DEVELOPMENTS IN SURVEYING TECHNOLOGY AND THEIR APPLICATION TO ENGINEERING GEOLOGY

A thesis submitted for the degree of Doctor of Philosophy in the University of London

by

T J M Kennie
BSc, MAppSci, ARICS, MInstCES

Department of Geology
Imperial College of Science, Technology and Medicine,
London, SW7 2BP

March, 1991

CONTENTS Page No.

Abstract i

Acknowledgements ii

1. **INTRODUCTION** 1

1.1 The Nature of Engineering Geology 1
1.2 The Relationship of Engineering Geology to the Stages of a Civil Engineering Project 5
1.3 Scope and Structure of Thesis 8
1.4 References 10

2. **INDIRECT INVESTIGATION AND MEASUREMENT TECHNIQUES** 13

2.1 REMOTE INVESTIGATION TECHNIQUES 14
2.1.1 Satellite Remote Sensing 14
2.1.2 Airborne Remote Sensing 22
2.1.3 Geophysics 26

2.2 FIELD MEASURING TECHNIQUES 37
2.2.1 Satellite Positioning Systems 37
2.2.2 Aerial Photogrammetry 38
2.2.3 Close Range Photogrammetry 39
2.2.4 Ground Surveying 43
2.2.5 Invasive and Non Invasive Measuring Techniques 44
2.3 Conclusion 46
2.4 References 49

3. **DISPLAY AND SYNTHESIS OF ENGINEERING GEOLOGICAL DATA** 55

3.1 Introduction 56
3.2 Digital Terrain Modelling 58
3.2.1 Methods and Software 58
3.2.2 Applications 59
3.3 DTM's and GIS in Engineering Geology 60
3.3.1 The scope of DTM's in GIS 60

3.4 Spatial Information Systems 64
3.5 Future Developments of 3D GIS 65
3.5.1 Developments in Data Acquisition 65
3.5.2 Developments in Data Generalisation and Transfer Standards 67
3.6 Conclusions 68
3.7 References 69

4. **RECONNAISSANCE STAGE CASE STUDY - DETECTION OF SOLUTION FEATURES USING AIRBORNE REMOTE SENSING** 71

4.1 Background to Study 72
4.2 Definitions of Solution Subsidence Features 73
4.3 Previous Work 74
4.4 Data Acquisition 76
4.4.1 Objectives 76
4.4.2 Case Study Data Acquisition Periods 78
4.4.3 Optimum Time of Day for Data Acquisition 78
4.5 Test Sites 79
4.5.1 Swanley 79
4.5.2 Southover Heath 80
4.6 Data Analysis 80
4.6.1 Air Photo Interpretation 80
4.6.2 Thermal IR Linescanning 81
4.7 Critical Analysis 82
4.7.1 Radiometric Restoration and Correction 82
4.7.2 Selection of Optimum Band Combination 85
4.7.3 Practical Detection Difficulties 86
4.8 Future Work 87
4.8.1 Objectives 87
4.8.2 Methodology and Programme 88

4.9 Conclusions 91
4.10 References 92

5. **INVESTIGATION DESIGN STAGE CASE STUDY - APPLICATIONS OF THERMAL VIDEO FRAME SCANNERS (TVFS)** 93

5.1 Background to Study 94
5.2 Previous Work 94
5.3 Data Acquisition 96
5.3.1 Design of TVFS 96
5.3.2 Operation of Sensor in Light Aircraft 97
5.4 Thermal Mapping of Ground Conditions 99
5.4.1 Hunterston Example 99
5.4.2 Lothian Example 99
5.5 Data Analysis 102
5.5.1 Hunterston Example 102
5.5.2 Lothian Example 102
5.6 Critical Analysis of Method 105
5.7 Conclusions 106
5.8 References 106

6. **FIELD MEASURING DESIGN STAGE CASE STUDY - MONITORING OF SLOPE INSTABILITY USING CLOSE RANGE PHOTOGRAMMETRY** 108
6.1 Background to Study 109
6.2 Analytical Photogrammetric Instrumentation 109
6.2.1 Officine Galieo Digicart 110
6.3 Monitoring of Erosion 110
6.3.1 Brighton Region - Embankment Monitoring 110
6.3.2 Lewes Area - Road Tunnel 112
6.4 Critical Analysis of Method 113
6.5 Conclusions 114
6.6 References 114

7. CONSTRUCTION STAGE CASE STUDY - MONITORING GROUND MOVEMENT DURING BRIDGE CONSTRUCTION 116

7.1 Background to Study 117
7.2 Data Acquisition 119
7.2.1 Geotechnical Methods 119
7.2.2 Surveying Methods 120
7.3 Results and Analysis of Deformation Monitoring 123
7.3.1 Column Deflections/Settlement During Backfilling 123
7.3.2 Deformation of Abutment During Construction 126
7.4 Critical Analysis of Project 133
7.5 References 135

8. POST CONSTRUCTION STAGE CASE STUDY - USE OF AN ELECTRONIC DISTANCE MEASURING INSTRUMENT FOR EARTH VIBRATION STUDIES 136

8.1 Background to Study 137
8.2 Previous Work 137
8.3 Data Analysis 138
8.3.1 Instrument Characteristics 138
8.4 Test Procedure 139
8.4.1 Sampling Requirements 140
8.4.2 Experimental Details 140
8.5 Results and Analysis 142
8.5.1 Model 1 - Single Pendulum 143
8.5.2 Model 2 - Variable Frequency Vibrator 143
8.5.3 Model 3 - Double Pendulum 144
8.5.4 Low Cost Retro-reflectors 145
8.6 Critical Analysis and Future Work 147
8.7 References 148

9. CONCLUSIONS 149

9.1 Introduction 150
9.2 Summary of Main Contributions and Recommended Use of Surveying Technology for the Solution of Engineering Geological Problems 150
9.2.1 Reconnaissance Stage 156
9.2.2 Feasibility Stage 157
9.2.3 Design Stage 158
9.2.4 Construction Stage 160
9.2.5 Post Construction Stage 161
9.3 Problems Associated with the Introduction of New Surveying Technology into Engineering Geology 163
9.3.1 Financial Investment 163
9.3.2 Black Box Technology 163
9.3.3 Practical Difficulties 164
9.3.4 Incompatibility 164
9.3.5 Training 165
9.4 References 165

APPENDICES - SUPPORTING PUBLICATIONS

Page No

APPENDIX A: DEVELOPMENTS IN AIRBORNE AND SATELLITE REMOTE SENSING AND APPLICATIONS TO ENGINEERING GEOLOGICAL INVESTIGATIONS A1

[1]. KENNIE T.J.M. 1985. An introduction to remote sensing. Chapter 2 in Remote Sensing in Civil Engineering, Kennie T J M and Matthews M C (Eds.), Surrey University Press, 356pp. A2

[2]. KENNIE T.J.M 1985. Remote sensing scanning systems. Chapter 3 in Remote Sensing in Civil Engineering, Kennie T J M and Matthews M C (Eds.), Surrey University Press, 356pp. A15

[3]. KENNIE T.J.M 1989. Remote sensing for topographic and thematic mapping. Chapter 6 in Engineering Surveying Technology, Kennie T J M and Petrie G (Eds.), Blackie Publishers, 485pp. A61

[4]. KENNIE T.J.M, MATTHEWS M.C. and BEAVEN P.J. 1985. Remote sensing in civil engineering practice. Chapter 5 in Remote Sensing in Civil Engineering, Kennie T J M and Matthews M C (Eds.), Surrey University Press, 356pp. A118

[5]. KENNIE T.J.M. and EDMONDS C.N. 1988. The location of potential collapse features in soluble carbonate rocks by remote sensing techniques. ASTM Special Technical Publication, on Geotechnical Appl. of Remote Sensing and Remote Data Transmission. Johnson I A and Pettersen B, 99-121. A131

[6]. KENNIE T.J.M, ROSENBAUM M.S. and EDMONDS C.N. 1988. Analysis of multitemporal airborne remote sensing for the location of solution features in the chalk of Southern England. 16th Congress of the Int. Soc. of Photgramm. and Remote Sensing, Kyoto, Japan. **Int. Archives of Photogram. and Remote Sensing,** Vol.27, Part B7, Commission VII, 323-332. A155

[7]. KENNIE T.J.M, DALE C.D. and STOVE G.C. 1986. A preliminary assessment of an airborne thermal video frame scanning system for environmental engineering applications. ISPRS Commission VII Int. Symp. Enschede, The Netherlands. **Int. Archives of Photogram. and Remote Sensing,** Vol. 26 Part 7/1, Balkema Press, 285-290. A165

[8]. STOVE G.C, KENNIE T.J.M. and HARRISON L. 1987. Airborne thermal mapping for highway maintenance engineers using the Barr and Stroud IR18 thermal video frame scanner. **Int. Journal of Remote Sensing,** Vol.6, 7, 350-357. A172

APPENDIX B: DEVELOPMENTS IN AERIAL AND GROUND BASED SURVEYING AND THEIR APPLICATION TO MONITORING CHARACTERISTICS OF ENGINEERING GEOLOGICAL IMPORTANCE

[9]. KENNIE T.J.M. 1989. Photogrammetry and Remote Sensing Chapter 7, **The Civil Engineer's Reference Book,** Blake L S (Ed.), 30pp. B181

[10]. KENNIE T.J.M. and McKAY M. 1986. Monitoring of geotechnical processes by close range photogrammetry. **Planning and Engineering Geology,** McCann P and O'Hara M (Eds.), Engineering Geology Special Publication, No 4, Geological Society, 163-170. B210

[11]. KENNIE T.J.M. 1987. Surveying. The Ground Engineer's Reference Book, Chapter 19, 25pp. B218

[12]. KENNIE T.J.M. 1989. Electronic angle and distance measurement. Chapter 1 in Engineering Surveying Technology, Kennie T J M and Petrie G (Eds.), Blackie Publishers, 485pp. B244

[13]. KENNIE T.J.M. 1984. The use of acrylic retro-reflectors to monitor the deformation of a bridge abutment. Civil Engineering Surveyor, Vol.9, 6, 10-15. B285

[14]. KENNIE T.J.M. 1983a. Some tests of retro-reflective materials for electro-optical distance measurements (Part 1). Survey Review, Vol.27, January, 3-12. B289

[15]. KENNIE T.J.M. 1983b. Some tests of retro-reflective materials for electro-optical distance measurements (Part 2). Survey Review, Vol.30, April, 67-81. B299

[16]. KENNIE T.J.M., HOWARD C. and HORKAN S. 1989. A preliminary assessment of the Tellurometer MA-200 for vibration measurement of flexible structures. Survey Review, Vol.30, April, 67-81. B310

[17]. KENNIE T. J. M., BRUNTON S., PENFOLD A. and WILLIAMS D. 1988. EDM Instruments: Calibration Methods and Facilities in the U.K. Land and Minerals Surveying, Vol.6, 27-35. B325

APPENDIX C: DEVELOPMENTS IN MODELLING SURFACE GEOLOGY AND TOPOGRAPHY BASED ON SURVEY DATA

[18]. PETRIE G. and KENNIE T. J. M. 1987. Terrain modelling in surveying and civil engineering. Computer Aided Design, Vol.4, 2, 177-189. B335

[19]. KENNIE T. J. M. and McLAREN R. 1988. Modelling for digital landscape and terrain visualisation. The Photogrammetric Record, Vol.XII, No 72, 711-745. B352

APPENDIX D: COMPLETE LISTING OF STATEMENTS SUMMARISING CONTRIBUTION OF AUTHOR TO JOINTLY AUTHORED PAPERS B387

APPENDIX E: BIBLIOGRAPHY OF WORK PUBLISHED DURING PERIOD OF REGISTRATION FOR PHD B400

ACKNOWLEDGEMENTS

The author wishes to thank the following individuals and institutions for their help and assistance during the preparation of this thesis:-

(a) Dr M.S. Rosenbaum, (Research Supervisor) of the Dept. of Geology, Imperial College for his guidance and support throughout my period as a part-time research student at the College.

(b) All of the authors who contributed towards the production of the papers listed in Appendices A to C, and in particular Professor G. Petrie of the University of Glasgow for the opportunity to work together on a number of projects and publications.

(c) The University of Surrey for providing the facilities and encouragement to pursue the research programmes discussed in the thesis and for the financial support towards th payment of academic fees.

(d) Maureen Albert for typing the thesis, and

(e) Last, but by no means least my family including my mother and father, my wife Sue and my two children Tamsin and Stuart for their patience and understanding during the preparation of this thesis.

DEVELOPMENTS IN SURVEYING TECHNOLOGY AND THEIR APPLICATION TO ENGINEERING GEOLOGY.

by

T J M KENNIE

Abstract

This thesis is concerned with an examination of problems concerning remote investigation and field based measurement techniques in engineering geology using modern land surveying methods. There are three distinct sections to the thesis and each is supported by a number of published papers.

The first section (Chapter 1) discusses the nature of engineering geology and reviews the most significant activities performed by professionals in this field at each stage of a civil engineering project. This structure is then further developed to illustrate the relationship between engineering geology, the Case Studies discussed in section 3 and the construction process.

The second section (Chapters 2 and 3) reviews the techniques which can be used in engineering geology to acquire, synthesise and display spatial data. In Chapter 2 the distinction is drawn between remote investigation techniques (remote sensing; from airborne and spaceborne platforms, and geophysics) and field measuring techniques (including ground and satellite positioning techniques and photogrammetry). The relative merits of each technique are outlined together with recommendations for their use during the construction process. Chapter 3 subsequently reviews the growing influence of computer hardware and software for the synthesis and display of engineering geological data. The current status of Digital Terrain Modelling (DTM) and Geographic Information Systems (GIS) are discussed.

Section 3 of the thesis (Chapters 4 to 8) contains the main scientific contributions and is presented in the form of a series of linked Case Studies. The solutions to specific engineering geological problems are presented together with a critical review of the methods used and results obtained. In a number of cases recommendations for future work are also presented. The problems addressed in section 3 include:-

(a) the detection of solution features in Chalk (Chapter 4)
(b) the thermal mapping of ground conditions (Chapter 5)
(c) the monitoring of slope instability (Chapter 6)
(d) the monitoring of ground movement (Chapter 7), and
(e) the measurement of earth vibration (Chapter 8).

The principal conclusions of the thesis are presented at the end of each chapter and summarised in Chapter 9. The main conclusions indicate that as far as remote sensing in the United Kingdom is concerned airborne sensors are more appropriate than

satellite sensors. Their use is also restricted primarily to the reconnaissance stage of a project. Ground based instrumentation is, however, more appropriate for monitoring, for example, the effect of erosion or the impact of construction on the underlying geology. For the future the use of computer based spatial information systems will become of increasing importance.

CHAPTER 1
INTRODUCTION

1. INTRODUCTION

This thesis is concerned with an examination of problems in engineering geology which involve remote investigation and field based surveying techniques. The nature of surveying technology has changed dramatically during the past 10 years and increasingly the instrumentation used is electronic in design and the output digital. In addition, computer processing of the acquired data has become an integral part of the overall data analysis and presentation process. Such changes in technology have had a considerable impact on the range of problems in engineering geology which may now be tackled using surveying techniques.

The thesis is presented using selected works published by the author during the period of registration for the Ph.D. The papers are appended in support of the thesis and a full bibliography of publications may be found in Appendices A to E.

1.1 THE NATURE OF ENGINEERING GEOLOGY

Attewell and Farmer (1976) identified 'Engineering Geology' as 'one of those terms that invite definition!'. Many subjects evolve from an interdisciplinary background (in this case geology and civil engineering) and it is often in such situations that the greatest difficulties of definition arise. Johnson and De Graff (1988) define engineering geology simply as 'the application of geology to engineering practice'. The American Geological Institute offer a more precise definition as 'the application of geological sciences to engineering practice for the purpose of assuring that the geological factors affecting the location, design, construction, operation and maintenance of engineering works are recognised and adequately provide for'.

The International Association of Engineering Geology proposed the following definition for Article 2 of its statute at its 1981 Amsterdam meeting: 'Engineering Geology is the science devoted to the investigation, study and solution of the engineering and environmental conditions resulting from the interaction of the work and activities of man on the geological environment and of the occurrence of geologic hazards'.

As a discipline engineering geology draws upon all the geological subjects notably petrology, sedimentology, stratigraphy, structural geology, geomorphology and hydrogeology. Considerable debate exists as to the extent of the boundaries between these disciplines and indeed some practitioners would argue that the perception of engineering geology as a primary discipline is inappropriate. For the purpose of this thesis the broad definition of engineering geology as used by the Engineering Group of The Geological Society of London for the purpose of accepting papers for inclusion within the Quarterly Journal of Engineering Geology will be used. The Engineering Group considers engineering geology to be 'concerned with the subject of geology as applied to civil engineering, mining practice and water resources including rock mechanics, soil mechanics and geotechnics, applied sedimentology, pedology, mineralogy and petrology, geohydrology and hydrology, engineering seismology and engineering applications of geophysics'.

Engineering geology may therefore be considered to be the broad umbrella term which interfaces with, and to some extent includes,

a) Soil and Rock Mechanics - the sub-disciplines which deal with the mechanical description of the strength and stress/strain behaviour of engineering soils and rocks.

b) Geotechnical Engineering - the sub-discipline defined by the American Society of Civil Engineers in the Journal of Geotechnical Engineering as '.... the relationship between the geologic environment and man made works; including foundations, earth retaining structures, earth supported structures, dam construction and slope stability as well as soil failures and their solutions'.

The distinction in the United Kingdom between the disciplines of engineering geology and geotechnical engineering can to some extent be traced to the educational development of the subjects. Engineering geologists often gain their initial training in Higher Education Departments of Geology whereas the educational development and research in geotechnical engineering has generally originated from Departments of Civil or Mining Engineering. Consequently the geotechnical engineer has tended to develop a more significant role in the design stage of civil engineering projects whereas the engineering geologist's role has developed more significance at the preliminary investigation stages prior to development, or in the post-construction monitoring of structures. The immediate pre-design stage of a project generally referred to as the 'site investigation' is primarily the domain of the engineering geologist.

c) Hydrology - A range of sub-disciplines associated with the interaction of water, man made structures and geology. Hydrogeology is generally concerned with groundwater conditions, hydrology (ss) with water supply, and hydraulics with river engineering and beach/coastal engineering, where the emphasis is more concerned with the civil engineering problems associated with river and sea defences.

A further term which is gathering some acceptance in the United Kingdom is 'Ground Engineering' (Bell 1987). The distinction between this more recently coined alternative to engineering geology and geotechnical engineering is not immediately apparent. Its main attraction would appear to be in the combining of the geotechnical engineers and engineering geologists areas of expertise, and hence help to minimise the polarising separation of the subjects which has become more apparent in recent years.

1.2 THE RELATIONSHIP OF ENGINEERING GEOLOGY TO THE STAGES OF A CIVIL ENGINEERING PROJECT

The role which engineering geology has to play in the development and construction process varies quite considerably at each stage of a civil engineering project. In broad terms five distinct stages of a civil engineering project can be identified and these are illustrated by Figure 1.1.

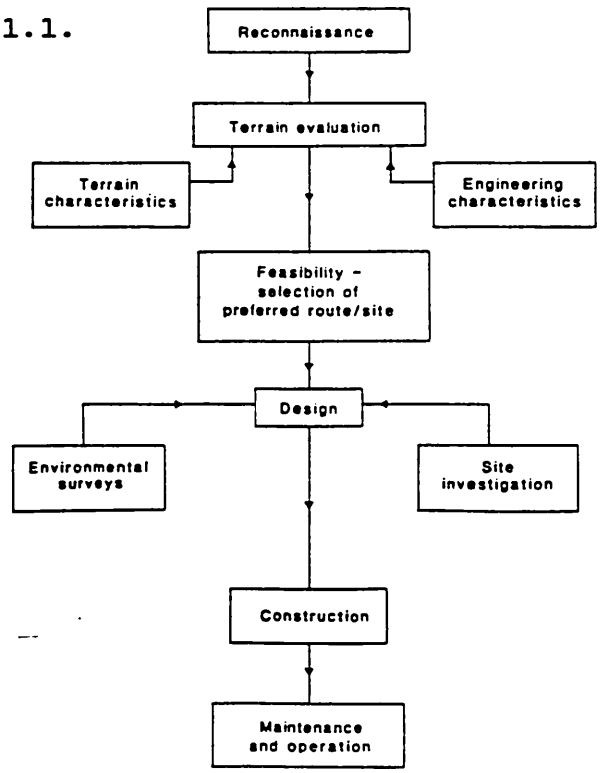


Figure 1.1
The Stages of a Civil Engineering Project (Kennie, 1989)

These stages are well established and the engineering geological involvement at each stage can be broadly stated as follows:

a) Reconnaissance Stage

At this stage the primary objective is the identification of possible alternative routes and sites. It will generally involve an investigation of existing information in the form of topographic and thematic maps (geological, hydrological, etc), aerial photographs and engineering project reports.

b) Feasibility Stage

The aim of the feasibility study is the selection of a preferred route or site representing the best available option from those identified in the Reconnaissance Stage. It may involve the use of structured terrain evaluation techniques to specify the areas of particular importance. A comprehensive report on the use of such techniques can be found in Edwards et al (1982).

c) Design Stage

At the design stage the role of the engineering geologist will be to carry out more detailed site investigations along the route specified or at the site identified during the feasibility study. Two forms of data can be of value at this stage.

The first is associated with the formal site investigation process. Data at this stage can be obtained about:-

- the site history and past land use to identify potential areas of instability or, for example,

areas previously covered by forestry.

- the location of features such as: abandoned mine shafts, infilled trenches, solution features and other localised features which may hinder construction.
- areas of instability caused by poor drainage.
- the location of fill aggregates and other construction materials.

These data represent some of the more common requirements during site investigation although other information may also be required during this phase.

The second form of data required during the design stage of a civil engineering project is that associated with what can be termed 'environmental' data. As the environmental damage caused by construction becomes more of an issue this stage is becoming increasingly important. Such environmental surveys may involve a comprehensive environmental impact assessment or may be restricted to an assessment of the visual impact of a particular design. It may also involve the identification of zones containing specific types of vegetation which are to be saved from the damaging effects of construction. Finally it may involve monitoring the effect of the environment on an existing structure or feature as a precursor to further design work. Two of the case studies which are reported in Chapters 5 and 6 illustrate this approach. In the first case (Chapter 5) the environmental survey provides data to aid in the design of highway maintenance operations. In the second case (Chapter 6) the environmental monitoring assists in the design of remedial measures to minimise the influence of slope instability.

d) Construction Stage

The engineering geologist's role during the construction phase of a civil engineering project is somewhat reduced by comparison with the previous three stages. There are however specific circumstances in which the data provided by an engineering geologist can be of value. Firstly the engineering geologist may be able to monitor the rate of construction progress in zones where remedial modifications to the original design may be required (e.g. areas with steep slopes where slope instability may be a problem). Secondly data may be required to monitor the influence of construction on the geology of an area. The case study discussed in Chapter 7 illustrates this type of problem.

e) Post Construction Stage

The post construction/maintenance stage of a civil engineering project also requires an understanding of the role of the engineering geologist. At this stage field monitoring may also be required, for example as in the case study reported in Chapter 8 to monitor the impact of earth vibration. Remote monitoring of slopes or examination of areas of instability may also form part of the engineering geological investigations at this stage.

1.3 SCOPE AND STRUCTURE OF THESIS

The previous two sections have outlined the nature of engineering geology and its relationship to the various stages of a civil engineering project. At all stages one of the most significant characteristics of the service provided by the engineering geologist is in the provision of detailed interpretative data (often provided by a remote investigation procedure) or field based measurement data.

Historically these data have been provided by the interpretation of aerial photography or the use of simple surveying techniques. This thesis is concerned with the potential benefits of utilising modern land surveying technology in the engineering geological investigations at the various stages of a civil engineering project identified in Section 1.2.

The thesis considers two distinct forms of surveying techniques. The first, which may be referred to as remote investigation techniques includes the use of remote sensing (both from satellites and aircraft) and geophysics. In this case the emphasis is on the thematic mapping of data relevant to the engineering geologist, generally during the Reconnaissance, Feasibility and Design Stages of a project. The second form of surveying technique may be referred to as field based measurement techniques. This category includes the traditional ground based surveying techniques using optical and electronic measuring devices together with the precise mensuration data which can be provided by photogrammetry (both aerial and close range) and satellite position fixing systems (primarily from the Global Positioning System). Chapter 2 of the thesis presents a review of these various surveying techniques.

A second role of engineering geology during a civil engineering project is to synthesise data from many sources and provide a critical analysis of such data. The processing techniques developed in the land surveying community can aid the engineering geologist in this process and Chapter 3 reviews the current trends in this field which may be of benefit to the engineering geologist.

The following five Chapters (4-8) of the thesis illustrate the manner in which the engineering geologist may use surveying techniques and technologies at the different stages of a project and Table 1.1 illustrates the interrelationship between the surveying technologies, the

stages of a civil engineering project and some typical engineering geological problems which may be addressed using land surveying technology. Table 1.2 further illustrates the relationship of the Case Studies discussed in Chapters 4 to 8 to the structure developed in Table 1.1.

The final Chapter (9) critically considers the nature of surveying technology in engineering geology and provides recommendations on the use of surveying during the various stages of a civil engineering project.

REFERENCES

ATTEWELL F.G. and FARMER I.W. 1976. Principles of Engineering Geology. Chapman and Hall, London, 1045pp.

BELL P.B. (editor) 1987. Ground Engineers Reference Book Butterworths.

EDWARDS R.J.G. (editor) 1982. Land Surface Evaluation for Engineering Practice: Report by a Working Party under the auspices of the Geological Society. Quarterly Journal of Eng. Geol, Vol. 15, pp 265-316.

JOHNSON R.B. and DE GRAFF J.V. 1988. Principles of Engineering Geology, Wiley, New York, 497 pp.

KENNIE T.J.M. 1989. Remote Sensing for Topographic and Thematic Mapping. Chapter 6 in Engineering Surveying Technology, Kennie TJM and Petrie G. (Eds), Blackie & Son, pp181 - 237.

TABLE 1.1: THE CONTRIBUTION OF ENGINEERING GEOLOGY DURING THE STAGES OF A CIVIL ENGINEERING PROJECT

	Engineering Geology	
	Remote Investigation Techniques	Field Measuring Techniques
Reconnaissance/Feasibility	Regional Terrain Evaluation	Regional subsidence/ seismographic studies
Design	Site Specific Geophysical and Remote Sensing Investigations	Ground investigation characteristics. Foundation design.
Construction	-	Ground improvement and consolidation
Post-Construction	-	Monitoring of ground conditions and stability of slopes, etc

TABLE 1.2: THE RELATIONSHIP OF THE CASE STUDY MATERIAL TO THE STAGES OF A CIVIL ENGINEERING PROJECT

Stage of Civil Engineering Project	Engineering Geology	
	Remote Investigation Techniques	Field Measuring Techniques
Reconnaissance/Feasibility Stage	Chapter 4 Detection of Solution Features Using Airborne Remote Sensing	-
Design Stage	Chapter 5 Applications of Thermal Video Frame Scanners (TVFS)	Chapter 6 Monitoring of Slope Instability Using Close Range Photogrammetry
Construction Stage	-	Chapter 7 Monitoring Ground Movement During Bridge Construction
Post-Construction Stage	-	Chapter 8 Use of an Electronic Distance Measuring Instrument for Earth Vibration Studies

CHAPTER 2

INDIRECT INVESTIGATION AND MEASUREMENT TECHNIQUES

2. INDIRECT INVESTIGATION AND MEASUREMENT TECHNIQUES

The engineering geologist requires a wide range of data about the earth's surface and its relationship to engineering projects. These data range from countrywide investigations concerned with regional geological trends to laboratory based measurements of soil samples. Clearly the instrumentation needed to perform these various analyses will differ considerably. Land Surveying technology is increasingly being used by the engineering geologist to satisfy these data needs at scales varying from regional reconnaissance investigations to post construction monitoring studies. The objective of this chapter is to review the various techniques currently available and illustrate the role of surveying techniques within the broader context of the other indirect investigation and measurement techniques available to the engineering geologist. Table 1.1 in Chapter 1 illustrated the contribution of engineering geology during the stages of an engineering project. Table 2.1 is a modification of this to illustrate the role of the various indirect investigation and measurement techniques.

2.1 REMOTE INVESTIGATION TECHNIQUES

A considerable number of techniques have been developed over the recent past to investigate the geological and environmental characteristics of the earth's surface relevant to the engineering geologist. Three techniques of particular relevance for such investigations are remote sensing (from spaceborne and airborne platforms) and ground based geophysics.

2.1.1 Satellite Remote Sensing

The technology of satellite remote sensing and its application to civil engineering projects has been extensively reviewed in the following publications by the author - Kennie [2], Kennie [3] and Kennie et al [4].

TABLE 2.1: CONTRIBUTION OF INDIRECT INVESTIGATION AND FIELD BASED MEASUREMENT TECHNIQUES DURING THE STAGES OF A CIVIL ENGINEERING PROJECT

Stages of Project	Engineering Geology		
	Remote Investigation Techniques	Field Based Measurement Techniques	
Reconnaissance/ Feasibility	Satellite Remote Sensing, Airborne Geophysics	Aerial Photogrammetry (small scale)	Satellite Position Fixing
Design	Airborne Remote Sensing, Ground Geophysics	Aerial and Close Range Photogrammetry	-
Construction	-	Close Range Photogrammetry	High Precision Field Surveying
Post-Construction	-	High Precision Deformation Studies. Ground Vibration Studies. Close Range Photogrammetry.	-

These publications considered the applications of the sensors summarised in Table 2.2, 2.3 and 2.4 for reconnaissance and feasibility scale topographic and thematic mapping.

a) Reconnaissance Scale Topographic Mapping

As far as topographic mapping from spaceborne sensors is concerned it is possible using the accuracy requirements for standard map scales produced by Doyle (1982) to conclude that only imagery from the French SPOT sensor can meet the accuracy demands for 1:50,000 mapping (Table 2.5). The other sensors are either unsuitable for mapping at this scale or can only provide more limited coverage at smaller map scales.

b) Thematic Mapping

In spite of the limited capacity of satellite remote sensing to meet the very high accuracy demands of topographic mapping in the field of thematic mapping it has proved much more successful. The advantages of such imagery in these cases include:-

i) Coverage of large areas. For example each Landsat MSS scene covers an area of some 33,000km . Since the first launching in 1972, the Landsat satellites have recorded all the land surface of the globe except those areas persistently covered by cloud.

ii) Repetitive cover. Many zones have been recorded more than once, enabling the interpreter to compare views of the terrain taken in different years, or at different seasons. Such comparisons are valuable for example in regional hydrological studies e.g. monitoring seasonal changes to river courses.

iii) Data are computer compatible. Apart from the photographic missions the data can be digitally

TABLE 2.2: CHARACTERISTICS OF A SELECTION OF SPACEBORNE PHOTOGRAPHIC MISSIONS

	NASA Skylab S-190B	ESA Metric Camera	NASA Large Format Camera
Focal Length (cm)	46	30	30
Format (cm)	12.5 X 12.5	23 X 23	23 X 46
Coverage (km)	110 X 110	190 X 190	225 X 450
Scale	1:950,000	1:820,000	1:1,000,000
Flying Height (km)	385	250	300
Planimetric Accuracy (m)	20 - 25	20 - 30	20 - 30

TABLE 2.3: CHARACTERISTICS OF A SELECTION OF SATELLITE SCANNERS

	Line Scanners		Pushbroom Scanners	
	Landsats 1-5 Multispectral Scanner (MSS)	Landsats 4,5 Thematic Mapper (TM)	Modular Optoelectronic Multispectral Scanner (MOMS)	Le Systeme Probatoire d'Observation de la Terre (SPOT)
Operated by (country)	EOSAT (USA)	EOSAT (USA)	DFVLR (West Germany)	SPOT Image (France)
Date of Launch	1972	1982,84	1983	1986
Orbital Altitude (KM)	900	705	300	830
Ground Resolution (m)	80	30	20	10, 20
Spectral Range (m)	0.5-12.6	0.45-12.5	0.575-0.975	0.5-0.89
Number of Wavebands	5	7	2	3

TABLE 2.4: CHARACTERISTICS OF A SELECTION OF SATELLITE BASED SAR SYSTEMS

	Past			Future	
	Seasat	Shuttle Imaging Radar (SIR)-A	SIR-B	European Radar Satellite (ERS-1)	Radarsat
Date of Launch	1978	1981	1984	1991	1991
Operated by	NASA	NASA	NASA	European Space Agency	Canada
Altitude (KM)	800	245	255, 274, 352	675	1000
Wavelength (mm)	230	230	540	540	540
Look angle (o)	20	50	57	-	30-45
Resolution (m)	25	40	30	25	30
Swath Width (KM)	100	50	20-25	80	150

TABLE 2.5: SUITABILITY OF VARIOUS SATELLITE SENSORS FOR TOPOGRAPHIC MAPPING

	Photographic Sensors		Scanning Sensors		Synthetic-aperture radar
	Metric Camera	Large Format Camera	Landsat Thematic Mapper	SPOT	Seasat SIR-A/B
Plan Accuracy	***	***	**	***	*
Height Accuracy	-	-	-	**	-
Map Compilation	**	**	*	**	*
Map Revision	***	***	*	****	*
Map Intelligence	***	***	***	****	**

\- Unsuitable
* Suitable for 1:250,000 mapping
** Suitable for 1:1,000,000 mapping at a relaxed specification (70% detail)
*** Suitable for 1:100,000 mapping
**** Suitable for 1:50,000 mapping

processed by computer in order to produce images that are easier to interpret, thereby improving the quality of the interpretation.

iv) Data are widely available.

v) Low Cost. The cost of satellite data depends largely on whether the image is examined on photographic film or paper, or on an image processing system. In the latter case the costs are obviously higher since access is required to both the computer compatible tapes and an image processing system. In both cases, however, the costs are relatively small when compared with the benefits which can be obtained from interpretation of the imagery.

In spite of the significant benefits which satellite remote sensing imagery offers to the engineering geologist its use is nevertheless restricted. The following are some of the main limitations of the imagery:-

i) Spatial resolution. The present generation of SPOT satellites offer a spatial resolution of 10m. Whilst this is a significant improvement over the early Landsat MSS imagery (80m) it is nevertheless only suitable for basic planimetric mapping at scales of less than 1:100,000. Map revision at scales up to 1:50,000 may be possible under limited circumstances. An associated disadvantage is the lack of or limited height information from the imagery which restricts its usefulness to the engineering geologist.

ii) Cloud cover. In spite of the repetitive nature of the satellite systems and the advent of SAR (Side Looking Radar) there are still several areas of the world, including the U.K, for which there is only

restricted coverage because of the problems of cloud coverage.

iii) Ground verification. In order to be able to make maximum use of the satellite data it is important that ground data are available both to improve the interpretation and also to verify the results which are obtained.

In spite of these difficulties satellite remote sensing has a valuable role to play in ground investigation, particularly overseas, and Table 2.6 demonstrates the recommended use of such data at the various stages of a civil engineering project.

2.1.2 Airborne Remote Sensing

a) Aerial Photography

Traditionally the most common form of airborne sensor to be used for engineering geological investigations has been the aerial camera. In general the imagery acquired has been of relatively large scale (1:1000 - 10,000) large format (23 x 23 cm), obtained with the camera axis pointing vertically and with a panchromatic emulsion. While small format photography may also be used (e.g. Tomlins, 1983, Graham and Read, 1984) or oblique photography acquired (Matthews and Clayton, 1988) the vast majority of photography utilised for engineering geological site investigations has been acquired using aerial photogrammetric cameras. One of the more common modifications to this approach has involved the use of multiple emulsion types in which the photographic emulsion is sensitised to a specific part of the EM spectrum (e.g. the near infrared). Several studies have been reported on the use of this approach (Beaumont, 1977) in some cases involving the use of multiple small format cameras (Heath, 1980). Table 2.7 summarises the relative advantages and limitations of the different emulsion types.

ENGINEERING PROJECT

Project Phase	Photograph/ Image Scale	Satellite					
		Photography	Landsat MSS	Landsat TM	SPOT	MOMS	Side-Looking Radar
Reconnaissance/ Planning	1:500,000 to 1:50,000	***	****	***	****	***	***
Feasibility	1:50,000 to 1:10,000	***	**	***	***	***	***
Design	1:10,000 to 1:2,500	-	-	*	*	-	-
Construction	1:2,500 to 1:500	-	-	-	-	-	-
Post-Construction Maintenance	1:5,000 to 1:500	-	-	-	-	-	-

**** Very Useful *** Very useful (but coverage limited) ** Useful
* Of limited use - inadequate

Characteristics of various photographic emulsions

Film type	Structure	Sensitivity	Advantages	Limitations	Cost
Black-and-white panchromatic	Single-layer emulsion	0.4–0.7 μm	Good definition and contrast, wide exposure latitude. Boundaries are often clearly portrayed. Subtle textural variations can be identified and geometric patterns are more obvious	Sometimes difficult to interpret ground features due to the inability of the eye to distinguish between subtle differences in grey tones	Low
Black-and-white infrared	Single-layer emulsion	0.4–0.9 μm	It provides specific enhancement of different forms of vegetation and clearly defines bodies of water and ground moisture distribution. Helps to penetrate haze, giving a sharper more contrasting image	Contrast may be excessive and detail is often lost in areas of shadow	Low
True-colour	Multi-layered emulsions. Top layer is sensitive to blue light, the 2nd layer to green and blue light and the 3rd to blue and red light. A blue absorbing filter is introduced between the 1st and 2nd layers.	0.4–0.7 μm	Good contrast and tonal range. Wide exposure latitude for negative film types. Materials are more identifiable and hence moisture conditions are also more identifiable	Less good definition than panchromatic film	Intermediate to high
False-colour infrared	Multi-layered emulsion incorporating a reflected infrared sensitive layer. Top layer is sensitive to blue and reflected infrared light, 2nd to green and blue light and 3rd to red and blue light. The film is normally used with a filter to block out blue light.	0.4–0.9 μm	Helps to penetrate haze, provides accurate identifiable data on vegetation, rocks, soils, water bodies and moisture distribution	Lower resolution than colour film. Exposure difficult to determine and processing expensive	High

TABLE 2.7 (Characteristics of photographic Emulsions (Kennie (1985))

b) Thermal Infrared Scanners (IR)

One of the main limitations of aerial photography for remote geological investigations is the inability to record variations in EM response outside of the visible and near IR regions of the spectrum. To record longer wavelength emitted radiation as opposed to shorter wavelength passively reflected IR radiation, it is necessary to utilise a device in which the variations in thermal radiation are recorded by variations in the response of a semi-conducting crystal (generally mercury cadmium telluride). The main characteristics of the more common thermal linescanners are discussed in detail in Kennie [3] pages A72-73. An alternative to thermal linescanners are thermal framescanners. Such devices whilst not designed specifically for civilian remote sensing applications have been adapted from their role as military night vision systems to this application. The design characteristics and application of such a device, the Barr and Stroud IR18 are discussed in Kennie [3] page A78 and also in Chapter 5.

c) Airborne Multispectral Scanners and Imaging Spectrometers.

Airborne multispectral linescanners (MSS) follow the same broad principle of operation to the sensors on board the Landsat satellites discussed in section 2.1.1. One of the more common designs in use in the UK is the Daedalus 1268 Airborne Thematic Mapper (ATM) which is discussed in Kennie [3] page A77.

A recent variant of the MSS is the imaging spectrometer. These generally operate in a 'pushbroom' mode of operation. Potentially they offer considerable flexibility in data acquisition by providing information on radiation response in many narrow contiguous wavebands.
The Moniteq Programmable Multispectral Imager (PMI)

discussed in Kennie [3] pages A84-86 can record up to 288 wavebands.

d) Side Looking Airborne Radar (SLAR)

SLAR Systems have been flown on a number of occasions over parts of the UK. They offer the advantages of recording terrain response to actively generated microwave radiation and the opportunity to acquire imagery in adverse weather conditions and/or at night. The characteristics of SLAR are documented in Kennie [3] pages A91-94.

2.1.3 Geophysics

Geophysics is concerned with the determination of the subsurface geological structure of the earth by measuring specific physical data either from an airborne platform, the ground surface or from boreholes. Whilst airborne geophysical techniques are widely used in oil and gas exploration they have not gained a wide audience in the field of civil engineering site investigations. Several reasons may be cited to account for this situation. One of the principal reasons is that several examples have been reported in the literature where poor or inconclusive results have been obtained. This is particularly true in the case of the detection of cavities in carbonate rocks. Although the low success rate of many of the techniques can be attributed to the failure of the geophysical technique to resolve the feature, the main reason for failure often stems from a lack of attention to initial planning and an awareness of the factors which influence the success, or otherwise, of a particular technique. Consequently this section will be restricted to a review of the role of ground based geophysical techniques.

In general terms geophysical surveys may be used in one of two roles in engineering geological investigations. Firstly, to enable a choice to be made rapidly and economically between a number of alternative sites which

are thought to contain features hazardous to construction i.e. as a reconnaissance tool, or secondly to supplement a boring programme at the detailed site investigation stage of a project.

In all cases, detection of the feature under investigation relies on a sufficient contrast in physical characteristics existing between the feature and the surrounding ground. Since the feature being investigated may be filled with air, water, soil, rock or a combination of these materials, the contrast with the in-situ rockmass will inevitably vary.

In view of the comments above the aim of this section of the thesis will be to provide a brief review of the literature which has reported the use of geophysics in engineering site investigations. To illustrate the role specific examples will be cited to indicate the use of the differing techniques for the detection of naturally occurring and man made subsurface features. Several reviews have been published on this topic (Fountain et al 1975, Greenfield 1979 and McCann et al 1982) and the following sections will summarise typical examples of previous work in this field.

It should be noted that the review will not provide a detailed explanation of all the various techniques. The reader is referred to Griffiths and King (1981) for further details.

a) Seismic Methods

Seismic survey methods are based on the generation of seismic waves, either on the surface or in a borehole, and the subsequent measurement of the time taken by the waves to travel from the source, through the rock mass to a series of receivers or geophones. The presence of a subsurface feature will tend to modify the way in which the seismic waves propagate by increasing the overall travel

time and reducing the amplitude of the seismic pulse. Thus, the geological discontinuity will manifest itself as a series of anomalous readings recorded in the vicinity of the feature. Several distinct methods of seismic surveying can be identified. These include seismic refraction, reflection, resonance and borehole seismic techniques.

i) Seismic Refraction

The seismic refraction method is widely used in site investigations to determine the depth of bedrock. Whilst several researchers have mentioned the use of the method for cavity detection few have reported any high degree of success. Ballard et al (1983), for example, reported the use of the technique during a site investigation for an arch dam in Honduras. The dam was founded on dolomite and limestone and many subsurface cavities were known to exist throughout the area. By using a 24 channel seismograph they were able to record the seismic refractions produced by a sledgehammer impact. The results obtained by the technique were 'as expected; no cavities were identified .... but major weakness zones between the two galleries could be detected'. The authors concluded that the technique should be used as a reconnaissance level technique rather than for detailed investigations.

A more successful case history is reported in McCann et al (1982). In this case the authors choose to examine the seismic response of an artificial cavity of known dimensions in order to develop techniques which could be applied to the detection of naturally occurring air filled cavities. The 'cavity' consisted of a disused railway tunnel 5 metres wide by 5 metres high situated at a depth of 14 metres below ground level. The survey was carried out using a 24 channel instrument with a falling weight of seismic source. The data obtained from the survey was analysed statistically by a least squares technique and the anomalous zones caused by changes in the superficial deposits and those associated with the presence of the

tunnel were identifiable.

ii) Seismic Reflection

The seismic reflection method is by far the most widely used geophysical technique for the determination of lithology and structure at great depths. Its use for engineering applications where one is concerned with features at "shallow" depths (<50 metres) has proved to be much less successful. One of the main reasons is the "trade off" which exists between depth penetration and spatial resolution. For example, in cases where the wavelength of the incident seismic energy is greater than the diameter of the feature being investigated, very little energy is reflected back to the surface and hence it will not appear on the seismic record. In addition, for shallow surveys, reflections from near surface interfaces may be totally obscured by large amplitude surface waves. In spite of these difficulties McCann et al (1982), have suggested that a "high frequency seismic source operating in the 700 to 1500 Hz bandwidth used in conjunction with common depth point processing has considerable potential."

Cook (1965) has also suggested that reflection methods may offer some benefits and suggests that the variation in amplitude attenuation can be used to delineate brine cavities in bedded salt at depths of up to 500 metres. Whilst several successful examples have been reported in the literature the technique of seismic reflection is seldom used in practice, primarily because it does not offer the high resolution required to detect shallow features. It is also a relatively expensive technique to employ.

(iii) Seismic Resonance

Seismic resonance is a relatively new technique based on some early work by Biot (1952). The exact mechanism by which seismic resonance occurs is not fully understood but it is thought to be caused by resonant oscillations being excited in the walls of the feature by the variable

frequency seismic source.

A commercial development of the seismic resonance principle has been used in the United Kingdom for a variety of cavity detection problems. The technique is known as SHRIMP (Savage and Heirli Resonant Investigation Method Patented). It is claimed by the distributors of the technique that a cavity one metre in diameter can be located at depths of up to 25 metres. Few successful case histories have however been reported in the literature. Arrowsmith and Rankilor (1981) refer to its use to locate mineshafts and workings along the proposed route of the Dalton By-pass. The results obtained were reported to have been inconclusive and the work was terminated.

(iv) Borehole Seismic Techniques

It has been suggested by McCann et al (1982) that surface geophysics can be of limited value in urban or redevelopment areas, both because of surface obstructions and also because the near surface geological strata may have been disturbed. One technique which overcomes these problems and which also can extend the range of the seismic technique is that of borehole seismography. Two distinct techniques can be identified which have been applied to subsurface cavity detection, interborehole acoustic scanning and the surface to borehole technique. Interborehole Acoustic Scanning surveys are carried out by transmitting an acoustic pulse through the rock mass between two adjacent boreholes and measuring the travel time of the acoustic pulse.

McCann et al (1982) also describe two case histories relating to the use of this technique for detecting cavities. The first relates to the detection of a disused railway tunnel. Two boreholes perpendicular to the tunnel and 10 metres from the centreline of the tunnel were used in order to obtain data. By determining the amplitude of the received pulse for a given wavelength they were able to produce amplitude "signatures" at different depths.

The second case history reported by McCann et al (1982) refers to a survey to investigate an area of disturbed ground below a residential dwelling, where it was believed a series of cavities existed. A similar procedure to that adopted in the previous case was used. Analysis of the results enabled the areal extent of an acoustic anomaly associated with cavity formation to be mapped.

The second approach is termed the 'surface to borehole technique' and is essentially an extension of the vertical seismic profiling method described by Kennet et al (1980). The technique consists of suspending a geophone at a known depth within a borehole. The seismic record is subsequently measured from a source positioned near the borehole. It is also common practice to position the source at various points around a circle of varying radius centred on the borehole.

(b) Magnetic Methods

Magnetic surveys are based on the measurement of local variations in the earths total magnetic field strength. Measurements of magnetic variation may be obtained using either a proton precession magnetometer or a fluxgate magnetometer. Both instruments are simple to operate and do not require accurate levelling or orientation. Whilst magnetometry has been used for the detection of both naturally occurring and man made cavities it has proved to have a much higher success rate when applied to the detection of man made cavities such as abandoned mineshafts. The main reason stems for the magnitudes of the magnetic anomalies associated with such features in comparison with naturally occurring subsidence features.

The use of magnetometry to locate natural cavities, such as solution features in chalk has also been examined by McDowell (1975). McDowell hypothesised that clay filled sinkholes would exhibit magnetic "highs" because the

magnetic susceptibility of clay is normally greater than that of chalk. He also suggested that "a magnetically clean site, free from metal objects is essential for success". The site in this case was located over the position of a proposed reservoir in Hampshire. A 5 metre grid of observations were recorded over the site using a proton precession magnetometer. Several magnetic anomalies associated with clay filled sinkholes were observed. In addition a magnetic traverse over a known sinkhole induced a magnetic "high" of almost 20 gammas.

(c) Electrical Resistivity Methods

This method of surveying relies on the measurement of variations in the terrains electrical resistance. There are many variations in the practical application of the technique, however, they all involve passing a current through the ground between a pair of "current electrodes" and measuring, with a resistivity meter, the resulting potential difference between a pair of "potential electrodes". By multiplying the value measured by the meter and the corresponding distance factor, it is possible to determine the "apparent resistivity" (pa) of the ground. Two distinct resistivity surveying techniques can be identified. The first, electrical sounding, enables vertical changes in resistivity to be measured by progressively moving the electrodes outwards with respect to a fixed central electrode. The second technique, resistivity profiling, enables lateral changes in resistivity to be detected by using a fixed electrode separation and moving the whole electrode array between each resistivity measurement.

A comparison of the results obtained by different electrode configurations over a known clay filled sinkhole is provided by McDowell (1975). A line between three boreholes was chosen for resistivity profiling. The profile obtained using a Wenner configuration indicated a small "low" masked by flanking highs over the clay filled

sinkhole.

Mention should also be made of the use of resistivity surveying for the location of abandoned mineshafts. Higginbotton (1976) specified two criteria that require to be met in order to enhance the chances of success when using this technique. Firstly, the workings must be above the water table and secondly, the workings should be open and at shallow depths. Unfortunately, this tends to restrict the use of the resistivity technique to working less than 5 metres below the surface.

Dearman et al (1977) also suggest that mineshaft detection may be possible using resistivity techniques. They do, however, view the requirement for even ground as a major limitation of the technique "particularly where the suspected shaft position is close to a pit heap".

(i) Inductive Conductivity Technique

A development of the electrical resistivity technique which has been applied to cavity detection is that of inductive conductivity. The technique involves measuring the conductivity of the terrain by means of a hand held instrument. The technique has several advantages over conventional resistivity techniques and these are discussed by McDowell (1981). McDowell also carried out a survey using such an instrument to locate sand lenses within Reading Beds in South Hampshire. The positions of such features has been suggested as being indicative of localised solution weathering of the underlying chalk and likely to be associated with cavities and sinkholes.

Zalasiewicz et al (1985), also reported on the application of ground conductivity measurements in general for geological mapping. Using the EM31 and EM34 instruments manufactured by Geonics they were able to delineate clay and clay rich sediments from sands, sandstones and limestones.

(d) Gravity Methods

Gravity surveying involves measuring variations in the earth's gravitational field. Measurements are obtained by means of extremely sensitive balances known as gravimeters. The resolution of most gravimeters is of the order of 1 part in 10^8 of the earth's gravitational field (0.01 milligal). The observations obtained with a gravimeter must be corrected so that the data solely reflects variations in the mass distribution of the earth and not, for example, latitude or altitude induced variations. This requirement for great care in reducing the gravity measurements is one of the main disadvantages of the technique. In particular, because of the influence of altitude on gravity measurements it is necessary to determine the elevations of the gravity observation points to an accuracy of 0.05 metres. This tends to make the technique much slower in practice than, for example, magnetometry. In addition to the problems associated with correcting the data, anomalies may also be induced into the readings by soil density variations. Thus, it has been generally accepted by most researchers that gravity techniques are only suitable for the detection of large voids where the induced gravimetric anomaly is greater than 0.05 milligal (Greenfield, 1979). Higginbottom (1976), however, suggests that "the repeatability of residual gravity values is likely to vary within 0.05 milligal, and the anomaly expected from the model should be at least twice this for reliable detection".

(e) Electromagnetic (EM) Methods

Electromagnetic (EM) methods based on the use of very low frequency radio signals are a relatively recent development in the field, particularly in the United Kingdom. The primary technique which has been considered as having potential is ground probing radar.

Ground probing radar (also known as EM subsurface

profiling, ground impulse radar and georadar) was developed in the early 1970's, primarily as a shallow subsurface exploration tool for engineering applications. The technique is based on the determination of the ranges to subsurface interfaces by using reflected VLF (100MHz) EM energy, as supplied from a moving transmitter. Several researchers have examined the use of the technique in engineering including Cook (1975), Moffat and Puskar (1976), Benson (1978), and Rubin and Fowler (1978).

Spencer et al (1976), describe the use of an "impulse radar profiler" over a series of known cavity locations. The radar pulses were transmitted into the ground by a small transmitter/receiver sled that was towed slowly across the ground. The return pulses were recorded on magnetic tape. Analysis of the data enabled several subsurface interfaces to be correlated with known voids. The depth of penetration was however limited to a maximum of 2.4 metres. This was mainly as a consequence of the attenuating influence of the moist, clay rich soils in the area of interests.

Darracott and Lake (1981) investigated in some detail the attenuation problems associated with ground probing radar. They suggested that the following factors influence the performance of such a system.

(i) the conductivity of the soil
(ii) the dielectric contrast between the target and soil
(iii) the shape of the target and its orientation with respect to the radar antenna
(iv) the density of scattering bodies within the soil that produce reflections similar to those from the target.

Whilst penetration can be increased by using a lower frequency the trade off in resolution terms would be such that it would be difficult to detect small cavities. Several case histories are reported by Darracott and Lake

including several related to the detection of subsurface cavities. A summary of the data relating to the most relevant examples is presented in Table 2.8.

Test Site	Objective	Depth of Penetration	Level of Success
Bagillt, Clwyd	Location of Abandoned Mineshafts	2-3 m	None
Edinburgh	Location of Cavities	10m	None
Marlborough, Wiltshire	Location Holes and Solution Cavities in Chalk	2m	Low
Aylesford, Kent	Location of Infilled Solution Features	unknown	None
Stevenson, Ayrshire	Location of Cavities Below Concrete	2 - 3 m	Very High

TABLE 2.8: SUMMARY OF GROUND PROBING RADAR CASE HISTORIES REPORTED BY DARRACOTT AND LAKE (1982).

The authors conclude that the technique "is a promising new method for site investigations" but would appear to be of most use in "the search for buried objects such as pipes, sewers and old foundations".

More promising results than those obtained by Darracott and Lake (1981) are reported by Leggo and Leach (1983). Using the same equipment as that used by the previous authors the researchers in this case were able to locate several abandoned mineshafts which had been backfilled. It was necessary however in all cases to ensure a very close spacing of the traverse lines. For example, in order to locate a 2 metre diameter shaft the traverse separation had to be less than 1 metre.

2.2 FIELD MEASURING TECHNIQUES

2.2.1 Satellite Positioning Systems

Satellite positioning is a relatively recent surveying technique and has become increasingly widespread within the last 10 years. As far as the engineering geologist is concerned it has only become a technique with likely applications in the field since the advent of the Global Positioning System (GPS). Although other positioning techniques have been developed (e.g. Doppler Transit System and PAGEOS) generally they did not provide the accuracy required for engineering geological projects at a speed appropriate to the needs of such projects. In contrast, GPS can provide the engineering geologist with high absolute and relative accuracy results in a short period of time. A detailed description of the GPS is not appropriate to this thesis and is well documented in Cross (1989).

The main attraction of GPS to the engineering geologist is that analysis of the results produced by such a system can provide valuable positional data relating to changes in the position and height of specific points over considerable distances. Its advantages compared with conventional terrestrial survey techniques (such as triangulation) are the speed of data acquisition, the ability to obtain both height and positional data simultaneously and the high accuracies achievable. Porter and McLellan (1987) reviewed a number of geotechnical engineering applications of GPS. One specific case cited the movement of a rock slide being monitored using GPS. Apart from providing results of improved accuracy compared to conventional methods the time to obtain the equivalent data by GPS was one third of that required for a conventional network of angle and distance measurements.

More recently Bordley et al (1989) have demonstrated the

potential of GPS to provide regional height data as an alternative to conventional spirit levelling. The main attractions of such a solution are the ability to provide height data over long lines irrespective of terrain, route or weather. The results have indicated that GPS levelling is within the allowable misclosures for tertiary levelling on baselines longer than 5km. On a cost bases the break even point is considered to be at the 10km level.

The main difficulties at present with GPS are:

a) the limited time 'windows' during which the system is available for use.

b) the introduction of selective availability (SA) in March 1990 which under certain circumstances reduces the accuracy considerably.

c) The lack of detailed knowledge of the geoid-ellipsoid separation which limits its ability to provide accurate absolute heights.

d) the practical line of sight difficulties of using GPA in congested urban areas or underground.

Its use is nevertheless becoming increasingly widespread and with the full implementation of the system within the next 3/4 years it will undoubtedly become the preferred method of providing positional data for many engineering geological projects in the late 1990's.

2.2.2 Aerial Photogrammetry

Aerial photogrammetry is primarily concerned with topographic mapping, where the location of detailed point and line feature information to a high degree of accuracy is paramount. Consequently, very high demands are set for both the planimetric accuracy of the objects to be mapped, and of the accuracy of the height information. Despite the

advances in electronic surveying instrumentation, which are discussed later, as far as mapping is concerned photogrammetry using aerial photography continues to be the primary method for medium to small scale mapping. Only at very large scales (1:1000 and larger), and over small areas, can ground surveying methods compete economically with aerial photogrammetry.

A comprehensive account of the principles and technology of photogrammetry and its applications to engineering geology is reported in Kennie [9]. The text, written as a reference source for civil engineers and engineering geologists emphasises both the simple, graphical and mechanical techniques in photogrammetry and the benefits of the more rigorous analytical techniques.

2.2.3 Close Range Photogrammetry

Close range photogrammetry refers to the sub-discipline which deals with the specific measurement problems which occur when the distance from the imaging source to the object is of the order of a few tens of metres at most (and generally much closer). It is not a new topic and in fact was developed before aerial photogrammetry. However, the development of analytical techniques and specialist instruments designed specifically for close range measurement has opened up a whole new range of applications, particularly in engineering geology and geotechnical engineering.

While the method overlaps with more direct methods of high precision measurement (such as those based on electronic theodolites), there are quite unique circumstances when close range photogrammetry is of particular benefit. The first occurs when remeasurement may be required at a later date. The ability to store the 'object' on film offers a major advantage over alternative methods. Similarly if the object is moving e.g. when monitoring laboratory samples, the ability of photography to 'freeze' the motion

of the object is also of significance.

The benefits of using analytical close range photogrammetry (as opposed to analogue techniques) are listed in Kennie and McKay [10] and include:-

a) Higher accuracy.

b) Photography with different format sizes and focal lengths can be used. In certain cases, provided suitable precautions are taken to eliminate any systematic errors such as lens distortion, non-metric cameras can be used.

c) Increased speed in the setting up of the photography prior to measurement.

d) Greater flexibility in the acquisition of photography.

e) Data are stored in digital form.

These benefits of close range photogrammetric approach have been exploited by a number of other workers in this field and Table 2.9, summarises a selected number of the applications of photogrammetry to engineering geology.

Topic	References to the use of close range photogrammetry
Rockface Mapping	Cheffins and Rushton, 1970 Torlegard and Dauphin, 1975 Robertson et al, 1982
Landslide Monitoring	Chandler et al, 1987 Heath et al, 1978 Heath and Dowling, 1980 Chandler et al, 1989
Earth and Rockfill Dam Monitoring	Planicka and Nosek, 1970 Moore, 1973 Brandenberger, 1974
Gabion Wall Monitoring	Veress et al, 1980
Channel/Gully Erosion Monitoring	Collins and Moon, 1979 Welch and Jordan, 1983
Laboratory Studies	El-Beik, 1973 Wong and Vonderoe, 1981 Davidson, 1985

TABLE 2.9: APPLICATIONS OF CLOSE RANGE PHOTOGRAMMETRY IN ENGINEERING GEOLOGY

In the case of both aerial and close range photogrammetry the nature of the instrumentation used for precise measurement has changed considerably in the last 30 years. The analogue stereoplotter, based on the use of optical or mechanical models and manufactured during the twenty year period between 1955 and 1975 is now in decline. The extent of this decline is such that today only one or two manufacturers are currently producing such instruments, and only in small numbers. Although the analytical stereoplotter was originally invented in the late 1950's it did not have a major impact in the civilian sector until the mid 1970's. However, since the technology has developed considerably (with significant real term reductions in price and increases in performance) and

such instruments are now manufactured by all major photogrammetric equipment manufacturers. The impact of 'all digital' acquisition and measurement systems is still in its infancy, although many anticipate that the current technological limitations to the widespread development of such instruments will be overcome in the next decade.

For the future, a number of different possible scenarios can be identified. The first, and easiest to identify is the further decline in the production of analogue instrumentation (Figure 2.1d), the second (Figure 2.1c) is the increasing interest and development of all digital measurement systems, although the rate of development is much less easy to predict. Furthermore it remains to be seen whether the development of all digital systems will be at the expense of analytical instruments (Figure 2.1b) in the short term (5-10 years), or whether the two approached will continue to develop in parallel (Figure 2.1a) for a longer period and eventually emerge in the longer term (10 years or more).

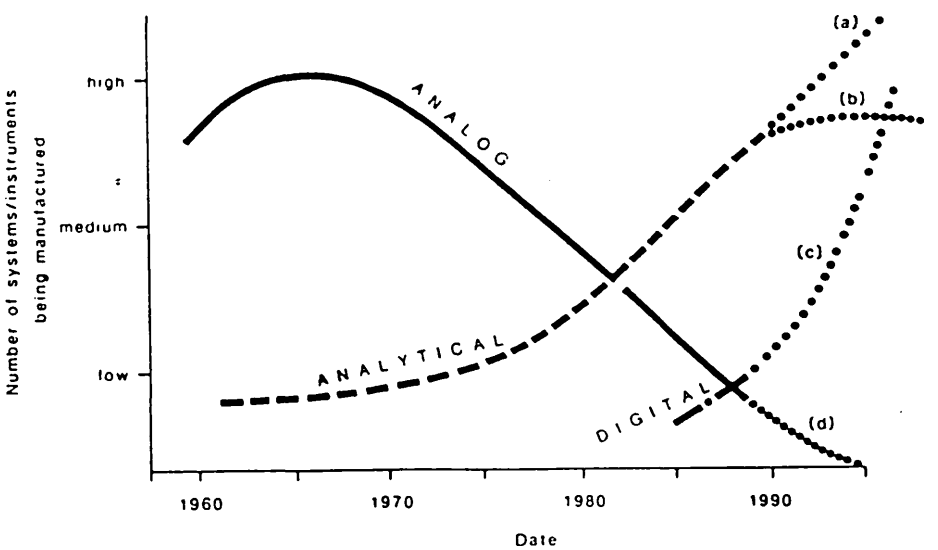


Figure 2.1: Development of analogue, analytical and digital photogrammetric instrumentation and future scenarios. (Kennie and Petrie, 1989)

2.2.4 Ground Surveying

a) Introduction

Engineering geologists have historically used the traditional optical and mechanical surveying instruments to produce simple site surveys, profiles and cross sections and to provide positional date for borehole investigations. A review of these techniques can be found in Kennie [11].

In recent years the design of such instrumentations has changed dramatically and it has become less common for engineering geologists to perform the field surveying activities. The impact of the recent technological developments on surveying instrumentation is documented in Kennie [12], and this earlier publication has recently been further extended in Kennie (1990). The influence of these developments in the surveying world has been considerable enabling date to e acquired at a speed and accuracy which was quite impossible to achieve using manual non-electronic methods. In addition, the new instruments also provide the capability to perform tasks which were not previously possible. For example the development of electronic coordinate determination systems (ECDS) Kennie [12] page B272-282 enables very high precision operations to be performed. Such techniques could have a useful role to play as a independent laboratory based measuring system.

b) Instrument Calibration

One of the most critical requirements if such instrumentation is to provide accurate results is the extent to which the equipment is regularly calibrated. Currently the procedures for the calibration of electronic surveying equipment are wide and varied. In the case of Electronic Distance Measuring (EDM) equipment some progress has been made and Kennie et al [17] review the current status of equipment calibration in UK and the future action required to ensure the measurements from such equipment are

traceable to national standards. As a consequence of this Study a Working Party of the Royal Institution of Chartered Surveyors (RICS) is now in the process of drafting guidelines for the standardisation of procedures for EDM instrument calibration. These guidelines will be published during the first quarter of 1991.

2.2.5 Invasive and Non Invasive Measuring Techniques.

a) Invasive Techniques

One of the most common invasive instruments used in engineering geology is the accelerometer. An accelerometer is a transducer which provides an output signal proportional to acceleration. The accelerometer is most frequently used to measure vibrations. The accelerometer can respond to a wide range of frequencies (4 Hz to 1500 Hz), is light (0.5 g) and compact (diameter = 5-6mm). These properties are important and ensure that the accelerometer has a minimal effect on the properties of the vibrating system which is being examined.

Many of the disadvantages of using accelerometers are associated with the use for monitoring ground vibrations around existing structures. Wiring up an existing structure is often difficult and may be impossible in some cases e.g. offshore structures or tall masts. Furthermore, the wiring may be easily damaged or alternatively because of the length of wiring, excess 'noise' may be set up. Often accelerometers are placed inside the structure to provide more reliable results. In these circumstances however problems with servicing and damage to the accelerometer may occur.

b) Non-Invasive Techniques

(i) Laser Interferometry

Ashkenazi et al. (1974) discussed the application of a laser interferometer for the measurement of continuous dynamic movement. The interferometer used was a Hewlett-Packard 5526A 700uW, Helium-Neon (He-Ne) gas laser emitting coherent radiation in the visible range of the spectrum with an average wavelength of 632.8nm.

The performance of the interferometer for vibration analysis was assessed by a laboratory rig which simulated frequencies in the range from 0.15 Hz to 0.9 Hz with amplitudes ranging from 2 to 30 mm. A conventional glass retro-reflector was attached to the test rig to enable remote measurement using the laser interferometer. Data acquisition was carried out using a paper tape data logger and the filtered data were processed using Fourier Transform analysis. The results of this project indicated that frequencies in the range 0.15 to 0.9 Hz, with amplitudes of up to 25 mm could be determined using such a system. The primary limiting factors to the more widespread use of the technique were identified as the slow paper tape data acquisition system and that the maximum operating range of the instrument was 50-60 metres.

(ii) Laser Vibrometer

An alternative instrument used to measure vibrations based on the measurement of the Doppler frequency shift is the Laser Vibrometer developed at the Institution of Sound and Vibration, University of Southampton (Halliwell (1981) and Pickering et al. (1986). The instrument is portable, relatively inexpensive and has been used in fluid mechanics and motor and sound engineering, to measure vibrations of high frequencies (200-1000 Hz).

However, the laser source employed in the instrument has an output of less than 5uW and consequently the operating range of the instrument is limited to 1 m approximately. It is therefore currently inappropriate for use in the determination of the vibrations normally associated with engineering projects.

(iii) Cine-Photogrammetry

The potential of high speed photography for monitoring oscillations of tall antennae was recently examined by Ethrog & Horamy (1988). In this case a non-metric 16 mm cine camera with an exposure interval of 1/60 S was examined. The authors carried out manual photogrammetric measurements of single frames of photography. Because of the considerable time required to measure the photograph the authors considered the technique to be impractical in its present form. However, some suggestions for future work to automate the point search and measurement procedures were provided.

2.3 CONCLUSIONS

The previous sections have discussed the role of the wide range of remote investigation and field measuring techniques for engineering geological applications. It is clear that all the techniques listed have their limitations and they are more suited to specific stages of a civil engineering project. Using the procedure previously introduced in Table 2.6, Tables 2.10 and 2.11 summarise the potential use of the techniques discussed at the various stages of a civil engineering project.

TABLE 2.10 POTENTIAL USE OF REMOTE INVESTIGATION TECHNIQUES AT VARIOUS STAGES OF A CIVIL ENGINEERING PROJECT.

	Remote Investigation Technique				
Stage of Project	Satellite Remote Sensing	Aerial Photography	Airborne Thermal	Airborne MSS	Geophysics
Reconnaisance	***	**	*	*	**
Feasibility	***	***	***	**	***
Design	-	**	**	**	*
Construction	-	*	-	-	*
Post-Construction	-	*	**	*	*

KEY

*** Very Useful

** Useful

* Of limited Use

- Inadequate

TABLE 2.11 POTENTIAL USE OF FIELD MEASURING TECHNIQUES
VARIOUS STAGES OF A CIVIL ENGINEERING PROJECT.

	Field Measuring Technique				
Stage of Project	Satellite Positioning	Ground Surveying	Aerial Photgrammetry	Close Range Photgrammetry	Indirect Techniques
Reconnaisance	***	**	***	-	-
Feasibility	***	**	***	**	-
Design	*	*	-	***	**
Construction	**	***	*	-	**
Post-Construction	*	**	*	***	***

KEY

*** Very Useful

** Useful

* Of limited Use

- Inadequate

REFERENCES - CHAPTER 2

ARROWSMITH E.J., and RANKILOR P.R., (1981). Dalton By Pass : Site Investigations in an Area of Abandoned Haematite Mine Workings. Q.J. Engineering Geology, Vol. 14, 201-210.

ASHKENAZI V. (1974). Measurement of dynamic deformation by laser interferometry. Civil Engineering Surveyor, Oct - Nov., 11-14.

BALLARD R.F., CUENOD Y., and JENNI J.P., (1983). Detection of Karst Cavities by Geophysical Methods. Bulletin of the Int. Assoc. of Engineering Geology, No. 26-27, Paris, 153-157.

BEAUMONT T. (1977) Techniques for the interpretation of remote sensing imagery for highway engineering purporses. TRRL Report 53, Transport and Road Research Lab., 24pp.

BENSON R.C., (1978). Applications of Ground Penetrating Radar to Geotechnical, Hydrologic and Environmental Assessments. Proc. of the Soc. of Expl. Geophy., San Francisco.

BIOT M.A., (1952). Propagation of Elastic Waves in a Cylindrical Bore Containing a Fluid. Applied Physics, Vol. 23, No. 9, 997-1005.

BORDLEY R., CHRISTIE R., DODSON A., GERRARD S. 1989. Taking the Spirit out of Levelling. Land and Minerals Surveying, Vol 7, 526-534.

BRANDENBERGER A.J. (1974). Deformation measuremnts of power dams. Photogrammetric Engineering, 40, 1051-8.

CHANDLER J.H. and COOPER M.A.R. 1989. The extraction of positional data from historical photographs and then application to geomorphology. Photogrammetric Record, Vol XIII, 73, 69-78.

CHANDER J.H., CLARK J.S., COOPER M.A.R. and STIRLING D.M. (1987). Analytical photogrammetry applied to Nepalese Slope morphology. Photogrammetric Record, 12 (70), 443-458.

CHEFFINS O.W. and RUSHTON J.E.M. (1970). Edinburgh Castle Rock, a survey of the north face by terrestrial photogrammetry. Photogrammetric Record, 8, 46, 417-33.

COLLINS, S.H. and MOON G.C. (1979). Stereometric measurement of stream bank erosion. Photogrammetric Engineering and Remote Sensing, 45, 183-90.

COOK J.C., (1965). Seismic Mapping of Underground Cavities using Reflection Amplitudes. Geophysics, Vol. 30, No. 4, 527-538.

COOK J.C., (1975). Radar Transparencies of Mine and Tunnel Rocks. Geophysics, Vol. 40, No. 5, 865-887.

CROSS P.A. 1989. Satellite Positioning fixing systems for land and offshore engineering surveying. Chapter 4 in Engineering, Surveying Technology, Kennie T J M and Petrie G (Eds), Blackie & Sons 111-145.

DARRACOTT B.W., and LAKE M.I., (1981). An Initial Appraisal of Ground Probing Radar for Site Investigation in Britain. Ground Engineering, Vol. 14, No. 3, 14-18.

DAVIDSON J.L. (1985). Stereophotogrammetry in geotechnical engineering research. Photogrammetric Enginerring and Remote Sensing, 51, 1589-96.

DEARMAN W.R., BAYNES F.J., and PEARSON R., (1977). Geophysical Detection of Disused Mineshafts in the Newcastle upon Tyne area, North East England. Q. J. Engineering Geology, Vol. 10, 257-269.

DOYLE F. (1982). Statellite systems for cartognaphy. Int. Archives of Photogrammetry, (24) 1, 180-188.

EL-BEIK A.H.A. (1973). Photogrammetry in centrifugal testing of soil models. Photogrammetric Record, 7, 41, 538-54.

ETHROG V. and HORAMY N. (1988). Measuring the harmonic oscillations of tall antennae by photogrammetry. Int. Archives of Photogrammetry and Remote Sensing, 27, 585-591.

FOUNTAIN L.S., HERZIG F.X., and OWEN T.E., (1975). Detection of Subsurface Cavities by Surface Remote Sensing Techniques. Final Report No. FHWA-RD-DOT-FH-11-8496. Prepared by the Southwest Research Institute, Houston, 124pp.

GRAHAM R.W. and READ R. (1984). Small format aenial photography from microlight platforms. J. Photog. Science, 32, 100-109.

GREENFIELD R.J., (1979). Review of Geophysical Approaches to the Detection of Karst. Bulletin of the Association of Engineering Geology, Vol. 16, No. 3, 393-408.

GRIFFITHS D.H., and KING R.F., (1981). Applied Geophysics for Engineers and Geologists. Pergamon Press, London, 2nd Edition.

HALLIWELL N.A. (1981). The laser torsional vibrometer, a new instrument. Journal of Sound and Vibration, 93 (4), 588-592.

HEATH W. (1980). Inexpensive aenial photography for highway engineering and traffic Studies. TRRL, Suppl. Report, 632, Transport and Road Research Report, 25pp.

HEATH W. and DOWLING J.W.F. (1980). Examples of the use of terrestrial photogrammetry in highway engineering. Proceedings of the 14th. Congress of the International Society of Photogrammetry, Hamburg, Commission V, 308-20.

HEATH W., PARSLEY L.L. & DOWLING J.W.F. (1978). Terrestrial photogrammetric surveys of unstable terrain in Columbia. Transport and Road Research Laboratory Report, LR 876.

HIGGINBOTTOM I.E., (1976). The Use of Geophysical Methods in Engineering Geology: Part 2, Electrical Resistivity, Magnetic and Gravity Methods. Ground Engineering, Vol. 9, No. 3, 34-38.

KENNETT P., IRESON R.L., and CONN P.J., (1980). Vertical Seismic Profiles: their application in Exploration Geophysics. Geophysical Prosp., Vol. 28, 676-699.

KENNIE T.J.M. and PETRIE G. (1989) Engineering Surveying Technology, Blackies Publishing Co. Ltd, Glasgow, 357pp.

KENNIE T.J.M. (1990). Field data collection for terrain modelling. Chapter 2 in Terrain Modelling in Surveying and Civil Engineering, Kennie TJM and Petrie G (Eds), 4-16.

LEGGO P.J., and LEACH C., (1983). Subsurface Investigation for Shallow Mine Workings and Cavities by the Ground Impulse Technique. Ground Engineering, Vol.16, No. 1, 20-23.

MATTHEWS M.C. AND CLAYTON C.R.I. (1988). The use of obligal photography to assess the extent and sequence of landslipping at Stag Hill, Guildford. Risc. 20th Annual Meeting of the Eng. Group of the Geolog. Society, Guildford, 8pp.

McCANN D.M., BARIA R., JACKSON P.D., CULSHAW M.G., GREEN A.S.P., SUDDABY D.L., and HALLAM J.R., (1982). The use of Geophysical Methods in the Detection of Natural Cavities, Mineshafts and Anomolous Ground Conditions. Engineering Geology Report No. 82/5, British Geological Survey, 272pp.

McDOWELL P.W., (1975). Detection of Clay Filled Sinkholes in the Chalk by Geophysical Methods. Q.J. Engineering Geology, Vol.8, 303-310.

McDOWELL P.W., (1981). Recent Developments in Geophysical Techniques for the Rapid Location of Near Surface Anomolous Ground Conditions. Ground Engineering, Vol. 14, 3, 20-23.

MOFFAT D.L., and PUSKAR R.J., (1976). A subsurface Electromagnetic Pulse Radar. Geophysics, Vol. 41, No. 3, 506-518.

MOORE J.F.A. (1973). The photogrammetric measurement of the constructional displacements of a rockfill dam. Photogram-constructional displacements of a rockfill dam. Photogrammetric Record, 7, 42,628-48.

PICKERING C.J.D. and HALLIWELL N.A. (1986). The laser torsional vibrometer: a portable instrument. Journal of Sound and Vibration, 107 (3), 471-485.

PLANICKA A. and NOSEK L. (1970). Terrestrial photogrammetry in measurement of deformations of rockfill dams. Proceedings of the 10th International Conference on the 10th International Conference on Large Dams, Montreal, 207-15.

PORTER T. and McLELLAN J. 1987. GPS Surveying Techniques for Deformation Analysis. ASP-ACSM Convention, Baltimore, 10pp

ROBERTSON G.R., MCRAE A.M.R., TRIBE J., SIBLEY D.W. and SMITH D.H. (1982). Use of photogrammetric methods for mine slope deformation surveys. Proceedings of the 4th Canadian Symposium on Mining and Deformation Monitoring.

RUBIN L.A., and FOWLER J.C., (1978). Ground Probing Radar for Delineation of Rock Features. Engineering Geology, Vol. 12, 163-170.

SPENCER J.W., KOCA J.F., RIB H.T., and FALLS C.P., (1976). Evaluation of Geophysical Systems for Remote Sensing of Subsurface Cavities in Kansas. Transportation Research Record, 581, 31-41.

TOMLINS G.F. (1983). Some considerations in the design of low cost remotely piloted aircraft for Civil remote sensing applications. Canadian Surveyor, 37, 157-167.

TORLEGARD A.K.I. and DAUPHIN E.L. (1975). Deformation measurement by photogrammetry in cut and fill mining. Proceedings of the ASP Symposium on Close Range Photographic Systems. University of Illinois, 24-39.

VERESS S.A., JACKSON N.C. and HATZOPOULOS J.N. (1980). Monitoring a gabion wall by inclinometer and photogrametry. Photogrammetric Engineering and Remote Sensing, 49, 771-8.

WELCH R and JORDAN T (1983). Analytical non-metric close range photogrammetry for monitoring stream channel erosion. Photogrammetric Engineering and Remote Sensing, 49, 367-74.

WONG K W. and VONDERHOE A.P. (1981). Planar displacement by motion parallax. Photogrammetric Engineering and Remote Sensing, 47, 769-77.

ZALASIEWICZ J.A., MATHERS S.J., and CORNWELL J.D., (1985). The Application of Ground Conductivity Measurements to Geological Mapping. Q.J. Eng. Geology, Vol. 18, 139-148.

CHAPTER 3

DISPLAY AND SYNTHESIS OF ENGINEERING GEOLOGICAL DATA

3. DISPLAY AND SYNTHESIS OF ENGINEERING GEOLOGICAL DATA

3.1 INTRODUCTION

The display of engineering geological data is a vital element of the work performed by professionals in this field. The range of 'thematic' data which may be recorded is vast and includes not only topographical information but also:-

- Geological data
- Soils data
- Hydrogeological data
- Geomorphological data
- Morphogenetic data

In addition to the display of the new data in the form of 'record maps' is the need for the display of second generation derivative or predictive maps based on the interpretation and synthesis of data from several sources. Typical examples in this case may include

a) Derivative maps formed by the superimposition, or combining of several data sources. Examples may include drainage pattern maps (catchment area analysis), site investigation planning maps and so on.
b) Predictive maps where the likelihood of some future action is displayed in a map form. Examples in this case might include the assessment of slope stability, environmental impact etc.

Historically such processes have been performed manually using conventional drafting techniques. Increasingly however, these processes are becoming automated and performed with the aid of computer technology.

Figure 3.1 illustrates the range of thematic information which may form part of the derivative or predictive data produced by the engineering geologist.

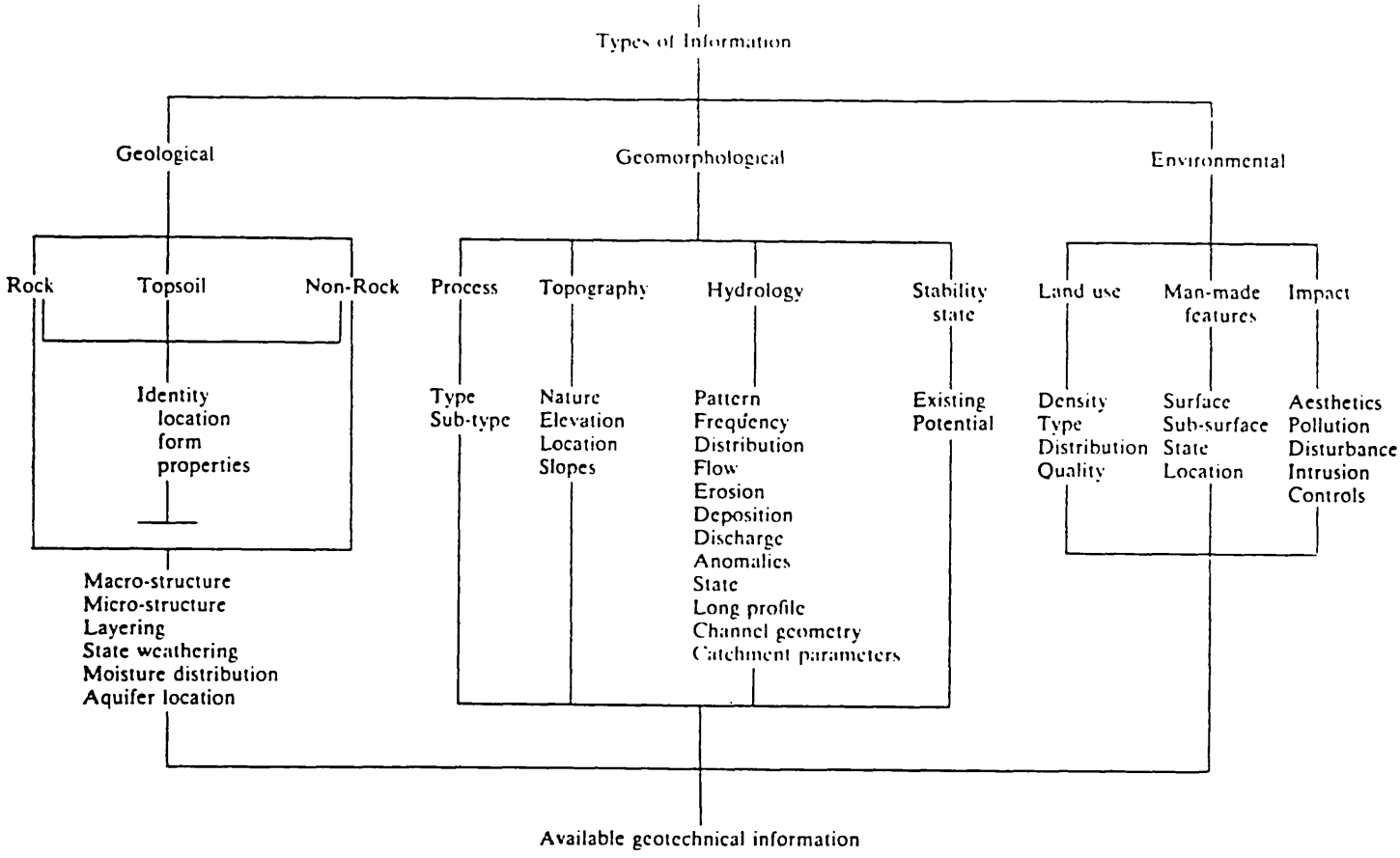


Figure 3.1
Types of Geotechnical Information
(Geological Society Working Party, 1982)

This section of the thesis reviews the opportunities offered to the engineering geologist by the advances which have occurred in the fields of digital terrain modelling (section 3.2) and Spatial Information Systems (sections 3.4).

3.2 DIGITAL TERRAIN MODELLING

The concept of creating digital models of the terrain is a relatively recent development and the introduction of the term Digital Terrain Model (DTM) is generally attributed to two American engineers (Miller and La Flamme, 1958) working at the Massachusetts Institute of Technology. Since then the concept has progressed considerably driven by a number of factors including the explosive developments in computer technology, the advances in computer graphics and imaging technology and the availability of data acquisition systems such as electronic tacheometers, digital and analytical photogrammetric instruments and automated digitising systems. As a consequence DTM techniques are used widely in many areas of engineering geology and in other fields ranging from landscape architecture to radio communications planning and aircraft simulators.

A plethora of terms have been devised to describe the process of representing the continuous surface of the ground by a selected number of discrete points of known position and height; these terms include Digital Elevation Model (DEM), Digital Ground Model, (DGM) Digital Height Model (DHM) and Digital Terrain Elevation Model (DTED) to name some of the more common. While DTM's are clearly related to and, indeed, may form part of a Geographic Information System (GIS) (Section 3.4) they are at present more commonly independent, stand alone software systems.

3.2.1 Methods and Software

A steadily increasing number of DTM software packages have been developed over the past 25 years or so, and in the UK alone, there are now in excess of 30 commonly used packages for civil engineering and engineering geological applications. The systems are broadly divisible into those which are grid based and those which are triangular based. The first approach assumes that the terrain has been approximated by a regular, systematic pattern of height

points e.g. on a regular grid, whereas the second approach assumes that points have been selectively chosen in a non regular manner with the vertices of these discrete points forming a triangular mesh. The formation of this Triangulated Irregular Network (TIN) is generally performed by one of two algorithms known as the radial sweep and Delaunay methods respectively. The TIN approach is generally perceived to provide a more precise definition of the terrain, although for large area projects the storage and retrieval benefits of the grid approach are often preferable.

The relative merits of both approaches has been discussed extensively in the literature and further details can be found in Petrie and Kennie [18]

3.2.2 Applications

The use of DTM's in engineering geology is wide and varied. Among the more common uses are:-

(a) The measurement of earthworks from borrow pits or spoil areas.

(b) Catchment/watershed analysis to assess the storage capacity of reservoirs, and

(c) Intervisibility analyses to assess the visual impact of a proposed development.

The software tools required in each case to produce a solution are quite different, although they all require the ability to interrogate a simple, single attribute spatial database. It is this particular feature which provides the link between DTM systems and the broader concept of a Spatial or Geographic Information System (GIS) (Section 3.3).

3.3 DTM'S AND GIS IN ENGINEERING GEOLOGY

GIS and the closely allied activities of LIS (Land Information Systems) and AM/FM (Automated Mapping/Facilities Management) are slowly beginning to have an impact in the field of engineering geology. Since the majority of GIS systems have so far been only designed to deal with 2D datasets (and not the 3D data generated for use in DTM systems), the majority of GIS applications have been based around the planimetric spatial data derived from large scale maps. A wide range of potential applications of GIS in engineering geology exist and some of these include:

(a) Groundwater Management
(b) Geotechnical Site Investigation
(c) Hazard Assessment/Environmental Impact Analysis
(d) Land Use/Environmental Planning

While a knowledge of the third dimension (of both the terrain and features on the terrain) would be helpful in all of these applications it is rare so far for such data to be included in any analysis. Yet in many instances such data may have a vital role to play in the solution of the specific problem being investigated.

3.3.1 The Scope of DTM's in GIS

The inclusion of DTM capabilities within a GIS provides the user with the opportunity to model, analyse and display phenomena related to specific 3D surfaces (e.g. the earths topography or some design surface). The scope of the tasks which may be addressed by terrain modelling may include one or more of the following activities. For a more comprehensive discussion please see Weibel and Heller (1990).

(a) DTM Creation

This process involves the capture of the discrete point data over the surface and the construction of the 'model' using various algorithms. This data capture stage may involve a wider range of surveying instrumentation including field surveying instrumentation (Kennie, 1990), analogue and analytical photogrammetric systems (Petrie, 1990) and digitising systems (Petrie, 1990).

(b) DTM Manipulation

A fundamental feature of most DTM systems is the capacity to edit, quality control and merge data captured in the first process. This stage may also involve the capacity for the conversion of data from one 'model' form to another. For example the PANACEA system (McCullagh, 1990) Fig 3.2 initially forms a triangulated model from the input which may then be transformed into a grid form for display or visualisation purposes.

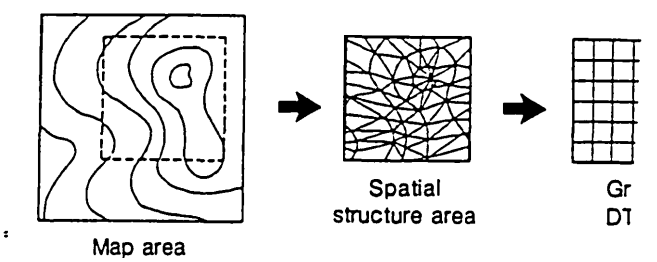


Figure 3.2:
PANACEA Triangulated and Grid DTM's (McCullagh, 1990)

(c) DTM Analysis

At the next level of development many systems have the capability to extract information from the raw point data. Derivative information such as slope, aspect or the extraction of geomorphological characteristics may form part

of this stage. The latter process is one area of considerable interest and may in time enable automated analysis of drainage networks to be performed to establish hydrological run off or soil erosion potential. At a more basic level this stage may include the capability to establish volumes, extract profiles and carry out intervisibility analyses.

(d) DTM Visualisation

Increasingly even the most basic DTM system has the capability to produce 2D or 3D displays of the terrain. At its simplest level this may involve the orthographic display of a wireline model and at its most sophisticated a fully rendered 3D image which can be interrogated as a dynamic, moving image. A comprehensive account of the processes involved in digital terrain and landscape visualisation can be found in Kennie and McLaren, [19] and Figure 3.3 indicates diagramatically the relationship between the various techniques used to visualise the terrain.

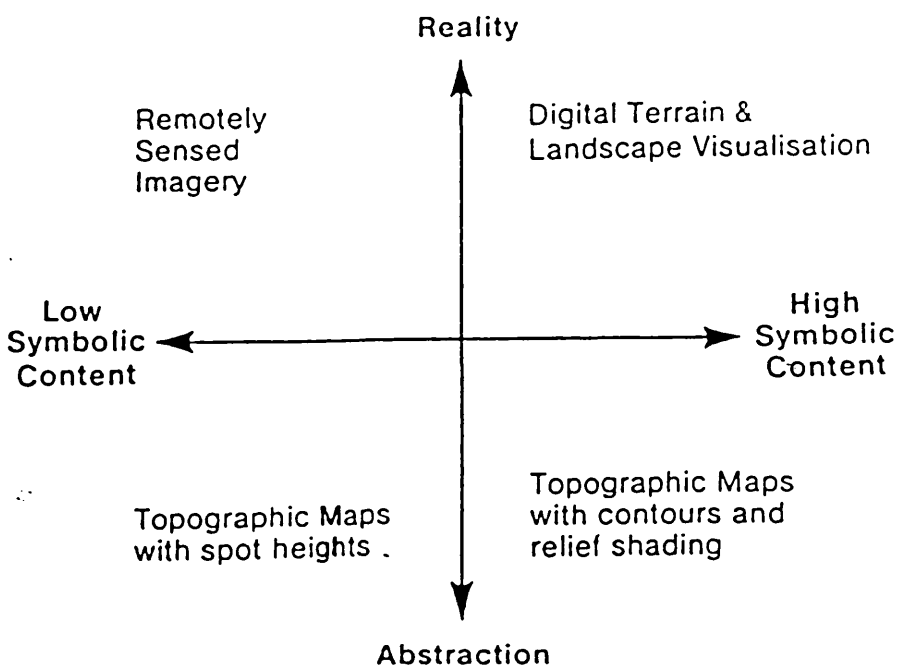


Figure 3.3
Terrain Visualisation (Kennie and McLaren [19]

(e) DTM Applications

At this stage are the discipline specific applications of DTM's. In the context of engineering geology this may include the production of

(i) contour maps and grid DTM's illustrating height variations

(ii) estimates of the volume of rock or soil within a specific area

(iii) borehole records and subsurface 3D models of geological strata.

(f) DTM Spatial Queries

All of the previously mentioned tasks are generally available within the stand alone DTM systems discussed in Section 3.2. They also form the heart of many DTM facilities within GIS software, although in many cases the linkage between the two is rather limited. This capability to perform fully three dimensional spatial queries is however, the integrating mechanism which will in the future provide the linkage between the two facilities. GIS software must have the potential to perform simple 2D spatial queries such as 'display all soil types of a specific type which are within the limits identified on the screen '. Increasingly for planning purposes there is a need to answer more complex queries such as 'display/list all development areas within 1 km of an A class road which have a value/hectare of less than a, are between 150 and 200m above sea level, are not visible from locations x y and z and display the results in perspective form from location W. The facilities for such queries in the current range of GIS systems are very limited, although in the future this must change.

3.4 SPATIAL INFORMATION SYSTEMS

Considerable developments have taken place in the recent past in the computer handling of spatial databases of geoscientific data. Such systems are generally referred to as Geographic Information Systems (GIS Burroughs, 1986), Land Information Systems (LIS, Parker, 1989) and less commonly Geoscientific Mapping and Modelling Systems (GMMS, Roper, 1989). All share a number of common characteristics and these are illustrated by Figure 3.4 in the context of the field of enquiry for this thesis geological/geotechnical spatial information systems.

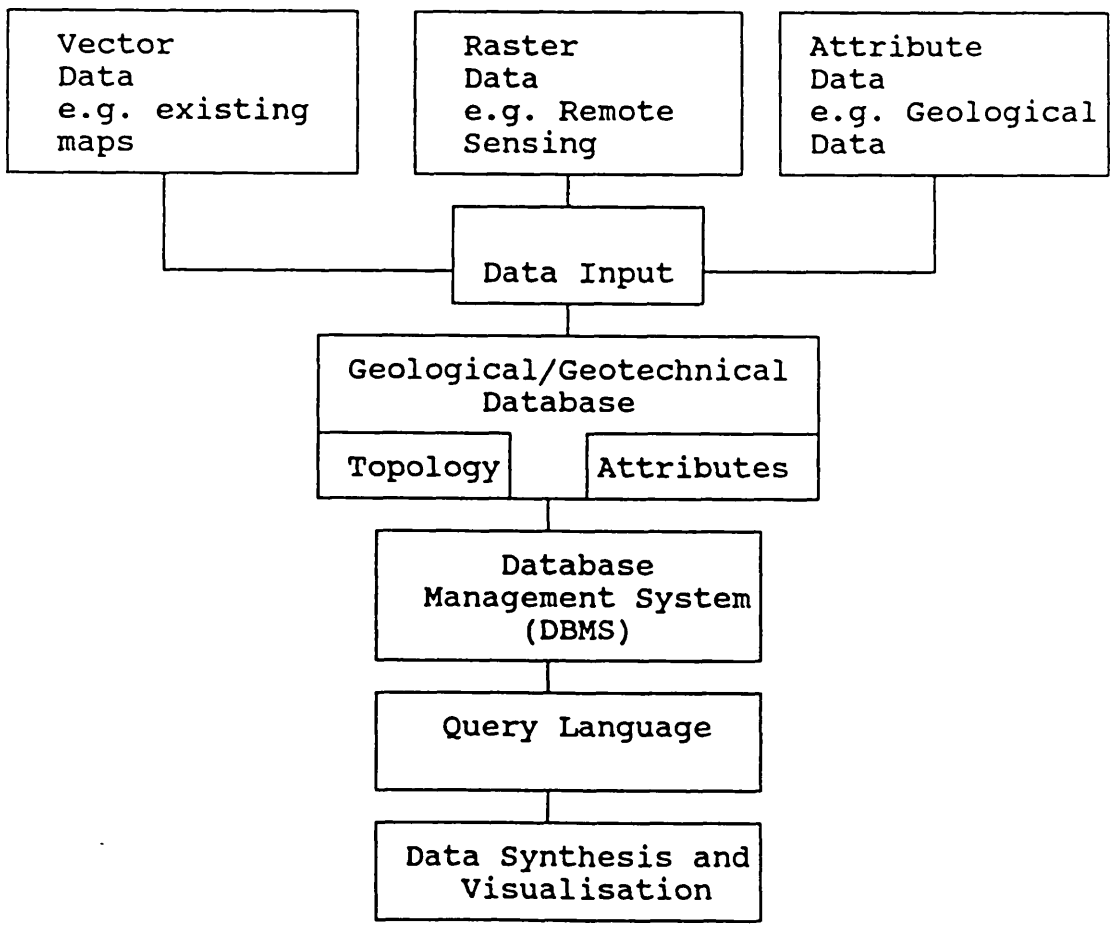


Figure 3.4
Components of a geological/geotechnical spatial information systems (GSIS)

The main characteristics of such systems are the ability to objectively synthesise considerable amounts of spatial data and visualise the results of specific spatial queries. The range of possible applications of such systems in engineering geology is vast particulary for reconnaissance and feasibility studies.

The study of spatial information modelling is now developing into a major inter-disciplinary topic of interest in the geoscientific community. At present its use in engineering geology is primarily restricted to regional planning and feasibility studies enabling the conventional thematic maps produced manually by engineering geologists to be produced using computer technology. Increasingly the use of such techniques will become an integral part of all engineering geological investigations at both small and larger scales.

3.5 FUTURE DEVELOPMENT OF 3D GIS

It would not seem too fanciful to suppose that in the near future (1-5 years time frame) GIS users in the engineering geological community will increasingly demand the capability to be able to handle 3D datasets and to make truly 3D spatial queries. To allow this to occur, a number of developments will need to take place. Two distinct areas for further work can be identified, firstly activities related to data acquisition and secondly progress in the field of data generalization and transfer standards.

3.5.1 Developments in Data Acquisition

(a) UK National Terrain Database

There is no doubt that one of the predominant limiting factors to the more widespread application of GIS in engineering geology is that of data capture and this is particularly relevant in the context of DTM data. Whereas the Ordnance Survey has been progressively transferring its paper based large scale maps into digital format it is only

relatively recently that the production of a National Terrain Database (NTDB) has been initiated. Clearly therefore the lack of widely available DTM coverage of the UK is an area of need for the future. A number of other countries already possess such data on a national or semi-national coverage including the USA (El Assal and Carusu, 1983) and Sweden (Moren, 1983).

The production of a UK NTDB was initially identified as a requirement for military purposes and hence the data capture process was driven by the Mapping and Charting Establishment (MCE), although it is now being marketed by the Ordnance Survey. The DTM is based on the digitised 5 or 10 metre contours form the 1:50,000 map series, and the data is provided on a 50m interpolated grid. To date coverage is available for large sections of England and Wales.

While the provision of a large scale dataset of the form of the NTDB will assist considerably in relation to the data capture requirements of engineering geologists, from a data accuracy point of view care needs to be taken over the use of such data. According to the Ordnance Survey, the accuracy of the DTM data produced to date from 1:50,000 contours lies between $\pm$1.7 to 2.5m. However separate tests at the University of Glasgow suggest that the accuracy which can be expected on a national basis may be considerably poorer. Using 1:50,000 10m contours (as opposed to 5m) in a rugged upland area indicated a standard error of $\pm$6.5m. Whatever the correct figures are it is clear that the use of such data by engineering geologists will be restricted to regional planning and will clearly be inappropriate for large scale design projects.

(b) Bare Earth v Cultural DTM's

A separate consideration relating to the future use of such data lies in the extent to which such data will be integrated with other 2D digital information. Ideally the

'bare earth' NTDB will need to be integrated and related to separate 2D datasets at similar or larger scales in order to provide a 'cultural' dataset.

(c) Urban DTM's

A further large scale data requirement is the need for the provision of 3D urban datasets which include not only terrain data but also 3D information about buildings. Such data is often acquired as part of the mapping/revision process and where available is being used by architects and planners for visualisations of proposed developments. City databases for such data are beginning to be created (Herbert, 1987) and the Ordnance Survey are now marketing such data on a limited basis.

3.5.2 Developments in data generalization and transfer standards

(a) Data Generalization

The ability to manipulate, analyse and visualise the very large datasets discussed in 3.5.1 will require further work if these operations are to be performed economically. In particular work needs to be carried out into methods of data generalisation. The ability to use the wide area data for visualisation purposes (where low resolution may be quite acceptable) with higher resolution datasets for intervisibility operations (where high accuracy is of more significance) will require the ability to generalise smaller scale datasets from the original NTDB.

(b) Information Exchange and Transfer Standards

The issue of tradeable geographic data is a relatively recent phenomena, particularly in engineering geological practice. Clearly however the value of such information (in this context DTM data) can be considerable. Hosken and Parker (1990) indicated that in a small survey which they

conducted approximately one third of the consulting engineers they contacted had data which they considered was tradeable. The facilities for the commercial exchange of such datasets needs to be established. Such a process would also help formalise the technical standards for the exchange of such data. In absence of such standards a number of vendor specific 'standards' have become established e.g. MOSS, Integraph etc. While this is quite satisfactory at present in the longer term it may be more appropriate to devise a non vendor specific standard form.

3.6 CONCLUSIONS

GIS is a relatively recent phenomenon in engineering geology although engineering geologists are extensive users of spatial data. The future development of the topic within this field must include the capability of handling 3D data as an integral element of the overall system. Capabilities must also exist for linking general purpose GIS facilities to more specialist engineering design oriented modelling software such as MOSS. The future use of such software in engineering geology offers exciting opportunities both in terms of enhanced productivity within existing projects and potentially for the provision of additional value added services to clients.

REFERENCES

BURROUGHS P.A. 1986. Principles of Geographical Information Systems for Land Resources Assessment Clarendon Press, Oxford, 193pp

EL ASSAL A.A. and CARUSO V.M. 1983. Digital Elevation Models, US Geological Survey Circular 895-B, 40pp

HERBERT A. 1987. A walk on the Clydeside. CADCAM International, 6(4), 41-42.

HOSKEN E. and PARKER D. 1990. Civil Engineers need Geographic Information. Paper presented at Association for Geographic Information Conference, Brighton, 6pp

KENNIE T.J.M. 1990. Field Data Collection for Terrain Modelling. Chapter 2 in Terrain Modelling in Surveying and Civil Engineering. Petrie G and Kennie TJM (Eds), Whittles Pub, 4-17,351pp

MILLER C. and LAFLAMME R.A., 1958. The digital terrin model-theory and applications - Photogrammetric Engineering, 24,3,433-442.

MOREN A. 1983. AutoKa, applications at the NLS. Land materiverket, Professionel Papers, LMV - Rapport 13pp

PARKER D. 1989. Land Information Databases. Chapter II in Engineering Surveying Technology, Kennie TJM and Petrie G (Eds) Blackies Publishing Co. Ltd, 485pp

PETRIE G. 1990. Modelling Interpolation and Contouring Procedures. Chapter 8 in Terrain Modelling in Surveying and Civil Engineering. Petrie G and Kennie TJM (Eds), Whittles Pub, 112-127, 351pp

PETRIE G. 1990. Photogrammetric Methods of Data Acquisition for Terrrain Modelling. Chapter 4 in Terrain Modelling in Surveying and Civil Engineering. Petrie G and Kennie TJM (Eds), Whittles Pub, 26-17, 351pp

PETRIE G. 1990. Terrain Data Acquisition from Existing Maps. Chapter 7 in Terrain Modelling in Surveying and Civil Engineering. 85-111, 351pp

ROPER J. 1989. Three dimensional Applications in Geographic Information Systems. Taylor and Francis, 189pp

WEIBEL R. and HELLER M. 1990. A Framework for Digital Terrain Modelling. Proc. 4th Int. Symp. on Spatial Data Handling, Zurich, 219-228

CHAPTER 4

RECONNAISSANCE STAGE CASE STUDY - DETECTION OF SOLUTION FEATURES USING AIRBORNE REMOTE SENSING

4. RECONNAISSANCE STAGE CASE STUDY: DETECTION OF SOLUTION FEATURES USING AIRBORNE REMOTE SENSING

4.1 BACKGROUND TO STUDY

The location of zones which contain either surface or subsurface subsidence features is a topic which has aroused considerable interest in Engineering Geology for many years. Several reasons may be cited to account for this concern. Firstly, the location of subsidence features is an extremely difficult task. Whilst many techniques have been used to locate such feature they all have limitations which restrict their application to very specific circumstances. Secondly, the consequences of not detecting these feature may be catastrophic. Many examples across the world may be quoted to illustrate the disastrous nature of events which may occur if the features are undetected. Simons (1977), for example, reported in South Africa the collapse in December 1962, of a three level crusher plant at a gold mine in South Africa, into a sinkhole. This resulted in the loss of 29 lives. Similar reports have also been recorded in the United States. Sowers (1977), reported the collapse in June 1972 of both lanes on Interstate Highway 4, north east of Lakeland, Florida. Inspection of previously obtained aerial photography disclosed a circular depression that was centred on the central reservation. The cause of the collapse was subsequently found to have been due to the presence of a subsurface cavity. In the United Kingdom Edmonds (1987), reported several cases of subsidence and collapse features which have been encountered during engineering investigations in the South East of England.

There is therefore considerable justification for examining methods of locating such hazards and, in particular, for examining in detail the possible benefits which may accrue from utilising remote sensing techniques.

The overall objective of this Case Study is therefore to consider the role of remote sensing in the process of detecting subsidence features, with particular reference to the United Kingdom.

Section 4.2 begins by defining certain terms which will be used in the thesis. In particular the nature of subsidence features. Section 4.3 reviews and critically assesses previous literature which has discussed this topic. This is followed by Sections 4.4 - 4.6 which discuss the characteristics of the test sites which were chosen in order to assess the performance of three airborne remote sensing systems; aerial photography, multispectral scanning and thermal infrared linescanning.

The results of the interpretation and processing of the airborne remote sensing data are then presented and finally in section 4.7 a series of conclusions are reached for this type of operation and a series of recommendations for future work are presented.

4.2 DEFINITIONS OF SOLUTION SUBSIDENCE FEATURES

Solution features are present in any carbonate rock as a result of dissolution, caused by acidic surface or ground water. Near surface solution produces funnel like depressions within the rock, and these may be partially or completely infilled by loose or soft materials derived from the cover deposits. Alternatively voids may form deeper within the bedrock. The infill to solution features is likely to be more compressible than the surrounding carbonate rock . If any voids are present any changes in water infiltration or loading can lead to sudden or unexpected ground collapses.

In order to discuss the forms of solution subsidence features which occur on chalk outcrops it is necessary to have a series of well defined and widely accepted definitions of the most commonly occurring features. Unfortunately this is not the case, and examination of the literature becomes very confusing because of the varied and often incorrect usage of several terms to describe essentially the same subsidence feature. For example, the British Standards Code of Practice for Foundation Design (CP 2004: 1972)

considers the term 'sinkhole' and 'swallow hole' to be synonymous: "in places (the joints) are connected to the surface by swallow holes (sinkholes) which may be masked by surface deposits". Similarly Lake (1975) uses the terms interchangeably: "in common with other limestones, pipes and swallow holes, sometimes called sinkholes, do occur in the chalk". Other authors have also used the terms shakeholes, swallet, swilly, solution pipe, watersink and watershallow to describe the same feature which has been referred to by others as a sinkhole or swallow hole. Thus, in order to avoid further confusion it is suggested that the definitions developed by Edmonds (1983) be used to describe the forms of the solution features which occur in chalk regions. Edmonds identified three distinct forms:-

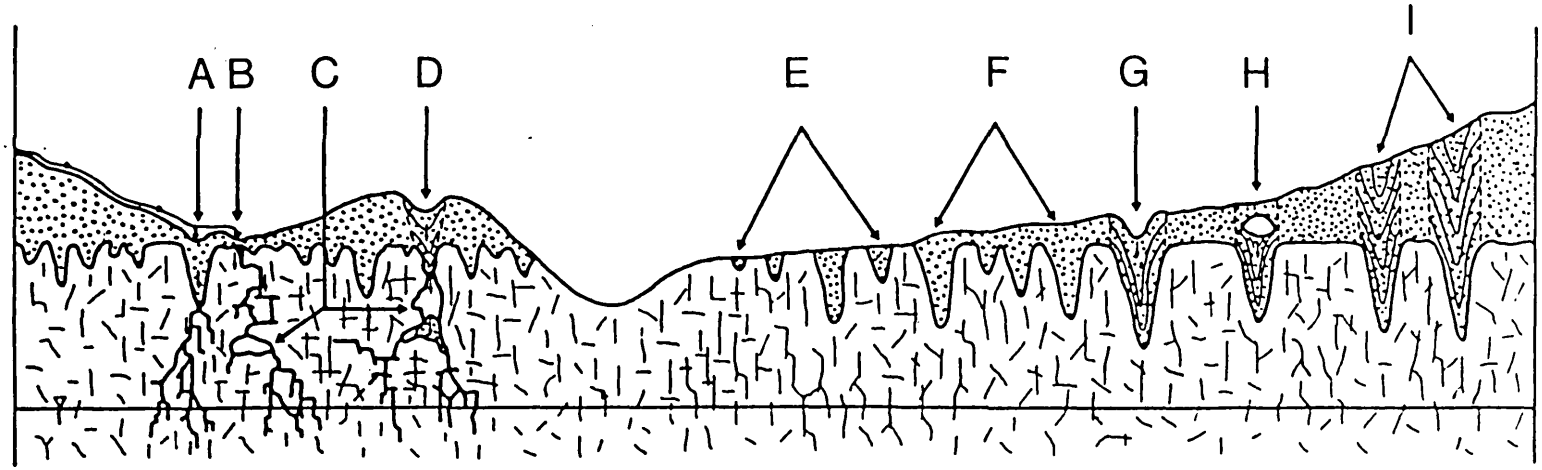
(a) Doline; a closed cone or bowl shaped depression. However, unlike other authors who have used this term, Edmonds suggests the term should only be used "solely as a surface morphological descriptive term".

(b) Solution Pipe; a cone or pipe-like cavity in vertical section typically infilled with overlying deposits that have subsided into the cavity created by dissolution of the soluble host rock. Solution pipes may, or may not be associated with a surface doline. Several types of solution pipe are identified and these are illustrated by Figure 4.1.

(c) Swallow Hole; a surface feature where a void in the chalk continuously or intermittently "swallows" wholly or partially a surface stream. Two different forms of swallow hole are identified and these are also illustrated by Figure 4.1.

4.3 PREVIOUS WORK

The use of airborne remote sensing techniques for the detection of surface, and, to a lesser extent, subsurface subsidence features has been a subject of considerable interest for several years. A wide variety of remote sensing systems have been applied



SCHEMATIC DIAGRAM TO ILLUSTRATE RANGE OF NATURAL SOLUTION FEATURES PRESENT ON LIMESTONE

A: Swallow hole where stream water permeates down through pre-existing solution pipe below surface hollow.

B: Swallow hole where stream water permeates down through solution widened joints and bedding

C: Solution cavity formed within chalk by solution widening of joints and bedding as ground water permeates downwards to reach the water table

D: Solution cavity formed by vadose or phreatic dissolution, where roof has become destabilised causing collapse into cavity, forming collapse doline at the surface

E: Solution pipe remnants, downslope from cover deposit margin, but infilled with disturbed cover deposits

F: Solution pipe formed below thin cover deposits, shallow zone above pipe mouth being disturbed by solution subsidence

G: Solution pipe present beneath cover deposit, where upward migration of metastable ground conditions has resulted in ground subsidence forming a subsidence doline at the ground surface

H: Solution pipe present beneath cover deposit, where downward solution subsidence movement of the cover deposit has ceased because of the presence of a resistant later, while material below this has continued to subside causing a metastable cavity

I: Solution pipe formed below thick cover deposits, zone above pipe mouth being disturbed by solution subsidence

Figure 4.1

to this problem including aerial photography and other systems which can record regions of the electromagnetic spectrum outside the visible and near infrared. The earliest attempts were concerned mainly with the potential of different photographic methods, and this was extended to include multispectral scanners, thermal infrared linescanning and radar systems such as microwave radiometry and side looking radar. A comprehensive review of this work can be found in Kennie and Edmonds [5]; pages A131 - 154 and Table 4.1 summarise the literature which has considered the use of thermal IR linescanning.

4.4 DATA ACQUISITION

4.4.1 Objectives

The results of a research programme to assess the success of airborne remote sensing for the detection of surface and near surface solution features in the Chalk of Southern England is reported in Kennie et al [6]. The Chalk provides an important example of a soluble carbonate rock since, because it outcrops over approximately 15% of the area of England, it is of considerable economic importance. Thus a great many foundations and a number of major highway projects are constructed on sites underlain by the Chalk each year.

The specific objectives of the research were;

i) To specify the geological conditions in which solution features occur.

ii) To critically assess the literature which had considered the use of remote sensing for such an application.

iii) To examine the use of multispectral scanner and thermal line scanning systems for solution feature detection.

iv) To assess the potential of new thermal sensing devices such as the Barr and Stroud IR18 thermal video frame scanner.

SUMMARY OF PREVIOUS RESEARCH: THERMAL IR LINESCANNING

Authors (Year)	Data Acquis.		Instrumentation			Environmental Conditions During Data Acquisition	Image Processing	General Comments and Level of Success
	Date	Time	Manuf.	Spectral range (μm)	Flying height (metres)			
Coker[24] (1969)	6th Sep	0647	Univ. of Michigan MSS	8-14	600	Data obtained during a drought	Density sliced (optically)	Data combined with near IR MSS data on veg. stress * * *
Metalucci + Abdel Hady[26] (1969)	-	-	Bendix LN-1	3-5.5	-	-	-	- * *
Noble[27] (1972)	April Dec.	700 - 1700	Univ. of Michigan MSS	4.5-5.5 10-11 8-13.5	600 and 1500	Raining during Dec. night-time flight	None	Ground radiometer measurements obtained *
DeLoach[28] (1972)	Nov.16	Night	-	8-14	300	-	Mainly visual but also microdensitometry	Interp. followed by boring * * *
Newton et al[15] (1973)	Jan.18 Mar.14	day/night day/night	-	8-14	500	Night flights suffered from effect of mist	None	Used primarily for lineament mapping *
Stohr and West[25] (1974)	Apr.7, 18, Dec.14, 15	1100-1200 2300-0200	Univ.of Michigan MSS	8-14	600 and 1500	-	Supervised and cluster analysis	Data combined with MSS *
Rib et al[29] (1973)	Mar.24 Mar.27 Apr.4	midday and night	Texas Ins. RS-14 ERIM MSS	8.2-10.7 9.4-12.1 8-14	450	Snow cover melted between 27 March and 4 April	Thermal contouring Thermal ratioing	Ground based radiometer readings *
Rinker[30] (1975)	Jun.11	0100	-	8-14	-	Clear sky, little wind	None	- * *
Harvey et al[31] (1977)	Mar.8, 9 Jun.20, 21	1200 and night		8-14	500	Data obtained during cloudy weather	None	Midday late summer suggested as optimum time *
Bakker et al[32] (1978)	-	Predawn	Texas Inst. RS-25	3.5-5.5 8-14	1000	-	None	Used in conjunction with API * * *
Warwick et al[33] (1979)	-	Predawn	Texas Inst. RS-25	3.5-5.5 8-14	1000	-	None	Used in conjunction with API * * *
Marino et al (1983[34] 1984)[35]	Nov.	Predawn noon	Daedalus 1230	1-2 3-5 9-11	800	-	Density slicing, texture analysis, ratioing	- * * *

KEY * of limited use ** useful *** very useful

Table 4.1 (See Kennie and Edmonds (5) for references)

v) To determine the optimum environmental and geological conditions for each detection technique.

vi) To determine the most appropriate image processing techniques to use in order to detect solution features and

vii) To compare the results from airborne remote sensing with those from conventional techniques of ground investigation.

4.4.2 Case Study Data Acquisition Periods

Airborne remote sensing data for the project was obtained during the period between April and September, 1986 as part of the Natural Environment Research Council (NERC) Airborne Remote Sensing Campaign. In order to examine the influence of seasonal variations and time of data acquisition on the detection of subsidence features a programme of multitemporal data acquisition was designed. The original specification indicated that data should be acquired during the months of April, July and September between the following time periods:-

i) 2000 and 0530 hours.
ii) 1030 and 1130 hours.
iii) 1400 and 1500 hours.

While it proved possible to obtain data over both sites during the campaign it was not possible to acquire coverage during the time periods specified.

4.4.3 Optimum Time of Day for Data Acquisition

In the past, pre-dawn has tended to be the preferred period for data acquisition for thermal data, primarily because the effects of direct downwelling radiation are at a minimum (Lynn 1984). However, according to Lynn, this may be the worst time to acquire data and Lynn suggested that daytime imagery acquired during mid-morning or mid-afternoon may be more appropriate for the following reasons. As the sun rises above the horizon, solar radiation increases heating up the ground. By mid-morning

the solar radiation is not at its maximum (and so degrading effects are reduced), but sufficient upwelling and downwelling radiation exist for both thermal and MSS images to be recorded. By mid-afternoon the atmosphere and the top surface of the ground cools. This causes the diurnal heat stored in the ground to move to the surface and causes maximum emissivity and temperature differentials so highlighting the sub-surface features on the resulting images. The imagery is also aided by the sun being at a low zenith angle which enhances subdued effects by shadows, especially if the flightline is perpendicular to the sun.

In the case of solution feature detection it was hypothesised that mid-afternoon was the optimum data acquisition period because of the reduction in downwelling radiation, the enhanced emissivity of the ground surface and the low sun angle.

4.5 TEST SITES

Two test sites were chosen for the investigations: -Swanley, Kent and Southover Heath, Dorset.

The sites were located in regions which exhibit a high density of solution features. In the Southover Heath area the features are primarily subsidence dolines, whereas, in the Swanley area the features are primarily solution pipes.

4.5.1 Swanley

The Swanley site can be found on Ordnance Survey 1:50,000 map sheet 177, map reference 539 692. It is an area of upper chalkland which has been completely covered with Blackheath beds. The adjacent motorway, is partly built over an area of the exposed chalk and solution features became evident at the time of construction. There are no surface features shown on the geological map, but there is evidence of there having been a drop in the water table. Several wells are marked on the maps of the area and are indicated as being dry.

4.5.2 Southover Heath

The Southover Heath site can be found on Ordnance Survey 1:50,000 map sheet 194, grid reference 789 924. It is an area of upper chalkland. At the northern end of the site, the chalk is overlain by an undefined thickness of Reading beds, with some plateau gravel. Solution features exist, in the main, on the sandy Reading beds. There is no evidence on the Southover Heath map to suggest that there has been any drop in water table, which may be a cause for initiating dissolution of the chalk. The site exists on top of a chalk escarpment running parallel with two rivers, the Piddle and Frome. A spot height of 95 metres (789 922) indicates that the hill top is some 65 metres above the level of the rivers.

Collapses have however been found to be numerous in parts of Dorset. Sperling et al (1975) records that in Bagshot beds on top of Bronkham Hill, "beside the largest tumult (SY624 873) a hole about a foot (0.3 metres) in diameter appeared on the surface. Beneath it was a large cave at least 20 foot (6 metres) in diameter and about 20 feet (6 metres) deep". Further "in a field about 630 yards (576 metres) west-south-west of the Hardy monument a solution pit was formed".

4.6 DATA ANALYSIS

4.6.1 Air Photo Interpretation

a) Southover Heath Site

The area of interest was examined on each run of panchromatic aerial photographs and any suspected solution features were identified. Features were identified most clearly from the photography which was taken in the early morning on the 19 September 1986.

b) Swanley Site

The Swanley photographs were visually examined in a similar manner to those obtained over as the Southover Heath. The April and

September photography indicated numerous circular features. On balance the autumn imagery provided to be more useful for the discrimination of possible solution features.

4.6.2 Thermal IR Linescanning

a) Southover Heath

Data was available from flights at three different times. Data was recorded at 0845 hours on July 17, 1986, 0920 hours on September 20, 1986 and 1600 hours on September 20, 1986. Two runs were available for each date. For each set of data, images of linear, equalised, normalised, piecewise and unstretched displays were recorded in photographic form as slides.

Unstretched data of the area illustrated six very large clearly defined features in the southern area, plus many other less well defined circular "shadows". The imagery obtained on the 20 September, at 0920 hours, was particularly well defined showing these six main features. The ground between these features appeared free from any features. The September, late afternoon imagery is however by far the clearest and indicates all the features.

The most appropriate time for normal, linear and equalised stretched imagery appeared to be the afternoon of September 20th, probably because the ground has had time by the afternoon to absorb more solar energy, has a greater emissivity, and a greater emissive contrast. July imagery is of limited use for the normal, linear and equalised stretches since the denser vegetation cover tends to give the area a uniform emissivity, so reducing the contrast.

b) Swanley

Data was available for flights at four different times for this site and were recorded at 1130 on 28 April 1986, 1330 pm 28 April 1986, 0930 on 17 July 1986 and 1330 on 20 September 1986.

The optimum time for unstretched, normal, linear and equalised stretched data was considered to be the afternoon data of 20 Septmber. The April afternoon data was also very clear in the equalised stretch.

c) Comparison of Thermal Data on Swanley and Southover Heath

The 1600 hours September 20th images proved to be most useful on the Southover Heath data, as they also did, on the 1330 hours September 20th data, of the Swanley site. The most useful contrast stretches on the Southover Heath data for the detection of features, were the equalised and linear stretched data. On the Swanley images the features of interest were also clearest on the equalised stretch. Figures 4.2 and 4.3 illustrate one specific example from each site of the Imagery obtained.

4.7 CRITICAL ASSESSMENT

The interpretation and analysis of the data acquired during this case study was limited to a visual interpretation of the unprocessed raw digital data. One of the main limiting factors was the inability to perform further image processing techniques to the data and investigate the use of more sophisticated classification techniques involving the multiple MSS bands. To do so would have required some initial pre-processing and the following section highlights some of the processes which could be applied to the raw data.

4.7.1 Radiometric Restoration and Correction

The radiance recorded by an airborne linescanning sensor such as that used in this case study is stored as a digital number (DN), typically in the range 1 to 255. The gain for each wavelength recorded can be preset to one of five settings in order to adequately utilise the 0-255 DN intensity range and to achieve maximum sensitivity under the prevailing viewing conditions of the flight. As a result of this, values of, say, 100 DN in one wave band do not necessarily represent the same radiance values of 100 DN, in another wave band. Consequently, ratios or further

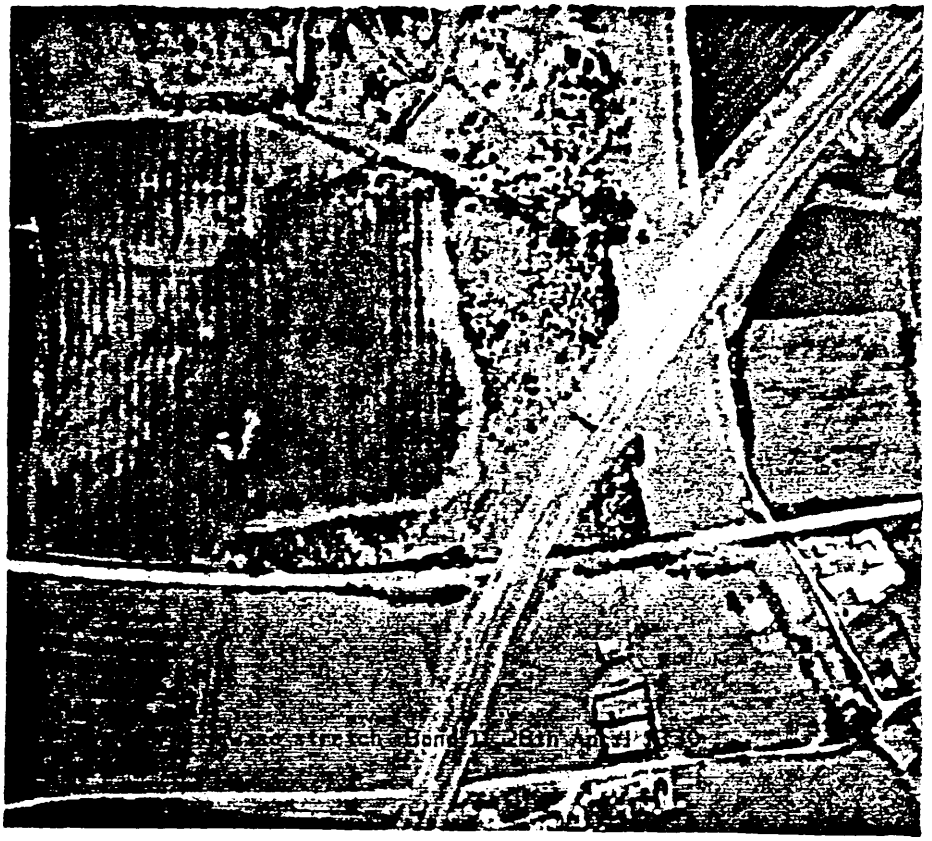


Figure 4.2
Swanley: Thermal Imagery 28th April (pm)



Figure 4.3
Southover Heath: Thermal Imagery 17th July (am)

mathematical transformations cannot be applied directly to the unprocessed data. To do so requires corrections for the influence of gain setting (calibration) atmosphere conditions and environmental variations. Radiometric correction involves the rearrangement of the DN in an image so that all areas of the image have the same linear relationship between the DN and either radiance or backscatter. This is important since only through such a process can the spatial or temporal changes which may be observed be considered as significant and not artifacts introduced by a change in the radiometric calibration of the sensor, or by atmospheric and illumination differences.

Corrections should be applied to compensate for the following effects on the spectral signatures of objects:-

i) Illumination conditions
ii) Observation geometry
iii) Atmospheric phenomena
iv) Topographic variations

Each of these four types of correction will usually have the same three prerequisites:-

i) Radiometric correction
ii) An atmospheric model
iii) A target reflectance model

In order to compare image data acquired from different sites, or at different times, radiance values require to be standardised and converted to reflectances. Such corrections centre on the adjustments of illumination and observation geometry to some common configuration with respect to a horizontal terrain. These adjustments effectively involve different paths of solar radiation through the atmosphere and different samplings of the bi-directional reflectance distribution functions of terrain cover types.

4.7.2 Selection of Optimum Band Combinations

As previously mentioned, the scanner detects 11 different bands of the EM spectrum including the thermal infrared. For the process of image analysis the choice of individual bands and band combinations play a vital part in maximising the information obtainable from a scene. Up to three bands can be processed together giving a possible 165 different combinations for the 11 channel data. It is therefore important to use the bands which do not all contain the same information or are correlated.

There are three general approaches to the selection of the optimum combination band. The first is to deduce the most likely combination by knowing the type of site, information required, etc. From this starting point, different variations could then be assessed on a "trial and error" basis to produce the best possible image. However, this method requires some degree of experience (and luck) and may lead to information being overlooked.

A second and more efficient method is to perform a statistical analysis of the correlation between bands to obtain the Optimum Index Factor (OIF) (Chavez et al, 1982). This provides a ranking of possible three band combinations in terms of their uncorrelated spectral data content. Chavez et al (1982) devised the OIF statistic which provides a measure of the uncorrelated information content of all possible 3-band combinations. The statistic was originally developed for Landsat MSS data but the statistic could equally be applied to any set of independent variables. The OIF is calculated for each 3-band combination by dividing the sum of the standard deviations by the sum of the absolute values of the correlation of the coefficients.

The third method, suggested by Sheffield (1985) uses the volume of pairs of principal axes (when the DN values of bands are plotted against one another) as a measure of information content in order to select those bands that are highly correlated. In a similar fashion to the OIF method this algorithm was developed for Landsat imagery but can be used for ATM data. The basis for the

procedure is principal component analysis, and involves the formation of a 11 by 11 variance-covariance matrix for a scene. Any triplet of bands will be represented within this 11 by 11 matrix by a 3 by 3 submatrix. If, using a three dimensional graph, the DN of each of the three bands were plotted against each other the result would be an ellipsoid with three principal axes. To discourage the selection of pairs of bands with high correlation an ellipsoid of maximum volume is selected. This is achieved by computing and ranking in order, the determinants of each 3 by 3 principal submatrix in the original 11 by 11 matrix. The band triplets associated with these determinants are ranked in order of decreasing overall information content.

The best band selection varies remarkably depending on whether or not the data has been calibrated. Therefore unless the data is calibrated it is not possible to determine the most suitable bands for analysis using either of the methods described. (Donoghue and Shennan, 1988).

4.7.3 Practical Detection Difficulties

A number of practical detection problems were encountered during the investigation which could severely limit the usefulness of airborne remote sensing for this type of application. Among the more significant were:

(i) masking of areas of potential subsidence by heavy vegetation and areas of woodland,
(ii) the effect of shadows cast along the boundaries of areas masking features of interest,
(iii) fields that have been ploughed for many successive years removing previously existing indications of solution features, and,
(iv) urban development which has also destroyed tonal patterns associated with the presence of solution features.

Other problems can arise when attempting to identify and classify the type of feature which can be detected on the images. For solution features which have no surface expression, it is

necessary to be cautious of other anomalous ground tonal patterns such as those caused by archaeological remains and agricultural crop watering patterns. One particularly difficult feature to interpret is a surface depression. On the Chalk outcrop of southern England there are numerous man-made surface depressions which have been created by the mining or quarrying of chalk for application on fields to neutralise acid soils. After a period of deterioration these old workings frequently assume a similar appearance to a naturally formed doline. Thus the user of remote sensing imagery should appreciate that the surface morphology of an identified circular ground depression can, by itself, often be insufficient to infer its origin. Other forms of spatial distribution analysis should help to discriminate between the presence of natural or artificial features. In the end, however, final proof of the origin of a potential solution feature can only be confirmed by direct ground investigation such as drilling a borehole or excavating a trial pit through the feature.

4.8 FUTURE WORK

As indicated in section 4.7.3 one of the primiary limitations of an investigation outlined in this case study was the lack of direct ground truth to validate the results of the remote Sensing investigations. Future work on this topic should therefore include the opportunity to carry out a detailed site investigation of the site to provide evidence of the existence of solution subsistence. The following proposal outlines one possible extension of the project which would enable this limitation to be overcome. The details form part of a research grant application to the Science and Engineering Research Council (SERC) which was made in 1987 by the author and two colleagues at the University of Surrey; Dr C.R.I. Clayton and Mr M.C. Matthews.

4.8.1 Objectives

The overall purpose of the proposed investigation would be:-

(a) to assess quantitatively the ability of a range of existing

remote sensing and geophysical techniques for the location and investigation of solution features in the chalk of Southern England, and

(b) to determine the cost effectiveness and most suitable conditions for the application of each technique.

4.8.2 Methodology and Programme

In order to guarantee the success of the project it would be desirable to carry out intensive ground investigations at the selected sites by auger boring before the indirect methods (i.e. remote sensing and ground geophysics) of investigation were tested.

The research programme would consist of three primary phases;

(a) Phase 1 - Desk study, subsidence hazard assessment and, remote sensing, leading to a prediction of solution feature locations.

(b) Phase 2 - Geophysical traversing of 2 sites, leading to a further prediction of solution feature locations.

(c) Phase 3 - Subsurface investigation to determine ground truth, and final reporting.

The programme is shown in Figure 4.4. The activities proposed under each stage would be as follows:-

Stage 1 - Desk study

- examine topographic maps, geological maps and literature, and existing air-photograph cover of three sites.
- subsidence hazard assessment using predictive model of Edmonds et al (1987) to assess subsidence risk on each site.

Stage 2 - Selection of 2 target sites

Stage 3 - Land Survey

- carry out detailed topographic survey for each site, to define

STAGE	ACTIVITY
	INDUCTION OF RESEARCH ASSISTANT
1	DESK STUDIES Desk study Subsidence hazard assessment Analysis of data
2	SELECTION OF TARGET SITES based on rotary augering
3	LAND SURVEY
4	REMOTE SENSING STUDIES Thermal Video Framescan (TVFS) survey Balloon mounted thermovision survey Ground temperature measurements Acquisition of oblique multispectral aerial photography Process and analyse thermal remote sensing data Interpretation of aerial photography Reprocess and analyse NERC MSS data
5	REPORT PREDICTIONS BASED ON STAGES 1,3 & 4
6	GEOPHYSICAL SURVEY Magnetometer survey Microgravity survey Electrical conductivity survey Ground impulse radar survey Process and interpret geophysical data
7	REPORT PREDICTIONS BASED ON STAGES 1,3,4 & 6
8	FINAL SUBSURFACE INVESTIGATION Plan subsurface investigation Carry out rotary auger boring Carry out cross-hole seismic profiling Carry out laboratory index tests on borehole samples Process and interpret subsurface investigation data
9	ANALYSIS OF DATA COLLECTED DURING STAGE 1,3,4,6 & 8 Reprocessing and reinterpretation of desk study data Reprocessing and reinterpretation of remote sensing data Reprocessing and reinterpretation of geophysical data
10	PRODUCTION OF FINAL REPORT

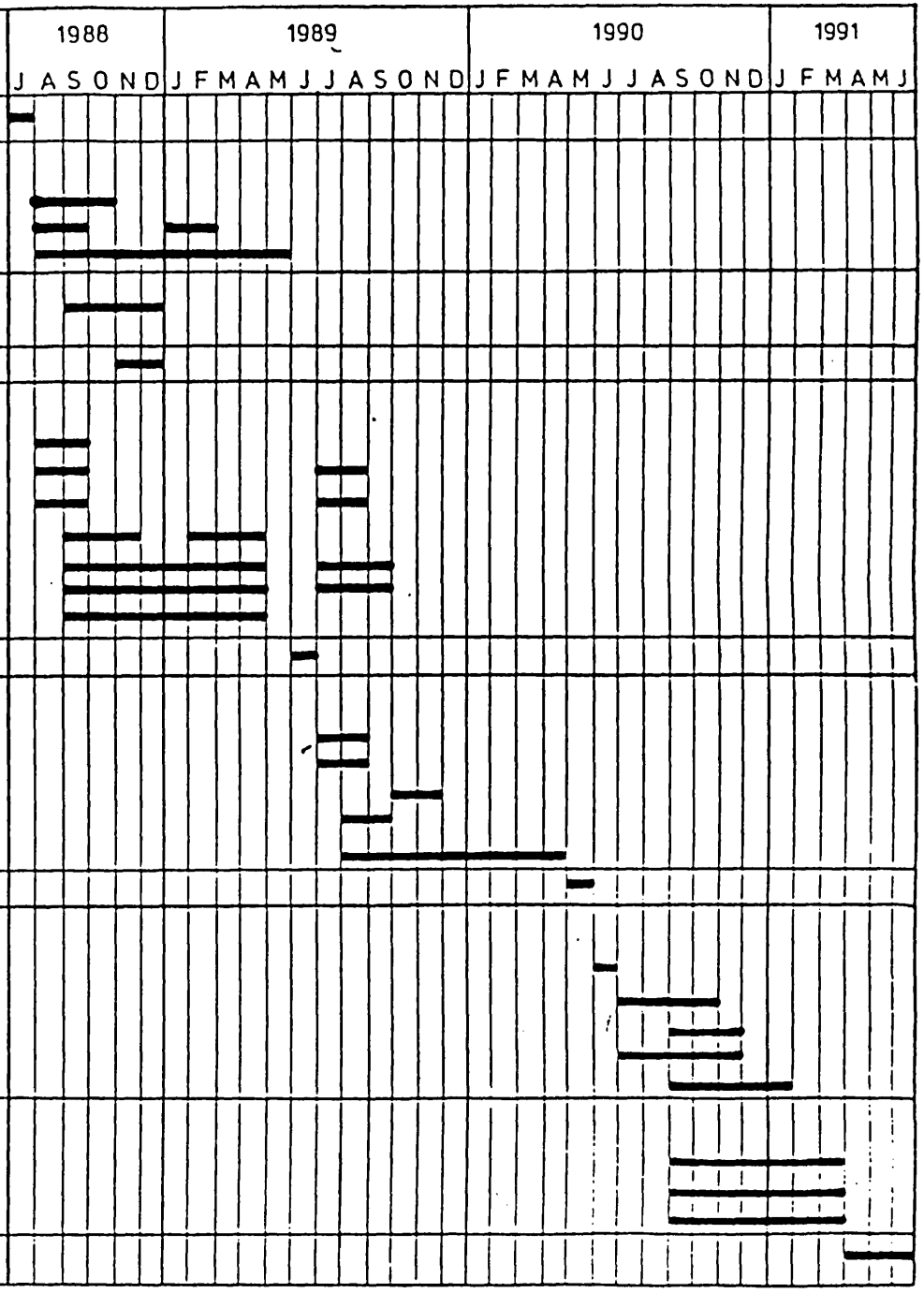


Figure 4.4

surface topography and give setting out control for later work.

Stage 4 - Remote Sensing Studies

- Thermal Video Framescan survey (TVFS) (from aircraft)
- balloon mounted thermovision survey
- ground temperature measurement for calibration of thermal infrared studies, process and analyse TVFS data, acquisition of oblique multispectral aerial photography, reprocess and analyse NERC MSS data.

Stage 5 - Report on Predictions based upon Desk Study and Remote Sensing

Stage 6 - Geophysical Traversing Surveys

- gridline surveys by the following methods:-
- proton precession magnetometer
- electrical conductivity meter
- ground impulse radar
- process and interpret geophysical data

Stage 7 - Report on Predictions based upon Desk Studies, Remote Sensing, and Geophysical traversing.

Stage 8 - Subsurface Investigation

- plan subsurface investigation
- direct investigation by auger boring on gridlines, and at locations of suspected solution features
- cross-hole seismic (geophysical) profiling of 2 major features in each site

Stage 9 - Comparison of predicted and observed frequency, sizes and positions of solution features on each site

Stage 10 - Final reporting

- detailed reporting of factual information obtained for each site
- explanation of interpretative methods used to process data
- examination of success rate and cost-effectiveness of each method used to locate solution features
- recommendations on suitable conditions for the application of each technique.

The project outlined above was submitted to the SERC and

considered for funding in April 1987. The project was graded as alpha however, funding was not provided on the first application. The letter from the SERC noted that "although the committee wished to make an award it is unable to do so from the very limited funds available".

4.9 CONCLUSIONS

Airborne remote sensing is recommended as a technique to be used for the detection of solution features at, or immediately after, the desk study stage of a site investigation. An assessment of the various remote sensing techniques indicates that aerial photography should generally be the first source of data to be used for the detection of solution features which have a surface expression. However, where it is necessary to use indirect evidence in order to infer the presence of solution features, it is suggested that airborne thermal IR and MSS imagery should be used to supplement the aerial photography. Indeed, thermal IR and MSS imagery are particularly helpful in detecting material responses of the ground to underlying karstic features where there is an absence of a surface topographic expression.

Several difficulties have been identified which can restrict the use of airborne remote sensing techniques, including the problem of determining the optimum conditions for obtaining imagery. Some recommendations for acquiring data have been provided and for future work concentrating on the inter-band spectral correlation of the MSS datasets and the need for ground investigation data to correlate with the remote sensing investigations.

For the future it is suggested that consideration be given to using airborne remote sensing data more extensively as a component of spatial subsidence hazard assessment modelling, such as that developed by Edmonds et al (1987). At present this particular form of hazard prediction is carried out manually, but the digital nature of the data lends itself to computer manipulation and it is suggested that the hazard prediction could be enhanced by using a Geographical Information System (GIS). This would effectively combine the spatial and spectral information from the remote sensing data with spatial data from existing map sources (topographic, geological etc.) and non-spatial sources of data (such as borehole information).

4.10 REFERENCES

CHAVEZ, P. S. (Jr), BERLIN, G.L., and SOWERS, L. B., 1982. "Statistical method for selecting Landsat MSS ratios", Journal of Applied Photographic Engineering, 8, p. 23.

DONOGHUE, D. N. M., and SHENNAN, I., 1988, "Calibration and processing of Morton Fen ATM data", Proceedings of the NERC 1986 Airborne Campaign Workshop, Natural Environment Research Council, pp. 175 - 186.

EDMONDS C.N. 1983. Towards the prediction of subsidence risk upon the chalk outcrop. Q.J. of Engineering Geology, vol 16, 4, 261-266.

EDMONDS C.N. 1987. The engineering geomorphology of Karst development and prediction of subsidence risk upon the Chalk outcrop in England. Ph.D. Thesis, University of London.

EDMONDS C.N., GREEN C.P. and HIGGINBOTTOM I.E. 1986. Subsidence hazard prediction for limestone terrains as applied to the English Cretaceous Chalk. Proc. 22nd Reg. Conf. of the Eng. Group of the Geol. Soc., Plymouth, 315 - 333.

LYNN, D. W., 1984, "The need for integrating ground thermal measurements with thermal aircraft imagery", Proceedings of the Conference on Integrated Approaches in Remote Sensing, European Space Agency (ESA), Special Publication, ESA SP-214, pp.327 - 334.

SHEFFIELD, C., 1985, "Selecting band combinations from multispectral data", Photogrammetric Engineering and Remote Sensing, 51, pp. 681 - 687.

SIMONS N.E. 1977. Slips, settlements and sinkholes. Inaugural Lecture, University of Surrey, 52pp.

SOWERS G. 1977. Failures in limestones in humid sub tropics. Journal of the Geotech. Div. of the ASCE, vol 101, GT8, 771-787.

SPERLING C.H.B., GOUDIE A.S., STODDARD D.R., and POOLE G.G, Dolines of the Dorset Chalklands and other areas in Southern Britain, Trans. Inst. Brit. Geog. 1977, pt.2.

CHAPTER 5

INVESTIGATION DESIGN STAGE CASE STUDY -

APPLICATIONS OF THERMAL VIDEO FRAME SCANNERS (TVFS)

5. INVESTIGATION DESIGN STAGE CASE STUDY - APPLICATIONS OF THERMAL VIDEO FRAME SCANNERS (TVFS)

5.1 BACKGROUND TO STUDY

The indirect measurement of the variations in the heat emitted by the earths surface is a surveying technique which has been used by engineering geologists for some time. Interpretation of such data can be used to infer the positions of areas of ground instability, zones of spontaneous combustion in areas with coal waste tips, positions of underground pipes, and cavities etc. The conventional methods of carrying out such surveys involve the use of thermal infrared line scanners (see Kennie [2] page A16).

Video based camera systems are an alternative to this and are being used by an increasing number of remote sensing users. These systems offer several advantages compared to the more conventional line scanner systems in terms of ease of use, speed of operation, portability and cost. Of particular interest are the thermal video frame scanning (TVFS) systems. These scanners, developed primarily for military purposes, differ considerably in design from the thermal infrared linescanning instruments traditionally used for thermal surveys in remote sensing. In the following Case Study one specific example, the Barr and Stroud IR18, is described together with the modifications necessary to mount such a system in a number of light aircraft. Methods of digitally enhancing and classifying single frames of data are then discussed. The Case Study considers one example of the use of the instrument for remotely determining ground temperature variations and highlights the potential of using such a device by highway maintenance engineers and engineering geologists.

5.2 PREVIOUS WORK

A summary of previous literature which has considered the use of thermal video systems previously is illustrated by Table 5.1 in Kennie et al [7], page A166.

TABLE 5.1: SUMMARY OF PREVIOUS LITERATURE - AIRBORNE VIDEO BASED DATA ACQUISITION SYSTEMS

SPECTRAL REGION	AUTHOR (S)	APPLICATION
(0.4 - 0.7um) Panchromatic	Mussakowski, 1983	Water Assessment
(0.4 - 0.7um) 3 band colour	Vlcek, 1983 Vlcek and Cheung, 1983	Forestry
(0.4 - 1.1um) 3 band 'false' colour	Meisner and Lindstrom, 1985 Everitt and Nixon, 1985 Nixon, Escobar and Menges, 1985	Rangeland Studies. Vegetation Studies
(0.55 - 1.1um) 4 band multispectral	Vlcek and King, 1984 Vlcek, King and Shemilt, 1984	Vegetation Vegetation Mapping
(3 - 5.6um, 8 - 14um) Thermal infrared	Tice and Larsson, 1978 Geraci and Caltabiano, 1983 Geraci, 1983 Clement, 1985 Amin, 1986	Heat loss Oil pollution Water quality Geothermal Mapping

5.3 DATA ACQUISITION

5.3.1 Design of Thermal Video Frame Scanners (TVFS)

A TVFS system consists of four basic components:

a) a scanning system which views the scene to be imaged and focuses the incoming infrared radiation,
b) a ferry-magnetic or semi-conductor detector which measures the thermal variations over the scene,
c) a CCTV compatible display system where the thermal variations are used to modulate the intensity of an electron beam, and
d) a video cassette recorder (VCR) to record the thermal imagery.

Unlike thermal IR linescanning systems which are designed specifically for airborne use, the TVFS systems under consideration are designed for both ground and airborne operation. Consequently, the optical configuration, scanning mechanisms and detector arrays differ from those used for thermal linescanning and this leads to significantly different imaging systems being employed.

The design of the Barr and Stroud IR18 used in this Case Study is discussed by Kennie et al [7], page A168. The optics of the IR18 are used to collimate the incoming radiation and, by using afocal telescopes enable a variety of magnification ratios to be used so producing various object field of views. Such an arrangement ensures that the incident radiation forms a bundle of parallel rays which can then be focused onto the frame scanning rotor. A concave mirror then re-collimates the image onto the rotating hexagonal rotor and subsequently onto the detector.

The detector used is an eight element Mullard 'SPRITE', although four elements are normally preselected for use. The scanning sequence is such that on each successive frame scan a different element (of the four active elements) scans the same part of the scene. A consequence of this is that if the image output is integrated over a number of frames, detector non-uniformity errors can be eliminated. Complex microprocessor controlled circuitry controls the synchronisation of the rotor

speeds and signal outputs and the transfer of the data onto video tape. A conventional U-Matic video tape recorder is used to store the imagery. Included in the scanner design is a thermal reference place which is placed at one end of the linescan locus so that referencing takes place at the scan rate. This arrangement enables the selection of five sensitivity ranges covering the following temperature ranges 2°, 4°, 8°, 16° and 32°.

5.3.2 Operation of Sensor in Light Aircraft

To date the Barr and Stroud IR18 TVFS has been operated in two light aircraft: a Cessna 172 (Figures 5.1, 5.2 and 5.3) and a Piper Aztec. Details relating to the use of the TVFS in the former case can be found in Ekin (1984,1987) and Kennie et al [7]. For some applications, however, it may be unacceptable to use the scanner from a small aircraft such as the Cessna 172, either because of the need for a more stable platform or the need to use the camera in conjunction with several other sensors. More recent experience of using the IR18 for defence and commercial applications by ERSAC Ltd, has involved using the IR18 in a purpose built camera mount in a twin engined Piper Aztec aircraft. For maximum flexibility of operations and integration with a standard aerial camera, a special adapter plate has been designed to enable the IR18 to be mounted and operated through a standard Wild RC8 aerial camera mount. The most significant advantage of this design is the capability for full yaw control; this enables higher precision to be achieved when flying parallel flightpaths for the acquisition of imagery over large areas.

The use of a light aircraft and an easily manipulated TVFS system of the type described offers several advantages to the user. Firstly, by using widely available light aircraft the operational costs of the survey can be significantly reduced. Secondly, once the site of interest has been located and the camera rig extended, the scanner operator can view, via a small monitor, the scene being scanned by the instrument. Thirdly, after the survey has been completed immediate confirmation that the data has been recorded successfully can be achieved by playing back the video tape, either in flight, or more commonly when the aircraft returns to base.

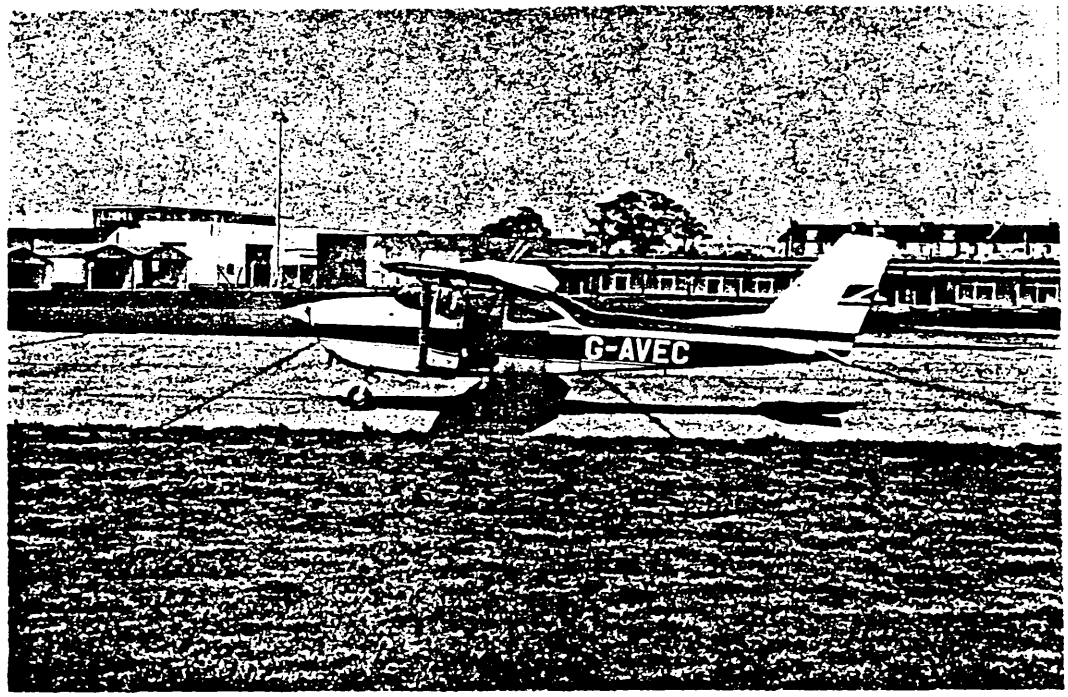


Figure 5.1
Cessna 172 Used for aerial survey

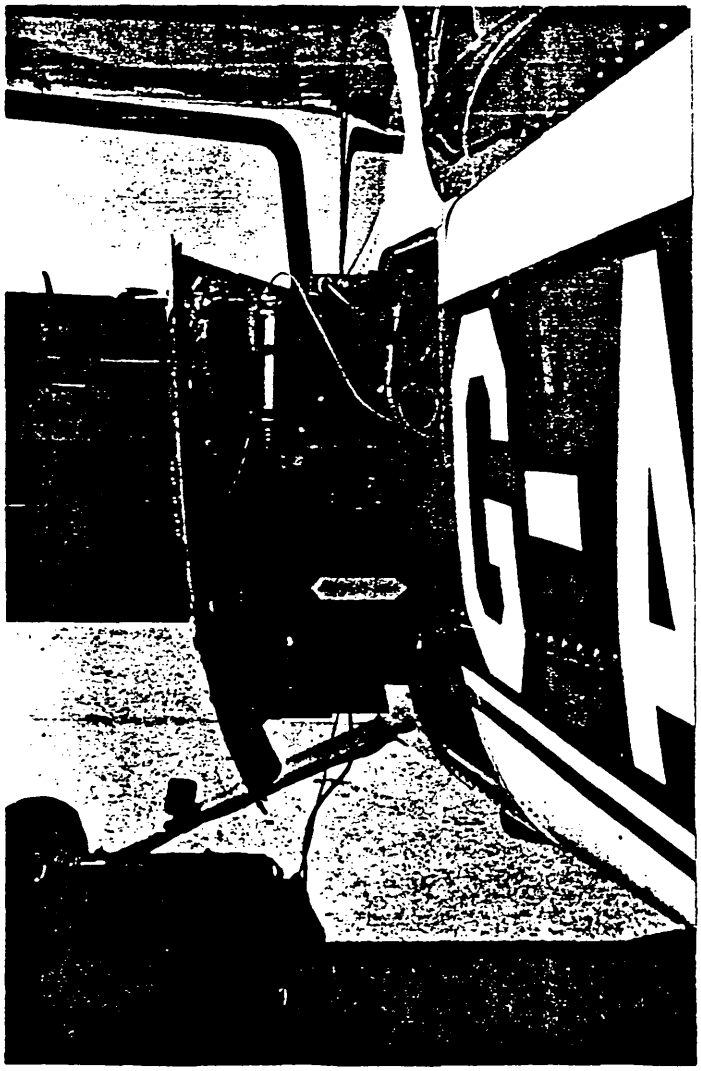


Figure 5.2
Barr and Stroud IR18 mounted in modified extending door frame

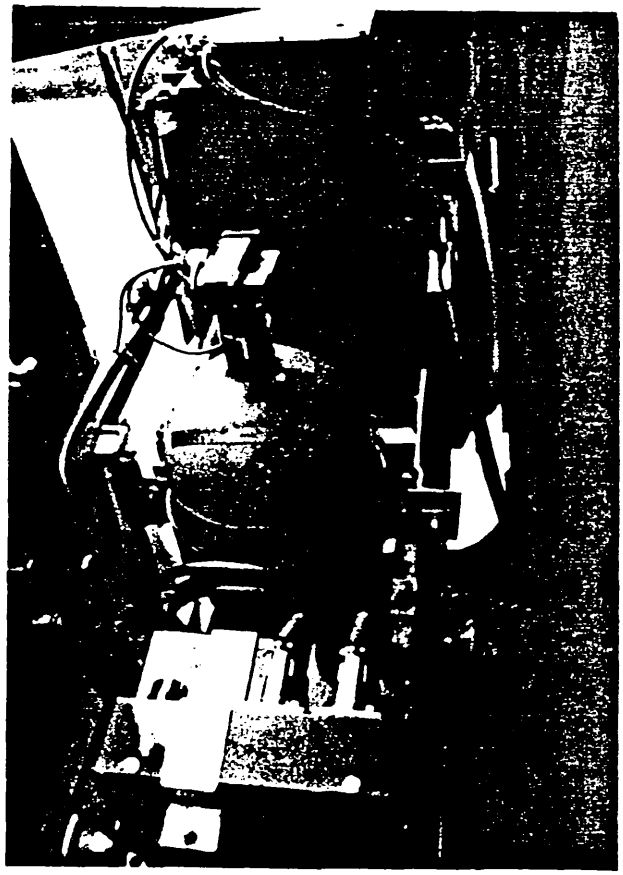


Figure 5.3
View from inside Cessna 172 illustrating IR18 and Cooling System

5.4 THERMAL MAPPING OF GROUND CONDITIONS

5.4.1 Hunterston Example

As part of a joint research project between the Universities of Surrey and Glasgow a preliminary flight of the Barr and Stroud IR18 for civilian remote sensing purposes took place during August 1984. The sensor was mounted in a Cessna 172 using the extendable rig illustrated in Figures 5.2 and 5.3.

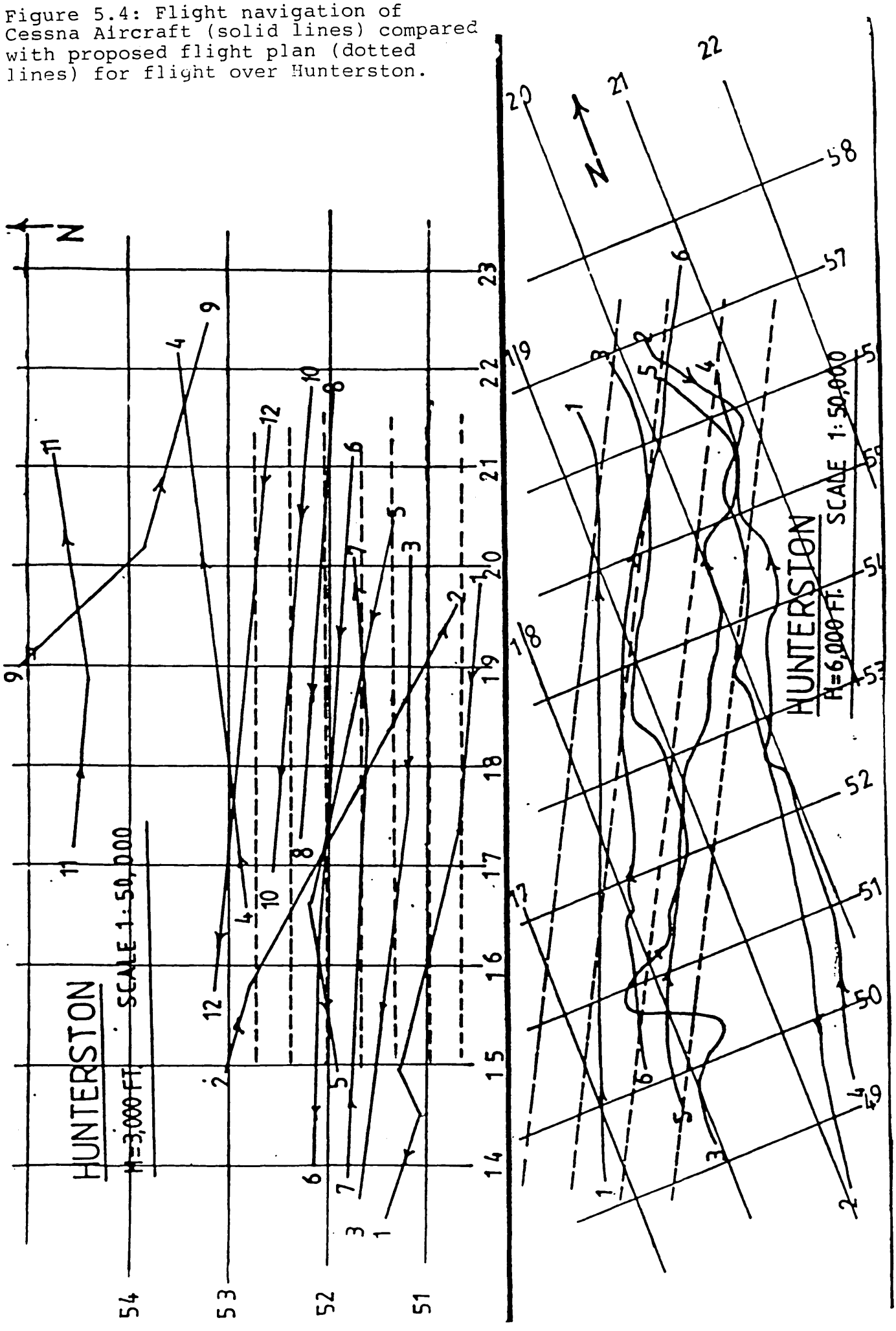
The flights over the Hunterston area were part of a larger survey which covered part of Central Glasgow. Two distinct flying heights were used; at 915m (3000 feet) and 1830 metres (6000 feet). The 6000 feet altitude flight was flown North to South whereas the 3000' flight was flown in a West to East Direction. Figure 5.4 indicates the area of the survey and the variations in flightlines. It can be seen that the reliance on visual navigation proved very inadequate and the variations on both flights were considerable. In terms of flight geometry using the 1.5x telescope the area of coverage was approximately 850m x 550m at 6000 feet and 425 x 275m at 3000 feet.

The flights were intended to take place immediately pre-dawn however because of early morning fog the flights did not take place until late morning. In view of the change of time the original objective to use the instrument to assess spontaneous combustion within coal waste tips was cancelled and the potential of the system for general hydrogeological and reconnaissance surveys was assessed.

5.4.2 Lothian Example

Each winter UK local authorities spend more than £100 million on measures to keep roads free of ice and snow. Furthermore, recent research has indicated that spending on rock salt, in particular, may be double that which is actually required (Thornes 1985). Thermal surveying is one of several new techniques which have been in the UK in recent years to help reduce these costs. Information provided by such surveys can be used not only to locate areas prone to icing but also identify ground sites where detailed temperature monitoring should be carried out by sensors embedded in the road surface.

Figure 5.4: Flight navigation of Cessna Aircraft (solid lines) compared with proposed flight plan (dotted lines) for flight over Hunterston.



A full report of a project to assess the use of the IR18 for such surveys is reported in Stove et al [18].

Optimum weather conditions for thermal mapping occurs in winter months when anticyclonic conditions and clear skies exist, giving maximum cooling to the road surface. These conditions also produce temperature inversions, the local meteorological effects of greatest concern to the highway maintenance engineer. Temperature inversions occur as a result of cold dense air draining from higher altitudes to settle in low-lying areas, producing an abnormal lapse rate where air temperature rises with increasing height. This condition may lead to the formation of frost or ice on the road surface. To assess the possibility of utilising airborne TVFS for such activities a survey was initiated in the Lothian region of Scotland. The survey was also not initiated until there was a high degree of confidence that the conditions outlined above would exist in the survey area.

The survey took place at 04.40 hours on the morning of 27 February 1986 towards the end of a prolonged cold period over the UK. Meteorological records at Edinburgh airport indicated that this period was the coldest recorded since 1948. Ground temperatures reported by the meteorological office ranged from +1°c on the coast to -8°c inland.

The survey was flown at a height of 800m at an airspeed of 160 knots, this enabled a complete recording of the survey on only 40 minutes of video tape. Navigation was accomplished using conventional radio location aids and a standard aerial navigation sight in combination with a monitor linked to the Barr and Stroud IR18 sensor. In order to enable road surface temperatures to be determined from the thermal imagery, calibrated ground temperature measurements were obtained by ground teams using digital thermometers. Observations were obtained at several road junctions where ice formation normally occurred. In addition measurements were also obtained at other randomly chosen interchanges and junctions at which it was possible to register the road temperature measurement with its location on the thermal imagery. In all cases temperature profiles were measured across the carriageway and slip roads. Reference air temperatures were also collected at 0.5m, 1.0m and 2.0m above the road surface to determine whether any local

eddy or current effects were present. Ground temperatures varied during the survey from +2 c to -8 c.

5.5 DATA ANALYSIS

5.5.1 Hunterston Example

The data acquired over the Hunterston area were initially stored on U-matic video tapes. Following prelimary editing 20 individual frames were chosen for digitising as 512 x 512 images. Digitising was performed using the National Remote Sensing Centres (NRSC) IPD 3000 Image processing system linked to a Quantel DFS 3001 digital frame store. Subsequent analysis of the individual frames was performed using a GEMS image processing system with GEMSTONE software.

The digitised frames were subsequently enhanced using the standard image processing techniques available within GEMSTONE including contrast stretching, de-striping and density slicing. Linear stretching and the influence of a 45 x 12 box filter to reduce the effect of a horizontal striping effect produced imagery with a high level of spatial resolution as is illustrated by Figures 5.5 and 5.6 which are two typical examples of the imagery produced by such processes. Further unquantitative density slicing analysis of a number of thermal plumes in the sea surrounding Hunterston provided images which clearly indicated thermal variations.

5.5.2 Lothian Example

In this second example of the use of TVFS imagery it was possible to apply more sophisticated analyses to the imagery including a semi-quantitative density slice to provide discrete temperation variations.

Specific frames of video imagery covering those areas where ground temperatures had been measured were selected and digitized. The relationship between the measured ground temperature and the digitized image grey scale was then used to create a semi-quantitative density slice at 1½c intervals (Figure 5.7). The limits of the density slice were subsequently applied to successive frames of the video imagery.

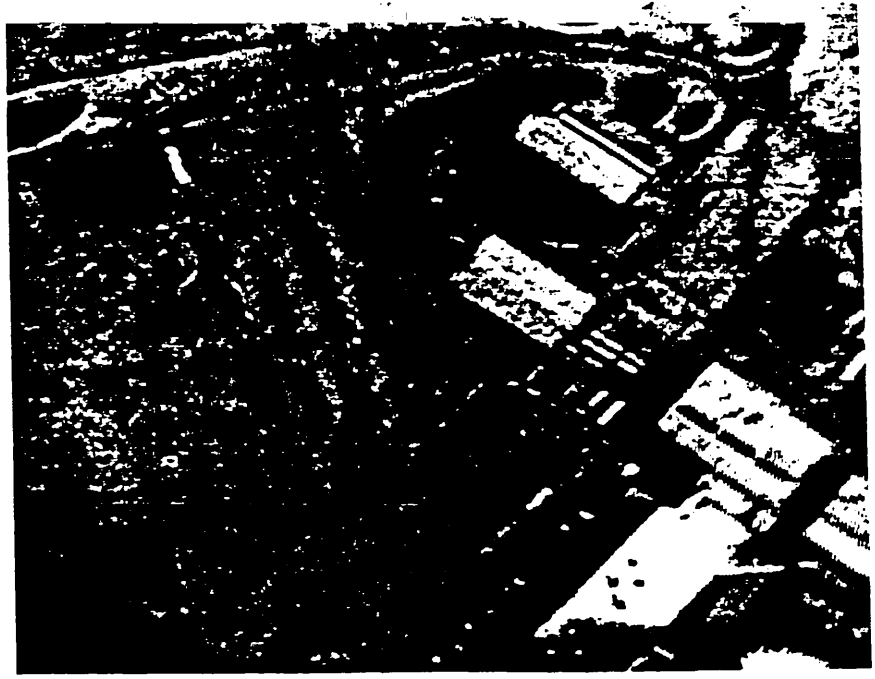


Figure 5.5
512 x 512 Pixel Digitised TVFS Image from Hunterston Area illustrating high spatial resolution of imagery

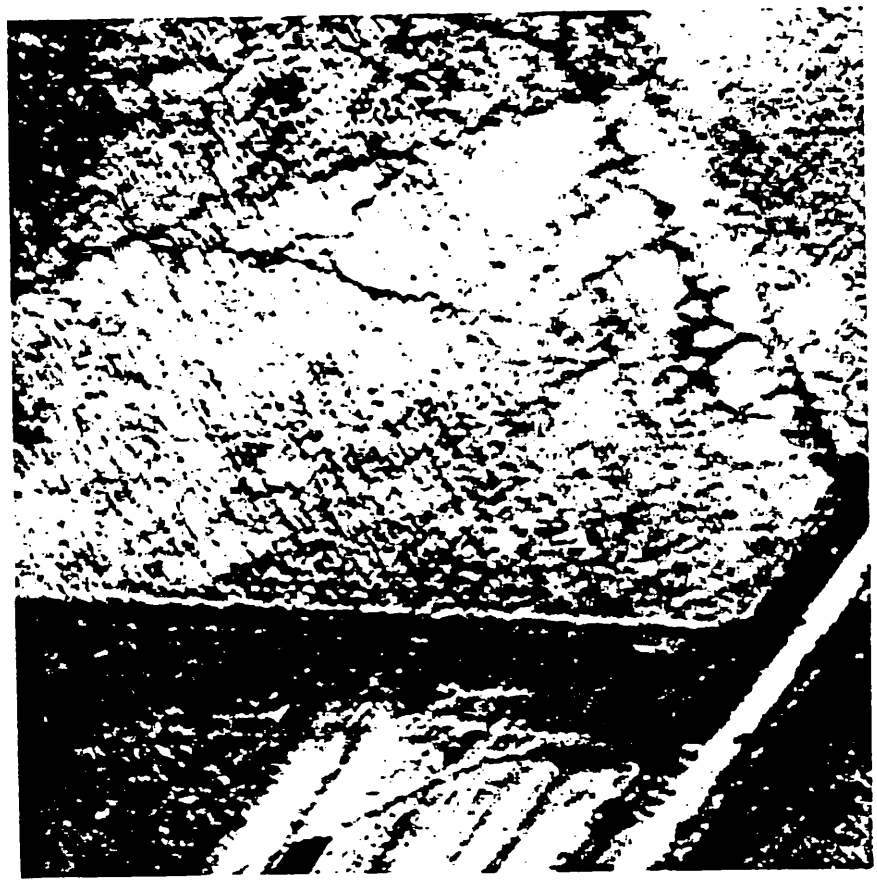


Figure 5.6
512 x 512 pixel digitised TVFS image from Hunterston Area illustrating hydrological potential of imagery



Figure 5.7
1½c Density sliced image of TVFS Scene
over M9 interchange, W. Lothian, Scotland
(original in colour)

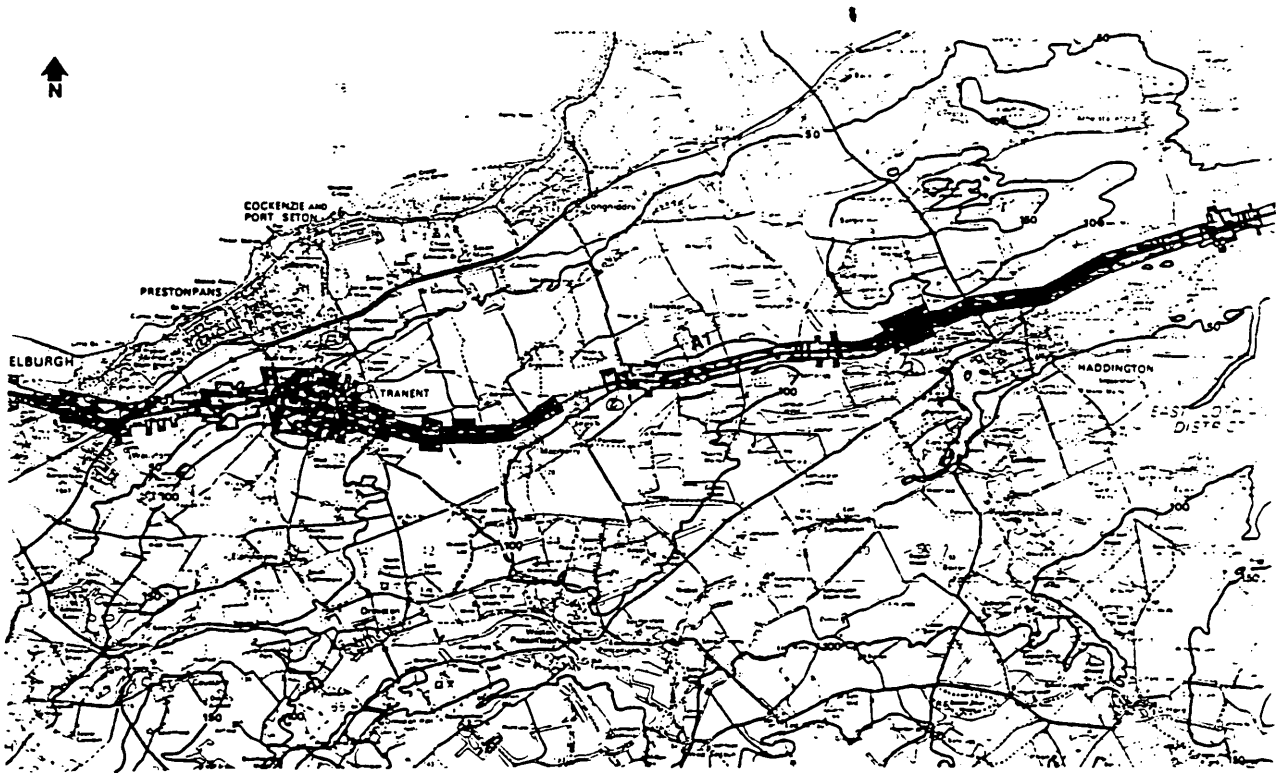


Figure 5.8
1:50,000 map Summarising results of TVFS
Survey and indicating temperature variations for section of
M9 motorway (original in colour)

Visual interpretation of the video tape was then carried out and transferred to 1:10,000 scale maps of road sections and finally onto 1:50,000 scale maps (Figure 5.8).

The primary aim of the survey was to recommend priority sites for the installation of remote ground temperature sensors. Twelve sites, located in both the coldest and warmest sections of the road network, were eventually recommended. Such a distribution of ground sensors would therefore enable the engineer to have knowledge of the full range of road temperatures likely to be experienced on any particular night.

5.6 Critical Analysis of Method

The results reported in this Case Study have provided the first evidence of the potential offered by low cost scanners such as the Barr and Stroud IR18. The advantages of such instrumentation include:-

(a) The opportunity to acquire thermal imagery from a light aircraft at low cost.

(b) The ability to obtain data in 'real time' without the need for further processing (as would be required using photographic techniques).

(c) The capability to use relatively low cost hardware for the acquisition, recording (VHS or U-matic video-tapes) and display (conventional VTR technology) of the acquired data.

(d) The ability to utilise low cost digital frame store and image processing equipment when specific single frames of data require more detailed investigation.

Against these advantages must however be judged the following disadvantages:-

(a) The imagery is of small format and consequently has limits to its spatial resolving power.

(b) The data is acquired in an analogue format and requires to be converted to digital form prior to processing.

(c) The lack of a black body reference severely limits the capability of the instrument for quantitative temperature analysis.

(d) Whilst indirect correlation techniques may be used to infer temperature variations, such as those used in the Lothian example, such techniques do rely on a relatively small number of temperatures samples to provide the relationship between grey scale level and temperature.

Overall the work performed to date would tend to weight against the use of the TVFS devices for quantitative remote sensing operations. Rather the device should be viewed as a highly cost effective reconnaissance tool to be used as part of a preliminary photographic mission. In this way the TVFS can be used to complement the work performed by thermal linescanners rather than, at least at present, being viewed as an alternative to such instruments.

5.7 CONCLUSIONS

TVFS systems differ considerably both in optical design and in radiometric characteristics from thermal linescanning instruments. Such instruments can be utilised as a highly cost effective alternative method of acquiring thermal data although their use should be perceived as complementary to the more common thermal linescanners. TVFS systems are of more benefit as reconnaissance tools, forming part of a more comprehensive imaging sortie. Although limited in their ability to provide quantitative thermal data the Case Study has illustrated how indirect methods may be used to infer temperature variations.

5.8 REFERENCES

EKIN W.H. 1984. The development of an inexpensive retractable vertical camera rig for a light aircraft. Photogrammetric Record, 11 (63) 311-317.

EKIN W.H. 1987. A new lightweight camera rig. Photogrammetric Record, 12, 343-348.

THORNES J.E. 1985. The prediction of ice formation on roads. Journal of the Institution of Highways and Transportaion, 32, 3-12.

CHAPTER 6

FIELD MEASURING DESIGN STAGE CASE STUDY - MONITORING OF SLOPE INSTABILITY USING CLOSE RANGE PHOTOGRAMMETRY

6. FIELD MEASURING DESIGN STAGE CASE STUDY - MONITORING OF SLOPE INSTABILITY USING CLOSE RANGE PHOTOGRAMMETRY

6.1 BACKGROUND TO STUDY

Disintegration of rock faces can be a major safety hazard to the public. Before remedial measures can be implemented, the ground conditions have to be quantitatively defined by an engineering geologist and it is vital to have data concerning the precise rate and volume of erosion which is occuring. A number of techniques may be used for such a project including: (a) direct measurement, (b) laser profiling and (c) close range photogrammetry (CRP). The benefits of the third approach have already been documented (2.2.3) and these recent developments in analytical photogrammetric instrumentation have been applied to two engineering geological investigations in which the rate and volume of erosion were required. The following Case Study, based on work performed in a joint project between the University of Surrey and Engineering Surveys Ltd and reported in Kennie and McKay [10], illustrates the methods used to solve the engineering geological problem of defining the rate of disintegration of a rock face in chalk.

6.2 ANALYTICAL PHOTOGRAMMETRIC INSTRUMENTATION

In the past surveying organizations have used conventional analogue stereoplotting equipment, designed principally for aerial mapping purposes, to compile plans for architectural, engineering and other close-range applications. The main drawback of this type of equipment lies in the constraints placed on such elements as focal length, lens distortion, camera orientation, film format size and depth variation. These optical and mechanical limitations to orienting stereomodels are well known to photogrammetrists. For example, the maximum allowable tilt movement in an analogue instrument normally restricts the attitude of the camera to a maximum of +5-6°. Orientating a stereomodel under these extreme conditions can be an arduous task which may take considerable time. This could be more efficiently used measuring the feature of interests.

With the advent of digital computers, especially mini and micro-computers together with electronic encoders for digitizing the output from stereoplotting equipment (Petrie 1989), a new generation of photogrammetric stereoplotter has emerged; the analytical stereoplotter (Helava 1958, 1980). With this type of instrument a mathematical model replaces the space rods and other mechanical linkages which are found in more conventional analogue stereoplotters. The Officine Galileo Digicart is an example of an instrument of this type.

6.2.1 Officine Galileo Digicart

The Officine Galileo Digicart is a well established analytical stereoplotter which has been in use in the United Kingdom for the past 5 years (McKay et al. 1985). The instrument consists of a stereoviewing system, linear encoders which digitally record the position of the carriages on which the photographs are mounted, and a minicomputer to process the data. Unlike a conventional analogue instrument where manual elimination of y-parallax is performed, with this design any parallax existing within the model is eliminated by a feedback loop which ensures that corrections are transmitted to the optics in real time (50 times a second) so that a model free from parallax is provided.

6.3 MONITORING OF EROSION

The facility to monitor rates of erosion and produce rock face maps of eroded slopes are two capabilities of direct relevance to engineering geologists involved in assessing unstable slopes as a precursor to instigating remedial measures. The following two case histories indicate how CRP can be used to provide this relevant information. The first example refers to work performed in the Brighton area and the second to a problem in the Lewes area of East Sussex.

6.3.1 Brighton region: Embankment Monitoring

In this case, the client had identified sites in the Brighton

area where erosion was progressing at a rate sufficient to give concern about public safety. A decision was made to adopt a two-dimensional approach to measuring the slope by observing a series of pre-determined section lines. The precise location of each section was identified on the ground and on the 1:1250 scale Ordnance Survey map sheets of the area. Points at the base and top of the embankment were monumented and surveyed using electronic distance measuring (EDM) instruments and theodolites in order to coordinate precisely the control. In cases where only one section was required per site, two additional points offset by 5 m at the base and top of the embankment were also surveyed to give full control in the stereomodel. A Zeiss (Jena) UMK 10/1318 camera (focal length = 100mm) was used to obtain the photography during this project.

The stereomodels were orientated using a minimum of four control points per model, but in some cases up to eight points were used. The photoscale varied throughout the stereomodel from a minimum of 1:250 at the base of the embankment to a maximum of 1:600. The maximum error on the absolute orientation of the control was found to be 25mm in three dimensions and the final accuracy of the sections was of the order of $\pm$4 cm. Field verification of these results proved difficult due to the site conditions, but a photogrammetric repeat of the exercise indicated that the level of completeness was high. The results were presented as 1:100 scale plots and Fig. 2 in Kennie and McKay [10], page 213 illustrates a typical profile.

This two-dimensional technique is a very practical method of carrying out a monitoring problem since the results from one year can easily be overlayed on the results for subsequent years giving a direct indication of the variations. The disadvantage of this approach is that it only measures variations along the selected sections. Thus, the location of these lines must be chosen very carefully if the maximum or average amount of erosion is to be determined. A further disadvantage is that it is only possible to measure accurately down slope (as opposed to lateral movements) and hence the volume determination is less precise unless many closely spaced sections have been measured.

6.3.2 Lewes area: Road Tunnel

The second case involved the monitoring of the shape and erosion occurring around the entrance to a road tunnel. The southern approach was cut into chalk and consisted of four separate and distinct faces varying in height from thirty to sixty meters. A mesh fence had been erected as a temporary measure to stop rock falls from impacting on to the road surface.

Control stations were monumented at both the top and base of the cliffs, ensuring that a minimum of six control points were visible within each stereomodel. As with the previous project the stations were surveyed by EDM and theodolite and targets positioned over the monuments for precise viewing and control pointing in the stereomodel. The control network was tied into the National Grid and all height values related to Ordnance Datum. Photography was performed using a Zeiss (Jena) UMK 10/1318 camera with glass negatives to provide greater stability. Figure 3 in Kennie and McKay [10] page B214 illustrates the site as well as giving the locations of the control and photographic stations. The camera was positioned in front of the cliff faces by the use of a hoist. Again, due to the variable distance of the cliff face from the camera a wide variation in photoscale from 1:400 to 1:900 resulted.

For each of the four faces, the control network was rotated about two points on the cliff top thus forming a datum line parallel to the face itself. The shape of the cliff was then defined by contours at 0.25 m intervals of depth - that is, as differences of horizontal distance from this datum plane. This contrasts with a conventional survey of vertical elevations above a height datum. The contours produced by this process were digitized directly using the Digicart and draughted off-line on a Kongsberg flat-bed plotter at a scale of 1:100. A typical contour plot and section are illustrated in Kennie and McKay [10] Figure 5, page 216. The resultant plans were then examined (see Kennie and McKay [10] page B216 and eight critical area were selected for detailed measurement. In order to achieve this, the stereomodels

were reintroduced into the analytical plotter and, with the aid of a program forming part of the basis software of the system, sections were digitized and subsequently plotted at a scale of 1:100. The accuracies achieved were found to be the same order as for the embankment at Brighton, both the mapping and sections were within +5 cm of true ground position.

6.4 CRITICAL ANALYSIS OF METHOD

The preliminary contouring of the cliff faces provides a comprehensive view of the rock surface permitting the optimum selection of sections for monitoring. The method used at Lewes, however, should be considered as two-dimensional because future measurements would be along fixed sections. An alternative approach for measuring erosion, and one which could have been applied to either of the case histories discussed here, would involve the use of a three-dimensional terrain modelling package such as those discussed in Petrie & Kennie [18].

In order to apply this technique an array of heights (either at grid intersections or randomly located) would be formed into a terrain model of the slope. The cliff face to be monitored would therefore be recorded as a series of x, y and z coordinates from which cross-sections and profiles could be extracted at user defined positions. Area, volume and other computed geometrical data could then be generated from this data base. Thus the survey results from successive years could be compared and the volume of rock that had been eroded be calculated. If such disintegration occurred in an unexpected area, the terrain model together with the photographic record could then be referred to, and precise measurements quickly obtained.

Since the initial work reported in this Case Study has been completed the concept of using a DTM to analyse the erosion of slopes has been examined further and some preliminary results are reported in Chandler and Cooper (1988, 1989) and Chandler and Moore (1989). The range of secondary geomorphological derivative data which may be provided by using such a process in illustrated by Figure 6.1.

6.5 CONCLUSIONS

Analytical close range photogrammetry can be of significant value to the engineering geologist in determining the extent and rate of slope instability. In its simplest form such monitoring may be restricted to the interpretation of conventional contour maps and profiles. More sophisticated data analysis techniques based on DTM's may be used and these can provide information on changes in slope morphology, more accurate volume calculations and more realistic visualisations of the slope.

6.6 REFERENCES

CHANDLER J.H. and COOPER M.A.R. 1988. Monitoring the development of landslides using archival photography and analytical photogrammetry. Land and Minerals Surveying, 6, 576-584.

CHANDLER J.H. and COOPER M.A.R. 1989. The extraction of positional data from historical photography and their application to geomorphology. Photogrammetric Record, 13(73), 69-78.

CHANDLER J.H. and MOORE R. 1989. Analytical photogrammetry: a method for monitoring slope instability. Quarterly Journal of Engineering Geology, Vol. 22, 99-110.

HELAVA U.V. 1958. New principles for photogrammetric plotters. Photogrammetiq, 14, 49 - 56.

HELAVA U.V. 1980. Analytical plotter basics. Analytical Plotter Symposium, ASP, 12 - 30.

McKAY W.M., DOWMAN I.J and FARROW J.E. 1985. Experience with analytical plotters. Photogrammetric Record, 11, 66, 729 - 45.

PETRIE G. 1989. Analogue, analytical and digital photogrammetric systems applied to aerial mapping. Chapter 7 in Engineering Surveying Technology, Kennie T.J.M. and Petrie G. (Eds), 238 - 287.

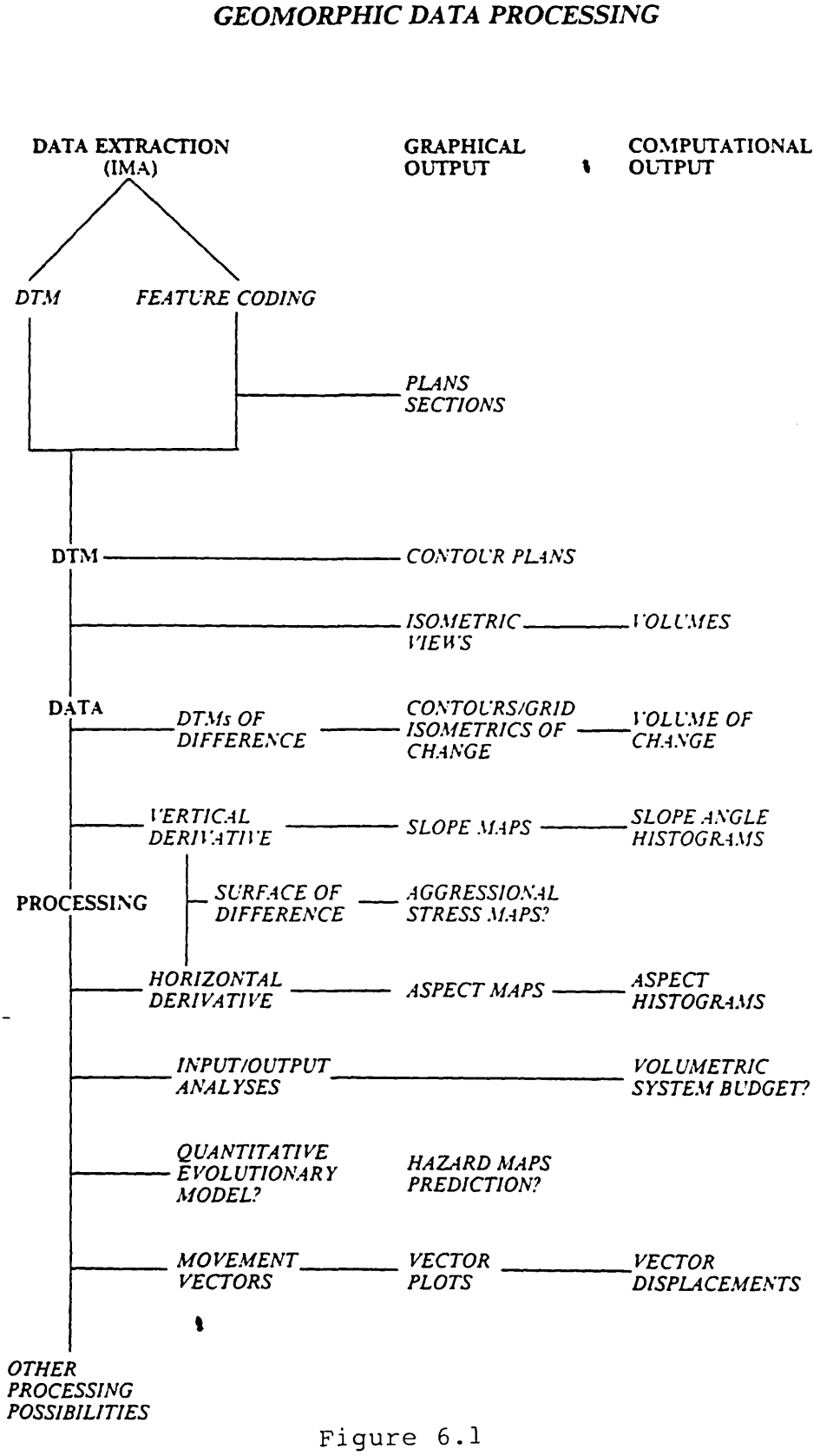


Figure 6.1

Use of DTM's for Slope Stability Studies (Chandler and Cooper, 1989)

CHAPTER 7

CONSTRUCTION STAGE CASE STUDY
MONITORING GROUND MOVEMENT DURING BRIDGE CONSTRUCTION

7. CONSTRUCTION STAGE CASE STUDY - MONITORING GROUND MOVEMENT DURING BRIDGE CONSTRUCTION

7.1 BACKGROUND TO STUDY

The design of spill-through bridge abutments has been a problem of concern to geotechnical engineers for many years. Knowledge of the bedding conditions around such structures has been very limited and no large scale assessment of the changes in geological conditions existing during their construction has been made. Consequently the Department of Civil Engineering at the University of Surrey were awarded a research contract by the Department of Transport to investigage the earth pressures acting on such structures and then use this data to infer the extent of ground movement during construction. Three publications relevant to this Case Study are included in Appendix B (Kennie [13], [14] and [15]).

The primary aim of the project was to gather data, during construction, of the earth pressures acting on the abutments of bridge number 15 on the Chertsey to Wisley section of the M25. This data would then be compared with the estimated earth pressures used by the designers and used to formulate some general guidelines for designing bridge abutments of this type.

The structure to be monitored consisted of a pair of spill-through abutments one at either end of the bridge. A spill-through abutment consists of a transverse sill beam supporting the end of the bridge deck, with two or more buried columns carrying the load down to a common base slab. In this particular case six columns transmitted the load to the base slab. The bridge itself consisted of a 76m continuous voided slab constructed in two halves separated by a central expansion joint. Figure 7.1 illustrates the primary features of the structure.

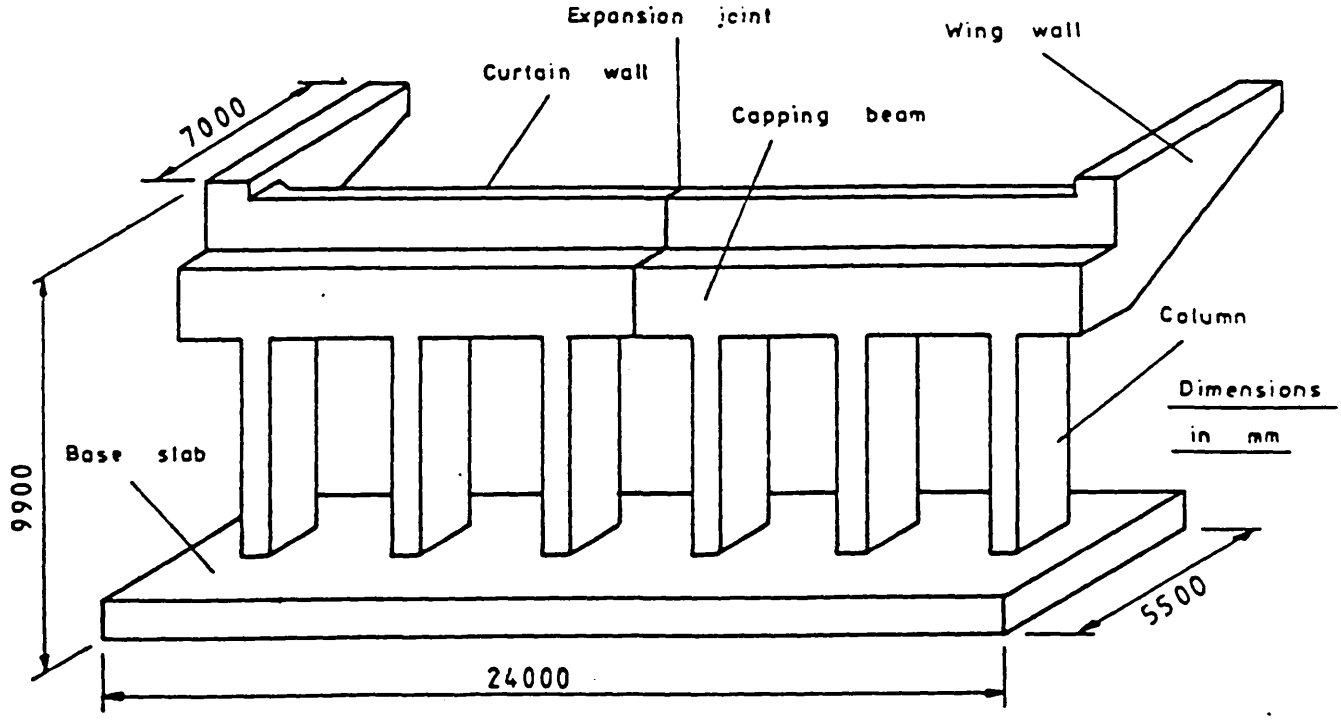


Figure 7.1

Primary Features of Bridge Abutment
on Chertsey/Wisley Interchange - on the M25

7.2 DATA ACQUISITION

Data relating to the movement and deformation of the structure were obtained using both geotechnical and surveying instrumentation:

7.2.1 Geotechnical Instrumentation

The detailed measurement of earth pressures was restricted to two columns on each abutment. Three different geotechnical instruments were used to provide details of the forces on all faces of the columns.

Firstly, the measurement of lateral earth pressures was carried out using vibrating wire earth pressure cells. It was considered that measurements obtained on the sides of the columns and at three different levels would enable not only the magnitude of earth pressures to be determined but also the extent of any 'spill through' of earth between the columns.

Secondly, vertical earth pressures were measured using boundary cells with Maihak embedment cells. Data was acquired at base slab level and at differing levels in the fill since it was anticipated that the pressures may have been considerably greater than that associated with the nominal weight of the backfill. The wires from the earth pressure cells were passed underground in trunking to a control panel situated about 30 metres from the abutments. The position of the control panel also coincided with the positions of the survey pillars.

Thirdly, two inclinometer tubes were installed. The first, in one of the columns, was used in order to determine the complete deflection profile down to base slab level and the second, installed in the backfill zone between the two columns, was used to enable an assessment to be made of the relative movement of the fill compared to the abutment.

7.2.2 Surveying Methods

The surveying requirements for this project consisted of three distinct stages. Firstly it was required that a network of surveying stations be established around the site in order to provide an absolute reference system. Secondly survey methods were to be used to provide an independent means of confirming any deflection and/or translation of the abutment during construction i.e. surveying techniques would be used to confirm the relative movements as determined by the earth pressure cells and the inclinometer. Thirdly the settlement of the earth behind the abutment was to be monitored when the bridge was opened in order to determine the compacting effect of traffic loading. The following sections deal specifically with the requirements for the second stage. Details of the other two parts of the project are reported in Kennie, 1984.

a) Layout of Survey Pillars

The initial specification required that six survey stations should be established within the site. In addition it was also agreed that the survey stations could be sited in the manholes which would be used to house the control unit for the pressure cells. Several other constraints were imposed on the positioning of the survey stations. Firstly the positions had to be sited so as to be able to measure the deflection of the abutments. In view of the instrumentation available (a Kern DKM-2A 1" theodolite and Aga Geodimeter EDM Unit 120) it was decided that the deflection would be best determined by measuring angular deflection from a position perpendicular to the expected direction of movement. Secondly the survey stations should be positioned so as to be intervisible both before construction began and also subsequently in order to enable remeasurement at a later date. Finally the stations should be sited outside the zone of movement associated with the abutments. The final layout of the survey stations is illustrated by Figure 7.2.

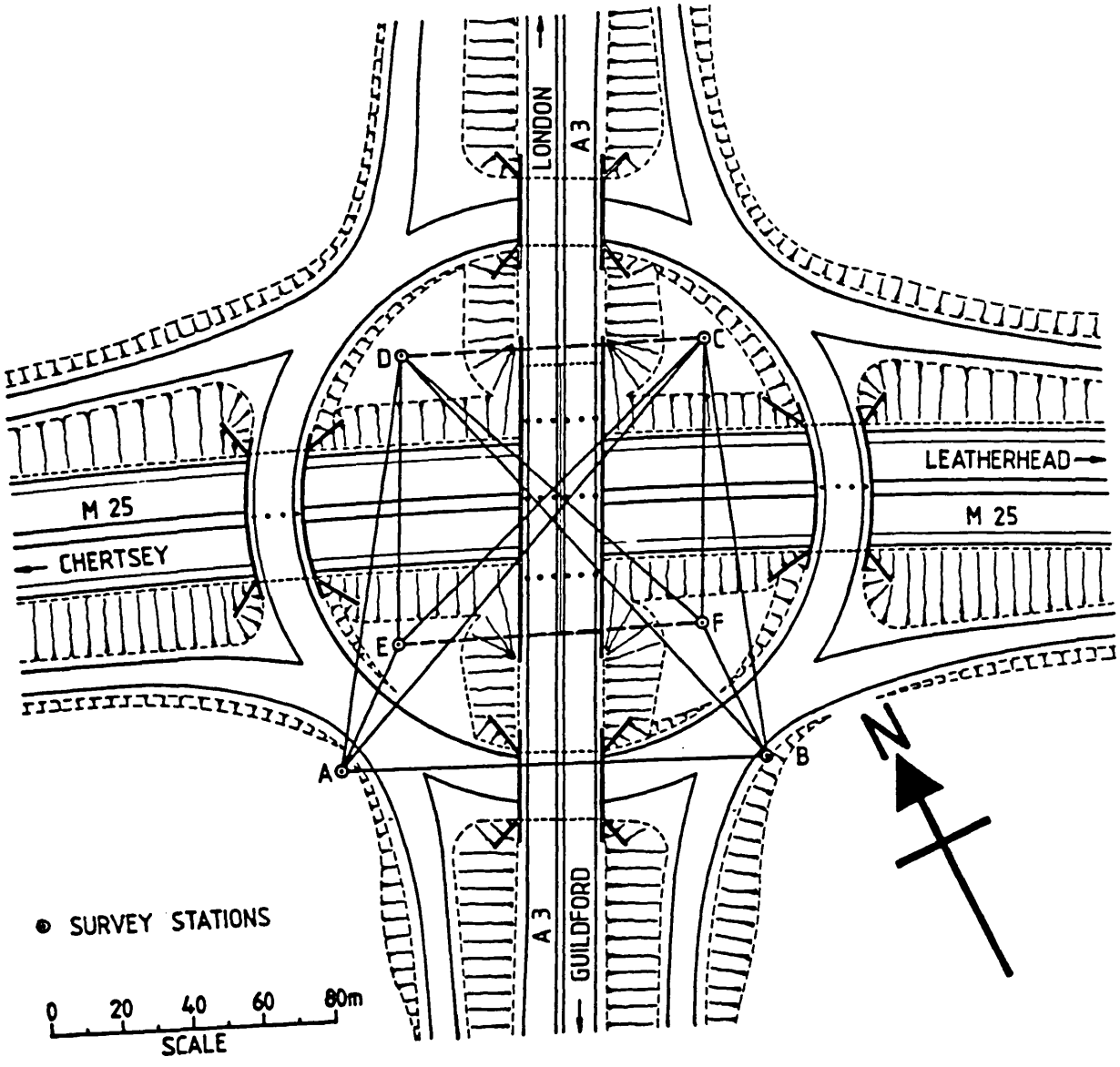


Figure 7.2

Layout of Survey Control Points

The survey stations consisted of a concrete pillar sited in a brick built manhole. The pillars had a steel surround and were founded on a 600 X 300 mm concrete slab. The actual survey point consisted of a small steel plate on which was welded a spherical bolt head, an indentation on the top surface of the bolt indicating the survey point. The entire manhole was sealed by a steel access cover.

b) Design of Survey Targets

Two types of survey target, one for angular measurement and one for distance measurement, were used on this particular project. The target used for the measurement of angular deflection consisted simply of a Demec stud attached permanently to the abutment by plastic padding. The stud itself while extremely small proved to be an ideal survey target over short distances (up to 80 m). From these distances the indentation in the stud was almost exactly identical to the width of the theodolite crosshairs. Thus repeat pointings to the point could be made very precisely.

For distance measurements, a small 1 cm reflex acrylic reflector was permanently affixed to the abutment. These reflectors were positioned 115mm above the Demec studs, thus ensuring that the reflector was offset vertically by the exact distance which existed between the optics of the Geodimeter 120 and the Kern DKM2A theodolite (Kennie [13]), page B286. The characteristics of this type of target is described in detail in Kennie [14], page B294 and Kennie [15], page 300. The work reported in both these publications was carried out prior to the period of registration and formed part of a dissertation for a Masters degree. They are nevertheless included within this thesis since they contributed to the development of the monitoring system used in this Case Study.

The targets also had the added benefit of being very unobtrusive and thus aesthetically acceptable, and, on a more practical level, less likely to attract the attention of vandals.

More recently the method has attracted interest elsewhere in the world. Rueger (1989, 1990) has recently reported further work on the use of these reflectors for monitoring applications.

7.3 RESULTS AND ANALYSIS OF DEFORMATION MONITORING

Measurements of the angular deflections of the spill through abutment were taken at two critical stages during the project. Firstly, measurements were recorded during the process of backfilling and compaction of earth around the twelve columns which formed part of the base of the abutment. Secondly, measurements were recorded during the period immediately after the bridge deck was cast.

7.3.1 Columns Deflections/Settlement During Backfilling

Whilst it was thought possible that the columns could be deflected by the backfilling and compaction operation, it was considered much more likely that the main effect of the compaction would be to cause a slight settlement of the entire structure. Thus a suitable remote means of measuring this settlement has to be devised. Deflection was measured by horizontal angle measurements from Point D using survey station A as a reference object.

From a practical point of view the most appropriate means of remotely measuring settlement over this distance was considered to be trigonometrical heighting. The procedure therefore adopted for this part of the project was to measure horizontal angles, on six zeros, vertical angles on four zeros and to take repeat distance measurements with the Aga Geodimeter 120 over the period during which backfilling was taking place. The results obtained are illustrated in Tables 7.1 and 7.2. As the level of backfill was raised around the columns frequent pressure cell readings were recorded so as to relate the increased overburden to the changes in soil pressures. The lateral

POINT	HORIZONTAL FROM POINT D (m)	HEIGHTS RELATIVE TO POINT D (M)	HORIZONTAL ANGLE	DEFLECTION (")
NW COLUMN 2 DEPTH OF FILL (M)		0 X X	0 ' "	
0	30.598	104.579	263 07 19	0
1	-		263 07 22	+3
1.5	-		263 07 31	+12
3.5	30.597	104.574	263 07 25	+6
5.0	30.597	104.574	263 07 31	+12
NW COLUMN 2				
0	30.751	104.608	265 10 18	0
1	-	-	265 10 17	-1
1.5		-	265 10 23	+5
3.5	30.751	104.603	265 10 19	+1
5.0	30.752	104.603	265 10 29	+1
NW COLUMN 3				
0	34.586	104.565	262 31 01	0
1	-	-	262 31 06	+5
1.5		-	262 31 10	+9
3.5	34.587	104.560	262 31 07	+6
5.0	34.586	104.560	262 31 11	+10

Table 7.1 - Column Deflections: Columns 1-3 on NW Abutment

COLUMN	REDUCED LEVEL OF REFERENCE POINT		SETTLEMENT
	HEIGHT OF FILL (0m)	HEIGHT OF FILL (5m)	
NW1	104.475	104.471	4mm
NW2	104.478	104.475	3mm
NE1	104.521	104.517	4mm
NE2	104.526	104.522	4mm
SW1	103.028	103.022	6mm
SE1	103.955	103.951	4mm
SE2	103.979	103.974	5mm

Table 7.2: Settlement of Abutment During Backfilling

earth pressures indicated similar profiles on each pair of opposite faces of the abutment indicating that as evidenced by the surveying data that the structure was reasonably well balanced. The vertical settlements of between 3-6mm for the abutments were due to the compression of the underlying strata caused by the loading from 5m of backfill.

7.3.2 Deformation of Abutment During Construction

The construction of the spill-through abutment continued with the casting of the capping beam across the top of the columns and this was finally completed with the construction of the wing walls. With the abutment complete the bridge deck falsework support was prepared and the bridge deck slab constructed in 2 halves, with the East deck being poured 6 weeks before the adjacent West deck. Placement of the in-site concrete commenced at the centre of the deck and proceeded outwards on two fronts simultaneously towards the abutments. During this stage of construction, the embankment only existed to a height level with a top of the abutment columns and as such was expected to provide only minimal resistance to any lateral loading. The expansion joint separating the two halves of each capping beam was expected to prevent the majority of the loading effects of the East section of an abutment from being transmitted to the West section. In reality a significant amount of interaction occurred. For this part of the project the surveying requirement involved the measurement of the angular and distance displacements immediately before the deck was cast and monitoring of these displacements during the following few days. Observation were taken from Survey Stations C and F to monitor the first pour and from Stations D and E for the second pour. Five sets of survey targets were established on the abutment wing walls. As before, horizontal angles were measured on six zeros, vertical angles on four zeros and repeat distance obtained with the Geodimeter 120. The results for this part of the project are illustrated by Figures 7.3 - 7.10. (see also Kennie [13], page 288).

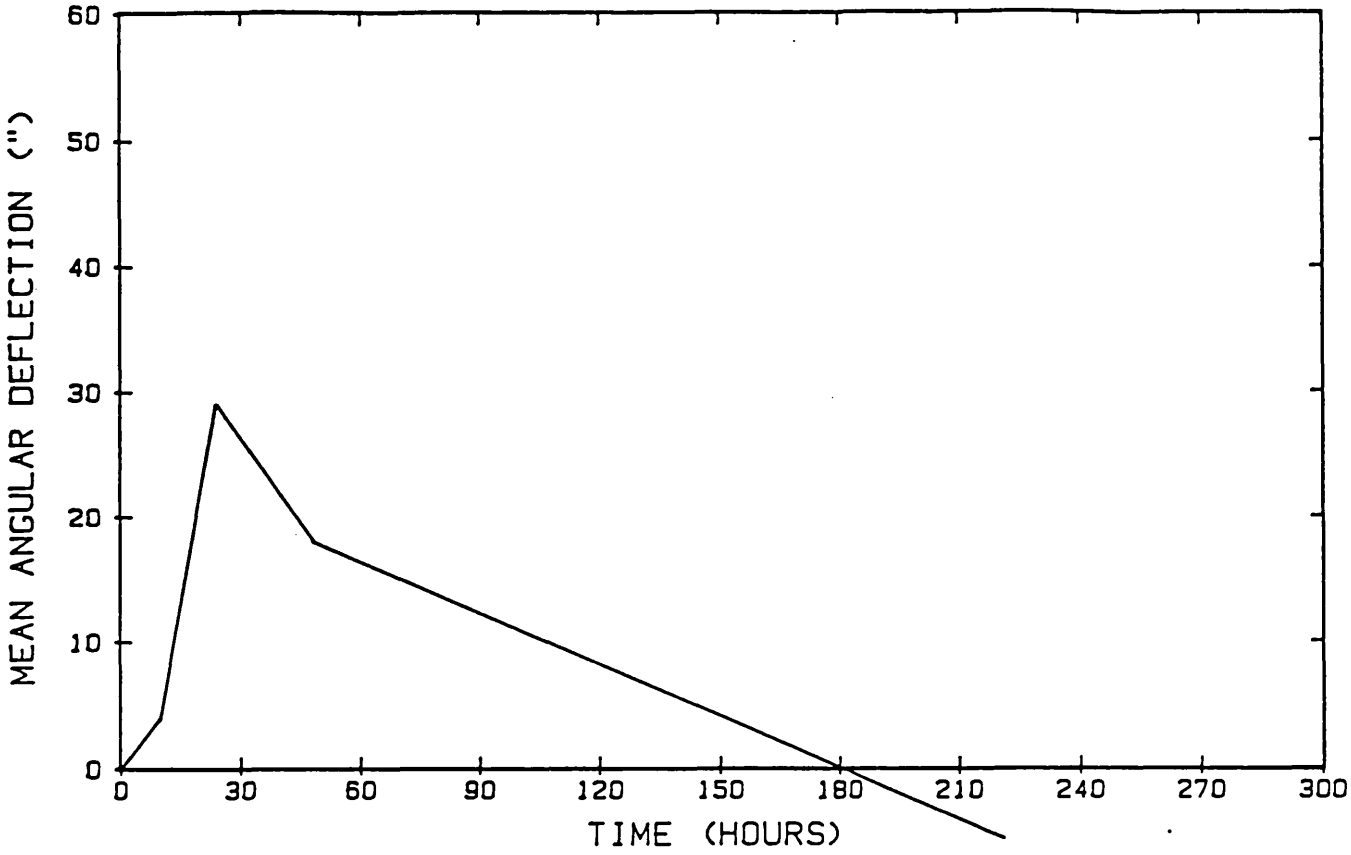
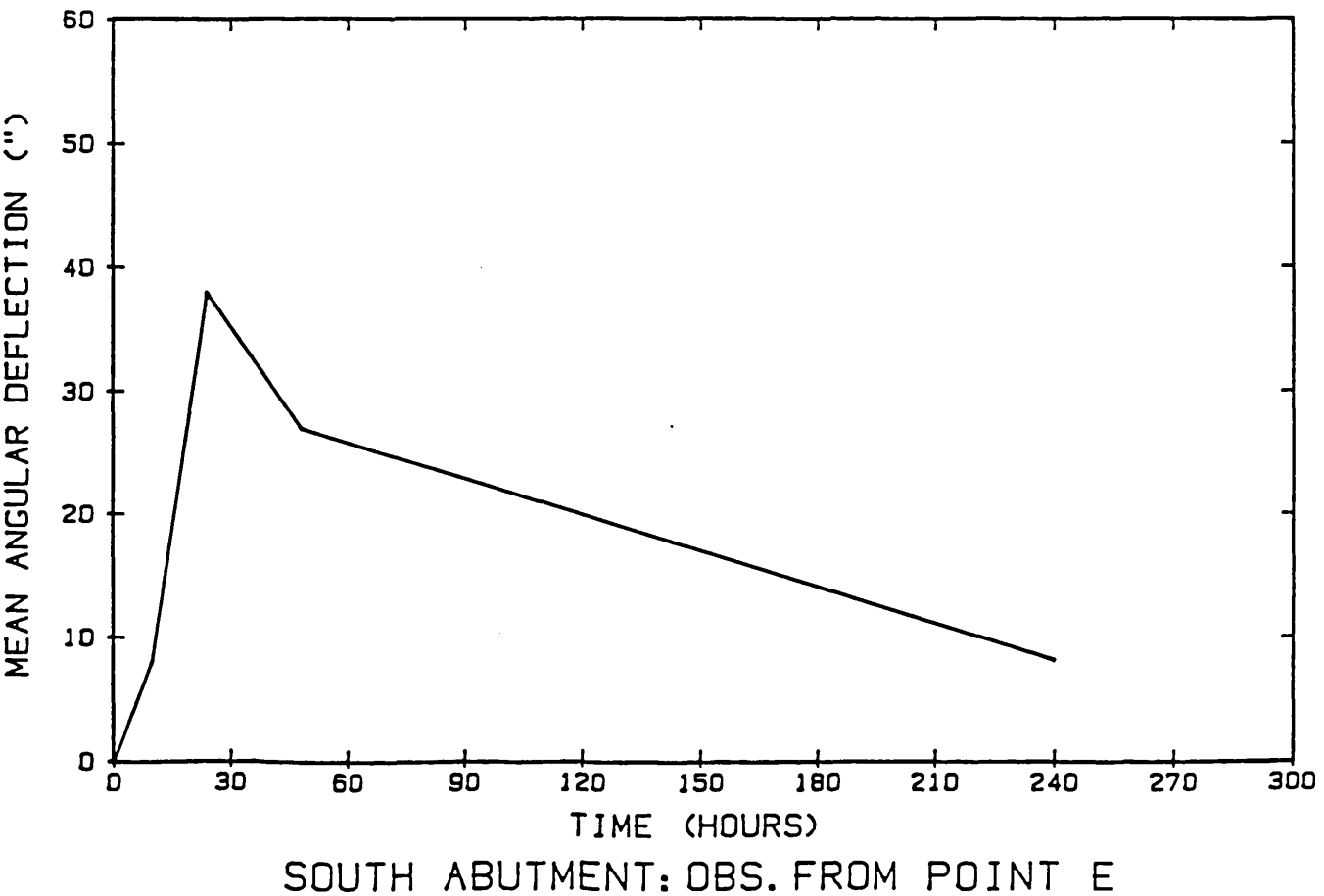
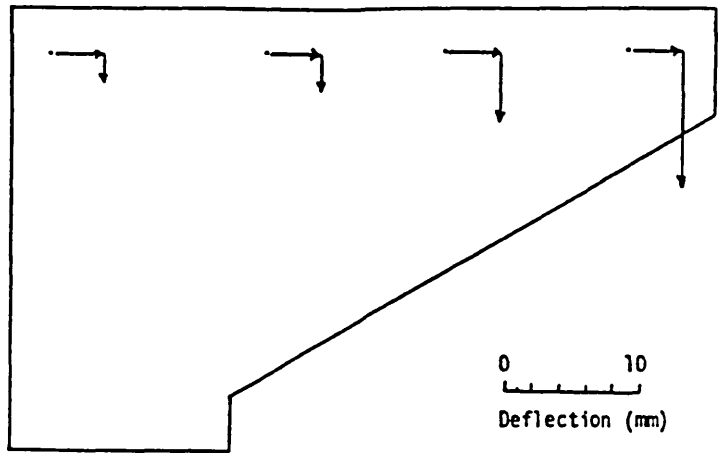


Figure 7.3: Deflection of West Abutment (North)

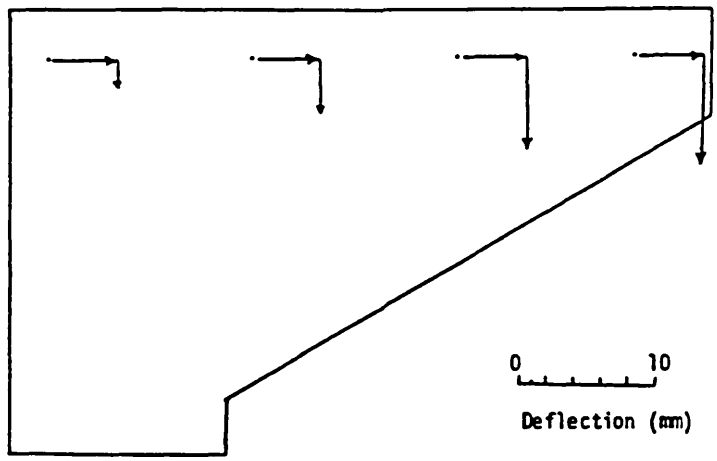
Figure 7.4: Deflection of West Abutment (North)





North Abutment; Vectors of Deformation
as Observed from Point D After 24 Hours.

Figure 7.5



South Abutment; Vectors of Deformation
as Observed from Point E After 24 Hours.

Figure 7.6

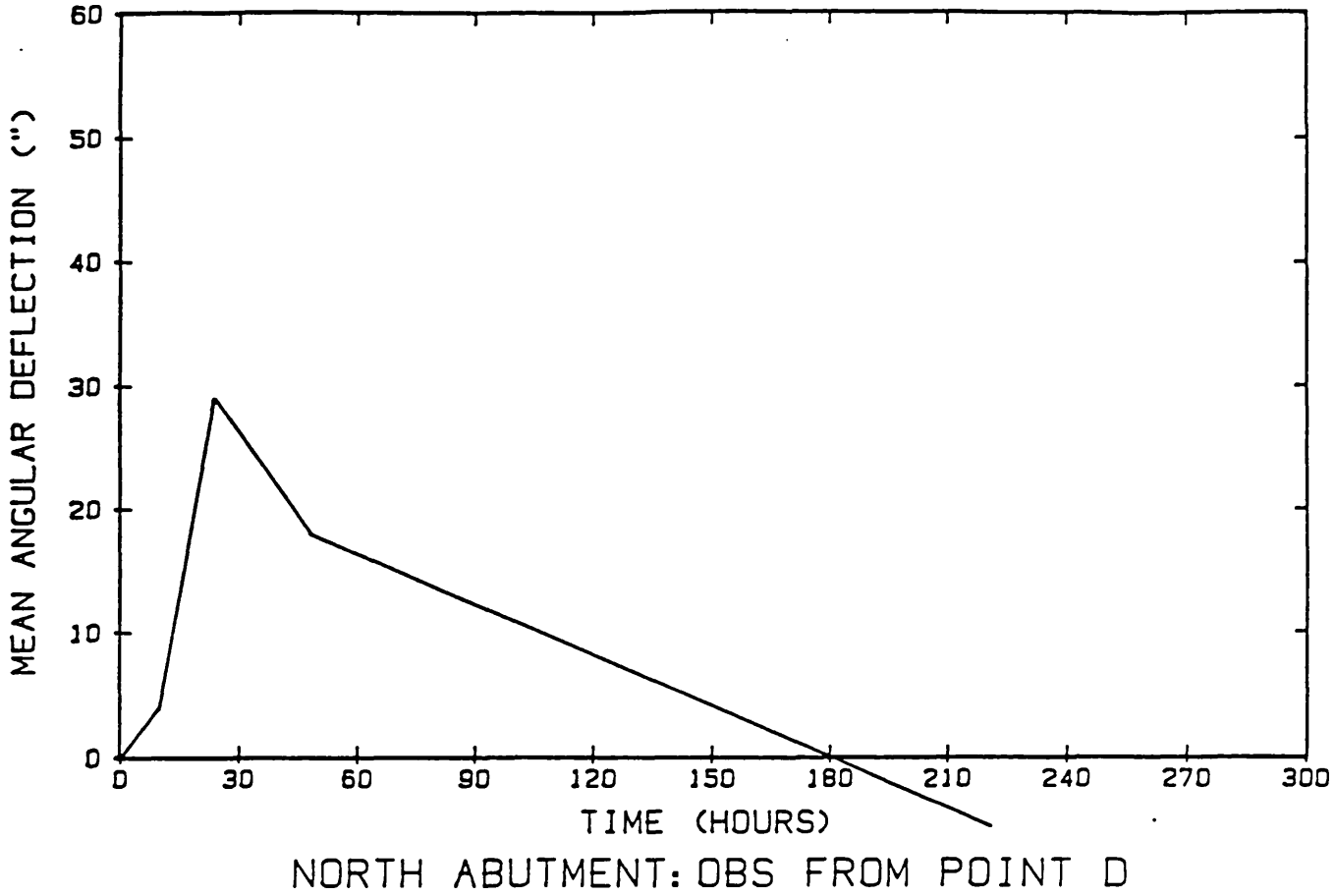
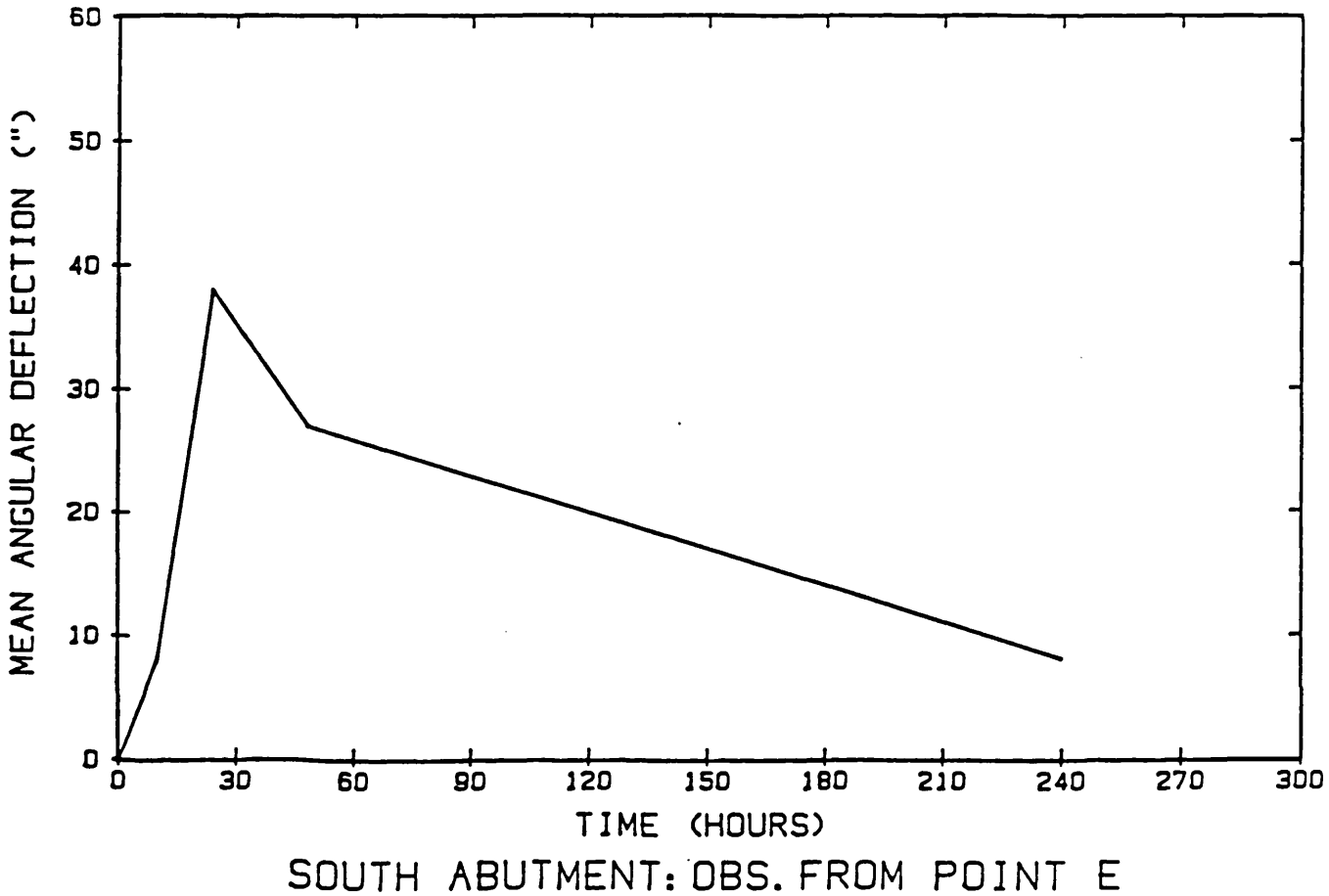
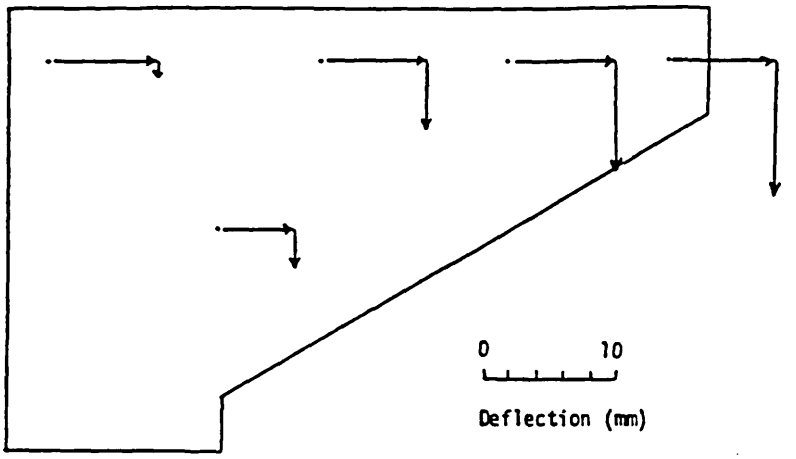


Figure 77: Deflection of East Abutment (North)

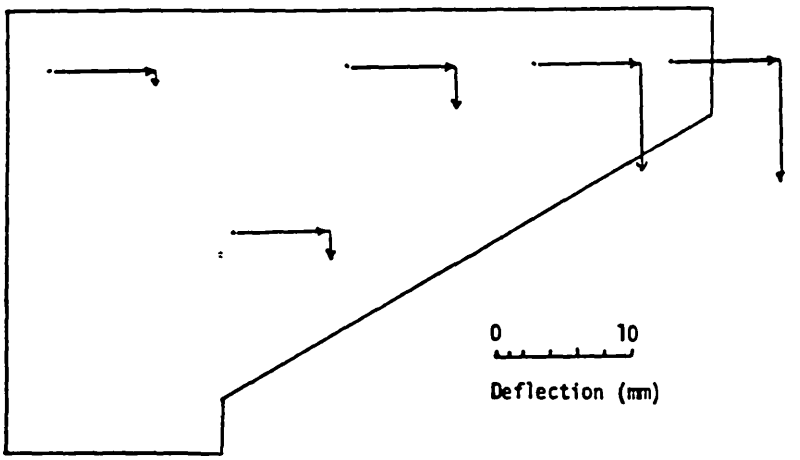
Figure 7.8: Deflection of East Abutment (South)





North Abutment; Vectors of Deformation
as Observed from Point C After 24 Hours.

Figure 7.9



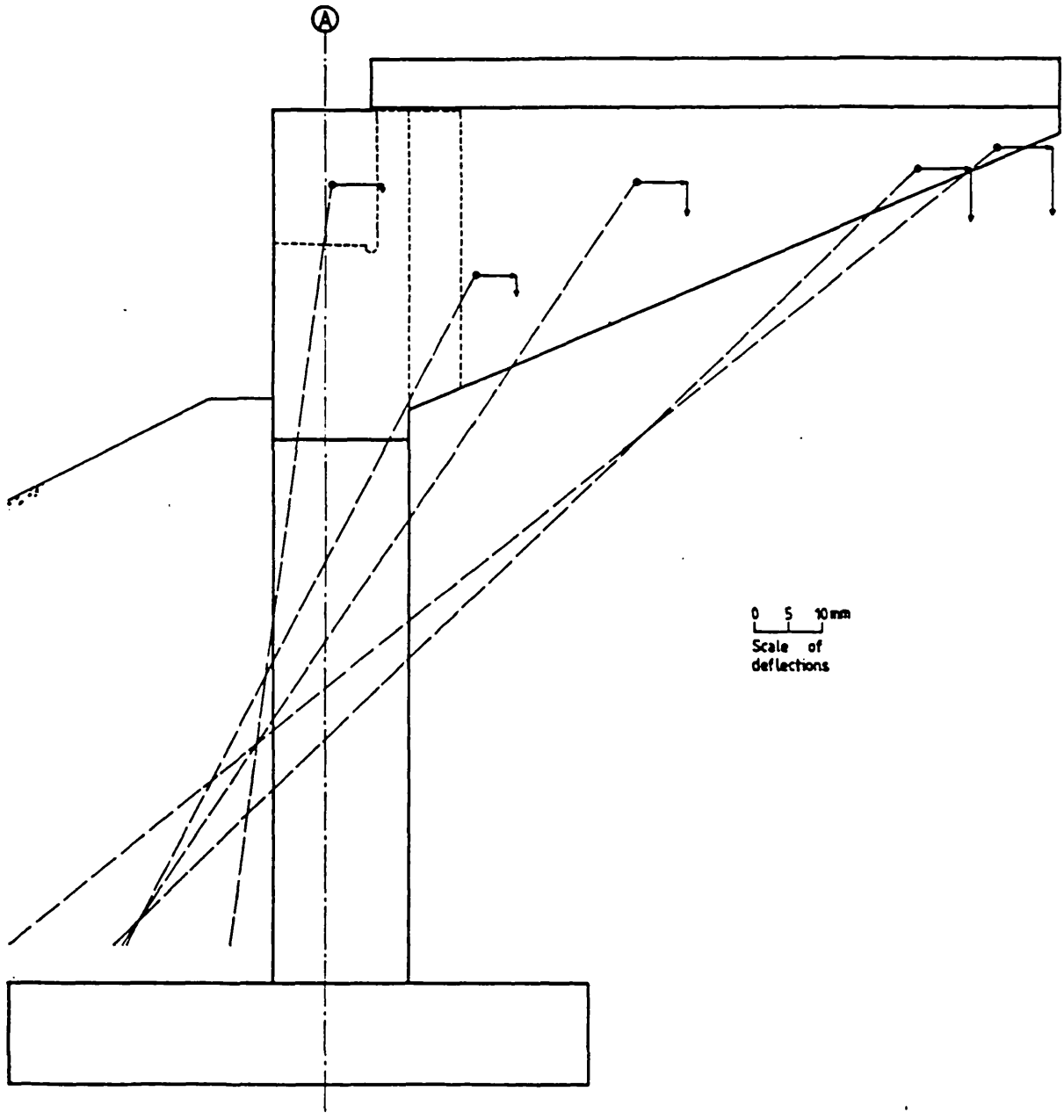
South Abutment; Vectors of Deformation
as Observed from Point F After 24 Hours.

Figure 7.10

Surveying of the East Wing Walls of both abutments, before and 24 hours after the deck pour indicated movements of the wing wall tips of 8 mm in a horizontally backwards direction and 10mm and 9mm in a vertical downwards direction for the North and South abutments respectively. The horizontal and vertical displacements of 4 and 3 other points across the face of the wing walls of the A and E abutments respectively were also measured in a similar manner. The recorded displacements were primarily due to a rotation of the abutments as a result of the lateral loading caused by the expansion of the deck. Lindsell et al (1984) estimated the position of the angle of rotation as the intersection of the lines perpendicular to the direction of movement of each reference point on a wing wall (Figure 7.11). These geometric constructions do not indicate a unique position for the centre of rotation but they do indicate that it occurred between 1 and 3m above the top of the base slab. This is contrary to the common design approaches in which it is often assumed that rotation occurs about a point at the junction between the columns and the base slab.

Deflections were also monitored for the West bridge deck and indicated that the tip of the West Wing Wall of the South abutment had moved 6mm backwards and had dropped vertically by 8mm. The A abutment displacements were 4mm backwards and 5mm vertically. The wing walls were also observed to have deflected inwards at the top by 7mm and 5mm for the North and South abutments respectively. These lower values were considered to be due to the restraint offered by the east deck and the torsional resistance of the base slab. Regular readings after the pour from the inclinometer tube located within the South column indicated that the maximum deflection of the column occurred 36 hours after casting the deck. A relative displacement of about 5.5mm occurred between the top and bottom of the North abutment. Again the inference of the centre of rotation from this data indicated a position between 1 and 2m above the upper surface of the base slab.

Figure 7.11: Estimated Centre of Rotation of Spill through abutment



Inspection of the changes in soil pressure after the pour indicated that the maximum effects occurred between 24 and 36 hours after casting. By studying by pressure plots and defining the height at which the interpolated pressures remained unchanged on a particular face indicated that the centre of rotation was consistently between 0.75 and 1.25mm above the top of the base slab (Lindsell at al, 1984).

The abutment deflections were observed to recover considerably after the initial period of expansion of the base slab.

The results emphasised the role of the compacted backfill around the columns in providing lateral resistance to the lateral load exerted by the expansion of the decks during the first 36 hours after casting. In addition the centre of rotation has been shown to occur at a considerable distance above the base and not at base level as is generally assumed for design purposes.

7.4 CRITICAL ASSESSMENT OF PROJECT

The main criticism of the monitoring approach outlined in Section 7.3 is that it was designed to be capable of detecting periodic movements from a single point in a relative rather than in an absolute sense.

Whilst several advantages accrue from this simple approach, there are several significant drawbacks. The main advantages include:

a) Simplicity - the method was not dependent on highly sophisticated measurement or analysis techniques and is thus advantageous in situations where limited resources in terms of equipment, computing facilities and skilled manpower exist;

b) Speed - in situations where dynamic measurements

are required, the speed and ease of detecting deformations using this approach can be extremely beneficial. One could for example envisage situations where the deformation was found to be close to some maximum acceptable deformation value in which case the rapid results produced by this method could help avoid serious damage occurring to the structure.

The approach does however suffer from several disadvantages:

a) Since monitoring any given target point is dependent on measurements from one survey station, if for any reason that point is damaged, then monitoring will no longer be possible.

b) Assumptions have to be made about the stability of the survey stations during the periods when monitoring is taking place. Whilst every care may be taken to situate points outside the zone of influence of the structure and to build stable reference points, an independent means of confirming this assumption is highly recommended.

c) Most importantly, the results relating to the magnitude and direction of the deformation are unable to be statistically assessed.

In view of the above mentioned deficiencies, it is suggested that given 'ideal world' resources, this type of project would be best tackled using the network approach, both for the control and deformation measurement. The main advantages this would offer over the method which was used for the project are:

a) Increased flexibility - since the observations would not necessarily have to be of the same type or number for each periodic measurement, any loss or addition of survey stations could be easily

accommodated.

b) Greater control - by taking measurements of both the main control network and the deformation target points during the same measurement period greater control could be exercised over the effects of any movement of the survey control stations.

c) An ability to statistically assess the significance of any movement detected - by utilising the full statistical analysis which is available as part of the adjustmnt process, decisions could be made as to the significance of any deformation which was detected.

d) Absolute rather than relative deformation measurements could be detected.

7.5 REFERENCES

KENNIE T.J.M. 1984. Deformation Monitoring of Engineering Structures: A review of Current Practice and an Examination of Two Case Studies. Test of Professional Competence, Royal Institution of Chartered Surveyor, Land Surveyors Division, 139pp.

LINDSELL P., MOORE S.R., KENNIE T.J.M., RAZAK H.A., and ROBINSON S. 1984. Monitoring of Spill through Abutments. University of Surrey. Final Report to Transport and Road Research Laboratory 40p.

RUEGER J.M. 1989. Permanent and Inexpensive reflectors for monitoring surveys, in Proc. Symp. on Surveillance and Monitoring Surveys (SAMS' 89) 63-73.

RUEGER J.M. 1990. On the use of inexpensive reflectors in EDM. The Australian Surveyor, Vol 35, 3, 238-247.

CHAPTER 8

POST CONSTRUCTION STAGE CASE STUDY
USE OF AN ELECTRONIC DISTANCE MEASURING INSTRUMENT FOR EARTH VIBRATION STUDIES

8. POST CONSTRUCTION STAGE CASE STUDY - USE OF AN ELECTRONIC DISTANCE MEASURING INSTRUMENT FOR EARTH VIBRATION STUDIES

8.1 BACKGROUND TO STUDY

The effects of earth vibration and the possible damage caused by such vibrations is a subject which has been of concern to engineering geologists for many years. The source of these vibrations can vary and may be due to natural causes such as wind or seismic activity. Man-made vibrations may also be of concern, for example, vibrations may need to be monitored by contractors piling or blasting in quarries adjacent to areas of development to ensure that acceptable levels of vibration are not exceeded.

Typically earthquake vibrations occur at around 1 Hz, oscillations caused by the wind are often within a frequency range of 1 Hz to 2 Hz and those induced by man-made vibrations in the range 0.1Hz to 10Hz.

8.2 PREVIOUS WORK

The measurement of earth vibration by non-contact methods has been investigated for some years and the following Case Study considers the application of a new approach to this problem based on the use of an electromagnetic distance measuring (EDM) instrument. The results are reported in Kennie et al [16] and more recently in Howard et al (1989).

Four quantities can be derived from a vibrational analysis: a) the fundamental and harmonic frequencies of the structure; b) the amplitude of the vibration; c) the modal shape of the vibration and; d) the degree of damping associated with the vibration.

The methods used for measuring these parameters can be categorised as either invasive techniques, that is those

which involve direct contact with the structure, or non-invasive techniques, or non-contact methods which measure vibrations by remote measurement. A review of such instruments is reported in Kennie et al [16], and in particular pages B311 to B313.

8.3 DATA ANALYSIS

8.3.1 Instrument Characteristics

To assess the capability of a new design of EDM equipment for vibration studes a study was carried out using the Tellurometer MA-200.

Distance measurement is determined using the MA-200 in a conventional manner by comparing the phase of the transmitted and reflected signals. The instrument uses an infra-red carrier signal of 0.78 um wavelength which is intensity modulated over a range of frequencies between 90 MHz and 110MHz. The frequencies are swept through this range until the distance is an integral number of an exact wavelength. This is then measured to give a sub-millimetre resolution of 0.1mm. In absolute distance measurement the reflected signal is compared with an internally reflected signal rather than by direct comparison with the source. Thus the distance measured represents the distance between a fixed point within the instrument and the reflecting prism. By this process measurements are made with reference to the fixed and known internal paths.

The internal signal is established by switching the emitted laser radiation into one of the two paths using a mirror. One path is directed to the receiver photo-diode to form the internal path while the other is directed towards the prism. Depending on the position of the internal/external shutter, the light incident on the receiver photo-diode will either by that from the

internal or external path. The quoted specification of the instrument for distance measurement is 'better than half a millimetre +/- 0.5 parts per million'.

For vibration analysis, the change in absolute distance measurement is not of prime importance, but rather the difference between consecutive phase measurements (since the difference between phase measurements will enable the amplitude of the vibration to be determined). In order to obtain as many samples of phase as possible from a vibration it is possible to neglect the internal path, i.e to maintain the shutter in such a position as to allow an external path for the emitted radiation only.

The MA-200 is normally operated from a hand held controller for normal distance measurement although it may also be interfaced with a hand held/lap top computer. The latter is particularly useful for field storage of phase data.

8.4 TEST PROCEDURE

The oscillatory vibration of the ground or structures placed upon it is a function of a number of factors including the geological nature of the site; the distribution of the mass and stiffness of the structure; the connectivity between the component parts; the support conditions; and the type and frequency of the applied forces or displacements. The dynamic displacement in such cases is a complex combination of a number of different harmonic frequencies with varying amplitudes and as such is a complex waveform.

In order to interpret vibrational data it is necessary to separate the signal into its component frequencies. In this way the harmonic, or resonant, frequencies can be identified. These would be apparent by their high amplitude. With a single degree of freedom system only

one harmonic frequency exists and may be evident from examination of displacement or acceleration time data. However, post data processing is required to characterise the harmonic frequencies for a multi-degree of freedom system.

8.4.1 Sampling Requirements

Clearly, for a given frequency of vibration, the number of samples and the rate of sampling are of critical importance. The frequency range is described by the 'Nyquist Sampling Theorum' which states that information with a band width of bHz must be sampled at a minimum of 2bHz if the information is to be retrieved. Since the MA-200 has a sampling frequency of 58Hz it was considered possible, in theory, to measure the vibrational displacements up to a maximum frequency of 29Hz. The resolution within the frequency domain is directly related to the frequency and the size of the sample obtained. The original prototype instrument had a sample size of 256 giving a frequency resolution of 0.226Hz. Currently, however, the MA-200 has a maximum sample size of 1024 samples and therefore yields a frequency resolution of 0.0566Hz.

8.4.2 Experimental Details

The evaluation of the Tellurometer MA-200 for earth vibration monitoring took the form of a number of simple experiments carried out both in the field and in the laboratory. The experimental work had four main objectives:-

1) to investigate the degree to which the MA-200 could measure the frequency and amplitude of:

   i) a simple harmonic vibration using a single degree of freedom model;

ii) simple harmonic vibration caused by a frequency vibrator;

iii) a more general periodic vibration using a multi degree of freedom model.

2) To consider the potential use of low cost retro-reflectors for vibrational analysis.

The results of the tests in 1(i), 1(iii) and 2 are reported in more detail in Kennie et al [16], pages B314 - B323 and those for 1(ii) in Howard et al (1989).

a) Model 1: Single Pendulum

In order to simulate simple harmonic motion, in other words a vibrating system with only one degree of freedom resulting in a single frequency and mode of displacement, a simple pendulum model was used. The model was constructed using a conventional retro-reflector as the suspended mass supported on a platform by two thin steel cable anchored to a reaction frame (see figure 1, page B315). The very simple sinusoidal displacement of the model allowed an initial evaluation of the measuring capacity of the instrument without the need for any post-processing of the acquired data.

b) Model 2: Variable Frequency Vibrator

In order to further simulate a simple harmonic displacement, a variable frequency vibrator was also used. An acrylic retro-reflector was mounted on the head of the vibrator (Howard et al, 1989). Some preliminary investigations (Kennie et al [16]), page B321 indicated that this type of low

cost retro-reflector could be used provided that the distance between the reflector and the instrument was within the range 10-70m. For the purpose of this initial test the MA-200 was set up on a conventional tripod at a distance of approximately 13m from the vibrator and in line with the axis of displacement. A number of tests were conducted at discrete frequencies of oscillation ranging from 1.55Hz to 28.99Hz (just below the cut-off frequency of 29Hz). The phase data collected for each test were transferred to an IBM-PC/AT computer for subsequent processing to give displacement-time data. In this case the more recent instrument with a sample size of 1024 was used.

c) Model 3: Double Pendulum

A more complex vibrational motion was achieved using a three degree of freedom model. The double pendulum used consisted of two glass retro-reflectors suspended from the reaction frame in series. The motion of the higher of the two masses was monitored and therefore three independent modes and associated frequencies of vibration detected. (see figure 2(a) and 2(b), page B315).

8.5 RESULTS AND ANALYSIS

In order to interpret the vibrational characteristics of the three models tested the frequency of the observed oscillation was required. For the single pendulum this was achieved by examination of the displacement-time data obtained. For the variable frequency vibrator and the double pendulum it was necessary to use Fourier analysis in order to separate out the frequencies which contributed to the observed oscillation.

8.5.1 Model 1: Single Pendulum

The displacement-time data collected using the MA-200 on the single pendulum model are shown in Figure 3, page B317. For simplicity only one complete period of the oscillation measured is shown. The motion of the pendulum can be seen to be sinusoidal in character as expected. The frequency of the oscillation was determined by interpolation of the data and was found to be 0.74Hz.

In order to independently assess the frequency measurement obtained, separate measurements were obtained using an accelerometer. This was processed using a Fast Fourier Transform (FFT) and the harmonic frequency was found to be 0.76Hz. The resultant data plotted in the frequency domain is shown in Figure 4, page B318. It can be seen from this data that a single harmonic frequency indicating simple harmonic or sinusoidal motion is evident. The difference in the value of the harmonic frequency obtained by the two measurements is thought to be attributable to the number of samples obtained. With a sampling frequency of 58Hz and a total number of samples of 256, the maximum frequency resolution (without interpolation) is 0.226Hz.

8.5.2 Model 2: Variable Frequency Vibrator

The results obtained were compared with the known frequency of oscillation of the vibrator and are given in Table 8.1. The data in the frequency domain for the case of an oscillator frequency of 20.02Hz is typical of the range of data obtained and is shown in Figure 8.1.

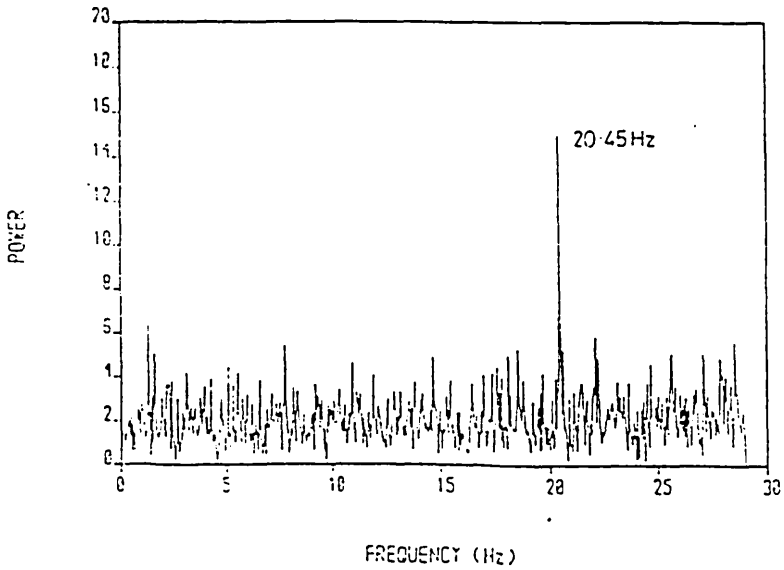


Figure 8.1

VIBRATOR OSCILLATOR FREQUENCY (Hz)	MA-200 FREQUENCY (Hz)
1.55	1.59
1.98	2.04
2.97	3.06
5.03	5.15
10.06	10.25
15.17	15.52
20.02	20.45
24.98	25.49
28.99	28.38

TABLE 8.1: FREQUENCY DATA

8.5.3 Model 3: Double Pendulum

The data collected using the MA-200 on the double pendulum model are shown in Figure 5, page B318. It can be seen that the motion depicted cannot be described by a single harmonic frequency but has two or more superimposed simple harmonic oscillations. Analysis of the displacement-time data using the FFT enabled definition of these compound frequencies. In order to demonstrate the effect of sample size on the resolution of the frequency data three Fourier transformations were performed on the data, using 64, 128 and 256 samples (Figures 6, 7, 8, pages B318 - B319). Part of the data collected using the accelerometer is shown in Figure 9, page B320. The sampling rate used was 100Hz and 4000 samples were taken over a period of 40s. Therefore, the frequency resolution is 0.025Hz. The data does not depict the smooth sinusoidal form of that collected by the MA-200. However, the periodic motion is at a lower frequency than the interfering noise. Hence, the relatively high frequency, which is probably attributable to electrical and vibrational noise, does not interfere with the harmonic frequency determination.

The FFT data for the acceleration response indicates there are three clear harmonic frequencies at 0.62Hz, 1.53Hz and 1.87Hz. In the case of Model 2 (with multiple harmonic

frequencies) the data from the accelerometer source confirms the existence of three fundamental frequencies. The data for the MA-200 also agrees closely to these results, the maximum variation being 0.06Hz, again well below the frequency resolution of the MA-200. It is interesting to note however the existence of two further high amplitude frequencies at 1.35Hz and 0.45Hz. This is again a product of the frequency resolution of the MA-200, since data points only exist at intervals of 0.226hz. In the FFT computation harmonics which exist between datapoints will become apparent in the analysis. For example in the case of the lowest of the three harmonics (0.62Hz) the MA-200 has identified 2 components adjacent to this at 0.452Hz (the second data point) and at 0.679Hz (the third data point). Given the relative magnitudes at these frequencies it is clear that the precise location of the harmonic lies closer to the latter of these results. A similar analysis can be used to explain the existence of a harmonic at 1.35Hz.

8.5.4 Investigation of Low Cost Retro-Reflectors

Virtually all EDM instruments require a method of reflecting the incident signal and this is usually achieved by means of precisely manufactured glass corner cube reflectors which return the signal along an identical path back to the instrument. Such reflectors are often referred to as retro-reflectors. However, it is also possible to use less precisely manufactured, and consequently cheaper, retro-reflective devices and the characteristics of a number of these are discussed in Kennie [14] and [15].

In the field of vibrational analysis such retro-reflectors have a number of advantages (in addition to cost) over conventional glass retro-reflectors. In particular from a practical point of view they are more robust, lighter and are easier to use than the more conventional glass

corner cube reflectors.

The particular retro-reflectors tested were:-

a) multiple, glass corner cube design (GCC) (26 and 85mm diameter).
b) clear, reflex acrylic corner cube design (ACC) (11 and 35mm diameter).

Measurements were obtained of distance, return signal strength and scatter of phase of each reflector type over a range of distances from 7m to 200m.

A summary of the results is provided in graph 11 of signal strength against distance and is shown on page B322 of Kennie et al [16]. It can be seen that as one would anticipate, signal strength decreases with increasing distance from the instrument to the prism. The loss of signal strength at longer distances is primarily caused by imperfect reflection and loss of energy of the signal.

The fall in signal strength for both designs as distance from the instrument is decreased below 22m is anomalous. These results may be explained by the biaxial geometry of the MA-200 transmitting and receiving optics. The receiving optic is located 63mm to the side of the transmitting optic. Therefore, to ensure reception at the receiving optic, a reflector would have to cause a sideways shift in the beam of 63mm. Over shorter distances when the beam has not diverged, the low cost retro-reflectors designs do not cause this shift. However, as the distance increases and the beam becomes wider, more of the reflected beam is directed into the receiving optic.

In summary, the results indicate that although the signal strengths are much lower for the low cost retro-reflectors, they are potentially capable of being used

to measure the vibration with the condition that the distance from the instrument is not greater than 70m and not less that 7m (in the case of the ACC design).

8.6 CRITICAL ANALYSIS AND FUTURE WORK

It can be concluded from the preliminary experiments on vibrational data acquisition and analysis using the MA-200 that:-

a) The maximum frequency resolution of the instrument in its original prototype form was 0.226Hz. For applications where determination of vibrational frequencies of the order of 0.5Hz is acceptable this limitation is clearly insignificant. In its current form the instrument now has a frequency resolution of 0.0566Hz.

b) Low cost retro-reflectors are capable of being used for vibrational measurement provided that the range does not exceed 70 metres.

Finally, a common problem in signal analysis, using discrete data, is aliasing. Aliasing results in harmonic frequency peaks, which exist at frequencies higher than the Nyquist cut-off frequency, being folded back into the frequency range below the cut-off frequency. Where the data is continuous filters can be used to remove higher unwanted frequencies before transformation of the data. However, with discrete data filtering is not possible. To avoid problems with aliasing in this case, aliased data needs to be identified as such and hence removed after the transformation. If displacement-time data is sampled at several different frequencies on the same vibrational signal, the aliased products will become apparent by the changing position on the frequency axis. At present the MA-200 does not allow for different sampling rates. It is considered, therefore, that a modification to the

instrument to facilitate this would be necessary to minimise such effects.

8.7 REFERENCES

HOWARD C.M., KENNIE T.J.M., HOWELL J., AND ALLAN A.L., 1989. Applications of the Tellurometer MA-200 in Industrial Metrology, Vibrational Analysis and for High Precision Field Measurements. CONSAS 1989. Paper 12.4, 16pp.

CHAPTER 9

CONCLUSIONS

9. CONCLUSIONS

9.1 INTRODUCTION

The data acquisition and processing needs of engineering geologists have changed considerably over the past decade. Concurrent with these changes have been parallel developments in the technology used in modern land surveying. This thesis has been concerned with the extent to which the changing user needs of the engineering geologist may be satisfied by modern land surveying technologies and procedures. This final chapter synthesises the main conclusions produced by the research projects discussed in the Case Studies and, on the basis of these results, provides some guidance on the recommended use of these new technologies at the various stages of an engineering project. (Section 9.2). Finally, the thesis concludes by discussing the main problems which may be encountered in introducing these new approaches to solving current and future engineering geological problems.

9.2 SUMMARY OF MAIN CONTRIBUTIONS AND RECOMMENDED USE OF SURVEYING TECHNOLOGY FOR THE SOLUTION OF ENGINEERING GEOLOGICAL PROBLEMS.

The role of engineering geology during the construction process has developed considerably over the past 10-15 years and in Chapter 1 the contribution of the discipline at the various stages of a civil engineering project was outlined.

In general terms it is possible to identify five distinct stages of a project and some of the possible areas of practice for the engineering geologist are summarised in Table 9.1. The data acquisition and analysis needs of the engineering geologist can be broadly categorised into those which involve remote investigation techniques (remote sensing from aerial and space platforms and geophysics) and field measuring techniques (using ground and satellite measuring devices and aerial/close range photogrammetry).

Stage of Project	Typical Areas of Practice
Reconnaissance	Terrain analysis Surficial materials mapping Regional subsidence studies General eng. geological mapping e.g. subsidence hazard assessment
Feasibility	Landslide hazard assessment Environmental impact analysis
Design	Slope instability studies Foundation condition studies
Construction	Ground movement studies
Post Construction	Monitoring of ground condition, and stability of slopes.

Table 9.1

Relationship between stages of a civil engineering project and typical areas of practice for engineering geologists

The unique characteristics of each of these distinct surveying techniques were reviewed in Chapter 2 and a number of recommendations made about the suitability of various satellite sensors for topographic mapping (Table 2.5, page 20) and thematic mapping for engineering projects (Table 2.7 page 23). These recommendations were based on work published in Kennie [1], [2], [3] and Kennie et al [4] and appended to the thesis (pages A1 - 130). All four

publications formed part of a book edited by the author which reviewed the subject of 'Remote Sensing in Civil Engineering'.

The role of ground geophysics for engineering geological site investigations was subsequently reviewed (section 2.1.3 pages 26 - 36). Although a number of successful cases have been reported in the literature, the history of ground geophysics in engineering geology has been relatively unsuccessful. Indeed whilst the technology is now reasonably mature, there is still a need for guidance on good practice to ensure that the correct techniques are used to address the engineering geological problems under investigation. This is the prime reason why poor or inconclusive results have often been obtained in practice. In addition, improved targetting methods are required to ensure that the geophysical techniques being used at the site investigation stage of a project are in very close vicinity to the features/phenomenum under investigation. Aerial remote sensing techniques have a key role to play at the reconnaissance stage in this process.

The following sections discuss in more detail the role of these remote investigation techniques and field measuring techniques at the various stages of an engineering project. To provide a basis for comparison each technique will be reviewed on the basis of three criteria. The first will be its operational level; many potentially useful techniques may only be available on a limited basis at present (because of coverage), or may only be at a research level of development. Secondly, an assessment of the techniques effectiveness to solve specific types of engineering geological problems will be made. Finally, the vitally important matter of cost effectiveness will be considered and an assessment made of the techniques cost/benefit relationship. The results of this comparative assessment are illustrated by Tables 9.2 and 9.3.

Tables 9.2 and 9.3 illustrate the conclusions reached for

both the remote investigation techniques and field measuring techniques. Examination of the conclusions provided in these tables can assist engineering geologist's in the determination of those techniques which are of potential benefit at each stage of an engineering project. More detailed examination of the data contained within the tables indicates the following general trends;

(a) Remote Investigation Techniques (Table 9.2)

(i) The pre-eminence of satellite remote sensing and aerial photography at the reconnaissance and feasibility stages of a project (c.f. airborne thermal scanners, MSS and ground geophysics).

(ii) The increased role of ground geophysics at the design stage of a project where it would normally form part of the site investigation process.

(iii) The diminished value of satellite imagery after the feasibility study is complete .

(b) Field Measuring Techniques (Table 9.3)

(i) The growing cost effectiveness of satellite positioning systems at a number of stages of the project, particularly at the reconnaissance and feasibility stages.

(ii) The key role of electronic ground surveying equipment throughout the entire process.

(iii) The potential for utilising close range photogrammetry, particularly at the design and construction stages.

The following sections discuss these general trends in more detail and in the specific context of the case studies reported in Chapters 4 - 8.

Table 9.1

Overall Effectiveness of Remote Investigation Techniques

Stage of Project and Typical Areas of Practice	Remote Investigation Techniques														
	Satellite Remote Sensing			Aerial Photog.			Airborne Thermal			Airborne MSS			Ground Geophysics		
Reconnaissance	O	E	C	O	E	C	O	E	C	O	E	C	O	E	C
- Terrain analaysis	o	*	x	o	*	x	o	*	x	o	*	x	-	-	-
- Regional subsidence studies	o	*	x	o	*	x	o	*		o	*	x			
- Surfical materials mapping	o	*	x	o	*	x									
- General thematic mapping															
Feasibility	o	*	x	o	*	x	o	*	x	o	*	x	o	*	x
- Landside hazard assessment	o	*	x	o	*	x	o	*	x	o	*	x	o	*	x
- Subsidence hazard assessment	o	*	x	o	*	x									
Design	o	*	x	o	*	x	o	*	x	o	*	x	o	*	x
- Slope instability studies				o	*	x	o	*	x				o	*	x
Environmental impact assessment															
Construction	-	-	-	o	*	x	-	-	-	-	-	-	o	*	x
- Ground movement studies				o											
Post Construction	-	-	-	o			-	-	-	-	-	-	-	-	-
- Vibration analysis				o	*	x									

Operational level

o In widespread use
o
o

o Use limited at present
o

o Research use only

- Not in use (at present)

Effectiveness

* Very effective
*
*

* Effective
*

* Of limited use

- Inadequate

Cost Effectiveness

x Highly
x cost effective
x

x Moderately
x cost effective

x Only in limited cases

- Not cost effective at present

Table 9.2 Overall Effectiveness of Field Measuring Techniques

Stage of Project and Typical Areas of Practice	Field Measuring Techniques														
	Satellite Positioning			Ground Surveying			Aerial Photogm.			Close Range Photogm.			Indirect Techniques		
	O	E	C	O	E	C	O	E	C	O	E	C	O	E	C
Reconnaissance															
- Terrain analaysis	o	*	x	o	*	x	o	*	x	o	*	x	-	-	-
- Regional subsidence studies	o	*	x	o		x	o	*	x						
- Surficial materials mapping							o	*	x						
- General thematic mapping															
Feasibility	o	*	x	o	*	x	o	*	x	o	*	x			
- Landside hazard assessment	o	*	x	o	*	x	o	*	x				-	-	-
- Subsidence hazard assessment				o			o	*	x						
Design	o	*	x	o	*	x	o	*	x	o	*	x	-	-	-
- Slope instability studies				o	*	x	o	*	x	o	*	x			
- Environmental impact assessment										o	*	x			
Construction				o	*	x				o	*	x	o	*	x
- Ground movement studies	o	*	x	o	*	x	o	*	x	o	*	x	o	*	x
				o	*	x				o	*	x	o	*	x
Post Construction				o	*	x							o	*	x
- Vibration analysis	-	-	-	o	*	x	-	-	-	-	-	-	o	*	x
				o	*	x							o	*	x

Operational level		Effectiveness		Cost Effectiveness	
o o o	In widespread use	* * *	Very effective	x x x	Highly cost effective
o o	Limited at present	* *	Effective	x x	Moderately cost effective
o	Research use only	*	Of limited use	x	Only in limited cases
-	Not in use (at present)	-	Inadequate	-	Not cost effective (at present)

9.2.1. Reconnaissance Stage

(a) Remote Investigation Techniques

The reconnaissance stage of a project is concerned with the analysis of broad regional trends and/or terrain analysis. Consequently satellite remote sensing can have a valuable role to play at this stage in the production of soil, vegetation, rock or hydrogeological data. Some of the factors influencing the choice of sensor to use in the developing and developed world are discussed in Kennie et al [4], (pages A120-127), and Kennie [3] (pages A106-A109) provides some typical examples of the use of satellite remote sensing at this stage of a project.

High altitude satellite or aerial photography may also be used to good effect at this stage of a project. Low altitude airborne remote sensing devices may also be used for example, to identify zones subject to subsidence or hazardous to construction. Techniques such as airborne multispectral scanners or airborne thermal linescanners are of specific relevance. The results of an investigation to assess the capability of such devices for the detection of solution features was discussed in Chapter 4; Kennie and Edmonds [5] and Kennie et al [6] in Appendix A.

The main conclusions of this piece of research were, as far as the detection of solution features in Southern England is concerned, that;

(i) Interpretation of aerial photography and thermal linescanner imagery acquired in autumn proved more effective than that acquired in mid-summer and spring.

(ii) Uncorrected multispectal scanner data is of very limited value for automated detection of solution features, although some practical recommendations

to enable such processes to be applied were provided (pages 82 to 86).

(iii) A number of practical difficulties to the detection of features were identified (section 4.7.3, page 86).

(b) Field Measuring Techniques

The main role of the various field measuring techniques at this stage is to provide ground control and detailed planimetric and height data for topographic mapping operations. The range of equipment used for modern land surveying operations is discussed in Kennie [11] (page B218-242) and Kennie [12] page B243-248) and for photogrammetric operations in Kennie [9], (pages B181 - B209). The role of satellite positioning systems such as the Global Positioning System (GPS) is reviewed in Chapter 2 pages 37 - 38.

Overall all of these techniques have a vital role to play at this stage of a project with aerial photogrammetry being of particular significance.

9.2.2 Feasibility Stage

(a) Remote Investigation Techniques

The purpose of the feasibility stage of a project is to select a preferred route or site for a project from the available options. To do so may require more detailed analysis of the engineering and terrain characteristics of the ground surface together with more detailed hazard assessment models. In this context the developing field of Geographic Information Systems discussed in Chapter 3 is of particular significance. Such systems enable the spatial data acquired from remote sensing or ground based thematic mapping to be combined with attribute data from previous engineering investigations or borehole/drilling campaigns.

Hazard assessment models based on such software can be used to determine zones subject to specific geological problems. At this stage the use of the high resolution satellite imagery such as that available from SPOT can be used more efficiently and integrated with conventional aerial photography.

The role of the field measuring techniques at this stage is very similar to their use at the reconnaissance stage; the main distinction being the scale of the topographic base maps being produced. Whereas the reconnaissance stage may involve the production of small scale regional mapping (at scales from 1:500,000 to 1:50,000), at the feasibility stage the map scale is more likely to be in the range from 1:50,000 to 1:10,000.

9.2.3 Design Stage

As outlined in Chapter 1 the role of the engineering geologists investigations at this stage may be for two distinct purposes. In the first case the data may be required as part of the formal site investigation process. In the second the data may be necessary to solve some of the environmental problems related to a construction project.

(a) Remote Investigation Techniques

The use of remote sensing at the design stage is more generally restricted to the potential of airborne sensors and Table 9.5 from Kennie [3] summarises a selected series of references to the use of this technology to satisfy some of the site investigation needs.

	Aerial photography	MSS	Thermal IR linescanning
Landslides/ slope instability	Matthews and Clayton (1987)	Liu (1985)	Chandler (1975)
Solution features	Norman and Watson (1975)	Coker *et al.* (1969)	Edmonds *et al.* (1986)
Derelict/ contaminated land	Bullard (1983)	Coulson and Bridges (1984)	Coulson and Bridges (1984)
Surficial subsurface materials surveys	Caiger (1970)	Lynn (1984)	Singhroy and Barnett (1984)

Table 9.4
Use of Remote Sensing During the Site Investigation Stage of a Project (Kennie [3])

In relation to the environmental engineering needs of construction projects the potential of a transportable thermal imaging devices for operational remote sensing is reported in Chapter 5, Kennie et al, [7] and Stove et al, [8]. In Kennie et al [7], the paper reported a joint research project between the Universities of Surrey and Glasgow to assess the capability of the Barr and Stroud IR-18 for topographic mapping and civilian remote sensing. This was the first published report of the use of this instrument from a light aircraft. The project considered the characteristics of this unique design of sensor and led to the coining of the term Thermal Video Frame Scanner (TVFS) to describe the scanning system. The initial experimental tests highlighted the potential of such relatively low operational cost thermal scanners for a range of engineering projects. This subsequently led to the instrument being used commercially for operational remote sensing by ERSAC (Environmental Remote Sensing Applications Centre) Ltd (Stove et al, [8]).

In the second case (Stove et al, [8]) the instrument was used to delineate areas of road surface susceptible to ice formation in the Lothian region of Scotland. This was the first application of this remote investigation technique in the United Kingdom and successfully proved the cost effectiveness of this approach for highway maintenance engineering.

A separate, but related engineering need at this stage of a project is the capability to provide engineers with data which illustrates the environmental impact of a project. A significant element of such a process may be the ability to consider the matter of visual impact analysis. Such a process is often based nowadays on the interpretation of a Digital Terrain Model (DTM). The basic principles of this topic were discussed in Chapter 3 and in Petrie and Kennie [18]. This work was further extended in Kennie and McLaren [19] which detailed the procedures necessary to create complex computer generated visualisations of the terrain and objects placed on the terrain.

9.2.4 Construction Stage

(a) Remote Investigation Techniques

The use of remote investigation techniques at this stage is somewhat limited in comparison to the previous three stages. The needs at this stage are for more detailed data related to the on-going construction process. One of the more common uses of for example, aerial photography at this stage is to record the process of construction. In some instances this may be solely to provide a visual record of progress whereas in other cases photogrammetric measurements may be taken to assess, for example, volumetric changes relating to gravel extraction or general earthworks measurements.

(b) Field Measuring Techniques.

A review of the basic principles of modern land surveying relevant to the engineering geologist and geotechnical engineer was reported in Kennie [11]. This theme was developed further in Kennie [12] which considered the development of electronic angle and distance instruments together with methods for storing field data. Instruments such as those described in this publication are now widely used to provide the base mapping for large scale geological investigations.

Such electronic surveying equipment may also be used for monitoring structures during the construction phase of a project and Chapter 7 summarised the application of a novel retro-reflective target for measuring the effect of soil 'spill-through' on a series of bridge abutments. This work formed part of a research contract to monitor the earth pressures and movement of the bridge abutments on the A3/M25 Wisley interchange near Guildford, Surrey. Sequential measurements from a series of control points were used to determine the relative movement of a series of retro-reflective targets attached to the wing wall which formed part of the abutment. The targets consisted of small acrylic reflectors. The performance of this form of retro-reflector (and others) was previously investigated by the author and reported in Kennie [14-15]. The results from the original work, which formed part of a dissertation for a Masters degree at the University of Glasgow, enabled this simple, cost effective technique to be developed and the technique is now finding increasing application in a number of deformation monitoring operations.

9.2.5 Post-Construction Stage

(a) Remote Investigation Techniques

Again as with the construction phase of a project the

opportunities to utilise satellite imagery or airborne sensing devices are very limited. Aerial photography may be used for monitoring purposes, for example, in relation to traffic flow rates. The use of the TVFS system outlined in Chapter 5 also indicated the potential of such devices to evaluate temperature variations across road surfaces at the post-construction phase.

(b) Field Measuring Techniques

The potential of a prototype electromagnetic distance measuring instrument for measuring the vibrational characteristics of the ground and structures was reported in Chapter 8 and Kennie et al, [16]. The results reported in Chapter 8 and the paper cited above were the first experimental investigations of the instrument's performance for measuring natural and man made sources of vibration. The results clearly identified the range of frequencies which the instrument could resolve and this subsequently led to a number of design modifications being made to the prototype instrument to improve the frequency resolution of the instrument.

The significance, and importance, of instrument calibration is relevant to all forms of electronic distance measurement and this was further investigated by the author (Kennie et al, [17]). This publication provided the results of two questionnaire surveys, by the University of Surrey and the National Physical Laboratory (NPL), to assess the EDM instrument calibration facilities available in the UK. The results of the survey highlighted the need for a national standard for EDM instrument calibration. The author subsequently carried out a further study for the Royal Institution of Chartered Surveyors (RICS) on the technical standards for calibration as part of a Quality Assurance (QA) schedule, (Kennie, 1988). QA is now finding increasing application in many fields of engineering and surveying and a growing number of organisations are establishing procedures which will allow them to satisfy the requirements

of the QA registering bodies and comply with the appropriate British Standard (BS 5750). This work has subsequently led to the formation of a joint NPL/RICS Working Party to draft regulations on EDM calibration for submission to the National Measurement Accreditation Service (NAMAS), an independent organisation within the NPL.

9.3 PROBLEMS ASSOCIATED WITH THE INTRODUCTION OF NEW SURVEYING TECHNOLOGY INTO ENGINEERING GEOLOGY

Technological developments of the forms described in the case studies can potentially lead to increased productivity for practising engineering geologists. To maximise these benefits, however, there are a number of difficulties which need to be considered by practitioners considering their use. To conclude the thesis this section highlights some of the most significant of these problems.

9.3.1 Financial Investment

The hardware and software costs required to invest in many of the new surveying technologies discussed in the preceding sections can be vast. Against this must be balanced the increased data volume and flexibility which can also be achieved. Users must be aware of being overwhelmed by the glossy benefits which may be promoted by the computer marketing personnel. A firm appreciation of the real user needs is absolutely vital if engineering geologists are to avoid costly purchasing mistakes.

9.3.2 Black box technology

The ease of use of modern field survey equipment, analytical photogrammetric equipment and image processing systems continues to improve. However, associated with these improvements and refinements are concerns about the inappropriate use of such technology. The specific case of EDM instruments has been discussed previously and highlights these concerns. Such instruments are now used widely,

although it has yet to become accepted practice to ensure that such equipment is regularly calibrated. Consequently there has been an increased tendency to blindly accept the results from EDM and other surveying devices without question i.e. to use the instrument as a 'black box'. Similar concerns also exist with regard to the use of other devices such an analytical photogrammetric equipment or DTM systems. Blind acceptance of the results can also occur because of a confused and often superficial understanding of the capabilities of such instruments or systems. In many cases, this confusion is further enhanced by the technical literature on the subject which frequently suffers from an over reliance on technical jargon and ill-defined terminology. The engineering geologist utilising such technology will need a much higher level of understanding of the basic principles of the technology than was the case in the past.

9.3.3 Practical difficulties

A number of basic conditions are necessary for the efficient use of these technological advances. Although many of these are more serious in developing countries, elementary matters such as unreliable mains electrical supplies, unavailability of spares, lack of foreign currency, poor servicing facilities, and distance from the systems supplier can also be major disincentives in the so-called developed world.

9.3.4 Incompatibility

Even with the simplest of electronic instruments and computer hardware problems of incompatibility can occur. In theory problems of hardware compatibility should be less of a problem since the introduction of internationally agreed standards. In practice hardware interfacing is a minefield for the unwary. In the field of software, problems of data interchange can arise, although some progress has been made in providing standard data interchange formats.

9.3.5 Training

There is no doubt that the use of electronic computer based technology of the types discussed in this thesis will become increasingly more important in the future for processing spatial data. To enable such devices to be used efficiently there is a need for significant levels of technical knowledge and this will increasingly necessitate staff retraining. The hidden cost of acquiring this technical knowledge is often not fully appreciated, yet it may be quite significant. Furthermore, rather than being a 'one off' process most new technology is itself constantly evolving and thus also requires the user to continually keep up to date. This is a major problem for the future. Personnel may view this process as either challenging or a continual battle. A willingness to accept, and become accustomed to, a high level of change will undoubtably become increasingly important for engineering geologists in the future. The educational challenge for the future is to ensure that the new skills required to operate in this new environment are integrated into existing degree courses for engineering geologists.

9.4 REFERENCES

KENNIE T.J.M. 1988. A Comparative Study of EDM Instrument Calibration Facilities in the UK and Australia. Unpublished Report to the Education Trust of the Royal Institution of Chartered Surveyors, 14pp.

APPENDIX A

DEVELOPMENTS IN AIRBORNE AND SATELLITE REMOTE SENSING AND APPLICATIONS TO ENGINEERING GEOLOGICAL INVESTIGATIONS

1 An introduction to remote sensing

T. J. M. KENNIE

1.1 Introduction

The use of aerial photography as a means of assessing ground conditions can be traced back to the middle of the last century. These early photographs, obtained by the French army from hot air balloons, were soon realized to offer considerable benefits. The instantaneous record of the terrain not only provided an overall view of the ground surface, but also enabled specific sites to be examined in great detail. It was not however until the invention of the aeroplane had taken place at the beginning of the 20th century that aerial photography could be obtained over large areas on a repetitive basis.

The need for strategic and tactical reconnaissance during World War I gave rise to several major advances in the technology of obtaining aerial photography. In addition, the need for up-to-date maps gave added impetus to the developments which were taking place in aerial photogrammetry. Following World War I the use of aerial photography increased dramatically in many fields, including geology, forestry, agriculture, land surveying and civil engineering and led to the manufacture of much-improved camera systems. Concurrent developments in aircraft technology meant that by the outbreak of World War II the sciences of aerial photo-interpretation and photogrammetry were well established. Again the war provided the impetus for further developments, both in conventional photographic systems, and also in non-photographic systems such as radar.

Following this period the pace of technological advance increased enormously, particularly during the early 1960s. This led to the coining of the term 'remote sensing' by a group operating in the US Office of Naval Research, to describe this 'new' branch of science which was developing.

In its broadest sense remote sensing could be defined as 'the science of acquiring information about some property of an object through the use of a measuring device which is not in physical contact with the object under investigation'. Clearly, techniques such as magnetometry, gravimetry, sonar and spectroscopy could all thereby be considered to be examples of remote

sensing systems. The definition of remote sensing has, however, developed a much more restricted meaning in the world of science. A more precise and more generally accepted definition of remote sensing therefore is that provided by the American Society of Photogrammetry (Anon., 1983).

> Remote Sensing is an application of photogrammetry in which imagery is aquired with a sensor, other than (or in addition to), a conventional camera through which a scene is recorded, such as by electronic scanning, using radiations outside the normal visual range of the film and camera—microwave, thermal infrared, ultraviolet as well as multispectral. Special techniques are applied to process and interpret remote sensing imagery for the purpose of producing conventional maps, thematic maps, resource surveys, and so on, in the fields of agriculture, archaeology, forestry, geography, geology and others.

Whilst this definition is a useful starting point it is rather misleading in that by stating 'remote sensing is an application of photogrammetry' the impression may be given that the subject is entirely quantitative, involving precise measurement of the image. In practice a substantial proportion of remote sensing work is qualitative in nature, involving subjective interpretation of the imagery which is produced.

In general terms however the subject can be seen to consist of two distinct stages: firstly that of data acquisition involving the use of different 'remote sensors' and, secondly, data analysis in order to process and interpret the data. The primary aim of this chapter will be to discuss in more detail the procedures required in order to process and interpret the data. However, before discussing the attributes of specific acquisition systems in detail, it is useful to briefly examine the basic physical principles underlying the subject of remote sensing.

1.2 Basic physical principles of remote sensing

In physical terms, remote sensing is concerned with the measurement and recording of the variations in electromagnetic (EM) energy which occur when energy of this type interacts both with the earth's atmosphere and with the earth's surface.

The term EM energy refers to all energy which travels in a periodic harmonic manner at the velocity of light. Using the well-known relationship

$$\lambda = c/f$$

where λ is the wavelength, f the frequency and c the velocity of EM energy, it is possible to categorize EM energy according to its position along a continuum of wavelengths commonly referred to as the EM spectrum.

1.2.1 *The electromagnetic (EM) spectrum*

The range of wavelengths, and the names generally assigned to those regions of the EM spectrum which are of interest for remote sensing purposes, are shown

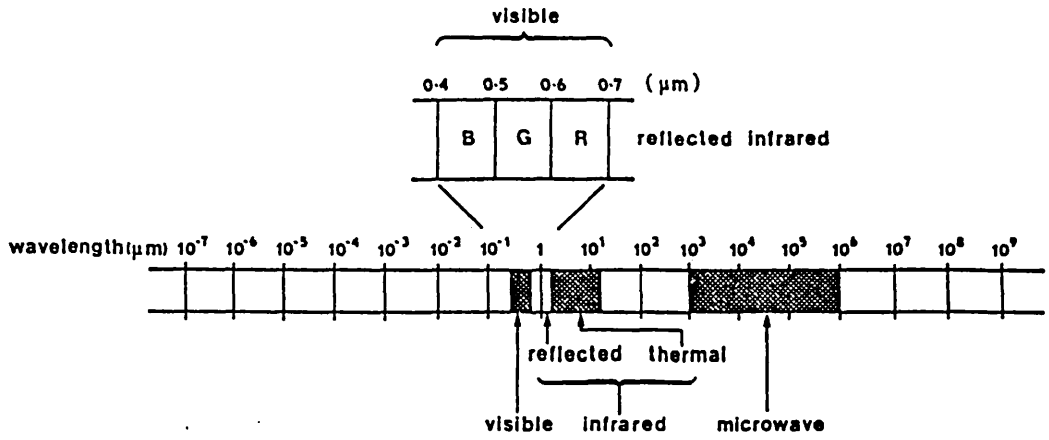


Figure 1.1 The electromagnetic spectrum and its relationship to remote sensing systems.

in Figure 1.1. The wavelength bands indicate broad regions of the EM spectrum and the boundaries between regions should be considered as general areas rather than as highly specific cut-off points.

Several regions of the EM spectrum are of particular interest for remote sensing, those of most importance being the visible, infrared and microwave regions. The visible region of the EM spectrum lies within the spectral range from about 0.36 to 0.75 μm. The significance of this region stems from the fact

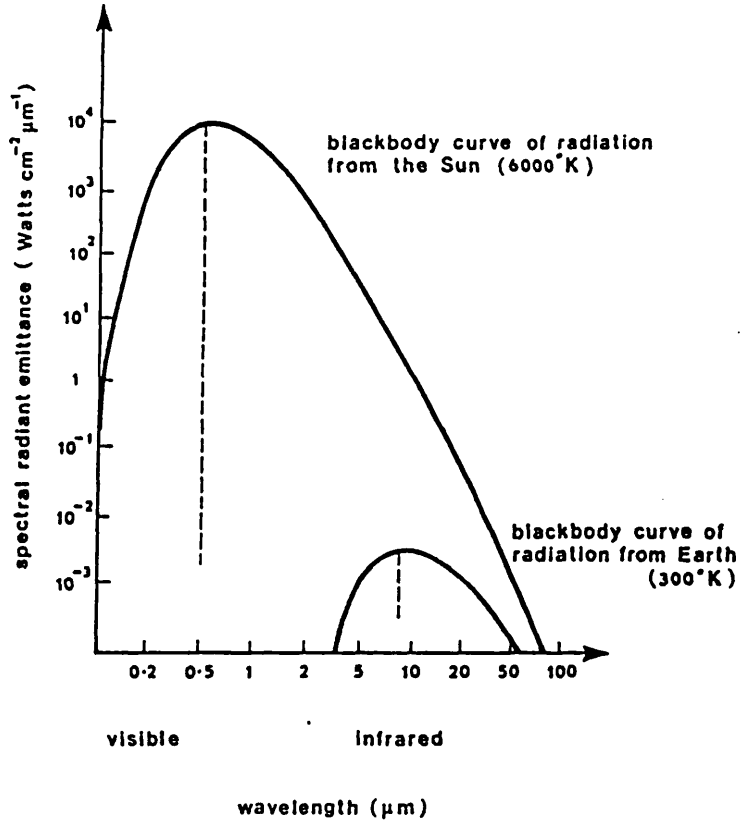


Figure 1.2 Spectral energy distribution of black bodies at temperatures approximating to those at the surface of the Sun and the Earth.

that the radiation emitted from the sun reaches a peak value within this region, at about 0.5 μm. This can be illustrated by Figure 1.2 which shows the spectral energy distribution for a theoretically perfect emitter or 'black body' at a temperature of 6000 K. Figure 1.2 also illustrates the spectral energy distribution associated with the earth's surface, assuming an ambient temperature of 300 K. In this case it can be seen that the peak value of emittance occurs at 9.7 μm, within the infrared region of the EM spectrum.

The infrared portion of the EM spectrum is represented by the spectral band which lies between the visible and microwave regions. Within this broad band there are two regions of significance. The first is termed the reflected infrared and the second the thermal infrared region. The latter is centred around the peak spectral emittance value associated with the surface of the earth.

The third significant region of the EM spectrum, from a remote sensing point of view, is the microwave region. In this case, the range of wavelengths lies between 1 mm (10^3 μm) and 1 m (10^6 μm).

1.2.2 *Reflected and emitted EM energy*

A secondary feature of EM energy to which reference has already been made is the distinction between reflected and emitted energy. During daylight, the radiant energy from a scene consists of two components. Firstly, at wavelengths up to about 3 μm the energy is predominantly *reflected* sunlight. Since human eyes and photographic film are sensitive to reflected energy within this wavelength range, it is possible to discriminate between features on the basis of the amount of energy reflected by the object. In contrast, at longer wavelengths greater than about 3 μm the dominant type of radiant energy is that which is *emitted* by a body rather than that which is reflected.

The distinction between reflected and emitted radiation can best be illustrated by examining the infrared region of the EM spectrum. Within this region it is apparent that as the wavelength increases, the emitted infrared component becomes progressively more significant. However, unlike the reflected infrared it cannot be sensed by photographic emulsion and has to be detected by the use of some other medium such as, for example, crystal detectors.

It is possible to quantify the amount of energy a body is emitting by means of the Stefan-Boltzmann Law, which states that

$$W = C\varepsilon T^4$$

where W is the radiant emittance in watts m^{-2}

ε is the emissivity of the object

C is the Stefan-Boltzmann constant equal to 5.7×10^{-8} watts $m^{-2} K^{-4}$ and

T is the absolute temperature of the object (K).

Three significant conclusions may be drawn from this law. Firstly, it can be

Table 1.1 Emissivities of a selection of materials (within the 8 to 14 μm wavelength band).

Material	Average emissivity
Clay soil	0.98
Water	0.97
Sandy soil	0.88
Grass	0.88
Buffed stainless steel	0.16

seen that the energy emitted from a body increases very rapidly with an increase in temperature. The second point relates to the emissivity (ε) of an object. Emissivity is a factor which describes how efficiently an object radiates energy in comparison to a 'black body' which has an emissivity value of one. Emissivity is defined as:

$$E(\lambda) = \frac{\text{radiant emittance from an object at a given temperature}}{\text{radiant emittance from a black body at the same temperature}}$$

where $E(\lambda)$ represents the emissivity at a particular wavelength value λ. Emissivity is quoted as having a value between 0 and 1. Table 1.1 indicates some typical emissivity values which may occur for a selection of materials.

Because of emissivity variations it is clearly possible for features to be at the same temperature and yet have completely different radiant emittance values. Thirdly, given a knowledge of the radiant emittance of a body, together with its emissivity, it should be possible to infer indirectly the temperature of the body. This represents the basis of the technique of thermal infrared remote sensing.

It is also possible, using Wien's displacement law, which states that

$$\lambda_{MAX} = \frac{2897}{T}$$

where λ_{MAX} = wavelength of maximum spectral emittance
and T = temperature (K),
to quantify the wavelength at which maximum spectral radiance will occur. This helps explain the wavelength shift, or displacement, discussed previously which occurs in the values associated with the energy emitted from the sun (0.5 μm) and from the earth's surface (9.7 μm).

1.2.3 *EM radiation and the atmosphere*

The discussion so far has dealt with the general characteristics of EM radiation, without regard to the influence of the medium through which the signal passes: the earth's atmosphere. The principal effects of the atmosphere on remotely sensed EM radiation are illustrated by Figure 1.3. It can be seen that the atmosphere influences EM radiation in two respects, firstly by scattering and secondly by absorbing the EM signal.

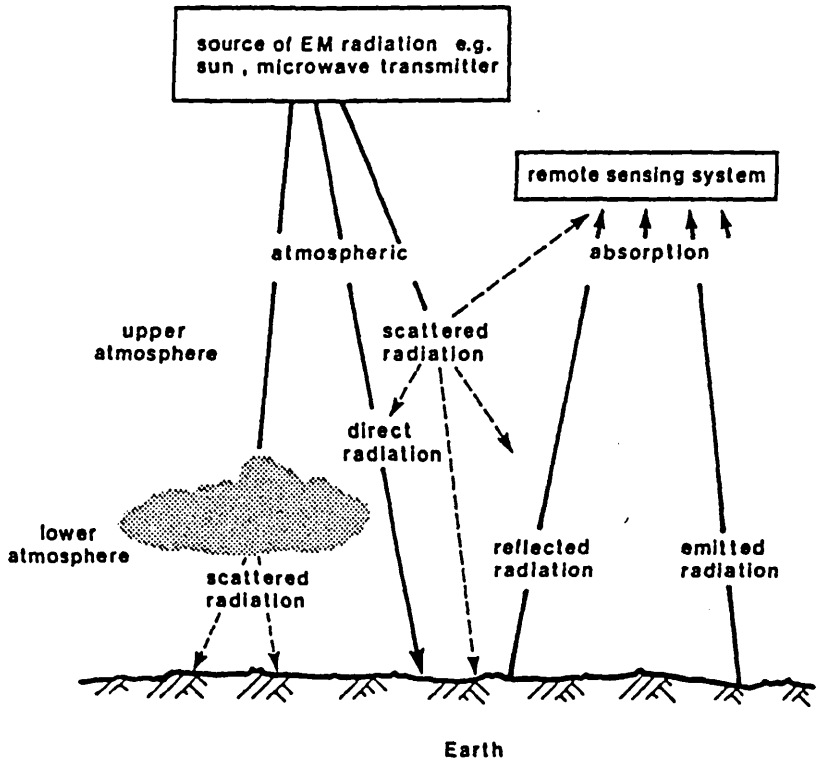


Figure 1.3 The influence of the atmosphere on EM radiation.

1.2.3.1 *Scattering.* Scattering of the incoming EM radiation to the earth, and of the reflected EM radiation from the earth's surface, is caused primarily by the presence of molecules of gas as well as dust and smoke particles in the atmosphere. In the upper atmosphere this scattering is caused predominantly by the interaction of gas molecules which have diameters much less than the radiation wavelength. Scattering of this type is termed 'Rayleigh scatter' and is one of the main reasons for the presence of haze on remotely sensed imagery. In contrast, scattering in the lower atmosphere is mainly a product of the interraction of dust and smoke particles with the EM signal. This type of scattering is termed 'Mie scatter'. A third type of scattering often referred to as 'non-selective scatter' occurs when the particles causing the scattering have diameters several times greater than that of the radiation wavelength. A good example of this type of scattering is that caused by water droplets (diameter $\sim$ 50 μm) as they interact with radiation within the visible spectrum ($\lambda \sim 0.5\ \mu$m).

1.2.3.2 *Absorption.* Absorption of EM radiation occurs primarily as a consequence of the attenuating nature of molecules of ozone, carbon dioxide and water vapour in the atmosphere. Because these gases absorb EM radiation in specific wavelength bands they govern which regions of the spectrum can be sensed.

Figure 1.4 explains this concept in relation to the visible and infrared parts of the spectrum.

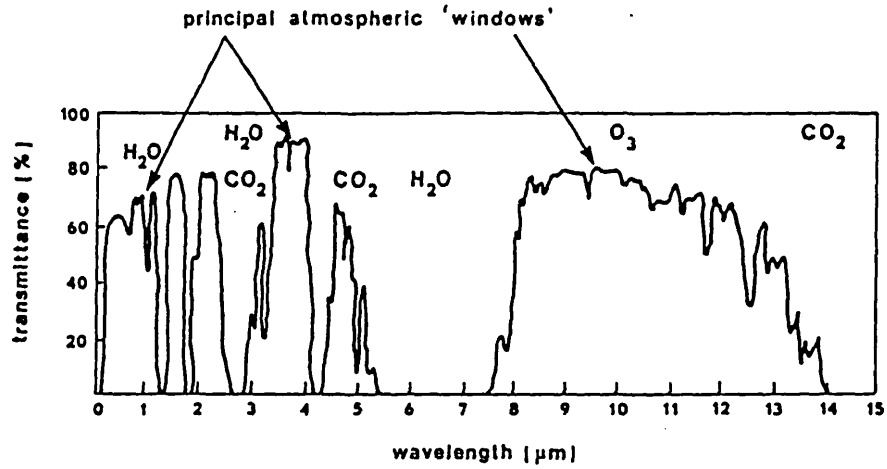


Figure 1.4 Atmospheric transmittance in the 0 to 15 μm wavelength range (the principal regions of absorption by gaseous elements of the atmosphere are also shown).

It can be seen from the graph that in several regions, for example the visible zone, the atmosphere is highly transmittive and consequently is almost totally free from the effects of absorption. Regions with a high atmospheric transmittance are generally referred to as atmospheric 'windows'. Not surprisingly, it is important that remote sensing systems should operate within those portions of the EM spectrum which coincide with these atmospheric 'windows'. A good example of this selective nature of remote sensing systems is the thermal infrared region. Figure 1.4 illustrates that within this wavelength region (3 to 14 μm), two atmospheric windows exist. The first is from 3 to 5.5 μm and the second from 8 to 14 μm. Not unexpectedly, these coincide with the operational bands for most thermal infrared remote sensing systems.

Similar atmospheric windows can be identified in other regions of the spectrum, such as for example, the microwave region. Whilst an understanding of the influence of the atmosphere is of critical importance in remote sensing, it is also essential to appreciate how the EM signal interacts with the terrain.

1.2.4 *EM radiation and the earth's surface*

When EM radiation is incident upon the earth's surface it will tend to be either reflected, absorbed or transmitted by that surface.

The magnitude and extent to which each of these processes occurs enables features to be differentiated from each other. For example, within the visible region of the spectrum, the concept of colour is simply a manifestation of the extent to which different features reflect different proportions of the visible spectrum.

It is also interesting to note that the same feature may appear very different when it is examined by sensors which are able to record reflectance or

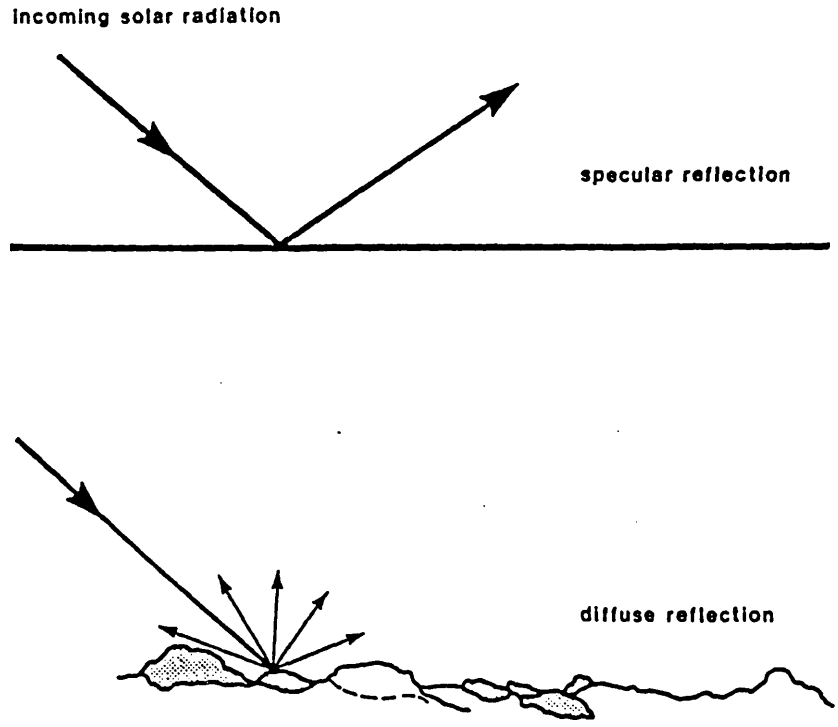


Figure 1.5 Specular and diffuse reflection and its relation to surface roughness.

emittance outside the visible spectrum. For example water surfaces appear black when imaged on to a photographic emulsion sensitive to reflected infrared radiation. This is a consequence of the almost total absorption of infrared radiation by water. In contrast, viewing the same water surface with a photographic emulsion sensitive to the visible region of the spectrum may provide details about submerged features which may not have been sensed previously.

The reflective properties of the terrain can be classified as being either specular or diffuse in nature. This concept is illustrated by Figure 1.5. It can be seen that this factor is essentially a function of the surface roughness. In cases where the surface is relatively smooth, a specular reflection occurs, whereas with rougher surfaces the reflection tends to be more diffuse in nature. Surface roughness in this context refers to the roughness of the terrain relative to the wavelength of the energy incident upon it. Rough surfaces tend to be those where the surface variations are equal to or greater than the wavelength of the source of illumination. An appreciation of this concept is crucial when radar imagery is being interpreted (see Chapter 3 for further details).

Finally, mention should also be made of the use of spectral reflectance patterns.

1.2.4.1 *Spectral reflectance patterns.* The reflectance characteristics of features on the earth's surface can be quantified by measuring the proportion of energy reflected by the feature at different wavelengths. For example, Figure 1.6

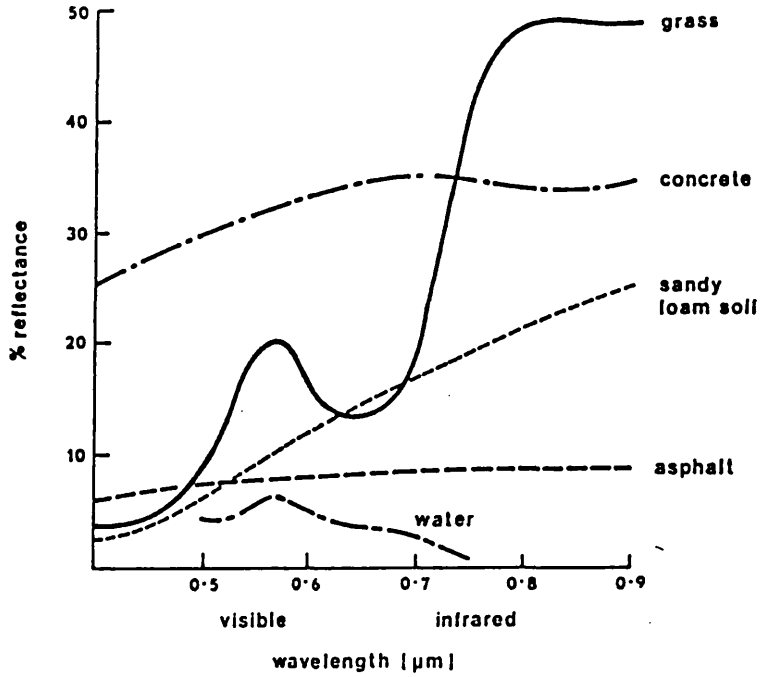


Figure 1.6 Spectral reflectance curves for selected features. (Adapted from Jensen, 1983.)

illustrates the spectral characteristics of several surfaces which may be of interest to the civil engineer.

Three significant points can be deduced from this example. Firstly, it can be seen that certain surfaces, such as grass, have considerably different reflectance characteristics in the visible and infrared regions of the spectrum. In contrast a surface such as asphalt has a relatively stable, low reflectance in both regions. Secondly it can be seen that the most appropriate region of the spectrum for differentiating between the surfaces is the infrared. In this region a clear hierarchy of reflectance exists, from grass with a high reflectance through concrete and sandy loam soil to asphalt and water which have a relatively low reflectance. In contrast, in the blue/green region of the visible spectrum ($\sim 0.5\,\mu$m) the discrimination between asphalt and sandy loam soil surfaces is very poor. This may indicate, for example, that difficulty may arise in the interpretation of these surfaces from conventional aerial photography. Thirdly, the graph again illustrates the very low reflectance of water surfaces in the infrared region of the spectrum.

This ability to spectrally define a feature or surface is often referred to as defining the 'spectral signature' of the feature. This term does, however, imply that features are uniquely and absolutely defined by measuring this parameter. There is, unfortunately, a degree of variability in these 'spectral signatures'. A more appropriate term which should be used is spectral reflectance pattern. This term is intended to suggest that features can be identified by a distinctive spectral response without implying that this response is necessarily unique.

1.3 Classification of remote sensing systems

Remote sensing systems can be classified on the basis of many different criteria. For example, one criterion which could be used as a basis for classification is the wavelength of the remotely sensed radiation. Using this criterion it would be possible to identify three broad categories: systems which sense either visible, infrared or microwave radiation. Classifications based on a single criterion such as this do, however, produce a rather general and also incomplete categorization.

A more complete classification can be obtained by categorizing systems on the basis of several different criteria, so producing a multidimensional classification. One such classification involves examining three different characteristics of remote sensing systems, namely whether the system is active or passive, photographic or scanning, and aircraft- or satellite-based.

1.3.1 *Active and passive systems*

Passive remote sensing systems are concerned with the detection of naturally occurring reflected or emitted radiation. In contrast, a remote sensing system which supplies its own source of energy is known as an active system. This system-generated energy, normally of microwave wavelengths (30 to 250 mm), is directed at the earth's surface and the multiple reflections from the terrain are recorded in order to produce an image.

1.3.2 *Photographic and scanning systems*

Two distinct methods can be used in order to detect and record the variations in reflected or emitted radiation which may occur when passive solar energy or actively generated microwave energy interact with the earth's surface.

The most common detection and recording technique used in remote sensing is the aerial camera. A lens and shutter assembly within the camera ensures that incoming radiation is focused and exposed instantaneously onto a photosensitive emulsion. The latent photographic image is subsequently developed by chemical processing, so producing a conventional aerial photograph. Photographic systems, however, only operate over a narrow wavelength band from 0.3 to 0.9 μm. They nevertheless produce relatively cheaply an image which is of high geometric fidelity and resolution.

An alternative detection and recording system which overcomes the spectral range limitations of the photographic systems is the airborne scanning system. The image produced by the airborne scanner differs fundamentally from the aerial photograph. Firstly, rather than the image being exposed at one instant by means of a shutter, the image is exposed sequentially, line by line, as shown by Figure 1.7. Thus a scanner image can be seen to consist of a series of adjacent scanlines perpendicular to the line of flight of the aerial platform. The second difference relates to the method of recording

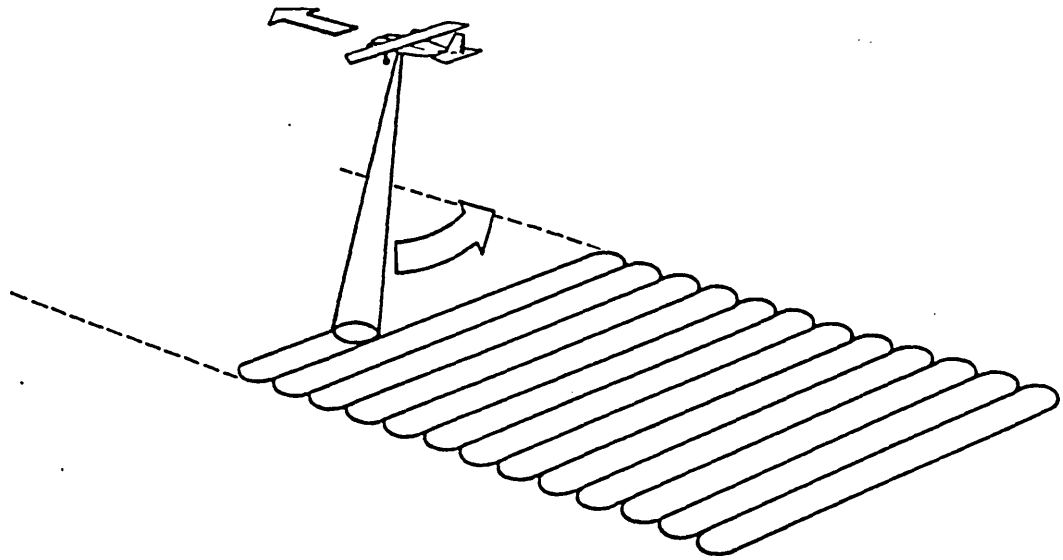


Figure 1.7 Airborne scanning.

the spectral variations over a scene. Conventional aerial photography records the variations in spectral reflectance by means of a film emulsion; in contrast, scanner systems employ crystal detectors. These detectors transform the incident reflected or emitted radiation into electrical signals. Consequently, using the scanner approach, the image is normally recorded directly in digital format on to magnetic tape. In addition, a much broader spectral range can be accommodated using these detectors. One significant disadvantage of the scanner approach is the much poorer spatial (ground) resolution associated with the image when compared with an equivalent-scale aerial photograph.

Finally, mention should be made of the distinction between the terms 'photograph' and 'image'. In remote sensing, the term 'photograph' is normally restricted to images that have been both detected and recorded by means of photographic film. In contrast, the more general term 'image' is normally used to describe any pictorial representation of image data. Thus, although scanner data may be used to create a photographic product, this should be referred to as a scanner image and not a scanner photograph, since the original detection mechanism was not a photographic film emulsion.

1.3.3 *Characteristics of sensor platforms*

The final characteristic to be used as a basis for classifying remote sensing systems is that of sensor platform.

Very broadly the distinction can be made between those remote sensing systems which are based on aircraft and those which are based on satellites as the sensor platform. The relative merits of each type of platform are summarized in Figure 1.8.

Finally, using these three characteristics of remote sensing systems, it is

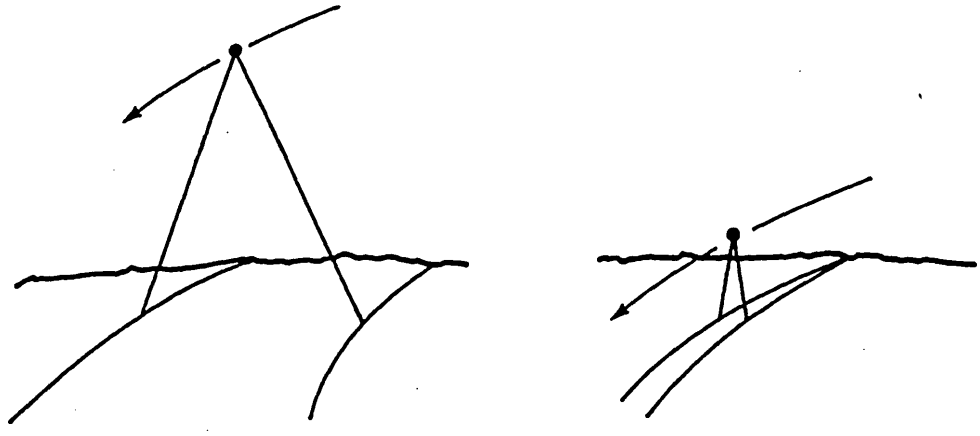


Figure 1.8 Relative merits of satellite and aircraft-based remote sensing systems.

Wide, synoptic view	Narrow, limited view
Very stable platform	Platform less stable
Small scale (> 1:500 000)	Large scales (1:1000 to 1:30 000)
Limited repeat coverage (16 to 18 days)	Flexible repeat coverage
Low spatial resolution (10 m to 500 m)	High spatial resolution
Highly cost effective	Less cost effective

Table 1.2 Classification of some typical remote sensing systems (see Chapter 3).

Mode of operation	Mode of recording			
	Photographic		Scanning	
Passive	Aircraft	Satellite	Aircraft	Satellite
	Wild RC-10 Mapping Camera	Skylab S-190B etc.	Daedalus 1230 Thermal Scanner	NOAA 1–8
		Spacelab		HCMM
	Zeiss (Jena) MKF-6M Multispectral Frame Camera	Zeiss (Ober.) RMK 30/23 Metric Camera	Daedalus 1268 MSS Scanner	Landsat MSS
			PNC MSS Scanner	SPOT MOMS
		Landsat RBV		
	TRRL Multi-spectral Camera			
Active			Radar	
			SAR 580	Seasat SIR-A/B ERS-1 Radarsat
			Lidar	
			LARSEN 500	

possible to devise a classification table which can be used in order to categorize any system. One particular classification is illustrated by Table 1.2.

Further details concerning remote sensing photographic systems are discussed in Chapter 2 and details of the scanning systems listed in Table 1.2 are discussed in Chapter 3.

References

Anon. (1983) What photogrammetric engineering and remote sensing is. *Photogrammetric Engineering and Remote Sensing*, 49:2.

Jensen, J. R. (1983) 'Urban/suburban land use analysis', in *Manual of Remote Sensing*, Vol. II, 2nd edn., Colwell, R. N. (ed.), American Society of Photogrammetry, Virginia, 2417 pp.

Bibliography

Barret, E. C. and Curtis, L. F. (1982) *Introduction to Environmental Remote Sensing*. 2nd edn., Chapman and Hall, London, 352 pp.

Bullard, R. K. and Lakin, P. J. (1981) *First Steps in Remote Sensing*. Working Paper No. 3, Dept. of Land Surveying, N. E. London Polytechnic, 61 pp.

Colwell, R. N., (Ed.) (1983) *Manual of Remote Sensing*. Vols. I and II, 2nd edn., American Society of Photogrammetry, Virginia, 2417 pp.

Lillesand, T. M. and Kiefer, R. W. (1979) *Remote Sensing and Image Interpretation*. John Wiley and Sons, New York, 612 pp.

Lintz, J. and Simonett, D. S. (1976) *Remote Sensing of Environment*. Addison-Wesley, Reading, Massachusetts, 694 pp.

Rudd, R. D. (1974) *Remote Sensing: A Better View*. Duxbury Press, N. Scituate, Massachusetts, 135 pp.

Slater, P. N. (1980) *Remote Sensing Optics and Optical Systems*. Addison-Wesley, Reading, Massachusetts, 575 pp.

Swain, P. H. and Davis, S. M. (1978) *Remote Sensing: The Quantitative Approach*. McGraw-Hill, New York, 396 pp.

3 Remote sensing scanning systems

T. J. M. KENNIE

3.1 Introduction

Photographic remote sensing systems are capable of sensing the visible and near infrared regions of the EM spectrum ($\lambda = 0.4$ to $0.9\,\mu$m). There are, however, other longer-wavelength regions of the EM spectrum, which are potentially useful for remote sensing purposes. Unfortunately, however, conventional glass lenses and photographic emulsions are not able to detect and record these regions of the EM spectrum.

An alternative method of detecting and recording both the visible, near infrared and longer-wavelength bands, such as thermal infrared and microwave energy, is the airborne scanner. The airborne scanner differs from an aerial camera in that it collects energy sequentially from the ground below, as a series of scanlines, each of which is perpendicular to the line of flight (see Figure 1.7). Scanning is carried out by a rotating mirror, which directs the incoming radiation on to a detector. The scanner detector is analogous to the emulsion of a conventional film camera. Its purpose is therefore to detect, measure and record EM energy. However, in contrast to film emulsions, scanner detectors are available which are sensitive to wavelengths longer than 1 μm, such as the thermal infrared and microwave regions of the EM spectrum.

Airborne scanning systems can be very broadly classified as being either passive, that is, sensing naturally occurring EM energy, or alternatively active, that is, sensing system-generated EM energy.

3.2 Passive scanning systems

Two distinct types of passive scanning system are currently in use. The first type of system is one which is only sensitive to a single spectral band, such as, for example, the thermal infrared and microwave regions of the EM spectrum. In contrast, the second, generally more complex, system is sensitive to several spectral bands. These bands may range from the visible to the microwave regions of the spectrum.

3.2.1 *Thermal infrared scanning systems*

The technique of airborne infrared scanning was originally developed after the Second World War for military purposes. Its use in the civilian field has, however, only developed during the past 10 to 15 years.

3.2.1.1 *General principles.* Thermal infrared scanning systems are concerned primarily with the measurement of earth-emitted energy within the spectral wavelength bands from 3 to 5.5 μm and 8 to 14 μm. Because these energy bands coincide with the wavelengths normally associated with heat emitted from the earth's surface, they are normally termed thermal infrared energy bands. Systems of this type therefore differ from the photographic infrared systems, which were discussed in section 2.3.2.2. Those systems were, in contrast, concerned with the detection and measurement of reflected infrared energy. One significant consequence of this independence from reflected energy is that thermal infrared systems can be operated both during the day and at night if required.

Although the technique is based on the detection of heat and the production of 'heat pictures', the system does not record surface temperatures directly. Rather, the system senses temperature differences between the various components of the scene.

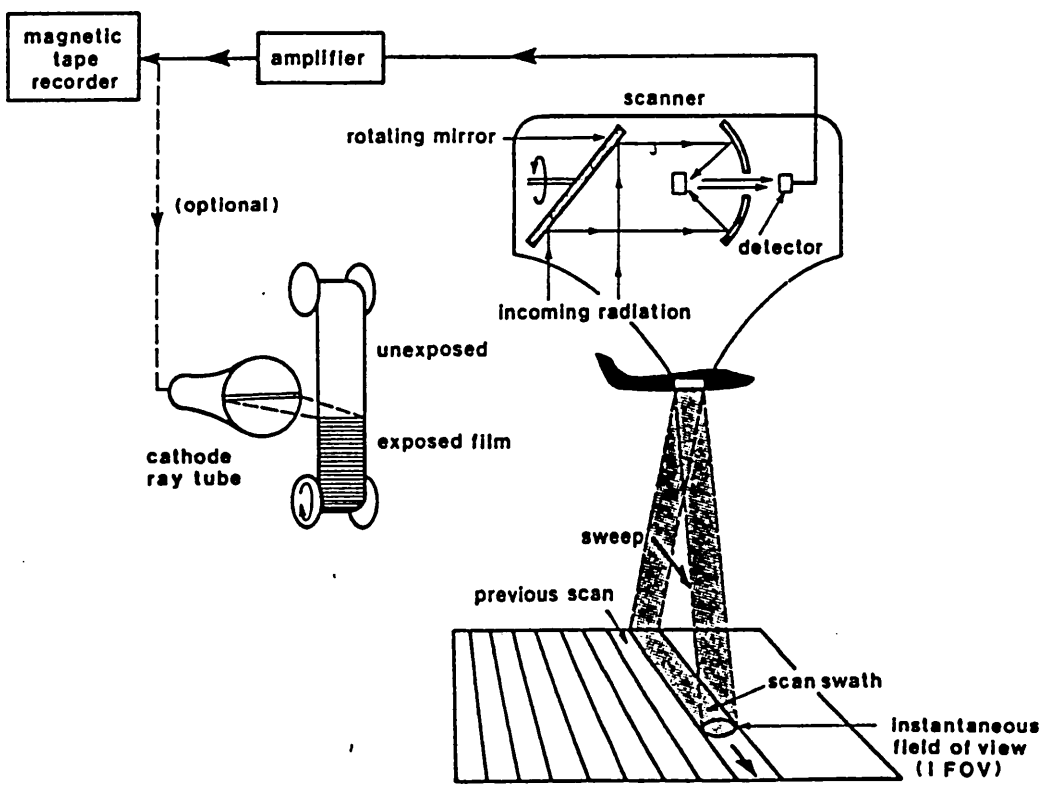


Figure 3.1 Simplified diagram of thermal scanning system (adapted from Rudd, 1974).

The design of a typical thermal infrared scanner is illustrated in Figure 3.1. It can be seen that the thermal scanner is, in very general terms, a telescope which focuses the earth-emitted thermal radiation on to a detector assembly. A selection of typical detectors is listed in Table 3.1. The function of the detectors is to convert the incoming radiation into electronic signals. These electronic signals are then normally recorded directly on to magnetic tape for processing after the flight. Some systems may also incorporate the facility to record a 'quick look' image during flight by means of a cathode ray tube (CRT) and film recorder.

In order to calibrate and quantify the thermal data, two black-body reference temperatures are normally recorded at either end of each scan line. These black-body temperatures are controlled and set by the operator. One is usually set at the ambient ground temperature, whilst the other is set above the highest expected temperature. Thus the data sets for the flight are captured ween two known temperatures. Since the detectors are recording thermal variations close to those within the aircraft, it is necessary to cool the detector assembly in order to reduce the effect of background radiation from the aircraft or scanner itself. Hence the temperature variations which are recorded are those which are due solely to temperature changes associated with the earth's surface. Temperature variations as small as 0.2° C can be resolved using this type of system. For absolute temperature determinations, information about the emissivity of the objects being sensed becomes a critical factor.

After the thermal data has been stored on magnetic tape it is necessary to play back the data to produce imagery from which interpretation can be undertaken. This involves transforming the stored electronic signal into a light beam by use of an optic fibre beam and exposing on to a continuous strip of 125 mm panchromatic black-and-white film which moves across the face of the beam. The intensity of the light beam is proportional to the electronic signal. This is illustrated by Figure 3.2. The scale along the line of flight is corrected by controlling the speed at which the film moves across the optic fibre. The scale oss the flight path is corrected by varying the speed at which the optic fibre crosses the film. In all cases the two speeds are correlated to ensure that x and y axes have approximately the same scale.

On the film negative, ground surfaces having high radiant temperatures give rise to light photographic tones on the image. The remaining areas are

Table 3.1 Characteristics of thermal scanner detectors (adapted from Lillesand and Kiefer, 1979).

Type of detector	Chemical abbreviation	Spectral sensitivity range (μm)
Mercury doped germanium	Ge:Hg	3 to 14
Indium antimonide	InSb	3 to 5
Mercury cadmium telluride	HgCdTe	8 to 14

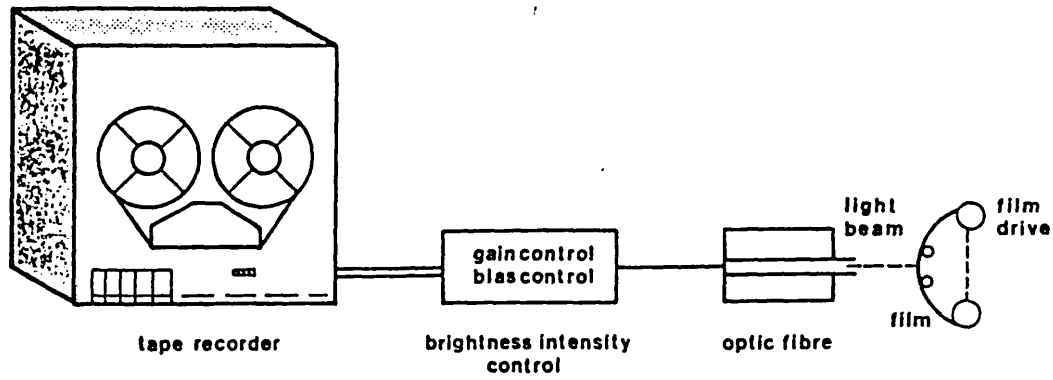


Figure 3.2 Playback of data for reproduction of image.

composed of grey tones which can be interpreted according to the data provided by the black-body reference sources. Figure 3.4 is an example of this type of imagery.

The data recorded in flight, later processed and printed in the form of black-and-white imagery, are essentially a 'qualitative' assessment of the ground temperature. For many applications there is also the requirement for accurate temperature measurement of specified points on the ground—'quantitative' assessment. Quantitative values can be obtained by converting the analogue data recorded on the magnetic tape to a digital format by computer processing. This analogue to digital (A to D) conversion involves transforming the continuous format of the analogue signal to a series of discrete numerical values.

The continuous analogue signal is sampled at a set time interval and the signal level at each sample point is numerically sampled. In computer compatible form the data may be processed, analysed and displayed in a variety of manners. For example, the image may be 'density sliced' (see Chapter 4) and a colour assigned to a range of grey levels. Since each range of

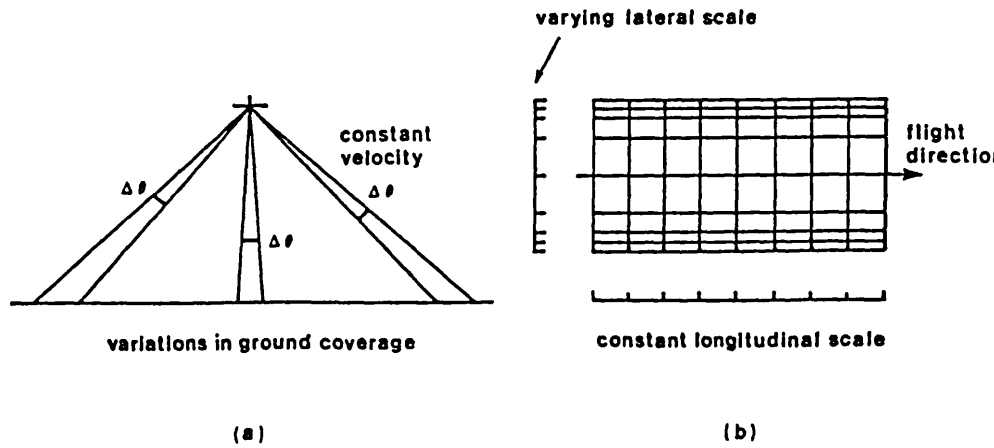


Figure 3.3 Scale variations in thermal scanning imagery. (*a*) Geometry of rotating scanning mirror. (*b*) Resulting scale distortion in unrectified scanner imagery. (After Lillesand and Kiefer, 1979.)

grey levels will have associated with it a particular temperature band, it is possible to produce a colour-coded image. Plate 18 illustrates an example of this technique in which each colour corresponds to a small temperature difference.

3.2.1.2 *Geometric characteristics.* Although a thermal infrared image has the appearance of an aerial photograph, it does, however, differ significantly in geometrical terms.

The first and most obvious characteristic difference is that of scale. Unless it is geometrically rectified, thermal scanner imagery manifests severe scale distortions, primarily in the direction perpendicular to the flight direction. In the main, this scale distortion is caused by the geometry of the scanner's mirror. The geometry of the scanning mirror and the consequent scale variation are illustrated by Figures 3.3 and 3.4. It can be seen from Figures 3.3 and 3.4 (*a*) that the scale along the scan direction is compressed at the edges of the image. This is because the scanning system, which has a nearly uniform angular velocity, images a greater length of ground in a given time interval when it is pointing obliquely down from the aircraft than when it is pointing to the nadir position. Fortunately, this source of scale variation is normally eliminated during the subsequent ground processing of the image, so producing a rectified image (illustrated by Figure 3.4 (*b*)). Smaller-scale variations may be induced by variations in the attitude of the aircraft. Again, however, these scale changes can be controlled, both in flight and also during processing.

The second fundamental geometrical difference between thermal scanner imagery and aerial photography is in terms of relief displacement. In the case of a truly vertical aerial photograph, relief displacement is radial from the centre of the image (from the principal point). In contrast, however, relief displacement on thermal scanner imagery is perpendicular to the centre line of the image. Therefore the displacement only occurs in one dimension, that is along the scan line.

3.2.1.3 *Aircraft-based systems.* Several aircraft-based thermal scanners are currently in use. A selection of the specifications relating to these scanners is given in Table 3.2. Further details regarding the practical operation of airborne thermal scanners are discussed in Chapter 10.

3.2.1.4 *Satellite-based systems.* Many thermal infrared images of low resolution have been acquired by meteorological satellites. The very low resolution severely limits the usefulness of these systems for engineering purposes. Nevertheless, two systems which are worthy of further comment are the National Oceanic and Atmospheric Administration (NOAA) satellites, and the Heat Capacity Mapping Mission (HCMM) satellite.

One example from the National Oceanic and Atmospheric Administration

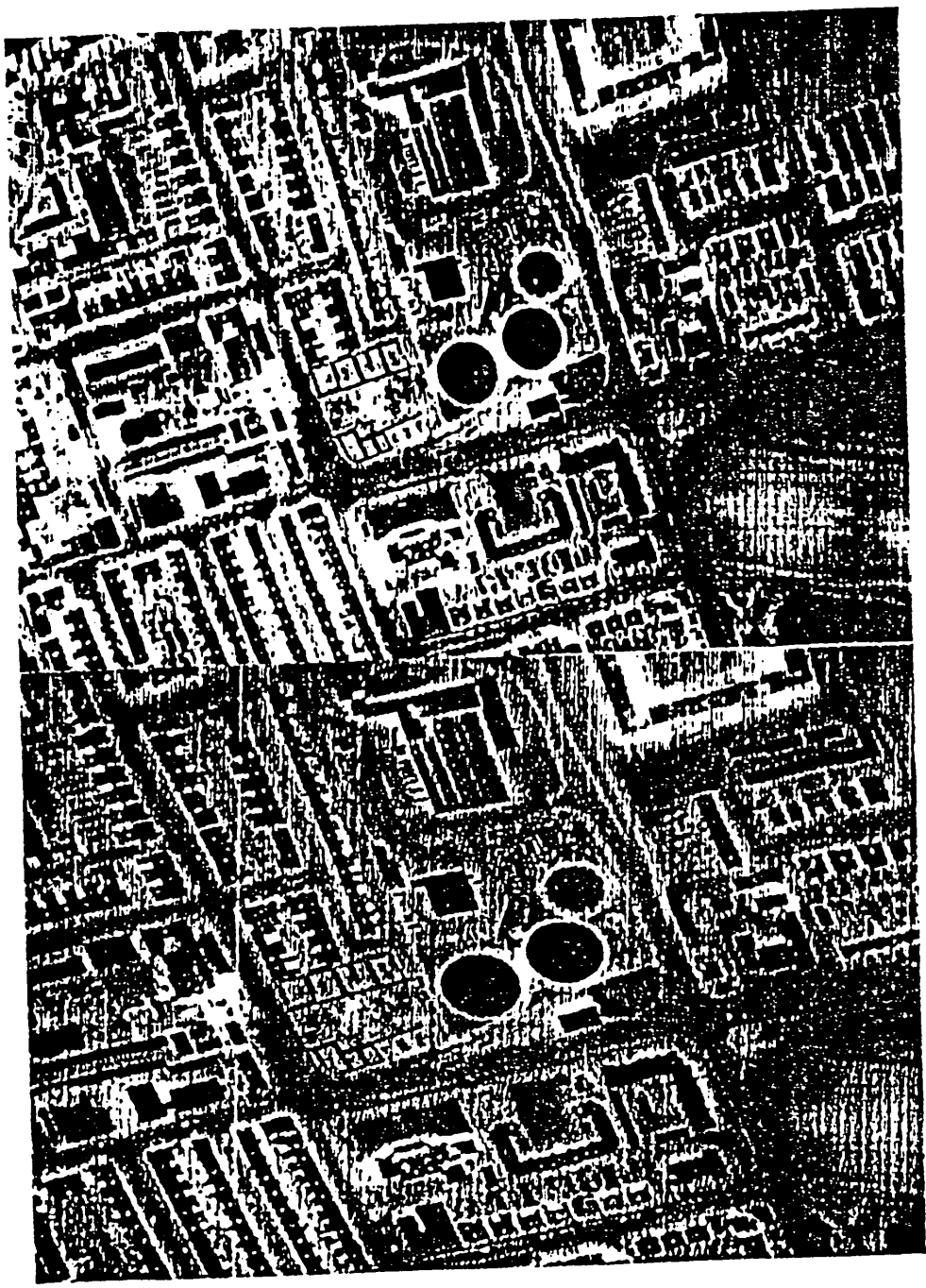


Figure 3.4 Thermal infrared images: (left), unrectified; (right) rectified. (Reproduced by kind permission of Clyde Surveys Ltd.)

Table 3.2 Thermal scanner specifications (after Bullard and Lakin, 1981)

Scanner	Spectral range (μm)	IFOV (mrad)	Resolution (m) at a flying height of 500 m	Temperature sensitivity (°C)	Field of view (°)	Scan rate (oscillations/s)	Detector
EMI							
Airscan	3.5 to 5.5	1.5	0.75	0.3	140	50	InSb
Hawker-Siddeley (British Aerospace)							
201/204/214	8 to 14	1.5	0.75	0.25	120	500	HgCdTe
401	8 to 14				120	500	HgCdTe
Daedalus							
DS-1250	8 to 14	2.5	1.25	0.25	77	80	HgCdTe
DS-1230	8 to 14	1.7	0.85	0.20	77	80	HgCdTe
Bendix							
BTM	3 to 5.5	2.5	1.25	0.50	120	—	InSb
	8 to 14	2.5	1.25	0.50	120	—	Ge:Hg
Texas Instruments							
RS-310	0.3 to 14.0	1.5	0.75	—	90	200	—

series of satellites was NOAA-5. This satellite was operational during the period 1976–1979 and circled in a near polar orbit at an altitude of 1510 km. It carried on board several sensors, including a Very High Resolution Radiometer (VHRR). This instrument was a scanning radiometer capable of producing thermal infrared images within the 10.5 to 12.5 μm wavelength band, with a spatial (ground) resolution of about 900 m.

One interesting application of NOAA-5 imagery is reported by Berg and McGinnis (1980). Berg discusses the use of computer-enhanced VHRR data for river flood mapping in the Kentucky river basin.

The Heat Capacity Mapping Mission satellite (HCMM) was launched in early 1978 and placed in a near circular orbit at an altitude of 620 km. On board it carried a two-channel scanning radiometer, which obtained data both in the visible near IR (0.5 to 1.1 μm) and thermal IR (10.5 to 12.5 μm) portions of the EM spectrum at times of maximum and minimum temperature (1330 and 0230 h). The field of view of the radiometer was rotated through a scan angle of 60° producing images with a swath width of 720 km. Ground resolution in the thermal channel was 600 m. HCMM was essentially built to test the feasibility of measuring thermal variations of the earth's surface features in order to infer their identity and condition. Temperature difference and thermal inertia model data were produced by first geometrically correcting the day and night data and then digitally overlapping the data sets. Thermal inertia data were derived by the application of a mathematical thermal model to the day/night data and visible band data.

These thermal inertia data have been used to discriminate rock types and mineral resource locations (Watson *et al.*, 1982). HCMM data has also been used to evaluate soil moisture and the depth to shallow groundwater (Heilman and Moore, 1982). One further application worthy of mention is that reported by Schowengerdt (1982); in this paper he discusses a digital processing method of combining the higher resolution Landsat band 7 MSS data (see section 3.2.3.3) with HCMM data to produce a thermal map of a lake in Virginia.

The future may possibly bring satellites with improved resolution in the thermal channels. Townsend (1981) rates the proposed Thermal Inertia Mapper, a successor to the HCMM satellite, as 'semi-credible to credible'. If built, the Thermal Inertia Mapper would have 10 m resolution in the 8–13 μm band from a 600 km orbit. It is intended that this satellite would be used to map the thermal inertia of the earth's terrain with sequential passes over the same area at 0400/1000 h and 1600/2200 h local time crossings. A satellite system of this type would have great potential for geological mapping in remote areas of the world where it is prohibitively costly to produce airborne thermal data.

3.2.2 *Microwave scanning radiometers*

A second, and much less common, passive scanning system, is the microwave scanning radiometer.

As with the thermal scanning systems, this type of system senses naturally-occurring radiation. However, in this case it is sensing the emitted EM energy at microwave wavelengths (λ = 1 mm to 1 m). All features emit energy at these wavelengths, although it is extremely faint in comparison to thermal infrared radiation. Also, much less is known about the characteristics of microwave scanner imagery, when compared with thermal imagery. From the literature it appears that the most promising application of passive microwave sensing is for the determination of soil moisture conditions (see Chapter 11). Further details about this application can be found in Estes *et al.* (1977).

3.2.3 *Multispectral scanners (MSS)*

The discussion so far has been concerned with either the photographic systems which sense reflected radiation, or the thermal and microwave scanning systems which detect and measure emitted radiation. An instrument which allows us to detect and record simultaneously both reflected and emitted radiation, in several spectral bands, is the MSS.

3.2.3.1 *General principles.* The technique of data collection used with this type of instrument is very similar to that previously described for thermal scanners (section 3.2.1.1). The system therefore produces imagery in a sequential fashion by recording the radiation (emitted or reflected), along a scan line, perpendicular to the line of flight. As before, the forward motion of the sensor platform advances the strip being scanned and so produces a linescan image.

The operation of a typical MSS can be illustrated by the flow diagram shown in Figure 3.5. Points on the earth's surface are scanned in a raster fashion, normally by means of a continuously spinning prism, or oscillating mirror in the case of a satellite system. This is usually termed an optical mechanical MSS. An alternative approach, the multilinear array or 'push-broom' scanner, is discussed in section 3.2.3.3. Radiation then passes through a converging optical system to focus it.

The radiation is next dispersed into its spectral components by means of a prism or diffraction grating system. This splits the incoming radiation into a series of spectral channels or bands. An array of electronic detectors, placed at

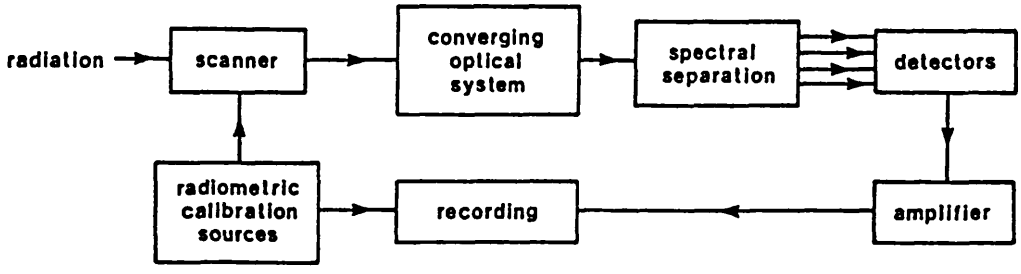


Figure 3.5 Components of a multispectral scanner system (adapted from Swain and Davis, 1978).

the appropriate geometric position behind the grating, detects the radiation within the wavelength region to which it is sensitive. The amplified signal is then recorded on to some suitable medium, commonly a tape recorder, or is telemetered to a ground or satellite receiving station. In addition, most MSS systems also incorporate an internal reference source to radiometrically calibrate the data collected in each channel. This calibration source is therefore scanned by the system and recorded at the beginning and end of each scan line with the radiance data.

3.2.3.2 *Aircraft-based systems.* In recent years a considerable amount of interest has been shown in the use of aircraft-based MSS systems. The early work in this field began in 1965 at the Environmental Institute of Michigan (ERIM). This led to the development of an eleven-channel scanner covering the spectral range from 0.38 to 12.25 μm. Several other organizations also developed scanners, but it was not until relatively recently, particularly in Europe, that airborne MSS data became available on a commercial basis. Two organizations which are currently operating airborne MSS systems in the United Kingdom are Programmed Neuro Cybernetics (PNC) and Hunting Geology and Geophysics Ltd. The specific features of the systems they are presently operating are given in Table 3.3.

The benefits of using an airborne system of the type outlined in Table 3.3 are several. Firstly, the spatial resolution is much higher than that which can be obtained from a satellite system. By flying at low altitude, it is possible to obtain a spatial resolution of the order of 1 to 2 metres. Secondly, in comparison with satellite systems, the user generally has available a greater

Table 3.3 Examples of aircraft-based MSS systems

Detail	Hunting system	PNC systems	
	Daedalus AADS 1268 Airborne Thematic Mapper	Mk II	Mk II(a)
No. of wavebands	11	4	3
Spectral ranges (μm)			
Channel 1	0.42 - 0.45	7.2- 8.9	2.05–2.15
2	0.45 - 0.52	8.8–10.2	2.15–2.25
3	0.52 - 0.60	10.1–10.9	2.28–2.40
4	0.605- 0.625	10.9–13.9	
5	0.63 - 0.69		
6	0.695- 0.75		
7	0.76 - 0.90		
8	0.91 - 1.05		
9	1.55 - 1.75		
10	2.08 - 2.35		
11	8.50 -13.00		
Spatial resolution (mrads)	2.5 (1.25 optional)	1.5	3.3
Swath width (°)	85	120	70

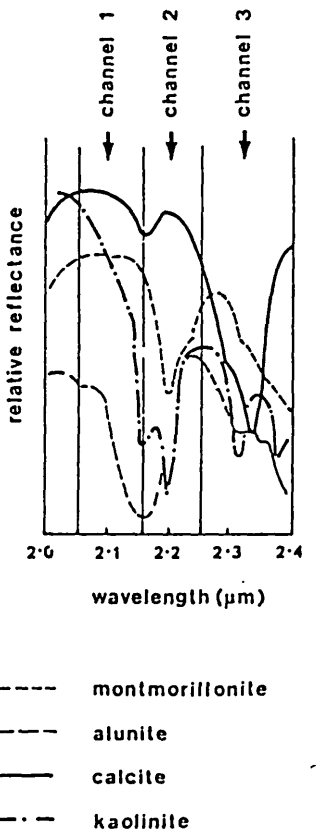


Figure 3.6 Spectral characteristics of clay minerals and carbonates. (After Anon., 1983.)

number of spectral wavebands. More importantly these bands may be chosen in order to aid the discrimination between features of interest. For example, the spectral bands chosen for use in the PNC Mk II(a) scanner enable certain clay minerals and carbonates to be detected by restricting the spectral range of the scanner to the 'clay band' which exists between 2.00 and 2.40 μm. The relationship between the spectral characteristics of certain clay minerals and carbonates and the PNC Mk II(a) wavebands is illustrated by Figure 3.6. Further discussion relating to the use of the PNC scanner system for geological purposes can be found in Barnett *et al.* (1984).

A third benefit of airborne MSS data accrues from the opportunity to use these data either to 'simulate' the appearance and form of imagery obtained directly from satellites, or to simulate the imagery which can be expected from future satellite systems such as SPOT (see section 3.2.3.3 and Figure 3.21). For example, by using bands 2, 4 and 5 of the Daedalus 1268 scanner it is possible to simulate the appearance of a 'true colour composite' which would be obtained using bands 1, 2 and 3 of the Thematic Mapper on board Landsat 5 (see section 3.2.3.3 for further details). Plate 1 illustrates an example of this

type of imagery. There is no doubt that in the future, airborne MSS will become an increasingly important source of terrain information for the civil engineer.

3.2.3.3 *Satellite-based systems.* In contrast to the limited use and availability of aircraft-based MSS imagery, satellite-based MSS imagery is used extensively and is also easily obtainable.

The development of satellite-based MSS systems for commercial rather than military use is a very recent phenomenon. Although a great deal of satellite photography had been taken from space from the early manned space programmes of the 1960s, it was not until the early 1970s that MSS imagery became available. Although some experimental imagery has been obtained from manned satellites, such as Skylab in 1973, the majority of satellite MSS imagery has been obtained from unmanned platforms, in particular from the Landsat series of satellites.

The *Landsat Satellite System*, previously known as the Earth Resources Technology Satellite (ERTS), was initiated in 1967 by NASA in conjunction with the US Department of the Interior. The system was initially designed as an experiment in order to assess the feasibility of collecting earth resource data from unmanned satellites.

To date, five of the six satellites originally planned in 1967 have been launched. Table 3.4 lists a selection of the satellite specifications for Landsats 1 to 5. The table illustrates that Landsats 1 to 3 have many common characteristics which

Table 3.4 Landsat satellite system specifications

	Landsat satellites				
Specification detail	1	2	3	4	5
Orbital details					
Launch date	23/7/72	22/1/75	5/3/78	16/7/82	1/3/84
Ceased to operate	6/1/78	25/2/82	7/9/83	2/83 (T.M. only)	—
Orbital altitude (km)	900	900	900	705	705
Orbital period (mins)	103	103	103	99	99
Equatorial spacing of orbits (km)	2760	2760	2760	2752	2752
No. of orbits/day	14	14	14	14	14
No. of days for repeat coverage	18	18	18	16	16
Equator crossing time	09.42	09.42	09.42	09.45	09.45
Sensor details					
Return Beam Vidicon (RBV)					
No. of spectral bands	3	3	2	—	—
Ground resolution (m)	80	80	30	—	—
Multispectral Scanner (MSS)					
No. of spectral bands	4	4	5	4	4
Ground resolution (m)	79	79	79	79	79
Thematic Mapper (TM)					
No. of spectral bands	—	—	—	7	7
Ground resolution (m) (Bands 1–6/7)	—	—	—	30/120	30/120

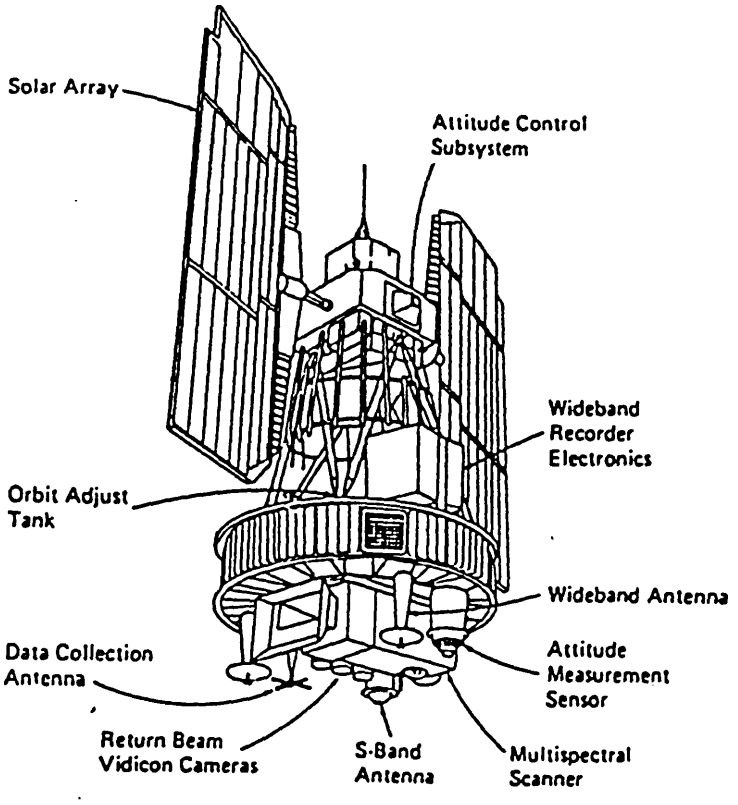


Figure 3.7 Physical characteristics of Landsats 1–3. (After NASA, 1978–83.)

differ significantly from the orbital and sensor characteristics of Landsat 4 and 5. In view of this we will deal specifically with the characteristics of Landsats 3 and 5.

The physical characteristics of Landsat 3 are illustrated by Figure 3.7. The satellite has a weight of about 815 kg and, with the solar panels extended, it has a diameter of about 4m. All Landsat satellites operate in near-polar sun synchronous orbits, which are designed to ensure that the satellite images the same area on the earth's surface at the same local time (about 0940 h). This ensures that the sun illumination conditions are similar for each image and thus repeatable during specific seasons. Repeatable illumination conditions are desirable when producing mosaics from adjacent tracks of imagery and comparing annual changes in land cover. The global orbital pattern for a single day's coverage of Landsat 3 is illustrated by Figure 3.8.

Landsat 3 was launched with two remote sensing systems on board: (a) a five-channel multispectral scanner (MSS) system and (b) a two-channel return beam vidicon (RBV) camera system. The MSS is a line scanning device which uses an oscillating mirror to scan at right angles to the satellite flight direction. This is illustrated in Figure 3.9. This approach contrasts with the spinning scan mirror normally adopted for aircraft MSS systems. The total field of view scanned is approximately 11.56°, with a scanner instantaneous field of view (IFOV) of 0.086 mrad (for the first four channels), or a nominal ground

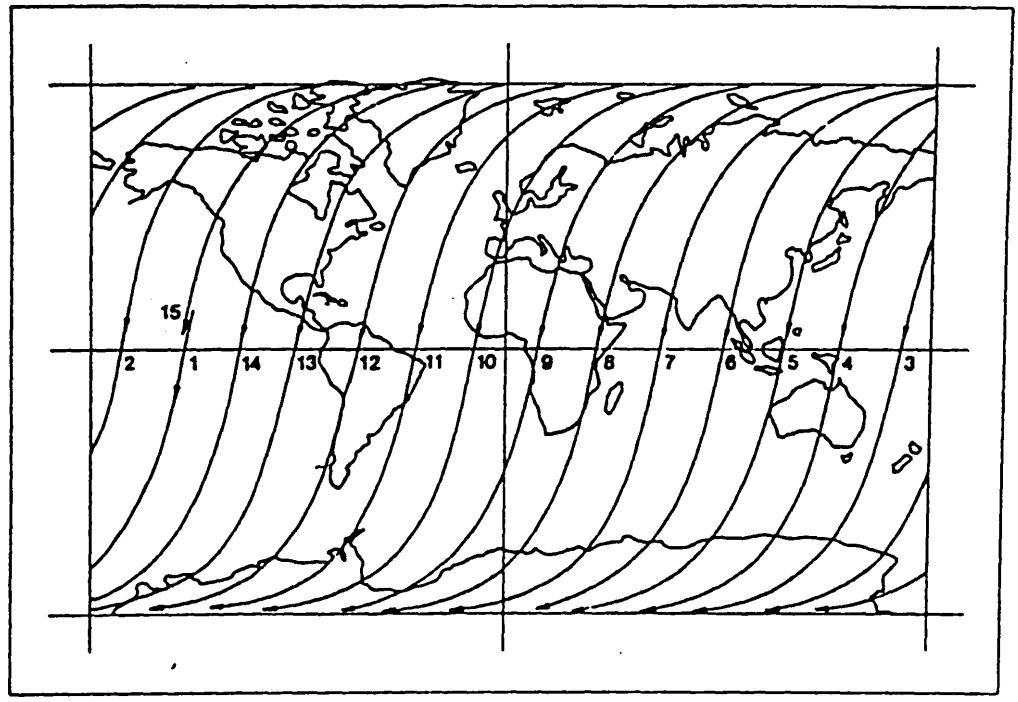


Figure 3.8 Daily orbital pattern of Landsat 3. (After NASA, 1978–83.)

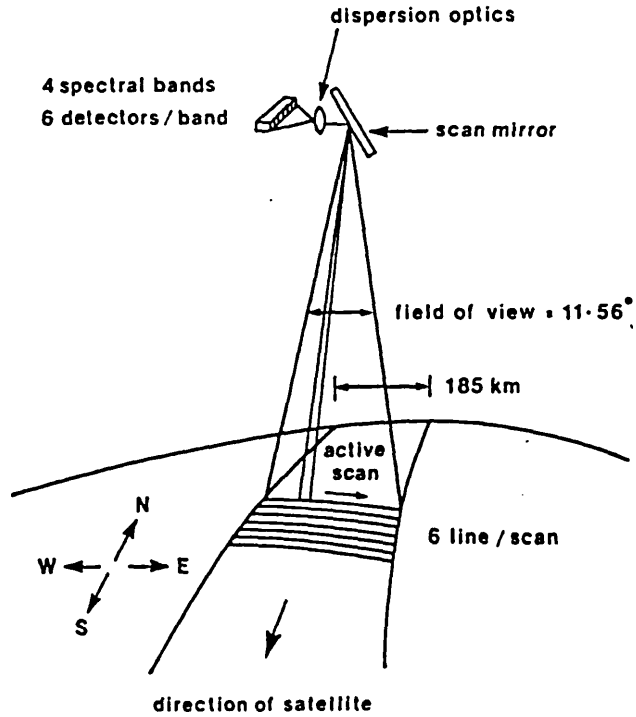


Figure 3.9 Landsat MSS. (After NASA, 1978–83.)

resolution cell of approximately 56 m × 79 m. The fifth channel has an IFOV of 0.258 mrad, or a ground resolution of about 235 m.

The MSS scans each line from west to east with the southward motion of the satellite providing the along track progression of the scan lines. Each Landsat MSS scene covers an area approximately 185 km × 185 km. In view of the extremely high mirror oscillation rate which would be required using this approach, the system is designed to scan six lines simultaneously with each oscillation of the mirror. This results in an area 474 m × 185 km being recorded with each sweep.

The spectral bands used in the Landsat 3 MSS are as follows:

Band number	Spectral range (μm)		Application
4	0.5 to 0.6	Visible (green)	Water penetration –sediment, turbidity studies.
5	0.6 to 0.7	Visible (red)	Vegetation discrimination
6	0.7 to 0.8	Near infrared	Land cover discrimination
7	0.8 to 1.1	Near infrared	Water discrimination
8	10.4 to 12.6	Thermal infrared	Thermal mapping

Unfortunately, band 8, the thermal channel of the MSS, developed operating problems that caused it to fail shortly after launch. The data relating to Landsat 3 MSS are therefore confined to bands 4, 5, 6 and 7. Detection of individual spectral radiances therefore requires four arrays (one for each band) of six detectors (one for each line). The analogue signal from each detector is then converted into digital form by an on-board analogue to digital converter. Each radiance is then associated with a digital number (DN) in the range 0 to 63.

A typical Landsat scene, 185 km × 185 km, consisting of 2340 scan lines with about 3240 picture elements or pixels per line therefore consists of over 7.5 million DN. With four spectral observations per pixel this amounts to over 30 million DN values which have to be recorded for every Landsat scene. Since a new scene is imaged every 25 s, the rate of data generation is enormous (about 15 megabits per second).

A further distinguishing feature of a typical Landsat scene is the parallelogram shape of the final processed image. This is a consequence of the effect of the earth's rotation during the 25 s interval required to image a 185 km × 185 km Landsat scene.

After the scene has been imaged, the data is then either telemetered to one of the twelve ground receiving stations in real time if the receiving station is within range, or recorded on to two on-board tape recorders. The recorded data can then be transmitted to one of the ground receiving stations when the satellite passes within range. Figure 3.10 illustrates the receiving ranges of the Landsat ground stations. European Landsat scenes are recorded at Fucino in Italy and also at Kiruna in Sweden. By means of the Space Informatics

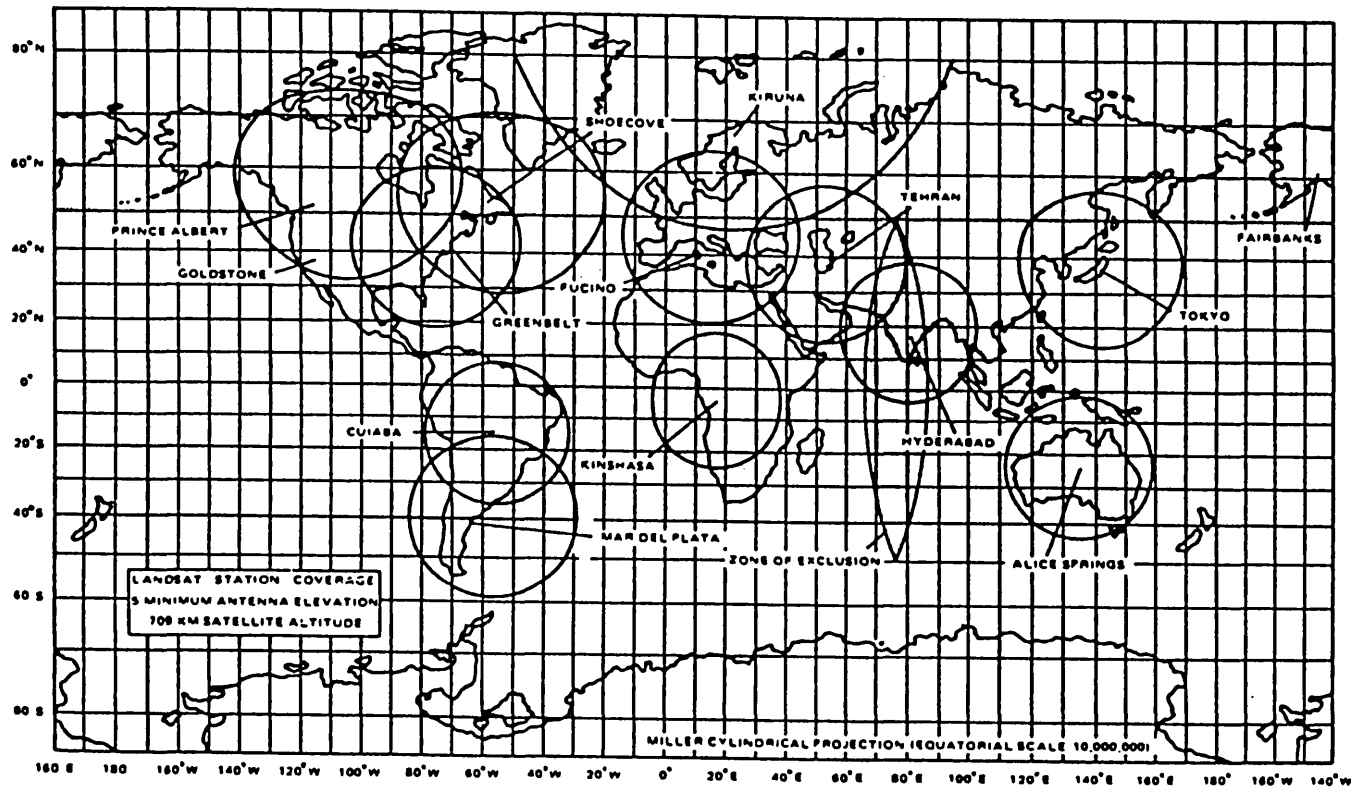


Figure 3.10 Range of Landsat ground stations. (After NASA, 1978–83.)

Network Experiment (SPINE), data can be transferred from Fucino and Kiruna via the European communications satellite OTS to RAE Farnborough within 24 hours of the data being received.

All imagery acquired by Landsat satellites is recorded and archived at the Earth Resources Observation Systems (EROS) Data Center at Sioux Falls, South Dakota. Customers can request imagery direct from EROS in the form of computer compatible magnetic tapes (CCTs) or as photographic products. For United Kingdom, European and a selected number of other areas, RAE Farnborough act as the UK National Point of Contact (NPOC). Further details on the procedure required in order to obtain data may be found in Appendix A.

The digital processing of the CCTs produced by Landsat is beyond the scope of this section and will be discussed in detail in subsequent chapters. Mention should, however, be made of the optical processing of Landsat imagery.

The simplest Landsat MSS photographic product consists of a black-and-white image for each spectral band. Although the resulting image is vaguely familiar, in that it is similar to a conventional panchromatic aerial photograph, much of the information content of the image is lost. An alternative approach is to assign a different colour to each spectral band and superimpose these to produce a colour composite image. Although there are a vast number of combinations which could be used, the one which is the most common is that which produces an image similar in form to the 'false' colour photography discussed in Chapter 2.

In order, therefore, to produce a 'false colour composite' satellite image, the following combination of spectral bands and colours are used:

Band 4	imaged in blue
Band 5	imaged in green
Band 6 or 7	imaged in red

As with false colour photography, this results in healthy vegetation, which reflects strongly in the reflected infrared part of the spectrum, appearing bright red on the image. Figure 3.11 illustrates the appearance of each of the four spectral bands, and its ability to penetrate the water surface for an area of the Red Sea.

This ability to penetrate the water surface may, in conditions where the water is free from suspended sediments and other material, enable the imagery to be used as an aid to bathymetric charting (Benny and Dawson, 1983). For example Plate 2 illustrates the same area as that shown by Figure 3.11, in the form of a false colour composite. By computer processing it is possible to construct lines of equal brightness using the band 4 (green) data of the sea, and colour code these to indicate areas of equal water depth. Using this approach the image shown in Plate 3 can be produced. By interpreting this image it is

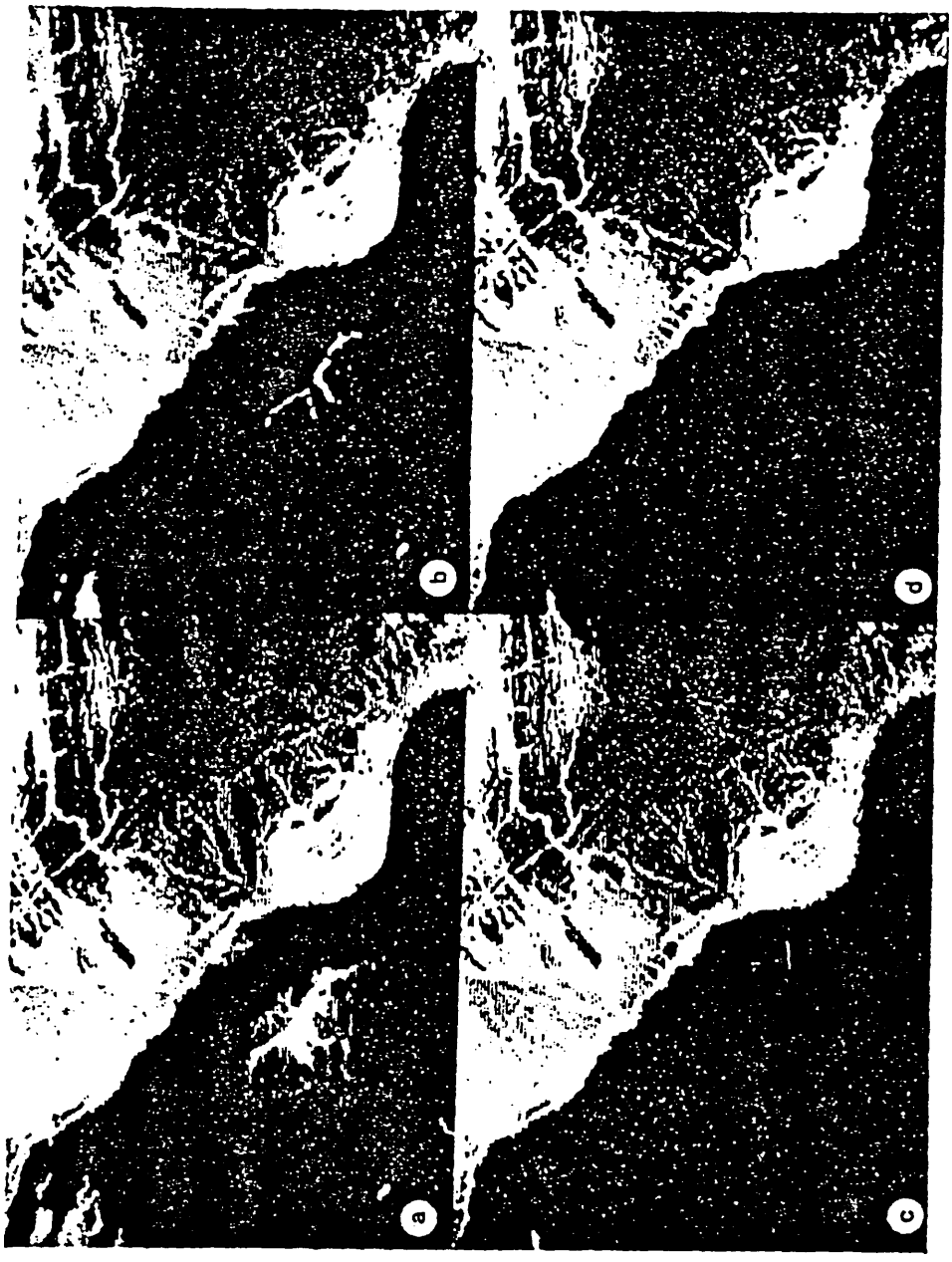


Figure 3.11 Landsat 3 MSS image of the northern region of the Red Sea. Note the transparency of the sea in the green part of the spectrum (*a*) and the strong absorption in the near infrared (*c*), (*d*). (*a*) band 4 (0.5–0.6 μm), visible (green); (*b*) band 5 (0.6–0.7 μm), visible (red); (*c*) band 6 (0.7–0.8 μm), near infrared; (*d*) band 7 (0.8–1.1 μm), near infrared. (Reproduced by kind permission of National Remote Sensing Centre, Royal Aircraft Establishment).

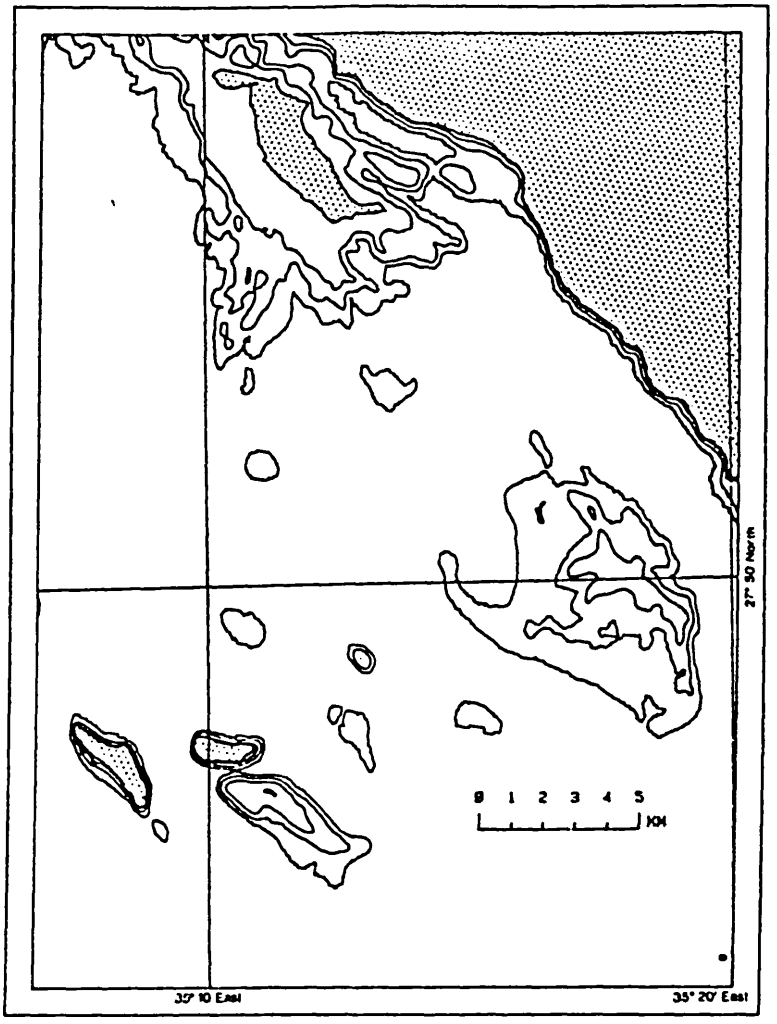


Figure 3.12 Bathymetric chart produced from Landsat 3 MSS image illustrated by Plate 3. (After Benny and Dawson, 1983.)

possible to produce a bathymetric chart of the area. Figure 3.12 illustrates an example of such a chart which extends over part of the area covered by the Landsat scene shown in Plate 3.

The second remote sensing system on board Landsat 3 was the two-channel RBV. The development of the vidicon camera has provided satellites with a method of obtaining high-resolution pictures without the need for conventional photographic film. The principle of operation of the RBV can best be described by first considering the simple direct beam vidicon as illustrated by Figure 3.13.

It can be seen that the direct beam vidicon consists of an evacuated glass envelope containing an electron gun which faces a thin photoconductive insulating layer (the target) which is deposited on the end window of the tube. An electron gun scans across the target surface and removes any positive charge, so stabilizing the target surface at zero voltage. An exposure is made in the conventional manner by means of the shutter and lens assembly. An

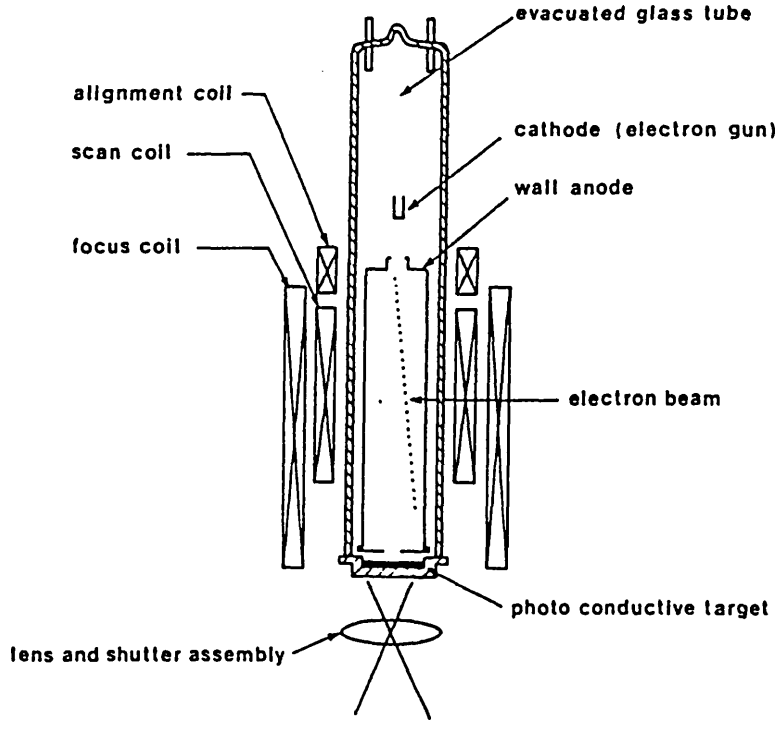


Figure 3.13 Direct beam vidicon tube.

image then forms on the target, the illuminated regions inducing a positive charge in proportion to the brightness of the reflected signal. This charged pattern is retained on the target surface until it is scanned a second time, in this case the magnitude of the pattern being measured and recorded.

The return beam vidicon differs slightly from the description above, in that it relies on the strength of the returning electron beam after it has neutralized the charge on the target as a means of measuring the charge pattern. The advantages of this design are firstly that the camera can operate in poorer illumination conditions, and secondly the resolution is generally higher than that associated with direct beam vidicons.

The RBV cameras on board Landsat 3 covered one spectral band 0.505 to 0.750μm, (green to near infrared) and imaged an area approximately one-quarter that of the MSS. Mounted side by side and operating alternately, the RBVs acquired four subscenes to provide the same ground coverage as the MSS scene, as shown in Figure 3.14. The resolution of the RBV imagery was about double that of the MSS and there was a réseau grid superimposed on the image. For these reasons several investigators have examined the potential of using RBV products for topographic mapping. See Chapter 6 for further details.

Landsat 5, the second of the new generation Landsat satellites, was launched in March, 1984, following the failure early in 1983 of the Thematic Mapper sensor on board Landsat 4. The design of this second-generation series of satellites began in the early 1970s and has led to the development of a

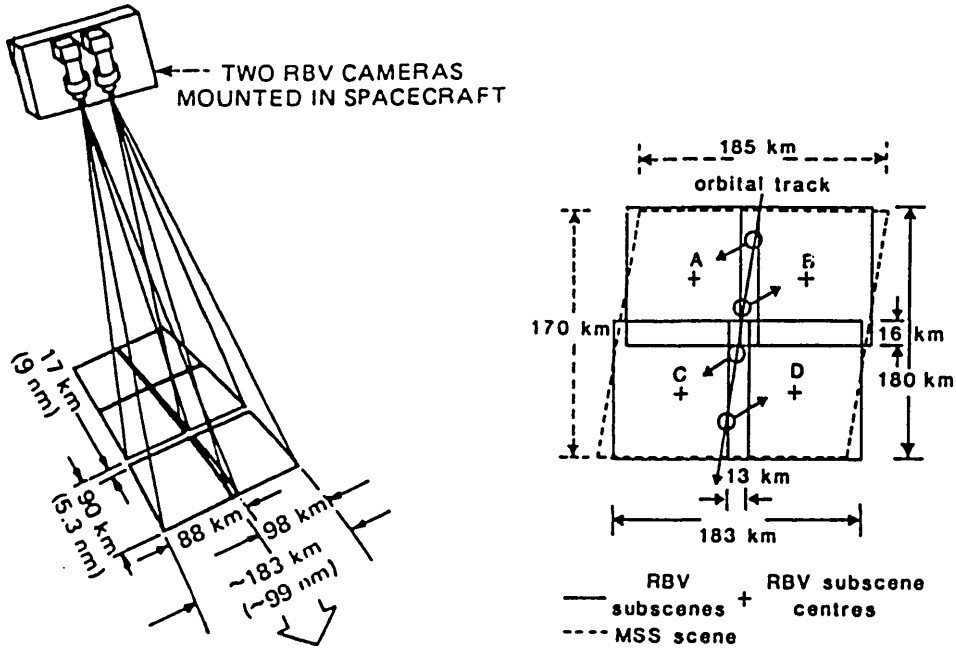


Figure 3.14(*a*) Landsat 3 RBV; (*b*) coverage. (After NASA, 1978-83.)

system which differs considerably from the first-generation Landsat satellites (numbers 1 to 3). By far the most significant change has been the introduction of a new sensor with improved spatial resolution, spectral separation, geometric fidelity and radiometric accuracy known as the Thematic Mapper (TM). In addition, and in common with the previous satellites, Landsat 5 also has a four-channel MSS with identical spatial and spectral characteristics to the previous Landsats' MSS.

In terms of basic design, the TM differs fundamentally from the MSS in five ways. Firstly, the TM obtains data on both the west to east and east to west scans, in contrast to the single-scan method employed on Landsats 1 to 3. This bidirectional approach was introduced in order to further reduce the mirror scan rate. Secondly, the TM detector arrays are located within the primary focal plane of the instrument, allowing incoming light to be reflected directly onto the detectors without transmission through the fibre optics employed in the MSS. By the introduction of this scanner design feature, it was hoped to minimize any loss in the intensity of the incoming radiation. An unfortunate consequence, however, is that the detector arrays for different spectral bands are spaced apart in the focal plane by a small amount; hence the same point on the ground is not simultaneously scanned in all spectral bands. Precise band-to-band registration is therefore an extremely important aspect of the subsequent data processing.

A third distinguishing design feature of the TM is the use of seven spectral bands in contrast to four for the MSS. The band designations, spectral ranges

and principal applications are as follows:

Band number	Spectral range(μm)	Application
1	0.45 to 0.52	Water penetration
2	0.52 to 0.60	Measurement of visible green reflectance
3	0.63 to 0.69	Vegetation discrimination
4	0.76 to 0.90	Delineation of water bodies
5	1.55 to 1.75	Differentiation of snow and cloud
6	10.40 to 12.50	Thermal mapping
7	2.08 to 2.35	Geological mapping

In addition to the increase in the number of spectral bands, the TM also has an increased radiometric sensitivity. The range of discrete radiometric levels which can be sensed had been increased from 64 to 256.

Finally, the spatial resolution of the TM is much improved over the MSS. A pixel size of 30 m is used in all bands except band 7 which has a pixel of 120 m. This contrasts with the 79 m pixel associated with the MSS. The effect of this improved spatial resolution is apparent from Plate 4 which illustrates the comparative spatial resolution capabilities of both Landsat MSS and TM imagery. Several alterations to the orbital parameters of the satellite had to be introduced in order to obtain this much improved ground resolution capability. The major alterations are listed in Table 3.4 and some of the orbital parameters are shown in Figure 3.15.

One consequence of the lower orbit of Landsat 5 has been a significant change in the earth coverage cycle. Figure 3.16 illustrates that for Landsat 5 the orbital strip adjacent to the initial orbital strip is scanned after a time lapse of seven days. This contrasts with the time lapse of a single day for the Landsat 3 system. A further alteration is the reduction in the repeat coverage period from 18 days to 16 days. Both alterations have necessitated the

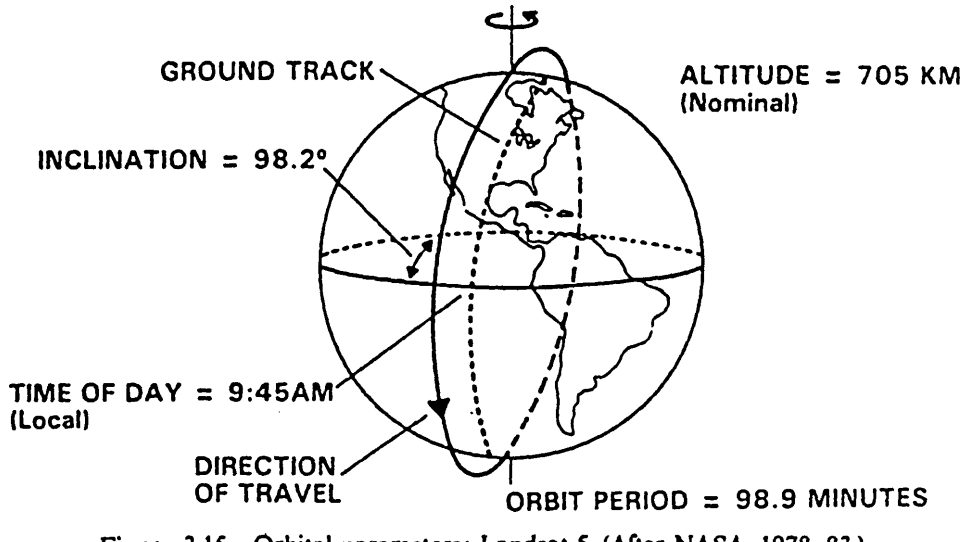


Figure 3.15 Orbital parameters: Landsat 5. (After NASA, 1978–83.)

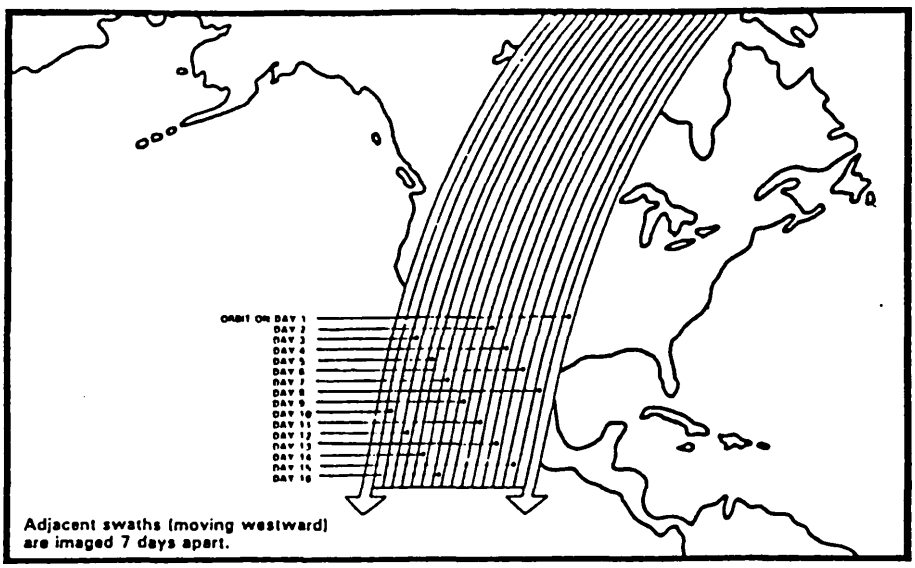


Figure 3.16 Landsat 5 orbital pattern. (After NASA, 1978–83.)

introduction of a new worldwide reference system for defining the path and row parameters of each Landsat TM image.

A further major advantage of the Landsat 5 system will be the satellite's capability to communicate with other complementary satellites. The various components of Landsat 5 and its relationship with the other communications satellites in the system are illustrated by Figures 3.17 and 3.18.

One of the most significant drawbacks from which Landsats 1 to 3 suffered was their requirement to store MSS data upon onboard tape recorders when not within range of a ground station. The recorders not only had a limited lifetime, but they also proved to be rather erratic in operation. In order to circumvent this problem, Landsat 5 will in the near future relay data to earth by communicating with one of the two Tracking and Data Relay Satellites (TDRS), which will then telemeter the data to a single ground receiving station. Most of the existing Landsat receiving stations should be able to receive TM data by mid-1985.

A domestic communications satellite (DOMSAT) system will then be used to transmit sensor data from this ground receiving station to the Goddard Space Flight Center (GSFC) and subsequently to the EROS Data Center. In addition, it is hoped that the Global Positioning System (GPS) navigation satellites will provide Landsat with highly accurate position and velocity data.

Landsat imagery is by far the most common source of satellite-based remote sensing data available to the civil engineer. The previous few pages have discussed some of the technical features of the system and subsequent chapters will discuss several examples of the application of Landsat imagery to civil engineering. It is therefore appropriate at this stage to summarize the principal advantages and limitations of the system.

The main advantages of Landsat imagery can be stated as follows:

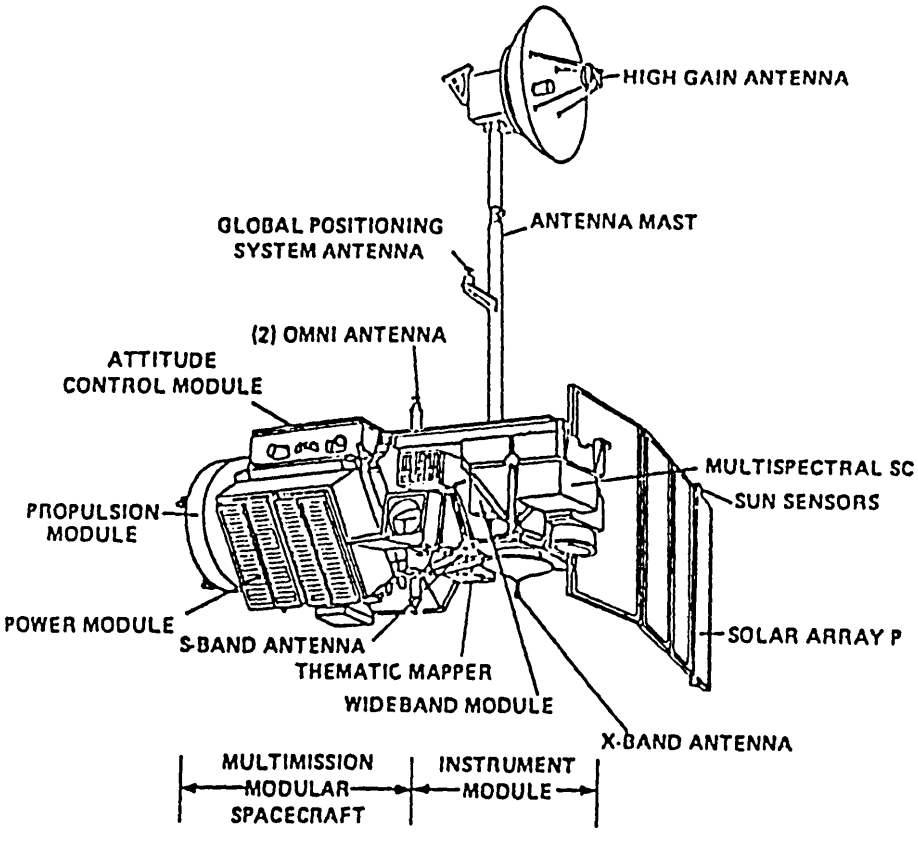


Figure 3.17 General configuration of Landsat 5. (After NASA, 1978–83.)

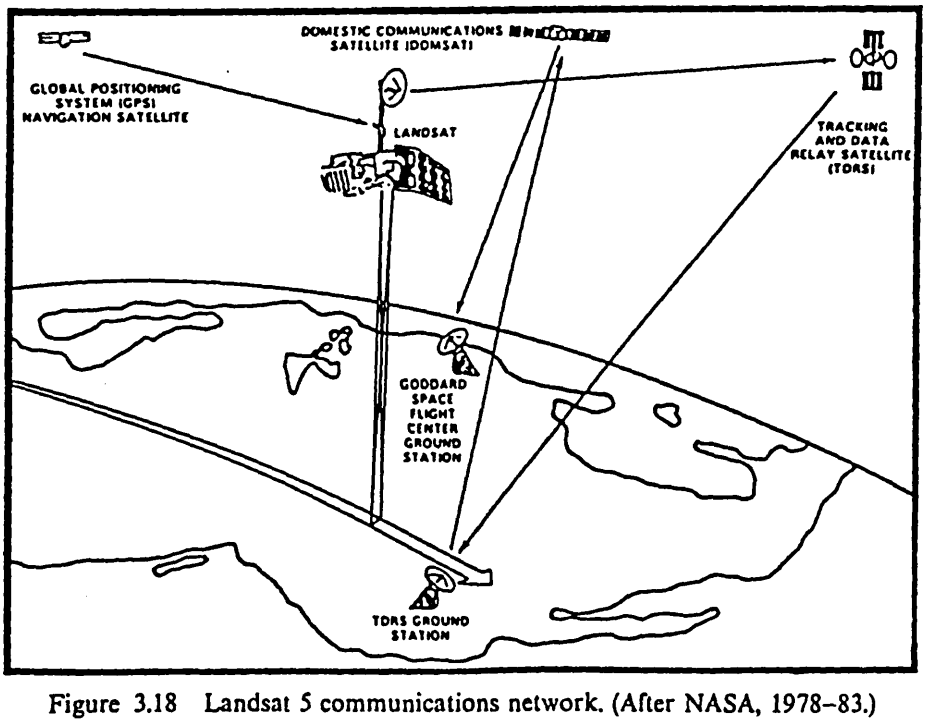


Figure 3.18 Landsat 5 communications network. (After NASA, 1978–83.)

(a) Coverage of large areas. Each Landsat scene, for example, covers an area of some 33000 km^2. Since the first launching in 1972, the Landsat satellites have recorded all the land surface of the globe except those areas persistently covered by cloud.

(b) Repetitive cover. Most Landsat scenes have been recorded more than once, enabling the interpreter to compare views of the terrain taken in different years, or at different seasons. Such comparisons are valuable in regional hydrological studies involving changing river courses, coastal modification or deforestation.

(c) Data are computer compatible. Landsat data can be digitally processed by computer in order to produce images that are easier to interpret, thereby improving the quality of the interpretation.

(d) Availability. Landsat data is easily obtainable either directly from the EROS Data Center in the USA or from any of the growing number of regional remote sensing centres which now exist in various parts of the world (for further details refer to Appendix A).

(e) Low cost. The cost of utilizing Landsat data depends largely on whether the image is examined on photographic film or paper, or on an image processing system. In the latter case the costs are obviously higher since access is required to both the computer-compatible tapes and an image processing system. In both cases however, the costs are relatively small when compared with the benefits which can be obtained from interpretation of the imagery.

In spite of the significant benefits which Landsat offers to the engineer its use in nevertheless restricted. The following are some of the main limitations of the imagery.

(a) Spatial resolution. The present generation of Landsat satellites offer a spatial resolution of ~ 30 m. Whilst this is a significant improvement over the MSS imagery it is nevertheless only suitable for basic planimetric mapping at scales of less than 1:100 000. Map revision at scales of up to 1:50 000 may be possible under limited circumstances. An associated disadvantage is the lack of height information from the imagery, which restricts its usefulness for the civil engineer.

(b) Cloud cover. In spite of the repetitive nature of Landsat there are still several areas of the world, including the UK, for which there is only restricted coverage because of the problems of cloud coverage.

(c) Ground verification. In order to be able to make maximum use of the Landsat system it is important that ground data are available both to improve the interpretation and also to verify the results which are obtained. In some instances problems can arise at this stage because, firstly, few or no data have been obtained, or secondly, the data which have been obtained are out of date.

Landsat 5 is one of many remote sensing satellites which will be operating in the 1980s. Several other systems are planned for launch in the next five years. One of these, which should prove to be of great interest, is the French *SPOT system.*

The SPOT ('Le Système Probatoire d'Observation de la Terre') satellite is due to be launched in January 1985 by the French developed Ariane launch vehicle. The payload will consist of two high resolution visible (HRV) imaging instruments employing the multilinear array or 'pushbroom' design of MSS. This design of MSS differs from the optical-mechanical design, discussed

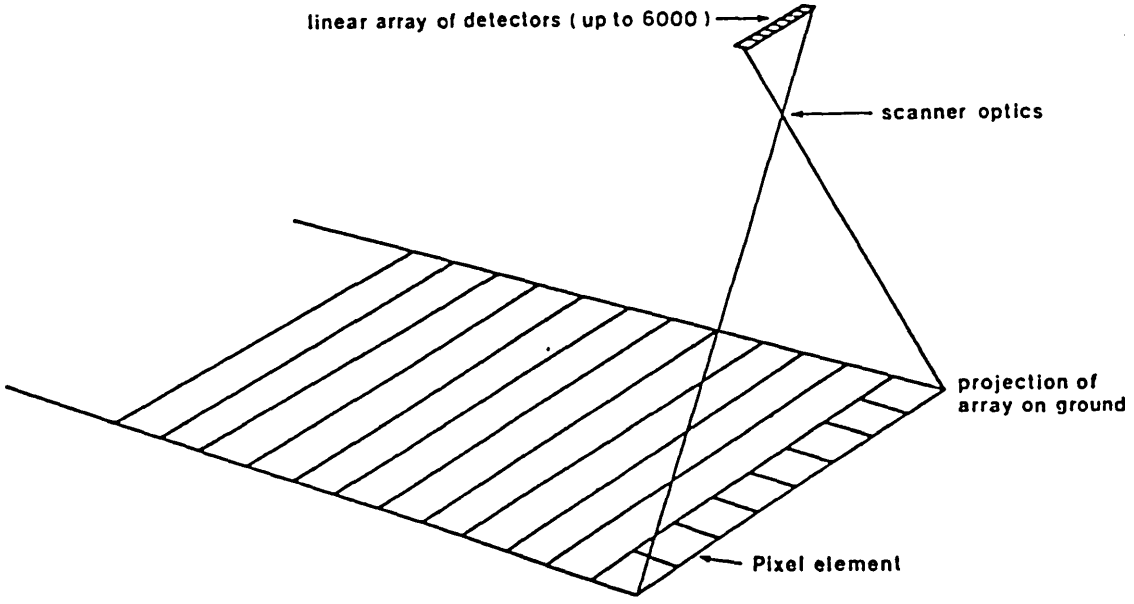


Figure 3.19 Pushbroom scanner.

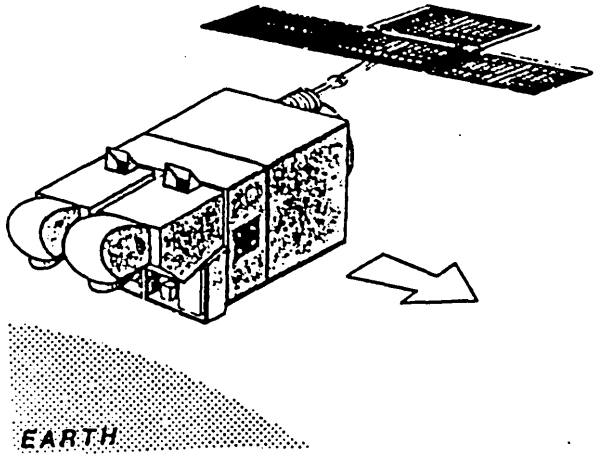


Figure 3.20 SPOT satellite with two HRV pushbroom scanners and attached solar panels

previously, in that it does not incorporate any moving parts such as scanning mirrors. In contrast, each line of the image is formed by measuring the radiances as imaged directly on to a one-dimensional linear array of small detectors located in the instrument's focal plane. Each line is then subsequently scanned electronically and the radiance values recorded onto magnetic tape. As before, successive lines of the image are produced by the forward motion of the satellite along its orbital path. This pushbroom concept is illustrated diagrammatically by Figure 3.19, and Figure 3.20 is an artist's impression of the satellite.

The main advantages of the pushbroom scanner are that it tends to be lighter, use less power, have a longer life expectancy and, most importantly, have a higher geometric fidelity than the equivalent optical-mechanical scanners.

Further significant orbital and sensor details relating to the SPOT system are given in Table 3.5. Figure 3.21 illustrates a simulated SPOT image which was produced using the Daedalus 1268 MSS scanner discussed in section 3.2.3.2.

A further salient feature of the SPOT system will be its capability to provide stereoscopic coverage of the earth's surface. A rotatable mirror in the SPOT sensor package will permit a given scene of the earth to be acquired over areas

Table 3.5 SPOT satellite system specifications

Orbital details	
Launch date	Early 1985
Altitude (km)	832
Period (min)	101
Equatorial spacing of orbits (km)	1084
No. of orbits/day	14.2
No. of days for repeat coverage	26
Equator crossing time	10.30
Swath width (km)	60
Sensor details	
High Resolution Visible (HRV) scanner—	
MSS mode	
No. of spectral bands	3
Spectral ranges (μm)	0.50–0.59 0.61–0.68 0.79–0.89
No. of individual sensor detectors	3000
Ground resolution (m)	20
No. of grey levels	256 (8 bit)
Data rate (megabits/s)	25
Black-and-white mode	
No. of spectral bands	1
Spectral range (μm)	0.51–0.73
No. of individual sensor detectors	6000
Ground resolution (m)	10
No. of grey levels	128 (6 bit)
Data rate (megabits per second)	25



Figure 3.21 Simulated SPOT image of Bakewell, Derbyshire. Image consists of 20 m pixels of band 3 (infrared) data. (Reproduced by kind permission of National Remote Sensing Centre, Royal Aircraft Establishment.)

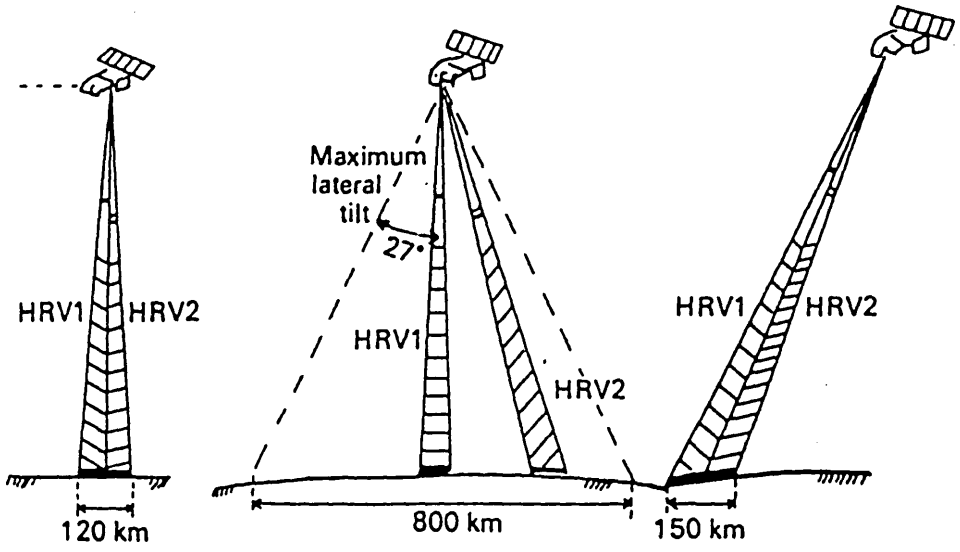


Figure 3.22 SPOT: rotatable mirror. The two HRV scanners can either be operated in a vertical position or tiltable mirrors can be used to point to areas located at some distance from the satellite ground track. (Adapted from Ducher, 1980.)

up to 400km left or right of the normal vertical vantage point of the satellite. Besides providing stereoscopic coverage, this feature will also permit much easier acquisition of high priority scenes. In addition, the two HRVs will be able to be operated jointly covering a swath width of 117km (3km overlap). These features are illustrated by Figure 3.22. It seems likely that SPOT will form the beginning of a new generation of high resolution earth observation satellites which should be of considerable benefit to the civil engineer. Further details relating to the SPOT system can be found in Chevrel *et al.* (1981) and Begni (1982).

MOMS (*Modular Optoelectronic Multispectral Scanner*) is a further example of the pushbroom design of scanner. This instrument, developed by the

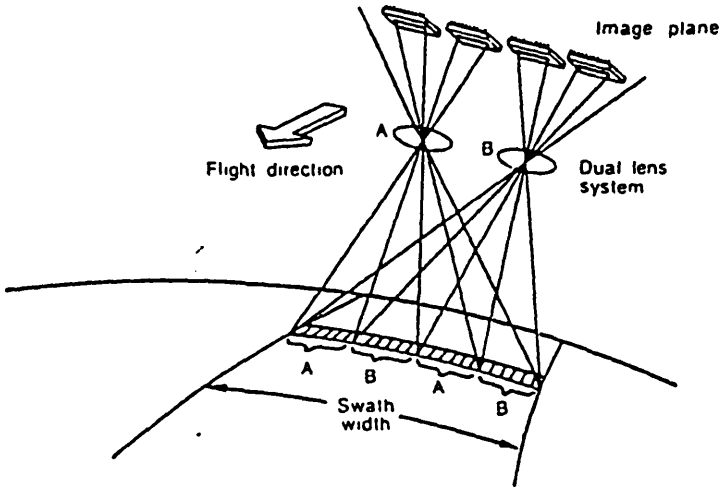


Figure 3.23 MOMS system: geometric arrangement of optical system.

Table 3.6 MOMS 01 system specifications

Orbital details	
Altitude (km) (on Space Shuttle)	300
Swath width (km)	140
Sensor details	
Lens focal length (mm)	237
No. of spectral bands	2
Spectral range (μm)	0.575–0.625 0.825–0.975
No of individual sensor detectors/array	6912
Ground resolution (m)	20
No of grey levels	128
Data rate (megabits s^{-1})	40

German MBB company for the German Aerospace Research establishment (DFVLR), is suitable for use on both aircraft and satellite platforms. The first satellite-borne imagery was obtained with the system during the seventh Space Shuttle flight in June, 1983. The scanner has several unique design features of which the most significant are the dual lens system and the four linear arrays of 1728 pixels each which enable a continuous line of 6912 pixels to be swept out by the scanner (Figure 3.23). Details of the system specification are given in Table 3.6 and further discussion relating to MOMS may be found in Hofmann (1982). Plate 5 illustrates one particular example of the type of imagery obtainable from MOMS.

A second generation MOMS system is currently under development by MBB which will offer the possibility of obtaining stereoscopic imagery and which will also have an extended range of spectral bands. It is expected that MOMS-02 will be flown on board the Space Shuttle in 1986.

Further details relating to the mapping possibilities offered by both SPOT and MOMS are discussed in Chapter 6.

3.3 Active scanning systems

All the systems which have been discussed so far have been of the passive type, that is, they sense and measure the naturally occurring reflected or emitted radiation from a body. There is, however, another group of systems which supply their own source of energy to illuminate features. An image is subsequently produced by recording the reflections of the illumination signal from the feature being examined. For example, a typical active photographic system would be a camera when used in conjunction with some external illumination source, such as a flash unit. Similarly, a good example of an active non-photographic system, which is relevant to remote sensing, is radar.

3.3.1 *Radar*

Radar (Radio Detection And Ranging) was initially developed as a means of using microwaves to detect the presence or absence of an object and then subsequently to derive its range from the transmitter. This process was achieved by transmitting pulses of radiation and recording the reflection which occurred from features within the beam of the signal.

Most of us are familiar with the standard cathode ray tube (CRT) radar display, produced by a rotating antenna, such as is found on the majority of ships. The display is a record of the echoes received from ships or other objects within the area of interest. A photograph of this trace could therefore provide us with a 'radar image' if required.

Although this system could be used on board an aircraft or satellite for imaging purposes, two problems preclude its application. Firstly, it would be rather impractical to operate a rotating antenna on board an airborne platform. This, however, is not the main difficulty. The basic problem is one of resolution. The spatial resolution of a radar system is a function of several factors including the wavelength of the signal, the pulse length and the antenna length. The pulse length relates to the time interval between successive microwave pulses. This parameter can be altered and is therefore not a critical problem. Antenna length is, however, a more difficult problem, because it has a direct relationship with the antenna beamwidth, which in turn is directly related to resolution. Antenna beamwidth can be evaluated by the following expression:

$$\beta = \frac{\lambda}{L}$$

where β is the antenna beamwidth, λ is the wavelength of the transmitted pulses and L is the length of the antenna. Consequently, a longer antenna will have a smaller beamwidth and therefore be of higher resolution. It is this point which precludes the use of rotating antenna systems. For example, to achieve a 10 mrad beamwidth with a 50 mm wavelength radar requires a 5 m antenna and for a 1 mrad beamwidth, a 50 m long antenna! Clearly this is impractical.

The solution to the problem involves reducing the coverage of the system by mounting a fixed antenna to the side of the aircraft or satellite. A system such as this is known as Side Looking Airborne Radar (SLAR), and is illustrated by Figure 3.24.

3.3.1.1 *Side Looking Airborne Radar (SLAR).* SLAR was first developed for military purposes, in the early 1950s. It was not however until the early 1970s that it became available commercially. The main impetus for the development of SLAR arose out of its two significant advantages over optical photographic or scanning systems: its ability to produce imagery both day and night and also its ability to 'sense' through haze, smoke, cloud and even rain.

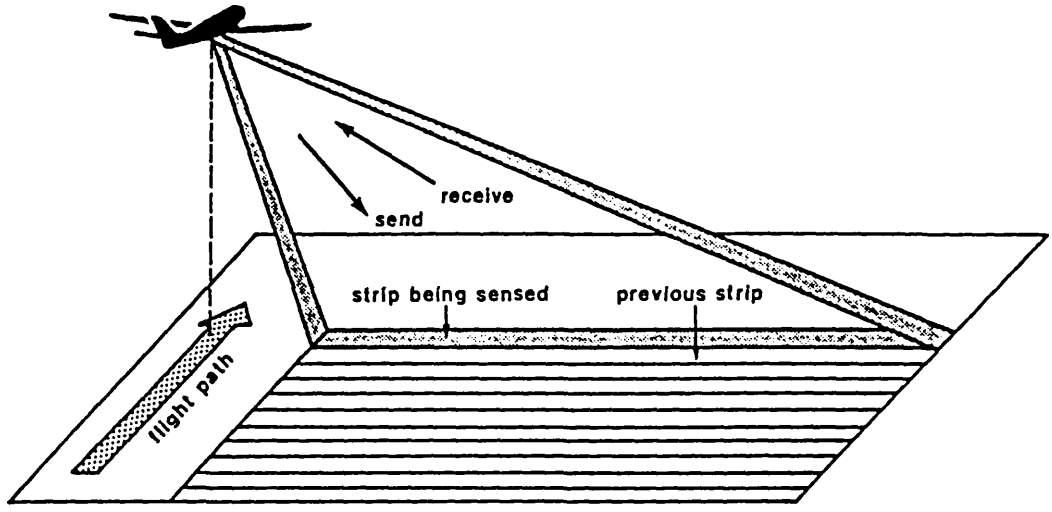


Figure 3.24 Side-looking airborne radar (SLAR). (After Rudd, 1974.)

Table 3.7 SLAR: common wavelength bands

Band designation	Wavelength (mm)
K	8 to 24
X	24 to 38
C	38 to 75
L	150 to 300

The main reason for these twin advantages stems from the SLAR system's use of microwaves as the imaging source. Several different microwave sources are commonly used with SLAR, as shown in Table 3.7, and all are capable of transmission through most atmospheric conditions. The shorter wavelengths tend to be attenuated by atmospheric conditions to a greater extent than the longer wavelengths.

The image which is produced by SLAR is very different from that produced by conventional optical systems. The view which is obtained is a record of the earth's reflective properties at microwave wavelengths. Consequently, the nature and intensity of the reflections by SLAR will be influenced by factors, such as ground conductivity and surface roughness, which are much less significant with optical systems. In addition, the geometrical properties of the image differ significantly from a conventional aerial photograph.

The main components of a typical SLAR system are illustrated in Figure 3.25. It can be seen that the basic principle of operation of a SLAR system involves the measurement of the time interval between the transmission and reflection of a microwave pulse. This indirectly provides the range from the aircraft to the ground feature. The information obtained is subsequently used in many systems to produce an image on a CRT display. The intensity of the returned signal controls the intensity of a bright spot

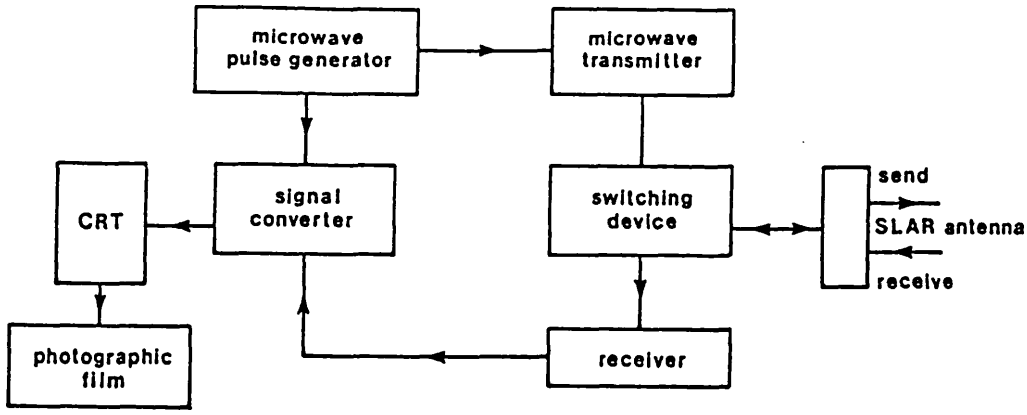
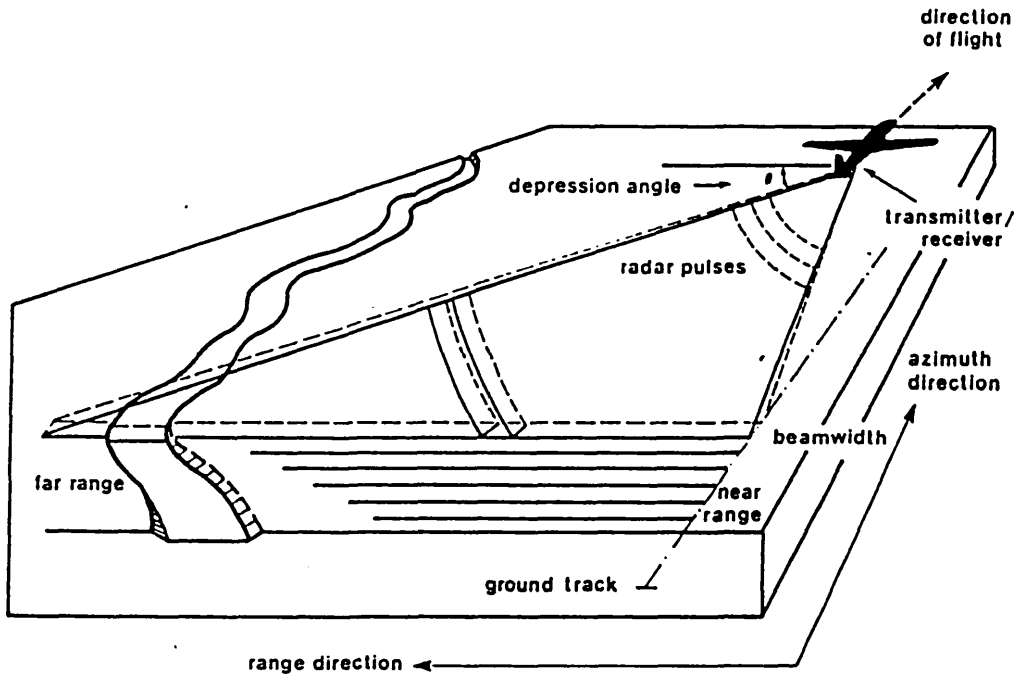


Figure 3.25 Components of a typical SLAR system.



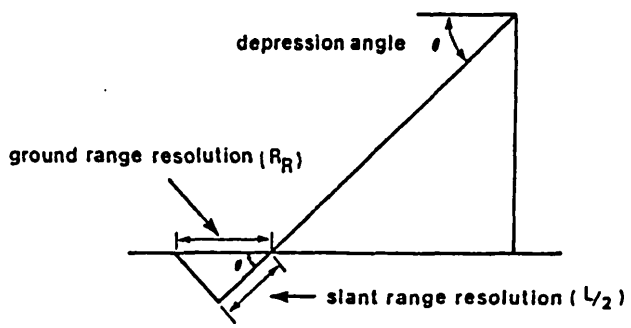


Figure 3.26 Geometry of a SLAR system.

moving across a CRT display at a velocity proportional to the aircraft velocity. Successive positions of the spot therefore correspond to successive ranges across the track. By this process a line of varying intensity is swept across the CRT with each pulse from the microwave transmitter. By synchronizing the velocity of a moving film to that of the aircraft, a continuous photographic image can be created.

A critical consideration in the design of a SLAR system is the question of ground resolution. As was mentioned previously, this is controlled by several factors, including wavelength. However, for a fixed wavelength system, the two most significant parameters are, as mentioned previously, firstly, pulse length and secondly, antenna beamwidth (with its associated parameter antenna length). Figure 3.26 indicates that the pulse length determines the spatial resolution in the direction of microwave propagation, that is, in the range direction.

Ground resolution in the range direction can be determined by the expression:

$$R_R = \frac{tv}{2\cos\theta}$$

where t is the pulse length e.g. $0.1\mu s$,

v is the velocity of a signal e.g. $3 \times 10^8\,ms^{-1}$, and

θ is the depression angle.

Resolution is therefore poorer in the near range than in the far range. In contrast, the antenna beamwidth dictates the resolution in the direction of flight or azimuth direction. The latter is the more critical parameter and has been one of the main limiting factors in the search for high resolution SLAR. However, in recent years new techniques have evolved in SLAR design which have led to the development of two distinct types of SLAR systems, namely Real Aperture Radar and Synthetic Aperture Radar systems.

The SLAR system whose components and geometry has been discussed so far in section 3.3.1 is conventionally referred to as a *real aperture*, 'brute force' or noncoherent radar system. The main distinguishing feature of this design of radar is the resolution limitation of the system in the azimuth direction. Figure 3.27 indicates the deterioration in resolution in the azimuth direction which is a consequence of the fanning out of the signal as the range increases. In this example the two objects close together in the near range position (a) will be discriminated as two individual objects since two separate signals will be reflected. At position (b), however, in the far range, the two objects the same distance apart will not be resolved, since they are both within the beam simultaneously.

This azimuth resolution limitation could be improved by using a longer physical antenna, which would result in a much narrower beamwidth and hence improve resolution in the far range. There is, however, a limitation to the length of antenna which can be mounted on the side of an aircraft (~ 5 m) and

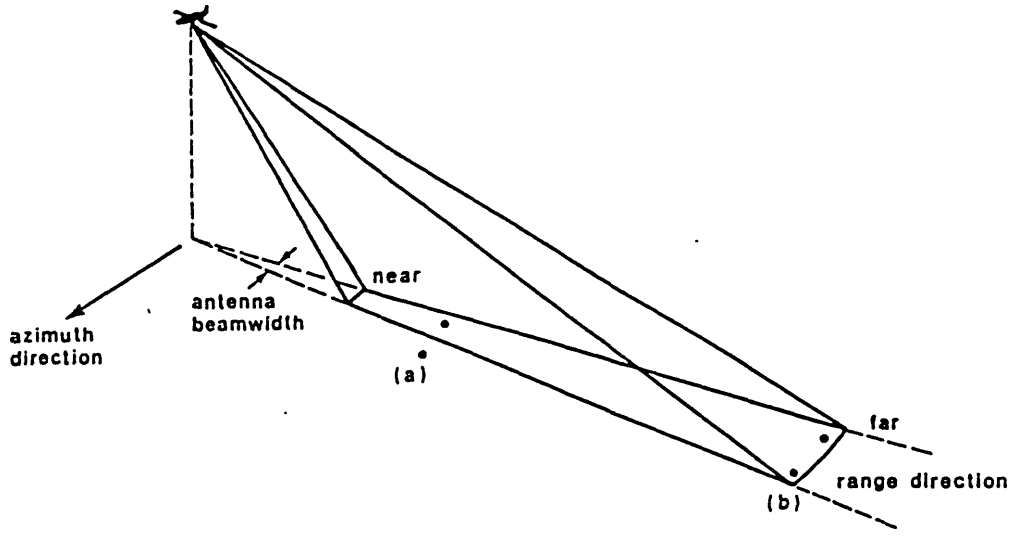


Figure 3.27 Azimuth resolution of a real aperture radar system. (Adapted from Lillesand and Kiefer, 1979.)

it is this problem which severely limits the application of real aperture radar.

An alternative design of radar which overcomes the azimuth resolution limitation of the previous system is the SAR (*synthetic aperture radar*) system. SAR systems are much more complex, both in operation and in subsequent data processing, than the equivalent real aperture radar systems. They are also much more expensive to operate. A detailed analysis of SAR is beyond the scope of this chapter; therefore the discussion will be restricted solely to the fundamental principles of operation of the system. For further details about SAR, see Jensen *et al.* (1977).

The main distinguishing feature of SAR is the much shorter antenna and hence wider beamwidth associated with the system when compared with the equivalent resolution real aperture system. This approach would appear initially to lead to a system of much lower resolution. However, by means of some complex data processing, the antenna is made to behave as if it were a very long antenna. The antenna is therefore electronically synthesized in order to effectively increase the antenna beamwidth and azimuth resolution. The procedure by which this is achieved is as follows.

A series of short microwave pulses are transmitted by the instrument, as with the real aperture systems. Due to the wide beamwidth, as the aircraft or satellite flies over the terrain, features will enter the antenna's beam, move through the beam and finally leave it after a relatively long time-period (depending on the range to the feature). The receiver on board the aircraft or satellite detects the delay between transmission and reception of each microwave pulse and is therefore able to determine the range to a particular feature on the ground. Signals from points on the surface that differ in range can therefore be discriminated on the basis of their arrival time at the antenna. However, for any given position of the aircraft, there will be many points which

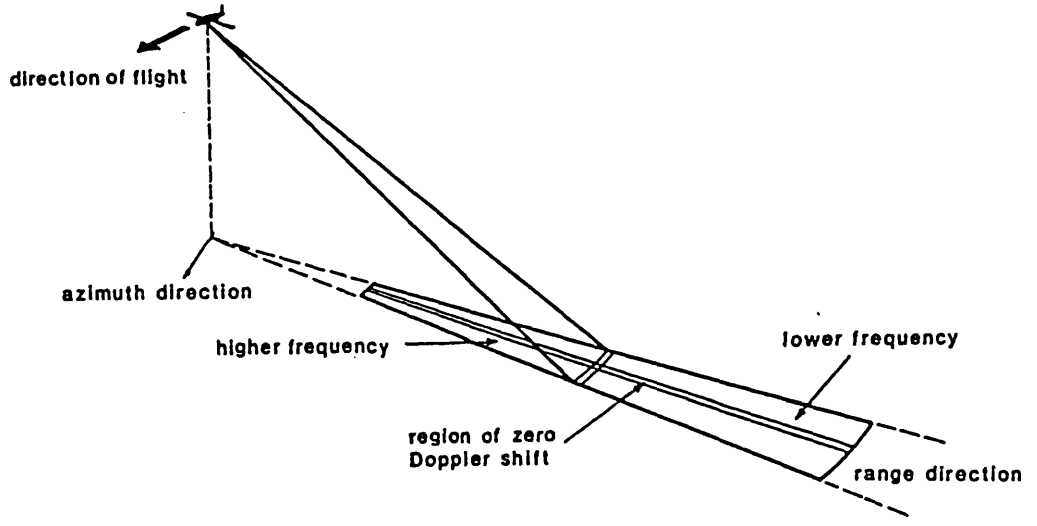


Figure 3.28 Resolution of SAR. (After Lillesand and Kiefer, 1979.)

will be at the same range from the antenna. In order to discriminate different points with the same range, the azimuth to all features is also measured. The azimuth details are determined by means of the Doppler shift in the return signals. The Doppler shift is the change in frequency of the return signal caused by the relative motion of the aircraft and target. For example, the echo from a point perpendicular to the flight path has a Doppler shift of zero, whilst points ahead of the perpendicular are shifted to high frequencies and those behind are shifted to lower frequencies, as shown in Figure 3.28. This Doppler shift information is obtained by comparing the reflected signals with a stable reference signal, generated internally by the system. The comparison is carried out by making the reflected signal interfere with the reference signal, so producing a radar interference pattern. This pattern is the radar equivalent of an optical hologram.

These 'radar holograms' may be recorded and processed, either photographically or digitally. In the case of photographic processing, a coherent light source, such as a laser beam, is normally employed in order to recreate the radar image. Digital processing is also possible. However, at present, it is very expensive and also rather slow. For example (see Elachi, 1982(*b*)), over one billion arithmetic and logic operations are required in order to digitally process the information recorded in one second from the SIR-A radar system (see section 3.3.1.3). Even when a processor such as the CRAY 1 (one of the fastest and most expensive computers currently in operation capable of over 100 million operations per s) is employed, over 500 seconds of processing are still required for each second of data capture.

Finally, the factors affecting the central problem of resolution must be re-examined. Firstly, the resolution of SAR is determined by the system's ability to accurately measure the minute differences in time delay and Doppler shift resulting from adjacent surface elements. Secondly, the azimuth resolution is

further improved by the system's capability of synthetically producing a much longer antenna than the real antenna actually employed. This synthetic antenna is produced because the signals received by the SAR system are recovered over a relatively long time period. During this time period the aircraft motion translates the real antenna over a corresponding distance. It is this distance which in effect becomes the length of the synthetic aperture. By this process a small real antenna is therefore able to act as if it were much longer, with a consequent increase in azimuth resolution.

3.3.1.2 *Aircraft-based systems.* The development of aircraft-based SLAR systems is a very recent phenomenon (Table 3.8). One of the first large-scale projects using SLAR was a survey carried out by the Westinghouse Electric Corporation using a real aperture system developed for the US Army. This survey, of Darien, a cloud-covered province between Panama and South America, produced images which formed a mosaic covering 20 000km^2. A more ambitious project involving the use of SLAR was the production of mosaics covering the entire Brazilian Amazon basin, an area of some 4 000 000km^2. This project was known as Project Radam. Imagery was acquired by a SAR system installed in a Caravelle jet, which was operated by the Goodyear Aerospace Corporation. On the basis of the mosaics, tentative routes were selected for sections of the Trans-Amazon Highway.

A more recent airborne radar project, relevant to the United Kingdom, was the *SAR 580 Experiment.* The European SAR 580 experiment took place during the summer of 1981. A multiband synthetic aperture imaging radar system was operated aboard a Convair 580 aircraft, as part of a research project, jointly sponsored by the European Space Agency (ESA) and the European Joint Research Council (JRC). The main objective of the project was to evaluate the potential role of airborne SAR for the collection of high spatial resolution data over cloudy regions of Europe. A secondary objective, (Haskell, 1981) was to acquire a greater understanding of the design of SAR in order to aid the development of a satellite-based SAR system.

A selection of parameters relating to the system is given in Table 3.9.

Table 3.8 Specifications of aircraft-based SLAR systems.

System	Operator	Type	Band	Resolution (m)	Comments
Westinghouse	—	Real aperture	K	10–20	Ceased commercial operation in 1973
Motorola	MARS	Real aperture	X	20–40	—
Goodyear	Aero Service	Synthetic aperture	X	10	Operated in Caravelle
SAR 580	Intertech	Synthetic aperture	X, C, L	3	Dual polarization
IRIS	CCRS*	Synthetic aperture	C	6	Dual polarization Available 1985

* Canadian Centre for Remote Sensing.

Table 3.9 Characteristics of the SAR 580 system (adapted from Wright, 1982).

	X Band	C band	L band
Wavelength (mm)	32.0	57.0	235.0
Azimuth resolution (m)	2.5	2.5	2.5
Range resolution (m)	2.0	2.0	3.0

A further feature of the system was its ability to transmit and receive radiation which is either horizontally (H) or vertically (V) polarized. Thus, the image resulting when both the transmitted and the received radiation polarization planes are the same (HH or VV) is known as a 'like polarized' image. Conversely, when the polarization planes are different (HV or VH), a 'cross polarized' image is produced. The SAR 580 system was capable of producing both HH and HV images. Figure 3.29 illustrates a typical example of this type of imagery.

Although the SAR 580 experiment took place in 1981, there was a considerable delay in the production of processed images. This, unfortunately,

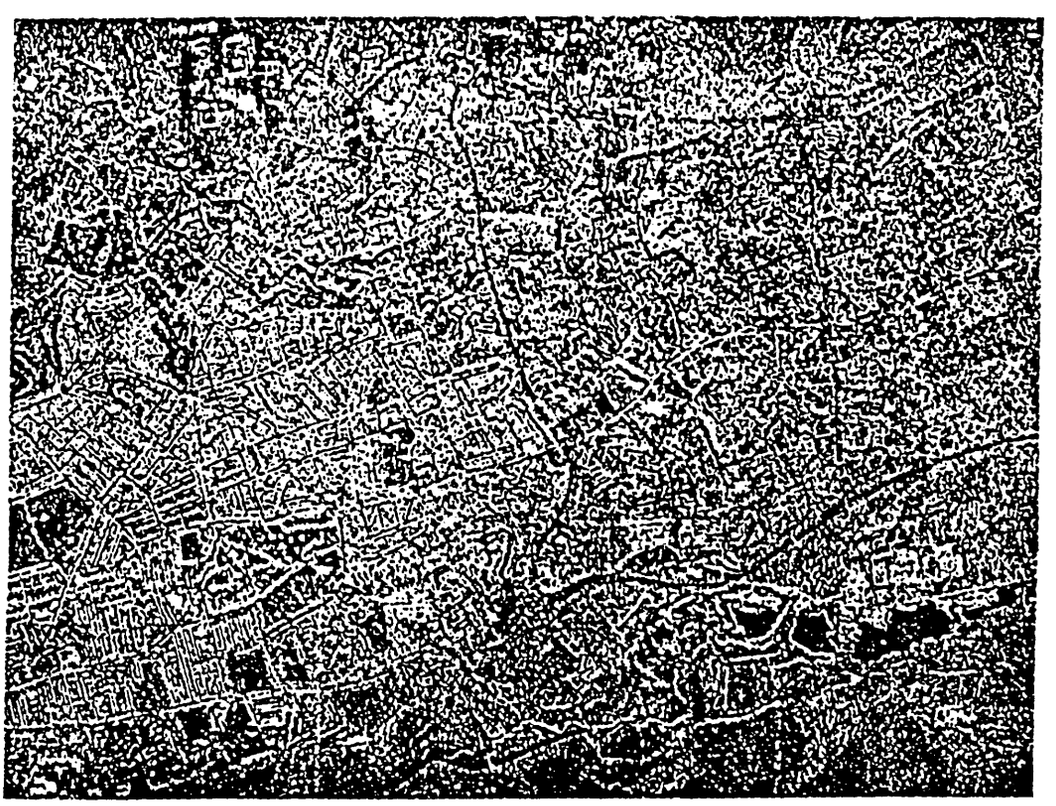


Figure 3.29 SAR 580 image of Northampton, X band, HH polarization. (Reproduced by kind permission of DFVLR.)

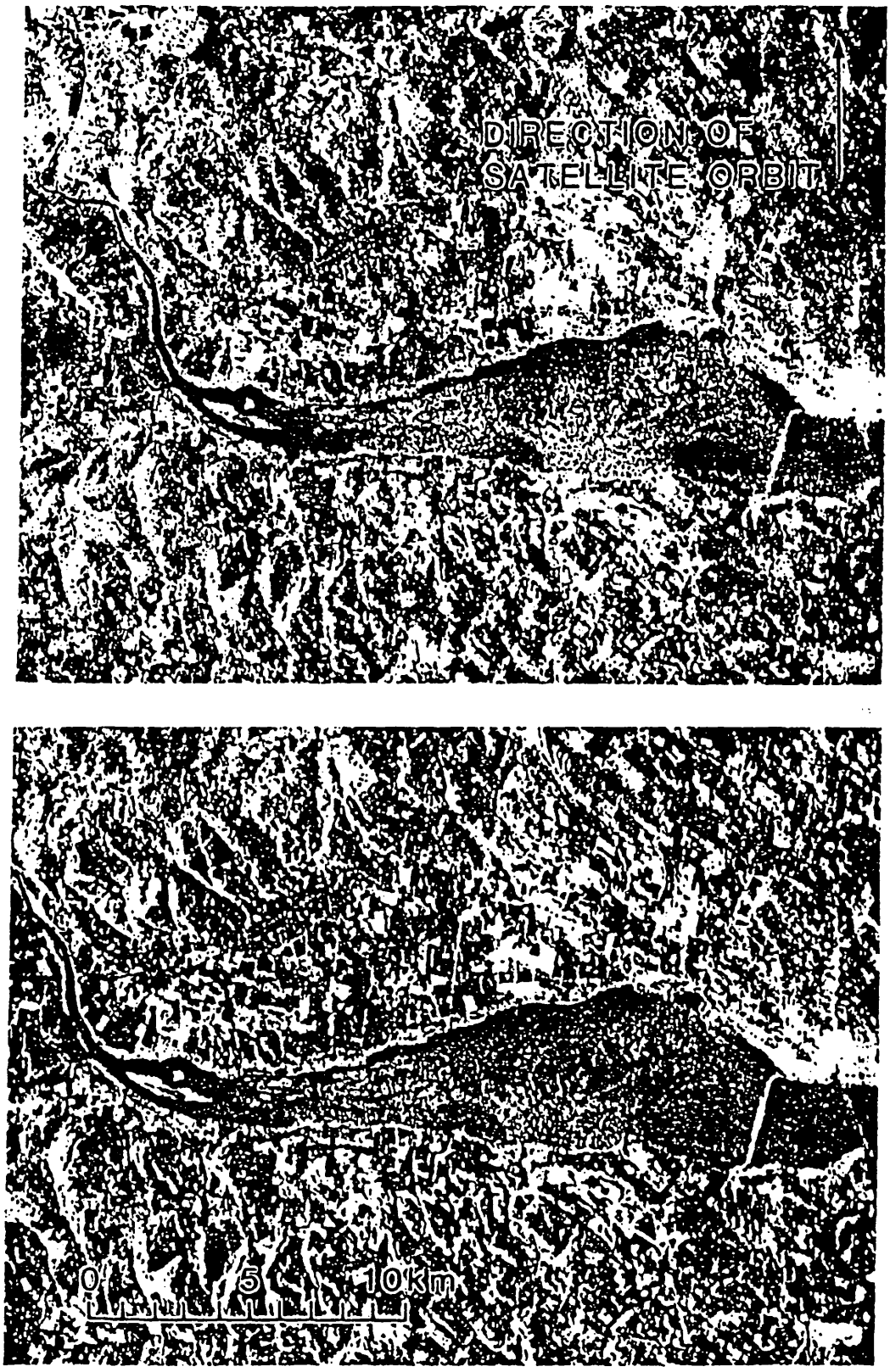


Figure 3.30 Seasat SAR images of River Tay/North Fife, Scotland, obtained in August 1978. (Scale approx. 1:250 000.) (Top) Optically processed image. (Reproduced by kind permission of NASA.) (Bottom) Digitally processed image. (Reproduced by kind permission of DFVLR.)

has led to a significant delay in the reporting of the conclusions of the experiment.

Further details relating to the coverage and availability of SAR 580 data can be obtained from the National Remote Sensing Centre at the Royal Aircraft Establishment (RAE).

3.3.1.3 *Satellite-based systems.* To date there have been two SLAR satellite programmes. Both were of the SAR type and were primarily designed for research purposes. During their operational lifetimes, a substantial amount of imagery was collected and this is now available for interpretive analysis.

The first of these systems to be launched was *Seasat.* Seasat was launched in June 1978 and was chiefly oriented towards oceanographic and geological research. The majority of the sensors on board Seasat were radar-based. Of these, the most important was the SAR system. The SAR operated in the L band ($\lambda = 230$ mm) and orbited at an altitude of 800 km. Each image path consisted of a 100 km wide band located between 230 km and 330 km to one side of the satellite groundtrack. The paths extended for over 400 km and had a resolution of 25 m. Figure 3.30 illustrates two examples of Seasat imagery obtained over Dundee, Scotland. Unfortunately, the Seasat system had an operational life of only 99 days, failing on 3rd October 1978. Nevertheless, during its lifespan, it acquired a significant amount of imagery, which gave a unique view of the earth's surface. Further details on the application of Seasat may be found in Elachi (1980), Foster and Hall (1981) and Allan (1983).

The *Shuttle Imaging Radar-A* (*SIR-A*), launched aboard the second flight of NASA's space shuttle Columbia on 12th November 1981, was an experiment which again had as its primary objective the investigation of spaceborne imaging radars for geological mapping. In common with Seasat, SIR-A employed a SAR, which operated in the L band ($\lambda = 230$ mm). The antenna, 9.4 m long, was fixed to produce a 50° incidence angle at the centre of the swath. This 50° incidence angle alleviated one of the most serious problems encountered in the analysis of Seasat SAR data in areas of high relief. In this type of terrain Seasat suffered from extreme foreshortening and layover, which was caused by its low incidence angle (24° at centre). The larger incidence angle associated with SIR-A therefore reduced this significantly, resulting in images which were more geometrically correct.

SIR-A operated from an altitude of 250 km over selected parts of the earth's surface from latitude 41° N to 36°S. Figure 3.31 indicates the areas of the world for which coverage is available. Data were recorded optically on to signal film carried in a cassette on board the Shuttle. The imagery was subsequently optically processed on to film with a swath width of 50 km and a resolution of 40 m. The scale of the imagery was 1:500 000. Figure 3.32 illustrates as example of this type of imagery. For further details see Elachi (1982) (*a*), (*b*) and Holmes (1983).

A successor to SIR-A, SIR-B, will be flown in the near future. The SAR

Numbers are Data Take References. Swaths are 50 km across.

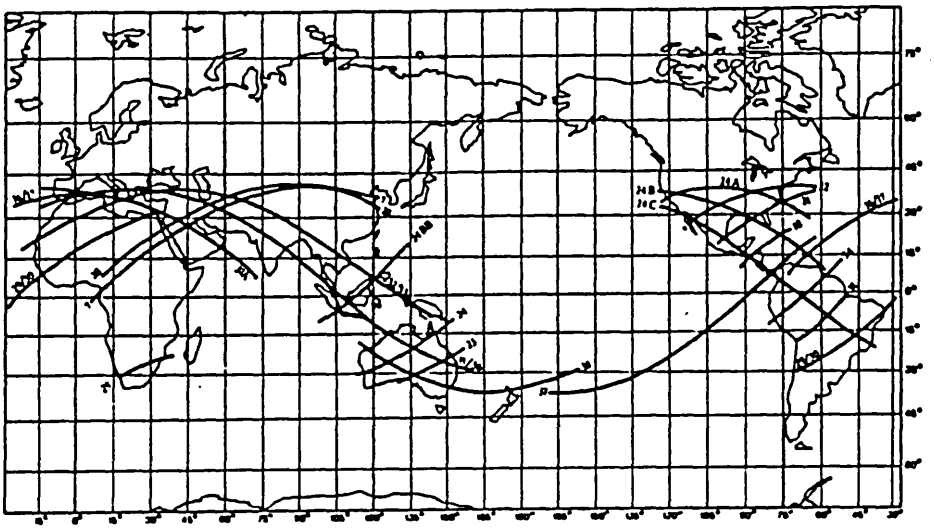


Figure 3.31 Coverage of Shuttle Imaging Radar-A (SIR-A). (After NASA.)

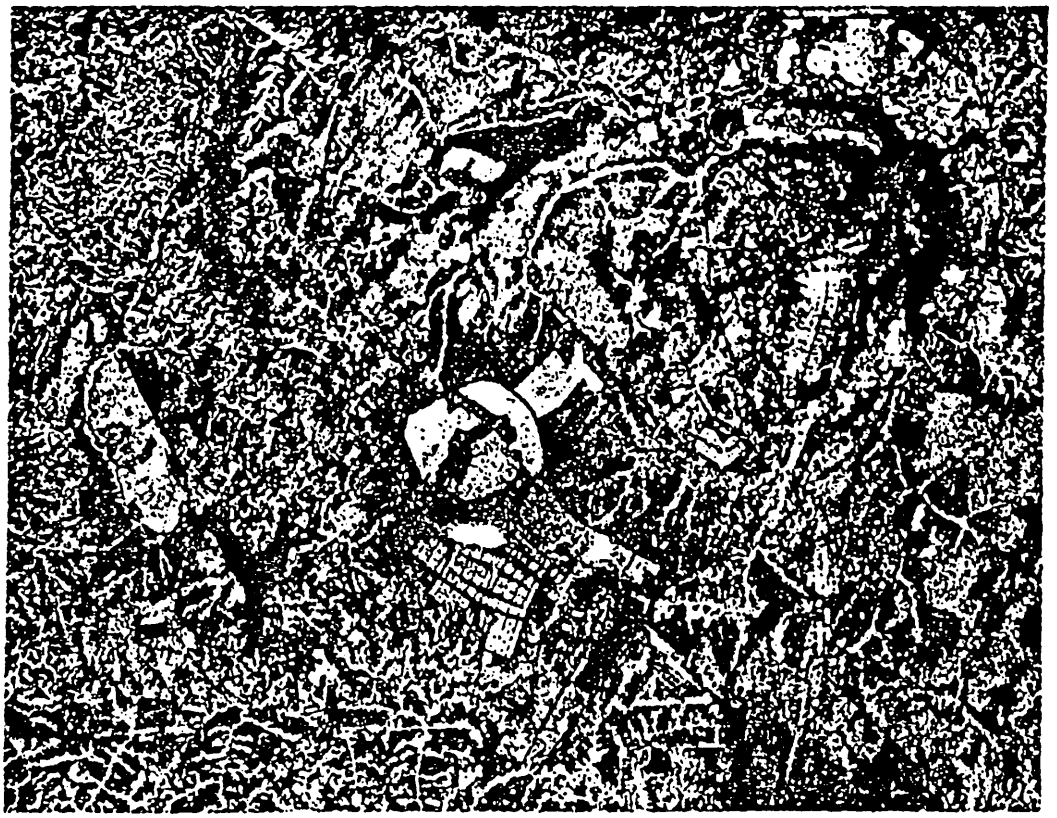


Figure 3.32 Shuttle Imaging Radar-A (SIR-A) image of Brasilia. (By kind permission of Hunting Geology and Geophysics Ltd.)

used will be a modified version of SIR-A. The principal modification is the incorporation of the facility to mechanically rotate the radar antenna in order to obtain imagery at different incidence angles, typically between 15° and 60° as measured from the gravitationally defined vertical. Digital rather than optical techniques will be used in order to produce the imagery.

Two future SAR satellite systems currently under development which are worthy of mention are the *ERS-1 system* and the *Radarsat project.*

ERS-1 (ESA Resource Satellite-1) is the first of a series of European remote sensing satellites which will come into operation in the late 1980s, and operate throughout the 1990s. Planned for launch in 1987 and operating in a 750 km near polar orbit, ERS-1 will carry on board several sensors, of which the most significant from a civil engineering point of view will be the Active Microwave Instrumentation (AMI).

The AMI, operating in the C band, will combine the functions of a SAR imaging system, a wave scatterometer and a wind scatterometer. The principal aim of this package of sensors is to provide high-resolution images of coastal zones, ice areas and land areas together with data about the surface condition of the sea. Two areas where the data could be of use are the fields of coastal and offshore engineering. In the former case the SAR imagery could be used to complement the optical data provided by other satellites such as Landsat 5 and SPOT, whilst in the latter case the continuous monitoring of sea states could provide useful statistical data on wave and wind conditions for the engineering design of oil platforms and other offshore structures.

The second SAR satellite system under consideration is Radarsat. Radarsat is a Canadian system which is planned for launch in 1990. The primary aim of the project is the development of a system to provide information about ice conditions for selected Arctic and coastal regions of Canada.

The overall concept of Radarsat is illustrated by Figure 3.33. The satellite is planned to operate in a near-polar orbit at an altitude of approximately 100 km. It is proposed that SAR data over a 150 km swath will be transmitted to a ground receiving station where the data will be converted digitally into imagery. The image data will then be relayed via a communications satellite to an ice information centre, where other data such as aircraft SAR and weather satellite data will be included to provide an ice forecast. This forecast will then be relayed direct to users such as offshore drilling rigs, ships and so on within the area of interest.

3.3.2 *Lidar*

Lidar (Light Detection and Ranging) is a further example of an active remote sensing system. Whilst much less common than radar the fundamental principle of operation is very similar and involves the determination of range by measuring the time interval between the transmission and reception of a pulse of EM energy. In this case, however, the source of energy is a laser, rather

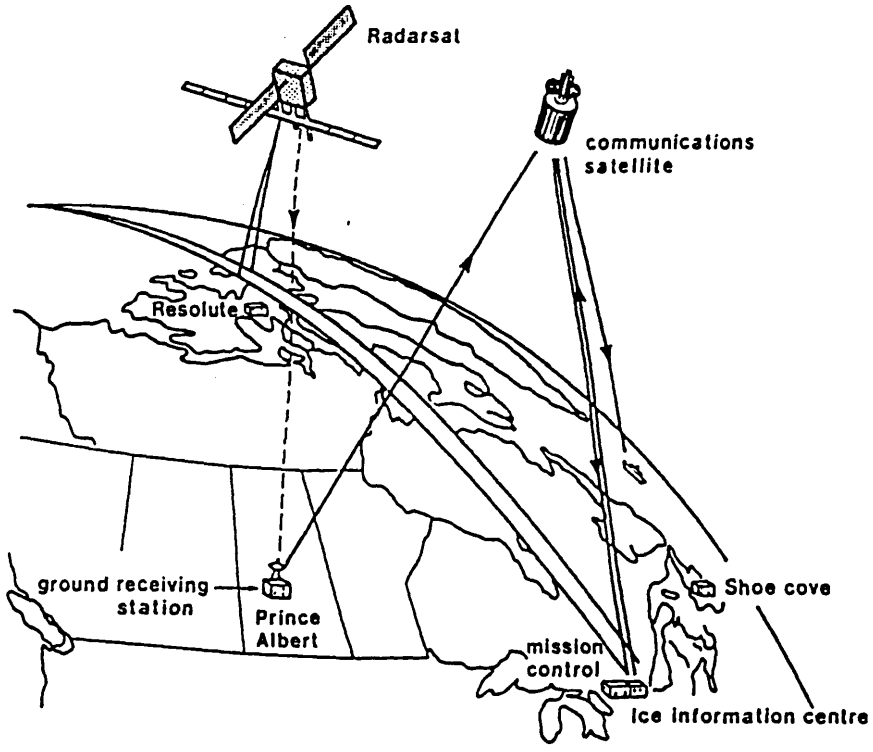


Figure 3.33 Radarsat (adapted from Raney, 1983).

than a pulse of microwaves. Lidar systems can be broadly classified as either those which operate in a profiling mode, or those which operate in a scanning manner. In both cases, however, the primary aim is to measure range rather than the production of imagery. The components and functional arrangement of a typical lidar profiling system are illustrated by Figure 3.34.

Accurate profiling of water depth is accomplished by transmitting a pulsed coherent laser light source from the sensor platform to strike the water at right angles. A strong reflected light signal is received from the water surface, followed closely by a weaker reflected signal from the floor of the water body. The calculation of water depth is carried out by determining the time of travel of a given laser pulse for the round trip from aircraft to water bed to aircraft minus the return round trip travel time from aircraft to water surface. Further details relating to the practical application of this technique to bathymetry are discussed in Chapter 11.

A recent development of the lidar technique is the scanning lidar. One example of this approach is the Larsen-500 system currently under development by the CCRS for use by the Canadian Hydrographic Service. The main components of the system are illustrated by Figure 3.35.

As with the profiling system range data are continuously being acquired by the system. In this case however the scanning facility enables large areas to be covered much more efficiently. The final processing is carried out on the ground at which time the lidar depth data is integrated with acoustic

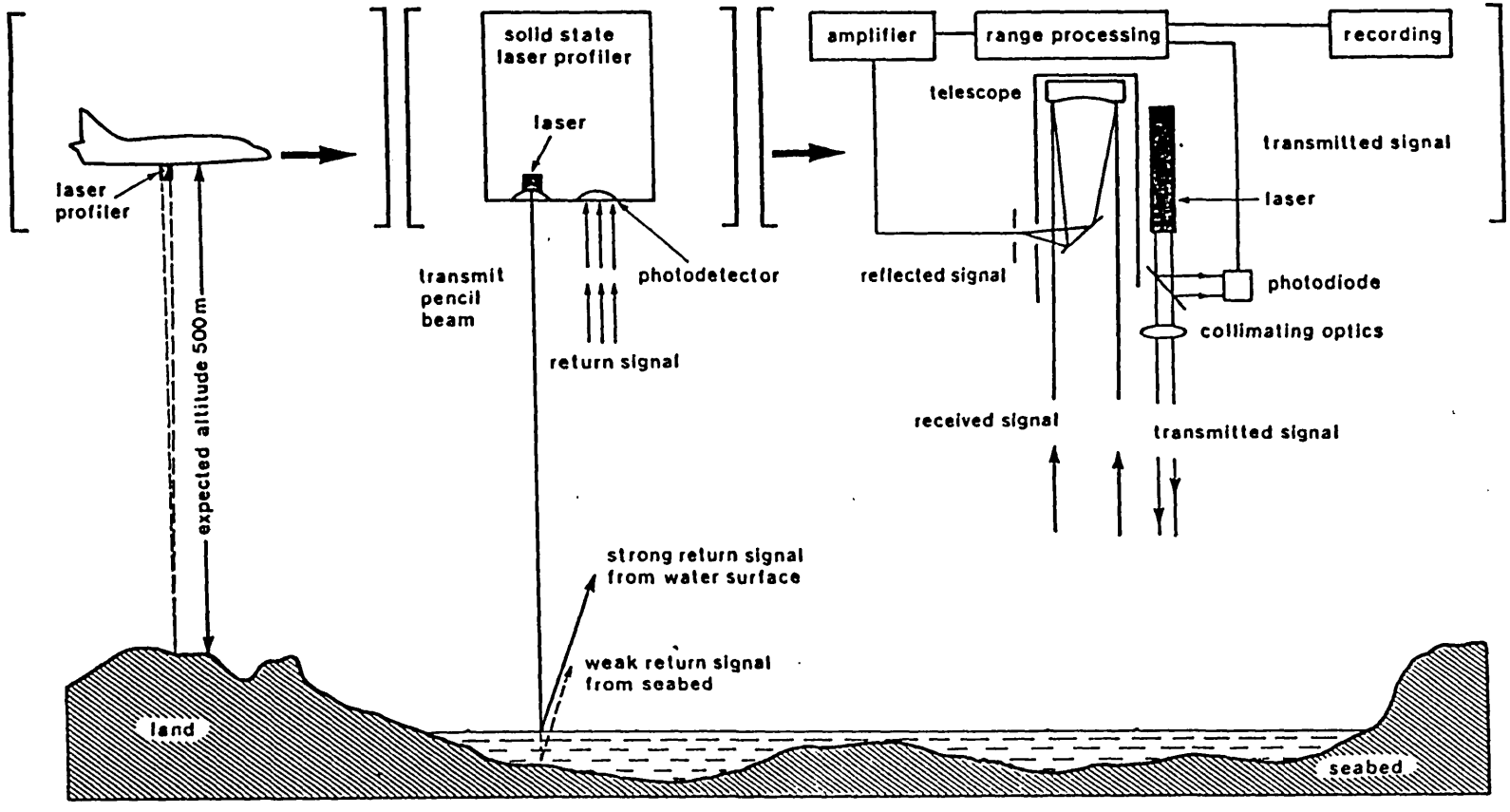


Figure 3.34 Mode of operation of lidar profiler.

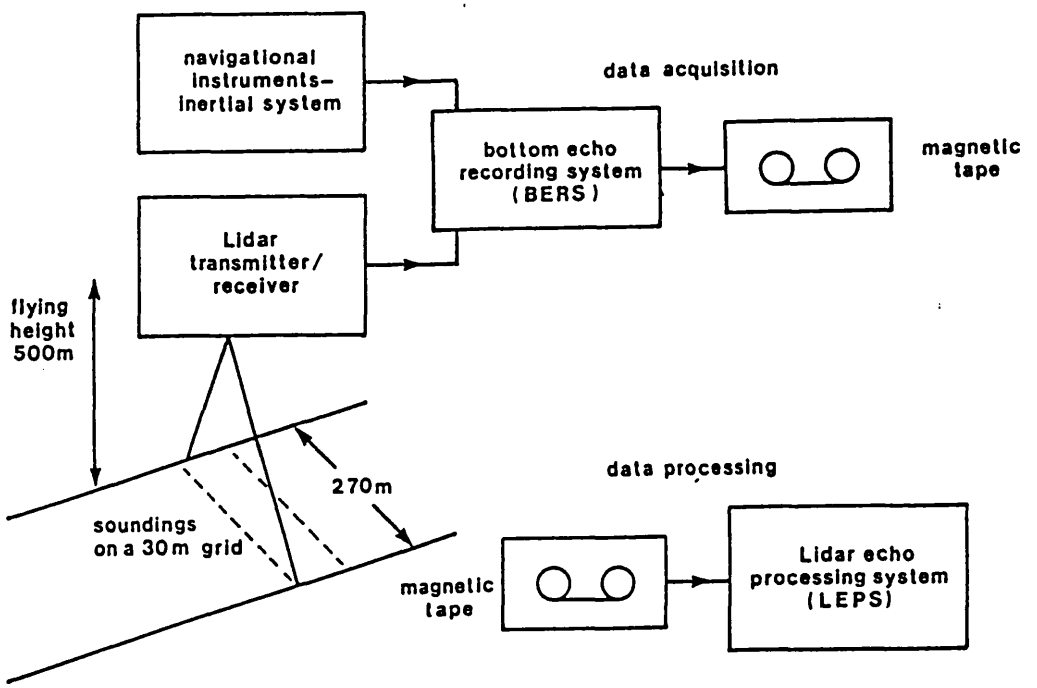


Figure 3.35 Components of the Larsen-500 Lidar Scanning System.

soundings, archived data, aids to navigation and so on to create a final hydrographic chart.

3.4 Conclusions

Remote sensing scanning systems of the type discussed in this chapter may be considered both as an alternative and as a means of complementing the conventional photographic systems which were discussed in Chapter 2.

Whilst, in general, the spatial resolution of scanning systems is not as good as that obtainable using photographic processes, they do nevertheless offer several major benefits. Firstly they offer the user the facility to record over a much wider spectral range, and to sense regions of the EM spectrum which are 'invisible' to conventional photographic processes. Secondly the ability of active systems, such as SLAR, to produce imagery in all weathers, regardless of the time of day is a major advantage in many regions of the world. However the principal benefit of the scanner approach is the opportunity to digitally process the remotely sensed data in order to enhance and classify the information content of the imagery. Further details relating to this aspect of remote sensing are dealt with in Chapter 4.

References

Allan, T. D. (1983) 'A Review of Seasat', in *Satellite Microwave Remote Sensing*, Allan, T. D. (ed.), Ellis Horwood, Chichester, 526 pp.

Anon. (1983) Exploration: new scanning system augments remote sensing techniques. *Mining Journal*, September.

Barnett, M. E., Bird, A. C. and Dawes, M. C. (1984) Multiband thermal infrared imagery as a discriminant for surface rock types. *Int. Journal Remote Sensing* 5(2) 511–515.

Begni, G. (1982) Selection of the optimum spectral bands of the SPOT satellite. *Photogrammetric Engineering and Remote Sensing* 48, 1613–1620.

Benny, A. H. and Dawson, G. J. (1983) Satellite imagery as an aid to bathymetric charting in the Red Sea. *The Cartographic Journal* 20(1) 5–16.

Berg, C. and McGinnis, D. F. (1980) 'Mapping of the 1978 Kentucky River flood from NOAA-5 satellite thermal infrared data', in *Proceedings of the 46th Annual Meeting of the American Society of Photogrammetry*, St Louis, Missouri, pp. 106–111.

Bullard, R. K. and Lakin, P. J. (1981) *First Steps in Remote Sensing*. Working Paper No. 3, North East London Polytechnic, Department of Land Surveying, 61 pp.

Chevrel, M., Courtois, M. and Weill, G. (1981) The SPOT satellite remote sensing mission. *Photogrammetric Engineering and Remote Sensing* 47, 1163–1171.

Ducher, G. (1980) Cartographic possibilities of the SPOT and Spacelab projects. *Photogrammetric Record* 10(56), 167–180.

Elachi, C. (1980) Spaceborne imaging radar: geologic and oceanographic applications. *Science* 216, 1073–1082.

Elachi, C., 1982(*a*) Shuttle imaging radar experiment. *Science* 218, 996–1003.

Elachi, C., 1982(*b*) Radar images of the Earth from space. *Scientific American* 247(6), 46–54.

Estes, J. E., Mel, M. E. and Hooper, J. O. (1977) Measuring Soil Moisture with an Airborne Imaging Passive Microwave Radiometer. *Photogrammetric Engineering and Remote Sensing* 43, 1273–1281.

Foster, J. L. and Hall, D. K. (1981) Multisensor analysis of engineering features. *Photogrammetric Engineering and Remote Sensing* 47, 655–664.

Haskell, A. (1981) 'The European SAR 580 Campaign as a preparation for the ERS-1 Satellite SAR', in *Matching Remote Sensing Technologies and Their Applications*, Remote Sensing Society, London, pp. 50–51.

Heilman, J. L. and Moore, D. G. (1982) Evaluating depth of shallow groundwater using HCMM Data. *Photogrammetric Engineering and Remote Sensing* 48, 1903–1906.

Hofmann, O. (1982) Digitale Aufnahmetechnik. *Bildmessung und Luftbildwesen* 50, 16–25.

Holmes, A. (1983) Shuttle Imaging Radar-A. 1983. Information and data availability. *Photogrammetric Engineering and Remote Sensing* 49, 65–67.

Jensen, H., Graham, L. C., Porcello, J. and Leith, E. N. (1977) Side looking airborne radar. *Scientific American* 237(6), 84–94.

Lillesand, T. M. and Kiefer, R. W. (1979) *Remote Sensing and Image Interpretation*. John Wiley and Sons, New York, 612 pp.

NASA, *Landsat Data Users Notes*, 1978–1983. U. S. Geological Survey, Sioux Falls, S. Dakota.

Raney, R. K. (1983) 'The Canadian SAR Experience', in *Satellite Microwave Remote Sensing*, Allan, T. D. (ed.), Ellis Horwood, Chichester, 526 pp.

Rudd, R. D. (1974) *Remote Sensing: A Better View*. Duxbury Press, N. Scituate, Massachusetts, 135 pp.

Schowengerdt, R. (1982) Enhanced thermal mapping with Landsat and HCMM digital data. *Proceedings of the 48th Annual Meeting of the American Society of Photogrammetry*, Denver, Colorado, 414–422.

Swain, P. H. and Davis, S. M. (1978) *Remote Sensing: The Quantitative Approach*. McGraw-Hill, New York, 396 pp.

Townsend, J. R. G. (ed.) (1981) *Terrain Analysis and Remote Sensing*. George Allen and Unwin, London, 232 pp.

Watson, K., Hummer-Miller, S. and Offield, T. (1982) Geological thermal inertia mapping using HCMM satellite data. *Proceedings of the IGARSS' 82 Symposium*, Munich.

Wright, R. (1982) Land Use Applications of imaging radar: retrospect and prospect. *Photogrammetric Record* 10, 697–703.

6 Remote sensing for topographic and thematic mapping

T.J.M. KENNIE

6.1 Introduction

The creation of maps by aerial photogrammetry and the interpretation of information from aerial photography are two well-established techniques in surveying and civil engineering. In recent years, however, developments in satellite and airborne sensors, together with improvements in computing speed and reductions in the costs of computer memory, have led to the creation of the relatively new discipline of remote sensing. Although still in an embryonic state, particularly in the field of image processing, remote sensing is nevertheless proving to be a cost-effective method of investigating regional engineering phenomena, such as surface drainage, and more localized phenomena such as landslides, building heat loss and so on. It is also being used, although to a much more limited extent, for topographic and hydrographic mapping.

This chapter is concerned with the fundamental principles of this new technology. Section 6.2 outlines the basic physical principles of the subject, while section 6.3 describes the characteristics of the airborne and satellite sensors which are used to acquire the remotely sensed data. The full potential of this type of data can only be fully exploited if appropriate hardware and software is available to enhance and classify the raw data. Section 6.4 therefore discusses the requirements of the digital image processing systems used in remote sensing. Finally, section 6.5 reviews a selection of the current applications of remote sensing relevant to the land surveyor and civil engineer.

6.2 Basic principles of remote sensing

Remote sensing basically refers to measurement at a distance. More precisely, it is concerned with the detection and measurement of variations in electromagnetic (EM) energy. Not all regions of the EM spectrum are of equal importance, since some are more affected by the attenuating effects of the Earth's atmosphere than others; Fig. 6.1 illustrates these attenuating regions and identifies those areas of the EM spectrum most significant to remote sensing.

6.2.1 *Characteristics of EM radiation*

The visible and reflected infrared regions of the spectrum represent the most important 'atmospheric windows', since they enable solar radiation reflected from the Earth to be

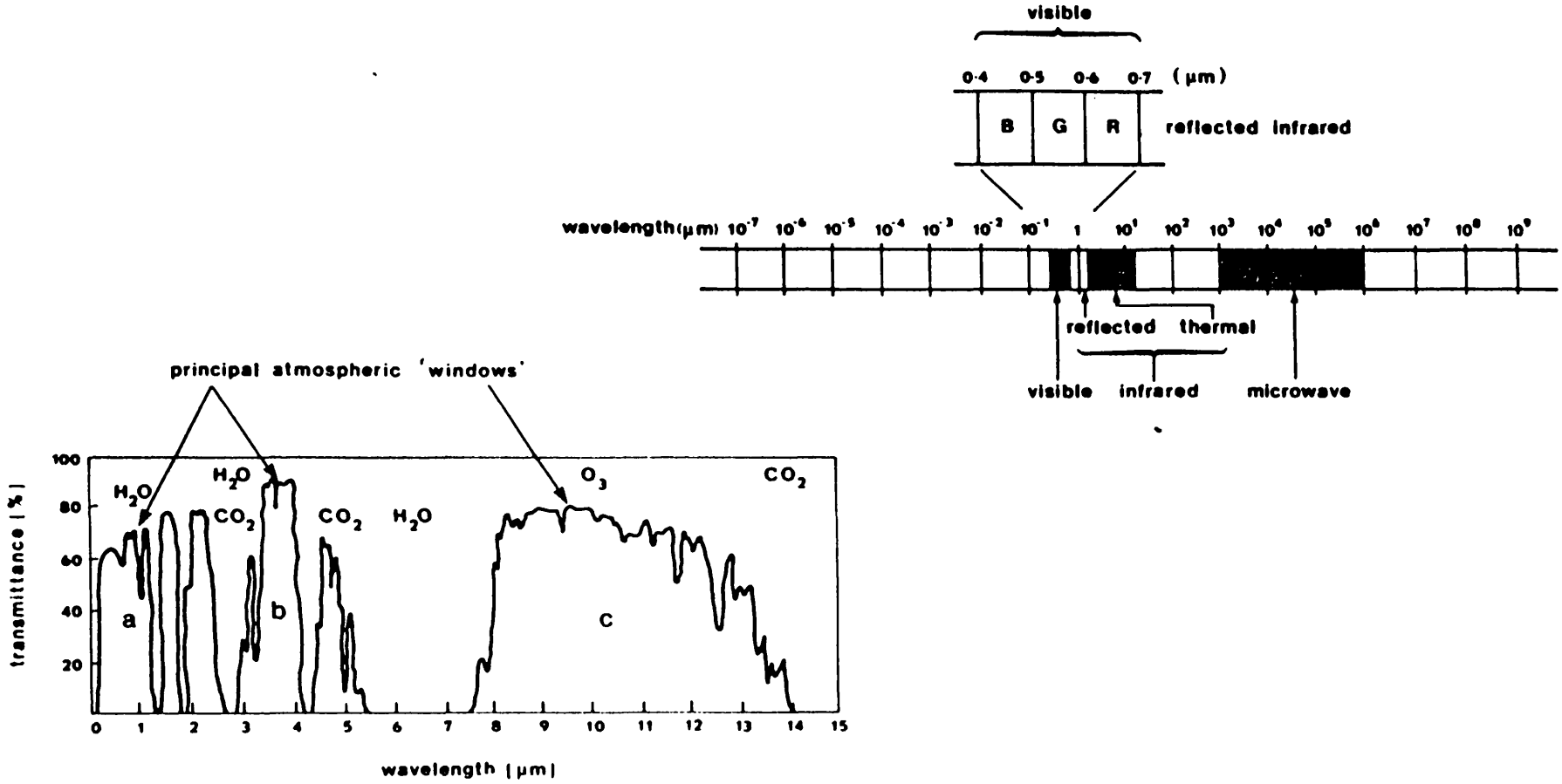


Figure 6.1 (Top) The electromagnetic spectrum (EM) and the principal regions relevant to remote sensing. (Bottom) Regions of the EM spectrum attenuated by the Earth's atmosphere. (*a*) Visible/near infrared; (*b*) far infrared (emitted); (*c*) far infrared (thermal).

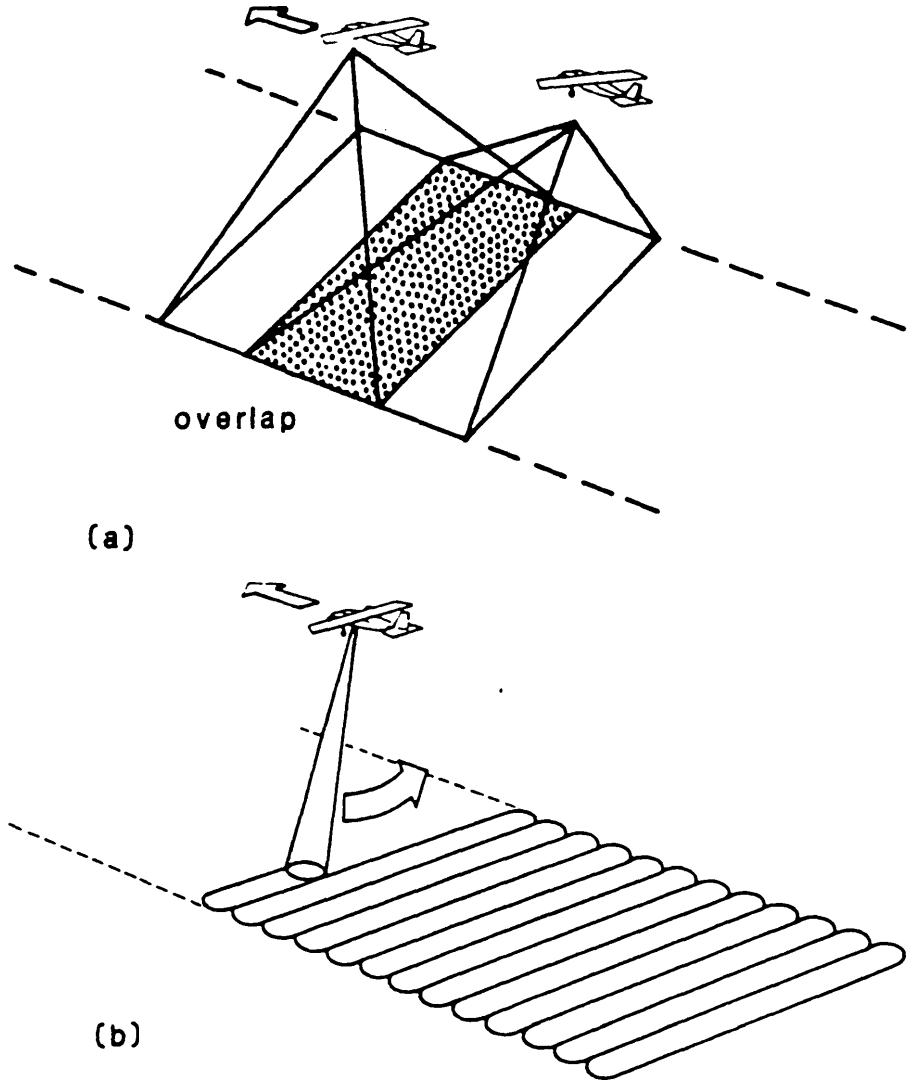


Figure 6.2 Modes of data acquisition. (*a*) Simultaneously using an aerial camera; (*b*) sequentially using a linescanner.

measured. At longer wavelengths, for example in the thermal infrared (8–14 μm) region, the measurement of radiation is of greater importance. The amount of energy emitted by a body can be quantified by means of the Stefan-Boltzmann law:

$$W = C\Sigma T^4 \tag{6.1}$$

where W is the radiant emittance in $\mathrm{W\,m^{-1}}$

Σ is the emissivity of the object (a variable quantity which describes how efficiently an object is radiating in comparison with a theoretically perfect radiator)

C is the Stefan-Boltzmann constant, equal to $5.7 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$ and

T is the absolute temperature of the object.

The application of this simple concept enables the temperature of an object to be inferred by remotely detecting the radiance from the feature being examined. However, before precise determinations of temperature can be achieved, a knowledge of the spatial variation in the emissivity of the surfaces being sensed is vitally important. More details of the basic physical principles of the subject can be found in Curran (1987).

6.2.2 *Characteristics of remote sensing systems*

Depending on the nature of the radiation being measured, it is possible to record the reflected or emitted EM energy, using either a lens and photographic emulsion or a scanner and crystal detector element. The geometrical distinction between the two approaches is illustrated in Fig. 6.2. The scanning technique enables EM energy outside the sensitivity range of photographic emulsions to be measured.

All measurements within both the visible and infrared regions (reflected and emitted) are normally referred to as *passive* in nature, since the radiation being recorded occurs naturally. Although it is possible to measure passive radiation at longer wavelengths, for example in the microwave region, it is generally easier to actively generate EM radiation of this wavelength and record the reflected radiation from the terrain. Systems of this type are referred to as *active*, and side-looking radar is a typical example of a system based on this technique.

Before discussing the specific characteristics of the various data acquisition systems in detail, mention should be made of the distinction between the terms *photograph* and *image*. A 'photograph' generally refers to an image which has been detected by photographic techniques and recorded on photographic film. An 'image', in contrast, is a more general term which is used to describe any pictorial representation of detected radiation data. Therefore, although scanner data may be used to create a photographic product, it would normally be referred to as an 'image', since the original detection mechanism involved the use of crystal detectors, rather than a lens and photographic emulsion.

6.3 Data acquisition

A wide and varied range of sensing devices has been developed for the acquisition of remote sensing imagery. In general terms, instruments can be classified according to two criteria, the mode of operation and their spectral range. Thus it is possible to differentiate between:

(i) Instruments which expose the entire format at a single instant and record visible and reflected (near) infrared radiation by photographic (or non-photographic) processes (6.3.1)
(ii) Instruments which operate by scanning the image and record radiation in the visible, reflected infrared and emitted infrared regions of the spectrum by means of crystal detector elements, (6.3.2 and 6.3.3)
(iii) Instruments which operate by measuring the time taken for actively generated pulses of microwave radiation to be reflected from the Earth's surface (6.3.4).

In all three instances, the sensors may be mounted either in airborne or in spaceborne platforms, most commonly in fixed-wing aircraft and unmanned satellites. Each has its own specific and largely complementary features. For example, airborne platforms enable small localized phenomena to be recorded at high spatial resolution. By contrast, spaceborne platforms enable wide synoptic views to be obtained, but at much lower spatial resolutions. In the latter case this may range from kilometres in the case of meteorological satellites, to a few tens of metres in the case of the most recent Earth resources satellites.

6.3.1 *Frame imaging systems*

A frame imaging system refers to a sensing device which exposes the entire format of an image at a single instant. The most common example is the aerial camera where

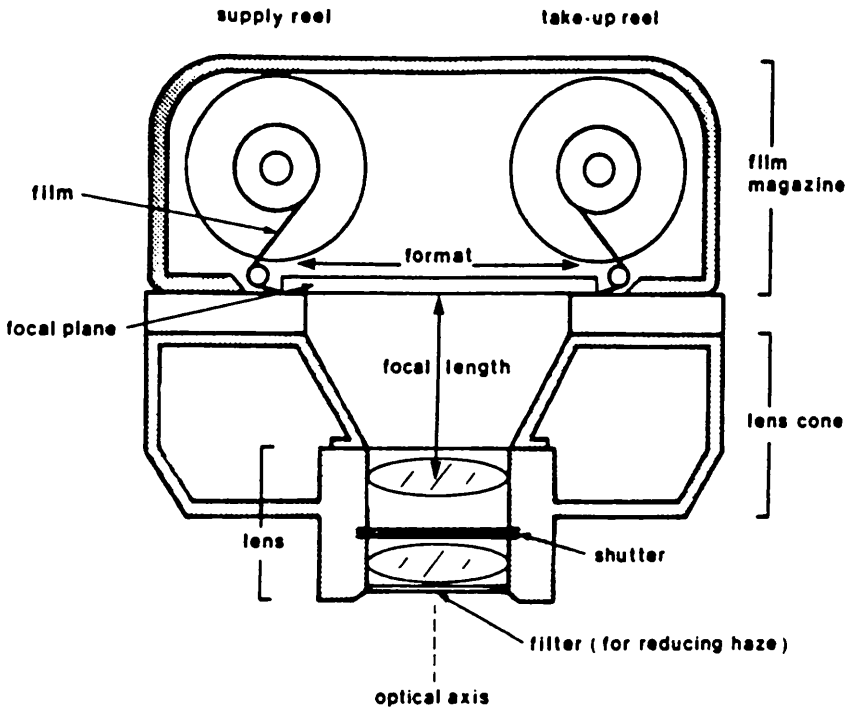


Figure 6.3 Components of a typical aerial photogrammetric mapping camera.

each frame is exposed by a camera shutter and the latent image is recorded on a photographic emulsion. Alternatively, the recording media may be non-photographic, and these are discussed further in section 6.3.1.2. The main distinguishing feature, however, is that in all cases the image produced has a perspective geometry and can therefore be analysed using conventional photogrammetric measuring techniques such as those described in Chapter 7 and discussed briefly in section 8.5.

6.3.1.1 *Photographic systems.* Vertical aerial photography obtained using a high-precision mapping camera is by far the most commonly available source of remotely sensed data. The design and construction of a modern photogrammetric mapping camera is given in Fig. 6.3.

The main components are:

(i) The main camera body which houses the motor drives and other mechanical components which are necessary to operate the film transport mechanism and the shutter
(ii) The lens-cone, which will normally be interchangeable, the lens, a diaphragm to regulate its aperture and a between-the-lens shutter
(iii) The magazine containing the unexposed and exposed spools of film, and the film flattening device which is usually of the vacuum suction type
(iv) The mount on which the camera will be supported and which enables the operator to level the camera, or to rotate it around its vertical axis to eliminate the effects of aircraft crabbing.

The current types of aerial camera produced by manufacturers such as Wild (Switzerland), Zeiss Oberkochen (West Germany) and Zeiss Jena (East Germany) and used for photogrammetric purposes are highly developed, and are characterized by:

(i) A relatively large format—230 × 230 mm is standard, although larger formats up to 230 × 460 mm have been used for civilian mapping from space platforms (Table 6.9)

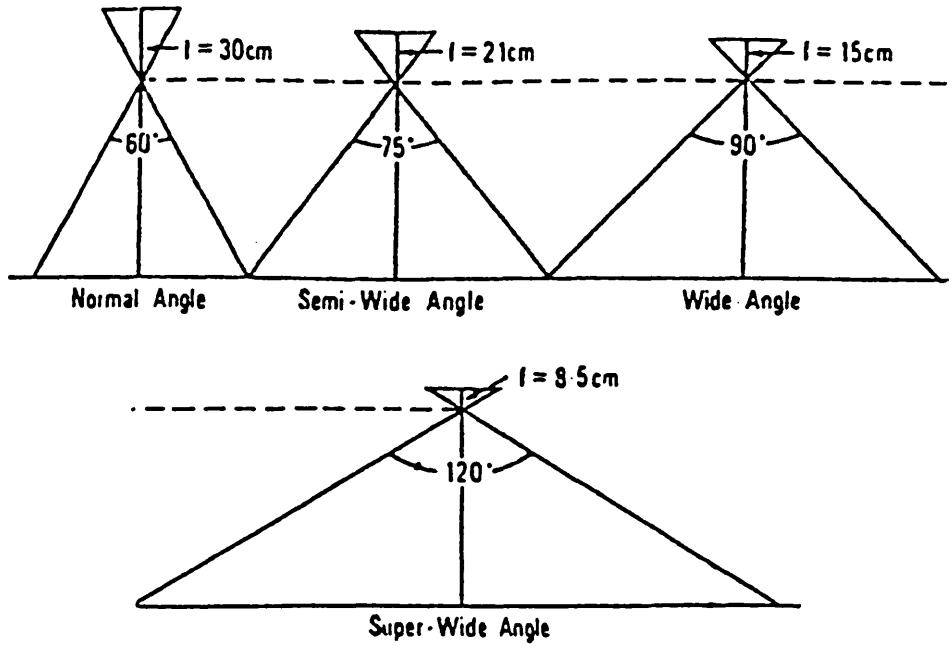


Figure 6.4 Angular coverage of aerial photogrammetric cameras of varying focal length.

Table 6.1 Geometric characteristics of typical designs of photogrammetric mapping cameras

	Normal-angle	Semi-wide-angle	Wide-angle	Super-wide-angle
Focal length	12" (30 cm)	$8\frac{1}{4}$" (21 cm)	6" (15 cm)	$3\frac{1}{2}$" (8.5 cm)
Angular coverage	56°	75°	93°	120°
Base:height ratio	0.3	0.45	0.6	1.0
Photo scale (assuming $H = 3000$ m)	1:10 000	1:14 300	1:20 000	1:34 300
Ground resolution (40 lp/mm)	0.8 ft (0.25 m)	1.05 ft (0.35 m)	1.6 ft (0.5 m)	2.8 ft (0.85 m)
Ground area covered	2.3 × 2.3 km	3.3 × 3.3 km	4.6 × 4.6 km	8.4 × 8.4 km

(ii) Very low geometric distortions (5–7 μm in the negative plane)
(iii) Large-aperture lenses ($f/4$ to $f/5.6$ at fullest opening) with high light-gathering powers at considerable angular values from the optical axis
(iv) Between-the-lens shutters of high efficiency (over 90% at large apertures) which also permit the use of short exposure times (0.5–1 ms) to cut down image motion
(v) Angular coverage of single exposures ranging from 60° (normal angle) to 120° (superwide angle) (Fig. 6.5) and Table 6.1
(vi) Use of a range of different emulsion types which are sensitive to specific wavebands.

For the purpose of mapping, it is necessary to ensure that the individual photographs overlap in their coverage of the terrain, primarily to enable stereo-viewing or stereo-measurements to be made. Although a 50% overlap in the forward direction of flight would be sufficient to provide a stereoscopic view for a number of reasons, illustrated by Fig. 6.4, it is common practice to specify that the overlap should be at least 60%. In

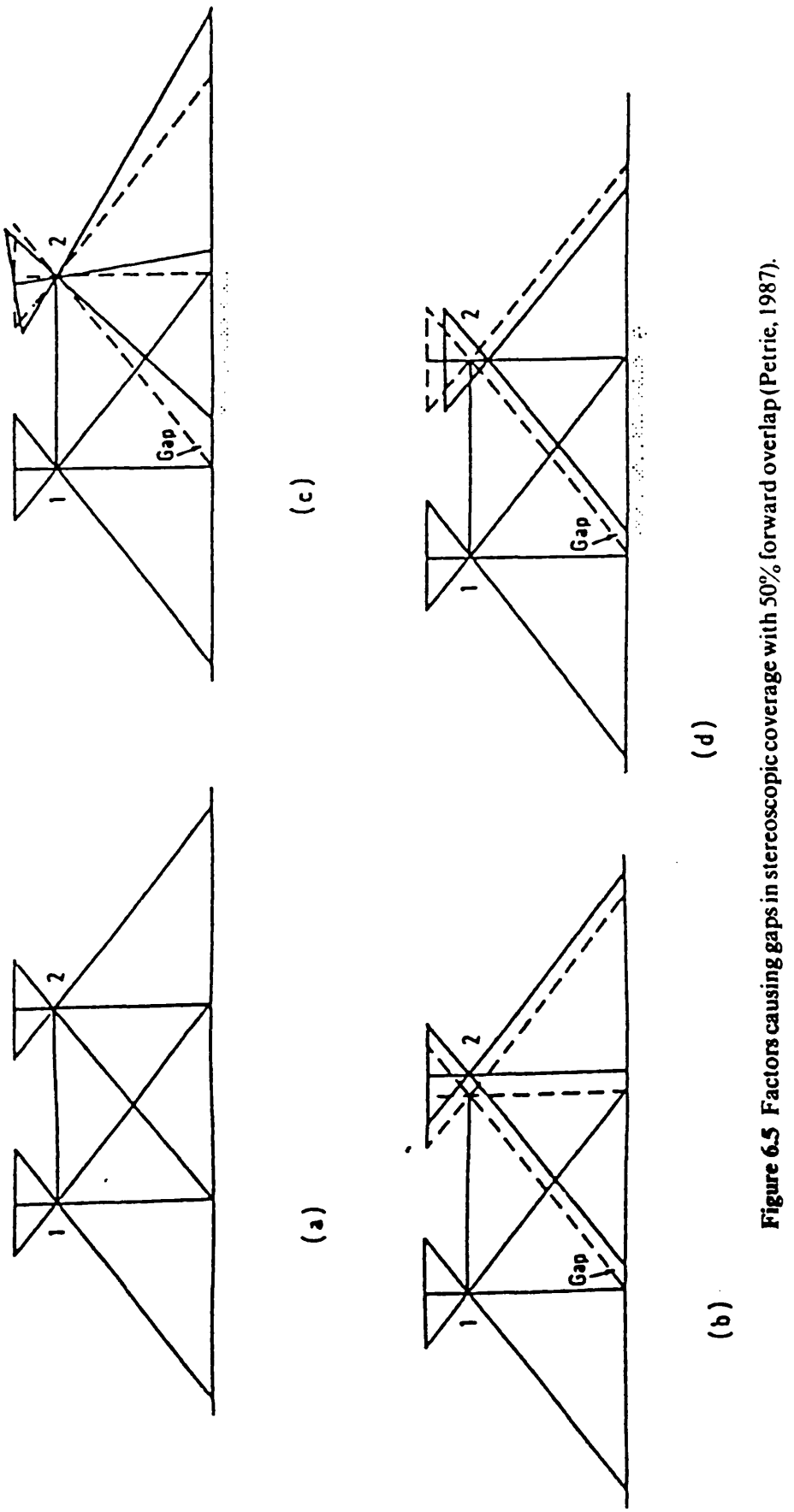


Figure 6.5 Factors causing gaps in stereoscopic coverage with 50% forward overlap (Petrie, 1987).

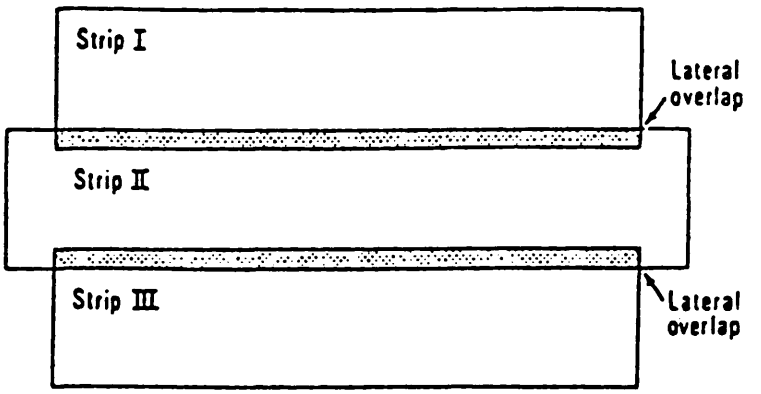


Figure 6.6 Lateral overlap between strips of aerial photography.

order to avoid gaps between adjacent strips of photography, it is again common procedure to specify a lateral overlap of between 15 and 20% (Fig. 6.6).

The ratio of the air base B to the flying height H is a parameter of particular importance for photogrammetric measurement, since it has a direct influence on the accuracy of heighting. Large B:H ratios, while preferable from a geometric standpoint, are also associated with super wide-angle coverage. The less favourable aspect of using such a camera is the loss of fine detail in the corners due to relief displacement and the 'dead' areas which occur in urban areas with tall buildings or regions of high relief.

Many photogrammetric cameras also include Image Movement Compensation (IMC) systems to further enhance the resolution of the photography. In essence, such systems operate by moving the film in its image plane at a rate corresponding to the aircraft speed. The importance of such a system at large scale can be highlighted by the following typical case. A standard high-resolution photogrammetric mapping camera producing 1:3000 scale photography will have a ground resolution of about 6 cm. If one assumes an exposure period of 1/250 second from a survey aircraft flying at 200 kph, then the aircraft will have moved a distance of 22 cm during the exposure interval. Thus the inherent high resolution of the large-scale photography will be degraded by the slight blurring of the image due to aircraft movement while the shutter is open. Meier (1985) provides a more exhaustive treatment of the theory of IMC, and Fig. 6.7 summarizes the effect of IMC for a number of typical photographic scales and aircraft speeds.

Small-format (typically 70 or 35 mm) low-precision cameras have also been used on occasions for interpretive work over small sites (Heath, 1980; Tomlins, 1983; Graham and Read, 1984). Coverage may involve the production of either vertical or oblique photography. In the latter case, the camera axis is intentionally tilted. Using the camera in this manner helps overcome the problems associated with the very narrow angular coverage of such camera systems when acquiring vertical photography. The oblique view may, however, mask features of interest, and the variation in scale across the format may make the accurate measurement of dimensions very difficult.

In instances when small-format cameras are being used, it is common to combine several (typically four) cameras to acquire photography of the same ground area, each camera having a different film/filter combination. The most common photographic films are black-and-white panchromatic (monochrome), black-and-white infrared, true-colour and false-colour infrared. Table 6.2 summarizes the characteristics of each type. Alternatively, filters may be placed over the each camera lens to restrict the

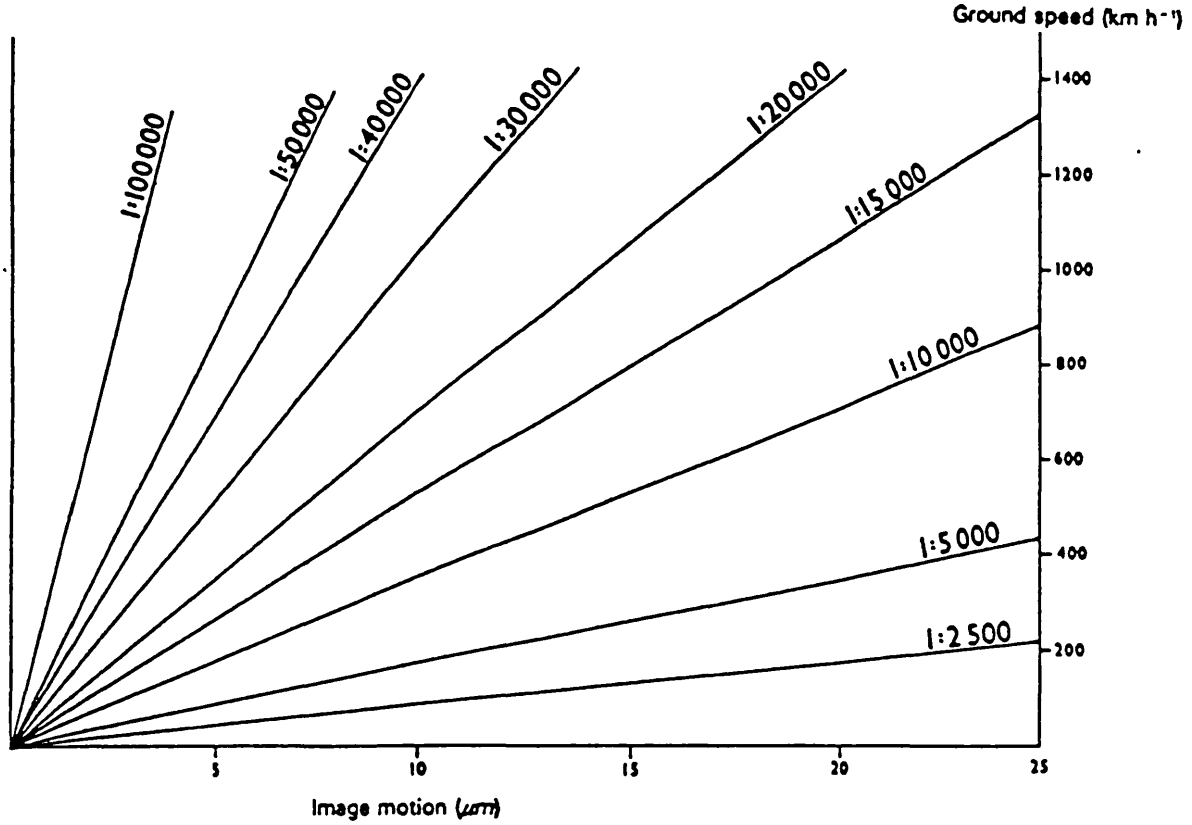


Figure 6.7 Image motion for a range of photographic scales and aircraft speeds (Meier, 1985).

responses to a narrow waveband, the results being recorded on a conventional panchromatic emulsion. These produce the so-called 'multispectral photography'. Such photography may be interpreted either by examining hardcopy or by using an 'additive viewer' to interactively examine the various film/filter combinations (Beaumont, 1977).

6.3.1.2 *Non-photographic systems.* It is also possible to record an image produced from a frame imaging system using non-photographic techniques. The most commonly used devices are video cameras, return beam vidicon (RBV) cameras, and cameras which use charge-coupled devices (CCD). With each of these devices, the image data are recorded either in an analogue manner on ,video tape, or digitally on high-density digitial tape. Hardcopy imagery can subsequently be generated, either by replaying the video tape and directly photographing the television screen on which the image is being displayed by use of a hardcopy video printer, or digitally by means of a raster scan laser-based film output device.

Video cameras mounted in an aerial platform have been used for a number of engineering applications, including the monitoring of traffic flow (Mountain and Garner, 1981), the recording of stages in the development of engineering projects, and the assessment of the impact of construction activities such as pipelines on the terrain.

The imaging system used in a video camera is based on a conventional lens and shutter arrangement. However, unlike photographic methods, the video image is formed by recording the charge falling on to a photosensitive screen (Fig. 6.8) rather than on to a silver halide emulsion. Incoming light falling on to the charge face of the vidicon tube induces a positive charge proportional to the brightness of the incoming

Table 6.2 Characteristics of various photographic emulsions

Film type	Structure	Sensitivity	Advantages	Limitations	Cost
Black-and-white panchromatic	Single-layer emulsion	0.4–0.7 μm	Good definition and contrast, wide exposure latitude. Boundaries are often clearly portrayed. Subtle textural variations can be identified and geometric patterns are more obvious	Sometimes difficult to interpret ground features due to the inability of the eye to distinguish between subtle differences in grey tones	Low
Black-and-white infrared	Single-layer emulsion	0.4–0.9 μm	It provides specific enhancement of different forms of vegetation and clearly defines bodies of water and ground moisture distribution. Helps to penetrate haze, giving a sharper more contrasting image	Contrast may be excessive and detail is often lost in areas of shadow	Low
True-colour	Multi-layered emulsions. Top layer is sensitive to blue light, the 2nd layer to green and blue light and the 3rd to blue and red light. A blue absorbing filter is introduced between the 1st and 2nd layers.	0.4–0.7 μm	Good contrast and tonal range. Wide exposure latitude for negative film types. Materials are more identifiable and hence moisture conditions are also more identifiable	Less good definition than panchromatic film	Intermediate to high
False-colour infrared	Multi-layered emulsion incorporating a reflected infrared sensitive layer. Top layer is sensitive to blue and reflected infrared light, 2nd to green and blue light and 3rd to red and blue light. The film is normally used with a filter to block out blue light.	0.4–0.9 μm	Helps to penetrate haze, provides accurate identifiable data on vegetation, rocks, soils, water bodies and moisture distribution	Lower resolution than colour film. Exposure difficult to determine and processing expensive	High

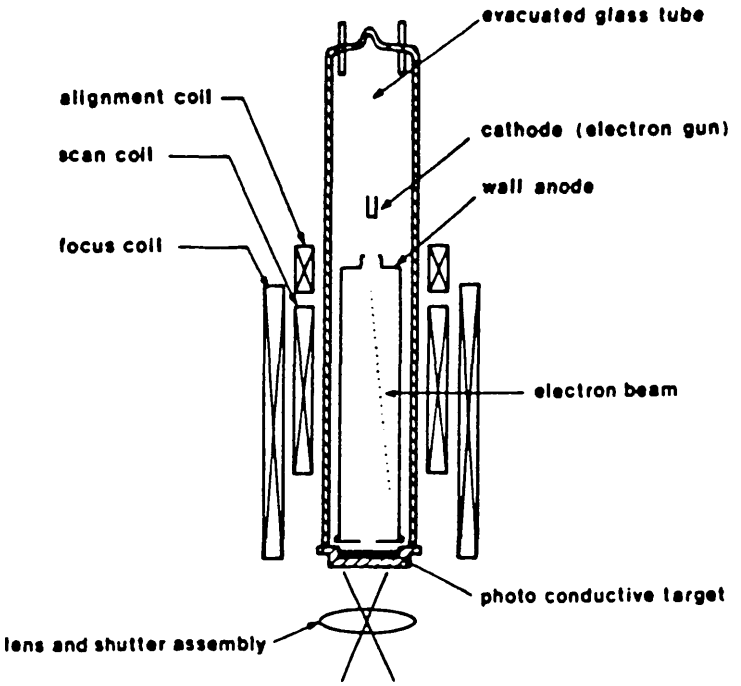


Figure 6.8 Principle of operation of a vidicon.

radiation, and hence a latent image is formed on the charge face. In order to 'develop' this latent image, an electron beam scans the charge face in a raster fashion, and in so doing discharges the initial positive charge.

The lenses used in video cameras normally have very short focal lengths (typically 8–35 mm), and the image format sizes are also very small in comparison with conventional photographic methods. Successive frames of imagery are recorded at rates which are typically between 25 and 30 Hz on to either U-Matic or VHS video tape.

Video cameras can be designed for a range of spectral sensitivities, although those which record the visible part of the EM spectrum, either in black-and-white or in colour, are the most common. Meisner and Lindstrom (1985) and Everitt and Nixon (1985), however, discuss the modifications which they designed for a conventional colour video camera in order to increase its spectral sensitivity and so enable the production of false-colour infrared imagery.

The *return beam vidicon* (*RBV*) camera operates in a manner similar to the more common vidicon. However, unlike the vidicon, the signal is continuously reflected back to the aperture of the electron gun. The main advantages of this design are the ability to produce higher-resolution images, and the ability of the photoconductive surface to remain in focus for longer periods.

The best-known example of an RBV camera system is that which operated on board Landsat 3 (section 6.3.2.2). Two RBV cameras with a spectral range of 0.505–0.750 μm (green to near infrared) were operated, and had a ground pixel equivalent to 30 m. Unfortunately, technical difficulties on board the satellite reduced the operational life of the sensor; nevertheless, considerable image coverage of the Earth's surface was acquired during this period.

A *charge coupled device* (CCD) camera differs from the previous two examples by having a rectangular, areal array of sensor elements situated in the imaging plane of the camera. Like the RBV, a latent image is formed on the surface of each array element, and the charge pattern formed is subsequently measured. Although the geometric

stability of the image is superior to that produced by a RBV camera (since a secondary scan is not required), at present the spatial resolution is significantly lower. Current arrays reported in the literature range from 100 × 100 elements (Hodgson *et al.*, 1981) to 288 × 385 elements (Traynor and Jeans, 1982), although an 800 × 800 element array has been reported for astronomical applications. Increasing interest has also been shown in their use as 'digital framing' or 'solid-state' cameras (Petrie, 1983; Burnside, 1985) to replace the conventional 230 × 230 mm format photographic camera. At present however, the size and density of arrays which are needed and the vast storage requirements (particularly at high resolution) preclude their widespread application.

CCDs are also appropriate to many other fields of surveying instrument design, and are discussed elsewhere in this book in relation to theodolites (1.5.1.3) and photogrammetric instrumentation (7.6.1, 8.7).

6.3.2 *Optical/mechanical line scanners*

The geometrical distinction between the instantaneous, photographic (or non-photographic) frame imaging technique and the sequential line scanning approach has been illustrated previously in Fig. 6.2. The other specific difference between such systems is the ability of the latter to measure radiation in regions of the EM spectrum which cannot be recorded on photographic emulsions, below 0.4 μm and above 0.9 μm wavelengths, in the ultraviolet and the longer infrared regions of the spectrum.

The basic principle of a line scanner involves the detection of the reflected or emitted radiation from objects present on the ground during an angular scan across the terrain, along a line perpendicular to the direction of flight. Scanning is performed by either a mechanically rotating or oscillating mirror. The area on the ground over which the radiation is measured is defined by the instantaneous field of view (IFOV) of the scanner. This also represents the size of the picture elements or pixels along each scan line. The IFOV is often used as a crude measure of the spatial resolution of the scanner. However, IFOV does not take into account other factors, particularly scene contrast, which can have a major impact on the detectability of features.

Two distinct designs of line scanner can be identified—those which record only a single waveband, generally in the thermal infrared part of the EM spectrum (in which case they are referred to as thermal scanners), and those which record images in several wavebands, the multispectral scanners (MSS).

6.3.2.1 *Thermal infrared (IR) line scanners.* Thermal IR line scanners operate by recording emitted IR radiation corresponding to the atmospheric window in the 8–14 μm regions of the EM spectrum. This region also coincides with the peak value of emitted energy from the Earth's surface.

The design of a typical Thermal IR linescanner is illustrated in Fig. 6.9. The incoming scanned radiation is focused on to a crystal detector, often of mercury cadmium telluride (HgCdTe), which transforms the radiation into an electronic signal. The analogue signal is recorded on video tape, and may be used to modulate a light beam moving across a cathode ray tube to produce a visible image. This 'quick look' image is also generally recorded on a continuous strip of moving film. It is also common practice to record the data digitally for later analysis in an image processing system.

In order to quantify the recorded signals, a blackbody reference is used to calibrate the raw data, and temperature variations as small as 0.2°C can be determined using such equipment. For absolute temperature measurements, however, factors such as wind speed and the spatial variations in the emissivity of different surface materials are

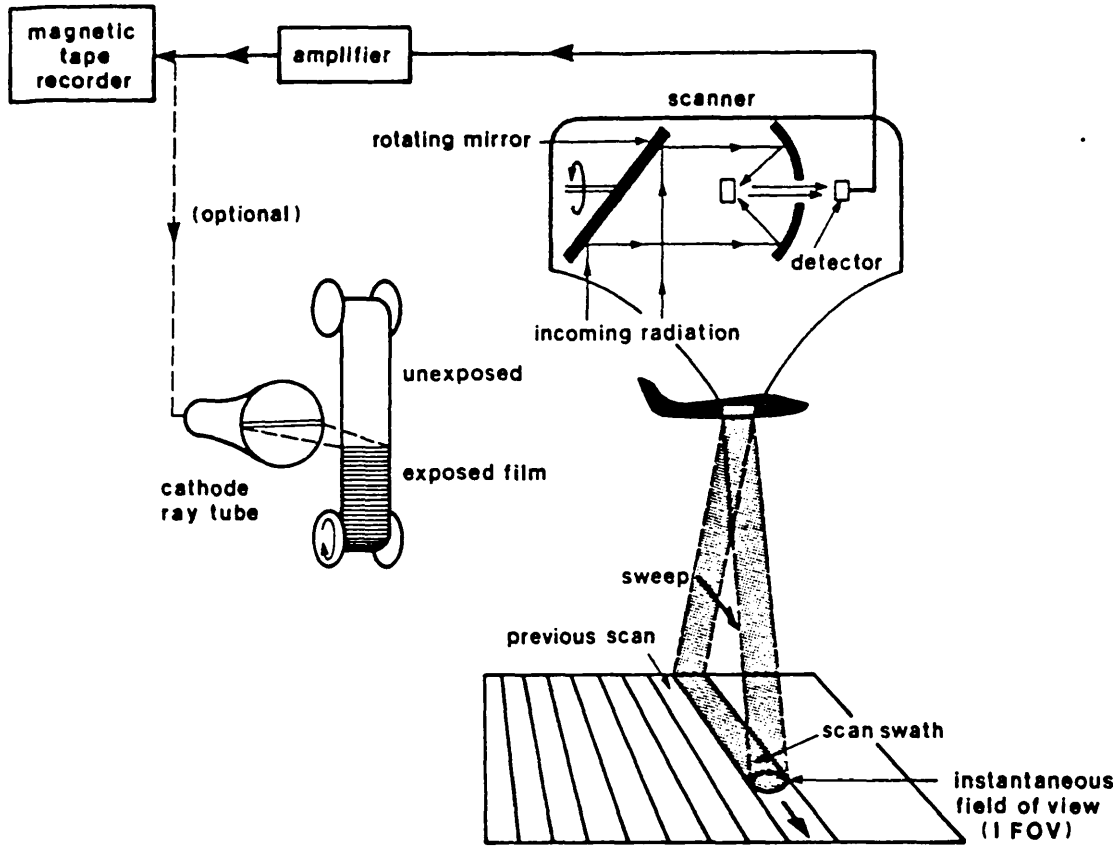


Figure 6.9 Components of a thermal linescanner (Rudd, 1974).

vitally important. Fig. 6.10 is a typical example of the type of image which may be created using this type of scanner, and which could, for example, be used for monitoring heat loss from buildings.

6.3.2.2 *Multispectral line scanners.* In general terms, multispectral scanners (MSS) can be classified into those of a mechanical design using, for example, rotating mirrors, or alternatively, those of the linear array design to be discussed in section 6.3.4. The following discussion is restricted to the principal features of the former design, although many of the characteristics are common to both.

The main components of an MSS are illustrated in Fig. 6.11. The line scanning mechanism is essentially the same as that used by thermal IR linescanners. The main difference lies in the separation of the incoming radiation into several spectral channels by means of a diffraction grating or prism. This enables the reflected or emitted radiation to be measured in several discrete wavebands for each individual pixel in the image. Both airborne and spaceborne platforms can be used to house MSS sensors.

The *Landsat* satellite system (Table 6.3), formerly known as the Earth Resources Technology Satellite (ERTS) was designed in the late 1960s by the US National Aeronautical and Space Administration (NASA), primarily to assess and monitor land resources. The first satellite was launched in 1972 and until 1985 the system was operated as a public service. The operational activities were then transferred to EOSAT, a joint company formed by the Hughes Aircraft Co. and RCA.

Three distinct sensors have been carried on board Landsat (up to 1987, 5 satellites have been in operation):

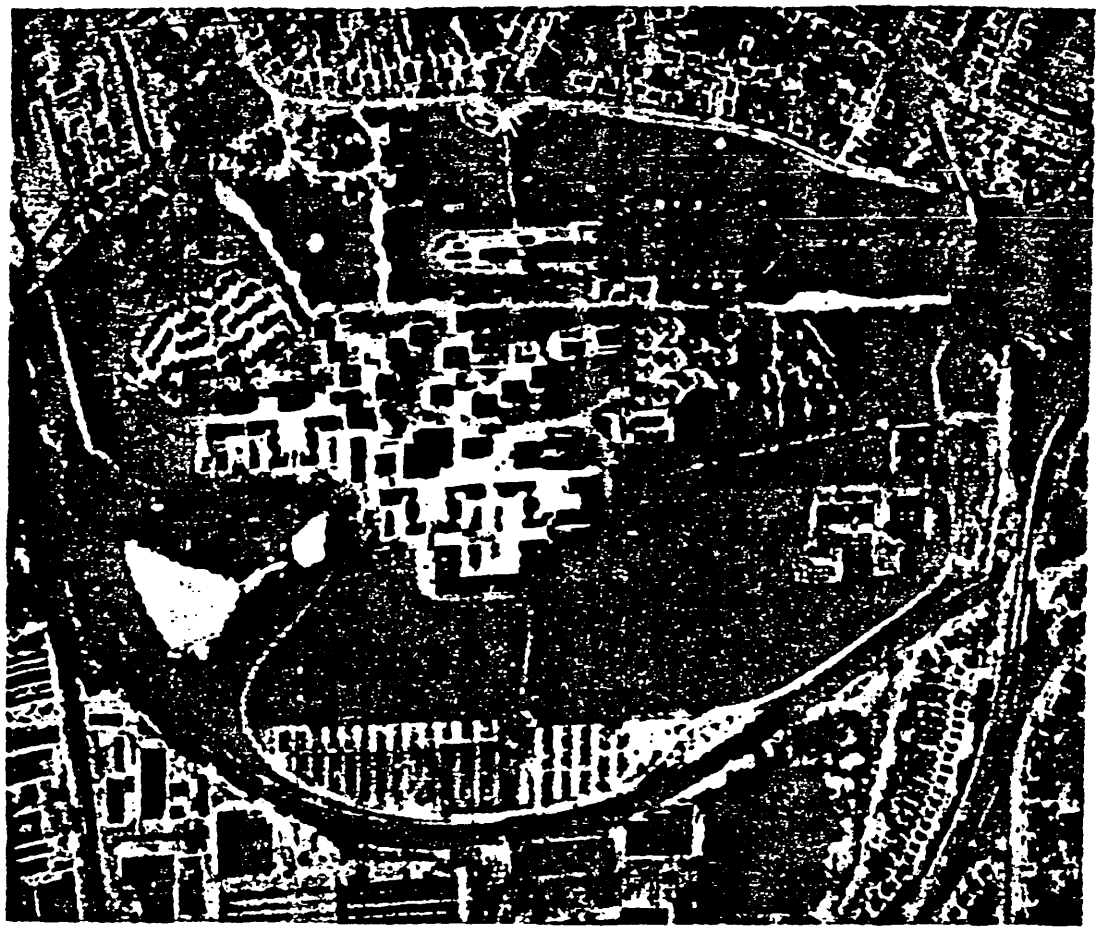


Figure 6.10 Thermal linescan image of the University of Surrey campus. Courtesy Clyde Surveys Ltd.

(i) A 4- or 5-waveband multispectral scanner (MSS)
(ii) A more advanced 7-waveband MSS termed the Thematic Mapper (TM)
(iii) A 3-waveband RBV camera (as discussed in section 6.3.1.2).

The former type of MSS has been carried on board all Landsat satellites. Scanning is performed using an oscillating mirror which scans six lines perpendicular to the orbital direction (Fig. 6.12). The swath width of a single image is 185 km, and the IFOV of the sensor produces a ground pixel equivalent to 56 × 79 m in the four visible and near infrared channels. The spectral range of these channels is from 0.5–1.1 μm. A fifth thermal channel covering the 8–14 μm spectral range has also been provided, and has operated on occasions, although with an enlarged IFOV giving a ground resolution equivalent to 235 m.

The *Thematic Mapper* (TM) sensor has only been operated on board Landsat 4 (launched 1982) and Landsat 5 (launched 1984). It is essentially an advanced MSS sensor which offers an improved spatial performance (with an IFOV equivalent to 30 m) and spectral range (7 wavebands) compared to the MSS. Nevertheless, in spite of these improvements, the high cost of TM data has proved to be a major limitation which has restricted the more widespread adoption of this type of data for engineering projects. The details of the specific wavebands covered by the TM are provided in Table 6.4, and the comparative spatial resolution of the two sensors is illustrated by Fig. 6.13.

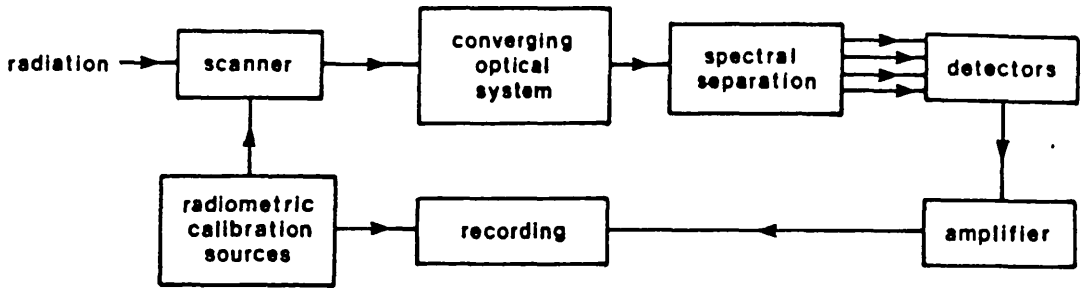


Figure 6.11 Principle of operation of a multispectral scanner.

Table 6.3 Characteristics of a selection of satellite scanners

	Line scanners		Pushbroom scanners	
	Landsats 1–5 Multispectral Scanner (MSS)	Landsats 4, 5 Thematic Mapper (TM)	Modular Optoelectronic Multispectral Scanner (MOMS)	Le Système Probatoire d'Observation de la Terre (SPOT)
Operated by (country)	EOSAT (USA)	EOSAT (USA)	DFVLR (W. Germany)	SPOT Image (France)
Date of launch	1972	1982, 84	1983	1986
Orbital altitude (km)	900	705	300	830
Ground resolution (m)	80	30	20	10, 20
Spectral range	0.5–12.6	0.45–12.5	0.575–0.975	0.5–0.89
No. of wavebands	5	7	2	3
Further reading	NOAA (1978–1986)	NOAA (1978–1986)	—	Chevrel *et al.* (1981)

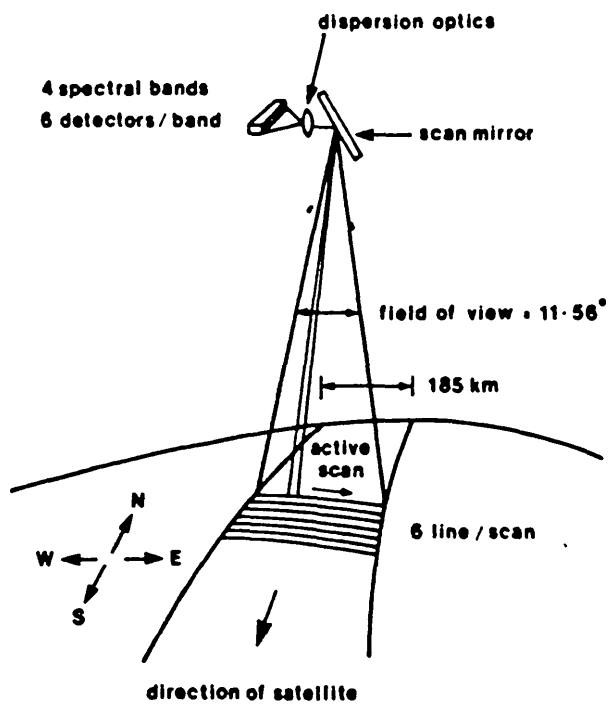


Figure 6.12 Landsat Multispectral Scanner (MSS) (after NASA).

Table 6.4 Landsat Thematic Mapper: spectral channels

Band number	Spectral range (μm)	Application
1	0.45–0.52	Water penetration
2	0.52–0.60	Measurement of visible green reflectance
3	0.63–0.69	Vegetation discrimination
4	0.76–0.90	Delineation of water bodies
5	1.55–1.75	Differentiation of snow and cloud
6	10.40–12.50	Thermal mapping
7	2.08–2.35	Geological mapping

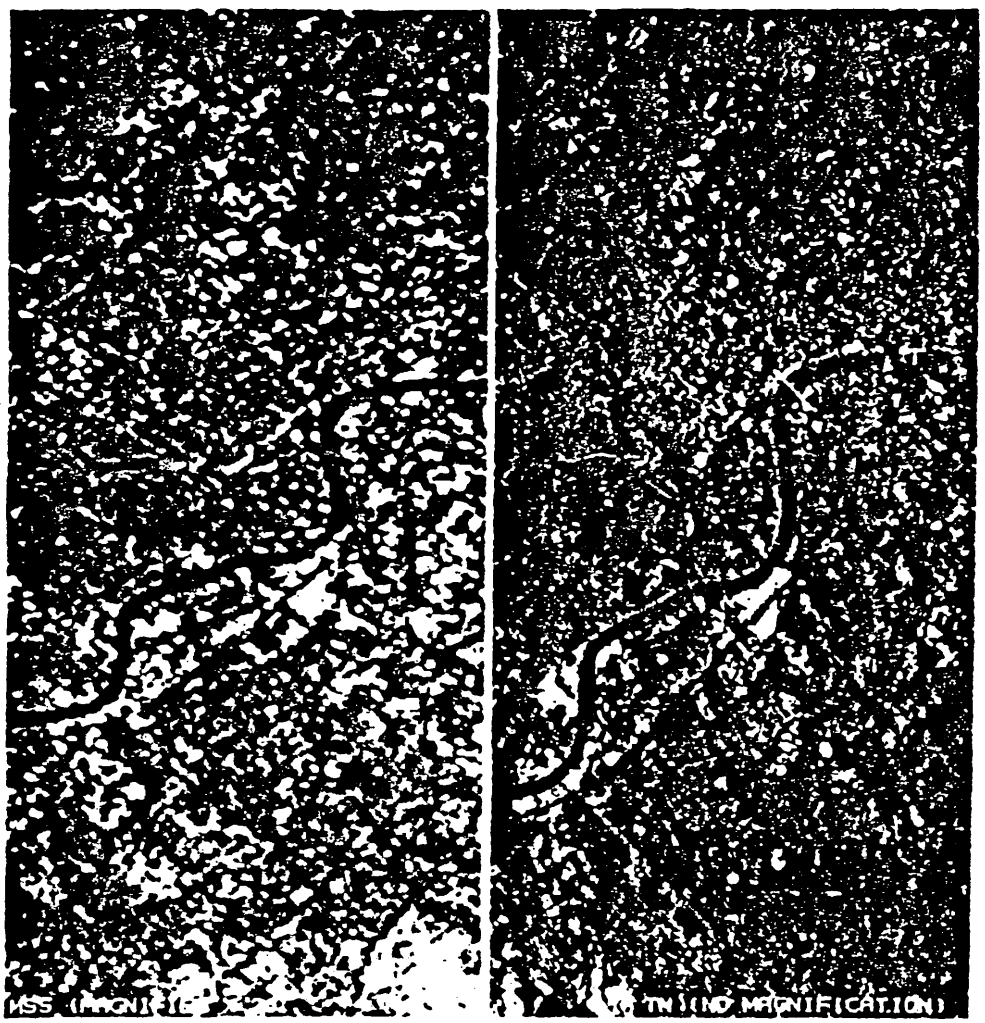


Figure 6.13 Comparative spatial resolution of (left) Landsat Multispectral Scanner (MSS) and (right) Landsat Thematic Mapper (TM) imagery. Courtesy National Remote Sensing Centre.

Figure 6.13 shows two typical single-waveband, black-and-white photographic images produced from Landsat digital data. Only a limited amount of information can be gleaned from the visual interpretation of single-band imagery. However, it is also possible to photographically combine three different wavebands to form a hardcopy colour 'composite' image. In the case of MSS imagery it is only possible to produce a 'false'-colour composite, while with TM it is also possible to produce a 'true'-colour composite image.

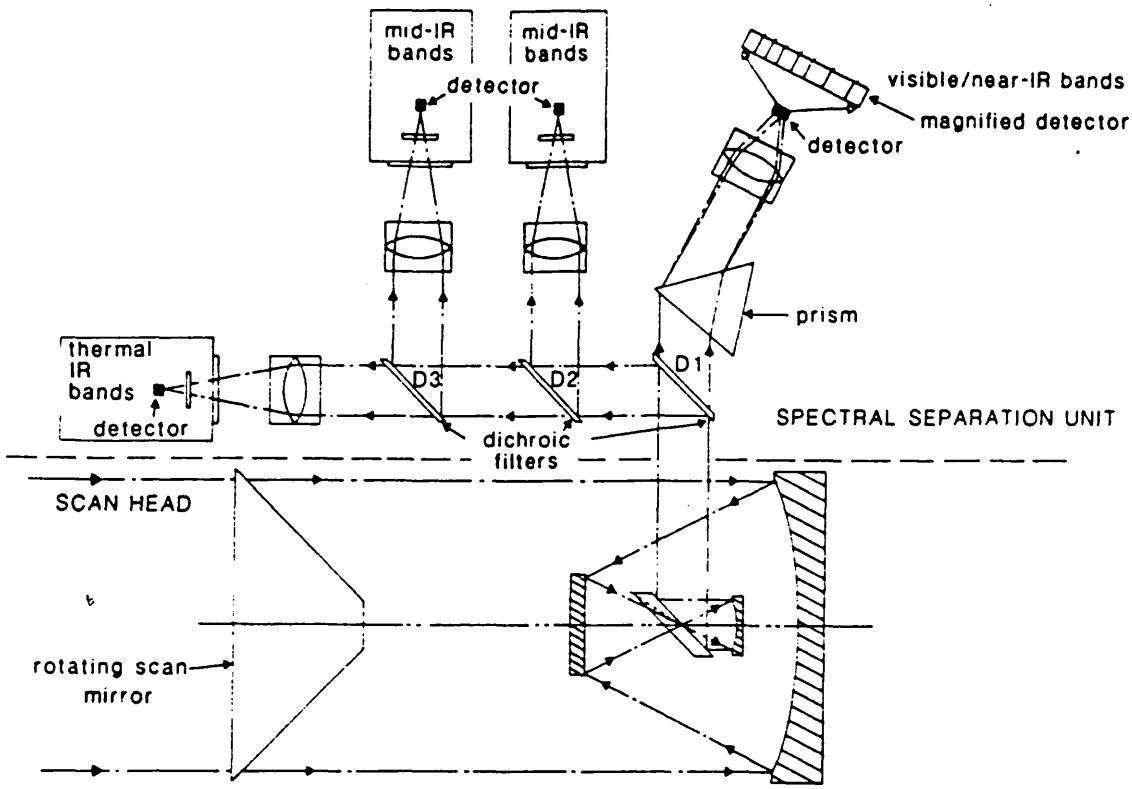


Figure 6.14 Optical components of Daedalus AADS 1268 Airborne Thematic Mapper.

Further benefits can be obtained if the imagery is examined interactively using an image processing system. However, to date, the high cost of such systems has meant that few surveying and engineering companies have invested in this technology. With the continued reductions in the cost of computer processing power, this is less likely to be a major limiting factor in the future.

The *Daedalus AADS 1268 Airborne Thematic Mapper* (ATM) is a typical airborne multispectral line scanner. Scanning is performed by a rotating mirror which operates at speeds of between 12.5 and 50 scanlines per second. The separation into 11 spectral channels is performed by three dichroic filters and a diffracting prism (Fig. 6.14). Table 6.5 summarizes the radiometric specifications of the scanner. The instrument has an IFOV of 2.5 m rad, which approximates to a ground resolution of a metre at a flying height of between 400 and 500 m. Data in the form of digitized analogue video signals for each channel are recorded on to high density digital tape (HDDT). These data are subsequently transferred on to computer compatible tape in order to read the data into an image processing system. Figure 6.15 is a typical example of the imagery which has been obtained with this type of scanner.

6.3.3 *Optical/mechanical frame scanners*

A more complex optical/mechanical design of scanner which has been used for acquiring remote sensing imagery is the frame scanner. In this instance, entire frames of data are captured, rather than single lines as described in section 6.3.2. Sensors of this design normally operate in the thermal part of the spectrum and may therefore be referred to as thermal video frame scanners (TVFS). Sensors of this type are particularly useful for engineering projects because they are portable and can be used in light aircraft

Table 6.5 Spectral channels for the Daedalus 1268 AADS ATM

Spectral band number	Wavelength (μm)	Spectral region
1	0.42–0.45	Blue
2	0.45–0.52	
3	0.52–0.60	Green
4	0.605–0.625	Red
5	0.63–0.69	
6	0.695–0.75	Near-IR
7	0.76–0.90	
8	0.91–1.05	
9	1.55–1.75	
10	2.08–2.35	Mid-IR
11	8.50–13.00	Far (thermal) IR

(Kennie *et al.*, 1986). This also reduces the operational costs of using the sensor. One particular TVFS system which has been used for remote sensing applications is the Barr and Stroud IR18.

6.3.3.1 *Barr and Stroud IR18 TVFS.* Unlike the line scanners previously described, which are designed specifically for airborne use, TVFS systems such as the Barr and Stroud IR18 (Fig. 6.16) are designed for both ground and airborne operations. Since such a system may therefore be used from both static and dynamic platforms, some mechanism for scanning in two orthogonal directions is necessary, and is described in detail in Amin (1986). Thus, incoming radiation is initially focused on to a frame scanning mirror oscillating at 50 Hz. The radiation reflected by this mirror is then directed on to a spherical focal plane. The line scanning is performed by a hexagonal prism which rotates at a rate of 39 000 rpm, scanning four lines of the image simultaneously. The image reflected by the line scan mirror is then focused by the relay lens onto a bank of four SPRITE (signal processing in the element) HgCdTe detector elements. The outputs from the detectors are stored in analogue form in a CCD storage unit and then read out in a TV-compatible PAL video format.

The main disadvantage of this design of scanner is the lack of a calibrated blackbody reference for quantitative temperature measurements. For some engineering applications, this is a severe limitation, but for others, for example where anomalous thermal patterns may indicate regions of subsidence (Edmonds *et al.*, 1986), a knowledge of the relative temperatures within a scene can be just as important. Quantitative temperatures can, however, be derived from the imagery by correlating known ground temperatures with corresponding image points. This technique has been used to assess the temperature variations throughout a road network with the aim of identifying those areas most prone to ice formation (Stove *et al.*, 1987). Figure 6.17 is an example of the type of imagery which was generated for this project.

6.3.4 *Linear array (pushbroom) scanners*

The linear, multilinear or 'pushbroom' scanner differs considerably from the optical-mechanical design. In fact, although such devices are referred to as scanners, the main



Figure 6.15 Airborne multispectral scanner image of the River Humber produced using the Daedalus AADS 1268 Airborne Thematic Mapper. Courtesy Hunting Technical Services Ltd.

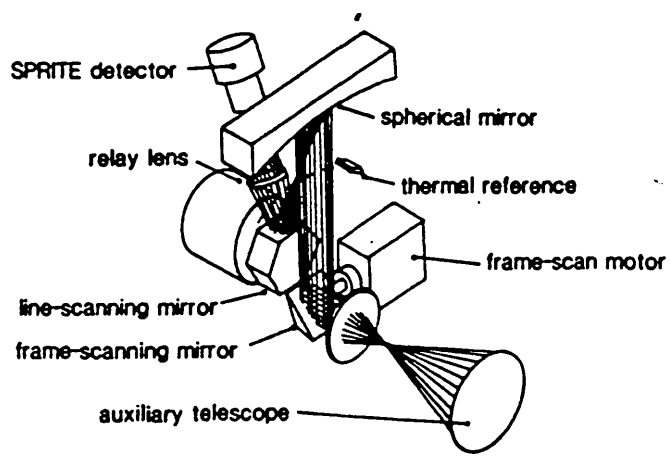


Figure 6.16 Components of the Barr and Stroud IR18 Thermal Video Frame Scanner (TVFS). Courtesy Barr and Stroud Ltd.

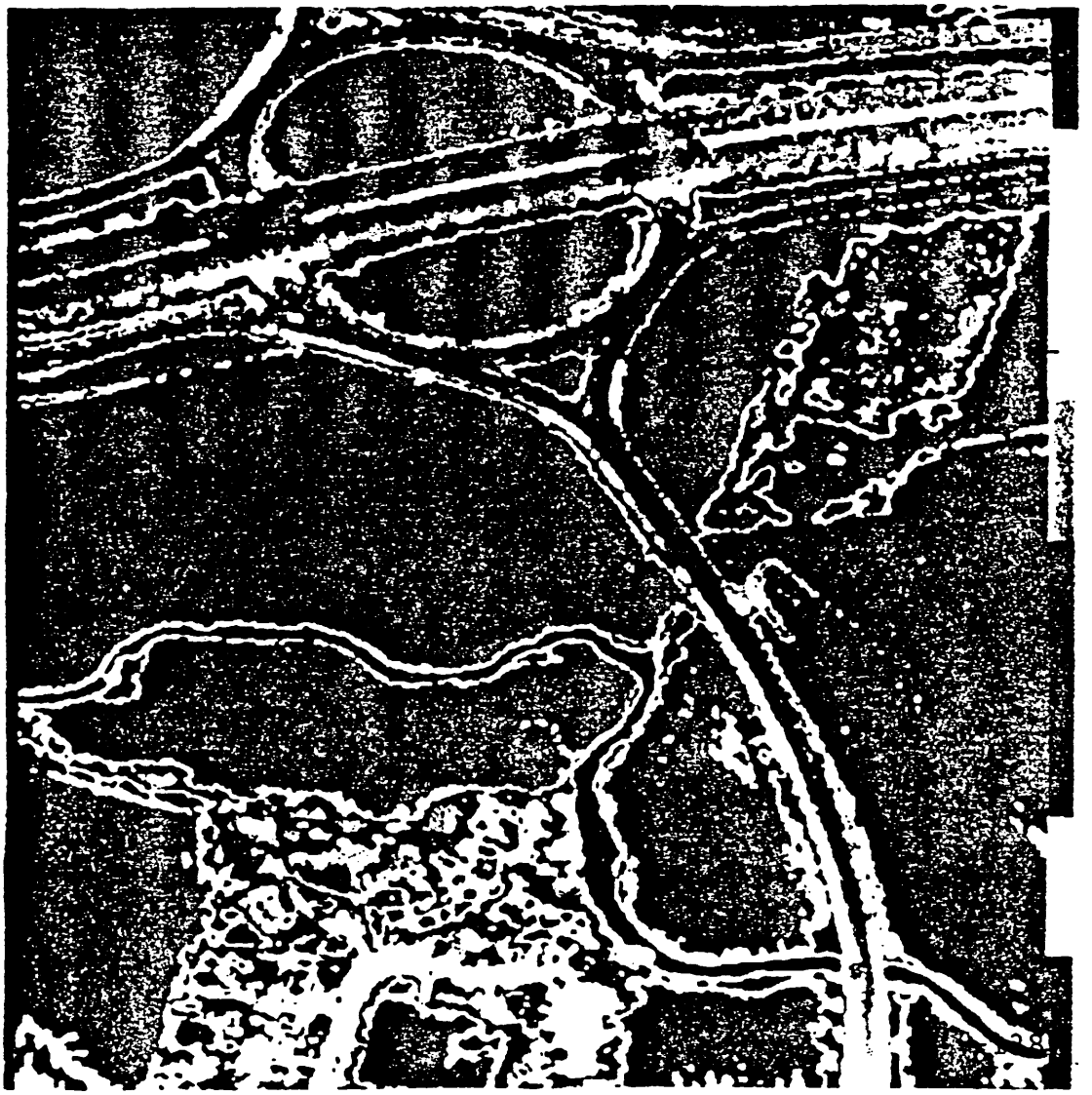


Figure 6.17 Density sliced Barr and Stroud IR18 TVFS image of part of the M9 motorway in Scotland, illustrating temperature variation over the road surface (original in colour). Courtesy ERSAC Ltd.

distinguishing feature of this design is the lack of any means of actually scanning the scene of interest. By contrast, each line of the image is formed by measuring the radiances as imaged directly on to a one-dimensional linear array of CCD detectors located in the instrument's focal plane (Fig. 6.18). Each line is then scanned electronically and the radiance values recorded on magnetic tape. As before, successive lines of the image are produced by the forward motion of the aircraft or satellite. Care has to be taken in the calibration of the individual detector elements to ensure that no degradation in the radiometric quality of the imagery occurs. Several satellite and airborne pushbroom scanners are currently in use, and the following sections describe a selected number of them.

6.3.4.1 *Le Système Probatoire d'Observation de la Terre (SPOT).* In common with Landsat, SPOT (Table 6.3) is also a commercial satellite remote sensing system, the original development being funded by the French Government and aerospace industries. SPOT was launched on 22 February 1986, and is the first of what is hoped

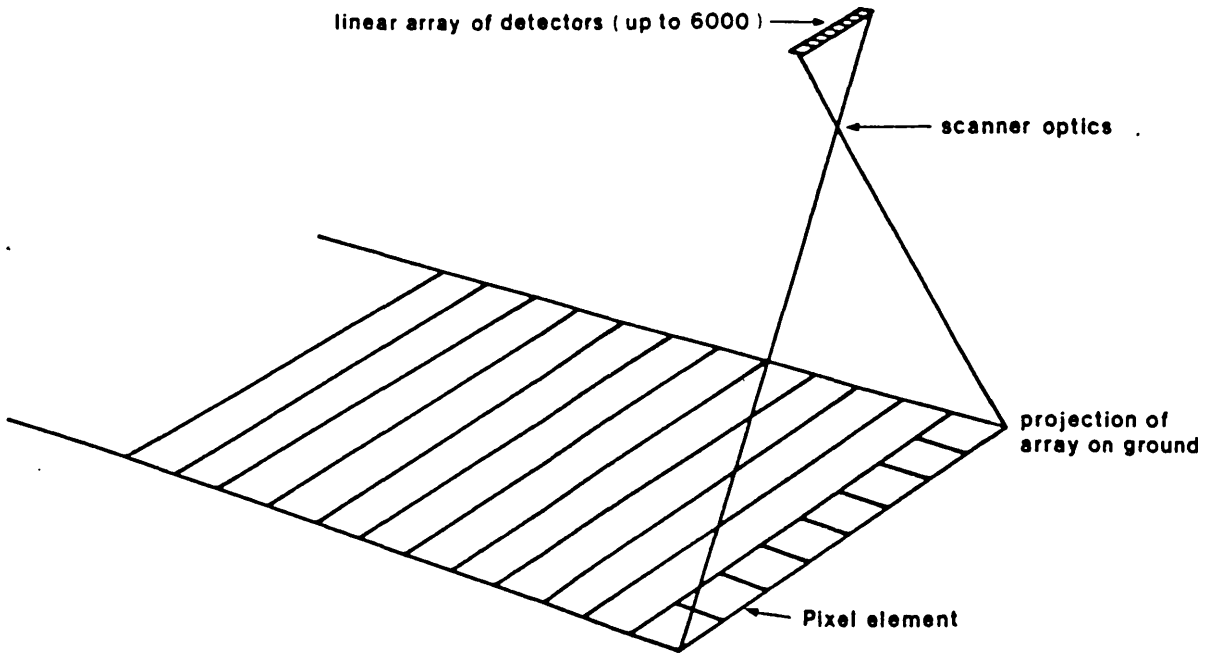


Figure 6.18 Pushbroom design of scanner.

will be a series of four satellites which will provide coverage until the turn of the century.

The sensors on board the satellite consist of two high resolution visible (HRV) pushbroom scanners. The HRV instrument is designed to operate in either a panchromatic or a multispectral mode. In the former case, the IFOV of the sensor produces a 10-m pixel, whereas in the latter case the IFOV produces a 20-m pixel. Figure 6.19 illustrates a typical panchromatic image.

The two HRV instruments are designed to be operated in either a vertical (nadir) or an oblique (off-nadir) position. Nadir pointing enables a swath of up to 117 km to be imaged (Fig. 6.20), whilst off-nadir viewing enables areas up to 425 km away from the satellite ground track to be imaged (Fig. 6.21). The ability to obtain coverage in this manner provides much greater flexibility in data acquisition, and enables the time interval between successive views of the same area on the ground to be reduced from 26 days to 2–3 days, assuming cloud-free conditions. This ability to point the sensor up to 27° off nadir also enables a range of scenes to be acquired by programming the sequence of sensor pointings from the ground (Fig. 6.22).

One of the most interesting features of SPOT imagery from a surveying and engineering point of view is the ability to provide stereoscopic coverage by using the lateral overlap between successive runs, as illustrated by Fig. 6.23. The use of stereoscopic imagery for the production of digital terrain models is discussed in section 6.5.1.

SPOT data are available to users in several different forms, ranging from 1:400 000 photographic negatives to geometrically corrected digital data. Figure 6.24 summarizes the different levels of correction which can be applied to the raw data. The operation and distribution of the satellite data is controlled by SPOT Image, supported by local distributors throughout the world. In the United Kingdom, for example, distribution rights are shared between the National Remote Sensing Centre (NRSC) and Nigel Press Associates (NPA).



Figure 6.19 SPOT—panchromatic image of part of Montreal. Courtesy SPOT Image.

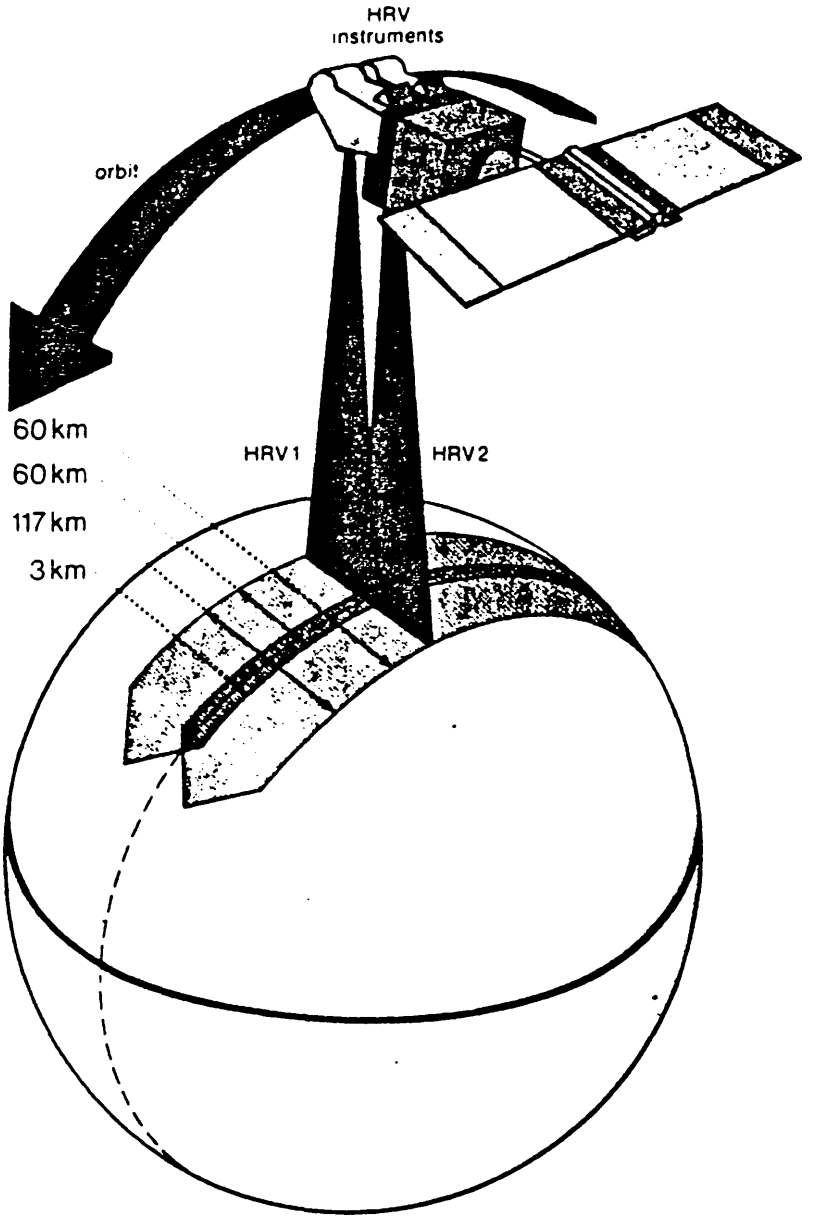


Figure 6.20 SPOT—Swath width of scanner in nadir pointing mode of operation. Courtesy SPOT Image.

6.3.4.2 *Modular optoelectronic multispectral scanner (MOMS).* MOMS is a further example of a satellite-based pushbroom scanner (Table 6.3). The sensors were designed by the West German MBB Company for the German Aerospace Research Establishment (DFVLR). Unlike SPOT, it is not an operational system, and was flown on board the Space Shuttle in June 1983 as a development system. The scanner design is more complex than for SPOT (Fig. 6.25), and consists of four linear arrays of 1728 pixels which enable a continuous line of 6912 pixels to be scanned using a dual-lens optical arrangement. Operating at the altitude of the Space Shuttle, this was equivalent to a ground pixel of some 20 m. A second-generation MOMS sensor currently under development will not only have an extended range of spectral bands but will also enable stereoscopic imagery to be obtained.

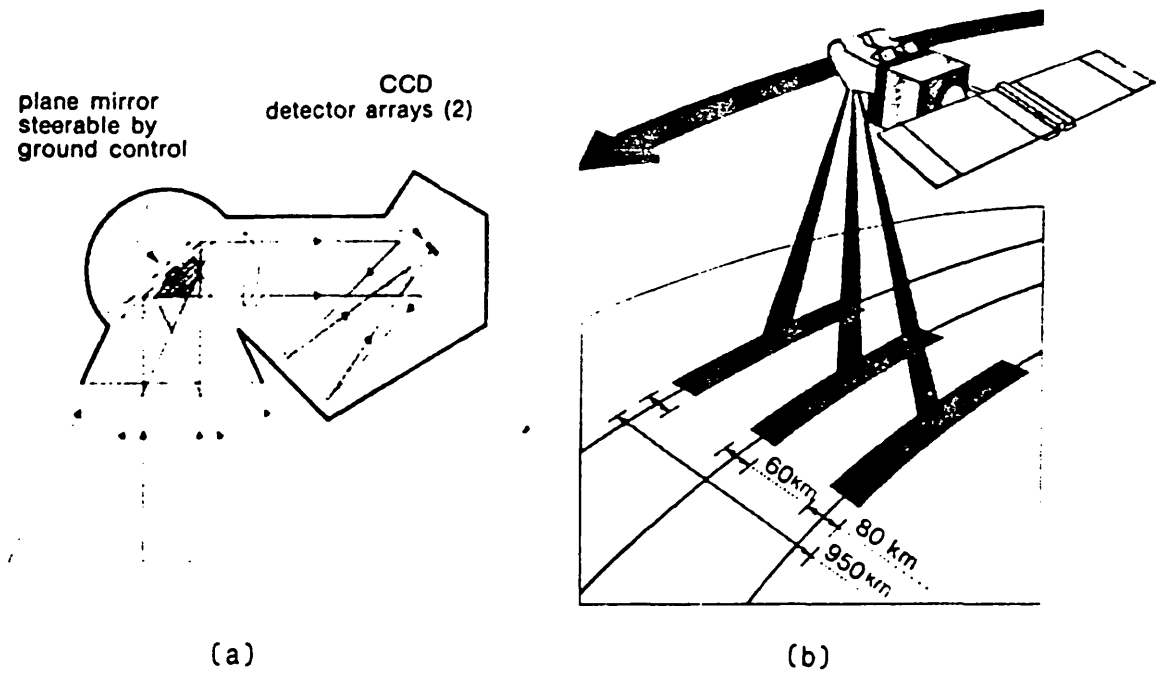


Figure 6.21 SPOT—(*a*) steerable mirror for off-nadir viewing; (*b*) ground coverage in off-nadir mode of operation. Courtesy SPOT Image.

6.3.4.3 *Pushbroom imaging spectrometers.* The requirement to be able to measure the many small variations in the spectral response of surface materials (particularly minerals) has been one of the primary factors leading to the development of very narrow waveband imaging systems. Such devices are similar in many respects to MSS. However, whereas in an MSS the spectrum of a surface material is recorded by a relatively small number of widely spaced samples, an imaging spectrometer records a much more detailed spectral pattern by simultaneously recording up to several hundred contiguous spectral measurements (Fig. 6.27). The primary application of such instruments is in the field of mineral exploration (Goetz *et al.*, 1985) but other applications include the assessment of damage to vegetation and forestry, as caused for example by acid rain.

The *Moniteq Programmable Multispectral Imager* (*PMI*) operates in a pushbroom mode, wide coverage being obtained by five registered units which enable a region up to 1925 pixels wide to be imaged. The instrument can be configured to specific applications by varying its spectral and spatial resolution. Figure 6.26 illustrates the variations which can be accommodated. Thus up to 288 narrow waveband spectral channels can be utilized by use of the transmission grating spectrometer.

The concept of the imaging spectrometer can also be utilized on board a spacecraft, and such a device, the *High Resolution Imaging Spectrometer* (HIRIS), is currently under development by the NASA Jet Propulsion Laboratory (JPL). It is planned that such an instrument could be operated on board the Space Shuttle in the early 1990s. A prototype airborne system, known as the Airborne Imaging Spectrometer (AIS), is

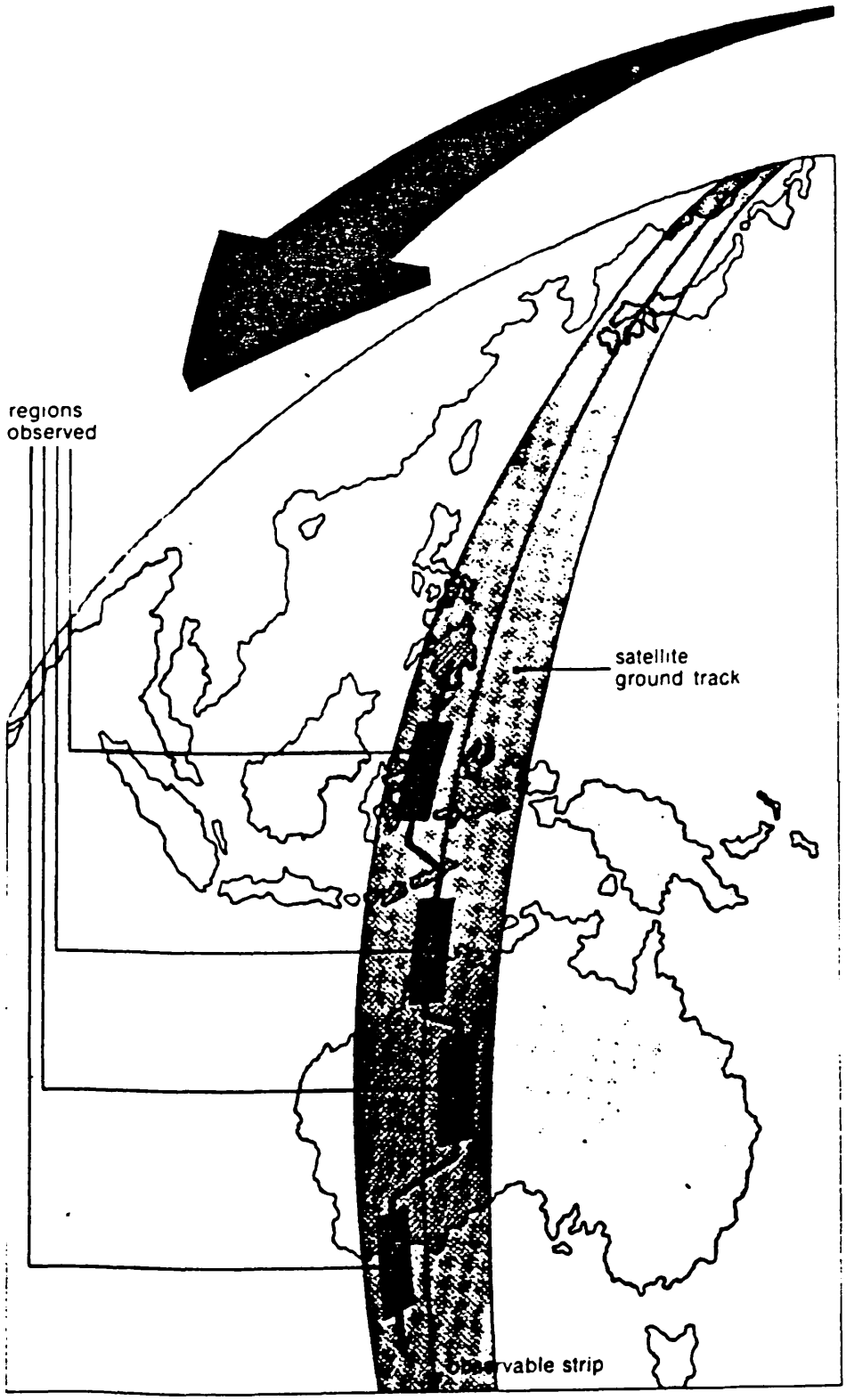


Figure 6.22 SPOT—programmed sensor pointing. Courtesy SPOT Image.

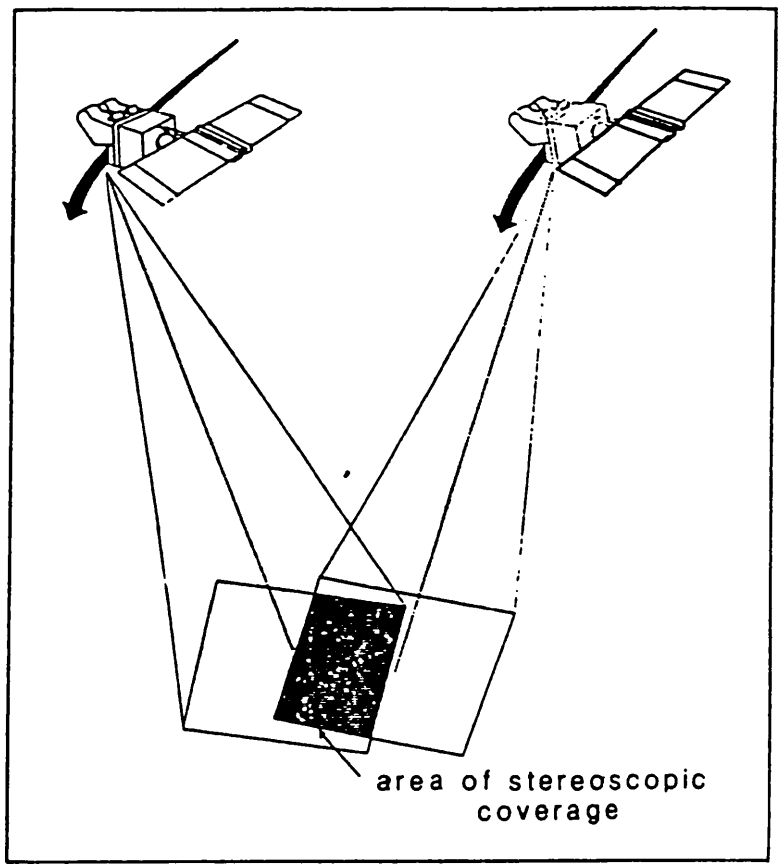


Figure 6.23 SPOT–Geometry of a typical pair of lateral stereoscopic images.

currently being tested. At present the instrument has a spectral range between 1.2 and 2.4 μm in 128 contiguous wavebands.

6.3.5 *Microwave sensors*

Both passive (sensing naturally-occurring radiation) and active (sensing reflected, system-generated radiation) designs of microwave sensor have been developed. Passive scanning microwave radiometers are not, however commonly used for engineering applications at present because of the difficulty of measuring naturally-occurring microwave radiation. Of much greater importance are the active microwave sensors such as side-looking radar (SLR).

6.3.5.1 *Side-looking radar (SLR).* Side-looking radar was initially developed for military purposes during the early 1950s, although it was some 20 years later before it became available for civilian use. The main attractions of SLR were, firstly, the sensor's ability to produce imagery both during the day and at night and, secondly, the ability to acquire imagery of the terrian through haze, smoke, cloud and rain. Two distinct designs of SLR are used; these are referred to as *real aperture radar* and *synthetic-aperture radar.*

The principle of operation of a typical real-aperture radar system is illustrated by Fig. 6.28. It can be seen that the sensor measures the time interval between the transmission and reflection of a microwave pulse from objects in the terrain. Since the reflections from an individual microwave pulse are recorded sequentially, this creates, along a single line, a microwave image of the terrain. Successive strips of imagery are

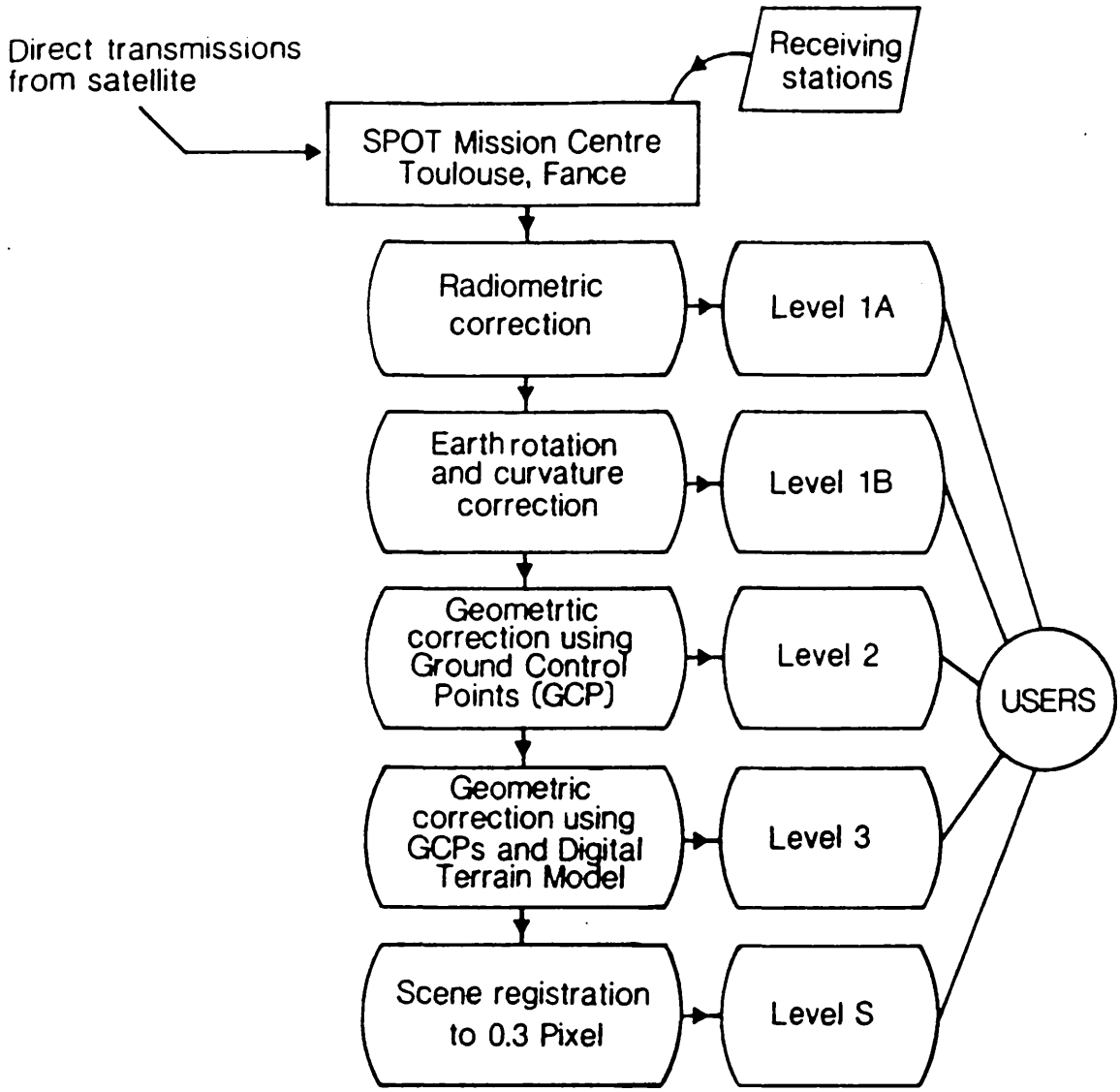


Figure 6.24 Forms of SPOT data illustrating different levels of correction applied to various products.

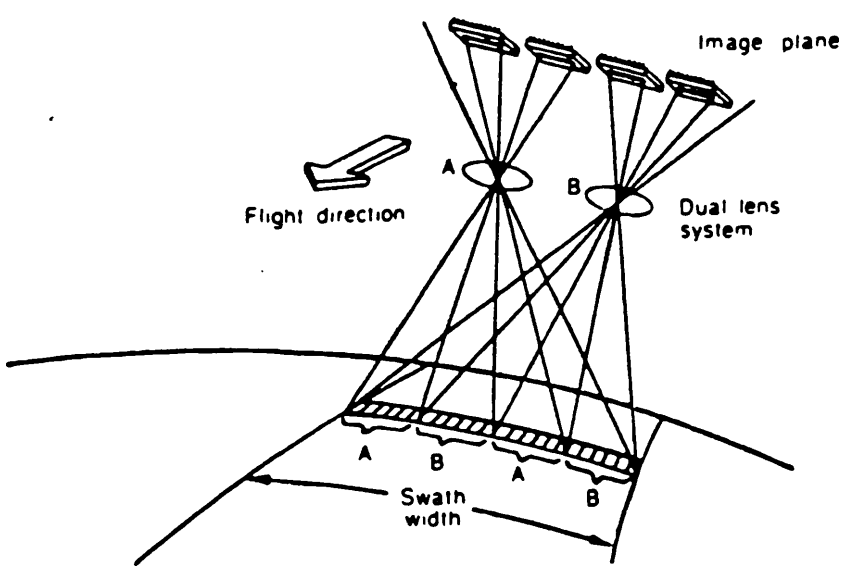
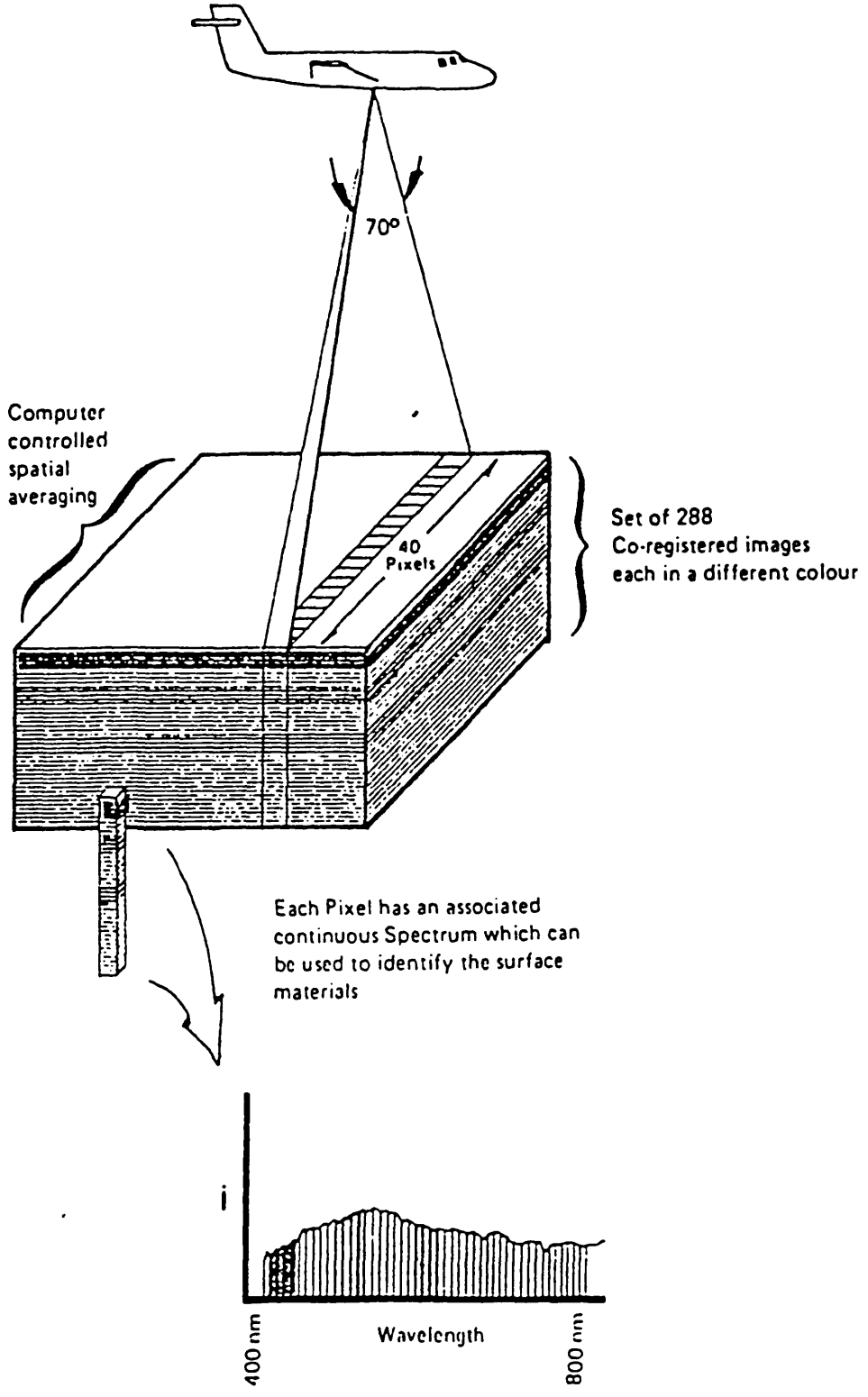
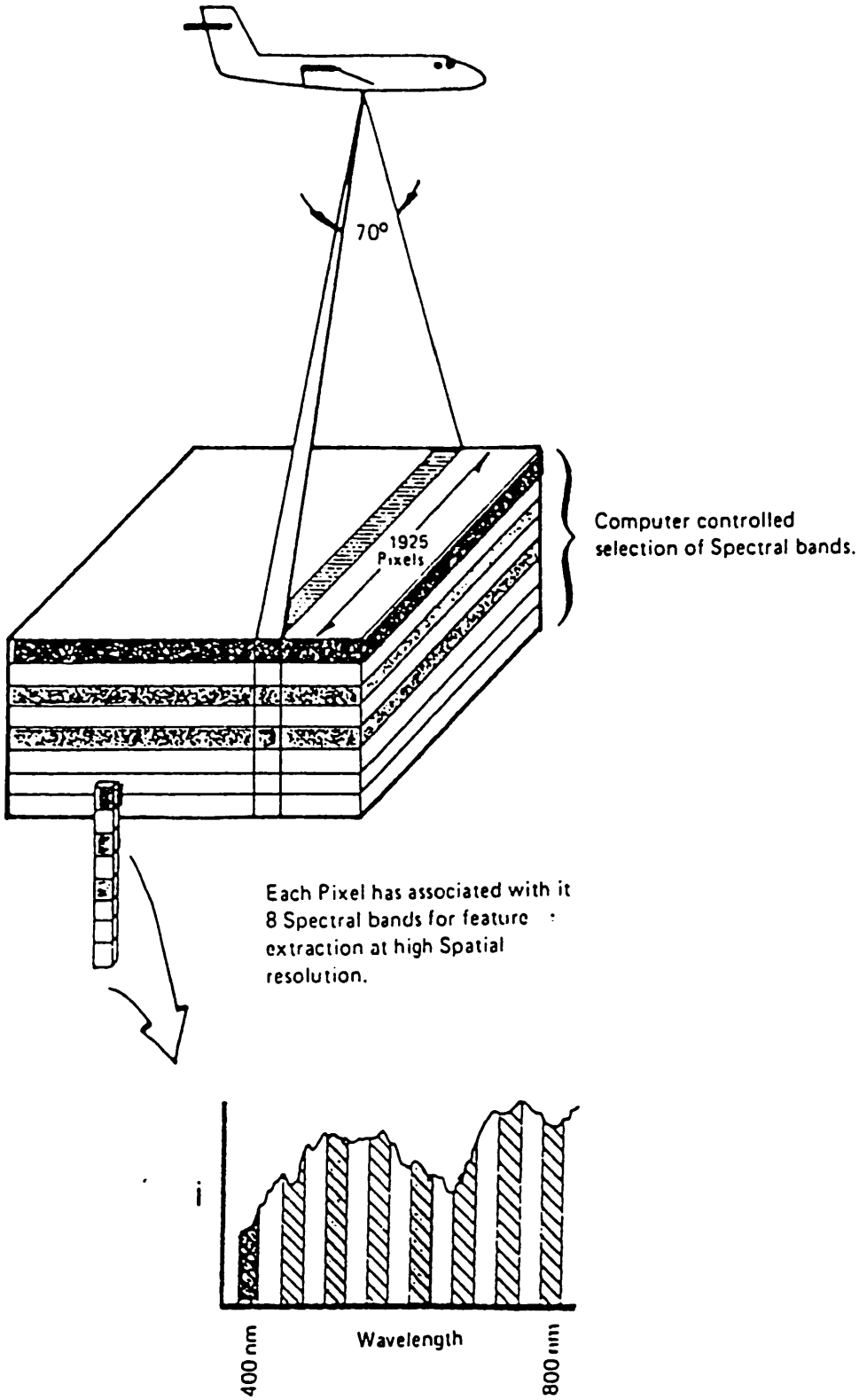


Figure 6.25 Optical and geometrical arrangement of MOMS scanner. Courtesy DFVLR.



(a)

Figure 6.26 Moniteq PMI: (*a*) full spectral mode; (*b*) full spatial mode. *i*, intensity.


70°
1925
Pixels
Computer controlled
selection of Spectral bands.
Each Pixel has associated with it
8 Spectral bands for feature
extraction at high Spatial
resolution.
i
Wavelength
400 nm
800 nm

(b)

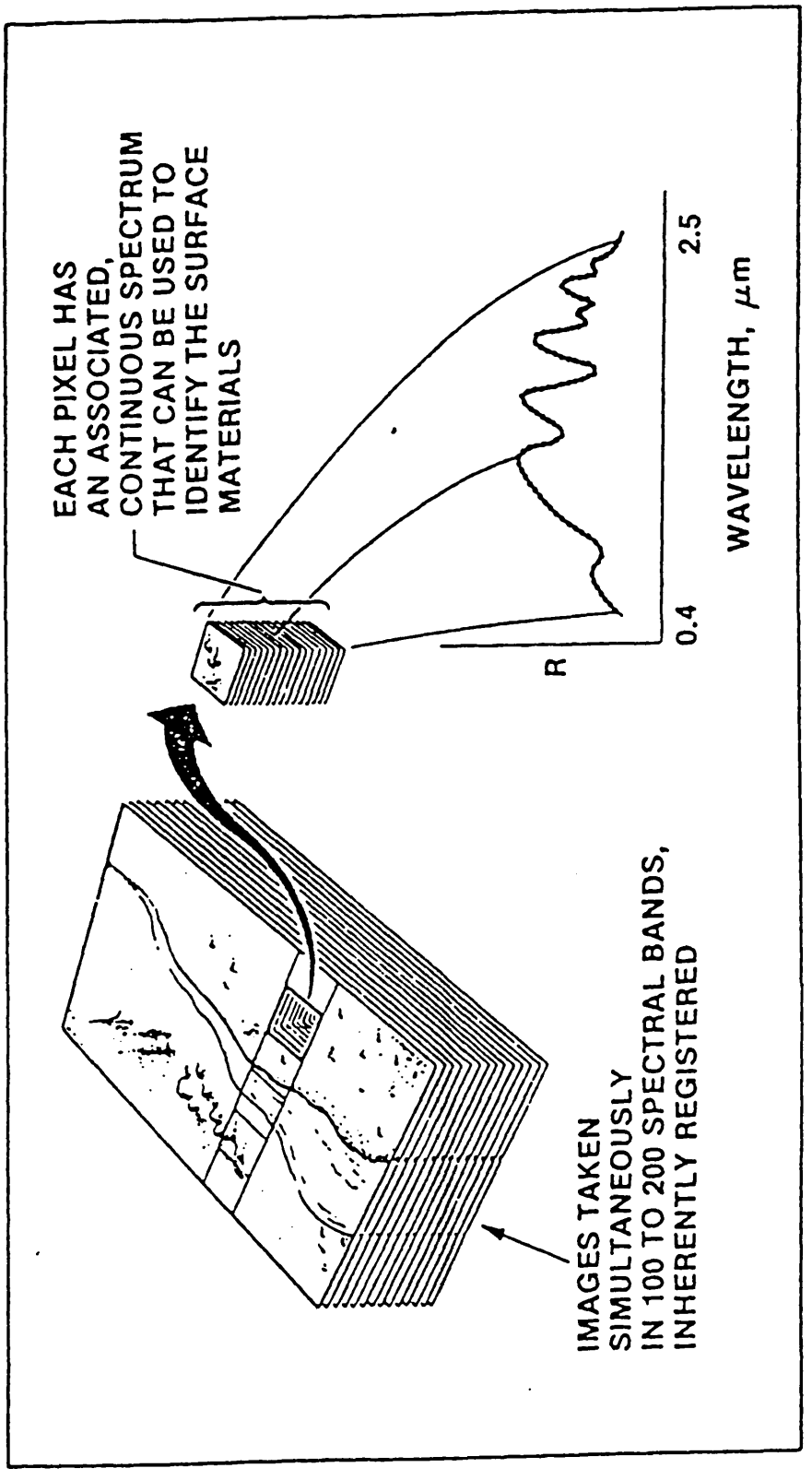


Figure 6.27 Principle of airborne imaging spectrometry.

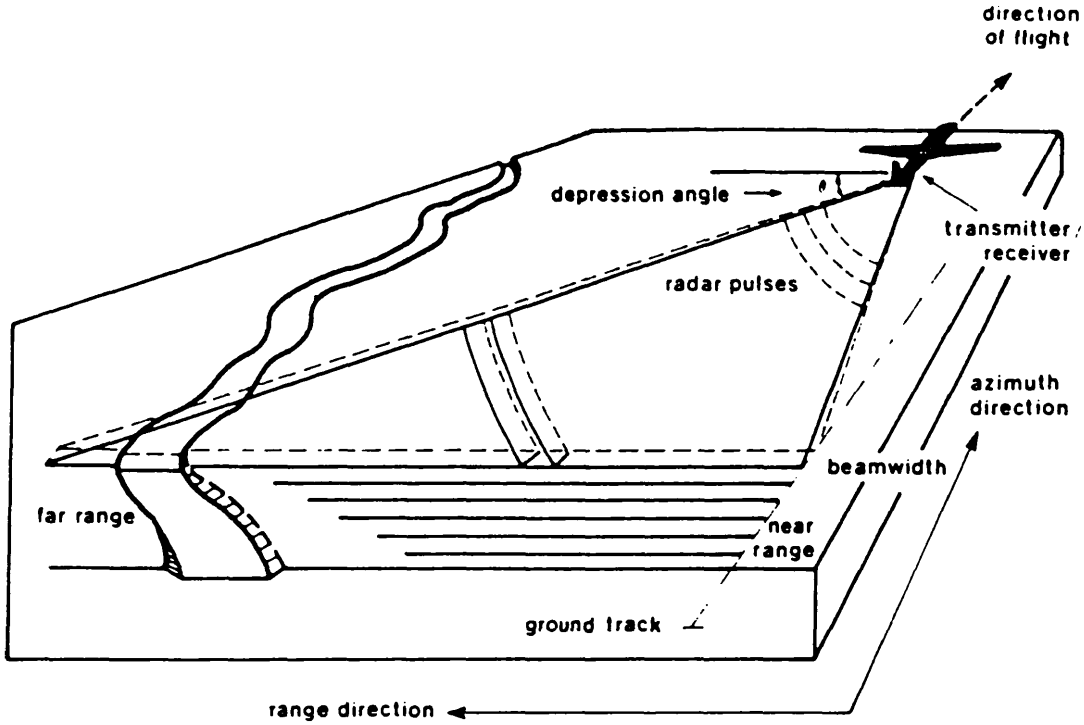


Figure 6.28 Principle of operation of a side-looking radar system.

then produced by the forward motion of the aircraft, thus creating, line by line, a continuous microwave image. Figure 6.29 is an example image produced from a side looking airborne radar (SLAR) system.

The image which is produced by a SLR system is very different from that produced by a conventional optical system or passive microwave radiometer. The view obtained is a record of the range to objects on the terrain and is based on the Earth's reflective properties at microwave wavelengths (8–300 mm). Consequently, the nature and intensity of the reflections are influenced by factors such as ground conductivity and surface roughness, which are much less significant at shorter wavelengths.

The spatial resolution of a real-aperture system is determined by two factors—the pulse length and the antenna beamwidth. The first of these factors determines the resolution in the range direction, whereas the second influences the discrimination between features in the direction of flight, in the azimuth direction (Fig. 6.28). A critical consideration is the antenna length, since this directly influences the beamwidth of the sensor and thus controls the resolving power of the sensor in the azimuth direction. The resolution of a real-aperture radar is consequently limited by the size of antenna which can be fixed to the side of the aerial platform. Although this restriction can be a serious deficiency when high-resolution imagery is required, these systems nevertheless have the ability to produce imagery in real time, without the need for sophisticated optical or computer processing. This has been a particularly important factor in the development of such systems for military use. The limitations in the resolution of real-aperture radars can, however, be overcome by using suitable signal processing techniques. Sensors of this design are referred to as synthetic-aperture radar (SAR) systems.

A detailed description of the principles of SAR is beyond the scope of this chapter, and readers should refer to Jensen *et al.* (1977) and Trevett (1986) for further

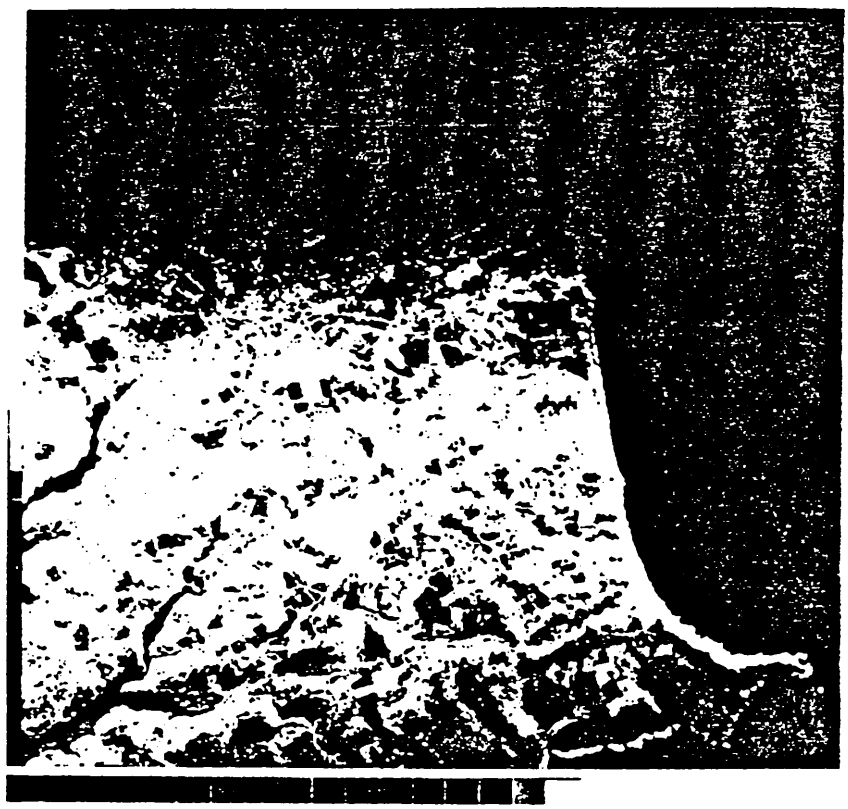


Figure 6.29 Typical SLAR image created by the SAR-580 system.

information on the technique. In very simple terms, the antenna used in SAR systems is synthesized electronically in order to increase its effective length. Thus a relatively short antenna, mounted on an airborne or spaceborne platform, is able to act as if it were a much longer antenna, with a consequent increase in resolving power in the azimuth direction. This improvement is achieved by processing not only the time interval between the outgoing and reflected signals but also the change in frequency (the Doppler shift) between the two signals. Figure 6.30 illustrates how this technique was used on board the Shuttle Imaging Radar-A (SIR-A) sensor.

Several airborne SAR systems have also been flown, and such systems have been used to generate imagery over extensive parts of South America (Von Roessel and De Godoy,, 1974), West Africa and Indonesia. The European Space Agency (ESA) sponsored the SAR-580 campaign providing the most recent airborne SAR imagery in the United Kingdom (Trevett, 1985). The campaign also provided valuable information for the design of the forthcoming European Radar Satellite (ERS-1). Up to 1988, three SAR sensors have been operated, for civilian use, from satellite platforms, with at least a further two planned for the near future. Table 6.6 summarizes the principal characteristics of these systems.

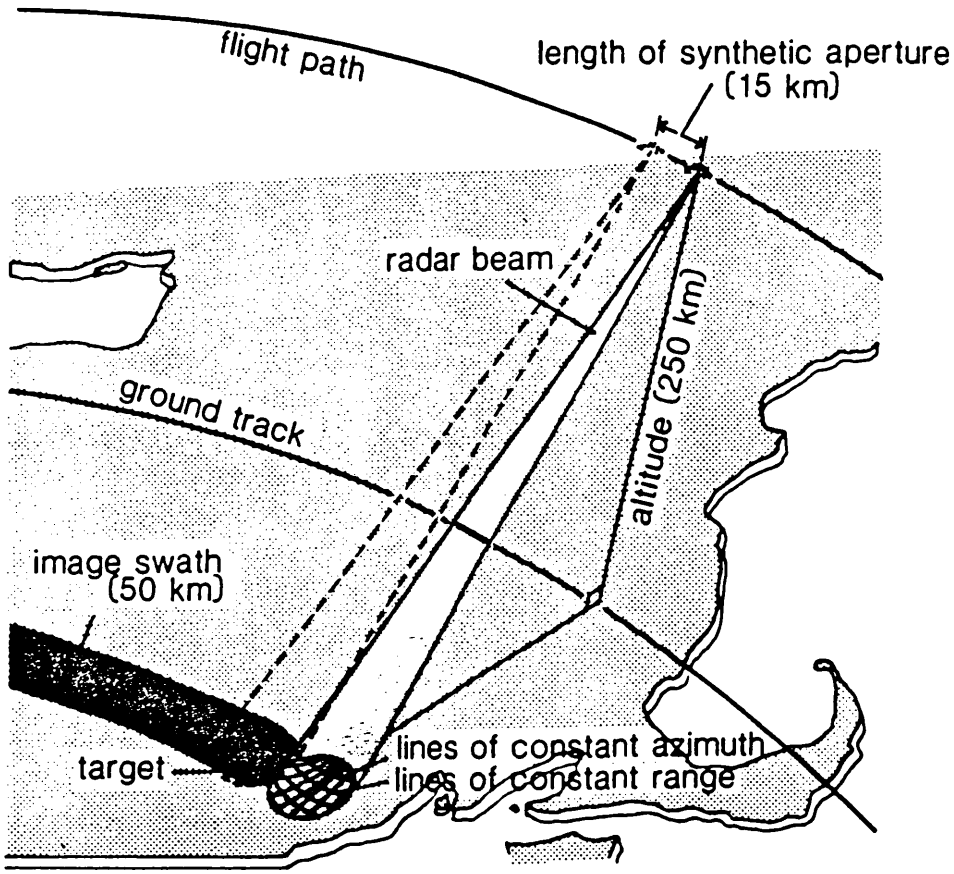


Figure 6.30 Shuttle imaging radar (SIR)–a synthetic aperture radar (Elachi, 1982).

Table 6.6 Characteristics of a selection of satellite-based SAR systems

	Past			Future	
	Seasat	Shuttle Imaging Radar (SIR)-A	SIR-B	European Radar Satellite (ERS-1)	Radarsat
Date of launch	1978	1981	1984	1989	1991
Operated by	NASA	NASA	NASA	European Space Agency	Canada
Altitude (km)	800	245	255, 274, 352	675	1000
Wavelength (mm)	230	230	540	540	540
Look angle (°)	20	50	57	—	30–45°
Resolution (m)	25	40	30	25	30
Swath width (km)	100	50	20–25	80	150
Further reading	Elachi (1980)	Elachi (1982)	—	Braun and Velten (1986)	Annett (1985)

It is important, however, to appreciate the limitations of SLR. These difficulties arise from two principal sources: background clutter on the image which can make interpretation quite difficult, and the geometry of the 'illuminating' signal relative to the terrain. The latter may lead to difficulties in the detection of linear features such as roads. Further details of the effects of these difficulties for topographic mapping can be found in section 6.5.1. In spite of these limitations, SLR sensors are important, particularly for monitoring in cloud-covered tropical regions and also in temperate and northern latitudes. For the future, as the technology for digitally processing SAR imagery improves, it is likely that such data will have a much more significant role to play in operational remote sensing activities than at present.

6.4 Digital image processing (DIP)

Digital image processing is central to the effective use of modern remote sensing data. Although a limited amount of information can be gleaned from the interpretation of a photographic image produced from Landsat imagery, for example, the full potential of the data can only be realized if the original data, edited and enhanced to remove systematic errors, are used with a suitably programmed image processing system.

6.4.1 *Hardware*

Until relatively recently, most DIP of remote-sensing data was carried out on a large expensive dedicated system such as that illustrated in Fig. 6.31. This enables data to be read into the system from a magnetic tape reader, to be displayed on a colour TV monitor and, after suitable processing, to be output to a colour filmwriter for the production of high-quality imagery. Typical of the modern dedicated image processing systems, which also incorporate local microprocessor support for image display and arithmetic processing, is the GEMSYS 35 system manufactured by GEMS of Cambridge (Fig. 6.32).

A trend in recent years has been the move towards smaller and cheaper image processing based on the new generation of 32-bit superminicomputers and 16-bit IBM-compatible personal computers (Fearns, 1984; Tait, 1987). Table 6.7 lists a selection of the remote sensing systems currently available.

6.4.2 *Software*

The range of algorithms which are used to restore and classify remote-sensing data is extremely wide and varied. Furthermore an appreciation of the choice of the most appropriate technique, and the advantages and limitations of each technique, is a subject of some complexity and beyond the scope of this chapter. Consequently, only a very brief description of the most commonly used techniques will be provided. For further details, Bernstein (1978) and Bagot (1985) should be consulted.

6.4.2.1 *Image restoration.* Several preliminary transformations are normally carried out on raw satellite data to reduce the effects of Earth rotation and curvature and variations in the attitude of the satellite. If sufficient ground control points are available, the imagery can be fitted to the control using least-squares techniques. This procedure is also necessary if the remote-sensing data are to be combined with existing map data. Geometric preprocessing may also be required for airborne imagery, for example, if multitemporal sources of data are being compared.

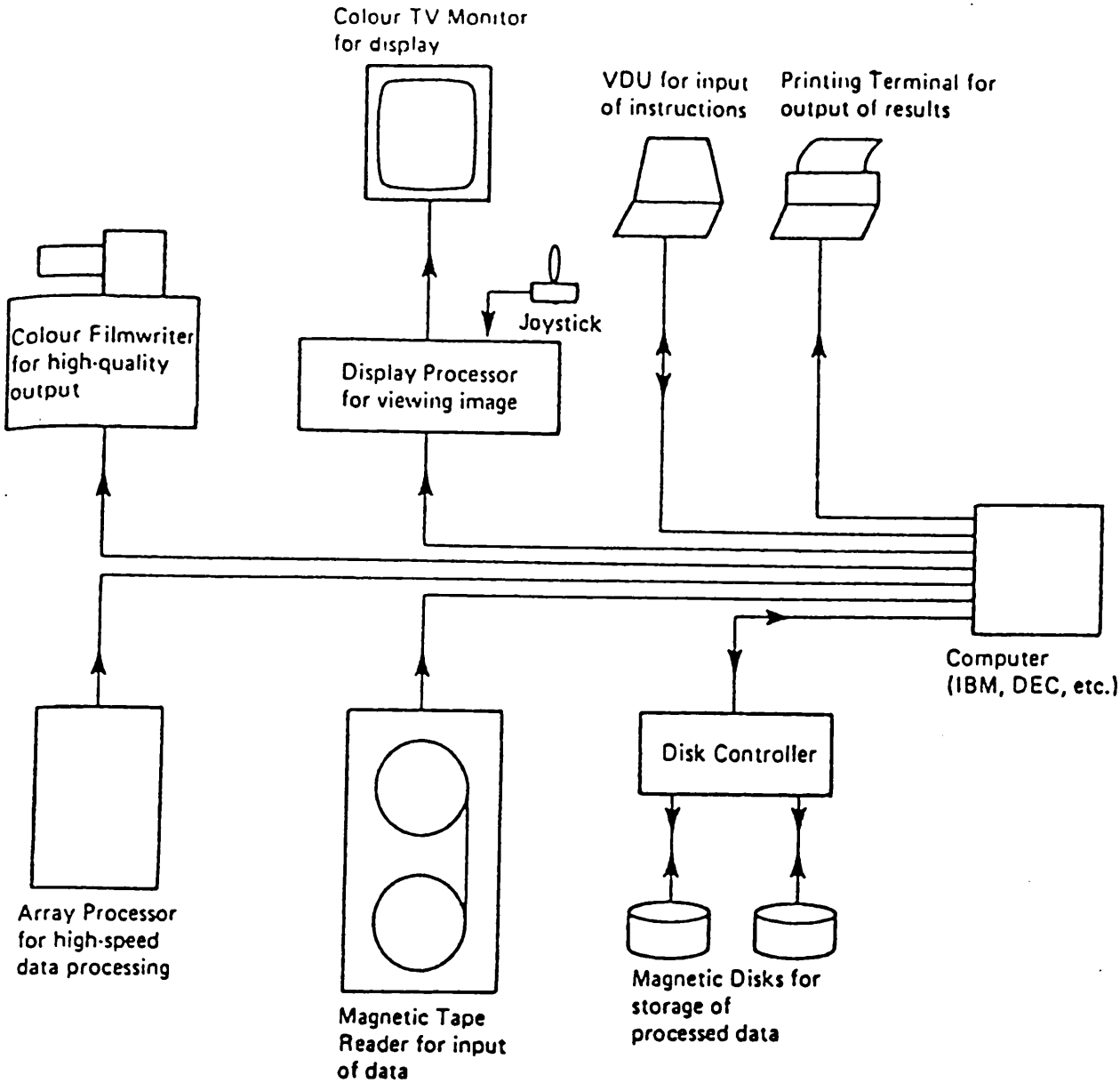


Figure 6.31 Components of a typical digital image processing system.

Variations in the output from the detectors used in an MSS may cause a striping effect on the final image. Such an effect can be eliminated by a process termed 'destriping'. In some cases, complete scan lines may be missing from the data, and again preprocessing can be carried out to overcome this type of problem. Radiometric preprocessing to eliminate the effects of atmospheric variations may also be necessary, particularly with airborne imagery.

6.4.2.2 *Image enhancement. Contrast stretching* is a technique which is used to brighten an image and produce one which allows the full potential range of the output device to be realized. It is often carried out automatically, and is generally one of the first operations performed by a user when viewing a new image. A range of different techniques can be used, and the choice generally depends on the distribution of the grey values or digital numbers within the scene. Figure 6.33, for example, illustrates one particular intensity histogram distribution.



Figure 6.32 GEMSYS 35 image processing system. Courtesy GEMS of Cambridge.

Table 6.7 A selection of remote sensing systems currently available

	Name of system	Company
Dedicated image processing system	Gemsys 33	GEMS, Cambridge, UK
	I^2S Model 75	International Imaging Systems, California, USA
	Vicom VDP Series	Vicom Systems, San Jose, USA
	Dipix Aries II	Dipix Systems, Ottawa, Canada
Personal computer-based systems	Diad Systems	Diad Systems, Edenbridge, Kent, UK
	LS10	CW Controls, Southport, Lancashire, UK
	Microimage	Terra Mar, California, USA
	ERDAS	Earth Resources Data Analysis System, Atlanta, USA
	ImaVision	RBA Associates, Ottawa, Canada

Two distinct forms of contrast stretching are normally used; the linear (single or two-part) and the more complex non-linear (e.g. normalization or gaussian). For example, Fig. 6.34 illustrates the effect of a simple linear stretch on the data shown in Fig. 6.33. This graph represents the 'transfer function' between the input and output digital numbers. If this transfer function is used on the original input data distribution, the effect would be to stretch the input range of 30 to 180 so that the output range varied linearly between 0 and 255.

Spatial filtering is carried out in order to either enhance boundaries between features (edge enhancement or high-pass filtering) or to smooth and eliminate boundaries

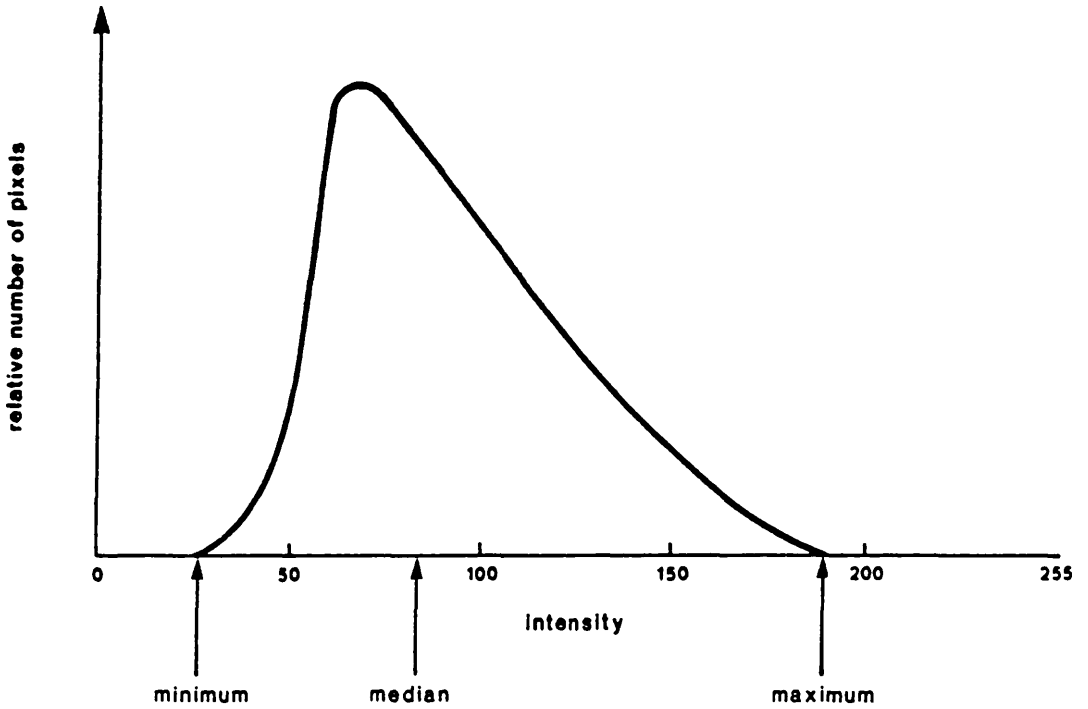


Figure 6.33 Pixel intensity histogram (adapted from Bagot, 1985).

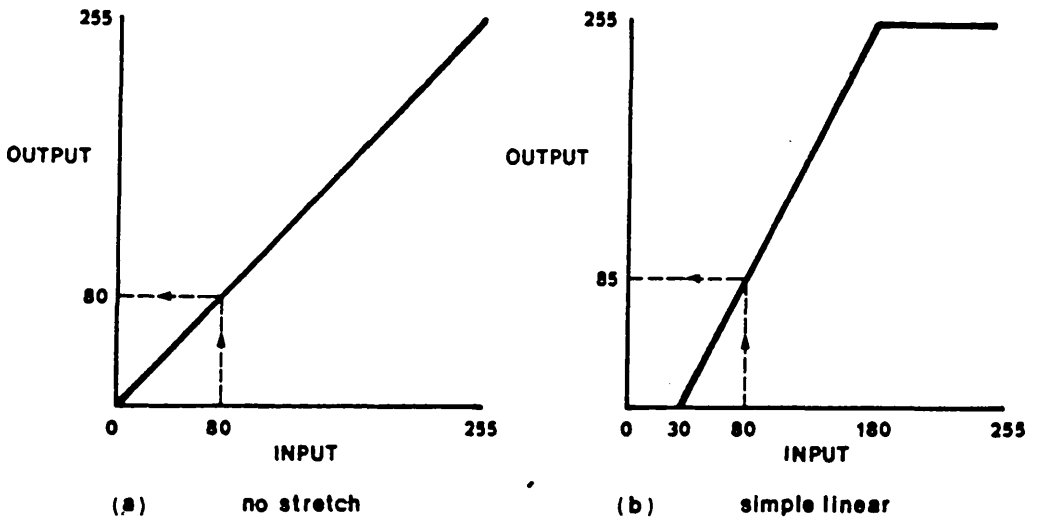


Figure 6.34 Contrast stretch-transfer function for simple linear stretch using data shown in Fig. 6.33.

(smoothing or low-pass filtering). The former approach is most suitable for enhancing geological features, particularly lineaments. The latter may be used to reduce random noise in an image, and Fig. 6.35 illustrates the concept.

Image ratioing is the process of dividing the radiance values of each pixel in one waveband by the equivalent values in another waveband. Ratioed images reduce radiance variations caused by local topographic effects, and consequently may be used to enhance subtle spectral variations between surface materials that are difficult to detect on single-band images. Figure 6.36 illustrates the concept in more detail. In this case the DN values for the same terrain class can be seen to vary, depending on the

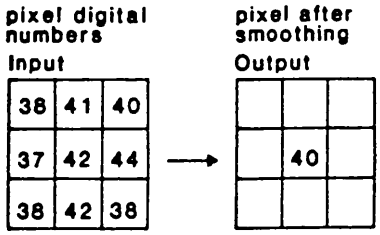
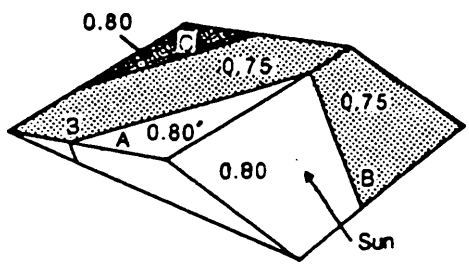


Figure 6.35 Concept of 3 × 3 low pass (smoothing) spatial filter. The central pixel value (40) is the mean of the nine surrounding values.



	Band 1 DN	Band 2 DN	Band ratio 1:2
Slope facing Sun			
Unit A	20	25	0.80
Unit B	30	40	0.75
Slope facing away from Sun			
Unit A	16	20	0.80
Unit B	24	32	0.75
Unit C	32	40	0.80

Figure 6.36 Use of image ratioing to eliminate systematic effects caused by varying illumination conditions (adapted from Taranik, 1978).

illumination conditions (sunlit or shaded slopes). The ratioed values remove this systematic bias in the original data.

Principal components analysis is a statistical technique which is used to improve the discrimination between similar types of ground cover. It is based on the formation of a series of new orthogonal axes which reduce the correlation existing between successive wavebands in a multispectral data set.

For example, Fig. 6.37 is a scatter plot showing the radiance values for two features *A* and *B* in two different wavebands *X* and *Y*. In the original wavebands (Fig. 6.37*a*) the variation between features *A* and *B* is very small in comparison to the original range of the data. Principal components analysis operates by generating a new set of axes along the lines of maximum variance; this helps exaggerate the difference between features such as *A* and *B*. Thus, in Fig. 6.37*b*, although the data range has been compressed, the variation between features *A* and *B* is now almost half the total range of the data and consequently eases the identification of these feature classes.

6.4.2.3 *Image classification. Density slicing* is the simplest form of automated classification. It enables a single band (generally thermal IR data) to be colour-coded according to the grey level of the individual pixels. Thus, for example, pixel values 0–20 may be coloured blue, 20–25 red and so on. This not only aids interpretation, but if the density slices are related to temperature levels, the resulting image can be used to quantitatively assess the variations in temperature across the scene (Fig. 6.17).

Supervised methods can be used in cases where ground data exists about the terrain

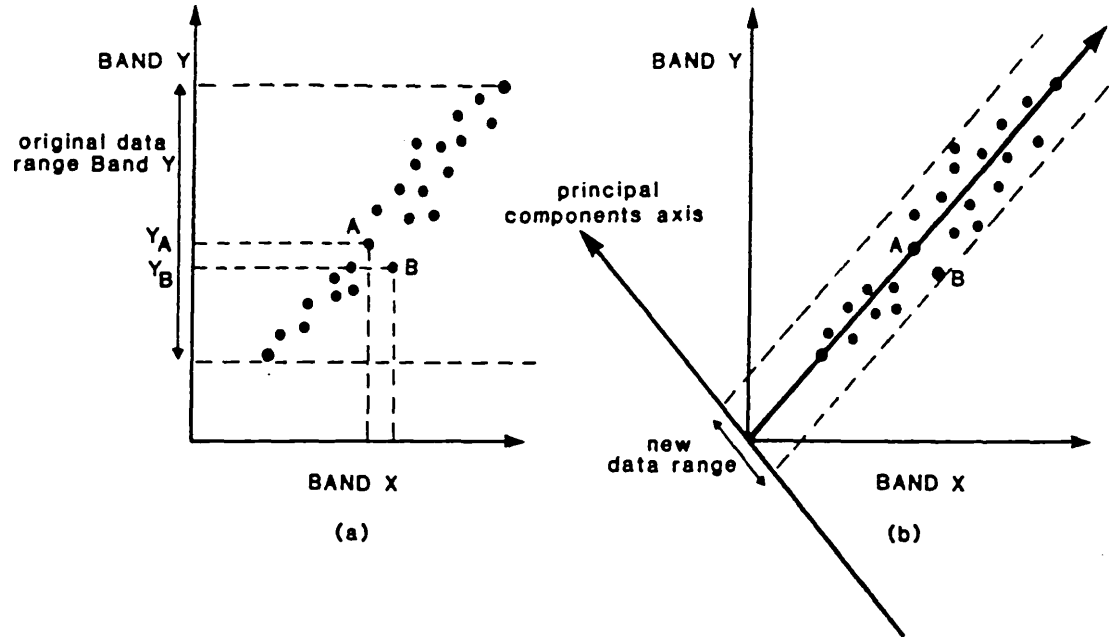


Figure 6.37 Graphical illustration of principal components analysis: (*a*) original wavebands; (*b*) new principal components wavebands.

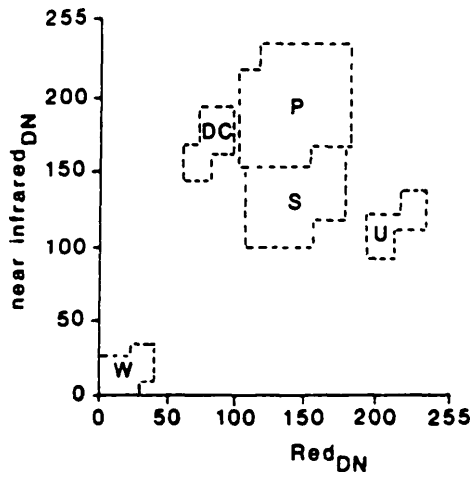


Figure 6.38 Supervised classification–parallelopiped or box classification of five land cover types (Curran, 1986). *W*, water, *DC*, deciduous woodland; *P*, pasture; *S*, bare soil; *U*, urban areas.

under investigation. It is possible to use this data to 'train' the computer to perform a simplified form of automated interpretation. For example, assume a particular group of pixels are known to represent an area of deciduous woodland. By examining the maximum and minimum numerical values of the radiances in each of the spectral bands being used (7 for Landsat TM, although only the least correlated bands would be used in practice) a 'spectral pattern' for deciduous woodland could be defined. If the computer then compared this multidimensional spectral pattern with the spectral pattern of every pixel in the scene, then it would be possible to display only those areas which satisfied the criteria set by the training sample. Consequently, all deciduous woodland over a large area could be automatically determined.

Generally, several different land cover types would be analysed and classified using this technique. Figure 6.38 illustrates a hypothetical set of data and the classifier which

could be used to discriminate each of the five spectral classes. The process outlined above refers to the simplest form of classifier, the 'box' or parallelopiped classifier. Other more sophisticated techniques, such as maximum likelihood, can also be used.

In *unsupervised classification* of multispectral data, no attempt is made to 'train' the computer to interpret common features. Instead, the computer analyses the pixels in the scene and divides them into a series of spectrally distinct classes based on the natural clusters which occur when the radiances in n bands are compared. The user then analyses other evidence to determine the nature of each of the distinctive classes which have been identified by the computer. Density slicing, for example, is a simple form of unsupervised classification. Unsupervised classification techniques are generally more expensive in terms of processing requirements than supervised methods.

6.5 Applications of remote sensing

The applications of remote sensing cover a wide range of scientific and engineering disciplines and many of these are beyond the scope of this chapter. Consequently two general topics relevant to the surveyor and engineer will be discussed. The first relates to the use of satellite imagery for topographic mapping and the second to the use of airborne and satellite imagery to generate thematic mapping information appropriate to engineering projects.

6.5.1 *Small-scale topographic mapping from satellite sensors*

The creation of topographic maps is based almost totally on the use of remotely sensed data. Only at very large scales can ground surveying compete economically. The majority of maps created in this manner are based on aerial photography obtained using a high precision mapping camera and measured using a photogrammetric stereoplotter. The principles of these techniques are well known and considered in detail by Burnside (1982) and Wolf (1983), and to a more limited extent in Chapters 7 and 8.

In recent years as the spatial resolution of satellite sensors has improved, the potential of spaceborne cameras, scanners and radar systems for small-scale topographic mapping has been considered. This section examines the progress which has been achieved in this field.

6.5.1.1 *Satellite sensors suitable for topographic mapping.* A wide range of satellite based sensors has been considered as potentially suitable for cartographic applications. Three distinct groups of sensors can be identified; photographic sensors, scanning sensors and synthetic aperture radar.

Several limited-term missions have been flown to examine the capability of satellite-based photographic cameras for mapping; examples include the Skylab Itek camera, and more recently the Spacelab Zeiss RMK 'metric camera' and the Itek Large Format Camera (LFC).

Unlike photographic sensors, the scanning systems are generally housed on board Earth resources satellites providing repeat coverage of the terrain. The most significant sensors in this category are Landsat MSS, Landsat TM and SPOT. Landsat RBV has also been used for mapping, and indeed was primarily designed for this application; however, because of the limited amount of data obtained from the sensor, it will not be considered.

Table 6.8 Requirements for standard map series (Doyle, 1982)

Map scale	Positional accuracy (m)	Elevation accuracy (m)	Typical contour interval
1:1 000 000	300	30	100
1:500 000	150	15	50
1:250 000	75	8	25
1:100 000	30	6	20
1:50 000	15	3	10
1:25 000	8	2	5

The cloud penetration benefits of *synthetic aperture radar* (SAR) have already been discussed, and this is a particularly important feature for space systems imaging in temperate latitudes, such as the UK, or in tropical regions, such as South America and West Africa. Three particularly important systems in this context have been Seasat and the Shuttle Imaging Radars A and B (SIR-A and SIR-B) (Table 6.6). The experience from both of these systems will be invaluable in maximizing the success of future systems such as ERS-1 and Radarsat.

The level of success of all these systems for topographic mapping is dependent on two factors; first, the geometrical accuracy of the data which are obtained, and second, the level of completeness of detail which can be interpreted from the data.

6.5.1.2 *Geometrical accuracy testing.* Before examining in detail the results which have been obtained from the various sensors described above, it is useful to consider the accepted accuracy requirements for various standard map scales. Comparing these with the geometrical accuracy results obtainable from satellite systems will then enable the potential of the different sources of data to be assessed. Table 6.8 (Doyle, 1982) summarizes these specifications.

The procedure for testing the geometric accuracy of satellite imagery normally involves a comparison of the coordinates of a series of well-defined control points on the imagery with their known ground coordinates (from an existing map).

Various mathematical models can be used to perform the transformation from image coordinates to ground coordinates. These range from simple linear conformal or similarity transformation (4 parameters, scale change, rotation and translation in X and Y), to affine transformations (6 parameters, as linear conformal but with different scale changes in X and Y) and polynomial transformations with different numbers of terms (see section 7.3). The choice of transformation depends largely on the type of imagery and the level of pre-processing which has been performed on the imagery.

A summary of the results which have been reported for the three photographic sensors discussed previously is provided in Table 6.9. Comparison of these results with those for the mapping specifications given in Table 6.8 indicates that, geometrically, this type of data could be suitable for planimetric mapping at scales between 1:50 000 and 1:100 000.

Table 6.10 summarizes a selection of the results which have been obtained by applying different transformations to MSS imagery obtained from Landsats 2 and 3. Comparative results using TM imagery of two test sites in the United States are provided by Welch *et al.* (1985), and are illustrated in Fig. 6.39. Similar results were reported by Fusco *et al.* (1985) and Isong (1987) (Table 6.11).

Table 6.9 Characteristics of spaceborne photographic missions

	NASA Skylab S-190B	ESA Metric Camera	NASA Large Format Camera
Focal length (cm)	46	30	30
Format (cm)	12.5 x 12.5	23 x 23	23 x 46
Coverage (km)	110 x 110	190 x 190	225 x 450
Scale	1:950 000	1:820 000	1:1 000 000
Flying height (km)	385	250	300
Planimetric accuracy (m)	20–25	20–30	20–30
Reference	Petrie (1985)	Menuquette (1985)	

Table 6.10 Selected accuracy tests for Landsat MSS imagery

Image source	No. of control points	Type of transformation	Root mean square errors X (m)	Y (m)	Reference
Landsat MSS					
Bulk Image	181	Uncorrected	215	195	Wong (1975)
"	"	Linear conformal (similarity)	94	67	"
"	"	20-term polynomial	37	44	"
Landsat MSS Precision Images	35	Linear conformal (similarity)	64	81	Mohammed (1977)
"	"	Affine	46	52	"
"	"	12-term polynomial	40	49	"

The results obtained in virtually all cases are remarkably good, and indicate that sub-pixel level accuracies can be achieved with relatively low-order processing. The results from Welch *et al.* (1985) also indicate the higher geometrical characteristics of Landsat 5 TM imagery.

These results indicate that, geometrically, Landsat MSS could be suitable for planimetric mapping at scales of up to 1:100 000 and for Landsat TM up to 1:50 000. Preliminary results using simulated SPOT imagery (Welch, 1985), indicate that, geometrically, it could be used for planimetric mapping at scales of up to 1:50 000.

Comprehensive reviews of the geometric and cartographic potential of satellite SAR systems for topographic mapping can be found in Petrie (1978), Dowman and Morris (1982), and Leberl *et al.* (1985).

A summary of the geometrical accuracy results obtained from four differently processed Seasat images of the area is provided in Ali (1986). The results indicate that optically processed Seasat images contain substantial scale errors compared to their digitally processed counterparts. These results also indicate that the effect of using higher-order polynomials is more noticeable on the optically processed images than on those produced digitally. The imagery produced by the German Aerospace Agency (DFVLR) produced the best results during the test process. In both cases, the effect of terrain relief on the accuracy of the results was noticeable.

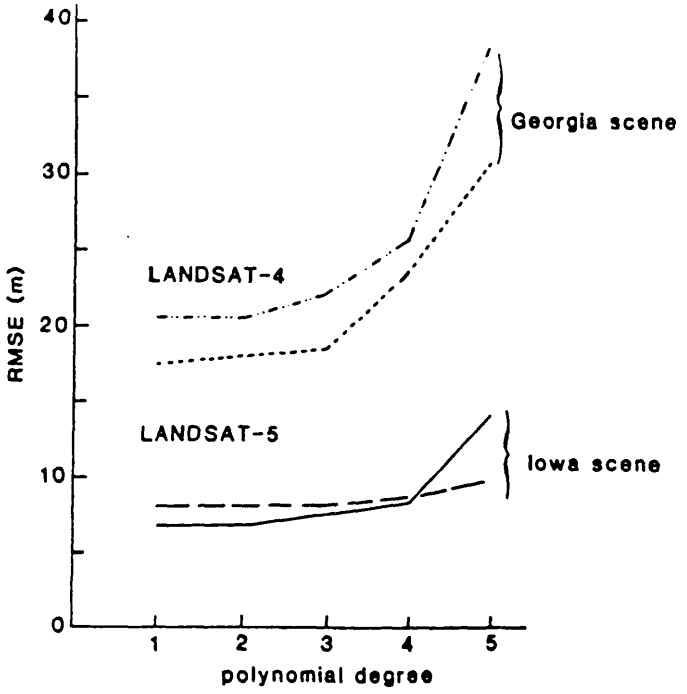


Figure 6.39 Landsat TM: comparison of RMSE values determined at check points for Landsat TM datasets in Georgia and Iowa as a function of the degree of polynomical used for rectification (adapted from Welch *et al.*, 1985).

Table 6.11 *Summary of results of various geometric transformations applied to Landsat Thematic Mapper imagery of an area around London, England (Isong, 1987)*

Transformation used	Terrain coordinates	
	Vector standard errors (m)	
	Control points $n = 20$	Check points $n = 20$
Linear conformal (similarity)	19.81	20.90
Affine	19.60	21.10
4-term polynomial	18.84	20.76
6 " "	18.71	20.49
8 " "	18.78	21.97
10 " "	19.49	22.13

Leberl and Domik (1985) report the results which have been achieved with Seasat and SIR-A imagery using an analytical stereoplotter and the SMART (stereomapping with radar techniques) bundle adjustment software system. Table 6.12 summarizes the results obtained.

6.5.1.3 *Level of completeness.* It should be clearly noted that all the results cited in the previous section refer to mathematical coordinate determinations for a few well selected and clearly identifiable points. These results do not necessarily reflect the planimetric mapping potential of the sensors, since the results do not account for the level of mapping detail which can be interpreted from the imagery. For example, variations in scene contrast in Landsat imagery or changes in radar illumination

Table 6.12 Root mean square (rms) accuracies obtained with an analytical plotter and SMART program using Seasat and SIR-A data. *X* denotes the along-track co-ordinate (Leberl *et al.*, 1985)

	Radar data and systems					
	Seasat (Los Angeles)			SIR-A (Greece)		
Number of ground control points	16	4	2	31	4	3
Number of orientation points	10	0	8	6	0	6
r.m.s. *X*(m)	58	73	72	62	89	86
r.m.s. *Y*(m)	134	174	169	113	132	119

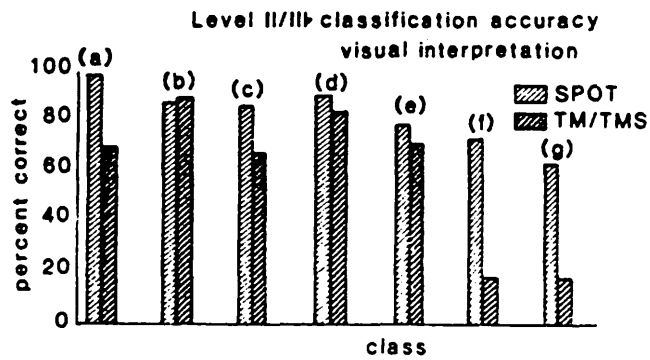


Figure 6.40 Landsat TM and SPOT–percent completeness for selected land cover classes. (*a*) Roads, highways and railways; (*b*) forested wetland; (*c*) commercial property; (*d*) single family housing; (*e*) rivers, streams and ponds; (*f*) barren transitional areas; (*g*) urban grassland. (Adapted from Welch, 1985.)

conditions can affect the interpretability of the features being mapped from each source of data. Consequently, several authors have assessed the level of mapping completeness which can be achieved from the three sources of data described.

The identification of planimetric features from photography produced by the ESA Metric Camera is reported in Baudoin (1985). The results indicate detectability of various features in areas of high and low contrast. For areas of high contrast, the threshold width for linear features such as roads, rivers and hedges which will ensure a 100% level of success is 18 m. For low-contrast targets, the values are two to three times greater. For areas of woodland and isolated buildings, it was not possible to state threshold values for identification, even in areas of high contrast. Generally, the level of detail which can be interpreted from the metric camera photography allows only about 50–60% of the detail required for mapping at 1:100 000 to be obtained. Similar results for the earlier Skylab mission are reported in Welch (1982).

The results of tests to assess the levels of completeness from Landsat MSS are reported in Welch and Pannell (1982). The tests, carried out using imagery of three test sites in China, indicate that overall only 41% and 37% of the detail contained on a 1:250 000 map can be obtained from visual interpretation of black-and-white Landsat band 5 and 7 imagery respectively.

More recent analysis of the relative performance of SPOT and Landsat TM imagery is reported in Welch (1985), and a selection of the results is illustrated in Fig. 6.40. The results indicate that accuracy levels of over 80% can be realized for selected land use/land cover types using SPOT imagery, whereas completeness levels of 65–70% are more typical of the results obtainable from Landsat TM imagery. It is suggested that

the level of completeness is suitable for limited planimetric mapping at scales from 1:250 000 to 1:100 000.

The difficulties associated with SAR imagery from Seasat are summarized by Petrie (1985). The two main difficulties experienced were, firstly, deficiencies in image quality, and secondly, detection difficulties associated with the orientation of features relative to the direction of the radar signal. In the first case, the effect of background clutter considerably reduces the interpretability of the imagery. In the second case, the detection of linear features in particular seems to be highly dependent on their orientation relative to the SAR signal. The consequence of this is to produce rather inconsistent sets of results where features may or may not appear, depending on their orientation. It is essential that methods of improving the interpretability of such imagery are developed, and some progress in this field has been reported by Rye and Wright (1985), using segmentation techniques. However, it would seem that, until some improvements occur in this area, SAR imagery from space is unlikely to be a significant source of information for planimetric mapping, and may be of greater use in mosaic form for planning purposes (Van Roessel and DeGodoy, 1974).

6.5.1.4 *Contouring and digital terrain modelling (DTM).* Ackermann and Stark (1985) concluded that, on the basis of an assessment of photography of Bavaria, the height accuracy from the Metric Camera was about ± 20 m, and this would be appropriate for contouring with an 80–100 m contour interval. Using a substantial number of known control heights, the results were reduced to ± 15 m and 50–60 m respectively.

A more rigorous investigation into the accuracy of DTMs produced automatically from the Metric Camera using digital image correlation techniques is reported in Boochs and Decker (1986). By comparing known and observed heights, and filtering the dataset to eliminate gross errors, height errors of between ± 15 m and ± 31 m were obtained, the variations being caused by changes in image contrast, object geometry and object types within the test sites. Engels *et al.* (1985) suggested that the DTM produced from this data may be suitable for the production of orthophotographs at a scale of 1:50 000.

The potential of Landsat imagery (both MSS and TM) for height determination is very limited, given the very poor geometry of the lateral overlap between successive passes. Generally, the accuracies achieved have been at best ± 60 m for Landsat MSS (Dowman and Mohammed, 1981), and similar results have been obtained with Landsat TM imagery (Cooper *et al.*, 1986).

More promising results using simulated SPOT imagery have been reported by Cooper *et al.* (1986) and Swann *et al.* (1986). Using an edge-based hierarchical matching algorithm, height errors of between ± 7 m and ± 10 m were obtained. Further discussion of the potential applications of DTMs and SPOT imagery can be found in Baudoin (1984).

The heighting accuracy of spaceborne SAR imagery is also very poor. Stereo radar imagery obtained from the 1984 SIR-B mission has been evaluated by several researchers. The results reported include both manual extraction of heights on a modified analytical plotter (Leberl *et al.*, 1986), and automated height determination using correlation techniques (Simard *et al.*, 1986). The results using the latter technique have proved inconsistent, and at present do not appear suitable for the operational determination of heights. The work carried out by the former researchers is also disappointing: differences between the radar-generated DTM and the map-derived

Table 6.13 Suitability of various satellite sensors for topographic mapping

	Photographic sensors		Scanning sensors		Synthetic-aperture radar
	Metric Camera	Large Format Camera	Landsat Thematic Mapper	SPOT	Seasat SIR-A/B
Plan accuracy	***	***	**	***	*
Height accuracy	—	—	—	**	—
Map compilation	**	**	*	**	*
Map revision	***	***	*	****	*
Map intelligence	***	***	***	****	**

— Unsuitable
* Suitable for 1:250 000 mapping
** Suitable for 1:100 000 mapping at a relaxed specification (70% detail)
*** Suitable for 1:100 000 mapping
**** Suitable for 1:50 000 mapping.

DTM are about ± 60 m. Clearly the results are not currently of use on an operational basis, although improvements in the imaging geometry of future satellite SAR systems, such as Radarsat, may enable greater levels of accuracy to be obtained (Derenyi and Stuart, 1984).

6.5.1.5 *Suitability of satellite sensors for topographic mapping.* It is evident from the results provided in the previous three sections that, at present, the potential of satellite-based remote-sensing systems for topographic mapping is restricted to scales of between 1:100 000 to 1:250 000. Table 6.13 summarizes the suitability of the various sensors which have been discussed. For the future, the more likely application of the satellite data will be for map revision and map intelligence operations. In the latter case, Landsat TM imagery is currently being used by the Canadian government for detecting change on its 1:50 000 series in the more isolated northern regions. Mapping of the areas where a sufficient level of land cover change has occurred is then performed using conventional aerial photogrammetric techniques.

6.5.2 *Thematic mapping for engineering projects*

The results reported in the previous section have highlighted some of the reasons why, at present, the operational use of satellite remote sensing for small-scale topographic mapping is limited. Remote sensing data has, however, been used more successfully to produce thematic maps for engineering projects, since for such applications users are satisfied with a relatively lower level of positional accuracy and completeness in comparison with the requirements for topographic mapping. At larger scales also, remote sensing has a role to play, and the following sections indicate some of the potential uses of such imagery during the reconnaissance/planning, feasibility, design, construction and maintenance and operation stages of a civil engineering project (Fig. 6.41).

6.5.2.1 *Reconnaissance/planning stage.* The first stage of any major civil engineering project generally involves some form of preliminary reconnaissance study of the region

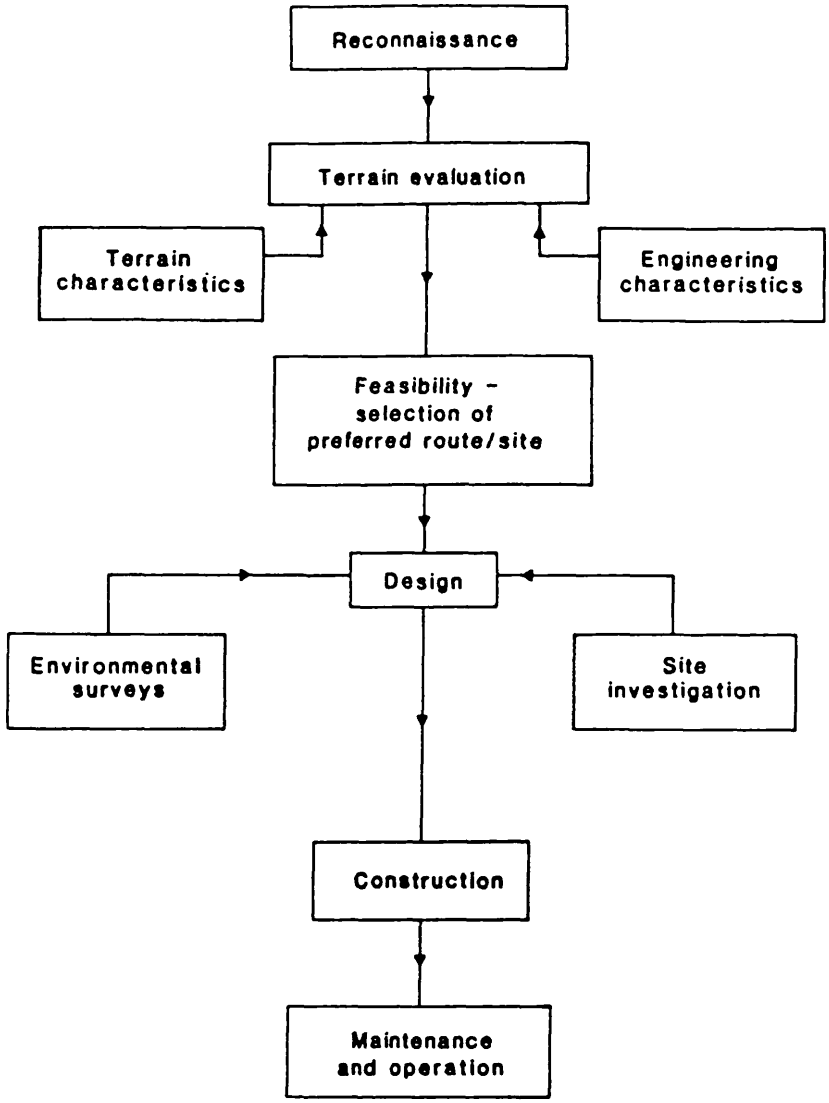


Figure 6.41 Main stages of a civil engineering project.

or project area. The primary aim of this study is to collect together all available data concerning the physical characteristics of the terrain in order to assess its likely influence on the overall design of the engineering works.

Traditionally, the engineer has carried out this reconnaissance stage by examining existing maps of the region; both topographic maps and specialist thematic maps (e.g. geological, geomorphological and pedological) are of greatest use at this initial stage of the project. In addition, examination of engineering reports produced for previous projects in the region may also provide valuable reference material. The main problems associated with this approach are, firstly, that in many regions existing maps may be at a very small scale (< 1:500 000), significantly out of date, and in some instances may not exist at any scale. Secondly, whilst engineering reports may give valuable data about a specific site, they may not be appropriate to or representative of regional terrain conditions. Consequently, many engineering projects now involve an examination of satellite and aerial imagery at this early stage.

The use of Landsat MSS data for projects of this type has been reported by Beaumont (1982). In this case, use of Landsat data for water resources planning in North-East Somalia is reported. Visual interpretation of Landsat data enabled a series of

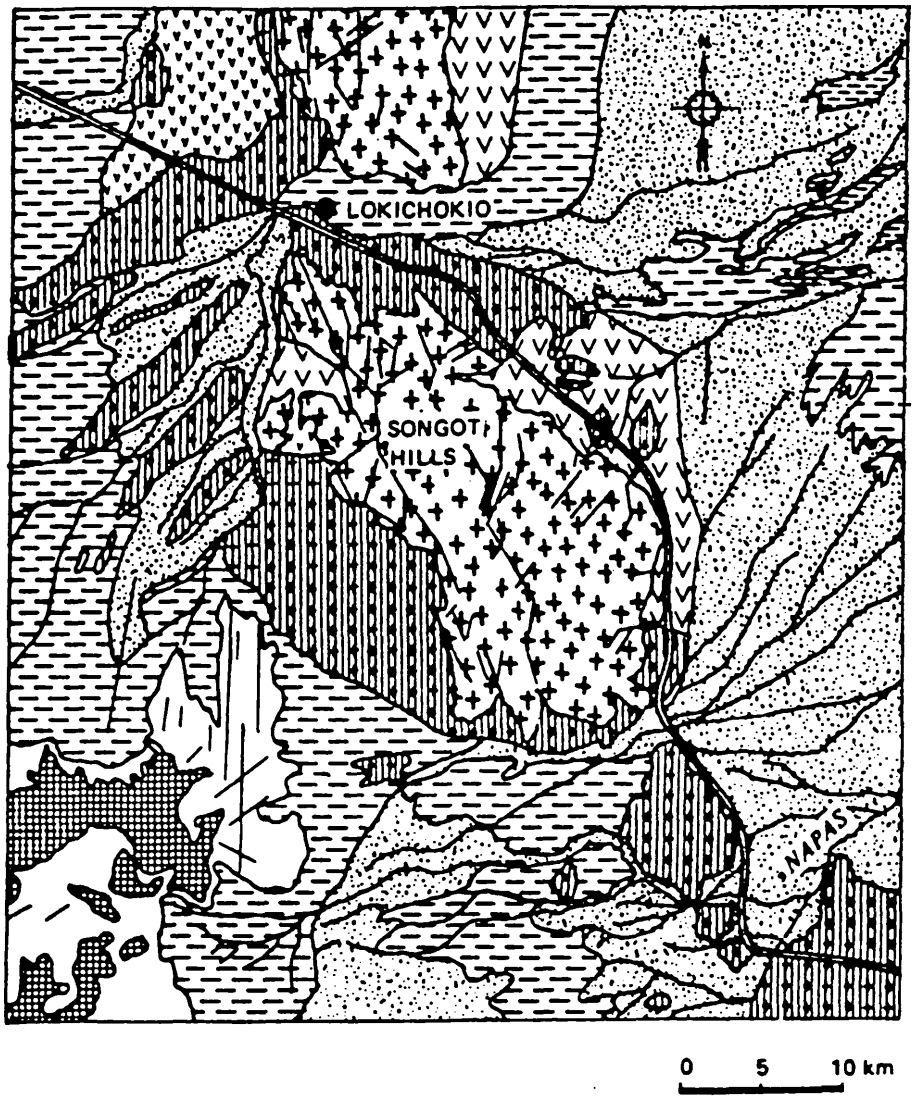


Figure 6.42 Geological map of Northern Kenya as interpreted from 1:250 000 Scale Landsat MSS false-colour composite image (Dick, 1982).

transparent overlays to be produced, illustrating drainage, surface water, groundwater potential and land capability over the region. A similar approach in relation to agricultural development planning in Southern Sudan is reported in Beaumont (1985).

The use of remote sensing for terrain evaluation purposes is also of considerable importance. Terrain evaluation is a form of thematic mapping in which the terrain is classified in a hierarchical manner into units having common landscape patterns and engineering characteristics. The former element of the process can be carried out very cost-effectively using satellite imagery in conjunction with aerial photography (Overseas Unit, 1982).

The use of Landsat MSS imagery for transport planning in Northern Kenya and Southern Sudan is reported in Dick (1982). Geomorphological and geological mapping was carried out for this project by visual interpretation of single-band and false-colour composite images. Figure 6.42 illustrates some of the results which were obtained.

Interpretation of Landsat data may also be useful for the provision of additional supplementary information on hydrological phenomena such as flooding (Deutsch and

Ruggles, 1978) and water quality monitoring (Alfoldi and Munday, 1978). A more general review of hydrological applications can be found in Blyth (1985).

Landsat has also been used for estimating urban populations over large areas (Forster, 1983), and for monitoring change, using multitemporal imagery (for example the expansion of urban areas, changes in soil erosion, variations in river channels, the impact of desertification and so on).

6.5.2.2 *Feasibility stage.* The aim of the feasibility stage of a project is firstly, to select potential routes/or sites, and secondly, to select from these the best from the available options. Beaven and Lawrance (1982) identified the following engineering requirements for the feasibility study where remote sensing could be of assistance:

(i) Determination of foundation conditions
(ii) Location of drainage patterns, calculation of catchment areas and positioning of culverts
(iii) Identification of spoil areas and possible borrow pits
(iv) Location of possible bridge sites
(v) Identification of major hazard areas, for example, poorly drained soils, spring lines, unstable ground
(vi) Location of possible sources of constructional materials.

In developing regions, the location of low-cost construction materials such as calcrete may be important (Beaven and Lawrance, 1982; Lawrance and Toole, 1984). Calcrete is a calcareous material which generally accumulates in large bare depressions in arid regions. It is a widespread source of low-cost road construction material, and consequently its location and distribution can have a considerable effect on the positioning of a new road alignment. Supervised classification techniques have been used to identify sources of calcrete, and Fig. 6.43 illustrates one such classification. The change in colour of the overlying sand deposits from their normal red or brownish hue to a neutral grey has been found to be correlated with calcrete deposits. Figure 6.44 illustrates the discrimination of these regions by the use of principal components analysis.

A comprehensive review of the role of remote sensing for highway feasibility studies can be found in Lawrance and Beaven (1985).

6.5.2.3 *Design.* Remote sensing has a particularly valuable role to play at the design stage of a project. At this stage a detailed site investigation of the proposed site or route will be required, together with supplementary environmental information which may influence the design of the project being considered. A critical stage of the site investigation stage is the desk study. The desk study provides the engineer with an overview of the project area, and is normally carried out using aerial photography. Not only can valuable information be obtained during this interpretation about possible engineering problems, but it can also be very important for the planning of the subsequent ground investigations. A general review of the potential of aerial photography in this respect is provided by Dumbleton (1983) and Matthews (1985), and a selected summary of the literature which has examined specific geotechnical problems is presented in Table 6.14.

Of increasing importance in recent years has been the role of remote sensing in providing environmental engineering data for input into the design process. A

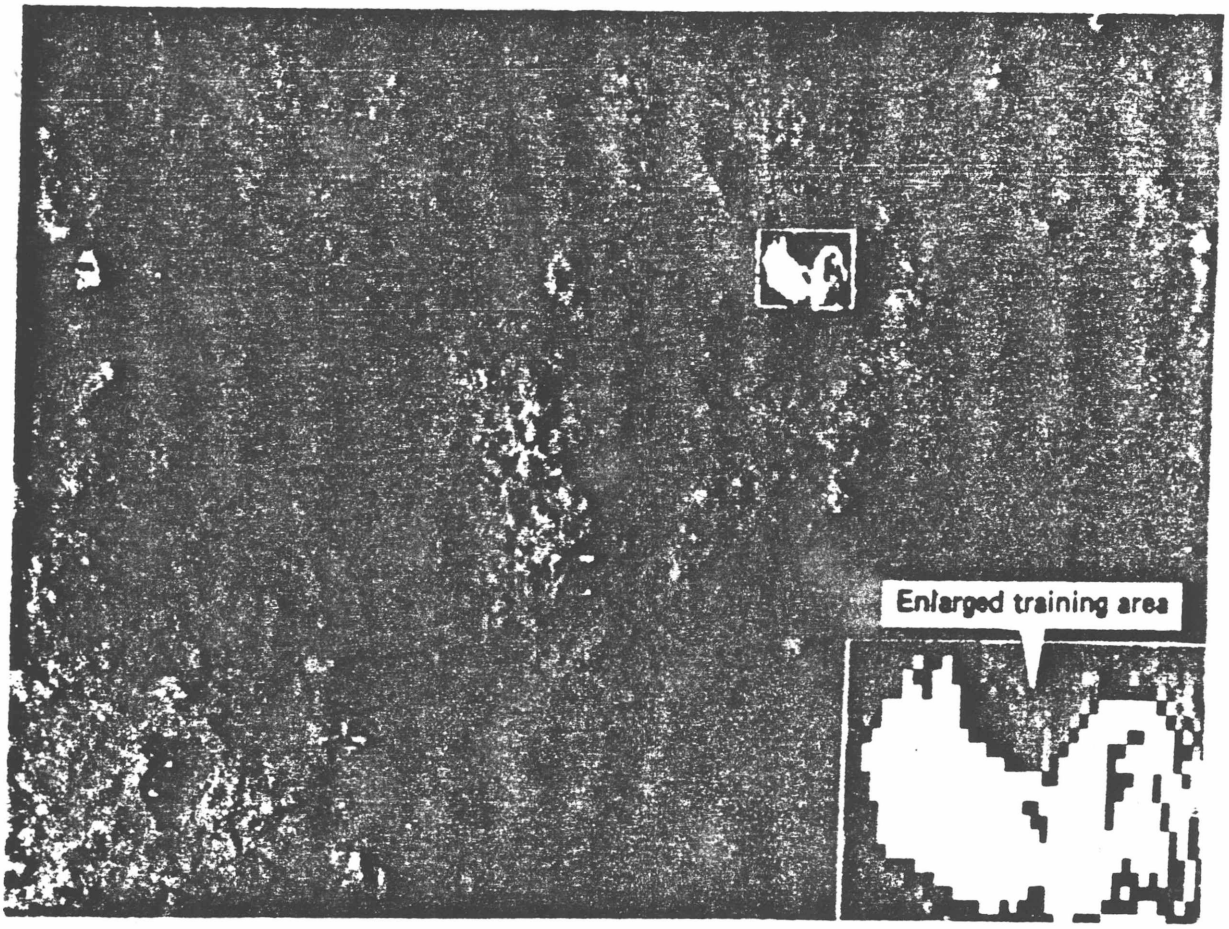


Figure 6.43 Location of calcrete deposits in Botswana–supervised classification and training area (Beaven and Lawrance, 1982).

comprehensive review of remote sensing in this context is provided by Mason and Amos (1985). Information which may be extracted from remote sensing imagery (often thermal IR linescan) includes data on energy conservation (heat loss), power-station cooling-water patterns, groundwater and spring detection, pipeline and field drainage patterns and hydraulic leakage from earth dams.

6.5.2.4 *Construction stage.* The role of remote sensing is of much less importance during construction than at the preceding planning and design stages. Nevertheless, aerial photography can be used to provide a valuable historial record of construction activities and may also be used to plan construction activities. Singhroy (1987) provides an interesting case history concerning the use of large-scale colour IR aerial photography and aerial video during pipeline construction in Canada. Both types of data were used to plan construction activities, determine site conditions and assess environmental effects before, during and after construction activity.

6.5.2.5 *Post-construction maintenance.* Post-construction applications of remote sensing are often related to the need to monitor the impact of the civil engineering project on the environment or to assist in the maintenance of the projcet. Examples which are relevant in this context include aerial traffic investigations using helicopters and airborne video cameras (Mountain and Garner, 1981), or the location of road surfaces suceptible to icing during winter months (Stove *et al.*, 1987). On a regional scale,

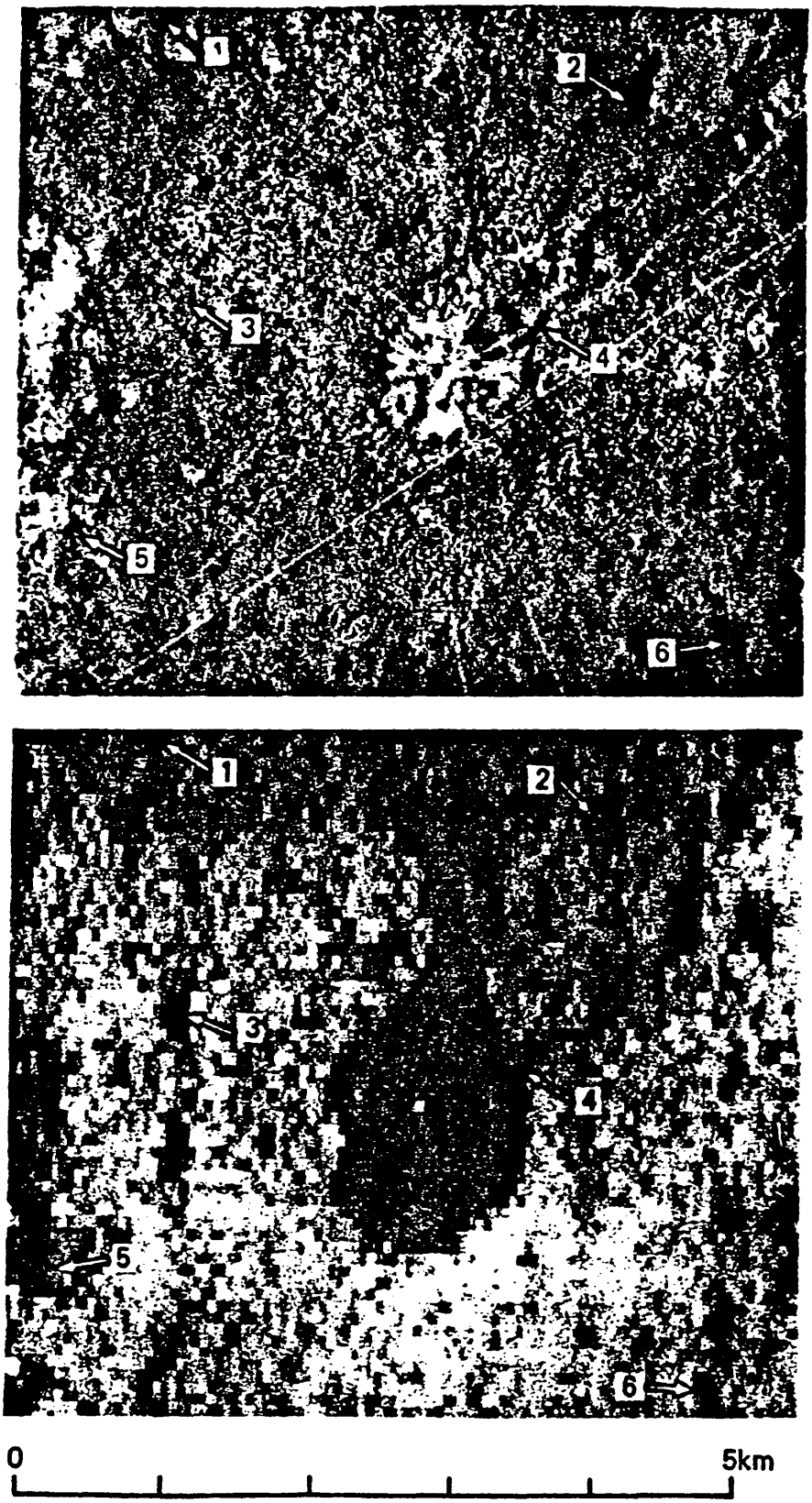


Figure 6.44 Location of calcrete deposits in Botswana–comparison of aerial photograph (top) and Landsat MSS image (bottom). Although the aerial photograph has a high spatial resolution, it is unable to distinguish areas of grey sand, where calcrete occurs, from the surrounding reddish brown sands. The Landsat image (after principal components analysis) enhances the difference and shows them as variations in dark tones (Beaven and Lawrance, 1982).

Table 6.14 Remote sensing and site investigations—selected references

	Aerial photography	MSS	Thermal IR linescanning
Landslides/ slope instability	Matthews and Clayton (1987)	Liu (1985)	Chandler (1975)
Solution features	Norman and Watson (1975)	Coker *et al.* (1969)	Edmonds *et al.* (1986)
Derelict/ contaminated land	Bullard (1983)	Coulson and Bridges (1984)	Coulson and Bridges (1984)
Surficial/ subsurface materials surveys	Caiger (1970)	Lynn (1984)	Singhroy and Barnett (1984)

satellite remote sensing may be used to evaluate the environmental impact of a project. For example, multitemporal Landsat MSS and TM imagery has been used to evaluate the historical development of irrigation practices in Morocco.

In conclusion, Table 6.15 illustrates the recommended use of remote-sensing techniques at each stage of a civil engineering project.

6.6 Conclusions and future trends

The primary objective of this chapter has been to discuss some of the surveying and civil engineering projects which may benefit from the use of airborne and spaceborne remote sensing. With a few exceptions, the use of remote sensing in these fields is currently still at the research and development stage. Before the technology is used routinely on an operational basis, it will require further advances, particularly with regard to the spatial resolution of spaceborne sensors.

For topographic mapping applications, the interpretability of the currently available satellite imagery is a major obstacle to its use even for very small-scale mapping. Similarly the accuracies attainable for height determination, even from SPOT imagery, limit its application to small-scale map revision. The sensors which are widely believed to offer the greatest benefits, particularly in the UK, are the SAR systems. Again, however, at present the discrimination of features on the imagery is very poor and generally inadequate for topographic mapping. It would seem therefore that aerial photography and the photogrammetric techniques discussed in Chapter 7 will remain the pre-eminent solution for routine topographic mapping operations.

In the creation of thematic maps for engineering projects, the development of more sophisticated processing and analysis techniques based on the use of geographical information systems (GIS) (Chapter 11), will be of considerable importance for the future. By using such systems, the engineer will eventually be able to examine a wide range of vector map data, and attribute textual data and raster-based remote sensing data. With the ability to interrogate this database and generate maps which consider the interrelationships between several different variables, the surveyor and civil engineer will have a very powerful capability at his disposal.

Table 6.15 Recommended use of remote-sensing techniques during a civil engineering project

Project phase	Photograph/ image scale	Satellite						Aircraft			
		Photography	Landsat MSS	Landsat TM	SPOT	MOMS	Side-Looking Radar	Photography	MSS	Thermal Scanners	SLAR
Reconnaissance/ planning	1:1 000 000 to 1:50 000	***	****	***	****	***	***	**	—	—	—
Feasibility	1:100 000 to 1:20 000	***	**	***	***	***	***	**	—	—	—
Design	1:20 000 to 1:2000	—	—	*	*	—	—	****	***	***	***
Construction	1:2000 to 1:500	—	—	—	—	—	—	****	*	*	*
Post-construction/ maintenance	1:2000 to 1:500	—	—	—	—	—	—	****	*	*	*

**** Very useful *** Very useful (but coverage limited) ** Useful * Of limited use — inadequate.

References

Ackermann, F. and Stark, E. (1985) Digital elevation model from Spacelab metric camera photography. *ESA Spec. Publ.* **SP-209**, 9–12.

Alfoldi, T.T. and Munday, J.C. (1978) Water quality analysis by digital chromaticity mapping of Landsat data. *Can. J. Remote Sensing* **4**, 2.

Ali, A.E. (1986) Optical and digital Seasat SAR processing techniques: a comparison of accuracy results. *Int. Arch. Photogramm. and Remote Sensing* **26**(3/1), 28–39.

Amin, A.M. (1986) Geometrical analysis and rectification of thermal infrared video frame scanner imagery and its potential applications to topographic mapping. Ph.D. Thesis, University of Glasgow, 257 pp.

Annett, R.J. (1985) Radarsat: low earth orbit platform. *Proc. Conf. on Advanced Technology for Global Environmental Monitoring*, CERMA and Remote Sensing Society, 111–123.

Bagot, K.H. (1985) Digital processing of remote sensing data. In *Remote Sensing in Civil Engineering*, eds. T.J.M. Kennie and M.C. Matthews, Surrey University Press [Blackie], Glasgow and London, 87–105.

Baudoin, A. (1984) Applications of SPOT and DTM to cartography. *Proc. 18th Int. Symp. on Remote Sensing of Environment, Paris*, **1**, 61–75.

Baudoin, A. (1985) Identification of planimetric features from the Metric camera. *ESA Special Publication* **SP-209**, Metric Camera Workshop, 69–75.

Beaumont, T.E. (1977) Techniques for the interpretation of remote sensing imagery for highway engineering purposes. *TRRL Rept* **753**, Transport and Road Research Laboratory, Crowthorne, Berks., 24 pp.

Beaumont, T.E. (1982) Land capability studies from Landsat satellite data for rural road planning in North East Somalia. In *Proc. OECD Symp. on Terrain Evaluation and Remote Sensing for Highway Engineering in Developing Countries. TRRL Rept* **SR 690**, Transport and Road Research Laboratory, Crowthorne, Berks., 86–95.

Beaumont, T.E. (1985) Interpretation of Landsat satellite imagery for regional planning studies. In *Remote Sensing in Civil Engineering*, eds. T.J.M. Kennie and M.C. Matthews, Surrey University Press [Blackie], Glasgow and London, 164–201.

Beaven, P.J. and Lawrence, C.J. (1982) Terrain evaluation for planning and design. *TRRL Suppl. Rept* **725**, Transport and Road Laboratory, Crowthorne, Berks., 24 pp.

Bernstein, R. (1978) *Digital Image Processing for Remote Sensing.* IEEE Press, New York, 473 pp.

Blyth, K.B. (1985) Remote sensing and water resources engineering. In *Remote Sensing in Civil Engineering*, eds. T.J.M. Kennie and M.C. Matthews, Surrey University Press [Blackie], Glasgow and London, 289–334.

Braun, H. and Velten, E. (1986) The first ESA Remote Sensing satellite: status and outlook. *Proc. ISPRS Comm. I Symp. Progress in Imaging Sensors, Int. Arch. of Photogramm. and Remote Sensing* **26**, 1, 329–334.

Bullard, R.K. (1983) Abandoned land in Thurrock: an application of remote sensing. *Working Paper* **8**, North East London Polytechnic, London.

Burnside, C.D. (1982) *Mapping from Aerial Photographs.* Granada, London, 303 pp.

Burnside, C.D. (1985) The future prospect of data acquisition by photographic and other airborne systems for large scale mapping. *Photogramm. Record* **11**(65), 495–506.

Caiger, P.B. (1970) Aerial photographic interpretation of road construction materials in S. Africa with special reference to its potential to influence route location in underdeveloped territories. *Photogrammetria* **25**, 151.

Chandler, P.B. (1975) Remote detection of transient thermal anomalies associated with the Portuguese Bend landslide. *Bull. Assoc. Eng. Geol.* **12**(3), 227–232.

Chevrel, M., Courtois, M. and Wells, G. (1981) The SPOT satellite remote sensing mission. *Photogramm. Eng. and Remote Sensing* **47**, 1163–1171.

Coker, A.E., Marshall, R. and Thompson, N.S. (1969) Application of computer processed multispectral data to the discrimination of land collapse (sinkhole) prone areas in Florida. *Proc. 6th Int. Symp. on Remote Sensing of the Environment*, Michigan, Vol. 1, 65–69.

Cooper, P.R., Friedmann, D.E. and Wood, S.A. (1986) The automatic generation of terrain models from satellite images by stereo. *Acta Astranautica* **14**, 10 pp.

Coulson, M.G. and Bridges, E.M. (1984) The remote sensing of contaminated land. *Int. J. Remote Sensing* **5**(4), 659–669.

Curran, P.J. (1985) *Principles of Remote Sensing*. Longman, London, 282 pp.

Derenyi, E.E. and Stuart, A.J. (1984) Geometric aspects of spaceborne stereo imaging radar. *Int. Arch. Photogramm. and Remote Sensing* **XXV**(A7), VII, 153–162.

Deutch, M. and Ruggles, F.H. (1978) Hydrological applications of Landsat imagery used in study of the 1973 Indus river flood, Pakistan. *Water Resources Bull.* **14**(2), 261–274.

Dick, O.B. (1982) Remote sensing for transport planning and highway planning. *TRRL Suppl. Rept* **690**, Transport and Road Research Laboratory, Crowthorne, Berks., 172 pp.

Doyle, F.J. (1982) Satellite systems for cartography. *Int. Arch. of Photogramm.* **24**(1), 180–18.

Dowman, I.J. and Mohamed, M.A. (1981) Photogrammetric applications of Landsat MSS imagery. *Int. J. Remote Sensing* **2**, 105–113.

Dowman, I.J. and Morris, A.H. (1982) The use of synthetic aperture radar for mapping. *Photogramm. Record* **10**(60), 687–696.

Dumbleton, M.J. (1983) Air photographs for investigating natural changes, past use and present condition of engineering sites. *TRRL Rept* **LR 1085**, Transport and Road Research Laboratory, Crowthorne, Berks.

Edmonds, C.N., Kennie, T.J.M. and Rosenbaum, M. (1986) The application of airborne remote sensing to the detection of solution features in limestone. *Proc. 22nd Ann. Meeting of the Engineering Group of the Geological Society, Plymouth Polytechnic*, 109–120.

Elachi, (1980) Spaceborne imaging radar—geologic and oceanographic applications. *Science* **216**, 1073–1082.

Elachi, C. (1982) Shuttle imaging radar experiences. *Science* **218**, 996–1004.

Engels, H., Muller, W. and Gonecny, G. (1985) Application of Spacelab Metric Camera imagery for map production. *Int. Arch. Photogramm. and Remote Sensing* **26**(4), 170–182.

Everitt, J.H. and Nixon, P.R. (1985) False colour video imagery: a potential remote sensing tool for range management. *Photogramm. Eng. and Remote Sensing* **51**(6), 675–679.

Fearns, D.C. (1984) Microcomputer systems for satellite image processing. *Earth Orient. Appl. Space Technol.* **4**(4), 247–254.

Forstner, G. (1983) Some urban measurements from Landsat data. *Photogramm. Eng. and Remote Sensing* **49**(12), 1693–1707.

Fusco, L., Frei, U. and Hsu, A. (1985) Thematic mapper: operational activities and sensor performance at ESA/Earthnet. *Photogramm. Eng. and Remote Sensing* **51**(9), 1299–1314.

Goetz, A.F.H., Vane, G., Solomon, J.E. and Rock, B.N. (1985) Imaging spectrometry for earth remote sensing. *Science* **228**, 1147–1153.

Graham, R.W. and Read, R. (1975) Small format aerial photography from microlight platforms. *J. Photogr. Sci.* **32**, 100–109.

Heath, W. (1980) Inexpensive aerial photography for highway engineering and traffic studies. *TRRL Suppl. Rept* **632**, Transport and Road Research Laboratory, Crowthorne, Berks., 24 pp.

Hodgson, R.M., Cady, F.M. and Pairman, D. (1981) A solid state airborne sensing system for remote sensing. *Photogramm. Eng. and Remote Sensing* **47**(2), 177–182.

Isong, M. (1987) An investigation of the geometric fidelity of spaceborne and airborne thematic mapper imagery and its potential application to topographic mapping. *M. App. Sci. Dissertation*, University of Glasgow, 242 pp.

Jensen, H., Graham, L.C., Porcello, J. and Leith, E.N. (1977) Side looking airborne radar. *Scientific American* **237**(6), 84–94.

Kennie, T.J.M., Dale, C.D. and Stove, G.C. (1986) A preliminary assessment of an airborne thermal video frame scanner. *Int. Arch. Photogramm. and Remote Sensing* **26**(7/2), 285–290.

Lawrance, C.J. and Beaven, P.J. (1985) Remote sensing for highway engineering projects in overseas countries. In *Remote Sensing in Civil Engineering*, eds. T.J.M. Kennie and M.C. Matthews, Surrey University Press [Blackie], Glasgow and London, 240–268.

Lawrance, C.J. and Toole, T. (1984) The location, selection and use of calcrete for bituminous road construction in Botswana. *TRRL Rept* **LR 1122**, Transport and Road Research Laboratory, Crowthorne, Berks.

Leberl, F.W., Domik, G. and Kobrick, M. (1985) Mapping with aircraft and satellite radar images. *Photogramm. Record* **11**(66), 647–666.

Leberl, F.W., Domik, G., Raggam, J. and Kobrick, M. (1986) Radar stereomapping techniques and application to SIR-B images of Mt. Shasta. *IEEE Trans. on Geoscience and Remote Sensing* **GE-24**(4), 473–480.

Liu, J.K. (1985) Remote sensing for identifying landslides and for landslide prediction—cases in

Taiwan. *Int. Conf. on Remote Sensing Data Acquisition, Management and Applications*, 223–232.

Lynn, D.W. (1984) Surface material mapping in the English fenlands using airborne multispectral scanner data. *Int. J. Remote Sensing* **5**(4), 699–713.

Matthews, M.C. (1985) Interpretation of aerial photography. In *Remote Sensing in Civil Engineering*, eds. T.J.M. Kennie and M.C. Matthews, Surrey University Press [Blackie], Glasgow and London, 204–239.

Matthews, M.C. and Clayton, C.R.I. (1987) The use of oblique aerial photography to assess the extent and sequence of landslipping at Stag Hill, Guildford, *Proc. 20th Ann. Meeting of the Engineering Group of the Geological Society*, Geological Society, London.

Mason, P.A. and Amos, E.L. (1985) Environmental engineering applications of thermal infrared imagery. In *Remote Sensing in Civil Engineering*, eds. T.J.M. Kennie and M.C. Matthews, Surrey University Press [Blackie], Glasgow and London, 269–288.

Meier, H.-K. (1985) Image quality improvement by forward motion compensation. Paper presented at 40th Photogrammetric Week, Stuttgart, Institute of Photogrammetry, University of Stuttgart, 20 pp.

Meisner, D.E. and Lindstrom, O.M. (1985) Design and operation of a color infrared video system. *Photogramm. Eng. and Remote Sensing* **51**(5), 555–560.

Menuguette, A.A.C. (1985) Evaluation of metric camera photography for mapping and coordinate determination. *Photogramm. Record* **11**(66), 699–710.

Mohamed, M.A. (1977) Photogrammetric analysis and rectification of Landsat MSS and Seasat SAR imageries. *Ph.D. Thesis*, University of London, 180 pp.

Mountain, L.J. and Garner, J.B. (1981) Semi-automatic analysis of small format photography for traffic control studies of complex intersections. *Photogramm. Record* **10**, 331–342.

NOAA (1978–1986) *Landsat Data User's Notes*, Issues 1–36. National Oceanographic and Atmospheric Administration, Washington DC.

Norman, J.W. and Watson, I. (1975) Detection of subsidence conditions by photogeology. *Eng. Geology* **9**, 359–381.

Overseas Unit (1982) Terrain evaluation and remote sensing for highway engineering in developing countries. *TRRL Suppl. Rept* **690**, Transport and Road Research Laboratory, Crowthorne, Berks., 172 pp.

Petrie, G. (1978) Geometric aspects and cartographic applications of side looking radar. *Proc. 5th Ann. Conf. Remote Sensing Society*, University of Durham, 21 pp.

Petrie, G. (1983) The philosophy of digital and analytical photogrammetric systems. *Proc. 39th Photogrammetric Week, Stuttgart*, Institute for Photogrammetry, University of Stuttgart, 27 pp.

Petrie, G. (1985) Remote sensing for topographic mapping. In *Remote Sensing in Civil Engineering*, eds. T.J.M. Kennie and M.C. Matthews, Surrey University Press [Blackie], Glasgow and London, 119–161.

Petrie, G. (1987) General photogrammetry: an introduction to basic concepts, instrumentation and procedures. *ASPRS Int. Conf. and Workshop on Analytical Instrumentation, Phoenix, Arizona*, 36 pp.

Rudd, R.D. (1974) *Remote Sensing: A Better View*. Duxbury Press, N. Scituate, Mass., 135 pp.

Rye, A.J. and Wright, A. (1985) Segmentation—the extraction of spatial information from radar images. *Int. Conf. on Advance Tech. for Monitoring and Processing Global Environmental Data, London*, 305–317.

Simard, R., Plourde, F. and Toutin, T. (1986) Digital elevation modelling with stereo SIR-B image data. *Int. Arch. Photogramm. and Remote Sensing* **26**(7/2), 161–166.

Singhroy, V. (1988) Case studies on the application of remote sensing data to geotechnical investigations in Ontario. In *Geotechnical Engineering Applications of Remote Sensing and Remote Data Transmission, ASTM Spec. Tech. Publ.* **967**, American Society for Testing and Materials, Philadelphia, 9–45.

Singhroy, V. and Barnett, P. (1984) Locating subsurface mineral aggregate deposits from airborne imagery: a case study in southern Ontario. *Int. Symp. on Remote Sensing for Exploration Geology, Colorado Springs*, 523–539.

Stove, G.C., Kennie, T.J.M. and Harrison, L. (1987) Airborne thermal mapping for winter highway maintenance using the Barr and Stroud IR18 thermal video frame scanner. *Int. J. Remote Sensing* **8**(7), 1077–1084.

Swann, R., McDonald, J.S., Westwell Roper, A., Wood, S. and Laing, W. (1986) The automated

extraction of digital terrain models from satellite imagery. *Proc. 20th Int. Symp. on Remote Sensing of the Environment, Nairobi*, 16 pp.

Tait, D.A. (1987) *Small Computers in Image Processing and Mapping*. Remote Sensing Society, London, 103 pp.

Taranik, J.V. (1978) Principles of computer processing of Landsat data for geological applications. *USGS Open File Report* **78-117**, 50 pp.

Tomlins, G.F. (1983) Some considerations in the design of low cost remotely piloted aircraft for civil remote sensing applications. *Canadian Surveyor* **37**, 157–167.

Traynor, C.P. and Jeans, T.G. (1982) The development of a satellite borne earth imaging system for UOSAT. *Radio and Electronic Engineer* **52**(8/9), 398–402.

Trevett, J.W. (Ed.) (1985) *The European SAR 580: Final Report*. ESA, Joint Research Centre (JRC), Ispra Publication.

Trevett, J.W. (1986) *Imaging Radar for Resources Surveys*. Chapman and Hall, London, 313 pp.

Van Roessel, J.N. and de Godoy, R.C. (1974) SLAR mosaics for Project RADAM. *Photogramm. Eng.* **40**(6), 583–595.

Welch, R. (1982) Image quality requirements for mapping from satellite data. In *Proc. ISPRS Commission 1, Symp. on Primary Data Acquisition*, 50–54.

Welch, R. and Panell, C.W. (1982) Comparative resolution of Landsat 3 MSS and RBV image data of China. *Photogramm. Record* **10**(59) 575–586.

Welch, R. (1985) Cartographic potential of SPOT image data. *Photogramm. Eng. and Remote Sensing* **51**(8), 1085–1090.

Welch, R., Jordan, T.R. and Ehlers, M. (1985) Comparative evaluations of the geodetic accuracy and cartographic potential of Landsat 4 and 5 thematic mapper data. *Photogramm. Eng. and Remote Sensing* **51**(11), 1799–1812.

Wolf, P.R. (1983) *Elements of Photogrammetry*. 2nd edn., McGraw-Hill, 628 pp.

Wong, K.W. (1975) Geometric and cartographic accuracy of ERTS-1 imagery. *Photogramm. Eng.* **41**(5), 621–635.

5 Remote sensing in civil engineering practice

T. J. M. KENNIE, M. C. MATTHEWS and P. J. BEAVEN

5.1 Introduction

The measurement and recording of data relating to the physical characteristics of the earth's surface is a critical phase of any major civil engineering project. The information which these 'terrain investigations' provide is an essential element of the initial planning stages of a project and it can also provide useful data in subsequent phases of a project. Remote sensing in the form of the photographic, scanning and processing systems discussed in earlier chapters is one of the most appropriate means of recording existing ground conditions, of assessing their potential for engineering projects and also of evaluating the effect, or potential effects, of the subsequent construction on the environment. It should however be stressed that remote sensing is only an aid to engineering investigations, providing information which is complementary to that obtained from field measurements, site visits and existing sources of data such as maps and project reports. Indeed, if maximum benefit is to be gained from remote sensing techniques it is essential that 'ground truth' information from external sources forms part of the survey in order to verify the results of the remote sensing investigation independently.

It is therefore the primary aim of this chapter to examine the role of remote sensing systems in terrain investigations for civil engineering projects. A secondary aim is to present a coherent summary of the present and future uses of remote sensing in civil engineering practice. Subsequent chapters will deal in specific terms with the application of particular remote sensing techniques to various branches of civil engineering practice.

5.2 The role of remote sensing in terrain investigations

Remote sensing methods are a highly cost-effective means by which the civil engineer can obtain complete, up-to-date and often repetitive information about terrain conditions. The data produced about the terrain may be in two

forms, either metric or interpretive. In the former case the information relates to the precise position of features, for example as provided by topographic maps produced by photogrammetric methods. In the latter case the information is primarily thematic in nature based on the subjective interpretation of the remotely sensed imagery, with much less emphasis on positional accuracy. Whilst this categorization is clearly not mutually exclusive it does nevertheless illustrate the broad distinction in the role which remote sensing plays in ground investigation.

5.2.1 *Photogrammetry*

The use of remote sensing imagery (almost totally photography) for the production of orthophotographs and topographic maps by photogrammetric means is a well established and highly sophisticated science. Whilst this aspect of the use of remote sensing in civil engineering is dealt with in greater detail in Chapter 6, it is useful to examine one particular example in order to illustrate the use of remote sensing in this role.

Route mapping for highway location and design is one of the most common forms of photogrammetric project in civil engineering. In this role the remote sensing data, typically in the form of medium to large scale black-and-white panchromatic aerial photography, enable the civil engineer to evaluate several alternative route alignments within a broad corridor of interest. In addition, by further restricting the area under consideration to a much narrower band of interest, more detailed preliminary design data can be examined before the final design data is produced for one particular alignment. By using different scales of aerial photography, as shown in Table 5.1, together with photogrammetric methods, topographic maps, digital ground models (DGM) and detailed design data can be produced. Further discussion of remote sensing in this particular role can be found in Leatherdale (1975) and Adolfsson (1972). More details regarding the further use of remote sensing in highway engineering can be found in Chapter 9.

Table 5.1 Typical specifications for route/highway mapping by photogrammetric methods

Stage of investigation	Photographic scale	Flying height (assuming wide angle lens)	Map scales	Contour interval (m)
Route location	1:30 000	4 600	1:20 000–1:10 000	5
Preliminary design	1:10 000	1 500	1:2 500	1
Final design	1:3 000	460	1:500	0.5

5.2.2 *Interpretation*

A second and vitally important role of remote sensing in terrain investigations is as a source of data on various themes of engineering significance. In this case the information provided by remote sensing involves either subjective human interpretation or automatic computer classification of the available imagery. Examples of the specific forms of thematic information which can be provided by remote sensing techniques include geology, geomorphology, geotechnical information such as site history, extent of reclaimed land and location of landslips, together with additional information, including drainage patterns, groundwater potential, urban/rural development, transportation patterns, energy monitoring and pollution control.

Whilst the interpretation of remote sensing data can be an extremely useful tool for the civil engineer, it is essential before embarking on a terrain investigation involving remote sensing to consider the following points. Firstly, consideration should be given to whether remote sensing is appropriate. This will depend largely on the extent of existing material and, if material does exist, on the reliability of that information. If remote sensing is considered a suitable technique the question of choosing the most appropriate system(s) arises. This also raises questions about further matters such as the most suitable method of viewing and interpreting the imagery, the means of verifying the results of any interpretation, and of obvious importance, the costs which will be incurred in a project of this nature.

The choice of the most suitable technique to use depends on several factors including (a) the scale of the project (regional or local), and (b) the stage of the investigation, (reconnaissance, feasibility, design, construction or post-construction). Both factors also tend to be related to the stage of development of a country. In general terms the distinction can be made between regional investigations concerned with the reconnaissance and feasibility stages of a civil engineering project, and more local investigations where the primary role of remote sensing is as an aid at the design, construction and post-construction stages of a project. In the former case the projects tend to be almost exclusively located in developing or Third World countries, whereas in the latter the projects are more predominant in the developed world.

Several reasons account for this situation. In developing countries the major economic needs are likely to involve the development of transportation, drainage and agriculture and also the exploitation of natural resources. Projects undertaken in such areas are therefore likely to be on a regional scale. In addition the basic information in the form of topographic and geological maps is likely to be limited or in some instances non-existent. Hence there is an immediate requirement at the reconnaissance stages of such projects to obtain an overview of the ground surface. This can be gained rapidly and at reasonable cost only by using remote sensing techniques. The resolution of the techniques which are used need not be very high, and hence Landsat MSS and

TM imagery and high-altitude aerial photography may be quite suitable. In some areas where the earth's surface is continually obscured by cloud the use of airborne or satellite-based radar (SAR) systems may be the only means of obtaining the necessary information. More detail relating to the application of Landsat imagery, in particular, for regional planning studies of this nature may be found in Chapter 7.

In contrast, within developed countries the majority of construction projects are of a more local nature. In addition since the demand for land may be high and the amount of land take restricted on economic or environmental grounds, the scope for selecting the most suitable sites for construction is much more limited. Also the amount of background information about the project area in the form of topographic maps, geological survey reports, site investigation reports, records of mines and mineral deposits, hydrological information and aerial photography is likely to be high. Much greater emphasis is therefore placed on investigating the ground conditions at a local level, particularly for design purposes. Further discussion relating to the interpretation of aerial photography for localized projects of this type may be found in Chapter 8.

5.3 The use of remote sensing for projects in the developing world

The range of development and economic status of countries outside Europe and the USA is diverse. However they all share the need to expand their civil engineering infrastructure in order to support development projects in areas such as agriculture, urban development, mineral extraction and forestry. All projects of this nature generate the need for terrain investigations for planning purposes. These are generally undertaken on a national scale, although more detailed investigations may also be necessary subsequently, when a particular project is agreed upon. The relative importance of the engineering terrain investigations depends primarily on the objective of the project, and specifically on whether it involves a planning or design study. A feature common to all such projects is that they tend to cover a very large area, and also have to be carried out within strict budgetary limits. Under these circumstances it is essential to employ techniques that can rapidly examine large areas at relatively low cost. Hence, the interpretation of remote sensing can make a significant contribution to improving the effectiveness of regional engineering investigations of this type.

It has already been noted that the use of background information such as maps of geology, topography, soil or climate can aid considerably in the accuracy of an interpretation. In addition the use of terrain evaluation techniques enables the engineer to focus the interpretation on to the significant parts of the terrain, further increasing the efficiency of the interpretation. The objective of a terrain evaluation is to subdivide an area of ground in such a way that it is possible to record the terrain properties of interest to a particular investigation (see Mitchell, 1973).

The basis of any terrain evaluation involves the classification of the ground into distinct mapping units, selected to meet the following criteria.

(i) The units must divide up the terrain in a way that is relevant for the user of the map.
(ii) The units must be reasonably homogeneous in terms of the purpose that they are intended to serve.
(iii) The units must be convenient to map, in terms of the effort required and the resources available.

These criteria must be considered before deciding which mapping units are most appropriate. The mapping scale is largely determined by factors (ii) and (iii).

The end result of a terrain evaluation study is usually a thematic map, which may be purely geological or soil-type, but should preferably be of more direct relevance to the project in hand. The Geological Society (1982) has produced a comprehensive report on terrain mapping which outlines the various systems and recommends the use of air-photo interpretation based on a framework of terrain classification or geomorphological terrain mapping. The most widely used system of terrain classification for land resource mapping is the land system/land facet concept, devised for use in both engineering and agricultural surveys (Beckett *et al.*, 1972). The advantage of this joint technique is that data gathered by one group can be used by others. For example the maps and reports edited by Hill (1978) for the Land Resources Development Centre have provided the basis of many engineering studies in Nigeria.

The type of terrain classification used in these surveys is based upon the fact that terrains can be divided up in a hierarchical fashion, in which large units are made up of associations of smaller units, in a manner similar to soils. The largest units in the classification occupy many thousands of square kilometres; the smallest may cover only a few square metres. Land systems and land facets occupy a position of central importance in a scheme that includes more generalized as well as more specific levels. A land system is a large area of terrain having characteristic landforms, drainage pattern and associations of materials, developed on a single geological unit or sequence. It is typified by a distinctive scenery, land use and social pattern, and it can generally be mapped at about 1:250 000 to 1:1 000 000 scale. Land systems form distinctive air photo patterns caused by their arrangement of drainage, topography and land cover. The component parts of a land system, called land facets, are defined in a similar way, but they are smaller and less variable, so that an engineer would normally expect a single design to be appropriate for sections of road built on each facet. The number of land facets in a land system is generally between three and six, but they always occur in the same relationship. Land facets can be subdivided into land elements (the smallest members of the classification hierarchy) if these small features are considered to be important, as they often are in engineering studies. The relation between land system, land facet and

land element is shown in Figure 5.1. Terrain classification studies have been carried out in Africa (Scott *et al.*, 1971) and the Far East (Lawrance, 1978; CSIRO, 1976) as well as other areas of the world.

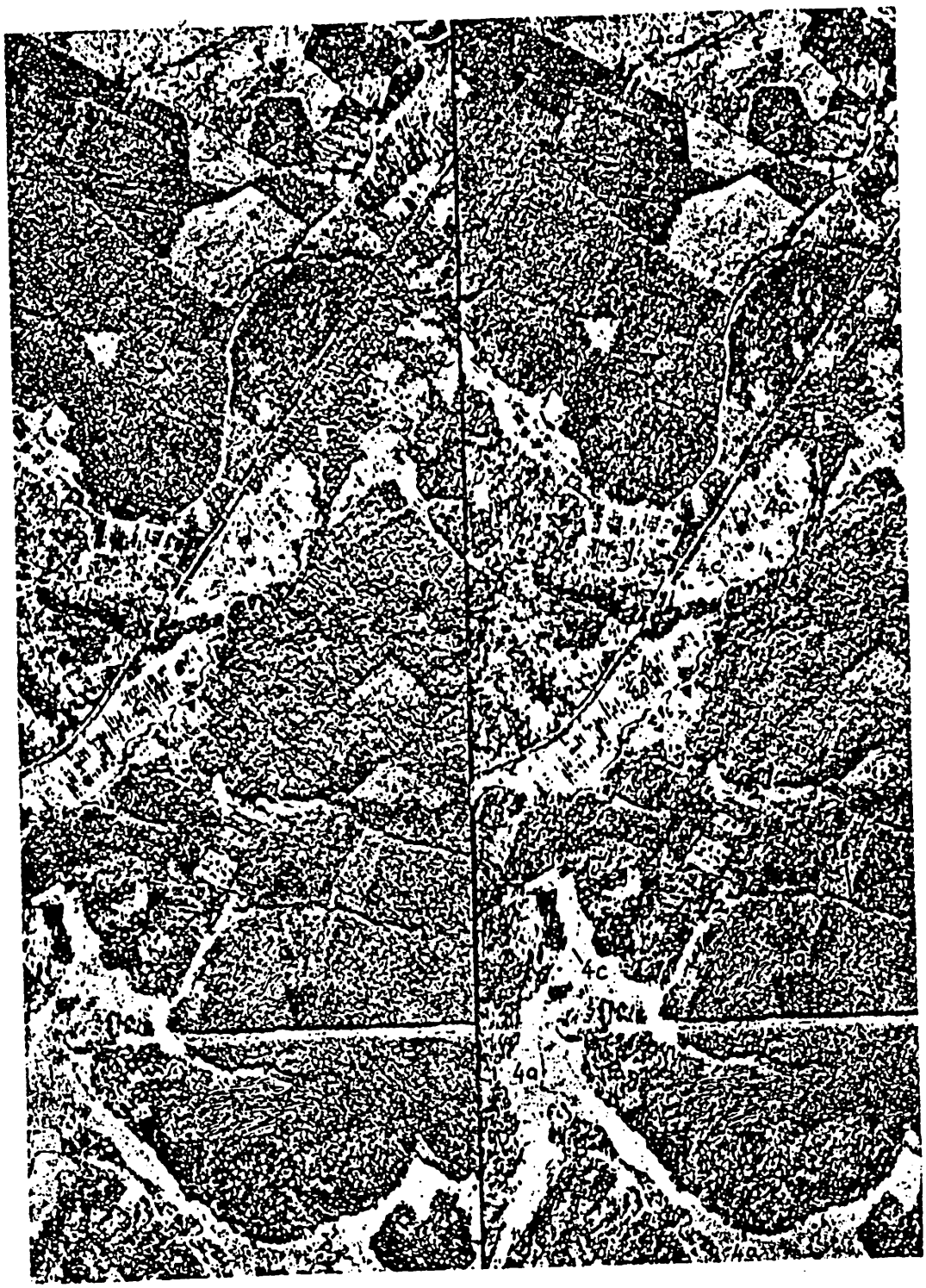
5.3.1 *Selection of appropriate techniques*

The interpretation of remote sensing imagery represents a technique that can be used to varying degrees in all types of survey. Its relative importance will depend on the scale of the survey, and also on the availability of other material. The availability of background material is different for every project and in some cases a reconnaissance survey is commissioned to collect data needed for planning. In addition the facilities available to an expatriate consultant may be different to those available to a local Ministry, e.g. the time scale of a project may make it impossible to seek out and collate data which is suspected to exist.

The importance of the interpretation of Landsat imagery for planning studies is set out in Chapter 7, with reference to a wide variety of surveys. Most of the studies described were carried out from Europe or America; one reason for this is that digital imaging processing techniques were used in several studies, and at present most of the equipment needed for this is both expensive and sophisticated and thus less readily available in developing countries. However photographic processing of Landsat data provides an effective source of imagery as demonstrated by the prints produced at the Regional Centre for Services in Surveying and Mapping in Nairobi. In addition the development of microcomputers and associated accessories means that small robust image processors are now being produced which give high-quality images suitable for interpretation in engineering surveys. In this way it will be possible to take the equipment to the country being surveyed, and by making the interpretation during the course of the project, demonstrate its usefulness locally and also enable more rapid correlation between interpretation and field checking. The theme of appropriate equipment is also discussed in Chapter 9 describing the use of light aircraft to take specialist photography for the interpretation of specific features. Although related to highway engineering, the techniques described are appropriate to many other types of survey. In addition to specialist surveys, this chapter describes the interpretation of satellite imagery and conventional air photography for highway projects ranging from a planning survey through feasibility and design studies to post-construction monitoring.

5.4 The use of remote sensing for projects in the developed world

Engineering projects in the developed areas of the world are associated primarily with the improvement and expansion of the infrastructure of that country. In such projects remote sensing techniques may be employed in a variety of applications including topographic mapping, site investigations,



(a)

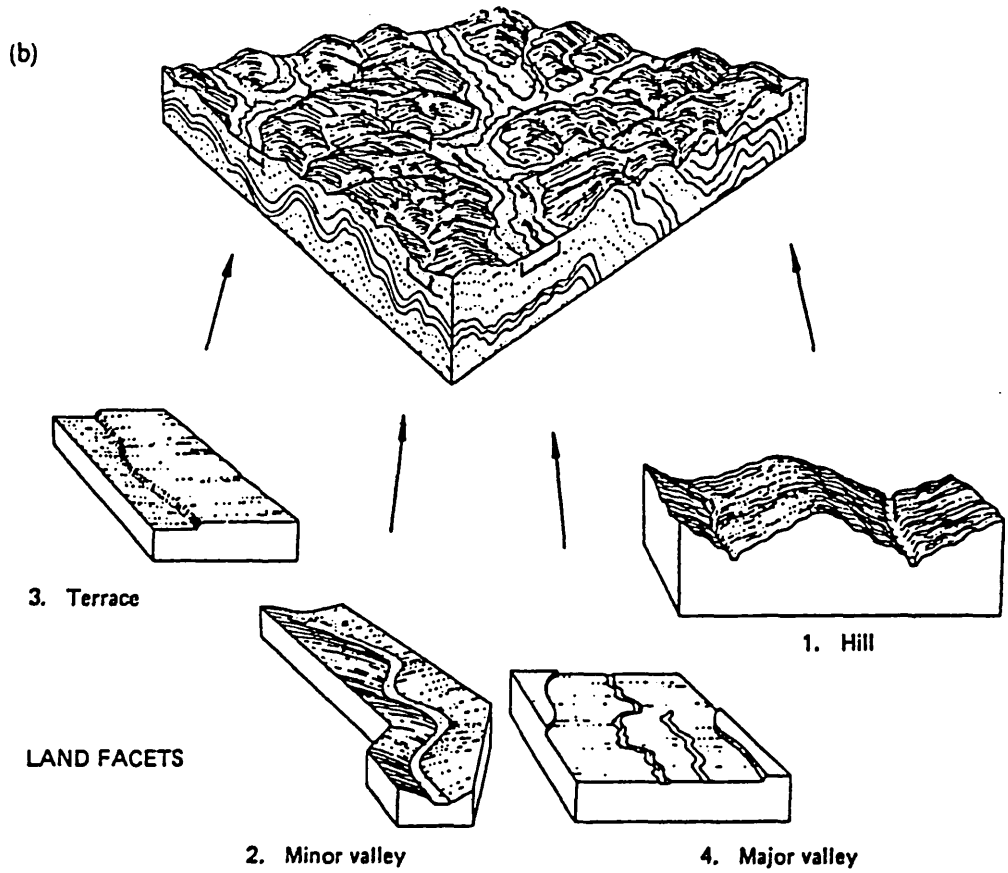


Figure 5.1 Stereopair and block diagrams showing the land facets of the Alor Gajah land system. (Air photos by courtesy of the Director of National Mapping, Malaysia, Government of Malaysia copyright reserved: DNMM 21.16/60).

and environmental engineering studies (e.g. energy conservation and pollution monitoring).

In topographic mapping applications much use is made of conventional black-and-white aerial photography. The demands on accuracy set by this application necessitate specialized equipment and skills. Expensive survey cameras of the types described in Chapter 2 are used to obtain the photographs and complex stereoplotting instruments are used in order to determine accurate three-dimensional coordinate data. The photogrammetric operations required for such processes are discussed in Chapter 6.

In contrast, the use of remote sensing for site investigations demands a much lower degree of accuracy with respect to the measurement process. In this case emphasis is placed on the interpretation of the imagery to locate features of geotechnical significance. The size of the project area will generally be small (from a fraction of a hectare to the order of one or two hundred hectares). In Europe and the USA small- and large-scale maps and documentary information are usually readily available, providing useful background information about the project area. However, the wealth of data may, in some instances, obscure the impact of some factor of vital importance, either

because it is too localized or specialized to be included in these data sources. A definitive image of the ground from an elevated platform can clearly provide comprehensive information which complements that obtained from other sources.The resolution of this imagery should however be compatible with the dimensions of the features of interest and hence at present aerial photography is the primary source of imagery for projects of this nature. Currently the spatial resolution of satellite-borne sensors is not sufficiently high to enable the detection of ground features likely to be of engineering significance in such investigations. The availability of existing aerial photography is limited largely to black-and-white panchromatic obtained for topographic mapping purposes. However, the type of photography used for this purpose need not be restricted to conventional black-and-white vertical photography, and may involve the interpretation of oblique photographs and those obtained using colour and 'false' colour infrared emulsions. Indeed, the increase in interest in small-format photographs obtained using remotely piloted aircraft (RPA) (Tomlins, 1983) and microlight aircraft (Graham and Read, 1984) may make the commissioning of oblique and colour or 'false' colour infrared photography a cost effective approach for small projects.

The benefits which can be derived from incorporating the interpretation of aerial photography into a site investigation project can be substantial. For example the information obtained by the aerial photography may be used for the following:

(a) The planning of the distribution of site investigation work in the field. Gates and breaks in hedges or fences, together with an overview of the general terrain to be covered, can be seen on photographs taken at the appropriate scale. This can be of great assistance in planning the movement of drilling rigs and other equipment over the site.
(b) The location and usage of fill, aggregates and other construction materials. The interpretation of the geology of the site and its environs will often provide this information.
(c) The identification of catchments, the evaluation of run-off, the pattern and density of streams and the location of spring lines and seepages. This information is normally clearly visible on aerial photographs and may be more comprehensive than that obtained from topographic maps.
(d) The identification of instability. Landslip activity whether recent or not can often be identified. Examination of photographs taken over a period of time may be used to define the most active zones in landslip areas. Also air photographs may be used to identify areas of potential instability.
(e) The determination of site history. Examination of the photographs may reveal evidence of the past use of the site or natural changes which may have affected the site. This may include the identification of landfill, areas of tree removal, and areas subject to occasional flooding. In coastal areas the rate of cliff recession may be apparent from a sequence of aerial

photographs. The use of aerial photography as a means of assessing site history is described in Dumbleton (1983).

(f) The identification of localized features of geotechnical interest. For example this may include features such as sinkholes, land drains, infilled trenches and natural drainage channels, disturbed drainage, gullies and abandoned mine shafts.

(g) The identification of general topographic and geological features in and around the project area.

Remote sensing also has an important role to play in environmental engineering. By using scanning systems it is possible to obtain data about objects outside the visible spectrum. This is particularly useful for energy conservation studies where heat loss is often an important area of interest. Thermal infrared techniques can be used very cost effectively to detect heat loss from structures and also from buried services (e.g. steam lines). Thermal and multispectral scanning systems may also be used for pollution and sediment transport studies. Clearly a high resolution is necessary for such investigations, and this, with a few exceptions, precludes the use of satellite-mounted systems. Hence the systems are generally mounted in an aircraft. More discussion relating to the use of thermal infrared imagery to environmental engineering in particular may be found in Chapter 10.

5.5 Future developments

Many of the recent and future developments in remote sensing have been discussed in the earlier chapters of the book. It is, however, worth examining how some of these technological advances may influence their future application in civil engineering.

One of the most significant advances which has occurred with satellite-based scanning systems has been the dramatic improvement in spatial resolution. This is clearly illustrated by Figure 5.2 which shows the progressive improvement which has occurred in the last decade.

For various reasons, both technical and political, it seems highly unlikely that there will be a major improvement in this area in the near future. Nevertheless one can envisage that substantial interest will be generated in satellite remote sensing when the full impact of the Landsat TM and SPOT data becomes apparent. Indeed the improved resolution and stereoscopic viewing facility of SPOT should prove to be an extremely useful aid for both reconnaissance and feasibility level investigations.

Turning attention to aircraft-based systems, it is possible to identify two major areas of future development. Firstly, from a photogrammetric viewpoint the possibility may exist in the near future for the use of data from high-resolution digital framing cameras which utilize solid-state areal arrays (Petrie, 1983). A radical development of this nature would have dramatic

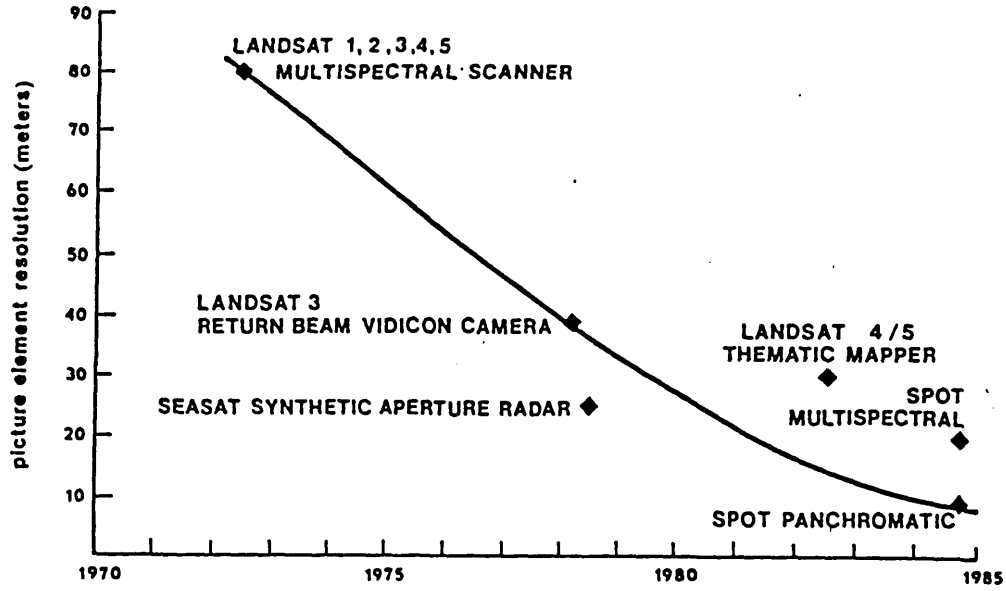


Figure 5.2 Improvement in resolution of satellite-based scanning systems. (Adapted from Settle, 1983).

repercussions on both the storage requirements for such data (e.g. video discs), and also on the measurement techniques which would be suitable (the video stereocomparator). The development of such a system would offer many benefits to mapping for civil engineering projects, some of which are currently applicable to analytical plotters. These include improved accuracy, greater flexibility both in terms of data input and also of output, generation of data for inclusion in geographic information systems, and finally decreased time and cost. A second future development related to airborne systems, specifically the scanner systems, will be an increase in the spatial and spectral resolution. Furthermore, as experience is gained in the analysis of high-resolution MSS imagery and coverage is extended and becomes more readily available the use of this source of data will become more widespread. Indeed it is quite possible that eventually imagery of this type will become a standard source of information for local site investigations and act in a complementary manner to conventional aerial photography.

Undoubtedly, however, the main advance which can be identified in the future, and one which will benefit both spaceborne and airborne systems, is the continued drop in the cost of computer memory and the increased speed of operation of computer systems. This will continue the movement of remote sensing away from analogue photographic methods to systems which make much greater use of digital technology. Two developments which are relevant in this respect are the development of low-cost image processing systems and also of geographic information systems. Brief mention has been made of the low-cost image processing systems in Chapter 4. This development will have

enormous impact on remote sensing and in particular on the expansion of remote sensing into fields such as civil engineering. The opportunity to use, at relatively low cost, only a few of the processing techniques discussed earlier will enable the user to make much greater use of the vast information content of most remotely sensed imagery than is possible by simple visual inspection. Finally brief mention should be made of geographic information systems (GIS). A GIS is a computer system designed specifically to accept large volumes of spatial data, derived from a variety of sources including remote sensing. The system is able to store, retrieve, analyse and display these data according to the requirements of the user and to create both tabular and cartographic output which reflects these requirements. Currently these systems are in their infancy and many problems have to be solved before they become widely available. They do, however, seem to offer an opportunity to make optimum use of remote sensing data by combining its unique attributes with those of other sources of spatial data. Further discussion of GIS may be found in Marble and Peuquet (1983).

5.6 Conclusions

Remote sensing systems whether mounted in an aircraft or a satellite provide a comprehensive overview of the ground surface and are thus a cost effective source of data for engineering investigations. These data, which are normally displayed in the form of an image, may be used either for the precise determination of planimetry and height or for the qualitative analysis of ground type and condition. The former application is restricted largely to the use of vertical aerial photography, whilst the latter commonly involves the examination of all forms of remotely sensed imagery.

Five stages of engineering investigation have been identified: reconnaissance, feasibility, design, construction and post-construction. In developing countries projects tend to be on a regional scale, often with limited background information, hence greater emphasis is placed on the reconnaissance and feasibility level investigations. In such cases small-scale aerial photography and/or satellite imagery is applicable. In contrast the needs of the industrialized nations are different and hence greater emphasis is placed on the design, construction and post-construction stages. In terms of remote sensing large-scale aerial photography and airborne scanning systems are the most appropriate forms of imaging.

The chapters which follow provide a state-of-the-art review of past and current applications of remote sensing in civil engineering. It is hoped that this will prove to be an interim review and that as future remote sensing systems become operational and the costs of processing and storing images are reduced, so the range of applications and their general usage will increase accordingly.

References

Adolfsson, B. (1972) 'The adaptation of photogrammetric techniques to road design needs', in *Computer Systems in Highway Design* (ed. Turner, A. K.) Proceedings NATO Advanced Study Course, Copenhagen, Denmark, pp. 52–64.

Beckett, P. H. T., R. Webster, G. M. McNeil and C. W. Mitchell (1972) Terrain evaluation by means of a data bank. *Geographical Journal* 138(4) 430–456.

CSIRO (1976) *Lands of the Ramu-Madang area, Papua New Guinea.* Land Research Series 37, Commonwealth Scientific and Industrial Research Organisation, Australia.

Dumbleton, M. J. (1983) *Air Photographs for Investigating Natural Changes and Present Conditions of Engineering Sites.* Dept. of the Environment, Dept. of Transport, Transport and Road Research Laboratory Report 1085, Crowthorne, Berks.

Geological Society Working Party (1982) Land surface evaluation for engineering practice. *Q. J. Eng. Geol.* 15, 265–316.

Graham, R. W. and Read, R. (1984) Small format aerial photography from microlight platforms. *J. Photographic Science* 32, 100–110.

Hill, I. D. (ed.) (1978) *Land Resources of Central Nigeria.* Ministry of Overseas Development, Land Resource Study 29, Land Resources Development Centre.

Lawrance, C. J. (1978) *Terrain Evaluation in West Malaysia, Part 2, Land Systems of South West Malaysia.* Dept. of the Environment, Dept. of Transport, Transport and Road Research Laboratory Report SR 378, Crowthorne, Berks.

Leatherdale, J. W. (1975) Surveys for engineering design. *The Consulting Engineer...*

Marble, D. F. and Peuquet, O. J. (eds.) (1983) 'Geographic information systems', in *Manual of Remote Sensing*, Vol. 1 (ed. Colwell, R. N.), American Society of Photogrammetry, pp. 923–958.

Mitchell, C. W. (1973) *Terrain Evaluation*, Longman, London, 165 pp.

Petrie, G. (1983) 'The philosophy of digital and analytical photogrammetric systems', in *Proceedings of the 39th Photogrammetric Week*, University of Stuttgart, W. Germany.

Scott, R. M., C. J. Lawrance, R. Webster and F. H. T. Beckett (1971) *Atlas of the Land Systems of Western Kenya.* Military Engineering Experimental Establishment, Christchurch, Hampshire.

Settle, M. (1983) 'Current trends and research challenges in land remote sensing', in *Proceedings Alpbach Summer School, European Space Agency Report* ESA SP-205, 3–9.

Tomlins, G. F. (1983) Some considerations in the design of low cost remotely piloted aircraft for civil remote sensing applications. *The Canadian Surveyor* 37, 157–167.

GEOTECHNICAL APPLICATIONS OF REMOTE SENSING AND REMOTE DATA TRANSMISSION

A symposium
sponsored by
ASTM Committee D-18
on Soil and Rock
Cocoa Beach, FL, 31 Jan.-1 Feb. 1986

ASTM SPECIAL TECHNICAL PUBLICATION 967
A. I. Johnson, International Committee on
Remote Sensing and Data Transmission, and
C. B. Pettersson, Brown and Root, Inc., editors

ASTM Publication Code Number (PCN)
04-967000-38



1916 Race Street, Philadelphia, PA 19103

Thomas J. M. Kennie[1] *and Clive N. Edmonds*[2]

Location of Potential Ground Subsidence and Collapse Features in Soluble Carbonate Rock by Remote Sensing Techniques

REFERENCE: Kennie, T. J. M. and Edmonds, C. N., "Location of Potential Ground Subsidence and Collapse Features in Soluble Carbonate Rock by Remote Sensing Techniques," *Geotechnical Applications of Remote Sensing and Remote Data Transmission, ASTM STP 967*, A. I. Johnson and C. B. Pettersson, Eds., American Society for Testing and Materials, Philadelphia, 1988, pp. 99-121.

ABSTRACT: The occurrence of subsidence and collapse dolines in soluble carbonate rock is a frequent geotechnical hazard throughout the world. Such features pose serious geotechnical problems to engineering and building construction and also to public safety.

Detection methods utilized during preliminary site investigations range from direct methods, such as drilling, to indirect methods, such as ground geophysics and airborne remote sensing.

The aim of this paper is to examine the role of remote sensing in these preliminary site investigations. Previous and current remote sensing research concerned with the location of solution features is reviewed, with particular emphasis on airborne multispectral and thermal infrared line scanning. A comparative assessment of the relative merits of the different techniques is given, together with an examination of the factors which influence whether a particular technique will be successful.

The authors recommend that the most favorable detection results will be obtained if remote sensing methods are used under optimum environmental conditions and in appropriate geological conditions.

KEY WORDS: ground subsidence, ground collapse, soluble carbonate rock, solution features, geotechnical problems, detection, remote sensing, aerial photography, thermal infrared line scanning, multispectral scanner

The widespread occurrence of soluble carbonate rock is discussed with special reference to Great Britain. Various types of natural solution features are present in limestone outcrops. They have characteristic forms and, to some extent, their development depends on whether the limestone is weak and chalky or strong and crystalline. According to the genesis of natural solution features, they can be of a stable or metastable nature. Stable forms do not pose serious subsidence risk to engineering and building structures but could result in some differential settlement. Metastable forms can, however, pose a serious subsidence risk, resulting in severe stress to structures or even complete destruction by collapse. This paper presents a review of the various forms of solution features and their geotechnical significance.

It would be advantageous to be able to detect metastable forms prior to urban development, and airborne remote sensing is suggested as a good, cost-effective precursor to ground investiga-

[1]Lecturer in engineering surveying, Department of Civil Engineering, University of Surrey, Guildford, Surrey, England GU2 5XH.

[2]Senior engineering geologist, Applied Geology, Warwickshire, England CV8 1FB

tion by indirect geophysical or direct exploratory techniques. The background experience gained in using the various remote sensing techniques is reviewed. Finally, in order to apply remote sensing techniques effectively as part of a site investigation, it is necessary to have an appreciation of the factors which influence the detection success of the various techniques; this topic is discussed later in this paper.

Geotechnical Significance of Soluble Carbonate Rocks

Forms of Solution Features

In Great Britain there is a wide diversity of soluble carbonate rock of various different geological ages (Fig. 1). Solution features form in such rock as a result of dissolution by aggressive surface and ground water. They are broadly divisible into surface and subsurface karst features. Surface karst features are present in the surface zone of the limestone bedrock and can also be present within the thickness of any superficial deposits. The subsurface karst features are those present below the surface zone of the limestone bedrock, deeper within the body of the host rock. Both Sweeting [*1*] and Bogli [*2*] discuss surface and subsurface karst landforms in some detail.

The remote sensing techniques to be discussed later in this paper are primarily concerned with the detection of surface karst features. The forms of such features have been widely discussed in the literature. However, one of the most problematic aspects is the lack of a series of unambiguous and well-defined terms to describe solution features. The difficulties associated with terminology have been identified by previous authors (for example, Walsh et al. [*3*], and Beck [*4*]). To avoid confusion, this section, based on field studies in Great Britain, outlines the terminology used by the authors to describe solution features.

Four distinct forms of solution feature can be identified (Fig. 2):

(*a*) *Solution pipes*—cone-like or pipe-like cavities extending from the surface downwards in vertical section. These cavities are typically infilled with overlying deposits that have subsided into the cavity by dissolution of the soluble limestone bedrock. They are generally circular or oval in plan and at the ground surface may or may not be overlain by a closed depression or doline. These can occur in isolation or in groups [*5*]. This form is mainly associated with weak, porous chalky limestones (Fig. 2, E, F, and I).

(*b*) *Solution widened joints*—V-shaped cavities extending from the surface downward in vertical section. Again, these cavities are generally infilled with overlying deposits. They are linear in plan, pinching and swelling along their length, and may or may not be associated at the ground surface with a closed depression or doline. These cavities can occur in isolation or they may be linked with others. This form is mainly associated with strong crystalline limestones. In cross section they appear similar to solution pipes.

(*c*) *Dolines*—closed depressions having a cone-like, bowl-like, or pipe-like shape in cross section, generally circular or elliptical in plan. Dolines can occur in isolation or in groups. The term "doline" is used solely as a surface morphological descriptive term without its genesis being inferred [*5*] (see Fig. 2, D and G.)

(*d*) *Swallow holes*—surface features where a void in the limestone below, continuously or intermittently, "swallows," wholly or partially, a surface stream—that is, the point at which the stream disappears from the surface to flow underground. Swallow holes can occur in isolation or in groups and are often recognizable as hollows along a valley floor stream channel. Although the general description of a doline includes a swallow hole, the latter term is preferred because it specifies the genesis of this particular type of doline. Thus, the term doline is only used in this paper to describe natural karstic closed depressions other than swallow holes (Fig. 2, A and 2B).

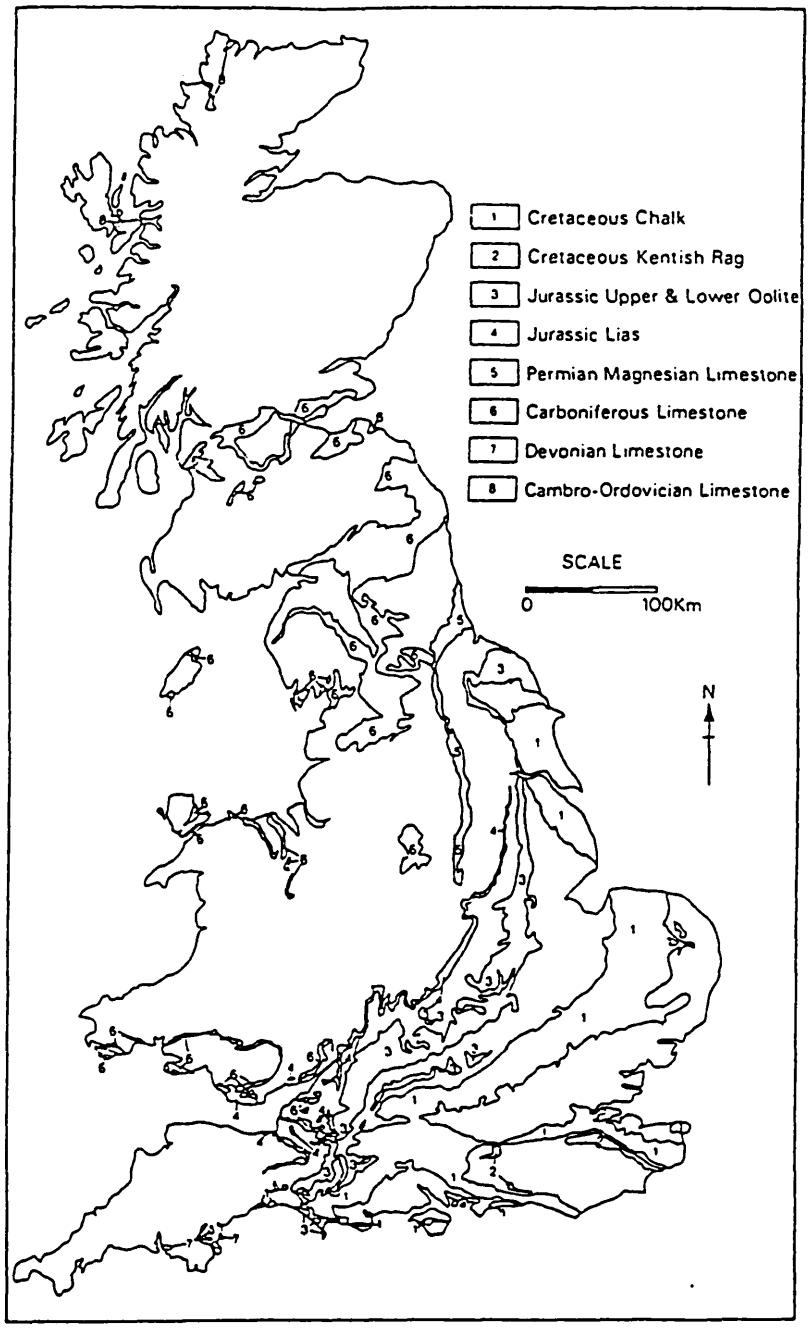


FIG. 1—*Distribution of soluble carbonate rock in Great Britain.*

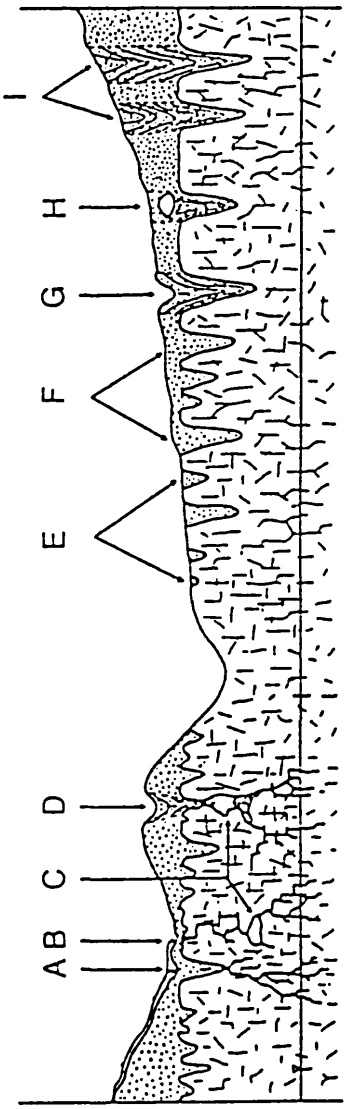


FIG. 2—*Schematic diagram to illustrate range of natural solution features present on limestone.*

Development of Metastable Forms of Solution Features

Dissolution of a limestone bedrock during the formation of surface and subsurface karst features can lead, in certain circumstances, to the formation of metastable forms of solution features. Destabilization of these can result in the production of ground subsidence hollows or dolines. Four main genetic forms of doline are recognized by Sweeting [*1*] (Fig. 3) as occurring in limestone. The presence of doline Types A and B in chalk outcrop is, however, uncertain, the majority being of Types C and D.

(*a*) *Solution dolines*—formed when limestone bedrock discontinuities are widened by dissolution beneath a soil veneer. As the discontinuities widen, the soil settles down into them, resulting in a localized surface lowering forming a shallow closed depression (Fig. 3, A).

(*b*) *Alluvial dolines*—formed when major limestone bedrock discontinuities are widened by concentrated dissolution beneath a cover deposit. As the joints widen, the deposit settles down into the open joints, and the gradual subsidence produces funnel-shaped depressions at the ground surface (Fig. 3, B).

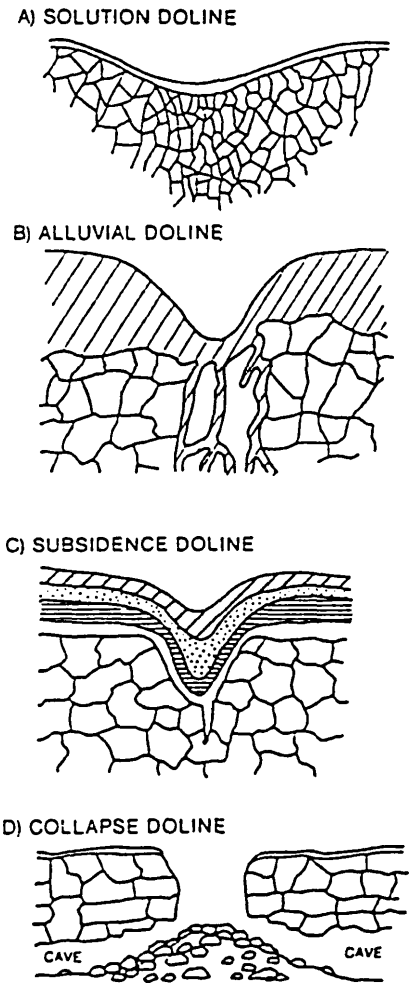


FIG. 3—*Types of doline (after Sweeting* [*1*]*).*

(c) *Subsidence dolines*—formed by dissolution of a limestone bedrock below a consolidated cover stratum or unconsolidated cover deposit. In this case, dissolution of the soluble bedrock takes place from the limestone rockhead interface downward along discontinuities. As the discontinuities widen by dissolution to form solution cavities, gravitational solution subsidence of the cover down into the enlarging cavities takes place. If the cover is a consolidated jointed rock stratum, then joint-bounded blocks of the rock stratum will only subside once the cavities in the limestone bedrock are greater than the joint pattern spacing of the rock cover stratum. Otherwise, the joint-bounded blocks will bridge the solution cavities. If the cover is an unconsolidated deposit, then it will subside at a more even rate, generally keeping pace with the creation of solution cavity space in the limestone bedrock. As the cover deposit loses basal support and subsides downward, the effects of the motion gradually migrate upward, eventually forming a closed depression at the ground surface (Fig. 3, C).

(d) *Collapse dolines*—formed by the dissolution of a limestone bedrock, which may or may not be overlain by a cover deposit. In this case dissolution of the soluble bedrock (vadose and phreatic) takes place below the limestone bedrock rockhead, forming natural caves. There comes a point when the stable span of a cave is exceeded, causing collapse of a cave roof, often in a single catastrophic movement. This collapse process can produce craters on the surface of bare limestone outcrops or cover strata if present (Fig. 2, D and Fig. 3, D).

The principal doline types encountered in both strong, crystalline limestones and weak, chalky limestones in Great Britain are the subsidence doline and collapse doline. The majority of dolines formed in Cretaceous Chalk, appear to be of the subsidence type, since the development of natural caves is poor [*6*]. During solution subsidence of unconsolidated Tertiary and Quaternary deposits overlying the chalk, downward subsidence into solution pipes may be resisted by a denser stronger layer. As the material below continues to settle, a metastable cavity is created, roofed over by a resistant layer (Fig. 2, H). The survival of this metastable cavity is largely governed by the *in situ* ground stresses and ground water movement. In time, the cavity roof collapses, either in a single movement or in multiple movements, which transmit the void upwards to the surface, resulting in ground subsidence, referred to as a subsidence doline (Fig. 2, G and Fig. 3, C).

Swallow hole locations represent sites of presently active solution and erosion of chalk or any other limestone bedrock. It is also possible that metastable cavities and solution cavities (both open and infilled with soft compressible materials) may be present in the vicinity of swallow holes, within the cover deposit and the limestone bedrock, respectively.

The results of this recent research complement the findings of other studies by Jennings et al. [*7*], Sowers [*8*], Brink [*9*], and Newton [*10–13*].

Geotechnical Problems

Where limestones are devoid of a superficial deposit, the potential hazard that solution features could pose to civil engineering and building projects is often visible. However, where a superficial cover deposit is present, the solution features posing a subsidence risk may not be obvious. Solution features formed within the surface zone of the limestock bedrock usually present a greater hazard than those at deeper levels within the body of the limestone bedrock. Thus, ground surface loading would not normally overstress a cave roof, leading to collapse, unless the cave system was at a shallow depth [*14*].

Metastable forms of solution features can, however, provide an unreliable bearing capacity, possibly leading to a foundation failure of the structure. Such subsidence case histories are widely reported in the literature in Europe, South Africa, Japan, and the United States [*15–18*]. Remedial measures such as grouting and underpinning can be undertaken to stabilize a structure that has suffered subsidence damage, but this can involve considerable expense. Occasionally, catastrophic damage is suffered by the structure and it has to be demolished, or alterna-

tively, the cost of stabilizing the site is uneconomic, leading to abandonment of the site. There is also the inherent danger to the public; a number of lives have been tragically lost because of unexpected ground collapses.

There is good evidence based on recent studies in Great Britain [*6*] to suggest that several subsidence triggers can act on metastable solution features to produce ground subsidence; these trigger mechanisms include:

(*a*) heavy rainstorms;
(*b*) the introduction into the ground of excessive concentrations of water via field drains, leaking water supply pipes or drains, soakaways, and garden hoses;
(*c*) artificial ground water lowering; and
(*d*) active or static surface loading.

Trigger *a* is associated with both urban and rural areas, while Triggers *b*, *c*, and *d* are mainly associated with urban areas. Once the natural drainage and metastable ground conditions are disturbed by development, the subsidence triggers listed may act to induce doline formation. Newton [*10-13*] also discusses similar trigger factors which can lead to the formation of "induced sinkholes." Without disturbance by development, the preexisting ground conditions would normally remain stable for a much longer time.

Two typical examples of induced doline formation are shown in Figs. 4*a* and 4*b*. It should be noted that these situations are only a representative sample of the possible subsidence mechanisms which may cause foundation failure. The diagram does, however, emphasize that several explanations may account for the same type of structural damage and, more importantly, that an understanding of the different mechanisms is essential if the correct remedial action is to be implemented. Sowers [*19*] gives a comprehensive review of the possible remedial measures available to the engineer to arrest structural damage.

Detection Methods

There are many different ways to detect the presence of natural solution features in limestone terrain. The success or failure of a particular detection method depends principally on detecting contrasts in physical character between the solution features and the surrounding ground. Whether or not these features have a surface expression can also be important. Generally, detection methods can only successfully identify features in plan or vertical section. Therefore, two or more successful methods have to be combined in order to achieve a three-dimensional picture of the feature being studied.

Three distinct methods of locating solution features exist: remote sensing, ground geophysics, and direct ground investigation. The use of these detection methods in normal site investigation is illustrated in Table 1. In broad terms, the distinction can be made between direct investigation techniques, such as drilling, and indirect techniques, such as remote sensing and ground geophysics. Direct techniques may provide conclusive evidence of the existence of features such as subsurface cavities, but they are expensive to operate over large areas, and are only appropriate for the detailed inspection of small sites. Indirect techniques, in contrast, are most cost-effective for examining large areas. Three distinct indirect techniques exist:

(*a*) airborne and satellite remote sensing,
(*b*) airborne and ground geophysics, and
(*c*) downhole geophysical techniques.

The following section will only be concerned with a review of the first technique. For further details about the latter two techniques the reader is directed to Fountain, Herzig, and Owen [*20*], Kirk and Werner [*21*], and McCann et al. [*22*].

Remote sensing should therefore be considered primarily a reconnaissance tool at the desk

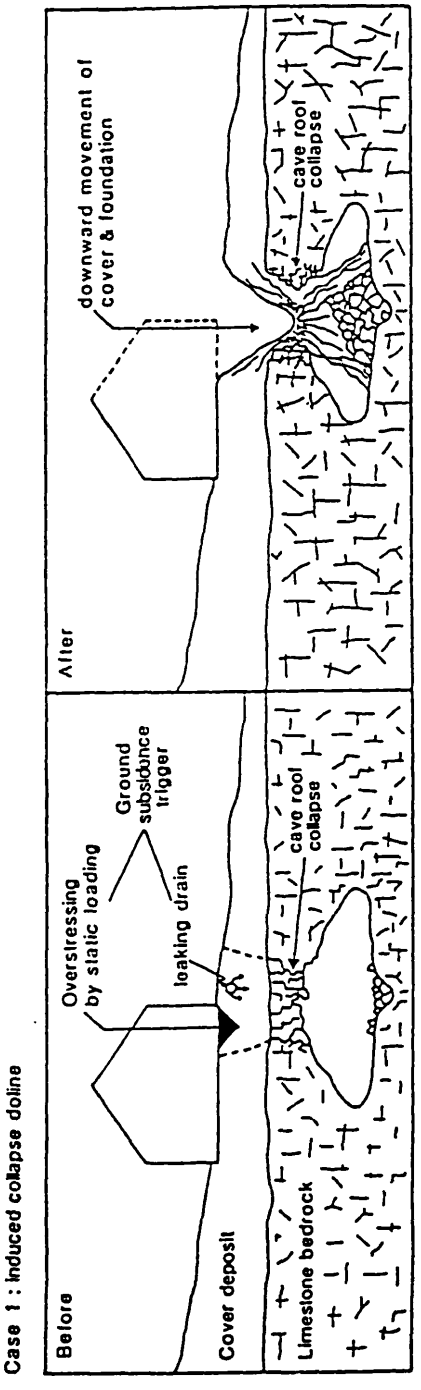


FIG. 4a—*Example of a foundation failure caused by an induced doline: Case 1—induced collapse doline.*

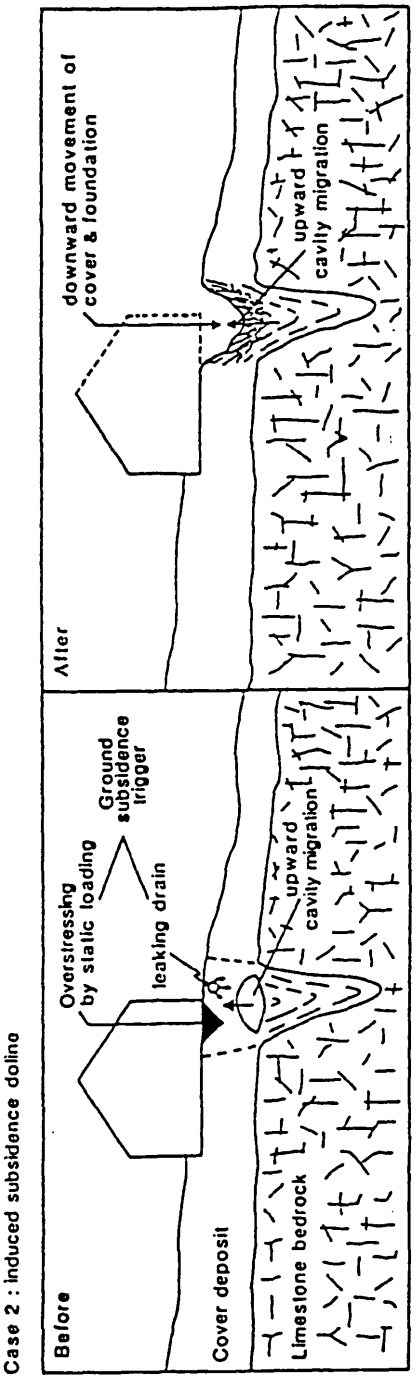


FIG. 4b—*Example of a foundation failure caused by an induced doline: Case 2—induced subsidence doline.*

TABLE 1—*Relationship between stages of site investigation and methods of detecting solution features.*

Stage of Site Investigation	Objective	Preferred Detection Method
Desk study	preliminary assessment of ground conditions and identification of subsidence risk target areas	remote sensing
Ground investigation	detailed investigation of target areas before construction	ground geophysics direct drilling

study stage rather than a technique to use during the detailed ground investigation stage. The aim of a survey utilizing remote sensing techniques should be to delineate areas where potentially metastable features may exist.

A Review of Remote Sensing Research Concerned with the Detection of Solution Features

The use of airborne remote sensing for the detection of surface and, to a lesser extent, subsurface solution features has been a subject of considerable interest for the past 20 years. The aim of this section of the paper is to provide a review of work in this field. In particular, emphasis is placed on the role of airborne and satellite multispectral scanners (MSS), airborne thermal infrared line scanners, and airborne and ground-based microwave radiometers. The importance of aerial photography should, however, not be overlooked, and the following section briefly reviews previous work in this field.

Aerial Photography

Aerial photography is the oldest and also the most frequently used source of remote sensing data for site investigations concerned with the location of solution features. Its use should not be underestimated because of the synoptic view provided by aerial photography, its high spatial resolution, and its ability to view overlapping photographs stereoscopically (Fig. 5). Furthermore, it is a widely available source of data at relatively low cost. A broad distinction can be made between "direct" evidence provided by aerial photography and "indirect" evidence of the existence of a potential solution feature.

Direct Evidence—Norman and Watson [*23*] have suggested that the following interpretation criteria (among others) could be used to provide direct evidence of a potential subsidence feature:

> dark tonal patches with sharp edges, usually rounded, small localized depressions, centripetal drainage over a small area, and the sudden disappearance underground of drainage.

West and Dumbleton [*24*] also discuss the use of aerial photography as a means of providing direct evidence. In particular, they stress the importance of stereoviewing to help exaggerate the surface morphological expression of dolines.

Indirect Evidence—One indirect criterion which has been suggested by several authors as having potential for locating solution features is the existence of stressed vegetation [*23-27*]. In this context, "stressed vegetation" refers to the abnormal growth patterns which may occur within vegetation caused by either an excess or deficiency of moisture. The delineation of vegetation which has been stressed by this process can be carried out by examining the reflectance characteristics of the vegetation in the near-infrared (IR) region of the electromagnetic (EM) spectrum. These reflectance patterns can be recorded photographically using a "false" color-infrared emulsion. Newton, Copland, and Scarborough [*26*], for example, successfully used this technique to delineate potential areas of collapse in a limestone region of Alabama.



FIG. 5—*Panchromatic vertical aerial photograph of Southover Heath, Dorset, England, illustrating solution pipes (A) and subsidence dolines (B) on cretaceous chalk (reproduced by permission of Aerofilms, Ltd.).*

A second criterion which has been suggested by other researchers as useful indirect evidence of the existence of areas susceptible to collapse is the existence of fracture traces or lineaments [*28-35*]. Fracture traces and lineaments have been defined [*33*] as natural linear features (topographic, vegetation, or soil tonal alignments) expressed continuously in aerial photography or satellite imagery. The distinction between the two terms lies in their magnitude, lineaments normally being restricted to features that extend for several miles. It is suggested by those who have adopted this approach that fracture traces and lineaments may be related to the positions of minor faults, regions of concentrated jointing, increased weathering, and high permeability and, hence, may indicate solution zones. It is also suggested that where a high density of traces or lineaments exists and, in particular, where two or more intersect, that the risk of subsidence will be greater. The main difficulties associated with fracture analysis are, first, the problem of delineating traces and lineaments which are specifically related to dissolution of carbonate bedrock and, second, the difficulty of objectively interpreting the data to enable subsidence risk regions to be located. Fracture analysis cannot be used by itself to predict precisely where subsidence is likely to occur. It may, however, in a limited number of situations, have a supplementary role to play in identifying areas where caution should be exercised in land use planning.

Multispectral Scanners

Both airborne and satellite based MSS systems have been evaluated by previous researchers. In the case of airborne MSS, conflicting results have been reported. Coker, Marshall, and Thompson [*36*], for example, investigated the use of an 18-channel MSS system over an area near Bartow, Florida. The aim of the investigation was to delineate zones of moisture-stressed vegetation, which were thought to be associated with the positions of sinkholes. The project concluded that sinkhole-prone areas could be located by this process. Stohr and West [*37*], in contrast, using a supervised classification of airborne MSS data obtained over an area in Virginia, "distinguished many cover types but failed to delineate sinkholes within the cover type containing them."

To date, the use of satellite MSS data has been restricted to Landsat MSS. Clearly the resolution capabilities of such systems do not enable individual solution features to be detected. Several authors [*31,32,35*], however, have reported on the use of Landsat for mapping regional structural patterns and lineaments, which may indicate regions that require further investigation.

Casper, Ruth, and Degner [*31,32*], for example, used Landsat MSS to examine the spatial distribution of sinkholes in Florida. This study was carried out by examining the positions of the sinkholes in relation to lineaments or fracture traces that were expressed on aerial photographs and satellite imagery. Both high-altitude aerial photography and Landsat MSS were used. The authors report "a strong correlation was found to exist between the positions of sinkholes and collapses and the intersections of the lineaments." They concluded the report by stating that, while it is not possible to predict sinkhole occurrence based solely on the data supplied by the fracture trace studies, the information nevertheless could be very useful for locating zones with the greatest potential for sinkhole development and, hence, could be used as an additional source of data for land use planning.

A similar approach was used by Sësoren [*37*]. In this case the area of interest was in Southern Turkey and interpretation was based on multitemporal Landsat MSS imagery at scales of 1:1 000 000 and 1:400 000.

Thermal Infrared Line Scanning

A substantial body of research has been carried out over the past 15 years on the potential use of thermal IR line scanning for the detection of solution features. Figure 6 summarizes the

details of these investigations and the relative levels of success that have been reported. This figure should be examined in conjunction with the following text.

Parker [*38*] was one of the first researchers to consider the use of thermal IR line scanning for engineering applications. He noted that while the technology for producing thermal images had been available for several years, little of engineering significance had been reported. Parker further concluded that although several papers could be cited to illustrate the potential of the technique, the economic benefits to be gained from such a system had yet to be demonstrated. Parker also suggested some key questions that needed to be answered in order to remedy this situation, which to a large extent still remain unanswered. These questions, which are highly significant and worth repeating, are the following:

1. How do the complex interactions of material properties and environment affect the radiation output and image characteristics of terrain features and materials?
2. Do they produce detectable thermal differences?
3. What are the optimum temporal and environmental conditions for recording these?
4. Are the results sufficiently predictable over a range of materials and environmental conditions?
5. What is the engineering significance of the information obtained?

Following this study and for a period of five to six years many other North American researchers considered the potential of this technique for locating solution features. The reports were varied in nature and often led to conflicting conclusions about the usefulness of the technique. Stingelin [*39*], for example, reported that infrared imagery did not prove to be significantly better than aerial photography in locating solution cavities and sinkholes. No details were reported regarding the type of instrument used, the conditions during data acquisition, or the method of analysis.

A more rigorous analysis of the problem was reported by Stohr and West [*37*]. Imagery was obtained over a limestone test site known to contain three sinkholes. (See Fig. 6 for details.) The main conclusion from this work was that, when using both manual and computer-assisted classification techniques, thermal infrared line scanning alone was not capable of delineating the positions of known sinkhole locations. Stohr and West further suggested that thermal imagery should only be used to provide supplementary evidence to the location of solution features in limestone terrains.

A further, more extensive review of the effectiveness of various remote sensing systems for detecting subsurface cavities in particular was provided by Rib et al. [*40*]. This research was carried out in the state of Kansas by a consortium consisting of the State Highway Commission, the Federal Highway Administration, and several contractors, including the Environmental Institute of Michigan. The systems evaluated included various types of aerial photography, an 18-channel MSS, a dual-channel thermal IR line scanner, a microwave radiometer, and a scatterometer profiling system. The emphasis of the data analysis was to assess the capability of the microwave radiometer and the infrared imagery to detect the presence of a subsurface limestone mine located in the test area. The overburden over the test site varied from 3 to 35 m in depth. Several subsurface cavities also existed in the test site. The conclusions reached by the researchers which are significant to thermal IR line scanning were the following:

1. No direct indication was noted on the imagery of the presence of the subsurface mine. Indications were only noted by secondary indirect criteria, such as variations in moisture content.
2. Use of enhancement techniques such as "thermal contouring" and "thermal ratioing" did not produce patterns related to the presence of the subsurface mine.

The authors concluded that more work was required to determine the optimum diurnal and seasonal times for data collection. Furthermore, consideration should be given to including in

Authors (Year)	Data Acquis. Date	Time	Instrumentation Manuf.	Spectral range (μ m)	Flying height (metres)	Environmental Conditions During Data Acquisition	Image Processing	General Comments and Level of Success
Coker[36] (1969)	6th Sep	0647	Univ. of Michigan MSS	8-14	600	Data obtained during a drought	Density sliced (optically)	Data combined with near IR MSS data on veg. stress * * *
Metalucci + Abdel Hady[41] (1969)	-	-	Bendix LN-1	3-5.5	-	-	-	- * *
Noble[43] (1972)	April Dec.	700 - 1700	Univ. of Michigan MSS	4.5-5.5 10-11 8-13.5	600 and 1500	Raining during Dec. night-time flight	None	Ground radiometer measurements obtained *
DeLoach[42] (1972)	Nov.16	Night	-	8-14	300	-	Mainly visual but also microdensitometry	Interp. followed by boring * * *
Newton et al[26] (1973)	Jan.18 Mar.14	day/night day/night	-	8-14	500	Night flights suffered from effect of mist	None	Used primarily for lineament mapping *
Stohr and West[37] (1974)	Apr.7, 18, Dec.14, 15	1100-1200 2300-0200	Univ.of Michigan MSS	8-14	600 and 1500	-	Supervised and cluster analysis	Data combined with MSS *
Rib et al[40] (1973)	Mar.24 Mar.27 Apr.4	midday and night	Texas Ins. RS-14 ERIM MSS	8.2-10.7 9.4-12.1 8-14	450	Snow cover melted between 27 March and 4 April	Thermal contouring Thermal ratioing	Ground based radiometer readings *
Rinker[44] (1975)	Jun.11	0100	-	8-14	-	Clear sky, little wind	None	- * *
Harvey et al[45] (1977)	Mar.8, 9 Jun.20, 21	1200 and night		8-14	500	Data obtained during cloudy weather	None	Midday late summer suggested as optimum time *
KEY	* of limited use		** useful		*** very useful			

FIG. 6—*Summary of previous research: thermal IR linescanning.*

Authors (Year)	Data Acquis. Date	Time	Instrumentation Manuf.	Spectral range (μ m)	Flying height (metres)	Environmental Conditions During Data Acquisition	Image Processing	General Comments and Level of Success
Bakker et al[46] (1978)	-	Pre-dawn	Texas Inst. RS-25	3.5-5.5 8-14	1000	-	None	Used in conjunction with API * * *
Warwick et al[47] (1979)	-	Pre-dawn	Texas Inst. RS-25	3.5-5.5 8-14	1000	-	None	Used in conjunction with API * * *
Marino et al (1983)[49] 1984)[50]	Nov.	Pre-dawn noon	Daedalus 1230	1-2 3-5 9-11	800	-	Density slicing, texture analysis, ratioing	- * * *
KEY	* of limited use ** useful *** very useful							

FIG. 6—*continued.*

the data analysis ground temperature measurements obtained during the data acquisition period.

More encouraging results have been reported by several other researchers. For example, Coker et al. [*36*] examined the problem over a test site in Florida. By combining thermal infrared line-scanner temperature patterns with moisture-stressed vegetation patterns (obtained by airborne MSS), Coker and his co-workers reported evidence of significant correlations between such data and areas of known sinkhole activity. He was also bold enough to suggest that "these techniques appear to have potential for locating hydrogeological phenomena occurring as much as 100 ft (~30 m) below the surface of the earth."

Matalucci and Abdel-Hady [*41*] also made mention of the success of thermal infrared line scanning in locating natural subsidence features, such as near-surface voids and "muck pockets," and also man-made features, such as drainage channels, pipelines, and conduits. In common with many of the reports, no details are provided about the type of instrument or of the environmental conditions which existed during data acquisition.

Further encouraging results from the United States were reported in DeLoach [*42*]. DeLoach reported "very positive correlations" between anomalous cool patterns observed on nighttime imagery and the positions of deep fissures in the bedrock or of cavities above the water table. The interpretation of these data were solely visual, with no computer processing. Of 16 borings made in regions selected by interpretation of the thermal imagery, 12 led to the detection of solution features considered hazardous to foundation design.

Noble [*43*] also reported on the use of this technique for the location of sinkholes, in this case in Virginia. Using both day and nighttime imagery, together with aerial photography, he was able to detect several features which exhibited a medium grey or "cool" tone on the nighttime imagery. Field checking confirmed that sinkholes or depressions associated with subsurface cavities were evident at seven of these sites. Noble concluded that thermal imagery may be useful as a guide to the location of subsurface cavities.

Thermal IR line scanning can also be useful for fracture or lineament mapping. Newton [*25, 26*] refers to the potential of line scanning in this role for sinkhole detection in Alabama. By enhancing the interfaces between adjacent temperature levels on daytime imagery, lineaments

were located which had not previously been identified on conventional aerial photography. As with aerial photography, regions with a high density of lineaments, or regions where two or more lineaments intersect, were assumed to indicate areas of higher subsidence risk.

The detection of large caves in limestone terrain using thermal IR data is a subject which has been investigated by Rinker [*44*]. In this case the thermal sensor was flown over a known cave system in Puerto Rico. By examining ground temperature measurements both within the cave and at the surface, Rinker reports that it was possible to choose the optimum time for obtaining imagery. He suggests that this period occurs in the early morning, when the interior of the cave is warmer than the outside air by up to 5°C. This time period also coincides with a reduction in atmospheric pressure. Thus, it was hypothesized that the cave system should be "exhaling" warm air through any dry sinkholes which linked the cave system to the ground surface. Examination of the imagery obtained over the test site confirmed this hypothesis.

A further report from the United States by Harvey et al. [*45*] has also indicated that thermal imagery may be a useful supplementary technique to use for locating sinkholes. The authors also stressed the need to be aware of local farming practices when interpreting the imagery. On several occasions it was reported that confusion arose between circular or elliptical anomalies on the imagery and similarly shaped areas on the ground. These areas were subsequently identified as grassy areas associated with cattle feeding methods in the area.

Concurrent with much of the work being carried out in the United States was a series of investigations in South Africa. Bakker et al. [*46*] and Warwick et al. [*47*] discuss the comparative performance of aerial photography and thermal line scanning for sinkhole detection. In this case, the area of interest was in the Western Transvaal and consisted of an area of dolomite overlain by a thin soil cover less than 0.5 m thick. The thermal imagery revealed the underlying geological structure. The dolomite with its high thermal inertia appears as a white tone on the image. Joints and faults, conversely, produce an almost black (cool) tone on the predawn imagery, primarily as a consequence of the moist soil which fills these openings. A map was subsequently produced from this imagery to illustrate the spatial distribution of the joints and faults and also the directional trends of the joints. Analysis of this map enabled several sites to be located for further detailed ground investigation.

Day and Wagener [*48*] also discuss the use of thermal data for projects in South Africa. In this case, the aerial data were integrated with gravimetric data to identify zones of high subsidence risk in a proposed site for residential development.

Work on the use of thermal IR line scanning in Italy has been reported by Marino et al. [*49,50*]. This work relates to methods of detecting shallow cavities in the limestone regions of southeastern Italy. These particular investigations formed part of a series of terrain investigations for a proposed housing development. For this project, thermal imagery was obtained both before dawn and at noon with a dual-channel scanner (9 to 11 and 3 to 5 μm range). The data were subsequently processed using several different image enhancement techniques.

The simplest processing technique applied to the data was that of density slicing the predawn imagery. More sophisticated and elaborate processing techniques are also discussed. For example, both "normalized gradient" and "texture analysis" (Tonneli [*51*]) are described. The latter technique would appear to be an edge enhancement technique; however, the technical details reported in both cases make detailed assessment difficult.

The papers cited in this section report work carried out in many different parts of the world and in varied geological conditions. Consequently, it is only possible to reach some very broad conclusions. First, the timing of the flight is critical; this conclusion is discussed in more detail in a later section of this paper. Second, periods following heavy rainfall should be avoided since the subtle variations in moisture content within an area used to infer the positions of solution features may be masked. Finally, at present little guidance can be offered about the most appropriate image processing techniques to use, and further work is currently being carried out in this field by the authors.

Microwave Radiometry

Compared with the large number of research reports published relating to subsidence detection using visible, near-IR and thermal IR remote sensing imagery, few papers have been written which have discussed the use of microwave sensing techniques. While both active and passive microwave systems exist, the main interest, to date, has been concentrated on the application of passive microwave radiometry. Kennedy [*52*], for example, reported the use of a triple-frequency, dual polarization, ground-based radiometer for ground profiling over a well-developed karst region in California. Geologically, the region consisted of carboniferous limestone overlain by between 3 and 15 m of alluvial material. Analysis of the results indicated "significant radiometric cold anomalies associated with void space development beneath several tens of feet of soil cover." Richer [*53*], however, in his comments about Kennedy's work, suggested that the observed "cold anomalies" could not be explained in terms of the presence of subsurface voids. He further concluded that the observed anomalies were more likely to have been caused by soil moisture or slope effects than by penetration to the subsurface voids.

Conflicting results have also been reported when using airborne microwave radiometry. Dedman and Culver [*54*] obtained day and nighttime data over a series of subsurface voids in Kansas. A high correlation was found to exist between cool anomalies on several temperature profiles and the positions of known subsurface voids. They concluded, however, that further surveys were required in order to identify the cause of these anomalies more positively. Much less success was reported by Rib et al. [*40*], who flew a 13.7 GHz radiometer over a known limestone mine. Unfortunately much of the data that were obtained during the flights were contaminated by interference from the Doppler navigational aids. While postflight filtering of the data was successful in eliminating this noise from the data, it also reduced the resolution of the data. Analysis of the results indicated that

> no direct relationship exists between microwave brightness temperatures and the presence of the subsurface mine. Both high and low brightness temperature anomalies occurred within the mine area as well as anomalies of equivalent magnitude outside the mine area.

Further discussion about some of the factors which may account for the variable success rates reported in this section are presented in the following section.

Factors Influencing the Effectiveness of Remote Sensing Methods

Geological Conditions

The most suitable geological conditions in which infilled solution features, without surface expression, can be detected are those in which good lithological contrasts exist. A favorable contrast, for example, would consist of clay surrounded by sand or gravel materials. If, following solution subsidence, a higher lying clay bed subsides downward to become juxtaposed against a lower lying sandy gravel, then subsequent surface erosion down to this level would reveal a good lithological contrast when viewed in plan. The lithological contrast could be expressed by soil color changes or by water content differences due to changes in permeability or porosity across the clay and sandy gravel boundary. Variations in the moisture content of the ground associated with disturbance by solution subsidence may also be discernible by ground temperature contrasts. Where lithological contrasts are not as strong, say a silt infill juxtaposed against a fine sand cover deposit, then the plan outline of these infilled solution features is consequently less likely to be detected.

Geobotanical contrasts may also be present; for example, different plant species may grow above the clay infill as opposed to those growing on the surrounding soils above the sandy gravel, thus reflecting the underlying geology..

Recent experience of applying remote sensing methods (aerial photography and airborne MSS) to the detection of solution features without surface expression, in the Cretaceous Chalk of England, has revealed the geological situations where detection has been successful. Most infilled solution pipes detected, occurred either on bare chalk surfaces just beyond a cover deposit margin (Fig. 3, E) or along the feathering margin of cover deposits (Fig. 3, F). Infilled solution pipes present where superficial cover deposits thickened to several metres or more were rarely detected (Fig. 3, I). Where solution pipes contained metastable infills (Fig. 3, H) the materials above cavities are undisturbed, hence, detection was not possible.

Solution features with surface expressions (Fig. 3, D and G) were generally successfully detected in all geological situations, on bare chalk surfaces or where a cover deposit was present. The main problems in these circumstances were the discrimination of natural ground subsidences from artificially created features. (Refer to the section on image interpretation.)

Resolution of Sensor

Spatial Resolution—Spatial resolution may be considered the ability of the sensor to distinguish small objects on the terrain from their surroundings. Photographic systems generally have high spatial resolution and the units used to describe the spatial resolution of a photographic system concern the film resolving power at a particular contrast (expressed in line pairs/mm) or ground resolution distance. The latter term is useful because it translates film resolving power to a ground distance at the contact scale of the photography. Ground resolution distances vary for different photographic scales and film resolving powers (Table 2). Altering the latter parameter to some extent enables the effects of varying contrast in the scene to be simulated.

In contrast, scanning systems have much lower spatial resolution at equivalent flying heights. There is some disagreement about the most suitable measure of spatial resolution for defining the performance of a scanner (Forshaw et al. [55]), but a widely used measure is that of instantaneous field of view (IFOV). It should, however, be noted that this measure does not account for scene contrast. The IFOV of an airborne scanner is typically either 1.25 or 2.5 mrad; the consequent variation in ground resolution for various flying heights is shown in Table 2. Therefore, if solution features exist which have a direct surface expression then the high spatial resolution capabilities of an aerial camera will make it the preferred technique to use.

Spectral Resolution—The principal advantages of a scanning system, particularly MSS systems, are their wider spectral range and higher spectral resolution. If an indirect criterion, such as stressed vegetation, is being used to detect potential solution feature target areas, then the spectral capabilities of an MSS are of considerable use. False color photographic emulsions can be used to extend the spectral range of an aerial photograph, but they do not have as high a radiometric resolution. Also, unlike MSS data, they cannot be enhanced and classified by image processing systems.

TABLE 2—*Comparison of the spatial resolution of photographic and scanning systems.*

Photographic Scale	Flying Height*	Photographic Resolution, line pairs/mm			Scanner Resolution, IFOV mrad	
		20	50	100	1.25	2.5
1:3 000	450	0.15 m	0.06 m	0.03 m	0.6 m	1.2 m
1:10 000	1500	0.50 m	0.20 m	0.10 m	1.9 m	3.8 m
1:20 000	3000	1.00 m	0.40 m	0.20 m	3.8 m	7.6 m

*Assuming the focal length, f, is 0.15 m.

However, the wide availability of aerial photographs in comparison with airborne scanning data may, nevertheless, make aerial photography the preferred technique for local applications.

Conditions During Data Acquisition

Temporal Effects—The variation, through time, of the reflected or emitted radiance from a solution feature can have a critical effect on its detection. An understanding of the diurnal and seasonal variation in response is crucial if the optimum time for detection is to be determined. At present, limited research has been carried out to measure the spectral variation of solution features *in situ*, or to record multitemporal imagery as a further aid to defining the optimum time for data acquisition.

For remote sensing imagery recording the visible and near-IR parts of the EM spectrum, several factors have been identified as being significant. Diurnally the optimum period will generally be within the time when solar reflectance is at a maximum, thus avoiding the hours when long shadows occur. Photography taken at low sun angles may, however, be beneficial in enhancing the surface expression of ground subsidence features. Seasonally, in temperate climatic zones, late summer following prolonged dry weather is suggested as the optimum period for detecting infilled solution features without surface expression [*23*].

An understanding of solution feature temporal response in thermal IR is particularly important, since diurnal ground temperature variation may easily exceed the size of the anomaly being detected. Data acquisition times can vary (Fig. 6), though the predawn period tends to be the preferred period, when the effects of direct solar radiation are at a minimum. However, Harvey [*45*] suggested that midday in late summer could be a very good time for data acquisition. Lynn [*56*] recently examined ground temperature measurements and multitemporal thermal imagery in humid temperate environments. He found that the maximum diurnal temperature range and greatest thermal variation between different soils and surface materials occurred during midmorning and after local noon, again supporting the case for obtaining daytime imagery. He also suggested that thermal IR data acquired during the postsunset/predawn period is likely to provide poor discrimination of surface materials.

Environmental Effects—The environmental conditions during data acquisition are important in determining the success rate of a particular remote sensing technique. The optimum conditions for obtaining imagery in both the visible and IR (near and thermal) parts of the spectrum can be summarized as follows: cloud free, low atmospheric turbulence, preferably with no wind at ground surface, low atmospheric humidity, and an absence of smoke and dust. It is also suggested that periods following excessive precipitation should be avoided since saturated ground conditions tend to produce an even tone on the imagery.

Image Processing

In situations where scanner imagery has been obtained, the use of image processing techniques can enable the user to extract more information from the data than is possible from direct visual interpretation. Simple techniques such as contrast stretching and zoom facilities can be of considerable use when interpreting MSS data, as can density slicing when interpreting thermal IR imagery. More sophisticated image processing techniques such as classification and band ratioing can also be undertaken; however, few results have been published about the use of such techniques for solution feature detection.

For the future it may be that further development is required into systems that can integrate the spatial and spectral information content of remote sensing data with the more localized data from geophysical sensors, and thus provide a multisensor detection capacity.

Image Interpretation

The factors discussed previously are guidelines to enable a user to plan the data acquisition and processing stages of a project. After image processing, it is normal for some subjective interpretation of the imagery to be carried out; however, there are several potential detection problems. Some of the detection problems encountered include masking of areas by heavy vegetation and forestation; shadows cast along the boundaries of some of these areas; fields that have been ploughed for many successive years, which have completely obliterated previously existing indications of solution features; and urban development, which has also destroyed tonal patterns associated with the presence of solution features.

Other problems can arise in correctly interpreting the type of feature detected. The most problematical feature is a ground depression. On the English Cretaceous Chalk outcrop there are numerous man-made surface depressions formed by the digging of chalk in order to neutralize acid soils. After deterioration these depressions may assume the same appearance as naturally formed dolines. The collapse of shaft cappings and backfill, or adit roofs of chalk and flint underground workings can also create surface depressions similar to naturally formed dolines.

The problems outlined above are not unique and must plague similar studies of other limestone terrains across the world. Thus, the remote sensing user should appreciate the fact that the surface morphology of an identified circular ground depression, by itself, is often insufficient to infer genesis. Other forms of spatial distribution analysis may help to distinguish between the presence of natural or artificial features; however, ultimately, doubtful surface depressions should be closely investigated in the field. Sometimes, final proof of origin might only be possible by sinking boreholes or excavating a trial pit through the feature.

For solution features without surface expression, it is also necessary to be wary of other analogous ground tonal patterns—for example, shaft backfills, well heads, hay stacks, archaeological remains, or agricultural crop watering patterns. Again, doubtful tonal patterns should be closely investigated in the field, by indirect geophysical or direct exploratory techniques, if required.

Conclusions and Recommendations

Soluble carbonate rock terrain is present in many parts of the world, and karst processes have resulted in the formation of a variety of solution features. Certain forms of solution features are metastable and can give rise to ground subsidence problems. Any engineering or building development to be located on limestone should take account of the potential subsidence risk posed by these solution features.

Remote sensing is recommended as a technique to use at the desk study stage of a site investigation. A review of the various remote sensing techniques indicates that aerial photography should be used for solution features which have a surface expression. However, where indirect evidence of the presence of solution features is being used for detection, then airborne MSS and thermal IR line scan should be used to supplement aerial photography.

The authors also recommend, for the most favorable detection results, that remote sensing techniques be used in appropriate geological conditions and under optimum environmental conditions. This work suggests that further research is required to determine the *in situ* spectral response of solution features. Data of this type would assist in the determination of the optimum time for detection.

Acknowledgments

The authors would like to express their thanks to the Natural Environment Research Council for their financial support and to V. Brown for her comments on parts of the original manuscript. They also wish to thank Aerofilms Ltd. for permission to reproduce Fig. 5.

References

[1] Sweeting, M. M., *Karst Landforms*, MacMillan, London, 1972.

[2] Bogli, A., *Karst Hydrology and Physical Speleology*, Springer-Verlag, New York, 1980.

[3] Walsh, P. T., Edler, G. A., Edwards, B. R., Urbani, D. M., Valentine, K., and Soyer, J., "Large Scale Surveys of Solution Subsidence Deposits in the Carboniferous and Cretaceous Limestones of Great Britain and Belgium and their Contribution to an Understanding of the Mechanisms of Karstic Subsidence," *Proceedings*, Association of Engineering Geologists Symposium on Sinkholes and Subsidence-Engineering Geological Problems Associated with Soluble Rocks, Hanover, England, 1973, pp. T2.A1-10.

[4] Beck, B. F., "Sinkhole Terminology," in *Sinkholes: Their Geology, Engineering and Environmental Impact*, B. F. Beck, Ed., Balkema Press, 1984, pp. 9-10.

[5] Edmonds, C. N., "Towards the Prediction of Subsidence Risk Upon the Chalk Outcrop," *Quarterly Journal of Engineering Geology*, Vol. 16, No. 4, 1983, pp. 261-266.

[6] Edmonds, C. N., "The Engineering Geomorphology of Karst Development and Prediction of Subsidence Risk upon the Chalk Outcrop in England," Ph.D. thesis, University of London, London, England, 1987.

[7] Jennings, J. E., Brink, A. B., Louw, A., and Gowan, G. D., "Sinkhole and Subsidence in the Transvaal Dolomite of South Africa," *Proceedings*, Sixth International Conference on Soil Mechanics, 1965, pp. 51-54.

[8] Sowers, G. E., "Failure in Limestones in Humid Subtropics," *Journal of the Geotechnical Division of the ASCE*, Vol. 101, GT8, 1975, pp. 771-787.

[9] Brink, A. B., *Engineering Geology of South Africa*, Vol. 1, *Building Publications*, Pretoria, South Africa, 1979.

[10] Newton, J. G., "Induced and Natural Sinkholes in Alabama—A Continuing Problem in Highway Corridors," *Transportation Research Record*, Vol. 612, 1977, pp. 9-16.

[11] Newton, J. G., "Induced Sinkholes: An Engineering Problem," *Journal of the Irrigation and Drainage Division of the ASCE*, Vol. IR2, June 1981, pp. 175-185.

[12] Newton, J. G., "Review of Induced Sinkhole Development," in *Sinkholes: Their Geology, Engineering and Environmental Impact*, Beck, B. F., Ed., 1984, Balkema Press, pp. 3-9.

[13] Newton, J. G., "Natural and Induced Sinkhole Development," IAHS Publication No. 151, International Association of Hydrological Sciences, 1986, pp. 549-564.

[14] Malkin, A. B. and Wood, J. C., "Subsidence Problems in Route Design and Construction," *Quarterly Journal of Engineering Geology*, Vol. 5, No. 3, 1972, pp. 179-194.

[15] Price, D. G., "Engineering Geology in an Urban Environment," *Quarterly Journal of Engineering Geology*, Vol. 4, No. 2, 1972, pp. 191-208.

[16] Partridge, T. C., Harris, G. M., and Diesel, V. A., "Construction upon Dolomite of the South-Western Transvaal," *Bulletin of the International Association for Engineering Geology*, No. 24, 1981, pp. 125-135.

[17] Nogushi, T., Takahashi, R., and Tokumitsu, Y., "Small Sinking Holes in Limestone Areas with Special Reference to Drainage of Coal Mines," IAHS Publication No. 89, International Association of Hydrological Sciences, 1970, pp. 467-474.

[18] Beck, B. F., Ed., *Sinkholes: Their Geology, Engineering, and Environmental Impact*, Balkema Press, 1984.

[19] Sowers, G. F., "Correction and Protection in Limestone Terrane," in *Sinkholes: Their Geology, Engineering, and Environmental Impact*, B. F. Beck, Ed., Balkema Press, 1984, pp. 373-378.

[20] Fountain, L. S., Herzig, F. X., and Owen, T. E., "Detection of Subsurface Cavities by Surface Remote Sensing Techniques," Report No. FHWA-RD-75-80, U.S. Department of Transportation, Federal Highway Administration, Office of Research and Development, Washington, DC, 1975.

[21] Kirk, K. G. and Werner, E., "Handbook of Geophysical Cavity Locating Techniques with Emphasis on Electrical Resistivity," Report No. FHWA-IP-81-3, Department of Transportation, Federal Highway Administration, 1981.

[22] McCann, D. M., Baria, R., Jackson, P. D., Culshaw, M. G., Green A. S. P., Suddaby, D. L., and Hallam, J. R., "The Use of Geophysical Methods in the Detection of Natural Cavities, Mineshafts and Anomalous Ground Conditions," Engineering Geology Report, Vol. 82, No. 5, British Geological Survey, 1982.

[23] Norman, J. W. and Watson, I., "Detection of Subsidence Conditions by Photogeology," *Engineering Geology*, Vol. 9, 1975, pp. 359-381.

[24] West, G. W. and Dumbleton, M. J., "Some Observations on Swallow Holes and Mines in the Chalk," *Quarterly Journal of Engineering Geology*, Vol. 5, 1972, pp. 171-177.

[25] Newton, J. G., "Early Detection and Correction of Sinkhole Problems in Alabama, with a Preliminary Evaluation of Remote Sensing Applications," Alabama Highway Research Report No. 76, University of Alabama, Tuscaloosa, AL, 1976.

[26] Newton, J. G., Copeland, C. W., and Scarborough, L. W., "Sinkhole Problem along Proposed Route

of Interstate Highway 459 near Greenwood, Alabama," *Geological Survey of Alabama Circular*, No. 83, 1973.

[27] Warren, W. M. and Wielchowsky, C. C., "Aerial Remote Sensing of Carbonate Terranes in Shelby County, Alabama," *Ground Water*, Vol. 11, No. 6, 1973, pp. 14-26.

[28] Trainer, F. W. and Ellison, R. L., "Fracture Traces in the Shenandoah Valley, Virginia," *Photographic Engineering*, Vol. 33, No. 2, 1967, pp. 190-199.

[29] Littlefield, J. R., Culbreth, M. A., Upchurch, S. B., and Stewart, M. T., "Relationship of Modern Sinkhole Development to Large Scale Photolinear Features," in *Sinkholes: Their Geology, Engineering and Environmental Impact*, B. F. Beck, Ed., 1984, Balkema Press, pp. 189-195.

[30] Parisek, R. R., "On the Nature and Significance of Fracture Traces and Lineaments in Carbonate and other Terranes," *Karst Hydrology and Water Resources*, V. Yerjervich, Ed., Vol. 1, Karst Hydrology, Water Resources Publications, 1976, pp. 3/1-3/108.

[31] Casper, J. W., Ruth, B., and Degner, J. D., "A Remote Sensing Evaluation of the Potential for Sinkhole Occurrence," Report by the Remote Sensing Applications Laboratory, University of Florida, Gainesville, FL, 1981.

[32] Casper, J. W., Ruth, B. E., and Degner, J. D., "Remote Sensing Study of Sinkhole Occurrence," Civil Engineering Applications of Remote Sensing, ASCE-ASP Speciality Conference, Madison, WI, 1980, pp. 1-14.

[33] Lattman, L. H., "Technique of Mapping Geologic Fracture Traces and Lineaments on Aerial Photographs," *Photogrammetric Engineering*, Vol. 24, 1958, pp. 568-576.

[34] Lattman, L. H., and Parisek, R. R., "Relationships Between Fracture Traces and Occurrence of Ground Water in Carbonate Rocks," *Journal of Hydrology*, Vol. 2, 1964, pp. 73-91.

[35] Sesoren, A., "Potential of Remote Sensing Use in a Karstic Area, Antalya Region in the South of Turkey," IAHS Publication No. 161, International Association of Hydrological Sciences, 1986.

[36] Coker, A. E., Marshall, R., and Thompson, N. S., "Application of Computer Processed Multispectral Data to the Discrimination of Land Collapse (Sinkhole) Prone Areas in Florida," *Proceedings*, Sixth International Symposium on Remote Sensing of the Environment, Vol. 1, Ann Arbor, MI, 1969, pp. 65-69.

[37] Stohr, C. J. and West, T. R., "Delineation of Sinkholes Using Thermal Infrared Imagery," LARS Information Note 032574, Purdue University, West Lafayette, IN, 1974.

[38] Parker, D. C., "Remote Sensing for Engineering Investigations of Terrain—Infrared Systems," *Proceedings*, Fifth Michigan Symposium on Remote Sensing of the Environment, 1968, pp. 701-709.

[39] Stingelin, R. W., "Aerial Infrared Imagery and Its Limitations in Civil Engineering Practice," *Highway Research Record*, Vol. 412, 1972, pp. 14-21.

[40] Rib, H. T., Spencer, J. W., Falls, C. P., and Koca, J. F., "Evaluation of Aerial and Field Remote Sensing Systems for Detecting Subsurface Cavities in Kansas," Report No. FHWA-RD-75-120, Federal Highway Administration, Washington, DC, 1975.

[41] Matalucci, R. V. and Abdel-Hady, M., "Surface and Subsurface Exploration by Infrared Surveys," *Highway Research Board Special Report*, No. 102, 1969, pp. 1-12.

[42] Deloach, W. C., "Remote Sensing Applications to Environmental Analysis," *Highway Research Record*, No. 412, 1972, pp. 29-37.

[43] Noble, D. F., "Utilization of Remote Sensing in the Preliminary Survey-Highway Planning Stage in Virginia," *Highway Research Record*, No. 412, 1972, pp. 41-49.

[44] Rinker, J. N., "Airborne Infrared Thermal Detection of Caves and Crevasses," *Photographic Engineering and Remote Sensing*, Vol. 41, No. 11, 1975, pp. 1391-1400.

[45] Harvey, E. J., Williams, J. H., and Dinkel, T. R., "Application of Thermal Imagery and Aerial Photography to Hydrologic Studies of Karst Terrane in Missouri," USGS/WRD/WRI-78/005, U.S. Geological Survey, Water Resources Division Report, Denver, CO, 1977.

[46] Bakker, P., Church, D. J., Feuchtwanger, T. F., Grotenboer, C. A., Lee, C. A., Longshaw, R. P., and Viljoen, R. P., "Some Practical Applications of Thermal Infrared Linescanning," *Mining Magazine*, Vol. 139, No. 4, 1978, pp. 398-413.

[47] Warwick, D., Hartopp, P. G., and Viljoen, R. P., "Application of Thermal Infrared Linescanning Techniques to Engineering Geological Mapping in South Africa," *Quarterly Journal of Engineering Geology*, Vol. 12, 1979, pp. 159-179.

[48] Day, P. W., and Wagener, F. M., "Investigation Techniques on Dolomites in South Africa," *Sinkholes: Their Geology, Engineering and Environmental Impact*, B. F. Beck, Ed., Balkema Press, 1984, pp. 153-158.

[49] Marino, C. M., Perna, G., and Tonelli, A. M., "Thermal I.R. Flights in Land Use Planning: Subsurface Cavities Detection in Urban Areas," *Proceedings*, 17th Symposium on Remote Sensing of the Environment, Ann Arbor, MI, 1983, pp. 853-862.

[50] Marino, C. M., Perna, G., and Tonelli, A. M., "Examples of Airborne Thermal Infrared Analysis Applied to Engineering Geology and Civil Works Site Investigation," *Proceedings*, 18th International Symposium on Remote Sensing of the Environment, Paris, France, 1984.

[51] Tonneli, A. M., "Surface Texture Analysis with Thermal and Near Infrared Scanners," *Photographic Engineering and Remote Sensing*, Vol. 44, 1978, pp. 1273-1278.
[52] Kennedy, J. M., "A Microwave Radiometric Study of Buried Karst Topography," *Geological Society of America Bulletin*, Vol. 79, 1968, pp. 735-742.
[53] Richer, K. A., "Comments on A Microwave Radiometric Study of Buried Karst Topography," *Geological Society of America Bulletin*, Vol. 8, 1970, pp. 585-588.
[54] Dedman, E. V. and Culver, J. L., "Airborne Microwave Radiometer Survey to Detect Subsurface Voids," *Highway Research Record*, No. 421, 1972, pp. 66-70.
[55] Forshaw, M. R. B., Haskell, A., Miller, P. F., Stanley, D. J., and Townsend, J. R. G., "Spatial Resolution of Remotely Sensed Imagery—A Review Paper," *International Journal of Remote Sensing*, Vol. 4, No. 3, 1983, pp. 497-520.
[56] Lynn, D. W., "The Timeliness of Thermal IR Data Acquisition for Soil and Surface Material Survey in Humid Temporate Environments," *Proceedings*, Fourth Symposium of the International Society of Soil Science, Working Group, Remote Sensing for Soil Survey, Enschede, The Netherlands, 1985.

ANALYSIS OF MULTITEMPORAL AIRBORNE REMOTE SENSING DATA FOR THE LOCATION OF SOLUTION FEATURES IN THE CHALK OF SOUTHERN ENGLAND

T J M Kennie,
Department of
Civil Engineering,
University of Surrey,
Guildford GU2 5XH,
ENGLAND.

M S Rosenbaum,
Department of
Geology,
Imperial College of
Science and Technology,
London, SW7 2BP, ENGLAND

C N Edmonds,
Applied Geology,
Camden House,
Kenilworth,
Warwickshire,
CV8 1FB, ENGLAND.

Commission VII

ABSTRACT

The potential use of airborne remote sensing techniques for the detection of collapse and subsidence features in chalk and other limestone rocks is examined. The detection of such features in the early stages of an engineering project is of crucial importance if serious geotechnical problems to building construction and public safety are to be avoided. Particular attention is paid to the use of aerial photography and thermal infrared (IR) data as a means of detection. Some preliminary results from the analysis of multitemporal data acquired over two test sites are presented.

1. INTRODUCTION

Natural solution features which develop upon soluble carbonate rock terrains have morphologies controlled by geological, hydrogeological and geomorphological conditions. Chalk, which has the largest outcrop area (250,000Km^2) of any limestone in Great Britain, is a weak, open textured non-crystalline carbonate rock. Such rocks have a characteristic range of solution features, some of which are metastable, and these are outlined in Section 2. The metastable forms pose a serious subsidence risk to engineering structures which can result in severe stress, or even their complete destruction.

It would clearly be advantageous to be able to detect metastable solution forms prior to the engineering development of an area. Remote sensing has been attempted for such detection, with varying degrees of success, as cost-effective precursors to land-based investigation using geophysical techniques (McCann et al, 1987) or standard ground investigation techniques (details of which can be found in BS 5930 (1981). In general, detection methods can only successfully identify solution features in plan or vertical section. Therefore, usually two or more successful methods will have to be combined in order to achieve a three dimensional picture of the feature requiring treatment.

Airborne remote sensing should be considered primarily as a reconnaissance tool to be used during, or immediately after, the

desk study stage, rather than as a technique to be used during the main detailed ground investigation stage. The aim of a survey utilising remote sensing techniques should therefore be to delineate areas where, potentially metastable solution features may exist and to identify specific areas where more detailed local investigations should be carried out.

There are three principal methods of airborne remote sensing which have been used for the detection of potential solution features: aerial photography, thermal IR linescanning and multispectral scanning (MSS). To date the latter technique has been used to a much lesser extent than the former two methods.

The application of aerial photography to the detection of collapse features is well known; a general account of the method being presented by Norman and Watson (1975). Aerial photography can be used to detect collapse or subsidence features either by direct (often stereo) observation of the photographs or by the use of indirect evidence; for example, the presence of stressed vegetation or structural weaknesses such as faults and fractures (Littlefield et al, 1984). In the latter two cases the use of false colour IR emulsions has also provided useful.

A number of authors have also considered the use of airborne thermal IR and, to a much lesser extent, MSS for solution feature detection. The following selective list of previously published work indicates the widespread interest in these techniques. In the US work has been carried out by Coker et al (1969), Newton (1976), in South Africa by Warwick et al (1979) and Day and Wagner (1984) in Italy by Marino et al (1983) and in the United Kingdom by Cooper (1988). A more comprehensive literature review and an assessment of the comparative performance of the different techniques is given in Kennie and Edmonds (1987). In general terms the results obtained by previous workers in this field have been quite variable and further highlight the need for an investigation of the factors which influence the success of the different techniques.

Details of a current project to acquire multitemporal data over two test sites in southern England is presented in Section 4. The results of the preliminary interpretation of the imagery over these two sites, and an outline of some of the difficulties associated with the use of such data for two test sites for solution feature detection is presented in Section 5. The paper concludes by considering the benefits which may accrue from the use of a Geographical Information System (GIS) for the assessment of subsidence hazard in chalk regions.

2. TYPES OF SOLUTION FEATURES

Surface and subsurface solution features develop in carbonate rocks as a result of dissolution caused by acidic surface and ground water; collectively known as karst. "Surface" karst features in a carbonate bedrock may also be reflected within any overlying superficial deposits. Thus solution-widened joints,

solution pipes, swallow holes and dolines ("sinkholes")can cause collapse features in the overburden. Subsurface karst features are developed deeper within the body of the carbonate bedrock and include both open and infilled solution pockets, cavities and cave systems.

Solution is the fundamental cause for the development of karst. However, it is subsidence, caused by the destabilisation of the infilling deposits or breakdown of a limestone roof to a cavity, which triggers the final collapse and causes damage to engineering structures.

The ambiguous and poorly defined terminology of karst provides one of the most difficult aspects of studying solution features as discussed by Edmonds (1983) and Beck (1984). In order to clarify the terminology, Figure 1 illustrates and defines schematically the range of terms used by the present authors to describe natural solution features that develop both within the carbonate bedrock and the overburden.

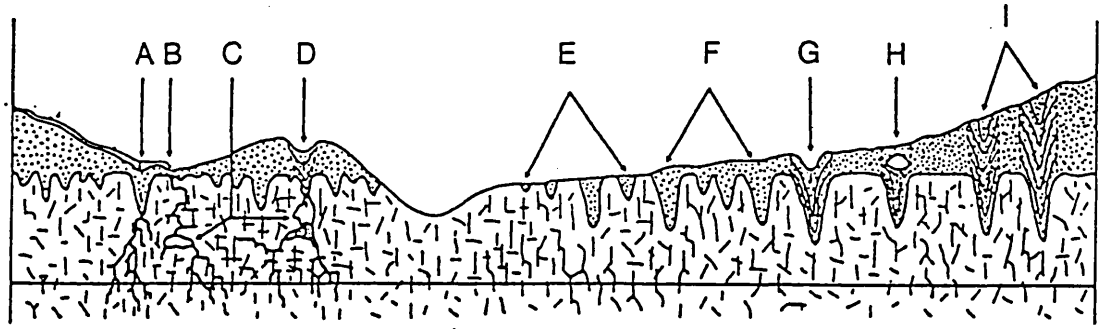


FIG.1. SCHEMATIC DIAGRAM TO ILLUSTRATE RANGE OF NATURAL SOLUTION FEATURES PRESENT ON LIMESTONE

A Swallowhole, downward percolation of ground water taking advantage of a pre-existing infilled solution pipe or solution widened joint.

B Swallowhole, downward percolation of ground water taking advantage of solution enlarged joints in the limestone bedrock.

C Solution cavity within the bedrock produced by phreatic and/or vadose dissolution.

D Collapse doline, produced by collapse of a solution cavity roof and subsequent upward cavity migration to the surface.

E Remnants of infilled solution widened joints or solution pipes just beyond the cover deposit outcrop margin.

F Solution widened joints or solution pipes infilled with deposits that have undergone solution subsidence below a thin cover deposit.

G Subsidence doline, produced by collapse of a metastable cavity within a cover deposit and subsequent upward cavity migration to the surface.

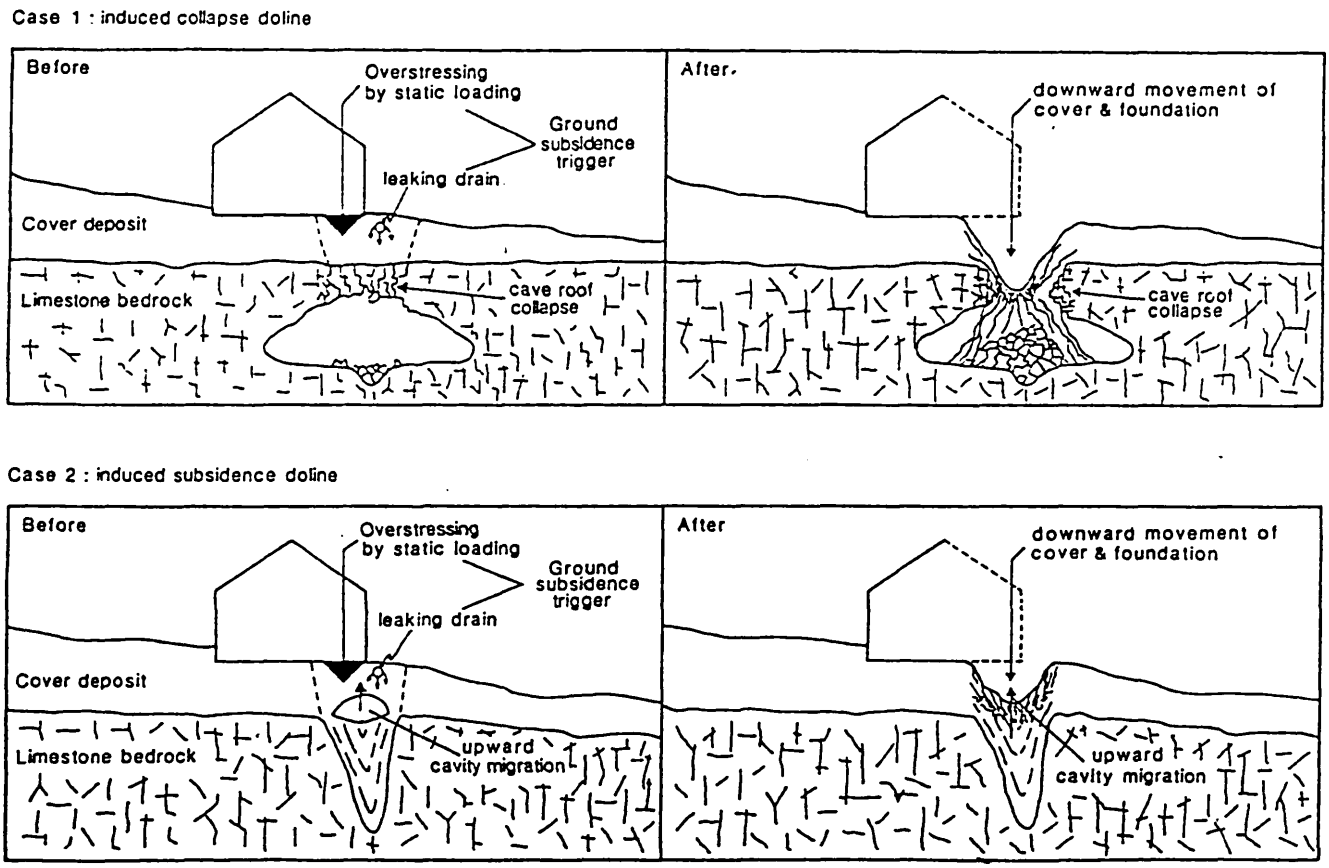
H Metastable cavity formed during solution subsidence of cover deposits down into developing solution widened joints or solution pipes in the limestone rock.

I Solution widened joints or solution pipes infilled with deposits that have undergone solution subsidence below a thick cover deposit.

3. GEOTECHNICAL PROBLEMS ASSOCIATED WITH SOLUTION FEATURES

The metastable karst features caused by the dissolution of carbonate rock generally provide unreliable bearing capacity for foundations leading to partial or complete structural failure. The time dependent nature of karstic solution makes prediction of the hazard risk uncertain. Once failure has occurred, remedial measures such as grouting and underpinning can be undertaken but this can involve considerable expense. Occasionally severe or even catastrophic damage is suffered by the structure which necessitates demolition and of course there is also the inherent danger to the public because of the possibility of unexpected ground collapses. There is therefore an urgent need to more reliably assess the risk of solution feature development in carbonate areas.

FIGURE 2 Examples of foundation failure caused by induced dolines



First, the areas that could potentially host active karstic development need to be identified. Then an assessment must be made of each of the possible trigger mechanisms. These include heavy rainstorms, water from leaking drains or pipes, artificial ground water lowering and both dynamic and static surface loading. Two examples which illustrate the potential effects of static surface loading and excessive concentrations of water drainage on the development of induced dolines are shown in Figure 2.

4. MULTITEMPORAL DATA ACQUISITION

4.1 Location of Test Sites

The first test site chosen for the project was located near Swanley, Kent; an area close to the highly populated and actively developing fringe around London. In this general area to the east of London, the Cretaceous Chalk limestone bedrock is overlain by a variety of Tertiary and Quaternary deposits. Beyond the London Clay margin these deposits are in many places disturbed by solution subsidence effects associated with the presence of solution features which have developed in the underlying Chalk.

The second test site was on Southover Heath, Dorset a rural area to the south west of London. Again, the Cretaceous Chalk limestone bedrock is overlain by Tertiary and Quaternary deposits. In particular the Palaeocene Reading beds are extensively piped into the underlying Chalk as are the high level fluvial gravels.

In the first site the solution features are predominantly solution pipes with no surface dolines being present. Evidence of solution features in this area was reported in the ground investigations carried out for the construction of the M25 London orbital motorway which passes through the central part of the test site Within the second site the solution features are also predominantly solution pipes, although surface dolines are also present.

In order to examine the influence of seasonal and diurnal spectral variations on the response from solution features developed on the chalk or its overburden it was necessary that the data should be acquired on a multitemporal basis.

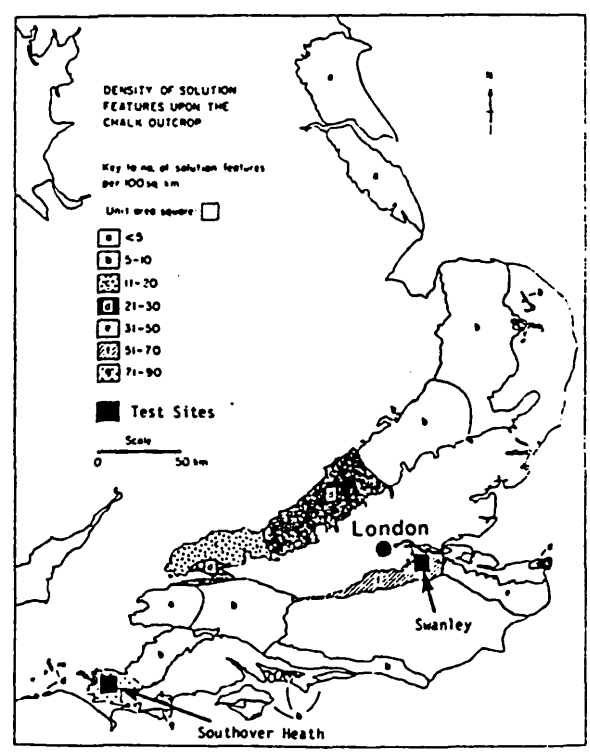


Figure 3

Location of Test Sites

4.2 Data Acquisition

Ideally the data was required during spring, summer and autumn, and for each day before dawn and during mid morning and mid afternoon. This would ensure the sampling of a characteristic cross section of ground temperatures and general ground conditions. Poor weather conditions over the test sites somewhat reduced the multitemporal coverage that could be achieved in practice, and the data acquired is summarised in Table 1.

Table 1

Season (Date)	Site 1 Southover Heath, Dorset	Site 2 Swanley, Kent
Spring (28th April, 1986)	None acquired	11:30
Summer (17th July, 1986)	08:45	09:30
Autumn (19th September, 1986)	09:20	None acquired
(20th September, 1986)	16:00	13:30

5. ASSESSMENT OF THE SUCCESS OF AIRBORNE REMOTE SENSING TECHNIQUES

5.1 General Observations

Preliminary visual analysis of selected sub scenes of the airborne remote sensing data (comprising aerial photography, MSS and thermal IR linescanning) has enabled the following general conclusions to be reached:

(1) solution features with a surface expression (Figure 1, Case D and G) were generally detectable in all geological situations, both upon bare chalk surfaces and where a superficial deposit was present. The main problem which arose in these circumstances was the problem of discriminating between natural ground subsidence and artificially created features,

(2) most infilled solution pipes which were detected occurred either on bare chalk surfaces just beyond a superficial deposit margin (Figure 1, Case E), or along the margin of superficial deposits (Figure 1, Case F). Where superficial deposits thickened to more than 2 to 3 metres such features

were rarely detected (Figure 1, Case I). Also, where solution pipes contained metastable infills (Figure 1, Case H), and the materials above the cavities were undisturbed, detection was generally not possible due to the lack of contrast in physical properties. However, the fact that in many cases solution features, both with and without, surface expression were able to be detected indicates the potential benefit of the technique to planners and engineering geologists.

Clearly, in terms of spatial resolution, it is important to ensure that the features under investigation are at least equivalent to the IFOV (Instantaneous Field of View) of the sensor if direct detection is to be successful. Furthermore, in practice the feature should normally be at least two to three times the IFOV to allow for the detection of low contrast features.

5.2 Aerial Photography

Two periods were identified as being significant when using panchromatic aerial photography for the detection of solution features. Diurnally, the optimum period for detection was generally found to be when solar reflectance was at a maximum. This avoids the hours when long shadows occur that would otherwise obscure significant tonal patterns. Nevertheless, photography taken at low sun angles may also be beneficial since it enhances the surface expression of ground subsidence features. Seasonally, in temperate climatic zones such as the UK, early Autumn following prolonged dry weather appears to provide the best contrast for the detection of infilled solution features which have no surface expression.

5.3 Thermal IR and MSS Imagery

The spectral resolution of the sensor is clearly of greatest importance when indirect criteria such as stressed vegetation or soil moisture variations are being used to infer the presence of a solution feature. In addition the multitemporal response from a solution feature can have a critical effect on its detection. This is particularly important if thermal IR data is being used since the diurnal ground temperature variation may easily exceed the size of the anomaly associated with, for example, moisture variations in the vicinity of a solution feature. Pre-dawn has tended to be the preferred period for such data acquisition, primarily because the effects of direct solar radiation are at a minimum. Lynn (1986), however, has suggested that the maximum diurnal temperature range and the greatest thermal variation between different soils and surface materials occurs during mid-morning and after local noon, supporting the case for obtaining daytime imagery. An assessment of the validity of the conclusions from this work is currently being investigated in order to determine the optimum time period for data acquisition.

The investigations to date suggest that thermal imagery acquired in the early to mid-morning period is superior to that acquired

during mid-afternoon. However, the inability to acquire imagery during the pre-dawn period limits the extent to which this can be specified as the optimum period for data acquisition. Seasonally the current work indicates that Autumn is the preferred period for acquiring thermal data. Further work is required with the MSS data to assess the value of different band combinations for solution feature detection. It is proposed to examine the potential use of the optimum index factor as applied to airborne MSS data acquired in the UK by Greenbaum (1987).

5.4 Limitations and Practical Detection Difficulties

A number of practical detection problems have been encountered which could severely limit the usefulness of airborne remote sensing for this type of application. Among the more significant are:

(1) masking of areas of potential subsidence by heavy vegetation and areas of woodland,
(2) the effect of shadows cast along the boundaries of areas masking features of interest,
(3) fields that have been ploughed for many successive years removing previously existing indications of solution features, and,
(4) urban development which has also destroyed tonal patterns associated with the presence of solution features.

Other problems can arise when attempting to identify and classify the type of feature which can be detected on the images. For solution features which have no surface expression, it is necessary to be cautious of other anomalous ground tonal patterns such as those caused by archaeological remains and agricultural crop watering patterns.

One particularly difficult feature to interpret is a surface depression. On the Chalk outcrop of southern England there are numerous man-made surface depressions which have been created by the mining or quarrying of chalk for application on fields to neutralise acid soils. After a period of deterioration these old workings frequently assume a similar appearance to a naturally formed doline. Thus the user of remote sensing imagery should appreciate that the surface morphology of an identified circular ground depression can, by itself, often be insufficient to infer its origin. Other forms of spatial distribution analysis should help to discriminate between the presence of natural or artificial features. In the end, however, final proof of the origin of a potential solution feature can only be confirmed by direct ground investigation such as drilling a borehole or excavating a trial pit through the feature.

6. CONCLUSIONS AND FUTURE WORK

Soluble carbonate rocks outcrop in many parts of the world and karst processes have resulted in the formation of a variety of solution features. Certain forms of solution feature are

metastable and can give rise to ground subsidence problems. Airborne remote sensing is recommended as a technique to be used at, or immediately after, the desk study stage of a site investigation. An assessment of the various remote sensing techniques indicates that aerial photography should generally be the first source of data to be used for the detection of solution features which have a surface expression. However, where it is necessary to use indirect evidence in order to infer the presence of solution features, it is suggested that airborne thermal IR and MSS imagery should be used to supplement the aerial photography. Indeed, thermal IR and MSS imagery are particularly helpful in detecting material responses of the ground to underlying karstic features where there is an absence of a surface topographic expression.

Several difficulties have been identified which can restrict the use of airborne remote sensing techniques, including the problem of determining the optimum conditions for obtaining imagery. Some recommendations for acquiring data have been provided in the paper and future work will concentrate on the inter-band spectral correlation of the MSS datasets with a view to determining the optimun three-band combination.

For the future it is suggested that consideration should be given to using airborne remote sensing data more extensively as a component of a spatial subsidence hazard assessment model, such as that developed by Edmonds et al (1987). At present this particular form of hazard prediction is carried out manually, but the digital nature of the data lends itself to computer manipulation and it is suggested that the hazard prediction could be enhanced by using a Geographical Information System (GIS). This would effectively combine the spatial and spectral information from the remote sensing data with spatial data from existing map sources (topographic, geological etc.) and non - spatial sources of data (such as borehole information). At present a variety of database management systems are being considered together with methods of optimising the data interpretation (Rosenbaum, 1987) and methods of presenting the results.

REFERENCES

BS 5930. 1981. Code of practice for site investigations. British Standards Institute. HMSO. 147pp.

BECK B.F. 1984. Sinkhole terminology, in Sinkholes: Their Geology, Engineering and Environmental Impact, Beck B.F., (Ed), Balkema Press, IX-X.

COKER A.E., MARSHALL R., and THOMPSON N.S. 1969. Application of computer processed multispectral data to the discrimination of land collapse (sinkhole) prone areas in Florida, Proc. of the 6th International Symposium on Remote Sensing of the Environment, Ann Arbor, Michigan, Vol. 1., 65-69.

COOPER A.H. 1988. Airborne multispectral scanning of subsidence caused by Permian gypsum dissolution at Ripon, North Yorkshire. Proc. of 1986 Natural Environment Research Council (NERC) Airborne Remote Sensing Campaign, Wallingford, Oxfordshire, February, 63-75.

DAY P.W., and WAGENER F.M. 1984. Investigation techniques on dolomites in South Africa. Sinkholes: Their Geology, Engineering and Environmental Impact, Beck B.F. (Ed.) Balkema Press, 153-158.

EDMONDS C.N. 1983. Towards the prediction of subsidence risk upon the chalk outcrop. Quarterly Journal of Engineering Geology, 6, 4, 261-266.

EDMONDS C.N., GREEN C.P. and HIGGINBOTTOM I.E. 1987. Subsidence hazard prediction for limestone terrains, as applied to the English Cretaceous Chalk. in Planning and Engineering Geology Culshaw et al (Eds.), Geological Society Engineering Geology Special Publication, No. 4, 283-293.

GREENBAUM 1987. Lithological discrimination in central Snowdonia using airborne multispectral scanner imagery. Int. Journal of Remote Sensing, 8, 6, 799-816.

KENNIE T.J.M. and EDMONDS C.N. 1988. The location of potential ground subsidence and collapse features in soluble carbonate rocks by remote sensing techniques. Proc. of the American Society for Testing and Materials (ASTM) International Symposium on Geotechnical Applications of Remote Sensing and Remote Data Transmission, ASTM Special Technical Publication 967, 45-60.

LYNN D.W. 1986. Monotemporal, multitemporal and multidate thermal IR data acquisition from satellites for soil and surface material survey. Int. Journal of Remote Sensing. 7, 2, 213-231.

LITTLEFIELD J.R., CULBRETH M.A., UPCHURCH S.B. and STEWART M.T. 1984. Relationship of modern sinkhole development to large scale photolinear features, in Sinkholes; Their Geology, Engineering and Environmental Impact, Beck B.F.(Ed.), Balkema Press, 189-195.

MARINO C.M., PERNAL G., and TONELLI A.M. 1983. Thermal Infrared flights in land use planning: subsurface cavities detection in urban areas. Proc. of the 17th Symposium on Remote Sensing of the Environment, Ann Arbor, Michigan, 853-862.

McCANN D.M., JACKSON P.D. and CULSHAW M.G. 1987. The use of geophysical surveying methods in the detection of natural cavities and mineshafts. Quarterly Journal of Engineering Geology, 20, 59-74.

NEWTON J.G. 1976. Early detection and correction of sinkhole problems in Alabama, with a preliminary assessment of remote sensing techniques. Alabama Highway Reseach Report No. 76, 83pp.

NORMAN J.W.and WATSON I. 1975. Detection of subsidence conditions by photogeology. Engineering Geology, 9, 359-381.

ROSENBAUM M.S. 1987. The use of stochastic models in the assessment of a geological database. Quarterly Journal of Engineering Geology, 20, 31-40.

WARWICK D., HARTOPP P.G. and VILJOEN R.P. 1979. Application of thermal infrared linescanning techniques to engineering geological mapping in South Africa. Quarterly Journal of Engineering Geology, 12, 159-179.

PROCEEDINGS OF THE SEVENTH INTERNATIONAL SYMPOSIUM ON REMOTE SENSING FOR RESOURCES DEVELOPMENT AND ENVIRONMENTAL MANAGEMENT
ISPRS COMMISSION VII / ENSCHEDE / 25–29 AUGUST 1986

Remote Sensing for Resources Development and Environmental Management

Edited by
M.C.J.DAMEN / G.SICCO SMIT / H.TH.VERSTAPPEN
ITC, Enschede, Netherlands

OFFPRINT

A.A.BALKEMA / ROTTERDAM / BOSTON / 1986

Symposium on Remote Sensing for Resources Development and Environmental Management / Enschede / August 1986

A preliminary assessment of an airborne thermal video frame scanning system for environmental engineering surveys

T.J.M.Kennie & C.D.Dale
University of Surrey, Guildford, UK

G.C.Stove
ERSAC Ltd, Livingston, Scotland, UK

ABSTRACT : Video based camera systems are being used by an increasing number of remote sensing users. These systems offer several advantages in terms of ease of use, speed of operation, portability and cost. Of particular interest are the thermal video frame scanning (TVFS) systems. These scanners, developed primarily for military purposes, differ considerably in design from the thermal infrared linescanning instruments traditionally used for thermal surveys in remote sensing. One specific example, the Barr and Stroud IR18, is described together with the modifications necessary to mount such a system in a number of light aircraft. Methods of digitally enhancing and classifying single frames of data are discussed. The paper concludes by considering the application of suitably processed imagery for environmental engineering projects such as heat loss monitoring, thermal mapping of roads, geotechnical site investigations and drainage, sewer collapse and utility duct surveys. Further potential areas of application are also identified.

RESUME: Les utilisateurs de télé-détection utilisent de plus en plus l'appareil photographique qui se fonde sur le video. Ces systems offrent plusieurs advantages - facilité d'usage, vitesse d'operation, portabilité, et cout. Les systemès qui se basent sur l'explorateur thermique de cadre video (TVSF) sont d'intéret particulier. Le dessin de ces appareils explorateurs qui étaient dévéloppés pour les buts militaires d'abord différent considérablement des instruments d'exploration thermique et lineaire qu'on emploie d'habitude pour les levés thermiques dans le télé-détection. On décrit un exemple spécifique, le Barr et Stroud IR18, et aussi les modifications nécessaires pour installer un tel système dans quelques avions légers. On discute les méthodes digitales de rehausser et classifier les cadres uniques de données. Le papier conclut de considerer l'application des images traités pour les projets de génie de l'environnement, par exemple, la surveillance de la perte de chaleur, la cartographie thermique de routes, les investigations géotechniques de site et la surface de captation des eaux, l'effondrement des égouts, et les levés des conduits utilitaires. On identifie aussi les autres domaines d'application potentiale.

1 INTRODUCTION

An increasing number of users in recent years have been considering the potential of video based aerial remote sensing systems. A wide variety of data acquisition systems are currently available and a selection of these are reported in section 2, together with some discusssion of the advantages of using video as a recording medium.

One of the most exciting developments however, is in the use of thermal video frame scanners (TVFS) operated from light aircraft. The design features of such instruments, with particular reference to the Barr and Stroud IR18, are discussed in section 3. This is followed by details relating to the mounting of such a system in a variety of light aircraft.

The use of digital image processsing techniques to filter and enhance single frames of thermal data is presented in section 4. The paper concludes by reporting some of the current projects for which the scanner has been used.

2 AIRBORNE VIDEO BASED DATA ACQUISITION SYSTEMS

A summary of previous literature which has reported on the use of video based data acquisition systems is given in Table 1. Several advantages can be cited to demonstrate the benefits of video based systems, which include system portability, near real time verification of data capture, rapid turnaround (since there is no processing stage), low cost of operation and ease of duplication of data (Meisner and Lindstrom, 1985). In view of these perceived advantages a research programme was initiated by the University of Surrey, in conjunction with the University of Glasgow, in August 1984 to evaluate the use of a portable TVFS system for remote sensing applications. Although low resolution thermal imaging systems such as the Aga Thermovision have been used previously (Tice and Larsson, 1978, Geraci, 1983, and Geraci and Caltabiano, 1983) little attention has been focussed on the more expensive, but much higher performance, military forward looking infrared (FLIR) imaging systems.

Table 1 Summary of previous literature - airborne video based data acquisition systems

Spectral Region	Author(s)	Application
(0.4-0.70μm) Panchromatic	Mussakowski, 1983	Water assessment
(0.4-0.7μm) 3 band colour	Vlcek, 1983 Vlcek and Cheung, 1983	Forestry "
(0.4-1.1μm) 3 band 'false' colour	Meisner and Lindstrom, 1985 Everitt and Nixon, 1985 Nixon, Escobar and Menges, 1985	Rangeland studies Vegetation studies
(0.55-1.1μm) 4 band Multispectral	Vlcek and King, 1984 Vlcek, King and Shemilt, 1984	Vegetation Vegetation Mapping
(3-5.6μm, 8-14μm) Thermal Infrared	Tice and Larsson, 1978 Geraci and Caltabiano, 1983 Geraci, 1983 Clement, 1985 Amin, 1986	Heat loss Oil pollution Water quality Water quality Geothermal Mapping

3 THERMAL VIDEO FRAME SCANNERS (TVFS)

A TVFS system consists of four basic components:
(a) a scanning system which views the scene to be imaged and focusses the incoming infrared radiation,
(b) a ferro-magnetic or semi-conductor detector which measures the thermal variations in the scene,
(c) a CCTV compatible display system where the thermal variations are used to modulate the intensity of an electron beam, and
(d) a video cassette recorder (VCR) to record the thermal imagery.

Unlike thermal IR linescanning systems which are designed specifically for airborne use, the TVFS systems under consideration are designed for both ground and airborne operation. Consequently, the optical configuration, scanning mechanisms and detector arrays differ from those used for thermal linescanning and this leads to significantly different imaging systems being employed.

3.1 Imaging techniques

In general terms a thermal imager operates by mechanically scanning a focussed beam of incoming radiation onto an infrared detector. In its simplest form, such as in a thermal IR linescanner, the scanning is performed in one direction, perpendicular to the direction of flight, by an oscillating mirror or rotating prism. The incoming radiation is then focussed onto a single detector, normally of mercury cadmium telluride (HgCdTe). The 'frame' scanning in this case is performed by the forward motion of the aircraft.

If the imager is to be used from a static platform then some method of scanning in two dimensions is required. There are, however, several practical difficulties with such a single element detector design if a flicker free, high spatial resolution image is to be obtained. Consequently, if high resolution is required it is often more efficient to use a multi-element detector design. Three distinctive arrangements are possible: parallel, serial and matrix scanning.

Parallel scanning involves using a single column of detectors arranged so that each detector element scans a single line in the image. This arrangement reduces the scanning speed. However, the performance of each of the detector elements needs to be similar if the formation of a 'streaky image' is to be avoided. Further significant reductions in scan speed can also be achieved by using band interlaced scanning techniques (Chiari and Morten, 1982).

Serial scanning systems, in contrast, operate by using a single row of detector elements. The cumulative output is obtained by summing the individual signals from each detector. In order to achieve this it is necessary to include separate pre-amplifiers and delay line circuitry for each detector element. Although the electronics of this design are more complex, since the same detector element scans all lines in the image the uniformity of the image is more consistent than with the parallel system.

A compromise arrangement which combines the scene uniformity advantages of the serial scan approach with the high scan speeds of the parallel arrangement is the mixed parallel/serial or 'matrix' design. Inevitably, as the number of detector elements increases, so the number of electrical connections also increases. Although this difficulty, and the consequent signal processing complications can be overcome, the design trend in recent years has been to use SPRITE detectors.

As mentioned previously, with conventional serial scanning the output from each IR detector is pre-amplified, delayed and then added to the signal which is generated in the following element. SPRITE (SIgnal PRocessing In The Element) detectors overcome the need for these separate connections. In the SPRITE design the row of individual detectors is replaced by a single strip of HgCdTe with only two connections and one pre-amplifier. Eight element SPRITE detectors are used in most TVFS systems, such an arrangement is equivalent to a conventional array consisting of 64 discrete elements.

Although SPRITE detectors eliminate many of the connections and much of the circuitry, they still require some form of mechanical scanning to be carried out. An area of considerable interest at present is the development of 'staring arrays' i.e. a matrix of infrared detectors. In this case the function of the optics is simply to focus the incoming radiation onto the matrix of detectors located in the focal plane of the camera. Whilst it is likely that such arrangements will eventually replace the SPRITE design, this is unlikely to occur until very dense matrices can be formed thus enabling high resolution imagery to be produced.

3.2 System review

A wide range of thermal imaging systems are currently available and Table 2 outlines the technical specifications associated with a selected sample.

Table 2: Technical Specifications : Thermal Video Frame Scanning Systems.

	Spatial Resolution (mr)	MRTD* (^{0}C)	Spectral Range (μm)
AGA Thermovision	5.8	0.1	3-5.6 8-12
GEC V1010 TICM 11	2.27	0.1	8-13
Hawkeye HT4	2.1	0.15	8-13
Rank Pullin Controls SS600	2.1	0.15	8-13
Inframetrics IRTV-445G	2.0	0.4	8-12
FLIR Systems 1000A	1.87	0.2	8-12
Barr and Stroud IR18	1.73	0.38	8-13

MRTD* = Minimum Resolvable Temperature Difference

As mentioned in the literature review several previous authors have discussed the use of the AGA Thermovision range of instruments for airborne use. These instruments, of which model 782 is the most recent, are single detector, low spatial resolution systems. Two versions of this instrument are available for sensing in the 3 to 5 μm and 8 to 14μm regions of the spectrum. An image consisting of 100 elements/line over a 280 lines/frame format refreshed at 25Hz and interlaced four to one is produced by the scanner. In addition to the low spatial resolution of the instrument the slow refresh rate creates image registration problems when sensing from a moving platform. A further disadvantage is the difficulty of producing a vertical image. In order to achieve this a 45^0 mirror has to be used so as to avoid the liquid nitrogen from the cooling system being expelled.

The Inframetrics IRTV 445G can however be gimbal mounted for airborne use. It uses a 4 element HgCdTe detector and creates a 400 elements/line image over a 445 lines/frame format refreshed at at a frame rate

of 30Hz. Both AGA and Inframetrics have also developed dedicated image processing software for enhancing the imagery and extracting quantitative temperature data.

The majority of high performance TVFS systems operating in the 8 to 13 μm band have been developed for avionic and military applications. Development of the current range of UK military imagers began in 1976 and at that time the decision was made to adopt a modular design philosophy. This is known as the Thermal Imaging Common Module (TICM) programme. The GEC V1010 TICM 11 is representative of this design. Instruments of this type can be mounted in either a fixed wing aircraft or a helicopter, in the latter case the scanner is mounted in a pod to allow a forward looking image to be obtained.

The Rank Pullin Controls (RPC) SS600 thermal imager differs from the conventional TICM scanning design which uses a flapping mirror to scan the field in elevation and a rotating polygon to provide the line scan. In contrast, the RPC design uses two coaxially mounted rotating polygons to form the two dimensional scan pattern. This unique design (Figure 1) is claimed to result in a lighter, more robust and more compact imager suitable for use where severe vibration may be experienced.

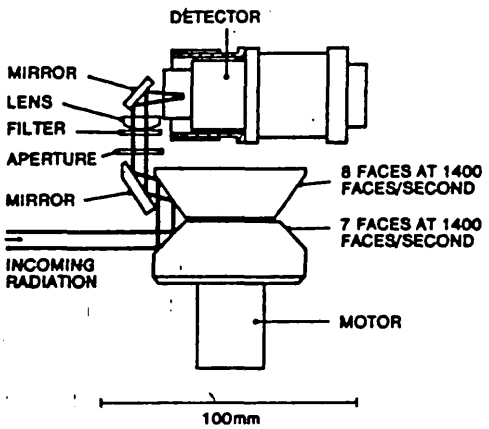


Figure 1. Rank Pullin Controls SS600-scanning system (Lettington, 1985)

The Barr and Stroud IR18 is another TVFS system based on a modular design approach. A narrower scanner field of view is used compared with several other equivalent systems. In this case also a novel optical-mechanical scanning design is used which is discussed in the following section of the paper.

3.3 The Barr and Stroud IR18 Mk II Thermal Video Frame Scanner

The scanner used in these tests was developed initially for military use, but it is now available for commercial remote sensing applications. Table 2 outlines some of the technical specifications of the sensor and Figure 2 illustrates the principle of operation of the scanner.

To achieve a video compatible image, two main criteria have to be met: firstly, the resultant image must be rectilinear and, secondly, it should have a 625 line CCTV output (525 lines in the USA). In order to achieve this, banded scanning is carried out using a high speed rotor, turning at 39,000 rpm, to give

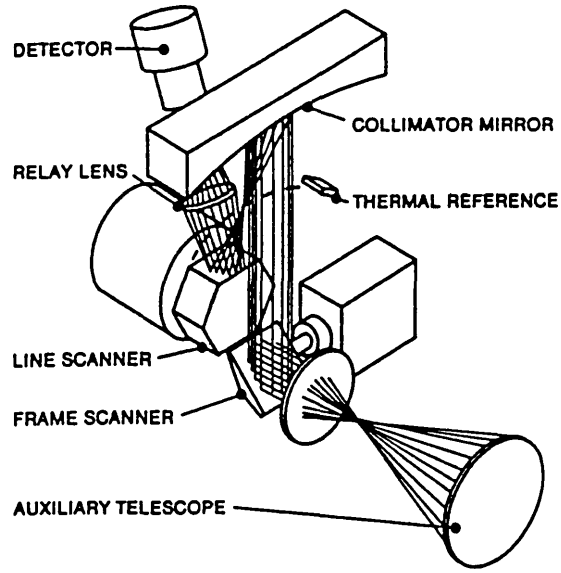


Figure 2. Barr and Stroud IR18 Mk 11 thermal video frame scanner - principle of operation (Berry and Runciman, 1981)

the line format, and a slower galvanometer driven rotor to achieve the frame transfer at 50Hz. Petrie (1983) has estimated that the system has a data collection rate equivalent to 19 X 10^6 pixels/second.

The optics of the IR18 are used to collimate the incoming radiation and, by using afocal telescopes, enable a variety of magnification ratios to be used, so producing various object field of views (Table 3). Such an arrangement ensures that the incident radiation forms a bundle of parallel rays which can then be focussed onto the frame scanning rotor. A concave mirror then re-collimates the image onto the rotating hexagonal rotor and subsequently onto the detector.

Table 3: Barr and Stroud IR18 - object field of view with various auxiliary telescopes

Magnification Ratio	Object Field of View (H^0 X V^0)
1	38.0 25.5
1.5	25.3 17.0
6	6.30 4.25
9	4.22 2.8
14	2.7 1.8

The detector used is an eight-element Mullard 'SPRITE', although four elements are normally preselected for use. The scanning sequence is such that on each successive frame scan a different element (of the four active elements) scans the same part of the scene. A consequence of this is that if the imager output is integrated over a number of frames, detector non-uniformity errors can be eliminated. Complex microprocessor controlled circuitry controls the synchronisation of the rotor speeds and signal outputs and the transfer of the data onto video tape. A conventional U-Matic video tape recorder is used to store the imagery.

Included in the scanner design is a thermal reference plate which is placed at one end of the linescan locus so that referencing takes place at the scan rate. This arrangement enables the selection of five sensitivity ranges covering the following temperature ranges: 2^0, 4^0, 8^0, 16^0 and 32^0. Cooling is carried out using high pressure compressed air supplied to a Joule-Thompson minicooler. Further design details can be found in Berry and Runciman, 1981.

4 DATA ACQUISITION

To date, both the Barr and Stroud IR18 and Rank Pullin Controls thermal scanners have been evaluated for remote sensing applications. The initial assessment of the former system was performed in August 1984 by flying the camera, in a Cessna 172 light aircraft, over several tests sites in the West of Scotland.

The Cessna 172 is a four seat aeroplane with two access doors and one luggage door immediately behind the pilot's door. Several modifications were necessary to the aircraft to enable the IR18 to be used for the production of vertical imagery. The main features of the design which was used are described in Ekin (1984). Briefly the approach adopted by Ekin was to design a retractable camera carriage, attached to the luggage door, which could be extended in flight by the camera operator. A pair of rails, on which four wheels with V - shaped groves ran ensured the smooth movement of the rig into the extended position. Figure 4 illustrates the IR18 mounted in a slightly modified version of the rig described.



Figure 4: Barr and Stroud IR18 mounted on retractable camera rig in a Cessna 172 light aircraft.

The adoption of a light aircraft and an easily manipulated TVFS system of the type described offers several advantages to the user. Firstly, by using widely available light aircraft the operational costs of the survey can be significantly reduced. Secondly, once the site of interest has been located and the camera rig extended, the scanner operator can view, via a small monitor, the area being imaged by the instrument. Thirdly, after the survey has been completed immediate confirmation that the data has been recorded successfully can be achieved by playing back the video tape, either in flight, or more commonly when the aircraft returns to base.

For other applications however, it may be unacceptable to use the scanner from a light aircraft, either because of the need for a more stable platform, or the need to use the camera in conjunction with several other sensors. More recent experience of using the IR18, for defence and commercial applications by ERSAC Ltd. has involved using the IR18 in a purpose built camera mount in a twin engined Piper Aztec aircraft. For maximum flexibility of operations and integration with a standard aerial camera, a special adapter plate was designed and deveolped by ERSAC and Barr and Stroud to enable the IR18 to be mounted and operated through a standard Wild RC8 aerial camera mount. This configuration is illustrated in Figure 4. The most significant advantage of this design is the capability for full yaw control; this enables higher precision to be achieved when flying parallel flightpaths for the acquisition of imagery over large areas



Figure 5: Barr and Stroud IR18 in an adapted Wild RC8 aerial camera mount in a Piper Aztec aircraft

5 DATA PROCESSING

The processing of video based thermal data consists of three distinct phases:

(a) conversion of the analogue video recording to a digital format,

(b) pre-processing of the digital data to remove image inconsistencies, and/or to rectify and geometrically correct the image, and

(c) processing and analysis of the data using conventional image processing techniques.

The analogue to digital (A to D) conversion of video tape is carried out using a digital frame store (DFS) or 'frame grabber'. These instruments are now widely available at relatively low cost. The particular model used for these investigations was the Quantel 3001 DFS controlled by an IDP 3000 image processing system.

The pre-processing, and processing stages, were carried out on a GEMS image processing system. One particularly useful pre-processing technique is that

of de-striping. This was performed by running a rectangular box filter across the image. The advantage of such a process is to produce a more consistent image. However, since the filter cannot discriminate between small scene variations and systematic and random noise, both tend to be removed or reduced in amplitude.

For applications where there is an interest in the detection of subtle linear features, (for example the identification of subterranean field drains), it has been found useful to perform a variable filter edge enhancement, the filter size being dependant on the particular feature of interest. The most commonly used filters range from 3 X 3 to 15 X 15 pixels in dimension. Once the feature of interest has been enhanced by this process it has also been found useful to apply a median filter; this resamples the data in a 3 X 3 box. By using this process isolated pixels caused by the edge enhancement procedure tend to be removed.

A further processing operation is required if the project requires quantitative absolute temperature values. In order to obtain such information calibration of the imagery needs to be carried out using ground temperature measurements obtained at the time of the aircraft overpass. Correlation of these ground measurements with the unstretched digital numbers can be used to establish the limits for a quantitative density slice representing discrete temperature levels. The application of this technique to heat loss monitoring is presented in section 6.

6 APPLICATIONS

As mentioned previously several flights of the Barr and Stroud IR18 TVFS were performed during August 1984 as part of a preliminary research assessment of the instrument. As a consequence of early morning mist these flights took place from late morning until late afternoon rather than, as planned, during the post-sunset/ pre-dawn period. The imagery obtained was, nevertheless, useful since it enabled the potential of the imagery to be assessed together with methods of processing the imagery (Dale, 1985)

More recently both the Barr and Stroud and RPC TVFS systems have been flown, for a variety of projects by ERSAC Ltd. The following material reports some of the most significant applications to date.

6.1 Heat Loss Monitoring.

Several projects have been carried out using the Barr and Stroud IR18 to assess heat loss from industrial and residential buildings. As mentioned above, in order to provide quantitative estimates of localised heat loss from TVFS imagery it is necessary to establish calibration data to relate the grey scale variations on the image to emitted radiance and eventually to estimates of heat emission. In addition to deriving emitted temperature it is also necessary to evaluate the variations in emissivity over the scene.

Ground temperature measurements can be obtained using either conventional in-situ temperature probes (e.g. thermistors or thermometers) or alternatively by using ground based radiometric measurements from instruments such as the AGA Thermovision. As mentioned earlier the AGA Thermovision is a low spatial resolution system; however, it is possible, by using a blackbody reference, to obtain ground temperature measurements. By using the Stefan-Boltzmann law the heat loss for a particular temperature value can be derived and a suitable density slice for the IR18 imagery obtained. The results of one particular set of data are shown in Table 4.

If the computed heat losses (Q) from Table 4 are regressed on the mean digital number indicating emitted radiance (R) from the IR18, then a regression coefficient (r) of 0.924 is obtained. The

Table 4. Calibration of Heat Loss from Barr and Stroud IR18 using data from the AGA Thermovision

Colour Coding	Mean Radiance(R)	AGA Temperature (^{0}c)	Mean Computed Heat Loss (Q) (Watts)
Blue	44	2.5	11
Cyan	74	4.0	17
Magenta	80	4.7	21
Grey	95	6.3	28
Black	110	6.5	29
Red	131	7.6	33
White	155	13.8	63

coefficient of determination (r^2) is 0.854, indicating that 85% of the computed heat losses from the calibration results above can be explained by the corresponding radiance values (R) as recorded and digitised from the IR18 sensor data. The resulting heat loss regression equation for the IR18, for the 8^0c gain setting was of the form;

$$Q = 0.4185R - 12.537 \quad ....(1)$$

with a standard error of ±6 watts.

6.2 Thermal Mapping of Roads

A novel environmental engineering application of the IR18 TVFS system has involved thermal mapping of road surfaces in winter to identify accident black spots. The primary objective of this night-time aerial survey work was to assist road engineers in the siting of ground sensors to monitor ice conditions around high accident risk sections of road. It is also hoped that the data will enable improvements and ecomomies in road gritting operations to be achieved.

The survey was carried out in Scotland in February 1986. Ground temperature measurements were obtained and used to determine calibration data for the IR18. By using this data in conjunction with the GEMS image processing system it was possible to digitise and calibrate a selected number of frames of data. Using a digital to analogue converter it was then possible to create from these selected scenes, a density sliced video image of the original analogue tape. This feature has been shown to be of considerable benefit to non specialists viewing the imagery.

6.3 Geotechnical Site Investigations

An assessment of the potential of the IR18 to detect solution features in chalk will be carried out during the summer of 1986. Solution features are a serious geotechnical hazard and generally provide unreliable bearing capacity for foundations. A review of the significance of solution features and the use of remote sensing techniques for their detection is provided in Kennie and Edmonds (1986). The relative performance of TVFS and thermal infrared linescanning techniques will also be investigated during this exercise.

6.4 Drainage, Sewer Collapse and Utilities Surveys

A recent application of the IR18 has involved the detection of subsurface utilities and problems associated with such features. In this case the TVFS was mounted on a boom attached to a car and was used in conjunction with a ground impulse radar system (λ = 6 to 330 centimetres).

This work was carried out in London during June 1985 (Stove, 1985) with the aim of identifying leaking water mains, sewer collapses and piperuns below the pavement surface. The main conclusion from this work was the complementary nature of the two systems since the impulse radar provides information relating to the depth of features beneath the terrain, that is in the vertical plane, whereas the TVFS provides data about the spatial extent of the phenomenon in the horizontal plane. For example, at one particular sewer site the ground impulse radar indicated a marked discontinuity between the structure of the road and the fill material beneath; normally this is indicative of poor compaction or voiding in this zone. Also, a reduction in the radar signal velocity occurred which indicated that the material in this zone was of a much higher moisture content than the surroundings. Analysis of the TVFS imagery over the same area was carried out by density slicing the thermal emissions to indicate 1^{0}c temperature intervals. A clear indication of the size of the sewer was obtained and a marked increase in emission was detected corresponding to the discontinuity recorded by the ground impulse radar.

The work on this particular application is at an early stage of development and a method of automating the measurement of any correlation between the data from these two complementary sensors is currently being considered. One possible solution may be to reformat the output from the ground impulse radar and record this on videotape (Stove, 1985). The radar data could then be digitally formed into image form and registered with the TVFS data.

6.5 Further Applications

In addition to the projects outlined above TVFS systems have potential application for many other environmental engineering projects including: oil pollution monitoring, monitoring of spontaneous combustion of waste tips, power station cooling water discharge monitoring, ground water and spring detection and hydraulic leakage from dams.

7 CONCLUSIONS

TVFS systems differ considerably both in optical design and in radiometric characteristics from thermal linescanning instruments. A wide range of instruments are currently available using distinctive imaging techniques. Such instruments provide an alternative method of acquiring thermal data and, in general, at lower cost, if a light aircraft is used as the measuring platform. By using suitable image processing techniques and ground temperature measurements for calibration purposes quantitative thermal characteristics of the terrain can be deduced. It is suggested that TVFS systems have a significant role to play in the acquisition of thermal data for environmental engineering projects particularly when used in conjunction with other complementary remote sensing or geophysical sensors.

8 ACKNOWLEDGEMENTS

The authors wish to thank the following organisations and individuals for their assistance with this project; Barr and Stroud Ltd, University of Surrey Research Committee, Professor G. Petrie, University of Glasgow, Mrs V. Brown, University of Surrey.

9 REFERENCES

Amin, A. 1986. Geometric aspects of airborne thermal video frame scanners. Ph.D. Thesis, University of Glasgow, (in preparation)

Berry, J.P. and Runciman H.M. 1981. Radiation scanning system. Patent Specification 1586099, The Patent Office, London,.

Clement, W.Z. 1985. Airborne video detection of geothermal areas on Mt. St. Helen's, Washington. 19th Int. Symp. on Remote Sensing of Environment, Ann Arbor

Chiari, J.A. and Morten F.D. 1982. Detectors for thermal imagers. Electronic Components and Applications, 4(4), 242-252.

Dale, C.D. 1985. The interpretation and processing of video based thermal infrared imagery. Project report, Dept. of Civil Eng., Univ. of Surrey, 94pp.

Ekin, W.H. 1984. The development of an inexpensive retractable vertical camera rig for a light aircraft. Photog. Record, 11(63), 311-317.

Everitt, J.H. and Nixon P.R. 1984. False colour video imagery: a potential remote sensing tool for range management. Photog. Eng. and Remote Sensing, 51(6), 675-679.

Geraci, A.L. and Caltabiano, T. 1983. Water quality monitoring of eastern Sicily's coastal zone near Taormina using remote sensing techniques Proc. of the 3rd Sypm. on Coastal and Ocean Management. ASCE/San Diego, California, 194-205.

Geraci, A.L. 1983. Remote sensing analysis of oil pollution in Augusta Bay, Sicily. Proc.of EARSEL/ESA Symp. on Remote Sensing Applications or Env. Studies. Brussels, Belgium, ESA-SP 188.

Kennie, T.J.M. and Edmonds, C. 1986. The location of potential ground subsidence and collopse features in soluble carbonate rocks by remote sensing techniques. ASTM Special Publication, Geotechnical Engineering Applications of Remote Sensing, 26pp.

Lettington, A.H. 1985. Infrared scanning techniques An Introduction to Military Thermal Imaging, SIRA, 5pp.

Meisner, D.E. and Lindstrom, O.M. 1985. Design and operation of a colour infrared aerial video system. Photog. Eng. and Remote Sensing, 51(5), 555-560.

Mussakowski, R.S. 1983. The application of video remote sensing to resource surveys and environmental monitoring Proc. of the 49th Annual Meeting of the ASP, Washington, 63-69.

Nixon, P.R., Escobar, D.E. and Menges, R.M. 1985. A multiband video system for quick assessment of vegetal conditions and discrimination of plant species. Remote Sensing of Environment, 17, 203-208.

Petrie, G. 1983. The philosophy of digital a d analytical photogrammetric systems. Proc. of the 39th Photogrammetric Week, Univ. of Stuggart, 27pp.

Tice, R.R. and Larsson, R.A. 1978. Experience with hybrid infrared-visible viewing from low altitude aircraft. ISP-IUFRO Symp. on Remote Sensing for Observation and Inventory of Earth Resources and the Endangered Environment. Freiburg 2-8th July.

Vlcek, J. 1983. Videography. some remote sensing resource surveys and environmental monitoring. Proc. 8th Canadian Symp. on Remote Sensing, Montreal, 91-99.

Vlcek, J. and Cheung, E. 1983. Video image analysis 8th Candaian Symp. on Remote Sensing, Montreal, 63-69.

Vlcek, J. and King, D. 1984. Digital analysis of multispectral video imagery. Proc. of the 50th Annual Meeting of the ASP, 628-632.

Vlcek,J., King, D. and Shevilt, S. 1984. Multiband video imaging: early trials and results. 9th Canadian Symp. on Remote Sensing.

Stove, G.C. 1985. Scottish penal establishments: aerial thermographic (heat loss) infrared surveys. Final Report by ERSAC Ltd to SDD Prisons Dir. 142pp

Stove , G.C. 1985. Remote sensing of subsurface water and sewerage problems from utility mains in London First Report by ERSAC Ltd to Thames Water, 32pp.

INT. J. REMOTE SENSING, 1987, VOL. 8, NO. 7, 1077–1084

Remote Sensing Letters

Airborne thermal mapping for winter highway maintenance using the Barr and Stroud IR18 thermal video frame scanner

G. C. STOVE

ERSAC Ltd., Peel House, Livingston, West Lothian, Scotland

T. J. M. KENNIE

Department of Civil Engineering, University of Surrey, Guildford, Surrey, England

and A. HARRISON

Scott, Wilson, Kirkpatrick and Partners, Edinburgh, Scotland

(Received 10 November 1986, in final form 1 May 1987)

Abstract. The use of a portable video-based airborne thermal sensor for thermal mapping of roads is described. The design and operational characteristics of the Barr and Stroud IR18 thermal video frame scanner are discussed together with image processing requirements for a project of this nature. Finally, a summary of the results obtained for a project carried out in the Lothian region of Scotland is presented.

1. Introduction

Each winter U.K. local authorities spend more than £100 million on measures to keep roads free of ice and snow. Furthermore, recent research has indicated that spending on rock salt, in particular, may be double that which is actually required (Thornes 1985). Thermal surveying is one of several new techniques which have been used by highway engineers in the U.K. in recent years to help reduce these costs. Information provided by such surveys can be used not only to locate areas prone to icing but also identify ground sites where detailed temperature monitoring should be carried out by sensors embedded in the road surface.

A thermal map, or more correctly thermal profile, is simply a longitudinal section along a road which illustrates the surface temperature of the highway pavement. Thornes (1985) has shown that given consistent ambient weather conditions thermal profiles can provide a representative assessment of the temperature along a road and may be used in conjunction with permanent remote ground sensors to assess the range of temperatures throughout a road network. The combined use of road ground sensors, weather radar, thermal profiling and road surface temperature models have all been used successfully to help reduce the level of use of de-icing agents and road clearing patrols.

Thermal profiles were originally prepared from ground surveys undertaken using special survey vehicles equipped with infrared sensors and microcomputer logging equipment. Although relatively cheap to operate and applicable under a wide variety of weather conditions, ground surveys inevitably required some time to execute and the thermal profiles produced were liable to generate unrepresentative results if the ambient weather conditions changed during the survey. Consequently this necessitated

several repeat surveys which, in some cases, may have offset the cost advantages of this approach.

Aerial methods offer the advantages of speed, ease of interpretation and provide direct evidence of the temperature of not only the road surface but also of the areas surrounding the road being investigated (Rosema and Welleman 1977). This additional spatial information may be of particular interest in areas where temperature inversions occur, since this condition is often associated with the formation of ice. A recent project examined the use of aerial remote sensing techniques to produce thermal maps of trunk roads in the Lothian region of Scotland using the Barr and Stroud IR 18 thermal video frame scanner (TVFS).

The Barr and Stroud IR 18 is a video-based thermal scanner which can be mounted in a light aircraft. Several practical benefits accrue from the use of video systems including system portability, almost real-time verification of data capture, rapid turnround (since there is no photographic processing stage), low cost of operation (since the sensor can be mounted to a light aircraft) and ease of data duplication (Meisner and Lindstrom 1985).

2. Thermal video frame scanners (TVFS)

A TVFS system consists of four basic components:

(*a*) an optical mechanical scanning system which views the scene to be imaged and redirects the incoming infrared radiation;
(*b*) a ferromagnetic or semiconductor detector which measures the thermal variations in the scene;
(*c*) a CCTV (close circuit television) compatible display system where the thermal variations are used to modulate the intensity of an electron beam;
(*d*) a video cassette recorder (VCR) to record the thermal imagery.

Unlike thermal IR linescanning systems which are designed specifically for airborne use, TVFS systems are designed for both ground and airborne operation. Since such a system may therefore be used from both static and moving platforms, some method of scanning in two orthogonal directions is required. This leads to significantly different optical configurations, scanning mechanisms and detector arrangements from those in thermal IR linescanners. The simplest approach to adopt involves the use of a scanning mechanism which images onto a single detector element (such as the AGA Thermovision) or onto a pyroelectric vidicon (such as the Insight 80 series). There are, however, several practical difficulties with such designs if a flicker-free, high spatial resolution image is to be obtained, particularly when imaging from a moving platform. If high spatial resolution is required it is more common to use a multi-element detector design, since much higher scanning speeds can be accommodated with such an arrangement. Three distinctive arrangements are possible: parallel, serial and matrix scanning.

Parallel scanning involves using a single column of detectors arranged so that each detector element scans a single line in the image. This arrangement reduces the scanning speed. However, the performance of each detector elements needs to be similar if the formation of a 'streaky image' is to be avoided.

Serial scanning systems, in contrast, operate by using a single row of detector elements. The output from each IR detector is pre-amplified, delayed and added to the signal which is generated in the following element. In order to achieve this it is necessary to include separate pre-amplifiers and delay line circuitry for each detector element. Although the electronics of this design are more complex, since the same

detector element scans all lines in the image, the uniformity of the image is more consistent than with the parallel system.

A compromise arrangement which combines the scene uniformity advantages of the serial scan approach with the high scan speeds of the parallel arrangement is the mixed parallel/serial or 'matrix' design. Inevitably, however, as the number of detector elements increases, so the number of electrical connections also increases. Although this difficulty, and the consequent signal processing complications can be overcome with matrix designs, the design trend in recent years has been to use SPRITE (Signal PRocessing In The Element) detectors (Elliott 1981).

As mentioned previously, with conventional serial scanning the output from each IR detector is obtained by summing the individual signals from each detector. Each element therefore requires a seperate set of delay and amplification connections. SPRITE detectors overcome the need for these separate connections. In the SPRITE design the row of individual detectors is replaced by a single strip of HgCdTe with only two connections and one pre-amplifier. A current source is applied to the strip of HgCdTe, by çarefully choosing the bias current the velocity of the induced carrier drift can be matched to the scan velocity. Consequently as each pixel traverses the length of the detector the carriers it excites also accumulate as they travel along the detector. Thus the time delay and integration is carried out within the element, rather than by a series of separate components. Eight-element SPRITE detectors are used in most TVFS systems; such an arrangement being equivalent to a conventional array consisting of 64 discrete elements.

Although SPRITE detectors eliminate many of the connections and much of the circuitry, they still require some form of mechanical scanning to be carried out. An area of considerable interest at present is the development of 'starring arrays' i.e. a matrix of infrared detectors. In this case the function of the optics is simply to direct the incoming radiation onto the matrix of detectors located in the focal plane of the camera. Whilst it is likely that such arrangements will eventually replace the SPRITE design, this is unlikely to occur until very dense matrices can be formed which will enable high spatial resolution imagery to be produced.

3. Barr and Stroud IR18 MKII TVFS

The Barr and Stroud IR18 TVFS was developed initially for military use, but it is now also available for commercial remote sensing applications. The principal technical specifications are summarized in the table.

The optics of the IR 18 are used to focus the incoming radiation. By using afocal telescopes, a variety of magnification ratios can be produced with various object field of views (see the table). This optical arrangement also ensures that the incident radiation forms a bundle of parallel rays which can be focused onto a relatively slow frame scanning rotor. A concave mirror then re-collimates the image onto the high speed rotating hexagonal rotor and subsequently onto the detector (figure 1).

The detector used is an eight-element Mullard 'SPRITE'; although four elements are normally preselected for use when the detector is fitted. The scanning sequence is such that on each successive frame scan a different element (of the four active elements) scans the same part of the scene. A consequence of this is that if the system output is integrated over a number of frames, detector non-uniformity errors can be eliminated. Complex microprocessor circuitry controls the synchronization of the rotor speeds, the signal outputs and the transfer of the data onto video tape. A conventional U-Matic video tape recorder is used to record the imagery.

Specifications of the Barr and Stroud IR18 TVFS.

Instantaneous field of view	1·73 mrad
Spectral range	8–13 μm
Minimum resolvable temperature difference (MRTD) with 1·5 × telescope, over 2°C temperature range	0·1°C
Detector	8-element SPRITE
Weight	7·4 kg
Object field of view (H° × V°)	
0·5 × magnification	76·0° × 51·0°
1 × magnification	38·0° × 25·5°
1·5 × magnification	25·3° × 17·0°
6 × magnification	6·3° × 4·2°
9 × magnification	4·2° × 2·8°
14 × magnification	2·7° × 1·8°

In order to achieve a video compatible image, two main criteria have to be met: firstly, the resultant image must be rectilinear, and secondly, it should have a 625 line CCTV output (525 lines in the USA). In order to achieve this, banded scanning is carried out using a high speed rotor, turning at 39 000 r.p.m., to give the line format, and a slower galvanometer driven rotor to achieve the frame transfer at 50 Hz. Petrie (1983) has estimated that the system has a data collection rate equivalent to 19×10^6 pixels/s.

Included in the scanner design is a thermal reference plate which is placed at one end of the linescan locus so that referencing takes place at the scan rate. This arrangement enables the selection of five sensitivity ranges covering the following temperature ranges: 2°, 4°, 8°, 16° and 32°. Cooling is carried out using high pressure compressed air supplied to a Joule–Thompson minicooler (Berry and Runciman 1981).

4. Operation of the Barr and Stroud IR18 TVFS in light aircraft

To date the Barr and Stroud IR18 TVFS has been operated in two light aircraft: a Cessna 172 and a Piper Aztec. Details relating to the use of the TVFS in the former case

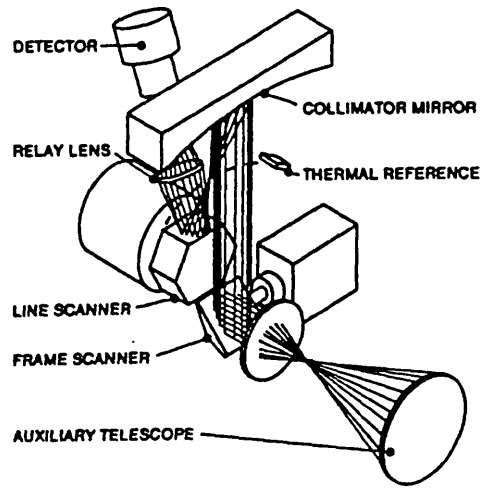


Figure 1. Optical design of the Barr and Stroud IR18 Thermal Video Frame Scanner.

can be found in Ekin (1984, 1987) and Kennie *et al.* (1986). For some applications, however, it may be unacceptable to use the scanner from a small aircraft such as the Cessna 172, either because of the need for a more stable platform or the need to use the camera in conjunction with several other sensors. More recent experience of using the IR 18, for defence and commercial applications by ERSAC Ltd., has involved using the IR 18 in a purpose-built camera mount in a twin-engined Piper Aztec aircraft. For maximum flexibility of operations and integration with a standard aerial camera, a special adapter plate has been designed to enable the IR 18 to be mounted and operated through a standard Wild RC8 aerial camera mount. This configuration is illustrated in figure 2. The most significant advantage of this design is the capability for full yaw control; this enables higher precision to be achieved when flying parallel flightpaths for the acquisition of imagery over large areas.

The use of a light aircraft and an easily manipulated TVFS system of the type described offers several advantages to the user. Firstly, by using widely avaiable light aircraft the operational costs of the survey can be significantly reduced. Secondly, once the site of interest has been located and the camera rig extended, the scanner operator can view, via a small monitor, the scene being scanned by the instrument. Thirdly, after the survey has been completed immediate confirmation that the data have been recorded successfully can be achieved by playing back the video tape, either in flight, or more commonly when the aircraft returns to base.

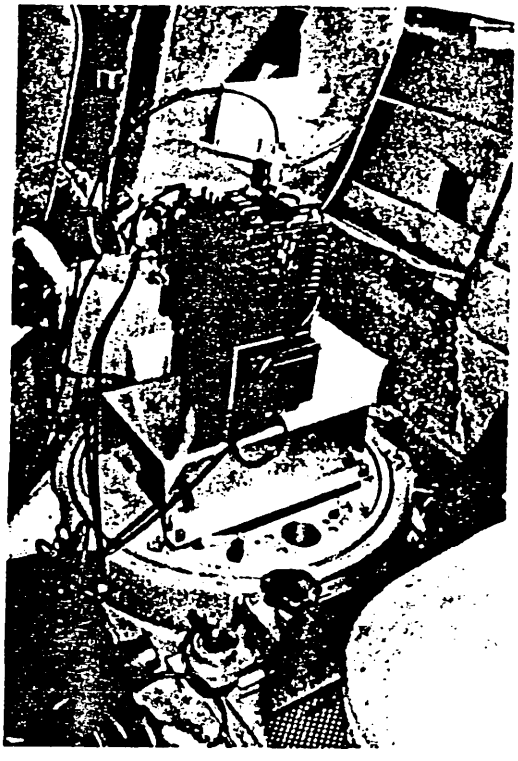


Figure 2. Barr and Stroud IR18 TVFS mounted in an Wild RC-8 aerial camera housing.

5. Processing of TVFS imagery

The processing of TVFS data consists of three distinct phases:

(*a*) conversion of the analogue video recording to a digital format;
(*b*) pre-processing of the digital data to remove image inconsistencies (for example using destriping techniques);
(*c*) processing and analysis of the data using conventional image processing techniques (in particular contrast stretching, density slicing and edge enhancement).

The analogue to digital (A to D) conversion of video tape is carried out using a digital frame store (DFS) or 'frame grabber', and the latter two phases carried out using a conventional image processing system. An optional phase which may be necessary is geometric rectification of the image using suitably positioned ground control.

Unlike linescanning systems, TVFS do not normally include a calibrated black body reference for quantitative temperature measurements. If a project, however, requires quantitative measures of the absolute temperature values, calibration of the imagery can be carried out using ground temperature measurements obtained at the time of the aircraft overpass. Correlation of these ground measurements with the unstretched digital numbers can be used to establish the limits for a semi-quantitative density slice representing discrete temperature levels. The results of such a process do not, however, account for spatially variable emissivity and the effects of atmospheric attenuation.

6. Thermal road mapping survey

Optimum weather conditions for thermal mapping occurs in winter months when anticyclonic conditions and clear skies exist, giving maximum cooling to the road surface. These conditions also produce temperature inversions, the local meteorological effects of greatest concern to the highway engineer concerned with winter road maintenance operations. Temperature inversions occur as a result of cold dense air draining from higher altitudes to settle in low-lying areas, producing an abnormal lapse rate where air temperature rises with increasing height. This condition may lead to the formation of frost or ice on the road surface. Identification of these areas is of primary importance in thermal mapping and the survey in the Lothian region of Scotland was not initiated until there was a high degree of confidence that such conditions would exist in the survey area.

Following several days on standby, the survey took place at 04.40 hours on the morning of 27 February 1986 towards the end of a prolonged cold period over the U.K. Meteorological records at Edinburgh airport indicated that this period was the coldest recorded since 1948. Ground temperatures reported by the meteorological office ranged from +1°C on the coast to −8°C inland. In order to enable road surface temperatures to be determined from the thermal imagery, calibrated, ground temperature measurements were obtained by ground teams using digital thermometers. Observations were obtained at several road junctions where ice formation normally occurred. In addition measurements were also obtained at other randomly chosen interchanges and junctions at which it was possible to register the road temperature measurement with its location on the thermal imagery. In all cases temperature profiles were measured across the carriageway and slip roads. Reference air temperatures were also collected at 0·5 m, 1·0 m and 2·0 m above the road surface to determine whether any local eddy or current effects were present. Ground temperatures varied during the survey from +2°C to −8°C.



Figure 3. Density sliced image (at 1 C intervals) over a temperature range of 8°C of a single frame of TVFS data. Area shown is at junction 4 of the M8 in Scotland.

The survey was flown at a height of 800 m at an airspeed of 160 knots, this enabled a complete recording of the survey on only 40 min of video tape. Navigation was accomplished using conventional radio location aids and a standard aerial navigation sight in combination with a monitor linked to the Barr and Stroud IR18 sensor. Specific frames of video imagery covering those areas where ground temperatures had been measured were selected and digitized. The relationship between the measured ground temperature and the digitized image grey scale was then used to create a semi-quantitative density slice at 1°C intervals (figure 3). The limits of the density slice were subsequently applied to successive frames of the video imagery. Visual interpretation of the video tape was then carried out and transferred to 1:10 000 scale maps of road sections and finally onto 1:50 000 scale maps.

The primary aim of the survey was to recommend to the highway authority priority sites for the installation of remote ground temperature sensors. Twelve sites, located in both the coldest and warmest sections of the road network, were eventually recommended. Such a distribution of ground sensors would therefore enable the highway maintenance engineer to have knowledge of the full range of road temperatures likely to be experienced on any particular night.

7. Conclusions

TVFS systems such as the Barr and Stroud IR18 differ considerably in design from thermal linescanning instruments. When mounted in a light aircraft such systems offer significant advantages to the remote sensing user in terms of ease of use, system portability and cost. The use of such an instrument provides highway maintenance engineers with a fast and cost-effective method of identifying areas where temperature inversions are likely to occur and to locate priority sites for detailed ground temperature monitoring. The complementary data provided by airborne remote sensing should also enable cost savings to be achieved by defining more effectively the policy to be adopted for de-icing and road clearing patrols over a road network.

Acknowledgments

The authors wish to acknowledge the assistance and cooperation of the Director of Highways, Lothian Regional Council, Scotland.

References

BERRY, J. P., and RUNCIMAN, H. M., 1981, Radiation scanning system. *Patent Specification 1586099* (London: The Patent Office).

EKIN, W. H., 1984, The development of an inexpensive retractable vertical camera rig for a light aircraft. *Photogrammetric Record*, **11**, 311–317.

EKIN, W. H., 1987, A new lightweight camera rig. *Photogrammetric Record*, **12**, 343–348.

ELLIOTT, C. T., 1981, New detector for thermal imaging systems. *Electronic Letters*, **17**, 312–313.

KENNIE, T. J. M., DALE, C., and STOVE, G. C., 1986, A preliminary assessment of an airborne thermal video frame scanner for environmental engineering applications. *Proceedings of the 7th ISPRS, Commission VII Symposium on Remote Sensing for Resource Development and Environmental Management*, Enschede, The Netherlands (International Archives of Photogrammetry Vol. 26 Part 7/1) Balkema Press, 285–290.

MEISNER, D. E., and LINDSTROM, O. M., 1985, Design and operation of a colour infrared aerial video system. *Photogrammetric Engineering and Remote Sensing*, **51**, 555–560.

PETRIE, G., 1983, The philosophy of digital and analytical photogrammetric systems. *Proceedings of the 39th Photogrammetric Week*, University of Stuttgart.

ROSEMA, A., and WELLEMAN, A. G., 1977, Microclimate and winter slipperiness: a study of the factors influencing slipperiness with application of thermal infrared observation techniques. *NIWARS Publication, No. 38*.

THORNES, J. E., 1985, The prediction of ice formation on roads. *Journal of the Institution of Highways and Transportation*, **32**, 3–12.

APPENDIX B

DEVELOPMENTS IN AERIAL AND GROUND BASED SURVEYING AND THEIR APPLICATION TO MONITORING CHARACTERISTICS OF ENGINEERING GEOLOGICAL IMPORTANCE

Civil Engineer's Reference Book

Fourth Edition

Edited by

L S Blake

BSc(Eng), PhD, CEng, FICE, FIStructE
Consultant; formerly Director of the
Construction Industry Research and
Information Association

With specialist contributors

7.1 Introduction

Photogrammetry and remote sensing are two indirect methods of obtaining both quantitative and qualitative data about the Earth's surface or other features of interest. Since some debate exists regarding the demarcation between the two subjects, the following broad distinction will be used throughout this chapter. Photogrammetry will be considered to be concerned with the scientific methods of obtaining reliable measurements from ground or airborne imagery (primarily photographic) in order to produce a precise representation (graphical or digital) of the feature of interest. Remote sensing, in contrast, will be defined as encompassing those methods of detecting variations in radiant energy from the Earth's surface using airborne or satellite sensors. Interpretation (either visually or using computer techniques) of these recorded patterns is used to create thematic maps. The aim of this chapter will therefore be to: (1) provide an introduction to the fundamental principles of photogrammetry and remote sensing; (2) review the current state of development of instrumentation in both fields; and (3) examine the application of both techniques to problems in civil engineering.

7.2 Principles of photogrammetry

Since measurements may be taken from both air and ground images (normally photographs) two separate branches of the discipline are generally recognized: aerial and close range (or terrestrial) photogrammetry.

Aerial photogrammetry is a well-established technique in civil engineering for the production of topographic maps. Aerial photographs produced for such purposes can be obtained either with the optical axis of the camera pointing, nominally, vertically downwards so producing vertical aerial photographs or, with the axis intentionally tilted, to produce oblique aerial photographs. The former are almost exclusively used nowadays for photogrammetric purposes, although the latter have received a revival in recent years with the advent of analytical techniques. Vertical aerial photographs are normally acquired to achieve a 60% forward and a 20 to 25% lateral overlap in coverage. This enables a stereoscopic view of the terrain to be obtained and also a stereoscopic 'model' to be produced. The latter forms the basis of almost all of the photogrammetric instruments and techniques discussed in section 7.3.

Close-range, terrestrial or, to use a less attractive but commonly used alternative, nontopographic photogrammetry, is concerned with photographs taken on, or near the ground, rather than from an aerial platform. This branch of photogrammetry has developed considerably in the past 20 yr or so and has been applied to many problems in civil engineering. A selected number of these applications are presented in section 7.4.

A further distinction which is becoming used increasingly as a means of classification, both in aerial and close-range photogrammetry (CRP), is that between the analogue and analytical approaches to the subject. Apart from a few exceptions, until relatively recently photogrammetric instruments and techniques consisted of optical and mechanical analogue systems which were used for the production of a graphical end product (map, plan, elevation, section, etc.). However, the emphasis in photogrammetry is now moving towards analytical systems which use rigorous mathematical models to simulate the problem under investigation. By using these mathematical models in conjunction with modern computers, new photogrammetric instruments have evolved which offer significant advantages over their analogue counterparts in terms of speed of operation, accuracy and flexibility. Furthermore, the output products of these computer-based systems consist not only of graphical products, but also of other information sources such as digital terrain models.

7.2.1 Geometry of a single photograph

A photograph differs fundamentally in geometric terms from a map (unless the terrain is flat and the photograph vertical). Image displacement causes scale variations over the format due to the alteration in the position of points, compared with their corresponding map positions, as a result of the effects of ground relief and tilt of the photograph at the moment of exposure. The geometrical influence of both factors is illustrated by Figure 7.1 which shows that a tall tower is not imaged as a single point, as it would be on a map, but rather as a displaced line t–t and t'–t' on the vertical and tilted photographs respectively. In this case, the displacement is greatest on the vertical photograph. A similar displacement will exist for all points above the datum level e.g. point X. It is also evident that the tilted and vertical photographs are coincident at point i, the isocentre, and that tilt displacement radiates from this point. This radial nature of the displacements can be used to devise a simple method of plotting from a pair of photographs.

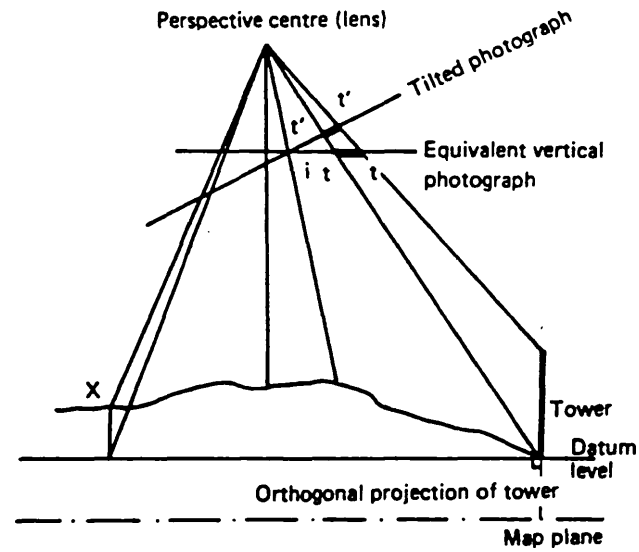


Figure 7.1 Relief and tilt image displacements on a vertical and near-vertical aerial photograph

The effect of image displacement on scale can be seen in Figures 7.2 and 7.3 which illustrate the separate results of tilt and variation in ground relief. It can be seen that, unlike a map which has a constant scale, the combined effects of relief and tilt produce a photograph which will be of constantly changing

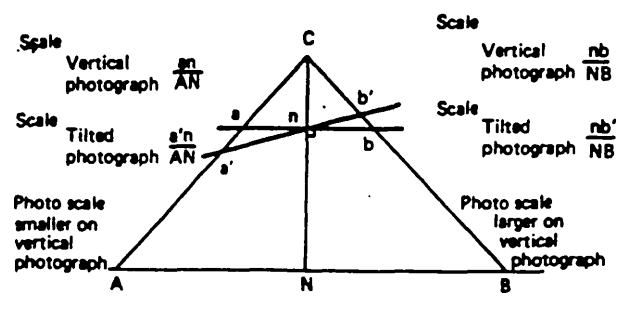


Figure 7.2 Scale changes on a vertical aerial photograph caused by variations in the elevation of the ground

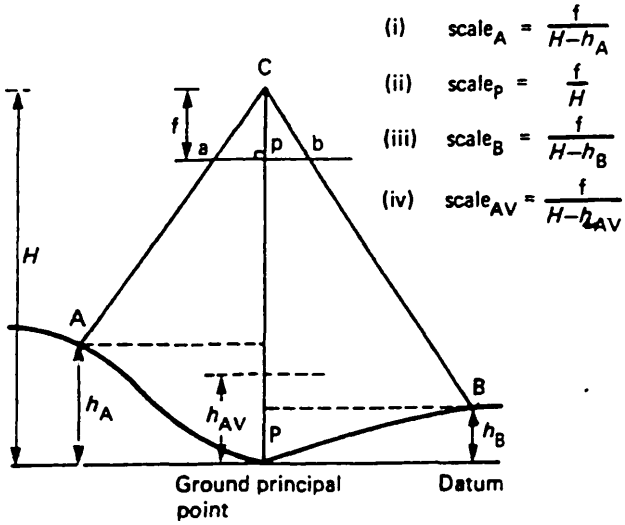


Figure 7.3 Scale changes on a tilted and vertical aerial photograph

scale, although an average nominal value for the scale can be calculated (often referred to as the contact scale).

7.2.2 Stereoscopy and parallax

Stereoscopic viewing and the measurement of the perceived stereomodel is fundamental to photogrammetry and thus of great importance. If two photographs taken from different viewpoints of the same area are viewed simultaneously, the difference in position of a common image point on the two photographs results in a discrepancy in image coordinates and this leads to the concept of x parallax and y parallax.

All points appearing on successive overlapping photographs exhibit x parallax in the direction of flight and this is the principal reason why a stereomodel exists. It can be seen from Figure 7.4 that the combined movement of point A across the focal plane of both photographs relative to the central axis is greater than for point B, thus point A is defined as exhibiting a greater absolute x parallax than B. It is also evident from the diagram that the absolute x parallax of a point is related to the height of that point and that it increases with increasing height. This simple concept forms the basis of the methods used to derive heights from aerial photographs. Section 7.3.3 outlines one such technique.

In contrast, y parallax is not related to height variations. It is produced if there is a difference in orientation of the two photographs at the time of exposure. The effect of y parallax on a stereomodel can be eliminated by reproducing the original camera orientations to ensure that corresponding rays from the two photographs will intersect. Such a process is termed the 'relative orientation' phase in setting up a stereoplotter.

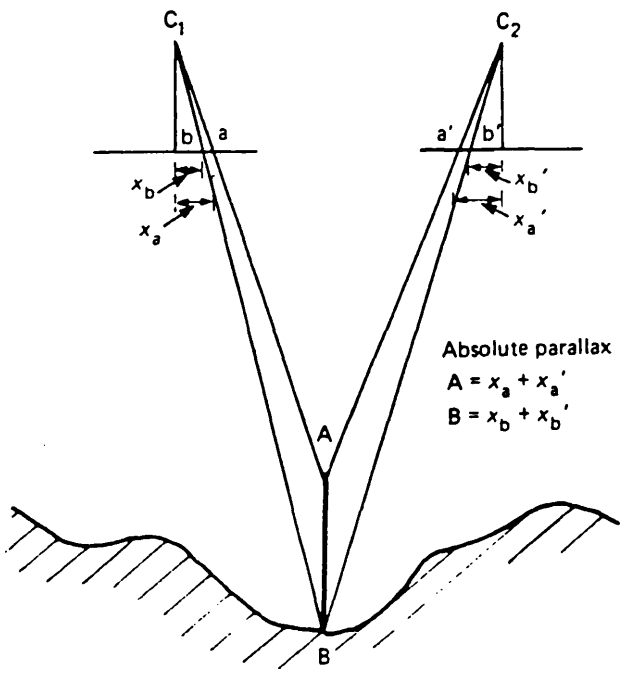


Figure 7.4 Variations in x-parallax on an overlapping pair of aerial photographs

7.2.3 Analytical photogrammetry

Many photogrammetric operations which were previously carried out by graphical or analogue methods are now being performed digitally using analytical instruments and computational techniques. Since most analytical methods involve some form of coordinate transformation, it is first necessary to examine the coordinate reference systems which are used. On the plane of the photograph, the reference system known as the 'image space' system is defined with respect to the principal point and the fiducial marks, as shown in Figure 7.5. It may also be important to take into account any variation in position which exists between the fiducial centre (the intersection of the fiducial marks) and the principal point (where the perpendicular line from the lens intersects the focal plane). This is defined by a camera calibration procedure. On the ground, positions are usually referred to a cartesian coordinate system, e.g. the National Grid. This is illustrated by Figure 7.6 in which X_c, Y_c and Z_c represent the 'object space' coordinates of the camera exposure station.

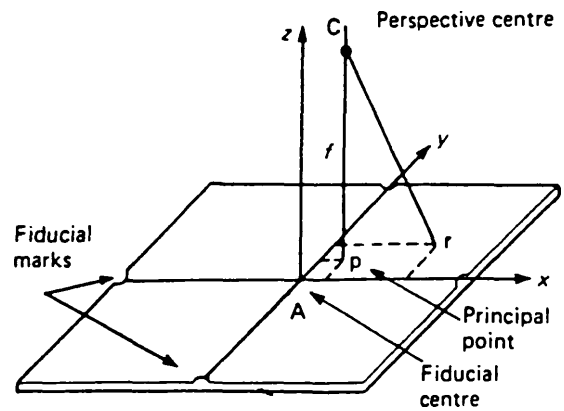


Figure 7.5 Image space coordinate reference system

An appreciation of the two basic assumptions which are used in order to define positions is also essential. Firstly, the line in space joining a point in the object space - the perspective centre of the camera lens and the position of the point on the photograph - is assumed to be a straight line, i.e. line p–C in Figure 7.6. In analogue stereoplotters, this assumption is imposed by optical projection or mechanical 'space rods'.

Analytically, it is expressed by means of the collinearity equations for a *single* photograph

$$x_c = -f\frac{M_{11}(X_c - X) + M_{12}(Y_c - Y) + M_{13}(Z_c - Z}{M_{31}(X_c - X + M_{32}(Y_c - Y + M_{33}(Z_c - Z} \qquad (7.1)$$

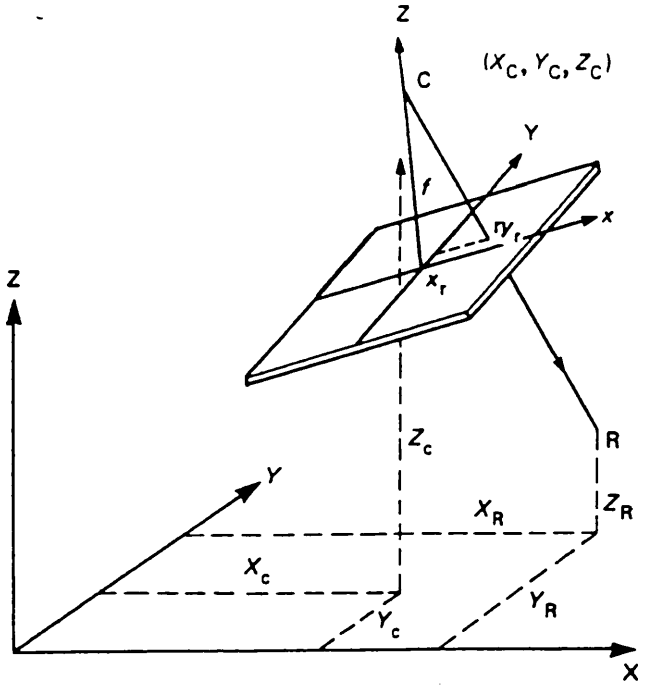


Figure 7.6 Object space coordinate reference system and collinearity condition

$$y_r = -f\frac{M_{21}(X_r - X_c) + M_{22}(Y_r - Y_c) + M_{23}(Z_r - Z_c)}{M_{31}(X_r - X_c) + M_{32}(Y_r - Y_c) + M_{33}(Z_r - Z_c)} \qquad (7.2)$$

where $M_{11} \ldots M_{33}$ represent terms of an orthogonal rotation matrix and the other terms are as defined in Figure 7.6. The collinearity equations are used extensively in analytical photogrammetry to relate image coordinates, ground coordinates and the coordinates of the perspective centre. The technique may then be used to derive the coordinates of a series of unknown positions.

A further analytical technique which is used by photogrammetrists is that defined by the coplanarity equations. Coplanarity equations can be established for a pair of photographs and they may be used to reconstruct the orientation of the cameras at the moment of exposure, i.e. an analytical method of relative orientation. In order to satisfy the condition, the two exposure stations C_1 and C_2, the object point R and its corresponding image points r_1 and r_2 are assumed to all lie in a common plane, known as the epipolar plane (Figure 7.7).

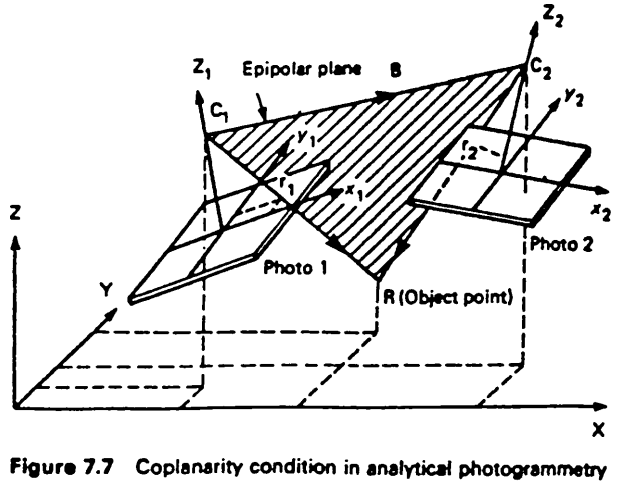


Figure 7.7 Coplanarity condition in analytical photogrammetry

7.3 Photogrammetric instrumentation

7.3.1 Cameras

Most photogrammetric measurements are recorded from photography produced by a high-quality camera specifically designed for photogrammetric purposes. The main feature which distinguishes this type of camera from others is that the internal geometrical characteristics are known precisely. Thus, data relating to the lens distortion, focal length and principal point location are established, and monitored periodically by a camera calibration procedure.

Both aerial and close-range cameras are used for data acquisition. Aerial mapping cameras are normally classified on the basis of the angular field of view of the lens, a parameter which relates directly to both the focal length (f) and the format size of the camera. The focal length of an aerial camera is defined as the distance from the optical centre of the lens to the image plane. For the general case of a 230 × 230 mm format, the classification shown in Table 7.1 can be produced. Normal angle (NA) photography is only occasionally used and it can be seen from Figure 7.8 that the principal advantage of the super wide

Table 7.1 Classification of aerial mapping cameras

Camera type	*Approximate angular field of view* (°)	*Approximate focal length* (mm)	*Examples*
Super wide angle (SWA)	120	90	Zeiss (Ober.) RMK A 8.5/23 Wild RC-10 (Super Aviogon lens) Zeiss (Jena) MRB 9/2323
Wide angle (WA)	90	150	Zeiss (Ober.) RMK A 15/23 Wild RC-8 Wild RC-10 (Universal Aviogon lens) Zeiss (Jena) MRB 15/2323
Normal angle (NA)	60	300	Zeiss (Ober.) RMK A 30/23 Wild RC-10 (Normal Aviogon lens)

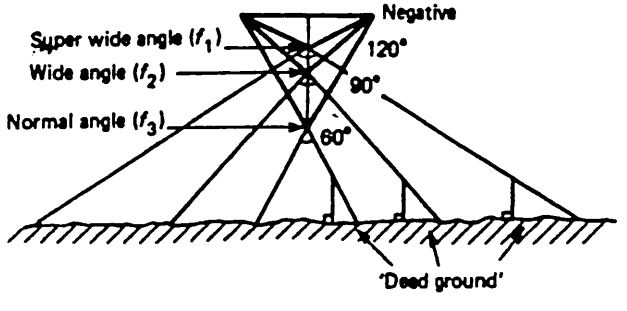


Figure 7.8 Variation in ground coverage of normal, wide and superwide angle cameras

angle (SWA) lens is its greater ground coverage for a given flying height. This is particularly important since it not only reduces the number of photographs required to cover an area, but also reduces the number of control points which are required for controlling the mapping. However, because SWA photography is more susceptible to the production of 'dead ground' (loss of detail because of the screening effect of features), a compromise solution is normally adopted by using a camera which enables wide angle (WA) photography to be produced.

Three distinct types of close range or terrestrial photogrammetric camera are currently available: (1) metric; (2) stereometric; and (3) nonmetric. The metric camera models include those in which a camera and a theodolite form one integrated unit, the classical phototheodolite, and those where the camera is a separate unit from the theodolite. The latter case tends to be more common and the UMK10/1318 is a typical example (Figure 7.9). Details of the characteristics of several cameras of this type are listed in Table 7.2.

The second design, the stereometric system, consists of two matched camera units mounted on each end of a base bar, typically 0.4 to 3 m long. The camera axes are set at right angles to the bar and since the relative position and orientation of the cameras at the instant of exposure are known, the subsequent analysis phase is considerably simplified. The Zeiss (Oberkochen) range of cameras illustrated by Figure 7.10 are representative of this design. Further details of other cameras of this type are listed in Table 7.3.

A nonmetric camera is one which was not designed specifically for photogrammetry. Although such cameras offer advantages in terms of size and cost, their use is limited, unless analytical techniques are used to compensate for the unstable interior geometry of the camera. A varied selection of nonmetric cameras has been used for photogrammetric purposes including



Figure 7.9 Zeiss (Jena) UMK 10 1318 close-range camera

Table 7.2 Summary of main features of a selection of close-range photogrammetric cameras

Manufacturer	*Model*	*Nominal focal length* (mm)	*Principal distance* (F = fixed, V = variable)	*Format* (mm)	*Minimum range* (m)
Wild	P30	165	F	100 × 150	20
Wild	P31	45, 100, 200	F, V, V	100 × 150	7, 1.4, 8
Wild	P32	64	F	65 × 90	3.3
Officine Galileo	Verostat	100	V	90 × 120	2
Zeiss (Jena)	UMK 10/1318	100	V	130 × 180	1.4
Zeiss (Ober.)	TMK 6	60	V	90 × 120	5

(Adapted from Atkinson (1978) 'Photography in non-topographic photogrammetry'. In: Newman (ed.) *Photographic techniques in scientific research*, vol. III. Academic Press, London)

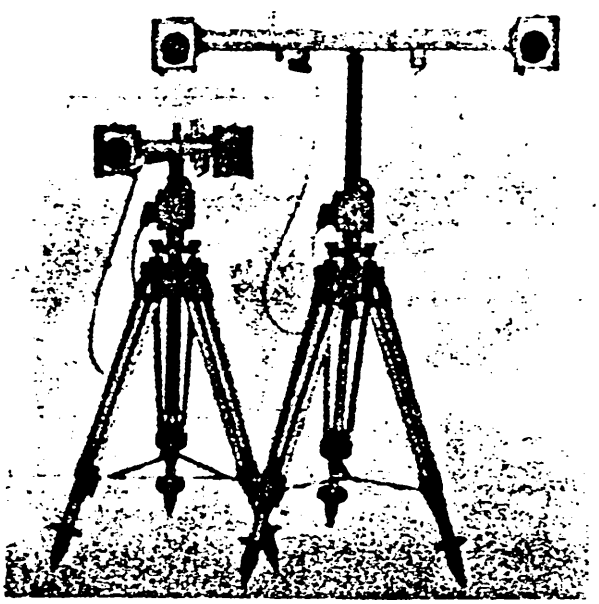


Figure 7.10 Zeiss (Oberkochen) SMK 120 and SMK 40 stereometric camera

medium-priced 'amateur' cameras, such as the Olympus OM-2, Canon TX, Minolta XG-7 and Rolleiflex SL66, as well as the more expensive 'professional' cameras such as the Linhof Technica and the Hasselblad 500EL.

7.3.2 Single-photograph-based instruments

In certain circumstances it may be appropriate to attempt to derive metric plan information from a single photograph. Reasonable results can be obtained provided that relatively flat ground has been imaged on an almost tilt-free photograph.

The camera lucida principle, whereby images of an existing map and the corresponding photograph are superimposed, has been adopted to produce a simple mapping instrument such as the Zeiss (Jena) Sketchmaster and the Bausch and Laumb Zoom Transfer Scope. Adjustments can be made either mechanically or optically in order to remove the effect of minor tilt distortions and to make a correction for small variations in scale between map and photograph. This type of instrument is of particular value for map revision. When detail shown on both photograph and map has been aligned, new details imaged on the photograph can be traced on to the map.

It is also possible to apply a digital approach to measurement on a single photograph using a monocomparator. This instrument essentially consists of a travelling microscope moving

Table 7.3 Summary of main features of stereometric close-range photogrammetric cameras

Manufacturer	*Model*	*Nominal principal distance* (mm)	*Principal distance* (F = fixed, V = variable)	*Format* (mm)	*Focusing range*
Wild	C40	64	F	65 × 90	1.5–7.0
Wild	C120	64	F	65 × 90	2.7–∞
Officine Galileo	Veroplast	100	V	90 × 120	2–∞
Zeiss (Jena)	SMK5.5/0808	56	F	80 × 80	5–∞
Zeiss (Ober.)	SMK120	60	F	90 × 120	5–∞

(Adapted from Atkinson (1978) 'Photography in non-topographic photogrammetry'. In: Newman (ed.) *Photographic techniques in scientific research*, vol. III. Academic Press, London)

along orthogonal axes in order to determine image coordinates. Various refinements to the basic principle have been made; the Surveying and Scientific Instruments PI-1a, for example, employs a miniature CCTV observation system and provides digital readout with the MDR 1S/3 display, whilst the Zeiss PSK-1 instrument uses optical scales based on Moiré fringes. With the aid of a microcomputer and suitable software, a plot can be produced from such equipment.

7.3.3 Approximate solution stereoscopic instruments

When reconnaissance mapping or revision of existing maps from new photography is being carried out, it is often more cost effective to use a simple, approximate stereoscopic instrument rather than a sophisticated and expensive rigorous stereoplotter.

Several manufacturers produce stereoscopic versions of the monoscopic instruments which are based on the camera lucida principle, e.g. the Cartographic Engineering Stereosketch and the Bausch and Laumb Stereo Zoom Transfer Scope. The image of the stereoscopic view of an overlap is superimposed on the map image and photographic plan detail can be traced; no facilities are provided for height measurement. The Cartographic Engineering Radial Line Plotter (Figure 7.11) is a simple, lightweight instrument which can be used for small-scale planimetric mapping. The instrument design is based on the radial line assumption which states that angles measured at the principal point are true if the photographic tilt and variation in ground relief are below some limiting value (normally 3° and 10% of the flying height). Other approximate instruments include the Zeiss (Oberkochen) G2 Stereocord and the Officine Galileo Stereomicrometer.

A parallax bar (Figure 7.12) is a portable device for obtaining measurements of differences in x parallax which can be used to calculate differences in height. Spot heights at changes of slope can hence be obtained and contours interpolated. The eye observes the measuring mark on the glass plates through a stereoscope. When adjusted to the correct x separation the two marks appear to fuse into a single point which may appear to 'float' above the level of the terrain. In order to make some correction for possible photographic tilt and variation in flying height for the stereopair, a minimum of five control points of known ground height are required for the overlap; ideally eight to ten points will be used. Each of the control points is observed in turn and the micrometer screw is turned to remove x parallax and, hence, bring the floating mark to ground level.

Any y parallax can be eliminated by adjusting the right-hand stage plate in the y direction. The micrometer readings are recorded for all the control points. In order to calculate spot heights, the mean flying height must be determined from scaled measurements on the photographs and the absolute parallax of one point must be determined using a travelling microscope. A computer program is then normally used to calculate crude

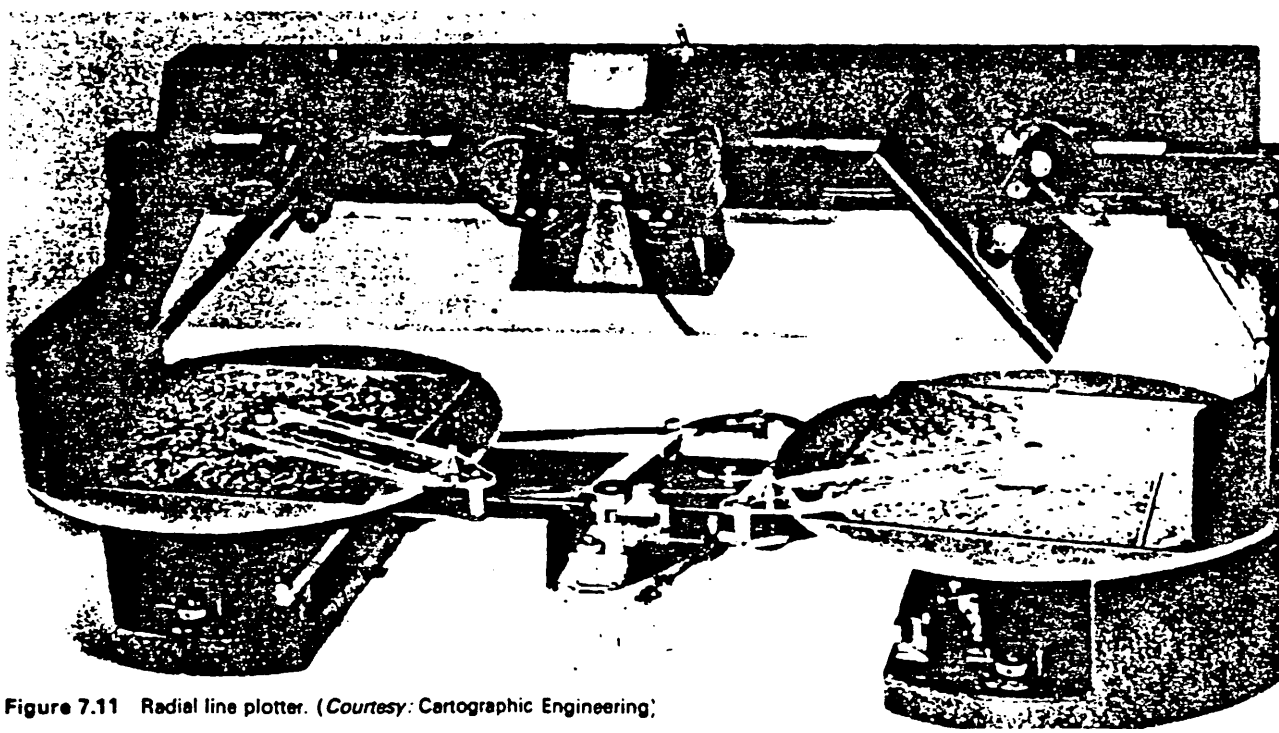


Figure 7.11 Radial line plotter. (*Courtesy*: Cartographic Engineering)

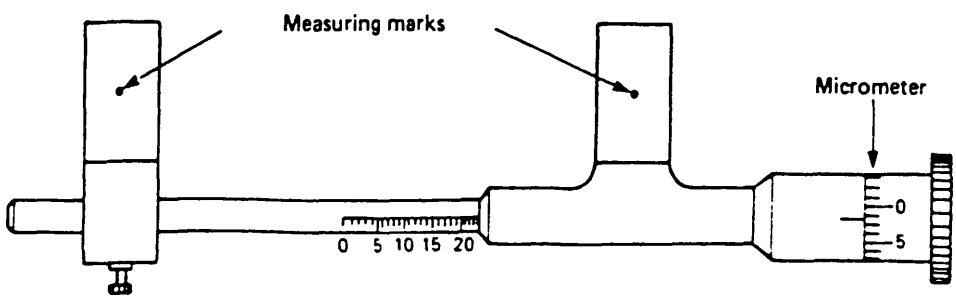


Figure 7.12 Parallax bar

heights from Equation (7.3) and to obtain corrected spot heights from Equation (7.4). If redundant control has been provided, residuals will be available and the heighting accuracy can be investigated.

$$h_B = h_A - \frac{(H - h_A)\Delta_{p_{AB}}}{p_A - \Delta_{p_{AB}}} \tag{7.3}$$

where h_A = height of A (a reference point of known height), h_B = crude height of point B (point of required height), $\Delta_{p_{AB}}$ = difference in parallax bar readings between A and B (care must be taken with sign) and p_A = parallax of A.

$$h_{B'} = h_B + a_0 + a_1 x + a_2 y + a_3 xy + a_4 x^2 \tag{7.4}$$

where $h_{B'}$ = corrected height of point B and x, y are image coordinates of B, and $a_0 \ldots a_4$ are coefficients, constant for a given overlap. This method of height correction was first proposed by Thompson[1] and was subsequently modified by Methley.[2] Contours may be interpolated between spot heights, continuous reference being made to the stereomodel as they are drawn.

7.3.4 Rigorous solution stereoscopic instruments

7.3.4.1 Analogue stereoplotters

Analogue stereoplotters create and measure an exact three-dimensional model of the ground. Direct optical projection in the form of Multiplex equipment was the earliest method of realizing this objective. The Zeiss (Ober.) DP 2 stereoplotter (Figure 7.13) is a modern version of Multiplex equipment. Plotters such as these, which employ direct optical projection, have the disadvantage of relatively poor viewing conditions and an inflexible plotting scale, usually approximately double the photograph scale. However, their uncomplicated operation and minimum of maintenance make them very suitable for medium-scale mapping and map revision, although in recent years they have tended to be superseded by mechanical and analytical instruments.

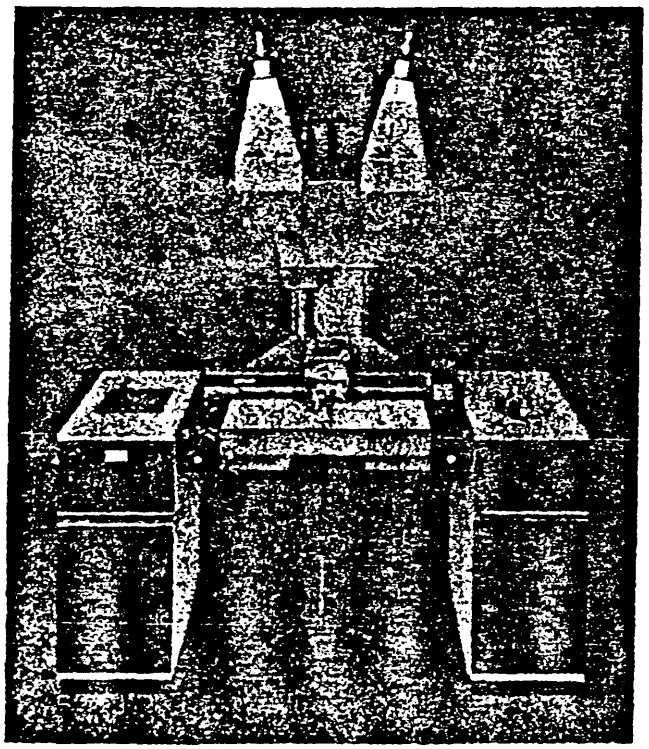


Figure 7.13 Zeiss (Oberkochen) DP-2 stereoplotter. (*Courtesy*: Zeiss (Oberkochen))

Mechanical rods replace light rays when a rigorous model is formed in instruments employing mechanical projection (Figure 7.14). Full-sized diapositives are placed in projector heads which can be tilted; a stereoview is obtained through a binocular eyepiece linked to viewing microscopes fitted to each projector. The movement of each microscope is communicated via a tie rod to a sleeve at the top of a space rod which pivots about a gimbal joint representing the projection centre. The two space rods take up the attitude of the space rays from the ground point to the camera stations; they normally intersect in a model point which is connected to a plotting pencil by gears. The principal distance of the taking camera is represented by the separation of the sleeve on the space rod which represents the image point and the gimbal joint; this distance can easily be varied so that mechanical projection plotters can accommodate photography taken with a wide variety of cameras. Considerable flexibility of plotting scales (up to 6 times the photographic scale) is provided by altering the gears linking the model point to the plotting point (Table 7.4).

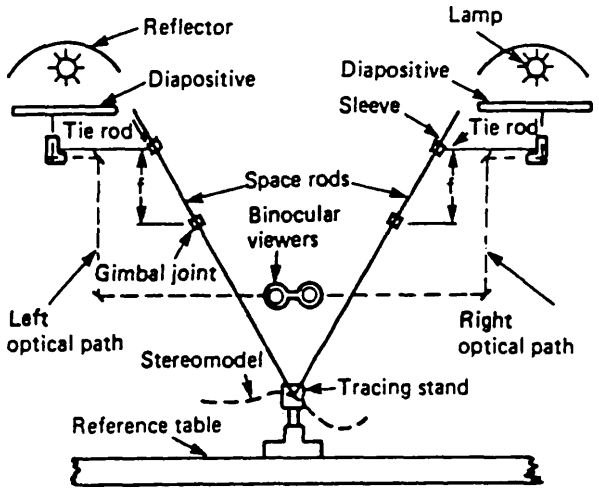


Figure 7.14 Mechanical projection stereoplotter

Table 7.4

Photographic scale	*Flying height* (m)	*Enlargement factor*	*Mapping scale*
1:3000	450	6	1:500
1:5000	750	5	1:1000
1:10 000	1500	4	1:2500
1:25 000	3750	2.5	1:10 000

Digital readout of model coordinates is now standard practice and recent developments include computer-assisted prompts, to guide the orientation procedures, and computer-aided drawing. The Wild A10 stereoplotter (Figure 7.15) is a typical example of a mechanical instrument.

7.3.4.2 Analytical stereoplotters

With the advent of digital computers, especially mini- and microcomputers together with electronic encoders for digitizing the output from stereoplotting equipment, a new generation of

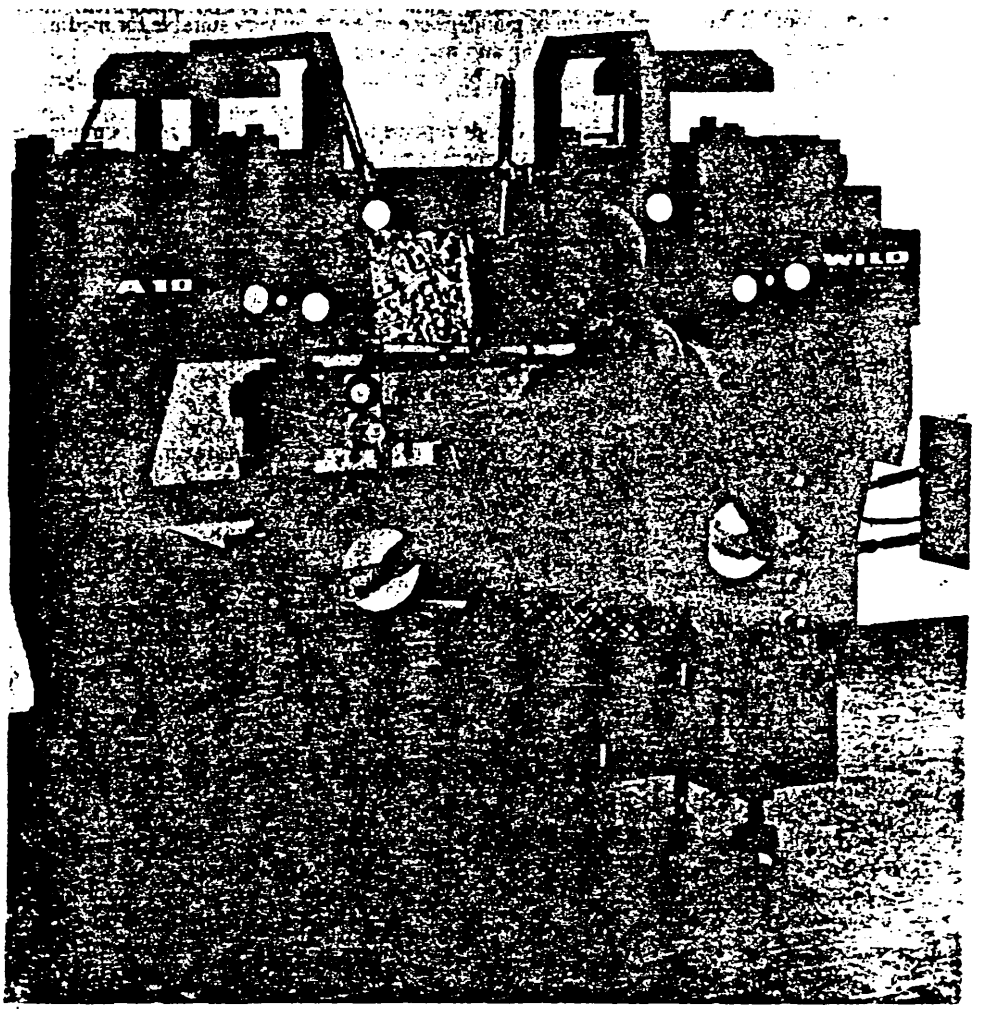


Figure 7.15 Wild A-10 stereoplotter. (*Courtesy*: Wild Heerbrug)

photogrammetric stereoplotter has emerged: the analytical stereoplotter. With this type of instrument a mathematical model replaces the space rods and other mechanical linkages which are found in more conventional analogue stereoplotters. An instrument which is representative of this design is the Officine Galileo Digicart. This instrument consists of a stereoviewing system, linear encoders to digitally record the position of the carriages on which the photographs are mounted, a minicomputer to process the data, and a plotting table. Unlike a conventional analogue instrument where manual elimination of y-parallax is performed initially over the entire model, with this design any parallax existing within the model is continuously eliminated by a feedback loop which ensures that corrections are transmitted to the optics in real time (50 times a second) to provide a model free from parallax. Figure 7.16 illustrates the principle of operation.

The advent of analytical techniques has had a considerable influence on the practical application of photogrammetry to engineering problems. In particular, analytical techniques offer the following advantages over conventional analogue techniques:

(1) Higher accuracy of measurement can be achieved than was previously possible (up to $\pm 3\,\mu m$ at photoscale with suitable care).
(2) Photography taken with differing format sizes and focal lengths can be used. In certain cases, provided suitable

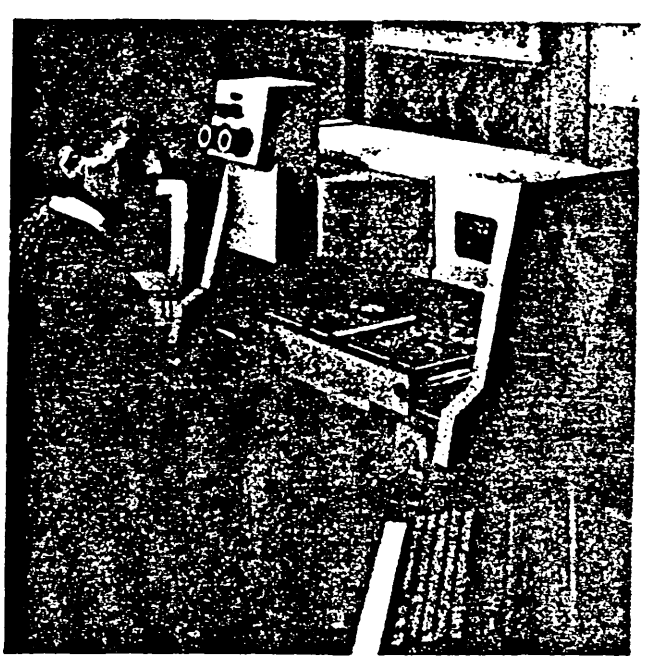


Figure 7.16 Officine Galileo Digicart analytical stereoplotter (*Courtesy* Officine Galileo)

precautions are taken to eliminate any systematic errors such as lens distortion, nonmetric cameras can be used.

(3) The setting up of any stereomodel presents little difficulty; most models can be set up in 10 to 15 min by an experienced operator, thus leaving more time for the actual measuring phase.

(4) Photography can be obtained in the field without too great a concern about the orientation of the camera. This can save a great deal of time (as much as 50% over the conventional approach). Where access is difficult, and vertical or normal case geometry is impeded, oblique photography may be taken. Such an approach would have been precluded using traditional methods.

(5) Data are captured in ground coordinates and stored on magnetic disks. This allows for plotting off-line on graphical plotters at any scale. Also, with data being collected digitally rather than graphically, much more flexibility exists in data handling, e.g. the data can be manipulated on a Computer Aided Drafting and Design (CADD) system.

7.3.4.3 Orientation of stereoplotters

Whatever type of equipment is used, before drawing can commence three stages of orientation must be carried out in order to set up the model correctly.

Inner orientation consists of the careful setting of the diapositives on glass stage plates. For analogue instruments, either the principal point or the fiducial marks on both diapositive and register glass must be aligned. Unless the principal distance is fixed (as is the case for direct optical projection instruments), this must be set to the value obtained from the camera calibration certificate. If an analytical plotter is used, the reference marker is taken to each fiducial mark in turn and the coordinates are recorded; the camera principal distance is entered into the computer as are any known lens distortion characteristics.

Relative orientation is next carried out in order to produce a true undistorted model. It does not require any reference to ground control and is achieved in analogue instruments by removing y parallax for clear points of detail selected at five standard positions. The entire model should then be clear of y parallax. For analytical models, y parallax is measured at five or more positions to enable the computer to deduce the movement of the servomotors subsequently required to view a parallax-free model.

Absolute orientation is required in order to scale the model and level it to ground datum. Scaling is basically accomplished by comparing a model distance with a known ground distance and adjusting the machine base in analogue instruments to achieve the required model scale. For heighting, the model and ground heights of a minimum of three points are compared and the model is rotated about the machine X and Y axes until differences in height agree. The machine datum will then be parallel to ground datum and its exact level can be determined. True ground heights can then be read directly from the instrument height scale. Analytically, photo coordinates and corresponding ground coordinates of three or more control points are compared in order to enable the required transformation coefficients to be computed.

The machine is now set up and plotting can commence. For plan detail, the operator moves the floating mark to follow image detail, keeping it in contact with the model surface by constantly adjusting its height. In correspondence, the plotting pencil traces out the detail on the map sheet. In order to plot a contour, the floating mark is kept at the required constant height and is moved in contact with the model surface so that the contour line is drawn on the plot. A similar procedure is followed for the analytical plotter as far as the movement of the floating mark by the operator is concerned, but computer-assisted plotting offers additional options such as defining the end points of a straight line and instructing the pencil to join them.

The machine plot will only contain information that can be obtained directly from the stereopair and field completion is required to add ground detail which might be obscured on the photographs, e.g. within woodland or below the eaves of houses, and also to assist annotation of, for example, place names and road classifications.

7.3.4.4 Orthophoto systems

An orthoprojector is used to generate a photograph that is the geometrical equivalent of a map by a process of differential rectification. An exact model is set up in a stereoplotter, the effects of tilt being removed by the standard procedures of relative and absolute orientation. The operator scans the model systematically with the floating mark, adjusting the effective height of the mark continuously so that it remains on the surface of the ground. The orthoprojector operates in the darkroom conditions and holds a duplicate of one of the photographs of the stereopair. The tilt of the corresponding plotter projector, and the linear motion and height movement of the floating mark are transmitted either mechanically (Zeiss (Oberkochen) DP Ortho-3), optically (Wild PPO-8) or analytically (Zeiss (Oberkochen) Orthocomp Z-2) to the orthoprojector which allows exposure through a small slit on to light-sensitive paper. Thus a corrected print of the overlap area, known as an orthophotograph, is gradually built up as the slit moves in sympathy with the floating mark, the effects of height difference being removed by the continuous adjustment of the orthoprojector. Cartographic enhancement may be added to the print (e.g. contour lines, grid lines and place names), resulting in an orthophotomap.

7.4 The use of photogrammetry in civil engineering

7.4.1 Aerial photogrammetry

7.4.1.1 Topographic mapping

The use of aerial survey is generally considered to be the standard method of producing a topographic map or plan at scales smaller than 1:500. For scales greater than 1:500, ground survey would almost invariably be used.

The basic sequence of operations required for the production of a topographic map is shown by the flow diagram in Figure 7.17. The major air survey companies will have the equipment and manpower to carry out all the stages indicated but smaller concerns might, for example, subcontract the photography. It is very important that the exact requirements for the survey output should be known at the outset and that detailed discussion should take place between the civil engineer and the air survey company in order to identify any specific problems which might arise in connection with a particular project.

Flight planning. Before a survey flight can commence, a flight plan must be prepared detailing all the information required by the pilot, navigator and camera operator.

Aerial photographs which are to be used for mapping purposes must be taken in a regular sequence along parallel lines of flight. Navigation is usually visual by identifying landmarks shown on the map, mosaic or satellite image on which the

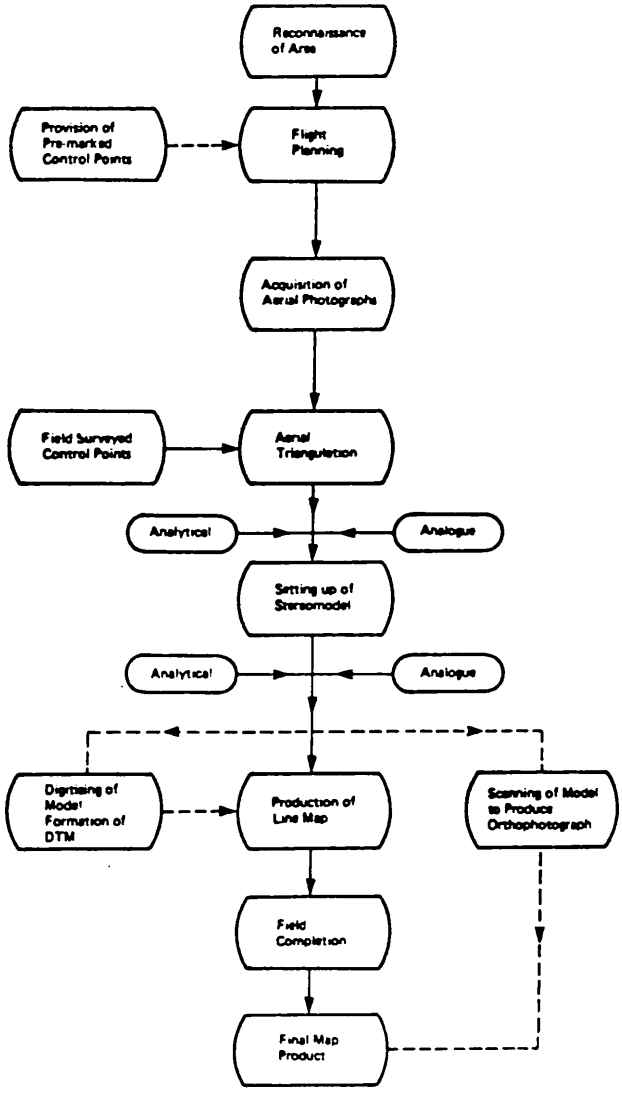


Figure 7.17 Flow diagram illustrating the sequence of photogrammetric operations involved in the production of a topographic map

required lines of flight have been marked. Increasingly, reliance may be placed on some form of radio navigation system. Flight lines will normally be drawn parallel to the longes edge of the area to be covered. However, if there is any marked grain of relief, this direction may be taken for the flight lines in order to attempt to maintain a constant scale along a given strip. Each line of flight produces a *strip* of photography, successive overlapping strips forming a *block* of photography.

Since a stereoscopic view of the entire area is needed, each ground point must be photographed from two camera stations giving a minimum overlap stipulation of 50% in the direction of the line of flight (the fore and aft direction). However, in order to give additional overlap to satisfy aerial triangulation requirements and to provide a safety margin, a value of 60% is usually specified.

A standard *Specification for vertical aerial photography* has been prepared by the Royal Institution of Chartered Surveyors' in collaboration with the British Air Survey Association. This specification is used by air survey companies throughout the world as a contract document for providing black-and-white aerial photography. The specification includes the air camera and its calibration, photographic coverage and operation, flying conditions, the type of aerial film, processing and drying conditions, the resultant products to be supplied, and documentation and annotation.

Aerial triangulation. The object of aerial triangulation is to establish a network of control points with known ground coordinates which are required before a map can be plotted from aerial photographs. In order to perform absolute orientation when setting up a model in a stereoplotter, the minimum theoretical control required for each model is 2 points with known plan coordinates and 3 points with known heights (in practice, 3 plan points and 4 height points are usually considered to be the minimum). Thus, a large project would require the provision of an extensive amount of control; it would be very expensive and time-consuming if this had to be carried out exclusively by field survey methods. Aerial triangulation also has the advantages that it is easy to provide more than the bare minimum of control and it will usually be possible to obtain points in very convenient positions for their subsequent use in a stereoplotter.

Analogue methods of aerial triangulation have evolved steadily with the development of analogue stereoplotters. The method of aerial triangulation by independent models (AIM) is now used almost universally with modern analogue instruments. Since the technique requires model coordinates to be measured in an analogue plotter, a substantial amount of computation is required to obtain ground coordinates and it could therefore be described as a semi-analytical method.[4]

Whereas model coordinates measured in an analogue plotter form the basic data for the calculation of aerial triangulation by independent models, x, y coordinates measured either in a mono- or stereocomparator provide the raw data for the fully analytical methods. Much less observation time is required than with the equivalent analogue methods, but much greater emphasis is, however, placed on the data processing phase, both for the location of gross and systematic errors and for the computation of the coordinates. As outlined in section 7.2.3, it is possible to establish the mathematical equations of the bundle of rays from the optical centre of the camera lens to all image points observed on a given photograph by using the collinearity method. The combination of all the bundles of rays for the entire block can be used to carry out relative and absolute orientation simultaneously.[5] The calculated ground coordinates obtained by this method are adjusted to a best mean fit by least squares. It is clearly advantageous if both the necessary hardware and software are available to consider the block as an entity rather than to adjust a strip and then connect adjacent strips in a separate operation.

Having carried out aerial triangulation, each stereomodel is subsequently set up in turn using the orientation procedures mentioned in section 7.3.4.3. Plotting can then commence, so creating a topographic map or the operator may measure spot heights to form a digital terrain model.

7.4.1.2 Production of digital terrain models

The term digital terrain model (DTM) was originally coined in 1958 by Miller and La Flamme. Since then, the subject has developed considerably and is currently an area of widespread activity in surveying, geology and geophysics, civil and mining engineering and other disciplines in the earth sciences.

Although a variety of different techniques exist for the creation of DTMs in general, the distinction can be made between those which make use of height data which have been collected in a regular grid and those based on a network of randomly located height points. The former, simpler, approach is generally adopted when photogrammetric equipment is being used for the generation of the DTM. While the square grid is the

most common form, rectangular-, hexagonal- and triangular-based grids are also used.

Although the regular grid is the simplest technique to adopt it has one serious limitation: the distribution of data points is not related to the characteristics of the terrain. If the data point sampling is conducted on the basis of a regular grid, then the density must be high enough to portray accurately the smallest terrain features present in the area being modelled. If this is done, then the density of data collected will be too high in most areas of the model, in which case there will be an embarrassing and unnecessary data redundancy in these areas.

One solution to this problem is to use progressive sampling and its development, composite sampling.[6] Instead of all points in a dense grid being measured, the density of the sampling is varied in different parts of the grid, being matched to the local roughness of the terrain surface. This approach has been widely implemented on a variety of grid-based terrrain modelling packages such as HIFI,[7] and SCOP.[8]

Terrain models derived by photogrammetric techniques are also used by modelling packages such as MOSS, ECLIPSE, and Intergraph DTMN. A comprehensive review of the use of such terrain modelling packages in surveying and civil engineering can be found in Petrie and Kennie.[9]

7.4.1.3 Monitoring

Aerial photogrammetry has been used extensively in the past as a means of monitoring changing conditions such as the depletion of coal stocks or the degree of coastal erosion occurring over a period of years. In more recent years, its use has also developed for monitoring the deformation of natural or man-made features.

Fraser[10] and Fraser and Gruendig[11] describe the planning and subsequent execution of a photogrammetric survey of the Frank landslide/rockslide region of Turtle Mountain in Alberta, Canada. A rockslide occurred in this area over 80 yr ago when more than 30 million m^3 of rock moved down the east face of the mountain covering an area 3 km square to a depth of 14 m. Since 1933, crack monitoring surveys of the area have been undertaken to determine the stability of the rock wedge forming the southern peak of the mountain. These surveys have been supplemented in recent years by an *in situ* monitoring program using crack motion detectors. Whilst these devices are capable of very high accuracy (±0.1 mm/yr), they can only, realistically, be deployed over a very small area. Also, if hazardous movements are detected it may be considered unsafe to continue recording measurements. Aerial photogrammetry, by covering large areas in a noncontact manner, can overcome these limitations. Therefore, it was considered appropriate to evaluate the potential of photogrammetry for this project. Fraser[10] established that the optimum flight plan consisted of obtaining large-scale (1:2000) multiple photographs (10 to 15) of the area containing the thirty monitoring target points. Photography was obtained using a Wild RC8 camera (focal length 152 mm) on two occasions approximately 1 yr apart. A subsequent analysis of the results indicated that statistically significant deformations had occurred at eight of the thirty points, the maximum movement being 63 mm.

7.4.2 Close-range photogrammetry (CRP)

Whereas it is possible to identify a standard procedure for mapping from aerial photographs, each project which requires the use of close-range photographs will tend to be regarded as unique due to the wide variation in circumstances which can arise for both taking the photographs and the subsequent analysis. For example, the positioning of the cameras for photography needs to be adapted to the particular demands of an individual task and may vary from a simulated aerial case with parallel, near vertical axes, through inclined and convergent axes to parallel horizontal axes. At the measurement stage, use of either analogue or analytical equipment may be specified according to the type of camera and whether a graphical or a numerical output is demanded.

Although some air survey companies also have a department which specializes in CRP, as do a few land survey organizations and universities, much of this type of work could be carried out by engineering firms themselves, perhaps within a research and development department as, for example, at Rolls Royce.[12] It is even more vital for CRP than for aerial photogrammetry that the engineer should be very closely associated with all stages of the work so that he can ensure that his exact requirements are met.

7.4.2.1 Monitoring

Deformation of structures. Close-range photogrammetry has been used widely in the field of structural engineering. The ability of the technique to record both the shape of a structure at one moment and the deformations which occur between two epochs have been exploited. Notable in the first case is the extensive work which has been carried out in architectural photogrammetry. In the latter case, comprehensive applications reviews have been written by Atkinson[13] and Cheffins and Chisholm.[14] A review of the various methods which can be adopted has been prepared by Cooper.[15] Cooper identifies four main methods including those which make use of a single camera (time parallax), controlled stereomodels, resection/intersection, and bundle adjustment. The latter is considered to be the most general case and most suitable for high-accuracy applications.

The range of large engineering structures which have been investigated by photogrammetric methods includes cooling towers, box girder bridges, elevation of St Paul's Cathedral, tower cranes, offshore platforms, and ships. In addition, photogrammetry has been used for the measurement and dimensional control of smaller structures such as microwave antennae, robots, and large compressors (see Welsh[16] for further details). Increasingly, the emphasis is being placed on the development of a 'real-time' photogrammetric system using video technology and solid-state cameras. Wong and Wei-Hsin[17] outline the development of such a system.

Earth and rockfill dam monitoring. The use of CRP for monitoring earth and rockfill dams has been discussed by Moore[18] and Brandenberger,[19] among others. In the case of Moore,[18] the author describes the use of a Wild P30 phototheodolite for monitoring the three-dimensional displacements of the Llyn Brianne rockfill dam, in mid Wales, during several stages of construction. The predicted displacements were in the range of ±0.5 to 1.0 m, and whilst high absolute positional accuracy was deemed to be important, it was considered that the definition of the direction of movement was of equal importance. By using a Wild A7 stereoplotter, measurements were taken at over 80 targets, at differing levels of fill. The results indicated a range of displacements, from 0.1 to 0.6 m (with a standard error of ±0.05 m).

Retaining wall monitoring. The use of a modified KA-2 aerial camera (focal length 610 mm, format 23 × 23 cm) to monitor the deflection of a gabion wall is described by Veress, Jackson and Hatsopoulos.[20] The wall, situated near Seattle, Washington State, was over 400 m long and varied in height from 2 to 18 m. The gabions were constructed of 1-m steel mesh rockfilled cubes. Photographs were obtained from fixed control points up

to 1000 m from the wall. Over 100 target points were observed with an analytical plotter and ground coordinates computed. The wall was measured on eleven occasions. The authors suggest that CRP used in conjunction with an internal monitoring system such as an inclinometer would be an ideal system for monitoring new structures.

7.4.2.2 Slope/rock stability studies

One of the earliest reports of rockface mapping is given in Cheffins and Rushton.[21] This paper describes the procedures used to produce 1:50-scale contoured elevations of the north face of Edinburgh Castle Rock. The elevations were required by a team of engineering geologists who were carrying out rock bolting and grouting as stabilizing measures to reduce the chance of rock falls from the face. The survey was carried out by taking four pairs of overlapping photographs from a baseline approximately parallel to the face. The photographs were obtained using a Wild RC5A aerial camera mounted on a hydraulic platform. The elevations with 0.25-m contours were subsequently used, in conjunction with ultrasonic data about the joint structure, to plan the positions of the necessary boreholes for rock bolting.

Moore[22] discusses the use of a phototheodolite for mapping vertical faces, in this case in clay pits. The problem under investigation involved the determination of information about the continuity and spacing of major joints in the clay face. Photogrammetric techniques enabled a three-dimensional model of the spatial disposition of the joints to be constructed. Other references which mention the use of photogrammetry for slope and rockface stability monitoring include Torlegard and Dauphin[23] and Robertson *et al.*[24]

Landslides are a commonly occurring hazard in road design and construction. Heath, Parlsey and Dowling[25] describe the use of a Wild P32 camera to obtain photographs of areas susceptible to landslipping in Columbia and Nepal. The photographs were observed using a Zeiss Topocart and Wild A40 Terrestrial Stereoplotter, and 1:200- to 1:10 000-scale contoured plots were produced. The plans were subsequently used to design remedial and preventative measures associated with the landslip areas. They also enabled a classification of landslide characteristics to be produced.

Kennie and McKay[26] discuss the use of CRP to monitor the erosion of chalk cliffs around a road tunnel in East Sussex. Control stations were monumented at both the top and base of the cliffs, ensuring that a minimum of six control points were visible within each stereomodel. The stations were surveyed by EDM and theodolite and targets positioned over the monuments for precise viewing and control pointing in the stereo model. Photography was taken with a Zeiss UMK 10/1318 camera using glass negatives for greater stability.

Due to the angle of slope of the cliff face from the camera a wide variation in photoscale resulted in this case, from 1:400 to 1:900. For each of the four faces under investigation, the control network was rotated about two points on top of the cliff thus forming a datum line parallel to the face itself. The shape of the cliff was then defined by contours at 0.25-m intervals of depth – i.e. as differences in the horizontal distance from this datum plane – rather than as more conventionally vertical elevations above a height datum. The accuracy of points shown on both the map and sections was within ± 5 cm of true ground position.

7.4.2.3 Tunnel profiling

Recent developments associated with the use of photogrammetry for producing tunnel profiles are described by Anderson and Stevens.[27] Anderson and Stevens outline the development and use of a 'mono-photogrammetric' tunnel profile measuring system. The two main elements of the system are a high-intensity light plane generator, which illuminates the section to be profiled, and a camera which records the line of light which is generated. By digitizing the line, applying corrections for lens distortion, and scaling the photographs, an accuracy of ± 10 mm is claimed. Furthermore, if the quoted progress rate of 100 sections per hour can be achieved in practice, the system would appear to be a highly cost-effective solution.

7.4.2.4 Laboratory-based applications

Close-range photogrammetry has also been applied to measurement problems in the laboratory; for example, Andrawes and Butterfield,[28] Wong and Vonderhoe,[29] and Davidson[30] have used the motion or 'false parallax' technique to measure the planar displacement fields associated with soil models. In the first two cases cited, the technique involved taking repeat photographs, from a single camera position, of a glass-sided tank which contained the soil under investigation. Photographs were taken using a nonmetric 35 mm camera before and after the sand within the tank was subjected to movement by a moving wall or wedge situated in the tank. By examination of enlargements of the photographs in an analogue stereoplotter, the relative positions of features could be measured and displacement contour diagrams produced. Analysis of these diagrams enabled the displacement component along the sand bed to be determined to within 5 μm. The latter two authors, in contrast, used analytical techniques to investigate the movement of soil particles around a model tunnel and soil penetration probe respectively.

7.5 Principles of remote sensing

The interpretation of aerial photography has for many years proved to be a valuable source of data for civil engineers. Until the early 1960s civilian 'remote sensing' was concerned primarily with the interpretation of such imagery. Since then, however, developments in orbiting satellites, sensor technology and computing have led to the creation of a discipline which now impinges on many areas in science and engineering. Although still in an embryonic state, particularly in the field of image processing, it has already proved to be a cost-effective method of investigating engineering phenomena of both large aerial extent, e.g. surface drainage, and those which are more localized, e.g. landslides, unstable land, etc.

7.5.1 The electromagnetic spectrum

Electromagnetic (EM) energy is all energy which travels in a periodic harmonic manner at the velocity of light. Electromagnetic energy is normally considered to consist of a continuum of wavelengths referred to as the EM spectrum (Figure 7.18).

It can be seen from Figure 7.18 that several regions of the EM spectrum are of particular importance in remote sensing. For example, the visible and reflected infra-red regions of the spectrum are important since they enable *reflected* solar radiation to be measured. In contrast, at longer wavelengths in the infra-red (8 to 14 μm waveband) the sensing of *emitted* thermal energy is of more significance. Measurements within the visible and infra-red (reflected and emitted) regions are considered to be *passive* in nature since the radiation being recorded occurs naturally. As the wavelength increases to the order of several millimetres it becomes more convenient *actively* to generate EM radiation of this wavelength and record the reflected radiation from the terrain. Thus, a typical *active* system would be side-

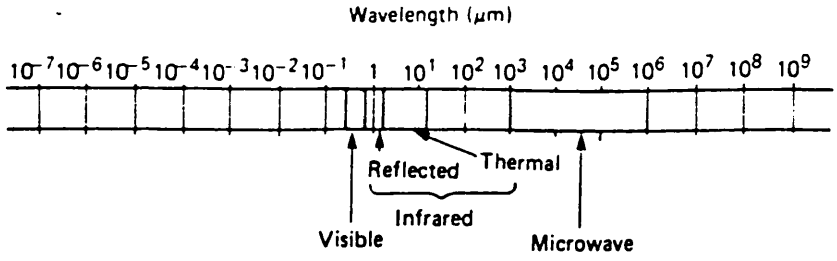
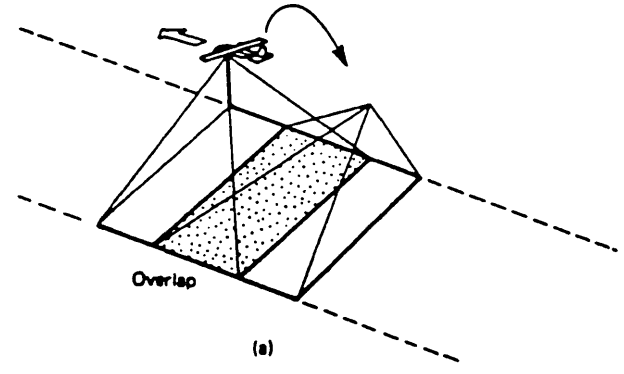


Figure 7.18 The electromagnetic spectrum

looking airborne radar (SLAR). It should be noted that instruments also exist for the measurement of *passive* microwave emission. However, since the emitted EM energy at this wavelength is very small, microwave radiometers are much less common than SLAR instruments.

Depending on the nature of the radiation being measured, it is possible to record the reflected or emitted energy by using either a lens and photographic emulsion or by using a linescanner and crystal detector. The geometrical distinction between the two approaches is illustrated in Figure 7.19. The primary advantage of using a linescanner approach is that it is possible to record radiation of wavelengths greater than about 0.9 μm. It is also possible to measure the variations in radiation within narrow spectral regions and to record directly these variations in digital form.

Mention should also be made here of the distinction between the terms 'photograph' and 'image'. A 'photograph' is an image which has been detected by photographic techniques and recorded on photographic film. In contrast, an 'image' is a more general term used to describe any pictorial representation of detected radiation data. Therefore, although scanner data may be used to create a photographic product, this result is normally referred to as an 'image' since the original detection mechanism involved the use of crystal detectors creating electrical signals, rather than a lens focused on to photographic film.



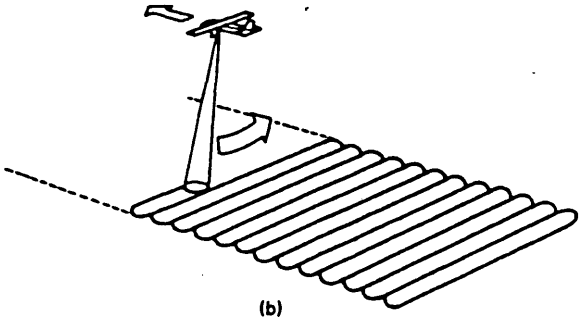


Figure 7.19 Geometry of data acquisition. (a) Camera; (b) linescanner

7.5.2 Classification of remote sensing systems

Remote sensing systems can be classified using various criteria, such as sensor sensitivity range, mode of recording (photographic or scanning), mode of operation (active or passive) or type of sensor platform (aircraft or satellite). Table 7.5 provides a framework for the classification of data acquisition systems in remote sensing by using the latter two criteria; it also provides some common examples of each category.

Table 7.5 Classification of data acquisition systems

	Aircraft	*Satellite*
Passive systems	Wild RC-10 camera Daedalus 1268 MSS Daedalus 1230 Thermal Scanner Barr and Stroud IR18 TVFS	Spacelab metric camera NASA large-format camera Landsat MSS, RBV, TM SPOT MOMS
Active systems	Goodyear GEMS Westinghouse SLAR SAR 580	Seasat SIR A/B ERS-1 Radarsat

7.6 Data acquisition

Data can be acquired for remote sensing from a variety of aerial platforms, although fixed-wing aircraft and unmanned orbiting satellites are the most common. Both have specific and complementary advantages. Aircraft, for example, enable small localized phenomena to be investigated at high levels of resolution, whereas satellites enable wide synoptic views of the terrain to be obtained, often on a repeatable basis, but at much lower resolution.

7.6.1 Airborne systems

7.6.1.1 Photography

Vertical panchromatic aerial photography taken with a high-

precision mapping camera is the most commonly available source of airborne remote sensing data (Figure 7.20(a) and Table 7.1). In addition to conventional aerial photography obtained using the system described above, it is also possible to utilize oblique aerial photography[31] (Figure 7.20(b)), low-cost or small-format photography,[32] or multiple-camera multispectral designs where each camera is filtered in order to record the reflected radiance in a particular waveband.[33] Alternatively, lower cost platforms such as microlight aircraft[34] or remotely piloted aircraft[35] may be used to obtain photography over a small site. A more recent development has been to produce photography from airborne video systems.[36]

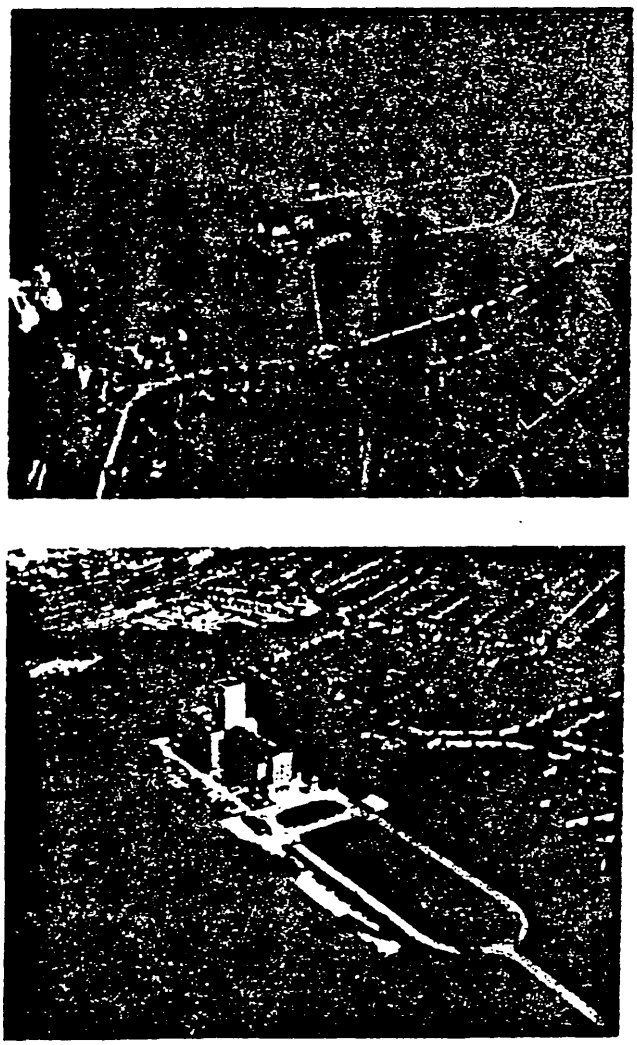


Figure 7.20 Aerial photographs of Guildford Cathedral. (a) Vertical; (b) oblique

7.6.1.2 Multispectral scanner

The principle behind the use of the multispectral scanner (MSS) concerns the detection of radiation from the Earth's surface. This radiance can be quantified by measuring the proportion of energy reflected or emitted by an element of the Earth's surface at various specific wavelengths. Each element is referred to as a pixel. An image can be created using an airborne linescanner by detecting the radiance in a linear pass across the ground perpendicular to the line of flight. Successive passes can be conducted at a rate commensurate with the aircraft flying speed, which can then be used to create an image. The image is therefore built up pixel by pixel in a sequential form, line by line. Crystal detectors in the MSS transform the incident radiation into electrical signals which are recorded in digital form on magnetic tape.

The technique used in the MSS is to detect both radiation *emitted*, and solar radiation *reflected*, from the Earth's surface. These forms of radiation, within several different bands of wavelength, can be detected simultaneously using this method. The advantage of using airborne systems rather than satellite systems (such as Landsat MSS) is the much higher spatial resolution of the former (less than a metre) and the greater number of available spectral wavebands.

One MSS which has been used in the UK for a variety of engineering projects is the Daedalus AADS 1268 Airborne Thematic Mapper (ATM). This instrument has an instantaneous field of view (IFOV) of 2.5 mrad and the spectral specifications of the instrument are presented in Table 7.6. Data are recorded in flight on high-density digital tape and subsequently reproduced in a standard form on to computer compatible tape (CCT) for analysis on a digital image processing system. Figure 7.21 illustrates a typical example of imagery obtained with this type of scanner.

Table 7.6 Spectral bands for the Daedalus 1268 AADS ATM

Spectral band	*Wavelength* (μm)	*Spectral region*
1	0.42–0.45	Blue
2	0.45–0.52	
3	0.52–0.60	Green
4	0.605–0.625	Red
5	0.63–0.69	
6	0.695–0.75	Near IR
7	0.76–0.90	
8	0.91–1.05	
9	1.55–1.75	
10	2.08–2.35	Mid IR
11	8.50–13.00	Far (Thermal) IR

7.6.1.3 Thermal infra-red (IR) scanning

Thermal IR scanning techniques are similar to those utilized by MSSs but concentrate on a single wavelength within the far IR region of the spectrum. Two regions can be sensed by such instruments, corresponding to atmospheric 'windows' in the 3 to 5 μm and 8 to 14 μm regions. The former region is used primarily for examining very hot objects, e.g. forest fires, and is rarely used for engineering applications. The latter region, corresponding to the peak value of emitted radiation from the Earth, is of much greater importance.

The most common thermal IR sensor is the linescanner which is a widely used instrument based on a scan geometry identical to that for MSS devices. A cadmium mercury telluride (CMT) detector is used to record the temperature variations over the scene and a calibrated internal blackbody reference enables precise quantitative temperature variations to be measured. Several thermal IR linescanners are currently in operation in the UK, including a Daedalus 1230 dual-channel scanner and as mentioned previously a Daedalus 1268. Figure 7.22 shows a typical example of a thermal image obtained with this scanner which could be used to assess heat loss.

An alternative technique for sensing in the thermal IR region of the spectrum which has recently been evaluated for remote sensing applications in environmental engineering is the thermal video frame scanner (TVFS).[37] Several benefits accrue from the

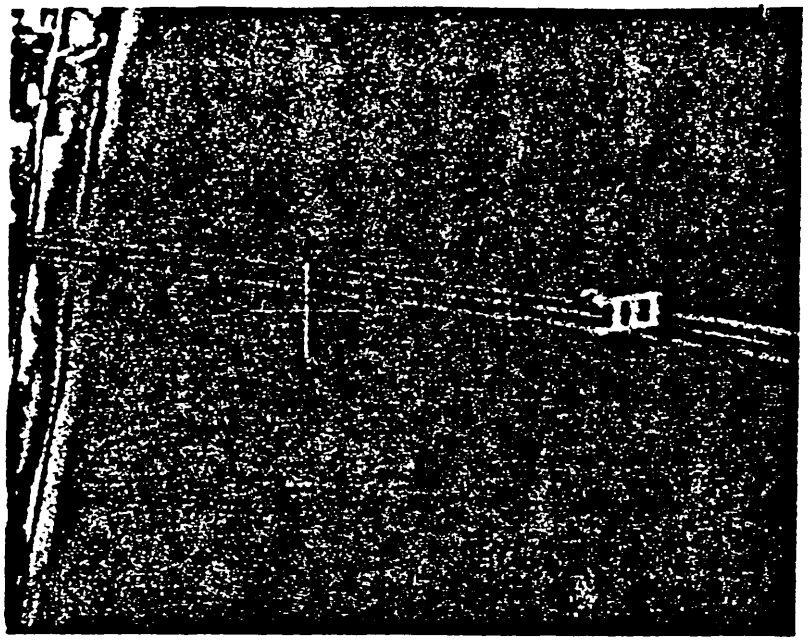


Figure 7.21 Airborne multispectral scanner image of the Humber Bridge illustrating sediment transport patterns. (*Courtesy*: Huntings Geology and Geophysics)

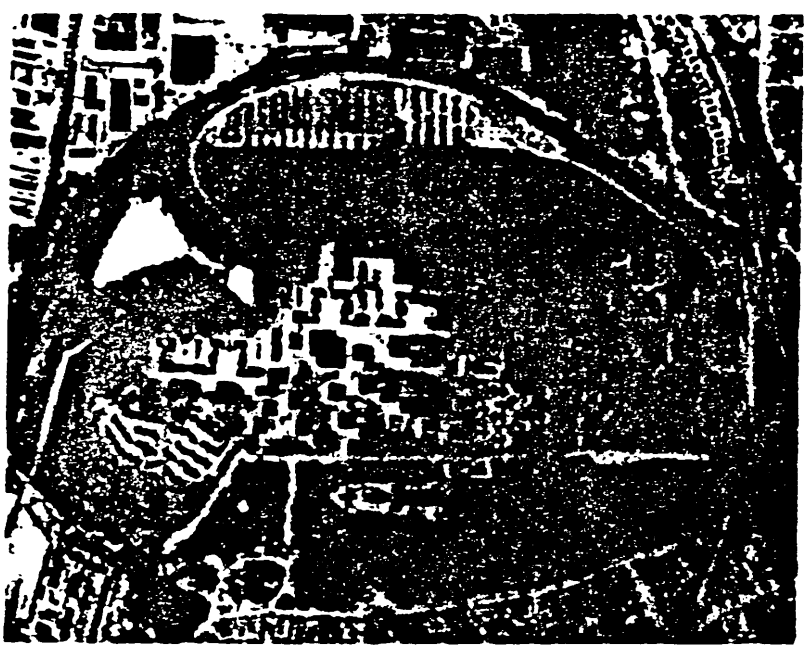


Figure 7.22 Thermal infra-red linescan image of the University of Surrey. (*Courtesy*: Clyde Surveys Ltd)

use of video-based instruments. These include system portability, almost real time verification of data capture, rapid turnaround (since there is no processing stage), low cost of operation (since the sensor can be operated from a light aircraft) and ease of duplication of data. Unlike linescanning systems, however, these instruments normally do not include a calibrated internal blackbody reference for quantitative temperature measurements. For some applications this could prove to be a severe limitation. However, for many engineering applications an assessment of the relative temperatures within a scene can be just as important as a detailed knowledge of the absolute ground temperatures. It should be noted that the quantitative determination of temperatures can, nevertheless, be carried out by correlating grey scale levels and ground control points of known temperature.[8] One example of this type of instrument which is currently being used for remote sensing applications is the Barr and Stroud IR18 TVFS system.

7.6.1.4 Side-looking airborne radar (SLAR)

Side-looking airborne radar was first developed for military purposes in the early 1950s. It was not, however, until the early 1970s that it became available commercially. The main impetus for the development of SLAR arose out of its two significant advantages over optical photographic or scanning systems: (1) its ability to produce imagery both day and night; and (2) its

ability to 'sense' through haze, smoke, cloud and even rain. These radar systems are consequently in great demand for the production of imagery in equatorial regions of the Earth's surface.

The main reason for these twin advantages stems from SLAR's use of microwaves as the imaging source. The image which is produced by SLAR is very different from that produced by conventional optical systems. The view which is obtained is a record of the Earth's reflective properties at microwave wavelengths. Consequently, the nature and intensity of the reflections by SLAR will be influenced by factors such as ground conductivity and surface roughness, which are much less significant with optical systems. In addition, the geometrical properties of the image differ significantly from a conventional aerial photograph.

The principle of operation of a typical SLAR system is illustrated in Figure 7.23.[7] It involves the measurement of the time interval between the transmission and reflection of a microwave pulse. This indirectly provides the range from the aircraft to the ground feature. Successive pulses can then be sensed sequentially so providing a radar picture of the terrain. Figure 7.24 is a large-scale image obtained using an airborne radar system.

Two distinct designs of SLAR exist. The earliest designs were termed real aperture radar (RAR) systems. Although able to sense through cloud their ground resolution was limited by the size of the antennae which could be mounted on the side of the aircraft. In order to overcome this limitation, an alternative design, synthetic aperture radar (SAR) has been developed. Using suitable computer processing such a system is able to provide much higher resolution imagery than that which can be provided by a RAR system. Table 7.7 outlines some of the characteristics of a selection of SLAR systems.

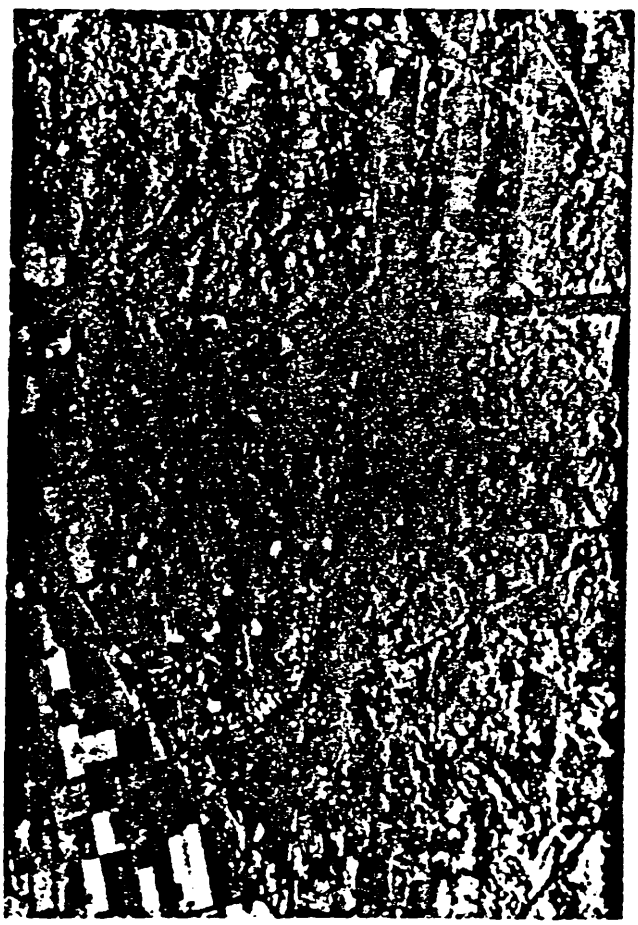


Figure 7.24 Side-looking airborne radar image

7.6.2 Satellite systems

Satellite remote sensing systems range from low-resolution systems such as the meteorological satellites to high-resolution photographic missions such as the recently flown metric camera on board the space shuttle. Although both have some application to civil engineering the low ground resolution of the former system and the limited coverage of the latter restrict their application considerably. Of much greater importance to civil engineers are the Earth resources satellites operating the visible, infra-red and microwave regions of the spectrum.

7.6.2.1 Earth resources satellites - passive

A selection of the characteristics of the most common Earth resources satellites is given in Table 7.8.

Landsat. The Landsat satellite system, previously known as the Earth Resources Technology Satellite (ERTS), was initiated in 1967 by NASA in conjunction with the US Department of the

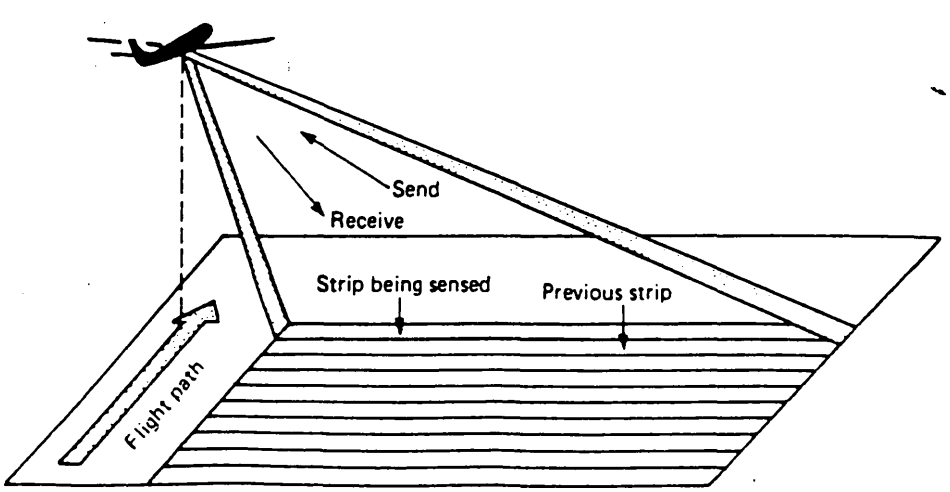


Figure 7.23 Side-looking airborne radar (After Rudd (1974) *Remote sensing: a better view* Duxbury Press Massachusetts

Table 7.7 Side looking airborne radar systems

	Motorola AN/APS	*Westinghouse*	*Goodyear GEMS*	*SAR-580*
Type	RAR	RAR	SAR	SAR
Wavelength band	× (3 cm)	K(0.8 cm)	× (3 cm)	× (3 cm)
Resolution (m)	15	15	12	3

Table 7.8 Characteristics of passive Earth resources satellites

	Landsat's 1-5 multispectral scanner (MSS)	*Landsat's 1-3 return beam vidicon (RBV)*	*Landsat's 4-5 thematic mapper (TM)*	*Modular optoelectronic multispectral scanner (MOMS)*	*Le Système Probatoire d'Observation de la Terre (SPOT)*
Operated by (country)	EOSAT (USA)	EOSAT (USA)	EOSAT (USA)	DFVLR (W. Germany)	SPOT Image (France)
Date of launch	1972	1972–81	1982	1983	1986
Orbital altitude (km)	900	900	705	300	830
Ground resolution (m)	80	80, 30	30	20	10, 20
Spectral range (μm)	0.5–12.6	0.505–0.75	0.45–12.5	0.575–0.975	0.5–0.89
No. of wavebands	5	1	7	2	3
Further reading	NOAA[40]	NOAA[40]	NOAA[40]	—	Chevrel, Courtois and Wells[41]

Interior. The system was initially designed as an experiment in order to assess the feasibility of collecting Earth resource data from unmanned satellites. However, following the commercialization of remote sensing activities in 1985, the current operational activities of Landsat have been transferred to EOSAT, a joint venture formed by Hughes and RCA.

Three distinct sensors have been carried by Landsat: (1) a MSS; (2) a return beam vidicon (RBV); and (3) a thematic mapper (TM). The MSS is a linescanning device which uses an oscillating mirror to scan at right angles to the satellite flight direction. The IFOV of the sensors operating in the visible and near IR produce a resolution cell of approximately 56 × 79 m. The fifth channel has an IFOV of 0.258 mrad, or a ground resolution of about 235 m.

The MSS scans each line from west to east with the southward motion of the satellite providing the along-track progression of the scan lines (Figure 7.25).[40] Each Landsat MSS scene covers an area approximately 185 × 185 km. In view of the extremely high mirror oscillation rate which would be required using this approach if only one line was scanned, the system is designed to scan six lines simultaneously with each oscillation of the mirror. This results in an area 474 m × 185 km being recorded with each sweep.

A typical Landsat scene, 185 × 185 km, consists of 2340 scan lines with about 3240 pixels per line; each image therefore consists of over 7.5 million digital values. With four spectral observations per pixel, this means that over 30 million values have to be recorded for every Landsat scene.

The simplest Landsat MSS product is a photograph consisting of a black-and-white image for each spectral band (Figure 7.26). Although the resulting image is vaguely familiar, because it is similar to a conventional panchromatic aerial photograph, much of the information content of the image is lost. An alternative approach is to assign a different colour to each

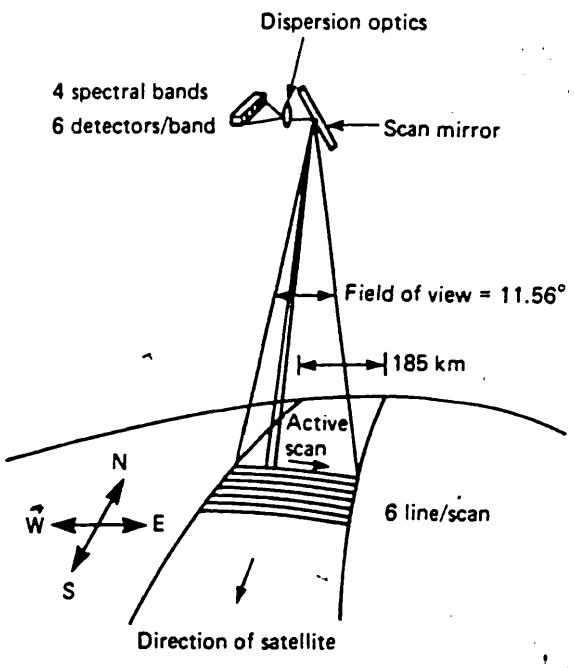


Figure 7.25 Landsat multispectral scanner[40]

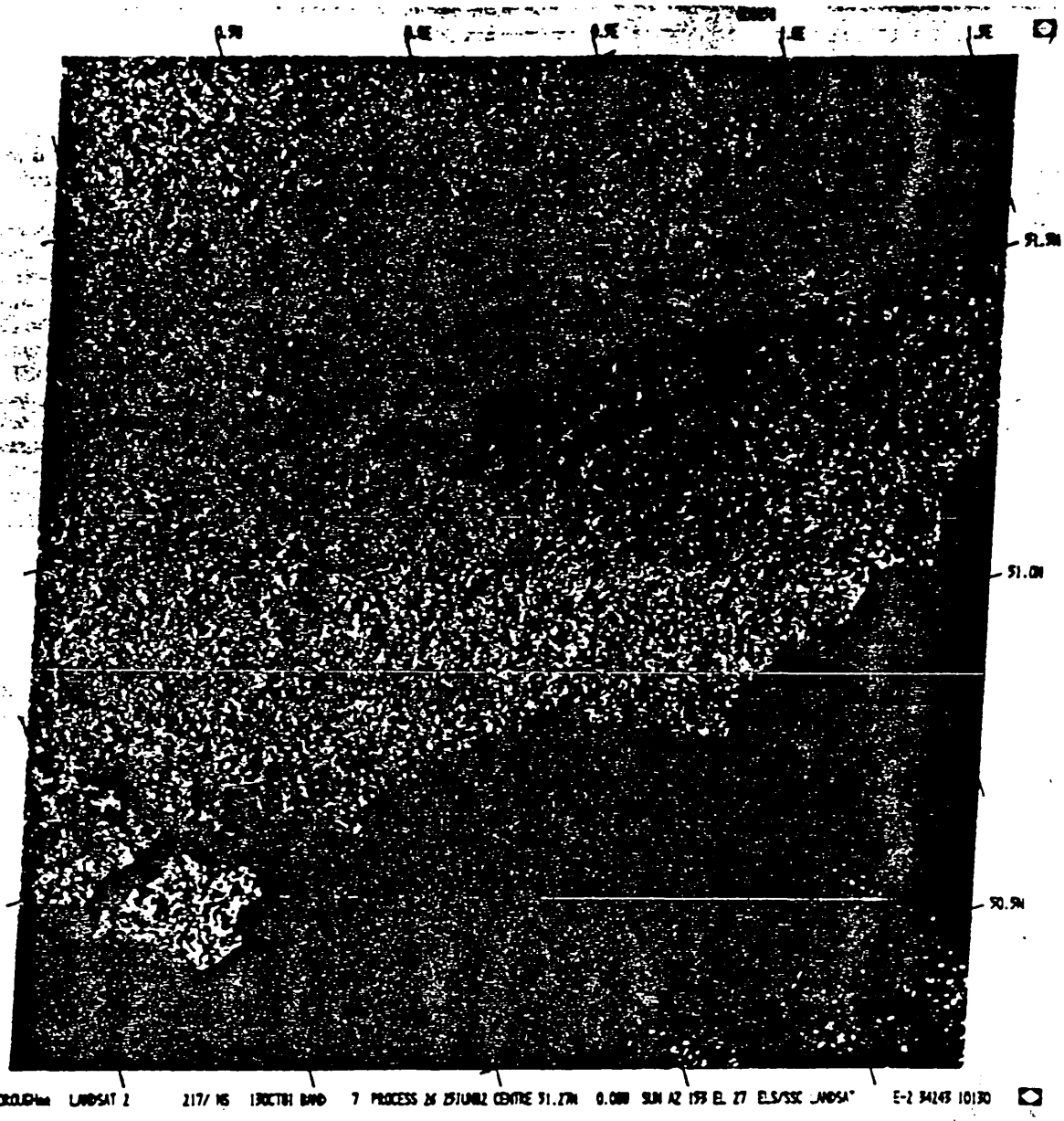


Figure 7.26 A black-and-white image as presented by Landsat of Southeast England created by a MSS. (*Courtesy*: National Remote Sensing Centre)

spectral and superimpose these to produce a colour-composite image. Since the sensor also senses-in the reflected infra-red region of the spectrum, the consequent composite does not provide a true colour-coded image of the terrain but rather a 'false' colour image.

The second sensor, the RBV, was designed to provide high-resolution imagery suitable for mapping. In addition to having a high ground resolution (30 m on Landsat 3) it also had a réseau grid superimposed on the image. In recent years however, the advent of the TM has tended to reduce the importance of both MSS and RBV imagery, apart from forming a historical record.

The thematic mapper has been operational since 1984 and it provides data of higher spatial resolution (30 m) and finer spectral resolution (seven bands) than that available from the MSS or TM sensors. The comparative spatial resolution of the sensors is illustrated in Figure 7.27.

Modular optoelectronic multispectral scanner (MOMS). The modular optoelectronic multispectral scanner (MOMS) was designed by the West German MBB company for the German Aerospace Research establishment (DFVLR), primarily as a research system. The first satellite-borne imagery was obtained with the system during the seventh space shuttle flight in June 1983. The scanner has several unique design features of which the most significant are the dual lens system and four linear arrays of 1728 pixels which enable a continuous line of 6912 pixels to be swept out by the scanner (equivalent to 20 m on the ground).

A second-generation MOMS system is currently under development by MBB which will offer the possibility of obtaining stereoscopic imagery and which will also have an extended range of spectral bands.

Le Système Probatoire d'Observation de la Terre (SPOT). This is a commercial remote sensing system developed by the French Government and aerospace industry and operated worldwide by SPOT image. The sensors on board the satellite consist of two high-resolution visible (HRV) imaging instru-

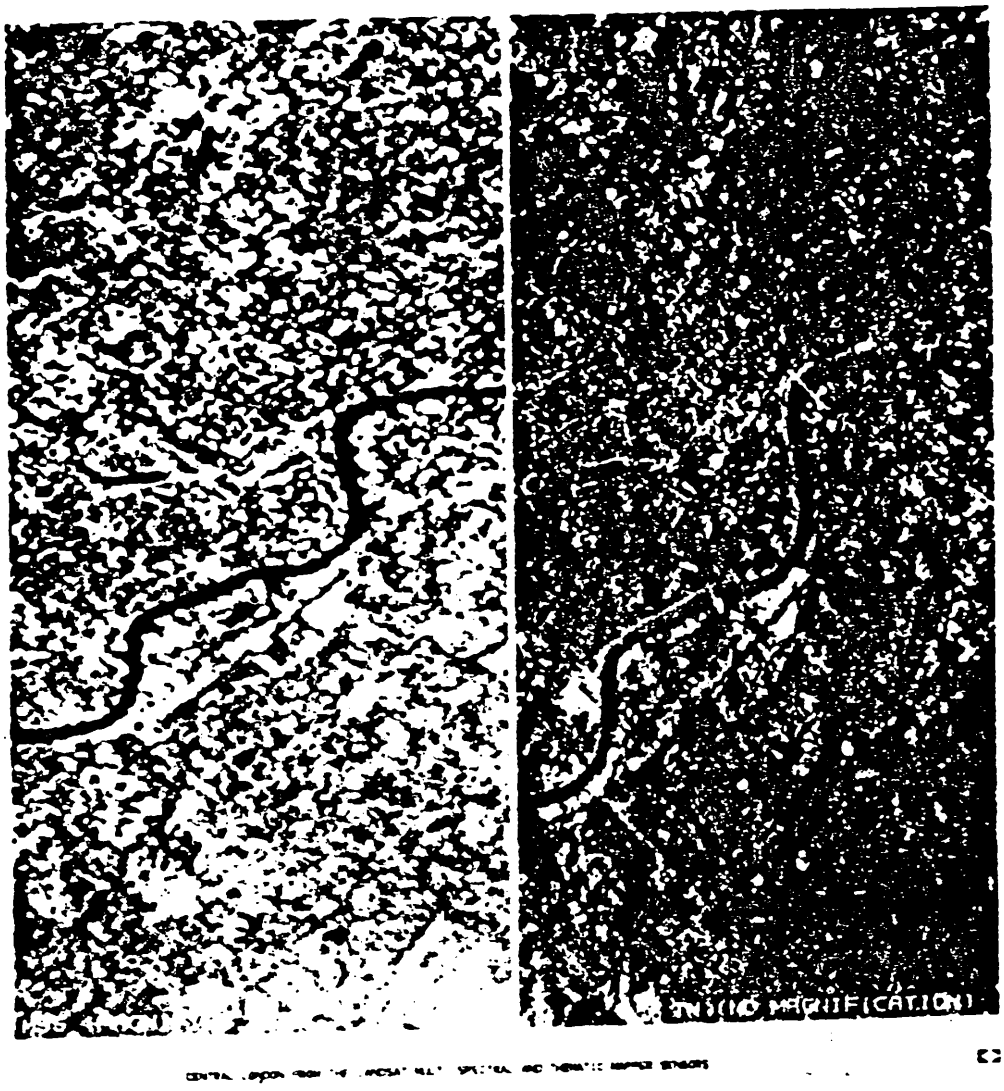


Figure 7.27 A comparative spatial resolution of the Landsat MSS (left) and thematic mapper sensors. (*Courtesy*: National Remote Sensing Centre)

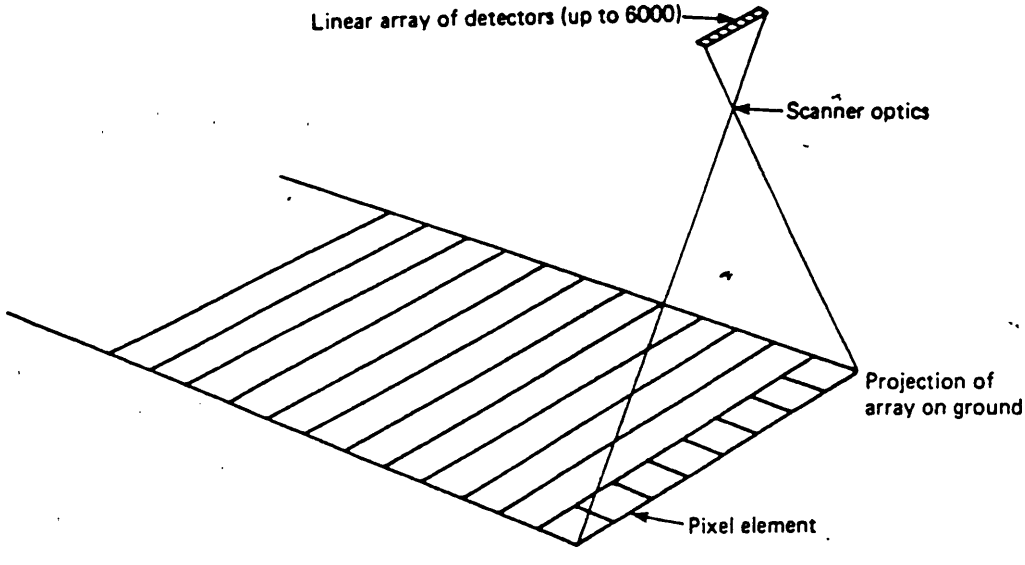


Figure 7.28 SPOT: pushbroom scanner design

ments employing the multilinear array or 'pushbroom' design of MSS. This design of MSS differs from the optical–mechanical design discussed previously. In this case each line of the image is formed by measuring the radiances which are imaged directly on to a one-dimensional linear array of small detectors located in the instrument's focal plane. Each line is subsequently scanned electronically and the radiance values recorded on to magnetic tape. As before, successive lines of the image are produced by the forward motion of the satellite along its orbital path. This pushbroom concept is illustrated diagrammatically in Figure 7.28.

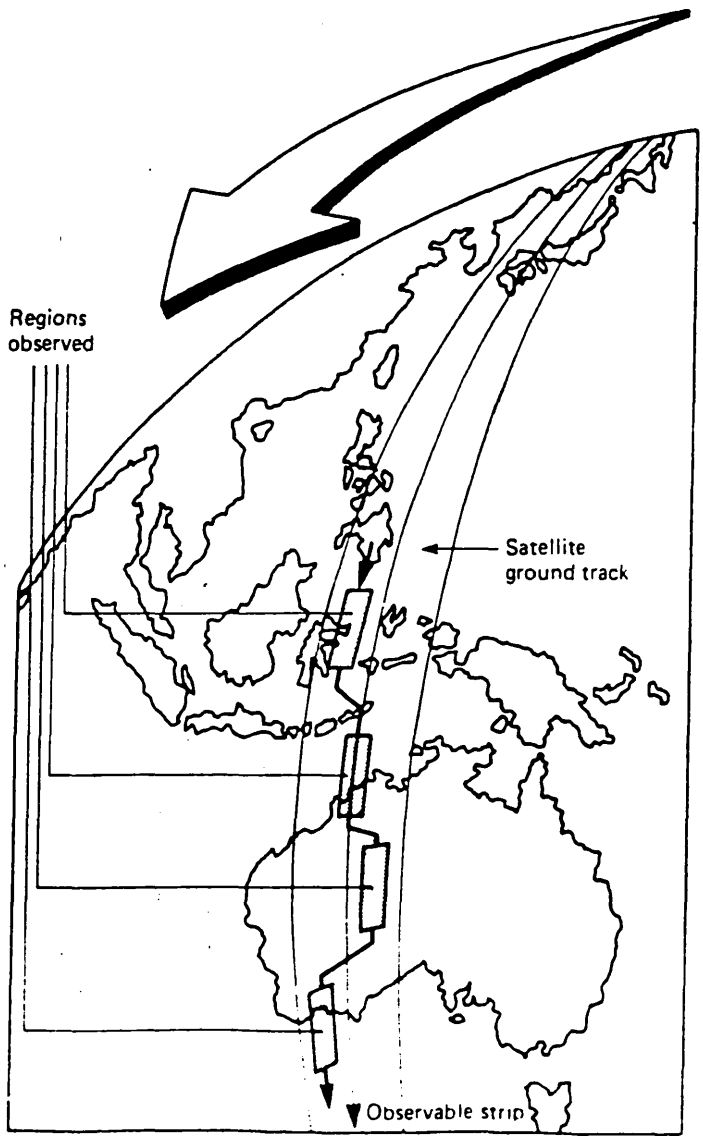
A further important feature of the SPOT system from an engineering point of view is its capacity to provide stereoscopic coverage of the Earth's surface from the lateral overlap between successive scenes. A rotatable mirror in the SPOT sensor package also permits scenes to be acquired over areas up to 400 km left or right of the normal vertical vantage point of the satellite. This feature will permit much easier acquisition of high-priority scenes (Figure 7.29). Figure 7.30 shows a typical example of a SPOT panchromatic image (10 m pixel).

7.6.2.2 Earth resources satellites – active

Following the success of the Seasat satellite SAR mission in 1978 (Table 7.9 and Figure 7.31), a great deal of interest has been shown in the development of further side-looking radar satellites. For example, the forthcoming ERS-1 satellite will be used to provide continuous monitoring of ocean parameters such as wave and wind height and pattern, which may help to improve the engineering design of oil platforms. The SAR data will also be used to aid the planning of shipping routes particularly in Arctic regions. Similarly, the Canadian Radarsat will provide imagery to enable the positions of icebergs to be monitored more precisely. This data will be used by oil tankers transporting oil and natural gas through the Northwest passages from oil fields in and around the Arctic islands of northern Canada.

7.7 Digital image processing (DIP)

Digital image processing is a crucial stage in the effective use of



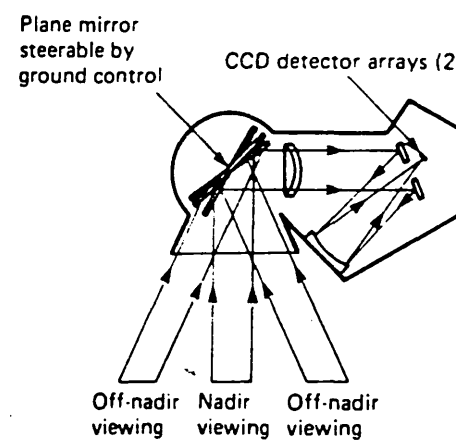


Figure 7.29 SPOT nadir and off nadir viewing (*Courtesy* SPOT Image)



Figure 7.30 SPOT: panchromatic image of part of Montreal (SPOT data copyright CNES, 1986, image provided by Nigel Press and Associates)

modern remote-sensing data. Although a great deal of information can be gleaned from the interpretation of a photographic image produced from, for example, Landsat, the full potential of the data can only be realized if the original data, edited and enhanced to remove systematic errors, is used with a suitably programmed image processing system.

7.7.1 Hardware

Until relatively recently, most DIP of remote sensing data was carried out on large, expensive, dedicated systems such as that illustrated in Figure 7.32. Such a system enables data to be read into the system from a magnetic tape reader, to be displayed on a colour TV monitor and, after suitable processing, to be output to a colour filmwriter for the production of high-quality imagery. The trend, however, is towards smaller and cheaper image processing based on the new generation of 32-bit super minicomputers and 16-bit IBM-compatible personal computers.[42] Table 7.10 lists a selection of the remote-sensing systems currently available.

7.7.2 Software

The range of algorithms which are used to restore and classify remote-sensing data is extremely wide and varied. Furthermore, an appreciation of the choice of the most appropriate technique,

Table 7.9 Characteristics of a selection of active Earth resources satellites

	Seasat	*Shuttle imaging radar (SIR)-A*	*SIR-B*	*European radar satellite (ERS-1)*	*Radarsat*
Date of launch	1978	1981	1984	1989	1991
Operated by	NASA	NASA	NASA	European Space Agency	Canada
Altitude (km)	800	245	255, 274, 352	675	1000
Wavelength (mm)	230	230	540	540	540
Look angle (°)	20	50	57	–	30–45
Resolution (m)	25	40	30	25	30
Swath width (km)	100	50	20–25	80	150

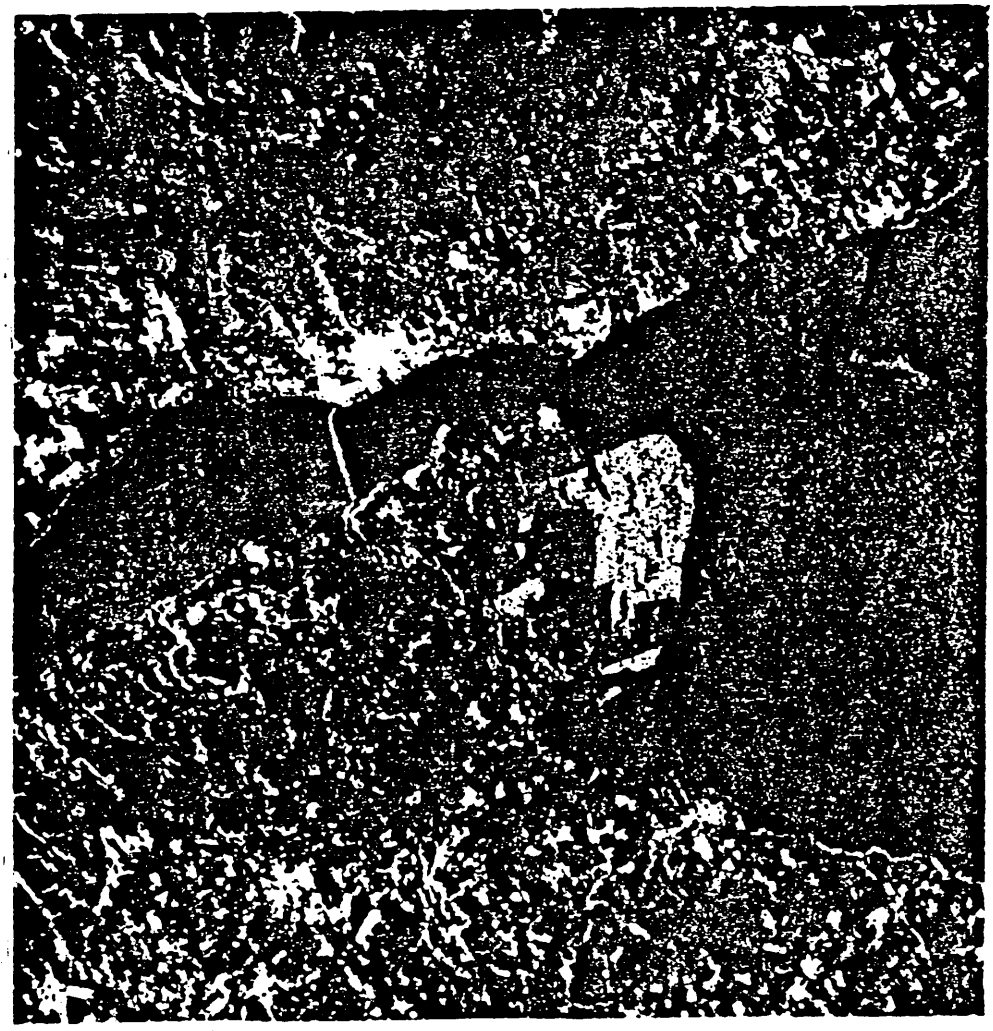


Figure 7.31 SEASAT SAR image of the Tay estuary (*Courtesy* National Remote Sensing Centre)

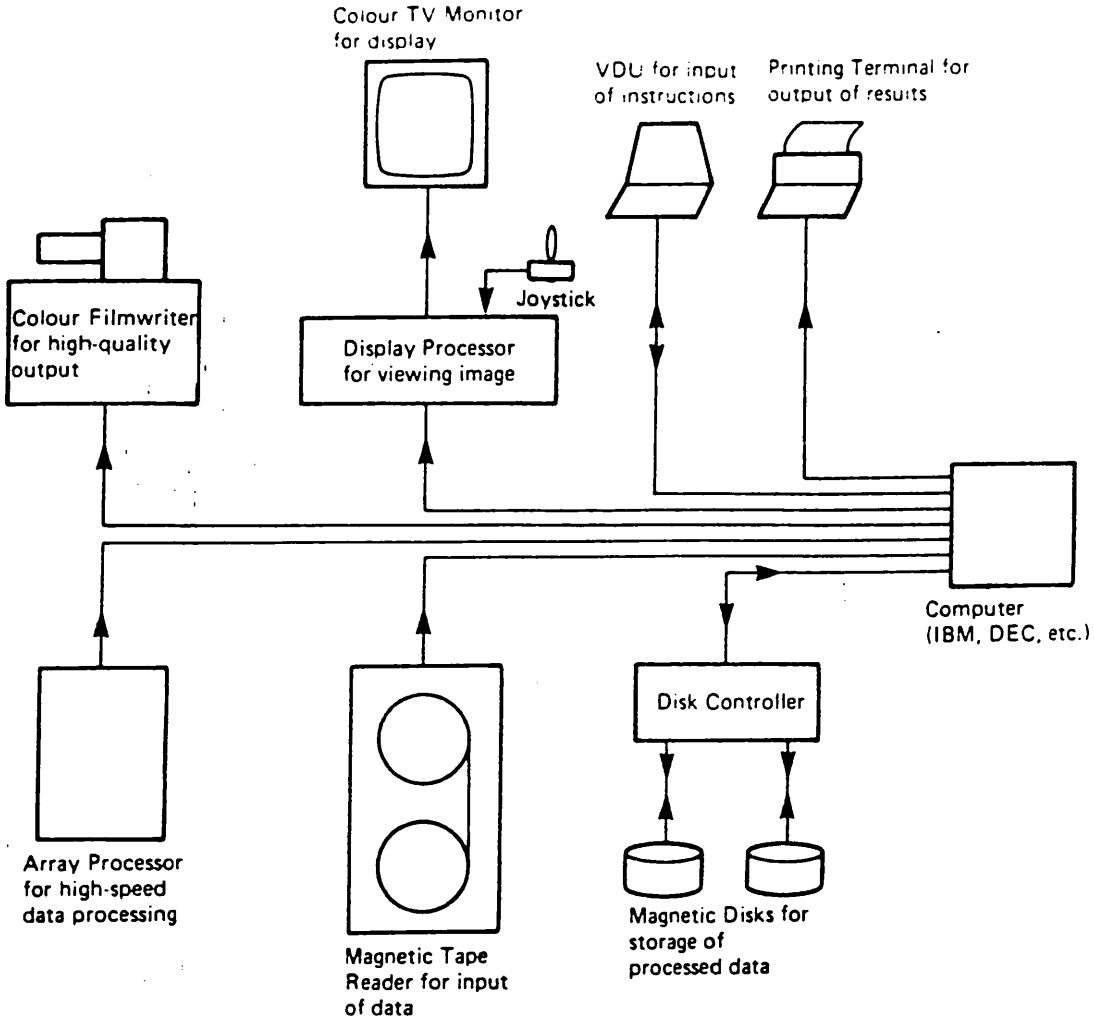


Figure 7.32 Digital image processing system

and the advantages and limitations of each technique, is a subject of some complexity beyond the scope of this chapter. Consequently, only a very brief description of the most commonly used techniques will be provided. For further details Bernstein,[43] Avery and Graydon[44] and Bagot[45] should be consulted.

7.7.2.1 Image restoration

Geometric correction. Several preliminary transformations are normally carried out on raw satellite data to reduce the effects of the Earth's rotation and curvature and variations in the attitude of the satellite. If a sufficient number of ground control points are available the imagery can be fitted to the control using a least squares technique. Such a procedure is also necessary if the remote sensing data is to be combined with existing map data.

Radiometric correction. Variations in the output from the detectors used in a MSS may cause a striping effect on the final image. Such an effect can be eliminated by a process termed 'destriping', which effectively interpolates the missing radiance values from the outputs of adjacent pixels. In some cases, complete scan lines may be missing from the data; preprocessing can also be carried out to overcome this type of problem.

Table 7.10 A selection of remote-sensing systems currently available

	Name of system	*Company*
Dedicated image-processing system	Gemstone 33	GEMS of Cambridge, UK
	I²S Model 75	Int. Imaging Systems, California, USA
	Vicom VDP Series	Vicom Systems, San Jose, USA
	Dipix Aries II	Dipix Systems, Ottawa, Canada
Personal computer-based systems	Diad Systems	Diad Systems, Edenbridge, Kent, UK
	LS10	CW Controls, Southport, Lancashire, UK
	Microimage	Terra Mar, California, USA
	ERDAS	Earth Resources Data Analysis System, Atlanta, USA
	ImaVision	RBA Associates, Ottawa

7.7.2.2 Image enhancement

Contrast stretching. Contrast stretching is a technique which is used to brighten an image and produce one which uses the full dynamic range of the data. It is often carried out automatically and is generally one of the first operations carried out by a user when viewing a new image.

Spatial filtering. Spatial filtering is carried out in order either to enhance boundaries between features (edge enhancement or high pass filtering) or to smooth and eliminate boundaries (smoothing or low pass filtering). The former approach is most suitable for enhancing geological features, whereas the latter may be used to reduce random noise from an image.

Image ratioing. Image ratioing is the process of dividing the radiance values of each pixel in one waveband by the equivalent values in another waveband. Ratioed images reduce radiance variations caused by local topographic effects and consequently may be used to enhance subtle spectral variations.

Principal components analysis. Principal components analysis is a statistical technique which is used to improve the discrimination between similar types of ground cover. It is based on the formation of a series of new axes which reduce the correlation existing between successive wavebands in a multispectral data set.

7.7.2.3 Image classification

Density slicing. Density slicing is the simplest form of automated classification. It enables a single band (often of thermal IR data) to be colour-coded according to the grey level of the individual pixels. Thus, for example, pixel values 0 to 19 may be coloured blue, 20 to 25 red and so on. This not only aids interpretation but if the density slices are related to temperature levels, the resulting image can be used quantitatively to assess the variations in temperature across the scene.

Supervised classification. In cases where ground data exists about the terrain under investigation, it is possible to use this data to 'train' the computer to perform a simplified type of automated interpretation. For example, a particular group of pixels may be known to represent an area of water. Examination of the maximum and minimum numerical values of the radiances in each of the spectral bands being used (seven for Landsat TM) would enable a 'spectral pattern' for water to be determined. If the computer then compared this multidimensional spectral pattern with the spectral pattern of every pixel in the scene it would be possible to display only those areas which satisfied the criteria set by the training sample. Consequently, all water areas over a large area (185 × 185 km for Landsat) could be determined.

The process outlined above refers to the simplest form of classifier, the 'box' classifier. Other more sophisticated techniques are also available.

Unsupervised classification. In unsupervised classification of multispectral data, no attempt is made to 'train' the computer to interpret common features. Instead, the computer analyses all the pixels in the scene and divides them into a series of spectrally distinct classes based on the natural clusters which occur when the radiances in *n*-bands are compared. The user may now analyse other evidence to determine the nature of each of the distinctive classes which have been identified by the computer. Unsupervised classification techniques are generally more expensive in terms of processing requirements than supervised techniques.

7.8 The use of remote sensing in civil engineering

Although aerial and satellite remote-sensing imagery (other than black-and-white aerial photography) have been used for topographic mapping, the primary role of this type of imagery has been for the production of thematic maps. Generally, for such applications users are satisfied with a relatively low level of positional accuracy and are also willing to accept a lower level of completeness than is the case with topographic maps. The type of remote-sensing data which is appropriate will depend largely on the stage of the civil engineering project and the degree of economic development of the region where the project is being carried out.

Five main stages can be identified for a civil engineering project: (1) reconnaissance; (2) preliminary planning/feasibility; (3) design; (4) construction; and (5) post construction/maintenance. The potential role of remote sensing at each stage will vary from project to project, but the following sections indicate some of the possible uses of the techniques at each stage.

7.8.1 Reconnaissance level investigations

The first stage of any major civil engineering project generally involves some form of preliminary reconnaissance study of the region or project area. The primary aim of this study is to collect together all available data concerning the physical characteristics of the terrain in order to assess its likely influence on the overall design of the engineering works.

Traditionally, the engineer has carried out this reconnaissance stage by examining existing maps of the region. Both topographic maps and specialist thematic maps (e.g. geological, geomorphological and pedological) are of greatest use at this initial stage of the project. In addition, examination of engineering reports produced for previous projects in the region may also provide valuable reference material. The main problems associated with this approach are that: (1) in many regions existing maps may be at a very small scale ($< 1:500000$), be significantly out of date and, in some instances, may not exist at any scale; and (2) that whilst engineering reports may give valuable data about a specific site they may not be appropriate or representative of regional terrain conditions. Consequently, many engineering projects have involved an examination of satellite and aerial imagery at this early stage.

The use of Landsat MSS data for projects of this type has been reported by Beaumont.[46] In this case, use of Landsat data for water resources planning in northeast Somalia is reported. Visual interpretation of Landsat data enabled a series of transparent overlays to be produced illustrating drainage, surface water, groundwater potential and land capacity over the region.

The use of remote sensing for terrain evaluation purposes is also of considerable importance. Terrain evaluation is a form of thematic mapping in which the terrain is classified in a hierarchical manner into units having common landscape patterns and engineering characteristics. This process can be carried out very cost-effectively using satellite imagery in conjunction with aerial photography.[47]

Interpretation of Landsat data may also be useful for the provision of additional supplementary information on hydrological phenomena, e.g. flooding, water quality, groundwater potential and water depth. A more general review of hydrological applications can be found in Blyth.[48]

Landsat has also been used for estimating urban populations

over large areas,[48] and using multitemporal imagery for monitoring change, e.g. the expansion of urban areas, changes in soil erosion, varying river channels, the impact of desertification and so on.

7.8.2 Preliminary planning/feasibility

The aim of the preliminary planning/feasibility stage of a project is: (1) to select potential routes/sites; and (2) to select from these the best from the available options. Depending on the state of economic development of the region, it is likely that different factors will influence the choice of the optimum route/site. For example, in developing regions the location of low-cost construction materials such as calcrete may be important[50] or the identification of potential bridge crossing-points. Both activities can be carried out cost-effectively using satellite imagery in conjunction with aerial photography. A comprehensive review of the role of remote sensing for highway feasibility studies can be found in Lawrance and Beaven.[51]

7.8.3 Design

Remote sensing has a particularly valuable role to play at the design stage of a project. At this stage a detailed site investigation of the proposed site or route will be required, together with supplementary environmental information which may influence the design of the project being considered. A critical stage of a site investigation is the desk study. The desk study provides the engineer with an overview of the project area and is normally carried out using aerial photography. Not only can valuable information be obtained during this interpretation about possible engineering problems but it can also be very important for the planning of the subsequent ground investigations. A general review of the potential of aerial photography in this respect is provided by Dumbleton[52] and Matthews,[53] and a selected number of papers which consider the use of remote sensing for geotechnical investigations are presented in Table 7.11.

Of increasing importance in recent years has been the role of remote sensing in providing environmental engineering data which may provide useful supplementary data to the engineer to input into the design process. A comprehensive review of remote sensing in this context is provided by Mason and Amos.[54] Information which may be extracted from remote-sensing imagery (often thermal IR linescan) includes data on energy conservation (heat loss), power station cooling-water patterns, groundwater and spring detection, pipeline and field drainage patterns, hydraulic leakage from earth and concrete dams and the location of ice-prone regions on roads.

7.8.4 Construction

The role of remote sensing is much smaller during construction than at the preceding planning and design stages. Nevertheless, aerial photography can be used to provide a valuable historical record of construction activities and may also be used to plan construction activities. Singhroy[47] provides an interesting case-history concerning the use of large-scale colour IR aerial photography and aerial video during pipeline construction in Canada. Both types of data were used to plan construction activities, determine site conditions and assess environmental effects before, during and after construction activity.

7.8.5 Post-construction maintenance

As mentioned in the previous section, remote sensing can be used to assess the environmental impact of construction activities. Remote sensing may also be used to assess the use of a new road by conducting aerial traffic investigations.[48] Nevertheless, the use of remote sensing at this stage is again of much less importance than during the previous four stages which have been discussed.

In conclusion, Table 7.12 illustrates the recommended use of remote-sensing techniques at each stage of a civil engineering project.

7.9 Sources of remote-sensing data

7.9.1 Satellite data

The most general source of satellite data and information on the availability and cost of data are the national points of contact (NPOC) which exist in most countries throughout the world. In the UK, for example, the NPOC is the National Remote Sensing Centre, Royal Aircraft Establishment, Farnborough.

Several other organizations may have archives of satellite data or may act as distributors for Landsat or SPOT data. Table

Table 7.11 Remote sensing and site investigations – selected references

	Aerial photography	*MSS*	*Thermal IR linescanning*
Landslides/slope instability	Norman, Leibowitz and Fookes[55] Matthews and Clayton[31]	Liu[56]	Chandler[57]
Solution features	Norman and Watson[58]	Coker, Marshall and Thompson[59]	Kennie and Edmonds[60]
Derelict/contaminated land	Bullard[61]	Coulson and Bridges[62]	Ellyet and Fleming[63]
Surfacial/subsurface materials surveys	Caiger[64]	Lynn[65]	Singhroy and Barnett[66]

Table 7.12

		Satellite						*Aircraft*				
Project phase	*Photograph/ image scale*	*Photog.*	*Landsat MSS*	*Landsat TM*	*SPOT*	*MOMS*	*Side-looking radar*	*Photog.*	*MSS*	*Thermal*	*SLAR*	*Video*
Reconnaissance	1:1 000 000 to 1:50 000	•••	••••	•••	•••	•••	•••	••	–	–	–	–
Planning/feasibility	1:100 000 to 1:20 000	•••	••	•••	•••	•••	•••	••	–	–	–	–
Design	1:20 000 to 1:2000	–	–	–	•	–	–	••••	•••	•••	•••	••
Construction	1:2000 to 1:500	–	–	–	–	–	–	••••	•	•	•	•••
Post construction maintenance	1:2000 to 1:500	–	–	–	–	–	–	••••	•	•	•	••

••••Very useful •••Very useful (but coverage limited) ••Useful •Of limited use – inadequate

7.13 lists a selected sample of the organizations who provide remote-sensing services and may also hold copies of satellite data.

7.9.2 Aerial photography

As mentioned in Table 7.13, several of the commercial remote-sensing organizations offer facilities for obtaining airborne thermal and MSS data and should be consulted if such data are required. However, aerial photography remains the most popular source of data for remote-sensing applications in civil engineering. Table 7.14 lists some of the organizations which have archives of aerial photographs. In addition, several of the organizations listed in Table 7.13 may also have holdings of aerial photographs.

Table 7.13 Sources of airborne and satellite data

Name of Company/Organization	*Based in*	*Comments*
Clyde Surveys Ltd	Maidenhead, Berks, England	Operate Daedalus thermal scanner
DFVLR	Oberpfaffenhofen, W. Germany	Distributors of MOMS and metric camera data
EOSAT Corp.	Arlington, Virginia, USA	Responsible for operation of Landsat
GeoSurvey Ltd	East Molesey, Surrey, England	–
Huntings Geology and Geophysics Ltd	Borehamwood, Herts, England	Operate Daedalus airborne MSS
EROS Data Centre	Sioux Falls, S. Dakota, USA	Prime inter. distribution centre for Landsat
Nigel Press and Associates	Edenbridge, Kent, England	Distributors of SPOT data in UK
SPOT Image	Toulouse, Cedex, France	Main operator for SPOT
SPOT Image Corp	Washington DC, Reston, Virginia, USA	Distributors of SPOT data in N. America

Table 7.14 Sources of aerial photography

Name of Organization	*Based in:*
Aerofilms Ltd	Borehamwood, Herts, England
BKS Surveys Ltd	Coleraine, Co. Londonderry, N. Ireland
Cartographical Services Ltd	Salisbury, Wilts, England
Central Register of Aerial Photography of N. Ireland (Ordnance Survey)	Belfast, N. Ireland
Central Register of Aerial Photography of Wales	Cardiff, Wales
ERSAC Ltd	Livingston, W. Lothian, Scotland
Scottish Development Department	Edinburgh, Scotland
J. A. Story and Partners	Mitcham, Surrey, England
Ordnance Survey and Overseas Survey Directorate	Southampton, Hants, England
Royal Air Force Film Library	Whitehall, London
University of Cambridge Committee for Aerial Photography	Cambridge, England

References

1 Thompson, E. H. (1954) 'Heights from parallax measurements.' *Photog. Rec.* 1, 4, 38–49.

2 Methley, B. D. F. (1970) 'Heights from parallax bar and computer.' *Photog. Rec.* 6, 35, 459–465.

3 Royal Institution of Chartered Surveyors (1984) *Specification for vertical aerial photography*, 2nd edn. The British Air Survey Association and the Land Surveyor's Division of the RICS, 13pp. Surveyors Publications, London.

4 Thompson, E. H. (1964) 'Aerial triangulation by independent models.' *Photogrammetria*, 19, 7, 262–274.

5 Schut, G. H. (1980) 'Block adjustment by bundles.' *'Canadian Surv.*, 34, 2, 139–151.

6 Makarovic, B. (1973) 'Progressive sampling for digital terrain models.' *ITC J.* 397–416.

7 Ebner, H. (1980) Hoffman-Wellenhof, B., Reiss, P. and Steidler, F. 'HIFI–a minicomputer program package for height interpolation by finite elements.' *14th Congress of the International Society of Photogrammetry*, Hamburg, Commission IV, 14pp.

8 Kosli, A. and Wild, E. (1984) 'A digital terrain model featuring varying grid size.' *Contributions to the XVth ISPRS Congress*, Rio de Janeiro, Institute of Photography, University of Stuttgart, 10, 117–126.

9 Petrie, G. and Kennie, T. J. M. (1986) 'Terrain modelling in surveying and civil engineering.' *Proceedings, British Computer Society Displays Group meeting on state of the art in stereo and terrain modelling*, London, 32pp.; *Computer Aided Design*, 19, 4, 171–188.

10 Fraser, C. S. (1983) 'Photogrammetric monitoring of Turtle mountain – a feasibility study.' *Photog. Engr. and Remote Sensing*, 49, 11, 1551–1559.

11 Fraser, C. S. and Gruendig, L. (1985) 'The analysis of photogrammetric deformation measurements on Turtle mountain.' *Photog. Engr. and Remote Sensing*, 51, 2, 207–216.

12 Stewart, P. A. E. (1979) 'X-ray photogrammetry of gas turbine engines at Rolls-Royce', *Photog. Rec.*, 9, 54, 813–821.

13 Atkinson, K. B. (1976) 'A review of close-range engineering photogrammetry'. *Photog. Eng. and Remote Sensing*, 42, 1, 57–69.

14 Cheffins, O. W. and Chisholm, N. W. T. (1980) 'Engineering and industrial photogrammetry.' In: K. B. Atkinson (ed.) *'Developments in close-range photogrammetry*, vol. I. Elsevier Applied Science, London, pp. 149–180.

15 Cooper, M. A. R. (1984) 'Deformation measurement by photogrammetry.' *Photog. Rec.*, 11, 63, 291–301.

16 Welsh, N. (1986) 'Photogrammetry in engineering.' *Photog. Rec.*, 12, 67, 25–44.

17 Wong, K. W. and Wei Hsin Ho. (1986) 'Close-range mapping with a solid state camera.' *Photog. Engr. and Remote Sensing*, 52, 1, 67–74.

18 Moore, J. F. A. (1973) 'The photogrammetric measurement of the constructional displacements of a rockfill dam.' *Photog. Rec.*, 7, 42, 628–648.

19 Brandenberger, A. J. (1974) 'Deformation measurements of dower dams.' *Photog. Engr.*, 40, 9, 1051–1058.

20 Veress, S. A., Jackson, N. C. and Hatsopoulos, J. N. (1980) 'Monitoring a gabion wall by inclinometer and photogrammetry.' *Photog. Engr. and Remote Sensing*, 46, 6, 771–778.

21 Cheffins, O. W. and Rushton, J. E. M. (1970) 'Edinburgh Castle Rock, a survey of the north face by terrestrial photogrammetry.' *Photog. Rec.*, 8, 46, 417–433.

22 Moore, J. F. A. (1974) 'Major mapping joints in the lower Oxford Clay using terrestrial photogrammetry.' *Q. J. Engrg. Geol.*, 7, 57–67.

23 Torlegard, A. K. I. and Dauphin, E. L. (1975) 'Deformation measurement by photogrammetry in cut-and-fill mining.' *Proceedings of the ASP Symposium on close-range photographic systems*, University of Illinois, pp. 24–39.

24 Robertson, G. R., MacRae, A. M. R., Tribe, J., Sibley, D. W. and Smith, D. H. (1982) 'Use of photogrammetric methods for mine slope deformation surveys.' *4th Canadian Symposium on Mining and Deformation Monitoring*, Toronto.

25 Heath, W., Parlsey, L. L. and Dowling, J. W. F. (1978) 'Terrestrial photogrammetric surveys of unstable terrain in Columbia.' Transport and Road Research Laboratory, Publication No. LR876, TRRL, London.

26 Kennie, T. J. M. and McKay, W. M. (1986) 'Monitoring of geotechnical processes by close-range photogrammetry.' *Proceedings, 22nd Conference of the Engineering Group of the Geological Society*, Plymouth Polytechnic.

27 Anderson, H. and Stevens, D. (1982) 'Mono photogrammetric tunnel profiling.' *International Archives of Photographers' Commission 5 Symposium, York, Precision and speed in close-range photogrammetry*, 24, 1, 23–30.

28 Andrawes, K. Z. and Butterfield, R. (1973) 'The measurement of planar displacements of sand grains', *Géotechnique*, 23, 4, 571–576.

29 Wong, K. W. and Vonderohe, A. P. (1981) 'Planar displacement by motion parallax.' *Photog. Engr. and Remote Sensing*, 47, 6, 769–777.

30 Davidson, J. L. (1985) 'Stereophotogrammetry in geotechnical engineering.' *Photog. Engr. and Remote Sensing*, 51, 10, 1589–1596.

31 Matthews, M. C. and Clayton, C. R. I. (1984) 'The use of oblique aerial photography to assess the extent and sequence of landslipping at Stag Hill, Guildford.' *Proceedings 20th Regional Conference of the Engineering Group of the Geological Society of London*, Guildford, vol I, pp. 319–330.

32 Heath, W. (1980) *Inexpensive aerial photography for highway engineering and traffic studies*. Transport and Road Research Laboratory Supplementary Report No. 632, TRRL, London, 24pp.

33 Beaumont, T. E. (1977) *Techniques for interpretation of remote sensing imagery for highway engineering purposes*. Transport and Road Research Laboratory Report No. 753, 24pp.

34 Graham, R. W. and Read, R. (1984) 'Small format aerial photography from microlight platforms.' *J. Photog. Sc.*, 32, 100–109.

35 Tomlins, G. F. (1983) 'Some considerations in the design of low cost remotely piloted aircraft for civil remote sensing applications.' *Can. Surveyor*, 37, 157–167.

36 Meisner, D. E. and Lindstrom, O. M. (1985) 'Design and operation of a colour infra-red aerial video system.' *Photog. Engr. and Remote Sensing*, 51, 5, 555–560.

37 Kennie, T. J. M., Dale, C. D. and Stove, G. C. (1986) 'A preliminary assessment of an airborne thermal video frame scanning systems for environmental engineering surveys.' ISPRS, Commission VIII. *Proceedings International Symposium on remote sensing for resources development and environmental management*, Enschede, Balkema Press, pp. 215–221.

38 Stove, G. C., Kennie, T. J. M. and Harrison, L. (1987) 'Airborne thermal mapping for winter highway maintenance using the Barr and Stroud IR18 thermal video frame scanner.' *Int. J. Remote Sensing*

39 Rudd, R. O. (1974) *Remote sensing–a better view*. Duxbury Press, Massachusetts.

40 NOAA *'Landsat data users' notes*. Issues 1–36.

41 Chevrel, M., Courtois, M. and Wells, G. (1981) 'The SPOT satellite remote sensing mission.' *Photog. Engrg and Remote Sensing*, 47, 1163–1171.

42 Fearns, D. C. (1984) 'Microcomputer systems for satellite image processing.' *Earth Orient. Appl. Space Technol.*, 4, 4, 247–254.

43 Bernstein, R. (1978) *Digital image processing for remote sensing*. IEEE Press, New York, 473pp.

44 Avery, T. E. and Graydon, L. B. (1985) 'Digital image processing.' In: *Interpretation of aerial photographs*, 4th edn, Chapter 15, Burgess Publishing Co., Minneapolis, pp. 451–536.

45 Bagot, K. H. (1985) 'Digital processing of remote sensing data.' In: T. J. M. Kennie and M. C. Matthews (eds) *Remote sensing in civil engineering*, Surrey University Press, London, pp. 87–105.

46 Beaumont, T. E. (1982) 'Land capability studies from Landsat satellite data for rural road planning in North East Somalia.' *Proceedings, OECD symposium on Terrain evaluation and remote sensing for highway engineering in developing countries*. Transport and Road Research Laboratory Report No. SR690, pp. 86–95.

47 Overseas Unit (1982) *Terrain evaluation and remote sensing for highway engineering in developing countries*. Transport and Road Research Laboratory Supplementary Report No. 690, 172pp.

48 Blyth, K. (1985) 'Remote sensing and water resources engineering.' In: T. J. M. Kennie and M. C. Matthews (eds), *Remote sensing in civil engineering*, Surrey University Press, London, pp. 289–334.

49 Forstner, G. (1983) 'Some urban measurements from Landsat data.' *Photog. Engr and remote sensing*, 49, 12, 1693–1707.

50 Beaumont, T. E. (1979) 'Remote sensing for location and mapping of engineering construction materials in developing countries.' *Q. J. Engrg Geology*, 12, 3, 147–158.

51 Lawrance, C. J. and Beaven, P. J. (1985) 'Remote sensing for highway engineering projects in developing countries.' In: T. J. M. Kennie and M. C. Matthews (eds), *Remote sensing in civil engineering*, Chapter 9, Surrey University Press, London, pp. 240–268.

52 Dumbleton, M. J. (1983) *Air photographs for investigating natural changes, past use and present condition of engineering sites*. Transport and Road Research Laboratory Report No. 1085, TRRL, London.

53 Matthews, M. C. (1985) 'Interpretation of aerial photography,' Chapter 8. In: T. J. M. Kennie and M. C. Matthews (eds) *Remote sensing in civil engineering*, Surrey University Press, London, pp. 204–239.

54 Mason, P. A. and Amos, E. L. (1985) 'Environmental engineering applications of thermal infrared imagery.' In: T. J. M. Kennie and M. C. Matthews (eds) *Remote sensing in civil engineering*, Chapter 10, Surrey University Press, London, pp. 269–288.

55 Norman, J. W., Leibowitz, T. H. and Fookes, P. G. (1975) 'Factors affecting the detection of slope instability with air photographs in an area near Sevenoaks, Kent.' *Q. J. Engrg Geol.*, 8, 3, 159–176

56 Liu, J. K. (1985) 'Remote sensing for identifying landslides and for landslide prediction-cases in Taiwan.' *International Conference on advanced technology for monitoring and processing global environmental data*, Remote Sensing Society, London, 8pp.

57 Chandler, P. B. (1975) 'Remote detection of transient thermal anomalies associated with the Portuguese Bend landslide.' *Bull. Assoc. Engrg Geol.*, 12, 3, 227–232.

58 Norman, J. W. and Watson, I. (1975) 'Detection of subsidence conditions by photogeology.' *Engrg Geol.*, 9, 359–381.

59 Coker, A. E., Marshall, R. and Thompson, N. S. (1969) 'Application of computer-processed multispectral data to the discrimination of land collapse (sinkhole) prone areas in Florida.' *Proceedings 6th International Symposium on remote sensing of the environment*, Michigan, vol. I, pp. 65–69.

60 Kennie, T. J. M. and Edmonds, C. N. (1986) 'The location of potential ground subsidence and collapse features in soluble carbonate rocks by remote sensing techniques.' *Proceedings, American Society for Testing and Materials International Symposium on geotechnical applications of remote sensing and remote data transmission*. ASTM Special Technical Publication, pp. 206.

61 Bullard, R. K. (1983) *Abandoned land in Thurrock: an application of remote sensing*. Working Paper No. 8, North East London Polytechnic, London.

62 Coulson, M. G. and Bridges, E. M. (1984) 'The remote sensing of contaminated land.' *Int. J. Remote Sensing*, 5, 4, 659–669.

63 Ellyet, C. D. and Fleming, A. W. (1974) 'Thermal infrared imagery of the burning mountain coal fire.' *Remote sensing of environment*, pp. 79–86.

64 Caiger, J. H. (1970) 'Aerial photographic interpretation of road construction materials in South Africa with special reference to its potential to influence route location in underdeveloped territories.' *Photogrammetria*, 25, 151.

65 Lynn, O. W. (1984) 'Surface material mapping in the English fenlands using airborne multispectral scanner data.' *Int. J. Remote Sensing*, 5, 4, 699–713.

66 Singhroy, V. and Barnett, P. (1984) 'Locating subsurface mineral aggregate deposits from airborne imagery.' A case study in southern Ontario, *International Symposium on remote sensing for exploration geology*. Colorado.

67 Singhroy, V. (1986) 'Case studies on the application of remote sensing data to geotechnical investigations in Ontario.' In: *Geotechnical applications of remote sensing and remote data transmission*. American Society for Testing and Materials Special Technical Publicaton No. 206.

68 Mountain, L. J. and Garner, J. B. (1981) 'Semi-automatic analysis of small-format photography for traffic control studies of complex intersections.' *Photogrammetric Rec.* 10, 331–342.

Bibliography

Atkinson, K. B. (ed.) (1980) *Developments in close-range photogrammetry*, Elsevier Applied Science, London, 222pp.

Burnside, C. D. (1979) *Mapping from aerial photographs*. Granada, London, 304pp.

Colwell, R. W. (ed.) (1983) *Manual of remote sensing*, vols I and II, 2440pp.

European Space Agency (1984) *Remote sensing applications in civil engineering*. ESA, Publication No. SP-216, 198pp.

Karara, H. M. (ed.) (1979) *Handbook of nontopographic photogrammetry*. American Society of Photogrammetry, 206pp.

Kennie, T. J. M. and Matthews, M. C. (eds) (1985) *Remote sensing in civil engineering*, Surrey University Press, London, 356pp.

Kilford, W. K. (1979) *Elementary air survey*, 4th edn, Pitman, 345pp.

Lillesand, T. M. and Kiefer, R. W. (1979) *Remote sensing and image interpretation*. Wiley, New York, 611pp.

Moffit, F. H. and Mikhail, E. (1980) *Photogrammetry*, 3rd edn., Harper and Row, New York, 648pp.

Slama, C. C. (ed.), (1980) *Manual of photogrammetry*, 4th edn., American Society of Photogrammetry, 1056pp.

Wolf, P. R. (1983) *Elements of photogrammetry*, 2nd edn., McGraw-Hill, New York, 628pp.

Monitoring of geotechnical processes by close range photogrammetry

T. J. M. Kennie & W. M. McKay

ABSTRACT: The role of close range photogrammetry (CRP) in engineering geology is described, the relative advantages and limitations of the technique being outlined. A comprehensive literature review reports on the use of CRP for *in situ* monitoring of earth and rockfill dams, retaining walls, rockfaces and landslides, channel and gully erosion and, also, laboratory based experiments. The historical development of analytical CRP and the types of project where such a technique would be appropriate are subsequently discussed. Two case histories illustrating the use of analytical CRP for monitoring a series of eroded slopes in southern England are reported. The paper also examines the benefits which occur when digital photogrammetric data of this type are integrated with suitable ground modelling software.

Introduction

Close range photogrammetry (CRP) has, for many years, proved to be both a versatile and efficient method of measurement for many problems in civil and mechanical engineering, although its use by planners has been much less extensive. Various areas of application have been identified and these have been reviewed by Atkinson (1976) and Cheffins & Chilsholm (1980). Not only has the ability of the photogrammetric method to provide accurate three-dimensional information about the shape of an object been exploited, but also its ability to monitor the stability or deformation of an object, such as an engineering structure (Erlandson & Veress 1975; Cooper & Shortis 1980; Cooper 1984).

CRP has also been applied to the solution of many problems which are of concern to planners and geotechnical engineers and this paper provides a review of the published literature in this field. Following this, the development of the analytical stereoplotter is discussed. The use of this equipment has increased considerably in recent years and several advantages accrue from its use which have a considerable influence on whether to consider CRP as a potential measurement system. Finally, the paper outlines a typical geotechnical project concerned with the monitoring of chalk rockfaces where CRP has been applied. The erosion of such surfaces can be a major safety hazard and thus of obvious concern to planners. Before remedial measures can be implemented it is essential that precise measurements of the rate and volume of erosion are obtained. Analytical CRP provides a cost-effective solution to this type of problem.

Close range photogrammetry and geotechnical engineering

The decision to apply photogrammetry to a particular problem depends on several factors including total cost, accuracy required, availability of equipment, and the expertise available. In many situations photogrammetry, and CRP in particular, can be a highly effective, and perhaps the only viable measurement technique. Thompson (1962), for example, considered that photogrammetry is an appropriate technique to use in the following circumstances:

(i) when the object to be measured is inaccessible or difficult to access;
(ii) when the object is not rigid and its instantaneous dimensions are required;
(iii) when it is not certain what measurements are required;
(iv) when the object is very small.

If photogrammetry is considered appropriate, it can offer several advantages since it is a non-contact method of measurement, data acquisition is fast, the photographs can be stored to provide a historical record or be used for remeasurement at a later date, and the technique can be optimized to provide appropriate accuracy levels. Drawbacks do, however, exist and some of the more significant include the delay in obtaining results (it cannot at present be considered a real time measurement technique), the possibility of mistakes at the photographic processing stage which can have disastrous consequences, and the need, generally, for expensive equipment and highly skilled personnel. Nevertheless, photogrammetry is a widely used technique and the following sections outline some of the applications which may be of interest to the engineering geologist or geotechnical engineer. Furthermore, as will become evident from the review, the technique is equally relevant to the measurement and monitoring of large features in the field (for example, rockfaces, landslides, retaining walls) and small features in the laboratory (for example, testing soil models).

Earth and rockfill dam monitoring

The use of close range photogrammetry for monitoring earth and rockfill dams has been discussed by Planicka & Nosek (1970) and Moore (1973). In the latter case the author describes the use of a Wild P30 phototheodolite for monitoring, during construction, the three dimensional displacements of the Llyn Brianne rockfill dam in mid-Wales. The predicted displacements were in the range of ± 0.5 to 1.0 m, and whilst high absolute positional accuracy was deemed to be important, it was considered that the definition of the direction of move-

From CULSHAW, M. G., BELL, F. G., CRIPPS, J. C. & O'HARA, M. (eds) 1987. *Planning and Engineering Geology*, Geological Society Engineering Geology Special Publication No. 4, pp. 163–70.

ment was of equal importance. By using a Wild A7 stereoplotter, measurements for differing levels of fill were taken for over 80 targets. The results indicated a range of displacements, from 0.1 to 0.6 m (with a standard error of ±0.05 m).

Bradenberger (1974) also discusses the use of photogrammetry as part of a dam deformation monitoring programme, in this case in Quebec, Canada. Data were acquired using a Wild P30 phototheodolite during the spring and autumn of 1972 from geodetically surveyed control points on the downstream side of the dam face. From photo-coordinates measured in a monocomparator, the ground coordinates of 17 common test points on the dam face were computed. The results indicated a downstream displacement of the centre section of the dam of the order of 2 to 3 cm.

Retaining wall monitoring

The use of a modified KA-2 aerial camera (focal length 610 mm, format 23 × 23 cm) to monitor the deflection of a gabion wall is described by Veress & Sun (1978), and Veress *et al.* (1980). The wall, situated near Seattle, Washington State, was over 400 m long and varied in height from 2 to 18 m. The gabions consisted of steel mesh rockfilled one metre cubes. Photographs were obtained from fixed control points up to 1000 m from the wall. Over 100 target points were observed on 11 occasions with an analytical plotter and hence ground coordinates computed. The authors suggest that close range photogrammetry used in conjuction with an internal monitoring system such as an inclinometer would be an ideal system for monitoring new structures. More recent work by Veress & Hatzopoulos (1981) indicates that a substantial improvement in accuracy can be achieved by incorporating aerial photogrammetric observations into measurement schemes.

Rockface mapping and landslide monitoring

One of the earliest reports of rockface mapping is given in Cheffins & Rushton (1970). This paper describes the procedures used to produce 1:50 scale contoured elevations of the north face of Edinburgh Castle Rock. The data were required by a team of engineering geologists who were carrying out rock bolting and grouting stabilization measures to reduce the chance of rock falls occuring from the face. The survey was carried out by taking four pairs of overlapping photographs from a baseline approximately parallel to the face. These photographs were obtained using a Wild RC5A aerial camera mounted on a hydraulic platform. The elevations with 0.25 m contours were plotted using a Zeiss Stereoplanigraph C8. They were subsequently used, in conjunction with ultrasonic data about the joint structure, to plan the positions of the necessary boreholes for the rock bolting work.

The effectiveness of photogrammetry as a method of providing data associated with the stability of rock slopes is demonstrated by Wickens & Barton (1971) who used a Wild P30 phototheodolite to obtain photographs of rock faces in open pit mines. Photogrammetric measurements provided the coordinates of three points on each exposed rock plane. From these measurements, the degree and direction of dip were determined and used to predict the stability of different parts of the slope. The authors also outlined a photogrammetric method to provide 'roughness profiles' of rock joints. It is suggested that data of this type may also be used to determine the stability condition of slopes.

Moore (1974) discusses the use of a phototheodolite for mapping vertical faces, in this case in clay pits. The problem under investigation involved the determination of information about the continuity and spacing of major joints in the face. Normally the joints were not visible on the top horizontal surface of the excavation unless a prolonged dry period occurred. By observing photographs for successive stages of excavation, a three dimensional model of the spatial disposition of the joints was constructed. Torlegard & Dauphin (1975) and Robertson *et al.* (1982) also discuss the application of photogrammetry to slope and rockface stability monitoring.

Landslides are a commonly occurring hazard in road design and construction. Heath *et al.* (1978) and Heath & Dowling (1980) describe the use of a Wild P32 camera to obtain photographs of areas susceptible to landslipping in Columbia and Nepal. The photographs were observed using a Zeiss Topocart and Wild A40 terrestrial stereoplotter, and 1:200 to 1:10 000 scale contoured plots were produced. The plans were subsequently used to design remedial and preventative measures associated with the landslip areas and a classification of landslide characteristics was also produced.

A low cost photogrammetric technique for measuring slope profiles for use in the early stages of an engineering investigation, is described by Wooding (1978). In this case a Hasselblad SW camera which was modified to incorporate a reseau plate in the film plane was used. This equipment was utilized during an engineering geomorphology study of slope instability in Columbia. The results were subsequently compared with those provided by the more expensive system described above and although a reduced order of accuracy was apparent, the results were considered adequate for preliminary design and investigation purposes.

Channel and gully erosion monitoring

Channel and gully erosion is a problem of major concern in many parts of the world. Photogrammetric investigations which aim to quantify the amount of erosion have been performed by Collins & Moon (1979) and Welch & Jordan (1983). In the latter case, by the use of a non-metric camera in conjuction with analytical techniques, the amount of stream channel erosion which occurred during a storm was determined. Stereoscopic photographs were obtained from a fixed platform using a Pentax Spotmatic camera and image coordinates were measured using a cartographic digitizer. The results indicated accuracies of the order of 3 to 6 mm.

Tunnel profiling

An early application of photogrammetric methods to recording tunnel measurements is described in Wright *et al.* (1969). More recent developments associated with the use of photogrammetry for profiling tunnels are described by Opteka (1975) and Anderson & Stevens (1982). The authors of the second paper outline the development and use of a 'mono-photographic' tunnel profile measuring system. The two main elements of the system are a high intensity light plane generator which illuminates the section to be profiled, and a camera which records the line of light generated. By digitizing the line, applying corrections for lens distortion, and scaling the photographs an accuracy of ± 10 mm is claimed. Furthermore, if the quoted progress rate of 100 sections per hour can be achieved in practice, the system would appear to be highly cost effective.

Laboratory based studies

Close range photogrammetry has also been applied to solve measurement problems in the laboratory. For example Butterfield *et al.* (1970), Andrawes & Butterfield (1973), Wong & Vonderhoe (1981) and Davidson (1985) used the motion or 'false parallax' technique to measure the planar displacement fields associated with soil models. In the first two cases cited, the technique involved taking from a single camera position repeat photographs of a glass sided tank which contained the soil under investigation. Photographs were taken using a non-metric 35 mm camera before and after the sand in the tank was subjected to displacement by a moving wall or wedge. By examining enlargements of the photographs in an analogue stereoplotter, the relative heights of features could be measured, and hence displacement contour diagrams plotted. From the analysis of these diagrams the displacement component along the sand bed was determined to within 5 μm. Wong & Vonderhoe (1981), in contrast, used a Wild STK-1 stereocomparator to measure the photographs and determine analytically the derived movement vectors of sand particles around a model tunnel. Davidson (1985) also used the technique, in this case to investigate soil disturbance during penetration testing in sand. Two different penetration probes, the cone penetrometer and the flat plate dilatometer, were used. Each probe was cut longitudinally in half and pushed into a glass walled container of sand. By obtaining photographs before and after penetration, the induced false parallax could be measured. Observations were recorded at grid points scored onto the glass wall using an optical projection stereoplotter. The results of the tests were presented in the form of displacement vector plots together with contoured plots of volumetric and shear strain.

A further soil mechanics example is presented in El-Beik (1973) who outlines the development of a system which utilizes a pair of Zeiss (Jena) UMK 10/1318 cameras for photographing soil models in a heavy centrifuge. Stereophotographs of soil models subjected to steadily increasing gravitational forces were recorded and measured using a modified Kelsh stereoplotter. Spot height measurements of between 9 and 20 points enabled settlement/heave curves to be constructed. From these graphs the nature of the failure of the soil models could be determined and they were used to predict the behaviour of the soil in practice.

Development of analytical close range photogrammetric techniques

In the past surveying organizations have used conventional analogue stereoplotting equipment, designed principally for aerial mapping purposes, to compile plans for architectural, engineering and other close-range applications. The main drawback of this type of equipment lies in the constraints placed on such elements as focal length, lens distortion, camera orientation, film format size and depth variation. These optical and mechanical limitations to orienting stereomodels are well known to photogrammetrists. For example, the maximum allowable tilt movement in an analogue instrument normally restricts the attitude of the camera to a maximum of ± 5–$6°$. Orientating a stereomodel under these extreme conditions can be an arduous task which may take considerable time. This could be more efficiently used measuring the feature of interest.

With the advent of digital computers, especially mini- and micro-computers together with electronic encoders for digitizing the output from stereoplotting equipment (Petrie 1980), a new generation of photogrammetric stereoplotter has emerged; the analytical stereoplotter (Helava 1958, 1980). With this type of instrument a mathematical model replaces the space rods and other mechanical linkages which are found in more conventional analogue stereoplotters. The Officine Galileo Digicart is an example of an instrument of this type.

Officine Galileo Digicart analytical stereoplotter

The Officine Galileo Digicart is a well established analytical stereoplotter which has been in use in the United Kingdom for the past two years (McKay *et al.* 1985). The instrument consists of a stereoviewing system, linear encoders which digitally record the position of the carriages on which the photographs are mounted, and a minicomputer to process the data. Unlike a conventional analogue instrument where manual elimination of y-parallax is performed, with this design any parallax existing within the model is eliminated by a feedback loop which ensures that corrections are transmitted to the optics in real time (50 times a second) so that a model free from parallax is provided. Figure 1 illustrates the principle of operation.

Practical benefits of analytical techniques in close range photogrammetry

The application of analytical techniques has had a considerable influence on the practical application of CRP to engineering problems. In particular, analytical techniques offer the following advantages over conventio-

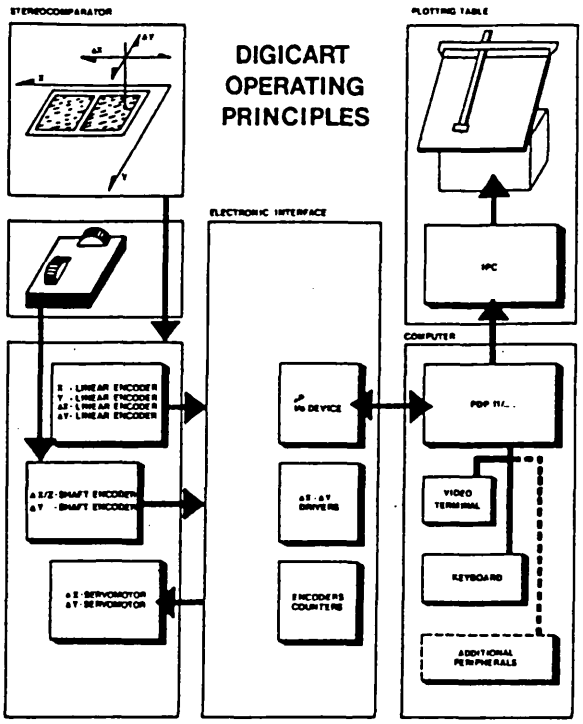


FIG. 1. Operating principles of a Digicart analytical stereoplotter.

nal analogue techniques:

(i) Higher accuracy than was previously possible (up to ± 3 μm at photoscale with suitable care).

(ii) Photography can be at differing format sizes and focal lengths. In certain cases, provided suitable precautions are taken to eliminate any systematic errors such as lens distortion, non-metric cameras can be used.

(iii) As mentioned previously, the setting up of even the most difficult stereomodels presents little difficulty with most models taking no more than ten to fifteen minutes for an experienced operator, thus leaving more time for the actual measurements.

(iv) Photography can be performed in the field without too great a concern about the orientation of the camera. This can save a great deal of time (as much as 50% over the conventional approach). Where access is difficult or normal case geometry is impeded, oblique photography may be used. This would not be possible using traditional methods.

(v) Data are captured in ground coordinates and stored on magnetic disk. This allows for plotting off-line on graphical plotters at any scale. Also, with data being collected digitally rather than graphically, a great deal more flexibility exists, for example, the data can be manipulated on a Computer Aided Drafting and Design (CADD) system.

Monitoring of erosion: case studies

The facility for monitoring rates of erosion and the production of rockface maps of eroded slopes are two capabilities of direct relevance to planners involved in remedial measures to combat such hazards. The following two case histories indicate how CRP can be used to provide relevant information. The first example refers to work performed in the Brighton area and the second to a problem in the Lewes area of East Sussex.

Brighton region: embankment monitoring

In this case, the client had identified sites in the Brighton area where erosion was progressing at a rate sufficient to give concern about public safety. A decision was made to adopt a two-dimensional approach to measuring the slope by observing a series of pre-determined section lines. The precise location of each section was identified on the ground and on the 1:1250 scale Ordnance Survey map sheets of the area. Points at the base and top of the embankment were monumented and surveyed using electronic distance measuring (EDM) instruments and theodolites in order to coordinate precisely the control. In cases where only one section was required per site, two additional points offset by 5 m at the base and top of the

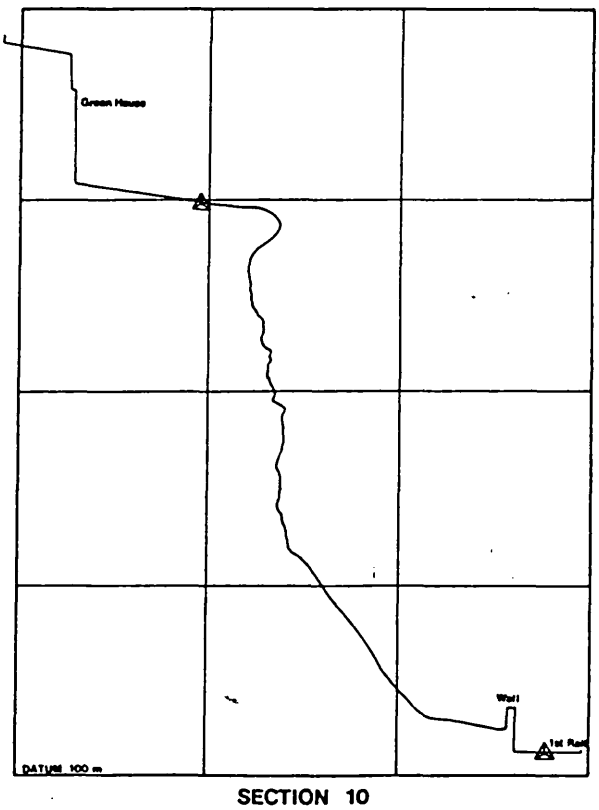


FIG. 2. Typical section through embankment at Brighton.

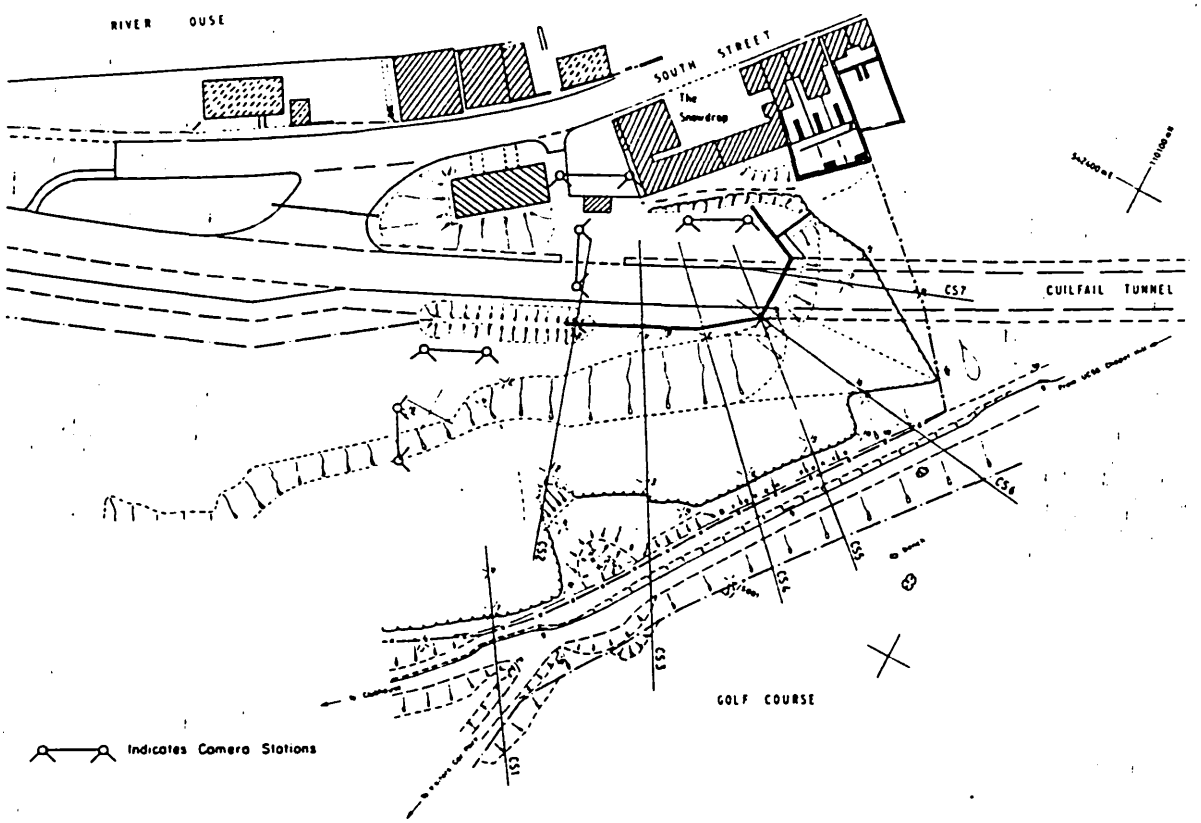


FIG. 3. Survey location map for Lewes road tunnel.

embankment were also surveyed to give full control in the stereomodel. A Zeiss (Jena) UMK 10/1318 camera (focal length = 100 mm) was used to obtain the photography during this project.

The stereomodels were orientated using a minimum of four control points per model, but in some cases up to eight points were used. The photoscale varied throughout the stereomodel from a minimum of 1:250 at the base of the embankment to a maximum of 1:600. The maximum error on the absolute orientation of the control was found to be 25 mm in three dimensions and the final accuracy of the sections was of the order of ±4 cm. Field verification of these results proved difficult due to the site conditions, but a photogrammetric repeat of exercise indicated that the level of completeness was high. The results were presented as 1:100 scale plots and Fig. 2 illustrates a typical profile.

This two-dimensional technique is a very practical method of carrying out a monitoring problem since the results from one year can easily be overlayed on the results for subsequent years giving a direct indication of the variations. The disadvantage of this approach is that it only measures variations along the selected sections. Thus, the location of these lines must be chosen very carefully if the maximum or average amount of erosion is to be determined. A further disadvantage is that it is only possible to measure accurately down slope movements; the volume determination is less precise unless many closely spaced sections have been measured.

Lewes area: road tunnel

The second case involved the monitoring of the shape and erosion occurring around the entrance of a road tunnel. The southern approach was cut into chalk on the eastern side and consisted of four separate and distinct faces varying in height from thirty to sixty meters. A mesh fence had been erected as a temporary measure to stop rock falls from inpinging on the road surface.

Control stations were monumented at both the top and base of the cliffs, ensuring that a minimum of six control points were visible within each stereomodel. As with the previous project the stations were surveyed by EDM and theodolite and targets positioned over the monuments for precise viewing and control pointing in the stereo model. The control network was tied into the National Grid and all height values related to Ordnance datum. Photography was performed using a Zeiss (Jena) UMK 10/1318 camera with glass negatives to provide greater stability. Figure 3 illustrates the site as well as giving the locations of the control and photographic stations.

The camera was positioned in front of the cliff faces by



FIG. 4. Photography of rock face in chalk at Lewes obtained using a Zeiss (Jena) UMK 10/1318 camera.

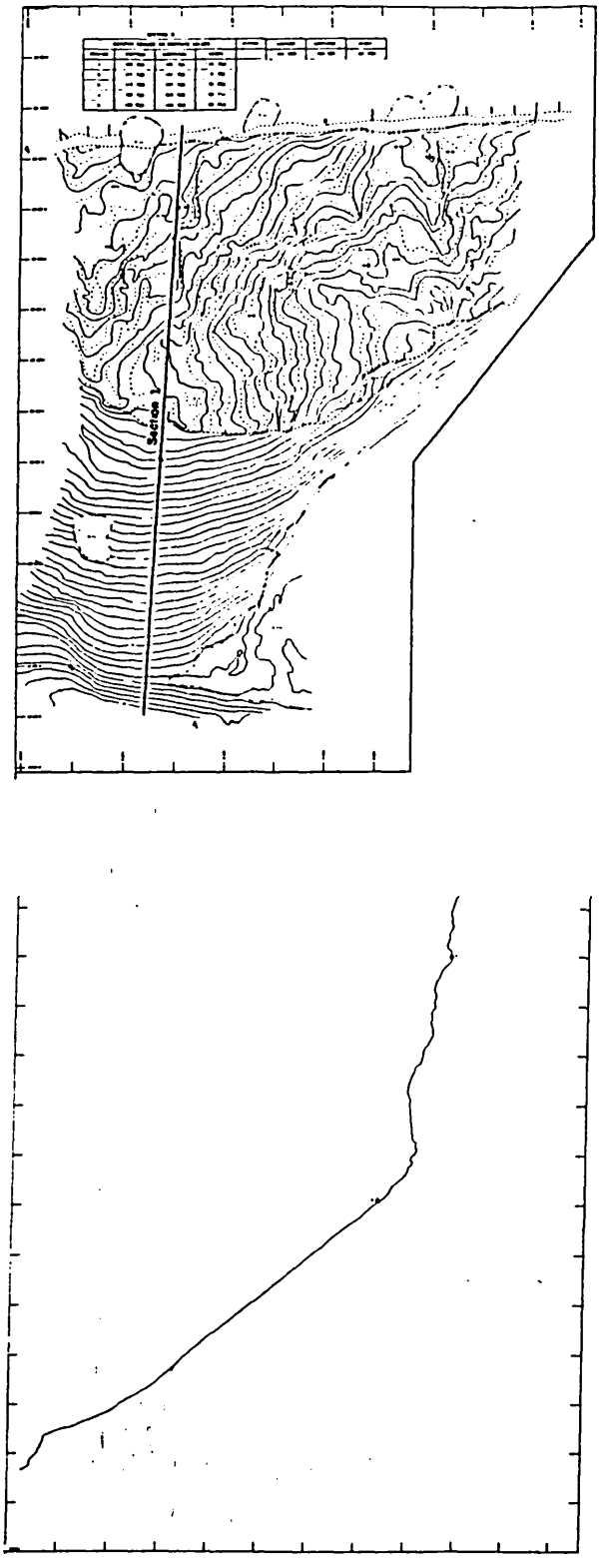


FIG. 5. Study of rock face at Lewes. (a) Typical contour plot (b) typical section.

the use of a hoist. Again, due to the variable distance of the cliff face from the camera a wide variation in photoscale from 1:400 to 1:900 resulted. Figure 4 is a typical photograph obtained with the Zeiss UMK camera.

For each of the four faces, the control network was rotated about two points on the cliff top thus forming a datum line parallel to the face itself. The shape of the cliff was then defined by contours at 0.25 m intervals of depth — that is, as differences of horizontal distance from this datum plane. This contrasts with a conventional survey of vertical elevations above a height datum. The contours produced by this process were digitized directly using the Digicart and draughted off-line on a Kongsberg flat-bed plotter at a scale of 1:100. A typical contour plot and section are illustrated in Figure 5.

The resultant plans were forwarded to the client and eight critical areas were selected for detailed measurement. In order to achieve this, the stereomodels were reintroduced into the analytical plotter and, with the aid of a program forming part of the basic software of the system, sections were digitized and subsequently plotted at a scale of 1:100. The accuracies achieved were found to be of the same order as for the embankment at Brighton, both the mapping and sections were within ± 5 cm of true ground position.

The preliminary contouring of the cliff faces provides a comprehensive view of the surface before the selection of the section positions for monitoring. The method used at Lewes, however, must also be considered as two-dimensional because future measurements are along fixed sections. An alternative approach to measuring erosion, and one which could have been applied to either of the previous cases, would involve the use of a three-dimensional terrain modelling package such as MOSS or ECLIPSE (Petrie & Kennie 1987). In order to apply this technique an array of heights (either in a grid, random locations of contour format) would have been formed into a terrain model of the slope. The cliff face requiring monitoring would therefore be stored as a series of x, y and z coordinates from which cross-sections and profiles could be extracted at user defined positions. Area, volume and other computed data could then be generated from this data base. Thus the results from successive years could be compared and the volume of rock erosion calculated. If slippage occurred in an unpredicted area, the terrain model together with the photographic record could be referred to, and precise measurements made.

Conclusions

Close range photogrammetry is a versatile measurement technique which has been used to solve a variety of problems of concern to planners and geotechnical engineers. It offers several advantages over direct surveying techniques, particularly, when the feature of interest is largely inaccessible. Furthermore, the ability to store a historical record of the site conditions in the form of a photographic negative is also an important feature of the technique. Recent developments in instrument design

now provide a computer based analytical solution to many measurement problems. Adopting a digitally based technique offers many advantages to a prospective user including the ability to integrate such data with plotting (e.g. CADD) and terrain modelling software.

ACKNOWLEDGEMENTS: The authors wish to thank the Department of Highways, East Sussex County Council and the Chief Engineer, British Rail Southern Region for their cooperation and access to source data. They also wish to thank Mrs V. Brown for her valuable comments on the paper.

References

ANDERSON, H. & STEVENS, D. 1982 Mono photographic tunnel profiling. *Fifth Symposium of the International Archives of Photogrammetry Commission, York, Precision and Speed in Close Range Photogrammetry*, **24**, 1, 23–30.

ANDRAWES, K. Z. & BUTTERFIELD, R. 1973 The measurement of planar displacements of sand grains. *Geotechnique*, **23**, 571–6.

ATKINSON, K. B. 1976. A review of close range engineering photogrammetry. *Photogrammetric Engineering and Remote Sensing*, **42**, 57–69.

BRANDENBERGER, A. J. 1974 Deformation measurements of power dams. *Photogrammetric Engineering*, **40**, 1051–8.

BUTTERFIELD, R., HARKNESS, R. M. & ANDRAWES, N. Z. 1970 Stereophotogrammetric method for measuring displacement fields. *Geotechnique*, **20**, 308–4.

CHEFFINS, O. W. & CHISHOLM, N. W. T. 1980. Engineering and Industrial Photogrammetry. *In*: ATKINSON, K. B. (ed.) *Developments in Close Range Photogrammetry*, **1**, 149–80, Applied Science.

CHEFFINS, O. W. & RUSHTON, J. E. M. 1970. Edinburgh Castle Rock, a survey of the north face by terrestrial photogrammetry. *Photogrammetric Record*, **8**, 46, 417–33.

COLLINS, S. H. & MOON, G. C. 1979. Stereometric measurement of stream bank erosion. *Photogrammetric Engineering and Remote Sensing*, **45**, 183–90.

COOPER, M. A. R. & SHORTIS, M. R. 1980. Analytical photogrammetry applied to the measurement of large structures. *Proceedings of the 14th Congress of the International Society of Photogrammetry, Hamburg*, Commission V, 137–48.

COOPER, M. A. R. 1984. Deformation measurement by photogrammetry. *Photogrammetric Record*, **11**, 63, 291–301.

DAVIDSON, J. L. 1985. Stereophotogrammetry in geotechnical engineering research. *Photogrammetric Engineering and Remote Sensing*, **51**, 1589–96.

EL-BEIK, A. H. A. 1973. Photogrammetry in centrifugal testing of soil models. *Photogrammetric Record*, **7**, 41, 538–54.

ERLANDSON, J. P. & VERESS, S. A. 1975. Monitoring deformations of structures. *Photogrammetric Engineering and Remote Sensing*, **41**, 1375–84.

HEATH, W. & DOWLING, J. W. F. 1980. Examples of the use of terrestrial photogrammetry in highway engineering. *Proceedings of the 14th Congress of the International Society of Photogrammetry, Hamburg*, Commission V, 308–20.

HEATH, W., PARSLEY, L. L. & DOWLING, J. W. F. 1978. Terrestrial photogrammetric surveys of unstable terrain in Columbia. *Transport and Road Research Laboratory Report*, LR 876.

HELAVA, U. V. 1958. New principles for photogrammetric plotters. *Photogrammetria*, **14**, 49–56.

——. 1980. Analytical plotter basics. *Analytical Plotter Symposium, Reston*, American Society of Photogrammetry, 12–30.

MCKAY, W. M., DOWMAN, I. J. & FARROW, J. E. 1985. Experience with analytical plotters. *Photogrammetric Record*, **11**, 66, 729–45.

MOORE, J. F. A. 1973. The photogrammetric measurement of the constructional displacements of a rockfill dam. *Photogrammetric Record*, **7**, 42, 628–48.

——. 1974. Major mapping of joints in the lower Oxford clay using terrestrial photogrammetry. *Quarterly Journal of Engineering Geology*, **7**, 57–67.

OPTEKA, G. 1975. Report on the use of the Wild P32 terrestrial camera in the construction of the Vienna underground. *Wild Reporter*, **8**, 14–16, Wild Heerbrugg.

PETRIE, G. 1980. Hardware aspects of digital mapping. *Proceedings of the Congress of the International Society of Photogrammetry*, Hamburg, Commission IV.

PETRIE, G. & KENNIE, T. J. M. 1987. Terrain modelling in surveying and civil engineering. *Computer Aided Design*, **19**, 171–88.

PLANICKA, A. & NOSEK, L. 1970. Terrestrial photogrammetry in measurement of deformations of rockfill dams. *Proceedings of the 10th International Conference on Large Dams, Montreal*, 207–15.

ROBERTSON, G. R., MCRAE, A. M. R., TRIBE, J., SIBLEY, D. W. & SMITH, D. H. 1982. Use of photogrammetric methods for mine slope deformation surveys. *Proceedings of the 4th Canadian Symposium on Mining and Deformation Monitoring.*

THOMPSON, E. H. 1962. Non-topographic photogrammetry. *Photogrammetric Record*, **5**, 27, 147–9.

TORLEGARD, A. K. I. & DAUPHIN, E. L. 1975. Deformation measurement by photogrammetry in cut and fill mining. *Proceedings of the ASP Symposium on Close Range Photographic Systems, University of Illinois*, 24–39.

VERESS, S. A. & SUN, L. 1978. Photogrammetric monitoring of a gabion wall. *Photogrammetric Engineering and Remote Sensing*, **44**, 205–11.

VERESS, S. A. & HATZOPOULOS, J. N. 1981. A combination of aerial and terrestrial photogrammetry for monitoring. *Photogrammetric Engineering and Remote Sensing*, **47**, 1725–31.

VERESS, S. A., JACKSON, N. C. & HATZOPOULOS, J. N. 1980. Monitoring a gabion wall by inclinometer and photogrammetry. *Photogrammetric Engineering and Remote Sensing*, **49**, 771–8.

WELCH, R. & JORDAN, T. 1983. Analytical non-metric close range photogrammetry for monitoring stream channel erosion. *Photogrammetric Engineering and Remote Sensing*, **49**, 367–74.

WICKENS, E. H. & BARTON, N. E. 1971. The application of photogrammetry to the stability of excavated rock slopes. *Photogrammetric Record*, **7**, 7, 46–54.

WONG, K. W. & VONDERHOE, A. P. 1981. Planar displacement by motion parallax. *Photogrammetric Engineering and Remote Sensing*, **47**, 769–77.

WOODING, M. G. 1978. Engineering geomorphological analysis of slope instability. *Proceedings of the 3rd Symposium on Geotechniques and Engineering Geology*, Bogata, Columbia.

WRIGHT, D. E., COX, D. E. & CHEFFINS, O. W. 1969. Photogrammetric measurement of rock surfaces in a power tunnel. *Water Power*, 230–4 and 274–9.

T. J. M. KENNIE, Department of Civil Engineering, University of Surrey, Guildford GU2 5XH, UK.
W. M. MCKAY, Automated Mapping and Cartography, Fordbrook Business Centre, Marlborough Road, Pewsey, SN9 5NU, UK.

Ground Engineer's Reference Book

Edited by
F G Bell

With specialist contributors

Butterworths

London · Boston · Durban · Singapore ·
Sydney · Toronto · Wellington

19.1 Introduction

The work of the land surveyor can be classified into three main areas of responsibility: firstly, the recording of measurements which allow the size and shape of the Earth to be determined; secondly, and primarily, the collection, processing and presentation of the information necessary to produce maps and plans; thirdly, it may be required to locate on the surface of the Earth the exact positions to be taken up by new roads, dams or other civil engineering works. As a consequence of the diverse nature of the land surveyor's duties, several distinct branches of the subject have evolved.

19.1.1 Branches of surveying

Geodetic surveys are carried out on a national or international basis to locate points large distances apart. This type of survey acts as a framework for 'lower order' surveys. In order to ensure high accuracy, the effect of factors such as the curvature of the Earth on observations must be considered and the necessary corrections applied.

Topographic surveys are concerned with the small-scale representation of the physical features of the Earth's surface. Frequently the data necessary for such an operation will be provided by the use of aerial photography. The science of taking measurements from photography to produce maps is known as photogrammetry. Topographic surveys are often the responsibility of a national organization such as, for example, the Ordnance Survey in the United Kingdom.

Hydrographic surveys, by contrast, involve the representation of the surface of the sea floor. The end product is normally a navigational chart. In recent years this branch has become increasingly important with the development of the offshore oil industry. In this case, in addition to the production of charts, the surveyor may be required to precisely position a large structure such as an oil production platform. This would normally necessitate the use of sophisticated electronic position fixing equipment.

Cadastral surveys relate to the location and fixing of land boundaries. In many countries in the world, for example, Australia, the information supplied by the cadastral surveyor may be an integral part of a land registration system.

Finally, engineering surveys are required for the preparation of design drawings relating to civil engineering works such as roads, dams or airports. The surveys are normally at a large scale, with scales of 1:500 and 1:1000 being most common.

Many of these branches of the subject, however, require highly specialized knowledge, beyond the scope of this chapter. It is therefore the aim of this chapter to discuss those aspects of the subject which are required in order to carry out simple engineering surveys.

19.1.2 Principles of surveying

In spite of the diverse nature of land surveying, it is possible to define certain basic principles which are common to all branches of the subject. These principles have proved over the years to be vital if accurate surveys are to be conducted.

The first and most important principle is the provision of an initial framework before observing and fixing the detail of a survey. This process is often known as providing control. It is essential to ensure that the positions of the control points are known to a higher order of accuracy than those of the subsidiary points. By satisfying this principle it is possible to ensure that errors, which inevitably occur, do not accumulate but are contained within the control framework.

A second and perhaps more obvious principle is that of planning. All too often it is tempting to rush into a survey without consideration for an overall scheme. For example, the definition of a specification is indispensable before a survey begins. The relationship between cost and accuracy is not linear and an increase in accuracy may have a disproportionate effect on cost. For instance, if a distance of 500 m is to be determined to an accuracy of either 5 mm or 0.5 mm, the cost ratio of the respective accuracies may be of the order of 1:300. It is important, therefore, to choose techniques and instruments appropriate to the survey specification. Of equal importance is the planning of the reconnaissance stage. Before starting a task it is essential to examine the area carefully, considering all the possible ways of doing the survey and then selecting the most suitable method. Time spent on reconnaissance and planning is never wasted.

The third and final principle is to ensure that sufficient independent checks are incorporated into the survey to eliminate or minimize errors. It is important that a checking system is included at all stages of the survey from fieldwork and computations to the final plotting. In addition, the checking system should be independent and not solely a repeat of the initial measurement. Examples of independent checks are:

(1) Fieldwork:
   (a) Measure both diagonals of a quadrilateral.
   (b) Measure distances in both directions.
   (c) Measure angles using different parts of the theodolite circle.
(2) Computations:
   (a) Use the summation check on angle observations of geometric figures, for example, the sum of the interior angles of a closed polygon $= (2n-4) \times 90°$ where n represents the number of sides.
   (b) Levelling booking cross checks.
(3) Plotting:
   (a) Plot positions of important points by using angles and distance and also using coordinates.

19.1.3 Errors in surveying

It is an unfortunate and often misunderstood fact that all measurements are affected by errors. So often, confronted with the question 'How accurate do you want the survey to be?' or 'How accurately do you want this point located?', the glib answer 'Exactly!, or 'Spot on!' is given as the reply by a prospective client. If, in addition, the question of errors is raised, it is quickly dispensed with the comment 'Errors don't occur if you do it properly'. The answer is correct in one respect; that is, in relation to mistakes, or, more correctly, gross errors. These should not occur if a survey is carried out according to the basic principles of surveying. However, other types of errors do occur which can be much more difficult to handle.

Systematic errors, as the name suggests, are errors which follow a pattern or system. Errors of this type are normally related to the variations in physical conditions which can occur when a measurement is made. For example, a steel tape is normally known to be a certain length at some standard temperature. If the temperature under which a measurement is made varies from this standard, a systematic error will occur. By knowing the coefficient of linear expansion a value for the expansion of the tape can be determined and a correction applied. Systematic errors are therefore errors whose effect can be modelled mathematically and hence eliminated.

Random errors, by contrast, are not constant and are entirely based on the laws of probability. These errors, or rather variations in measurement, will occur after gross and systematic errors have been eliminated. The measurement of a distance by taping can again be taken as an example. It is often not appreciated that the same distance measurement made under the same physical conditions with the same tape will produce different answers. It is normally assumed that the results follow a

normal distribution pattern and hence can be examined using standard statistical techniques. The following formula can therefore be applied to the analysis of random errors:

$$\text{arithmetic mean} = \bar{x} = \frac{\sum x_i}{n} \tag{19.1}$$

where x_i ($i = 1, 2, \ldots, n$) are the observed values and n denotes the number of observed values.

The arithmetic mean is significant because it is often taken to be the closest approximation to the 'true' value and as such is known as the most probable value (m.p.v.). The difference between the m.p.v. and the observed value is known as a residual (v).

A term often used to estimate the precision of a series of measurements is the standard error, where

standard error of a single observation

$$= \sigma_s = \pm\left(\frac{\sum v^2}{n-1}\right)^{1/2} \tag{19.2}$$

standard error of the arithmetic mean

$$= \sigma_m = \frac{\pm\sigma_s}{n^{1/2}} \tag{19.3}$$

For surveying purposes the terms standard error, standard deviation and root mean square error are synonymous. All such terms are used to give an indication of the precision of the result; that is, the degree of agreement between successive measurements. High precision may not be indicative of high accuracy, since accuracy is related to the proximity of the measurement to the true value. If, however, all the effects of the bias caused by systematic errors have been eliminated, these indices of precision may also be used as indices of accuracy.

For example, suppose an angle has been measured nine times and the subsequent error analysis indicates that $\sigma_s = \pm 3''$ and $\sigma_m = \pm 0.81''$. What does this information tell us? Firstly, it indicates that the angle has been measured to a high precision. Secondly, it indicates that, statistically, there is a 68% chance or probability of the standard error of a single measurement being less than $\pm 3''$. However, if one extends the confidence limit to a value three times the standard error or $\pm 9''$, then statistically the probability that the error will be less than $\pm 9''$ is now 99.7% with only a 0.3% chance of the error being greater than $\pm 9''$. This confidence limit is often applied as a rejection criteria to a group of observations. Any observation with a residual greater than three times the standard error is then rejected on the basis that it is highly unlikely that the variation is solely a consequence of random effects. Similar reasoning would apply to the standard error of the arithmetic mean. Further information on errors in surveying can be found in Cooper[1].

19.2 Surveying instrumentation

Surveying is essentially concerned with the direct measurement of three fundamental quantities; the angle subtended at a point, the distance between two points and the height of a point above a datum, such as mean sea level. From the measurement of these three quantities, it is possible to compute the three-dimensional positions of points.

With the exception of electronic methods of determining distance, the instruments used by the surveyor have not radically changed in principle over the last 40 to 50 years. Advances in technology may have reduced the size and increased the efficiency of the instruments, but the fundamental principles remain unchanged.

19.2.1 Angular measurement using the theodolite

The theodolite is used for the measurement of horizontal and vertical angles. In simple terms a theodolite consists of a telescope mounted on a platform which may be levelled to form a horizontal plane by means of a simple spirit bubble. Angles are measured by pointing the telescope at targets and establishing the difference between readings on a circular protractor mounted on the level platform.

There is a bewildering choice of theodolites available. *Table 19.1* lists the characteristics of a selection of commonly available modern theodolites. A broad distinction can be made between those instruments which measure angles and those such as compass and gyro-theodolites which measure bearings, relative to magnetic north and to true north respectively.

There are certain fundamental relationships and components which are common to all theodolites. Before examining the detailed construction of a modern glass arc theodolite, it is important to appreciate the geometrical arrangement of the axes of a theodolite, as illustrated in *Figure 19.1*.

In this ideal arrangement the vertical axis is vertical, the trunnion axis is perpendicular to it and hence horizontal, and the line of collimation is perpendicular to the trunnion axis. Unfortunately, it is not possible during the manufacturing process to ensure that these orthogonal relationships occur exactly. Similarly, during use over a period of years, wear may occur which may also alter these conditions. The extent to which a theodolite fails to satisfy them can be measured by a series of instrument tests which may be carried out in the field. If, subsequently, the instrument is found to be out of adjustment, the instrument should be returned to the manufacturer or a specialist instrument technician for adjustment. Details of the field tests and methods of adjustment are provided by Cooper[2]. If, however, a modern theodolite is treated with care, and a suitable observational technique is employed, regular servicing should be all that is required to obtain good results.

Table 19.1 Characteristics of some modern theodolites

Type of theodolite:	1" Precise	20" Engineers	10' Builders	Compass	Electronic
Typical example:	Kern DKM 2A-E	Sokkisha TM20ES	Zeiss (Ober.) TH51	Wild T0	Kern E-1
Country of manufacture:	Switzerland	Japan	W. Germany	Switzerland	Switzerland
Direct reading to	1"	20"	10'	1'	1"
By estimation to	0.1"	5"	1'	30"	—
Telescope magnification	32×	28×	20×	20×	32×
Telescope aperture (mm)	40	45	30	28	45
Sensitivity of plate level per 2 mm run	20"	30"	45"	8'	—
Weight of instrument (kg)	6.2	4.2	2.2	2.9	8.7

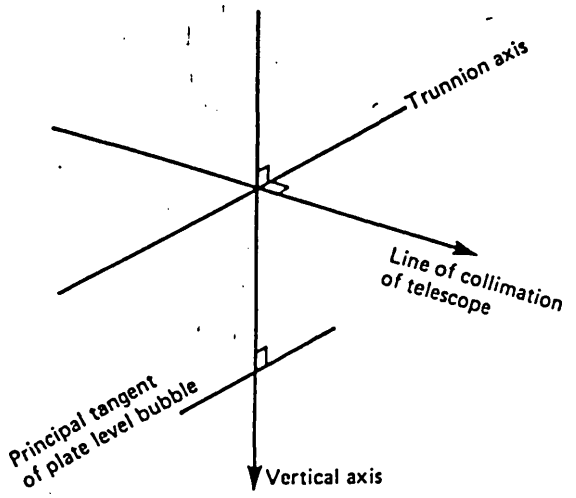


Figure 19.1 Theodolite axes

The detailed construction of a modern one-second precise theodolite is shown in *Figures 19.2* and *19.3*. Examination of these figures illustrates that the theodolite consists essentially of three distinct parts:

(1) *Base*. This consists of two main components; namely, the tribrach and the horizontal circle. The tribach can be screwed securely to the tripod and, by means of three footscrews, the instrument may be levelled. The circle is made of glass with photographically etched graduations. It is normally graduated in a clockwise manner. A circle setting screw is also usually provided and the base will generally house an optical plummet. This consists of a small eyepiece with a line of sight which is deviated by 90° in order to point vertically down. By this process it is possible to centre the instrument precisely over a ground point. In some cases the optical plummet may be housed in the alidade.

(2) *Alidade*. This rotatable upper part of the theodolite is also known as the upper plate. The alidade rotates about the vertical axis. Mounted on the alidade is the plate level bubble which indicates whether the instrument is level. By means of clamps and slow-motion screws it is possible to rotate and clamp the alidade relative to the base.

(3) *Telescope*. Attached to the trunnion axis of the theodolite is the telescope. The telescope magnifies the object and, by the use of cross-hairs, allows the exact bisection of the target. Focussing of the object and the cross-hairs are carried out using separate focussing screws. A further clamp and slow-motion screw allow precise pointing of the telescope in a vertical plane.

Angles of elevation or depression are measured using a vertical circle also attached to the trunnion axis. Prior to measuring a vertical angle it may be necessary to set the altitude bubble. However, the latest theodolites all employ an automatic compensating mechanism. In these cases vertical angles may be recorded after the plate level has been set, without recourse to an additional bubble setting.

When the vertical circle is to the left of the telescope, the theodolite is in what is conventionally called the face left (FL) position. Conversely when the vertical circle is to the right of the telescope as it views an object, the theodolite is in the face right (FR) position.

By projecting daylight through the standards of the theodolite, it is possible to illuminate the glass scale of both the horizontal and vertical circles.

In order to resolve a direction to a higher precision than that to which the circle has been graduated, an optical micrometer is employed. Optical micrometers are the modern equivalent of verniers. The principle of operation involves the use of a plane parallel-sided block of glass as shown in *Figure 19.4*. When the glass is in the normal position, as shown by position (a), light

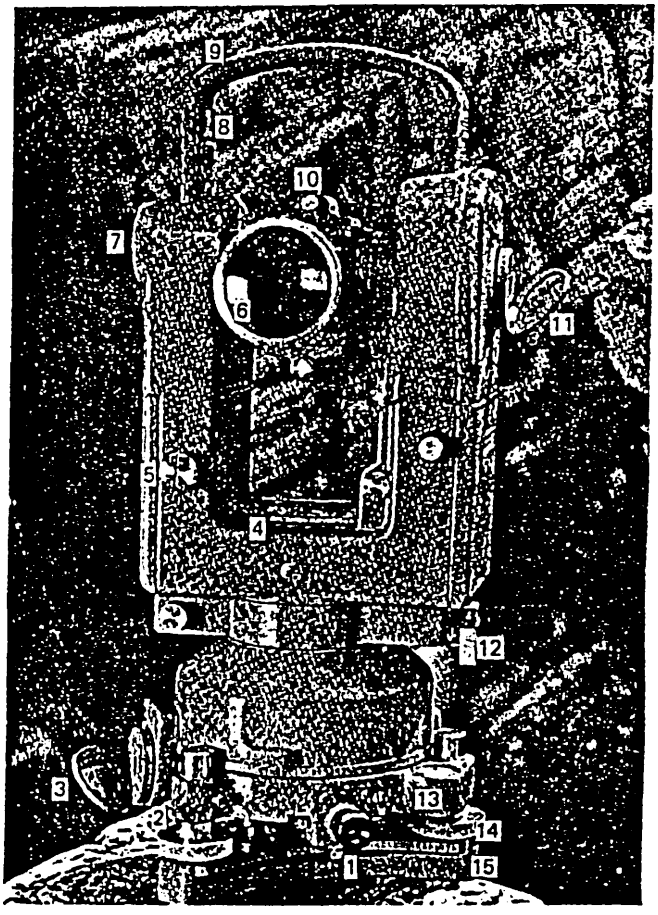


Figure 19.2 Wild T-2 one-second theodolite (Wild Heerbrugg). 1. Eyepiece for optical plummet. 2. Tribrach. 3. Illumination mirror for horizontal circle. 4. Plate level. 5. Selector knob for viewing either horizontal or vertical circle. 6. Telescope objective. 7. Optical micrometer knob. 8. Locking screw for carrying handle. 9. Carrying handle. 10. Optical sighting device. 11. Illumination mirror for vertical circle. 12. Horizontal clamp for alidade. 13. Circular bubble. 14. Footscrew. 15. Trivet stage

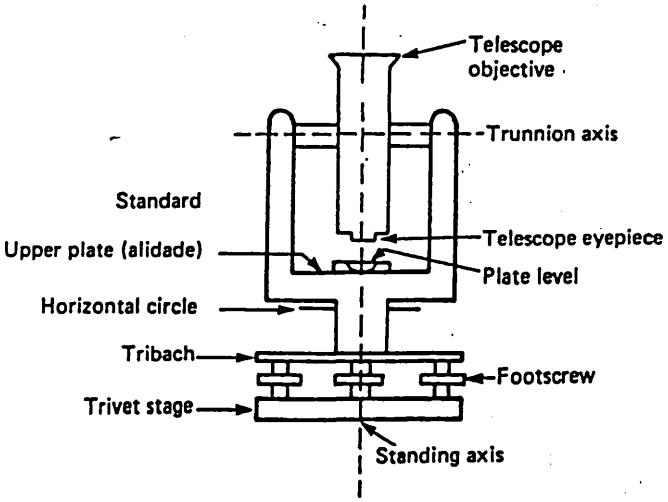


Figure 19.3 Constructional components of a theodolite

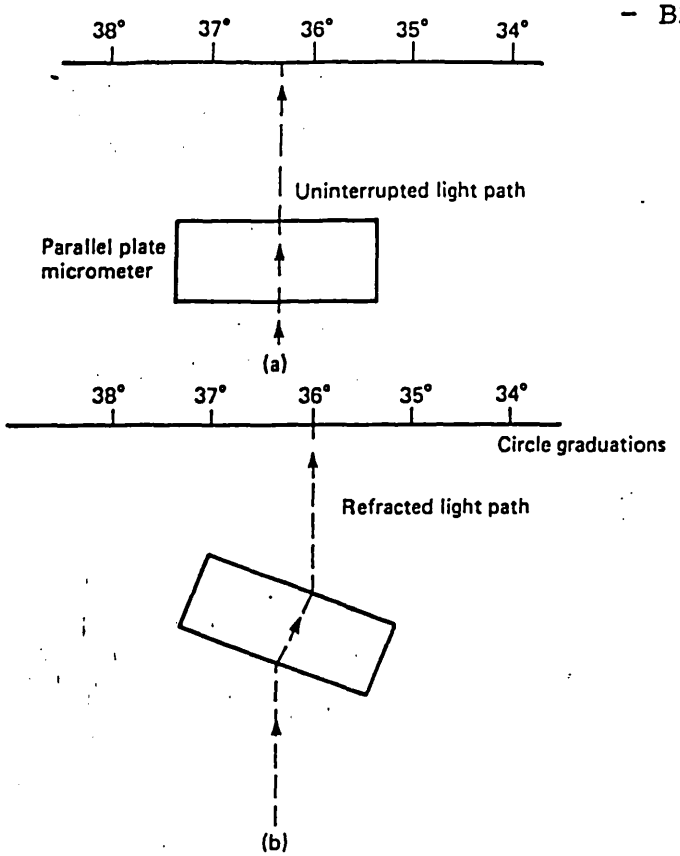


Figure 19.4 Parallel plate micrometer

passing through will be uninterrupted. Rotation of the block of glass, however, produces a lateral shift of the incident beam as shown by position (b). This rotation is controlled by the micrometer screw of the theodolite. Movement of this screw enables the observer to read, on an auxiliary scale, the lateral shift required in order to bring the image of the main scale degree graduations into coincidence with the index marks which are built into the optical path. Using this technique it is possible to resolve directly to 20" of arc if the micrometer is reading from one side of the circle. Resolution direct to 1" is possible if a mean reading optical micrometer is used. In this case readings from two points diametrically opposite are meaned in order to eliminate the effects of any circle eccentricity. *Figures 19.5* and *19.6* illustrate two typical examples of the circle reading systems for both the single and mean reading optical micrometers.

Potentially the theodolite is a very precise instrument. It is, however, necessary to follow a strict procedure both in setting up the instrument and in observing, if this potential is to be realized. Incorrect use of a theodolite will undoubtedly yield poor results, regardless of how precise the instrument may be.

Setting up a theodolite prior to observations being taken basically consists of three separate operations; centring, levelling and focussing. Centring involves positioning the instrument exactly over a ground point. This may be achieved by means of either a plumb bob suspended from the instrument or an optical plummet. The process of centring and levelling should be considered as iterative in nature, becoming increasingly more precise after each operation. Levelling the theodolite carefully is a necessary prerequisite for precise measurements. The following sequence of operations must be carried out in order to level a theodolite:

(1) Approximately level the instrument using the small circular bubble.
(2) Set the plane level bubble parallel to any two footscrews, such as A and B in *Figure 19.7a*. Rotate both footscrews together or apart until the bubble is in a central position.
(3) Rotate the alidade until the bubble is now approximately perpendicular to the initial position, as shown in *Figure 19.7b*. Using footscrew C only, centralize the bubble.
(4) Return to the initial position and again centralize the bubble using footscrews A and B. Repeat (2) and (3) until the bubble is in a central position.
(5) Rotate the alidade through 180° until the position shown by *Figure 19.7c* is achieved. If the bubble does not remain in a central position, move the bubble until it is in a position midway between a central position and its initial position.
(6) Rotate the alidade until the position illustrated by *Figure 19.7d* is achieved. Using footscrew C, move the bubble into the same position as in *Figure 19.7c*. The bubble should then remain in the same off-centre position for any alignment of the alidade.

The final step before observations begin is to focus both the cross-hairs and the object to which observations will be made. It is important to ensure that both images appear clear and sharp. In addition, it is critical that parallax does not exist. Parallax

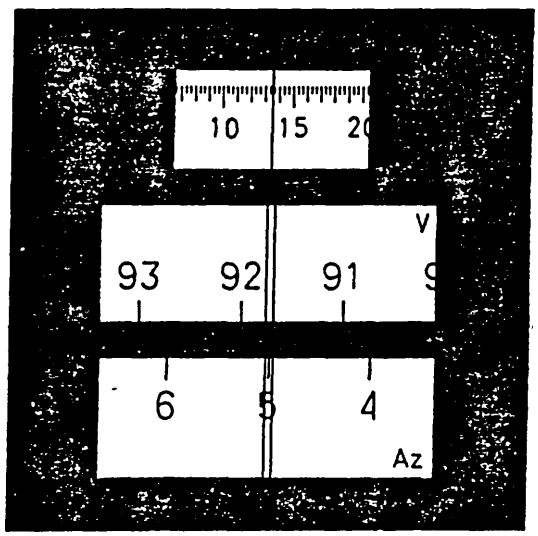


Figure 19.5 Single reading optical micrometer: circle reading Wild T-1A 05°13'30" (Wild Heerbrugg)

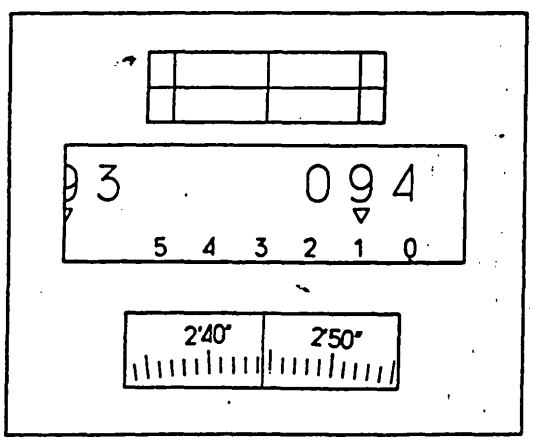


Figure 19.6 Mean reading optical micrometer: circle reading Wild T-2 94°12'44.3" (Wild Heerbrugg)

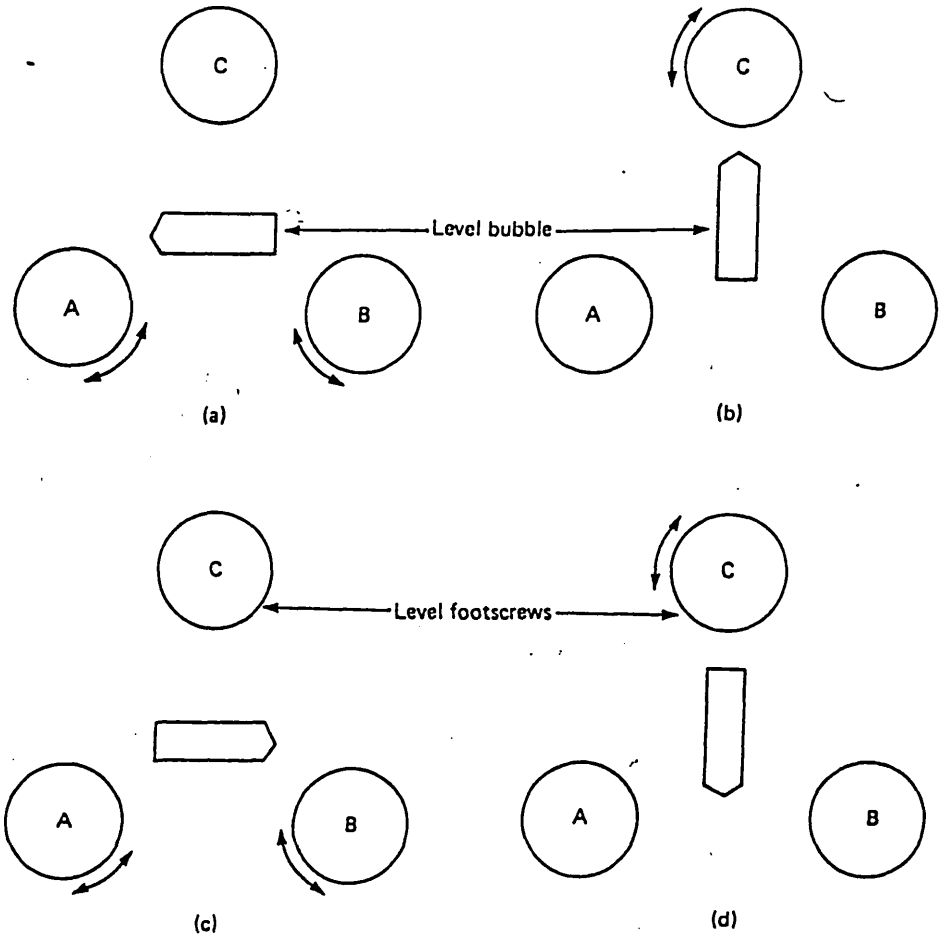


Figure 19.7 Levelling the theodolite

refers to the apparent movement of the cross-hairs and objects relative to each other when the observer moves his head. It is caused by the image of the object not lying in the same vertical plane as the cross-hairs. If this occurs, the focussing operation must be repeated until it is eliminated.

A strict observational procedure is essential if both human and instrumental errors are to be reduced to a minimum. Consider the problem of measuring the angle shown in *Figure 19.8*. The observational procedure which should be adopted is as follows. (The booking procedure is illustrated by *Table 19.2*.)

(1) Point with telescope in the face left position to the target at X and record the horizontal circle reading, for example, 90°20'30".
(2) Point on FL to target Z and record, for example, 130°25'40".

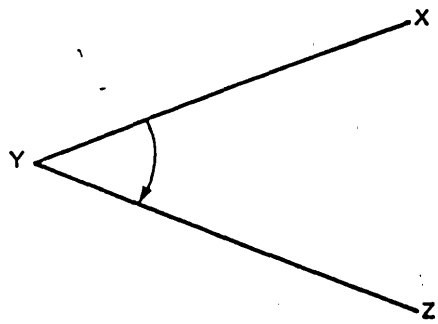


Figure 19.8 Angle measurement

Table 19.2 Booking and reduction of theodolite readings

Station Y — Observer ..... — Date .....
Height of Inst: ..... — Booker ..... — Weather .....

	To	*FL*	*FR*	*Mean*	*Angle*	*Comments*
		° ' "	° ' "	° ' "	° ' "	
Round 1	X	90 20 30	270 20 40	90 20 35	40 05 10	
	Z	130 25 40	310 25 50	130 25 45		
Round 2	X	135 30 15	315 30 25	135 30 20	40 05 15	
	Z	157 35 30	355 35 40	175 35 35		
	Mean angle: 40°05'13"					

(3) Change face to FR and point to target Z and record circle reading 310°25'50".
(4) Point on FR to target X and record direction 270°20'40".

To reduce the effect of instrument maladjustments to a minimum, the mean of the FL and FR circle readings to the same point is determined and the value entered in the mean column. The difference begween the mean circle readings is then derived and entered in the angle column. This constitutes one round of angles.

The base setting screw should then be adjusted and the process repeated in order to increase the precision of the angle measurement. A minimum of two rounds is necessary for the least precise measurements; up to 16 rounds may be required for very precise operations.

19.2.2 Distance measurement

The second fundamental quantity which it is necessary to measure is distance. A wide variety of techniques can be used for the determination of distance. The general distinction can, however, be made between direct, optical and electronic methods. All of the techniques discussed are capable of varying levels of precision depending on the degree of sophistication of the instrumentation and of the observational techniques adopted.

The simplest method of measuring distance is that of physically measuring the distance with a tape. In the past, invar tapes were used for the precise measurement of baselines for triangulation networks. Nowadays direct distance measurement (DDM) is generally confined to either the precise measurement of short distances for setting out or control purposes or the less precise measurement of the detailed dimensions of a building or land parcel.

There are basically two types of tape in common use. Fibreglass measuring tapes are manufactured from multiple strands of fibreglass coated with PVC. They are waterproof and normally either 30 m or 50 m in length. Fibreglass tapes are generally used for detail measurements and have largely superseded the linen tapes which were available previously. For more precise measurements it is necessary to use steel bands. These are typically 30 m, 50 m or 100 m in length.

In order to obtain high precision with either type of tape, it is essential that it is periodically checked against a standard reference tape, the length of which is known to a higher order of accuracy than that of the tape being checked. If a significant variation exists, a standardization correction should be applied. In addition it is vital that suitable attention is paid to the effect of variations in slope, temperature and tension which may necessitate appropriate corrections being applied to the measured distance. The corrections (C_1, C_2 and C_3 respectively) are:

Slope

$$C_1 = -L(1-\cos\theta) \qquad (19.4)$$

where θ = slope angle, L = measured slope distance; or

$$C_1 = -\Delta h^2/2L$$

where Δh = height difference between end points.

Temperature

$$C_2 = \pm\alpha L(t_m - t_s) \qquad (19.5)$$

where t_m = measured temperature in the field,
t_s = temperature at which the tape was standardized, usually 20°C,
α = coefficient of linear expansion (0.0000112 for steel bands).

Tension

$$C_3 = \pm L(T_m - T_s)/AE \qquad (19.6)$$

where
T_m = measured tension,
T_s = standard tension,
A = cross-sectional area of tape,
E = Young's modulus of elasticity for the tape, typically 200 kN/mm² = 200 000 N/mm².

Miller[3] detailed the typical accuracy levels which can be achieved with steel tapes.

As an alternative to the direct method of measuring distance, it is also possible to measure distance indirectly by optical methods. The development of optical distance measurement (ODM) began over two centuries ago. James Watt is recorded as having used this approach in his survey of the west of Scotland in 1774. Although many instruments and improvements have been introduced since then, they all essentially involve the solution of the same geometrical problem.

All methods of ODM are based on the solution of an isosceles triangle, as shown in *Figure 19.9*. The triangle consists of three important components; the parallactic angle α, the base length B (which may be either in a horizontal or vertical plane), and D, the horizontal bisector of the base of the triangle. By knowing the relationship between the three components, the horizontal distance D between two points can be determined.

Two methods of solution are possible; either an instrument with a fixed parallactic angle is used and the variable base B is measured (*Figure 19.10*), or a base of fixed length is set up and the variable parallactic angle is measured (*Figure 19.11*). In both cases the variable quantity is proportional to the horizontal distance. By defining the mathematical relationship between the fixed and variable quantities it is therefore possible to determine the horizontal distance.

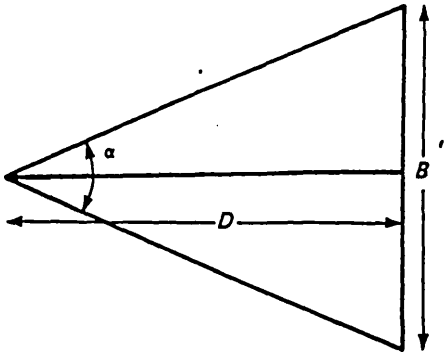


Figure 19.9 Optical distance measurement (ODM)

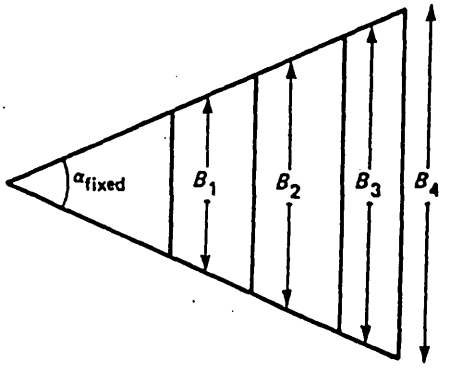


Figure 19.10 ODM: fixed angle, variable base

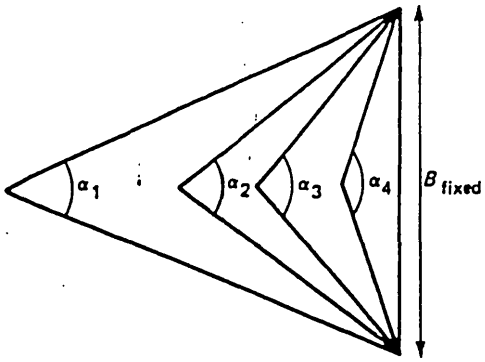


Figure 19.11 ODM: fixed base, variable angle

The first approach described above (fixed angle, variable base) is commonly known as *tacheometry* or more correctly as *vertical staff stadia tacheometry*. It is normally used for the measurement of distance where a proportional error of between 1/500 and 1/1000 is acceptable; for example, in picking up survey detail points.

All modern theodolites have a diaphragm consisting of a main horizontal cross-hair and two horizontal stadia lines spaced either side of it. These stadia lines define the fixed parallactic angle. If the theodolite telescope is sighted onto a levelling staff and the readings of the outer lines noted, the difference in the readings, the staff intercept (s), will be directly proportional to the horizontal distance between the instrument and the staff. Generally, the distance between the stadia lines is designed in such a manner that the horizontal distance D_H between the instrument and staff is given by

$$D_H = 100s \tag{19.7}$$

For inclined sights the geometry is as shown in *Figure 19.12*. Hence

$$D_s = 100(s \cos \theta) \tag{19.8}$$

where D_s = slope distance,
θ = vertical angle measured by the theodolite.

$$\therefore\ D_H = 100s \cos^2 \theta \tag{19.9}$$

$$\Delta h_{AB} = h_I + V - m \tag{19.10}$$

where Δh_{AB} = difference in height between A and B,
h_I = height of instrument (trunnion axis to ground),
m = middle hair reading,
V = difference in height between middle hair reading and trunnion axis = $50s \sin 2\theta$ (19.11)

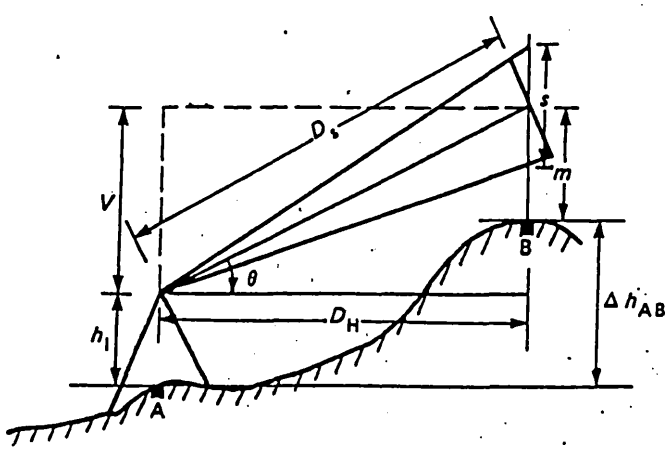


Figure 19.12 Stadia tacheometry: inclined sights

Several self-reducing tachometers have also been designed. The main advantage of these instruments is their ability to compensate for the effect of the inclination of the theodolite telescope and hence allow the direct determination of horizontal distance without additional computation.

Two notable examples of this type of instrument are the Wild RDS vertical staff self-reducing tachometer and the Kern DK-RT horizontal bar double image self-reducing tachometer. Details of the construction and use of these instruments has been described by Smith[4]. In recent years the manufacture of these precise optical devices has ceased, their place being taken by low-cost electronic measuring devices.

The second approach (fixed base, variable angle) is commonly known as the *subtense* or *horizontal subtense bar method*. The method is normally confined to the measurement of distance for control purposes. Using this approach distances may be determined with a proportional error of up to 1/10000.

The instrumentation required consists of a subtense bar, normally 2 m long, and a one-second theodolite, such as the Wild T2. The bar has targets mounted at each end of an invar strip. The strip is protected by a surrounding aluminium strip in order to ensure that, for all practical purposes, the length of the bar remains constant at 2 m. The bar is set up and oriented at right angles to the line of sight of the theodolite, as shown in *Figure 19.13*.

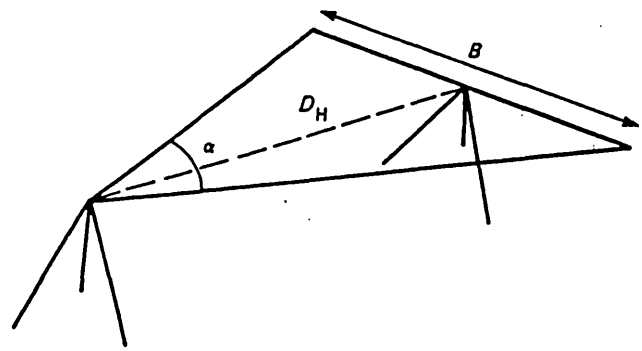


Figure 19.13 Subtense bar

The horizontal parallactic angle α is measured with the theodolite. Irrespective of the vertical angle to the bar, the horizontal distance is given by:

$$D_H = B/2 \cot \alpha/2 \tag{19.12a}$$

with

$$B = 2\text{ m}$$

$$D_H = \cot \alpha/2. \tag{19.12b}$$

For distances greater than 100 m it is advisable to subdivide the distance to be measured or alternatively to use the auxiliary base method[5].

Subtense methods are also tending to be superseded by low-cost electronic methods. Nevertheless many organizations still possess this type of equipment and for many projects it is a very suitable technique to adopt.

Electronic distance measurement (EDM) instruments were developed in the early 1950s. Typical of the early developments were the Swedish Geodimeter (GEO DIstance METER) and the South African Tellurometer instrument. The former, an electro-optical instrument, used visible light measurement, whilst the latter used high-frequency microwaves. Both instruments were primarily developed for military geodetic survey purposes and had the ability to measure long distances (up to 80 km in the case of the Tellurometer) to a precision of a few centimetres. They were also, however, bulky, heavy and expensive in comparison to their modern-day equivalents.

During the late 1960s, developments in microelectronics and low-power light-emitting diodes led to the emergence of a second generation of EDM instruments. These electro-optical instruments utilized infrared radiation as the measuring signal and were developed for the short range (<5 km) market. In addition they were considerably smaller, lighter and less expensive than their predecessors. Probably the best-known example is the Wild DI-10 Distomat.

The introduction of microprocessors into the survey world in the early 1970s led to the introduction of a third series of EDM instruments. With this group it became possible not only to determine slope distance, but also to carry out simple computational tasks in the field. For example, the facility became available to automatically compute the corrected horizontal distance and difference in height between two points by manual input of the vertical angle read from the theodolite. EDM instruments of this type had also been reduced in size to the extent that the EDM unit could be theodolite mounted. The Wild DI-3 is a typical example of this type of instrument.

The latest short-range EDM instruments are similar to the previous group, but have several additional features worthy of mention. Firstly, the technology now exists to automatically sense the inclination of the EDM unit and therefore to be able automatically to compute the horizontal distance between two points. The AGA Geodimeter 122 (*Figure 19.14*) has this facility. This instrument also has the ability to measure to a moving target, or track, a useful feature for setting-out purposes. By using an additional unit it is also possible to have one-way speech communication between the instrument and the target, again valuable when setting out. This instrument can also be connected to a GEODAT hand-held data collector (*Figure 19.14*), which is able to automatically store distance information from the EDM unit. Other relevant information (numeric or alphanumeric) can be input manually via the keyboard. The AGA 122 has a range of up to 2.5 km with one prism, determined to a standard error of ±5 mm ± 5 parts per million of the distance (p.p.m.).

Finally, mention should be made of high-precision EDM instruments. These instruments have been designed for projects such as dam deformation or foundation monitoring, where extremely high precision is necessary. The Comrad Geomensor 204 DME, for example, has a range of up to 10 km with a standard error of ±0.1 mm ± 0.5 p.p.m. Burnside[6] provides up-to-date technical information on many modern EDM instruments.

Although there are a wide variety of EDM instruments on the market, they all measure distance using the same basic principle. This can be most clearly illustrated by means of the flow diagram (*Figure 19.15*), which relates specifically to electro-optical instruments.

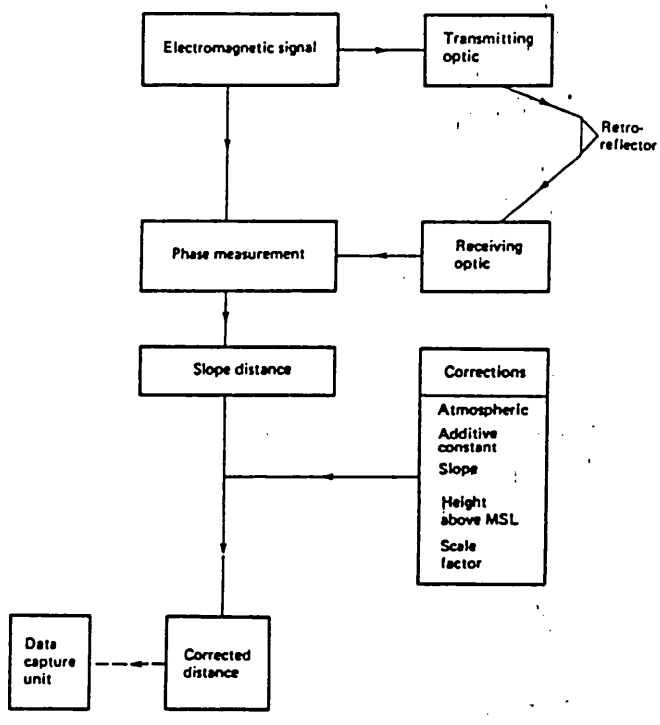


Figure 19.15 Principle of operation: electro-optical distance measurement

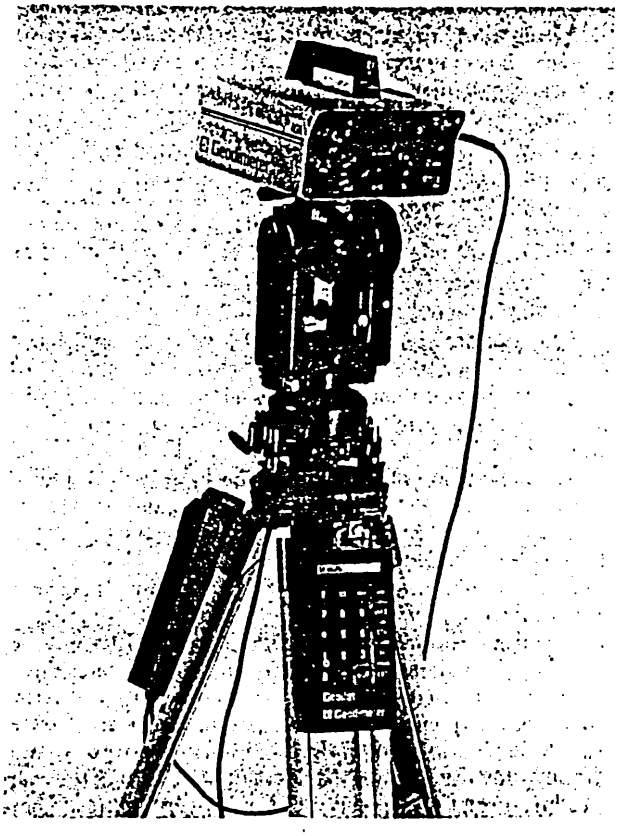


Figure 19.14 Geodimeter 122 and Geodat data collector (Geotronics)

An electromagnetic (EM) signal of wavelength equal to either 560 nanometres (nm) (visible light), 680 nm (HeNe laser) or 910 nm (infrared) is generated. This signal is subsequently amplitude modulated before being transmitted through the optical system of the instrument towards a retro-reflector mounted at the end of the line to be measured. The signal is then retro-reflected, or redirected through 180°, by a precisely ground glass corner cube. Cheaper acrylic corner cubes may also be used[7]. This reflected signal is consequently directed towards the receiving optical system of the instrument. On entering the optical system of the instrument, the signal is converted by means of a photomultiplier into an electrical signal.

The next stage involves the measurement of the phase difference between the transmitted and received signals and the conversion of this information into distance. *Figure 19.16* shows the path taken by an EM signal radiated by an EDM instrument together with the instantaneous phase of the signal. It is apparent that the distance X–Y–X travelled by the EM signal is equivalent to twice the distance to be measured. Also, this distance can be seen to be related to the modulation wavelength (λ) and the

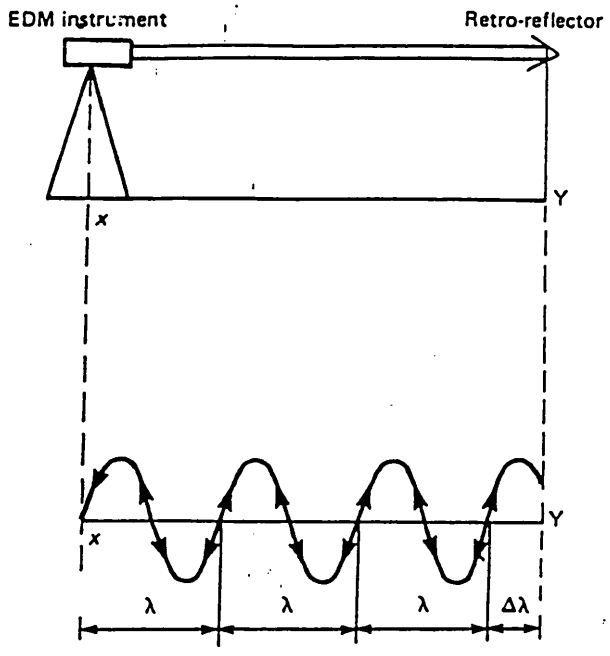


Figure 19.16 Electromagnetic signal path

fraction of the wavelength ($\Delta\lambda$) by the following relationship:

$$2D = n\lambda + \Delta\lambda \quad (19.13)$$

where n is an unknown integer number of wavelengths. The determination of the distance therefore involves resolving both $\Delta\lambda$ and n. Phase detectors are used to determine $\Delta\lambda$ which effectively measures the phase difference between the transmitted and received signal and hence allows the fractional part of the distance less than one full wavelength to be determined. The value of n can be determined by using two or more EM waves of slightly varying wavelengths. For example, using a simple example, assume $2D = 25.5$, $\lambda_1 = 2.500$ m and $\lambda_2 = 2.400$ m, then:

$$2D = 2.5\,n_1 + 0.5 \quad (19.14)$$

and

$$2D = 2.4\,n_2 + 1.5 \quad (19.15)$$

Assuming $n_1 = n_2$ for short distances and solving for n leads to

$$n = 10 \quad (19.16)$$

Substituting in Equation (19.14) or (19.15),

$$D = 12.75\text{ m} \quad (19.17)$$

This entire process is fully automatic in modern instruments, taking approximately 10–20 s to complete.

As with any other method of distance measurement, it is necessary to apply several corrections to the measured slope distance in order to determine the corrected horizontal distance. The first correction to be applied is the atmospheric or refractive index correction. Just as a steel tape varies in length with variations in temperature and pressure, so, too, does the modulation wavelength of an EDM instrument. It is therefore necessary to measure the temperature, pressure and, in some cases, relative humidity during measurements. A correction is then applied to compensate for the variation in modulation wavelength caused by variations in atmospheric conditions.

Many instruments have the facility to automatically compute and apply this correction to observations directly in the field. Temperature, pressure and relative humidity readings are taken and the appropriate reading to be set on the refractive index correction dial is read from a nomogram.

A second important correction is that known as the additive constant. This correction represents the difference between the electro-optically determined distance and the correct length of line. It is a combination of the errors due to prism offset and the variation in the physical and electrical centres of the EDM instrument. Many manufacturers design their corner cube reflectors to eliminate this correction totally. However, if several different types of corner cube are being used it is essential that a full field calibration be undertaken in order to determine the correction[8]. The slope correction is the same as for DDM. For distances measured above or below mean sea level (MSL) a correction is necessary in order to reduce the distance to its equivalent at MSL. The correction (C_4) is given by

$$C_4 = \frac{-LH_m}{R} \quad (19.18)$$

where H_m is the mean height of the instrument and reflector above MSL and R is the radius of the Earth (6370 km). Finally, if the distance is to be used for computation of coordinates on the National Grid, the horizontal distance at MSL must be multiplied by the local scale factor. For the Transverse Mercator projection, the local scale factor (F) may be approximately calculated from:

$$F = 0.999\,601\,27 + (1.228 \times 10^{-14} \times (E - 400\,000)^2) \quad (19.19)$$

where E is the mean local National Grid easting in metres of the line to be measured.

19.2.3 Height measurement using the level

The third and final quantity measured is height or, more correctly, height difference. This is achieved by means of a level.

The fundamental principle of the level is illustrated by *Figures 19.17* and *19.18*. *Figure 19.17* represents the situation which normally exists when the level is initially set up. In this case the

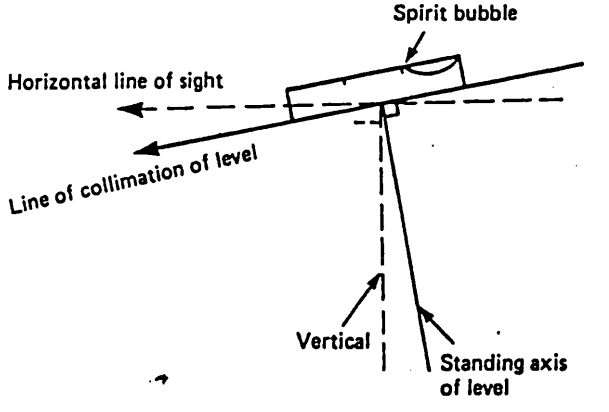


Figure 19.17 Geometry of the level axes: before levelling

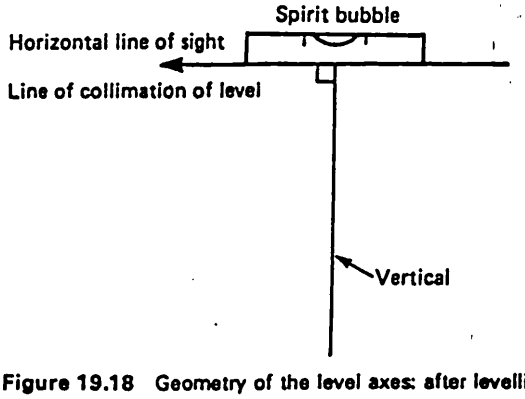


Figure 19.18 Geometry of the level axes: after levelling

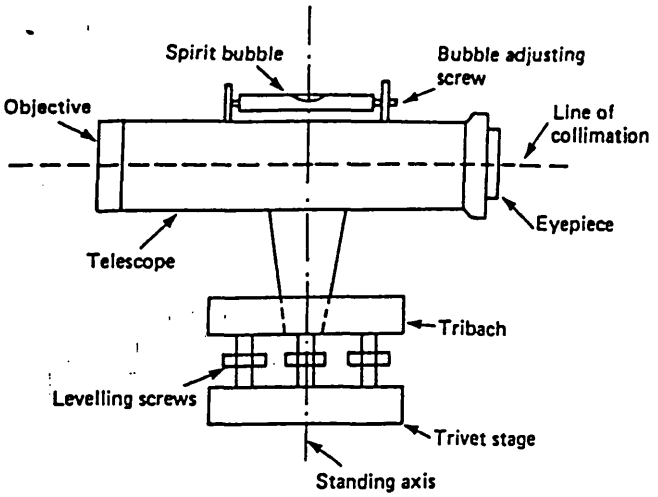


Figure 19.19 Level construction: dumpy level

standing axis of the level and the vertical do not coincide. Hence the line of collimation of the level will not be horizontal. *Figure 19.18* represents the geometrical arrangement of the axis when the instrument has been levelled using the procedure outlined in Section 19.2.1. It can be seen that completion of this procedure ensures that, firstly, the standing and vertical axes are made coincident and, secondly, if the instrument is in perfect adjustment, the line of collimation of the level is coincident with a horizontal line of sight.

Three distinct types of level are available for engineering survey purposes.

Dumpy levels were so named because of the rather short telescopes which were used by early versions of this instrument. The construction of a typical dumpy level is shown in *Figure 19.19*. The most distinctive feature of this type of level is that the axis of the telescope is fixed rigidly to the standing axis of the instrument. In order to satisfy the condition that both the vertical and standing axes are coincident, the standard levelling procedure outlined in Section 19.2.1 is carried out. Rotation of the telescope will now define a horizontal plane.

In the past this type of level was very popular for general engineering work. It has, however, been replaced in recent years by the automatic level.

The *tilting level* is a more precise instrument than the dumpy level. *Figure 19.20* illustrates the main features of this type of level. In contrast to the dumpy level, the telescope is not rigidly attached to the standing axis but is able to be tilted in a vertical plane about a pivot point X, by means of a tilting screw.

Prior to recording an observation, the instrument is approximately levelled. This is normally achieved by means of a

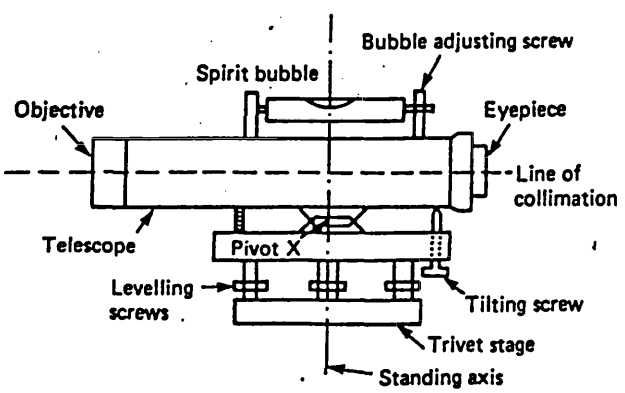


Figure 19.20 Level construction: tilting level

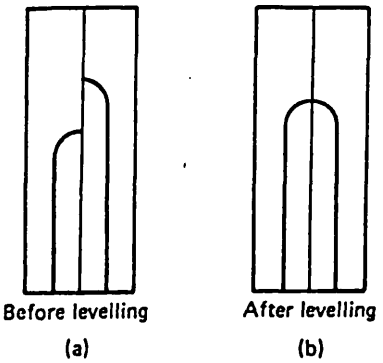


Figure 19.21 Coincidence bubble reading system

ball and socket arrangement and a small circular bubble. In order to set the standing axis exactly vertical, the tilting screw is turned and the main bubble altered until a coincident position (*Figure 19.21*), as viewed through a small auxiliary eyepiece, is achieved. If the telescope is now rotated horizontally to sight a second or subsequent point, it is important to relevel the main bubble by means of the tilting screw.

An *automatic level* is not, as the name suggests, totally automatic. Human intervention is still necessary. However, one major source of human error, that of setting the bubble, is replaced by an automatic compensating system. In common with the tilting level, approximate levelling is still necessary. The tedious and error-prone bubble-setting process, however, is eliminated. As with the dumpy level, the instrument when rotated defines a horizontal plane. The automatic level therefore combines the speed of operation of the dumpy level with the precision of the tilting level.

Figure 19.22 illustrates the main components of this type of level. The essential feature of the instrument is the incorporation

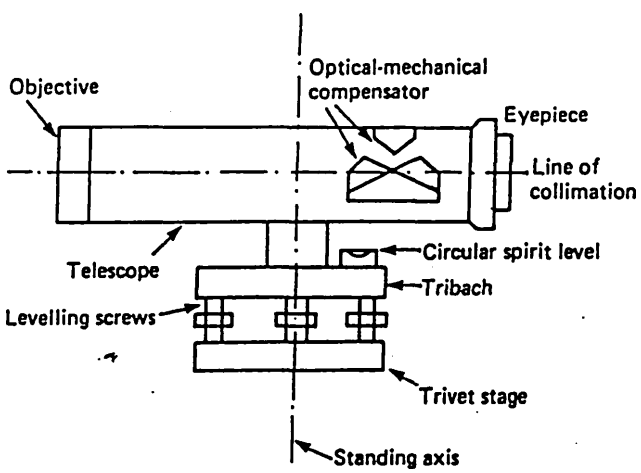


Figure 19.22 Level construction: automatic level

of an automatic optical-mechanical compensating mechanism. The use of such a system ensures that the line of collimation as defined by the centre cross-hair will trace out a horizontal plane irrespective of the fact that the optical axis of the instrument may not be exactly horizontal. It is, however, necessary to level the instrument approximately to ensure that the line of sight is within range of the compensating mechanism.

So far the assumption has been made that, once the standing axis of the level has been set truly vertical, the line of collimation will be horizontal. This may not always be the case. If this condition does not occur, a collimation error is said to exist. This

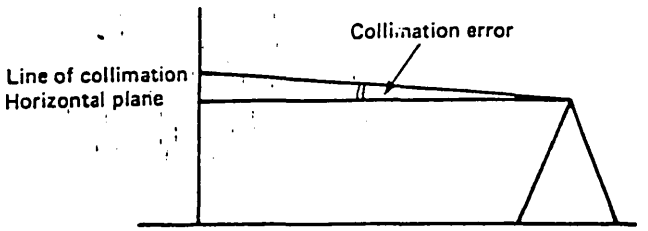
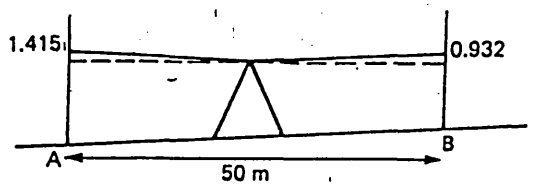


Figure 19.23 Collimation error

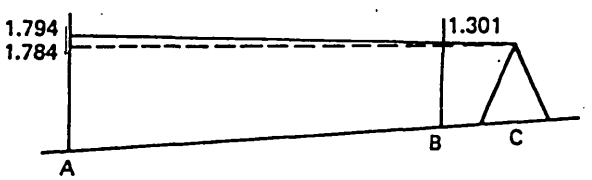
is illustrated by *Figure 19.23*. If accurate levelling is to be achieved, it is essential that a regular testing procedure is established in order to check the magnitude of any collimation error that may exist.

A common field procedure which can be used to test a level is known as the 'two-peg test'. The procedure is as follows:

(1) Set out two points A and B approximately 50 m apart, as shown in *Figure 19.24a*. The level is set up at the mid-point of AB and levelled as in Section 19.2.1.



(a)



(b)

Figure 19.24 Two-peg test

(2) A reading is taken onto a staff held at points A and B and the difference between the two readings calculated. This value represents the true difference in height between A and B. Any collimation error which exists will have an equal effect on both readings and hence will not affect the difference between the readings. In this case the difference in height is 1,415 − 0.932 = 0.483 m.

(3) The instrument is now moved to a point C close to the staff at B (about 3–5 m away), *Figure 19.24b*. The reading on staff B is recorded (1.301). If no collimation error exists the reading on staff A should be equal to the reading on staff B ± the true difference in height as established in step 2; that is, 1.301 + 0.483 = 1.784 m.

(4) The actual observed reading on staff A is now recorded (1.794). Any discrepancy between this value and that derived previously in step 3 indicates the magnitude and direction of any collimation error. For example, in this case, the error would be 1.794 − 1.784 = 10 mm per 50 m.

An error of up to 2–3 mm over this distance would be acceptable. If, however, the error is greater than this, the instrument should be adjusted. Unlike theodolite adjustments, this type of adjustment can normally be performed without any great difficulty by the engineer and the procedure is as follows:

(1) For the dumpy and automatic level, alter the position of the cross-hairs until the centre cross-hair is reading the value which should have been observed from step 3 above. This is achieved by loosening the small screws around the eyepiece which control the position of the cross-hairs.

(2) For the tilting level, again alter the position of the centre cross-hair until it is reading the value previously determined in step 3, in this case by tilting the telescope using the tilting screw. Unfortunately, this will displace the bubble. The bubble must, therefore, be centralized by means of the bubble adjusting screw.

19.3 Surveying methods

19.3.1 Horizontal control surveys

Any engineering survey or setting-out project, regardless of its size, requires a control framework of known coordinated points. Several different control methods are available as described below. The choice of which method to use depends on many factors; for example, the purpose for which the control is required, the accuracy required, the density of control points required, the type of equipment and computing facilities available and, lastly, the physical nature of the ground.

Triangulation is the oldest, and in the past was the most common, method of control for large civil engineering projects. The principles are well known and essentially involve the establishment of a measured baseline from which a network of triangles are formed, all of the angles of the triangle being measured. The development of EDM has, however, led to the establishment of several alternative control methods, such as trilateration and traversing. The introduction of EDM has therefore tended to make 'classical triangulation' obsolete as a method of control.

Trilateration is a method of establishing control which involves the direct measurement, normally using EDM, of all the sides of a network of triangles, in contrast to triangulation which involves the measurement of angles. Although the method has been used in this 'classical form', it does not offer any significant advantages over the method of triangulation. The method has therefore not become particularly common.

Traversing is a method of establishing control by measuring the distance between successive points and also the horizontal angle between adjacent stations, as shown in *Figure 19.25*. The method is very popular for several reasons. Firstly, it is a more flexible method than triangulation. In the case of triangulation, the positions of the control stations must be chosen so that not only are they intervisible, but also the triangles formed are well conditioned. For this reason the reconnaissance stage in triangulation projects is extremely important and often very time-consuming. In contrast, with traversing much less attention has to be paid to the reconnaissance stage, since it is only required that adjacent stations be intervisible. This allows the surveyor much greater flexibility in the choice of control station

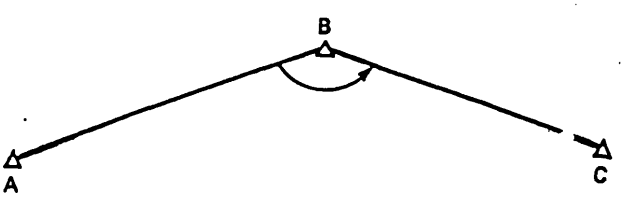


Figure 19.25 Traversing

positions. The stations can then be positioned in areas close to the detail to be picked up, or close to the project for which they are required. This can be of enormous benefit in areas which are either very flat or, alternatively, heavily forested. A second reason for the popularity of traversing is the computational simplicity, both of determining provisional coordinates and also of adjusting any misclosure which may exist. It should, however, be mentioned that in recent years more rigorous techniques of adjustment, based on the principle of least squares, have become more common for the adjustment of traverse.

Traversing does, however, suffer from one serious drawback; that is, the lack of redundant observational data. As a consequence, the effect of small errors of measurement is not only difficult to detect but is also cumulative in nature. To counteract this problem, additional angle and distance observations are often taken in order to strengthen the control framework. In ths past this additional information tended to be used solely for the detection of gross errors. Nowadays, by using a suitable adjustment technique, these additional observations can be used to improve the precision of the coordinates.

The normal procedure adopted for traverse adjustment is, firstly, to determine and adjust the angular misclosure and, secondly, to determine and adjust the misclosure in easting and northing. The angular misclosure is determined, in the case of a closed polygon, by adding the internal angles. These should sum to $(2n-4)$ right angles, where n is the number of traverse sides. A misclosure of $>(20''\sqrt{n})$ would not be acceptable for site traverses.

The question of which method to use for the adjustment of any misclosure in eastings and northings is a matter of some controversy. Traditionally the Bowditch and Transit methods have been used:

Bowditch:

$$\Delta E_c = M_E \times \frac{d}{\sum d} \quad \text{and} \quad \Delta N_c = M_N \times \frac{d}{\sum d} \tag{19.20}$$

Transit:

$$\Delta E_c = M_E \times \frac{\Delta E}{\sum |\Delta E|} \quad \text{and} \quad \Delta N_c = M_N \times \frac{\Delta N}{\sum |\Delta N|} \tag{19.21}$$

where M_E, M_N is the misclosure in easting, northing,
d is the length of a traverse leg,
$\sum|\Delta E|$, $\sum|\Delta N|$ is the absolute sum of the provisional ΔE, ΔN,
ΔE_c, ΔN_c is the correction to be applied to the provisional ΔE, ΔN.

Schofield[9] and Lam[10] have discussed the problems which arise when semi-rigorous methods of adjustment such as Bowditch and Transit are used to adjust modern EDM traverses. The main conclusion they reach is that both methods are based on assumptions which are not applicable to EDM. For example, it is assumed that the expected error in an EDM measurement is proportional to the distance measured, which is clearly not true. It is suggested that all EDM-based traverses should be adjusted by a rigorous method such as variation of coordinates[9].

The combination of angle, distance and orientation measurements to form a control framework is now commonly referred to as a survey network. The advantages of such an approach are considerable. Firstly, the scale and orientation errors associated with classical triangulation can be reduced by the inclusion of additional distance and azimuth measurements. Secondly, the control framework does not suffer from the serious propagation of errors which can occur in traversing. Thirdly, the optimum number of observations (angular and distance) can be determined before the field work commences, using computer simulation techniques. Finally, by using the method of variation of coordinates a least squares procedure can be used to determine the most probable values of the coordinates.

The principle of least squares states that the most probable value of a sample is that for which the sum of the squares of the residuals is a minimum. *Variation of coordinates* is a computational method, based on this principle, which is used for the determination of coordinates in a survey network.

Consider *Figure 19.26* which illustrates a small network in which all the eight angles α^o (1 to 8) and all distances l^o (1 to 6) have been observed. Point A is fixed, the bearing to point D is

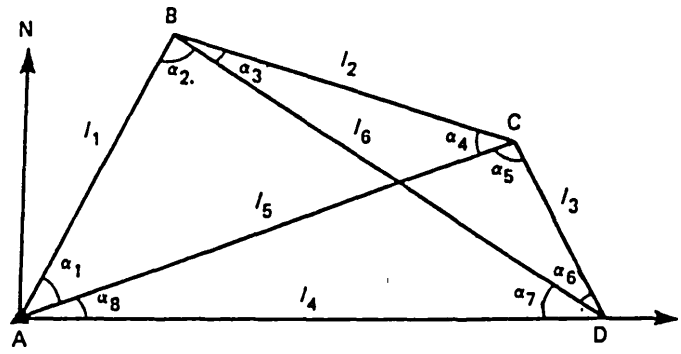


Figure 19.26 Adjustment by variation of coordinates

fixed in order to orient the network and the *a priori* standard errors of the observations are estimated. There are, therefore, 15 observations and six unknowns to be determined, i.e. the eastings and northings of points B, C and D. The method of solution is as follows:

(1) Assume provisional coordinates for all points. These may be either scaled graphically from a plan or computed using a selection of the measurements.
(2) Using the provisional coordinates, compute values for the angles α^c 1–8 and the distance l^c 1–6.
(3) Set up an observation equation for each individual observation.

For a distance l_{ij}

$$-\cos\beta_{ij}\,dN_i - \sin\beta_{ij}\,dE_i + \cos\beta_{ij}\,dN_j + \sin\beta_{ij}\,dE_j = (l_{ij}^o - l_{ij}^c) + v \tag{19.22a}$$

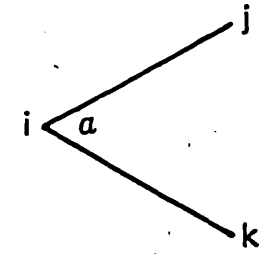
$$= P\,dN_i + Q\,dE_i + R\,dN_j + S\,dE_j = (l_{ij}^o - l_{ij}^c) + v \tag{19.22b}$$

For bearing ϕ_{ij}

$$\frac{\sin\beta_{ij}}{l_{ij}}dN_i - \frac{\cos\beta_{ij}}{l_{ij}}dE_i - \frac{\sin\beta_{ij}}{l_{ij}}dN_j + \frac{\cos\beta_{ij}}{l_{ij}}dE_j = (\phi_{ij}^o - \phi_{ij}^c) + v \tag{19.23a}$$

$$= K\,dN_i + L\,dE_i + M\,dN_j + N\,dE_j = (\phi_{ij}^o - \phi_{ij}^c) + v \tag{19.23b}$$

For angle α_{ijk}



B231

$$\left(\frac{\sin\beta_{ik}}{l_{ik}}-\frac{\sin\beta_{ij}}{l_{ij}}\right)dN_i+\left(\frac{-\cos\beta_{ik}}{l_{ik}}+\frac{\cos\beta_{ij}}{l_{ij}}\right)dE_i$$
$$+\frac{\sin\beta_{ij}}{l_{ij}}dN_j-\frac{\cos\beta_{ij}}{l_{ij}}dE_j$$
$$-\frac{\sin\beta_{ik}}{l_{ik}}dN_k+\frac{\cos\beta_{ik}}{l_{ik}}dE_k$$
$$=(\alpha^o_{jik}-\alpha^c_{jik})+v \qquad (19.24)$$

For a position

$$dN_i=(N_i^o-N_i^c)+v \qquad (19.25)$$

$$dE_i=(E_i^o-E_i^c)+v \qquad (19.26)$$

where β_{ij} = direction of line ij,
dN_i, dE_i = the unknowns, the corrections to the provisional coordinates,
v = residual.

These observation equations can be expressed in matrix form:

A		X		b		v
Matrix of coefficients	6	Vector of unknowns	1	Vector of *O–C* terms	1	Vector of residuals
15		6		15		15

(19.27)

The solution is found by forming the normal equations.

$$\mathbf{A^TWAX=A^TWb} \qquad (19.28)$$

where W is the weight matrix, the diagonal elements of which are equal to $1/\sigma^2$ where σ is the *a priori* standard error of the observation.

The normal equations may then be solved using a Choleski triangular decomposition method or by matrix inversion:

$$\mathbf{x=(A^TWA)^{-1}A^TWb} \qquad (19.29)$$

The column vector x therefore contains the corrections (dN_i, dE_i) to the provisional coordinates of the unknown points. Using these new values for the coordinates, the entire computational procedure is repeated until there is no further change in the coordinates.

Usually when this method is used a complete error analysis of the results is carried out. By this process information about the precision and reliability of the coordinates can be obtained. For further reading on this aspect of the method see Ashkenazi[11].

19.3.2 Detail surveys

After the main control survey has been observed and computed, a detail survey is carried out to locate the positions of features which are to be presented on the map.

The most common method of carrying out a detail survey by ground methods at scales greater than 1:500 is that of stadia tacheometry. The basic principles have been outlined in Section 19.2.2.

In practice the observational and booking procedure should be as follows. The theodolite should be set up and levelled over the point from which observations are to be taken. This point may coincide with one of the main control survey stations, or more commonly be one of a series of subsidiary control stations which have been established by traversing from the main control. By this process not only are any errors contained within the control network, but also the positions of the subsidiary control will be closer to the detail which is to be surveyed.

The circle of the theodolite should then be oriented until a horizontal circle reading of 0° coincides with the direction to an adjacent control point. This is known as the reference object (RO). This procedure simplifies the subsequent plotting of the field observations. The height of instrument should be recorded.

The staffman is then directed to the various points which are to be mapped. The staff is held vertically over the point and the stadia readings, vertical angle and horizontal circle reading to the point are observed by the theodolite. Several techniques for speeding up the field recording of the stadia hair readings have been advocated[12]. Assuming that a pocket calculator, preferably programmable, will be used to reduce the observations, the most efficient field method is to set the bottom hair to the 1 m or 2 m point of the staff. Using this method the mental calculation of the stadia intercept becomes much quicker.

The booking procedure for stadia observations is illustrated in *Table 19.3*. It is also essential to draw, whilst in the field, a good sketch map of the area being surveyed. This is invaluable when the results are being plotted.

Chaining or *offsetting* is one of the simplest forms of detail surveying. The method involves measuring the lengths of the sides of a series of triangles or braced quadrilaterals. Points of detail are then picked up by measuring offsets from these lines. The procedure is illustrated in *Figure 19.27*. In this case two traverse stations from an existing control survey are within the boundary of the site which is to be surveyed.

The first stage in the survey involves breaking down this existing control into either well-conditioned triangles or braced quadrilaterals. In this case a braced quadrilateral is chosen. The inclusion of diagonal measurements provides an independent field check on the measurements. These 'chain lines' are normally measured with a steel band.

Table 19.3 Booking and reduction of tacheometric readings

Observer: TJMK — Station B
Booker: DRG — Height of instrument: 1.690 m
Date: 22/9/82 — Reduced level: 58.35 m

Point	*Horizontal circle*	*Staff readings*	*Staff intercept*	*Vertical circle*	*Horizontal distance*	*V*	*Reduced level*
		U M L	(S)		$(100\cos^2\theta)$ (D_H)	$(50\sin 2\theta)$	
A(RO)	0°00′	—	—	—	—	—	—
1	10°30′	1.430 1.215 1.000	0.430	91°00′	42.99	0.75	59.58

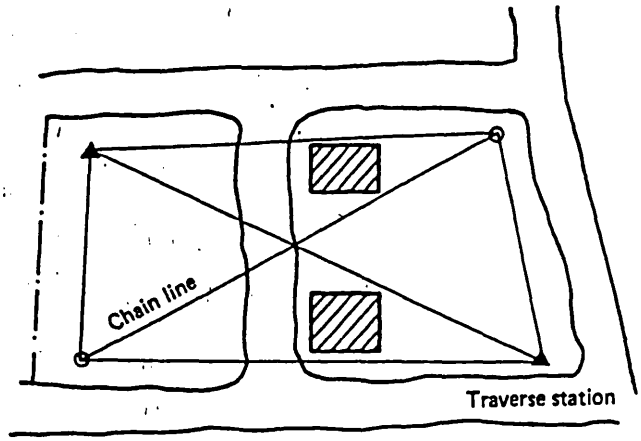


Figure 19.27 Chaining

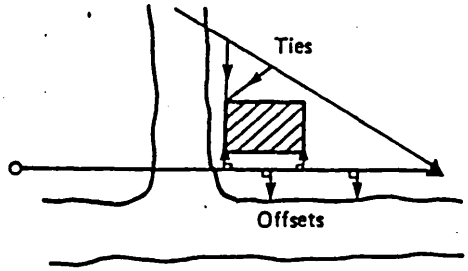


Figure 19.28 Offsets and ties

The second stage in the survey process involves the measurement of offset or ties from these chain lines to pick up the survey detail. *Figure 19.28* illustrates the distinction between offsets and ties. Ties, which involve two measurements to one point, are normally employed when the offset distance is large and hence the accuracy with which the perpendicular to the chain line can be set out is low. Offset and tie measurements are normally made with a fibreglass tape.

A standard method of booking is conventionally adopted for chain surveying. Details of this and other points relating to chain surveying have been given by Irvine[13].

19.3.3 Vertical control surveys

In general, all civil engineering projects require not only planimetric control, but also vertical or height control points. The two most common methods of obtaining this height control are by levelling and by trigonometrical heighting.

Levelling is the name given to the process of determining, by means of a surveyor's level, the height of a point above some datum, normally mean sea level. As is the case with horizontal control, it is important to work from 'the whole to the part'. It is therefore normal practice to design a levelling control framework in a hierarchical manner, in order to contain small errors. Typical maximum allowable misclosures for level loops in one such hierarchical design are shown in *Table 19.4*.

The basic principle of levelling involves taking horizontal backsight (BS), foresight (FS) and intermediate sight (IS) readings, as defined by the line of collimation of the level, onto vertical staves as shown in *Figure 19.29*. The difference between successive readings indicates the difference in height between points. By this process it is therefore possible to determine the reduced level (RL) of a series of points.

The reduced level of a point is defined as its height above some datum. In the UK the fundamental datum established by the

Table 19.4 Accuracy of levelling

Order	*Maximum allowable misclosure* (m)
1st	$0.004\sqrt{K}$*
2nd	$0.008\sqrt{K}$
3rd	$0.012\sqrt{K}$
4th	$0.24\sqrt{K}$

*K is the distance in kilometres

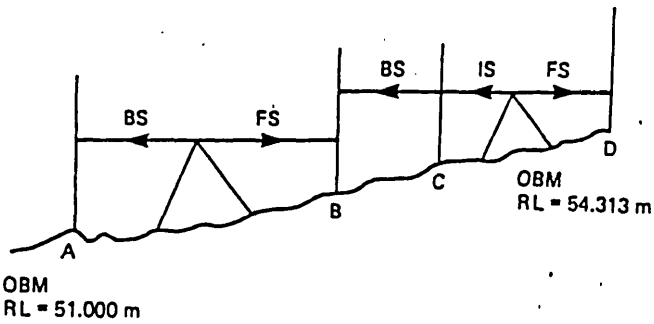


Figure 19.29 Levelling

Ordnance Survey is mean sea level at Newlyn, Cornwall, known as Ordnance Datum. A further series of points of known height have been established throughout the UK and these are known as Ordnance Bench Marks (OBM).

The results of a levelling operation are, by convention, booked in a standard manner. Two methods of booking are used. The rise and fall method (*Table 19.5*) is usually employed when running lines of levels between bench marks in order to establish additional supplementary height points. The height of collimation method (*Table 19.6*), on the other hand, is more suitable for tasks such as recording cross-sectional information, in which many intermediate sights have been taken.

For projects where the acceptable accuracy requirements of the height control are lower than would be obtained by levelling, the method of *trigonometrical heighting* is normally used. The method is based on the measurement of the vertical angle between two points by means of a theodolite, together with either the slope or horizontal distance. By simple geometry, the difference in height can therefore be calculated. As the distance between the points increases, however, the effect of two phenomena, Earth curvature and refraction, become more significant. *Figures 19.30* and *19.31* illustrate the, highly exaggerated, effect of these phenomena.

In *Figure 19.30* a distant point B' appears too low by an amount Δc. A positive Earth curvature is therefore necessary. It can also be seen in *Figure 19.31* that the effect of refraction is to refract the line of sight of the telescope so that a distant point C appears too high by an amount Δr. A negative refraction correction is therefore required.

It can be shown that the difference in height between A and C can be determined by the equation

$$\Delta h_{AC} = d \tan \theta + (\Delta c - \Delta r) \qquad (19.30)$$

where

$$(\Delta c - \Delta r) = d^2 \frac{(1-2K)}{2R} \qquad (19.31)$$

and K = coefficient of refraction $\simeq 0.07$,
R = radius of Earth = 6378 km,
d = horizontal distance,
θ = vertical angle.

Table 19.5 Rise and fall method of booking

	Backsight	*Intermediate sight*	*Foresight*	*Rise*	*Fall*	*Reduced level*	*Remarks*
	2.345					51.000	OBM
	1.935		0.632	1.713		52.713	B
		1.213		0.722		53.435	C
			0.335	0.878		54.313	OBM
$\sum$	4.280		0.967	3.313			

Checks: $\sum$BS − $\sum$FS = 3.313 m. $\sum$rise − $\sum$fall = 3.313 m.
First RL − last RL = 3.313 m

Table 19.6 Height of collimation method of booking

	Backsight	*Intermediate sight*	*Foresight*	* *Height of collimation*	*Reduced level*	*Remarks*
	2.345			53.345	51.000	OBM
	1.935		0.632	54.648	52.713	B
		1.213			53.435	C
			0.335		54.313	OBM
$\sum$	4.280		0.967			

Checks: $\sum$BS − $\sum$FS = 3.313 m. First RL − last RL = 3.313 m
* Height of collimation = reduced level + backsight
RL = HC − FS or IS

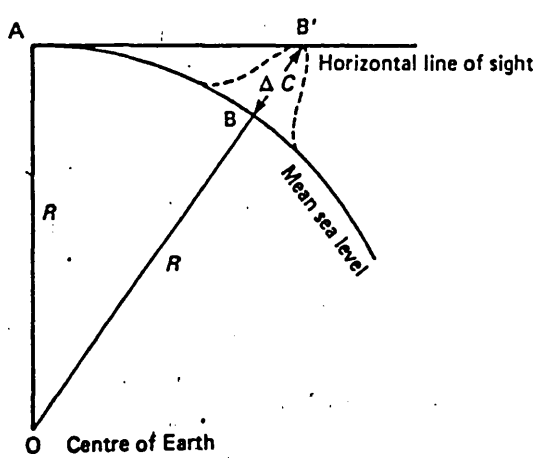


Figure 19.30 Trigonometrical heighting: earth curvature

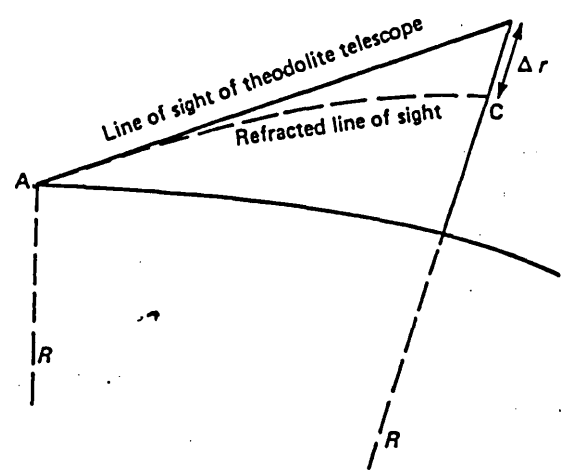


Figure 19.31 Trigonometrical heighting: refraction

If $K = 0.07$ and d is in km, then $(\Delta c - \Delta r) = 0.0675d^2$ m. The results obtained using this simple approach can be increased significantly by observing vertical angles from both ends of a line. Reciprocal observations of this type, particularly if they are observed simultaneously, can completely eliminate the necessity for Earth curvature and refraction corrections. In this case the difference in height between A and C could be computed by the general expression

$$\Delta h_{AC} = d \tan \frac{\theta_A + \theta_c}{2} + \frac{(h_{S_A} - h_{T_C}) - (h_{S_C} - h_{T_A})}{2} \qquad (19.32)$$

where h_{S_A} = height of signal at A,
h_{S_C} = height of signal at C,
h_{T_A} = height of theodolite at A,
h_{T_C} = height of theodolite at C.

19.3.4 Further uses of levelling

A *contour* is an imaginary line joining points of equal elevation. The level can be used in a variety of ways for contouring. One of the simplest methods is by means of grid levelling. With this technique, the area to be contoured is covered by an imaginary grid of lines forming squares of 10, 20 or 30 m. The level is set up in a central position and levels are then taken to a site temporary bench mark (TBM) and at the intersections of the grid lines, as shown in *Figure 19.32*. The grid can be set out by various methods. The method illustrated in *Figure 19.32* involves setting out two lines of ranging rods along adjacent sides of the grid. The staffman can then line himself in and place the staff at the appropriate intersection point. Contours may then be interpolated either graphically or mathematically from the grid of levels.

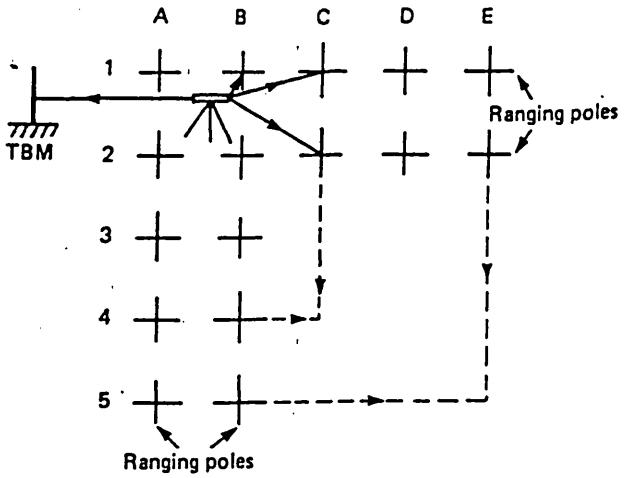


Figure 19.32 Grid levelling

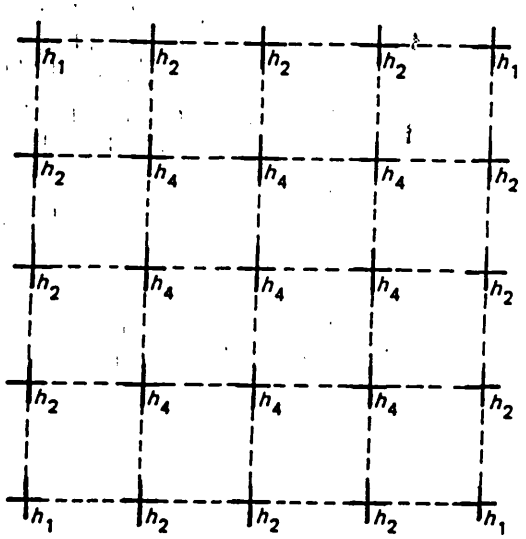


Figure 19.33 Quantity determination

The method of grid levelling provides a convenient means of determining earthwork quantities of, for example, borrow pits. The depth of cut (h) at each intersection point is established. This will be equal to the difference between the ground level and the proposed formation level of the borrow pit. *Figure 19.33* illustrates a simple example for a grid of levels. The volume of excavation consists of a series of rectangular prisms each having a base area (A). The total volume (V) for the general case is therefore:

$$V=\frac{A}{4}(\sum h_1+2\sum h_2+3\sum h_3+4\sum h_4) \qquad (19.33)$$

where h_1 = depth of cut used once,
h_2 = depth of cut used twice,
etc.

The volume determination assumes that the ground slope between grid intersections is constant. It is therefore evident that, by reducing the size of the grid, the accuracy of the quantity determination will be increased.

19.3.5 Deformation monitoring surveys

Deformation monitoring surveys using conventional land surveying methods, photogrammetric survey methods or specialist geotechnical methods are used to quantify the amount by which an engineering structure has moved both vertically and horizontally over specific periods of time.

The information which surveys of this type provide can be of critical importance and may be used to indicate that:

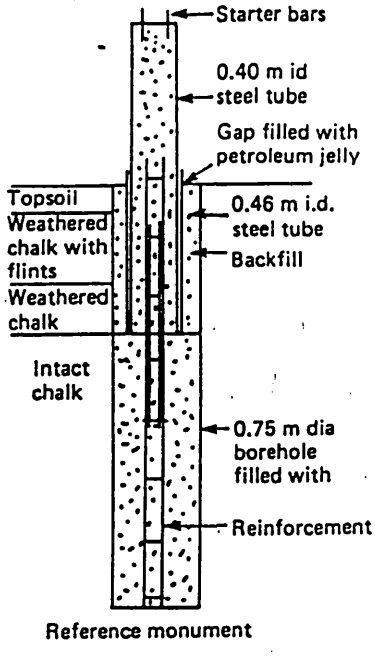
(1) the structure is stable and consequently safe;
(2) the structure is experiencing small random movements which are not imposing significant forces on the structure;
(3) the structure is experiencing small localized systematic deformations, for example as caused by seasonal effects, which may or may not be significant;
(4) lastly, and most significantly, the structure is experiencing deformation which is increasing as a function of time. This may indicate that remedial measures have to be taken, in some cases immediately, in order to avoid catastrophic consequences.

For a variety of reasons surveys of this type have become increasingly important in recent years. Several interrelated factors have been responsible for this increase in interest. Firstly, for many types of structures it is now a mandatory element of the civil engineering process that a monitoring survey be commissioned. Notable in this respect in this country are, for example, reservoirs, which now have to be monitored following the implementation of the 1975 Reservoirs Act. A further example, from Switzerland, is the requirement by the Swiss Federal Government for all dams with a height greater than 15 m and a cubic capacity greater than 50 000 m^3 to be monitored[14]. A second factor which has led to this increase in interest has been the speed of development both in the manufacture of precise survey instrumentation and in high-speed, low-cost computing facilities. This, in conjunction with improvements in very elegant and highly sophisticated software for the design and analysis of surveying observations, now provides the surveyor and engineer with a highly accurate measurement system. Thirdly, the tolerances to which civil engineers are now designing and constructing many modern structures necessitates a much higher order of accuracy in the initial dimensional control and also in the subsequent deformation monitoring. Indeed, this factor may have acted as a springboard for many of the developments in software and instrument design. It is also hoped that to some extent civil engineers are now more aware of the possibilities offered by modern survey techniques and are therefore requesting this type of survey more frequently than was the case in the past.

Most of the instruments used for deformation monitoring projects have been discussed in previous sections of this chapter. It will therefore be the aim of this section to concentrate on two other important aspects of deformation monitoring surveys. The first is the design of suitable reference and monitoring points. Clearly when very small displacements are being measured it is crucial that the points from which measurements are being recorded are not subject to movement. The second aspect which will be discussed is the computational process associated with the horizontal control networks which are often used to quantify the extent of any structural deformation. Finally it should be noted that deformation monitoring can often be carried out, in some cases more efficiently, using close-range photogrammetric techniques. Reference should be made to Chapter 21 for further details.

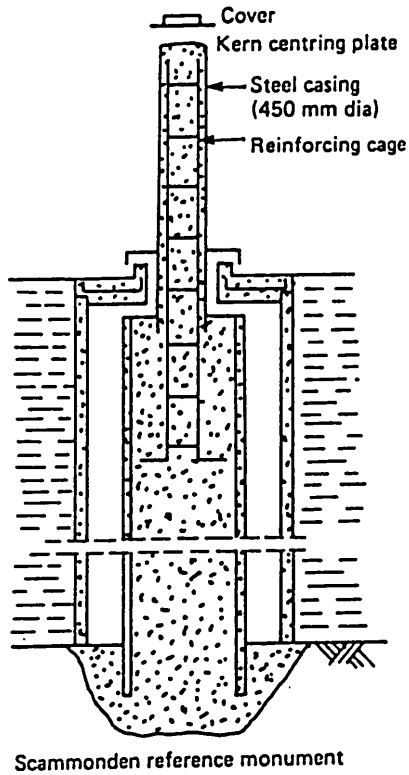
The full accuracy potential currently offered by modern surveying equipment for measuring small structural displacements can only be fully realized if care and attention is paid to the design of the reference points from which observations will be taken. Equally important is the need to design appropriate monitoring points to be placed on the structure under investigation.

Two distinct types of survey reference point can be identified. The first, typically a survey pillar, forms the reference framework



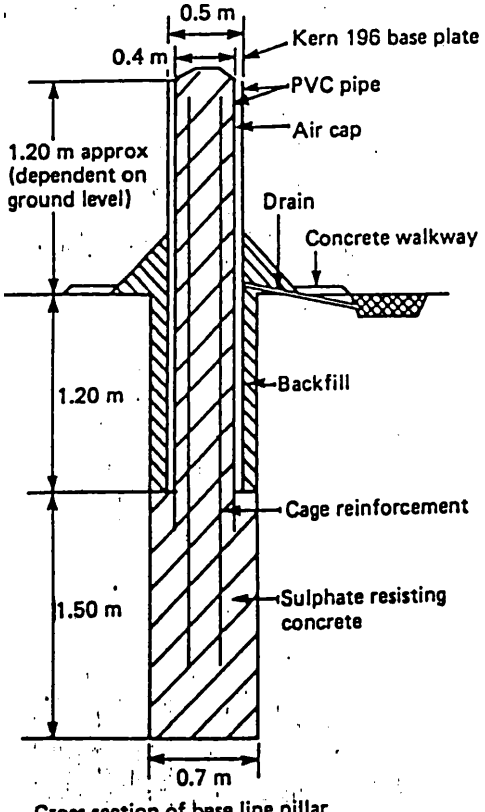
Reference monument

(a)



Scammonden reference monument

(b)



Cross section of base line pillar

(c)

Figure 19.34 Survey pillar designs: (a, b) from Penham and Charles[28]; (c) from Deeth *et al.*[29]

the horizontal and vertical control measurements, whilst the nd, a steel or concrete pile, forms the datum for levelling servations.

hree distinct designs of survey pillar are illustrated by *Figure 4*. The main requirement in siting pillars is the need to ensure t they are founded on stable ground outside the zone of uence of the structure under investigation. It is, however, n difficult to assess whether this is the case before rvations are recorded. It is therefore important to rporate into the design an insulating gap which ensures that the central pillar is not in contact with the surrounding ground. This should ensure that the effects of diurnal or seasonal earth movements are minimized.

A further common design feature is the incorporation of some system of forced centring. The two most common forced centring systems for deformation monitoring are the Wild ball centring system and the Kern system. Both designs are illustrated by *Figure 19.35*. Use of either system should ensure that errors from this source do not exceed ± 0.1 mm.

This concept of an insulating sleeve around the reference point

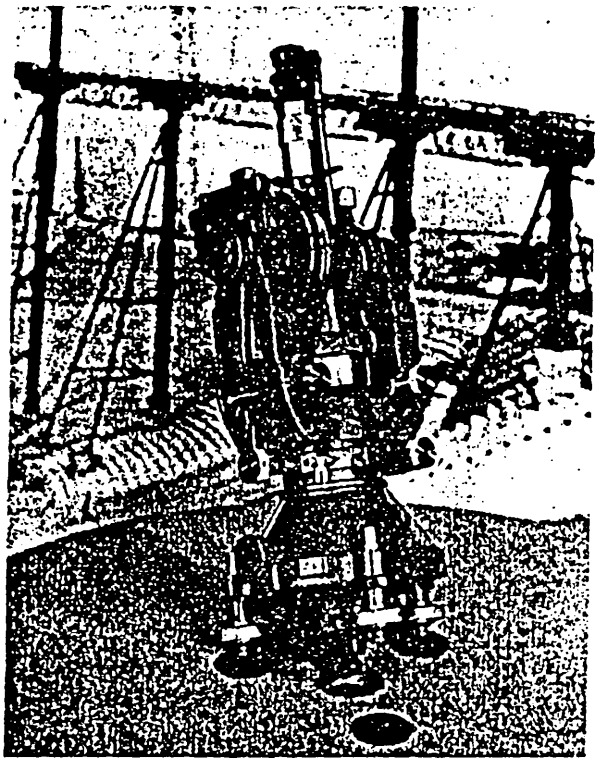
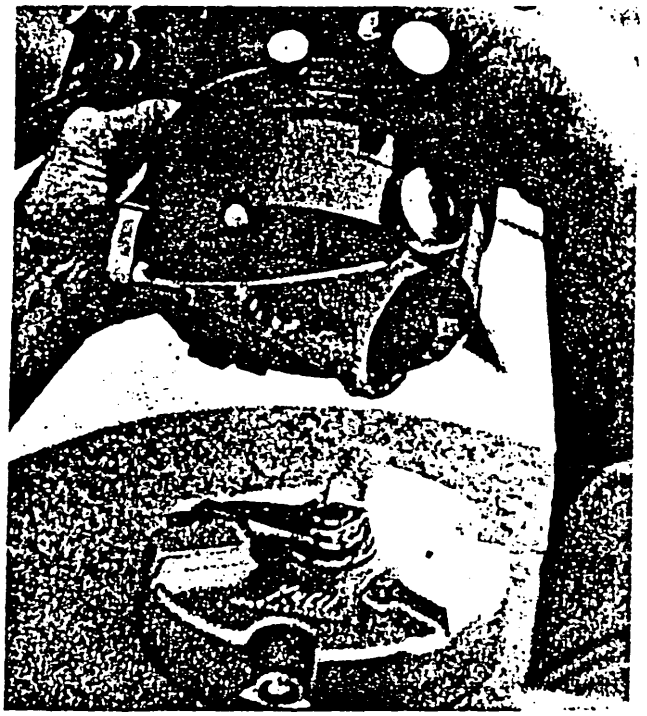


Figure 19.35 Pillar centring systems: (a) Wild ball centring system



(b) Kern centring system

is also evident in the levelling datum designs illustrated by *Figure 19.36a*. *Figure 19.36b* can be seen to consist of a steel foot driven into the ground at the bottom of a 5 m deep borehole. This steel foot is connected to a central rod which extends almost to ground level. To minimize the potential effect of ground movement influencing the datum, the central rod is surrounded by a telescopic sleeve. An alternative design is illustrated by *Figure 19.36b*. In this case the inner datum rod consists of a 10 mm bore galvanized tube surrounded by a guard tube of 25 mm bore. The guard tube is driven into a 150 mm diameter hole which has been bored to a depth of 5.5 m. The datum point consists of a dome-shaped steel ball about 0.1 m below ground level. The ball is covered by a protective cap and by a sealing plate (possibly a manhole cover) at ground level. Further details relating to the installation procedure are given by Cheney[15]. It is also important to install a sufficient number of datums to enable any settlement or uplift of the datums to be detected. Ideally three should be installed.

The main requirement of a survey monitoring point is that it can be either permanently affixed to or precisely relocated on the structure being investigated. Again, two distinct types of monitoring point can be identified; those to which angle/distance measurements will be taken and those which will be used as precise levelling settlement points.

The measurement of angles and distances to monitoring points on the structure will normally require that the target points be designed so that they are capable of accepting both conventional survey targets and also corner cube reflectors. The simplest approach, and that commonly used for dam deformation work, is to build a series of pillars on the structure. The pillars are fitted with a suitable forced centring system which enables both survey targets and corner cube reflectors to be interchanged very accurately. An example of such a system is illustrated by *Figure 19.37*. An alternative approach may be used in situations where the distances to be measured are short. In these cases it may be possible to affix targets and reflectors to the structure permanently. Details of one particular arrangement which involves the use of reflex acrylic reflectors is reported in Kennie[16]. Cheney[17] also discusses the use of permanently mounted reflectors, in this instance for use with the Kern Mekometer ME 3000.

For settlement measurements one of the most common designs of monitoring point is that which has been designed by the Building Research Establishment (BRE). The components of the settlement system are illustrated by *Figure 19.38*. It can be seen that the system consists of four components: a stainless steel socket 65 mm long by 22 mm diameter which is grouted into the structure, a detachable settlement bolt, a protective Perspex cap and an alloy wrench to enable removal of the cap. The main advantages of such a system are, firstly, a high degree of accuracy in relocating the settlement point. Cheney[15] states that the settlement bolts may be repositioned in the socket to within 0.03 mm. The second advantage of such an arrangement is the level of protection offered to the settlement point from either accidental or other damage.

Horizontal control surveys for deformation monitoring purposes are essentially no different from control surveys for other purposes (e.g. national control surveys). Whilst they may differ in size (generally much smaller) and accuracy requirements (generally much higher), the basic principles in terms of observational and computational techniques are common. It is not surprising to note, therefore, that in recent years the type of control survey generally adopted for deformation purposes consists of a mixed set of angle and distance observations or a 'survey network'. The advantages of such an approach in reducing scale and orientation errors have been discussed previously. However, the primary advantage of this approach occurs when the data is subjected to a rigorous least squares adjustment using the method of variation of coordinates (see Section 19.3.1). This not only enables the most probable values of the positions to be determined, but it also enables statistical data about the precision and reliability of the network to be determined. Furthermore, by interpreting these statistical indices and the results of a network adjustment from two different epochs it is possible to evaluate the statistical significance of any changes which are observed. The various

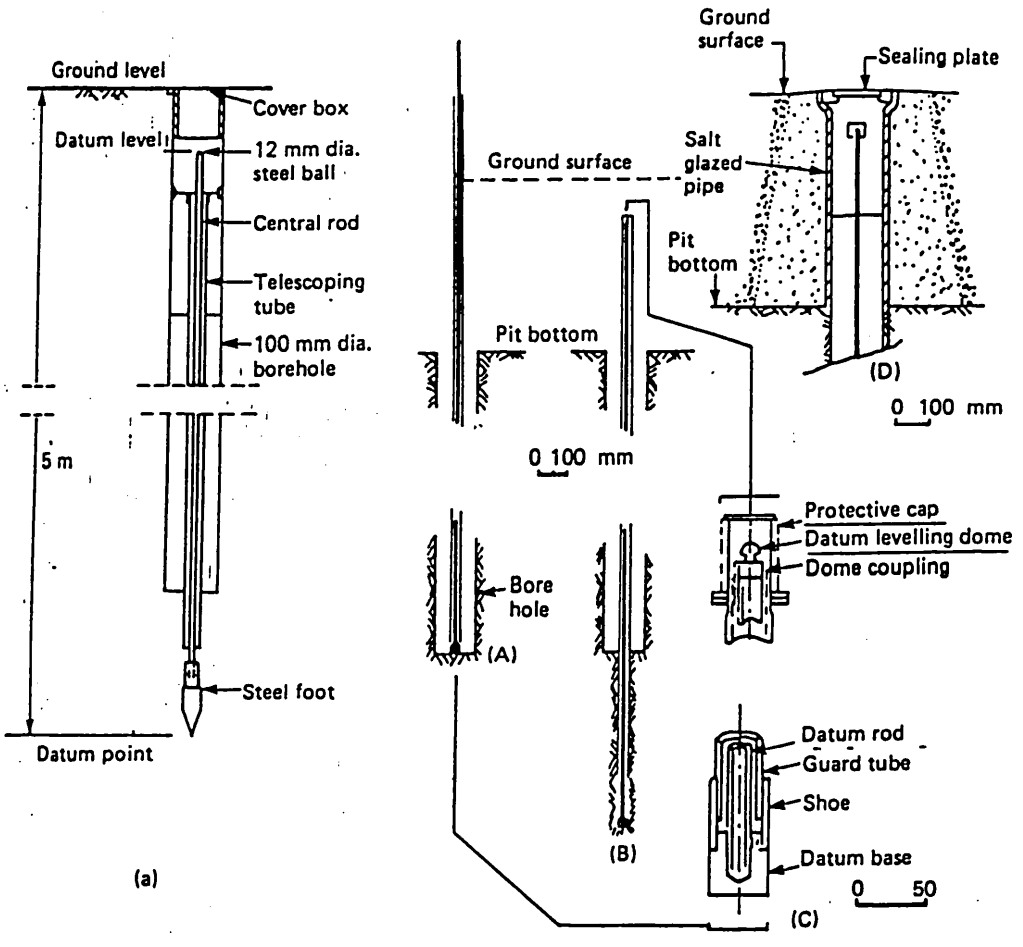


Figure 19.36 Levelling datum designs: (a) Penham and Charles[28], (b) Cheney[15]

stages in the computational process are shown in *Figure 19.39*.

The various statistical indices which can be derived from a variation of coordinates adjustment are well documented in the literature (e.g. Cooper[1]). It is therefore intended to only briefly mention the primary features of each statistical index.

Three distinct types of statistical indices can be identified; firstly, those concerned with network precision; secondly, those concerned with network reliability; and, thirdly, those for deformation analysis.

Network precision:

(1) The *residuals* (v) of the observation equations are obtained from

$$\mathbf{v} = A\mathbf{X} - \mathbf{b} \tag{19.34}$$

where A, $\mathbf{x}$ and b are as defined in Equation (19.27). In cases where a large number of observations exist, the ratio v/σ, where σ refers to the standard error of the observation, may be used in order to reject suspect observations. For example, if v/σ is greater than 2.5, this indicates that there is only a 1.2% probability that this variation is caused by random observational effects and is more likely to be indicative of a poor observation. On this basis, therefore, the observation may be rejected.

(2) *Unit variance* (σ_0^2) is computed from the expression

$$\sigma_0^2 = \frac{\mathbf{v}^T W \mathbf{v}}{m-n} \tag{19.35}$$

where m is the number of observations, n the number of unknowns and W is the weight matrix. It can be shown that, theoretically, σ_0 (the standard error of unit weight) should equal 1. A large departure from unity is usually indicative of some gross blunder, whilst a small departure may give an indication whether the initial *a priori* standard errors of the observations have been correctly estimated. In certain circumstances it is possible to correct the initial standard errors by a trial and error process in order to bring the value of σ_0 closer to unity.

(3) The *variance–covariance matrix* (σ_{xx}) is an extremely useful by-product of the adjustment process. It is computed from the following expression:

$$\sigma_{xx} = \sigma_0^2 (\mathbf{A}^T \mathbf{W} \mathbf{A})^{-1} \tag{19.34}$$

The diagonal elements of the matrix ($\sigma_{E_i}^2, \sigma_{N_i}^2 \ldots$), refer to the variances or squares of the standard errors associated with the unknowns E_i, N_i, etc. It is therefore an extremely useful means of measuring the precision of the network. The

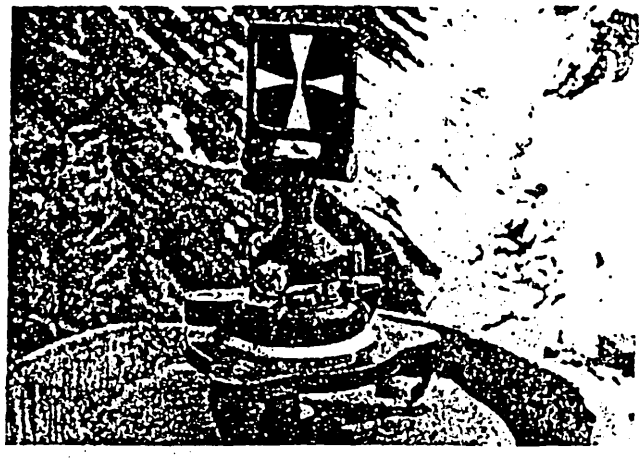


Figure 19.37 Pillar and forced centring system for theodolite target

off-diagonal elements or covariances ($\sigma_{E_iN_i}\ldots$) may be used in order to construct error ellipses.

(4) The 'absolute *error ellipse*' of a calculated point is an ellipse defined by σ_{min} and σ_{max} with the semi-major axis $a=\sigma_{max}$ and the semi-minor axis $b=\sigma_{min}$. When drawn around a point the ellipse is generally interpreted as depicting the region within which there is 39% confidence that it contains the position of the point. The values of σ_{max} and σ_{min} are determined by the following expressions:

$$\sigma_{max}=(\sigma_E^2+\sigma_N^2)/2+[(\sigma_E^2-\sigma_N^2)/4+\sigma_{EN}]^{1/2} \quad (19.35)$$

$$\sigma_{min}=(\sigma_E^2+6_N^2)/2-[(\sigma_E^2-\sigma_N^2)/4+\sigma_{EN}]^{1/2} \quad (19.36)$$

Two drawbacks exist with error ellipses of this type. Firstly, the size of the error ellipse is not 'invariant' and depends on the position of the point which is chosen as the origin. Secondly, no information is given about any inter-station correlation which may exist. An alternative type of error ellipse which overcomes the first drawback is the 'relative error ellipse'. In this case the ellipse is computed on the basis of the variances and covariances of the differences in coordinates between points. It is therefore a measure of the relative positional accuracy between the points.

(5) The *a posteriori* or post-adjustment *standard errors* of the adjusted quantities are a further set of 'invariant' statistics which can be computed. For example, the *a posteriori* standard error associated with the adjusted distance between any two points in the network is given by

$$\sigma_l^2=[P\ Q\ R\ S]\begin{bmatrix}4\times4\\ \text{sub-matrix}\\ \text{of }\sigma_{xx}\end{bmatrix}\begin{bmatrix}P\\Q\\R\\S\end{bmatrix} \quad (19.37)$$

where P, Q, R and S are the trigonometric functions defined in Equation (19.22b).

Network reliability in this context is defined as the ability of the network to detect gross errors in the observed quantities. Ashkenazi[18] has suggested that the following statistic should be used in order to assess the reliability of the network:

$$\text{reliability}=\frac{a\ posteriori\ \text{standard error of the observation}}{a\ priori\ \text{standard error of the observation}} \quad (19.38)$$

Theoretically, the greater the number of observations indirectly affecting a particular observation, the lower the ratio should be. Thus low ratios are indicative of high reliability, whilst a ratio of one would indicate complete unreliability. Ashkenazi suggests that the maximum acceptable value of this ratio should be 0.9.

Deformation analysis. The statistical indices discussed so far relate to a single set of network observations. A much more common problem in deformation monitoring work, however, is the situation where two or more sets of observations of the same network, taken at different times, exist. In these circumstances further analysis is required to determine whether any difference in coordinates which exist are statistically significant.

Currently this particular problem is the subject of considerable research interest. Following the second FIG symposium on Deformation Measurements in Bonn, 1978, a committee was established to investigate different approaches into the analysis of deformation measurements using the same measurement data. The provisional results of this committee's work were reported

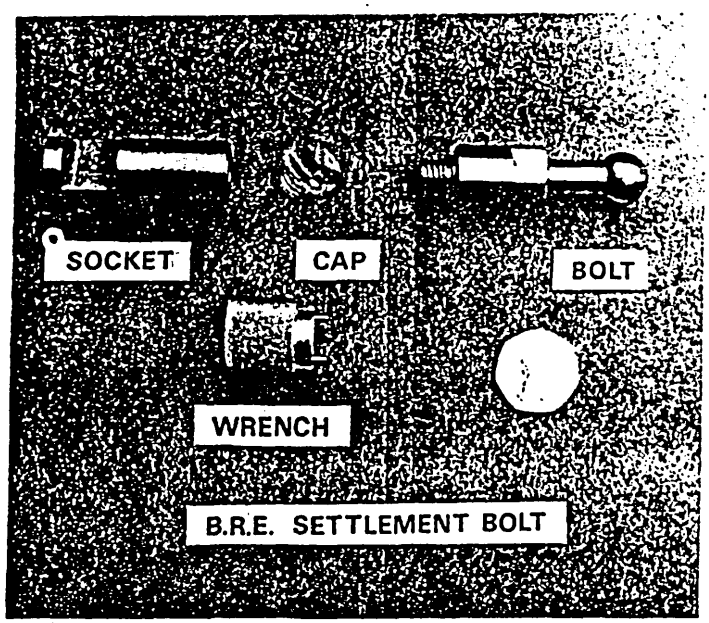




Figure 19.38 (a) Components of BRE settlement bolt system; (b) settlem

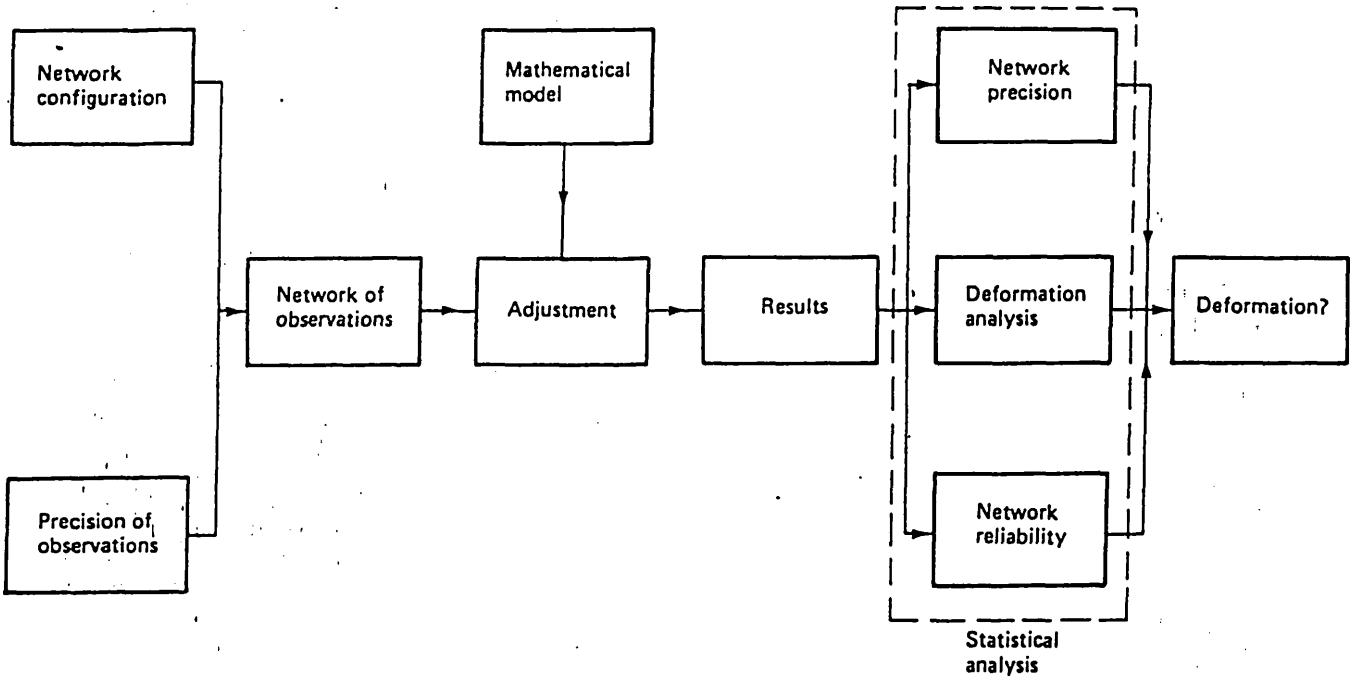


Figure 19.39

by Chrzanowski[19] and subsequently expanded in Chrzanowski *et al.*[30]. The primary problems which this group, and others, have been examining are, firstly, what is the most appropriate means of eliminating systematic effects from the data sets and, secondly, what are the most appropriate types of statistical test to apply to the recorded deformations in order to test their significance.

The first problem, that of eliminating systematic effects, such as a scale bias in EDM measurements, has been discussed by Ashkenazi *et al.*[20] and Ashkenazi[18]. In cases where a systematic bias is considered to be present in one set of data, the systematic bias parameters (scale, orientation and translation) may be determined by means of a least squares four parameter Helmert transformation. By the use of this procedure it is therefore possible to eliminate from one data set a series of systematic effects, which if not eliminated may be mistaken as systematic deformations.

The main danger associated with the application of this technique to networks is that it may lead to the removal of a genuine systematic deformation under the mistaken belief that it is caused by observational error. It has therefore been suggested by Dodson[21] that it is preferable for any systematic effects to be eliminated by careful instrument calibration, or by designing the network so that 'the influence of such effects does not influence the acceptance or rejection of any hypothesis of deformation'.

The second problem, that of devising the most appropriate statistical tests, has also been examined by Ashkenazi and Dodson[22]. The approach they have adopted involves comparing the detected coordinate differences with the corresponding coordinate standard errors. The basis of this approach can be illustrated by considering two points P and Q in a fictitious network. If the position of P in the network is stronger (i.e. more accurately determined because of a good geometrical configuration and higher precision observations) than point Q, then quite clearly a smaller difference between two successive values of the coordinates of P may be more significant than a larger difference corresponding to point Q. Thus periodic coordinate differences may only be considered to be significant if the standard error associated with the deformation at that point is also examined. Therefore if

$$\Delta^2 = (\Delta E^2 + \Delta N^2)^2 \qquad (19.39)$$

where Δ is the computed deformation based on planimetric coordinate differences ΔE and ΔN, and the standard error in Δ is given by

$$\sigma_\Delta^2 = (\sigma_{\Delta E}^2 + \sigma_{\Delta N}^2)/2 \qquad (19.40)$$

where

$$\sigma_{\Delta E} \simeq \sqrt{2}\sigma_E \quad \text{and} \quad \sigma_{\Delta N} \simeq \sqrt{2}\sigma_N \qquad (19.41)$$

then the ratio Δ/σ may then be used in order to test the significance of any deformation. For example, if the ratio Δ/σ is 2 (95% probability level), there is only a slight chance (19 to 1 against) that the movement has been caused by random observational errors and it is much more likely to have been caused by movement of the point. In contrast, a ratio of 0.7 would indicate that there was an even chance (50% probability level) that the difference was caused by movement, it being equally likely that the variation was simply a manifestation of a random observational error. Ashkenazi *et al.*[23] discusses the application of this technique to sets of data observed at Cattleshaw reservoir.

19.4 Modern developments

Technological change is an ever-increasing challenge for the engineering community, surveying being no exception. In recent years many traditional techniques and instruments have been replaced as technological developments have taken place. Electronic tacheometers, digital mapping systems and lasers are three of the most significant of these developments.

19.4.1 Electronic tacheometers

'Electronic tacheometer' is the term which has been adopted to describe instruments which combine an EDM unit with an electronic theodolite. The Wild T2000/DI4 (*Figure 19.40*) is typical of this type of instrument. In addition to being able

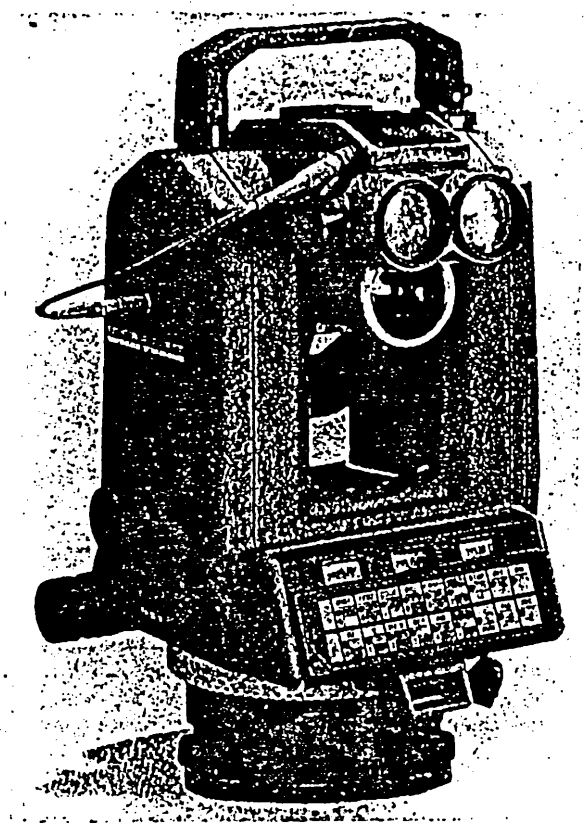


Figure 19.40 Wild T2000 and DI4

automatically to measure angles and distances electronically, the instrument also has the facility to automatically record distance, angle and coordinate information directly onto the GRE3 data collector. The main advantages of this approach are that, firstly, the speed of data capture is increased and, secondly, the field data is available in a form which is computer compatible and can therefore be directly input (via the GLE cassette reader) into a digital mapping system for processing and plotting.

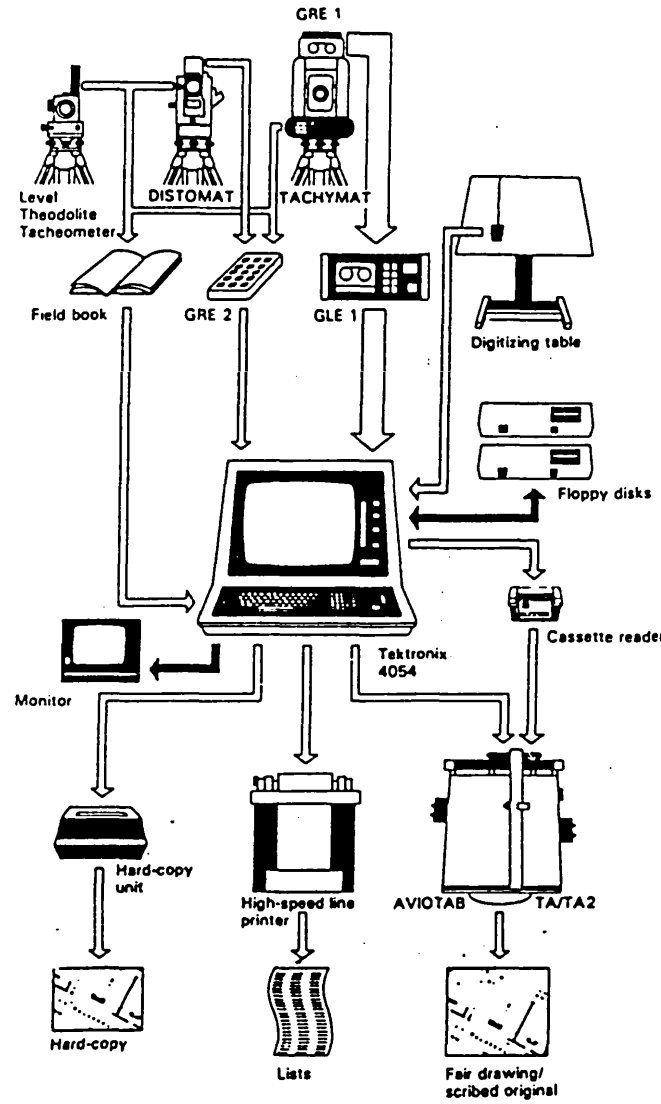


Figure 19.41 Wild GEOMAP system (Wild Heerbrugg)

19.4.2 Digital mapping systems

The use of computers for computational purposes in surveying is well established. Recently, however, the development of relatively low-cost interactive graphics facilities, together with flat bed and drum plotters, has led to the production of several digital mapping systems.

Many systems are currently in use. They range from large government systems, such as that developed by the Ordnance Survey[24], through large commercial systems based on mainframe computers such as MOSS, to minicomputer- and microcomputer-based systems, such as the Wild GEOMAP[25]. Several survey organizations have also developed their own in-house systems.

The Wild GEOMAP system illustrates many of the fundamental features of the digital mapping concept. *Figure 19.41* shows the main hardware components of the system. Data input is flexible, although biased towards the Wild TC1 electronic tacheometer. Field data from the TC1 can be input directly into the processor by means of the GLE 1 cassette reader. Other field data can be manually input via the standard keyboard of the computer. In some cases the input to the system may be in the form of existing maps or plans. By using a tablet digitizing table, this graphic information can be converted into a digital format and input into the system. Data processing is carried out by a Tektronix 4054 minicomputer. The computer has 64K of memory, twin floppy disks for storage of data and a high-resolution CRT screen. The software for the system consists of three interrelated software packages. The first allows the user to edit the input data in order to obtain a clean data file. The second consists of several application packages for survey computational purposes; for example, traverse or triangulation network adjustment, formation of digital ground models and computation of areas and volumes. A third software package controls the interactive graphics facilities. Using the clean input data file, the plane is created on the CRT screen. By means of the interactive graphics facilities, information can be added or deleted and the corrections automatically executed on the screen. When this final editing phase is complete, the resultant plan can be automatically drawn by a digital plotter such as the Wild Aviotab TA. Data output in the form of hard copy from the CRT display is also possible.

19.4.3 Lasers

Lasers are monochromatic, coherent and highly collimated light sources, initially developed in the 1940s. Until relatively recently, their use has tended to be restricted to the field of pure scientific research. Nowadays, however, the laser is a widely used tool in land surveying for distance measurement, alignment and levelling purposes.

There are essentially two types of laser in use in civil engineering, namely the fixed-beam and the rotating-beam laser. The fixed-beam laser projects a single highly collimated light beam to a single point. This type of laser is extremely useful for alignment purposes. The rotating-beam laser, by contrast, takes the fixed-beam source and rotates it at high speed, so forming a plane (horizontal or vertical) of laser light. This type of laser is useful for levelling or grading purposes.

For projects such as tunnelling, where alignment is a critical problem, the laser is an extremely useful device. The Kern K0-S2 laser theodolite, as shown in *Figure 19.42*, may be taken as an example of this type of laser. The instrument is a combination of a Kern K0-S theodolite and a separate low-power HeNe gas laser, strapped to one of the tripod legs. The parallel rays of the laser are collimated to a beam width of 0.08 mm and transmitted via a fibre optic cable to the telescope of the theodolite which is fitted with a special eyepiece attachment.

Within the telescope, the light is focussed on a reticle which marks out a cross. The laser beam then passes through a beam-splitting cube which directs the laser along the optical axis of the telescope. The projected cross may be brought to focus onto a target by focussing the theodolite telescope. Under favourable conditions with suitable targets, a range of 400 m is possible.

The laser eyepiece may also be attached to a conventional optical plummet to provide a vertical reference line. A Wild GL02 eyepiece attached to a ZNL plummet was used in the construction of the Dinorwic pumped storage scheme[26].

The fixed-beam laser can also be used for pipelaying. The Spectra-Physics Dialgrade laser, shown in *Figure 19.43*, is typical of this type of instrument.

A rotating-beam laser, the Spectra-Physics EL-1 is shown in *Figure 19.44*. The laser beam in this case forms a 360° horizontal plane which can be picked up by a portable sensing device also shown in *Figure 19.44*. The laser unit will automatically correct for any dislevelment of the instrument, providing it has been roughly levelled to within ±8° of the vertical. An accuracy of ±5–6 mm per 100 m up to a maximum range of 300 m can be achieved with this type of instrument.

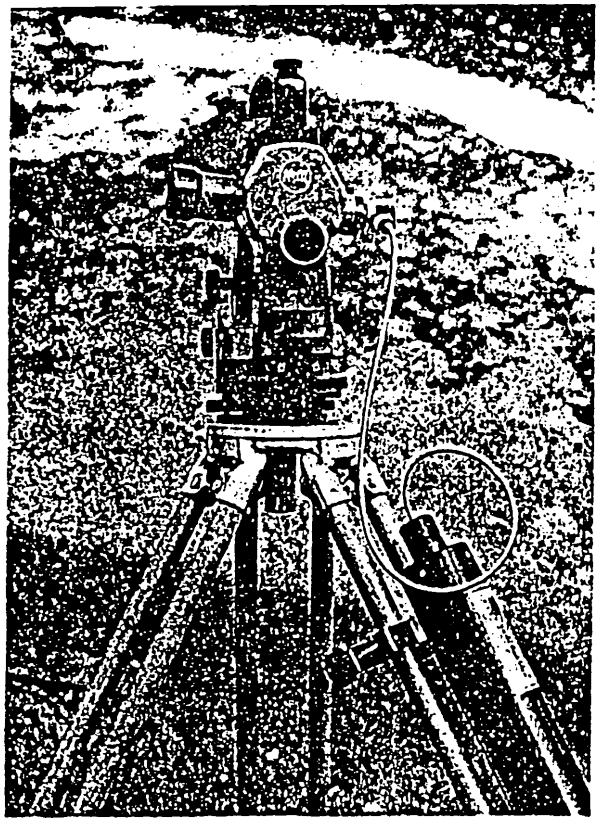


Figure 19.42 Kern K0-S2 laser theodolite (Survey and General Instruments).

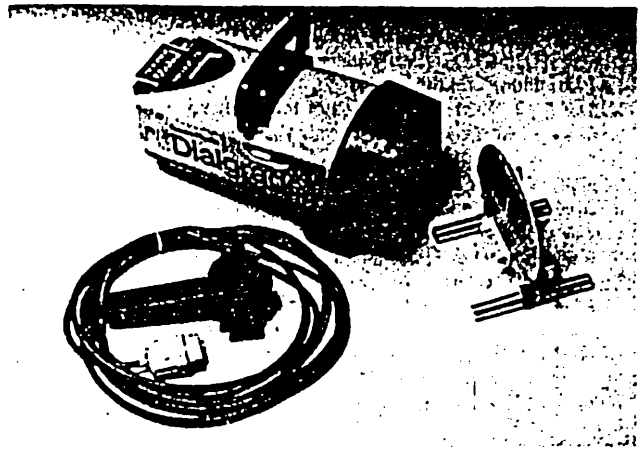


Figure 19.43 Spectra-Physics Dialgrade laser (Spectra Physics)

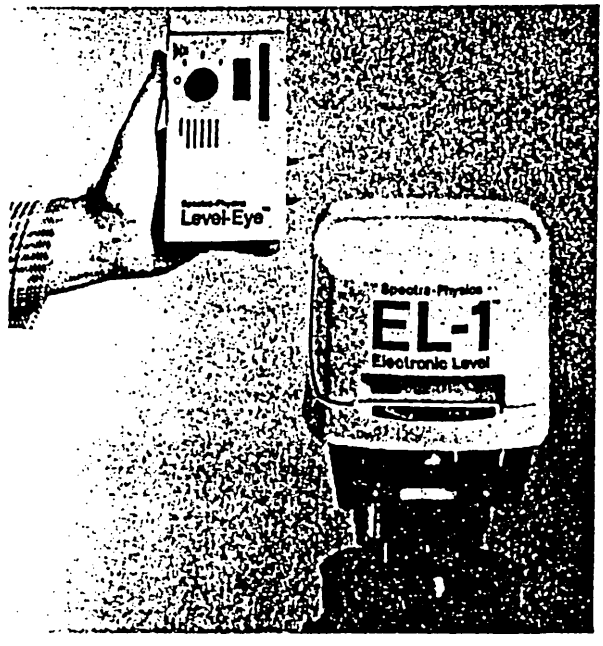


Figure 19.44 Spectra-Physics EL-1 (Spectra-Physics)

A novel application of the laser levelling principle is site grading[27]. Laser levelling sensors attached to conventional scrapers pick up the laser signal being emitted from the laser unit which is placed on site. The sensor itself is mounted on the blade of the scraper, and by signals which activate the hydraulic system of the scraper, the blade is automatically raised or lowered to the appropriate height in order to ensure that the correct grade is obtained.

References

1 COOPER, M. A. R., *Fundamentals of Survey Measurement and Analysis*, Crosby Lockwood Staples, London (1974)

2 COOPER, M. A. R., *Modern Theodolites and Levels*, 2nd Ed., Granada, London (1982)

3 MILLER, R. M., 'Accuracy of measurement with steel tapes', *Building Research Paper CP51/69*, Building Research Station, Watford (1969)

4 SMITH, J. R., *Optical Distance Measurement*, Crosby Lockwood Staples, London (1970)

5 HODGES, D. J. and GREENWOOD, J. B., *Optical Distance Measurement*, Butterworths, London (1971)
6 BURNSIDE, C. D., *Electromagnetic Distance Measurement*, 2nd Ed., Granada, London (1982)
7 KENNIE, T. J. M., 'Some tests of retro-reflective materials for electro-optical distance measurement', *Survey Review*, 207, 3–12 (1983)
8 SCHWENDENER, H. R., 'Electronic distancers for short range: accuracy and checking procedures', *Survey Review*, 164, 273–281 (1972)
9 SCHOFIELD, W., 'The effect of various adjustment procedures on traverse networks', *Civ. Engng. Surv.*, 4, 4, 13–19 (1979)
10 LAM, C. P., 'Chaks rule—a method for traverse adjustment', *Civ. Engng. Surv.*, 3, 5, 20–25 (1978)
11 ASHKENAZI, V., 'The solution and analysis of large geodetic networks', *Survey Review*, 146, 166–173 (1968)
12 REDMOND, F. A., *Tacheometry*, The Technical Press Ltd., London (1951)
13 IRVINE, W., *Surveying for Construction*, McGraw Hill, London (1974)
14 EGGER, K., 'Geodetic measurement and the unusual behaviour of the Zeuzier arch dam', *Land and Min. Surv.*, 1, 10, 15–21 (1983)
15 CHENEY, J. E., 'Techniques and equipment using the surveyors level for the accurate measurement of building movements', *Proc. of the British Geotechnical Society Symposium on Field Instrumentation*, London (1973)
16 KENNIE, T. J. M., 'The use of acrylic retroreflectors for monitoring the deformation of a bridge abutment', *Civ. Engng. Surv.*, 9, 6, 10–15 (1984)
17 CHENEY, J. E., 'Some requirements and developments in surveying instrumentation for civil engineering monitoring', *Proc. of the FIG Commission 6 Symp. on Engineering Surveying*, London (1980)
18 ASHKENAZI, V., 'Least square adjustment: signal or just noise', *Chartered Land Surveyor*, 3, 42–49 (1981)
19 CHRZANOWSKI, A., 'A comparison of different approaches into the analysis of deformation measurements', *Proc. of the FIG Conference*, Montreux (1981)
20 ASHKENAZI, V., DODSON, A. H. and CRANE, S. A., 'Monitoring deformations to millimetre accuracy', *Proc. of the FIG Commission 6 Symp. on Engng. Surv.*, London (1980a)
21 DODSON, A. H., 'Pre-analysis and design of a measurement scheme', *Land and Min. Surv.*, 2, 1, 13–19 (1984)
22 ASHKENAZI, V. and DODSON, A. H., *Measuring Deformations by Surveying Techniques*, Seminar, University of Nottingham (1978)
23 ASHKENAZI, V., DODSON, A. H., SKYES, R. M. and CRANE, S. A., 'Remote measurement of ground movements by surveying techniques', *Civ. Engng. Surv.*, 5, 4, 15–22 (1980b)
24 LOGAN, I. T., 'Ordnance Survey digital mapping', *Civ. Engng. Surv.*, 6, 6, 13–23 (1981)
25 SCHENK, T. and HENAUT, M., 'Wild GEOMAP interactive graphics surveying system', *Civ. Engng. Surv.*, 5, 9, 21–23 (1980)
26 MURRAY, G. A., 'Lasers and Dinorwic', *Civ. Engng. Surv.*, 5, 6, 6–11 (1980)
27 ANON., 'Britain sets its sites on lasers', *Cons. Engr.*, 45, 1, 39–40 (1981)
28 PENHAM, A. D. M. and CHARLES, J. A., 'Measuring movements of engineering structures', *Building Research Station, Current Paper CP 32/71* (1971)
29 DEETH, C. P., DODSON, A. H. and ASHKENAZI, V., 'EDM: accuracy and calibration', *Symp. on EDM, Polytechnic of the South Bank*, London (1979)
30 CHRZANOWSKI, A., YONG-QI, C. and SECORD, J. M. 'On the analysis of deformation surveys. *4th Can. Symp. Mining Survey. and Deformation Movements*, Banff (1982)

1 Electronic angle and distance measurement

T.J.M. KENNIE

1.1 Introduction

The measurement of angles and distances is central to almost all land surveying activities, since from these observations it is possible to derive the relative positions and heights of points. Historically, angles have been measured using theodolites equipped with engraved metal or glass circles, and accuracies of a single second of arc have been achieved with such instruments. Recent developments in microelectronics, however, now enable these reading mechanisms to be replaced by electronic components. Also, the inclusion of microprocessors into the instrument design now allows a less rigorous observational procedure to be adopted without reducing the accuracy of the angular measurement. The basic principles and construction of these electronic theodolites are discussed in section 1.2.

Similarly, methods of distance measurement have changed dramatically in the recent past. Section 1.3 therefore examines the basic principles of measurement using electromagnetic waves. This section also examines the vitally important process of calibration. This has become of even greater significance with the birth of the surveying 'black box', an instrument which automatically outputs numerical values of distance and angle. Methods of independently checking the accuracy and integrity of such data are of crucial importance if systematic measurement errors are to be avoided.

The development of electronic surveying instruments which enable three-dimensional coordinates to be derived, either in the field or in the laboratory are also discussed. In the first case, the ability to electronically measure and store data has had a major impact on productivity levels in surveying. In the second case, the determination of coordinates to an accuracy of a fraction of a millimetre, over short ranges, has led to the greater use of surveying in the field of industrial measurement.

1.2 Principles of electronic angle measurement

The overall design of the traditional optical theodolite has remained virtually unchanged for the past 25–30 years. Although over this period improvements in materials and optical machining tolerances have enabled the instruments to become smaller and lighter and the optics more efficient, the general design has not altered significantly. A major change has, however, occurred in recent years with the introduction of electronic circle reading systems and electronic tilt sensing systems into

the design of the theodolite. While electronic theodolites which incorporate such features have been available for the past 10–15 years, it is only in the 1980s that the cost of these instruments has become comparable with their optical counterparts, and consequently they have become more common in engineering surveying applications.

1.2.1 *Electronic circle measuring systems*

Most automatic circle measuring systems used in modern electronic theodolite designs are based on the use of a photolithographically coded circle and a photoelectric detector element or photodiode. Two measurement systems are used: absolute and incremental. Absolute systems provide unique angular values, normally by means of a coded circle; incremental designs, in contrast, normally involve measurement relative to a zero or index mark.

The concept of electronic angle measurement can be best illustrated by initially considering a glass circle on which a series of opaque sections has been photographically etched. If a light source is used to illuminate the glass circle, and a photodiode, with a width equivalent to one section, is used as an index mark, it is possible to measure the number of light and dark units which pass below the diode as the alidade is rotated. Hence the angle which has been turned can be determined. Unfortunately the degree of accuracy which can be achieved with such a single-track system is very limited. For example, assuming that the photodiode had an area of 0.5 mm^2 and the circle diameter was 80 mm, then the maximum number of transparent and opaque sections which can be accommodated is approximately 500, providing an angular resolution of about 40′ of arc.

An alternative to employing a single track of transparent and opaque sections is to use several concentric tracks. If each transparent section then represents an 'on' signal and is designated as the number 1, and each opaque section represents an 'off' signal and is represented by 0, then it is possible to represent circle readings as individual binary codes. Again, however, practical considerations restrict the use of this technique for high resolution. Firstly, it is essential that all the detector elements are precisely aligned, since a small error in the alignment of the detector (particularly when reading the innermost tracks) could have serious consequences on the accuracy of the circle reading. Secondly, the interval between successive transparent and opaque sections becomes extremely small, and hence difficult to manufacture, at high resolutions. The magnitude of this problem is illustrated by Table 1.1, which lists the number of tracks and the distance between sections on the innermost track for various angular resolutions. Even to achieve a direct angular resolution of 20″ of arc the distance between elements on the innermost track must be less than 2μm. For higher resolutions, the problem becomes even more acute. Various designs have been proposed to overcome these problems (Gorham, 1976; Rawlinson, 1976). One of the most common techniques to be used for high-accuracy interpolation involves the use of moiré fringe patterns. Two typical instruments which use this approach are the Kern E2 and Wild T-2000 electronic theodolites.

1.2.1.1 *Kern E2 system.* The Kern E2 electronic theodolite (Fig. 1.1) was launched in 1981 at the Montreux Congress of FIG (Fédération International de Géomètres), and is a development of an earlier design, the Kern E1.

The horizontal circle measuring system is an incremental design based on resolving, dynamically, a coarse measurement to about 30″ of arc, and, statically, a fine

Table 1.1 Pure binary coded electronic circle: distance between elements at varying angular resolutions (assuming circle radius of 80 mm and a track width of 1 mm)

Resolution	1′	20″	1″	0.1″
No. of sectors	2160	64 800	1 296 000	12 960 000
No. of tracks	12	16	20	24
Distance between elements on the innermost track (m)	1×10^{-4}	2×10^{-6}	1×10^{-8}	1×10^{-8}

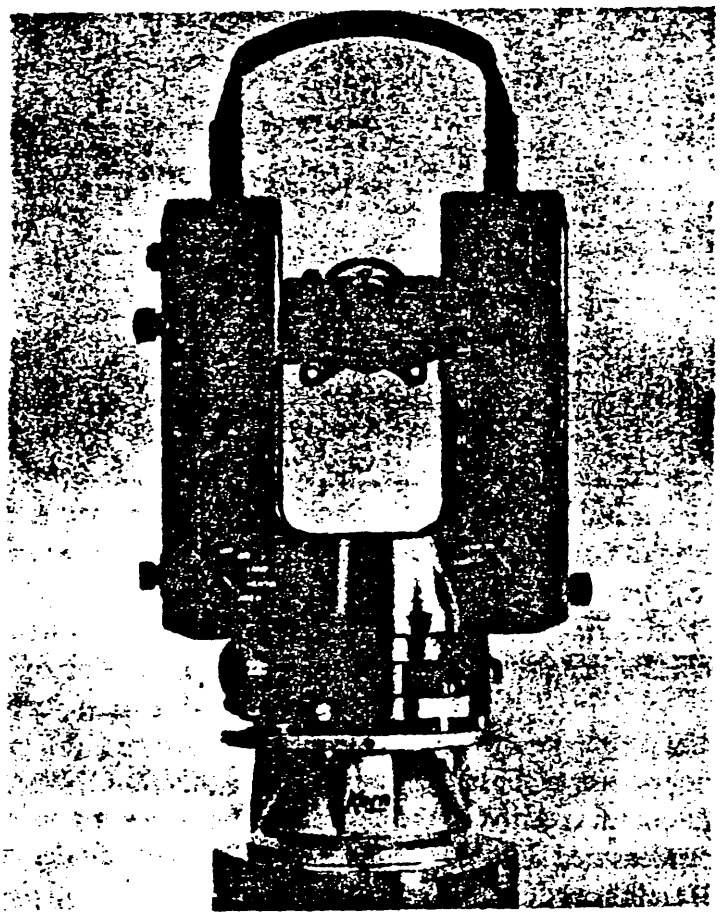


Figure 1.1 Kern E2 electronic theodolite. Courtesy Kern (Aarau) Ltd.

measurement to the nearest 0.3″. Both measurements are performed by using a moiré fringe pattern. Interference patterns of this type can be formed when closely ruled parallel lines are superimposed at a slight angle to each other. In the case of the Kern E2, the pattern is formed by focusing a slightly magnified section of one side of the circle on to an identical section from the region of the circle which is diametrically opposite. Since the circle consists of some 20 000 opaque and transparent lines approximately 6μm wide, a moiré pattern is formed (Fig. 1.2). Four photodiodes monitor the sinusoidal brightness distribution of the moiré pattern, and the outputs are used to obtain a coarse and a fine measurement.

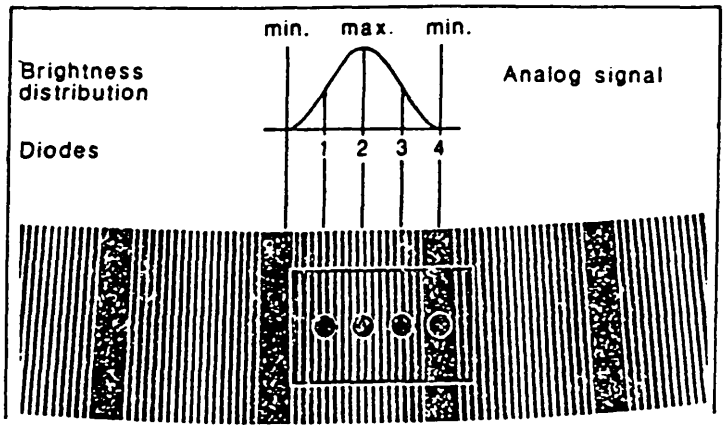


Figure 1.2 Kern E2 electronic angle measurement: formation and measurement of moiré fringe pattern.

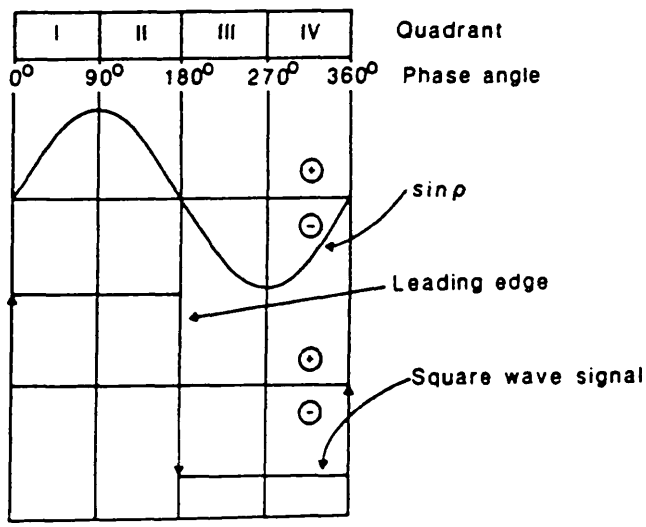


Figure 1.3 Conversion of sinusoidal output from photodiodes to a square wave and measurement of leading edge.

The coarse component of the circle reading is obtained by measuring the rotation of the array of photodiodes relative to the circle. This process is carried out by converting the dynamically measured sinusoidal output from the photodiodes into a square wave signal (Fig. 1.3) and then counting the number of leading edges (changes from negative to positive status) which occur during the transition from the initial to the final pointing. By this process it is possible to derive the angular change to approximately 30″ of arc.

The fine measurement is performed when the photodiodes are stationary in relation to the circle. The fine measurement aims to determine, by analysing the output from the four photodiodes, the location of the measuring point in relation to the leading edge of the last moiré period. The interpolation is carried out by measuring the phase angle of the moiré signal. This is achieved by processing the output from the photodiodes which are spaced at intervals of 0.25 of the moiré period (90°). Using the outputs from diodes 1 and 3, and from 2 and 4 and the process described by Munch (1984) it is possible to discriminate the position of the measuring mark to within 1% of the moiré period or to approximately 0.3″ of arc.

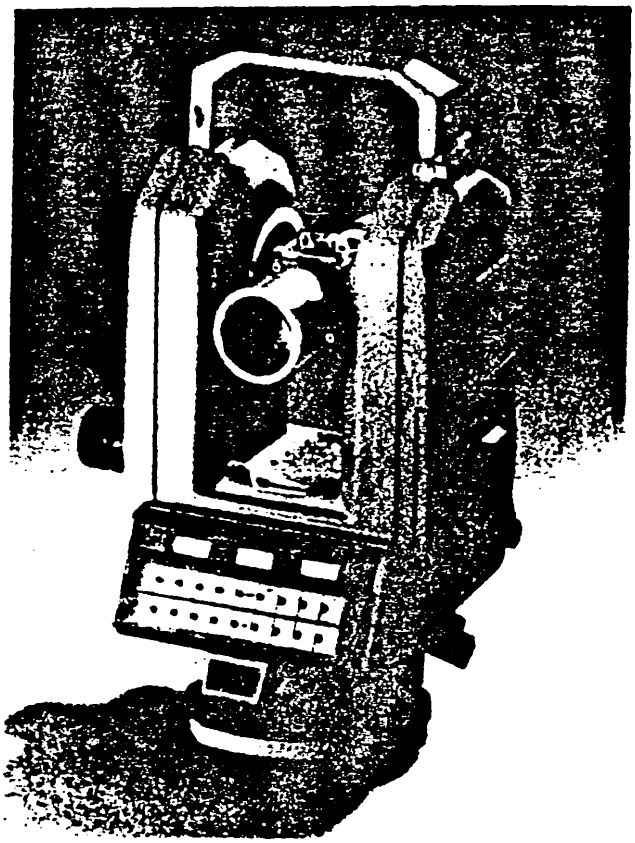


Figure 1.4 Wild T-2000 electronic theodolite. Courtesy Wild (Heerbrugg).

The vertical circle reading system is similar to that described above. However, in order to ensure that the vertical angles are measured with respect to a horizontal plane (and not relative to the arbitrary position of the telescope when the instrument is switched on) it is necessary to initially electronically orient the circle. This is achieved by an electronic zero pulse which is identified by the instrument when the telescope is rotated through the horizontal.

Accuracy testing, using the technique described by Maurer (1984) indicates that the extent of any circle graduation errors arising from the use of the electronic system described is less than 0.5″ of arc.

1.2.1.2 *Wild T-2000 system.* The Wild T-2000 electronic theodolite (Fig. 1.4) uses a dynamic, circle scanning technique to derive horizontal and vertical angles. In this case the circle is divided into 2096 equally spaced opaque and transparent segments (or 1048 phase units, ϕ_0) as illustrated by Fig. 1.5. A drive motor rotates the glass circle past two pairs of photodiodes which are positioned diametrically opposite each other (and thus enable circle eccentricity errors to be minimized). Figure 1.5 shows the case for one pair of photodiodes. One diode L_F is fixed, and is analogous to the zero mark of a conventional theodolite. The other, movable diode, L_M, records the direction of the telescope.

The determination of the angle ϕ between the two diodes is carried out by measuring the phase difference between the signals L_F and L_M. If, for example, the two signals are in phase, then the angle will consist of an unknown integer number (n) of phase units (ϕ_0). If, however, the signal from L_M is displaced relative to that from L_F, then the angle

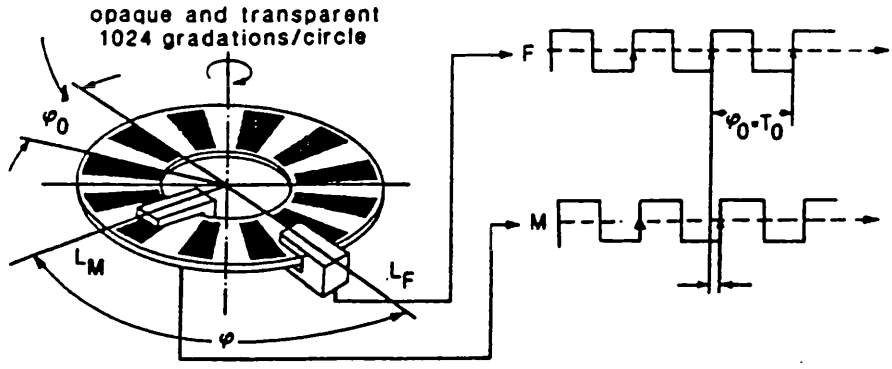


Figure 1.5 Wild T-2000 electronic circle measuring system.

will be defined as

$$\phi = n\phi_0 + \Delta\phi \tag{1.1}$$

where $\Delta\phi$ is the phase difference between the two signals. In order to resolve ϕ it is therefore necessary to derive a value for n, the coarse measurement, and a value for $\Delta\phi$, the fine measurement. Both measurements are carried out simultaneously after the circle has reached its specified rotation speed.

The coarse measurement is carried out by an electronic counter which sums the number of periods of the phase unit ϕ in the angle being measured. Since each phase unit is equivalent to a time interval (T_0) of about 300 μs, this measurement can be obtained directly and enables the angle to be resolved to the nearest 20 minutes of arc. The fine measurement is also carried out using an electronic counter in order to resolve $\Delta\phi$. Over 500 measurements (alternately from opposite sides of the circle) of $\Delta\phi$ are evaluated during a single rotation of the circle. The time taken for the entire process of obtaining a single measurement is 0.6 s, half of which is used to accelerate the motor to its design speed. A full description of the circle measuring system and of the laboratory tests to assess the accuracy of the instrument can be found in Katowski and Saltzmann (1983).

1.2.2 *Electronic tilt sensors*

A second unique feature of electronic theodolites compared with their optical counterparts is the inclusion of electronic tilt sensing devices. Components of this type can be used to automatically compensate horizontal and vertical angles for any residual inclinations of the vertical axis of the theodolite. Not only do these devices increase the accuracy of the instrument, but they also enable corrections to be applied directly to single face observations, so eliminating, for all but the most precise measurements, the need to record observations in the face left and face right positions of the telescope. Compensators may be referred to as single or dual axis devices. In the first case their function is normally only to correct vertical angles, whereas in the second case, corrections to horizontal angles may also be applied. The dual axis tilt sensor used by the Kern E2 is illustrated by Fig. 1.6(*a*) and (*b*).

The compensation system in this design is based on the reflection of a light spot from a liquid surface. The surface of the liquid will always remain horizontal irrespective of the inclination of the instrument, and can therefore be used as a reference surface. By redirecting the light source through various optical components and on to a photodiode, it is possible, by measuring the induced current, to determine the

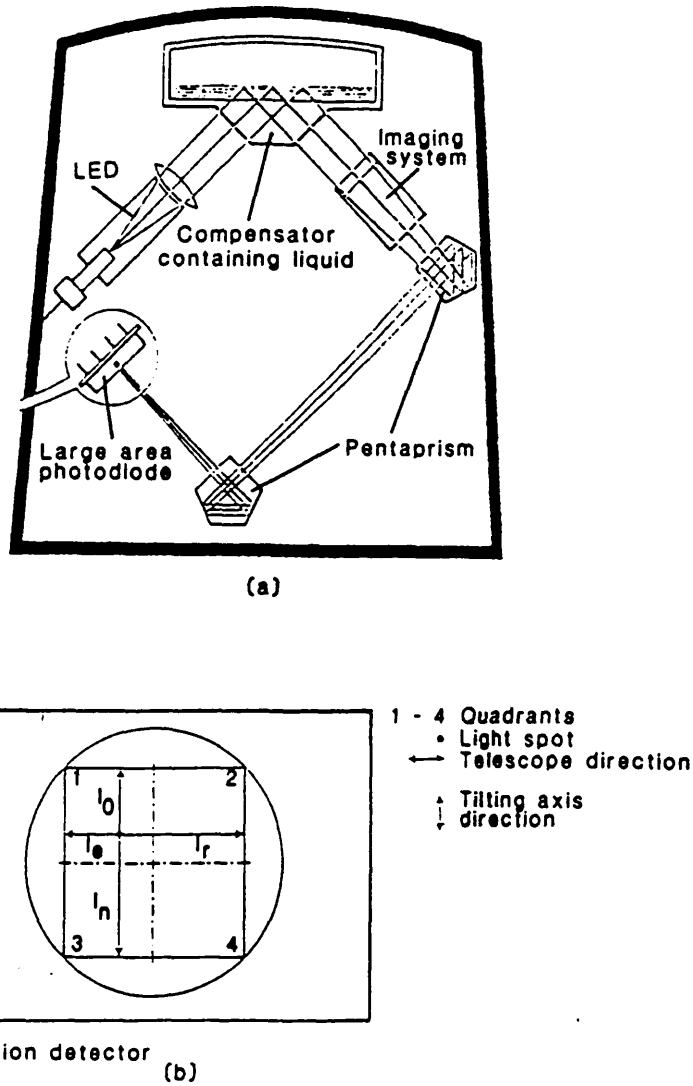


Figure 1.6 Kern E2: (*a*) dual axis electronic tilt sensor; (*b*) position detector (courtesy Kern Aarau).

position of the reflected light in relation to the zero point of the compensator. Values can then be derived for the inclination of the theodolite along the line of collimation and also along the trunnion axis of the instrument.

1.3 Principles of electronic distance measurement (EDM)

1.3.1 *Principles of distance measurement using phase measurements*

Although a wide variety of EDM instruments are currently available, most derive distances by measuring either the time delay between pulses of EM energy, normally referred to as the *pulse echo* technique, or by comparing transmitted and received signals and deriving the *phase difference* between them. The remainder of this section will deal specifically with instruments which utilize the latter technique. The former technique is normally restricted to laser-based instruments, and further details can be found in Chapter 2.

The fundamental stages involved in the determination of a distance electronically can be represented by a flow diagram (Fig. 1.7). The basic EM signal generated by the

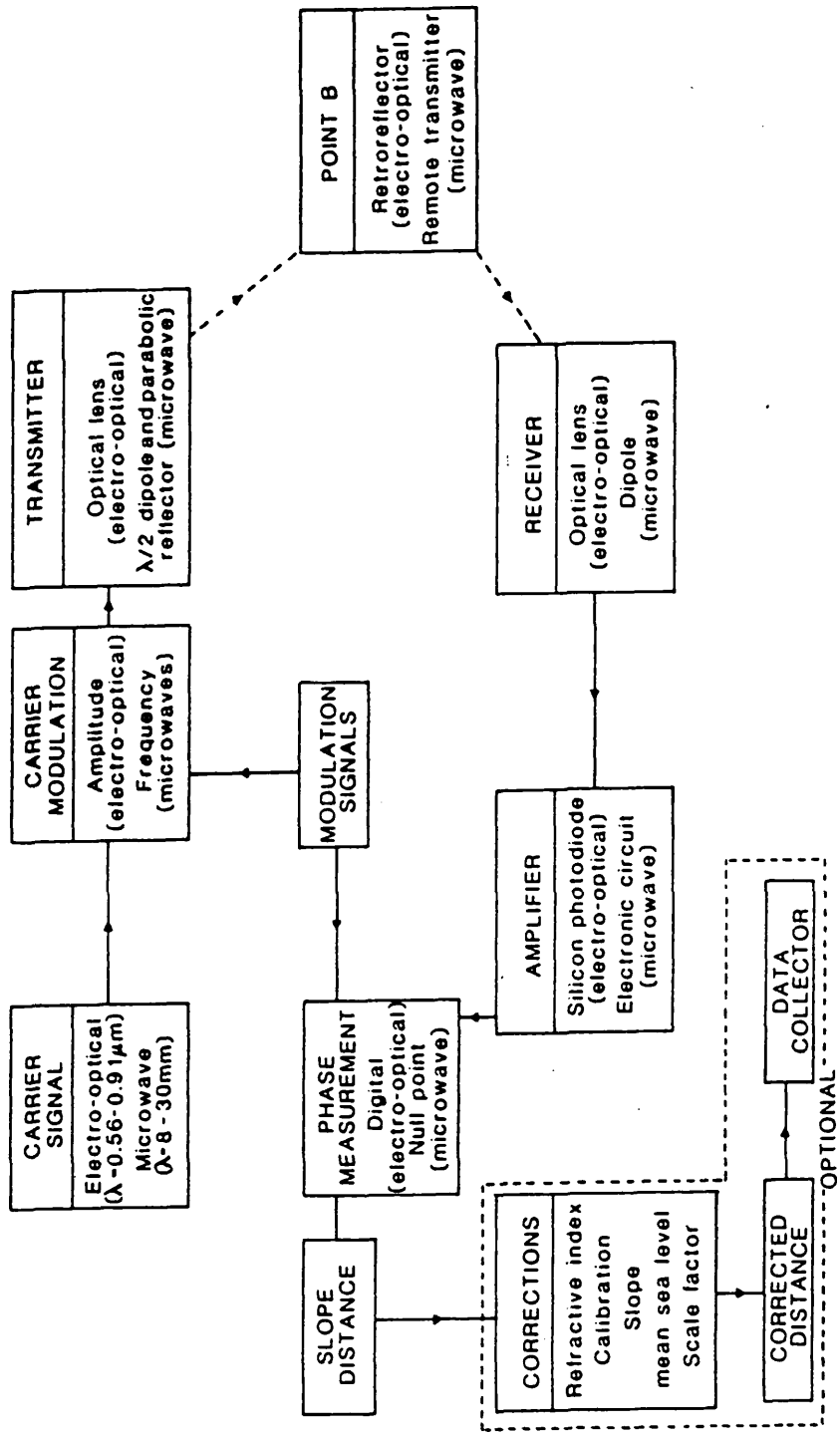


Figure 1.7 Flow diagram illustrating principle of operation of an EDM instrument.

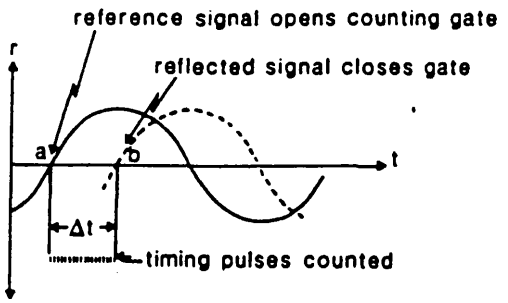


Figure 1.8 Phase measurement using digital pulse counting (Burnside, 1982).

instrument is termed the carrier signal, and may be of visible, infrared or microwave wavelengths. The first two cases may be termed *passive* EDM instruments, in which the EM signal is reflected from an unmanned reflector, whereas the latter forms the basis of *active* EDM instruments in which the return signal is generated by a second manned instrument at the terminal point.

Since the wavelength of the carrier signal is very short (0.5 μm–30 mm), in order to ensure that the phase difference can be measured precisely, a measurement or modulation signal is superimposed on to the carrier signal. This process is termed 'carrier modulation'. The transmitted signal is then either reflected or retransmitted along an identical path back to the instrument. With electro-optical devices, reflection of the signal is achieved by means of precisely manufactured retroreflective prisms, also referred to as corner cube reflectors. For short-range operations, acrylic retroflectors may also be used (Kennie, 1983, 1984). Reflectors of this type return the incident signal along an identical path back to the instrument. Generally speaking, alignment of the reflector is not a critical problem, although precise alignment of the measuring instrument is essential. In the case of microwave instruments, the signal is received by a second 'remote' instrument and retransmitted back to the master instrument.

The weak reflected signal is then amplified and compared in phase with the original transmitted beam. The difference in phase between the two signals is derived using either a digital pulse counting approach (electro-optical instruments) (Fig. 1.8) or a null point approach (microwave instruments) in which the phase difference between the two signals is brought to zero by introducing an additional phase shift (the required phase shift being indicative of the original phase difference).

The measurement of the phase difference between the two signals is representative of the fractional part of the total distance less than the integer value of the modulation wavelength. Figure 1.9 graphically represents the measurement of the distance AB. It can be seen that the distance AB is given by

$$D = M\lambda + \Delta\lambda \tag{1.2}$$

where M is an integer (2 in this case), and $\Delta\lambda$ is a fraction of a wavelength. Both M and $\Delta\lambda$ are unknowns. As mentioned previously, the signal is normally reflected back to the instrument, resulting in

$$2D = N\lambda + \Delta\lambda \tag{1.3}$$

where N is an integer (4 in this case).

This is equivalent to N revolutions of the vector OA (Fig. 1.9) plus the fractional excess equal to angle ϕ. Angle ϕ is a measure of the phase difference between the

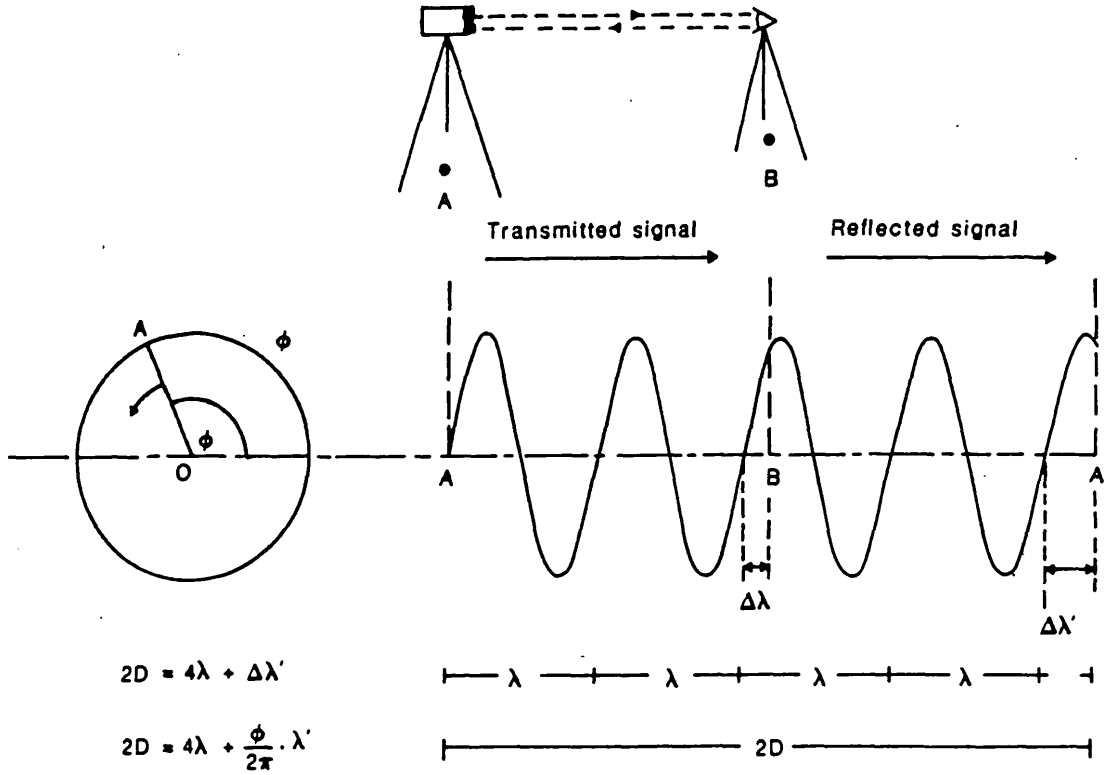


Figure 1.9 EDM distance determination: double path measurement.

outgoing and reflected signals and can be used to derive a value for

$$\Delta\lambda = \frac{\phi}{2\pi}\lambda \tag{1.4}$$

thus

$$D = N\lambda/2 + \frac{\phi}{2\pi}\frac{\lambda}{2}. \tag{1.5}$$

This is a rather idealized model, and equation (1.6) describes the situation more precisely.

$$D = N\frac{C}{2fn_a} + \left[\frac{\phi}{2\pi} \times \frac{C}{2fn_a}\right] + k_2 + k_3 \tag{1.6}$$

where C is the velocity of the EM signal in a vacuum
f is the modulation frequency (assumed error-free)
n_a is the refractive index of the atmosphere (see section 1.3.2.2)
k_2 is the zero error of the instrument (see section 1.3.2.3)
k_3 is the cyclic error of the instrument (see section 1.3.2.3).

The integer N in equations (1.3), (1.5) and (1.6) cannot be derived directly by a single measurement, and is normally obtained by introducing successive changes to λ, which lead to a series of equations of the general form

$$\overbrace{D}^{\text{unknown}} = N_1\overbrace{\frac{\lambda_1}{2}}^{\text{known}} + \overbrace{\left(\frac{\phi_1}{2\pi}\cdot\frac{\lambda_1}{2}\right)}^{\text{measured}} \tag{1.7}$$

$$D = N_2\frac{\lambda_2}{2} + \left(\frac{\phi_2}{2\pi}\cdot\frac{\lambda_2}{2}\right) \tag{1.8}$$

$$D = N_3\frac{\lambda_3}{2} + \left(\frac{\phi_3}{2\pi}\cdot\frac{\lambda_3}{2}\right) \tag{1.9}$$

In the case of microwave instruments, the solution to these equations is obtained by manually measuring $\Delta\lambda$ using five values of λ which increase progressively by a factor of 10. Dodson (1978) provides the following example to illustrate this principle:

λ(m)	$\lambda/2$(m)	$\Delta\lambda = \frac{\phi}{2\pi}\cdot\frac{\lambda}{2}$(m)
2	1	.1243
20	10	6.124
200	100	76.12
2000	1000	376.1
20000	10000	2376.0

Distance = 2376.1243 m

With modern electro-optical instruments it is not necessary to manually alter the value of λ. It is changed automatically in a continuous manner, and phase difference values are measured several hundred times on both an internal calibration path and over the external light path.

For very high-precision distance measurement, the principle of measurement differs in some respects from that described so far. The Kern Mekometer ME 3000 is a typical example of this category of instrument.

The Mekometer was developed in the early 1960s at the United Kingdom National Physical Laboratory (NPL), was subsequently manufactured commercially by Kern of Switzerland and became available as a production model in the early 1970s. Two models are currently in use, the ME3000 and the ME5000, the latter being a recent development which uses a He-Ne laser.

The Kern Mekometer ME 3000 is one of the few instruments which uses a xenon flash tube as the source of the carrier signal. Also, it uses polarization modulation, with an extremely high modulation frequency and short modulation wavelength (30 cm). Furthermore, to reduce the effect of atmospheric variations on the measurements, the modulation wavelength is fixed by the resonance of a small microwave resonator containing air under pressure. It is a well-established instrument and has been used throughout the world for deformation monitoring projects. For example, Murname (1982) and Keller (1978) have used the instrument for monitoring reservoirs and arch dams respectively. A full account of the principle of measurement using EDM instruments can be found in Burnside (1982) and Rüeger (1988).

1.3.2 *Errors influencing EDM measurements*

Many different sources of systematic error can affect EDM measurements, including those caused by the instrument operator, the atmosphere and instrument maladjustment.

1.3.2.1 *Instrument operator errors.* Clearly it is important that, for reliable, precise results, due care and attention should be paid to the operation of an EDM instrument.

In particular, care should be taken with the following operations:

(i) Precise centring–regular checking of the optical plummet of a theodolite (if the EDM is theodolite-mounted), or of an integral optical plummet, is important; inaccurate centring is often one of the most common reasons which account for distance variations occurring between points measured using different EDM instruments

(ii) Pointing–careful pointing to the correct position on a reflector target is particularly important when using an EDM which has the facility for automatic slope reduction, since any pointing error may not become evident, particularly if the results of the slope reduction are not independently checked using, for example, the observed vertical angle

(iii) Setting automatic correction values—for instruments which have the facility to store atmospheric and other correction factors, it is essential to ensure that the correct values for the prevailing conditions are being used.

1.3.2.2 *Atmospheric errors.* The effect of the prevailing atmospheric conditions (temperature, pressure and relative humidity) on EDM measurements is analogous to the influence of temperature when measuring distances with a steel tape. Variations in temperature cause an expansion or contraction of the tape from its standardized length, which, if not corrected, will lead to a systematic error in the measured distances. Similarly, variations in atmospheric conditions from those which are assumed by the instrument will cause an increase or reduction in the measuring wavelength and lead to a systematic error if a correction is not applied.

Normally the effect of the atmosphere is defined by the change in the refractive index (n), where n is defined as

$$n = \frac{c}{v} \tag{1.10}$$

where c = velocity of electromagnetic waves in a vacuum (299 792.5 km/s)
v = the velocity in the medium (atmosphere).

The value of n is normally close to unity and is assumed by many EDM instruments to be 1.000320. In the visible and infrared parts of the EM spectrum, n is calculated using the Barrell and Sears formula:

$$(n_t - 1) = (n_s - 1)\frac{273}{273 + t}\cdot\frac{P}{760} - \frac{15.02E}{273 + t} \times 10^{-6} \tag{1.11}$$

where

$$(n_S - 1) \times 10^{-6} = 287.604 + \frac{1.6288}{\lambda_0^{\,2}} + \frac{0.0136}{\lambda_0^{\,4}} \tag{1.12}$$

and n_t = prevailing refractive index
n_S = refractive index of air at 0°C and 760 mm Hg pressure with a 0.03% CO_2 content
t = temperature during observation (°C)
P = pressure during observation (mm Hg)
E = water vapour pressure (mm Hg)
λ_0 = wavelength of signal in a vacuum.

Table 1.2 Precision required in measurements of t, P and E to obtain 1 ppm precision in n (after Burnside, 1982)

	Precision in measurement		
Carrier signal	t(°C)	P(mm Hg)	Humidity (mm Hg)
Microwave	0.8	±2.9	±0.17
Visible/infrared	1.0	±2.7	±20

Equation (1.11) is normally used to produce a nomogram or slide rule by the manufacturers which illustrates the variation in n (in parts per million of the distance, ppm) for various values of t and P. The variation can then be entered into the instrument as a scale correction.

Although equation (1.11) physically models the variations in n to an accuracy of 1×10^6, it is nevertheless based on observed values of t, P and E (normally at the terminal points of the line being measured). It is therefore essential that care is taken to accurately measure these quantities. Burnside (1982) has shown that to obtain a precision of 1ppm in the determination of n, the three quantities must be measured to the degrees of precision shown in Table 1.2.

Table 1.2 clearly indicates (i) that temperature and pressure need to be determined to about 1°C and 3 mm Hg, and (ii) that microwave instruments are very susceptible to changes in relative humidity and that this quantity should be measured to a much higher precision than for electro-optical instruments.

1.3.2.3 *Instrument errors.* EDM devices, like other surveying instruments, require careful use and regular calibration if they are to provide reliable and accurate results. Although it is rare for a modern EDM instrument to display results which are grossly incorrect (most instruments normally do not display any results if severely out of adjustment), most instruments are subject to relatively small but vitally important systematic errors caused by instrumental maladjustment.

Scale errors (k_1) occur if the modulation frequency of the EDM instrument does not correspond exactly with the design frequency value for the instrument. The error is proportional to the distance measured and may often be expressed in parts per million (ppm) of the distance. Ideally, it should be negligible (< 1 ppm); however it may be as high as 20 to 30 ppm (20–30 mm/km) in extreme cases.

Zero error (k_2) (also referred to as additive constant error, index error or reflector/prism offset) represents the difference between the EDM measured distance between two points (corrected for scale, cyclic and atmospheric errors) and the known distance between the two points. It is caused if the internal measurement centre of the instrument (and reflector) does not coincide with the physical centre of the instrument/reflector which is plumbed over the measuring mark. It also accounts for the variation in the path length through the retroflector caused by refractive index variations at the glass–air interface. It is a systematic error of constant magnitude and therefore *not* proportional to the distance measured.

The value of zero error obtained during calibration refers to a particular EDM instrument and retroflector. If either the instrument or reflector are changed, then a new zero constant will exist.

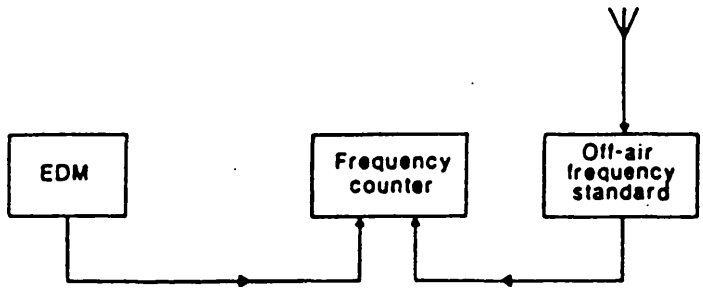


Figure 1.10 Laboratory determination of scale error using standard frequency.

Cyclic errors (k_3) (also termed periodic errors, resolver errors or non-linearity errors) are analogous to circle graduation errors in theodolites in the sense that they are periodic and can have a systematic effect, particularly at short distances. They are generally the smallest of the three sources of instrument error and are caused by internal electronic contamination between the transmitter and receiver circuitry, the effect being cyclic over the modulation wavelength. The effect of this source of error is reduced by the instrument manufacturers by electrical isolation and shielding of components within the instrument, although the effect becomes more apparent with a weak return signal (Bevin, 1975).

1.3.3 *Methods of instrument calibration*

In general terms the distinction can be made between the need for quick and simple calibration tests which could, for example, be carried out on site, and more sophisticated calibrations (normally involving the use of a laboratory or permanently sited baseline) for more accurate determinations of the state of health of an EDM instrument.

Three distinct approaches to instrument calibration can be identified: laboratory methods, field methods using known baseline lengths, and field methods using unknown baseline lengths.

1.3.3.1 *Laboratory methods.* Laboratory tests can be carried out to assess both the scale and cyclic error components. Their main advantages over field tests are, firstly, that they are quick and convenient to apply in practice, and secondly, they enable the surveyor to precisely control the measurement conditions. For practical reasons, however, they are limited to short-range calibration measurements.

Scale error is determined in the laboratory by comparing the observed modulation frequency of the EDM signal with a true reference frequency (Fig. 1.10). The reference frequency may be obtained either from off-air radio transmissions, or from a crystal generated laboratory standard.

Cyclic errors can be determined in the laboratory by recording distance measurements to a series of reflector positions which extend over a distance equivalent to the halfwavelength of the instrument. The simplest and cheapest arrangement involves the use of a graduated metal bar (Hodges, 1982).

The calibration bar is constructed so that the prism can be moved in increments of 10 mm (with an accuracy better than 0.02 mm). The bar is placed approximately 40 to 50 m from the EDM instrument to minimize pointing errors. If the effects of scale error are presumed to be negligible over such short distances, then each measurement to the

bar (L^M) will be subject only to zero and cyclic errors, so that

$$L^M = L^T + k_2 + k_3 \tag{1.13}$$

where L^M is measured distance
L^T is true distance (unknown)
k_2 is zero error
k_3 is cyclic error.

For any other point i,

$$L_i^M = L_i^T + k_2 + k_{3i}, \tag{1.14}$$

Subtracting equation (1.13) from (1.14) to eliminate constant zero error:

$$\Delta L^M = \Delta L^T + (k_3 - k_{3i}) \tag{1.15}$$

or

$$(\Delta L^M - \Delta L^T) = (k_3 - k_{3i}) \tag{1.16}$$

A graph may then be drawn illustrating the differences between the distances (ΔL) at various values of L. The resulting graph will then illustrate the amplitude and phase of any cyclic error.

In situations where space is limited, the arrangement described by Deeth *et al.* (1979) may be adopted. A measurement range of 50 m was obtained by bending the path of the EDM signal using surface-silvered mirrors. The calibrated bar in this case was an 11 m track on which a rectroreflector was mounted. A laser interferometer (Hewlett Packard 5526A) provided a precise output of the movement of the reflector which could then be compared with that obtained using the EDM instrument.

1.3.3.2 *Field methods using unknown baseline lengths.* Measurements on a calibration baseline can be used to determine all three error components (scale, zero and cyclic errors). Both unknown and known baseline lengths may be used, although the former tends to be the more popular technique (Kennie *et al.*, 1988). The main *advantage* of the unknown-baseline-length approach is that, although stability of the markers or pillars indicating the baseline is important, the points are required to be stable only for the duration of the survey observations. If a known baseline length is used, long-term pillar stability and periodic monitoring of the interpillar distance are particularly important. The main *disadvantage* of the unknown baseline length method is that it is not possible to derive a value for scale error by field observation. Two techniques which are based on the use of an unknown-length baseline are the three-point and Schwendener methods.

The simplest procedure for a quick and approximate determination of the zero error involves the use of a three-point baseline (Dodson, 1978), illustrated in Fig. 1.11. If the observed distances (l) are each in error by the same zero error (k_2), and if the actual distances from 1 to 2 and 2 to 3 are d_1 and d_2 then, ignoring all other errors,

$$l_{12} = d_1 + k_2 \tag{1.17}$$

$$l_{23} = d_2 + k_2 \tag{1.18}$$

$$l_{13} = d_1 + d_2 + k_2 \tag{1.19}$$

therefore

$$k_2 = l_{12} + l_{23} - l_{13} \tag{1.20}$$

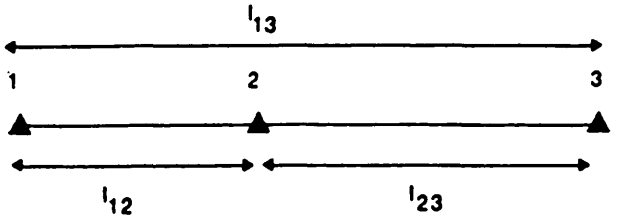


Figure 1.11 Zero error determination using three-point baseline method (Dodson, 1978).

If the baseline consists of more than two sections and each section l is measured, then if the total length is (L), a more precise value of k_2 can be derived using

$$k_2 = \frac{L - \sum l_i}{n - 1} \tag{1.21}$$

where L is the total length of baseline
l_i is the length of each baseline section
n is the number of baseline sections.

The standard error of the zero constant is given by

$$\sigma k_2 = \frac{(n+1)\sigma_d}{(n-1)^2} \tag{1.22}$$

where σ_d = standard error of the measured distances.

A further improvement in accuracy can be achieved by measuring not only the individual section distances, but all possible combinations of distances on a baseline, and using least squares methods to derive the most probable values for the zero error and (in some cases) cyclic error. One of the most common methods used is that devised by Schwendener (1972).

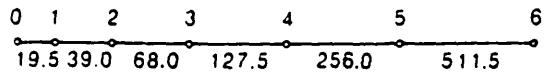
The Schwendener method improves the determination of k_2 by making use of a seven-point baseline (six sections), which, taking all combinations of distances, results in a total of 21 measurements. Since seven are required for a unique solution, the solution is overdetermined, with 14 redundant measurements.

The particular baseline design and interpillar distances suggested by Schwendener are illustrated in Fig. 1.12, and a typical pillar design and centring system in Fig. 1.13. The choice of interpillar distances is of vital importance and should, firstly, ensure that the distances to be measured are spread over the measuring range of the instrument, and secondly, that the fractional elements of the interpillar distances are evenly distributed over the half wavelength of the modulation signal (normally 10 m). Schwendener's basic observation equation is of the form

$$\left(\sum l_j\right)_i + k_2 = M_i + v_i \tag{1.23}$$

where l_j are the unknown interpillar distances ($l_{01} \ldots l_{06}$)
k_2 is the zero error
M_i are the observed values (21 distances)
v_i are the residuals (21 values).

A tabular form of solution to determine the most probable values of l_{01}, l_{02}, l_{03}, l_{04}, l_{05}, l_{06}, and k_1 is presented in Schwendener (1972), together with equations to



Distances measured

01	12	23	34	45	56	
02	13	24	35	46		Total 21
03	14	25	36			
04	15	26				
05	16					
06						

Figure 1.12 Schwendener baseline design and inter-pillar distances (Schwendener, 1972).

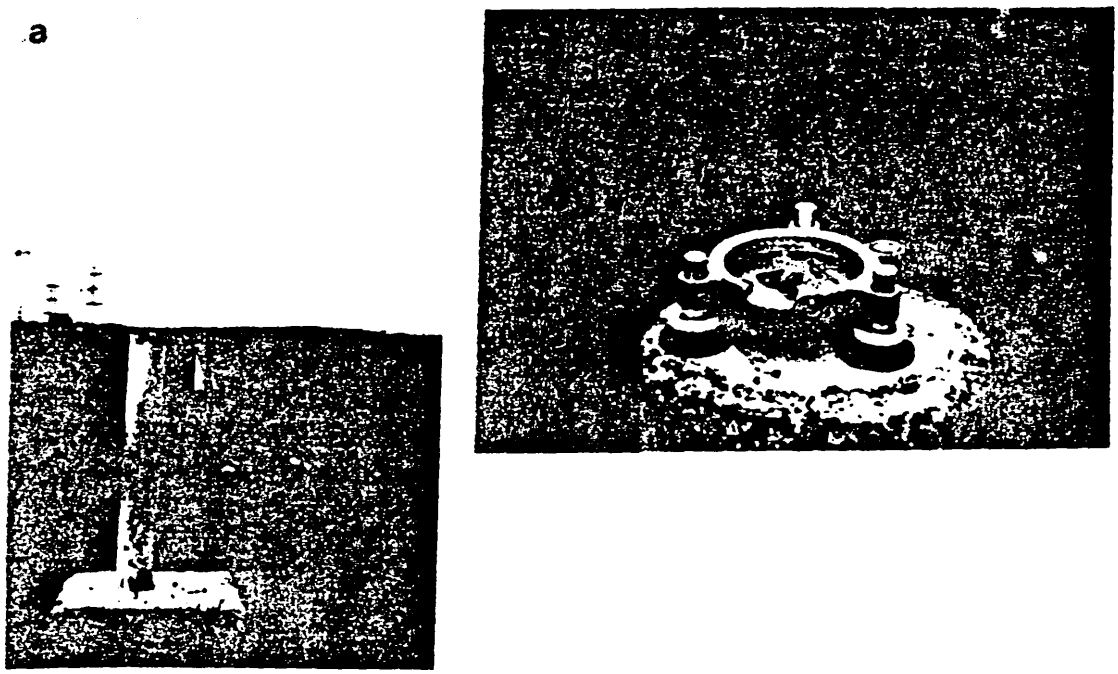


Figure 1.13 (*a*) Typical pillar design for multi-point baseline; (*b*) Wild centring system. Courtesy Paisley College.

determine the standard error of a single measured distance and the zero error. An indication of any cyclic errors can be obtained by plotting the residuals (most probable values of the distances – observed values) as a function of the unit metres and fractions of a metre of the individual measured distances (Fig. 1.14). Although the method provides useful results, it should be noted that the cyclic error curve illustrates not only the effect of cyclic errors, but also the influence of random observational errors. Consequently, if the cyclic error is not significantly greater than the random variations, it may not be determined by this procedure.

1.3.3.3 *Field methods using known baseline lengths.* The use of known values for the length of the baseline and baseline sections is advantageous since it enables the value of the scale error to be derived in addition to zero and cyclic errors. A variation on the Schwendener method which includes measured distances and overcomes these limitations has been developed at the University of Nottingham (Ashkenazi and Dodson, 1975; Deeth *et al.*, 1979).

The Nottingham method uses a baseline consisting of seven pillars which span a total

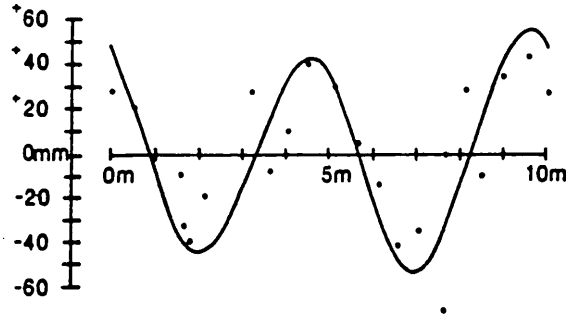
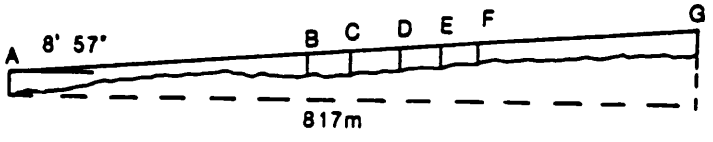


Figure 1.14 Graphical representation of cyclic error determination using Schwendener baseline (Schwendener, 1972).



Profile of baseline

	Approximate interpillar distance (m)					
Pillar	A	B	C	D	E	F
B	380.5					
C	422.2	41.7				
D	473.0	92.5	50.8			
E	526.1	145.6	103.9	53.1		
F	577.5	197.0	155.3	104.5	51.4	
G	817.8	437.3	395.6	344.8	291.7	240.3

Figure 1.15 Nottingham University multipoint baseline (Ashkenazi and Dodson, 1975).

distance of 817 m (Fig. 1.15). By employing a Hewlett Packard 5526A laser interferometer, the four central bays (197 m) were measured to an accuracy of ± 0.1 mm. It has been shown (Ashkenazi and Dodson, 1975) that the basic Schwendener model can be varied to include the effects of scale and cyclic error to form the observation equation

$$\left(\sum l_j\right)_i + k_2 + k_1\left(\sum l_j\right)_i = M_i - (k_3)_i + v_i \qquad (1.24)$$

where l_j are the known or observed distances

k_2 is the zero error

k_1 is the scale error

k_3 is the known cyclic error (from laboratory measurements)

M_i are the observed distances

v_i are the residuals.

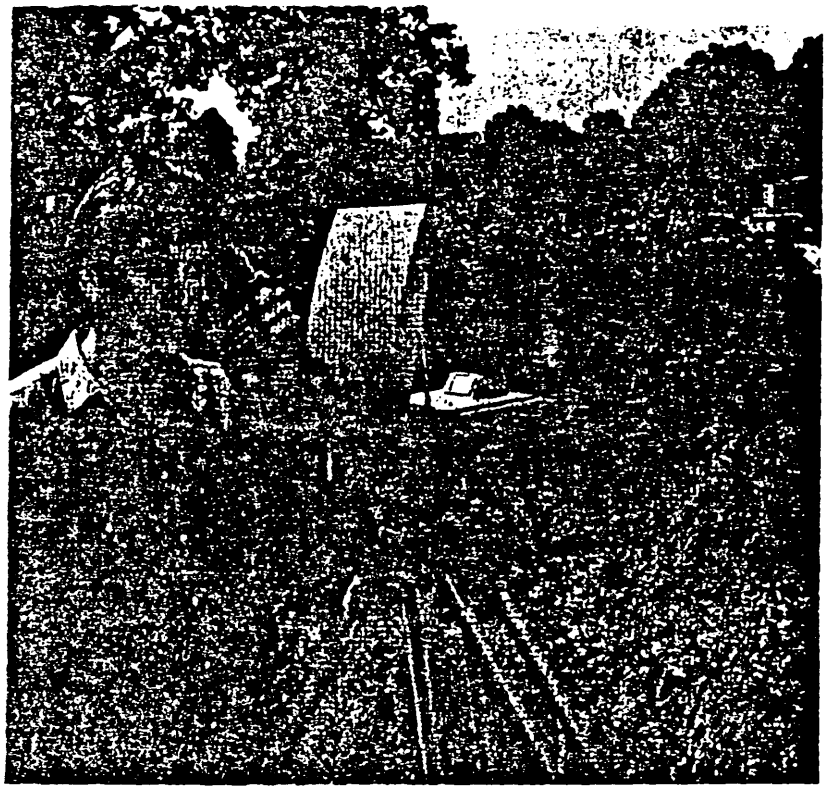


Figure 1.16 Tellumat CMW 20. Courtesy Tellumat.

The inclusion of known distances into a Schwendener-type baseline has also been adopted by the Ordnance Survey (Toft, 1975) and the USA National Geodetic Survey (Spofford, 1982).

An alternative approach to the deduction of scale, zero and cyclic errors using known baseline lengths and least squares regression techniques has been developed by Sprent and Zwart (1978) and Sprent (1980) and is used widely in Australia.

1.3.4 *EDM instruments and applications*

Since their introduction, EDM instruments have had an enormous impact over the whole range of survey measurement processes. The following section highlights some of the features of the EDM instruments which may be used for these activities.

1.3.4.1 *Geodetic control networks.* Electronic distance measurements over long ranges (> 10 km) have been used since the original development of EDM instruments to measure baselines and to strengthen existing geodetic control measurements. The direct measurement of distances in excess of 15 to 20 km is almost always performed nowadays using active microwave measuring instruments such as the Tellumat CMW 20 (Fig. 1.16). Distance measurement using microwaves requires a second instrument at the end of the line to be measured, rather than a retro-reflector, to retransmit the microwave signal back to the instrument position. A further distinguishing feature of microwave instruments is the provision of a two-way speech link between the 'master'

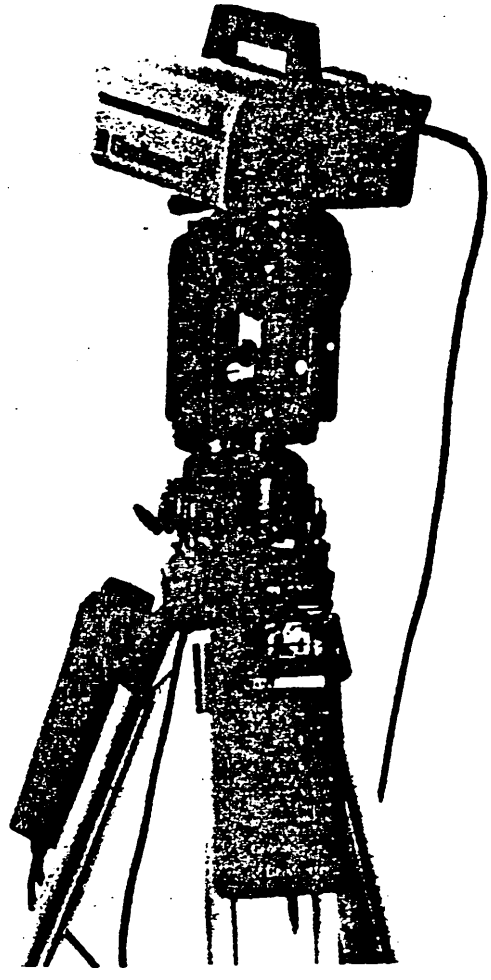


Figure 1.17 Geodimeter 122 EDM instrument. Courtesy Geotronics.

and 'remote' instruments. 'Master' in this context refers to the instrument which is controlling the distance-measuring process.

Instruments in this category are also attractive because of their ability to measure in poor visibility conditions, for example fog, conditions in which electro-optical instruments would find it difficult to operate.

1.3.4.2 *Large-scale surveys.* Large-scale (< 1:1000 scale) route and site surveys form the central activity of most survey organizations. Although traditional techniques such as theodolite and tape traversing and stadia tacheometry are still used, they have been largely superseded in most large organizations by EDM traversing using instruments such as the Geodimeter 122 (Fig. 1.17) and polar detailing using electronic tacheometers and data recorders (section 1.4). The use of such EDM instruments not only decreases the observation time but also enables higher orders of precision to be achieved in comparison to more conventional techniques.

1.3.4.3 *Setting out.* The widespread adoption of EDM instruments in the construction industry has led to significant changes in the techniques used for setting out. Of particular importance has been the move towards polar methods of setting out using coordinates (either on a local or national grid) rather than the more traditional

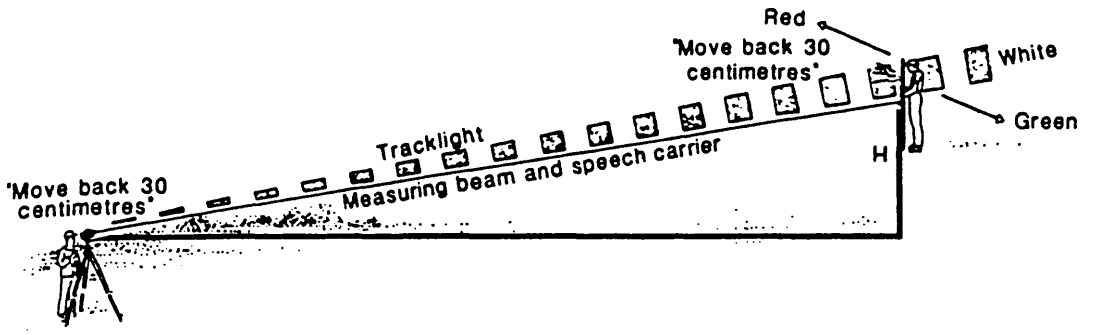


Figure 1.18 Instrument/prism reflector communication using UNICOM and Tracklight. Courtesy Geotronics.

techniques based on site grids and offsets. In order to improve the efficiency of such techniques in setting out, several additional facilities have been incorporated into many modern instruments.

Tracking: almost all modern electro-optical EDM instruments are capable of measuring to a moving target. Generally the level of accuracy is of a lower order of accuracy than for static measurements, typically 10–20 mm. Variations in slope or horizontal distance can thus be displayed and updated at intervals of a fraction of a second.

Integral speech communication: one-way communication between the instrument and target can be achieved by modulating the infrared carrier signal with speech information. The Geotronics UNICOM system uses a microphone built into the EDM instrument and a microphone mounted above the target prism to receive the transmitted instructions (Fig. 1.18).

Visible guidance system: a further facility offered by Geotronics is the tracklight system. In this case a flashing light source is built into the instrument. The correct alignment for the target occurs when the prism is located exactly along the direction being defined by the tracklight. When the target is in this position, a white pulsing light will be viewed by the survey assistant (Fig. 1.18). Any deviation off line will be indicated by a change in the colour of the pulsing signal, which turns red with movement to the right and green if the deviation is to the left of the correct position.

Remote receiver: an alternative approach to communication is adopted by the Kern (Aarau) company. In this case, rather than providing verbal instructions or visible clues to alignment, measured information is transmitted by the EDM instrument to a small receiver adjacent to the prism. A liquid crystal display (LCD) indicates up to five measured parameters, such as horizontal distance or northing and easting. An acoustic signal is also generated when the receiver is within the transmitted beam.

1.3.4.4 *Deformation monitoring.* The analysis of survey observations from different epochs to determine whether any structural deformation has occurred is the subject of Chapter 5. Several different measurement techniques can be used for such operations, including those based primarily on angle measurement, distance measurement or close-range photogrammetry. The determination of precise distances is of fundamental importance in all cases, and is generally performed using specialist instruments such as the Kern Mekometer (described previously) or the Comrad Geomensor. For very short-range distance measurements, laser interferometers may also be used (section 2.2).

1.4 Electronic tacheometers

The term electronic tacheometer refers to an instrument which is able to electronically measure both angles and distances and to perform some limited computational tasks using an integral microprocessor, such as reducing a slope distance to the horizontal or calculating coordinates from a bearing and distance. The instrument may also, but not necessarily, be able to electronically store data, either in an internal memory unit, or more commonly, on an external solid-state data recorder.

A further term which is in widespread use and is used to describe instruments in this category is 'total station'. While this latter term normally implies that the instrument is capable of electronic angle and distance measurement, it has also been applied to instruments which measure angles optically.

While there exist individual design features which are unique to a particular manufacturer, it is generally possible to differentiate between the integral and modular designs of electronic tacheometer. In view of the rate of development of microprocessor technology and the ever-increasing range of instruments on the market, it would, however, be inappropriate to attempt to describe the whole range of instruments in this field. Consequently, two examples which are representative of instruments in each group will be described in order to enable the general characteristics of each design to be more fully appreciated.

1.4.1 *Integrated design*

Many of the earlier examples of electronic tacheometers, such as the AGA Geodimeter 700 and the Zeiss (Oberkochen) Reg Elta, were of an integrated design. The primary feature of such a design is that the electronic theodolite and EDM instrument form a single integrated unit. One of the main advantages of such an approach is the need to transport only a single unit, and also to be able to dispense with auxiliary cables which are often required to link separate units. Furthermore, since the units are matched together during manufacture, and often use coaxial optics for telescope pointing and transmitting the EDM signal, instruments of this type rarely suffer from a lack of collimation of the telescope line of sight and the EDM signal. Two instruments which are typical of this design are the Sokkisha SET2/3 and the Geodimeter 440. The general characteristics of these instruments are summarized in Table 1.3.

The Sokkisha SET 2 (Fig. 1.19) and SET 3 are typical of the relatively low-cost electronic tacheometers which are being manufactured by Japanese electronics companies. Other recent examples include the Topcon ET-1 and the Nikon DTM-5. The system can also be linked to a data recorder such as the Sokkisha SDR2.

The Geodimeter 440 is a further example of an integrated instrument. The instrument has several unique features including an electronic level display, and the one way speech communication and visible guidance system discussed previously (see Fig. 1.18). It may also be linked to a data recorder.

1.4.2 *Modular design*

In this case the electronic theodolite and EDM instrument are separate units which can be operated independently. This arrangement is more flexible, since theodolites and EDM units with differing accuracy specifications can be combined. It may also be a more cost-effective solution, since the individual units may be replaced and upgraded

Table 1.3 Integrated electronic tacheometers

	Sokkisha SET 2	Sokkisha SET 3	Geodimeter 440
Angle measurement			
Standard deviation of mean pointing (face I and II)			
Horizontal	± 2″	± 5″	± 1″
Vertical	± 2″	± 5″	± 1″
Distance measurement			
Range (m) average conditions:			
1 prism	1300	1000	2300
3 prisms	2000	1600	3500
Maximum	2600	2200	5000
Standard deviation	± 3 mm ± 2 ppm	± 5 mm ± 3 ppm	± 5 mm ± 5 ppm

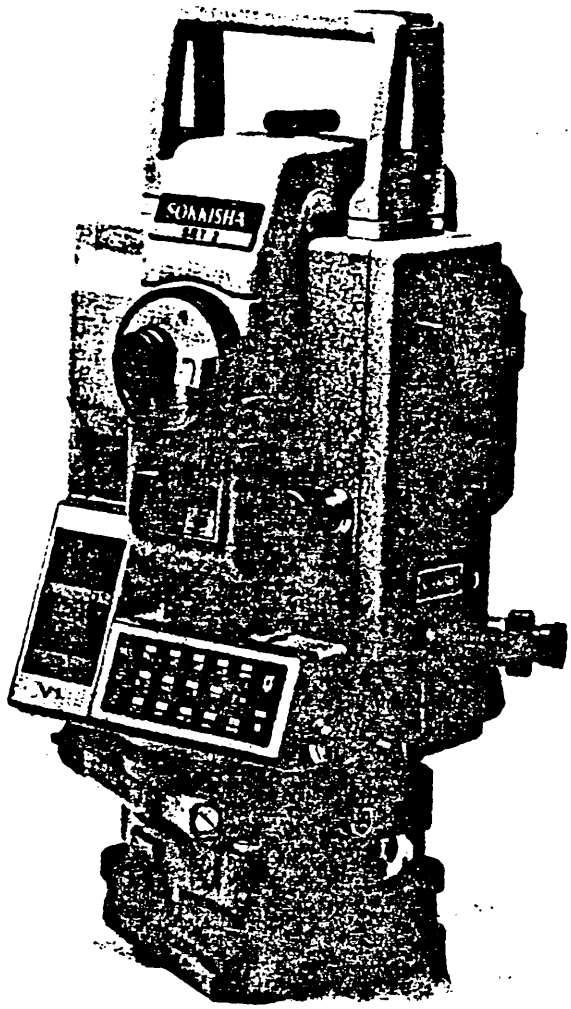


Figure 1.19 Sokkisha SET 2 integrated electronic tacheometer. Courtesy Sokkisha Instruments.

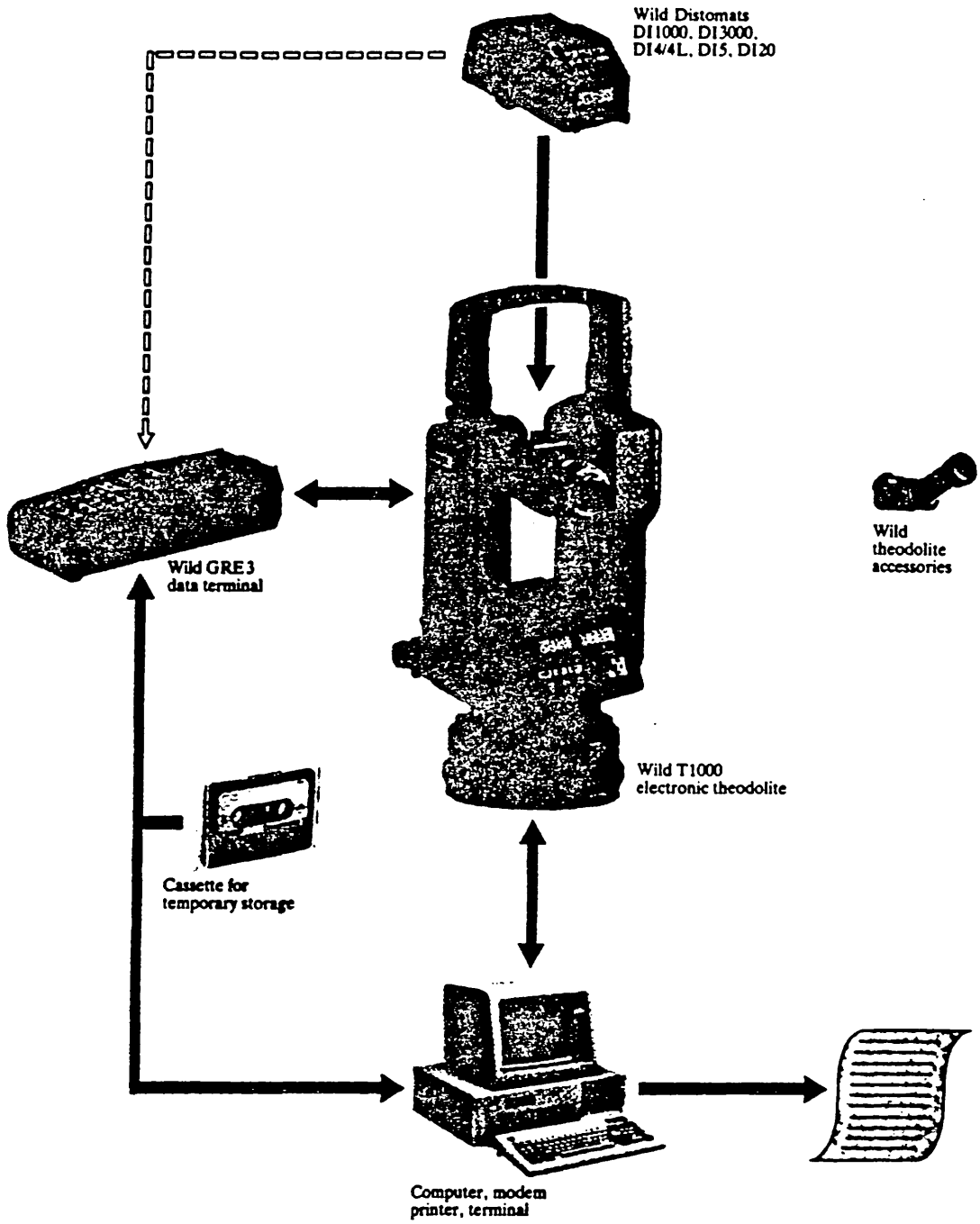


Figure 1.20 Elements of the Wild T-1000 modular electronic tacheometer. Courtesy Wild (Heerbrugg).

as developments occur. The Wild (Heerbrugg) and Kern (Aarau) designs are typical of instruments in this category.

Wild (Heerbrugg) offer two modular systems based around the Wild T-2000 (section 2.1.2) and T-1000 electronic theodolites (Fig. 1.20). Several combinations of theodolite and EDM are available, but the most popular arrangements are the Wild T-1000/DI-1000 and the Wild T-2000/DI-5. Both would normally be linked to a data recorder such as the Wild GRE3. The specifications of both systems are given in Table 1.4.

Kern (Aarau) also offer a range of components which can be combined to form a

Table 1.4 Modular electronic tacheometers

	Wild (Heerbrugg)		Kern (Aarau)	
	T-1000/ DI-1000	T-2000/ DI-5	E1/ DM 550	E2/ DM 503
Angle measurement				
Standard deviation of mean pointing (face I and II)				
Horizontal	±3″	±0.5″	±2″	±0.5″
Vertical	±3″	±0.5″	±2″	±0.5″
Distance measurement				
Range (m) average conditions:				
1 prism	500	2500	1800	2500
3 prisms	800	3500	3000	3500
Maximum	800	5000	3300	5000
Standard deviation	±5 mm ± 5 ppm	±3 mm ± 2 ppm	±5 mm ± 5 ppm	±3 mm ± 2 ppm

modular electronic tacheometer. Two of the most common arrangements are given in Table 1.4.

1.4.3 *Data recorders*

Traditionally the surveyor has recorded survey observations manually in a fieldbook. Standard fieldbook layouts have developed over the years, and these attempt to identify gross booking or reading errors. Some form of numbering and a field sketch are also normally required, particularly for recording survey detail. Conventional techniques such as these are quite adequate for many survey projects. However, with developments in digital mapping and terrain modelling in recent years, traditional methods of recording survey observations and subsequently transferring them via a computer keyboard have been found, in practice, to be slow and error-prone. Consequently the need arose for a device which could electronically store survey field observations. Such a device is referred to as a *data recorder*, or alternatively a data collector or data logger.

The data recorder and its associated peripherals for data transfer (Fig. 1.21) form a crucial link between the field observations and the final end product. In general, the distinction can be made between recorders dedicated solely to the logging and processing of survey observations, and the more general purpose hand-held computers which have been adapted to perform survey data collection but which can also operate as stand-alone computers.

1.4.3.1 *Dedicated surveying data recorders.* The design of data recorders dedicated solely to recording and processing surveying data has altered dramatically in the past 15 years. The first designs of data recorder, such as the Geodat 700, used punched paper tape as the recording medium. This approach suffered from many practical difficulties, including breakage of the tape (from mishandling or excessive humidity), a cumbersome and heavy recording mechanism and high power consumption. Although crude in design by present-day standards, this device was a significant development and

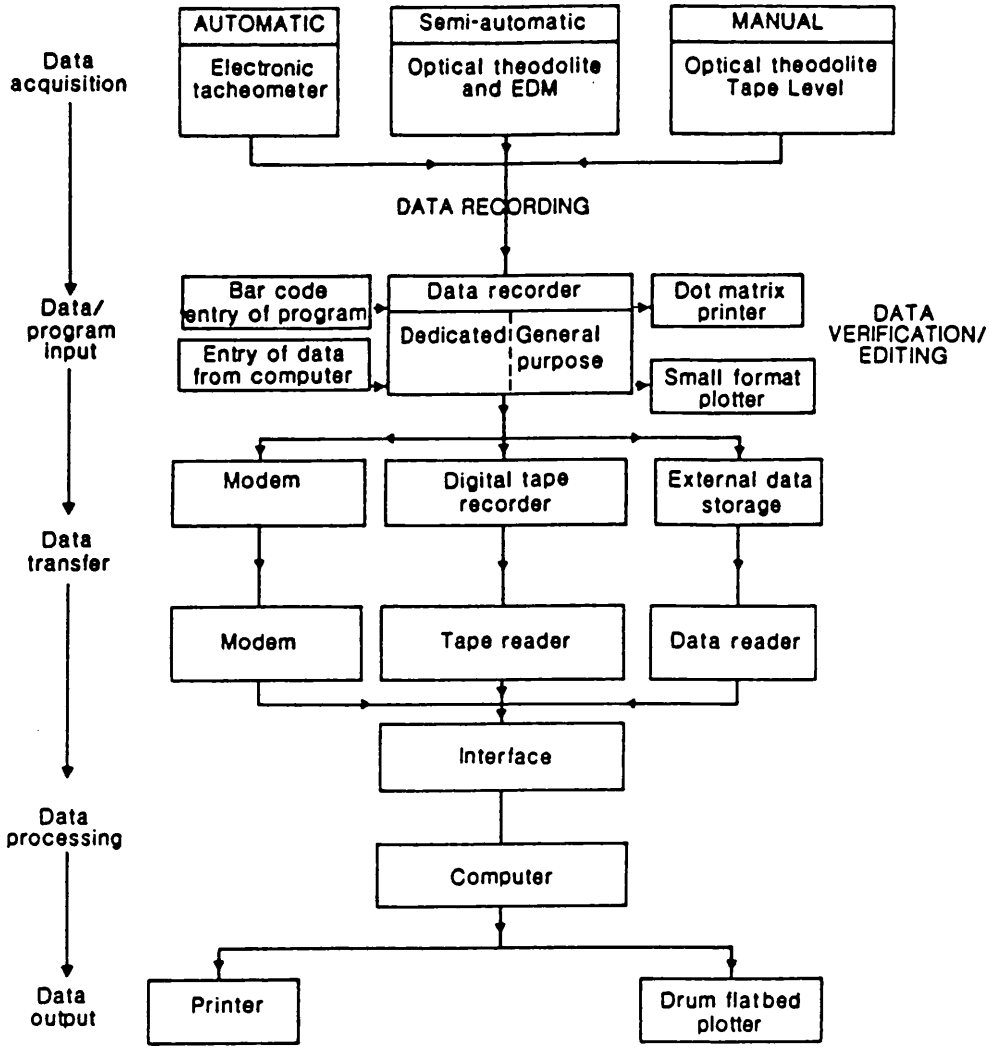


Figure 1.21 Diagram illustrating information flow from data acquisition to final output using data recorders.

provided the first effective method of automatically recording field observations.

The next development, which overcame many of the problems associated with paper tape recorders, involved the use of magnetic tape as the recording medium. The Wild TC-1 used this system. A digital tape recorder, protected against damp and dust by a sealed cover, enabled about 1200 points to be recorded on one digital tape. One of the main advantages cited by users of this technique was the relative ease of sending data recording from a remote location to a central processing facility for plotting. Although the magnetic tape units are no longer manufactured they are still used by several surveying organizations.

The most recent development has been the production of solid-state electronic data recorders. Most data recorders available at present are of this design. The distinction can be made between solid-state data storage modules which plug into an electronic tacheometer (such as the Wild GRM 10), and hand-held units with an integral keyboard. The latter are very much more common and range from devices based on electronic calculators (such as the Geodat 126), to more sophisticated devices such as the Kern Alphacord (Fig. 1.22). Kennie (1987) and Fort (1988) discuss the characteristics of these devices in more detail.

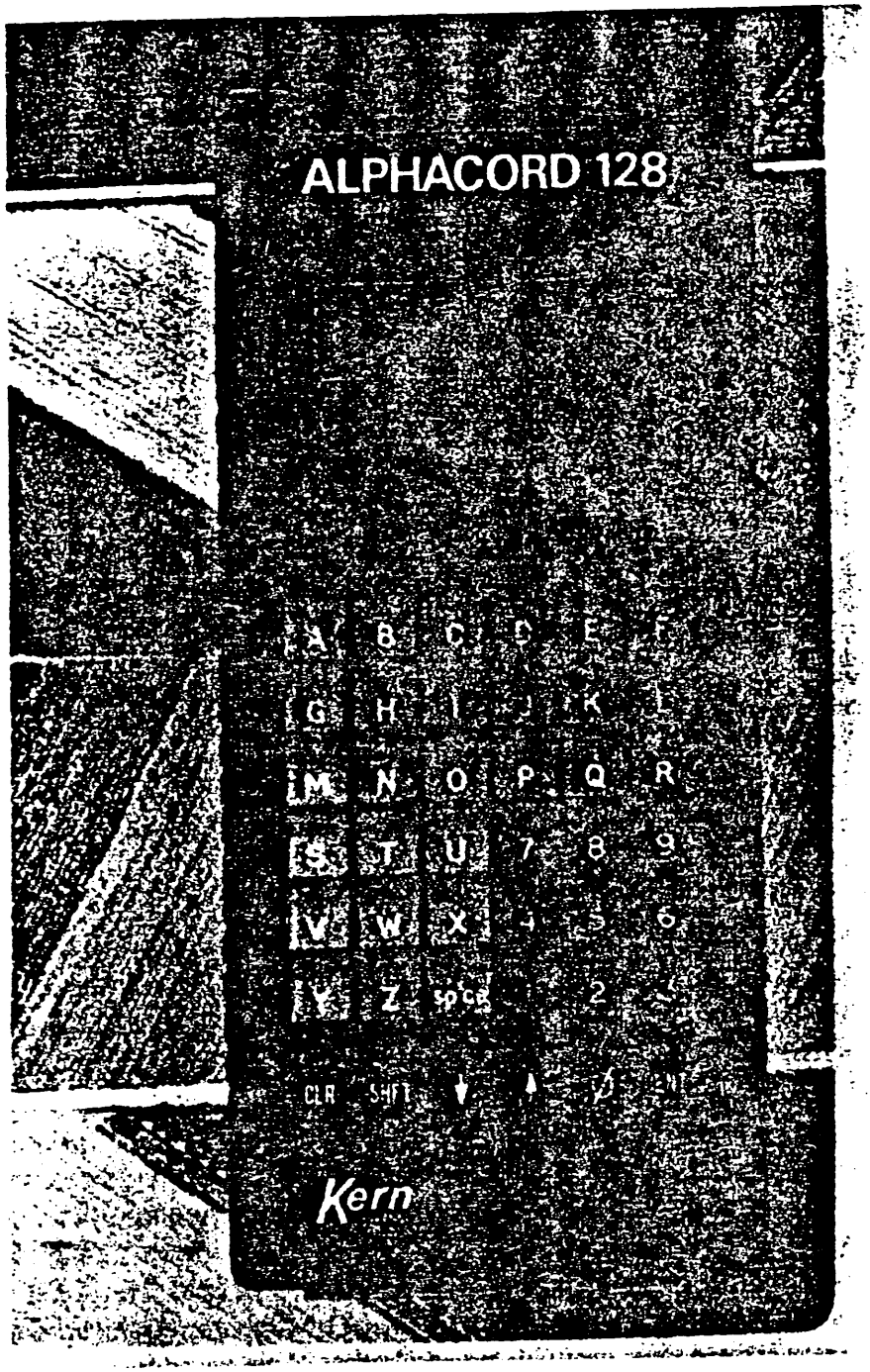


Figure 1.22 Kern Alphacord dedicated surveying data recorder. Courtesy Kern (Aarau).



Figure 1.23 Husky Hunter general-purpose data recorder.

In general terms, however, all dedicated designs suffer from two areas of weakness. First, users are often tied to a particular manufacturer's design of data recorder, since it can be difficult to interface recorders to different electronic tacheometers. Secondly, it may be difficult, if not impossible, to alter the software within the recorder.

1.4.3.2 *General purpose hand-held computers.* A more flexible approach which has become particularly common in recent years is the use of general purpose hand-held computers for data recording. Two different types of computer may be used: those which are fully weatherproof, of rugged construction and suitable for outdoor use (such as the Husky Hunter, Fig. 1.23), and those which are not designed for use outdoors (such as the Epson HX-20). This approach offers several advantages to the surveyor.

The first advantage is the ability to reprogram the form of data entry, for example, for specialist surveying operations or for other types of surveying equipment, which dedicated recorders would find difficult to accommodate. The second advantage is the flexibility to create specialist programs to process data in the field. Thirdly, in view of the general purpose nature of these devices it is also possible to use the computer for conventional surveying computational tasks and other office-based tasks such as word processing.

1.4.4 *Feature coding*

The method of storing data varies considerably, depending on the type of data recorder which is being used. It is possible, however, to distinguish between stored data which relates to the position of the object, and information which describes the object being surveyed.

Various options exist for recording field observations, ranging from recording raw observations (horizontal angle, slope distance and vertical angle), to storing three-dimensional coordinates. The former approach is preferred by many users, since it avoids the possibility of erroneously stored corrections (e.g. refractive index variations) inadvertently being applied to a series of measured distances.

Feature coding is the process of describing objects in the data recorder by means of some numeric or alphanumeric description. The coding process is normally carried out as an integral part of the field data acquisition phase, although instances of coding being performed after the survey has been performed have also been reported. While full alphanumeric coding may eliminate the need for a filed sketch over small sites, for large projects it is vitally important that a numbered sketch be prepared in the field. Coding systems vary considerably, and may reflect either the design of the data recorder being used or the in-house standards of a particular organization. Generally, however, they should enable the surveyor to describe certain characteristics of the object or how it should be represented on the final map. Thus a coding system could be used to define the type of annotation to be attached to a point feature. Similarly it may be used to describe whether points on a string (or series of linked points) should be joined by a straight or curved line.

1.5 Electronic coordinate determination systems (ECDS)

Electronic coordinate determination systems (ECDS), based on the use of two electronic theodolites and (normally) a portable computer, are becoming increasingly important for industrial metrology and deformation monitoring applications. They provide a portable, non-contact and real-time method of acquiring three-dimensional coordinates about objects ranging in size from vehicle sub-assemblies to oil rigs. While in theory objects smaller than about a metre in size can be measured by such systems, it is generally more appropriate in these circumstances to use direct methods such as coordinate measuring machines.

An alternative non-contact surveying technique for the measurement of objects such as those described above is close-range photogrammetry, and this approach is discussed in more detail in Chapter 8. Each method, however, offers complementary advantages to the surveyor and engineer. For example, photogrammetric techniques require only a relatively short time for data acquisition. This may be of value if it is important to reduce the interruption in the workplace to minimum during the survey. Photogrammetry also provides a permanent historical record of the raw data in the form of a pair of photographs. Thus, remeasurement can take place at a later stage if required. ECDS, in contrast, although requiring full access to the measurement surface during the survey, can provide real-time coordinate data. Real-time photogrammetric systems are also becoming available (Brown and Fraser, 1986) and section 8.7.2. However, they are at present generally research oriented, require substantial data processing facilities and are not widely available for routine measurements.

1.5.1 *Hardware*

The first commercial system developed for real-time coordinate determination was the Hewlett Packard CDS (Coordinate Determination System). This system, developed during the mid-1970s, was based on two HP 3820A electronic theodolites linked to a HP 9845 microcomputer (Johnson, 1980*b*). Although several systems were sold,

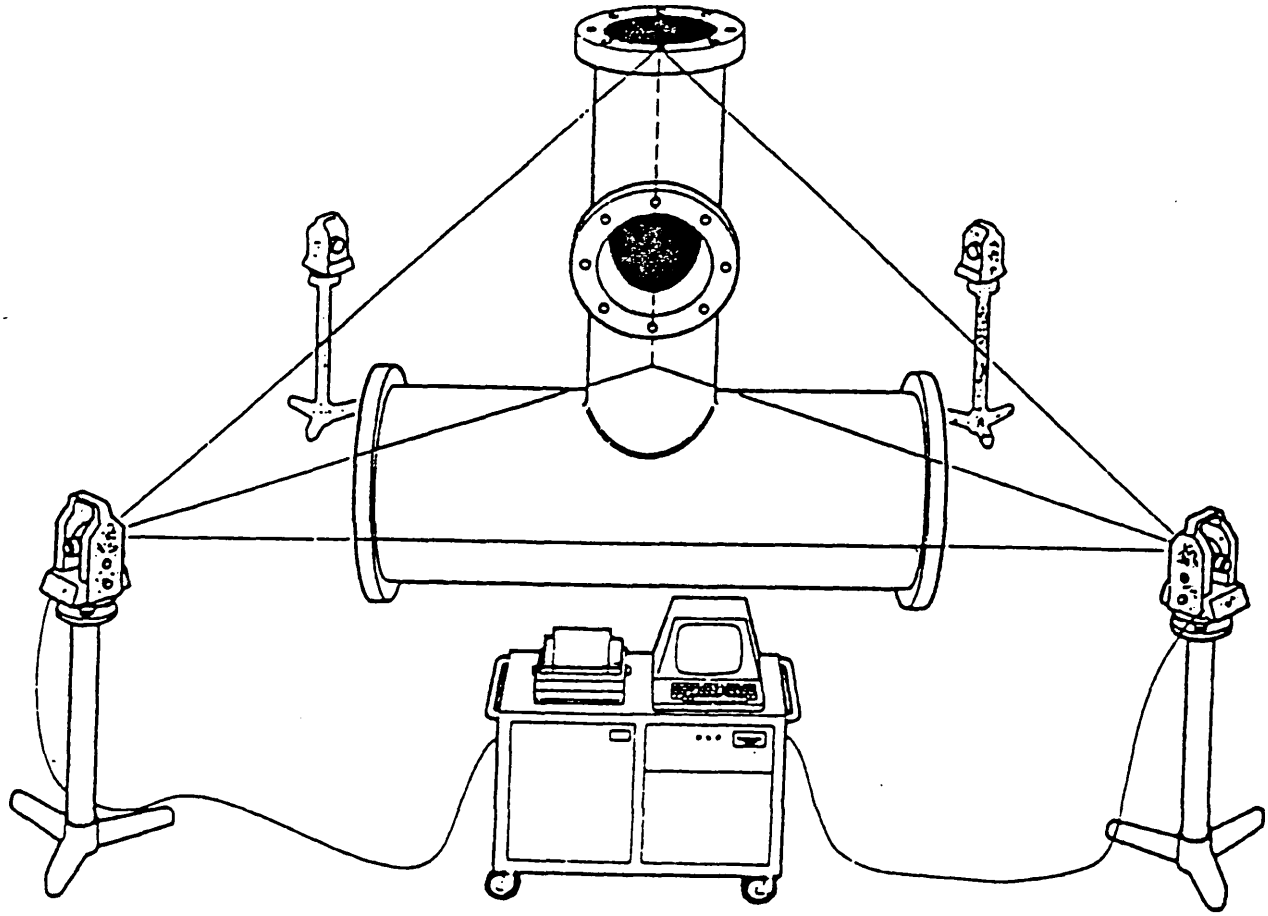


Figure 1.24 Wild RMS-2000 ECDS. Courtesy Wild (Heerbrugg).

primarily in the USA, development of the system ceased in 1980 following the decision by Hewlett Packard to discontinue its interest in surveying instruments. Two other systems which were also developed during this period were the Zeiss (Oberkochen) IMS (Industrial Measuring System) (Leitz, 1982), and the Keuffel and Esser AIMS (Analytical Industrial Measuring System) (Vyner and Hanold, 1982). Two systems which are currently particularly popular are the Wild Leitz RMS 2000 and the Kern ECDS 1.

1.5.1.1 *Wild Leitz Remote Measuring System (RMS).* The Wild Leitz Remote Measuring System (Fig. 1.24) consists of between two and four Wild T-2000 or T-2000S (a high-magnification, short-focus model) electronic theodolites linked to a Wang 2200 microcomputer for on-line real-time operations, or to a Wild GRE3 data collector for off-line storage of angular measurements. (See Katowski, 1986.)

1.5.1.2 *The Kern ECDS 1.* The Kern ECDS 1 (Electronic Coordinate Determination System) (Lardelli, 1984; Grist, 1986) consists of two or more Kern E1 or E2 electronic theodolite linked to a DEC Micro PDP 11/23 or IBM PC/AT computer for data processing. For applications where it is difficult to provide mains power, it is also possible to use a Husky Hunter portable computer (Fig. 1.23) together with the Dimensional Control System (DCS) software package to perform most of the processing requirements.

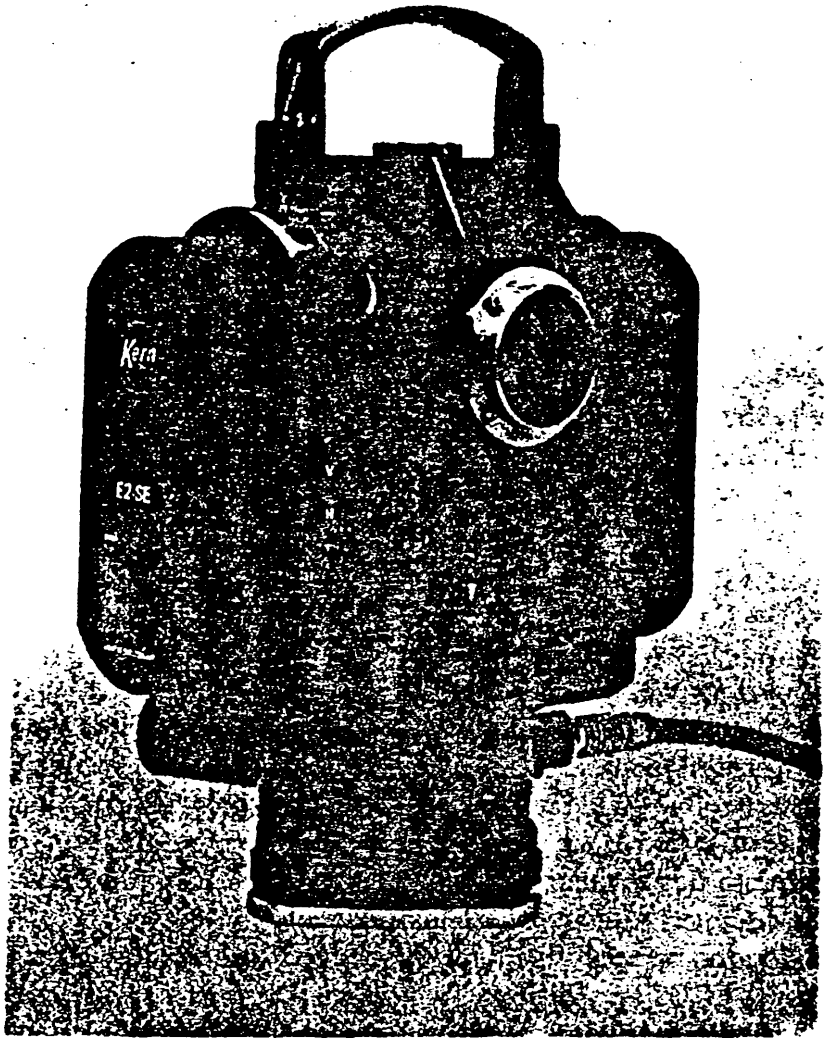


Figure 1.25 Kern E2-SE electronic theodolite. Courtesy Kern (Aarau).

1.5.1.3 *Kern SPACE system.* A development of the technique of real-time intersection (section 1.5.2) which attempts to automate the coordinate determination process is presented in Gottwald and Berner (1987) and Roberts and Moffitt (1987). Both papers discuss the development of a 'System for Positioning and Automated Coordinate Evaluation (SPACE)'.

The data acquisition element of SPACE is based on a modified version of the Kern E-2 electronic theodolite, while the processing phase is carried out using image processing techniques on an IBM PC/AT microcomputer. The theodolite, the Kern E2-SE (Fig. 1.25), incorporates several unique design features which aim to automate the target pointing and angle measuring process. First, servo-motors drive the horizontal and vertical axes of the instrument and also focus the telescope. The motors are designed to enable a positioning accuracy of up to $\pm 0.1''$ to be achieved, at speeds of up to 54° per second. Second, the theodolite includes within the telescope a Charge Coupled Device (CCD) sensor. The output from the CCD arrays are then fed to a video

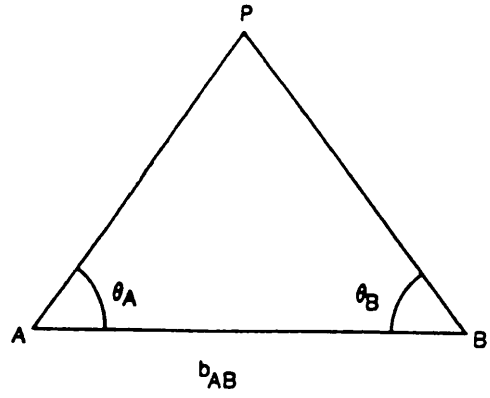


Figure 1.26 Two-dimensional intersection.

frame store in the computer. Image processing algorithms are then used to automatically drive the theodolite to either predetermined or targeted points (Zhou, 1986). A similar instrument has been designed by Wild (Heerbrugg) and is discussed in Katowski (1987).

1.5.2 *Principles of coordinate determination using ECDS*

1.5.2.1 *Real-time three-dimensional intersection.* All ECDS are based on the surveying technique known as intersection. In its simplest form this involves the simultaneous measurement, using two theodolites, of horizontal angles θ_A and θ_B from either end of a precisely measured baseline (Fig. 1.26). Standard surveying computational techniques can then be used to derive the two-dimensional coordinates of the unknown point *P*. Data acquired in this manner can also be more rigorously processed off-line. This approach is discussed subsequently in more detail.

The solution for the real-time three-dimensional coordinates of remote points is central to most ECDS applications. Two different approaches are used to solve this problem; the first uses trigonometric functions, and the second is based on the use of vector algebra. Both approaches can utilize least-squares estimation techniques if redundant observations have been recorded.

A set of equations can be derived by trigonometric methods for the general case in which the three-dimensional coordinates of the point *P* are required. Johnson (1980*a*) provides the following solution based on the measurement of the horizontal and vertical angles shown in Fig. 1.27.

$$\begin{aligned} X_P &= \frac{b}{2}\sin(\phi_T) \quad 1 + \frac{\sin(\theta_B - \theta_A)}{\sin(\theta_B + \theta_A)} \\ Y_P &= b\sin(\phi_T) \qquad \frac{\sin\theta_B \sin\theta_A}{\sin(\theta_B + \theta_A)} \\ Z_P &= b\sin(\phi_T) \qquad \frac{\sin\theta_B \cot\phi_A}{\sin(\theta_B + \theta_A)} \end{aligned} \tag{1.25}$$

The solution of these equations provides unique values in *X* and *Y* for the position of *P*.

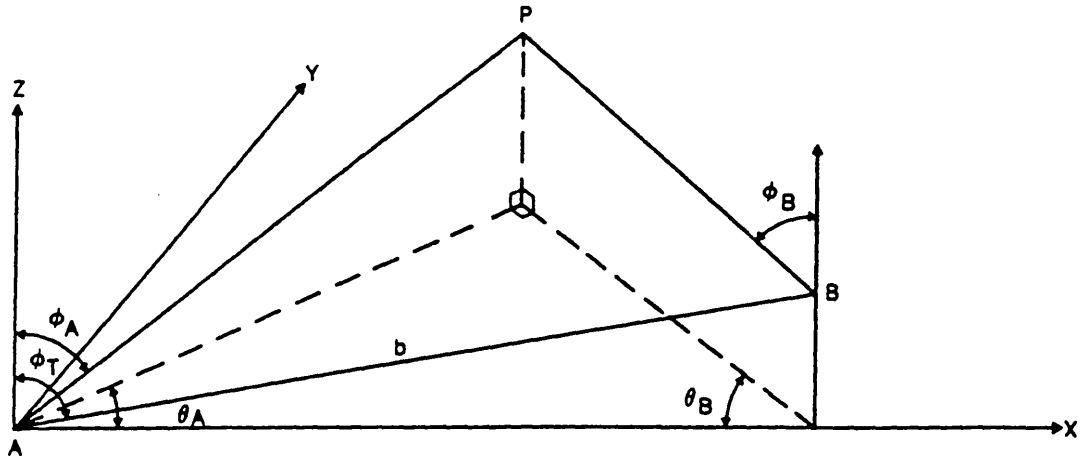


Figure 1.27 Three-dimensional intersection. At *A*, record horizontal angle *BAP* and zenith angles to *P* and *B*. At *B*, record horizontal angle *ABP* and zenith angle to *P* only.

It is possible, however, to derive two values for Z_P using the zenith angles ϕ_A and ϕ_B. A mean value for Z_P may then be derived. The difference between the two values, dZ, is used by several systems to check for gross errors in the pointing of the theodolites in the Z direction.

The solution of the intersection problem using vectors has been examined by Allan (1987) and Cooper (1986). Cooper derives the solution using the following argument. Consider two unit vectors p and q with scalar components δ and μ as illustrated in Fig. 1.28. The rays at the unknown point P will not in general intersect, since p and q are skew as a consequence of random observational errors. The position of P will, however, lie at some point along the vector e joining δp and μq, where e is given by

$$e = -\delta p + b + \mu q \tag{1.26}$$

Geometrically, the most probable position of P is at the midpoint of the vector of minimum length. The values of δ and μ which arise when e is a minimum are given by

$$\begin{bmatrix} \delta_0 \\ \mu_0 \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} 1 & p \cdot q \\ p \cdot q & 1 \end{bmatrix} \begin{bmatrix} b \cdot p \\ -b \cdot q \end{bmatrix} \tag{1.27}$$

If the coordinate system and angles measured in Fig. 1.27 are used, then

$$b = \begin{bmatrix} X_b \\ 0 \\ Z_b \end{bmatrix} \quad p = \begin{bmatrix} \sin\phi_A & \cos\theta_A \\ \cos\phi_A & \sin\theta_A \\ \cos\phi_A & 0 \end{bmatrix} q = \begin{bmatrix} -\sin\phi_B & \cos\theta_B \\ \sin\phi_B & \sin\theta_B \\ \cos\phi_B & 0 \end{bmatrix} \tag{1.28}$$

The solution of equation (1.28) for δ and μ can then be used to derive a value for the position vector of P relative to A from the following expression:

$$r = \delta_0 p + \frac{e}{2} = \tfrac{1}{2}(\delta_0 p + b + \mu_0 q) \tag{1.29}$$

The values of b, p and q can also be substituted in equation (1.26) to determine the components of the residual vector e.

The coordinates of point P can also be determined more precisely, using either of the approaches outlined above, by least-squares methods. Estimates of the unknown

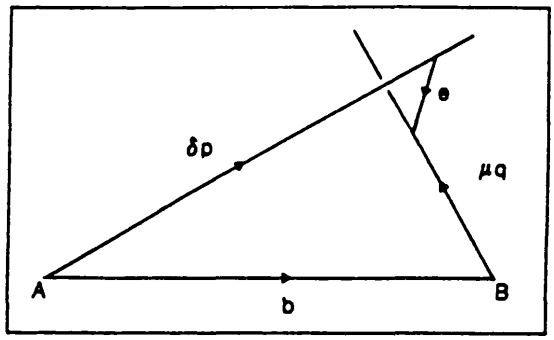


Figure 1.28 Analytical intersection using vectors (Cooper, 1986).

coordinates of P in the X, Y plane are initially determined using the horizontal angles measured from either end of the baseline. Next, one of the vertical angles and the estimated horizontal distance to the target (based on the approximate coordinates obtained previously) is used to determine the approximate height. Observation equations of the form described in Chapter 5 are then produced for each of the observed quantities (two vertical and two horizontal angles). These four equations in three unknowns (the corrections to the provisional coordinates) can then be determined by forming normal equations. The corrections obtained from the solution of the normal equations are then applied to the provisional coordinates, the process being iterated until the corrections become negligible. Although in theory it is possible to determine various statistical parameters to describe the precision of the coordinates (see Chapter 5), in practice these are not normally computed, since they would be of little practical value because statistically inconclusive.

The solution of equations (1.25) and (1.29) require a value for the length of the baseline b. This may be derived either by direct measurement or by the introduction into the object space of a scaling mechanism. The most common technique involves the use of a bar of known length. Calibration of the scaling bar is essential if systematic errors are to be avoided. It is also important to ensure that the bar is manufactured from a material with a low coefficient of linear expansion, such as carbon fibre, invar or in some cases glass. Alternatively, distances may be measured by laser interferometry (Chapter 2), or known distances on the object may be used. The scaling 'bar' is used to indirectly establish the length of the baseline. The technique is analogous to scaling a pair of photographs during the absolute orientation phase of setting up a stereoplotter. The use of a scaling bar generally reduces the measuring time and, more importantly, enables the baseline to be determined to a high level of accuracy with relative ease.

The simplest case occurs when a single distance is measured in the object space, as illustrated by Fig. 1.29. In this case an initial estimate of the length of the baseline (b') is used in conjunction with the observed horizontal angles θ_A and θ_B to determine the X and Y coordinates of the targets placed at either end of the scaling bar. From these coordinates the distance between the targets can be computed. By comparing this computed length of the bar S' with the known calibrated length S, a corrected value for the base b can be determined from the following expression.

$$b = \frac{S\,b'}{S'} \tag{1.30}$$

This revised value of b can then be used in equations (1.25) and (1.29) to determine the coordinates of the unknown points.

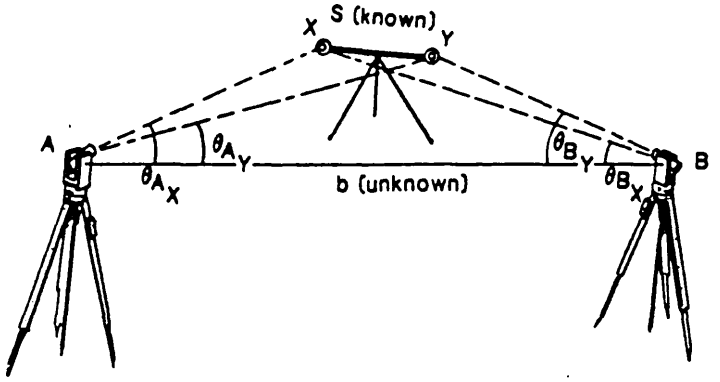


Figure 1.29 Indirect scaling using scaling bar.

An alternative approach to the processing of electronic theodolite data may be adopted in those situations where the simultaneous use of two theodolites providing real-time coordinates is not required. In such cases, the data acquisition may be accomplished at lower capital cost by using a single theodolite. The use of a single instrument is also advantageous since more flexibility in the positioning of the control stations is possible, and also there is no requirement to align the theodolite in relation to another instrument prior to obtaining measurements. More importantly, however, is the ability to rigorously process the data by least-squares techniques to fully exploit any redundant data. One particular approach which has been advocated involves the use of the bundle adjustment technique (Brown, 1985). The bundle technique is a computational method which is used in photogrammetry to simultaneously solve for the positions of unknown ground points and camera orientations using measured x and y photographic coordinates of known control points and unknown ground points. The method rigorously accommodates redundant data and provides a full error analysis. The application of the method to close-range photogrammetry is discussed in Chapter 8.

In order to utilize the bundle adjustment method in this context, it is currently necessary to transform angular measurements obtained from an electronic theodolite into equivalent photographic observations. These 'fictitious' photographs may then be processed using software packages such as STARS (Simultaneous Triangulation and Resection) (Brown, 1982), to provide ground coordinates and statistical information about the precision and accuracy of the position determinations.

Offline processing of data is also used in cases where it is necessary to transform coordinates from one coordinate reference system to another. For example, a common requirement is to transform from the 'theodolite system' to the 'object system' coordinates. In order to perform such a transformation it is necessary to know the coordinates, in both reference systems, of a minimum of three points. If more than three sets of coordinates are known, then a least-squares solution for each of the seven transformation parameters (three rotations, three translations and one scale change) can be carried out.

An alternative process which avoids the necessity of transforming coordinates from one system to another is to carry out all measurements in either the theodolite or the object reference frame. In the former case, if it is necessary to move the theodolites, their positions are re-established by recording angular measurements to a minimum of three points on the object and performing a resection calculation. In the latter situation,

particularly if the object to be measured is relatively small, the object may be mounted on a rotary table and repositioned relative to two fixed theodolite stations. Offline processing of the data can then be performed to compute theodolite-based coordinates (Griffiths, 1985).

1.5.2.2 *Factors influencing the accuracy of real-time intersection.* The overall accuracy of an ECDS based on real-time intersection is a function of several interrelated factors. These factors are, firstly, the instruments being used for angular measurements, and secondly the geometry of the measurement problem, and thirdly the influence of the environment.

The use of high-precision electronic theodolites able to resolve to a fraction of a second is essential if high-accuracy measurements are to be obtained. Two instruments which are particularly well suited to this are the Wild T-2000 and Kern E-2 which have been discussed in section 1.1. The internal precision of such theodolites will, however, only be translated into high-accuracy measurements if, firstly, the effect of any instrumental errors is eliminated, and secondly, if the pointing of the telescope to targets on the structure is precise.

The first requirement is normally achieved by conventional observational techniques (double face readings), although in some cases where a reduced-accuracy specification is applicable, single face pointings electronically corrected for instrument maladjustments may be acceptable. The second requirement is equally important and is primarily dependent on the design of the targets. A wide range of possibilities exist for targeting an object, including

(i) Natural points of detail
(ii) Adhesive targets of known dimensions
(iii) Points of detail artificially created during manufacture
(iv) Artificially created targets formed by projecting a laser beam through the objective of one theodolite.

Whatever technique is used it is important that the design of the targets enables precise centring of the telescope crosshairs over a wide angular range (60–120°). It is also important that the software can compensate for any systematic errors which may be introduced by the target design, as, for example, in (ii) above.

The second group of interrelated factors are those concerned with the geometrical relationship between the theodolites and the points on the object to be measured. The decrease in accuracy caused by an increase in the length of the *baseline distance* between the two theodolites is illustrated in Fig. 1.30, in which a nominal angular error is assumed. In order to limit this effect for high-accuracy measurements, the length of the baseline is normally restricted to between 5 and 10 m. The *angle of intersection* on the object of rays from adjacent theodolites has a major impact on the accuracy of the coordinate determination. Figure 1.31 illustrates this in two dimensions for a particular case where the baseline is 10 m and the standard error of the angular measurements is $\pm$ 1.5″. The intersection angle range suggested by Wild (Heerbrugg) for high-accuracy projects is between 78° and 142°.

The position and orientation of the scaling bar is also a significant consideration. It is important not only to locate the bar so that the angle of intersection at the target points on the bar is close to 90°, but also to ensure that the orientation of the bar enables a clear view of the targets to be achieved.

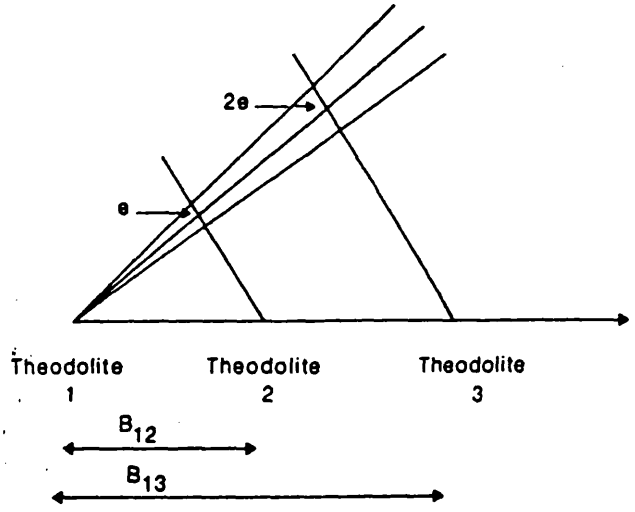


Figure 1.30 Influence of baseline length on coordinate determination.

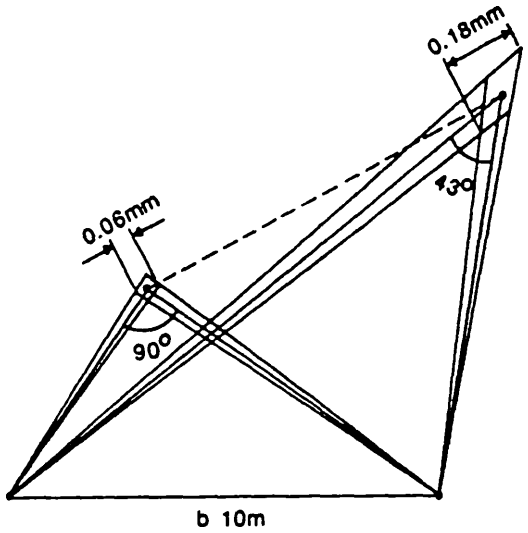


Figure 1.31 Influence of intersection angle on coordinate determination (adapted from Leitz, 1982).

Environmental effects such as vibration, wind, temperature fluctuations, atmospheric refraction and varying lighting conditions are some of the most influential factors which limit the accuracy of this technique. While several of the factors can be controlled, for example by ensuring that machinery adjacent to the object to be measured is not operated during measurement, ultimately uncontrolled environmental fluctuations which cause varying refraction conditions limit the accuracy achievable with such a measuring system.

1.5.3 *Applications*

As mentioned previously, ECDS are predominantly used in large-scale industrial metrology. The following two examples illustrate the use of these techniques in this field.

1.5.3.1 *The Wild Leitz RMS 2000 and the measurement of a parabolic antenna.*

Katowski (1985) describes the use of the Wild Leitz RMS system to measure,

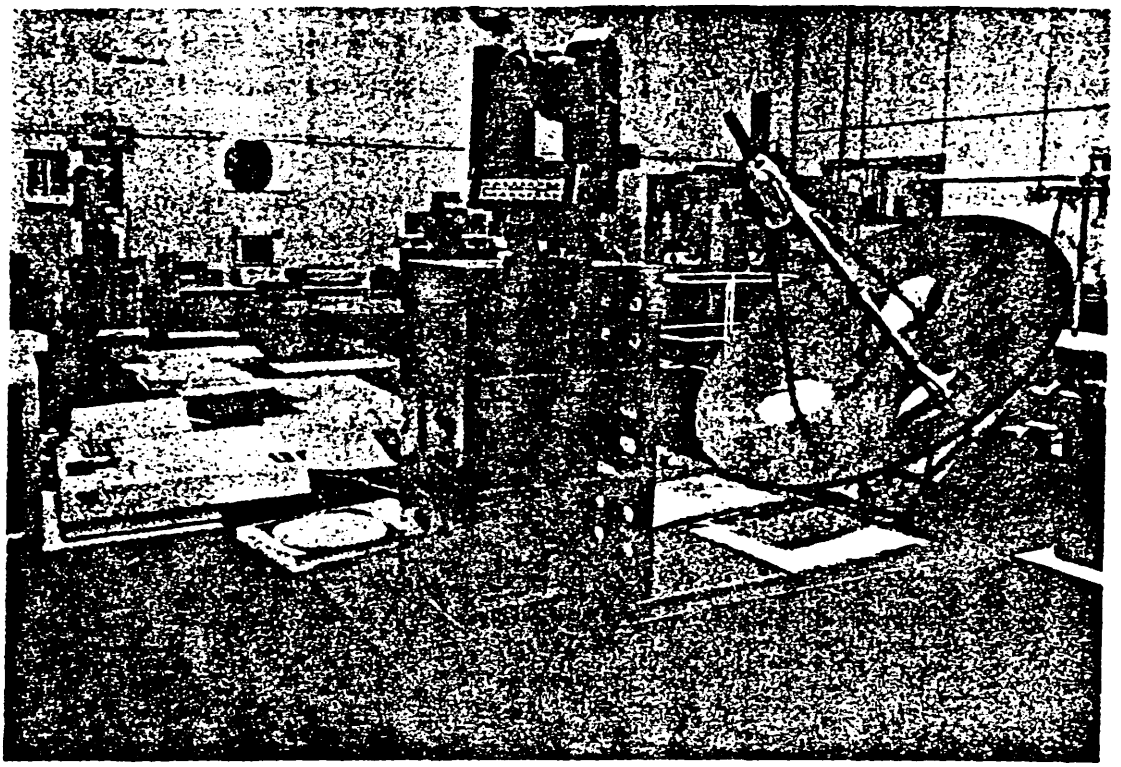


Figure 1.32 Use of Wild-Leitz RMS 2000 to monitor paraboloid reflector. Photograph courtesy European Space Research and Technology Centre (ESTEC).

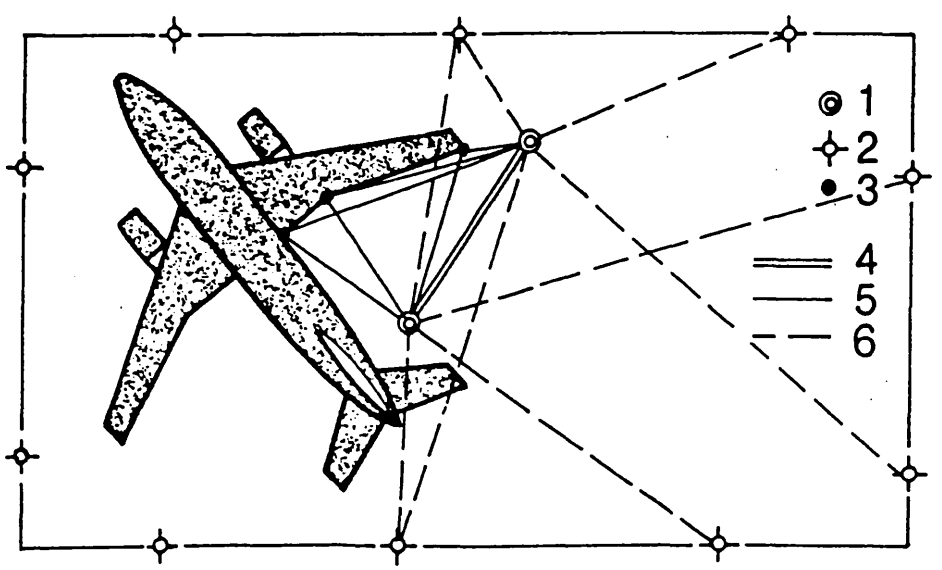


Figure 1.33 Kern ECDS-1: network configuration for aircraft dimensional control. Free choice of both base points (1), computation of its coordinates and the baseline data (4) by means of the directions (6) at the known fixed points (2), intersection (5) at the object points (3). Courtesy Kern (Aarau).

under varying environmental conditions, the shape of the paraboloid reflector and antenna for the Ulysses satellite by the European Space Research and Technology Centre (ESTEC) in the Netherlands. The reflector measured had a diameter of 1.65 m and measurements were recorded to 26 pre-targeted points using the measuring arrangement illustrated in Fig. 1.32. The measurements were subsequently processed to define the reflector focal point, the focal distance using the procedure outlined in Griffiths and Thomsen (1985). Several other applications of the Wild-Leitz system are discussed in Woodward (1987).

1.5.3.2 *The Kern ECDS and aircraft dimensional control.* In the assembly of large aircraft it is now becoming common practice to use systems such as the Kern ECDS to check the dimensional quality control of the various sub-assemblies which form the structure of the aircraft. Also, when the aircraft has been fully assembled, a full dimensional survey of the entire structure is carried out to ensure that the final dimensions are within tolerance and consequently unlikely to generate excessive stresses within the structure. Figure 1.33 illustrates how a network of control points (2) can be used as a basis for resecting theodolite stations (1) from which measurements can be taken onto the aircraft (3) to determine its dimensions. More recent applications of the ECDS have involved submarine hull mapping (Bethel and Sedej, 1986), and the alignment and positioning of the magnets for the Stanford Linear Collider (Oren *et al.*, 1987).

1.6 Conclusions

This chapter has discussed the current technology for the measurement and storage of angle and distance observations. The impact of these developments on surveying practice and productivity levels has been significant. Although such devices are now widely used in the land surveying community, it is only recently that they have become more common in general civil engineering and building projects. However, if the cost of these instruments continues to drop, this trend is likely to increase.

Finally, for the future it has been suggested (Ashkenazi, 1986), that the current hierarchical geodetic control frameworks based on angle and distance measurements may eventually be replaced by point measurements to satellite systems such as the Global Positioning System (GPS) (Chapter 4) or by inertial systems (Chapter 3). This seems unlikely to occur during the next 10–15 years, and in any case it has been argued that it would be inappropriate to become dependent on a system such as GPS which could, for political reasons, become inaccessible to non-military users. Consequently, the need for angle and distance measurements will continue, providing complementary data to those produced by satellite and inertial surveying systems.

References

Allan, A. (1987) The principles of theodolite intersection systems. *Proc. 2nd FIG/ISPRS Industrial and Engineering Survey Conf., September 2–4, London*, 16–31.

Ashkenazi, V. (1986) Satellite geodesy—its impact on engineering surveying. *Proc. 18th FIG Congress, Toronto*, Paper 502/1, 70–85.

Ashkenazi, V. and Dodson, A. (1975) The Nottingham multi-pillar baseline. *Proc. 16th UGG General Assembly, Grenoble*, 22 pp (also reprinted in *Civil Engineering Surveyor*, June 1977, 1–8.

Bethel, J. and Sedej, J. (1986) Submarine hull mapping and fixture set out: an operation of the ECDS system. *46th ASP-ACSM Convention, Washington DC*, 92–96.

Bevin, A.J. (1978) Accuracy and calibration of short range EDM. *New Zealand Surveyor* **28** (1) 15–33.

Brown, D.C. (1982) STARS: a turnkey system for close range photogrammetry. *Int. Arch. Photogramm.* **24**, V/1, *Precision and Speed in Close Range Photogrammetry*, 68–89.

Brown, D.C. (1985) Adaptation of the bundle method for triangulation of observations made by digital theodolites. *Proc. Conf. South African Surveyors (CONSAS)*, Paper 43, 18 pp.

Brown, D.C. and Fraser, C.S. (1986) Industrial photogrammetry—new developments and recent applications. *Photogramm. Record* **12** (68) 197–217.

Burnside, C.D. (1982) *Electromagnetic Distance Measurement*. 2nd edn., Granada, London, 240 pp.

Cooper, M.A.R. (1986) Spatial positioning, vector notation and coordinates. *Land and Minerals Surveying* **4** (12) 623–625.

Deeth, C.P., Dodson, A.H. and Ashkenazi, V. (1978) EDM accuracy and calibration. *Proc. Symp. on EDM*, Polytechnic of the South Bank, London, 24 pp.

Dodson, A. (1978) EDM systematic errors and instrument calibration. *ASCE Journal* (now *Civil Engineering Surveyor*) **3** (4) 20–22.

Fort, M. (1988) Data recorders and field computers. *Civil Engg Surveyor*, July/August, 8 pp.

Gorham, B.J. (1976) Electronic angle measurement using coded elements with special reference to the Zeiss Reg Elta 14. *Survey Review* **23** (180) 271–279.

Gottwald, R. and Berner, W. (1987) The new Kern system for positioning and automated coordinate determination—advanced technology for automated 3-D coordinate determination. *47th ASP-ACSM Spring Convention, Baltimore*, 7 pp.

Griffiths, J. (1985) Measuring an object with the aid of a rotary table. Wild RMS-2000 Supplement No. 8, 11 pp.

Griffiths, J. and Thomsen, P. (1985) Measuring a parabolic antenna using a 3D surface element. Wild RMS-2000 Supplement No. 2, 9 pp.

Grist, M.W. (1986) Real time spatial coordinate measuring systems. *Land and Minerals Surveying* **4** (Sept.) 458–471.

Hodges, D.J. (1982) Electro-optical distance measurement. *University of Nottingham Mining Magazine* **34**, 19 pp.

Johnson, D.R. (1980*a*) An approach to large scale non-contact coordinate measurements. *Hewlett Packard Journal*, September, 16–20.

Johnson, D.R. (1980*b*) An angle on accuracy. *Quality*, November, 4 pp.

Katowski, O. (1986) RMS-2000, a system for non-contact measurement of large objects. Wild Heerbrugg, 15 pp.

Katowski, O. (1987) Automation of electronic angle measuring instruments. *Proc. 2nd FIG/ISPRS Industrial and Engineering Survey Conf., September 2–4, London*, 230–240.

Katowski, O. and Salzmann, W. (1983) The angle measurement system in the Wild Theomat T-2000. Wild Heerbrugg, 10 pp.

Keller, W. (1978) Deformation measurements on large dams. Kern (Aarua) Publication, 15 pp.

Kennie, T.J.M. (1983) Some tests of retroreflective materials for electro-optical distance measurement. *Survey Review* **27** (207) 3–12.

Kennie, T.J.M. (1984) The use of acrylic retro-reflectors for monitoring the deformation of a bridge abutment. *Civil Engineering Surveyor*, July/August, 10–13.

Kennie, T.J.M. (1987) Field data collection for terrain modelling. *Proc. Short Course on Terrain Modelling in Surveying and Civil Engineering*, University of Surrey, 12 pp.

Kennie, T.J.M., Brunton, S., Penfold, A. and Williams, D. (1988) Electro-optical distance measurement: calibration methods and facilities in the United Kingdom. *Land and Minerals Surveying*, January, 12–16.

Lardelli, A. (1984) ECDS-1: an electronic coordinate determination system for industrial applications. Kern Aarau, 11 pp.

Leitz, H. (1982) Three dimensional coordinate measurement with electronic precision theodolites. *Proc. NELEX'82, 14–16 September*, Paper 4/1, 10 pp.

Maurer, W. (1985) Kern E2: accuracy analysis of the graduated circle. *Kern Bulletin* **37**, 3–6.

Munch, K.H. (1984) The Kern E2 precision theodolite. *Proc. 17th FIG Congr. Sofia*, Commission 5, 22 pp.

Murname, A.B. (1982) The use of the Kern Mekometer ME3000 in the Melbourne and Metropolitan Board of Works. *Proc. 3rd Int. Symp. on Deformation Measurement by Geodetic Methods, Budapest.*

Oren, W., Pushor, R. and Ruland, R. (1987) Incorporation of the Kern ECDS-PC software into a project oriented software environment. *47th ASPRS-ACSM Convention, Baltimore*, 7 pp.

Rawlinson, C. (1976) Automatic angle measurement in the AGA 700 Geodimeter—principles and accuracy. *Survey Review* **23** (180) 249–270.

Roberts, T. and Moffitt, N.M. (1987) Kern system for positioning and automated coordinate evaluation: a real time system for industrial measurement. *47th ASPRS-ACSM Convention, Baltimore*, 5 pp.

Rüeger. J.M. (1988) *Introduction to Electronic Distance Measurement*. 2nd edn, 128 pp.

Schwendener, H.R. (1972) Electronic distances for short ranges: accuracy and calibration procedures. *Survey Review* **21** (64) 273–281.

Spofford, P.R. (1982) Establishment of calibration baselines. NOAA Technical Memo NOS/NGS/8, National Geodetic Information Centre, 21 pp.

Sprent, A. (1980) EDM calibration in Tasmania. *Australian Surveyor*, **30**(4) 213–227.

Sprent, A. and Zwart, P.R. (1978) EDM calibration—a scenario. *Australian Surveyor* **29** (3) 157–169.

Toft, D. (1975) Tests of short range EDM equipment in the Ordnance Survey. *Commonwealth Survey Officers Conference*, Paper C-1, 15 pp.

Vyner, N.A. and Hanold, J.H. (1982) AIMS- analytical industrial measuring system. *Int. Arch. Photogramm.* **24**, V/1, *Precision and Speed in Close Range Photogrammetry*, 524–532.

Woodward, C.A.W. (1987) Practical experiences with the Wild-Leitz RMS 2000 system. *Proc. 2nd FIG/ISPRS Industrial and Engineering Survey Conf., September 2–4, London*, 42–53.

Zhou, G. (1986) Accurate determination of ellipse centres in digital imagery. *ASPRS-ACSM Convention, Washington DC*, 256–264.

THE USE OF ACRYLIC RETROREFLECTORS FOR MONITORING THE DEFORMATION OF A BRIDGE ABUTMENT

T.J.M. Kennie,
Lecturer in Engineering Surveying,
University of Surrey

INTRODUCTION

Deformation monitoring surveys using conventional land surveying methods, photogrammetric survey methods or specialist goetechnical methods are used to measure the amount by which an engineering structure has moved either vertically or horizontally over specific periods of time.

The information which the survey provides can be of critical importance and may be used to indicate either:

(a) The structure is stable and consequently safe.
(b) The structure is experiencing small random movements which are not imposing significant forces on the structure.
(c) The structure is experiencing small, localised systematic deformations, for example as caused by seasonal effects, which may, or may not, be significant.
(d) Lastly the most significantly, the structure is experiencing deformation which is increasing as a function of time. This may indicate that remedial measures have to be taken, in some cases immediately, in order to avoid catastrophic consequences.

For a variety of reasons, surveys of this type have become increasingly important in recent years. Several interrelated factors have been responsible for this increase in interest. Firstly, for many types of structures it is now a mandatory element of the Civil Engineering process that a monitoring survey be commissioned. Notable in this respect are, for example in this country, reservoirs, which now require to be monitored following the implementation of the 1975 Reservoirs Act. A further example, from Switzerland, is the requirement by the Swiss Federal Government for all dams with a height greater than 15m and a cubic capacity greater than 50,000m³ to be monitored. A second factor which has led to this increase in activity has been the speed of development both in the manufacture of precise survey instrumentation and also in high speed, low cost computing facilities. This, in conjunction with improvements in very elegant and highly sophisticated software for the design and analysis of surveying observations, now provides the surveyor with a highly accurate measurement system. Thirdly, the tolerance to which civil engineers are now designing and constructing many modern structures necessitates a much higher order of accuracy in the initial dimensional control and also in the subsequent deformation monitoring. Indeed this factor may have acted as a "springboard" for many of the developments in software and instrument design. Finally, it is hoped that to some extent civil engineers are now more aware of the possibilities offered by modern survey techniques and are, therefore, requesting this type of survey more frequently.

The aim of this paper is, therefore, to describe briefly one such deformation monitoring survey and to discuss the potential offered by an alternative retroreflective target which was used during part of the project.

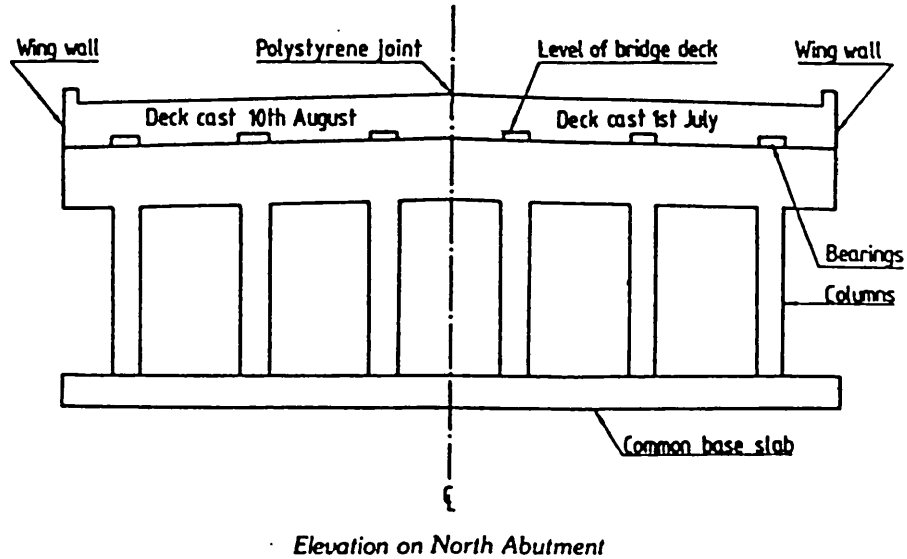
PROJECT OUTLINE

The deformation monitoring survey formed part of a research project carried out by the Department of Civil Engineering, University of Surrey. The primary aim of the research project was to gather data relating to the effect of earth pressures on the abutments of bridge No. 15 on the Chertsey to Wisley section of the M25 during several stages of the construction process. These data were subsequently to be compared with the estimated earth pressures used by the designers and eventually lead to the formulation of some general guidelines for the design of bridge abutments of the type under investigation.

The structure to be monitored consisted of a pair of skeleton abutments, one at either end of the bridge. A skeleton abutment consists of a transverse cill beam supporting the end of the bridge deck, with two or more buried columns carrying the load down to a common base slab. In this particular case, six columns transmitted the load to the base slab. Figure 1 illustrates the general layout of this particular design of abutment. The detailed measurement of earth pressures was restricted to two columns on each abutment. Several geotechnical instruments were used to provide details of the forces on all faces of the columns.

The measurement of lateral earth pressures was carried out using vibrating wire earth pressure cells. Pressure cells were installed on all sides of the columns and at three different levels. Analysis of these data was used to determine the magnitude of earth pressures and also the extent of any "spill through" of earth between the columns.

Vertical earth pressures were checked at base slab level and at differing levels in the fill since it was anticipated that the observed pressures may be considerably greater than those associated with the nominal weight of the backfill. The wires from the earth pressure cells were passed underground in trunking to a control panel situated about 30m



Elevation on North Abutment

Figure 1: Design of a Skeletal Abutment

from the abuntments. The position of the control panel also coincided with the positions of the survey pillars. Finally, two inclinometer tubes were installed. The first, in one of the columns was used in order to determine the complete deflection profile down to base slab level and the second, installed in the backfill zone between the two columns, was used to enable an assessment to be made of the relative movement of the fill compared to the abutment.

The primary purpose of the surveying operation in this particular project was to provide independent confirmation of the results obtained using the geotechnical instrumentation. Therefore, whilst periodic monitoring with high absolute accuracy was considered highly desirable, it was felt that a more appropriate and cost effective solution to adopt was to obtain high precision measurements over the short time periods when significant pressures were being exerted on the abutments. These periods occurred during the backfilling operation to bury the base slab and columns and also during the exertion of pressure which occurred when the bridge deck was cast. Thus the approach adopted was concerned primarily with the measurement of the relative rather than the absolute deformation of the structure.

SURVEYING OPERATIONS

The surveying requirements for the project consisted of two main operations.

Firstly it was required that a network of six surveying stations be established around the site in order to provide an absolute reference system which would be used for determining the positions of the inclinometers and pressure cells and which could also be used as additional survey control by the site staff for setting out purposes.

Several constraints were imposed on the positioning of the survey stations. Firstly the position had to be sited so as to be able to measure the deflection of the abutments. In view of the instrumentation available, (Kern DKM—2A one second theodolite and Geodimeter 120 EDM) it was decided that the deflection would be best determined by measuring angular deflection and distances from a position perpendicular to the expected direction of movement. Secondly, the survey stations should be positioned so as to be intervisible both before construction began and also subsequently in order to enable remeasurement at a later date. Finally it was important that the stations should be sited outside the zone of movement associated with the abutments. The final layout of the survey stations is illustrated by Figure 2.

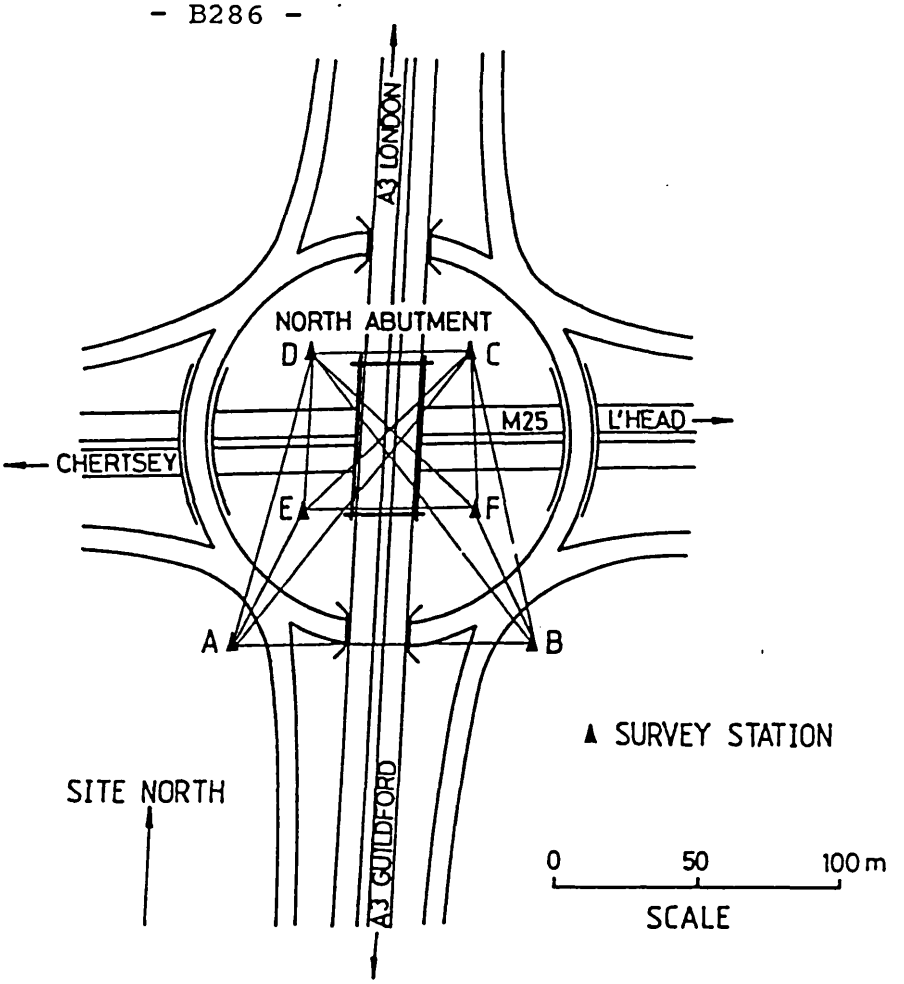


Figure 2: Site Plan of Wisley Interchange and Positions of Survey Stations

The second surveying operation was concerned with the measurement of the deflection and/or translation of the abutment during backfilling and also immediately following the casting of the bridge deck. In order to carry out this operation, extensive use was made of acrylic retroreflectors as a monitoring target for distance measurement.

RETROREFLECTIVE TARGET

The primary requirement of a survey monitoring target is that either it is able to be precisely relocated on the structure under investigation or alternatively that it can be permanently affixed to the structure. If the latter approach is adopted it is also important that the target is unobtrusive and aesthetically pleasing and thus, hopefully, on a more practical level, less likely to attract the attention of vandals.

In this particular case the latter

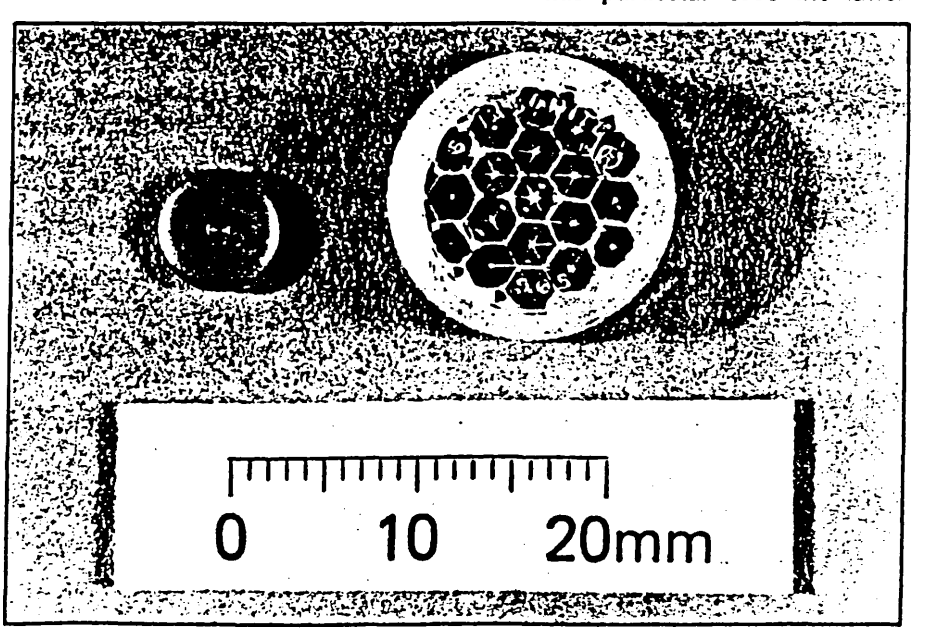


Figure 3: Demec Stud and Acrylic Retroreflector

method was adopted and a target used which consisted of a "Demec stud" for angular measurements and a small 1cm acrylic retroreflector for distance measurements (see Figure 3).

It can be seen from Figure 3 that both targets are extremely small. However, the Demec stud is an extremely good target over short distances (up to 80m) since the indentation in the stud is almost identical to the width of the crosshairs on a one second theodolite. Thus repeat pointings to the same point can be made very precisely. The possibility of using acrylic retroreflectors as survey targets has been discussed by Done and Methley (1980). Furthermore, tests carried out by Kennie (1983) indicated that for short distances the results obtained with retroreflectors of this nature do not differ to any significant degree from those measured using conventional glass retroreflectors. It has been shown, however, that for high precision the entrance angle (angle between EDM signal and a line perpendicular to the face of the retroreflector) should be no greater than 10°.

The monitoring targets for the latter part of the backfilling operation and the bridge deck survey were permanently affixed to the wing walls of the abutment. Five sets of targets were established with the acrylic retroreflector being offset vertically by the distance which exists between the optics of the Geodimeter 120 and those of the Kern DKM—2A theodolite. This arrangement is illustrated by Figure 4.

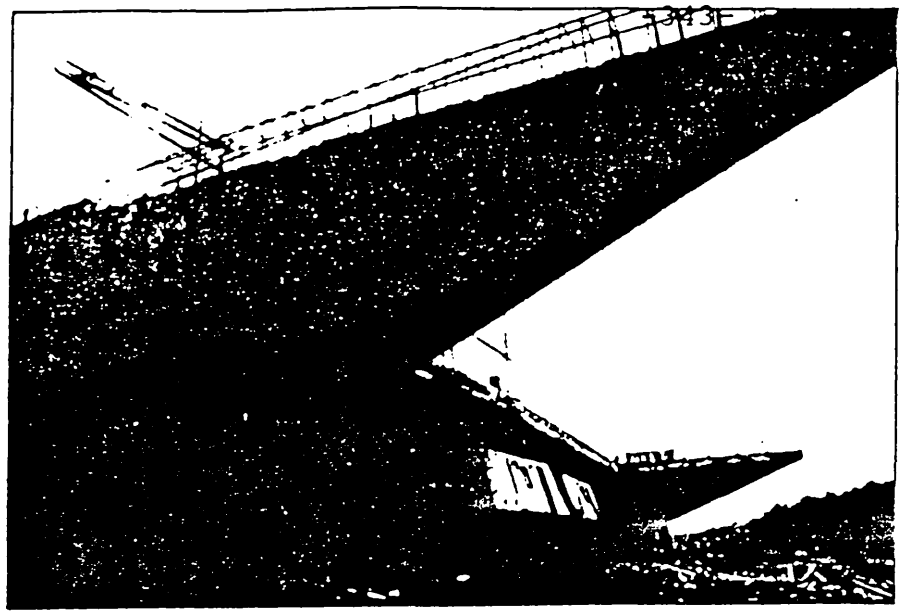


Figure 4: Positions of Survey Targets on Wing Walls

RESULTS

The results of the horizontal control survey were computed in a standard manner using the method of variation of coordinates. A total of thirty-six observation equations were formed and solved to provide final coordinates for stations A to F. Subsequent error analysis indicated a high internal precision for the survey, with the average standard error for the final coordinates being ± 1.9mm.

The results for the deformation monitoring operation can best be illustrated by examining one particular set of results obtained during the casting of the bridge deck. It should be mentioned that the deck was cast in two sections from the centre outwards, and Figure 5 illustrates the bridge deck during the first concrete pour.

Angle and distance observations were taken from survey stations C and D in order to monitor the North abutment. By measuring the angular changes which occurred relative to stations A and B, initially at twelve hour and subsequently longer intervals, it was possible to determine the angular deflection of the bridge abutment. One particular set of data for the North abutment is illustrated in Figure 6.

Examination of the graph clearly illustrates that the maximum deflection occurred approximately 24 hours after the beginning of the deck pour.

A clearer indication of the deformation can however, be inferred from Figure 7 which illustrates the vectors of deformation at each of the target points on the abutment after 24 hours, the vertical components being calculated from the vertical angled and EDM distances measured to the Demec stud and acrylic retroreflectors. The horizontal components were derived in a similar manner.

It would appear from the magnitude and direction of the results that the abutment was subjected to two distinct movements, involving both translatory and rotational elements.

Correlation of the survey data with that obtained independently by geotechnical methods seems to suggest that the translatory element of the movement was caused by the bending of the columns of the abutment. The rotational element, on the contrary, seems to be related to a rotation of the common base slab of the abutment. The results obtained for the deck cast subsequently follow the trends associated with the first set of results. In this case, however, the magnitude of the column bending appears to be less than that obtained during the first deck pour; this would seem to suggest that the adjacent deck was restraining the action of the forces on the abutment.

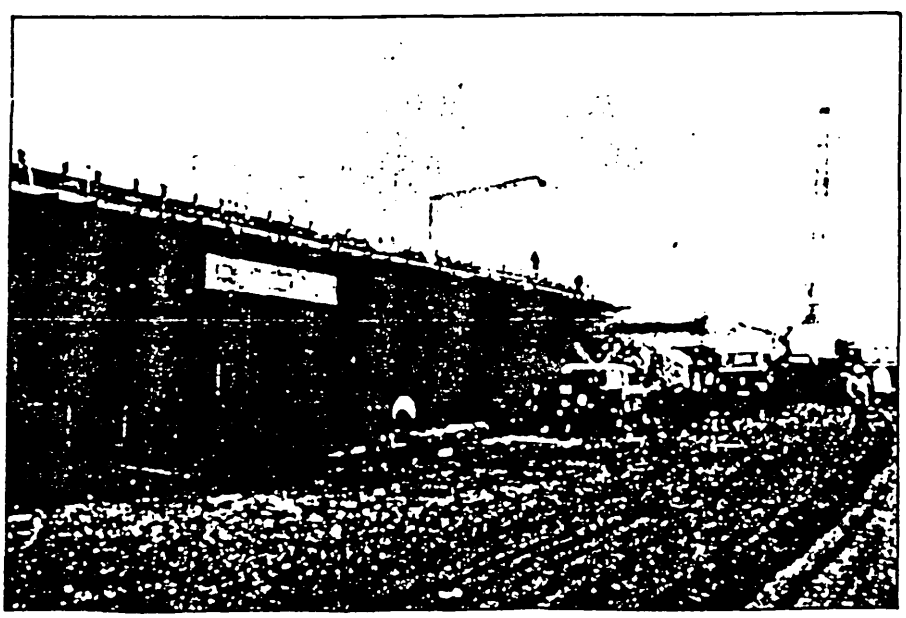


Figure 5: Bridge Deck Survey First

CONCLUSIONS

A method of monitoring a bridge abutment during construction using acrylic retroreflectors has been

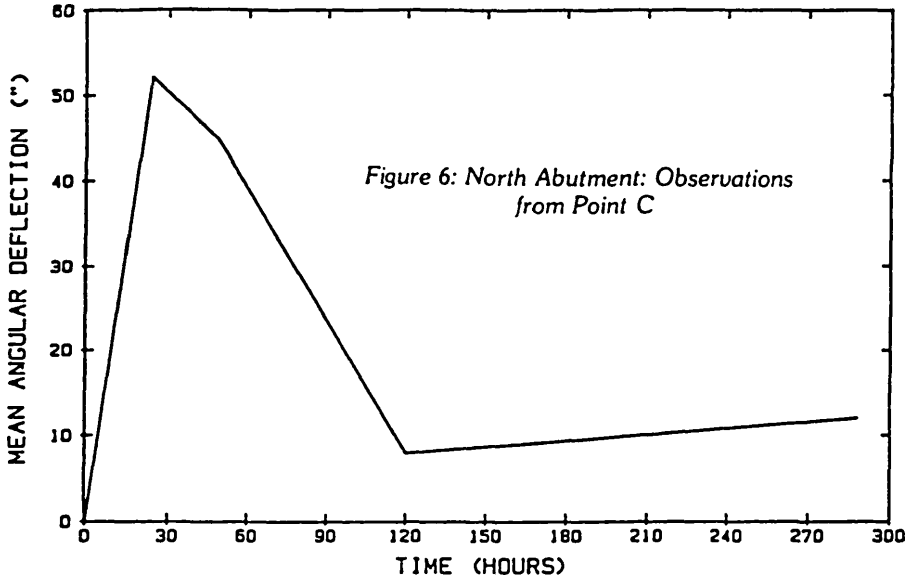


Figure 6: North Abutment: Observations from Point C

outlined. Whilst the techniques adopted were relatively unsophisticated, they offered two principle advantages: simplicity and speed, the latter being particularly important during the latter part of the project.

The results obtained related to the deformation of the structure over a short time span and relative to a set of datum values. Thus the results obtained, whilst very precise, may not necessarily be of high accuracy. For situations where this limitation is acceptable the use of acrylic retroreflectors and conventional EDM instruments can be an appropriate means of assessing structural deformation.

REFERENCES

Done, P. and Methley, B.D.F., 1980 - The Use of Acrylic Reflectors with Electro-optic Distance Measuring Instruments. ***Survey Review*** *25, 195, 215-222.*

Kennie, T.J.M., 1983 - Some Tests of Retroreflective Materials for Electro-optical Distance Measurements - Part 2, ***Survey Review,*** *27, 207, 3-12.*

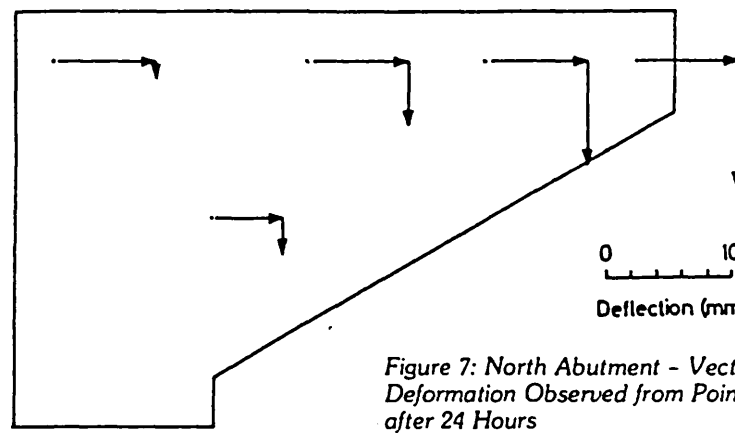


Figure 7: North Abutment - Vectors of Deformation Observed from Point C after 24 Hours

Survey Review Vol. 27, 207, January 1983

SOME TESTS OF RETRO-REFLECTIVE MATERIALS FOR ELECTRO-OPTICAL DISTANCE MEASUREMENTS

T. J. M. Kennie

University of Surrey

Due to the length of this paper it has been divided into two parts. Part 1 describes materials and equipment used in the tests. Part 2, which will appear in the next issue of the Survey Review, includes the analysis and results of the tests and describes some applications.

PART 1

Abstract

A series of materials are tested in order to assess their retro-reflective performance over short ranges with electro-optic distance measuring instruments.

Introduction

In recent years much attention has been focused on the development of short range electronic distance measuring (EDM) instruments, particularly those of the electro-optic type employing visible light and infra-red radiation. The principle of measurement is well documented (Burnside, 1970) and involves the measurement of the phase difference between a transmitted and reflected signal. Reflection of the signal is usually achieved by means of precisely manufactured glass corner cubes which return the signal along an identical path back to the instrument. Generally speaking, alignment of the reflector is not a critical problem although precise alignment of the measuring instrument is essential.

It is also possible to employ less precisely manufactured and consequently cheaper retro-reflective devices. However, random losses of energy can be great, resulting at times in a substantial loss in effective range of the instrument. Provided that this limitation is known and acceptable for the particular application involved, an opportunity exists for the development of alternatives to the currently produced glass corner cubes.

Objectives

In view of the previous remarks, a study was undertaken in which the following objectives were defined:

(*a*) the physical and optical properties of different retro-reflective materials and devices should be investigated;

(*b*) the ability of different devices to provide reliable measurements at different ranges and orientations should be assessed; and

(*c*) the possible applications to surveying practice of a selection of these retro-reflective devices should be examined.

EDM Equipment

Three instruments—a Hewlett Packard 3800B Distance Meter, an AGA Geodimeter 12A and an AGA Geodimeter 120—were available for the experiments. These instruments are well tried and are representative of modern EDM instruments currently in use. A full technical description of each instrument would be inappropriate and may be found elsewhere. The Hewlett Packard 3800 Distance

Meter has been reviewed extensively by Gort (1970), Moffitt (1970) and Methley (1974). The AGA Geodimeter 12 similarly has been reviewed by Allan (1975). It is however instructive at this stage to note certain characteristic features of each instrument which have direct relevance to the investigations.

Hewlett Packard 3800B Distance Meter

The carrier wavelength of the Hewlett Packard 3800B is 910 nanometres emitted in the standard manner by a gallium-arsenide diode. The distance is determined by consecutively recording phase difference measurements on four different frequencies.

An important feature of the instrument particularly relevant to retro-reflection is the co-axial optical configuration of the transmitting and receiving lenses. An indication of the power output and potential range of the instrument is provided by Methley (1974). The following table abstracted from this publication illustrates the maximum range obtainable with different numbers of corner cube prisms.

TABLE 1

No. of Prisms	Maximum Range (Km)
1	2·6
4	4·5
10	6·3

AGA Geodimeter 12A

Although the AGA Geodimeter 12A is similar to the Hewlett Packard 3800B in principle, it differs from it in detail. In contrast to the manual measuring procedure of the Hewlett Packard, the measuring cycle of the AGA Geodimeter takes place automatically with a digital readout on an LED display occurring after a period of approximately 15 seconds. In addition, the instrument employs an audio visual pointing system. An audible tone is heard when the target is reflecting a signal and this note assumes a high pitched tone when the pointing is correct. Both features tend to improve the speed of field operation and the accuracy of pointing. A third and important difference lies in the optical configuration. Unlike the co-axial design of the Hewlett Packard 3800B, the AGA Geodimeter 12A employs a bi-axial design, that is, the transmitting and receiving optics are offset laterally from each other. A further contrasting feature is the beam divergence of the instrument. A figure of 0·25 m/km is quoted for the AGA Geodimeter in contrast to the value of 1·0 m/km given for the Hewlett Packard 3800B. Finally, the comparative power output of both instruments is evident by comparing the following maximum ranges for the AGA Geodimeter 12A using different configurations of corner cube reflectors.

TABLE 2

No. of Prisms	Maximum Range (Km)
1	1
3	2
6	2·6

AGA Geodimeter 120

The AGA Geodimeter 120 was introduced in 1978. This particular instrument provided a new concept in electronic distance measurement, namely that of the "Semi-total Station". By incorporating the additional facilities of a calculator and an automatic reduction sensor, it provided directly in the field both horizontal distance and height difference in addition to slope distance. More significantly, by employing the optional Goedat solid state data collector, measured data could be automatically recorded in the field. The AGA Geodimeter 120 has much in common with the AGA Geodimeter 12A, for example, the same bi-axial optical configuration is employed. The beam divergence is also identical to the Geodimeter 12A. Finally the comparative power output of this instrument can be illustrated by a similar table to those given above.

TABLE 3

No. of Prisms	Maximum Range (Km)
1	1·2
3	2·3
8	3·3

One problem which arose with all of the instruments used was the difficulty of quantifying the return signal strength from various retro-reflectors. Although each of the instruments has a return signal strength meter, none are graduated in absolute units (the AGA instruments having no graduations). In addition, the scales employed give the impression that the variation in signal strength is a linear variation whereas in fact it varies in a logarithmic manner.

THE RETRO-REFLECTORS

Retro-reflecting optical systems have been known for a considerable time. Certain animals, in particular the cat, provide good examples of nature adapting the eyes of some animals to enable them to see clearly at night. The eye of the cat focuses incoming light from a narrow cone angle ahead of the animal onto the

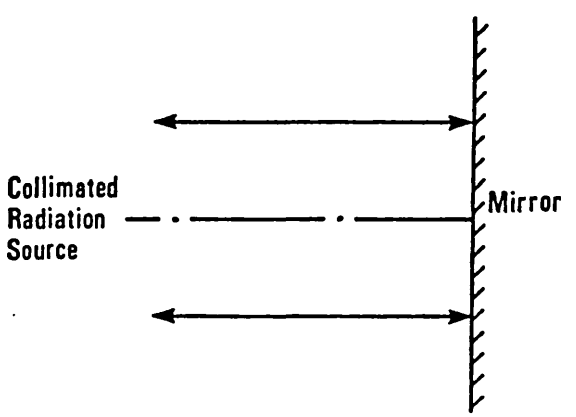


Fig. 1

retina of the eye which is provided with a highly sensitive bright smooth layer known as the tapetum lucidum. By this process, very low levels of illumination from behind an observer's head will produce an extremely intense glow of retro-reflected light from the cat's eyes when the animal's eyes are directed towards the observer. A cat's eye is thus a very efficient retro-reflector. Man has also created ingenious retro-reflective systems, several of which have been used for experimental testing purposes in this present study.

An ideal retro-reflector should turn or redirect by reflection all the incident light rays through 180°. The degree to which this is achieved indicates the efficiency of the system. One of the simplest approaches would be to employ a plane mirror as the retro reflective surface, retro-reflection occurring as in Fig. 1.

The problem arises, of course, that the mirror surface must be held in a position normal to the source of radiation. Several previous authors, notably MacVilly (1966) and Allan (1978), have discussed the use of certain observational procedures in order to circumvent this problem and hence employ the plane mirror with EDM instruments. However, although the plane mirror has the advantages of cheapness and simplicity, the alignment problems are a considerable disadvantage for field survey work. For these reasons, the plane mirror was not tested. In order to overcome the problems associated with the alignment of the reflector several other designs of retro-reflector were tested; the following are some of the characteristic features of each type.

(1) *Retro-reflective Paint and Tape*

This material, marketed by the 3M Company Ltd. as *High Intensity Scotchlite Reflective Tape* or when in paintform as *Codit Paint*, consists of a series of small glass beads embedded in a transparent or pigmented binder which holds the beads together. The geometrical arrangement for a single bead is shown in Fig. 2.

The beads are of diameter 0·25 mm–0·025 mm. In order to ensure that total internal reflection occurs, the material is altered in such a manner so as to produce

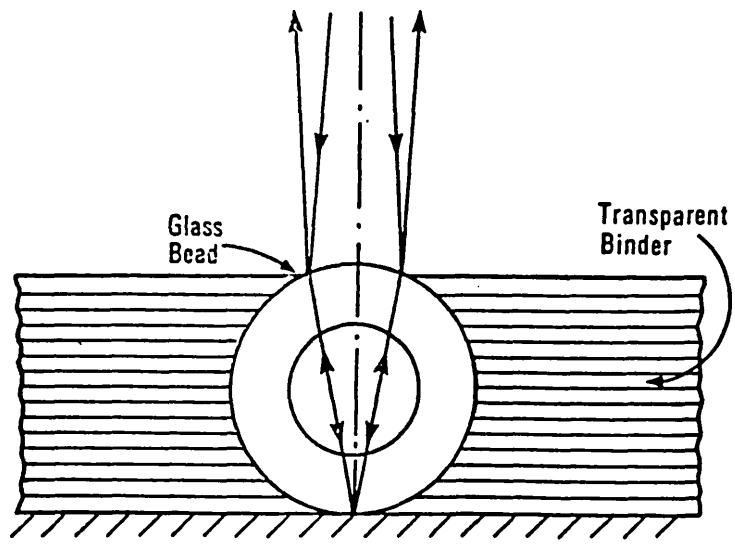


Fig. 2

a refractive index of 2·9. This is achieved by leaching the beads in dilute nitric acid to make their outer layers porous. The porous outer layer then has a lower refractive index than the unetched central core, the combined refractive index being 2.9. The combined effect is shown in Fig. 3.

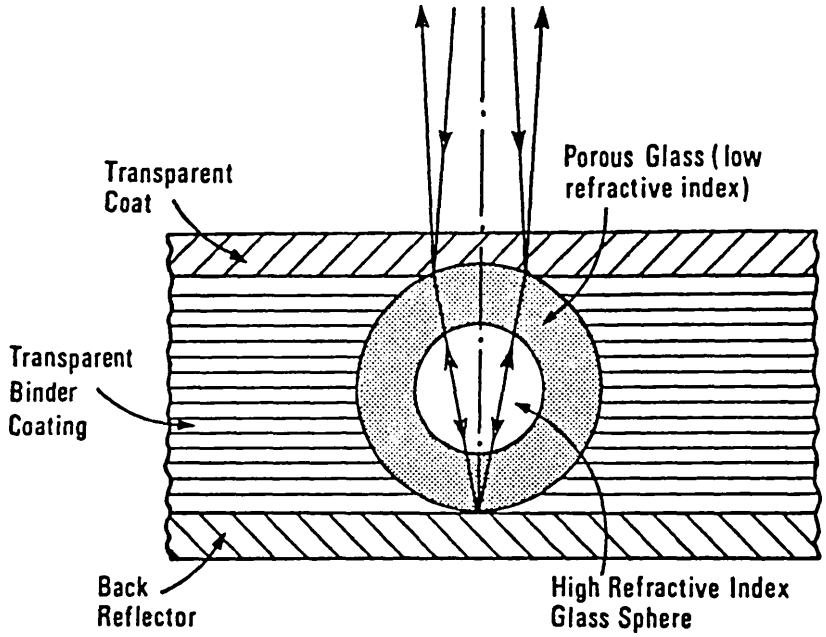


Fig. 3

The material was patented by Taylor (1955) and has been used extensively in recent years for roadside marking on motorways. When used for this application the material is a very efficient reflector of the light from car headlamps and enables motorists to see pavement markings very effectively. They are, however, optimised for this specific application. Thus they do not provide a true retro-reflection but reflect substantially in a reflex reflective manner a shown in Fig. 4.

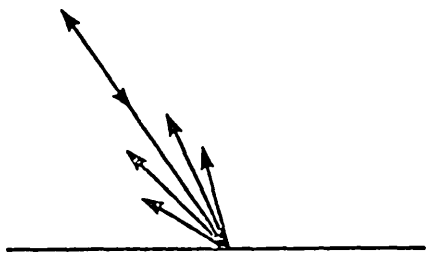


Fig. 4

Reflex reflection consists of reflection which returns the greater part of the incident signal in a direction very close to that from which the signal was emitted, irrespective of the angle the reflector makes with the source direction. In general, the reflected signal tends to diffuse into a cone shaped beam. By this process, most

of the reflected light is directed towards the driver's eyes rather than back towards the headlamp as would be the case for ideal retro-reflection. Nevertheless the material does exhibit strong retro-reflective properties and was therefore considered as suitable for use with EDM instruments.

(2) *Reflex Acrylic Reflectors*

Several designs of reflex acrylic reflectors are available on the market. The particular design employed for experimental purposes and that which was found to be the most suitable for use with EDM units was the Stimsonite range as marketed by the Signal Products Division of the Amerace Corporation. The 8·25 cm and 3·5 cm diameter centremount designs as shown in Photograph 1 were considered the most useful designs for EDM purposes. Six reflectors were employed for the tests—the red, amber, clear and green 8·25 cm convex centremount reflectors, an 8·25 cm clear flat reflector with no centremount and a small 3·5 cm clear convex centremount reflector.

A full description of the reflector characteristics can be found in Done and Methley (1980). It is however useful to briefly review some of the pertinent features of the material. The unit consists of hexagonally-shaped acrylic prisms housed in embossed aluminium. The lens portion is fused by heat on to a plastic coated metallic foil backing which acts as a permanent seal against dust, water and water vapour. The front surface is smooth, the rear surface having a multiple prismatic configuration consisting of a series of prisms of the corner cube design. The corner cubes form two complementary halves. Reflectors of this pattern are known as *split field reflectors.*

A consequence of the design is to reduce the depth of the reflector. For example if a single prism of comparable aperture were used, it would require a much greater depth, as shown in Fig. 5.

As with the retro-reflective tape, the reflection is of the reflex variety, again a consequence of the specific application for which these reflectors were designed,

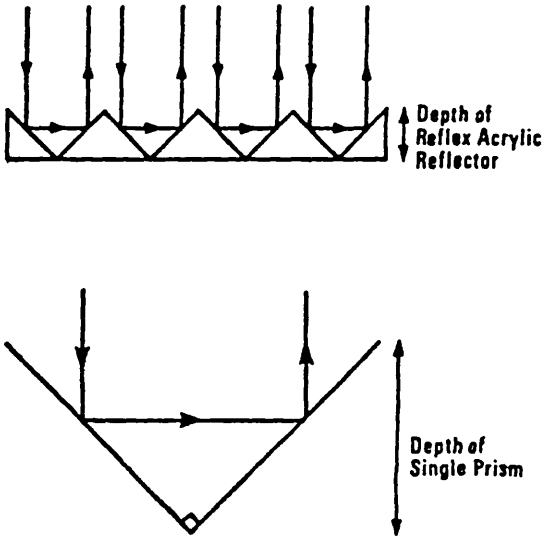


Fig. 5

that is, as roadside markings. The dispersion of the return signal into a cone shape is produced by several means. Firstly, the front surface of the reflector forms a convex spherical surface. The consequence of this is that reflection of the beam at the spherical surface forces the signal to converge to a focus from which it disperses into a conical shaped beam. Secondly, the degree of precision possible in maintaining orthogonality between the faces of the corner cubes is limited, being a function of the quality of the moulding tool and of the quality control in the moulding process. The effect of this is to sub-divide the reflected beam into six separate beams, each in itself being a parallel beam, but diverging somewhat from the other five sub-beams. Finally it appears that, as the moulded plastic cools, it shrinks, causing the faces to cave in so that, instead of the faces being flat, they become concave and hence further disperse the beam. Empirical testing has shown that the aperture of the dispersion cone is approximately 2°.

(3) *Glass Lens or "Cat's Eye" Reflector*

The third design of retro-reflector tested was that commonly referred to as the cat's eye. The unit employed for test purposes is shown in Photograph 1. As can be seen, the unit consists of 30 individual cat's eye reflecting roadstuds attached to an 8·25 cm diameter base, which in this instance is a reflex acrylic reflector.

The production of a retro-reflected signal is obtained by means of a combination of a short focus converging lens and a mirror surface conforming to its principal focal surface. The reflector is normally made of moulded glass dipped in a silvering bath to produce surface M. The geometrical arrangement can be illustrated by Fig. 6.

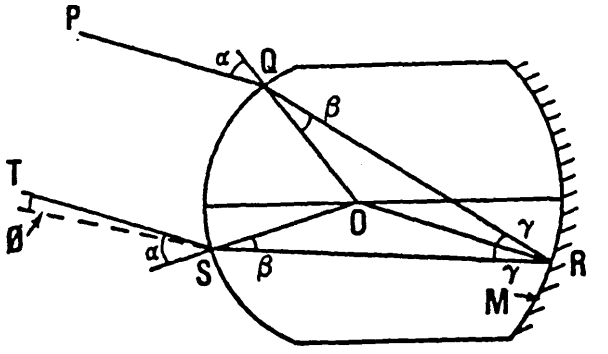


Fig. 6

A parallel incident beam, whether parallel to the lens axis or not, after passing through the external spherical surface at point Q is converged to a point on the spherical mirror surface at point R. Reflection at point R produces an emergent return beam at point S on the external spherical surface, which is parallel to the incident beam. If the lens of the cat's eye actually functioned in the manner described, the cat's eye would be ideally suited for use with EDM instruments. Unfortunately the defects inherent in simple spherical lenses of this type are sufficient to produce a spreading of the beam so that the return beam deviates from the incident beam by angle ϕ as shown in Fig. 6. However, having established

that this type of reflector possessed retro-reflective properties it was again decided to carry out some simple tests to evaluate its performance.

(4) *Glass Corner Cube Reflectors*

Conventionally the technique employed to ensure retro-reflection of an EDM signal has been the glass corner cube reflector which Grubb introduced in his British patent of 1903 (Grubb, 1903). The prism equivalent or tetrahedral prism was patented by Forbes also in 1903 (Forbes, 1903). The fundamental unit of the corner cube design is a piece of glass cut in such a manner as to produce three mutually perpendicular faces. One distinct advantage of such a design relevant to EDM measurements is the substantial range of incident angles which the reflector can accommodate whilst still ensuring retro-reflection. Two distinct designs of glass corner cube were tested.

(*a*) The single, glass corner cube design

All single glass corner cube reflectors designed for use with EDM instruments are manufactured from a Borosilicate or Boron-doped glass designated BK-7, which has a refractive index of 1·509 at 910 nanometres and 1·515 at 630 nanometres. Unfortunately, the refractive index of the glass is sufficiently high to induce significant amounts of incident energy to be lost by reflection, particularly from the front surface of the prism. A thin anti-reflection coating can be applied to this front surface by vacuum evaporation in order to eliminate, almost totally, reflection of this type at a particular wavelength. Coatings of this type must meet two criteria. Firstly the refractive index must be equal to the square root of the index of the optical material to be coated, and secondly the optical thickness must be equal to one quarter of the wavelength at which minimum reflection occurs. Several materials can be used as coatings. For example, the coating applied to Hewlett Packard reflectors is magnesium fluoride. This is used because it has resistance to attack by chemicals and moisture and is also hard enough to resist cleaning. Some reflectors also have high reflection coatings applied to the rear surfaces of the reflector in order to improve the efficiency at extreme incident angles. For example, the rear surfaces of the Hewlett Packard prisms are coated with metallic silver and silicon monoxide.

In addition to the optical characteristics of corner cube reflectors there are two distinct design features which are worthy of mention. Firstly the efficiency of a single corner cube can be directly related to the degree of precision to which the faces of the adjacent sides have been ground and the consequent expected divergence of the return beam. Most manufacturers do not make available information of this

TABLE 4. Divergence characteristics of AGA glass corner reflectors

Type of Prism	Maximum Range Obtainable (Km)	Divergence (Seconds)
Super	>10	2
Standard	10	8
Short	4	20
Zero Index	4	12

nature. Fortunately however the AGA Company has been able to provide information relating to the comparative performance of the range of reflectors which it manufactures.

The second design feature is that associated with the transmitting/receiver optical configuration. Previously the distinction was made between co-axial and bi-axial optical designs. As a consequence of this, it is necessary to employ different corner reflectors with these optical systems. The general distinction can be illustrated by the Fig. 7 (Allan, 1976).

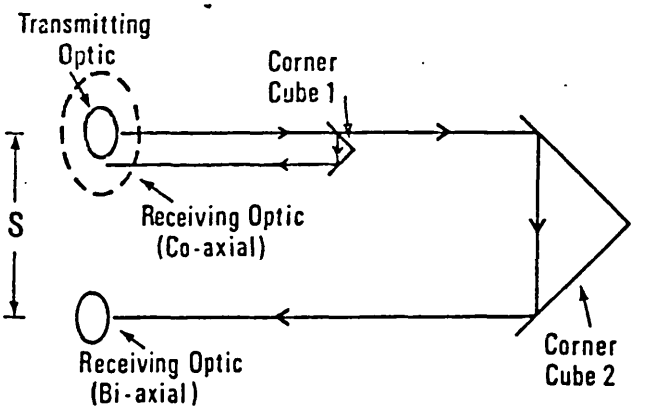


Fig. 7

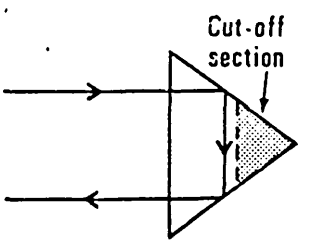


Fig. 8

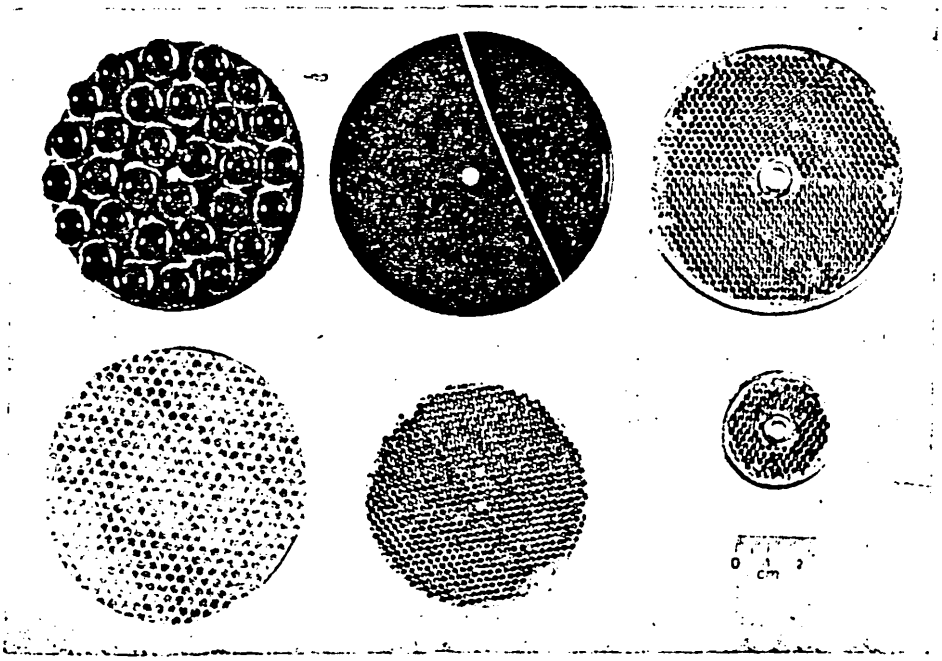
The diagram illustrates the situation which exists with a bi-axial optical system when the transmitting and receiving optics are separated by distance S. It is evident that the small corner cube 1 is unable to introduce the necessary lateral shift S required by the separation. By contrast, the larger corner cube 2 has a sufficiently large aperture and hence is able to introduce the lateral shift. In order to make the corner cube of design 2 reasonably portable, the depth of the corner cube is reduced by cutting off the unused portion as shown in Fig. 8. By this process the distinctive shape of the Wild and Kern corner cubes is produced.

Testing was restricted to the performance of three distinct corner cubes—the AGA standard prism, the AGA super prism and a less expensive American prism manufactured by Retro-ray.

(*b*) The multiple glass corner cube design

A recently developed reflector based on the single corner cube is the Swarcflex multiple glass corner cube design as shown in Photograph 1. The reflector is

manufactured by an Austrian company, D. Swarovski and Co., and is being marketed in the United Kingdom by the English Glass Company. The unit was designed for optoelectronic applications.



Photograph 1

Lens or Cat's Eye Reflector (8·25 cm)	Multiple Glass Corner Cube Design (8·25 cm)	Clear Reflex Acrylic Reflector (8·25 cm)
Retro-reflective Tape (8·25 cm)	Clear Reflex Acrylic Reflector (no centremount) (8·25 cm)	Clear Reflex Acrylic Reflector (3·5 cm)

Examination of Photograph 1 illustrates that the unit consists of 156 small individual glass corner cubes embedded in a plastic holder. A variety of reflector sizes is available. In order to provide useful comparative results, the 8·25 cm diameter unit was chosen as the most suitable size for further investigation.

In common with the single glass corner cube, the multiple cube reflector is based on the principle of retro-reflecting the signal from three mutually perpendicular faces. In contrast however, the degree of precision possible in producing the three mutually perpendicular faces is of a much lower order of accuracy than that possible with the larger single corner cube. The consequence of this is that the divergence of the return beam is much greater, forming a conical beam similar to that produced by the reflex acrylic reflector. Three grades of reflector were obtained with values of 1°, 2° and 3° for the aperture of the dispersion cone formed by reflection. All three grades were employed for testing purposes.

To be continued.

Survey Review Vol. 27, 208, April 1983

SOME TESTS OF RETRO-REFLECTIVE MATERIALS FOR ELECTRO-OPTICAL DISTANCE MEASUREMENTS

T. J. M. Kennie

Continued from No. 207 p. 12.

PART 2

Test Procedure

The test procedure adopted involved two simple experiments. The first determined the maximum range obtainable with each retro-reflector. The second established the variation in return signal strength and distance as the entrance angle was varied. Entrance angle refers to the angle between the reflector central axis and a line from the EDM unit. The procedure adopted was therefore as follows.

(*a*) Maximum Range

A flat open area near Glasgow was chosen for testing purposes. Each single reflector was mounted on the reflector unit and the signal strength measured at distances from 25 m until a maximum range was established. Increments of 25 m were established up to 200 m after which 50 m intervals were employed.

The signal strength was estimated on the return signal strength scale to the nearest five units. The scale on the Geodimeter models is much smaller than that on the Hewlett Packard. This resulted in coarser interpolation.

Maximum range was established by a trial and error procedure and was taken as the range at which reliable readings could be established. In the case of the Geodimeter instruments, this was quantified as the range at which it was possible to maintain an audible tone, this occurring at approximately 12 on the signal strength scale. With the Hewlett Packard, the reliability criterion which was adopted was the absence of gross errors, the signal strength at which this occurs being approximately 15 units on the signal strength meter. As mentioned previously, the signal strength scales on different instruments are not compatible, normally being logarithmic in nature. Hence the signal strength values only allow a useful comparison of general trends rather than absolute comparisons.

(*b*) Variation in Signal Strength and Distance with Respect to Entrance Angle

This second test was carried out to establish the variation in signal strength with respect to entrance angle. The objective of this test was therefore two-fold:—

(i) to establish the variation in signal strength and hence the maximum angular value at which it is possible to record measurements; and

(ii) to determine the variation in distance as the entrance angle increases and hence determine the maximum angular value at which it is possible to obtain reliable measurements.

The procedure involved observing a series of signal strengths at intervals of 15° up to a maximum of 60°. In addition, in order to determine the distance variation, three distance measurements were taken with the Geodimeter 12A at each of the 15° intervals.

Test Results

A total of 15 samples was tested and a summary of the maximum range and maximum entrance angles which can be accommodated, using both EDM units is shown in Table 5.

Analysis

(1) *Category* 1

(*a*) Retro-reflective Tape—The Scotchlite tape has potential for short range measurement up to a maximum distance of 140 m with the Hewlett Packard 3800B instrument. One signfiicant feature of the material is its ability to return a substantial part of the signal at extreme entrance angles.

(*b*) Retro-reflective Paint—Tests were carried out using retro-reflective "Codit" paint. No return signal was evident with any of the three EDM instruments. It is thought that a more elaborate procedure than that adopted may be necessary in order to produce a suitable target surface, such as, for example, electrostatic spraying.

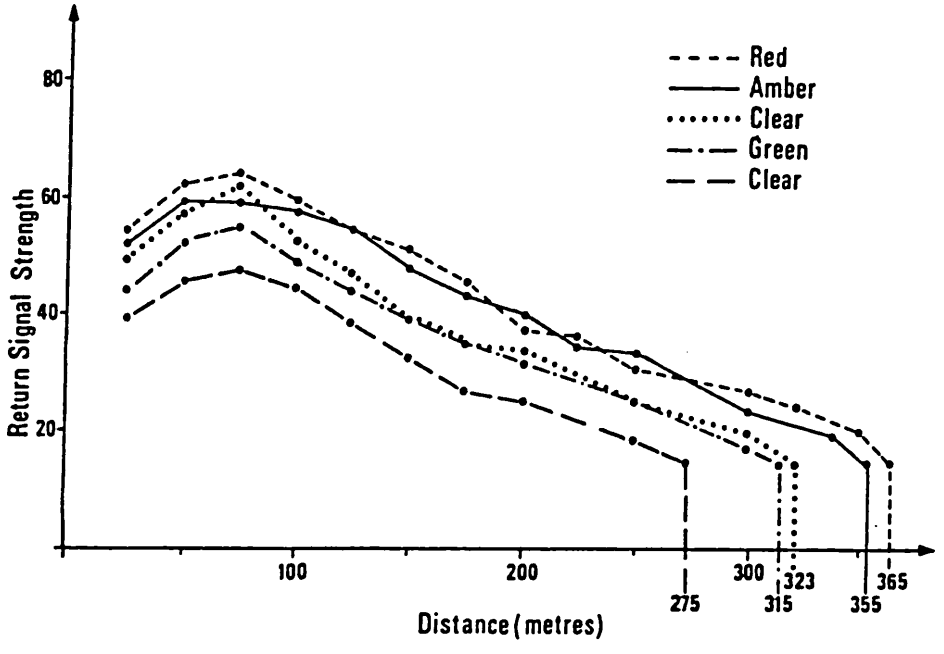


Fig. 9

(2) *Category* 2—*Reflex Acrylic Reflections*

The most significant result evident from the data is the influence of colour on reflector performance. This is illustrated graphically in Fig. 9 which shows the variation in signal strengths for various distances with the four colours of reflector using the Hewlett Packard 3800B. Results for the AGA instruments show a similar trend.

The results shown in Fig. 9 indicates that there is a consistent trend in range performance which closely parallels the visible spectrum, the red reflector giving

TABLE 5. Maximum range and entrance angles obtained with retro-reflectors using the Hewlett Packard 3800B, AGA Geodimeter 12A and AGA Geodimeter 120

Retro-reflector Category	EDM Instrument / Retro-reflector	Hewlett Packard 3800B Maximum Range (m)	Hewlett Packard 3800B Maximum Entrance Angle (°)	AGA Geodimeter 12A Maximum Range (m)	AGA Geodimeter 12A Maximum Entrance Angle (°)	AGA Geodimeter 120 Maximum Range (m)	AGA Geodimeter 120 Maximum Entrance Angle (°)
1	Retro-reflective tape	140	30	110	45	120	45
	Retro-reflective paint	NO RETURN SIGNAL				NO RETURN SIGNAL	
2	Reflex acrylic reflectors 8·25 cm centremount						
	(i) Red	365	30	310	60	310	45
	(ii) Amber	355	30	295	45	280	45
	(iii) Clear	323	30	270	45	260	45
	(iv) Green	315	30	240	45	230	45
	8·25 cm no centremount						
	(i) Clear	275	30	180	45	170	45
	3·5 cm centremount						
	(i) Clear	165	15	120	15	130	15
3	Lens cat's eye reflector	165	15	110	30	NO TEST UNDERTAKEN	
4	Glass corner cube						
	(*a*) Single Corner Cube Design						
	(i) Retro-ray prism	2650	—	2340	—	NO TEST UNDERTAKEN	
	(ii) AGA super prism	2500	—	1950	—		
	(iii) AGA standard prism	2380	—	1730	—		
	(*b*) Multiple Corner Cube Design						
	(i) 1° Dispersion	265	45	190	45	215	45
	(ii) 2° Dispersion	210	30	150	45	160	45
	(iii) 3° Dispersion	190	30	140	45	140	45

noticeably more efficient reflection than the amber and so on. This pattern has been confirmed independently by tests carried out by Lang (1980). The following extract is taken from a short report by Lang to AGA Geotronics—"four colours of Amerace reflector were tried, the red being 10% better than the orange which was 10% better than the crystal which was 10% better than the green. The tests were not carried out under controlled conditions. The instrument used was the AGA Model 120".

Following exhaustive research of the available literature and discussions with many individuals, the author is unable to provide a definite scientific explanation of this pattern. So far, the evidence suggests that the red plastic reflector absorbs the infra-red signal less than the amber and so on.

A second unexpected result is evident when one compares the data for the 8·25 cm diameter clear centremount and clear non-centremount reflectors. Contrary to initial expectations, the centremount reflector is approximately 15% more efficient. Apart from the centremount, which reduces the surface area slightly, the only other distinguishing feature is the slightly convex cross-sectional shape of the centremount compared to the non-centremount. It would appear that the convex shape enhances the return signal by focusing the signal slightly.

Distance variation as a function of entrance angle was carried out using the red, amber, clear and green reflectors at distances of 25 m, 75 m, 125 m and 175 m. The results for the red reflector with the AGA Geodimeter 12A are shown in Fig. 10. Results using the other reflectors illustrate a similar trend.

Two significant conclusions may be drawn from Fig. 10. Firstly apart from short distances (<25 m) at entrance angles greater than 15°, the variation in distance increases substantially and gives erroneous results. Secondly, although a systematic distance variation exists at the 15° entrance angle position, the maximum variation is of the order of +5 mm. It is evident therefore that, in cases where high

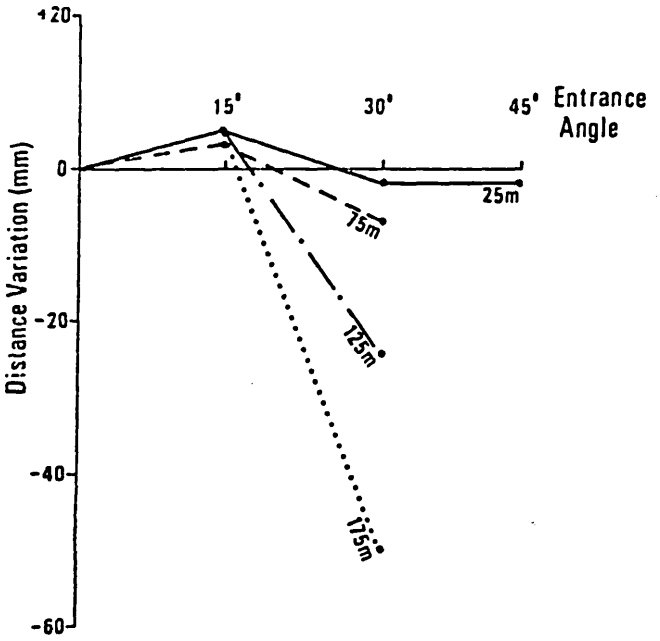


Fig. 10

accuracy is required, alignment of the reflector by eye to within 10° is necessary. However, in applications where the accuracy specifications are lower, entrance angles of up to 15° may be accommodated without incurring significant error.

Finally it should be noted that Reflex Acrylic Reflectors may be used in conjunction with most short range EDM instruments employing amplitude modulation. However, with instruments such as the Kern Mekometer ME 3000, the acrylic reflectors tend to degrade the polarisation modulation associated with this type of instrument and therefore cannot be used (Bradsell, 1980).

(3) *Category 3—Lens Cats Eye Reflector*

Two significant conclusions are evident from the results. Firstly, the cat's eye reflector is significantly less efficient than the reflex acrylic reflectors. Secondly, the reduction in range and signal strength as one alters the entrance angle is quite substantial. For example, with the Hewlett Packard 3800B, the reflector will not reflect at entrance angles greater than 15° and at this value the reduction in range is about 50%. Undoubtedly the beam divergence of the return signal caused by the imperfections in the glass are strongly associated with the results obtained.

(4) *Category 4—Glass Corner Cube*

(*a*) Single Corner Cube Design

The full results for this test are shown in Fig. 11 which illustrates the results obtained using the Hewlett Packard 3800B. Results obtained with the AGA Geodimeter 12A show a similar trend.

Again two significant conclusions are evident. Firstly the results obtained with each prism far exceed that expected from the manufacturers' literature. Secondly, as would be expected, the AGA Super prism performs better than the AGA Standard prism, the more precisely ground faces minimizing the divergence of the return signal. However, the cheaper Retro-ray prism outperformed both AGA

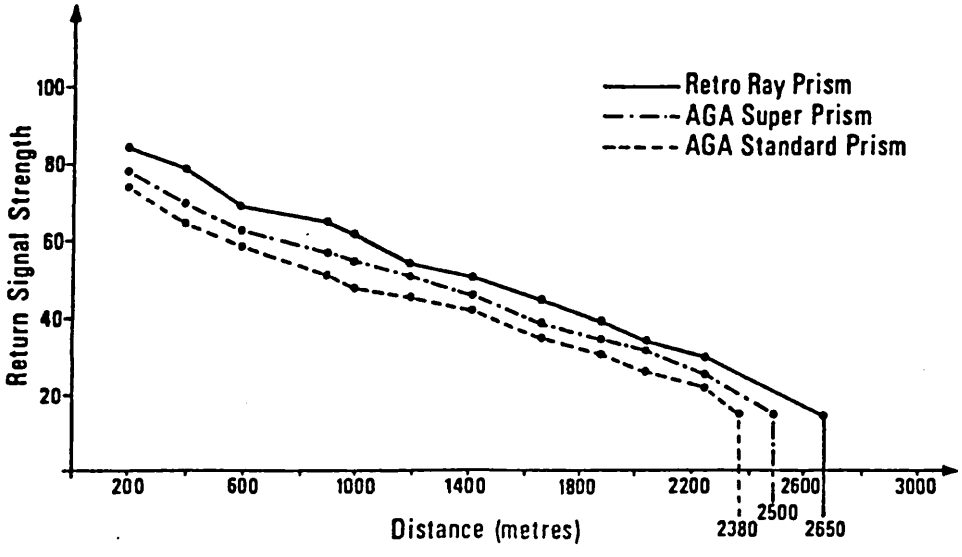


Fig. 11

prisms. Unfortunately, technical specifications have not been supplied (as yet) for the Retro-ray prism which may help to explain the results. However, two possible explanations are—that either the angles between the faces of the prism have been ground to a higher degree of precision than the AGA Super prism (i.e. <2″ of arc), or alternatively the Retro-ray prism having an aperture of 7·0 cm compared with the 6·3 cm for the AGA prisms is providing a greater surface area from which reflection may take place.

(*b*) Multiple Corner Cube Design

As with the previous reflectors, the more powerful Hewlett Packard instrument produces ranges approximately 25% greater than those obtained using the AGA instrument. Not unexpectedly, a pattern emerges in the results which parallels the dispersion characteristics of the reflectors. Therefore the narrower the cone of dispersion the greater the return signal strength and hence the range.

However, contrary to initial expectations, the Multiple Corner Cube Design performed much more poorly than the Reflex Acrylic Reflectors. Two factors may possibly help to explain this result. Firstly the Multiple Corner Cube Design is mass-produced and the degree of orthogonality between the faces of the individual corner cubes is not high. Therefore, although the cone of dispersion is limited to 1° at best, this is not better than that which exists with the Reflex Acrylic Reflectors. Secondly there exists a substantial reduction in reflective surface area between the two designs, the Multiple Corner Cube Design having less than 40% of the reflective area associated with the Reflex Acrylic Reflector design.

Applications

There is no doubt that for many survey projects involving short-range distance measurements, the use of a robust, inexpensive alternative to the conventional corner cube prism could be extremely advantageous. As has been shown, both the reflex acrylic reflector and retro-reflective tape have potential for short range EDM measurement. Although few results have been published, several organizations have introduced these types of reflector. The following is a review of specific survey tasks on which these alternative reflector systems have been employed.

(1) *Dimensional Control*

During the construction of many large engineering structures it is imperative that accurate measurement of the "as-built" shape of the structure is available to the engineer or designer. This information is invaluable in allowing prefabricated units to be constructed in such a manner that the fit between units is exact, rather than depending on the original design drawings. This aspect of engineering survey work is commonly referred to as dimensional control. This area is ideally suited to the use of the alternative reflectors since, firstly, the distances to be measured are short and, secondly, distances often have to be measured to inaccessible points on the structure to which it may be difficult, if not impossible, to attach conventional glass corner cubes. One particularly good example of the use of these alternative reflectors in dimensional control is in the area of *shipbuilding*. The British Ship Research Association (BSRA) carried out tests with both alternative reflectors, specificially for use with their BRITCHECK system [Hacking (1978),

TABLE 6. BSRA reflector test results

Reflector	Maximum Range (m)		Max. Entrance Angle
	Tellurometer MA-100	AGA Geodimeter 710	
1 cm Scotchlite	No test Undertaken	100	40°
8·25 cm Stimsonite	90	No test Undertaken	20°
1·25 cm Stimsonite	No test Undertaken	250	20°

Lamb (1980)]. This system employs an AGA Geodimeter 710 laser-based total station for the processing of ship structure measurements. Experiments were also undertaken using a Tellurometer MA-100 EDM unit. Three distinct retro-reflective targets were tested:—

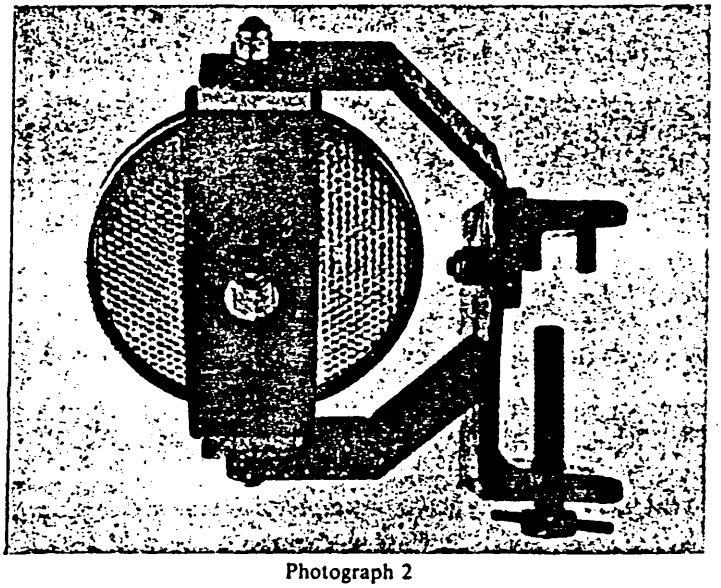
(*a*) 1 cm diameter High Intensity Grade Scotchlite retro-reflective tape;

(*b*) 8·25 cm diameter Stimsonite reflex acrylic reflectors; and

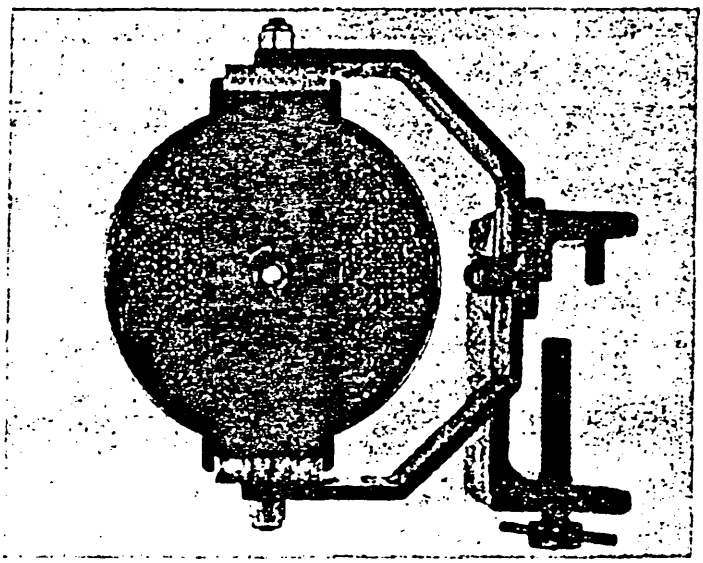
(*c*) 1·25 cm diameter Stimsonite reflex acrylic reflectors.

It is interesting to note the substantial increase in range achieved with the laser carrier signal.

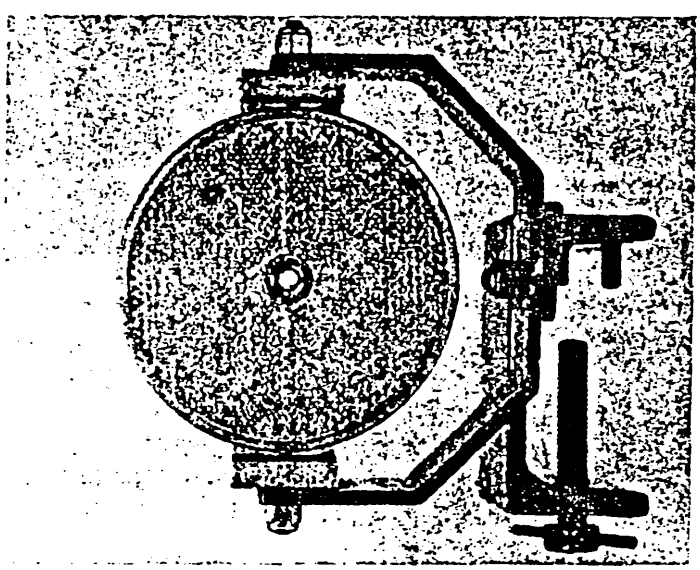
Having established the feasibility of using the alternative reflectors, a gimbal mounted unit was designed, as shown in Photograph 2. The process of aligning the reflector required an operator who rotated the target until a position was reached as shown in Photograph 3. Finally, by removing the central flexible metal alignment aid, a measurement was undertaken. Photograph 4 illustrates this position.



Photograph 2



Photograph 3



Photograph 4

(2) *Open-Pit Mine Operations*

A second paper in which the investigation of alternative reflectors is reported is that by MacRae (1976). He investigated procedures which would allow *slope failures* in open-pit mine operations to be predicted.

The conventional surveying methods employed for this type of work consists of observing to permanently mounted targets using theodolites and levels, any change in co-ordinate value or height being indicative of ground movement. For a

variety of reasons such as equipment cost, accuracy and safety, the author decided to investigate the possibility of using an EDM instrument for measurement.

The instrument chosen was the Keuffel and Esser Ranger III employing a laser carrier signal, $\lambda = 560$ nanometres. The maximum range of the instrument is quoted as 12 km using several standard 63 mm glass reflectors. MacRae investigated the range attainable with the Stimsonite reflex acrylic reflector and reported that "it performed with accuracy equal to the standard prism but at greatly reduced range". A maximum range of 370 m was achieved with one clean reflector, and 180 m with a dust covered reflector. Unfortunately no size of reflector is quoted. However, the photograph of the reflector indicates an approximate size of 1" (2·5 cm) in diameter.

MacRae also designed a high quality 25 mm diameter glass cube reflector costing between $60 and $70, approximately one-quarter the price of conventional units. Results using laser and infra-red instruments were reported to be favourable with only a slight reduction in range as compared with the conventional prism. No results are given in the paper. The author also reported tests with a 63 mm diameter "hollow" reflector costing $135 (1976 prices). Initial results indicated that this unit performed better than the conventional 63 mm unit. Again no specific results are reported.

(3) *Urban Surveys*

Much attention has been focused in recent years on the consequences of neglecting survey activities in a highly complex urban environment. In order to redress the balance, some countries have instituted a program of control densification in urban areas. The reflex acrylic reflector is well suited to this type of application and has been used successfully in Canada as *a wall marker for EDM measurement* (Chrzanowski and Steeves, 1977).

The idea of wall monumentation was developed a long time ago and has been used in many European countries. Different approaches have been developed and can be classified into two groups. In the first group, the wall markers serve only as reference points for re-constructing the original location of traverse stations. In the second group, the wall markers serve directly as co-ordinated control points. Using this approach traverses are run with forced centring without permanently marking the traverse stations on the ground. Instead, three or more small markers are monumented and targeted on the buildings in the vicinity of each traverse station. The wall points are established as the traverse progresses by measuring distances and directions to each wall marker.

The design of a reflector target suitable for use as a wall marker is shown by Fig. 12. The target was used for test purposes in Fredricton, Canada (Chrzanowski and Steeves, 1977).

The target may be rotated about its vertical axis and is equipped with a small reflex acrylic reflector allowing distance measurements to be recorded. Experience of a test network indicates that four markers (one traverse station) could be monumented in less than half an hour.

The only apparent problem with the target is that the co-ordinated point is not the hole of the marker but the centre of the cross on the removable target. In effect, all users of the control points must have the same type of targets or must remember that the control point is 45 mm in front of the marker.

(4) *Crane Positioning*

A novel and efficient solution to a problem of *crane positioning* has been developed by Sir Robert McAlpine and Sons Ltd., civil engineering contractors (Lang, 1980). In order to ensure that cutting and dumping of material is performed efficiently, positional data is determined by the crane operator using an AGA Geodimeter 120 mounted on the crane. Using a gyro compass and distances measured to an approx. 8" ball covered with retro-reflective tape suspended from the jib, the actual position of the jib can be determined. Using this method, a more efficient control system is available. In order that the data determined from the crane can be quantified in absolute values it is necessary to determine the position of the cab of the crane. This may be undertaken by a number of methods depending on the circumstances, e.g. theodolite intersection or a combination of angular and distance measurements.

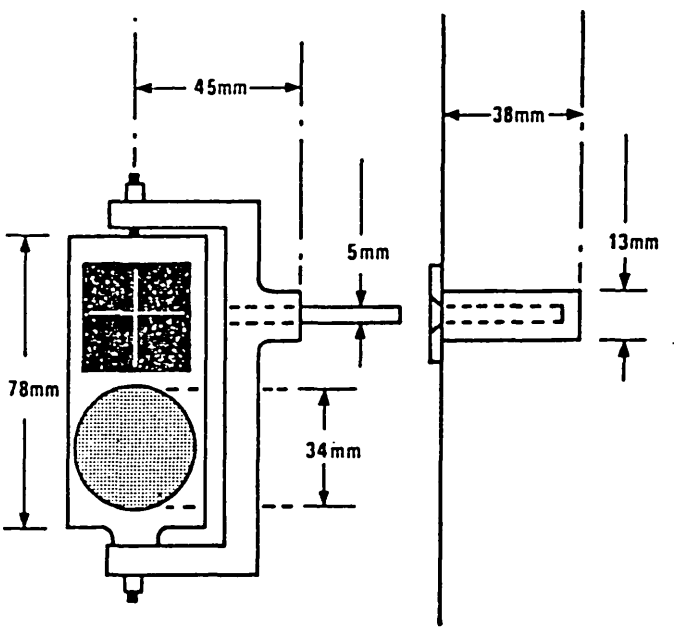


Fig. 12

CONCLUSIONS

Several types of retro-reflective material and devices were tested, including the conventional glass corner cube. The results of the test indicate that, for short range applications, several of the alternative systems are potentially very useful, particularly when used in conjunction with an EDM instrument with co-axial optics. Of particular importance are the reflex acrylic reflectors, specifically the red coloured reflector, and also the retro-reflective tape.

As has been shown, for specific areas of application where the distances to be measured are short or where multiple measurements are required to pick-up points of detail, the conventional glass corner cube may be an expensive and unnecessary means of carrying out the measurement.

In conclusion therefore the alternative systems should not be considered as replacements for the conventional glass corner cubes, but rather as alternatives to the glass corner cube for specific types of measurement problems.

ACKNOWLEDGEMENTS

The author is particularly grateful to the following companies who provided technical information on their products: Amerace Ltd.; AGA Geotronics; English Glass Company; Reflecting Roadstuds Ltd.; Swarovski and Co. Ltd.; and the 3M Company. I am also grateful to Mr. C. B. Lamb of the BSRA for permission to publish his organization's results, Mr. S. J. Lang for permission to publish details of his investigation and to the Department of Geography, University of Glasgow, for their help in producing the line drawings included in this paper. Finally, my thanks are due to Mr. B. D. F. Methley and Professor G. Petrie of the University of Glasgow for their constant encouragement and useful comments in the writing of this paper.

References

Allan, A. L., 1975. The AGA Geodimeter 12. *Survey Review* XXIII (178), 191–193.

Allan, A. L., 1976. Electro-optic Distance Measurement. *Ph.D. Thesis*, University of London.

Allan, A. L., 1978. The Plane Mirror in Electro-optical Distance Measurements. *Survey Review* XXIV (188), 268–274.

Bradsell, R. H., 1980. Private correspondence.

Burnside, C. D., 1970. Electromagnetic Distance Measurement. Crosby Lockwood Staples. London.

Chrzanowski, A. and Steeves, P., 1977. Control Networks with Wall Monumentation. A Basis for Integrated Surveys in Urban Areas. *Canadian Surveyor*, Vol. 31, No. 3, 211–222.

Done, P. and Methley, B. D. F., 1980. The Use of Acrylic Reflectors with Electro-optic Distance Measuring Instruments. *Survey Review* XXV (195), 215–222.

Forbes, G., 1903. Improvements in Signalling in Apparatus Thereof. British Patent Number 9129.

Gort, A. F., 1970. Electronic Distance Measurements with the Hewlett Packard Model 3800. *Technical Publication of the Hewlett Packard Company.*

Grubb, H., 1903. Improvements in and Means of Signalling and Indicating Position of Objects. British Patent Number, 21865.

Hacking, H., 1978. A Survey System for Marine Structures with Automatic Data Processing and Printout. *Proceedings of the 2nd International Symposium on Deformation Measurements by Geodetic Methods*, Bonn, East Germany.

Lamb, C. B., 1980. Private correspondence.

Lang, S. J., 1980. Correspondence with AGA Geotronics.

MacRae, A. M. R., 1976. Pit Slope Displacement Monitoring as an Aid to Open Pit Mine Operation. *Engineering and Mining Journal*, Sept. 1976, 105–107.

MacVilly, R. B., 1966. An Adaptation of the Plane Mirror for Use in Geodimeter Work. *Survey Review* XVIII (142), 359–361.

Methley, B. D. F., 1974. The Hewlett Packard 3800B Distance Meter. *Survey Review* XXII (174), 337–346.

Moffitt, F. H., 1970. Field Evaluation of the Hewlett Packard Model 3800 Distance Meter. *Proceedings ASP-ACSM Conference*, Denver, Colorado.
Meter. Proceedings ASP-ACSM Conference, Denver, Colorado.

Taylor, N. W., 1955. Reflex Light Reflectors and Glass Beads thereof. U.S. Patent No. 2713286.

Survey Review, 30, 232 (April 1989)

A PRELIMINARY ASSESSMENT OF THE TELLUROMETER MA200 FOR VIBRATIONAL MEASUREMENT OF FLEXIBLE STRUCTURES

T. J. M. Kennie, C. M. Howard and S. P. Horkan

Department of Civil Engineering, University of Surrey

Abstract

The use of static or dynamic tests to determine the vibrational characteristics of civil engineering structures is a well established procedure. Traditionally this process has been carried out by invasive techniques, normally involving the use of accelerometers. More recently a non-invasive technique based on a laser based Electromagnetic Distance Measuring (EDM) instrument has been developed. The paper reviews these different techniques and, in particular, discusses the potential of the Tellurometer MA200 EDM instrument for the measurement of structural vibration. Descriptions of some simple experiments to assess the instrument's capabilities for measuring such vibrations are presented. From the results of these preliminary investigations it is concluded that the Tellurometer MA200 has some potential for such operations although further work is required in order to fully assess the instrument's capabilities.

Introduction

The effects of structural vibration and the possible damage caused by such vibrations is a subject which has been of concern to civil engineers for many years. The source of these vibrations can vary from those induced by natural causes such as wind or seismic activity to man-made vibrations caused by traffic, industrial machinery, pile driving or blasting.

The need to assess the effect of such vibrations on structures has become more significant in the past two decades as structures such as tall buildings, radio masts and so on have become increasingly more 'flexible'. Flexible structures have developed for a number of reasons including: (a) improved analytical techniques which provide a better understanding of the behaviour of structures, (b) the use of lighter and stronger materials, (c) changes in the design procedures used, including the elimination of partitions and non-structural components, and (d) an increasing need for more cost effective designs.

The vibrational behaviour of a structure is a function of its stiffness and mass distribution. In general, the harmonic frequency of a structure is proportional to the stiffness and inversely proportional to the mass. One of the most significant effects of these design changes has been to decrease both the stiffness and mass of structures. In certain circumstances this may have also increased the possibility of structural problems arising from excessive lateral vibration. Furthermore, as a concrete structure cracks, or a steel structure corrodes, its stiffness will change. The stiffness will also alter if the support conditions change. Generally these variations in stiffness cannot be detected by visual inspection although it is suggested by Ward [12] that they may be detected by vibration analysis.

A recent report by the SERC and the DOE [11] also emphasises the need for information on the behaviour of structures in service. This report highlights the importance of dynamic tests in increasing our knowledge of how structures

Table 1. *Some examples of the vibrational characteristics of civil engineering structures*

Structure	Fundamental Harmonic Frequency (Hz)	Maximum Amplitude of Vibration Recorded (mm)
Severn Suspension Bridge, UK	0·14	—
Long Creek Bridge, Canada	0·60	200
Tinsley Viaduct, UK [6]	1·4–1·8	20
Empire State Building (USA)	0·12	—
Reinforced Concrete 11 Storey Building, UK [13]	1·3–1·95	—
190 m Radio Tower, UK [4]	0·15	195
Eiffel Tower, France	—	125

perform during construction and in service. Vibrational analysis may also help to assess the integrity of a structure which may need to be checked, for example, if a structure is reaching its design life or after earthquakes and fires.

Vibrations may also need to be monitored by contractors piling or blasting in quarries adjacent to areas of development to ensure that acceptable levels of vibration are not exceeded.

In general, the fundamental harmonic frequencies of many civil engineering structures lie within a frequency range of 0·1 Hz to 10 Hz. Although, in practice, the frequencies of forced vibrations are at the lower end of this range. Typically earthquake vibrations are at around 1 Hz and oscillations caused by the wind are often within a frequency range of 1 Hz to 2 Hz. Table 1 gives some examples of the vibrational characteristics recorded for different types of structure.

Review of Current Techniques for Measuring Structural Vibrations

Four quantities can be derived from a vibrational analysis: (a) the fundamental and harmonic frequencies of the structure; (b) the amplitude of the vibration; (c) the modal shape of the vibration and; (d) the degree of damping associated with the vibration.

The methods used for measuring these parameters can be categorised as either invasive techniques, that is those which involve direct contact with the structure, or non-invasive techniques, or non-contact methods which measure vibrations by remote measurement.

Invasive Techniques

The most common instrument used to measure vibrations is the accelerometer. An accelerometer is a transducer which provides an output signal proportional to acceleration. The accelerometer most frequently used to measure vibrations is based on a seismic mass principle and is called the piezo-electric accelerometer. The accelerometer can respond to a wide range of frequencies (4 Hz to 1500 Hz), is light (0·5 g) and compact (diameter = 5–6 mm). These properties are important and ensure that the accelerometer has a minimal effect on the properties of the vibrating system.

Using accelerometers to measure structural vibrations is also a tried and tested

technique and the results have generally been found to be reliable. Accelerometers can be placed at several places on a structure and so modes of vibration can be determined. Also since the accelerometer is an invasive technique, it is not susceptible to effects outside the structure.

Many of the disadvantages of using accelerometers are associated with the wiring and maintenance. Wiring up an existing structure is often difficult and may be impossible in some cases e.g. offshore structures or tall masts. Furthermore, the wiring may be easily damaged or alternatively because of the length of wiring, excess 'noise' may be set up. Often accelerometers are placed inside the structure to provide more reliable results. In these circumstances however problems with servicing and damage to the accelerometer may occur.

Non-Invasive Techniques

1. *Laser Interferometry*

Ashkenazi *et al.* [2] discussed the application of a laser interferometer for the measurement of continuous dynamic movement. The interferometer used was a Hewlett–Packard 5526A 700 μW, Helium–Neon (He–Ne) gas laser emitting coherent radiation in the visible range of the spectrum with an average wavelength of 632·8 nm.

The performance of the interferometer for vibration analysis was assessed by a laboratory rig which simulated frequencies in the range from 0·15 Hz to 0·9 Hz with amplitudes ranging from 2 to 30 mm. A conventional glass retro-reflector was attached to the test rig to enable remote measurement using the laser interferometer. Data acquisition was carried out using a paper tape data logger and the filtered data were processed using Fourier Transform analysis. The results of this project indicated that frequencies in the range 0·15 to 0·9 Hz, with amplitudes of up to 25 mm could be determined using such a system. The primary limiting factors to the more widespread use of the technique were identified as the slow paper tape data acquisition system and that the maximum operating range of the instrument was 50–60 metres.

2. *Laser Vibrometer*

An alternative instrument used to measure vibrations based on the measurement of the Doppler frequency shift is the Laser Vibrometer developed at the Institution of Sound and Vibration, University of Southampton (Halliwell [7] and Pickering *et al.* [10]). The instrument is portable, relatively inexpensive and has been used in fluid mechanics and motor and sound engineering, to measure vibrations of high frequencies (200–1000 Hz).

However, the laser source employed in the instrument has an output of less than 5 μW and consequently the operating range of the instrument is limited to 1 m approximately. It is therefore currently inappropriate for use in the determination of the vibrations normally associated with civil engineering structures.

3. *Cine-Photogrammetry*

The potential of high speed photography for monitoring oscillations of tall antennae was recently examined by Ethrog & Horamy [5]. In this case a non-metric 16 mm cine camera with an exposure interval of $\frac{1}{60}$ s was examined. The authors carried out manual photogrammetric measurements of single frames of photography. Because of the considerable time required to measure the photography the authors considered the technique to be impractical in its present form. However, some suggestions for future work to automate the point search and measurement procedures were provided.

THE TELLUROMETER MA200

As discussed in Allan [1] the Tellurometer MA200 electronically determines distance in a conventional manner by comparing the phase of the transmitted and reflected signals. In absolute distance measurement the reflected signal is compared with an internally reflected signal rather than by direct comparison with the source. Thus the distance measured represents the distance between a fixed point within the instrument and the reflecting prism. By this process measurements are made with reference to the fixed and known internal paths.

The internal signal is established by switching the emitted laser radiation into one of two paths using a mirror. One path is directed to the receiver photo-diode to form the internal path while the other is directed towards the prism. Depending on the position of the internal/external shutter, the light incident on the receiver photodiode will either be that from the internal or external path.

For vibration analysis, the change in absolute distance measurement is not of prime importance, but rather the difference between consecutive phase measurements (since the difference between phase measurements will enable the amplitude of the vibration to be determined). In order to obtain as many samples of phase as possible from a vibration it is possible to neglect the internal path, i.e. maintain the shutter in such a position as to allow an external path for the emitted radiation only.

The MA200 is normally operated from a hand held controller for normal distance measurement; however, because it was necessary to process raw phase data and the shutter had to be set to the external position, the MA200 was controlled by, and interfaced to, an IBM PC/AT. Through a BASIC program it was therefore possible to 'point' the instrument (i.e. obtain a maximum value for signal strength of the returning signal), set the shutter to external and capture/retrieve the raw phase data. Currently the maximum number of samples which can be captured is 256 although this may be increased in the future.

Clearly, for a given frequency of vibration, the number of samples and the rate of sampling is of critical importance. This requirement is described by the 'Nyquist Sampling Theorem' which states that information with a band width of bHz must be sampled at a minimum of 2 bHz if the information is to be retrieved. Since the MA200 has a sampling rate of 58 Hz it is considered possible, in theory, to measure vibrations up to a maximum frequency of 29 Hz.

T. J. M. KENNIE, C. M. HOWARD AND S. P. HORKAN

Evaluation of the Tellurometer MA200 for Vibrational Measurement

The evaluation of the Tellurometer MA200 took the form of a number of simple experiments carried out both in the field and in the laboratory, at the University of Surrey. The experimental work had two main objectives:

(*a*) To investigate the degree to which the MA200 could measure the frequency and amplitude of:
  (i) a simple harmonic vibration using a single degree of freedom model;
  (ii) a more general periodic vibration using a multi degree of freedom model.
(*b*) To consider the potential use of low cost retro-reflectors for vibrational analysis.

The oscillatory vibration of a structure is a function of: the distribution of mass and stiffness; the connectivity between the component parts; the support conditions; and the type and frequency of the applied forces or displacements. Practical structural forms represent complex combinations of these variables and, as a result, represent complex multi-degree of freedom systems. The dynamic displacement of such structures is often a complex waveform composed of a number of different frequencies with varying amplitudes.

Clearly, in order to interpret vibrational data it is necessary to separate out the signal into its component frequencies. In this way the harmonic, or resonant, frequencies of the structure can be identified. These would be apparent by their high amplitude. With a single degree of freedom system only one harmonic frequency exists and is normally evident from an examination of the displacement- or acceleration-time data. However, data post-processing is required to characterise the harmonic frequencies for a multi-degree of freedom system.

To evaluate the capacity of the MA200 for the measurement of vibrational displacement, of the type described, two experimental models have been used: a single degree of freedom simple pendulum; and a multi-degree of freedom double pendulum.

Model 1: Single Pendulum

In order to simulate simple harmonic motion, in other words a vibrating system with only one degree of freedom resulting in a single frequency and mode of displacement, a simple pendulum model was used. The model was constructed using a conventional retro-reflector as the suspended mass supported on a platform by two thin steel cables anchored to a reaction frame, as is shown in Fig. 1. The very simple sinusoidal displacement of the model allowed an initial evaluation of the measuring capacity of the instrument without the need for any post-processing of the acquired data.

Model 2: Double Pendulum

A more complex vibrational motion was achieved using a three degree of freedom model. The double pendulum used consisted of two glass retro-reflectors suspended from the reaction frame in series as shown in Fig. 2(*a*). The motion of the higher of the two masses was monitored and therefore three independent modes and associated frequencies of vibration could be detected, as is shown in Fig. 2(*b*).

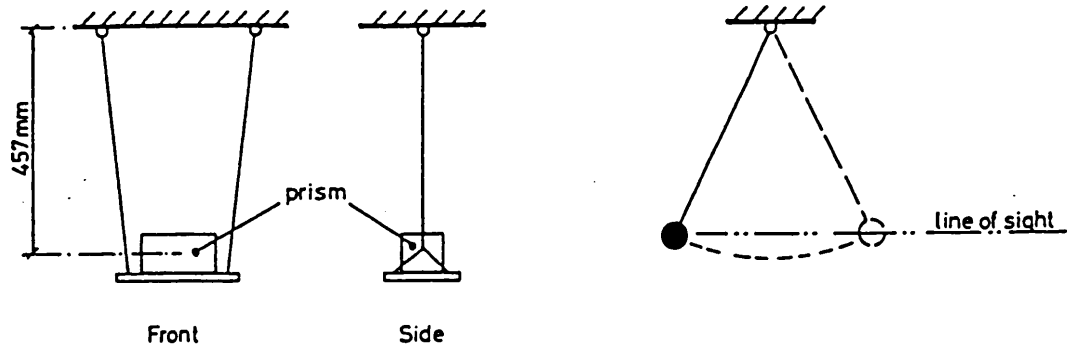


Fig. 1. Single pendulum test arrangement, Model 1.

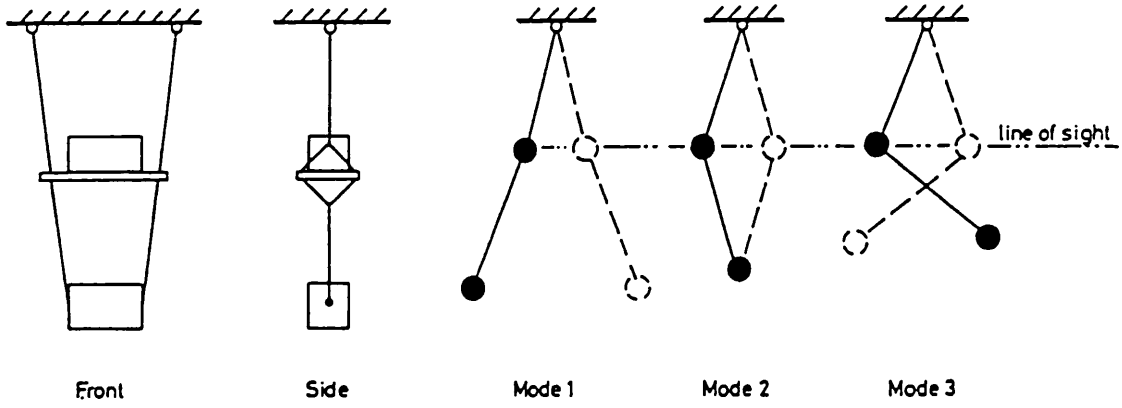


Fig. 2.(*a*): Double pendulum test arrangement, Model 2, and (*b*) modes of oscillation.

This multi-degree of freedom model was used in order to examine the additional requirements on the measuring parameters to enable full characterisation of the vibrational behaviour of real structures.

Data Acquisition

An independent measuring system using conventional accelerometers was used in tandem with measurement using the MA200. This enabled correlation checks to be made with respect to an alternative measuring system.

MA200

For the purpose of these initial tests the MA200 was set-up on a conventional theodolite tripod at a distance of approximately 7 m in line with the pendulum models. The test procedure used was identical for each of the two models. The pendulum was set in motion by giving the mass an initial displacement of approximately 50 mm. The phase data was then sampled (256 samples were taken at a rate of 58 Hz yielding a total period of measurement of 4·41 s). The phase data collected was then dumped from the instrument to an IBM-PC/AT and subsequently processed to give displacement data.

Accelerometer

The accelerometer was attached to the back of the glass corner cube reflector and acceleration-time data was recorded simultaneously with the remote measurement. The data was monitored and recorded using a Datalab DL1200 datalogging waveform recorder. The data was sampled at a rate of 100 Hz over a period of 40 s. In this way 4000 data points were obtained for each of the two tests. The data was then dumped to a Hewlett Packard 200 series PC for post-processing.

Data Analysis and Post-Processing

In order to interpret the vibrational characteristics of the two models tested the frequency of the observed oscillation was required. For the single pendulum this was achieved by examination of the displacement-time data obtained. For the double pendulum it was necessary to use Fourier analysis in order to separate out the three frequencies which contributed to the observed oscillation.

Fourier Analysis

Fourier showed that any periodic motion can be represented by a series of harmonically related sines and cosines. It therefore follows that given a set of discrete data of finite length a discrete Fourier series of sine and cosine terms can be found such that the series will pass through all the data points. This process is known as a Discrete Fourier Transform (DFT) and is detailed in Appendix 1. In order to reduce the computational time necessary to achieve the transform an adapted type of algorithm known as the Fast Fourier Transform (FFT) has been developed (Bendat and Piersol [3]).

A FFT was used to analyse the discrete data samples obtained from the two test procedures. In the case of the data obtained using the accelerometer and waveform recorder post-processing was achieved using the FFT routine within the Acquire software operating on the Hewlett Packard 200 series PC. The data collected using the MA200 was analysed using a FFT routine within the NAG-library [9], mounted on a Prime mini-computer.

Results and Discussions

The displacement-time data collected using the MA200 on the single pendulum model are shown in Fig. 3. For simplicity only one complete period of the oscillation measured is shown. The motion of the pendulum can be seen to be sinusoidal in character as expected. The frequency of the oscillation was determined by interpolation of the data and was found to be 0·74 Hz.

In order to independently assess the frequency measurement obtained, the data collected using the accelerometer was processed using a FFT and the harmonic frequency was found to be 0·76 Hz. The resultant data plotted in the frequency domain is shown in Fig. 4. It is noted that this data shows a single harmonic frequency which indicates simple harmonic or sinusoidal motion. As is discussed in detail later in reference to the model 2 test results, the difference in the value of the harmonic frequency obtained by the two measurement techniques is thought

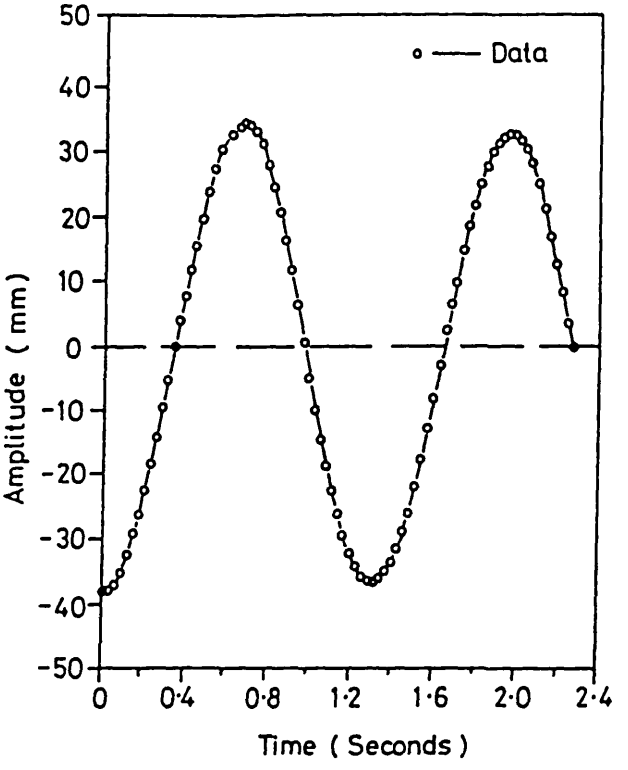


Fig. 3. Displacement-time data acquired using MA200, Model 1.

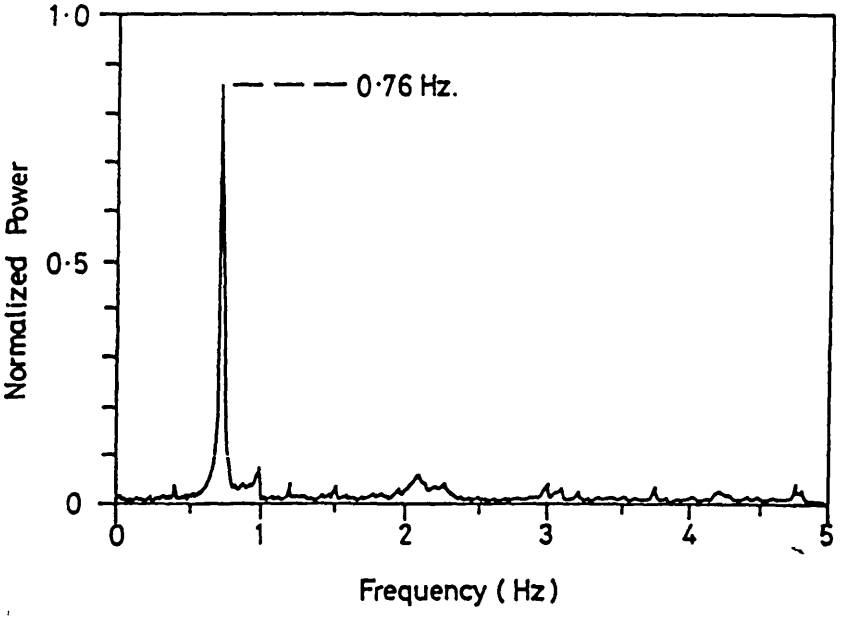


Fig. 4. Frequency response acquired using accelerometer data, Model 1.

to be attributable to the number of samples obtained. With a sampling frequency of 58 Hz and a total number of samples of 256, the maximum frequency resolution (without interpolation) is 0·226 Hz.

The data collected using the MA200 on the double pendulum model are shown in Fig. 5. It can be seen that the motion depicted cannot be described by a single harmonic frequency but has two or more superimposed simple harmonic oscillations. Analysis of the displacement-time data using the FFT enabled

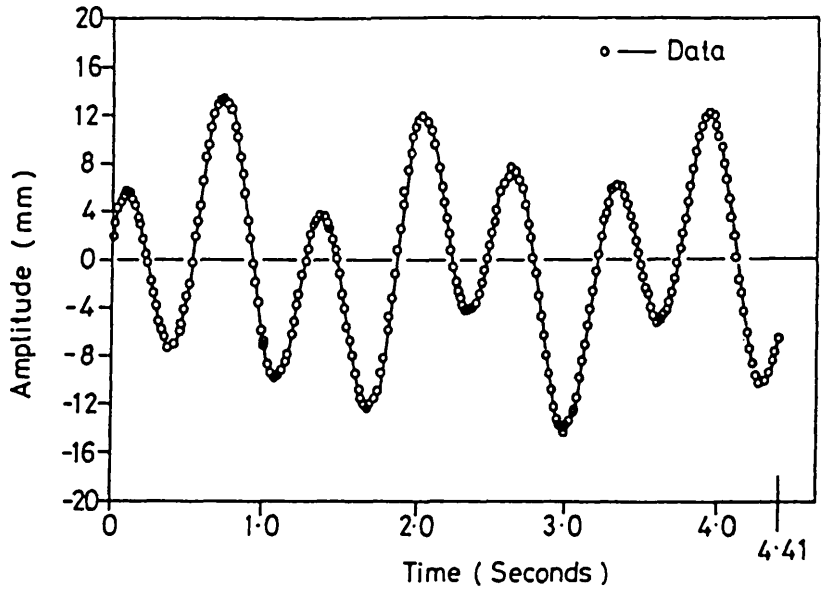


Fig. 5. Displacement-time data acquired using MA200, Model 2.

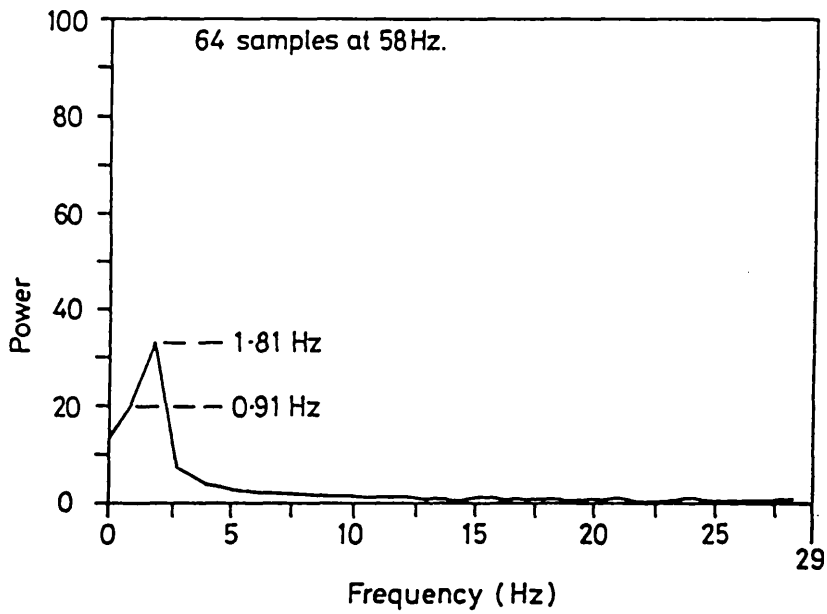


Fig. 6. Frequency response acquired using MA200 data (64 samples), Model 2.

definition of these compound frequencies. In order to demonstrate the effect of sample size on the resolution of the frequency data three Fourier transformations were performed on the data, using 64, 128 and 256 samples. Figs 6, 7 and 8 show the result of the post-processing of the data in the frequency domain. It should be noted that linear interpolation of the frequency data has been used.

Three independent modes of oscillation and associated harmonic frequencies were expected for the motion of model 2, as is shown in Fig. 2(*b*). However, the resolution within the frequency domain is directly related to the frequency and the size of the sample, as is described by the relation:

$$\text{Frequency resolution} = \frac{\text{Sampling frequency}}{\text{Number of samples}}.$$

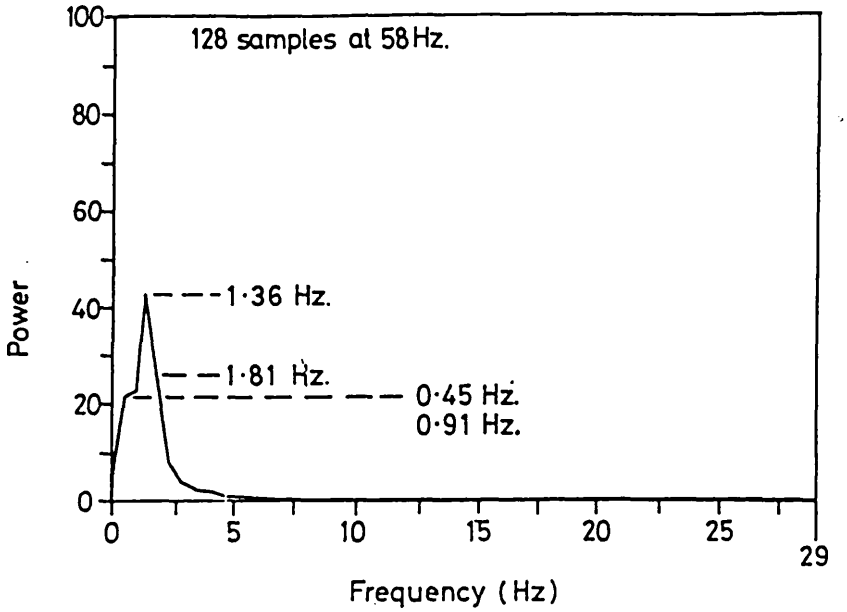


Fig. 7. Frequency response acquired using MA200 data (128 samples), Model 2.

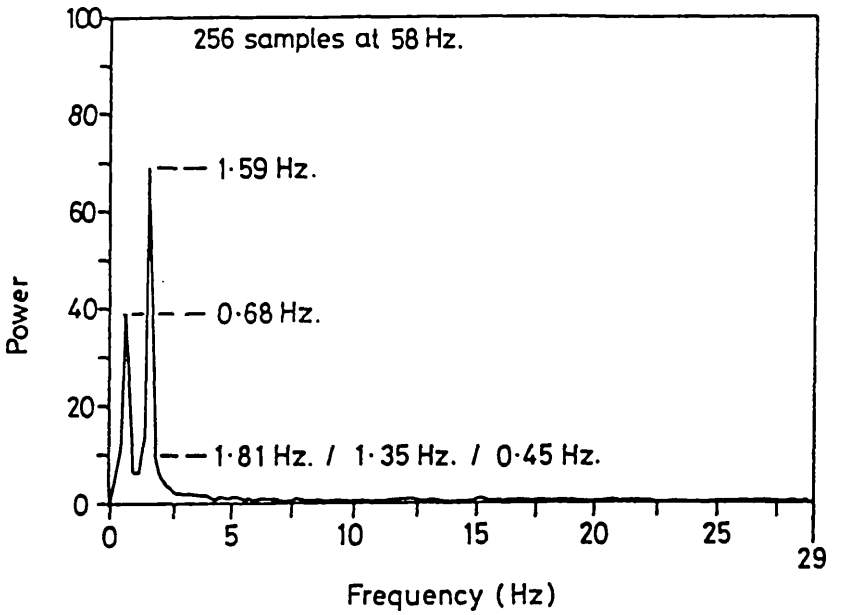


Fig. 8. Frequency response acquired using MA200 data (256 samples), Model 2.

This is clearly demonstrated by the frequency data obtained for the three sample sizes. When 64 samples are used with a sampling frequency of 58 Hz, the maximum frequency resolution is 0·906 Hz. The resultant number of points up to the 'Nyquist cut off frequency' of 29 Hz is 32. Only two distinct harmonic frequencies are identifiable with this amount of data. However, when larger sample sizes are used the definition within the frequency domain increases enabling the identification of three or more possible harmonic frequencies. It must be noted, however, that the clear definition at only two high amplitude frequencies is possible from the data generated using the maximum sample size of 256 samples in Fig. 8. Three further harmonics are identifiable but are not distinct.

Part of the data collected for Model 2 using the accelerometer is shown in Fig. 9. The sampling rate used was 100 Hz and 4000 samples were taken over a period

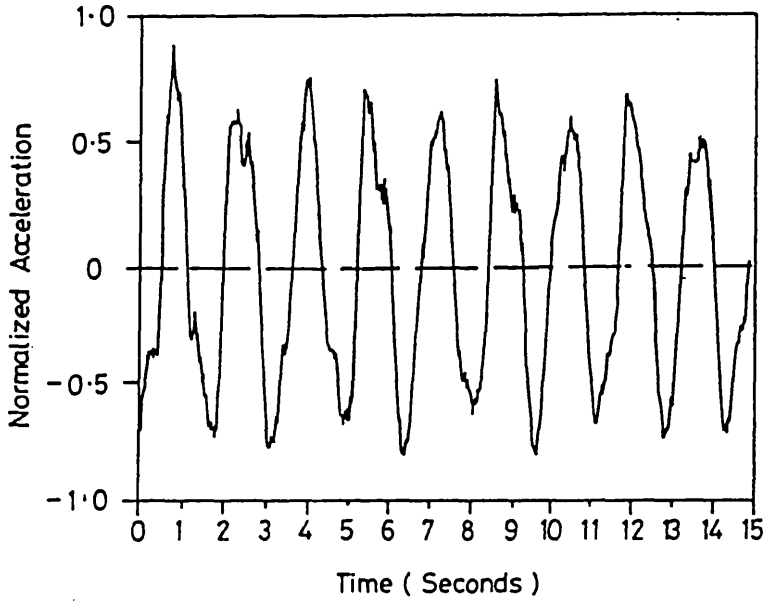


Fig. 9. Acceleration-time data acquired using accelerometer, Model 2.

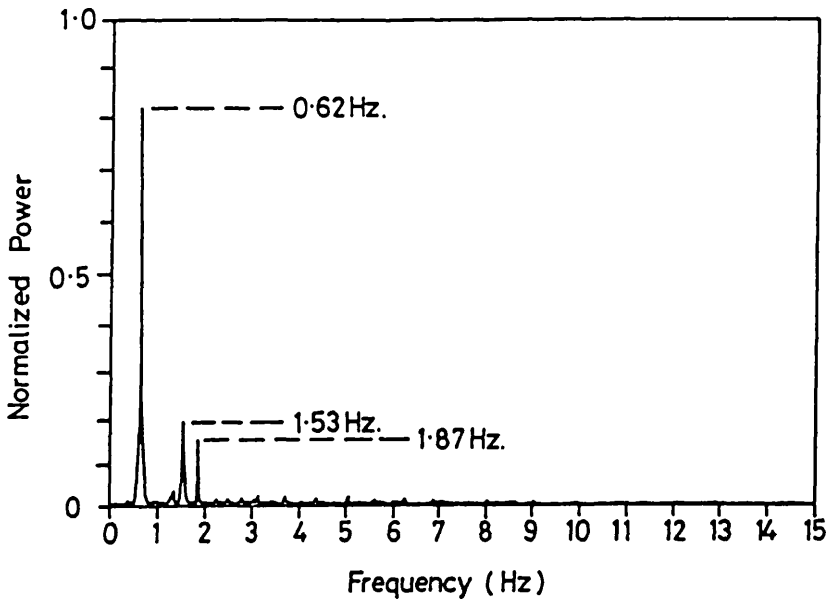


Fig. 10. Frequency response acquired using accelerometer data, Model 2.

of 40 s. Therefore, the frequency resolution is 0·025 Hz. The data does not depict the smooth sinusoidal form of that collected by the MA200. However, the periodic motion is at a lower frequency than the interfering noise. Hence, the relatively high frequency, which is probably attributable to electrical and vibrational noise, does not interfere with the harmonic frequency determination. Fig. 10 shows the FFT data for the acceleration response. There are three clear harmonic frequencies at 0·62 Hz, 1·53 Hz and 1·87 Hz.

A comparison of the harmonic frequency data acquired using the two measurement techniques illustrates a significant level of agreement. In the case of Model 1 (the single harmonic frequency) the difference between the two measurement techniques is 0·02 Hz. This is considered insignificant given that as mentioned previously the maximum frequency resolution of the MA200 is

0·226 Hz. In the case of Model 2 (with multiple harmonic frequencies) the data from the accelerometer source confirms the existence of three fundamental frequencies. The data for the MA200 also agrees closely to these results, the maximum variation being 0·06 Hz, again well below the frequency resolution of the MA200. It is interesting to note however the existence of two further high amplitude frequencies at 1·35 Hz and 0·45 Hz. This is again a product of the frequency resolution of the MA200, since data points only exist at intervals of 0·266 Hz. In the FFT computation, harmonics which exist between datapoints will become apparent in the analysis. For example in the case of the lowest of the three harmonics (0·62 Hz) the MA200 has identified 2 components adjacent to this at 0·452 Hz (the second data point) and at 0·679 Hz (the third data point). Given the relative magnitudes at these frequencies it is clear that the precise location of the harmonic lies closer to the latter of these results. A similar analysis can be used to explain the existence of a harmonic at 1·35 Hz.

Investigation of Low Cost Retro-Reflectors

Virtually all EDM instruments require a method of reflecting the incident signal and this is usually achieved by means of precisely manufactured glass corner cube reflectors which return the signal along an identical path back to the instrument. Such reflectors are often referred to as retro-reflectors. However, it is also possible to use less precisely manufactured, and consequently cheaper, retro-reflective devices and the characteristics of a number of these are discussed in Kennie [8].

In the field of vibrational analysis such retro-reflectors have a number of advantages (in addition to cost) over conventional glass retro-reflectors. In particular from a practical point of view they are more robust, lighter and are easier to attach to a structure. In order to investigate the modes of vibration of a tall structure, several reflectors would be needed and in this case there would therefore be significant benefits associated with the use of such reflectors.

The particular retro-reflectors tested were

(*a*) multiple, glass corner cube design (GCC) (26 and 85 mm diameter)
(*b*) clear, reflex acrylic corner cube design (ACC) (11 and 35 mm diameter)

Measurements were obtained of distance, return signal strength and scatter of phase for each reflector type over a range of distances from 7 m to 200 m.

Test Results

A summary of the results is provided in a graph of signal strength against distance and and is shown in Fig. 11. It can be seen that as one would anticipate, signal strength decreases with increasing distance from the instrument to the prism. The loss of signal strength at longer distances is primarily caused by imperfect reflection and loss of energy of the signal.

The fall in signal strength for both designs as distance from the instrument is decreased below 22 m is anomalous. These results may be explained by the biaxial geometry of the MA200 transmitting and receiving optics. The receiving optic is located 63 mm to the side of the transmitting optic. Therefore, to ensure reception at the receiving optic, a reflector would have to cause a sideways shift in the beam

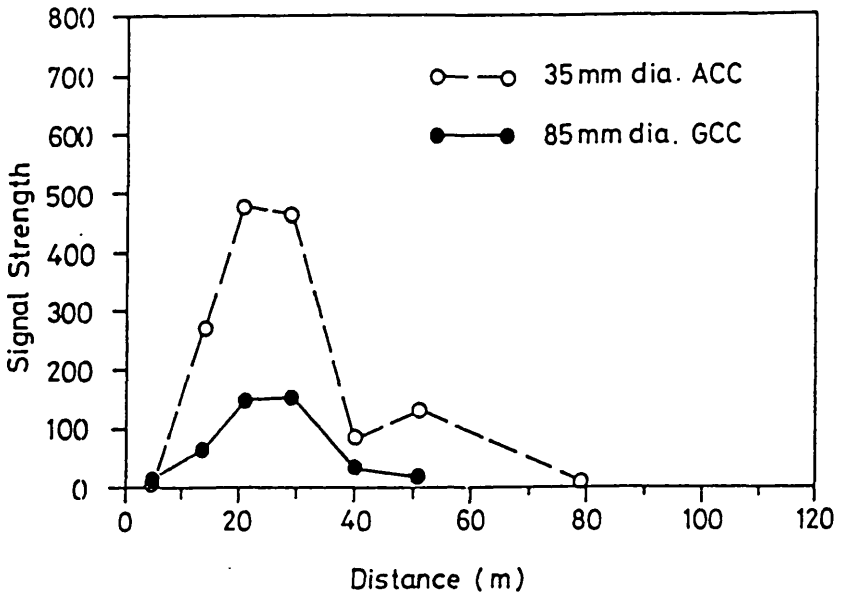


Fig. 11. MA200 return signal strength as a function of distance for two retro-reflectors.

of 63 mm. Over shorter distances when the beam has not diverged, the low cost retro-reflector designs do not cause this shift. However, as the distance increases and the beam becomes wider, more of the reflected beam is directed into the receiving optic.

In summary, the results indicate that although the signal strengths are much lower for the low cost retro-reflectors, they are potentially capable of being used to measure the vibration with the condition that the distance from the instrument is not greater than 70 m and not less than 7 m (in the case of the ACC design).

Conclusions

It can be concluded from the preliminary experiments on vibrational data acquisition and analysis using the MA200 that;

(*a*) The maximum frequency resolution of the instrument in its current form is 0·226 Hz. For applications where determination of vibrational frequencies of the order of 0·5 Hz is acceptable this limitation is clearly insignificant. However, in circumstances such as the determination of changes in vibrational behaviour of structures, a higher frequency resolution would be required since the changes may be well below the resolving capability of the MA200.

(*b*) In order to resolve harmonic frequencies more precisely there is a need to increase the number of samples acquired by the instrument, since the frequency resolution is inversely proportional to the number of samples obtained. For example, if a vibrational monitoring survey to determine changes in structural behaviour required a resolution of 0·05 Hz then using a sampling frequency of 58 Hz the sample size would need to be increased to a least 1160 samples. This would also require modifications to the storage device used in the field to record the data. Alternatively some method of continuous data acquisition would provide a very flexible solution.

(*c*) Low cost retro-reflectors are capable of being used for vibrational measurement provided that the range does not exceed 70 metres.

Finally it is considered that further work should be undertaken to investigate the potential use of this instrument for structural inspection. This should include field studies on existing flexible structures in conjunction with conventional inspection techniques.

Appendix I: Fourier Series for Discrete Data

A sample $x(t)$ is at finite length T which consists of an even number of records N at equal spacing h, where h is selected to produce a sufficiently high cut-off frequency, $f_c = h/2$.

then: $$x_n = x(n.h) \quad n = 1, 2, \ldots N$$

$$\text{Tp} = N.h = \text{total period}$$

the finite Fourier series which will pass through these N data values is:

$$x_n = A_0 + \sum_{q=1}^{N/2} A_q \cos\frac{(2\pi.n.q)}{N} + \sum_{q=1}^{(N/2)-1} B_q \sin\frac{(2\pi.n.q)}{N}$$

and the coefficients are given by:

$$A_0 = \frac{1}{N}\sum_{n=1}^{N} x_n = \bar{x}$$

$$A_q = \frac{2}{N}\sum_{n=1}^{N} x_n \cos\frac{(2\pi.n.q)}{N} \quad q = 1, 2, \ldots, \left(\frac{N}{2} - 1\right)$$

$$A_{N/2} = \frac{1}{N}\sum_{n=1}^{N} x_n \cos q\pi$$

$$B_q = \frac{1}{N}\sum_{n=1}^{N} x_n \sin\frac{(2\pi.n.q)}{N} \quad q = 1, 2, \ldots, \left(\frac{N}{2} - 1\right).$$

References

1. Allan, A., 1988. Modifications to the Tellurometer MA200 for use in large scale metrology. Paper presented at Seminar on Large Scale Metrology, University College London, September.
2. Ashkenazi, V. *et al.*, 1979. Measurement of dynamic deformation by laser interferometry. *The Civil Engineering Surveyor*, Oct.-Nov.: 11–14.
3. Bendat, J. S. & Piersol, A. G., 1971. *Random data: Analysis and measurement procedures*. Wiley Interscience, pp. 299–309.
4. Creasy, L. R. *et al.*, 1965. Museum Radio Tower, London. *Proceedings of the Institution of Civil Engineers*, Jan.: 33–78.
5. Ethrog, U. & Horamy, N., 1988. Measuring the harmonic oscillations of tall antennae by photogrammetry. *International Archives of Photogrammetry and Remote Sensing*, **27**: 585–591.
6. Eyre, R. & Smith, I. J., 1977. Wind and traffic induced behaviours of some steel box girder systems. *TRRL*, Suppl. Rep. 275.
7. Halliwell, N. A., 1984. The laser torsional vibrometer: a new instrument. *Journal of Sound and Vibration*, **93** (4): 588–592.
8. Kennie, T. J. M., 1983. Some tests of retro-reflective materials for electro-optical distance measurements. Parts 1 and 2. *Survey Review*, **27** (207): 3–12.
9. N.A.G., 1984. Numerical algorithms user manual. *NAGLIB*: 1851/1491: Mk II, Jan.
10. Pickering, C. J. D. & Halliwell, N. A., 1986. The laser torsional vibrometer: a portable instrument. *Journal of Sound and Vibration*, **107** (3): 471–485.
11. SERC/DOE, 1987. Exciting opportunities for the dynamic testing of structures at full scale. *The Civil Engineering Research Newsletter*, No. 4, April.

12. Ward, H. S., 1984. Traffic generated vibrations and bridge integrity. *Journal of Structural Engineering, ASCE*, **110** (10): 2487–2499.
13. Williams, C., 1983. Some dynamic characteristics of a tall building. *Proceedings of the Institution of Civil Engineers*, **75**: 751–759.

EDM CALIBRATION

EDM INSTRUMENTS: CALIBRATION METHODS AND FACILITIES IN THE UK

Tom Kennie BSc MAppSci ARICS MInstCES *and Susan Brunton* BSc,
Department of Civil Engineering, University of Surrey, and
Arthur Penfold BSc *and David Williams* BSc PhD MInstP,
Division of Mechanical and Optical Metrology,
National Physical Laboratory

The importance of evaluating the effects of systematic instrumental errors in electromagnetic distance measuring equipment is discussed. In particular, a review of current methods of calibrating equipment is presented, together with the results of two questionnaires to assess the distribution and nature of calibration facilities in the United Kingdom. Evidence from the second questionnaire indicates a requirement for a national EDM instrument calibration facility providing traceability to the SI unit of length as maintained at the National Physical Laboratory.

Modern electromagnetic distance measurement (EDM) instruments are generally very reliable instruments which operate well within the manufacturers' accuracy specifications. However, in recent times these devices have become cheaper and virtually automatic in operation. They are consequently more common in the construction industry, and there has been a growing tendency to accept the values obtained from such instruments without question, unless they differ considerably from the results obtained by other techniques. For many engineering survey operations this approach may lead to acceptable results. However, for more precise surveys and setting-out operations it is imperative that all EDM instruments and reflectors are regularly calibrated to evaluate the influence of any instrument errors on the field observations. These instrument errors are termed scale (k_1), zero (k_2) and cyclic (k_3) errors. A comprehensive account of the reasons for their existence, together with their effects, can be found in [4].

The magnitudes of the possible errors, which may be undetected without regular instrument calibration, are difficult to determine. However, some recent results on an Australian calibration baseline [3] indicated that 29 per cent of instrument calibrations carried out showed scale errors greater than 10 ppm (10 mm/km) and in 20 per cent of the calibrations the amplitude of the cyclic error curve was greater than 6 mm. Therefore some means of assessing the state of adjustment of an EDM instrument is necessary.

In general, the distinction can be made between quick and simple tests for accuracy which could be carried out on site, and the more thorough calibrations, involving the use of a laboratory or permanently sited baseline, for more accurate determinations of the state of adjustment of an EDM instrument. The former type of test is currently under consideration by the International Standards Organisation (ISO) [16]. Whatever method is used, it will only successfully determine the instrument errors if suitable precautions have been taken to reduce to a minimum other possible sources of error, such as incorrect centring or pointing.

Section 2 of this paper therefore outlines briefly a number of methods of carrying out EDM calibration. Considerable efforts have been made in recent years to provide EDM instrument calibration facilities in the United Kingdom

but these are not formally traceable to the SI unit of length. Sections 3 and 4 summarise the results of two questionnaires which were sent to a number of surveying organisations to determine the number and nature of the facilities currently available. Finally, consideration is given in section 5 to the creation of a national calibration service for the UK.

2: Methods of calibration

In general three distinct approaches to instrument calibration can be identified, namely, laboratory methods, field methods using either unknown or known baseline lengths, and field methods using networks.

Under laboratory methods, laboratory tests can be carried out to assess both the scale and cyclic error components. Tests of this nature would normally be carried out by either an instrument manufacturer or an agent who offers calibration services. Their main advantages over field tests are that firstly they are quick and convenient to apply in practice and, secondly, they enable the surveyor to control precisely the measurement conditions. For practical reasons, however, they are limited to a range of about 50 m.

Scale error is determined in the laboratory by comparing the observed modulation frequency of the EDM signal with a reference frequency. The reference frequency may be obtained either from off-air radio transmissions, for example, Droitwich MSF in the UK, or from a crystal-generated laboratory standard.

Cyclic errors can be determined in the laboratory by recording distance measurements to a series of reflector positions which extend over a distance equivalent to the half wavelength of the instrument. The simplest and cheapest arrangement involves the use of a graduated metal bar [10]. Other authors who have investigated the use of a short baseline (10 m to 30 m) include [1, 9, 6].

In situations where space is limited, the arrangement in [7] may be adopted. A measurement range of 50 m was obtained by bending the path of the EDM beam, using surface-coated mirrors. The calibrated length in this case was an 11 m track on which an EDM reflector was mounted. A laser interferometer provided a precise measurement of the movement of the reflector, which could be compared with that obtained using the EDM instrument. This apparatus considerably simplified the practical difficulties and time required to carry out a cyclic error test.

Under field methods measurements on a calibration baseline can be used to determine all three error components (scale, zero and cyclic errors). Both unknown and known baseline lengths may be used, although the former tends to be the more popular technique in some countries such as Australia [13].

When *unknown* baseline lengths are considered, the main *advantage* of the unknown baseline length approach is that although stability of the markers or pillars is important, the points are only required to be stable for the duration of the survey observations. If a known baseline length is used, the long-term pillar stability and periodic monitoring of the interpillar distances are of particular importance. The main *disadvantage* of the unknown baseline length method is that it is not possible to derive a value for scale error by field observation.

The simplest procedure for a quick and approximate determination of the zero error involves the use of a three point baseline [8]. If the observed distances l_{12}, l_{23}, and l_{13} each have the same zero error k_2, then, ignoring all other errors:

$$k_2 = l_{12} + l_{23} - l_{13}. \qquad (1)$$

If the baseline, of total length L, consists of more than two sections and the

length l_i of each of the n sections is measured, then a more precise value of k_2 can be derived using:

$$k_2 = \frac{L - \Sigma l_i}{n - 1} \quad (2)$$

A further improvement in accuracy can be achieved by measuring not only the individual section distances but all possible combinations of distances on a baseline. Least squares methods are then used to derive the most probable values for the zero error and in some cases the cyclic error. One of the most common methods used is in [15].

The Schwendener method improves the determination of k_2 by making use of a seven point baseline (six sections). Measuring all possible combinations of distances results in a total of 21 measurements. Since seven measurements are required for a unique solution (six interpillar distances and k_2) the solution is overdetermined with 14 redundant measurements. The choice of interpillar distances should ensure firstly that the distances to be measured are spread over the measuring range of the instrument and secondly that the fractional elements of the interpillar distances are evenly distributed over the half wavelength of the modulation signal, normally 10 m.

An indication of the cyclic error can be obtained by plotting the residuals, the differences between the most probable values of the distances and the observed values, as a function of the excess over a whole number of cycles of the individual measured distances.

When using *known* baseline lengths, the use of defined values for the length of the baseline and its sections is advantageous since it enables the value of the scale error to be derived in addition to zero and cyclic errors.

A simple field test to evaluate scale error is described in [4]. The technique is based on the use of a single baseline of known length which is also equal to an integral number of modulation half wavelengths. By taking measurements to retroreflectors on the end points A and B, from a point C which is on line with them, a value for the distance AB can be determined from the difference between these measurements. This differential measurement of the distance will eliminate any zero error. Thus the ratio of this measured length to the known length of the baseline will indicate the scale error. For more precise determinations, cyclic effects may be reduced by taking further differential measurements from a series of points C_2 or C_3 which span a distance of half a wavelength. Although this simple test is sufficient for a quick assessment of the scale error of an instrument, it does not provide any information on zero or cyclic errors.

A variation on the Schwendener method which includes measured distances and overcomes these limitations has been developed at the University of Nottingham [2, 7]. The Nottingham method uses a baseline consisting of seven pillars which span a total distance of 817 m. By employing a laser interferometer, the four central bays (197 m) were measured to an accuracy of ± 0.1 mm. It has been shown [2] that the basic Schwendener model can be altered to include the effects of scale and cyclic error. The inclusion of known distances into a Schwendener type of baseline has also been adopted by the UK Ordnance Survey and the US National Geodetic Survey [17].

An alternative approach to the deduction of scale, zero and cyclic errors using known baseline lengths and least squares regression techniques has been proposed [11, 18, 19]. The Sprent and Zwart method is based on a seven-point (six-section) baseline. Unlike the Schwendener method, only ten field measurements need to be recorded since the precision obtained has been found

to be equivalent to that obtained using 21 measurements. This method is based on the differential measurement from two pillars (1 and 2) 5 m apart, to a series of points *A* to *E* at distances of 174 m, 393 m, 505 m, 752 m and 931 m from pillar 1. Each of the measurements from 1 and 2 to the points on the base differ by 5 m, which is equivalent to the half wavelength of most modern instruments. The effect of cyclic errors with only odd harmonics will be eliminated by this differential method because, irrespective of the phase of the signal, the signs of the errors in distances differing by half a cycle will be opposite.

When using *networks*, the determination of the cyclic error by means of a network adjustment is discussed in [6]. This method is based on an adaptation of the fundamental distance observation equation used in least squares estimation by variation of coordinates. Conradie and Merry used a mixed network of 42 distances and 36 directions to evaluate the cyclic error of a Tellurometer MA-100. The results were in general agreement with those obtained using the graduated bar technique. However, the method suffers from the following *disadvantages:*

(i) great care is required in the design of the network to ensure that the distances measured are evenly spread over the half wavelength of the instrument. It would seem impracticable to design such a network specifically for calibration;

(ii) the effects of mis-centring the instrument can seriously degrade the accuracy of the determination of the cyclic error components. Unless forced centring is used, the technique would not appear to be significantly better than the methods already discussed. The technique does have some merit in that no special baseline is needed, or calibration takes place under field conditions.

3: EDM calibration facilities – the Surrey survey

In order to assess the current situation in the United Kingdom regarding instrument calibration, two questionnaires were sent out in 1986 – the first by the University of Surrey and the second, independently, by the National Physical Laboratory (NPL). The first questionnaire was sent to a selected sample of 60 surveying organisations. It was adapted from a similar survey which was carried out in 1980 in Australia [13]. A response rate of over 50 per cent was achieved and a summary of the main findings from the returned questionnaires is given in tables 1 and 2. The results of the questionnaire indicate:

(a) the existence of 15 multi-point baselines or networks which are used for EDM calibration. A further five organisations reported the use of short two-point baselines of known lengths (100 m – 300 m) for approximate checking of instruments;

(b) the majority of the baselines are similar to the seven point design originally suggested by Schwendener (nos 1, 7, 9 and 12), although the interpillar distances are not identical in any case and a further four (nos 8, 10, 13 and 14) are variants of the classic seven point design;

(c) the maximum length of baseline is 1 100 m (no 12) and the minimum 34 m (no 15), the average length being 529 m;

(d) pillars with forced centring are used in four cases (nos 1, 8, 9, and 15); fixed leg tripods specifically adapted for this type of operation are used by the Ordnance Survey (no 12), and semi-permanent tripods are used by the University of Newcastle (no 4);

(e) in virtually all cases the responses indicate that all three error sources could be determined by calibration. In one case scale error was not determined by baseline measurements (no 7).

EDM CALIBRATION

Organisation	Laboratory Testing Frequency	Cyclic	Contact	General Use
(a) Universities and polytechnics				
1 City University	No	Yes	Mr M A R Cooper, Department of Civil Engineering, The City University, Northampton Sq, London EC1V OHB. Tel: 01-253-4399	No
2 Nottingham University	Yes	Yes (by laser interferometer)	Dr A H Dodson, Dept of Civil Engineering, University of Civil Engineering, University Park, Nottingham NG7 2RD. Tel: (0602) 506101, ext 2667	No
3 University College London	Yes	Yes	Dr A L Allan, University College, London, Dept of Photogrammetry and Surveying, Gower St, London. Tel: 01-387-7050	No
4 Kingston Polytechnic	No	By steel tape	Mr D H Simm, Canbury Park Rd, Kingston upon Thames KT2 6LA. Tel: 01-549 0151	/
5 NE London Polytechnic	Yes	Yes	Mr R Fisher, NE London Polytechnic, Longbridge Road, Dagenham, Essex RM8 2AS. Tel: 01-590-7722	No
6 Trent Polytechnic	No	Yes	Mr F A Shepard (Principal Lecturer Civil Eng), Dept of Civil and Structural Eng, Trent Polytechnic, Burton St, Nottingham NG2 6JH. Tel: (0602) 418248, ext 2032	/
(b) Central Government				
7 Ordnance Survey	Yes	By steel tape	OS Accuracy Testing Station, C528, Romsey Road, Maybush, Southampton. Tel: (0703) 792483	No
(c) Instrument suppliers				
8 Geotronics Ltd	Yes	Yes (each time a diode is replaced)	Mr M Peak, Geotronics Ltd, West Newlands Estate, St Ives Rd, Somersham, Huntingdon, Cambridgeshire PE17 3EB. Tel: (0487) 842282	No
9 Survey and General Inst Co Ltd	Yes	Yes	Dr M Grist, Survey & General Inst Co Ltd, Fircroft Way, Edenbridge, Kent TW8 6HA. Tel: (0732) 864111	/

Table 1. *Summary of responses to laboratory calibration facilities*

4: EDM calibration facilities requirements: the NPL survey

Independent of the survey of existing EDM calibration facilities in the UK carried out by the University of Surrey, the NPL carried out a nationwide survey to establish the demand for a UK EDM calibration service which would ensure that measurements were traceable to the SI unit of length. A number of other countries such as Australia [5, 14], the USA [20] and Canada [12] have already created government-sponsored facilities for the accurate calibration of EDM instruments.

The survey was conducted by circulating a questionnaire to EDM instrument manufacturers or their agents, land surveyors, government and educational establishments and other users of EDM instruments. Out of 104 questionnaires 58 were returned (55 per cent). Out of the completed questionnaires 75 per cent considered that a service should be set up and said they would use it. The 43 organisations which wanted the calibration service had between them approximately 200 EDM instruments of varying types requiring calibration once per year. At present many of these instruments are checked by the

Table 2. *Summary of responses to University of Surrey questionnaire on EDM calibration facilities – multi-point baselines*

Name	Locality	Overall Length	No of Stations	Section Distance	Pillars	Distance Known	Errors Determined Scale	Zero	Cyclic	General Use	Fee Payable	Contact
(a) Universities and polytechnics												
1 **Bunny**	Nottingham	818	7	380, 422, 473, 526, 577, 818	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Dr A H Dodson, Dept of Civil Engineering, University Park, Nottingham NG7 2RD. Tel: (0602) 506101, ext 2667
2 **Quorn Base**	Leicestershire	942	12	80, 130, 150, 170, 375, 445, 600, 700, 820, 865, 942	No	Yes	Yes	Yes	Yes	No	Yes	Professor D J Hodges, Dept of Mining Engineering, University Park, Nottingham NG7 2RD. Tel: (0602) 506101, ext 2259
3 **Mining Lab Base**	Nottingham	43	5	13, 23, 33, 43	No	Yes	Yes	Yes	Yes	No	/	Professor D J Hodges, Dept of Mining Engineering University Park, Nottingham NG7 2RD. Tel: (0602) 506101, ext 2259
4 **Newcastle University**	Newcastle	445	4 (+2)	84, 154, 207 (this base is a network and also has additional offset lengths)	Semi perm Tripods	Yes	Yes	Yes	Yes	No	/	Dr D Parker Dept of Surveying, Univ of Newcastle-upon-Tyne, Newcastle NE1 7RU. Tel: 091-232-8511, ext 2447
5 **Oxford University**	Oxford	420	3	180, 420	No	Yes	Yes	Yes	Yes	No	/	Dr J G Olliver, Dept of Earth Sciences, 62 Banbury Road, Oxford OX2 6PN. Tel: (0865) 57816
6 **Strathclyde Campus**	Glasgow	108	3	59, 108	No	Yes	Yes	Yes	Yes	Yes	/	Dr P H Milne, University of Strathclyde, Dept of Civil Engineering, John Anderson Building, Glasgow G4 ONG. Tel: 041-552-4400, ext 3281
7 **Coventry Airport**	Coventry	1 000	7	410, 460, 515, 564, 610, 1 000	No	Yes	No	Yes	Yes	No	Yes	Mr P A Greengrass, Dept of Civil Eng & Building, Coventry (Lanchester) Polytechnic, Priory St, Coventry. (0203) 24166

8 NE London Polytechnic	Dagenham	294	6	15, 30, 81, 162, 294	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Mr R Fisher, NE London Polytechnic, Longbridge Road, Dagenham, Essex RM8 2AS. Tel: 01-590-7722
9 Robertson Baseline	Paisley	1 010	7	162, 248, 263, 393, 935, 1 010	Yes	Yes	Yes	Yes	Yes	No	/	Mr J Paul, Dept of Civil Engineering, Paisley College of Technology, Paisley PA1 2BE. Tel: 041-887-1241
10 Trent Polytechnic	Clifton Nottingham	254	5	45, 103, 175, 254	No	Yes	Yes	Yes	Yes	Yes	/	Mr F A Shepard (Principal Lecturer Civil Eng), Dept of Civil & Structural Eng., Trent Polytechnic, Burton St, Nottingham NG2 6JH. Tel: (0602) 418248, ext 2032
(b) Central Government												
11 Barton Stacey Calibration Bay	Andover	276	4	88, 178, 276	No	Yes	Yes	Yes	Yes	Yes	/	Officer Commanding 19 TOPO Sqn RE, 42 Svy Engr Gp, Hermitage, Newbury, Berks. Tel: (0635) 200371, ext 269
12 Ladycross Baseline	New Forest	1 100	7	200, 400, 600, 800, 1 000, 1 100	Fixed Tripods	Some	Yes	Yes	No	No	Yes	OS Accuracy Testing Station C528, Romsey Road, Maybush, Southampton. Tel: (0703) 792483
13 Kenley	RAF Kenley	600	5	100, 200, 300, 600	No	Yes	Yes	Yes	Yes	No	/	DCES Central Survey Branch, 8th Floor, Apollo House, 36 Wellesley Road, Croydon. Tel: 01-686-5622
(c) Instrument suppliers												
14 SGJ Baseline	Bromley	600	8	60, 100, 103, 105, 108, 146, 200, 600	No	Yes	Yes	Yes	Yes	No	/	Dr M Grist, Survey & General Inst Co Ltd, Fircroft Way, Edenbridge, Kent TW8 6HA. Tel: (0732) 864111
15 Geotronics	Cambridgeshire	34	3	11, 34	Yes	Yes	Yes	Yes	Yes	No	/	Mr M Peak, Geotronics Ltd, West Newlands Estate, St Ives Rd, Somersham, Cambs PE17 3EB. Tel: (0487) 842282

manufacturers to an accuracy of ± (5 mm + 5 ppm), but many organisations have asked for a higher accuracy of ± (1 mm + 1 ppm), or better. The highest accuracy asked for was ± 0.1 mm, this being for the calibration of a Mekometer. The ranges over which the EDM instruments were required to be calibrated were also requested. Nineteen respondents asked for up to 1 000 m which would cover normal civil engineering distances for EDM instruments using infra-red sources; the remaining 24 asked for over 1 000 m, the longest being 20 km.

The survey showed 17 organisations which have their own baselines, although in many cases they only consisted of two reference points with a known separation. The locations and types of multi-pillar baselines had already been identified by the University of Surrey and are listed in table 2. Also shown are organisations which have laboratory facilities for determining the cyclic error of EDM instruments (table 1).

A new venture by Thames Water is their proposal to set up an EDM calibration facility both for their own use and for outside organisations on a repayment basis. They have a very suitable site available at their Ashford Common works on which to build an 800 m multi-pillar baseline, similar to the baseline at Nottingham University (no 1) and to provide a 25 m laboratory facility for determining the cyclic error of EDM instruments. This baseline would be situated close to the NPL whereas the longest baseline that can be accommodated on the NPL site would be approximately 300 m.

None of the facilities identified in tables 1 and 2 has traceability to the SI unit of length as maintained at the NPL. Arrangements for ensuring measurement traceability already exist through the British Calibration Services (BCS) which is part of the National Measurement Accreditation Service, an organisation which is part of NPL. Accreditation to carry out calibrations to national standards is essential. After accreditation, the organisation is entitled to carry out calibrations for customers on a repayment basis, to an agreed level of accuracy, and to issue authorised calibration certificates.

For the calibration of extremely high accuracy distance measuring instruments, the NPL is currently setting up a 50 m indoor laboratory facility in which the instrument will be calibrated with reference to a laser interferometer which will be traceable to the SI unit of length. An instrument can then be used as a transfer standard by NPL for calibrating baselines of outside organisations wishing to become approved. The 50 m laboratory would also be equipped to measure the cyclic error of EDM instruments to a high accuracy.

5: Conclusions

The measurement of distances by electronic methods is now commonplace and accepted practice not only in land surveying but also in the civil engineering and building industries. The trend towards further operational simplicity and automation of the measurement process may lead, however, to inaccurate results being obtained, particularly if the effects of systematic instrumental errors are not determined.

There are no British Standards dealing with the use of EDM instruments although, as noted in the introduction, an ISO standard on calibration checks for EDM instruments is under consideration. In the UK, quality assurance procedures are currently being drafted for land survey work which will require traceable calibration facilities. It is therefore recommended that a working party of experts from industry, government and academe be set up under the auspices of the British Standards Institution to consider the use of EDM instruments in working practice and to make particular recommendations on:

(a) the requirements and methods for EDM instrument calibration;
(b) the frequency of calibration of instruments;
(c) methods for calibrating baselines; and
(d) field procedures for calibration checks.
Such recommendations should take account of any existing specifications and standards approved in other countries or by other international standards bodies.

The response to the NPL questionnaire has shown the need for an authenticated service for the calibration of EDM instruments in the UK for scale, zero and cyclic errors. The NPL intends to arrange the setting up of a BCS service for the calibration of these instruments in the UK.

Acknowledgements

The authors wish to extend their thanks to the organisations who completed the questionnaires on EDM calibration.

References

1. Anon, 1974. Test lines for electro-optical distance measuring. Kern Bulletin, no 20.
2. Ashkenazi A, and Dodson A, 1977. The Nottingham multi-pillar baseline. Proceedings of the XVI General Assembly of the Association of Geodesy, Grenoble, 22 pp (also reprinted in Civil Engineering Surveyor, June, pp 1-8).
3. Benwell G L, Murnane A B and Sprent A, 1985. Results of EDM calibrations. The Australian Surveyor, vol 32, (no 6): pp 122-432.
4. Burnside C D, 1982. Electromagnetic distance measurement. 2nd Edition, Granada.
5. Ciddor P E, Edensor K H, Loughry K J and Stock H M P, 1987. A 70-metre laser interferometer for the calibration of survey tapes and EDM equiopment. The Australian Surveyor, vol 33, (no 6): pp 493-502.
6. Conradie A S, and Merry C S, 1983. A comparison of two methods of determining cyclic errors. South African Survey Journal, vol 19, 114 (no 3): pp 16-25.
7. Deeth C P, Dodson A H, and Ashkenazi V, 1979. EDM accuracy and calibration. Proceedings of the Symposium on EDM, South Bank Polytechnic, 24 pp.
8. Dodson A, 1978. EDM – systematic errors and instrument calibration. ASCE Journal (now Civil Engineering Surveyor) vol 3 (no 4): pp 20-22.
9. Emenike E N, 1982. On the calibration of EDM instruments. The Australian Surveyor, vol 31 (no 3): pp 175-185.
10. Hodges D J, 1982. Electro-optical distance measurement. University of Nottingham Mining Magazine, vol XXXIV, 19 pp.
11. Kassan A M, 1986. EDM baseline calibration for Alberta survey control. Proceedings of the XVIII FIG Congress, Toronto, Paper 510.1, 20 pp.
12. Kelly K, 1986. Precise calibration baselines in Ontario: the program and its status. The Ontario Land Surveyor, Spring, pp 12-14.
13. Rüger J M, 1981. Survey on Australian EDM instrument calibration facilities – 1980. The Australian Surveyor, vol 30 (no 7): pp 458-472.
14. Rüger J M, 1985. Traceability of electronic distance measurement to National Standard. Proceedings 27th Australian Survey Congress, Alice Springs, pp 149-163.
15. Schwendener H R, 1972. Electronic distancers for short ranges: accuracy and calibration procedures. Survey Review, vol XXI (no 64): pp 273-287.
16. Smith J R, 1986. ISO Draft standards on accuracy in use of survey instruments on construction sites, Proceedings of the XVIII FIG Congress, Toronto, Paper 602.1.
17. Spofford P R, 1982. Establishment of calibration baselines. NOAA Technical Memo. NOS/NGS/8. National Geodetic Information Centre, 21 pp.
18. Sprent A, and Zwart P R, 1978. EDM calibration – a scenario. The Australian Surveyor, vol 29 (no 3): pp 157-169.
19. Sprent A, 1980. EDM calibration in Tasmania. The Australian Surveyor, vol 30 (no 4): pp 213-217.
20. Yorczyk, R A, 1986. The National Geodetic Surveys EDMI calibration baseline program. ASP-ACSM Convention, Washington, pp 32-41.

APPENDIX C

DEVELOPMENTS IN MODELLING SURFACE GEOLOGY AND TOPOGRAPHY BASED ON SURVEY DATA

Terrain modelling in surveying and civil engineering

G Petrie and T J M Kennie*

This paper is a review of the application of terrain modelling to surveying and civil engineering. The first section considers the theoretical background to the subject and two distinct approaches are identified based on the use of either grid or randomly located height data. The relative merits of each are discussed and the most frequently used algorithms for interpolation and contouring are examined. This is followed by a review of the characteristics of a selected number of terrain modelling and contouring packages. Five distinct groups of packages are identified. Finally, a review of applications is presented together with some views on the future trends in this field.

terrain modelling, civil engineering, landscape planning

Surface modelling is a general term used to describe the process of representing a physical or artificially created surface by means of a mathematical expression. Terrain modelling is one particular category of surface modelling which deals with the specific problems of representing the surface of the earth.

The term 'digital terrain model' was originally coined in 1958 by Miller and LaFlamme[1]. Since then, the subject has developed considerably and is currently an area of widespread activity in surveying, geology and geophysics, civil and mining engineering and other disciplines in the earth sciences.

The paper is divided into three distinct sections. The first section examines the theoretical background and considers the advantages and limitations of the various interpolation and contouring techniques which are used by the program packages described in the second section. The paper concludes by listing the range of projects where terrain modelling may be applied and considers the most likely trends for the future in this field.

THEORETICAL BACKGROUND

As the review in the later part of the paper will show, a very large number of program packages have been devised and written for terrain modelling applications in surveying and civil engineering. Each program package has been written independently and usually has been optimized for a specific application or for use on a specific computer or uses a different programming language, or has a distinctive user interface. However, in spite of this diversity, when their characteristics are analysed, it can be seen that basically they follow one or two main approaches (see Figure 1). Either they are based on, or make use of, height data which has been collected or arranged in the form of a regular (e.g. rectangular or square) grid or they are based on a triangular network of irregular size, shape and orientation, and use randomly-located height data.

Department of Geography, University of Glasgow, Glasgow G12 8QQ, UK
*Department of Civil Engineering, University of Surrey, Guildford GU2 5XH, Surrey, UK

This paper was presented to the Conference on State of the Art in Stereo and Terrain Modelling held by the British Computer Society – Displays Group, 28th May, 1986.

Grid based terrain modelling

The first approach is in many ways the simplest, in that the data comprising the terrain model is measured or collected in the form of a regular grid. This is in widespread use, e.g. in the grid levelling of building sites carried out by surveyors and civil engineers and in the acquisition of profile data measured photogrammetrically in stereoplotting instruments, in the latter case, especially when derived automatically or semi-automatically during the production

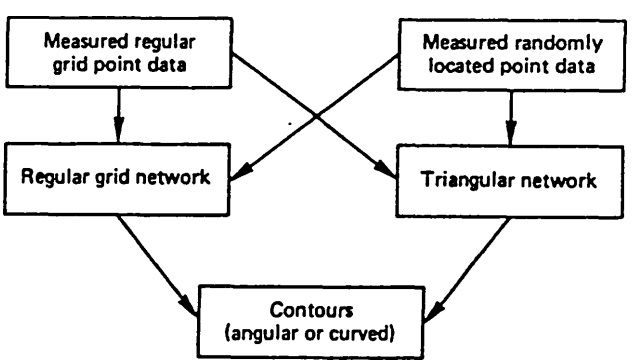


Figure 1. Overall relation between measured point data, networks and contours in terrain modelling

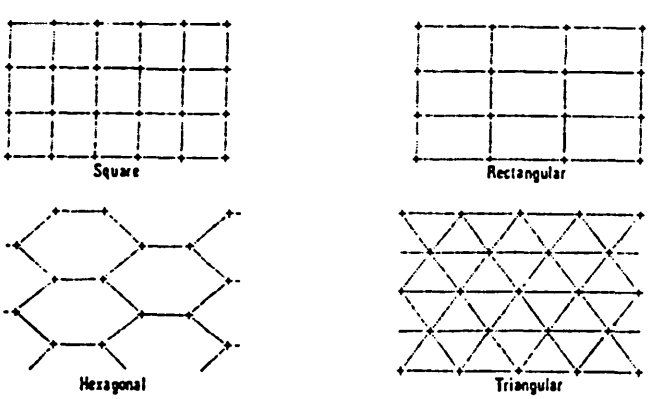


Figure 2. Regular grid patterns

 0010–4485/87/040171–17 $03.00 © 1987 Butterworth & Co (Publishers) Ltd

of orthophotographs or under computer control using an analytical plotter. While the square grid is the most common form, rectangular, hexagonal and triangular-based grids are also encountered and these forms are illustrated by Figure 2.

On the other hand, a shortcoming of this regular grid-based approach is that the distribution of data points is not related to the characteristics of the terrain itself. If the data point sampling is conducted on the basis of a regular grid, then the density must be high enough to accurately portray the smallest terrain features present in the area being modelled. If this is done, then the density of data collected will be too high in most areas of the model, in which case there will be an embarrassing and unnecessary data redundancy in these areas. In this situation, filtering of the measured data may need to be carried out as a pre-processing activity before the terrain model can be defined.

Progressive sampling

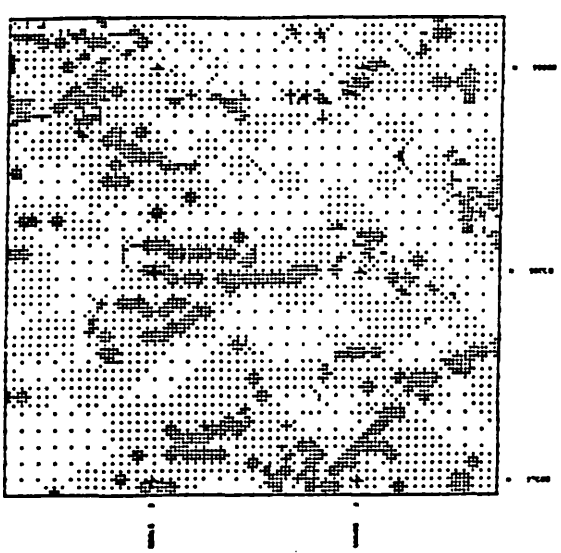
Since the grid-based approach is most easily implemented in computer controlled photogrammetric instruments, a solution to the above mentioned shortcoming has come from the photogrammetrists in the form of progressive sampling, and its development composite sampling, originally proposed by Makarovic[2-4]. Instead of all points in a dense grid being measured, the density of the sampling is varied in different parts of the grid, being matched to the local roughness of the terrain surface.

The basis of the method is that one starts with a widely spread (low resolution) grid that will give a good general coverage of height points over the whole area of the model. Then a progressive increase in the density of sampling (or measurement) takes place on the basis of an analysis of terrain relief and slope using the online computer attached to the photogrammetric instrument. Thus the basic grid is first increased in density by halving the size of the grid-cell in certain limited areas, based on the results of the preceding terrain analysis. Measurements of the height points at the increased density are carried out under computer control only in these predefined areas. A further analysis is then carried out for each of these areas for which the measured data has been increased. Based on this second data analysis, an increased density of points may be prescribed for still smaller areas. Normally, three such runs or iterations are sufficient to acquire the terrain data necessary to define a satisfactory model. Figure 3 illustrates two examples of progressive sampling applied to different data sets.

In this way, the progressive sampling technique attempts to optimize automatically or semi-automatically the relationship between specified accuracy, sampling density and terrain characteristics. Recently, the method has been widely implemented in grid-based terrain modelling packages devised by photogrammetrists, e.g. Hifi (Heights interpolation by finite elements)[5,6], Scop (Stuttgart contour program)[7-9], the package devised in the ITC by Huurneman and Tempfli[10] and the Graz Terrain Model (GTM) package implemented at the Technical University of Graz in Austria.

Random-to-grid interpolation

Especially in land surveying operations, terrain data will be collected by optical or electronic distance-based methods



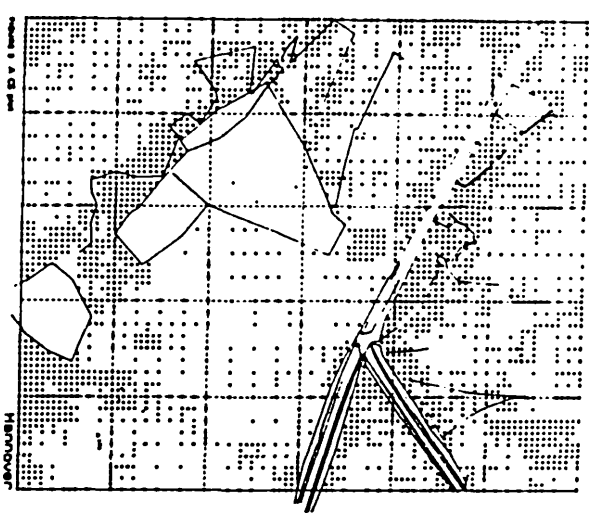


Figure 3. Examples of progressive sampling

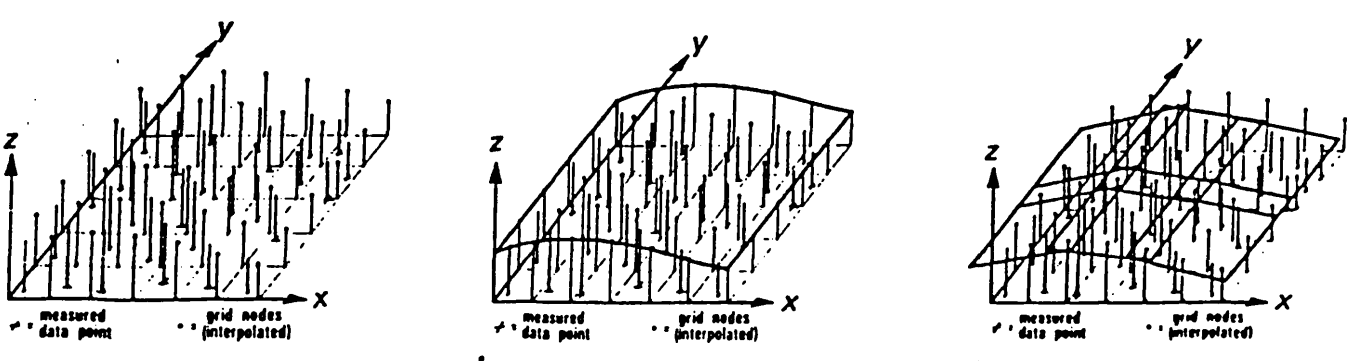


Figure 4. Random-to-grid interpolation methods

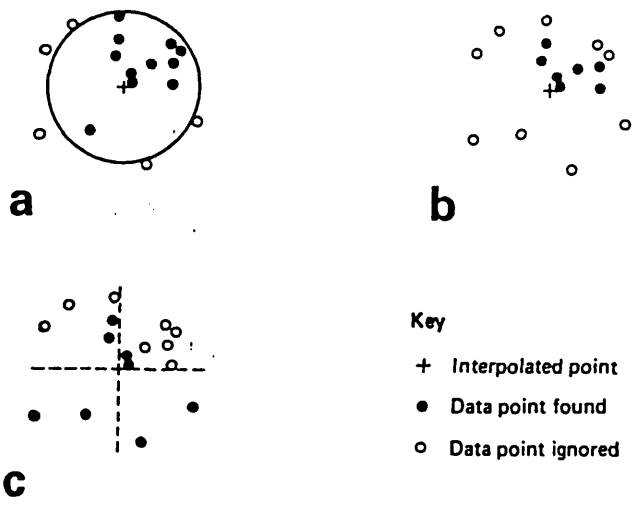


Figure 5. Pointwise interpolation : area search techniques

combined with the necessary angular measurements. Since the surveyor or civil engineer must visit every measured point to set up the measuring staff or retroprism, the opportunity exists to ensure their positioning on points that are important in terms of terrain morphology or representation, e.g. on hill tops, in pits, hollows or saddles, along breaklines such as ridges, breaks of slope, rivers etc. besides giving an appropriate general spread of height points. If a brid-based terrain modelling package is to be used with such specific, but randomly located points, then a preliminary interpolation must be carried out which converts the measured data to a suitably dimensioned regular grid.

The distinguishing feature between different random-to-grid interpolation procedures is the range of the interpolation function which is employed in the interpolation. Usually the following are distinguished:

- Pointwise methods which involve the independent determination of different functional parameters and height values for each and every grid node being interpolated (see Figure 4(a)).
- Global methods which involve the establishment of a single complex three dimensional surface through the complete data set of measured height points with the subsequent interpolation of terrain heights at all the required nodes or points on the regular grid (see Figure 4(b)).
- Patchwise methods in which a series of local 3D surfaces or patches are established from which the elevations of a series of neighbouring grid points lying within each patch can be interpolated (see Figure 4(c)).

Pointwise methods

These involve the interpolation of the values of the height at a specific grid node from its neighbouring randomly located measured height points. Since each point, or node, on the final grid is determined independently of any other node, it has no effect or impact on the adjacent points in the terrain model and a continuous surface may be generated through all the derived grid nodes without any discontinuities or boundary problems.

Almost all the algorithms used for the determination of the height of each individual grid point are based on a search for a set of nearest neighbours, followed by the averaging of their heights weighted inversely by some function of their respective distances, d from the position of the grid node. This weight, $w = 1/d^m$ where m is the power used and is typically in the range 0.5—4. The various search methods are illustrated in Figure 5.

The search for the nearest neighbours may involve a simple area search (see Figure 5(a)) where a circle of predefined radius, or a box of predefined size, is used to select the data points from which the grid point value is determined. A variant of this technique, in which the n nearest neighbours are searched for (see Figure 5(b)), may also be used, n being user defined but usually in the range from 6 to 10. Since these two methods offer no control over the distribution of the points used for the interpolation of the grid node point, a further variant is the sectored nearest neighbour technique in which the area around the grid point, or node, is divided into equal sectors – usually four (quadrants), or eight (octants), with the nearest two or four neighbours being searched for in each case (Figure 5(c)).

Global methods

These methods involve the fitting of a single three dimensional surface defined by a high-order polynomial through all of the measured randomly located terrain height points existing within the model. Once this global surface has been defined, and the specific values of the polynomial coefficients have been determined, the values of the heights for each grid node can then be interpolated. Difficulties may arise if many terms are used (as will be the case with high order polynomials), and if a large data set is required to form the model, resulting in large amounts of computation time. Also the unpredictable nature of the oscillations produced by such high-order polynomials may produce poorly interpolated values for the grid node points.

Patchwise methods

These lie in an intermediate position between the pointwise method and the global method. The whole area to be modelled is divided into a series of equal sized patches of identical shape. The shapes of each patch are regular in form, typically square or rectangular. Quite separate mathematical functions are then generated to form the surface for each patch. Thus separate sets of fitting parameters are computed for each individual patch.

Two distinct methods of patchwise interpolation exist and are illustrated by Figure 6.

Exact fit patches (see Figure 6(a)) may be defined in which each patch abuts exactly onto its neighbours. The difficulty that may result from the use of such patches is that they may result in sharp discontinuities along their junctions, which show up markedly when the isolines or contours are finally produced. The patches are larger than the cells defined by the grid nodes so several grid nodes will fall within a single patch.

The alternative is to use an arrangement of overlapping

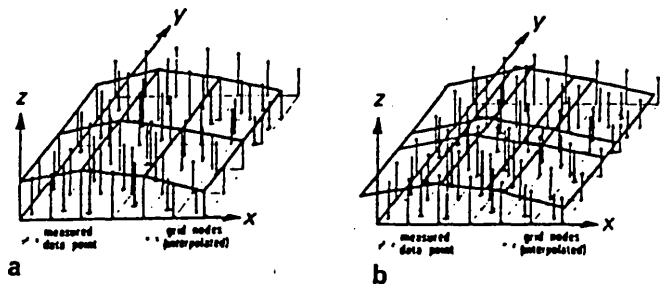


Figure 6. Patchwise interpolation

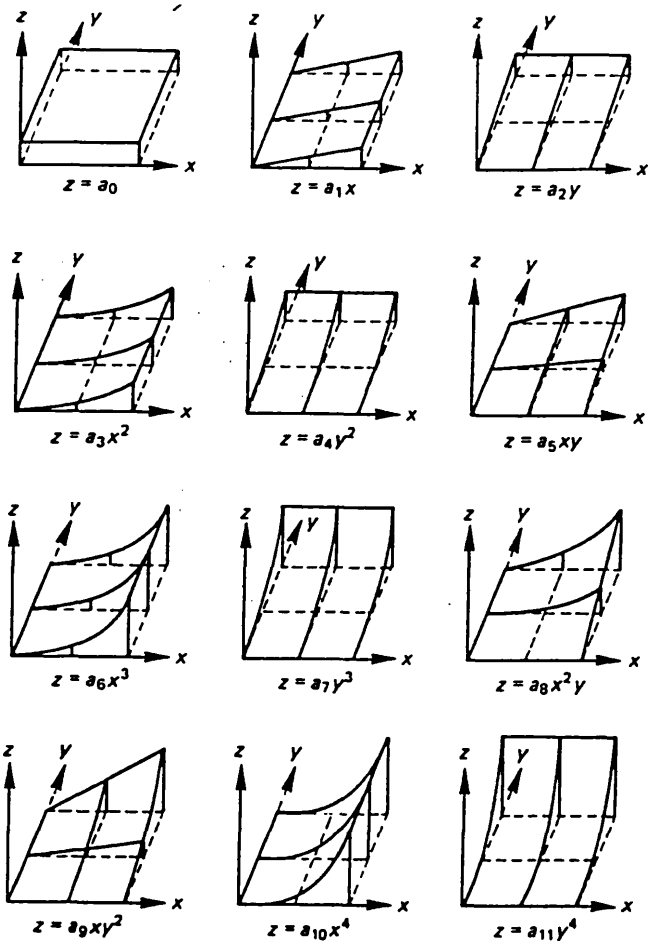


Figure 7. Surface shapes produced by individual terms in the general polynomial equation

patches (see Figure 6(b)), in which case there will be common points lying within the overlap which will be used in the computation of the parameters for each patch.

The advantages of using patchwise methods over global methods are that quite low order terms (parameters) can be used to describe satisfactorily each patch. So only a few unknowns have to be solved via simultaneous equations using least square methods for each patch. Also, once the unknowns have been solved for, it is easy to calculate the derived points, i.e. the grid nodes, by back substitution in the functions or equations describing the patch.

However, there are also some disadvantages of the patchwise method. In the first place, it needs greater organisation of data and of its processing than pointwise or global methods. Also the subdivision of the model surface into patches needs to be carried out with care. If the data is poorly distributed towards the patch corners, then this affects the computed parameters and, in turn, the accuracy of the subsequent heights determined for the grid node points.

Polynomials used for surface representation

As noted above, polynomial equations are used to represent the terrain surfaces in the global and patchwise methods of interpolation. The basic general polynomial equation for surface representation is shown in Table 1, where z_i is the height of an individual point i, x_i and y_i are the rectangular coordinates of the point i and a_0, a_1, a_2, . . . etc. are the coefficients of the polynomial. One such equation will be generated for each individual point i with coordinates x_i, y_i, z_i occurring in the terrain model.

Table 1. General polynomial equations used for interpolation

Individual terms	Order of term	Descriptive term	Number of terms
$z_i = a_0$	zero	planar	1
$+ a_1x + a_2y$	first	linear	2
$+ a_3x^2 + a_4y^2 + a_5xy$	second	quadratic	3
$+ a_6x^3 + a_7y^3 + a_8x^2y + a_9xy^2$	third	cubic	4
$+ a_{10}x^4 + a_{11}y^4 + a_{12}x^3y + a_{13}x^2y^2 + a_{14}xy^3$	fourth	quartic	5
$+ a_{15}x^5 + \ldots\ldots$	fifth	quintic	6

In the first step, the values of x, y and z are known for each measured point present in the overall data set or patch. Thus the values of the coefficients a_1, a_2, a_3 and so on, can be determined from the set of simultaneous equations which have been set up, one for each data point. Once the values of the coefficients a_1, a_2, a_3 and so on have been determined, then for any given grid node point with known coordinates x, y the corresponding height value z can be calculated.

To make a correct selection of the terms which will best represent or model the terrain surface, the surveyor or civil engineer must keep in mind the shape of the individual terms produced in the polynomial equation, as illustrated by Figure 7.

Typical of the simpler types of surface used to model patches are:

- the 4 term bilinear polynomial with the form

$$z = a_0 + a_1x + a_2y + a_3xy$$

(as used in BIPS, Hifi etc)

- the 9 term biquadratic polynomial

$$z = a_0 + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2 + a_6x^2y + a_7xy^2 + a_8x^2y^2$$

(as used in Scop)

- the 16 term bicubic polynomial

$$z = a_0 + a_1x + a_2y + a_3xy + a_4x^2 + a_5y^2 + a_6x^2y + a_7xy^2 + a_8x^2y^2 + a_9x^3 + a_{10}y^3 + a_{11}x^3y + a_{12}xy^3 + a_{13}x^3y^3 + a_{14}x^3y^2 + a_{15}x^2y^3$$

(as used in Hifi and Rijkswaterstadt DTM)

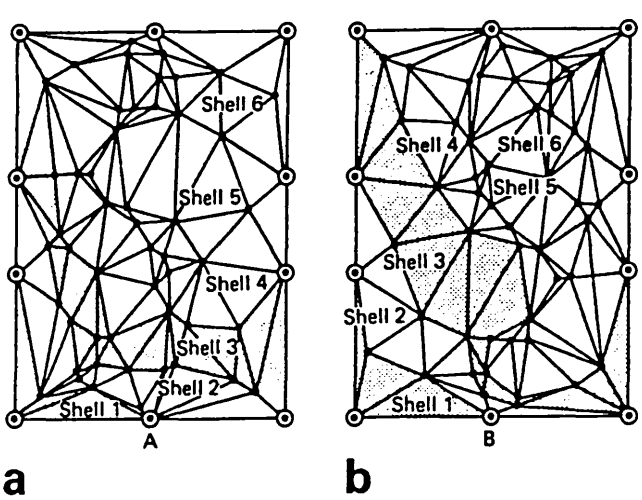


Figure 8. Advancing shells of a Delaunay triangulation[11] *(a) orientation as in Figure 9; (b) rotated through 180°*

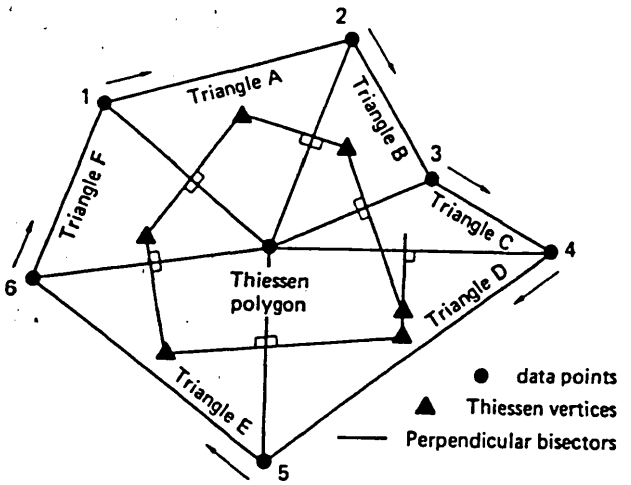


Figure 9. Construction of the Thiessen polygon

In spite of the several shortcomings or disadvantages of the grid-based terrain models discussed above, they are still widely used. Furthermore the development of progressively sampled data sets using computer controlled photogrammetric instruments has ensured that the principal disadvantage of grid-based models of lack of variation in sampling density with terrain type has been overcome to a considerable extent. Also the incorporation of terrain breaklines in program packages such as Hifi and Scop has given further credibility or life to grid-based terrain modelling methods especially when based on photogrammetric data.

Triangulation-based terrain modelling

This particular method is being used to an ever increasing extent in terrain modelling and many recently devised program packages are based on this method. The reasons for this development are that every measured data point is being used and honoured directly since they form the vertices of the triangles used to model the terrain, to determine the heights of additional points by interpolation and to carry out the construction of contours. Furthermore, the use of triangles offers a relatively easy way of incorporating break-lines, fault-lines and so on.

Originally the triangulation method suffered from the fact that it proved difficult to ensure that the same network of triangles would be generated from a single set of randomly-located measured data points no matter from which point in the data set the triangulation started. Also automatic triangulation often took an exorbitant time to execute in a computer. However these problems have been overcome (see Figure 8), hence the present popularity of the triangle-based modelling of the terrain.

Any triangular-based approach should attempt to produce a unique set of triangles that are as equilateral as possible and with minimum side lengths[12]. One or other of two main algorithms are used to implement these requirements. These are:

- the Delaunay triangulation method
- the radial sweep algorithm

Delaunay triangulation

Associated with Delaunay triangulation[13] is the Thiessen polygon (Figure 9), which endeavours to define geometrically the region of influence of a point on an areal basis. This is done by constructing a series of perpendicular bisectors on each of the triangles formed around that specific point. These intersect at the Thiessen vertices. The polygon so defined is the Thiessen polygon. Because of their distinct pattern when viewed on a computer graphics terminal, they are often referred to as tiles. Furthermore the data points surrounding a specific data point (e.g. 0) are known as its Thiessen neighbours.

A preliminary step before triangulation begins is to define a set of (artificial) boundary points to form a perimeter around the edges of the data set area. This is necessary to create a frame to the terrain model and a set of boundary triangles which allow contours to be extrapolated outside the area of the data set itself. Once these boundary points (which may have arbitrary values) are defined and have been added to the data set, the whole area can then be triangulated starting with a pair of the artificial boundary points known as the initial known neighbours (A and B, located in the bottom left hand corner of the area illustrated in Figure 10).

The search for the next neighbour is then made by constructing a circle with the base AB as diameter and searching to the right (i.e. clockwise) to find if any point falls within this circle (see Figure 11). This search can be carried out quite quickly using the computer. If no data point lies within the circle, it is increased in size to perhaps twice the area of the original circle with AB now a chord

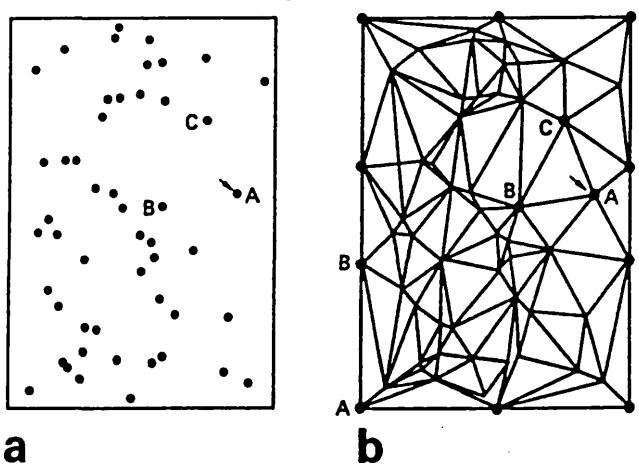


Figure 10. Delaunay triangulation[14] (a) data points, (b) triangulated surface and boundary points

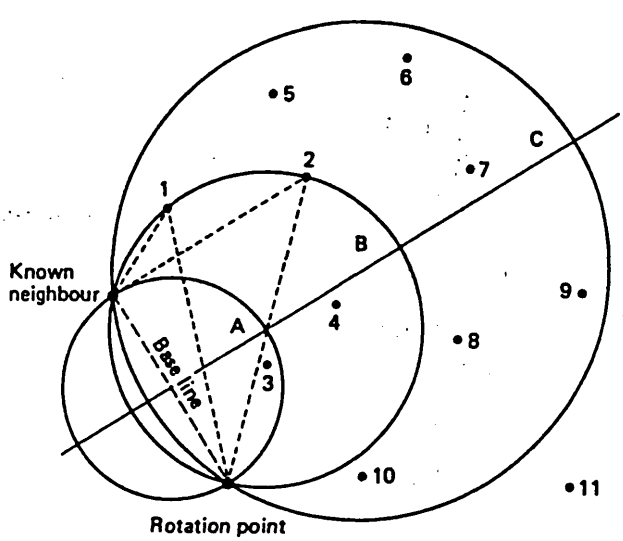


Figure 11. Expanding circle search for nearest neighbour[11]

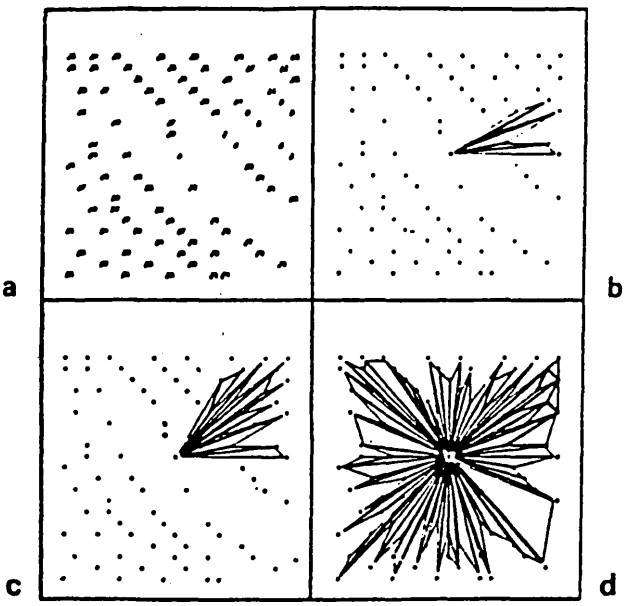


Figure 12. Sorting and triangle generation (a) nodes sorted by heading, distance, and pitch, (b) beginning of radial sweep, (c) radial sweep past two points with same heading, (d) beginning of concavity infill[14]

in the larger circle (see Figure 11). Any data points lying within the new circle are tested to discover which meets the criteria set for the nearest Thiessen neighbour.

Once this has been achieved, the search for the next neighbour then continues to the right, (i.e. clockwise). Then onto the next neighbour and so on till the next boundary point is reached. The triangles so formed constitute a so-called shell. The process of the triangulation then continues with each point in the shell being used in turn as the starting point for the search until the next set of Thiessen neighbours. This continues in a systematic manner until the neighbours for all the points existing in the data set have been found and the corresponding triangles formed as illustrated by Figure 8.

Now that it is better known and understood, the Delaunay method is the one used to form triangles in the majority of terrain modelling packages based on the triangulation method.

Radial sweep algorithm

The radial sweep algorithm is an alternative to the Delaunay triangulation which was devised and first published by Mirante and Weingarten in 1982[14].

As before, the input data is in the form of randomly located (as distinct from systematically located) points with x, y and z coordinates. The points will have been located on summits, on ridges and breaklines, etc. in the usual way. The point or node which is located nearest to the centroid of the data is selected as the starting point for the triangulation. From this central point, the distance and bearing to all the other points in the data set is calculated and the points are then sorted and placed in order by bearing.

Once this has been done, the radiating line to each point is established and a long thin triangle formed by connecting a line between the new point and the previous point (see Figure 12). If two points have the same bearing, then they are used to form a pair of triangles on each side of the common line. As each point is accessed, it is added to a linked list which forms the other boundary of the triangulated network. Once the initial sweep has been achieved, the concavities created by the initial radial sweep triangulation must be filled by new triangles. Each point or node on the boundary list is combined and compared with the next two nodes on the list and checked to see if an inside triangle can be formed. If so, then the new triangle is added to the database and the second (inside) node is removed from the list of outer boundary points. After the process has been completed, the list will comprise the points or nodes forming the convex edge of the terrain model. As can be seen from Figure 13, all the data points are now triangulated with nonoverlapping triangles. However, the shapes and connections are far from desirable. To optimize the shapes, each triangle is now tested against each of its neighbours.

A quadrilateral is formed by a pair of triangles. This is tested by calculation of the two distances by the opposite pairs of nodal points. If the distance between the two common points is greater than the distance between the two unique nodes, then the triangle indices are switched and the database pointers updated. This process is repeated with successive passes through the database till an entire pass through the database produces no changes.

Contouring

Considering a single grid cell as illustrated by Figure 14, a simple linear interpolation is carried out along each of

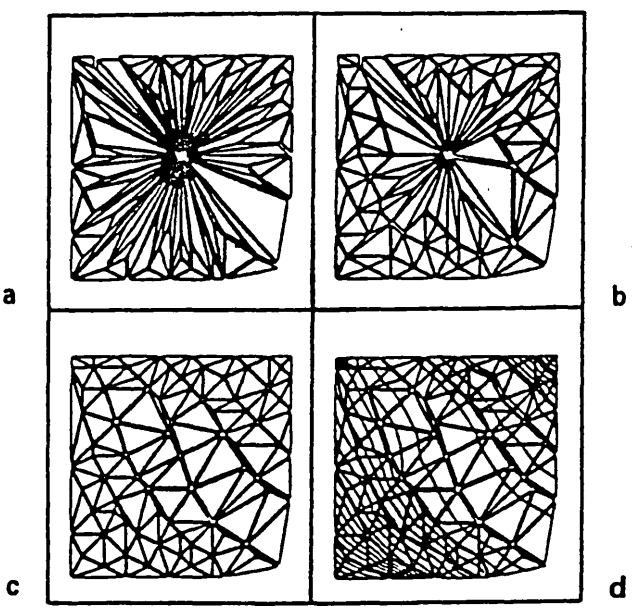


Figure 13. Triangle switching and contouring (a) mesh before first switching pass, (b) after first switching pass, (c) after all switching passes, (d) with contour cuts[13]

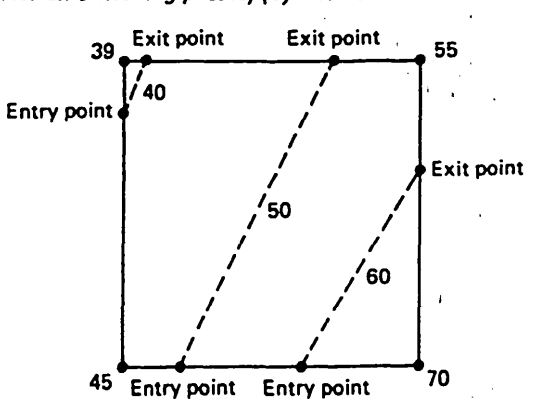


Figure 14. Linear contour interpolation in a single cell

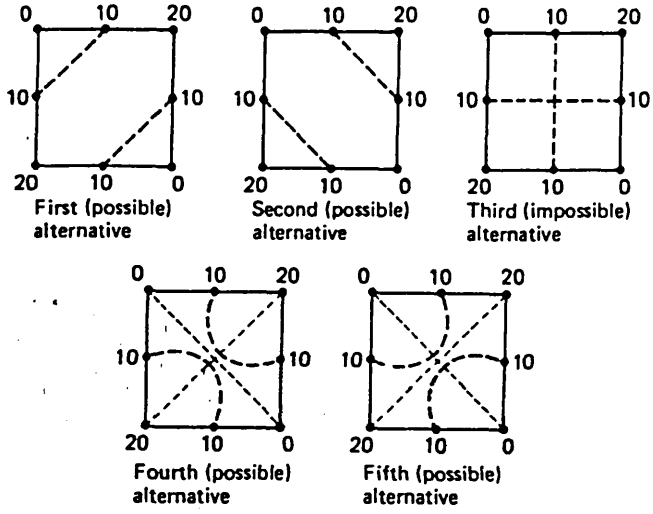


Figure 15. Ambiguity in contour threading in a single cell

the four sides in turn, based on the values at the nodes. The positions of all the contour values are determined for each side. Next they are connected up – by straight lines or vectors, since for every entry point there must also be an exit point.

However, ambiguities may also occur with alternative solutions and impossible situations. Taking the data points given in the example in Figure 15 there are two possible solutions which give quite different positions for the contour and a third (impossible) alternative. A solution to the problem is to revert to a centre point figure, i.e. splitting the grid cell into four triangles, assigning the average of the four grid nodes to the central point and then implementing an arbitrary rule such as 'keep the high ground to the right of the contour'. This is, of course, quite an arbitrary solution to overcome the difficulty.

In view of these potential difficulties, in many contouring packages based on regularly gridded data, no attempt is made to carry out direct threading of the contour. Instead, the interpolation of the contour is based on some type of function fitting, e.g. typically using the bicubic polynomial already discussed in the context of random-to-grid interpolation. Thus a new patch is formed based on a group of grid cells. If the full 16 term bicubic polynomial is employed, then the minimum size of patch that can be used is one made up of 4 x 4 = 16 nodes with 9 cells (see Figure 16). If a smaller number of coefficients are used, e.g. as in the 9 term biquadratic polynomial, then there will be redundancy and a least squares solution will be employed in the determination of the parameters. This can help to give extra smoothing and a better fit or continuity between cells. The interpolation of the contour through an individual grid cell is then carried out with reference to the whole patch rather than by simple linear interpolation in each individual grid cell.

Thus, the values of the coefficients $a_0, a_1, a_2, \ldots$ of the local surface patch are determined using the 16 grid node values first. In some contouring packages, e.g. Calcomp's GPCP (General Purpose Contouring Package), additional height values are then determined using the known values of the coefficients for all the points located on a subgrid which is generated within each grid cell. The contours are then interpolated linearly and threaded between the points on the subgrid cell.

The length of each of the straight line vectors will be quite short in each case, and the quite abrupt changes in direction which would be seen with contours derived from linear interpolation on the whole cell are lessened considerably, thus improving the visual appearance of the contours as seen by the user. It is possible to further divide the subgrid cell itself into centre point triangles and conduct the contour threading on the cutting points of the diagonals, in which case, the vectors will be still shorter and the appearance even more acceptable, see Figure 17.

Quite a number of contour packages using grids try to get more pleasing contours by an alternative approach, contour smoothing, which involves fitting a series of cubic splines between the entry and exit points. This produces smooth contours that are visually more acceptable than the straight line type produced from the same data. However, a separate curve or spline will be fitted to each segment of each contour line to ensure its smoothness. A difficulty which can then arise in an area of steep terrain with close contouring is that the curved smooth contours may then cross, which of course is not permissible in terms of actual terrain.

Contours from triangulated data

As with contouring of regular gridded height points, so with randomly located triangulated height data, there are two main options for the contour threading:

- simple linear interpolation of the contours
- generation of curved smoothed contours using some type of function

Unlike the situation with gridded data, the use of direct linear interpolation for the contour generation is very common when the terrain model is based on triangulated data. The potential difficulty with ambiguities regarding the direction that the contours might take is either not

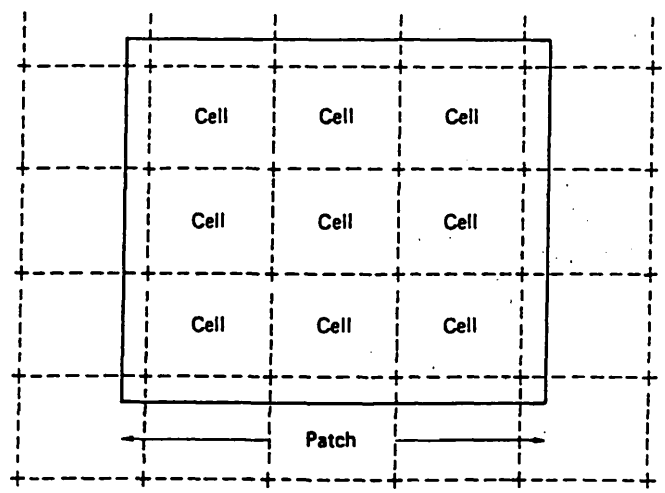


Figure 16. Patch formed by 16 grid nodes (nine grid cells)

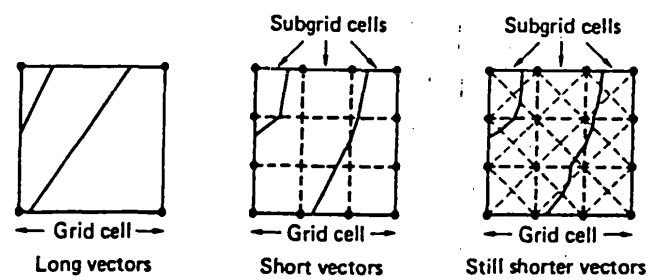


Figure 17. Contour threading through grid cells, subgrid cells and centre point triangles

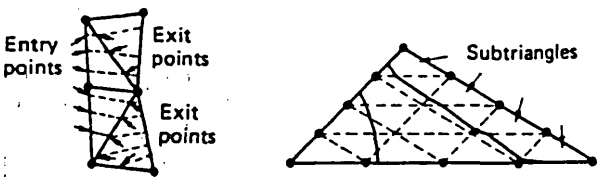


Figure 18. Contour threading through triangles and subtriangles

present or can be resolved. So direct linear interpolation gives a simple and robust solution. Contour threading usually begins at the boundary triangles. All the entry points along the perimeter are located using linear interpolation and the corresponding exit points found on the interior sides of each boundary triangle. These then act as entry points for the next internal triangle.

Just as with the grid-based method, so in the triangle-based method, the area of an individual triangle can be subdivided into smaller triangles (subtriangles) to get over the difficulty of long vectors with abrupt changes of direction along the common line between two adjacent triangles. The values at the vertices of each subtriangle can be calculated by simple linear interpolation. Then the cutting points are determined for each required contour along the sides of each subtriangle. Joining up the cutting points gives the required contour; again this looks smoother since it consists of a series of short vectors, instead of long vectors spanning the whole triangle. Obviously the threading decisions in any subtriangle are also simplified to a choice between the two remaining sides of the triangle as against the three possible sides within the grid cell (see Figure 18).

If curved smooth contours are required, then, as before, they can be achieved using a series of cubic splines or polynomials fitted through the string of interpolated cutting points along the triangle boundaries. Again in areas of very close contours in steep areas, crossing contours may occur if this procedure is adopted.

The alternative approach is again to fit some type of curved 3D surface patch to each triangle. This ensures a smooth transiton from one triangle to the next instead of having a series of planar triangular facets with abrupt changes in the direction of contours between facets along the lines common to adjacent triangles. This will entail forming a polygon patch using the central point and its Thiessen neighbours (see Figure 19). There is no difficulty in generating the polygon patch, but it will be irregular in shape (since there is no regular grid) so the data points will have a rather intractable format that is not easy to handle. The distances from the central point will all be unequal and the fitting function could be weighted inversely with increasing distance. So an inverse distance-weighted, low-order polynomial surface could be fitted through those irregularly located data points that comprise the patch.

Terrain model data derived from contours

The present discussion has concentrated till now on the situation most commonly encountered in surveying and civil engineering, namely that the individual points comprising the digital terrain model have been measured either by land surveying or by photogrammetric methods. The contours (and any perspective views) representing the model surface are then constructed or derived from measured spot height data.

However, the reverse situation may also occur, in which the terrain model comprising a set of individual points is formed from existing contours already recorded on a topographic map. This situation is not likely to be encountered at large scales where the accuracy requirements of the terrain model for road, dam design and construction, volume determination, etc. are such that the spot height data must be directly measured to a high degree of accuracy. However, in other applications where a wide area has to be modelled or represented at small scales, e.g. in landscape visualization, aircraft simulators, landform or geological representation, geophysical work etc., the situation is quite different. Accuracy requirements are much lower and less stringent and data has to be acquired for a large terrain area. In such situations, the terrain modelling will often be conducted on the basis of data derived from contours on existing maps.

Since in the first plane, the contour data represented on the topographic map will be less accurate than the measured spot heights produced via surveying or photogrammetric procedures, it follows that the individual spot heights derived by interpolation from such contours to form a terrain model will be still less accurate. Thus, although the appearance of the terrain data derived from contours is the same, its quality in terms of accuracy is decidedly lower. As long as this limitation is realised by and is acceptable to the users, then terrain models may readily be constructed from contours. The contours themselves will be measured by one of three methods:

- using manual line-following techniques
- using semi-automatic line-following methods
- through raster-scan digitizing of the contours present on a complete map sheet

Manual digitizing is normally carried out using a solid-state digitizing tablet equipped with a measuring cursor with which the operator follows the contour either in point

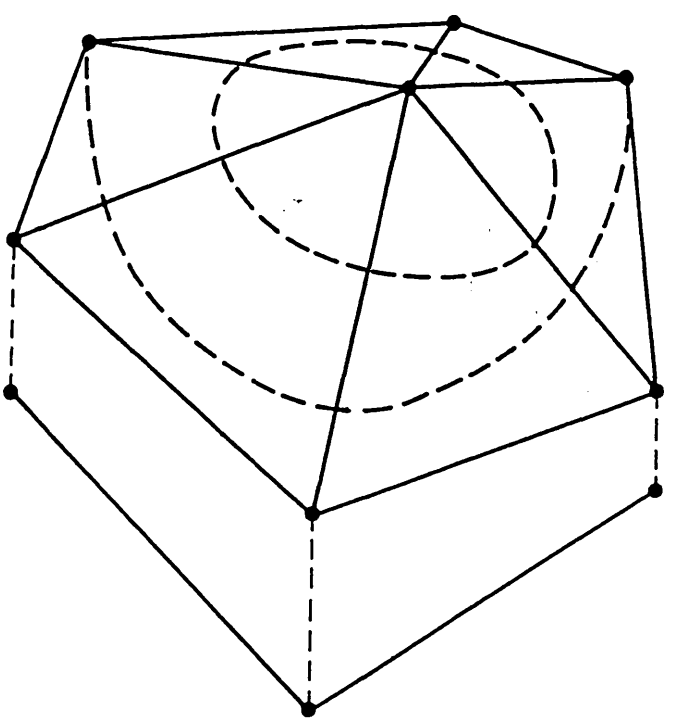


Figure 19. Surface and contours on a centre polygon that comprises several triangles

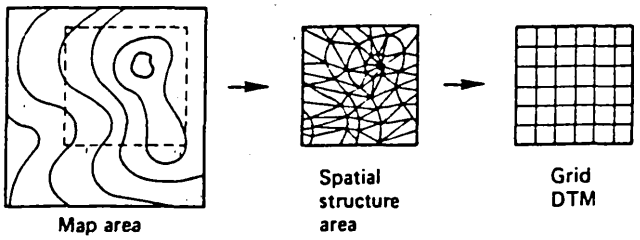


Figure 20. Contours to spot height data[11]

mode or in stream mode. Blind digitizing with direct collection and recording of the data onto a storage medium such as magnetic tape has been the traditional method used in the past. However, with the advent of comparatively inexpensive microcomputers and graphics workstations, the online display and editing of the digitized contour information is now quite common.

Semi-automatic line-following of contours is practised principally through the Fastrak digitizing system produced by Laser-Scan of Cambridge. A laser beam steered by a computer-controlled mirror is used to track each contour at high speed with direct display of the digitized lines and the possibility of operator intervention and interactive ending if the line-following should falter or fail.

Raster-scanning digitizers are fully automatic devices capable of rapid scanning and production of raw coordinate data for all the contours present in an individual map sheet. Both drum and flat-bed raster scanners exist and may be used for this purpose. The advantage of using such devices or systems is the sheer speed of the digitizing process, the time taken being solely a function of the size of the map sheet and quite independent of the length of the contour line to be digitized. However the time and expense needed to convert the contour data from the scanned raster format to the vector format required for data compaction and display are major drawbacks to this method.

The enormous capital investment required to purchase and operate either semi-automatic line-following or full-automatic raster-scan digitizing systems ensures that they are only acquired and operated by very large topographic mapping agencies which have huge programs involving the digitizing of complete national map series. Users with smaller requirements may obtain access to such devices through the bureau services offered by the system manufacturers. For most users, however, manual digitizing is still the principal method of digitizing contour information.

Once the contour lines have been measured and converted to digital form, the usual procedure for deriving, or forming terrain model data, is first to select points at appropriate intervals along the contours using a suitable algorithm implemented as a computer program. These individual points located on the contours are then treated as the randomly located height points or nodes in a triangular network (see Figure 20) with the obvious constraint that the triangles so formed must not cross the contour lines (which are treated as strings). From this triangular network of spot heights, the height values of the points or nodes in a regular grid may then be interpolated[11].

Currently, due to the work involved in digitizing contours, this technique is usually implemented only for individual map sheets or parts of map sheets. One may expect however, as time goes on, that digitized contours and the terrain model data derived from them will become available from national mapping agencies on a regional or national basis, in which case this would undoubtedly be regarded as the primary source of terrain model data at small scales and for large areas. For large scales and small areas, the primary source of terrain data for use in surveying and civil engineering applications of terrain modelling is likely to remain directly measured spot height data, in which case, the grid and triangular based models will be implemented as discussed earlier.

CHARACTERISTICS OF TERRAIN MODELLING AND CONTOURING PACKAGES

The discussion so far has focused on the alternative procedures and algorithms for both the regular grid- and triangular-based techniques, together with the various methods employed for the generation of contours. The aim of this section is to consider how these techniques are used by a selection of the software packages currently available.

A number of previous papers have discussed the various methods used for the formation of DTMs and also the interpolation procedures which can be adopted[15-17]. Inevitably however, these papers do not report some of the most recent developments in this field. Indeed the published literature has increased considerably in the past decade. To provide a framework for the analysis and discussion of the range of program packages currently available, the classification scheme illustrated by Figure 21 will be used.

General purpose modelling/contouring packages

The most general, and often least flexible approach to adopt for terrain modelling is to use one of the many general purpose modelling/contouring packages. Most of these were written in the early 1970s. Almost all are essentially grid-based and usually contain random-to-grid routines to convert nongrid data to the required grid format. Most organizations that possess a large computer installation will normally find that this software is readily available. These packages can be used to model any regularly or irregularly distributed scientific dataset. The characteristics of some of the most common grid-based packages are illustrated by Table 2.

As previously mentioned, any one of three interpolation procedures can be used if randomly positioned data is being used:

- pointwise
  - involving the determination of different functional parameters for each point, using either the distance weighted average, paraboloid or projected weighted average technique

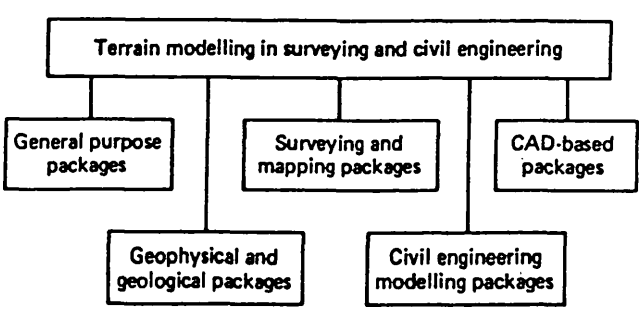


Figure 21. Classification of terrain modelling software packages in surveying and civil engineering

Table 2. Comparison of the interpolation and contouring procedures for a selected number of general purpose modelling/contouring packages (adapted from Grassie[18])

Name of package	Method of interpolation from random points to grid nodes	Method of contouring
Ghost	Pointwise (distance weighted average)	Triangle (diagonal method)
Surface II Graphics	Pointwise (distance weighted average)	Rectangle (linear)
Ginosurf	Pointwise (paraboloid)	Triangle (centre point method)
GPCP (General Purpose Contouring Package)	Pointwise (projected weighted average)	Polynomial (bicubic)
SACM (Surface Approximation and Contouring Mapping)	Patchwise	Rectangle
Surface II Graphics (Trend)	Global (up to 6th order)	Rectangle (linear)
GPCP (Trend)	Global (up to 6th order)	Polynomial (bicubic)

- patchwise
  - involving the establishment of a series of local surfaces
- global
  - based on a single polynomial function (including up to 10th order terms)

Similarly, various contouring functions can be used to thread contours through the interpolated grid of data points. An evaluation of the accuracy of some of these general purpose packages for cartographic applications has been presented by Grassie[18]. Grassie carried out an empirical study using various topographic data sets. Using contours measured photogrammetrically as a reference, he compared the accuracies obtained using different packages. Triangulation methods produced consistently better results than random-to-grid methods of which the patchwise and global techniques often produced poor results. Gridded or evenly distributed data sets were more satisfactory than poorly distributed linear data.

Surveying and mapping modelling/contouring packages

While the packages outlined in the previous section were designed to handle digital data from a large number of different technical fields, there are also several program packages which have been designed specifically for handling topographic and terrain data.

Height interpolation by finite elements (Hifi)

Hifi[5,6] is a suite of programs developed at the Technical University of Munich, which is available commercially from Carl Zeiss (Oberkochen) as part of the software associated with their Planicomp C100 analytical plotter.

The finite element method has been applied to many problems in civil and mechanical engineering, but to date it has not been used extensively in the surveying and mapping industry. In essence however, Hifi is a grid-based package and was developed especially to deal with the grid-based data sets generated by photogrammetric methods, e.g. during profile scanning for orthophotograph production and under computer control using an analytical plotter. Thus in some respects it is not too different from the general purpose modelling packages described earlier. If nongridded data is being used it performs firstly a random-to-grid interpolation, followed by contouring based on the interpolated heights at the grid nodes. In contrast to the pointwise methods used by most of the programs discussed in the previous section, it uses a patchwise interpolation technique. Furthermore, it is able to include the effects of breaklines which are important in terrain modelling, but are not generally included in the general purpose programs. Breaklines are terrain discontinuities (such as ridges and valley floors) measured by the surveyor or photogrammetrist. The inclusion of such data prevents a model forming, for example, across a valley as illustrated by Figure 22(a). Two versions of Hifi are available: the first uses bicubic finite elements, or patches, whereas the second, which allows for the inclusion of breaklines, uses bilinear elements. In areas where breaklines are not considered necessary, the use of bicubic patches enables a significant reduction in computational effort to be realised. The form of the two surfaces is illustrated by Figure 22. The software is written in Fortran and is designed to run on Hewlett-Packard, DEC and Prime hardware.

Contour Interpolation Program (CIP)

CIP[19] is the current terrain modelling/contouring package from Wild Heerbrugg (previously they offered the Scop package). It is written in Fortran and is designed to run

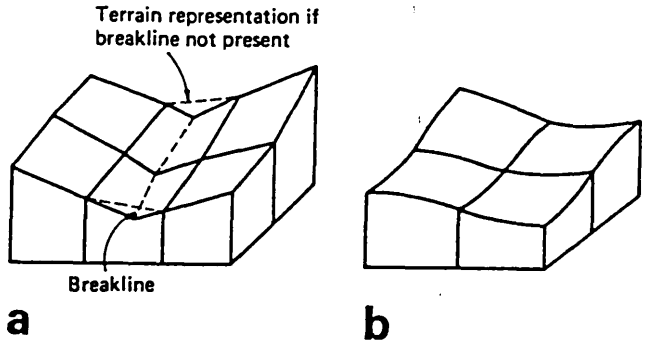


Figure 22. Hifi interpolation surfaces (adapted from Ebner et al.[6]) (a) bilinear finite elements, (b) bicubic finite elements

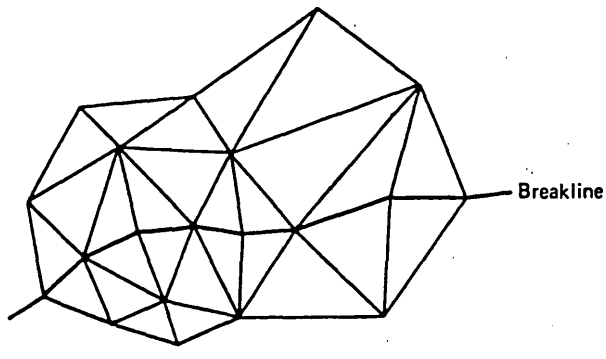


Figure 23. CIP triangular mesh, including breaklines[19]



Figure 24. CPS-1 contour plots (courtesy of Radian Corporation)

on the Data General Nova series (used by the Wild AC-1 and BC-1 analytical plotters), DEC VAX 11 series (used by the Wildmap system) and the Sun-3 workstations (used by the new Wild System 9 digital mapping and geographic information system[20]).

In contrast to Hifi, the program is based on the use of triangles as the finite elements or patches (see Figure 23). The triangular method fits the measured data exactly and it also lends itself to the incorporation of breaklines.

The interpolation procedure involves the use of a 10 term cubic polynomial which forms a smooth surface across each triangular patch. Continuity is ensured at the boundaries, except across breaklines. Contouring is carried out using 16 subtriangles. The cutting points along the edges of the subtriangles are then determined. Other options include facilities for generating the heights of single points, a regular grid DTM, profiles, perspective views and the volume relative to a reference surface.

MBB-DGM

A further package based on triangles and very similar to that developed by Wild Heerbrugg has been written by Messerschmitt-Bolkow-Blohm (MBB), as part of their Integrated Civil Engineering System (ICES)[21]. Again, a series of triangular patches are used to represent the terrain and depending on the degree of accuracy required, a linear or cubic model is used. Various output options are available including contours, profiles, perspective views and volume calculations.

Intergraph

The US systems' supplier Intergraph has developed several terrain modelling packages, generally as part of an integrated digital mapping system. One particular system, developed in 1982 for the US Defense Mapping Agency (DMA), specifically for viewing topographic surfaces and mapping operations is the Terrain Edit System/Elevation Matrix Processing System (TES/EMPS). The hardware installation, probably one of the largest in the world dedicated to terrain modelling, consists of two Intergraph 780 systems each equipped with an array processor connected to 12 graphics workstations, with twin 1280 x 1024 pixel high resolution displays.

Data can be input in either grid, random point or string (contour) form. In the latter two instances the data are reformatted into a grid or matrix of elevations. Input in vector or raster form is also possible with the facility to convert between both forms of data structure. Up to eight scenes of the terrain can be viewed at any one instant together with full interactive editing of the data.

Intergraph has also developed a digital terrain modelling nucleus (DTMN) as part of its civilian mapping system. Again various forms of data input can be accommodated including random point data from field instruments, grid data from photogrammetric stereoplotters and vector or raster cartographic data acquired from contours. A combination of the Interactive Coordinate Geometry (Cogo) System (ICS) with the DTMN enables users to carry out horizontal and vertical road design and site planning operations.

Geological and geophysical modelling/ contouring packages

While general purpose modelling/contouring packages can be used for examining geological and geophysical phenomena, they are limited in their ability to deal with surface discontinuities such as geological faults. Consequently, a number of packages have been developed to specifically deal with problems of this type.

Contour Plotting System (CPS)

CPS is a widely established modelling and contouring package which was originally developed in 1972 by Unitech and is now available from the Radian Corporation. It is written in Fortran and was developed, primarily, for users in the oil, gas and mining industries. It is available in three

forms: CPS-1, which is designed to operate on mainframe and superminicomputers, CPS/PC which operates on the IBM PC XT and AT range of personal computers and CPS-1/G, an interactive version for the display and editing of projects that have been run on CPS-1.

Unlike the majority of the packages discussed so far which are generally restricted to one or two methods of interpolation, CPS-1 offers users a choice of various interpolation algorithms. These include global trend surfaces using polynomials and local interpolation by a number of methods. However, the major benefit of packages such as CPS is their ability to handle geological faults and other surface discontinuities. Figure 24(a) represents a contour plot that models a complex faulted region. Other features of CPS-1 include the facility to generate 'orthocontour maps' in which regions of maximum gradient are illustrated (see Figure 24(b)). This feature is of particular use in water runoff and oil migration studies. 3D raised contour plots can also be created (see Figure 25(a)).

Unimap

Another package that has been used for the depiction of geological and geophysical surfaces is Unimap. Unimap and its subsidiary products Geoint and Geopak enable the user to access a range of interpolation procedures for the creation of data sets from gridded input, the formation of gridded data sets from randomly distributed observations, and for the production of contoured maps. Interpolation can be of a global nature, patchwise using bicubic polynomials and local (pointwise) using distance weighted average methods. The ability to deal with faults is also an element of the package. Figure 25(b) illustrates a perspective contour map produced using this package.

Further packages

In addition to the packages mentioned, geological and geophysical modelling and contouring has also been performed using the Panacea system[11,12]. Another package that is in common use is Z-Map from the Zycor Corporation.

For some users the opportunity to tailor a suite of programs to meet the inhouse requirements of an organization may be important. Thus, in addition to the packages already mentioned, several libraries of subroutines are also available. Two examples of these libraries are those available from Precision Visuals (DI-3000 subroutines) and from Northwest Digital Research (Interactive Surface Modelling, ISM). The latter, available in Fortran 77 and Pascal has been used for landfill operations and is discussed by Garvis and Gee[22].

Civil engineering modelling, design and drafting packages

Civil engineers have been using digital terrain model data for commercial operations since the mid 1960s. The early systems, developed primarily for highway design, enabled the engineer to replace the traditional manual design process by an automated system. Thus, several different road alignments could be examined within a narrow corridor of interest, a process which was not practical before the advent of these computer-aided highway design (CAHD) packages. One of the most popular of these early systems was the British Integrated Program System (BIPS) for highway design, which was primarily grid-based. By producing links to other suites of programs such as the Highway Optimization Program System (HOPS) for cost minimization[23], the engineer had at his disposal a powerful design tool.

These early systems were however limited in several respects:

- restricted terrain modelling capabilities, with data normally held in a square grid or cross-section form (consequently terrain definition was poor)
- inability to deal with complex geometrical conditions at junctions and interchanges

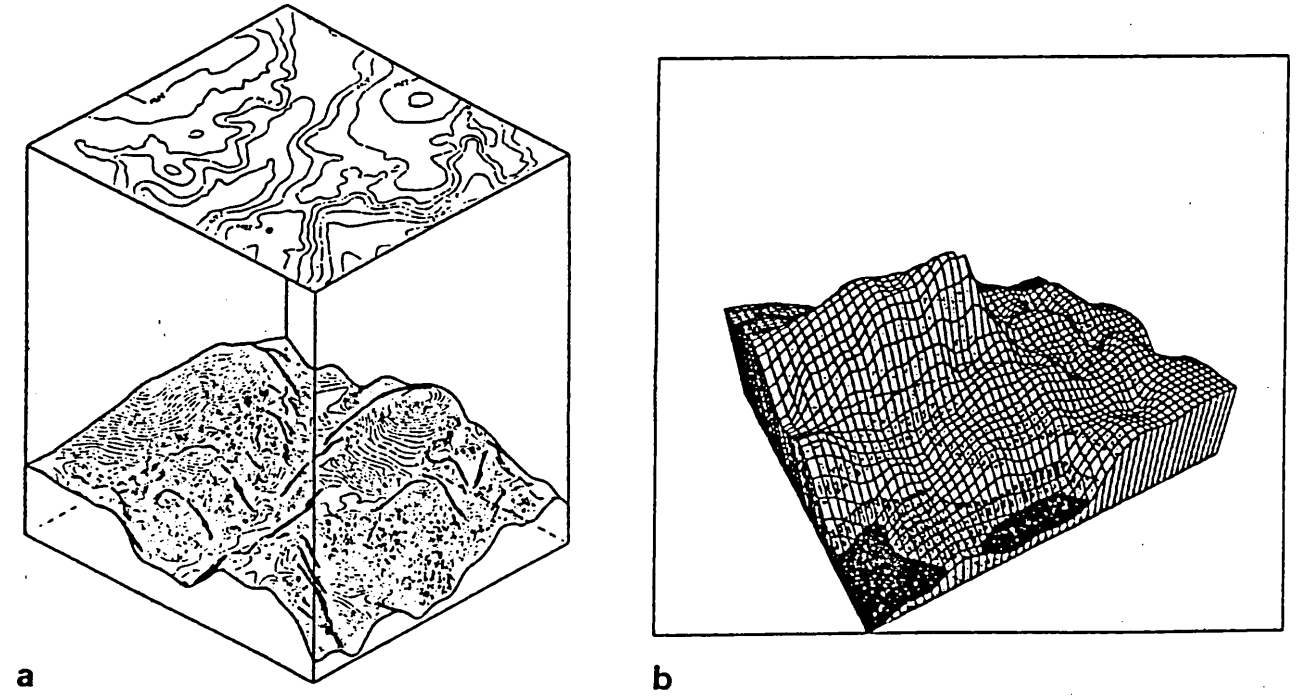


Figure 25. Three dimensional raised contour plots (a) raised contours with ceiling plot (courtesy of Radian Corporation), (b) perspective view of terrain surface with displayed contours (courtesy of UNIRAS)

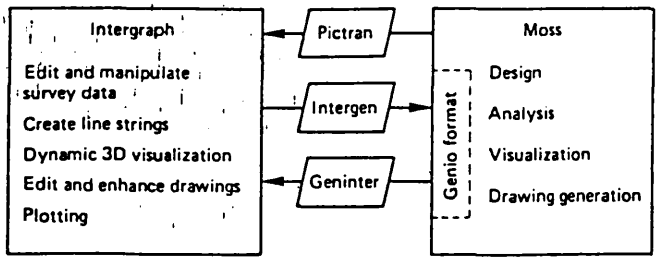


Figure 26. Intergraph/Moss interface[27]

- the terrain model and the civil engineering design could not be merged to produce a composite model
- limited drafting facilities (although automatic output could be generated, it was generally not possible to produce either cartographically acceptable drawings, or final contract design drawings, without considerable manual intervention)
- no interactive graphics capability for interrogating and editing the database

One of the first systems to address many of the limitations of these early systems was the Moss Modelling System[24-26].

Mainframe and minicomputer-based packages

Moss was developed from 1974 onwards by a consortium of UK County Councils (Durham, Northamptonshire and West Sussex) acting on behalf of the County Surveyors Society. The system continued to be developed by this consortium until 1983 when the privately owned Moss Systems Ltd was formed. Its use is widespread in the civil engineering industry and it is particularly dominant in local authorities and consulting engineers offices and has over 150 users worldwide.

The concepts of 'point' and 'string' data are fundamental to an understanding of the Moss System. Both survey and design data are considered to consist of either single points (such as lamp-posts and trees) or strings, which are a series of linked 3D points (kerb lines, contours, breaklines, etc). The use of this string data together with triangular modelling techniques enabled a much better terrain surface representation to be achieved than with cross-section or grid-based techniques. Furthermore, defining both the terrain and the design surface (e.g. road, railway, dam, quarry) in this way within a common database, overcame the problems of merging and complexity to a considerable degree. The drafting and interactive problems have also been overcome to a large extent with the advent of graphics workstations, such as those based on the Apollo and DEC MicroVAX range. With these a new range of interactive facilities termed Interactive Moss have been added to the current system. These facilities allow the user to create graphic windows and display all, or part, of a model on a graphics screen and by means of mouse driven screen menus to alter the way in which graphic information is drawn.

Other users have interfaced Moss to standard computer-aided drafting and design (CADD) systems, for example, to the Intergraph CADD system[27]. In this case, the interactive capabilities provided by the Interact graphics workstation enables the user to delete and modify text or model data. The interface programs (see Figure 26) include Intergen which transfers lines from the Intergraph format to a Moss general input/output file (Genio), Pictran which transfers Moss picture files to Intergraph and Geninter which transfers lines in Moss to the Intergraph system.

Visualization of the impact of new structures by using photomontage techniques[28] has been a feature of Moss for several years. A new development is the production of colour shaded perspective views using image processing techniques[29]. Systems such as Vosco from the CAD Centre and Vigil from the Transport and Road Research Laboratory (TRRL) enable a more realistic impression of the relationship between the terrain and the structure to be obtained. Furthermore, it is possible to simulate the effects of varying lighting and weather conditions (e.g. fog), and also to examine the visual intrusion caused by the proposed structure on the terrain.

Although Moss was initially developed for highway design projects, it has developed into a general purpose package for civil engineering modelling, design and drafting. It has been applied to a wide range of projects including the design of roads, railways, airports, dams, together with applications in land reclamation and mining operations. Further details on applications are presented in a later section.

Axis Surface Modelling System (SMS) is a similar modelling design and drafting package to Moss. Written in Fortran-77, the package has been installed on a variety of mainframes (IBM and ICL), minicomputers (DEC PDP 11/23) and engineering workstations (ICL Perq and Apollo Domain). Recently it has also been rewritten to run on the IBM PC XT, AT and Olivetti M24 range of microcomputers.

The facilities offered by the package are similar to those available from other manufacturers and include triangular modelling (using the radial sweep method), area and volume analysis, horizontal and vertical road design, together with various output options including perspective viewing. The system can be linked to a separate mapping package. Figure 27 illustrates an isometric projection of sections extracted from a typical model.

Closely associated with the SMS package is the Axis Mapping System (AMS) which, as the name suggests, is an interactive digital mapping system designed to handle field survey, photogrammetric and cartographic data specifically for topographic and utility mapping purposes.

Road CAD is a terrain modelling, road design and drafting system developed by Sito, a firm of Finnish consulting engineers. Unlike the other systems described in this section, it is designed specifically for road design and setting out operations rather than being a more general purpose civil engineering package. The package is designed to run on a Data General Eclipse MV/6000 minicomputer.

Packages of this type are generally not as attractive to potential users as some of the systems outlined previously for several reasons. First, they do not offer operational flexibility since they are dedicated to one single task, road design. Second, the package is specific to one particular

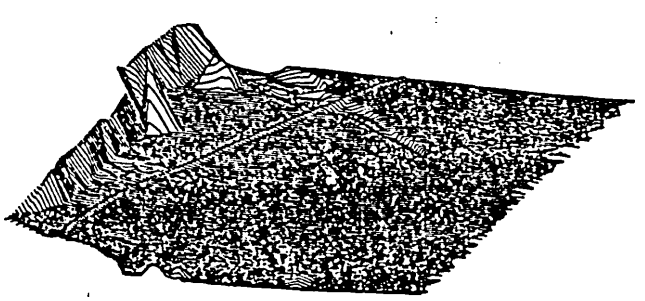


Figure 27. Isometric view of terrain model (courtesy of AXIS-SMS Ltd)

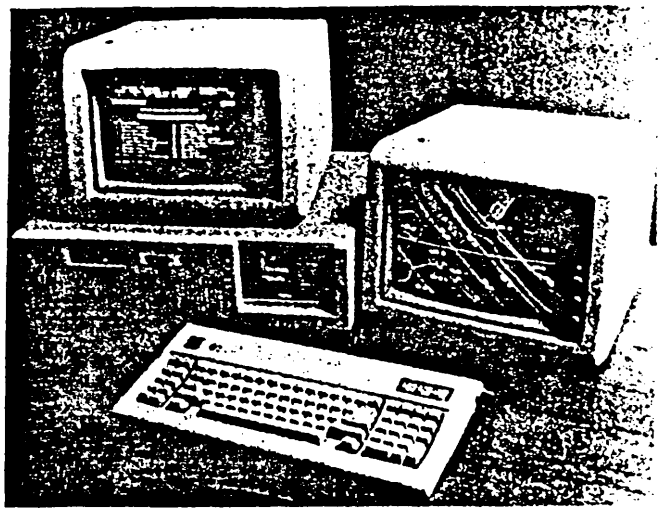


Figure 28. Eclipse package on the IBM PC AT (courtesy of Eclipse Associates Ltd)

hardware configuration. Third, given the conflicting demands of production activities, the backup and upgrading of software may not be as good as from a well known software supplier specializing in this field. Conversely, if an organization specializes in highway engineering activities, then a package of this type may be appropriate.

Personal computer- and desktop computer-based packages

During the past five years or so a variety of terrain modelling packages have become available on PC and desktop computers. Some of the most common systems currently in use in the UK are listed below.

Eclipse (Entire Computer Loaded Integrated Program for Surveyors and Engineers) was initially developed in the late 1970s as a workstation system based on the Wang 2200 series of desktop computers. Recently however, the system has also been installed on a variety of personal computers, e.g. the Wang PC and APC, together with the IBM PC XT and AT[30], and as illustrated in Figure 28. Like Moss it is widely used, particularly for surveying operations, and has over 100 users worldwide.

Eclipse is an interactive package, written largely in Basic, which enables the user to input survey data from a wide variety of sources (field data recorders, stereoplotters and digitizers), process the data using standard surveying computational techniques and create a 3D database. This database can subsequently be used for the creation of digital mapping data or be integrated with road, building layout or drainage design information. Triangular terrain models can be created and area and volume reports be generated. Output can be cartographically enhanced and models can be transformed to Moss, GDS, Rucaps (an architectural package) and Autocad formats.

Hasp-Digical 11S consists of a series of program modules developed for surface modelling, volumetric calculations, road and building design, digital mapping and cadastral surveying[31].

The modules although originally written in HPL are now available in Pascal, are designed to run on the Hewlett Packard 200 and 300 series of desktop computers and the HP 9816 and 9825 series of personal computers. Modelling is carried out using the triangular irregular network technique using the radial sweep algorithm developed by Mirante and Weingarten[13]. Breaklines can be incorporated either at the data collection stage or during the data processing. A typical hardware configuration using the HP 200 series for processing is illustrated by Figure 29.

SURVPAK[32] developed by D J Herriott Ltd is another system which operates on the HP 200 and 300 series of desktop computers. The facilities of the system are very similar to those of the systems already mentioned above and the package is based on the principles of triangular modelling. However, one unique feature is the provision of a database facility specifically for pipeline laying operations.

Acsis is a suite of surveying and modelling programs designed to run on the Apricot PC and Xen computers. Again the facilities include triangular modelling, volume calculations and road design capabilities.

CADD-based modelling systems

The cost effective benefits of using CADD systems are being increasingly realised by the civil engineering industry. Many organizations now possess a CADD facility for the production of engineering drawings. For many of these users, the opportunity to use a terrain modelling package within the context of a more general purpose CADD system has some obvious advantages, particularly in terms of data compatibility and speed of data transfer.

Two distinct types of CADD system are available to the engineer, those based on surface modelling and those which form a solid model. Terrain modelling packages are currently available for both types of system.

Surface modelling

Surface modelling CADD systems are the most common type in use in the civil engineering industry. One particular company which has developed a terrain modelling capability within a conventional drafting system is McDonnell Douglas which has developed an interactive site terrain system (Sites) to operate within its General Drafting System (GDS).

GDS-Sites is a command driven drafting and modelling system which was launched in 1980 and enables the interactive creation, recall and editing of drawings to be carried out. Two complementary packages known as Survey and Sites have recently been added to the system for the creation and analysis of terrain data.

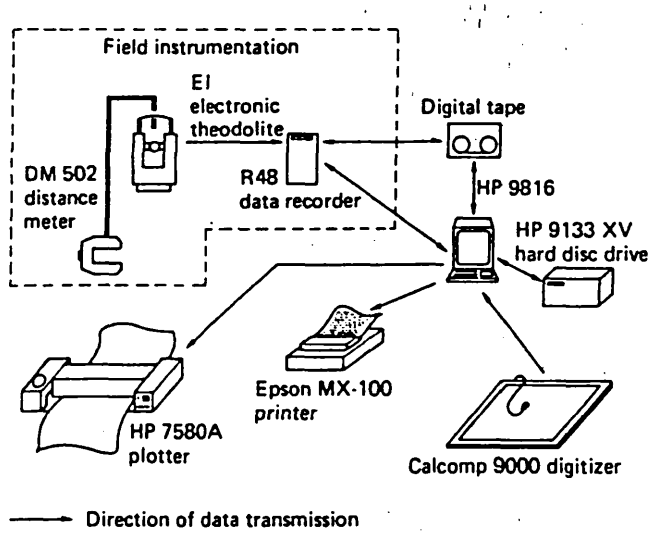


Figure 29. Hasp-Digical system flowchart

Survey[33] provides a means of entering raw field surveyed data into the modelling part of the system. Control data is processed by standard methods, an interesting feature being the inclusion of an interactive preanalysis package for the optimization of control networks. Survey details can be entered by various methods including data from field data recorders. One product of the use of the Survey package is a database of spot heights which forms the terrain model.

Sites[34] is a separate element of the system which carries out the interpolation and analysis of the model. The basic algorithm which are used were developed for an earlier package known as CIIS (Contouring, Interpolation and Integration of Surfaces) described by Grayer[35]. Essentially it consists of a modified Delaunay triangulation to allow for the inclusion of breaklines. Contours are determined by a linear interpolation along each triangular face; smoothed contours can also be created by fitting a quadratic B-spline through the points.

At present, the system does not include any design capabilities for road design, for example, although volumetric analyses can be carried out. One feature of particular interest to hydrologists is the facility to display watershed lines and to indicate, using cross slope vectors, the direction of runoff. This ability to extract ground geometry and interrogate the terrain for physical attributes, other than simple spot height information, has led to the coining of the term intelligent ground models (IGM) for Strodachs and Durrant[36] to describe systems which offer these additional features.

Acropolis is a further example of a surface modelling system which forms part of an integrated CADD system from BDP Computing Services. Although developed primarily for architectural operations, the system can also generate ground models for visualization purposes.

Solid modelling

Solid modelling systems have been developed primarily for the display and manipulation of mechanical engineering components. Systems of this type can be classified, according to Pratt[37] and Requicha and Voelcker[38], into constructive solid geometry (CSG), boundary representation (B-Rep) and simple sweep types. In the first case, shapes are represented in terms of simple volumetric primitives such as the cuboid, cylinder, cone, sphere and torus. The limitations of this type of modelling system for terrain modelling operations are discussed in Strodachs and Durrant[36]. The second approach, which involves the storage of all the elements which make up an object's boundary (the faces, edges and vertices together with topological information about relationships between elements) is currently under investigation by researchers at Loughborough University of Technology[36,39]. The system which they are examining is based on the Medusa CADD system developed by Cambridge Interactive Systems (CIS). The third type of solid modelling system has not been assessed for terrain modelling.

Medusa Ground Modelling (MGM) package is currently a development system. Facetting techniques are used by the system to develop a series of triangular elements (using the Delaunay triangulation) to form a solid model of the terrain. At present, however, the system does not allow for the inclusion of breaklines, although this has been suggested as an improvement for future versions of the package. Boolean or set operations can be invoked to enable new structures (such as roads and dams) to be superimposed onto the terrain model. From this composite model, volumetric data could subsequently be determined. Figure 30 illustrates the use of these techniques to simulate the effect of a proposed highway on an existing ground model. Although the system uses a different modelling technique, the final results are not too dissimilar from those obtained from many of the packages discussed earlier.

APPLICATIONS

The main impetus for the development of many of the packages outlined in the previous section, and in particular for design, drafting and modelling packages, was the demand for major highways during the late 1960s. However, since then, the use of these packages has diversified into many other fields of surveying, geology and civil and military engineering. The following list, while unlikely to be complete, provides an indication of the wide range of projects where terrain modelling methods have been used.

Surveying and mapping

Terrain models can be created both by field surveying and photogrammetric techniques and generally form the basis for the applications in geology, geophysics and civil engineering. In addition to the standard production of square grid, triangular or string terrain models for engineering projects, terrain models have also been used to:

- model the seabed from echo soundings, and positions obtained from electronic position fixing equipment
- rectify satellite remote sensing imagery (for example as created from the Landsat or Seasat series of satellites)
- enable the superimposition of satellite imagery onto a realistic 3D surface

Geology and geophysics

As mentioned in the section on geological and geophysical modelling packages, modelling systems have been developed to deal with the specific problems of representing heavily faulted geological structures and of modelling geophysical data. Some applications in this area include:

- oil reservoir simulation studies

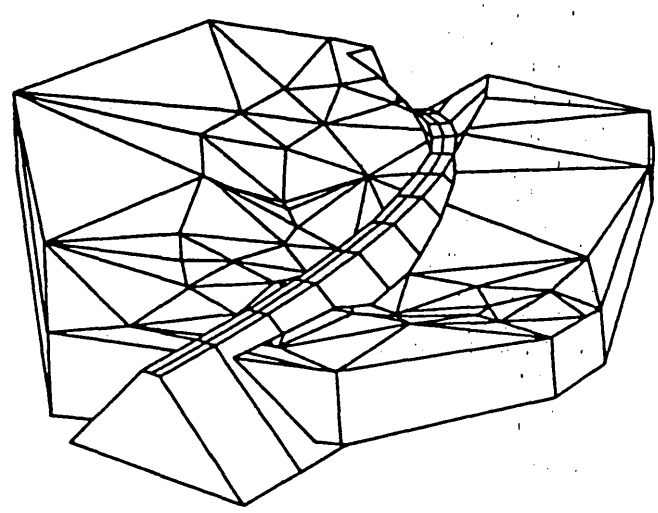


Figure 30. Simulated highway model created using the MGM[39]

- the determination of reservoir capacity of an oil field from geological and geophysical measurements
- the modelling of aeromagnetic and seismic data

Civil engineering

Highway engineering

Modelling systems can be used to solve a variety of highway engineering problems including:

- the measurement of earthworks from borrow pits, spoil areas or to produce volumes of different strata for precise costing
- the design of major roads
- the design of interchanges and junctions or complex urban redevelopment schemes
- the visualization of structural elements (for example bridge decks or concrete foundation slabs)

Other linear civil engineering projects

Many of the techniques developed for highway engineering projects can also be applied to other linear projects including:

- the design of major railways and rapid transit systems
- the design of canals and river retraining schemes
- airfield design
- monitoring and visualization of tunnels using tunnel profiling systems

Areal civil engineering projects

Large areal (defined space) projects can also benefit from the use of terrain modelling techniques. Some recent examples include:

- land reclamation and rehabilitation operations associated with mineral extraction, flattening of spoil tips or the filling of gravel pits
- catchment/watershed analysis to assess the storage capacity of reservoirs
- dam design and monitoring of tailings dams

Military engineering

In addition to the use of terrain modelling techniques for military engineering projects similar to those outlined above, the techniques have also been applied to solve specific military engineering problems including:

- the siting of ground based radar stations
- line of sight and intervisibility analyses for artillery planning
- analysis of exposure and fallout regions associated with nuclear, biological or chemical attack
- flight simulation over restricted airspace or ground corridors
- simulation of the performance of missile systems which use terrain comparison guidance techniques

CONCLUSIONS AND FUTURE TRENDS

Terrain modelling has been applied to a wide range of problems in the fields of surveying and mapping, geology and civil and military engineering. Various possibilities exist for the creation of DTMs ranging from field survey, aerial photogrammetry and direct digitizing of existing data sources such as topographic maps. Depending on the form of these input data, a variety of interpolation techniques can be used to create triangular or gridded data sets. Five distinct types of DTM package which have common characteristics are identified.

For the future, the use of terrain modelling methods will continue to develop and expand into other areas of activity. Although trends are difficult to identify and extrapolate from it seems likely that the creation of a national terrain model of the UK is a likely product for the future. Such a model could be created from existing 1:50 000 and 1:10 000 contour maps formatted into a grid, the grid dimensions depending on the variability of the terrain. While the technical problems of creating such a database would be formidable, a much greater problem would be the cost, revision and general management of such information. However, while such a derived model would be sufficiently accurate for general visualization, or flight simulation, it would be unlikely to meet the needs of surveyors and civil engineers. For such users, the primary source of terrain data is likely to continue to be directly measured spot heights and the creation of grid- and triangular-based models. Finally, if such data are to be widely used in industry it will require the cooperation and understanding of both the military and the public and private sectors.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the assistance given by many manufacturers and individuals. They have supplied a great deal of relevant and interesting material, much of which has had to be condensed considerably to fit within the scope of this review paper. Nevertheless, sincere thanks are due to all of those who have contributed information used in this paper.

REFERENCES

1 Miller, C and LaFlamme, R A 'The digital terrain model – theory and applications' *Photogrammetric Eng.* Vol 24 No 3 (1958) pp 433–442

2 Makarovic, B 'Progressive sampling for digital terrain models' *ITC J.* No 3 (1973) pp 397–416

3 Makarovic, B 'Amended strategy for progressive sampling' *ITC J.* No 1 (1975) pp 117–128

4 Makarovic, B 'Composite sampling for DTMs' *ITC J.* No 3 (1977) pp 406–433

5 Ebner, H, Hoffman-Wellenhof, B, Reiss, P and Steidler, F 'HIFI – a minicomputer program package for height interpolation by finite elements' *14th Congress of the Int. Soc. of Photogrammetry* Hamburg, FRG (1980)

6 Ebner, H and Reiss, P 'Experience with height interpolation by finite elements' *Photogrammetric Eng. Remote Sensing* Vol 50 No 2 (1984) pp 177–182

7 Assmus, E 'Extension of the Stuttgart contour program to treating terrain breaklines' *13th Int. Soc. of Photogrammetry (ISP) Congress* Helsinki, Finland (1976)

8 Stanger, W 'The Stuttgart Contour Program (SCOP) – further development and review of its applications'

Proc. 13th Int. Soc. Photogrammetry Congress Helsinki, Finland (1976)

9 Kostli, A and Wild, E 'A digital terrain model featuring varying grid size' *Contribution to the XVth ISPRS Congress.* Rio de Janeiro, Institute of Photogrammetry, University of Stuttgart (1984) pp 117–126

10 Huurnemann, G and Tempfli, K 'Composite sampling using a minicomputer supported analogue instrument' *ITC J.* No 2 (1984) pp 129–133

11 McCullagh, M J 'Transformation of contour strings to a rectangular grid based digital elevation model' *Proc. Euro-Carto II* (1983)

12 McCullagh, M J 'If you're all sitting comfortably then we'll begin' *Workshop Notes on Terrain Modelling* Australian Computing Society, Siren Systems (1983)

13 Delaunay, B 'Sur la sphere vide' *Bulletin of the Academy of Sciences of the USSR* (1934) pp 793–800

14 Mirante, A and Weingarten, N 'The radial sweep algorithm for constructing triangulated irregular networks' *IEEE Computer Graph. Appl.* Vol 2 No 3 (1982) pp 11–21

15 Grist, M 'Digital ground models – an account of recent research' *Photogrammetric Record* Vol 7 No 40 (1972) pp 424–441

16 Jancaitis, J R and Junkins, J L 'Modelling irregular surfaces' *Photogrammetric Eng.* Vol 39 No 4 (1973) pp 413–420

17 Schut, G H 'Review of interpolation methods for digital terrain models' *Proc. Congress of the Int. Soc. Photogrammetry* Helsinki, Finland (1973)

18 Grassie, D N D 'Contouring by computer: some observations' *British Cartographic Society* Special Publication 2 (1982) pp 93–116

19 Steidler, F, Zumofen, G and Haitzmann, A 'CIP – a program package for interpolation and plotting of digital height models' *Proc. ACSM-ASP Meeting* Washington DC, USA (reprinted by Wild Heerbrugg) (1984)

20 Sykes, D and Crane, S 'System 9 of the new concept from Wild Heerbrugg' *Civil Engineering Surveyor* Vol 11 No 3 (1986) pp 20–23

21 Bauhuber, V F, Erlacher, V and Gunther, P 'Programmsystem fur die behandlung digitaler hohenmodelle' *Bildmessung und Luftbildwesen* Vol 3 (1975) pp 103–108

22 Garvis, G K and Gee, B C 'Computer graphics plot the lie of the land' *Surveyor* (May 1985)

23 Davies, H E H 'Optimizing of highway vertical alignments to minimize construction costs: program Minerva' *Transport and Road Research Laboratory Report, LR463,* (1972)

24 Craine, G S 'MOSS' *Civil Engineering Surveyor* Vol 10 No 10 (1985) pp 17–21

25 Houlton, J M 'Computer aided drafting in highway engineering' *Current Issues in Highway Design. Proc. Conf. Highway Design Inst. of Civil Engineers,* Telford Publications (1985) pp 149–162

26 Houlton, J M 'MOSS integrated modelling, design and drafting for civil engineers' *Proc. CADCAM'86* Birmingham, UK (1986) pp 235–242

27 Hill, D 'MOSS/Intergraph at Minenco' *Proc. MOSS Users Conference* Melbourne, Australia (1985)

28 Siddans, D R 'Photomontages from oblique aerial photographs' *The Highway Eng.* (March 1980) pp 14–19

29 Cobb, J 'A new route for graphics' *Proc. of Computer Graphics '85* On-Line Publications (1985) pp 173–183

30 Gould, S 'ECLIPSE' *Civil Engineering Surveyor* Vol 11 No 1 (1986) pp 12–16

31 Hogan, R E 'HASP Digical survey design system' *Proc. of the ASP-ACSM Fall Convention* San Antonio, TX, USA (1984) pp 458–475

32 Hirst, D 'SURVPAK' *Civil Engineering Surveyor* Vol 11 No 2 (1986) pp 15–16

33 Bray, D and McGarry, M 'From total station to digital model – the role of CAD' *Proc. CADCAM'86,* Birmingham, UK (1986) pp 249–254

34 McGarry, M 'Surface modelling in the context of a general draughting system' *Proc. 2nd Int. Conf. Civil Struct. Eng. Comput.* (1985) pp 195–199

35 Grayer, J L 'A site modelling system' *Proc. 4th Int. Conf. on Computers in Design Engineering (CAD 80)* Butterworths, UK (1980) pp 724–732

36 Strodachs, J J and Durrant, A M 'Surface and solid modelling for civil engineering applications – an introduction to 'intelligent ground models'' *Proc. CADCAM'86* Birmingham, UK (1986) pp 243–248

37 Pratt, M J 'Interactive geometric modelling for integrated CAD/CAM' *Eurographics tutorial notes* Springer-Verlag (1984)

38 Requicha, A A G and Voelcker, H B 'Solid modeling: a historical summary and contemporary assessment' *IEEE Computer Graph. Appl.* Vol 2 No 2 (1982) pp 9–24

39 Mayo, R H, Durrant, A M and Finniear, L J 'The application of the new Medusa ground modeller, preliminary findings of the beta-test' *Proc. Medusa User Group Meeting* Cambridge, UK (1986)

BIBLIOGRAPHY

Hodges, D J and Alderson, J S 'Automated volumetric surveys' *University of Nottingham Mining Dept. Magazine* No 37 (1985) pp 71–76

Photogrammetric Record, 12(72): 711–745 (October 1988)

MODELLING FOR DIGITAL TERRAIN AND LANDSCAPE VISUALISATION

By T. J. M. KENNIE
University of Surrey
and R. A. MCLAREN
Know Edge Limited

(Paper read at the Technical Meeting of the Society on 16th February, 1988)

Abstract

Methods of creating digital terrain and landscape models for regional and site specific visualisations are described. A number of alternative approaches are discussed and their relative accuracies are considered. This is followed by a review of the computer graphics procedures which can be used to render such models. Particular attention is directed towards methods for depth cueing, hidden surface removal, shading and texturing. The effects of the atmosphere and methods of producing shadows are also examined. The paper then presents varied applications of such visualisations, including small scale, regional examples and large scale, site specific applications. In conclusion, the role of photogrammetry in the process of visualisation is assessed and possible future developments in this field are identified.

INTRODUCTION

THE REPRESENTATION of relief is a fundamental component of the cartographic process. A wide range of techniques representing the topographic variations of the earth's surface on a two dimensional surface have been developed and these vary both in their symbolic content and in their degree of realism (Fig. 1).

Recently, the use of digital terrain models (DTM) and computer graphics techniques has become a more common method not only for generating computer visualisations of the terrain but also for assessing the impact of manmade objects on the landscape. Thus in many cases the terrain model providing the geometric description of the earth's surface will be supplemented by descriptions of significant landscape features. Vegetation and cultural information as well as design objects will often be combined with the terrain model to create a scene with a higher degree of realism. Models of this form can be referred to as landscape models. In small scale applications, this landscape information will probably be polygonal land use data, while at larger scales it may include explicit three dimensional geometric descriptions of individual features or blocks of features. This paper therefore reviews the techniques used to create and enhance these digital terrain and landscape visualisations and considers some of their small and large scale applications.

There are many interrelated reasons which account for the growth of interest in this area of visualisation. Certainly one of the most significant has been the improvement in the price/performance ratio of modern computers. The display architectures and memory facilities available in the current generation of graphics workstations enables results to be achieved at a price which could not have been contemplated 10 years ago (Gelberg and Stephenson, 1987).

The general improvements in the field of computer science have also had an immense impact on the design of the instrumentation used to acquire terrain data.

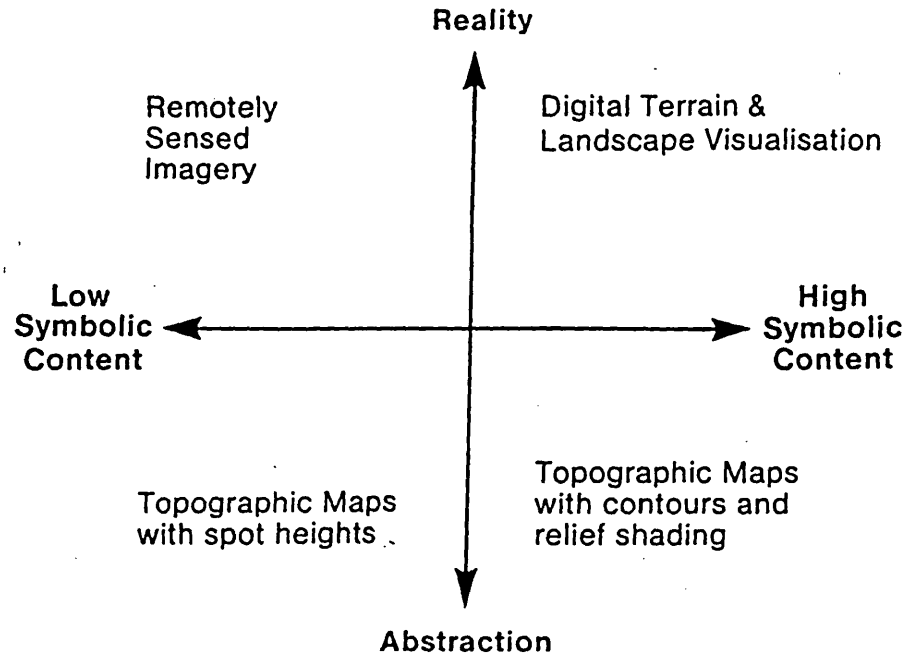


FIG. 1. Classification of techniques for visualising terrain.

Electronic tacheometers, analytical stereoplotters and automated cartographic digitisers are now widely used to produce grid or triangular terrain models. This ability to create DTMs at an economic cost has had an enormous effect on the use of DTMs generally for mapping and civil engineering design and has also created a demand for more sophisticated methods of displaying the data. Several organisations (mostly military) have also created small scale regional, national and international terrain databases. The provision of these terrain databases has also been a significant influence on the development of methods of visualising the terrain.

Concurrent with the improvements in computer hardware have been some extraordinary developments in the algorithms used for general scientific visualisation. The impetus for many of these software developments has come from the needs of the simulation, CAD/CAM, advertising and entertainment industries. The requirements of the film industry for computer generated animations have been particularly influential in this respect. The importance of this area of technology is indicated by the recent formation by the USA National Science Foundation of a group to examine user needs in the general field of "Visualisation in Scientific Computing" (McCormick *et al.*, 1987).

While there is no shortage of techniques for general visualisation purposes, the characteristics of the earth's surface can limit the applicability of many of these techniques. Among the more important differences between terrain and landscape visualisation and other forms of visualisation are the following considerations.

(a) Natural phenomena are inherently more complex than manmade objects and thus more difficult to model.

(b) The earth's surface is not geometric in character and cannot be modelled effectively by using higher order primitives such as those used in CAD/CAM applications.

(c) The terrain and landscape model dataset sizes are considerably larger than the datasets in many other forms of visualisation.

(d) There are generally higher constraints on geometric accuracy than in many other applications.

(e) The scenes are not spatially compact. Therefore, the modelling may involve multiple levels of detail based on the object/viewpoint relationship.

(f) The optical model is more complex due to the effects of atmospheric refraction and earth curvature which are encountered in extensive datasets.

Finally, user requirements have also influenced developments in this field. For example, as a consequence of recent changes in EEC contract regulations (European Economic Community, 1985), visual impact analysis is now a legal requirement of certain major construction projects. The needs of users in the field of geographic information systems (GIS) for display technologies to illustrate the results of sophisticated spatial analyses have also been a significant factor. Military users have also been a major driving force and a selected number of applications of terrain and landscape visualisation are presented in the latter half of the paper.

SECTION I: CREATION OF DIGITAL TERRAIN MODELS FOR VISUALISATION

The acquisition of terrain data is a critical stage in the process of visualisation since it will have a considerable influence on the fidelity and hence the degree of realism which can subsequently be achieved. Four distinct approaches to the creation of DTMs for terrain and landscape visualisation can be identified:

(a) ground surveying techniques using electronic tacheometers and data collectors for the production of high accuracy models of limited areas;

(b) photogrammetric techniques using either analytical stereoplotters, possibly with correlators attached, or instruments for orthophotograph production for medium to high accuracy models of large areas;

(c) graphics digitising techniques where contour lines on existing maps are transformed into digital co-ordinate data using either manual or automatic, line following digitisers or automatic, raster scan digitisers (in both cases the accuracy is significantly lower than the previous two techniques although the coverage can be very large); and

(d) generation of simulated terrain using either computer based geomorphological models of the earth's surface or fractal geometry to generate more abstract surfaces.

Approaches to Sampling Real Terrain

Regardless of the technique which is used to measure the terrain, the matter of determining the optimum number of points to acquire is a central issue, since the density of sampling has a direct influence on the accuracy of the terrain representation.

Several theoretical models have been developed in recent years in an attempt to resolve this problem. The most commonly reported methods are those based on the analysis of either the power spectrum of the terrain using Fourier analysis (Tempfli, 1980), variograms of the terrain (Frederiksen *et al.*, 1985) or self similarity using fractal techniques (Muller and Saksono, 1986). These theoretical techniques are currently limited in their applicability although they show some potential in the determination of sampling strategies for photogrammetric mapping over large areas (Balce, 1987).

In practice it is possible to distinguish three commonly occurring approaches to data sampling: selective (irregularly located) sampling using both ground and photogrammetric survey techniques, systematic (regularly located) sampling using photogrammetric techniques and hybrid sampling using cartographic digitising.

Selective (Irregular) Sampling using Ground Surveying Techniques

For high accuracy terrain representation over relatively small areas, selective ground sampling methods are used to acquire data. In this case the surveyor chooses points at significant changes in terrain form, for example at the tops of hills, breaks of slope, ridges and so on. The outcome is eventually a series of irregularly located points representing the terrain. Surface representation is subsequently derived by using either planar or curvilinear triangular elements or by grid interpolation.

Systematic (Regular) Sampling using Photogrammetric Techniques

An alternative strategy is normally adopted when sampling is performed by close range or aerial photogrammetric techniques. In this case a systematic pattern of heights is measured, normally in a grid fashion. The primary advantage of this technique is that all, or part, of the process can be automated, for example by using analytical stereoplotters to drive to pre-set positions and/or by using correlators to automate the height measurement. The well known disadvantage of this approach, that the distribution of the datapoints may not be related to the nature of the terrain, can be overcome by using the progressive and composite sampling techniques derived by Makarovič and reported more recently by Tempfli (1986).

Hybrid Sampling using Cartographic Techniques

Digitising of existing maps leads to a different approach to sampling from those discussed in the previous two examples. For example, in the case of manual line following digitising, the sampling of the terrain often combines features of both the selective and systematic procedures outlined previously. The digitising of maps as an input to the Panacea software suite (McCullagh, 1988) involves initially forming an irregular triangulated data structure followed by the generation of a grid DTM from the triangulated data.

Regional Terrain Databases

A development in recent years, which is based on cartographic digitising techniques and to a more limited extent on photogrammetric measurements, has been the creation of a series of regional terrain databases. These have been produced by a number of organisations, including the Defense Mapping Agency (DMA), the United States Geological Survey (USGS) and the national mapping agencies in Sweden, West Germany, Belgium, Canada and Japan among others. In the United Kingdom, the Mapping and Charting Establishment (MCE) has also created regional and national DTMs from both 1:250 000 and 1:50 000 scale OS topographic maps, although their use is currently restricted to military projects. Selected parts of this database have, however, recently become available for civilian projects and are now available from the Ordnance Survey (OS). Full coverage of the UK at 1:50 000 scale is anticipated by 1991. Many applications which involve elements of terrain and landscape visualisation make use of these databases and therefore a brief description will be provided of the most commonly used databases.

DMA Terrain Databases

The Digital LandMass System (DLMS) is a database of terrain and associated cultural information which has been created during the past 15 years by the DMA and the military survey units of Belgium, France, Germany, Italy, The Netherlands, Norway and the United Kingdom. Approximately one third of the globe has been covered to date.

The DLMS consists of two independent components: the Digital Terrain Elevation Data (DTED) and the Digital Feature Analysis Data (DFAD). The latter database is concerned with identifying surface material types and their likely radar signature. DTED consists of a uniform grid of elevations arranged in a geographical reference system. The data are generated either by cartographic digitising of existing 1:250 000 contour sheets to produce low resolution (Level 1) elevations on a 100 m grid or by analytical photogrammetry (Level 2) to produce higher resolution data in areas of strategic military importance.

USGS Terrain Databases

The USGS provides two separate DTM or, more correctly, Digital Elevation Model (DEM) products. Firstly, the USGS distributes the 1:250 000 scale DTED

data produced by the DMA for the USA and Alaska. Their second product is a DEM of the 7'·5 quadrangle topographic map sheets, which are at a scale of 1:24 000. In this case the heights are derived from GPM-2 correlator devices or from profiles measured by analytical stereoplotters and structured into a regular grid of elevations at 30m intervals. Further details of the data structure of these databases can be found in El Assal and Caruso (1983).

Accuracy of DTMs for Visualisation

The relative accuracies of DTMs derived by ground survey, photogrammetry and digitising existing map sources are illustrated by Table I. It is clear from this table that the fidelity of the elevations derived by different techniques can vary substantially. For small scale visualisation applications (such as flight simulation), these inaccuracies in the absolute values in the database may be relatively unimportant provided that the dataset is free from gross errors and the relative heights of points are within an acceptable tolerance. As long as these limitations in accuracy are recognised by and are acceptable to users lower quality data are tolerable.

For large scale applications in landscape architecture, civil engineering and opencast mining, however, the inaccuracies in elevation resulting from the data derived from small scale map sources described in Table I would generally be unacceptable. This would be particularly important if the terrain model was being used for visual impact analyses.

Table I. Relative accuracies of DTM elevation values as derived by various surveying techniques (adapted from Petrie and Kennie (1988)). (Accuracy factor 1 = high accuracy; factor 500 = low accuracy).

Source	*Data measured*	*Accuracy factor*
Ground Survey		
Large scale model (1:500)	1. Directly measured spot heights	1
Photogrammetric measurement		
Aerial photographs for topographic mapping (for example 1:10 000)	2. Photogrammetric spot heights —measured in a stationary mode	5
	3. Photogrammetric contours —measured in a dynamic mode	15
Cartographic digitising		
Medium scale topographic maps (for example 1:50 000)	4. Contours generalised from medium scale maps	25
	5. Contours measured by field survey in the 19th century	50
Small scale topographic maps (for example 1:250 000)	6. Spot heights at grid nodes derived by interpolation from digitised contours	500

Approaches to Creating Simulated Terrain

An alternative approach to the creation of terrain models for visualisation purposes is to produce a simulated surface rather than measure a real surface. Two distinct techniques, which have been used for such exercises, are those which simulate geomorphological processes and those which use fractal geometry.



PLATE I. Simulated fractal landscape with fractal dimension of 2.2. *R. Voss, IBM Ltd.*



PLATE V. Modelling of trees using particle systems. *Reeves and Blau, 1985.*



(a) (b)

PLATE II. (a) Area fill colour coded heights. (b) Hill shading of same region as in (a) to simulate radar reflectance. Both images were produced by the Laser-Scan Terrain Visualisation and Exploitation Software (TVES). *Laser-Scan Laboratories Ltd.*



PLATE III. Shading. (a) Facetted shading. (b) Gouraud shading. Both images were produced by the CIS Medusa software. *Cambridge Interactive Systems Ltd.*



PLATE IV. Terrain model of area near Los Angeles generated using ChapTerrains Software Data by Jet Propulsion Laboratories. © *Pixar 1987.*



PLATE VI. Aeronautical chart of part of Alaska covering Anchorage and Prince William Sound illustrating automated hill shading with colour coded height variations. *USGS EROS Data Center.*



PLATE VII. Scene generated using flight simulator. *Rediffusion Ltd.*



PLATE VIII. Photomontage of transmission line routing. *Turnbull Jeffrey Partnership.*

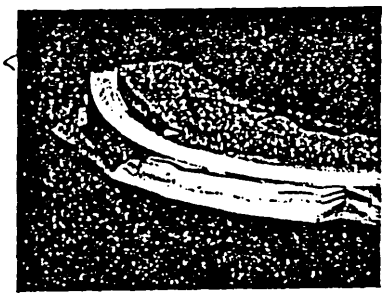
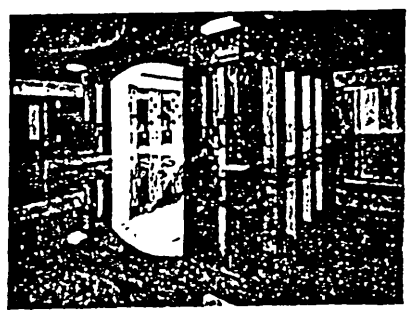


PLATE IX. Colour fill view of highway model produced by the MOSS System and Auto-Trol colour fill software. *Colorado Dept. of Highways* and *MOSS Systems Ltd.*



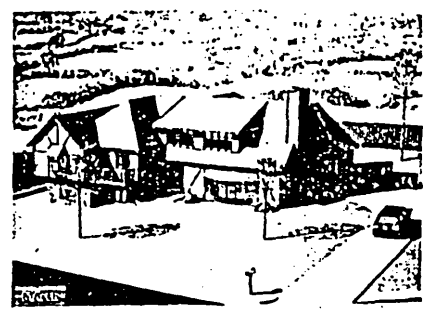


PLATE X. Intergraph Render examples. *Intergraph (UK) Ltd.*

Geomorphological Modelling

Computer simulations of terrain attributes such as slope development (Armstrong, 1976) or of large scale erosional landforms (Craig, 1980) have attracted considerable interest in the geomorphological community in recent years. A novel application of several of these concepts to the creation of a synthesised terrain surface for a vehicle simulation exercise can be found in Griffin (1987). The aim of the project was to simulate a fluvial landscape by the establishment of a network of streams over an initial surface. The simulation was based on a primary surface (a horizontal/tilted plane consisting of a grid of spot heights) with different surface lithologies (to simulate resistance to erosion) and the movement of streams based on a random walk algorithm.

Durrant (1987) also discusses the use of simulation, in this case to assess the dynamic nature of the discharge from tailings dams. These dams are structures built from mine waste which are continuously constructed during their service lives. The paper considers methods of modelling the nature of the depositional cycle from the discharge points by simulating the flow pattern. The CIS Medusa system was used for these operations. Fig. 2 shows a wireframe example of one such simulation as well as a colour shaded simulation.

Most of the geomorphological models include both deterministic and stochastic (for example random walk) elements. Realistic simulated terrain has also been produced solely by stochastic processes of which the most common involves the use of fractals.

Stochastic Modelling using Fractal Geometry

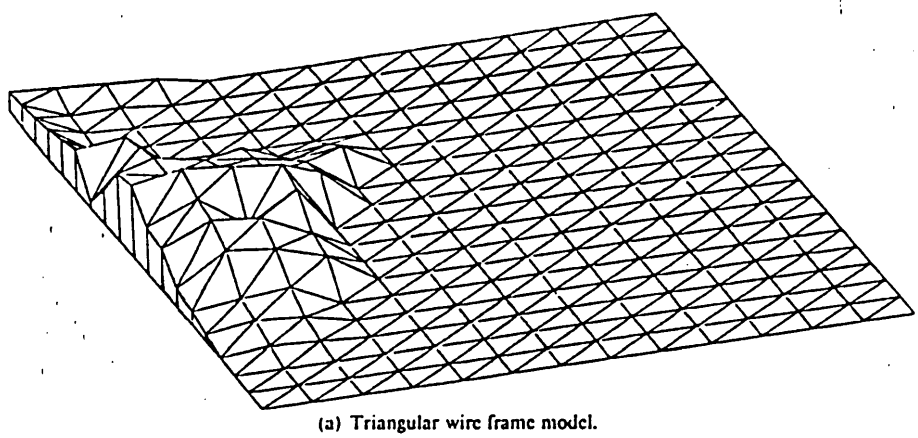
The terms fractal and fractional Brownian motion were originally devised by Mandelbrot and van Ness (1968) to denote a family of stochastic models which could be used to represent many spatial or temporal phenomena. Since then the concept has been applied by Mandelbrot and many other researchers to the process of terrain and landscape simulation (Mandelbrot, 1975; Fournier *et al.*, 1982; Voss, 1985; Clarke, 1987).

One of the central themes of fractal geometry is the concept of self similarity. This property of an object describes the invariance of the object under changes of scale, or alternatively the degree to which objects appear similar, but nevertheless different, at varying levels of magnification. This idealised form of fractal geometry rarely, if ever, occurs in nature although statistical or stochastic self similarity is considered by many to be evident in real landscapes. Mandelbrot and others have shown that landscape form can be described by the fractal dimension of the surface, a non-integer value between 2 (a perfectly smooth surface) and 3 (an infinitely variable surface). Real landscapes have been found to have fractal dimensions between 2·1 and 2·3, although values as high as 2·75 have been reported.

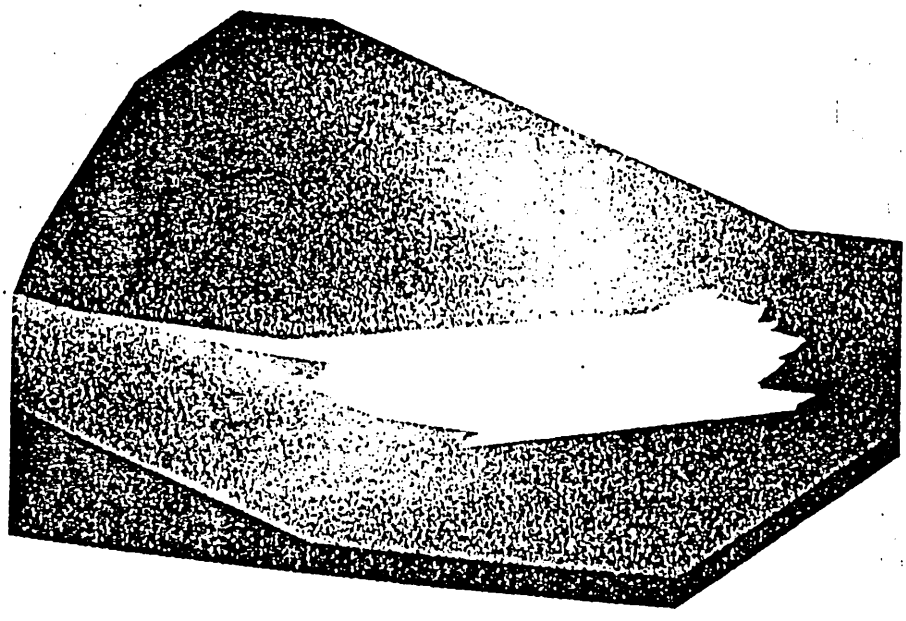
The consequence of accepting self similarity is to suggest that the spatial structures inherent within the landscape repeat themselves at all scales. The evidence from several sources seems, however, to indicate that self similarity does not occur in nature. One explanation suggested by Roy *et al.* (1987) is that the earth forming processes have operated on the landscape for varying amounts of time and consequently the earth's surface manifests the scale at which these different processes operated. Burrough (1985) also confirms this opinion and suggests that "most landscapes are not the result of a single dominant process, but are the result of the complex interaction and superimposition of many processes". In spite of these limitations, simulated landscapes generated with fractal dimensions between 2·2 and 2·3 have been found to produce realistic looking images and have been used in several animated sequences for science fiction films and advertising (Plate I).

SECTION II: COMPUTER GRAPHICS TECHNIQUES FOR DISPLAYING DIGITAL TERRAIN AND LANDSCAPE VISUALISATIONS

This section of the paper examines the various techniques which can be used to



(a) Triangular wire frame model.



(b) Shaded simulation.

FIG. 2. Simulation of flow regime associated with outflow from tailings dams. *Durrant (1988).*

create computer visualisations of the terrain and associated landscape features. The degree of realism which can be achieved in such a process is dependent upon several factors including the nature of the application, the objective of the visualisation, the capabilities of the available software and hardware and the amount of detail recorded in the model of the scene. At one end of the realism spectrum are highly abstract, simple, monochrome wireframe models while at the opposite end are highly realistic, fully textured and coloured images. While the quest for realism ultimately attempts to produce images which are indistinguishable from reality, it is generally impractical at present to use such images in dynamic environments such as landscape design or avionics. For the future, however, it seems likely that there will be a trend towards the creation of more sophisticated dynamic imagery.

Assumed Display and Output Devices

Raster scan displays have become the *de facto* technology for visualisation applications. Such devices are typically a 19 inch diagonal CRT screen with a resolution of 1024 × 1280 pixels and a capability of displaying up to 16·8 million colours. Typical examples of this technology are the Sun-3 and Sun-4 families of engineering workstations from Sun Microsystems, Inc. The scene generation approaches described in the paper will concentrate on techniques which rely on raster technology. These techniques will also continue to be applicable as the current CRT technology is superseded by new devices such as flat panel displays.

Despite this predominance of raster display technology in forming images for visualisation, the hard copy output is often produced by a vector plotter. For many applications the level of abstraction produced by such devices is sufficient for the scene analysis, especially at the planning and early design phases. Therefore, vector techniques will also be explored in describing scene generation approaches.

Methods of Displaying DTM Data

Some relatively simple techniques have been developed for displaying DTMs on a computer graphics display. More sophisticated approaches which involve the use of perspective transformations and rendering techniques are discussed subsequently.

Probably the most common method which is used to indicate height variations on a two dimensional surface is the contour. Traditional graphical maps normally portray contours in a single colour and use a variable line width to assist in the identification of major height variations. Colour graphics displays may also be used to portray contour lines, although colour variations are more commonly used to give an appreciation of the height variations rather than variable line widths. An extension of the colour coding of contour lines is to colour code the areas between contour lines as an aid to improving the appreciation of the topographic variations (Plate II(a)). In both cases the choice of colours which are used to represent the lines or areas is of crucial importance. Often a single colour of varying hue and intensity can be particularly effective.

In addition to the display of contours it is possible to derive additional information from a DTM which can improve the visualisation of a particular attribute of the earth's surface. Thus it is quite common to derive the slope or gradient of the terrain and display this either in the form of slope vectors or as a colour coded image. Aspect or direction of maximum slope is a further attribute which can be derived from the terrain data and displayed as a colour coded image. In both cases the choice of colours can also have a major impact on the interpretability of the resulting images. By combining both slope and aspect information with the position of the sun as the illumination source, it is possible to generate, in monochrome, the effect of hill shading or shaded relief (Plate II(b)).

These techniques are capable of meeting the display needs of many users. They do, however, exhibit a number of deficiencies including the production of a highly abstract view of the terrain from a fixed viewing position with few facilities for including significant landscape features in the displayed image. The following sections describe methods of addressing these limitations.

Components of an Information Model for Visualisation

The majority of computer graphics techniques have been developed to address the visualisation of objects that have their geometry defined using a mesh of planar surfaces such as triangles. Some algorithms are available to process higher order surfaces such as bicubics. However, these still tend to be the exception, forcing most approaches to decompose higher order surfaces into their component planar surfaces prior to graphics processing (Snyder and Barr, 1987).

These geometric descriptions of a surface can also be supplemented by parameters describing the surfaces' visual characteristics such as colour, reflectivity and texture that in effect clothe the terrain. The allocation of the correct colours to

objects is crucial in achieving realism and the process is not as simple as it may initially appear. Difficulties can arise when attempting to select a computer graphics generated colour, on a screen or hard copy device, to match the colour of a natural object. In part, this is due to the "artificial" colour models used to present colour to the user of graphics systems. Schwarz *et al.* (1987) conducted a revealing experimental comparison of a number of different colour models encountered in the computer graphics world.

Ideally, a scene generation system should be an integral facility of a GIS where the objects composing a scene could be extracted from the database, together with their associated three dimensional graphical representations, to create a required scene.

Methods of Rendering Terrain and Landscape Models

Having assembled the geometric model of the terrain and landscape objects to be visualised, the next stage of visualisation is to *render* this three dimensional description of a scene onto a two dimensional display device.

As well as the model of the terrain and associated landscape features contained within the required scene, the rendering process requires a number of other parameters to be defined including:

(a) the viewing position and direction of view of the observer;

(b) a lighting model to describe the illumination conditions;

(c) a series of "conditional modifiers" (McAulay, 1988), parameters which describe the viewing condition of the landscape objects (under wet conditions for example, the surface characteristics of objects are quite different from dry conditions);

(d) a set of "environmental modifiers" (McAulay, 1988), parameters which describe atmospheric conditions and may model effects such as haze; and

(e) a sky and cloud model representing the prevailing conditions.

All, or part, of the information above may then be used by the scene rendering process to generate a two dimensional array of intensities or pixel values that will be displayed on the raster display device. The complexity of the rendering process is directly dependent upon the degree of image realism required by the user. There is always a trade off between quality of image and performance/cost. For example, in the development of the Reyes rendering systems (Cook *et al.*, 1987) the performance objective was to create a two hour film in one year. At 24 frames per second, this is equivalent to rendering a frame in three minutes.

An overview of the rendering process is shown in Fig. 3 and the remainder of this section describes the steps involved in the scene rendering procedure. The sequence of steps may vary depending on the rendering approach chosen.

Geometric Transformations

The three dimensional terrain and landscape information is normally mapped into two dimensional space by a perspective projection, where the size of an object in the image is scaled inversely as its distance from the viewer. An example of this form of projection is illustrated by Fig. 4. In this class of projection, the interpretation of depth is usually based on the assumption that the largest of a class of objects is closest to the viewer and, in effect, it models a "pin hole" camera.

Occasionally, the isometric projection (Butland, 1979; Dubayah and Dozier, 1986) is used to visualise terrain and the modelling of sub-surfaces. This is an axonometric projection that has the useful property that all three principal axes are equally foreshortened, providing equal scale in all three directions. In addition, the principal axes are projected so that they make equal angles with one another. This is useful for some forms of analyses.

For applications where the geometric fidelity of the rendered scene is of vital importance, for example the creation of a photomontage product in visual impact assessment, it is necessary to incorporate both earth curvature and atmospheric refraction corrections into the viewing model.

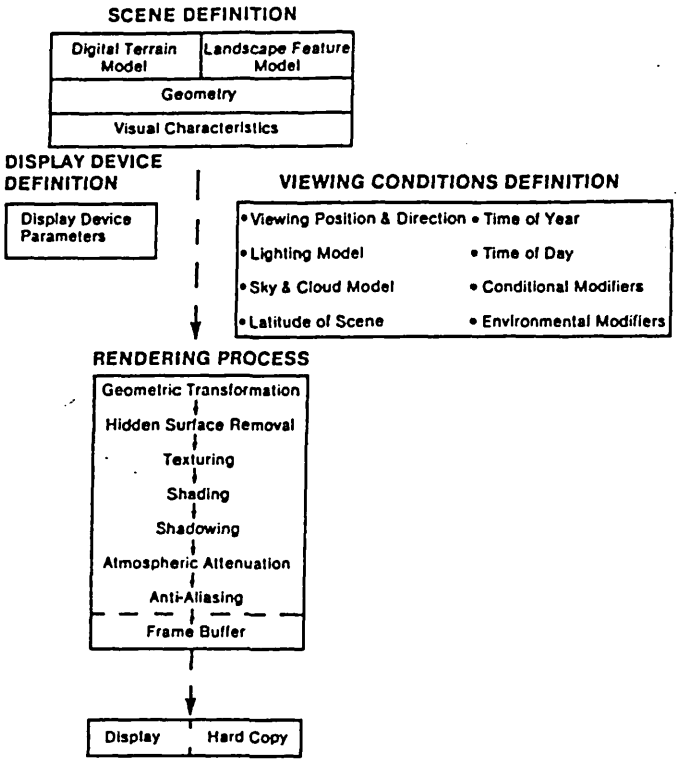


FIG. 3. Overview of the rendering process.

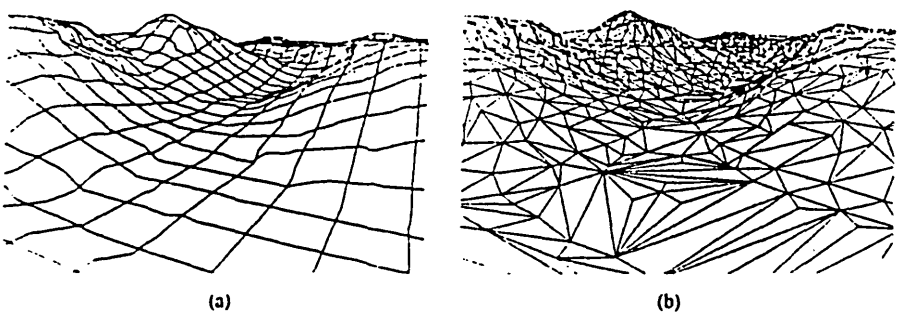


FIG. 4. Depth cueing using (a) gridded structure and (b) triangular structure. *Cambridge Interactive Systems Ltd.*

Depth Cueing

When a three dimensional scene is rendered into two dimensional space with any level of abstraction, an ambiguous image will probably be portrayed. This is compounded by the fact that human eyes are not a window into the world, but instead the world is created in the mind based on preconceived models that vary from person to person (Gregory, 1977). Therefore, if new computer graphics presentation concepts do not match these preconceived models, then they are open to misinterpretation. A wireframe model presents the viewer with the maximum degree of ambiguity. To compensate for this loss of inherent three dimensional information, techniques have been developed to increase the three dimensional

interpretability of the scene using depth cueing techniques that attempt to match the perceived computer generated image to human "natural" visual cue models.

Perspective projection. Depth cues are inherent in the perspective projection used to create the two dimensional image. When a projected object is known to contain parallel lines this type of projection is excellent at depicting depth, since parallel lines seem to converge at their vanishing points. This very effective visual cueing is highlighted when comparing the projection of a triangular DTM with its equivalent square grid derivative, in their wireframe forms (Fig. 4). The depth cueing is pronounced in the square grid form due to the combined effect of parallel lines and the diminishing size of the uniform squares. The triangular form, with its lack of uniformity and randomness of triangular size, presents a very ambiguous and confusing image that requires further cueing techniques to allow adequate interpretation.

Intensity depth cueing. When a viewer analyses a natural scene, one of the most important depth cues corresponds to the relative lighting intensities of objects. Through atmospheric attenuation, the contrast of objects with the sky diminishes with distance; light objects become dimmer while dark ones increase in brightness. This lighting graduation effect can also be simulated in the rendering process. Although an important component of a rendered scene, it cannot be used to depict small differences in distance from the viewer since the eye's intensity resolution is much lower than its spatial resolution.

If the sky is modelled in the scene, this can also add to the depth cues by ensuring that the sky appears lighter overhead than at the horizon and thus conforms to reality.

Variable object representation. As an object recedes from a viewer, the object diffuses through several forms of apparent simplification until it eventually disappears from sight. This is due to the limited resolution of the eye's optical system and is influenced by atmospheric distortion and the shape and texture of the object. The ability to resolve detail is a key depth cue and the techniques of object image simplification are analogous to the problems of scale generalisation in cartography. Therefore, any image that is attempting to simulate realism in a spatially deep scene should include a mechanism to vary the perceived level of detail of an object based on its distance from the viewer.

This facility is an integral component of a flight simulator where the scene database contains multiple models of objects together with rules which control when and how a particular model will be displayed. When flying towards a distant object, the system fades out one model of the object and merges in a more detailed model based on the distance from the object. Dynamic overlapping translucency techniques are employed to smooth the transition between object changes, avoiding distracting scene flicker.

This simplification approach is used in simulating pen drawings, derived from the DTM, to highlight specific geomorphological characteristics of the terrain form that are relevant for a particular application. Sasada (1987) describes a number of techniques including mountain range outlines, ridge horizon lines and natural waterflow lines. In many ways these techniques form a decomposition of the landscape into fundamental structural components that also correspond with the elementary visual cues (the mental model of landscape forms). Therefore, these very abstract graphics allow the landscape scene to be reconstructed with relative ease. Possibly they could be combined with the more "sophisticated" computer graphics visualisation techniques to achieve acceptable realism with a higher degree of abstraction than would normaly be allowed, for example a smoothly shaded triangular network with shadows and ridge lines superimposed.

Dynamic projections. Many of the current display devices allow users to dynamically change the viewing position and direction, simulating either a walk through the scene or a rotation of the scene depending on the frame of reference. This ability to vary the parallax allows the user to resolve many of the depth and geometric ambiguities apparent in the static image.

Dynamic display facilities are an inherent capability of flight simulators and have become an essential tool in many design applications. For terrain and

landscape visualisation with accompanying large data volumes, dynamic modelling can pose significant problems for non-customised display hardware.

Stereopsis. Computer graphics can also be used to generate stereoscopic images that provide a very powerful depth cue. Several products are available which enable stereoviewing to be achieved on a single CRT. Alternating refresh cycles display left- and right-eye views synchronised with mechanical or optical shutters on customised stereoviewing glasses. Waldern (1986) provides a review of such techniques.

A further example of the use of this technique in photogrammetry is the employment of "stereo-image injection" as demonstrated on the Wild (Heerbrugg) S9-AP analytical plotter (Beerenwinkel *et al.* (1986)). The procedure is implemented by projecting the image of two small format CRTs into the optical trains of the two eyepieces. The co-ordinate systems of the CRTs are set to correspond to each of the photoco-ordinate systems of the analytical plotter, allowing existing digital information in world co-ordinates to be displayed and viewed stereoscopically, superimposed on the stereomodel. This device only allows monochrome vector information to be displayed.

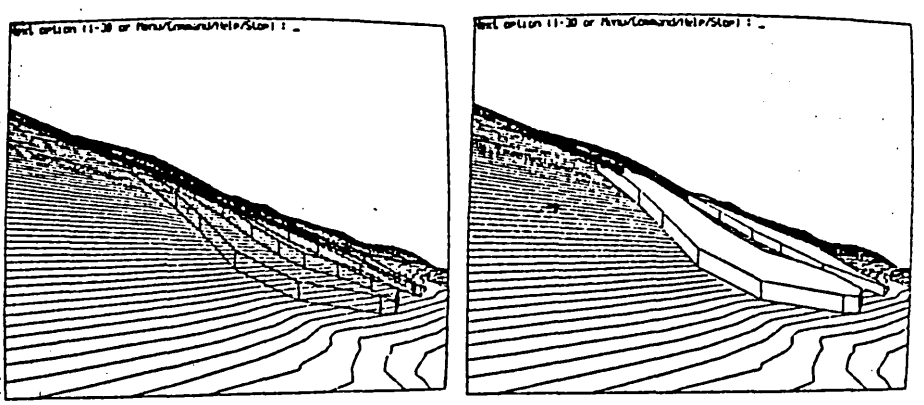
Hidden Edge Removal

At the highest level of abstraction, objects can be modelled by their edge components, creating a three dimensional skeleton form called a wireframe model. When this is subsequently viewed as a perspective projection, the result is confusing and ambiguous due to edges being visible that would normally be hidden if a solid model had been used. The interpretability of depth relationships of the image can be improved by removing the hidden edges from the image. The approach to hidden edge removal is equivalent to hidden surface removal which is described in the following section.

Although images formed by wireframe models with hidden edge removal are primitive images which no pretence at being realistic, they still portray form and geometric fidelity and provide an inexpensive technique for visualisation. Major advantages are the low overheads in their production and their ability to be output on standard vector plotting devices. An example of hidden line removal is shown in Fig. 5. The visualisation of terrain models in wireframe form is not limited to the direct output of the grid or triangular network; ridge lines and other visual forces, derived from the DTM, may be used discretely or in combination with the DTM structure lines.

Hidden Surface Removal

An extension of the wireframe display is to 'fill' the component surfaces



(a) (b)

FIG. 5. Hidden edge removal. (a) No edges removed. (b) With edges removed. *Turnbull Jeffrey Partnership.*

defining the objects, giving the impression of solid objects. Just as edges were removed in the context of wireframe models to eliminate ambiguities, hidden surface removal techniques are employed to remove those edges and surfaces that are obscured by other visible surfaces. The hidden surface capability provides a further step forward in the quest for realism.

The implementation of the technique of hidden surface removal is computationally expensive, especially for complex landscape scenes, where the rendering process can involve hundreds of thousands of surfaces. Therefore, the challenge has encouraged a wide variation of algorithms. These can be categorised into two fundamental approaches: *image space* and *object space* algorithms.

One of the simpler and more popular object space algorithms is the *depth sort* algorithm (Newman and Sproull, 1972; Sechrest and Greenberg, 1981), in which the polygons to be displayed are sorted by their distance from the viewpoint and scan converted into the refresh buffer of the display device in order of decreasing distance. Since the nearest polygons are placed into the refresh buffer last, they obscure the polygons that are further away by overwriting them in the refresh buffer. The priority for displaying the polygons cannot be simply based on a single maximum z value since depth ambiguities may occur, which have to be resolved. Although this algorithm uses the structure of the display device to aid in the surface removal and is conceptually simple, it does have the limitation of requiring the polygons to be sorted prior to display.

One of the most popular algorithms used for the elimination of surfaces in a landscape scene is the *z buffer* or *refresh buffer* image space algorithm. This approach assumes that the display device has a z buffer, in which z values for each of the pixels can be stored, as well as the frame buffer in which the intensity levels of the pixels are stored. With the reduced cost of memory, this is becoming a standard facility of display devices. Each polygon is scan converted into the frame buffer where its depth is compared with the z value of the current pixel in the z buffer. If the polygon's depth is less than the z buffer value, then the pixel's frame buffer and z buffer values are updated. This is repeated for each polygon, without the need to pre-sort the polygons as in other algorithms. Hence, objects appear on the screen in the order they are processed and not necessarily either front to back or back to front.

A further example of the image space category of algorithms is the *scan line* algorithm. The approach has the strategy of finding the visible polygons for each scan line of the image. This simplifies the domain of the problem from polygons to lines. In order to reduce the number of polygons that have to be considered for each scan line, the polygons are sorted in one direction (from the top downwards, for example). There are many variations of this algorithm, with different strategies in finding visible lines and how they propagate relevant information to neighbouring scan lines and capitalise on scan line and edge coherence.

Anti-aliasing

Many computer graphics images displayed on raster display devices exhibit disturbing image defects such as jagging of straight lines, distortion of very small or distant objects and the creation of inconsistencies in areas of complicated detail. These distortions are caused by improper sampling of the original image and are called *aliasing* artifacts. Techniques known as *anti-aliasing*, which have their roots in sampling theory, have been developed to reduce their influence (Crow, 1977).

The fundamental concept is that a pixel, which has a non-zero area on the screen, should be used to represent the corresponding non-zero area of the world that is mapped onto the pixel. Therefore, many objects in the world may contribute to the intensity function at a pixel. In an effort to reduce processing overheads, most of the anti-aliasing algorithms assume that aliasing effects will only be encountered at the boundary of polygons where high frequencies occur. However, high frequencies also occur in the interior of polygons where there is an abrupt change in the displayed objects surface normal and when highly directional light sources introduce abrupt intensity changes. These aspects of the problem have been addressed by several authors (Crow, 1982; Williams, 1983).

Anti-aliasing techniques do add high overheads to the rendering process. However, hardware implementations using A buffer anti-aliased, area averaged accumulation buffer techniques (Carpenter, 1984) and parallel hardware architectures have also been used (Gupta *et al.*, 1981). An alternative approach that bypasses the problem is to perform a scan conversion of a line at double the resolution, which is generally faster than anti-aliasing the line at the original resolution.

Smoothing of the lines and polygonal boundaries by applying anti-aliasing techniques is achieved at the expense of image sharpness. There is a slight blurring of the line edges. However, the resulting images are less noisy and more pleasing to the eye.

Shading

The next step towards the goal of realism is the shading of visible surfaces within the scene (Newell *et al.*, 1972). The appearance of a surface is dependent upon the type of light source(s) illuminating the object, the condition of the intervening atmosphere, the surface properties including colour, reflectance and texture, and the position and orientation of the surface relative to the light sources, other surfaces and the viewer. The objective of the shading stage is to evaluate the illumination of the surfaces within the scene from the viewer's position. The more oblique the surface is to the light rays, the less the illumination. This variation in surface illumination is a powerful cue to the three dimensional structure of an object.

The effectiveness of the shading algorithm is related to the complexity of the model of the light sources. Computer graphics light models can include an infinite number of light sources, a variety of wavelengths of light, as well as complex geometries of light sources including point, linear, area, polyhedra and sky light (Nishita and Nakamae, 1985 and 1986). Recently the modelling of light sources has been extended (Nishita *et al.*, 1987) to account for the luminous light intensity variations of light sources.

Natural lighting models, in the case of landscapes under daylight conditions, are normally simplified by assuming that there is only a single parallel or isotropic light source located at infinity (the sun) and conditions are calculated under direct sunlight. However, Nishita and Nakamae (1986) have extended the model to include both direct sunlight and scattered light caused by clouds and other forms of water vapour. This creates a much more realistic image of landscapes under overcast conditions. The lighting model is considered to be a hemisphere with a large radius that acts as a source of diffuse light with non-uniform intensity, thus simulating the varying intensity of sky lighting.

There are two types of light sources apparent in the environment: ambient and direct. Ambient light is light reaching a surface from multiple reflections from other surfaces and the sky, and is normally approximated by a constant illumination on all surfaces regardless of their orientation. This simplified model produces the least realistic results since all surfaces of an object are shaded the same as experienced, for example, under heavily overcast cloud conditions. However, more complex modelling of the ambient or indirect component of lighting has produced enhanced realism using ray tracing techniques to model the contribution from specular inter-reflections and transmitted rays and radiosity techniques to account for complex diffuse inter-reflections.

Direct light is light striking a surface directly from its source without any intermediate reflection or refraction. When direct light, from a point light source say, hits a surface, light reflects off the surface. This can be broken into two reflected components: diffuse and specular, depending on the surface characteristics. The diffuse component is reradiated equally in all directions while the specular component is reradiated in the reflected direction. Both reflected components must be modelled to create realistic images.

The diffuse component is simple to compute since it is independent of the viewing position and is based on Lambert's Law. The specular component is more

complex to model since real objects are not ideal reflectors, reflecting a portion of the light off the axis of the ideal light direction. The first reasonable approximation was an empirical approach (Phong, 1975) which has subsequently been refined by Blinn (1977) and Cook and Torrance (1982). However, due to the computational expense of the later versions, the original approach is still in widespread use.

Images are created by determining the visible surface at each screen pixel location, and then computing the intensity of light leaving a point in the direction of the viewpoint (the eye). In simple shading models, the ambient illumination component is considered to be constant throughout the environment and only directly reflected light is considered, without considering inter-reflections from other surfaces.

Having modelled the intensity of colour at a point, this must now be expanded to encompass a surface. Most rendering algorithms require all surfaces to be decomposed and described as polygonal meshes. Since the normal of a polygon is constant over its area, the shade will also be uniform. Therefore, at polygonal boundaries there will be shade discontinuities that reveal the polygonal approximation of an object's structure. An example of this facetted shading is shown in Plate III(a). Techniques have been developed to overcome this limitation. In Gouraud's approach (Gouraud, 1971) the true surface normals at the vertices of each of the polygons are calculated. When the polygon is converted into pixels, the correct intensities are computed at each vertex and these values are used to linearly interpolate values across the polygonal surface. This approach eliminates the impression of underlying polygons. However, the approach is prone to distorting highlights and the production of Mach band interference effects.

An alternative, but computationally more expensive approach was developed by Phong (1975). This applies linear interpolation of the surface normals over the polygonal area and performs an explicit shading calculation at each of the pixels instead of interpolating the value. The result is normally superior to the Gouraud approach. Plate III(b) is an example of the Gouraud approach and disguises the underlying polygonal surface structure.

Ray tracing. The approaches to shading so far described have used approximations to model the illumination environment. These ignore or excessively approximate illumination components resulting from light reflected from other surfaces. The model does not include the refracted component of light and becomes convoluted in treating complex multiple light sources. A more elegant, but computationally expensive solution is to use ray tracing techniques evolved by Appel (1968), Whitted (1980) and Cook *et al.* (1984).

This approach, which is view dependent, involves tracing a ray of light from the viewpoint through a pixel and into the model where its interaction with objects is analysed (Fig. 6). As the ray strikes the first object in its path, the ray is broken into three components: diffusely reflected light, specularly reflected light and transmitted (refracted) light. Similarly, a ray of light leaving the surface of an object is in general the sum of the three components. Therefore, each collision with an object generates three new rays that should ideally be traced. A diffuse reflection generates an infinite number of rays. Therefore, only rays from specular reflection and refraction are normally continued to be traced. However, Kajiya (1986) has included the modelling of complex diffuse interreflections in ray tracing, previously only modelled by radiosity techniques. This collision and ray splitting procedure is continued until there is a cut-off point imposed based on the number of surfaces encountered or on a minimum energy level. This in effect grows a tree of rays from which the intensity at the pixel can be calculated.

This ray tracing procedure is repeated for rays passing through each pixel centre until the entire scene is generated. To avoid aliasing effects, multiple rays can be passed through a pixel and an adaptive process can be implemented to fire a varying number of rays per pixel, based on the complexity of the objects encountered by the rays.

The highest computational overhead in ray tracing is finding the intersection between rays and the objects in the scene. Performance advances in this area have been achieved by Snyder and Barr (1987) using a hierarchy of lists and three

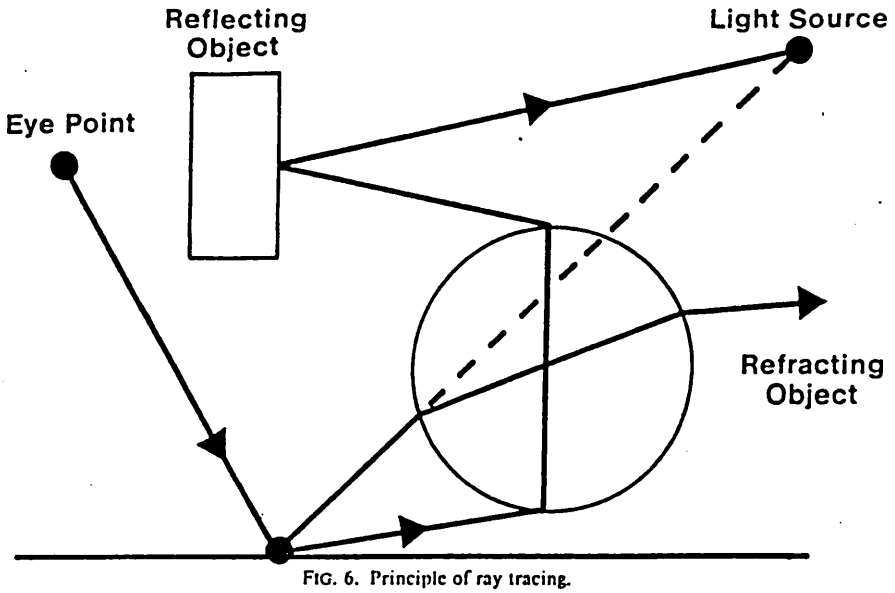


FIG. 6. Principle of ray tracing.

dimensional grids to store the triangular tessellations defining the objects. Using this technique they have rendered a scene containing 400 billion triangles, at a resolution of 512×512 (16 rays per pixel), using 12 hours CPU time on an IBM 4381/Group 12 computer. Further advances will be made using the application of coherence or other properties of the objects being displayed and will ultimately be addressed using Very Large Scale Integration (VLSI) hardware since the problem is ideally suited to parallel architectures. Plate IV illustrates an example of a ray traced image produced by the Jet Propulsion Laboratory using the Pixar modelling system.

Radiosity. Although ray tracing methods produce strikingly realistic images, only radiosity techniques can adequately model the complex, diffuse inter-reflections within an environment. The radiosity method, originally developed for thermal engineering, is based on the Fundamental Law of Conservation of Energy within a closed system. It was first applied to computer graphics by Goral *et al.* (1984) and further extended into more complex environments by Cohen and Greenberg (1985) and Kajiya (1986). The method entails performing a simultaneous global solution for the intensity of light leaving each surface by constructing and solving a set of linear equations describing the transfer of diffuse light energy between all surfaces. The advantages of this method are the increased accuracy in simulating the global illumination effects and the fact that all of the intensity calculations are independent of viewing parameters (Immel *et al.*, 1986), allowing efficient rendering of dynamic sequences.

Recently, Wallace *et al.* (1987) have devised an approach to rendering that combines ray tracing and radiosity techniques in a two pass complementary solution.

Shadows

Shadows are an essential scene component in conveying reality in a computer graphics image. A scene that appears "flat" suddenly comes to life when shadows are included in the scene, allowing the comprehension of spatial relationships amongst objects. Shadows provide the viewer with one of the strongest visual cues needed to interpret the scene. They are not always present in a scene, such as an overcast day simulation, but are often essential.

The intricacies of the shadow algorithm used are related to the complexity of the model of the light sources. Computer graphics shadowing techniques can now handle light sources inside and outside the field of view, an infinite number of light

sources, umbra and penumbra effects, a variety of wavelengths of light, as well as complex geometries of light sources including point, linear, area, polyhedra and sky light (Nishita and Nakamae, 1986).

The fundamental approach is equivalent to the hidden surface algorithm approach. The shadow algorithm determines what surfaces can be "seen" from the light source. Those surfaces that are visible from both the viewpoint and the light source are not in shadow. Those that are visible from the viewpoint, but not from the light source are in shadow. A variety of shadow algorithms have been developed and can be categorised into five groups (Max, 1986): z buffer, area subdivision, shadow volumes, reprocessing and ray tracing, with the most recent advances being made by Reeves *et al.* (1987) using depth maps in the z buffer algorithm.

The modelling of atmospheric illumination has been refined by Max (1986) to simulate the effects of cloud shadow on the terrain and also the atmosphere. Sunlight scattering from water or dust in the air causes the atmosphere to glow. This glow is particularly visible in beams or columns of light in a shadowed atmospheric environment.

Surface Texture Detail

Natural landscape scenes are characterised by features with a wide variety of complex textures. Computer graphics visualisations of landscapes can only achieve an acceptable level of realism if they can simulate these intricate textures. The "flat" shading algorithms, described in the previous section, do not meet this requirement directly since they produce very smooth and uniform surfaces when applied to planar or bicubic surfaces (Plate III(b)). Therefore, the shading approach must be supplemented by other techniques to either directly model or approximate the natural textures.

Explicit modelling. This approach involves creating a more detailed polygonal and colour model of the landscape and surface features to enable a higher level of detail and texture to be visualised. For landscape visualisation, explicit modelling has so far proved impractical due to the size and intricacy of the model that would have to be created to reflect the required level of detail, although a field of grass has recently been modelled using 400 billion polygons (Snyder and Barr, 1987). However, it may be necessary to employ this technique for certain scene objects such as those closest to the viewpoint.

In circumstances where the exact modelling of a terrain surface is not required, then realistic impressions of the microlevel of terrain detail can be derived using data amplification primitives such as fractals which automatically densify the model. These techniques are described further in the section on procedural stochastic models.

Texture mapping. This technique provides the illusion of texture complexity at a reasonable cost. Texture mapping shades a surface mathematically, greatly increasing the visual complexity and richness of an image without the overhead of explicitly modelling those properties. The approach introduced by Catmull and further refined by Blinn and Newell (1976) is a method of "wallpapering" existing polygons with a user defined texture map. This texture map can represent designs, photographs or, more applicable to terrain visualisation, frame grabbed images of natural textures.

The technique has been more generally applied to the rendering process to approximate local lighting effects such as reflections (Blinn and Newell, 1976), shadows (Reeves *et al.*, 1987) and refractions (Paramount Picture Corp., 1987). It is versatile and has become one of the key tools in the fast rendering of realistic images, since it can adequately approximate non-local lighting effects. Stunning landscape images have been created by Cook (1984) using shade tree techniques that more closely integrate the shading and texturing processes. The shader is adaptive and textures can be combined in a flexible environment to produce scene components such as rainbows.

The mapping of a texture map onto a polygon surface causes problems associated with aliasing and inconsistencies between the scale of the polygon and the associated texture map. This has been addressed by the use of symbol filters (Blinn and Newell, 1976) and Gaussian filters (Feibush *et al.*, 1980).

The technique is ideal for the texturing of terrain and surface features. Gardner (1984) has developed a mathematical texturing function, requiring 25 parameters, for application in landscape scenes and has included a translucency component softening the edges of objects when applied to trees and clouds. The results are similar to those achieved by landscape painters. The concept of varying the resolution of the texture map has been developed by Williams (1983). Texture maps at several levels of resolution can be stored and the most appropriate one chosen for the scale of viewing. This technique has been further extended in flight simulation to include dynamic texturing that allows, for example, the simulation of wave patterns caused by helicopter downdraft (Yan, 1985).

Bump mapping. Mapping of pattern arrays in texture mapping simply affects the colouration of the surface. Another way of using texture mapping to increase the illusion of surface complexity is to store further surface properties. Bump mapping (Blinn, 1978) stores surface normal perturbations in the texture map, achieving roughness without explicitly modelling the geometry. The surface normal perturbation should be applied to a surface prior to shading procedures. Both random and regular textures can be achieved, but it is not quite realistic since the texture does not affect the silhouette edges of textured objects. However, it is very effective for many terrain visualisations.

Stochastic fractal models. A further approach for texturing terrain models involves the use of fractal surfaces, a class of irregular shapes that are defined according to the laws of probability and can accurately model the natural terrain (Mandelbrot, 1982). Simple models of the terrain are defined using quadrilaterals or triangles, which are subsequently recursively subdivided to produce more detailed terrain models as described previously. So far the fractal technique has been most effectively applied to the creation of rough and jagged, almost "Tolkein"-like terrain (Miller, 1986). An example is shown in Plate I.

Unlike the texture and bump mapping approaches, this technique results in the creation of a more detailed geometric model of the terrain. Therefore, it has to be applied prior to the rendering process.

Fractal stochastic models (Oppenheimer, 1986) as well as graftals (Smith, 1984) and particle systems (Reeves and Blau, 1985) can also be used to generate landscape surface features and have been effectively used to model trees and grass to a stunning level of realism as shown in Plate V.

Satellite imagery texturing. An approach to directly "clothing" the terrain model with natural textures is to enhance perspective views of the terrain model directly with satellite imagery data. The satellite imagery will contain shadows related to the platform's position and the time of scene capture. These shadows may not correspond with the shadow configuration required to match the enhanced perspective view. Using the terrain geometry, the shadows can be removed from the original image and recast in the new scene. Despite this possibility for inconsistencies in the image, excellent results have been achieved for small scale terrain visualisations.

Atmospheric Attenuation

Due to atmospheric moisture content, objects undergo an exponential decay of contrast with respect to distance from the viewpoint. The decay converges to the sky luminance at infinity. This reduction rate is dependent upon the season, weather conditions, level of air pollution and time. The result is a hazing effect.

The attenuation can be effectively modelled to simulate a variety of atmospheric conditions (Middleton, 1952). The intensity of the component, primary colours decays exponentially by varying amounts, with blue being the most prominent at distance resulting in the blueness of distant objects. By introducing

this atmospheric model into the rendering process, more realistic landscape scenes can be achieved and it introduces an important cue for depth interpretation.

Particles of moisture or dust in the atmosphere, when illuminated, become visible through the scattering and absorption of light. Sunlight falling between clouds or landscape objects causes the atmosphere to glow and is visible in beams or columns of light. These resulting illumination volumes have been modelled by Max (1986) and recently refined by Nishita *et al.* (1987) to include the distribution of luminous intensity of the light sources and variable density of particles.

The interaction of light with suspended particles in the atmosphere also produces beautiful effects on the colours of the sun and sky. Klassen (1987) has recently devised a lighting model that allows the colours of the sun and sky to be calculated for any sun position above the horizon under a variety of atmospheric conditions. This adds an essential degree of realism to the synthesised scene.

SECTION III: APPLICATIONS OF DIGITAL TERRAIN AND LANDSCAPE VISUALISATION

The use of visualisation techniques for both military and civilian applications is an area of significant growth at present and some of the factors which account for this have been outlined previously. Generally these applications are either for small scale, regional visualisation or conversely for large scale, site specific projects. The former group are oriented primarily towards military applications whereas the latter tend to be more biased towards engineering, mining and landscape planning.

Small Scale, Regional Digital Terrain and Landscape Visualisation

For the majority of small scale regional applications the accuracy requirements of the height data are relatively low and may therefore be acquired from existing map sources. Due to the need for wide coverage, it is also common practice for such applications to use one or more of the regional terrain databases which have been discussed previously.

Cartographic Applications

Quality control. One of the main difficulties associated with the production of large terrain databases is that of identifying gross errors at the data acquisition phase. Visualisation techniques have been developed to assist with this activity and a number of alternative strategies are discussed by Faintich (1984). For low resolution analyses of datasets, the use of pseudo-stereoscopic pairs of shaded relief images is suggested as having potential for the identification of spike points. For higher resolution anomaly analyses, the use of convolution (edge enhancement) filters and Fourier analysis techniques are highlighted.

At a simpler level many organisations validate their terrain data by regenerating contours from the terrain database and comparing this with the original contour sheet.

Relief shading. Variations in elevations are normally represented symbolically by contours on topographic maps. Occasionally hill or relief shading is added manually by the cartographer in order to enhance the topography. Several authors have considered the use of regional DTMs for automating the relief shading process and Schachter (1980) and Horn (1982) provide some research results.

More recently Scholz *et al.* (1987) have discussed the procedure adopted to automate the production of relief shading on 1:250 000 scale aeronautical charts (Plate VI). The algorithm used simulates the effect of illumination by computing the slope and aspect for each point on the surface. The authors estimate cost savings of up to 75 per cent compared with the manual techniques previously used.

Remote Sensing Applications

DTM data have several possible applications in remote sensing, particularly

when regional data are available. For example they have been used to improve the geometric quality of satellite imagery, or to improve the accuracy of the supervised classification techniques which are currently being used. In the latter case the DTM is used to remove the variations in radiance caused by varying surface gradients, since these effects are additional to the radiance caused by changes in land cover. In the field of visualisation, the main use of remotely sensed imagery is in the production of realistic perspective views of the terrain using DTM data merged with aerial or satellite imagery.

Registration of DTMs and satellite imagery. Various theoretical difficulties associated with the precise alignment of Landsat MSS imagery and DTMs are considered in Horn and Bachman (1978). Although no perspective views were created, they discuss the geometric and radiometric transformations which are necessary in order to relate the two data sets.

Perspective scene generation. Many of the basic techniques discussed in Horn and Bachman (1978) have been used to pre-process imagery before generating perspective views from the merged data sets. Work on this activity using satellite imagery has been carried out by IBM (perspectives of Landsat MSS (Bernstein *et al.*, 1984)), the Institute for Image Processing and Computer Graphics in Graz (perspectives using SIR-B imagery (Leberl *et al.*, 1986)), the Remote Sensing Laboratory at the University of Zurich (Meier, 1987) (perspectives of Landsat TM imagery) and the Geospectra Corp. (perspectives from SPOT panchromatic imagery (Anon, 1987)) among others. This process has recently been taken a stage further by the production of video based perspective animations of the terrain (Muller *et al.*, 1988). In this case the triangulated terrain model is textured using either panchromatic SPOT imagery or false colour composite Landsat TM data. Animation is produced by generating multiple views along a defined "flight path".

Whiteside *et al.* (1987) discuss the use of a digital perspective scene generation (PSG) technique to simulate stereoscopic views of the terrain using vertical aerial photography. The photography was initially digitised using an Optronics scanner and merged with DTED values. Monoscopic views were then generated using a modified perspective transformation applied to each input pixel. Stereoscopic viewing of the two scenes was then generated using electronic shutters to alternately display single 512×512 pixel images to the user.

Military Applications

DTMs are used extensively by the armed services for a number of applications ranging from general surveillance and intelligence gathering to radar simulation, battlefield planning and simulation and missile guidance. Although the techniques used by the military do not differ substantially, their needs in this field differ significantly from those of the civilian users. Firstly, there is a need for more accurate and up to date terrain data. Secondly, there is a requirement for fast and effective methods of displaying the analyses of the terrain database. Finally it is essential to have efficient methods of communicating and distributing these results to users. Given the wide ranging requirements of this group, the following section will only highlight a selected number of examples of the use of terrain and landscape visualisation in military situations.

Military simulation. The ability to be able to simulate the tactical scenarios used in battle is a technique which is being used by increasing numbers of military organisations. Systems such as the Battle Area Tactical Simulation (BATS) from Merit Technology enable a pilot or commander to plan and execute a mission using sophisticated display techniques and simulate the effect of enemy forces. Other systems such as the Scicon Viewfinder (Thorpe, 1986) can also be used to simulate, for example, the deployment of a ground radar and enable the coverage of the sensor to be compared for various locations.

Terrain analysis. The development of portable field systems which can generate terrain information such as intervisibility analyses, perspective views, slope and aspect maps and assessments of cross country movement is a topic which has been widely reported in the USA. Tindall *et al.* (1981) describe some early work on the

Field Exploitation of Elevation Data (FEED). More recently Sither (1987) has discussed the Terrain Analyst Work Station (TAWS) being developed by the US Army Engineer Topographic Laboratories (USAETL).

Command, Control, Communications and Intelligence (C^3I) Systems. For the future it seems likely that some of the most sophisticated uses of visualisation are likely to be in the field of C^3I. These computer systems are responsible for collecting and processing information on enemy locations, strength and movements and using this information to determine the necessary military actions which will be required. A terrain model is an essential component for making such decisions. Furthermore the use of many of the display and enhancement techniques described so far are central to the effective communication of the decisions which are reached.

Avionics

Unlike the static, single image views of the landscape which have been discussed so far, the visualisation needs in the field of avionics require dynamic, multiple displays of the terrain. Whitehead *et al.* (1986), for example, provide details of work which involved the generation of ridge lines from DTED data for overlaying onto forward looking infrared (FLIR) imagery. This provided a simplified appreciation of the terrain ahead, which is particularly useful in situations where the quality of the FLIR imagery becomes degraded (such as in poor weather conditions). This image and line display superimposition is normally accomplished by using a head up display (HUD).

Discussions of the development of fully digital cockpits of the future are provided by Warwick (1986) and Broadbent (1986). The latter author describes the development of a large 500 × 380mm combat aircraft cockpit display which will, in real time, create fully coloured and textured perspective views of the terrain.

Flight simulation

In terms of computer processing requirements, the most sophisticated forms of visualisation are undoubtably those required for flight simulation. A detailed discussion of these systems is beyond the scope of this paper and thus only a brief description of the main characteristics of the data and processing techniques will be provided. Further details can, however, be obtained from Yan (1985) and Moxon (1986).

Terrain modelling for flight simulation. Although flight simulators were invented over 50 years ago to provide a more cost effective method of training aircraft pilots, it is only in the past decade that computer generated imagery (CGI) has been used extensively. Prior to this, terrain and cultural information were normally created by photographing model boards using servo-driven television cameras. Nowadays the models are generated from either DTED databases or by digitising existing topographic maps.

In contrast to virtually all other applications of terrain and landscape visualisation, it is common practice in flight simulation to create a hierarchical series of DTMs to represent the region to be modelled rather than utilising a single data set. This is necessary in order to model the perceived change in the degree of detail on the terrain as the pilot moves across the model. The DTM is arranged to consist of a series of triangular facets. Precise details of the algorithms used for this process are not available in the literature. Other scene elements are normally defined by a series of polygonal elements.

Data processing for flight simulation. Data processing for flight simulators involves a combination of special purpose, distributed computing hardware together with sophisticated software to ensure that a flicker-free, full colour, textured, hidden surface image can be computed in 1/30 s. In order to achieve this, a system consisting of the components illustrated by Fig. 7 is used.

The scene manager is the overall control system for the other three main components of the simulator and is responsible for retrieving data from disc

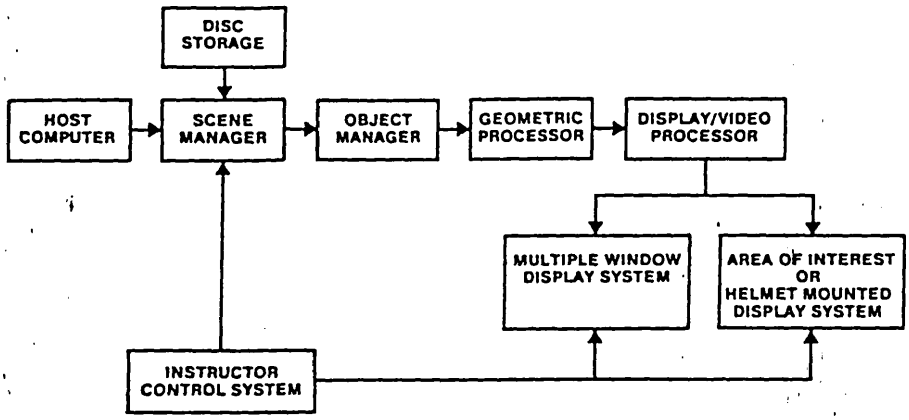


FIG. 7. Main system components of a flight simulator.

associated with the current pilot position. The next component is the object manager which determines the scene elements visible to the pilot from a particular eyepoint position. This process involves an ordering of the scene elements into visual priority order so as to solve the problems associated with hidden surface removal. The object manager processer requires about 40 million instructions per second (MIPS) of computing resources.

Having approximately sorted the polygonal scene elements, the geometric processor transforms the scene polygons to channel co-ordinates clipped to the display window. For this activity approximately 60 MIPS of processing power are required. Finally the video/display processor computes the colour and shade of each pixel. Atmospheric attenuation and the addition of texture may also be modelled at this stage. Since a single frame of imagery will typically consist of 1000 lines of 1000 pixels and the refresh rate is 1/30s, simply to model the colour of each pixel involves changes to 30 million pixels every second. Yan (1985) estimates that the processing requirement of a typical video/display processor capable of producing realistic textured and anti-aliased imagery is of the order of 10000 MIPS. Plate VII illustrates a typical scene from a modern flight simulator.

The final component of the system is the display sub-system. Two distinct methods of displaying CGI are currently available: those which use multiple windows (normally three) and those which use a helmet mounted display system. The former are the more common and provide fields of view of 120° to 180° horizontally and 40° to 90° vertically. Although such display systems are adequate for the current generation of simulators, as the resolution requirements and processing facilities improve these systems will be unable to display the wide fields of view and high resolutions which will be required, particularly for military applications. In order to accommodate these needs, advanced area-of-interest (AOI) helmet mounted display systems are currently being developed which only provide high resolution within a small field of view around the line of sight of the pilot. Typical examples are the eye slaved projected raster inset (ESPRIT) system (Link) and the eye slaved variable acuity lens display system (McDonnell Douglas).

LARGE SCALE, SITE SPECIFIC TERRAIN AND LANDSCAPE VISUALISATION

Landscape Planning: Visual Impact Analysis

Growing public awareness of environmental issues has recently been strengthened by the European Economic Community's Directive on the "Assessment of the Effects of Certain Public and Private Projects on the Environment". This Directive will force certain proposed changes to the landscape to be publicly assessed for environmental impact. A component of this environmental audit is a statement on the visual intrusion of proposed landscape changes. Consequently, projects such as

road construction, transmission line routing and open cast mining as well as more dynamic phenomena such as forestry may need to be judged visually.

Traditionally, landscape visualisation techniques have involved the building of physical models or the creation of artists' impressions. However, these are time consuming to create and are inherently inaccurate and inflexible once created. In order to more accurately quantify the level of visual intrusion, computer graphics modelling and visualisation techniques are increasingly being used in the planning and design of landscape projects. These new approaches allow more accurate visualisations and more analytical assessments of visual intrusion to be determined. Due to the flexibility of the approach, many more proposed designs can be evaluated, resulting in a more refined design solution. Turnbull *et al.* (1987) pioneered the development of a computer aided visual impact analysis system (CAVIA) that has been used, for example, to provide evidence at public inquiries related to electricity transmission line routing through environmentally sensitive landscapes (Turnbull *et al.*, 1986) (Plate VIII). Projects are typically performed at the sub-regional level with areas up to 80×40 km being analysed. The approach uses DTMs, landscape features and proposed design objects to produce an estimate of the visual intrusion. This visual intrusion toolkit includes intervisibility analysis to produce levels of visual impact, dead ground analysis, identification of the portions of the landscape forming a backcloth for the design object, situations where the design object appears above the landscape horizon and the identification of optimal locations for vegetation screen placement.

One of the visualisation techniques used by CAVIA and other landscape planning systems is photomontaging. A photomontage is a physical or image composite of photographs of the existing landscape with a registered computer generated image of the proposed design object(s). Plate VIII shows a photomontage of a proposed transmission line route. In this approach only the proposed design objects have to be rendered, avoiding the rendering of the intricate terrain and landscape detail. However, to achieve total image fidelity, the computer rendered portion of the image must effectively "merge" with the photographic image. Therefore, the atmospheric and distancing effects apparent in the photographic image must be inherited by the computer generated image. Nakamae *et al.* (1986) have developed techniques for merging image components to compensate for fog and aliasing effects. However, problems still occur when the design object's interaction with the existing landscape is complex, such as road earthworks, and where foreground features, such as trees, partially obscure the design objects. Future solutions to this problem will use a hybrid approach to rendering a photomontage image, incorporating ray tracing, frame grabbed images and textures, image processing and pixel painting techniques.

Road/Traffic Engineering

Visualisation has found a number of interesting applications in the field of road design. Many road engineering design systems are now offering visualisation capabilities. These form an integral part of the design process and allow the design to be subjectively assessed and refined for safety and visual intrusion in the context of its environment.

The Transport and Road Research Laboratory of the UK has developed (Cobb, 1985) a system to model and visualise road designs. There are various applications of the system.

Road lighting scheme design. In complex road designs or under environmental constraints, the design of an efficient road lighting scheme can be enigmatic. This is the perfect design environment for the application of visualisation techniques where the designer can directly examine the results of his design under a variety of atmospheric conditions.

Road safety. Visualisation tends to imply aesthetic appearance. However, in this application, visualisation is concerned with the perceptual problems encountered by road users. Factors contributing to potential perceptual problems could be line of slight difficulties, incorrectly positioned street furniture or poor lane

markings. In accident prone sites, the system can be used to identify possible contributing factors and to evaluate solutions. Plate IX illustrates an example of a road visualisation.

Street furniture design and placement. New designs for road signs and street furniture and their optimal positioning can be evaluated.

This application has been taken one step further by the West German car manufacturer, Daimler-Benz, who have added dynamics and created a car simulator. The driver can experience driving a range of cars under a variety of driving conditions.

Urban Design

In recent years urban renewal has become an activity increasingly exposed to and controlled by public and royal opinon. Architects, in an attempt to alleviate public fears of a continuation of the "Kleenex box" era, have turned to computer generated images to convince the public of the merits of their proposed building designs. Computer generated visualisations have become a fashionable marketing tool.

Although the architectural industry was one of the first application areas where computer aided design (CAD) techniques were applied, it is only recently that tools for creating high quality visualisations of the resulting building designs have been made available. This capability is a natural extension of the CAD process and many CAD system vendors are now supplying this capability as an integral part of their system or as an interface to foreign visualisation packages. Plate X shows an example of Intergraph's ray tracing package (Render) for their architectural customers.

Typically, architectural visualisations are not just isolated previews of the proposed building, but also include the contextual surroundings to allow appraisal of its applicability to the existing character of its urban environment. This usually involves the creation of a three dimensional model of the terrain, streets, street furniture and buildings in the immediate vicinity of the site. This approach was recently adopted by Arup Associates in their submission for the development of Paternoster Square in London. The computer model of the development was created on a McDonnell Douglas GDS system and supplemented with the surrounding urban details through photogrammetric measurement and direct input from the Ordnance Survey's digital map series. An example of a derived visualisation looking over St. Paul's Cathedral is shown in the Frontispiece.

Since national large scale map series of urban areas normally lack the essential height information of the buildings, users either are being forced into the expense of an original survey or are being inhibited in the use of these techniques. In an attempt to encourage the use of these techniques, the City of Glasgow in Scotland sponsored the creation of a digital model of the downtown core of the city (Herbert, 1987). This was performed by the University of Strathclyde and the model included the terrain, all major buildings, streets and rivers. Potential users are now being encouraged to use the "city database" to site communications equipment, to derive tourist information as well as to create visualisations for planning purposes.

Conclusions

Computer generated visualisations of digital terrain and landscape scenes are now widely accepted in many applications as efficient technical analysis, design and marketing tools. Visualisation techniques have released the world from its traditional two dimensional approaches to display and, in so doing, have highlighted the three dimensional deficiencies in our sources of data in terms of availability and accuracy.

Many large scale maps have minimal terrain elevation data and the majority ignore the three dimensional aspects of surface features. Although the photogrammetric acquisition of DTM data is well established, to date much less attention has

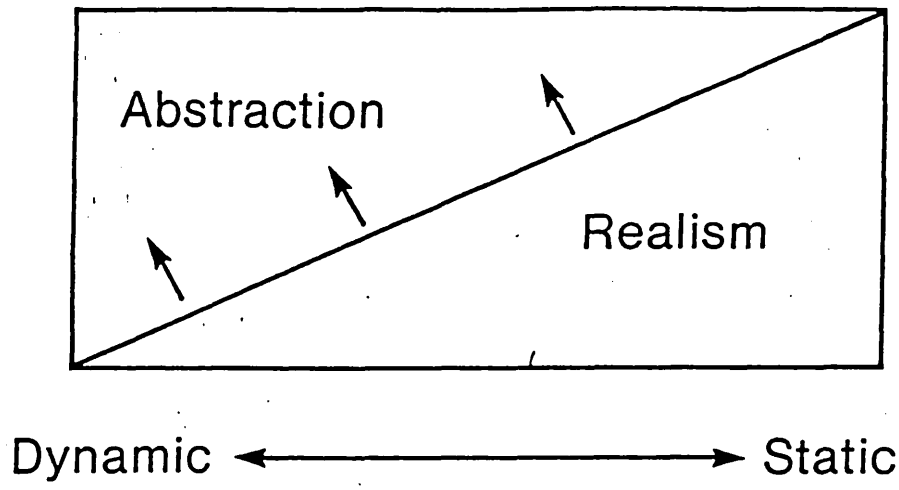


FIG. 8. Future development of visualisation techniques.

been directed towards the photogrammetric acquisition of three dimensional descriptions of significant objects on the terrain. In many cases the three dimensional co-ordinates of features such as buildings may be observed, although often the Z co-ordinate values are deemed superfluous and do not form part of the recorded dataset. Photogrammetrists therefore have an important role to play in the acquisition of this form of data, particularly in urban environments and at close range. Indeed the lack of data is currently inhibiting the wider application of many of these techniques.

In the GIS environment, visualisation techniques are recognised as an invaluable system component, aiding in the interpretation of spatially related phenomena and complex data analyses that takes the GIS a step beyond two dimensional polygonal overlay analyses. Many of the GIS vendors are including this capability in their systems to help cope in our understanding of the "fire hose" of data being produced by contemporary sources such as satellites.

Over the past decade, substantial advances in the realism of computer graphics generated visualisations have been achieved. The techniques of ray tracing and radiosity in modelling global illumination are currently forming the leading edge. Despite this progress, the visualisation of natural phenomena inherent in landscape scenes are still simplistic, forming "realistic abstractions" or "abstract realism". Our visual system is specialised and highly skilled at recognising them in all their subtle forms, making the objective even more elusive.

At present, apart from flight simulation, the production of realistic scenes using the visualisation techniques discussed in this paper has been largely restricted to static applications (Fig. 8). Similarly at present, dynamic visualisations are primarily restricted to wireframe "abstractions". For the future, however, the ability to perform dynamic visualisation with high levels of scene realism will become an increasingly attainable goal, although the explicit modelling of many natural phenomena will continue to be a significant problem. For example, many of the current landscape visualisation products use photographs, frame grabbed images or texturing techniques of scene components to avoid the computer graphics synthesis of natural phenomena. Some creators of the most realistic landscape images synthesised to date have used their artistic licence by over approximating the simulation of natural phenomena and "faking" low interest scene components with stochastic or texture map processes. Although not creating a scientifically correct image, they have achieved surprisingly effective results using what could be

termed artistic techniques. This emphasises the need for the scientific and artistic communities to work together towards this goal. Certainly more research is needed in perceptual comparison of real scenes and their synthetic image counterparts, with particular emphasis on the synthesis of natural colours and textures.

Inevitably, the successful approach will resort to true simulation of the phenomena based on the laws of physics. However, this is a computationally expensive option that in most cases is currently prohibitive, but will become feasible with the guaranteed increase in processing power. This is already happening with the recent release of parallel processing architectures, minisupercomputer based workstations, high bandwidth links between supercomputers and workstations and the availability of customised VLSI for specific applications. The present techniques of approximating or faking (Frenkel, 1988) will be displaced by progressive refinements of the simulation model. This approach has been endorsed by the McCormick *et al.* (1987) initiative on visualisation in scientific computing.

Despite realism being a distant target, it acts as a convenient measure of our techniques and understanding and will continue to be relentlessly pursued to our continuing benefit.

Acknowledgement

The authors wish to express their sincere thanks to the many individuals who have provided information for inclusion in this review paper. In particular, they would also like to thank the following companies who very kindly provided financial support which has enabled a number of colour photographs to be included in the paper: Cambridge Interactive Systems (CIS), Intergraph (GB), Laser-Scan Laboratories, McDonnell Douglas Information Systems, MOSS Systems, Sun Microsystems and the Turnbull Jeffrey Partnership.

REFERENCES

Anon., 1987. A new view produced from SPOT imagery. *Spotlight*, 1(4): 4–5.

Appel, A., 1968. Some techniques for shading machine renderings of solids. *AFIPS JSCC Proceedings*, 32: 37–45.

Armstrong, A., 1976. A three-dimensional simulation of slope forms. *Zeitschrift für Geomorphologie*, N. F. Supplementband, 25: 20–28.

Balce, A. E., 1987. Determination of optimum sampling interval in grid digital elevation models (DEM) data acquisition. *Photogrammetric Engineering and Remote Sensing*, 53(3): 323–330.

Beerenwinkel, R., Bonjour, J.-D., Hersch, R. D. and Kölbl, O., 1986. Real-time stereo raster injection for photogrammetric plotting. *International Archives of Photogrammetry and Remote Sensing*, 26(4): 99–109.

Bernstein, R., Lotspiech, J. B., Myers, H. J., Kolsky, H. G. and Lees, R. D., 1984. Analysis and processing of Landsat-4 sensor data using advanced image processing techniques and technologies. *IEEE Transactions on Geoscience and Remote Sensing*, GE-22(3): 192–220.

Blinn, J. F., 1977. Models of light reflection for computer synthesized pictures. *ACM Computer Graphics*, 11(2): 192–198.

Blinn, J. F., 1978. Simulation of wrinkled surfaces. *Ibid.*, 12(3): 286–292.

Blinn, J. F. and Newell, M. E., 1976. Texture and reflection in computer generated images. *Communications of the ACM*, 19(10): 542–547.

Broadbent, S., 1986. McDonnell Douglas' big picture show. *Jane's Defence Weekly*, 6(18): 1125–1126.

Burrough, P., 1985. Fakes, facsimilies and facts: fractal models of geophysical phenomena. *Science and Uncertainty* (Editor S. Nash). Science Reviews Ltd. 219 pages: 151–169.

Butland, J., 1979. Surface drawing made simple. *Computer-Aided Design*, 11(1): 19–22.

Carpenter, L., 1984. The A-buffer, an antialiased hidden surface method. *ACM Computer Graphics*, 18(3): 103–108.

Clarke, K. C., 1987. *Scale-based simulation of topography*. Presented paper to Auto-Carto 8. 9 pages.

Cobb, J., 1985. A new route for graphics. *Computer Graphics '85*. Online Publications, Pinner, UK. Pages 173–183.

Cohen, M. F. and Greenberg, D. P., 1985. The hemi-cube: a radiosity solution for complex environments. *ACM Computer Graphics*, 19(3): 31–40.

Cook, R. L., 1984. Shade trees. *Ibid.*, 18(3): 223–231.

Cook, R. L. and Torrance, K. L., 1982. A reflectance model for computer graphics. *ACM Transactions on Graphics*, 1(1): 7–24.

Cook, R. L., Porter, T. and Carpenter, L. 1984. Distributed ray tracing. *ACM Computer Graphics.*, 18(3): 137–145.

Cook, R. L., Carpenter, L. and Catmull, E., 1987. The Reyes image rendering architecture. *Ibid.*, 21(4): 95–102.

CRAIG, R. G., 1980. A computer program for the simulation of landform erosion. *Computers and Geosciences*, 6: 111–142.

CROW, F. C., 1977. The aliasing problem in computer-generated shaded images. *Communications of the ACM*, 20(11): 799–805.

CROW, F. C., 1982. Computational issues in rendering anti-aliased detail. *COMPCON Conference Proceedings*. Pages 238–244.

DUBAYAH, R. O. and DOZIER, J., 1986. Orthographic terrain views using data derived from digital elevation models. *Photogrammetric Engineering and Remote Sensing*, 52(4): 509–518.

DURRANT, A., 1987. *Intelligent terrain modelling—the CV Medusa GIS Approach.* Presented paper to Short Course on Terrain Modelling in Surveying and Civil Engineering, University of Surrey and University of Glasgow. 10 pages.

DURRANT, A., 1988. *Computer aided simulation of hydrologic tailings disposal.* Loughborough University of Technology thesis in preparation.

EL ASSAL, A. A. and CARUSO, V. M., 1983. Digital elevation models. *U.S. Geological Survey Circular, 895-B.* 40 pages.

EUROPEAN ECONOMIC COMMUNITY, 1985. Council directive of the 27th June on the assessment of the effects of certain public and private projects on the environment. *(85/337/EEC).*

FAINTICH, M. B., 1984. State-of-the-art and future needs for development of digital terrain models. *International Archives of Photogrammetry and Remote Sensing*, 25(3a): 180–196.

FEIBUSH, E. A., LEVOY, M. and COOK, R. L., 1980. Synthetic texturing using digital filters. *ACM Computer Graphics*, 14(3): 294–301.

FOURNIER, A., FUSSEL, D. and CARPENTER, L., 1982. Computer rendering of stochastic models. *Communications of the ACM*, 25(6): 371–384.

FREDERIKSEN, P., JACOBI, O. and KUBIK, K., 1985. A review of current trends in terrain modelling. *I.T.C. Journal*, 1985-2:101–106.

FRENKEL, K. A., 1988. The art and science of visualizing data. *Communications of the ACM*, 31(2): 111–121.

GARDNER, G. Y., 1984. Simulation of natural scenes using textured quadric surfaces. *ACM Computer Graphics*, 18(3): 11–20.

GELBERG, L. M. and STEPHENSON, T. P., 1987. Supercomputing and graphics in the earth and planetary sciences. *IEEE Computer Graphics and Applications*, 7(7): 26–33.

GORAL, C. M., TORRANCE, K. E., GREENBERG, D. P. and BATTAILE, B., 1984. Modeling the interaction of light between diffuse surfaces. *ACM Computer Graphics*, 18(3): 213–222.

GOURAUD, H., 1971. Continuous shading of curved surfaces. *IEEE Transactions on Computers*, C-20(6): 623–629.

GREGORY, R. L., 1977. *Eye and brain* (Third Edition). Weidenfeld and Nicolson, London. 256 pages.

GRIFFIN, M. W., 1987. A rapid method for simulating three-dimensional fluvial terrain. *Earth Surface Processes and Landforms*, 12(1): 31–38.

GUPTA, S., SPROULL, R. and SUTHERLAND, I., 1981. AVLSI architecture for updating raster scan displays. *ACM Computer graphics*, 15(3): 71–78.

HERBERT, M., 1987. A walk on the Clydeside. *CADCAM International*, 6(4): 41–42.

HORN, B. K. P., 1982. Hill shading and the reflectance map. *Geo-Processing*, 2(1): 65–144.

HORN, K. P. and BACHMAN, B. L., 1978. Using synthetic images to register real images with surface models. *Communications of the ACM*, 21(11): 914–924.

IMMEL, D. S., COHEN, M. F. and GREENBERG, D. P., 1986. A radiosity method for non-diffuse environments. *ACM Computer Graphics*, 20(4): 133–142.

KAJIYA, J. T., 1986. The rendering equation. *Ibid.*, 20(4): 143–150.

KLASSEN, R. V., 1987. Modeling the effect of the atmosphere on light. *ACM Transactions on Graphics*, 6(3): 215–237.

LEBERL, F. W., DOMIK, G., RAGGAM, J. and KOBRICK, M., 1986. Radar stereomapping techniques and application to SIR-B images of Mt. Shasta. *IEEE Transactions on Geoscience and Remote Sensing*, GE-24(4): 473–480.

MANDELBROT, B. B., 1975. Stochastic models for the earth's relief, the shape and fractal dimensions of coastlines, and the number area rule for islands. *Proceedings of the National Academy of Science U.S.A.*, 72(10): 2825–2828.

MANDELBROT, B. B., 1982. *The fractal geometry of nature.* W. H. Freeman, San Francisco. 468 pages.

MANDELBROT, B. B. and VAN NESS, J. W., 1968. Fractional Brownian motions, fractional noises and applications. *SIAM Review*, 10(4): 422–437.

MAX, N. L., 1986. Atmospheric illumination and shadows. *ACM Computer Graphics*, 20(4): 117–124.

MCAULAY, I. C., 1988. *Visual descriptors. A design tool for visual impact analysis.* Ph.D. Thesis., Plymouth Polytechnic. (Unpublished). 458 pages.

MCCORMICK, B. H., DEFANTI, T. A. and BROWN, M. D., 1987. Visualisation in scientific computing—a synopsis. *IEEE Computer Graphics and Applications*, 7(7): 61–70.

MCCULLAGH, M. J. 1988. Terrain and surface modelling systems: theory and practice. *Photogrammetric Record*, 12(72): 747–779.

MEIER, E., 1987. TM colour composite in perspective viewing. *Earth Observation Quarterly*, 18: 2 pages.

MIDDLETON, W. E. K., 1952. *Vision through the atmosphere.* University of Toronto Press, Toronto. 250 pages.

MILLER, G. S. P., 1986. The definition and rendering of terrain maps. *ACM Computer Graphics*, 20(4): 39–48.

MOXON, J., 1986. The race for realism. *Flight International*, 22nd February. 129: 34–42.

MULLER, J.-P. and SAKSONO, T., 1986. An evaluation of the potential role of fractals for digital elevation model generation from spaceborne imagery. *International Archives of Photogrammetry and Remote Sensing*, 26(4): 63.

MULLER, J.-P., DAY, T., KOLBUSZ, J., DALTON, M., RICHARDS, S. and PEARSON, J. C., 1988. Visualisation of topographic data using video animation. *Ibid.*, 27(4): 602–615.

NAKAMAE, E., HARADA, K., ISHIZAKI, T. and NISHITA, T., 1986. A montage method: the overlaying of the computer generated images onto a background photograph. *ACM Computer Graphics*, 20(4): 207–214.

NEWELL, M. E., NEWELL, R. G. and SANCHA, T. L., 1972. A new approach to the shaded picture problem. *Proceedings of the ACM National Conference*, Boston. Pages 443–452.

NEWMAN, W. M. and SPROULL, R. F., 1979. *Principles of interactive computer graphics* (Second Edition). McGraw-Hill International Book Company, London. 541 pages.

NISHITA, T., and NAKAMAE, E., 1985. Continuous tone representation of three-dimensional objects taking account of shadows and interreflection. *ACM Computer Graphics*, 19(3): 23–30.

NISHITA, T., and NAKAMAE, E., 1986. Continuous tone representation of three-dimensional objects illuminated by sky light. *Ibid.*, 20(4): 125–132.

NISHITA, T., MIYAWAKI, Y. and NAKAMAE, E., 1987. A shading model for atmospheric scattering considering luminous intensity distribution of light sources. *Ibid.*, 21(4): 303–310.

OPPENHEIMER, P. E., 1986. Real time design and animation of fractal plants and trees. *Ibid.*, 20(4): 55–64.

PARAMOUNT PICTURE CORP., 1985. Stained glass man sequence by Pixar and Lucas Films Ltd.

PETRIE, G. and KENNIE, T. J. M., 1988. Digital terrain modelling. Chapter 10 in *Engineering surveying technology* (Editors T. J. M. Kennie and G. Petrie). Blackie Publishing Co. In press.

PHONG, B. T., 1975. Illumination for computer generated pictures. *Communications of the ACM*, 18(6): 311–317.

REEVES, W. T. and BLAU, R., 1985. Approximate and probabilistic algorithms for shading and rendering structured particle systems. *ACM Computer Graphics*, 19(3): 313–322.

REEVES, W. T., SALESIN, D. H. and COOK, R. L., 1987. Rendering anti-aliased shadows with depth maps. *Ibid.*, 21(4): 283–291.

ROY, A. G., GRAVEL, G. and GAUTHIER, C., 1987. *Measuring the dimension of surfaces: a review and appraisal of different methods.* Presented paper to Auto-Carto 8. Pages 68–77.

SASADA, T. T., 1987. Drawing natural scenery by computer graphics. *Computer-Aided Design*, 19(4): 212–218.

SCHACHTER, B., 1980. Computer generation of shaded relief maps. In *Map data processing* (Editors H. Freeman and G. G. Pieroni). Academic Press, New York. 374 pages: 355–368.

SCHOLZ, D. K., DOESCHER, S. W. and HOOVER, R. A., 1987. *Automated generation of shaded relief in aeronautical charts.* Presented paper to ASPRS Annual Convention. 4: 14 pages.

SCHWARZ, M. W., COWAN, W. B. and BEATTY, J. C., 1987. An experimental comparison of RGB, YIQ, LAB, HSV, and opponent colour models. *ACM Transactions on Graphics*, 6(2): 123–158.

SECHREST, F. S. and GREENBERG, D. P., 1981. A visible polygon reconstruction algorithm. *ACM Computer Graphics*, 15(3): 17–27.

SITHER, M. A., 1987. *Terrain analyst work station (TAWS) demonstrations.* Presented paper to ASPRS Annual Meeting, Baltimore. 9 pages.

SMITH, A. R., 1984. Plants, fractals, and formal languages, *ACM Computer Graphics*, 18(3): 1–10.

SNYDER, J. M. and BARR, A. H., 1987. Ray tracing complex models containing surface tessellations. *Ibid.*, 21(4): 119–128.

TEMPFLI, K., 1980. Spectral analysis of terrain relief for the accuracy estimation of digital terrain models. *I.T.C. Journal*, 1980–3: 478–510.

TEMPFLI, K., 1986. Composite/progressive sampling—a program package for computer supported collection of DTM data. *Papers at ACSM-ASP Meeting*, Washington, D.C., 4 (Photogrammetry). 519 pages: 202–209.

THORPE, L. W., 1986. Graphical display and manipulation in two and three dimensions of digital cartographic data. *Proceedings Auto Carto London*, 2: 418–430.

TINDALL, T. O., ROSENTHAL, R. L. and JONES, C., 1981. Three dimensional terrain graphics for the battlefield. *41st ACSM-ASP Convention*, Washington. Pages 538–544.

TURNBULL, W. M., MAVER, T. W. and GOURLAY, I., 1986. Visual impact analysis: a case study of a computer based system. *Proceedings Auto Carto London*, 1: 197–206.

TURNBULL, W. M., MCAULAY, I. C. and MCLAREN, R. A., 1987. *The role of terrain modelling in computer aided landscape design.* Presented paper to Short Course on Terrain Modelling in Surveying and Civil Engineering, University of Surrey and University of Glasgow. 19 pages.

VOSS, R., 1985. Random fractal forgeries: from mountains to music. *Science and Uncertainty* (Editor S. Nash). Science Reviews Ltd., 219 pages: 69–86.

WALDERN, J. D., 1986. *Computer imaging using stereo techniques.* Paper presented to British Computer Society Conference on Documentation Displays Group Seminar on State of the Art in Stereo and Terrain Modelling, London. 16 pages.

WALLACE, J. R., COHEN, M. F., and GREENBERG, D. F., 1987. A two-pass solution to the rendering equation: a synthesis of ray tracing and radiosity methods. *ACM Computer Graphics*, 21(4): 311–320.

WARWICK, G., 1986. Navigating by stealth. *Flight International*, 15th November. 131: 34–37.

WILLIAMS, L., 1983. Pyramidal parametrics. *ACM Computer Graphics*, 17(3): 1–11.

WHITEHEAD, A. M., THOMAS, W. P. and WILKINS, M. P., 1986. *Integrated flying aid and mission displays for modern combat aircraft incorporating a digital data base.* Paper presented to IEEE/AIAA 7th Digital Avionics Systems Conference, Fort Worth, Texas. 15 pages.

WHITESIDE, A., ELLIS, M. and HASKELL, B., 1987. *Digital generation of stereoscopic perspective scenes.* Paper presented to SPIE Conference on True Three Dimensional Imaging Techniques and Display Technologies. 761: 7 pages.

WHITTED, T., 1980. An improved illumination model for shaded display. *Communications of the ACM*, 23(6): 343–349.

YAN, J. K., 1985. Advances in computer-generated imagery for flight simulation. *IEEE Computer Graphics and Applications*, 8(6): 37–51.

Résumé

On décrit dans cette communication des méthodes permettant de constituer les modèles numériques du terrain et ceux qui concernent le paysage et d'en tirer des visualisations particulières, propres à un site ou à une région.

On examine un certain nombre d'approches possibles en comparant leur précision relative. Puis on analyse les processus infographiques que l'on peut utiliser pour représenter ces modèles.

Une attention particulière est portée aux méthodes qui introduisent la perception de la profondeur, éliminent les parties cachées, s'attachent à l'estompage et aux textures. On étudie également les effets de l'atmosphère et les méthodes de tracé des ombres. Divers exemples de visualisation sont fournis en ce sens, comprenant des applications à grande échelle particulières à un site et des exemples régionaux à petite échelle.

On évalue en conclusion les procédés de visualisation en mettant en évidence les développements que l'on peut attendre à l'avenir.

Zusammenfassung

Beschreibung von Verfahren zur Erzeugung digitaler Gelände- und Landschaftsmodelle für regionale und bestimmte örtliche Visualisierungen. Diskussion einer Anzahl alternativer Lösungen und ihrer relativen Genauigkeit. Dem schließt sich ein Überblick über Verfahren der Digitalgrafik an, die genutzt werden kann, um solche Modelle wiederzugeben. Besondere Aufmerksamkeit gilt den Verfahren zum Aufzeigen von Tiefen, der Beseitigung verdeckter Oberflächen, der Schattierung und Texturierung. Die Einflüsse der Atmosphäre und Methoden zur Schattenerzeugung werden ebenfalls überprüft. Dann erfolgt eine Darstellung vielfältiger Anwendungen solcher Visualisierungen, worunter sich kleinmaßstäbige regionale Beispiele und großmaßstäbige lokalspezifische Anwendungen befinden. Im Ergebnis wird der Prozeß der Visualisierung eingeschätzt, und es werden mögliche künftige Entwicklungen auf diesem Gebiet aufgezeigt.

DISCUSSION

Mr. J. E. Farrow (Chairman): We have just had a fascinating preview of what can happen to the digital data that many of us are involved in capturing.

Mr. Chandler: You mentioned some of the computational problems involved with the production of realistic graphical images. Which particular component of the graphical image is computationally the most difficult? Is it multiple light sources or shadows or some other aspect?

Mr. McLaren: One of the major problems associated with rendering a landscape scene is the complexity and volume of data necessary to model all the scene components. During the talk I mentioned an example of explicitly modelling a field of grass using 400 billion polygons. This volume of data would challenge even the most efficient of rendering algorithms. If we consider the example of the ray tracing technique, where the ray is being traced from the eye through a pixel into object space, there is a high overhead in tracing the reflected and refracted rays through the scene's object. The optimisation of this ray/object intersection problem is currently the subject of intense research. To date, data structure solutions using hierarchies of lists and three dimensional grids have substantially increased the performance of ray tracing.

Mr. Chandler: Are there any specific areas which are causing major problems?

Mr. Kennie: I think that many of the fundamental problems associated with unusual light sources and shadowing have been addressed although of the two the former is the more problematical in terms of speed. The most serious difficulties with regard to speed arise when detailed explicit textures of natural phenomena need to be modelled or alternatively where animated video sequences are being

created. Both are computationally expensive tasks. Also when very large datasets are being used, these can also be a problem in visualisation particularly when it is necessary to be able to produce very detailed and also more generalised visualisations. I'm thinking of cases where the level of detail included in the image changes depending on the position of the observer in relation to the object, for example situations where a building is displayed either in great detail or as a block outline.

Mr. Chandler: Was the video in real time?

Mr. Kennie: If you mean were the images displayed in the flight simulation sequence produced at video rates of display, then the answer is yes. However, one has to be aware of the enormous resources required in order to achieve this level of performance.

Mr. Chandler: Obviously more impressive equipment would be required, say capable of 20 MIPS.

Mr. Kennie: It is difficult to be precise about the number of MIPS required but something of the order of 40 to 60 is typically specified in many flight simulators to solve the hidden surface removal and display window problems. The power of the video/display processor to carry out the full rendering process including texturing, atmospheric attentuation modelling and so on has been estimated at the order of 10000 MIPS by some authors. Even with these resources it is surprising how low a level of terrain and landscape detail can be included in the scene. In fact many of the earlier flight simulators provided a more detailed representation of the terrain because they were based on a video camera passing over a physical model of the terrain. Of course in such cases it was only possible to cover a relatively small area.

Mr. McLaren: Coming back to the point about the degree of detail shown within a scene, I think that in any visualisation application the scene can be split into different components based on their role and hence their priority within the application. Since realism is currently expensive to create, only those scene components with the highest priority are modelled and rendered with high levels of realism. Within the flight simulation environment it is important to model the textures of the runway very precisely compared to the trees that were shown in the video flying by as green cotton wool. They were recognisable as trees, but didn't have to be modelled down to the level of branches and leaves. Similarly in the Daimler Benz road simulator example, the road surface is very important and that is modelled precisely. So realism is currently being addressed by analysing the application and deciding which components of the scene merit great realism and creating a very crude, but effective, approximation for the rest.

Mr. Coote: Can you tell me how quickly you envisage that parallel processor driven displays will reach the microcomputer market. I believe they are already in use in flight simulators but when will a £10000 computer have them?

Mr. McLaren: Certainly some of the techniques like ray tracing are very applicable to parallel processing because each of the rays can be processed individually. So high levels of realism at a low cost unit will be achieved in the near future.

Mr. Coote: When am I likely to be able to purchase (say) a Sun 3/60 equivalent with an array of Inmos transputer chips controlling the display?

Mr. McLaren: I would guess in about three or four years time, as soon as that. If we can extrapolate from what has happened over the past two years, then I'm confident about my prediction. The flight simulator achieves its high performance using specialised and expensive hardware. However, it's now becoming easier and cheaper to customise one's own VLSI for relatively small production runs. This is already being reflected in some current graphic hardware products; Silicon Graphics now have a proprietary Geometry Engine which is VLSI based and Hewlett-Packard have recently incorporated the radiosity approach to rendering in VLSI. In the near future there will be more and more customised VLSI tackling each of the components of the rendering system. Then you'll start having the performance required for low cost realism: a Cray on your desk.

Mr. Kay: One of the problems that's beginning to be mentioned in remote sensing literature is how to model anisotropic reflectance. If that can be achieved,

all the problems involved in texturing and rendering surfaces can be solved. How close are you to doing that? Is ray tracing the only answer or do you have other solutions that might work?

Mr. McLaren: I believe there are other solutions. There's one area called radiosity. This is based on modelling heat dissipation and it's being used a lot to model, or more exactly model, the ambient light sources. Until now these have just been assumed to be a constant illumination, which they are not. The sky light has variable intensities. So this is one key area that over the past three years has received a lot of attention in the approach to shading.

Can I ask the audience if anybody here uses computer graphics at all in their business? Have they penetrated this marketplace to any extent?

Chairman: You showed the architectural use in planning and engineering. Is that percolating through into practice?

Mr. McLaren: I believe that architectural companies are realising that computer visualisations are a very powerful tool when it comes to winning competitions, winning public enquiries or even avoiding public enquiries. If you can present the general public with these types of images, giving them an impression of what a structure will look like when it's built, then in many cases you can actually avoid a public enquiry. Many of the power companies and planning authorities are using these tools before the projects are perceived as environmental problems to save money because public enquiries are very expensive. So, yes, the tools that we have presented tonight are being used in practice. Mr. Kennie and I, as part of our research for this paper and talk, visited companies to see where the action was taking place and we found considerable activity. Many system vendors, such as Intergraph who provide architectural design facilities, are extending their systems to include or interface to visualisation packages. The industry certainly perceives visualisation as a high growth area.

Mr. Kennie: There is an enormous amount of activity in this field, particularly in the architectural world. Apart from a few notable exceptions, the use of such techniques in civil engineering is much less developed. The photomontage technique has been used to a limited extent so far, but the fact that many design packages are now offering visualisation options or interfaces to external visualisation packages may change this in the future.

Dr. Dowman: Could you tell us what the role of the photogrammetrist is going to be in all this? Are we going to be measuring buildings in three dimensions from aerial photographs for incorporation into this type of model? Is it possible that we are going to be using unconventional geometry, oblique photographs for example, so that they can be used in this type of model?

Mr. Kennie: I certainly think that aerial photogrammetry has a role to play in the provision of height information, both of the terrain and of buildings in urban scenes. Close range photogrammetry is also very important, particularly for measuring interiors of buildings and for large scale modelling in general. Oblique photography has traditionally been used for producing photomontages; this could be extended by digitising the photograph and providing a facility for interactively superimposing design and other surface feature data onto the "model". Photogrammetrists would seem to be the most suitable group of professionals to develop such a facility.

Mr. McLaren: One way we could solve the problem of making a photomontage is to take it one stage further and introduce dynamics through the use of a video. A "walk through" of the site prior to a proposed project being built in the landscape could be achieved by superimposing the computer graphics image of the new design objects on the video. This is certainly one potential area for photogrammetric research: modelling the geometry of the video camera, and tracking the position and orientation of the camera to ensure the precise superimposition of the computer graphics, not just in static mode, but in the dynamic environment of the video. One of the cities in France, which applied for the Olympic Games, created a video of their proposed stadium so that you "flew" by imaginary helicopter over the stadium and then flew into and around the stadium. It was impressive. This would be another major step forward in the photomontage technique.

Mr. Phillips: The picture of St. Paul's Cathedral gives some idea of what surveyors and photogrammetrists are providing at the moment. This image should probably be credited to Arup Associates.

Mr. McLaren: That's correct.

Mr. Phillips: Although we haven't produced three dimensional data for that particular model, Plowman Craven and Associates have produced them for the same client from survey data and photogrammetric data. Surveyors are being asked to produce three dimensional data for architects.

Mr. Kennie: But are you being asked to take the process a stage further and provide visualisations from the basic data?

Mr. Phillips: No, the architects have the tools and they're using them themselves.

Chairman: But you're supplying the measured information to them?

Mr. Phillips: Yes.

Mr. McLaren: In urban environments we're being forced to perform original surveys since the 1:1250 scale map series or its digital equivalent from the Ordnance Survey doesn't provide sufficient information to perform these types of analyses. There is insufficient terrain elevation information and the heights of buildings are missing, as well as their three dimensional shape.

Chairman: If you ask the Ordnance Survey, we'll produce the data you require!

Mr. McLaren: For the time being, however, we'll be forced to go out to people like Mr. Phillips to collect that information.

Chairman: If you ask for that information, we'll provide it.

Mr. McLaren: It must be a low priority since the provision of two dimensional information from the OS is far from complete in the UK.

Mr. Laing: To what exent do you find DTMs at large scale available or not available for the type of work that you're doing?

Mr. Kennie: In the United Kingdom I understand that a national DTM has been produced by the Mapping and Charting Establishment for military projects. There has been some discussion about releasing these data into the public domain but to my knowledge they are not yet available. I am sure that the lack of national or more specifically regional DTM data is a major limiting factor to the widespread application of the techniques we have been discussing this evening.

Mr. Laing: Perhaps I can make a small contribution in that case, which is to say that MacDonald Dettwiler has been doing work recently in the extraction of DTMs from raw satellite data. We do this automatically using computer based artificial vision techniques within the Meridian software. However, we also make imaging spectrometers and we're really looking for some encouragement to equip the aircraft carrying these in order to produce large scale stereoscopic electronic images in digital form which can be run through our DTM extractors. The techniques for providing these large scale DTMs are available. It's just a question of funding. We see there's a need but being business people we have to wait until there's actually a demand.

Mr. Kennie: I'm really not sure where the answer lies. The technology and techniques for providing the DTM data exist, using both conventional digitising techniques and stereomatching of digital images. Certainly one problem is that of the education gap. Do we as surveyors and photogrammetrists need to educate our clients so that they are more aware of what is achievable? I think we probably do. I assume also that the process depends to a large extent on a need from the user community for specialised products such as computer visualisations. The EEC Directive on environmental impact analysis may act as a catalyst to generate this demand.

Mr. McLaren: I've been involved with regional visual impact analyses associated with transmission line routing and there we've been forced to go back to the original Ordnance Survey contours to provide the terrain information. This, as Mr. Kennie has pointed out, is potentially an inaccurate source of elevation data, but for this broad type of planning application the clients have accepted the inaccuracies. However, if the application is, say, visual analysis of road design in undulating terrain, then we can't really tolerate such inaccuracies and we're forced to go to original surveys.

Chairman: What scale were you taking the contours from?

Mr. McLaren: For the regional level analysis a scale of 1:50 000 was used and for larger scale projects, 1:10 000.

Mr. Turner: What sort of inaccuracy are you talking about in using the Ordnance Survey contours?

Mr. McLaren: We're talking about the accepted contour accuracy specifications which state that 95 per cent of the contours must lie within plus or minus half the contour interval. For many purposes that's insufficient because having digitised the contours, we then interpolate a terrain model. The resulting DTM is potentially highly inaccurate for the application for which it's being adopted. That's why we apply these data to planning, but not design purposes.

Mr. Davies: Could the Ordnance Survey tell us whether they do have any long term plan for providing the general public with a DTM of Britain?

Mr. Newby: Contours extracted from the 1:50 000 scale map information and a DTM with 50 m spacing derived from those contours are on sale. The first area covered was Wales and then south-west England and the south-east were also covered. Progress is being made northwards across Great Britain at the moment. Data for Wales and England south of the Wash are already available. The whole of Great Britain will be available in 1992.

Mr. Kennie: Could you give any indication of the accuracy of the DTM?

Mr. Newby: We don't advertise an accuracy, but in a small area of digital 1:50 000 scale contours that I've compared with the underlying photogrammetric 5 m contours (from which, of course, the 1:50 000 scale contours were originally derived), the digital contours which we're selling appear to be extremely faithful to the original data, with very little degradation from the original accuracy. I was quite surprised and we haven't yet repeated this over a large area but it certainly does not bear out the suggestion of a great degradation of accuracy which emerged from a very small student project recently. What I found was rather better than a 2 m standard error.

Mr. McLaren: It certainly requires considerably more research to find out the quality associated with the OS DTM product.

Mr. Newby: Of course the underlying contours have a standard error of the order of 1 m and they've been cartographically generalised to go from 1:10 000 scale to 1:50 000 scale. Apart from this generalisation it is basically the same contours that are being used.

Chairman: We still have the original film information or the diapositives which created the contours in the first place and if a better terrain model is required from that photogrammetric data, we'll give a quote for doing it. If there are customers who want better quality terrain information, then I am sure that there are organisations, including the OS, who will supply it. Similarly whether a three dimensional building or manmade features that sit on top of that terrain model are needed, as photogrammetrists we're capable of supplying that information if there are customers that want to pay for it.

Mr. Newby: Meanwhile it is the contours and derived DTM from the 1:50 000 scale map that are on sale. Data are available for 20 × 20 km square cells. Details of licensing fees may be obtained from Ordnance Survey Digital Marketing (tel: 0703 792773).

Mr. McLaren: I can certainly suggest that if this information had been available, say two or three years ago, while working on some of the projects presented here tonight, we would have come to you and used your data. There is certainly a market for your DTM product and it's growing.

Chairman: I think we're all agreed that we have had an interesting look into the future concerning the potential market for the digital mapper. Up to now we have shied away from depicting the third dimension in any great detail, but perhaps there is a market for it.

I should like to thank our two speakers for their excellent presentation.

APPENDIX D

STATEMENTS SUMMARISING CONTRIBUTION OF AUTHOR TO JOINTLY AUTHORED PAPERS

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

DEPARTMENT OF GEOLOGY

STATEMENT OF CONTRIBUTION TO PUBLISHED PAPERS

CO-AUTHORED WITH TJM KENNIE

Title of Paper : REMOTE SENSING IN CIVIL ENGINNERING PRACTICE

Authors : TJM KENNIE, MC MATTHEWS and PJ BEAVEN

Publication Details : Chapter 5 in REMOTE SENSING IN CIVIL ENGINNERING (Ed. Kennie and Matthews, 1985)

I certify that to the best of my knowledge the details cited above are correct, and that the contribution of each author to this publication was as follows;

TJM KENNIE : 35%

Prime Contribution : Role of photogrammetry and use of aerial photography in civil engineering General coordination, introduction and conclusions.

MC MATTHEWS : 35%

Prime Contribution : The use of remote sensing for projects in the developed world

PJ BEAVEN : 30%

Prime Contribution : The use of remote sensing for projects in the developing world

Address : Transport and Road Research Laboratory
Crowthorne RG11 6AU

Signature : P J B Beaven

Date : 9 November 1989

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

DEPARTMENT OF GEOLOGY

STATEMENT OF CONTRIBUTION TO PUBLISHED PAPERS

CO-AUTHORED WITH TJM KENNIE

Title of Paper : REMOTE SENSING IN CIVIL ENGINNERING PRACTICE

Authors : TJM KENNIE, MC MATTHEWS and PJ BEAVEN

Publication Details : Chapter 5 in REMOTE SENSING IN CIVIL ENGINNERING (Ed. Kennie and Matthews, 1985)

I certify that to the best of my knowledge the details cited above are correct, and that the contribution of each author to this publication was as follows:

TJM KENNIE : 35%

Prime Contribution : Role of photogrammetry and use of aerial photography in civil engineering General coordination, introduction and conclusions.

MC MATTHEWS : 35%

Prime Contribution : The use of remote sensing for projects in the developed world

PJ BEAVEN : 30%

Prime Contribution : The use of remote sensing for projects in the developing world

Address : Department of Civil Engineering,
University of Surrey, Guildford, GU2 5XH

Signature : M.C. Matthews

Date : 6/11/89

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

DEPARTMENT OF GEOLOGY

STATEMENT OF CONTRIBUTION TO PUBLISHED PAPERS

CO-AUTHORED WITH TJM KENNIE

Title of Paper : ANALYSIS OF MULTITEMPORAL AIRBORNE REMOTE SENSING FOR THE LOCATION OF SOLUTION FEATURES IN THE CHALK OF SOUTHERN ENGLAND

Authors : TJM KENNIE MS ROSENBAUM and CN EDMONDS

Publication : International Archives of Photogrammetry and Remote Sensing, Vol. 27, Part B7, Commission VII, 323-332. (1988)

I certify that to the best of my knowledge the details cited above are correct, and that the contribution of each author to this publication was as follows;

TJM KENNIE : 40%

Prime Contribution : Design of data acquisition phase of project. Analysis of results

MS ROSENBAUM : 30%

Prime Contribution : Project supervision

CN EDMONDS : 30%

Prime Contribution : Forms of solution features, geotechnical problems associated with solution features.

Address : Geology Department, Prince Consort Road, Imperial College London SW7 2BP

Signature : M. Rosenbaum

Date : 6 Nov 89

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

DEPARTMENT OF GEOLOGY

STATEMENT OF CONTRIBUTION TO PUBLISHED PAPERS

CO-AUTHORED WITH TJM KENNIE

Title of Paper : ANALYSIS OF MULTITEMPORAL AIRBORNE REMOTE SENSING FOR THE LOCATION OF SOLUTION FEATURES IN THE CHALK OF SOUTHERN ENGLAND

Authors : TJM KENNIE MS ROSENBAUM and CN EDMONDS

Publication : International Archives of Photogrammetry and Remote Sensing, Vol. 27, Part B7, Commission VII, 323-332. (1988)

I certify that to the best of my knowledge the details cited above are correct, and that the contribution of each author to this publication was as follows;

TJM KENNIE : 40%

Prime Contribution : Design of data acquisition phase of project. Analysis of results

MS ROSENBAUM : 30%

Prime Contribution : Project supervision

CN EDMONDS : 30%

Prime Contribution : Forms of solution features, geotechnical problems associated with solution features.

Address : APPLIED GEOLOGY (CENTRAL) LTD., SOUTH LODGE, CRANFORD, KENILWORTH ROAD, BLACKDOWN, LEAMINGTON SPA, WARWICKS., CV32 6RG.

Signature : C.N. Edmonds

Date : 14/11/89

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

DEPARTMENT OF GEOLOGY

STATEMENT OF CONTRIBUTION TO PUBLISHED PAPERS

CO-AUTHORED WITH TJM KENNIE

Title of Paper : LOCATION OF POTENTIAL GROUND SUBSIDENCE AND COLLAPSE FEATURES IN SOLUBLE CARBONATE ROCK BY REMOTE SENSING TECHNIQUES

Authors : TJM KENNIE and CN EDMONDS

Publication : Geotechnical Applications of Remote Sensing and Remote Data Transmission , ASTM Special Technical Publication 967, Johnson A.I. and Pettersson C.B. (Eds.), 99-121 1986, (Published 1988).

I certify that to the best of my knowledge the details cited above are correct, and that the contribution of each author to this publication was as follows;

TJM KENNIE : 50%

Prime Contribution : Review of remote sensing research associated with the detection of solution features

CN EDMONDS : 50%

Prime Contribution : Forms of solution features, geotechnical problems associated with solution features.

Address : APPLIED GEOLOGY (CENTRAL) LTD., SOUTH LODGE, CRANFORD, KENILWORTH ROAD, BLACKDOWN, LEAMINGTON SPA, WARWICKS., CV32 6RG.

Signature : C.N. Edmonds

Date : 14/11/89

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

DEPARTMENT OF GEOLOGY

STATEMENT OF CONTRIBUTION TO PUBLISHED PAPERS

CO-AUTHORED WITH TJM KENNIE

Title of Paper : A PRELIMINARY ASSESSMENT OF AN AIRBORNE THERMAL VIDEO FRAMESCANNING SYSTEM FOR ENVIRIONMENTAL ENGINEERING APPLICATIONS

Authors : TJM KENNIE, CD DALE and GC STOVE

Publication : International Archives of Photogrammetry and Remote Sensing, Vol. 26, Part 7/1, Balkema Press, 285-290.

I certify that to the best of my knowledge the details cited above are correct, and that the contribution of each author to this publication was as follows;

TJM KENNIE : 50%

Prime Contribution : Original concept, project planning, General methodology. Prime author.

CD DALE : 25%

Prime Contribution : Preliminary processing and applications.

GC STOVE : 25%

Prime Contribution : Discussion of Applications

Address : 12 HILLBROOKE CRESCENT

MAIDENHEAD BERKS

Signature : C. DALE

Date : 5/11/89

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

DEPARTMENT OF GEOLOGY

STATEMENT OF CONTRIBUTION TO PUBLISHED PAPERS

CO-AUTHORED WITH TJM KENNIE

Title of Paper : AIRBORNE THERMAL MAPPING FOR WINTER HIGHWAY MAINTENANCE USING THE BARR AND STROUD IR18 THERMAL VIDEO FRAME SCANNER

Authors : GC STOVE, TJM KENNIE and A HARRISON

Publication : International Journal of Remote Sensing Vol. 8, No. 7, 1077-1084 (1987)

I certify that to the best of my knowledge the details cited above are correct, and that the contribution of each author to this publication was as follows;

GC STOVE : 40%

Prime Contribution : Project planning and implementation

TJM KENNIE 40%

Prime Contribution : Analysis of data discussion of results

A HAARISON : 20%

Prime Contribution : Interpretation of results for highway engineering

Address : The MVA Consultancy
27 York Place Edinburgh

Signature : Alexander Harrison

Date : 9/11/89

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

DEPARTMENT OF GEOLOGY

STATEMENT OF CONTRIBUTION TO PUBLISHED PAPERS

CO-AUTHORED WITH TJM KENNIE

Title of Paper : MONITORING OF GEOTECHNICAL PROCESSES BY CLOSE RANGE PHOTOTGRAMMETRY

Authors : TJM KENNIE and WM McKAY

Publication : Planning and Engineering Geology, McCann P and O'Hara M (Eds.), Engineering Geology Special Publication No. 4, Geological Society, 163-170 (1986)

I certify that to the best of my knowledge the details cited above are correct, and that the contribution of each author to this publication was as follows;

TJM KENNIE : 50%

Prime Contribution : Review of applications, analysis and discussion of results

WM McKAY : 50%

Prime Contribution : Data acquisition, discussion of results

Address : A.M.C. LTD., FORD BROOK BUS. CEN. MARLBOROUGH RD., PEWSEY, WILTS.

Signature : .....

Date : 5/11/87

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

DEPARTMENT OF GEOLOGY

STATEMENT OF CONTRIBUTION TO PUBLISHED PAPERS

CO-AUTHORED WITH TJM KENNIE

Title of Paper : A PRELIMINARY ASSESSMENT OF THE TELLUROMETER MA200 FOR VIBRATIONAL MEASUREMENT OF FLEXIBLE STRUCTURES

Authors : TJM KENNIE, CM HOWARD and S HORKAN

Publication : Survey Review, 30, 232, (April, 1989)

I certify that to the best of my knowledge the details cited above are correct, and that the contribution of each author to this publication was as follows;

TJM KENNIE : 40%

Prime Contribution : Original concept, review of alternative techniques, design of experiments and discussion of results

CM HOWARD 40%

Prime Contribution : Design of experiments, analysis of data and discussion of results

SP HORKAN : 20%

Prime Contribution : Data acquisition and investigation of retro-reflectors

Address : 44 Howard Road

Coulsdon Surrey

Signature : Sea Horkan.

Date : 23/11/89.

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

DEPARTMENT OF GEOLOGY

STATEMENT OF CONTRIBUTION TO PUBLISHED PAPERS

CO-AUTHORED WITH TJM KENNIE

Title of Paper : A PRELIMINARY ASSESSMENT OF THE TELLOROMETER MA200 FOR VIBRATIONAL MEASUREMENT OF FLEXIBLE STRUCTURES

Authors : TJM KENNIE, CM HOWARD and S HORKAN

Publication : Survey Review, 30, 232, (April, 1989)

I certify that to the best of my knowledge the details cited above are correct, and that the contribution of each author to this publication was as follows;

TJM KENNIE : 40%

Prime Contribution : Original concept, review of alternative techniques, design of experiments and discussion of results

CM HOWARD 40%

Prime Contribution : Design of experiments, analysis of data and discussion of results

SP HORKAN : 20%

Prime Contribution : Data acquisition and investigation of retro-reflectors

Address : Civil Eng. Dept., University of Surrey, Guildford, Surrey

Signature : CMHoward

Date : 3/11/89

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

DEPARTMENT OF GEOLOGY

STATEMENT OF CONTRIBUTION TO PUBLISHED PAPERS

CO-AUTHORED WITH TJM KENNIE

Title of Paper : TERRAIN MODELLING IN SURVEYING AND CIVIL ENGINEERING

Authors : G PETRIE and TJM KENNIE

Publication : Computer Aided Design, Vol. 19, No.4, May (1987)

I certify that to the best of my knowledge the details cited above are correct, and that the contribution of each author to this publication was as follows;

G PETRIE : 50%

Prime Contribution : Theoretical background

TJM KENNIE : 50%

Prime Contribution : Characteristics of terrain modelling packages

Address : DEPARTMENT OF GEOGRAPHY & TOPOGRAPHIC SCIENCE
UNIVERSITY OF GLASGOW,
GLASGOW, G12 8QQ.

Signature : Gordon Petrie. (Professor of Topographic Science)

Date : 8th November 1989

IMPERIAL COLLEGE OF SCIENCE, TECHNOLOGY AND MEDICINE

DEPARTMENT OF GEOLOGY

STATEMENT OF CONTRIBUTION TO PUBLISHED PAPERS

CO-AUTHORED WITH TJM KENNIE

Title of Paper : MODELLING FOR DIGITAL TERRAIN AND LANDSCAPE VISUALISATION

Authors : TJM KENNIE and RA McLAREN

Publication : Photogrammetric Record, 12(72), 711-745 (1988)

I certify that to the best of my knowledge the details cited above are correct, and that the contribution of each author to this publication was as follows;

TJM KENNIE : 50%

Prime Contribution : Data acquisition procedures, applications and elements of rendering

RA McLAREN : 50%

Prime Contribution : Rendering procedures and applicationss

Address : 33 LOCKHARTON AVENUE

EDINBURGH EH14 1AY SCOTLAND

Signature : Robert A. McLaren

Date : 14th November 1989

APPENDIX E

BIBLIOGRAPHY OF WORK PUBLISHED DURING PERIOD OF REGISTRATION FOR P.H.D.

PUBLICATIONS

KENNIE T J M 1984. The use of acrylic retro-reflectors to monitor the deformation of a bridge abutment. Civil Engineering Surveyor, Vol.9, 6, 10-15.

KENNIE T J M 1984. Deformation monitoring of engineering structures: a review of current practice and an examination of two case studies. Test of Professional Competence Report, Royal Institution of Chartered Surveyors, 139pp.

LINDSELL P, KENNIE T J M, MOORE S R, RAZAK H A and ROBINSON S, 1984. Monitoring of spill through abutments. Final Report to the Transport and Road Research Laboratory, 210pp.

KENNIE T J M and MATTHEWS M C (Eds) 1985. Remote Sensing in Civil Engineering. Surrey University Press and Halstead Press, London and New York, 357pp.

KENNIE T J M 1985. An introduction to remote sensing. Chapter 2 in Remote Sensing in Civil Engineering, Kennie T J M and Matthews M C (Eds.), Surrey University Press, 1-13

KENNIE T J M 1985. Remote sensing photographic systems. Chapter 2 in Remote Sensing in Civil Engineering, Kennie T J M and Matthews M C (Eds) Surrey University Press, 14-40.

KENNIE T J M 1985. Remote sensing scanning systems. Chapter 3 in Remote Sensing in Civil Engineering, Kennie T J M and Matthews M C (Eds.), Surrey University Press, 41-85.

KENNIE T J M, MATTHEWS M C and BEAVEN P J 1985. Remote sensing in civil engineering practice. Chapter 5 in Remote Sensing in Civil Engineering, Kennie T J M and Matthews M C (Eds.), Surrey University Press, 106-117.

KENNIE T J M 1985. Leadership training for the land surveyor. Land and Minerals Surveying, VOL.3, 3, 143-150.

EDMONDS C N, KENNIE T J M and ROSENBAUM M S 1986. The application of remote sensing to the detection of solution features in limestone. Planning and Engineering Geology, McCann P and O'Hara M (Eds), Engineering Geology Special Publication No 4, Geological Society, 125-131.

KENNIE T J M and McKAY M 1986. Monitoring of geotechnical processes by close range photogrammetry. Planning and Engineering Geology, McCann P and O'Hara M (Eds.), Engineering Geology Special Publication, No 4, Geological Society, 163-170.

KENNIE T J M, EDMONDS C N and ROSENBAUM M S 1986. The success of airborne MSS and thermal IR imagery for the detection of solution features in chalk. BGRG/IHA/RSS Joint Conference on Remote Sensing Applications in Geomorphology and Hydrology, University of Reading, 6-9 January, 8pp.

KENNIE T J M 1986. Management training for the land surveyor: a case for leadership. Paper 203/21 Vol 10, 18th Conference of the International Federation of Surveyors (FIG), Toronto 3-11 June, 17pp.

PETRIE G and KENNIE T J M 1986. Terrain modelling in surveying and civil engineering. Invited paper to British Computer Society Displays Group Meeting on State of the Art in Stereo and Terrain Modelling, Cafe Royal, London, 5 April, 25pp.

KENNIE T J M, DALE C D and STOVE G C 1986. A preliminary assessment of an airborne thermal video frame scanning system for environmental engineering applications. ISPRS Commission VII Int. Symp. Enschede, The Netherlands. Int. Archives of Photogram. and Remote Sensing, Vol. 26 Part 7/1, Balkema Press, 285-290.

KENNIE T J M and GARDINER P F 1986. Leadership and Leadership development for land surveyors. RICS Land Surveying Division Ordinary General Meeting, London, 6 November, 10pp.

KENNIE T J M and MATTHEWS M C 1986. The role of short post experience courses in remote sensing for practising civil engineers. European Space Agency Workshop on Education and Training in Remote Sensing, Denmark, 7pp.

KENNIE T J M 1986. Some thoughts on management and leadership training for chartered surveyors. Winner of the Ryde Memorial Prize and the President's Prize in the Royal Institution of Chartered Surveyors Prize Papers Competition, 25pp.

KENNIE T J M 1987. Surveying. Chapter 19 in The Ground Engineer's Reference Book, Chapter 19, 27pp.

BROWN V and KENNIE T J M 1987. Photogrammetry. Chapter 21 in The Ground Engineers Reference Book, Bell F G (Ed), 29pp.

STOVE G C, KENNIE T J M and HARRISON L 1987. Airborne thermal mapping for highway maintenance engineers using the Barr and Stroud IR18 thermal video frame scanner. Int. Journal of Remote Sensing, Vol.6, 7, 350-357.

KENNIE T J M and EDMONDS C N 1987. The location of potential collapse features in soluble carbonate rocks by remote sensing techniques. ASTM Special Technical Publication on Geotechnical Application of Remote Sensing and Remote Data Transmission. Johnson I A and Pettersen B, 99-121.

PETRIE G and KENNIE T J M 1987. Terrain modelling in surveying and civil engineering. Computer Aided Design, Vol.4, 2, 177-189.

KENNIE T J M and GARDINER P F 1987. Management training and approaches to motivation in surveying organisations. Land and Minerals Surveying, Vol 5, 10, 510-518.

PETRIE G and KENNIE T J M 1987. An introduction to terrain modelling, terminology and applications. Proc. Short Course on Terrain Modelling in Surveying and Civil Engineering, University of Surrey, April and University of Glasgow, September, 4pp

KENNIE T J M 1987. Field methods of data collection for terrain modelling. Proc. Short Course on Terrain Modelling and Civil Engineering, University of Surrey, April and University of Glasgow, September, 10pp.

KENNIE T J M 1987. Software for terrain modelling applications. Proc. Short Course on Terrain Modelling in Surveying and Civil Engineering, University of Surrey, April and University of Glasgow, September, 18pp.

KENNIE T J M, BRUNTON S, PENFOLD A and WILLIAMS D 1988. EDM Instruments: Calibration Methods and Facilities in the U.K. Land and Minerals Surveying, Vol.6, 27-35.

KENNIE T J M and McLAREN R 1988. Modelling for digital landscape and terrain visualisation. Invited lecture to the Photogrammetric Society, London, February and to the British Computer Society, Newcastle Upon Tyne, March. The Photogrammetric Record, Vol.XII, No 72, 711-745.

KENNIE T J M and McLAREN R 1988. Computer graphics techniques for terrain and landscape visualisation. 16th Congress of the International Society of Photogrammetry and Remote Sensing, Kyoto, Japan, July. International Archives of Photogrammetry and Remote Sensing, Vol. 27, Part B3, Commission III, 527-536.

KENNIE T J M, EDMONDS C N and ROSENBAUM M R 1988. Analysis of multitemporal airborne remote sensing data for the location of solution features in the chalk of Southern England. 16th Congress of the International Society of Photogrammetry and Remote Sensing, Kyoto, Japan, July. International Archives of Photogrammetry and Remote Sensing, Vol. 27, Part B6, Commission VII, 236-245.

KENNIE T J M 1988. National standards for EDM instrument calibration. RICS Land Surveyor's Biennial Conference, University of Nottingham, September 23rd.

KENNIE T J M and GARDINER P F 1988. Leadership training for Chartered Surveyors in the United Kingdom - A review of a programme of continuing education. International Federation of Surveyors (FIG) Commission 2 Symposium on 'University Education for Surveyors', October, Madrid, Spain. 22pp.

KENNIE T J M 1989. Quality assurance and national standards for EDM instrument calibration. Invited contribution to Seminar on 'Quality Assurance for the Surveying Profession', University of Nottingham, 25 January, 11pp.

HOWARD C M, KENNIE T J M, HOWELL J and ALLAN A L 1989. Applications of the Tellurometer MA-200 in industrial metrology, vibrational analysis and for high precision field measurements. Proceeding, CONSAS 1989, Paper 12.4, 15pp.

KENNIE T J M, HOWARD C and HORKAN S 1989. A preliminary assessment of the Tellurometer MA-200 for vibration measurement of flexible structures. Survey Review, Vol.30, April, 67-81.

KENNIE T J M and McLAREN R 1989. Visualisation for planning and design. Seminar on Surveying Information Technology, Royal Institution of Chartered Surveyors, 1 March, 12pp.

KENNIE T J M 1989. Engineering Surveying. Chapter 6 in the Civil Engineer's Reference Book, Blake L S (Ed), Butterworths Publications.

KENNIE T J M 1989. Photogrammetry and Remote Sensing. The Civil Engineer's Reference Book, Blake L S (Ed), Butterworths Publications.

McLAREN R and KENNIE T J M 1989. Visualising Digital terrain models: techniques and applications. 'Three Dimensional Geographic Information Systems' Raper J (Ed), Taylor and Francis, 134-156.

KENNIE T J M and PETRIE G (Eds), 1989. Engineering Surveying Technology. Blackie Publishing Ltd. 485pp.

KENNIE T J M 1989. Electronic angle and distance measurement. Chapter 1 in Engineering Surveying Technology, Kennie T J M and Petrie G (Eds.), Blackie Publishers, 485pp.

KENNIE T J M 1989. Remote sensing for topographic and thematic mapping. Chapter 6 in Engineering Surveying Technology, Kennie T J M and Petrie G (Eds.), Blackie Publishers, 485pp.

PETRIE G and KENNIE T J M 1989. Digital Terrain Modelling. Chapter 10 in Engineering Surveying Technology. Kennie T J M and Petrie G (Eds). Blackie Publishing Ltd. 485pp.

PETRIE G and KENNIE T J M 1990. Terrain Modelling in Surveying and Civil Engineering. Whittles Publishers, Thomas Telford. 350pp.

COOTE A and KENNIE T J M 1990. New Skills for a Changing Role. 19th Congress of the International Federation of Surveyors, Commission 2, Paper 202.1. 10pp.

DODSON A H and KENNIE T J M 1990. National Standards for EDM Instrument Calibration. 19th Congress of the International Federation of Chartered Surveyors, Commission 5, Paper 501.1. 10pp.

KENNIE T J M 1991. Digital Terrain Modelling and Geographic Information Systems in Civil Engineering. Seminar on GIS in the Civil Engineering Industry, Nottingham Polytechnic, 13pp