COMPUTATIONAL INTELLIGENCE SUPPORTING
ANATOMICAL SHAPE ANALYSIS AND
COMPUTER-AIDED DIAGNOSIS

by

Aaron D. Ward
M.Sc., University of Regina, Canada, 2004
B.Sc.Hons., University of Regina, Canada, 2000

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the School
of
Computing Science

© Aaron D. Ward 2008
SIMON FRASER UNIVERSITY
Fall 2008

All rights reserved. This work may not be
reproduced in whole or in part, by photocopy
or other means, without the permission of the author.
APPROVAL

Name:  

Aaron D. Ward

Degree:  

Doctor of Philosophy

Title of thesis:  

Computational Intelligence Supporting Anatomical Shape Analysis and Computer-Aided Diagnosis

Examining Committee:  

Dr. Greg Mori
Chair

Date Approved:
Abstract

Medical imaging technologies allow the collection of remarkable, three-dimensional pictures of the inside of the body, and have led to noninvasive means of disease diagnosis and treatment planning. However, the proliferation of this technology has resulted in the production of a huge number of medical images, increasing the demand on the radiology work force to a critical level. There is therefore an important need to provide a means of transforming medical image data into information that increases the accuracy and efficiency of radiologists’ work. This dissertation focuses on the problem of transforming medical image data in order to provide high-level information about the shapes of anatomical structures. Such information is useful to medical researchers addressing hypotheses relating anatomical shape and pathology, and is also useful to the development of computer-aided diagnosis systems based on shape. This dissertation describes two studies relating shape to pathology of musculoskeletal structures in the shoulder, and uses these studies to motivate research into further interesting questions in shape analysis. Techniques from computational intelligence, such as machine learning, graph matching, feature selection, manifold learning, optimization, and pattern recognition, are used in novel approaches to the steps of a shape analysis pipeline supporting medical research. We describe a machine learning-based approach to eliciting expert knowledge about feature saliency, for use in establishing shape correspondence. We propose a novel approach to medial shape description that localizes shape deformations, and demonstrate a manifold learning-based approach to computing the basic building blocks of it and other medial shape descriptions. Finally, we propose a groupwise paradigm for the computation of a pruning order for the components of medial shape representations, in order to remove unwanted components arising from noise. These contributions to the shape analysis pipeline enable more accurate and intuitively-understood results, enabling medical researchers to gain further understanding into pathological processes in the body.
To my family.
“Explanations exist: they have existed for all times, for there is always an easy solution to every problem – neat, plausible and wrong.”


“What was that?” hissed Arthur.

“Something red,” hissed Ford back at him.

“Where are we?”

“Er, somewhere green.”


— Douglas Adams, Life, the Universe, and Everything, 1982
Acknowledgments

The work described in this dissertation would not have been possible without the generous support of the many people who were involved. First, I would like to sincerely thank Dr. Ghassan Hamarneh, my senior supervisor, for his intellectual support throughout the years that I have been a part of his lab. Ghassan has been extremely generous with his valuable time, ideas, and skills in supporting this work, and has arranged numerous collaborations and valuable opportunities for me to interact with the wider research community. I am extremely grateful for his very reliable, unwavering support and hard work in this research endeavour, and I owe him a sincere debt of gratitude for helping me to lay the foundation for a career in medical image analysis research.

I am also extremely grateful to the members of my supervisory committee, Dr. Stella Atkins and Dr. Richard Zhang, who have devoted substantial time and effort to the critical analysis of this work, and have provided numerous helpful comments and constructive criticisms. I want to thank Dr. Atkins in particular for going beyond the call of duty to provide me with very valuable career advice on numerous occasions. I want to thank Dr. Zhang for his attention to the mathematical details of this work and for his numerous, helpful and detailed comments on my thesis proposal. His perspective was very refreshing, and caused me to take a critical, dispassionate second look at this work throughout the writing of this dissertation.

In all of the medical research studies described in chapter 2, I enjoyed an excellent working collaboration with Dr. Mark Schweitzer, Professor of Radiology and Orthopedic Surgery, Hospital for Joint Diseases, New York University Medical Center (now Chair of Diagnostic Imaging at the University of Ottawa and Chief of Radiology at The Ottawa Hospital). Dr. Schweitzer and two of his radiology residents, Dr. Reem Ashry, and Dr. Sarah Shock, were instrumental in the completion of the medical research studies described
chapter 2. My collaboration with them has kept me grounded, as a reminder of the practical needs to be served by our research. For both musculoskeletal studies, they performed all of the necessary data collection and ground truth diagnosis. For the supraspinatus project given in section 2.1, Dr. Ashry provided manual delineations of the muscles for every patient.

Many of the members of our research lab have been of tremendous help in discussions and brainstorming of ideas related to medical image analysis research. I wish to extend sincere thanks to all of the lab members with whom I have had the pleasure to interact during my years as a graduate student. Additionally, I would like to thank the following specific lab members, in no particular order, for their assistance: Mr. Dylan Tisdall, for his constant willingness to discuss anything and everything of a mathematical nature, and for offering his unique perspective on my work whenever I asked for it; Mr. Chris McIntosh, whose drive to ensure strong theoretical foundations for his research served as a helpful inspiration; Mr. Paul Wighton, for his helpful advice related to statistical analysis; Mr. Ahmed Saad, for our many interesting discussions on the practical impacts of our research; Ms. Lisa Tang, for her very helpful input regarding the shape correspondence problem; Mr. Eli Gibson, for his manual segmentations, and valuable contributions of ideas related to automating the classical measurements of the bicipital groove, appearing in section 2.2; Mr. Richard Frank, for his extensive work, both creative and implementational, on the medial patch shape description in section 4.1; and Mr. Ali Khan and Mr. Ben Smith for our many enlightening discussions regarding the meaning of making truly useful contributions to scientific research. I have gained tremendously from my interactions with everyone in our lab, and this is reflected in the research presented in this dissertation.

Finally, I want to thank my family for the endless support they have provided to me throughout my graduate career. My chosen path has kept me at my desk in front of a computer well beyond normal working hours for most my many years as a graduate student, limiting the amount of time I have had available to spend with the ones I love. To Tania, Galyna, Denise, Gerry, Ashley, and especially to my daughter, Alina, thank you so much for the patience, support, and love during this very difficult journey. Without your support, this dissertation, and the tremendous intellectual growth I have experienced in developing the work behind it, would not have been possible.
## Contents

**Approval** ii

**Abstract** iii

**Dedication** iv

**Quotation** v

**Acknowledgments** vi

**Contents** viii

**List of Tables** xi

**List of Figures** xii

### 1 Introduction 1

1.1 Motivation of this dissertation ................................. 1
1.2 Computational intelligence ........................................ 3
1.3 Anatomical shape analysis and computer-aided diagnosis ........ 4
1.4 Shape representation ............................................... 7
  1.4.1 Global shape representations .................................. 7
  1.4.2 Boundary-based shape representations ......................... 9
  1.4.3 Medial-based shape representations ............................. 14
  1.4.4 Other approaches ............................................... 21
1.5 Shape correspondence .............................................. 22
1.6 Outline of this dissertation ....................................... 28
List of Tables

2.1 Supraspinatus pathology groups ........................................ 35
2.2 Types of global supraspinatus shape measurements ................. 40
2.3 Results of supraspinatus shape measurements ........................ 45
2.4 Supraspinatus measurements’ p values ............................... 45
2.5 Classification accuracies based on supraspinatus shape measurements .... 47
2.6 Global shape measurements of the bicipital groove .................. 69
2.7 Bicipital groove global measurements’ p values ...................... 69
2.8 Bicipital groove measurements, by patient age ....................... 69
2.9 Results of validation of the medial shape descriptor for the bicipital groove . 70
3.1 Quantitative accuracies and timings of MDL vs. our method ....... 92
5.1 SVM classification accuracies based on skeletons of silhouettes in the Brown LEMS database ............................................ 150
List of Figures

1.1 Anatomy of the bicipital groove of the proximal humerus .................. 5
1.2 Schematic of a boundary-based shape description .......................... 8
1.3 Schematic of a medial-based shape representation ......................... 15
1.4 Schematic of the M-reps shape representation .............................. 17
1.5 Illustration of the instability of the medial axis transform ............... 19
1.6 Illustration of the process of constructing a statistical shape model ..... 22
1.7 Illustration of type I, II, and III landmarks ................................. 24
1.8 Chart depicting the components of the research in this dissertation .... 29
2.1 Anatomy of the supraspinatus muscle ..................................... 33
2.2 Computational pipeline for the supraspinatus study ....................... 36
2.3 Outlining tool used in the supraspinatus study ............................ 37
2.4 Supraspinatus muscle surface segmentation ............................... 38
2.5 Samples of segmented supraspinati ....................................... 39
2.6 Segmented supraspinatus visualized in anatomical context ............. 40
2.7 Illustrative examples of principal components analysis .................. 41
2.8 Illustrative example showing use of multiple features for classification 43
2.9 Separability of supraspinatus pathology groups based on shape measurements 46
2.10 Anatomy of the bicipital groove of the proximal humerus ............. 52
2.11 Classical, 2D measurements of bicipital groove ......................... 53
2.12 Appearance of humeral bone surfaces on MRI ............................ 55
2.13 Segmented bicipital groove visualized in anatomical context .......... 56
2.14 Computational pipeline used to compute the differences between bicipital groove pathology groups .......................... 56
2.15 Examples of bicipital grooves with different elongation measures .... 57
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.16</td>
<td>Computational pipeline for computing a medial shape descriptor for the bicipital groove</td>
<td>57</td>
</tr>
<tr>
<td>2.17</td>
<td>Fitting the intertubercular and medial sheets to the bicipital groove</td>
<td>59</td>
</tr>
<tr>
<td>2.18</td>
<td>Schematic of the medial shape descriptor for the bicipital groove</td>
<td>60</td>
</tr>
<tr>
<td>2.19</td>
<td>Computational pipeline for validating the medial shape descriptor for the bicipital groove</td>
<td>62</td>
</tr>
<tr>
<td>2.20</td>
<td>Diagram depicting the computation of the depth of the bicipital groove</td>
<td>64</td>
</tr>
<tr>
<td>2.21</td>
<td>Process of computing point correspondence across bicipital grooves</td>
<td>66</td>
</tr>
<tr>
<td>2.22</td>
<td>Illustration of the normalization of thickness fields</td>
<td>68</td>
</tr>
<tr>
<td>2.23</td>
<td>Illustration of possible plateaus in the medial wall of the bicipital groove</td>
<td>71</td>
</tr>
<tr>
<td>2.24</td>
<td>Illustration of the interpretation of the thickness fields of the bicipital groove</td>
<td>72</td>
</tr>
<tr>
<td>2.25</td>
<td>Illustration of the quantitative exploration of the thickness fields</td>
<td>73</td>
</tr>
<tr>
<td>3.1</td>
<td>Correspondences given by MDL on rectangles with two peaks</td>
<td>79</td>
</tr>
<tr>
<td>3.2</td>
<td>Computational pipeline for establishing correspondence using expert knowledge</td>
<td>82</td>
</tr>
<tr>
<td>3.3</td>
<td>Illustration of two types of correspondence error</td>
<td>84</td>
</tr>
<tr>
<td>3.4</td>
<td>Expert-given landmarks on three training samples</td>
<td>85</td>
</tr>
<tr>
<td>3.5</td>
<td>Cost functions obtained from the training process</td>
<td>86</td>
</tr>
<tr>
<td>3.6</td>
<td>Three sets of features computed for three shapes of corpora callosa</td>
<td>88</td>
</tr>
<tr>
<td>3.7</td>
<td>Comparison of correspondences established by MDL vs. our method on a textural example</td>
<td>93</td>
</tr>
<tr>
<td>3.8</td>
<td>Comparison of correspondences established by MDL vs. our method on corpora callosa</td>
<td>94</td>
</tr>
<tr>
<td>3.9</td>
<td>Comparison of correspondences established by MDL vs. our method on brain ventricles</td>
<td>95</td>
</tr>
<tr>
<td>4.1</td>
<td>Illustration of the medial patch representation of a bent and extruded slab</td>
<td>101</td>
</tr>
<tr>
<td>4.2</td>
<td>Terminology and variables of medial patches</td>
<td>103</td>
</tr>
<tr>
<td>4.3</td>
<td>Detailed view of the differential spherical coordinates used in medial patches</td>
<td>103</td>
</tr>
<tr>
<td>4.4</td>
<td>Medial patch representation of a caudate nucleus</td>
<td>108</td>
</tr>
<tr>
<td>4.5</td>
<td>Deformations applied to a bicipital groove using medial patches</td>
<td>109</td>
</tr>
<tr>
<td>4.6</td>
<td>Illustration of the “crossing normals” problem in medial representations</td>
<td>114</td>
</tr>
<tr>
<td>4.7</td>
<td>Computational pipeline for computing the components of a medial shape representation using manifold learning</td>
<td>116</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Motivation of this dissertation

Medical imaging modalities such as magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound enable the capturing of rich data about the structure of the inside of the body. The proliferation of medical imaging technology, and its benefits to disease diagnosis and treatment planning, have resulted in the ongoing collection of a very large number of medical images. As an illustration of the scale of medical image data collection, consider that in 2003, there were 60 million MR scans performed worldwide [192], and 80 million ultrasound scans performed in the United States [113]. In 2007, 62 million CT scans were performed in the United States [32]. The pace of medical image collection continues to increase; between 1998 and 2002 in the United States, there was a 110% increase in the number of CT scans performed, a 66% increase in MRI scans, and a 45% increase in ultrasound scans [111]. The aging baby boom generation is expected to create an even greater demand for diagnosis based on medical images, with the total number of imaging examinations expected to increase from 350 million to 500 million between 2003 and 2023 [148].

After each of these millions of images is collected, it needs to be directly inspected by a highly-trained medical specialist, a radiologist, in order to obtain an accurate diagnosis leading to a treatment plan. The huge number of produced images is placing an enormous strain on the work force in radiology, and on the education system that produces new radiologists to join that work force. The following facts illustrate the situation:
CHAPTER 1. INTRODUCTION

- In 1998, there were at least 600 more positions in radiology available in the United States than there were radiologists to fill them [200].

- The ratio of number of radiology positions available to the number of qualified people seeking those positions was 1.3:1 in 1998, and increased to 3.8:1 in 2000 [201]. This illustrates that the situation is not improving with time.

- A survey in 2000 of radiologists in the United States revealed that 51% of radiologists believed their workloads exceeded their capacity, compared with only 5% who responded that they had too little work to do [199].

- In a 2002 survey of hospital administrators, 63% identified radiologists as the most difficult physician specialists to recruit [229].

- This situation is expected to continue to deteriorate. The number of radiologists entering the field is increasing by 2% annually, whereas the use of medical images is increasing by 6% per year [6].

The pressure on radiologists to diagnose an excessive number of cases in a very limited time leads to concern about errors in diagnosis. It has been shown that the speed at which radiologists scan through 3D images in cine-mode (quickly paging through 2-dimensional [2D] slices of 3-dimensional [3D] images on a computer screen) has an impact on diagnostic accuracy: paging through the images too quickly increases mistakes [80]. The issue of “satisfaction of search”, where a radiologist stops examining a case when an anomaly is found, rather than searching for further problems, is also exacerbated under tight time constraints [80]. Clearly, there is a problem in diagnostic radiology today: the number of images generated every year is increasing much more rapidly than the number of radiologists available to examine them, the existing work force is under heavy strain, and this can lead to errors in diagnosis and treatment.

This problem drives research in medical image analysis, which is the transformation of medical image data into useful, higher-level information that increases the speed, accuracy, precision and/or repeatability of radiologists’ diagnosis and subsequent treatment plans. Medical image segmentation is an example of such a transformation, where the result is a division of the medial image data into relevant anatomical parts. Automatic and semi-automatic segmentation reduces the workload on the radiologist attempting to find
structures of interest in medical images, and in attempting to precisely locate their coordinates as part of a surgical or radiation treatment plan. The automatic transformation of segmented structures into information describing their shapes, leading to automated or semi-automated diagnosis of their pathological conditions, yields even higher level diagnostic information to the radiologist. It is this last transformation, from segmented structures to shape and diagnostic information, that is the topic of this dissertation.

1.2 Computational intelligence

Computational intelligence (CI) [58, 161] is related to more conventional artificial intelligence (AI) in that both disciplines somehow aim to reproduce human intelligence characteristics in a machine. However, whereas AI techniques are generally focused on formulating and solving problems in a symbolic way, CI expands beyond this to include more diverse perspectives and approaches to the problem of achieving human qualities of intelligence artificially. CI is an umbrella term encompassing many areas, including machine learning, neural networks, evolutionary computing, artificial life, swarm intelligence, machine vision, fuzzy systems, and graph learning and matching, among many others. The focus of this dissertation is on the development of novel approaches to problems in anatomical shape analysis and computer-aided diagnosis, involving elements of CI. Although we use different CI techniques for different purposes throughout the dissertation, automated learning runs as a common thread throughout the chapters. We believe this is a critical step for a system which is to transform raw data into meaningful information that helps to make radiologists’ work more efficient.

We employ techniques from machine learning to the problem of computer aided diagnosis of shoulder disorders, and a dynamic programming technique to the computation of a shape descriptor for a structure in the shoulder (chapter 2). We design a system that automatically learns the characteristic shape and appearance features of important anatomical landmarks from an expert user, and uses what it has learned in order to provide refined shape correspondences (chapter 3). We also propose a manifold learning-based technique for computing the building blocks of any medial shape representation, using an approach that automatically learns about the intrinsic parameterization of the nonlinear surfaces involved (chapter 4). We propose an approach that automatically discovers commonalities within a group of shapes in order to distinguish parts of those shapes arising from true signal from
1.3 Anatomical shape analysis and computer-aided diagnosis

Anatomical shape analysis is the study of the shapes of structures inside the body. Such a study can involve a single shape, wherein one describes the shape using both global (e.g. surface area, volume) and local (e.g. surface curvature near a point on the surface) measures, providing visualizations for qualitative analysis and numbers to quantify the computed measures. It can also involve multiple shapes, where one is interested in computing and analyzing the differences between individuals and groups. In this dissertation, the source of the shapes takes the form of a set of segmented structures from medical image data, usually in the form of binary 2D and 3D images (where 1s represent the segmented structure, and 0s represent the background). The goal of anatomical shape analysis is to compute useful information about their shapes, given 2D or 3D data depicting anatomical structures of interest.

Anatomical shape analysis is a useful technique for a number of applications within the domain of medical image analysis. Here, we will describe two such applications: deformable model segmentation and medical studies of the relationship between shape and pathology.

As a first step in many studies involving medical image analysis, the structure of interest must be segmented from the rest of the medical image data surrounding it. A common approach to this problem is to use deformable model segmentation [66, 85, 121, 155, 231], wherein, for example, a triangle mesh is iteratively guided from an initial state in the medical image to a state in which it delineates the region of interest from the remainder of the image. This usually involves an energy minimization scheme that balances an external energy term and an internal energy term. The external energy value is computed based on the medical image data lying at or near the mesh vertices. The internal energy value is computed based on some prior knowledge or heuristics about what it means for a mesh to be “reasonable” [126]. For example, one heuristic might be that the surface implied by the mesh should be smooth. Following that heuristic, the internal energy term would penalize meshes with non-smooth boundaries. The overall idea is that the external energy term should drive the mesh to the correct loci in the image, and the internal energy term should help the optimizer to overcome noise. A more sophisticated internal energy heuristic can be based on the plausibility of the shape of the mesh according to a precomputed shape
model (please see appendix A for details on computing shape models). For example, if we are segmenting a hippocampus (brain structure) from a MR image, we could use a shape model of the hippocampus to ask, about our mesh, the question, “How plausibly does this mesh depict a hippocampus?”. The answer to this question is used to compute the internal energy value, ensuring that we do not deform the mesh such that it no longer resembles a hippocampus. The use of these competing energy terms presents a dilemma of how much weight to give to each of them; regularization using a shape prior or other means increases robustness to noise, at the expense of making the approach inflexible to fitting to cases that differ significant from the prior. Constructing an accurate shape model based on a set of exemplars for use in deformable model segmentation is one of the important applications of anatomical shape analysis.

Medical researchers testing hypotheses involving a relationship between the shape of an anatomical structure and pathological conditions of that structure require accurate, meaningful descriptions of anatomical shape. As we shall explore in detail in chapter 2, there exist several studies of the relationship of the shape of the bicipital groove of the proximal
humerus to the incidence of injury (dislocation, subluxation, tearing) of the long biceps tendon [3, 100, 137, 153, 214]. Figure 1.1 depicts the anatomy of the humeral head, describing the bicipital groove. The bicipital groove is formed by two tuberosities (bony protrusions) on the head of the humerus (upper arm bone) and serves to retain the long biceps tendon, which extends from the biceps, over the humeral head, and into the shoulder, as the arm moves. The referenced studies discovered that the shape of the bicipital groove, as seen on 2D radiographic (X-ray) views or in excised bones, has a bearing on the tendency to injure the long biceps tendon. These studies demonstrate both the importance of accurate shape measurements of the anatomy, and also reinforce the need for shape analysis in 3D. Relationships between 3D shapes of anatomical structures and their associated pathologies can lead to powerful new hypotheses about disease progression and diagnosis, and also to computer-aided diagnosis techniques. An example of this type of medical problem, also covered in chapter 2, is the diagnosis of the type of damage sustained in a supraspinatus (a shoulder muscle) injury. After a supraspinatus tear, the muscle can atrophy, retract, or both. It is important to distinguish these conditions since one of them (retraction) is correctable by surgery, whereas the other (atrophy) is not. Since both of these conditions result in a reduction of the apparent size of the muscle, they are often difficult to distinguish on medical images, and shoulder arthroscopy, involving an incision and insertion of an optical camera, is the gold standard for diagnosis. There is much active research into finding a reliable, non-invasive means for supraspinatus diagnosis using ultrasound or MRI [10, 120, 162, 169, 240], and a correlation between the 3D shape of the muscle extracted from medical images and the various pathological conditions of the muscle is of great assistance in developing a reliable, non-invasive approach to diagnosis.

Performing an analysis of anatomical shapes requires that we address three interesting problems. First is the choice of a shape representation suitable to the task at hand; an overview of this research area is provided in section 1.4. Second, depending on the choice of shape representation, we may be confronted with the shape correspondence problem, wherein an accurate, meaningful correspondence must be established between each point on each shape and a point on every other shape in a group of structures. We give an overview of this problem in section 1.5. Third, when a medial axis-based shape representation is used, we are confronted with the notorious instability of the representation with respect to boundary noise, and this must be addressed. We discuss this problem further in section 1.4, since it is pertinent to medial-based shape representations.
1.4 Shape representation

When performing anatomical shape analysis, or segmentation using a shape model, one must make an appropriate choice of shape representation. That is, what kinds of variables should be used to capture the important and relevant information about a shape? This choice leads to a number of compromises in terms of the complexity of computing the chosen representation from a binary image of a shape, the stability of the chosen representation with respect to noise and boundary perturbations of a shape, the ability to compactly encode typical shape deformations, and the intuitive interpretability of the resulting analysis.

Thompson [210] made the earliest recorded attempt at the morphometric analysis of biological structures, positing for the first time that the physical laws of mechanics play a strong role, alongside evolution, for determining the shape of an organism. He also explored, for the first time, the degree to which shape differences between animals could be described by mathematical transformations. In this dissertation, we focus mainly on the use of global, boundary-based, and medial-based (the latter two being local) approaches to shape description. Consequently, the next three sections are devoted to describing these types of approaches. Following this, we describe some alternative approaches to shape representation that are given in the literature.

1.4.1 Global shape representations

A global means of describing a shape provides a single descriptor capturing properties of the entire shape. Examples of global shape descriptors are the shape’s volume, surface area, and overall elongation. There are also 3D moment invariants [33, 176] that characterize the distribution of the mass of a shape in space. More details about these approaches are provided in chapter 2, where they are used in two medical studies. Global approaches to shape description are arguably the simplest to work with, mainly due to the fact that they are typically invariant to object translation and rotation. This allows their direct comparison across shapes, regardless of their alignment. The obvious drawback to using global shape descriptors is that they aggregate all information about an object’s shape into a single value. However, as we shall see in chapter 2, a collection of global shape descriptors can sometimes provide a powerful means of distinguishing pathology groups in medical studies.
Figure 1.2: A salamander shape using a boundary representation. Points are sparsely distributed on the boundary for clarity of this figure; points can be distributed with any desired density.
1.4.2 Boundary-based shape representations

Boundary-based shape representations capture shape information by describing the shape’s surface, as in the 2D example of points sampling the boundary of the salamander shape in figure 1.2. Broadly, boundary-based shape representations can be classified in (at least) three ways: explicit representations, parametric representations, and implicit representations. Explicit representations capture information about the shape’s boundary directly; capturing the points sampling the boundary in figure 1.2 would yield an explicit representation of this boundary. Implicit approaches represent shapes by embedding them in a higher-dimensional space, and defining a function \( \phi : \Omega \rightarrow \mathbb{R} \), where \( \Omega \) represents the domain of this space, and the boundary is defined by the set of loci \( \{ x | \phi(x) = c \} \) for some constant \( c \in \mathbb{R} \). Parametric representations encode important shape information by creating functions mapping a few important variables/parameters from a low-dimensional space into the intrinsic space of the object surfaces. A shape is thus captured in small number of values that weight the basis functions; basic examples of this kind of shape representation are B-spline surfaces and Bezier surfaces [143]. In the following we illustrate some representative examples of boundary-based shape representations falling into each of these categories.

Explicit boundary representations

One approach to explicit, boundary-based shape representation is to use a polygon mesh to represent the object’s surface, with mesh vertices lying on the surface and edges connecting vertices. Such a representation permits the construction of deformable shape models [87, 89, 133, 135, 136, 231], which allow the polygon mesh to be deformed, but also provides constraints on the deformation. For example, the edges of the mesh can be modeled as springs that regularize deformations of the mesh. Deformations of the mesh can also be governed by a finite element model (FEM) [134]. Such mesh-based models can also be organized hierarchically, in order to speed the search for shapes in image segmentation and to allow dynamic topological adaptations [114, 132, 138, 142]. Mesh-based deformable models have the advantage of great flexibility and straightforward construction from a binary image of an object; the marching cubes algorithm [125] can be used to convert an isosurface from a volume into a mesh. However, such polygon meshes can have digitization artifacts arising from their binary image sources.

The curvature scale space (CSS) [140, 141] approach explicitly captures the shape of a
boundary as a function

\[ c = f(t, \sigma), \]

(1.1)

where \( c \) represents a curvature value, \( t \) is the arc length along the boundary from a defined starting point, and \( \sigma \) is the scale of the curvature computed at position \( t \), represented as the standard deviation of a Gaussian function used to smooth the curve. Thus a contour is represented using its local curvature values, and at a variety of different spatial scales; hence the curvature scale space moniker. This approach is therefore a parametric shape representation focusing on curvature as the feature of interest in defining a contour, and has been used in 2D medical image analysis for diagnosis of malignant melanomas in [118]. For closed contours, however, this approach presents the problem of defining the starting point in some consistent way; such a definition is crucial in shape matching applications using this representation. It has been shown, however, that a phase-correlation approach is useful in aligning the start points of two closed curves represented in CSS, and this representation has been shown to be effective in shape matching [117, 174]. Recognition of 3D objects based on a CSS-representation has been investigated by learning about the 3D shape by capturing an optimal number of 2D views, and then recognizing a 3D object by its 2D silhouette taken from a single view [139].

Landmark-based approaches to boundary-based shape description sample a contour with a set of points, called landmarks, as depicted in figure 1.2. Often, a set of landmark-based shape representations is collected to form a point distribution model (PDM) [37], which describes the observed variability in the landmarks in the sample using a mean shape and the major modes of observed variation in the sample. This is a useful statistical model of shape because it permits the computation of the plausibility of a shape, according to the distance from the mean. In appendix A, the details of how to construct a statistical shape model are discussed. Landmark-based approaches have the advantage of a very straightforward representation, being a simple collection of points for each shape. They also carry the advantage of straightforward statistical analysis, since a shape represented by a set of landmarks lives in a Euclidean space, permitting linear approaches to analysis such as PCA for dimensionality reduction. The main drawbacks to this approach are that it is not a continuous shape description, two landmark-based shape representations need to be aligned into a common coordinate frame to be compared, and good landmark correspondences must be established in order to obtain a meaningful PDM.
CHAPTER 1. INTRODUCTION

Parametric boundary representations

A combination of spherical harmonic basis functions can be used to represent a shape surface [31]. Using this approach, a shape is represented using a set of coefficients of these basis functions. This is analogous to using Fourier descriptors to specify the contour of a 2D shape; spherical harmonics are the analogous basis functions for surfaces of 3D objects. The approach involves mapping every point on the object’s surface to a point \((\theta, \phi)\) on the unit sphere. This approach has the benefit of a continuous representation of the object surface, free of any staircasing artifacts arising from digitization. Surfaces can also be aligned according to their low-order harmonics, and correspondence can then be established across surfaces in the coordinates \((\theta, \phi)\) ([197] evaluates this method for correspondence establishment). However, the drawbacks to using this representation are twofold. First, since the basis functions are spherical harmonics, no combination of these functions can represent a shape with a non-spherical topology. However, since many organs and structures in the body have a spherical topology, this representation can be used for a large number of structures despite this limitation. Another drawback to using this approach involves the mapping of every point on the object’s surface to a point on the unit sphere. Since this operation may (and usually does) result in a deformation of the object’s surface, we must decide on a heuristic constraint on this warping that maintains some useful aspect(s) of the object’s shape. If we have a polygon mesh representing the object, and we wish to compute a spherical harmonic representation of the object, two typically-used heuristics include computing a mapping that preserves the angles between adjacent polygon edges, and computing a mapping preserving the areas of the polygons. Since these different constraints lead to different spherical harmonic representations of the same shape, this choice of heuristic constraint represents a free parameter in the computation of the shape representation.

Wavelet-based representations of shape provide a hierarchical view of a shape’s boundary by computing the wavelet transform [130] of the shape’s boundary [62, 166, 211]. The representation is then formed by the resulting wavelet coefficients, and is therefore related to the spherical harmonics and Fourier-based approaches in that basis function coefficients form the representation. At a high level, the hierarchical representation of the object’s shape is computed by passing the boundary through a wavelet filter bank which, at each stage, puts the boundary through a low-pass filter and a high-pass filter. The high-pass filtered result from each stage is used as the input to the filters in the next stage of the filter bank.
CHAPTER 1. INTRODUCTION

This results in a hierarchical representation of the boundary that has a space requirement no greater than that for the original shape boundary. However, by storing only the most significant wavelet coefficients, significant compression has been achieved [53]. Wavelet-based representations are favourable in that they give a precise measure of the error of the representation in approximating the original boundary, and they permit the localization of frequency changes in small domains [166]. This last advantage has led to their use in the construction of hierarchical active shape models, where the wavelet transform provides an elegant way to partition the spatial domain of the shape captured by the model [47]. This coarse-to-fine hierarchy allows a decomposition of the general configuration of an object boundary from its detailed, higher-frequency components [62].

Implicit boundary representations

The level set method [150] is an approach to shape representation that captures the boundary, the inside, and the outside of a shape. Using this approach, a shape of an object (restricted in this discussion to 2D shapes, but generalizable to nD) is captured implicitly as a function $f(x, y)$ on the Cartesian grid. The boundary of the shape is given by the loci $(x, y)$ where $f(x, y) = 0$. At the loci corresponding to the inside of the shape, $f(x, y) > 0$, and outside of the shape $f(x, y) < 0$. At every locus $(x, y)$, the absolute value of $f(x, y)$ is the distance to the nearest point $(x', y')$ where $f(x', y') = 0$. Thus, essentially, the shape of an object is captured using its signed distance transform, the values of which are given by $f$. The level set approach avoids parameterizing the boundary of the shape, and in doing so, it simplifies the modeling of topological changes to shapes. This is helpful to both the evolution of the boundary of a shape, as when segmenting a structure, and the computation of statistics of multiple shapes. Because shapes of different topologies all inhabit the same space, namely, that of all level set functions, shape correspondence can be established in the $(x, y)$ domain of level set functions. The drawback to using this shape representation arises from the fact that the space of distance transforms is not a Euclidean vector space [173]. Therefore, linear methods for statistical shape modeling do not apply, and more complex approaches are required.

In a related work, Cremers et al. [43] introduce a definition of shape which is a relaxation of a binary image representation. A shape embedded in a $d$-dimensional image is represented by a function $q : \mathbb{R}^d \rightarrow [0, 1]$, giving the probability that a pixel $x \in \mathbb{R}^d$ is part of the shape. Essentially, this allows the relaxation of the specification of the precise position of the shape.
boundary, in contrast to the level set approach wherein the boundary is given precisely as the loci \((x, y)\) where \(f(x, y) = 0\). This approach has two advantages over the level set approach. First, the shape space using the proposed representation is linear, in contrast to that of the level set representation, which is nonlinear. When an image is segmented with a shape prior based on this representation, the resulting cost functional is convex; this is not the case using the level set method.

**Hybrid methods**

Deformable superquadrics [206] is an approach to boundary representation that permits the use of few parameters to represent a large variety of shapes. The basic idea of this approach is to start with a primitive, such as a supertoroid, superhyperboloid, or superellipsoid, and then adjust its parameters such that it represents the desired shape. For example, the superellipsoid is used to represent any shape with spherical topology, and is given by the implicit surface [20] equation [35]

\[
\begin{aligned}
f(x, y, z) &= \left( \frac{x}{a_1} \right)^{\frac{2}{\epsilon_1}} \left( \frac{y}{a_2} \right)^{\frac{2}{\epsilon_2}} + \left( \frac{z}{a_3} \right)^{\frac{2}{\epsilon_1}}. \\
&= \left( \frac{x}{a_1} \right)^{\frac{2}{\epsilon_1}} \left( \frac{y}{a_2} \right)^{\frac{2}{\epsilon_2}} \left( \frac{z}{a_3} \right)^{\frac{2}{\epsilon_1}}. 
\end{aligned}
\]  

This equation has an ellipsoid form, with additional parameters \(\epsilon_1\) and \(\epsilon_2\) that give additional control over the shape’s curvature. The shape is defined such that \(f(x, y, z) = 1\) when the point \((x, y, z)\) lies on the shape surface. One of the main benefits to this approach is that the parametric representation of the superellipsoid can be written as [35]

\[
S(\eta, \mu) = \begin{bmatrix} a_1 \cos^{\epsilon_1}(\eta) \cos^{\epsilon_2}(\mu) \\ a_2 \cos^{\epsilon_1}(\eta) \sin^{\epsilon_2}(\mu) \\ a_3 \sin^{\epsilon_1}(\eta) \end{bmatrix}, \quad -\frac{\pi}{2} \leq \eta \leq \frac{\pi}{2}, \quad -\pi \leq \mu \leq \pi. 
\]  

Deformable superquadrics are thus a hybrid implicit/parametric shape representation, and allow the capturing of both global and local shape variations with a small number of parameters [206]. They have also been shown to support physically-based and (user) interactive deformations [206].

In general, boundary-based shape representations have the advantage of being relatively straightforward to compute from a binary image of a structure, since the boundary of the structure is explicitly represented in a binary image resulting from segmentation. The main
CHAPTER 1. INTRODUCTION

1.4.3 Medial-based shape representations

Medial shape representations capture shape information by describing the set of curves (for shapes embedded in 2D) or surfaces (for shapes embedded in 3D) lying medial to the object, as in the points sampling the medial axes of the salamander shape in figure 1.3. The boundary of the shape is then implied by a set of vectors emanating from each point on the medial axes. The medial shape representation has been of considerable interest to the research community in recent years [9, 25, 26, 46, 54, 63, 73, 85, 91, 92, 93, 108, 122, 127, 129, 155, 157, 159, 180, 189, 195, 196]. This is mainly due to the fact that such representations encode shape information in an object-centered coordinate system, yielding variables that directly encode intuitive notions of shape such as bending, thickness, and elongation. Such representations are therefore very useful in anatomical shape studies, since the results of such studies need to communicate clearly and intuitively the relationship between shape and pathology, often to an audience with a medical, rather than computational, background. Another reason for the attention received by medial shape representations is that there exist numerous psychovisual studies that demonstrate that human perception of shape tends to focus on medial aspects [190]. Chapter IX of Thompson’s seminal work [210] provides an early look at the mathematical description of skeletal structures of biological organisms.
Figure 1.3: A salamander shape using a medial representation. In contrast to the boundary representation, points are distributed along axes lying medial to the object, and the boundary is implied by vectors (blue) emanating from each medial point. Points are sparsely distributed on the medial axes for clarity of this figure; points can be distributed with any desired density.
An approach to medial shape representation that has seen substantial application in medical studies of shape is the M-Reps shape representation developed at the University of North Carolina [155]. Using this representation, shapes are decomposed into individual parts called figures, where each figure has within it a single medial surface that represents the figure. Figure 1.4 illustrates this, with two medial surfaces lying within a two-figure 3D object. Each medial surface is sampled using a grid of loci called medial atoms, indicated by the small circles in figure 1.4. Each atom is described by its position, and the orientation and length of two thickness vectors, or spokes, emanating from the atom and terminating at the object’s surface. Thus, each atom implies the object’s surface at the endpoints of its spokes. Each medial atom lies on an intrinsically 2-dimensional surface, and so is indicated using a set of 2-dimensional integer coordinates. Given a set of such coordinates \( (i, j) \), we can retrieve from the corresponding medial atom its location in 3-dimensional space, \( p \in \mathbb{R}^3 \); the thickness of the object at that location, \( r \in \mathbb{R}^+ \); and its two spoke directions, \( U^+, U^- \in S^2 \) (where \( S^2 \) refers to the space of all coordinates lying on the 2-sphere). Due to the inclusion in each atom of the spatial location \( p \in \mathbb{R}^3 \), in a global coordinate system, the M-Reps representation is not invariant to translation and rotation; it mixes information about shape and pose\(^1\) information. The encoding of the spatial coordinates relative to a global system results in poor localization of gross bending transformations of an object; we delve into this problem and propose a solution in chapter 4.

Next, we turn to the problem of how to compute a medial shape representation of a segmented anatomical structure. This is usually a more complicated operation, compared with computing boundary-based descriptors, since an object’s medial axes are not explicitly given by its segmentation. Computing an M-Reps representation of an object is an optimization process involving the initialization of a fixed-topology skeletal structure within the object, followed by its iterative deformation until it fits the structure, lying medial to it and accurately implying its boundary [155]. However, historically, work on medial shape representations began with Blum’s proposal of the medial axis transform (MAT) for extraction of medial shape features [21, 22]. The MAT of a shape is defined as the loci of the centers of the maximal disks lying within the object, and their radii. A maximal disk is defined as one

\(^1\)In the shape analysis community, the pose of an object is defined according to the parameters of rigid translation and rotation in the coordinate space [55]. In performing a study of shapes, we are not interested in comparing pose parameters. We must therefore normalize the data set to eliminate variations due to pose prior to analysis, usually using Procrustes analysis (see appendix A) [55].
Figure 1.4: To compute a M-Reps medial shape representation, surfaces medial to the object surface are computed, and then evenly sampled to obtain medial loci, or *atoms* in M-reps nomenclature, indicated by the small circles. Multi-part objects require multiple medial sheets, as indicated in the above two-part object.
which lies entirely within the object and is tangent to the object’s boundary at more than one point. Given a collection of such loci and radii, one can reconstruct the original shape boundary. Another way to think about the MAT is using the grassfire analogy. If a grassfire were lit at the boundary of the object and proceeded to burn inward at uniform speed, in a direction normal to the object boundary, the loci of the MAT are specified by the points at which the flame front self-extinguishes. More formally, the MAT is represented by the shocks of a partial differential equation whose boundary conditions are given by the object boundary.

A useful property of the MAT is that any given shape has exactly one medial representation as given by the MAT, and given a set of loci and radii, the inverse of the MAT reconstructs exactly the one object boundary that generated this MAT. This is an attractive property, for example, if one wishes to use medial representations for object retrieval, since no two MATs imply the same object boundary. However, the use of the MAT for computing medial shape representations carries a significant drawback: medial representations given by the MAT are extremely sensitive to small boundary perturbations. Figure 1.5 illustrates this problem, showing the effect on the result of the MAT of adding a small bump to the surface of an object. The medial representation is drastically altered, with a very different branching structure. If one were performing object retrieval based on the MATs of these two objects, one might conclude that they were quite different objects when in fact they are identical except for the small bump. This difficulty led to the choice in M-reps to use a priori knowledge to fix the topology of the medial representation, and use an optimization scheme to fit a skeleton to an object. Although this is a very practical approach that avoids the instability problem in the MAT, it does not handle abnormal shapes very well. Given, for example, a medical data set where abnormal patients differ from normal patients by some topological variation in their medial axis representations, it is impossible to choose a priori a single, fixed topology for the medial representation. One would need to choose multiple topologies for the different pathology groups and apply the right topologies to the right groups, defeating one of the purposes of using anatomical shape analysis, which is to automatically discover shape variations between normal and abnormal populations. We address this problem in chapter 5.

Voronoi methods for computing skeletons of 2D and 3D objects have also received considerable attention [8, 30, 144, 145, 165]. The basic idea is to compute the Voronoi diagram of the set of points sampling the object’s boundary, yielding a set of Voronoi edges. After
removing all Voronoi edges that cut across the object’s boundary, the remaining edges form an approximation to the MAT. It is been shown that as the number of points sampling the boundary goes to infinity, the Voronoi skeleton becomes the MAT [190]. Voronoi skeletons have the added advantage of requiring no connectivity between boundary points. This is highly useful if the input data is in the form of a point cloud, for example from a range camera, or from points sampled from an anatomical structure. It is also useful for boundaries that are disconnected, for example a set of edges resulting from edge detection on a grayscale image.

Another main category of approaches for producing skeletons is that of thinning algorithms [7, 115, 151, 232]. All such algorithms are based on layer by layer erosion of the binary image of the object, where pixels are only removed if their removal would not change the topology of the object implied by the remaining skeleton. The algorithms mainly differ in their heuristics for determining the removability of a pixel, and a substantial effort has gone into developing fast thinning algorithms. Speed is their main advantage, and their main disadvantage is that they are not guaranteed to produce skeletons which are truly medial, unlike Voronoi approaches and the MAT.

Gorelick et al. [76] use an approach that fits in this section because, although not strictly medial, it is an interior representation of an object that has qualities not unlike a medial representation. Their idea is to compute, for every point internal to the object, a value indicating the average time required for a random walk from that point to any point on the boundary of the object. They compute these values by solving Poisson’s equation, with boundary conditions provided by the object’s boundary. By examining the curvature of the level sets of the Poisson equation solution passing through each point inside the shape,
they are able to extract a rough skeleton-like structure. These skeletons, however, are often disconnected and may be of questionable utility. The authors do show, however, an ability to extract a hierarchical, parts-to-whole shape representation from their method, as well as detecting useful entities/quantities such as corners, and the local orientation and aspect ratio of a shape.

Another, quite different, approach to computing medial representations of 3D objects is the curve skeleton [41]. Whereas, in general, the skeleton (MAT) of a 3D object consists of both surfaces and space curves, the goal in computing the curve skeleton of a 3D object is to obtain a 1 pixel thick skeletal structure, where the skeleton consists entirely of connected space curves. Such a representation is advantageous because it captures the essential topology of an object in a very compact form. It is very useful in medical applications for quasi-tubular structures such as airways, blood vessels, and the spinal cord and canal, since although the true skeletons of these structures may contain some narrow surfaces, it is more natural to think of their skeletons as curves. Curve skeletons also have applications in animation and virtual navigation. However, the drawback to using this approach is that the curve skeleton is not very well defined. Since it is by definition a subset of the medial axis transform, there can be many different “legal” curve skeletons for a single object, and different algorithms for computing curve skeletons will arrive at different results. Cornea et al. [41] give several helpful properties that curve skeletons should obey, in addition to being thin. First, they should be homotopic to the shapes they represent. Second, the curve skeleton of an object should be invariant to isometric transformations of the object. That is, if an object $O$ is transformed by an isometric transformation $T(O)$, and the curve skeleton of $O$ is $C$, then the curve skeleton of $T(O)$ should be $T(C)$. Third, a curve skeleton should allow reconstruction of its generating object, insofar as it is possible to do so. Again, since the curve skeleton is a subset of the MAT, it is not possible to perform a faithful object reconstruction from the curve skeleton. However, one can measure the accuracy of reconstruction and judge one curve skeleton to be better than another if it yields a better reconstruction. Fourth, a curve skeleton should be centered within the object. Fifth, every boundary point on the object should be visible from at least one point on the curve skeleton. Sixth, the parts of the curve skeleton, as separated by its junction points, should logically reflect the parts of the object from which it was computed. Seventh, the curve skeleton should be robust to noise, giving it an advantage over the MAT.
Despite the open problems in stably computing medial shape representations, we continue to seek good solutions because the results of shape analysis using these constructs can be intuitively understood by the non-expert. The ability to report shape variations in intuitive, object-centered terms such as bending, thickness, and elongation raises the possibility of making such analysis useful for physicians in practice. Also, because object surfaces can be reconstructed from their medial representations, it is possible to report on object surface characteristics (e.g. local curvature) as well as on the overall object-centered features given above [46, 92].

1.4.4 Other approaches

There exist a large number of alternative approaches to shape description. Although, to the best of the author’s knowledge, no recent survey of such methods exists, [70] provides a useful, albeit brief, overview and reference list. Here, we examine two alternative approaches which are quite different from those covered thus far, as a means of broadening the spectrum of the present discussion.

An interesting approach that shares many elements of the spherical harmonics approach described in section 1.4.2 is described in [70]. They represent a 3D shape by first constructing a set of spherical shells centered at the center of mass of the shape, with equally spaced radii. These shells are represented as spherical functions, with the values of the functions set according to intersection with the object at each point on each shell. Thus, if there are \( R \) shells constructed, then there are \( R \) spherical functions representing the intersections of those shells with the 3D object. The spherical Fourier transform [94] is then used to compute the coefficients of the spherical harmonic basis functions for each of the shells, and these \( R \) sets of coefficients form the shape representation. Whereas the spherical harmonics-based approach described in section 1.4.2 is the 3D analogue to the Fourier description of contours embedded in 2D, the present approach provides a means of describing the shape’s interior as well.

Another approach to shape modeling that is quite different from those discussed so far is the treatment of shapes as solids [12, 181]. When these are constructed out of a set of primitives, this approach is referred to as constructive solid geometry (CSG) [14, 15, 230]. The idea is to begin with a set of primitives, such as cuboids, cylinders, prisms, pyramids, spheres, and cones, and construct objects from them according to set operations. For example, one can position two primitives such that they overlap, and then take the set
CHAPTER 1. INTRODUCTION

Figure 1.6: The corresponding coordinates (correspondence indicated by the colours of the landmarks) are used to form a matrix, where each row depicts a single shape. Note that coordinates in a given column of the matrix all correspond (have the same colour). Principal component analysis is performed on this matrix, yielding the mean shape, and the eigenvectors and eigenvalues corresponding to the major directions and magnitudes of observed shape variation.

union of them to produce an object filling the space that was occupied by either of the two primitives. Or, one can do the same but take the set intersection, producing an object filling the space that was occupied by both primitives. The set difference operation can be used to subtract one primitive from another as well. A shape is then represented as the set of transformations that position, rotate, and scale primitives in space, and the set operations that combine the primitives to form the desired shape. The CSG approach is useful because it is straightforward to ensure that the resulting shape is “water-tight”; this is less straightforward with surface-based representations. To the best of the author’s knowledge, however, this type of shape representation has not been used in anatomical shape analysis; its primary applications are in graphics and engineering.

1.5 Shape correspondence

Given a set of pose-normalized shapes, one wishes to build a statistical model describing the observed variability amongst those shapes. Broadly, this involves three steps, depicted in figure 1.6. The first step is the establishment of correspondence between each surface point on each shape and a surface point on every other shape in the set (the coloured points on the left of figure 1.6 indicate, very sparsely here, a set of corresponding landmarks). Second, the corresponded point coordinates are used to construct a set of points in a higher-dimensional shape space. Third, the mean shape is computed from these points in shape
space, and the eigenvalues and eigenvectors of their covariance matrix give the major modes of observed shape variation (more details on this type of statistical shape modeling are given in appendix A).

The question of how to establish correspondence between surface landmarks is called the shape correspondence problem. Without correctly established shape correspondence, statistical models of shape carry very little meaning and are of limited utility. It is important that established correspondences are homologous, and there are various definitions of homology for this purpose [24], illustrated in figure 1.7. Type I landmarks are those which carry biological significance, such as interface points between different tissues. These landmarks can usually be named in accordance with anatomic labels given in standard anatomy references. Type II landmarks are defined geometrically and may not directly correspond to a specific anatomic location. For example, we might select a point with greatest surface curvature in a region, or an apex of a region, as a type II landmark. Type III landmarks are those which are distributed with even spatial distances between type I and type II landmarks. In figure 1.7, the junction point between the parietal, sphenoidal, and temporal skull bones is a type I landmark because it corresponds to an unambiguously defined, anatomically meaningful point on the skull. The mandible apexes are not type I landmarks since they do not have a specific anatomic meaning, but they are defined geometrically in such a way as to be reliably found on each specimen, so they are type II landmarks. The evenly-spaced landmarks between the type II landmarks are type III landmarks since they do not correspond to any specific geometric features, but instead are placed relative to type I and type II landmarks (type II landmarks only, in this case). An ideal shape correspondence will correctly map all homologous type I and type II landmarks to each other across all shapes, and then map some evenly distributed type III landmarks according to the user’s desired density of landmarks. It is especially critical that type I landmarks be mapped correctly, as these landmarks are most strongly homologous for medical and biological applications. Failure to correctly map type I landmarks can easily invalidate a medical study drawing conclusions from the resulting statistical analysis of shapes.

Most methods for approaching the shape correspondence problem rely on the minimization of some energy function whose domain is the set of possible correspondences between shapes. Usually, this function has a component measuring fine-scale differences between corresponding points, and another, regularizing component to penalize correspondences that cause unwanted tearing or folding of one shape as it is warped to correspond to another.
Figure 1.7: A type I landmark (red dot) at the junction of the parietal, sphenoidal, and temporal skull bones. Type II landmarks (orange dots) at the lower anterior and posterior apexes of the mandible (jaw bone). Type III landmarks (brown dots) equally spaced between the two type II landmarks. Skull diagram is based on that provided by Gray’s Anatomy [79].
There are several approaches to optimizing these cost functions, including greedy optimization as in the approach of Jain et al. [105], genetic optimization as in Davies et al. [52], dynamic programming as in the work of Scott et al. [179], and ant colony optimization as in the work of van Kaick et al. [215]. The correspondence problem can also be formulated as a bipartite graph matching problem to be solved by combinatorial optimization [152].

Template-based methods form one category of approaches to this problem. These methods parameterize a reference shape, and then establish correspondence between this template and all of the shapes in the study. Bookstein et al. and Wang et al. align each shape to a template using generalized Procrustes analysis, and then slide landmarks along the shape contour until a thin-plate spline (TPS) bending energy function is minimized [23, 218]. The TPS energy is computed by considering that the template landmarks lie on a 2D grid, and observing the amount of deformation required of this grid in order to align the template’s landmarks with those of the new shape. Heitz et al. also use a template-based approach, meshing a template shape, and then deforming the template mesh to segment a new structure [97]. They promote their method as landmark-free, since correspondence is established at each mesh vertex between the original template and the deformed version. These vertices could, in effect, be considered as type III landmarks, but it is not clear which of them would be type I or II landmarks. Vos et al. also use a template- and mesh-based method, but use an iterative closest point technique [19] to find a correspondence between mesh vertices.

Another class of approaches uses feature-wise matching of points to form a correspondence. Tagare [204] optimizes an objective function that measures local curvature and arc length at each landmark, and seeks a correspondence yielding a best match of these features. Shen et al. [187] also use a feature-based approach (called Hierarchical Attribute Matching Mechanism for Elastic Registration, HAMMER), computing an attribute vector that measures aspects of the image locally at each point, and then performing elastic registration of images based on these attribute vectors. Although this is an approach to image registration (a survey of image registration methods is out of the scope of this dissertation; please see [128, 160, 239] for surveys), the result of registration is an establishment of correspondence between image pixels and therefore also points on the contours of the shapes in the images. This approach is of interest to us due to the use of locally computed features to drive the registration/correspondence process, and is related to our work in chapter 3. Zheng et al. [236] use more elaborate features called shape contexts, which are log-polar (i.e. dartboard-like) histograms, attributed to Belongie et al. [17]. They establish correspondence
in two passes. In the first pass, they compute shape contexts, at all landmarks, and perform landmark matching based on the best match of these histograms. In the second pass, they use relaxation labelling [172] in order to enforce that if landmarks $x_i$ and $x_j$ in shape $X$ are mapped to landmarks $y_i$ and $y_j$ (respectively) in shape $Y$ and $x_j$ is a neighbour of $x_i$, then $y_j$ should be a neighbour of $y_i$ (i.e. neighbouring landmarks should map to neighbouring landmarks). Jain et al. [105] compute a geodesic shape context for shape matching, where the bins are distributed geodesically along the shape contour. Grauman et al. [78] use an approximation to the earth mover’s distance (EMD) in order to measure the feature-wise mismatch between contours, and achieve contour matching. The basic idea behind EMD is to view the process of transforming one weighted point set into another analogically to that of transforming one pile of dirt into another, and measuring the work required to do so according to the amount of dirt moved, times the distance that it was moved. Scott et al. [179] augment an approach based on bipartite matching of features extracted from contour points, in order to ensure that the matching preserves order. Van Kaick et al. [215] note that order preservation is not sufficient; it is also important to preserve proximity. That is, if point $P$ is close to point $Q$ on a contour, point $P' = \pi(P)$ should be close to point $Q' = \pi Q$ where $\pi$ is a mapping of points from one contour to another. They therefore formulate the matching as a quadratic assignment problem, and use ant colony optimization to solve it.

Another class of approaches is based on diffeomorphic image registration, as in the work of Cootes et al. [40] and Joshi et al. [107]. A diffeomorphism, in this context, is a function that maps coordinates of one image into coordinates of another image, such that the function itself is smooth and invertible. These properties are desirable because they prohibit undesirable folding and tearing of one image as it is warped into the other, preserving neighbourhood information surrounding corresponding points. However, these approaches are typically very expensive to compute.

Hill et al. use a hierarchical approach to correspondence establishment, building a shape hierarchy by identifying and eliminating non-critical points [237] at each step of the hierarchy [99]. Their approach is interesting in that in addition to obtaining good correspondences between landmarks, it generates the landmarks themselves. It was demonstrated in their paper that the generated landmarks are very similar to those placed by a human expert.

In section 1.4.2 we briefly discussed the idea of using the low-order harmonics of a spherical harmonics-based shape representation in order to establish correspondence across
CHAPTER 1. INTRODUCTION

shapes. This idea is attributed to Brechbuehler et al. [31], wherein they find a mapping from all of a 3D object’s landmarks to points on a sphere, such that areas of polygons specified using landmarks as vertices are preserved in the mapping. This permits the use of spherical harmonic basis functions for describing the shapes, and the low-order harmonics are used for aligning the shapes since they represent the shape most coarsely. Once the coarse shapes are aligned, correspondence is established in the spherical domain.

A different perspective on the shape correspondence problem is to look to an information theoretic measure of the statistical shape model resulting from the correspondence, and use it as an indirect measure of the correspondence itself, as in the work of Kotcheff et al. [109]. They optimize to find a correspondence yielding a model that concentrates the observed shape variance into a few modes with large variances, by measuring the determinant of the covariance matrix. Davies et al. have devised a similar method that measures the statistical shape model constructed from the correspondence. Called the minimum description length (MDL) approach, it measures the size of the model, in the information theoretic sense, as the number of bits required for a sender to transmit the model to the receiver [48, 49, 51, 52]. An optimizer (a genetic optimizer in the original paper [52]) seeks the correspondence yielding the shape model transmittable using the smallest number of bits. The MDL approach is based on the Occam’s Razor principle, which suggests that the simplest model for describing shape variation is the best. Thodberg et al. [207] have extended this approach to include the use of local curvature features in the model, in order to establish correspondence between points where curvature is the salient feature, and this approach has received considerable attention by several groups [50, 60, 59, 193, 208].

Styner et al. [197] evaluated the correspondence establishment methods of Kotcheff et al. (evaluating the correspondence by examining the determinant of the covariance matrix of the result statistical shape model) [109], of Brechbuehler et al. (using low order spherical harmonics to establish correspondence) [31], and of Davies et al. (evaluating the correspondence by examining the description length of the resulting statistical shape model) [52, 51, 49, 48]. The evaluation of the correspondence of these three methods was against that established manually by experts on the femoral head and the brain ventricles. The result was that the method based on spherical harmonics consistently trailed the other two methods, but correspondence errors were between 3 and 4 mm for the best methods. This is a substantial correspondence error given the overall sizes of the femoral head and the brain ventricles. Such errors could severely disrupt statistical analysis of shape models
built using these correspondences, suggesting a need for a more accurate approach to the problem of shape correspondence. It should also be noted that inter-expert variability in establishing correspondence was 1.9 mm for the brain ventricles, and 2.5 mm for the femoral head; this high variability motivates the need for more consistent, automatic and semi-automatic methods for correspondence establishment. Ericsson and Karlsson [61] propose a benchmarking approach to shape correspondence evaluation which, although useful in principle, is sensitive to the accuracy of the manually-established ground truth correspondences against which algorithms are measured.

1.6 Outline of this dissertation

Figure 1.8 gives an overview of work presented in this dissertation. The gray blocks depict a general pipeline for medical research studies investigating the relationship between anatomical shape and pathology. The process begins with the formation of a medical hypothesis about such a relationship. This hypothesis drives the collection of the needed medical image data for the experiment; at this stage decisions are made about the number of data sets required, the criteria for patient selection, parts of the body to be imaged, and the imaging parameters to be used. The next step is to obtain segmentations of the anatomical structures of interest from each of the collected data sets. In figure 1.8, this step straddles the boundary between computing and medicine, since segmentation can be done semi-automatically, relying on medical personnel to provide input. Next, a representation of the shapes of the segmented structures is computed and stored. If the representation includes local shape measures, then shape correspondence must be established across the group of patients. We can then compute a statistical model of observed shape variability and/or use shape measures to train a classifier for computer-aided diagnosis. This can be used to test the hypothesis regarding the relationship between shape and pathology, to draw an appropriate conclusion from the result of the test. The conclusion may lead to a refinement of the hypothesis, commencing another cycle of this process. The blue blocks in figure 1.8 indicate our contributions, and the dotted lines indicate the steps (gray blocks) that each of these contributions supports. The remainder of this dissertation will describe the contributions in each of these blue blocks, within the chapters indicated.

The high-level progression of ideas in this dissertation is as follows:

- We begin in chapter 2 with two medical research studies, investigating the relationship
Shape measures can distinguish pathologies. Local regions of shapes must meaningfully correspond.

Medical image data (e.g. CT, MRI) directs data collection. Hypothesis tests a relationship between anatomical shape and rate of injury/pathology. Yields a statistical significance of the relationship. Conclusion leads to formation of refined hypothesis, process repeats.

Figure 1.8: A chart depicting the components of the research presented in this dissertation. The gray blocks depict a general pipeline for conducting studies of anatomical shape. The blue blocks depict the contributions of this dissertation and the dashed lines indicate the general pipeline steps to which the contributions are related. The yellow blocks in the green frame show some components of computational intelligence on which this dissertation is built.
between the shapes of shoulder structures, and the pathological conditions of those and related structures. Section 2.1 describes our work on the shape-based assessment of the supraspinatus muscle of the shoulder, in order to assist the radiologist in determining the need for surgery in a noninvasive manner. In this study, we employ global measures of shape, and train a classifier to distinguish the pathology groups. Section 2.2 describes our work on the shape-based assessment of the bicipital groove of the proximal humerus, which is a trough in the head of the upper arm bone. We assess the relationship between its shape and the propensity for injury of the long biceps tendon that sits within it. In addition to using global shape measures in this study, we also describe the computation and use of an anatomy-specific, medial-based shape representation for the bicipital groove, which captures aspects of shape that are of interest to the medical community studying this structure. These studies, in addition to contributing to medical research, serve to motivate our development of novel techniques in medial shape representation and computation, shape correspondence, and medial axis pruning in chapters 3, 4, and 5, respectively. It was evident when we conducted these studies that they and other similar studies would benefit from improved approaches to shape description and correspondence, and these studies serve as a strong practical motivator and reminder of the potential human health impact of this research.

- In chapter 3, we describe our work on the shape correspondence problem, wherein we utilize a machine learning approach to automatically elicit shape and appearance features characterizing important anatomical landmarks. The automated learning approach is based on a training set provided by an expert; thus, our approach incorporates, in an implicit manner, expert knowledge into the shape correspondence process.

- In chapter 4, we examine two problems associated with computing medial shape representations: how to represent the medial shape information intuitively, and how to compute the basic components, namely a parameterized medial surface and thickness vectors emanating from it, for any medial shape representation. We introduce our medial patch shape representation, which addresses a problem with M-Reps wherein shape deformations and differences are not encoded in a localized way. Using our shape
representation, gross deformations are encoded directly and only at their source, making comparison of such deformations across shapes straightforward. We also describe our manifold learning-based approach to medial and object surface parameterization, and show that it can be used to compute the basic building blocks of both the medial patch and M-Reps representations.

- In chapter 5, we describe our novel, groupwise approach to pruning unwanted branches from medial shape representations. Our philosophy is that when presented with a group of shapes, as is almost always the case in medical shape studies, information from the entire group should be used in order to decide on which branches of a medial representation to prune. We show that shapes from the same class contain substantial redundancy (in the information theoretic sense), and we posit that this redundancy arises from the shape information that these shapes have in common. Thus, this redundancy yields a way to determine which branches arise from the true, uncorrupted shape of an object, and which arise from noise and should be pruned.

- Each of the above chapters concludes with a section which summarizes the contributions of the chapter, discusses the limitations and failure points of the work, and points the way to interesting future work based on the chapter’s contents. Chapter 6 provides an overall summary of the contributions described in this dissertation, with references to the specific publications arising from each chapter.
Chapter 2

Motivational studies of shoulder shape vs. pathology

This chapter describes two studies of anatomical shape of structures in the human shoulder. The purpose of this chapter is to illustrate the practical importance and utility of applying shape analysis techniques to musculoskeletal problems, and to motivate the design and development of more sophisticated shape analysis techniques than those used here; such techniques will be the subject of later chapters. Section 2.1 describes a study of the use of global 3D shape descriptors in examining the relationship between shape and pathology of a rotator cuff muscle. Section 2.2 describes a study of the relationship between the bicipital groove of the proximal humerus (a groove in the head of the upper arm bone) and pathological conditions of the long biceps tendon, which sits in the bicipital groove. This study explores both global and medial shape representations of the bicipital groove, providing a practical motivation for the further study of medial shape representations in chapters 4 and 5.

2.1 Shape analysis of the supraspinatus to determine the need for surgery

2.1.1 Motivation and hypothesis

The rotator cuff comprises several muscles and tendons that stabilize the shoulder, including the supraspinatus muscle, indicated in figure 2.1. Disorders of the rotator cuff are
prevalent; incidence of disorder has been found to be 34% in asymptomatic individuals in a study where diagnosis was performed on MR images [188], and 30% of individuals over 60 years of age in a cadaveric study [119]. The symptoms of rotator cuff disorder can be debilitating, including pain, weakness, and limited range of motion, especially for overhead work [57, 68, 71]. Disorders of the supraspinatus muscle often involve tearing, which can lead to muscle retraction, atrophy, or both [84, 103, 169, 209]. It is particularly important to be able to distinguish between retraction and atrophy because retraction is a condition that is repairable by pulling the muscle forward in surgery, whereas atrophy is a condition uncorrectable by surgery. Since both of these conditions result in a reduction of the apparent size of the muscle and therefore are difficult to distinguish by size alone, it is important to investigate the utility of analyzing the 3D shape of the supraspinatus to discover shape characterizations that may assist the physician in distinguishing between these groups. So, the hypothesis under investigation is: “There exist 3D shape characteristics of the supraspinatus muscle that can be used to distinguish the normal supraspinati from the pathological, with specific pathologies of interest being tearing, atrophy, and retraction of the muscle.”
In order to test the above hypothesis, the specific objectives of this study are to explore the relationship between shape and pathology of the supraspinatus, and to carry out a preliminary study of the effectiveness of an automated machine learning algorithm for distinguishing the pathological cases of the supraspinatus based on its shape \[223, 224\].

At a high level, we proceed as follows. The 3D surfaces of the supraspinatus muscle were segmented by an expert from magnetic resonance imaging (MRI) data of 73 patients. The patients were then grouped by experts according to a quantitative procedure to label each patient’s data set according to pathology; the procedure is described in \[67, 234\]. Since shoulder arthroscopy (insertion of an optical camera into the shoulder through an incision, for visual inspection) is considered to be the gold standard for evaluation of the rotator cuff, all MRI-based diagnoses were also verified using shoulder arthroscopy \[10, 162, 202, 240\]. We then compute eleven different shape descriptors of each patient’s segmented supraspinatus surface. We perform an analysis of variance (ANOVA) testing the null hypothesis that the population means of all of the pathology groups are the same, for each measurement. We select the most discriminatory measurements to train a support vector machine-based classifier to automatically assign new supraspinatus surfaces into the correct pathology groups. We evaluate our results according to two criteria: (1) the \(p\) values reported by the ANOVA test, which indicate the statistical significance of the differences in the measurements across pathology groups, and (2) the accuracy of the trained support vector machine in classifying new supraspinati, which gives a preliminary indication of the utility of this approach in designing a computer-aided diagnosis (CAD) system for the supraspinatus. Ultimately, the predicted clinical use of such a system would be to act as a second, supporting reader to a radiologist attempting to make the diagnosis based on MRI.

2.1.2 Description of patient data

T2-weighted MR images of the shoulder were acquired from 73 patients at 1.5T (1.5 Tesla), with imaging parameters as follows: repetition time 4000-5500; echo time 36; field of view 14. The in-plane (sagittal) resolution of the data was between 0.3 mm and 0.6 mm, with a matrix of \(256 \times 256\), and the slice thickness ranged from 3 to 5 mm. Because these data sets were acquired from real patients in a clinical setting, the resolutions of the data sets were not

---

consistent across all patients. Patients were consistently imaged in the supine position (i.e. lying on their backs), relaxed, and with the arm in minimal external rotation (i.e. minimal rotation about the axis of the upper arm bone, away from the midline of the body) in order to minimize the effects of pose and gravity on the shape of the muscle. The patients were selected according to diagnoses made by examining the MR images of the shoulder. The pathology groups, and the number of patients in each group, are summarized in table 2.1.

2.1.3 Method

The overall processing performed on the 73 MR images is given in figure 2.2, and is described in the following subsections.

Muscle segmentation

Due to the low contrast with surrounding tissues on many slices, expert anatomical knowledge is required to segment the supraspinatus from the rest of the MR volume. Expert manual segmentation of the supraspinatus muscle was performed on the sagittal MR images in a slice-by-slice manner, taking between 3 and 10 minutes per patient. The software tool used for the outlining was custom-written by me for this purpose (see figure 2.3) and allows the resident to select control points lying on the surface of the supraspinatus on each slice. The tool automatically fits a parametric cubic spline curve \[143\] to these points to guarantee smoothness, although sharper turns in the curve are made possible by allowing the user to add more control points near the turns. The tool performed the spline fitting interactively (i.e. every addition, deletion, and modification of a control point updates the

<table>
<thead>
<tr>
<th>Pathology Group</th>
<th>Abbr.</th>
<th>No. Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal: No pathology</td>
<td>N</td>
<td>14</td>
</tr>
<tr>
<td>Abnormal: T, TA, TR, TAR</td>
<td>A</td>
<td>59</td>
</tr>
<tr>
<td>Abnormal Subgroups</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tear: Full/partial tear</td>
<td>T</td>
<td>20</td>
</tr>
<tr>
<td>Tear + atrophy</td>
<td>TA</td>
<td>13</td>
</tr>
<tr>
<td>Tear + retraction</td>
<td>TR</td>
<td>15</td>
</tr>
<tr>
<td>Tear + atrophy + retraction</td>
<td>TAR</td>
<td>11</td>
</tr>
</tbody>
</table>

Table 2.1: Descriptions of the pathology groups.
Figure 2.2: Computational pipeline used in this study. Slice-by-slice segmentation of the supraspinatus from MR images leads to a set of sparse surface points which are interpolated to create a 3D surface. Each muscle is assigned to a pathology group by an expert, and 3D shape measures are computed for each group. Group difference significance is then assessed using ANOVA testing and training and evaluation of a machine learning algorithm for automated diagnosis.
CHAPTER 2. MOTIVATIONAL STUDIES OF SHOULDER SHAPE

Figure 2.3: Screenshots of the custom tool for supraspinatus outlining. (a) Control points define a spline surrounding the muscle. (b) The user can correct a mistake by clicking a point (green) and clicking the position to which to move it (c). (d) Right-clicking anywhere closes the contour at the yellow point.

display of the parametric cubic spline curve immediately) so that the user could manipulate the control points until the curve accurately followed the contour of the muscle. In previous work, it has been shown that intra- and inter-observer variation in supraspinatus contouring is less than 5% [120].

3D interpolation

Due to the 3-5 mm slice thickness of the data, slice-by-slice segmentation results in a set of points which are dense within the (sagittal) imaging planes but sparse in the out-of-plane direction (contours in figure 2.4). The next step is to perform 3D interpolation of the contours to obtain a set of object surface points that is dense along all axes. To this end we used the Interpolation Module of the segmentation editor in the Amira software (Mercury Computer Systems, Inc), based on an implicit, level-set based shape representation, similar to work by Turk and O’Brien [212]. Figure 2.4 illustrates the interpolated muscle surface between the contours. Figure 2.5 shows samples of 3D renderings of three representative muscles from each pathology group. Figure 2.6 places one segmented sample in anatomical context, showing an example of a segmented supraspinatus surface in context of the MR image from which it was segmented.

Expert diagnosis of pathology

Independently of the 3D shape analysis process, the condition of the supraspinatus of each patient was assessed by experts based on MRI and shoulder arthroscopy, and assigned to
CHAPTER 2. MOTIVATIONAL STUDIES OF SHOULDER SHAPE

Figure 2.4: Supraspinatus muscle surface segmentation. Contours resulting from expert segmentation of the supraspinatus are shown, rendered (as dark bands) using physical space coordinates. Large spaces between contours are due to the low out-of-plane resolution of the data. Note that contours appear non-parallel because of perspective projection. Also shown is the result of 3D interpolation of the contours, yielding a dense set of points lying on the surface of the supraspinatus, rendered as a surface. This surface has been made translucent so that the contours can be clearly seen surrounding the muscle.

one of the following groups: normal, tear, tear and atrophy, tear and retraction, and tear and atrophy and retraction.

Computation of shape measures

We computed eleven different 3D shape measures for each data set, summarized in table 2.2:

Ratios of eigenvalues (3 measures): We perform principal components analysis (PCA) [106] (described in appendix A) on the points lying on the surface of each shape. This is done by first subtracting the mean point from all of the shape surface points. Next, a $3 \times 3$ covariance matrix is calculated for these centered points, and the eigenvectors $e_1, e_2, e_3$ and eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ of this covariance matrix are calculated. The eigenvalues $\lambda_1, \lambda_2, \lambda_3$ represent the variances of the supraspinatus surface points in the directions of the eigenvectors (which describe the main directions of variation) for each supraspinatus shape. The eigenvalues give the lengths of the major axes of an ellipsoid approximating the surface of each supraspinatus. Computing the three values of the ratios $\frac{\lambda_1}{\lambda_2}, \frac{\lambda_1}{\lambda_3}$, and $\frac{\lambda_2}{\lambda_3}$ yields measures of elongation of the object. For a spherical object we expect that $\lambda_1 \approx \lambda_2 \approx \lambda_3$. For a cylindrical object we expect that $\lambda_1 \gg \lambda_2 \approx \lambda_3$, and for a disk-like object we expect that $\lambda_1 \approx \lambda_2 \gg \lambda_3$. See figure 2.7 for an illustration of the meanings of these elongation measures.
Figure 2.5: Sample supraspinatus surfaces. Each row shows three samples for each pathology group. Row 1: Normal. Row 2: Tear. Row 3: Tear and atrophy. Row 4: Tear and retraction. Row 5: Tear and atrophy and retraction. The surfaces are shaded according to the computed mean curvature at each point; lighter shading corresponds to higher curvature values.
Figure 2.6: Supraspinatus surface, in context. Four views, from different angles, of the same segmented supraspinatus, in context of the MR image from which it was segmented. (a) The semitransparent sagittal slices viewed face-on, with the supraspinatus rendered as a hollow region. (b) Image (a) rotated slightly for a different perspective. (c) A view from the side of the shoulder. (d) A view from the top of the shoulder.

<table>
<thead>
<tr>
<th>No.</th>
<th>Description of Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Eigenvalue ratio $\lambda_1 / \lambda_2$</td>
</tr>
<tr>
<td>2</td>
<td>Eigenvalue ratio $\lambda_1 / \lambda_3$</td>
</tr>
<tr>
<td>3</td>
<td>Eigenvalue ratio $\lambda_2 / \lambda_3$</td>
</tr>
<tr>
<td>4</td>
<td>Mean of distances to centroid (cm)</td>
</tr>
<tr>
<td>5</td>
<td>Standard deviation of distances to centroid (cm)</td>
</tr>
<tr>
<td>6</td>
<td>3D moment $J_1$</td>
</tr>
<tr>
<td>7</td>
<td>3D moment $J_2$</td>
</tr>
<tr>
<td>8</td>
<td>3D moment $J_3$</td>
</tr>
<tr>
<td>9</td>
<td>Surface area (cm$^2$)</td>
</tr>
<tr>
<td>10</td>
<td>Volume (cm$^3$)</td>
</tr>
<tr>
<td>11</td>
<td>Surface area / Volume (1 / cm)</td>
</tr>
</tbody>
</table>

Table 2.2: Descriptions of the measurements taken, with their associated measurement numbers used in this section.
Figure 2.7: Illustration of the meaning of the relationship of the eigenvalues (resulting from PCA on shape surface points) to each other. (a) PCA on a set of 2D points: Eigenvalues $\lambda_1$ and $\lambda_2$ represent the lengths of the vectors along the directions of greatest and second-greatest variation in the data points, respectively. (b) When PCA is done on a set of points lying on a sphere-like shape, eigenvalues are approximately equal. (c) When PCA is done on a set of points lying on a cylinder-like shape, $\lambda_1$ tends to be much larger than $\lambda_2$ and $\lambda_3$, both of which are approximately equal. (d) PCA on a set of points lying on a disk-like shape yields $\lambda_1$ approximately equal to $\lambda_2$, both of which being much larger than $\lambda_3$.

Mean and standard deviation of distances to centroid (2 measures): Here, we compute the centroid of all of the supraspinatus surface points, and compute the mean Euclidean distance from each surface point to this centroid, as a measure of size. We also compute the standard deviation of these distances as a measure of non-sphericity or surface irregularity. This measure, although capturing non-sphericity in some sense, is different from that given by the ratios of eigenvalues, however, because a rough surface that, grossly speaking, is spherical, will give $\lambda_1 \approx \lambda_2 \approx \lambda_3$, but the measure described here will be large.

3D moment invariants (3 measures): We compute three 3D moments, $J_1, J_2, J_3$, that have been shown to be invariant to translation and rotation. Moments are measures that capture characteristics of the spatial distribution of the mass of a shape [176]. They are computed as

$$ J_1 = \mu_{200} + \mu_{020} + \mu_{002}, \quad (2.1) $$

$$ J_2 = \mu_{200} + \mu_{020} + \mu_{002} + \mu_{200}\mu_{002} + \mu_{020}\mu_{002} - \mu_{110}^2 - \mu_{101}^2 - \mu_{011}^2, \text{ and} $$
\[ J_3 = \mu_{200} \mu_{020} \mu_{002} + 2 \mu_{110} \mu_{101} \mu_{011} - \mu_{002} \mu_{110}^2 - \mu_{020} \mu_{101}^2 - \mu_{200} \mu_{011}^2, \]

where \( \mu_{pqr} \) is given by

\[ \mu_{pqr} = \sum_x \sum_y \sum_z (x - \bar{x})^p (y - \bar{y})^q (z - \bar{z})^r p(x, y, z), \]

\( p \) is given by

\[ p(x, y, z) = \begin{cases} 
1 & \text{if } (x, y, z) \text{ is a surface point} \\
0 & \text{otherwise}
\end{cases}, \]

\( \bar{x}, \bar{y}, \bar{z} \) are given by

\[ \bar{x} = m_{100}/m_{000} \]

\[ \bar{y} = m_{010}/m_{000} \]

\[ \bar{z} = m_{001}/m_{000}, \]

and \( m_{pqr} \) is given by

\[ m_{pqr} = \sum_x \sum_y \sum_z x^p y^q z^r p(x, y, z). \]

**Surface area, volume, and their ratio (3 measures):** We compute the surface area and volume of each supraspinatus in physical units, and take the ratio of surface area to volume. The surface area and volume are computed by extracting a triangular mesh corresponding to the iso-surface at the boundary of the shape surface, using the marching cubes algorithm [125]. The volume contained by the mesh is then computed, and the surface area of the mesh is computed as the sum of the areas of the mesh faces.
Figure 2.8: A plot showing two features, $f_1$ and $f_2$, in 2D feature space. Black points correspond to one group, and white points to another. The points are projected onto the $f_1$ and $f_2$ axes in order to illustrate that the group separability is very poor for either of these features taken separately. However, a classifier based on a quadratic decision curve is able to achieve perfect class separation when both features are considered together.

**Computing group differences**

As a test of the significance of the measurements taken, for each measurement type, we performed a one-way ANOVA to test the null hypothesis that the means of the measurements of the normal and pathological groups were the same. Thus, we are testing eleven separate hypotheses (one for each measurement) on our data set. We therefore need to take care in selecting our statistical significance level to account for these multiple comparisons since, by chance alone, we may encounter a favourable significance level simply due to the number of tests that we are conducting. In this work, the multiple comparison correction we apply is the Bonferroni correction [1]. The Bonferroni correction states that if we test $n$ hypotheses on a set of data, then we should adjust our statistical significance level by a factor of $\frac{1}{n}$. Choosing a standard significance level of $p < 0.05$ for a single experiment, our Bonferroni-corrected significance level in this work is $p < \frac{1}{11}(0.05)$, or $p < 0.0045$.

To determine whether differences observed in the shape measures are substantial enough for practical purposes, we test their classification power in discriminating the pathology groups. To do so, we consider the machine learning problem wherein we attempt to train a classifier to distinguish the supraspinatus pathology groups based on extracted shape descriptors. It is well-known in the machine learning community that although a set of features
may lead to very poor classification performance when each feature is considered individually, considering these features in combination may lead to much better classification. This idea is illustrated in figure 2.8, where we illustrate a simple problem involving two pathology groups (white and black), and two shape descriptors, or features, f1, and f2. It can be seen in the figure that either of f1 or f2, taken individually, would result in very poor classification performance; there is no place on either of these axes where a decision point could be placed that results in good classification. However, taking both features into account, a quadratic decision curve can be computed in 2D feature space that perfectly separates the pathology groups. Based on this idea, we train a support vector machine [64] using the most significant features as determined by the ANOVA test above, and evaluate its classification accuracy in differentiating the supraspinatus pathology groups. We perform leave-one-out cross-validation, averaging the classifier’s accuracy across all rounds.

2.1.4 Results

Table 2.3 gives the mean and standard deviation values of the measurements taken for each pathology group. Measurements whose ANOVA tests favoured the rejection of the null hypothesis ($p < 0.0045$) are shown in table 2.4. These measurements include the mean of distances to centroid, the 3D moment invariants, surface area, volume, and the ratio of surface area to volume.

To visualize the separability of the pathology groups based on the most discriminatory (according to $p$ value) of a variety of the measurements taken, we provide box plots showing the first, second, and third quartiles, the extent of non-outliers, and the outliers for the mean of distances to centroid, the best of the 3D moment features, and the best of the surface area and volume measurements in figure 2.9.

Table 2.5 shows the results of support vector machine classification. At each cell of the table, a value is given that indicates the classification accuracy achieved by the support vector machine in differentiating the pathology groups indicated by the row and column of that cell. This accuracy is computed as a ratio of the number of correct classifications made by the SVM, divided by the total number of classifications in the cross-validation experiment.
CHAPTER 2. MOTIVATIONAL STUDIES OF SHOULDER SHAPE

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>T</th>
<th>TA</th>
<th>TR</th>
<th>TAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.4 ± 1.2</td>
<td>2.6 ± 0.90</td>
<td>4.0 ± 2.4</td>
<td>3.2 ± 1.4</td>
<td>3.0 ± 2.1</td>
</tr>
<tr>
<td>2</td>
<td>5.7 ± 2.2</td>
<td>6.7 ± 2.3</td>
<td>10 ± 5.7</td>
<td>7.6 ± 2.6</td>
<td>7.4 ± 3.3</td>
</tr>
<tr>
<td>3</td>
<td>2.5 ± 0.60</td>
<td>2.6 ± 0.69</td>
<td>2.7 ± 0.75</td>
<td>2.5 ± 0.85</td>
<td>3.0 ± 1.5</td>
</tr>
<tr>
<td>4</td>
<td>2.0 ± 0.44</td>
<td>1.8 ± 0.32</td>
<td>1.8 ± 0.34</td>
<td>1.6 ± 0.29</td>
<td>1.6 ± 0.22</td>
</tr>
<tr>
<td>5</td>
<td>0.65 ± 0.22</td>
<td>0.66 ± 0.13</td>
<td>0.66 ± 0.11</td>
<td>0.56 ± 0.13</td>
<td>0.6 ± 0.17</td>
</tr>
<tr>
<td>6</td>
<td>1.2 × 10^5</td>
<td>8.7 × 10^4</td>
<td>6.5 × 10^4</td>
<td>4.5 × 10^4</td>
<td>3.7 × 10^4</td>
</tr>
<tr>
<td></td>
<td>±9.1 × 10^4</td>
<td>±5.2 × 10^4</td>
<td>±4.2 × 10^2</td>
<td>±2.8 × 10^4</td>
<td>±2.2 × 10^4</td>
</tr>
<tr>
<td>7</td>
<td>5.3 × 10^9</td>
<td>2.7 × 10^9</td>
<td>1.5 × 10^9</td>
<td>6.8 × 10^8</td>
<td>4.6 × 10^8</td>
</tr>
<tr>
<td></td>
<td>±6.4 × 10^9</td>
<td>±2.8 × 10^9</td>
<td>±1.9 × 10^9</td>
<td>±7.3 × 10^8</td>
<td>±5.5 × 10^8</td>
</tr>
<tr>
<td>8</td>
<td>7.6 × 10^13</td>
<td>2.6 × 10^13</td>
<td>1.3 × 10^13</td>
<td>3.2 × 10^12</td>
<td>1.7 × 10^12</td>
</tr>
<tr>
<td></td>
<td>±1.1 × 10^14</td>
<td>±3.7 × 10^13</td>
<td>±2.2 × 10^13</td>
<td>±4.5 × 10^12</td>
<td>±2.8 × 10^12</td>
</tr>
<tr>
<td>9</td>
<td>60 ± 24</td>
<td>49 ± 18</td>
<td>45 ± 22</td>
<td>34 ± 7.6</td>
<td>34 ± 12</td>
</tr>
<tr>
<td>10</td>
<td>23 ± 13</td>
<td>15 ± 8.3</td>
<td>15 ± 10</td>
<td>9.1 ± 2.8</td>
<td>7.1 ± 4.7</td>
</tr>
<tr>
<td>11</td>
<td>2.8 ± 0.56</td>
<td>3.5 ± 0.80</td>
<td>4.8 ± 1.7</td>
<td>4.1 ± 1.3</td>
<td>5.1 ± 0.89</td>
</tr>
</tbody>
</table>

Table 2.3: Mean ± standard deviation values of each of the measurements for each of the groups. Please refer to tables 2.1 and 2.2 for the meanings of the column and row labels, respectively.

<table>
<thead>
<tr>
<th></th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of distances to centroid (cm)</td>
<td>0.0039</td>
</tr>
<tr>
<td>3D moment J1</td>
<td>0.0014</td>
</tr>
<tr>
<td>3D moment J2</td>
<td>0.0018</td>
</tr>
<tr>
<td>3D moment J3</td>
<td>0.0027</td>
</tr>
<tr>
<td>Surface area (cm²)</td>
<td>0.0010</td>
</tr>
<tr>
<td>Volume (cm³)</td>
<td>0.000041</td>
</tr>
<tr>
<td>Surface area / Volume (1 / cm)</td>
<td>0.000034</td>
</tr>
</tbody>
</table>

Table 2.4: p values resulting from a one-way ANOVA test for each measurement, testing the null hypothesis that the means of the measurements of all of the pathology groups are the same. Only p values leading to rejection of the null hypothesis (p < 0.0045) are shown.
Figure 2.9: Box plots showing the separability of the pathology groups using the mean of distances to centroid (top left), 3D moment $J_1$ (top right), and ratio of surface area to volume (bottom). The bottom and top of each displayed box plot indicate the first ($Q_1$) and third ($Q_3$) quartiles of the measurements, respectively. The line in the middle of the box shows the median. The whiskers emanating from the box show the largest and smallest non-outliers. The + symbols show outliers (defined as being smaller than $Q_1 - P$ or larger than $Q_3 + P$ where $P = 1.5 \times Q$, $Q = |Q_1 - Q_3|$).
Table 2.5: Table showing automated classification results. Each cell shows the accuracy (1 being perfect accuracy) of a support vector machine trained to distinguish the pathology groups corresponding to the row and column of that cell. Leave-one-out cross-validation was performed, with the results averaged over all rounds. The support vector machine was trained using shape features corresponding to the three smallest \( p \) values: surface area, volume, and the surface area to volume ratio.

2.1.5 Discussion

Based on our results, we address the main hypothesis by concluding that 3D shape is indeed useful in distinguishing the pathological conditions of the supraspinatus. From the plots in figure 2.9 it can be qualitatively observed that the normal group is generally separable from the abnormals taken as a whole, and also separable from the TAR group. It appears that the surface area to volume ratio provides the best separability between the normal and abnormal groups. This result is interesting because the surface area to volume measurement captures aspects of both shape non-sphericity and size. Geometrically, spheres have the lowest surface area to volume ratio, so it can be said that high values of this measurement indicate non-sphericity. This non-sphericity could be in the form of an elongation of the object. It is arguable in this study that differences in elongation may not be involved, since the elongation measures (ratios of eigenvalues) produced poor group separability (failure to reject the null hypothesis in the ANOVA test). It could also be the case that the separability associated with the surface area to volume ratio is a consequence of differences in the amount of folding in the surface between the normal and pathological groups, increasing its total surface area without a corresponding increase in volume. Informally, the samples shown in figure 2.5 seem to bear this out. However, this hypothesis would require a more detailed investigation before any conclusions could be drawn, especially because the size of an object also affects this measurement; the larger an object (of the same shape) becomes, the lower becomes its surface area to volume ratio. It is less clear how these measurements, taken separately,
would separate the individual groups corresponding to pathological conditions. The SVM classification results from the cross-validation experiment are generally promising, with the exception of the classifier’s ability to distinguish a tear from a tear with retraction, and a tear with atrophy from a tear with atrophy and retraction. Specifically encouraging, however, are the classifier’s ability to distinguish normal from abnormal (79% accurate), and most importantly, tear and atrophy from tear and retraction (82% accurate). Although these accuracies are not sufficient for immediate medical use, they provide optimism that further investigations on the relationship between shape and pathology of the supraspinatus would be fruitful. The seemingly below-chance accuracy of the classifier (44%) in distinguishing torn supraspinati from those which are torn and retracted should be investigated further in order to determine if this was caused by an outlier accuracy score in the cross-validation, or if this is a general trend.

2.1.6 Contributions, limitations, and future work

The main contributions of the work described in this section and published in [223, 224] are:

- **The establishment of a relationship between supraspinatus shape and pathology:** We performed an examination of the relationship between shape and pathology of the supraspinatus, and showed that several global shape features (mean of distances to centroid, 3D moments, surface area, volume, and surface area to volume ratio) were shown to differentiate the pathology groups in a statistically significant manner ($p < 0.0045$ resulting from an ANOVA).

- **The exploration of CAD for the supraspinatus:** We carried out a preliminary exploration of the feasibility of a CAD system for the supraspinatus, based on its 3D shape. We obtained encouraging results from a trained support vector machine attempting to automatically diagnose supraspinatus pathologies. In particular, the classifier’s accuracy in differentiating the tear with atrophy group from the tear with retraction group is very encouraging, since it is this very diagnosis that is critical to determining whether surgery is required (retraction) or not (atrophy), and this diagnosis is difficult to make clinically and noninvasively.

The main limitations of the work described in this section are:
• **Size and shape are confounded:** Several of the shape measures that we computed include an element of size (i.e. are not invariant to scale). This is intentional; domain expertise suggests that aspects of size are important in distinguishing these pathology groups (e.g. atrophy is, by definition, a reduction in size). In a small or poorly-chosen sample of patients, this inclusion of size measures could be problematic. Consider, for instance, a scenario where all of the patients with normal supraspinati happened to be large individuals, and all of the patients with torn supraspinati were small individuals. We might then erroneously conclude that size is a distinguishing factor between these groups, when in fact this is an artifact of our patient selection process. To compensate for this, it may be tempting to normalize for patient size (for example, by using the size of the joint), but this would fail to compensate for differences in fitness and muscle mass across individuals (e.g. large individuals with small muscles and small individuals with large muscles). On the other hand, if all groups consist of a representative cross section of different-sized people, and yet size remains a distinguishing factor between the groups, then it is safe to conclude that the size differences are related to the muscle pathologies. We believe this to be the case in our study, which involves a relatively large group of real patients.

• **Shape descriptors are global:** We use global shape descriptors, and in doing so we discard all information about regional variations in shape. We therefore are unable to discover or report whether the differences between supraspinatus atrophy and retraction pertain to any specific part of the muscle. The positive side of this approach is that it avoids having to solve the shape correspondence problem, eliminating a source of error in the analysis.

Future work based on the work described in this section includes:

• **Deeper investigation:** A deeper investigation of the relationship between shape and pathology of the suprapsinatus requires the use of a richer shape description. This description should be local, characterizing qualities of the shape (e.g. curvature, concavity) at each point on the surface. Doing so requires a solution to the shape correspondence problem, discussed further in chapter 3. A medial shape description could also be used, in order to get object-centric measurements of thickness, bending, and elongation for the supraspinati. This requires a stable means of computing medial shape representations, which is the subject of chapters 4 and 5.
• **Clinical application:** This work represents only the first steps toward the overall goal of the development of CAD for the diagnosis of the supraspinatus. To make such a system useful in a clinical setting, several avenues of future work need to be explored. First, a reliable, highly automated method for segmentation of the supraspinatus on MR images must be developed. Second, the accuracy of the automated classifier needs to be increased to a level that is comfortable to practicing radiologists. With these two goals achieved, a clinical work flow incorporating this work can be envisioned: the patient presents with symptoms of rotator cuff disorder, the primary care physician performs a diagnosis based on palpation and range of motion tests and orders an MRI. The radiologist performs a first reading of the MR image and diagnoses the condition of the patient’s supraspinatus. The CAD system segments the supraspinatus from the MR images, computes 3D shape measures, and performs an automated diagnosis, effectively acting as a second reader. If the system’s diagnosis differs from that of the radiologist, the radiologist is alerted so that he/she may give the diagnosis further consideration to decide whether or not to accept the CAD system’s recommendation. Design of such a CAD system is a long-term goal; the results of this work suggest that CAD based on 3D shape of the supraspinatus is indeed possible with substantial future work.
2.2 Shape analysis of the bicipital groove to determine the predisposition to injury

2.2.1 Motivation and hypothesis

The bicipital groove (BG) of the proximal humerus is a groove in the humeral head (the top of the upper arm bone), formed by the medial and lateral tuberosities (bony protrusions on the humeral head; see figure 2.10). The long biceps tendon (LBT) sits within the BG and connects the biceps muscle to the shoulder; the BG prevents the LBT from dislocating during movement of the arm. The BG and the LBT are intimately related; the shape of the BG has a great impact on the tendency for the LBT to become dislocated, subluxated (partially dislocated), frayed, or torn. It is understood that a shallow, wide BG can promote subluxation and/or dislocation of the LBT; a deep, narrow BG can cause LBT irritation and tenosynovitis (inflammation of the fluid-filled sheath surrounding the LBT); osseous spurs (bony protrusions) in the BG can cause LBT fraying, and the presence of the supratubercular ridge of Meyer (a ridge of bone projecting immediately proximal to the medial wall of the BG and continuous with it) is suspected to promote dislocation [13, 100, 137, 147, 214].

According to our medical collaborator, the clinical accuracy of the diagnosis of LBT disorders by the examination of medical images is very limited. There is therefore a clinical need for an indirect method for the assessment of the LBT. Because of the anatomical relationship between the BG and the LBT, many researchers have devoted efforts to shape analysis of the BG, with the aim of using its shape as an indirect measure of LBT health. Previous studies of BG shape variation primarily involved taking 2D measurements from a single axial cross section of the humerus [3, 100, 153, 214] (figure 2.11). Because of the large variation in BG shape from one axial slice to the next within a single patient (figure 2.10(c) and (d)), such measurements risk overlooking shape features important to LBT pathology. This motivates the development of an approach to the 3D shape description of the BG; such a description is one of the main contributions of this section. Although the shape of the proximal humerus has been studied previously in 3D [102, 168], these studies are primarily concerned with capturing the shape of the humeral head for arthroplasty (improving the function of a joint by reshaping its surfaces, or replacing it altogether). Our aim differs from that of these previous works, and motivates the use of shape descriptors capturing different aspects of BG shape. Our work also differs in that our study is based on in-vivo MR images,
Figure 2.10: (a) Profile view of a humeral head (segmented from an MRI of the shoulder) with tuberosities and BG indicated. (b) View similar to Ahovuo’s “groove view” [3], showing the profile of the BG. (c) An axial slice from a 1.5T MRI showing the BG. (d) A distal slice from the same patient, illustrating the large intra-patient variability in BG shape (the dark oval pointed to by the arrow is the LBT).
CHAPTER 2. MOTIVATIONAL STUDIES OF SHOULDER SHAPE

Figure 2.11: BG measurements taken from a single cross section in previous literature: medial wall angle, width, and depth.

whereas the previous studies are based on excised humeri.

In this section, we describe our work [227, 225, 226, 228] testing two main hypotheses about the relationship between the shape of the BG and pathological conditions of the LBT. First, we test the hypothesis that global, surface-based shape descriptors of the BG have a discernible relationship to pathological conditions of the LBT. To test this hypothesis, we group the data sets according to expert diagnosis of LBT condition for each data set, and use the 3D shape of their BGs to determine shape differences between the groups. We test the null hypothesis that the means of the groups are the same, rejecting it in favour of the alternative hypothesis that there are statistically meaningful differences in the 3D shapes in the normal group (i.e. in subjects not having LBT pathology) as compared to the pathological groups (i.e. in subjects having a torn, subluxated, or dislocated LBT).

Second, we test the hypothesis that an anatomy-specific, medial-based shape descriptor for the BG captures the essence of the BG measurements classically taken in 2D (figure 2.11), but provides a richer set of information and provides useful qualitative and quantitative information about the condition of the BG. We validate the discriminatory power of our shape descriptor by testing the performance of different classifiers attempting to classify normal vs. pathological BG shapes (defined according to established 2D measurements taken in previous literature) [3, 95, 100, 214].

---

2.2.2 Description of patient data

32 MR shoulder images were obtained, 19 of male patients, 13 of female patients. The images were taken at 1.5T, with imaging parameters as follows: repetition time 4000-5000, echo time 40-50, field of view 14. As these data sets were obtained clinically by a variety of physicians for different purposes, their pulse sequences varied. The 256 × 256 images were axially oriented, with an in-plane resolution of 1.5 mm × 1.5 mm, with a slice thickness of 4 mm. The mean age of the patients was 50 years (standard deviation 18.4 years), with the youngest patient being 24 and the oldest being 87 years of age. We segmented the bicipital groove surface manually under the direction of our collaborator. It was critical to distinguish between the cortical and endosteal surfaces during segmentation, to ensure that the surface that comes in contact with the LBT was the one extracted (figure 2.12). The result of segmentation of one BG is rendered in the context of the MR image from which it was extracted in figure 2.13. Based on the appearance of the LBT on MR, each patient was diagnosed by experts to belong to one of the following six groups: normal, tear, subluxation, dislocation, tear and subluxation, and tear and dislocation. The criteria used in making their diagnoses are described in [34, 216].

2.2.3 Method: Global shape descriptors

In this section, we describe our method for computing and validating global 3D BG shape differences between groups of patients with normal LBTs, as compared with groups having various LBT abnormalities (e.g. tear, subluxation, dislocation, etc.). The process is given in figure 2.14 and is as follows:

1. **Group normal and pathological cases.** As a first step, we obtain pathology groupings from our medical collaborators.

2. **Compute global 3D shape measures.** We compute three shape measures of interest, for each BG. The first is elongation, computed according to the eigenvalues of the covariance matrix of the BG surface points, as in section 2.1.3. Figure 2.15 illustrates the elongation measurement with two example BGs. The second shape measure we compute is surface irregularity or non-sphericity, which is computed as the standard deviation of the Euclidean distances from all of the BG surface points to their centroid, also as in section 2.1.3. The third shape measure is a 3D moment invariant, defined
Figure 2.12: Bone surfaces to be distinguished during segmentation. The inner surface where bone meets bone marrow is the endosteal surface. The outer surface where bone meets surrounding tissue is the cortical surface. We must distinguish between these during surface extraction, because the LBT contacts the cortical surface.

3. Compute significance of 3D shape measure differences. We perform an ANOVA to determine the statistical significance of the differences between the means of the measurements of the normal and pathological groups. This tests the null hypothesis that the population means of the different groups are equal, based on the observations from our 32 data sets.

2.2.4 Method: Medial shape descriptor

Computing a medial shape descriptor for the BG

Here, we describe a medial shape representation of the BG that captures its overall shape (not limited to a single slice), allows for quantitative 3D shape measurements extending the 2D classical measurements, and allows for intuitive visualization and exploration of the BG. A medial-based representation is ideal for this purpose because it generalizes to 3D the classical measurements taken in 2D: width, depth, and wall angles. These are object-centric
CHAPTER 2. MOTIVATIONAL STUDIES OF SHOULDER SHAPE

Figure 2.13: A BG (gold) rendered in context of the 3D MR image from which it was segmented (gray), with a few intersecting axial slices (red) highlighted.

Figure 2.14: Process used to compute differences between groups having normal vs. pathological LBTs. First, we compute the 3D shape descriptor from the segmented BG surface. Next, each data set is assigned a group label by an expert who has diagnosed the LBT based on a reading of the original MRI, resulting in N different groups (e.g. normal, tear, subluxation, dislocation, etc.). 3D shape measures are then computed for each group, and they are compared to determine their differences.
CHAPTER 2. MOTIVATIONAL STUDIES OF SHOULDER SHAPE

Figure 2.15: (a) An example of a BG with a large elongation ($\lambda_1/\lambda_2$) score of 16.1. (b) A BG with a much smaller elongation score of 3.1.

Figure 2.16: The pipeline used to compute the 3D shape representation proposed in this paper, a schematic of which is shown in figure 2.18. Beginning with a segmented BG surface (figure 2.17(a)), we first compute the intertubercular surface, which intersects both tuberosities and closes the BG (figure 2.17(b)). Next, we compute the medial surface to lie perpendicular to the intertubercular surface on each slice, and lie near the deepest region of the BG while remaining smooth (figure 2.17(d)). Last, we compute vector fields which we call thickness fields from the medial surface to form width and medial wall fields (an example of which is given in figure 2.22), and from the intertubercular surface to form a depth field.
measurements that benefit from the computation of a medial surface, and a representation of shape measurements relative to that surface. The process used to compute the 3D shape representation presented in this paper is given in figure 2.16, and is as follows:

1. **Compute intertubercular surface.** Starting with a BG surface (figure 2.17(a)), we automatically locate all of the points at the apexes of the tuberosities, and define an intertubercular surface to close the BG, intersecting with these apexes (figure 2.17(b)). There are, of course, many surfaces meeting these criteria; we choose the smoothest surface that intersects with all of the segments joining the tuberosities on each axial slice of the MRI from which the BG was extracted. The purpose of the intertubercular surface is to allow for the computation of the depth of the BG at any point. The depth of the BG at any point \( P \) on its surface is defined to be the length of the segment normal to the intertubercular surface, having one endpoint on the intertubercular surface and the other at \( P \) (indicated by \( T_D \) on figure 2.18).

2. **Compute medial surface.** Next, we use the intertubercular surface to compute the depth at every point on the BG surface. We use this depth information to position a medial surface, normal to the intertubercular surface, and intersecting a space curve \((x, y, z) = s(t)\). This space curve is constrained to lie on the BG surface, and pass through the deepest point on the BG surface at each axial slice indexed by \( t = 1 \ldots n \). However, due to noise, digitization artifacts, and roughness of the bone, such a space curve is not smooth. Since the medial surface must intersect this curve, this results in a discontinuous medial surface (figure 2.17(c)). We therefore add a smoothness (i.e. minimal total curvature) criterion to the computation of this curve. We define the optimal curve as that which is smooth and close to the deepest points of the BG. We formally define this curve as the optimum of the objective function

\[
\text{argmin}_{s} c(s) = \alpha \sum_{t=1}^{n} \kappa(s(t)) - (1 - \alpha) \sum_{t=1}^{n} D(s(t)), \tag{2.2}
\]

where \( D(s(t)) \) is the depth of the BG at the 3D point given by \( s(t) \), and

\[
\kappa(s(t)) = \frac{|s'(t) \times s''(t)|}{|s'(t)|^3}, \tag{2.3}
\]
Figure 2.17: (a) The segmented surface of a BG. (b) Intertubercular surface fitted to BG surface. (c) Medial surface fitted using BG depth as the only criterion. Note discontinuities in the medial surface. (d) Medial surface fitted using BG depth and medial surface smoothness as criteria, each with 50% weight. The medial surface now lies generally in the deepest part of the BG, while remaining smooth.
that is, $\kappa(s(t))$ is the standard definition of curvature of a space curve $s$ at $t$. The curve $s$ yielding the global optimum of $c$ in equation (2.2) can be found by considering a multiple stage path finding problem, where each stage involves finding one point along the path in a single axial slice (i.e. we proceed through the BG one axial slice at a time, choosing the optimal path on that slice). Having reformulated this problem as a multi-stage decision process, we can use dynamic programming to find the optimal path [104]. We obtained smooth paths approximately bisecting the BG for all data sets by setting $\alpha = 0.5$.

3. **Compute thickness fields.** Finally, we compute vector fields capturing information about the depth, width, and medial wall of the BG. The vectors in the field capturing the depth information are normal to the intertubercular surface, oriented in the direction of the BG surface, and terminate at the BG surface. The vectors in the fields capturing information about BG width and the medial BG wall are normal to the medial surface, oriented in the direction of the BG surface, and terminate at the BG surface. See figure 2.18 for a schematic diagram. In all of these vector fields, the distance between the starting points of neighbouring vectors is 0.1 mm. This resolution was chosen to be higher than that of the original 3D medical images, guaranteeing
that the thickness fields capture the smallest features perceptible in the those images.

To summarize, we begin with a segmented BG surface, consisting of a set of 3D points lying on the surface. From this, we compute a set of three thickness fields capturing information about the depth, width, and medial wall of the BG. These fields are computed to be the 3D analogues of the 2D depth, width, and medial wall angle BG measurements taken in previous studies (figure 2.11).

Validating the medial shape descriptor for the BG

Previous studies typically take a single set of 2D measurements of width, depth, and medial wall angle for each BG, either from a dry bone or a radiograph. Additionally, previous researchers note the presence/absence of the supratubercular ridge of Meyer (henceforth referred to as “the ridge”) on each BG, which is considered to have an impact on the tendency for dislocation of the LBT. These measurements of width, depth, medial wall angle, and presence/absence of the ridge, shall henceforth be referred to as the “classical measurements” for brevity. Next, we test whether our shape representation captures, at minimum, the classical measurements, by validating against distributions of 2D measurements taken of large numbers of bones in previous studies [3, 100, 214]. This validation confirms the correctness of the calculation of our shape descriptor, and its alignment with previously computed measures. This is important in order to ensure that we are conveying, in 3D, information which is intuitively similar to that which physicians and researchers are accustomed to using in 2D on individual cross sections of the BG. We are particularly interested in the ridge, since its identification poses difficulty even for the expert observer. To test whether the proposed 3D shape descriptor captures the classical measurements, we follow the process given in figure 2.19:

1. **Compute classical measurements.** Given a set of points sampled from a segmented BG surface, we automatically compute the width, depth, and medial wall angle (figure 2.11) of each data set. To take the measurements, the BG surfaces are first rotated so that slicing along the axial direction yields cross sections orthogonal to the humerus. An intertubercular surface is fitted to the tuberosities (figure 2.17). At each axial slice \( S \), the intertubercular surface appears as a line segment \( PQ \) with endpoints \( P \) and \( Q \) touching the tuberosities. The depth of the \( BG \) on \( S \) is determined as shown in figure 2.20. The width of the BG at \( S \) is the length of \( PQ \). Starting
Figure 2.19: Process used to validate the proposed 3D shape descriptor. Points sampled from the surface of the BG are used to compute the descriptor (figure 2.18), and also used to compute “classical measurements” of BG shape used in previous work (figure 2.11). Based on distributions of classical measurements determined in previous studies, a diagnosis of normal versus pathological shape is made for each data set based on its classical measurements. These diagnoses form labels for a set of training shapes given to a classifier, which attempts to correctly label BG shapes in the test set. We evaluate the accuracy of the classifier based on a comparison of these labels to known labels for the test set.
from the deepest point on the BG surface on $S$, points sampled along the medial wall define vertices of a polyline approximating the wall. Angles of these line segments with respect to the segment $PQ$ are recorded for $S$. So, for each BG, we have a set of depth and width values and a set of medial wall angles for every axial slice. Since measurements taken in previous work are of a single slice, and previous authors are not specific in describing how the slice is chosen [3, 100, 214], we unfortunately are forced to make an arbitrary choice as to how to aggregate our measurements. In this work, we choose to aggregate all of our measurements by taking their mean in each data set to yield a single depth, width, and medial wall angle for each data set.

2. **Label normal and pathological cases.** Three previous studies have performed measurements of depth, width, and medial wall angle on dry bones and radiographs [3, 100, 214]. Counting all of these studies, each measurement was taken on 130 bones. The results are as follows. Medial wall angle mean: 60.02 degrees, standard deviation (SD): 15.32 degrees. Depth mean: 4.19 mm, SD: 0.96 mm. Width mean: 7.9 mm, SD: 1.42 mm. Although it is understood that there is a relationship between BG shape and LBT pathology, the exact nature of this relationship is unknown. It is therefore not possible for an expert to label each training set as normal or pathological according to the shape of its BG; a “pathological” BG does not produce patient symptoms in and of itself, but rather through its effect on the LBT. For example, although it is known that a shallow, wide BG can promote LBT dislocation, it is not known precisely how shallow and how wide a BG has to be in order to guarantee LBT dislocation with high confidence. Consequently, we turn to previous studies taking measurements of the BG on dry bones and radiographs to obtain a mean and standard deviation (SD) of these measurements for a sample of the population, and specify a SD cutoff to divide the sample into approximately equal-sized “normal” (closer to the mean obtained from previous studies) and “abnormal” (further from the mean) groups. We choose a standard deviation threshold of 1.5 SD for our data, for all measurement types (i.e. depth, width, medial wall angle). Thus, for example, if a data set’s classical depth measurement is further than 1.5 SD from the mean of the distribution of depth values reported in previous studies, it is deemed to be abnormal; otherwise it is normal. This threshold, applied to all measurement types, results in the following labeling: medial wall angle: 15 normal data sets, 17 abnormal data sets; width: 14 normal, 18 abnormal;
Figure 2.20: (a) Schematic of a humeral head with the bicipital groove shown. (b) An expanded view of the bicipital groove. Because the BG curves with the contour of the humeral head, it is important in measuring depth to ensure that depth is measured along vectors orthogonal to the intertubercular surface, lying in search planes exemplified in (a). (b) is an enlarged view of a single search plane from (a) showing the details of the definition of the search plane at a given slice of the BG. Segment $PQ$ is defined to join the endpoints of the intertubercular surface, and vector $v_1$ is defined from $P$ to $Q$. Vector $v_2$ is defined to originate from the midpoint $M$ of $PQ$ and be normal to the intertubercular surface at $M$. The search plane is defined to contain $v_1$ and $v_2$. Segments are computed to originate from points sampled uniformly along and orthogonal to $PQ$, constrained to lie in the search plane, and terminating at the BG surface. The length of the longest such segment (indicated by $D$) is determined to be the depth of the BG on this slice.
depth: 18 normal, 14 abnormal. The significance of this choice of threshold is the
challenge that it presents to the classifier, since there are roughly as many normals as
abnormals (47%, 44%, and 56% of the sets are normals for medial wall angle, width,
and depth, respectively). For instance, a different choice of threshold could yield 30
normals and 2 abnormals. A classifier that blindly classifies all test sets as normal
would give the correct answer $30 / 32 = 93.8\%$ of the time. By choosing a suitable
threshold we ensure that a classifier cannot achieve high accuracy by consistently
outputting a single response. Each data set was also labeled according to observation,
by a board-certified radiologist, of the supratubercular ridge of Meyer [137].

3. **Computation of 3D shape descriptor.** This is computed as described in the
previous section, according to the process given in figure 2.16.

4. **Anatomical correspondence.** To prepare the thickness field data for classifiers, we
establish anatomically meaningful correspondence between elements. Each thickness
field has a 2D parameterization, where each point $(i, j)$ maps to a single scalar which
is the magnitude of one vector in the field. We must establish a correspondence
between thickness fields, across patients, such that point $(i, j)$ in each thickness field
of a given type (i.e. depth, width, medial wall) should correspond anatomically with
points $(i, j)$ in the thickness fields of the same type in all other data sets. Establishing
correspondence is challenging for the BG as it does not have well-defined anatomical
landmarks. Due to large slice thickness (4mm), the proximal end of the BG is not
reliably determined, and a method for reliably determining the distal end is the subject
of debate [214]. Our approach avoids having to define the beginning and the end of the
BG: since the BG is formed by parts of the humerus (tuberosities), we align the humeri,
carrying alignment of the BGs along consequently. Our shape descriptor is invariant to
rigid transformations except translation along the axial dimension (recall that all BG
surfaces are rotated so that the humerus is axially aligned), and it is not invariant to
changes in scale. To establish correspondence, we compute the center and radius of the
best-fit sphere to points sampled on the humeral head using the Hough transform [56].
The alignment process that follows is given in figure 2.21. After aligning the humerus
bones along a common axis (figure 2.21(b)), we uniformly scale thickness fields to
normalize for humeral head size (figure 2.21(c)), using the best-fit sphere radii as
the indicator of humeral head size. We align all thickness fields such that the axial
CHAPTER 2. MOTIVATIONAL STUDIES OF SHOULDER SHAPE

Figure 2.21: An illustration of the anatomical alignment process. (a) Unaligned humeri. (b) Three humeri of different sizes and at different positions along the axial dimension (to which the humerus is aligned). Wire mesh spheres fit to the humeral heads illustrate the size differences between the heads. (c) The same humeri after normalization of humeral head size. Note that the axial coordinates $z_1$, $z_2$, $z_3$ of the sphere centers are different. (d) The same humeri after alignment along the axial dimension. The axial coordinate of the sphere center is now the same for each humerus.
coordinate of these spheres is the same, thus aligning the bones (figure 2.21(d)). Our rationale for this approach is based on the need of an approach to alignment that is based on humeral anatomical landmarks, accounts for different-sized bones, and aligns the grooves axially without relying on poorly-defined groove boundaries along the axial dimension. To the best of our knowledge, an approach to alignment meeting these criteria does not exist in the literature, except in our publication [226]. For example, Procrustes Alignment [75], a widely used approach, mutually aligns shapes in a group such that the sum of the boundary differences between each each shape and the mean shape of the group is minimized. For our application, this approach to alignment would result in a scaling of the thickness fields so that they are normalized to each other without respect to the size of the humeral head. It would similarly adversely affect the translation of each thickness field along the axial dimension. After alignment, we have a set of thickness fields (e.g. figure 2.22(a)), normalized for scale and aligned along the bone (axial) dimension. The final step is a rescaling of the field along the depth dimension to make the fields rectangular by resampling thickness values at locations from 0% to 100% along the depth of the BG (figure 2.22(b)) in the case of the width and medial wall fields, and along the width of the BG in the case of the depth field. This step establishes correspondences between thickness field vectors which sit at the same fractional position with respect to the depth (in the width and medial wall fields) and width (in the depth field) of the BG.

5. **Classification.** Each thickness field can be viewed as a single, high-dimensional observation vector representing the entire field. Each element of such a vector contains the magnitude of a single thickness vector in the field. We perform principal components analysis [106] (described in appendix A) on these observation vectors to avoid performing supervised classification against such large observations. We then trained several classifiers against the principal components of the depth, width, and medial wall field data capturing 95% of the variation. We also trained classifiers to recognize the presence/absence of the supratubercular ridge of Meyer [137] based on the depth fields. Due to our small sample size, testing of the classifiers was performed in a leave-one-out cross-validation, with classification errors averaged over all rounds.
2.2.5 Results

Global shape descriptors

Table 2.6 gives the mean and standard deviation, for each pathology group, of the three measurements described in section 2.2.3. The first row of table 2.7 gives the ratio of the measurements for the normal groups against the mean of the measurements for all of the abnormal groups (indicated in rows 2-6 of table 2.6). The purpose of the first row of table 2.7 is to yield a picture of the ways in which the groups are different. To test the null hypothesis that there are no statistically significant differences between the groups, we performed a one-way ANOVA test for every measurement type on every thickness field type. Resulting $p$ values are given in the second row of table 2.7. We also stratified the measurements according to age in table 2.8.

Medial shape descriptor

Table 2.9 shows the results of our classification experiment described in section 2.2.4. The methods shown gave the best performance against our data. Each row corresponds to a single type of measurement (width, depth, medial wall angle) or observation (presence or absence of supratubercular ridge). The second column shows the classifier yielding the best
Table 2.6: Results of 3D shape measures taken of the bicipital groove surfaces, for 32 patients. Results reported are the mean for each group, with the standard deviation following in parentheses. For instance, the value in row Tear, column Elongation shows the average of the ratios of the 2nd eigenvalue to the 3rd eigenvalue for all BGs whose long biceps tendon was diagnosed as torn; the closer the depth and width of the BG are to being the same, the closer to 1 is this measure.

Table 2.7: For each measurement, all of the abnormal groups are aggregated by taking the mean. Normals are then compared with abnormals by taking the ratio given in the first row of the table. The second row shows p values resulting from a one-way ANOVA test for each measurement, to test the null hypothesis that there are no significant differences between the means of the normal and pathological groups.

Table 2.8: The mean value of each measurement within five different age groups. We observe a difference in the surface roughness and 3D moment measurements for the groups with patients older than 30, as compared with the 30 and under group. The elongation measurement shows a difference in the 61 and over group compared to the others.
CHAPTER 2. MOTIVATIONAL STUDIES OF SHOULDER SHAPE

Table 2.9: Results of validation of the proposed 3D shape descriptor.

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Error</th>
<th>No. Principal components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Width</td>
<td>Quadratic Bayes</td>
<td>0.1875</td>
</tr>
<tr>
<td>Depth</td>
<td>Min. LS Linear</td>
<td>0.1875</td>
</tr>
<tr>
<td>Medial wall angle</td>
<td>Quadratic Bayes</td>
<td>0.3750</td>
</tr>
<tr>
<td>Supratubercular ridge</td>
<td>Min. LS Linear</td>
<td>0.1250</td>
</tr>
</tbody>
</table>

performance in performing a correct diagnosis of the normality/pathology of the given measurement. The third column shows the classification error rate, and the fourth column shows the number of principal components used (after principal components analysis is performed on the BG surfaces) to train the classifiers. This error is computed as the ratio of the number of incorrect classifications made by the SVM, divided by the total number of classifications in the cross-validation experiment. Accuracy was over 80% for most classification tasks, although different types of classifiers performed best for different tasks. We experimented with all classifiers in the PRTools v.4 package\(^3\), and we obtained best results with a minimum least square linear classifier and a quadratic Bayes normal classifier. The former bases its decision on a linear function of the inputs, whereas the latter finds a quadric hypersurface that best separates the classes. For further details, we refer the reader to Fukunaga’s book on pattern recognition [69].

2.2.6 Discussion

Global shape descriptors

The results in the first row of table 2.7 indicate that differences between normals and abnormals, as evidenced by the 3D shape measures taken, range between 30% (for elongation and surface roughness) and 250% (for 3D moment invariant). Some of the \(p\) values in the second row of table 2.7 are also encouraging. For the measurement of elongation, we can reject the null hypothesis that the means of the groups are the same with 98% confidence. This agrees with what is known about the relationship of BG dimensions to LBT pathology: a deep, narrow groove and a shallow, wide groove both have a tendency to cause LBT pathology, and both of these conditions are reflected in the elongation measure as it was

taken in this work. The confidence of null hypothesis rejection for the surface roughness measure is slightly weaker at 94%, but still encouraging. The 3D moment invariant measure clearly requires further investigation into its usefulness as we can reject the null hypothesis with only 73% confidence with this measure. However, the difference between the normals and the mean of the abnormals for the 3D moment invariant measure given in the first row of table 2.7 is encouraging enough that we should continue to investigate the properties and value of this measurement of the 3D shape of the BG. In table 2.8, the surface roughness and 3D moment measurement show a discernible difference between the 30 and under and the 31 and over groups.

**Medial shape descriptor**

No classifier performed adequately in diagnosis of medial wall angle abnormalities; we attribute this to the imprecise definition given in previous studies [3, 100, 214]. Usually, this is given as a diagram (such as figure 2.11), which does not cover frequent cases where the medial wall may have one or more plateaus rather than having a uniform slope (figure 2.23). Also, previous authors are vague regarding the locations of BG measurements; such inconsistencies between studies may have an adverse effect on classification. Considering these obstacles, the results are encouraging in that they suggest that the majority of important BG shape features are captured by our representation. Especially encouraging is the classification performance for the supratubercular ridge of Meyer, which proved to be difficult for the human expert to identify.
CHAPTER 2. MOTIVATIONAL STUDIES OF SHOULDER SHAPE

Figure 2.24: (a) BG visualized with cutaway view of intertubercular surface and medial surface. Small spheres indicate 3D positions on the intertubercular surface and the medial surface. (b) The small sphere indicates the corresponding 2D location on the visualized depth field. (c) The small sphere indicates the corresponding 2D location on the visualized width field.

The proposed medial shape description permits visualizations of the BG measurements which are useful to the physician. It is of value to establish visual correspondence between points on the visualized 3D BG surface, and points on the thickness fields, as shown in figure 2.24, to give the physician an intuitive sense of how to visualize and navigate the thickness and depth fields. The physician can then observe, at a glance, qualitative features of depth, width, and medial wall in isolation from one another; something that is difficult to do when simply visualizing the entire BG surface at once. The physician can, for example, distinguish a rapid widening on the medial wall side of the BG from an overall widening of the BG in a straightforward fashion by observing a rapid change in brightness on the thickness fields. This distinction would be of clinical interest because it is known that the slope of the medial wall has a strong impact on the tendency for LBT subluxation and dislocation [13]. Should the physician wish to explore the fields qualitatively, this shape representation permits such exploration using tools such as Matlab’s imtool (Natick, MA: The Mathworks, 2006), as shown in figure 2.25. The physician can browse to an area of interest in the field, zoom it in tightly, and see corresponding thickness values in physical units.
2.2.7 Contributions, limitations, and future work

The main contributions of the work described in this section and published in [225, 226, 228] are:

- **A study of the relationship between global shape descriptors of the BG and pathology of the LBT:** We demonstrated separability of the pathology groups using global measures of shape (tables 2.6, 2.7, and 2.8). This is useful because it demonstrates that 3D shape of the BG, measured in ways very different from the classical measurements 2.11, does in fact have a relationship to LBT disorders. This study is different from that of the supraspinatus described in section 2.1; the previous study examined the relationship between the shape of a structure and pathological conditions of *that same structure*. In contrast, the present study examines the relationship between the shape of a structure (BG) and pathological conditions of a functionally related structure (LBT).
• A novel, anatomy-specific medial shape descriptor for the BG: We describe the computation of a medial-based representation of the BG that intuitively captures the depth, width, and medial wall angle of the BG; classical measurements which previously were taken in 2D (figure 2.11). Our generalization of these measurements to 3D eliminates the need for the difficult selection of a single slice, or small group of slices, to be used in shape analysis. We also developed an anatomy-specific means of establishing correspondence between coordinates in the proposed medial representation. We validated our 3D shape representation and correspondence by showing that the representation captures the 2D classical measurements, and enables the detection of the supratubercular ridge of Meyer by an automatic classifier (table 2.9). We also demonstrated that our representation enables the physician to conduct qualitative, quantitative, and independent explorations of the depth, width, and medial wall angle of the BG.

The main limitations of the work described in this section are:

• Manual segmentation: The shape analysis methods described depend on the provision of segmentations of the BG from medical image data. Since this segmentation problem is out of the scope of this thesis, we performed segmentations manually. Any clinical application of this work would require, at minimum, an accurate, semi-automated approach to the segmentation of the BG in order to eliminate this time-consuming manual step.

• Anatomy-specific shape representation and correspondence: The methods for computing the medial representation of the BG, and for establishing correspondence across patients, are highly specific to this anatomy. They would apply generally to any trough- or groove-like structures in the body where measures of depth and width are sought, but this is a quite limited range of applications.

Future work based on the work described in this section includes:

• Handedness identification in forensic science: It is understood that handedness may play a role in determining the shape of the BG [182]. The hypothesis is that the dominant arm is used differently and more heavily than the less dominant arm, and this difference in use causes a difference in the interaction between the LBT and the
BG in the dominant arm. This different interaction causes shape changes in the BG which may be detectable. The exact nature of the relationship between the shape of the BG in one arm, and whether or not that arm is the subject’s dominant arm, would be useful to forensic scientists attempting to gain information about people based on their remains. A large-scale study investigating the relationship between the shape descriptors computed in this section and handedness would be useful in answering this question.
2.3 Looking forward

The two studies presented in this chapter represent contributions to medical research, through the development and application of existing and novel shape analysis techniques to specific anatomical structures. However, the shortcomings of these studies motivate further work in the following areas, which are the subjects covered in the remainder of this dissertation:

- **The shape correspondence problem**: With the exception of the shape descriptor computed in section 2.2.4, the descriptors computed in this chapter were global. Although this approach is simple and has produced reasonably good results in both studies, it discards the rich set of shape information provided at every point on the surface of each shape. Conducting a study using local shape descriptors produces insights as to specific regional variations in anatomical structures that are associated with pathological conditions. However, in order to use local shape descriptors, the shape correspondence problem must be solved. Although we approach this problem in section 2.2.4, our approach is highly anatomy specific. In chapter 3, we describe our more generic approach to this problem.

- **The computation of medial shape representations**: The use of a medial shape representation in section 2.2.4 enabled the generalization to 3D of the set of measurements typically taken of the BG in 2D. This is useful because it maintains the communication of shape measurements to physicians and medical researchers using terms which are intuitive to them in this problem domain, such as width, depth, and wall angle. However, the method for computing the medial shape representation described in section 2.2.4 is very anatomy-specific. In chapter 4 we describe our approach to computing and representing medial shapes in the general case, and in chapter 5 we describe our approach to pruning unwanted axes resulting from boundary noise-sensitive skeletonization operations.
Chapter 3

Using expert knowledge for improved shape correspondence

3.1 Motivation and problem

The use of local shape descriptors permits a rich description of shape, and the ability to localize differences between shapes within a sample, or across different classes of shapes. In medical applications, this gives the ability to communicate to physicians the regions of anatomical structures whose shape variations are related to pathological conditions of those or related structures. Using local shape descriptors, however, requires a solution to the shape correspondence problem, which involves the computation of a mapping between each point on each shape and a point on every other shape. The exact meaning of “point” in this context depends on the chosen shape representation. In boundary-based shape representations, points are taken to lie on the surface of the shape. In medial-based shape representations, points lie on the medial manifolds. In medical applications, correspondence must be established in a way that is anatomically meaningful; that is, they must respect and follow correspondences which would be given by a trained anatomist. This is a very challenging problem because the set of features that best characterizes important and salient points varies across different classes of object, and also varies locally along individual objects.

An information-theoretic approach to this problem has recently gained substantial attention in the research community [48, 49, 50, 51, 52, 163, 197, 207, 208]. Given a group of object surfaces to be brought into correspondence, this approach consists of searching the
space of all possible correspondences between the objects’ surface points, and choosing the
correspondence yielding the most compact statistical shape model (for details on computing
a statistical shape model, please see appendix A). The compactness of the shape model is
evaluated by computing its information description length, which is the minimum number
of bits required to transmit the model from a sender to a receiver. The intuition behind
this method is that dissimilarity between corresponding object surface points arises from
two sources: (1) actual shape differences between the objects, and (2) poorly-established
shape correspondences, resulting in incorrect parts of objects being compared. This method
measures shape variability by measuring the dispersion of points in shape space using the
description length of the shape model. By seeking the correspondence yielding the mini-
imum description length, the method attempts to remove as much variability arising from
incorrect correspondence as possible, so that the only remaining variability is that which
arises from the true differences in shape between the objects. This is based on the “Occam’s
razor” principle, the “law of parsimony”: all other things being equal, the simplest solution
is the best.

Approaches based on optimizing shape model description length are simple and have
produced very promising results [48, 49, 50, 51, 52, 163, 207, 208]. A study measuring
the performance of this method relative to ground truth established by experts on medici-
al images revealed that the minimum description length (MDL) based approach exceeded
other contemporary approaches in its agreement with landmarks placed by a human ex-
pert [197]. However, this same study revealed that for correspondences established between
human brain ventricles, the mean disagreement between landmarks placed by MDL and
expert landmarks was $\approx 4$ mm. To put this error in perspective, consider that a study
of 100 normal adults found a median ventricle width of 18.2 mm for men and 16.5 mm
for women [83]; a 4 mm correspondence error could certainly have a significant impact on
the results of a shape study of structures of this size. Inter-expert variability in establish-
ing correspondence was 1.9 mm for the brain ventricles, indicating that there is room for
improvement relative to human performance, and further motivating the need for a more
consistent method for correspondence establishment. The left column of figure 3.1 illus-
trates the problem with DL-based correspondence, showing the DL-based correspondences
of a set of rectangles with two peaks; one large, and one small. The results shown are given
by given by Thodberg’s implementation of DL-based shape correspondence [207]. Although
the gross established correspondence appears qualitatively good according to the coloured
Figure 3.1: Comparison of correspondences established on a group of 5 shapes (one per row) by [207] (left column) with those established by our method (right column) on an artificial example of a rectangle with two peaks. Corresponding points have the same colour. The points of interest are indicated by diamonds in different shades of gray, with numbers indicating correspondence. The bottom-most pair in the final row is zoomed for clarity.
visualization, highlighting specific point correspondences reveals errors at the peak apexes. In particular, although the large peak is generally brought into correspondence correctly (albeit with some errors), the small peak suffers from severe miscorrespondences across shapes. This is due to the fact that the spatial coordinates of the small peak are not very different from those of the nearby points on the side of the rectangle. Thus, a miscorrespondence that corresponds the small peak with non-peak points on the rectangle does not introduce much variability into the statistical shape model, and thus the impact of this miscorrespondence on the description length of the model is small. A miscorrespondence of the large peak, on the other hand, introduces much greater variability into the shape model. A correspondence minimizing the description length of the shape model thus favours proper correspondences of the large peak, at the expense of the small peak, as observed in the left column of figure 3.1. This is problematic since, using curvature as a measure of saliency (for example), both peaks are equally salient.

The fundamental issue here is that the human notion of point saliency is not necessarily encoded by the spatial coordinates of the points. There exists research into building a shape model using other geometric measures of the shape surface, and finding the correspondence minimizing the description length of such a model [96, 207]. These approaches are steps in the right direction and speak to the need for the integration of features other than spatial coordinates. However, they suffer the shortcomings of the arbitrary choice of geometric features and their uniform use throughout the object surface, failing to consider that different surface points may be characterized by different features.

To overcome these problems, we propose the use of expert knowledge, in the form of a set of training examples, to guide the selection of shape and appearance features\(^1\). These features are used in the refinement of correspondences given by a description length optimizer, to establish correspondences that agree with the human intuitions of point saliency given in the training set. Since, in the medical domain, we expect that human anatomists establish meaningful correspondences based both on local shape and appearance information (from the underlying medical images), we acquire shape and appearance information about the points chosen by the user during the training phase. The training information consists of a

set of corresponding points, or landmarks, on the boundaries of the shapes in the training set. The expert selects and corresponds the particular landmarks that are deemed critical to bring into precise correspondence for the shape study. We then automatically learn, for each landmark, the specific shape and appearance features that best distinguish the landmark. To the best of our knowledge, ours is the first approach that uses spatially adaptive feature selection for establishing correspondence. This is important because it stands to reason that due to local variability in shape and appearance at different points on a shape’s surface, it is not reasonable for a single set of features to properly characterize every landmark on a shape. We use the learned features to define cost functions whose optimization refines the correspondence, guiding corresponding landmarks to their correct locations on each shape in accordance with the information obtained during training. Although there exist previous approaches incorporating machine learning [154], to the best of our knowledge, ours is the first incorporation of machine learning to elicit local, characteristic features of anatomically meaningful points, and to make use of those features in a MDL-based approach to the correspondence problem [220]. Our rationale for developing a semi-, rather than fully-automatic method for correspondence establishment is that there exists an abundance of expert knowledge about anatomically meaningful landmarks, and our philosophy is that this information should not be ignored, particularly in the medical domain. When conducting medical research studies of shape versus pathology, such as those described in chapter 2, the medical research team is always heavily involved, so it is not onerous to collect a training sample to harvest their expertise. What we propose here is a means of eliciting the salient shape and appearance features implied by the provided training data, so that the medical expert does not need to explicitly express this information, and encode it in such a way as to be useful to a computer algorithm. This is useful in part because it reduces the burden on the medical researcher, but also because the human expert is not always able to introspect sufficiently as to be explicit about the visual features that cause him/her to select and correspond landmarks; much of this expertise is internalized, as with any human activity in which one develops proficiency.
3.2 Method

3.2.1 Overview

At a high level, our overall process consists of two steps, given by the blocks in figure 3.2: (1) acquiring knowledge about distinguishing shape and appearance features for each expert-chosen landmark ("Training, feature selection"), and (2) applying that knowledge to drive the point correspondence process toward a better solution ("Correspondence via MDL + landmark features").

The goal of the training step is to find, for each landmark chosen by the user, a set of shape and appearance features that, when utilized to refine a point correspondence, (1) drives landmarks toward anatomically meaningful targets as indicated by the user, and (2) reduces as much as possible the chances of a landmark being driven toward a wrong target having similar shape and appearance features to those indicated by the user. The high-level process used in training is to select a subset of shape and appearance features that best satisfies the above criteria. To this end, we have designed a cost function that evaluates each subset. The cost function is a combination of two types of errors: a target error $\epsilon_t$ and a basin width error $\epsilon_b$. The target error $\epsilon_t$, for a given subset of features, is inversely related to the accuracy with which a classifier trained on a feature subset selects target contour points in the training set as belonging to the class of expert-labeled landmarks. Ideally, the classifier would select only the expert-selected points as targets, yielding $\epsilon_t = 0$. The basin width error $\epsilon_b$ measures, for a given subset of features, the spacing between detected targets along the contour. Ideally, this spacing should be as large as possible in order to eliminate confusion between targets during correspondence establishment, when landmarks are driven
The goal of the correspondence establishment step is to refine the point correspondence given by an MDL-based process, resulting in a shape model that is as compact as possible while still anatomically meaningful. To this end, we utilize Thodberg’s implementation of DL-based shape correspondence establishment [207] until it converges, which provides a good initialization for a subsequent minimization of a new cost function derived from the classifiers and feature subsets learned during training. This new cost function, explained in detail in section 3.2.3, drives key (according to the expert) landmarks to anatomically meaningful locations, according to information gained in training. In fact, our optimization is completely decoupled from the MDL-based correspondence; it can be used to refine correspondences given by any method. We choose to use MDL-based correspondence as our initialization here due to its demonstrated performance in [197].

3.2.2 Training

Here, we explain the training step in more detail. Given a set of medical images and contour points lying on the boundaries of segmentations of an object of interest on each medical image, we manually collect \( n \) landmarks \( l_{ij}, i \in \{1, \ldots, n\}, j \in \{1, \ldots, m\} \) on each of \( m \) training examples. To this end, we present the user with not only the object contours, but also with the underlying medical images. The rationale for this is that the user may select points based on both shape and appearance features of the object. Figure 3.4 illustrates the collected training samples, showing three landmarked corpora callosa, with two landmarks (on the rostrum and fornix) each.

Next, for each landmark, we train a classifier, using the best possible subset of shape and appearance features to distinguish the landmark as best as possible from all of the other points on the training objects’ contours. The result is \( n \) different classifiers, one for each landmark. For the moment, we treat the feature selection problem [82] of choosing the best possible subset of features as a black box; we will return to it in section 3.2.4.

3.2.3 Correspondence establishment

Given an initial approximate correspondence from an MDL-based approach, the task is to refine the correspondence to make it more accurate with respect to the corresponded landmarks given by the expert in the training set. We do this via a subsequent minimization
On target, but other targets are confusing

Figure 3.3: Top: Target error $\epsilon_t$ is zero, because the landmark pointed to by the arrow is precisely on the rostrum of the corpus callosum. Middle: $\epsilon_t \neq 0$ because the landmark is off target. Bottom: $\epsilon_t = 0$, but the basin width error $\epsilon_b$ is high due to the presence of many confusing targets nearby.
Figure 3.4: Expert-given landmarks on three training samples. On each example, the rostrum of the corpus callosum is landmarked in red, and the fornix is in green.
Figure 3.5: Cost functions obtained from the training process, applied to a single shape. Since each training example had two landmarks, one for the rostrum and one for the fornix, there are two cost functions computed for each test shape. Zeros are indicated with the red and green circles, with cool colours corresponding to smaller values.
CHAPTER 3. EXPERT KNOWLEDGE FOR SHAPE CORRESPONDENCE

of a new cost function derived from the classifiers and feature subsets learned during training. This new cost function drives key (according to the expert) landmarks to anatomically meaningful locations, according to the information obtained in training. Here, we describe our approach to computing this cost function. Given a set of classifiers (one per landmark) and feature subsets (one per classifier), each classifier assigns each contour point on all contours to one of two classes: points similar to the landmark chosen by the user (which we shall denote as class \( U \)), and all of the other contour points (class \( \tilde{U} \)). For illustrative purposes here, we represent curve points parameterized from beginning to end as strings of digits. For example, curve 0000000000000000 has 16 points, each labeled with a zero. We denote points belonging to \( U \) with ones and points belonging to \( \tilde{U} \) with zeros. Consider, for illustrative purposes, a case where a classifier, for one landmark, given a curve and feature subset outputs 0000011100000000 (indicating that three points on the curve are similar to the user’s chosen landmark). We need a function such that, given a position \( x \) on the curve, the cost function returns the distance to the point best matching the features of the user’s chosen landmark. To do this, we first thin the cluster of ones, yielding 0000000100000000, under the assumption that the central element of each cluster is the desired target (in the case of an even number of ones, we choose an arbitrary center from the two ones resulting from thinning). In order to obtain a cost which is a function of distance from the target, we then take the distance transform of this result, yielding 7654321012345678, where every element shows the distance to the target. We denote the result of this distance transform, for landmark \( i \) on curve \( j \) as \( c_{ij}(x) \). Thus \( c_{ij}(x) \) is a function for landmark \( i \) on a shape \( j \) whose minima lie on points that are similar, according to the selected feature subset, to the user’s given landmarks. For example, consider a case where after the convergence of an MDL-based correspondence, a landmark exists at location \( x_0 \), 3 positions away from the correct target on the curve; \( c_{ij}(x_0) = 3 \). An energy minimization procedure using a cost function based on \( c_{ij}(x) \) will thus drive points to the bottoms of the basins (catchment regions surrounding minima) given by the cost functions \( c_{ij}(x) \). Figure 3.5 illustrates the cost functions computed for a single corpus callosum. Specifically, after the MDL-based correspondence converges, we locate on all of the shape contours the points \( x_{i1} \) corresponding to the landmarks on the first training sample \( l_{i1} \), \( \forall i = 1 \ldots n \). We also locate their corresponding points \( x_{ij} \) on all of the other shape contours. We then find the correspondence minimizing the objective function.
Feature selection

Here, we explain our approach to feature selection during the training step. We choose the feature subset minimizing a cost function which is a combination of a target error $\epsilon_t(p, i)$ and a basin width error $\epsilon_b(p)$, for a feature subset $p$ and a given landmark $i$. It is important to note that we choose a possibly different feature subset for every corresponding set of landmarks $l_{ij}, \forall j = 1 \ldots m$. This allows for different landmarks to be characterized by different subsets of features.

For a candidate feature subset $p$, we train a classifier (which uses a quadratic decision hypersurface) using the features $p$ gathered from all of the points along the contours in the training set. Figure 3.6 illustrates the variability of some computed features along the contours of corpora callosa. The points are divided into two classes for training: class $U$, containing the expert’s selected points, and class $\bar{U}$ containing all other points. For example, consider a case where we have 5 training shapes, each with 100 contour points. The expert user selects one landmark on each shape for training, such that these landmarks correspond across shapes. Consider a feature subset $p$ involving two shape features: curvature and its

Figure 3.6: Three sets of features computed for three shapes of corpora callosa. Different features values are illustrated by different colours.

$$\epsilon_e = \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij}(x_{ij}).$$ (3.1)

This minimization is straightforward; at each time step, each $x_{ij}$ is moved one pixel along the contour so as to decrease $c_{ij}(x_{ij})$. 

3.2.4 Feature selection

Here, we explain our approach to feature selection during the training step. We choose the feature subset minimizing a cost function which is a combination of a target error $\epsilon_t(p, i)$ and a basin width error $\epsilon_b(p)$, for a feature subset $p$ and a given landmark $i$. It is important to note that we choose a possibly different feature subset for every corresponding set of landmarks $l_{ij}, \forall j = 1 \ldots m$. This allows for different landmarks to be characterized by different subsets of features.

For a candidate feature subset $p$, we train a classifier (which uses a quadratic decision hypersurface) using the features $p$ gathered from all of the points along the contours in the training set. Figure 3.6 illustrates the variability of some computed features along the contours of corpora callosa. The points are divided into two classes for training: class $U$, containing the expert’s selected points, and class $\bar{U}$ containing all other points. For example, consider a case where we have 5 training shapes, each with 100 contour points. The expert user selects one landmark on each shape for training, such that these landmarks correspond across shapes. Consider a feature subset $p$ involving two shape features: curvature and its
first derivative. We thus have 500 points in 2D feature space, 5 of which (chosen by the user) are in $U$, with the remainder in $\bar{U}$. A classifier is trained on this data, attempting to find a decision boundary that best separates the two classes. All points are then reclassified into classes $U$ and $\bar{U}$ according to this classifier. Of course, in the ideal case, the point classifications do not change, but this is unlikely with an imperfect classifier. It is very important to note here that the choice of classifier type is critical; a very sophisticated classifier may overtrain to the training set, giving good performance for many different feature subsets, with poor generalization ability. Essentially, the goal of this step is to determine how separable the expert’s selected points are from all other points, using feature subset $p$ and the limited training flexibility of a classifier using a quadratic decision hypersurface.

The subset is then assessed according to the output of the classifier. Recall that this assessment is according to two error measures: the target error $\epsilon_t$ and the basin width error $\epsilon_b$. The target error for the $i$th set of corresponding landmarks is defined as

$$\epsilon_{t_i}(p) = \sum_{j=1}^{m} d(l_{ij}, \hat{u}_{ij}),$$

(3.2)

where $d(p, q)$ is the geodesic distance along the object contour between points $p$ and $q$, and $\hat{u}_{ij}$ is the geodesically nearest point in class $U$ on the contour to $l_{ij}$. For example, if the user’s selected landmark $i$ (indicated with a 1) is 0000000100000000 and a classifier trained using subset $p$ yields a classification of 0000000001000000 on shape $j$ (the 1 is 3 positions away from the user’s landmark), then $d(l_{ij}, \hat{u}_{ij}) = 3$. If the classification were identical to the user’s selection, 0000000100000000, then $d(l_{ij}, \hat{u}_{ij}) = 0$. This measure indicates how far a cost function defined by the output of a classifier, trained on a shape subset, will move a landmark off target.

The basin width error for the $i$th set of corresponding landmarks is defined as

$$\epsilon_{b_i}(p) = \sum_{j=1}^{m} \left( \frac{N_p}{2} - d(\hat{u}_{ij}, u_{ij}) \right),$$

(3.3)

where $N_p$ is the number of points on the longest contour in the training set (and thus the maximum basin width error is $\frac{N_p}{2}$) and $u_{ij}$ is the geodesically nearest point in class $U$ on the contour to $\hat{u}_{ij}$. For example, if the user’s selected landmark $i$ is 0000000010000000 and a classifier trained using subset $p$ yields a classification of 0100000100000000 on shape $j$, $d(\hat{u}_{ij}, u_{ij}) = 5$. A classification of 0001000101000000 yields $d(\hat{u}_{ij}, u_{ij}) = 2$. Clearly the
former is to be preferred over the latter; the latter results in a possibly confusing minimum value of $\epsilon_e$ during correspondence establishment. For each subset $p$, for a set of corresponding landmarks $i$, we compute the overall error measure

$$
\epsilon_i(p) = \begin{cases} 
\epsilon_{t_i}(p) + \frac{\epsilon_{b_i}(p)}{N^2} + 1 & \text{if } \exists \epsilon_{t_i}(p) < T \land \exists \epsilon_{b_i}(p) < \frac{N^2}{2} - T \\
\frac{\epsilon_{t_i}(p)}{\alpha(\epsilon_{t_i}(p) - T)^2 + 1} + \epsilon_{b_i}(p) & \text{otherwise}
\end{cases}
$$

The error measure $\epsilon_i$ is a nonlinear combination of $\epsilon_{t_i}$ and $\epsilon_{b_i}$. In (3.4), $T$ is a threshold specifying the relative impacts of $\epsilon_{t_i}(p)$ and $\epsilon_{b_i}(p)$ on $\epsilon_i(p)$. The intention is that if there exists a feature subset $p$ such that $\epsilon_{t_i}(p)$ and $\epsilon_{b_i}(p)$ are sufficiently small, it is desirable for $\epsilon_{t_i}(p)$ to dominate the cost function so that we find a feature subset providing the highest accuracy possible in positioning the landmarks. This is expressed by the function $\epsilon_{t_i}(p) + \frac{\epsilon_{b_i}(p)}{N^2} + 1$, where the first term is in units of pixels and the second term is normalized to be between 0 and 1. Therefore, $\epsilon_{t_i}(p)$ dominates this function. If $\epsilon_{t_i}(p)$ and $\epsilon_{b_i}(p)$ are not sufficiently small for any feature subset $p$, then the cost function becomes a nonlinear combination of $\epsilon_{t_i}(p)$ and $\epsilon_{b_i}(p)$ such that the dominance of $\epsilon_{t_i}(p)$ diminishes as $\epsilon_{t_i}(p)$ grows. This avoids the choice of subsets with relatively large $\epsilon_{t_i}(p)$ and small $\epsilon_{b_i}(p)$ in cases where the smallest possible $\epsilon_{t_i}(p) \geq t$. Intuitively, we are expressing here that if no subset $p$ of features can provide very accurate positioning of landmarks with respect to the training set, then we prefer instead to seek a feature subset yielding the widest catchment regions (basins) possible, so that the optimizer is not confused by nearby targets. The rate at which the dominance of $\epsilon_{t_i}(p)$ diminishes is directly proportional to $\alpha$, which was set to $10^{-2}$ for all experiments in this chapter. The threshold $T$ can be set automatically by iterating the MDL-based correspondence (or any other chosen method) until convergence and then computing the maximum geodesic distance between any landmark and its correct position given by the training data. By brute force optimization, we select the best feature subset $\hat{p}_i$ for the $i$th set of corresponding landmarks as

$$
\hat{p}_i = \arg\min_p \epsilon_i(p).
$$

Shape features used in our experiments include curvature, the absolute value of the first derivative of curvature, and the local area integral invariant [131]. Appearance features include the average image intensity in a circular region centered at each contour point, the average intensity along a segment extending normal to the contour into the object, and the
average intensity along a segment extending normal to the contour out of the object. Each feature was computed at a variety of scales. In our implementation we computed a set of 12 different kinds of feature-scale information, yielding a search space of size $2^{12} = 4096$ for the optimization problem in (3.5) (we discuss the scalability of this approach in section 3.5). Note that our method does not dictate the specific choices of these features; the features used here are chosen to exemplify the effectiveness of the method.

3.3 Results

Figures 3.1, 3.7, 3.8, and 3.9 demonstrate the performance of our method, compared to the performance of [207] alone. In all of these figures, the left column shows the performance of [207], and the right column shows the refined correspondence given by the optimization of (3.1). The gray diamonds indicate points of interest which should be brought into correct correspondence.

Figure 3.1 demonstrates our performance on rectangles (the bottoms of the rectangles are cut off to make the figure fit one page) with two protruding peaks, one small and one large. In this example, training involved a single example with two training points. Figure 3.7 demonstrates our performance on rectangles, each with a point of interest defined by the texture of the surrounding area in the image. Training in this example involved a single example with one training point. Figure 3.8 demonstrates our performance on some sample corpus callosum (CC; a structure joining to two hemispheres of the brain) shapes, segmented from 2D midsagittal MR images. These points should lie on the tip of the rostrum (the apex of the “hook” of the CC on the anterior side; see the dark gray diamond in the first row of the right column), and at the interface between the CC and the fornix, which emanates below the CC in the middle of its body (see light gray diamond in the first row of the right column). Training was performed on 5 examples, with testing on 15. Figure 3.9 compares our method to the performance of [207] on some sample brain ventricle shapes, segmented from axial MR images. The left column shows the performance of [207], and the right column shows the refined correspondence given by our algorithm. Training was performed on 5 examples, with testing on 14.

---

In all results of this paper, we use the implementation of MDL-based correspondence establishment given in [207] which optimizes the correspondence based on the description length of a statistical shape model formed using the spatial coordinates of boundary landmarks.
Table 3.1: Training and correspondence times, and errors, shown as ([207], our method). LP = large peak, SP = small peak, R = rostrum, F = fornix.

Table 3.1 shows training times, correspondence times, and mean geodesic landmark errors (defined as the mean of the geodesic distances, along the object contour, between landmarks and their correct, manually-landmarked locations). Note that the correspondence times shown for our method include the time required to obtain an initialization from [207]. Our method shows a decrease in landmarking error, and training and correspondence execute reasonably quickly (using a 2.4GHz AMD Opteron processor). Note that for the ventricles, the mean error for our method is skewed by outliers; the median error for [207] is 6.0 pixels, compared to a median error of 1.5 pixels for our method.

3.4 Discussion

For the small peaks in figure 3.1, optimization of (3.5) resulted in the selection of curvature and the absolute value of its derivative as features, which is expected given the high curvature defining the apex of the peak. Interestingly, for the large peak, curvature features were not selected; the local area integral invariant was chosen instead, because this feature differentiates the large peak from the small peak, thus yielding a lower basin width error $\epsilon_b(p)$. The small peak apex’s proximity to the bulky rectangle makes its area integral invariant measure substantially higher than that for the apex of the large peak. This is a revealing example of our approach’s ability to elicit distinguishing landmark features without requiring the expert to explicitly communicate them; it is reasonable to expect that a human, if asked, would say that both peak apexes are best characterized by their curvatures, although this is not true for the large peak.

The poor performance of [207] on the example in figure 3.7 is unsurprising since the method does not utilize appearance information in shape matching. Since the landmark of
Figure 3.7: Comparison of correspondences established on a group of 5 shapes (one per row) by [207] (left column) with those established by our method (right column) on an artificial example of a rectangle surrounded by textural appearance information. Corresponding points have the same colour. The points of interest are indicated by the numbered diamonds. The bottom-most pair in the final row is zoomed for clarity.
Figure 3.8: Comparison of correspondences established on a group of shapes (one per row) by [207] (left column) with those established by our method (right column) on anatomically meaningful points of the corpus callosum. Corresponding points have the same colour. The points of interest are indicated by the numbered diamonds. The bottom-most pair in the final row.
Figure 3.9: Comparison of correspondences established on a group of shapes (one per row) by [207] (left column) with those established by our method (right column) on anatomically meaningful points of the brain ventricles. Corresponding points have the same colour. The points of interest are indicated by the numbered diamonds. The bottom-most pair in the final row is zoomed for clarity.
interest is defined only by appearance, with no distinguishing shape characteristics, simply establishes its correspondence by arc length interpolation between the salient corners of the rectangles. Optimization of (3.5) yielded a single feature that best characterized the point of interest, which was the average local image intensity value.

On the CC shapes (figure 3.8), optimization of (3.5) resulted in the selection of curvature and the absolute value of its derivative as features characterizing the rostrum, and the average intensities along segments normal to the contour emanating into and out of the object to characterize the fornix. Note once again that our refinement of the correspondence has a large impact at the fornix, which is mainly characterized by intensity variations in the medical image, rather than shape.

For the brain ventricle shapes (figure 3.9), optimization of (3.5) resulted in the selection of curvature, its derivative, and the local area integral invariant as features characterizing the point of interest. This is not unexpected given that the point of interest lies at a “corner” of the ventricle, and variations in local intensity around the contour are small.

3.5 Contributions, limitations, and future work

The main contributions of the work described in this chapter and published in [220] are:

- **Feature selection for correspondence establishment:** We have developed a means of automatically selecting, from a variety of shape and appearance features computed at different scales, a subset of these features that best characterizes a set of corresponding landmarks given by an expert. We propose a framework allowing different feature subsets to characterize landmarks in different regions of a shape, addressing shortcomings of “one size fits all” approaches that use the same features everywhere. Our approach eliminates any necessity for the expert to explicitly know or declare the distinguishing features for each landmark, automatically eliciting this information from the correspondences in the training set.

- **Demonstration of superior performance, compared with a leading approach:** We demonstrate, both qualitatively and quantitatively, using artificial and real examples, that our method does indeed provide a refinement of established correspondences for important landmarks. Establishing correspondence as that which minimizes the
description length of a shape model built out of spatial coordinates of landmarks ignores salient shape and appearance features not captured through spatial coordinates alone. Our method compensates for this by selecting the most distinguishing features for each landmark.

- **Incorporation of appearance features in the shape correspondence problem:** With anatomical shape analysis as our main target application, we reason that trained anatomists use not only shape but also appearance of an anatomical structure on the underlying medical image (e.g. CT, MRI) in order to establish anatomically meaningful correspondences. We assert that this information should therefore be utilized in establishing anatomically meaningful landmark correspondences, and we enable this by computing local appearance features along the shape boundaries and providing them to the feature selection process. In a sense, this begins to blur the boundary between shape correspondence, where contours/surfaces are corresponded based on their geometric and topological qualities, and image registration, where image pixels are corresponded based on their appearance information.

- **Decoupling from the MDL approach:** Our approach is entirely decoupled from the method used to provide an initializing correspondence. In this work, we use an MDL-based method due its reported good performance, but any other method can be used as an initializer. This is also a detriment, however, as we shall see below.

The main limitations of the work described in this chapter are:

- **Brute force optimization of (3.5):** This approach to optimization means that our approach is of exponential computational complexity in the number of shape and appearance features in the domain of (3.5). In this work, our total number of features is sufficiently small that the optimization of (3.5) by brute force is tractable. Although this brute force optimization serves to demonstrate the efficacy of the method when the global optimum of (3.5) is found, clearly this approach does not scale very well. This is problematic since the discriminative power of the feature subset selected by optimizing (3.5) is limited by the features that are available. If no features are available that do a good job of distinguishing an expert’s landmarks, then a poor correspondence will result. It is therefore important in future explorations of this idea to optimize (3.5) by a more efficient means.
• **Decoupling from the MDL approach:** Although the decoupling of our approach from the means used to provide an initial correspondence can be seen as a benefit (please see above), there is also a drawback to this decoupling. Were our approach more fully integrated with the MDL approach, we could optimize a single cost function that balances the description length of the model against the shape and appearance feature match given by (3.1). This would allow the tuning of a parameter giving more weight to the one or the other; in a sense, the MDL criterion could be seen as a regularizer of (3.1). Such an approach would establish correspondence by optimization of a single cost function, rather than using a two-stage process as we do here, and so would be more elegant.

Future work based on the work described in this chapter includes:

• **An integrated approach:** An integrated approach to correspondence establishment can be pursued in two main directions. The first would be to use a hybrid cost function as described above, with the DL of the shape model being one term, and (3.1) being the other term. This approach is advantageous in that it would still be possible to substitute the term(s) of the cost function of any other optimization-based correspondence establishment method for the DL term, to hybridize any method with ours. The second approach, specifically in the context of MDL-based correspondence, would be to utilize the features given by the optimum of (3.5) to regionally augment the shape model such that their mismatch would increase the description length of the model. MDL-based correspondence establishment would then proceed as usual using this modified shape model. Such an approach would be very elegant, but suffers from a very tight coupling to the MDL-based approach.

• **Large scale testing with more features:** Because the domain of (3.5) consists of feature subsets, the optimization of (3.5) represents a combinatorial optimization problem. Genetic algorithms [44, 65] are well-suited to finding the global optima for such problems, as selection and merging of feature subsets can easily be modeled using genetic mutation and crossover operations. Such an approach would permit the expansion of the available set of shape and appearance features (e.g. saliency features [203]) for large-scale testing of our approach. Large-scale testing also requires ground-truth correspondences, which can be challenging to obtain from experts. We
expect that an emerging technique for simulation of ground-truth data will be helpful to this task [86].

• **Generalization to 3D:** Extending this work to 3D involves generalizing to 3D the geodesic distances calculated in (3.2) and (3.3), as well as the thinning and distance transform operations described in section 3.2.3. The main challenge, however, in moving to 3D is enabling the reliable identification of landmarks by experts. There are significant visualization and user interface challenges involved in obtaining reliable landmarking in 3D, and since our method is dependent upon the training set, these problems would need to be addressed.
Chapter 4

Computing medial shape representations

Medial shape representations are useful in that they provide intuitive, object-centered coordinates showing deformations or differences in terms of bending, thickness, and elongation. However, there are challenging questions to be answered in developing a medial representation of shape, and in computing the components of any medial shape representation. In this chapter, we describe our approaches to both of these problems. In section 4.1, we describe our approach to medial shape representation that improves the localization of shape deformations, leading to a more intuitive representation of shape differences. In section 4.2, we describe our manifold learning-based approach to the computation of the basic building blocks of any medial shape representation, namely, a parameterization of the medial manifolds, and the thickness vectors emanating from those manifolds, implying the object surface.

4.1 The medial patch shape representation: Localizing shape differences

4.1.1 Motivation and problem

When engaged in shape analysis, one is obviously interested in inspecting, quantifying, and comparing information about objects’ shapes. For example, one might wish to describe and quantify the differences in the shapes of two objects. It is common for the basic elements
Figure 4.1: (a) A bent slab with a protrusion, with only the bending component $\delta(i,i)$ highlighted. (b) Same as (a), with thickness components $(T_+ (i,j) + T_- (i,j))$ highlighted. (c),(d) medial patch range components of bending and thickness, corresponding to (a) and (b), rendered as grayscale images.

of shape representations to include coordinates which are relative to some global coordinate system. For example, the landmarks in the widely used point distribution model (PDM) [37] are relative to a global frame of reference. The M-Reps medial shape representation [155] also encodes elements of both pose and shape. It decomposes medial representations of shapes into a tree of medial manifolds, each of which is sampled by medial atoms. Three of the coordinates of each medial atom give the location in $\mathbb{R}^3$ of the medial atom, relative to some global coordinate frame of reference. Thus, prior to performing statistical analysis of shapes represented by a PDM or M-Reps, one needs to perform a normalization of pose. Procrustes analysis [55] is a typical preprocessing step used for this purpose (see appendix A).

However, even after the normalization for pose, the use of global coordinates in a shape
representation produces a diffusion of deformation and difference information throughout the representation. Consider, for example, the shape differences between a planar slab, and a bent slab such as that shown in figure 4.1 (top left). The two bends of the slab have a ripple effect on an M-Reps representation of this manifold, since many of the \( \mathbb{R}^3 \) coordinates of the medial atoms need to change as a consequence of these deformations. This is counterintuitive; these bends occur along specific local crests in the deformed shape, so the change to the shape representation should be local to these regions also. As a preview, our proposed medial shape representation does localize these changes, as shown in figure 4.1 (top left, and bottom left).

Inspired by the use of differential coordinates [123] for encoding polygon meshes in computer graphics, we present a new medial-based 3D shape representation that avoids the use of global coordinates wherever possible\(^1\). It allows for analysis of 3D anatomical shapes, where shape differences are decomposed into independent, intuitive deformation types (bending, stretching, and thickening). We represent the shape of a 3D object as a collection of parts. The different parts are related by an undirected graph encoding the object’s topology. To capture the object’s geometry, each part is described using a medial surface with thickness values encoding the distances from the medial surfaces to the object’s surfaces. Two parameters traverse each medial surface and are mapped to orientation, elongation, and thickness values, which describe the shape of each surface and the object boundary implied by the representation. Given the analogy of this mapping to a patch in differential geometry and the medial-based nature of our shape representation, we refer to our representation as the Medial Patch shape representation.

4.1.2 Method

The medial patch shape representation

We represent the shape of a 3D object using a coarse-to-fine approach through a collection of medial surfaces, joined via an undirected graph \( G(V, E) \) where each vertex \( V \) represents a surface, and each edge \( E \) represents a connection between two surfaces. Each medial surface

CHAPTER 4. COMPUTING MEDIAL SHAPE REPRESENTATIONS

Figure 4.2: Terminology and variables of medial patches. Elliptic dashed region scaled up in figure 4.3.

Figure 4.3: Detailed view of (a) longitudinal angle $\lambda(i,j)$, latitudinal angle $\delta(i,j)$ and elongation $r(i,j)$, (b) the distances above $T_+(i,j)$ / below $T_-(i,j)$ the medial surface.
is parameterized on \((i, j)\) where \(i, j \in \mathbb{N}\), and comprises a base medial curve sampled at base medial nodes and several medial curves sampled at medial nodes and attached to the base medial nodes (figure 4.2). Associated with each of the medial nodes is a description of the relative orientation and spacing between nodes, and thickness values giving two surface nodes, which are point samples lying on the object’s surface. Medial surfaces are joined via overloaded nodes which serve as medial nodes for one surface and as base medial nodes for another (figure 4.2).

A medial patch is a map \(M : \mathbb{N}^2 \to \mathbb{R}^+ \times S^1 \times S^1 \times \mathbb{R}^+ \times \mathbb{R}^+\). We encode each medial surface, and the boundary surface it implies, using a medial patch with medial surface parameters \((i, j)\) forming the patch’s domain. The patch’s range has five scalar components, three of which encode relative positional information about the medial nodes: (1) the elongation (distance), \(r(i, j) \in \mathbb{R}^+\), between neighbouring nodes, (2) the longitudinal and (3) latitudinal angles, \(\lambda(i, j) \in S^1\) and \(\delta(i, j) \in S^1\) respectively, between neighbouring nodes (figure 4.3(a)). Two additional thickness components, \(T_+(i, j) \in \mathbb{R}^+\) and \(T_-(i, j) \in \mathbb{R}^+\), store the distance above and below the medial surface, respectively, to the upper and lower parts of the surface of the object, defining surface nodes (figure 4.3(b)). The thickness values \(T_+(i, j)\) and \(T_-(i, j)\) need not be the same.

**Computing a medial patch of a 3D object**

Although our shape representation does not prescribe a method for its computation for a given shape, we describe here an approach to computing the medial patch representation of a 3D object, where the desired topology of the medial representation is known a priori. Our approach is in many ways similar to that taken in M-Reps [155] and fixed topology skeletons [73]. The rationale behind this type of approach is that when the topology of the anatomical structure is known in advance, spurious skeleton branches can be avoided by fixing the topology of the skeleton. We address the question of how to proceed when the topology of the skeleton is unknown in advance in chapter 5.

Our approach to computing a medial patch representation of a 3D object is described in

---

2Without loss of generality, we make the arbitrary choice that nodes where \(j = 1\) are base medial nodes. An \(O(n)\) operation is required to transform a given shape to place the base medial nodes along \(i = 1\) (and vice-versa).

3Based on the definition of a patch in differential geometry; a map \(M : \mathbb{R}^2 \to \mathbb{R}^n\).
algorithms 4.1 and 4.2, and involves the deformation of an initialized planar surface according to the shape of the object. Note that algorithm 4.1 can be trivially adapted to terminate the loop once the medial representation approximates the object to a desired level of precision, reducing or eliminating the effects of boundary noise on the medial representation. Similarly, multiple connected planar surfaces can be initialized and deformed to fit to complex objects. We encode small, simple protrusions and indentations of objects as changes to thickness values rather than changes to the medial surface itself. We represent our shape with a sufficiently large number of medial surfaces to ensure that these indentations and protrusions can be described by changes in thickness. Specifically, if a protrusion is too complex to be described as an increase in thickness, then this warrants a new additional medial surface to the graph.

Algorithm 4.1 Deforming a flat surface to a medial surface of a binary object

**Input:** Binary volume \( V \); \( V(x, y, z) = 1, \forall (x, y, z) \) inside the object

**Output:** The 3D coordinates \( N(i, j) \) of a medial surface and \( N_{+}(i, j), N_{-}(i, j) \) of upper and lower implied surfaces of the input object

1: \( \lambda_1, \lambda_2, \lambda_3, e_1, e_2, e_3 \leftarrow \) Eigenvectors \( \lambda_i \) and eigenvalues \( e_i \) resulting from principal component analysis of \( \{x, y, z \mid V(x, y, z) = 1\} \)
2: \( N(i, j) \leftarrow \) A uniform rectilinear grid of medial nodes lying on a medial surface with principal directions along \( \lambda_1, \lambda_2 \), and extents proportional to \( e_1, e_2 \)
3: repeat
4: for all nodes \((i, j)\) in the medial surface do
5: \( \overrightarrow{Q} \leftarrow \) Normal vector to the medial surface at \((i, j)\)
6: Cast rays from \( N(i, j) \) along \( \overrightarrow{Q} \) and \( -\overrightarrow{Q} \) until they exit the object. Record the two object exit points as surface nodes \( N_{+}(i, j) \) and \( N_{-}(i, j) \)
7: \( N(i, j) \leftarrow \) Midpoint of segment from \( N_{+}(i, j) \) to \( N_{-}(i, j) \)
8: end for
9: until no change in \( N(i, j), N_{+}(i, j), N_{-}(i, j), \forall (i, j) \)

Given the 3D coordinates, in a global coordinate system \( \Phi_G \), of nodes \( N(i, j) \) sampled from (a) medial surface(s) lying within a 3D object and corresponding surface nodes \( N_{+}(i, j) \) and \( N_{-}(i, j) \) (e.g. from algorithm 4.1), algorithm 4.2 computes the object’s medial representation. At each medial patch domain element \((i, j)\) we compute the patch’s range: \( r(i, j), \lambda(i, j), \delta(i, j), T_{-}(i, j), \) and \( T_{+}(i, j) \).

We use algorithm 4.3 to construct\(^4\) a local Cartesian coordinate system \( \Phi(i, j) \) with origin \( N(i, j) \) and orthonormal basis \( \overrightarrow{x}_1, \overrightarrow{x}_2, \) and \( \overrightarrow{x}_3 \) at each node \((i, j)\). Using coordinate system

\(^{4}\)In algorithm 4.3, we make an arbitrary yet consistent choice of basis, without loss of generality: for nodes where \( i = 1 \) or \( j = 1 \), we are faced with the problem of how to define the local frame of reference at \( N(i, j) \) since \( N(0, j) \) or \( N(i, 0) \) are nonexistent. We replace \( N(0, j) \) with \((1, 0, 0)\) and \( N(i, 0) \) with \((0, 1, 0)\).
Algorithm 4.2 Computing a medial patch from medial and surface nodes

**Input:** The 3D coordinates \( N(i,j) \) of a medial surface and \( N_+(i,j), N_-(i,j) \) of upper and lower implied surfaces of the object; the global coordinate system of those coordinates, \( \Phi_G \)

**Output:** A medial patch: \( \lambda(i,j), \delta(i,j), r(i,j), T_+(i,j), T_-(i,j) \)

1: for all nodes \((i,j)\) in the medial surface do
2: \( \Phi(i,j) \leftarrow \) Local coordinate system computed using algorithm 4.3
3: \((x_1,x_2,x_3) \leftarrow \) 3D coordinates \( N(i,j) \) transformed from global coordinate system \( \Phi_G \) to \( \Phi(i,j) \)
4: \( r(i,j) \leftarrow \sqrt{x_1^2 + x_2^2 + x_3^2} \)
5: \( \lambda(i,j) \leftarrow \tan^{-1}\left(\frac{x_1}{x_2}\right) \)
6: \( \delta(i,j) \leftarrow \cos^{-1}\left(\frac{x_3}{r(i,j)}\right) \)
7: \( T_+(i,j) \leftarrow \|N(i,j) - N_+(i,j)\| \)
8: \( T_-(i,j) \leftarrow \|N(i,j) - N_-(i,j)\| \)
9: end for

\( \Phi(i,j) \), the spherical coordinates \( r(i,j), \lambda(i,j), \) and \( \delta(i,j) \) define \( N(i,j) \), as in figure 4.3(a).

The position of the shape in space relative to a global coordinate system (figure 4.2) is controlled by values \( \lambda(1,1), \delta(1,1) \) and \( r(1,1) \).

**Shape reconstruction from medial patches**

Given a medial patch, we perform two reconstructions, each mapping \( \mathbb{N}^2 \to \mathbb{R}^3 \). The first maps each \((i,j)\) to 3D coordinates of a medial node. The second maps each \((i,j)\) to the 3D coordinates of the object surface above \( (N_+) \) and below \( (N_-) \) the surface. The reconstruction process is given in algorithm 4.4.

**4.1.3 Results**

Figure 4.1 uses deformed synthetic slabs to illustrate the ability of this shape representation to decompose shape variations into intuitive components. Figure 4.4 illustrates the bending of a single caudate nucleus (a brain structure), as well as the ability to highlight intuitive deformations in local regions. In the figure, medial patch range components are visualized using colours on the surface nodes to which they correspond. Figure 4.5 shows the modeling of shape deformations of the bicipital groove (BG) the subject of the study described in section 2.2.
CHAPTER 4. COMPUTING MEDIAL SHAPE REPRESENTATIONS

Algorithm 4.3 Computing a local Cartesian coordinate system at a medial node

Input: Parameters \((i, j)\), medial nodes \(N(m, n)\), \(\forall m = 1..i, n = 1..j\)

Output: The orthonormal Cartesian coordinate system \(\Phi(i, j)\)

1: if \(i = 1\) and \(j = 1\) then \{Base node\}
   2: \(\Phi(i, j) \leftarrow\) Global coordinate system \(\Phi_G\); Stop
3: else if \(i > 1\) and \(j = 1\) then \{Base medial nodes\}
4: \((i, j) \leftarrow (i - 1, j)\)
5: else \{Medial nodes\}
6: \((i, j) \leftarrow (i, j - 1)\)
7: end if
8: if \(i = 1\) then
9: \(N(i - 1, j) \leftarrow (1, 0, 0)\)
10: end if
11: if \(j = 1\) then
12: \(N(i, j - 1) \leftarrow (0, 1, 0)\)
13: end if
14: \(\vec{x}_1 \leftarrow \left( N(i, j) - N(i - 1, j) \right) \left( N(i, j) - N(i, j - 1) \right) \cdot \vec{x}_2 \)
15: \(\vec{x}_2 \leftarrow \left( N(i, j) - N(i - 1, j) \right) \cdot \vec{x}_1 \)
16: \(\vec{x}_3 \leftarrow \vec{x}_1 \times \vec{x}_2\)
   \{\(\vec{x}_3\) is the unit vector normal to \(\vec{x}_1\) and \(\vec{x}_2\) ensuring a right-handed coordinate system\}
20: \(\Phi(i, j) \leftarrow \) Origin \(N(i, j)\), orthonormal basis vectors \(\vec{x}_1, \vec{x}_2, \vec{x}_3\)

Algorithm 4.4 Shape reconstruction from medial patches

Input: A medial patch: \(\lambda(i, j), \delta(i, j), r(i, j), T_-(i, j), T_+(i, j)\)

Output: The 3D coordinates \(N(i, j)\) of the medial surface and \(N_+(i, j), N_-(i, j)\) of upper and lower implied surfaces of the object

1: for all nodes \((i, j)\) in the medial surface do
2: \(\Phi(i, j) \leftarrow\) Local coordinate system computed using algorithm 4.3
3: \(x_1 \leftarrow r(i, j) \sin(\lambda(i, j)) \sin(90^\circ - \delta(i, j))\)
4: \(x_2 \leftarrow -r(i, j) \cos(\lambda(i, j)) \sin(90^\circ - \delta(i, j))\)
5: \(x_3 \leftarrow r(i, j) \cos(90^\circ - \delta(i, j))\)
6: \(N(i, j) \leftarrow 3D\) coordinates \((x_1, x_2, x_3)\) transformed from \(\Phi(i, j)\) to global coordinate system \(\Phi_G\)
7: end for
8: for all nodes \((i, j)\) in the medial surface do
9: \(\vec{Q} \leftarrow\) Normal vector to the medial surface at \((i, j)\)
10: \(N_+(i, j) \leftarrow T_+(i, j) \vec{Q}\)
11: \(N_-(i, j) \leftarrow -T_-(i, j) \vec{Q}\)
12: end for
Figure 4.4: (a) A caudate nucleus, with warmer colours indicating more bending. (b) A second caudate nucleus, with a small (artificial) protrusion. Note that the protrusion is difficult to see in this view. (c) If the user rotates the view to the correct angle, it may be possible to see this protrusion. (d) The protrusion is captured by our shape representation as a thickness difference as compared with (a), and readily highlighted.
Figure 4.5: Deformations applied to the BG, with the medial surface shown. (a) The original segmented BG surface. (b) A spur (bony protrusion) created in the lateral wall. (c) The medial wall’s slope is steepened. (d) The entire width of the BG is reduced. In (c) and (d), the original BG is shown in light gray.
CHAPTER 4. COMPUTING MEDIAL SHAPE REPRESENTATIONS

4.1.4 Discussion

Localization of deformations

The key contribution of this work is described at lines 2 through 6 of algorithm 4.2. The values \( r(i, j), \lambda(i, j), \) and \( \delta(i, j) \) are a set of spherical coordinates in a local coordinate system given by algorithm 4.3. Thus, the nodes on a medial sheet are given an order according to their \( (i, j) \) parameterization, and the position of each node is represented with coordinates that are relative to the immediately previous nodes in the ordering. Consequently, deformations of bending and elongation in objects are localized to the nodes in which they occur. This allows for clear, local descriptions of differences between shapes, computed very simply by differencing the ranges of their medial patches. This is exemplified by the results in figures 4.1 and 4.4. This straightforward encoding of deformations directly extends to several of the shape variations of interest for the BG, namely overall width, width of a single wall, and the presence of osseous spurs (bony protrusions) on the walls. These deformations are exhibited in figure 4.5, which graphically illustrates the effects of this transformations on the BG.

Comparison to M-Reps

The medial patch representation has benefits compared to the M-Reps representation extending beyond the localization of shape deformations and differences through the use of differential spherical coordinates. Multi-figural M-Reps represent complex objects as trees of single-figure M-Reps (where each figure represents one object part), where subfigures are attached to the surface, rather than the medial atoms, of host figures. This presents two problems addressed by our approach. First, the use of a tree implies the existence of a root, and parent-child relationships. There exist many anatomical structures that do not conform well to this model. For example, in the human heart we have a collection of roughly equal-sized parts working cooperatively (ventricle, atria, and valves). Which of these would be considered to be the primary structure? Furthermore, some body systems, such as the circulatory system, being circuitous in nature, cannot be represented as a tree. We therefore choose an undirected graph to represent relationships between medial surfaces; this data structure encodes adjacency without imposing parent-child relationships. Second, M-Reps’ attachment of subfigures to the surfaces of primary figures makes subfigure transformations dependent on the approach taken to surface reconstruction, and fails to adequately model
attachment of components in biological systems (e.g. one’s arms are not stitched to one’s torso at the skin, but rather are more deeply connected in the skeleton). Our approach, using overloaded medial nodes at connection points (figure 4.2) between surfaces, results in a biologically meaningful skeletal structure, with transformations of one figure having natural implicit consequences on connected figures.

4.1.5 Contributions, limitations, and future work

The main contributions of the work described in this section and published in [90] are:

- **Localization of shape deformations/differences:** Our use of spherical coordinates relative to local coordinate systems within the shape results in the representation of each node relative to its neighbours. This provides a useful localization of shape differences. Deformations of shapes have an effect on the shape representation only at the locations where they occur.

- **Straightforward control of global positioning:** Global positioning of a shape is accomplished by adjusting the spherical coordinates of the base node only.

- **Intuitive visualization of deformations/differences:** The localization of shape deformations/differences can be effectively visualized in both 2D (by examining the medial patch range components individually, as in figure 4.1 (bottom)) and in 3D (by colouring 3D surfaces according to medial patch range components, as in figure 4.4).

The main limitations of the work described in this section are:

- **The need to impose an order on the medial nodes:** In order to construct a local coordinate system at each node, there needs to be an ordering of the nodes on each medial sheet so that we have a pair of “previous” nodes to use in forming axes of a local coordinate system. This creates problems at the edges of the sheets, where no previous node exists in one direction, and an arbitrary choice of coordinate system needs to be made. This also creates problems of asymmetry of encoding of certain kinds of deformations. For example, a bend in the medial surface which is exactly orthogonal to the base medial curve (the curve joining nodes where \( j = 1 \)) can be encoded by a modification to one node on that curve. The same is not true for bends orthogonal to the curve joining nodes where \( i = 1 \). We overcome this with an \( O(n) \)
transformation that places the base medial curve through nodes where \( i = 1 \), but this is somewhat inelegant.

- **Crossing normals:** In step 5 of algorithm 4.1, we compute the normal vector to the medial sheet node \((i, j)\), and position the corresponding surface node at the point where this vector passes through the object surface. In shapes containing regions with coincident high curvature and large thickness, such an approach can cause these normal vectors to cross, producing self-intersection of the surface of the object. An approach to this problem is the subject of section 4.2.

Future work based on the work described in this section includes:

- **Algorithms for multi-figural objects:** The fitting approaches given here apply only to single-figural objects. There is substantial further work to be done in developing ways of computing medial patch representations for complexes of objects. Much of the work that has been done at the University of North Carolina on this topic will be helpful to this task [190].

- **Fitting to grayscale images:** Future work includes automatic fitting of medial patch representations to non-binary data (segmentation), to be achieved by extending our automatic surface fitting approach using energy minimization and complex features rather than transitions of binary intensities, as with deformable organisms [88].
4.2 Computing medial shape components using manifold learning

4.2.1 Motivation and problem

As illustrated in section 4.1, medial shape representations such as medial patches and M-Reps [155] typically decompose an anatomical structure into one or more single-figural objects. Each single figural object is represented using a connected set of loci (medial nodes in medial patches nomenclature; atoms in M-Reps) which lie within the object on or near its medial surface forming a medial surface, and a set of thickness vectors (medial patches nomenclature; spokes in M-Reps) which emanate from the medial nodes and terminate on the surface of the object. A medial shape representation for an anatomical structure can be computed directly from a grayscale volumetric medical image (e.g. CT, MRI), effectively segmenting the structure from the image as the shape representation is computed [155]. However, often a segmentation step precedes the computation of the medial shape representation, and the representation is based on a 3D binary image of the anatomical structure (i.e. with 1s representing the object, and 0s elsewhere). Two problems then arise: given a binary image of an anatomical structure of interest, (1) how to position the loci that form the medial surface, and (2) in which direction to emanate each thickness vector from each medial node?

In this section, we explore the utility of manifold learning-based surface parameterization in solutions to both of the aforementioned problems\(^5\). When addressing the problem of how to find the medial loci, skeletonization algorithms such as the MAT [21] and others [7, 116, 151, 232] perform a transformation of the binary representation of the object boundary, yielding the medial loci. Another approach is to begin with a set of loci in the form of nodes of a mesh representing a medial surface, and deform the mesh such that the nodes become medial (or near-medial) loci of the binary object, as in algorithm 4.1 in section 4.1.2, as well as [73, 155]. Both approaches have disadvantages. Skeletonization results in a dense set of 3D points, whereas what is generally required is a simpler mesh sampled from the surface implied by the skeleton. Initializing a mesh and deforming it to the medial region of an object yields the desired data structure, but possibly at the cost of true medialness of the

Figure 4.6: Illustration, in 2D, of problems with computing thickness vectors to be normal to the surface (black) or normal to the medial surface (red). (a) An object with high curvature and high thickness causes thickness vectors (blue) normal to the object surface to fail to intersect the medial surface. Similarly, thickness vectors (green) normal to the medial surface can fail to intersect with the object surface. (b) By parameterizing the surfaces according to their intrinsic coordinate system, this problem is avoided (red vectors shown emanating from the medial surface to the object surface). (c) A bump on an object causes thickness vectors (blue) normal to the object surface to cross. (d) Surface parameterization according to the intrinsic coordinate system solves this problem.

mesh. The approach taken in this work is a compromise between these two approaches. We perform skeletonization of the shape in order to find its precise medial loci, and then apply manifold learning techniques (e.g. ISOMAP [205], Locally Linear Embedding (LLE) [175], Laplacian Eigenmaps [16]) in order to discover the intrinsic 2-dimensional \((i, j)\) coordinate system of the medial surface sampled by the skeleton points. These techniques are related to surface-flattening techniques applied to arbitrary meshes [238] and voxel-based surfaces [81] (the core of the latter cited work being essentially the same as ISOMAP [124]). The loci are then determined by overlaying a regular lattice of nodes on the flattened medial surface.

To address the second problem of specifying the direction of each of the thickness vectors, we propose an approach that splits the original surface into upper and lower sections (relative to the medial surface). We parameterize each section, also using a manifold learning approach, using a regular lattice of the same resolution as that which parameterized the medial surface. Thickness vectors emanating from point \((i, j)\) on the medial surface then
CHAPTER 4. COMPUTING MEDIAL SHAPE REPRESENTATIONS

terminate at points \((i, j)\) on the upper and lower surfaces. That is, vectors emanating from nodes on the medial surface terminate at corresponding nodes on the upper and lower surfaces, where the correspondence is determined in the intrinsic 2D coordinate systems of the surfaces. The motivation behind this approach is illustrated in figure 4.6. When an object has a region of high thickness and also of high curvature, such as may occur in the intestinal tract in virtual colonoscopy, or in the major blood vessels, thickness vectors computed to be normal to the surface (as in M-Reps) may fail to encounter the medial surface, and conversely, vectors normal to the medial surface may fail to encounter the intended region of the object surface. Also, concavities in real anatomical objects can cause normal thickness vectors to cross. Several attempts have been made to solve this problem in previous work. Oda et al. [146] use a spring model to modify directions of planes cutting through the colon in a virtual colonoscopy study, the aim being to minimize the total length of springs. This approach is an iterative technique, and the authors state that “Updating these [plane] parameters is performed by an iterative process that is terminated if the number of iterations reaches a predefined value.” The approach taken in this work, by contrast, is noniterative and arguably simpler. Wang et al. [217] propose a solution that models gastrointestinal tract cross sections as electrical force lines. This approach is computationally complex, to the extent where the authors suggest means of approximating their full algorithm, at the expense of a guarantee that the cross sections do not intersect. Han et al. [92] utilized a radial shape operator to address this problem, but their formulation makes the assumptions that the medial shape representation’s parameters form a Riemannian symmetric space (although this is usually true) and that the thickness vectors lie normal to the shape boundary, which can be problematic, as illustrated in figure 4.6(c).

4.2.2 Method

The overall process followed in this work is described in figure 4.7. Given a binary image of an anatomical structure, its skeleton is computed and pruned by any applicable method [116]. Skeleton pruning is out of the scope of the work in this section, but is the subject of chapter 5. In this work, we use the BinaryThinningImageFilter from the Insight Toolkit (ITK) [233], which is based on a 2D algorithm given in [74]. Skeleton points are then clustered such that each cluster contains a single medial surface. To do this, we eliminate all voxels having more than 8 26-connected neighbours, and find the resulting connected components, treating each component as a separate surface. This is justified because it
Figure 4.7: Flowchart depicting the overall process used in this work. Starting with a binary image containing an anatomical structure, we first skeletonize the object. If the skeleton contains $N$ figures, the skeleton points are clustered into $N$ groups; one group per figure. Points on the surface of the binary object are classified, based on the position of the skeleton, as belonging to either the upper or the lower surface of the object. Next, the surfaces represented by the skeleton, upper, and lower surface points are all parameterized via nonlinear dimensionality reduction (see figure 4.8). This yields, for each figure, meshes for the skeleton, the upper surface, and the lower surface. Correspondence between nodes in these meshes is established in the 2D coordinate systems established during nonlinear dimensionality reduction, and this correspondence is used to compute the thickness vectors.
Figure 4.8: The process used to parameterize a 2D surface embedded in 3D, given a set of disconnected 3D points sampled from the surface. Given a binary image depicting a surface, 3D points are extracted that lie on that surface. Next, these points serve as input to an algorithm for nonlinear dimensionality reduction (ISOMAP [205] is used in this work) in order to map the points into their intrinsic 2D space. Points are then sampled from this 2D space based on a regular lattice, and these point samples are mapped back to 3D using information derived from the mapping discovered by ISOMAP. The mapped 3D points form the vertices of a surface mesh, and the edges of the mesh connect neighbours according to the 2D lattice.

is generally agreed upon that a skeletonization algorithm should yield a single-pixel-wide connected skeleton [116]. However, we acknowledge that this approach is not sufficiently general to separate arbitrary skeletons of 3D objects into their constituent manifolds; work by Saha et al. [177] explores this problem in detail. During the clustering process, we obtain for free the loci of junction points of the surfaces: the loci of the removed voxels. These can be used, for example, to locate overloaded nodes at a connection curve in medial patches [90], or hinge atoms in multi-figural M-Reps [91]. Thus we have a set of clusters of medial voxels, where the voxels in each cluster are taken to be samples from a single (likely non-planar) surface in 3D, corresponding to a single-figural part of the anatomical structure. Next, we assign a label to each object surface voxel as belonging to the upper or lower surface of the object, depending on which side of the object’s skeleton the point lies. Thus, for each figure, we have three sets of 3D points: medial surface, upper surface, and lower surface. Next, we learn the parameters of the intrinsic 2D coordinate systems of the surfaces from which each of these point sets were sampled via nonlinear dimensionality reduction. This process is depicted in figure 4.8 and is discussed below; for the moment we will treat this as a black box. The output of surface parameterization is, for each figure, three meshes, each with $N \times M$ vertices, depicting the medial surface, the upper surface, and the lower surface. Remembering that each 3D mesh vertex corresponds to a set of 2D $(i, j)$
parameters obtained during manifold learning, we form thickness vectors emanating from each vertex \((i, j)\) on the medial surface to corresponding vertices \((i, j)\) on the upper and lower surfaces. We have thus computed the basic building blocks for a medial representation of the figure: a medial surface consisting of a set of connected nodes, and a set of thickness vectors emanating from each of the nodes to the surface.

We now turn our attention to the surface parameterization process depicted in figure 4.8. Given a 3D binary image depicting a surface of interest, we extract the 3D points lying on the surface. Next, we perform nonlinear dimensionality reduction, also known as manifold learning, using the ISOMAP method [205]. ISOMAP forms a neighbourhood graph with the 3D points as nodes, with edges between each node and its nearest neighbours in 3D. There are two possible ways to choose the nearest neighbours. Using \(k\)-ISOMAP, the \(k\) nearest neighbours are chosen. Using \(\epsilon\)-ISOMAP, all of the neighbours lying within a sphere of radius \(\epsilon\) are chosen. The edges are weighted according to the Euclidean distance between neighbouring nodes. The geodesic distance between each pair of points is computed by integrating the graph edge weights along the shortest path between the points. A distance matrix is thus formed and used in classical multi-dimensional scaling [42] in order to perform dimensionality reduction preserving (insofar as it is possible to do so) inter-point geodesic distances, both locally and globally. Here, we use ISOMAP to reduce the dimensionality of the 3D points \(P\) to 2D points \(Q\), illustrated in figure 4.9. Because ISOMAP is a geodesic distance-preserving mapping, the resulting 2D coordinates are intrinsic to the 2-manifold sampled by the skeleton points. Next, we center an \(N \times M\) regular lattice of 2D nodes on the region containing points \(Q\). This is done by setting a margin size for the region,
and overlaying the regular lattice to fill the entire region specified by the margin size. The
motivation for the use of a margin is that in practice, we do not want medial or surface
nodes to fully extend to the edges of the skeleton or upper and lower surfaces. These nodes
form vertices of a quad mesh where edges join 4-connected neighbours. Lastly, each vertex
\( V \) of the mesh is mapped back to 3D by trilinear interpolation of points \( P \) corresponding
to neighbours \( N \in Q \) of \( V \). The end result of this process is that mesh vertices have \textit{equal geodesic spacing} along the surface. The cost of our method is bound by the cost of the
multidimensional scaling, which is \( O(n^3) \).

4.2.3 Results

Figure 4.10 illustrates the parameterization of the medial surfaces of images of the thalamus
and caudate nucleus segmented from an MRI of the brain. Specifically, this figure shows
the result of applying the process given in figure 4.8 to a binary image containing a single
medial surface. Parts (a) and (c) show the skeleton points in green, and parts (b) and (d)
show the meshes resulting from medial surface parameterization on a \( 6 \times 8 \) grid. Manifold
learning was conducted with \( k\)-ISOMAP \([205]\), \( k = 8 \).

Figure 4.11 is an artificial example illustrating the parameterization of an object’s skele-
ton in the multi-figural case. Figure 4.11(b) shows results obtained for an artificial multi-
surface skeleton (a), using \( \epsilon\)-ISOMAP \([205]\), \( \epsilon = 4 \).

Figure 4.12 illustrates the results of applying the entire process given in figure 4.7
to images containing a caudate nucleus and a supraspinatus muscle. In both cases, \( k\)-
ISOMAP \([205]\) was used, with \( k = 12 \). Note the even, regular distribution of medial and
surface nodes that is a consequence of spacing them along equal geodesic distances along
each surface. It is due to this equal geodesic spacing, in combination with the correspon-
dence established in the 2D coordinate systems intrinsic to each surface, that the thickness
vectors emanating from the medial surface always contact the correct object surface and do
not cross.

4.2.4 Contributions, limitations, and future work

The main contributions of the work described in this section and published in \([219]\) are:
Figure 4.10: (a) Left thalamus binary image (blue) with skeleton points (green) shown within. (b) Mesh resulting from medial surface parameterization of (a). (c),(d) Similar to (a) and (b), for left caudate nucleus image.
Figure 4.11: (a) A multi-surface skeleton of an artificial object. (b) The result of medial surface parameterization after clustering. Different medial surfaces are indicated with different colours.

- **A manifold learning approach to surface parameterization:** We applied manifold learning to discover the intrinsic parameterization of the medial and object surfaces, and used that parameterization to distribute medial nodes with a desired resolution, and to correspond them to object surface nodes. This information allows for the creation of thickness vectors that do not cross. Thus, we have the basic building blocks of any medial shape representation, such as medial patches (section 4.1) or M-Reps [155].

- **Correspondence establishment:** The equal geodesic spacing of the thickness vectors demonstrated in figure 4.12 leads to a type of correspondence establishment across shapes, which is the 3D analogue of arc-length correspondence in 2D. Such an approach could serve as an initialization to or regularization of a 3D version of the correspondence establishment method described in chapter 3.

The main limitations of the work described in this section are:

- **Non-normal thickness vectors:** Thickness vectors generated by our method are unlikely to be normal to either the medial or the object surfaces. This requires any medial shape representation built from these components to explicitly store the angle of each thickness vector within each medial node.
Figure 4.12: (a) A left caudate nucleus segmented from an MRI of the brain, rendered semitransparently as an isosurface. The medial surface is rendered as a surface within, with thickness vectors emanating to the parameterized object surface. (b) Similar to (a), for a supraspinatus muscle (part of the rotator cuff in the shoulder) extracted from MRI. (c) The parameterized surfaces of the caudate nucleus shown. The top surface is made translucent and the parameterized skeleton is removed for clarity. (d) Similar to (c), for the supraspinatus.
• **1:1 aspect ratio surfaces are problematic**: The multidimensional scaling step of the ISOMAP process takes a matrix of distances between points, and produces an embedding of those points into a lower dimensional space such that those inter-point distances are preserved (insofar as this is possible). Where the bounding box surrounding the low-dimensional representation of the set of points has an aspect ratio other than 1:1, it is trivial to place the intrinsic representations of the medial, upper, and lower object surfaces into a common coordinate system where these surfaces are aligned. However, when the bounding box has a 1:1 aspect ratio, its orientation is ambiguous and the alignment of these surfaces is problematic.

• **Need for ISOMAP parameter setting**: The success of this approach depends on accurate manifold learning. This involves setting the correct $k$ or $\epsilon$ parameter to ISOMAP. Setting these parameters too low causes the manifold to become disconnected, whereas setting these parameters too high can cause parts of the manifold to become connected when they should not be, altering the topology of the manifold.

• **No guarantee of non-intersection**: Although ISOMAP aims to find a global distance-preserving dimensionality reduction, this may not always be possible. In particular, it is possible for the map between the input high-dimensional points and the output low-dimensional points to not be one-to-one. It is also possible for a folding to occur in a surface as it is mapped to a lower dimensionality, resulting in crossing thickness vectors.

Future work based on the work described in this section includes:

• **Addressing the 1:1 orientation problem**: This problem can appear frequently enough that it requires addressing. One straightforward approach that should be attempted is to find the orientations of the medial, upper, and lower surfaces that yield the shortest thickness vectors; we expect this will yield the correct orientations.

• **Automatic ISOMAP parameter setting**: A robust system built on this work will require the manifold learning parameters to be set automatically. An approach that may be promising is to choose the smallest possible parameter that yields a single, connected manifold. An approach to choosing $k$-ISOMAP versus $\epsilon$-ISOMAP is not as straightforward, as it would depend (at least) on a measure of the even distribution of points along the manifold.
Chapter 5

Groupwise pruning of medial axes

5.1 Motivation and problem

Since the introduction of the medial axis transform (MAT) by Blum [21], there has been an intense research effort to make use of medial shape representations in approaches to problems involving shape analysis, shape retrieval, and shape optimization for image segmentation [63, 73, 85, 90, 92, 108, 116, 155, 157, 183, 190, 191, 194, 195, 219]. As described in section 1.4.3, medial shape representations permit the use of an object-centered coordinate system whose coordinates directly yield intuitively understood notions of shape, such as bending, thickness, and elongation. However, the use of medial shape representations presents a challenge arising from the well-known instability of the MAT with respect to small perturbations of the object boundary. This results in skeletons that, due to boundary noise and digitization artifacts, contain many unwanted branches. These extra branches pose a problem to both statistical shape modeling and object retrieval because they do not capture useful shape information and result in same-class objects having substantially different skeletal representations, both in terms of geometry and topology. It is this problem that is the focus of this chapter.

To address the problem of unwanted branches, many approaches to skeleton pruning have been developed. Most approaches reduce to computing a significance measure for every branch, followed by pruning the least significant branches first (see [36, 73, 149, 183, 190] and references within). Recently, Bai and Latecki have proposed a very promising approach based on discrete contour evolution [11]. Their algorithm works by iteratively partitioning the object’s contour into segments, and eliminating all skeleton branches whose generating
points lie on a single segment. Contour partitioning, and hence pruning, is based on a contour vertex relevance score that is a combination of the curvature of the contour at the vertex, and the lengths of the edges incident to the vertex. Bai and Latecki's algorithm is effective and efficient, requiring the tuning of a single parameter: the desired number of vertices in the simplified boundary polygon. Their method compares very favourably to several leading approaches, including [36, 73, 149].

Approaches to computing skeleton branch significance measures [11, 36, 73, 108, 145, 149, 183, 190, 194] utilize heuristics aimed at distinguishing branches corresponding to signal (i.e. information about an object's shape) from branches arising from noise (e.g. boundary noise, digitization artifacts). Thus, skeleton pruning can be viewed as a denoising problem. Approaches to computing skeleton branch significance can generally be dichotomized as being either local or global [190]. Local approaches compute a branch's significance measure using only information from a small neighbourhood of that branch, such as the length of the branch, the length and curvature of the boundary generating the branch, or the separation angle formed by connecting a point on the medial axis to closest points on the boundary [11, 149, 198]. Global approaches compute a branch's significance measure as the impact of removing the branch on the overall shape of the object, for example by measuring the impact on the boundary resulting from branch removal [108, 145, 194].

Despite substantial research into different local and global approaches to branch significance computation, improvements to branch pruning methods are still required to make them suitably powerful and general for use in computing medial representations for shape studies and object retrieval applications. The importance of this problem is highlighted by the research effort toward approaches requiring a priori knowledge of the expected, fixed topology of the skeletons, and deforming such skeletons to become approximately medial to the objects [73, 155, 157]. These approaches represent very substantial contributions, because by circumventing the problem of MAT instability with respect to boundary perturbations, they have made the use of medial shape representations practically applicable, and shown that, indeed, very useful and intuitive shape statistics are produced from medial shape models [63]. However, the requirement of a priori knowledge of the medial topology is onerous to the researcher. It can be impractical to satisfy in exploratory shape research, where the expected skeleton topology is unknown and to be discovered, or, more importantly, in cases where objects of a single class vary in skeleton topology. The latter can happen, for example, in medical studies of anatomical shapes where two patient groups
differ by skeletal topology, and it is precisely this difference that is to be discovered or tested in the medical hypothesis. Requiring the researcher to know and specify the required skeletal topology in these cases is undesirable. Thus, despite the significant progress made in using fixed-topology medial shape representations, it is worthwhile to continue research into automatic discovery of intrinsic skeletal topologies of objects via denoising or pruning of the MAT.

Inspired by Martin Styner’s approach to computing a common groupwise topology for shape studies using M-Reps [194], we propose a third class of approaches to computing branch significance: *groupwise* approaches. With the view that skeleton pruning is essentially a problem of distinguishing signal (i.e. branches arising intrinsically from the object’s noise-free shape) from noise (i.e. branches arising from anything else), we assert that a group of skeletons of objects from the same class contains a substantial amount of redundancy. We posit that this redundancy primarily conveys information related to the objects’ shapes, rather than to noise, since it stands to reason that objects drawn from the same class of shapes have common shape information encoded into their skeletons, whereas noise is uncorrelated and carries no redundancy. This idea is illustrated in figures 5.1 and 5.2, which demonstrate quantitatively and qualitatively that more redundancy exists in skeletons of groups of objects drawn from the same class. We propose that a groupwise approach to skeleton branch significance measurement can utilize this redundancy in order to distinguish branches corresponding to signal from those corresponding to noise. Our requirement of a group of objects is satisfied in many applications, including object retrieval, where a database of objects of a given class exists, and medical studies of anatomical shapes, where several groups of patients exist. Our philosophy is that valuable information is provided by the group of objects, so where such a group exists, this information should be utilized to yield more accurate pruning. Figure 5.3 contains an illustrative example showing the importance of groupwise information in computing branch significance, and the difficulty of making correct pruning decisions when skeletons are considered one at a time. The figure illustrates that the decision to prune a small branch depends on the context of the overall group; if that branch appears in every member of the group, it likely arises from signal, not noise, and should not be pruned early.

We describe here a framework, called the *Groupwise Medial Axis Transform* (G–MAT), for skeleton pruning. The G–MAT eliminates branches that are deemed to arise from noise,
Figure 5.1: Examples of the overlay of binary images of all skeletons from four different groups. The overlay is a histogram that approximates a probability density function of the likelihood location of a skeleton pixel. Top left: a group of shapes where every shape in the group is from a different class. Top right: fountain. Bottom left: corpus callosum (the corpus callosum is a brain structure that connects the two hemispheres of the brain). Bottom right: child. Lighter coloured points indicate greater amounts of skeleton overlap (i.e. higher probability). Overlap of binary images of the objects is shown in the background in grayscale for context; lighter shades of gray indicate greater amounts of object overlap. Note the dispersed histogram of the heterogeneous group, indicating higher entropy and less redundancy, and the opposite for the homogeneous groups. The binary images of the objects are similarity-transform registered prior to skeletonization.
Figure 5.2: A graph of the Shannon entropy [184] of a group of binary images of skeletons, versus the number of skeletons in the group. The coloured curves correspond to groups of skeletons all drawn from the same class of shapes, whereas the black curve corresponds to groups of skeletons where every skeleton in the group is drawn from a different shape class from the Brown LEMS database [186]. The Shannon entropy is computed based on a histogram where each bin corresponds to a 10 × 10-pixel region of the image, containing a count of skeleton pixels appearing in the region (figure 5.1).
Figure 5.3: Two groups of four skeletons each, one group on the left side of the vertical line, and one group on the right. Local and global pruning approaches may be inclined to prune the branches pointed to by the arrows, due to their seemingly small contributions to the overall shape. In the left-hand group, this pruning may be acceptable, but in the right-hand group, this branch corresponding to a small feature appears in every skeleton. Since it is unlikely that such a branch would appear so consistently due to noise, it more likely corresponds to an important (albeit small in size) feature arising from the shape of the objects, and it should be retained. A groupwise approach detects that this branch represents redundancy within the group, and therefore likely represents signal and not noise. In figure 5.7, we show an example where the G–MAT addresses this problem.
by measuring their support in the form of evidence from the group. The G–MAT approach also discovers common skeletal topologies within the group. The G–MAT relies on computing branch significance, or confidence measures, based on the information provided by a group of objects. At a high level, the G–MAT procedure consists of computing the conventional MAT of each object within the group. The procedure then locates skeleton junction points to separate each skeleton into individual branches, and then computes a set of topological and geometric features for each branch, and uses those features in a graph matching approach to establish branch correspondence across all of the skeletons. The set of corresponding branch features is used to compute a confidence value for each branch, based on the impact of removing the branch on the group. Figure 5.4 gives a qualitative preview of the results generated by our approach, in comparison to [11]. We observe an intuitively common topology and correct branch correspondence from our method in the bottom row, in contrast to the row above it, generated by [11]. We perform our computations and report results on 2D examples in this chapter, and comment on the generalization to 3D in section 5.6.

5.2 Method

The proposed method is iterative, where at each iteration a single branch is pruned from the entire group. The branch is selected according to the optimum of an objective function (5.1) defined in section 5.2.1. The objective function requires the calculation of a groupwise confidence measure for each branch. This confidence measure is computed using the G–MAT, described in section 5.2.2.

5.2.1 Skeleton pruning using groupwise information

We start with a set of $n_s$ binary images $\beta = \{\beta_1 \ldots \beta_{n_s}\}$ containing segmented and similarity transform-registered structures to be skeletonized. We then compute the conventional MAT [21] of each image in the set $\beta$, yielding a set of skeletons $S = \{s_1 \ldots s_{n_s}\}$, where $s_i$ is a set of loci and radii of the maximal disks that imply the boundary of the object $\beta_i$. Next, we locate each of the distinct branches in each skeleton, yielding $B = \{B_1 \ldots B_{n_s}\}$, where $B_i = \{b_{i1}, b_{i2}, \ldots, b_{ini}\}$, the set of $n_{bi}$ branches of skeleton $i$. For skeletons in 2D images, this is done by splitting the skeleton at its junction loci, which are defined as loci with a Hilditch crossing number [98] greater than 2.
Figure 5.4: First row: A group of 3 corpus callosum images. Second row: Conventional skeleton of those images, showing extraneous branches caused by small boundary perturbations. Third row: Results of a pruning using [11]. Fourth row: Pruned, branch-corresponded (correspondence indicated by colour) skeletons given by our pruning method based on the proposed G–MAT.
We then prune skeleton branches according to an iterative procedure. At each iteration, a single branch is pruned from the entire group of skeletons. This branch is selected such that it maximizes the reduction in group variability in the pruned skeletons; this choice yields the maximum increase in skeleton similarity within the group. The indices \((i^*, j^*)\) of the branch \(b_{ij}^*\) to be pruned are chosen as

\[
(i^*, j^*) = \operatorname{argmin}_{(i, j)} \gamma(b_{ij}),
\]

where

\[
(i, j) = \{(i, j) \in \mathbb{Z}^+ \times \mathbb{Z}^+ | i \leq n_s, j \leq n_b, b_{ij} \in B^e\},
\]

\(B^e\) is the set of all branches containing one or more end points (i.e. having a Hilditch crossing number [98] of 2). We require that candidate branches for pruning have at least one end point in order to guarantee that their pruning does not change the topology of the shape implied by the skeleton. The value \(\gamma(b_{ij})\) indicates the confidence that the branch arises from signal and not noise, according to the groupwise support given to the branch. We compute \((i^*, j^*)\) in (5.1) by brute force optimization. It is not onerous to do so since this involves the computation of \(\gamma(b_{ij})\) for \(|B|\) branches (\(|\cdot|\) denotes the cardinality of a set). Each pruning iteration \(i\) produces a new set of skeletons \(\tilde{S}^i\) and branches \(\tilde{B}^i\). Next, we address the question of how to compute \(\gamma(b_{ij})\) using the G–MAT.

### 5.2.2 The Groupwise Medial Axis Transform (G–MAT)

The pruning algorithm given in section 5.2.1 requires the calculation of the amount of overall support given to that branch by the group. We compute these values using the proposed G–MAT, which augments a group of conventional skeletons with a confidence value at each medial locus. Higher confidence values associated with a skeleton branch indicate a greater probability that the branch represents signal rather than noise, as determined by the support given to that branch by the group. The G–MAT begins with the computation of skeleton branches \(B\), as described in section 5.2.1. Next, a set of branch features \(F = \{F_1, F_2, \ldots, F_{ns}\}\) is computed, where \(F_i = \{F_{i1}, F_{i2}, \ldots, F_{in_b}\}\), the set of features computed for the branches in skeleton \(i\). \(F_{ij} = \{f_{ij1}, f_{ij2}, \ldots, f_{ijn_f}\}\), the set of \(n_f\) scalar features computed for branch \(b_{ij}\). In this work, we compute the following six features (i.e. \(n_f = 6\)):

1. the number of loci in the branch;
2. the \(x\) coordinate and
3. the \(y\) coordinate of
the branch centroid, expressed as a fraction of the width and height, respectively, of the structure; (4) the sum of the radius values along the branch (reflecting the size of the object part enclosed by the boundary implied by the branch); (5) the angle of the branch relative to the horizontal axis; and (6) the number of other branches sharing a junction point with the branch. The feature values are normalized to lie in the range \([0, 1]\). Note that our framework does not constrain the set of branch features used, and that the described features are chosen in this work in order to demonstrate the efficacy of the use of groupwise information in pruning.

The next step is to establish correspondence between each branch on each skeleton and a branch on every other skeleton, if such a correspondence exists. This correspondence is captured in \(\Pi = \{\pi_{ij} : 1 \leq i \leq n_s, 1 \leq j \leq n_s, i \neq j\}\), where \(\pi_{ij} : y \rightarrow z\) is a function mapping each branch index \(y\) in skeleton \(s_i\) to a branch index \(z\) in skeleton \(s_j\). This correspondence is established based on the agreement of features between branches. This is done by solving \(n_s \times (n_s - 1)\) bipartite graph matching problems, where each graph is considered to have two sets of disjoint vertices. The first set of vertices corresponds to the branches in one skeleton, and the second set of vertices to branches in another. We use the polynomial-time Hungarian algorithm [112] to solve these bipartite graph matching problems, thus computing the correspondences \(\Pi\).

We then compute confidence values \(\gamma(b_{ij})\) for every branch \(j\) of every skeleton \(i\), augmenting the skeletons \(S\) to become fuzzy skeletons \(\tilde{S}\). The tilde symbol appearing above an entity indicates that the entity is fuzzy, containing additional confidence value(s). The goal here is to compute \(\gamma(b_{ij})\) that reflect the groupwise support for each branch, in terms of its feature similarity to corresponding branches. We use these confidence values in skeleton pruning; see (5.1). We propose and test four different approaches to computing the confidence values, summarized by (5.2), (5.7), (5.18), and (5.22). These approaches each measure, in some way, the branch feature variability in the group of skeletons, and compute confidence values with the aim of reducing this variability by the pruning of low confidence branches. The four approaches differ in computational complexity and real computation speed (discussed further in section 5.5.2), and differ in the specific methods for computing group variability. Our goal in proposing and evaluating these four approaches is to provide the reader with sufficient information to select an approach presenting a reasonable compromise of speed versus performance. The four approaches are described in the following sections.
Confidence according to medial shape model description length (DL)

The description length (DL) [167] of a data set is the minimum number of bits required to transmit the data set from a sender to a receiver. We compute a medial shape model from a group of skeletons using \( F \) and \( \Pi \). Inspired by the work of Davies et al. [52] on DL optimization for shape boundary correspondence establishment, we compute the DL of the medial shape model to measure its compactness, and then set \( \gamma_{DL}(b_{ij}) \) according to the impact of removing \( b_{ij} \) on the DL. Branches that cause larger decreases in DL are more likely to be more dissimilar to the group (e.g. if they arise from noise) and therefore are accorded lower confidence values. Specifically,

\[
\gamma_{DL}(b_{ij}) = \frac{-(DL - DL_{ij}) - \gamma_{\min}}{\gamma_{\max} - \gamma_{\min}},
\]  

(5.2)

where \( DL \) is the description length of the model including \( b_{ij} \), \( DL_{ij} \) is the description length of the model with \( b_{ij} \) removed, and \( \gamma_{\min} \) and \( \gamma_{\max} \) are the smallest and largest confidence values observed in the group; thus \( 0 \leq \gamma(b_{ij}) \leq 1 \). We compute the DL as in [207], which requires the eigenvalues in the shape model. To obtain these eigenvalues, our first step is to form an \( n_s \times d_s \) matrix of observations in medial shape space, where each observation represents a single skeleton (each coordinate represents a single feature of a skeleton branch) and is of dimensionality \( d_s = n_f \times n_{b_{max}} \), where \( n_{b_{max}} = \max\{|B_i|, i = 1 \ldots n_s\} \), the largest number of branches observed in any skeleton. We arrange the mean-centered observations into an \( n_s \times d_s \) block matrix \( X \), \( X(i,j) = x_{ij} \), where each block \( X_{ij} \) is a \( 1 \times n_f \)-dimensional row vector of branch features from a single skeleton branch. The matrix \( X \) is arranged such that corresponding branches (according to \( \Pi \)) appear in the same column. We compute the contents of each block as \( X_{ij} = f_i(\pi_{k_i(j)}) \), where \( k = \arg\max_q(|B_q|, q = 1 \ldots n_s) \). In mean-centering the points, all coordinates other than those containing branch angles are averaged conventionally, in Euclidean space. Coordinates containing angle information are averaged along \( S^1 \). Since not all skeletons have the same number of branches, there exist some null values in \( X \), for which no information is available from \( F \). In order to penalize branches which do not correspond to branches in other skeletons (since such branches lack groupwise support, and should be pruned first), we replace these null entries with values that maximize the variance in the coordinates in which the null values occur (hence also maximally increasing the DL of the resulting model). For each column \( j \) of \( X \), we perform...
the following procedure once for each null entry. First, we compute the mean $m$ of all non-null values $x_{ij}, i = 1\ldots n_s$ in column $j$. Then, since all computed features lie in the range $[0, 1]$, we replace the null with 1 if $m < 0.5$; 0 otherwise.

Next, we perform a principal component analysis (PCA) \[106\] of $X$. We show the steps of this computation in detail here, as the intermediate quantities in the calculations will be used later in the chapter. The first step is to compute the $d_s \times d_s$ covariance matrix $C$ of $X$, where

$$C(i, j) = c_{ij} = \frac{1}{n_s - 1} \sum_{k=1}^{n_s} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)$$

and

$$\bar{x}_j = \frac{1}{n_s} \sum_{i=1}^{n_s} x_{ij}.$$  

The eigenvectors $p_i = (p_{i1}, \ldots, p_{id_s})^T, i = 1\ldots d_s$ of $C$ form the columns of the $d_s \times d_s$ matrix $P$. The $d_s \times 1$ vector of eigenvalues of $C$ is

$$(\lambda_1 \ldots \lambda_{d_s})^T = diag(P^T P).$$  

Using these eigenvalues, the DL is calculated (using \[207\]), the confidence of each branch is evaluated (according to (5.2)), and the branch to prune at the current iteration is selected (according to (5.1)).

**Confidence according to contribution to model variance (MV)**

In this method, we compute the contribution of each branch in each skeleton to the total variance of the medial shape model computed from the group of skeletons. Higher confidence values are given to branches making smaller contributions to model variance, since these branches are more similar to the other branches in the model. This confidence measure is based on the rationale that branches contributing large amounts of variance to the shape model are more likely to be outliers, dissimilar to corresponding branches in the group, and therefore more likely arising from noise. These confidence values are computed as

$$\gamma_{MV}(b_{ij}) \approx -\frac{\hat{\sigma}_{ij}}{\sigma_{ij}} \hat{\lambda}_j,$$  

where $\hat{\lambda}_j$ is the $j$-th largest eigenvalue and $\hat{\sigma}_{ij}$ is the estimated standard deviation of $x_{ij}$. 


where $\hat{\sigma}_{ij^*}$ is the squared feature difference of $b_{ij}$ from the mean of the features of $b_{ij}$ and all of its corresponding branches, $\sigma_{j^*}$ is the variance of the features of $b_{ij}$ and all of its corresponding branches, and $\hat{\lambda}_{j^*}$ is the contribution of $b_{ij}$ and all of its corresponding branches to the total variance $\sum_{i=1}^{d_s} \lambda_i$ in the medial shape model. Intuitively, $j^*$ is the index of the column of values $X_{j^*}$ containing the features of branch $b_{ij}$ and its corresponding branches. Formally, since $\sigma_{j^*}$ is the variance of the features of $b_{ij}$ and all of its corresponding branches, $j^* = \pi_{ik}(j)$, where $k = \arg\max_q(|B_q|, q = 1 \ldots n_s)$. The equality in (5.6) is approximate only due to the omission of a normalization that we show shortly. Next, we define each of the three components, $\hat{\sigma}_{ij^*}$, $\sigma_{j^*}$, and $\hat{\lambda}_{j^*}$, of (5.6) in (5.8), (5.10), and (5.11), culminating in the normalized definition such that $0 \leq \gamma_{MV}(b_{ij}) \leq 1$,

$$
\gamma_{MV}(b_{ij}) = \frac{-\hat{\sigma}_{ij^*} \hat{\lambda}_{j^*} - \gamma_{\text{min}}}{\gamma_{\text{max}} - \gamma_{\text{min}}}. \tag{5.7}
$$

The variance of the features of $b_{ij}$ and all of its corresponding branches is defined as follows. We compute $\sigma_{j^*}$ as

$$
\sigma_{j^*} = \frac{1}{n_s - 1} \sum_{i=1}^{n_s} ||X_{ij^*} - \bar{X}_{j^*}||_2, \tag{5.8}
$$

where

$$
\bar{X}_{j^*} = \frac{1}{n_s} \sum_{i=1}^{n_s} X_{ij^*}. \tag{5.9}
$$

The squared feature difference of $b_{ij}$ from the mean of the features of $b_{ij}$ and all of its corresponding branches is computed and normalized as

$$
\hat{\sigma}_{ij^*} = \frac{1}{n_s - 1} (X_{ij^*} - \bar{X}_{j^*})^2. \tag{5.10}
$$

The contribution of $b_{ij}$ and all of its corresponding branches to the total variance in the medial shape model is defined as

$$
\hat{\lambda}_{j^*} = \sum_{d=(j^*-1)n_f+1}^{j^*n_f} \hat{\lambda}_d, \tag{5.11}
$$

where $\hat{\lambda}_d$ is the total variance contribution of dimension $d$ ($1 \leq d \leq d_s$) of $X$ to the shape model and is defined to be
CHAPTER 5. GROUPWISE PRUNING OF MEDIAL AXES

\[ \hat{\lambda}_d = \sum_{i=1}^{d_s} \hat{\lambda}_{di}, \]  
(5.12)

where

\[ \hat{\lambda}_{di} = p_{di} \sum_{j=1}^{d_s} p_{ji} (c_{dj}^d + c_{jd}^d). \]  
(5.13)

The quantity \( \hat{\lambda}_{di} \) represents the contribution of dimension \( d \) of \( X \) to the variance represented by eigenvalue \( \lambda_i \) of the covariance matrix \( C \). Computing \( \hat{\lambda}_{di} \) therefore requires that we determine the individual contributions of dimensions \( i \) and \( j \) to each covariance \( c_{ij} \). We denote the individual contribution of dimension \( i \) to \( c_{ij} \) as \( c_{ij}^i \), and compute it as

\[ c_{ij}^i = g_{ij}^i c_{ij} \]  
(5.14)

(similarly for \( c_{ij}^j \) using \( g_{ij}^j \)). The value \( 0 \leq g_{ij}^i \leq 1 \) represents the fraction of contribution to \( c_{ij} \) by dimension \( i \). We define this fraction to be the relative magnitude of the factor arising from dimension \( i \) to each term of (5.3),

\[ g_{ij}^i = \sum_{k=1}^{n_s} \frac{|x_{ki} - \bar{x}_i|}{|x_{ki} - \bar{x}_i| + |x_{kj} - \bar{x}_j|}; \]  
(5.15)

\[ g_{ij}^j = 1 - g_{ij}^i. \]  

Having now defined \( c_{ij}^i \) and \( c_{ij}^j \), we justify (5.13) by observing that, using (5.5), each eigenvalue \( \lambda_i \) can be written as

\[ \lambda_i = \sum_{j=1}^{d_s} p_{ji} (\sum_{k=1}^{d_s} c_{jk}^i P_{ki}). \]  
(5.16)

The variance contribution of dimension \( d \) of \( X \) to \( \lambda_i \) can therefore be written as

\[ \hat{\lambda}_{di} = (\sum_{j=1}^{d_s} p_{ji} (\sum_{k=1}^{d_s} c_{jk}^d P_{ki}), j = d) + (\sum_{j=1}^{d_s} p_{ji} (\sum_{k=1}^{d_s} c_{jk}^d P_{ki}), k = d); \]  
(5.17)

simplification yields (5.13).
Confidence according to corresponding branch similarity (CBS)

In this approach, the confidence value computed for a branch is directly proportional to the agreement, according to a feature similarity metric, with the corresponding branch features in the group. To compute these confidence values, we first compute an error value $\epsilon_{ij}(\Pi, F)$ for each branch, which measures the mismatch between that branch’s features and the features of all other branches to which that branch corresponds, according to $\Pi$. We then compute the confidence value for all fuzzy skeleton loci on branch $b_{ij}$ as

$$
\gamma^{CBS}(b_{ij}) = 1 - \frac{\epsilon_{ij}(\Pi, F)}{(n_s - 1)n_f}.
$$

The denominator $(n_s - 1) \times n_f$ used to normalize the confidence values represents the maximum possible error that can be observed for a given branch; the maximum feature error $n_f$ multiplied by all $n_s - 1$ skeletons in which we attempt to find support from a matching branch.

The error value is computed for branch $b_{ij}$ as

$$
\epsilon_{ij}(\Pi, F) = \sum_{l=1}^{n_s} \sum_{k=1}^{n_s} d_F(F_{ij}, F_{ik} | \pi_{lk}(\pi_{il}(j))),
$$

where

$$
d_F(F_{pq}, F_{rs}) = \begin{cases} 
\sum_{k=1}^{n_f} d_f(k_{pq}, k_{rs}) & \text{if } b_{rs} \neq \emptyset, \\
\frac{1}{n_f} & \text{otherwise}
\end{cases}
$$

and $d_f$ is a distance function appropriate to the manifold of feature $k$. In this work, $d_f$ represents Euclidean distance for all features except for the branch angle feature, where $d_f$ is $2$ times the geodesic distance$^1$ on the manifold $S^1$. Since all feature values were normalized to lie in the range $[0, 1]$ prior to computing $d_f$, it follows that $n_f$ is the maximum observable feature difference for a pair of branches. Intuitively, $d_F(F_{pq}, F_{rs})$ is the feature distance between $b_{pq}$ and corresponding branch $b_{rs}$, if any. In the case where $b_{pq}$ has no such corresponding branch (i.e. $b_{rs} = \emptyset$), then the distance is computed to be the maximum possible error $n_f$.

$^1$We double the geodesic distance along $S^1$ since the maximum distance between angles normalized to the range $[0, 1]$ is $0.5$, rather than $1$ for the Euclidean features.
Although a correspondence between all branches of all skeletons can be established with a set of correspondences $\pi_{ij}$ for a fixed $j$ and $\forall i \neq j$, the accuracy of such a correspondence is affected by the choice of $j$. Instead, we perform all possible bipartite matchings in order to accumulate as much evidence of branch confidence as possible, without regard to the choice of a reference skeleton $j$.

Confidence according to overall branch similarity (OBS)

This approach is similar to the previous approach, except that we compute the overall branch dissimilarity of the group of skeletons as

$$E(\Pi, F) = \sum_{i=1}^{n_s} \sum_{j=1}^{\left| B_i \right|} \epsilon_{ij}(\Pi, F). \quad (5.21)$$

Then, to compute the confidence for a branch $b_{ij}$, we tentatively prune the branch, and compute a new set of branch features and correspondence for the skeletons with the branch removed. We then compute the above overall confidence $E_{ij}(\Pi, F)$ of this revised skeleton group. The confidence value assigned to $b_{ij}$ is then

$$\gamma_{OBS}(b_{ij}) = \frac{- (E - E_{ij}) - \gamma_{\min}}{\gamma_{\max} - \gamma_{\min}}. \quad (5.22)$$

Intuitively, $\gamma_{OBS}(b_{ij})$ then reflects the impact of the branch on the overall feature-wise agreement between corresponding branches in the skeleton group. Branches that cause larger decreases in $E$ are likely to be dissimilar to the group, and therefore are more likely to arise from noise. They are therefore accorded lower confidence values.

Regardless of which of the four described methods is chosen, the result is a measure of branch confidence $\gamma(b_{ij})$ that reflects the amount of groupwise support indicating that the branch arises from signal (shape) and not noise. As stated in section 5.2.1, $\gamma(b_{ij})$ is used in (5.1) to select the branch to prune at each iteration.

5.3 Validation

We perform both quantitative and qualitative validation of the proposed groupwise approach. Our validation methods are described in the following sections.
5.3.1 Skeletal denoising

The goal of this validation method is to evaluate the skeletal denoising performance of our approach, and to show that it compares favourably to a recent leading approach [11]. Here, ideal denoising refers to the pruning of skeletal branches arising from anything other than the shape of the object represented by the skeleton. In our experiment, we generate objects with known ground truth skeletons, and then add branches arising from noise. We then evaluate the capability of our method and [11] to yield the known ground truth skeletons via pruning.

The details of this validation are as follows. We construct a set of objects consisting of rectangles with peaks added to their boundaries (figure 5.5 gives examples). We start with a group of 5 identical such objects (i.e. the 5 rectangles have exactly the same distribution of peaks), yielding identical ground truth skeletons $S_{GT}$. We test using groups of only 5 objects since this small group size contains less data redundancy than, for example, a group of 50 objects. This tests the performance of our approach when minimal redundancy is present\(^2\). Next, we generate a set of 5 images $\beta$ by randomly adding noisy branches to the ground truth skeletons. We accomplish this by adding boundary noise to each shape with a specified noise probability $P_{Noise}$ (figure 5.5 gives representative samples of skeletons with different values of $P_{Noise}$). We then measure the denoising performance of a pruning algorithm $A$ by using it to prune the skeletons of $\beta$, yielding pruned skeletons $S_A$, and comparing $S_A$ to $S_{GT}$. The metric used to compare these skeletons is

$$V_{Noise} = |DT(S_A) - DT(S_{GT})|,$$  \hspace{1cm} (5.23)

where $DT(\cdot)$ is the distance transform of a skeleton represented as a binary image, with skeleton points as ones and background as zeros. Our rationale for comparing the distance transforms of the binary images of the skeletons is that this metric is zero where the skeletons are identical, and its value increases as skeletal branches in $S_A$ and $S_{GT}$ are further apart from each other. We perform an identical number of single-branch pruning operations (pruning until the skeletons each contain only one branch) using both the G–MAT and [11] (by varying the single parameter of [11] such that one branch is pruned per iteration), and select the iterations yielding the lowest $V_{Noise}$ for both approaches.

\(^2\)In section 5.3.5, we test the performance of our method for a range of group sizes.
5.3.2 Skeleton classification

This validation asks the further question of whether a pruning approach’s denoising capability leads to better classification accuracy of pruned skeletons. This test validates that differences in $V_{\text{Noise}}$ observed for different pruning methods have a material impact on a practical problem. In this test, we start with 2 groups of 5 objects, constructed in the same fashion as those in the denoising test. Within each group, the objects are identical (i.e. rectangles have the same size peaks in the same locations within each group), but the objects are different between the two groups. We then generate a set of 10 images $\beta$ by randomly adding boundary noise to each shape (which leads to noisy skeleton branches), with a specified noise probability $P_{\text{Noise}}$. We then measure the classification performance yielded by a pruning algorithm $A$ by using it to prune the skeletons of $\beta$, yielding pruned skeletons $S_A$ prior to classification. Each of these $m \times n$ images of skeletons is a single $m \times n$-dimensional feature vector representing the skeleton ($m \times n > 10000$ in our experiments). We train a support vector machine (SVM) to distinguish the groups based on skeleton feature vectors. In order to increase efficiency and reduce overtraining when feature vectors are large, it is common practice to perform feature selection or extraction prior to classifier training in order to reduce the dimensionality of the feature vectors. In our experiments, we perform feature extraction by dimensionality reduction, by performing PCA [106] on the $m \times n$ feature vectors. We retain enough coordinates to explain 95% of the variance in $S_A$, yielding a set of reduced-dimensionality skeletons $S'_A$ in the space of the retained eigenvectors from PCA. We then perform a cross-validation experiment (leave-one-out, performance averaged over 10 rounds) to measure the performance of a SVM in classifying $S'_A$ into the two ground truth skeleton groups. The metric used to measure the performance is the classification accuracy.
\[ V_{\text{Classification}} = \frac{N_{\text{Correct}}}{N}, \]  
where \( N \) is the number of trials in the cross-validation experiment, and \( N_{\text{Correct}} \) is the number of correct classifications given by the SVM. The best-case performances of the algorithms are compared, as in the previous validation.

### 5.3.3 Skeleton similarity

This validation is motivated by the rationale that skeletons of objects from the same class should be similar to each other, owing to the shape similarity between objects of the same class. In this test, we start with a group of 5 identical objects (i.e. the 5 rectangles have exactly the same distribution of peaks). We then generate a set of 5 images \( \beta \) by randomly adding boundary noise to each shape (which leads to noisy skeleton branches), with a specified noise probability \( P_{\text{Noise}} \). We then measure the skeleton-similarity performance of a pruning algorithm \( A \) by using it to prune the skeletons of \( \beta \), yielding pruned skeletons \( S_A \), and computing the similarity of the skeletons in \( S_A \) to each other. The metric used to compare these skeletons is

\[ V_{\text{Dissimilarity}} = DL(DT(S_A)), \]  
where \( DT(\cdot) \) is the set of distance transforms of the binary images of the skeletons in \( S_A \), and \( DL \) is the description length of a model created by PCA on the images of the distance transforms. Since the description length measures the compactness of the shape model, skeleton groups with smaller model description lengths are more internally similar. The best-case performances of the algorithms are compared, as in the previous validation. Note that in this validation, we measure the description length of a model built from distance transforms of binary images of skeletons; for fairness, this stands in contrast to our optimization of the description length of a medial shape model built from corresponding branch features using (5.2).

### 5.3.4 Brown LEMS silhouette classification

This validation is similar to that in section 5.3.2, but with silhouettes of objects from the Brown LEMS database [186] rather than artificial objects. The test is identical to that in section 5.3.2, except that the 10 images used come from two groups, where 5 of the
CHAPTER 5. GROUPWISE PRUNING OF MEDIAL AXES

Figure 5.6: This figure shows the fuzzy skeletons of corpus callosum images before pruning (\(\tilde{S}\)), and at pruning iterations 2, 17, and 21 in the first four columns (colourbar indicates map from colours to confidence values), and the correspondence between skeleton branches (same colours correspond) at iteration 21 in the fifth column. Notice that the two branches with lowest confidence (pointed to by arrow A) are pruned after the first two iterations (arrow B).

images are from one database category, and 5 of the images are from another. Classification accuracies for all combinations of 11 categories (see table 5.1) are evaluated.

5.3.5 Performance with respect to group size

In this test, we investigate the performance of the G–MAT relative to [11] for groups containing different numbers of skeletons, ranging from 2 to 50. We conduct the same validation as in section 5.3.2, using \(P_{\text{Noise}} = 0.05\) (see figure 5.5).

5.4 Results

Figure 5.6 illustrates the groupwise pruning process (using (5.18)) on a set of 5 images of corpora callosa. Note the pruning of low confidence branches first, eventually reaching a state where all skeletons are identical and branch correspondences appear to be correct.

Figure 5.7 illustrates the need for groupwise pruning, using fish silhouettes. The skeleton
CHAPTER 5. GROUPWISE PRUNING OF MEDIAL AXES

Figure 5.7: This figure shows the initial skeletons of the group of 4 fish in the first row, and the skeletons of the fish after pruning in the second row. Within each row, colours indicate correspondence. The small branches within the fish in the circled area (zoomed in the fifth column) are retained due to their groupwise support.

of each fish contains two small branches representing their heads, but in some fish these branches are extremely small; as small as two pixels each. An algorithm that computes the significance of these branches using a local, or even global, approach, would be tempted to prune them, but our groupwise approach (using (5.18)) retains them on account of their support from the remainder of the group. These branches, albeit small, are useful to retain because they represent a common skeletal topology within the group.

Figure 5.8 (top) shows the result of the skeleton denoising experiment described in section 5.3.1. For every noise probability $0 \leq P_{\text{Noise}} \leq 0.2$ (figure 5.5), 40 trials were run wherein a set of randomly noisy shapes was constructed with probability $P_{\text{Noise}}$, and their skeletons were pruned by each of the five pruning methods indicated in the figure (the four proposed methods and [11]). The denoising measure $V_{\text{Noise}}$ (5.23) was calculated for each method, and the means and standard deviations of $V_{\text{Noise}}$ are shown for each noise probability level. Figure 5.8 (bottom) is a visualization of the statistical significance of (our confidence in) the differences between every pair of curves in figure 5.8 (top), for every noise probability. In these experiments, we use a Bonferroni-corrected [1] statistical significance value of $\alpha = \frac{0.05}{20} = 0.0025$ due to the fact that for each pair of curves, we perform 20 comparisons for the 20 different values of $P_{\text{Noise}}$ tested. Bonferroni correction is the most conservative correction of the $\alpha$ level in multiple comparison tests, so the results in figure 5.8 (bottom) are therefore pessimistic with regard to the statistical significance of the differences between the curves in figure 5.8 (top). It is important to note that red markers in figure 5.8 (bottom) do not indicate inferiority of our method relative to [11]; relative performances are given at the top of the figure.

Figures 5.9, 5.10, and 5.11 show the results of the validation experiments described
Figure 5.8: Top: The means and standard deviations of $V_{\text{Noise}}$ for 40 random trials of a branch noise removal experiment (section 5.3.1) at increasing noise levels. The methods $\Gamma_{DL}$, $\Gamma_{MV}$, $\Gamma_{SBA}$, and $\Gamma_{OBA}$ use (5.2), (5.7), (5.18), and (5.22), respectively. Bottom: The statistical significance ($\alpha = 0.0025$) of the differences between each pair of curves in the top graph, at each noise level. Statistical significance is according to the colourbar, with green circles indicating statistically significant differences.
Figure 5.9: Top: The means and standard deviations of $V_{Classification}$ for 40 random trials of a branch noise removal experiment (section 5.3.2) at increasing noise levels. Meaning of legend is the same as in figure 5.8. Bottom: Statistical significance ($\alpha = 0.0025$) of the curve differences; meaning is the same as in figure 5.8.
Figure 5.10: Top: The means and standard deviations of $V_{\text{Dissimilarity}}$ for 40 random trials of a branch noise removal experiment (section 5.3.3) at increasing noise levels. Meaning of legend is the same as in figure 5.8. Bottom: Statistical significance ($\alpha = 0.0025$) of the curve differences; meaning is the same as in figure 5.8.
Figure 5.11: Top: The means and standard deviations of $V_{\text{Noise}}$ for 40 random trials of a branch noise removal experiment (section 5.3.3) at increasing skeleton group sizes, for $P_{\text{Noise}} = 0.05$. Meaning of legend is the same as in figure 5.8. Bottom: Statistical significance ($\alpha = 0.0042$) of the curve differences; meaning is the same as in figure 5.8.
in sections 5.3.2 (classification), 5.3.3 (similarity), and 5.3.5 (group size), respectively. In figure 5.11, $\alpha$ is Bonferroni-corrected to $\alpha = \frac{0.05}{12} = 0.0042$ because 12 different group sizes were tested.
Table 5.1: Table of classification accuracies in a cross-validation experiment where an SVM is trained using skeletons of Brown LEMS database [186] silhouettes. The skeletons are pruned using our groupwise approach (5.18) and [11]. Classification accuracies are indicated in each cell as (G–MAT, [11]). Green, underlined values indicate the winning approach, with black text indicating a tie.

<table>
<thead>
<tr>
<th></th>
<th>camel</th>
<th>dino</th>
<th>crown</th>
<th>stef</th>
<th>ethand</th>
<th>cat</th>
<th>elephant</th>
<th>heart</th>
<th>flightbird</th>
<th>hand</th>
<th>ray</th>
</tr>
</thead>
<tbody>
<tr>
<td>camel</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>dino</td>
<td>1.00</td>
<td>0.50</td>
<td>1.00</td>
<td>0.50</td>
<td>1.00</td>
<td>0.70</td>
<td>0.60</td>
<td>1.00</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>crown</td>
<td>1.00</td>
<td>0.50</td>
<td>1.00</td>
<td>0.40</td>
<td>0.90</td>
<td>0.50</td>
<td>0.50</td>
<td>0.80</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>stef</td>
<td>1.00</td>
<td>0.50</td>
<td>0.80</td>
<td>0.30</td>
<td>0.90</td>
<td>0.70</td>
<td>0.50</td>
<td>0.70</td>
<td>0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ethand</td>
<td>1.00</td>
<td>0.70</td>
<td>0.80</td>
<td>0.60</td>
<td>1.00</td>
<td>0.70</td>
<td>0.80</td>
<td>0.90</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cat</td>
<td>0.80</td>
<td>0.50</td>
<td>1.00</td>
<td>0.40</td>
<td>0.60</td>
<td>0.50</td>
<td>0.40</td>
<td>0.80</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>elephant</td>
<td>1.00</td>
<td>0.60</td>
<td>0.70</td>
<td>0.50</td>
<td>0.90</td>
<td>0.90</td>
<td>0.60</td>
<td>0.80</td>
<td>0.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>heart</td>
<td>1.00</td>
<td>0.90</td>
<td>1.00</td>
<td>0.80</td>
<td>1.00</td>
<td>0.80</td>
<td>0.80</td>
<td>1.00</td>
<td>0.60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>flightbird</td>
<td>1.00</td>
<td>0.50</td>
<td>0.90</td>
<td>0.30</td>
<td>0.60</td>
<td>0.70</td>
<td>0.60</td>
<td>0.90</td>
<td>0.90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hand</td>
<td>0.80</td>
<td>0.60</td>
<td>0.80</td>
<td>0.70</td>
<td>0.70</td>
<td>0.60</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ray</td>
<td>1.00</td>
<td>0.50</td>
<td>0.80</td>
<td>0.50</td>
<td>1.00</td>
<td>0.60</td>
<td>0.50</td>
<td>1.00</td>
<td>0.70</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.1 gives the results of the Brown LEMS validation experiment described in section 5.3.4. Each cell of the table shows the result of the groupwise approach (based on (5.18)), followed by the result obtained using [11]. Green, underlined values indicate the winning approach, with black text indicating a tie.

Figures 5.12 and 5.13 give qualitative results of our groupwise approach. Figure 5.13 (top two groups) in particular is intended to illustrate the effects of object occlusion on our groupwise method.

5.5 Discussion

5.5.1 Validation results

Because all of our validation tests take as input a set of binary images containing pruned skeletons, the tests are blind to any of the details of the pruning methods and are therefore fair tests. It should be noted, however, that this approach of fully decoupling the validation tests from the methods may lead to suboptimal results on the validation tasks, compared to what could be achieved if the validations were more tightly coupled to the pruning methods. For example, we might gain better object classification performance if, instead of training the SVM to classify binary images of skeletons, we provided it with our computed features $F$, corresponded by $\Pi$. This, however, would not represent a fair test of our performance against [11] since the validation procedure would be based on a set of corresponded features which our method optimizes, and [11] understandably does not. So, in order to be fair, all of our validation tests start with binary images of skeletons, the direct output of [11].

Figure 5.8 (top) indicates that our groupwise methods based on (5.18) and (5.22) have superior denoising performance to [11]. Moreover, the first and second rows of figure 5.8 (bottom) indicate that this difference is statistically significant (i.e. unlikely to be due to chance). Additionally, we observe in the fifth row of figure 5.8 (bottom) that the performance of the corresponding branch similarity method (5.18) is not significantly different from that of the overall branch similarity method (5.22), suggesting that the decision as to which of these approaches to use should be made based on their computational complexity, which will be discussed later in this section.

Figure 5.9 (top) indicates that our groupwise method based on (5.22) yields pruned skeletons that result in the most superior SVM classification performance compared to [11]. Figure 5.9 (bottom) indicates that its performance is statistically significantly better than
CHAPTER 5. GROUPWISE PRUNING OF MEDIAL AXES

Figure 5.12: Qualitative pruning results on silhouettes of bricks (top), maple leaves (middle), and hands (bottom). Within each group, the first row indicates the skeleton group before pruning, and the second row indicates the skeleton group after pruning by our groupwise method (5.18). Within each row, the computed branch correspondence is indicated by colour.
Figure 5.13: Qualitative pruning results on silhouettes of people with occlusions, forks, and hammers. Within each group, the first row indicates the skeleton group before pruning, and the second row indicates the skeleton group after pruning by our groupwise method (5.18). Within each row, the computed branch correspondence is indicated by colour.
that of [11] for $P_{\text{Noise}} \geq 0.02$. Our methods based on (5.7) and (5.2) generally provide significantly better performance than [11] for $P_{\text{Noise}} \geq 0.04$ and $P_{\text{Noise}} \geq 0.12$, respectively.

Figure 5.10 (top) indicates that our approaches based on (5.7), (5.18), and (5.22) yield skeletons which are more similar to each other, compared to [11]. The first three rows of figure 5.10 (bottom) indicate that these differences are statistically significant, although statistical significance is not achieved for (5.7) at high noise levels. As in figure 5.8 (bottom), we observe that the performance of the corresponding branch similarity method (5.18) is not significantly different from that of the overall branch similarity method (5.22).

Figure 5.11 shows the performance of the corresponding branch similarity method (5.18), compared with the performance of [11], as the size of the skeleton group increases. We observe a statistically significant difference in performance across group sizes ranging from 2 to 50 skeletons, in favour of our approach.

The results of these extensive validation experiments on artificial data are corroborated by the performance of our groupwise approach in providing skeletons suitable for classifying silhouettes from the Brown LEMS database [186]. Our results in table 5.1 show superior classification performance to [11] across a variety of shape classes from the database. This experiment, as well as the experiment resulting in figure 5.9, serve to illustrate that our method can handle class inhomogeneity within the group of skeletons being pruned. In both of these experiments, two different classes of objects’ skeletons are being pruned simultaneously by our groupwise method, and we report superior performance nevertheless. This is due to the fact that there exists sub-groupwise support within each homogeneous group.

5.5.2 Computational complexity

The time complexity of our groupwise approach to skeleton pruning is as follows. We begin at the stage where the skeletons $S$ have been computed, since the cost of computing the skeletons themselves is out of scope of the pruning process. The process of finding the branches of the set of skeletons $S$ is linear in the number of pixels in the skeletons, and thus is $O(|S|)$. Computing the features $F$ for the branches $B$ is also a linear operation and thus is $O(|B|)$. Solving the bipartite graph matching problem using the Hungarian algorithm [112] has complexity $O(n^3)$ where $n$ is the number of vertices in each graph. In our problem, each graph vertex is a skeleton branch. Computing the set of correspondences $\Pi$ between features using the Hungarian algorithm involves solving $|S| \times (|S| - 1)$ bipartite
CHAPTER 5. GROUPWISE PRUNING OF MEDIAL AXES

Problem 155

graph matching problems, where each graph has at most \( n_{b_{max}} \) vertices. Thus finding the correspondence \( \Pi \) is \( O(|S| \times (|S| - 1) \times n_{b_{max}}^3) \). Thus far, the complexity is \( O(|S| + |B| + |S| \times (|S| - 1) \times n_{b_{max}}^3) \). For simplicity in the following analysis, since the computation thus far yields a set of corresponding branch features, we write the preceding complexity as \( O(\Pi) \); a fifth-degree polynomial. The main contributor to this complexity is the computation of the correspondence \( \Pi \).

The remaining operation is to compute the confidence values. Computing \( \gamma_{DL}(b_{ij}) \) requires that we recompute the features and correspondence once for every tentative branch removal, requiring \( O(|B| \times (|B| + |S| \times (|S| - 1) \times n_{b_{max}}^3)) \). We must then build a shape model. This operation is bound by the time required to compute the eigenvalue decomposition of the covariance matrix, which is at best \( O(d_3^3) \) [185]. So, the overall process then requires \( O(\Pi) + O(|B| \times ((|B| + |S| \times (|S| - 1) \times n_{b_{max}}^3) + d_3^3)) \); a sixth-degree polynomial.

The time to compute \( \gamma_{MV}(b_{ij}) \) is bound by the time required to compute the eigenvalue decomposition. It is performed only once, requiring \( O(d_3^3) \). Thus, the overall process using this equation requires \( O(\Pi) + O(d_3^3) \); a fifth-degree polynomial.

The time to compute \( \gamma_{SBA}(b_{ij}) \) is bound by the need to compute the feature-wise differences between each branch in each skeleton and the corresponding (according to \( \Pi \)) branch in every other skeleton. This operation requires \( O(|S|^2) \). Thus, the overall process using this equation requires \( O(\Pi) + O(|S|^2) \); a fifth-degree polynomial.

Similarly as with computing \( \gamma_{DL} \), computing \( \gamma_{OBA}(b_{ij}) \) requires that we recompute the features and correspondence once for every tentative branch removal, requiring \( O(|B| \times (|B| + |S| \times (|S| - 1) \times n_{b_{max}}^3)) \). We then need to accumulate the feature-wise difference (computed as for \( \gamma_{SBA}(b_{ij}) \)) across the whole group for each branch, requiring \( O(|S| \times n_{b_{max}}^3) \). Thus, the overall process using this equation requires \( O(\Pi) + O(|B| \times (|B| + |S| \times (|S| - 1) \times n_{b_{max}}^3)) + O(|S|^3 \times n_{b_{max}}^3) \); a sixth-degree polynomial.

Although the total time complexities for all four groupwise approaches is a fifth- or sixth-degree polynomial, the speed of computation on a computer is substantially faster for (5.18) (SBA) than for the other approaches. On a 3.0GHz Pentium IV computer, one iteration of our Matlab (Natick, MA: Mathworks, 2008) implementation of the G–MAT for the group of 5 corpora callosa in figure 5.6 took 9.3 seconds using (5.2) (DL), 6.2 seconds using (5.7) (MV), 0.9 seconds using (5.18) (SBA), and 9.6 seconds using (5.22) (OBA). While these timings are certainly sensitive to differences in implementation across the methods, the main reason for the relatively faster computation of (5.18) (SBA) is evident when one considers
that $|S| \ll d_s$ and $|S| \ll n_{b_{\text{max}}}$ in most practical situations.

For comparison, the Bai and Latecki’s implementation of the method in [11] required 2.2 seconds to prune the same skeletons such that each of them had 4 end points. Our approach using (5.18) (SBA) required 6.9 seconds to iteratively prune the skeletons such that they each had four end points. This result is not surprising since our approach incorporates more information into the pruning decision. However, the computation time for our method using (5.18) is not onerous. We also note that there remain substantial speed gains to be made via further optimization of our implementation.

5.5.3 Multiple skeleton topology groups

The approach described in [11] has a single parameter to be tuned, which is the desired number of end points of the pruned skeleton. In essence, this parameter determines the extent to which skeletons should be pruned. Our approach has a similar parameter: the pruning iteration at which to stop. Just as a skeleton would vanish if the user requested that [11] prune a skeleton until it had zero end points, our iterative pruning approach, if allowed to run indefinitely, will eliminate all of the skeleton branches in the group. What we and the authors of [11] have proposed thus far is an approach to setting the pruning order of skeleton branches, not a means of deciding how many branches should remain in the skeletons after pruning.

Our groupwise approach, however, permits the computation of a quantity $\Gamma_{\text{intra}}$, the mean of the intra-class confidences observed in the skeletons. This measure reports the similarity of the skeletons at each iteration within their respective topological classes. Iterations having high values of $\Gamma_{\text{intra}}$ are useful because many applications, such as shape analysis and object retrieval, require that medial representations of objects of a given class be similar to each other. The details of the computation of $\Gamma_{\text{intra}}$ are as follows. At each iteration $i$ of the pruning algorithm given in section 5.2.1, we classify the skeletons $S^i$ into $n_c$ topological classes $\tilde{S}^{i1} \ldots \tilde{S}^{inc}$ according to the number of branches in each skeleton. We then compute the G–MATs of the skeletons in each of these classes individually, yielding sets of branches $\tilde{B}^{i1} \ldots \tilde{B}^{inc}$, features $F^{i1} \ldots F^{inc}$ and sets of correspondences $\Pi^{i1} \ldots \Pi^{inc}$. We then compute

$$\Gamma_{\text{intra}} = \frac{1}{n_c} \sum_{j=1}^{n_c} \frac{\Gamma(\tilde{B}^{ij})}{|\tilde{B}^{ij}|},$$

(5.26)
Figure 5.14: Plots of the mean intra-class confidence for iterations containing 1 (top) and 2 (bottom) topological classes of skeletons, respectively. The iterations with the top two values are indicated with squares, and topological samples from these iterations are given.
where $\Gamma(\tilde{B}^{ij})$ is the sum of all confidence values in the branches of $\tilde{S}^{ij}$. If an application requires all skeletons to have the same topology, then one is only interested in this measure for iterations where $n_c = 1$. An application such as anatomical shape analysis in medicine may be interested in iterations where $n_c > 1$, since such applications typically involve two or more patient groups whose anatomical shapes are hypothesized to vary with respect to some disease state. In medical research studies of this kind, the medial topologies of the different pathology groups is typically not known a priori; these and other types of shape differences are often exactly the information sought by the hypothesis. Therefore, approaches which presume the existence of an expert-determined, fixed-topology skeleton for the entire group are inapplicable here [155, 73]. On the other hand, using our approach, since it is often known that there exist $n$ different patient groups in the study, one can examine pruning iterations where $n_c = n$, and select a pruning iteration with maximal value of $\Gamma^{\text{intra}}$. Figure 5.14 illustrates this idea using the corpora callosa in figure 5.6. Graphs of $\Gamma^{\text{intra}}$ are shown for the pruning iterations where $n_c = 1$ and $n_c = 2$. Iterations corresponding to the top two values of $\Gamma^{\text{intra}}$ in both the one- and two-class cases are selected, and random samples of skeletons from these iterations can be shown to the user. These samples show the user the skeleton topologies (if more than one) present in the iterations shown. This approach permits a kind of focused, exploratory shape research that drastically reduces the need for a user to examine all skeletons of all iterations to decide on the best iteration, while still providing enough information for the user to make an informed decision. For applications requiring a fully-automated approach to selecting the correct pruning iteration, all that is required is that the desired value for $n_c$ be selected, and the iteration corresponding to the maximal value of $\Gamma^{\text{intra}}$ for iterations with $n_c$ topological classes is used.

5.6 Contributions, limitations, and future work

The main contributions of the work described in this chapter and published in [221] and [222] (submitted article) are:

- **Groupwise approach to computing branch pruning order:** In this work, we have proposed the Groupwise Medial Axis Transform (G–MAT), which is a novel approach to the determination of the pruning order of skeleton branches, using geometric and topological branch feature information provided by the group as a whole. This groupwise approach represents a third paradigm in computation of branch pruning
order, in addition to the existing paradigms of using local (single branch-based) and
global (whole shape-based) approaches.

- **Demonstration of superior performance to a leading method:** We tested four
approaches to computing the pruning order based on groupwise information, using
the description length of the medial model, the total variance of the medial model,
and the overall and individual groupwise branch similarities. We demonstrated, using
quantitative analysis, that one or more of the four variations of our approach is more
effective than a leading approach [11] in terms of removal of noisy skeleton branches,
classification of noisy artificial and real skeletons into different object classes, and in
producing skeletons which are similar for similar objects. We also demonstrated the
performance of the G–MAT using many qualitative examples from the Brown LEMS
database and elsewhere.

- **Detection of multiple topology groups:** We also demonstrated that we can use
the G–MAT to measure the internal similarity of skeletons within topological classes
at each pruning iteration, and use this information either to automatically select a
pruning iteration, or to provide a user/researcher with a sample of promising iterations
from which to choose. Thus, this approach allows for the computation of topologically
equivalent groups, or subgroups, of skeletons. This stands in contrast to the approach
in [11] where the user must know *a priori* the desired number of end points in the
skeleton after pruning, as well as to the approaches in [73] and [155] where the entire
skeleton topology needs to be known beforehand.

The main limitations of the work described in this chapter are:

- **Pathology groups must be well-represented:** Consider a hypothetical study with
99 normal patients and 1 abnormal patient, where the abnormal patient’s structure
of interest has a skeleton which differs from those of the normal patients by one or
more branches. Such a situation poses a problem for the G–MAT, since the additional
branches in the abnormal patient’s skeleton have little to no groupwise support; those
branches are likely to be erroneously pruned. Typically, however, patients are selected
for a study such that every pathology group is well-represented; otherwise it is difficult
to achieve statistical significance in the detected differences between groups.
• **Correspondence establishment is not truly groupwise:** We establish branch correspondence using a set of bipartite graph matchings between pairs of skeletons. This is not the same as establishing a single correspondence that takes into account the entire group of skeletons. We make our choice for reasons of computational complexity; whereas our approach to correspondence runs in polynomial time, the groupwise inexact graph matching problem is known to be NP-complete [18]. Thus, we were presented with two alternatives when deciding on an approach to branch correspondence establishment: finding a possibly suboptimal solution to the NP-complete problem of truly groupwise correspondence, or finding the optimal solution to the simplified problem of finding the set of optimal pairwise correspondences. We opted for the latter, but the former may give better results with enough computation time, and this may be desirable.

Future work based on the work described in this chapter includes:

• **Extension to 3D:** The main modification required for this work to be extended to 3D is a different means of splitting the 3D skeleton into its component manifolds and curves. There exists work on this problem that could be applied here [177].

• **Statistics on a group of skeletons with varying topology:** Another future direction is the application of this work to studies testing hypotheses relating anatomical shape to pathology, to explore the ability of the G–MAT to yield pruned skeletons having different topologies across groups, but similar topology within each group. This raises the interesting question of how to do the necessary statistics of shape on the sample when the medial topologies vary across pathology groups, and this problem will need to be addressed.

• **Feature selection:** There is also the issue of feature selection; although the geometric and topological branch features used in this work provided superior performance, there remain substantial opportunities for improvement using more sophisticated, pose-invariant features. In this direction, the possibility of applying automated feature selection to choose those features which best highlight redundancy in a particular group should be considered.

• **Initialization for existing skeleton fitting methods:** Finally, it would be of interest to test the utility of the pruned skeletons given by our approach as inputs to
existing approaches [155, 73] that deform predefined skeletons to fit new structures. Whereas previously, their work requires \textit{a priori} knowledge of the skeleton topologies to be used, our approach could serve as the initialization to theirs, eliminating the need for user-specified skeletal topology.
Chapter 6

Summary of contributions

This dissertation began with a description of very practical, health-oriented medical research on the relationship between shape and shoulder disorders in chapter 2, and delved deeper into the problems of shape correspondence (chapter 3), computing medial shape representations (chapter 4), and medial axis pruning (chapter 5). The details of all of the contributions pertinent to each chapter appear in sections 2.1.6, 2.2.7, 3.5, 4.1.5, 4.2.4, and 5.6. A summary of the overall contributions of this work is as follows:

- A medical research study relating the shape of the supraspinatus muscle to its pathological conditions. Specifically, we demonstrated that global shape descriptors can be used to train a classifier to distinguish retraction from atrophy, which can aid in the decision to proceed with surgery. The impact of this contribution is that it is a first step toward non-invasive diagnosis of the supraspinatus; currently the gold standard for diagnosis is arthroscopy. This work was described in section 2.1, and in our published papers [223, 224].

- A medical research study relating the shape of the bicipital groove of the proximal humerus to the pathological conditions of the long biceps tendon. We demonstrated the utility of global shape descriptors for this purpose, and designed a 3D, anatomy-specific medial-based shape descriptor for the bicipital groove. We demonstrated that our descriptor intuitively captures the measurements classically taken by medical researchers in 2D, and supported the automatic detection of the supratubercular ridge of Meyer [137]. We also showed that the representation permits useful qualitative and quantitative exploration of relevant measurements in an independent fashion. The
impact of this contribution is that it was the first 3D study of the bicipital groove as it relates to the long biceps tendon, thus eliminating for future medical researchers the difficult task of choosing specific 2D slices to use for analysis. Our approach is straightforward for medical researchers in this area to use, since it very naturally builds on the 2D measurements that have been taken of this structure for the past 80 years. This work was described in section 2.2, and in our published papers [225, 226, 228].

- An approach to the shape correspondence problem incorporating both shape and appearance information, with an approach to optimal section of features according to an expert-provided training set. This approach keeps the medical researcher involved in the shape study, and leverages his/her substantial experience and training in human anatomy, without the onerous requirements of fully manual landmarking or the explicit elicitation of shape and appearance features important to the structures. The approach also permits and encourages different sets of features to be used for corresponding different local regions of shapes, which is reasonable since it is expected that characterizing features are not the same for every part of a shape. The impact of this work is that it makes good, usable shape correspondences for statistical shape analysis a practical reality until a suitably accurate, fully automatic method for doing so is found. This enables much richer studies of shape, with the ability to report the relationship between pathological conditions and regional shape variations. This work was described in chapter 3, and in our published paper [220].

- An approach to medial shape representation that localizes shape deformations and differences to the specific regions where they occur; we call this the medial patch representation. Through the use of local coordinate systems constructed at each locus on every medial surface of the object, we relegate coordinates in a global frame of references to a single locus. This permits the straightforward global positioning of a shape in space, and eliminates the confounding effect of the global coordinate frame on the encoding of deformations and differences throughout the rest of the shape. The impact of this work is that it enables more straightforward, direct comparisons of objects, especially where they are different in terms of gross bendings. This work was described in section 4.1, and in our published paper [90].

- A manifold learning-based approach to the parameterization of medial and object
surfaces, in order to compute the basic building blocks of any medial shape representation. Using manifold learning, we distribute nodes on the medial surfaces of an object’s skeleton with equal geodesic spacing. We similarly distribute corresponding nodes on the upper and lower surfaces of the object, giving the termination points for thickness vectors emanating from the medial nodes. The impact of this work is to provide a straightforward means of computing the components of an object’s medial representation, based on its pruned skeleton. Because our approach contains an element of nonlinear dimensionality reduction, it respects the natural curves and foldings of the medial and object surfaces, and ensures that thickness vectors do not cross and cause self-intersection of the implied boundary. This work was described in section 4.2, and in our published paper [219].

• A groupwise approach to skeleton pruning, that uses the redundancy in the skeletons of same-class objects as a means of differentiating branches corresponding to signal (true, uncorrupted shapes) from those arising from noise, digitization artifacts, or other sources. We demonstrate that this redundancy can also be used to determine the existence of multiple topological classes of objects in a group, making possible the detection of pathology groups whose shapes differ according to skeletal topology. We also contribute several means of fairly evaluating the performance of skeleton pruning algorithms, in terms of their ability to provide denoised skeletons which are similar for same-class objects, and enable accurate automated classification of objects of different classes. The validations, being based on binary images of pruned skeletons, are blind to and decoupled from any parameterization or other internal elements of the algorithms that generated the skeletons, and are thus fair and equitable benchmarks. The impact of this work is that it represents a new paradigm with respect to the scope of information used in determining the pruning order of branches, and we quantitatively demonstrated superior pruning results relative to a recent, leading method. This work was described in chapter 5, in our technical report [221], and in our submitted journal article [222].
Appendix A

Statistical shape modeling

A.1 What is a statistical shape model?

Given a set of shapes, one can construct a statistical shape model of these shapes, which serves to indicate the range and types of shape variability seen in the set. Statistical shape models are useful because if they are constructed with a sufficiently large, random sample of shapes, they can give information about the plausibility of a new shape. This is useful, for example, in deformable model-based segmentation, as an internal constraint on the deformable model, to ensure that image noise does not drive the optimizer to an implausible result. It can also be useful in object classification, where one can evaluate the plausibility of a new shape relative to a database of statistical shape models of different object classes, choosing the object class whose model indicates highest plausibility of the new shape.

A.2 Computing point distribution models

The point distribution model (PDM) is a widely-adopted approach to statistical shape modeling [37]. The first step in computing a point distribution model is to eliminate variations of pose (defined as global translation, rotation, and scale) from the set of shapes under consideration. This is done using generalized Procrustes analysis, wherein the variations in overall translation, rotation, and scale are removed in the set of shapes [55]. This process is illustrated in figure A.1. To perform this normalization, a set of corresponding landmarks, points lying on the object surfaces, is required. Normalization of translation is performed by translating all landmarks such that their centroid becomes the origin. Normalization of
Before Procrustes analysis, pose and shape information is combined. After Procrustes analysis, pose information is normalized.

Figure A.1: Procrustes analysis normalizes for pose parameters of translation, scaling, and rotation, so that the only variations that remain between landmark coordinates (coloured circles) are due to shape differences.

Scale is performed by computing the size $s$ of the object, and then dividing all landmark coordinates by $s$ (after normalization for translation). Normalization of rotation is performed by a least-squares technique yielding minimal distances between corresponding landmarks across the group. Sometimes the shapes are not normalized for scale, if the size of the object is also important in the analysis (in this case we are doing size-and-shape analysis). As an interesting historical note, the name Procrustes refers to a bandit from Greek mythology, who is said to have lured guests to lie down in his iron bed. Should a guest be too tall for the bed, Procrustes would amputate his limbs so that he fit, and should a guest to be too short, Procrustes would stretch him on the rack until he was the right size. In doing Procrustes analysis we are doing essentially the same thing to shapes, so that they all have approximately the same size, position, and orientation.

Once pose has been normalized, all of the variation between landmark coordinates is due to shape differences. As this stage we construct a model of these shape variations. The process is depicted in figure A.2. The $m$ $d$-dimensional landmarks on each shape are serialized to form a single $p = m \times d$ length vector. These vectors are then mean-centered and arranged such that each of the vectors forms a row of an $n \times p$ matrix.
Each triangle is a 6D point in shape space. Note that coordinates correspond.

Principal component analysis yields a shape model.

Figure A.2: The corresponding coordinates (correspondence indicated by the colours of the landmarks) are used to form a matrix, where each row depicts a single shape. Note that coordinates in a given column of the matrix all correspond (have the same colour). Principal component analysis is performed on this matrix, yielding the mean shape, and the eigenvectors and eigenvalues corresponding to the major directions and magnitudes of observed shape variation.

\[
X = \begin{pmatrix}
    x_{11} & x_{12} & \cdots & x_{1p} \\
    x_{21} & x_{22} & \cdots & x_{2p} \\
    \vdots & \vdots & \ddots & \vdots \\
    x_{n1} & x_{n2} & \cdots & x_{np}
\end{pmatrix},
\] (A.1)

where each row of \(X\) represents a single observation (shape) in \(p\)-dimensional shape space. Next, we perform a principal components analysis (PCA) \([106]\) of the points in \(X\). The first step is to compute the \(p \times p\) covariance matrix \(C\) of \(X\) as

\[
C = \begin{pmatrix}
    c_{11} & c_{12} & \cdots & c_{1p} \\
    c_{21} & c_{22} & \cdots & c_{2p} \\
    \vdots & \vdots & \ddots & \vdots \\
    c_{p1} & c_{p2} & \cdots & c_{pp}
\end{pmatrix},
\]

where

\[
c_{ij} = \frac{1}{n-1} \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \quad \text{and} \quad \bar{x}_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}.
\]

The eigenvectors \(p_i, i = 1 \ldots p\) of \(C\) form the columns of the \(p \times p\) matrix
The $p \times 1$ vector of eigenvalues of $C$ is

$$(\lambda_1 \ldots \lambda_p)^T = \text{diag}(P^T C P).$$

The shape model then consists of the mean shape, the eigenvectors given in $P$, and the eigenvalues $\lambda_i$, $1 \leq i \leq p$. The eigenvectors represent the major modes of observed variation in the shape sample, and the eigenvalues give their magnitudes; they show the amount of variation accounted for by each of the eigenvectors. Typically, one sets a threshold $t$, where for example if $t = 95$, one retains all eigenvectors and eigenvalues accounting for 95% of the total observed variation. The lower $t$ is set, the coarser is the shape model. If $t$ is set to 100, then the computed shape model captures all of the variation in the observed shapes.

There are several alternative approaches to computing statistical shape models. In all approaches the goal is to build a model capturing shape variability, usually by estimating a probability density function of different shapes. Many of these approaches have the objective of addressing the issue of non-linearity of the arrangement of points in shape space [27, 28, 29, 39]. Rogez et al. propose a combined approach based on a PDM and a Gaussian mixture model which captures remaining nonlinearity not described by the PDM [170]. Al-Shaher et al. followed up on this research with a proposal to learn a mixture of PDMs to describe non-linear shape distributions [4, 5]. Other approaches have aimed to address the nonlinearity problem by using different approaches to the dimensionality reduction step in constructing the PDM. Cootes et al. [37] construct a model similar to a PDM that captures both shape and appearance. Davatizkos et al. [47] construct a hierarchical active shape model, using the wavelet transform to build the hierarchy.

There exist several alternative approaches to PCA. Independent component analysis (ICA) [101, 235] is a type of blind source separation, where the input data is separated into statistically independent parts. Factor analysis [2, 77] is a method originating in psychometrics, and is very similar to PCA in that the observations are modeled as linear combinations of other variables called factors. The main difference in factor analysis is that the variances in the errors (the parts of the observations not explained by the factors) can be unequal.
Kernel PCA [45, 164, 171, 178, 213] is an extension of PCA to use kernel methods. The basic idea is to apply a function called a kernel to the data, to transform it into a different space prior to dimensionality reduction. The main reason for doing this is that the intrinsic modes of variation in the observations is nonlinear and therefore not discoverable by PCA alone. By mapping the data into another space via a nonlinear kernel function, the main modes of variation in the data can become linear and thus discoverable by PCA. Nonlinear manifold learning techniques [16, 175, 205] are related to this, in that they attempt to find a mapping of the observations into a lower-dimensional space, such that the inter-observation distances, along the nonlinear manifold in which they intrinsically lie, remain the same after the dimensionality reduction.

Where a medial-based, rather than boundary-based, shape representation is used, the coordinates of the landmarks frequently do not live in a linear space. In the case of the M-Reps shape representation (section 1.4.3), the components of the representation have been shown to form a nonlinear space [63]. Consequently, PCA is not applicable as a dimensionality reduction method for building statistical shape models based on the M-Reps representation. A generalization of PCA, called Principal Geodesic Analysis (PGA) has been developed to address this problem [63]. Using PGA, the notions of the mean shape and the principal modes of variation have been generalized to their intrinsic counterparts within the nonlinear space spanned by M-Reps. Thus, PGA determines the mean shape and principal geodesics of variation in the nonlinear space.
Bibliography


