Imperial College London
Department of Computing
The Hamlyn Centre for Robotic Surgery

Hybrid Tissue Surface Shape Measurement and Hyperspectral Imaging Using a Multispectral Structured Lighting Endoscope

Jianyu Lin

Supervisors:
Prof. Daniel S. Elson
Prof. Guang-Zhong Yang

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Computing and the Diploma of Imperial College
May 2018

Statements of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for other degrees or purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Jianyu Lin

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Jianyu Lin

Abstract

Surgical guidance and decision making could be improved with accurate measurement of intra-operative data including shape and hyper/multispectral information of the tissue surface. In this thesis, a dual-modality endoscopic system based on the structured light (SL) technique has been proposed to enable real-time tissue surface reconstruction and pixel-level dense multispectral imaging. The system is described in a chronological order of development for both the hardware and software aspects.

Starting from the previously developed ICL SL system that enabled tissue surface shape measurement, a naive 3D reconstruction pipeline is described. This includes a SL system calibration method, normalized cut and local rigid registration based pattern decoding algorithm. In order to enable a faster, denser, and more robust surface reconstruction, further hardware and algorithmic improvements have been proposed, including replacing the RGB camera with a multispectral camera and stroboscopically switching between white light (WL)/SL modes. Software improvements have allowed on-the-fly calibration and reconstruction, and improved accuracy through the integration of local normal information, the use of convolutional neural networks (CNN) based pattern decoding with a specially designed feature matching algorithm, and the combination of SL and structure-from-motion (SfM).

The system has also been extended to enable sparse hyperspectral signal measurements, by integrating a slit hyperspectral camera. A CNN based model has been developed to estimate pixel-level dense multispectral images from RGB images and the captured sparse hyperspectral signals.

The whole system has been validated using phantoms and *ex vivo* tissue, with *in vivo* demonstrations in a large animal trial and in patients, demonstrating its potential in real-world clinical applications.

Acknowledgements

I am grateful and proud to be a member of the Hamlyn Centre for Robotic Surgery, which provided me with many opportunities to be exposed to cutting edge research in surgical technologies. I also would like to thank the Institute of Global Health Innovation for supporting my PhD with the IGHI scholarship.

I would like to express my gratitude to everyone who helped and companied me during my PhD:

First of all, I would like to thank my supervisor Prof. Daniel Elson for providing my with this wonderful experience, for his patient supervision and support, and for granting me enormous freedom to explore unknown research areas. His vision towards surgical imaging truly enlightened and inspired me during my PhD.

I also offer my heartfelt thanks to my co-supervisor Prof. Guang-Zhong Yang, for his inspiring pioneering notions towards robotic surgery and great support.

During my PhD I had the honour to work with a fantastic optical scientist, Dr. Neil Clancy, who used his patience and solid knowledge to enlighten me, discussed each small detail with me, and spent a lot of time in building the optical system. I wouldnt have learnt and achieved so much without his help.

I also want to thank my colleagues and collaborators, who made my time in the past five years. Thanks to Dr. Ji Qi for his support in optical design and the exploration of different Chinese restaurants; Dr. Mohan Singh for his great sense of humour; Dr. Maria Tziraki for her delicious cakes; Dr. Stamatia Giannarou and Dr. Su-lin Lee for their support in my research project; Yang Hu for CAD model design. I am also grateful to my collaborators Dr. Danail Stoyanov, Prof. Lena Maier-Hein, Dr. Taran Tatla, and Dr. Emil Walsted — without them it would be impossible for me to successfully achieve and publish in different projects. My thanks also go to Fernando Avila, Elham Nabavi, Maria Leiloglou, Lin Zhang, Mali Shen, Xiao-Yun Zhou and all other colleagues and friends for their company.

Last but foremost, I would like to express our sincere gratitude to my parents, who are always the greatest support and source of inspiration for me, and give me infinite power and courage to take unknown adventures.

Contents

Abstract **v**

Acknowledgements **vii**

1 Introduction **1**

1.1 Motivation and Objectives 1

1.2 Thesis Structures and Contributions 3

1.3 Publications 6

1.4 Technical and Clinical Acknowledgements 8

2 Literature Review **9**

2.1 Introduction 9

2.2 Surface Shape Reconstruction Techniques 9

2.2.1 3D Reconstruction with Optical Techniques 11

2.2.2 Structured Light 18

2.3 Medical Hyperspectral Imaging 24

2.3.1 Light-tissue Interaction 25

2.3.2 Introduction to HSI Systems 26

2.3.3 Medical Applications 30

2.4 Convolutional Neural Networks 33

2.4.1 From Neurons to Layers 33

2.4.2 Network Training 38

2.5 Discussion and Conclusions 40

3 The Naive 3D Reconstruction Pipeline for the ICL SL System 43

3.1 Introduction 43

3.2 The ICL SL System 44

3.3 The Naive 3D Reconstruction Pipeline 45

3.3.1 SL System Calibration 45

3.3.2 Pattern Decoding 49

3.3.3 Triangulation 55

3.4 Experimental Results 56

3.4.1 Light Pattern Decoding 57

3.4.2 Surface Reconstruction Accuracy in the Naive Reconstruction Pipeline . 59

3.5 Discussion and Conclusions 60

4 Improvements of the ICL System and the Naive Reconstruction Pipeline 63

4.1 Introduction 63

4.2 Increasing the Colour Discrimination Using a Multispectral Camera 64

4.3 On-the-fly Self-calibration and Reconstruction 65

4.4 Orientation Estimation According to Pattern Deformation 69

4.5 Experimental Results 71

4.5.1 Comparison of Multispectral and RGB Cameras in Light Pattern Decoding 71

4.5.2 Reconstruction Results Using Self-calibration 72

4.5.3 Incorporation of Local Normal Information 74

4.6 Discussion and Conclusions 75

5 Towards Faster, Denser, and More Robust Surface Reconstruction 77

5.1 Introduction 77

5.2 Improvements of the ICL SL Optical System 78

5.2.1 Interleaved Switching Between WL and SL Images 78

5.2.2 Fixation of the Probe 80

5.3 Light Pattern Decoding 80

5.3.1 Neural Network based Image Segmentation 82

5.3.2 FCN based Feature Detection 87

5.3.3 Fast Feature Matching with Epipolar Constraints 90

5.4 Quasi-dense Reconstruction with SL and SfM 91

5.4.1 Monocular SfM 92

5.4.2 Fusing SL and SfM Data 93

5.5 Experimental Results 94

5.5.1 Pattern Decoding Validation 94

5.5.2 A Clinical Case . . . 97

5.6 Discussion and Conclusions . . . 98

5.6.1 Hardware . . . 98

5.6.2 Algorithm . . . 99

6 Hybrid Structured Light and Hyperspectral Imaging System 101

6.1 Introduction . . . 101

6.2 From SL to SLHSI . . . 102

6.2.1 System Configuration . . . 103

6.2.2 Hyperspectral Signal Calibration and Processing . . . 104

6.3 Pixel-level Dense Hypercube Estimation . . . 107

6.3.1 State-of-The-Art on Super-Resolution Methods . . . 108

6.3.2 Super-Spectral-Resolution . . . 115

6.4 Experimental Results . . . 118

6.4.1 HSI Acquisition . . . 119

6.4.2 Dense Hypercube Estimation . . . 120

6.5 Discussion and Conclusions . . . 123

7 Discussion and Conclusions 127

7.1 Summary of Achievements . . . 127

7.2 Future Research Directions . . . 130

7.2.1 Improvements on the ICL SLHSL System . . . 130

7.2.2 Further System Design and Development 132

7.3 Conclusion 132

Bibliography **135**

List of Tables

2.1 Comparison of HSI acquisition modes: whiskbroom, pushbroom, staring, and snapshot. Table reproduced from [162]. 28

3.1 Heart phantom surface reconstruction accuracy using manual and automatic pattern decoding, respectively. The mean, average, and maximum value of mean, standard deviation, and maximum errors of all the reconstructed surfaces are listed. Table reproduced from [165]. 61

4.1 Reconstruction errors using the self-calibration technique on benchmark data. Table reproduced from [168]. 73

5.1 Validation of feature matching. Manual annotation is used as the ground truth. 95

List of Figures

2.1 (a) The epipolar plane (grey) with two image planes. (b) The correspondence is searched along the epipolar line on the right image plane. (c, d) The image captured from the two cameras, with red points as features and white lines as epipolar lines. Figures reproduced from [109]. 12

2.2 (a) schematic of a binocular endoscope pointing at tissue surface. The point X has projections x and x' on the two image planes. (b) The binocular endoscope from the Da Vinci® surgical system. (c)(d) The surgical scenes (upper row) and their corresponding disparity maps (lower row). Figure (a) is from [182], (c)(d) are from [264]. 15

2.3 Schematics for different optical techniques: (a). monocular SfM for deformable tissue surface shape reconstruction. (b). Time-of-flight (ToF). Comparision between stereoscopy (c) and structured light (d). Figures reproduced and adapted from [182]. 19

2.4 (a) Binary coding. (b) Phase shifting. (c) De Bruijn sequence. (d) M-array. Figure source: [241, 215]. 21

2.5 A hypercube (left) and an RGB image (right). A hypercube contains hundreds of bands along the spectral dimension, normally ranging from UV to NIR; An RGB image contains three bands (red, green, and blue) in VIS. 25

2.6 Examples on light-tissue interaction. With the incident light shed on the tissue, reflection, scattering and absorption events may happen multiple times and produce sepcular reflection, diffuse reflection, and fluorescence emission out from the tissue surface. 26

2.7 (a) A typical fully connected artificial neural networks. (b) A linear computation in a neuron followed by an activation function. 34

2.8 (a) A Conv layer with 10 channels. Input: $512 \times 512 \times 3$ RGB image; output: $512 \times 512 \times 10$ volume. The visually protruding cubic is used to show the receptive field of one neuron. (b) A convolution example with kernel size 3, stride 2, and zero-padding. 36

2.9 The schematic of AlexNet used for image classification. Figure source: [151] . . 37

3.1 (a) Schematic of the ICL SL system (from Lin et al., copyright permission to be resolved). (b) The miniaturised probe tip. (c) The probe inserted through the biopsy channel of a flexible endoscope. (d) *In vivo* pig bowel surface under SL. Figure source: [165] 45

3.2 (a). A schematic of our probe-based SL system showing a similar triangulation geometry (b) "Full" checkerboard used for camera calibration (c) "Hollow" checkerboard used for projector calibration under white light. (d) Light pattern projected on the "hollow" checkerboard in SL mode. (e) Probe ray direction calibration result from method 1: locations of spots for different plane positions (red) and rays estimated by line fitting (blue). (f) Probe ray direction calibration result from method 2: rays (blue line), projector image plane (green), coordinates of features (red dots) on the image plane, the probe centre (red circle). (g) Light pattern projected on a white plane in SL mode. (h) Synthetic reference SL image that simulates the probe image plane image, with manually defined intrinsic parameter and without any distortion. Figure source: [165, 168] 46

3.3 (a) Original SL image; (b) the foreground (light pattern) of the SL image; (c) spots delineated by boundaries (red). Two partially overlapping spots with similar colours can be separated (see the red arrow). 51

3.4 (a) Captured SL image; (b) captured SL image with initially identified spots; (c) the identification procedure is propagated to more spots by local rigid registration based method. 54

3.5 (a) The SL image (left) and the reference image (right). (b) Neighbour detection for the SL image (left) and reference image (right): segmented spot boundaries (white) and connections between neighbours (red). (c) Labelled SL image (left) according to the labels assigned manually in the reference image (right). (d) Correspondences of spots between the SL image (left) and the reference image (right). 55

3.6 Light pattern decoding results using an RGB camera, on *ex vivo* porcine kidney, heart, liver, and liver with blood, silicone heart phantom, and *in vivo* porcine large bowel. The average value (blue bar) and standard deviation (black error segment) are provided. (a). Spot detection sensitivity. (b). Spot detection precision. (c). Spot identification sensitivity. (d). Spot identification precision. . 58

3.7 ICP based rigid registration to evaluate the reconstruction error. First row: silicone heart phantom under white light; second row: the reconstructed surface; third row: registration of the reconstructed surface (green) with the ground truth mesh (red). Figure reproduced from [165] 60

4.1 (a) RGB camera; (b) normalised filter transmission spectra for the RGB camera and the overlapping area (yellow shadow); (c) Multispectral camera; (d) Filter transmission spectra for the multispectral camera and the overlapping area (yellow shadow). 65

4.2 This figure shows the ambiguity of colour discrimination in different wavelengths, for both the RGB and multispectral camera. (a) Multispectral camera: estimated wavelengths (red); estimation error bar (black); the camera filter transmission spectra (blue). (b) RGB camera: estimated wavelengths (red), estimation error bar (black), and the camera filter transmission spectra (blue). Longer black lines indicate higher ambiguity at corresponding wavelengths (left x axis). 66

4.3 The spot detection and identification results on different objects. (a). Spot detection sensitivity. (b). Spot detection precision. (c). Spot identification sensitivity. (d). Spot identification precision. In all charts, the experimental results using RGB camera are shown in blue; while those using multispectral camera are shown in red. 72

4.4 (a) Liver under white light; (b) Liver under SL; (c) Reconstructed surface using self-calibration where colour indicates the surface depth (in mm); (d) Relative error map on the reconstructed surface using self-calibration where colour indicates the reconstruction error (%). Figure reproduced from [168]. 73

4.5 (a). Rectangular box edge under white light and SL. (b). Reconstructed rectangular box edge (blue) aided by normals (red). (c). cylinder under white light and SL. (d). Reconstructed cylinder (blue) aided by normals (red). (e). Liver phantom under white light and SL. (f). Reconstructed liver phantom (blue) aided by normals (red). (g). Reconstructed step surface (blue) without normal information. (h). Reconstructed step surface (blue) aided by normals (red). . . . 75

5.1 The schematic of the upgraded ICL SL system with chopper wheel and probe adaptor. Figure adapted from [166]. 78

5.2 (a) The chopper wheel. The inter-fan area allows light to pass through. (b) Trigger timing schematic. The 5 V output signal from the signal generation device is used to trigger acquisitions from the camera and the chopper wheel. The chopper period is indicated by T, which is double the camera trigger signal. The rising and falling edges (e.g. t_1 and t_2) triggers the camera, and the trigger latency, exposure and readout times define the upper limit of the acquisition framerate. 79

5.3 (a) The adaptor consists of two parts that can be joined by screws (a). (b)(c) The CAD model showing adaptor fixing the probe (blue) to the tip of the 5 mm rigid endoscope (black), from different viewing angles. (f) A photograph of the printed prototype. 81

5.4 The fully convolutional network (FCN) model. Replacing the FC layers with Conv layers enables low-resolution heatmaps that highlight the existence of target objects; extending the network with deconvolution layers produces pixel-wise class scores. Figure source: [252] (copyright ©2017 IEEE). 84

5.5 The data mergence strategies in different models. SegNet uses the corresponding max-pooling indices followed by spatial domain convolution; FCN uses element-wise sum; U-Net uses concatenation and convolution along the channel dimension. Figure reproduced and adapted from [18, 87] (copyright ©2017 IEEE). . . 85

5.6 The exponential increase of receptive field by stacking firstly 1-dilated convolution (a), then 2-dilated convolution (b), and finally 4-dilated convolution (c). Figure source: [302]. 85

5.7 (a) The deployed FCN model; (b) Cropped SL image; (c) Density map of the cropped SL image; (d) Feature matching between the captured (left) and the reference SL image (right). Figure source: [167] 88

5.8 (a) The reference SL image, with a target spot delineated in white outline. (b) A captured SL image, with candidate matched spots close to the epipolor line, with false matching delineated by yellow, and positive matching by white. (c) The working space marked by white segments. The normalised ratios between red, green, and blue in each segment are used as elements in the feature vector. . 91

5.9 The workflow for combining SL and SFM. Figure source: [166]. 92

5.10 Examples of pattern decoding results using the proposed method. The estimated probability maps and the matching between the captured and reference SL images are shown on four example images. The upper two are for the ovine heart, and the lower two for the ovine liver. 96

5.11 (a) The ICL SL system set up on a two layer trolley. (b) The surgeon holding the endoscope with probe attached, and inserting it through a standard trocar. (c) the SL pattern projected on the patient's larynx. 97

5.12 (a) The reconstructed surface from an *in vivo* experiment on patient larynx. (b) The reconstructed surface from an *ex vivo* experiment on excised patient larynx tissue. 98

6.1 The spatial mapping between the randomised end (2D), the linear array end (1D), and the slit hyperspectral sensor. 103

6.2 Left: setup in SL mode. Supercontinuum laser is used to generate the structured light, and the white light is blocked. Right: setup in HSI mode and the captured spectrograph image (lower right). Supercontinuum laser is blocked and white light is shed on the tissue surface. The mirror mounted on a flipper is used to redirect the reflectance to the slit hyperspectral imager. 104

6.3 A linear combination of oxy and deoxyhaemoglobin absorption spectra results in the final absorption spectrum. The relative ratio between the oxy and deoxyhaemoglobin can be estimated via linear least squares. 106

6.4 The schematic of the SRCNN model. The first layer extracts features from the input LR image; the second layer maps those features into HR; Finally the feature map is merged along the channel dimension to generate the output HR image. Figure reproduced from [62] (copyright ©2016 IEEE). 111

6.5 The schematic of the ESPCN model. The first stage extracts features from the input LR image and maps the feature maps to different dimensions; the second layer reshapes the $M \times N \times r^2$ feature map into the output scale ($rM \times rN$). Figure reproduced from [254] (copyright ©2016 IEEE). 112

6.6 The schematic of the VDSR model. The Conv layers from the VGG-net is adopted, aided by a skip-connection to avoid gradient vanishing/ explosion. Figure reproduced from [144] (copyright ©2016 IEEE). 112

6.7 The schematic of the SRGAN model. The model contains a generator that predicts the HR image, and a discriminator that distinguish between the generated and the reference image. Figure source: [160] (copyright ©2017 IEEE). 113

6.8 The schematic of the CNN model that predicts a hypercube from an input RGB image. The model contains two steps: 1) spectrally upscaling the RGB image; 2) Extraction of high-frequency signals and mergence with the low frequency ones. Figure reproduced from: [166]. 116

6.9 The schematic of the CNN model that predicts a hypercube from an input RGB image and sparse hyperspectral signals. This model can be seen as Model 1 with an extra mergence phase that integrates the hyperspectral signals. Figure reproduced from: [166]. 117

6.10 (a) Macbeth colour chart illuminated with SL. (b) Reflectance spectra from selected colour panels of the chart, with R and C indicating the row and column, respectively. Data from the hybrid system (SLHSI) are plotted alongside the high resolution spectrometer results (GS). Figure source: [167]. 119

6.11 (a) Three-coloured cylindrical target with SL. (b) Mean reflectance spectra for each of the targets regions. (c) 3D reconstruction coloured by RGB values generated from spectra in (b). (d) Murine abdomen with StO2 overlay. (e) Tissue absorbance spectrum and model fit. Figure source: [167]. 120

6.12 (a) The ground truth spectra (black), estimated spectra using method 1 (blue) and method 2 (red), on five representative locations from a porcine sample image. (b) comparison (in PSNR) between the performances of model 1 (blue) and model 2 (red) along different wavelengths in LOOCV. PSNR map of the prediction results for model 1 (c) and model 2 (d) regarding the same sample (a). (b) PSNR change along number of training iterations (model 1: red; model 2: blue) in LOOCV. The average relative error on different wavelengths regarding (a) are shown in (f) and (g). Figure reproduced from: [166]. 121

6.13 The transfer learning results (indicated by PSNR) with model 1 (blue) and model 2 (red). Both models were trained and tested on 4 datasets: pig bowel, rabbit uterus, sheep uterus, and mixed rabbit/sheep uterus. 5-fold cross validation was applied to calculate the PSNR when training and testing on data from the same source. 122

6.14 Example human (a) in vivo and (d) ex vivo reconstructed tissue surface. (b, e) synthetic narrow band images and (c, f) synthetic oxygen saturation maps overlaid on these surfaces. Figure reproduced from: [166]. 123

List of Acronyms

API	Application Programming Interface
AOTF	Acousto-Optical Tunable Filter
API	Application Programming Interface
ANN	Artificial Neural Network
AR	Augmented Reality
BRIEF	Binary Robust Independent Elementary Features
CPU	Central Processing Unit
CCD	Charge-Coupled Device
CT	Computed Tomography
CGH	Computer-Generated Hologram
CRF	Conditional Random Field
CW	Continuous Wave
CONV	Convolutional
CNN	Convolutional Neural Network
DSfM	Deformable SfM
DOE	Diffractive Optical Elements
DoF	Degree-of-Freedom
DP	Dynamic Programming
ELU	Exponential Linear Unit
FAST	Features from Accelerated Segment Test
FPGA	Field Programmable Gate Arrays
FOV	Field-of-View
FPS	Frames per Second
FCN	Fully Convolutional Network
FC	Fully-Connected
GAN	Generative Adversarial Network
GRIN	Gradient Refractive Index
GPU	Graphics Processing Unit
HR	High Resolution
HF	High-Frequency
HSI	Hyperspectral Imaging
IR	Infrared
KRR	Kernel Ridge Regression
LTCF	Liquid Crystal Tunable Filter
LR	Low Resolution

MRI	Magnetic Resonance Imaging
MSER	Maximally Stable Extremal Regions
MIS	Minimally Invasive Surgery
MSI	Multispectral Imaging
NIR	Near Infrared
NCC	Normalised Cross-Correlation
OS	Operating System
OMP	Orthogonal Match Pursuit
PAN	Panchromatic
PSNR	Peak Signal-to-Noise Ratio
PGP	Prism-Grating-Prism
RANSAC	Random Sample Consensus
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
ROI	Region-of-Interest
ORB	Rotated BRIEF
SIFT	Scale Invariant Feature Transform
SfS	Shape-of-Shading
SNR	Signal-to-Noise Ratio
SNR	Signal-to-Noise Ratio
SVD	Singular Value Decomposition
SURF	Speeded Up Robust Features
SL	Structured Light
SfM	Structure-from-Motion
SR	Super-Resolution
SVM	Support Vector Machine
ToF	Time-of-Flight
US	Ultrasound
UV	Ultraviolet
VIS	Visible Spectrum
WL	White Light
WTA	Winner-Takes-All

Chapter 1

Introduction

1.1 Motivation and Objectives

"Surgery is not a place or procedure, but an intellectual discipline, characterized by operative procedures but defined by an attitude of responsibility toward the care of the sick."

This frequently used quote is by Dr. Mark M. Ravitch, an internationally recognised surgeon who is a pioneer in minimally invasive surgery (MIS) and deserves enormous credit to the development of stapling devices [69]. In the past several decades, the development of technology has boosted the emergence of new surgical approaches. MIS, for example, has limited the physiological damage of surgery by smaller incisions, reduced blood loss, less pain and fewer infections, etc. All these contribute to the fast recovery of patients, benefiting both individual patients and society. However, MIS also brings challenges to surgeons, such as reduced tactile feedback and stereo perception, limited surgical instruments, harder hand-eye coordination, higher learning curve, and so on. In this thesis, the problem of insufficient stereo information is addressed by the proposed system.

The accurate imaging of patient anatomy provides useful information for surgical decision making both pre- and intra-operatively. The quantification of tissue surface morphology is valuable in medicine and surgery as it may indicate tissue pathology. For instance, the detection

of colonic polyps often depends on their shapes [182]. Also, tissue surface shape has the potential to be used in MIS to facilitate navigation, for example, generating augmented reality (AR) [225]. However, due to the difficulty in matching pre-operative images with the laparoscopic view intra-operatively, the pre-operative images are infrequently used during the procedure itself and are consulted mainly for pre-operative planning. One of the challenges preventing their usage is the lack of depth information on the 3D tissue surface during MIS. In comparison, tomographic imaging modalities such as computed tomography (CT), and magnetic resonance imaging (MRI) have been developed which can detect diseased tissue as well as measuring tissue morphology and defining surface orientations pre-operatively. However, these cannot be easily integrated with ordinary surgical instruments intra-operatively.

Besides the tissue surface shape, multi/hyperspectral imaging (MSI/HSI) also provides important information for image-guided MIS. MSI/HSI measures the spectrum for different locations in the image of a scene, resulting in hypercubes (x-y-λ volumes) with more channels than RGB images do, containing richer spectral information to indicate the object properties. MSI/HSI has the potential to be used in various clinical applications, such as narrow-band imaging for vascular visualisation, oxygen saturation estimation for intra-operative perfusion monitoring and clinical decision making [50], tissue classification and pathology identification [174], and so on. However, in MSI/HSI system design there is always a trade-off between the spatial, spectral resolution, acquisition time, and the cost, especially for a miniaturised endoscopic system.

Due to the importance of tissue surface shape and multi/hyperspectral information, and the need to not disrupt the surgical workflow, it will be beneficial to integrate the functions of 3D object surface reconstruction and MSI/HSI into a normal endoscopic system. Developing one clinically feasible endoscopic system that is capable of 3D reconstruction and MSI/HSI is non-trivial work, since we need to not only overcome the challenges mentioned above, but also miniaturise the extra hardware to enable system integration. These problems are addressed in this thesis.

Previously, Clancy et al. developed a miniaturised endoscopic system (ICL SL system) for tissue surface reconstruction, based on SL techniques [51]. However, the previous work on this

system mainly focus on the hardware design, lacking the software platform development. So starting from this ICL SL system, this research centres around the continuous improvement and extension of the hardware, as well as the establishment of the software platform for fast and robust tissue surface imaging.

The aims of this thesis are: 1) propose a customised 3D reconstruction pipeline to enable an accurate real-time tissue surface shape measurement; 2) hardware improvement on the original system for the adaptation to clinical environments; 3) integration of HSI into the ICL SL system; 4) combine hardware and state-of-the-art algorithms to enable a near real-time dense multispectral image measurement. In general, this thesis aims to develop a miniaturised system that is capable of fast 3D reconstruction and hyperspectral imaging in clinical environments.

1.2 Thesis Structures and Contributions

Since the project centres around the building and extension of the ICL SL system, the thesis is arranged in a chronological order. The outline of the thesis is listed as follows.

Chapter 2 gives a general literature review towards 1) object surface 3D reconstruction and 2) hyperspectral imaging. It starts with a review of optical techniques for 3D reconstruction, including stereoscopy, structure-from-motion (SfM), shape-from-shading (SfS), and Time-of-Flight (ToF). This is followed by an overview of SL, arranged in terms of pattern categories. Previous works on laparoscopic SL systems are also reviewed. The second part of this chapter provides an introduction to MSI/HSI, followed by examples of previous MSI/HSI systems, and the application of MSI/HSI techniques in medical applications.

Chapter 3 starts with a brief introduction to the ICL SL optical system. Next, a naive 3D reconstruction pipeline is proposed. This pipeline includes customised SL system calibration, pattern decoding based on normalized cut and local rigid registration, and the triangulation for 3D reconstruction as the final step. Validations are provided based on phantom, *ex vivo*, and *in vivo* experiments.

Chapter 4 illustrates the improvements to the original ICL SL system and the naive 3D reconstruction pipeline. Three aspects have been covered, including replacing the RGB camera with a multispectral camera, on-the-fly self-calibration and reconstruction, and the integration of local normal information for more accurate 3D reconstruction. Validations are also provided based on phantom, *ex vivo*, and *in vivo* experiments, which show potential of applying the ICL SL system in real-world clinical environments.

Chapter 5 aims at further system improvements to enable a faster, denser, and more robust surface reconstruction. A chopper wheel has been used to stroboscopically switch between the SL and white light (WL) modes. The SL frames are used for 3D reconstruction, and the WL frames are provided to the surgeons for surgical scene visualization. Then a 3D printed adaptor is proposed to fix the probe by the side of the endoscopic tip, and enable the insertion of the probe through a normal clinically used trocar. Next, a new light pattern decoding algorithm that enables real-time 3D reconstruction are described. This is realised by convolutional neural networks (CNN) for feature detection, and a customised feature descriptor and epipolar constraints for feature matching. In addition, a review of CNN based image segmentation methods is also provided. Furthermore, we manage to combine the reconstructed surface with SL and the counterpart with SfM to increase the density of the 3D reconstruction. In the validation, phantom and *ex vivo* ovine tissues are used to evaluate the performance of the pattern decoding algorithm. Besides, a clinical experiment is provided to demonstrate the applicability of using the improved ICL system in the operating theatre.

Chapter 6 proposes a system for hybrid 3D reconstruction and HSI, by functional extension based on the ICL SL system. A slit hyperspectral camera is integrated in to the original optical system, in the way that the reflectance from the tissue surface under white light is fed into the slit for sparse hyperspectral signal measurement. This ensures the measurement of sparse hyperspectral signal from the tissue surface. A CNN based super-resolution model has also been deployed to estimate pixel-level dense hypercubes from the RGB images and the sparse hyperspectral signals. In the experiments, the accuracy HSI signal measurement was evaluated by experiments on Macbeth color checker, and phantoms. *ex vivo* tissues like pig bowel, rabbit uterus, and lamb uterus were used to validate the performance of the proposed dense hypercube

estimation algorithm.

The contributions of this work are listed as follows:

1. Detailed illustration of a customised SL system calibration method.
2. Development of patten decoding method with normalized-cut and local rigid registration.
3. Replacement of the RGB camera with a multispectral camera for potential multispectral imaging.
4. On-the-fly self-calibration that enables free-hand manipulation.
5. Integration of local normal information for more accurate 3D surface reconstruction.
6. Stroboscopic switching between the SL and WL modes.
7. Adaptor design to enable the real-world application of the system.
8. Real-time 3D reconstruction with SL aided by CNN and customised feature descriptor.
9. Combination of SL and SfM for quasi-dense surface reconstruction.
10. Integration of the function of HSI to the ICL system.
11. Development of super-resolution solution for near real-time pixel-wise dense hypercube generation.
12. Validation on Phantom, *ex vivo* animal, *in vivo* animal, and patient experiments.

To sum up, in this thesis, we demonstrate a hybrid system that enables fast and quasi-dense 3D reconstruction, and a pixel-wise dense hypercube measurement, of the tissue surface. The evaluation of the system demonstrates its potential to be used in clinical environments.

1.3 Publications

J. Lin, N. T. Clancy, J. Qi, Y. Hu, T. Tatla, D. Stoyanov, L. Maier-Hein and D. S. Elson, "Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks", *[submitted for publication]* , (2018).

J. Lin, E. S. Walsted, V. Backer, J. H. Hull and D. S. Elson, "Quantification and analysis of laryngeal closure from endoscopic videos", *[submitted for publication]* , (2018).

X-Y. Zhou, J. Lin, C. Riga, G-Z. Yang and S-L. Lee, "Real-time 3D shape instantiation from single fluoroscopy projection for fenestrated stent graft deployment", *The IEEE Robotics and Automation Letters (RA-L)* , (2018).

J. Qi, H. He, J. Lin, Y. D, D. Chen, H. Ma and D. S. Elson, "Assessment of tissue polarimetric properties using Stokes polarimetric imaging with circularly polarized illumination", *Journal of Biophotonics* , (2017).

J. Lin, N. T. Clancy, Y. Hu, J. Qi, T. Tatla, D. Stoyanov, L. Maier-Hein and D. S. Elson, "Endoscopic depth measurement and super-spectral-resolution imaging", *20th International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI)* , (2017).

J. Lin, N. T. Clancy, and D. S. Elson, "Recovering dense tissue multispectral signal from in vivo RGB images", *The Hamlyn Symposium on Medical Robotics (HSMR)* , (2017).

J. Lin, N. T. Clancy, X. Sun, J. Qi, M. Janatka, D. Stoyanov and D. S. Elson, " Probe-based rapid hybrid hyperspectral and tissue surface imaging aided by fully convolutional networks", *19th International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI)* , (2016).

J. Lin, N. T. Clancy, X. Sun, J. Qi, D. Stoyanov and D. S. Elson, " Hybrid tissue surface hyperspectral imaging and shape recovery", *Hyperspectral Imaging and Applications conference (HSI)* , (2016).

J. Lin, N. T. Clancy, D. Stoyanov and D. S. Elson, " Tissue surface reconstruction aided by local normal information using a self-calibrated endoscopic structured light system", *18th International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI)* , (2015).

J. Lin, N. T. Clancy, and D. S. Elson, " An endoscopic structured light system using multi-spectral detection", *Int J CARS* **10**, 1941-1950 (2015).

N. T. Clancy, J. Lin, S. Arya, G. B. Hanna and D. S. Elson, "Dual multispectral and 3D

structured light laparoscope", *Proceedings of SPIE* , (2015).

J. Lin, N. T. Clancy, G. Boissonnat and D. S. Elson, "Detection and identification of multispectral structured light patterns for minimally invasive surgery", *The Hamlyn Symposium on Medical Robotics* , (2013).

K. Leibrandt, P. Wasuntapichaikul, J. Lin, A. Cakir, K. Shetty, K. Sriskandarajah, M. Sodergren, D. Leff, A. Darzi and G. Z. Yang, "Classification of safe and erroneous surgery using visual gaze behaviour and cortical brain function", *2013 Annual Meeting - Organization for Human Brain Mapping (OHBM)* , (2013).

D. Harris-Birtill, M. E. Gallina, Y. Zhou, J. Lin, M. Singh, T. Cass and D. S. Elson, "Automatic image analysis of nanorod dimensions", *Bio-Nano-Photonics Symposium* , (2013).

1.4 Technical and Clinical Acknowledgements

Chapter 3 and 4: Neil Clancy provided hardware support for the hardware modification, as well as the device setting up during the experiments.

Chapter 5: Neil Clancy and Ji Qi provided hardware support for the integration of chopper wheel, as well as the experiments include the clinical trial. Yang Hu assisted in the design and 3D printing of the adaptor. The clinical trial on patients was arranged and applied with the help of Taran Tatla.

Chapter 6: Neil Clancy provided hardware support for the integration of the slit HSI camera, as well as the *in vivo* animal experimental data collection.

Chapter 2

Literature Review

2.1 Introduction

As is described in Section 1.1, this thesis proposes a system that is capable of tissue surface shape reconstruction and hyperspectral imaging. In this chapter, previous relevant works are reviewed to provide the readers with an overview of the problems to be solved in this project, from both the hardware and software aspects. For hardware, the state-of-the-art methods of optical 3D reconstruction and HSI/MSI, especially those medical related ones, are reviewed. Section 2.2 starts with the description of the applications and necessarity for intra-operative soft tissue surface reconstruction, followed by a review of various optical based reconstruction techniques, and previous SL systems in laparoscopic surgery. Section 2.3 focuses on reviewing state-of-the-art HSI/MSI systems categorised according to the acquisition mode, dispersive devices, and different medical applications. For software, since convolutional neural networks (CNNs) have been heavily deployed for algorithm development, in terms of aiding multiple key steps such as pattern decoding and super-spectral-resolution, a general overview of CNNs is given in section 2.4.

2.2 Surface Shape Reconstruction Techniques

Accurate imaging of tissue anatomy is important both pre- and intra-operatively. It benefits clinical applications.

In MIS, due to the indirect visualisation through endoscopes, surgeons lose the sense of depth context when visualising the intra-operative scene. This leads to instrument mis-location and consequently unwanted tissue interaction, and further results in higher learning curves for instrument manipulation for the surgeons [280]. In robotic surgery, the application of active constraints which constrain the maneuver of manipulator end-effectors, also depends on accurate tissue surface shape measurement.

Tissue surface morphology also carries important information and indicates some pathologies. For example, colonic polyps could be detected and discriminated from each other depending on their shapes [202].

Also, tissue surface shape has the potential to be used in MIS to facilitate navigation, for example, generating augmented reality (AR) to enhance the surgical scene visualisation, as well as to alleviate the drawback of lacking direct view and tactile feedback during MIS [24]. A typical workflow of surface-based laparoscopic AR approach has been summarised in [185]:

- Surface extraction for the organ of interest from the pre-operative data, e.g. CT, MRI, ultrasound, etc.
- Tissue surface reconstruction from the intra-operative scenes.
- Rigid registration of data that are acquired pre- and intra-operatively, via manual, semi-/fully automatic methods.
- Registration refinement with rigid or non-rigid models.

The intra-operative surface can be reconstructed either sparsely or densely, depending on adopted equipment and algorithms. The accuracy of registration between two surfaces relies on the intra-operative reconstruction accuracy: a small error in intra-operative reconstruction might lead to a big registration error, due to the continuous deformation of tissues. The surface shape could be recovered via optical techniques which observe the superficial tissue; while volume could be recovered by intra-operative tomographic imaging modalities, such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US). AR based on volume registration, although can be achieved with richer 3D correspondences, is be a non-trivial problem in practice, given challenges in the application of intra-operative tomographic imaging

modalities: 1) cost; 2) ergonomics; 3) the integration with ordinary surgical instruments; 4) the trade-off between image acquisition speed, resolution, and signal-to-noise ratio (SNR) [181].

Recently in MIS, using endoscopes with adequate algorithms or integrated external hardware for 3D soft tissue reconstruction, has become more popular. These methods are referred to as optical techniques, which utilise the geometry of the reflected light from the tissue surface for reconstruction. They are in general advantageous due to: 1) lower cost; 2) little change to the theatre environment; 3) little extra manipulation for the surgeons; 4) free from radiation. In this chapter, an overview of optical 3D reconstruction techniques is given.

2.2.1 3D Reconstruction with Optical Techniques

Object surface reconstruction, which refers to the process that captures the shapes of objects, has been a popular area in computer vision and computer graphics for decades [109]. In general the optical based 3D reconstruction techniques can be divided into two categories: passive and active methods, although the practical MIS realisation of either method is yet to be achieved reliably. The former uses images from the camera without an extra light source, and the latter acquires image data under controlled lighting conditions [192]. In the following, several typical techniques are concisely introduced and compared: passive methods including stereoscopy, Structured-from-Motion (SfM), and Shape-from-Shading (SfS), and active methods including Time-of-Flight (ToF) and Structured Light (SL), with their applications in MIS.

Stereoscopy

Stereoscopy refers to the techniques making use of parallax between observations from two different viewpoints to calculate the 3D location [182]. It is also called passive stereo in some previous works. Here a brief introduction towards the 3D reconstruction with stereoscopy is give, since other optical techniques such as SfM and SL, share much in common in geometrical concepts with stereoscopy.

A typical stereoscopy system consists of two cameras positioned in 3D space and facing the target object with parallax, sharing partially overlapping Field-of-View (FOV). An image example of FOVs from the left and right cameras is given in Fig. 2.1(c)(d). In MIS, modeling

cameras on the tip of the endoscopes with pin-hole model, "triangulation" is used to describe the procedure of calculating the intersection of light for 3D reconstruction. This procedure with stereoscopy is shown intuitively in Fig. 2.1(a)(b). The camera is normally modelled as a pin-hole model in the reconstruction. A point X in the 3D space, its projection on both the camera image planes, and the camera centre lie on the same line. For a two camera system, X with its two projections x and x' and camera centres C and C', are located on one plane named the epipolar plane π. The intersection of π and the camera image plane is called the epipolar line l' on which x' is located. The line connecting camera centres C and C' is called the baseline. The intersection between the baseline and the image plane are epipoles e and e'. By measuring the relative position between the two cameras, as well as the correspondence between projections x and x', the 3D location of X can be estimated via triangulation.

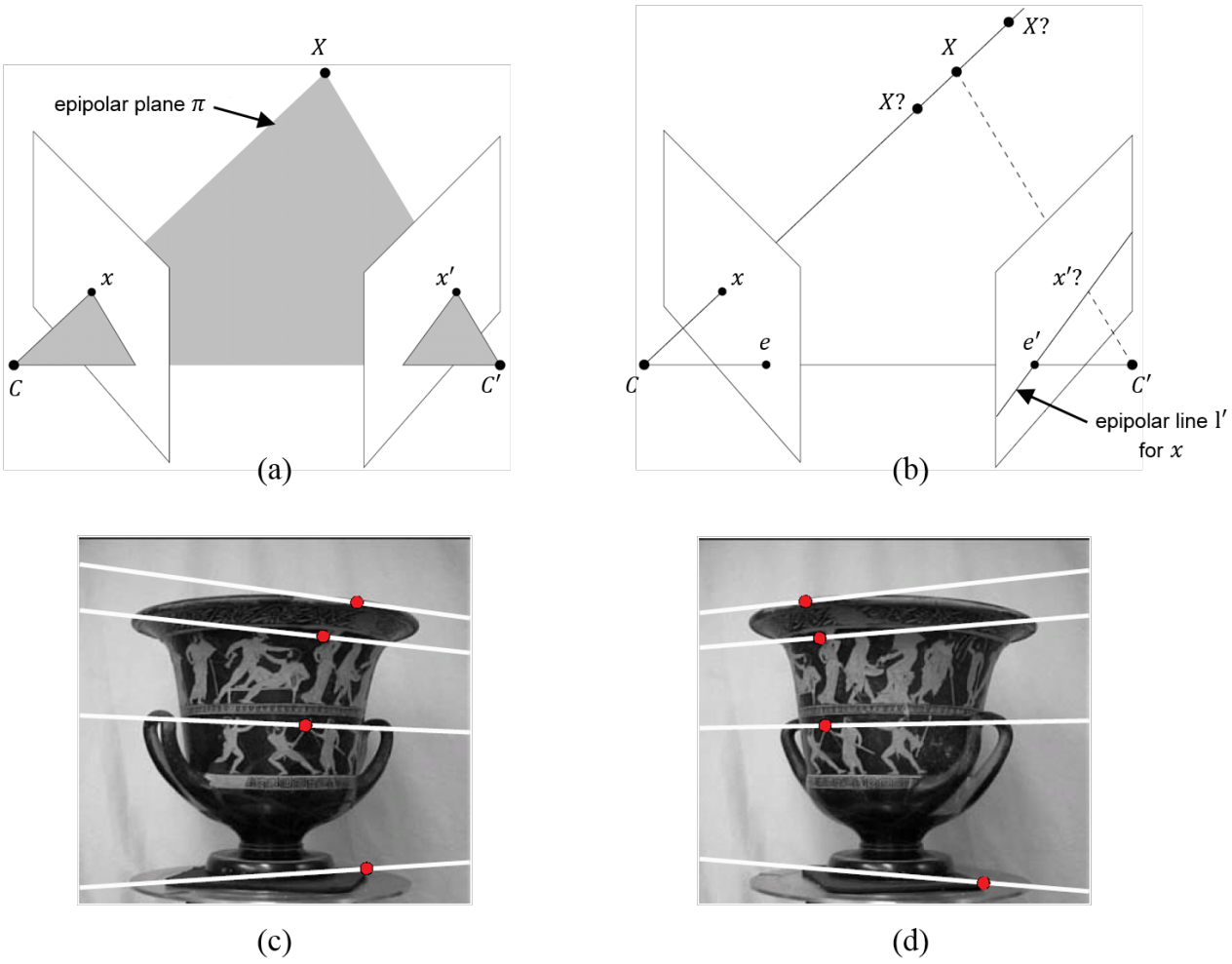


Figure 2.1: (a) The epipolar plane (grey) with two image planes. (b) The correspondence is searched along the epipolar line on the right image plane. (c, d) The image captured from the two cameras, with red points as features and white lines as epipolar lines. Figures reproduced from [109].

Here a concise mathematical description of the reconstruction procedure is provided, whose details can be found in [109]. The reconstruction starts with a calibration step which aims to find camera intrinsic parameters and relative positions. Considering cameras as pin-hole models, intrinsic parameters can be represented with 3×3 matrices K and K' for the two cameras, respectively. A 3D coordinate system is built using one of the cameras (camera 1). The coordinate system origin is set at one of the camera centres C, with X and Y axes along the row and column dimension of the camera image plane, Z axis along the camera principal axis. Through calibration the rotation R and translation t of the second camera (camera 2) can be estimated. The camera centre for camera 2 can be calculated by $C' = -Rt$. Then the camera matrices P and P' are constructed by $P = K[I|\mathbf{0}]$ and $P = K[R|t]$, for camera 1 and 2, respectively. Given a point X in the 3D space, its projections on the two camera image planes can be computed as $x = PX$, and $x' = P'X$. A fundamental matrix F is used to describe the correspondence between x and x':

$$x'^{\mathrm{T}} F x = 0 \tag{2.1}$$

where F is defined as $F = x'^{-\mathrm{T}}[t]_{\mathrm{x}} R K^{-1}$, with $[\quad]_{\mathrm{x}}$ standing for the skew symmetry operation. It is noticeable that equation 2.1 is suitable for cameras not at infinity, which is the case in normal 3D reconstruction with stereoscopy. The triangulation between x and x' can be carried out with different methods such as linear triangulation, Sampson approximation, or optimisation by minimising the reprojection error. The detailed triangulation methods are illustrated in [109].

As the triangulation follows a standardised paradigm given calibrated camera parameters, the main challenge of using stereoscopy is to find unique correspondence of image primitives on both the stereo images [264]. Once the correspondence is built, either sparsely [277] or densely [20], the 3D location can be estimated with standard triangulation algorithms.

There are many feature detection and matching algorithms proposed previously for sparse correspondence search, such as Scale Invariant Feature Transform (SIFT) [173], Speeded Up Robust Features (SURF) [22], Features from Accelerated Segment Test (FAST) [235], Maximally Stable Extremal Regions (MSER) [188], Oriented FAST and Rotated BRIEF (ORB) [238], etc. These methods manage to find sparse correspondences between paired images, while the dense correspondence is found by estimating the disparity map [270].

Rectification which projects paired images on a common image plane, is the prerequisite for

disparity map generation. Through rectification, the correspondence of each pixel can be found by searching along the same Y axis on the other image, which greatly simplifies the matching problem and reduces the computational cost. There are in general four steps for computing a disparity map [243]: 1) cost computation, 2) cost aggregation, 3) disparity computation, 4) disparity refinement. Barnard and Fischler proposed a local algorithm with slide-window style normalised cross-correlation (NCC) based similarity measurement between paired images with winner-takes-all (WTA) strategies [20]. However, the local methods aren't able to utilize regional constraint and therefore lead to large mis-matching areas in areas with relatively homogeneous appearances. Alternatively, global algorithms have been proposed to optimise the cost function of disparity with global constraints, by means of dynamic programming (DP) [56], graph cuts [149], graph cuts with belief propagation [273], variational methods [268], etc. Recently, some researchers have also exploited deep learning algorithms with CNN models to aid the disparity, scene flow, or optical flow estimation [65, 190]. There are many other works on sparse/ dense stereo matching for 3D reconstruction, so it is a highly developed and mature research area. The main challenge is adapting existing algorithms in detailed applications, which often requires further task-specific algorithm extension.

In MIS, using two individual monocular endoscopes for stereo vision is not appreciated, due to problems like difficulty in sharing overlapping FOVs, large post-rectification image distortion introduced by relative axial and large transverse movement between the two individual scopes. So commercial binocular endoscopes that are specially designed for stereo vision are used instead. An example is the binocular scope in the Da Vinci® surgical system (Fig. 2.2(b)).

Due to the free-form tissue deformation and complexity of the constantly changing surgical scenes (e.g. smoke, bleeding, specular highlights, appearance of surgical instruments, etc.), the applicability and robustness of common feature detection and tracking or stereo matching algorithms is questionable in MIS. Even though some specific methods have been proposed to adapt to this extreme situation [91], they still suffer from high computational load. Besides, monocular scopes are used in the majority of MIS procedures where stereoscopy is not able to be applied.

It has been found that the reconstruction accuracy depends on the image resolution and the camera calibration accuracy [137], given an accurate feature/ stereo matching. The ratio between the working distance and the baseline length also affects the accuracy, e.g., maximally

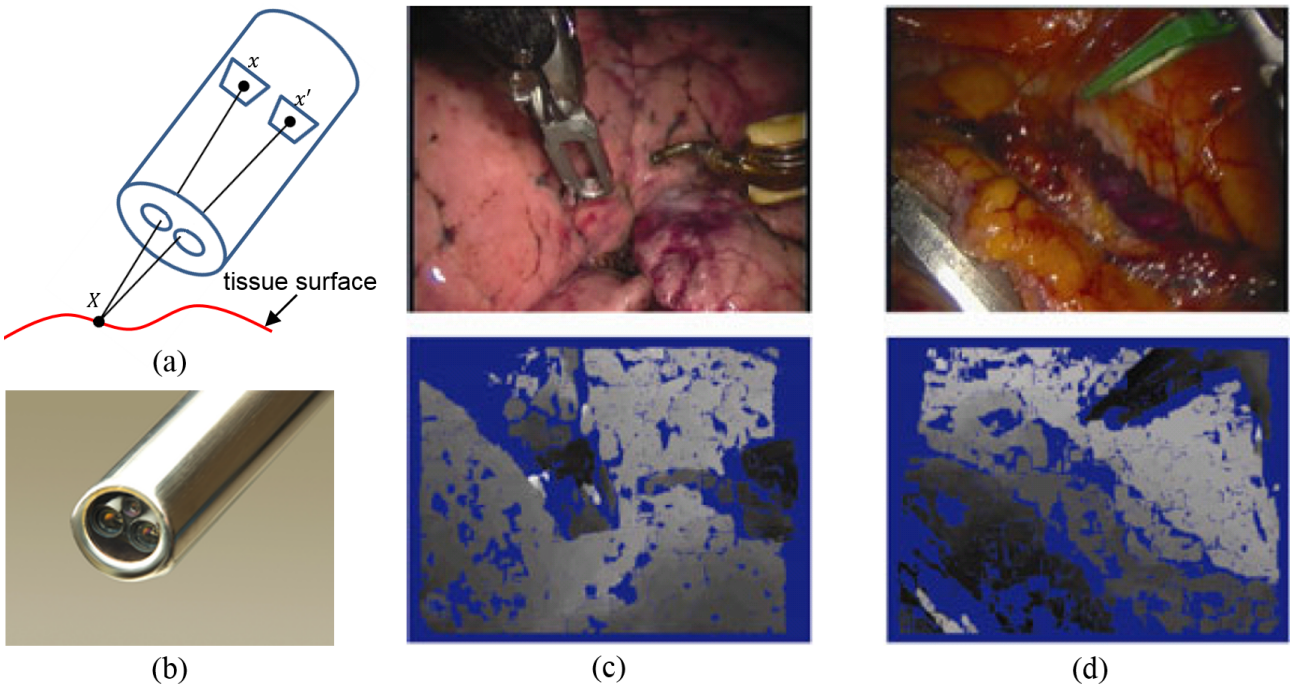


Figure 2.2: (a) schematic of a binocular endoscope pointing at tissue surface. The point X has projections x and x' on the two image planes. (b) The binocular endoscope from the Da Vinci® surgical system. (c)(d) The surgical scenes (upper row) and their corresponding disparity maps (lower row). Figure (a) is from [182], (c)(d) are from [264].

the working distance should not exceed $10 - 15$ times of the baseline length [182]. The main benefit of using stereoscopy for surgical scene 3D reconstruction is no requirement for additional hardware integration on the surgeon's end. And (near) real-time performance can be realised with algorithmic implementation on GPUs. However, the robustness of proposed algorithms are still to be tested, while images from different surgical applications vary. Those CNN based methods developed recently, although showing high potential in smoothness and speed, are also questionable in robustness given that the features learnt during training is dataset-specific. Most importantly, stereoscopy depends on the tissue surface texture information, so will lead to poor results in the case of textureless scenes or the existence of repetitive and ambiguous texture patterns. Last but not the least, monocular scopes are used in the majority of MIS procedures where stereoscopy is not able to be applied, which might hinder the application of stereoscopy in theatre.

Monocular SfM

Monocular SfM refers to the method for object surface reconstruction with multiple observations, provided by camera movements (Fig. 2.3(a)). Rigid SfM is applied in rigid scenes, with a moving camera. Once the displacement of the camera is known, it is equivalent to stereoscopy with a changing baseline [182]. However, in MIS or general surgery, the tissue keeps moving and deforming, which makes the reconstruction into a deformable SfM (DSfM) problem. Relatively more reliable DSfM utilises templates, of which the previous work will be reviewed concisely in this section.

DSfM follows the following three steps in general [182]: 1) reconstruction of a 3D deformable template with parameters that describe the target object; 2) correspondence search between the template and the captured image; 3) deform the template to match the image with the built correspondences. Step 1 is a one-off procedure carried out before the experiments, and step 3 is the key step which deforms the template with additional physical constraints. Salzmann and Fua, as well as Perriollat et al., proposed to model the template with the properties of paper [242, 217]. Keypoints with pixel colours were made use of to match the template to the captured image [219, 220, 88]. In these methods, the matching was carried out by minimising a cost function consisting of reprojection error and the physical constraints as the regulariser. Examples of applying DSfM can also be found in laparoscopic surgery [32, 183].

Despite several examples that reported applying SfM in surgical scenes, no stable and robust result was achieved in these works, as heavy task-specific parameter tuning is still needed to enable a promising reconstruction [182]. The main challenges are: 1) the textureless tissue surface; 2) the appearance of bleeding and smoke during surgery; 3) inefficient correspondence search between the captured images and the templates due to high difficulty in rectification; 4) tissue physical properties that are hard to model for template building. It is also claimed that SfM is capable of leading to dependable global shape recovery with limited manual interactions, but with the expense of losing surface details, especially in textureless/featureless regions [57].

SfS

SfS also uses only one camera for surface reconstruction. It models the relationship between the pixel intensities and the surface normals. Given a calibrated light source and camera position,

and the camera radiometric response, the surface can be reconstructed in two steps: firstly the constraints on surface normals are found; then these constraints are integrated for pixel-wise depth estimation [120]. It is noticeable that most SfS based algorithms make four assumptions: 1) the target object surface is modelled with Lambertian reflectance; 2) the reflectance (albedo) of the surface is known; 3) there is only one light source; 4) the shape is differentiable and its projection is continuous.

The fixed relative position between the light source and the scope in laparoscopic surgery ensures the feasibility of applying SfS [203, 300, 291]. Visentini-Scarzanella et al. proposed a novel formulation of SfS with a near point light source, aided by specular highlights [281]. A non-parametric light model for any light source shape has been proposed for SfS reconstruction [55].

In general SfS reconstruction is not a reliable technique since it highly depends on several assumptions like the Lambertian reflectance, known tissue albedo, and certain light source. Unlike SfM, SfS provides accurate surface detail, but may also result in distorted global shapes [57]. Recently, DSfM and SfS have been combined since they are complementary to each other [196]. SfX is advantageous that it can be integrated seamlessly to common surgical hardware. However, image matching is a major problem for SfM when the structures in laparoscopic images are poor and repeated [182]; the application of SfS is also restricted to the above assumption constraints which are not realistic for tissue surface in real-world situation.

ToF

In ToF, the distance between the target object and the detector is calculated by measuring the time that light needs to travel to the object surface and back to the detector (Fig. 2.3(b)). ToF cameras also have other names such as ranging camera, flash lidar, and depth camera. ToF cameras are able to acquire dense depth information with video-level framerate without a scanning mechanism. Pulsed and continuous wave (CW) modulation are the two most widely used approaches. Pulsed approaches send a pulse of light and measure its ToF for depth estimation [124, 296]. In medical applications a commonly used CW modulation measures the phase difference between the emitted and captured modulated incoherent light [295, 102]. There are already commercial mature ToF cameras available for the purpose of general applications

including gaming and gesture capturing, e.g. Microsoft® Kinect™ 2.0 (512×424 3D resolution), IFM® Efector™ O3D303 (176 × 132 3D resolution), SICK® Visionary-T™ (144 × 176 3D resolution), etc.

The ToF camera or so called range finder is combined with rigid standard endoscope optics, scanning the whole FOV and updating the 3D location of the tissue surface area by area [216, 103]. The spatial resolution of the depth map, however, is normally lower than that of the coupled RGB images. There are generally fewer works proposed on ToF systems adapted to laparoscopic surgery, compared with other optical techniques. But due to its fast acquisition time and non-reliance on object surface texture information, ToF should have high potential in MIS once the integration and miniaturisation of the device is achieved. Foix et al. summarised the error sources of ToF based reconstruction into systematic and nonsystematic errors. The former can be alleviated by calibration, while the latter by filtering [81]. These engineering challenges on the hardware building (integration to the endoscopic system, miniaturisation, measurement error, cost, low spatial resolution, etc.) hinders the application of ToF in MIS.

2.2.2 Structured Light

SL is another active method for 3D reconstruction. In SL, a specific setup is designed to generate a certain artificial pattern of light, and project it through a projector onto the target object (Fig. 2.3(d)). Then according to the deformation of the projected light pattern recorded by a mono camera, the object surface is estimated. Similar to stereoscopy, prior to the 3D reconstruction calibration should be applied to establish the spatial relationship between the projector and camera. In laparoscopy, some works have been proposed where the projector is integrated to the laparoscope [51, 148, 292]. SL systems are generally real-time and accurate, and also robust when reconstructed objects with featureless/ homogeneous surface [182]. However, a certain amount of offset between the projector and camera is required to establish the geometry of triangulation, and the designed light pattern could affect the accuracy of pattern detection to a large extent, which further influences the reconstruction accuracy. Also, a light pattern of insufficient density could result in missing of shape details.

In summary, as long as the appropriate light pattern is designed, SL seems to be a more promising and robust method with high potential, even when the surface of the target object is

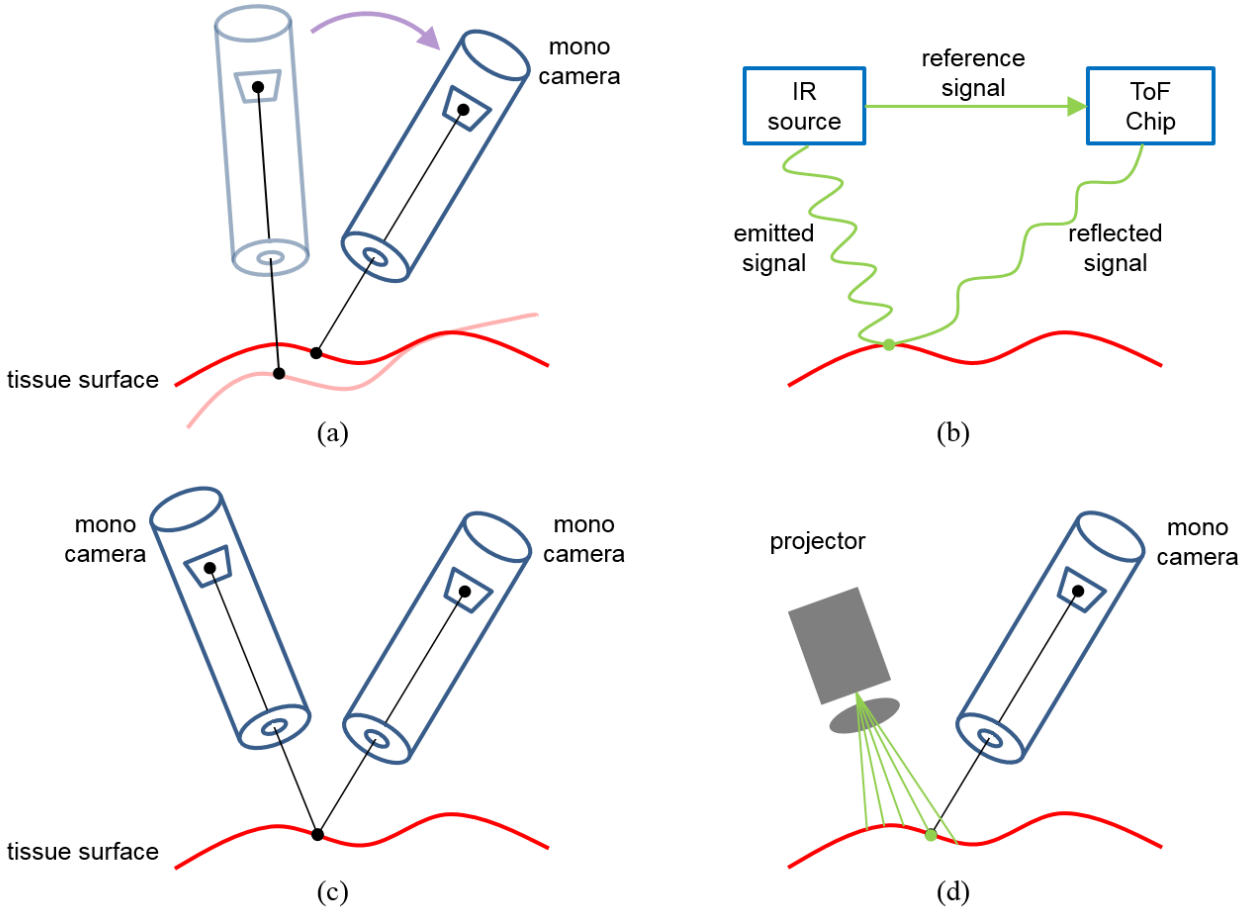


Figure 2.3: Schematics for different optical techniques: (a). monocular SfM for deformable tissue surface shape reconstruction. (b). Time-of-flight (ToF). Comparision between stereoscopy (c) and structured light (d). Figures reproduced and adapted from [182].

textureless; while other techniques have obvious drawbacks in certain situations which cannot be alleviated easily due to their own mechanisms [181].

Classification of SL Systems

SL techniques can be categorised according to their pattern codification strategies. Surveys regarding the classification were proposed previously [21, 240, 241, 90]. In this thesis, pattern classification based on Gengs categorization system is adopted, combined with some part of the work from Salvi et al. SL patterns can be divided into two categories: multi-shot (sequential) and single-shot. In multi-shot methods a sequence of light patterns are projected onto the object, either discretely or continuously, forming different codewords for specific locations on the surface. Single-shot applications, on the other hand, project one single unchanged pattern onto the target surface, and reconstruct the 3D shape according to the snapshot image.

Multi-shot methods include several subdivisions such as binary coding, grey coding, phase shifting, and frequency multiplexing [90].

Binary coding uses sequential black and white stripes or dots for projection, so that unique and non-repetitive codewords are assigned towards all different surface regions, as is shown in Fig. 2.4(a). For example, N patterns lead to 2^N codes; therefore 2^N points could be extracted after projection. Posdamer and Altschuler proposed the first work using binary code [221]. Their special projection system projects an $n \times n$ dot array by $n \times n$ laser rays onto the target surface, and a microprocessor is used to control the shutter for occlusion of part of laser rays at microsecond speed, creating specific codewords for all the dots. Sun et al. and Ishii et al. have also published papers regarding binary coding SL system [267, 127]. Binary coding techniques are highly reliable and insensitive to surface characteristics, since only white (1) and black (0) values are used to represent a bit. Whereas they require a large number of patterns for dense 3D reconstruction, leading to longer image acquisition time in which the target object should stay still.

In order to overcome the drawback of binary coding, in 1984, Inokuchi et al. proposed a SL system in which grey-level patterns are used instead of binary ones [125]. Grey-level coding was developed in order to reduce the number of patterns in one sequence. For instance, if M levels of intensities are employed in N patterns, then M^N codes could be generated. In practice, a vital issue when using binary or grey-level coding is defining the accurate location of stripes or dots. There are two ways to detect stripe edges with high accuracy: one way is to find the zero-crossing of the images second derivative; the other method is to find the intersection of both normal and inverse stripe patterns [241]. Some researchers studied optimization methods to find the balance between pattern number and reconstruction performance. Caspi et al. developed a grey coding SL system based on colour [39]. Horn and Kiryati proposed a work studying the optimization problem under certain noisy conditions [121].

Reliable as it is, binary/grey coding techniques also suffer from resolution limitation due to the discrete property of the patterns themselves. Phase shifting which achieves higher spatial resolution, was developed. A set of continuous sinusoidal patterns/gratings are projected onto the object with varying intensities of pixels along the coding axis for different phases (Fig. 2.4(b)). In order to convert the wrapped phase to absolute phase, phase deviation is calculated first by shifting light patterns several times (Fig. 2.4 (b)). Ono et al. designed a system which

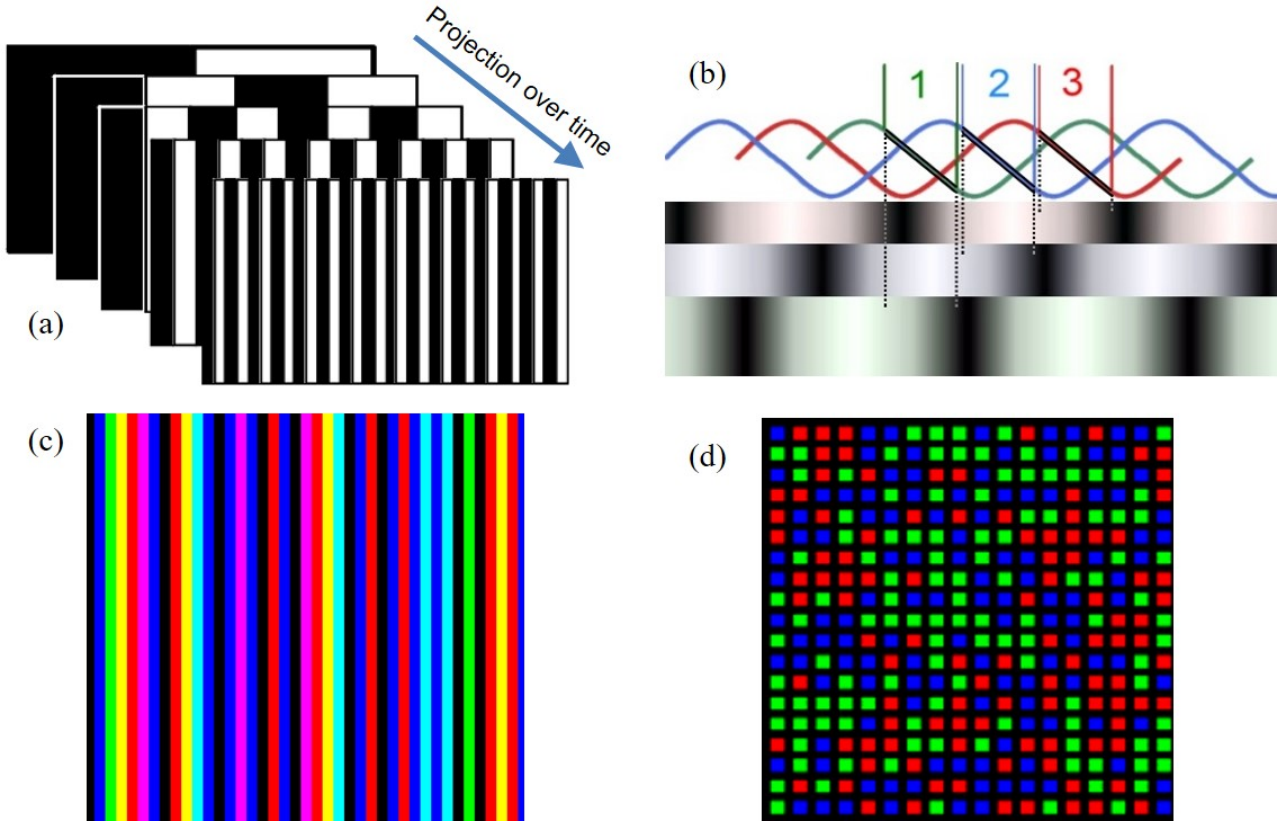


Figure 2.4: (a) Binary coding. (b) Phase shifting. (c) De Bruijn sequence. (d) M-array. Figure source: [241, 215].

makes use of phase shifting and unwrapping [206]. However, in reality phase shifting requires more than the theoretical minimum of 3 patterns for calculation. In order to overcome this drawback, Wust and Capson used overlapping patterns in red, green and blue colours. Thus, the camera can record all the three patterns respectively at the same time, reducing image acquisition time [293]. Unwrapping algorithm can only give a relative unwrapping and the absolute phase cannot be calculated, especially when the phase difference between two points is larger than 2π [90]. This so-called "ambiguity" problem can be solved by combing grey coding and phase shifting, since the former determines absolute phase and the latter provides higher resolution [122, 98] However, this hybrid method demands a larger number of patterns and therefore is unsuitable for moving objects.

In addition to methods mentioned above, there are also some other multi-shot SL patterns, like continuously varying colour coding and frequency multiplexing, which also require certain time length for image acquisition [90, 272, 53].

To sum up, although multi-shot methods lead to reliable and accurate reconstruction, they are

still not the optimal patterns for in MIS due to the continuous deformation of soft tissue and movement of the endoscope during surgery. Single-shot techniques, on the other hand, were proposed to deal with the moving object surface reconstruction problem.

Since only a snapshot image is recorded and employed to reconstruct the object 3D shape, different characteristics as well as neighbours are used in single-shot to label different areas. Given that only one image is used, detection and identification of each individual pattern unit becomes the key point for accurate reconstruction. These methods can be distinguished by the pattern unit's appearance — stripes, grids, dot matrix and so on.

Stripes are commonly applied in SL, either in 1D or 2D. There are different ways to identify each individual stripe. Previous works have been proposed to identify unites by using grey-scale, colours or segment patterns [31, 186, 68]. In most cases, the distribution of stripes with different characteristics (grey level, colour, segment, etc.) is arranged so that in a certain window no repetition of stripe combination can appear. An example is displayed in Fig. 2.4(c). Among different methods, the most representative algorithm is based on the De Bruijn sequence. A De Bruijn of rank n on a number set with a size k is a cyclic order in which words in k^n with length n is unrepeatable if counted around the cycle [180, 90]. This nature of uniqueness makes the De Bruijn sequence an algorithm with high potential in pattern distribution design. Many works were proposed based on this [123, 194, 304].

Grid indexing is another pattern strategy. This technique assigns unique labels to all sub-windows, with different combinations of binary or multivalued patterns, or by different colours [278, 101, 218].

Likewise, 2D array of colour coded dots which utilise spatial neighbours for identification were also applied in previous research. Payeur and Desjardins introduced an exhaustive way to generate unrepeatable sub-windows for the dot array (Fig.2.4(d)) [215, 59]. While some other researchers employed M-arrays (perfect maps) to arrange dots. Perfect maps, which can be seen as an extension of De Bruijn sequence in 2D space, were presented by Etzion in 1988 [71]. Griffin et al. and Adan et al. proposed their work by using perfect maps [101, 3].

Generally speaking, if the target object is static and no strict time constraint is required for image acquisition, multi-shot is more reliable and accurate. While for moving or deforming target, more errors would be introduced into codewords by using multi-shot; more accurate 3D

reconstruction results could be achieved according to images acquired in a short time instance by employing single-shot methods.

SL in Surgical Applications

Although SL has been used mostly for macroscopic imaging, early in 1994, Haneishi et al. proposed the first SL system aiming at gastrointestinal surface profilometry by integrating a laser beam projector with fiber bundle onto a normal electronic endoscope [107]. In this system a beam scanner is used to project one single dot sequentially onto different areas of the object surface, and an experiment was carried out on a synthetic resin, showing satisfying results (depth error < 0.16 mm). Laser beam scanners were used in works of Knaus et al. and Hayashibe et al. [148, 112]. In the Knaus approach, scanning to create a dense reconstruction (6400 points) resulted in long scanning time (27 seconds), making this system infeasible in real surgical situations. While in the Hayashibe method, scanning is much faster since a scanner with higher frequency was employed (1 Hz). According to the experiment result, when the distance between laparoscope and object is 10–15 cm, the estimated maximum error is 1.0–1.5 mm. Although these laser scanning based SL systems offer high accuracy, they also require a large number of images for 3D reconstruction. Ackerman et al. designed and enhanced their laparoscopic SL system for abdominal organ scanning [141, 1]. They used coded stripe patterns for projection and a 30 Hz camera for imaging, testing the equipment in an open surgery environment. It was claimed that high-speed cameras operating at up to 220 Hz could be integrated and image processing could be accomplished online. Hasegawa et al. developed a flexible SL system with fibres, using a spatial coding strategy of seven patterns to reconstruct the target surface [110]. The shape information could be acquired in about 0.5 second. But this system was only tested on ordinary objects such as tablet and vase. Chan et al. proposed another endoscopic SL system which uses a single-shot of a dot matrix as projection pattern [40]. The calibration method is based on Tsais algorithm [275], and the reconstruction experiments were applied on both the reflective model surface and tissues from forearm and oral cavity, showing satisfactory results: the error is smaller than 3% for reflective model and comparable results were found for curved tissue surface without quantitative analysis. For the purpose of improving colposcopy diagnosis accuracy of cervical cancer, another system using a grid pattern was used to reconstruct the tissue surface [292]. Recently, Maurice et al. developed a

rigid laparoscopic system with single-shot grid-based pattern designed according to the M-array (perfect maps) algorithm [189]. The setup was tested under an open surgery situation, and promising results were generated. Another recent one-shot endoscopic SL system was developed by Schmalz et al. [244]. This scanner was specially designed for tissue reconstruction in tubular cavities, so both the camera with a curved mirror and ring-shaped slide projector were designed in a very specific way to produce and record ring-like colour patterns. The instrument can already be used with a flexible endoscope and has been tested in *ex vivo* experiments.

To summarise, SL has advantages over other 3D reconstruction techniques in the following aspects: 1) non-reliance on tissue surface texture information; 2) user-defined patterns to facilitate correspondence search; 3) enables 3D reconstruction with one-shot SL system. These well indicate the potential of using SL in robustly reconstructing tissue surfaces during MIS. The challenges, however, are the miniaturisation and integration of the SL device to the endoscopic system.

2.3 Medical Hyperspectral Imaging

Hyperspectral imaging (HSI) aims to acquire the spectrum from all locations in a scene. As an RGB image contains three bands from the visible spectrum (VIS) — red, green, and blue, HSI results in a hypercube for a scene, with a much higher number of bands on the spectral dimension [290] (Fig. 2.5). In practice, HSI and multispectral imaging (MSI) are similar concepts except that normally the former contains hundreds of bands, while the latter contains more than three but less than one hundred bands. Due to the rich spectral information it carries, HSI/ MSI has vast applications in remote sensing [95, 77], crime detection [153, 248], water resource control [99, 2], food quality control [100, 75], etc.

Potentially HSI/MSI can also be used to aid medical applications, such as clinical diagnostics, surgical guidance, pathology detection, and so on, since it is a non-invasive approach. In the tissue the incident light undergoes scattering due to tissue inhomogeneity, and is absorbed by different structures like water, haemoglobin, melanin, etc. [310, 283]. The scattering, absorption and the resulting fluorescence characterise the tissue properties, and thus contains useful diagnostic information [276, 76, 74].

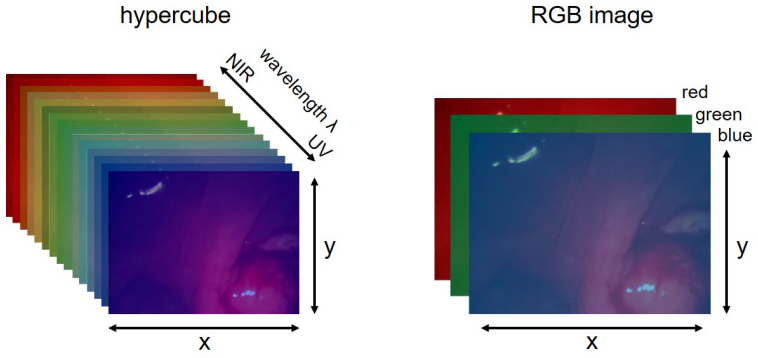


Figure 2.5: A hypercube (left) and an RGB image (right). A hypercube contains hundreds of bands along the spectral dimension, normally ranging from UV to NIR; An RGB image contains three bands (red, green, and blue) in VIS.

2.3.1 Light-tissue Interaction

As the light-tissue interaction characterises the tissue properties, in this section a brief introduction to light-tissue interaction, especially those factors related to HSI, is given. Scattering and absorption occur for multiple times when the light propagates through the tissue [213] (Fig. 2.6). The scattering is caused by the heterogeneity of the tissue's reflective index [132]. Multiple scattering may cause re-emission of the light from the surface, leading to diffuse reflection. Scatter structures include: collagen fibrils, macromolecular aggregates, lysosomes, vesicles, mitochondria, nuclei, etc.

Absorption limits the penetration depth of light in the tissue. Most tissues contains weaker absorbers (chromophores) in the spectrum ranging from 600–1300 nm [132]. Photons are absorbed by molecules when their energy meets the energy difference between adjacent electric energy levels. Since different types of molecules have unique energy level intervals, the wavelength of the absorbed light can be considered as unique fingerprints that can describe tissue properties [305]. For instance, the absorption difference between the oxyhaemoglobin and deoxyhaemoglobin can be used to estimate the oxygen saturation (StO_2) of the blood, and can be used to aid clinical decision making during organ transplantation or colorectal surgery [50]. Haemoglobin and melanin are the two most significant chromophores in the visible spectrum (VIS), while amino acids and protein dominate the ultraviolet (UV) spectrum, water for the infrared (IR) spectrum [289]. Either heat or luminescence including fluorescence and phosphorescence is produced in absorption. Fluorescence stems from the endogenous chromophores and

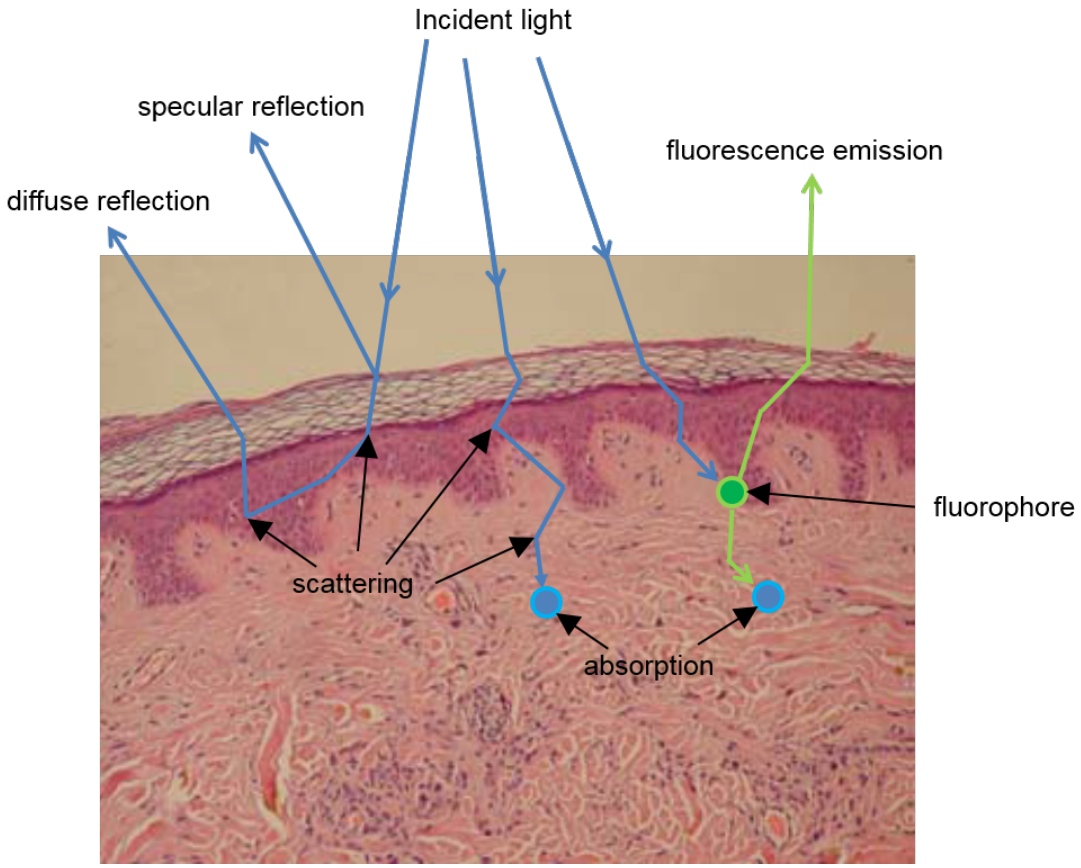


Figure 2.6: Examples on light-tissue interaction. With the incident light shed on the tissue, reflection, scattering and absorption events may happen multiple times and produce sepcular reflection, diffuse reflection, and fluorescence emission out from the tissue surface.

is called auto-fluorescence, and generally covers the spectrum from the UV to VIS. In general, reflectance reveals the local tissue properties in terms of scattering from structures and absorption by blood, while auto-fluorescence indicates the biochemical composition of the tissue [233]. Therefore, HSI which measures the spectrum of the fluorescence, or light reflected by/ transmitted through the tissue, containing rich medically valuable information.

2.3.2 Introduction to HSI Systems

A HSI system mainly consists of a light source, a spectral dispersive device, and a detector. The light source emits light onto the tissue surface; the dispersive device selects the excitation wavelengths either for the light source or the detector; the detector contains sensors to capture the reflectance and/or fluorescence. In the following an overview is given for medical HSI systems, including the acquisition mode, the dispersive device, and medical applications.

Acquisition Mode

This section provides a concise introduction to different acquisition modes of normal HSI systems. Generally most HSI systems can be divided into three types: spatial scanning, spectral scanning, and snapshot. HSI with spatial scanning refers to the system which generates the hypercube by sensing the whole spectrum through all locations of a scene, either point by point (whiskbroom) or line by line (pushbroom). Spectral scanning, on the contrary, scans the whole scene at a certain bandwidth at a time, and stacks all the 2D images along the spectral dimension. Snapshot methods manage to record all the spatial and spectral information by the detector in a snapshot, without scanning along any dimension.

Whiskbroom is widely used in remote sensing for earth resource technology [162]. In this mode the direction of the sensor is often directed by rotating a mirror to scan through different locations of the scene. The sample is scanned, area by area, along both X and Y dimensions. In each scanning the light is dispersed by a prism and a spectrum is detected by a linear array detector. After scanning all areas a hypercube can be built. In whiskbroom, a motorised stage with at least two degree-of-freedom (DoF) needs to be integrated for spatial scanning, leading to an increased system complexity. Besides, pixel-wise scanning in the spatial dimension makes complete imaging relatively time-consuming and inefficient. The main advantage of whiskbroom is the fewer number of sensor elements to calibrate for data acquisition.

Pushbroom, as the alternative with a spatial scanning mechanism, scans the whole scene line by line. A slit spectrum image is acquired in one exposure, with a prism-grating-prism (PGP) as the dispersive device. For a full scene data acquisition, the scanner is directed to scan along the spatial dimension perpendicular to the slit. Compared with whiskbroom, a pushbroom based system requires a significantly longer exposure time for one-shot spectrum capturing, as well as higher spectral resolution, and signal with lower signal-to-noise ratio (SNR) [6].

Staring is a spectral scanning approach for HSI acquisition. It measures the whole scene at a certain wavelength bandwidth in terms of a grey-level image in one exposure, and then spectrally stacks all the images to form a hypercube, normally enabled by filters. Liquid crystal tunable filters (LCTFs), for examples, use electronically controlled liquid crystal elements to select the wavelength range of the transmitted light. Compared with spatial scanning, staring grants more freedom in spectrum selection. Furthermore, while spatial scanning only displays

the scene after the completion of scanning, staring enables a live update since the full view at a certain wavelength can be updated at one exposure, facilitating aiming and focusing [132, 176].

Instead of scanning through either the spatial and spectral dimension to build the hypercubes, some work proposed optical systems for hypercube capturing in a snapshot. Snapshot based HSI systems directly image the remapped and dispersed scene image onto the detector/camera [288]. Despite the fast data acquisition, snapshot system generally outputs hypercubes with low spatial and spectral resolutions. The resolution is limited by the number of sensors on the camera [86].

To summarise different HSI acquisition modes, a comparison is provided in Table 2.1 that is reproduced from the review paper of Li et at. [162].

Table 2.1: Comparison of HSI acquisition modes: whiskbroom, pushbroom, staring, and snapshot. Table reproduced from [162].

Comparison	Whiskbroom	Pushbroom	Staring	Snapshot
Scanning	Spatial X and Y	Spatial X or Y	Spectral	N/A
Wavelength ranges	Wide	Wide	Medium	Wide
Wavelength selectability	Partial	Partial	Yes	No
Spectral resolution	High	High	Medium	Low
Band numbers	Hyperspectral	Hyperspectral	Hyperspectral	Multispectral
Acquisition time	Long	Long	Short	Shorter
Cost	Low	Low	High	Medium

Dispersive Device

The dispersive device is a key component in a HSI system. It is either used for excitation wavelength selection or emission light dispersion. In the former case it is positioned in between the light source and the sample, while in the latter case between the sample and the detector [174]. In general there are three types of dispersive devices: monochromator, bandpass filter, and snapshot imager.

Monochromators are capable of separating polychromatic light into beams with different spectra [208], and are widely used in spatial scanning HSI systems. Two representative monochromators use diffraction gratings and prisms. In general diffraction gratings lead to simpler system design, but prisms are advantageous in higher light transmission and lower scattering in VIS and NIR. A hybrid prism-grating-prism (PGP) monochromator containing a transmission grating fixed between two prisms with filters for wavelength selection, has been used in industry and research, with a small and compact structure, to sense the light with the wavelength between 320 and 2700 nm. It has been adopted in several medical HSI systems [10, 9, 163, 7, 143].

Optical bandpass filters are normally used in spectral scanning HSI systems, and are either fixed or tunable. The fixed filters can be mounted on a wheel to select the light with certain wavelengths. These filter wheels can be used for MSI, by capturing sequential grey images in spectral dimension by rotation (e.g. SpectroCam Multispectral Camera, Ocean Optics, USA). However, these systems only enable hypercube collection with limited bands [4, 232, 42, 135]. Despite the convenient utilisation, the filter wheel still suffers from low spectral resolution, slow image acquisition, and consequently the mis-registration between images at different wavelengths [66]. To the contrary, tunable bandpass filters including liquid crystal tunable filter (LTCF) and acousto-optical tunable filter (AOTF) are widely deployed in HSI systems, due to their high spatial resolution and faster acquisition [89]. They work from VIS to NIR, and UV to IR, respectively. In general LCTF leads to higher image qualities but slower acquisition, compared with AOTF [174]. It is also noticeable that in many applications both are still too slow due to the required switching time.

A typical snapshot imager is a computer-generated hologram (CGH). CGH is based on a 2D grating formed by pixels arranged in an array. It enables a snapshot hypercube acquisition with lower resolution than other scanning based dispersive devices. Examples of its application

in HSI system can be found in [134, 108].

2.3.3 Medical Applications

The reflectance and fluorescence from the tissue reveal the local scattering and absorption properties, as well as the biochemical composition [233], so they can be used in different medical applications including pathology detection and surgical guidance.

Pathology Detection

The metabolism and morphology of cells, as well as their scattering and absorption properties, can be altered in the case of pathology [76, 23]. Thus, potentially, HSI which acquires reflectance and fluorescence could provide medically meaningful information for disease diagnosis.

Much previous research has focused on cancer detection with HSI. Kaposi's Sarcoma (KS) is a type of cancer that occurs in skin and the mouth. As a highly vascularised tumour, KS is indicated by cutaneous lesions which refers to skin patches characterised by texture, thickness, and pigment in contrast [256]. A 6-band MSI system has been developed to monitor the blood volume and oxygen saturation before anti-angiogenesis surgery [111]. Dicker et al. proposed to distinguish malignant and benign skin tissue in a histology study with high resolution HSI [60]. Hypoxia in breast cancer contains prognostic value for radiation and surgery, and is led by insufficient oxygen supply. HSI has been used by Sorg et al. to monitor the *in vivo* haemoglobin of the mouse mammary carcinoma. A spectral mixture analysis was then performed on the acquired data for protein classification. Boucheron et al. applied MSI on H& E-stained breast cancer biopsy samples, and was found that MSI provided richer information to aid classification than normal RGB images do [29]. Cervical cancer detection is another application where HSI can potentially outperform the current gold standard method, the Pap smear test which might lead to 15–40 % false positive rate. Some HSI based methods were proposed given the fact that the auto-fluorescence signal is an important indicator for cancer [214, 229]. Ferris et al. proposed a HSI imaging system to capture both the reflectance and fluorescence from the cervical epithelium on the ectocervix, on a diverse population of women [76]. The wavelength-of-interest covered the UV and VIS, and is capable of differentiating high-grade cervical lesions, less-severe lesions, and normal tissues. Benavides et al. incorporated MSI with a colposcope to

obtain reflectance and auto-fluorescence images of the cervix [23]. It was demonstrated that by combining reflectance and fluorescence the cervical intraepithelial neoplasia lesions and normal tissues could be distinguished.

HSI has also been used to aid retinal disease diagnosis. Chohen et al. integrated an HSI system to the fundus camera for the evaluation of the effectiveness of eye surgery [54]. Hirohara et al. used a fundus camera with an LCTF to quantify the absorption of vessels, by measuring the reflected light that was transmitted through the eye vessels and their surrounding tissues [117]. HSI has also been adopted to evaluate the progress of age-related macular-degeneration (AMD) [73, 249].

Diabetic foot is another condition whose diagnosis can be aided by HSI techniques. One of the major complications of diabetes is diabetic foot ulceration [258], which might lead to partial amputation if untreated. Diabetic foot can be diagnosed and monitored according to the microcirculation and large vessel changes. By controlled experiments, Khaodhiar et al. utilised HSI to measure the oxy-/deoxyhaemoglobin to predice diabetic foot ulcer healing with high sensitivity and specificity [142]. Another work on the healing prediction was carried out by Yudovsky et al., using HSI to capture images between 450–700 nm.

HSI has been used to assist diagnosing other diseases such as heart and circulatory pathology [147, 46], shock [38], anaemia [314, 315], etc.

Surgical Guidance

Visual feedback guides surgical decision making [84]. However, in either traditional open surgery or MIS, the visualisation of the patient anatomy is usually the scene of tissue surface in VIS. As a technique that captures molecular and cellular level information, the function of HSI/ MSI has been explored to enrich the surgical visual feedback for surgical guidance. The applications of HSI in surgical guidance include but are not limited to: 1) oxygen saturation monitoring [311, 50]; 2) the visualisation of tissue under vascular structures [195]; 3) cancer detection for tumour removal [23, 214]. Since HSI/ MSI provide rich diagnostic information and is generally easy to integrat with surgical devices, it has been used to aid different surgeries in previous research.

For small renal cortical tumour removal, laparoscopic partial nephrectormy has drawn more

and more attractions compared with the traditional gold standard open surgery. One of the most important information for decision making in this MIS procedure is via monitoring ischemia. Holzer et al. proposed a HSI system to measure renal parenthetical StO_2 by measuring reflectance from 520 to 645 nm, in order to find kidney response during the open partial nephrectormy [119]. Digital light processing based HSI has been used for StO_2 measurement to aid robotic laparoscopic partial nephrectormy [205]. Zukak et al. also worked on the project of LCTF based HSI guided renal surgeries, developing a real-time system that can has been tested on *in vivo* porcine kidney [313].

Laparoscopic gall bladder surgery on the other hand, is another operation that can be aided by HSI. Due to the loss of tactile feedback, the identification of biliary tree suffers from low image contrast through the laparoscope. An endoscopic HSI system has been developed for swine tissue identification at molecular level [312]. It was shown in this work that the bile ducts which contains lipid can be recognised, based on the fact that lipid absorbs 930 nm light. Another work was carried out by Mitra et al. for biliary anatomy classification with classifiers working on the reflectance and fluorescence data captured by HSI [193]. This work was validated on *ex vivo* swine.

In mastectomy, the existence of residual tumour needs to be examined to assess the completion of tumour removal intra-/post- operatively [79]. Panasyuk et al. proposed to use HSI for tumour detection. On the rat breast model, they demonstrated the successful detection of tumour with sizes of 0.5–1 mm with 89% sensitivity and 94% specificity.

There are many other studies that utilised HSI/ MSI for surgical guidance, e.g. abdominal surgery [50, 8] and intestinal surgery [10].

To summarise, by comparing different HSI systems in terms of their acquisition modes, structures, and medical applications, we find some hints about endoscopic HSI system design. HSI has high potential to be used in disease detection and surgical guidance, as it yields signals with rich information on tissue, cellular, and molecular level. There is no "gold standard" HSI system structure, as different system have different advantages and disadvantages. Thus, the integration of HSI into the surgical workflow is a task-specific problem and should be planned accordingly. Since real-time imaging is unaffected by tissue deformation and movement is a general requirement, HSI systems with snapshot acquisition mode are preferred. However, the resolution of the acquired hypercube needs to be increased, e.g. by means of algorithmic sup-

port. The minimisation of the system is another requirement that the endoscopic HSI system should satisfy in order to use them in real surgical environments.

This work is a true combination of medical image computing (MIC) and image guided intervention (IGI), thus includes innovative work on both the hardware and algorithmic development. The convolutional neural networks (CNNs) based algorithms were developed to establish the software platform of the hybrid system that captures both tissue surface shape and hyperspectral signals. Therefore, in the following a brief introduction to CNN is provided.

2.4 Convolutional Neural Networks

CNNs have been used as tools to implement key procedures in this project, including segmentation for real-time structured light pattern decoding, and super-spectral-resolution for fast pixel-level-dense hypercube estimation. Therefore, a brief introduction towards CNN is given in this chapter, and concise literature reviews regarding CNN based segmentation and super-resolution algorithms are provided in Chapter 5 and Chapter 6, respectively.

2.4.1 From Neurons to Layers

Artificial neural networks (ANNs) are parametrised computational systems that map an input to an output, and are inspired by biological animal brains [116]. A typical ANN can be structured as input layers, hidden layers, and output layers, where the hidden layer consists of neurons as basic computational units (Fig. 2.7(a)). Neurons take the input and apply linear computations to generate a corresponding output, followed by an activation function, as is shown in Fig. 2.7(b). Each neuron is connected to neurons in the previous layer, and its output is fed into neurons in the next layer. For a single neuron connected to N neurons in the previous layer, assume its inputs as x_n, weights w_n, where $n \in (0, \cdots, N)$, bias b set to w_0 (assume $x_0 = 1$), then its output can be represented as:

$$u = \sum_n w_n x_n + w_0 \tag{2.2}$$

With an activation function, the output y can be calculated as $y = h(u) = f(x)$. Different activation functions can be chosen, such as sigmoid ($y = 1/(1 + e^{-u})$), rectified linear unit (ReLU, $y = max(0, u)$) [93], Leaky ReLU [179], exponential linear units (ELUs) [52], and so on.

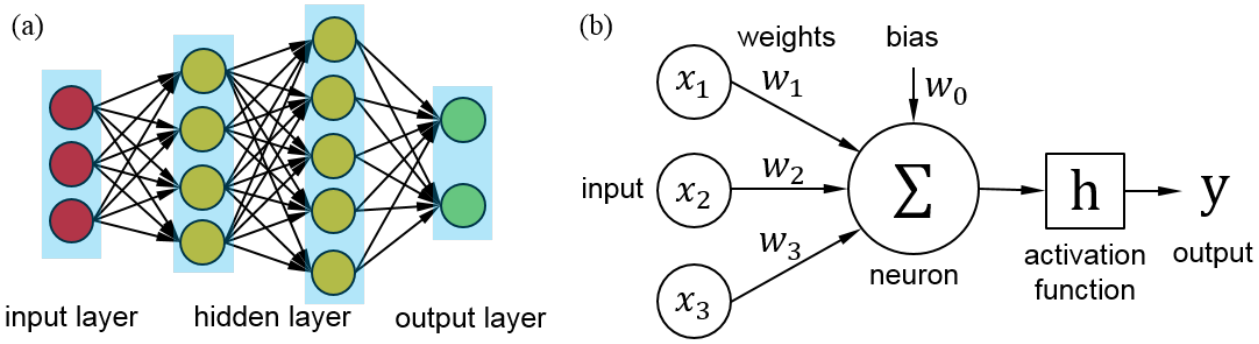


Figure 2.7: (a) A typical fully connected artificial neural networks. (b) A linear computation in a neuron followed by an activation function.

For the hidden layer m, the outputs y_j can be calculated as:

$$y_j^{(m)} = h^{(m)}(\sum_n w_{nj}^{(m)} x_{nj} + w_0) \tag{2.3}$$

Piling up M layers ($M > 1$),

$$y = (f^{(M)} \circ \cdots \circ f^{(1)})(x) \tag{2.4}$$

where $\circ$ denotes function composition and $f^{(m)}$ denotes a mapping from $\mathbb{R}^{D_{m-1}}$ to $\mathbb{R}^{D_m}$, for layer m.

The neuron connections between adjacent layers determine the layer function. A traditional and widely used layer is the fully connected (FC) layer, in which neurons are connected to all the neurons in the previous layer. This layer extracts features from given input vectors and feeds its output data to the next hidden layer. When the FC layer is used as the output layer, it typically generates scores for classification or regression. However, an FC layer is not the optimal layer if directly used for image feature extraction, since it doesn't scale well to input image size [161]. Every neuron in the first hidden FC layer needs to connect with every pixel in the input image, resulting in an explosive increase in connection number with an increase of image size. Therefore, it is wasteful and computationally expensive to process the image directly with FC layers, leading to severe overfitting. Instead, independent data pre-processing

including feature extraction is a commonly applied to feed the extracted "feature vector" into the FC layer, in image classification [152, 27, 262] or segmentation [207, 231, 228] which can be modelled as pixel/patch-wise classification. Different features have been proposed to describe an image patch, e.g. derivatives, Hessian matrices, Scale-Invariant Feature Transform (SIFT) [173], Speeded-Up Robust Features (SURF) [22], KAZE [11], and so on. However, as hand-craft features with specific performance, their generalisation and robustness aren't optimal in many cases, since image segmentation tasks vary and need customised feature descriptors.

In order to reduce overfitting as well as to prevent information loss using hand-craft features, networks containing convolutional (Conv) layers are introduced to extract features from images, facilitating different problems such as classification, detection, segmentation, super-resolution, etc. A Conv layer consists of neurons arranged in 3D: two spatial dimensions and one channel (feature) dimension. This can be seen as a FC layer with an infinitely strong prior [96]. For a typical Conv layer this prior says 1) the weights w_n ($n \in (1, \cdots, N)$) and bias w_0 of every neuron is the same as those of others in the same channel, which is referred to as the "parameter sharing" property; 2) all the weights except for those in a small local receptive field, are set to 0, for a single neuron, which is referred to as the "local connectivity" property. Through a learning procedure, different sets of weights and biases can be learnt by different channels. During a feed-forward operation the computation can be seen as convolving the input in spatial dimensions using different convolutional kernels (filters) along the channel dimension.

A convolution between continuous functions x and w is defined as:

$$(x * w)[t] \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} w(\tau) w(t - \tau) d\tau \tag{2.5}$$

When discretised:

$$(\mathbf{x} * \mathbf{w})[t] \stackrel{\text{def}}{=} \sum_{\tau} x[\tau] w[t - \tau] \tag{2.6}$$

If an image is denoted as $\mathbf{I}$, its ith row and jth column value as $\mathbf{I}(i, j)$, the convolutional kernel as $\mathbf{C}$, then convolving $\mathbf{I}$ with $\mathbf{C}$ can be written as:

$$(\mathbf{I} * \mathbf{C})[i,j] = (\mathbf{C} * \mathbf{I})[i,j] = \sum_m \sum_n \mathbf{C}[m,n]\mathbf{I}[i-m, j-m] \tag{2.7}$$

A schematic of a typical Conv layer is shown in Fig. 2.8(a).

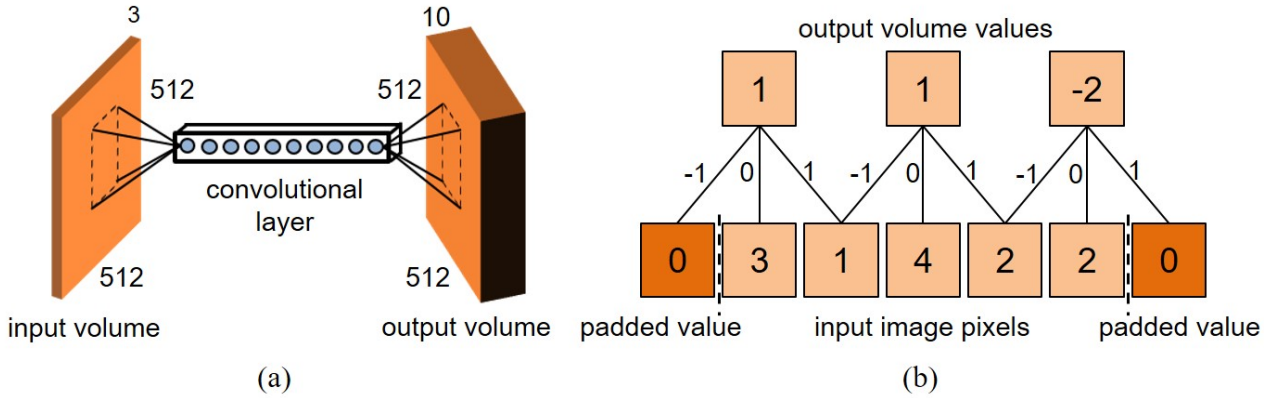


Figure 2.8: (a) A Conv layer with 10 channels. Input: $512 \times 512 \times 3$ RGB image; output: $512 \times 512 \times 10$ volume. The visually protruding cubic is used to show the receptive field of one neuron. (b) A convolution example with kernel size 3, stride 2, and zero-padding.

The convolutional behaviour of the Conv layer therefore greatly reduces the number of parameters to estimate during the training procedure, thus preventing overfitting introduced by the FC layer. Since different convolutional kernel are widely used as feature detectors in low-level edge/object detection problems [259, 35], the trained kernels in the layer can be considered as feature detectors in different hierarchies, where the existence of a desired image feature is expected to lead to a high response after convolution.

There are four key hyperparameters to define in describing a typical Conv layer: depth, kernel size, stride, and padding. Depth determines the number of different convolutional kernels in a layer, and hence the channels of the output volume; kernel size defines the size of the filter and hence the local connectivity of the neurons; stride is the interval between neighbouring convolution; padding decides the padded values around the input volume in order to equalise the size of the data volume before and after convolution. A convolution with kernel size 3, stride 2, and zero-padding is shown in Fig. 2.8(b). If the input image/ feature map contains multiple channels (e.g. the RGB image), as is shown in Fig. 2.8(a), every filter is expanded to 3D with the same number of channels as the input. Then correlation between the input and the filter occurs channel-wisely, and all the values on different pixel locations are summed up in the channel dimension to form a feature map with single channel. Stacking single-channel feature

maps generated by other filters in the same Conv layer, the output multi-channel feature map is obtained.

Conv layers are usually combined with other layers including activation layers, pooling layers, FC layers, etc., to solve image processing tasks. Taking an image classification problem for example, multiple Conv layers are chained and applied on different image scales. The downscaling of the data is realised by pooling layers. This typical "Conv-downsampling-Conv-downsampling" structure, named "encoder" in some works, enables a "deep" task-specific feature extraction on different image scales. The FC layers are appended after a series of Conv and pooling layers, to use the extracted features for classification, where neurons in the last output layer stand for different class scores. A classic CNN model used for image classification is AlexNet [151] (Fig. 2.9). As a milestone, AlexNet won the ILSVRC 2010 by a large margin, with a 10.8% higher top-5 accuracy than the second best method.

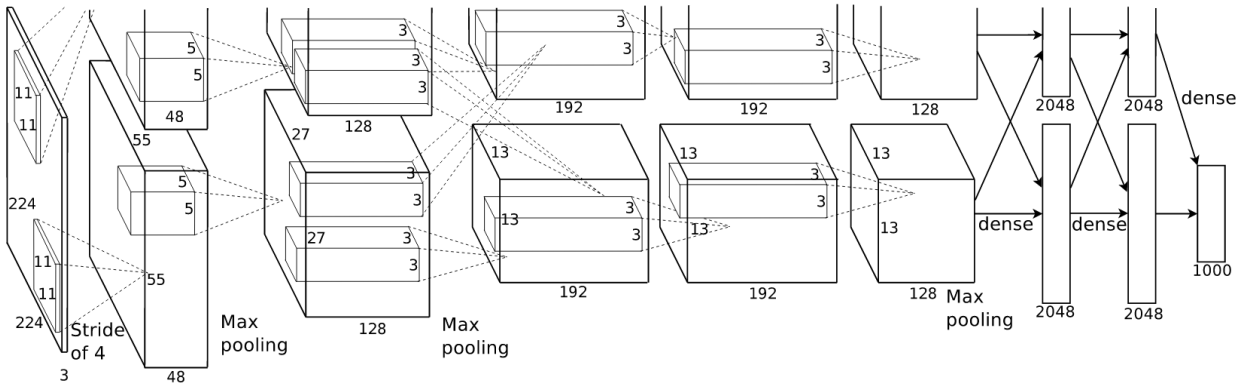


Figure 2.9: The schematic of AlexNet used for image classification. Figure source: [151]

Similar to FC layers, a typical Conv layer is normally followed by an activation function.

The pooling layer is normally used to progressively downsample the input volume in the spatial dimensions (width, heigth), enabling feature extraction on different scales. Two commonly adopted pooling layers are max pooling and average pooling: the former replaces the pixels in the receptive field with the maximum value in a local area; while the latter uses the average value. In practice generally max pooling with 2×2 or 3×3 filter size and a stride of 2 is preferred in normal CNN architectures.

There are several other milestones in CNN model development, e.g. VGG [257], GoogLeNet [269], ResNet [114], and so on. All of those models follow a encoder-classification pipeline, and differ in terms of detailed layer arrangement and connection manner. Although they were

proposed to be applied to different tasks— mainly image classification— their architectures can be adapted to be applied to other tasks such as image detection, segmentation, and super-resolution.

2.4.2 Network Training

Network training refers to the procedure of looking for optimal weights $\mathbf{W}$. This procedure can be realised by training via minimising a user-defined loss function. A typical loss function consists of two parts, data loss L_d and regularisation loss L_r. L_d depicts the difference between the predicted output $\hat{\mathbf{y}}$ and the ground truth $\mathbf{y}$; while L_r is mainly used to prevent overfitting [96]. The loss function L is defined by the sum of L_d and L_r:

$$L = L_d + \lambda L_r \tag{2.8}$$

where λ is called "decay weight" used to adjust the weight of the regularisation loss in the final loss function. The data loss L_d is defined according to detailed tasks.

In a single value regression problem, given one example $(\mathbf{x}, \mathbf{y})$, and a CNN model architecture $\mathbf{y} = f(\mathbf{x}, \mathbf{W})$, then a commonly used data loss function, the $L2$-loss, is defined as:

$$L_d = \frac{1}{2}(\hat{\mathbf{y}} - \mathbf{y})^2 = \frac{1}{2}(f(\mathbf{x}, \mathbf{W}) - \mathbf{y})^2 \tag{2.9}$$

In a classification problem, the cross-entropy loss function is used:

$$L_d = -\sum_i y_i log(\hat{y}) \tag{2.10}$$

where i denotes the class index. if y_i stands for the probability of class i, with a given ground truth class i_{GT},

$$y_i = \begin{cases} 1 & if \quad i = i_{GT} \\ 0 & \text{otherwise} \end{cases} \tag{2.11}$$

Two commonly used regularisation losses are the $L1$-regulariser:

$$L_r = \sum_i |w_i| \tag{2.12}$$

and the $L2$-regulariser:

$$L_r = \sum_i w_i^2 \tag{2.13}$$

Denoting the first-order derivative (gradient) of the loss function L at $\mathbf{w}$ as $\nabla L(W)$,

$$\nabla L(\mathbf{W}) = (\frac{\partial L}{\partial w_1}, \cdots, \frac{\partial L}{\partial w_N})^T \tag{2.14}$$

The global minimum of $L(\mathbf{W})$ is located in the points where $\nabla L(\mathbf{W}) = 0$. However, since $L(\mathbf{W})$ is non-convex, so many points including local maxima, local minima, and saddle points, satisfy $\nabla L(\mathbf{W}) = 0$. So instead of a global minimum it is more feasible to find a local minimum instead, by iterative optimisation algorithms. In practice the local minimum found by iterative optimisation methods is good enough for general applications [47].

A widely used method for loss function minimisation is gradient descent optimisation. As an iterative method, it starts with a initialised $\mathbf{W}^{(0)}$, and updates $\mathbf{W}$ by the gradient of $L(\mathbf{W})$ iteratively:

$$\mathbf{W}^{(\tau+1)} = \mathbf{W}^{(\tau)} - \alpha \nabla L(\mathbf{W}^{(\tau)}) \tag{2.15}$$

where $\alpha > 0$ is the learning rate that controls the update magnitude of $\mathbf{W}$ in every iteration. In order to prevent convergence at small local minima, momentum is introduced for the update of $\mathbf{W}$, by linear combination of current gradient and previous update [239]:

$$\mathbf{V}^{(\tau+1)} = m\mathbf{V}^{(\tau)} - \alpha \nabla L(\mathbf{W}^{(\tau)}) \tag{2.16}$$

$$\mathbf{W}^{(\tau+1)} = \mathbf{W}^{(\tau)} + \mathbf{V}^{(\tau+1)} \tag{2.17}$$

where $\mathbf{V}$ denotes the velocity and $0 \leq m \leq 1$ the momentum that adjusts $\mathbf{V}$.

Making use of the differentiability of most of the layers in the neural network, the back-propagation (backprop) algorithm [239, 157, 158] is widely used to compute the gradients

$\nabla L(\mathbf{W})$ and consequently update weights $\mathbf{W}$ in iterations in neural networks. The name "back-propagation" comes with the information flow direction: starting with a loss that is computed via forward propagation, the information flow backwards from the last layer to the first layer to update the gradient $\nabla L(\mathbf{W})$, making use of the chain rule in derivative computation. Details are not provided in this thesis frame but the three general steps of backprop are listed as follows:

1. Feed forward the input $\mathbf{X}$ and compute the loss L using a defined loss function.
2. Compute the loss L_j for hidden layers.
3. Compute the partial derivatives $\nabla L(\mathbf{W})$ for weights using the chain rule.

Due to the application of the backprop algorithm, a "vanishing gradient" problem can be found when using traditional activation functions such as sigmoid function, especially in deep neural networks (DNN) [245]. Therefore in DNN, ReLU or its variants like Leaky ReLU and ELU are generally preferred.

2.5 Discussion and Conclusions

The aim of this project is to develop a hybrid system which enables both tissue surface shape reconstruction and hyperspectral imaging (HSI). The hardware design and algorithmic development are included in detail. The literature provided in this chapter covers both aspects. For the hardware, an introduction and review of state-of-the-art techniques for surface shape reconstruction and HSI was demonstrated; for software, a brief introduction of convolutional neural networks (CNNs) was given, as CNNa are important techniques used to handle key challenges of this work.

Tissue surface reconstruction provides clinically rich information that can be potentially used in medical applications, such as pathology detection, robotic surgery, AR based surgical guidance. In MIS where the tactile force feedback and depth perception are missing for surgeons, the intra-operative tissue shape enriches surgeons' visualisation of the surgical scenes.

Concise reviews were provided for different optical techniques for 3D reconstruction, including stereoscopy, monocular Structure-from-Motion (SfM), Shape-from-Shading (SfS), Time-of-Flight (ToF), and Structured Light (SL).

In surgical scenes, the main challenges are the deformation and moving of tissue as well as low image qualities, which requires a fast one-shot tissue reconstruction. This requirements can only be potentially met by stereoscopy, ToF, and SL. As stereoscopy depends on texture information, and ToF suffers from high building difficulty and systematic errors that cannot be compensated by calibration, SL has been chosen as the 3D reconstruction techniques in this work. A miniaturised endoscopic SL system is described in Chapter 4 and Chapter 5.

Although light-tissue interaction might lower the image quality for optical based 3D reconstruction techniques, it contains medically rich information. The reflectance reveals the tissue properties in terms of scattering and absorption, while auto-fluorescence indicates the biochemical composition of the tissue. Therefore, hyperspectral imaging (HSI) that is capable of capturing the reflectance and fluorescence signal as a 3D hypercube ($x - y - \lambda$), has potential to be used to aid medical applications including disease diagnosis and surgical guidance.

There is no "gold standard" HSI system that is suitable for all applications, as different systems have their own advantages and disadvantages. So the HSI systems should be customised depending on the target applications. While the endoscopic system in this project should couple with surgical scenes where tissue movement and deformation exist all the time, the two main challenges for HSI to address are: 1) the real-time hypercube acquisition; 2) the miniaturisation of the system. Therefore, snapshot acquisition mode is considered as it provides a one-shot hypercube collection and is unaffected by tissue movements. To further compensate the low hypercube resolution, an algorithm that manages to predict a pixel-level dense hypercube has been developed (section 6.3).

An introduction to convolutional neural networks (CNNs) is given, as CNN is an important tool in two key technical contributions in this work: the real-time SL pattern decoding (section 5.3.2), and the super-spectral-imaging algorithm (section 6.3). Literature reviews are provided for CNN based segmentation and super-resolution methods in section 5.3.1 and section 6.3.1, respectively.

Chapter 3

The Naive 3D Reconstruction Pipeline for the ICL SL System

3.1 Introduction

In this chapter an endoscopic structured light (SL) system (ICL SL system) [51] for tissue surface shape measurement is described. Making use of structured light techniques, the 3D reconstruction with the ICL system does not rely on object surface texture information. This chapter describes both the hardware and software platforms, and focuses on the latter, i.e., a naive 3D reconstruction pipeline which applies general 3D reconstruction, with a fixed relative position between the probe and the endoscope. Detailed descriptions of system calibration, pattern decoding, and triangulation are included.

The content of this chapter has been published in:

1. J. Lin, N. T. Clancy and D. S. Elson, "An endoscopic structured light system using multispectral detection", International Journal of Computer Assisted Radiology and Surgery, vol. 10, pp. 1941-1950, 2015.

3.2 The ICL SL System

The whole tissue surface shape reconstruction pipeline has been enabled on the previously developed ICL SL optical system [51]. This system centres around an incoherent fibre bundle. In total 171 fibres are arranged in a linear array at one end of the fibre bundle, and randomised at other end. This randomised end is called the probe in the following text. During SL imaging, the collimated output of the visible wavelengths ($420-750$ nm) of a 4 W supercontinuum laser (SC400-4, Fianium Ltd., Southampton, UK) was dispersed by a SF-11 prism before being focussed onto a linear array end. Those fibres, each carrying light of a unique narrowband spectrum, passed through a custom-made converter to form a random distribution at the distal end. A gradient refractive index (GRIN) objective lens formed an image of the bundle end-face resulting in a random spot pattern at the target sample, in which each spot contains unique wavelength information. The brass ferrule of the fibre linear array end has a diameter of 12.5 mm, but the fibre bundle and distal tip of the probe has a maximum outer diameter of 2.1 mm. This miniaturised size enabled the insertion of the probe through the biopsy channel of standard flexible endoscopes. The system schematic and the projected pattern are shown in Fig. 3.1.

By analysing the projected SL images, i.e. detecting and identifying individual spots, the 3D locations of those spots can be calculated, enabling object surface measurement. The reconstruction density therefore depends on the number of fibres. Developing robust algorithm to decode the light pattern is vital for recovering the 3D shape of the target object, however, several challenges need to be addressed to achieve robust light pattern decoding in real-world applications: in general using the ICL SL system the colour and neighbourhood information is required for decoding. However, spots with similar wavelengths cannot always be differentiated unambiguously due to the broad-band RGB spectral response of the camera. Furthermore, at the blue end of the spectrum light is absorbed strongly by molecules like haemoglobin, resulting in dim spots, while the higher penetration depth at the red end results in increased lateral diffusion and a decrease in edge contrast of the projected spots. The appearance of liquid and fat also leads to image blurring during imaging. Furthermore, tissue areas with high curvature deform the shapes of the projected spots. All of the above problems lead to an increased difficulty in pattern decoding. In the following, a naive 3D reconstruction pipeline is proposed to recover the object surface shape with the ICL SL system.

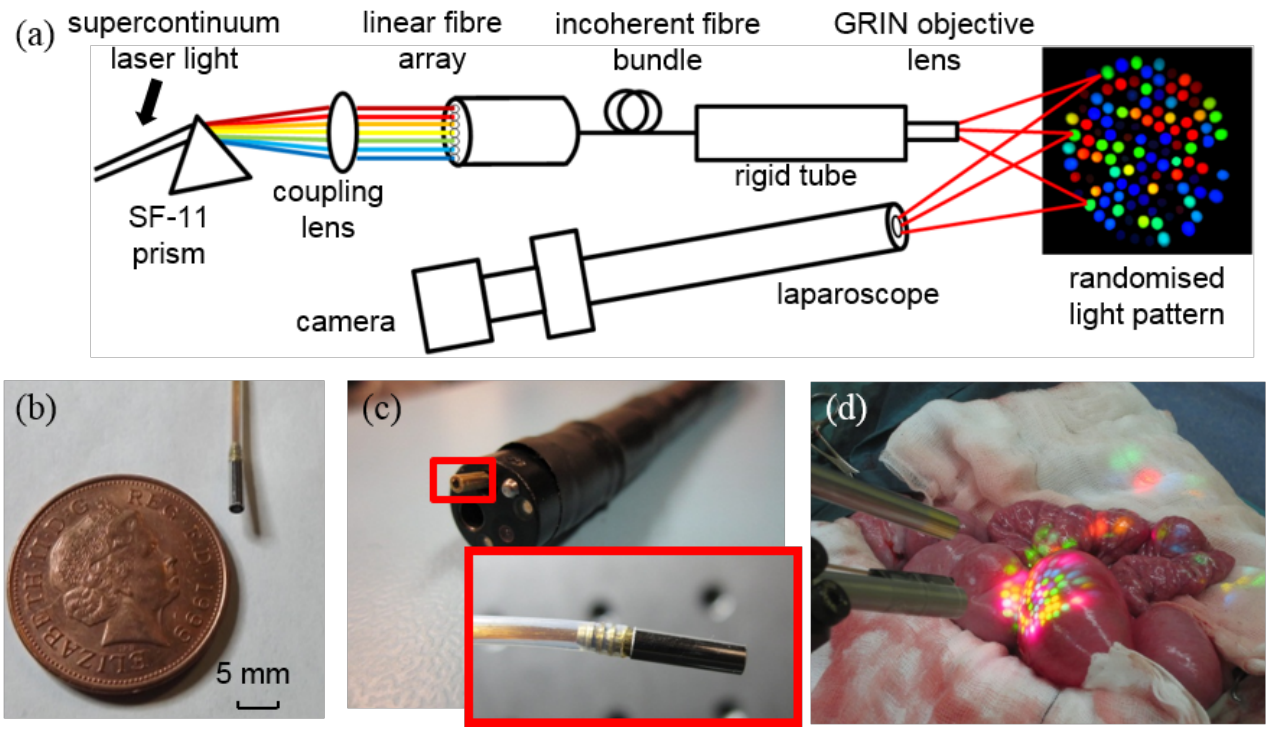


Figure 3.1: (a) Schematic of the ICL SL system (from Lin et al., copyright permission to be resolved). (b) The miniaturised probe tip. (c) The probe inserted through the biopsy channel of a flexible endoscope. (d) *In vivo* pig bowel surface under SL. Figure source: [165]

3.3 The Naive 3D Reconstruction Pipeline

An SL-based object surface reconstruction pipeline can be divided into three steps: system calibration, pattern decoding, and triangulation. The detailed implementation is illustrated in the following subsections.

3.3.1 SL System Calibration

Calibration is an essential step in computer vision approaches to extract 3D information from 2D images. The camera can be modelled as a pinhole model which assumes all the light rays go through the camera centre [306]. Compared with passive stereoscopy, a typical SL system replaces one camera with a projector — the probe in the ICL SL system. Similarly, during calibration, the probe is treated as an inverse pinhole camera (Fig. 3.2(a)). The calibration of a SL system is divided into two steps: camera calibration, and probe calibration including extrinsic parameter calibration.

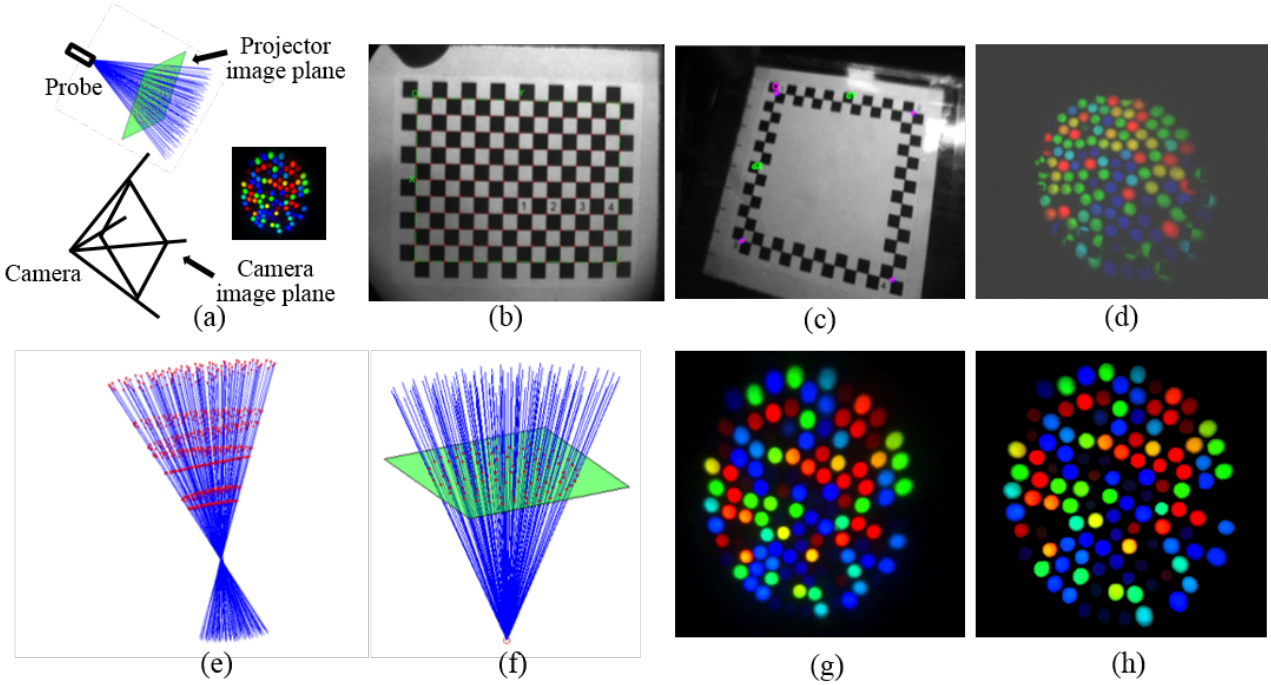


Figure 3.2: (a). A schematic of our probe-based SL system showing a similar triangulation geometry (b) "Full" checkerboard used for camera calibration (c) "Hollow" checkerboard used for projector calibration under white light. (d) Light pattern projected on the "hollow" checkerboard in SL mode. (e) Probe ray direction calibration result from method 1: locations of spots for different plane positions (red) and rays estimated by line fitting (blue). (f) Probe ray direction calibration result from method 2: rays (blue line), projector image plane (green), coordinates of features (red dots) on the image plane, the probe centre (red circle). (g) Light pattern projected on a white plane in SL mode. (h) Synthetic reference SL image that simulates the probe image plane image, with manually defined intrinsic parameter and without any distortion. Figure source: [165, 168]

Camera Calibration

The camera calibration is applied with the aid of the camera calibration toolbox developed by Bouguet [30], where the camera is considered as a pinhole model. Here we start with defining notations used in camera calibration:

Assume a point P with a 3D coordinate $[X;Y;Z]$ in the camera coordinate system, in which the origin is located at the camera centre, and the Z axis aligns on the camera principal axis. Let P_n be the normalised coordinates which are projected from P, then

$$P_n = [x_n; y_n] = \left[\frac{X}{Z}; \frac{Y}{Z}\right] \tag{3.1}$$

The normalised coordinates including lens distortion are defined as:

$$P_d = [x_d; y_d] = 3 \begin{bmatrix} (1 + k_{r(1)}r^2 + k_{r(2)}r^2 + k_{r(3)}r^4)x_n + 2k_{t1}x_n y_n + k_{t(2)}(r^2 + 2x_n^2) \\ (1 + k_{r(1)}r^2 + k_{r(2)}r^2 + k_{r(3)}r^4)y_n + 2k_{t2}x_n y_n + k_{t(1)}(r^2 + 2y_n^2) \end{bmatrix} \tag{3.2}$$

where k_r and k_t stand for radial distortion and tangential distortion parameters, respectively, and $r^2 = x_n^2 + y_n^2$. A camera′s calibration matrix consisting of intrinsic parameters can be defined as:

$$K = \begin{bmatrix} f_x & 0 & x_p \\ 0 & f_y & y_p \\ 0 & 0 & 1 \end{bmatrix} \tag{3.3}$$

where f_x and f_y are the focal lengths in x and y directions, $[x_p, y_p]$ is the principal point, which is the intersection of the principal axis and the image plane. The final image coordinates can be defined as:

$$\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = K \begin{bmatrix} x_d \\ y_d \\ 1 \end{bmatrix} \tag{3.4}$$

where $[x_i; y_i]$ is the image coordinates of point P.

Based on the previous work, the camera intrinsic parameters and distortion parameters K, k_r and k_t can be estimated according to homography transform by observing different orientations of a planar checkerboard [306, 33, 115]. The dimensions of the checkerboard are determined according to detailed applications, while in this project, a 12×14 grid array consisting of grids with sizes of 5×5 mm^2 is adopted (Fig. 3.2(b)).

Projector Calibration

The homography transform is also used in projector calibration. Since a sparse light pattern is used in this project, and every spot on the SL image shows the intersection between the object surface and one ray from the projector, to calibrate the projector means to find the directions of all the rays from the probe. So there are in total 171 rays to be estimated. In this step a "hollow pattern" checkerboard was used instead of the "full pattern" one in camera calibration (Fig.3.2(c)), with the same grid size. For projector calibration this checkerboard was manually positioned at different distances from the projector. At each position, both the white light (WL)

and SL images were taken. According to the homography transformation based on the rectangle formed by four of corners on the checkerboard, the rotation matrix R_i and the coordinate of one point on the plane M_i could be estimated in the camera coordinate system. Assuming the plane is defined as $Ax + By + Cz + D = 0$, the normal to the plane could be found as $n_i = R_i(:, 3)$, where $A = n_i(1)$; $B = n_i(2)$; $C = n_i(3)$; $D = -M_i n_i$, and $\sqrt{A^2 + B^2 + C^2} = \text{norm}(R_i(:, 3))$. Thus, the distance from the camera centre to the checkerboard could be derived as:

$$d = \frac{|Ax_c + By_c + Cz_c + D|}{\sqrt{A^2 + B^2 + C^2}} = |D| = M_i R_i(:, 3) \tag{3.5}$$

For a certain spot S_j, which stands for a ray that has been detected from the image captured by the camera, its ray direction pointing from the camera centre $[x_c; y_c; z_c] = [0; 0; 0]$ to its projection on the checkerboard P_{ij} can be calculated and represented as a unit vector $r_{dij} = [x_{uij}; y_{uij}; z_{uij}]$:

$$P_{ij} = dr_{dij} \tag{3.6}$$

Assuming that the checkerboard has been located in multiple positions, and that we know that there are 171 spots/rays in total, P_{ij} can be computed with all $i \in (1, \cdots, N)$ and $j \in (1, \cdots, 171)$. As a result, coordinates along its direction have been acquired for all the spots/rays S_j, $j \in (1, \cdots, 171)$. Then a least squares based line fitting algorithm was applied, calculating all the ray directions emitted from the probe (Fig. 3.2(e)). It is noticeable that these fitted rays contain the relative position information of every fibre and the camera coordinate system. So for the naive 3D reconstruction pipeline, these calibrated projector rays could be used to reconstruct the object surface shape by intersecting them with those captured by the camera. This probe calibration method is referred to as method 1.

However, if a more advanced reconstruction pipeline is adopted, e.g. on-the-fly self-calibration (section 4.3), these estimated rays should be forced to go through a virtual "probe centre", assuming the projector to be an inverse pinhole model without distortion. A synthetic probe image plane image (Fig. 3.2(h)) can be used to enable more mature triangulation solutions for other reconstruction problems, by simulating stereoscopy.

In this case, using the S_j estimated above as initializations, an optimisation technique is applied to force all the rays to go through one point which indicates the projector centre

$PC = [PC_x; PC_y; PC_z]$ Fig. 3.2(f). Assuming the normalised gradients R_j of S_j are defined by $R_j = [a_j; b_j; c_j]$, where $j \in (1, \cdots, 171)$, the distance between the spot location P_{ij} and the corresponding ray is defined as:

$$d_{ij} = \frac{|(P_{ij} - PC) \times (P_{ij} - (PC + R_j))|}{|R_j|} \tag{3.7}$$

The cost function E_j is therefore defined as:

$$E_j = \sum_i d_{ij}^2 \tag{3.8}$$

By minimising Equation 3.8 the final directions R_j can be estimated for all the rays, as well as the probe centre PC.

Assuming a virtual projector image plane with normal set to the mean direction of the rays, and located 1 mm in front of the projector centre, by intersecting the rays and this virtual image plane, the relative position between the projector and the camera can be determined by estimating the rotation matrix R and translation vector t. Consequently, feature coordinates on the projector image plane were also acquired. By taking another SL image projected onto a white plane and calculating the homography between the feature coordinates on the projector image plane and those from the white plane, all the pixels on the white plane SL image can be transformed into the projector image plane image, with a manually defined projector intrinsic matrix K_p. The generated reference SL image contains no distortion, since it is simulated. So during triangulation, similar to the passive stereoscopy system, our SL system provided two images from the projector and the camera: one has a fixed pattern, while the other one contains spot patterns changing from frame to frame. This probe calibration approach is called method 2. Compared with method 1, the reconstruction accuracy is sacrificed to a small extent, but more flexible choices of triangulation methods are enabled in the upgraded 3D reconstruction pipeline (section 4.3).

3.3.2 Pattern Decoding

Pattern decoding refers to a procedure that identifies and "understands" the projected light pattern. Similar to passive stereoscopy, this can be done by finding correspondences between

the captured SL image to the reference SL image. The reference image can be acquired by imaging a flat white plane using the SL system before the experiment. Generally speaking, the colour and the spatial distribution of spots are used for pattern decoding.

Pattern decoding can be divided into three steps: pre-processing, spot detection, and spot identification.

Colour Representation

Colour is the most important feature for spot identification using the ICL SL system, since each fibre contains different wavelength information. The images collected contain multiple colour channels. The number of channels depends on the camera used and three components are contained for each pixel from images captured by RGB cameras.

In this project, "colour vectors" were utilised to represent pixel colours, which were defined as unit vectors containing normalised pixel values in red, green, and blue channels. The colour difference between two pixels i and j can be quantified by the angle θ between their corresponding colour vectors C_i and C_j, ranging from 0° to 90°:

$$\theta_{ij} = \arccos \frac{c_i c_j}{\|c_i\| \|c_j\|} \tag{3.9}$$

This approach of using only one indicator to describe the colour difference brings convenience to the following light pattern decoding procedure.

Pre-processing

Due to strong light-tissue interactions, the captured SL images suffer from the appearance of specular reflective highlights, and large shape deformations in different situations. A simple pre-processing algorithm was adopted to remove specular reflection highlights and to smooth the whole image. A threshold T ($T = 210$ for 8 bit gray level images converted from RGB images) was set to segment the highlighted area, and the OpenCV inpaint function [128] was adopted to "recover" the selected area from the boundary pixel values. This was then followed

by Gaussian (standard deviation = 0.3) and a median filtering (kernel size = 5×5) to smooth the SL images.

Feature Detection Using Normalized Cut

In stereoscopy, feature detection and matching are essential steps. Searching for spot centroids in SL images can be seen as feature detection [270]. However, feature detection in typical SL images is a non-trivial problem. It can be found from a typical SL image (Fig. 3.3(a)) of a tissue surface that: 1) the spots generally have similar sizes; 2) the spot centres should be located at regional maxima of the intensity map, assuming the included pixel values are assumed to follow a Gaussian distribution; 3) in some cases the spots are closely distributed, overlapping each other; 4) some spots might suffer from large deformation; 5) spots with shorter wavelengths are generally darker, and their colour could be "contaminated" by the nearby brighter spots due to light diffusion; 6) there are even some "missing" spots due to breaks in corresponding fibres. For instance, some of the blank areas of the SL tissue image in Fig. 3.3(a) is caused by either darker blue or "missing" spots. These observations provide hints for designing robust spot segmentation algorithms which enable promising results in all the environments.

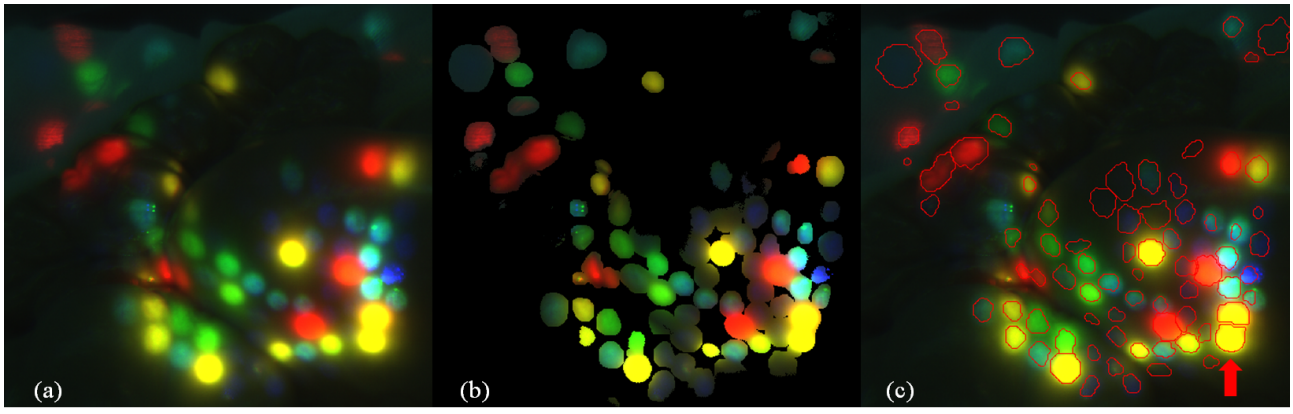


Figure 3.3: (a) Original SL image; (b) the foreground (light pattern) of the SL image; (c) spots delineated by boundaries (red). Two partially overlapping spots with similar colours can be separated (see the red arrow).

In this work, an algorithm based on normalized cut is proposed to detect the spots. Shi and Malik developed the normalized cut algorithm for image segmentation, which takes both the global and local features into consideration [253]. This algorithm treats image segmentation as a graph partitioning problem, by setting a criterion according to inter-group difference and intra-group similarity. By considering one image as one graph, one pixel is assumed to be one

node, while the differences between them are formulated into edges. A graph can be represented as $G = (V, E)$, where V stands for all the nodes, E stands for all edges. Partitioning a graph into two disjoint sets S_1 and S_2, so that $S_1 \cup S_2 = V$, $S_1 \cap S_2 = \varnothing$, is equivalent to find and remove the edges between these two sets, which is called a "cut" in graph theory. A cut can be formulated as:

$$\mathrm{cut}(S_1, S_2) = \sum_{i \in S_1, j \in S_2} w(i, j) \tag{3.10}$$

Unlike other graph cut algorithms, normalized cut partitions a graph by finding the trade-off between weights of edges and the total number of nodes, ending up with partitions in reasonably similar sizes, neither too large nor too small [253]:

$$\mathrm{Ncut}(S_1, S_2) = \frac{\mathrm{cut}(S_1, S_2)}{\mathrm{assoc}(S_1, V)} + \frac{\mathrm{cut}(S_1, S_2)}{\mathrm{assoc}(S_2, V)} \tag{3.11}$$

where $\mathrm{assoc}(S_1, V) = \sum_{i \in S_1, t \in V} w(i, t)$ is the number of connections from nodes in S_1 to all other nodes in the whole graph and $\mathrm{assoc}(S2, V)$ is defined similarly. Shi and Malik formulated it into a spectral clustering problem.

Spectral clustering makes use of the similarity matrix of a graph and standard linear algebra to cluster data points according to eigenvectors of matrices, and is simple to implement, leading to better performance compared to some traditional clustering algorithm like k-means [199, 177]. To deal with the SL images, firstly a mask was used to make sure that the normalized cut was only applied in the foreground which contains spots. This mask was generated based on regional maxima detection and morphological transform, in order to cover as many spots (foreground) as possible (Fig. 3.3 (b)). Then spectral clustering was applied to the pixels inside this mask to cluster them into different spots. Based on the previous works on spectral clustering [253, 199, 303], the steps are listed in the following:

1. Use all the pixels inside the mask to build a weighted graph $G = (V, E)$ by establishing a similarity matrix W. One entry of the matrix $w_{i,j}$, represents the similarity between two nodes (pixels) i and j, and is defined as follows:

$$W_{i,j} = e^{\frac{-\theta_{ij}^2}{\epsilon_1^2}} * \begin{cases} e^{\frac{-\|x_i - x_j\|^2}{\epsilon_2^2}} & \text{if } \|x_i - x_j\| < r \\ 0 & \text{otherwise} \end{cases} \tag{3.12}$$

where x_i and x_j stand for the image coordinates of the pixels, r the threshold of Euclidian distance between two pixels, ϵ_1 and ϵ_2 constants.

2. Calculate K eigenvectors corresponding to the K largest eigenvalues of:

$$D^{-\frac{1}{2}}(D-W)D^{-\frac{1}{2}}x = \lambda x \tag{3.13}$$

where D is a diagonal matrix, whose (i,j) entry is the sum of Ws ith row, K is the number of clusters. There exists a method to select K value automatically according to the "eigen gap" [177]. However, this method is sensitive to the choice of ϵ_1 and ϵ_2, and could be unreliable when clustering noisy data. Therefore, in the experiment the K value could be set empirically related to the number of detected local regional maxima.

3. Stack the K eigenvectors $v_1, v_2, \cdots, v_k$ in columns to form a matrix A, and normalise it in row direction.

4. Use k-means to cluster all the rows in A.

5. Assign labels to all the rows, so all the pixels are labelled.

In this way, the spots were clustered and segmented in the SL image (Fig. 3.3(c)). Fig. 3.3(c) also shows that two adjacent and partially overlapping spots with similar colours could be separated using this algorithm. The gravities of the segmented spots were treated as the spot centres and used for the following matching and triangulation steps.

Feature Matching by Spot Identification

Spot identification for this SL system refers to the feature matching procedure. The aim is to find the correspondences between the the collected SL image and the reference SL image. In this work, the reference image was generated using the method described in section 3.3.1 by probe calibration and a SL image projected on a white plane on which all the spots are manually identified beforehand. By comparing it with the captured SL image, unique labels should be assigned to individual detected spots. Colour was the main information that could be used in order to discriminate different spots. Due to the hardware limitation, identifying those spots only according to the colour information was difficult. Furthermore, the projected light

patterns suffered from strong light-tissue interactions, affecting colour vector changes. Thus, an algorithm based on colour and neighbourhood information was developed:

1. Find neighbours for all the spots in both the reference and SL images. Delaunay triangulation is applied to build the connections between adjacent spots followed by a connection pruning process to keep only the "real" neighbourhood connections according to the shapes, sizes and orientations of those spots (Fig. 3.5(b)).

2. Calculate the average colour vector, and establish a stack of neighbours' colour vectors in a clockwise sequence for every spot.

3. Compare the colour vectors with the neighbours to determine unique matches between the SL image and the reference image.

4. Propagate the spot matching procedure is propagated from the initially identified spots to their neighbours based on to colour information and local rigid registration. (Fig. 3.4 and Fig. 3.5(c, d)). Thus, the spots are identified and correspond to the rays emitted from the probe.

5. Remove incorrect spot-ray correspondences according to the calibration.

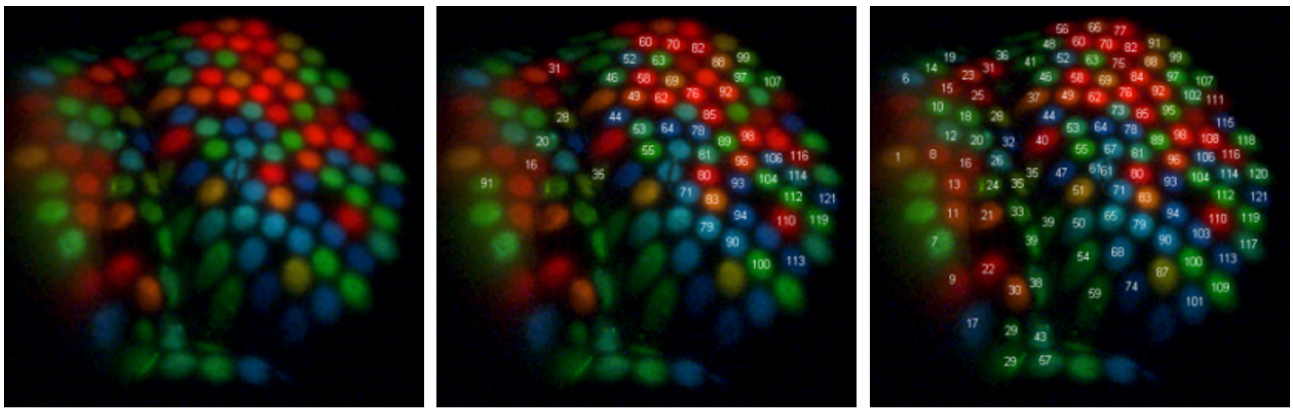


Figure 3.4: (a) Captured SL image; (b) captured SL image with initially identified spots; (c) the identification procedure is propagated to more spots by local rigid registration based method.

In this way, the spots in the captured SL images are matched to those from the reference image. This matching result can be used for the triangulation (section 3.3.3) for object surface 3D reconstruction.

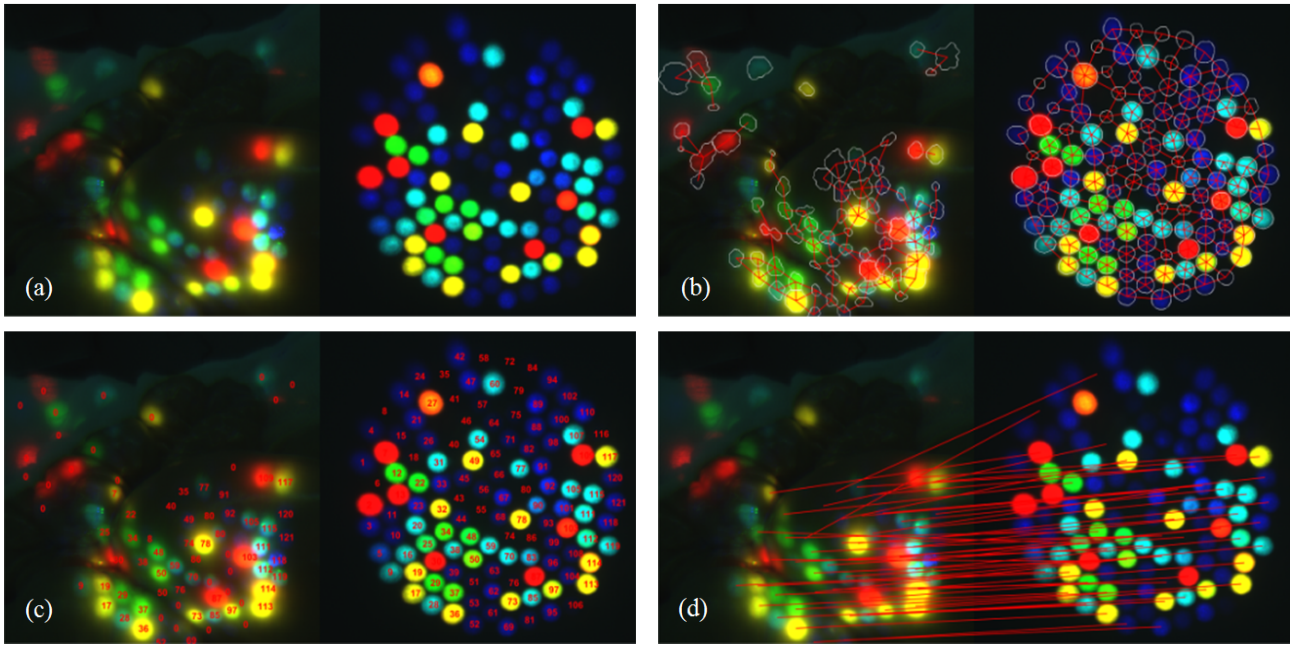


Figure 3.5: (a) The SL image (left) and the reference image (right). (b) Neighbour detection for the SL image (left) and reference image (right): segmented spot boundaries (white) and connections between neighbours (red). (c) Labelled SL image (left) according to the labels assigned manually in the reference image (right). (d) Correspondences of spots between the SL image (left) and the reference image (right).

3.3.3 Triangulation

Following feature matching, a generic triangulation procedure is adopted to compute the 3D locations of the matched projected spots. Compared with classic triangulation methods which triangulate rays from two cameras in the passive stereoscopy problem [109], in this naive reconstruction pipeline the geometric intersections between rays from the camera and the probe are used as the triangulation results. As is discussed in section 3.3.1, this triangulation method leads to higher accuracy, since simply treating the probe as an inverse pinhole model by forcing all the rays to go through a virtual "probe centre", introduces reconstruction error.

According to Section 3.3.1, the direction of rays captured by the camera can be calculated with the camera calibration results. Since the camera is considered as a pinhole model, all the rays go the camera centre, which is $[0; 0; 0]$ in the camera coordinate system. Thus, the rays

$S_j, j \in (1, \cdots, 171)$ can be represented as:

$$\begin{cases} x = a'_j t \\ y = b'_j t \\ z = t \end{cases} \tag{3.14}$$

$[a'_j, b'_j, 1]$ indicates the direction for rays captured by the camera. While the rays emitted from the probe are:

$$\begin{cases} x = x_{0j} + a_j t \\ y = y_{0j} + b_j t \\ z = z_{0j} + t \end{cases} \tag{3.15}$$

where $[x_{0j}, y_{0j}, z_{0j}]$ stands for the offsets, and $[a_j, b_j, 1]$ the direction, regarding S_j.

In real-world applications, the two corresponding rays emitted from the probe and captured by the camera normally do not intersect, so the middle point of their common normal is calculated and considered as the intersection.

By intersecting all the identified rays the 3D locations of the projected spots on the object surface can be estimated. For the ICL SL system a maximum of 171 3D locations can be calculated from the projected spots, depending on the number of spots detected. The reason for using this naive triangulation method instead of the classic approach is that, our probe can only be considered as an inverse pinhole model camera by approximation, as the rays emitted do not go through one exact point. However, the classic way, which assumes the probe as a pinhole model, can be used in the case of the further improvement of the software platform, as illustrated in section 4.3.

3.4 Experimental Results

For validation experiments on different objects have been carried out. SL images are acquired on: silicone heart phantom, *ex vivo* porcine liver, kidney, and heart, and *in vivo* porcine large bowel. In order to evaluate the robustness of the software platform in challenging surgical scenes, SL images on liver with blood are also captured during the *ex vivo* experiments.

This section is divided into two parts: evaluation of pattern decoding algorithm with normalized cut and spot matching; 3D reconstruction accuracy affected by pattern decoding. During the experiments, rather than inserting the probe through the biopsy channel, the probe was mounted on a rigid support. The angle between the probe and each camera was about 15°, the baseline about 5 cm, and the optical axis of both endoscopes roughly perpendicular to the object surface. Images were acquired at 30 FPS, with a maximum 33 ms exposure time per frame. Practically this fame rate is high enough to avoid image blurring caused by respiration movements. It takes about 50 s to process a single frame using MATLAB (Intel i7-3770, 8 GB RAM), in the naive reconstruction pipeline.

3.4.1 Light Pattern Decoding

Given an accurate calibration, the light pattern decoding results determine the reconstruction accuracy for a typical SL system. As light pattern decoding is divided into spot detection and identification with ICL SL system, images acquired from phantom, *ex vivo*, and *in vivo* experiments using an RGB camera, have been used to evaluate the performance and robustness of these two steps. The indicators used for the evaluation include the sensitivities and specificities of detection and identification results. They are calculated by comparing with the ground truth data (manual detection and identification results). Given the automatically detected number T_{da}, the manually detected number T_{dm}, and the correctly automatically detected number T_{dc}, the sensitivity of spot detection is defined as $\mathrm{Sen_d} = \frac{T_{dc}}{T_{dm}}$, while the precision $\mathrm{Pre_d} = \frac{T_{dc}}{T_{da}}$. Given the automatically identified number T_{ia}, the manually identified number T_{im}, and the correctly automatically identified number T_{ic}, the sensitivity of spot identification is defined as $\mathrm{Sen_i} = \frac{T_{ic}}{T_{im}}$, while the precision $\mathrm{Pre_i} = \frac{T_{ic}}{T_{ia}}$.

In this section the pattern decoding results using RGB camera are demonstrated. The average and standard deviation of these indicators in different experiments are shown in Fig. 3.6. The proposed pattern decoding algorithm aims to achieve identification precisions as high as possible while maintaining identification sensitivities. As the wrongly detected spot will be excluded in the spot identification step, empirically on the experimental samples in general 70% detection precision can guarantee a good performance for surface reconstruction.

For all the images, the average sensitivities are always higher than 60% (Fig. 3.6(a)), while

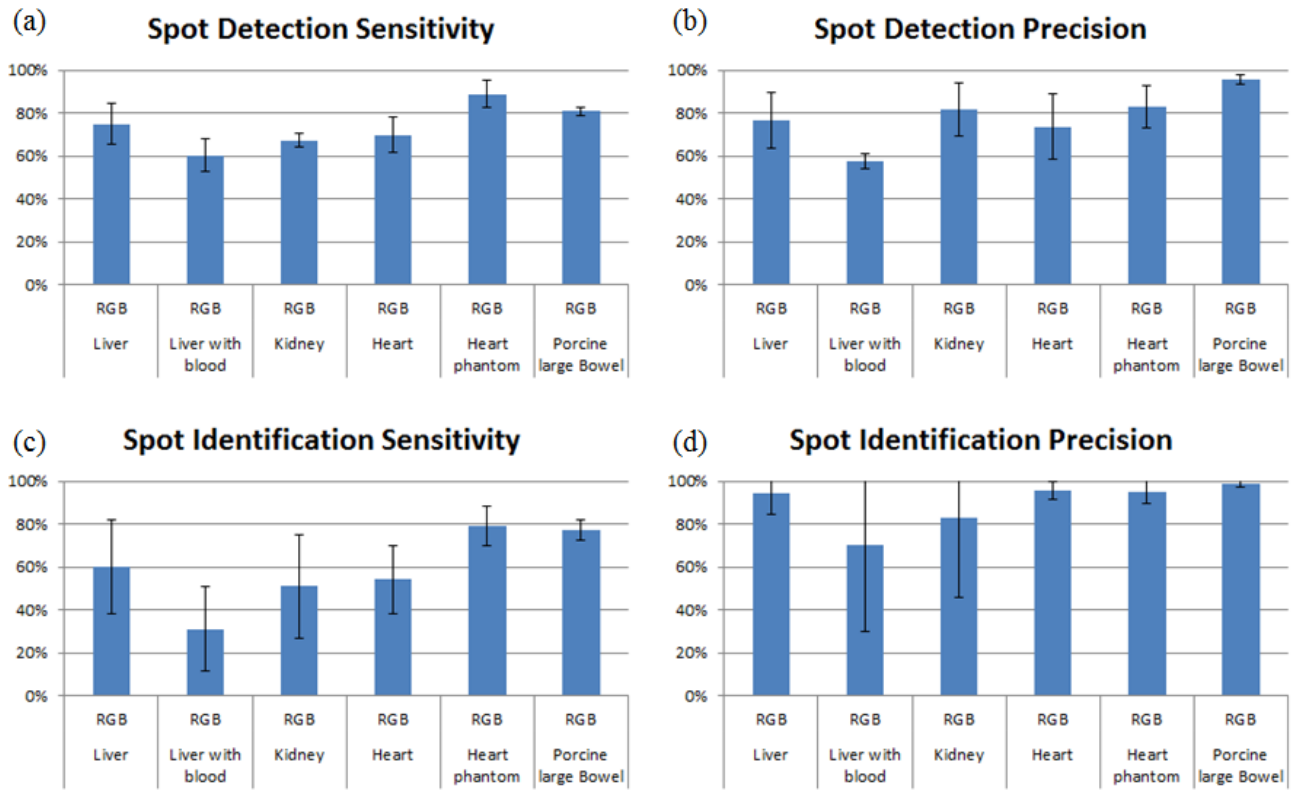


Figure 3.6: Light pattern decoding results using an RGB camera, on *ex vivo* porcine kidney, heart, liver, and liver with blood, silicone heart phantom, and *in vivo* porcine large bowel. The average value (blue bar) and standard deviation (black error segment) are provided. (a). Spot detection sensitivity. (b). Spot detection precision. (c). Spot identification sensitivity. (d). Spot identification precision.

the average precisions are around 80% except for the case of liver with blood (Fig. 3.6(b)). Inevitably, the identification sensitivities are slightly lower compared with the detection results, but the identification precision has been improved for all cases. This is because the proposed spot identification method eliminates the wrongly detected spots, leaving only those correct ones. The average identification precisions for all the SL images are higher than 95%. It is noticeable that very little error will be introduced even if the identification precision is low. In the case of low identification precision, that corresponding SL image will be abandoned.

Taking all the SL images collected by the RGB camera into consideration, our algorithm leads to identification sensitivities higher than 40% as well as precision higher than 90% in 63 out of 75 SL images. Among those images, in general our algorithm performs better on the silicone heart phantom and most of the *ex vivo* tissues. However, as expected, both the detection and identification sensitivity and specificity decreased and became unstable in the appearance of blood on the tissue surface, as the SL image quality is generally much worse than that in other

cases.

3.4.2 Surface Reconstruction Accuracy in the Naive Reconstruction Pipeline

Given the difficulty in acquiring the ground truth of tissue surface due to its softness/deformability, only the reconstruction result of the heart phantom experiment has been used to measure the reconstruction accuracy. This experiment aimed at evaluating the surface reconstruction accuracy using the naive reconstruction pipeline, so the relative position between the projector and camera was calibrated before the experiment. Eight SL images of the heart phantom from different angles on different local areas are captured, so eight reconstructed surfaces can be acquired. By registering the ground truth mesh acquired by a laser scanner and the reconstructed surface using rigid iterative closest point (ICP) method [25], the reconstruction error can be indicated by the distance between these two registered surfaces Fig. 3.7. During ICP, the ground truth was manually cropped to keep only the corresponding local area for registration, in order to reduce the error introduced by the ICP based registration itself. The average, the mean, and the maximum distance are used as the indicators. Furthermore, in order to find the influence of the pattern decoding algorithm, the result of using the automatic decoding is compared to that of manual pattern decoding.

Table 3.1 shows the reconstruction result using manual light pattern decoding. The statistics demonstrates that the accuracy of surface reconstruction using SL system under strict calibration and perfect pattern decoding is very high, with average error smaller than 0.7 mm and maximum errors under 3 mm from seven out of eight SL images. It is also demonstrated that the reconstruction accuracy drops if the automatic pattern decoding algorithm is adopted. However, if one of the cases, whose identification precision is lower than 90%, is excluded, the mean values of the average distance and maximum distance reach to 0.7 mm and 2.98 mm, respectively, making a very small difference from the results using manual pattern decoding. Given that satisfying identification results (sensitivity $> 40\%$ and precision $> 90\%$) can be achieved in about 84% of all our SL images, we show the robust and promising performances of tissue surface reconstruction using the proposed naive 3D reconstruction pipeline.

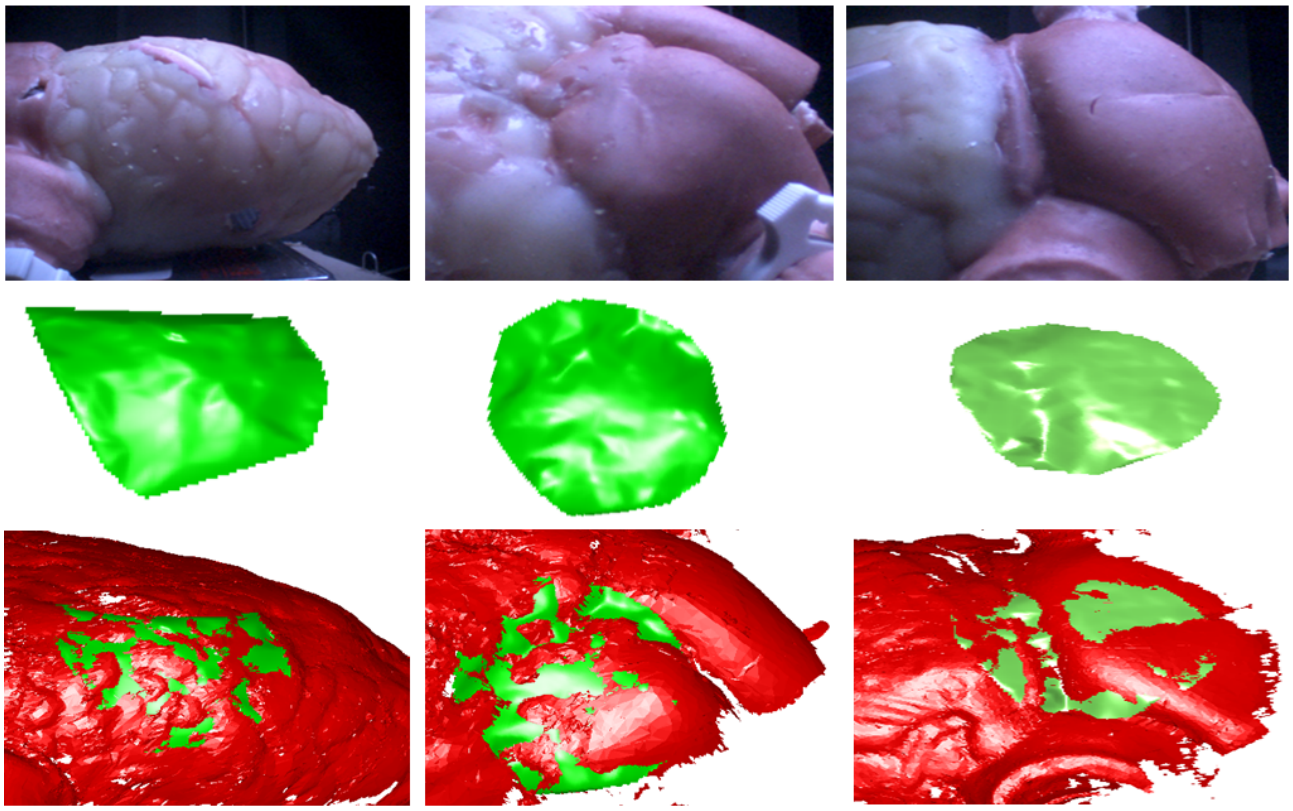


Figure 3.7: ICP based rigid registration to evaluate the reconstruction error. First row: silicone heart phantom under white light; second row: the reconstructed surface; third row: registration of the reconstructed surface (green) with the ground truth mesh (red). Figure reproduced from [165]

3.5 Discussion and Conclusions

In this chapter, the ICL SL system is introduced, including both the hardware and software aspects: the ICL SL optical hardware system, previously developed by Clancy et at. [51] is firstly demonstrated, followed by a description of a naive 3D reconstruction pipeline. Finally validation on phantom, *ex vivo*, and *in vivo* experiments is provided.

Due to its miniaturised size, the probe of the ICL system could be conveniently integrated into an endoscopic system. In the experiments the probe was always mounted on another support detached from the endoscope. However, there are still other options to integrate the probe into an endoscopic system: e.g. the probe can be inserted through the biopsy channel of a flexible endoscope; the probe can also be attached beside the endoscope using a specially designed 3D-printed adaptor (section 5.2.2).

As implied from the name, the naive reconstruction pipeline describes an initial and "clean"

Table 3.1: Heart phantom surface reconstruction accuracy using manual and automatic pattern decoding, respectively. The mean, average, and maximum value of mean, standard deviation, and maximum errors of all the reconstructed surfaces are listed. Table reproduced from [165].

	Identification sensitivity	Identification precision	Mean error (mm)	Std error (mm)	Max error (mm)
(Manual) Median	100.00%	100.00%	0.65	0.32	2.52
(Manual) Mean	100.00%	100.00%	0.64	0.35	2.78
(Manual) Max	100.00%	100.00%	0.72	0.44	5.04
(Automatic) Median	70.51%	98.33%	0.71	0.40	3.08
(Automatic) Mean	68.74%	96.70%	0.88	0.67	5.58
(Automatic) Max	90.76%	100.00%	2.17	2.65	23.76

version of the software platform, which refers to the simplest, most straightforward, and basic algorithmic implementation. The following improvements (Chapter 4 and Chapter 5) have been proposed to extend the functionality as well as to accelerate the computation based on this pipeline.

In this pipeline, two specially designed projector calibration methods were proposed to cope with different situations. The main difference between the two methods is: it is assumed in method 2 that all the rays emitted from the probe go through a virtual "probe centre". Method 1 applies independent line fitting to find the direction of the rays; while method 2 finds the probe centre together with the ray directions by optimising an objective function that takes into consideration all the ray directions, and ends up with a virtual projector image plane image with manually defined projector intrinsic parameters. If the relative position between the probe and the endoscope is fixed, the projector calibration results from method 1 is used since it leads to higher reconstruction accuracy; Otherwise in the case of algorithmic extensions like on-the-fly self-calibration (section 4.3) and location normal integration (section 4.4), method 2 is adopted since it facilitates mathematical computation in those applications.

The key factor in robust surface reconstruction, the light pattern decoding algorithm described in the naive pipeline, has been validated by phantom, *ex vivo*, and *in vivo* experiments. Based on normalized cut and local rigid registration, the proposed algorithm led to promising performance in most datasets except for the liver with blood. The properties of normalized cut ensures that it always leads to separated clusters with similar sizes and benefit spot detection, especially in the case that the adjacent spots contain a similar overlapping spectrum. The local rigid registration based spot identification algorithm also showed its ability to remove wrongly detected spots and propagate the matching procedure to the whole image. However, there are still two main drawbacks of this algorithm: 1) normalized cut is an unsupervised method which requires manual interaction (e.g. number of spots which can only be achieved approximately by pre-processing); 2) it is generally slow since it requires eigenvalues and eigenvectors computation from a sparsity matrix, which could be to large to enable real-time computation in most cases. Therefore, in order to address these challenges a deep learning based method is proposed for pattern decoding, which can be found in section 5.3.

To sum up, in this chapter, a brief introduction is given to the ICL SL system, and the naive 3D reconstruction pipeline. The overall and step-wise performance of the hardware and software platforms have been evaluated by phantom, *ex vivo*, and *in vivo* experiments, demonstrating a potential of using the system in clinical applications. The extension of the system functionality can be found in Chapter 4. The main challenge of this system, i.e., the non-real-time performance and the requirements of parameter tuning in the light pattern decoding method, will be address in Chapter 5 by using deep learning based models and integrating monocular Structure-from-Motion (SfM) algorithms.

Chapter 4

Improvements of the ICL System and the Naive Reconstruction Pipeline

4.1 Introduction

In this chapter further extensions on the ICL SL system, including the hardware and software, were illustrated. Based on the original ICL SL optical system [51], and the naive 3D reconstruction pipeline (section 3.3), improvements have been proposed in three aspects: 1) replacing the RGB camera by a multispectral camera to facilitate the pattern decoding; 2) on-the-fly self-calibration to handle changing relative position of the endoscope and the probe; 3) Integration of local surface orientation to refine the 3D reconstruction. Validation on phantom, *ex vivo*, and *in vivo* experiments have been provided.

The content of this chapter has been published in:

1. J. Lin, N. T. Clancy, D. Stoyanov and D. S. Elson, "Tissue Surface Reconstruction Aided by Local Normal Information Using a Self-calibrated Endoscopic Structured Light System", In: *Proc. MICCAI, Part I.*, 2015, pp. 405-412.

2. J. Lin, N. T. Clancy and D. S. Elson, "An endoscopic structured light system using multispectral detection", International Journal of Computer Assisted Radiology and Surgery, vol. 10, pp. 1941-1950, 2015.

4.2 Increasing the Colour Discrimination Using a Multispectral Camera

As an image acquisition technique, sequential multispectral imaging (MSI) captures images of certain objects at diverse wavelengths, compiling spectra for all the pixels. This technique is used in a number of medical applications. For instance, chromophores like oxy- and deoxyhaemoglobin can be detected and differentiated, and their relative concentrations used to determine measures of tissue perfusion and oxygen saturation [261, 274]; molecular targeting and spectral unmixing of multiple fluorophores can also be achieved using multispectral fluorescence imaging [299, 105]. Compared to normal RGB cameras, which only record three wavebands, a multispectral camera results in images corresponding to more wavelengths, providing higher colour discrimination ability. However, according to our knowledge, there is no work in which a multispectral camera is integrated with an SL system.

In the original ICL SL system, a standard colour CCD camera (DCU 223C, Thorlabs Ltd., UK), was used to capture the SL images, indicating three different wavebands in the blue, green and red ranges, respectively. However, due to strong light-tissue interaction (section 3.3.2), the captured SL images were generally blurry and hard to distinguish in cases of containing close wavelength information. By analysing the reflectance response of the RGB camera (Fig. 4.1), it can be found that only those overlapping areas of the different spectra have the potential to be used to distinguish those spots, since the ratio between three channels is sensitive to the wavelength change. In order to increase the overlapping spectral area, a multispectral camera (SpectroCam Multispectral Camera, Ocean Optics, USA) was used in this work, which contains 8 different bandpass filters, leading to eight channels in the output images. Fig. 4.1(d) shows the spectrum and the increase in overlapping area compared to the RGB camera, indicating a stronger ability to identify colour or different wavelengths. Besides, the multispectral camera can also be potentially used for tissue classification, which might further facilitate real-world clinical applications (see section 2.3).

In order to intuitively demonstrate the superiority of the multispectral camera to the RGB camera in colour discrimination ability, both the cameras were applied to capture SL images of a white plane. The colour vectors were recorded for all the spots, and consequently their corresponding wavelengths were roughly estimated according to the filter transmission spectra

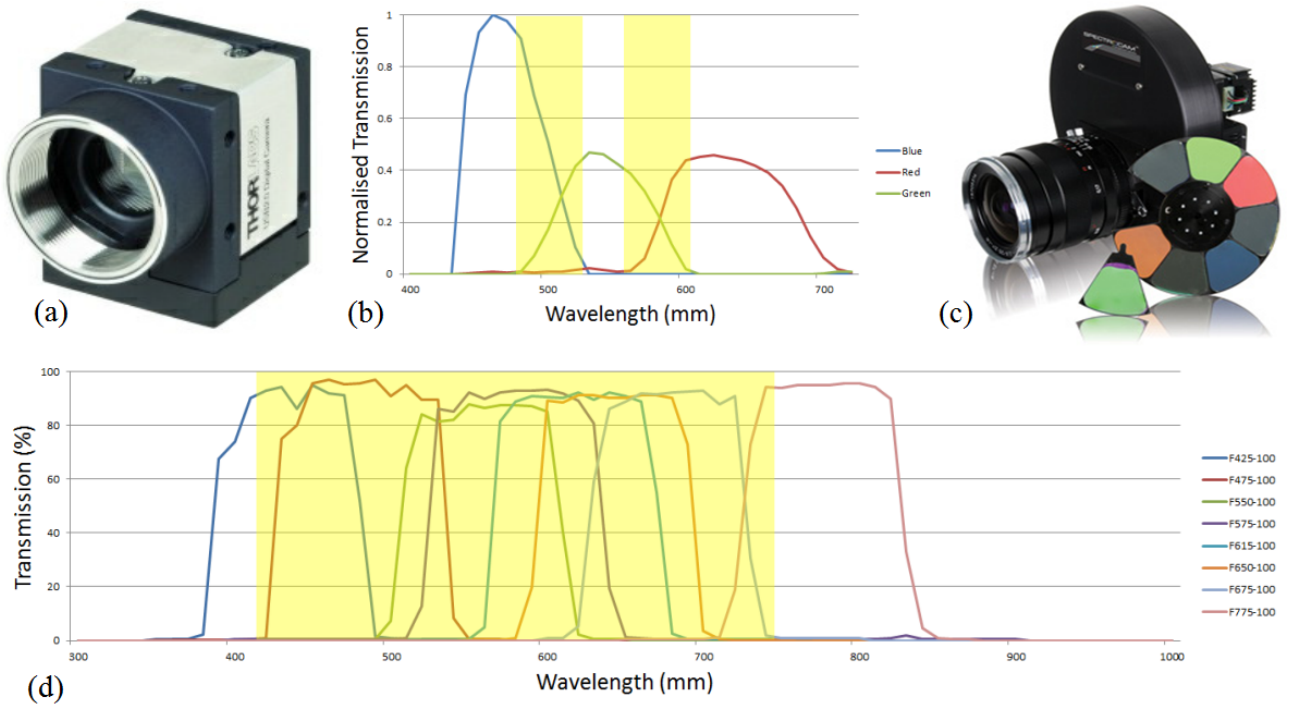


Figure 4.1: (a) RGB camera; (b) normalised filter transmission spectra for the RGB camera and the overlapping area (yellow shadow); (c) Multispectral camera; (d) Filter transmission spectra for the multispectral camera and the overlapping area (yellow shadow).

shown in Fig. 4.1. The estimated wavelengths were sorted in ascending order and plotted together with the estimation error bar and the camera filter transmission curves (Fig. 4.2). The result showcases the difference between two cameras: due to the smaller spectral overlapping area, the colour estimation for the RGB camera image suffers from larger errors, and those colours in the blue end are not estimated or distinguished correctly since there are rarely responses for green and red channels in that band. On the contrary, because of a larger overlapping spectral area, the estimation ambiguity for the multispectral camera is much lower, and the effective discriminable band is much wider. The detailed comparison between the two cameras are provided in section 4.5.1.

4.3 On-the-fly Self-calibration and Reconstruction

In the naive 3D reconstruction pipeline (section 3.3.1), the extrinsic parameters standing for the relative position between the projector and the camera, is estimated before the experiment, and stay unchanged. However, one of the main drawbacks of using the naive pipeline is that calibration is required once the relative position has been changed, which may happen during

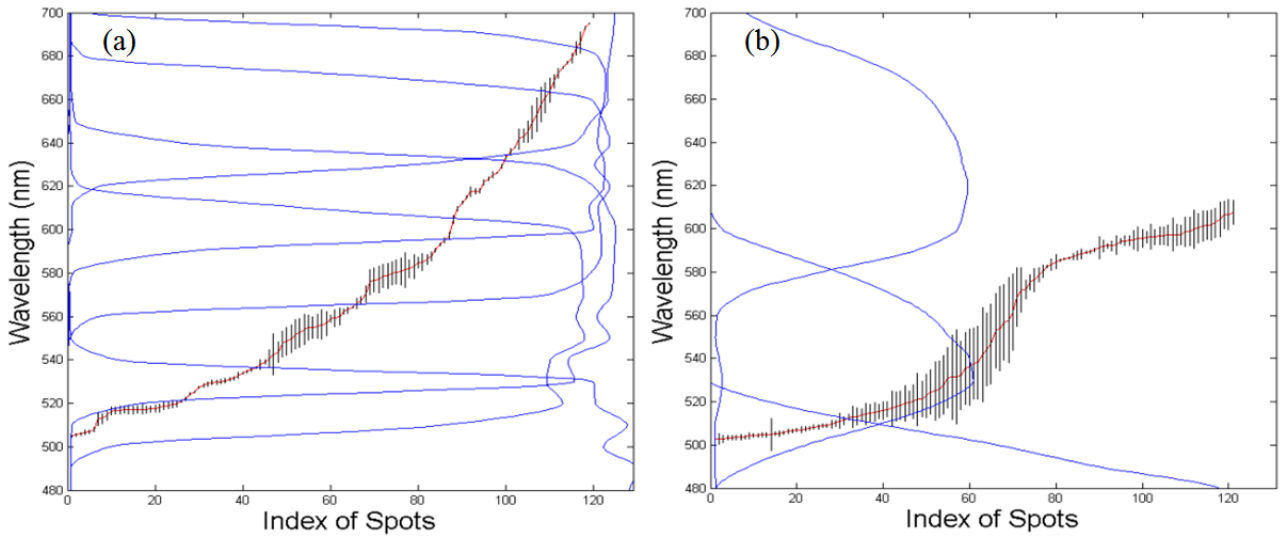


Figure 4.2: This figure shows the ambiguity of colour discrimination in different wavelengths, for both the RGB and multispectral camera. (a) Multispectral camera: estimated wavelengths (red); estimation error bar (black); the camera filter transmission spectra (blue). (b) RGB camera: estimated wavelengths (red), estimation error bar (black), and the camera filter transmission spectra (blue). Longer black lines indicate higher ambiguity at corresponding wavelengths (left x axis).

experiments, hindering the feasibility of using it in real-world applications: the probe and endoscope may inserted through trocars located on different locations of the patient's body. In this case, the relative position between the probe and the camera will not be fixed during the surgery. Thus, in order to make the SL system more practical, an on-the-fly self-calibrated method is proposed to enable a free-hand manipulation of the system during experiments.

As the projector is considered as an inverse camera without distortion, the estimation of the relative position is very similar to that in normal passive stereoscopy [109]. In two-view geometry, the information from two image planes corresponding to both the cameras is required for estimating the correspondence geometry, camera motion, and structure. While in SL, the image plane of the projector can be seen as the image plane of a second camera. According to the projector calibration procedure, the coordinates for all the spots/rays X_p on the image plane can be calculated through the intersection of the rays and a plane which is located 1 mm away from the projector Fig. 3.2(d). While on the camera image plane, the image coordinates of corresponding spots X_c after distortion correction and normalization by the camera matrix are also available according to the camera calibration result. Assume that X_p and X_c are the projections of a 3D point X, the projective matrices of the projector P_p and the camera P_c

which satisfy $X_p = P_p X$ and $X_c = P_c X$ can be defined as:

$$P_p = [R|t] \tag{4.1}$$

$$P_c = [I|0] \tag{4.2}$$

Here R stands for the rotation matrix of the projector, and $t = -RC$, where C is the coordinates of the projector centre. Therefore, R determines the orientation of the projector, while C determines the projector centre position. Then there is a fundamental matrix F indicating the correspondences between these two image planes can be defined as:

$$X_p^{\mathsf{T}} F X_c = 0 \tag{4.3}$$

Together with the constraint that $\det F = 0$, F can be derived using the Maximum Likelihood (ML) estimation which minimizes the reprojection error. For the detailed procedure please refer to Hartley and Zisserman [109]. However, in practice during the experiment, it was found that due to the error in the matching procedure in Sec. 3.3.2, not all the correspondences found can be used. Therefore, we applied RANSAC to exclude those incorrect correspondences (outliers) [78], and used only those reliable correspondences to estimate F.

The essential matrix E is a special case of F when the image coordinates are normalised. While in real-world applications, E is not exactly equal to F even after normalisation, due to errors like calibration error and feature matching error. But E can still be computed from F: Let F be a matrix with SVD given by $F = UDV^{\mathsf{T}}$, where $D = \mathrm{diag}(a, b, c)$, where $a \geq b \geq c$. Then reconstructed $D_E = \mathrm{diag}((a+b)/2, (a+b)/2, 0)$, therefore $E = UD_EV^{\mathsf{T}}$. Also, E can be defined as:

$$E = [t]_x R \tag{4.4}$$

where $[t]$ represents the corresponding skew-symmetric matrix of t. Given a known E, four possible projector projective matrices can be derived:

$$\begin{aligned} P_p = \ & [UWV^{\mathsf{T}}|U(:,3)] \\ \text{or} \quad & [UWV^{\mathsf{T}}|-U(:,3)] \\ \text{or} \quad & [UW^{\mathsf{T}}V^{\mathsf{T}}|U(:,3)] \\ \text{or} \quad & [UW^{\mathsf{T}}V^{\mathsf{T}}|-U(:,3)] \end{aligned} \tag{4.5}$$

where

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{4.6}$$

Based on the fact that all the points should always be located in front of both the image planes, only one solution is selected and subsequently P_p, R, and C can be determined. As a result, the locations of spots in the camera coordinate system can be derived. For the detailed procedure please refer to a previous book [109].

Therefore, the 3D positions of the spots can be estimated up to scale by using this method, which means that the real scale of the object surface recovered by this self-calibration technique cannot be determined unless there exists another constraint. But this pure shape information is still very useful in some clinical applications like augmented reality or polyp detection. In order to validate its reconstruction accuracy, an experiment which measured the error by using self-reconstruction was applied by comparing its results with those under known extrinsic parameters. Please see section. 4.5.2 for the detailed information.

In some cases, due to strong optical scattering or absorption, the spot identification method failed to provide a perfect feature matching result. Therefore, the following iterative procedure was applied to refine both the epipolar geometry and spot identification, during on-the-fly self-calibration procedure: in each iteration, the fundamental matrix F was estimated with RANSAC outlier rejection and F optimisation. More correspondences were then searched for using F as a constraint. A tolerance is set as the stopping criterion. Then F and correspondences were both updated.

This automatic on-the-fly self-calibration technique allowed free-hand manipulation of the SL instruments during experiment, with varying projector poses. This technique can potentially be used in surgery, providing tissue shape information up to scale. However, the stability of this depends on the accuracy of the pattern decoding results, and is further discussed in section 4.6.

4.4 Orientation Estimation According to Pattern Deformation

In two view geometry, points on one plane can be related to corresponding points on another plane by a homography uniquely induced by a plane at a general position in 3D space [109]. Based on this theory an algorithm is proposed to estimate the local normals of individual spots in 3D space, and use this extra information to enhance the 3D reconstruction results.

Given the projective matrices P_p and P_c, and a plane defined by $N^\mathsf{T}X = 0$ with $N = (n^\mathsf{T}, 1)^\mathsf{T}$ in 3D space, a point X on the plane was projected as X_p and X_c on the projector and camera planes, respectively. Then $X_p = HX_c$, where

$$H = R - tn^\mathsf{T} \tag{4.7}$$

Normally, H can be estimated according to at least four 2D to 2D point correspondences. If H is known and R and t are calibrated beforehand, the normal of the plane n can be estimated.

Based on this, we propose a method to estimate the orientation of the object surface at the reconstructed spot centres. Given that the individual spots are very small in 3D space, we assume that all the pixels inside spot S_i are located on one plane defined by $N_i^\mathsf{T}X = 0$, with $N_i = (n_i^\mathsf{T}, 1)^\mathsf{T}$, then the image coordinates of those pixels on two image planes can be linked by a homography H_i. However, since pixelwise correspondences cannot be built between the two image planes in our SL system, another method is adopted to estimate H_i. It is noticeable that the shapes of spots on the projector image plane, which represents the intersection between the cone-like rays and the image plane, are all ellipses. It can also be observed that the shapes of most spots on the camera image plane are similar to ellipses, showing the feasibility of the assumption, since a conic can be transformed into another conic by a given homography. An ellipse on a 2D plane can be represented as

$$Ax^2 + 2Bxy + Dy^2 + 2Ex + 2Fy + G = 0 \tag{4.8}$$

Equation 4.8 can be written as $X^\mathsf{T}CX = 0$, where$X = [x, y, 1]^\mathsf{T}$, and

$$C = \begin{bmatrix} A & B & C \\ B & D & F \\ E & F & G \end{bmatrix} \tag{4.9}$$

In our SL system, if a pair of corresponding conics C_p and C_c on two image planes are known, they can be related by a homography H:

$$C_c \sim H^\mathsf{T} C_p H \tag{4.10}$$

Combining Equation 4.7 and Equation 4.10,

$$C_{ci} \sim (R - tn^\mathsf{T})^\mathsf{T} C_{pi} (R - tn\mathsf{T}) \tag{4.11}$$

Since Equation 2.11 defines an overdetermined system, n^T can be estimated through optimisation as long as the conics are non-degenerate. Based on the assumption that each pair of projected spots are located on one single plane in 3D space, corresponding spots in the two image planes of the SL system can be seen as pairs of corresponding conics C_p and C_c. Therefore, we start with ellipse fitting for all the spots to acquire the conic coefficient matrices, using the method proposed by Fitzgibbon et al. [80]. Based on the fitting cost, some spots are excluded since either they are not accurately delineated or they are not located in single planes, hence do not affect the reconstruction. Because all the fitted ellipses are non-degenerate, the local normal n where the spot is located can always be estimated by optimising Eq. 4.11. The optimisation procedure is applied with the constraint that the estimated plane should always go through the 3D location of the spot centre, which is acquired in the previous self-calibration step. The initial guess of a local normal can be estimated from the normal of plane determined by the three surrounding spot locations in 3D space. After optimisation, the estimated normals are filtered using the cost of the optimised objective function.

Given the 3D locations of spot centres, the object surface is reconstructed using a thin-plate spline (TPS) [82]. A simple but effective method is used to constrain the surface orientation: for one point with estimated orientation, six new neighbouring points close to it on the estimated plane are added. All the points including both the original and the newly added ones are

used to fit the TPS model. In this way, those newly added neighbouring points constrain the local surface direction when TPS fitting takes place. The intuitive demonstration of using local normal information to improve the reconstruction accuracy could be found in section 4.5.3.

4.5 Experimental Results

In order to evaluate the feasibility of the software platform and its practicality, experiments on different objects have been carried out. SL images are acquired on: a silicone heart phantom using the RGB and multispectral camera; porcine liver, kidney, and heart samples during *ex vivo* experiments using an RGB camera; porcine large bowel samples during an *in vivo* open surgery experiment using an RGB and a multispectral camera.

This section is divided into three parts to demonstrate and analyse the experiment results: 1) comparison of multispectral and RGB cameras; 2) reconstruction accuracy by using self-calibration technique; 3) reconstruction with estimated local normal information. The multispectral camera ran at 2 FPS, while the RGB camera ran at 30 FPS.

4.5.1 Comparison of Multispectral and RGB Cameras in Light Pattern Decoding

The comparison of the multispectral and RGB cameras have been carried out on the silicone heart phantom and the porcine large bowel in the *in vivo* experiment, demonstrating that the former performs slightly better than the latter in the light pattern decoding procedure (Fig. 4.3).

In the phantom experiment, both of the cameras performed well. The RGB camera led to higher identification sensitivity, while the multispectral camera outperformed in identification precision. Although both of them worked well in the in *in vivo* experiment, the multispectral camera outperforms the RGB camera in both the two identification indicators, especially precision. This is due to the higher importance of colour information in the spot identification step compared with the detection step. As the multispectral camera has undoubtedly much higher

colour discrimination ability, it potentially leads and actually led to slightly better identification results.

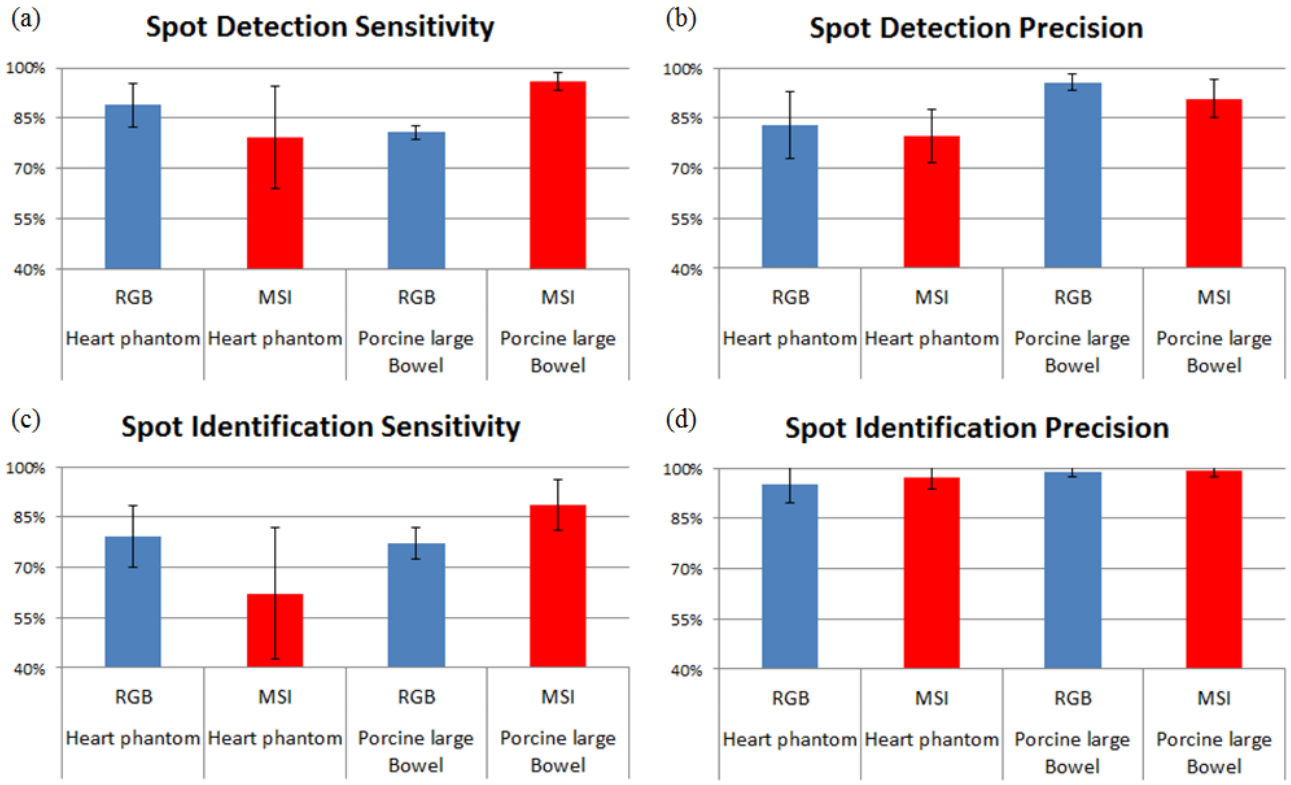


Figure 4.3: The spot detection and identification results on different objects. (a). Spot detection sensitivity. (b). Spot detection precision. (c). Spot identification sensitivity. (d). Spot identification precision. In all charts, the experimental results using RGB camera are shown in blue; while those using multispectral camera are shown in red.

The experimental results, however, also shows the instability of the proposed pattern decoding algorithms, whose performance could be affected by factors such the surface texture, topology, background illumination, and so on. This is further discussed in section 4.6.

4.5.2 Reconstruction Results Using Self-calibration

The pattern decoding and reconstruction results using self-calibration are compared with those using the manual calibration, which has demonstrated a high accuracy in section 3.4.2. In the experiment, 63 out of the total 75 SL images are used, for which the pattern decoding algorithm leads to promising results (precision $> 90\%$, sensitivity $> 40\%$). The results were evaluated by rigidly registering the reconstructed surface from manual and self-calibration. Since the surfaces are reconstructed up to scale, the scaling factor is the only variable during registration. After

registration, the distances between the two surfaces are used as indicators of reconstruction error (Fig. 4.4).

For individual SL images with available identification results, the average and maximum values for absolute distances are calculated. Since the real scale of surface cannot be acquired by the self-calibration method, these two indicators are both represented as relative errors as a percentage. Table 4.1 shows the statistics regarding images from different sources. To demonstrate the overall results, the median value of the average distance and the maximum distances regarding all the SL images were acquired and compared.

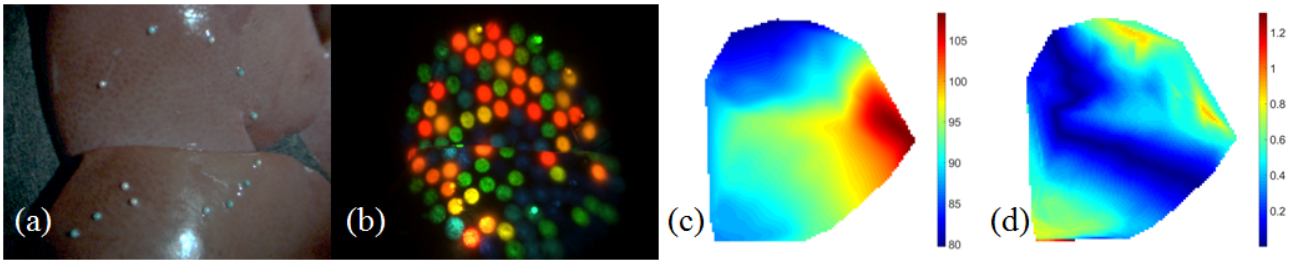


Figure 4.4: (a) Liver under white light; (b) Liver under SL; (c) Reconstructed surface using self-calibration where colour indicates the surface depth (in mm); (d) Relative error map on the reconstructed surface using self-calibration where colour indicates the reconstruction error (%). Figure reproduced from [168].

Table 4.1: Reconstruction errors using the self-calibration technique on benchmark data. Table reproduced from [168].

Object	Mean distance	Median of mean distance	Max. distance	Median of Max. distance
Heart phantom	0.32% ± 0.13%	0.28%	1.03% ± 0.51%	0.88%
Liver	2.36% ± 5.26%	0.80%	3.84% ± 2.54%	2.68%
Kidney	0.40% ± 0.37%	0.20%	1.76% ± 1.70%	0.88%
Heart	0.81% ± 0.68%	0.44%	2.93% ± 2.39%	1.43%
Large Bowel	3.58% ± 4.38%	2.07%	16.70% ± 23.29%	7.55%

The mean and median value of both the average and maximum distance demonstrate the

promising result of reconstruction using the self-calibration technique. But it is noticeable that according to the maximum distance, despite generally good performance for the case of the liver with blood and the porcine large bowel, possible large errors could still happen in some areas of the reconstructed surface. This phenomenon has been studied and it is found that it is caused by the pour imaging of light patterns on the tissue surface, due to strong scattering of light at long wavelengths and absorption of light at short wavelengths. It also indicates that the current light pattern might not be the most suitable one to create unambiguous SL images. But the current SL system generally works well and shows its potential to be applied with self-calibration technique in surgery environments.

4.5.3 Incorporation of Local Normal Information

The estimated local normals provide additional surface detail. The TPS interpolation [**sec:improvement˙c** is able to provide a reconstruction with high quality for objects with smooth surfaces. However, adding normal information helps to improve the surface detail only when the object surface is relatively complex, in which case the sparse projected light pattern is not able to reconstruct local shapes. In the case of imaging an object, the estimated local normal information does not affect much the final reconstructed surface shape whose surface is generally smooth. Therefore, it is hard to quantify the benefit brought by this algorithm. We tested the normal estimation algorithm on a rectangular box, a cylinder, and a liver phantom, to showcase the reconstruction improvement intuitively. Fig. 4.5(b,d,f) demonstrates the reconstruction of an area around one edge of a rectangular box, a cylinder, and the liver phantom, respectively. The local normals (red arrows) are well estimated using the proposed method.

Since local normals do not benefit the reconstruction much on relatively smooth surfaces, a small object consisting of two steps was used to intuitively show the improvement of surface reconstruction after incorporating normal information. Light patterns were projected onto the object, and Fig. 4.5(g,h) show the reconstructed surfaces before and after addition of local normal information. The step shape is only accurately defined when incorporating normal information, demonstrating that this method could improve the reconstruction when a sparse light pattern is projected on a non-smooth object surface.

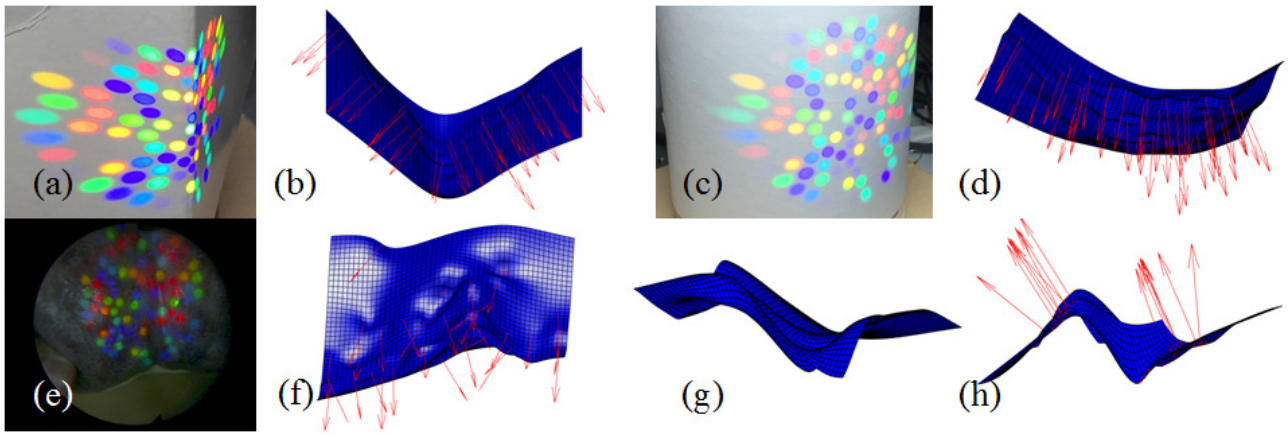


Figure 4.5: (a). Rectangular box edge under white light and SL. (b). Reconstructed rectangular box edge (blue) aided by normals (red). (c). cylinder under white light and SL. (d). Reconstructed cylinder (blue) aided by normals (red). (e). Liver phantom under white light and SL. (f). Reconstructed liver phantom (blue) aided by normals (red). (g). Reconstructed step surface (blue) without normal information. (h). Reconstructed step surface (blue) aided by normals (red).

4.6 Discussion and Conclusions

In this chapter, further systematic extensions have been proposed to improve the practicality and performances of the original ICL SL system and the naive 3D reconstruction pipeline that are described in Chapter 3. The improvements mainly contain three aspects: 1) replacing the RGB camera into a multispectral camera to facilitate pattern decoding; 2) one-the-fly self-calibration based reconstruction that enables a changing relative position between the probe and the endoscope; 3) Integration of local normal information to improve the reconstruction accuracy.

In order to enhance the light pattern decoding procedure, a multispectral camera was proposed. The *in vivo* experiment demonstrated its superiority over the RGB camera in the spot identification step. Except for leading to a better colour discrimination ability, the multispectral camera can also be used in other medical applications, such as tissue classification, oxygen level estimation, synthetic narrow-band image generation, and so on.

An on-the-fly self-calibration technique along with an iterative algorithm, was also implemented to update both the epipolar geometry and feature matching to enhance SL inference of 3D structural information. Comparison between the resulting reconstruction and that from manual calibration demonstrates the feasibility and robustness of this technique in object surface

reconstruction. However, further studies are needed to localise depth in metric space. This self-calibration based reconstruction method could be extended to be used for other SL systems with a known project light pattern.

Based on the assumption that some projected spots are located on one plane in 3D space, local normal information is estimated according to the homography between two corresponding ellipses. The experimental results show that the improvement of incorporating normal information is notable and that this normal incorporation technique can also be adapted in other SL systems using light patterns with specific known shapes.

However, the experiments in section 4.5.1 and section 4.5.2 also show that the robustness of the proposed pattern decoding algorithm needs to be improved. Further improvements on the computational speed and the robustness of the algorithm are illustrated in Chapter 5.

To sum up, in this chapter, further algorithmic developments based on the original system have been provided. The overall and step-wise performance of the hardware and software platforms have been evaluated by phantom, *ex vivo*, and *in vivo* experiments, demonstrating a potential of using the system in clinical applications. The main challenge of this system, i.e., the non-real-time performance and the requirements of parameter tuning in the light pattern decoding method, will be addressed in Chapter 5 by using deep learning based models and integrating monocular Structure-from-Motion (SfM) algorithms.

Chapter 5

Towards Faster, Denser, and More Robust Surface Reconstruction

5.1 Introduction

In Chapter 3 and Chapter 4, the ICL SL system was introduced, as well as a naive reconstruction pipeline and corresponding further extensions. Although the proposed system is generally capable of measuring tissue surface shape in *in vivo* experiments, it still suffers from problems such as no white light (WL) visual feedback to surgeons, slow computation, a requirement for manual interaction, and sparse reconstruction. To address these problems, in this chapter some of the key elements in the reconstruction pipeline have been upgraded: 1) To guarantee the reconstruction accuracy, instead of on-the-fly calibration the probe is fixed on the tip of the endoscope by a 3D printed adapter, and the system can switch between the WL and SL modes

The content of this chapter has been published in:

1. J. Lin, N. T. Clancy, X. Sun, J. Qi, M. Janatka, D. Stoyanov and D. S. Elson, "Probe-based rapid hybrid hyperspectral and tissue surface imaging aided by fully convolutional networks", In: *Proc. MICCAI, Part III.*, 2016, pp. 414-422.
2. J. Lin, N. T. Clancy, Y. Hu, J. Qi, T. Tatla, D. Stoyanov, L. Maier-Hein and D. S. Elson, "Endoscopic Depth Measurement and Super-Spectral-Resolution Imaging", In: *Proc. MICCAI, Part II.*, 2017, pp. 39-47.

via introducing a chopper wheel. 2) A new light pattern decoding algorithm has been proposed, using convolutional neural networks (CNNs) based spot detection followed by specially designed feature matching. 3) The reconstruction result for SL is combined with structure-from-motion (SfM) for denser reconstruction. The new pipeline enables a real-time, quasi-dense and more robust tissue surface reconstruction.

5.2 Improvements of the ICL SL Optical System

The optical hardware system have been upgraded in this work to enable a stroboscopic illumination of WL and SL, and a fixed relative position between the probe and the endoscope tip, by introducing an optical beam chopper and a customised 3D printed adaptor. The schematic of the upgraded system is demonstrated in Fig. 5.1.

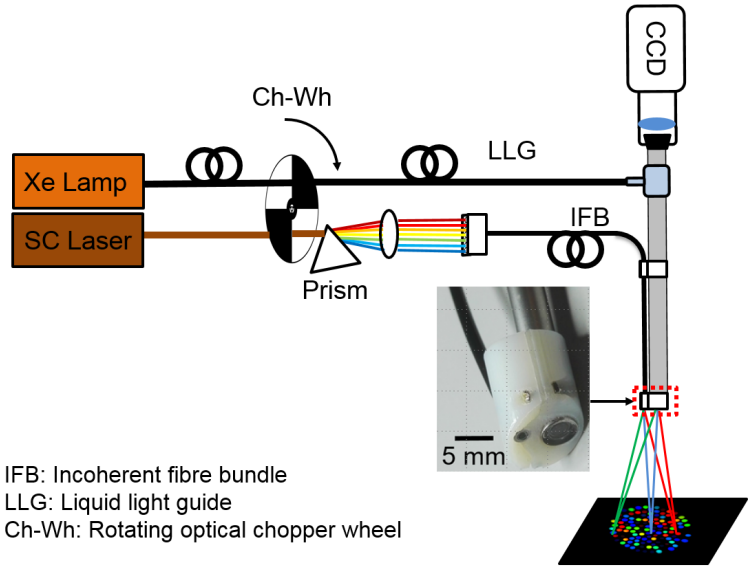


Figure 5.1: The schematic of the upgraded ICL SL system with chopper wheel and probe adaptor. Figure adapted from [166].

5.2.1 Interleaved Switching Between WL and SL Images

Rapid stroboscopic switching between WL and SL was achieved using an optical chopper wheel (3501 Optical Chopper; New Focus, Inc., USA), as shown in Fig. 5.2(a). This chopper wheel spins at a user-defined frequency, thus enabling an interleaved illumination given both the light

sources were set to pass the inter-fan area. Two fibre optic light cables were used for WL: one was connected to a xenon lamp and the other to the laparoscope. Their free ends were then positioned opposing each other, separated by a 2 mm air gap through which the chopper wheel could pass. The chopper was mounted so that the output of the supercontinuum laser also passed through the wheel, and the SL and WL beam paths were alternately blocked or transmitted as it turned. The result was that the light emerging from the tip of the instrument switched between SL and xenon illumination at the chopping frequency.

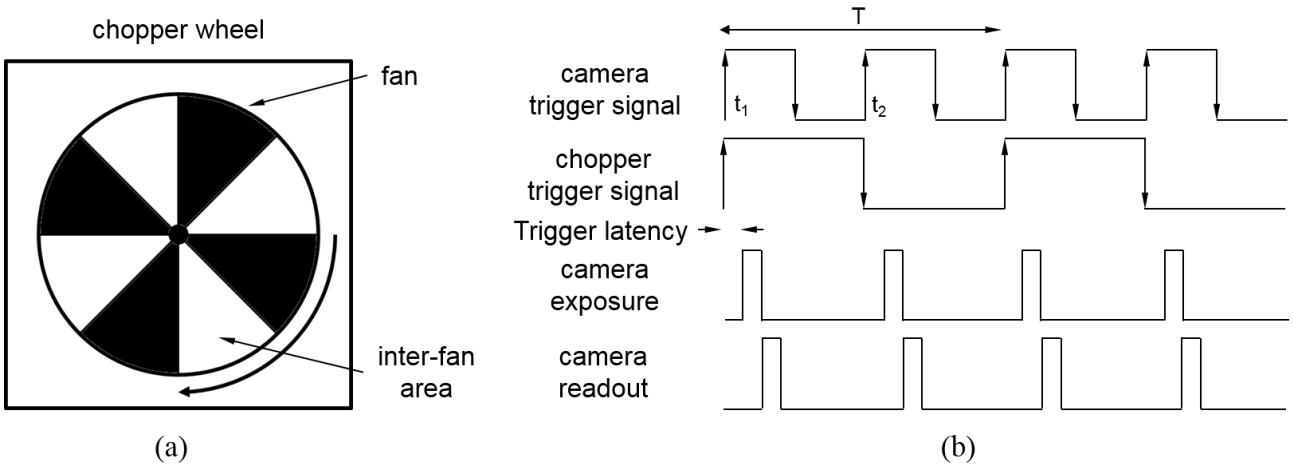


Figure 5.2: (a) The chopper wheel. The inter-fan area allows light to pass through. (b) Trigger timing schematic. The 5 V output signal from the signal generation device is used to trigger acquisitions from the camera and the chopper wheel. The chopper period is indicated by T, which is double the camera trigger signal. The rising and falling edges (e.g. t_1 and t_2) triggers the camera, and the trigger latency, exposure and readout times define the upper limit of the acquisition framerate.

Separately, a computer-controlled signal generation device (NI USB-6211; National Instruments Corporation, USA) was used to produce two synchronized square waveforms of variable frequency and phase. One was used to trigger image acquisition by the CCD camera, while the other controlled the rotation frequency and the phase of the chopper wheel. The trigger frequency was set to twice that of the chopper and the phase adjusted so that the acquired frames comprised of alternating SL and WL-illuminated images. In this way, both the WL and SL images were captured at a frequency of 8FPS. The WL was provided to the surgeons for visualisation as well as used for SfM reconstruction (section 5.4), and the SL was fed to the PC for SL based tissue surface 3D reconstruction.

5.2.2 Fixation of the Probe

In the previous ICL SL system, the probe was mounted on a support beside the endoscope in the naive 3D reconstruction pipeline (section 3.3), in which case minor movement of the instruments in real-world applications might reduce the reconstruction accuracy; or the on-the-fly calibration was applied to estimate the relative position between the probe and the endoscope tip, which is a relatively inaccurate reconstruction method due to the instability of the feature matching step (section 4.5.2). Therefore, the system needs to be upgraded to guarantee an accurate calibration data acquisition.

A 3D printed adapter was developed to mount the SL probe on the tip of the rigid endoscope (5 mm diameter Hopkins II Optick 30°, Karl Storz GmbH, Germany). This adapter is cylindrical (11 mm diameter), with two channels to house the endoscope and SL probe, respectively (Fig. 5.3). According to the field of view (FoV) of the probe ($\approx 65°$) and the endoscope ($\approx 70°$), the angle and relative positions (12° angle, 3 mm offset) of these two channels were designed to maximise the overlapping area of these two FoVs, enabling the SL system to reconstruct tissues surfaces in approximately 1.5–4cm working distances, with a baseline ≈ 6 mm. The small size of the adaptor enables the insertion of the probe and endoscope tip with enough working space, through a typical laparoscopic trocar that can be, for example, used in laryngeal surgeries.

5.3 Light Pattern Decoding

In the ICL SL system, projected spots with different colours are considered as features. In surgical environments spot detection is the key step for accurate pattern decoding, and is the bottleneck for surface reconstruction using the ICL SL system, in terms of both accuracy and efficiency. The strong light-tissue interactions, appearance of mucus, and large curvature of the tissue surface increases the difficulty in efficient and accurate processing in surgical scenes. Pattern detection aiming to divide all the pixels in a SL image into two groups: spots (foreground) and background, is a binary segmentation problem.

Image segmentation has been a popular research field in computer science for decades. New algorithms emerge frequently. Compared with the traditional algorithms (thresholding, region growing, normalized cut, etc.), that lead to promising performance in general but require elab-

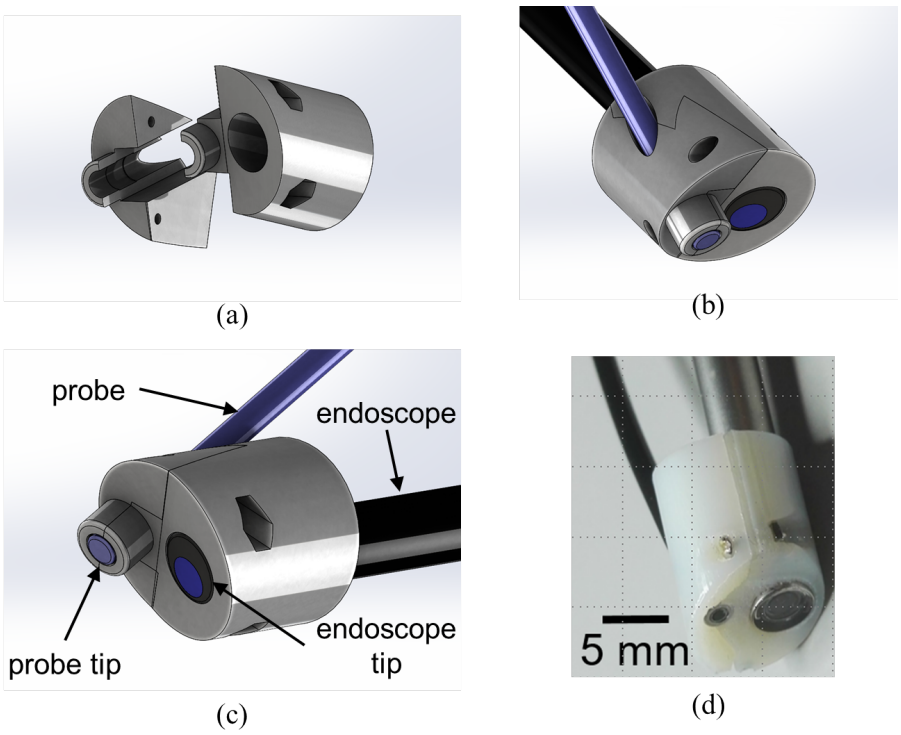


Figure 5.3: (a) The adaptor consists of two parts that can be joined by screws (a). (b)(c) The CAD model showing adaptor fixing the probe (blue) to the tip of the 5 mm rigid endoscope (black), from different viewing angles. (f) A photograph of the printed prototype.

orative manual parameter tuning, supervised machine learning methods are in general more suitable for automatic segmentation. These methods benefit from relatively simpler manual interaction, by formulating the majority of parameter tuning into model training. Therefore, supervised machine learning algorithms are generally more suitable for semi-/fully-automatic segmentation given an adequate training set. In recent years, advances in hardware have made solving large-scale problems (training/testing on large datasets) using CNN possible in a reasonable amount of time. CNN based methods dominate different large-scale image classification [58], object detection [169], and segmentation [16, 37] challenges. Due to its superiority in performances over other algorithms in solving these general computer vision problems, a CNN based model is utilised in this work to realise a real-time and accurate spot detection, without manual interaction. To show the advantages of using CNN based algorithms, background information on general CNN based segmentation techniques are provided in the following section.

5.3.1 Neural Network based Image Segmentation

Image segmentation tasks aim at assigning labels to all image pixels. In this section a brief literature research on supervised ANN based image segmentation techniques is provided. According to the label assignment schemes the existing image segmentation methods based on ANN can be divided into two categories: sliding-window style pixel/ patch-wise classification and fully convolutional network (FCN) based segmentation.

The pixel/patch-wise based classification follows a classic approach for image segmentation. The basic idea is to use ANN models as classifiers to assign labels pixel by pixel or patch by patch. Before the flourishing of CNN, neuron networks using FC layers were widely used to segment the image via pixel classification. At this stage, task-specific hand-crafted feature extraction is applied first and then followed by ANNs. This type of method has been widely used in different tasks such as textual content detection [156], remote sensing image recognition [250], finger print segmentation [184], food contamination detection [211], brain image segmentation in MRI images [106, 136], mammongram segmentation [38], chest radiograph segmentation [191], and so on.

The application of CNN models enables pixel/patch-wise classification without hand-crafted feature extraction, since the convolutional layers in the models are capable of extracting task-specific "deep" features automatically via network training. Such examples can be found in knee cartilage segmentation in MRI images [224], human action recognition [130], text recognition [285], cells and nuclei segmentation [201]. Likewise, classic CNN classification model such as LeNet [28, 159], AlexNet [151], VGG [257], or their variants, can be utilised.

However, sliding-window style pixel/patch-wise classification based segmentation using CNN has two main drawbacks: 1) local features, limited by the patch size, rather than global ones are extracted and used for the segmentation; 2) looping through all pixels/patches decreases the segmentation efficiency, due to the large number of repetitive computations on the overlapping areas between adjacent patches. Therefore, segmentation pipelines that take entire images as inputs, are more suitable, as they normally integrate more global information from a larger receptive field, and lead to higher efficiency in processing the whole image at once instead of separate single image patches. So this section mainly focuses on the review of segmentation methods based on FCN based CNN models.

Most of the successful FCN based segmentation methods are rooted in milestone CNN models developed for image classification problems. As is described in section 2.4.1, these classic models share similar architectures that consist of a "Conv-pooling-Conv-pooling" feature extraction encoder phase, followed by a FC layer based classification scoring phase. The main differences between them are the depth of the CNN and the detailed connection strategies. Using typical classification models, by replacing the FC layers with Conv layers, the resulting architecture enables an end-to-end pixel-wise classification both in training and prediction. This greatly improves the segmentation efficiency and avoids hand-crafted feature selection and consequently reduces the dependency on image pre-processing.

One of the most successful and representative models is fully convolutional network (FCN) developed by Long et al. [172, 252]. Its architecture was inspired by the classification models (Fig. 5.4). FCN takes advantage of powerful feature extraction ability to learn "deep" features in different hierarchies, producing a feature map that is small in the spatial dimension but deep in the channel dimension. This feature map is then upsampled spatially by a decoder mainly consisting of deconvolution layers (transposed convolutional layers). In this phase, accurate location and deep feature information is combined to refine the label scoring results by "shortcuts", through which the intermediate feature maps at each scale level from the encoder phase are merged to the corresponding layers in the decoder. The data mergence is applied via element-wise sum.

There are other two popular FCN variants — SegNet [18] and U-Net [234], for which the cross-entropy loss function for all the pixels is adopted as the loss function:

$$L_d = -\frac{1}{2N}\sum_{n}^{N}\sum_{i} {y_i}^{(n)} log(\hat{y}^{(n)}) \tag{5.1}$$

where N is the total pixel number. All these models follow an encoder-decoder architecture, but differs in terms of detailed layer arrangement, upsampling manner, and the data mergence method in the shortcuts. Both FCN and U-Net use deconvolution layers for upsampling. The former merges shortcut feature maps with element-wise add, while the latter concatenates feature maps from corresponding encoder and decoder layers in channel dimension, and merges them via convolution. Instead of using deconvolution layers, SegNet upsamples feature maps using the max-pooling indices from the corresponding layers in the encoder phase, then gener-

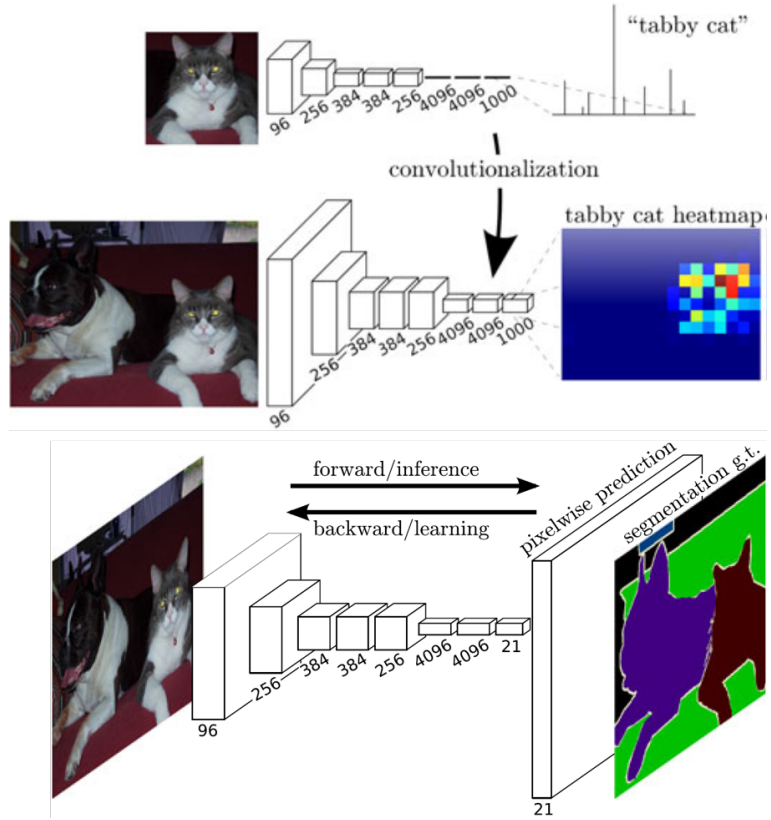


Figure 5.4: The fully convolutional network (FCN) model. Replacing the FC layers with Conv layers enables low-resolution heatmaps that highlight the existence of target objects; extending the network with deconvolution layers produces pixel-wise class scores. Figure source: [252] (copyright ©2017 IEEE).

ates dense feature maps via convolution. Different strategies adopted by these different models are shown in Fig. 5.5. These 2D FCN based models have also been extended into 3D, mainly for solving medical image segmentation problems [49, 61, 236].

Compared to sliding-window fashion pixel/patch wise classification methods, the FCN based network structures are more efficient and are capable of making use of features from larger local areas for label assignment, due to the relatively large receptive field of the filter in the final layer of the encoder. However, they still lack global context, due to the local connectivity property of the convolutional layers: Using normal Conv layers, the receptive field only increases linearly with the number of Conv layers with stride 1. With limited number of Conv layers, the receptive field would not be large enough to extract global information. Therefore, many studies have been carried out to extend and apply the FCN based models with other post-processing techniques to integrate global context information [87].

Dilated convolution [309], has been proposed to increase the receptive field of the CNN models.

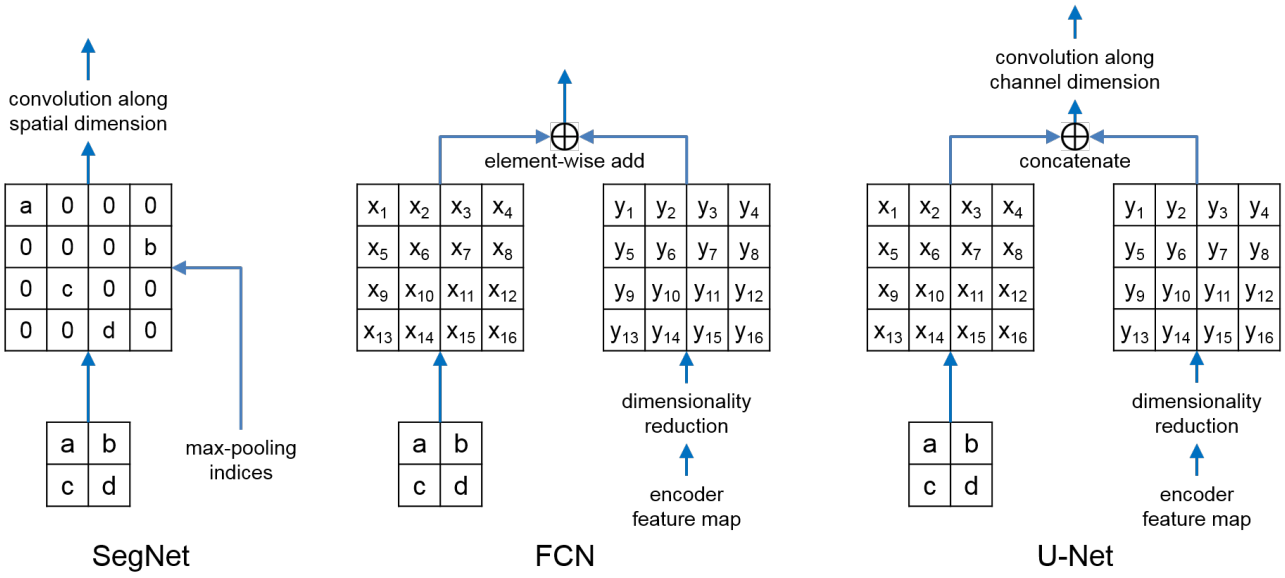


Figure 5.5: The data mergence strategies in different models. SegNet uses the corresponding max-pooling indices followed by spatial domain convolution; FCN uses element-wise sum; U-Net uses concatenation and convolution along the channel dimension. Figure reproduced and adapted from [18, 87] (copyright ©2017 IEEE).

By adjusting the "dilation rate" l, dilated convolution enables interval changes between filter intervals. By piling up several layers, the receptive field for the filter in the last layer increases exponentially (Fig. 5.6). Yu et al. adopted dilated convolution for multi-scale context aggregation [302]. Another work by Paszke et al. mixed the dilated convolution with normal Conv layers to construct a model with parallel strided convolution and max-pooling downsampling [210]. DeepLab [44, 45] combines dilated convolution with a post-processing technique called conditional random fields (CRFs), which will be illustrated in the next paragraph.

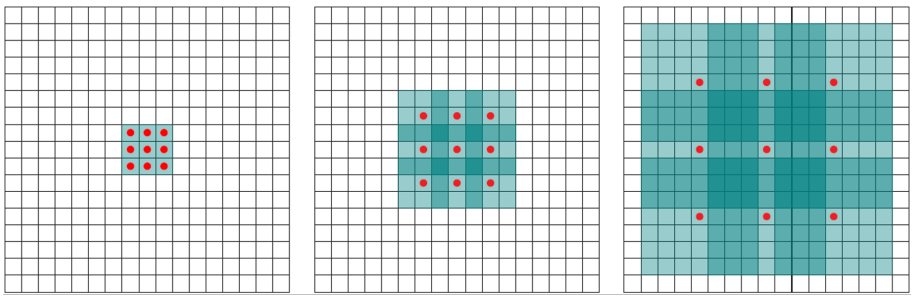


Figure 5.6: The exponential increase of receptive field by stacking firstly 1-dilated convolution (a), then 2-dilated convolution (b), and finally 4-dilated convolution (c). Figure source: [302].

Conditional random field (CRF) is a statistical model that can be used in machine learning

problems for prediction [155]. Taking pixels as nodes and interaction between pixels as edges, an energy function can be built. Minimising the function enables assigning the same labels for similar pixels, and different labels for pixels with large differences. This manages integration of local and global context information in the image. Chen et al. used a fully connected CRF model [150] as a separated post-processing step to refine the segmentation results from the CNN [44, 45]. Kamnitas et al. proposed to combine fully connected CRF with an efficient multi-scale 3D CNN for accurate brain lesion segmentation [138]. Another significant work was done by Zheng et al. [307]. In this work the CRF with Gaussian pairwise potentials and mean-field inference was formulated as recurrent neural networks (RNNs), and merged into the CNN pipeline, enabling an end-to-end training and prediction.

Combining multi-scale prediction results is another solution to integrate global information. In practice, the FCN based models are apt to detect objects in "certain scales", which affect the generalisation of the model. Therefore, some researchers developed models that process images on different scales and merge the results as the final prediction [87]. Bian et al. proposed a n FCNs structure which combines n FCN models and merges their prediction results. The individual FCNs were trained separately and then the whole model was trained [26]. In the work of Roy et al. [237], four networks of different scales were chained in a coarse-to-fine manner for label assignment. A two-scale VGG-16 based model was developed by Raj et al., in which two way processing with different network depths, was applied followed by extra convolutional layers to merge predictions on different scales. Another significant work, the nested U-Net model with RNNs, has been illustrated in a previous study for 3D biomedical image segmentation [43]. This work will be described in the following.

Recurrent neural networks (RNNs) are powerful tools to integrat temporal information, and enable global information to be kept and delivered along the video frames. As 3D data such as CT or MRI also have continuation between adjacent slices, similar to videos, RNNs have also been applied to aid biomedical image recognition and segmentation tasks as well. Due to higher efficiency and lower memory occupation, RNNs are advantageous over 3D convolution models in processing 3D data. Vanilla RNNs can not well model long-term dependencies due to the gradient vanishing problem, so different gating mechanisms have been used instead. For example, an intuitive combination of U-Net and Gated Recurrent Units (GRUs) [48] was adopted to segment MRI cardiac images [222]; ReSeg model used GRUs with a VGG-16 network for

segmentation [282]. An alternative of gating mechanism is Long Short-Term Memory (LSTM) [118]. Valipour et al. has proposed a FCN with convolutional recurrent units for video segmentation [279]. Bi-directional LSTM was applied with nested U-Net for 3D biomedical image segmentation to segment 3D images from Serial Block-Face Scanning Electron Microscope [43]. Another example is the parallel multi-dimensional LSTM based architecture for Electron Microscope image segmentation [263]. Unlike the previous one, in this work bi-directional LSTM based segmentation was applied in all directions, resulting in six predictions. The final results were achieved by merging the output of all the branches.

There are new CNN based segmentation algorithms emerging everyday, so it is impossible to give a full review in this thesis, and only the most representative works including those milestones are included. From the analysis, it can be concluded that: 1) FCN based segmentation models are appreciated due to their higher efficiency compared to sliding-window style pixel/patch-wise classification methods; 2) the CNN models can be combined with other post-processing techniques if the global context information is expected to be integrated.

5.3.2 FCN based Feature Detection

In this work an algorithm based on FCN has been deployed to detect the SL spots in endoscopic images robustly in near real-time despite poor image quality in *ex vivo*/*in vivo* animal experiments.

Pre-processing

The captured SL images were pre-processed using the same method that was described in section 3.3.2 for specular highlight removal and image smoothing, followed by a linear interpolation based image scaling operation to resize the images to $[512 \times 384]$ $(\text{width}, \text{height})$ [128], before being fed into an FCN model for further prediction.

The FCN Model

According to a previous work [234], FCN models are capable of achieving promising segmentation with a relatively small training set, if applied on images with repetitive patterns. While

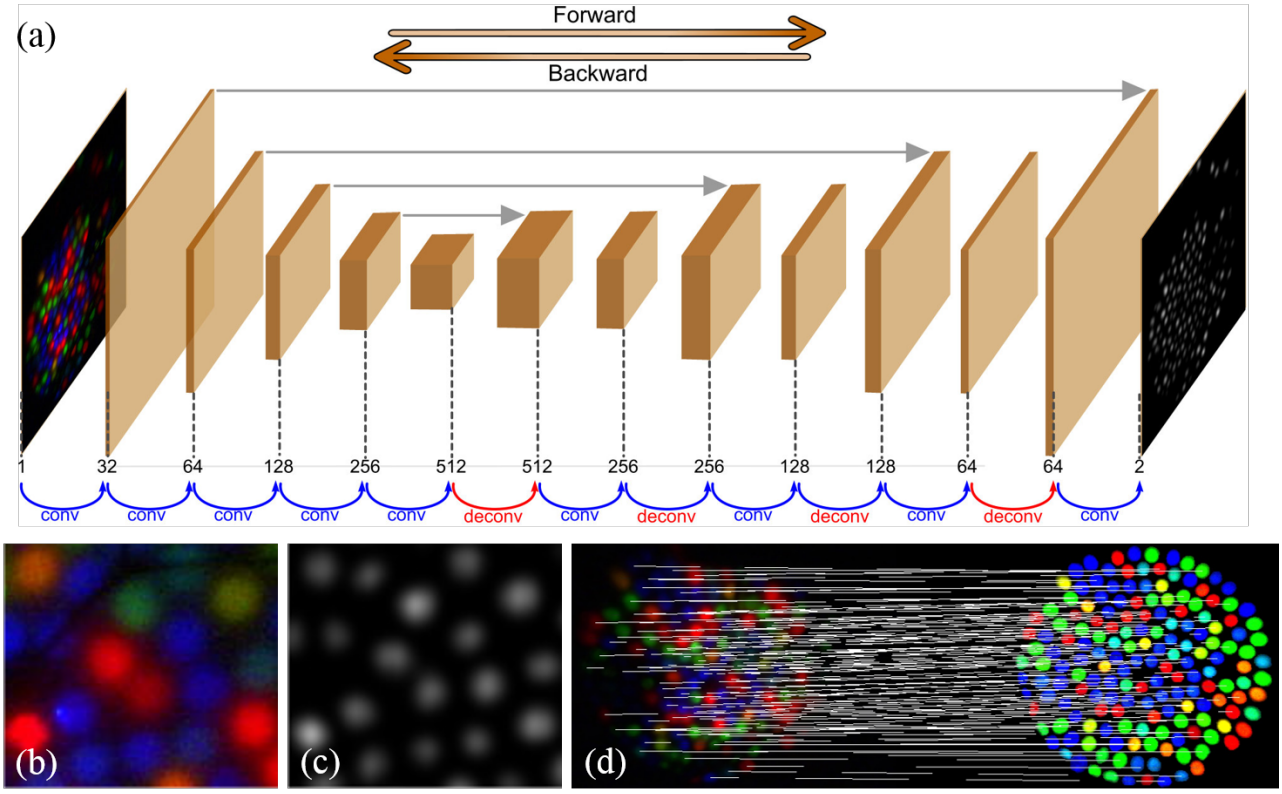


Figure 5.7: (a) The deployed FCN model; (b) Cropped SL image; (c) Density map of the cropped SL image; (d) Feature matching between the captured (left) and the reference SL image (right). Figure source: [167]

the spots in SL images produced by the ICL system are generally uniform in morphology, but with colour differences, it is expected that FCN is a suitable model that can be used for spot detection. The proposed FCN model followed the encoder-decoder structure, with U-Net style shortcuts for feature map mergence between the encoder and decoder, as is shown in Fig. 5.7(a). The encoder consists of consecutive units with the same layer combination. Each unit contains one 3×3 Conv layer followed by an ReLU activation layer and a max-pooling layer with a size 2×2 and a stride 2 . This halves the input image size while doubles the channel dimension. In the last layer of the encoder, each neuron has an effective receptive field of 46×46 pixels, and the feature dimension of its output volume is 512. The decoder also contains four consecutive units, where each step upsamples the image, fuses the upsampled feature maps and the convoluted counterpart from the encoder, and then convolutes to decrease the feature dimension. This feature mergence thus supplements the loss of the location information in the encoder due to consecutive downsampling, combining the coarse "what" and fine "where" information for a pixel-wise prediction of the whole image. The output feature map is cropped, resulting in a 2-channel image with the same size as the input image. Each channel indicates the probabilities

of being foreground (spots) or background.

Training

For training a loss function consisting of cross-entropy loss and $L2$-regulariser was used. From equation 2.13 and 5.1:

$$L_d = -\frac{1}{2N}\sum_{n}^{N}\sum_{i} {y_i}^{(n)} log(\hat{y}^{(n)}) + \lambda \sum_{j} w_j^2 \tag{5.2}$$

where n, i, and j indicate the pixel index, the class index, and the weight index, respectively. Empirically the decay weight λ was set to 0.0005.

The training set came from approximately 200 SL images captured from phantom, *ex vivo*, and *in vivo* experiments, with different spot sizes and different working distances. The segmentation ground truth were generated manually, with the software "ilastic" [260]. Data augmentation was applied to increase the training set size to more than 20k. This includes resizing in three scales, left-right and top-bottom flipping, rotations, and affine transformations. Stochastic Gradient Descent (SGD) solver [28] was used for training. The training included a user-interactive coarse-to-fine tuning, with batch size 1 and momentum 0.9. The learning rate was gradually changed from 0.01 to 0.0005, in 500000 iterations. The final loss converged at about 0.006.

Prediction and Post-processing

In a normal FCN model used for segmentation, in the prediction stage an "arguments of the maxima" (argmax) layer is stacked behind the multiple-channel output. This layer outputs a single channel image whose pixels represent the indices of maximum values of the input data along the channel dimension. However, in our situations due to poor image quality from the tissue surface, generating binary output with the argmax layer leads to "overlapping" or connection between adjacent detected spots, which requires further complex post-processing to separate them before the spot matching step. Therefore, a different approach was adopted. Through observation, it was found that the the subtraction between the foreground and background channels produced a "Gaussian-like" distribution for each spot, while the locations with highest values indicated the spot centre. So this subtracted image was deployed, with a binary

mask generated by the argmax layer for background removal, to generate the "probability map" (Fig. 5.7(b,c)). The "probability" values in every isolated connected component of the mask were normalised to the range $[0 - 255]$. Then the local maxima of the map was detected via adaptive thresholding, and their centres of mass were treated as spot centres.

The training and prediction of the FCN model was realised by Caffe [131], a deep learning framework. Using its C++ API, the prediction with the FCN model was integrated into the 3D reconstruction pipeline.

5.3.3 Fast Feature Matching with Epipolar Constraints

Matching spots between the captured and reference SL images is similar to the feature matching step in stereoscopy. Two main improvements to the feature matching method have been developed, compared with the previous pipeline: 1) a pattern-specific feature detector to describe the detected spots for fast matching; 2) the epipolar constraint is utilised to constrain the search space of the matching. These methods are described in the following.

Based on the spot detection results, correspondences between the captured and reference SL images were found using a customized feature descriptor and epipolar constraints. An algorithm based on Delaunay triangulation with false connection pruning was applied to establish the neighborhood of detected spots, which were each allocated a customized $32 \times 32 \times 3$ feature descriptor to indicate the color distribution in 32 directions at 32 orientations (Fig. 5.8(c)). The 1st dimension was the index of the starting direction (orientation), the 2nd indicated the direction and the 3rd indicated the average color (normalized red, green, and blue values).

For every spot in the reference image, candidate matched spots in the captured image were firstly found using the epipolar constraints, according to the distance from the spot centres to the epipolar line (Fig. 5.8(b)). The smallest distance between a spot descriptor on the captured image and all 32 descriptors of a spot on the reference image was used to describe the distance between them. The match with closest distance smaller than a threshold was chosen, followed by a pruning procedure based on neighborhood information. An iterative method was then applied to propagate matches to other neighboring unmatched spots, followed by pruning, until the number of matches remained constant.

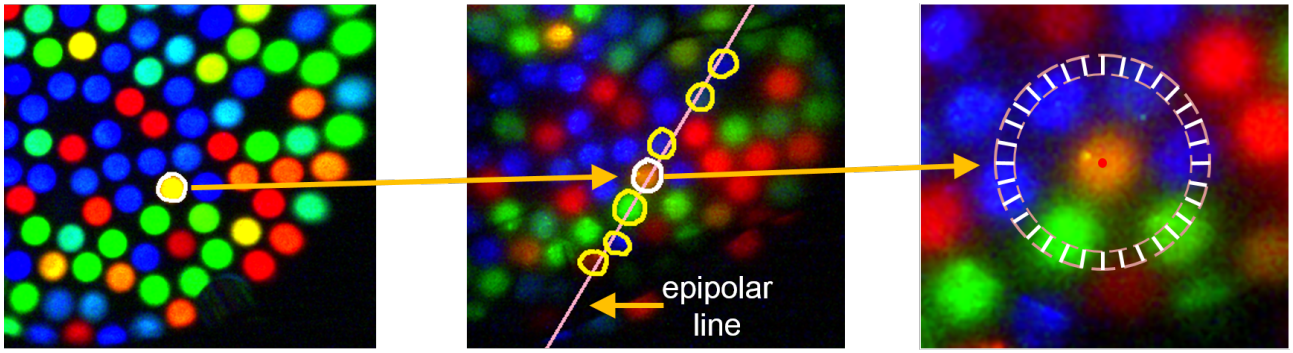


Figure 5.8: (a) The reference SL image, with a target spot delineated in white outline. (b) A captured SL image, with candidate matched spots close to the epipolor line, with false matching delineated by yellow, and positive matching by white. (c) The working space marked by white segments. The normalised ratios between red, green, and blue in each segment are used as elements in the feature vector.

After spot matching (Fig. 5.7(d)), the target object surface shape was recovered by triangulation using calibrated parameters, as illustrated in section 3.3.3. Neither the classic linear method nor the optimal one that minimises the reprojection error [109] was adopted in this work for triangulation, since errors would be introduced if the ICL SL probe was simulated as an inverse camera, as the emitted rays from the fibres did not intersect exactly at one "probe centre". So the line-intersection method was used for triangulation, same as the naive reconstruction pipeline (section 3.3.3).

The 3D reconstruction pipeline was implemented in C++ with external libraries including OpenCV [128], Caffe [131], and VTK [246]. The computation time for reconstruction from a single SL image was $\approx$ 80 ms on a PC (OS: Ubuntu 14.04; processor: i7-3770; graphics card: NVIDIA GTX TITAN X).

5.4 Quasi-dense Reconstruction with SL and SfM

Due to the reduced image quality for *in vivo* data, the SL reconstruction does not always return a sufficiently dense reconstruction. Therefore, SL was combined with monocular SfM on a GPU to increase the density and robustness of the reconstruction. SfM attempts to reconstruct an object surface from different viewing angles achieved via camera movements [109]. If no external information is provided then the reconstructed object surface shape can only be found up-to-

scale given correspondences between different views. The density of the reconstruction using SfM depends on the feature detection and matching procedure (correspondence search) on the WL images, but is generally much more dense than that can be achieved by the ICL SL system ($\approx$ 1000 vs. $<$ 171 points). However, SL provides the surface scale, for the reconstructed surface using SfM. Thus, in this section, a method combining SL and monocular SfM is proposed for denser tissue surface shape reconstruction. The workflow is shown in Fig. 5.9 and is described in further details in section 5.4.2.

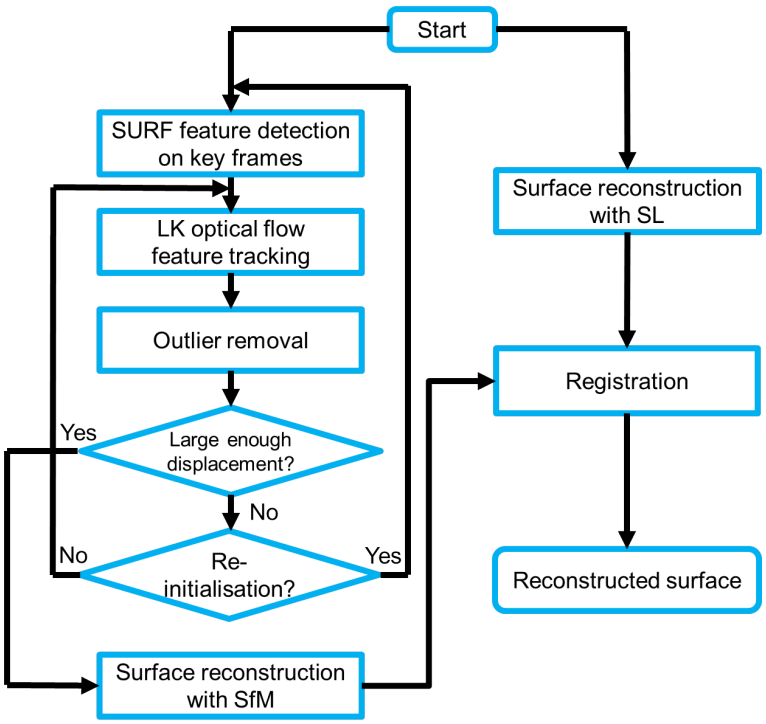


Figure 5.9: The workflow for combining SL and SFM. Figure source: [166].

5.4.1 Monocular SfM

In this work, a method combining SURF based feature detection [22] and LK optical flow based tracking [175] has been proposed to perform a correspondence search. Several criteria were applied to exclude tracking outliers, including the feature descriptor difference, flow vector length, temporal smoothness, symmetric optical flow, and random sample consensus (RANSAC) for essential matrix estimation. For the surface reconstruction we assumed the surface was rigid during a small time window. Considering two normalised points x_p and x'_p from adjacent frames corresponding to the same point X in a 3D coordinate system whose origin was established

based on frame 1, it was assumed that there exist projective matrices $P_p = [I|0]$ for frame 1 and $P'_p = [R|t]$ of frame 2, where R and t indicate the camera relative position, and $t = -RC$ with C representing the camera centre in frame 2 in frame 1's camera coordinate system. These points should follow $x'^T Ex = 0$. The essential matrix E describes the relative position between the cameras in the two frames. By estimating E using Singular Value Decomposition (SVD), and selecting from four possible solutions using the criterion that the points can only exist in front of the camera in both frames, X can be found. In theory, given enough correspondences, the surface can be measured using two adjacent frames. However, in the real-world application, the per-frame reconstruction is limited by 1) small camera movement; 2) inaccurate camera parameters; 3) imperfections of the pinhole model. These elements lead to unstable reconstruction, thus we only applied reconstruction when the features had been tracked with large enough displacements (> 20 pixels).

5.4.2 Fusing SL and SfM Data

Since interleaved SL and WL images were grabbed, the surface reconstructed with SL was used to determine the scale of the reconstructed surface using SfM in the following WL frame. Given the point cloud (PC) $[\mathbf{X}_s, \mathbf{Y}_s, \mathbf{Z}_s]$ reconstructed by SL, $[\mathbf{X}_m, \mathbf{Y}_m, \mathbf{Z}_m]$ reconstructed by SfM, in the camera coordinate system, their normalised coordinates can be written as $[\mathbf{X}_{sn}, \mathbf{Y}_{sn}, 1]$ and $[\mathbf{X}_{mn}, \mathbf{Y}_{mn}, 1]$, respectively. The algorithm that registers the two PCs is shown as follows:

1. For the SfM PC the convex hull S on the x - y plane is firstly found and defined by a subset of $[\mathbf{X}_{mn}, \mathbf{Y}_{mn}]$, $S := \{x_{mn}^{(l_i)}, y_{mn}^{(l_i)}\}$, where $\{l_1, \cdots, l_K\}$ denotes the ordered indices of 2D points.

2. Loop through all $[\mathbf{X}_{sn}, \mathbf{Y}_{sn}]$ to select those points $[\mathbf{x}_s, \mathbf{y}_s, \mathbf{z}_s]$ that are located inside the convex hull on the x - y plane, by checking the sign consistency of the cross product between the the query $[\mathbf{X}_{sn}^n, \mathbf{Y}_{sn}^n, 0]$ and all the segments that define the boundary of convex hull.

3. Interpolate the SfM PC to a denser PC $[\mathbf{x}_m, \mathbf{y}_m, \mathbf{z}_m]$ [14].

4. Find nearest neighbours of all the selected SL points $[x_s^{(n)}, y_s^{(n)}, y_s^{(n)}]$ in the interpolated SfM PC $[x_m^{(n)}, y_m^{(n)}, y_m^{(n)}]$ in the x-y plane.

5. Register the two PCs using the established correspondences in step 4, with the "scaling" t of the SfM PC as the only transformation parameter. The registration can be formulated as a cost function minimisation problem:

$$cost = \sum_{n}^{N} (x_s^{(n)} - tx_m^{(n)})^2 + (y_s^{(n)} - ty_m^{(n)})^2 + (z_s^{(n)} - tz_m^{(n)})^2 \tag{5.3}$$

   which can be further simplified as a quadratic equation: $cost = t^2 + at + b$, whose minimum value can be found at $t = -\frac{a}{2}$, which is also the closed-form solution of the SfM PC scaling factor.

6. Interpolate both the SL PC and scaled SfM PC with the same specified x and y coordinates.

7. Estimate the final surface by averaging the z coordinates between the two interpolated PCs.

In this way, one quasi-dense surface reconstruction with known scale can be achieved for every pair of SL/ WL frames, at a framerate of 8 FPS, limited by the modality-switching frequency of the optical hardware.

5.5 Experimental Results

5.5.1 Pattern Decoding Validation

For stereo based reconstruction methods including the ICL SL system, the reconstruction accuracy is mainly limited by two factors: calibration and feature matching. As a SL calibration technique based on standardised camera calibration with homography has been proposed and validated earlier in this work (section 3.3.1), pattern decoding is the only procedure that is validated in this section.

Experiments on a silicone heart phantom, *ex vivo* ovine heart, and *ex vivo* ovine liver have been carried out for validation, with the probe adaptor used to fix the probe next to the endoscope at an angle $\approx 12°$, and baseline ≈ 6 mm. Videos were recorded for all the experiments, in which

the working distance was always kept at $1.5 - 4$ cm, with the endoscope optical axis roughly perpendicular to the tissue surface. The endoscope movements were planned to simulate real ones that could happen in theatre, by including sudden fast movements and movements along the axial direction.

Fig. 5.10 shows the intuitive "probability map" estimation and feature matching examples for four representative SL images during *ex vivo* experiments. It is clearly demonstrated in the figure that spots can be detected regardless of poor image quality, even for inhomogeneous illumination between spots or specular highlights.

The corresponding "probability" for brighter and darker spots differs in size instead of value, due to the normalisation step which forces the maximum value in individual binary masks to be 255. This algorithm would certainly fail in the case that images suffer from over-exposure or massive specular reflections, but it generally guarantees robust performance on normal tissue surfaces. Also, it is noticeable that despite being successfully detected, some "isolated" spots induced by large surface curvatures could not be matched.

Table 5.1: Validation of feature matching. Manual annotation is used as the ground truth.

Object	Annotated matches	True positive	Sensitivity	Precision
Silicone heart phantom	170 ± 1	154 ± 17	0.907 ± 0.101	0.997 ± 0.004
Ovine heart	134 ± 8	113 ± 10	0.844 ± 0.038	0.997 ± 0.006
Ovine liver	128 ± 11	110 ± 13	0.861 ± 0.052	0.994 ± 0.005

Statistical results are provided in Table 5.1, with 10 randomly chosen frames per recorded video on different objects. Automatic feature matching results were compared with those from manual annotation. The true positive (correctly detected and matched spots), annotated matches (manually annotated spot matches), together with the matching sensitivity and precision are used as indicators.

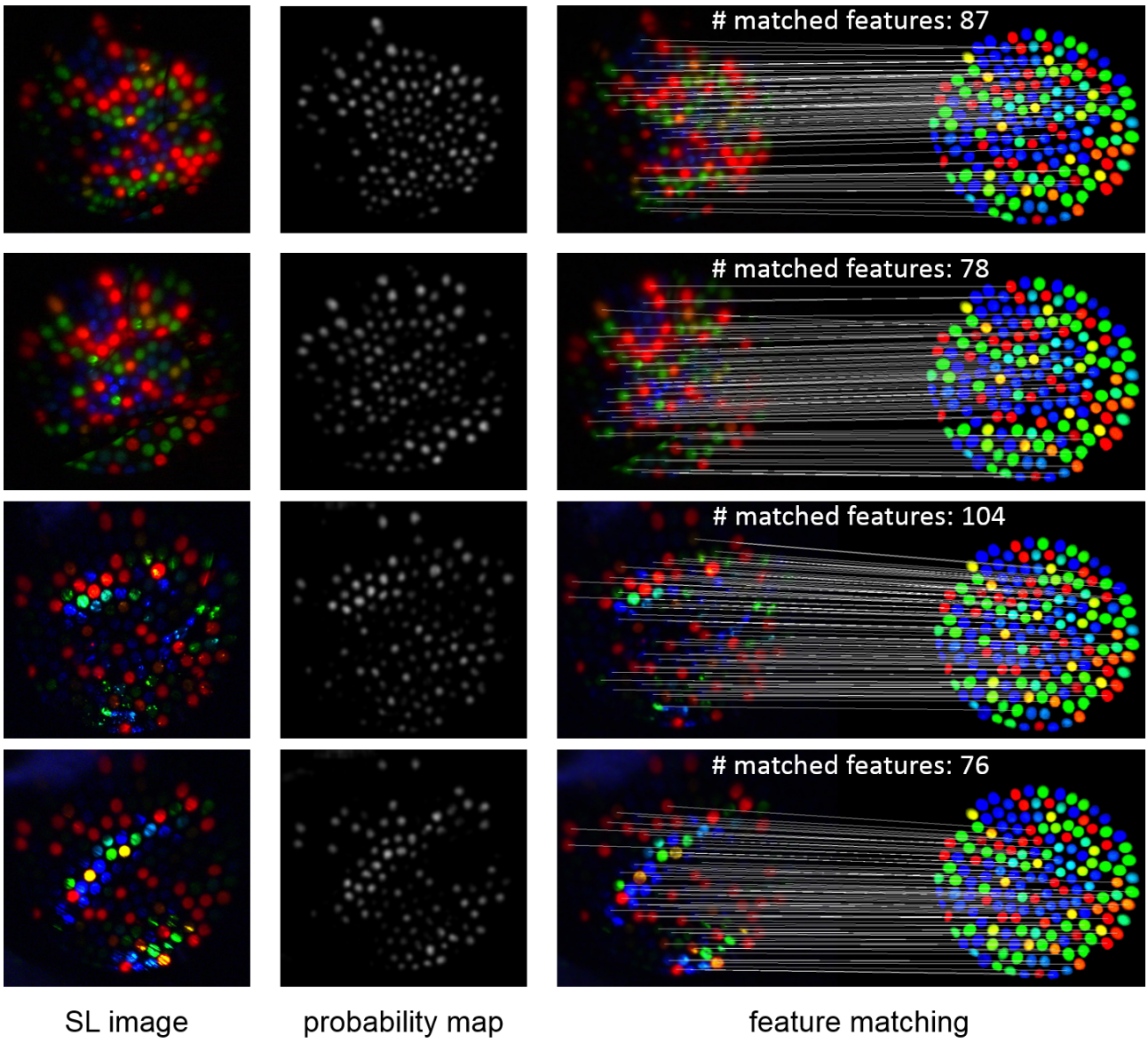


Figure 5.10: Examples of pattern decoding results using the proposed method. The estimated probability maps and the matching between the captured and reference SL images are shown on four example images. The upper two are for the ovine heart, and the lower two for the ovine liver.

Table 5.1 shows that the pattern decoding algorithm functions robustly in general. The average sensitivity on the phantom experiments was higher than 0.9 due to high image quality; it was also around 0.85 in both the *ex vivo* experiments, where the image quality was affected by the appearance of mucosa, and specular reflection, etc. This indicates the capability of the algorithm for spot detection. It is also noticeable that for all data the feature matching precision was higher than 0.99, showing the reliability of the spot detection using the proposed method. Compared with previous work (section 3.4.1), this shows a much higher sensitivity

for feature matching. Meanwhile, the near real-time performance (12 FPS) guarantees its real-world practicality. These improvements are owning to the deployment of FCN based detection algorithm, and the efficient feature matching enabled by customised feature descriptors and the epipolar constraint. Further discussion is provided in section 5.6.

5.5.2 A Clinical Case

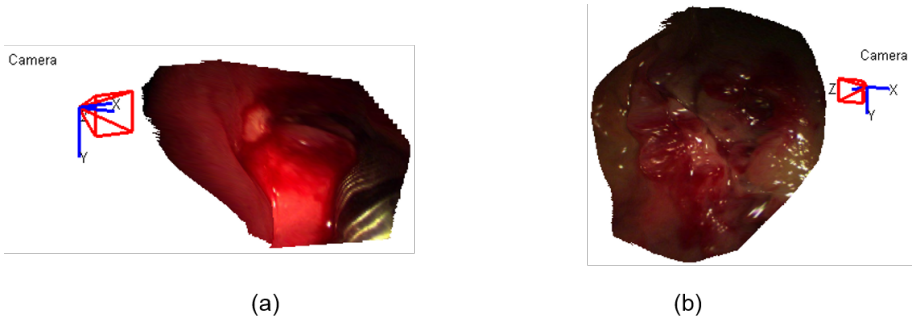
The system has been tested on clinical cases. During Laryngeal surgery in the operating theatre, the ICL SL system was placed on a two layer trolley beside the operation table. The imaging part of the system was set up on a breadboard with size $40 \times 40\ cm^2$, surrounded by a light-tight structure constructed from blackout materials (Fig. 5.11(a)). The surgeon held the endoscope with a customised CCD camera attached (DCU 223C, Thorlabs Ltd., UK), and the probe. The endoscope could be inserted through a standard clinically used trocar for optical imaging during the laryngeal surgery (Fig. 5.11(b)).



(a) (b) (c)

Figure 5.11: (a) The ICL SL system set up on a two layer trolley. (b) The surgeon holding the endoscope with probe attached, and inserting it through a standard trocar. (c) the SL pattern projected on the patient's larynx.

In vivo SL and WL videos on the patient larynx, as well as *ex vivo* ones on the excised laryngeal tissue, were recorded for further processing. The reconstruction of two frames from *in vivo* and *ex vivo* cases, are shown in Fig. 5.12.



(a) (b)

Figure 5.12: (a) The reconstructed surface from an *in vivo* experiment on patient larynx. (b) The reconstructed surface from an *ex vivo* experiment on excised patient larynx tissue.

5.6 Discussion and Conclusions

In this chapter, improvements for the ICL SL system were proposed, which enabled a faster, denser, and more robust tissue surface reconstruction. The discussion on the hardware and algorithm aspects are provided as separate sections:

5.6.1 Hardware

The implementation of the interleaved SL and WL imaging provided WL views with shape and texture information that could be extracted for further applications e.g., object tracking or visual servoing, to benefit MIS and robotic surgery. Currently it enables a maximum recording framerate of 8 FPS for both the SL and WL videos. Due to the relatively small FOVs and fast movements of normal endoscopes, the captured images might suffer from image blurring at this framerate. Analysing the SL images does not depend on fine texture information, so the reconstruction with SL would not be affected, in contrast to the applications using WL images. In the future, switching to another RGB camera that can be triggered at higher frequency could be a solution.

A 3D printed adaptor was designed to house the probe beside the endoscope tip with optimised angle and baseline that enabled an accurate triangulation at working distance $1.5 - 4$ cm. This technical improvement is an important step towards real-world clinical application of the ICL SL system, since self-calibration led to larger reconstruction error (section 4.5.2). The current

adaptor has a diameter of around 11 mm, which enables the endoscope tip and probe to be inserted through a clinical trocar. In order to adapt the system to laparoscopic surgeries that only allow smaller trocars, both endoscope and adaptor with smaller size need to be designed and adopted.

5.6.2 Algorithm

An accurate pattern decoding algorithm has been proposed to facilitate the near real-time tissue surface reconstruction using the ICL system. An real-time 3D reconstruction algorithmic pipeline has been implemented to solve spot detection and identification. Treating spot detection as a binary segmentation problem, together with pre-/post-image processing an FCN model was applied to estimate a "probability" that indicated spot locations, which were then utilised for spot identification with a specific feature descriptor and epipolar constraints.

For detection an FCN model with U-Net style was adopted to estimate the "probability" map, whose local maxima were regarded as the spot centres. The selection of the U-Net style model was done empirically, and according to the its advantage that smaller size of training set is needed if the image pattern is repetitive [234]. Through training with augmented data from initial 200 SL images, the proposed method leads to a significant improvement in both the accuracy and speed compared with the previous normalized-cut based method (section 3.3.2). Furthermore, the supervised training and prediction procedure avoided manual parameter tuning that is frequently required in traditional image segmentation pipelines.

An upgraded spot matching method has been integrated into the 3D reconstruction pipeline. Compared with the counterpart in the naive reconstruction pipeline (section 5.3.3), the proposed method makes use of the epipolar constraint to constrain the correspondence search space, and a customised feature descriptor for Brute-Force matching, followed by a robust pruning step to remove false positives. All of these improve the matching results and can be executed in several ms when implemented in C++.

Although highly promising results have been achieved, there is still space to improve the pattern decoding. For examples, more SL images from different tissue surfaces should be collected to improve the generalisation of the training set; the detection can be upgraded with state-of-the-art CNN algorithms; the recent research on CNN based stereo matching [190] or optical flow

estimation [65] could also give us some hints to design a model that is capable of end-to-end training and pattern decoding.

Merging SfM with SL provided a quasi-dense surface reconstruction with known scale. As this method was also implemented with C++ on a GPU, it guaranteed a real-time reconstruction performance together with the pattern decoding step. However, the two-view geometry based SfM is not the state-of-the-art and is based on the assumption that the tissue remains static in a small time window, which introduces errors. Furthermore, the tracking based on SURF and LK optical flow is not robust enough for tissue surfaces in surgical scenes. Thus the monocular SfM should be upgraded to e.g. a state-of-the-art SLAM method such as ORB-SLAM [197] which has been developed to handle many practical difficulties. The fusion of SL surfaces and SfM surfaces is intuitive, simple, and fast. However, in the future work, it still needs to be extended to address more difficult situations, like reconstruction errors that could occur in both SL and SfM, the optimisation of point cloud fusion, etc.

With the improvements proposed in this chapter, the ICL SL system has been pushed towards its target — a medical device that can be reliably deployed in real clinical environments. Besides the future work for each step described above, the hardware and software need to be integrated into one framework for consistency. Currently, the synchronised mode switching and video recording is realised via LabVIEW, while the reconstruction pipeline including the calibration, pattern decoding, and final data fusion and 3D reconstruction is realised in C++. One of the most important tasks in the future work will be the "gluing" and integration of the hardware and the software with C++ under one framework. Besides the engineering part, more animal trials and clinical experiments are also expected to continuously validate the ICL system at different development stages.

Chapter 6

Hybrid Structured Light and Hyperspectral Imaging System

6.1 Introduction

In this chapter, the functionality of the ICL SL system was further extended, to enable tissue surface hyperspectral imaging (HSI). A slit hyperspectral imager was integrated in to the system, in the way that the reflectance from the tissue surface under white light is fed into the slit for hyperspectral signal measurement. This ensures the measurement of sparse hyperspectral signals from the tissue surface. The SL and HSI modes can be switched by controlling a "flipper" mirror. In order to estimate pixel-level dense hypercubes for imaging purposes such as narrow band imaging (NBI) and tissue oxygen saturation (StO_2) measurement, a CNN based super-resolution model has been developed to estimate hypercubes from the RGB images and

The content of this chapter has been published in:

1. J. Lin, N. T. Clancy, X. Sun, J. Qi, M. Janatka, D. Stoyanov and D. S. Elson, "Probe-based rapid hybrid hyperspectral and tissue surface imaging aided by fully convolutional networks", In: *Proc. MICCAI, Part III.*, 2016, pp. 414-422.

2. J. Lin, N. T. Clancy, Y. Hu, J. Qi, T. Tatla, D. Stoyanov, L. Maier-Hein and D. S. Elson, "Endoscopic Depth Measurement and Super-Spectral-Resolution Imaging", In: *Proc. MICCAI, Part II.*, 2017, pp. 39-47.

the sparse hyperspectral signals. In the experiments, the accuracy of the captured hyperspectral signal was evaluated by experiments on a Macbeth color checker, and phantoms. *Ex vivo* tissues like pig bowel, rabbit uterus, and lamb uterus were used to validate the performance of the proposed dense hypercube estimation algorithm.

6.2 From SL to SLHSI

As is demonstrated in section 2.3, HSI has high potential to aid medical applications including disease detection and surgical guidance, as it yields signals with rich information at tissue, cellular, and molecular levels. Different HSI acquisition modes have also been reviewed, and from the analysis it is concluded that generally a miniaturised snapshot based HSI system that produces hypercubes with reasonably high resolution is highly suited to MIS related applications.

Considering the structure of the fibre bundle that is used for in the ICL SL system, the integration of HSI is possible:

1. The fibre bundle realises a 1D $\leftrightarrow$ 2D spatial mapping (Fig. 6.1). The fibres are arranged in a linear array and randomised 2D pattern at the two ends, respectively. In SL mode the dispersed laser is fed into the linear array end and emitted from the randomised end to be projected on the tissue surface as artificial patterns. As the light propagation is reversible, light coming from the tissue surface, e.g. the reflectance or fluorescence, can travel in the opposite direction — be fed into the randomised end and come out of the linear array end. In this case, the light emitted from different fibres is matched to different locations on the tissue surface.

2. The "linear array" fibre pattern coincides with the shape of the sensor on a slit hyperspectral imager. Therefore, if focused accurately, the light from the tissue surface can be directly fed into the hyperspectral imager with an extra lens combination. In this way, a 1D $\leftrightarrow$ 2D spatial mapping can be established between the slit sensor and the projected light patterns on the tissue surface. The benefit of this is that sparse HSI can be carried out with a region-of-interest (RoI) covering the majority of the ICL SL system's field-of-view (FoV), by the imager in a snapshot.

3. As the fibre bundle distal tip (randomised end) that projects light patterns in the ICL system in SL mode has already been miniaturised (2.0 mm diameter), no additional modifications need to be made on this end to adapt the device to an endoscope in surgical applications.

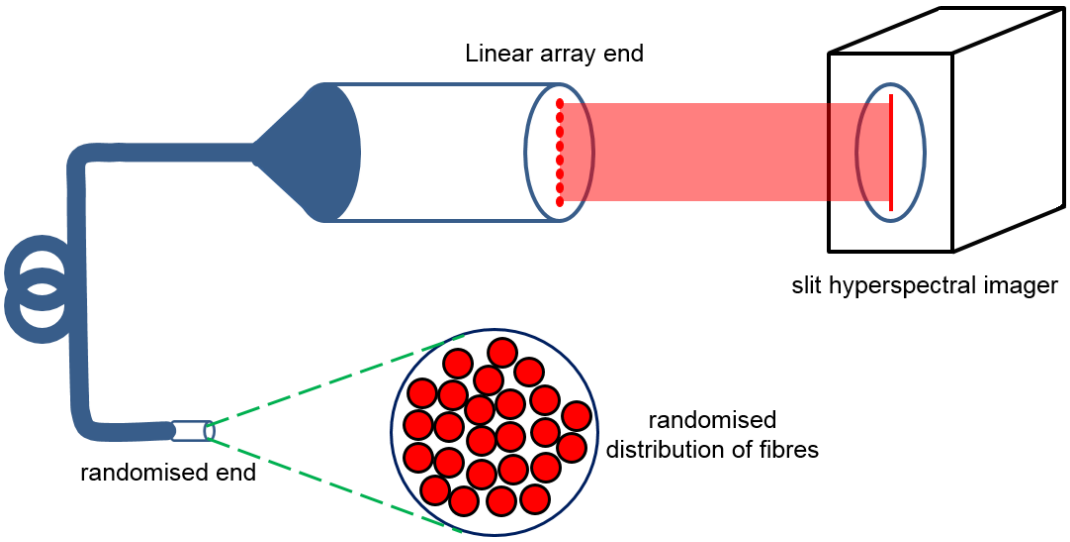


Figure 6.1: The spatial mapping between the randomised end (2D), the linear array end (1D), and the slit hyperspectral sensor.

6.2.1 System Configuration

The extended system, namely the ICL SLHSI system, was built based on the ICL SL system (section 3.2) [51, 165]. A flipper mount (Motorized Filter Flip Mounts, Thorlabs Ltd., UK) mounted with a mirror is used to switch between the SL the HSI mode. It is positioned between the lens and the linear array end of the fibre bundle (Fig. 6.2(right)).

In the SL mode (Fig. 6.2(left)), the flipper mount is fixed in the horizontal direction out of the light propagation path. As is described in section 5.2, the supercontinuum laser passes through the inter-fan space of the chopper wheel and is dispersed and coupled into the probe's linear array end. Rays containing different wavelength information are then emitted form the randomised end and projected onto the tissue surface to form a SL image. This image is captured by a CCD attached to an endoscope and further analysed for 3D reconstruction.

In the HSI mode (Fig. 6.2(right)), the supercontinuum laser is blocked, and a white light source (Xenon 300; Karl Storz GmbH, Germany) is activated to shed light onto the tissue

sample. The reflectance propagates into the randomised end and is emitted out of the linear array end. Meanwhile the flipper is flipped to the vertical position to the linear array end at about 45 °angle, to redirect the reflectance towards a 250/50 mm focal length lens combination to form a demagnified image of the fibre array on the slit of a HSI imager (Nano-Hyperspec; Headwall Photonics, Inc., USA).

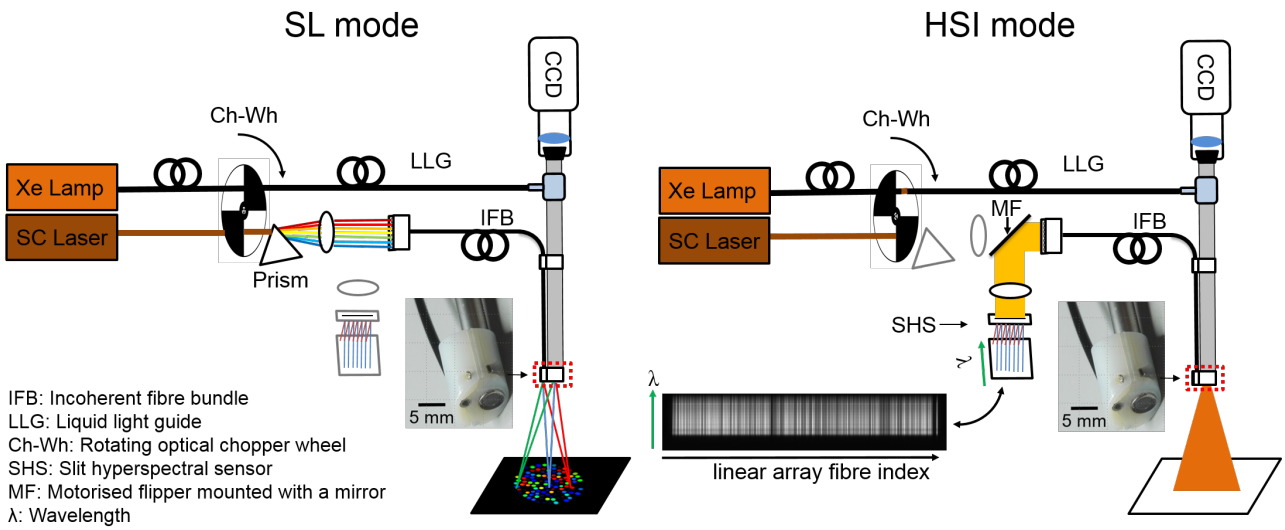


Figure 6.2: Left: setup in SL mode. Supercontinuum laser is used to generate the structured light, and the white light is blocked. Right: setup in HSI mode and the captured spectrograph image (lower right). Supercontinuum laser is blocked and white light is shed on the tissue surface. The mirror mounted on a flipper is used to redirect the reflectance to the slit hyperspectral imager.

6.2.2 Hyperspectral Signal Calibration and Processing

The output of the HSI imager is a grey-level image whose X axis indicates the position on the slit, and Y axis the wavelength. The system was manually adjusted to ensure a full display of all the reflectance from different fibres. So the hyperspectral signals from 171 fibres were displayed in the output image (Fig. 6.2(lower right)).

Hyperspectral Signal Calibration

In order to find the correspondence between the x coordinate and the actual positions where the reflectance comes from the tissue surface, a one-off calibration step before the experiment was carried out. Since the indices of signals captured by the HSI imager follow the order the

fibre arrangement at the linear array end, the problem was transformed into correspondence search between fibres at the two ends of the fibre bundle. By feeding white light into the linear array end and recoding a video showing the projected light patterns on white paper, a translation stage was utilised to initially cover all the fibres and then gradually uncover them in one direction. According to the order of emergence of the projected spots on the white paper, the 1D $\leftrightarrow$ 2D correspondence between the slit sensor and the captured SL/ WL images was established.

The X-axis locations, where the spectrum of light from a certain fibre is located, need to be determined every time after the system is re-assembled. Presumably an automatic peak search algorithm that looks for local maxima could be applied to allocate the x coordinates for individual fibres. However, manual annotation was adopted to avoid miss-detection, since the the calibration is a one-off procedure. Then the spectrum of different fibres can be acquired along the detected x coordinates. Each fibre$'$ s reflectance spectrum R was achieved using this extracted (detected) spectrum I, the "dark spectrum" (I_D) from the ambient light, and the reflectance (I_O) from a white reference target (Spectralon; Labsphere, Inc., USA), to correct for wavelength-dependent transmission characteristics of the system:

$$R = \frac{I - I_D}{I_O - I_D}, \tag{6.1}$$

from which the absorption spectrum (A) is calculated:

$$A = -\ln(R). \tag{6.2}$$

Oxygen Saturation Estimation

Previous research shows that one of the most important medical applications of HSI are oxygen saturation (StO_2) estimation, to aid clinical decision making and surgical guidance [283, 7, 50]. A linear model can been used to interpret the absorption spectrum (A) of the tissue surface. A can be seen as a linear combination between the absorption spectra from oxy and deoxyhaemoglobin [261, 230], as is demonstrated in Fig. 6.3. This is a model that works under several assumptions that are realistic in many cases [50]: 1) the haemoglobin is the main absorber with affects the absorption spectrum; 2) light with different wavelengths travels with

equal distances; 3) wavelengths of interest have similar scattering losses.

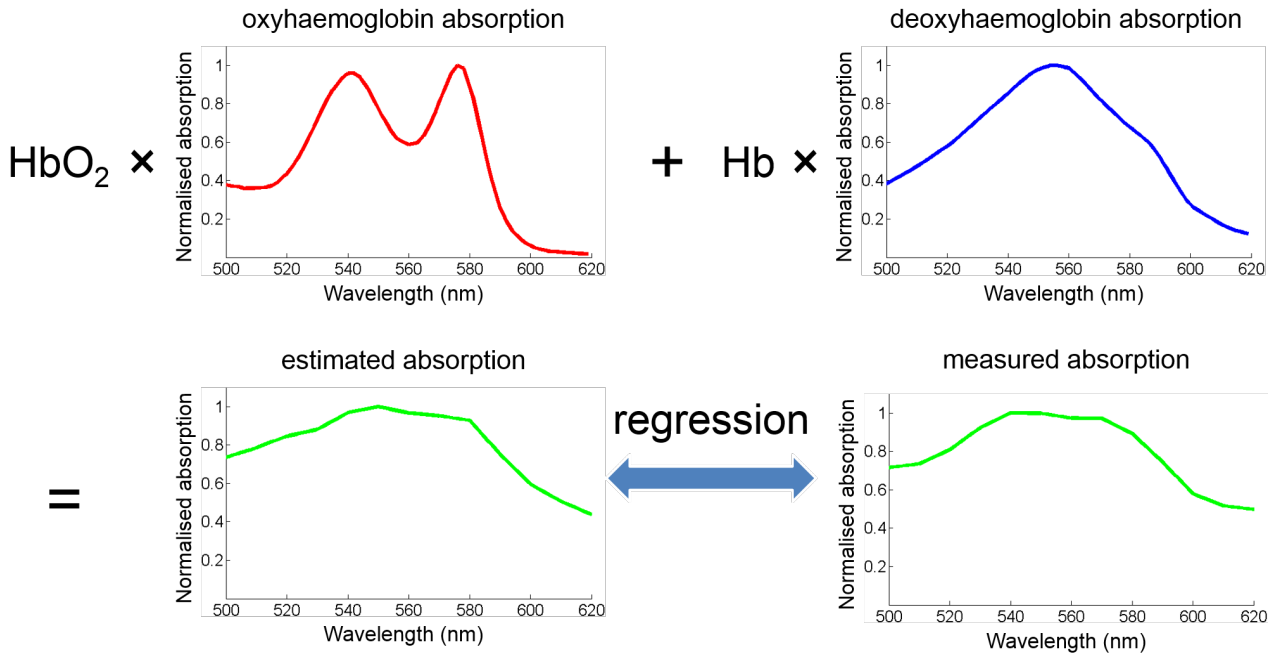


Figure 6.3: A linear combination of oxy and deoxyhaemoglobin absorption spectra results in the final absorption spectrum. The relative ratio between the oxy and deoxyhaemoglobin can be estimated via linear least squares.

Given the concentrations of oxy and deoxyhaemoglobin (HbO_2 and Hb), their joint absorption can be represented as

$$A' = [HbO_2]A_{HbO_2}(\lambda) + [Hb]A_{Hb}(\lambda) + \alpha, \tag{6.3}$$

where $A_{HbO_2}(\lambda)$ and $A_{Hb}(\lambda)$ are the absorption spectra for oxy and deoxyhaemoglobin, with α as a constant. $A_{HbO_2}(\lambda)$ and $A_{Hb}(\lambda)$ are measured molar extinction coefficients provided in [223]. Using least squares linear regression to approximate A with A', $[HbO_2]$ and $[Hb]$ can be estimated with closed-form solution by Singular Value Decomposition (SVD). Finally, the oxygen saturation StO_2 can be calculated as

$$StO_2 = \frac{HbO_2}{HbO_2 + Hb}. \tag{6.4}$$

Fusing SL and HSI

For the purpose of detailed medical applications such as AR, the extracted information from HSI signal should be combined with the tissue surface recovered by SL. By calibration (section 6.2.2), a unique mapping was established between spectra and SL spots in the reference SL image. As the captured SL image can be identified by the algorithm proposed in section 5.3, the spectra collected in the HSI mode can be mapped onto different locations in the SL image captured in the previous frame. Then the extracted information such as the oxygen saturation can be shown on the reconstructed surfaces.

6.3 Pixel-level Dense Hypercube Estimation

There is always a trade-off between spatial resolution, spectral resolution, acquisition time, and cost, to be considered in HSI system design. The main benefit of the ICL SLHSI system is its snapshot based data acquisition mode. However, it is also noticeable that only a spatially sparse spectrum can be directly collected in a snapshot. Although this still provides medically meaningful information that covers the FOV, the majority of the area is still left blank after HSI falling between the spot locations. However, generally a hypercube with pixel-level dense spatial resolution is preferred. Thus, in this section, a CNN based algorithm is described for dense hypercube generation, which is referred to as "super-spectral-resolution" in the following context.

There are two data sources that can be used to recover the hypercube, the RGB image captured by the CCD camera (DCU 223C, Thorlabs Ltd., UK), and the spatially sparse spectral signals captured by the slit hyperspectral imager. Previously, there have been algorithms proposed to recover high-resolution (HR) images from low-resolution (LR) images, or hypercubes from RGB images. Inspired by the existing algorithms, in this project a CNN based model has been proposed to recover the hypercubes from both the RGB images and captured sparse spectrum signals. Generally speaking, this workflow includes a spectral super-resolution (SR) operation for the RGB image, and a spatial SR for the sparse HSI signals. Before describing the detailed algorithm, a literature review is given on the SR techniques that aim to recover HR data from LR data, either in the spatial or spectral dimension.

6.3.1 State-of-The-Art on Super-Resolution Methods

As a popular research area in the past two decades, SR algorithms have been developed for different applications such as facial image analysis, satellite imaging, road sign plates recognition, etc. In this section previous algorithms of single-image super-resolution (SISR) are reviewed. SISR refers to the procedure that finds the mapping from a single input LR image to an output HR image. Despite that it mainly refers to a procedure in the spatial dimensions (X and Y), those algorithms should be able to be adapted for spectral SR, since the hypercube is a 3D matrix.

Non-CNN Based Spatial SR Techniques

Several survey papers [198, 297] have been published on SR algorithms, covering traditional methods. Recently, CNN based deep learning algorithms also showed superior performances. In the following, works based on these different methods are reviewed.

Among the earliest algorithms are the non-learning based methods, which assume that the mapping from LR to HR images is predefined, so a certain fixed mathematical formula is usually adopted to represent the mapping. Typical algorithms are interpolation-based methods, including bilinear, bicubic, and Lanczos filtering [67], that can interpolate the HR images with weighted combination of existing pixels from the LR images. However, these methods fails to recover fine details such as texture and gradient information, only leading to overly smooth HR images. Irani and Peleg proposed a method, that starts with a user-defined downsampling model and an initial guess of the HR image, and iteratively update the HR images to approximate the LR image [126]. As a known downsampling model is used, this method preserves more high contrast edge information than normal interpolation-based methods do. Generally the non-learning based methods assume a fixed mapping from LR to HR, so it cannot recover sufficient details in HR images.

Compared with non-learning based methods, learning based methods that learn the HR reconstruction prior or the direct mapping from the paired LR and HR images are the dominant methods to solve SR problems.

Shan et al. [251] developed a method that simulated a reasonable inversion of a previously

proposed model that maps an HR image to an LR one [19]. A feedback-control loop was used to evaluated the inverse model and provide the hint for model update. This was claimed to be a fast algorithm free from artifacts. In the work of Kim and Kwon [146], kernel ridge regression (KRR) was adopted to learn the LR to HR mapping, aided by sparse coding with matching pursuit algorithms. The sparsity of the gradient distribution has also been used in [294] for regularisation to remove artifacts. In the work of Aly and Dubois [13], total variation was used as the regulariser to reconstruct the objective function of mapping that can be minimised based on level-set method.

Some algorithms are based on learning edge structures for the LR images. Edge information is important for visual perception. Therefore, the edge can be learnt to generate the prior for the HR image reconstruction. Fattal learnt the depth and width features of the edges [72], while the gradient profile was used by Sun et al. [265] and Tai et al. [271]. Although edge based methods lead to high-quality edge structures and less artifacts, it has been reported that these methods are not very effective for other high-frequency (HF) structures like the image texture [297].

More methods were proposed to directly learn the mapping from LR images to HR ones, by training their models on paired LR and HR patches/images. Freeman et al. proposed a modelling method based on Markov network with Bayesian belief propagation [85]. Chang et al. used locally linear embedding to recover local geometry from its neighbours in the feature space [41]. Based on the observation that similar patches recur in both the same and different scales for natural images, Glasner et al. developed an algorithm which recovered detailed information from the patch redundancy from different scales [92]. A similar strategy was adopted in [83], where a local self-similarity assumption was made on natural images. These self-similarity based methods are advantageous on their non-reliance on extensive training on external datasets. Other learning methods includes sparse coding [104, 298, 63, 226, 284], support vector machine (SVM) [200], Gaussian process regression [113], weighted average [266], etc.

As there are numerous previous works related to SR problems, only some representative non-CNN based ones are reviewed briefly above. For more comprehensive reviews, the reader can refer to [198, 297].

CNN Based Spatial SR Techniques

In recent years with the advancement of hardware as well as the algorithms, CNN models have been deployed to solve image analysis problems like classification, segmentation, and SR. In SR the CNN based methods also demonstrate superior performance.

There are previous ANN based methods proposed for image restoration, a problem similar to SR. Most of these algorithms target denoising/deblurring. For example, NNs with FC layers networks have been used for the purpose of denoising in [34, 247]; CNNs were also applied for denoising the natural images [129] and dirt/rain removal [70]. However, these algorithms are not specially designed as an end-to-end style, and normally require complex training procedure. Since the end-to-end CNN models are more convenient to apply and allow a simultaneous optimisation of the model parameters during training, these structures are much more preferred in algorithmic development. In the following several representative end-to-end CNN models for SR problems are listed and reviewed.

To the author's knowledge, the first end-to-end CNN model used for SR problems is the SRCNN model proposed by Dong et al. [62]. In this work, the authors divided the SR workflow into three steps: 1) patch extraction that extracts feature vectors out of patches; 2) non-linear mapping that maps the extracted features to other high-dimensional vectors corresponding to the HR patches; 3) reconstruction that aggregates patch-wise feature vectors into a HR image. A 3-layer CNN model that simulates a sparse coding procedure has been developed (Fig. 6.4) to refine the details in the upscaled image with bicubic interpolation. This model can be seen as a milestone, since similar network structures have been adopted in the following CNN based HR image prediction models. Simply structured as it is, SRCNN has a relatively shallow structure and small receptive field (13×13), hindering its HR image reconstruction accuracy. Furthermore, the authors claimed that "the deeper the better" principle is not suitable for network design in SR problems, since gradient vanishing/ explosion can happen very easily for deeper networks. This might be caused by non-existence of FC/pooling layer and the sensitivity of parameter initialisation in the Conv layers.

Kim et al. proposed a deeper recursive neural network (RNN) with skip connection (DRCNN) to increase the receptive field to (41×41) [145]. There are two main differences from SRCNN: 1) in the non-linear mapping step a same convolution is applied for 16 times recursively via a single

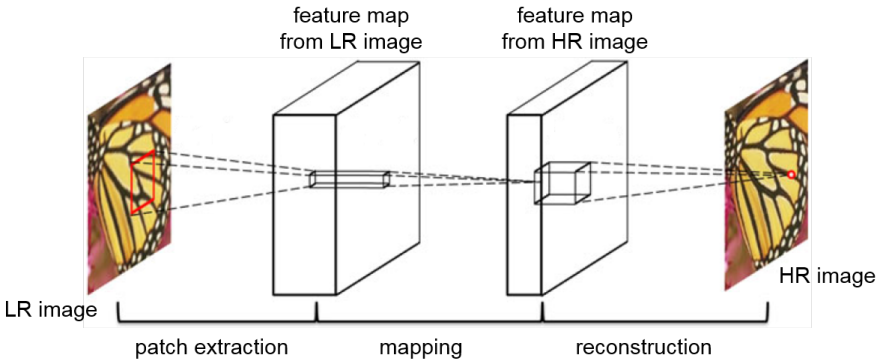


Figure 6.4: The schematic of the SRCNN model. The first layer extracts features from the input LR image; the second layer maps those features into HR; Finally the feature map is merged along the channel dimension to generate the output HR image. Figure reproduced from [62] (copyright ©2016 IEEE).

recursive layer to increase the receptive field. Moreover, inspired by the ResNet [114], the skip-connection is realised by merging the outputs from all convolutions in the final reconstruction layer. With this model, the authors demonstrated a significant improvement to SRCNN on the Timofte dataset [298], while still kept a small number of parameters for optimisation.

In both SRCNN and DRCNN, before being fed into the network, the input LR image needs to be upscaled to the size of the desired output HR image via bicubic interpolation. Therefore, all the computation occurs at the HR scale, hindering their processing speed. In order to realise real-time SR, efficient sub-pixel convolutional neural network (ESPCN) was proposed by Shi et al. [254]. ESPCN processes the input LR image with consecutive Conv layers in LR scale, and upscale the feature map in the last layer. In the last upscaling step, a sub-pixel convolution was applied, which is in fact a reshaping procedure rather than real "convolution". In this way, the computation mainly takes place in the previous layers in the LR scale. A schematic of ESPCN can be found in Fig. 6.5.

In CVPR 2016, besides DRCNN, Kim et al. proposed another CNN based model deep structure to address the SR problems [144]. A very elegant solution has been found to deal with problems of prone-gradient vanishing/explosion in deep neural networks (DNNs). The key factor is the application of residual blocks that enables the application of DNNs [114]. With a skip-connection, the VGG layers were utilised to enhance the feature extraction and non-linear mapping. This model has also been validated on SR problems with different scales, demonstrating its robustness. The same strategy has been adopted by Okay et al. to solve MRI

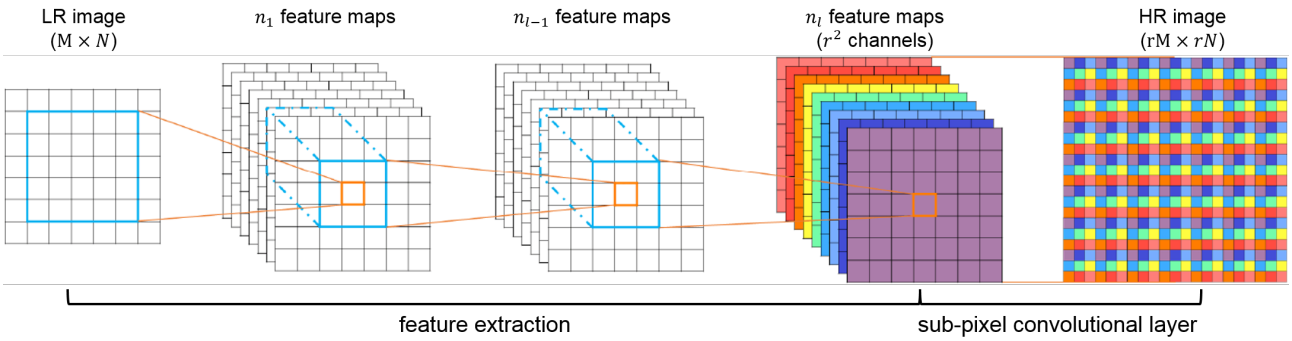


Figure 6.5: The schematic of the ESPCN model. The first stage extracts features from the input LR image and maps the feature maps to different dimensions; the second layer reshapes the $M \times N \times r^2$ feature map into the output scale ($rM \times rN$). Figure reproduced from [254] (copyright ©2016 IEEE).

cardiac image SR problems [204].

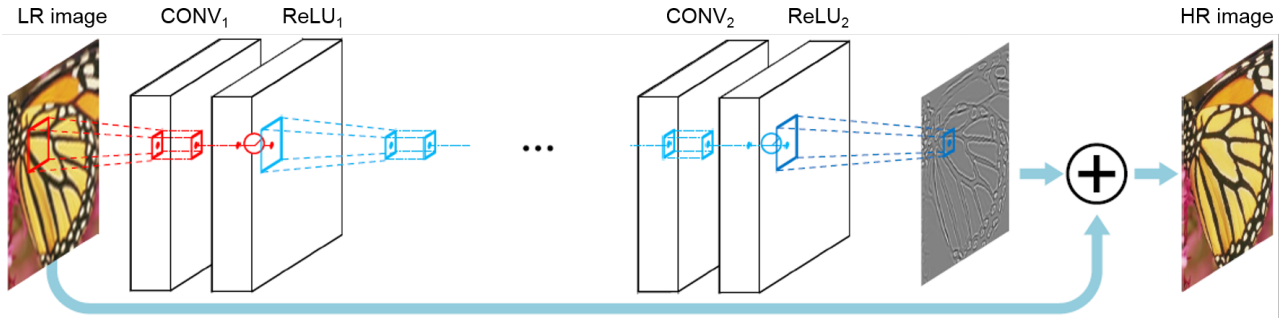


Figure 6.6: The schematic of the VDSR model. The Conv layers from the VGG-net is adopted, aided by a skip-connection to avoid gradient vanishing/ explosion. Figure reproduced from [144] (copyright ©2016 IEEE).

Besides the layer and connection arrangement, the choice of loss function is also a key factor of CNN design. In the models presented above [62, 145, 254, 144, 204], pixel-wise loss functions like $L1$ loss, $L2$ loss, or mean square error (MSE) have been deployed for network training. However, it was found that using pixel-wise loss functions leads to overly-smooth prediction since pixel-wise averaging of plausible solutions is encouraged [160, 198, 133, 64].

In image generation problems, the generated image is expected to be close to the reference image in many aspects like low/ high level features and overall styles. Therefore, instead of pixel-wise loss functions, using another classifier to identify the correctness of the generated image would be a more reasonable choice. Based on this, generative adversarial networks (GANs) have been adopted for the purpose of image generation. The detailed explanations and examples of GANs are not provided in this thesis, but can be found in [97, 227]. Here only

a brief introduction to GANs is given. A typical GAN consists of two networks, the generator (G) and the discriminator (D). While G aims to generate images with a random input, D aims to distinguish between the output of the G and the real reference image. G and D are trained iteratively until the Nash equilibrium is reached. The training of GANs is relatively unstable, but many methods have been proposed to stabilise the training procedure of GAN, such as deep convolutional generative adversarial networks (DCGAN) [227] and Wasserstein generative adversarial networks (WGAN) [17].

The super-resolution generative adversarial network (SRGAN) is a model proposed to solve the SR problems with GANs [160]. It was reported as the first CNN based model that is capable of dealing with a upscaling factors of 4. The loss function of the model consists of a content loss and an adversarial loss. Compared with other models such as ESPCN, DPCN, or SRCNN, although the resulting pixel-wise indicators such as peak-to-signal-ratio (PSNR) and structural similarity (SSIM) index were not the highest, evaluation with mean opinion score (MOS) showed that SRGAN led to more "photo-realistic" results. The model structure is provided in Fig. 6.7.

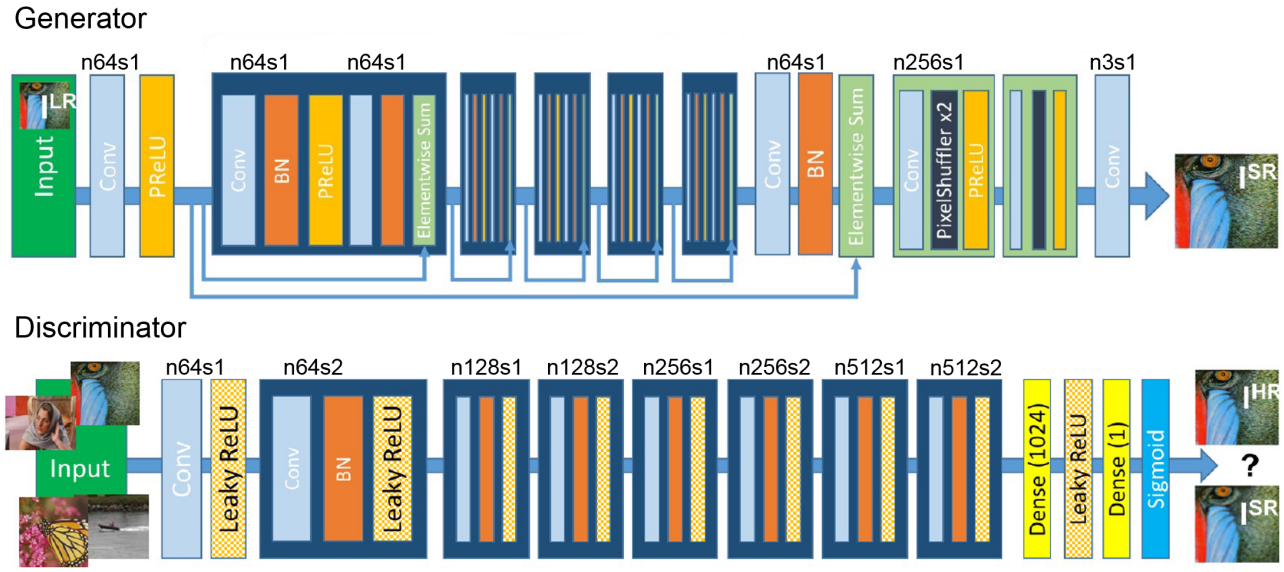


Figure 6.7: The schematic of the SRGAN model. The model contains a generator that predicts the HR image, and a discriminator that distinguish between the generated and the reference image. Figure source: [160] (copyright ©2017 IEEE).

The above five representative works on CNN based model for SR generally lead to better performance compared with traditional algorithms. The network structures also provide us with hints about the model design for our SR problems. However, it is also noticeable that

none of the above CNN based models is the "gold standard", including the latest SRGAN. The model design should still be task specific and finely adjusted accordingly.

Spectral SR Techniques

As is analysed in section 2.3.2, due to the trade-off between spatial/spectral resolution, acquisition time, and cost, estimating hypercubes from normal RGB images algorithmically provides an alternative for constructing dense hypercubes.

Early works for recovering hypercubes from RGB counterparts requires two inputs: an HR RGB image and an LR hypercube. Based on matrix factorisation a method has been proposed to estimate the hypercube [140]. Despite of high estimation accuracy, this method suffers from heavy computational expense and has been reported to cost hours on one image. By replacing the matrix factorisation with simpler propagation, a hybrid system enabling several frames per second computation has been proposed by Cao et al. [36].

More recent studies focus on methods that recover hypercubes from RGB images only. A system aided by time-multiplexed illumination has been proposed by Goel et al. [94], which managed to estimate scene reflectance along a number of wavelengths. This system enabled 17 spectral band hypercube generation at 9 FPS. With a similar system, a 31 spectral band hypercube generation was realised using 5 narrow-band LED light sources [209]. However, these systems depend on controlled lighting, and artifacts could be introduced when imaging moving objects, hindering their real-world applications especially the outdoor usage, and surgical applications.

Arad and Ben-Shahar proposed a sparse coding based method to predict the hypercubes [15]. In this work the hyperspectral prior was first learnt from different hypercube datasets, to build a hyperspectral dictionary consisted of non-orthogonal vectors as the basis spectra. This was followed by building an RGB dictionary by simply computing the projections of hyperspectral dictionary into the RGB space. Next, using the orthogonal match pursuit (OMP) [212] algorithm, a linear combination of RGB basis can be found to approximate the input RGB image, described by weights. Then the same weights can be applied to linearly combine basis spectra in the hyperspectral dictionary to finally reconstruct the corresponding hypercube.

Rather than recovering hypercubes from RGB images, studies on recovering HR hypercubes from LR hypercubes and HR panchromatic (PAN) images drew more attention, namely pan-

sharpening, especially in the field of remote sensing [171]. PAN images refer to grey-level images with information from a wide range of wavelengths, and contains more spatial details. Non-CNN based pan-sharpening can be divided into five categories [171]: component substitution [255, 154, 5], multiresolution analysis [170, 12], hybrid methods [139, 164], Bayesian methods [286, 287], and matrix factorisation [301]. More recently, CNN models have also been adopted for pan-sharpening [308, 187]. Since these works adopted similar network structures or design strategies from the spatial SR CNN models, a detailed review is not provided in this thesis.

6.3.2 Super-Spectral-Resolution

SR, which aims to recover HR images from their LR counterparts, is an ill-posed method, where one LR input can be mapped to multiple possible HR outputs. To deal with this ill-posed problem, two assumptions were made in this work: 1) The HR information is redundant in the HR images and could be partially extracted from the LR ones. 2) The mapping from LR to HR can be learnt from training sets containing data similar to the unseen data. Based on the fact that the captured RGB images are spatially dense while spectral signals are spatially sparse, the hypercube should be predicted mainly based on the RGB images, and the sparse spectral signals can be used to refine the prediction results. In this work, a CNN model has been adopted to predict the dense hypercube. The workflow consists of two steps: hypercube estimation using an RGB image, and hypercube refinement by integrating sparse spectral signals.

Recover Hypercubes from RGB Images

An RGB image was considered as a hypercube with 3 spectral channels. The proposed model, namely model 1, looked for a mapping from an $M \times N \times 3$ hypercube to an $M \times N \times 24$ hypercube, where $M \times N$ stands for the image spatial resolution. This CNN model consists of two main stages:

1. Upscaling the input RGB image along the spectral dimension. This was realised by three 3D deconvolution layers (transposed convolutional layers) piled up to map a $M \times N \times 3$ matrix to a $M \times N \times 24$ one.

2. High-frequency-signal-extraction (HFE). This extracts and combines the high frequency signal with LR stacks. This stage was implemented using a residual block which introduces a "shortcut" to reduce the degradation of the training accuracy problem when deep networks are used. It has been proven that this structure simplifies the optimization. The output of the shortcut is combined with the output of three Conv layers, by element-wise add operation. with this structure, we expected a mergence of the low-frequency and high-frequency data to refine the final predicted hypercube.

The structure of model 1 can be found in Fig. 6.8. This mapping achieved generally good spectral prediction but still with inevitable noticeable errors, since the problem is highly ill-posed. To refine the predication, we extended model 1 to incorporate spatially sparse HSI signals captured using the system′s HSI mode.

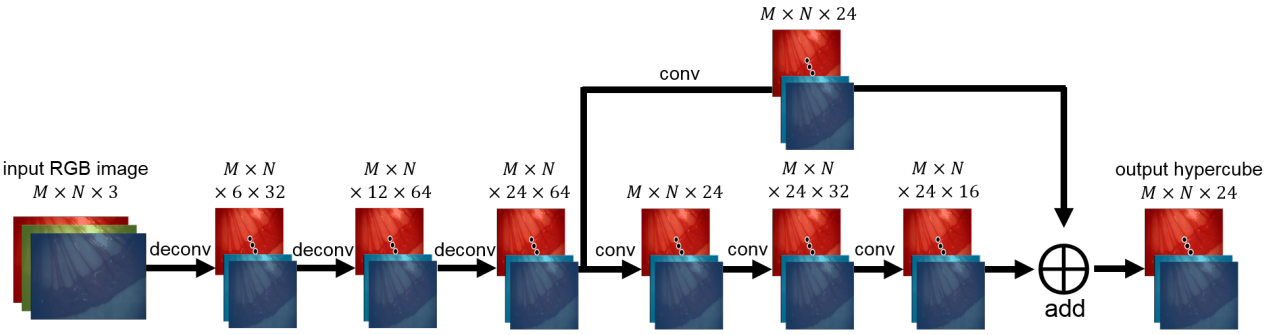


Figure 6.8: The schematic of the CNN model that predicts a hypercube from an input RGB image. The model contains two steps: 1) spectrally upscaling the RGB image; 2) Extraction of high-frequency signals and mergence with the low frequency ones. Figure reproduced from: [166].

Refine Hypercube Prediction by Integrating the Sparse Hyperspectral Signals

RGB images provided high spatial but low spectral resolution; while HSI mode had low spatial but high spectral resolution. Due to the overly sparse hyperspectral signal, RGB was used as the main contributor for hypercube estimation, then the HSI signal was applied to refine the estimation.

Model 2 took three inputs: an RGB image ($M \times N \times 3$), a density map ($M \times N$) indicating the locations where HSI took place, and a sparse stack ($M \times N \times 24$) containing the sparse HSI signals. In the density map, the background was set to 0, while the locations where the

hyperspectral signals come from were simulated with individual Gaussian distributions with maximum value as 1. The sparse stack was built from a $M \times N \times 24$ matrix with all background pixels as 0, and filled with hyperspectral signals according to their corresponding locations.

This model was extended from Model 1, by adding a merging stage on top of it, where all inputs were integrated (Fig 6.9). Popular methods that merge inputs from different sources in CNN models, include element-wise add and concatenation along the feature channel followed by convolution. In this work, the latter way was adopted due to better outcome in practice. The sparse HSI data was concatenated with the element-wise product between the density map and the HSI stack recovered from RGB. The final spatially dense hypercube was estimated following a convolution.

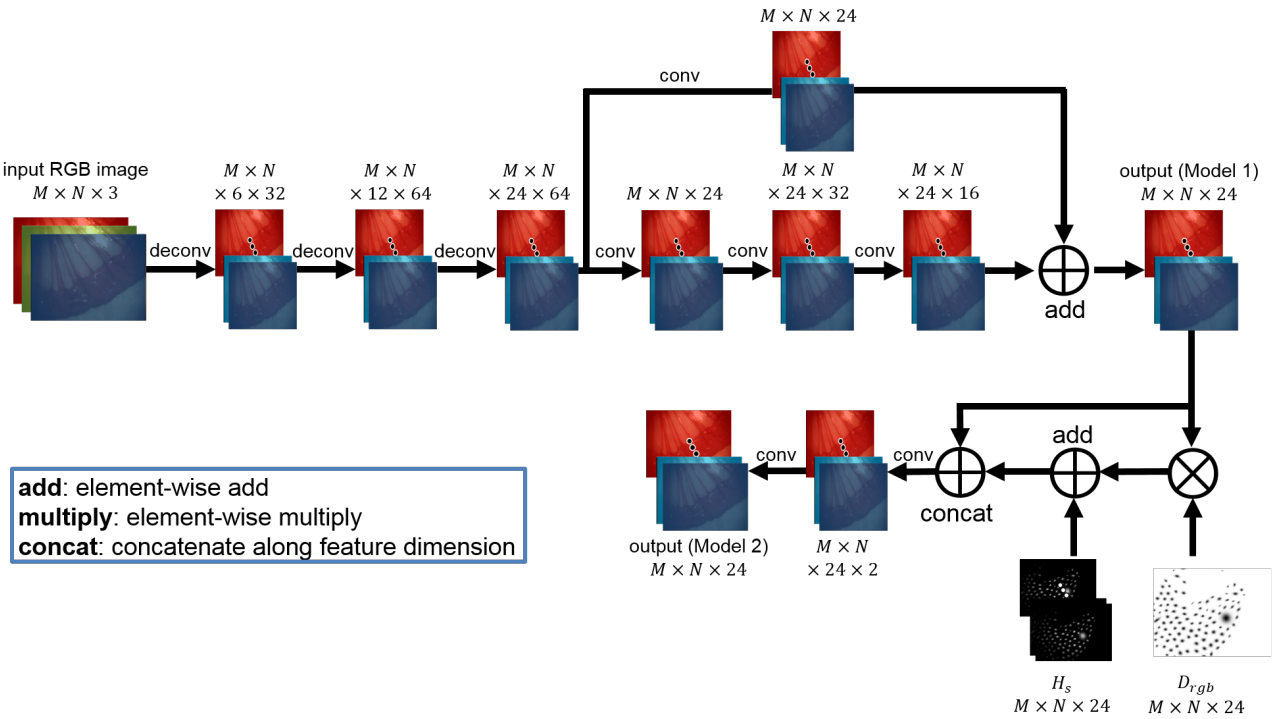


Figure 6.9: The schematic of the CNN model that predicts a hypercube from an input RGB image and sparse hyperspectral signals. This model can be seen as Model 1 with an extra mergence phase that integrates the hyperspectral signals. Figure reproduced from: [166].

Training

Choosing a training set that provides adequate prior knowledge is of great importance for accurate MSI recovery. In this work, hypercubes collected *in vivo* during animal trials have been used for training and testing. These hypercubes were collected from porcine bowel, rabbit

uterus, and sheep uterus, using an LCTF endoscopic imager [50]. The mis-alignment of images in adjacent wavelengths due to tissue movement has been compensated with the registration method proposed by Du et al. [66]. In order to prepare the RGB images for training, the transmission spectrum (h) for an RGB camera (Thorlabs DCU223C) was utilised to generate the synthetic RGB images R from HSI stacks, with $R = h * H$. The density map D_{hsi} for the sparse HSI signal was produced using previous spot segmentation results, with Gaussian distribution ($max = 1$) filled at each spot location (where the sparse spectral signals came from). The density map for the RGB image D_{rgb} was defined by $D_{rgb} = 1 - D_{hsi}$. The sparse HSI stack (H_s) was the element-wise product between the density map and the HSI stack ($H_s = D_{hsi.} * H$).

To guarantee sufficient training samples, Model 1 was trained on individual pixel spectral vectors instead of whole hypercubes, based on the assumption that adjacent pixels do not affect the spectra prediction. In this case convolutions were applied along the spectral dimension, so that the trained network can be directly applied to inputs with arbitrary spatial dimensions, since no convolution on the spatial dimensions was adopted in the model. When training Model 2, the network was initialised by the trained parameters from Model 1. The training set for Model 2 was prepared by data augmentation including flipping and rotation to increase the size as well as the generalisation. A two-stage training strategy was adopted instead of training directly from scratch: the parameters in the shared layers with Model 1 were frozen while the "mergence" layers were updated; then all the parameters were updated until convergence. Both models were trained using the Adam optimizer and $L2$-norm loss function. In prediction, RGB images were captured by the same CCD camera used for generating synthetic RGB images, while the sparse HSI signals and the consequent density map came from the slit hyperspectral imager. Training and prediction were implemented using Tensorflow [178]. Using the proposed CNN model, pixel-wise hypercubes can be estimated at ≈ 8 FPS on a PC (OS: Ubuntu 14.04; CPU: i7 – 3770; GPU: NVIDA GTX TITAN X).

6.4 Experimental Results

In this section validation on pantom, *ex vivo*, and *in vivo* tissue experiments has been carried out to validate the HSI function of the ICL SLHSI system, in two aspects: 1) the HSI acquisition;

2) the performance of the hypercube prediction algorithm.

6.4.1 HSI Acquisition

For quantitative evaluation of the HSI accuracy, a Macbeth colour chart, whose spectra was known, was used. By imaging the Macbeth chart, reflectance spectra from different regions could be compared with those previously measured using a spectrometer (USB4000HR; Ocean Optics, Inc., USA). The spot locations where the hyperspectral signals originated were determined by the SL image acquired from the corresponding frame in the SL mode (Fig. 6.10(a)). The comparison was plotted in Fig. 6.10(b). It can be found that in each panel the SLHSI systems mapping procedure returned the correct reflectance spectrum for each location, with a mean spectral error of 10% between the SLHSI and the gold standard.

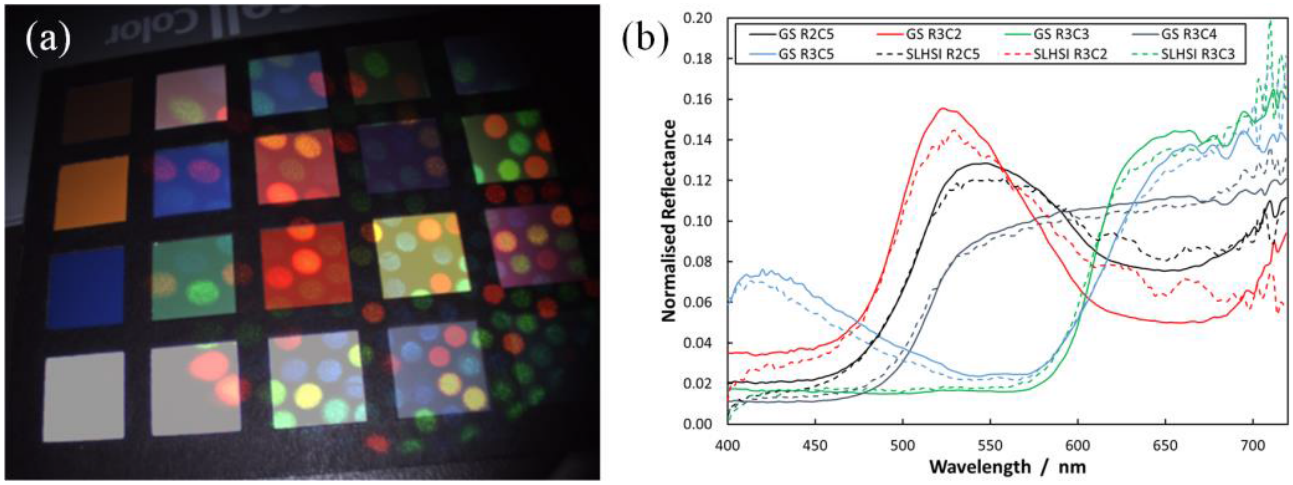


Figure 6.10: (a) Macbeth colour chart illuminated with SL. (b) Reflectance spectra from selected colour panels of the chart, with R and C indicating the row and column, respectively. Data from the hybrid system (SLHSI) are plotted alongside the high resolution spectrometer results (GS). Figure source: [167].

To demonstrate combined shape and spectral information acquisition with the ICL SLHSI system, a cylindrical object was imaged (Fig. 6.11(a)). The surface of this object has three colours, red, green, and blue. Mean reflectance curves for each region show peaks in the corresponding wavelengths of red, green and blue (Fig. 6.11(b)). The captured hyperspectral signals were convolved with a colour CCD's spectral sensitivity, simulating the CCD behaviour in recovering RGB data. The recovered colour was displayed on the surface mesh that was reconstructed in the SL mode (Fig. 6.11(c)). This experiment intuitively demonstrated the capability of the ICL SLHSI system in object surface reconstruction and HSI.

Reflectance spectra from a murine abdomen, imaged post mortem, were used to calculate the absorbance spectrum for each spot location and extract the tissue oxygen saturation (StO_2) using the method described in section 6.2.2. StO_2 at various spot locations is shown in Fig. 6.11(d), along with an absorbance spectrum from one region (Fig. 6.11(e)). Spectra that did not match the model well ($r^2 < 0.8$ in linear fitting) were rejected. Mean StO_2 in the abdomen is low ($14 \pm 10\%$), as expected for an *ex vivo* experiment.

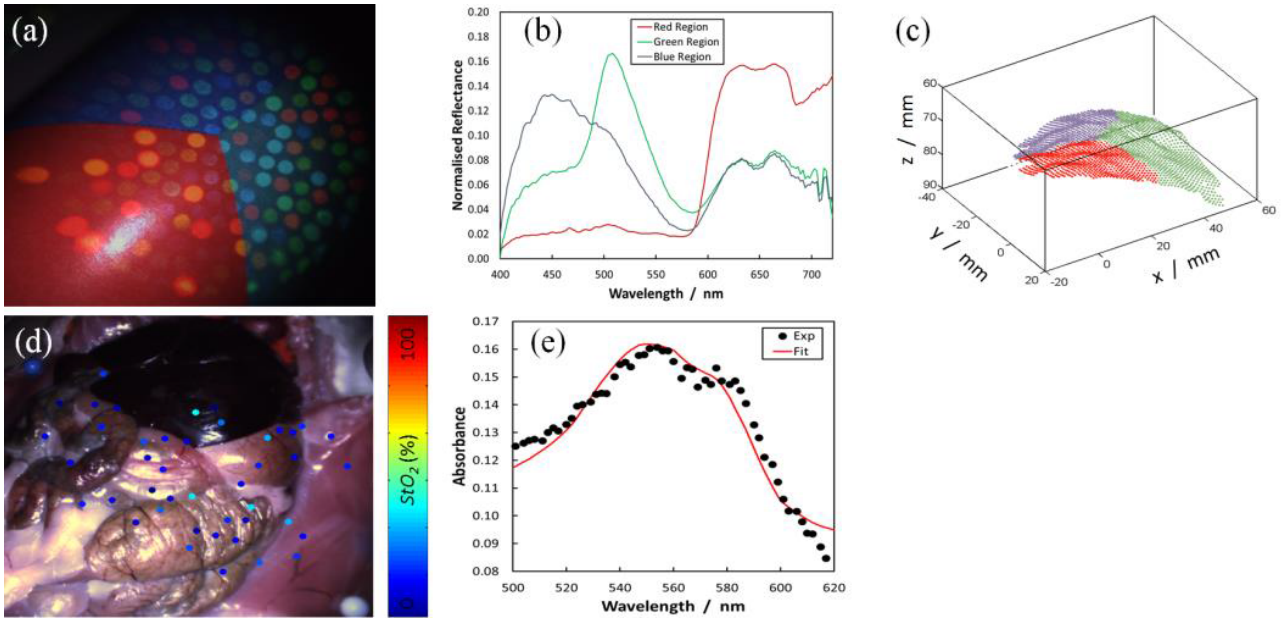


Figure 6.11: (a) Three-coloured cylindrical target with SL. (b) Mean reflectance spectra for each of the targets regions. (c) 3D reconstruction coloured by RGB values generated from spectra in (b). (d) Murine abdomen with StO2 overlay. (e) Tissue absorbance spectrum and model fit. Figure source: [167].

According to the experiments, this system is able to measure sample reflectance spectra accurately. This conclusion should apply to other objects including *in vivo* tissue, since the spectra accuracy is not affected by the target object. The main differences to be expected *in vivo* are higher StO_2 values and breathing/peristalsis-related movement. The difference in StO_2 will only affect the measured spectral shape. Tissue motion will not introduce much spectral artefact as each fibres signal is acquired in a single snapshot (100 ms).

6.4.2 Dense Hypercube Estimation

In vivo animal experimental data (hypercubes from 50 pig bowel, 21 rabbit uterus, 10 sheep uterus) were used to train and validate both Model 1 and Model 2. By mixing the data from

different sources and data augmentation, a 5-fold leave one-out cross-validation (LOOCV) was applied on a dataset containing 243 hypercubes; each fold contained 200 hypercubes for training and the remaining for testing.

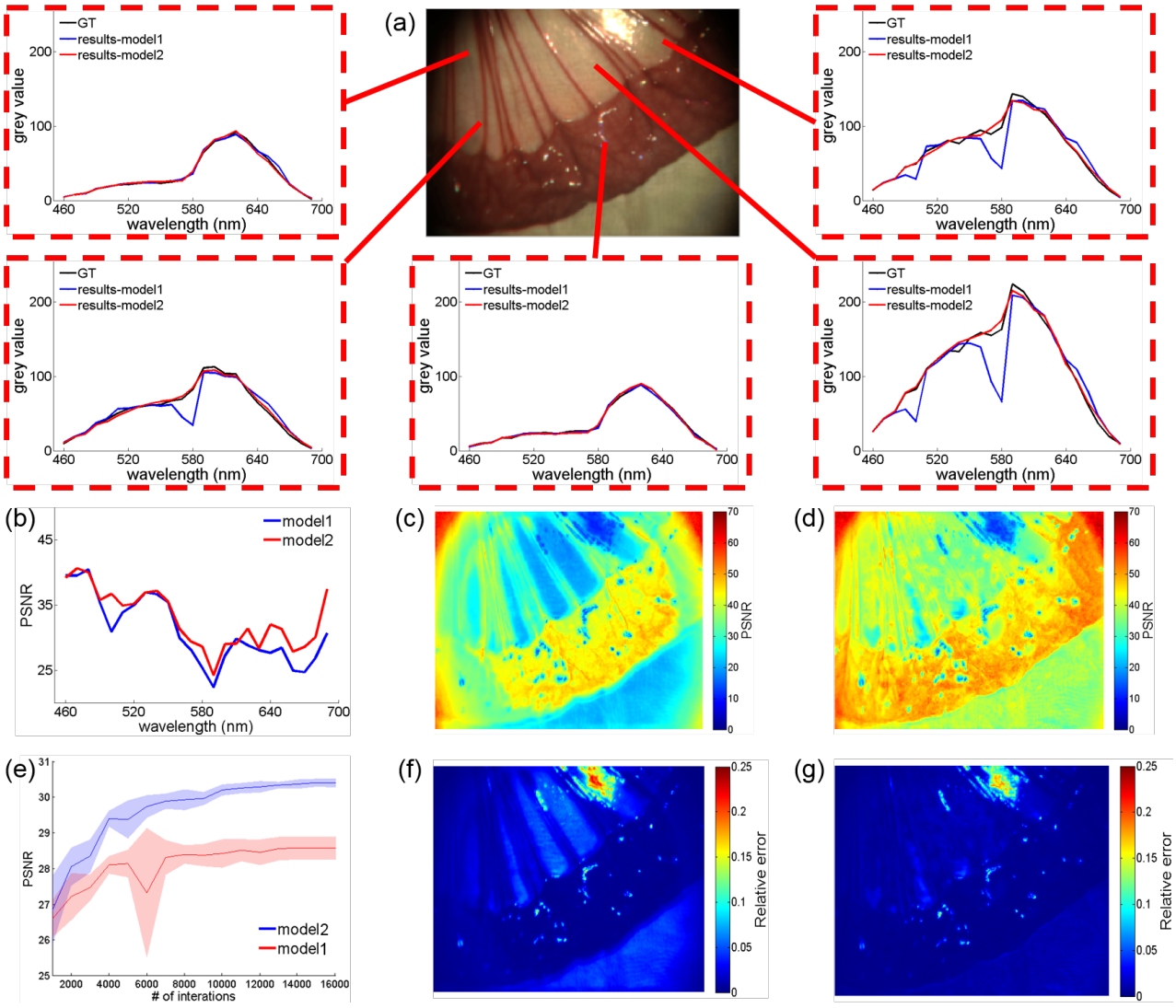


Figure 6.12: (a) The ground truth spectra (black), estimated spectra using method 1 (blue) and method 2 (red), on five representative locations from a porcine sample image. (b) comparison (in PSNR) between the performances of model 1 (blue) and model 2 (red) along different wavelengths in LOOCV. PSNR map of the prediction results for model 1 (c) and model 2 (d) regarding the same sample (a). (b) PSNR change along number of training iterations (model 1: red; model 2: blue) in LOOCV. The average relative error on different wavelengths regarding (a) are shown in (f) and (g). Figure reproduced from: [166].

Given the ground truth the peak signal-to-noise ratio ($PSNR = 20\log_{10}(\frac{255}{MSE})$, where MSE stands for mean square error) was adopted to evaluate the performances of both models. Fig. 6.12(e) demonstrated the PSNR change of the estimated hypercube on the test set with number of iterations. Model 2 demonstrated a higher PSNR compared to Model 1 throughout the whole

training procedure. In the validation on average model 2 also demonstrated higher PSNR ($\approx$ 30.4) compared with model 1 ($\approx$ 28.5). The average PSNR on different wavelengths are shown in Fig. 6.12(b).

In order to intuitively show the difference between two models, the estimated multispectral signals from 5 points, randomly chosen from representative areas in one pig bowel image, are compared (Fig. 6.12(a)). Although model 1 provided an estimation that generally fitted the ground truth, it suffered from large errors at some wavelengths. On the contrary model 2 provided improved accuracy over the entire spectral range. The pixel level PSNR maps for the estimated MSI regarding the same image are shown in Fig. 6.12(c)(d). Excluding the saturation area, the minimum and mean PSNR are 13.1 and 34.7 for model 1, and 14.1 and 42.0 for model 2. The relative error maps of estimated hypercubes from Model 1 and Model 2 are also provided in Fig. 6.12(f)(g).

Evaluation of transfer learning results is of great importance on machine learning problems, especially the clinical ones, where high model generalisation capability is required. Thus, we also trained our models on data from different sources and tested them on each other. Fig. 6.13 compares the transfer learning results on two models, showing the importance of integrating the sparse hyperspectral signal for more accurate MSI estimation.

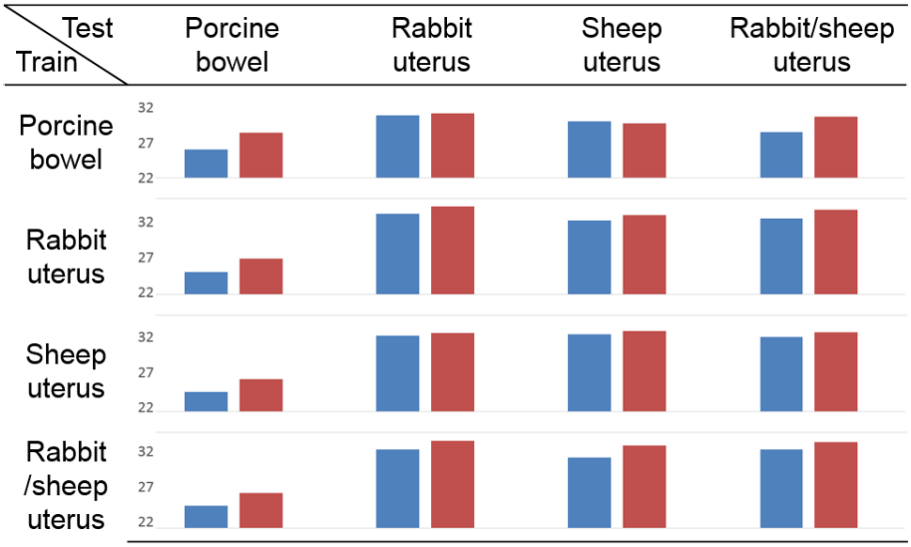


Figure 6.13: The transfer learning results (indicated by PSNR) with model 1 (blue) and model 2 (red). Both models were trained and tested on 4 datasets: pig bowel, rabbit uterus, sheep uterus, and mixed rabbit/sheep uterus. 5-fold cross validation was applied to calculate the PSNR when training and testing on data from the same source.

Acquiring the hypercube is a prerequisite for imaging modalities like oxygen saturation and narrow-band imaging, which could provide information to aid diagnosis and surgical navigation. As image examples, the synthetic oxygen saturation and narrow band images, estimated from the predicted hypercube, are overlaid onto the 3D reconstructed surfaces from *ex vivo* and *in vivo* human experiments (Fig. 6.14). According to Fig. 6.12, the hypercube estimation only from the RGB image suffers from big errors in certain wavelengths ($\approx 570 - 590$ nm), which belongs to the range important for StO_2 calculation. Thus, the StO_2 maps in Fig. 6.14(c)(f) look unrealistic. This point is further discussed in section 6.5.

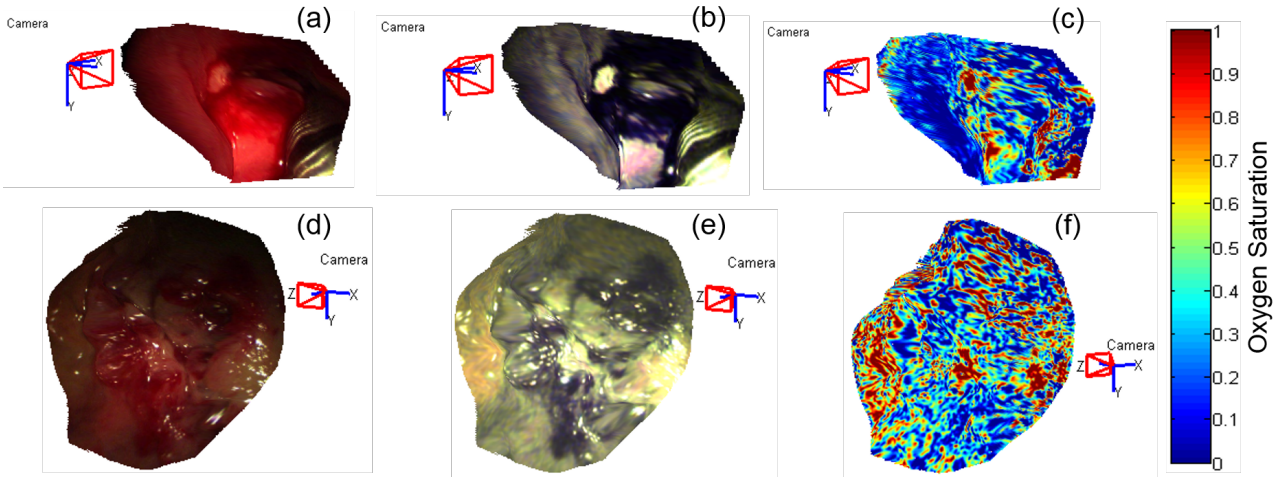


Figure 6.14: Example human (a) in vivo and (d) ex vivo reconstructed tissue surface. (b, e) synthetic narrow band images and (c, f) synthetic oxygen saturation maps overlaid on these surfaces. Figure reproduced from: [166].

6.5 Discussion and Conclusions

A flexible and hybrid system, i.e. the ICL SLHSI system, has been developed for both 3D sensing and HSI, using a motorised flipper mirror to switch between the SL and HSI modes. The probe's size and flexibility mean that it is compatible with standard clinical endoscopic tools for assessing the gastrointestinal tract and abdomen, while its imaging capabilities can enable clinical studies based on previous work in measurement of oxygenation dynamics [50], tissue classification [174] and augmentation of the clinicians view with pre-operative data [182].

Making use of the special fibre distribution at the two ends of the fibre bundle, a slit hyperspectral imager has been integrated into the system to enable hyperspectral signal acquisition,

aided by a motorised flipper and a lens combination. The HSI can be carried out in one exposure, and the FoV covers a reasonably large area of the target object surface. By 1D $\leftrightarrow$ 2D according to the fibre mapping, the locations where the hyperspectral signals come from can be determined using the adjacent SL frame. Therefore, although the signal acquisition is fast, the measurement resolution is limited to the number of fibres in the fibre bundle. The motorised flipper was controlled manually via the LabVIEW API, so the switching between different modes requires manual interaction.

The algorithmic contribution demonstrated in this chapter is the prediction of pixel-wise dense hypercubes using the RGB images and the captured sparse hyperspectral signals, for the purpose of compensating the low spatial resolution of the captured hyperspectral signals using the SLHSI system. In the proposed method, a hypercube was firstly predicted only using the RGB image, then refined by the sparse hyperspectral signals. A CNN model was adopted to enable a fast end-to-end training and prediction ($\approx$ 8 FPS). It was also shown that the proposed method can be potentially used in the generation of narrow-band images or oxygen saturation maps. It is also noticeable that, the estimation result using RGB only leads to obvious errors in certain wavelength bands that covers those important for StO_2 estimation. One of the reasons is the deconvolution layers that upscaled the RGB image into a 24 channel hypercube. With this structure, the spectra prediction in between red and green, green and blue is inaccurate, since the prediction only depends on one of the RGB channels. A significant improvement can be observed after integrating the sparse hyperspectral signals, as is shown in section 6.4.2.

The future work should focus on several aspects:

1. Refinement of the software platform to integrate all procedures of the hybrid imaging into one workflow. Currently, capturing of SL, HSI, and the switching between these two modes, are realised manually. In the future work, all the steps should be combined together with C++ API with a user-friendly interface.

2. Improvements of the super-spectral-resolution algorithm. The current CNN model uses a kernel with size 2 to upscale in the spectral dimension, thus does not use correlations between red, green, and blue channels to predict hypercubes from RGB images, leading to relatively big errors in the overlapping wavelengths between these channels. The mergence manner of two inputs can also be changed to potentially guarantee a better combination

of different data sources. The CNN model will be upgraded to improve the prediction accuracy.

3. System reliability and practicality enhancements will be considered, including compatibility with existing clinical instruments and sterilisability. Pilot studies on freshly excised tissue from human surgical procedures are planned and will allow testing of the system against histology, while further preclinical work will enable evaluation and optimisation of performance *in vivo*.

Chapter 7

Discussion and Conclusions

7.1 Summary of Achievements

Technological advancements have enabled minimally invasive surgery (MIS) with smaller incision, less pain and infection, and faster patient recovery. However, MIS also brings challenges for the surgeons, such as the loss of tactile feedback and direct vision. As the surgeon can only observe the anatomy through endoscopic images, assisting them by providing extra intra-operative information that is usable with available surgical instruments may benefit surgery by improving clinical decision making, pathology diagnosis, and surgical navigation.

The depth of patient tissue anatomy, i.e. the distance between the tip of the endoscope and the surface of the tissue, is important intra-operative information. If recorded across a whole field of view, this depth has the potential to be used in different applications, such as pathology detection (e.g. polyps detection), augmented reality (AR) (by registering pre- and intra-operative data), and active constraint determination in robotic surgery, etc.

Another intra-operative imaging modality carrying rich biomedical information is hyperspectral/ multispectral imaging (HSI/MSI). For convenience, both HSI and MSI are referred to as HSI in the following context. The potential applications include pathology detection, narrow-band image generation, oxygen saturation estimation, etc. Integration of depth measurement and HSI into the endoscopic system may be beneficial and would benefit from further research studies.

The main challenges of integrating depth measurement and HSI, however, include hardware miniaturisation, real-time data acquisition, and robust signal/ image processing. In this work, a system that is capable of sensing shape and HSI, is proposed to address these problems. The 3D reconstruction was based on structured light (SL) techniques, while HSI was realised via integrating a slit hyperspectral imager to the system. This system has been presented from a combined approach considering hardware and software, as well as experimental validation on phantoms, *ex vivo*, and *in vivo* tissues. A summary of the achievements and contributions described in each chapter are given in the following sections.

Chapter 2 provided an overview on the three main research topics developed, namely 3D reconstruction techniques, HSI systems, and neural networks. The reviewed 3D reconstruction techniques included stereoscopy, monocular Structure-from-Motion (SfM), Shape-from-Shading (SfS), Time-of-Flight (ToF), and structured light (SL). It was concluded that SL has an advantage in its non-reliance on tissue surface texture. Detailed examples of SL systems used in MIS showed the potential of adapting SL to normal endoscopic systems. Then a concise overview was given on HSI systems in terms of the acquisition modes, dispersive devices, and medical applications. The hyperspectral signal contains rich information on tissue, cellular, and molecular level, and so can be used in disease diagnosis and surgical guidance. It was also shown that, in MIS applications, generally a HSI system that can be miniaturised and enables snapshot based hyperspectral signal acquisition is preferred. Last but not least, chapter 2 provided a brief introduction to neural networks, from basic neural, commonly used fully-connected (FC) and convolutional (Conv) layers, to global parameters in network training. This provided a basis for the understanding of convolutional neural networks (CNNs) based algorithms proposed in section 5.3 and section 6.3.

In chapter 3, the ICL SL optical hardware system was firstly demonstrated, followed by a description of a naive 3D reconstruction pipeline. Due to its miniaturised size (2.0 mm diameter), the probe of the SL system could be conveniently integrated into an endoscopic system, either by being inserted through the biopsy channel or being fixed beside the endoscope tip. The proposed naive reconstruction pipeline described an initial version of the algorithmic workflow that managed to reconstruct the tissue surface shape from the captured SL images. This pipeline consisted of system calibration, pattern decoding, and triangulation. A system specific calibration method was described, including camera calibration and probe calibration. The

pattern decoding algorithm was based on normalized cut and feature matching with local rigid registration. The step-wise performance of the system was validated on phantoms, *ex vivo*, and *in vivo* experiments, demonstrating the potential of using this system in clinical environments.

In chapter 4, further improvements on the original ICL SL optical system and the naive reconstruction pipeline were developed, including: 1) using a multispectral camera to enhance the light pattern decoding, 2) an on-the-fly self-calibration technique to enable a free-hand manipulation of the endoscope, 3) integrating the local normal information to improve the 3D reconstruction results. These incremental improvements aim to improve the practicability and accuracy of 3D reconstruction using the ICL SL system. Validation has been applied with phantoms, *ex vivo*, and *in vivo* experiments.

In chapter 5, improvements for the ICL SL system were proposed, to enable a faster, denser, and more robust tissue surface reconstruction. A chopper wheel was integrated into the system to stroboscopically switch between SL and WL images. The SL images were further processed for 3D reconstruction, while the WL images were provided to surgeons for visualisation. A 3D printed adaptor was designed to house the probe beside the endoscopic tip with optimised angle and baseline that enabled an accurate triangulation at working distances of $1.5 - 4.0$ cm. The small size of the adaptor (11 mm) enabled the insertion of the device through a trocar that is clinically used in e.g. laryngeal surgery. A new pattern decoding algorithm was implemented with a GPU, for the purpose of fast tissue surface reconstruction. This algorithms consisted of two parts: pattern detection with a FCN style model, and pattern matching aided by a customised feature descriptor and epipolar constraint. Through experiments this proposed algorithm showed a robust and real-time performance ($\approx$ 12 FPS) in 3D reconstruction. A two-frame based SfM method was implemented and combined with SL for surface reconstruction. Since SfM leads to a quasi-dense up-to-scale reconstruction while SL leads to sparse reconstruction, merging recovered surfaces from SfM and SL yielded a quasi-dense surface with known scale.

In chapter 6, the ICL SL system was further extended to enable HSI of the target tissue surface. Together with algorithmic implementations, the resulting ICL SLHSI system is capable of producing pixel-level dense hypercubes. The sparse HSI was enabled by integrating a slit hyperspectral imager with a lens combination. An extra "flipper" mirror was used to switch between the SL and HSI mode. Experiments on a Macbeth colour chart demonstrated the

accuracy of the acquired sparse hyperspectral signals. The algorithmic contribution in this chapter was on the prediction of pixel-wise dense hypercubes using the RGB images and the captured sparse hyperspectral signals, for the purpose of compensating the low spatial resolution of the captured hyperspectral signals using the SLHSI system. In the proposed method, a hypercube was firstly predicted only using the RGB image, then refined by the sparse hyperspectral signals. A CNN model was adopted to enable a fast end-to-end training and prediction on a GPU at about 8 FPS. It was also shown that the proposed method can be potentially used in the generation of narrow-band images or oxygen saturation maps, by experiments on an *in vivo* patient. Last but not least, the "super-spectral-resolution" algorithm was shown to have benefits other general HSI acquisition modalities to greatly reduce scanning time with little compromise in performance.

7.2 Future Research Directions

As this work aimed at providing 3D sensing and HSI in MIS, future research will focus on the improvement of the feasibility and practicality of the ICL SLHSI system in real clinical environments. In the following, some future research directions are listed:

7.2.1 Improvements on the ICL SLHSL System

Although highly promising results have been achieved, there is still space to improve the pattern decoding procedure. As over-exposure occurs during the movement of endoscopes and hinders the acquisition of clear SL images, real-time auto-exposure algorithms need to be developed to guarantee high quality images for further processing. For the pattern decoding algorithm several options are also available to enhance this procedure, e.g. more SL images from different tissue surfaces should be collected to improve the generalisation of the training set; the detection can be upgraded to state-of-the-art CNN algorithms; the recent research on CNN based stereo matching or optical flow estimation also give us some hints to design a model that is capable of end-to-end training and pattern decoding.

The proposed two-view based SfM algorithm also requires improvement, since it is not a state-of-the-art algorithm. Monocular SfM should be upgraded to e.g. deformable SfM, or state-

of-the-art SLAM such as ORB-SLAM which has been developed to handle many practical difficulties. The fusion of SL surface and SfM surface still needs to be extended to address more difficult situations, like reconstruction errors that could occur in both SL and SfM, the optimisation of point cloud fusion, etc.

The CNN model for dense hypercube prediction can be enhanced. In the current CNN model used to predict the hypercubes from RGB images, correlations between red, green, and blue channels are not used, leading to relatively big errors in the overlapping area between these channels. The deconvolution layers will be re-designed to improve the prediction accuracy. In the experimental results, relatively inaccurate prediction in the $580-600$ nm wavelength range has also been observed, due to the unexpected spectrum fluctuation in this range. This could be potentially alleviated by increasing the spectral resolution of the ground truth/predicted hypercubes, increasing the number of training sets, or assigning higher weights for these bands during training.

One of the potential clinical application of this system is the estimation of StO_2 maps. However, in this work, the StO_2 was predicted using linear regression on the estimated hypercubes, not directly from the acquired RGB and sparse hyperspectral signals. This "indirect" approach inevitably introduced errors as recovering hypercubes is an ill-posed problem. Therefore, techniques predicting StO_2 directly from the acquired information by the ICL SLHSI system would be more straight-forward and less erroneous. GAN is one of the potential approaches.

The ICL SLHSI optical system has been validated via different experiments including an *in vivo* patient experiment. However, the system is still at a proof-of-concept stage, and requires manual interaction during the experiments. Currently, the SL mode, the HSI mode, and the switching between these two modes, are controlled via different software platforms, including the manufacturer-provided control software and the LabVIEW programs. In order to enable a practical system that can be used in clinical environments, fully automatic operation of the system with minimum manual interaction is required. Therefore, one of the important following works would be the integration of all imaging modes as well as the corresponding data processing into one workflow using the C++ API.

7.2.2 Further System Design and Development

The ICL system depends on using a supercontinuum laser as the light source for structured lighting. However, this laser is expensive and emits laser light with high energy that might harm the eyes if not correctly enclosed. Therefore, alternatives for the purpose of a safer and more cost-effective optical system are desirable.

One of the alternatives is to replace the supercontinuum laser with multiple laser diodes with different wavelengths. Multiple lasers can be dispersed by devices, e.g. prisms, and then fed into the linear array of the fibre bundle simultaneously to create randomly distributed spot with distinctive colour and neighbourhood information to facilitate structured light. A proof-of-concept of this device was developed in a Masters project running in parallel with this PhD.

Another possible solution would be creating the artificial light pattern with diffractive optical elements (DOEs). Using a laser diode as the light source, light can be delivered through a single-mode fibre and diffracted by the DOE, projecting certain patterns on the tissue surface. The main benefit of this design is that the projector can be miniaturised since a fibre is used to deliver the light. However, the main challenge would be the DOE design, since creating a unique identifiable pattern is extremely important, especially when only one or two laser diodes are integratable to the system as light sources.

Besides the system development, more experiments, especially *in vivo* experiments need to be carried out, not only to objectively evaluate the system, but also to provide with CNN based algorithms with more generalised training set.

7.3 Conclusion

To summarise, in this thesis, a hybrid optical system has been proposed for both tissue surface 3D sensing and hyperspectral imaging. A detailed description of the system has been provided, arranged according to the chronological order of the system development. The miniaturised shape of the probe enabled the system to be coupled with a normal endoscope system, and the algorithmic implementation enabled a real-time quasi-dense tissue surface reconstruction and a fast pixel-level dense hypercube acquisition. Validation using phantoms, *ex vivo*, and *in vivo*

tissue has proved the potential of using the ICL SLHSI system in clinical applications.

Further improvements on hardware and algorithms mentioned in section 7.2, as well as more experimental validation is also planned to enhance the feasibility and practicality of using this system in real clinical environments.

Bibliography

[1] J. D. Ackerman, K. Keller, and H. Fuchs. "Surface reconstruction of abdominal organs using laparoscopic structured light for augmented reality". In: (2002), pp. 39–46. DOI: 10.1117/12.460179. URL: +http://dx.doi.org/10.1117/12.460179.

[2] E. Adam, O. Mutanga, and D. Rugege. "Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review". In: *Wetlands Ecology and Management* 18.3 (2010), pp. 281–296. ISSN: 1572-9834. DOI: 10.1007/s11273-009-9169-z. URL: https://doi.org/10.1007/s11273-009-9169-z.

[3] A. Adan et al. "3D feature tracking using a dynamic structured light system". In: *Computer and Robot Vision, 2005. Proceedings. The 2nd Canadian Conference on.* 2005, pp. 168 –175. DOI: 10.1109/CRV.2005.1.

[4] M. A. Afromowitz et al. "Multispectral imaging of burn wounds: a new clinical instrument for evaluating burn depth". In: *IEEE Transactions on Biomedical Engineering* 35.10 (1988), pp. 842–850. ISSN: 0018-9294. DOI: 10.1109/10.7291.

[5] B. Aiazzi, S. Baronti, and M. Selva. "Improving Component Substitution Pansharpening Through Multivariate Regression of MS +Pan Data". In: *IEEE Transactions on Geoscience and Remote Sensing* 45.10 (2007), pp. 3230–3239. ISSN: 0196-2892. DOI: 10.1109/TGRS.2007.901007.

[6] B. Aiazzi et al. "Noise modelling and estimation of hyperspectral data from airborne imaging spectrometers". In: *Annals of Geophysics* 49.1 (2006).

[7] M. Aikio and V. T. Tutkimuskeskus. *Hyperspectral prism-grating-prism imaging spectrograph.* Technical Research Centre of Finland, 2001.

[8] H. Akbari et al. "Blood vessel detection and artery-vein differentiation using hyperspectral imaging". In: *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society.* 2009, pp. 1461–1464. DOI: 10.1109/IEMBS.2009.5332920.

[9] H. Akbari et al. "Cancer detection using infrared hyperspectral imaging". In: *Cancer Science* 102.4 (2011), pp. 852–857. ISSN: 1349-7006. DOI: 10.1111/j.1349-7006.2011.01849.x. URL: http://dx.doi.org/10.1111/j.1349-7006.2011.01849.x.

[10] H. Akbari et al. “Detection and Analysis of the Intestinal Ischemia Using Visible and Invisible Hyperspectral Imaging”. In: *IEEE Transactions on Biomedical Engineering* 57.8 (2010), pp. 2011–2017. ISSN: 0018-9294. DOI: 10.1109/TBME.2010.2049110.

[11] P. Alcantarilla, A. Bartoli, and A. Davison. “KAZE features”. In: *Computer Vision – ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VI*. Ed. by A. Fitzgibbon et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 214–227. ISBN: 978-3-642-33783-3. DOI: 10.1007/978-3-642-33783-3_16. URL: https://doi.org/10.1007/978-3-642-33783-3_16.

[12] L. Alparone et al. “Sharpening of very high resolution images with spectral distortion minimization”. In: *IGARSS 2003. 2003 IEEE International Geoscience and Remote Sensing Symposium. Proceedings (IEEE Cat. No.03CH37477)*. Vol. 1. 2003, 458–460 vol.1. DOI: 10.1109/IGARSS.2003.1293808.

[13] H. A. Aly and E. Dubois. “Image up-sampling using total-variation regularization with a new observation model”. In: *IEEE Transactions on Image Processing* 14.10 (2005), pp. 1647–1659. ISSN: 1057-7149. DOI: 10.1109/TIP.2005.851684.

[14] I. Amidror. “Scattered data interpolation methods for electronic imaging systems: a survey”. In: *Journal of Electronic Imaging* 11 (2002), pp. 11 –11 –20. DOI: 10.1117/1.1455013. URL: http://dx.doi.org/10.1117/1.1455013.

[15] B. Arad and O. Ben-Shahar. “Sparse Recovery of Hyperspectral Signal from Natural RGB Images”. In: *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII*. Ed. by B. Leibe et al. Cham: Springer International Publishing, 2016, pp. 19–34. ISBN: 978-3-319-46478-7. DOI: 10.1007/978-3-319-46478-7_2. URL: https://doi.org/10.1007/978-3-319-46478-7_2.

[16] I. Arganda-Carreras et al. “Crowdsourcing the creation of image segmentation algorithms for connectomics”. In: *Frontiers in Neuroanatomy* 9 (2015), p. 142. ISSN: 1662-5129. DOI: 10.3389/fnana.2015.00142. URL: https://www.frontiersin.org/article/10.3389/fnana.2015.00142.

[17] M. Arjovsky, S. Chintala, and L. Bottou. “Wasserstein generative adversarial networks”. In: *International Conference on Machine Learning*. 2017, pp. 214–223.

[18] V. Badrinarayanan, A. Kendall, and R. Cipolla. “SegNet: a deep convolutional encoder-Decoder architecture for image segmentation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.12 (2017), pp. 2481–2495. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2016.2644615.

[19] S. Baker and T. Kanade. “Limits on super-resolution and how to break them”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24.9 (2002), pp. 1167–1183. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2002.1033210.

[20] S. T. Barnard and M. A. Fischler. “Computational stereo”. In: *ACM Comput. Surv.* 14.4 (Dec. 1982), pp. 553–572. ISSN: 0360-0300. DOI: 10.1145/356893.356896. URL: http://doi.acm.org/10.1145/356893.356896.

[21] J. Batlle, E. Mouaddib, and J. Salvi. "Recent progress in coded structured light as a technique to solve the correspondence problem: a survey". In: *Pattern Recognition* 31.7 (1998), pp. 963 –982. ISSN: 0031-3203. DOI: 10.1016/S0031-3203(97)00074-5. URL: http://www.sciencedirect.com/science/article/pii/S0031320397000745.

[22] H. Bay et al. "Speeded-up robust features (SURF)". In: *Computer Vision and Image Understanding* 110.3 (2008). Similarity Matching in Computer Vision and Multimedia, pp. 346 –359. ISSN: 1077-3142. DOI: https://doi.org/10.1016/j.cviu.2007.09.014. URL: http://www.sciencedirect.com/science/article/pii/S1077314207001555.

[23] J. M. Benavides et al. "Multispectral digital colposcopy for in vivo detection of cervical cancer". In: *Opt. Express* 11.10 (2003), pp. 1223–1236. DOI: 10.1364/OE.11.001223. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-11-10-1223.

[24] Sylvain Bernhardt et al. "The status of augmented reality in laparoscopic surgery as of 2016". In: *Medical Image Analysis* 37.Supplement C (2017), pp. 66 –90. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2017.01.007. URL: http://www.sciencedirect.com/science/article/pii/S1361841517300178.

[25] P. J. Besl and N. D. McKay. "A method for registration of 3-D shapes". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2 (1992), pp. 239–256. ISSN: 0162-8828. DOI: 10.1109/34.121791.

[26] X. Bian, S. N. Lim, and N. Zhou. "Multiscale fully convolutional network with application to industrial inspection". In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2016, pp. 1–8. DOI: 10.1109/WACV.2016.7477595.

[27] C. M. Bishop. *Neural networks for pattern recognition*. New York, NY, USA: Oxford University Press, Inc., 1995. ISBN: 0198538642.

[28] L. Bottou, Y. Bengio, and Y. Le Cun. "Global training of document processing systems using graph transformer networks". In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1997, pp. 489–494. DOI: 10.1109/CVPR.1997.609370.

[29] L. E. Boucheron et al. "Utility of multispectral imaging for nuclear classification of routine clinical histopathology imagery". In: *BMC Cell Biology* 8.1 (2007), S8. ISSN: 1471-2121. DOI: 10.1186/1471-2121-8-S1-S8. URL: https://doi.org/10.1186/1471-2121-8-S1-S8.

[30] J. Y. Bouguet. *Camera calibration toolbox for Matlab*. 2008. URL: http://www.vision.caltech.edu/bouguetj/calib_doc/..

[31] K. L. Boyer and A. C. Kak. "Color-Encoded Structured Light for Rapid Active Ranging". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* PAMI-9.1 (1987), pp. 14 –28. ISSN: 0162-8828. DOI: 10.1109/TPAMI.1987.4767869.

[32] C. Bregler, A. Hertzmann, and H. Biermann. "Recovering non-rigid 3D shape from image streams". In: *Computer Vision and Pattern Recognition*. Vol. 2. 2000, pp. 2690–2696. DOI: 10.1109/CVPR.2000.854941.

[33] D. C. Brown. "Close-range camera calibration". In: *Photogrammetric Engineering* 37.8 (1971), pp. 855–866.

[34] H. C. Burger, C. J. Schuler, and S. Harmeling. "Image denoising: Can plain neural networks compete with BM3D?" In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. 2012, pp. 2392–2399. DOI: 10.1109/CVPR.2012.6247952.

[35] J. Canny. "A computational approach to edge detection". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-8.6 (1986), pp. 679–698. ISSN: 0162-8828. DOI: 10.1109/TPAMI.1986.4767851.

[36] X. Cao et al. "High resolution multispectral video capture with a hybrid camera system". In: *CVPR 2011*. 2011, pp. 297–304. DOI: 10.1109/CVPR.2011.5995418.

[37] A. Cardona et al. "An integrated micro- and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy". In: *PLOS Biology* 8.10 (Oct. 2010), pp. 1–17. DOI: 10.1371/journal.pbio.1000502. URL: https://doi.org/10.1371/journal.pbio.1000502.

[38] D. Cascio et al. "Mammogram segmentation by contour searching and mass lesions classification with neural network". In: *IEEE Transactions on Nuclear Science* 53.5 (2006), pp. 2827–2833. ISSN: 0018-9499. DOI: 10.1109/TNS.2006.878003.

[39] D. Caspi, N. Kiryati, and J. Shamir. "Range imaging with adaptive color structured light". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20.5 (1998), pp. 470 –480. ISSN: 0162-8828. DOI: 10.1109/34.682177.

[40] M. Chan et al. "Miniaturized three-Ddimensional endoscopic imaging system based on active stereovision". In: *Appl. Opt.* 42.10 (2003), pp. 1888–1898. DOI: 10.1364/AO.42.001888. URL: http://ao.osa.org/abstract.cfm?URI=ao-42-10-1888.

[41] H. Chang, D.-Y. Yeung, and Y. Xiong. "Super-resolution through neighbor embedding". In: *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* Vol. 1. 2004, pp. I–I. DOI: 10.1109/CVPR.2004.1315043.

[42] A. J. Chaudhari et al. "Hyperspectral and multispectral bioluminescence optical tomography for small animal imaging". In: *Physics in medicine and biology* 50.23 (2005), pp. 5421–5441. ISSN: 0031-9155. DOI: 10.1088/0031-9155/50/23/001. URL: https://doi.org/10.1088/0031-9155/50/23/001.

[43] J. Chen et al. "Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation". In: *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*. 2016, pp. 3036–3044. URL: http://papers.nips.cc/paper/6448-combining-fully-convolutional-and-recurrent-neural-networks-for-3d-biomedical-image-segmentation.

[44] L.-C. Chen et al. "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs". In: *CoRR* abs/1606.00915 (2016). arXiv: 1606.00915. URL: http://arxiv.org/abs/1606.00915.

[45] L. C. Chen et al. "DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP.99 (2017), pp. 1–1. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2017.2699184.

[46] J. A. Chin, E. C. Wang, and M. R. Kibbe. "Evaluation of hyperspectral technology for assessing the presence and severity of peripheral artery disease". In: *Journal of Vascular Surgery* 54.6 (2011), pp. 1679 –1688. ISSN: 0741-5214. DOI: https://doi.org/10.1016/j.jvs.2011.06.022. URL: http://www.sciencedirect.com/science/article/pii/S074152141101384X.

[47] A. Choromanska et al. "The loss surface of multilayer networks". In: *CoRR* abs/1412.0233 (2014). arXiv: 1412.0233. URL: http://arxiv.org/abs/1412.0233.

[48] J. Chung et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling". In: *CoRR* abs/1412.3555 (2014). arXiv: 1412.3555. URL: http://arxiv.org/abs/1412.3555.

[49] O. Çiçek et al. "3D U-Net: learning dense volumetric segmentation from sparse annotation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II*. Ed. by S. Ourselin et al. Cham: Springer International Publishing, 2016, pp. 424–432. ISBN: 978-3-319-46723-8. DOI: 10.1007/978-3-319-46723-8_49. URL: https://doi.org/10.1007/978-3-319-46723-8_49.

[50] N. T. Clancy et al. "Intraoperative measurement of bowel oxygen saturation using a multispectral imaging laparoscope". In: *Biomed. Opt. Express* 6.10 (2015), pp. 4179–4190. DOI: 10.1364/BOE.6.004179. URL: http://www.osapublishing.org/boe/abstract.cfm?URI=boe-6-10-4179.

[51] N. T. Clancy et al. "Spectrally encoded fiber-based structured lighting probe for intraoperative 3D imaging". In: *Biomed. Opt. Express* 2.11 (2011), pp. 3119–3128. DOI: 10.1364/BOE.2.003119. URL: http://www.opticsinfobase.org/boe/abstract.cfm?URI=boe-2-11-3119.

[52] D. Clevert, T. Unterthiner, and S. Hochreiter. "Fast and accurate deep network learning by exponential linear units (ELUs)". In: *CoRR* abs/1511.07289 (2015). arXiv: 1511.07289. URL: http://arxiv.org/abs/1511.07289.

[53] P. Cobelli et al. "Global measurement of water waves by Fourier transform profilometry". English. In: *Experiments in Fluids* 46 (6 2009), pp. 1037–1047. ISSN: 0723-4864. DOI: 10.1007/s00348-009-0611-z. URL: http://dx.doi.org/10.1007/s00348-009-0611-z.

[54] D. Cohen et al. "The use of spectral imaging for the diagnosis of retinal disease". In: *LEOS '99. IEEE Lasers and Electro-Optics Society 1999 12th Annual Meeting*. Vol. 1. 1999, 220–221 vol.1. DOI: 10.1109/LEOS.1999.813558.

[55] T. Collins and A. Bartoli. "Towards Live Monocular 3D Laparoscopy Using Shading and Specularity Information". In: *Information Processing in Computer-Assisted Interventions: Third International Conference, IPCAI 2012, Pisa, Italy, June 27, 2012. Proceedings*. Ed. by P. Abolmaesumi et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 11–21. ISBN: 978-3-642-30618-1. DOI: 10.1007/978-3-642-30618-1_2. URL: https://doi.org/10.1007/978-3-642-30618-1_2.

[56] A. Criminisi et al. "Efficient Dense Stereo with Occlusions for New View-Synthesis by Four-State Dynamic Programming". In: *International Journal of Computer Vision* 71.1 (2007), pp. 89–110. ISSN: 1573-1405. DOI: 10.1007/s11263-006-8525-1. URL: https://doi.org/10.1007/s11263-006-8525-1.

[57] J.E. Cryer, PS Tsai, and M. Shah. "Integration of shape from shading and stereo". In: *Pattern Recognition* 28 (7 1995), pp. 1033–1043. DOI: 10.1016/0031-3203(94)00183-M.

[58] J. Deng et al. "ImageNet: a large-scale hierarchical image database". In: *CVPR09*. 2009.

[59] D. Desjardins and P. Payeur. "Dense stereo range sensing with marching pseudo-random patterns". In: *Computer and Robot Vision, 2007. CRV '07. Fourth Canadian Conference on*. 2007, pp. 216 –226. DOI: 10.1109/CRV.2007.22.

[60] D. T. Dicker et al. "Differentiation of normal skin and melanoma using high resolution hyperspectral imaging". In: *Cancer biology & therapy* 5.1538-4047 (Aug. 2006), pp. 1033–1038. URL: http://www.landesbioscience.com/journals/cbt/article/3261/.

[61] J. Dolz, C. Desrosiers, and I. Ben Ayed. "3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study". In: *CoRR* abs/1612.03925 (2016). arXiv: 1612.03925. URL: http://arxiv.org/abs/1612.03925.

[62] C. Dong et al. "Image super-resolution using deep convolutional networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38.2 (2016), pp. 295–307. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2015.2439281.

[63] W. Dong et al. "Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization". In: *IEEE Transactions on Image Processing* 20.7 (2011), pp. 1838–1857. ISSN: 1057-7149. DOI: 10.1109/TIP.2011.2108306.

[64] A. Dosovitskiy and T. Brox. "Generating Images with Perceptual Similarity Metrics based on Deep Networks". In: *Advances in Neural Information Processing Systems 29*. Ed. by D. D. Lee et al. Curran Associates, Inc., 2016, pp. 658–666. URL: http://papers.nips.cc/paper/6158-generating-images-with-perceptual-similarity-metrics-based-on-deep-networks.pdf.

[65] A. Dosovitskiy et al. "FlowNet: learning optical flow with convolutional networks". In: *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015, pp. 2758–2766. DOI: 10.1109/ICCV.2015.316.

[66] X. Du et al. “Robust surface tracking combining features, intensity and illumination compensation”. In: *International Journal of Computer Assisted Radiology and Surgery* 10.12 (2015), pp. 1915–1926. ISSN: 1861-6429. DOI: 10.1007/s11548-015-1243-9. URL: https://doi.org/10.1007/s11548-015-1243-9.

[67] C. E Duchon. “Lanczos filtering in one and two dimensions”. In: *Journal of Applied Meteorology* 18.8 (1979), pp. 1016–1022.

[68] N.G. Durdle, J. Thayyoor, and V.J. Raso. “An improved structured light technique for surface reconstruction of the human trunk”. In: *Electrical and Computer Engineering, 1998. IEEE Canadian Conference on*. Vol. 2. 1998, 874 –877 vol.2. DOI: 10.1109/CCECE.1998.685637.

[69] S. Dutta, R. K. Woo, and T. M. Krummel. “Chapter 4 - advanced and emerging surgical technologies and the process of innovation”. In: *Pediatric Surgery (Seventh Edition)*. Ed. by A. G. Coran. Seventh Edition. Philadelphia: Mosby, 2012, pp. 37 –75. ISBN: 978-0-323-07255-7. DOI: https://doi.org/10.1016/B978-0-323-07255-7.00004-0. URL: https://www.sciencedirect.com/science/article/pii/B9780323072557000040.

[70] D. Eigen, D. Krishnan, and R. Fergus. “Restoring an Image Taken through a Window Covered with Dirt or Rain”. In: *2013 IEEE International Conference on Computer Vision*. 2013, pp. 633–640. DOI: 10.1109/ICCV.2013.84.

[71] T. Etzion. “Constructions for perfect maps and pseudorandom arrays”. In: *Information Theory, IEEE Transactions on* 34.5 (1988), pp. 1308 –1316. ISSN: 0018-9448. DOI: 10.1109/18.21260.

[72] R. Fattal. “Image Upsampling via Imposed Edge Statistics”. In: *ACM SIGGRAPH 2007 Papers*. SIGGRAPH ’07. San Diego, California: ACM, 2007. DOI: 10.1145/1275808.1276496. URL: http://doi.acm.org/10.1145/1275808.1276496.

[73] A. A. Fawzi et al. “Recovery of macular pigment spectrum ¡i¿in vivo¡/i¿ using hyperspectral image analysis”. In: *Journal of Biomedical Optics* 16 (2011), pp. 16 –16 –10. DOI: 10.1117/1.3640813. URL: http://dx.doi.org/10.1117/1.3640813.

[74] B. Fei et al. “Label-free reflectance hyperspectral imaging for tumor margin assessment: a pilot study on surgical specimens of cancer patients”. In: *Journal of Biomedical Optics* 22 (2017), pp. 22 –22 –7. DOI: 10.1117/1.JBO.22.8.086009. URL: http://dx.doi.org/10.1117/1.JBO.22.8.086009.

[75] Y.-Z. Feng and D.-W. Sun. “Application of Hyperspectral Imaging in Food Safety Inspection and Control: A Review”. In: *Critical Reviews in Food Science and Nutrition* 52.11 (2012). PMID: 22823350, pp. 1039–1058. DOI: 10.1080/10408398.2011.651542. eprint: https://doi.org/10.1080/10408398.2011.651542. URL: https://doi.org/10.1080/10408398.2011.651542.

[76] D. G. Ferris et al. “Multimodal hyperspectral imaging for the noninvasive diagnosis of cervical neoplasia”. In: *Journal of Lower Genital Tract Disease* 5.2 (2001), pp. 65–72.

[77] C. Fischer and I. Kakoulli. "Multispectral and hyperspectral imaging technologies in conservation: current research and potential applications". In: *Studies in Conservation* 51.sup1 (2006), pp. 3–16. DOI: 10.1179/sic.2006.51.Supplement-1.3. eprint: https://doi.org/10.1179/sic.2006.51.Supplement-1.3. URL: https://doi.org/10.1179/sic.2006.51.Supplement-1.3.

[78] M. A. Fischler and R. C. Bolles. "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography". In: *Commun. ACM* 24.6 (June 1981), pp. 381–395. ISSN: 0001-0782. DOI: 10.1145/358669.358692. URL: http://doi.acm.org/10.1145/358669.358692.

[79] B. Fisher et al. "Twenty-Year Follow-up of a Randomized Trial Comparing Total Mastectomy, Lumpectomy, and Lumpectomy plus Irradiation for the Treatment of Invasive Breast Cancer". In: *New England Journal of Medicine* 347.16 (2002). PMID: 12393820, pp. 1233–1241. DOI: 10.1056/NEJMoa022152. eprint: http://dx.doi.org/10.1056/NEJMoa022152. URL: http://dx.doi.org/10.1056/NEJMoa022152.

[80] A. W. Fitzgibbon, M. Pilu, and R. B. Fisher. "Direct least square fitting of ellipses". In: *IEEE Transaction on Pattern Analysis and Machine Intelligence* 21 (1999), pp. 476 –480.

[81] S. Foix, G. Alenya, and C. Torras. "Lock-in Time-of-Flight (ToF) Cameras: A Survey". In: *IEEE Sensors Journal* 11.9 (2011), pp. 1917–1926. ISSN: 1530-437X. DOI: 10.1109/JSEN.2010.2101060.

[82] R. Franke. "Smooth interpolation of scattered data by local thin plate splines". In: *Computers & Mathematics with Applications* 8.4 (1982), pp. 273 –281. ISSN: 0898-1221. DOI: http://dx.doi.org/10.1016/0898-1221(82)90009-8. URL: http://www.sciencedirect.com/science/article/pii/0898122182900098.

[83] G. Freedman and R. Fattal. "Image and Video Upscaling from Local Self-examples". In: *ACM Trans. Graph.* 30.2 (Apr. 2011), 12:1–12:11. ISSN: 0730-0301. DOI: 10.1145/1944846.1944852. URL: http://doi.acm.org/10.1145/1944846.1944852.

[84] J. Freeman et al. "Multispectral and hyperspectral imaging: applications for medical and surgical diagnostics". In: *Engineering in Medicine and Biology Society, 1997. Proceedings of the 19th Annual International Conference of the IEEE.* Vol. 2. 1997, 700–701 vol.2. DOI: 10.1109/IEMBS.1997.757727.

[85] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. "Learning Low-Level Vision". In: *International Journal of Computer Vision* 40.1 (2000), pp. 25–47. ISSN: 1573-1405. DOI: 10.1023/A:1026501619075. URL: https://doi.org/10.1023/A:1026501619075.

[86] L. Gao, R. T. Kester, and T. S. Tkaczyk. "Compact Image Slicing Spectrometer (ISS) for hyperspectral fluorescence microscopy". In: *Opt. Express* 17.15 (2009), pp. 12293–12308. DOI: 10.1364/OE.17.012293. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-17-15-12293.

[87] A. Garcia-Garcia et al. "A review on deep learning techniques applied to semantic segmentation". In: *CoRR* abs/1704.06857 (2017). arXiv: 1704.06857. URL: http://arxiv.org/abs/1704.06857.

[88] R. Garg, A. Roussos, and L. Agapito. *Robust trajectory-space TV-L1 optical flow for non-rigid sequences.* 2011, pp. 300–314. DOI: 10.1007/978-3-642-23094-3_22.

[89] N. Gat. *Imaging spectroscopy using tunable filters: a review*. 2000. DOI: 10.1117/12.381686. URL: http://dx.doi.org/10.1117/12.381686.

[90] J. Geng. "Structured-light 3D surface imaging: a tutorial". In: *Adv. Opt. Photon.* 3.2 (2011), pp. 128–160. DOI: 10.1364/AOP.3.000128. URL: http://aop.osa.org/abstract.cfm?URI=aop-3-2-128.

[91] S. Giannarou, M. Visentini-Scarzanella, and G.-Z. Yang. "Probabilistic Tracking of Affine-Invariant Anisotropic Regions". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.1 (2013), pp. 130–143. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2012.81.

[92] D. Glasner, S. Bagon, and M. Irani. "Super-resolution from a single image". In: *2009 IEEE 12th International Conference on Computer Vision*. 2009, pp. 349–356. DOI: 10.1109/ICCV.2009.5459271.

[93] X. Glorot, A. Bordes, and Y. Bengio. "Deep sparse rectifier neural networks". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Vol. 15. PMLR, 2011, pp. 205 –208.

[94] M. Goel et al. "HyperCam: Hyperspectral Imaging for Ubiquitous Computing Applications". In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '15. Osaka, Japan: ACM, 2015, pp. 145–156. ISBN: 978-1-4503-3574-4. DOI: 10.1145/2750858.2804282. URL: http://doi.acm.org/10.1145/2750858.2804282.

[95] A. F. H. Goetz. "Three decades of hyperspectral remote sensing of the Earth: A personal view". In: *Remote Sensing of Environment* 113.Supplement 1 (2009). Imaging Spectroscopy Special Issue, S5 – S16. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2007.12.014. URL: http://www.sciencedirect.com/science/article/pii/S003442570900073X.

[96] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[97] I. Goodfellow et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani et al. Curran Associates, Inc., 2014, pp. 2672–2680. URL: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

[98] S. S. Gorthi and P. Rastogi. "Fringe projection techniques: whither we are?" 2010.

[99] M. Govender, K. Chetty, and H. Bulcock. "A review of hyperspectral remote sensing and its application in vegetation and water resource studies". English. In: *Water SA* 33.2 (2007), pp. 145–151. ISSN: 0378-4738. URL: https://journals.co.za/content/waters/33/2/EJC116430.

[100] A. A. Gowen et al. "Hyperspectral imaging: an emerging process analytical tool for food quality and safety control". In: *Trends in Food Science & Technology* 18.12 (2007), pp. 590–598. ISSN: 0924-2244. DOI: https://doi.org/10.1016/j.tifs.2007.06.001. URL: http://www.sciencedirect.com/science/article/pii/S0924224407002026.

[101] P. M. Griffin, L. S. Narasimhan, and S. R. Yee. "Generation of uniquely encoded light patterns for range data acquisition". In: *Pattern Recognition* 25.6 (1992), pp. 609 –616. ISSN: 0031-3203. DOI: 10.1016/0031-3203(92)90078-W. URL: http://www.sciencedirect.com/science/article/pii/003132039290078W.

[102] C. A. Grimes, E. C. Dickey, M. V. Pishko, et al. *Encyclopedia of sensors*. Vol. 10.

[103] A. Groch et al. "3D surface reconstruction for laparoscopic computer-assisted interventions: comparison of state-of-the-art methods". In: *SPIE Medical Imaging: Visualization, Image-Guided Procedures, and Modeling*. 2011, pp. 796415–796415–9. DOI: 10.1117/12.878354. URL: +http://dx.doi.org/10.1117/12.878354.

[104] S. Gu et al. "Convolutional Sparse Coding for Image Super-Resolution". In: *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015, pp. 1823–1831. DOI: 10.1109/ICCV.2015.212.

[105] B. W. Rice H. Xu. "In-vivo fluorescence imaging with a multivariate curve resolution spectral unmixing technique". In: *Journal of Biomedical Optics* 14 (2009), pp. 14 –14 –9. DOI: 10.1117/1.3258838. URL: http://dx.doi.org/10.1117/1.3258838.

[106] L. O. Hall et al. "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain". In: *IEEE Transactions on Neural Networks* 3.5 (1992), pp. 672–682. ISSN: 1045-9227. DOI: 10.1109/72.159057.

[107] H. Haneishi, T. Ogura, and Y. Miyake. "Profilometry of a gastrointestinal surface by an endoscope with laser beam projection". In: *Opt. Lett.* 19.9 (1994), pp. 601–603. DOI: 10.1364/OL.19.000601. URL: http://ol.osa.org/abstract.cfm?URI=ol-19-9-601.

[108] J. P. Hartke and E. L. Dereniak. "Snapshot dual-band visible hyperspectral imaging spectrometer". In: *Optical Engineering* 46 (2007), pp. 46 –46 –7. DOI: 10.1117/1.2434348. URL: http://dx.doi.org/10.1117/1.2434348.

[109] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Second. Cambridge University Press, 2003.

[110] K. Hasegawa, K. Noda, and Y. Sato. "Electronic endoscope system for shape measurement". In: *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. Vol. 1. 2002, 792 –795 vol.1. DOI: 10.1109/ICPR.2002.1044878.

[111] D. Hattery et al. "Hyperspectral imaging of Kaposi's Sarcoma for disease assessment and treatment monitoring". In: *Applied Imagery Pattern Recognition Workshop, 2002. Proceedings.* 2002, pp. 124–130. DOI: 10.1109/AIPR.2002.1182265.

[112] M. Hayashibe, N. Suzuki, and Y. Nakamura. "Laser-scan endoscope system for intraoperative geometry acquisition and surgical robot safety management". In: *Medical Image Analysis* 10.4 (2006). Special Issue on Functional Imaging and Modelling of the Heart (FIMH 2005), pp. 509 –519. ISSN: 1361-8415. DOI: 10.1016/j.media.2006.03.001. URL: http://www.sciencedirect.com/science/article/pii/S1361841506000156.

[113] H. He and W. C. Siu. "Single image super-resolution using Gaussian process regression". In: *CVPR 2011*. 2011, pp. 449–456. DOI: 10.1109/CVPR.2011.5995713.

[114] K. He et al. "Deep residual learning for image recognition". In: *CoRR* abs/1512.03385 (2015). arXiv: 1512.03385. URL: http://arxiv.org/abs/1512.03385.

[115] J. Heikkila and O. Silven. "A four-step camera calibration procedure with implicit image correction". In: *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. 1997, pp. 1106–1112. DOI: 10.1109/CVPR.1997.609468.

[116] G. E. Hinton and T. Shallice. "Lesioning an attractor network: investigations of acquired dyslexia". In: *Psychological Review* 98.1 (1991), pp. 74–95. ISSN: 0033-295X. URL: https://www.learntechlib.org/p/79180.

[117] Y. Hirohara et al. "Validity of retinal oxygen saturation analysis: Hyperspectral imaging in visible wavelength with fundus camera and liquid crystal wavelength tunable filter". In: *Optical Review* 14.3 (2007), p. 151. ISSN: 1349-9432. DOI: 10.1007/BF02919416. URL: https://doi.org/10.1007/BF02919416.

[118] S. Hochreiter and J. Schmidhuber. "Long short-term memory". In: *Neural Comput.* 9.8 (Nov. 1997), pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. URL: http://dx.doi.org/10.1162/neco.1997.9.8.1735.

[119] M. S. Holzer et al. "Assessment of Renal Oxygenation During Partial Nephrectomy Using Hyperspectral Imaging". In: *The Journal of Urology* 186.2 (2011), pp. 400 –404. ISSN: 0022-5347. DOI: https://doi.org/10.1016/j.juro.2011.03.162. URL: http://www.sciencedirect.com/science/article/pii/S0022534711035750.

[120] B. K. P. Horn. "Height and gradient from shading". In: *International Journal of Computer Vision* 5.1 (1990), pp. 37–75. ISSN: 1573-1405. DOI: 10.1007/BF00056771. URL: https://doi.org/10.1007/BF00056771.

[121] E. Horn and N. Kiryati. "Toward optimal structured light patterns". In: *3-D Digital Imaging and Modeling, 1997. Proceedings., International Conference on Recent Advances in*. 1997, pp. 28 –35. DOI: 10.1109/IM.1997.603845.

[122] P. S. Huang and S. Zhang. "Fast three-step phase-shifting algorithm". In: *Storage and Retrieval for Image and Video Databases*. 2006.

[123] H. Hugli and G. Maitre. "Generation and use of color pseudo random sequences for coding structured light in active ranging". In: (1989), pp. 75–82. DOI: 10.1117/12.949215. URL: http://dx.doi.org/10.1117/12.949215.

[124] G. J. Iddan and G. Yahav. "3D imaging in the studio". In: *Proc. SPIE*. Vol. 4298. 1, pp. 48–55.

[125] S. Inokuchi, K. Sato, and F. Matsuda. "Range imaging system for 3-d object recognition". In: *International Conference on Pattern Recognition*. 1984.

[126] M. Irani and S. Peleg. "Improving resolution by image registration". In: *CVGIP: Graphical Models and Image Processing* 53.3 (1991), pp. 231 –239. ISSN: 1049-9652. DOI: https://doi.org/10.1016/1049-9652(91)90045-L. URL: http://www.sciencedirect.com/science/article/pii/104996529190045L.

[127] I. Ishii et al. "High-speed 3D image acquisition using coded structured light projection". In: *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*. 2007, pp. 925 –930. DOI: 10.1109/IROS.2007.4399180.

[128] Itseez. *Open source computer vision library*. https://github.com/itseez/opencv. 2015.

[129] V. Jain and S. Seung. "Natural Image Denoising with Convolutional Networks". In: *Advances in Neural Information Processing Systems 21*. Ed. by D. Koller et al. Curran Associates, Inc., 2009, pp. 769–776. URL: http://papers.nips.cc/paper/3506-natural-image-denoising-with-convolutional-networks.pdf.

[130] S. Ji et al. "3D convolutional neural networks for human action recognition". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.1 (2013), pp. 221–231. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2012.59.

[131] Y. Jia et al. "Caffe: convolutional architecture for fast feature embedding". In: *arXiv preprint arXiv:1408.5093* (2014).

[132] M. Joel and V.-D. Tuan. *Biomedical Photonics Handbook*. CRC Press, 2003.

[133] J. Johnson, A. Alahi, and F.-F. Li. "Perceptual Losses for Real-Time Style Transfer and Super-Resolution". In: *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II*. Ed. by B. Leibe et al. Cham: Springer International Publishing, 2016, pp. 694–711. ISBN: 978-3-319-46475-6. DOI: 10.1007/978-3-319-46475-6_43. URL: https://doi.org/10.1007/978-3-319-46475-6_43.

[134] W. R. Johnson et al. "Snapshot hyperspectral imaging in ophthalmology". In: *Journal of Biomedical Optics* 12 (2007), pp. 12 –12 –7. DOI: 10.1117/1.2434950. URL: http://dx.doi.org/10.1117/1.2434950.

[135] R. Jolivot, P. Vabres, and F. Marzani. "Reconstruction of hyperspectral cutaneous data from an artificial neural network-based multispectral imaging system". In: *Computerized Medical Imaging and Graphics* 35.2 (2011). Advances in Skin Cancer Image Analysis, pp. 85 –88. ISSN: 0895-6111. DOI: https://doi.org/10.1016/j.compmedimag.2010.07.001. URL: http://www.sciencedirect.com/science/article/pii/S0895611110000662.

[136] D. M. Joshi, N. K. Rana, and V. M. Misra. "Classification of brain cancer using artificial neural network". In: *2010 2nd International Conference on Electronic Computer Technology*. 2010, pp. 112–116. DOI: 10.1109/ICECTECH.2010.5479975.

[137] Mikko K., Mikko N., and Pirkko O. *Method for measuring stereo camera depth accuracy based on stereoscopic vision*. 2011. DOI: 10.1117/12.872015. URL: http://dx.doi.org/10.1117/12.872015.

[138] K. Kamnitsas et al. "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation". In: *Medical Image Analysis* 36.Supplement C (2017), pp. 61 –78. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2016.10.004. URL: http://www.sciencedirect.com/science/article/pii/S1361841516301839.

[139] X. Kang, S. Li, and J. A. Benediktsson. "Spectral-Spatial Hyperspectral Image Classification With Edge-Preserving Filtering". In: *IEEE Transactions on Geoscience and Remote Sensing* 52.5 (2014), pp. 2666–2677. ISSN: 0196-2892. DOI: 10.1109/TGRS.2013.2264508.

[140] R. Kawakami et al. "High-resolution hyperspectral imaging via matrix factorization". In: *CVPR 2011*. 2011, pp. 2329–2336. DOI: 10.1109/CVPR.2011.5995457.

[141] K. Keller and J. D. Ackerman. "Real-time structured light depth extraction". In: (2000), pp. 11–18. DOI: 10.1117/12.380037. URL: +http://dx.doi.org/10.1117/12.380037.

[142] L. Khaodhiar et al. "The Use of Medical Hyperspectral Technology to Evaluate Microcirculatory Changes in Diabetic Foot Ulcers and to Predict Clinical Outcomes". In: *Diabetes Care* 30.4 (2007), pp. 903–910. ISSN: 0149-5992. DOI: 10.2337/dc06-2209. eprint: http://care.diabetesjournals.org/content/30/4/903.full.pdf. URL: http://care.diabetesjournals.org/content/30/4/903.

[143] B. Khoobehi, J. M. Beach, and H. Kawano. "Hyperspectral imaging for measurement of oxygen saturation in the optic nerve head". In: *Investigative ophthalmology & visual science* 45.5 (2004), pp. 1464–1472. ISSN: 0146-0404. DOI: 10.1167/iovs.03-1069. URL: https://doi.org/10.1167/iovs.03-1069.

[144] J. Kim, J. K. Lee, and K. M. Lee. "Accurate Image Super-Resolution Using Very Deep Convolutional Networks". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 1646–1654. DOI: 10.1109/CVPR.2016.182.

[145] J. Kim, J. K. Lee, and K. M. Lee. "Deeply-recursive convolutional network for image super-resolution". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 1637–1645. DOI: 10.1109/CVPR.2016.181.

[146] K. I. Kim and Y. Kwon. "Single-Image Super-Resolution Using Sparse Regression and Natural Image Prior". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.6 (2010), pp. 1127–1133. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.25.

[147] C. Kittrell et al. "Diagnosis of fibrous arterial atherosclerosis using fluorescence". In: *Appl. Opt.* 24.15 (1985), pp. 2280–2281. DOI: 10.1364/AO.24.002280. URL: http://ao.osa.org/abstract.cfm?URI=ao-24-15-2280.

[148] D. Knaus et al. "System for laparoscopic tissue tracking". In: *Biomedical Imaging: Nano to Macro, 2006. 3rd IEEE International Symposium on*. 2006, pp. 498 –501. DOI: 10.1109/ISBI.2006.1624962.

[149] V. Kolmogorov et al. "Probabilistic fusion of stereo with color and contrast for bilayer segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.9 (2006), pp. 1480–1492. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2006.193.

[150] P. Krähenbühl and V. Koltun. "Efficient inference in fully connected CRFs with Gaussian edge potentials". In: *Advances in Neural Information Processing Systems 24*. Ed. by J. Shawe-Taylor et al. Curran Associates, Inc., 2011, pp. 109–117. URL: http://papers.nips.cc/paper/4296-efficient-inference-in-fully-connected-crfs-with-gaussian-edge-potentials.pdf.

[151] A. Krizhevsky, I. Sutskever, and G. E Hinton. "ImageNet classification with deep convolutional neural networks". In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[152] S. Y. Kung and J. S. Taur. "Decision-based neural networks with signal/image classification applications". In: *IEEE Transactions on Neural Networks* 6.1 (1995), pp. 170–181. ISSN: 1045-9227. DOI: 10.1109/72.363439.

[153] J. Kuula et al. *Using VIS/NIR and IR spectral cameras for detecting and separating crime scene details*. 2012. DOI: 10.1117/12.918555. URL: http://dx.doi.org/10.1117/12.918555.

[154] C. A. Laben and B. V. Brower. *Process for enhancing the spatial resolution of multispectral imagery using pan-sharpening*. US Patent 6,011,875. 2000.

[155] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. "Conditional random fields: probabilistic models for segmenting and labeling sequence data". In: *Proceedings of the Eighteenth International Conference on Machine Learning*. ICML '01. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001, pp. 282–289. ISBN: 1-55860-778-1. URL: http://dl.acm.org/citation.cfm?id=645530.655813.

[156] D. X. Le, G. R. Thoma, and H. Wechsler. "Classification of binary document images into textual or nontextual data blocks using neural network models". In: *Machine Vision and Applications* 8.5 (1995), pp. 289–304. ISSN: 1432-1769. DOI: 10.1007/BF01211490. URL: https://doi.org/10.1007/BF01211490.

[157] Y. Lecun. "A theoretical framework for Back-Propagation". In: (). URL: http://yann.lecun.com/exdb/publis/pdf/lecun-92b.pdf.

[158] Y. LeCun et al. "Effiicient BackProp". In: *Neural networks: tricks of the trade, This Book is an Outgrowth of a 1996 NIPS Workshop*. London, UK, UK: Springer-Verlag, 1998, pp. 9–50. ISBN: 3-540-65311-2. URL: http://dl.acm.org/citation.cfm?id=645754.668382.

[159] Y. Lecun et al. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324. ISSN: 0018-9219. DOI: 10.1109/5.726791.

[160] C. Ledig et al. "Photo-realistic single image super-resolution using a generative adversarial network". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2017, pp. 105–114. DOI: 10.1109/CVPR.2017.19.

[161] Fei fei Li. *CS231n: Convolutional Neural Networks for Visual Recognition*. 2017.

[162] Q. Li et al. "Review of spectral imaging technology in biomedical engineering: achievements and challenges". In: *Journal of Biomedical Optics* 18 (2013), pp. 18 –18 –29. DOI: 10.1117/1.JBO.18.10.100901. URL: http://dx.doi.org/10.1117/1.JBO.18.10.100901.

[163] Q. Li et al. "Study on microscope hyperspectral medical imaging method for biomedical quantitative analysis". In: *Chinese Science Bulletin* 53.9 (2008), pp. 1431–1434. ISSN: 1861-9541. DOI: 10.1007/s11434-008-0199-3. URL: https://doi.org/10.1007/s11434-008-0199-3.

[164] W. Liao et al. "Processing of Multiresolution Thermal Hyperspectral and Digital Color Data: Outcome of the 2014 IEEE GRSS Data Fusion Contest". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 8.6 (2015), pp. 2984–2996. ISSN: 1939-1404. DOI: 10.1109/JSTARS.2015.2420582.

[165] J. Lin, N. T. Clancy, and D. S. Elson. "An endoscopic structured light system using multispectral detection". In: *International Journal of Computer Assisted Radiology and Surgery* 10.12 (2015), pp. 1941–1950. ISSN: 1861-6429. DOI: 10.1007/s11548-015-1264-4. URL: http://dx.doi.org/10.1007/s11548-015-1264-4.

[166] J. Lin et al. "Endoscopic depth measurement and super-spectral-resolution imaging". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II*. Ed. by M. Descoteaux et al. Cham: Springer International Publishing, 2017, pp. 39–47. ISBN: 978-3-319-66185-8. DOI: 10.1007/978-3-319-66185-8_5. URL: https://doi.org/10.1007/978-3-319-66185-8_5.

[167] J. Lin et al. "Probe-based rapid hybrid hyperspectral and tissue surface imaging aided by fully convolutional networks". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part III*. Ed. by S. Ourselin et al. Cham: Springer International Publishing, 2016, pp. 414–422. ISBN: 978-3-319-46726-9. DOI: 10.1007/978-3-319-46726-9_48. URL: https://doi.org/10.1007/978-3-319-46726-9_48.

[168] J. Lin et al. "Tissue surface reconstruction aided by local normal information using a self-calibrated endoscopic structured light system". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part I*. Ed. by N. Navab et al. Cham: Springer International Publishing, 2015, pp. 405–412. ISBN: 978-3-319-24553-9. DOI: 10.1007/978-3-319-24553-9_50. URL: http://dx.doi.org/10.1007/978-3-319-24553-9_50.

[169] T.-Y. Lin et al. "Microsoft COCO: common objects in context". In: *Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V*. Ed. by D. Fleet et al. Cham: Springer International Publishing, 2014, pp. 740–755. ISBN: 978-3-319-10602-1. DOI: 10.1007/978-3-319-10602-1_48. URL: https://doi.org/10.1007/978-3-319-10602-1_48.

[170] J. G. Liu. "Smoothing Filter-based Intensity Modulation: A spectral preserve image fusion technique for improving spatial details". In: *International Journal of Remote Sensing* 21.18 (2000), pp. 3461–3472. DOI: 10.1080/014311600750037499. eprint: https://doi.org/10.1080/014311600750037499. URL: https://doi.org/10.1080/014311600750037499.

[171] L. Loncan et al. "Hyperspectral Pansharpening: A Review". In: *IEEE Geoscience and Remote Sensing Magazine* 3.3 (2015), pp. 27–46. ISSN: 2473-2397. DOI: 10.1109/MGRS.2015.2440094.

[172] J. Long, E. Shelhamer, and T. Darrell. "Fully convolutional networks for semantic segmentation". In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2015, pp. 3431–3440. DOI: 10.1109/CVPR.2015.7298965.

[173] D. G. Lowe. "Object recognition from local scale-invariant features". In: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Vol. 2. 1999, 1150–1157 vol.2. DOI: 10.1109/ICCV.1999.790410.

[174] G. Lu and B. Fei. "Medical hyperspectral imaging: a review". In: *Journal of Biomedical Optics* 19.1 (2014), p. 010901. DOI: 10.1117/1.JBO.19.1.010901. URL: http://dx.doi.org/10.1117/1.JBO.19.1.010901.

[175] B. D. Lucas and T. Kanade. "An iterative image registration technique with an application to stereo vision". In: *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*. IJCAI'81. Vancouver, BC, Canada: Morgan Kaufmann Publishers Inc., 1981, pp. 674–679. URL: http://dl.acm.org/citation.cfm?id=1623264.1623280.

[176] H. Luo et al. "A simple calibration procedure for structured light system". In: *Optics and Lasers in Engineering* 57.0 (2014), pp. 6 –12. ISSN: 0143-8166. DOI: http://dx.doi.org/10.1016/j.optlaseng.2014.01.010. URL: http://www.sciencedirect.com/science/article/pii/S0143816614000116.

[177] U. von Luxburg. "A tutorial on spectral clustering". English. In: *Statistics and Computing* 17.4 (2007), pp. 395–416. ISSN: 0960-3174. DOI: 10.1007/s11222-007-9033-z. URL: http://dx.doi.org/10.1007/s11222-007-9033-z.

[178] M. et al. "TensorFlow: A system for large-scale machine learning". In: *CoRR* abs/1605.08695 (2016). arXiv: 1605.08695. URL: http://arxiv.org/abs/1605.08695.

[179] A. L. Maas, A.Y. Hannun, and A. Y. Ng. "Rectifier Nonlinearities Improve Neural Network Acoustic Models". In: 2013.

[180] F.J. MacWilliams and N.J.A. Sloane. "Pseudo-random sequences and arrays". In: *Proceedings of the IEEE* 64.12 (1976), pp. 1715 –1729. ISSN: 0018-9219. DOI: 10.1109/PROC.1976.10411.

[181] L. Maier-Hein et al. "Comparative Validation of Single-Shot Optical Techniques for Laparoscopic 3-D Surface Reconstruction". In: *Medical Imaging, IEEE Transactions on* 33.10 (2014), pp. 1913–1930. ISSN: 0278-0062. DOI: 10.1109/TMI.2014.2325607.

[182] L. Maier-Hein et al. "Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery". In: *Medical Image Analysis* 17.8 (2013), pp. 974 –996. ISSN: 1361-8415. DOI: http://dx.doi.org/10.1016/j.media.2013.04.003. URL: http://www.sciencedirect.com/science/article/pii/S1361841513000480.

[183] A. Malti, A. Bartoli, and T. Collins. "Template-Based Conformal Shape-from-Motion-and-Shading for Laparoscopy". In: *International Conference on Information Processing in Computer-Assisted Interventions (IPCAI)*. Ed. by Purang Abolmaesumi et al. Vol. 7330. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2012, pp. 1–10. ISBN: 978-3-642-30617-4. DOI: 10.1007/978-3-642-30618-1_1. URL: http://dx.doi.org/10.1007/978-3-642-30618-1_1.

[184] A. C. Pais Barreto Marques and A. C. Gay Thome. "A neural network fingerprint segmentation method". In: *Fifth International Conference on Hybrid Intelligent Systems (HIS'05)*. 2005, 6 pp.–. DOI: 10.1109/ICHIS.2005.12.

[185] B. Marques et al. "Framework for augmented reality in Minimally Invasive laparoscopic surgery". In: *2015 17th International Conference on E-health Networking, Application Services (HealthCom)*. 2015, pp. 22–27. DOI: 10.1109/HealthCom.2015.7454467.

[186] M. Maruyama and S. Abe. "Range sensing by projecting multi-slits with random cuts". In: *Industrial Applications of Machine Intelligence and Vision, 1989., International Workshop on*. 1989, pp. 163 –168. DOI: 10.1109/MIV.1989.40543.

[187] G. Masi et al. "Pansharpening by Convolutional Neural Networks". In: *Remote Sensing* 8.7 (2016). ISSN: 2072-4292. DOI: 10.3390/rs8070594. URL: http://www.mdpi.com/2072-4292/8/7/594.

[188] J. Matas et al. "Robust wide-baseline stereo from maximally stable extremal regions". In: *Image and Vision Computing* 22.10 (2004). British Machine Vision Computing 2002, pp. 761 –767. ISSN: 0262-8856. DOI: https://doi.org/10.1016/j.imavis.2004.02.006. URL: http://www.sciencedirect.com/science/article/pii/S0262885604000435.

[189] X. Maurice et al. "A structured light-based laparoscope with real-time organs' surface reconstruction for minimally invasive surgery". In: *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. 2012, pp. 5769 –5772. DOI: 10.1109/EMBC.2012.6347305.

[190] N. Mayer et al. "A Large Dataset to Train Convolutional Networks for Disparity, Optical Flow, and Scene Flow Estimation". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 4040–4048. DOI: 10.1109/CVPR.2016.438.

[191] M. F. McNitt-Gray, H. K. Huang, and J. W. Sayre. "Feature selection in the pattern classification problem of digital chest radiograph segmentation". In: *IEEE Transactions on Medical Imaging* 14.3 (1995), pp. 537–547. ISSN: 0278-0062. DOI: 10.1109/42.414619.

[192] D. J. Mirota, M. Ishii, and G. D. Hager. "Vision-based navigation in image-guided interventions". In: *Annu Rev Biomed Eng* 1523-9829 (Linking) (2011), pp. 279–319.

[193] K. Mitra et al. “Indocyanine-green-loaded microballoons for biliary imaging in cholecystectomy”. In: *Journal of Biomedical Optics* 17 (2012), pp. 17–17–8. DOI: 10.1117/1.JBO.17.11.116025. URL: http://dx.doi.org/10.1117/1.JBO.17.11.116025.

[194] T.P. Monks and J.N. Carter. “Improved stripe matching for colour encoded structured light”. In: *Lecture Notes in Computer Science*. Ed. by D. Chetverikov and W. Kropatsch. Vol. 719. Address: Berlin. Springer Verlag, 1993, pp. 476–485. URL: http://eprints.soton.ac.uk/250335/.

[195] S. T. Monteiro, Y. Kosugi, and K. Uto. “Towards Applying hyperspectral Imagery as an Intraoperative Visual Aid Tool”. In: *Proceedings of the fourth International Conference on Visualization, Imaging and Image Processing*. 2004, pp. 483–488.

[196] F. Moreno-Noguer, J. M. Porta, and P. Fua. “Exploring Ambiguities for Monocular Non-rigid Shape Estimation”. In: *European Conference on Computer Vision*. 2010, pp. 370–383. DOI: 10.1007/978-3-642-15558-1_27.

[197] R. Mur-Artal and J. D. Tardos. “ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras”. In: *IEEE Transactions on Robotics* 33.5 (2017), pp. 1255–1262. ISSN: 1552-3098. DOI: 10.1109/TRO.2017.2705103.

[198] K. Nasrollahi and T. B. Moeslund. “Super-resolution: a comprehensive survey”. In: *Machine Vision and Applications* 25.6 (2014), pp. 1423–1468. ISSN: 1432-1769. DOI: 10.1007/s00138-014-0623-4. URL: https://doi.org/10.1007/s00138-014-0623-4.

[199] A. Y. Ng, M. I. Jordan, and Y. Weiss. “On Spectral Clustering: Analysis and an algorithm”. In: *Advances in Neural Information Processing Systems*. MIT Press, 2001, pp. 849–856.

[200] K. S. Ni and T. Q. Nguyen. “Image Superresolution Using Support Vector Regression”. In: *IEEE Transactions on Image Processing* 16.6 (2007), pp. 1596–1610. ISSN: 1057-7149. DOI: 10.1109/TIP.2007.896644.

[201] Feng Ning et al. “Toward automatic phenotyping of developing embryos from videos”. In: *IEEE Transactions on Image Processing* 14.9 (2005), pp. 1360–1371. ISSN: 1057-7149. DOI: 10.1109/TIP.2005.852470.

[202] J. C. Obuch, C. M. Pigott, and D. J. Ahnen. “Sessile Serrated Polyps: Detection, Eradication, and Prevention of the Evil Twin”. In: *Current Treatment Options in Gastroenterology* 13.1 (2015), pp. 156–170. ISSN: 1534-309X. DOI: 10.1007/s11938-015-0046-y. URL: https://doi.org/10.1007/s11938-015-0046-y.

[203] T. Okatani and K. Deguchi. “Shape Reconstruction from an Endoscope Image by Shape from Shading Technique for a Point Light Source at the Projection Center”. In: *Computer Vision and Image Understanding* 66.2 (1997), pp. 119 –131. ISSN: 1077-3142. DOI: https://doi.org/10.1006/cviu.1997.0613. URL: http://www.sciencedirect.com/science/article/pii/S1077314297906135.

[204] O. Oktay et al. "Multi-input Cardiac Image Super-Resolution Using Convolutional Neural Networks". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part III*. Ed. by S. Ourselin et al. Cham: Springer International Publishing, 2016, pp. 246–254. ISBN: 978-3-319-46726-9. DOI: 10.1007/978-3-319-46726-9_29. URL: https://doi.org/10.1007/978-3-319-46726-9_29.

[205] E. O. Olweny et al. "Renal oxygenation during robot-assisted laparoscopic partial nephrectomy: characterization using laparoscopic digital light processing hyperspectral imaging". In: *Journal of Endourology* 27.3 (2013), pp. 265–269. ISSN: 0892-7790. DOI: 10.1089/end.2012.0207. URL: https://doi.org/10.1089/end.2012.0207.

[206] N. Ono et al. "Real-time 3D imager based on spatio-temporal phase unwrapping". In: *SICE 2004 Annual Conference*. Vol. 3. 2004, 2544 –2547 vol. 3.

[207] N. R. Pal and S. K. Pal. "A review on image segmentation techniques". In: *Pattern Recognition* 26.9 (1993), pp. 1277 –1294. ISSN: 0031-3203. DOI: https://doi.org/10.1016/0031-3203(93)90135-J. URL: http://www.sciencedirect.com/science/article/pii/003132039390135J.

[208] C. Palmer and E. G Loewen. *Diffraction grating handbook*. 2005.

[209] M. Parmar, S. Lansel, and B. A. Wandell. "Spatio-spectral reconstruction of the multispectral datacube using sparse recovery". In: *2008 15th IEEE International Conference on Image Processing*. 2008, pp. 473–476. DOI: 10.1109/ICIP.2008.4711794.

[210] A. Paszke et al. "ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation". In: *CoRR* abs/1606.02147 (2016). arXiv: 1606.02147. URL: http://arxiv.org/abs/1606.02147.

[211] D. Patel, E.R. Davies, and I. Hnnah. "The use of convolution operators for detecting contaminants in food images". In: *Pattern Recognition* 29.6 (1996), pp. 1019 –1029. ISSN: 0031-3203. DOI: https://doi.org/10.1016/0031-3203(96)00135-5. URL: http://www.sciencedirect.com/science/article/pii/0031320396001355.

[212] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad. "Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition". In: *Proceedings of 27th Asilomar Conference on Signals, Systems and Computers*. 1993, 40–44 vol.1. DOI: 10.1109/ACSSC.1993.342465.

[213] M. S. Patterson, B. C. Wilson, and D. R. Wyman. "The propagation of optical radiation in tissue I. Models of radiation transport and their application". In: *Lasers in Medical Science* 6.2 (1991), pp. 155–168. ISSN: 1435-604X. DOI: 10.1007/BF02032543. URL: https://doi.org/10.1007/BF02032543.

[214] I. Pavlova et al. "Microanatomical and Biochemical Origins of Normal and Precancerous Cervical Autofluorescence Using Laser-scanning Fluorescence Confocal Microscopy". In: *Photochemistry and Photobiology* 77.5 (2003), pp. 550–555. ISSN: 1751-1097. DOI: 10.1562/0031-8655(2003)0770550MABOON2.0.CO2. URL: http://dx.doi.org/10.1562/0031-8655(2003)0770550MABOON2.0.CO2.

[215] P. Payeur and D. Desjardins. "Structured Light Stereoscopic Imaging with Dynamic Pseudo-random Patterns". In: *Proceedings of the 6th International Conference on Image Analysis and Recognition*. ICIAR '09. Halifax, Nova Scotia, Canada: Springer-Verlag, 2009, pp. 687–696. ISBN: 978-3-642-02610-2. DOI: 10.1007/978-3-642-02611-9_68. URL: http://dx.doi.org/10.1007/978-3-642-02611-9_68.

[216] J. Penne et al. "Time-of-Flight 3-D Endoscopy". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2009, pp. 467–474. DOI: 10.1007/978-3-642-04268-3_58.

[217] M. Perriollat, R. Hartley, and A. Bartoli. "Monocular Template-based Reconstruction of Inextensible Surfaces". In: *International Journal of Computer Vision* 95.2 (2011), pp. 124–137. ISSN: 1573-1405. DOI: 10.1007/s11263-010-0352-8. URL: https://doi.org/10.1007/s11263-010-0352-8.

[218] E. M. Petriu et al. "Object recognition using pseudo-random color encoded structured light". In: *Instrumentation and Measurement Technology Conference, 2000. IMTC 2000. Proceedings of the 17th IEEE*. Vol. 3. 2000, 1237 –1241 vol.3. DOI: 10.1109/IMTC.2000.848675.

[219] J. Pilet, V. Lepetit, and P. Fua. "Fast Non-Rigid Surface Detection, Registration and Realistic Augmentation". In: *International Journal of Computer Vision* 76 (2 2008), pp. 109–122. DOI: 10.1007/s11263-006-0017-9.

[220] D. Pizarro and A. Bartoli. "Feature-Based Deformable Surface Detection with Self-Occlusion Reasoning". In: *International Journal of Computer Vision* (2012), pp. 1–17. DOI: 10.1007/s11263-011-0452-0.

[221] J. L. Posdamer and M. D. Altschuler. "Surface measurement by space-encoded projected beam systems". In: *Computer Graphics and Image Processing* 18.1 (1982), pp. 1 –17. ISSN: 0146-664X. DOI: 10.1016/0146-664X(82)90096-X. URL: http://www.sciencedirect.com/science/article/pii/0146664X8290096X.

[222] R. P. K. Poudel, P. Lamata, and G. Montana. "Recurrent Fully Convolutional Neural Networks for Multi-slice MRI Cardiac Segmentation". In: *Reconstruction, Segmentation, and Analysis of Medical Images: First International Workshops, RAMBO 2016 and HVSMR 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 17, 2016, Revised Selected Papers*. Ed. by M. A. Zuluaga et al. Cham: Springer International Publishing, 2017, pp. 83–94. ISBN: 978-3-319-52280-7. DOI: 10.1007/978-3-319-52280-7_8. URL: https://doi.org/10.1007/978-3-319-52280-7_8.

[223] S. A. Prahl. *Tabulated molar extinction coefficient for hemoglobin in water*. 2017. URL: http://omlc.ogi.edu/spectra/hemoglobin/summary.html (visited on 12/14/2017).

[224] A. Prasoon et al. "Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural Network". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part II*. Ed. by K. Mori et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 246–253. ISBN: 978-3-642-40763-5. DOI: 10.1007/978-3-642-40763-5_31. URL: https://doi.org/10.1007/978-3-642-40763-5_31.

[225] P. Pratt et al. "An effective visualisation and registration system for image-guided robotic partial nephrectomy". In: *Journal of Robotic Surgery* 6.1 (2012), pp. 23–31. ISSN: 1863-2491. DOI: 10.1007/s11701-011-0334-z. URL: https://doi.org/10.1007/s11701-011-0334-z.

[226] P. Purkait and B. Chanda. "Image Upscaling Using Multiple Dictionaries of Natural Image Patches". In: *Computer Vision – ACCV 2012: 11th Asian Conference on Computer Vision, Daejeon, Korea, November 5-9, 2012, Revised Selected Papers, Part III*. Ed. by K. M. Lee et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 284–295. ISBN: 978-3-642-37431-9. DOI: 10.1007/978-3-642-37431-9_22. URL: https://doi.org/10.1007/978-3-642-37431-9_22.

[227] A. Radford, L. Metz, and S. Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks". In: *arXiv preprint arXiv:1511.06434* (2015).

[228] M. I. Rajab, M. S. Woolfson, and S. P. Morgan. "Application of region-based segmentation and neural network edge detection to skin lesions". In: *Computerized Medical Imaging and Graphics* 28.1 (2004), pp. 61 –68. ISSN: 0895-6111. DOI: https://doi.org/10.1016/S0895-6111(03)00054-5. URL: http://www.sciencedirect.com/science/article/pii/S0895611103000545.

[229] N. Ramanujam. "Fluorescence Spectroscopy of Neoplastic and Non-Neoplastic Tissues". In: *Neoplasia* 2.1 (2000), pp. 89 –117. ISSN: 1476-5586. DOI: https://doi.org/10.1038/sj.neo.7900077. URL: http://www.sciencedirect.com/science/article/pii/S1476558600800238.

[230] J. C. Ramella-Roman and S. A. Mathews. "Spectroscopic Measurements of Oxygen Saturation in the Retina". In: *IEEE Journal of Selected Topics in Quantum Electronics* 13.6 (2007), pp. 1697–1703. ISSN: 1077-260X. DOI: 10.1109/JSTQE.2007.911312.

[231] W. E. Reddick et al. "Automated segmentation and classification of multispectral magnetic resonance images of brain using artificial neural networks". In: *IEEE Transactions on Medical Imaging* 16.6 (1997), pp. 911–918. ISSN: 0278-0062. DOI: 10.1109/42.650887.

[232] T. E. Renkoski, U. Utzinger, and K. D. Hatch. "Wide-field spectral imaging of human ovary autofluorescence and oncologic diagnosis via previously collected probe data". In: *Journal of Biomedical Optics* 17 (2012), pp. 17 –17 –14. DOI: 10.1117/1.JBO.17.3.036003. URL: http://dx.doi.org/10.1117/1.JBO.17.3.036003.

[233] D. Roblyer et al. "Multispectral optical imaging device for in vivo detection of oral neoplasia". In: *Journal of biomedical optics* 13.2 (2008), p. 024019. ISSN: 1083-3668. DOI: 10.1117/1.2904658. URL: http://europepmc.org/articles/PMC3970814.

[234] O. Ronneberger, P. Fischer, and T. Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III*. Ed. by N. Navab et al. Cham: Springer International Publishing, 2015, pp. 234–241. ISBN: 978-3-319-24574-4. DOI: 10.1007/978-3-319-24574-4_28. URL: http://dx.doi.org/10.1007/978-3-319-24574-4_28.

[235] E. Rosten and T. Drummond. "Machine Learning for High-Speed Corner Detection". In: *Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I*. Ed. by A. Leonardis, H. Bischof, and A. Pinz. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 430–443. ISBN: 978-3-540-33833-8. DOI: 10.1007/11744023_34. URL: https://doi.org/10.1007/11744023_34.

[236] H. R. Roth et al. "Hierarchical 3D fully convolutional networks for multi-organ segmentation". In: *CoRR* abs/1704.06382 (2017). arXiv: 1704.06382. URL: http://arxiv.org/abs/1704.06382.

[237] A. Roy and S. Todorovic. "A Multi-scale CNN for Affordance Segmentation in RGB Images". In: *Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV*. Ed. by B. Leibe et al. Cham: Springer International Publishing, 2016, pp. 186–201. ISBN: 978-3-319-46493-0. DOI: 10.1007/978-3-319-46493-0_12. URL: https://doi.org/10.1007/978-3-319-46493-0_12.

[238] E. Rublee et al. "ORB: An efficient alternative to SIFT or SURF". In: *2011 International Conference on Computer Vision*. 2011, pp. 2564–2571. DOI: 10.1109/ICCV.2011.6126544.

[239] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. "Neurocomputing: Foundations of Research". In: (1988). Ed. by James A. Anderson and Edward Rosenfeld, pp. 696–699. URL: http://dl.acm.org/citation.cfm?id=65669.104451.

[240] J. Salvi, J. Pages, and J. Batlle. "Pattern codification strategies in structured light systems". In: *Pattern Recognition* 37.4 (2004), pp. 827 –849. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2003.10.002. URL: http://www.sciencedirect.com/science/article/pii/S0031320303003303.

[241] J. Salvi et al. "A state of the art in structured light patterns for surface profilometry". In: *Pattern Recognition* 43.8 (2010), pp. 2666 –2680. ISSN: 0031-3203. DOI: 10.1016/j.patcog.2010.03.004. URL: http://www.sciencedirect.com/science/article/pii/S003132031000124X.

[242] M. Salzmann and P. Fua. "Linear Local Models for Monocular Reconstruction of Deformable Surfaces". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.5 (2011), pp. 931–944. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.158.

[243] D. Scharstein, R. Szeliski, and R. Zabih. "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms". In: *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*. 2001, pp. 131–140. DOI: 10.1109/SMBV.2001.988771.

[244] C. Schmalz et al. "An endoscopic 3D scanner based on structured light". In: *Medical Image Analysis* 16.5 (2012), pp. 1063 –1072. DOI: 10.1016/j.media.2012.04.001. URL: http://www.sciencedirect.com/science/article/pii/S1361841512000473.

[245] J. Schmidhuber. "Learning Complex, Extended Sequences Using the Principle of History Compression". In: *Neural Comput.* 4.2 (Mar. 1992), pp. 234–242. ISSN: 0899-7667. DOI: 10.1162/neco.1992.4.2.234. URL: http://dx.doi.org/10.1162/neco.1992.4.2.234.

[246] W. Schroeder, K. M. Martin, and W. E. Lorensen. *The Visualization Toolkit (4th ed.)* Kitware, 2006. ISBN: 978-1-930934-19-1.

[247] C. J. Schuler et al. "A Machine Learning Approach for Non-blind Image Deconvolution". In: *2013 IEEE Conference on Computer Vision and Pattern Recognition*. 2013, pp. 1067–1074. DOI: 10.1109/CVPR.2013.142.

[248] R. L Schuler, P. E Kish, and C. A Plese. "Preliminary observations on the ability of hyperspectral imaging to provide detection and visualization of bloodstain patterns on black fabrics". In: *Journal of forensic sciences* 57.6 (2012), pp. 1562–1569. ISSN: 0022-1198. DOI: 10.1111/j.1556-4029.2012.02171.x. URL: https://doi.org/10.1111/j.1556-4029.2012.02171.x.

[249] J. Schweizer et al. "Hyperspectral imaging-A new modality for eye diagnostics". In: *Biomedical Engineering/Biomedizinische Technik* 57.SI-1 Track-P (), pp. 293–296.

[250] S. B. Serpico, L. Bruzzone, and F. Roli. "An experimental comparison of neural and statistical non-parametric algorithms for supervised classification of remote-sensing images". In: *Pattern Recognition Letters* 17.13 (1996). Special Issue on Non-conventional Pattern Analysis in Remote Sensing, pp. 1331–1341. ISSN: 0167-8655. DOI: https://doi.org/10.1016/S0167-8655(96)00090-6. URL: http://www.sciencedirect.com/science/article/pii/S0167865596000906.

[251] Q. Shan et al. "Fast Image/Video Upsampling". In: *ACM SIGGRAPH Asia 2008 Papers*. SIGGRAPH Asia '08. Singapore: ACM, 2008, 153:1–153:7. ISBN: 978-1-4503-1831-0. DOI: 10.1145/1457515.1409106. URL: http://doi.acm.org/10.1145/1457515.1409106.

[252] E. Shelhamer, J. Long, and T. Darrell. "Fully convolutional networks for semantic segmentation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.4 (2017), pp. 640–651. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2016.2572683.

[253] J. Shi and J. Malik. "Normalized cuts and image segmentation". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.8 (2000), pp. 888–905. ISSN: 0162-8828. DOI: 10.1109/34.868688.

[254] W. Shi et al. "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 1874–1883. DOI: 10.1109/CVPR.2016.207.

[255] S. C. Sides, J. A. Anderson, et al. "Comparison of three different methods to merge multiresolution and multispectral data- Landsat TM and SPOT panchromatic". In: *Photogrammetric Engineering and remote sensing* 57.3 (1991), pp. 295–303.

[256] R. Siegel, D. Naishadham, and A. Jemal. "Cancer statistics, 2012". In: *CA: A Cancer Journal for Clinicians* 62.1 (2012), pp. 10–29. ISSN: 1542-4863. DOI: 10.3322/caac.20138. URL: http://dx.doi.org/10.3322/caac.20138.

[257] K. Simonyan and A. Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *CoRR* abs/1409.1556 (2014). arXiv: 1409.1556. URL: http://arxiv.org/abs/1409.1556.

[258] N. Singh, D. G. Armstrong, and B. A. Lipsky. "Preventing foot ulcers in patients with diabetes". In: *JAMA* 293.2 (2005), pp. 217–228. DOI: 10.1001/jama.293.2.217. eprint: /data/journals/jama/4959/jcr40054.pdf. URL: +http://dx.doi.org/10.1001/jama.293.2.217.

[259] Irwin Sobel. "An Isotropic 3 3 Image Gradient Operator". In: (Feb. 2014).

[260] C. Sommer et al. "Ilastik: Interactive learning and segmentation toolkit". In: *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. 2011, pp. 230–233. DOI: 10.1109/ISBI.2011.5872394.

[261] B. S. Sorg et al. "Hyperspectral imaging of hemoglobin saturation in tumor microvasculature and tumor hypoxia development". In: *Journal of Biomedical Optics* 10.4 (2005), pp. 044004–044004–11. DOI: 10.1117/1.2003369. URL: http://dx.doi.org/10.1117/1.2003369.

[262] D. Stathakis and A. Vasilakos. "Satellite image classification using granular neural networks". In: *International Journal of Remote Sensing* 27.18 (2006), pp. 3991–4003. DOI: 10.1080/01431160600567779. eprint: http://dx.doi.org/10.1080/01431160600567779. URL: http://dx.doi.org/10.1080/01431160600567779.

[263] M. F. Stollenga et al. "Parallel Multi-dimensional LSTM, with Application to Fast Biomedical Volumetric Image Segmentation". In: *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*. NIPS'15. Montreal, Canada: MIT Press, 2015, pp. 2998–3006. URL: http://dl.acm.org/citation.cfm?id=2969442.2969574.

[264] D. Stoyanov et al. "Real-Time Stereo Reconstruction in Robotically Assisted Minimally Invasive Surgery". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Ed. by Tianzi Jiang et al. Vol. 6361. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, pp. 275–282. ISBN: 978-3-642-15704-2. DOI: 10.1007/978-3-642-15705-9_34. URL: http://dx.doi.org/10.1007/978-3-642-15705-9_34.

[265] J. Sun, Z. Xu, and H.-Y. Shum. "Image super-resolution using gradient profile prior". In: *2008 IEEE Conference on Computer Vision and Pattern Recognition*. 2008, pp. 1–8. DOI: 10.1109/CVPR.2008.4587659.

[266] J. Sun et al. "Image hallucination with primal sketch priors". In: *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.* Vol. 2. 2003, II–729–36 vol.2. DOI: 10.1109/CVPR.2003.1211539.

[267] J. Sun et al. "Large 3D free surface measurement using a mobile coded light-based stereo vision system". In: *Sensors and Actuators A: Physical* 132.2 (2006), pp. 460 –471. ISSN: 0924-4247. DOI: 10.1016/j.sna.2006.02.048. URL: http://www.sciencedirect.com/science/article/pii/S0924424706001944.

[268] Y. Swirski, Y. Y. Schechner, and T. Nir. "Variational stereo in dynamic illumination". In: *2011 International Conference on Computer Vision*. 2011, pp. 1124–1131. DOI: 10.1109/ICCV.2011.6126360.

[269] C. Szegedy et al. "Going Deeper with Convolutions". In: *CoRR* abs/1409.4842 (2014). arXiv: 1409.4842. URL: http://arxiv.org/abs/1409.4842.

[270] R. Szeliski. *Computer Vision: Algorithms and Applications*. 1st. New York, NY, USA: Springer-Verlag New York, Inc., 2010. ISBN: 1848829345, 9781848829343.

[271] Y. W. Tai et al. "Super resolution using edge prior and single image detail synthesis". In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2010, pp. 2400–2407. DOI: 10.1109/CVPR.2010.5539933.

[272] M. Takeda and K. Mutoh. "Fourier transform profilometry for the automatic measurement of 3-D object shapes". In: *Appl. Opt.* 22.24 (1983), pp. 3977–3982. DOI: 10.1364/AO.22.003977. URL: http://ao.osa.org/abstract.cfm?URI=ao-22-24-3977.

[273] M. F. Tappen and W. T. Freeman. "Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters". In: *Proceedings Ninth IEEE International Conference on Computer Vision*. 2003, 900–906 vol.2. DOI: 10.1109/ICCV.2003.1238444.

[274] C. R. Tracy et al. "Characterization of Renal Ischemia Using DLP Hyperspectral Imaging: A Pilot Study Comparing Artery-Only Occlusion Versus Artery and Vein Occlusion". In: *Journal of Endourology* 3.24 (2010), pp. 321–325. DOI: 10.1089/end.2009.0184.

[275] R. Tsai. "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses". In: *Robotics and Automation, IEEE Journal of* 3.4 (1987), pp. 323–344. ISSN: 0882-4967. DOI: 10.1109/JRA.1987.1087109.

[276] V. Tuchin. "Tissue Optics : Light Scattering Methods and Instruments for Medical Diagnosis". In: *IS11, SPIE, Bellingham, 2000* (2000). URL: http://ci.nii.ac.jp/naid/10015325325/en/.

[277] S. Ullman. *The Interpretation of Visual Motion*. MIT Press, 1979.

[278] A. O. Ulusoy, F. Calakli, and G. Taubin. "One-shot scanning using De Bruijn spaced grids". In: *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*. 2009, pp. 1786 –1792. DOI: 10.1109/ICCVW.2009.5457499.

[279] S. Valipour et al. "Recurrent Fully Convolutional Networks for Video Segmentation". In: *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2017, pp. 29–36. DOI: 10.1109/WACV.2017.11.

[280] V. Velanovich. "Laparoscopic vs open surgery". In: *Surgical Endoscopy* 14.1 (2000), pp. 16–21. ISSN: 1432-2218. DOI: 10.1007/s004649900003. URL: https://doi.org/10.1007/s004649900003.

[281] M. Visentini-Scarzanella, D. Stoyanov, and G.-Z. Yang. "Metric depth recovery from monocular images using Shape-from-Shading and specularities". In: *2012 19th IEEE International Conference on Image Processing*. 2012, pp. 25–28. DOI: 10.1109/ICIP.2012.6466786.

[282] F. Visin et al. "ReSeg: A Recurrent Neural Network-Based Model for Semantic Segmentation". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2016, pp. 426–433. DOI: 10.1109/CVPRW.2016.60.

[283] L. V. Wang and H.-I. Wu. "Introduction". In: *Biomedical Optics*. John Wiley & Sons, Inc., 2009, pp. 1–15. ISBN: 9780470177013. DOI: 10.1002/9780470177013.ch1. URL: http://dx.doi.org/10.1002/9780470177013.ch1.

[284] S. Wang et al. "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. 2012, pp. 2216–2223. DOI: 10.1109/CVPR.2012.6247930.

[285] T. Wang et al. "End-to-end text recognition with convolutional neural networks". In: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. 2012, pp. 3304–3308.

[286] Q. Wei, N. Dobigeon, and J. Y. Tourneret. "Bayesian Fusion of Multi-Band Images". In: *IEEE Journal of Selected Topics in Signal Processing* 9.6 (2015), pp. 1117–1127. ISSN: 1932-4553. DOI: 10.1109/JSTSP.2015.2407855.

[287] Q. Wei et al. "Hyperspectral and Multispectral Image Fusion Based on a Sparse Representation". In: *IEEE Transactions on Geoscience and Remote Sensing* 53.7 (2015), pp. 3658–3668. ISSN: 0196-2892. DOI: 10.1109/TGRS.2014.2381272.

[288] L. Weitzel et al. "3D: The next generation near-infrared imaging spectrometer". In: *Astronomy and Astrophysics Supplement Series* 119.3 (1996), pp. 531–546.

[289] A. J. Welch, M. J.C. van Gemert, and Willem M. Star. "Definitions and Overview of Tissue Optics". In: *Optical-Thermal Response of Laser-Irradiated Tissue*. Ed. by Ashley J. Welch and Martin J.C. van Gemert. Dordrecht: Springer Netherlands, 2011, pp. 27–64. ISBN: 978-90-481-8831-4. DOI: 10.1007/978-90-481-8831-4_3. URL: https://doi.org/10.1007/978-90-481-8831-4_3.

[290] W. L. Wolfe. *Introduction to imaging spectrometers / William L. Wolfe*. English. SPIE Bellingham, Wash, 1997, xiii, 148 p. : ISBN: 0819422606.

[291] C. Wu, S. G. Narasimhan, and B. Jaramaz. "A Multi-Image Shape-from-Shading Framework forNear-Lighting Perspective Endoscopes". In: *International Journal of Computer Vision* 86.2 (2010), pp. 211–228. ISSN: 1573-1405. DOI: 10.1007/s11263-009-0207-3. URL: https://doi.org/10.1007/s11263-009-0207-3.

[292] T. T. Wu and J. Y. Qu. "Optical imaging for medical diagnosis based on active stereo vision and motion tracking". In: *Opt. Express* 15.16 (2007), pp. 10421–10426. DOI: 10.1364/OE.15.010421. URL: http://www.opticsexpress.org/abstract.cfm?URI=oe-15-16-10421.

[293] C. Wust and D. W. Capson. "Surface profile measurement using color fringe projection". In: *Mach. Vision Appl.* 4.3 (July 1991), pp. 193–203. ISSN: 0932-8092. DOI: 10.1007/BF01230201. URL: http://dx.doi.org/10.1007/BF01230201.

[294] Z. Xiong, X. Sun, and F. Wu. "Robust Web Image/Video Super-Resolution". In: *IEEE Transactions on Image Processing* 19.8 (2010), pp. 2017–2028. ISSN: 1057-7149. DOI: 10.1109/TIP.2010.2045707.

[295] Z. Xu et al. "Smart pixel : photonic mixer device (PMD) ; new system concept of a 3D-imaging camera-on-a-chip". In: *Proceedings : M2VIP '98; Nanjing, China, 10th - 12th September 1998*. Ed. by Edmund H. M. Cheung. 1998, pp. 259 –264.

[296] G. Yahav, G. J. Iddan, and D. Mandelboum. "3D Imaging Camera for Gaming Application". In: *2007 Digest of Technical Papers International Conference on Consumer Electronics*. 2007, pp. 1–2. DOI: 10.1109/ICCE.2007.341537.

[297] C.-Y. Yang, C. Ma, and M.-H. Yang. "Single-Image Super-Resolution: A Benchmark". In: *Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV*. Ed. by D. Fleet et al. Cham: Springer International Publishing, 2014, pp. 372–386. ISBN: 978-3-319-10593-2. DOI: 10.1007/978-3-319-10593-2_25. URL: https://doi.org/10.1007/978-3-319-10593-2_25.

[298] J. Yang et al. "Image Super-Resolution Via Sparse Representation". In: *IEEE Transactions on Image Processing* 19.11 (2010), pp. 2861–2873. ISSN: 1057-7149. DOI: 10.1109/TIP.2010.2050625.

[299] V. X. D. Yang et al. "A multispectral fluorescence imaging system: design and initial clinical tests in intra-operative Photofrin-photodynamic therapy of brain tumors". In: *Lasers in surgery and medicine* 32.3 (2003), pp. 224–232. ISSN: 0196-8092. DOI: 10.1002/lsm.10131. URL: http://dx.doi.org/10.1002/lsm.10131.

[300] S. Y. Yeung, H. T. Tsui, and A. Yim. "Global Shape from Shading for an Endoscope Image". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI'99: Second International Conference, Cambridge, UK, September 19-22, 1999. Proceedings*. Ed. by Chris Taylor and Alain Colchester. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 318–327. ISBN: 978-3-540-48232-1. DOI: 10.1007/10704282_35. URL: https://doi.org/10.1007/10704282_35.

[301] N. Yokoya, T. Yairi, and A. Iwasaki. "Coupled Nonnegative Matrix Factorization Unmixing for Hyperspectral and Multispectral Data Fusion". In: *IEEE Transactions on Geoscience and Remote Sensing* 50.2 (2012), pp. 528–537. ISSN: 0196-2892. DOI: 10.1109/TGRS.2011.2161320.

[302] F. Yu and V. Koltun. "Multi-Scale Context Aggregation by Dilated Convolutions". In: *CoRR* abs/1511.07122 (2015). arXiv: 1511.07122. URL: http://arxiv.org/abs/1511.07122.

[303] L. Zelnik-manor and P. Perona. "Self-tuning spectral clustering". In: *Advances in Neural Information Processing Systems 17*. Ed. by L.K. Saul, Y. Weiss, and L. Bottou. MIT Press, 2005, pp. 1601–1608. URL: http://papers.nips.cc/paper/2619-self-tuning-spectral-clustering.pdf.

[304] L. Zhang, B. Curless, and S. M. Seitz. "Rapid shape acquisition using color structured light and multi-pass dynamic programming". In: *3D Data Processing Visualization and Transmission, 2002. Proceedings. First International Symposium on*. 2002, pp. 24 –36. DOI: 10.1109/TDPVT.2002.1024035.

[305] Y. Zhang et al. "Visible and near-infrared spectroscopy for distinguishing malignant tumor tissue from benign tumor and normal breast tissues in vitro". In: *Journal of Biomedical Optics* 18 (2013), pp. 18 –18 –8. DOI: 10.1117/1.JBO.18.7.077003. URL: http://dx.doi.org/10.1117/1.JBO.18.7.077003.

[306] Z. Zhang. "A flexible new technique for camera calibration". In: *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.11 (2000), pp. 1330 –1334. ISSN: 0162-8828. DOI: 10.1109/34.888718.

[307] S. Zheng et al. "Conditional Random Fields as Recurrent Neural Networks". In: *2015 IEEE International Conference on Computer Vision (ICCV)*. 2015, pp. 1529–1537. DOI: 10.1109/ICCV.2015.179.

[308] J. Zhong et al. "Remote Sensing Image Fusion with Convolutional Neural Network". In: *Sensing and Imaging* 17.1 (2016), p. 10. ISSN: 1557-2072. DOI: 10.1007/s11220-016-0135-6. URL: https://doi.org/10.1007/s11220-016-0135-6.

[309] S. Zhou et al. "Exploiting Local Structures with the Kronecker Layer in Convolutional Networks". In: *CoRR* abs/1512.09194 (2015). arXiv: 1512.09194. URL: http://arxiv.org/abs/1512.09194.

[310] G. Zonios et al. "Diffuse reflectance spectroscopy of human adenomatous colon polyps in vivo". In: *Appl. Opt.* 38.31 (1999), pp. 6628–6637. DOI: 10.1364/AO.38.006628. URL: http://ao.osa.org/abstract.cfm?URI=ao-38-31-6628.

[311] K. J. Zuzak et al. "Active DLP Hyperspectral Illumination: A Noninvasive, in Vivo, System Characterization Visualizing Tissue Oxygenation at Near Video Rates". In: *Analytical Chemistry* 83.19 (2011). PMID: 21842837, pp. 7424–7430. DOI: 10.1021/ac201467v. eprint: http://dx.doi.org/10.1021/ac201467v. URL: http://dx.doi.org/10.1021/ac201467v.

[312] K. J. Zuzak et al. "Characterization of a Near-Infrared Laparoscopic Hyperspectral Imaging System for Minimally Invasive Surgery". In: *Analytical Chemistry* 79.12 (2007). PMID: 17492839, pp. 4709–4715. DOI: 10.1021/ac070367n. eprint: http://dx.doi.org/10.1021/ac070367n. URL: http://dx.doi.org/10.1021/ac070367n.

[313] K. J. Zuzak et al. *Hyperspectral imaging utilizing LCTF and DLP technology for surgical and clinical applications*. PMID: 21842837. 2009. DOI: 10.1117/12.816279. URL: http://dx.doi.org/10.1117/12.816279.

[314] K. J. Zuzak et al. "Visible and infrared hyperspectral visualization of normal and ischemic tissue". In: *Proceedings of the First Joint BMES/EMBS Conference*. Vol. 2. 1999, 1118 vol.2–. DOI: 10.1109/IEMBS.1999.804280.

[315] K. J. Zuzak et al. *Visible spectroscopic imaging studies of normal and ischemic dermal tissue*. 2000. DOI: 10.1117/12.384958. URL: http://dx.doi.org/10.1117/12.384958.

Image Reuse Permissions

IEEE Images

All images replicated in this thesis from IEEE publications have been referenced according to the following statement from the IEEE:

"The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, the following statement may be used as a permission grant: In the case of illustrations or tabular material, we require that the copyright line ©[Year of original publication] IEEE appear prominently with each reprinted figure and/or table".

The following figures were reproduced from IEEE images:

- Fig. 5.4
- Fig. 5.5
- Fig. 6.4
- Fig. 6.5
- Fig. 6.6
- Fig. 6.7

Springer Images

The following figures were reproduced from Springer images with permissions:

- Fig. 2.2(a), license number: 4359201289130
- Fig. 2.4(d), license number: 4359210596634
- Fig. 3.1, license number: 4359210764884
- Fig. 3.2(b)(c)(e), license number: 4359210764884

- Fig. 3.2(a)(f)(g)(h), license number: 4359211066395
- Fig. 3.5, license number: 4359210764884
- Fig. 3.7, license number: 4359210764884
- Fig. 4.3, license number: 4359210764884
- Fig. 4.4, license number: 4359211066395
- Fig. 4.5, license number: 4359211066395
- Fig. 5.1, license number: 4359211498523
- Fig. 5.7, license number: 4359211361044
- Fig. 5.9, license number: 4359211498523
- Fig. 6.8, license number: 4359211498523
- Fig. 6.9, license number: 4359211498523
- Fig. 6.10, license number: 4359211361044
- Fig. 6.11, license number: 4359211361044
- Fig. 6.12, license number: 4359211498523
- Fig. 6.14, license number: 4359211498523

Elsevier Images

The following figures were reproduced from Elsevier images:

- Fig. 2.2(c)(d), license number: 4359220766980
- Fig. 2.3, license number: 4359220766980
- Fig. 2.4(b)(c), license number: 501404556

Images from Cambridge University Press

Cambridge University Press grants permission freely for the reproduction in another work of a short prose extract (less than 400 words), a single figure or a single table in which it holds rights (see the important caveat in the Notes below). In such cases a request for permission need not be submitted, but the reproduced material must be accompanied by a full citation of the original source.

The following figures were reproduced from images from Cambridge University Press:

- Fig. 2.1