Analysis of MRI for Knee Osteoarthritis using Machine Learning

Claire Rebecca Donoghue

A dissertation submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

of

Imperial College London

August 2013

Department of Computing

Imperial College London

To Sheila, Rose and James

Declaration of originality

I declare that the work presented in this thesis is my own, unless specifically acknowledged.

Claire Donoghue

Copyright

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.

Abstract

Approximately 8.5 million people in the UK (13.5% of the population) have osteoarthritis (OA) in one or both knees, with more than 6 million people in the UK suffering with painful osteoarthritis of the knee. In addition, an ageing population implies that an estimated 17 million people (twice as many as in 2012) are likely to be living with OA by 2030. Despite this, there exists no disease modifying drugs for OA and structural OA in MRI is poorly characterised.

This motivates research to develop biomarkers and tools to aid osteoarthritis diagnosis from MRI of the knee. Previously many solutions for learning biomarkers have relied upon hand-crafted features to characterise and diagnose osteoarthritis from MRI. The methods proposed in this thesis are scalable and use machine learning to characterise large populations of the OAI dataset, with one experiment applying an algorithm to over 10,000 images. Studies of this size enable subtle characteristics of the dataset to be learnt and model many variations within a population.

We present data-driven algorithms to learn features to predict OA from the appearance of the articular cartilage. An unsupervised manifold learning algorithm is used to compute a low dimensional representation of knee MR data which we propose as an imaging marker of OA. Previous metrics introduced for OA diagnosis are loosely based on the research communities intuition of the structural causes of OA progression, including morphological measures of the articular cartilage such as the thickness and volume. We demonstrate that there is a strong correlation between traditional morphological measures of the articular cartilage and the biomarkers identified using the manifold learning algorithm that we propose ($R^2 = 0.75$). The algorithm is extended to create biomarkers for different regions and sequences. A combination of these markers is proposed to yield a diagnostic imaging biomarker with superior performance. The diagnostic biomarkers presented are shown to improve upon hand-crafted morphological measure of disease status presented in the literature, a linear discriminant analysis (LDA) classification for early stage diagnosis of knee osteoarthritis results with an AUC of 0.9.

From the biomarker discovery experiments we identified that intensity based affine registration of knee MRIs is not sufficiently robust for large scale image analysis, approximately 5% of these registrations fail. We have developed fast algorithms to compute robust affine transformations

of knee MRI, which enables accurate pairwise registrations in large datasets. We model the population of images as a non-linear manifold, a registration is defined by the shortest geodesic path over the manifold representation. We identify sources of error in our manifold representation and propose fast mitigation strategies by checking for consistency across the manifold and by utilising multiple paths. These mitigation strategies are shown to improve registration accuracy and can be computed in less than 2 seconds with current architecture.

Acknowledgements

I would first like to express my sincere thanks to my supervisors; Daniel Rueckert, Anthony Bull and Anil Rao for all of their enthusiasm, encouragement and advice. It has been a pleasure to work on this project.

I would like to thank all of my collaborators, colleagues and friends at Imperial College for the companionship, the support and the celebrations. Specifically to those who have or are working in office 302 and the members of the BioMedIA group for providing a stimulating working environment.

I would like to thank all of my friends. Above all, I would like to thank my parents (Charles and Judith), my brother (Alex) and my grandparents (James, Rose and Sheila) for a lifetimes worth of support and encouragement. And thanks to my boyfriend (Crispin) for 10 years of happiness and inspiration.

Contents

1 Introduction **23**

1.1 Motivation 23

1.2 Biomarkers Discovery 26

1.2.1 Machine Learning 28

1.3 Registration 29

1.3.1 Intensity based Affine Registration of the Knee 29

1.4 Thesis Contributions 31

1.4.1 Learning Osteoarthritis Diagnostic Markers 31

1.4.2 Affine Registration of Knee MRI 32

1.5 Thesis Overview 32

2 Background **35**

2.1 Introduction 35

2.2 Machine Learning 36

2.2.1 Terminology and Notation 37

2.2.2 Classification 38

2.2.3 Classification Performance Metrics 43

2.2.4 Cross Validation 45

2.2.5 Regression 47

2.2.6 Manifold Learning 49

2.2.7 Non-linear Manifold Learning Algorithms 51

2.3 Registration 55

2.3.1 Definitions and Basic Terminology 56

2.3.2 Transformation Models 57

2.3.3 Similarity Metrics 63

2.3.4 Optimisation Methods 67

2.3.5 Measuring registration accuracy 69

2.4 Knee Osteoarthritis 71

2.4.1 Knee Anatomy 72

2.4.2 Osteoarthritis Definition 74

2.4.3 The Osteoarthritis Initiative 74

2.4.4 MR Imaging of Osteoarthritis 77

2.4.5 MR Positioning 77

2.4.6 MR Sequences 77

2.4.7 Quantifying Osteoarthritis 78

2.4.8 Osteoarthritis Imaging Measures and Biomarkers 82

2.4.9 Summary 88

3 Manifold Learning Based Regression of Clinical Variables 90

3.1 Introduction 91

3.2 Method 92

3.2.1 Overview 92

3.2.2 Registration 94

3.2.3 Region of Interest (ROI) Definition 95

3.2.4 Manifold Learning 96

3.2.5 Predictor of Cartilage Morphology 98

3.2.6 Incorporating Multiple MR Sequences 99

3.3 Experiments 100

3.3.1 Data 100

3.3.2 Selecting the Graph Neighbourhood 102

3.3.3 Selecting Number of Predictors 103

3.3.4 Results 104

3.4 Discussion 106

3.4.1 Contributions 107

4 Osteoarthritis Diagnosis on Low Dimensional Embeddings 109

4.1 Introduction 110

4.1.1 Discovery of Novel Biomarkers For OA 110

4.2 Method 116

4.2.1 Dataset 116

4.2.2 Registration 117

4.2.3 Approximate Region of Interest Selection 119

4.2.4 Laplacian Eigenmap Embedding 120

4.2.5 Combining Multi-Region and Multi-Sequence Data 123

4.3 Experiments 126

4.3.1 LDA Classification Experiments 127

4.3.2 Results 128

4.3.3 Analysis 132

4.3.4 Visualisation of Laplacian Eigenmap Embedding 134

4.4 Conclusion and Discussion 136

4.4.1 Contribution Of Work 138

4.4.2 Limitations 138

4.4.3 Future Work 139

5 Accurate Global Geodesic Registrations 140

5.1 Introduction 141

5.1.1 Motivation 141

5.1.2 Preliminary Experiment 144

5.1.3 Intermediate Template Selection 144

5.2 Geodesic Registration Framework 146

5.2.1 Overview 146

5.2.2 Representation of the population 147

5.2.3 Edge Weights 148

5.2.4 Connectivity of k-Nearest Neighbour Graph 148

5.2.5 Intermediate Transformations 149

5.2.6 Geodesic Pairwise Registrations 149

5.2.7 Simple Refinement of Registrations 150

5.3 Experiments 152

5.3.1 Data 152

5.3.2 Implementation Details 154

5.3.3 Results and Analysis 156

5.4 Discussion 158

5.4.1 Limitations 160

5.4.2 Future Work 160

6 Fast and Accurate Global Geodesic Registrations 162

6.1 Introduction 163

6.1.1 Notation 164

6.1.2 Sources of Registration Error in Geodesic Registration 164

6.1.3 Strategies to Reduce Registration Error 166

6.1.4 Contributions 166

6.2 Increasing Geodesic Registration Accuracy through Refinements 167

6.2.1 Graph Construction for Geodesic Registration 167

6.2.2 Registration Fusion : Fusing Registration on Multiple Paths 168

6.2.3 Cycle Consistency : Globally consistent transformations 171

6.3 Experiments 175

6.3.1 Data 175

6.3.2 Results and Analysis 176

6.4 Discussion 184

6.4.1 Computational Cost 185

6.5 Future Work 186

7 Conclusion **187**

7.1 Contributions 187

7.1.1 Diagnostic Biomarkers 187

7.1.2 Robust and Efficient Affine registration of the Knee 188

7.2 Future Work 189

7.2.1 Diagnostic Biomarkers 190

7.2.2 Robust and Efficient Affine registration of the Knee 190

8 Publications **192**

Bibliography **194**

List of Tables

1.1 Details of existing imaging biomarkers, all are based upon the community's current understanding of osteoarthritis progression. †where SQ=semi-quantitative and Q=quantitative 27

2.1 Confusion Matrix 44

2.2 MR sequences, their abbreviations, intended uses and example MRI. 80

2.3 K&L scale for OA grading 83

2.4 Summary state-of-the-art OA biomarkers. "cartilage longevity" is a combination of biochemical (CTX-II) and imaging measures (cartilage volume, area, thickness, congruity, roughness, and homogeneity). 86

3.1 Results of multiple linear regression, using the co-ordinates from the manifold as predictor for known quantitative clinical variables of cartilage morphology published in Eckstein et al. (2009) , where a neighbourhood size is chosen for each manifold individually. p is always $< 0.1e - 13$ for the R^2 values quoted. . . 106

4.1 Previous contributions for developing OA biomarkers, they are all measures of cartilage morphology computed from segmentations. These contributions are hand-crafted and based upon the communities understanding of OA. 111

4.2 Regions of interest as defined by Wirth and Eckstein (2008); Eckstein et al. (2006a) 122

4.3 Composite regions of interest 125

4.4 Results for Sag 3D DESS and Cor T1 3D FLASH sequences: area under receiver operator curve from LDA, separating data between K&L grades listed in the separated columns. The top half of the table contains the three methods presented for combining the regions of interest for the specific sequences. The bottom half of the table contains an ordered list of the top five regions for classifying early OA for both MR sequences. * indicates those ROIs which are composites of other region of interest 131

4.5 Results for ensemble methods for all regions and both sequences. Results reported using area under receiver operator curve from LDA, separating data between K&L grades listed in the separated columns. 132

5.1 Registration quality of pairwise registration between subjects with very high BMI (a and b) and very low BMI (c and d). Registration qualitative success indicated by colour of square; green indicates a successful registration, orange indicates the registration had limited success with the results being ambiguous and red indicates the registration failed. Quantitative success measured by similarity metrics cross correlation and normalised mutual information are also displayed. 144

5.2 Mean and standard deviation of registration error and the percentage of failed registrations. A failed registration is considered to have an error of greater than 10mm. 157

6.1 Statistics of errors (mean and standard deviation) in mm where $k = 60$ 179

List of Figures

1.1 Example MR image of the knee for a healthy subject from the OAI dataset. The MR sequence is a sagittal 3D DESS. 25

2.1 A linear classifier of a single feature is represented in sub-figure 2.1 (a). The left curve constitutes the sampling distribution of subjects from the negative class, the curve on the right is the sampling distribution of subjects from the positive class. A separation threshold (green arrow) classifies a proportion of both sub-populations incorrectly; either false positive (FP, pink area) or false negative (FN, light blue area), these proportion vary based upon the location of the threshold. Sub-figure 2.1 (b) shows the corresponding ROC curve. The point and green arrow represents the separation threshold in sub-figure 2.1 (a) †. 46

2.2 Example of a rigid transformation. 59

2.3 Example of an affine transformation. 60

2.4 Registration Demo: A source image is being registered and warped to a target image. The source image is displayed in its native co-ordinate space, after an affine registration, a B-spline non-rigid freeform deformation and finally with the deformation grid. 64

2.5 Caption for LOF 75

2.6 Diagram of a normal knee on the left compared with an osteoarthritic degree on the right†. 76

2.7 Photo of MRI knee-imaging coil with liners and pad †. 78

2.8 Examples of the MR images acquired in the coronal plane. The images are of a healthy candidates right knee at baseline (id=9093584). 79

2.9 Examples of the MR images acquired in the sagittal plane. The images are of a healthy candidates right knee at baseline (id=9093584). 81

3.1 Pipeline of the framework for a single MR sequence. In step a), the region of interest has been identified in every image using registration (sections 3.2.2 and 3.2.3). Steps b) and c) depict the knn graph construction, where edges of low similarity are discarded. The spectral embedding is represented by step d) (section 3.2.4). Multiple linear regression is used to learn a linear combination of embedding coordinates to predict clinical data in step e) (section 3.2.5). . . . 93

3.2 Weight bearing region of the medial femoral condyle (cMF) region of interest shown on the reference image. 96

3.3 Overview of manifold learning. Laplacian Eigenmap embedding finds an appropriate low dimensional embedding $\boldsymbol{Y}$ from input images $\boldsymbol{X}$, by representing all images based on their pairwise similarities $\boldsymbol{W}$, and computing the k-nearest neighbours such that connections are retained for most similar images to find $\boldsymbol{W'}$. $\boldsymbol{Y}$ is a low dimensional embedding such that images which are represented as similar in $\boldsymbol{W'}$ are also close in $\boldsymbol{Y}$. , ... 97

3.4 Schematic of registration pipeline used to incorporate multiple MR sequences using the non-rigid components of the principle sequence. 101

3.5 R^2 computed from multiple linear regression using the first 14 eigenvectors as predictors for quantitative variable mean cartilage thickness (excluding denude bone area) with a varying neighbourhood of the similarity graph (where $k < 300$). This has been performed for five different MR sequences. Statistically insignificant regions are 'greyed-out', such that they can be identified and ignored. 103

3.6 Mean R^2 (a local average of R^2 over a range of $k \pm 5$) computed from multiple linear regression with a variable number of predictors (m). This is visualised for five MR sequences. When the rate of change of R^2 reduces the extra information being contributed by the predictors is having less influence on the prediction result. 104

3.7 MR images embedded in a 2D manifold. The different symbols describe the OA status of the knee. 105

4.1 Combining multi-region and multi-sequence data workflow 115

4.2 Sample of knee MR sequences from the OAI dataset: Sag 3D DESS (left) and COR T1 3D FLASH (right). 117

4.3 Registration pipeline; MRI of the same subject acquired at the same time-point are rigidly registered to the corresponding Sag 3D DESS image. All Sag 3D DESS images are registered non-rigidly to the reference image. 118

4.4 Mean image, $\boldsymbol{I}^{\alpha}_{mean}$, where $n = 1131$; where α is sequence Sag 3D DESS on right, Cor T1 FLASH on left 120

4.5 Regions of interest as defined and illustrated in Wirth and Eckstein (2008). . . 121

4.6 Sag 3D DESS MR image with dilated Region of Interest (ROI) for the lateral tibia (yellow), medial tibia (red), central lateral femur (blue) and central medial femur (green). 122

4.7 Grid search parameter selection visualisation for MR sequences; 4.6 (a) SAG 3D DESS and 4.6 (a) Cor T1 3D FLASH, to choose suitable parameters for the number of nearest neighbours in the graph required for nearest neighbourhood analysis and the intrinsic dimensionality of the manifold embedding. The metric for success is the AUC over the ROC which measures the overall success of the classifier. The measure is computed for each regional manifold and the mean of this score is computed. 129

4.8 ROC curve for LDA classification of *Ensemble by Laplacian eigenmap embedding*$\mathcal{M}$ (LEE), *Ensemble by PCA* (PCA) and *Composite of all regions of interest* for the sequence Sag 3D DESS (since it outperformed sequence Cor 3D T1 Flash for early diagnosis). 133

4.9 Classification results including correct rate, sensitivity and specificity for three different approaches for combining regional for the Sag 3D DESS sequence; *Ensemble by Laplacian eigenmap embedding*$\mathcal{M}$ (LEE), *Ensemble by PCA* (PCA) and *composite of all regions* . Each circle represents a sample, specificity can be read by the proportion of the left (green) bar shaded, sensitivity by the percentage of the right (red) bar shaded and the classification rate is the percentage of both of the shaded sections of the bars. 135

4.10 Visualisation of ensemble of Laplacian eigenmap embeddings, displaying only the first two eigenvectors, based on ascending non-zero eigenvalue order. It should be remembered that this visualisation is only illustrative of the population in two dimensions, whilst classification occurs in a low dimensional manifold space, dimensionality would be greater than two. The visualisation shows the population of knee MRI focusing on the articular cartilage for sequence a) Sag 3D DESS and b) Cor T1 3D FLASH c) both sequences 137

5.1 MRI of four subjects right knee, subjects were selected for having the highest and lowest BMIs in the database, the BMI of each are as follows a) 47.7, b) 45.4, c) 16.9, d) 17.6. 143

5.2 Registration scheme of standard sparse graphical registration 147

5.3 Illustration of accumulation of affine registration error and the refinement step to reduce errors 151

5.4 Histograms of the registration error (mean Euclidean distance between a set of landmarked points) distribution for all pairwise registrations 156

5.5 Histograms of the registration error distribution of 95th-100th percentile of ordered errors, showing the worst 5% of affine registrations for each approach. . . 157

5.6 A quantitative assessment of both direct registration and geodesic registration with refinement. Each point on the graph represents a registration error. Points are red when the direct registration error is lower than geodesic registration with refinement error. Points are green when the direct registration error is higher than geodesic registration with refinement error. Points are orange when both errors associated with both methods are approximately equivalent. 158

5.7 Qualitative example of inaccurate registrations improved by geodesic framework. 159

6.1 A registration between a source and target image across the shortest path in the graph (6.1 (a)) is used to represent a path upon the empirical manifold as depicted in 6.1 (b). However this method incurs two types of registration errors as described in subfigures 6.1 (c) and 6.1 (d). 165

6.2 Martins and Pascoal (2003) is used to compute the multiple shortest paths in a graphs, these paths are then fused to find a registration between two subjects using a Frechet mean. 6.2 (a), 6.2 (b) and 6.2 (c) show all the paths selected by the multiple shortest paths algorithm. 6.2 (d) shows the three shortest paths represented as a loopless tree and the frechet mean registration fusion. 172

6.3 Cycle consistency 173

6.4 Mean registration error in mm for values of k, where the number of paths $\mathcal{K} = 3$ 177

6.5 Graphs illustrating the performance in terms of computational speed and accuracy for the method presented in this chapter (Geodesic registration with cycle consistency and registration fusion) 180

6.6 These box plots show the distribution of errors for 3 registration methods, where $n = 2743$. Note the scales on these plots are different to illustrate the vast improvement of our method upon direct registration but more subtle improvement upon geodesic registration. 181

6.7 Pairwise comparison of registrations methods; direct registration versus the method proposed here, registrations to the upper-left of the line have reduced error with geodesic registration with cycle consistency and registration fusion, registrations to the lower-right have reduced error using direct registration. . . . 182

6.8 Box plot of the presented method without (left) and with (right) a refinement step. The results from the algorithm presented in this chapter are considered to be an approximation of an affine registration. The refinement step (as presented in chapter 5) is a further affine registration after the registration algorithm with the expectation of improving the accuracy. It can be seen that the accuracy decreases after the refinement, this suggests that a refinement step after the method is of no value. 183

6.9 Example of registration using direct registration and geodesic registration with cycle consistency and registration fusion. The images on the left are the target and source to be registered. Images on the right show the results of registration with the contours of the transformed source overlaid on the target. 184

Chapter 1

Introduction

1.1 Motivation

Osteoarthritis (OA) is a debilitating and prevalent pathology of the joints. It is a multifactorial disease which results in constant pain for 71% of sufferers in the UK (Arthritis Care (2012)). Osteoarthritis is structurally characterized by loss of articular cartilage within synovial joints in addition to hypertrophy of bone, i.e. osteophytes and subchondral bone sclerosis and thickening of the capsule. Clinically, the condition is characterized by joint pain, tenderness, limitation of movement, crepitus, occasional effusion, and some local inflammation. Risk factors for knee OA have been suggested to include ageing (DeGroot et al. (2004)), being female (Srikanth et al. (2005)), obesity (Lohmander et al. (2009)), history of knee injury (Wildera et al. (2002)), genetic factors (Neame et al. (2004)), smoking (Amin et al. (2007)) and altered biomechanics due to malalignment (Sharma et al. (2001)). It can occur in any joint but is most common in the hip, knee, the joints of the hand and foot and spine.

OA affects a large number of people in the western world, Arthritis Care (2012) approximate that 8.5 million people in the UK have OA in one or both knees, with more than 6 million people in the UK suffering with painful osteoarthritis of the knee. With a population of 62.5 million people, this is a sizeable proportion of the UK population. Due to an ageing population, an estimated 17 million people (twice as many as in 2012) are likely to be living with OA by

2030 (Arthritis Care (2012)).

There is a large economic and societal burden associated with osteoarthritis: the total cost to the UK economy is estimated at 1% of the gross national product and an estimated 36 million working days are lost as a result of osteoarthritis (Arthritis Care (2012)). Due to the likely increased prevalence of OA attributed to an ageing population, a large socio-economic impact in the future can be expected (Felson et al. (2000); March and Bachmeier (1997)). Presently it is a leading cause of disability and work limitation among adults resulting in enormous costs to society (Felson et al. (2000)), both through lost working hours and healthcare expenses. Approximately 77,578 primary knee replacements were performed in England and Wales in 2011-2012 for osteoarthritis (which makes up 98% of all primary knee replacements) (National Joint Registry (NJR) (2012)). Approximately 33% of people with osteoarthritis retire early, stop working or reduce their working hours (Arthritis Care (2012)). These reduced working hours contributes to arthritis being the leading cause of disability living allowance provided for under 65 years old from the Department for Work and Pensions (2012).

Not only does osteoarthritis have a very significant societal impact with regards to the economy but OA often has substantial impact upon sufferer's daily routines and overall quality of life. The findings of Arthritis Care (2012) indicate that osteoarthritis has a detrimental impact on an individual's social life, due to simple tasks such as getting in and out of a chair presenting challenges. With 20% of sufferers abstaining from holidays, leisure activities and hobbies; 13% of OA sufferers find socialising and meeting friends a challenge; and 13% of sufferers experiencing unbearable pain on a daily basis (Arthritis Care (2012)).

The knee is the focus of the work presented here. Osteoarthritis of the knee is extremely debilitating due to large mechanical stress experienced because of high leverage of the joint. The knee structure is strictly two joints, the tibiofemoral joint and the patellofemoral joint. The work presented here addresses osteoarthritis of the tibiofemoral joint, which is the articulating joint of the femoral (thigh bone) and tibial (shin bone) bone. It should be noted however despite the focused scope with regards to anatomy, the methodology and techniques described could be applied to other joints associated with osteoarthritis including the patellofemoral joint,

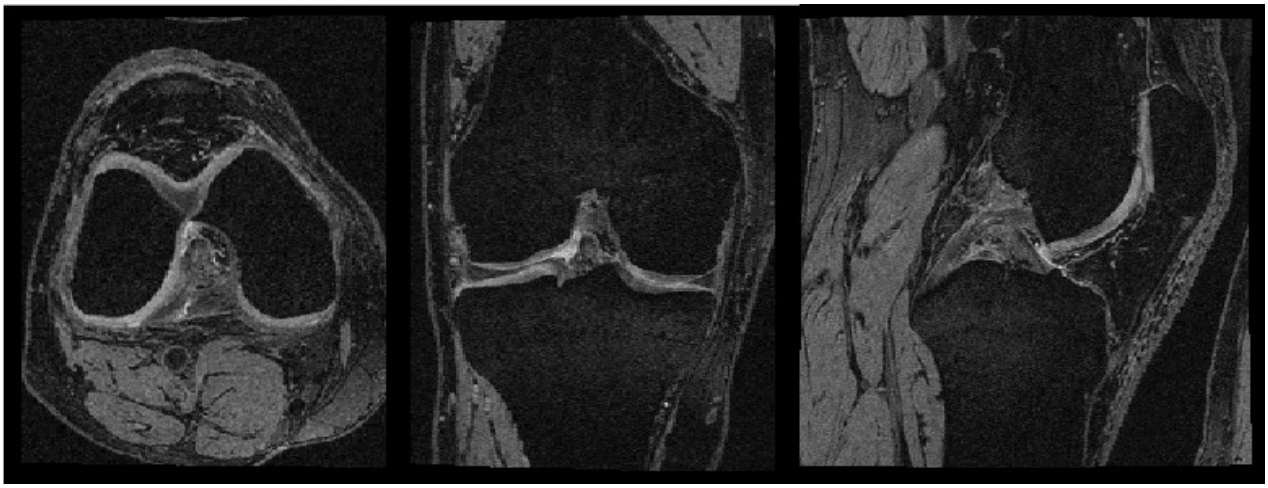


Figure 1.1: Example MR image of the knee for a healthy subject from the OAI dataset. The MR sequence is a sagittal 3D DESS.

joints in the hands and feet, hip joint or many more. No specific effort was made to tailor any techniques for the joint in question.

Magnetic resonance (MR) imaging is argued by many to be one of the most useful imaging modalities for structural evidence of knee osteoarthritis (Conaghan et al. (2006)). Its key advantages lie in its ability to visualise soft tissue structures in 3D. In addition to imaging the bone as with plain radiographs, MR is able to capture and visualise ligaments, synovium, menisci and subchondral bone. Magnetic resonance imaging (MRI) is also often acquired at high spatial resolution and is advantageous when compared to x-ray because there is no ionising radiation. An example of a knee MRI used within the thesis is presented in figure 1.1. The 3D MR image is visualised using three cross-sectional views; the first image shows an axial slice, the second image a coronal slice, and the final image is a sagittal slice through the MR image. For a sagittal view, as in figure 1.1, assuming the participant is standing, the first image shows the knee as if viewed from above, the second is from the front and the third is from the side.

The Osteoarthritis Initiative dataset (OAI) is used exclusively throughout this thesis. In this study thousands of participants with progressive osteoarthritis have been recruited and observed over four years to create a large scale longitudinal dataset. The OAI provides a longitudinal public dataset with 4796 participants consisting of medical images, clinical data and biospecimens. It is an ethnically diverse dataset of men and women with participants ranging from 49 to 79 in age, collated at four sites. Despite this wealth of data, large population imaging

studies are yet to become commonplace for OA investigation. This thesis presents scalable methodologies to automatically analyse very large population studies of OA.

The methods presented are scalable and use machine learning to characterise large populations of the OAI dataset, with one experiment using over 10,000 images. Studies of this size enable subtle characteristics of the dataset to be learnt and model many variations within a population. Many of the methods presented here require computationally expensive training phases due to the large scale of the experiments. This has been time consuming for some experiments and the thesis construction but yielded novel and demonstrably scalable methods.

1.2 Biomarkers Discovery

It is advantageous to be able to quantify or rank subjects based upon a diagnosis (the severity of pathology of the joint) and a prognosis (likely rate of progression of the pathology). This enables clinicians to select appropriate treatment options or for clinical trials to determine the efficacy of drugs in a large population of subjects.

A biological marker or a biomarker is used to measure the status of biological systems. A definition for biomarkers has been curated by an expert working group organised by the National Institutes of Health which has been charged to propose terms, definitions, and a conceptual model

> A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention (Biomarkers Definitions Working Group (2001)).

Biomarker applications include use as a diagnostic tool for the identification of those patients with disease, to classify the extent of disease, as an indicator of disease prognosis and use for prediction and monitoring of clinical response to an intervention. The definition and classification of biomarkers has been further developed by Bauer et al. (2006) to be more specifically applicable as a guide to developing and validating osteoarthritis biomarkers.

Biomarker Name	Modality	SQ/Q †	Description	Citation
K&L grade	X-Ray	SQ	Radiologists grades to describe joint space thickness and osteophytes	Kellgren and Lawrence (1957)
BLOKS	MRI	SQ	Aggregation of radiologist scores for various joint structures	Hunter et al. (2008)
WORMS	MRI	SQ	Aggregation of radiologist scores for various joint structures	Peterfy et al. (2004)
articular cartilage volume	MRI	Q	Articular cartilage morphology computed from segmentation	Wluka et al. (2004)
articular cartilage thickness	MRI	Q	Articular cartilage morphology computed from segmentation	Williams et al. (2003)
articular cartilage curvature	MRI	Q	Articular cartilage morphology computed from segmentation	Hohe et al. (2002)
articular cartilage surface area	MRI	Q	Articular cartilage morphology computed from segmentation	Hohe et al. (2002)

Table 1.1: Details of existing imaging biomarkers, all are based upon the community's current understanding of osteoarthritis progression. †where SQ=semi-quantitative and Q=quantitative

The work presented here will focus on structural biomarkers derived from image analysis. Diagnostic imaging biomarkers are defined as a quantifiable score which classifies individuals as either diseased or non-diseased from the analysis of medical images. If an accurate early diagnostic biomarker was developed, it could be suitable to aid clinicians to select appropriate early interventions to treat symptoms of osteoarthritis quickly and cost effectively.

Imaging biomarkers that have been presented previously are detailed in table 1.1. The reader should be aware, that in addition to these imaging biomarkers, biochemical biomarkers have been discovered for quantification of diagnostic and prognostic status (Tanishi et al. (2009); Garnero (2006); Clark et al. (1999)) but are not mentioned further since they are not in the scope of this thesis. Semi-quantitative measures are derived from a reader visually assessing the appearance of the image. Quantitative biomarkers for osteoarthritis are derived from manual or automated segmentations, for which accuracy and repeatability is problematic. All of these metrics are based upon the community's current understanding of osteoarthritis progression, which is currently incomplete.

We present an alternative approach using machine learning algorithms which learn early diagnostic markers from a large population of medical images. This is in contrast to previous

methods which define biomarkers based on human observation of disease progression. Moreover, the methods we present are automated, three dimensional imaging data takes a vast amount of time to thoroughly analyse by humans. Due to human errors, the data extracted is liable to inconsistencies and problems with reproducibility, since inter-reader variability has been shown to be high (Schneider et al. (2012)). An automated solution can add both speed and consistency to the analysis, saves time and ensures reproducibility. .

1.2.1 Machine Learning

Machine learning describes a set of tools which enables data driven analysis by representing, identifying or interpreting patterns. The algorithms generalise the input data such that a solution can be determined without being explicitly designed. Osteoarthritis is a complex condition which is challenging to define by human observation, machine learning provides tools which can be trained to recognise patterns and provide reproducible results over large datasets. We use classification tools, regression tools and manifold learning for representing data.

Manifold Learning

Medical images are very high dimensional data. Each image can be thought of as a feature vector embedded in a high dimensional space with as many dimensions as voxels. This is approximately 25 million dimensions for knee MRI from the OAI dataset. Since all of the MR images are of the knee and acquired under a strict protocol, the content of each image is strongly constrained and the feature vectors exhibit strong patterns and trends. Thus, it is reasonable to expect that no or very few feature vectors will occupy much of the high dimensional space. Instead many of the feature vectors lie on or near a lower dimensional manifold embedded in higher dimensions. Manifold learning describes a class of algorithms which learn the underlying low dimensional structure of the manifold given sufficient sampling of the manifold. Within the lower dimensional manifold embedding some important relationships between images are preserved such that the data can be analysed, visualised and explored more effectively.

Manifold learning features extensively in the chapters of the thesis with many of the manifold learning algorithms providing the facility to non-linearly reduce the dimensionality of high dimensional data. Manifold learning algorithms preserve desirable relationships in a low dimensional embedding which renders them useful for understanding dynamics of large population studies without the "curse of dimensionality" (Houle et al. (2010)).

1.3 Registration

It is important to find spatial correspondences between images, such that it is possible to compare local regions in a given image to the corresponding anatomical region in other images, this is commonly achieved by using a registration algorithm. A registration algorithm is defined as "the determination of a one-to-one mapping between the coordinates in one space and those in another such that points in the two spaces that correspond to the same anatomical point are mapped to each other" (Maurer and Fitzpatrick (1993)). An affine registration operates globally and transforms the co-ordinate system of an image to achieve a global alignment of structures. A non-rigid registration operates locally to deform small structures in the images to maximise joint similarity non-linearly.

1.3.1 Intensity based Affine Registration of the Knee

Our experiments in this thesis show that an affine registration algorithm which establishes correspondences based on intensity based similarity metric (Studholme et al. (1998)) is not sufficiently robust for knee MRI. We found that approximately 4-5% of registrations failed. An affine registration which fails to align images will make a subsequent non-rigid alignment of the images fail also. Over a large scale dataset, this problem becomes significant.

The three main challenges experienced for affine registration of the knee are due to :

1. Variation in volume and spatial distribution of various tissues,

2. Pose variation,

3. The globally cylindrical shape of many of the structures.

The following paragraphs elaborate upon these challenges.

The volume and spatial distribution of soft tissues can also vary based on the subject's lifestyle and genetics. Inter-subject scale variation of the entire joint is pronounced. The appearance of subjects who have high muscle quantity around the knee differ significantly from obese subjects who have higher volumes of fatty tissues because the intensity of these tissues in MR varies. An affine registration which uses an intensity based similarity measure might fail in the case where a *good* affine alignment (as defined by a human observer) has a low similarity .

Pose variation of a subject within the scanner also provides challenges as the field of view may vary and the flexion of the joint is variable. The OAI dataset provides a strict protocol for the MRI scans, knees are externally constrained to maintain a fixed flexion of the joint, however inter-subject anatomical variation means that the external constraint does not impose an identical internal flexion angle in two subjects. Despite an external constraint imposing a fixed flexion angle, internal soft tissue anatomy can not be as strictly constrained between scans. Additionally, a subject with severe OA may not be able to maintain the pose required.

A further problem with registration of the knee is the global shape of the anatomical structures. The global shape of many of the structures in knee MRI are approximately cylindrical. Registration of two knee MRI can result in the tibial bone and femoral bone overlapping. This might yield a high similarity measure even though it does not produce anatomically meaningful correspondences.

Comparison of Musculoskeletal and Neuroimaging Affine Registration

Affine registration is widely used for and applied to brain MR images, with great success. Affine registration for brain MR imaging poses different challenges, and is arguably less challenging when compared with affine registration of musculoskeletal images. The global shape and scale of

a human brain varies less across a population than the human knee. Moreover, the distribution of tissue types varies more significantly in musculoskeletal imaging, with specific reference to inter-subject variation in volume of muscle and fat tissue in knee MRI. Brain tissue has high local variability but this is less significant when considering maximising the similarity of images for affine registration.

The pose of an articulated joint is more variable than that of the brain, since the brain is constrained by the skull it can be expected that the global shape or pose will not vary significantly between scans. Registration of brain MRI is usually performed after skull stripping, a process which extracts brain tissue from skull and the background (Hajnal and Hill (2010)). Such a process does not exist for knee registration moreover, extracting joint tissue would be an ambiguous goal because it is not clear which tissues would be considered part of the joint .

Affine registration may have been considered a solved problem in the context of brain registration. Registration of joints has many unsolved challenges which are yet to have been addressed. This thesis presents strategies which overcome some of the obstacles that are exposed when registering joints.

1.4 Thesis Contributions

This work uses structural MR images to learn the characteristics of the early onset of osteoarthritis. In the first part of this work machine learning algorithms are proposed to learn the appearance of healthy and diseased knees with a greater success than previous contributions in the literature. The second part of this works aims to solve a practical problem of robust global registration of knee MRI.

1.4.1 Learning Osteoarthritis Diagnostic Markers

The first part of the thesis presents an automated method to learn biomarkers for osteoarthritis diagnosis using manifold learning. A high dimensional representation of all images is con-

structed based upon pairwise similarities of all images in the sample. Manifold learning is used to reduce the dimensionality of the dataset, classification and regression are used to predict clinical variables and outcomes with high levels of accuracy. The thesis addresses the combination of multiple local regions of interest and multiple sequences .

1.4.2 Affine Registration of Knee MRI

To learn local diagnostic markers as described in section 1.4.1, it is important to find spatial correspondences between images. For the purposes of biomarker discovery, a manual global registration can be initialised to ensure the knee joints are approximately aligned and thus it is likely for the registration to converge to an accurate solution. However, this solution is not practical for large scale studies of knee OA.

The second part of this thesis is aimed at finding a fast, accurate and robust method for affine registration of knee MRI. It was found when learning diagnostic markers that intensity based affine registration of the knee was unreliable and susceptible to registration failures. This part of the thesis improves the accuracy to mitigate a practical problem which was identified. By ensuring the method is fast it is feasible to compute pairwise registrations as opposed to registering all images to a single target, this removes bias of a single image being selected as the registration target.

1.5 Thesis Overview

The thesis begins by introducing the themes and theory underpinning the key contributions. The thesis contributions (section 1.4) broadly divides the thesis into two parts, the first part (chapters 3 & 4) addresses biomarker discovery and the second part (chapters 5 & 6) addresses accurate and fast affine registrations of knee MRI.

The remainder of the thesis can be outlined as follows:

Chapter 1, the current chapter, has presented the motivations for the thesis. The major contributions and themes are introduced and this section provides a reference of the contents of the chapters in the thesis.

Chapter 2 explores the state of the art techniques in the literature and provides a critique of methods available. Further motivations for the thesis are explored in addition to details of specific challenges which are addressed in subsequent chapters. This chapter presents background information on registration, machine learning and manifold learning.

Chapter 3 proposes a method which uses manifold learning to learn a low dimensional representation of a large dataset of knee MR images. We use a manifold learning technique which aims to best preserve local Euclidean distances between all pairs of images. Further details on manifold learning can be found in section 2.2.6. The co-ordinates of the low dimensional embedding are used as explanatory variables in a multiple linear regression, with clinical data regarding OA diagnosis as the dependant variable. The clinical data used in this study are quantitative physical outcome measures regarding articular cartilage morphology, previous studies have suggested these measure are suitable biomarkers. A high correlation coefficient between the manifold embedding and clinical data suggests that the co-ordinates are indicative of disease severity and can be used as a biomarker. The weight bearing region of the medial femur is selected as the region of interest for this study. Multiple MR sequences are explored to find which sequence contains the most useful information for OA diagnosis.

Chapter 4 further explores manifold learning as a method to establish automatically generated biomarkers of osteoarthritis of the knee. Here, multiple regions of interest and two MR sequences are explored. Two MR sequences selected are those that performed best in chapter 3. This chapter presents a way to combine the diagnostic capabilities of the different regions and different MR sequences, with an application of classifying disease status to discover novel diagnostic imaging biomarkers for osteoarthritis. This is a data driven approach which is derived from pairwise similarity between subjects.

Chapter 5 addresses an issue identified in chapters 3 and 4. It was found that the scalability was impaired by poor affine registration between subjects. This chapter seeks to rectify this

by proposing a method to improve affine registration accuracy in a large population of MRI images. A graphical representation of the population of images is constructed such that a registration between any pair of images can be derived from the shortest path across the graph. The accuracy of this is enhanced through a refinement step which is proposed.

Chapter 6 seeks to find a better refinement of the graphical registration algorithm introduced in chapter 5, which aims to boost both computational efficiency and accuracy. The refinement in chapter 5 achieves accurate results but with a computational cost. The refinements proposed a) fuse the transformations from multiple paths across the graph and b) apply a global consistency constraint to identify erroneous affine transformations.

Chapter 7 concludes the thesis with a summary of contributions and proposal for future directions of study.

Chapter 2

Background

2.1 Introduction

The aim of this chapter is to outline the background theory, motivations and some of the state of the art methods behind this thesis. The background is broadly divided into three main sections: machine learning algorithms, registration and a review of state of the art methods for analysis of radiographical imaging of Osteoarthritis. In addition, the datasets used within the thesis are described.

In the first part of the chapter, machine learning methods which are used through out the thesis in chapters 3, 4, 5 and 6 are reviewed; namely regression, classification and dimensionality reduction are discussed. Regression analysis allows for predicting and analysing observed continuous variables from training data. Classification enables observed data to be categorised into a finite set of classes. Finally, dimensionality reductions aims to define a low dimensional representation of data, which is better suited for further analysis or visualisation.

This chapter shall then introduce image registration. We will focus on global transformation models and their properties because chapters 5 and 6 present a robust global registration method designed for knee MRI. Non-rigid registration models are also discussed due to their application in chapters 3 and 4. In addition to transformation models, similarity metrics, optimisation and the application of registration are also described.

Finally, the chapter presents a clinical introduction to Osteoarthritis, along with a motivation and a review of state of the art techniques for biomarker extraction and analysis. An introduction to the osteoarthritis initiative (OAI) dataset, which is used through out the thesis, is provided.

2.2 Machine Learning

Challenging problems in computer vision have been addressed traditionally using hand-crafted solutions which are carefully designed by a human to achieve high performance. Another option to solve these challenges is by designing algorithms which have the capability to "learn" the solution. This can be referred to as machine learning. Machine learning algorithms are often data-driven and have the ability to generalise data to form effective and evidence-based solutions to the problem.

In general machine learning algorithms use feature vectors which describe samples from a training dataset to derive either a prediction or interpretation of the underlying data. A feature vector is a representation of each sample in the training dataset which satisfactorily describes the extraction of the data. The solution which has been learnt from the training data is used as a powerful tool to predict, understand or interpret unseen data. It is therefore important that the solution is generalisable to unseen data and produces accurate results.

Two common applications of machine learning in medical image analysis are segmentation and pathology diagnosis or prognosis. Segmentation is the delineation of organs or structures in a medical image which enables future processing such as shape analysis, biomechanical modelling or morphemetric analysis (such as structure volume). The input for a segmentation application are generally the voxels in the image, such that each voxel can be assigned a label and the feature vector for the input will generally be a mathematical description of the properties of each voxel in the image. Pathology diagnosis or prognosis is the task of predicting the disease status of a subject using information or features derived from the MRI. The input for a diagnostic application is generally the entire medical image and possibly additional clinical

data, such that each subject can be assigned a label. A simple feature vector of a structure of interest could be constructed using the image intensities, such that for an image consisting of M voxels, it is represented using a D-dimensional feature vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_M)$. These applications will be discussed further throughout the chapter.

Machine learning is a broad field of study but we focus only upon areas that are relevant to the thesis. We briefly discuss regression, classification and dimensionality reduction. A more thorough insight into machine learning can be found in Mitchell (1997).

2.2.1 Terminology and Notation

All machine learning algorithms can be separated into three classes: *supervised*, *unsupervised* and *semi-supervised* learning. *Supervised* learning uses labelled training data to predict a class label for previously unseen data. *Unsupervised* learning algorithms discover and explain the underlying structure of unlabelled data. *Semi-supervised* learning combines both labelled and unlabelled data to learn an appropriate function to predict unseen data. Semi-supervised learning algorithms are not explored further in this work.

Supervised Learning Algorithms

The input for *supervised* learning algorithms is a set of n pairs of training data $\{(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)\}$ where each feature vector $\boldsymbol{x}_i$ belongs to some feature or instance space $\mathbb{R}^M$, associated with each $\boldsymbol{x}_i$ is a class label (or property or response) y_i. We denote training data as the matrix $\boldsymbol{X} = [\boldsymbol{x}_{i,j}]_{i=1,\ldots,M;j=1,\ldots,n}$ and the response vector as $\boldsymbol{y} = [y_1, \ldots, y_n]^T$. During a training phase a function or relation f is learnt which predicts response variable y; $f(\boldsymbol{x}) : \boldsymbol{x} \rightarrow y$. After training, classification algorithms predict the class label $\widehat{y}$ for previously unseen data $\widehat{\boldsymbol{x}}$.

The two broad types of supervised learning algorithms that are reviewed here are classification in section 2.2.2 and regression in section 2.2.5. The performance of supervised algorithms can be assessed by using performance metrics described in section 2.2.3, test and train data can be divided for validation as described in section 2.2.4.

Unsupervised Learning Algorithms

The input data for *unsupervised* algorithms are unlabelled and so the input is a set of n feature vectors $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \in \mathbb{R}^M$. As before, feature vectors will be denoted as matrix $\boldsymbol{X} = [\boldsymbol{x}_{i,j}]_{i=1,\ldots,M;j=1,\ldots,n}$.

Unsupervised algorithms summarise and explain key properties of the data by trying to determine the underlying structure of the data, such that further statistical analysis, processing or other machine learning algorithms can be applied to the data. This is achieved by either learning an alternative representation of the features or finding structurally appropriate clusters of data. An alternative representation of features can be learnt using manifold learning and dimensionality reduction algorithms. These techniques are discussed in section 2.2.6. Examples of clustering algorithms include k-means (MacQueen (1967)) and hierarchical clustering (Johnson (1967)). These algorithms assume that there a k clusters and each feature vector is assigned to a cluster $y_i \in \{C_1, C_2, \ldots, C_k\}$ iteratively based on clusters centroids. Clustering is not discussed further in this work but more details about clustering techniques can be found a review by Xu and Wunsch (2010).

2.2.2 Classification

Classification algorithms are types of supervised learning, where a classifier function, f, is learnt to predict, y, as a member from a set of discrete class labels $\{C_1, C_2, \ldots, C_{\mathcal{N}}\}$. A classifier learns a mapping f, that maps data, $\boldsymbol{x}_i$, to a class label; $f : \boldsymbol{x} \rightarrow y$. For unseen data $\widehat{\boldsymbol{x}}$, function f predicts $\widehat{y} \in \{C_1, C_2, \ldots, C_{\mathcal{N}}\}$ where $\widehat{y} = f(\widehat{\boldsymbol{x}})$.

k-Nearest Neighbour (k-nn)

The k-Nearest Neighbour (k-nn) classification algorithm (Cover and Hart (1967)) is a supervised, instance-based learning algorithm. k-nn classifies an unseen data point based on its proximity to previously observed data.

The k-nn algorithm defers all the computation until classification where the feature space is only approximated locally. Prior to classification, features vectors ($\boldsymbol{x}_i$) are stored with their class label (y_i) in a database. Given unseen data point $\widehat{\boldsymbol{x}}$, the k-nearest neighbour classifier ($f(\widehat{\boldsymbol{x}})$) searches for the k data points in the database with the smallest distance $d(\boldsymbol{x}_i, \widehat{\boldsymbol{x}}), \forall \boldsymbol{x}_i \in \boldsymbol{X}$. This results in a probability distribution of class labels from the k samples. The class label with the highest probability is selected for $\widehat{y}$.

The distance function $d(\boldsymbol{x}_i, \widehat{\boldsymbol{x}})$ can be defined in a number of ways but the most common choices include

1. **Manhattan distance (L_1 norm)** : $d_{manh}(\boldsymbol{x}_i, \widehat{\boldsymbol{x}}) = |\boldsymbol{x}_i - \widehat{\boldsymbol{x}}|_1 = \sum_{j=1}^{d} |x_{i,j} - \widehat{x}_j|$. The Hamming distance for non-numeric strings is equivalent and often used as fast distance metric for text analysis.

2. **Euclidean distance (L_2 norm)** : $d_{eucl}(\boldsymbol{x}_i, \widehat{\boldsymbol{x}}) = |\boldsymbol{x}_i - \widehat{\boldsymbol{x}}|_2 = \sqrt{\sum_{j=1}^{d} (x_{i,j} - \widehat{x}_j)^2}$ is a standard distance function in Euclidean space.

3. **Mahalanobis distance** : $d_{mahal}(\boldsymbol{x}_i, \widehat{\boldsymbol{x}}) = \sqrt{\sum_{j=1}^{d} (x_{i,j} - \widehat{x}_j)^T \boldsymbol{S}^{-1} (x_{i,j} - \widehat{x}_j)}$ where $\boldsymbol{S}$ is the covariance matrix of the data $\boldsymbol{X}$. The euclidean distance is adjusted such that the distribution of the $\boldsymbol{X}$ is considered and feature vectors are weighted based on their variance.

4. **Supervised distance metric learning** : The classification accuracy of k-nn can be improved if the distance metric ($d_{learn}(\boldsymbol{x}_i, \widehat{\boldsymbol{x}})$) is learnt specifically for the underlying data (Yang (2006)). Here feature vectors are weighted based upon their significance, such as in neighbourhood component analysis (NCA) (Goldberger et al. (2005)).

5. **Image Simiarity Metrics** : If feature vector $\boldsymbol{x}_i$ is an image, the distance measure can be defined based upon the image similarities, such as those which are described in section 2.3.3.

The k-nn algorithm is attractive due to its simplicity, however it has drawbacks. Some of these have been discussed and contributions in the literature have improved upon the basic algorithm.

Firstly, exhaustively searching the database for the nearest neighbours for each classification can be costly. If many classifications are required, much of this search will be duplicated on subsequent classification. It can therefore be quicker to implement algorithms with sub-linear complexity. There are many contributions in the literature to improve upon the efficiency of the search including *kd*-trees (Bentley (1975)) and locality sensitive hashing (Indyk and Motwani (1998)).

A further drawback to the basic k-nn classification algorithm which works via majority vote, is that class labels with more examples ($\boldsymbol{x}_i$) in the database are likely to dominate the classification of the unseen data ($\widehat{\boldsymbol{x}}$), since they tend to appear at a higher rate within the k nearest neighbours. The effect of this can be reduced by weighting the contribution of each of the k nearest neighbours based upon the distance between $\widehat{\boldsymbol{x}}$ and $\boldsymbol{x}_i$, where $\boldsymbol{x}_i \in \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\}$ (indexing is based upon nearest neighbours to $\widehat{\boldsymbol{x}}$).

Linear Discriminant Analysis (LDA)

Linear discriminant analysis (LDA), or Fisher linear discriminant (Fisher (1936)), is a simple classifier which determines a linear combination of features which separates two or more classes. LDA seeks to reduce dimensionality while preserving as much of the class discriminatory information as possible.

LDA achieves linear classification by projecting the data on to a single dimension which minimises the within class scatter matrix ($\boldsymbol{S_W}$) and maximises the between class scatter matrix ($\boldsymbol{S_B}$), see equation 2.1. Each training data sample $\boldsymbol{x}_i$ has an associated label from the set of class labels $\{C_1, \ldots, C_{\mathcal{N}}\}$. A class C_α has n_α training samples. The mean of all samples from class C_α has a mean μ_α. $\bar{\boldsymbol{x}}$ is the mean of all training samples.

$$\boldsymbol{S_W} = \sum_{\alpha=1}^{\mathcal{N}} \sum_{i=1}^{n_\alpha} (\boldsymbol{x}_i - \mu_\alpha)(\boldsymbol{x}_i - \mu_\alpha)^T, \boldsymbol{S_B} = \sum_{\alpha=1}^{\mathcal{N}} (\mu_\alpha - \bar{\boldsymbol{x}})(\mu_\alpha - \bar{\boldsymbol{x}})^T \tag{2.1}$$

The between class scatter matrix is maximised whilst the within class scatter matrix is minimised simultaneously by maximising the ratio of the two. This leads to the objective function

in equation 2.2 in which W denotes the projection matrix:

$$\arg\max_{x} \frac{\boldsymbol{W}^T \boldsymbol{S_B} \boldsymbol{W}}{\boldsymbol{W}^T \boldsymbol{S_W} \boldsymbol{W}} \tag{2.2}$$

This reduces to the generalised eigen-problem in equation 2.3. For two class labels there will be at most one non zero eigenvalue. The corresponding eigenvector forms the discriminant vector $\boldsymbol{w}$ and the non-zero eigenvalue defines the value of separation between the class labels:

$$(\boldsymbol{S_W}^{-1} \boldsymbol{S_B})\boldsymbol{v} = \boldsymbol{\lambda v} \tag{2.3}$$

The classifier $f(\hat{\boldsymbol{x}})$ is defined as in equation 2.4, where b defines the location of the discriminatory hyperplane.

$$f(\hat{\boldsymbol{x}}) = \boldsymbol{W}^T \hat{\boldsymbol{x}} + b \tag{2.4}$$

Adaboost

Adaboost is a greedy algorithm which finds a linear combinations of *weak* classifiers to construct a *strong* classifier (Freund and Schapire (1997)). A *weak* classifier is a classifier which has a low classification accuracy, typically just over 50% for a binary classification problem. A *strong* classifier is a classifier which achieves high classification accuracy.

In the training phase, Adaboost greedily selects *weak* classifiers over a series of T rounds. Upon each round the weak hypothesis ($h(\boldsymbol{x})$) that best separates data points based upon their class labels is added to the final classifier $f(\boldsymbol{x})$. A weight α is used to explain its contribution of the weak hypothesis to $f(\boldsymbol{x})$.

$$f(\boldsymbol{x}) = \sum_{t=1}^{T} \alpha_t h_t(\boldsymbol{x}) \tag{2.5}$$

Data points which are unsuccessfully classified in previous rounds are considered to be more challenging exemplars to classify. Thus it is considered to be more important to find a good weak

classifier to discriminate these data points in subsequent rounds. Upon each training round, Adaboost weights each training example based on how challenging it has been to classify. Initially, all weights are set equally, but in each round, the weights of incorrectly classified examples are increased so that the weak learner is forced to focus on the challenging examples in the training set.

Despite each *weak* classifier only achieving low classification accuracy, its failings are compensated by the linear combination of the *weak* classifiers in the ensemble. The computational time of classifying unseen data is low since it is a linear combination of very simple classifiers. It has been successfully used as a face detector (Viola and Jones (2001)). Adaboost has been used for medical image analysis for detecting the knee joint in MR images (Yin et al. (2010)) and has been compared to other classifiers, eg. support vector machines (SVM) for the detection of Alzheimer's disease (Morra et al. (2010)).

Adaboost is referred to as an ensemble learning method. Ensemble learning combines multiple classifiers and weights their contribution towards an overall classification. Other ensemble learning approaches include bagging (Breiman (1996)) which was later extended to random forests (Breiman (2001)).

Support Vector Machines (SVM)

Support vector machines (SVM) is a supervised algorithm which divide a set of points into classes with a hyperplane that maximises the margin or the distance between the closest points to the class boundaries. It is also called a wide margin classifier. This section gives an overview of SVM. For a thorough tutorial please refer to Smola and Schölkopf (2004).

Each data point is represented by a vector in a vector space. The distance between those vectors that are closest to the hyperplane is maximized for the training data. This wide margin ensures that the test data points, which do not exactly correspond to the data used for training, are classified as reliably as possible. The hyperplane depends only on the closest vectors. The vectors which are further away from the hyperplane do not affect the location and position

of the hyperplane. The closest points to the boundary which define the margin are known as support vectors, hence giving the name support vector machine.

The concepts were first introduced with the Generalized Portrait algorithm (Vapnik and Chervonenkis (1964)), which is a linear classifier. More recent contributions provide a nonlinear generalization for data which can not be classified with a linear hyperplane. For non-linearly separable data, support vector machines use the 'kernel trick' (Boser et al. (1992)). The idea behind the 'kernel trick' is to transform the vector space and thus the training vectors into a higher-dimensional space. With a sufficiently high number of dimensions (or even an infinite dimensional space) the classification task is converted to a linearly separable problem, so that it is possible to compute a linear hyperplane to separate the classes. The kernel function enables the linear hyperplane to be computed at high-dimensions without it being necessary to explicitly perform the mapping into a high dimensional space. As part of the 'kernel trick', the solution is transformed back to the lower dimensional space, which converts the linear hyperplane to a non-linear hypersurface.

As a preprocessing step before the support vector machine classifier is applied, it is important to scale the features in both the training and test data, such that those with high variance do not dominate those with lower variances.

2.2.3 Classification Performance Metrics

It is essential to be able to assess the performance of a classifier so that is possible to decide whether a classifier is appropriate for the data.

Accuracy, Sensitivity and Specificity

Classification results can be represented by a confusion matrix (in table 2.1). It is also called contingency table or error matrix in other fields of study. It consists of two rows and two columns that reports the number of false positives (incorrectly identified subjects), false negatives (incorrectly rejected subjects), true positives (correctly identified subjects) and true

		Predicted Classes	
		Positive (patient)	Negative (control)
Actual Classes	Positive (patient)	True Positive (TP)	False Negative (FN)
	Negative (control)	False Positive (FP)	True Negative (TN)

Table 2.1: Confusion Matrix

negatives (correctly rejected subjects).

It is useful to understand the accuracy of the classifier, i.e. the proportion of correct classification, to give some indication of the overall performance. Accuracy is computed as

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN} \tag{2.6}$$

However, the percentage of correct classifications is not a reliable metric for the real performance of the classifier. It can be misleading if the dataset is unbalanced (the number of samples in each classes varies significantly). For example if the classifier predicted the disease status of every sample as healthy, this would be a very poor classifier. If there were only a small number of diseased subjects in the training dataset, an accuracy metric would appear high because the class distributions of the training samples were not considered. Additionally, accuracy on its own does not provide information about what the classifier is likely to misclassify. For an example of pathology screening, where a large group is screened for a disease, it may be important to ensure that a subject with the disease is not misclassified (false negative), since this could have a severe health impact. However, it might be less problematic if a subject who is healthy is misclassified (false positive).

Sensitivity and specificity are statistical measures of the performance of a binary classification test which enable analysis in an dataset with unbalanced labels. Sensitivity (also called the true positive rate; $\frac{TP}{TP + FN}$) measures the proportion of actual positives which are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the

condition). Specificity (also called the true negative rate; $\frac{TN}{TN+FP}$) measures the proportion of negatives which are correctly identified as such (e.g. the percentage of healthy people who are correctly identified as not having the condition, sometimes called the true negative rate).

Area Under Receiver Operating Characteristic Curve (AUC ROC)

The receiver operating characteristic or ROC (Fawcett (2006)) is a graphic to represent the performance of a binary classifier. The x-axis corresponds to $1 - specificity$ (or the false positive rate (FPR)), the y-axis shows the *sensitivity* (or the true positive rate (TPR)). It is a comparison of the two operating characteristics (FPR vs TPR) as the classifier threshold is altered. It represents the trade off between the FPR and TPR. It shows how well the classifier can model the underlying data and demonstrates its performance over a range of classifier thresholds. As described previously, some applications benefit from maximising the specificity (whilst of course maintaining a reasonably high sensitivity) as opposed to purely maximising the accuracy. Figure 2.1 shows an example of a ROC curve for a linear classification of a one dimensional feature, where samples from each classes are normally distributed with differing means and identical standard deviations .

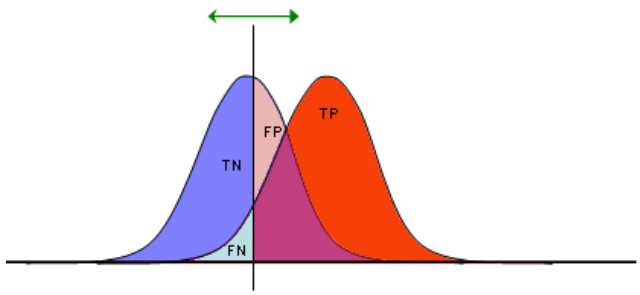
The area under the curve (AUC) corresponds to the probability that a classifier will rate a randomly chosen positive instance higher than a randomly chosen negative one (assuming 'positive' rates higher than 'negative'). AUC of the receiver operator curve is often reported to describe the overall performance of a classifier as a measure to explain the trade of between the FPR and TPR.

2.2.4 Cross Validation

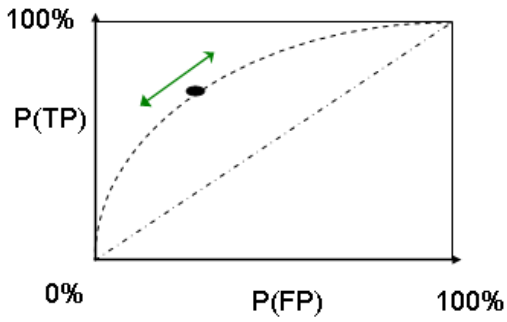
Cross validation is an approach used to compute a reliable value for the quality of a machine learning solution, testing its ability to generalise unseen data. It is important to ensure that a machine learning solution is tested on different data to which it was trained. In practise studies frequently suffer from a small sample size which will not faithfully reflect the entire population.

Figure 2.1: A linear classifier of a single feature is represented in sub-figure 2.1 (a). The left curve constitutes the sampling distribution of subjects from the negative class, the curve on the right is the sampling distribution of subjects from the positive class. A separation threshold (green arrow) classifies a proportion of both sub-populations incorrectly; either false positive (FP, pink area) or false negative (FN, light blue area), these proportion vary based upon the location of the threshold. Sub-figure 2.1 (b) shows the corresponding ROC curve. The point and green arrow represents the separation threshold in sub-figure 2.1 (a) †.

(a) Sample distribution



(b) ROC curve



† *Illustrations from http://upload.wikimedia.org/wikipedia/commons [accessed 29/03/13]*

Given a small sample of data, it is possible that, by chance, the distribution of the test and training data could bias the test and be a poor indicator of performance. Cross validation is employed to guard against such eventualities by repeatedly partitioning the data into different test and training sets. For each partition the algorithm is trained and then testing on the specified datasets.

k-fold Cross Validation

In k-fold cross-validation, the original dataset is randomly partitioned into k equal size subsets. The cross-validation process is repeated k times. Upon each iteration of the k-fold cross validation, a single subset is selected as the validation data for testing the model and the remaining $k-1$ subsets are used as training data. Each of the k subsets are used exactly once as the validation data. The k results from the folds are then averaged to produce a single estimation of performance. All observations are used for both training and validation, and each observation is used for validation exactly once.

Stratified k-fold Cross Validation

Stratified k-fold cross validation is similar to k-fold cross validation except that the partition selection is not random but determined such that each fold contains roughly the equal proportions of the types of class labels.

Leave one out cross validation

Upon each iteration of a leave-one-out cross-validation framework, a single sample from the original dataset is selected as the validation data and the remaining observations are used as the training data. This is equivalent to k-fold cross validation were k is the number of samples in the dataset (n) . As such there are n iterations and each observation in the dataset is used once as the validation data.

Summary

Cross-validation is a tool which predicts the fit of a solution to a hypothetical validation set when an explicit validation set is not available by "taking turns" with the test or training data. The goal of cross-validation is to estimate the performance of a model to a dataset that is independent of the training data.

2.2.5 Regression

Regression describes algorithms which model the relationship between explanatory variables (or observed data or in our case feature vector), $\boldsymbol{X}$ and the scalar response $\boldsymbol{y}$. Linear regression models the relationship of a single explanatory variable or a single feature and a single response variable. Multiple linear regression describes the case where the relationship between explanatory variables or multiple features and a single response variable are modelled. In some case a linear model is not appropriate for the data. In this case, non-linear regression models can be used but these are not covered here.

Multiple linear regression models the relationship between $\boldsymbol{y}$ and $\boldsymbol{X} = [x_{i,j}]_{n \times M}$ by fitting a linear model to the observed data, see equation 2.7. The coefficients $\boldsymbol{\beta}$ explain the contribution of each explanatory variable where the associated error in the model is explained by $\boldsymbol{\epsilon}$.

$$y_i = \beta_0 + \beta_1 x_{i,1} + ... + \beta_n x_{i,M} + \epsilon_i \tag{2.7}$$

The coefficients are described as *partial* regression coefficients because they allow for the effect of other variables and in combination can explain the response data with some residual error. The unknown coefficients β_0, β_1,..., β_n are computed using a least squares approach, so that the squared deviations of the observed data to the predicted data is minimised.

The applications of regression are two-fold:

1. To predict the response for unseen data. If an unseen feature vector $\widehat{\boldsymbol{x}}$ is given and the response value is unknown, the model can be used to predict the response $\widehat{y}$. In order to be used for prediction the model must be validated with cross-validation. This is described in section 2.2.4.

2. To learn the relationship between the explanatory and response variables, by discovering the strength of the relationship by recording the residuals R^2 and the p-value, associated with the F-test. The F-test is used to test the significance of the model.

Statistics for Analysing Regression

The R^2 statistic or the coefficient of determination indicates how well the data points fit the model. R^2 is defined as the proportion of variability in a dataset that is accounted for by the statistical model. It can be computed as from the sum of squared error $SSE = \sum_i (y_i - \hat{y}_i)^2$ and the sum of square total $SST = \sum_i (y_i - \bar{\boldsymbol{y}})^2$. Where $\bar{\boldsymbol{y}}$ is the mean and $\hat{y}_i$ is the value of y_i predicted using the model from $\boldsymbol{x}_i$.

$$R^2 = 1 - \frac{SSE}{SST} \tag{2.8}$$

The R^2 statistic is often accompanied by a p-value to indicate statistical significance. For multiple linear regression there are two associated p-values one is associated with an F-test and the other with a t-test. It is possible to compute a t-test p-value for each coefficient in the model. A low p-value indicates the variable which is multiplied by that coefficient is important to the model. A single F-test p-value can be computed for the entire regression model. A $p < 0.05$ indicates a significant linear regression relationship between the response variable and the predictor variables.

2.2.6 Manifold Learning

Dimensionality reduction seeks to find a lower dimensional representation of a dataset by finding a small number of features to represent a large number of observed dimensions. Given a dataset $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \in \mathbb{R}^M$, dimensionality reduction aims to learn a set of points $\{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n\} \in \mathbb{R}^m$ such that the dimensionality is reduced $m \ll M$ and $\boldsymbol{x}_i$ "corresponds" to $\boldsymbol{y}_i$.

Data can generally be explained in fewer dimensions than the number of features which are used to observe the data. Natural data is often highly constrained where patterns and structures can be expected in the data. For example an MR image of the knee will always contain boney structures of comparable intensity. Intuitively, it can be expected that samples within the feature space will form strong trends or clusters (along a low dimensional manifold), while other regions of the high dimensional feature space might not contain any sample points. As a results data samples lie on or near a low dimensional manifold in ambient space, $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\} \in \mathcal{M}$. Given a sufficiently-many data samples (n), manifold learning approximates the m-dimensional manifold $\mathcal{M}$ embedded in $\mathbb{R}^M$, where $m \ll M$. The data is unlikely to lie directly upon the manifold in high dimensional space as we expect to incur some observational noise. Experiments on hypothetical data have demonstrated their ability to learn manifold embeddings even in the presence of observational noise (Belkin and Niyogi (2003)). A detailed overview can be found at Van der Maaten et al. (2009); Cayton (2005).

The manifold learning algorithms described here aim to preserve different properties of the underlying high dimensional data. *Principle component analysis* and *Multidimensional Scaling*

are linear algorithms that compute a global projection that preserves both global and local distances using a linear transformation of the data to a new basis. All other methods learn a non-linear relationship between the high and low dimensional representations. *Local linear embedding* and *Laplacian eigenmap embedding* are both classed as algorithms which preserve local neighbourhoods. *Isomap embedding* learns a manifold representation which preserves global distances.

Principle Component Analysis (PCA)

Principle component analysis (PCA) is a linear dimensionality reduction algorithm. It was first introduced by Pearson (1901), and developed independently by Hotelling (1933). A thorough overview has been provided by Joliffe (1986).

PCA reduces the dimensionality of the data by embedding it into a lower dimensional linear subspace. The low-dimensional representation is constructed by finding a linear basis of reduced dimensionality which explains as much of the variance in the data as possible.

More precisely, PCA finds a linear basis $\boldsymbol{A}$ which explains as much of variance of $\boldsymbol{X}$ as possible. This is achieved by maximising the cost function $trace(\boldsymbol{A}^T\boldsymbol{C}\boldsymbol{A})$, where $\boldsymbol{C} = [cov(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{M\times M}$ is defined as the zero-mean adjusted sample covariance matrix of data $\boldsymbol{X} = [x_{i,j}]_{n\times M}$. The ith column of the data $\boldsymbol{x}_i$ is a feature vector of length n.

$$\boldsymbol{C}(i,j) = cov(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{\sum_{k=1}^{n}(x_{i,k} - \bar{\boldsymbol{x}}_i)(x_{j,k} - \bar{\boldsymbol{x}}_j)}{n-1} \tag{2.9}$$

The cost function is maximised by solving the eigenproblem (2.10) for the m principle (largest) eigenvalues λ. The corresponding m principle eigenvectors $\boldsymbol{v}$ form the linear mapping $\boldsymbol{A}$.

$$\boldsymbol{C}\boldsymbol{V} = \boldsymbol{\lambda}\boldsymbol{V} \tag{2.10}$$

The low dimensional representation of data points are computed as a mapping $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{A}$.

Multidimensional Scaling (MDS)

Multidimensional scaling (MDS) (Cox and Cox (2000)) computes the low dimensional embedding that best preserves pairwise Euclidean distances between data points. MDS uses the distance or weight matrix $\boldsymbol{W}$ with its elements $w_{i,j}$ which represent the proximity between two high-dimensional data points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. MDS seeks to find a linear transformation of the data on to a lower dimensional space that best preserves the pairwise distances in the high dimensional space. This is computed by minimising the objective function

$$\sum_{i,j=0}^{n} (w_{i,j}^2 - || \boldsymbol{y}_i - \boldsymbol{y}_j ||^2) \tag{2.11}$$

Solving this optimisation problem reveals the position of data points $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ in the low dimensional embedding such that when $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are close/far in the high dimensional space, they will also be so in the low dimensional space. This constraint is imposed by squaring weights from the weight matrix $\boldsymbol{W}$.

2.2.7 Non-linear Manifold Learning Algorithms

PCA and MDS are called classical scaling algorithms. They both aim to retain pairwise Euclidean distances through a linear transformation of the high dimensional feature space. Complex natural data often lies on or near a curved manifold. In this case classical scaling is no longer appropriate. Instead it is important for the manifold learning algorithm to consider the distribution of neighbouring data points.

Neighbourhood Definitions

For non-linear manifold learning it is important to represent the neighbourhood of data points, we present some background on graphs which are used to define data point relationships.

Graphs A graph $G = (V, E)$ can be constructed to explain local neighbourhoods or pairwise relationships between data points. V defines the set of vertices which represents the n data points (often corresponding to $\boldsymbol{X}$). E defines the set of edges which explain the relationship between the data points (often indicating similarity or distance). Associated with each edge $e_{i,j} \in E$ is a weight $w_{i,j}$. The weight matrix $\boldsymbol{W} = [w_{i,j}]_{n \times n}$ is defined as the distances between pairs of data points at vertex v_i and v_j via the function $w_{i,j} = d(i, j)$.

An **undirected** graph is a graph where edges between two vertices are bidirectional, explicitly $\forall e_{i,j} \in E \iff \exists e_{j,i} \in E$ and the weight matrix $\boldsymbol{W}$ is symmetrical. In a **directed** graph the edges connecting vertices are unidirectional, explicitly $\exists e_{i,j} \in E \not\Longrightarrow \exists e_{j,i} \in E$ and the weight matrix $\boldsymbol{W}$ is not a symmetric matrix.

A **Fully Connected Graph** is a graph in which all vertices are connected via an edge and $\boldsymbol{W}$ is dense.

A **Sparse Neighbourhood Graph** has a sparse set of edges and corresponding weights, vertices are connected to their neighbours as defined by neighbourhood function ($N(v_i)$). The neighbourhood function $N(v_i)$ can be defined as:

- **ϵ- or radius-neighbourhood** creates an undirected graph in which vertices are only connected if the distance between data points is less than ϵ; $d(i, j) < \epsilon \implies \exists e_{i,j} \in E$.
- **k-nearest neighbourhood** generates a directed graph which selects the k most similar neighbours of v_i in V. Where the graph is required to be undirected or symmetric ($e_{i,j} = e_{j,i}$) the definitions of neighbourhood have to be adjusted and referred to as:
  - *the k-nearest neighbour graph*, by making every edge that exists bidirectional. If $\exists e_{i,j} \in E$ then create $e_{j,i}$ and $w_{j,i} = w_{i,j}$.
  - *mutual k-nearest neighbour graph*, by removing edges which are not bidirectional. If $\exists e_{i,j} \in E$ and $\nexists\, e_{j,i} \in E$ then $e_{i,j}$ is removed from the set of edges E.

Fully connected graphs define pairwise relations between all points, where as a sparse neighbourhood graph encodes information regarding local neighbourhoods.

Isomap Embedding

Isomap (Tenenbaum et al. (2000)) is a non-linear manifold embedding algorithm which uses the MDS algorithm as a key component. Isomap preserves pairwise geodesic distances between data points. A geodesic distance is defined as the distance between two points measured along the manifold. To achieve this, Isomap approximates pairwise geodesic distance ($d_g(\boldsymbol{x}_i, \boldsymbol{x}_j)$) between all data points, represented as matrix, $GeoDist(\boldsymbol{X}) = [d_g(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{n\times n}$. The geodesic distances are approximated for data $\boldsymbol{X}$ by constructing a sparse symmetric k-nearest neighbour (or ϵ-neighbourhood) graph G. The edge weights $w_{i,j}$ are defined by an appropriate similarity metric for the data and $d_g(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is the sum of the weights of the edges of the shortest path between v_i and v_j. The shortest path can be computed using Dijkstra's algorithm (Dijkstra (1959)) for finding single shortest paths or Floyd's algorithm (Floyd (1962)) for finding all pairwise shortest paths. The final embedding coordinates $\boldsymbol{y}_i$ are obtained by applying classical MDS to the geodesic distance matrix $GeoDist(\boldsymbol{X})$. MDS preserves the geodesic distances computed by the shortest path algorithms.

Isomap is computed using a dense matrix, the following will discuss manifold learning solutions which are sparse algorithms.

Locally Linear Embedding (LLE)

Locally Linear Embedding (LLE) introduced by Roweis and Saul (2000) learns a low dimensional embedding which preserves the local neighbourhoods of high-dimensional data. The data is assumed to be locally linear but globally non-linear, which means that the algorithm can characterise the localised linear geometry of a set of patches. All data points $\boldsymbol{x}_i$ in $\boldsymbol{X}$ are represented as a weighted combination of their neighbours (neighbourhoods are defined in section 2.2.7) in high dimensional space. A reconstruction error can be computed by the cost function in equation 2.12 which sums the squared distances between all data points ($\boldsymbol{x}_i$) and their reconstructions ($\sum_{j=1}^{n} w_{i,j}\boldsymbol{x}_j$). The weights $w_{i,j}$ summarise the contribution of the jth data point to the ith reconstruction.

Since data point $\boldsymbol{x}_i$ is reconstructed only from its neighbours, this implies that $w_{i,j} = 0$ if $\boldsymbol{x}_j$ is not contained in the neighbourhood set of $\boldsymbol{x}_i$. The matrix $\boldsymbol{W}$ is learnt by an optimisation of the reconstruction error. This can be posed as a least squares minimisation problem.

$$\sum_{i=1}^{n} || \boldsymbol{x}_i - \sum_{j=1}^{n} w_{i,j} \boldsymbol{x}_j ||^2 \tag{2.12}$$

The new embedding co-ordinates Y are then learnt from the weight matrix $\boldsymbol{W}$ which is computed using equation 2.12 such that the reconstruction error is minimised:

$$\sum_{i=1}^{n} || \boldsymbol{y}_i - \sum_{j=1}^{n} w_{i,j} \boldsymbol{y}_j ||^2 \tag{2.13}$$

Laplacian Eigenmap Embedding (LEE)

Laplacian eigenmap embedding (Belkin and Niyogi (2003)) learns a low dimensional manifold representation which aims to preserve the local similarities. Locality is preserved by ensuring that the local neighbourhood of each data point in the high dimensional space is reflected in the low dimensional space. Laplacian eigenmap embeddings are closely related to spectral clustering for which von Luxburg (2007) presents a detailed tutorial. Local neighbourhoods for each data point are expressed using a weight matrix $\boldsymbol{W}$ and a graph representation. Laplacian eigenmaps uses a sparsely connected graph G which can be defined using a symmetrical k-nearest neighbour graph or an ϵ-nearest neighbour graph. The weight matrix $\boldsymbol{W}$ which defines the connectivity of each edge is defined as a dissimilarity function $(d(\boldsymbol{x}_i, \boldsymbol{x}_j))$.

The manifold is learnt by minimising the objective function $\sum_{i,j=1}^{n} w_{i,j}(\boldsymbol{y}_i - \boldsymbol{y}_j)^2$. The objective function ensures that when data points $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are close in high dimensional space, that the corresponding data points $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ are also close in the low dimensional manifold representation. The objective function can be minimised via a closed form solution. Assuming the graph Laplacian is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$ and $\boldsymbol{D}$ is the diagonal degree matrix, $\boldsymbol{D}_{i,i} = \sum_j w_{i,j}$ and $\boldsymbol{D}_{i,j} = 0$ where $i \neq j$, we can write:

$$
\begin{aligned}
\sum_{i,j=1}^{n} w_{i,j}(y_i - y_j)^2 &= \sum_{i=1}^{n} d_i y_i^2 - 2\sum_{i,j=1}^{n} y_i y_j w_{i,j} + \sum_{j=1}^{n} d_j y_j^2 \\
&= 2(\sum_{i=1}^{n} d_i y_i^2 - \sum_{i,j=1}^{n} y_i y_j w_{i,j}) \\
&= 2(\boldsymbol{Y}^T \boldsymbol{D} \boldsymbol{Y} - \boldsymbol{Y}^T \boldsymbol{W} \boldsymbol{Y}) \\
&= 2\boldsymbol{Y}^T(\boldsymbol{D} - \boldsymbol{W})\boldsymbol{Y} \\
&= 2\boldsymbol{Y}^T \boldsymbol{L} \boldsymbol{Y}
\end{aligned}
\tag{2.14}
$$

Equation 2.14 is minimised subject to the constraint $\boldsymbol{y}^T \boldsymbol{D} \boldsymbol{y} = 1$. This constraint removes arbitrary scaling factors in the embedding, this prevents trivial solution where all $\boldsymbol{y}_i$ are zero. This is solved using the generalised eigenvalue problem $\boldsymbol{L}\boldsymbol{V} = \boldsymbol{\lambda}\boldsymbol{D}\boldsymbol{V}$. The eigenvectors corresponding to the smallest non-zero eigenvalues form the embedding coordinates $\boldsymbol{Y}$.

Summary

The non-linear manifold learning algorithms reviewed here are powerful methods which provide insight by learning the low dimensional structure of the data. Complex natural world data can rarely undergo an effective linear transform which describes the nuances of the data and maximises the information present. Non-linear manifold learning algorithms go some way to explaining this data.

2.3 Registration

Image registration aims to find an optimal spatial transformation or deformation that maximises correspondences between a pair of images containing comparable objects. Medical image registration is applied to tomographic images, images of the interior of the body, these include CT, MRI, SPECT, PET or x-ray. The pair of images selected for registration can be of differing modalities or acquired from different subjects but must contain approximately the same anatomical region of interest so that correspondences can be found.

Registration for medical image analysis has been extensively reviewed in the literature. Specifically Maintz and Viergever (1998) provide an early review of work on image registration. Crum et al. (2004) present a more recent review of non-rigid registration, while Pluim et al. (2003) present a review of registration with a focus on mutual information metrics as a similarity measure.

2.3.1 Definitions and Basic Terminology

Given a pair of medical images, a registration algorithm determines a suitable transformation which warps (transforms, moves or deforms) the *source* image (moving image) such that a similarity metric between itself and the *target* image (fixed image) is maximised. Many registration algorithms can be subdivided into three key components. For a given application a suitable combination of components can be assembled to tackle the registration challenge. The three key components are;

1. **A Transformation model** which defines a class of transformations or deformations which can be used to transform the images.
2. **A Similarity metric** which evaluates the degree of spatial correspondence between images.
3. **A Optimisation method** which aims to optimise an objective function whilst avoiding local maxima.

Notation

For the purposes of explaining the concept of registration we refer to the source image as $\boldsymbol{I}_S$ and the target image as $\boldsymbol{I}_T$. The transformation $\boldsymbol{T}$ maps a point in $\boldsymbol{I}_T$ at voxel location $\boldsymbol{p} = (i, j, k)$ to another point in $\boldsymbol{I}_S$ at voxel location $\boldsymbol{p'} = (i', j', k')$.

$$\boldsymbol{T} : \boldsymbol{p} \implies \boldsymbol{p}' \tag{2.15}$$

The similarity metric of the registration is computed over all voxels in the image overlap domain $\boldsymbol{\Omega}_{T,S}$, which includes voxels in both the source and the target image that are overlapping, for a given transformation $\boldsymbol{T}$, $\boldsymbol{p} \in \boldsymbol{\Omega}_{T,S} \Leftrightarrow \boldsymbol{p}' \in \boldsymbol{\Omega}_{T,S}$. The intensity at voxel location $\boldsymbol{p}$ in the target image is defined as $\boldsymbol{I}_T(\boldsymbol{p})$ and the intensity at the corresponding voxel location in the source image is defined as $\boldsymbol{I}_S(\boldsymbol{p}')$ or $\boldsymbol{I}_S(\boldsymbol{T}(\boldsymbol{p}))$.

World and Image co-ordinates

The voxels of a medical image may be referred to in multiple different ways. Two standards are widely adopted, *world co-ordinates* and *image co-ordinates*. World co-ordinates (sometimes called scanner co-ordinates) define the origin in a real world position (this might not exist in the field of view of the image), the axis are defined relative to the scanner and the position of any point is indexed in mm. Image co-ordinates define the origin to be at a fixed location in any given image, commonly a corner or the center of the image, the axes are defined relative to the image and the position of any point is indexed by the voxel count from the origin. Sometimes it is useful to work in mm and other occasions voxel co-ordinates are more convenient, a transformation matrix exists for converting between the two standards.

2.3.2 Transformation Models

A transformation model describes a set of transformations that can be imposed upon an image to alter its shape and structure. The transformation model defines which spatial operations can be performed during the registration.

There are three core choices of transformation models which define varying degrees of expressive freedom; a *rigid transformation model* preserves distances and angles in the image; an *affine transformation model* preserves parallel lines and a *non-rigid transformation model* does not preserve straight or parallel lines.

The domain of the transformation can be defined as *global* if it applies to the entire image, and the transformation is defined as *local* if the deformation varies over regions of the image.

Both rigid and affine transformation models express global spatial transformations, whereas non-rigid transformation models explain local, non-linear variations in anatomy.

Non-rigid transformation models are useful for formulating the local changes in the image. To achieve this it is often necessary to ensure there is a initial global correspondence between the image pair first. Initially, a registration with global transformation model is applied to the image pair to resolve global alignment. The result from this is used as an initialisation for a further registration with local or non-rigid transform model. The resultant transformation from such a combination of transformations can be written as:

$$\boldsymbol{T}(\boldsymbol{p}) = \boldsymbol{T}_{global}(\boldsymbol{p}) + \boldsymbol{T}_{local}(\boldsymbol{p}) \tag{2.16}$$

In the subsequent sections we shall discuss each of these three classes of transformation models in more detail.

Rigid Transformation

The rigid transformation model includes translations and rotations. When the transformation model is considered for a 3D image, it has six degrees of freedom. Translations in direction of the x, y and z axis in addition to rotation about each of these axis. A rigid transformation can be defined by a rotation matrix $\boldsymbol{Ro}$ plus a translation $\boldsymbol{tr}$:

$$\boldsymbol{p}' = \boldsymbol{Ro} \cdot \boldsymbol{p} + \boldsymbol{tr} \tag{2.17}$$

The translation vector $\boldsymbol{tr}$ can be used to define a translational displacement of size t_x in the x axis, t_y in the y axis and t_z in the z axis.

Rotations in 3D can be represented as $\boldsymbol{Ro} = [ro_{i,j}]_{3\times3}$. A rotation around the x, y and z axis can be defined as $\boldsymbol{Ro}_x(\theta_x)$, $\boldsymbol{Ro}_y(\theta_y)$ and $\boldsymbol{Ro}_z(\theta_z)$. The rotation matrix $\boldsymbol{Ro}$ is constructed through composition of the rotations matrices $\boldsymbol{Ro}(\theta_x, \theta_y, \theta_z) = \boldsymbol{Ro}_x(\theta_x)\boldsymbol{Ro}_y(\theta_y)\boldsymbol{Ro}_z(\theta_z)$. It is

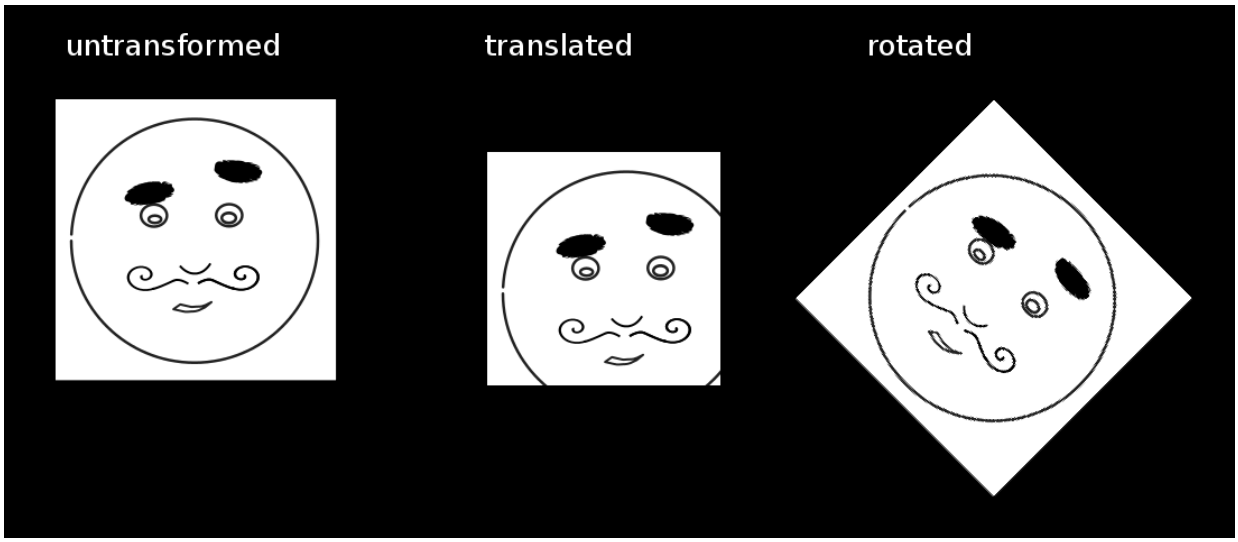


Figure 2.2: Example of a rigid transformation.

important to maintain a consistent ordering when composing rotation matrices because of the non-commutativity property of matrices. For mathematical convenience both the translational and rotational operations are formulated as a single matrix using homogeneous coordinates.

$$\boldsymbol{T}_{rigid}(\boldsymbol{p}) = \boldsymbol{p}' = \begin{pmatrix} i' \\ j' \\ k' \\ 1 \end{pmatrix} = \begin{pmatrix} ro_{00} & ro_{10} & ro_{20} & tr_x \\ ro_{01} & ro_{11} & ro_{21} & tr_y \\ ro_{02} & ro_{12} & ro_{22} & tr_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} i \\ j \\ k \\ 1 \end{pmatrix} \tag{2.18}$$

Affine Transformation Model

The affine transformation model is an extension of the rigid transformation model described in section 2.3.2. Additional operations include skewing (in xy, yz and xz) and scaling (in x, y and z). The resulting transformation model has 12 degrees of freedom.

An affine transformation written as an affine matrix ($\boldsymbol{Af}$) plus a translation vector ($\boldsymbol{tr}$).

$$\boldsymbol{p}' = \boldsymbol{Af} \cdot \boldsymbol{p} + \boldsymbol{tr} \tag{2.19}$$

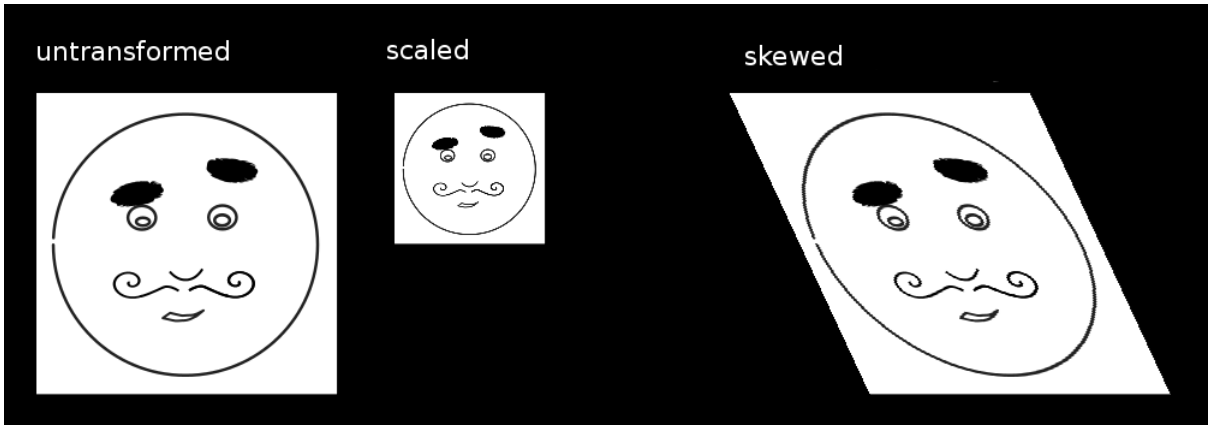


Figure 2.3: Example of an affine transformation.

The affine transformation matrix $\boldsymbol{Af}$ is a composition of a rotation matrix transformation ($\boldsymbol{Ro}$), a skew transformation ($\boldsymbol{Sk}$) and a scale transformation ($\boldsymbol{Sc}$). Again, order is important due the non-commutative property of matrices.

$$\boldsymbol{Af} = \boldsymbol{Ro} \cdot \boldsymbol{Sk} \cdot \boldsymbol{Sc} \tag{2.20}$$

As with the rigid transformation model, the affine transformation model can also be defined in homogeneous coordinates:

$$\boldsymbol{T}_{affine}(\boldsymbol{p}) = \boldsymbol{p}' = \begin{pmatrix} i' \\ j' \\ k' \\ 1 \end{pmatrix} = \begin{pmatrix} af_{0,0} & af_{1,0} & af_{2,0} & tr_x \\ af_{0,1} & af_{1,1} & af_{2,1} & tr_y \\ af_{0,2} & af_{1,2} & af_{2,2} & tr_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} i \\ j \\ k \\ 1 \end{pmatrix} \tag{2.21}$$

Properties of Rigid and Affine Transformation Models

Global transformations can be written as $\boldsymbol{A} = [a_{i,j}]_{4\times4}$ where $\boldsymbol{T}(\boldsymbol{p}) = \boldsymbol{A} \cdot \boldsymbol{p}$. As such, many simple properties of matrices apply to global transformations that can be written in matrix form. Explicitly we draw the reader's attention to composition, inversion, mean transformations and the identity property.

Transformation Composition Homogeneous co-ordinates enable efficient composition of transformations. An image, $\boldsymbol{I}_S$, initially transformed by transformation $\boldsymbol{T}_A$ and then $\boldsymbol{T}_B$ is equivalent to $\boldsymbol{I}_S$ undergoing transformation $\boldsymbol{T}_A \circ \boldsymbol{T}_B$, or $\forall \boldsymbol{p} \in \boldsymbol{\Omega}_{S,T} : \boldsymbol{A} \cdot (\boldsymbol{B} \cdot \boldsymbol{p}) \implies (\boldsymbol{A} \cdot \boldsymbol{B}) \cdot \boldsymbol{p}$. The operation however, is not commutative, i.e. $\boldsymbol{T}_A \circ \boldsymbol{T}_B \neq \boldsymbol{T}_B \circ \boldsymbol{T}_A$.

Inverse Transformations The inverse of a transformation, $\boldsymbol{T}^{-1}(\boldsymbol{p}') = \boldsymbol{p}$, can computed by inverting the transformation matrix. The inverse of a matrix can only be computed if the matrix is non-singular. However, since singular matrices do not represent a valid affine transformation which can be decomposed into a linear transformation, this does not pose any practical problem.

Mean Transformation The mean transformation ($\bar{\boldsymbol{T}}$) of n transformations $\boldsymbol{T}^1, \boldsymbol{T}^2, \ldots, \boldsymbol{T}^n$ can not be computed trivially by averaging the elements of the transformation matrix; $\bar{\boldsymbol{T}} \neq [\bar{t}_{i,j}]_{4\times4}$ where $\bar{t}_{i,j} = \frac{\sum_k^n t_{i,j}^k}{n}$. This is because affine transformations do not exist in a linear space. Instead the Frechet mean can be computed. This is discussed in more detail in section 6.2.2, where it is also applied.

Non-rigid Transformation Model

Non-rigid transformations are defined by a local transformation. The number of degrees of freedom can be large (up to the number of voxels in the image). In the case where the number of degrees of freedom equals the number of voxels, a deformation is defined for each voxel. It is useful for the deformations computed to be both smooth and invertible, such that there are no discontinuities in the transformed images.

Smooth deformations can be achieved by modelling the deformation to be elastic (Bajcsy and Kovačič (1989)), fluid (Crum et al. (2005)) or diffeomorphic. The latter models the deformations as geodesic flows expressed as vector field which is constrained to be smooth (Beg et al. (2005)). Within this thesis the free-form deformation (FFD) model (Rueckert et al. (1999)) is used to

achieve good anatomical correspondences between globally registered images. The FFD model is briefly reviewed here.

Free-form Deformation Model (FFD) Free-form deformations (FFDs) were first developed and used by Sederberg and Parry (1986) for computer graphics applications and first applied to medical image registration by Rueckert et al. (1999). FFDs are defined as a regular mesh of control points which are uniformly spaced. Manipulation of a control point defines a deformation to the mesh and this can be used to deform the image. B-splines are locally-controlled blending functions. They are used to smoothly approximate the displacements of the control points and so can be used to model FFDs. B-splines are computationally efficient with a large number of control points since manipulating a control point only affects the transformation in the local neighbourhood of that control point. B-splines also are more effective for modelling local deformations than those which use radial basis function, as these have infinite support and a global influence upon the deformation field.

The displacement field $\boldsymbol{u}$ can be computed at each voxel using a B-spline function in equation 2.22. The function is a 3D tensor product of 1D cubic B-splines. It blends the effect of neighbouring control points on the mesh of $n_x \times n_y \times n_z$ control points $\phi_{i,j,k}$ with uniform spacing δ:

$$\boldsymbol{u}(x,y,z) = \sum_{l=0}^{3}\sum_{m=0}^{3}\sum_{n=0}^{3} \theta_l(u)\theta_m(V)\theta_n(w)\phi_{i+l,j+m,k+n} \tag{2.22}$$

Here $i = \left\lfloor \frac{x}{\delta} \right\rfloor - 1$, $j = \left\lfloor \frac{y}{\delta} \right\rfloor - 1$, $k = \left\lfloor \frac{z}{\delta} \right\rfloor - 1$, $u = \frac{x}{\delta} - \left\lfloor \frac{x}{\delta} \right\rfloor$, $v = \frac{y}{\delta} - \left\lfloor \frac{y}{\delta} \right\rfloor$, $w = \frac{z}{\delta} - \left\lfloor \frac{z}{\delta} \right\rfloor$ and θ_l represents the lth B-spline basis function;

$$\begin{aligned}
\theta_0(s) &= (1-s)^3/6 \\
\theta_1(s) &= (3s^3 - 6s^2 + 4)/6 \\
\theta_2(s) &= (-3s^3 + 3s^2 + 3s + 1)/6 \\
\theta_3(s) &= s^3/6
\end{aligned}$$

FFDs based B-splines can introduce folding into the overall deformation. This means that the

transformation is non-invertible and may not reflect the true anatomy of structures in medical images. Regularisation can be used to enforce smoothness to prevent such undesirable effects Rueckert et al. (1999).

2.3.3 Similarity Metrics

Similarity metrics are used in the registration framework to evaluate the quality of spatial correspondence between two images. To achieve the best spatial correspondence, the similarity metric is maximised through optimisation methods.

Similarity metrics are not only used for registration in this thesis. After registration the image intensities still differ with regards to pathological status or inter-subject variations. Similarity metrics are applied to observe variation of pathology across a large population, in chapters 3 and 4.

Sum of Squared Differences

The Sum-of-squared difference (SSD) can be defined as the Euclidean distance of the intensities at each voxel between both images $\boldsymbol{I}_T$ and $\boldsymbol{I}_S$. It is computed where images overlap $\boldsymbol{p} \in \boldsymbol{\Omega}_{T,S}$.

$$SSD(\boldsymbol{I}_T, \boldsymbol{I}_S) = \frac{1}{N} \sum_{\boldsymbol{p} \in \boldsymbol{\Omega}_{T,S}} \mid \boldsymbol{I}_T(\boldsymbol{p}) - \boldsymbol{I}_S(\boldsymbol{p}') \mid^2 \tag{2.23}$$

SSD is minimised since a perfect registration would have SSD of zero. Sum-of-squared differences is an appropriate measure if the images differ only by Gaussian noise (Viola and Wells (1995)). When registering images of different modalities (e.g. CT/MR), the intensities of structures differs by significantly more than Gaussian noise. Futhermore the difference between two MR images acquired on different occasions is non-Gaussian, due to the bias field and artefacts of MR. SSD can only be used on mono-modal registration. It is also sensitive to a small number of voxels that have very large intensity differences (Crum et al. (2004); Hajnal and Hill (2010)), i.e. it is sensitive to outliers. In the case where it is necessary to reduce the influence

Figure 2.4: Registration Demo: A source image is being registered and warped to a target image. The source image is displayed in its native co-ordinate space, after an affine registration, a B-spline non-rigid freeform deformation and finally with the deformation grid.

of outliers, the L_1 norm distance between the intensities can be computed. This is referred to as the sum of absolute differences.

Normalised Cross Correlation

A registration algorithm which uses sum of square differences (SSD) as similarity metric relies on the assumption that once the images are registered, they differ only by Gaussian noise. When normalised cross correlation (NCC) is used as a similarity metric the assumption is relaxed and it is assumed that corresponding intensities in the images have a linear relationship (Crum et al. (2004); Hajnal and Hill (2010)). Normalised cross correlation is also referred to as the correlation coefficient and is the normalised version of the cross correlation measure. Normalised cross correlation (NCC) is defined as

$$NCC(\boldsymbol{I}_T, \boldsymbol{I}_S) = \frac{\sum_{\boldsymbol{p}\in\boldsymbol{\Omega}_{T,S}}(\boldsymbol{I}_T(\boldsymbol{p}) - \bar{\boldsymbol{I}}_T)(\boldsymbol{I}_S(\boldsymbol{p}') - \bar{\boldsymbol{I}}_S)}{\sqrt{\sum_{\boldsymbol{p}\in\boldsymbol{\Omega}_{T,S}}(\boldsymbol{I}_T(\boldsymbol{p}) - \bar{\boldsymbol{I}}_T)^2 \sum_{\boldsymbol{p}\in\boldsymbol{\Omega}_{T,S}}(\boldsymbol{I}_S(\boldsymbol{p}') - \bar{\boldsymbol{I}}_S)^2}} \tag{2.24}$$

Here $\bar{\boldsymbol{I}}_T$ and $\bar{\boldsymbol{I}}_S$ are defined as the mean intensity of the images and $\boldsymbol{p}$ is the index of a voxel within image $\boldsymbol{I}_T$

Normalised Mutual Information (NMI)

Multi-modal image registration requires computing spatial correspondences between images where a linear relationship can not be assumed between intensities of spatially corresponding voxels. Multi-modal images can include MR imaging with a different acquisition sequence or two different tomographic images, including MRI, CT or PET.

Registration can be thought of as reducing the amount of shared information content in both images. If the images are aligned then corresponding structures overlap which reduces duplication, therefore a successful registration reduces the total information in the combined image pair. Information theory is used to determine non-linear correspondences between voxel intensities in registered images and *information* can be used as similarity metric.

The most commonly used measure of information is called marginal or Shannon entropy (Shannon et al. (1949)). The entropy of a single image can be computed by constructing a histogram, where each voxel is binned based on the intensity value. The marginal probability of a voxel belonging to intensity bin is denoted as $p(t)$. The marginal entropy measure, $H(\boldsymbol{I}_T)$, between image $\boldsymbol{I}_T$ and $\boldsymbol{I}_S$ is the average information supplied by a set of n intensity bins , computed in equation 2.25.

$$H(\boldsymbol{I}_T) = -\sum_{t}^{n} p(t)\log(p(t)) \tag{2.25}$$

The marginal entropy is maximum if all intensity bins have equal probability (i.e. $p(t) = \frac{1}{n}\forall t$). Entropy has a minimum value of 0 if the probability of one intensity bin is 1 (i.e. $\exists i \mid p(i) = 1$). An image contains less information if the intensity histogram is uniform.

Joint entropy measures the amount of shared information content in two or more images. If image $\boldsymbol{I}_T$ and image $\boldsymbol{I}_S$ are very different then the joint entropy ($H(\boldsymbol{I}_T, \boldsymbol{I}_S)$) will equal the sum of entropies in both images. A joint histogram is constructed for image $\boldsymbol{I}_T$ and $\boldsymbol{I}_S$, for any given voxel location in both image domains intensity values from both images are simultaneously binned in the 2D histogram. The 2D histogram is represented as a matrix, each entry corresponds to a histogram bin, $h(t, s)$ represents the number of times a binned intensity pair (t, s) has co-occurred at the same voxel in images T and S. The joint probability distribution function, $p(t, s)$ is computed by normalising the bins of the joint histogram by the total number of voxels ($p(t, s) = \frac{h(t, s)}{N}$ where N is the number of voxels in the image). Joint entropy is computed as $H(\boldsymbol{I}_T, \boldsymbol{I}_S)$ in equation 2.26.

$$H(\boldsymbol{I}_T, \boldsymbol{I}_S) = -\sum_{t}\sum_{s} p(t, s)\log p(t, s) \tag{2.26}$$

A similarity metric that maximises the information content of each image whilst minimising the shared information content can be defined. Mutual information (MI) is a metric which calculates to what extent one image can be explained by another (Viola and Wells (1995));

$MI = H(\boldsymbol{I}_T) + H(\boldsymbol{I}_S) - H(\boldsymbol{I}_T, \boldsymbol{I}_S)$.

NMI corresponds to the ratio of the sum of the marginal entropies to the joint entropy. NMI assumes a probabilistic relationship between intensities. This means that a correspondence can be found from a tissue featuring in images acquired from different modalities and with differing intensities (Crum et al. (2004)). Several versions of the normalised mutual information measure have been proposed, all of which are closely related. In this thesis we adopt the normalisation approach proposed by Studholme et al. (1998).

$$NMI(\boldsymbol{I}_T, \boldsymbol{I}_S) = \frac{H(\boldsymbol{I}_T) + H(\boldsymbol{I}_S)}{H(\boldsymbol{I}_T, \boldsymbol{I}_S)} \tag{2.27}$$

Other Similarity metrics

There are other similarity metrics which have been introduced for registration. A recent contribution is a multi-modal similarity measure introduced by Wachinger and Navab (2010), they find new structural representations of images using manifold learning, such that the similarity of the images can be defined as a simple L_1 or L_2 norm. This is computationally more efficient than normalised mutual information.

2.3.4 Optimisation Methods

In most registration algorithms an optimisation process is required to find transformation parameters which maximises or minimises a cost function for the two images. The cost function is optimised by either a maximising a similarity measure (such as CC or NMI) or minimising a dissimilarity measure (such as SSD). The optimal transformation is one which most faithfully maps or warps one image to another using the operations permissible by the transformation model.

Optimisation algorithms determine the solution by iteratively generating a series of estimates starting from an initial guess. The starting guess is a transformation which has to be sufficiently

close to the optimal transformation for the algorithm to converge to the correct solution. The range between the initial starting point and the optimal solution required for convergence is known as the "capture range". Following the initial guess, further transformations are estimated iteratively with the cost function recalculated at each iteration. A cost function or similarity measure explains the quality of registration between two images. An optimal registration is achieved by finding a transformation which maximises (or minimise) the cost function. A review of optimisation algorithms can be found in Press et al. (2007).

For generality, we shall from hence forth refer to optimisation of cost functions as maximisations of similarity.

Gradient Descent

In this thesis gradient descent is used to optimise rigid, affine and non-rigid registrations. For consistent terminology, Gradient Ascent would be a more appropriate nomenclature if we consider the optimisation process maximises image similarity.

The set of transformation parameters $\boldsymbol{\phi}$ are adjusted iteratively such that upon each iteration the similarity is maximised. The optimisation steps along the direction of the steepest gradient upon each iteration. For a user specified step size of δ and a similarity metric S, the transformations parameters are updated upon the ith iteration as follows:

$$\mathbf{\Phi}_{i+1} \leftarrow \mathbf{\Phi}_i + \delta \nabla_{\mathbf{\Phi}} S \tag{2.28}$$

The gradient $\nabla_{\mathbf{\Phi}} S$ can be determined analytically for some similarity metrics such as SSD. Otherwise it can be found using a finite difference method. As the optimisation converges to the solution, successively smaller step sizes can be used to refine the solution.

Local optimum

A common limitation of optimisation algorithms is a tendency to be trapped by local optima with the solution failing to converge to the global maximum similarity.

The space of all solutions of the similarity function can be viewed as a similarity landscape, where each transformation parameter describes a dimension in the solution space. Sampling this space with a set of transformation parameters yields the similarity of the image. The landscape is a surface in the solution space, with peaks representing maxima and valleys representing minima. If the algorithm converges to a local maximum, this point on the similarity landscape is where the similarity is not the highest possible (a peak of a foothill). The transformation parameters to which this corresponds yields a sub-optimal or locally optimal registration.

Some strategies exist to avoid the optimisation getting trapped in local maxima. These include a multi-scale framework, centring and regions of interest, which will be briefly overviewed.

Multi-resolution Strategies Registration algorithms commonly use a hierarchical coarse-to-fine strategy to reduce the likelihood of convergence to a local optimum. Initially, transformation parameters are approximated coarsely to resolve correspondences between large structures. The coarse solution should be closer to the global optimum than the initial guess. The coarse solution forms the initial guess for a series of subsequent registration where the transformation parameters are refined. This strategy has been applied in two ways: with multi-resolution images (Studholme et al. (1997)) and a hierarchy of B-spline control points for non-rigid FFD registration (Schnabel et al. (2001)).

For registration using a multi-resolution approach, both images are blurred using a Gaussian kernel of width σ and then downsampled. An initial coarse registration is computed with a large step size. The transformation parameters are subsequently re-optimised using high resolution images with a smaller σ and smaller step size.

2.3.5 Measuring registration accuracy

It is important to be able to assess the quality of registration between different methods but it is very challenging to provide a suitable measure for registration accuracy. Whilst common anatomical structures exist between subjects, a one to one mapping of anatomy across subjects simply does not exist. Structures might either differ substantially or simply not be present in

an image. With higher resolutions this limitation becomes more apparent. For example there certainly is not a one to one mapping between cells in histology imaging. It is thus challenging to define what a *correct* transformation between two subjects would be. Despite challenges, it is useful to devise a metric which provides a simplified measure of success of registration.

Surrogate measure of registration accuracy

It is possible to use patient outcomes measures to estimate registration accuracy. Measuring registration success on the basis of patient outcomes convolves the quality of the alignment with the quality of the diagnosis pipeline (Fitzpatrick (2001)), for which the registration is just a component. As a result it is desirable to decouple the testing of predicted patient outcomes and registration accuracy.

Registration Error Metrics

Registration error metrics measure the error between markers or points located in two medical images after a registration. These can be measured in two different ways: fiducial registration errors (FRE) and target registration errors (TRE).

Fiducial Registration Error (FRE)

A fiducial marker is an physical object attached to subject's anatomy which is visible in the acquired medical image and can be used as a point of reference or a measure. Attached fiducial markers can be used to establish accurate fiducial points for registration. Fiducial co-ordinates are selected for their locatability but not for their clinical significance.

The Fiducial Registration Error (FRE) is the distance, after registration, between the measured position of the fiducial markers in both images. FRE is commonly reported as a distance between the N fiducial markers, where the ith fiducial markers are referred to as $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$:

$$FRE^2 = \sum_{i}^{N} |\boldsymbol{T}(\boldsymbol{p}_i) - \boldsymbol{q}_i|^2 \tag{2.29}$$

Target Registration Error (TRE)

Target registration errors (TRE) are calculated based using clinically relevant landmarks. The target registration error is a measure of displacement between corresponding landmarks after registration. Let $\boldsymbol{p}$ represent a target point in $\boldsymbol{I}_T$ and $\boldsymbol{q}$ represent a target point in $\boldsymbol{I}_S$. TRE is computed as in equation 2.30 as a vector, however since the direction of the error is frequently unimportant, often the magnitude is reported alone.

$$TRE = \boldsymbol{T}(\boldsymbol{p}) - \boldsymbol{q} \tag{2.30}$$

Localisation Errors

To compute accurate registration errors target points or fiducial markers are required to be accurately located. The errors associate with locating these points are referred to as localisation errors. The Fiducial Localisation Error (FLE) is defined as the distance between the true position of a fiducial marker and its measured position. The Target Localisation Error (TLE) is defined as the distance between the true target point and its manually located position in the MRI.

Localisation errors are present in images prior to registration but these errors will be propagated via registration. Localisation errors are hard to measure directly because it is challenging to determine the true location of such points.

2.4 Knee Osteoarthritis

Osteoarthritis (OA) is a debilitating pathology and the most prevalent amongst joint diseases (Dunlop et al. (2003); Conaghan et al. (2006)). OA imposes a large expense upon society

through medical treatment and inability of individuals to generate an independent income due to disability (Dunlop et al. (2003); March and Bachmeier (1997)). Despite its prevalence and severity, OA still remains poorly understood and a condition for which there are limited effective symptomatic treatments available (Hunter et al. (2009)).

OA is poorly defined and understood. Researchers are as yet unsure what initiates osteoarthritis or in which tissues the pathology originates (Aspden (2008)). It can be challenging to separate the structural changes due to ageing from pathological changes as a result of OA (Conaghan et al. (2006)). It is often considered to be a condition with multiple causes, making the condition harder to define and to establish a relationship between cause and effect.

OA historically has been considered to be a disease which is characterised by the erosion of articular cartilage. OA is now no longer perceived to be a disease solely of the articular cartilage and is recognised to be a multi-factorial process which alters both structure and function of the entire joint (Hunter et al. (2008)). OA is a complex disease where the pathophysiology involves both biomechanics and biochemistry and the treatments range from surgery to nutritional supplements and pharmaceutical products to education of the patient (Felson et al. (2000)). Due to this many different academic research fields are expressing interest in OA. This ranges from imaging sciences to biomechanical engineering, genetics, epidemiology and biochemistry.

2.4.1 Knee Anatomy

An illustration of healthy knee anatomy can be found in figure 2.5. It is important to draw the readers attention to the following structures; bone (tibia, femur, fibular and patella), meniscus (medial and lateral), articular cartilage (tibial, femoral and patellar) and ligaments (the cruciate and collateral ligaments). These are just some of the structures which are relevant upon studying Osteoarthritis. In this section, we shall introduce the function of these structures in a healthy individual.

The knee joint is the largest synovial joint in the body. It consists of what can be regarded as two joints, the joint of the tibia and the femur (tibiofemoral joint), the second is of the

patella and femur (patellofemoral joint). The tibiofemoral joint is the weight-bearing joint whose movement enables the leg to bend, its stability is assisted by the presence of ligaments tethering the joint as its bends. The patella is attached to the quadriceps femoris muscles (at the front of the upper thigh) via the quadriceps tendon this facilitates the straightening as the patella increases the moment and provides a platform for a muscle to pass over the front of the knee without wearing the tendons (Drake et al. (2005)). The stability of the patellofemoral joint is ensured by the geometry of the bone structures and the patella's position in the trochlear grove of the femur.

The **articular cartilage** has hyaline surfaces , which has extremely low friction. The two main surfaces on the tibiofemoral joint are on the lateral and medial femoral condyle and the adjacent surface of the tibial plateaus. The condyles are the two protruding surfaces of the femur and the plateau is the corresponding surface which meet at the joint on the tibia. There are similar surfaces for the patellofemoral joint (Drake et al. (2005)). The articular cartilage provides a smooth and low-friction surface which is able to withstand very large loads with minimal wear (Mow and Huiskes (2005)).

There are two **menisci**, which are "C" shaped piece of cartilage in the knee joint. Both of the menisci (medial and lateral) are attached at each end to the tibial plateau. The menisci improves congruency between the surfaces of the tibia and femur, so that during knee flexion the femoral articular surface is in contact with the tibia. This facilitates load transmission, stability and lubrication (Vedi et al. (1999)).

Tendons attach muscles to bone and ligaments attach bone to bone. There are four key ligaments involved in the tibiofemoral joint. There are two **collateral ligaments**, on each side (lateral and medial) of the joint, they stabilise the hinge motion of the knee. There are also two **cruciate ligaments** (anterior [ACL] and posterior [PCL]). The cruciate ligaments cross each other on the sagittal plane, in the intercondyle region and act to tether the join anteriorly and posteriorly. A torn ACL is a common injury and if not repaired, can be a risk factor for osteoarthritis incidence.

The **synovial membrane** is attached to the two membranes of the articular surfaces and the

outer margin of the menisci. The synovial membrane forms two pouches to provide low friction surfaces for the movement of the tendons (Drake et al. (2005)).

2.4.2 Osteoarthritis Definition

Osteoarthritis is generally characterised by joint pain, cartilage wear and the presences of boney growths. However, the definition of OA is inconsistent in the literature. We adopt the definition of knee OA used by the Osteoarthritis Initiative for the progression subcohort. Subjects have symptomatic tibiofemoral knee OA at baseline if they have both of the following in at least one native knee at baseline:

1. Radiographic tibiofemoral knee OA, defined as definite tibiofemoral osteophytes equivalent to Kellgren and Lawrence (K-L) grade ≥ 2 (Kellgren and Lawrence (1957) and 2.4.8) or OARSI atlas grades 1-3 (Altman and Gold (2007)), on a fixed flexion radiograph
2. Frequent knee symptoms in the past 12 months defined as "pain, aching or stiffness in or around the knee on most days" for at least one month during the past 12 months

This concise definition is limited, other structures visible in MR are known to either be contributors or symptoms of OA. Symptomatic responses also present a challenge to define, i.e. pain is an ambiguous term which is challenging to assess in a reproducible and comparable way.

2.4.3 The Osteoarthritis Initiative

This thesis solely uses data from the Osteoarthritis Initiative (OAI) [http://oai.epi-ucsf.org/datarelease/]. The OAI (Peterfy et al. (2008)) provides a longitudinal public dataset with 4796 participants consisting of medical images, clinical data and biospecimens. It is an ethnically diverse dataset of men and women with participants ranging from 49 to 79 in age, collected at five clinical centres in four sites in the United States: Rhode Island, Maryland, Ohio and Pittsburgh.

The OAI selected the knee joint as a principle focus of the study because this is the site where OA symptoms most frequently cause significant loss of function and disability. The OAI's

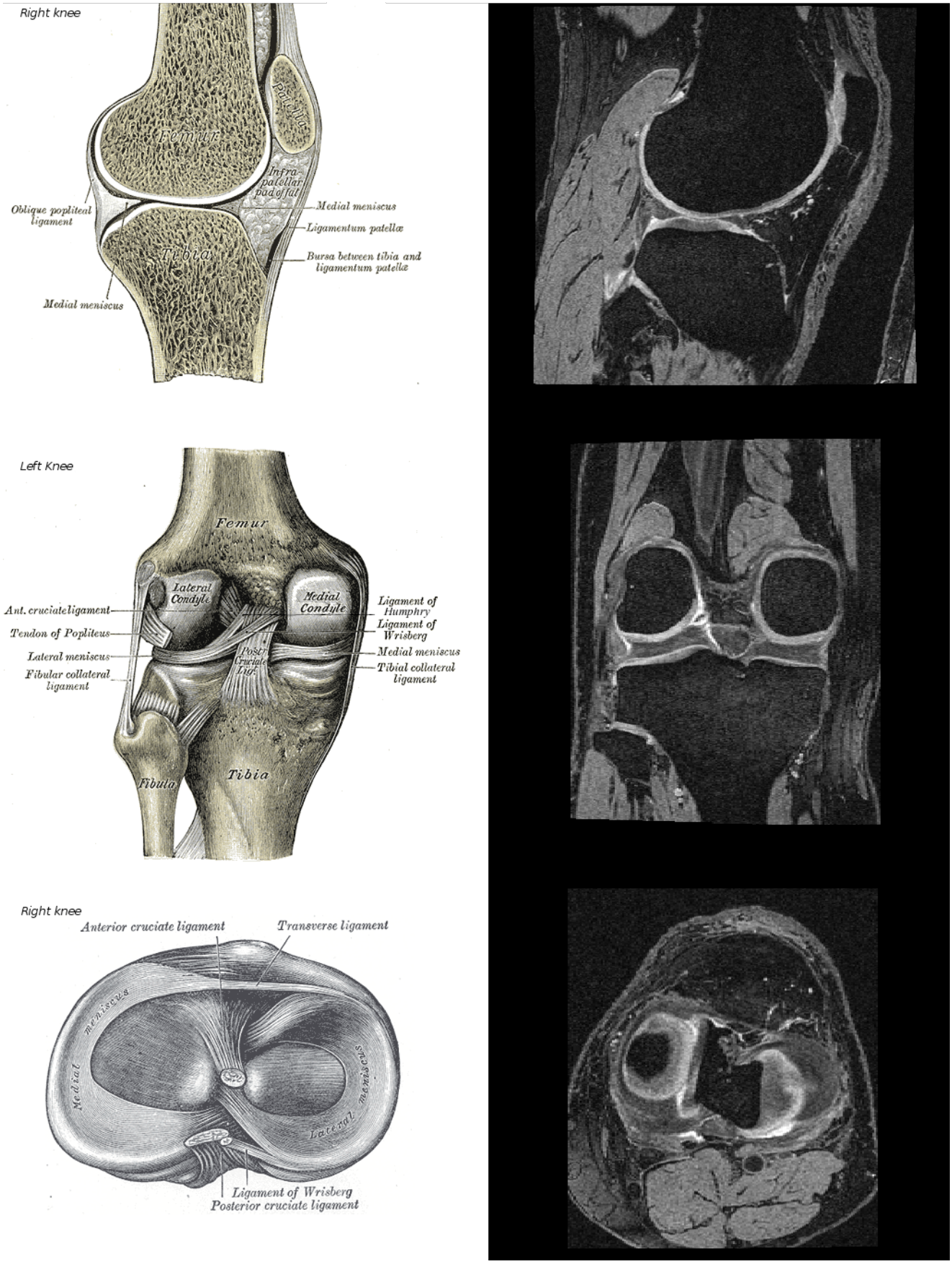


Figure 2.5: Knee anatomy diagram (left), matching radiological view MRI of a subject (right). Note that due to availability of knee illustrations, both left and right knees are displayed but clearly labelled†.

† *Illustrations from http://upload.wikimedia.org/wikipedia/commons [accessed 20/03/13].*

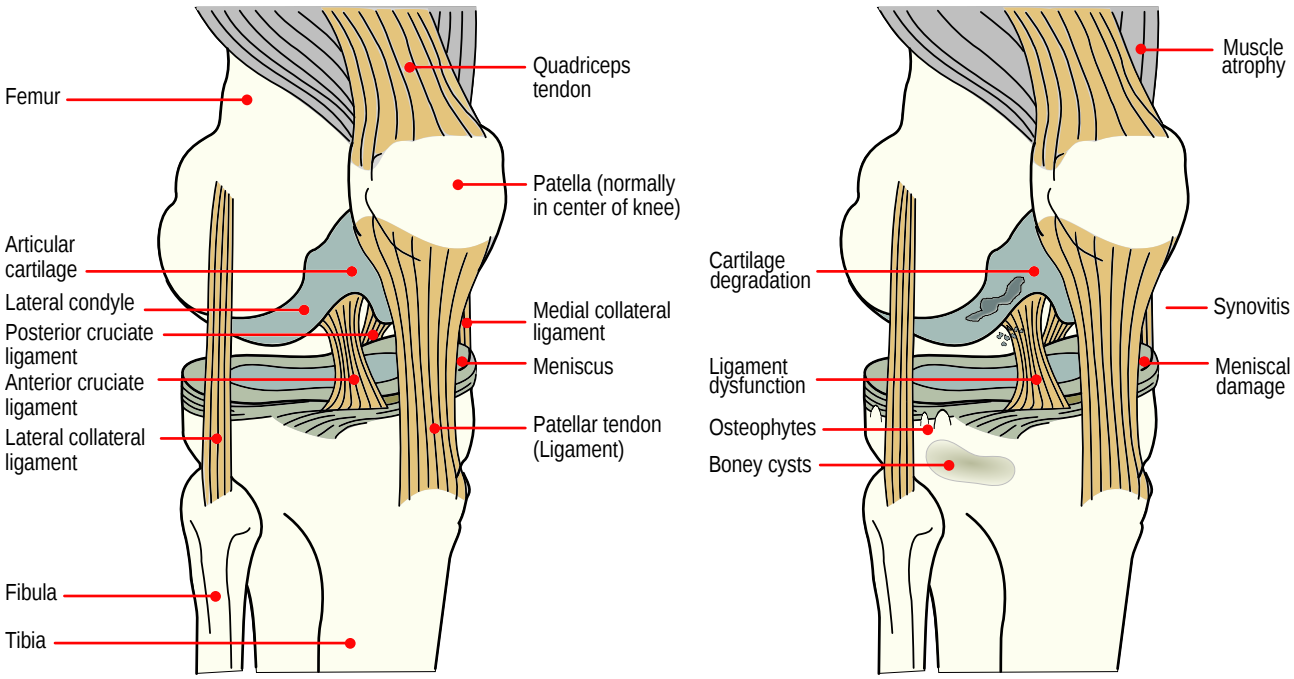


Figure 2.6: Diagram of a normal knee on the left compared with an osteoarthritic degree on the right†.

† *Illustration adapted from commons.wikimedia.org based on Hunter (2011).*

primary objectives are to evaluate radiographic and MR joint images as biomarkers for OA and explore their potential as surrogate endpoints for clinical studies and treatment trials of knee OA.

Joint imaging biomarkers (MR imaging and radiography), biochemical and genetic markers (from blood and urine) are collected at baseline and at all follow-up visits. There are four planned annual follow-up visits. Participants are followed for changes in the clinical status of the knee and other joints, including worsening and onset of symptoms and disabilities, worsening and onset of knee structural abnormalities, and changes in other imaging and biochemical markers of OA.

There are three main subgroups enrolled in the OAI dataset;

1. **Progression** : those with clinically significant tibiofemoral knee OA who are at risk of disease progression. Participants must have pain or risk factors for OA (total=1390).

2. **Incident** : individuals who are at high risk of initiation of clinically significant knee OA (total=3284).

3. **Controls** : non-exposed control subjects who have no pain or risk factors (total=122).

Exclusion criteria are: where there is evidence or suggestion of Rheumatoid Arthritis (RA) in any joint, the subject has severe joint space narrowing in both knees on the baseline knee radiograph (very late stage OA), bilateral total knee joint replacement or plans to have bilateral knee replacement are excluded, a positive pregnancy test, is unable to undergo a 3.0T Tesla MRI exam of the knee or they are unwilling to sign a consent form.

2.4.4 MR Imaging of Osteoarthritis

MR imaging is argued by many to be one of the most useful imaging modalities for structural evidence of knee OA (Conaghan et al. (2006)). Its key advantages lie in its ability to visualise soft tissue structures in 3D. In addition to imaging the bone as with plain radiographs, MR is able to capture ligaments, synovium, menisci and subchondral bone. MRI is also often acquired at high-spatial resolution and is advantageous when compared to X-Ray because there is no ionising radiation.

2.4.5 MR Positioning

The OAI positioning protocol is strict and so some spatial a-priori alignment of the knee can be assumed. The patella is aligned with a point on the knee coil (figure 2.7), all of the knees are in a fixed angle of flexion, the leg is in a relaxed neutral position and the position of the foot and big toe is specified in the protocol. This enables some assumptions to be made about the appearance of the knee MRI in the OAI dataset.

2.4.6 MR Sequences

MR images are acquired using Siemens 3T Trio scanners. Details of the MR sequences from the OAI dataset are described in table 2.2. Examples of coronal MR images can be seen in

Figure 2.7: Photo of MRI knee-imaging coil with liners and pad †.

† *from https://oai.epi-ucsf.org/datarelease/operationsManuals/MRI_knee_coil_size_screenV1_0p.pdf [accessed 15/04/13].*

figure 2.8, examples of sagittal MR images can be found in figure 2.9. Further information can be found in the MRI manual provided by the OAI and from Peterfy et al. (2008).

Peterfy et al. (2008) discusses the design rationale of the various acquisitions and concludes that for quantitative cartilage morphometry, fat-suppressed, 3D dual-echo in steady state (DESS) acquisitions appear to provide the best universal cartilage discrimination.

Some sequences such as Cor T1W 3D FLASH WE and Sag T2 MAP are acquired in the right knee only, unless the right knee is a knee replacement or there are metallic implants or foreign bodies seen on the right knee localizer, in which case these two sequences are performed on the left knee. Since there are more sequences acquired for the right knee, much of the work in this thesis uses the right knee MRIs.

2.4.7 Quantifying Osteoarthritis

It is useful to be able to quantify OA severity and progression both structurally and functionally.

(a) Coronal IW TSE

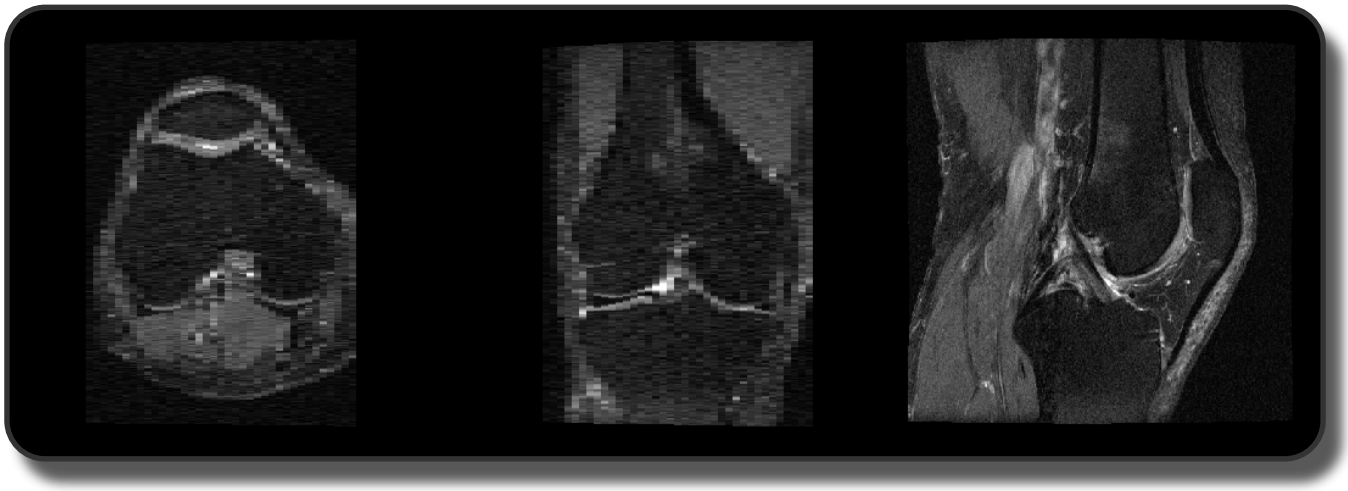
(b) Coronal T1 3D Flash

Figure 2.8: Examples of the MR images acquired in the coronal plane. The images are of a healthy candidates right knee at baseline (id=9093584).

Sequence	Abbreviation	Intended Use	Figures
Coronal Intermediate-Weighted Turbo Spin Echo	Cor IW TSE	used for evaluation of joint alignment, cartilage morphology, osteophytes, the body of the menisci, collateral ligaments and for the presence / extent of subchondral bone cysts and attrition.	2.7 (a)
Coronal T1-weighted 3D Water-Excitation Fast Low Angle Shot	Cor T1 3D FLASH	commonly used for cartilage thickness measurements and volume segmentation.	2.7 (b)
Sagital Intermediate-Weighted Turbo Spin Echo	Sag IW TSE	used for evaluation of the effusion volume, the anterior and posterior femoral and tibial osteophytes and for the presence / extent of subchondral bone cysts and attrition.	2.8 (a)
Sagital Water-Excitation 3D Dual Echo Steady State	Sag 3D DESS	provides information for total joint cartilage thickness and volume. In addition, information about osteophytes, subarticular bone cysts and bone attrition, and possibly collateral ligaments will be available.	2.8 (b)
Sagital T2 MAP	Sag T2 MAP	used to assess the cartilage morphology and condition, the anterior and posterior meniscal horns, the cruciate ligaments, anterior / posterior femoral and tibial osteophytes, superior / inferior patellar osteophytes, as well as subchondral bone cysts and attrition.	2.8 (c)

Table 2.2: MR sequences, their abbreviations, intended uses and example MRI.

(a) Sagital IW TSE



(b) Sagital 3D WE DESS

(c) Sagital T2 MAP

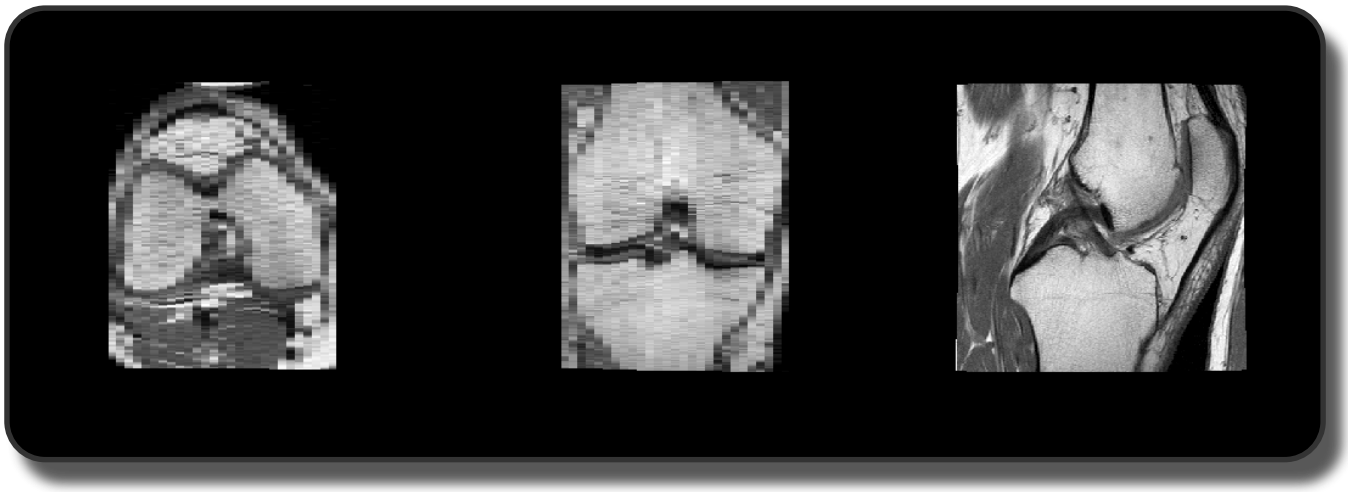


Figure 2.9: Examples of the MR images acquired in the sagittal plane. The images are of a healthy candidates right knee at baseline (id=9093584).

Functional Quantification

In this work we focus upon structural abnormalities and progression of pathology as can be observed with MR imaging. Whilst articular cartilage loss is strongly associated with OA, the cartilage is aneural and as a result does not sense pain. Identifying subjects experiencing pain using MR imaging is an extremely challenging problem.

Pain has been defined by the International Association for the Study of Pain (IASP) as

> "An unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage." (Bonica (1979) (president of IASP)).

Structural Quantification

Medical imaging can capture structural information which could potentially help to extract biomarkers and may lead to an understanding of OA function. To analyse the structural causes and effects of OA pathology upon the knee joint it is highly desirable to extract quantitative metrics or biomarkers from MR images. These metrics can be divided into two categories; quantitative and semi-quantitative scores which are derived from MR scans or X-rays.

2.4.8 Osteoarthritis Imaging Measures and Biomarkers

Biomarkers for OA can be employed to learn risk factors or predictors to help understand the manifestation and progression of the pathology. Biomarkers can also be used for drug discovery, both for stratifying a population for a well-defined sample and to test the efficacy of disease modifying OA drugs (DMOADs). All of these applications benefit from a method that can be used with a large dataset, thus scalability of biomarker extraction is paramount. Accuracy and precision are also important for any inter-subject comparisons.

Many quantitative imaging techniques to assess OA progression or diagnosis rely on the segmentation of the articular cartilage from MRI (Dam et al. (2007); Eckstein et al. (2009)). However,

K&L Grade	Description
0	Normal
1	Doubtful narrowing of joint space and possible osteophytic lipping
2	Possible narrowing of joint space and definite osteophytes
3	Definite narrowing of the joint space, moderate multiple osteophytes, some sclerosis and possible deformity of bone ends
4	Marked narrowing of joint space, large osteophytes, severe sclerosis and definite deformity of bone ends

Table 2.3: K&L scale for OA grading

manual segmentation is often tedious and time-consuming and current automatic segmentation approaches have limited accuracy and robustness. Semi-quantitative assessments such as BLOKS (Hunter et al. (2008)) are also employed and whilst these are less time consuming to read the images, their discretised nature leads to poor precision, limiting their applicability. Additionally, as manual approaches rely heavily on human input, the size of a feasible study is restricted and its results are susceptible to intra-rater and inter-rater variability.

Kellgren and Lawrence grades (K&L)

Kellgren and Lawrence grades (K&L) can be used for quantifying radiographic OA (Kellgren and Lawrence (1957)). Table 2.3 details the grades assigned for varying severities of OA. In summary, it assesses joint space width, osteophytes and subchondral bone sclerosis but many more structures than just these are involved in OA, where knees whose K&L grade is $\geq$2 are considered to be osteoarthritic. It is a course grading system from 2D radiographs and so it is not possible to assess soft tissue pathology, therefore it is a fairly crude measure of OA severity.

Semi-Quantitative OA Biomarkers

Trained readers read and assess medical images to give scores of OA status as a predefined combination of discretised ratings of the status of several tissues associated with OA, not limited to the articular cartilage. Examples of semi-quantitative OA scores include the Kellgren and Lawrence grade (K&L) (Kellgren and Lawrence (1957)), as well as Boston Leeds osteoarthritis knee score (BLOKS) (Hunter et al. (2007)) and whole-organ MRI scoring method (WORMS) (Peterfy et al. (2004)) which are derived from reading MRI.

Semi-quantitative approaches have the advantage of being faster than metrics which require a manual segmentation. However, they still require trained readers which imposes some limits (financial and time) upon the scale of the study. Additionally, the granularity of the results are lower than their quantitative counterparts due to its discrete scale and the technique can still fall foul of manual errors. Manual methods inherently struggle with inter- and intra-reader reproducibility when compared with automated methods.

Quantitative OA Scores

Quantitative measures typically extract biomarkers from a segmentation of the articular cartilage. Quantitative methods initially obtain a segmentation or delineation of organ interfaces and then extract morphometric measurements which can be used as OA biomarkers. These segmentations are either acquired manually (Eckstein et al. (2009); Hunter et al. (2009)), semi-automatically (Stammberger et al. (1999); Grau et al. (2004)) or fully-automatically (Yin et al. (2010); Fripp et al. (2010); Folkesson et al. (2007); Vincent et al. (2010)).

Cartilage morphology metrics computed from segmentations have been shown to differ significantly for different segmentations teams (Schneider et al. (2012)). Due to the extent of this variation the author recommends readers do not pool results from different teams unless equivalence can be demonstrated. Ideally a good biomarker could be computed by multiple teams and the results would be comparable, this is a limitation associated with this approach.

Manual Segmentation Manual delineation of the articular cartilage interface is performed by trained readers and it is reported to take approximately 2 hours to segment the medial tibial and femoral cartilage (Folkesson et al. (2007)). Additionally some groups choose to perform a quality control step to ensure inter-scan consistency (Eckstein et al. (2009)), which increases costs and reduces the potential scale of a study due to time constraints. Quality control is designed to make assurances about reproducibility but it is still inherently problematic due to human decision making being a core element of the procedure. Due to the time consuming nature of manual segmentation many studies are unable to exploit the potential value of large cohorts.

Automated Segmentation Automated segmentation of the articular cartilage is a challenging problem and how to perform accurate segmentations automatically still remains an open question and an active area of research. The challenges are predominantly attributed to its convex, thin and flat structure, partial volume effects and low contrast with surrounding tissues and fluid.

Current state of the art segmentation techniques for articular cartilage involve voxel-wise classifiers (Folkesson et al. (2007); Dam and Loog (2008); Yin et al. (2010)), active appearance models (Vincent et al. (2010); Fripp et al. (2010)), graph cuts (Yin et al. (2010)), watershed (Grau et al. (2004)) and many other methods. At MICCAI 2010, a Grand Challenge was introduced to segment knee MR images of both bone and articular cartilage. The winning team presented a fully automatic method based on active appearance models (Vincent et al. (2010)). The MICCAI Grand Challenge has now been extended under the name SKI10, the data provided for the challenge and the collated results are presented as a league table on the webpage (http://www.ski10.org/).

Leading state-of-the-art automated segmentation techniques typically achieve excellent specificity but poor sensitivity (sensitivity is reported as 75%-87% (Folkesson et al. (2007); Fripp et al. (2010); Yin et al. (2010)). This reveals a strong trend of under-segmentation, or misclassification of positive voxels. This can have a deleterious effect on the accuracy and reliability of morphometric measures, such as thickness, surface area and volume, which are extracted

Author	Biomarkers	Ground Truth	AUC (95% Confidence Interval)	Validation Set
Dam et al. (2009)	"cartilage longevity" automated segmentation required Folkesson et al. (2007)	diagnosis from K&L grade	0.84 (0.77 to 0.92) classifying K&L <2 0.82 (not stated) classifying K&L <1	159 participants (287 knees) - 51% healthy (K&L=0), - 30% boarderline OA (K&L=1), - 10% mild OA (K&L=2), - 9% moderate/severe OA (K&L= 3 or 4)
Eckstein et al. (2011)	dGEMRIC manual segmentation required	diagnosis from K&L grade (table 1)	0.738 classifying healthy vs radiographic OA (ROA)	152 participants (152 knees) - 49% healthy (K&L 0) - 51% ROA (K&L 2 or 3)

Table 2.4: Summary state-of-the-art OA biomarkers. "cartilage longevity" is a combination of biochemical (CTX-II) and imaging measures (cartilage volume, area, thickness, congruity, roughness, and homogeneity).

from the segmentations. Typically, global segmentation results are reported whilst regional segmentation results are often omitted. Automated segmentation methods that only report global accuracies could mask regional inaccuracies, potentially further harming the accuracy of morphometric measures.

The size of the validation set for automated segmentation techniques reported in the literature has been fairly small to date, with Folkesson et al. (2007) having the largest sample size of 114 participants, the work by Yin et al. (2010) tested 60 subjects and Fripp et al. (2010) was only tested on 20 subjects. The limited sample sizes can probably be attributed to the fact that the ground truths for these experiments require manual segmentation. Since this ground truth is time consuming to acquire, the feasible test sizes are limited. Many experiments have been performed predominately or entirely healthy knees. In Folkesson et al. (2007) the test sample was constructed of 70% with K&L grade 0 or 1. In Yin et al. (2010) the sample contained 80% K&L grade 0 or K&L grade 1 and in Fripp et al. (2010) the sample consisted of 100% K&L grade 0. Whilst this sample was probably determined by available data, it should be remembered that these experiments generally avoided participants with severe OA, which are more challenging to segment. This is likely to result in a poorer segmentation results and thus poorer late stage biomarkers.

OA Imaging Biomarker Validation Protocol

Bauer et al. (2006) suggests new diagnostic tests should be evaluated by comparison against an established gold standard in an appropriate spectrum of subjects. Bauer et al. (2006) specifically suggests that a diagnostic test using radiographs are a suitable option, typically a Kellgren-Lawrence (K&L) grade > 2 is required for a diagnosis of OA. K&L is a radiological measure primarily concerned with the presence of osteophytes (boney spurs on the bone margin). K&L grading is limited as it does not consider all manifestations of the pathology, it is derived from an X-ray which is 2-D, soft tissues are not visualised and it is inconsistently applied across many studies (Felson et al. (2011)). Despite these limitations, K&L grades are widely used and provided for many subjects in the OAI dataset therefore it is still important for our algorithms to be validated using this grading system.

State of the Art Quantitative Imaging Biomarkers

Studies have been performed to assess change in cartilage morphology as a biomarker with varying rates of success. After acquiring an accurate segmentation through either manual or automated approaches, morphometric biomarkers can be extracted. Morphometric biomarkers reported are using volume (Raynauld et al. (2004); Dam et al. (2007); Wildi et al. (2013)), surface area (Dam et al. (2007)), thickness (Dam et al. (2007); Eckstein et al. (2009)), homogeneity (Qazi et al. (2007)), curvature (Folkesson et al. (2008); Dam et al. (2007)), congruity (Dam et al. (2009)), roughness (Dam et al. (2009)), percentage of denuded bone area (Graichen et al. (2004)) and a combination study of some of the previously mentioned techniques (Dam et al. (2009)).

A summary of the strongest biomarkers in the literature can be found in table 2.4. To date, the MRI based biomarker to achieve the highest AUC is presented by Dam et al. (2009). It combines multiple MRI imaging biomarkers with known biochemical biomarkers to find improved classification results for identifying subjects with OA. Much of the literature focuses on increasing the accuracy of automated segmentation. Very few of these works follow on the work to produce quantitative biomarkers for OA. Since the ultimate goal of segmenting

osteoarthritic knees is diagnostics or prognostics, it is disappointing to not see this application of the segmentations reported.

Limited work has been presented to date regarding subregional biomarkers. A regional approach for acquiring morphometric biomarkers has been recognised as important by Buck et al. (2009); Wirth et al. (2010). Wirth and Eckstein (2008) first proposed regional analysis, dividing the articular cartilage into 16 arbitrary anatomically regions or interest using a method which they find to be highly reproducible. This is later developed by the author such that the regions are defined so that they are meaningful with respect to cartilage change for subjects with OA progression (Wirth et al. (2010)). Buck et al. (2009) addresses how to analyse regional data. They propose ranking change in thickness of the articular cartilage for subregional compartments by magnitude and direction. This enables comparison of regions with greatest change regardless of anatomical location and is called ordered values. This ordered value approach was found to be more sensitive than comparing thickness for corresponding regions across subjects. The ordered values enables studies to observe OA progression independent of the specific anatomic location in the joint, but in its current form does not make any attempt to combine information from multiple regions.

To our knowledge no biomarkers have been proposed for combining information from multiple MR sequences. There is evidence that different sequences provide useful and complementary information for OA diagnosis. However, to date the discussions have mostly focused on finding the most appropriate sequences (Peterfy et al. (2008); Eckstein et al. (2006b)) as opposed to combining information from multiple imaging sequences.

2.4.9 Summary

OA is a multifactorial disease with a complex inter-relation between structure and function, much of the recent contributions regarding OA imaging focus on the articular cartilage. With this in mind, many of the methods presented in the thesis focus on the articular cartilage to enable comparison and validation with results in the literature. For generality of future applicability to OA diagnosis the methods presented are designed to be generalisable and can

be applied to any structure within the entire joint that is visible on MRI.

Chapter 3

Manifold Learning Based Regression of Clinical Variables

Work in this chapter has, in part, been presented in:

C. Donoghue, A. Rao, A. M. J. Bull, and D. Rueckert. Manifold learning for automatically predicting articular cartilage morphology in the knee with data from the osteoarthritis initiative (OAI). *SPIE Medical Imaging 2011 : Image Processing*, 7962, 2011.

C. Donoghue, A. Rao, A. M. J. Bull, and D. Rueckert. Articular Cartilage as an Automatic Predictor for Cartilage Morphology with Data from the Osteoarthritis Initiative. *4th International Workshop on Imaging Based Measures of Osteoarthritis*, 2010.

(2nd prize) C. Donoghue, A. Rao, A. M. J. Bull, and D. Rueckert. Articular Cartilage as an Automatic Predictor for Cartilage Morphology with Data from the Osteoarthritis Initiative. *Poster Competition Research Students' Summer Symposium 2010, Graduate School of Engineering and Physical Sciences, Imperial College*, 2010.

Abstract

This chapter looks to manifold learning as an automatic approach to harness the plethora of imaging data provided by the Osteoarthritis Initiative (OAI). We construct spectral embeddings of articular cartilage appearance from MR images of the knee using multiple MR sequences. A region of interest (ROI) defined as the weight bearing medial femur is automatically located in all images through non-rigid registration. A pairwise intensity based similarity measure is computed between all images, resulting in a fully connected graph, where each vertex represents an image and the weight of the edges expresses the similarity of the images. Spectral analysis is then applied to these pairwise similarities, which acts to reduce the dimensionality non-linearly and embeds these images in a manifold representation. In the manifold space, images that are close to each other are considered to be more "similar" than those far away. We use manifold learning to automatically predict the morphological changes in the articular cartilage by using the co-ordinates of the images in the manifold as independent variables for multiple linear regression. In the study presented here five manifolds are generated from five MR sequences. We observe statistically significant correlations (up to $R^2 = 0.75$) between our predictors and the results presented in the literature.

3.1 Introduction

This chapter aims to identify a concise, low dimensional representation of a large population of MR images, such that the relationship between images can be described by only a few explanatory variables. It is possible to define a high dimensional representation of the imaging dataset by observing and recording a large number of features. However high dimensional data can often be expressed with fewer degrees of freedom than the number of features used to observe the data. The observed data is likely to incorporate redundant features and redundant combinations of features which can be excluded from a concise representation.

The OAI dataset contains a very large sample of subjects with a large variation in natural appearance and varying appearance due to pathology. These variations in the appearance of knee MRI can be regarded as continuous. As a result we can expect the data points to lie on or

near a lower dimensional manifold in the observed feature space. Due to noise in the model we would not expect the data points to lie directly on the manifold. In this work we attempt to approximate the structure of the manifold using manifold learning or dimensionality reduction. Manifold learning can be used to construct a representation of data lying on a low dimensional manifold embedded in a high-dimensional space (Cayton (2005)).

Manifold learning has been applied across the field of medical imaging, Wolz et al. (2010) used manifold learning for segmenting brain MR images, Wachinger et al. (2010) used manifold learning to detect the position of the body from full body MRI and Yang et al. (2008) analysed cyclical motion patterns such as beating of the heart in ultrasound images.

Imaging biomarkers of osteoarthritis can potentially enable clinicians to determine severity of disease and speed of progression. Some biomarkers have been proposed for which a segmentation is used to obtain morphological measures such as thickness and volume (Dam et al. (2007); Eckstein et al. (2009)). However, under-segmentation is common problem for automated segmentation of the articular cartilage. If the segmentation is inaccurate, this can result in an unreliable estimate of morphology.

Manifold learning is used here to learn a concise low dimensional representation of a large dataset of knee MR images with and without pathology. With the help of machine learning and statistical analysis these feature vectors from the new representation can be used to generate effective image based predictors and classifiers.

3.2 Method

3.2.1 Overview

This chapter proposes a predictor for cartilage morphology from a single region of interest for a large population study. Initially, an approximate automatic segmentation based non-rigid registration to a reference image is used to identify a well known region of interest, the central medial femur. A spectral embedding is computed based on the pair-wise similarities of the

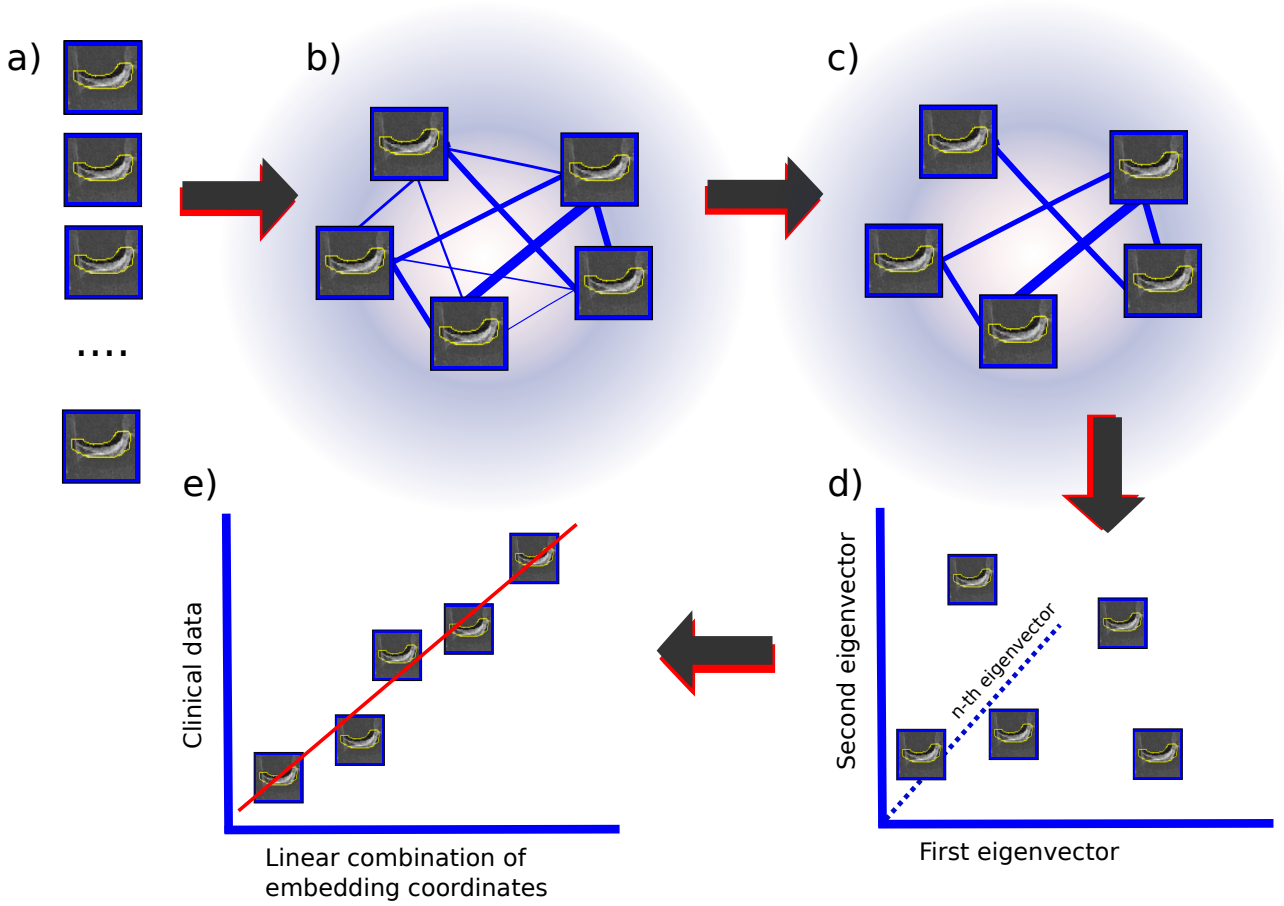


Figure 3.1: Pipeline of the framework for a single MR sequence. In step a), the region of interest has been identified in every image using registration (sections 3.2.2 and 3.2.3). Steps b) and c) depict the knn graph construction, where edges of low similarity are discarded. The spectral embedding is represented by step d) (section 3.2.4). Multiple linear regression is used to learn a linear combination of embedding coordinates to predict clinical data in step e) (section 3.2.5).

region of interest for every image. The embedding co-ordinates are used in a multiple linear regression framework to find a linear combination of embedding components which have a strong correlation with clinical data. The clinical data chosen is a morphological measure of articular cartilage which is often used as diagnostic biomarkers for osteoarthritis diagnostics in the knee. The method is well suited to a large dataset, such as that provided by the OAI. An overview of the proposed approach is depicted in diagram 3.1.

3.2.2 Registration

This chapter aims to compare a pathologically significant anatomical region of interest across many subjects with and without symptomatic and radiographic osteoarthritis. To achieve this it is efficacious to compute a transformation between all subjects such that a mapping exists to a common coordinate space which is suitable for voxel-wise comparison. Given a sample of n distinct knees and corresponding MR images, $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_n\}$, a single subject is selected as the reference image $\boldsymbol{I}_i$ to which all other images will be registered to derive n transformations. These transformation map all images into a common space. The reference image is chosen randomly from patient knees with Kellgren and Lawrence (K-L) grade 0 (radiographically normal) (Kellgren and Lawrence (1957)) and who experience no symptoms.

Each image is initially rigidly transformed to allow for position and orientation variation within the scanner. This is followed by an affine registration of the knee joint region to allow for scale discrepancies between images. Finally, each image is coarsely non-rigidly registered to $\boldsymbol{I}_i$, by a free-form deformation based on B-splines (Rueckert et al. (1999)) using normalised mutual information as a voxel similarity measure (Studholme et al. (1998)) and a 10mm B-spline control point spacing. It is important that the image is aligned with a coarse control point spacing to ensure the region of interest contains the area of interest whilst preserving some of the relevant structural differences.

Region of Interest to Improve Registration Accuracy In some circumstances it can improve registration results to exclude some regions from registration. By excluding irrelevant tissues higher accuracy results maybe obtained for the region of study. Hajnal et al. (1995) found it useful to exclude the soft tissue of the face to achieve precise rigid registration of the brain. By selecting a smaller region of interest, a speed increase is achieved through a faster similarity computation and in the case of FFD registration, the number of parameters to be optimised is reduced. As a result non-rigid registrations are computed over the joint space as defined by a bounded box. Figure 2.4 shows registration pipeline for a single MR sequence, from affine registration to a non-rigid deformation with the deformation grid. The bounding

box around the joint region can be seen in the freeform deformation grid visualised in the bottom sub-figure. The bounding box is where the grid appears deformed.

3.2.3 Region of Interest (ROI) Definition

A segmentation of the reference image, $\boldsymbol{I}_i$, can be propagated to all other images through non-rigid registration. This technique is referred to atlas-based segmentation (Dawant et al. (1999)). Atlas-based segmentation is well studied, and uses registration to yield accurate segmentations. Many extensions to the basic framework have been proposed to improve the quality of segmentations (Heckemann et al. (2006); Aljabar et al. (2009)). However our method does not require exceptionally accurate delineation of organs and as such the approach adopted here is chosen for its simplicity.

The region of interest in the reference image, $\boldsymbol{I}_i$ was segmented from a mean image in the reference space. The mean image is computed as the mean voxel-wise intensity from all images non-rigidly registered to $\boldsymbol{I}_i$ (described in 3.2.2). When registrations are of high quality, a high contrast mean image is produced, this lends itself to a semi-automatic region-growing segmentation (Adams and Bischof (1994)). This suffices as an approximate segmentation of all images. The segmentation is then dilated by 5 voxels to allow for small errors in the registration and to ensure it encompasses the defined region in every image.

For the experiments presented in this chapter, the region of interest selected is the weight-bearing region of the medial femoral condyle shown in figure 3.2. It was selected because it was found to incur the greatest rate of change in articular cartilage morphology in two recent one year follow-up studies (Hunter et al. (2009); Eckstein et al. (2009)). Eckstein et al. (2006a) proposed a nomenclature which defines anatomical labels of the knee joint cartilage. We follow the definition for the central medial femoral condyle (cMF). Furthermore, Eckstein et al. (2006a) have released publicly available data regarding the cartilage morphology of this well defined ROIs.

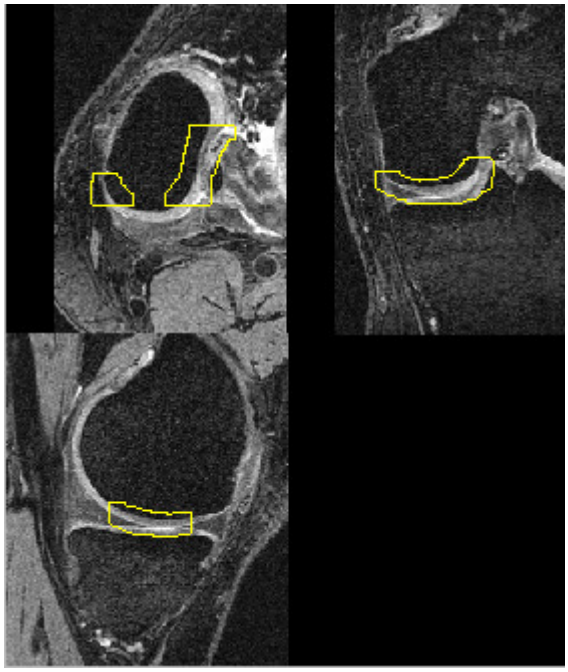


Figure 3.2: Weight bearing region of the medial femoral condyle (cMF) region of interest shown on the reference image.

3.2.4 Manifold Learning

Spectral analysis learns low dimensional features based upon similarities between data points of high dimensions. It is an unsupervised algorithm which preserves the underlying local structure to yield a lower dimensional embedding. More information can be found in Chung (1997) and a thorough tutorial is provided by von Luxburg (2007).

In the high dimensional space each data point represents an image, explicitly the images $\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_n$ are represented as a high dimensional point cloud $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \in \mathbb{R}^M$. The algorithm finds an embedding which preserves local similarities between images. The positions of images in the embedding are described as $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n \in \mathbb{R}^m$ where $m \ll M$. An overview of the manifold learning procedure employed in this section is visualised in figure 3.3.

The data is represented in high dimensions as a fully connected graph where each vertex of the graph ($\boldsymbol{x}$) corresponds to an image ($\boldsymbol{I}$) and each edge is associated with a weight which defines the similarity between the connected pair of vertices. An affinity matrix $\boldsymbol{W} = [w_{i,j}]_{n \times n}$ is computed, $w_{i,j} \geq 0$ is computed by a pairwise similarity measure between two images $\boldsymbol{I}_i$ and $\boldsymbol{I}_j$. For this work the similarity metric is defined as normalised mutual information, NMI

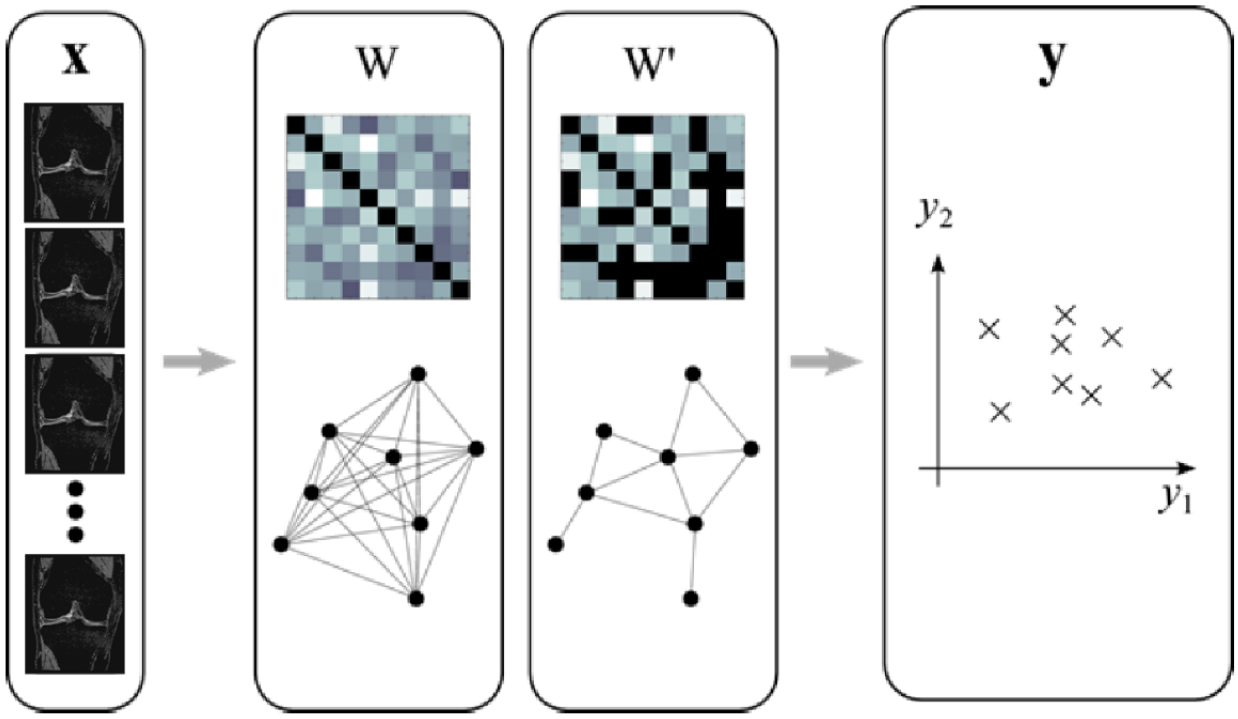


Figure 3.3: Overview of manifold learning. Laplacian Eigenmap embedding finds an appropriate low dimensional embedding $\boldsymbol{Y}$ from input images $\boldsymbol{X}$, by representing all images based on their pairwise similarities $\boldsymbol{W}$, and computing the k-nearest neighbours such that connections are retained for most similar images to find $\boldsymbol{W'}$. $\boldsymbol{Y}$ is a low dimensional embedding such that images which are represented as similar in $\boldsymbol{W'}$ are also close in $\boldsymbol{Y}$. ,

(Studholme et al. (1998)). The similarity measure is computed over an ROI which is defined in section 3.2.3.

From this high dimensional representation a sparse k-nearest neighbour graph is constructed (the parameter choice of k is discussed in section 3.3.2). The connectivity and edge weights of the graph are represented as matrix $\boldsymbol{W}'$ where $w'_{i,j} = 0$ if $\boldsymbol{I}_i$ is not within k nearest neighbours of $\boldsymbol{I}_j$. The weight matrix encodes the local similarities in the graph. The degree matrix $\boldsymbol{D}$ is a diagonal matrix defined by the degree or contribution of each vertex $d_{i,i} = \sum_{j=1}^{n} w_{i,j}$

We compute spectral analysis to preserve image similarities with the normalised graph Laplacian. The normalised graph Laplacian matrix $\mathfrak{L}$ is defined as

$$\mathfrak{L} = \boldsymbol{D}^{-1/2}(\boldsymbol{D} - \boldsymbol{W'})\boldsymbol{D}^{-1/2} \tag{3.1}$$

The embedding co-ordinates are derived by solving the generalised eigenvalue problem. This algorithm is closely related to a Laplacian eigenmap embedding (section 2.2.7).

3.2.5 Predictor of Cartilage Morphology

The representation computed by manifold learning can be used to quantify disease severity in previously undiagnosed subjects. We determine whether there is a linear relationship between these co-ordinates and clinical data. To investigate this link we use multiple linear regression. The clinical data that we are interested in are quantitative morphological changes in the articular cartilage, which are considered to be biomarkers of structural osteoarthritis. If such a relationship exists, we can predict the biomarkers using our automated learning approach.

Multiple linear regression can be used to model the relationship between a response variable (biomarker or clinical data) and more than one explanatory variables (manifold embedding co-ordinates) by fitting a linear model to the observed data. The m explanatory variables are from the manifold embedding are $\boldsymbol{Y}$. The response vector are clinical variables which have been observed for each subject ($\boldsymbol{\alpha}$). Multiple linear regression is used to compute the coefficients (

β_0, β_1,..., β_m) for the explanatory variables that are required to satisfy equation 3.2 with an associated error $\boldsymbol{\epsilon}$ (assuming $i \in [1, n]$).

$$\alpha_i = \beta_0 + \beta_1 y_{1,i} + \ldots + \beta_m y_{m,i} + \epsilon_i \tag{3.2}$$

The coefficients are described as *partial* regression coefficients because they allow for the effect of other variables and in combination can explain clinical data with some residual error. The unknown coefficients β_0, β_1, ..., β_m are computed using a least squares approach, so that the squared deviations of the observed data to the predicted data is minimised. The quality of the model is assessed by observing the R^2 and p-values. The R^2 explains how well the data points fit the model. We observe the F-test p-value to check for a significant linear regression relationship between the response variable and the explanatory variables. More information can be found in the background chapter section 2.2.5.

3.2.6 Incorporating Multiple MR Sequences

We have proposed how to compute a manifold for a single MR sequence. In this section we propose an extension for efficiently computing a manifold for other sequences.

For all n knees we have MR images acquired in ρ sequences. The images are indexed as $\boldsymbol{I}_\phi^\varphi$ where $\phi \in [1, n]$ and $\varphi \in [1, \rho]$. The reference image described in 3.2.2 was acquired using MR sequence j and subject index i is now referred to as $\boldsymbol{I}_i^j$. To compute embeddings for the $\rho - 1$ other sequences the ROI needs to be located in all $n \times \rho$ images.

Registration and ROI definition

The OAI protocol specifies that each sequence is acquired during the same visit, with strict alignment and positioning with the aid of an extremity coil and without removing the subject from the scanner between sequence acquisitions. As a result, we can expect there to be no non-rigid variation caused by pose or anatomical variation between MRI sequences acquired from

the same subject at the same time-point. It follows that we can resolve alignment discrepancies between any two images of the same subject index ϕ via a rigid-registration.

Non-rigid registrations have been computed between all images acquired using the reference sequence j, $\{\boldsymbol{I}_1^j, \boldsymbol{I}_2^j, \ldots \boldsymbol{I}_n^j\}$ and the reference image $\boldsymbol{I}_i^j$. Image $\boldsymbol{I}_\phi^\varphi$ with subject index ϕ, where $\phi \in [1, n]$ and MR sequence φ, where $\varphi \neq j$, is rigidly registered to the reference knee in the selected sequence $\boldsymbol{I}_\phi^j$. This approach ensures there are no global alignment between MR images of the same subject on the same visit, and reduces the total number of non-rigid registrations computed. This rigid registration is composed with a non-rigid registration between $\boldsymbol{I}_\phi^j$ and the reference image $\boldsymbol{I}_i^j$, so that a transformation exists between image $\boldsymbol{I}_\phi^\varphi$ and the reference image. The new registration pipeline which incorporates multiple sequences is illustrated in figure 3.4.

Manifold Learning and Predicting Morphology

All other steps in the algorithm described remain identical to the process for the reference sequence. This includes computing a manifold embedding for the sequence over the ROI and analysis of the correlation of the embedding coordinates with clinical data.

3.3 Experiments

3.3.1 Data

Our study focuses on five MR sequences of 390 right knees from the OAI public use dataset from groups 1.C.0 and 1.E.0. Of these 50 knees have evidence of radiographic and symptomatic OA, 218 are at risk of developing OA or have OA in their other knee (some of these knees may in fact also have radiographic evidence of joint degradation but not OA) and 122 are healthy control participants. Subject data selected for the study are required to have all five sequences acquired.

For this experiment we propose to use cartilage thickness and volume as predicted values. The

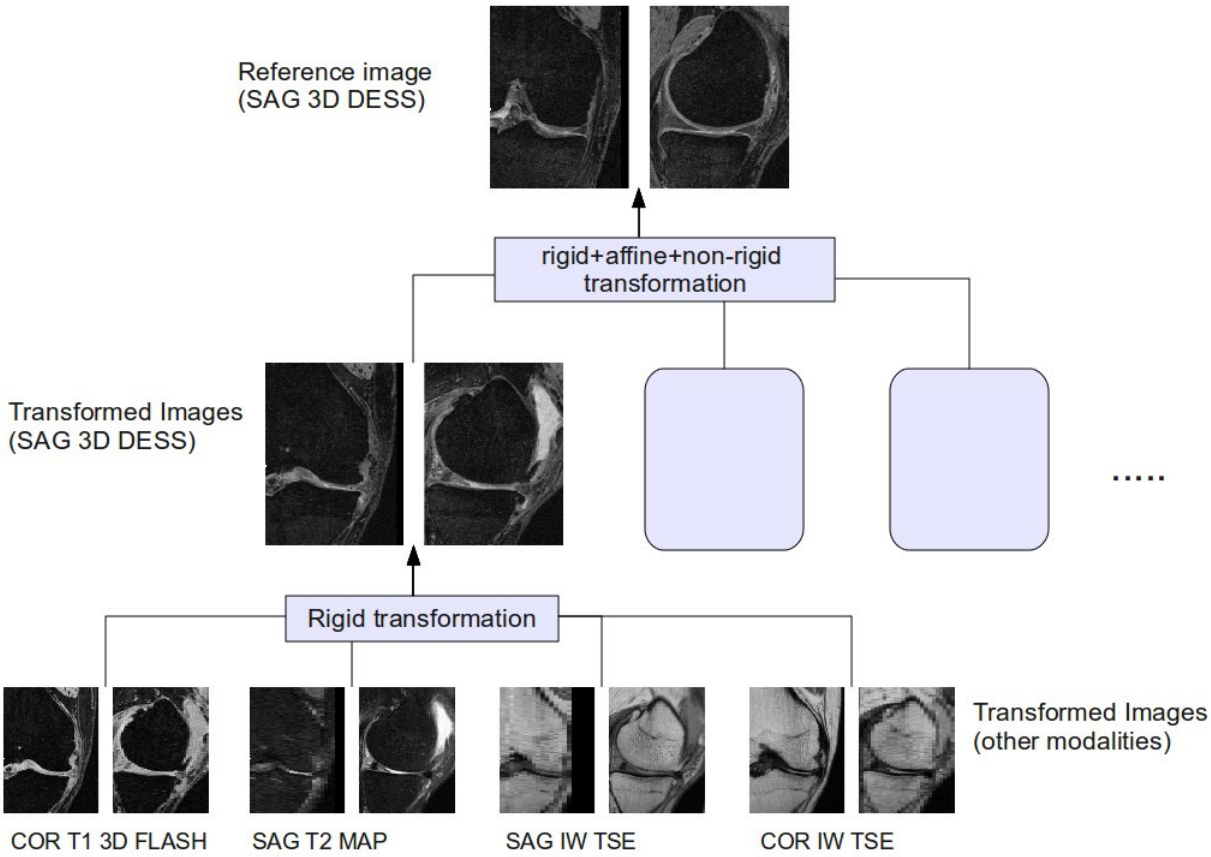


Figure 3.4: Schematic of registration pipeline used to incorporate multiple MR sequences using the non-rigid components of the principle sequence.

data selected as the ground truth for cartilage morphology is provided by the OAI and published in Eckstein et al. (2009) (project 7) for a subcohort of 120 progression enrollees. Additionally, 270 more knees for which clinical data is unavailable are selected to ensure that the manifold is well sampled and thus can be modelled more reliably.

The Osteoarthritis Initiative (OAI) provides five MR sequences ($\rho = 5$) for the right knee of each subject at the baseline time point. The Sag 3D DESS is selected as the reference sequence j because it is a high-resolution image and thus more faithfully represents the anatomy than 2D sequences for 3D registration. The other four sequences which are used are Sag IW TSE, Cor IW TSE, Sag T2 MAP and Cor T1 3D FLASH. Details of these MR sequences can be found in the section 2.4.6 and their use within the registration scheme can be seen in figure 3.4.

3.3.2 Selecting the Graph Neighbourhood

It is important to identify a suitable neighbourhood size (k) for the definition of pairwise similarities in the high dimensional space such that the representation is geometrically meaningful with regards to the underlying manifold. Since the underlying structure varies between manifolds, k is chosen individually for each manifold. Figure 3.5 shows R^2 values for varying neighbourhoods used to create five different manifolds. The observed clinical data used for parameter selection of k is mean thickness excluding denuded bone area.

k is selected for each manifold by maximising the predictive power of multiple linear regression against observed clinical data for a sample of possible k values. k is selected when R^2 is large, however as can be seen in figure 3.5 R^2 fluctuates as k varies. Therefore, we find a k where over a small range of $k \pm 5$, the mean R^2 is maximised. This enables us to avoid picking a k where the corresponding R^2 value is an unrealistically high. It is important that all results are statistically significant and as such it is ensured that $p < 0.05$.

From figure 3.5 it can be seen that a high R^2 value can be found for a fully connected similarity graph ($k = n$). Despite this, R^2 is less when $k = n$ than the peak observed R^2 value when $k \ll n$. Additionally, a fully connected graph is not geometrically meaningful on a non-linear

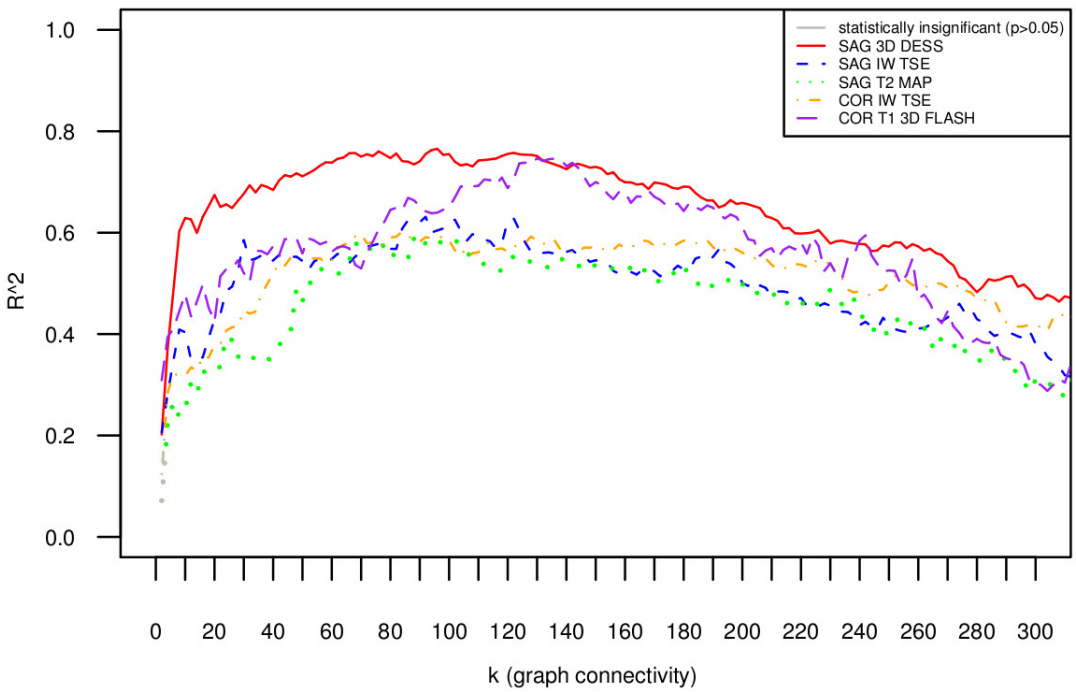


Figure 3.5: R^2 computed from multiple linear regression using the first 14 eigenvectors as predictors for quantitative variable mean cartilage thickness (excluding denude bone area) with a varying neighbourhood of the similarity graph (where $k < 300$). This has been performed for five different MR sequences. Statistically insignificant regions are 'greyed-out', such that they can be identified and ignored.

manifold where local measures of similarity express the manifold structure, as a result $k > 300$ is considered invalid.

3.3.3 Selecting Number of Predictors

It is necessary to select the number of predictors for multiple linear regression. In our case the number of predictors is the same as the number of dimensions used from the low dimensional manifold space (m). It is advisable that $m \ll M$ both for the purposes of manifold learning and multiple linear regression.

In figure 3.6, we observe the relationship between the number of predictors m and the predictive power of the model R^2. The mean R^2 is presented, which is the mean computed over the range $k \pm 5$ (where R^2 was maximised), as described in section 3.3.2. This is to ensure that the selection of k is robust to fluctuations. Figure 3.6 shows results for fixed k with results for each

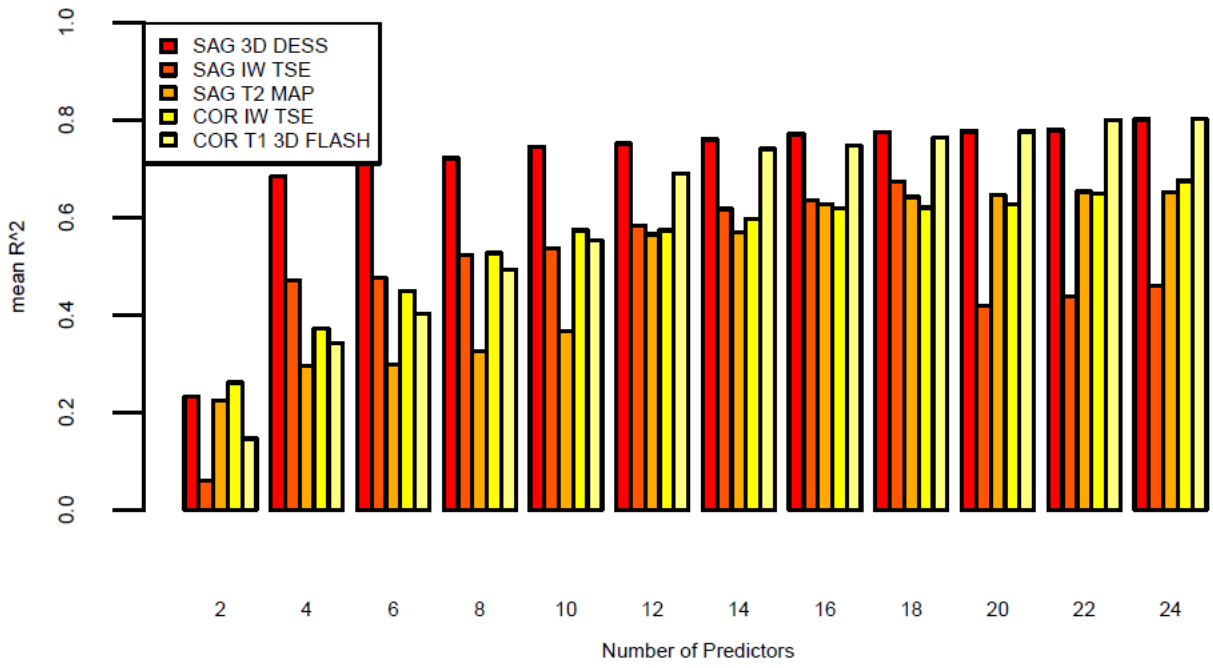


Figure 3.6: Mean R^2 (a local average of R^2 over a range of $k \pm 5$) computed from multiple linear regression with a variable number of predictors (m). This is visualised for five MR sequences. When the rate of change of R^2 reduces the extra information being contributed by the predictors is having less influence on the prediction result.

sequence.

As the number of predictors (m) increases the R^2 value increases initially, however this increase later decays. m is chosen where the rate of change in R^2 reduces since as the improvement in R^2 plateaus, additional dimensions do not contribute notable extra information from the manifold. However, including extra predictors will generally improves R^2 but this does not necessarily mean those predictors are important for the prediction of the outcome of interest.

3.3.4 Results

Visualising the Low Dimensional Embedding

Figure 3.7 depicts a low dimensional embedding ($m = 2$). Only two dimensions are selected purely for visualisation. However, subsequent dimensions still contain valuable information regarding the manifold and are utilised for data analysis. Despite this the visualisation is

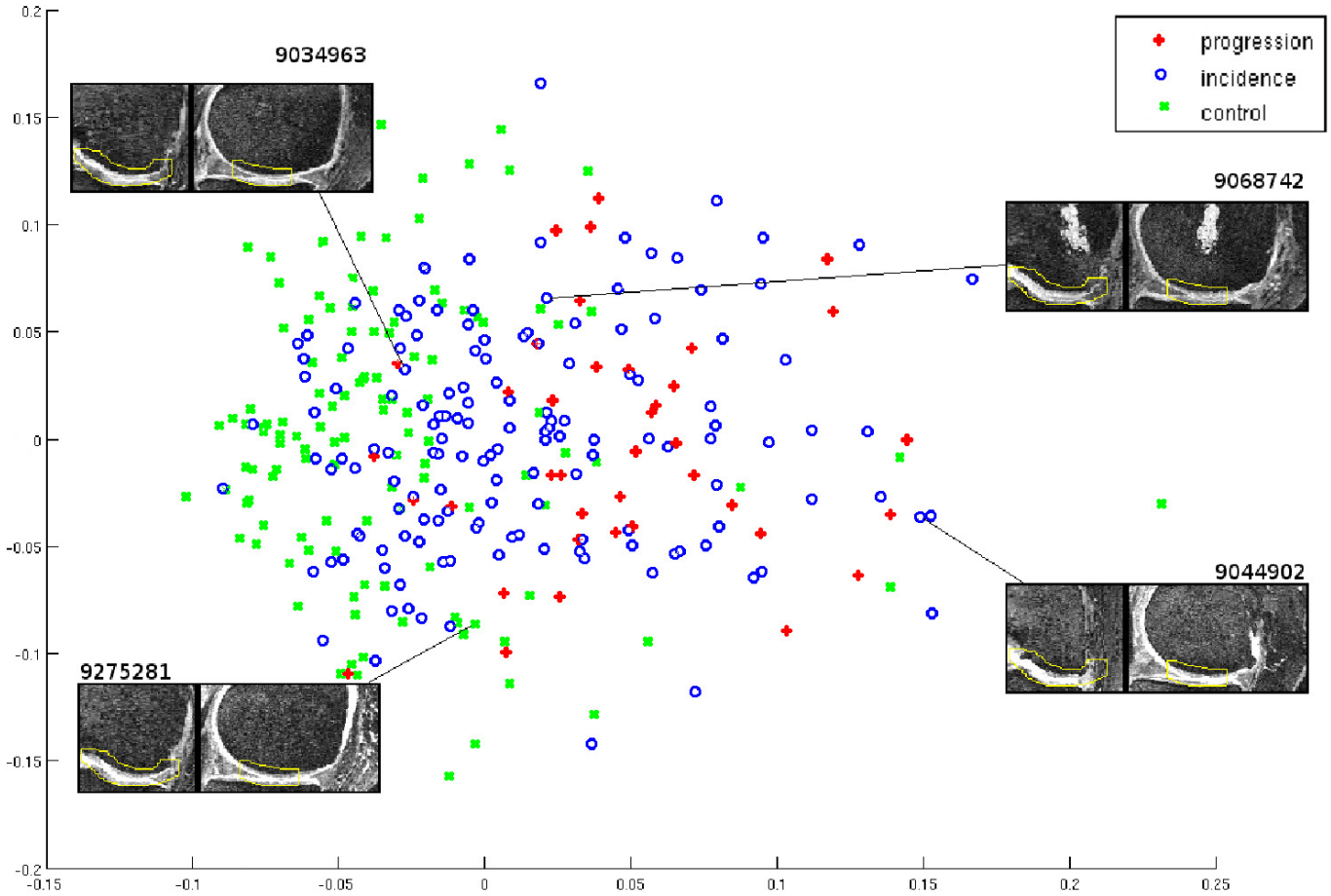


Figure 3.7: MR images embedded in a 2D manifold. The different symbols describe the OA status of the knee.

able to show the images embedded in lower dimensions and clearly shows a cluster of non-symptomatic knees on the left side of the figure .

Multiple Linear Regression

Our results from the multiple linear regression demonstrate there is a high correlation between the structure of the manifold space and the cartilage morphology in the region of interest. Specifically the results display strongest correlation with volumetric images and clinical variables which observe thickness or normalised volume. A more thorough analysis can be found in table 3.1

This enables the co-ordinates of the images in the manifold space to act as a predictor of cartilage morphology for other images in the cohort.

	Sag 3D DESS ($k = 134 \pm 5$, $n = 10$) Mean R^2	Cor T1 3D FLASH ($k = 84 \pm 5$, $n = 14$) Mean R^2	Sag IW TSE ($k = 96 \pm 5$, $n = 14$) Mean R^2	Cor IW TSE ($k = 78 \pm 5$, $n = 16$) Mean R^2	Sag T2 MAP ($k = 102 \pm 5$, $n = 16$) Mean R^2
ThCcAB	0.746	0.741	0.617	0.619	0.628
ThCtAB	0.737	0.740	0.593	0.618	0.609
VCtAB	0.745	0.738	0.596	0.609	0.610
VC	0.644	0.642	0.555	0.572	0.590

ThCcAB – mean thickness excluding denuded bone area;
ThCtAB – mean thickness including denuded bone area;
VCtAB – normalised volume
VC - volume (not normalised)

Table 3.1: Results of multiple linear regression, using the co-ordinates from the manifold as predictor for known quantitative clinical variables of cartilage morphology published in Eckstein et al. (2009) , where a neighbourhood size is chosen for each manifold individually. p is always $< 0.1e - 13$ for the R^2 values quoted.

3.4 Discussion

Manifold learning, to our knowledge, has not been applied to MRI of the knee and provides an excellent way to analyse a large cohort of MR images automatically. Spectral analysis maps the images from a high dimensional space to a low-dimensional manifold representation, whilst preserving local similarity in high dimensions. Relationships between images are represented as a graph with local similarities encoded as edges and the images as vertices. The manifold representation respects local neighbourhood similarity relationships. We are then able to efficiently analyse these relationships in this lower dimensional space and use the manifold representation to derive predictors of morphology.

Our results suggest that the Sag 3D DESS sequence out-performs the other MR sequences for predicting cartilage morphology in the central medial femoral region of the articular cartilage in the right knee. The results also show that fewer dimensions (m) from the manifold are required to predict the clinical variables from Sag 3D DESS images. Despite this it should be noted that the registration technique applied relies on only computing a non-rigid transformation between the Sag 3D DESS acquisitions and all other transformation are derived from this computation. Whilst this produces good results for the other sequences, it is possible the difference in accuracy could be attributed to a poor non-rigid transformation.

The method presented here lends itself to a large dataset. The dimensionality of the embedding space be significantly smaller than the observed feature space ($m \ll M$). The experiments here found that $m = 10$ is a suitable dimensionality for the embedding space. We suggest that this might be an appropriate approximation of the intrinsic dimensionality of the manifold.

It should be noted that approximately 4% of images do not register successfully during the first stage of the method. Within this sample of images, errors were always due to poor initial rigid and affine registration, usually as a result of large difference in initial joint location or large variation in image appearances. This was fixed by adjusting rigid and affine parameters and then re-running the pipeline, specifically adjusting the region of interest to the joint area in image $\boldsymbol{I}_j$. Chapters 5 and 6 seek to automatically test and improve the success of rigid and affine registration.

3.4.1 Contributions

This chapter demonstrates that a data driven manifold learning algorithm can learn a representation of the appearance of knee MRI without supervision. For each data point $\boldsymbol{x}$ in the high dimensional representation, a corresponding data point, $\boldsymbol{y}$, is learnt using spectral analysis. The manifold embedding preserves the local similarities between data points in the high dimensional space when represented in a lower dimensional space. The resulting manifold is non-linear. The data points in the manifold representation are compared to clinical patient outcomes which are considered important in osteoarthritis progression. It is observed that there is a high correlation between the embedding co-ordinates and the clinical data. This initial piece of work on exploring biomarkers has also looked at discovering a suitable dimensionality for the manifold embedding and finding a suitable k-nearest neighbour parameter.

This proposed algorithm could potentially lead to using the embedding co-ordinates as a predictor of cartilage morphology. This contribution could be clinically significant as a fully automated method as a surrogate measure of cartilage morphology. The metric would be reproducible and not susceptible to errors associated with human error from manual segmentation or automated segmentation errors.

As future work, cross validation could be used to demonstrate how well the method would generalise to an independent dataset. The work in this chapter should be considered a proof of concept for using Laplacian eigenmap embeddings as a predictor for accepted osteoarthritis disease indicators. It would be of greater use to predict osteoarthritis disease status as opposed to cartilage morphology which are surrogate measures of the disease. The next chapter explores the classification of osteoarthritis diagnostic scores using structural measures of OA severity with fully cross validated results.

Chapter 4

Osteoarthritis Diagnosis on Low Dimensional Embeddings

Work in this chapter has, in part, been presented in:

C. R. Donoghue, A. Rao, A. M. J. Bull, and D. Rueckert. Learning Osteoarthritis Imaging Biomarkers Using Laplacian Eigenmap Embeddings with Data from the OAI. *IEEE International Symposium on Biomedical Imaging*, 2014.

(prize won : young investigators award) C. Donoghue, A. Rao, A. M. J. Bull, and D. Rueckert. Automatically Generated Novel Diagnostic Imaging Biomarkers with Data from the OAI. *4th Imaging Workshop for Osteoarthritis*, 2011.

Abstract

We propose a data-driven approach to learn diagnostic biomarkers of osteoarthritis (OA) status using multi-region and multi-sequence data for a large population study. This study looks to manifold learning as an automatic approach to harness a large subset of the plethora of data provided by the Osteoarthritis Initiative (OAI). We explore three methods of combining multi-region and multi-sequence data to derive a consolidated score or biomarker for OA diagnosis. We show that combining each of the embeddings of multi-region and multi-sequence data to create an ensemble of manifolds is superior to creating an embedding of a single large region of interest. The efficacy of the novel biomarkers presented in this chapter is tested using Linear Discriminant Analysis (LDA), which linearly projects the diagnostic biomarkers onto a discriminant hyperplane. The area under the receiver-operator curve (AUC) for strongest early diagnostic biomarker of 0.904 (95% confidence interval 0.887-0.920) if the population is separated by K&L<2. The results demonstrate that these techniques improve upon results reported in the literature. This improvement on previous works opens the door to a single unified imaging biomarker of osteoarthritis.

4.1 Introduction

4.1.1 Discovery of Novel Biomarkers For OA

The method presented in this chapter is scalable and results are demonstrated using a very large subset of 1131 knee scans from the OAI dataset. It is automatic and does not rely on an accurate segmentation algorithm or a strict manual segmentation protocol to achieve accurate results. The features used for the diagnostic biomarkers are learnt using machine learning algorithms and not designed by humans. It combines information from multiple regions and multiple MR sequences to derive the global diagnosis of the knee. Additionally the results presented show an improvement compared to those previously presented in the literature.

Methods to compute OA biomarkers in the literature typically require a precomputed segmentation. Features are computed from the segmentation, these features are hand-crafted measures

Feature type	Citation
volume	Wluka et al. (2004); Eckstein et al. (2006b); Dam et al. (2007)
thickness	Williams et al. (2003); Eckstein et al. (2006b); Dam et al. (2007)
surface area	Hohe et al. (2002); Dam et al. (2007)
homogeneity	Qazi et al. (2007)
curvature	Hohe et al. (2002); Dam et al. (2007); Folkesson et al. (2008)
congruity	Dam et al. (2009)
roughness	Dam et al. (2009)
a linear combination (LDA) of those listed above	Dam et al. (2009)

Table 4.1: Previous contributions for developing OA biomarkers, they are all measures of cartilage morphology computed from segmentations. These contributions are hand-crafted and based upon the communities understanding of OA.

of cartilage morphology, which are selected based upon the communities understanding of OA. These features are detailed in table 4.1. Schneider et al. (2012) showed that cartilage morphology metrics including volume and mean thickness computed by independent segmentation teams differ significantly. These segmentation teams used manual, semi-automated and automated approaches. It suggests there are some limitations with creating a biomarker which requires segmentation approaches.

There are further problems associated with segmentations, this is discussed in the background section 2.4.8 but summarised here. Manual methods require significant human interaction which is both financially and time costly, potentially reducing the scale of a study. Scalable methodologies appropriate for large cohort OA studies are valuable for both drug development and understanding the manifestation of pathology over a large population. Whilst automated segmentation methods do not suffer from these constraints, low sensitivity is currently a problem amongst state of the art algorithms. The reported segmentation results do not generally assess subregional accuracy of the methods. The algorithms are mostly tested on knees which are not severely diseased and the number of subjects for evaluation have previously been small. Additionally many recent automated segmentation studies do not calculate biomarkers based on their segmentation results, which means we are unable to compare our results. Since the method presented here does not rely on an accurate segmentation it is not limited by these

shortcomings.

This chapter provides methodology for combining both multi-regional and multi-sequence data. Previously state of the art methods only use results from single MRI sequences for OA studies. Eckstein et al. (2006b) shows that cartilage morphology measures including thickness and volume are correlated in sequences Cor T1 3D FLASH and Sag 3D DESS from the OAI. Schneider et al. (2012) also compares thickness and volume metrics in sequences Cor T1 3D FLASH and Sag 3D DESS from the OAI, they found the sequences had similar precision, were generally equivalent, and may be combined for cross-sectional analysis with some adjustments. Whilst biomarkers computed from these sequences are compared, to the best of our knowledge cartilage morphology biomarkers have not previously been combined to provide a multi-sequence approach, instead the most appropriate MR sequence is selected. In this chapter we try to combine biomarkers from Sag 3D DESS and Cor T1 3D FLASH sequences to create a unified multi-sequence biomarker.

The methods presented in the literature commonly use diagnostic biomarkers or morphological measures derived from large regions of interest which often encompass the entire joint. In some cases sub-regional analysis of cartilage morphology is computed (Wirth and Eckstein (2008); Wirth et al. (2009)). However, these sub-regional metrics are not combined into a unifying biomarker. An approach introduced by Buck et al. (2009, 2011) differs from this, they propose ranking change in thickness of the articular cartilage for subregional compartments by magnitude and direction. This enables comparison of regions with greatest change regardless of anatomical location and is called ordered values. This ordered value approach was found to be more sensitive than comparing thickness for corresponding regions across subjects. Despite this ordered values still results in multiple diagnosis per subject, rather than a unified score of disease severity for the joint. We propose a method which uses regional information to identify images which exhibit similar pathology across each region independently. We then combine all of this regional data to identify images which are globally similar. The algorithm presented combines information from multiple ROI and MR sequences as opposed to selecting the most informative ROI or sequence.

Manifold Learning

Manifold learning is a non-linear technique that can be used to characterise low dimensional datasets which lie within a high dimensional space.

Manifold learning has been applied widely to many machine learning problems including face recognition (Wu et al. (2004)), handwriting recognition (Belkin and Niyogi (2002)) and speech signal analysis (Belkin and Niyogi (2003)). In medical imaging it has been used widely in brain and cardiac imaging (Gerber et al. (2009)), including for atlas-based segmentation (Wolz et al. (2010)). Laplacian eigenmaps and other manifold techniques have also been explored for patient position detection in MRI (Wachinger et al. (2010)) and to characterise neonatal brain development (Aljabar et al. (2011)). The methodology proposed here specifically uses Laplacian eigenmap embeddings.

In this work we use the Laplacian eigenmap algorithm, a non-linear dimensionality reduction tool which preserves local information within the data. Two images which are similar to each other are close in the high dimensional feature space. Laplacian eigenmaps preserves close spatial proximity between images which are similar to each other in the manifold embedding. Laplacian eigenmaps are relatively insensitive to outliers and noise (Belkin and Niyogi (2003)).

Overview of Proposed Framework

The principle goal of this chapter is to describe the population of knee MRIs present in the OAI dataset as a low dimensional manifold embedding, from which diagnostic biomarkers can be discovered and extracted. The Osteoarthritis Initiative (OAI) released a plethora of images into the public domain which are further supplemented with detailed clinical data. To utilise this dataset to its full potential it is important to consider methods which are both automatic and scalable. Our proposed framework can be described as follows.

Initially all knee MR images are aligned using non-rigid registration (Rueckert et al. (1999)), such that an approximate region of interest (ROI) can be defined automatically for all images in the dataset. We construct Laplacian eigenmap embeddings (von Luxburg (2007)) for several

ROIs of the articular cartilage, in multiple MR sequences, based on their appearance. Intuitively images that are close to each other in the low dimensional embedding space are considered to be more "similar" than those far away. Having generated multiple low dimensional embeddings for each ROI, using multiple sequences, we are able to derive multiple regional diagnosis of OA status for a given region and sequence. Each embedding provides region-specific and sequence-specific diagnostic information. We present three approaches of combining multi-region and multi-sequence data to provide a global diagnosis for a single knee joint (illustrated in figure 4.1) ;

1. Computing a Laplacian eigenmap embedding of a large region of interest composed of multiple regions.

2. Concatenating the embedding co-ordinates from multiple manifolds generated using different regions of interest and MR sequences followed by a linear dimensionality reduction of the concatenation.

3. Concatenating the embedding co-ordinates from multiple manfiolds generated using different regions of interest and MR sequences followed by a non-linear dimensionality reduction of the concatenation, by a further Laplacian eigenmap embedding.

The experiments presented explore a variety of different pairwise similarity measures, different image sequences, multiple anatomical ROIs and different definitions of the ROI.

The efficacy of the technique is demonstrated through Linear Discriminant analysis (LDA) in the embedding space. The biomarkers are validated by demonstrating strong classification results with K&L grades from the OAI which, for these purposes, we consider to be the ground truth of OA diagnosis. It should be remembered that K&L grades is a crude measure of OA because the grades are very coarse and read from a 2-dimensional image which does not visualise soft tissues. Despite, this we choose to use K&L as our ground truths because of the availability in the OAI dataset and results can be compared to previous work in the literature (Dam et al. (2009)). The results improve on previously presented MRI biomarkers in the literature. Bauer et al. (2006) propose the classification of biomarkers for osteoarthritis into

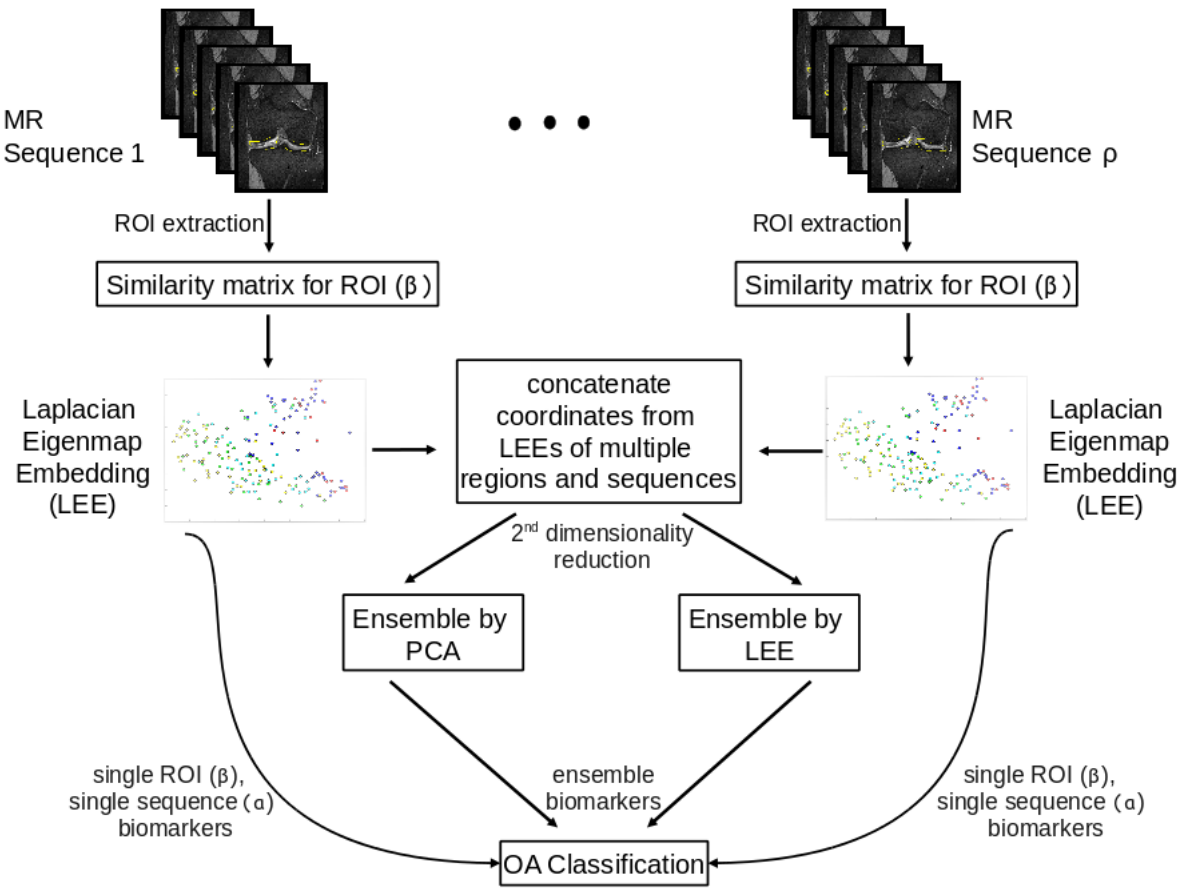


Figure 4.1: Combining multi-region and multi-sequence data workflow

five classes: investigative, diagnostic, prognostic, burden of disease and efficacy of intervention. Some of the proposed standards have been adopted for previous work investigating OA imaging biomarkers (Dam et al. (2009); Eckstein et al. (2009)). The biomarkers we explore are diagnostic biomarkers, which predict the current pathological status of the structure of interest. Bauer et al. (2006) suggests diagnostic biomarkers should be assessed by AUC, sensitivity and specificity.

4.2 Method

4.2.1 Dataset

The experiments presented use 1131 right knees of 445 males and 686 females from the Osteoarthritis Initiative (OAI) public use dataset from groups 1.C.0 and 1.E.0 at baseline. Of this subset 29.7% are healthy (K&L grade 0), 11.3% boarderline/mild OA (K&L grade 1), 23.8% have moderate OA (K&L grade 2), 16.3% have severe OA (K&L grade 3), 2.8% have end point OA (K&L grade 4), 16.1% are unknown because the K&L grades are not as yet available.

For this work, Sag 3D DESS and Cor T1 3D FLASH have been selected as image sequences (described in section 2.4.6). These MR sequences are selected because they have the highest resolutions and because they have been shown to correlate well with morphological measures of the articular cartilage (Chapter 3). resolution, for Sag 3D DESS the in plane resolution is $0.365 \times 0.365\text{mm}^2$ with 0.7mm slice thickness and the in-plane resolution for COR T1 3D FLASH is $0.313 \times 0.313\text{mm}^2$ with 1.5mm slice thickness. Figure 4.2 shows some sample images.

Clinical comparison - Kellgren and Lawrence grade

For comparative purposes the biomarker is assessed through its ability to diagnose OA where the subjects K&L grade is considered to be ground truth. The work presented in this chapter can only be compared to biomarkers which have been assessed for diagnostic potential in the literature (Dam et al. (2009)). It is not possible to compare our biomarkers directly with

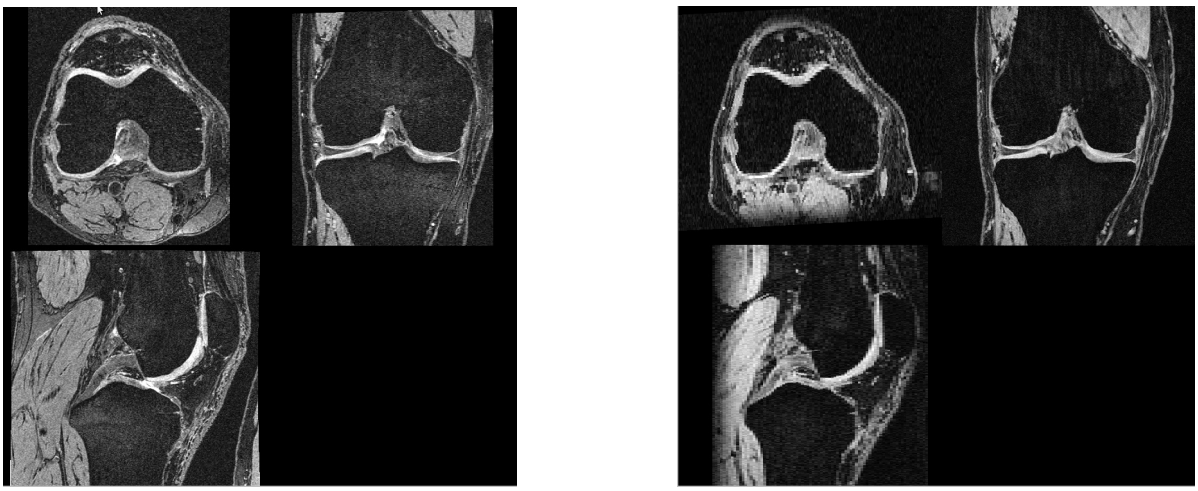


Figure 4.2: Sample of knee MR sequences from the OAI dataset: Sag 3D DESS (left) and COR T1 3D FLASH (right).

results from automated segmentation algorithms where only segmentation quality is assessed (Yin et al. (2010); Fripp et al. (2010)).

The K&L scale is recommended for initial validation of OA biomarkers by Bauer et al. (2006). The OAI provides K&L grades for 949 subjects of the subjects selected in this chapter, whilst this not a full complement of results, the additional subjects embedded in the manifold increase the sampling of the manifold, thus enriching the information set used to describe the entire population.

4.2.2 Registration

A randomly selected sample of n subjects with corresponding MRI are selected from the OAI dataset. The set of images is denoted as $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \ldots, \boldsymbol{I}_n\}$. For each subject, ρ MR images of different sequences are acquired. The set of images for the αth sequence is denoted as the set of images $\mathfrak{I}^\alpha = \{\boldsymbol{I}_1^\alpha, \boldsymbol{I}_2^\alpha, \ldots, \boldsymbol{I}_n^\alpha\}$.

In the same way as described in section 3.2.2 and 3.2.6, a reference image of a specified sequence is selected. For the current study the MR sequence chosen is Sag 3D DESS. This is discussed further in section 4.3. We assume without loss of generality that the reference subject is randomly selected with the following constraints: the subject is from the healthy cohort (with K&L grade 0, i.e. radiographically normal and whom experience no symptoms) and the image

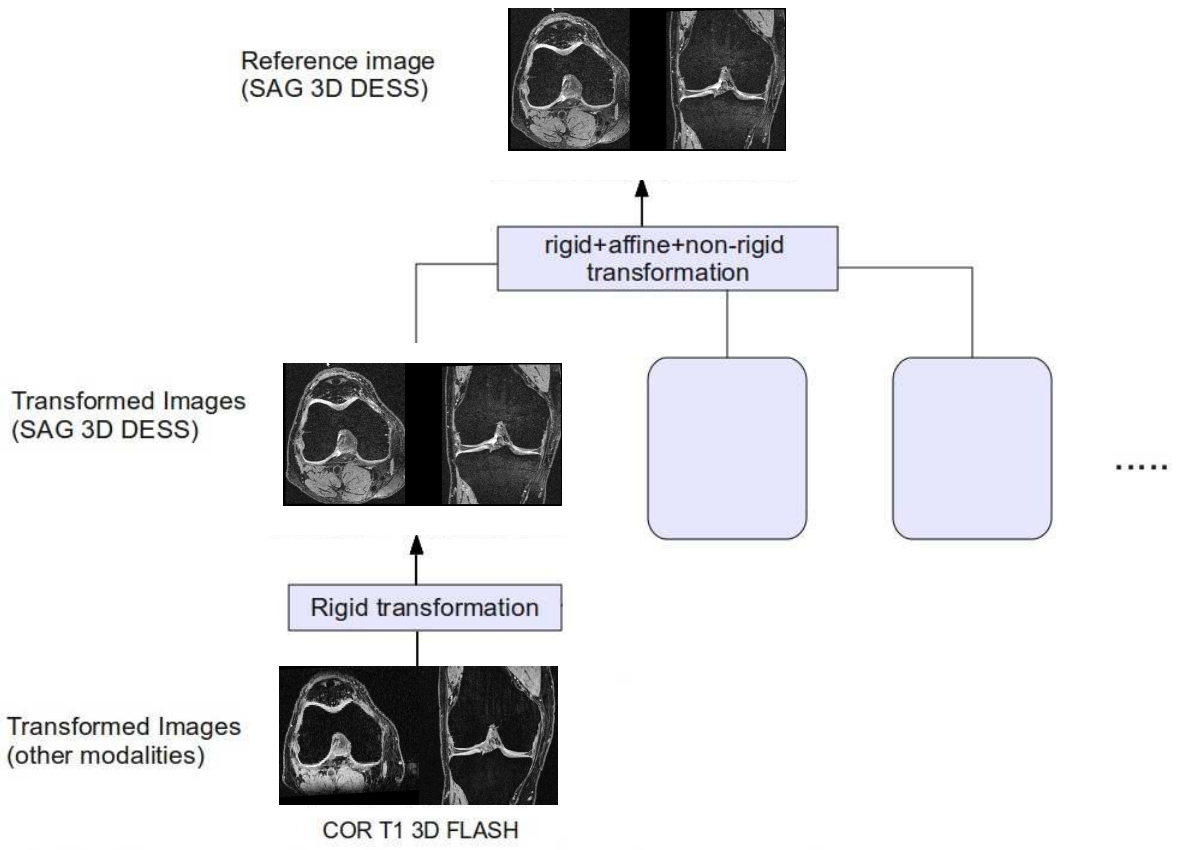


Figure 4.3: Registration pipeline; MRI of the same subject acquired at the same time-point are rigidly registered to the corresponding Sag 3D DESS image. All Sag 3D DESS images are registered non-rigidly to the reference image.

is visually inspected to ensure the image is both of high contrast and appears typical of the population.

As shown in figure 4.3 each image of the same sequence is automatically rigidly and affinely aligned to the reference image to allow for position variation within the scanner and scale discrepancies. These are all then coarsely non-rigidly registered to this reference image by a free-form deformation based on B-splines with a 10mm B-spline control point spacing (Rueckert et al. (1999)). A coarse registration is chosen as opposed to fine registration to ensure that the region of interest contains the tissue structure of interest while preserving some of the relevant shape differences through preventing excessive deformation.

The OAI protocol specifies that each sequence is acquired during the same visit, with strict patient alignment and positioning with the aid of an extremity coil and without removing

the subject from the scanner between sequence acquisitions (section 2.4.5 for details). It is not necessary to compute non-rigid and affine registration for all ρ MR sequences. Instead rigid registrations are computed for each subject between the different sequences as in figure 4.3. Through combining transformations from the precomputed coarse non-rigid registration between subjects (described above) and rigid registration of each subject between sequences, all $n \times \rho$ images can be aligned to the space of the reference image.

4.2.3 Approximate Region of Interest Selection

Atlas $\mathfrak{I}^{\alpha}_{mean}$ is created from the voxel-wise mean for all images which have been transformed to the reference image, for a given MR sequence, α. $\mathfrak{I}^{\alpha}_{mean}$ is high contrast, with well defined boundaries to the articular cartilage. An example is shown in figure 4.4, thus a region growing segmentation is well suited to select the appropriate region of interest in the atlas. Since the approximate segmentation is generated in the space of the reference image, the region of interest can be propagated to all images of all MR sequences.

Since the segmentation generated is in atlas space it can be thought of as a region of interest selection as opposed to a segmentation since only one very rough segmentation is required to represent the region of interest across the population. It is important to stress that regardless of the size of the dataset (in this case $n = 1131$ and $\rho = 2$), only one region growing segmentation is generated semi-automatically. The propagation of the region of interest to all $n \times \rho$ images is performed fully automatically and thus is scalable.

The approximate segmentation was subdivided into 16 distinct anatomical regions using an automated algorithm described by Wirth and Eckstein (2008); Eckstein et al. (2006a). To achieve this I replicated their algorithm in C++, which divides each of the tibial plateau into five regions, a central region which covers 20% of the bone surface and four peripheral regions. The segmentation respects the structure of the articular cartilage through finding the principle orthogonal axis of cartilage distribution across the bone cartilage interface (see figure 4.5 (a)). The femoral ROI are defined as the weight bearing region on the femoral condyle, which is then evenly divided into three strips from medial to lateral (see figure 4.5 (b)). All of the regions

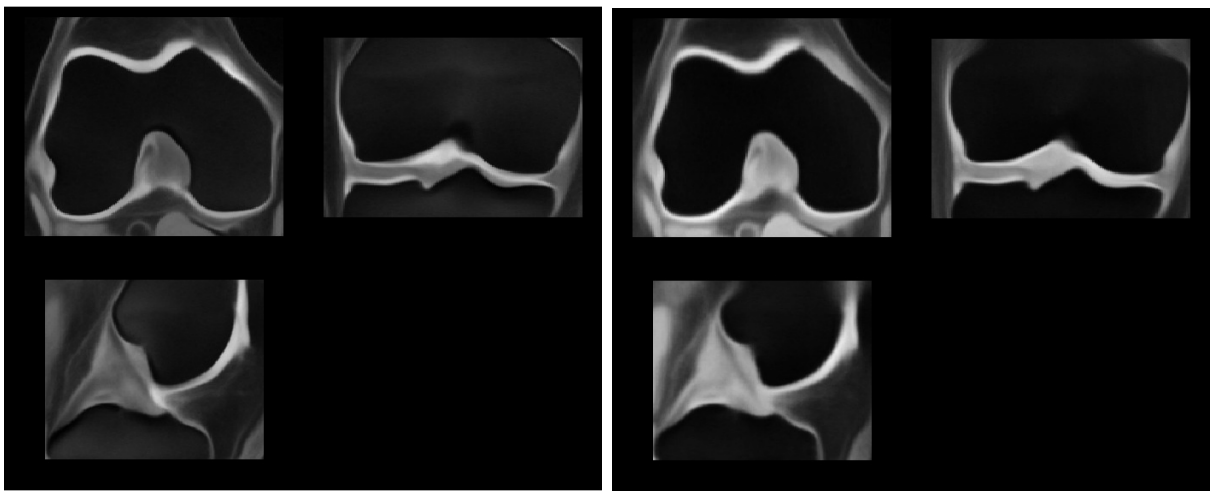


Figure 4.4: Mean image, $\boldsymbol{I}^{\alpha}_{mean}$, where $n = 1131$; where α is sequence Sag 3D DESS on right, Cor T1 FLASH on left

of interest are listed in table 4.2. The set of images for region of interest β and sequence α is expressed as $\mathfrak{J}^{\alpha,\beta} = \{\boldsymbol{I}^{\alpha,\beta}_1, \boldsymbol{I}^{\alpha,\beta}_2, \ldots, \boldsymbol{I}^{\alpha,\beta}_n\}$.

The regions are dilated to compensate for small errors in the coarse non-rigid registration but also to allow for the contribution of contextual information from the area around the articular cartilage. Figure 4.6 shows a dilated region of interest over the articular cartilage within the medial and lateral tibial plateaus and the medial and lateral femoral condyles.

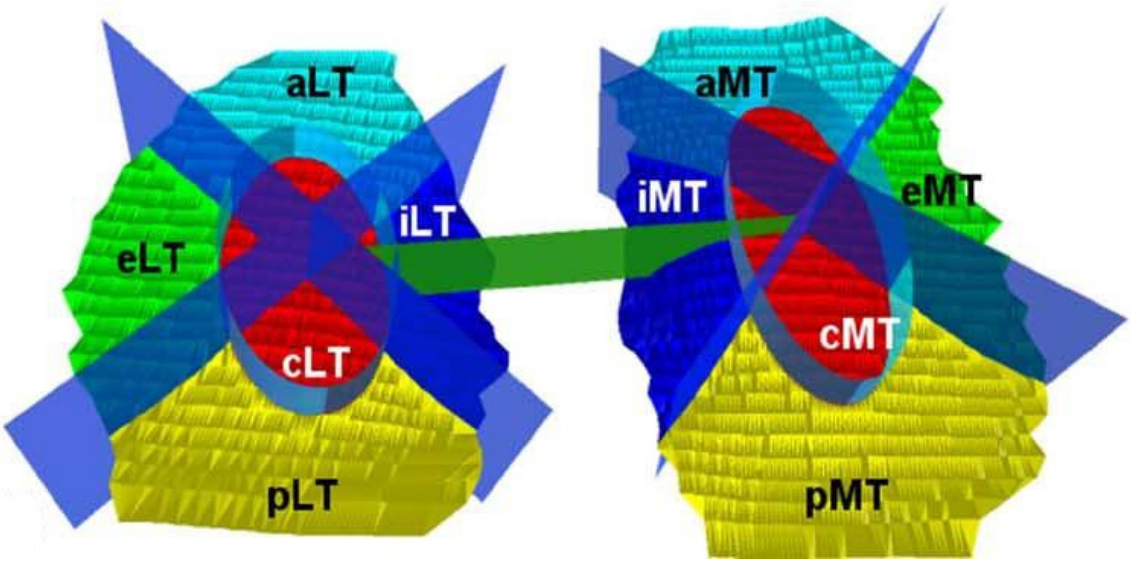
4.2.4 Laplacian Eigenmap Embedding

For all n MR images of a given sequence α and region β, a low-dimensional manifold embedding $\mathcal{M}^{\alpha,\beta}$ is learnt. Images in set $\mathfrak{J}^{\alpha,\beta}$ are used to construct a graph $G^{\alpha,\beta} = (V^{\alpha,\beta}, E^{\alpha,\beta})$ such that vertex $v^{\alpha,\beta}_i$ corresponds to image $\boldsymbol{I}^{\alpha,\beta}_i$. An edge $e^{\alpha,\beta}_{i,j}$ defines the relative pairwise relationships or similarity between the images $\boldsymbol{I}^{\alpha,\beta}_i$ and $\boldsymbol{I}^{\alpha,\beta}_j$. The underlying m-dimensional manifold $\mathcal{M}^{\alpha,\beta}$ (where $m \ll M$) is then approximated or learnt from the sampled imaging data using Laplacian eigenmaps (Belkin and Niyogi (2003)).

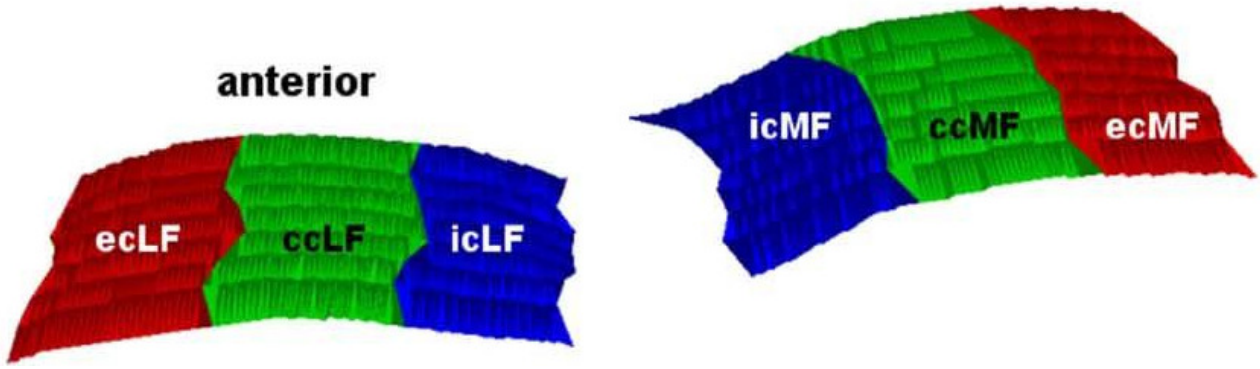
$G^{\alpha,\beta}$ is represented by an $n \times n$ affinity matrix $\boldsymbol{W}^{\alpha,\beta}$, such that $w^{\alpha,\beta}_{i,j}$ encodes the similarity of image $\boldsymbol{I}^{\alpha,\beta}_i$ and image $\boldsymbol{I}^{\alpha,\beta}_j$. The goal is to model the local neighbourhood properties of the data and so $\boldsymbol{W}^{\alpha,\beta}$ is sparsified by only retaining connections to the nearest neighbours. The k-nearest neighbour graph is required to be symmetrical. This is implemented by assuming

Figure 4.5: Regions of interest as defined and illustrated in Wirth and Eckstein (2008).

(a) Regions of interest on the tibial plateau.



(b) Regions of interest on the central femoral condyles.



Region	Abbreviation
external central medial femur	ecMF
central central media femur	ccMF
internal central media femur	icMF
external central lateral femur	ecLF
central central lateral femur	ccLF
internal central lateral femur	icLF
lateral medial tibia	lMT
posterior medial tibia	pMT
central medial tibia	cMT
medial medial tibia	mMT
anterior medial tibia	aMT
lateral lateral tibia	lLT
posterior lateral tibia	pLT
central lateral tibia	cLT
medial lateral tibia	mLT
anterior lateral tibia	aLT

Table 4.2: Regions of interest as defined by Wirth and Eckstein (2008); Eckstein et al. (2006a)

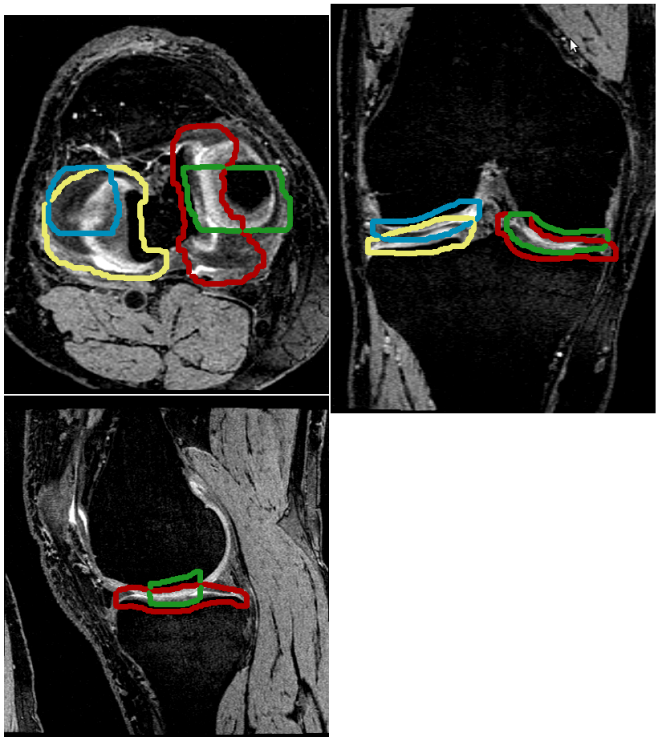


Figure 4.6: Sag 3D DESS MR image with dilated Region of Interest (ROI) for the lateral tibia (yellow), medial tibia (red), central lateral femur (blue) and central medial femur (green).

$\boldsymbol{I}_i^{\alpha,\beta}$ and $\boldsymbol{I}_j^{\alpha,\beta}$ are connected if $\boldsymbol{I}_i^{\alpha,\beta}$ is a neighbour of $\boldsymbol{I}_j^{\alpha,\beta}$ or if $\boldsymbol{I}_j^{\alpha,\beta}$ is a nearest neighbour of $\boldsymbol{I}_i^{\alpha,\beta}$, as described in section 2.2.7.

Laplacian eigenmap embedding (Belkin and Niyogi (2003)) is a spectral method which aims to minimise the following objective function:

$$\sum_{i,j} w_{i,j}^{\alpha,\beta}(\boldsymbol{y}_i^{\alpha,\beta} - \boldsymbol{y}_j^{\alpha,\beta})^2 \tag{4.1}$$

Here $\boldsymbol{y}_i^{\alpha,\beta}$and $\boldsymbol{y}_j^{\alpha,\beta}$ are points in the new embedding. The objective function essentially minimises the distances between pairs of coordinates in the embedded space, with greater importance assigned to images with a higher weight $w_{i,j}^{\alpha,\beta}$. A constraint is imposed to prevent the trivial solution where $\boldsymbol{y}_i^{\alpha,\beta} = 0$. More details with regards to both the theoretical and implementation details can be found in section 2.2.7.

The new coordinates of the images in the manifold embedding $\boldsymbol{y}_1^{\alpha,\beta}, \boldsymbol{y}_2^{\alpha,\beta}, \ldots, \boldsymbol{y}_n^{\alpha,\beta} \in \mathbb{R}^m$ are used as features to extract novel biomarkers for diagnosis of structural pathology. Images are embedded in a lower dimensional manifold space where closer images are more "similar" to each other.

The k-nearest neighbour graph is an important component of the algorithm. For this a value of k needs to be selected. It is also important to discover the intrinsic dimensionality of the embedding space m. There currently exists no theoretical basis for determining the intrinsic dimensionality of the manifold learnt by the Laplacian eigenmap algorithm. Therefore, we choose to determine this parameter empirically. This is discussed further in section 4.3.1.

4.2.5 Combining Multi-Region and Multi-Sequence Data

There has been evidence that osteoarthritis affects the articular cartilage non-uniformly where both thickening and thinning is experienced. Buck et al. (2010) observed that medial femorotibial cartilage thickening was observed as frequently as cartilage thinning in K&G 2 knees. Due to

the non-uniformity of the articular cartilage it is important to identify regional variations. Additionally, several MR sequences have been suggested as appropriate for studying osteoarthritis of the articular cartilage, yet in each study a single sequence is typically used. In this part of the method we wish to address the utilisation of both multi-sequence and multi-regional data.

Embedding co-ordinates are computed for 16 regions of the articular cartilage and for two different MRI sequences selected from the OAI. Therefore there are 32 different embeddings which could each be used as feature vectors to learn an osteoarthritis diagnostic biomarker. The next section explores three different ways of combining the embedding co-ordinates to provide a unified biomarker for automated diagnostic prediction of OA;

1. composites of Regions of Interest (ROI),
2. concatenation of biomarkers with PCA dimensionality reductions and
3. concatenation of biomarkers with Laplacian eigenmap dimensionality reductions

Composites of Regions of Interest

We propose creating several composites of regions of interest, which are each constructed of several distinct regions of interest. These are grouped such that they are anatomically relevant, such as "all regions of the articular cartilage" or "all regions of cartilage on the tibial surface" or "medial side of the joint", etc. The full list can be found in table 4.3. A single embedding is created for each region in that same way as described in section 4.2.4. There are a further 9 regions defined through the composites of regions. The total number of manifold embedding for regions of interest and their composites is 25. The manifold embeddings created from regions are also used as components in the ensembles described in the next two sections.

Concatenation of Biomarkers

For R anatomical regions of interest and ρ MR sequences, there are $R \times \rho$ sets of images $\mathfrak{I}^{\alpha,\beta}$ for which we have described how to compute a set of low-dimensional manifolds $\{\mathcal{M}^{\alpha,\beta} \mid 1 \leq \alpha \leq$

Composite ROI	Composition	Abbrev.
central medial femur	ecMF + ccMF +icMF	cMF
central lateral femur	ecLF + ccLF +icLF	cLF
central femur	cLF + cMF	F
medial tibia	lMT + pMT + cMT + mMT + aMT	MT
lateral tibia	lLT + pLT + cLT + mLT + aLT	LT
tibia	LT + MT	T
lateral side	LT + LF	L
medial side	MT + MF	M
all regions	T + F (or L + M)	All

Table 4.3: Composite regions of interest

ρ, $1 \leq \beta \leq R\}$. Each image $\boldsymbol{I}_i^{\alpha,\beta}$, is represented within every manifold embedding $\mathcal{M}^{\alpha,\beta}$ as the embedding coordinate $\boldsymbol{y}_i^{\alpha,\beta}$, which is a feature vector of length m, the intrinsic dimensionality of the embedding. The feature vector $\boldsymbol{y}_i^{\alpha,\beta}$ is scaled such that the first eigenvector (embedding dimension) is centred at zero and of unit variance. The new feature vectors for each subject i are then concatenated such that $\left[\boldsymbol{y}_i^{1,1}, ..., \boldsymbol{y}_i^{\rho,R}\right]$ becomes the new feature vector. The concatenated feature vector is not used directly as a biomarker because it is in a very high dimensional space $(n \times \rho \times R)$, so dimensionality reduction is required so that dimensionality is sufficiently low to apply a classification algorithm successfully.

Ensemble by Principle Component Analysis (PCA) The concatenated feature vectors are reduced in dimensionality linearly using PCA.

Ensemble by Laplacian Eigenmaps Embeddings (LEE) The pairwise distances between subjects can be computed via the L_2 norm of concatenated feature vectors. These distances are then converted from a dissimilarity measure to a similarity measure to find an affinity matrix. Finally we impose another Laplacian eigenmap step to reduce the dimensionality and find a generalised manifold $\boldsymbol{\mathcal{M}}$ corresponding to R regions of interest and ρ sequences. This approach has been influenced by Aljabar et al. (2011).

4.3 Experiments

Data Selection

In the previous chapter five manifolds were constructed for five MR sequences provided by the OAI and multiple linear regression was used to observed how well the manifold coordinates predicted cartilage morphology. The results suggested that the 3D MR sequences were better predictors of cartilage morphology than the other 2D sequences. In addition Peterfy et al. (2008) find that for quantitative cartilage morphometry, Sag 3D DESS provides the best universal cartilage discrimination. Therefore Sag 3D DESS and Cor T1 3D FLASH are used in the following experiments, with Sag 3D DESS chosen as the reference sequence.

Region of Interest Dilation

We also experimented with different dilations of the region of interests. We used both a 5 voxel and 2 voxel dilation. Generally the diagnostic power of the manifolds generated using a 5 voxel region of interest performed better. It is speculated that this could be due to ensuring that all of the articular cartilage is contained within the ROI or because the larger ROI provides extra contextual information with respect to the subject's anatomy for the learning algorithm

Similarity Metrics

Different similarity metrics were used as a measure of similarity between two images to describe neighbourhood information including normalised mutual information (NMI), normalised cross correlation (NCC) and sum of squared differences (SSD), described in section 2.3.3.

From preliminary experiments it was found that sum of squared differences performed very poorly in tests to separate candidates with different OA diagnosis. However, it was not conclusive whether Normalised Mutual Information or Cross Correlation were better able to separate subjects based on their OA diagnosis. Normalised Cross Correlation appears to perform marginally better and as a result this measure is adopted as the similarity metric going forward.

4.3.1 LDA Classification Experiments

Linear discriminant analysis (LDA) is used here to assess the quality of the biomarkers produced by the data representations. LDA is a simple and robust classifier which finds a discriminatory hyperplane that minimises the within class distances and the between class distances. More details about LDA can be found in section 2.2.2.

Whilst more sophisticated non-linear classification methods could be employed here, this chapter aims to show the strength of a novel data representation method in the field. If the method presented here outperforms previous results reported in the literature with a simple classification method, then there is reduced ambiguity about the reasons for the algorithms improved performance, and the successes can be attributed to the data representation.

LDA classifications have been computed for all disease severities on the K&L scale. Whilst the main focus of the chapter is for early diagnostics as this is beneficial for early intervention and for comparison with previous techniques in the literature, it is worth displaying the capability of this method across the disease spectrum.

Statistics for Assessing LDA Classification Accuracy

For all experimental results from the LDA classifier a 20-fold cross validation is computed for all results and the mean results from all folds are reported. The area under the receiver operator curve (AUC) and a 95% confidence interval is reported to provide an overall representation of the performance of the classifier. For those with high AUC, classification rates including accuracy, sensitivity and specificity are recorded. This is the recognised approach for assessing the performance of a diagnostic biomarker (Bauer et al. (2006)) and also allows comparison with other methods that have been previously published in the literature (Dam et al. (2009)).

Embedding Parameter Selection

It is important to select two parameters for Laplacian eigenmap embeddings, namely the neighbourhood for the affinity matrix k and the intrinsic dimensionality of the manifold m. We are interested in finding an accurate early diagnostic imaging biomarker to aid early detection for clinical interventions. We define this to be classifying subjects whose K&L grade is <2. To ensure a good overall classifier performance, we are interested in optimising the area under the ROC curve (AUC).

These parameters are selected empirically using cross validation and a 2D grid search. The data is divided into two sets a validation set and a test set, 20% and 80% of the dataset respectively. The parameters for the Laplacian eigenmap embedding are determined from the validation dataset. The test set is reserved for further experiments, using the parameters selected from the validation dataset.

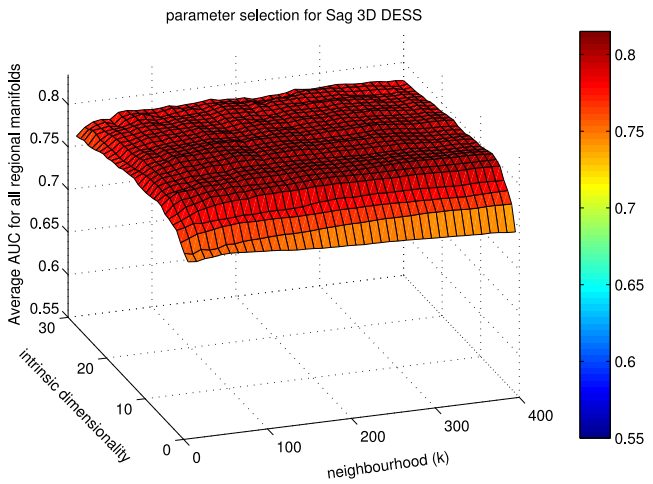
The parameter space is discretised for a bounded range, for which optimal parameters are selected using an exhaustive search within this 2D grid. The search aims to find the vicinity of the maximum average AUC across all regional manifolds. The selected parameters are assumed to be sufficiently robust for each of the Laplacian eigenmap embeddings. Since each element on the grid is computed independently, this process can be easily parallelised for increased performance. Figure 4.7 is a visualisation of the discretised parameter space grid. The parameter landscape is fairly stable when the connectivity neighbourhood is greater than 100 and the intrinsic dimensionality is greater than 5. For all future experiments, the connectivity neighbourhood of the graph was set to 120 and the intrinsic dimensionality to 20.

4.3.2 Results

Mono-Sequence Experiment

Table 4.4 shows the area under the receiver operator curve (AUC) results with a 95% confidence interval for all the classifications of OA status. All the results displayed in this table are

(a)



(b)

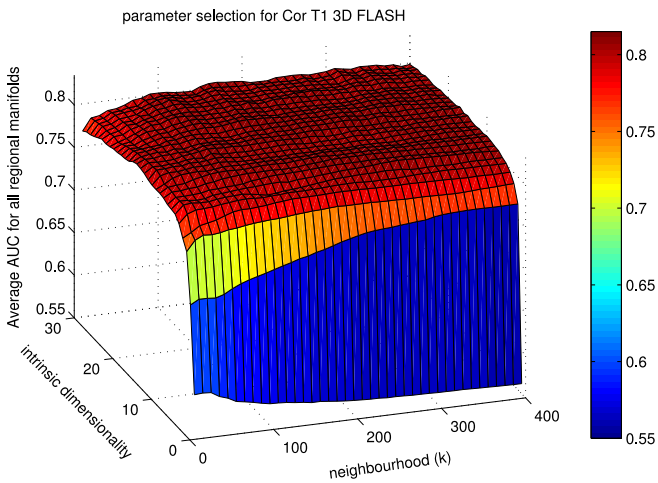


Figure 4.7: Grid search parameter selection visualisation for MR sequences; 4.6 (a) SAG 3D DESS and 4.6 (a) Cor T1 3D FLASH, to choose suitable parameters for the number of nearest neighbours in the graph required for nearest neighbourhood analysis and the intrinsic dimensionality of the manifold embedding. The metric for success is the AUC over the ROC which measures the overall success of the classifier. The measure is computed for each regional manifold and the mean of this score is computed.

computed for just one MR sequence, $\rho = 1$. Results for each sequence are displayed in two separate columns.

Results presented in the bottom half of table 4.4 contains the top five embeddings which are computed from using a single Laplacian eigenmap embedding. * denotes those regions that are defined as a composites of regions of interest as described in section 4.2.5. The embedding rankings are defined based on their performance for early OA diagnosis for both MR sequences.

Results presented in the top half of table 4.4 refer to the different combination methods in this chapter;

- *Composite of all regions * :* composite of regions as described in 4.2.5 for the entire articular cartilage.
- *Ensemble by PCA :* combination of manifold embeddings, reduced in dimensionality linearly by PCA.
- *Ensemble by LEE $\mathcal{M}$:* combination of manifold embeddings, reduced in dimensionality non-linearly by Laplacian eigenmap embeddings.

	AUC (95% CI)							
	Sag 3D DESS				Cor T1 3D FLASH			
Region of Interest	K&L<1	K&L<2	K&L<3	K&L<4	K&L<1	K&L<2	K&L<3	K&L<4
Ensemble by LEE $\mathcal{M}$	0.866 (0.848,0.885)	0.886 (0.864,0.905)	0.950 (0.937,0.962)	0.887 (0.880,0.977)	0.832 (0.802,0.862)	0.840 (0.813,0.867)	0.920 (0.900,0.940)	0.905 (0.832,0.978)
Ensemble by PCA	0.876 (0.857,0.894)	0.904 (0.887,0.920)	0.945 (0.931,0.960)	0.896 (0.803,0.990)	0.848 (0.821,0.876)	0.862 (0.837,0.887)	0.927 (0.911,0.944)	0.910 (0.837,0.983)
Composite of all regions *	0.858 (0.835,0.880)	0.882 (0.862,0.902)	0.929 (0.913,0.946)	0.873 (0.784,0.963)	0.833 (0.805,0.861)	0.854 (0.830,0.878)	0.920 (0.905,0.936)	0.891 (0.790,0.991)
Tibia *	0.860 (0.834,0.885)	0.880 (0.859,0.902)	0.922 (0.902,0.942)	0.877 (0.791,0.963)	0.814 (0.780,0.848)	0.832 (0.803,0.861)	0.901 (0.882,0.920)	0.900 (0.830,0.971)
Medial *	0.855 (0.831,0.879)	0.868 (0.849,0.888)	0.907 (0.883,0.931)	0.863 (0.795,0.931)	0.823 (0.791,0.855)	0.834 (0.817,0.852)	0.884 (0.856,0.912)	0.756 (0.619,0.893)
Central Femur *	0.838 (0.816,0.860)	0.856 (0.829,0.883)	0.938 (0.919,0.957)	0.849 (0.764,0.935)	0.827 (0.800,0.855)	0.845 (0.822,0.868)	0.924 (0.905,0.943)	0.851 (0.735,0.967)
Medial Tibia *	0.855 (0.832,0.879)	0.868 (0.849,0.887)	0.902 (0.877,0.928)	0.873 (0.800,0.945)	0.816 (0.788,0.845)	0.819 (0.793,0.845)	0.872 (0.845,0.899)	0.781 (0.667,0.896)
posterior Medial Tibia (pMT)	0.842 (0.808,0.876)	0.836 (0.813,0.860)	0.852 (0.817,0.887)	0.833 (0.770,0.896)	0.825 (0.788,0.862)	0.805 (0.780,0.830)	0.836 (0.808,0.865)	0.910 (0.837,0.983)

Table 4.4: Results for Sag 3D DESS and Cor T1 3D FLASH sequences: area under receiver operator curve from LDA, separating data between K&L grades listed in the separated columns. The top half of the table contains the three methods presented for combining the regions of interest for the specific sequences. The bottom half of the table contains an ordered list of the top five regions for classifying early OA for both MR sequences. * indicates those ROIs which are composites of other region of interest

	AUC (95% CI)			
	Ensemble of Sag 3D DESS and Cor 3D T1 FLASH			
Region of Interest	K&L<1	K&L<2	K&L<3	K&L<4
Ensemble by LEE $\mathcal{M}$	0.830 (0.800-0.861)	0.839 (0.811-0.867)	0.920 (0.900-0.940)	0.913 (0.836-0.989)
Ensemble by PCA	0.848 (0.821-0.876)	0.862 (0.837-0.887)	0.927 (0.911-0.944)	0.910 (0.837-0.983)

Table 4.5: Results for ensemble methods for all regions and both sequences. Results reported using area under receiver operator curve from LDA, separating data between K&L grades listed in the separated columns.

Multi-Sequence Experiment

Table 4.5 shows the results for the *ensemble by Laplacian eigenmap embeddings* and *ensemble by PCA*, where all MR sequences and regions are concatenated into a single feature vector.

4.3.3 Analysis

It appears that in general the multi-sequence, multi-region ensemble methods perform worse than the mono-sequence, multi-region ensemble method for AUC. However, for late stage diagnosis (K &L < 4) performance is slightly improved.

As can be seen from table 4.4, the *ensemble by Laplacian eigenmap embedding* and the *ensemble by PCA* methods outperform all other manifolds. This demonstrates the advantage of combining the regions through the concatenation of features from multiple manifold embeddings, followed by a further dimensionality reduction step.

Of the mono-sequence experiments, those which use the Sag 3D DESS sequence outperform the results for Cor 3D T1 Flash experiments. This could be indicate that the Sag 3D DESS sequence contains more relevant information in the intensity patterns in the articular cartilage. However, the enhanced performance could be attributed to Sag 3D DESS being the selected reference MR sequence. This could be investigated further as future work. Since we note

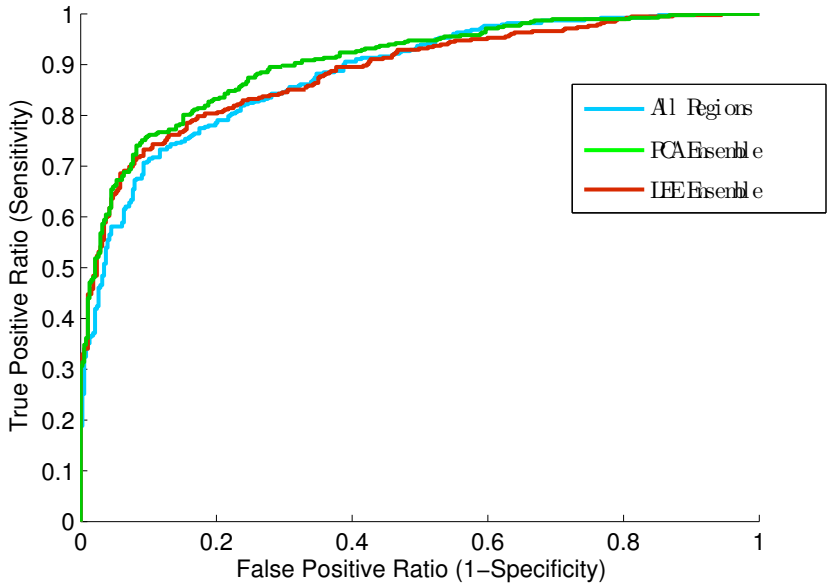


Figure 4.8: ROC curve for LDA classification of *Ensemble by Laplacian eigenmap embedding* $\mathcal{M}$ (LEE), *Ensemble by PCA* (PCA) and *Composite of all regions of interest* for the sequence Sag 3D DESS (since it outperformed sequence Cor 3D T1 Flash for early diagnosis).

higher AUC for experiments with the sequence Sag 3D DESS, future analysis will concentrate on mono-sequence, multi-region experiments with this MR sequence.

Early OA diagnosis (Diagnostic Imaging Biomarkers) The strongest biomarker in the literature to date was presented by Dam et al. (2009). Their cartilage longevity marker, which is a linear combination of cartilage morphology measures acquired from MRI, achieved an AUC of 0.84 (0.77 to 0.92), when assessed for diagnosing K&L<2 . The strongest marker presented here was through the *Ensemble by PCA* using regions in the ensemble, it achieved an AUC of 0.90 (0.89 to 0.92). Not only is the AUC improved by 6%, the confidence interval is narrower which suggests the method presented here is more robust.

Figure 4.8 visualises the ROC curve for the three methods of combining the data from multiple regions of interest using Sag 3D DESS. Here the diagnostic threshold is set to K&L<2. It suggests that *ensemble by PCA* slightly outperforms *composite of all regions* and *Ensemble by Laplacian eigenmap embedding*.

Late stage OA diagnosis (Burden of Disease Imaging Biomarkers) Whilst not at the heart of the work in this thesis, the reader's attention should be drawn to the results for late stage OA classification, achieving up to AUC= 95%. For previous studies which required an accurate segmentation to extract morphometric biomarkers, segmentations are notoriously difficult for subjects which look different from the healthy population and thus accurate morphometric measurements might not be feasible. Since the solution proposed here profits from dissimilarities between images from different classes, a manifold learning based biomarker is well suited to such a problem, as accurate segmentations are not required and the affinity matrix is generated through measurement of similarity of image appearance.

Classification Accuracy Results

Figure 4.9 shows a visualisation of the classification rate, sensitivity and specificity achieved by the methods for combining regions with the Sag 3D DESS sequence.

4.3.4 Visualisation of Laplacian Eigenmap Embedding

It is possible to visualise two (or up to three) dimensions of the Laplacian eigenmap embeddings. Typically the first two are visualised since these contain the most information regarding the embedding. In a visualisation, each point represents a subject and the clinical data can be represented through the colour of the point. Here the colour corresponds to OA status (K&L grade) of the subject.

Figure 4.10 shows the manifold embeddings for the ensemble of Laplacian eigenmap embeddings. The ensemble for visualisation 4.10 (a) uses only regions in the sequence Sag 3D DESS and the visualisation in 4.10 (b) combines regions from only Cor T1 3D FLASH sequences. The visualisation in 4.10 (c) shows the ensemble is constructed using all regions of interest from both sequences Sag 3D DESS and Cor T1 3D FLASH.

It should be remembered that the intrinsic dimensionality of the embedding (m) is likely to be higher than displayed in the visualisation. Despite the incomplete nature of the visualisation,

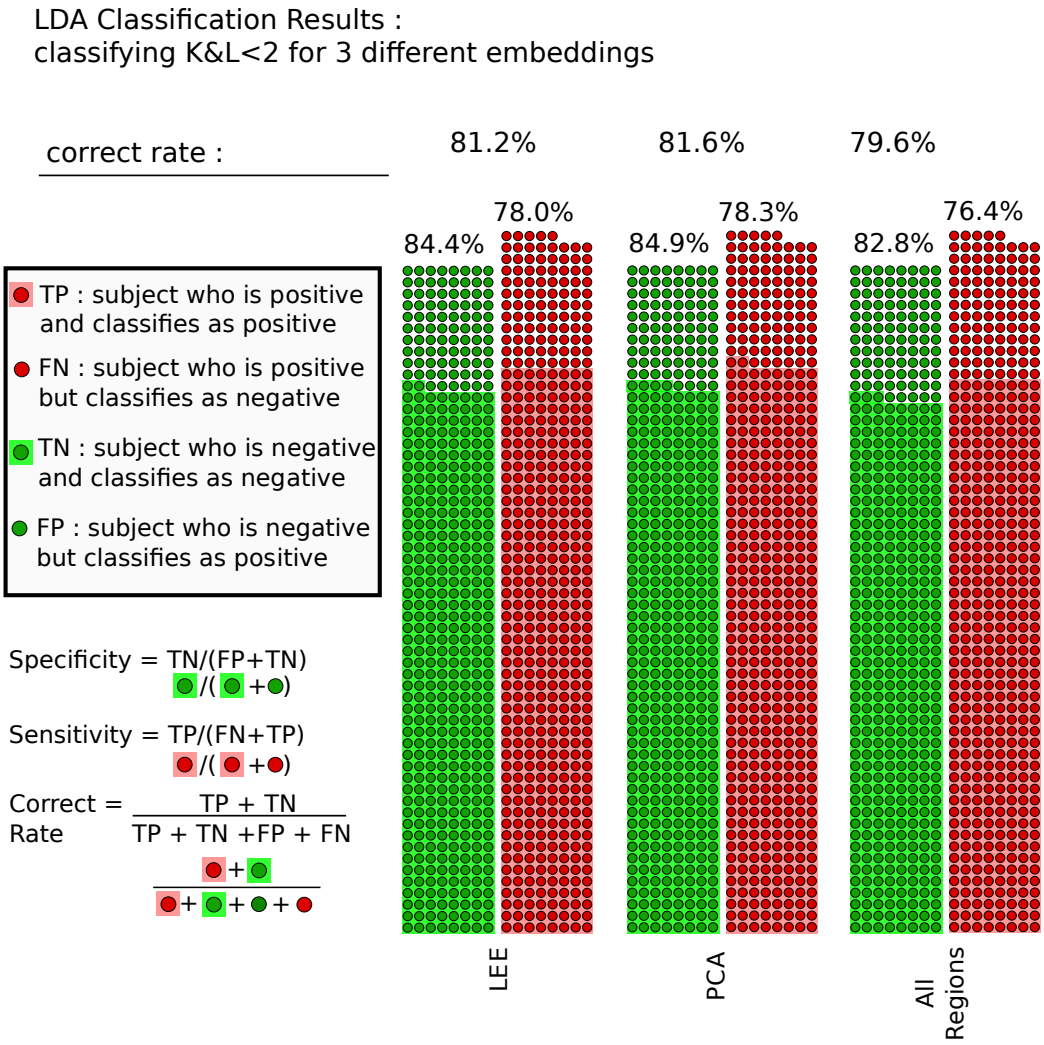


Figure 4.9: Classification results including correct rate, sensitivity and specificity for three different approaches for combining regional for the Sag 3D DESS sequence; *Ensemble by Laplacian eigenmap embedding*$\mathcal{M}$ (LEE), *Ensemble by PCA* (PCA) and *composite of all regions* . Each circle represents a sample, specificity can be read by the proportion of the left (green) bar shaded, sensitivity by the percentage of the right (red) bar shaded and the classification rate is the percentage of both of the shaded sections of the bars.

visualisations can still be used as a powerful tool for understanding the key structure of the data, identifying trends and outliers.

4.4 Conclusion and Discussion

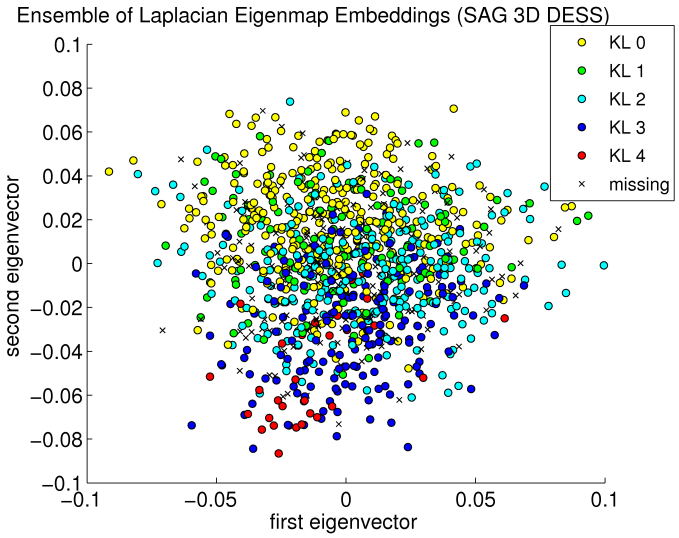
This chapter uses a machine learning algorithm to discover new diagnostic biomarkers directly from image appearance without requiring an accurate segmentation. Previous approaches create handcrafted and designed features which are thought to have a relationship with OA progression, these include thickness, volume and texture measures (Dam et al. (2009); Eckstein et al. (2009); Buck et al. (2009); Qazi et al. (2007); Raynauld et al. (2004)). Previously studies which have observed the articular cartilage over reduced regions of interest generally do not combine data across multiple-regions. Buck et al. (2009) has been successful in isolating the most descriptive region of interest in an ordered values approach, in contrast we combine these measures through machine learning to allow us to exploit the diversity and richness of data available.

The biomarkers proposed in this chapter improve upon previously proposed biomarkers. The current state of the art biomarker based on the articular cartilage in MRI (Dam et al. (2009)) reports an AUC (95% confidence interval) for LDA classification of K&L grade<1 as 0.82 (0.77-0.92) and of K&L grade<2 as 0.84. Our results are for K&L grade<1 0.88 (0.86-0.89), K&L grade<2 0.90 (0.89 to 0.92). The biomarkers presented here perform well and are very competitive for early stage OA diagnosis.

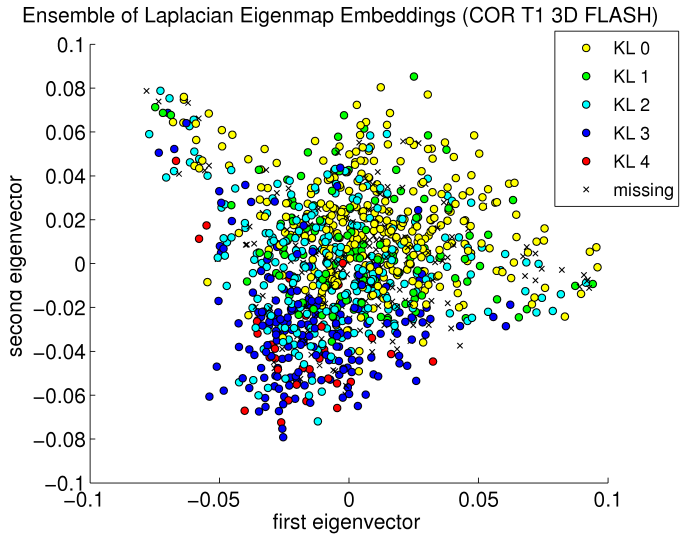
Bauer et al. (2006) discusses biomarkers for late stage OA diagnosis which they define as K&L grade<3. Bauer et al. (2006) describe these as burden of disease imaging biomarkers, as the name suggests the biomarker indicates the severity of OA. Such results for late stage OA are rarely presented in the literature. It could be speculated that biomarkers which rely on accurate segmentations might not perform very well, since inaccurate or noisy measurements are likely in cases where the articular cartilage is badly damaged due to OA. Our results show very good results for late stage OA classification, K&L grade<3 0.95 (0.94-0.96).

Figure 4.10: Visualisation of ensemble of Laplacian eigenmap embeddings, displaying only the first two eigenvectors, based on ascending non-zero eigenvalue order. It should be remembered that this visualisation is only illustrative of the population in two dimensions, whilst classification occurs in a low dimensional manifold space, dimensionality would be greater than two. The visualisation shows the population of knee MRI focusing on the articular cartilage for sequence a) Sag 3D DESS and b) Cor T1 3D FLASH c) both sequences

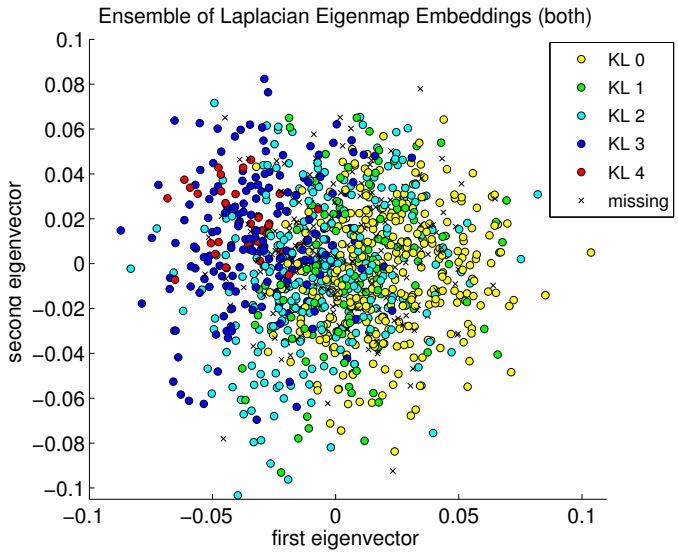
(a)



(b)



(c)



We have found in these experiments that combining multiple regions of interest from the Sag 3D DESS sequence yields higher diagnostic performance than using a multi-sequence biomarker for early OA diagnosis. However, for very late stage OA classification (K&L grade<4), a multi-sequence, multi-regional embedding (the multi-sequence *Ensemble by LEE*) yields the highest AUC when compared with all other manifold embeddings presented.

4.4.1 Contribution Of Work

We have proposed a novel way of combining all regions of interest and MR sequences into a single unified biomarker and we have demonstrated empirically that this unified biomarker has strong diagnostic performance. The automated algorithm presented is demonstrated on a large dataset. With further validation and experimentation the work presented in this chapter could be used to form a clinically applicable diagnostic OA biomarker at all stages of OA for knee MRI.

4.4.2 Limitations

The K&L scale is used as the ground truth data for classification of radiographic OA in this chapter as recommended by Bauer et al. (2006). However, K&L have been shown to have a weak relationship with pain and symptomatic OA, Hannan et al. (2000) showed that radiographic evidence of structural osteoarthritis does not indicate a symptomatic response. The K&L scale is limited as it is a one dimensional biomarker which describes three structural features which can be observed in radiographs of knees and are associated with osteoarthritis (more details in table 2.3). Additionally the K&L scale is difficult to apply due to inexact wording of the descriptors (Schiphof et al. (2008)) which affects reproducibility. K&L grades are derived from radiographs which are unable to detect soft tissue and so the grades focus on bone structures which are apparent in the radiographs. The methodology proposed here focuses on the articular cartilage (and its immediate neighbourhood in the 5 voxel dilation). Radiographs are further limited as a diagnostic tool because of the 2D nature of the images. The methodology proposed

in this chapter which generates biomarkers from MRI can capture subtle differences in cartilage appearance that the methodology for the K&L grades are unable to detect. It is therefore possible that the method presented here could outperform the K&L grade as a diagnostic biomarker. Since we are training the classifier using K&L grades some of the errors presented could be partially attributed to the limitations of the K&L grade.

Other limitations are introduced and addressed in future work (section 4.4.3).

4.4.3 Future Work

A linear classifier was employed to test the efficacy of the biomarkers derived in this chapter. The chapter primarily focuses upon biomarker discover, a linear classifier enables the study to test the expressive power of a biomarker generated by non-linear dimensionality reduction tools. Additionally, the LDA classification experiments in this chapter replicate those performed by Dam et al. (2009), this enables a fair comparison between our study and theirs. A non-linear classifier, such as SVM, might increase classification accuracy, this should be explored.

To address the limitations of the K&L grades as indicators of structural OA severity as observed on radiographs, it would be beneficial to use WORMS (Peterfy et al. (2004)) or BLOKS (Hunter et al. (2007)) as ground truth measures. WORMS and BLOKS are richer semi-quantitative measures of structural OA as observed by radiologists from MRI, this data has recently been released by the OAI.

It could be clinically useful to generate a prognostic biomarker which predicts patient outcome data. Outcome data could be structural such as whether patients need knee replacements or functional such as pain experienced by the patient. The relationship between structure and function as yet has been very challenging to determine.

It was found that occasionally region of interest selections were insufficiently accurate. This was attributed to a poor initial affine registration, which lead to implausible non-rigid registrations. The rate of affine registration failures and methods to compute robust and accurate affine transformations are presented in the subsequent chapters.

Chapter 5

Accurate Global Geodesic Registrations

Work in this chapter has, in part, been presented in:

C.R. Donoghue, A. Rao, A.M.J. Bull, and D. Rueckert. Robust Global Registration through Geodesic Paths on an Empirical Manifold with Knee MRI from the Osteoarthritis Initiative (OAI). *Biomedical Image Registration*, 7359:1-10, 2012.

Abstract

Registration is important for many applications in medical image analysis. Affine registration of knee MR images can suffer failures due to large anatomical, scale and articulated pose variations. Within the OAI dataset we have observed a failure rate of approximately 4.2% for direct affine registrations of knee MRI without manual initialisation. With the increase in large population datasets any manual interventions to aid registration are not feasible and so full automation with high accuracy is of paramount importance. Additionally, computing exhaustive pairwise registrations across the OAI dataset is very computationally expensive.

We present a sparse 'geodesic' registration method that increases the accuracy of pairwise registration and also enables fast online computation of registration. We model the population of unregistered knee MR images as a sparse k-nearest-neighbour graph to capture the empirical low dimensional manifold of the images. Affine registrations are computed in advance for nearest neighbours only. When a pairwise image registration is required the shortest path across the graph is extracted to find a geodesic path on the empirical manifold. The precomputed affine transformations along this path are composed to estimate the transformation. We propose an additional refinement step to boost the registration accuracy.

5.1 Introduction

5.1.1 Motivation

Affine registration is necessary for many applications in medical image analysis. In some cases this is achieved through the affine or rigid registration techniques described in chapter 3 and 4 as well as in the literature. Fripp et al. (2010) employs an affine initialisation prior to using a 3D active shape model for bone segmentation. In work by Carballido-Gamio and Majumdar (2011) an affinely aligned atlas is used to assess local cartilage changes of the patella. Atlases are further used in work by Tamez-Pena et al. (2011) for multi-atlas based segmentation of the articular cartilage, for which all images are affinely registered prior to atlas construction and non-rigid deformation. Despite the prominence of affine registration in the literature for

registration between inter-subject knee MRI, little exploratory work has been undertaken to boost performance for both accuracy of results and speed.

Yin et al. (2010) present an alternative approach, an adaboost sliding window detector to locate the tibia, femur and patella independently as a precursor to segmentation. The authors report that no erroneous detections were observed but do not report precision of the localisation. The method is tested only on 20 subjects and so it is unclear whether the method is truely robust to a large population of subjects. The authors report that the algorithm requires 6 minutes on a IntelCore Duo 2.6GHz with 4GB RAM. A similar but faster method marginal space learning (Zheng et al. (2009)) which subdivides the algorithm to reduce the search space has been demonstrated on cardiac images.

Numerous contributions recognise affine registration to be a prerequisite step for accurate spatial alignment (Carballido-Gamio and Majumdar (2011); Fripp et al. (2010)). Despite this affine registration is commonly assumed to be a solved problem. Affine registration may have appeared to be insignificant in studies where the sample size is approximately 20 subjects (Yin et al. (2010); Fripp et al. (2010)).

With large datasets it is infeasible to expect that registration failures can be corrected manually. The Osteoarthritis Initiative (OAI) provides public access data for large scale studies with about 4796 subjects. Robust automated tools that yield accurate and consistent results are essential for analysis of datasets of this scale and fundamental to aid understanding of pathology and biomarker discovery. Pairwise affine registration between images is highly desirable to remove positional, scale and orientational bias from an experiment but exhaustive pairwise registration is computationally unrealistic with thousands of images. Scalable solutions to address issues of speed and accuracy are of extreme importance.

In a small study with only very few registration failures, it is trivial to manually correct for errors. Based upon our findings of a 4.2% error rate, the impact of affine registration failures on real world large scale projects is likely to become significant. Additionally manually correction means that there is not consistent preprocessing for all subjects.

Figure 5.1: MRI of four subjects right knee, subjects were selected for having the highest and lowest BMIs in the database, the BMI of each are as follows a) 47.7, b) 45.4, c) 16.9, d) 17.6.

figure	a	b	c	d
a		CC : 0.35 NMI : 1.014	CC : 0.22 NMI : 1.008	CC : 0.29 NMI : 1.011
b	CC : 0.35 NMI : 1.014		CC : 0.31 NMI : 1.013	CC : 0.32 NMI : 1.014
c	CC : 0.22 NMI : 1.009	CC : 0.31 NMI : 1.012		CC : 0.63 NMI : 1.042
d	CC : 0.29 NMI : 1.011	CC : 0.32 NMI : 1.014	CC : 0.63 NMI : 1.042	

Table 5.1: Registration quality of pairwise registration between subjects with very high BMI (a and b) and very low BMI (c and d). Registration qualitative success indicated by colour of square; green indicates a successful registration, orange indicates the registration had limited success with the results being ambiguous and red indicates the registration failed. Quantitative success measured by similarity metrics cross correlation and normalised mutual information are also displayed.

5.1.2 Preliminary Experiment

To demonstrate the potential difference in structure of anatomy in different subjects in the OAI dataset we have selected some subjects with vastly different BMI. These are shown in figure 5.1. The maximum BMI in the database is 47.7 and the minimum BMI is 16.9. Since we would expect that these subjects are likely to be dissimilar and therefore yield a poor registration, we have conducted an experiment comparing the similarity of pairs of image pre- and post-registration. The results can be found in table 5.1. If a and c are manually aligned so that the images appear to have a very good overlap the cross correlation metric between the images is poorer than after an automated registration which doesn't appear to be so well aligned (0.24, 0.22 respectively).

5.1.3 Intermediate Template Selection

It is challenging to find anatomically meaningful correspondences between subjects with vastly different anatomy. One challenge is finding correspondences between subjects where some structures are not present in both images. Such differences in anatomy can be explained as either natural healthy variation but more often differences can be incurred due to pathology or developmental changes. Moreover, whilst it may be possible to determine correspondences between

healthy controls on a gross structure level, at higher imaging resolutions such correspondences become ambiguous.

Recently a new class of registration algorithms has been pioneered using intermediate templates to aid accurate registration (Serag et al. (2012); Hamm et al. (2009, 2010); Jia et al. (2010, 2011, 2012)). The methods use intermediate templates to avoid direct registrations between a pair of dissimilar images. A pair of dissimilar images can be registered via a series of intermediate templates. Each intermediate template is similar to its neighbouring images in the series, a registration between similar images is more likely to find accurate anatomical correspondences.

Serag et al. (2012) proposes registering neonatal brain MRI by selecting intermediate templates based on their gestational age. This ensures that the deformations reflect the typical developmental changes of subjects. Hamm et al. (2009) propose registering subjects using an empirical manifold which is represented as a graph. Upon requiring a registration a shortest geodesic path is computed across the empirical manifold. Jia et al. (2011) uses intermediate template selection to initialise registrations. This method differs from other methods because each registration in the tree is precomputed. Further related work has been employed to model an empirical manifold of non-rigid transformations of medical images (Hamm et al. (2009, 2010); Jia et al. (2010, 2012)). Prior to this large deformation registration methods have been proposed to preserve the topology of non-rigid deformations Christensen et al. (1996) but these do not model the biological anatomical variations of the population of images.

In addition to applying the ideas of intermediate templates to affine registration of the knee, we also look at refining the results of such a registration. Serag et al. (2012) propose refining the final transformation by using the transformation from the geodesic shortest path in the graph as an initialisation for a further registration. Section 6.2 addresses refinements of registration in more detail.

5.2 Geodesic Registration Framework

5.2.1 Overview

Upon affine registration of pairs of images without manual initialisation in a large dataset, a high failure rate has been observed (4.2%), where failure is defined in section 5.3.1. This chapter reports a registration framework which is robust to challenging pairwise registrations. We firstly propose a framework to register any pair of images in a dataset and then a refinement step is proposed in order to increase accuracy.

In the first section of this chapter we will address improving the accuracy of registration using a variation of the data driven, intermediate template selection approach. We propose modelling the empirical manifold of unaligned knee MR images in order to find a series of intermediate templates for improved affine registrations. We propose a method for sparse graph construction of an empirical manifold which is well suited to the registration of a large set of images. The method proposed here, also differs because we compute all registrations in the sparse graph offline.

Registrations may fail because the optimisation only finds a local minima and will not reliably locate the global minimum. We hypothesise that pairs of images with greater similarities are more likely to register successfully because the initial transformation is closer to the final solution. Therefore, we assume that images with high similarity will generally achieve an accurate affine registration. A large and varied dataset maybe considered advantageous as it is likely to result in a dense sampling of the manifold. Knee MRI are acquired in a variety of orientations, scales and positions (in addition to non-rigid variations in anatomy and pose), in a densely sample manifold a path of intermediate templates may exist between two very different subjects.

The low dimensional empirical manifold is represented as a graph with similar knee MRI connected via an edge. The proposed method suggests a way for a small number of registrations only between similar images to be precomputed. This means the computational cost is low.

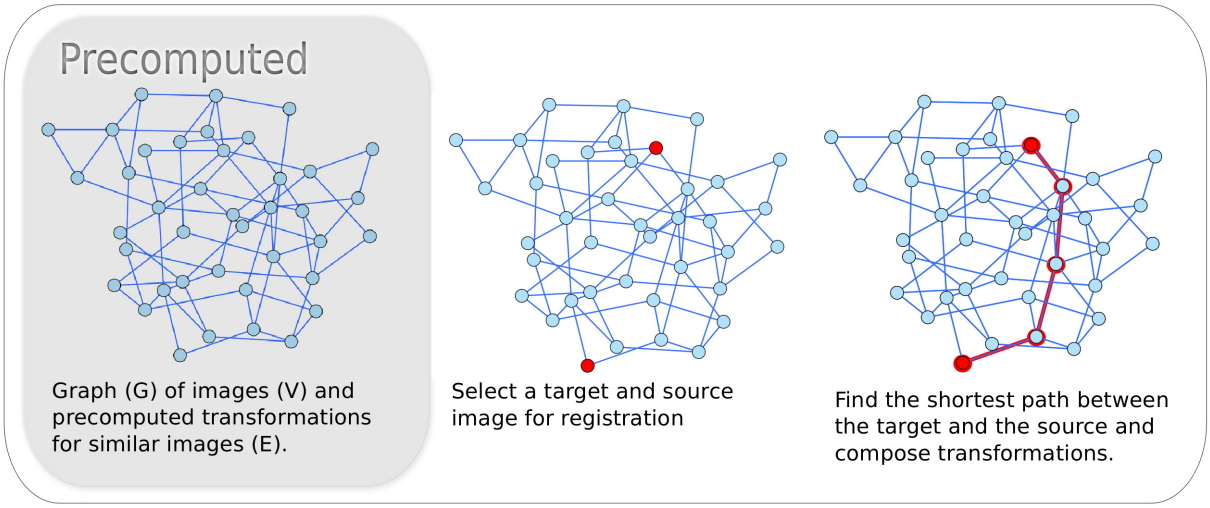


Figure 5.2: Registration scheme of standard sparse graphical registration

Figure 5.2 depicts an overview of this framework.

In the second section of this chapter we propose a refinement to geodesic registration. Any intermediate transformation has a small error associated with it and so the shortest path in the empirical manifold incorporates the registration errors. We consider the geodesic registration as an initialisation of registration. The refinement step is a further affine registration to compensate for such accumulated errors. The idea of registration refinement is further explored and discussed in more detail in chapter 6.

5.2.2 Representation of the population

A graph $G = (V, E)$ is constructed to model the low dimensional manifold of all MRIs in native space for an anatomical region of interest. Each vertex in V is a sample on the manifold and V represents the set of images $\{\boldsymbol{I}_1, \boldsymbol{I}_2, ..., \boldsymbol{I}_n\}$ in the dataset. Each edge in E is weighted $w_{i,j}$ according to the similarity of images $\boldsymbol{I}_i$ and $\boldsymbol{I}_j$ (see section 5.2.3 for more details). Each edge $e_{i,j}$ also has a transformation, $\boldsymbol{T}_{i,j}$ associated with it describing a spatial mapping from image $\boldsymbol{I}_i$ to image $\boldsymbol{I}_j$.

G is chosen to be an undirected, sparse k-nearest neighbour graph. The representation G is designed to model the local neighbourhood properties and preserve local relationships, this

is why G is a sparse k-nearest neighbour graph, the connectivity of G is discussed in more details in section 5.2.4. G is chosen to be an undirected graph since affine transformations are invertible and so transformations are only computed in one direction. The inverse is computed on the fly when required, this is discussed in more detail in section 5.2.5.

The registration transformation $\boldsymbol{T}_{i,j}$ is assumed to be more likely to be successful if edge $e_{i,j}$ is present and weigth $w_{i,j}$ is low, since this indicates that images $\boldsymbol{I}_i$ and $\boldsymbol{I}_j$ have high similarity. Images with high similarity are likely to be closer to a registration minimum and so have an increased likelihood of a successful registration. Conversely, very dissimilar subjects will not be nearest neighbours and so a direct transformation is not computed between such images.

5.2.3 Edge Weights

We weight the each edge $e_{i,j}$ in graph G based on the dissimilarity of images $\boldsymbol{I}_i$ and $\boldsymbol{I}_j$. The weighting $w_{i,j}$ reflects the quality of an affine registration, where a low weight is more likely to be successful since the transformation is theoretically closer to the global minima. A lower weighted edge is more likely to be travelled through in a shortest path computation.

5.2.4 Connectivity of k-Nearest Neighbour Graph

The graph G is computed as a k-nearest neighbour graph, thus an edge exists between two images $\boldsymbol{I}_i$ and $\boldsymbol{I}_j$ if the similarity metric associated with edge $e_{i,j}$ is amongst the k highest edge weights connected to either v_i or v_j. The similarity metric to define the nearest neighbours is application specific. In our preliminary experiments we observed that normalised cross correlation resulted in better registrations than normalised mutual information. We therefore use normalised cross correlation in the remainder of this chapter.

To construct the graph, pairwise similarities are computed for all pairs of images, due to the size of the OAI dataset $|V|$ is very large. Computing all pairwise registrations in advance is intractable and has complexity of $\binom{|V|}{2}$ registrations. For example if $|V| = 2000$, the number of registrations needed would be $1,999,000$. Assuming that an affine registration take 6 minutes,

the computations would take 8329 days on a single core machine. As a result it is not feasible to compute the post-registration similarity metric to determine connectivity.

The images are blurred and downsampled prior to computing similarities. It is computationally expensive to compute pairwise similarities of full image resolution using the available architecture when the experiment was devised, mostly due to network bandwidth when reading images from disk. Since the goal of the method is to achieve a strong global transformation and local details are immaterial at this stage, pairwise similarities of images at this lower resolution are sufficient.

5.2.5 Intermediate Transformations

Intermediate transformations are computed between images which are nearest neighbours on the empirical manifold. For each edge $e_{i,j} \in E$, the intermediate transformation $\boldsymbol{T}_{i,j}$ is computed as an affine registration between images $\boldsymbol{I}_i$ and $\boldsymbol{I}_j$. $\boldsymbol{T}_{i,j}$ is stored in a database, with its direction explicitly specified such that a matrix inversion of the transformation matrix $\boldsymbol{T}_{i,j}^{-1}$ can be computed in real time to yield $\boldsymbol{T}_{j,i}$. Registrations for all E are precomputed so any intermediate transformation can be retrieved instantaneously. All $\boldsymbol{T}_{i,j}$ are computed using an intensity based affine registration algorithm (Studholme et al. (1998)).

5.2.6 Geodesic Pairwise Registrations

To register any pair of images $\boldsymbol{I}_i$ and $\boldsymbol{I}_j$, which may be very dissimilar, the empirical manifold G can be used to find a sequence of similar intermediate templates. A corresponding sequence of intermediate transformations can be composed to find a geodesic transformation ($\boldsymbol{\mathcal{T}}_{i,j}$).

The sequence of intermediate transformations is computed as the geodesic shortest path in G from $\boldsymbol{I}_i$ to $\boldsymbol{I}_j$ using Dijkstra's algorithm (Dijkstra (1959)). The elements of the shortest path shall be indexed sequentially based on the element's position in the path and shall be denoted as $\bullet'$, to distinguish it from indexing in G. A path in the graph $< v'_1, e'_1, \ldots, e'_{s-1}, v'_s >$ of length s has an associated cost of the sum of the weights on each of the edges passed $\sum_{r=1}^{s} w'_r$. A path

constructed of vertices and edges in the graph encodes a representation of a series of similar images connected by intermediate templates $\boldsymbol{I}$ and intermediate transformations $\boldsymbol{T}$. Therefore this path can be expressed as $< \boldsymbol{I}'_1, \boldsymbol{T}'_1, \ldots, \boldsymbol{T}'_{s-1}, \boldsymbol{I}'_s >$.

Geodesic transformation, $\boldsymbol{\mathcal{T}}_{i,j}$, between images $\boldsymbol{I}_i$ and $\boldsymbol{I}_j$ is computed as the composition of the intermediate transformations along the shortest geodesic path $< \boldsymbol{I}'_1, \boldsymbol{T}'_1, \ldots, \boldsymbol{T}'_{s-1}, \boldsymbol{I}'_s >$, where $\boldsymbol{I}'_1 = \boldsymbol{I}_i$ and $\boldsymbol{I}'_s = \boldsymbol{I}_j$.

$$\boldsymbol{\mathcal{T}}_{i,j} = \boldsymbol{\mathcal{T}}'_{1,s} = \boldsymbol{T}'_1 \circ \boldsymbol{T}'_2 \circ \ldots \circ \boldsymbol{T}'_{s-1} \tag{5.1}$$

Any geodesic pairwise registration can be retrieved rapidly. The composition of geodesic transformations is achieved by a short series of matrix multiplications since all of the intermediate transformations are computed in advance.

5.2.7 Simple Refinement of Registrations

When several affine transformations are composed, small errors in each transformation will accumulate. Therefore the geodesic composed transformation $\boldsymbol{\mathcal{T}}$ is suboptimal and it is possible that it will have a considerable error associated with it. A refinement step can be employed in which $\boldsymbol{\mathcal{T}}$ is considered to be an initialisation for a further affine registration step. The refined geodesic transformation is referred to as $\hat{\boldsymbol{\mathcal{T}}}$, the refinement transformation is referred to as $\boldsymbol{T}_{refined}$. This is illustrated in figure 5.3.

$$\hat{\boldsymbol{\mathcal{T}}} = \boldsymbol{\mathcal{T}} \circ \boldsymbol{T}_{refined} \tag{5.2}$$

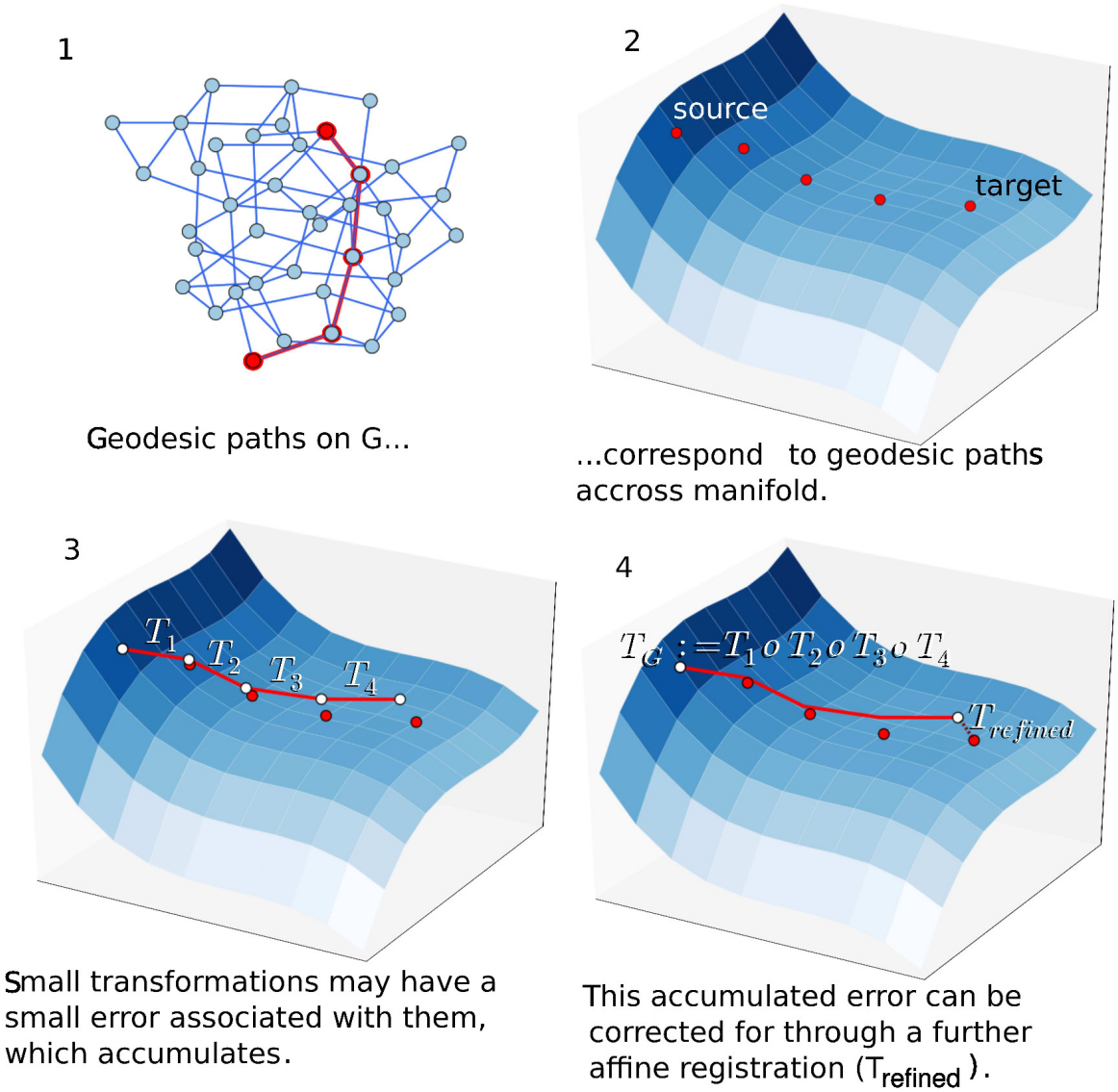


Figure 5.3: Illustration of accumulation of affine registration error and the refinement step to reduce errors

5.3 Experiments

5.3.1 Data

Image Data

The experiments presented use 10,307 MRIs of left and right knees obtained from the OAI. Specific datasets used are 1.C.0 and 1.E.0 at baseline and 1.C.1 and 1.E.1 at 12 months after the baseline date. The fat-suppressed, sagittal 3D dual-echo in steady state (DESS) sequence with selective water excitation (WE) has been selected since it has both high in plane resolution (0.36 x 0.36mm) and a small slice thickness (0.7mm).

Validation Data

Pairwise registration accuracy is validated for a set of 97 randomly selected and manually annotated images. The size of the validation set is significantly smaller than the size of the whole dataset due to the overhead of manual annotation. Since the graph is symmetric, registrations are only computed in one direction which means that 4656 registrations are evaluated.

Each image is annotated with four distinct landmarks, at the ACL and PCL ligament insertions on the femur and the tibia. The middle voxel of each ligament is selected just before it meets the bone. These landmarks are selected as they are visible in most subjects and exhibit a high degree of placement reproducibility by readers. Osteoarthritis is predominantly associated with structures in the joint space (articular cartilage, meniscus, synovium etc) therefore, observing registration error in the joint space is clinically relevant.

It is time consuming and challenging to annotate landmarks on knee images which are not affinely aligned. As a result, only four landmark points are annotated on the dataset. Landmarks which are harder to place are likely to have a higher target localisation error, which adds noise when measuring the target registration error.

Computing Target Error

The registrations are validated by computing the error as the mean L_2 norm of the target registration error (TRE), from N annotated landmarks (here $N = 4$). $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$ are the ith annotated landmarks in the source and target image respectively. Landmark $\boldsymbol{p}_i$ undergoes transformation $\boldsymbol{\mathcal{T}}$, where $\boldsymbol{\mathcal{T}}$ can be substituted with other transformations discussed in this chapter (eg $\boldsymbol{T}$, $\hat{\boldsymbol{\mathcal{T}}}$)

$$err = \frac{1}{N}\sum_{i}^{N} ||\boldsymbol{\mathcal{T}}(\boldsymbol{p}_i) - \boldsymbol{q}_i|| \tag{5.3}$$

Registration Failure

We have defined registrations with mean target registration error (err) greater than 10mm as failures. This error threshold is recognised to be fairly arbitrary. Despite the arbitrary nature of this chosen threshold it is convenient to be able to quantify the number of registration errors which are unacceptably high. The threshold has been set due to experimental findings. The errors have been observed to form a log-normal distribution with a large tail for high registration errors.

Assessing Target Localisation Error

The annotated landmarks are the center of the cruciate ligaments as they attach to the bone. These landmarks were selected since landmarks on thin and cylindrical ligaments were easy to identify. The mean target localisation error is computed using equation 5.3 where $\mathcal{T}$ is the identity . It is computed to quantify the reproducibility of manual annotations for a single reader.

Initially we tested landmark localisation on an MRI annotated twice by a single reader. 13 subjects were annotated twice, the mean localisation error computed between landmarks was 1.1mm. This suggests there is an expected reader error of 1.1mm, when all other possible

sources of errors are eliminated.

We also tested landmark localisation on a single subject, where the MRI is acquired at different time points. The anatomy is consistent between scans but they are not affinely registered. It is challenging to accurately localise landmarks on images which are not affinely registered. Unregistered baseline and 12 month MRI of 13 subjects were annotated. The 12 month MRI were affinely registered to the corresponding baseline MRI. All registrations were manually verified for failures and could be assumed successful. The mean localisation error is 2.1mm. This explains a lower bound for same subject localisation error with pose variation.

The errors explored can predominantly be attributed to target localisation error. These experiments provide lower bound error expectations for target registration errors computed in this chapter.

5.3.2 Implementation Details

Choosing k for Graph Construction

To determine the connectivity (the set E) of the graph G we use a k-nearest neighbour approach. When k is very large, G is fully connected and thus as k grows the shortest path across the graph is more likely to be equivalent to a direct registration. However, if k is too small, there would be multiple connected components and a shortest path would not exist between some vertices. In these experiments we chose the smallest k such that there is only one connected component.

Centring

A common initial guess for a registration is derived from aligning the images using the real world scanner co-ordinates. However, this is not an appropriate initialisation for affine registration of the knee which is susceptible to convergence to local minima. This was verified via preliminary experiments from the OAI dataset. Instead the central voxel on each image is aligned as an

initialisation by using a translation operation. This appears to be more successful, we speculate that this is due to the effort made by the radiologist to ensure that the joint is approximately at the centre of each image.

This initial transformation is used to define the connectivity of the graph, the edge weights and as an initial guess when computing the intermediate transformations.

Degrees of Freedom

Experiments regarding the number of degrees of freedom appropriate for computing intermediate transformations with this dataset were performed and nine degrees of freedom were deemed sufficient, these included translations, rotations and scaling. It was found from visually inspecting registration results that occasionally large skews would produce an inappropriate transformation. These transformations maximised the similarity of the images but generated transformations which were not anatomically meaningful with regards to the knee. Typically in these cases, the transformations were represented more faithfully as a scale plus a rotation operation. We speculate that this is due to the vastly differing scales of knees in the human population.

Geodesic transformations might incorporate skew transformations. A composition of transformation where a rotation is composed with a subsequent scale will yield a skew. However, we considered such a skew in the geodesic registration's transformation model to be anatomically meaningful because we assume that each intermediate transformation is anatomically meaningful, the model is constrained such that that large skew transformations which do not reflect anatomy are unlikely.

Similarity Metric

The similarity metric is used to define the connectivity of the graph, the edge weights and is used for computing intermediate transformations. In all cases we use measures which are based

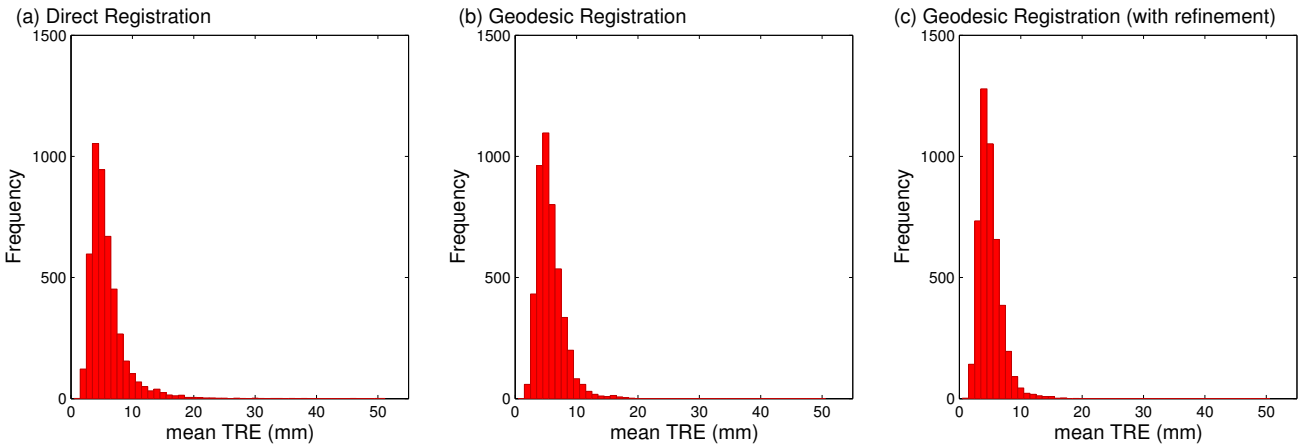


Figure 5.4: Histograms of the registration error (mean Euclidean distance between a set of landmarked points) distribution for all pairwise registrations

on normalised cross correlation (NCC), more details in section 2.3.3. To define edge weightings a dissimilarity metric is employed, this is computed as $w_{i,j} = 1 - NCC$.

5.3.3 Results and Analysis

Three registration approaches between each image pair (image $\boldsymbol{I}_i$ and image $\boldsymbol{I}_j$) are compared here :

1. Direct intensity based registration, $\boldsymbol{T}_{i,j}$
2. Geodesic registration, $\boldsymbol{\mathcal{T}}_{i,j}$
3. Geodesic registration with refinement, $\hat{\boldsymbol{\mathcal{T}}}_{i,j}$

Figure 5.4 shows the distribution of errors for all the pairwise registration. Table 5.2 shows the mean and standard deviation of the registration errors and the proportion of failed registrations. As can be seen from the results the geodesic registration improves upon direct registration, the standard deviation of the errors is reduced (4.69mm, 3.17mm respectively) and figure 5.4 indicates there are fewer outliers. However, errors appears to accumulate as transformations are composed along the geodesic shortest path. The mean and the standard deviation of the errors are further reduced due to a refinement step being added to geodesic registration. It can be seen in figure 5.4 (c) that the mean target registration error is smaller, in addition to

Affine registration algorithm	Mean Euclidean distance [mean (sd)] (mm)	Failed registrations (%)
Direct registration (Studholme et al. (1998))	5.80 (3.17)	4.17%
Geodesic registration	5.71 (2.17)	1.63%
Geodesic registration (refined)	5.01 (1.89)	0.97%

Table 5.2: Mean and standard deviation of registration error and the percentage of failed registrations. A failed registration is considered to have an error of greater than 10mm.

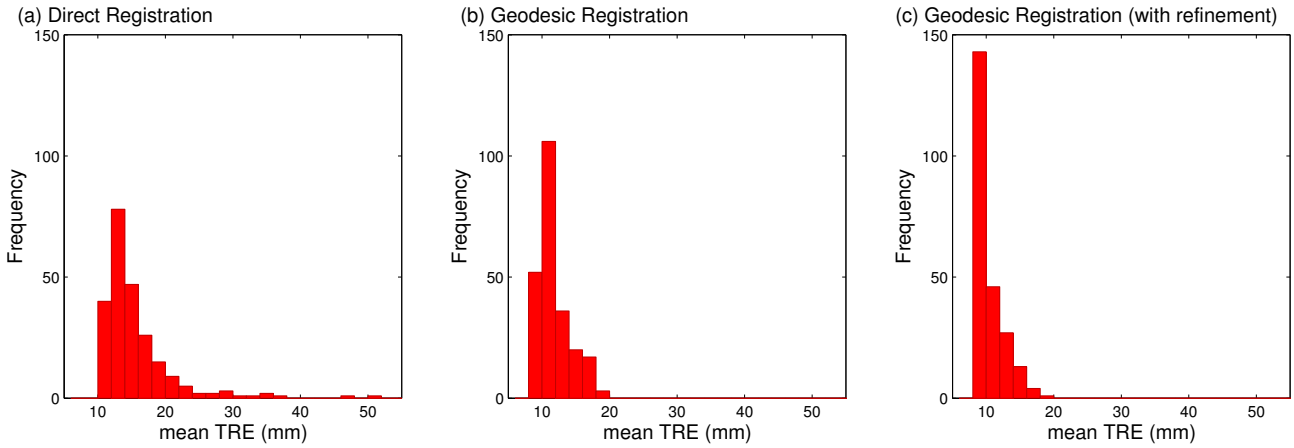


Figure 5.5: Histograms of the registration error distribution of 95th-100th percentile of ordered errors, showing the worst 5% of affine registrations for each approach.

fewer outliers with high registration error. This is further confirmed by the results in table 5.2 which show a significant reduction in mean registration error (-0.79mm), standard deviation (-1.28mm) and the number of failed registrations (-3.2%).

The focus of this work is to reduce the failure rate of affine registration. Figure 5.5 shows the error distribution across the worst 5% of affine registration (the 95th - 100th percentile). It can be observed that the geodesic registration method proposed here improves upon the direct registration, reducing the number of failed registrations. This is also evident in the statistics reported in table 5.2.

Figure 5.6 shows how the registration error changes when the geodesic registration with refinement is used compared to direct registration. Green indicates geodesic registration with refinement is more accurate. Orange indicates that both methods are of similar accuracy. Red indicates direct registration is more accurate. It can be seen that there are many more green points, which indicates that geodesic registration with refinement is generally more accurate

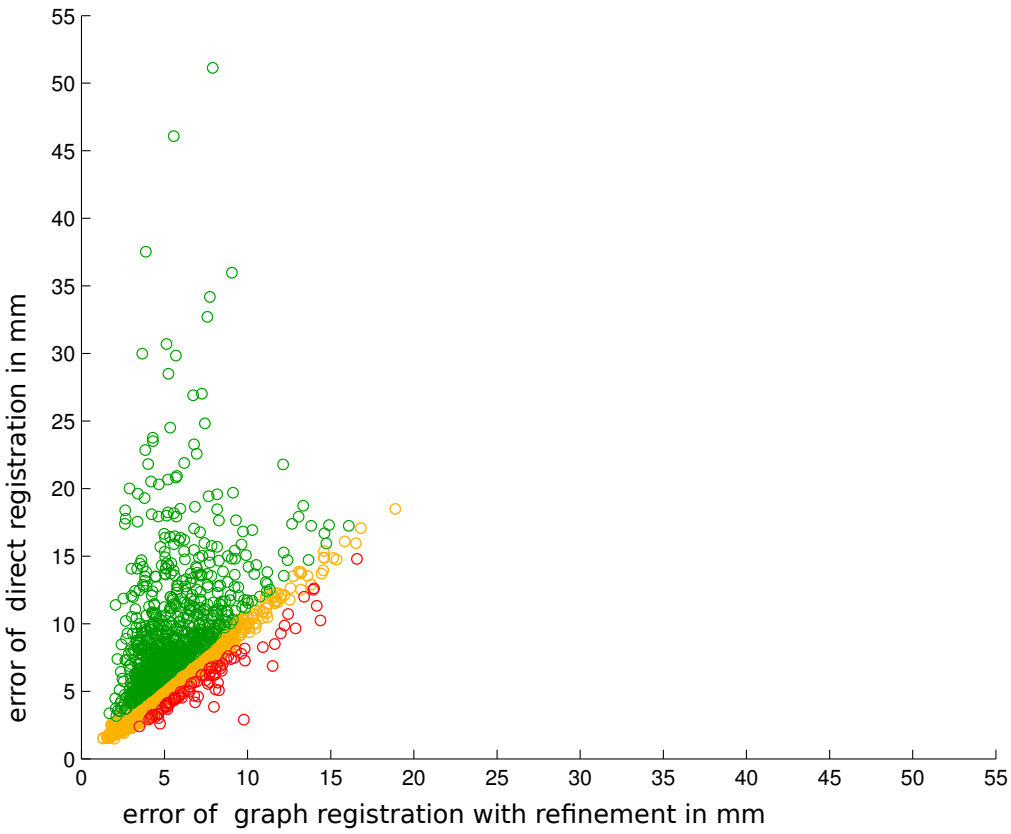


Figure 5.6: A quantitative assessment of both direct registration and geodesic registration with refinement. Each point on the graph represents a registration error. Points are red when the direct registration error is lower than geodesic registration with refinement error. Points are green when the direct registration error is higher than geodesic registration with refinement error. Points are orange when both errors associated with both methods are approximately equivalent.

than direct registration.

Figure 5.7 gives a qualitative example of failed direct affine registration which was improved by the geodesic registrations framework.

5.4 Discussion

In this chapter we have proposed modelling the population of unregistered knee MRI as an empirical manifold using a graph, where the local properties of the manifold are modelled at each vertex. This enables registrations to be precomputed for each edge of the graph. Since

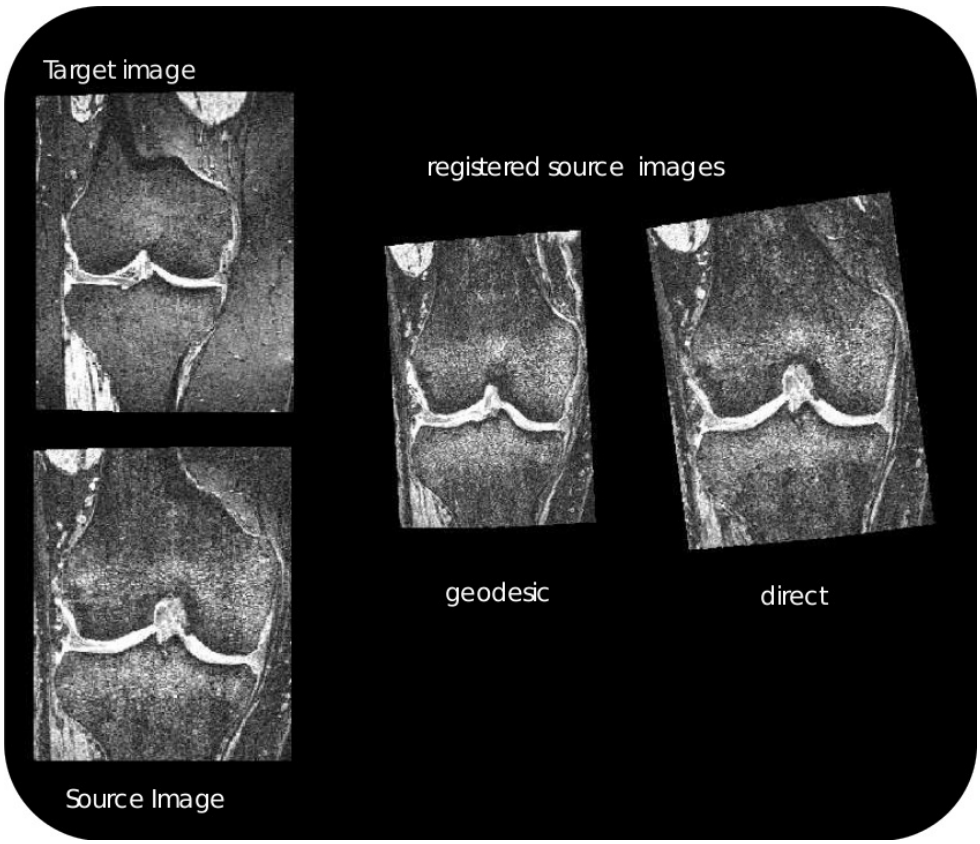


Figure 5.7: Qualitative example of inaccurate registrations improved by geodesic framework.

the graph is a sparse k-nearest neighbour graph where k is small, the maximum number of registrations required is $k \times |V|$. However, it is frequently observed that nearest neighbours might be common between vertices and so this upper bound is unlikely to be reached. All of the intermediate transformations are precomputed as part of the training phase and Dijkstra's shortest path algorithm is highly efficient, which means the run time computational cost is low. A registration computed via the shortest path on the graph can be considered an initialisation which is quickly computed between any pair of images. A refinement step is also proposed, leading to to dramatically increased registration accuracy relative to direct affine registrations.

Mean target registration errors were shown to be approximately $5mm$. However, some of this error could be attributed to the target localisation error of the manually placed landmarks. Mean target localisation error was estimated to be around $2mm$ when markers were placed on the same subject at different poses. However, this estimation does not include anatomical variation that is likely to be included in the mean target registration errors. This interpretation enables us to suggest that on average these localisation errors are fairly small. Despite this,

there are still outliers present and this motivates future work to reduce the number registrations with large errors.

5.4.1 Limitations

The search for k-nearest neighbours for the graph construction is exhaustive and the number of comparisons grows quadratically with the number of images. There are many contributions in the literature which discuss improvements, including hashing (Indyk and Motwani (1998)) or tree based searches (Bentley (1975)). However, since the graph only needs to be constructed once, this was not considered to be a major limitation.

An out-of-sample extension could be added to this framework so that an unseen image could be accurately registered to any image in the population. Initially, the nearest neighbour of the unseen image would need to be determined, then the method proposed here can be applied to find the registration. This scenario would benefit from an efficient nearest neighbour algorithm to find the nearest image in the empirical manifold, a sub-linear search would be preferable to an exhaustive linear search.

Currently, similarity metrics can only be computed at low resolutions for the graph construction due to the large cost of pairwise computations. For the purposes of this application, the computations at low resolution appears to be sufficient since the methods achieves boosted performance when compared with direct registration. However, the effect of similarities at higher resolutions could be explored using a multiscale approach. The l-nearest neighbours of G determined by pairwise similarity of low resolution images can be found, where $|V| \gg l > k$. From the set of l selected neighbours, the k-nearest neighbours can be determined using similarity metrics of higher resolution images.

5.4.2 Future Work

The geodesic registration method presented here performs well only as an initialisation. This was investigated by manually inspecting registrations with poor accuracy that were computed

using the geodesic framework. It appears that the registration failures can typically be assigned to occasional erroneous edges on the graph. The next chapter shall explore approaches to omit such erroneous edges or improve paths through more sophisticated machine learning algorithms.

Chapter 6

Fast and Accurate Global Geodesic Registrations

Work in this chapter has, in part, been presented in:

C. R. Donoghue, A. Rao, A. M. J. Bull, and D. Rueckert. Fast and Accurate Global Geodesic Registrations using Knee MRI from the Osteoarthritis Initiative. *Medical Computer Vision Workshop (MCV) at Computer Vision and Pattern Recognition (CVPR)*, 2012

Abstract

This chapter follows on from chapter 5 and addresses some of the limitations. In chapter 5 it was found that the geodesic registration method provides small accuracy improvements and faster registration compared with direct registration. The refinement method proposed significantly increases accuracy of the geodesic registrations but reduces the speed. In this chapter we present two novel refinement steps which boost registration accuracy over the graph but also preserve speed advantages of the geodesic registration. We use all precomputed transformations to infer transformation errors for each edge, through assuming global registration cycle consistency across a sparse graph. In conjunction with this, we propose fusing multiple successful registrations as a strategy to mitigate small errors in each intermediate transformation of the graph. It is shown that, in combination, these techniques achieve more accurate pairwise registrations than both geodesic registration and direct pairwise registration. This chapter addresses accuracy of registrations, speed of computation and scalability to a large scale dataset.

6.1 Introduction

The previous chapter (chapter 5) aimed to reduce the frequency of 'failed' registrations using an empirical model of the manifold of unregistered knee MR images. A shortest geodesic path across the empirical manifold is used to find accurate approximations of the correct transformation between a pair of images embedded in the manifold, this is fast to compute. A refinement step to increase the accuracy of geodesic registrations is also presented but the refinement step increases the computational costs so that it is equivalent to a direct registration.

This chapter provides a method which enables fast, robust and accurate pairwise registrations to be computed for all images embedded in the empirical manifold. It combines the speed benefits of the transformation being computed using a geodesic path across an empirical manifold (sections 5.2.1 - 5.2.6) together with the accuracy benefits of the refinement (section 5.2.7).

This chapter begins by identifying the key causes of registration errors which are incorporated into the model presented in chapter 5. If the registration errors which are corrected using the

refinement step (section 5.2.7) can be identified and mitigated during the geodesic registration algorithm presented then the refinement step can be eliminated.

6.1.1 Notation

We shall overview notation for consistency with the previous chapter. Geodesic transformation $\boldsymbol{\mathcal{T}}$ is computed from a shortest path over an empirical manifold represented as sparse graph $G = (V, E)$. The shortest path is of length s; $< v'_1, e'_1, \ldots, e'_{s-1}, v'_s >$, which has cost $c = \sum_{r=1}^{s} w'_r$, $\bullet'$ defines elements of a path which are indexed sequentially. The shortest path encodes a representation of a series of highly similar images connected by transformations $\boldsymbol{T}$. This path can be expressed as $< \boldsymbol{I}'_1, \boldsymbol{T}'_1, \ldots, \boldsymbol{T}'_{s-1}, \boldsymbol{I}'_s >$. Intermediate transformations $\boldsymbol{T}_{i,j}$ associated with edge $e_{i,j} \in E$ are all precomputed using a standard affine registration algorithm. Transformation $\boldsymbol{\mathcal{T}}_{\boldsymbol{I}'_1, \boldsymbol{I}'_s}$ between images $\boldsymbol{I}'_1$ and $\boldsymbol{I}'_s$ is computed as the composition of the intermediate transformations $\boldsymbol{\mathcal{T}}_{\boldsymbol{I}'_1, \boldsymbol{I}'_s} = \boldsymbol{T}'_1 \circ \boldsymbol{T}'_2 \circ \ldots \circ \boldsymbol{T}'_{s-1}$.

6.1.2 Sources of Registration Error in Geodesic Registration

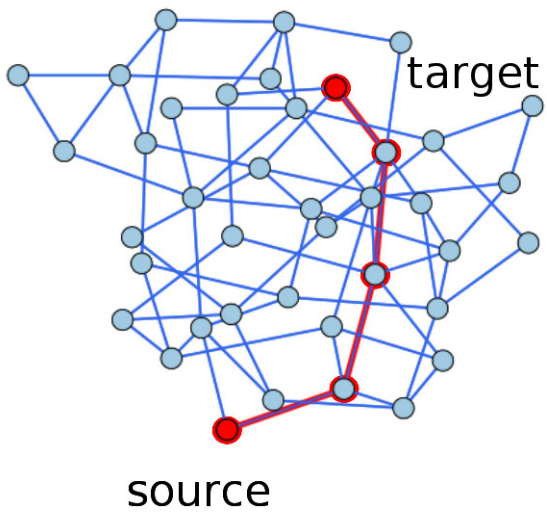
Registrations errors which are incorporated in a shortest geodesic path of the graph can be two-fold (illustrated in figure 6.1):

Error type 1 : Each intermediate transformation $\boldsymbol{T}$ are likely to incorporate small errors due to local minima, ambiguities and imperfect registrations on a discretised grid. These errors can accumulate along a path and become significant for a geodesic registration resulting with error in $\boldsymbol{\mathcal{T}}$.

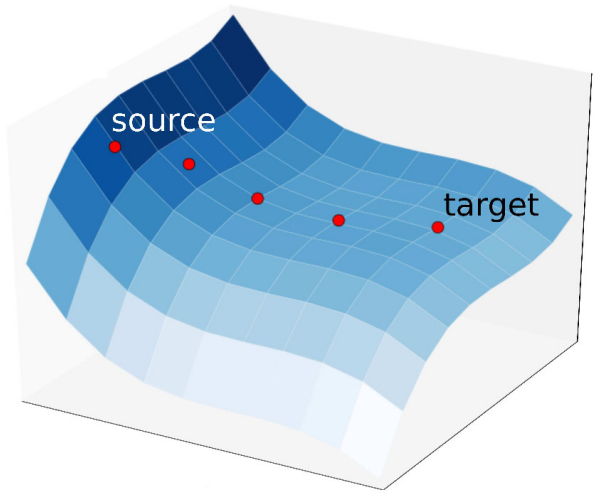
Error type 2 : It is possible that there exists intermediate transformations $\boldsymbol{T}$ which have failed to find correspondences between subjects. This can result in a large error in any $\boldsymbol{\mathcal{T}}$ which contains an erroneous intermediate transformation.

Figure 6.1: A registration between a source and target image across the shortest path in the graph (6.1 (a)) is used to represent a path upon the empirical manifold as depicted in 6.1 (b). However this method incurs two types of registration errors as described in subfigures 6.1 (c) and 6.1 (d).

(a) geodesic shortest path across the discrete representation of empirical manifold

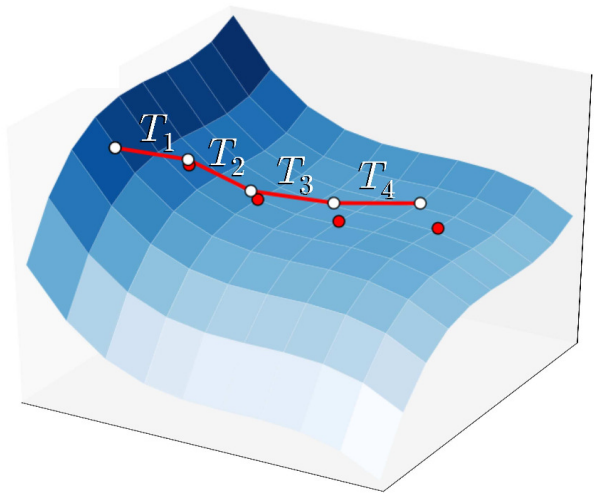


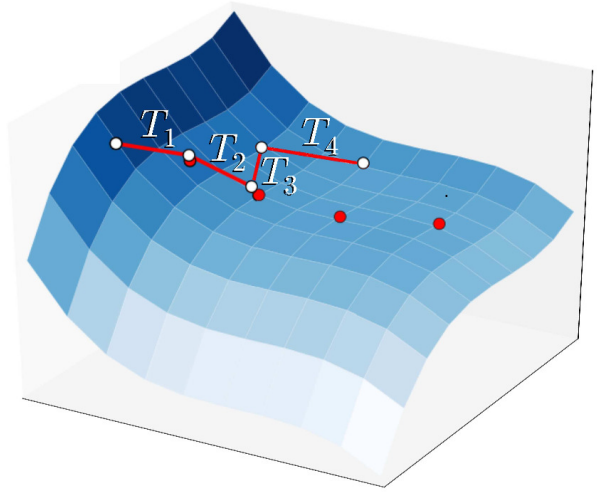
(b) images in the geodesic shortest path over the empirical manifold



(c) ***Error type 1 :*** small errors are associated with each registration since an affine registration can not define a perfect spatial mapping between two subjects with different anatomy, this error can accumulate across a path. These errors are addressed using *Registration Fusion*.

(d) ***Error type 2 :*** occasionally some of the intermediate transformations within the graph are erroneous registrations, these errors can be quite significant, as with transformation $\boldsymbol{T}_3$ in the diagram. Such registrations disrupt any registration path that passes this edge. The diagram also contains ***error type 1***, since ***error type 1*** is inherently ubiquitous. *Cycle Consistency* is used to try to mitigate the effects of ***error type 2***.





6.1.3 Strategies to Reduce Registration Error

This chapter presents two solutions devised to combat each of these error types. We mitigate ***error type 1*** by computing transformations for several shortest geodesic paths in the graph and computing the average of these transformations. This is analogous to selecting several approximations and fusing to yield an accurate transformation. Future references to this concept shall be referred to as *registration fusion*.

As described by ***error type 2***, a single erroneous transformation in G has the potential to pollute a number of geodesic registrations if it is used in multiple shortest geodesic paths. It is assumed that the majority of the intermediate transformations in G are reasonably accurate because only similar images are registered. Therefore very inaccurate intermediate transformations are not consistent with most intermediate transformations in the G. Inaccurate intermediate transformation can be identified offline prior to geodesic registration since all intermediate transformations $\boldsymbol{T}$ of the empirical manifold are precomputed. The edge weight matrix $\boldsymbol{W}$ can be adjusted such that intermediate transformations for which a high error is predicted can be penalised. Consistency is established by observing cycles in the graph, hence future references shall refer to this concept as *cycle consistency*.

6.1.4 Contributions

This chapter contributes a method which is fast, robust and tested on a large scale dataset.

By mitigating error types 1 and 2 using *registration fusion* and *cycle consistency*, a geodesic registration has been demonstrated to offer improved *robustness* relative to a direct registration. We find that the mean error is reduced by 1.0mm and the standard deviation is reduced by 1.9mm . We observe the number of outliers to be significantly decreased.

The computational speed of the algorithm is improved compared to geodesic registration with refinement (section 5.2.7). The proposed method computes a registration between a pair of images in approximately 2 seconds. Moreover, *cycle consistency* can boost accuracy offline and

thus does not affect runtime computational speed of geodesic registrations. *Registration fusion* computes multiple shortest paths when computing a registration between a pair of images. We show that the computational cost grows linearly with respect to the number of paths (in section 6.2.2).

The method we present has been demonstrated to perform well on a modern *large-scale dataset*, in addition there is reason to believe that performance is enhanced due to the dense sampling of subjects on the manifold.

6.2 Increasing Geodesic Registration Accuracy through Refinements

The geodesic registration algorithm in chapter 5 proposes computing a transformation composed from the shortest geodesic path in a graph connected by image similarity. A refinement step (section 5.2.7) was applied to enhance the accuracy of the algorithm. This chapter investigates two alternative refinement methods instead; registration fusion and cycle consistency.

6.2.1 Graph Construction for Geodesic Registration

The geodesic registration algorithm presented here is very similar to the basic algorithm presented in sections 5.2.1 - 5.2.6.

The set of edges E in the G are the k-nearest neighbours of V. The nearest neighbours are defined by using the pre-registration similarity metric between two connected images. The choice of connectivity parameter k varies between chapters, in this chapter the selection of the parameter is explored in more detail in section 6.3.2. All other details of the algorithm remain constant between chapters.

The weightings W of edges E in G are defined as the post-registration normalised-cross correlation similarity. This is in contrast to the approach in described section 5.2.3 where pre-

registration similarity is used for both selecting k-nearest neighbours and weightings W. The modification is made because the variance of pre-registration similarity is low and it is hypothesised that post-registration similarity assesses registration quality more accurately.

6.2.2 Registration Fusion : Fusing Registration on Multiple Paths

In general, an affine registration error is present at each edge in G and small errors can accumulate along a path in G. Registration errors here can be attributed to convergence to a local minima, numerical inaccuracies or that no single affine transformation exists for all anatomy. We propose eliminating some of the accumulated deviations of $\boldsymbol{\mathcal{T}}$ by fusing multiple transformation approximations $\boldsymbol{\mathcal{T}}^1, \ldots, \boldsymbol{\mathcal{T}}^{\mathcal{K}}$. The fusion is expected to average out random noise .

A similar idea has previously been adopted to fuse segmentations created by multiple raters (Warfield et al. (2004)) or multiple atlases where each segmentation is viewed as a classifier (Heckemann et al. (2006)). In the segmentation scenario, each classifier votes on the final segmentation outcome. In this work multiple registration estimates can be combined to generate improved registration. An overview of the registration fusion algorithm is presented in figure 6.2.

Multiple Paths using Yen's loopless $\mathcal{K}$-shortest paths

We propose finding the $\mathcal{K}$-shortest-paths[1] on the graph using a fast implementation of the Yen's loopless $\mathcal{K}$-shortest paths algorithm (Martins and Pascoal (2003)). Yen's algorithm is known as a deviation algorithm. Firstly it finds the shortest path and then seeks to find $\mathcal{K}-1$ deviations of the best path. When $\mathcal{K}$-shortest paths are requested between images associated with vertices v_i and v_j over graph G, the algorithm constructs a "psuedo"-tree representation of the graph. The "pseudo"-tree has v_i at the root of the tree and v_j at the leaf nodes. A route between the root and a leaf defines a path q_i. It is called a "psuedo"-tree because it has repeated nodes in

[1]Within this section $\mathcal{K}$ is used to denote the number of shortest paths which should not be confused with connectivity parameter k.

the tree. However, the tree does not contain repeated nodes in any path q_i which is why the algorithm is considered to be loopless. No edges can have a non-negative weight associated with them. This is consistent with the work in this thesis since all weights in graph G are non-negative. An illustration of the "pseudo"-tree can be seen in figure 6.2 (d).

The implementation employed here[2] differs from the standard Yen's $\mathcal{K}$-shortest path algorithm by analysing the nodes in a different order. This enables improved worst case computational complexity. The algorithm has complexity of $\mathcal{O}(\mathcal{K}n(m+n\log(n)))$, where $n = |V|$ and $m = |E|$. Assuming that a suitable empirical manifold is learnt and represented as G with fixed n and m, the algorithm is of linear complexity with respect to the number of paths selected. Each shortest path computed by Yen's algorithm can be considered a guess at an accurate global transformation. We shall refer to the $\mathcal{K}$ transformation approximations as the set $\{\boldsymbol{\mathcal{T}}^1, \boldsymbol{\mathcal{T}}^2, \ldots, \boldsymbol{\mathcal{T}}^{\mathcal{K}}\}$.

Fusing Transformations

Registration fusion computes a single transformation from $\{\boldsymbol{\mathcal{T}}^1, \boldsymbol{\mathcal{T}}^2, \ldots, \boldsymbol{\mathcal{T}}^{\mathcal{K}}\}$. The transformations are fused by taking the mean of the transformations ($\bar{\boldsymbol{\mathcal{T}}}$).

Affine transformations do not occupy a linear space and as a result $\bar{\boldsymbol{\mathcal{T}}}$ can not simply be computed by calculating an element-wise mean of the affine matrices, where $\bar{\boldsymbol{\mathcal{T}}} = [\bar{\tau}_{i,j}]$, $\bar{\tau}_{i,j} \neq \frac{\sum_{k=1}^{\mathcal{K}} t_{i,j}^k}{\mathcal{K}}$. This can be illustrated by computing the element-wise mean of a 180° rotation (as transformation $\boldsymbol{\mathcal{T}}^1$) and a 0° rotation (as transformation $\boldsymbol{\mathcal{T}}^2$), where the average transformation would be expected to be a 90° rotation (as transformation $\bar{\boldsymbol{\mathcal{T}}}$). However an element-wise average of $\boldsymbol{\mathcal{T}}^1$ and $\boldsymbol{\mathcal{T}}^2$ results in the matrix in equation 6.1, which we denote as $\underline{\boldsymbol{\mathcal{T}}}$. It can be seen that $\bar{\boldsymbol{\mathcal{T}}} \neq \underline{\boldsymbol{\mathcal{T}}}$. Moreover, $\underline{\boldsymbol{\mathcal{T}}}$ is not an affine transformation since it is not a rotation, scaling, translation or shearing operation and it is singular and so thus is not constructed from these.

[2] C++ implementation of $\mathcal{K}$-shortest paths algorithm version 2.0 downloaded from http://code.google.com/p/k-shortest-paths/downloads/list (accessed December 2011)

$$\underline{\mathcal{T}} = \begin{pmatrix} (\cos(180)+1)/2 & 0 & 0 & 0 \\ 0 & (\cos(180)+1)/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.1}$$

These effects are because affine transformations exist in a Riemannian manifold. A Riemannian manifold (or metric space) is a space for which there is a defined distance metric and the topology is uniform. The distance between two points in a Riemannian manifold can not be computed as the Euclidean distance between the points, instead a geodesic distance is computed. Any measurements that need to be computed can not be computed in a vector space but instead measurements need to be generalised for Riemannian manifolds (Pennec (2006)).

To find the mean of $\mathcal{K}$ transformations we introduce the Frechet mean (Fréchet (1944), the original paper is in French but a high-level summary in English can be found in Pennec (2006)). The mean transformation as defined by the Frechet mean is the transformation which minimises the log-Euclidean distance between $\mathcal{K}$ transformations. It can be computed iteratively by minimising the distance between $\mathcal{K}$ transformations and the current estimate of the average transformation $\bar{\boldsymbol{\mathcal{T}}}$

$$\arg\min_{\bar{\boldsymbol{\mathcal{T}}}} \{\exp(\frac{1}{\mathcal{K}} \sum_{i=1}^{\mathcal{K}} |\log(\boldsymbol{\mathcal{T}}^i) - \log(\bar{\boldsymbol{\mathcal{T}}})|)\} \tag{6.2}$$

In this case $\boldsymbol{\mathcal{T}}^i$ refers to the ith transformation, the transformation is not raised to the power of i. The approach presented here using log-Euclidean distances has previously been presented for averaging transformations by Alexa (2002); Aljabar et al. (2008) and Arsigny et al. (2006) who uses the Frechet mean to average tensors.

We propose that each of these registration estimates should be voting for an accurate global registration. In this implementation we demonstrate the simplest voting scheme where each estimate has an equally weighted vote but further extensions could investigate the benefits of a weighted scheme since these are found to be advantageous for atlas based segmentation fusion

(Artaechevarria et al. (2009)).

We can increase the speed of pairwise affine registrations. Since registrations in G are precomputed, upon requesting a pairwise registration only the shortest path needs to be computed followed by a composition of the affine transformations along this path, which is a matrix multiplication of the 4x4 transformation matrices traversed on the path. However, we have observed that errors can be incorporated into registrations using the graphical model. In the previous chapter this was solved with a simple refinement of the registrations.

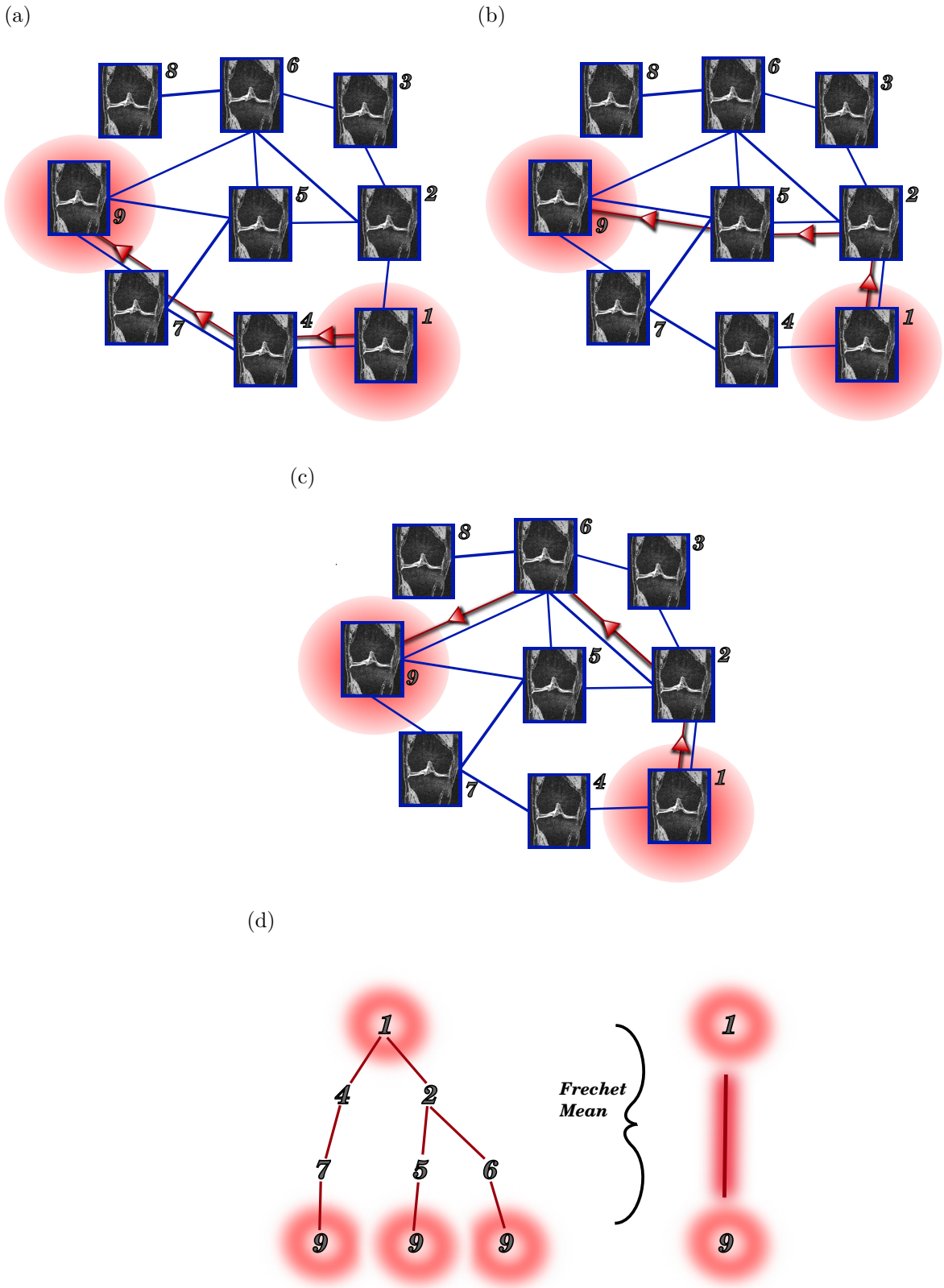
6.2.3 Cycle Consistency : Globally consistent transformations

The basic geodesic registration method presented in section 5.2 assumes that similar images are more likely to yield an accurate registration than two dissimilar ones. However, occasionally two similar images might result in convergence to a local minimum as opposed to the true global minimum. Large registration errors are less likely in graphs connected by only similar images but such errors exist. Additionally, in this graph mis-registered images can have a high similarity since registration optimises the similarity. Thus, one should be sceptical about using the same similarity measure as an indicator of registration accuracy. We propose inferring the error of each edge by observing known errors in the graph using a technique which we call cycle consistency. We use global cycle consistencies to improve the weight of edges, which indicates transformations quality, on graph G.

The idea of registration consistency has been introduced previously Christensen and Johnson (2001). However, more recently Datteri et al. (2011) has suggested using cycles to validate registrations and to improve on multi-atlas segmentation. The method presented by Datteri et al. (2011) uses a dense graph of images and therefore limits the number of subjects which can be included due to computational reasons. The method presented here is a sparse representation which can be solved for large numbers of subjects. The registration quality in this chapter is learnt offline, resulting in benefits in both speed and accuracy.

We extend on this method by assessing registration errors in the empirical manifold. This is

Figure 6.2: Martins and Pascoal (2003) is used to compute the multiple shortest paths in a graphs, these paths are then fused to find a registration between two subjects using a Frechet mean. 6.2 (a), 6.2 (b) and 6.2 (c) show all the paths selected by the multiple shortest paths algorithm. 6.2 (d) shows the three shortest paths represented as a loopless tree and the frechet mean registration fusion.



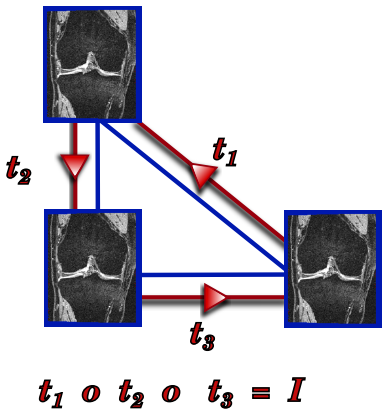


Figure 6.3: Cycle consistency

used as a similarity metric to weight the edges in the graph. This enables more reliable geodesic paths to be selected and in turn improved registration accuracy across the empirical manifold.

Cycle Consistency

If a path on G is taken such that it starts and finishes at the same vertex to form a cycle, we expect that the composed transformations on this path should be the identity transformation. Since the operation is effectively equivalent to registering an image to itself. To take this a step further, we can assume that all cycles in the graph should produce an identity transformation. Each edge is part of many cycles, so we use this intuition to infer the error of each edge in the graph by assuming that each cycle should adhere to this consistency constraint.

A cycle of length 3 is constructed of images $\boldsymbol{I}_i$, $\boldsymbol{I}_j$, $\boldsymbol{I}_k$ which are all connected by edges, $e_{i,j}$, $e_{j,k}$, $e_{k,i}$, with corresponding precomputed transformations , $\boldsymbol{T}_{i,j}$, $\boldsymbol{T}_{j,k}$, $\boldsymbol{T}_{k,i}$, in sparse graph G. The cyclic transformation is constructed by composing the transformations, $\boldsymbol{T}_c := \boldsymbol{T}_{i,j} \circ \boldsymbol{T}_{j,k} \circ \boldsymbol{T}_{k,i}$, as shown in figure 6.3.

d_c is used to express the distance between $\boldsymbol{T}_c$ and the identity which defines the consistency error. The distance function to compute d_c can be defined on an application specific basis. We assume that the error $x_{i,j}$ of each edge $e_{i,j}$ additively contributes to the cycle error and that $x_{i,j}$ is non-negative. We model the cycle error as $d_c = x_{i,j} + x_{j,k} + x_{k,i} + \epsilon$, where ϵ represents the random unmodelled registration errors. Considering all three-node cycles extracted from

our sparse graph, we globally model the cycle consistency as

$$\boldsymbol{d} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon} \quad s.t. \quad \boldsymbol{x} \geqslant 0\,, \tag{6.3}$$

in which $\boldsymbol{x}$ is the vector of unknown edge errors which we wish to recover. Here, $\boldsymbol{A}$ is a sparse matrix with elements $a_{i,j} \in \{0, 1\}$ which encode the connectivity of all cycles, where each row is a cycle and the columns are edge indexes. An entry of $\boldsymbol{A}$ is 1 if and only if the corresponding edge is present in the corresponding cycle. We adapted the method presented by Datteri et al. (2011) by formulating (6.3) as a least squares minimisation problem which enables us to introduce the convex non-negativity constraint:

$$\min_{x} ||\boldsymbol{A}\boldsymbol{x} - \boldsymbol{d}||_2^2 \quad st \quad \boldsymbol{x} \geqslant 0 \tag{6.4}$$

We optimise (6.4) using the method proposed in Coleman and Li (1996).

For this application, the distance between cycle transformation $\boldsymbol{T}_c$ and I is computed by measuring the mean error of the eight corner points of the 3D knee MR, $\boldsymbol{p}$, after transformation $\boldsymbol{T}_c$,

$$d_c = \frac{1}{8}\sum_{i}^{8} ||\boldsymbol{T}_c(\boldsymbol{p}_i) - \boldsymbol{p}_i||_2 \tag{6.5}$$

This approach is based on sparse precomputed pairwise transformations which are used to infer edge errors from known cycle errors (distances from the identity). We propose using the estimated edge errors to weight the transformations in our graphical registration framework. As shown later in the experiments, it is more reliable to use the estimated edge errors than other measures of registration accuracy such as the normalised cross correlation (NCC). However, many of the edge errors are very close to zero, which means that these weights are not appropriate for finding the shortest path, since a large path length is likely to be selected. We adjust for this effect by defining the edge weights $w_{i,j}$ as

$$w_{i.j} = x_{i,j} + \sigma(\boldsymbol{x}) \,, \tag{6.6}$$

where $\sigma(\boldsymbol{x})$ is the standard deviation of all learnt edge errors. The standard deviation is selected to reflect the distribution of $\boldsymbol{x}$. If the weight $w_{i,j}$ is too low for those edges with low error, it is likely that many edges will be selected when finding a shortest path. A long path is not desirable because it is possible that small errors will accumulate therefore is is necessary for some penalty $\sigma(\boldsymbol{x})$ to be associated with an edge even if the error is predicted as very low. Finally, it is important to mention that it is straightforward to generalise this approach to cycles composed of more than three nodes.

6.3 Experiments

6.3.1 Data

Image Data

The experiments presented use 2743 MRIs of right knees from the Osteoarthritis Initiative (OAI) public use dataset from groups 1.C.0 and 1.E.0 at baseline. The fat-suppressed, sagittal 3D dual-echo in steady state (DESS) sequence with selective water excitation (WE) (referred to here as Sag 3D DESS) has been selected, since it has both high in plane resolution (0.36mm by 0.36mm) and a small slice thickness (0.7mm).

The dataset used for experiments in section 5.3.1 uses subject's left and right knees, from baseline and 12 month scans. It is computationally expensive to compute the graph and registration error reduction methods for a very large dataset. To speed up experiments for this chapter, the dataset has been reduced to just include subject's right knee at baseline.

Validation Data

For validation of registration accuracy, 75 MRIs were randomly selected and annotated with landmark points at the insertions of the cruciate ligaments. The methods are tested for 2775 unique pairwise registrations. The selected landmarks and annotation protocol for the validation data is the same as in section 5.3.1.

6.3.2 Results and Analysis

We compare several approaches for registering image pairs:

Method 1 Direct registrations, where a pair of images are registered using a standard affine registration algorithm;

Method 2 A geodesic registration approach using a sparse graph described in section 5.2, which uses one geodesic registration path and a normalised cross correlation based distance metric;

Method 3 A geodesic registration approach using a sparse graph with cycle consistency to choose suitable weights for edges in the graph and a single registration path, as described in section 6.2.3;

Method 4 A geodesic registrations approach using a sparse graph with registration fusion, described in 6.2.2, using three fused registration paths and a normalised cross correlation based distance metric;

Method 5 A combination of the approaches, discussed in sections 5.2, 6.2.3 and 6.2.2, using a geodesic registrations approach using a sparse graph with three fused registration paths and cycle consistency to choose suitable weights for edges in the graph.

Analysis will begin by exploring the robustness of the algorithm to parameters k and $\mathcal{K}$.

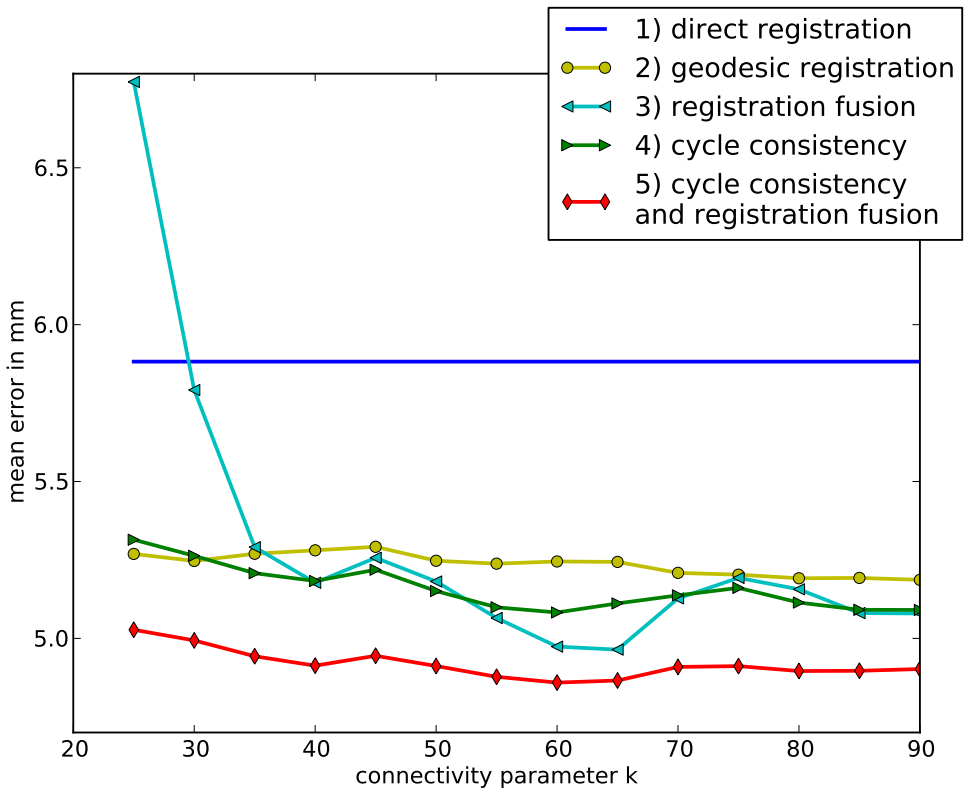


Figure 6.4: Mean registration error in mm for values of k, where the number of paths $\mathcal{K} = 3$

Method Comparison with Respect to Varying Connectivity k

Parameter $\mathcal{K}$ is fixed to 3 for these analyses, these conclusions generalise to other values of $\mathcal{K}$. Parameter $\mathcal{K}$ is necessary for registration fusion (method 3) and cycle consistency and registration fusion (method 5).

The connectivity parameter of the graph, k, determines the sparsity of graph G by specifying the number of edges connected to each vertex. At most $k \times |V|$ registrations are required but frequently many vertices have common nearest neighbours and so in practise the number of registrations required is much lower. If the selected k is very large there is an increased probability that a geodesic registration between two subjects will be equivalent to a direct registration, since the shortest path becomes more likely to be a single edge. However, k needs to be suitably large such that unnecessarily long paths incorporating many edges and therefore many small errors can be avoided.

Results for affine registration for the methods proposed are presented in figure 6.4 for vary-

ing parameter k. It can be observed that direct registration has a higher mean error when compared to all of the methods which use geodesic registration as a basis (methods 2-5). Direct registration is not dependant upon k. When the connectivity of the graph $k \geqslant 35$, the geodesic registration with registration fusion (method 3) has lower mean error when compared to geodesic registration (method 2). Geodesic registration with cycle consistency (method 4) has slightly lower mean error than standard geodesic registration (method 2) when the connectivity of the graph $k \geqslant 30$. We found that using geodesic registration with cycle consistency and registration fusion (method 5) has the lowest mean error for all k.

Geodesic registration with registration fusion (method 3) is very sensitive to parameter k. However, when registration fusion is used with cycle consistency (method 5), the results are less sensitive to parameter k. This supports the hypothesis that image similarity is not suitable for the edge weights. Since when multiple paths are selected bad edges are included because the weightings are not appropriate. When cycle consistency is used as edge weights, more reliable paths are selected and the error is more consistent over changes in k.

The standard deviation of the error for methods over this range was also observed and similar trends were noted, with fluctuations over varying k for registration fusion (method 3). Additionally, when geodesic registration is combined with cycle consistency and registration fusion (method 5) the standard deviation of the error is consistent over all k. Also the standard deviation of error over all k is lower than all other methods proposed.

Having observed the effect of k upon the methods, subsequent analysis will fix k=60, where a strong performance is observed for all methods.

Method Comparison with Respect to Varying Paths $\mathcal{K}$

The number of $\mathcal{K}$-shortest paths for the *registration fusion* is an important parameter that needs to be selected. Figure 6.5 shows the effect of varying $\mathcal{K}$ upon the mean error and computations time. As can be seen from figure 6.5 (a), the mean error is reduced dramatically with the first 3 paths but the gain tails off. The mean computational time is measured on one core of an Intel Core2 Duo E8400 3.00GHz with data stored on a network file system. It can be seen from

Affine registration algorithm	Mean Euclidean distance [mean(sd)] (mm)
Method 1	5.88 (3.46)
Method 2	5.25 (1.84)
Method 3	5.08 (1.71)
Method 4	4.97 (1.66)
Method 5	4.86 (1.59)

Table 6.1: Statistics of errors (mean and standard deviation) in mm where $k = 60$

6.5 (b) that computational cost grows linearly with the number of paths. A multi-threaded CPU implementation of the software would render this constant time if the number of paths is less than or equal to the number of threads.

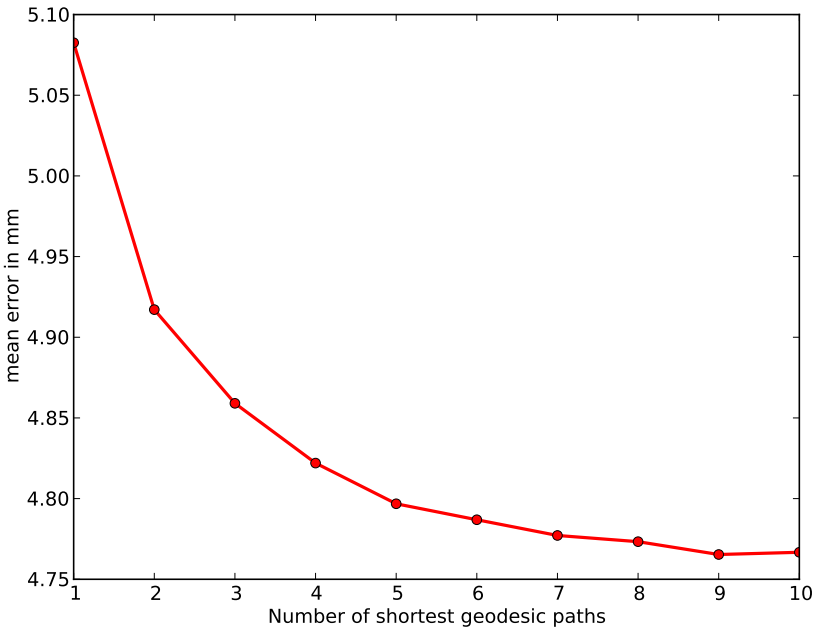
Method Comparison with fixed k and $\mathcal{K}$

A summary of results for $k = 60$ are presented in table 6.1. It can be seen that using the geodesic registration with cycle consistency and registration fusion (method 5) , has lower mean and standard deviations than other methods. Figure 6.6 compares the distribution of errors for different registration methods. Figure 6.6 (a) shows geodesic registration with cycle consistency and registration fusion (method 5) compared to direct registration (method 1) it can be seen that outliers are significantly reduced by using method 5. This demonstrates an increase in robustness. A more subtle reduction in outliers is observed in figure 6.6 (b).

Figure 6.7 visualises how the error of pairwise registrations changes with two registration methods, namely direct registration versus geodesic registration with cycle consistency and registration fusion. Each point represents a registration between a pair of images and its position represents the error for both methods. If a point lies upon the black line, it can be seen that this registration has not changed despite using different methods. We draw the readers attention to registrations with high errors using direct registration that are improved using geodesic registration with cycle consistency and registration fusion. It can be observed in a minority of cases that geodesic registration with cycle consistency and registration fusion increases the error of the registration by up to 5mm. Geodesic registration with cycle consistency and registration

Figure 6.5: Graphs illustrating the performance in terms of computational speed and accuracy for the method presented in this chapter (Geodesic registration with cycle consistency and registration fusion)

(a) Mean registration error in mm for number of paths selected in *registration fusion*.



(b) Computational time cost in seconds for number of paths selected in *registration fusion*.

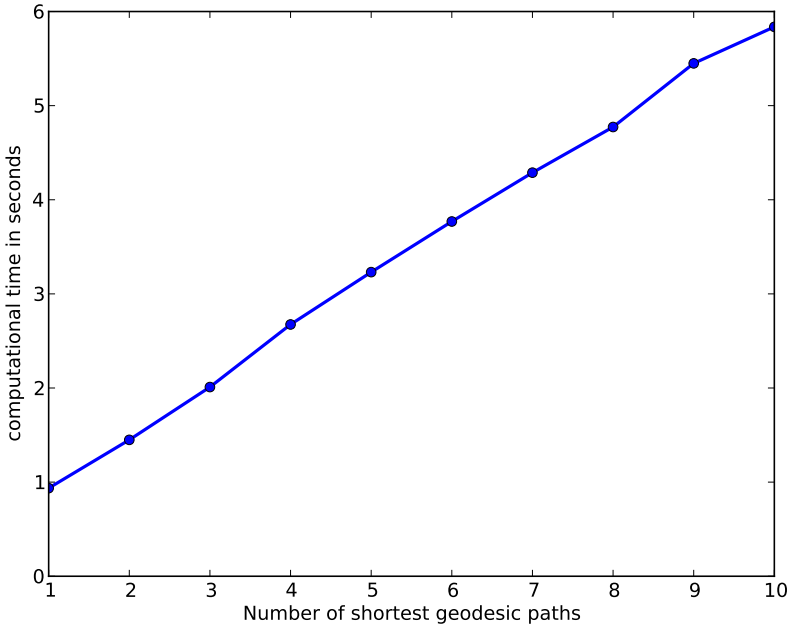
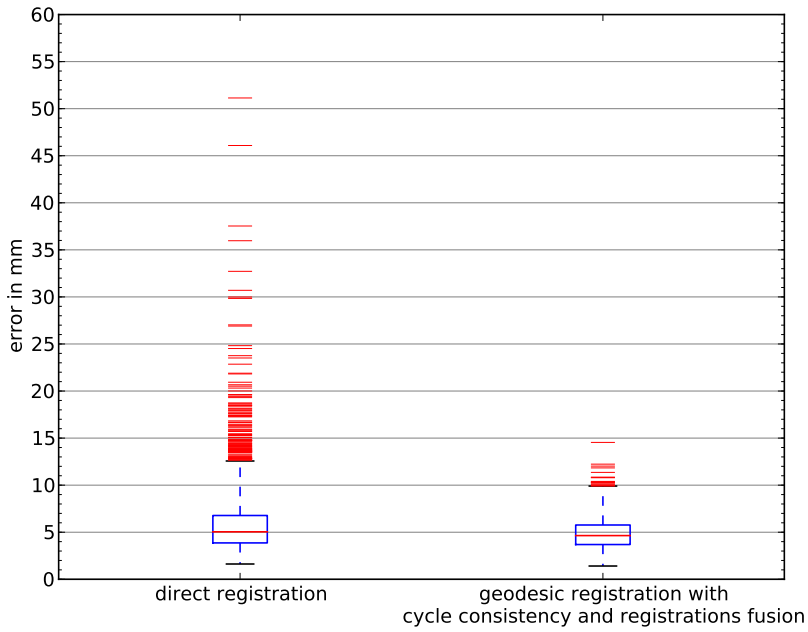
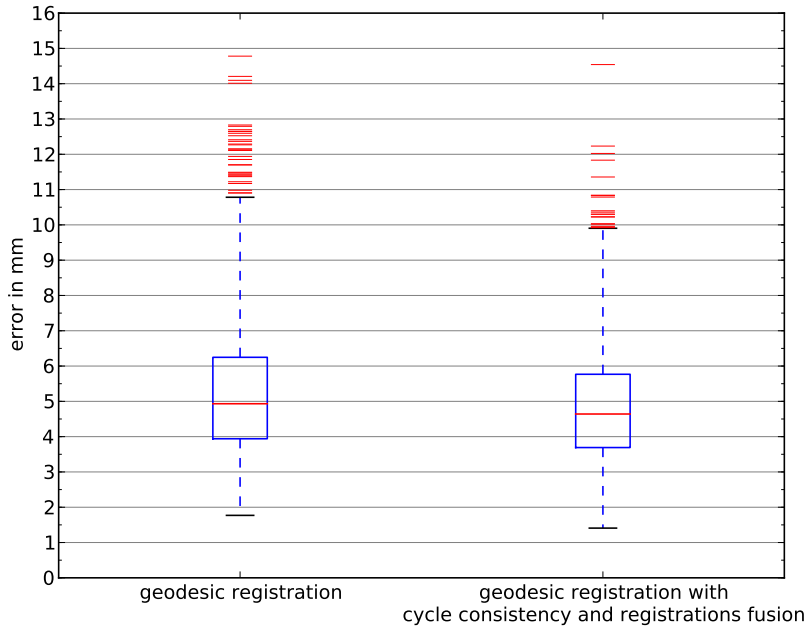


Figure 6.6: These box plots show the distribution of errors for 3 registration methods, where $n = 2743$. Note the scales on these plots are different to illustrate the vast improvement of our method upon direct registration but more subtle improvement upon geodesic registration.

(a) This box plot compares direct registration with our method.



(b) This box plot compares geodesic registration to our method.



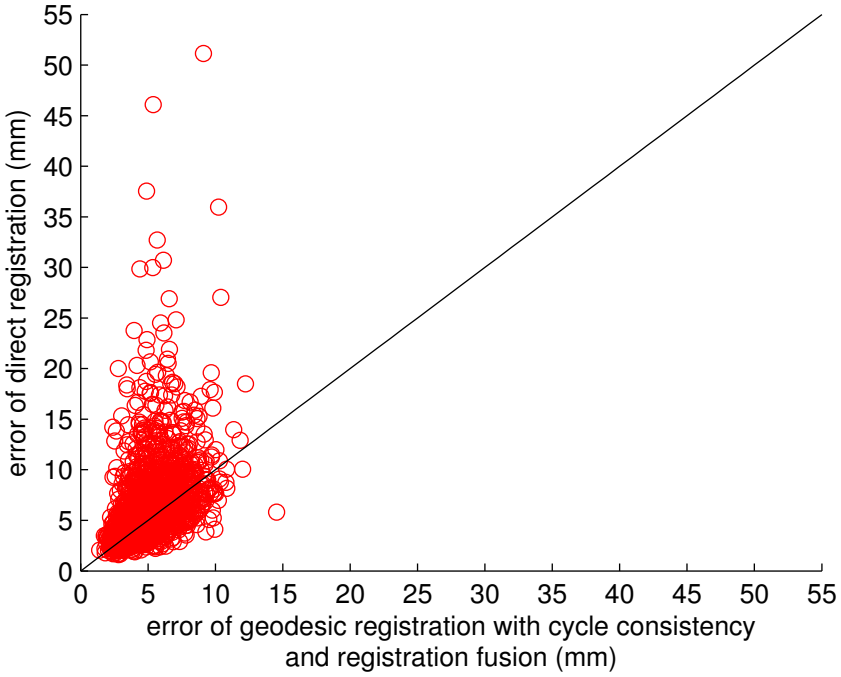


Figure 6.7: Pairwise comparison of registrations methods; direct registration versus the method proposed here, registrations to the upper-left of the line have reduced error with geodesic registration with cycle consistency and registration fusion, registrations to the lower-right have reduced error using direct registration.

fusion drastically reduces the number of registrations which have registration error greater than 1cm.

A contribution in chapter 5 was to propose the refinement step. However, some differences exist between this chapter and chapter 5, making it inappropriate to compare results directly between chapters. We test the efficacy of the refinement step proposed in chapter 5 by applying it to geodesic registration with cycle consistency and registration fusion (method 5). When applying the refinement step to these transformations, we find the accuracy on average degrades, the distribution of these results is presented in a box plot in figure 6.8

A Qualitative example of registration accuracy improvement is visualised in figure 6.9.

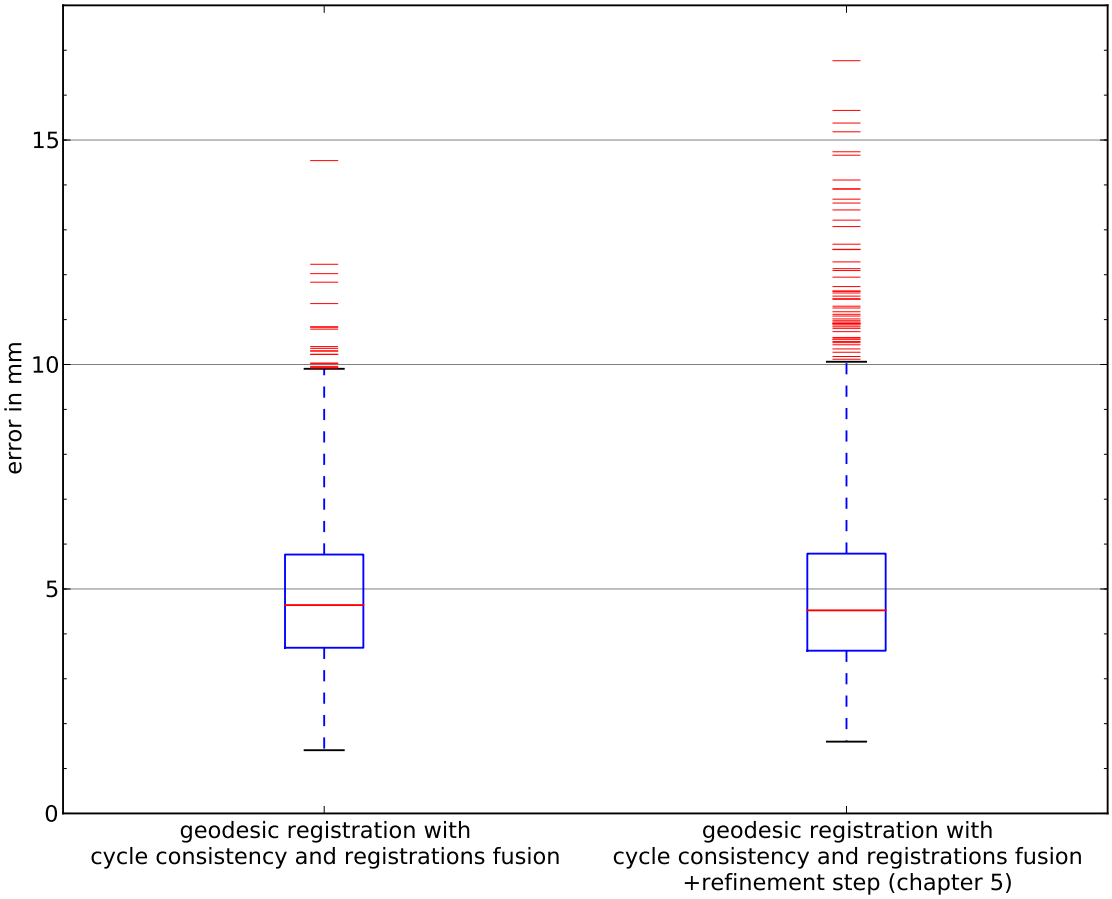


Figure 6.8: Box plot of the presented method without (left) and with (right) a refinement step. The results from the algorithm presented in this chapter are considered to be an approximation of an affine registration. The refinement step (as presented in chapter 5) is a further affine registration after the registration algorithm with the expectation of improving the accuracy. It can be seen that the accuracy decreases after the refinement, this suggests that a refinement step after the method is of no value.

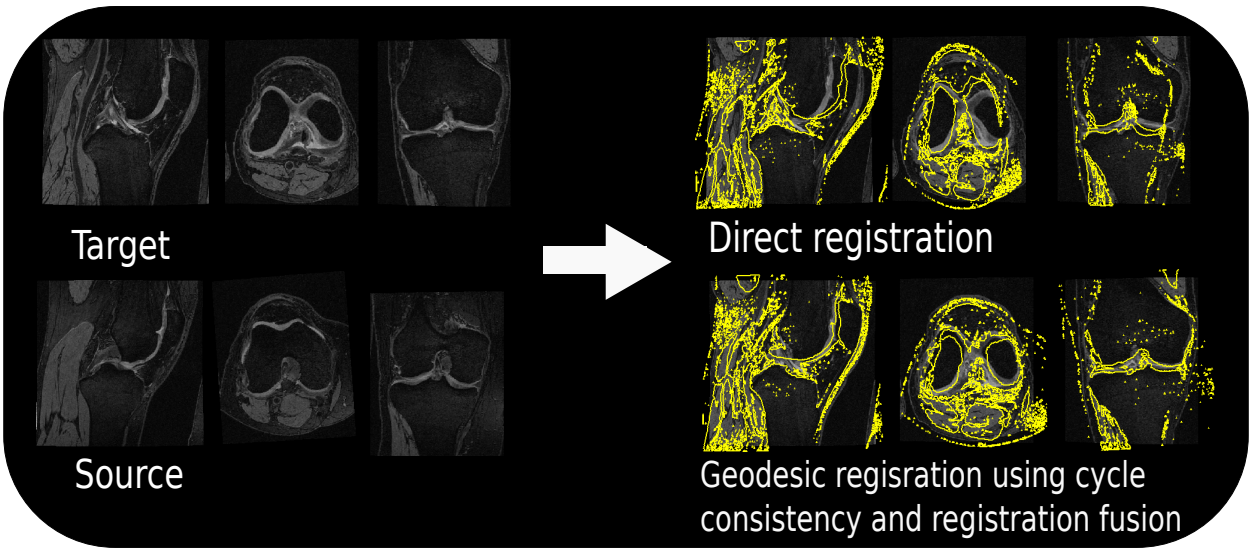


Figure 6.9: Example of registration using direct registration and geodesic registration with cycle consistency and registration fusion. The images on the left are the target and source to be registered. Images on the right show the results of registration with the contours of the transformed source overlaid on the target.

6.4 Discussion

In this chapter cycle consistency is used as an approach to infer a more reliable distance metric to weight the quality of a transformation connecting any two vertices (or images) in the empirical manifold. When a single shortest path is found in the empirical manifold, the transformations on the path traversed should be of higher quality and therefore able to perform a more meaningful global registration of a pair of images. With this approach alone, we show that the registration error reduces when compared to the standard geodesic registration model where a normalised cross correlation based metric is used instead. Direct affine registration is an optimisation which aims to minimise normalised cross correlation to find a suitable alignment. In the case where the registration gets trapped in a local minimum, the similarity metric between the images will be high without there being an accurate registration, which suggests a normalised cross correlation based metric will not perform well. We present evidence to suggest that assuming global cycle consistency across the empirical manifold to rate transformation reliability appears to be effective in identifying strong and weak transformations in the graph.

Fusing multiple shortest paths on the graph of affine transformations enables voting between

many successful registration estimates. It appears to have a degree of success using normalised cross correlation as a similarity metric to weight the edges in the graph. Although the performance of this approach appears to be unstable across different experiments where the connectivity parameter k is varied. For some k the mean error is much worse than similar graphical registrations approaches.

When both cycle consistency and registration fusion are combined both the mean and standard deviation of the error is reduced. Moreover, we observe the error to be stable over varying k. Geodesic registration with cycle consistency and registration fusion performs better than each approach in isolation. We find that the mean error is reduced by 1.0mm and the is reduced by 1.9mm when compared with direct registration this is because outliers are significantly reduced. It can be seen in figure 6.6 that very few registrations computed using a geodesic registration with cycle consistency and registration fusion have registration errors greater than 1cm.

6.4.1 Computational Cost

The graph used in this work is well sampled with 2743 images. However, it is impractical to work with a fully connected graph. The work presented in Datteri et al. (2011) uses cycle consistency as a measure of registration quality for multi-atlas segmentation on a fully connected graph but experiments are limited to 10 subjects. The work presented here demonstrates that the techniques presented here can be applied to a large, sparse k-nn graph.

The number of registrations required to compute the large, sparse graph offline is large. In the case of the experiment for the 60-nearest-neighbour graph with 2743 vertices there are approximately 125,000 unique registrations that need to be computed. Whilst this may seem excessive, it should be remembered that the method will yield results with improved accuracy upon direct registration for any pairwise registration. To compute all pairwise registration using the direct registration approach approximately 3,750,000 unique registrations would be computed.

The method presented here is fast for online computation of registrations. Geodesic registra-

tions with cycle consistency and registration fusion can be computed in less than 2 seconds on one core of an Intel Core2 Duo E8400 3.00GHz with data stored on a network file system. This compares favourably to direct registration which takes approximately 150 seconds. If registration fusion is not used, geodesic registration with or without cycle consistency can be computed in approximately 1 second. Whilst geodesic registrations using registration fusion provides more accurate results, the speed of computation is slower. Therefore the preferred approach is likely to vary depending whether the application requires high accuracy or fast online computation.

6.5 Future Work

We present and demonstrate a scalable algorithm for geodesic affine registration because affine registration has been shown to not be sufficiently robust when applied to the knee. For other anatomical structures, such as the brain, affine registration is a robust solution however, non-rigid registration is not. Future work could be embarked upon to extend this work for a non-rigid geodesic registration model. The affine geodesic registration technique in this chapter serves as a proof of concept to demonstrate that registrations can be improved through offline learning and fusion. In addition the pairwise registrations computed here could be used as an initialisation to non-rigid registrations across the graph.

It is possible to weight the contribution of the transformation matrices when computing the Frechet mean. The $\mathcal{K}$ shortest paths each have a different associated weight which reflects the quality of the components in the composed affine transformation. The transformations could be weighted accordingly when computing the average of these. This extension takes inspiration from the multi atlas based segmentation literature, where weighted fusion is employed (Artaechevarria et al. (2009)).

The accurate registrations computed in this work can be used as an input for further studies. This includes increasing registration accuracy for atlas based segmentation or biomarker discovery.

Chapter 7

Conclusion

7.1 Contributions

The contributions of this thesis can be broadly separated into two categories; discovering diagnostic biomarkers of Osteoarthritis from knee MRI and developing a computationally efficient method for robust intensity based registration for knee MRI.

7.1.1 Diagnostic Biomarkers

We have introduced novel fully automated algorithms, based upon manifold learning, which differ significantly from current state of the art methodologies. Typically, hand-crafted quantitative morphological measures are computed from segmentations (Dam et al. (2007); Eckstein et al. (2009)) or observed by a radiologist (Hunter et al. (2008); Peterfy et al. (2004)). These morphological measures are selected and designed to reflect the current understanding of the progression of OA. However, it is recognised that OA is a complex disease (Aspden (2008); Conaghan et al. (2006)) which is yet to be fully understood. These approaches could omit subtle but key features in the MR images. Cartilage morphology metrics computed from segmentations have been shown to differ significantly for different segmentations teams (Schneider et al. (2012)), due to this limitation the author recommends readers do not pool results from different teams.

To tackle these issues we introduce data-driven algorithms which learn features that best predict OA. These features are learnt from a very large dataset of healthy knee MR and MR of knees at various stages of OA. In this thesis diagnostic biomarkers are learnt from the appearance of the articular cartilage in MR. Automation removes errors which can be attributed to human-readers of images and yields results which are reproducible and makes large scale studies feasible. The algorithms presented avoid the need for an accurate segmentation. This is beneficial since segmentation is a common source of errors especially when studying the articular cartilage which is a challenging structure to segment. Moreover, the segmentation becomes more challenging for subjects with more severe OA.

The diagnostic biomarkers presented have been shown to provide superior AUC results compared with state of the art algorithms (Dam et al. (2009)) (chapter 4). It has also been shown that there is a strong correlation between clinical variables and the manifold embedding (chapter 3). The algorithm presented has potential for applications in a clinical setting, for drug discovery and as a research tool.

7.1.2 Robust and Efficient Affine registration of the Knee

We identified that intensity based affine registration of the knee is not sufficiently robust for large scale image analysis. On average, approximately 5% of affine registrations between a pair of knee MRI fail catastrophically. We developed fast algorithms to compute robust affine transformations of the knee, taking inspiration from Serag et al. (2012); Hamm et al. (2009, 2010); Jia et al. (2010, 2011, 2012).

We model the population of images as a non-linear manifold, where the similarity is defined by a prediction of registration success. We compute geodesic shortest paths on the graph to guide registration which increases robustness (chapter 5). We also identify sources of error for which we seek to find fast mitigation strategies using cycle consistency and registration fusion (chapter 6).

The method proposed here has been demonstrated to offer improved accuracy relative to a

direct registration between a pair of images. We find that the mean error is reduced by 1.02mm and the standard deviation is reduced by 1.87mm. We observe the number of outliers to be significantly decreased. The algorithm computes a registration between a pair of images in under 2 seconds. The method we present has been demonstrated to perform well on a modern large-scale dataset. In addition we believe that performance is enhanced due to the dense sampling of subjects on the manifold.

7.2 Future Work

At present the algorithms lend themselves to large scale clinical trials, where a large cohort is recruited and analysis to assess the efficacy of DMOADS is required. However, for a clinical application it is useful to provide a rapid computer aided diagnosis for a subject. Currently computing robust affine registration and diagnostic biomarkers for previously unseen subjects requires significant computation. This limitation can be addressed with out-of-sample solutions which exist on a theoretical basis (Bengio et al. (2004)) but would need to be incorporated and applied to the problem presented.

All algorithms presented are data driven and rely upon pairwise similarity metrics. It was found that cross correlation provided a good metric, both for registration accuracy and biomarker discovery. Cross correlation performed well because it was sufficient to assume a linear relationship between voxel intensities as all images are acquired as single sequences. However, if the similarity metric could be learnt for the specific task, performance may be improved. This might be achieved using recent research in sparse learning, which has been used to learn regions of important voxels (Janousova et al. (2012)) or by using Adaboost to boost feature comparison for distance metric learning (Yang et al. (2010)). It is likely that a different similarity metric might be required to compute robust affine registration and diagnostic biomarkers.

7.2.1 Diagnostic Biomarkers

The ground truth selected to validate the biomarkers presented was K&L grades. K&L grades are a crude and coarse way to define structural disease status. To improve upon K&L grades it would be beneficial to validate biomarkers against semi-quantitative metrics of disease severity which have greater granularity, such as BLOKS or WORMS (Hunter et al. (2008); Peterfy et al. (2004)).

Diagnostic biomarkers are a useful tool to aid understanding of the current pathological status of a subject. This can be complemented by effective prediction of outcome or prognosis. An accurate prognosis may lead to early surgical or non-invasive interventions. Bauer et al. (2006) suggests a framework for prognostic biomarker specification and validation. Recently Eckstein et al. (2013) has published a case-control study predicting joint replacement in patients using cartilage loss as a biomarker of OA progression. The outcome data used here has very recently been released by the OAI. A study comparing the manifold learning biomarkers presented in this thesis with cartilage loss would be interesting.

Understanding pain is an important motivation in Osteoarthritis research. It is challenging to find a relationship between function and structure but achieving some understanding would be a large contribution to the field. Addressing this with data driven approaches is more likely to yield positive results than hand-crafted solutions because of the complexity of the relationship.

7.2.2 Robust and Efficient Affine registration of the Knee

Pairwise affine transformations for this large sample of subjects from the OAI can be computed very quickly and accurately. Using this data it would be possible to build an unbiased atlas which establishes a common space for comparisons of knee MRI. Such a facility already exists for the brain MRI and is referred to as MNI standard space (Evans et al. (1993)). Moreover, it is possible to make pairwise comparisons of subjects in their native co-ordinate space.

It would be valuable to extend this work to include non-rigid registration in the framework. Pairwise affine registrations can currently be established for the entire population. Image pairs

which are similar after the affine transformation has been applied have an increased likelihood of a successful non-rigid registration. This non-rigid graph based algorithm would take further inspiration from Serag et al. (2012); Hamm et al. (2009, 2010); Jia et al. (2010, 2011, 2012). One could also explore adopting cycle consistency and registration fusion for a non-rigid version. One could consider adopting a B-spline non-rigid transformation model or a polyaffine model. The applications of this would be two-fold, firstly as an input for finding a transformation between any pair of images which would be useful for discovering biomarkers (as in chapter 4) or as a fast initialisation for an automated segmentation algorithm.

Chapter 8

Publications

Work presented in this thesis, has in part, been presented in:

C. R. Donoghue, A. Rao, A. M. J. Bull, and D. Rueckert. Learning Osteoarthritis Imaging Biomarkers Using Laplacian Eigenmap Embeddings with Data from the OAI. *IEEE International Symposium on Biomedical Imaging*, 2014.

C. R. Donoghue, A. Rao, A. M. J. Bull, and D. Rueckert. Fast and Accurate Global Geodesic Registrations using Knee MRI from the Osteoarthritis Initiative. *Medical Computer Vision Workshop (MCV) at Computer Vision and Pattern Recognition (CVPR)*, 2012

C.R. Donoghue, A. Rao, A.M.J. Bull, and D. Rueckert. Robust Global Registration through Geodesic Paths on an Empirical Manifold with Knee MRI from the Osteoarthritis Initiative (OAI). *Biomedical Image Registration*, 7359:1-10, 2012.

(prize won : young investigators award) C. Donoghue, A. Rao, A. M. J. Bull, and D. Rueckert. Automatically Generated Novel Diagnostic Imaging Biomarkers with Data from the OAI. *4th Imaging Workshop for Osteoarthritis*, 2011.

C. Donoghue, A. Rao, A. M. J. Bull, and D. Rueckert. Manifold learning for automatically predicting articular cartilage morphology in the knee with data from the osteoarthritis initiative (OAI). *SPIE Medical Imaging 2011 : Image Processing*, 7962, 2011.

C. Donoghue, A. Rao, A. M. J. Bull, and D. Rueckert. Articular Cartilage as an Automatic Predictor for Cartilage Morphology with Data from the Osteoarthritis Initiative. *4th International Workshop on Imaging Based Measures of Osteoarthritis*, 2010.

(2nd prize) C. Donoghue, A. Rao, A. M. J. Bull, and D. Rueckert. Articular Cartilage as an Automatic Predictor for Cartilage Morphology with Data from the Osteoarthritis Initiative. *Poster Competition Research Students' Summer Symposium 2010, Graduate School of Engineering and Physical Sciences, Imperial College*, 2010.

Bibliography

R. Adams and L. Bischof. Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6):641–647, 1994.

M. Alexa. Linear combination of transformations. *ACM Transactions Graph.*, 21(3):380–387, 2002.

P. Aljabar, K. Bhatia, M. Murgasova, J. Hajnal, J. Boardman, L. Srinivasan, M. Rutherford, L. Dyet, A. Edwards, and D. Rueckert. Assessment of brain growth in early childhood using deformation-based morphometry. *NeuroImage*, 39(1):348–358, 2008.

P. Aljabar, R. Heckemann, A. Hammers, J. Hajnal, and D. Rueckert. Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *NeuroImage*, 46(3):726, 2009.

P. Aljabar, R. Wolz, L. Srinivasan, S. Counsell, M. Rutherford, A. Edwards, J. Hajnal, and D. Rueckert. A combined manifold learning analysis of shape and appearance to characterize neonatal brain development. *IEEE Transactions on Medical Imaging*, 30 (12):2072–1086, 2011.

R. Altman and G. Gold. Radiographic Atlas for Osteoarthritis of the Hand, Hip and Knee. *Osteoarthritis and Cartilage*, 15, Supplement 1:0, 2007.

S. Amin, J. Niu, A. Guermazi, M. Grigoryan, D. J. Hunter, M. Clancy, M. P. LaValley, H. K. Genant, and D. T. Felson. Cigarette smoking and the risk for cartilage loss and knee pain in men with knee osteoarthritis. *Annals of the Rheumatic Diseases*, 66(1):18–22, 2007.

V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Log-Euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine*, 56(2):411–421, 2006.

X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de Solorzano. Combination strategies in multi-atlas image segmentation: Application to brain MR data. *IEEE Transactions on Medical Imaging*, 28(8):1266–1277, 2009.

Arthritis Care. *The OANation 2012 Survey.* iYouGov (May 2012), 2012.

R. M. Aspden. Is Generalized Osteoarthritis a Problem of Growth not Decay? 2nd Annual Workshop on Imaging Based Measures of Osteoarthritis. (In Proceedings) 2nd Annual Workshop on Imaging Based Measures of Osteoarthritis, 2008.

R. Bajcsy and S. Kovačič. Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, 46(1):1–21, 1989.

D. Bauer, D. Hunter, S. Abramson, M. Attur, M. Corr, D. Felson, D. Heinegard, J. Jordan, T. Kepler, and N. Lane. Classification of osteoarthritis biomarkers: a proposed approach1. *Osteoarthritis and Cartilage*, 14(8):723–727, 2006.

M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. *International Journal Computer Vision*, 61(2):139–157, 2005.

M. Belkin and P. Niyogi. Using Manifold Structure for Partially Labelled Classification. (In Proceedings) NIPS, page 929, 2002.

M. Belkin and P. Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, 15 (6):1373–1396, 2003.

Y. Bengio, J.-F. Paiement, P. Vincent, O. Delalleau, N. Le Roux, and M. Ouimet. Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering. *Advances in neural information processing systems*, 16:177–184, 2004.

J. L. Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9):509–517, 1975.

Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical pharmacology and therapeutics*, 69(3):89–95, 2001.

J. Bonica. Editorial The need of a taxonomy. *Pain*, 6(3):247–248, 1979.

B. E. Boser, I. M. Guyon, and V. N. Vapnik. A training algorithm for optimal margin classifiers. (In Proceedings) Proceedings of the fifth annual workshop on Computational learning theory, COLT '92, 144–152, ACM, New York, NY, USA, 1992.

L. Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

R. Buck, B. Wyman, M. H. L. Graverand, D. Hunter, E. Vignon, W. Wirth, and F. Eckstein. Using ordered values of subregional cartilage thickness change increases sensitivity in detecting risk factors for osteoarthritis progression. *Osteoarthritis and Cartilage*, 19(3):302–308, 2011.

R. Buck, B. Wyman, M.-P. H. L. Graverand, M. Hudelmaier, W. Wirth, and F. Eckstein. Osteoarthritis may not be a one-way-road of cartilage loss - comparison of spatial patterns of cartilage change between osteoarthritic and healthy knees. *Osteoarthritis and Cartilage*, 18(3):329–335, 2010.

R. J. Buck, B. T. Wyman, M. pierre hellio Le Graverand, M. Hudelmaier, W. Wirth, F. Eckstein, and A. Investigators. Does the use of ordered values of subregional change in cartilage thickness improve the detection of disease progression in longitudinal studies of osteoarthritis? *Arthritis & Rheumatism*, 61(7):917–924, 2009.

J. Carballido-Gamio and S. Majumdar. Atlas-based knee cartilage assessment. *Magnetic Resonance in Medicine*, 66(2):575–581, 2011.

L. Cayton. Algorithms for Manifold Learning. 2005.

G. Christensen and H. Johnson. Consistent image registration. *IEEE Transactions on Medical Imaging*, 20(7):568 –582, 2001.

G. Christensen, R. Rabbit, and M. Miller. Deformable Templates Using Large Deformation Kinematics. *IEEE Transactions on Image Processing*, 5(10):1435–1447, 1996.

F. Chung. *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92)*. American Mathematical Society, 1997.

A. G. Clark, J. M. Jordan, V. Vilim, J. B. Renner, A. D. Dragomir, G. Luta, and V. B. Kraus. Serum cartilage oligomeric matrix protein reflects osteoarthritis presence and severity: the Johnston County Osteoarthritis Project. *Arthritis & Rheumatism*, 42(11):2356–2364, 1999.

T. Coleman and Y. Li. A Reflective Newton Method for Minimizing a Quadratic Function Subject to Bounds on Some of the Variables. *SIAM Journal on Optimization*, 6(4):1040–1058, 1996.

P. G. Conaghan, D. Felson, G. Gold, S. Lohmander, S. Totterman, and R. Altman. MRI and non-cartilaginous structures in knee osterarthritis. *Osteoarthritis and Cartilage*, 14 (1):87–94, 2006.

T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1):21–27, 1967.

T. F. Cox and M. A. A. Cox. *Multidimensional Scaling, Second Edition.* Chapman and Hall/CRC, 2nd edition, 2000.

W. R. Crum, T. Hartkens, and D. L. G. Hill. Non-rigid image registration: theory and practice. *The British journal of radiology*, 77 Spec No 2:S140–153, 2004.

W. R. Crum, C. Tanner, and D. J. Hawkes. Anisotropic multi-scale fluid registration: evaluation in magnetic resonance breast imaging. *Physics in Medicine and Biology*, 50(21):5153–5174, 2005.

E. Dam, J. Folkesson, P. Pettersen, and C. Christiansen. Automatic morphometric cartilage quantification in the medial tibial plateau from MRI for osteoarthritis grading. *Osteoarthritis and Cartilage*, 15(7):808–818, 2007.

E. Dam and M. Loog. Efficient Segmentation by Sparse Pixel Classification. *IEEE Transactions on Medical Imaging*, 27(10):1525–1534, 2008.

E. B. Dam, M. Loog, C. Christiansen, I. Byrjalsen, J. Folkesson, M. Nielsen, A. A. Qazi, P. C. Pettersen, P. Garnero, and M. A. Karsdal. Identification of progressors in osteoarthritis by

combining biochemical and MRI-based markers. *Arthritis Research & Therapy*, 11(4):R115, 2009.

R. D. Datteri, A. J. Asman, B. A. Landman, and B. M. Dawant. Estimation of Registration Accuracy Applied to Multi-Atlas Segmentation. (In Proceedings) Proceedings of the 2011 MICCAI Workshop on Multi-Atlas Labeling and Statistical Fusion, 78–87, 2011.

B. Dawant, S. Hartmann, J.-P. Thirion, F. Maes, D. Vandermeulen, and P. Demaerel. Automatic 3-D segmentation of internal structures of the head in MR images using a combination of similarity and free-form transformations. I. Methodology and validation on normal subjects. *IEEE Transactions on Medical Imaging*, 18(10):909 –916, 1999.

J. DeGroot, N. Verzijl, M. J. G. Wenting-van Wijk, K. M. G. Jacobs, B. Van El, P. M. Van Roermund, R. A. Bank, J. W. J. Bijlsma, J. M. TeKoppele, and F. P. J. G. Lafeber. Accumulation of advanced glycation end products as a molecular mechanism for aging as a risk factor in osteoarthritis. *Arthritis & Rheumatism*, 50(4):1207–1215, 2004.

Department for Work and Pensions. Disability Living Allowance - parameters: all entitled cases Caseload (Thousands) : Main Disabling Condition by Region (Time Series=FEB12). online, 2012. [accessed on 23rd October 2012].

E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1(1):269–271, 1959.

R. L. Drake, A. W. Vogl, and A. W. M. M. M. B. FRCS. *Gray's Anatomy for Students*. Churchill Livingstone, 2 edition, 2005.

D. D. Dunlop, L. M. Manheim, E. H. Yelin, J. Song, and R. W. Chang. The costs of arthritis. *Arthritis Care & Research*, 49(1):101–113, 2003.

F. Eckstein, G. Ateshian, R. Burgkart, D. Burstein, F. Cicuttini, B. Dardzinski, M. Gray, T. M. Link, S. Majumdar, T. Mosher, C. Peterfy, S. Totterman, J. Waterton, C. S. Winalski, and D. Felson. Proposal for a nomenclature for magnetic resonance imaging based measures of articular cartilage in osteoarthritis. *Osteoarthritis and Cartilage*, 14 (10):974–983, 2006a.

F. Eckstein, M. P. H. L. Graverand, H. C. Charles, D. J. Hunter, V. B. Kraus, T. Sunyer, O. Nemirovskyi, B. T. Wyman, R. Buck, and investigators for the A9001140. Clinical, radiographic, molecular and MRI-based predictors of cartilage loss in knee osteoarthritis. *Annals of the Rheumatic Diseases*, 70(7):1223–1230, 2011.

F. Eckstein, M. Hudelmaier, W. Wirth, B. Kiefer, R. Jackson, J. Yu, C. B. Eaton, and E. Schneider. Double echo steady state magnetic resonance imaging of knee articular cartilage at 3 Tesla: a pilot study for the Osteoarthritis Initiative. *Annals of the Rheumatic Diseases*, 65(4):433–441, 2006b.

F. Eckstein, C. K. Kwoh, R. M. Boudreau, Z. Wang, M. J. Hannon, S. Cotofana, M. I. Hudelmaier, W. Wirth, A. Guermazi, M. C. Nevitt, M. R. John, D. J. Hunter, and for the OAI investigators. Quantitative MRI measures of cartilage predict knee replacement: a case-control study from the Osteoarthritis Initiative. *Annals of the Rheumatic Diseases*, 72(5):707–714, 2013.

F. Eckstein, S. Maschek, W. Wirth, M. Hudelmaier, W. Hitzl, B. Wyman, M. Nevitt, M. H. Le Graverand, and the OAI Investigator Group. One year change of knee cartilage morphology in the first release of participants from the Osteoarthritis Initiative progression subcohort: association with sex, body mass index, symptoms and radiographic osteoarthritis status. *Annals of the Rheumatic Diseases*, 65 (5):674–679, 2009.

A. Evans, D. Collins, S. R. Mills, E. D. Brown, R. L. Kelly, and T. Peters. 3D statistical neuroanatomical models from 305 MRI volumes. (In Proceedings) IEEE Nuclear Science Symposium and Medical Imaging Conference, volume 3, 1813–1817, 1993.

T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006.

D. T. Felson, R. C. Lawrence, P. A. Dieppe, R. Hirsch, C. G. Helmick, J. M. Jordan, R. S. Kington, N. E. Lane, M. C. Nevitt, Y. Zhang, M. Sowers, T. McAlindon, T. D. Spector, A. R. Poole, S. Z. Yanovski, G. Ateshian, L. Sharma, J. A. Buckwalter, K. D. Brandt, and J. F.

Fries. Osteoarthritis: New Insights. Part 1: The Disease and Its Risk Factors. *Annals of Internal Medicine*, 133(8):635–646, 2000.

D. T. Felson, J. Niu, A. Guermazi, B. Sack, and P. Aliabadi. Defining radiographic incidence and progression of knee osteoarthritis: suggested modifications of the Kellgren and Lawrence scale. *Annals of the Rheumatic Diseases*, 70(11):1884–1886, 2011.

R. A. Fisher. The use of multiple measurement in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936.

J. M. Fitzpatrick. *Medical Image Registration*, chapter Detecting Failure, Assessing Success, 117–139. CRC Press, 2001.

R. W. Floyd. Algorithm 97: Shortest path. *Commun. ACM*, 5(6):345, 1962.

J. Folkesson, E. Dam, O. Olsen, P. Pettersen, and C. Christiansen. Segmenting Articular Cartilage Automatically Using a Voxel Classification Approach. *IEEE Transactions on Medical Imaging*, 26(1):106–115, 2007.

J. Folkesson, E. B. Dam, O. F. Olsen, M. A. Karsdal, P. C. Pettersen, and C. Christiansen. Automatic quantification of local and global articular cartilage surface curvature: Biomarkers for osteoarthritis? *Magnetic Resonance in Medicine*, 59(6):1340–1346, 2008.

M. Fréchet. *L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application a la moyenne d'un élément aléatoire de nature quelconque.* La Revue Scientifique, 1944.

Y. Freund and R. E. Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1):119 – 139, 1997.

J. Fripp, S. Crozier, S. K. Warfield, and S. Ourselin. Automatic Segmentation and Quantitative Analysis of the Articular Cartilages From Magnetic Resonance Images of the Knee. *IEEE Transactions on Medical Imaging*, 29(1):55–64, 2010.

P. Garnero. Use of biochemical markers to study and follow patients with osteoarthritis. *Current rheumatology reports*, 8(1):37–44, 2006.

S. Gerber, T. Tasdizen, S. Joshi, and R. Whitaker. On the manifold structure of the space of brain images. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 12(1):305–312, 2009.

J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov. Neighbourhood Components Analysis. In L. K. Saul, Y. Weiss, and L. Bottou (editors), *Advances in Neural Information Processing Systems 17*, 513–520. MIT Press, Cambridge, MA, 2005.

H. Graichen, V. Eisenhart-Rothe, Rüdiger, T. Vogl, K.-H. Englmeier, and F. Eckstein. Quantitative assessment of cartilage status in osteoarthritis by quantitative magnetic resonance imaging: Technical validation for use in analysis of cartilage volume and further morphologic parameters. *Arthritis & Rheumatism*, 50(3):811–816, 2004.

V. Grau, A. Mewes, M. Alcaniz, R. Kikinis, and S. Warfield. Improved Watershed Transform for Medical Image Segmentation Using Prior Information. *IEEE Transactions on Medical Imaging*, 23(4):447–458, 2004.

J. V. Hajnal and D. L. G. Hill. *Medical Image Registration*. CRC Press, 2010.

J. V. Hajnal, N. Saeed, A. Oatridge, E. J. Williams, I. R. Young, and G. M. Bydder. Detection of subtle brain changes using subvoxel registration and subtraction of serial MR images. *Journal of computer assisted tomography*, 19(5):677–691, 1995.

J. Hamm, C. Davatzikos, and R. Verma. Efficient Large Deformation Registration via Geodesics on a Learned Manifold of Images. In G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor (editors), *Medical Image Computing and Computer-Assisted Intervention*, volume 5761 of *Lecture Notes in Computer Science*, 680–687. Springer Berlin / Heidelberg, 2009.

J. Hamm, D. H. Ye, R. Verma, and C. Davatzikos. GRAM: A framework for geodesic registration on anatomical manifolds. *Medical Image Analysis*, 14(5):633–642, 2010.

M. T. Hannan, D. T. Felson, and T. Pincus. Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. *The Journal of Rheumatology*, 27(6):1513–1517, 2000.

R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage*, 33(1):115 – 126, 2006.

J. Hohe, G. Ateshian, M. Reiser, K.-H. Englmeier, and F. Eckstein. Surface size, curvature analysis, and assessment of knee joint incongruity with MRI in vivo. *Magnetic Resonance in Medicine*, 47(3):554–561, 2002.

H. Hotelling. Analysis of complex statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933.

M. E. Houle, H.-P. Kriegel, P. KrÃűger, E. Schubert, and A. Zimek. Can Shared-Neighbor Distances Defeat the Curse of Dimensionality? In M. Gertz and B. LudÃďscher (editors), *Scientific and Statistical Database Management*, volume 6187 of *Lecture Notes in Computer Science*, 482–500. Springer Berlin Heidelberg, 2010.
`http://dx.doi.org/10.1007/978-3-642-13818-8_34`

D. J. Hunter. Osteoarthritis. *Best Practice & Research Clinical Rheumatology*, 25(6):801–814, 2011.

D. J. Hunter, G. H. Lo, D. Gale, A. J. Grainger, A. Guermazi, and P. G. Conaghan. The development and reliability of a new scoring system for knee osteoarthritis MRI: BLOKS (Boston Leeds Osteoarthritis Knee Score). *Annals of the Rheumatic Diseases*, 6:206–211, 2007.

D. J. Hunter, G. H. Lo, D. Gale, A. J. Grainger, A. Guermazi, and P. G. Conaghan. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds Osteoarthritis Knee Score). *Annals of the Rheumatic Diseases*, 67 (2):206–211, 2008.

D. J. Hunter, J. Niu, Y. Zhang, S. Totterman, J. Tamez, C. Dabrowski, R. Davies, M. H. Le Graverand, M. Luchi, Y. Tymofyeyev, C. R. Beals, and for the OAI Investigators. Change in cartilage morphometry: a sample of the progression cohort of the Osteoarthritis Initiative. *Annals of the Rheumatic Diseases*, 68(3):349–356, 2009.

P. Indyk and R. Motwani. Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality. *STOC '98 Proceedings of the thirtieth annual ACM symposium on Theory of computing*, 604–613, 1998.

E. Janousova, M. Vounou, R. Wolz, K. Gray, D. Rueckert, G. Montana, and T. A. D. N. I. (ADNI). Biomarker discovery for sparse classification of brain images in Alzheimer's disease. *Annals of the BMVA*, 2012(2):1–11, 2012.

H. Jia, G. Wu, Q. Wang, M. Kim, and D. Shen. iTree: Fast and accurate image registration based on the combinative and incremental tree. (In Proceedings) IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 1243–1246, 2011.

H. Jia, G. Wu, Q. Wang, and D. Shen. ABSORB: Atlas Building by Self-organized Registration and Bundling. *NeuroImage*, 51(3):1057–1070, 2010.

H. Jia, G. Wu, Q. Wang, Y. Wang, M. Kim, and D. Shen. Directed graph based image registration. *Computerized Medical Imaging and Graphics: The Official Journal of the Computerized Medical Imaging Society*, 36(2):139–151, 2012.

S. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.

I. Joliffe. *Principle Component Analysis*. Springer, 1986.

J. H. Kellgren and J. S. Lawrence. Radiological Assessment of Osteo-Arthrosis. *Annals of the Rheumatic Diseases*, 16(4):494–502, 1957.

L. S. Lohmander, M. Gerhardsson de Verdier, J. Rollof, P. M. Nilsson, and G. Engström. Incidence of severe knee and hip osteoarthritis in relation to different measures of body mass: a population-based prospective cohort study. *Annals of the Rheumatic Diseases*, 68(4):490–496, 2009.

J. MacQueen. Some methods for classification and analysis of multivariate observations. (In Proceedings) Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1, 281–297, California, USA, 1967.

J. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1–36, 1998.

L. M. March and C. J. Bachmeier. Economics of osteoarthritis: a global perspective. *Bailliere's Clinical Rheumatology*, 11(4):817–834, 1997.

E. Q. V. Martins and M. M. B. Pascoal. A new implementation of Yen's ranking loopless paths algorithm. *4OR: A Quarterly Journal of Operations Research*, 1:121–133, 2003.

C. Maurer and J. Fitzpatrick. *Interactive Image-Guided Neurosurgery*, chapter A review of medical image registration, 17–44. American Associatin of Neurological Surgeons, Park Ridge IL, 1993.

T. M. Mitchell. *Machine Learning*. McGraw-Hill Higher Education, new edition edition, 1997.

J. H. Morra, Z. Tu, L. G. Apostolova, A. E. Green, A. W. Toga, and P. M. Thompson. Comparison of AdaBoost and Support Vector Machines for Detecting Alzheimer's Disease through Automated Hippocampal Segmentation. *IEEE Transactions on Medical Imaging*, 29(1):30–43, 2010.

V. C. Mow and R. Huiskes. *Basic orthopaedic biomechanics & mechano-biology*. Lippincott Williams & Wilkins, 2005.

National Joint Registry (NJR). Full 9th NJR Annual Report. online, 2012. [accessed on 23rd October 2012].

R. L. Neame, K. Muir, S. Doherty, and M. Doherty. Genetic risk of knee osteoarthritis: a sibling study. *Annals of the Rheumatic Diseases*, 63(9):1022–1027, 2004.

K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.

X. Pennec. Intrinsic Statistics on Riemannian Manifolds: Basic Tools for Geometric Measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.

C. Peterfy, E. Schneider, and M. Nevitt. The osteoarthritis initiative: report on the design rationale for the magnetic resonance imaging protocol for the knee. *Osteoarthritis and Cartilage*, 16(12):1433–1441, 2008.

C. G. Peterfy, A. Guermazi, S. Zaim, P. F. J. Tirman, Y. Miaux, D. White, M. Kothari, Y. Lu, K. Fye, S. Zhao, and H. K. Genant. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. *Osteoarthritis and Cartilage*, 12(3):177–190, 2004.

J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing.* Cambridge University Press, 3 edition, 2007.

A. Qazi, J. Folkesson, P. Pettersen, M. Karsdal, C. Christiansen, and E. Dam. Separation of healthy and early osteoarthritis by automatic quantification of cartilage homogeneity. *Osteoarthritis and Cartilage*, 15(10):1199–1206, 2007.

J. Raynauld, J. Martel-Pelletier, M. Berthiaume, F. Labonté, G. Beaudoin, J. A. de Guise, D. A. Bloch, D. Choquette, B. Haraoui, R. D. Altman, M. C. Hochberg, J. M. Meyer, G. A. Cline, and J. Pelletier. Quantitative magnetic resonance imaging evaluation of knee osteoarthritis progression over two years and correlation with clinical symptoms and radiologic changes. *Arthritis & Rheumatism*, 50(2):476–487, 2004.

S. T. Roweis and L. K. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500):2323–2326, 2000.

D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18 (8)(8):712–721, 1999.

D. Schiphof, M. Boers, and S. M. Bierma-Zeinstra. Differences in descriptions of Kellgren and Lawrence grades of knee osteoarthritis. *Annals of the rheumatic diseases*, 67(7):1034–1036, 2008.

J. Schnabel, D. Rueckert, M. Quist, J. Blackall, A. Castellano-Smith, T. Hartkens, G. Penney, W. Hall, H. Liu, C. Truwit, et al. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. (In Proceedings) Medical Image Computing and Computer-Assisted Intervention-MICCAI 2001, 573–581, Springer, 2001.

E. Schneider, M. Nevitt, C. McCulloch, F. Cicuttini, J. Duryea, F. Eckstein, and J. Tamez-Pena. Equivalence and precision of knee cartilage morphometry between different segmentation teams, cartilage regions, and MR acquisitions. *Osteoarthritis and Cartilage*, 20(8):869–879, 2012.

T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometric models. *SIGGRAPH Comput. Graph.*, 20(4):151–160, 1986.

A. Serag, P. Aljabar, S. Counsell, J. Boardman, J. Hajnal, and D. Rueckert. LISA: Longitudinal image registration via spatio-temporal atlases. (In Proceedings) IEEE International Symposium on Biomedical Imaging (ISBI), 334–337, IEEE, 2012.

C. E. Shannon, W. Weaver, R. E. Blahut, and B. Hajek. *The mathematical theory of communication*, volume 117. University of Illinois press Urbana, 1949.

L. Sharma, J. Song, D. T. Felson, S. Cahue, E. Shamiyeh, and D. D. Dunlop. The role of knee alignment in disease progression and functional decline in knee osteoarthritis. *JAMA: the journal of the American Medical Association*, 286(2):188–195, 2001.

A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.

V. K. Srikanth, J. L. Fryer, G. Zhai, T. M. Winzenberg, D. Hosmer, and G. Jones. A meta-analysis of sex differences prevalence, incidence and severity of osteoarthritis. *Osteoarthritis and Cartilage*, 13(9):769 – 781, 2005.

T. Stammberger, F. Eckstein, M. Michaelis, K. Englmeier, and M. Reiser. Interobserver reproducibility of quantitative cartilage measurements: comparison of B-spline snakes and manual segmentation. *Magnetic Resonance Imaging*, 17(7):1033–1042, 1999.

C. Studholme, D. Hill, and D. Hawkes. Automated 3D registration of MR and PET brain images by multi-resolution optimisation of voxel similarity measures. *Med. Phys*, 24(1):25–35, 1997.

C. Studholme, D. L. G. Hill, and J. Hawkes, D. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32 (1):71–86, 1998.

J. Tamez-Pena, P. Gonzalez, J. Farber, K. Baum, E. Schreyer, and S. Totterman. Atlas based method for the automated segmentation and quantification of knee features: Data from the osteoarthritis initiative. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 1484–1487, 2011.

N. Tanishi, H. Yamagiwa, T. Hayami, H. Mera, Y. Koga, G. Omori, and N. Endo. Relationship between radiological knee osteoarthritis and biochemical markers of cartilage and bone degradation (urine CTX-II and NTX-I): the Matsudai Knee Osteoarthritis Survey. *Journal of bone and mineral metabolism*, 27(5):605–612, 2009.

J. B. Tenenbaum, V. De Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

L. Van der Maaten, E. Postma, and H. Van Den Herik. Dimensionality reduction: A comparative review. *Journal of Machine Learning Research*, 10:1–41, 2009.

V. Vapnik and A. Chervonenkis. A note on one class of perceptrons. *Automation and Remote Control*, 25(1), 1964.

V. Vedi, E. Spouse, A. Williams, S. Tennant, D. Hunt, and W. Gedroyc. Meniscal movement An in-vivo study using dynamic MRI. *Journal of Bone & Joint Surgery, British Volume*, 81(1):37–41, 1999.

G. Vincent, C. Wolstenholme, I. Scott, and M. Bowes. Fully automatic segmentation of the knee joint using active appearance models. *Medical Image Analysis for the Clinic: A Grand Challenge*, 224–230, 2010.

P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. *IEEE Computer Vision and Pattern Recognition (CVPR)*, 1:511–518, 2001.

P. Viola and I. Wells, W.M. Alignment by maximization of mutual information. (In Proceedings) Fifth International Conference on Computer Vision, 16–23, 1995.

U. von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

C. Wachinger, D. Mateus, A. Keil, and N. Navab. Manifold Learning for Patient Position Detection in MRI. (In Proceedings) IEEE International Symposium on Biomedical Imaging (ISBI), 1353–1356, Rotterdam, The Netherlands, 2010.

C. Wachinger and N. Navab. Manifold Learning for Multi-Modal Image Registration. (In Proceedings) Proceedings of the British Machine Vision Conference, 82.1–82.12, BMVA Press, 2010.

S. Warfield, K. Zou, and W. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*, 23(7):903–921, 2004.

F. Wildera, B. Hall, J. B. Jr, and N. Lemrow. History of acute knee injury and osteoarthritis of the knee: a prospective epidemiological assessment The Clearwater Osteoarthritis Study. *Osteoarthritis and Cartilage*, 10(8):611–616, 2002.

L. M. Wildi, J. Martel-Pelletier, F. Abram, T. Moser, J.-P. Raynauld, and J.-P. Pelletier. Assessment of Cartilage Changes Over Time in Knee Osteoarthritis Disease-Modifying Osteoarthritis Drug Trials Using Semiquantitative and Quantitative Methods: Pros and Cons. *Arthritis Care & Research*, 65(5):686–694, 2013.

T. G. Williams, C. J. Taylor, Z. Gao, and J. C. Waterton. Corresponding Articular Cartilage Thickness Measurements in the Knee Joint by Modelling the Underlying Bone. In R. E. Ellis and T. M. Peters (editors), *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003, 6th International Conference, Montréal, Canada, November 15-18, 2003,*

Proceedings, Part II, volume 2879 of *Lecture Notes in Computer Science*, 480–487, Springer, 2003.

W. Wirth, O. Benichou, C. K. Kwoh, A. Guermazi, D. Hunter, R. Putz, F. Eckstein, and for the OAI Investigators. Spatial patterns of cartilage loss in the medial femoral condyle in osteoarthritic knees: Data from the osteoarthritis initiative. *Magnetic Resonance in Medicine*, 63(3):574–581, 2010.

W. Wirth and F. Eckstein. A technique for regional analysis of femorotibial cartilage thickness based on quantitative magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 27(6):737–744, 2008.

W. Wirth, M.-P. H. L. Graverand, B. Wyman, S. Maschek, M. Hudelmaier, W. Hitzl, M. Nevitt, and F. Eckstein. Regional analysis of femorotibial cartilage loss in a subsample from the Osteoarthritis Initiative progression subcohort. *Osteoarthritis and Cartilage*, 17(3):291–297, 2009.

A. Wluka, R. Wolfe, S. Stuckey, and F. Cicuttini. How does tibial cartilage volume relate to symptoms in subjects with knee osteoarthritis? *Annals of the Rheumatic Diseases*, 63(3):264–268, 2004.

R. Wolz, P. Aljabar, J. V. Hajnal, A. Hammers, and D. Rueckert. LEAP: Learning embeddings for atlas propagation. *NeuroImage*, 49 (2):1316–1325, 2010.

Y. Wu, K. L. Chan, and L. Wang. Face recognition based on discriminative manifold learning. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR)*, 4:171–174, 2004.

R. Xu and D. C. Wunsch. Clustering algorithms in biomedical research: a review. *IEEE Reviews in Biomedical Engineering*, 3:120–154, 2010.

L. Yang. Distance Metric Learning: A Comprehensive Survey. 2006.

L. Yang, B. Georgescu, Y. Zheng, P. Meer, and D. Comaniciu. 3-D ultrasound tracking of

the left ventricles using one-step forward prediction and data fusion of collaborative trackers. *Proceedings of the IEEE Conference Computer Vision and Pattern Recognition*, 1–8, 2008.

L. Yang, R. Jin, L. Mummert, R. Sukthankar, A. Goode, B. Zheng, S. C. Hoi, and M. Satyanarayanan. A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(1):30–44, 2010.

Y. Yin, X. Zhang, R. Williams, X. Wu, D. Anderson, and M. Sonka. LOGISMOS - Layered Optimal Graph Image Segmentation of Multiple Objects and Surfaces: Cartilage Segmentation in the Knee Joint. *IEEE Transactions on Medical Imaging*, 29(12):2023–2027, 2010.

Y. Zheng, B. Georgescu, H. Ling, S. Zhou, M. Scheuering, and D. Comaniciu. Constrained marginal space learning for efficient 3D anatomical structure detection in medical images. (In Proceedings) IEEE Computer Vision and Pattern Recognition (CVPR), 194–201, 2009.