

# Estimation of Probability Distribution on Multiple Anatomical Objects and Evaluation of Statistical Shape Models

Ja-Yeon Jeong

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill  
2009

Approved by:

Stephen M. Pizer, Advisor

Surajit Ray, Reader

J. S. Marron, Reader

Martin A. Styner, Reader

Edward L. Chaney, Reader

© 2009  
Ja-Yeon Jeong  
ALL RIGHTS RESERVED

## **Abstract**

**Ja-Yeon Jeong: Estimation of Probability Distribution on Multiple Anatomical Objects and Evaluation of Statistical Shape Models  
(Under the direction of Stephen M. Pizer)**

The estimation of shape probability distributions of anatomic structures is a major research area in medical image analysis. The statistical shape descriptions estimated from training samples provide means and the geometric shape variations of such structures. These are key components in many applications.

This dissertation presents two approaches to the estimation of a shape probability distribution of a multi-object complex. Both approaches are applied to objects in the male pelvis, and show improvement in the estimated shape distributions of the objects. The first approach is to estimate the shape variation of each object in the complex in terms of two components: the object's variation independent of the effect of its neighboring objects; and the neighbors' effect on the object. The neighbors' effect on the target object is interpreted using the idea on which linear mixed models are based.

The second approach is to estimate a conditional shape probability distribution of a target object given its neighboring objects. The estimation of the conditional probability is based on principal component regression.

This dissertation also presents a measure to evaluate the estimated shape probability distribution regarding its predictive power, that is, the ability of a statistical shape model to describe unseen members of the population. This aspect of statistical shape models is of key importance to any application that uses shape models. The measure can be applied to PCA-based shape models and can be interpreted as a ratio of the variation of new data explained by the retained principal directions estimated from training data. This measure was applied to shape models of synthetic warped ellipsoids and right hippocampi. According to two surface distance measures and a volume overlap measure it was empirically verified that the predictive measure reflects what happens in the ambient space where the model lies.

## Acknowledgments

I would like to express my gratitude to people who have been important to this work and to my life as a graduate student at UNC.

First and foremost I am thankful to Steve Pizer, my doctoral advisor. Without his guidance and encouragement I do not think I would be able to finish this dissertation. His patience in reading draft after draft of my dissertation chapters amazed me. I sometimes felt he had more confidence in me than I had in myself. I thank him for his willingness to make time for me whenever I needed his help. I have learned a great deal from him, not only from his immense knowledge and understanding of the field but also from his positive attitude, patience, and passion for his work. He has been a great mentor to me.

I am grateful to my committee members for their comments and suggestions. Surajit Ray has been actively involved in my dissertation research and has been of great help in developing major ideas in this dissertation. Steve Marron has lent me freely his invaluable insight into statistical issues. I believe he can visualize any statistical concepts with several color pens and white papers. I want to thank Ed Chaney for his critical comments and advices as a medical physicist and Martin for his feedback and practical advices. I also want to thank Keith Muller at UF for investing time and energy discussing ideas with me. Many of the ideas in this dissertation originated from discussions with them.

I have been fortunate to have the people in MIDAG as my colleagues. They are good, hard-working, and smart. Especially, I want to thank Josh Levy, Joshua Stough, Qiong Han, Graham Gash, and Gregg Tracton who I pestered with questions, and my last office mate Rohit Saboo, two female members Xiaoxiao Liu, and Ilknur Kabul whose office I frequently visited whenever I wanted to take break from work.

I would like to thank faculty members and staffs in the department. My special thanks go to Russ Taylor and Leandra Vicci for their encouragements during my gloomy first year, and to Janet, Tammy, Linda, and Tim for their help in all (non-research related) problems that I encountered as a graduate student.

I wish to thank elders and my friends in First Baptist Korean Church of Raleigh for providing a loving and supportive environment for me.

Lastly, and most importantly, I am grateful to my parents who have supported me with constant love and understanding. To them I dedicate this thesis.

# Table of Contents

|  |           |
|--|-----------|
| <b>List of Tables</b> . . . . .                                | <b>x</b>  |
| <b>List of Figures</b> . . . . .                               | <b>xi</b> |
| <b>1 Introduction</b> . . . . .                                | <b>1</b>  |
| 1.1 Motivation . . . . .                                       | 1         |
| 1.1.1 Statistical Shape Analysis of Multiple Objects . . . . . | 4         |
| 1.1.2 Quality Measures of Statistical Shape Models . . . . .   | 7         |
| 1.2 Thesis . . . . .   | 11        |
| 1.3 Claims . . . . .   | 11        |
| 1.4 Overview of Chapters . . . . .                             | 12        |
| <b>2 Background</b> . . . . .                                  | <b>13</b> |
| 2.1 Statistical Background . . . . .                           | 13        |
| 2.1.1 Probability Distributions . . . . .                      | 13        |
| 2.1.1.1 Definitions . . . . .                                  | 13        |
| 2.1.1.2 Expectations of a Random Variable . . . . .            | 16        |
| 2.1.1.3 Special Distributions . . . . .                        | 17        |
| 2.1.1.4 Statistics: Sample Mean and Variance . . . . .         | 18        |
| 2.1.1.5 Multivariate Distributions . . . . .                   | 19        |
| 2.1.1.6 The Multivariate Normal Distribution . . . . .         | 20        |
| 2.1.2 Principal Component Analysis . . . . .                   | 22        |
| 2.1.3 The Multivariate Linear Model . . . . .                  | 26        |
| 2.1.3.1 Principal Component Regression . . . . .               | 27        |
| 2.2 The Statistical Theory of Shape . . . . .                  | 29        |
| 2.2.1 Kendall's Shape Space . . . . .                          | 30        |
| 2.2.2 Estimation of a Shape Probability Distribution . . . . . | 31        |

|          |   |           |
|----------|---|-----------|
| 2.2.3    | Alignment . . . . .   | 33        |
| 2.2.3.1  | Alignment for multi-object complexes . . . . .                          | 34        |
| 2.2.4    | Correspondence . . . . .  | 35        |
| 2.3      | Probabilistic Deformable Models . . . . .                               | 36        |
| 2.3.1    | Deformable Models . . . . .   | 36        |
| 2.3.2    | Probabilistic Deformable Models: Segmentation by Posterior Optimization | 37        |
| 2.3.3    | Object Representations via Landmarks . . . . .                          | 38        |
| 2.3.3.1  | Point Distribution Model . . . . .                                      | 39        |
| 2.3.4    | Object Representations via Basis Functions . . . . .                    | 39        |
| 2.3.4.1  | Elliptic Fourier Representation . . . . .                               | 40        |
| 2.3.4.2  | Spherical Harmonics Shape Description . . . . .                         | 40        |
| 2.4      | Medial Representations . . . . .  | 42        |
| 2.4.1    | The Medial Locus . . . . .  | 43        |
| 2.4.2    | M-rep . . . . .   | 48        |
| 2.4.2.1  | Discrete M-reps . . . . .   | 48        |
| 2.4.3    | Statistical Shape Analysis of M-reps . . . . .                          | 50        |
| 2.4.3.1  | Shape Space of M-reps and PGA . . . . .                                 | 50        |
| 2.4.3.2  | Tangent Space at a Point of a Manifold . . . . .                        | 50        |
| 2.5      | Construction of Training Models . . . . .                               | 51        |
| 2.6      | Segmentation of M-reps . . . . .  | 52        |
| <b>3</b> | <b>Evaluation of Statistical Shape Models</b> <sup>1</sup> . . . . .    | <b>55</b> |
| 3.1      | Background . . . . .  | 56        |
| 3.1.1    | Decomposition of the Covariance Matrix . . . . .                        | 56        |
| 3.1.2    | PCA for Statistical Shape Analysis . . . . .                            | 57        |
| 3.2      | Goodness of Prediction . . . . .  | 58        |
| 3.2.1    | PCA Input to Multivariate Regression . . . . .                          | 58        |
| 3.2.2    | Measure of Association: Second Moment Accuracy . . . . .                | 59        |
| 3.2.3    | Procedure for Iterative Calculation of $\rho^2$ . . . . .               | 62        |
| 3.3      | Derivation of B-reps from M-reps . . . . .                              | 63        |
| 3.4      | Application of $\rho^2$ to Models in Linear Space . . . . .             | 63        |

|          |  |           |
|----------|--|-----------|
| 3.4.1    | Simulated Ellipsoid M-reps . . . . .                               | 64        |
| 3.4.2    | Experiments on Simulated Ellipsoid B-reps . . . . .                | 65        |
| 3.4.3    | Right Hippocampus M-reps . . . . .                                 | 67        |
| 3.4.4    | Experiments on Right Hippocampus B-reps . . . . .                  | 68        |
| 3.5      | Distance Measures for $\rho^2$ Evaluation . . . . .                | 69        |
| 3.5.1    | Application of $D_*^2$ to Right Hippocampus B-reps . . . . .       | 71        |
| 3.6      | Goodness of Prediction $\rho_d^2$ for Curved Manifolds . . . . .   | 72        |
| 3.6.1    | Two Possible Extensions of $\rho^2$ . . . . .                      | 72        |
| 3.6.2    | $\rho_d^2$ for Nonlinear Shape Models . . . . .                    | 73        |
| 3.7      | Application of $\rho_d^2$ on Models in Nonlinear Space . . . . .   | 74        |
| 3.7.1    | Deformed Binary Ellipsoids . . . . .                               | 74        |
| 3.7.2    | Experiment on M-rep Fits to Deformed Binary Ellipsoids . . . . .   | 75        |
| 3.7.3    | Experiment on Right Hippocampus M-rep . . . . .                    | 76        |
| 3.7.4    | Evaluation of a Coarse-to-fine Shape Prior . . . . .               | 77        |
| 3.8      | Conclusion & Discussion . . . . .                                  | 79        |
| <b>4</b> | <b>Multi-Object Statistical Shape Models<sup>2</sup> . . . . .</b> | <b>82</b> |
| 4.1      | M-rep Operations . . . . .   | 85        |
| 4.1.1    | Residues of the Object Variations . . . . .                        | 86        |
| 4.2      | Inter-Object Relations . . . . .                                   | 87        |
| 4.2.1    | Augmentation . . . . .   | 87        |
| 4.2.2    | Prediction . . . . .   | 88        |
| 4.3      | Propagation of Sympathetic Changes . . . . .                       | 89        |
| 4.3.1    | Residues of Objects in Order . . . . .                             | 92        |
| 4.3.2    | Training the Probability Distribution per Object . . . . .         | 92        |
| 4.3.3    | Geometrically Proper Objects . . . . .                             | 93        |
| 4.4      | Decomposition of Shape Variations . . . . .                        | 97        |
| 4.4.1    | Self Variations & Neighbor Effects . . . . .                       | 98        |

|          |  |            |
|----------|--|------------|
| 4.4.2    | An Iterative Method . . . . .                                    | 99         |
| 4.4.3    | Joint Probability of Multiple Objects . . . . .                  | 101        |
| 4.4.4    | Shape Prior in MAP-Based Segmentation . . . . .                  | 102        |
| 4.4.5    | Segmentation of Male-Pelvis Model . . . . .                      | 103        |
| 4.4.5.1  | Probability Density Estimation . . . . .                         | 103        |
| 4.4.5.2  | Segmentation Results . . . . .                                   | 105        |
| 4.5      | Discussion and Conclusion . . . . .                              | 108        |
| <b>5</b> | <b>Conditional Shape Statistics<sup>3</sup></b> . . . . .        | <b>111</b> |
| 5.1      | Conditional Probability Distribution . . . . .                   | 112        |
| 5.2      | Estimation of Conditional Shape Distribution using PCR . . . . . | 113        |
| 5.3      | Applications of CSPD to Deformable M-rep Segmentation . . . . .  | 115        |
| 5.3.1    | Simulated Multi-objects . . . . .                                | 115        |
| 5.3.1.1  | Training Results . . . . .                                       | 117        |
| 5.3.2    | Objects in the Pelvic Region of Male Patients . . . . .          | 119        |
| 5.3.2.1  | Results of Within-patient Successive Segmentations . . . . .     | 120        |
| 5.4      | Discussion and Conclusion . . . . .                              | 123        |
| <b>6</b> | <b>Discussion and Conclusion</b> . . . . .                       | <b>125</b> |
| 6.1      | Summary of Contributions . . . . .                               | 125        |
| 6.2      | Evaluation of Statistical Shape Models . . . . .                 | 131        |
| 6.2.1    | Discussion . . . . .   | 131        |
| 6.2.2    | Future Work . . . . .  | 137        |
| 6.3      | Multi-Object Statistical Shape Models . . . . .                  | 137        |
| 6.3.1    | Discussion and Future Work . . . . .                             | 137        |
|          | <b>Bibliography</b> . . . . .                                    | <b>140</b> |

## List of Tables

|     |   |     |
|-----|---|-----|
| 4.1 | Total variations of self and neighbor effects per bladder, prostate, and rectum . | 104 |
| 5.1 | Number of fractions (sample size) per patient . . . . .                           | 118 |
| 5.2 | Mean volume overlaps of figure 5.3 . . . . .                                      | 121 |
| 5.3 | Number of principal geodesics estimated from patient 3106 data . . . . .          | 121 |

## List of Figures

|      |   |    |
|------|---|----|
| 2.1  | Different classes of medial points . . . . .  | 45 |
| 2.2  | Medial geometry of end atoms . . . . .  | 47 |
| 2.3  | A discrete m-rep . . . . .  | 49 |
| 2.4  | Segmented kidney m-reps . . . . .   | 54 |
| 3.1  | Simulated deformed ellipsoids . . . . .   | 64 |
| 3.2  | Two bar graphs of the first 10 eigenvalues in percentage estimated from 5000 simulated warped ellipsoids b-reps . . . . . | 65 |
| 3.3  | Box plots of $\rho^2$ for PCA on warped ellipsoid b-reps . . . . .  | 66 |
| 3.4  | Landmarks of hippocampus and 3 fitted hippocampus m-reps . . . . .  | 67 |
| 3.5  | Two bar graphs of the first 40 eigenvalues in percentage estimated from 290 right hippocampus b-reps. . . . .             | 68 |
| 3.6  | Box plots of $\rho^2$ for PCA on right hippocampus b-reps . . . . .   | 69 |
| 3.7  | A box plot of $D_m^2$ estimated from right hippocampus b-reps . . . . .   | 71 |
| 3.8  | Two box plots of $D_p^2$ for PCA on right hippocampus b-reps . . . . .  | 72 |
| 3.9  | Graphical interpretation of $\rho^2$ . . . . .  | 73 |
| 3.10 | A $(128 \times 128 \times 128)$ bent and tapered binary ellipsoid . . . . .   | 74 |
| 3.11 | Two box plots of $\rho_d^2$ for PGA on m-rep fits to deformed binary ellipsoids . . . . .                                 | 75 |
| 3.12 | Two box plots of $\rho_d^2$ for PGA on m-rep fits to right hippocampus . . . . .  | 76 |

|      |   |     |
|------|---|-----|
| 3.13 | Four simulated ellipsoids with local deformation . . . . .  | 77  |
| 3.14 | Two box plots of $\rho_d^2$ for multi-scale PGA on simulated ellipsoid m-reps with local deformation . . . . .              | 77  |
| 3.15 | Two box plots of $\rho_d^2$ for multi-scale PGA on simulated hippocampus m-reps . . . . .                                   | 79  |
| 4.1  | An m-rep 3-object complex for the bladder, the prostate, and the rectum of a patient in different view. . . . .             | 84  |
| 4.2  | Augmented atoms of the prostate and the bladder. . . . .  | 88  |
| 4.3  | Illustration of the sympathetic change of $\mathbf{R}'_1$ caused by $\mathbf{N}_1$ . . . . .                                | 90  |
| 4.4  | Augmented atoms in the bladder and the prostate . . . . .   | 91  |
| 4.5  | M-rep sampling . . . . .  | 94  |
| 4.6  | Illustration of first modes of variation of two patients . . . . .  | 96  |
| 4.7  | Predicted Prostates of a patient B163 . . . . .   | 100 |
| 4.8  | A bladder and a prostate m-reps on the left and the predicted prostate shape caused by the changes in the bladder . . . . . | 105 |
| 4.9  | A log plot of total variances over 200 iterations . . . . .   | 106 |
| 4.10 | Segmentation of the bladder and the prostate of patient 3106 and 3109 . . . . .   | 107 |
| 4.11 | Segmentation of the prostate and the bladder of patient 3109 . . . . .  | 107 |
| 4.12 | Plots of sorted volume overlaps and average distances of the prostate segmentations . . . . .                               | 108 |
| 5.1  | Simulated quasi-bladder, quasi-prostate, quasi-rectum complexes . . . . .   | 114 |
| 5.2  | Shape variability of simulated multi-object complex . . . . .   | 116 |

|     |   |     |
|-----|---|-----|
| 5.3 | Plots of mean prostate volume overlaps and average surface distance in mm . . .   | 119 |
| 5.4 | Plots of segmented JSC and average surface distance in mm . . . . .   | 122 |
| 6.1 | (a) Eigenvalue plots of PGAMR and PCASE of simulated ellipsoid m-reps. (b)<br>Box plots of $\rho_d^2$ on the simulated ellipsoid m-reps and the derived b-reps. . . . .   | 132 |
| 6.2 | (a) Eigenvalue plots of PGAMR and PCASE of deformed ellipsoid binaries. (b)<br>Two box plots of $D_p^2$ on warped ellipsoid binaries. . . . .   | 133 |
| 6.3 | (a) Eigenvalue plots of PCA on hippocampus b-reps and of PGA on the cor-<br>responding hippocampus m-rep. (b) Two box plots of $\rho^2$ on the hippocampus<br>b-reps and m-reps. . . . .  | 133 |
| 6.4 | Comparison of projections of 50 m-rep fits to deformed binary ellipsoids and of<br>the corresponding 50 b-reps on each shape space . . . . .  | 135 |
| 6.5 | Comparison of projections of 50 m-rep fits to deformed binary ellipsoids on a<br>shape space (a) with m-reps optimized in geodesic distance from the mean, (b)<br>with m-reps further optimized in Euclidean distance starting from the mean<br>m-reps. . . . . | 136 |

# Chapter 1

## Introduction

### 1.1 Motivation

Medical imaging has developed rapidly for the last several decades. Physicians can see inside a human body without cutting the body open via medical imaging. Various imaging modalities have emerged such as magnetic resonance imaging (MRI), ultrasound imaging, x-ray computed tomography (CT) and functional magnetic resonance imaging (fMRI). These imaging techniques produce high-resolution images of different properties of a human body that the naked human eye cannot perceive. The goal of medical image analysis is to make the best use of the information contained in these images. To that end it provides advanced tools that can help to improve diagnosis of disease, to understand the physiology of a human body, or to plan therapy such as radiation treatment.

One of the major research areas in medical image analysis is the geometric representation and statistical shape analysis of biological objects like human organs in medical images. Statistical shape analysis of anatomical objects can be defined as the study of the structure and variability of the shape of organs of humans or animals by means of statistical methods. Interest in the area has been growing because of its potential in many medical applications. For example, an automatic segmentation algorithm that identifies a human organ quickly and accurately can improve planning radiation treatment by minimizing the radiation exposure to normal and functional areas of the organs while maximizing the exposure to areas where tumors grow. A statistical shape description of such organs can be incorporated as a prior in a maximum a posteriori (MAP) framework based segmentation. Such a description provides a

probability distribution of the geometric shape variation of organs observed in training data.

Another application of statistical shape analysis can be found in the study of mental illnesses such as autism and schizophrenia. These illnesses have been reported to affect the shape of some brain structures as the illnesses progress. In the brain of schizophrenics, shapes of some parts of the hippocampus are different from those of non-schizophrenics. With brain structures extracted from the images, statistical shape analysis on their shapes and the inter-relation between structures can aid the understanding of the illnesses by providing a means to compare the shapes of brain structures between the ill and the healthy. It can ultimately serve clinicians as a diagnostic tool for early detection of such illnesses.

Statistical shape analysis hinges on the geometric representation of objects. The representation needs to capture the geometric properties and be consistent across all instances of the object of interest in describing the structure and features of the object. Many efforts have been made to find an appropriate geometric representation of either 2D or 3D objects. Among various models that have been actively investigated, some of the well-known ones are the following: active contour models [Malladi et al., 1995a] and its variant geodesic active contour models [Caselles et al., 1995], diffeomorphisms from atlases [Joshi, 1997, Christensen et al., 1997], level set models [Malladi et al., 1995b, Tsai et al., 2003], point distribution models (PDM) [Cootes et al., 1995] and m-reps [Pizer et al., 2003]. Most shape models of an object are based on a representation of the boundary that lives in a vector space; in that case the shape variations of the object are captured by changes in the boundary. In contrast, the m-rep is derived from a medial representation that focuses on representing the interior of an object. The m-rep describes local shape changes of an object uniquely by thickening, elongation, bending, and twisting.

On the basis of these various shape representations the estimation of the shape statistics of a single object has received most attention. Lately, there has been growing interest in characterizing probabilistically the population of geometric shape representations of a complex of objects instead of concentrating on just one object. It is usually the case that multiple organs are grouped together for a given anatomical region. Consequently, one organ can have a significant effect on the shapes of its abutting organs as it undergoes some changes. In such a situation, the probability distribution of an object's shape itself is not likely to give sufficient

information on the variation of the object unless the variation of its neighboring objects is considered together. A shape probability distribution that properly includes the information of its neighboring objects can play a crucial part in identifying the target object in a new image, especially when the boundaries of objects are so indistinct that even clinicians have difficulty in identifying a target object among its neighboring objects.

This dissertation develops novel alternative techniques to estimate a shape probability distribution of multi-object complexes. A new aspect of this work is to include in a shape probability distribution the relation of neighboring objects to the target object by means of augmentation, that is, augmenting features of its neighboring objects that are highly correlated with the target object to features of the target object. This work consists of two parts. The first one is a method to estimate a conditional shape probability distribution of a target object given its neighboring objects. This method is based on principal component regression (PCR) [Shawe-Taylor and Cristianini, 2004] in reflecting the effect of neighboring objects on the target object. PCR corresponds to performing principal component analysis and regressing in the feature space given by the few first principal directions that capture major variations explained in a given sample. The second part is an approach to estimate shape variation of each object in a grouped multi-object complex where each object's variation is considered to have two components: one being the object's variation independent of its neighbors's effect and the other being the object's interaction with its neighbors. This explicit separation of shape variation into two components is based on the concept of residue introduced in [Lu et al., 2007] for multi-scale shape analysis.

The development of new approaches to statistical multi-object shape analysis on top of existing statistical shape analysis methods of a single object requires criteria and measures to compare different statistical shape descriptions. Considering the goal of the applications using statistical shape models, their ability to describe a new member of the population is an important criterion in evaluating statistical shape descriptions. Muller [2007] proposed a new formal statistical correlation measure called *goodness of prediction* that allows us to judge the predictive power of statistical shape models. This dissertation, extending the measure to a nonlinear space that forms a manifold, provides a thorough study of goodness of prediction measure in both linear and nonlinear spaces.

In the study, goodness of prediction is analyzed as a function of training sample size. This approach of analyzing the correlation measure with respect to sample size enables us not only to deduce the statistical properties of the stability but also to determine an approximate minimum sample size that guarantees a given extent of predictive power. This approach also addresses the issue of sample size that arises in estimating statistics of shape descriptions.

Most statistical shape analysis methods suffer from small sample size. Available training samples are limited due to the cost and time involved in the manual segmentation of medical images. In addition, the dimension of shape feature spaces is in general very high, about several hundreds or thousands, due to the complexity of shape of objects in medical images. Samples of an object's shape in most shape representations are thus typical high dimensional low sample size (HDLSS) data sets. That is, the dimension of the data features is much larger than the sample size. This HDLSS nature of shape data presents a challenge because estimation of shape probability distributions can be very unstable as training samples from the population change. Moreover, many classical data analysis methods in statistics cannot be blindly applied to HDLSS data.

It is hard to avoid the HDLSS problem in statistical shape analysis. In such a situation, a method that can decide the minimum size of training sample that ensures some degree of predictive power holds practical importance, especially to applications of statistical shape analysis. To my knowledge, this is the first work to present an evaluation method of estimated shape probability distributions that has the capability to indicate the predictive power of the estimated distribution at a given sample size.

This dissertation focuses on estimation of shape probability distribution of multiple objects and evaluation of the estimated shape probability distribution. The following subsections describe more in detail the driving problems, challenges, and scopes of two main topics in this dissertation.

### **1.1.1 Statistical Shape Analysis of Multiple Objects**

Here the major thesis is that reflecting interaction between objects in shape probability distributions of multiple objects can help to adequately deal with neighbor relations. In the deformable template model framework, the approaches that have been taken to do shape anal-

ysis of multi-object complexes can be classified in terms of object features as follows: 1) by object complex, 2) object by object, 3) hierarchically from the object complex to the individual objects. The object complex approach has been applied to many current shape models since it is simple and straightforward. This global approach fails to capture the local variation of an object itself and sometimes gives misleading information about the inter-relation among objects because information at global and local scales are mixed in the features.

The object-by-object approach may capture sufficient information about the variability of each object’s shape itself in an estimated shape probability distribution. However, the shape probability distribution of an object itself does not have explicit knowledge about the interaction among objects, so it cannot be used to infer the interaction.

The hierarchical approach addresses the issue of scales in shape models. It computes multi-object shape statistics from the scale of an object complex down to the level of individual objects [Kapur et al., 1998, Vaillant and Davatzikos, 1999]. [Davatzikos et al., 2003] even consider the residue from a larger scale successively. The problem in the hierarchical approach is that it is difficult to come up with a reasonable criterion to define what the variations in object-complex scale and in object-itself scale are, and to distinguish one from the other. The weaknesses in these approaches suggest that we need to somehow reflect the effects from neighboring objects in describing the variation of each object in the object complex.

This dissertation presents two novel approaches to capture the inter-object relation using augmentation in estimating the shape probability distribution of each object in multi-object complexes. The first approach is to use a conditional shape probability distribution of a target object given its neighboring objects. The idea is to incorporate the relation of the target object to its neighboring objects into shape priors so that the shape variation of the target object is constrained not only by its own shape variations but also by its positional, orientational information, and even by shape deformations relative to neighbor objects over a population. A natural way to include this supplementary information in the shape prior is by means of the conditional probability distribution that enables finding a most probable shape of the target object given the configuration of its neighboring objects in relation to the target object. The structural information of a multi-object complex that the conditional shape probability distribution captures can give better guidance for the segmentation of a target object in a

multi-object complex especially when organs are clumped together and touch each other and when the low physical contrast in the image between such structures makes significant parts of the boundary of the target organ ambiguous.

My approach to estimation of the conditional shape distribution is based on the principal components regression (PCR) method [Shawe-Taylor and Cristianini, 2004]. Estimating both the conditional mean and the shape distribution involves computing an inverse of a sample covariance matrix. The limited sample size and high dimensional feature space of shape descriptors provides a challenge in computing an appropriate pseudo-inverse of the covariance matrix. As described before, PCR is equal to regressing in the feature space given by the major principal directions that PCA on the input data produces. Using PCR in reflecting the effect of neighboring objects on the target object in the conditional term of the conditional probability distribution has the following advantages. First, doing PCA on the neighboring shape features reduces the dimension of the feature shape space of neighbor objects. In the reduced features shape space, it is possible to compute a non-singular sample covariance matrix and get a stable estimate of its inverse. Second, it can help to tighten the distribution of the estimated conditional probability distribution since major variation of the neighboring objects is explained in the reduced feature space.

The second approach to shape analysis of multi-object complexes is to divide shape variation of each object in grouped multiple objects into two components: one being the object's inherent shape variation and the other being the object's shape change caused by the interaction with its neighbors, which must sum to the overall shape variation of each object. Thus, one component must be regarded as the residue of the other from the total variation of an object observed.

The use of residue to handle an object's variation can lead on to a more efficient analysis of shape variation. As mentioned before, a difficulty that any method faces to produce a statistical shape model is the HDLSS situation. The first approach uses PCR to solve the problem of inverting a singular covariance matrix caused by the HDLSS nature of shape data. The second approach takes advantage of locality in within-scale object variation by residue and describes each component of total shape variation by a smaller number of parameters so as to ease the serious effect of the HDLSS problem and get better estimate of the shape

probability distribution of each object. Since both approaches estimate the shape probability distribution object-by-object, the HDLSS problem does not get worse as it would in the global approach. [Lu et al., 2007] provides a detailed explanation about the advantages in considering various degrees of locality in the shape variation of an object and in describing an object’s variation via the residue within and between scales.

### 1.1.2 Quality Measures of Statistical Shape Models

Due to the variety of statistical shape models, one can desire some criteria to evaluate a statistical shape model for its effectiveness and some procedures to compare different statistical shape models so as to help choosing an appropriate one that meets one’s need. Many models for single objects are already available to geometrically represent shape and to statistically analyze geometric representations. In addition, new models for multi-object complexes are emerging, as mentioned in the previous section.

The main objective of statistical shape models is to provide a probability distribution on the shape that an object or other geometric entity can take. To that end, there are several properties desired for statistical shape models. First, for geometric representations of an object made from a tuple of spatially sampled primitives there needs to be reasonably good correspondence of the primitives across training cases. Poor correspondence can add noise to training samples that masks the real variation of shape of an object, resulting in an unreliable estimation of the shape probability distribution. Second, geometric representations need to be efficient so that they can describe the shape of the geometric entity of interest with a minimal number of parameters. For example, to measure the geometric efficiency of representation for curves in 2D, Leonard [2007] introduced the concept of  $\varepsilon$ -entropy that is the minimum number of  $\varepsilon$ -balls required to cover the space of a totally bounded metric space and construct an adaptive coding scheme that allows codewords of varying lengths for shape elements in a non-compact space. On the basis of  $\varepsilon$ -entropy and the adaptive encoding scheme, she theoretically determined conditions in which the medial axis is more efficient than the boundary curves in representing an object. She shows in her experiments that the medial axis holds a tenable position as a shape model in 2D: for all but three out of the 2,322 2D-shapes she analyzed, the medial representation is more efficient. The efficiency of the geometric representations of

shape can help to alleviate the HDLSS problem as well as to avoid the over-fitting problem. Third, an estimated shape probability distribution needs to be tight. Fourth, it needs to be unimodal since most statistical data analysis methods employed in shape analysis are based on the assumption of Gaussian distribution of the data. Fifth, a statistical shape model must be able to represent only real instances in the population of the geometric entity. Sixth, it must be able to describe a member of the population unseen in a training sample.

Among the few studies in which statistical shape models were evaluated, a key study done by Styner et al. [2003] defines three criteria that can assess some of these properties and then compares correspondence of shape models on the basis of the three criteria. The three criteria are defined as follows: compactness as the ability to use a minimal set of parameters, generalization as the ability to describe instances outside of the training set, and specificity as the ability to represent only valid instances of the object in its population. Generalization ability is assessed by doing leave-one-out reconstruction and computing approximation errors of unseen models averaged over the complete set of trials. These measures are defined as functions of the number of shape parameters. Generalization ability and specificity are defined in the *ambient space*, where the models lie physically. In regard to these criteria, Styner et al. [2003] examined and compared four methods: a manually initialized subdivision surface method for direct correspondence and three automatic methods - spherical harmonics, minimum description length, and minimum covariance determinant - for model-implied correspondence across instances of an object.

To compute the three criteria, [Styner et al., 2003] proposed to use an approximation error based on a small set of anatomical landmarks that a human expert selects manually on each object. The approximation error is defined as the mean absolute distance (MAD) between the manual landmarks and points corresponding to the same landmarks of the four shape models.

The approach of Styner et al. [2003] in the evaluation of the correspondence methods is grounded in an important observation: statistical shape models of good correspondence are highly likely to have good compactness, good generalization, and good specificity. In fact, these are qualities that a statistical shape model obtains as results of some optimizations to establish correspondence. They do not directly indicate the quality of correspondence of statistical shape models.

While these measures offer legitimate criteria to evaluate correspondence methods of different statistical shape models, they are short on the predictive power of the statistical shape models: the ability to describe unseen members of the population and to describe their frequency of occurrence. This ability of statistical shape models is critical since most applications of statistical shape models heavily rely on their predictive power. One example is model-based segmentation in the maximum a posteriori (MAP) framework, which uses a prior distribution of shape of an geometric entity to extract the geometric entity from a new image. Another example is classification of an object on the basis of its shape, using trained shape prior distributions.

Muller [2007] proposed a novel statistical correlation measure that evaluates the predictive power of statistical shape models derived using principal component analysis (PCA). Muller first shows that PCA can be recast as a multivariate regression model by treating the observed variables as responses and principal directions as predictors. Then, goodness of prediction is derived from goodness of fit, a standard statistical measure for second moment accuracy. As for possible goodness of prediction measures, he proves that canonical correlations and related measures of association degenerate to constants and that the “univariate approach to repeated measures” test (average squared correlation, generalized variance explained) provides a simple and useful measure of association. He also suggests another measure - squared multiple correlation - to provide more detailed information. Among the several measures he proposes, in this work we adopt the average squared correlation measure as a goodness of prediction measure to evaluate the predictive power of both linear and nonlinear statistical shape models. The detailed development of this measure is given in chapter 3.

This correlation measure has a clear statistical interpretation in terms of the predictability of statistical shape models. This feature facilitates evaluation and comparison of different approaches to estimate a shape description or different statistical shape models. Furthermore, the average squared correlation is a simple direct measure defined in the shape feature space and is quick and easy to compute. In contrast, the generalization measure in [Styner et al., 2003] is an indirect measure defined in the ambient space where the models lie physically, and it takes a long time to compute.

The main goal of this work is to form a systematic procedure for measuring and analyzing

the predictive power of statistical shape models as a function of training sample size using goodness of prediction. This correlation measure is first extended to nonlinear manifolds as some shape representations like the m-rep do not belong to a vector space. This work then demonstrates the procedure with two different shape representations: the m-rep and the PDM.

The PDM is a well-known and popular shape model that represents an object by concatenated points sampled on the surface of an object. In training its shape probability distribution, PCA is applied to reduce the dimension of the representation as well as to find modes of variation that are significant in the training sample. The m-rep on the other hand takes the object interior into account, and the boundary of the object is derived indirectly from the representation. Unlike the PDM, the m-rep does not lie in Euclidean space because of the radius and the rotational components of the medial atoms. Fletcher et al. [2004] found that an m-rep can be understood as a point in so called “symmetric space” where the properties of vector space do not hold. They developed principal geodesic analysis (PGA) that generalized PCA in the abstract, curved space known as manifolds. Due to their distinctive contrast in their perspectives to describe an object, the PDM with PCA and the m-rep with PGA are chosen for this work.

The extended correlation measure can be applied to the global and the hierarchical approach of multi-object shape analysis described in the previous section. For the hierarchical approach taken in the m-rep framework [Liu et al., 2008], goodness of prediction is proven to be a useful measure for comparing a fine-scale shape prior with an object-scale one. However, as will be discussed in a later chapter, goodness of prediction cannot be applied to the object-by-object approaches for multi-object shape analysis proposed in this thesis because of the inter-object relation reflective in the shape probability distribution. As a result, the estimated shape probability distribution that reflects the interaction among objects is compared with one that does not reflect the inter-object relation by 1) viewing the deformation of an object in each estimated shape probability distribution or by 2) comparing the segmentation results with the indirect measures such as volume overlap and surface distance [Crum et al., 2006] between the segmentation results and their truth.

## 1.2 Thesis

Thesis: (1) *Reflecting interaction among objects in statistical shape models for multi-object complexes via augmentation yields a shape probability distribution that can capture the configuration of objects and shape variability caused by neighboring objects.*

(2) *A systematic procedure to evaluate the predictive power of statistical shape models as a function of training sample size using the correlation measure provides a means to determine an approximate minimum size of sample that ensures a certain degree of predictive power and stability.*

## 1.3 Claims

The contributions of this dissertation are as follows:

1. It is shown that in multi-object statistical shape analysis augmentation of highly correlated geometric features of neighboring objects to the target object can be used to reflect interaction among objects in shape probability distribution of the target object.
2. A new method to estimate the shape probability distribution of an object conditioned on its geometric primitives of the neighboring objects has been developed. The method relies on principal component regression to have stable and reliable estimates of the conditional mean and the conditional covariance matrix.
3. The conditional shape probability distribution can usefully provide a shape prior for maximum posterior segmentation with the conditional mean as the initial template model. This conditional shape prior was applied to segment anatomical objects in the male pelvis.
4. A method to decompose shape variation of each object in multi-object complexes into two components has been developed on the basis of the concept of the residue.
5. It is shown that the probability distributions on the decomposed shape variations described in chapter 4 can be incorporated as a shape prior for maximum posterior segmen-

tation. These shape priors of the decomposed shape variations were applied to segment anatomical objects in the male pelvis.

6. A new correlation measure called goodness of prediction initially proposed by Muller [2007] in linear space has been extended to shape representations that live in a nonlinear manifold.
7. As a tool to evaluate the predictive power of statistical shape models, an iterative procedure that estimates the goodness of prediction in a given set of samples has been developed. This procedure provides a means to analyze the correlation measure for a given statistical shape model as a function of training sample size.

## 1.4 Overview of Chapters

Chapter 2 presents the background topics in medical image analysis that are relevant in this dissertation. Included topics are various probabilistic deformable models, medial representation, especially m-rep models and its statistical shape analysis. Chapter 3 explains in detail the statistical background for the goodness of prediction measure, its derivation, and the application to statistical shape models. Chapter 4 presents the new method to estimate the conditional shape probability distribution of multi-object data and shows the result from using the new shape probability distribution as a geometric prior for segmentation. Chapter 5 introduces the novel approach to decompose an object's shape variations into two components. Chapter 6 concludes with a discussion of what remains to be done to improve the approaches reported in this dissertation.

## Chapter 2

# Background

This chapter presents the background material from statistics and image analysis that is relevant to this dissertation. Section 2.1 gives an overview of basic concepts of probabilities and statistical methods that are central to this dissertation. Section 2.2 explains the statistical theory of shape, which lays the background for the next section. Section 2.3 explains probabilistic deformable models and a few geometric representations of objects. Then, section 2.4 describes in detail the medial object representation, a deformable model used in this dissertation. Section 2.5 describes transforming human expert segmentations into models that will be used to learn the probability distribution on the anatomic shape of interest. The last section: 2.6 explains the segmentation of m-reps.

## 2.1 Statistical Background

### 2.1.1 Probability Distributions

This section starts from the basic definitions of probability, random variables, probability distributions and elementary statistics, and it ends with the well-known multivariate normal distribution that appears repeatedly in this dissertation. Most materials in this section are from [Hogg and Craig, 1995, Muirhead, 1982].

#### 2.1.1.1 Definitions

Let  $\mathcal{C}$  be the sample space, that is, the set of every possible outcome of a random experiment. The *probability*  $P(C)$ ,  $C \subset \mathcal{C}$  is a function of the outcome of the random experiment defined

on subsets of the space  $\mathcal{C}$  such that [Hogg and Craig, 1995]

- (a)  $P(C) \geq 0$ ,
- (b)  $P(C_1 \cup C_2 \cup \dots) = P(C_1) + P(C_2) + \dots$ , where the sets  $C_i$ ,  $i = 1, 2, 3, \dots$ , are such that no two have a point in common (that is, where  $C_i \cap C_j = \emptyset$ ,  $i \neq j$ ), and
- (c)  $P(\mathcal{C}) = 1$ .

The probability function  $P(C)$  tells us how the probability is distributed over various subsets  $C$  of the sample space  $\mathcal{C}$ .

In some experiments we are concerned only with outcomes that are elements of a subset  $C_1$  of the sample space  $\mathcal{C}$ . The sample space then becomes in effect the subset  $C_1$ . For  $C_2$  another subset of  $\mathcal{C}$ , the *conditional probability*  $P(C_2|C_1)$  of the event  $C_2$  given the event  $C_1$  is defined as

$$P(C_2|C_1) = \frac{P(C_1 \cap C_2)}{P(C_1)}, \quad (2.1)$$

provided that  $P(C_1) > 0$ . That is,  $P(C_2|C_1)$  considers only those outcomes in  $C_2$  that are elements of  $C_1$ .  $P(C_2|C_1)$  also has the following properties:

1.  $P(C_2|C_1) \geq 0$ ,
2.  $P(C_2 \cup C_3 \cup \dots | C_1) = P(C_2|C_1) + P(C_3|C_1) + \dots$ , when  $C_2, C_3, \dots$  are mutually disjoint sets, and
3.  $P(C_1|C_1) = 1$ .

These are the precise conditions that a probability function must satisfy as described above. Occasionally, the occurrence of an event  $C_1$  does not change the probability of an event  $C_2$ , i.e.,

$$P(C_2|C_1) = P(C_2).$$

In this case, the events  $C_1$  and  $C_2$  are said to be *independent*. From the definition of conditional probability, we can see that two independent events satisfy

$$P(C_2 \cap C_1) = P(C_1)P(C_2|C_1) = P(C_1)P(C_2),$$

which can be used as an alternative definition of independence.

The famous *Bayes' rule* is also derived from the definition of conditional probability. The conditional probability of the event  $C_1$  given the event  $C_2$  is

$$P(C_1|C_2) = \frac{P(C_1 \cap C_2)}{P(C_2)}. \quad (2.2)$$

Since  $P(C_1 \cap C_2)$  is the common term both in (2.1) and (2.2), (2.1) and (2.2) can be rearranged and combined to obtain Bayes' rule:

$$P(C_1|C_2) = \frac{P(C_1 \cap C_2)}{P(C_2)} = \frac{P(C_2|C_1)P(C_1)}{P(C_2)}, \quad (2.3)$$

which relates two conditional and *marginal* probabilities of the events  $C_1$  and  $C_2$ . The definition of a marginal probability will be given later in section 2.1.1.5.

A *random variable*  $X$  is a function that assigns to each element  $c \in \mathcal{C}$  one and only one real number  $X(c) = x$ . The space of  $X$  is the set of real numbers  $\mathcal{A} = \{x \in \mathbb{R} : x = X(c), c \in \mathcal{C}\}$ . Thus, a random variable  $X$  is a function that carries the probability from a sample space  $\mathcal{C}$  to a space  $\mathcal{A}$  of real numbers. The probability  $P(A)$  now defined on  $\mathcal{A}$ ,  $A \subset \mathcal{A}$ .  $\mathcal{A}$  is also referred to as the sample space.

If  $\mathcal{A}$  contains a finite number of points or the points of  $\mathcal{A}$  can be put into a one-to-one correspondence with the positive integers,  $X$  is called a *random variable of the discrete type*. If the random sample space  $\mathcal{A}$  consists of an interval or a union of intervals in  $\mathbb{R}$ ,  $X$  is said to be a *random variable of the continuous type*. This work deals with only random variables of the continuous type, so topics about discrete random variables will be not be discussed here. Random variables will refer to continuous random variables from now on.

An example of a continuous random variable is found in the shape studies of this dissertation. The example is one component in the feature vector of m-reps, the main probabilistic deformable model used in this dissertation. This component describes the local magnification of shape of an object and is called the *spoke length*, denoted by  $r$ , which is non-negative. Its log,  $X = \ln(r)$ , is considered as a continuous random variable in  $\mathbb{R}$ .

The probability  $P(A)$ ,  $A \subset \mathcal{A}$  can be expressed in terms of a nonnegative function  $f(x)$  such that

$$\int_{\mathcal{A}} f(x)dx = 1,$$

by

$$P(A) = \int_A f(x)dx.$$

$f(x)$  is called the *probability density function* (p.d.f.) of  $X$ . When  $A$  is an unbounded set from  $-\infty$  to a real number  $x$ , the *distribution function* (or sometimes, *cumulative distribution function*)  $F(x) = P(X \leq x)$  is defined as

$$F(x) = \int_{w \leq x} f(w)dw.$$

### 2.1.1.2 Expectations of a Random Variable

Let  $X$  be a random variable having a p.d.f.  $f(x)$ . A central value of the probability is given by the *expectation of the random variable  $X$* :

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx.$$

For example, the expectation of the random variable  $X = \ln(r)$  gives the central value of  $\ln(r)$  when its p.d.f. is known. For a function of  $X$ ,  $Y = u(X)$ , the expectation of  $u(X)$  is  $E(u(X)) = \int_{-\infty}^{\infty} u(x)f(x)dx$ . When  $u(X) = X$ ,  $E(X)$  is the arithmetic mean of the value of  $X$ , and is called the *mean value* of  $X$ . Another special expectation is obtained by taking  $u(X) = (X - E(X))^2$ , that is,

$$E(X) = \int_x (x - E(X))^2 f(x)$$

This expectation is called the *variance* of  $X$ , and is the mean value of the square of the deviation of  $X$  from its mean value  $E(X)$ .

### 2.1.1.3 Special Distributions

Two important distributions are introduced in this section. They are frequently used in statistics and mentioned in this dissertation.

#### The Normal Distribution

A random variable  $X$  that has a p.d.f. of the form

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty \quad (2.4)$$

is said to have a *normal* or *Gaussian distribution* with mean  $\mu$  and variance  $\sigma^2$ . The normal p.d.f. occurs so often in many parts of statistics that it is denoted  $N(\mu, \sigma^2)$ . A useful property of the normal p.d.f is that given a random variable  $X$  that is  $N(\mu, \sigma^2)$ , the random variable  $W = (X - \mu)/\sigma$  is  $N(0, 1)$ . This fact helps with simplifying the calculations of probabilities concerning normally distributed random variables.

The log of the scale component  $\ln(r)$  in m-reps is assumed to follow the normal distribution, i.e.,  $\ln(r) \sim N(0, \sigma^2)$  of mean zero and a standard deviation  $\sigma$ , when treated as a random variable in shape analysis of m-reps.

#### The Gamma and Chi-Square Distributions

A random variable  $X$  that has a p.d.f

$$\begin{aligned} f(x) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} \exp(-x/\beta), & 0 < x < \infty \\ &= 0 & \text{elsewhere} \end{aligned}$$

is said to have a *gamma distribution* with parameters  $\alpha$  and  $\beta$ . The *gamma function* of  $\alpha$  is defined as  $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} \exp(-y) dy$ , which exists for  $\alpha > 0$ . The value of the integral is a positive number. A *chi-square distribution* is a special case of the gamma distribution in which  $\alpha = \gamma/2$ , where  $\gamma$  is a positive integer, and  $\beta = 2$ . The parameter  $\gamma$  is called the number of degrees of freedom of the chi-square distribution.  $X$  is  $\chi^2(\gamma)$ , i.e.,  $X \sim \chi^2(\gamma)$  means that the random variable  $X$  follows a chi-square distribution with  $\gamma$  degrees of freedom.

The chi-square distribution plays an important role in statistics partly because of its close tie with the normal distribution. The random variable  $V = (X - \mu)^2/\sigma^2$  is  $\chi^2(1)$  if  $X$  is

$N(\mu, \sigma^2)$ . So the square of the log of the spoke length components in m-reps  $(\ln(r)/\sigma)^2$  is also  $\chi^2(1)$ .

#### 2.1.1.4 Statistics: Sample Mean and Variance

A *statistic* is a function of one or more random variables that does not depend on any unknown parameters although the distribution of the statistic may depend upon the unknown parameters. Let the random variables  $X_i$ ,  $i = 1, \dots, n$ , be independent, each having the same p.d.f  $f(x)$ . For example, if  $X_1$  is  $N(\mu, \sigma^2)$ , then a random variable  $Y = (X_1 - \mu)/\sigma$  is  $N(0, 1)$  and is a function of  $X_1$  that depends on the two parameters  $\mu$  and  $\sigma$ , so it is not a statistic. In contrast,  $Y = \sum_{i=1}^n X_i$  does not depend on any unknown parameters and thus is a statistic.

A statistic can be used to infer information about the unknown parameters. Let a random variable  $X$  be defined on a sample space  $\mathcal{A}$ . In many situations, the distribution of  $X$  is known except for the value of the parameters of the distribution. To obtain information about the unknown parameters, the random experiment is repeated  $n$  independent times under identical conditions. If the random variable  $X_i$  is a function of the  $i$ -th outcome,  $i = 1, \dots, n$ , then  $X_1, X_2, \dots, X_n$  are called the *observations* of a random sample from the distribution under consideration. If we have a statistic  $Y = u(X_1, \dots, X_n)$  whose p.d.f is known and the p.d.f reaches its maximum when  $Y$  has a value close to the unknown parameter, then the statistic  $Y$  can be used to draw information about the unknown parameters.

The most common statistics are the *mean* and the *variance of the random sample*. Let  $X_1, X_2, \dots, X_n$  denote a random sample of size  $n$  from a given distribution. The mean of the random sample is defined as

$$\hat{\mu} = \frac{X_1 + \dots + X_n}{n} = \sum_{i=1}^n \frac{X_i}{n}.$$

The variance of the random sample is defined as

$$\hat{\sigma}^2 = \sum_{i=1}^n \frac{(X_i - \hat{\mu})^2}{n-1} = \sum_{i=1}^n \frac{X_i^2}{n-1} - \frac{n}{n-1} \hat{\mu}^2.$$

### 2.1.1.5 Multivariate Distributions

Describing shapes of objects in 2D or 3D usually requires multiple features, such as the ordered tuple of points on the surface, a set of control points of spline functions that represent the surface, or a set of curvatures. Statistical shape analysis considers these geometric features as random variables and focuses on understanding their relation, variability and geometrical properties. The analysis of more than one variable requires multivariate statistics.

Let  $X_i, i = 1, \dots, p$  be a set of random variables. The space of these random variables is the set of ordered  $p$ -tuples  $\mathcal{A} = \{(x_1, \dots, x_p) : x_1 = X_1(c), \dots, x_p = X_p(c), c \in \mathcal{C}\}$  where  $\mathcal{C}$  is the sample space. For  $A \subset \mathcal{A}$  the probability function of these  $p$ -variate random variables  $P(A)$  can be expressed as

$$P(A) = \int \cdots \int_A f(x_1, \dots, x_p) dx_1 \dots dx_p.$$

In parallel to the definition of a p.d.f for a single random variable,  $f$  is a p.d.f if  $f$  is defined and is nonnegative for all values in  $\mathcal{A}$ , and if its integral over all real values in  $\mathcal{A}$  is 1. The distribution function of  $X_i, i = 1, \dots, p$  is

$$F(x_1, \dots, x_p) = P(X_1 \leq x_1, \dots, X_p \leq x_p)$$

If  $X_1, \dots, X_p$  are continuous type random variables,  $f_1(x_1)$  is called the marginal p.d.f of  $X_1$  when

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_p) dx_2 \dots dx_p.$$

The marginal probability density functions  $f_2(x_2), \dots, f_p(x_p)$  for  $X_2, \dots, X_p$  are defined similarly as  $(p - 1)$ -fold integrals. The marginal p.d.f can be generalized to a joint marginal p.d.f for any  $k < p$  group of random variables.

To simplify the notation, let  $\mathbf{x}$  be a  $p \times 1$  random vector  $\mathbf{x} = (X_1, \dots, X_p)'$ . The mean of

the vector  $\mathbf{x}$  is defined to be the vector of expectations:

$$E(\mathbf{x}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{pmatrix}.$$

When dealing with a  $p$ -variate random vector  $\mathbf{x} \in \mathbb{R}_p$  with mean  $E(\mathbf{x}) = \boldsymbol{\mu}$ , the variance is defined by a covariance matrix. The covariance matrix of  $\mathbf{x}$  is defined to be the  $p \times p$  matrix

$$\Sigma = Cov(\mathbf{x}) = E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'].$$

The  $i, j$ th element of  $\Sigma$  is

$$\sigma_{ij} = E[(X_i - \mu_i)(X_j - \mu_j)],$$

and the  $i$ th diagonal element is

$$\sigma_{ii} = E[(X_i - \mu_i)^2].$$

Clearly,  $\Sigma$  is symmetric, i.e.,  $\sigma_{ij} = \sigma_{ji}$ . It is also a *non-negative definite* matrix. A  $p \times p$  symmetric matrix  $A$  is called non-negative definite if

$$\alpha' \Sigma \alpha \geq 0 \quad \text{for all } \alpha \in \mathbb{R}_p$$

and *positive definite* if

$$\alpha' \Sigma \alpha > 0 \quad \text{for all } \alpha \in \mathbb{R}_p, \quad \alpha \neq \mathbf{0}.$$

### 2.1.1.6 The Multivariate Normal Distribution

The  $p \times 1$  random vector  $\mathbf{x}$  is said to have a  $p$ -variate *normal* or *Gaussian distribution* if, for every  $\alpha \in \mathbb{R}_p$ , the distribution of  $\alpha' \mathbf{x}$  is univariate normal (2.4) [Muirhead, 1982]. From this definition it can be shown that  $\mathbf{x}$  has the following density function [Arnold, 1981]:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right),$$

where  $\boldsymbol{\mu}$  and  $\Sigma$  are the mean and the covariance matrix of  $\mathbf{x}$  respectively.  $\det(\Sigma)$  denotes the determinant of the covariance matrix  $\Sigma$ . The  $p$ -variate normal distribution of  $\mathbf{x}$  is denoted by  $N_p(\boldsymbol{\mu}, \Sigma)$ .

A multivariate normal random vector  $\mathbf{x}$  has important properties as to the marginal distributions and the relation between its subvectors. These properties play an important part in the method developed in chapter 5 for multi-object shape analysis. If  $\mathbf{x}$  is  $N_p(\boldsymbol{\mu}, \Sigma)$ , the marginal distribution of any subvector of  $k$  elements of  $\mathbf{x}$  is  $k$ -variate normal ( $k \leq p$ ). Moreover,  $\mathbf{x}$ ,  $\boldsymbol{\mu}$ , and  $\Sigma$  can be partitioned as

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

where  $\mathbf{x}_1$  and  $\boldsymbol{\mu}_1$  are  $k \times 1$  and  $\Sigma_{11}$  is  $k \times k$ . The subvectors  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are independent if and only if  $\Sigma_{12} = 0$ . In order to determine the independence of two subvectors of a normally distributed vector, it suffices to check that the covariance matrix of the two subvectors is zero.

Let  $\Sigma_{11.2}$  be  $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^- \Sigma_{21}$ , where  $\Sigma_{22}^-$  indicates a *generalized inverse* of  $\Sigma_{22}$ , i.e.,  $\Sigma_{22}\Sigma_{22}^-\Sigma_{22} = \Sigma_{22}$ . Then

- (a)  $\mathbf{x}_1 - \Sigma_{12}\Sigma_{22}^-\mathbf{x}_2$  is  $N_k(\boldsymbol{\mu}_1 - \Sigma_{12}\Sigma_{22}^-\boldsymbol{\mu}_2, \Sigma_{11.2})$  and is independent of  $\mathbf{x}_2$ , and
- (b) the conditional distribution of  $\mathbf{x}_2$  given  $\mathbf{x}_1$  is  $N_k(\boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^-(\mathbf{x}_1 - \boldsymbol{\mu}_1), \Sigma_{11.2})$ .

Item **b** implies that the covariance matrix  $\Sigma_{11.2}$  of the conditional distribution of  $\mathbf{x}_1$  given  $\mathbf{x}_2$  does not depend on  $\mathbf{x}_2$ . The independence property of  $\Sigma_{11.2}$  on  $\mathbf{x}_2$  comes in useful later when the conditional distribution is used to estimate a multi-object shape probability distribution. Also, the mean of the conditional distribution of  $\mathbf{x}_1$  given  $\mathbf{x}_2$  is

$$E(\mathbf{x}_1|\mathbf{x}_2) = \boldsymbol{\mu}_1 + \Sigma_{12}\Sigma_{22}^-(\mathbf{x}_2 - \boldsymbol{\mu}_2).$$

$E(\mathbf{x}_1|\mathbf{x}_2)$  is called the *regression function* of  $\mathbf{x}_1$  on  $\mathbf{x}_2$  with the *regression coefficients*  $\Sigma_{12}\Sigma_{22}^-$ . It is a linear regression function since it depends linearly on the variables  $\mathbf{x}_2$  [Muirhead, 1982].

Another important property of a multivariate normal random vector  $\mathbf{x}$  is that the diagonalization of the covariance matrix  $\Sigma$  by eigendecomposition produces a new set of statistically

independent coordinates. Let  $\mathbf{\Upsilon}\mathbf{\Lambda}\mathbf{\Upsilon}'$  be the eigendecomposition of  $\Sigma$ , where  $\mathbf{\Lambda}$  is a diagonal matrix and  $\mathbf{\Upsilon}$  is an orthonormal matrix whose columns are the new coordinate axes. In fact,  $\mathbf{\Upsilon}$  is a rotation matrix in  $SO(p)$ . The covariance of the rotated vector of  $\mathbf{x}$ , namely,  $\mathbf{y} = \mathbf{\Upsilon}\mathbf{x}$ , is

$$\begin{aligned}\Sigma_{\mathbf{y}} &= E[(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})(\mathbf{y} - \boldsymbol{\mu}_{\mathbf{y}})'] \\ &= E[(\mathbf{\Upsilon}\mathbf{x} - \mathbf{\Upsilon}\boldsymbol{\mu})(\mathbf{\Upsilon}\mathbf{x} - \mathbf{\Upsilon}\boldsymbol{\mu})'] \\ &= E[\mathbf{\Upsilon}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})'\mathbf{\Upsilon}'] \\ &= \mathbf{\Upsilon}\Sigma\mathbf{\Upsilon}' = \mathbf{\Lambda},\end{aligned}$$

which shows that the random variables in the random vector  $\mathbf{y}$  are independent.

### 2.1.2 Principal Component Analysis

*Principal component analysis* (PCA) is a technique developed by Hotelling [1933] to reduce the dimension of feature space without losing too much of the information about the variables contained in the covariance matrix. In most practical situations it is useful to describe a simple model for the structure of the corresponding covariance or correlation matrix for observations made on a large number of correlated random variables. In PCA, the original coordinate axes (variables) are rotated to give a new coordinate system having some optimal variance properties. The first principal component (variable) is the linear combination of the original variables with maximum variance; the second principal component is the linear combination of the variables having maximum variance uncorrelated with the first principal component; and so on. Thus, PCA attempts to find a new set of variables that can express larger variance with fewer variables.

There is a family of probability distributions called *elliptical* distributions. Elliptical distributions are unimodal and symmetric around the mean. It follows that the contours of equal probability density form ellipsoids. The *principal components* of such a distribution describe a rotation of the coordinate axes to the *principal axes* or *principal directions (vectors)* of the ellipsoid. [Muirhead, 1982]

PCA can be understood as a way to estimate the multivariate Gaussian distribution, which

forms an elliptical distribution. If random variables of feature space are considered to follow a multivariate normal distributions, the principal directions of observations taken on the random variables and their corresponding variances give a way to find these contours of equal probability along each principal direction, which can be interpreted as fitting a Gaussian distribution.

PCA can also be considered as a special case of a *factor analysis* (FA) in the context of *linear models*. Factor analysis is a method to uncover the latent structure (dimensions) of a set of observed variables and thus to reduce the attribute space from a larger number of variables to a smaller number of so called *factors*. [Tucker and R., 1993] The FA model is often preferred over the PCA model because the PCA model underfits when the FA model holds. The underfitting always leads to bias and an invalid model, no matter how large the sample size is [Muller and Stewart, 2006]. PCA is also sensitive to the choice of a particular coordinate system or units of measurements of the variables because the method is not invariant under linear transformations of the variables so a linear transformation changes the eigenstructure of the covariance matrix [Muirhead, 1982]. In spite of these drawbacks of PCA, in medical image analysis PCA has proven its usefulness for describing the variability of linear shape data as shown in the early work of Cootes et al. [1995], Bookstein [1999].

The usefulness of PCA in shape analysis is twofold: (1) producing linear combinations of shape feature variables that give an efficient reparametrization of the original shape feature variables by the variability observed on the training data and (2) reducing the dimension of the original feature space into a subspace, a statistical shape space that is described by the estimated principal directions of major variances.

The objective of PCA is to find a subspace centered at the mean that best explains the variability observed in the data. PCA can be formulated in two ways (1) an approach to minimize the sum-of-squares of the residuals to the data and (2) an approach to maximize the total variance of the projected data. Jolliffe [1986] shows that both approaches produce the same subspace in linear space. A concise summary of both approaches is given in [Fletcher, 2004]. Here only the second approach is described.

Let  $\|\cdot\|$  be the usual Euclidean norm and  $\langle \cdot, \cdot \rangle$  be the inner product of  $p$ -dimensional real space  $\mathbb{R}_p$ . Then  $\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle^2$  for an  $\mathbf{v} \in \mathbb{R}_p$ . Let  $\mathbf{y}_1, \dots, \mathbf{y}_N \in \mathbb{R}_p$  be  $N$  observed multivariate

data with their mean  $\boldsymbol{\mu}$ . The total variance is defined as

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N \|\mathbf{y}_i - \boldsymbol{\mu}\|^2 = \frac{1}{N} \sum_{i=1}^N \langle \mathbf{y}_i - \boldsymbol{\mu}, \mathbf{y}_i - \boldsymbol{\mu} \rangle^2. \quad (2.5)$$

Principal directions, or *modes of variation*  $\mathbf{v}_j$ , spanning the  $k$ -dimensional linear subspace  $V_k$ ,  $j = 1 \dots k$  and  $k \leq p$ , are given by

$$\mathbf{v}_j = \operatorname{argmax}_{\|\mathbf{v}\|^2=1} \sum_{i=1}^N \langle \mathbf{y}_i^j, \mathbf{v} \rangle^2, \quad (2.6)$$

where  $\mathbf{y}_i^j$ ,  $i = 1, \dots, N$ , satisfies the recurrence relation:

$$\begin{aligned} \mathbf{y}_i^1 &= \mathbf{y}_i - \boldsymbol{\mu} \\ \mathbf{y}_i^j &= \mathbf{y}_i^{j-1} - \langle \mathbf{y}_i^{j-1}, \mathbf{v}_{j-1} \rangle \mathbf{v}_{j-1}. \end{aligned}$$

That is,  $\mathbf{y}_i^j$  are the points of  $\mathbf{y}_i - \boldsymbol{\mu}$  projected on to the subspace orthogonal to  $V_{j-1}$ . These principal directions  $\mathbf{v}_i$  define a subspace that maximizes the total variance of the projected points. Principal components are the projection coefficients of the multivariate data  $\mathbf{y} \in \mathbb{R}_p$  onto the principal directions, i.e.,  $\langle \mathbf{y} - \boldsymbol{\mu}, \mathbf{v}_j \rangle$ .

### PCA in Matrix Notation

PCA can also be understood as a way to decompose the sample covariance matrix  $\mathbf{S}$  of the  $\mathbf{y}_i$ 's. Let  $\mathbf{Y}$  be a  $N \times p$  data matrix where each row is  $\mathbf{y}_i'$ , that is,

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}'_1 \\ \vdots \\ \mathbf{y}'_N \end{pmatrix}.$$

Let  $\hat{\boldsymbol{\mu}}$  be the  $p \times 1$  sample mean vector of  $\mathbf{Y}$ .  $\hat{\boldsymbol{\mu}}$  can be written in matrix notation as

$$\hat{\boldsymbol{\mu}} = \frac{1}{N} \sum_{i=1}^N \mathbf{y}_i = \frac{1}{N} \mathbf{Y}' \mathbf{1}' \quad (2.7)$$

where  $\mathbf{1}$  is the  $N \times 1$  vector of 1's. The  $p \times p$  sample covariance matrix  $\mathbf{S}$  can be written as

$$\begin{aligned}\mathbf{S} &= \frac{1}{N-1} \sum_{i=1}^N (\mathbf{y}_i - \widehat{\boldsymbol{\mu}}) (\mathbf{y}_i - \widehat{\boldsymbol{\mu}})' \\ &= \frac{N}{N-1} \left( \frac{1}{N} \mathbf{Y}'\mathbf{Y} - \widehat{\boldsymbol{\mu}}\widehat{\boldsymbol{\mu}}' \right).\end{aligned}\tag{2.8}$$

Recall that the covariance matrix is symmetric ( $\mathbf{S} = \mathbf{S}'$ ) and non-negative definite, which guarantees that its spectral decomposition or eigendecomposition produces only real and non-negative eigenvalues  $\lambda_1, \dots, \lambda_p$ , with  $\lambda_1 \geq \lambda_2 \dots \geq \lambda_p$ , and corresponding ordered eigenvectors  $\mathbf{v}_1, \dots, \mathbf{v}_p$ . This eigendecomposition of  $\mathbf{S}$  can be written as

$$\boldsymbol{\Upsilon}'\mathbf{S}\boldsymbol{\Upsilon} = \mathbf{D}(\boldsymbol{\lambda}) ,$$

where

$$\mathbf{D}(\boldsymbol{\lambda}) = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_p \end{pmatrix} ,$$

the  $p \times p$  diagonal matrix of the eigenvalues  $\{\lambda_i\}$ ,  $i = 1 \dots p$ , and where

$$\boldsymbol{\Upsilon} = \begin{pmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_p \end{pmatrix} ,$$

the column matrix of eigenvectors  $\mathbf{v}_1, \dots, \mathbf{v}_p$ . These eigenvectors are equivalent to the principal directions defined in (2.6) defining a new set of coordinate axes.  $\lambda_1, \dots, \lambda_p$  are estimates of the variances of the population principal components of the population covariance matrix  $\Sigma$ .

### Dimension Reduction and Approximation of Data

The estimated total variability or variance in  $\mathbf{Y}$  is defined as

$$\text{tr}(\mathbf{S}) = \text{tr}(\mathbf{D}(\boldsymbol{\lambda})) = \sum_{i=1}^p \lambda_i ,$$

the trace of the sample covariance matrix  $\mathbf{S}$ . This definition of total variance is equivalent to the previous definition in (2.5).

The data can be approximated with some number of principal directions, say  $p_a$ ,  $p_a < p$ ,

that, for example, explains a maximal percentage of the total variance. Projecting the data into the subspace  $V_{p_a}$  spanned by the first  $p_a$  principal directions allows us to approximate the original  $\mathbf{y}$  by the linear combination of the first  $p_a$  principal directions weighted by their corresponding principal components. The projection  $\mathbf{y}_a$  is then

$$\mathbf{y}_a = \hat{\boldsymbol{\mu}} + \sum_{i=1}^{p_a} \alpha_i \mathbf{v}_i, \quad (2.9)$$

where  $\alpha_i = \langle \mathbf{y} - \hat{\boldsymbol{\mu}}, \mathbf{v}_i \rangle$  for  $i = 1 \dots p$ . These  $\alpha_i$ 's are the principal components.

As an aside,  $\det(\mathbf{S})$  can be another measure of total variability. However,  $\det(\mathbf{S})$  is not as popular as  $\text{tr}(\mathbf{S})$  since it is very sensitive to any small eigenvalues (variances) even though the others may be large. In PCA the hope is that for some small  $p_a$ ,  $\lambda_1 + \dots + \lambda_{p_a}$  is close to  $\text{tr}(\mathbf{S})$  so the first  $p_a$  principal components explain most of the variation in  $\mathbf{Y}$  and the remaining principal components contribute little. Section 3.1.2 of chapter 3 revisits the approximation of the shape data by estimated principal components.

PCA can be recast as the *multivariate linear model* by treating the observed variables as *responses* and principal directions as *predictors* [Muller, 2007]. One major contribution in this dissertation is the introduction of a formal correlation measure called goodness of prediction described in chapter 3. The correlation measure is designed to evaluate statistical shape models that use principal component analysis (PCA) to characterize shape variability, in terms of their predictive power. This alternative interpretation of PCA as the multivariate linear model is a critical step in presenting the correlation measure. Another contribution on the analysis of object-relation in a multi-object complex is also based on a multivariate linear regression method modified by PCA. Therefore, the next section briefly touches on the multivariate linear model.

### 2.1.3 The Multivariate Linear Model

The multivariate linear model is an advanced topic in multivariate analysis. Here only the minimum background that is necessary to this dissertation is imparted. Refer to [Muller and Stewart, 2006, Kleinbaum et al., 1997, Timm, 2002] for more details about multivariate regression analysis and linear models and to [Muirhead, 1982, Arnold, 1981] for the theory of

multivariate analysis.

The multivariate linear model allows two or more responses to be measured on each independent sampling unit. The definition given in [Muller and Stewart, 2006] is as follows.

**Definition 1.** *A general linear multivariate model  $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$  with primary parameters  $\mathbf{B}$ ,  $\Sigma$  ( $\mathbf{B}$  is called the regression coefficient) has the following assumptions:*

1. *The rows of the  $N \times p$  random matrix  $\mathbf{Y}$  correspond to independent sampling units, that is,  $\mathbf{Y}_i$ ,  $i = 1, \dots, N$  are mutually independent where  $\mathbf{Y}_i$  is the  $i$ th row of the matrix  $\mathbf{Y}$ .*
2. *The  $N \times q$  design matrix  $\mathbf{X}$  has  $\text{rank}(\mathbf{X}) = r \leq q \leq N$  and is fixed and known without appreciable error, conditional on knowing the sampling units, for data analysis.*
3. *The  $q \times p$  parameter matrix  $\mathbf{B}$  is fixed and unknown.*
4. *The mean of  $\mathbf{Y}$  is  $\mu_{\mathbf{Y}} = \mathbf{XB}$ .*
5. *The mean of  $\mathbf{E}$  is  $\mu_{\mathbf{E}} = \mathbf{0}$ .*
6. *The rows of the response matrix  $\mathbf{Y}$  has finite covariance matrix  $\Sigma$ , which is fixed, unknown, and positive definite or positive semidefinite.*

Each  $\mathbf{Y}_i$ ,  $i = 1, \dots, N$ , is assumed to follow a multivariate normal distribution with the same covariance matrix  $\Sigma$ . Then the *maximum likelihood estimates* of  $\mathbf{B}$  and of  $\Sigma$  are given as

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\hat{\mu}_{\mathbf{Y}} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y},$$

and

$$\hat{\Sigma} = \frac{1}{N}(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}),$$

if  $N \geq q + p$ .

### 2.1.3.1 Principal Component Regression

As discussed in chapter 1.1 most shape representations are in the HDLSS situation, which gets more serious for multi-object complexes. Principal component regression (PCR) is a multivariate linear regression method that regresses on the new features given by major principal

directions instead of the original features. Principal components that PCA produces from the input data becomes the new input data, which allows the dimension reduction and associated stability of the features in the input data. Shape analysis of multi-object complexes described in chapter 5 makes use of PCR in estimating a conditional covariance matrix of a target object given selected features in its neighboring objects of the multi-object complex. As can be seen in the equation of  $\Sigma_{11.2}$  in section 2.1.1.6 the estimation of the conditional covariance matrix involves computing an inverse covariance matrix. PCA on the selected features in the neighboring objects reduces the dimension of the shape features and allows the stable estimation of the inverse covariance matrix.

Following the notation in definition 1, let  $\mathbf{X}$  be a data matrix in which each row is an input data sample feature vector (*independent variables*, predictors) and  $\mathbf{Y}$  be the desired output (*dependent variables*, responses) matrix. General linear multivariate regression can be considered as the optimization problem of the least squares equation

$$\hat{\mathbf{B}} = \underset{\mathbf{B}}{\operatorname{argmin}} \|\mathbf{XB} - \mathbf{Y}\|^2, \quad (2.10)$$

where  $\mathbf{B}$  is the  $q \times p$  regression coefficient matrix and the norm is taken as the *Frobenius matrix norm*, i.e., the sum of the squared norms of the individual errors.

By *singular value decomposition* (SVD) or PCA,  $\mathbf{X}'$  can be factored into a diagonal matrix  $\mathbf{\Lambda}$  between two orthogonal matrices  $\mathbf{\Upsilon}$  and  $\mathbf{V}$ , i.e.,  $\mathbf{X}' = \mathbf{\Upsilon}\mathbf{\Lambda}\mathbf{V}'$ , where  $\mathbf{\Upsilon}$  is  $q \times q$ ,  $\mathbf{\Lambda}$  is  $q \times N$ , and  $\mathbf{V}$  is  $N \times N$ .  $\mathbf{\Lambda}$  has nonnegative numbers on the diagonal and zeros off the diagonal. The columns of  $\mathbf{\Upsilon}$  are the principal directions (eigenvectors). Let  $\mathbf{\Upsilon}_k$  be the first  $k \leq q$  columns of the matrix  $\mathbf{\Upsilon}$  and  $\mathbf{\Lambda}_k$  be the  $k \times N$  diagonal matrix of the first  $k$  rows of  $\mathbf{\Lambda}$ . PCR takes the first  $k$  eigenvectors of  $\mathbf{X}'\mathbf{X}$  as the features while leaving  $\mathbf{Y}$  unchanged. That is, the data matrix now becomes  $\mathbf{X}\mathbf{\Upsilon}_k$  and then the least squares equation (2.10) for PCR becomes

$$\min_{\mathbf{B}} \|\mathbf{X}\mathbf{\Upsilon}_k\mathbf{B} - \mathbf{Y}\|^2, \quad (2.11)$$

where  $\mathbf{B}$  is the  $k \times p$  regression coefficient matrix. (2.11) can be further simplified using these identities:  $\mathbf{I}_N = \mathbf{V}'\mathbf{V}$ ,  $\mathbf{I}_q = \mathbf{\Upsilon}'\mathbf{\Upsilon}$ , and  $\mathbf{\Lambda}_k = \mathbf{\Upsilon}'_k\mathbf{\Upsilon}\mathbf{\Lambda}$ , and  $\mathbf{X}' = \mathbf{\Upsilon}\mathbf{\Lambda}\mathbf{V}'$ . Both  $\mathbf{V}$  and  $\mathbf{\Upsilon}$  are

orthogonal matrices. Since the norm does not change by pre-multiplication of an orthogonal matrix, (2.11) can be reduced as follows:

$$\begin{aligned}
\min_{\mathbf{B}} \|\mathbf{X}\Upsilon_k\mathbf{B} - \mathbf{Y}\|^2 &= \min_{\mathbf{B}} \|\mathbf{V}'(\mathbf{X}\Upsilon_k\mathbf{B} - \mathbf{Y})\|^2 \\
&= \min_{\mathbf{B}} \|\mathbf{V}'(\mathbf{V}\Lambda'\Upsilon_k\mathbf{B} - \mathbf{Y})\|^2 \\
&= \min_{\mathbf{B}} \|\Lambda'\Upsilon_k\mathbf{B} - \mathbf{V}'\mathbf{Y}\|^2 \\
&= \min_{\mathbf{B}} \|\Lambda'_k\mathbf{B} - \mathbf{V}'\mathbf{Y}\|^2.
\end{aligned}$$

The solution of  $\mathbf{B}$  with minimal Frobenius norm is now simply

$$\begin{aligned}
\hat{\mathbf{B}} &= (\Lambda_k\Lambda'_k)^{-1}\Lambda_k^{-1}\mathbf{V}'\mathbf{Y} \\
&= \Lambda_k^{-1}\mathbf{V}'\mathbf{Y},
\end{aligned} \tag{2.12}$$

where  $\Lambda_k^{-1}$  is a  $k \times N$  matrix whose diagonal elements are the reciprocals of the diagonal elements of  $\Lambda_k$ .

## 2.2 The Statistical Theory of Shape

In the past, image analysis has been focused on modeling and analysis of local features such as edges and colors at the pixel or voxel scale to characterize complex objects in the image with limited success. Now, characterizing the complex objects by their global shape features has become a major tool for understanding images in general computer vision [Srivastava et al., 2005], and especially for understanding anatomical structures in medical image analysis.

Shape analysis in the study of biological objects dates back to the classic work of Thompson and Bonner [1992] in 1917. Thompson and Bonner [1992] first looked at the way organic things grow and the form they take, and related the anatomical shape to the growth and function of the biological objects. Statistical shape analysis is gaining interest in medical research as a tool to analyze and understand the processes behind growth and the relation between shape change during the growth and the progress of disease.

Statistical shape analysis aims to characterize the population of the shape of an object

within a class of images. To that end, one needs to identify the shape space in which the object (its geometric representation) lives, and then one needs to examine the probability structures induced on such a shape space, which requires advanced mathematical and statistical tools. This section gives an overview of fundamental concepts in the statistical theory of shape.

The usages of the term “shape” in statistical shape analysis usually fall into two categories. Shape in the first category means geometric variations of an object that exclude any similarity transformations present in the population of the object. Shape in the second category includes some or all parts of geometric properties: the scale, the orientation and/or the location. The original Kendall’s shape space for a set of  $p$  points in  $\mathbb{R}_m$ , briefly introduced in section 2.2.1, belongs to the first category. This introduction is from his survey paper [Kendall, 1989].

The rest of this section is organized as follows. Section 2.2.2 describes how probability density estimation and PCA fits in modeling shape variability. Section 2.2.3 explains methods of aligning training models of an object to remove global similarity transformations that do not explain shape changes of the object. The last section: 2.2.4 discusses the issue of correspondence of the geometric shape representations across a population of objects.

### 2.2.1 Kendall’s Shape Space

The object shape representation that Kendall [1989] considered is a set of  $p$  labeled points  $x_1, \dots, x_p \in \mathbb{R}_m$  that are not totally coincident, i.e.,  $x_1 = \dots = x_p$ . Let  $\mathbf{X}$  be the  $m \times p$  matrix of the points  $x_1, \dots, x_p$ . (A vector  $\mathbf{x}$  of these points can also be considered as a point in  $\mathbb{R}_{pm}$ .) To remove the effects of a similarity transformation, the origin is moved to the centroid  $G$  of the  $p$  points, and the points are scaled by making  $L = \sum_{j=1}^p \|G - x_j\|^2$  to 1. After this standardization the rank of the matrix  $\mathbf{X}$  is at most  $p - 1$ , so the matrix is multiplied on the right by a linear transformation  $T$  of the orthogonal group  $O(p)$  that maps the column vector  $(0, \dots, 0, 1)$  to  $(1/\sqrt{p}, \dots, 1/\sqrt{p})$ . This rotational transformation leaves the new matrix to have the final column to be zeros. The remaining non-zero  $m(p - 1)$  elements can be identified with a point on a sphere of unit radius and  $m(p - 1) - 1$  dimensions since the squares of all  $m(p - 1)$  elements sum to 1. This sphere is called the sphere of *presshapes* and denoted by  $S^{m(p-1)-1}$ . The special orthogonal (rotation) group  $SO(m)$  acts on points of the unit sphere  $S^{m(p-1)-1}$  and moves them on  $S^{m(p-1)-1}$ . Finally, the shape space denoted by  $\Sigma_m^p$  is defined

as the *quotient* of the preshape sphere  $S^{m(p-1)-1}$  by  $SO(m)$ . Each  $SO(m)$  equivalence class, that is, points in  $S^{m(p-1)-1}$  that can be transformed into each other by a rotation in  $SO(m)$ , is considered as a single point in this quotient space. Thus this quotient group agrees with the suggested concept of the shape space.

There is a variety of approaches to statistical analysis of shape as there are diverse approaches to geometric description of shape (see section 2.3). Bookstein [1986] developed independently the theory of shape on a finite set of points in  $\mathbb{R}_m$ . His statistical analysis of shapes proceeds within a *tangent space* of  $\Sigma_m^p$  around a sample mean. A tangent space is a linear subspace that best approximates a neighborhood of the sample mean in  $\Sigma_m^p$ . While Kendall [1989] asks questions of an entire shape space, Bookstein [1986] focuses on aspects of multivariate statistical analysis such as differences of mean shape or covariance structures [Bookstein, 1989]. Another important issue in the theory of shape space is the choice of a metric on shape space, although it is not discussed here. For a complete overview about various topics in shape theory and its application refer to the books [Bookstein, 1999, Small, 1996, Dryden and Mardia, 1998]. Bookstein [1999] provides a systematic survey of the statistical study and method for the landmark data, and Small [1996] explains mathematical details of shape theory. [Dryden and Mardia, 1998] provide an easier and more accessible summary of statistics for landmark geometric data. More details on Kendall's shape space can be found in [Kendall, 1984].

## 2.2.2 Estimation of a Shape Probability Distribution

Let  $\Omega_1, \dots, \Omega_N$  be  $N$  training objects drawn from a population and  $m_1, \dots, m_N$  be their geometric representations.  $m_i$  is called a feature vector of the object  $\Omega_i$ ,  $i = 1, \dots, N$ . If the dimension of  $m_i$  is  $p$ , the data matrix of the  $N$  training models is  $N \times p$  matrix whose row is  $m_i'$ .

Statistical shape analysis treats these feature vectors  $m_1, \dots, m_N$  as the observations of a random sample from a  $p$ -variate random variable  $m$ . The most common approach to model  $m$  is to assume that  $m$  follows a multivariate normal distribution  $N_p(\boldsymbol{\mu}, \Sigma)$  of mean  $\boldsymbol{\mu}$  and covariance  $\Sigma$ . These two parameters can be estimated by the sample mean  $\hat{\boldsymbol{\mu}}$  in (2.7) and the sample covariance matrix  $\mathbf{S}$  in (2.8) when the sample size is large, i.e.,  $N \gg p$ . If the  $m_i$ ,  $i = 1, \dots, N$  are independent, (2.7) is a maximum likelihood estimate of the mean  $\boldsymbol{\mu}$ . (2.8) can

be a maximum likelihood estimate of the covariance  $\Sigma$  when multiplied by  $\frac{N-1}{N}$ . The p.d.f. of  $m$  is

$$p(m) = \frac{1}{(2\pi)^{p/2} \det(\mathbf{S})^{1/2}} \exp\left(-\frac{1}{2}(m - \hat{\boldsymbol{\mu}})' \mathbf{S}^{-1} (m - \hat{\boldsymbol{\mu}})\right),$$

which gives the geometric *prior* in the probabilistic deformable model explained later in section 2.3.

Taking  $-\ln$  of both sides gives

$$-\ln p(m) = \ln(2\pi)^{p/2} \det(\mathbf{S})^{1/2} + \frac{1}{2}(m - \hat{\boldsymbol{\mu}})' \mathbf{S}^{-1} (m - \hat{\boldsymbol{\mu}}). \quad (2.13)$$

When this geometric prior is used as a constraint on the shape deformation of an object in the *posterior optimization* (section 2.3.2), the first term in (2.13) is usually ignored since it is a constant. The second term without the constant  $1/2$  is called *Mahalanobis distance* from the mean, i.e.,

$$d(m, \boldsymbol{\mu})^2 = (m - \boldsymbol{\mu})' \mathbf{S}^{-1} (m - \boldsymbol{\mu}) \quad (2.14)$$

The number  $d(m, \boldsymbol{\mu})$  is a probabilistic measure of distance between  $m$  and  $\boldsymbol{\mu}$ . It corrects Euclidean distance such that points along directions of higher variance become closer to the mean. Thus points nearer to the mean in Mahalanobis distance have higher probabilities, which indicates that they are more the probable shapes in shape space [Fletcher, 2004].

However, the estimated sample covariance matrix  $\mathbf{S}$  is likely to be singular because the sample size is not large enough to estimate all  $p(p+1)/2$  unknown parameters in  $\mathbf{S}$ . As mentioned previously most geometric representations  $m$  of shape are high-dimensional, and their features can be strongly correlated due to the complexity of object shape. Available training models are often limited as well due to the cost and time involved in the manual segmentation of images.

In this situation PCA (section 2.1.2) is an attractive method because of its capability to reduce the dimension of the feature space of  $m$ . Principal components become the new uncorrelated random variables that can accurately capture the variability in the training sample. When the normal assumption on the underlying distribution holds, principal components are indeed independent and principal directions describe the independent shape variations of an

object.

The normal assumption is often made to model the shape probability distribution because the normal distribution has nice mathematical properties and only two parameters (mean and covariance) to estimate. It is also a distribution that occurs frequently in natural phenomena. If there is strong evidence that shape of an object cannot be assumed to have normal distribution, PCA may not be an adequate method to use. However, PCA is still a good choice to estimate the probability distribution of shape when the shape of an object can be assumed to follow an elliptical distribution.

### 2.2.3 Alignment

Alignment is a key preprocessing step that is carried out before estimating a shape probability distribution to control the degrees of freedom of transformations in the training models. This step usually mods out global transformations in training data that do not describe shape changes of an object and puts all objects into a common coordinate space. The variation of global transformations tends to be larger than that of the real shape changes. When PCA is applied to the unaligned training data, the estimated major modes of variation can be dominated by these global transformations since the real shape variation is swamped by the global transformations and is not captured by major principal directions.

Alignment method is based on *Procrustes distance*. For two sets of points  $\mathbf{x} = (x_1, \dots, x_p)$  and  $\mathbf{y} = (y_1, \dots, y_p)$  in  $\mathbb{R}_{pm}$  Procrustes distance is defined as the sum of squared distances between their corresponding points,

$$d_p(\mathbf{x}, \mathbf{y}) = \left( \sum_{i=1}^p \|x_i - y_i\|^2 \right)^{\frac{1}{2}}. \quad (2.15)$$

In fact, Procrustes distance is the same as the Frobenius matrix norm in section 2.1.3.1

The process to align one set of points to the other proceeds as follows:

1. Translate the centroid of each point set to the origin,
2. Normalize the spread of both point sets,
3. Rotate one point set to minimize the Procrustes distance defined in (2.15).

This alignment process is called *Ordinary Procrustes Alignment*. The rotation that this alignment produces is equivalent to the least square solution of (2.10) with one additional constraint that  $\mathbf{B} \in SO(m)$ .

Ordinary procrustes alignment is generalized to align a set of  $N$  points  $\mathbf{x}_1, \dots, \mathbf{x}_N \in \mathbb{R}_{pm}$ , which is called *Generalized Procrustes Alignment*. Generalized Procrustes Alignment proceeds as follows: [Gower, 1975]

1. Translate the centroid of each point set to the origin,
2. Choose a reference point set
3. Align each point set  $\mathbf{x}_i$  to the reference by the ordinary Procrustes alignment,
4. Compute the average of the point sets,
5. Set the reference to the average and repeat steps 3 and 4 until the Procrustes distance between the average and the reference is below a threshold.

### 2.2.3.1 Alignment for multi-object complexes

So far the alignment of one object is discussed. In a setting of multi-object complexes the goal of the alignment can vary from situation to situation in terms of transformations to filter out. If one chooses to model both shape and pose, i.e., the relation of a pair of objects, the pair should be aligned relative to some reference object. A mean is usually taken as a reference model, but any template model can be a reference model. In this situation aligning each object independently would lose any correlations between shapes and relative position, scale, and rotation. Thus, the resulting training models after the alignment do not have any information on the relation of the pair of objects. As discussed in [de Bruijne et al., 2006], an alignment method for multi-object complexes must be designed to accommodate a geometric conformation that one wants to capture in the shape space of the multi-object complexes.

Gorcowski et al. [2007] took a multi-scale approach to align the complexes in their study of the discrimination between populations of multi-object complexes. All objects are aligned as a whole first by a similarity transform, which is called a global alignment in [Gorcowski et al., 2007]. After this global alignment each object is aligned individually, which is referred to as

a local alignment. The pose differences of each object between the global and local alignment were combined with the shape change as features for discriminant analysis. In [Gorcowski et al., 2007] the combination of pose and shape is shown to perform better than pose or shape separately. Levy [2008] also used a multi-scale alignment of the bladder models from different patients to produce tight probability distributions that can be estimated from a relatively few training models and to produce clinically usable shape statistics of bladder for segmentation. As shown in these works, the choice of an alignment method directly determines what is captured by shape statistics of multi-object complexes.

### 2.2.4 Correspondence

Correspondence is the problem of establishing consistent anatomical or mathematical matching between primitives of a geometric shape representation across a population of objects. For example, the well-known hand model in [Cootes et al., 1995] was represented by 72 points along the boundary. These points were gathered by first identifying 12 landmark points at the ends and bases of the fingers, and placing equally spaced points in each segment between successive landmarks across the training samples. While it is fairly easy to establish correspondence of the 12 landmarks in this simple 2D hand model, it is challenging and difficult even to define what criteria to use to establish the correspondence of more complex anatomical 3D objects such as the rectum in the abdomen or the cerebral cortex in the brain.

The problem of correspondence can be formulated as a reparameterization of the object representations, which is often solved as an optimization problem of some kind of an objective function. Let  $m_1(x), \dots, m_N(x)$  be the object representations that map from a domain  $X$  to the object representation space  $Y$ .  $X$  is usually  $\mathbb{R}_2$ , and  $Y$  is  $\mathbb{R}_3$  for a surface representation of objects in 3D. For example, with m-reps which describe the interior of object (in section 2.4.2),  $X$  is  $\mathbb{R}_3$ , and  $Y$  is either  $\mathbb{R}_3 \times \mathbb{R}^+$  or  $\mathbb{R}_3 \times \mathbb{R}^+ \times S^2 \times S^2$ . Establishing correspondence in  $m_i(x)$ ,  $i = 1, \dots, N$  is equivalent to reparameterizing them by a set of mappings  $\phi_1(x), \dots, \phi_N(x)$  that maximize an objective function  $F$  matching some selected features of the object over all

values of  $x$  [Yushkevich, 2003], i.e.,

$$\phi_1, \dots, \phi_N = \operatorname{argmin}_{\phi_1, \dots, \phi_N} \int F(m_1(\phi_1(x)), \dots, m_n(\phi_N(x))) dx$$

Davies et al. [2002a,b] proposed an automatic method to establish correspondence of a PDM (in section 2.3.3.1) by minimizing an objective function that includes a *description length* of the training models: the coding length of the data when the data is transmitted as a coded message. The objective function also includes a regularity control that rewards a good object coverage. Without the regularity control term the global minimum of the objective function degenerates each PDM of the training models to a point that gives the least variation across the training models. Twining et al. [2007] recently presented a method to reduce the computational complexity of the optimization by a non-parametric surface-based regularization method. Brechbühler et al. [1995] also posed the parameterization (the embedding of the object surface graph in the surface of the unit sphere) as a constrained optimization problem whose goal is to minimize the distortion of the surface in the mapping in the parametric representation based on spherical harmonic basis functions (in section 2.3.4.2). The constraints that Brechbühler et al. [1995] used are 1) area preservation in the mapping such that any object surface region must map to a region of proportional area on the sphere, and 2) no angle of any spherical quadrilateral becomes negative or exceeds  $\pi$ . These constraints try to keep the shape of all mapped faces onto the unit sphere as similar to their original square as possible.

## 2.3 Probabilistic Deformable Models

### 2.3.1 Deformable Models

Deformable models are an actively researched model-based approach to medical image analysis. They provide a framework to exploit both constraints derived from the image data and prior geographical and geometric knowledge about the objects in finding anatomic structures in 3D medical images. Incorporating the prior knowledge makes deformable models able to handle the significant variability of anatomical structures over time and across people existing in 3D medical images.

The mathematical foundations of deformable models are geometry, physics, and approximation theory [McInerney and Terzopoulous, 1996]. Object shape is described by some geometric representations while object shape is restricted to vary within some physical constraints. Typically, these physical constraints are formulated by deformation energy functions that increase the energy as the model deforms away from a specified natural shape. Approximation theory provides mechanisms to fit the models to the image data by minimizing the energy functions. For example, Barr [1981], Ferrie et al. [1993] introduced superquadric solids, the spherical product of two basis curves called the superellipse and the superhyperbola to represent 3D object shape. Terzopoulous [1991] formulated deformable superquadrics that incorporate global deformation of a superellipsoid with local degrees freedom of a spline so that the local deformation fills up the details that the global deformation missed. Kass et al. [1988] presented the "snake" or active contour model for boundary detection. The deformation of the contour is obtained by minimizing the energy of a spline that is attracted to image features such as lines and edges whereas internal spline forces impose a smoothness constraint. There are other well-known deformable shape models such as the geodesic active contour, [Caselles et al., 1995] a variant of [Kass et al., 1988]; a level set approach of shape evolving along its gradient field [Malladi et al., 1995a]; and an explicit parametric representation of shape using a series of spherical harmonics [Brechtbühler et al., 1995]. A concise review of the deformable models can be found in [McInerney and Terzopoulous, 1996].

### 2.3.2 Probabilistic Deformable Models: Segmentation by Posterior Optimization

Probabilistic deformable models have emerged from viewing the model fitting process of deformable models in a probabilistic framework. This probabilistic framework permits the incorporation of a prior distribution  $p(m)$  of the model  $m$  and provides a measure of the uncertainty of the estimated shape parameters after fitting the model to image data [Szeliski, 1990].

The *posterior probability*  $p(m|I)$  to find  $m$  for a given image  $I$  can be expressed in terms of the prior and *likelihood* by Bayes' rule

$$p(m|I) = \frac{p(I|m)p(m)}{p(I)}, \quad (2.16)$$

where  $p(I|m)$  is the conditional probability or likelihood of producing an image  $I$  for a given model  $m$ . The true model  $m$  for an image can be inferred by maximizing the posterior probability, that is, by finding a model  $m$  that has maximum posterior probability given the image data  $I$

$$p(\hat{m}|I) = \max_m p(m|I) = \max_m \frac{p(I|m)p(m)}{p(I)}. \quad (2.17)$$

This expression can be simplified by taking the log and leaving out the prior distribution of the image data  $p(I)$  that is equal for all  $m$ . Thus, it suffices to maximize  $\ln p(I|m) + \ln p(m)$  to find  $\hat{m}$ , i.e.,

$$\hat{m} = \operatorname{argmax}_m F(m, I) = \operatorname{argmax}_m \{\ln p(I|m) + \ln p(m)\}. \quad (2.18)$$

The objective function  $F(m, I)$  is optimized to find the maximum a posteriori estimate of  $m$  given the image  $I$ , which trades off between the prior information of shape  $p(m)$  and the likelihood information of the image  $p(I|m)$ . When the prior  $p(m)$  follows a uniform distribution,  $F(m, I)$  can be simplified further by eliminating  $p(m)$ . Then  $\hat{m}$  is reduced to just a maximum likelihood estimate.

### 2.3.3 Object Representations via Landmarks

*Landmark representations* use geometrically and biologically salient locations on an object such as the bridge of the nose to describe the object. Landmark representations tend to be a sparse set of points due to the difficulty of extracting landmarks from 2D or 3D image data, which limits the object description only to a global scale [Siddiqi and Pizer, 2008]. However, the strong correspondence between each specified landmark across the population gives a strong advantage to the representations in the study of shape and shape change. For example, combined with other representations, landmark representations can be utilized in the generation of other geometric object representations to improve the correspondence. Merck et al. [2008], Han et al. [2007] made use of landmarks in their methods to create training m-reps. Joshi and Miller [2000] also developed methods to generate large deformation diffeomorphisms by inexact and exact matching of landmarks. For the early work on geometric and statistical properties of landmark representations refer to [Bookstein, 1986, Kendall, 1977, Small, 1988].

### 2.3.3.1 Point Distribution Model

A *point distribution model* (PDM) has been introduced in [Cootes et al., 1995, Cootes and Taylor, 2001] as a statistical shape model for their *Active Shape Model* (ASM). Each object in a PDM is captured by a dense set of landmark points sampled on the boundary of the object. In [Cootes et al., 1995, Cootes and Taylor, 2001], these boundary points are manually placed in a consistent manner on each of training models and are automatically aligned to minimize the sum of squared distances between corresponding points across the training models. After the alignment, the PDM is obtained by estimating the average positions of the points and major modes of variation by PCA on a training set. In later work the PDM was preprocessed to increase correspondence before using it in the ASM.

### 2.3.4 Object Representations via Basis Functions

A parametric shape representation with the probability distribution defined on the parameters provides strong prior information on the shape that introduces a bias toward an expected range of shapes. The spread in the distribution captures the variability in the instances of the object. To that end, the parametrization must be descriptive enough to represent any potential shape of a given geometric entity and to capture global as well as local shape features. It must also be general enough to handle different types of shapes.

The motivation to represent an object shape using some basis functions is that any object can be described as a weighted sum of the given basis functions. These weights (coefficients) are taken as the parametrization of the object representation. These basis functions capture different spatial variations of the object shape and provide a way to decompose the shape variation according to these different spatial variations.

Two parametric deformable models are described below. One model uses *Fourier basis functions* to represent 2D curves. The other model is based on *spherical harmonics* to represent a 3D object of spherical topology.

### 2.3.4.1 Elliptic Fourier Representation

To describe a list of the coordinates  $(x(t), y(t))$  of 2D curves, Staib and Duncan [1992] used the elliptic Fourier representations [Kuhl and Giardina, 1982]

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a_0 \\ c_0 \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \begin{bmatrix} \cos kt \\ \sin kt \end{bmatrix} \quad (2.19)$$

where  $a_0 = \frac{1}{2\pi} \int_0^{2\pi} x(t) dt$ ,  $c_0 = \frac{1}{2\pi} \int_0^{2\pi} y(t) dt$ ,  $a_k = \frac{1}{\pi} \int_0^{2\pi} x(t) \cos kt dt$ ,  $b_k = \frac{1}{\pi} \int_0^{2\pi} x(t) \sin kt dt$ ,  $c_k = \frac{1}{\pi} \int_0^{2\pi} y(t) \cos kt dt$ ,  $d_k = \frac{1}{\pi} \int_0^{2\pi} y(t) \sin kt dt$ .  $m = (a_0, c_0, a_1, b_1, c_1, d_1, \dots)$  gives the raw parameter to represent a 2D curve. The parameter  $m$  is refined to represent the shape in terms of the ellipse geometric properties: semi-major axis length, semi-minor axis length, rotation, and phase shift. The refined parameter is then converted to relative quantities from absolute quantities to isolate the overall rotation parameter and remove the overall phase shift.

This decomposition of 2D curves provides a geometric interpretation of the representation. The first two coefficients  $a_0, b_0$  in (2.19) describe the overall translation of the shape. When represented by the elliptic Fourier representations, 2D curves can be viewed as a sum of phasors that rotate with a speed proportional to their harmonic number  $k$  since each term of the sum in (2.19) defines an ellipse.

In order to use this representation, the sum must be truncated. Since the sinusoid basis functions introduce the notion of frequency in the representation, the series can be truncated to include a certain range of frequencies. Although the elliptic Fourier representation gives an accurate description of most real object boundaries, there is no clear solution to decide the frequency range.

### 2.3.4.2 Spherical Harmonics Shape Description

Parameterized surfaces of 3D objects can be described using spherical harmonics, extending elliptic Fourier representations for 2D closed curves to 3D objects. Brechbühler et al. [1995] first presented a method to parameterize arbitrarily shaped connected objects by a uniform mapping from the object surface into a 2D coordinate space with spherical topology and a method to generate the spherical harmonic descriptors that are invariant to translation,

rotation and scaling given the parametrization of the surface. This invariance property is critical for comparative analysis for different shape of an object. Szekely et al. [1996] combined the desired properties of physical models and the statistical shape variability with the spherical harmonic representation and came up with a coarse-to-fine segmentation procedure. Kelemen et al. [1999] then describes shape deformation by applying PCA to the coefficients of spherical harmonic basis functions and use the estimated shape deformation in segmentation by a variant of the active shape model.

As described in [Brechtbühler et al., 1995], the key step in the spherical harmonic shape description is the mapping of the surface to the unit sphere, i.e., the embedding of the object surface graph into the surface of the unit sphere. Each point on the surface must be mapped to exactly one point on the sphere and vice versa to parameterize every surface point with the unique polar coordinates  $\theta$  and  $\phi$ . This mapping is solved by a constrained optimization that tries to preserve areas to achieve a correspondence across training models. In general, the distortion of the mapped area cannot be avoided, but the optimization keeps the distortion minimal. With the surface parametrization by two polar coordinates  $\theta$  and  $\phi$  the cartesian coordinates  $\mathbf{x}(\theta, \phi)$  of an object surface can be described by three explicit functions

$$\mathbf{x}(\theta, \phi) = \begin{pmatrix} x(\theta, \phi) \\ y(\theta, \phi) \\ z(\theta, \phi) \end{pmatrix} \quad (2.20)$$

where  $\theta \times \phi \in [0, \pi] \times [0, 2\pi)$ .

These Cartesian coordinate functions are then decomposed by the spherical harmonics

$\mathbf{Y}_l^m(\theta, \phi)$

$$\mathbf{x}(\theta, \phi) = \sum_{l=0}^{\infty} \sum_{m=-1}^l \mathbf{c}_l^m \mathbf{Y}_l^m(\theta, \phi), \quad (2.21)$$

where

$$\mathbf{c}_l^m = \begin{pmatrix} c_{x_l}^m \\ c_{y_l}^m \\ c_{z_l}^m \end{pmatrix}. \quad (2.22)$$

Like the elliptic Fourier representation higher frequency components capture more detailed

features of the object.

To use this representation to describe the shape of an object, the sum also must be truncated by limiting  $l$  to  $n_l \geq 0$ . These coefficients  $\mathbf{c}_l^m$  up to the degree  $n_l$  then define a shape representation vector of an object

$$m = (\mathbf{c}_0^{0'}, \mathbf{c}_1^{-1'}, \mathbf{c}_1^{0'}, \mathbf{c}_1^{1'}, \dots, \mathbf{c}_{n_l}^{n_l}). \quad (2.23)$$

These parametric shape descriptions via basis functions have some advantages over the landmark-based representations. The normal and curvature information of the object surface can be derived analytically since all derivatives of the object surface can be computed as derivatives of basis functions. Landmark-based representations must rely on some interpolation methods to generate a smooth surface and to compute the normal and curvature information. Also, parametric shape descriptions establish implicitly a correspondence from an arbitrary set of surface points while landmark-based representations require an explicit correspondence of landmark points across the training models. However, the parametric shape descriptions are sensitive to a small local perturbation of the surface in the sense that all the coefficients need to be recomputed to describe the local shape change. This recomputation of all the coefficients is likely to affect the represented shape of an object. PDM needs only to change the tuples of surface landmarks covering the local perturbation, which does not affect the shape of an object that is not related to the local perturbation.

## 2.4 Medial Representations

Unlike the boundary representations explained in the previous sections medial representations describe an object interior, motivated by the idea that the shape of an object can be better described by the local deformations of the interior such as elongation, bending, twisting, swelling or contraction, and displacement rather than local deformations of the surface of the object. In addition, medial representations not only capture the same information as boundary representations but they also directly capture the boundary normal at each sample point.

This section gives an overview of medial representations and introduces m-reps, one medial

representation. Section 2.4.1 reviews the basic background of medial representations, including the definition of medial loci and mathematical properties of the medial geometry. Section 2.4.2 describes m-reps that are used as the deformable model of this dissertation.

Other medial representations proposed by Amenta et al. [2001], Kimia et al. [1995] and many others are derived from the boundary of the object. M-reps take a view that the boundary of the object can be more stably derived from the medial representation than the reverse. M-reps also take a view that a fixed topology of the medial locus can be derived from a population of an object. This view allows m-reps to have desirable properties for statistical shape analysis. For example, the fixed topology medial structures that m-reps capture enable the estimation of their probabilities from training samples. Section 2.4.3 reviews the statistical analysis of m-reps. All the methods developed in this dissertation are based on the methods described in this section. Siddiqi and Pizer [2008] provide a thorough and extensive overview of methods, algorithms, and applications that are based on medial representations in computer vision and image analysis. Most materials in this section are from [Siddiqi and Pizer, 2008].

### 2.4.1 The Medial Locus

The medial locus is a set of points midway between two sections of the boundary of an object together with the half widths of the two sections. It was introduced by Blum [1967], Blum and Nagel [1978] and has been used widely since in computer vision, image analysis and computer graphics [Bloomenthal and Shoemake, 1991, Igarashi et al., 1999, Amenta et al., 2001]. In spite of the simple definition of the medial locus, the general mathematical properties of the medial locus have been found recently [Giblin and Kimia, 2004, Damon, 2003, 2004] because the understanding of its mathematical properties requires a careful mathematical treatment from differential geometry and singularity theory. The terms *medial axis*, *symmetric axis*, and *skeleton* are often used interchangeably in these works. The book by Siddiqi and Pizer [2008] provides a clear summary of this work without too much mathematical detail.

The definition of Blum’s medial locus is based on the *maximal inscribed ball*.

**Definition 2.** *Let  $S$  be a connected closed set in  $\mathbb{R}_n$ . A closed ball  $B \subset \mathbb{R}_n$  is called a maximal inscribed ball in  $S$  if  $B \subset S$  and there does not exist another ball  $B' \neq B$  such that  $B \subset B' \subset S$ .*

A closed ball is defined as  $B_r(\mathbf{p}) = \{\mathbf{x} \in \mathbb{R}_n \mid \|\mathbf{x} - \mathbf{p}\| \leq r\}$ .

When  $S$  is *homeomorphic* to the  $n$ -dimensional closed ball, i.e., there exists a one-to-one, onto, invertible continuous mapping to the ball, the set  $S$  is called an  $n$ -dimensional object. This definition of the  $n$ -dimensional object guarantees that the object does not intersect itself and is bounded, which are the typical properties that most real-world objects would have. But this  $n$ -dimensional object cannot represent objects like rings that have a hole.

**Definition 3.** *The medial locus of  $S$  is the set of centers  $\mathbf{p}$  and radii  $r$  of all the maximal inscribed balls in  $S$ .  $(\mathbf{p}, r)$  is called a medial point of  $S$ .*

The medial locus can also be defined analytically using the *grassfire analogy*. When the boundary of a patch of grass is set on fire simultaneously, the fire fronts propagate inwardly from the boundary along the normal to the fire front with a constant speed. As the fire progresses, the fire fronts from different parts of the boundary meet and quench each other at points called *shocks*. These shock points make up the medial locus together with the time  $t$  at which each shock is formed. This definition can be formulated as

$$\frac{\partial C(t, x)}{\partial t} = -\alpha \mathbf{N}(x) ,$$

where  $C(t, x)$  is the fire front at time  $t$ , parameterized by  $x$ ,  $\mathbf{N}(x)$  is the unit outward normal to the fire front.  $\mathbf{N}(x)$  can be computed by the gradient of  $C(t, x)$  when  $C(t, x)$  is the distance map, that is, the Euclidean distance from  $x$  to the nearest point on the boundary of the object.

Each medial point can be classified according to the multiplicity  $k$  and the order of contact  $m$  between the surface of an object and the maximal inscribed ball centered at the medial point [Giblin and Kimia, 2004]. A medial point  $(\mathbf{p}, r)$  is labeled as an  $A_m^k$  point when the inscribed maximal ball  $B_r(\mathbf{p})$  is tangent at  $m$  distinct points and the order of tangent contact between the  $B_r(\mathbf{p})$  and the boundary is  $k$ .  $k$  indicates how tightly  $B_r(\mathbf{p})$  is fit to the object's surface. No superscripts means that  $B_r(\mathbf{p})$  contacts at a single point. In 2D and 3D, almost all the possible types of medial points that can occur generically are illustrated in Fig. 2.1. The following theorem in [Siddiqi and Pizer, 2008] specifies all the types of medial points in 3D.

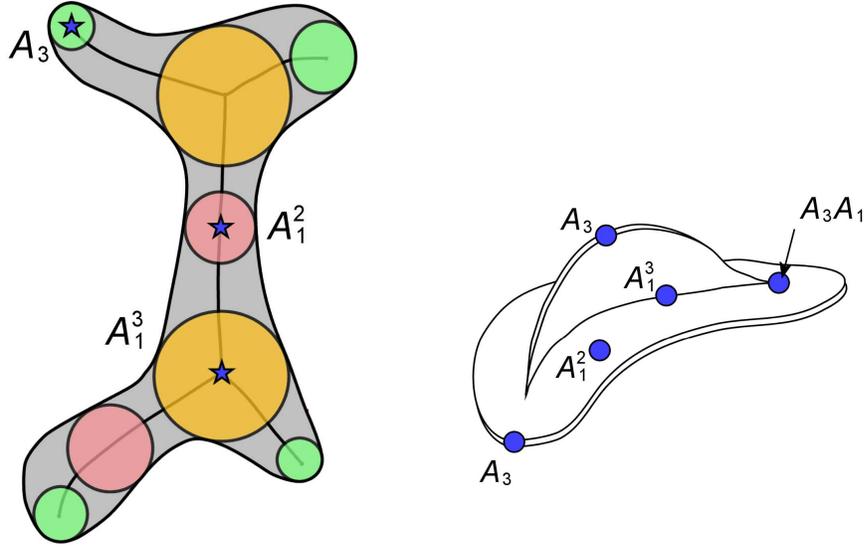


Figure 2.1: Different classes of medial points classified by Giblin and Kimia [2004]. Left: for a 2D object. Right: for a 3D object. (This figure is from the medial book. [Siddiqi and Pizer, 2008])

**Theorem 4.** *The medial locus in 3D consists of*

1. *sheets of  $A_1^2$  medial points;*
2. *curves of  $A_1^3$  points where three sheets join;*
3. *curves of  $A_3$  points, which bound the unconnected edges of the sheets and for which the corresponding boundary points fall on a crest;*
4.  *$A_1^4$  medial points, which occur when four  $A_1^3$  curves meet;*
5.  *$A_1A_3$ , that is,  $A_1$  and  $A_3$  contact at distinct points, which occur when an  $A_3$  curves meets an  $A_1^3$  curve.*

The last two types,  $A_1^4$  and  $A_1A_3$  medial points, do not occur in 2D. In 2D, bitangent  $A_1^2$  points form curves,  $A_1^3$  points occur when three curves of  $A_1^2$  points meet, and  $A_3$  points are generically the ends of  $A_1^2$  point curves.

Let  $\mathbf{b}^{\pm 1}$  be two boundary points where  $B_r(\mathbf{p})$  meets and  $\mathbf{U}^{\pm 1}$  be unit-length vectors orthogonal to the boundary. Then  $\mathbf{b}^{\pm 1} = \mathbf{p} + r\mathbf{U}^{\pm 1}$ . Following the analogy between the bitangent disc and a wheel, the two vectors  $\mathbf{S}^{\pm 1} = r\mathbf{U}^{\pm 1}$  are called *spokes* in the m-rep literature.

In addition to  $\mathbf{p}$  and  $r$ , Blum and Nagel [1978] describe two more first order properties of the medial curve in 2D to characterize points in the medial curve. The first is a unit-tangent vector  $\mathbf{U}_0$  along the medial curve, and the second is the object angle between  $\mathbf{U}_0$  and  $\mathbf{U}^{\pm 1}$ . The curvature along  $\mathbf{U}_0$  describes how the figure bends at  $\mathbf{p}$ . The object angle indicates the rate at which the object narrows along the  $\mathbf{U}_0$  direction as  $r$  describes locally the thickness of the figure at  $\mathbf{p}$ . The two spokes  $\mathbf{U}^{\pm 1}$  come together and meet at  $\mathbf{U}_0$  at the end of the medial curve in 2D, i.e., the object angle is 0. The corresponding boundary point is called a crest point.

The medial geometry for 3D objects has a lot common with 2D medial geometry. In 3D medial geometry we have bitangent spheres instead of bitangent discs in 2D. The two spokes are still perpendicular to the object surface, and their difference is still normal to the medial locus. Also, the end curve of the medial sheet is formed by centers  $\mathbf{p}$  of medial points whose spokes have come together. The locus of the corresponding spoke ends forms a crest curve on the boundary. The normal curvature at the points on the crest curve reaches a local maximum in the *principal direction* across the crest. Principal directions are the directions in which the maximum and minimum values of the normal curvature of a surface happen O'Neill [1997].

Based on this observation, a *medial atom* is defined to describe the medial locus of an object up to two differential orders.

**Definition 5.** An  $n$ -dimensional medial atom of order 0 is a tuple  $\mathbf{m} = (\mathbf{p}, r) \in \mathbb{R}_n \times \mathbb{R}^+$ .

A medial atom of order 0 simply represents the center and the radius of a maximally inscribed ball. Given medial atoms of order 0 sampled from the medial locus, the object can be approximately reconstructed by "shrink-wrapping". But the order 0 medial atoms lack the direct information as to the local symmetric relationships between the pair of points of contact where  $B_r(\mathbf{p})$  and the object boundary meet. To supplement the drawback of order 0 medial atoms, the next definition of the order 1 medial atoms adds the local symmetry information to the order 0 medial atom.

**Definition 6.** An  $n$ -dimensional medial atom of order 1 is a tuple  $\mathbf{m} = (\mathbf{p}, r, \mathbf{U}^{+1}, \mathbf{U}^{-1}) \in \mathbb{R}_n \times \mathbb{R}^+ \times S^{n-1} \times S^{n-1}$  where  $S^{n-1}$  is the unit  $n - 1$ -dimensional sphere.

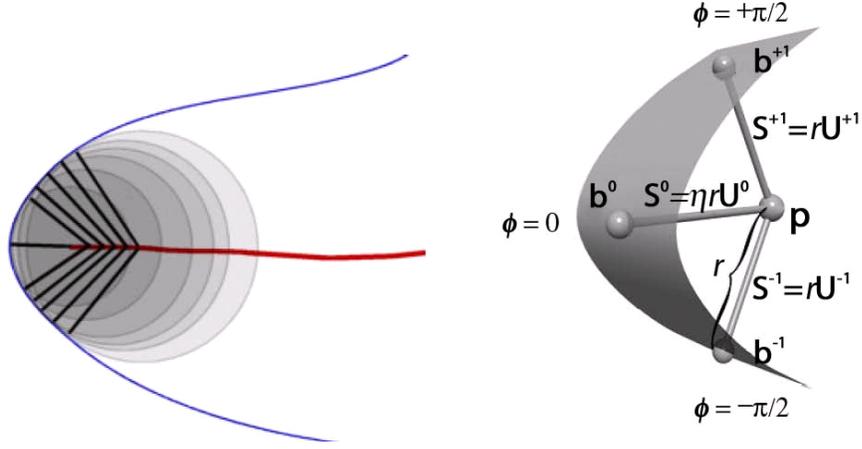


Figure 2.2: Left: In cross-section, the continuous relationship between a point on a medial surface and the points of contact between the disk inscribed at the point and the boundary of the object asymptotes at the crest point. Right: To account for this asymptotic relationship m-reps describe ends of figures using special end atoms with three spokes that describe the entire end-cap of the figure, symmetric about the crest at  $\mathbf{b}^0$ . (This figure is from the medial book. [Siddiqi and Pizer, 2008])

With these two unit-length vectors  $\mathbf{U}^{\pm 1}$  a medial atom of order 1 can reconstruct exactly the object boundary that its corresponding bitangent ball  $B_r(\mathbf{p})$  describes since the boundary points  $\mathbf{b}^{\pm 1} = \mathbf{p} + r\mathbf{U}^{\pm 1}$ .

The two spokes  $\mathbf{U}^{\pm 1}$  at  $\mathbf{p}$  move towards each other at an infinite rate in the limit as  $\mathbf{p}$  approaches the end of the medial sheet as shown in Fig. 2.2. This property poses some problems to the discrete m-rep representation, one of the medial representations that is described later in section 2.4.2.1. The problems and the approach taken to deal with the problems are discussed in the next section.

For some smooth objects the locus of  $\mathbf{p}$  of their medial atoms can form a manifold with boundary. The object or the part of the object represented by such a manifold is called a *figure*. For most 3D objects the manifold is two-dimensional in which case the figure is called a *slab* figure. However, the manifold can degenerate to a curve for long narrow objects with circular cross-sections or to a point for spheres. The collection of figures can form a multi-object complex in which objects are clustered together.

## 2.4.2 M-rep

M-reps are representations of object interiors that consist of hierarchies of continuous sheets of 1<sup>st</sup> order medial atoms. For a given reference atom each medial atom is expressed in terms of 1) a translation of the hub position  $\mathbf{p}$  (3 parameters), 2) a magnification of the spoke length  $r$  (1 parameter), and 3) two rotations of the spokes  $\mathbf{U}^{\pm 1}$  (4 parameters) of the reference atom. This transformation is called the *medial atom transformation*. By representing the object interior using a medial atom transformation, m-reps have the following properties:

1. By medial atom transformations they explicitly capture local bending and twisting (rotation), local magnification, and local elongation, and they separate these from one another.
2. They are based on the subdivision of an object into figures, i.e., main bodies, protrusions, and indentations. Moreover, they provide a fixed topology of such branching for a population of objects and thus allow statistics on this population.
3. They provide a local coordinate system for object interiors that can provide correspondences across instances of an object.
4. They allow neighboring 1st order medial atoms to be understood as medial atom transformations of each other. This allows rich characterization of neighbor relationships, for situations internal to a figure, between figures, or between objects.

### 2.4.2.1 Discrete M-reps

Two different representations of m-reps have been developed and implemented. One is an analytic form of m-reps that describe medial atoms of order 0 using B-splines [Yushkevich et al., 2002, 2005, Terriberry and Gerig, 2006]. The one used primarily in this dissertation, called the *discrete m-rep* [Pizer et al., 2003], represents each sheet by a discrete grid of medial atoms of order 1, from which we can interpolate a continuous sheet.

A discrete m-rep is formed by sampling the medial sheet over a spatially regular lattice to form a mesh of medial atoms. As pointed out in the previous section 2.4.1, the spokes of the atoms move towards each other at an infinite rate as the medial atoms approach the end of

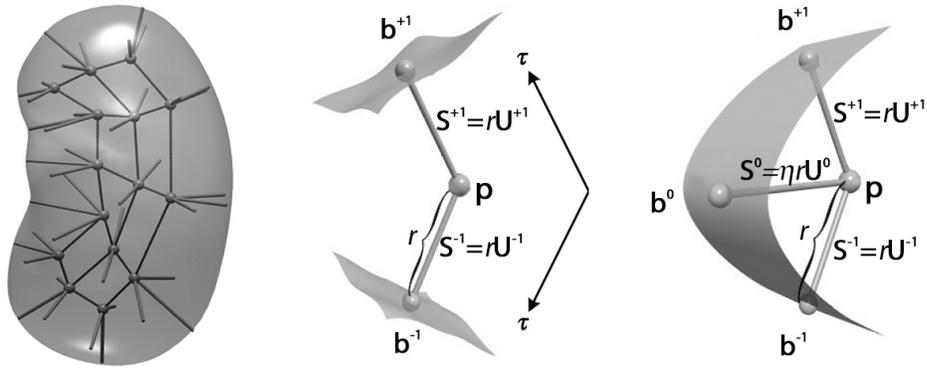


Figure 2.3: Left: a single figure m-rep for a kidney and the object boundary implied by it. Middle: an internal atom of two spokes  $\mathbf{S}^{+1/-1}$ , with  $\tau$  to parameterize the object interior along the spokes. Right: an end atom with an extra bisector spoke  $\mathbf{S}^0$ . (This figure is from the medial book. [Siddiqi and Pizer, 2008])

the medial sheet. Representing sampled points on the edge of the medial sheet by 1<sup>st</sup> order medial atoms is problematic since at these points the two spokes collapse, which makes the representation unstable. Moreover, the edge curve is a critical portion of the medial sheet in describing the shape of an object. To make the discrete m-rep stable, the interior description is cut off while the object angle is still significant, and the medial axis of the end portion in discrete m-reps continues straight from the place where the interior description stopped. Thus, a new spoke  $\mathbf{S}^0$  along the bisector  $\mathbf{U}^0$  is added to the representation of the end medial atom. The end of  $\mathbf{S}^0$  is incident to the crest of the implied boundary. The length of  $\mathbf{S}^0$  is  $\eta r$  and is determined by a crest sharpness parameter  $\eta \in \mathbb{R}^+$  additional to the parameters of the interior medial atoms. As a result, internal medial atoms have 8-parameters and edge atoms have one more parameter  $\eta$  in addition to the 8-parameters.

With the discrete m-reps Merck et al. [2008], Han et al. [2007] developed a principled approach to achieve correspondence at low cost while building training m-rep models. This approach will be explained in section 2.5. For the rest of this dissertation m-reps means discrete m-reps.

## 2.4.3 Statistical Shape Analysis of M-reps

### 2.4.3.1 Shape Space of M-reps and PGA

Fletcher et al. [2004] realized that a medial atom  $\mathbf{m} = (\mathbf{p}, r, \mathbf{U}^{+1}, \mathbf{U}^{-1})$  can be defined as an element of the “Riemannian symmetric space”  $G = R_3 \times R^+ \times S^2 \times S^2$  where the positional translation  $\mathbf{p} \in R^3$ , the spoke length magnification  $r \in R^+$ , and two rotations of unit spoke directions  $\mathbf{U}^{\pm 1} \in S^2$  the unit sphere. Any rotation in  $SO(3)$  leaves two antipodal points of  $S^2$ , that the axis of the rotation passes through, fixed, which implies  $S^2 = SO(3)/SO(2)$ .

If an m-rep has  $d$  medial atoms, the m-rep shape space becomes  $\mathcal{M} = G^d$ . The direct product of symmetric spaces is still a symmetric space. Because the spoke directions in medial atoms are values on the unit sphere, the symmetric space in which m-reps live is not a flat (linear) feature space, but instead is a curved differential manifold.

Fletcher et al. [2004] developed a generalized version of PCA called principal geodesic analysis (PGA) for probability density estimation of geometric entities that form a symmetric space including nonlinear transformations. PGA involves computing a Fréchet mean on the actual curved manifold via a gradient descent method, and then doing PCA on the linear space that best approximates  $\mathcal{M}$  around the Fréchet mean. The principal directions estimated on this linear space are projected back down onto the curved space  $\mathcal{M}$  and are called the *principal geodesics*.

### 2.4.3.2 Tangent Space at a Point of a Manifold

Riemannian symmetric spaces are associated with the Riemannian metric  $d(\cdot, \cdot)$ . The main property of differential manifolds is that locally they behave like Euclidean space. Thus, for every point  $p$  in a given differential manifold  $M \in \mathbb{R}^d$ , a linear subspace that best approximates a neighborhood of  $p$  in  $M$  can be associated. The linear subspace is called a tangent space at the point  $p$  and is denoted by  $T_p M$ .

Two key functions that map points between  $T_p M$  and  $M$  are the Riemannian exponential map and the Riemannian log map. The Riemannian exponential map at  $p \in M$  denoted by  $Exp_p : T_p M \rightarrow M$  is a diffeomorphism in a neighborhood  $U \subset T_p M$  mapping a point  $x \in U$  to a point  $Exp_p(x) \in M$  along the geodesic from  $p$  to  $Exp_p(x)$ . A geodesic in a manifold

$M$  is the shortest smooth curve segment between two points in  $M$ . In the Euclidean space a straight line is the geodesic path between two points. The inverse mapping of the Riemannian exponential map is called the Riemannian log map and denoted by  $Log_p : Exp_p(U) \rightarrow T_pM$ .

It is useful to select a metric on  $T_pM$  such that distances to  $p$  on the Riemannian manifold are equal to those on  $T_pM$ . That is, distances from a point  $x$  on the tangent plane to  $p$  denoted as  $\|x\|$  are equal to the geodesic distance  $d(Exp_p(x), p)$  on the manifold.

## 2.5 Construction of Training Models

Fitting shape representations to the binary images formed from human expert segmentations generates training models from which probability distributions of the shape of anatomic objects can be estimated. The quality of training models determines the quality of the estimated probability distribution that in turn affects any applications using the shape statistics.

Merck et al. [2008] presented a fitting method that iteratively estimates the best geometric model for a given binary image and shape probability distribution using a set of soft constraints. The method proceeds through iterations by successively relaxing the geometric constraints relative to the match to the binary image. The goal of this iterative scheme is to produce models tightly fitted to the target binary image. This approach was tested over a variety of complexes of single figure objects to fit m-reps and proved to work quite effectively.

An important issue in fitting m-reps is to guarantee correspondence across the training m-reps. M-reps are stable in the sense that small changes in the medial atoms of an m-rep do not bring large changes in its implied boundary. This is a strength of discrete m-reps. However, this property implies that multiple arrangements of medial atoms can describe the same boundary of an object. This ambiguity is dealt with by fitting in a predictable fashion into the boundary and by minimizing a penalty function that produces regular spacing between medial atoms.

In Merck et al. [2008]’s method fitting is achieved by an optimization process that minimizes an objective function  $F_{obj}(m, I)$  of a weighted sum of two data match functions and of three geometry penalties [Han et al., 2007]. The two data match functions are an image match function  $F_{img}$  and a landmark match function  $F_{lmk}$ .  $F_{img}$  forces the implied surface of  $m$  to

match with the boundary voxels of the binary.  $F_{lmk}$  enforces explicit correspondences between the landmarks in the binary image either identified by experts or by some programs and the matching surface points of the model.

The three geometry penalties are an irregularity penalty function  $F_{reg}$ , an illegality penalty function  $F_{leg}$ , and a reference model penalty  $F_{ref}$ .  $F_{reg}$  penalizes non-uniform spacing of the grid of medial atoms and non-uniform changes in spoke lengths and spoke directions of medial atoms.  $F_{reg}$  implicitly contributes to establish correspondence of medial atoms across the training cases.  $F_{leg}$  is a penalty unique to discrete m-reps, making use of a shape operator called  $\mathbf{S}_{rad}$  introduced by Damon [2003] for medial geometry. The illegality penalty function tries to prevent local self-intersections or creases from happening in the implied surface of a discrete m-rep.  $F_{ref}$  penalizes deforming too much from a given reference model.  $F_{reg}$  and  $F_{leg}$  prevent irregular deformations of the initial template model and the folding of its implied surface while  $F_{lmk}$  keeps the landmarks of the deforming model and their corresponding points in the binary image together.

At the end of each iteration during the fitting process, the shape space and the probability distribution on the shape space are updated, so optimizations in one step are over a shape space trained from models produced in the previous step. Each iteration produces better estimates of the mean model and the principal modes, and the fitting process stops when the estimated mean changes little.

## 2.6 Segmentation of M-reps

Section 2.3.2 introduced segmentation by posterior optimization of probabilistic deformable models. The segmentation framework of m-reps described in [Pizer et al., 2003] also follows the posterior optimization approach, i.e. deforming a template model by minimizing or maximizing an objective function that consists of a log prior term (a geometric typicality term) and a log likelihood term (a geometry-to-image match term). The geometric typicality measures how probable the shape of the deforming template model is in the population of the object, and the geometry-to-image match measures how likely the target image values relative to the deforming template model are in the population of the image.

The geometric typicality is a Mahalanobis distance (2.14) in a tangent space to the symmetric space  $M = G^d$  as shown in section 2.4.3.1 when PGA is used to estimate the shape prior. The Mahalanobis distance (2.14) is simplified to the sum of coefficients along all principal geodesics. During optimization the geometric template model (usually mean) is deformed by varying these coefficients.

The geometry-to-image match comes from the regional intensity summary based match method developed by Broadhurst et al. [2006]. The image regions are determined by object-relative coordinates that training m-reps produce since object-relative coordinates of m-reps provide correspondence to intensities of training gray-scale images. The method represents each regional intensity collection by the curve of intensity values versus quantile called an RIQF (regional intensity quantile function) [Pizer et al., 2005a, Broadhurst et al., 2005]. An RIQF is the inverse of the cumulative distribution function of image intensity in the region. Broadhurst et al. [2006] showed that 1) QFs do not suffer from quantization effects like most intensity histogram-based methods, 2) the analysis of RIQFs can be captured by linear statistics on DRIQFs (Discrete sampling of RIQFs), 3) Because the Earth-mover's distance on histograms (probability distributions) is equivalent to a Euclidean distance on quantile functions (QFs) [Levina, 2001], QFs form a linear space that PCA can be applied to, and 4) DRIQFs of a continuous distribution function that is parameterized by locations and scales form a 2-d linear space. In practice, DRIQFs also suffer from the HDLSS situation, so PCA is used to estimate the probability distribution of DRIQFs. Then the geometry-to-image match is a Mahalanobis distance of a DRIQF obtained from a target image region, which is the sum of coefficients along all principal directions in the region.

M-rep based segmentation is inherently multi-scale starting from the scale of multi-object complexes or a single figure down to the scale of medial atoms. As mentioned in section 2.2.3.1, the relevant shape properties at each scale and the relation to the neighboring geometric entities (figures or section of medial atoms) need to be taken into account in the geometric typicality term. One main topic of this dissertation is to estimate the probability shape distribution of multiple objects, which is related to the segmentation in the scale of multi-object complexes.

Segmentation of single figures is implemented in multi-scale. Initially a template model is placed into the target image by matching a few either manually or automatically chosen

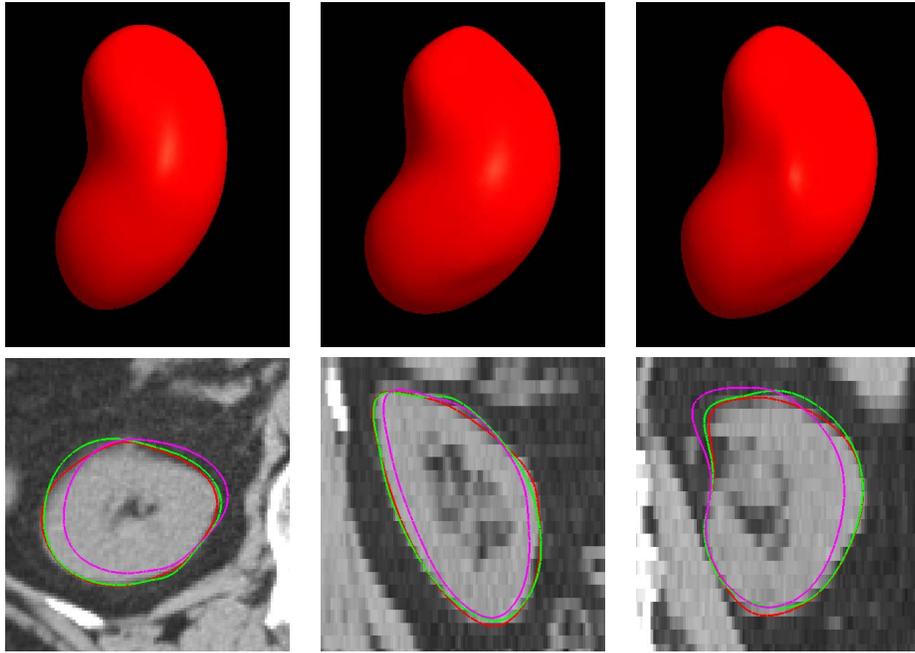


Figure 2.4: Segmented kidneys, aligned via landmark, figure stage, and atom stage result (This figure is from the medial book. [Siddiqi and Pizer, 2008])

landmarks. The segmentation proceeds through a number of stages at successively smaller levels of scale. Fig. 2.4 shows the segmentation results of a kidney at each stage: model alignment by landmarks, the figure stage, and the medial atom stage.

## Chapter 3

# Evaluation of Statistical Shape Models <sup>1</sup>

Principal component analysis (PCA) has become a very popular method used to analyze shape variability. Cootes et al. [1995], Bookstein [1999] were early users of PCA for shape analysis. The usefulness of PCA is two-fold: 1) decomposition of population variables into an efficient reparametrization of the variability observed on the training data and 2) dimension reduction of population variables that allows a focus on the subspace of the original space of population variables. Especially, #2 is a major advantage of PCA in shape analysis because most shape representations presented in the literature have very high dimensional feature spaces due to the complexity of object shape. On the other hand, available training samples are limited due to the cost and time involved in the manual segmentation of images. This kind of data is called high dimension, low sample size (HDLSS) in statistics. The measure we propose in this chapter applies to statistical shape models that use PCA as their method to describe shape variability.

Given a set of training samples from a population, PCA allows us to extract important directions (features) from these training samples and to use these features to describe new members of the population. The predictability of statistical shape models refers to this power of statistical shape models to predict a new member in the population.

There exist in statistics several criteria to judge the appropriateness of any dimension reduction technique. However, we will mainly concentrate on the criteria of predictability in view of the many practical applications of statistical shape models as explained in section 1.1.2.

---

<sup>1</sup>This chapter is coauthored with Surajit Ray and Keith Muller [Jeong et al., 2008a].

In addition, we will touch on the questions of the interpretability and the stability of the extracted directions that are equally important as the predictability: Does the direction have a meaningful interpretation or are they mere mathematical objects?; How do these directions differ from sample to sample, and how many training samples do we need to get a stable estimate of the important directions?

## 3.1 Background

This section provides the background necessary to understand goodness of prediction of PCA in a multivariate regression setting. PCA was covered in section 2.1.2, and the basic definition of the multivariate linear model was given in section 2.1.3 of chapter 2. Section 3.1.1 explains the decomposition of a covariance matrix by PCA taken from [Muller, 2007]. Section 3.1.2 describes the approximation of sample objects given by the mean and major principal coefficients.

### 3.1.1 Decomposition of the Covariance Matrix

Let  $\mathbf{Y} = Y_1, \dots, Y_N$  be  $N$  sample vectors from a  $p$ -variate distribution. Let  $\mathbf{Y}$  be a  $N \times p$  data matrix with  $Y_i'$  as rows, and let  $\Sigma$  be a corresponding  $p \times p$  variance covariance matrix.

By PCA (equivalently spectral decomposition),  $\Sigma$  can be written as  $\Sigma = \mathbf{Y} \mathbf{D}(\boldsymbol{\lambda}) \mathbf{Y}'$ , where  $\mathbf{D}(\boldsymbol{\lambda})$  is a  $p \times p$  diagonal matrix of nonnegative eigenvalues  $\{\lambda_i\}$  for  $i = 1 \dots p$  and where  $\mathbf{Y}$  is a matrix of column eigenvectors of the non-negative symmetric matrix  $\Sigma$ .

Most of the time  $N \ll p$  due to HDLSS situation, i.e., the sample size is much smaller than the dimension of the shape features. In general, some number of principal directions less than  $N$ , say  $p_a$ , covers most of the sample's variation, e.g., 80% or 90% of the total variation. Partly because the estimated principal directions explaining the smaller variation of the data are unreliable, we usually take the first  $p_a$  eigenvectors to approximate the covariance matrix. For  $i > p_a$ , the  $i$ -th eigenvector estimated from one sample is likely to be different from the  $i$ -th eigenvector estimated from another sample, and these later eigenvectors might not appear in the same order. Also, taking only the first  $p_a$  eigenvectors reduces the dimension of the original shape feature space considerably, which can be useful in applications of statistical shape models.

Let  $p_b = p - p_a$ . Considering  $p_a$  and  $p_b$  columns of matrices  $\mathbf{Y}$  and  $\mathbf{D}(\boldsymbol{\lambda})$  as two partitions gives  $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_a & \mathbf{Y}_b \end{bmatrix}$ ,  $\boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{\lambda}'_a & \boldsymbol{\lambda}'_b \end{bmatrix}'$ , and

$$\begin{aligned}
\boldsymbol{\Sigma} &= \mathbf{Y}\mathbf{D}(\boldsymbol{\lambda})\mathbf{Y}' \\
&= \begin{bmatrix} \mathbf{Y}_a & \mathbf{Y}_b \end{bmatrix} \begin{bmatrix} \mathbf{D}(\boldsymbol{\lambda}_a) & \mathbf{0} \\ \mathbf{0} & \mathbf{D}(\boldsymbol{\lambda}_b) \end{bmatrix} \begin{bmatrix} \mathbf{Y}'_a \\ \mathbf{Y}'_b \end{bmatrix} \\
&= \mathbf{Y}_a\mathbf{D}(\boldsymbol{\lambda}_a)\mathbf{Y}'_a + \mathbf{Y}_b\mathbf{D}(\boldsymbol{\lambda}_b)\mathbf{Y}'_b \\
&= \boldsymbol{\Phi}_a\boldsymbol{\Phi}'_a + \boldsymbol{\Phi}_b\boldsymbol{\Phi}'_b \\
&= \boldsymbol{\Phi}\boldsymbol{\Phi}' ,
\end{aligned}$$

where  $\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}_a & \boldsymbol{\Phi}_b \end{bmatrix}$ ,  $\boldsymbol{\Phi}_a = \mathbf{Y}_a\mathbf{D}(\boldsymbol{\lambda}_a)^{1/2}$ , and  $\boldsymbol{\Phi}_b = \mathbf{Y}_b\mathbf{D}(\boldsymbol{\lambda}_b)^{1/2}$ . Without loss of generality  $\mathbf{Y}'\mathbf{Y} = \mathbf{I}_p$  and  $\{\lambda_i | i = 1 \dots p\}$  are sorted from largest to smallest. Hence if  $\text{rank}(\boldsymbol{\Sigma}) = p_a$ , then  $\boldsymbol{\lambda}_b = \mathbf{0}$  and  $\boldsymbol{\Sigma} = \mathbf{Y}_a\mathbf{D}(\boldsymbol{\lambda}_a)\mathbf{Y}'_a = \boldsymbol{\Phi}_a\boldsymbol{\Phi}'_a$ .  $\boldsymbol{\Phi}_b\boldsymbol{\Phi}'_b \approx \mathbf{0}$  and  $\boldsymbol{\Sigma}_Y \approx \mathbf{Y}_a\mathbf{D}(\boldsymbol{\lambda}_a)\mathbf{Y}'_a$  are assumed when the covariance matrix is approximated with the first  $p_a \ll p$  components.

### 3.1.2 PCA for Statistical Shape Analysis

In statistical shape analysis, the  $N$  rows of  $\mathbf{Y}$  correspond to people or images, and the  $p$  columns of  $\mathbf{Y}$  correspond to features in shape space. With a  $p \times 1$  mean shape feature vector  $\boldsymbol{\mu}$  and an  $N \times 1$  column vector  $\mathbf{1}_N$  of 1's, the full set of component scores is  $(\mathbf{Y} - \mathbf{M})\mathbf{Y}$ , where  $\mathbf{M} = \mathbf{1}_N\boldsymbol{\mu}'$ . Component scores for retaining  $p_a$  components are computed by  $(\mathbf{Y} - \mathbf{M})\mathbf{Y}_a$ . Let  $\mathbf{Y}_c$  be the approximating set of component scores  $(\mathbf{Y} - \mathbf{M})\mathbf{Y}_a$ . Approximating the data with the components gives

$$\begin{aligned}
\mathbf{Y} \approx \mathbf{Y}_a &= \mathbf{M} + \mathbf{Y}_c\mathbf{Y}'_a \\
&= \mathbf{M} + (\mathbf{Y} - \mathbf{M})\mathbf{Y}_a\mathbf{Y}'_a ,
\end{aligned} \tag{3.1}$$

with the  $N \times p$  matrix  $\mathbf{Y}_a$  of rank  $p_a \ll p$ , while  $\mathbf{Y}_a\mathbf{Y}'_a$  is  $p \times p$  and of rank  $p_a \ll p$ .

## 3.2 Goodness of Prediction

In most applications of statistical shape models, a mean shape and modes of shape variation (eigenvectors, principal directions) estimated from a set of training models are used to approximate a shape in a new image. Considering how estimated shape statistics are used in applications, the prediction accuracy of estimated shape models can be assessed properly by using estimates from one set to predict shapes in a different set of models. We call a set of training models used for estimating shape statistics a training set and the different set of models a test set.

In this section, we first describe a modified interpretation of the original approach that is proposed in [Muller, 2007] to meet our need. Our focus here is measuring goodness of prediction of the estimated covariance (second moment accuracy), not of the estimated mean (first moment accuracy). We simplify the original approach to that end. Then we present the average squared correlation as the measure of association between the training set and the test set, that is, the goodness of prediction of the estimated covariance. The term "goodness of prediction" here is used in this restricted sense. Full details and proofs of the original approach can be found in [Muller, 2007] and in a series of forthcoming papers.

### 3.2.1 PCA Input to Multivariate Regression

Let  $\mathbf{Y}_t$  and  $\mathbf{Y}_s$  be the training data and the test data matrices respectively. The subscripts  $t$  and  $s$  indicate the training and the test set. Our objective is to measure the degree to which the probability distribution estimated from  $\mathbf{Y}_t$  describes the probability distribution that appears in  $\mathbf{Y}_s$ . In this process, the mean estimated from  $\mathbf{Y}_t$  is considered to be a true mean. i.e.  $E(Y) = \widehat{M}_t$ . Hat indicates the random estimator of a parameter. We assume that the training mean  $\widehat{\mathbf{M}}_t$  is already subtracted from the two data matrices  $\mathbf{Y}_t$  and  $\mathbf{Y}_s$  in the rest of this subsection.

With  $p_a \ll p$  approximating eigenvectors  $\widehat{\mathbf{Y}}_{at}$  estimated from a training set  $\mathbf{Y}_t$ , the component scores  $\mathbf{Y}_{cs|t}$  of  $\mathbf{Y}_s$  on  $\widehat{\mathbf{Y}}_{at}$  are  $\mathbf{Y}_s \widehat{\mathbf{Y}}_a$ . Then, a multivariate multiple regression model can be formulated by treating the test data matrix  $\mathbf{Y}_s$  as responses and the component scores

$\mathbf{Y}_{cs|t}$  of the test data as predictors, i.e.,  $\mathbf{X} = \mathbf{Y}_{cs|t} = \mathbf{Y}_s \hat{\mathbf{\Gamma}}_{at}$  with

$$\mathbf{Y}_s = \mathbf{X}\mathbf{B}_{as|t} + \mathbf{E}. \quad (3.2)$$

The inverse of  $\mathbf{X}'\mathbf{X}$  exists when the rank of the data matrix  $\mathbf{Y}_s$  is at least  $p_a$ . The least squares estimates of  $\mathbf{B}_{as|t}$  and the responses  $\mathbf{Y}_s$  are respectively

$$\begin{aligned} \hat{\mathbf{B}}_{as|t} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}_s = (\hat{\mathbf{\Gamma}}'_{at} \mathbf{Y}'_s \mathbf{Y}_s \hat{\mathbf{\Gamma}}_{at})^{-1} \hat{\mathbf{\Gamma}}'_{at} \mathbf{Y}'_s \mathbf{Y}_s \quad \text{and} \\ \hat{\mathbf{Y}}_s &= \mathbf{X}\hat{\mathbf{B}}_{as|t} = \mathbf{Y}_s \hat{\mathbf{\Gamma}}_{at} (\hat{\mathbf{\Gamma}}'_{at} \mathbf{Y}'_s \mathbf{Y}_s \hat{\mathbf{\Gamma}}_{at})^{-1} \hat{\mathbf{\Gamma}}'_{at} \mathbf{Y}'_s \mathbf{Y}_s. \end{aligned} \quad (3.3)$$

When  $\mathbf{Y}_t = \mathbf{Y}_s$ , that is, when we take the training data as responses (thus omitting subscripts  $t, s$  in the derivation below),  $\hat{\mathbf{B}}_{as|t}$  and  $\hat{\mathbf{Y}}_s$  can be simplified as follows:

$$\begin{aligned} \hat{\mathbf{B}}_{as|t} &= (\hat{\mathbf{\Gamma}}'_a \mathbf{Y}' \mathbf{Y} \hat{\mathbf{\Gamma}}_a)^{-1} \hat{\mathbf{\Gamma}}'_a \mathbf{Y}' \mathbf{Y} = (\hat{\mathbf{\Gamma}}'_a (N-1) \hat{\mathbf{\Sigma}} \hat{\mathbf{\Gamma}}_a)^{-1} \hat{\mathbf{\Gamma}}'_a \mathbf{Y}' \mathbf{Y} \\ &= (\hat{\mathbf{\Gamma}}'_a (N-1) \hat{\mathbf{\Gamma}} D(\hat{\boldsymbol{\lambda}}) \hat{\mathbf{\Gamma}} \hat{\mathbf{\Gamma}}_a)^{-1} \hat{\mathbf{\Gamma}}'_a (N-1) \hat{\mathbf{\Gamma}} D(\hat{\boldsymbol{\lambda}}) \hat{\mathbf{\Gamma}} = D(\boldsymbol{\lambda}_a)^{-1} D(\hat{\boldsymbol{\lambda}}_a) \hat{\mathbf{\Gamma}}'_a \\ &= \hat{\mathbf{\Gamma}}'_a, \\ \hat{\mathbf{Y}}_s &= \mathbf{X}\hat{\mathbf{B}}_{as|t} = \mathbf{Y} \hat{\mathbf{\Gamma}}_a \hat{\mathbf{\Gamma}}'_a, \end{aligned}$$

which is the usual approximation of the data as described in Eq. (3.2) for zero mean. Our goal is to measure the association between the estimates  $\hat{\mathbf{Y}}_s$  of the test data set in Eq. (3.3) and the test data  $\mathbf{Y}_s$  itself.

### 3.2.2 Measure of Association: Second Moment Accuracy

To measure the association between  $\hat{\mathbf{Y}}_s$  and  $\mathbf{Y}_s$ , we follow the approach suggested in [Muller, 2007]. Let  $\mathbf{S}_h$  be the sample covariance matrix of  $\hat{\mathbf{Y}}_s$  and  $\mathbf{S}_y$  be the sample covariance matrix of  $\mathbf{Y}_s$ . In the multivariate linear regression model,  $\mathbf{S}_h$  is the covariance under the regression, and  $\text{tr}(\mathbf{S}_h)$  is the amount of variance explained by the regression [ $\text{tr}(\mathbf{X})$  indicates the trace of  $\mathbf{X}$ ].  $\mathbf{S}_y$  is divided into two parts:  $\mathbf{S}_y = \mathbf{S}_h + \mathbf{S}_e$ .  $\text{tr}(\mathbf{S}_y)$  is the amount of total variance, and  $\text{tr}(\mathbf{S}_e)$  is the amount of unexplained variance left in  $\mathbf{Y}_s$ . In other words,  $\text{tr}(\mathbf{S}_e)$  represents the amount of variance in  $\mathbf{Y}_s$  that remains after accounting for the linear effect of  $\mathbf{X}$ .

Our goodness of prediction measure is the ratio of variation explained according to the principle of goodness of fit. A univariate approach to repeated measure of goodness of fit  $\hat{\tau}$  (when  $t = s$ ) is given by

$$\hat{\tau} = \frac{\text{tr}(\mathbf{S}_h)}{\text{tr}(\mathbf{S}_h + \mathbf{S}_e)}, \quad (3.4)$$

and is equivalent to the proportion of generalized variance controlled as shown below:

$$\hat{\tau} = \frac{\text{tr} \left[ \hat{\mathbf{Y}}_a \mathbf{D}(\hat{\boldsymbol{\lambda}}_a) \hat{\mathbf{Y}}_a' \right]}{\text{tr} \left[ \hat{\mathbf{Y}}_a \mathbf{D}(\hat{\boldsymbol{\lambda}}_a) \hat{\mathbf{Y}}_a' + \hat{\mathbf{Y}}_b \mathbf{D}(\hat{\boldsymbol{\lambda}}_b) \hat{\mathbf{Y}}_b' \right]} = \frac{\sum_{k=1}^{p_a} \hat{\lambda}_k}{\sum_{k=1}^p \hat{\lambda}_k}.$$

The property of decomposition of the total variation into  $\mathbf{S}_e$  and  $\mathbf{S}_h$  makes this measure very attractive as it now can be interpreted as the amount of variation explained by the retained directions, whereas  $\mathbf{S}_e$  measures the magnitude of the remaining variation.

In general, a goodness of fit test  $\hat{\tau}$  can be calculated as

$$\hat{\tau} = \frac{\text{tr} \left( (\hat{\mathbf{Y}} - \hat{\mathbf{M}})' (\hat{\mathbf{Y}} - \hat{\mathbf{M}}) \right)}{\text{tr} \left( (\mathbf{Y} - \hat{\mathbf{M}})' (\mathbf{Y} - \hat{\mathbf{M}}) \right)} = \frac{\sum_{i=1}^N (\hat{Y}_i - \hat{\mu})^2}{\sum_{i=1}^N (Y_i - \hat{\mu})^2}. \quad (3.5)$$

Our goodness of prediction  $\rho^2$ , a measure of association, is derived from the goodness of fit by applying  $\hat{\tau}$  to the proposed regression model (3.2) (when  $t \neq s$ ) and can be written as follows:

$$\hat{\rho}^2 = \frac{\text{tr} \left( (\hat{\mathbf{Y}}_s - \hat{\mathbf{M}}_t)' (\hat{\mathbf{Y}}_s - \hat{\mathbf{M}}_t) \right)}{\text{tr} \left( (\mathbf{Y}_s - \hat{\mathbf{M}}_t)' (\mathbf{Y}_s - \hat{\mathbf{M}}_t) \right)} = \frac{\sum_{i=1}^N (\hat{Y}_{si} - \hat{\mu}_t)^2}{\sum_{i=1}^N (Y_{si} - \hat{\mu}_t)^2}, \quad (3.6)$$

where  $\hat{\mu}_t$  is the sample mean estimated from a training set. We can break the numerator of (3.6) into two parts as follows:

$$\begin{aligned} (\hat{Y}_{si} - \hat{\mu}_t)^2 &= (\hat{Y}_{si} - \hat{\mu}_s)^2 + (\hat{\mu}_s - \hat{\mu}_t)^2 \\ &\quad + 2(\hat{Y}_{si} - \hat{\mu}_s)(\hat{\mu}_s - \hat{\mu}_t), \end{aligned} \quad (3.7)$$

$$\begin{aligned}
\sum_{i=1}^N (\widehat{Y}_{si} - \widehat{\mu}_t)^2 &= \sum_{i=1}^N (\widehat{Y}_{si} - \widehat{\mu}_s)^2 + \sum_{i=1}^N (\widehat{\mu}_s - \widehat{\mu}_t)^2 \\
&= \text{tr}(\mathbf{S}_h) + N(\widehat{\mu}_s - \widehat{\mu}_t)^2,
\end{aligned}$$

where  $\widehat{\mu}_s$  is the sample mean estimated from a test set. The cross product term of (3.7) disappears after summation since  $\sum_{i=1}^N (\widehat{Y}_{si} - \widehat{\mu}_s) = \sum_{i=1}^N \widehat{Y}_{si} - N\widehat{\mu}_s = 0$  and  $\widehat{\mu}_s - \widehat{\mu}_t$  is constant. Similarly, its denominator is decomposed into two parts as follows:

$$\begin{aligned}
\sum_{i=1}^N (Y_{si} - \widehat{\mu}_t)^2 &= \sum_{i=1}^N (Y_{si} - \widehat{\mu}_s)^2 + \sum_{i=1}^N (\widehat{\mu}_s - \widehat{\mu}_t)^2 \\
&= \text{tr}(\mathbf{S}_y) + N(\widehat{\mu}_s - \widehat{\mu}_t)^2 \\
&= \text{tr}(\mathbf{S}_h + \mathbf{S}_e) + N(\widehat{\mu}_s - \widehat{\mu}_t)^2.
\end{aligned}$$

Then  $\widehat{\rho}^2$  becomes

$$\begin{aligned}
\widehat{\rho}^2 &= \frac{\sum_{i=1}^N (\widehat{Y}_{si} - \widehat{\mu}_s)^2 + N(\widehat{\mu}_s - \widehat{\mu}_t)^2}{\sum_{i=1}^N (Y_{si} - \widehat{\mu}_s)^2 + N(\widehat{\mu}_s - \widehat{\mu}_t)^2} \\
&= \frac{\text{tr}(\mathbf{S}_h) + N(\widehat{\mu}_s - \widehat{\mu}_t)^2}{\text{tr}(\mathbf{S}_h + \mathbf{S}_e) + N(\widehat{\mu}_s - \widehat{\mu}_t)^2}.
\end{aligned} \tag{3.8}$$

The reason we choose to evaluate the deviation of both  $\widehat{Y}_{si}$  and  $Y_{si}$  from the mean  $\widehat{\mu}_t$  estimated from a training set instead of the mean  $\widehat{\mu}_s$  estimated from a test set is to be true to the applications of statistical shape models. In the applications of statistical shape models,  $\widehat{\mu}_t$  becomes a template for a new object since there is no way of estimating the mean of objects that are not included in the training set. We can still interpret  $\widehat{\rho}^2$  as the amount of variation of a test set explained by the retained principal directions estimated by a training set as long as the mean estimated from a training set is close to the mean estimated from a test set, that is,  $\widehat{\mu}_s \approx \widehat{\mu}_t$ .

$\rho^2$  has a value that is between 0 and 1. High values of  $\rho^2$  indicate that the retained modes of shape variation estimated from a training set capture the shape variation of new models well because the amount of total variance explained by the factors in the regression model yields a high value of  $\text{tr}(\mathbf{S}_h)$  as a proportion of the fixed total variation. On the other hand, the estimated modes of shape variation that explain less shape variation of new models give

lower  $\rho^2$ .

The theoretical distribution of  $\rho^2$  can be of special interest for further analysis of the measure of association between  $\widehat{\mathbf{Y}}_s$  and  $\mathbf{Y}_s$ : whether  $\rho^2$  has a unimodal or bimodal distribution; whether its mean or median gives a better summary statistic; whether the distribution is symmetric or skewed. However, we leave this topic as future research.

### 3.2.3 Procedure for Iterative Calculation of $\rho^2$

The goal is to analyze the predictive power of statistical shape models as the size of training sample changes. For each training sample size we calculate quartiles of the distribution of  $\rho^2$ , as follows.

**INPUT:** A sample pool  $\mathcal{P}$  of  $N$  objects of  $p$  geometric features

**SETTING:** Set the following parameters:

- 1) The test sample size  $\alpha$
- 2) The list  $\mathcal{L}$  of  $nt$  training sample sizes
 
$$\mathcal{L} = \{\beta_i | \beta_1 < \dots < \beta_{nt}, \beta_{nt} + \alpha \leq N, i = 1, \dots, nt\}$$
- 3) The number of retained principal directions  $p_a$
- 4) The number of repetitions  $R$

**OUTPUT:**  $\rho^2$  values calculated for  $R$  times at each training sample size  $\beta_i$ .  $R \times nt$  number of  $\rho^2$  values are computed.

**PROCEDURE:**

For  $i = 1, \dots, nt$

For repetitions =  $1, \dots, R$

Step1 Randomly select two disjoint sets: a test set  $\mathcal{S}_s$  of size  $\alpha$ ,  
a training set  $\mathcal{S}_t$  of size  $\beta_i$

Step2 Compute  $\widehat{\boldsymbol{\mu}}_t$ , eigenvectors  $\widehat{\mathbf{Y}}_{at}$  from  $\mathcal{S}_t$

Step3 Construct three  $\alpha \times p$  matrices: a data matrix  $\mathbf{Y}_s$  from  $\mathcal{S}_s$   
an estimate of the response matrix  $\widehat{\mathbf{Y}}_s$  in (3.3), and a mean matrix  $\widehat{\mathbf{M}}_t$ ,

Step4 Compute  $\rho^2$  using (3.6)

The reason that  $\mathcal{S}_s$  and  $\mathcal{S}_t$  must be disjoint is to reduce the sampling bias.

We use a box plot to visualize the output  $\rho^2$  values of the procedure. As illustrated in Fig. 3.3, a box plot has lines at the lower quartile (25th percentile), median (50th percentile), and upper quartile values (75th percentile). Whiskers extend from each end of the box to the adjacent values in the data; by default, the maximum whisker length is 1.5 times of the interquartile range. (The difference between the upper and lower quartiles is called the interquartile range.) Outliers, displayed with a ‘+’ sign, are data with values beyond the ends of the whiskers.

### 3.3 Derivation of B-reps from M-reps

Given an m-rep figure, a subdivision surface method [Thall, 2004] is presently used to generate a smooth object surface. Thall [2004] modified the Catmull–Clark subdivision surface algorithm to interpolate the boundary positions and the normals implied by the spokes.

Deriving a b-rep from an m-rep involves calculating the spoke ends of the m-rep. That is, for the spokes  $\mathbf{S}^{+1/-1}$  in each medial atom of the m-rep, its spoke ends are computed as  $\mathbf{b}^{+1/-1} = \mathbf{p} + r\mathbf{U}^{+1/-1}$  (Fig. 2.3-middle). The crest spoke ends of the end medial atoms are computed as  $\mathbf{b}^0 = \mathbf{p} + \eta r\mathbf{U}^0$  (Fig. 2.3-right). Although boundary points other than spoke end points can be sampled from the surface, we decided to choose only the spoke and bisector spoke end surface points from medial atoms so as not to add redundant dimensions to the derived b-rep. Thus, if an m-rep has  $n$  interior atoms and  $m$  end atoms, the dimension of the corresponding b-rep will be  $6 \times n + 9 \times m$  while that of the m-rep is  $8 \times n + 9 \times m$ . These two representations are not equivalent in the sense that m-reps cannot be constructed from b-reps. B-reps lack necessary nonlinear information of normal directions at the spoke ends to compute the hub positions of the medial atoms. In spite of the inequality, we use the b-rep as the linear representation corresponding to the m-rep.

### 3.4 Application of $\rho^2$ to Models in Linear Space

We tested the goodness of the prediction measure (3.6) through the procedure described in section 3.2.3 on two data sets. One set is made up of synthetic objects, b-reps of simulated

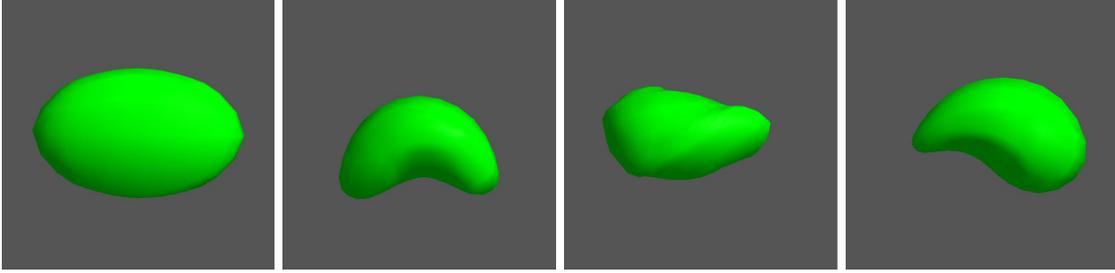


Figure 3.1: From left to right, a base ellipsoid m-rep, randomly bent, twisted, and tapered ellipsoid m-reps are shown reflecting the nonlinear variation in the population.

ellipsoid m-reps. The other set is made up of real anatomical objects, b-reps of m-reps fitted to right hippocampi. We began with the synthetic ellipsoid data because it allows us to generate as many samples as we want and control the kind of deformations in the samples, thus providing a means of checking properties of the  $\rho^2$  such as convergence.

The simulation of ellipsoid m-reps and the experimental results are described in sections 3.4.1 and 3.4.2. The training of right hippocampus binaries and the experimental results are described in the following sections 3.4.3 and 3.4.4.

### 3.4.1 Simulated Ellipsoid M-reps

We have a simulation program [Han et al., 2007] for generating random ellipsoid deformations as illustrated in Fig. 3.1. The program applies a composition of random bending, random twisting, and random tapering to a base ellipsoid m-rep  $\mathbf{M}_0$  sampled from the Blum medial axis of a standard ellipsoid centered at the origin.

Starting from the base ellipsoid m-rep  $\mathbf{M}_0 = \{\mathbf{p}_i, r_i, \mathbf{U}_i^{+1/-1} \mid i = 1, \dots, N\}$ , where  $\mathbf{p}_i = (x_i, y_i, z_i)$  and  $N$  is the number of medial atoms of  $\mathbf{M}_0$ , the three deformations are applied to the medial atoms of  $\mathbf{M}_0$  in the order of bending, twisting, and tapering.

1. Bending: each atom is translated by  $\delta|x_i|^2$  along the  $z$ -axis, and then rotated around the  $y$ -axis by the angle between  $(1, 0, 0)$  and the tangent vector  $(1, 0, 2\delta|x_i|)$ ;
2. Twisting: each atom is rotated around the tangent vector  $(1, 0, 2\delta x_i)$  of a parabola  $(x, 0, \delta x^2)$  at  $x_i$ . The rotation angle is  $\epsilon x_i$ ;
3. Tapering: the radius  $r_i$  is scaled by a factor of  $e^{\zeta x_i}$ , where  $|x_i|$  is the distance from the

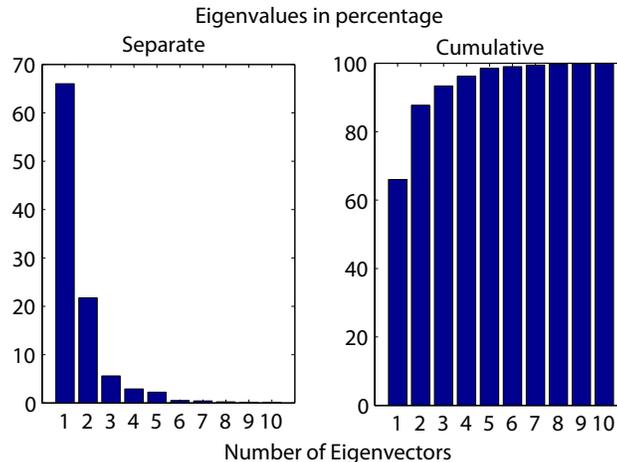


Figure 3.2: Two bar graphs of the first 10 eigenvalues in percentage estimated from 5000 simulated warped ellipsoids b-reps. Left: each eigenvalue per mode. Right: cumulative eigenvalues.

center of the ellipsoid to each atom along the  $x$ -axis,;

where  $\delta, \epsilon, \zeta$  are three independent random variables following Gaussian distributions with zero means. Each set of  $(\delta_j, \epsilon_j, \zeta_j)$ , sampled independently from Gaussian distributions, determines a deformed ellipsoid m-rep  $\mathbf{M}_j$  where  $j$  is the index to the series of deformed ellipsoid m-reps.

### 3.4.2 Experiments on Simulated Ellipsoid B-reps

The base ellipsoid m-rep  $\mathbf{M}_0$  consists of a  $3 \times 7$  grid of medial atoms, where 16 of them are end atoms and 5 of them are internal atoms. The dimension of the m-rep feature vector is  $16 * 9 + 5 * 8 = 184$ , and that of its corresponding b-rep feature vector is  $16 * 9 + 5 * 6 = 174$ . The radial lengths of the principal axes of the base ellipsoid  $\mathbf{M}_0$  are  $(0.2625, 0.1575, 0.1181)$  with a ratio of 10:6:4.5. The three parameters  $\delta, \epsilon, \zeta$  were sampled from three independent Gaussian distributions of standard deviations 1.5, 1.047, and 2.12 respectively. We generated 5000 warped ellipsoid m-reps and made warped ellipsoid b-reps from those m-reps.

Fig. 3.2 shows the variances of the first ten modes of variation estimated from the 5000 simulated warped ellipsoid b-reps. As there are three independent transformations - bending, twisting, and tapering - applied to the base ellipsoid, PCA on the b-reps of 174 feature dimensions produces three major principal eigenmodes with two trailing eigenmodes. The first three principal eigenmodes explain more than 90% of the total variance.

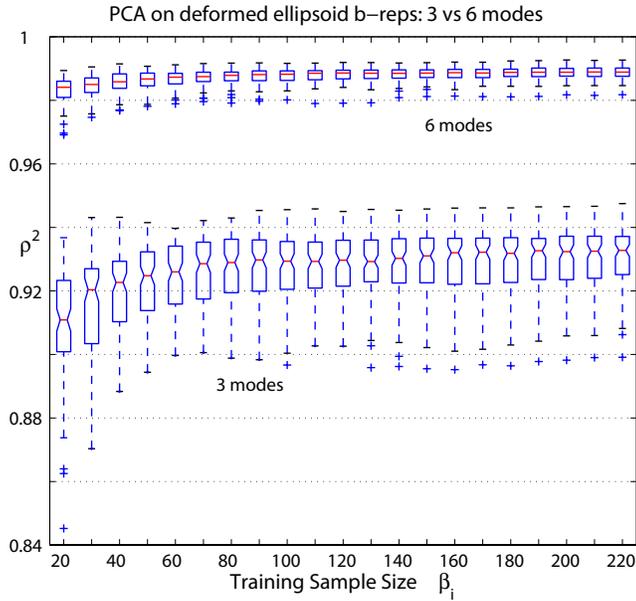


Figure 3.3: Box plots of  $\rho^2$  vs. training sample sizes  $\beta_i$  for PCA on warped ellipsoid b-reps with 3 and 6 eigenmodes. 100 independent trials  $R$  for each training sample size  $\beta_i$  from 20 to 220 were done using a fixed test sample of size  $\alpha = 100$ .

Fig. 3.3 shows two box plots of  $\rho^2$  where the test set size  $\alpha$  is 100. One box plot is for the probability distribution captured by 3 principal modes, and the other is for that captured by 6 principal modes. For each of these box plots the training sample sizes  $\beta_i$  range from 20 to 220 with the increment of 10. A training sample of size  $\beta_i$  and a test sample of size  $\alpha$  are randomly drawn 100 times.  $\rho^2$  is computed at each draw.

As expected, the values of  $\rho^2$  are higher at 6 modes than at 3 modes.  $\rho^2$  reaches near convergence at approximately 80 training samples for 3 modes and at about 60 training samples for 6 modes. Since there are only three true deformations in these synthetic data, we can see that  $\rho^2$  converges at a training size much smaller than the feature space dimension. We can also see that  $\rho^2$  values in the box plot (Fig. 3.3) correspond to the cumulative estimated eigenvalues in Fig. 3.2. The median  $\rho^2$  starts around 0.91 and converges to around 0.93 for 3 modes, and the median  $\rho^2$  starts a little bit above 0.98 and converges to around 0.99 for 6 modes. Moreover, the range of  $\rho^2$  values (interquartile range) is more spread out at 3 modes than at 6 modes, which indicates that the variation of new cases is captured more stably by 6 modes than by 3 modes.

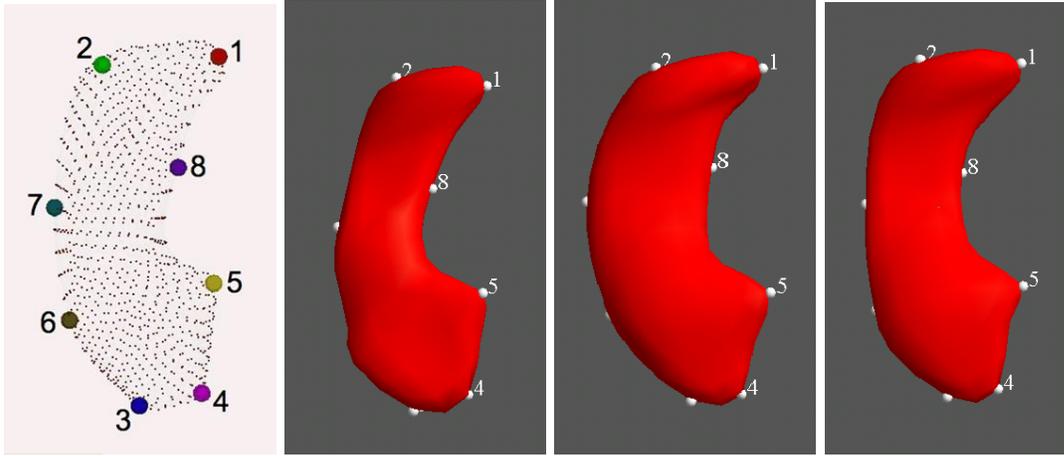


Figure 3.4: From left to right, landmark points of hippocampus and three fitted m-reps are shown.

### 3.4.3 Right Hippocampus M-reps

We ran the procedure on real anatomical object models, 290 right hippocampi b-reps. To have such a large number of samples, we pooled manually segmented binaries of hippocampi from many people regardless of their ages, their mental conditions, or their medications. An m-rep template was fitted into the binaries to extract hippocampus m-rep models through several steps.

Given 8 landmarks per binary (Fig. 3.4-left), we followed steps described as follows to fit hippocampus binaries.

*Hippocampi Fitting Steps:*

- Step1 Fit an initial template model to binaries with high weight on  $F_{lmk}$ ,  $F_{reg}$ , and  $F_{leg}$ , resulting in roughly fitted models.
- Step2 Discard bad fits resulting from Step1.  
Train a mean model and a shape space with the remaining fits.  
Fit the mean model by deforming over the shape space using the same configuration as Step1.
- Step3 Refine the fits from Step2 by deforming each medial atom separately with low weight on  $F_{lmk}$ ,  $F_{reg}$ , and zero weight on  $F_{leg}$ .

The fitted m-reps are used by PGA to produce the final shape space.

Step1 produces preliminary models that are roughly fitted to the binaries. High weight

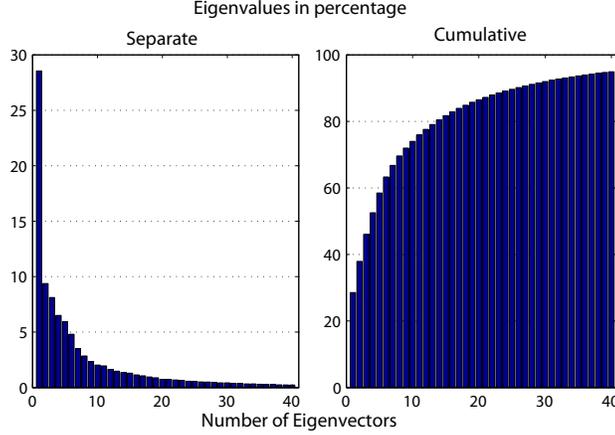


Figure 3.5: Two bar graphs of the first 40 eigenvalues in percentage estimated from 290 right hippocampus b-reps. Left: each eigenvalue per mode. Right: cumulative eigenvalues.

on  $F_{reg}$  and  $F_{leg}$  prevent irregular deformations of the initial template model and the folding of its implied surface while high weight on  $F_{lmk}$  keeps the landmarks of the deforming model and their corresponding points in the binary image together. Step2 repeats the fitting process with a mean model and a shape space trained from the resulting fits from Step1. The mean model usually provides a better initial template, and the shape space provides constraints for the mean model to deform toward the object in the binary. Step3 refines the local fits of the models from Step2 by letting each medial atom freely deform one-by-one.

### 3.4.4 Experiments on Right Hippocampus B-reps

The hippocampus m-rep consists of a  $3 \times 8$  grid of medial atoms, where 18 of them are end atoms and 7 of them are internal atoms. The dimension of m-reps is  $18 * 9 + 6 * 8 = 210$ , and that of the corresponding b-reps is  $18 * 9 + 6 * 6 = 198$ . Three panels on the right side of Fig. 3.4 show the boundary surfaces of the 3 fitted right hippocampus m-reps.

Fig. 3.5 shows the variances of the first 40 modes of variation estimated from the 290 right hippocampus b-reps. Unlike the simulated warped ellipsoids, PCA on the right hippocampus produces principal eigenmodes of slowly decreasing variances. It takes more than 30 modes to reach 90% of the total variance.

Fig. 3.6 shows two box plots of  $\rho^2$  when  $\alpha = 100$  and  $R = 100$ . One box plot is for the probability distribution captured by 18 principal modes, and the other is for that captured by

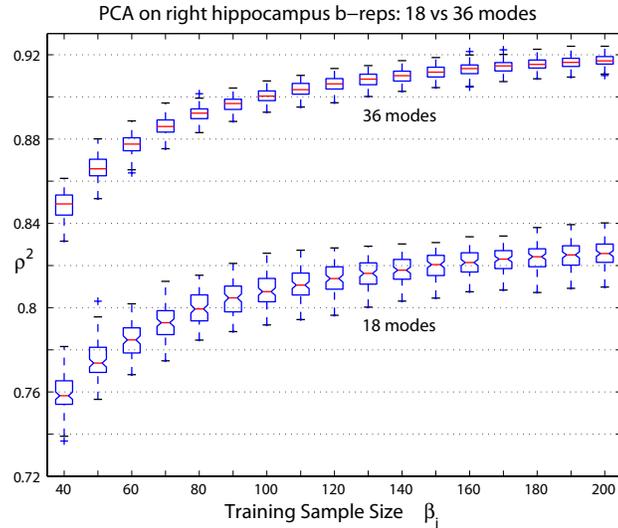


Figure 3.6: Box plots of  $\rho^2$  vs. training sample sizes  $\beta_i$  for PCA on right hippocampus b-reps with 18 and 36 eigenmodes. 100 independent trials for each training sample size were done using a fixed test sample of size  $\alpha = 90$ .

36 principal modes. The training sample sizes  $\beta_i$  range from 40 to 200 with the increment of 10.

As expected, the values of  $\rho^2$  are higher at 36 modes than at 18 modes. The interquartile range of  $\rho^2$  values is slightly more spread out at 18 modes than at 36 modes as well, which indicates that the estimated shape subspace captures the shape variation in the population more stably at 36 modes than at 18 modes. As for the convergence of  $\rho^2$ , it is hard to determine from Fig. 3.6 the training sample size at which  $\rho^2$  begins to converge within the range of the training sample size tested. However, we can judge the trade-off between the training sample size and the increase of  $\rho^2$  value from Fig. 3.6. For example, we can see that for 18 modes about 100 samples are enough to estimate the shape variation since  $\rho^2$  value increases very slowly after  $\beta_i = 100$ .

### 3.5 Distance Measures for $\rho^2$ Evaluation

It is important to assess the validity of our new measure,  $\rho^2$ . While it measures the closeness of estimated populations to the real population in the feature space, it is desirable to verify whether the measure's indications in the feature space concur with what happens in the am-

bient space. To that end, in this section we present other measures based on the standard surface-to-surface distance in the ambient space. With both the correlation measure and the distance measure as functions of training sample sizes, the relation between the predictive power measured in the shape feature space and in the ambient space is analyzed.

For a shape model  $Y$ , let  $B(Y)$  be the set of vertices of its corresponding b-rep in triangular 3D meshes. Let  $S(Y)$  be its surface. The mean absolute surface-to-surface distance between two shape models  $Y_i$  and  $Y_j$  is defined as follows [Aspert et al., 2002]:

$$d_{mad}(Y_i, Y_j) = \frac{1}{N_i + N_j} \left( \sum_{v \in B(Y_i)} \min_{p' \in S(Y_j)} \|v - p'\| + \sum_{v' \in B(Y_j)} \min_{p \in S(Y_i)} \|v' - p\| \right),$$

where  $N_i$  and  $N_j$  indicate the numbers of points in  $B(Y_i)$  and  $B(Y_j)$  respectively, and  $\|*\|$  represents the usual Euclidean norm. The surface-to-surface distances from  $S(Y_i)$  to  $S(Y_j)$  and from  $S(Y_j)$  to  $S(Y_i)$  are averaged since they are not equal.

We compute two kinds of surface-to-surface distance measures. One measure is the minimum of all squared mean absolute distances between a model  $Y$  in a test set  $\mathcal{S}_s$  and all models in a training set  $\mathcal{S}_t$ :

$$d_m^2(Y, \mathcal{S}_t) = \min_{Y' \in \mathcal{S}_s} (d_{mad}^2(Y, Y')) .$$

Another measure is the squared mean absolute distance between a model  $Y$  in a test set  $\mathcal{S}_s$  and its projection  $Y_a$ , Eq. (3.2) on the shape space estimated from a training set  $\mathcal{S}_t$ :

$$d_p^2(Y, \mathcal{S}_t) = d_{mad}^2(Y, Y_a) .$$

We compute these two distance measures following the same procedure for  $\rho^2$  computation described in section 3.2.3. In applying the procedure, we need a summary statistic for these distance measures to see their change as the training sample size increases, since at each iteration  $d_m^2$  and  $d_p^2$  are computed for every model in a test set. The output of the procedure at the end of each iteration is  $\alpha$  numbers of  $d_m^2$  and  $d_p^2$  values. So, we use the median as a

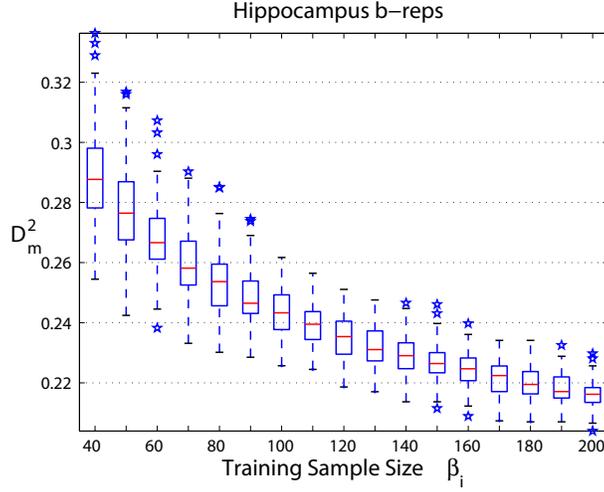


Figure 3.7: A box plot of  $D_m^2$  vs. training sample sizes  $\beta_i$  with the same setting as in section 3.4.4.

summary statistic:

$$D_*^2(\mathcal{S}_t, \mathcal{S}_s) = \text{median}_{Y \in \mathcal{S}_s} (d_*^2(Y, \mathcal{S}_t)) ,$$

where  $*$  indicates subscripts  $m$  or  $p$  for  $d_m^2$  or  $d_p^2$  respectively.

$D_p^2$  is more interesting to us than  $D_m^2$ .  $D_m^2$  indicates how close the test models are to training models and depends only on the choice and the size of  $\mathcal{S}_t$ . On the other hand,  $D_p^2$  indicates the ability of the estimated shape space to approximate a new model in its ambient space.  $D_p^2$  depends not only on the choice and the size of  $\mathcal{S}_t$  but also on the number of retained principal directions from  $\mathcal{S}_t$ .

### 3.5.1 Application of $D_*^2$ to Right Hippocampus B-reps

We tested the two distance measures on the right hippocampus b-reps. Figs. 3.7 and 3.8 show box plots for  $D_m^2$  and  $D_p^2$  respectively. Since  $D_m^2$  is independent of the number of retained eigenmodes, Fig. 3.7 has only one box plot of  $D_m^2$ . As the training sample size increases, we can see that the values and the interquartile range of  $D_m^2$  decreases. Fig. 3.8 clearly indicates that with more principal modes and larger training samples we get more accurate approximations of new instances in the population, which is consistent with what we have observed in the  $\rho^2$  plot (Fig. 3.6). The tests with  $D_*^2$  distance measures provide empirical evidence that our

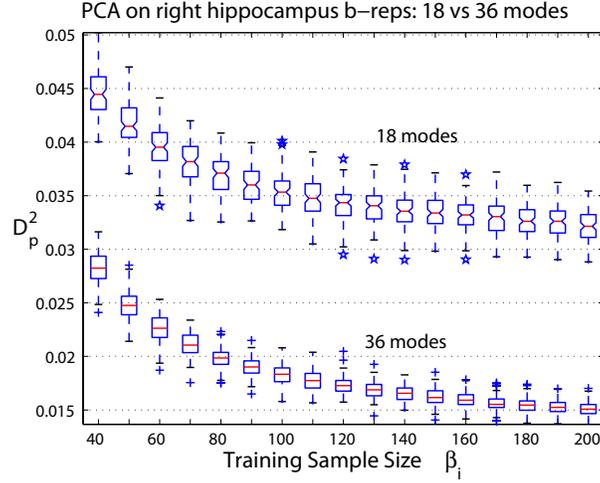


Figure 3.8: Two box plots of  $D_p^2$  vs. training sample sizes  $\beta_i$  for PCA on right hippocampus b-reps with 18 and 36 eigenmodes with the same setting as in section 3.4.4.

proposed goodness of prediction measure defined in the shape feature space reflects the shape variation appearing in the ambient space.

## 3.6 Goodness of Prediction $\rho_d^2$ for Curved Manifolds

### 3.6.1 Two Possible Extensions of $\rho^2$

Our goodness of prediction measure (3.6) (equivalently (3.8)) does not directly apply to models such as m-reps that live in a nonlinear curved manifold. However, as already indicated in the equation (3.6), the numerator and the denominator of (3.6) can be interpreted via distances from an estimated training mean  $\hat{\mu}_t$ . The following two equations simply show rewriting of (3.6) and (3.8) in terms of a more general metric function  $d$ :

1. from (3.6)

$$\hat{\rho}_d^2 = \frac{\sum_{i=1}^N d^2(\hat{Y}_{si}, \hat{\mu}_t)}{\sum_{i=1}^N d^2(Y_{si}, \hat{\mu}_t)}, \quad (3.9)$$

2. from (3.8)

$$\hat{\rho}_d^2 = \frac{\sum_{i=1}^N d^2(\hat{Y}_{si}, \hat{\mu}_s) + N d^2(\hat{\mu}_s, \hat{\mu}_t)}{\sum_{i=1}^N d^2(Y_{si}, \hat{\mu}_s) + N d^2(\hat{\mu}_s, \hat{\mu}_t)}. \quad (3.10)$$

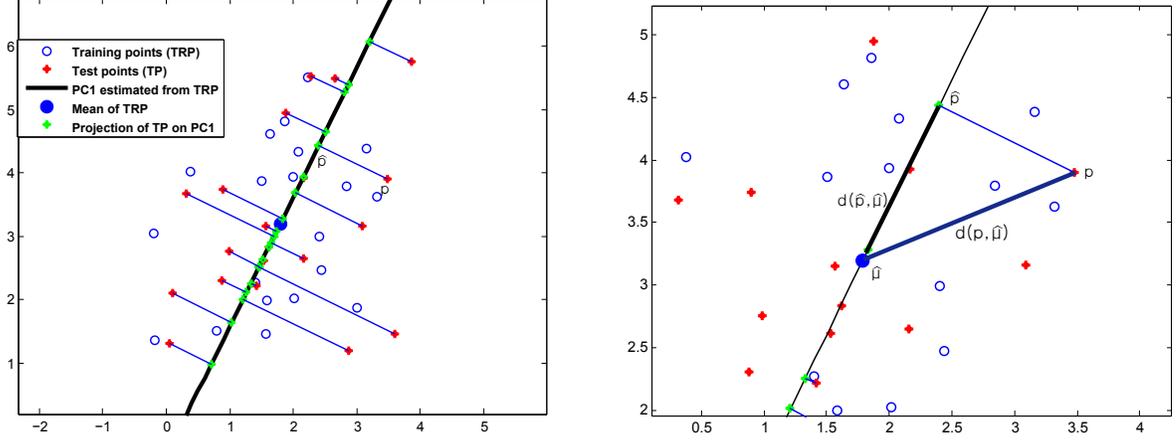


Figure 3.9: Left: Two sets of 20 random vectors chosen from a multivariate normal distribution. Hollow blue dots indicate training points from which the mean and the first principal direction are estimated. Red dots indicate test points projected onto the first principal direction going through the estimated mean. Green dots on the first principal directors are projections of test points (red dots), that is, approximations of test points with the training mean and the first principal component. Right: Zoomed plot of figure on the left. This plot shows the distances in the numerator and denominator of the correlation measure formula for one test point  $\mathbf{P}$ .

These two expressions for  $\hat{\rho}_d^2$  show a natural extension of the distance decomposition of the total variance  $\mathbf{S}_y$  into  $\mathbf{S}_h$  and  $\mathbf{S}_e$ . However, the equality of these two expressions, which holds for linear feature spaces, does not strictly hold for nonlinear spaces. Also, for nonlinear spaces the numerator and the denominator in equation (3.9) cannot be interpreted as  $\text{tr}(\mathbf{S}_h) + N(\hat{\mu}_s - \hat{\mu}_t)^2$  and  $\text{tr}(\mathbf{S}_h + \mathbf{S}_e) + N(\hat{\mu}_s - \hat{\mu}_t)^2$  respectively, as they can for linear spaces (equations (3.6) and (3.8)).

### 3.6.2 $\rho_d^2$ for Nonlinear Shape Models

(3.9) and (3.10) provide the two possible extensions of  $\rho^2$  to a curved manifold suggested in section 3.6.1. Recall that (3.9) and (3.10) are not equivalent in the curved manifold. We choose to use (3.9) as the goodness of prediction for nonlinear shape models in the curved manifold because it has a nice interpretation in the tangent space: the geodesic distance  $d^2(Y, \hat{\mu}_t)$  for  $Y \in M$  is equal to  $\| \text{Log}_{\hat{\mu}_t}(Y) \|$  in  $T_{\hat{\mu}_t}M$ . Also, the decomposition of the total variance  $\mathbf{S}_y$  holds in the tangent space  $T_{\hat{\mu}_t}M$ .

Fig. 3.9 shows the graphical view of the equation (3.9). One set of points indicates a

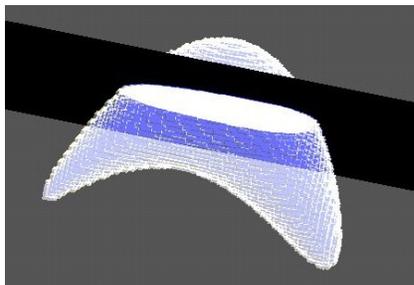


Figure 3.10: A  $(128 \times 128 \times 128)$  bent and tapered binary ellipsoid

training set, and the other set of points indicates a test set. The line is the first principal direction going through a mean estimated from the training set. In the subspace (line in this example) spanned by the first principal direction, points in a test set are approximated by their projections on the first principal direction. The denominator in equation (3.9) is the sum of the distances from the training mean to each point in a test set, and the numerator is the sum of the distances from the training mean to projections of points in a test set onto the subspace.

## 3.7 Application of $\rho_a^2$ on Models in Nonlinear Space

### 3.7.1 Deformed Binary Ellipsoids

A synthetic test population was created from an ellipsoid deformed from the original ellipsoid  $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} \leq 1$  by a set of 1000 diffeomorphisms of the form [Han et al., 2007]

$$\Psi_{\delta, \epsilon, \zeta}(x, y, z) \equiv \begin{bmatrix} x \\ e^{\zeta x}(y \cos(\epsilon x) + z \sin(\epsilon x)) \\ e^{\zeta x}(y \cos(\epsilon x) + z \sin(\epsilon x) + \delta x^2) \end{bmatrix},$$

where  $\delta$ ,  $\epsilon$ , and  $\zeta$  are parameters to control bending, twisting, and tapering respectively.  $\delta$ ,  $\epsilon$ , and  $\zeta$  follow  $N(0, (1.5)^2)$ ,  $N(0, (1.047)^2)$ , and  $N(0, (2.12)^2)$  respectively, where  $N(\mu, \sigma^2)$  is the normal distribution with mean  $\mu$  and standard deviation  $\sigma$ . In this experiment, we again used a standard ellipsoid with axis lengths of  $(0.2625, 0.1575, 0.1181)$  centered at the origin, i.e.,  $a = 0.2625$ ,  $b = 0.1575$ ,  $c = 0.1181$ . These parameters were sampled 1000 times from the three

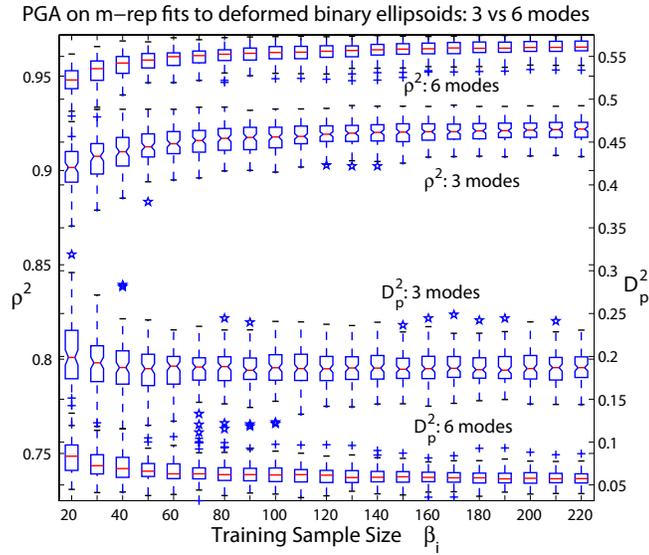


Figure 3.11: Two box plots of  $\rho_d^2$  vs. training sample sizes  $\beta_i$  for PGA on m-rep fits to deformed binary ellipsoids with 3 and 6 eigenmodes (left y-axis). 100 independent trials for each training sample size were done using a fixed test sample of size  $\alpha = 100$ . Two box plots of  $D_p^2$  vs. training sample sizes  $\beta_i$  with the same setting (right y-axis).

normal distributions, and the resulting deformations were applied to the standard ellipsoid. The results were 1000 ( $128 \times 128 \times 128$ ) binary images of warped ellipsoids. Fig. 3.10 shows a case of deformed ellipsoid binary.

The deformed binary ellipsoids are different from the deformed ellipsoid m-reps described in section 3.4.1. The difference lies in whether the deformations are applied to medial atoms of the base ellipsoid m-rep or to the ambient space of the base ellipsoid.

The fitting of m-reps to the binary ellipsoids follows the same steps taken for fitting the right hippocampus (section 3.4.3). Six landmarks are used for the fitting: two end points of three ellipsoid axes.

### 3.7.2 Experiment on M-rep Fits to Deformed Binary Ellipsoids

The fitted ellipsoid m-reps consist of a  $3 \times 7$  grid of medial atoms. The settings of the procedure here are the same as the settings for the simulated warped ellipsoid b-reps.

The results here are consistent with those for the simulated warped ellipsoid b-reps. As shown in Fig. 3.11, the values of  $\rho_d^2$  are higher at 6 modes than at 3 modes, and the values of

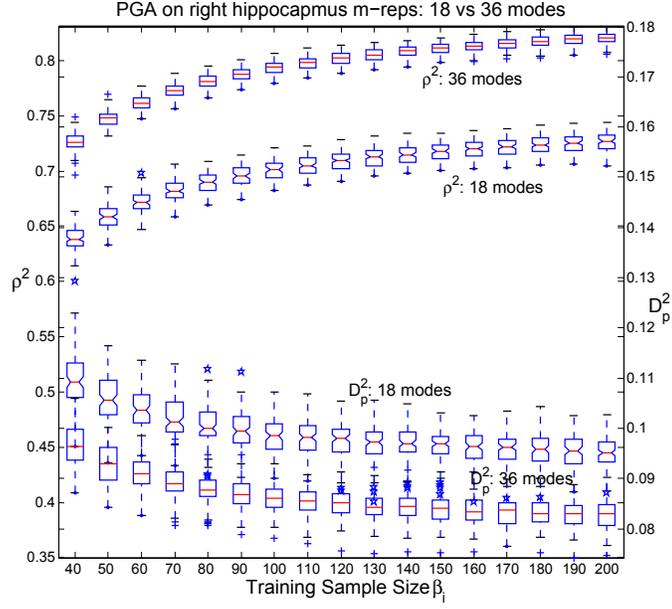


Figure 3.12: Two box plots of  $\rho_d^2$  vs. training sample sizes  $\beta_i$  for PGA on m-rep fits to right hippocampus with 18 and 36 eigenmodes (left  $y$ -axis). 100 independent trials for each training sample size were done using a fixed test sample of size  $\alpha = 90$ . Two box plots of  $D_p^2$  vs. training sample sizes  $\beta_i$  with the same setting (right  $y$ -axis).

$D_p^2$  are smaller at 6 modes than at 3 modes. Both measures begin to converge at approximately 60 training samples. The convergence starts at the training size much smaller than the feature space dimension since there are only 3 true deformations in this population of synthetic data. The interquartile range of  $\rho^2$  and  $D_p^2$  values is more spread out at 3 modes than at 6 modes, as well.

### 3.7.3 Experiment on Right Hippocampus M-rep

$\rho_d^2$  was tested through the procedure (section 3.2.3) on the 290 right hippocampus m-reps (section 3.4.3). Fig. 3.12 shows the results of the procedure as two box plots of  $\rho_d^2$  and  $D_p^2$  when  $\alpha = 90$  and  $R = 100$ . One box plot is for the probability distribution captured by 18 principal modes, and the other is for that captured by 36 principal modes. The training sample sizes  $\beta_i$  range from 40 to 200 with the increment of 10.

The results of  $\rho_d^2$  on the right hippocampus m-reps are very similar to the results of  $\rho^2$  on the corresponding b-reps. The values of  $\rho_d^2$  are higher at 36 modes than at 18 modes and the

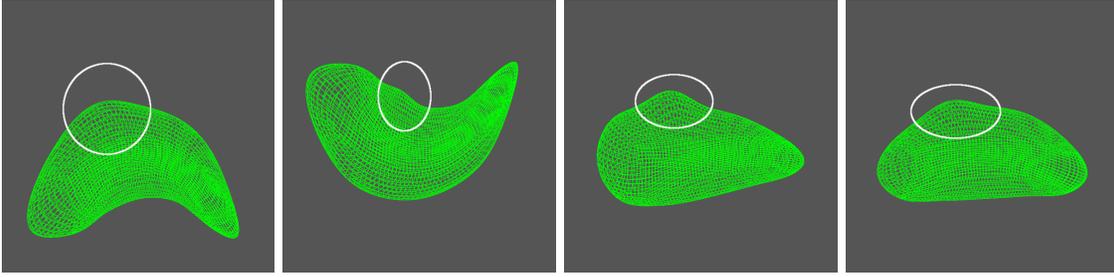


Figure 3.13: Four simulated ellipsoids with local deformation (bump - circle) (This figure is the courtesy of Xiaoxiao Liu and appears in [Liu et al., 2008].)

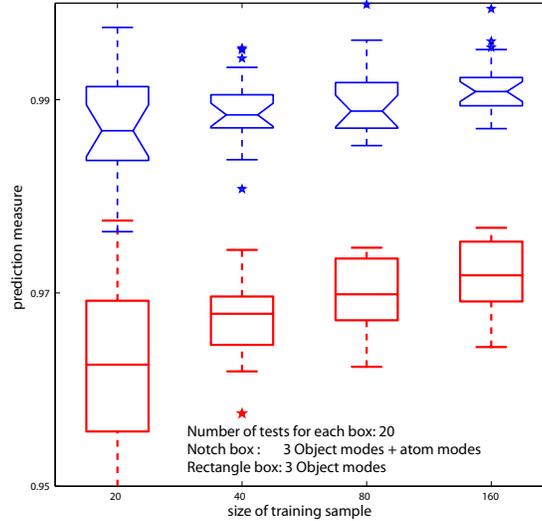


Figure 3.14: Two box plots of  $\rho_d^2$  vs. training sample sizes  $\beta_i$  for multi-scale PGA on simulated ellipsoid m-reps with local deformation with 3 object modes followed by 4 atom modes.

values of  $D_p^2$  are smaller at 36 modes than at 18 modes. It is still difficult to determine when  $\rho_d^2$  and  $D_p^2$  converge within the range of the training sample size tested.

### 3.7.4 Evaluation of a Coarse-to-fine Shape Prior

In work led by Xiaoxiao Liu with significant collaboration by me, our goodness of prediction measure  $\rho_d^2$  has also proven to be effective in the evaluation of the multi-scale shape priors. Liu et al. [2008] proposed a coarse-to-fine shape prior for the probabilistic segmentation to enable local refinement in the m-rep framework, which aims to capture the small level of detail components of the object shape variation that PCA-based approximations are likely to miss, as pointed out in [Nain et al., 2005] and [Davatzikos et al., 2003]. The approach developed

in [Liu et al., 2008] is to decompose the variation space into two scale levels: object-scale and atom-scale prior. As the object-scale prior describes the shape changes of the object as a whole, the atom-scale prior captures the small level of detail components of the shape changes that are left from the object-scale shape space. The atom-scale prior is constructed using the residual space, that is, the remainders of the actual shape space from the shape subspace described by the object-scale prior.

To show the robustness of the estimated coarse-to-fine shape prior, Liu et al. [2008] adopted our procedure and viewed  $\rho_d^2$  against training sample sizes.  $\widehat{Y}_{si}$  in (3.9) becomes the multi-scale projection that refines the object-scale approximation of  $Y_{si}$  by adding each atom-scale approximation in the residual space.

The two scale shape priors were first tested on a synthetic data set of 1000 warped ellipsoids m-reps. These model ellipsoids shown in Fig. 3.13 were produced by applying a relatively small amount of local perturbation on the hub position of one selected atom on top of the three global deformations described in section 3.4.1.

Fig. 3.14 shows the two box plots of  $\rho_d^2$  computed using the object-scale shape prior vs the coarse-to-fine shape prior. One box plot shows the  $\rho_d^2$  values for the shape prior captured by three object-scale principal modes, and the other shows those for the multi-scale shape prior captured by three object-scale principal modes followed by a few eigenmodes of the selected atom.  $\rho_d^2$  clearly indicates the improvement of the predictability when the coarse-to-fine shape prior is used.

Liu et al. [2008] then examined the advantage of the coarse-to-fine shape prior on simulated real anatomical structures. The simulation took 51 eigenmodes estimated from the 290 well-fitted right hippocampus m-reps described in section 3.4.3 and did Gaussian random sampling on the 51 eigenmodes. The 51 eigenmodes explain 95% of the total variation observed in the 290 fitted right hippocampus m-reps. With this data, the object-scale shape prior (the 33 object eigenmodes) was compared with the coarse-to-fine shape prior (the 10 object eigenmodes followed by the 9 atom eigenmodes) using the  $\rho_d^2$  and the volume overlap measure. The Dice Similarity Coefficient [Crum et al., 2006] on the volumes of the test models and the corresponding object-scale and atom-scale approximations was used as the volume overlap measure. Fig. 3.15 shows the comparison.

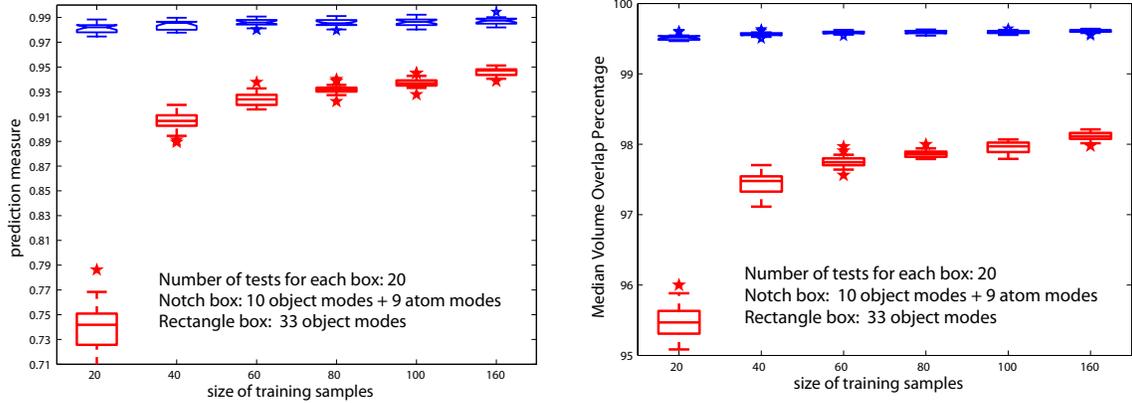


Figure 3.15: Left: Two box plots of  $\rho_d^2$  vs. training sample sizes  $\beta_i$  for multi-scale PGA on simulated hippocampus m-reps with 10 object modes followed by 9 atom modes. Right: Corresponding box plots of volume overlap between the test models and their projected models.

We can clearly see not only the benefit of the coarse-to-fine shape prior over the object shape prior by  $\rho_d^2$  but also the consistency between the volume overlap measure and the  $\rho_d^2$  measure. In addition to the distance measures we introduced in section 3.5 to show the validity of our  $\rho_d^2$  measures in ambient space, this consistency compared the volume overlap measure and the  $\rho_d^2$  measure, and confirms that our goodness of prediction measure  $\rho_d^2$  defined in the feature space does indeed reflect what happens in the ambient space.

### 3.8 Conclusion & Discussion

Our work has been motivated by the need to have a quantitative measure to evaluate the predictive power of a statistical shape model. We proposed a novel statistical correlation measure  $\rho^2$  called the goodness of prediction. It is designed to judge the predictive power of a PCA-based statistical shape model to analyze the major shape variation observed in the training sample. The measure is formally derived by interpreting PCA in terms of the multivariate linear regression model, and it is interpreted as a ratio of the variation of new data explained by the retained principal directions estimated from training data.

The major shift of our perspective in evaluating a statistical shape model is that we are more concerned about evaluating how well an estimated shape model fits the new data rather than how well an estimated shape model fits the parameters of the population probability

distribution. It is a reasonable stance to take considering that the ability of an estimated shape probability distribution to describe a new object is the major concern in most applications of statistical shape models.

The novelty of the measurement lies in its being computed using two disjoint sets of samples. One set of samples is used for the estimation of a shape space by PCA, and the other set of samples is used for the evaluation of the estimated shape space, which exactly reflects the situation happening in the applications of statistical shape models.

Moreover, we proposed a procedure to compute the goodness of prediction against the training sample sizes, which allows inferring the training sample size that ensures capturing a certain amount of shape variation present in objects unseen from training samples. The procedure was experimentally evaluated on synthetic warped ellipsoid b-reps and real anatomical right hippocampus b-reps. The results were visualized using box plots that show the median and the interquartile range of the  $\rho^2$  values of a large number of independent trials.

We tested a slight variation of the proposed procedure although we did not report the results in this paper. We further aligned each training set to tighten the distribution before doing PCA (Step2 in section 3.2.3) since we were concerned that the use of pre-aligned data might bring a bias in our estimation of  $\rho^2$ . This additional alignment on the right hippocampus data did not bring any noticeable difference in  $\rho^2$  values. Thus we did not include this additional alignment step in the procedure, concluding that the pre-alignment on the pooled data is sufficient to remove any non-shape related transformations in the models.

We extended  $\rho^2$  for linear shape representations to  $\rho_d^2$  for nonlinear shape representations that form Riemannian symmetric spaces.  $\rho_d^2$  carries the same statistical meaning as  $\rho^2$  because the geodesic distance from a training mean in a Riemannian space is equivalent to the Euclidean distance in a corresponding tangent space at the training mean.  $\rho_d^2$  was also tested on the synthetic ellipsoid m-reps and real anatomical right hippocampus m-reps. The results of  $\rho_d^2$  on these m-rep data are consistent with those of  $\rho^2$  on the corresponding b-rep data.

The  $\rho_d^2$  measure was also empirically verified by two surface distance measures and a volume overlap measure to prove that  $\rho_d^2$  really reflects what happens in the ambient space where the model lies.

Our experiments showed the usefulness and the versatility of the procedure. It yields an

appropriate training sample size for some retained number of eigenvalues and judging the trade-off between training sample size and the amount of variation explained by the retained number of eigenvalues. However, the major drawback of the proposed procedure is that it needs many data samples to see the convergence of median  $\rho_d^2$  values and of the interquartile range of  $\rho_d^2$  values with respect to the training sample size. Another issue with the procedure is the number of independent trials  $R$ . In Bayesian statistics,  $R$  must be several thousand to have a statistical significance, which is impractical in many applications.

In spite of these disadvantages, the measure itself is easy and fast to compute, and its statistical interpretation is simple to understand. It is also quite flexible to apply for any statistical shape model. We already showed in other work that applying the measure to evaluate multi-scale shape priors is straightforward. We hope this measure will be found to be useful in evaluating other statistical shape models and will motivate further exploration in the evaluation of statistical shape models.

## Chapter 4

# Multi-Object Statistical Shape Models<sup>1</sup>

Statistical shape models have been proven very effective in a number of applications, including image segmentation and characterization of the anatomic differences between the classes of normal and diseased patients. In segmentation, a shape prior restricts the deformation of the shape model within the variations learned from training data in this optimization process. In characterization of anatomic differences shape statistics provide the basis for a test of a null hypothesis that the probability densities for the two classes are the same. Obtaining an accurate estimation of shape statistics is thus essential.

Since multiple objects form a given anatomic region, there has been a desire to characterize probabilistically populations of multi-object anatomic geometry. Pizer et al. [2005a] makes the case that probabilities on populations of geometric regions are an essential part of multi-scale probabilities on geometric-model-based image intensities. In the schema described there regions of space with predictable intensity histograms are placed in a neighbor relationship, and this is done at a number of different discrete scale levels. In the work of Lu et al. [2007], three scale levels are under consideration: that global to the image, that consisting of only objects without interstitial regions, and that consisting of certain through-object subsections. Here the main focus is on the critical issue of how to produce probability distributions that reflect not only region (object) shape but also inter-object geometric relationships.

As explained in section 1.1.1, the approaches to estimate probability distributions of multi-object anatomic geometry that have been tried so far can be summarized in 3 categories:

---

<sup>1</sup>Much of the material in this chapter has appeared in [Pizer et al., 2005b, Jeong et al., 2006].

a global approach, an object-by-object approach and a hierarchical approach. The global approach concatenates the features of all objects and estimates the probability distribution on the concatenated feature vectors. The object-by-object approach treats each object separately. The hierarchical approach estimates multi-scale shape probability distributions from the scale of the multi-object complex to the scale of the individual objects or even to the smaller scale of the individual objects. Each approach has some drawbacks which make a comparison difficult to resolve.

The global approach fails to capture the local variation of an object itself. Moreover, the global approach is likely to give misleading information about inter-object relations when the inter-object correspondence is not well established in the geometric representation of each object across the training samples. As pointed out in section 2.2.4 establishing correspondence of an object itself, not to mention the inter-object correspondence, is a difficult problem to solve. This global approach has been applied on a variety of shape representations: PDMs, diffeomorphisms from atlases, and distance functions or their levels sets including m-reps.

The object-by-object approach totally ignores the interaction between objects although the scale level of this approach is appropriate to capture the shape variation of an object itself. The hierarchical approach can be most favorable to describe statistically multi-object anatomic geometry due to its consideration of multiple scales in shape variations [Vaillant and Davatzikos, 1999, Kapur et al., 1998, Davatzikos et al., 2003]. However, this advantage of the hierarchical approach becomes a disadvantage because of the difficulty in deciding an appropriate level of each scale in the shape variation of the multi-object complexes. The weakness of these approaches has led us to borrow the idea of the mixed model approach to handle the inter-object relation in describing variations of each object separately but within the multi-object complex.

In brief, the method based on the mixed model decomposes the shape variation of each of the objects into three components: the mean of the variation from some base state, self variations, and variations due to the inter-object relations. Few attempts have been made to describe the inter-object relations statistically. Pohl et al. [2005] describes inter-relations by representing them via the distance function to objects' boundaries. This approach however does not explicitly separate out inherent variation of object from the inter-object effects.

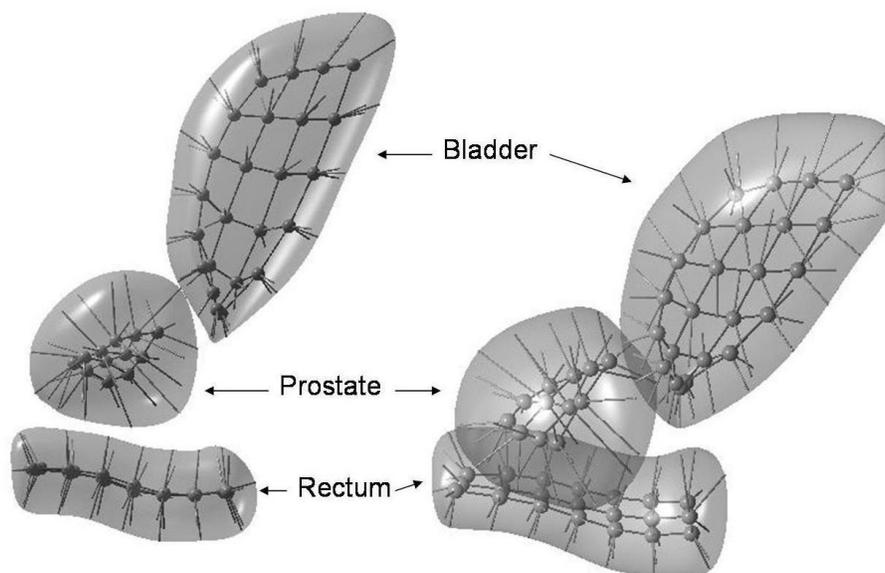


Figure 4.1: An m-rep 3-object complex for the bladder, the prostate, and the rectum of a patient in different view.

The self term describes the variation of the object itself, not affected by other objects in the multi-object complex. Shape variations of an object closely surrounded by other objects are also caused by the other objects, especially its neighboring objects, as well as by the internal changes within the object. The inter-object relation that is taken into account here is restricted to immediate neighboring parts of neighboring objects. A shape variation of an object caused by the neighboring parts is called a neighbor effect which is described as a function of the neighbors' geometric descriptors. The method estimates probability densities on each of these components per object. Section 4.4 details this method based on a mixed model approach.

A single-figure m-rep model [Pizer et al., 2003] is used to represent each object in a multi-object complex, as illustrated by a male pelvic organ model in Fig. 4.1. M-reps provide a fixed topology of the medial locus for a population of an object and the local coordinate system for object interiors that in turn provide correspondences across cases of the object. These properties of m-reps facilitate the characterization of the neighbor relationship among objects, allowing medial atoms in neighboring objects to be understood in terms of medial atom transformations of each other.

Both the self and neighbor probability densities are estimated on the nonlinear manifold by

PGA as described in section 2.4.3.1. The estimation of the neighbor probability distribution relies on two related methods: an augmentation method and a prediction method that are described in sections 4.2.1 and 4.2.2 respectively. The rest of this chapter is organized as follows. Section 4.1.1 explains the concept of the residue that is the means to analyze the inter-object relation. Section 4.1 describes the difference and addition operations through which residues are computed.

Section 4.3 explains a preliminary method to handle inter-object relations. This preliminary approach is a precursor to an iterative method explained in the next section 4.4. The major difference between the two methods is the way that the inter-object relation is handled. The preliminary method handles the inter-object relation in one direction: from the target object to its neighbors. The iterative method is not limited by this directional approach since the iterative method separates out the neighbor effect on each object from the total variation of the object. Section 4.4 details the iterative method to estimate self variations and neighbor effects of multi-objects. Section 4.4.4 explains the process to estimate these probability densities for the male-pelvis data to which the new method is applied.

Section 4.5 discusses the results as well as the work yet to be done.

## 4.1 M-rep Operations

Recall that a medial atom  $\mathbf{m} = (\mathbf{p}, r, \mathbf{U}^{+1}, \mathbf{U}^{-1})$  is defined as an element of the symmetric space  $G = R_3 \times R^+ \times S^2 \times S^2$  where the hub position  $\mathbf{p} \in R_3$ , the spoke length  $r \in R^+$ , and the two unit spoke directions  $\mathbf{u}, \mathbf{v} \in$  the unit sphere  $S^2$ . Let  $\mathbf{R}_{\mathbf{w}}$  represent the rotation along the geodesics in  $S^2$  that moves a point  $\mathbf{w} \in S^2$  to the north pole  $\mathbf{p} = (0, 0, 1) \in S^2$ . For given any two medial atoms  $\mathbf{m}_1, \mathbf{m}_2 \in G$  where  $\mathbf{m}_i = (\mathbf{p}_i, r_i, \mathbf{U}_i^{+1}, \mathbf{U}_i^{-1})$ ,  $i = 1, 2$ , the difference between them can be described as follows:

$$\mathbf{m}_1 \ominus \mathbf{m}_2 := (\mathbf{p}_1 - \mathbf{p}_2, \frac{r_1}{r_2}, \mathbf{R}_{\mathbf{U}_2^{+1}}(\mathbf{U}_1^{+1}), \mathbf{R}_{\mathbf{U}_2^{-1}}(\mathbf{U}_1^{-1})) . \quad (4.1)$$

Its corresponding addition operator  $\oplus$  is thus defined as

$$\mathbf{m} \oplus \Delta \mathbf{m} := (\mathbf{p} + \Delta \mathbf{p}, r \cdot \Delta r, \mathbf{R}_{\mathbf{U}^{+1}}^{-1}(\Delta \mathbf{U}^{+1}), \mathbf{R}_{\mathbf{U}^{-1}}^{-1}(\Delta \mathbf{U}^{-1})) \quad (4.2)$$

for a given  $\mathbf{m} = (\mathbf{p}, r, \mathbf{U}^{+1}, \mathbf{U}^{-1})$  and difference  $\Delta \mathbf{m} = (\Delta \mathbf{p}, \Delta r, \Delta \mathbf{U}^{+1}, \Delta \mathbf{U}^{-1})$ . For two m-reps  $\mathbf{M}_1, \mathbf{M}_2$  that consist of medial atoms, the difference  $\mathbf{M}_1 \ominus \mathbf{M}_2$  relative to the  $\mathbf{M}_2$  coordinates is defined as the collection of the differences of corresponding individual atoms in two m-reps. These operations and their properties are explained in detail in [Lu et al., 2007].

#### 4.1.1 Residues of the Object Variations

The probabilistic analysis of the two methods proceeds object by object in order. After some object has been described probabilistically and its effect has been applied to its neighbor objects, there is a further change in the remaining objects to be described. This further change is called the residue of the remainder objects with respect to the probability distribution on the first. Let  $\mathbf{M} \in \mathcal{M}$  be an m-rep or an m-rep residue of one object fitting a particular training case where  $\mathcal{M}$  is a symmetric space of  $\mathbf{M}$ . Let  $p(\mathbf{M}')$  be a probability distribution on  $\mathbf{M}' \in \mathcal{M}$  describing part of the variation of  $\mathbf{M}$ . Notice that if  $D(p)$  represents the domain of  $p$ , then  $D(p)$  is a submanifold of  $\mathcal{M}$ . Relative to the probability distribution  $p$ ,  $\mathbf{M}^0$ , the closest m-rep to  $\mathbf{M}$  in  $D(p)$ , is

$$\mathbf{M}^0 = \operatorname{argmin}_{\mathbf{M}' \in D(p)} d(\mathbf{M}, \mathbf{M}'), \quad (4.3)$$

where  $d(\mathbf{M}, \mathbf{M}')$  is the geodesic distance on  $\mathcal{M}$ . Then the residue  $\Delta \mathbf{M}$  of  $\mathbf{M}$  with respect to  $p$  can be defined as

$$\Delta \mathbf{M} := \mathbf{M} \ominus \mathbf{M}^0. \quad (4.4)$$

In the two methods to build a statistical shape model for multi-object complexes, a method called prediction is used to produce an approximation to  $\mathbf{M}^0$ . The prediction method is expected to give a reasonably good estimate of  $\mathbf{M}^0$  because the prediction is made on the shape space of  $\mathbf{M}$ . The residue is computed through the difference operation on m-reps described in the previous section.

## 4.2 Inter-Object Relations

There is evidence that atoms in one object that are near another object are most highly correlated with that other object. Ray et al. [2006], Jung and Marron [2008] On the basis of that evidence the inter-relation of a multi-object complex is described using these nearby atoms through the methods of augmentation and prediction. Augmentation pools medial atoms across objects to reflect the mixed effects local to an object (e.g., bladder filling with urine) with neighbor effects (e.g., bladder push).

### 4.2.1 Augmentation

Given a multi-object m-rep of  $l$  objects, i.e.,  $\{\mathbf{M}_k\}_{k=1}^l$  where  $\mathbf{M}_k$  is an ordered set of medial atoms per object, an object's inter-relation with other atoms is dealt with by augmenting highly correlated atoms of its neighboring objects. Let  $\mathbf{T}_i$  be a set of atoms in a target object  $\mathbf{M}_i$  that are located near its neighboring objects, and let  $\mathbf{N}_i$  be a set of atoms in neighboring objects that are close to  $\mathbf{T}_i$ .  $\mathbf{T}_i \subseteq \mathbf{M}_i$ , and  $\mathbf{N}_i \subseteq \cup_{k \neq i} \mathbf{M}_k$ . The “augmented” representation of the  $i$ -th object is then

$$\begin{aligned} \mathbf{U}_i &= \mathbf{T}_i \cup \mathbf{N}_i \\ &= \begin{pmatrix} \mathbf{T}_i \\ \mathbf{N}_i \end{pmatrix}. \end{aligned}$$

$\mathbf{N}_i$  indicates atoms used to predict the neighbors effect on the target object (predicting atoms), and  $\mathbf{T}_i$  indicates atoms in the target object affected by the variation of its neighboring objects (predicted atoms). Fig. 4.2 shows an example of these two sets of atoms selected for bladder and prostate in their multi-object complexes.

As will be described in the next section, this augmentation allows us to predict the changes brought on a target object by the change of the neighboring objects. After this deterministic effect of its neighbor on the target object is taken off from the total variation of the object, what remains in the total variation is considered as the combination of the variable part in the neighbor effect and the self variation of the object. A method called prediction to estimate this

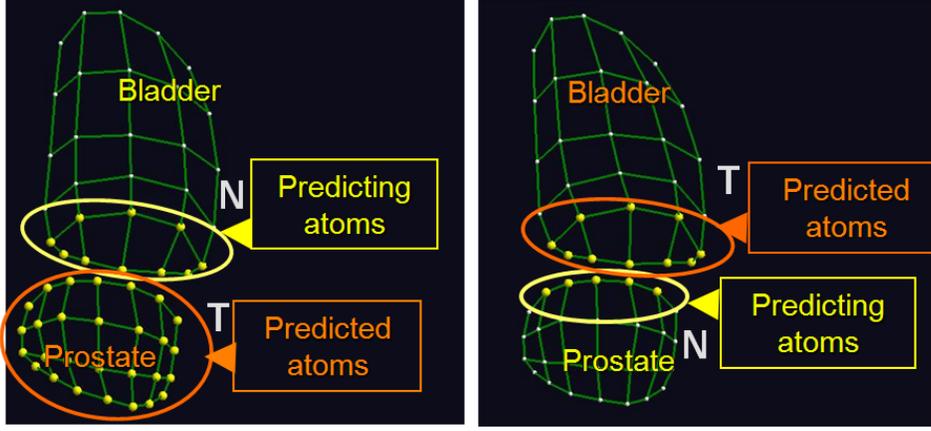


Figure 4.2: Prostate atoms combined with its predicting medial atoms (left). Bladder atoms combined with its predicting medial atoms (right).

deterministic effect is proposed in the next section. The prediction method helps to extract the deterministic effect from its neighbor on the object and to concentrate on the variable part in the neighbor effects. Augmentation is also used in the estimation of the conditional probability of the target object in a multi-object complex described in the next chapter 5.

#### 4.2.2 Prediction

In estimating the probability distribution of  $\mathbf{M}_k$ , prediction reflects changes in  $\mathbf{M}_{i \neq k}$  by predicting how  $\mathbf{M}_k$  bends, twists or warps from the change of  $\mathbf{M}_{i \neq k}$  through its augmenting atoms  $\mathbf{N}_k \subseteq \mathbf{M}_{i \neq k}$  and the predicted atoms  $\mathbf{T}_k \subseteq \mathbf{M}_k$ . In doing so, the shape space of the augmented atoms is taken into account as suggested in [Rajamani et al., 2004], but by using PGA in a nonlinear symmetric space rather than PCA as used in [Rajamani et al., 2004].

Now consider an augmented m-rep figure  $\mathbf{U}_k = (\mathbf{T}_k \cup \mathbf{N}_k)$ . Let  $\mu$  and  $H$  be the mean and the shape space generated by  $h$  principal geodesics in the symmetric space of  $\mathbf{U}_k$ . The deformation of  $\mathbf{U}_k$ , i.e., how  $\mathbf{T}_k$  and  $\mathbf{N}_k$  change together, can be used to predict the effect of the other objects  $\mathbf{M}_{i \neq k}$  on the target object  $\mathbf{M}_k$ .

Let  $\mathbf{U}^* = \begin{pmatrix} \mu|_{T_k} \\ \mathbf{N}_k \end{pmatrix}$ , where  $\mathbf{u}|_A$  indicates taking the elements in the vector  $\mathbf{u}$  that correspond to the subset  $A$  of all the features of the vector  $\mathbf{u}$ .  $\mathbf{U}^*$  is a vector that concatenates the mean of  $\mathbf{T}_k$  and the known variation in  $\mathbf{N}_k$ . Then, the effect on  $\mathbf{T}_k$  is predicted by finding an element

in  $H$  closest to  $\mathbf{U}^*$ , that is,

$$Pred(\mathbf{T}_k) = \underset{\hat{\mathbf{U}}|_{T_k \in H}}{\operatorname{argmin}} d(\mathbf{U}^*, \hat{\mathbf{U}}) \quad (4.5)$$

Prediction is an attempt to find an element of the highest probability in the shape space  $H$  given the known variation in  $\mathbf{N}_k$ . The issues in this definition of prediction will be discussed later in section 4.5.

This prediction can be easily computed by the projection operation [Fletcher, 2004] on the shape space  $H$ . The projection of  $\mathbf{U}^*$  on  $H$  is

$$Proj_H(\mathbf{U}^*) = \exp_{\mu} \left( \sum_{l=1}^h \langle \log_{\mu}(\mathbf{U}^*), v_l \rangle v_l \right), \quad (4.6)$$

where  $\{v_l\}_{l=1}^h$  are principal directions in the tangent space at  $\mu$  corresponding to the principal geodesics in  $H$ .  $Proj_H(\mathbf{U}^*)$  produces also an m-rep. Then the prediction for  $\mathbf{M}_{i \neq k}$  is defined as

$$Pred(\mathbf{T}_k) := Proj_H(\mathbf{U}^*)|_{T_k}. \quad (4.7)$$

The next section describes a preliminary approach to build multi-object shape models on the basis of these methods. In this preliminary approach the objects are assumed to be in an order of decreasing stability, i.e., whose posterior probability, based on both geometric and intensity variability and edge sharpness, are in decreasing levels of tightness. The probability distribution of each object is estimated in this order. Thus, the neighbor relation among objects is not mutual but directional from the most stable object to the least stable object. The direction of prediction is modified accordingly. The details are explained in the next section.

### 4.3 Propagation of Sympathetic Changes

The main idea of this preliminary approach is that if an object changes position, pose, size, or shape, its neighboring objects will change sympathetically. So far only the effect of neighboring objects on the target object has been discussed. However, their interaction is in fact mutual. The same argument about the neighbor effect on the target object holds by switching the

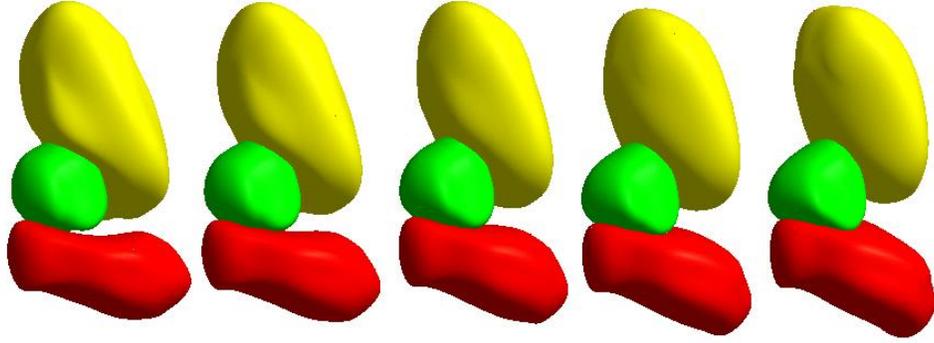


Figure 4.3: Assuming we have produced statistics for the augmented bladder  $\mathbf{U}_1$ , which has augmenting atoms  $\mathbf{N}_1$  in the prostate ( $M_2$ ), the sympathetic change of  $\mathbf{R}'_1$  caused by  $\mathbf{N}_1$  is illustrated.

target object and its neighboring object. This approach considers the inter-object relation in one direction from the target object to its neighboring objects. So the prediction propagates the augmenting atoms' movement in the statistics of one augmented object to the remainder of the objects to be processed. The index  $k$  of a multi-object m-rep  $\{\mathbf{M}_k\}_{k=1}^l$  indicates the order in which each  $\mathbf{M}_k$  is processed.

Let  $\mathbf{R}_k$  be the remainder of the objects i.e.,  $\mathbf{R}_k = \cup_{i>k} \mathbf{M}_i$ . Let  $\mathbf{R}'_k = \mathbf{R}_k \setminus \mathbf{N}_k$ . The probability distribution of  $\mathbf{U}_k$  and  $\mathbf{R}'_k$  can be written as  $p(\mathbf{U}_k, \mathbf{R}'_k) = p(\mathbf{U}_k)p(\mathbf{R}'_k|\mathbf{U}_k)$  by Bayes' theorem.  $p(\mathbf{R}'_k|\mathbf{U}_k)$  is divided into a deterministic prediction from  $\mathbf{U}_k$  and an  $\mathbf{U}_k$ -independent probability on the residue of  $\mathbf{R}'_k$  from that prediction. So the prediction reflects a change in  $\mathbf{M}_k$  in the probability distribution of  $\mathbf{R}_k$  by predicting how  $\mathbf{R}'_k$  bends, twists or warps from the change of  $\mathbf{M}_k$  through augmenting atoms  $\mathbf{N}_k$ .

Thus, the prediction (equation 4.7) is applied not to  $\mathbf{T}_k \subseteq \mathbf{M}_k$  but to  $\mathbf{R}'_k$ , i.e.,

$$Pred(\mathbf{R}'_k) = Proj_H(\mathbf{U}^*)|_{R'_k}, \quad (4.8)$$

where  $\mathbf{U}^* = \begin{pmatrix} \mu|_{R'_k} \\ \mathbf{N}_k \end{pmatrix}$ , and  $\mu$  and  $H$  are the mean and the shape space of  $\mathbf{R}_k$ . Again,  $\mathbf{R}_k$  includes all of the atoms in the other objects whose probability distributions are yet to be determined. The changes in  $\mathbf{N}_k$  will be reflected in sympathetic changes in  $\mathbf{R}'_k$  before the probability distribution on  $\mathbf{R}_k$  is estimated. Fig. 4.3 illustrates the sympathetic change of

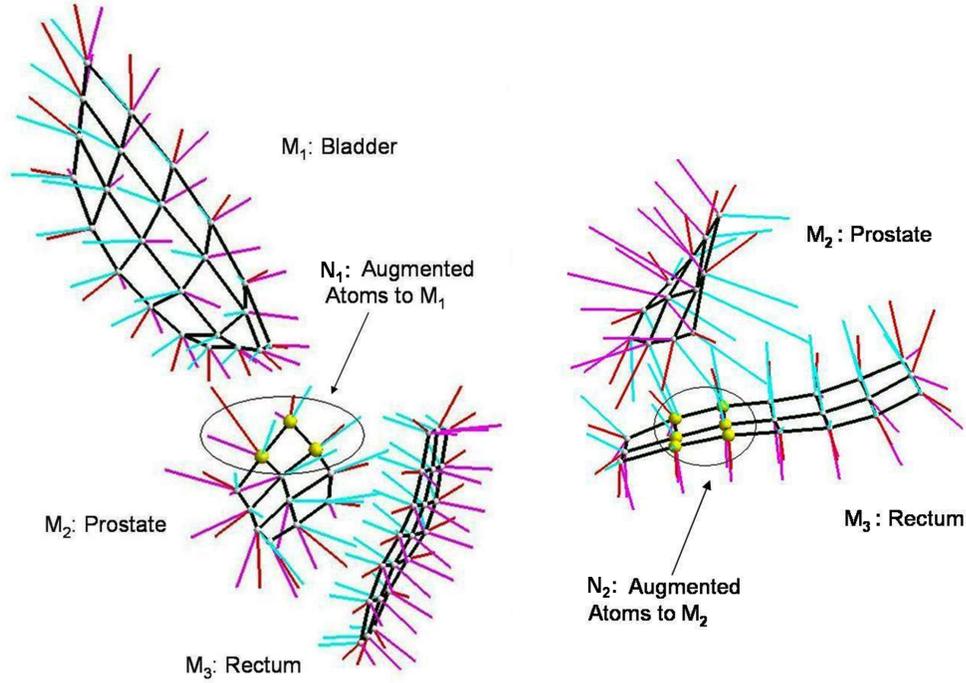


Figure 4.4: A discrete m-rep for the bladder ( $\mathbf{M}_1$ ), the prostate ( $\mathbf{M}_2$ ), the rectum ( $\mathbf{M}_3$ ) 3-object complex of a patient. The augmented atoms in the prostate forming  $\mathbf{N}_1$  are shown with their hub enlarged (left). The prostate ( $\mathbf{M}_2$ ), the rectum ( $\mathbf{M}_3$ ) of the same patient the enlarged atoms in the rectum form  $\mathbf{N}_2$  (right).

prostate and rectum by bladder in a multi-object complex of male-pelvic organs.

In the male-pelvis example shown in Fig. 4.4, medial atoms in bladder  $\mathbf{M}_1$  should be more highly correlated with medial atoms nearby in the prostate than with those in the rest of the prostate or in the rectum.  $\mathbf{N}_1$  indicates the nearby prostate atoms. The representation of the augmented bladder becomes  $\mathbf{U}_1 = \mathbf{M}_1 \cup \mathbf{N}_1$ . As for the prostate medial atoms in the rectum  $\mathbf{M}_3$  that are right below the prostate are more affected by changes in the prostate. Let  $\mathbf{N}_2$  be the nearby rectum atoms. The augmented prostate is then  $\mathbf{U}_2 = \mathbf{M}_2 \cup \mathbf{N}_2$ .

The rest of the section consists of three parts. Section 4.3.1 describes inter-object residues to describe the variation remaining after the propagation of effects from other objects. Section 4.3.2 explains how to train probabilities for objects by successive PGA's on object residues, once the object complexes have already been aligned across the cases, and the medial atoms correspond across the cases.

In this work a geometric model for a complex of non-interpenetrating objects is said to be

*proper* if a) the topology of the objects is retained, b) each object in the model does not have singularities or folds of its boundary or interior, and c) the non-interpenetration of objects is retained within the tolerances appropriate for the scale of the description. Many previous methods for estimating inter-object probability distributions have produced samples some of which are decidedly improper. In section 4.3.3 the method is tested by illustrating that models sampled from the estimated probability distributions on within-patient bladder, prostate, and rectum deformations are nearly proper and that the means and principal modes of variation of these distributions are intuitively reasonable. The application of these ideas to segmentation by posterior optimization is also briefly discussed.

### 4.3.1 Residues of Objects in Order

Once the changes in  $\mathbf{U}_k$  and the sympathetic changes in  $\mathbf{R}'_k$  are described, all that is left to describe statistically is the remaining changes in  $\mathbf{R}_k$  after the sympathetic changes have been removed. If the objects are treated in order and each object has augmenting atoms only in the next object, this will mean that  $l$  probability distributions will need to be trained, namely, for  $\mathbf{U}_1$ , for  $\mathbf{U}_2$  after the sympathetic changes from  $\mathbf{U}_1$  have been removed, ... , for  $\mathbf{U}_l$  after the sympathetic changes from  $\mathbf{U}_1, \mathbf{U}_2, \dots$ , and  $\mathbf{U}_{l-1}$  have been removed. The removal of sympathetic changes is accomplished via the residue idea described in [Lu et al., 2007]. Now the next section describes how to train the probability distribution of each object in a multi-object complex that reflects the propagation of sympathetic changes using the methods of augmentation, prediction, and the concept of the residues.

### 4.3.2 Training the Probability Distribution per Object

Training the probabilities for the object is done via successive PGA's on the object residues. Let  $\mathbf{O}^i$  be a multi-object m-rep residue in case  $i$  from which any truly global variations are removed from  $\{\mathbf{M}_k^i\}_{i \in I, k \in K}$ , where  $I = \{1, \dots, N\}, K = \{1, \dots, l\}$  are index sets for the  $N$  training cases and  $l$  objects. Then  $\mathbf{O}^i = \{\Delta \mathbf{M}_k^i\}_{i \in I, k \in K}$  forms a multi-object m-rep residue of the  $i^{th}$  training case.

The residues  $\{\mathbf{O}^i\}_{i \in I}$  are treated in the order of objects  $\mathbf{M}_k$  from  $k = 1$  to  $n$ . First PGA is applied on  $\{\Delta \mathbf{U}_1^i\}_{i \in I}$ , the residue of the first object, to get the mean  $\mu_1$  and a set of principal

variances and associated principal geodesics  $\{\exp_{\mu_1}(v_1^j)\}_{j=1}^{n_1}$ , where  $v_1^j \in T_{\mu_1}\mathbf{M}_1$ . This mean, principal variances, and principal geodesics provide our estimate of the probability distribution of  $\Delta\mathbf{U}_1$ . Let  $H_1$  be a submanifold of  $\mathbf{M}_1$ , where  $\mathbf{M}_1$  is the symmetric space for  $\Delta\mathbf{U}_1$ . The projection of  $\Delta\mathbf{U}_1^i$  onto the geodesic submanifold  $H_1$ ,  $Proj_{H_1}(\Delta\mathbf{U}_1^i)$ , describes the variation unique to  $\Delta\mathbf{U}_1^i$  in  $H_1$ . Now the residue  $\{\Delta\mathbf{R}_1^i\}_{i \in I}$  is updated to reflect the sympathetic effect from  $\Delta\mathbf{M}_1$  on  $\Delta\mathbf{R}'_1$  by  $\Delta\mathbf{N}_1$ . That is done using the prediction  $Pred(\Delta\mathbf{R}'_1)$  as described in the beginning of this section 4.3.

So the residue for the next object (the second object) is no longer  $\{\mathbf{O}^i\}_{i \in I}$ . The updated residue of the remainder to the first object becomes

$$\Delta^2\mathbf{R}_1^i = \Delta\mathbf{R}_1^i \ominus \begin{pmatrix} Pred(\Delta\mathbf{R}'_1) \\ \Delta\mathbf{N}_1^i \end{pmatrix} \quad i \in I. \quad (4.9)$$

With the new updated residue  $\Delta^{k-1}\mathbf{U}_k^i \subset \Delta^{k-1}\mathbf{R}_k^i$  for the  $k^{th}$  object,  $k = 2, \dots, l$ , the same steps are repeated: 1) applying PGA on  $\Delta^{k-1}\mathbf{U}_k^i$  and 2) updating the residue of the remainder, which produces a set of means  $\{\mu_k\}_{k \in K}$  and sets of principal geodesics  $\{\{\exp_{\mu_k}(v_k^l)\}_{l=1}^{n_k}\}_{k \in K}$  on object residues.

The next section, 4.3.3, describes the study done with the bladder, the prostate and the rectum models in the male pelvis on the probability distributions in the male pelvis that reflect the effect of the deformation of the bladder on the augmenting atoms and the relation of changes in the augmenting atoms  $\mathbf{N}_1$  to these of the rest of the prostate and the rectum.

### 4.3.3 Geometrically Proper Objects

Samples being geometrically improper has been a problem for other methods such as PCA on distance functions or on dense PDMs. Examples of geometrically improper objects are wrong topology, interpenetration of separated objects, folding, and singularities such as unwanted corners and cusps. There are two reasons why the proposed method is expected to avoid geometrically improper samples from their probability distributions.

1) M-reps are founded on the idea that using primitive transformations including local twisting and bending of objects will yield an economical representation of the single and

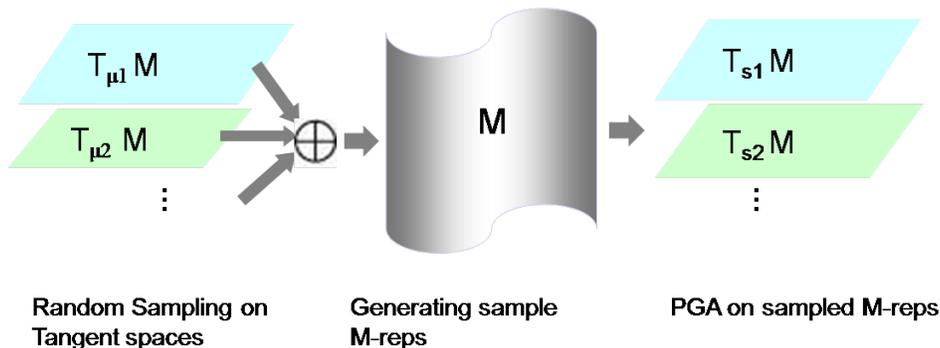


Figure 4.5: Left: tangent spaces at object residue means from real patient data. Middle: m-rep parameter space. Right: object residue means from generated training data. The movie of 100 sampled m-reps from patient 1 and patient 2 data is at <http://midag.cs.unc.edu/pubs/papers/movies/100SamplesPat1and2.avi>. In the movie the point of view changes from time to time.

multi-object transformations of anatomy between individuals or within an individual over time. When using such transformations in the representation methods and in particular in the methods of description of object inter-relations via augmentation and prediction, nonlinear PGA is necessary to produce sample object complexes that are geometrically proper.

2) The regular grids of medial atoms that were generated from training binary images of objects [Merck et al., 2008] are designed to have large geodesic distance to improper entities on the manifold  $\mathcal{M}$ . Thus, it is likely that objects within  $[-2, +2]$  standard deviations will also be geometrically proper. Analysis of these objects using a criterion based on the radial shape operator of [Damon, 2005] could be used to avoid improper models, but this criterion has not yet been applied.

The most basic test of the estimated probability distributions is to visually judge whether those generated samples are proper and whether the principal geodesic directions derived from real patient data explain variations observed in the training samples. Because the training set is just a particular sample subset of a population of m-reps, it would be useful to know how the proposed method would fare on other training sample subsets. This can be accomplished by generating new random samples from the estimated probability distributions and test whether training from these samples produces a probability distribution whose samples are proper.

The new samples are generated by assuming that each tangent plane principal component

from the original training follows the standard normal distribution after the principal directions are scaled by the square root of corresponding eigenvalues in the tangent space. Thus, for each object residue each principal component following the standard normal distribution is randomly sampled to generate random points on each tangent space about the mean  $\{\mu^k\}_{k \in J}$ . By taking exponential maps of those points, m-reps and residues are generated so that they can be combined by  $\oplus$  to produce new training sample m-reps. PGA on such a new sampled training set yields a new mean and set of principal directions and variances, whose samples can be judged as to how proper they are.

The method was applied to obtain the probability distributions from two training sets, each of which was obtained from bone-aligned male-pelvis CT images of a real patient over several days. A single-figure m-rep was fit to each organ:  $4 \times 6$  grids of medial atoms for the bladder,  $3 \times 4$  grids for the prostate, and  $3 \times 7$  grids for the rectum. The total number of medial atoms is 57, so the dimension of the m-rep parameter space is 456. Our software to fit the single figure m-reps to a binary image of each organ provides reasonable correspondence of medial atoms across cases by penalizing irregularity and rewarding correspondence to one case [Merck et al., 2008]. Inter-penetrations among m-reps of the three objects were restricted during the fitting of each training case. One patient (patient 1) had 11 training models (m-reps), and another patient (patient 2) had 17 training models.

Fig. 4.6 displays the first modes of variation of patients 1 and 2 at PGA coefficients -2, -1, 1, 2 standard deviations of bladder with prediction, prostate with prediction and rectum in Fig. 4.6 from the top row to the bottom row.

These movies, as well as the ones seen in Fig. 4.5, show the following. 1) The m-reps produced as samples or chosen along principal geodesics yield limited inter-object penetration, as desired since the training samples have small inter-object penetration. 2) The surfaces of the m-rep sample implied objects are smooth, with few exceptions. Folding is not observed, and the introduction of sharp ridges happens seldom, only at crest positions which are sharp in some of the training cases. 3) The principal geodesics seem to correspond to anatomically observed changes, such as strong growth in the bladder corresponding to filling and strong bulging of the rectum corresponding to the introduction of bowel gas. Also, the prostate residue shows only modest shape changes, a behavior expected from the fact that the prostate is typically quite

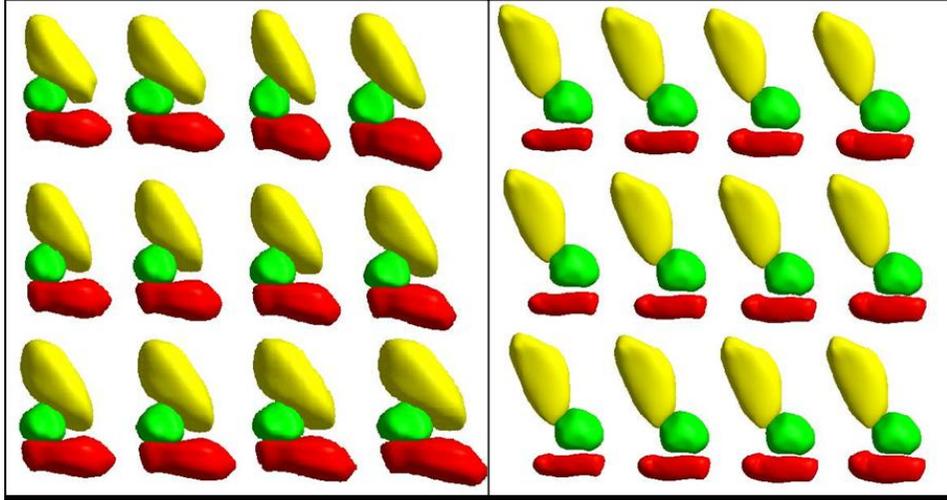


Figure 4.6: Illustration of first modes of variation of patient 1 in the box on the left and that of patient 2 in the box on the right. The movie that shows the first modes of variations of patient 1 and then patient 2 is at <http://midag.cs.unc.edu/pubs/papers/movies/VariationsPat1and2.avi>.

hard. Thus, the proposed method provides samples that are “nearly geometrically proper and means and principal modes of variations that are intuitively reasonable.”

In addition to the evaluation of m-rep probabilities just described, the probabilities can be judged A) on the ability of the PGA to extract known independent random deformations in simulations and B) by their usefulness in segmentation. A) In simulations with compositions of independent random bendings, twistings, and magnifications of an ellipsoid, PGA extracted these basic deformations very well. B) PGA m-rep probabilities trained from images of a given patient on a variety of days were used as the prior in segmentation by posterior optimization of m-reps of the bladder, prostate, rectum complex in target images of the same patient on different days. The details of the application to segmentation are given in [Pizer et al., 2005a] and [Pizer et al., 2005c], object segmentation using histogram statistics is described in [Broadhurst et al., 2005], and the results on a few cases, agreeing well with human segmentations, have been reported in [Chaney et al., 2004]. Briefly, the results are anecdotal but encouraging.

As described in [Pizer et al., 2005a], a schema involving neighboring regions at multiple scales has much to recommend it. At each scale level in this schema, except the global level, a means is needed to produce statistics reflecting region shape and inter-region relations for neighboring regions. In conclusion, the preliminary approach has shown the viability of a

particular method for producing these statistics.

The augmentation and prediction method has also been applied to estimate the probability distribution of m-reps of multi-figure objects, the structure of which is described in [Han et al., 2004]. The hinge atoms are taken as augmented atoms, and the sympathetic change of a subfigure from the change of its host figure is predicted via these augmented atoms [Han et al., 2005].

The ordering of the objects can be avoided by considering the mutual neighbor relation through augmentation. As can be seen in real situations such as the male-pelvis example used in the previous section, not only can the bladder induce a change in the prostate and rectum, but also the change of a prostate can induce sympathetic change in the bladder and rectum, etc. The next section, 4.4, describes an extension of this preliminary method that overcomes the ordering of the objects in estimating the probability distribution of each object in a multi-object complex.

## 4.4 Decomposition of Shape Variations

Two major improvements on the earlier method are 1) an explicit separation between the self variation and the neighbor effect of an object's variation and 2) a mutual neighbor relation through the augmentation and prediction methods in sections 4.2.1 and 4.2.2, allowing each object to have any of the others as neighbors. This mutual neighbor relation is more realistic and is clearly suggested in the male-pelvis data to which the new method is applied.

Two sets of atoms per object are chosen manually: a set of atoms  $\mathbf{T}$  in the target object that are located near its neighboring objects and a set of atoms  $\mathbf{N}$  in neighboring objects that are close to  $\mathbf{T}$ . The augmentation  $\mathbf{U} = \mathbf{T} \cup \mathbf{N}$  allows us to predict the changes brought on a target object by the change of the neighboring objects, i.e., the prediction (Eq. (4.7)) of the effect on  $\mathbf{T}$  conditioned by its neighbor atoms  $\mathbf{N}$ . Then variations from this prediction can be analyzed statistically.

#### 4.4.1 Self Variations & Neighbor Effects

The main idea here is to view the shape variation of an object in terms of two variations: one that is independent of the shape variations of neighboring objects and the other that is dependent on the shape variations of neighboring objects. The first is called the self variation, and the second is called the neighbor effect.

Let  $\Delta\mathbf{M}^i := \mathbf{M}^i \ominus \mathbf{M}^b$  where  $\mathbf{M}^b$  is a base figure in a multi-object base model such as the mean or some reference model, where  $i$  indexes the training cases. All figures in  $\mathbf{M}^i$  are aligned to figures in the base model.  $\{\Delta\mathbf{M}^i\}$  are variations of m-reps from the base m-rep model. The variations are decomposed into two parts as follows:

$$\Delta\mathbf{M} := \Delta\mathbf{M}^s \oplus \Delta\mathbf{T}^n$$

where  $\Delta\mathbf{M}^s$  measures self variation and  $\Delta\mathbf{T}^n$  measures neighbor effects. The neighbor effect term  $\Delta\mathbf{T}^n$  reflects the interaction among objects. Recall that in section 4.2.1  $\mathbf{T}$  is defined to be a set of atoms in  $\mathbf{M}$  that are affected by the variation of  $\mathbf{N}$  in its neighboring objects.

The neighbor effect is modeled by following the idea of the linear mixed model. Davidian [2001] That is,

$$\Delta\mathbf{T}^n := \mathbf{F}(\mathbf{N}) \oplus \mathbf{r} \oplus \mathbf{e},$$

where  $\mathbf{F}(\mathbf{N})$  is the systematic effects from neighbors,  $\mathbf{r}$  is the random effects, and  $\mathbf{e}$  is the vector of errors. The systematic effects from neighbors are also referred to as “the fixed effects”. The fixed effects  $\mathbf{F}(\mathbf{N})$  are estimated by the prediction method (4.7) described in section 4.2.2, i.e.,

$$\mathbf{F}(\mathbf{N}) = \text{Pred}(\Delta\mathbf{T}^n).$$

Thus,  $\Delta\mathbf{T}^n$  in the Eq. (4.4.1) is further subdivided as the prediction  $\text{Pred}(\Delta\mathbf{T}^n)$ , and the neighbor residue,  $\Delta\mathbf{T}^n \ominus \text{Pred}(\Delta\mathbf{T}^n)$ . Note that if  $\mathbf{M}$  is a geometric representation in linear space,

$$\sigma^2(\Delta\mathbf{T}^n) = \sigma^2(\mathbf{r}) + \sigma^2(\mathbf{e}), \tag{4.10}$$

under the normal assumption on  $\mathbf{r}$  and  $\mathbf{e}$ . However, the relation (4.10) does not hold for m-reps

since m-reps are not linear representations.

Several assumptions are made in this approach. First, within each object, the self variations  $\{\Delta\mathbf{M}^s\}$  and the residues from the predictions  $\{\Delta\mathbf{T}^n \ominus \text{Pred}(\Delta\mathbf{T}^n)\}$  are considered to be statistically uncorrelated. Second, the self variations among objects are assumed to be uncorrelated. Third, the neighbor effects among objects are also assumed to be uncorrelated. Fourth, the effect of the neighboring objects is local. Based on the fourth assumption the sets  $\mathbf{N}$ ,  $\mathbf{T}$  and hence  $\mathbf{U}$  are defined. These atoms are chosen currently based on Euclidean distance between atoms in nearby objects. In addition, section 4.4.3 shows the simplification of the joint probability distribution  $p(\mathbf{M}_1, \dots, \mathbf{M}_l)$  of  $l$  objects on the basis of these four assumptions.

The objects' variation decomposed into the self variation and the neighbor effect must sum to the overall difference of the object from its base state. Thus the self variation must be the residue describing overall change from the base state after the neighbor effect is removed, and the neighbor effect must be the residue describing overall change from the base state after the self variation is removed. The two residues are computed by the difference operator on m-reps, and the summation of them is computed by the addition operator on m-reps. The difference operation and its complementary addition operation are described in section 4.1.

#### 4.4.2 An Iterative Method

Beginning with a simple assumption on the separation of each object's  $\Delta\mathbf{M}$  into self and neighbor components, the method refines the separation by repeating the following steps over all figures in the multi-object m-reps. In the following description of the estimation of statistics for an object, the index over the objects and the index over the training cases is skipped.

**Shape Step.** Apply PGA on  $\Delta\mathbf{M} \ominus \widehat{\Delta\mathbf{M}}^n$  which gives the shape space and estimate of the self part of each training case  $\widehat{\Delta\mathbf{M}}^s$ . The hat  $\widehat{(\cdot)}$  indicates the best estimate of either neighbor or self components up to the previous iteration.

#### Neighbor Step.

1) **Augmentation.** Subtract the estimate of the self part from each training case not to

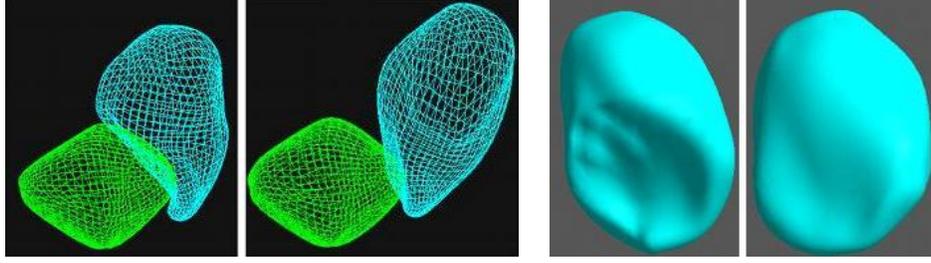


Figure 4.7: These four figures show the prediction of the deformation on the prostate brought by the change of the patient B163. The two objects in the left two panels are the bladder (cyan) and the prostate (green) in wire frame, and the right two panels show separately the bladder of the left two panels in solid. The bladders in the right two panels are respectively in positive and in negative two standard deviations from the mean along an estimated principal mode of the self variation. The left two panels show the predicted prostate corresponding to the two positions of the bladder shown in the right two panels.

corrupt the effect from the neighbor by the effect to other neighboring objects from the object, i.e.,

$$\Delta \mathbf{T}^n := \Delta \mathbf{T} \ominus \widehat{\Delta \mathbf{T}}^s.$$

Then, form an augmented set of differences to predict  $\Delta \mathbf{T}^n$  based on  $\widehat{\Delta \mathbf{N}}^s$ :

$$\Delta \mathbf{U} := \Delta \mathbf{T}^n \cup \widehat{\Delta \mathbf{N}}^s.$$

$\Delta \mathbf{N}^s$  is used rather than  $\Delta \mathbf{N}$  because the initial assumption of local effect leads to  $\Delta \mathbf{T}^n$  of the object and  $\Delta \mathbf{N}^n$  of its neighbors being statistically independent.

**2) Prediction. (1) Predictor function:** Perform PGA on  $\Delta \mathbf{U}$  to find the shape space of the augmented atoms  $\Delta \mathbf{U}$ . The shape space is used to find the deterministic effect from the neighbors as follows:

$$\begin{aligned} Proj_H(\Delta \mathbf{U}^*) &= \exp_\mu \left( \sum_{l=1} \langle \log_\mu(\widehat{\Delta \mathbf{U}}^*), v_l \rangle \cdot v_l \right), \\ Pred(\Delta \mathbf{T}^n) &:= Proj_H(\Delta \mathbf{T}^n)|_T \end{aligned}$$

where  $\Delta \mathbf{U}^* = \begin{pmatrix} \mu|_T \\ \widehat{\Delta \mathbf{T}}^s \end{pmatrix}$ , and  $H$  is the shape space of  $\Delta \mathbf{U}$ ,  $\{v_l\}_{l=1}$  are principal directions in the tangent space of  $H$  at its mean  $\mu$ .

**(2) Updated augmented set:** Now form a newly updated augmented set after removing the prediction from the residue from the self estimate.

$$\begin{aligned}\Delta\mathbf{T}^{rmdr} &= \Delta\mathbf{T}^n \ominus \text{Pred}(\Delta\mathbf{T}^n), \\ \Delta\mathbf{U}' &:= \Delta\mathbf{T}^{rmdr} \cup \widehat{\Delta\mathbf{N}}^s.\end{aligned}$$

Do PGA on the new augmented set  $\Delta\mathbf{U}'$  to obtain the shape space and the estimate  $\widehat{\Delta\mathbf{T}}^n$  of the neighbor part of each training case. As a result, the estimate of the neighbor part comprises the two components: prediction and the estimate of the variation from the prediction.

$$\widehat{\Delta\mathbf{T}}^n = \text{Pred}(\Delta\mathbf{T}^n) \oplus \exp_{\mu} \left( \sum_{k=1} \langle \log_{\mu}(\Delta\mathbf{T}^{rmdr}), v_k \rangle \cdot v_k \right).$$

#### 4.4.3 Joint Probability of Multiple Objects

Using this decomposition of self variations and neighbor effects of multiple objects, the joint p.d.f. of the multiple objects can be interpreted as follows:

$$\begin{aligned}p(\mathbf{M}_1, \dots, \mathbf{M}_l) &= p(\mathbf{M}_1^s, \mathbf{M}_1^n, \dots, \mathbf{M}_l^s, \mathbf{M}_l^n) \\ &= p(\mathbf{M}_1^n, \dots, \mathbf{M}_l^n | \mathbf{M}_1^s, \dots, \mathbf{M}_l^s) \times p(\mathbf{M}_1^s, \dots, \mathbf{M}_l^s),\end{aligned}\tag{4.11}$$

where the first term is the joint conditional probability of the neighbor effects and where the second term is the joint probability of the self variations.  $\Delta$  before  $\mathbf{M}$  will be omitted in this section for clarity of the expressions.

Under the assumption of normality on  $p(\mathbf{M}_1, \dots, \mathbf{M}_l)$  the first three assumptions made on self variations and neighbor effects in section 4.4.1 imply the independence of the self variation and the neighbor effect within an object, the independence of the self variations among objects, and the independence of the neighbor effects among objects. These independence assumptions and the locality assumption simplify both terms in the Eq. (4.11). The joint probability of the

self variations becomes a product of the probabilities of the self variation per object:

$$p(\mathbf{M}_1^s, \dots, \mathbf{M}_l^s) = \prod_{i=1}^{i=l} p(\mathbf{M}_i^s).$$

The joint conditional probability of the neighbor effects becomes also a product of the conditional probabilities of the neighbor effect per object:

$$\begin{aligned} p(\mathbf{M}_1^n, \dots, \mathbf{M}_l^n | \mathbf{M}_1^s, \dots, \mathbf{M}_l^s) &= \prod_{i=1}^{i=l} p(\mathbf{M}_i^n | \mathbf{M}_1^s, \dots, \mathbf{M}_l^s) \\ &= \prod_{i=1}^{i=l} p(\mathbf{M}_i^n | \mathbf{N}_i^s). \end{aligned}$$

Then the joint p.d.f. of the multiple objects can also be simplified into

$$p(\mathbf{M}_1, \dots, \mathbf{M}_l) = \prod_{i=1}^{i=l} p(\mathbf{M}_i^s) \times p(\mathbf{M}_i^n | \mathbf{N}_i^s). \quad (4.12)$$

#### 4.4.4 Shape Prior in MAP-Based Segmentation

In a maximum a posteriori segmentation framework as discussed in section 2.3.2 segmentation finds an object in a given image  $i$  that maximizes the posterior probability:

$$\begin{aligned} \widehat{\mathbf{M}}_i &= \operatorname{argmax}_{\mathbf{M}_i} p(\mathbf{M}_1, \dots, \mathbf{M}_l | I) \\ &= \operatorname{argmax}_{\mathbf{M}_i} p(I | \mathbf{M}_1, \dots, \mathbf{M}_l) p(\mathbf{M}_1, \dots, \mathbf{M}_l). \end{aligned}$$

Given an image likelihood distribution  $p(I | \mathbf{M}_1, \dots, \mathbf{M}_l)$  of the multi-object complex, the joint p.d.f. (4.12) in the previous section provides a simple shape prior of multiple objects for MAP-based segmentation. This interpretation of the joint p.d.f. is not affected by the order in which objects are estimated. However, the joint p.d.f. can be interpreted as

$$p(\mathbf{M}_1, \dots, \mathbf{M}_l) = p(\mathbf{M}_1 | \mathbf{M}_2, \dots, \mathbf{M}_l) \times p(\mathbf{M}_2 | \mathbf{M}_3, \dots, \mathbf{M}_l) \times \dots \times p(\mathbf{M}_{l-1} | \mathbf{M}_l) p(\mathbf{M}_l).$$

If the joint p.d.f. is estimated in this fashion, the estimation is sensitive to the order of objects in which each object is estimated.

#### 4.4.5 Segmentation of Male-Pelvis Model

A set of training models were obtained from male-pelvis CT images of real patients taken over a series of days who underwent radiotherapy. Three organs, namely the bladder, prostate, and a section of the rectum that is adjacent to the prostate are modeled. Both ends of the rectum model are arbitrary. A clinician contoured each organ slice by slice to generate binary images for all three organs. A single-figure m-rep was then fit to each of the bladder, prostate, and rectum separately in binary images:  $5 \times 6$  grids of medial atoms for the bladder,  $7 \times 4$  grids of medial atoms for the prostate, and  $15 \times 3$  grids of medial atoms for the rectum are used. This fitting process prevents penetration among fitted m-reps, prevents folding of the interior of the object represented, and maintains regularity of the grid across cases for correspondence of medial atoms [Merck et al., 2008]. The fitted three m-reps for bladder, prostate, and rectum of each patient were aligned by a similarity transformation that is computed from two landmarks, at the apex and base of the prostate. Then those aligned m-reps for bladder, prostate, and rectum were combined into one ensemble m-rep. The total number of the medial atoms is 103, and the number of parameters in the ensemble m-rep is 927.

##### 4.4.5.1 Probability Density Estimation

The iterative approach was applied to the male-pelvis models of five patients' m-rep fits. The patients are numbered as 3101, 3106, 3108, 3109, and B163. There were 14 m-rep fits for 3101, 17 for 3106, 18 for 3108, 18 for 3109, and 15 for B163. Models fitted to the first treatment image were used as the reference model from which the variation of the rest of models was taken.

In the first iteration, the neighbor parts in variations of the bladder and rectum were assumed to be zero because the self part of the bladder and rectum changes dominate the neighbor parts. Similarly, the self part in the prostate was set to zero for the shape of the prostate changes little except as affected by the bladder and rectum.

Starting with  $\Delta \mathbf{M}_b$ ,  $\Delta \mathbf{U}_b$ ,  $\Delta \mathbf{M}_p$ ,  $\Delta \mathbf{U}_p$ ,  $\Delta \mathbf{M}_r$ ,  $\Delta \mathbf{U}_r$  where the subscripts  $b$ ,  $p$ ,  $r$  represent the bladder, prostate, and rectum respectively, the estimates of neighbor and self effects in each organ in the first iteration were computed as follows.

|                    | <i>total variation per object</i> |                  |             |                 |                   |               |
|--------------------|-----------------------------------|------------------|-------------|-----------------|-------------------|---------------|
| <i>patient no.</i> | bladder self                      | bladder neighbor | rectum self | rectum neighbor | prostate neighbor | prostate self |
| 3101               | 0.379383                          | 0.014782         | 0.360542    | 0.004983        | 0.179949          | 0.006880      |
| 3106               | 0.082828                          | 0.002576         | 0.102897    | 0.001188        | 0.031047          | 0.003648      |
| 3108               | 0.391689                          | 0.010346         | 0.181145    | 0.012492        | 0.135748          | 0.009021      |
| 3109               | 0.080293                          | 0.006388         | 0.132851    | 0.003501        | 0.043438          | 0.005439      |
| B163               | 0.067945                          | 0.002470         | 0.137690    | 0.002819        | 0.053831          | 0.003205      |

Table 4.1: Total variations of two effects, self and neighbor per organ that are estimated after 2nd iteration.

1. Self effect on the bladder, and rectum differences  $\Delta\mathbf{M}_b, \Delta\mathbf{M}_r$ , assuming  $\Delta\mathbf{M}_b^n, \Delta\mathbf{M}_r^n$  are zero.
2. Neighbor effect on prostate differences  $\Delta\mathbf{U}_p$ , assuming  $\Delta\mathbf{T}_p^s$  are zero. For prediction from step 1 the estimate of the self variations of bladder and rectum was used.
3. Self effect on prostate residue from the estimate of the neighbor effect. The residue is  $\Delta\mathbf{M}_p^s := \Delta\mathbf{M}_p \ominus \widehat{\Delta\mathbf{M}}_p^n$ .
4. Neighbor effect on bladder and rectum residue from the estimate of the self effect  $\Delta\mathbf{U}_b, \Delta\mathbf{U}_r$  where  $\Delta\mathbf{T}_b := \Delta\mathbf{T}_b \ominus \widehat{\Delta\mathbf{T}}_b^s$  and  $\Delta\mathbf{T}_r := \Delta\mathbf{T}_r \ominus \widehat{\Delta\mathbf{T}}_r^s$ . Again, the estimate of the self part of the prostate  $\widehat{\Delta\mathbf{M}}_p^s$  from step 3 was used to compute the prediction of the bladder and rectum from the change of their neighboring object prostate.

For the later iterations, the steps described in section 4.4 were repeated with the updated estimates of self and neighbor effects from the previous iteration.

Table 4.1 compares the estimates of total variations of the two effects for each organ after the 2nd iteration. They are the sum of eigenvalues that are estimated in each step described in the previous section. Fig. 4.7 shows the primary mode of self variation of the bladder and the associated prediction of the deformation of the prostate in patient B163. It also shows a dent formed in the bladder in which the prostate fits as the bladder fills.

The results are consistent with what is known about the anatomy and seen in the training data. The self terms for the bladder and rectum dominate the neighbor terms, reflecting the fact that these organs' variations are principally due to changes in their contents. On the other

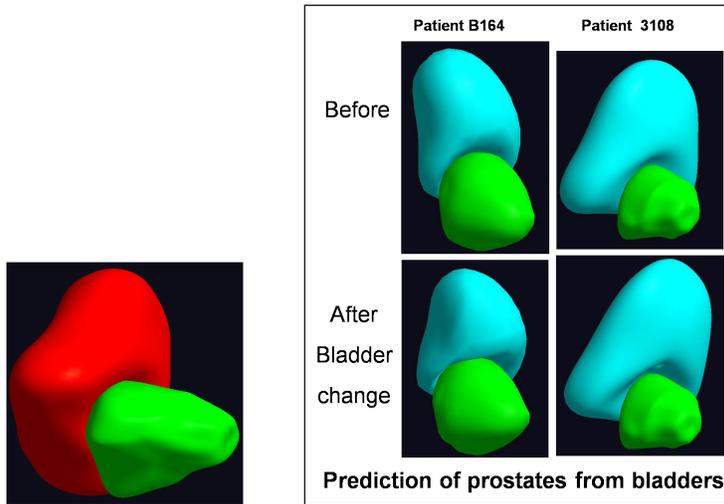


Figure 4.8: (left) A multi-object complex of the bladder and the prostate. (right) The predicted prostates brought on by the changes in the bladders of the patient B163 and 3109.

hand, the neighbor term for the prostate dominates the self term, reflecting the fact that the prostate is a rather hard organ that is pushed upon by the rectum and bladder.

Moreover, the predictions of the prostate from the bladder and the rectum seem realistic. Also, the prostate changes from its prediction are smaller than the self changes in the bladder and rectum, which are known to be larger.

The bladder self changes include modes corresponding to lengthening, widening, and lapping around the prostate, all anatomically observed processes. Also encouraging is that the prostate predictions of the bladder self modes have the prostate following the change in the indentations of the bladder while keeping the prostate almost entirely nonintersecting. Furthermore, the prostate self modes also make consistent predictions on the bladder indentation.

#### 4.4.5.2 Segmentation Results

On the basis of the positive result from the pilot study on the male-pelvis model with the iterative method, it was reasonable to check the convergence of the variances as the number of iterations increases and to segment the bladder and the prostate of the male-pelvis models using the estimated shape prior distribution.

A different set of improved m-rep fits to the male-pelvis CT images of the same five patients was used for training and segmentation. The rectum fits were not included in this experiment

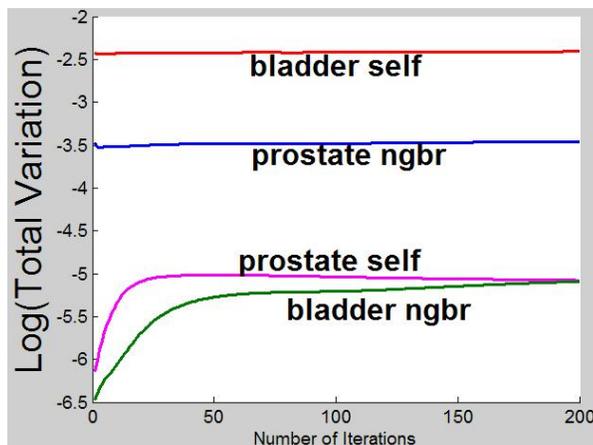


Figure 4.9: A log plot of total variances of the self variation and the neighbor effect per organ of the patient 3109 that are estimated over 200 iterations

because both ends of the rectum models are cut arbitrarily, which introduces noise in the estimation. Training of the self variations and the neighbor was done in the same manner. In the first iteration, the neighbor effect on the bladder was set to zero, and the self variation of the prostate was set to zero. The first treatment images were still used as the reference model. One modification to the iterative method described in section 4.4.2 is that the neighbor effects  $\Delta \mathbf{T}^n$  and the self variations  $\Delta \mathbf{M}^s$  of each organ were assumed to have zero means since the reference models were considered as the means. Thus, PGA on these two residues produces only the modes of variations.

Fig. 4.8 shows a multi-object complex of the bladder and the prostate on the left side and the prostate predicted by the changes in the bladder on the right. The prediction was done using the principal geodesics of the self variations and the neighbor effects after 200 iterations. The estimation of the self and neighbor effect shown in eigenmodes and their variances were biologically realistic. The self variation of the bladder displayed a large variation independent from changes of its neighboring object prostate while the prostate shows a small self variation in its own shape compared to the effect from its neighboring object bladder.

Their variances, the sum of eigenvalues that are estimated, also converge. Fig. 4.9 shows a log of the total variances of the self variation and the neighbor effect of the patient 3109 over 200 iterations. The variances of each organ converge pretty quickly except the variances of the neighbor effects on the bladder.

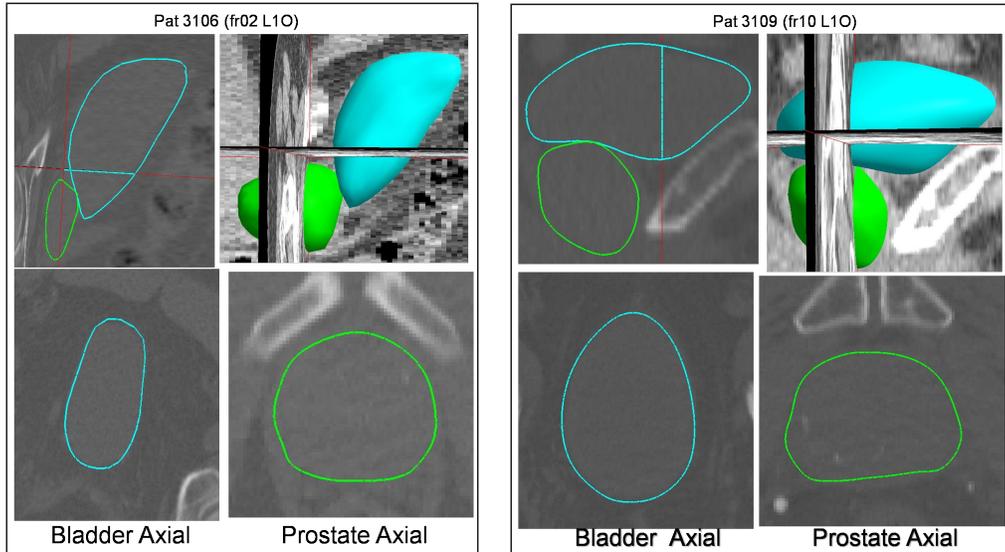


Figure 4.10: Segmentation of the bladder and the prostate of patient 3106 and 3109. The contour of the bladder is in cyan, and the contour of the prostate is in green.

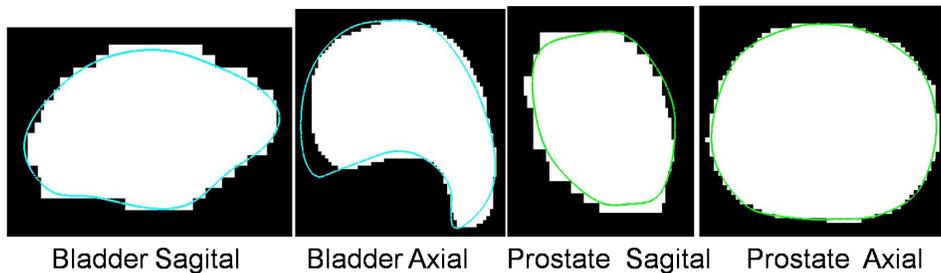


Figure 4.11: Segmentation results of the bladder and the prostate of patient 3109 are shown in slices against human expert segmentations.

The estimated geometric shape probability distributions of the male pelvic organs were incorporated as the prior within m-rep segmentation method's posterior objective function. This prior was evaluated by the effectiveness of the segmentation of these organs in leave-one-out segmentations of the bladder and the prostate of 5 patients over 75 cases. Fig. 4.10 shows the 3D view and the axial slices of the segmentation results of the bladder and the prostate of patients 3108 and 3109 among best results. The CT image in Fig. 4.10 shows the challenge in the prostate segmentation. The image intensity of the prostate is hardly distinguishable from the background. Fig. 4.11 compares the segmentations against human expert segmentations.

Two summary plots of the prostate segmentations are given in Fig. 4.12 in terms of the

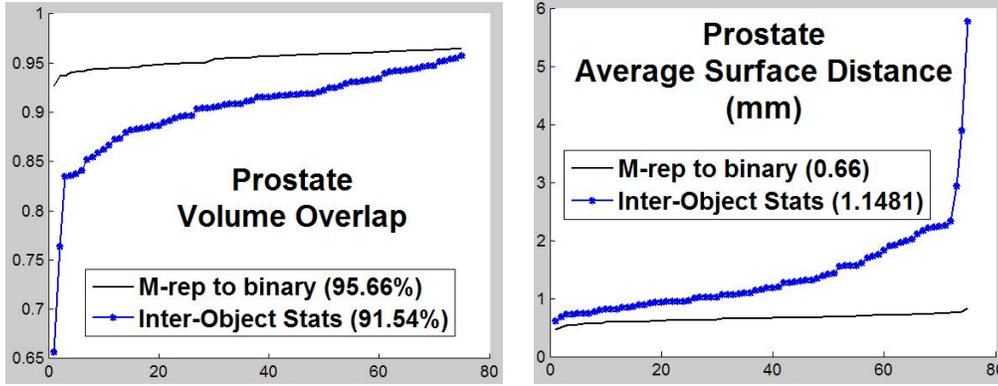


Figure 4.12: Plots of sorted volume overlaps and average distances of the prostate segmentations over 5 patients, total 75 cases.

volume overlaps and the average surface distance in mm. There are three or four worst cases whose volume overlaps and the average surface distances are very poor compared to the binary fits. However, the segmentation results are in general encouraging.

## 4.5 Discussion and Conclusion

In the nonlinear manifold on which m-reps are situated, the addition operation is not commutative. The separation of  $\Delta\mathbf{M}$  into  $\Delta\mathbf{M}^s \oplus \Delta\mathbf{T}^n$  is thus not equivalent to  $\Delta\mathbf{T}^n \oplus \Delta\mathbf{M}^s$ . In our iterative algorithm that estimates the self variations  $\Delta\mathbf{M}^s$  and the neighbor effects  $\Delta\mathbf{T}^n$ , this non-commutativity is ignored and  $\Delta\mathbf{M}^s$ ,  $\Delta\mathbf{T}^n$  are treated as if they are interchangeable. It should be tested whether the effect of our method's assumption of commutativity is significant.

It would also be useful to test the four assumptions stated in section 4.4.1 to show our interpretation on joint probability holds. The interpretation and the simplification of joint probability of multiple objects rely on these assumptions about correlations among the self variations and the neighbor effects of objects in a multi-object complex. The validity of these assumptions in turn provides ground of using the estimated self variations and the neighbor effects as a shape prior for segmentation of a multi-object complex since the simplified joint probability (Eq. (4.12)) is not sensitive to the order of objects.

The applicability of the method described in this chapter depends on the type of statistical variation in the population. Problems to which statistical shape analysis can apply can be

classified according to the target data, namely within patient and across patient. Within-patient data are relatively easier to deal with than cross-patient data since the amount of variations in the configuration and shapes of multiple anatomic objects within a patient are small compared to those of anatomic objects across patients. Objects in a region of a patient are bound together by muscles and tissues among the objects, so the positioning and the distribution of objects in a patient are rather stationary, and the shape changes of the objects in a patient are limited within the configuration. As a result, the configuration and the shape changes of multiple objects within a patient are diffeomorphic. However, this diffeomorphic property in the configuration and the shape changes of objects within a patient cannot be guaranteed when the same multiple objects are considered across patients. The configuration of the objects changes. For example, sliding of one object along the other object can be observed between two adjacent objects across different patients. But also the shape of each object in the same multi-object complexes is much more variable in cross-patient data than in within-patient data.

Thus, the configuration and the shape variations of objects of a patient can be predicted reasonably well from some mean or reference configuration and shapes of the objects even with a limited number of training models of the patient. Cross-patient data have much more non-diffeomorphic and geometric variations than within-patient data in the configuration of anatomic objects in a region, and in the shape variation of each object, which makes it challenging to capture the inter-object relation in different patients.

The experiments reported in this chapter are limited to data within a patient. They suggest that the new approach to separate out inherent variation of an object itself and effects from its neighboring objects is promising for within-patient cases. It is interesting to consider whether the new approach also applies to cross-patient cases.

For both within-patient and cross-patient data a further analysis is required to verify that the estimates truly reflect the self and neighbor effects of multi-objects. One possible approach is to simulate the obvious neighbor effects and self variations independent from neighbor effects on multiple ellipsoids  $m$ -reps and apply this approach on the simulated multiple ellipsoids  $m$ -reps to determine what are self variations and neighbor effects.

An open issue is how to choose augmented atoms. Clearly, that choice affects our estimates

of the self variation and the neighbor effects. For within-patient data the obvious choice is the most highly correlated neighboring atoms with those in the target object. Separate research is now being carried out to measure this correlation [Jung and Marron, 2008], and the results of that research will provide a firmer basis for this choice than the distance criterion presently being used. For cross-patient data the choice is not so simple because of the issue of correspondence.

The methods of augmentation and prediction assume the correspondence of selected atoms in neighboring objects to augment the target object across training models since this correspondence captures the relation of the neighboring objects to the target object. This assumption is the limiting factor of the proposed shape models for cross-patient data. As mentioned before, it is fair to assume this correspondence for training models within a patient since the changes in the configuration and shapes of organs are diffeomorphic and the amount of geometric variations are relatively limited. On the other hand, it is not reasonable to assume that the relation that the selected atoms in neighboring objects have relative to the target object in one patient carries over to the other patient. For example, positions of dents that can happen on the bladder by the nearby prostate of one patient can be explained by some atoms in the prostate. However, it is not feasible to expect that the same set of atoms (atoms of the same indices) in the prostate of the other patient can predict positions of dents on the bladder, considering that the positions of dents on the bladder vary from patients to patients. For the proposed shape model to work for cross-patient data, the methods of augmentation and prediction need an improvement that solves this issue of the correspondence of augmented objects across patients.

## Chapter 5

# Conditional Shape Statistics<sup>1</sup>

This chapter presents an indirect approach to the challenging problem of automatic segmentation of an object of ambiguous boundary when its nearby objects are easier to segment. As mentioned throughout this dissertation, segmentation of human organs is one of the central problems in medical image analysis. Anatomical structure is complex: some organs are clumped together and touch each other, and low physical contrast in the image between such structures can make significant parts of the boundary of the target organ ambiguous. Segmenting such an object slice-by-slice manually can be a particularly hard and laborious task even to clinicians, and certainly also to automatic segmentation methods. However, if image intensities at the boundaries of some objects near the target object are easier to identify than the target object, then positions, orientations or material properties of these objects relative to those of the target object can be used as guides to pick out the target object.

The goal of the approach is to make use of this observation and to incorporate into shape priors the relations of the target object to its neighboring objects that can be easily identified so that the shape variation of the target object is constrained not only by its own shape variations but also by its positional, orientational information, and even shape deformations relative to neighbor objects over a population. A natural and intuitive way to include this supplementary information into the shape prior is by means of conditional probability distribution. Thus, the conditional shape probability distribution (CSPD) of the target shape model given neighboring shape models is considered for segmentation of the target object. The CSPD provides a shape

---

<sup>1</sup>A version of this chapter has published in [Jeong et al., 2008b]

prior for finding a most probable shape of the target object given the configuration of its neighboring objects in relation to the target object.

In chapter 4 a novel approach to estimate the shape probability distribution of multi-objects was introduced. The basic objective of the conditional model is similar to that described in the previous chapter in the sense that both approaches aim to capture the inter-relations of objects in the shape probability distribution of each object. The main difference is that this conditional approach is geared towards finding the target object using its neighboring objects. That is, the conditional approach relies heavily on the segmentation quality of the neighboring objects and is likely to succeed when the segmentation quality of the neighboring objects is as good as that of the training models. The multi-object statistical shape model in the previous chapter is more general than the conditional shape model since the inter-dependency among objects is reflected on the self variation and the neighbor effect per object that is estimated iteratively.

Section 5.1 reviews the Gaussian conditional probability distribution. Section 5.2 presents a method to estimate the CSPD using the principal components regression method that was described in section 2.1.3.1 of chapter 2. Finally, section 5.3 demonstrates the application of the conditional shape model to a set of synthetic multi-object complex data and to real multi-object complex data.

## 5.1 Conditional Probability Distribution

The augmented representation of the target object  $U = (T, N)$  is assumed to follow a multivariate normal distribution. Recall that  $N$  is defined as the selected features in the neighboring objects and that  $T$  is defined as the shape features of the target object affected by the changes of  $N$  in section 4.2.1. Then the conditional probability distribution of  $T$  given  $N$ ,  $p(T|N)$ , also follows the multivariate normal distribution [Muirhead, 1982] whose mean  $E(T|N)$  and covariance matrix  $Cov(T|N)$  are

$$E(T|N) = \Sigma_{TN}\Sigma_{NN}^{-1}N \quad (5.1)$$

$$Cov(T|N) = \Sigma_{TT} - \Sigma_{TN}\Sigma_{NN}^{-1}\Sigma_{NT} \quad (5.2)$$

where  $\Sigma_{NN}^-$  is a generalized inverse of the covariance matrix of  $N$ , i.e.,  $\Sigma_{NN}\Sigma_{NN}^-\Sigma_{NN} = \Sigma_{NN}$  and the covariance matrix of the combined tuple  $U$  is

$$\Sigma = \begin{bmatrix} \Sigma_{TN} & \Sigma_{NN} \\ \Sigma_{NN} & \Sigma_{NT} \end{bmatrix}. \quad (5.3)$$

Since the atoms in  $T$  and  $N$  are elements in nonlinear symmetric space, the Riemannian log map of  $U$  at the Fréchet mean  $\mu$  of  $U$  is taken to map  $N$  and  $T$  into a corresponding linear tangent space at  $\mu$ . That is,  $T$  and  $N$  in the equation above are corresponding feature vectors of  $T$  and  $N$  in the tangent space at  $\mu$  of  $U$ . Thus, in the equation for the conditional mean (5.1) both  $T$  and  $N$  must be centered (mean zero) in the tangent space.

Estimating both the conditional mean and the shape distribution involves computing an inverse of a sample covariance matrix. The limited sample size and high dimensional feature space of shape descriptors provide a challenge in computing an appropriate pseudo-inverse of the covariance matrix. To deal with the HDLSS problem, the principal components regression (PCR) method [Shawe-Taylor and Cristianini, 2004] is employed in estimating the target mean shape given its neighboring shapes and the conditional covariance. The next section describes how PCR is used for the estimation of the CSPD.

## 5.2 Estimation of Conditional Shape Distribution using PCR

PCR corresponds to performing principal component analysis (PCA) and regressing in the feature space given by a few first principal directions that capture major variations explained in a given sample. Using PCR to estimate the CSPD has the following advantages. First, doing PCA on the neighboring shape features reduces the dimension of the feature shape space of neighbor objects. As will be shown in this section, in the reduced features shape space the sample covariance matrix becomes non-singular, and its inverse can be stably estimated. Second, using PCR can help to tighten the distribution of the estimated CSPD since the major variation of the neighboring objects is explained in the reduced features shape space.

Let  $\mathbf{N}$  and  $\mathbf{T}$  be data matrices of  $N$  and  $T$  respectively. PCA (or SVD) gives principal axes of  $\mathbf{N}$ . Let  $\mathbf{N}' = \mathbf{U}\mathbf{\Lambda}\mathbf{V}'$  by SVD;  $\mathbf{U}_k$  has columns of the first  $k$  principal axes. Let  $\tilde{\mathbf{N}}$  be

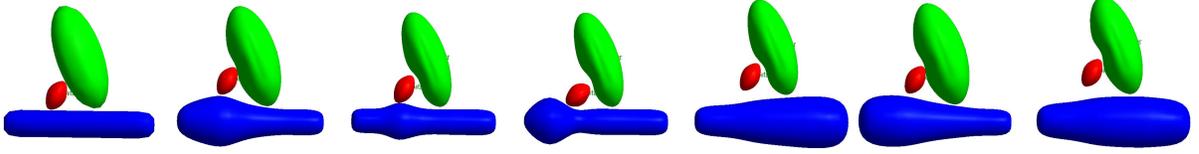


Figure 5.1: Samples from simulation of a quasi-bladder, quasi-prostate, quasi-rectum complex. The leftmost one is the base case of the multi-object complex.

the principal components of  $\mathbf{N}$ , i.e.,  $\tilde{\mathbf{N}} = \mathbf{N}\mathbf{U}_k$ . When the sample size is  $n$ , the covariance matrix of  $\tilde{\mathbf{N}}$  is

$$\begin{aligned}
 \Sigma_{\tilde{\mathbf{N}}\tilde{\mathbf{N}}} &= \frac{1}{n-1} \tilde{\mathbf{N}}' \tilde{\mathbf{N}} \\
 &= \frac{1}{n-1} \mathbf{U}'_k \mathbf{N}' \mathbf{N} \mathbf{U}_k \\
 &= \frac{1}{n-1} \mathbf{U}'_k (\mathbf{U} \mathbf{\Lambda} \mathbf{V}') (\mathbf{V} \mathbf{\Lambda} \mathbf{U}') \mathbf{U}_k \\
 &= \frac{1}{n-1} \mathbf{\Lambda}_k^2,
 \end{aligned} \tag{5.4}$$

where  $\mathbf{\Lambda}_k$  is the diagonal matrix of the first  $k$  diagonal components of  $\mathbf{\Lambda}$ . It appears reasonable that estimating the inverse of  $\Sigma_{\tilde{\mathbf{N}}\tilde{\mathbf{N}}}$  would be more stable compared to the ridge regression approach [de Bruijne et al., 2006, Shawe-Taylor and Cristianini, 2004], which tries to regularize  $\Sigma_{TT}$  by adding a typically positive small constant and computing the inverse.

Now it can be shown that the estimated conditional mean  $E(T|\tilde{\mathbf{N}})$  is equivalent to the estimator  $\hat{\mathbf{T}}$  by PCR, that is,  $\hat{\mathbf{T}} = \mathbf{N}\mathbf{U}_k \hat{\mathbf{W}}$ . With Eq. (5.4) and  $\Sigma_{T\tilde{\mathbf{N}}} = 1/(n-1) \mathbf{T}' \tilde{\mathbf{N}}$ , the conditional mean is

$$\begin{aligned}
 E(T|\tilde{\mathbf{N}}) &= \Sigma_{T\tilde{\mathbf{N}}} \Sigma_{\tilde{\mathbf{N}}\tilde{\mathbf{N}}}^{-1} \tilde{\mathbf{N}}' \\
 &= \mathbf{T}' \tilde{\mathbf{N}} \mathbf{\Lambda}_k^{-2} \tilde{\mathbf{N}}' \\
 &= \mathbf{T}' \mathbf{N} \mathbf{U}_k \mathbf{\Lambda}_k^{-2} \mathbf{U}'_k \tilde{\mathbf{N}}'.
 \end{aligned} \tag{5.5}$$

It can be easily seen that Eq. (5.5) is the same as  $\hat{\mathbf{T}}'$  when  $\hat{\mathbf{W}}$  is replaced with the identity Eq. (2.12).

The CSPD’s covariance matrix  $Cov(T|\tilde{N})$  (Eq. (5.2)) is computed in the tangent space. Once the conditional covariance matrix is estimated, SVD on the covariance matrix produces the principal directions. A subset of the conditional principal directions that is sufficient to describe the variability of the target m-rep shape space is used as an estimate of the conditional shape probability distribution of  $T$  given  $N$ .

### 5.3 Applications of CSPD to Deformable M-rep Segmentation

The new conditional shape model was applied to two data sets. A set of simulated multi-object complexes was first used to check that the estimated conditional means of the target object have a reasonable shape and position relative to the neighboring objects and that the estimated conditional principal geodesics capture the input transformations. After the conditional shape model of the simulated multi-object complexes was examined, the conditional shape model was applied as a shape prior in an m-rep deformable segmentation framework to segment the prostate in the male-pelvis area.

Section 5.3.1 describes the synthetic multi-object models to which simple transformations were applied and illustrates the principal modes of variation of the CSPD estimated from the synthetic data. Section 5.3.2 describes segmentation of the prostate given neighbor atoms from the bladder and the rectum of real patients using the conditional principal geodesics with the conditional means as initializations of the segmentation.

#### 5.3.1 Simulated Multi-objects

The first data set used to apply the conditional shape model is a randomly generated multi-object complex whose shape changes and motions simulate the bladder, the prostate, and the rectum in the male pelvis. The complex is made up of the pelvic arrangement of two slab ellipsoid (quasi-bladder and quasi-prostate) m-reps and one tube (quasi-rectum) m-rep. Fig. 5.1 shows the base model and sample deformations of the base model. The quasi-bladder slab m-rep is represented by a  $3 \times 7$  grid of medial atoms, the quasi-prostate slab m-rep is represented by a  $3 \times 5$  grid of medial atoms, and the quasi-rectum tube m-rep is represented by a column of 8 medial atoms. The transformations applied to each object imitate shape and motion changes

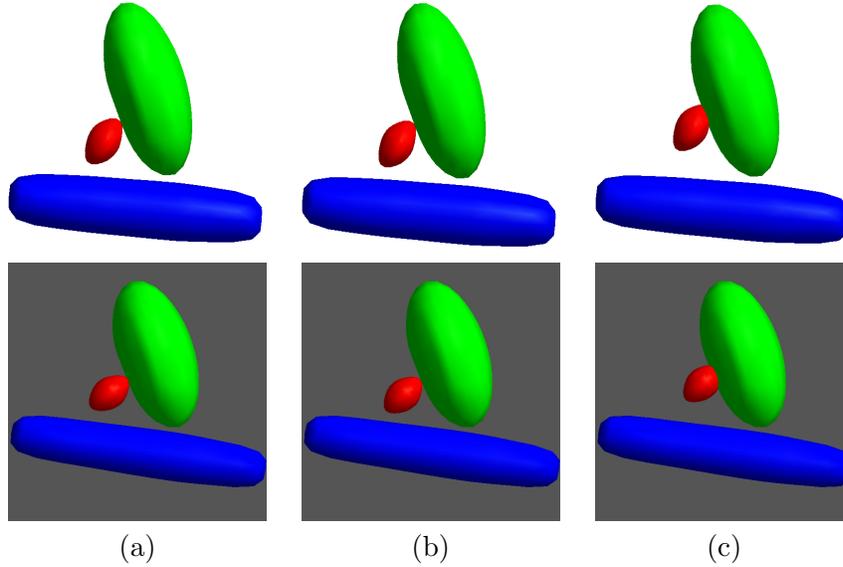


Figure 5.2: Shape variability of simulated multi-object complex. Top row: (a) mean- $1\sigma$  along the 1st principal geodesics of conditional distribution of the quasi-prostate given augmented atoms from quasi-prostate and quasi-bladder (b) mean of simulated multi-object complex (c) mean+ $1\sigma$  along the 1st principal geodesics of conditional distribution of the quasi-prostate. Bottom row: (a) mean- $1\sigma$  along the 1st principal geodesics of non-conditional distribution of the quasi-prostate (b) mean of simulated multi-object complex (c) mean+ $1\sigma$  along the 1st principal geodesics of non-conditional distribution of the quasi-prostate

that are observed in real organs and interaction between objects. The transformations that simulate the effect of neighboring objects depend on one random parameter  $\alpha$  that indicates the stiffness of a quasi-prostate.

**Quasi-bladder Ellipsoid M-rep** A point  $x_{bp}$  is set on the base ellipsoid quasi-bladder surface to simulate the location through which the urethra goes from quasi-bladder through the quasi-prostate. One random transformation is applied on the base quasi-bladder. This transformation imitates filling or emptying of the bladder with urine. The transform scales the lengths of the major axes of the base ellipsoid to produce a uniform thickening or thinning of the ellipsoid. If  $a_b, b_b, c_b$  are lengths of the major axes, then their ratio  $a_b : b_b : c_b$  follows a log-normal distribution, i.e.,  $a_b : b_b : c_b = e^l$  where  $l \sim N(0, \sigma_l)$ .  $l$  is one independent parameter of the transformation.  $x_{bp}$  is relocated to  $x'_{bp}$  by this transformation.

**Local Effect from Quasi-bladder to Quasi-prostate** As the quasi-bladder expands, the quasi-prostate shifts and in turn dents the quasi-bladder when the expansion of the quasi-

bladder is large. If  $x_{bp}$  would have moved to  $x'_{bp}$  in the absence of a quasi-prostate,  $x_{bp}$  is restricted to move only by the fraction  $\alpha$  of the distance between  $x_{bp}$  and  $x'_{bp}$ , and the quasi-prostate is shifted by the same amount along the same direction. The amount of dent in the quasi-bladder at  $x_{bp}$  by the quasi-prostate is determined deterministically to prevent interpenetration between the quasi-bladder and the quasi-prostate.

**Quasi-Rectum Tube M-rep** One random transformation is applied to the quasi-rectum tube to imitate the bulging of the rectum due to gas. The simulated bulge is cylindrical along the axis of tube, so the cross-section of the bulge is a circle. The transformation has 3 random parameters: the center position  $u_0$  of the bulge along the tube axis, the width  $w$  of the bulge, and the height  $h$  of the bulge.  $u_0$  follows a uniform distribution along the range of the quasi-rectum,  $w$  follows a normal distribution  $N(0, \sigma_w)$ , and  $h$  also follows a normal distribution  $N(0, \sigma_h)$ . Then the radius of the tube is scaled by  $1 + e^{h^2 - \frac{(u-u_0)^2}{2w^4}}$  where  $u$  is a parameter of the tube axis. The last two random parameters  $h$  and  $w$  determine the shape of the bulge.

**Local Effect from Quasi-rectum to Quasi-prostate** Bulging of the quasi-rectum causes the quasi-prostate to rotate and to shift. Consider a coronal plane that goes through the center of the quasi-prostate. If bulging happens on one side of the coronal plane, the quasi-prostate tilts toward the other side of the plane. If bulging happens right below the quasi-prostate, it pushes up the quasi-prostate. The direction and the displacement of the quasi-prostate is constrained so that these two objects do not interpenetrate each other. The displacement direction of the quasi-prostate is from  $x_t$  to  $m_p$ , where  $m_p$  is the center of the quasi-prostate and  $x_t$  is a top point of the bulge on the quasi-rectum. The displacement amount is also determined deterministically.

### 5.3.1.1 Training Results

The radii of the major axes  $a_b, b_b, c_b$  of the quasi-bladder base ellipsoid are 0.1597, 0.1076, and 0.0811 respectively. The radii of the major axes  $a_p, b_p, c_p$  of the quasi-prostate base ellipsoid are 0.0490, 0.0414, and 0.0325 respectively. The length of the quasi-rectum base tube is 0.4858. The radius of the quasi-rectum base tube is 0.0392.  $u_0$  is sampled from the  $\frac{1}{7}$  point to the  $\frac{6}{7}$  point of the tube axis along its length. 200 simulated multi-objects were generated by random sampling of parameters except  $\alpha$ .  $\alpha$  was not randomly sampled but was set to 1 to simplify

the experiment.  $\sigma_l$  was set to 0.1,  $\sigma_h$  was set to 0.2, and  $\sigma_w$  was set to 1.0. Out of 200 models, 12 models were thrown out because of interpenetration among the objects.

The CSPD of the quasi-prostate given selected neighbor atoms from the quasi-bladder and quasi-rectum was trained from the 188 synthetic multi-object m-reps. As neighbor atoms for the quasi-prostate, 4 atoms were selected from the quasi-rectum located right below the quasi-prostate, and 9 atoms were selected from the quasi-bladder near the end of the quasi-prostate meeting the quasi-bladder. Since these are simulated objects, no alignments were applied to the training models before training. The dimensions of the tube atom and slab atom in tangent space are 8 and 9 respectively [Saboo et al., 2008, Fletcher et al., 2004]. So the dimension of the combined  $T$  and  $N$  atoms in tangent space is  $4 \times 8 + 9 \times 9 = 113$ . In computing the conditional means and shape distribution of the quasi-prostate, the first four principal components of  $N$  atoms were chosen to reflect the four random transformations of the quasi-bladder and the quasi-rectum; this is a large reduction in dimension of the  $N$  from 113 to 4. This reduction allowed us to have a stable estimation of the conditional covariance matrix.

Table 5.1: Number of fractions (sample size) per patient

| Patient Id          | 3101 | 3106 | 3108 | 3109 |
|---------------------|------|------|------|------|
| Number of fractions | 14   | 16   | 17   | 17   |

Fig. 5.2 shows a comparison of the first principal geodesics of the conditional and non-conditional distributions of the quasi-prostate. To do the comparison, the first mode of conditional shape variation was applied to the non-conditional mean of the quasi-prostate. As can be seen in Fig. 5.2, the first non-conditional shape mode is not constrained by the neighboring two objects, and at 1 standard deviation from the mean the deformed quasi-prostate penetrates the quasi-bladder. In contrast, the 1<sup>st</sup> mode of variation from CSPD is restrained by the neighboring objects, which prevents the penetration between the quasi-bladder and the quasi-prostate at 1 standard deviation from the mean. The other modes of variations from the CSPD show that 1) they capture the tilting and the shift toward the quasi-bladder caused by the bulge of the quasi-rectum and 2) they are within the confines of neighboring objects. Although the non-conditional modes capture the tilting and shift, the range of their variation,

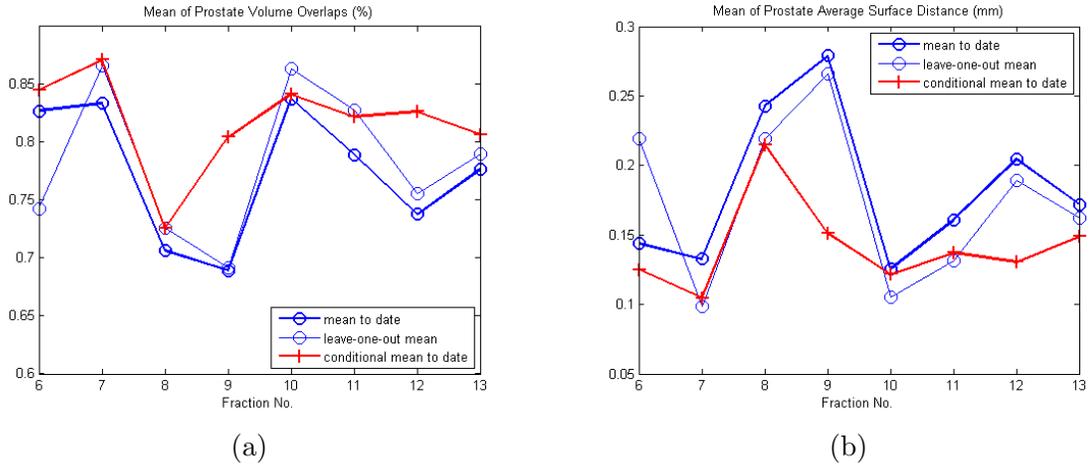


Figure 5.3: (a) Mean *Jaccard similarity coefficients (JSC)* [Crum et al., 2006] of mean prostate volume overlaps with training prostate binary images; (b) Mean of average surface distance in mm between mean prostate and training prostate binary images. Mean of these measures are taken over 4 patients at each fraction, from fraction 6 to fraction 13. The thin blue line with the round marker indicates leave-one-out mean prostate, the thick blue line with the round marker indicates mean prostate to date, and the red line with the cross marker indicates conditional mean given neighbor atoms from the bladder and prostate.

especially for the shift of the quasi-prostate, is too large. The eigenvalues of non-conditional modes are all at least an order of magnitude larger than those of conditional modes.

### 5.3.2 Objects in the Pelvic Region of Male Patients

The data used consist of four patients' image sets of which each is a series of approximately 16 CT scans taken during radiation treatments. They were fractionated over a series of days, and a new image was acquired before each fraction. Each image has an in-plane resolution of  $512 \times 512$  with voxel dimension of  $0.98 \text{ mm} \times 0.98 \text{ mm}$  and 3 mm distance between slices. Each image contains the bladder, prostate, and rectum. Expert manual segmentations of these three objects were provided with each image. M-reps were fitted to the manual segmentations:  $5 \times 6$  grids of medial atoms for the bladder,  $7 \times 4$  for the prostate, and 17 columns for the rectum. The fitting procedure prevents penetration among models and folding of the interior of the object, and it maintains the regularity of the grid of medial atoms for correspondence of the atoms [Merck et al., 2008]. These training models of the bladder and the rectum were used as the neighboring objects for prostate segmentation in this experiment.

**Segmentation Structure** The prostate in each image was segmented by Bayesian optimization over m-rep principal geodesic coefficients. The objective function for the optimization has two terms: the geometric penalty and the image match. Mahalanobis distance function from a mean model to a deformed model was used for the geometric penalty term. The regional intensity quantile functions (RIQF) [Broadhurst et al., 2006] was used for the image match term. The m-rep defines a coordinate system that provides an explicit correspondence between deformation of the mean m-rep model and the 3D volume in the object boundary region. The fitted m-reps were placed into the associated grayscale images, and image intensity quantile functions (inverse cumulative histograms) in regions interior to and exterior to the target object boundary were recorded in each training image. The image match function measures the statistical distance between observed quantile functions in a target image and trained quantile functions over the population for the corresponding region of the training set. Segmentation starts with the mean model placed in a new image of a patient by a similarity transformation computed by aligning the pubic bone of the patient. Segmentation then proceeds by a conjugate gradient optimization of the posterior of the geometric parameter given the new image data.

### 5.3.2.1 Results of Within-patient Successive Segmentations

Table 5.1 shows the sample size per patient. At any fraction  $i$ , the CSPD of each patient was trained using training fits from fraction 1 up to fraction  $i - 1$ . To compare with leave-one-day-out experiments that were done previously, the RIQF per fraction  $i$  was trained by all days' grayscale images of the patient except for fraction  $i$ . Fig. 5.3 shows the comparison of mean volume overlaps of all 4 patients at each fraction from fraction 6 to 13 among conditional means to date, non-conditional means to date, and leave-one-out means. It also shows the same comparison by mean average surface distances. The first 5 fractions were not included in the comparison because less than six training models are not enough to train covariance matrices to compute conditional means. About 6 to 9 different sets of atoms from the bladder per patient and 1 fixed atom from the rectum across patients  $i$  were selected as neighbor atoms to the prostate from the bladder and rectum. The coefficients of the first three principal axes of the neighbor atoms were taken as new features for neighbor atoms in estimating the CSPD

using PCR as described in section 5.2.

Table 5.2: Mean volume overlaps of figure 5.3

| Fraction No.        | 6      | 7      | 8      | 9      | 10     | 11     | 12     | 13     |
|---------------------|--------|--------|--------|--------|--------|--------|--------|--------|
| Mean to date        | 0.8272 | 0.8336 | 0.7064 | 0.6893 | 0.8376 | 0.7890 | 0.7378 | 0.7766 |
| Leave-one-out mean  | 0.7429 | 0.8664 | 0.7261 | 0.6915 | 0.8634 | 0.8278 | 0.7560 | 0.7900 |
| Cond'l mean to date | 0.8449 | 0.8713 | 0.7255 | 0.8044 | 0.8415 | 0.8221 | 0.8267 | 0.8068 |

**Result 1: Initialization** The comparison clearly indicates that the conditional mean prostate to date gives much better initialization than the mean to date. Compared with the mean to date, the conditional mean prostate to date improved the volume overlap (intersection over union) in 32 out of 46 cases. Even compared with leave-one-out mean, the conditional mean prostate to date improved the volume overlap in 27 out of 46 cases. (46 is the total fractions of all patients from fraction 6.)

**Result 2: Segmentation** Since it was clear that the conditional mean to date outperforms the non-conditional mean to date, segmentations were initialized with the conditional means to date at fraction  $i$ . Segmentation proceeded with both conditional and non-conditional principal geodesics. Fig. 5.4 shows segmentation results in volume overlap and average surface distance: segmentation initialized with the conditional mean to date using conditional principal geodesics trained from models of fraction 1 to fraction  $i - 1$ , segmentation initialized with the conditional mean to date using non-conditional principal geodesics trained from models of fraction 1 to fraction  $i - 1$ , and segmentation initialized with a leave-one-out mean using non-conditional principal geodesics trained from all models except fraction  $i$ , at fraction  $i$ .

Table 5.3: Number of principal geodesics estimated from patient 3106 data

| Fraction No.    | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|-----------------|---|---|---|---|----|----|----|
| Conditional     | 0 | 1 | 2 | 3 | 4  | 5  | 6  |
| Non-conditional | 3 | 4 | 5 | 6 | 6  | 6  | 6  |

On average, the segmentation using the conditional distribution to date performed better from around fraction 9 than on the best leave-one-out segmentations so far. However, it

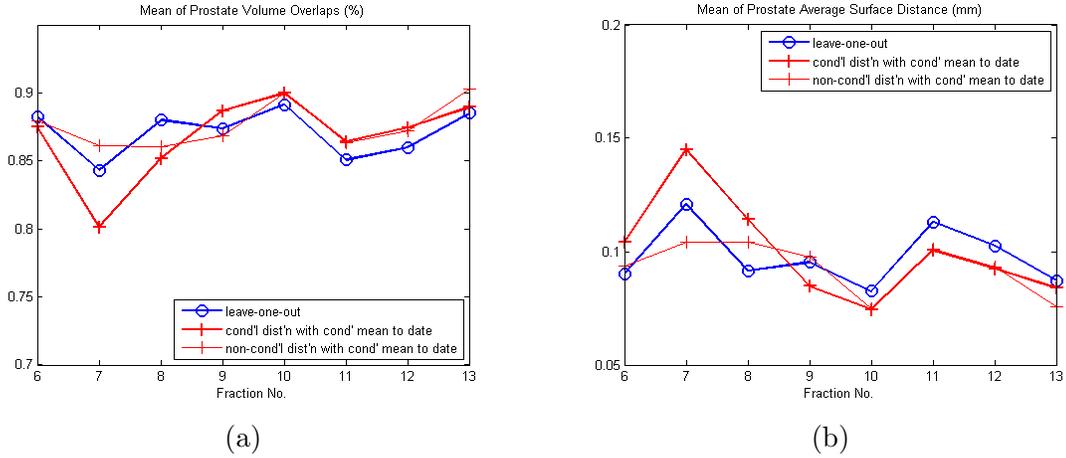


Figure 5.4: (a) Mean *Jaccard similarity coefficients* (*JSC*) of segmented prostate volume overlaps with training prostate binary images; (b) Mean of average surface distance in mm between segmented prostate and training prostate binary images. Mean of these measures are taken over 4 patients at each fraction, from fraction 6 to fraction 13. The thin blue line with the round marker indicates leave-one-out segmentation initialized with leave-one-out mean, the thick red line with the cross marker indicates successive segmentation starting with conditional mean to date and the conditional mean model deformed by conditional principal geodesics also trained up to a target fraction, and the thin red line with the round marker shows successive segmentation starting with the conditional mean to date but the conditional mean deformed by non-conditional principal geodesics of the prostate trained up to a target fraction.

was roughly equivalent to the segmentation using non-conditional principal geodesics, and it used the least number of eigenmodes. From fractions 6 to 8, the conditional distribution did worst, but for later fractions non-conditional and conditional distributions were superior to the leave-one-out non-conditional distributions.

The training sample size determines the number of possible principal geodesics (eigenmodes) used in estimating a probability distribution. During the training a threshold of  $10^{-10}$  was set for eigenvalues, and principal geodesics of higher eigenvalue were kept. The maximum number of eigenmodes to keep was set to 6. Table 5.3 shows the number of eigenvalues kept for patient 3106 under these conditions. The same pattern appears for the other patients. As can be seen in Table 5.3, at fraction 6 the number of principal geodesics is zero since the previous five fractions were not enough to estimate the conditional covariance matrix. The second term in the Eq. (5.2) for the covariance matrix of the CSPD is designed to correct the non-conditional covariance matrix using the reduced features of neighbor atoms. When there are too few training samples, the correction used to form the conditional probability distribution

inappropriately removes shape variability from the non-conditional probability distribution.

The segmentation results confirmed the importance of the initialization. When segmentation was initialized with the conditional mean to date and was proceeded with the non-conditional principal geodesics trained cumulatively, its results were better than the results of the leave-one-out segmentation.

## 5.4 Discussion and Conclusion

This chapter described a new method to estimate the shape probability distribution of an object conditioned on its neighboring objects. Using principal component regression analysis provides a way to get a stable and robust estimation of the conditional mean and shape distribution. Since this approach uses additional information extracted from the relation between the target object and its neighboring objects, it is capable of segmenting an object of insufficient boundary intensity information when good segmentations of neighboring objects are available.

The experiments reported in this chapter are limited to data within a patient, and the discussion here is mainly in the context of within patient cases.

Segmentation starting from conditional means also looks promising. However, as indicated in Table 5.3, the sample size is a critical factor in estimating conditional principal geodesics. While conditional means provided a better initialization than non-conditional means in the experiment with objects in the male-pelvis, the shape variation that the conditional principal geodesics capture was not adequate enough to further deform the initial conditional mean when the sample size was just enough to estimate the conditional means. In this connection it also remains as an issue how to determine the number of principal directions from neighboring atoms in estimating the conditional mean and covariance.

Another open issue is the choice of neighbor atoms. The choice affects the estimate of the conditional mean and conditional principal directions. These atoms can be chosen by looking at the correlations of every pair of atoms. However, the computational cost of this approach can be very high.

The most challenging issue is correspondence. Although the correspondence derived from the fitting mechanism works reasonably well to represent the shape of the object itself, it may

not be the right correspondence to describe the interrelation among the objects.

These last two issues - the selection of augmented atoms and the correspondence for multi-object complexes - of the conditional shape model are also issues shared by the multi-object statistical shape model presented in the previous chapter. For more discussion of these issues for cross-patient data, refer to section 4.5 in the previous chapter. In brief, the conditional shape model also needs a better method of augmentation that can handle correspondence of the augmented object across patients to be applicable to cross-patient data.

Finally, the initial models for bladder and rectum used in segmentation of the prostate were the training bladder and rectum models. It should be verified that the current initialization and segmentation results still hold when the training bladder and rectum models are replaced by good segmentations of the bladder and rectum.

# Chapter 6

## Discussion and Conclusion

This dissertation has made contributions both on the estimation of shape probability distributions on multi-object complexes and on the evaluation of statistical shape models. The first contribution is new methods to estimate a shape probability distribution on multi-object complexes. The two proposed methods allow the estimated shape probability distribution to characterize the inter-object relations using the methods of augmentation, prediction, principal component regression, and the concept of the residue. Effectiveness of these shape models as shape priors has been shown in the maximum posterior segmentation of m-reps. The second contribution is a novel measure and a procedure for evaluating statistical shape models in regard to their predictive power. The new measure can be applied to statistical shape models that use PCA to characterize the shape variability. It was extended to statistical shape models whose geometric representations are elements on a curved manifold. A novel procedure using the measure is also proposed to see the predictive power of the shape models as a function of the training sample sizes.

This chapter summarizes and discusses the contributions. Section 6.1 reviews the claims made in chapter 1. It is followed in section 6.2 and 6.3 by a discussion on issues related to the proposed methods and directions for future work.

### 6.1 Summary of Contributions

This section restates the claims and thesis in chapter 1 and discusses the contributions of this dissertation regarding the claims.

1. *It is shown that in multi-object statistical shape analysis augmentation of highly correlated geometric features of neighboring objects to the target object can be used to reflect interaction among objects in shape probability distribution of the target object.*

Augmentation is the combination of the geometric features of an object with the geometric features from the object's neighboring objects that have an effect on the shape of the object. In chapters 4 and 5 augmentation was used to reflect the inter-relations between objects on the probability distribution of multi-object complexes. In chapter 4 each object in a multi-object complex is augmented by adding to the geometric representation of the object some geometric primitives in its neighboring objects that are highly correlated with the object. Then augmentation was used as a means to estimate the joint p.d.f. of multi-object complexes in which the effect of neighboring objects on each object was estimated via prediction. In chapter 5 the target object is augmented with selected geometric features of its neighboring objects. The conditional probability distribution of the target object given the neighboring geometric primitives is estimated under the assumption of the normality of the augmented shape features of the target object.

The main reasons for using these selected geometric primitives instead of using the whole neighboring object in modeling the inter-relation is to ease the HDLSS problem that is already serious in a statistical shape model of a single object. Also, the reduction in the dimension of geometric shape feature space reduces the effect of noise in the estimation of shape probability distributions.

2. *A new method to estimate the shape probability distribution of an object conditioned on its geometric primitives of the neighboring objects has been developed. The method relies on principal component regression to have a stable and reliable estimation of the conditional mean and the conditional covariance matrix.*

Chapter 5 presented a method to estimate the conditional mean shape and covariance matrix of an object under the assumption that the object augmented with the features from neighboring objects follow the multivariate normal distribution. It was shown that when the feature space of the selected geometric primitives of the objects neighboring the target object is represented by the first few major principal components of the

neighboring geometric primitives, the conditional mean of the object given the principal components of the neighboring geometric primitives is equivalent to the response estimator of principal component regression on the object and the geometric primitives selected from the neighboring objects. It follows that the inverse covariance matrix of the selected geometric primitives in the neighboring objects that needs to be estimated to compute the conditional mean and covariance as in Eq. (5.1) and Eq. (5.2) can be replaced by the inverse covariance matrix of the principal components of the selected neighbor geometric primitives. In the HDLSS situation it is not possible to estimate the covariance matrix of the geometric primitives, let alone its inverse because of the number of parameters of the covariance matrix. However, the covariance matrix of the principal components can be estimated as long as the number of principal components are less than the sample size. The estimation with the inverse covariance matrix can be then made more stable by choosing only principal components with relatively large variances. This can help reducing noise in the estimation of the inverse covariance matrix.

3. *The conditional shape probability distribution can usefully provide a shape prior for maximum posterior segmentation with the conditional mean as the initial template model. This conditional shape prior was applied to segment anatomical objects in the male pelvis.*

As shown in chapter 5, the conditional shape probability distribution of multi-object complexes can be incorporated into maximum posterior segmentation of deformable models as a conditional shape prior. In the m-rep segmentation framework the conditional shape prior of m-reps can be incorporated as the principal geodesics of an estimated conditional covariance matrix. The geodesic Mahalanobis distance is used as the geometric prior in the objective function. The method to estimate a conditional mean and covariance matrix in chapter 5 was applied to two-object complex m-rep models of the bladder and the prostate in the male-pelvis area. The resulting conditional shape priors were used to segment both organs. The experiment was reported in section 5.3 of chapter 5. Its result showed that the conditional shape prior is effective in within-patient segmentations. Especially, the successive within-patient conditional mean of the  $k$ -th fraction given the first up to  $k - 1$ -th fractions of radiation treatment provides a better initialization than

the leave-one-out within-patient mean.

4. *A method to decompose shape variation of each object in multi-object complexes into two components has been developed on the basis of the concept of the residue.*

Shape variation of each object in multi-object complexes can be interpreted as having two components. One component is shape variation caused by the changes in the adjacent neighboring objects, and another component is variation independent of the effect from the neighboring objects. Chapter 4 presented an iterative method to separate out each component using the concept of residue. Borrowing the idea of the mixed model, the neighbor effect on the shape variation was understood in terms of a deterministic effect and a random effect.

5. *It is shown that the probability distributions on the decomposed shape variations described in chapter 4 can be incorporated as a shape prior for maximum posterior segmentation. These shape priors of the decomposed shape variations were applied to segment anatomical objects in the male pelvis.*

It was shown in chapter 4 that the probability distributions of the two components per object can be combined to describe the joint p.d.f. of multiple objects under a few assumptions on the statistical properties of the two components. The probability distributions of the two components are estimated iteratively, each iteration refining the estimations. When used as shape priors in maximum posterior segmentation, segmentation also proceeds iteratively among objects. These shape priors were applied to segment the bladder and the prostate in the male pelvis in the m-rep segmentation framework. The experiment and the result was described in section 4.4.5. Segmentation results were encouraging and showed that the proposed method is promising.

6. *A new correlation measure called goodness of prediction initially proposed by Muller [2007] in linear space has been extended for shape representations that live in a nonlinear manifold.*

Chapter 3 introduced a new correlation measure that can be used to evaluate a statistical shape model using PCA. The correlation measure originally proposed by [Muller, 2007]

is the ratio of the variance retained by the selected principal directions to the total variance of a set of training models. Geometrically interpreted, the numerator is the sum of the distances from training model approximations (Eq. (3.2)) to an estimated mean model, and the denominator is the sum of the distances from the training models to the mean model. The mean model is estimated from the same training models, and the approximations are also done in the shape space estimated from the same sample. This measures the predictability of the PCA method itself. However, what we are interested to measure is how well a learned probability distribution approximated by PCA can predict new cases that are not in the training set. To that end, the correlation measure proposed in chapter 3 is a modified version of the original correlation measure. Given two sets of models, one set is used to learn a probability distribution, i.e., mean and principal directions. The modified correlation measure is then defined as the proportion of summation of the distances from the approximated models from the other set to the estimated mean from the training set to summation of the distances from models from the other set to the mean. The key difference in the modified version is that the models are approximated in the shape space trained from a different set of models.

The limitation of the measure is that the measure is valid only for shape models that live in a linear space. The measure can be extended to evaluate shape models that live in a nonlinear symmetric space by replacing the Euclidean distance with the geodesic distance from the mean using the fact that the geodesic distance in a symmetric space is equal to the Euclidean distance in the corresponding tangent space at the training mean. This extension was described in detail in section 3.6.1.

7. *As a tool to evaluate the predictive power of statistical shape models, an iterative procedure that estimates the goodness of prediction in a given set of samples has been developed. This procedure provides a means to analyze the correlation measure for a given statistical shape model as a function of training sample size.*

Chapter 3 introduced a procedure using the correlation measure that can evaluate a statistical shape model as a function of sample sizes. The output of the procedure can be used to judge an approximate number of samples for a statistical shape model to have

a certain degree of predictability and stability.

The proposed correlation measure is defined in a shape feature space while other measures used to compare different shape models are defined in the ambient space  $\mathbb{R}_3$ . To verify that the correlation measure reflects what happens to the shape of an object, the correlation measure was compared with other measures such as surface distance and volume overlap using the procedure with various data sets. The results in chapter 3 confirmed that the correlation measure does reflect accuracy of the estimated shape in the ambient space  $\mathbb{R}_3$ .

*Thesis: (1) Reflecting interaction among objects in statistical shape models for multi-object complexes via augmentation yields a shape probability distribution that can capture the configuration of objects and shape variability caused by neighboring objects.*

Claim 1 shows that the simple idea of augmenting an object in a multi-object complex is a basis for the two statistical shape models for multi-object complexes proposed in this dissertation. Claim 2 pertains to a method to construct a conditional statistical shape model using PCR. Claim 3 shows that the conditional statistical shape model can be applied as a prior in a deformable m-rep segmentation framework, and the conditional mean can improve initialization of segmentation. Claims 4 and 5 pertain to an iterative method to decompose shape variation into two components: neighbor effects and self variations, and to the application of the probability distributions of these two components to the maximum posterior segmentation as shape priors. These claims prove the feasibility of capturing the arrangement of objects in multi-object complexes. They also indicate capturing the neighbor effect on shape change of each object in multi-object complexes and indicate the strong possibility to improve applications such as segmentation using the proposed shape models.

*(2) A systematic procedure to evaluate the predictive power of statistical shape models as a function of training sample size using the correlation measure provides a means to determine an approximate minimum size of sample that ensures a certain degree of predictive power and the stability.*

This dissertation raises the issue of evaluating statistical shape models and provides an evaluation method as to the predictability of statistical shape models. Claims 6 and 7 explain

a correlation measure and the use of the measure in a procedure that allows the evaluation of a statistical shape model as a function of sample sizes.

## 6.2 Evaluation of Statistical Shape Models

### 6.2.1 Discussion

A major problem of the evaluation method is that understanding the proposed correlation measure requires a large number of samples, which is the inherent limitation of any statistical estimation method. Applying the measure to an real anatomical shape can be hampered by this limitation due to the difficulty in acquiring many training samples of the anatomical object.

Additional concerns are brought out in the uses of the prediction measure. The prediction measure was initially intended for 1) comparing two different geometric shape models: PCA on b-reps and PGA on m-reps in terms of their predictive power and 2) comparing the new statistical shape models for multi-object complexes with previous models. However, there are serious difficulties involved in using the correlation measure for both comparisons.

In using the correlation measure for the comparison between m-rep with PGA and PDM with PCA, there are several serious issues that are difficult to resolve. The first issue is related to obtain training models that have an optimal correspondence for each geometric model. One experiment in chapter 3 used simulated deformed ellipsoid m-reps to derive b-reps at the m-reps' spoke ends. The three deformations of bending, twisting, and tapering were applied directly to a base ellipsoid m-rep, so the generated simulated m-reps have almost ground-truth optimal correspondence. This is clearly indicated in the plot of eigenvalues in Fig. 6.1 (a). The eigenvalue plot of the simulated m-reps on the left clearly distinguishes the input three deformations while the eigenvalue plot of the derived b-reps on the right shows the mixing of the three deformations and the spread of the deformations in five eigenvalues. The corresponding box plots in Fig. 6.1 (b) shows that in terms of the  $\rho^2$  measure the simulated ellipsoid m-reps with 3 eigenmodes are comparable to the derived b-reps with 6 eigenmodes. The data is biased toward m-rep representations because m-reps have a better correspondence than b-reps in this data.

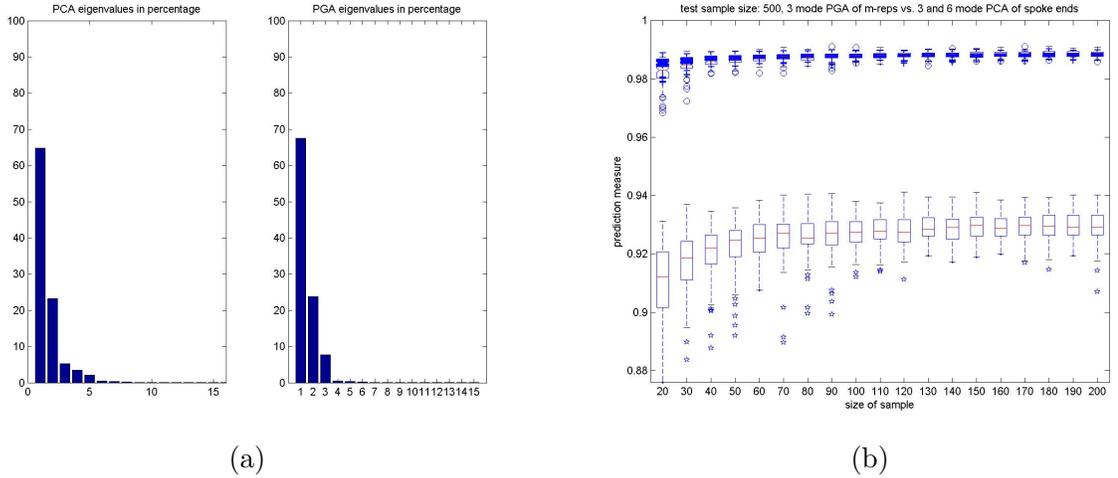
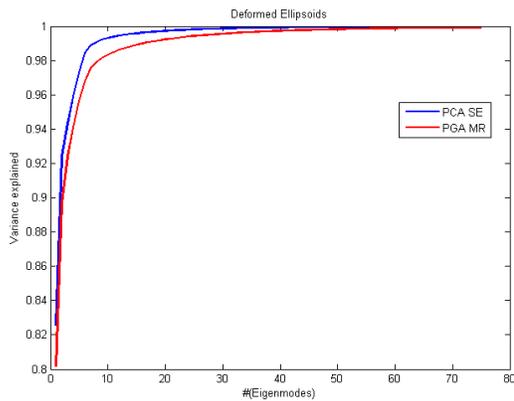


Figure 6.1: (a) Eigenvalue plots of PGA on the simulated ellipsoid m-reps and PCA on b-reps derived from the m-reps. (b) Box plots of  $\rho^2$  on the simulated ellipsoid m-reps and the derived b-reps.

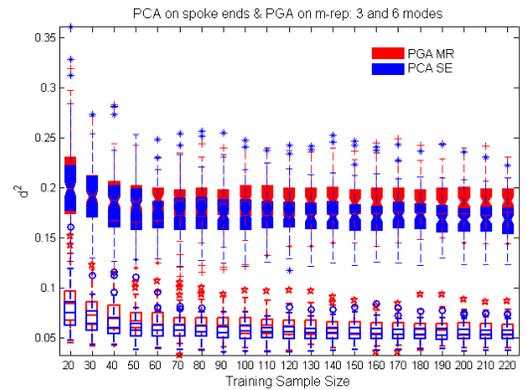
Another data set used in the evaluation of the correlation measure is m-rep fits to the deformed ellipsoid binaries and the derived b-reps from the m-rep fits. The fit was done following the procedure described in section 3.4.3. When PCA on b-reps and PGA on m-reps are compared with this data set, fits to the deformed ellipsoid binaries, the result with the  $\rho^2$  measure is the opposite to the result obtained with the previous data. The box plots of  $\rho^2$  for PCA on b-reps formed by m-reps' spoke ends (PCA SE) and PGA on m-reps (PGA MR) showed that PCA on the b-reps have higher median  $\rho^2$  values than PGA on the m-reps. Fig. 6.2 illustrates the result with the eigenvalue plot and the box plots of  $D_p^2$  of PCA on b-reps and PGA on m-reps. The box plots of  $D_p^2$  are consistent with those of  $\rho^2$  though box plots of  $\rho^2$  are not shown here. As for the hippocampus data set that was used in chapter 3 the comparison result showed that the PGA on m-reps gave lower media  $\rho^2$  values than PCA as shown in Fig. 6.3. (Also, refer to Fig. 3.6 and Fig. 3.12 in chapter 3.)

As shown with these examples, when one wants to compare two different object representations it is hard to know how to generate training models of two representations that are comparable in terms of correspondence and fits to the binaries. Thus, we cannot draw any conclusion from these results as to the advantage of one model over the other shape model.

There are also other issues involved in comparing statistical shape models in different

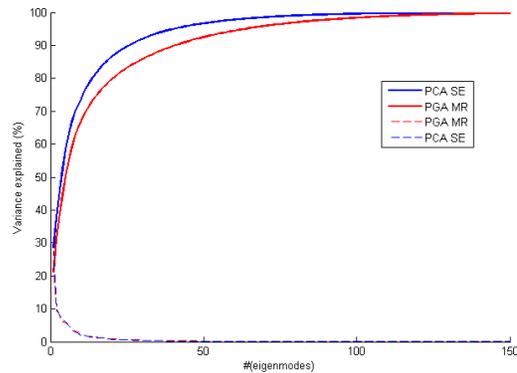


(a)

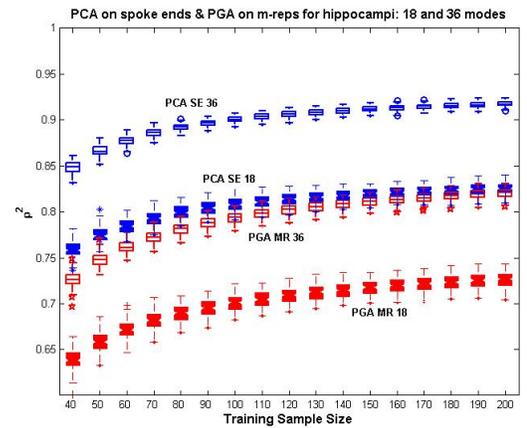


(b)

Figure 6.2: (a) Eigenvalue plots of PGA on the ellipsoid m-rep fits to deformed ellipsoid binaries and PCA on b-reps derived from the ellipsoid m-rep fits. (b) Two box plots of  $D_p^2$  for PCA on warped ellipsoid b-reps and for PGA on warped ellipsoid m-reps.



(a)



(b)

Figure 6.3: (a) Eigenvalue plots of PCA on hippocampus b-reps and of PGA on the corresponding hippocampus m-rep. (b) Two box plots of  $\rho^2$  for PCA on hippocampus b-reps and for PGA on hippocampus m-reps.

geometric representations. The correlation measure itself can be biased toward random vectors of small inherent dimension. M-reps carry surface normal information at its implied boundary points while b-reps must be interpolated to approximate the surface normal at the boundary points. The information contained in b-reps is not sufficient to restore corresponding m-reps. Thus, the inherent dimension of m-reps is likely higher than that of b-reps.

Also, there is an issue of a distance metric. M-reps and b-reps belong to mathematically different spaces, so the metric used to compute the distance between elements in the space are different. The distance metric determines the estimated principal directions and geodesics and thus the shape space spanned by these directions. The projection operation [Fletcher, 2004] involved in computing  $\rho^2$  (refer to Eq. (3.10)) depends on the distance metric. Approximation of models in shape space is done through the projection operation: the approximation in b-reps is projection in Euclidean space that minimizes Euclidean distance, and the approximation in m-reps is projection in nonlinear Riemannian space that minimizes geodesic distance.

Fig. 6.4 compares 50 projections of m-rep fits to the deformed binaries on a shape space spanned by 8 principal geodesics and 50 projections of their derived b-reps on a shape space spanned by 8 principal directions. This test was to see whether the shape space represented by principal geodesics describes approximately the same shapes that the shape space represented by principal components describes. We can see that the surface distances between those projected models both onto m-rep shape space and onto b-rep shape space, and the corresponding truth (training models) are comparable. This fact implies that shapes that both spaces spanned by 8 principal modes of variation cover are comparable although the result indicated in the box plot of  $\rho^2$  measures with 3 and 6 modes indicates that the shape space that principal directions describe have the better predictive power than the shape space that principal geodesics describe. Also, this test raises the issue of the selection of principal modes that are justifiable for comparison.

There is also an issue specifically related to m-reps. PGA has two definitions: the maximum variance definition and the least-squares definition. In linear space, i.e., for PCA, both definitions are equivalent. These two definitions for PGA are not equivalent, and we have taken the maximum variance definition for PGA. The shape space that PGA produces is the space where the projections of training models have maximum variance while the shape space

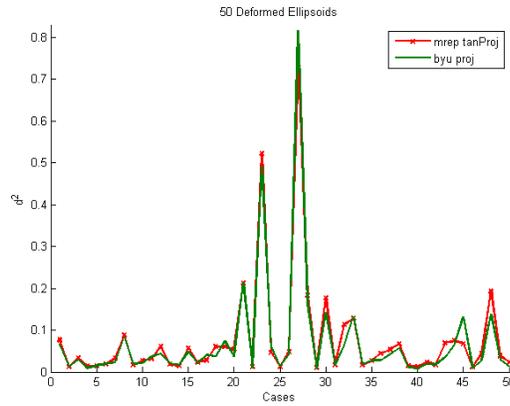


Figure 6.4: Comparison of projections of 50 m-rep fits to deformed binary ellipsoids on a shape space formed by 8 principal geodesics and projections of 50 b-reps on a shape space formed by 8 principal directions

that PCA produces is the space where the distance between models and their projections are minimized. In this respect the prediction measure can really favor PCA on b-reps since the least-squares property of PCA shape space makes  $\rho^2$  closer to 1. (Refer Fig. 3.9 in chapter 3.)

In addition there is no analytic solution to compute the projection operation onto an m-rep shape space, and an approximation to the definition of the projection operation is used as explained in [Fletcher, 2004]. A test was done to see whether the approximate projection operation produces models that are really close to the minimum geodesic distance from the original models to the projected models and whether there exists a model in the estimated shape space of m-reps that have a better surface distance. Fig. 6.5 shows the results of the test. Fig. 6.5 (a) illustrates the comparison of projections of 50 m-rep fits to deformed binary ellipsoids on shape space formed by 8 principal geodesics with m-reps optimized in geodesic distance from the mean; it shows that for these 50 cases the approximate projection operation produces models that have roughly minimum distances from the original models.

Another test was done to see whether there exist models in the estimated shape space of m-reps that have better surface distances than the projected models. Fig. 6.5 (b) illustrates the comparison of projections of 50 m-rep fits to deformed binary ellipsoids on shape space formed by 8 principal geodesics with m-reps further optimized in Euclidean distance starting from the mean m-reps. The result shown in Fig. 6.5 (b) shows that at least some models among these 50 cases in the estimated shape space of m-reps have better surface distances

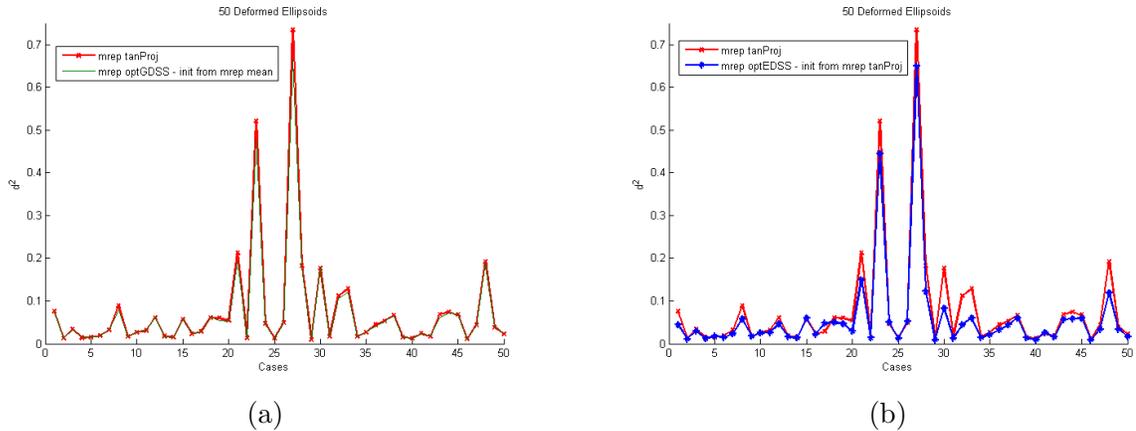


Figure 6.5: Comparison of projections of 50 m-rep fits to deformed binary ellipsoids on a shape space formed by 8 principal geodesics (a) with m-reps optimized in geodesic distance from the mean m-reps, (b) with m-reps further optimized in Euclidean distance starting from the mean m-reps.

than corresponding projected models.

As for the second goal of using the prediction measure to compare the two statistical shape models for multi-object complexes proposed in this dissertation with other models described in chapter 4, there are issues that are specific to each shape model. The problem in applying the prediction measure to the statistical shape model for multi-object complexes that estimates self variations and neighbor effects lies in the interpretation of the measure in terms of probability distributions of the self variation and the neighbor effect per object. The measure itself can be computed blindly (Eq. (3.9)) since the mean and the approximations (or projections) of the models can be respectively computed by adding these two components: self and neighbor mean; self and neighbor projections of each model. However, it is not clear what is a valid shape space in which the measure can be interpreted. Also, it is not guaranteed that the value of the measure will be within a  $[0, 1]$  interval.

The problem in applying the prediction measure to the conditional shape model is that the estimated shape space is centered at the conditional mean per model which changes across test models. The conditional mean per model can replace the mean in Eq. (3.9), and the measure can be computed as well. However, the modified measure should be examined to understand the meaning and interpretation in the context of statistical shape models. More work needs to be done to use the correlation measure for the comparison between the proposed

statistical shape models of multi-object complexes and the previous approaches for multi-object complexes listed in chapter 1.1.1.

## 6.2.2 Future Work

As described in the previous section, the current measure is not applicable to comparing the proposed approaches and the previous approaches to multi-object statistical shape models. One area for future work is, if possible, to extend or modify the current measure to evaluate the proposed multi-object statistical shape models.

Also, the distribution of the measure needs to be examined. It is assumed that the distribution is unimodal, and the median of the measure is used as a summary estimate of the measure. We can also draw other information on the statistical shape models from the distribution of the measure. For example, the variance of the measure at a sample size can show how reliable the median of the estimated measures can be in its indication of the predictive power of a shape model.

It will also be interesting to explore other possible correlation measures mentioned in [Muller, 2007] such as squared multiple correlation, which can provide more detailed information.

## 6.3 Multi-Object Statistical Shape Models

### 6.3.1 Discussion and Future Work

There are several areas of future work to improve the proposed statistical shape models for multi-object complexes. First, a systematic way that identifies geometric primitives (medial atoms) in the neighboring object needs to be developed. Currently, neighboring atoms to augment the target object are handpicked on the basis of the Euclidean distance from the atoms to the target object because of the preliminary evidence that those nearby atoms are highly correlated. Although this manual selection of the neighboring atoms works reasonably well when training data are obtained from a single patient, there are some issues that this simple manual approach cannot handle. The issues become more difficult to handle when the training data are from multiple patients.

One issue is correspondence. During the model fitting stage described in section 2.5 the

correspondence enforced on the training models is focused on minimizing the noise in shape statistics learned later from the generated training models while the training models capture the shape of each organ. However, this correspondence induced during the fitting stage may not be ideal for capturing the inter-object relation among the objects. This issue requires coming up with a new set of criteria for a new correspondence to capture inter-object relation and generating a new set of training models that can still represent the shape of the object across the training cases as well as the original training models do. M-reps have an advantage regarding this problem. As pointed out in section 2.5, the same boundary of an object can be captured by multiple layouts of medial atoms. One possible approach to this problem suggested by S. Pizer is to generate new training models from the original training models by interpolating the original training models using the new set of criteria for the correspondence Han et al. [2006].

Another issue is the selection of neighboring atoms to augment the target object for cross-patient data. As seen in the multi-object complex models in male-pelvis area, the position and the size of the region in which one organ is near the other organ vary across patients. As discussed in section 4.5, it is not clear how to select the neighboring atoms of correspondence across patients and how to decide the number of neighboring atoms to augment the target object so that the selected neighboring atoms are neither redundant nor insufficient to cover the nearby regions across the different patients.

Second, another area for future work is to develop a way to handle the non-transitivity in applying the neighbor effect on an object. As illustrated in Fig. 4.2 in section 4.2.1, the set of primitives  $\mathbf{T}$  that changes by the neighbor effects can apply to either all the primitives in the object or to selected primitives of the object. In the latter case the non-selected primitives of the object are ignored when the neighbor effect on the object is estimated, which can introduce an abrupt change in shape, i.e., a non-smooth boundary between the selected and non-selected area after the neighbor effect is applied to the selected primitives.

Third, a method to initialize self variations and neighbor effects needs to be developed. In training the self variations and neighbor effects of the bladder and the prostate in the male-pelvic area, the anatomical properties of each organ were made use of to initialize the self variations and neighbor effects. The main shape change of the bladder is caused by the

filling of urine. Although the shape of the bladder can easily be affected by its neighboring organs since it is basically a sack of fluid, the effect is local to the bladder. Thus, it was fairly clear that for the bladder the self change dominates the effect from neighboring organ. On the contrary, the shape of the prostate rarely changes because it is a firm organ. Most changes that happen in the prostate are rotational and transitional movement of the organ. On the basis of these facts the neighbor effect of the bladder was initialized to zero, and the self variation of the prostate was initialized to zero. However, for some organs we may not have good understanding about their properties as we have for the bladder and the prostate. It will be important to have a good method to initialize these variations since the initialization makes a serious impact on the resulting estimations.

Finally, using other real data the effectiveness of the iterative method and conditional method for multi-object complexes needs to be further validated and compared.

# Bibliography

- Ravikanth Malladi, James A. Sethian, and Baba C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2):158–175, 2 1995a. 2, 37
- V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *iccv95*, pages 694–699, Los Alamitos, CA, 1995. IEEE Computer Society Press. 2, 37
- S. Joshi. *Large Deformation Diffeomorphisms and Gaussian Random Fields for Statistical Characterization of Brain SubManifolds*. PhD thesis, Dept. of Electrical Engineering, Sever Institute of Technology, Washington Univ., 8 1997. 2
- Gary E. Christensen, Sarang C. Joshi, and Michael I. Miller. Volumetric transformation of brain anatomy. *IEEE Trans. Med. Imaging*, 16(6):864–877, 1997. 2
- R. Malladi, J. A. Sethian, and B. C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(2):158–175, 1995b. 2
- A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, E. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transactions on Medical Imaging*, 22(2):137–153, 2 2003. 2
- T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham. Active shape models – their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1 1995. 2, 23, 35, 39, 55
- S.M. Pizer, T. Fletcher, Y. Fridman, D.S. Fritsch, A.G. Gash, J.M. Glotzer, S. Joshi, A. Thall, G. Tracton, P. Yushkevich, and E.L. Chaney. Deformable m-reps for 3d medical image segmentation. *International Journal of Computer Vision - Special UNC-MIDAG issue, (O. Faugeras, K. Ikeuchi, and J. Ponce, eds.)*, 55(2):85–106, 2003. 2, 48, 52, 84
- John Shawe-Taylor and Nello Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, 2004. 3, 6, 113, 114
- C. Lu, S.M. Pizer, S. Joshi, and Ja-Yeon Jeong. Statistical multi-object shape models. *International Journal of Computer Vision*, 2007. 3, 7, 82, 86, 92
- K. E. Muller. Goodness of prediction for principal components, including high dimension, low sample size. work in progress, 2007. 3, 9, 12, 26, 56, 58, 59, 128, 137
- T. Kapur, P.A. Beardsley, S.F. Gibson, W.E.L. Grimson, and W.M. Wells. Model based segmentation of clinical knee MRI. In *Model-based 3D Image Analysis workshop (in conjunction with ICCV)*, pages 156–162, 1998. 5, 83
- M. Vaillant and C. Davatzikos. Hierarchical matching of cortical features for deformable brain image registration. In *Information Processing in Medical Imaging (IPMI)*, pages 182–195, 1999. 5, 83

- C. Davatzikos, T. Xiaodong, and D. Shen. Hierarchical active shape models, using the wavelet transform. *IEEE Transactions on Medical Imaging*, 22(3):414–423, 2003. 5, 77, 83
- Kathryn Leonard. Efficient shape modeling: -entropy, adaptive coding, and boundary curves -vs- blums medial axis. *International Journal of Computer Vision*, 74(2):183–199, August 2007. ISSN 0920-5691. doi: 10.1007/s11263-006-0010-3. URL <http://dx.doi.org/10.1007/s11263-006-0010-3>. 7
- Martin A. Styner, Kumar T. Rajamani, Lutz-Pter Nolte, Gabriel Zsemlye, Gabor Szekely, Christ J. Taylor, and Rhodri H Davies. Evaluation of 3d correspondence methods for model building. In *Information Processing in Medical Imaging (IPMI)*, 2003. 8, 9
- P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, 8 2004. 10, 50, 118
- Xiaoxiao Liu, Ja-Yeon Jeong, Joshua H. Levy, Rohit Saboo, Edward L. Chaney, and Stephen M. Pizer. A large-to-fine-scale shape prior for probabilistic segmentations using a deformable m-rep. In *IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, 2008. 10, 77, 78
- William R. Crum, Oscar Camara, and Derek L. G. Hill. Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transactions on Medical Imaging*, 25(11):1451–1461, 11 2006. 10, 78, 119
- Robert V. Hogg and Allen T. Craig. *Introduction of Mathematical Statistics*. Prentice-Hall, 5th edition, 1995. 13, 14
- Robb J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, 1982. 13, 20, 21, 22, 23, 26, 112
- Steven F. Arnold. *The Theory of Linear Models and Multivariate Analysis*. Wiley, New York, 1981. 20, 26
- H. Hotelling. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, 23:417–441, 498–520, 1933. 22
- L. Tucker and MacCallum R. Exploratory factor analysis, 1993. URL <http://www.unc.edu/~rcm/book/factornew.htm>. A Book Manuscript. 23
- Keith E. Muller and Paul W. Stewart. *Linear Model Theory: Univariate, Multivariate, and Mixed Models*. Wiley-Interscience, 2006. 23, 26, 27
- Fred L. Bookstein. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, 1999. 23, 31, 55
- I. T. Jolliffe. *Principal Components Analysis*. Springer Verlag, New York, 1986. 23
- P. Thomas Fletcher. *Statistical Variability in Nonlinear Spaces: Application to Shape Analysis and DT-MRI*. PhD thesis, Dept. of Computer Science, UNC-CH, 5 2004. 23, 32, 89, 134, 135
- David G. Kleinbaum, Lawrence L. Kupper, Keith E. Muller, and Azhar Nizam. *Applied Regression Analysis and Multivariable Methods*. Duxbury Press, 1997. 26

- Neil H. Timm. *Applied Multivariate Analysis*. Springer, 2002. 26
- A. Srivastava, S. H. Joshi, W. Mio, and Xiuwen Liu. Statistical shape analysis: clustering, learning, and testing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(4):590–602, 2005. doi: <http://dx.doi.org/10.1109/TPAMI.2005.86>. URL <http://dx.doi.org/10.1109/TPAMI.2005.86>. 29
- D’Arcy Wentworth Thompson and John Tyler Bonner. *On Growth and Form*. Cambridge University Press, 1992. 29
- D. G. Kendall. A survey of the statistical theory of shape. *Statistical Science*, 4(2):87–120, 1989. 30, 31
- F.L. Bookstein. Size and shape spaces for landmark data in two dimensions. *Statistical Science*, 1:181–242, 1986. 31, 38
- Fred L. Bookstein. [a survey of the statistical theory of shape]: Comment. *Statistical Science*, 4(2):99–105, 1989. ISSN 08834237. doi: <http://dx.doi.org/10.2307/2245332>. URL <http://dx.doi.org/10.2307/2245332>. 31
- Christopher G. Small. *The Statistical Theory of Shape*. Springer, 8 1996. 31
- I.L. Dryden and K. Mardia. *Statistical Shape Analysis*. Wiley, 1998. 31
- D. G. Kendall. Shape manifolds, procrustean metrics and complex projective spaces. *Bull. London Math. Soc.*, 16(2):81–121, 1984. 31
- J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40:33–51, 1975. 34
- M. de Bruijne, M. T. Lund, L. B. Tanko, P. P. Pettersen, and M. Nielsen. Quantitative vertebral morphometry using neighbor-conditional shape models. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 1–8, 2006. 34, 114
- Kevin Gorczowski, Martin Styner, Ja-Yeon Jeong, J. S. Marron, Joseph Piven, Heather Cody Hazlett, Stephen M. Pizer, and Guido Gerig. Statistical shape analysis of multi-object complexes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 6 2007. 34, 35
- Joshua H. Levy. *Refinement of Object-Based Segmentation*. PhD thesis, Dept. of Computer Science, UNC-CH, 8 2008. 35
- Paul A. Yushkevich. *Statistical Shape Characterization Using the Medial Representation*. PhD thesis, Dept. of Computer Science, UNC-CH, 5 2003. 36
- Rhodri H. Davies, Carole J. Twining, Tim F. Cootes, John C. Waterton, and Chris J. Taylor. A minimum description length approach to statistical shape modeling. *IEEE Transactions on Medical Imaging*, 21:525–537, 2002a. 36
- Rhodri H. Davies, Carole J. Twining, Tim F. Cootes, John C. Waterton, and Chris J. Taylor. 3d statistical shape models using direct optimisation of description length. pages 3–20. Springer, 2002b. 36

- Carole Twining, Rhodri Davies, and Chris Taylor. Non-parametric surface-based regularisation for building statistical shape models. In *Information Processing in Medical Imaging (IPMI)*, pages 738–750, 7 2007. 36
- Ch. Brechbühler, G. Gerig, and O. Kübler. Parametrization of closed surfaces for 3-d shape description. *Comput. Vis. Image Underst.*, 61(2):154–170, 1995. ISSN 1077-3142. doi: <http://dx.doi.org/10.1006/cviu.1995.1013>. 36, 37, 40, 41
- Tim McNerney and Demetri Terzopoulos. Deformable models in medical image analysis: A survey. *Medical Image Analysis*, 1(2):91–108, 1996. 37
- A.H. Barr. Superquadrics and angle-preserving transformations. *IEEE Computer Graphics and Applications*, pages 11–23, 1 1981. 37
- F.P. Ferrie, J. Lagarde, and P. Whaite. Darboux frames, snakes, and super-quadrics: Geometry from the bottom up. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (8): 771–783, 8 1993. 37
- D. Terzopoulos. Dynamic 3d models with local and global deformations: Deformable superquadrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (7):703–714, 7 1991. 37
- M. Kass, A. Witkin, and Terzopoulos D. Snakes: Active contour models. *International Journal of Computer Vision*, pages 321–331, 1988. 37
- Richard Szeliski. Bayesian modeling of uncertainty in low-level vision. *Int. J. Comput. Vision*, 5(3):271–301, December 1990. ISSN 0920-5691. doi: 10.1007/BF00126502. URL <http://portal.acm.org/citation.cfm?id=106014.106017>. 37
- Kaleem Siddiqi and Stephen M. Pizer. *Medial Representations: Mathematics, Algorithms and Applications (Computational Imaging and Vision)*. Springer-Verlag Berlin Heidelberg, 2008. 38, 43, 44, 45, 47, 49, 54
- Derek Merck, Gregg Tracton, Rohit Saboo, Joshua Levy, Edward Chaney, Stephen Pizer, and Sarang Joshi. Training models of anatomic shape variability. *Medical Physics*, 35(7), 2008. 38, 49, 51, 94, 95, 103, 119
- Q. Han, D. Merck, J. H. Levy, C. Villarruel, E. L. Chaney, and S. M. Pizer. Geometrically proper models in statistical training. In *Proc. of Information Processing in Medical Imaging, (Nico Karssemeijer and Boudewijn Lelieveldt, eds.)*, volume 4584, pages 751–762, 2007. 38, 49, 51, 64, 74
- S. Joshi and M.I. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Image Processing*, 9:1357–1370, 2000. 38
- D. G. Kendall. The diffusion of shape. *Advances in Applied Probability*, 9(3):428–430, 1977. ISSN 00018678. URL <http://www.jstor.org/stable/1426091>. 38
- C.G. Small. Techniques of shape analysis on sets of points. *Internat. Statist.*, 56:243–257, 1988. 38
- T. F. Cootes and C. J. Taylor. Statistical models of appearance for medical image analysis and computer vision. In *Proc. SPIE Medical Imaging*, 2001. URL [citeseer.ist.psu.edu/cootes01statistical.html](http://citeseer.ist.psu.edu/cootes01statistical.html). 39

- L. H. Staib and J. S. Duncan. Boundary finding with parametrically deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (11):1061–1075, 11 1992. 40
- F. P. Kuhl and C. R. Giardina. Elliptic fourier features of a closed contour. In *Comput. Graphics Image Processing*, pages 236–258, 1982. 40
- G. Szekely, A. Kelemen, C. Brechbuhler, and Gerig G. Segmentation of 2-d and 3-d objects from mri volume data using constrained elastic deformation of flexible fourier contour and surface models. *Medical Image Analysis*, 1(1):19–34, 1996. 41
- András Kelemen, Gábor Székely, and Guido Gerig. Elastic model-based segmentation of 3-d neuroradiological data sets. *IEEE Transaction on Medical Imaging*, 18(10), 10 1999. 41
- N. Amenta, S. Choi, and R. Kolluri. The power crest, unions of balls, and the medial axis transform. *Computational Geometry: Theory and Applications*, 19(2-3):127–153, 2001. 43
- B. B. Kimia, A. Tannenbaum, and S. W. Zucker. Shape, shocks and deformations i: The components of two-dimensional shape and the reaction-diffusion shape. *International Journal of Computer Vision*, 15:189–224, 1995. 43
- H. Blum. *A transformation for extracting new descriptors of shape*. MIT Press, 1967. 43
- H. Blum and R. N. Nagel. Shape description using weighted symmetric axis features. *Pattern Recognition*, 10:167–180, 1978. 43, 46
- J. Bloomenthal and K. Shoemake. Convolution surfaces. In *Computer Graphics (SIGGRAPH '91)*, volume 25, pages 251–256, 1991. 43
- T. Igarashi, S. Matsuoka, and H. Tanaka. Teddy: A sketching interface for 3d freeform design. In *SIGGRAPH '99*, pages 409–416, 1999. 43
- P. J. Giblin and B. B. Kimia. A formal classification of 3d medial axis points and their local geometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):238–251, 2004. 43, 44, 45
- James Damon. Smoothness and geometry of boundaries associated to skeletal structures i: Sufficient conditions for smoothness. *Annales de Institut Fourier*, 53(6):1941–1985, 2003. 43, 52
- James Damon. Smoothness and geometry of boundaries associated to skeletal structures ii: Geometry in the blum case. *Composito Mathematica*, 140(6):1657–1674, 2004. 43
- Barrett O’Neill. *Elementary Differential Geometry*. Elsevier, 1997. 46
- Paul Yushkevich, P. Thomas Fletcher, Sarang Joshi, Andrew Thall, and Stephen M. Pizer. Continuous medial representations for geometric object modeling in 2D and 3D. Technical report TR02-003, University of North Carolina, Chapel Hill, 2002. 48
- Paul A. Yushkevich, Hui Zhang 0005, and James C. Gee. Parametric medial shape representation in 3-d via the poisson partial differential equation with non-linear boundary conditions. In *IPMI*, pages 162–173, 2005. 48

- Timothy B. Terriberry and Guido Gerig. A continuous 3-d medial shape model with branching. In *Proc. of the International Workshop on Mathematical Foundations of Computational Anatomy (MFCA 06)*, 10 2006. 48
- Robert Broadhurst, Joshua Stough, Stephen Pizer, and Edward Chaney. A statistical appearance model based on intensity quantiles histograms. In *IEEE International Symposium on Biomedical Imaging*, pages 422–425, 4 2006. 53, 120
- S.M. Pizer, Ja-Yeon Jeong, R. Broadhurst, S. Ho, and J. V. Stough. Deep structure of images in populations via geometric models in population. In *International Workshop on Deep Structure, Singularities and Computer Vision (DSSCV)*, (O.F. Olsen, L. Florack, and A. Kuijper, eds.), LNCS, 6 2005a. 53, 82, 96
- Robert Broadhurst, Joshua Stough, Stephen Pizer, and Edward Chaney. Histogram statistics of local image regions for object segmentation. In *International Workshop on Deep Structure, Singularities and Computer Vision (DSSCV)*, 4 2005. 53, 96
- Elizaveta Levina. The earthmovers distance is the mallows distance: Some insights from statistics. In *IEEE International Conference on Computer Vision*, pages 251–256, 7 2001. 53
- Ja-Yeon Jeong, Surajit Ray, Qiong Han, Xiaoxiao Liu, Keith E. Muller, and Stephen M. Pizer. Goodness of prediction for principal components of shape: A correlation measure. Submitted to *International Journal of Computer Vision*, 2008a. 55
- A Thall. *Deformable Solid Modeling via Medial Sampling and Displacement Subdivision*. PhD thesis, Dept. of Computer Science, University of North Carolina, 8 2004. 63
- Nicolas Aspert, Diego Santa-Cruz, and Touradj Ebrahimi. Mesh: Measuring errors between surfaces using the hausdorff distance. In *IEEE International Conference in Multimedia and Expo*, volume 1, pages 705–708, 2002. URL <http://mesh.epfl.ch/>. 70
- Delphine Nain, Steven Haker, Aaron Bobick, and Allen Tannenbaum. Multiscale 3d shape analysis using spherical wavelets. In *Medical Image Computing and Computer Assisted-Intervention*, 2005. 77
- S.M. Pizer, Ja-Yeon Jeong, C. Lu, K. Muller, and S. Joshi. Estimating the statistics of multi-object anatomic geometry using inter-object relationships. In *International Workshop on Deep Structure, Singularities and Computer Vision (DSSCV)*, (O.F. Olsen, L. Florack, and A. Kuijper, eds.), Springer LNCS, pages 60–71, 6 2005b. 82
- Ja-Yeon Jeong, Stephen M.Pizer, and Surajit Ray. Statistics on anatomic objects reflecting inter-object relations. In *Mathematical Foundations of Computational Anatomy*, 10 2006. 82
- K.M. Pohl, J. Fisher, J.L. Levitt, M.E. Shenton, R. Kikinis, W.E.L. Grimson, and W.M. Wells. A unifying approach to registration, segmentation, and intensity correction. In *Medical Image Computing and Computer Assisted-Intervention*, pages 310–318, 2005. 83
- S. Ray, S. Simpson, and K. Muller. Structural covariance for multi-object image segmentation. 2006. 87

- Sungkyu Jung and Steve Marron. Spatial dependency of bladder based on reparameterization of m-reps, 2008. Presentation material at the internal seminar. 87, 110
- K. Rajamani, S. Joshi, and M. Styner. Bone model morphing for enhanced surgical visualization. In *IEEE International Symposium on Biomedical Imaging*, pages 1255–1258, 4 2004. 88
- James Damon. Determining the geometry of boundaries of objects from medial data. *International Journal of Computer Vision*, 63(1):45–64, 6 2005. 94
- S.M. Pizer, P.T. Fletcher, S. Joshi, A.G. Gash, J. Stough, A.Thall, G. Tracton, and E.L. Chaney. A method and software for segmentation of anatomic object ensembles by deformable m-reps. *Medical Physics (William Hendee, ed.)*, 32(5):1335–1345, 5 2005c. 96
- E.L. Chaney, S.M. Pizer, S. Joshi, R. Broadhurst, P.T. Fletcher, A.G. Gash, Q. Han, Ja-Yeon Jeong, C. Lu, D. Merck, J. V. Stough, G. Tracton, J. Bechtel, J. Rosenman, YY. Chi, and K. E. Muller. Automatic male pelvis segmentation from ct images via statistically trained multi-object deformable m-rep models. In *Presented at American Society for Therapeutic Radiology and Oncology*, 2004. 96
- Q. Han, C. Lu, G. Liu, S.M. Pizer, S. Joshi, and A. Thall. Representing multi-figure anatomical objects. In *IEEE International Symposium on Biomedical Imaging*, pages 1251–1254, 4 2004. 97
- Q. Han, S.M. Pizer, D. Merck, S. Joshi, and Ja-Yeon Jeong. Multi-figure anatomical objects for shape statistics. In *Information Processing in Medical Imaging (G. Christensen and M. Sonka, eds.)*, pages 701–712, 7 2005. 97
- M. Davidian. Applied longitudinal data analysis, 2001. Lecture Notes at Department of Statistics, North Carolina State University. 98
- Ja-Yeon Jeong, Joshua V. Stough, J. Stephen Marron, and Stephen M. Pizer. Conditional-mean initialization using neighboring objects in deformable model segmentation. In *Proc. SPIE Medical Imaging*, 2008b. 111
- Rohit R Saboo, Joshua H. Levy, Edward M Chaney, and Stephen M Pizer. Modeling populations of nearly tubular objects through medial models. Technical report, 2008. 118
- Qiong Han, Stephen M. Pizer, and James N. Damon. Interpolation in discrete single figure medial objects. In *IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2006)*, 7 2006. 138