Refinement of Object-Based Segmentation

Joshua Howard Levy

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
2008

Approved by:

Stephen M. Pizer, Advisor

Edward L. Chaney, Reader

Mark Foskey, Reader

J. Stephen Marron, Reader

Martin A. Styner, Reader

© 2008
Joshua Howard Levy
ALL RIGHTS RESERVED

ABSTRACT

Joshua Howard Levy: Refinement of Object-Based Segmentation
(Under the direction of Stephen M. Pizer)

Automated object-based segmentation methods calculate the shape and pose of anatomical structures of interest. These methods require modeling both the geometry and object-relative image intensity patterns of target structures. Many object-based segmentation methods minimize a non-convex function and risk failure due to convergence to a local minimum.

This dissertation presents three refinements to existing object-based segmentation methods. The first refinement mitigates the risk of local minima by initializing the segmentation closely to the correct answer. The initialization searches pose- and shape-spaces for the object that best matches user specified points on three designated image slices. Thus-initialized m-rep based segmentations of the bladder from CT are frequently better than segmentations reported elsewhere. The second refinement is a statistical test on object-relative intensity patterns that allows estimation of the local credibility of a segmentation. This test effectively identifies regions with local segmentation errors in m-rep based segmentations of the bladder and prostate from CT. The third refinement is a method for shape interpolation that is based on changes in the position and orientation of samples and that tends to be more shape-preserving than a competing linear method. This interpolation can be used with dynamic structures and to understand changes between segmentations of an object in atlas and target images.

Together, these refinements aid in the segmentation of a dense collection of targets via a hybrid of object-based and atlas-based methods. The first refinement increases the probability of successful object-based segmentations of the subset of targets for which such methods are appropriate, the second increases the user's confidence that those object-based segmentations are correct, and the third is used to transfer the object-based segmentations to an atlas-based method that will be used to segment the remainder of the targets.

ACKNOWLEDGMENTS

I thank my advisor Dr. Stephen M. Pizer for his guidance and mentorship and for welcoming me into the Medical Image Display and Analysis Group (MIDAG). Steve has gone to great lengths to ensure that MIDAG fosters effective interdisciplinary collaborations. Even in its earliest stages, my research was driven by the needs of Dr. Edward L. Chaney, a medical physicist from the Department of Radiation Oncology. I am thrilled that my next endeavor is a collaboration with Steve and Ed to translate technologies developed at MIDAG, including pieces of this dissertation, into clinically usable products.

The interdisciplinary nature of MIDAG is how the remainder of my dissertation committee came to consist of Dr. Mark Foskey, a mathematician turned computer scientist whose appointment is in the Department of Radiation Oncology; Dr. J. Stephen Marron from the Department of Statistics and Operations Research; and Dr. Martin A. Styner, a computer scientist who holds joint appointments in the departments of Computer Science and Psychiatry. Their diverse backgrounds are reflected in the breadth of my dissertation.

I thank all of the MIDAG members who have helped me along the way. I would like to explicitly thank three of my former office mates: Dr. Eli Broadhurst, Rohit Saboo, and Dr. Joshua Stough and three colleagues whose offices I frequently invaded: Gregg Tracton, Ja-Yeon Jeong, and Dr. Graham Gash for helping when I needed help and for pushing when I needed a push.

Finally, I thank my family for their love and support, especially Erin and Hazel who have tolerated my spending nights and weekends working on the laptop for far too long.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ART	adaptive radiotherapy
ASM	active shape model
CDF	cumulative density function
CT	computed tomography
DVH	dose-volume histogram
DRIQF	discrete regional intensity quantile functions
HDLSS	high dimension, low sample size
IID	independent and identically distributed
MAP	maximum *a postori* estimation
MLE	maximum likelihood estimation
MR	magnetic resonance
PCA	principal component analysis
PDM	point distribution model
PGA	principal geodesic analysis
QF	quantile function
RIQF	regional intensity quantile functions
ROC	receiver operating characteristic
RT	radiotherapy
SDM	signed distance map
SDSM	statistical deformable shape model
SPHARM	spherical harmonic

Chapter 1

Introduction

1.1 Motivation: Object-Based Medical Image Segmentation

The use of medical images to guide diagnostics, therapies, and therapy-evaluations is improving the clinical outcome and quality of life for patients with a variety of conditions. In these situations the clinician must understand the images in order to treat the patient.

Fig. 1.1 shows examples of medical images. The left pane is an X-ray image showing a severely fractured femur [1]. This image is fairly easy to understand. The bone is the bright object in the foreground and the fracture is the large-scale discontinuity in it. The right pane is the mid-sagittal slice of a computed tomography (CT) image of the male pelvis. Locating the prostate and bladder (green and cyan outlines, respectively) in this image is a challenging task.

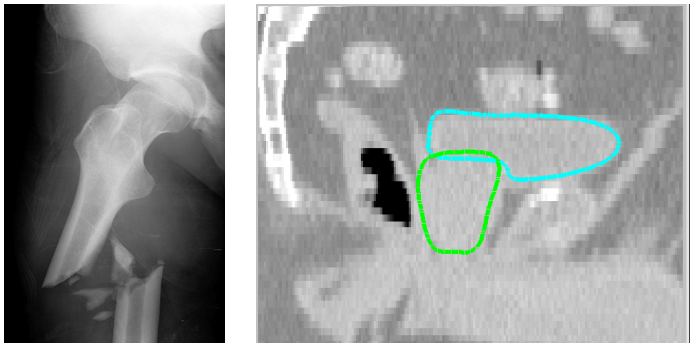


Figure 1.1: Examples of medical images

In applications that require an understanding of a challenging image it is a common practice

to *segment* the image. Here *segmentation* refers to the process of assigning labels to the voxels of the image that belong to an object or structure of interest. Once the image has been segmented the end users can easily understand it in the context of the specific medical task at hand via a casual glance at its label-map. However segmentation itself can be a difficult and time consuming task.

Segmentation methods are characterized in the following three ways based on the level of control the user has over the segmentation. *Manual* segmentation methods give the user the flexibility to control the labeling of each voxel, although they may feature tools that allow the user to assign a label to a region larger than a single voxel. *Semi-automatic* segmentations rely on limited user interaction in conjunction with automated image processing in order to label the image voxels. *Automatic* segmentation methods assign labels to voxels entirely on the basis of computerized image processing without any human interaction at all.

Automatic and semi-automatic segmentation methods are desirable for several reasons. The time and attention of the clinician is a scarce resource, whereas computational cycles are a commodity. If the medical expert can be relieved of the burden of producing segmentations, she can reallocate that time to better serve her patients. Furthermore, computers can produce reproducible segmentations, but human experts are subject to intra- and inter- human variabilities.

A powerful subclass of (semi-) automatic methods are those that are *object-based*. These methods are programmed to understand the geometry of an object (or structure) of interest. The set of voxels assigned the label for that object will have geometry that is consistent with the object itself. An object-based segmentation method that follows Bayes' rule for posterior optimization is a principled way to mimic manual segmentations. If the target object is the prostate, Bayes' rule instructs the computer to find a prostate-shaped object with prostate-like intensity patterns in and around it in the image. The "prostate" label is assigned to the voxels in the interior of that object.

Bayesian segmentation relies on estimated probability distributions on the objects themselves and on the image intensity patterns relative to the objects. The former is known as a *prior distribution* or a *shape prior*. The latter is known as a *data likelihood*. These prob-

ability distributions are *generalizable* if a novel object, or image intensity pattern relative to the object, from the target population is admitted with a high probability. These probability distributions are *specific* if all shape instances (resp. image patterns) that are admitted with high probability are representative of the target population. High generalizability and specificity are both desirable properties for the probability distributions used during Bayesian segmentation.

Returning to the example of prostate segmentation, a shape prior with high specificity gives the end user confidence that the final result will be "prostate-shaped". A highly generalizable shape prior gives the user further confidence that a suitably large space of "prostate-shaped" objects will have been examined. A data likelihood with high generalizability allows the user to believe that the target object will be recognized as having an appropriate intensity pattern. When the data likelihood is also highly specific, it becomes less likely that the segmentation will converge to a false optimum elsewhere in the image.

Bayesian segmentation methods should not be used when the shape prior and data likelihood distributions cannot be learned. For example, when working with CT images of the head and neck, regions of the lymphatic system cannot be seen directly. They are understood implicitly based on known relationships to nearby structures such as muscles, glands, blood vessels, and bones. In this application a segmentation method that transfers labels from a manually segmented *atlas image* to the target image is more appropriate. This transfer function is often defined by producing a warp of the ambient space that minimizes an energy function on the warp and the difference between the atlas image and the warped target image. I use the term *atlas-based segmentation* to refer to this method of producing a warp in order to transfer labels from the voxels of the atlas.

If an atlas-based segmentation method is forced to conform to object-based segmentation for the subset of targets that have already been well segmented, the segmentation of the entire complex of targets will be more relevant. This is true because errors involving that subset of pre-segmented targets will be bounded by the object-based segmentation and because the locations of these objects constrain the possible locations of the others.

This application of object-based segmentation is based on the assumptions that the segmen-

tations are successful and that the atlas-based segmentations will adhere to those object-based segmentations. In this dissertation I address those assumptions with an initialization strategy that produces high-quality object-based segmentations, with a statistical test that allows the clinician to be confident in the quality of the object-based segmentations she uses, and with a geometric tool for warping between object instances that can be used to drive a diffeomorphic warp between an atlas and a target image.

1.2 Semi-automatic Segmentation Initialization via a Sparse Set of Contours

A known limitation of object-based segmentation methods is that they frequently are equivalent to the optimization of a non-convex objective function and are susceptible to local optima problems. It is generally assumed that the objective function has its global optimum at the desired answer and that in some local neighborhood around that answer there are no other local optima. When such an object-based segmentation is initialized with a shape instance that is outside of this neighborhood, there is a risk that the method will hallucinate a segmentation in some other region of the image that happens to be locally optimal. An initialization strategy that begins the optimization within the capture range of the global optimum mitigates this risk.

One would expect that when the segmentation routine is seeded with an initial shape that is in the correct image region the final segmentation result will also be in that region. Although this assumption typically is correct, it does not produce a practical solution as it requires a segmentation of the target image in order to produce the initial shape instance that is used to segment the target image. However, in the setting of Bayesian segmentation, a shape prior with high generalizability can be used to produce a shape instance that is close to the desired answer without requiring a full segmentation of the target image.

In Chapter 3, I present an algorithm for semi-automatically producing a shape instance that can be used to initialize segmentation. The user is asked to identify points on the boundary of the target object, and the shape prior is used to interpolate a shape near those manually

selected points.

The goodess of fit of this approximated shape typically increases with the number of manually specified points. But, because the user's time and attention are scarce resources, it is important to limit the number of points the user is asked to identified. However, when too few points are specified, this approximated shape is likely to be inadequate and the final segmentation is likely to not be clinically usable. Consequently, the cost of manually correcting erroneous segmentations must also be taken into account when specifying the points the user is to identify at initialization time.

I propose a strategy for approximating a bladder segmentation from a subset of the boundary points found on only three axial slices of a CT image. A shape prior trained on bladder segmentations is used to find a bladder shaped object with boundary points near the ones identified on the three image slices. This shape model is suitable to initialize an object-based segmentation of the bladder that typically needs little or no manual correction to be clinically usable. Initialization strategies for the prostate and rectum are also discussed in this chapter, as is an alternative semi-automatic segmentation procedure that refines this approximated object by first having the user adjust the boundary and then generating a new candidate object from the shape prior that is matched to the modified boundary points.

1.3 Estimating the Local Credibility of a Segmentation

The medical image analysis community has improved automatic and semi-automatic segmentation methods to the point that for certain applications and for certain images, an appropriate method can produce a medically relevant segmentation for a great fraction of target images. The contributions of Chapter 3 help to make this true for segmentations of the organs in the male pelvis from CT images. The community's continued progress will likely produce segmentation methods for an increasingly large variety of medical applications that produce acceptable results for a great fraction of input images. This success brings with it a serious risk: the danger that users will become too accepting of computer-generated segmentations and in the event of a failure an inaccurate segmentation will corrupt a medical procedure. This

corruption can occur even if the failure is localized to a small region in the image.

Perhaps the safest behavior would be to have a trained expert, or better yet a team of trained experts, carefully review the validity of every segmentation produced. Unfortunately the scarcity of these experts and the fact that their time and attention can not possibly scale with number of images in need of analysis makes this option unfeasible.

Just as computerized segmentation methods are expanding the number of medical images that users can segment, computerized validation methods are needed to expand the number of segmentations that users can validate. The relationship between correct segmentations and the associated image intensity patterns captured by a Bayesian data likelihood distribution can be an important tool validating the segmentation of a new target image. If the image intensity pattern in and around the segmentation of the target image is inconsistent with what is known from the data likelihood distribution, further validation of the segmentation is warranted.

A mathematical function that can be used to evaluate the data likelihood for a new segmentation is often known as a *geometry-to-image match* function or simply as an *image match* function. When an image-match function can be decomposed into independent terms for different regions of the segmentation or its boundary, each such term is known as a *local* (geometry-to-) image match function.

In Chapter 4, I present a method that identifies statistical outliers of local geometry to image match functions as *non-credible* regions where the user should concentrate her validation efforts. This method has been applied to segmentations of the bladder and prostate from CT, and the identified non-credible regions were correlated with large disparities between the segmentations being tested and reference manual segmentations of the same images.

Knowing that a local image match function yields an outlier value for a given segmentation allows the end user to evaluate the segmentation for that local region and to manually make a decision about what, if any, corrections are needed. A variation on this method also considers the local image match function from the opposite side of the object boundary. Suppose that a particular neighborhood on the exterior of the object has been identified as non-credible and that the image intensity in that region is consistent with what is expected for the corresponding interior region of the object. This information is a strong cue as to how the segmentation could

be corrected. In this case, pushing the boundary further outward is likely to produce the desired segmentation.

By applying the tests described in Chapter 4, the end user can increase her confidence in the quality of a given segmentation, and she can manually correct it when needed to produce a medically relevant segmentation.

1.4 Interpolation of Oriented Geometric Objects via Rotational Flows

The tools proposed in Chapters 3 and 4 allow a medical user to produce an object-based segmentation of an appropriate target structure and to become confident in the quality of the segmentation before using it.

One possible application for high quality object-based segmentations is to provide a means for segmenting other structures that are not amenable to object-based methods. Suppose the user has an atlas image in which a variety of structures have been labeled. Assume that object-based segmentations are feasible for some, but not all, of the labeled structures in the atlas. The transformation that brings an object (or set of objects) from a configuration that segments the atlas into a configuration that segments the target image, could be extrapolated to define a mapping from the atlas image volume to the target image volume. This mapping when applied to the segmentation of one of the other structures in the atlas produces a segmentation of that structure in the target image.

Another application for high quality object-based segmentations is to provide an initialization for the object-based segmentation of related images. Suppose the user has a time series of images in which the target structures of interest are deforming as part of a physiological process. For example the lungs may be inflating and deflating, or the chambers of the heart may be filling and emptying. Let us assume the user has object-based segmentation of the images at key times, for example models of the lung at peak inspiration and peak expiration. An intermediate model configuration that is somehow "between" these two reference configurations approximates the configuration of the object at an intermediate time, and may be a

suitable initialization for object-based segmentation of the intermediate image.

In Chapter 5, I present a novel method for interpolating between two objects. The interpolation produced by this method can be used to describe the transformation between an object in the atlas and an object in the target image. A sampling of boundary points on these interpolated objects can serve the landmark paths which are commonly used to define mappings between images. The interpolation can also be used to produce intermediate objects to initialize object-based segmentations for a time series of images.

This interpolation method is based upon a concept that I call *rotational flows*. The interpolation requires correspondence between samples of the initial and target objects as well as a complete frame to orient each sample. The rotational flows cause each sample to move along a circular arc, rotating about a uniquely determined *rotational center* so that the angle swept out by the arc matches rotation needed to transform the reference frame into its target configuration.

Because rotational-flows interpolation connects position with orientation, it tends to be more shape maintaining than methods that interpolate those two independently. Specifically when the atlas and target objects are related by a similarity transformation, all interpolated objects are also similar to the reference set. When the transformation from the atlas to target object requires a shape change, the method of rotational flows frequently produces a visually satisfying interpolation.

1.5 Thesis and Contributions

Thesis: *Atlas-based segmentation of a dense region consisting of multiple objects and the spaces between them can be performed effectively given high quality segmentations of the objects in the target space and a means to transfer those segmentations to the entire space. Initializing an object-based segmentation by matching the object to a sparse set of boundary contours can produce results that are appropriate for this purpose. Estimating the credibility of the segmentation by detecting statistical outliers of the image match in local regions relative to the object allows the clinician to be confident in the segmentation results. Rotational-flows*

interpolations from atlas object instances to credible segmentations produce a set of landmark paths that can be used in the large-scale diffeomorphism framework to map the full set of labels on the atlas to the target image.

The contributions of this dissertation are as follows:

1. Evidence that given a shape prior with adequately high generalizability and specificity, a dense object boundary can be recovered from a sparse sampling of boundary points.

2. Strategies for how a user can simply produce such a boundary sampling to recover a shape model for the bladder from easily identifiable points found on relatively few CT image slices.

3. Applications of this recovered shape to initialize or refine a semi-automatic segmentation.

4. A novel tool that identifies statistical outliers of a localized geometry-to-image-match function as regions where a segmentation is not credible.

5. A novel method for shape interpolation that synchronizes local changes of position and orientation. This method is driven by "rotational flows" about "rotational centers"

6. A method for computing the mean of a shape population based on pairwise interpolations via the method of rotational flows.

1.6 Overview of Chapters

This dissertation is organized as follows. Chapter 2 is a discussion of background material on the production and evaluation of medical image segmentations. In Chapter 3, I present the semi-automatic initialization scheme that has been outlined in Sec. 1.2. This method has been applied to a shape space of bladders with high generalizability to allow high quality segmentations of the bladder from CT. In Chapter 4, I expand on Sec. 1.3 to present the method for evaluating a segmentation with respect to a set of geometry-to-image match functions with local support. This evaluation provides the end user with cues of when and where a (semi-) automatic segmentation may not be credible, and in some cases provides cues of how to correct

the segmentation boundary. In Chapter 5, I present the method for shape interpolation that was introduced in Sec. 1.4. This method interpolates between shape models (for example, segmented objects) in a way that synchronizes changes in position with changes in orientation. The interpolations it produces corresponding points on the object along curved paths. These paths can be extrapolated to define a warping of the entire volume from one image to the other. When one of these images is an atlas, this function defines a way to transfer the full set of labels to the target image. Another application of this interpolation is in the calculation of a mean shape for a population. In Chapter 5, I also present a method for constructing a population mean that is based on interpolations between pairs of shapes. Chapter 6 concludes this dissertation, and discusses the methods presented within and opportunities for future work.

Chapter 2

Background

Image analysis algorithms can be classified as *low-level* methods that simulate how the human visual system responds to signals in the images, such as detectors for edges [10] and other image features [28, 53], and *high-level* methods that provide greater abstraction to perform more complicated tasks, such as the segmentation, tracking, and recognition of objects. High-level methods are often built by combining low-level feature detectors with other constraints based on the physics or geometry [42, 36] of a target object, or the statistics of a population of objects [19, 26].

Medical image segmentation is an important tool for helping clinicians and researchers to use images to understand the anatomy of a patient or subject. In this chapter I survey high-level methods for segmenting medical images and metrics for evaluating segmentation quality.

2.1 Medical Image Segmentation

Throughout this dissertation I will use following notation. Let Ω denote an imaging target. To simplify notation I will assume that

$$\Omega \subset \mathbb{R}^3, \tag{2.1}$$

i.e., the image describes a three-dimensional volume. It should be noted that medical images exist in spaces with other dimensionalities. For example, sonography is often used to produce

one- and two-dimensional images, and a time series of three-dimensional images can be thought of as a single four-dimensional image.

An image, denoted by $\underline{I}$, is a map

$$\underline{I} : \Omega \quad \rightarrow \quad V \tag{2.2}$$

where V is the image value. For many medical imaging modalities the image has scalar values:

$$V \quad \subset \mathbb{R}. \tag{2.3}$$

As an example, the value of each voxel in a CT image is the average linear attenuation coefficient for the material in that volume and describes the density of that material. Images that assign a tuple of values to each voxel are used in medicine as well. For example, diffusion tensor magnetic resonance (MR) imaging takes a scalar measurement from each of a set of magnetic gradient directions to produce a multiple values for each voxel. However, I will make the simplifying assumption that $\underline{I}$ is scalar-valued.

Let L denote a label and let S_L denote the segmentation of L for $\underline{I}$.

$$S_L \quad \subset \quad \Omega \tag{2.4}$$

is the set of voxels for which L is an appropriate label. The label L may identify an organ, such as the kidney, or a tissue class, such as white matter in the brain. In those examples, the segmentation S_L would be the subset of voxels in the image that correspond to the kidney or to white matter. In many segmentation applications, only a single label L is of interest. To simplify notation in such cases, the index L will be omitted, and S will be used to denote the segmentation.

A manual segmentation process requires a human with domain knowledge to identify the voxels in S_L. This human is often known as the *rater*. In a semi-automatic segmentation process the computer identifies S_L with limited human interaction, and in a (fully-) automatic segmentation process the computer identifies S_L with no human input.

Segmentation methods that are (semi-) automatic have several potential advantages over manual processes. Manual segmentation of medical images can be a time consuming process, but (semi-) automatic segmentation can be much faster. In a clinical setting this speed-up could allow medical professionals to reallocate time spent manually segmenting images to other aspects of patient care. It could also reduce the time between image acquisition and the treatment of a patient. In some applications, the snapshot of the patient's anatomy in the image is only relevant for a limited time. In these cases, a fast computer-generated segmentation may mean the difference between having a clinically usable image or not.

The speed advantage of (semi-) automatic segmentation is important in a research setting as well. It can allow researchers to study a larger data set in a fixed time frame than would be possible with manual segmentations. It can allow researchers to produce their results faster. Computerized segmentation methods are typically able to scale with the number of images to be studied, so it can be relatively easy and inexpensive for researchers in a retrospective study to reduce their image processing time by purchasing more computers. In contrast, if the same researchers relied on manual segmentations, hiring additional raters is likely to be difficult and expensive. Even if the study can afford and has access to additional raters, inter-rater variability can lead the researchers to false conclusions.

Computer-generated segmentations can offer greater reproducibility than a manual process. Given a fixed input, a deterministic method will always produce the same segmentation. However multiple manual segmentations of an image will be subject to inter- and intra-rater variabilities.

Inter-rater variability is known to be quite large [44, 52, 79]. Different institutions may have different protocols for segmenting the same objects, leading to gross differences when two raters segment the same image. Even when the raters follow the same protocol, they may exhibit systematic biases based on their training.

Intra-rater variability is also known to be a significant issue [79, 76] due to the following factors. It can be difficult for the rater to see the boundary between S_L and $\Omega \ominus S_L$ in regions of some images. The rater has trouble reproducing the segmentation in those regions. Understanding S_L when Ω is a space with three or more dimensions is a difficult task for

the rater. In three-dimensions, the rater typically produces the segmentation by examining two-dimensional slices of the image. It is difficult for the rater to produce a segmentation that is consistent across slices. The quality of the segmentation is also dependent on amount of time the rater can devote to it and will be negatively affected if the rater is distracted or interrupted.

Because of the reproducibility and speed advantages that computerized segmentation methods hold over manual segmentations, a successful automatic segmentation method will be of great value to clinicians and researchers alike. The most general methods for segmenting images rely only on the image intensities and assign the same label to a set of nearby-voxels with similar values [91], with different labels on opposite sides of a boundary in the image. However, there are many applications where these methods fail due to a lack of strong intensity differences between nearby structures of interest. For example, assigning separate labels to the bladder and prostate in a (non-contrast enhanced) CT image of the male pelvis is difficult because there is very little change in appearance between the two objects.

In these challenging cases a segmentation method that is tailored to the specific target labels has the potential for greater success. The geometric relationships between the targets and between each target and other nearby structures are valuable information that can be used at segmentation time. Using the male pelvis as an example, a segmentation method that incorporates *a priori* models of what it means to be bladder-shaped or prostate-shaped, of the relative positioning of the two organs, and of their positions relative to the pubic bones and the rectum is likely to produce better segmentations of those organs than a segmentation method that relies on image intensity alone.

Object-based segmentation methods are a class of segmentation methods that adjust an *a priori* geometric model, denoted by $\underline{\underline{m}}_L$, to match the target L in the image data. The segmentation S_L is the set of voxels on the interior of $\underline{\underline{m}}_L$. I discuss the theory of object-based segmentation in Section 2.2. I provide a background on statistics relevant to this theory in Section 2.3 and a background on medial geometry in Section 2.4. I then survey several popular object-based segmentation algorithms, discussing competing models for how objects are represented and how the model is adapted to a novel image, in Section 2.5.

2.2 Object-Based Segmentation

Many object-based segmentation methods have been proposed in the literature, including the methods described in [12, 18, 42, 43, 65, 82]. All of these methods segment a novel image, which I will denote by $\underline{I}$, by adjusting $\underline{\underline{m}}_L$ so that it bests matches the image data subject to method-specific constraints. Thus object-based segmentation is an optimization problem.

There are two common formulations for this optimization problem. It may be posed as an *energy-minimization* problem,

$$\begin{aligned} \underline{\underline{m}}_L &= \arg\min_{\underline{\underline{m}}} C\left(\underline{\underline{m}}\right) \\ C\left(\cdot\right) &\geq 0. \end{aligned} \tag{2.5}$$

The function $C\left(\underline{\underline{m}}\right)$ is the cost, or energy level, associated with that configuration of the model. Typically this cost function includes a term based on how well the model is supported by the image data and a term based on the geometry of the model in this configuration. Specific examples of $C\left(\cdot\right)$ are given in Section 2.5.1 and Section 2.5.4.

Segmentation can also be posed as a statistical estimation problem.

$$\underline{\underline{m}}_L = \arg\max_{\underline{\underline{m}}} p\left(\underline{I}|\underline{\underline{m}}\right) \tag{2.6}$$

or

$$\underline{\underline{m}}_L = \arg\max_{\underline{\underline{m}}} p\left(\underline{\underline{m}}|\underline{I}\right) \tag{2.7}$$

The optimization (2.6) is known as maximum likelihood estimation (MLE), as the likelihood of the model given the image is being maximized. The optimization (2.7) is known as maximum *a postori* estimation (MAP), as it is the posterior probability of the model conditioned on the image that is being maximized.

The likelihood function, $p\left(\underline{I}|\underline{\underline{m}}\right)$, is typically learned from a set of training of exemplars. This function is minimized directly to produce an MLE segmentation. $p\left(\underline{I}|\underline{\underline{m}}\right)$ also plays an important role for MAP segmentation, because the posterior probability distribution cannot be observed from the training data. Rather, posterior probabilities can be recovered by application

of Bayes' Rule.

$$p\left(\underline{\underline{m}}|\underline{I}\right) = \frac{p\left(\underline{I}|\underline{\underline{m}}\right) \cdot p\left(\underline{\underline{m}}\right)}{p\left(\underline{I}\right)}, \tag{2.8}$$

where $p\left(\underline{\underline{m}}\right)$ is known as the prior distribution on the model. This probability distribution can be learned from training data. Specific examples of $p\left(\underline{I}|\underline{\underline{m}}\right)$ and $p\left(\underline{\underline{m}}\right)$ are given in Section 2.5.2 and Section 2.5.3.

Substituting (2.8) into (2.7), we see that for MAP estimation

$$\begin{aligned} \underline{\underline{m}}_L &= \arg\max_{\underline{\underline{m}}} \frac{p\left(\underline{I}|\underline{\underline{m}}\right) \cdot p\left(\underline{\underline{m}}\right)}{p\left(\underline{I}\right)} \\ &= \arg\max_{\underline{\underline{m}}} p\left(\underline{I}|\underline{\underline{m}}\right) \cdot p\left(\underline{\underline{m}}\right) \end{aligned} \tag{2.9}$$

because $p\left(\underline{I}\right)$ can be assumed to be constant for the image that is being segmented, even as $\underline{\underline{m}}$ varies.

There is a relationship between the energy-minimizing and probability-maximizing forms of segmentation. A cost function can be transformed into a probability by assuming an exponential distribution on the cost and then setting

$$p\left(\cdot\right) \propto \exp\left(-C\left(\cdot\right)\right) \tag{2.10}$$

Because we are interested in the maximizer of the probability distribution rather than the distribution itself, we need not solve for the exact scalar factor that relates $p\left(\cdot\right)$ to $\exp\left(-C\left(\cdot\right)\right)$.

It is possible to convert a probability distribution into a cost function simply by negating it. But it is often more convenient to use a cost function that is equal to the negated log of

the probability density function.

$$C\left(\underline{\underline{m}}\right) = -\log\left(p\left(\underline{\underline{I}}|\underline{\underline{m}}\right)\right) \tag{2.11}$$

or

$$C\left(\underline{\underline{m}}\right) = -\log\left(p\left(\underline{\underline{I}}|\underline{\underline{m}}\right)\right) - \log\left(p\left(\underline{\underline{m}}\right)\right) \tag{2.12}$$

where (2.11) applies to the MLE framework and (2.12) applies to the MAP framework. This conversion is especially useful when the probability distributions in question are Gaussian.

The statistics of multivariate Gaussian distributions are reviewed in 2.3. When data can be assumed to follow a multivariate Gaussain probability distribution, the parameters of that distribution can be learned via principal component analysis (PCA). PCA is discussed in greater detail in Section 2.3.1 and is an important statistical tool used by many of the object-based segmentation methods surveyed in Section 2.5. The -log probability terms of (2.11) and (2.12) can be understood as measures of distance from the mean of a population and can be expressed in a simple form for Gaussian data in general and with respect to PCA in particular.

2.3 Statistics of Multivariate Gaussians

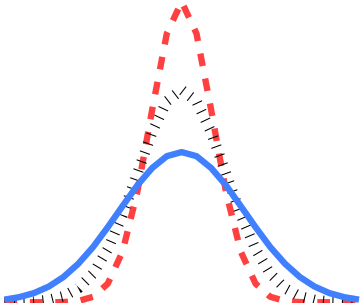
Suppose that $\underline{x}$ is an n-vector drawn from the multivariate Gaussian distribution. By definition, for any vector of coefficients $\underline{a}$, the linear combination $\underline{a} \cdot \underline{x}$ is normally distributed. The notation

$$\underline{x} \sim N\left(\mu, \Sigma\right) \tag{2.13}$$

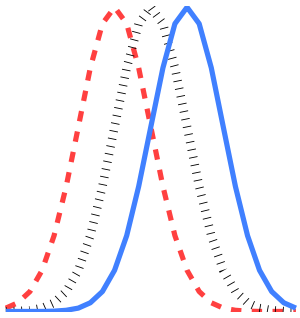
indicates $\underline{x}$ is drawn from a Gaussian distribution that has a mean $\underline{\mu}$ and non-singular covariance matrix Σ. The probability density function for this distribution is

$$p\left(\underline{x}\right) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}\left(\underline{x}-\underline{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\underline{x}-\underline{\mu}\right)\right) \tag{2.14}$$

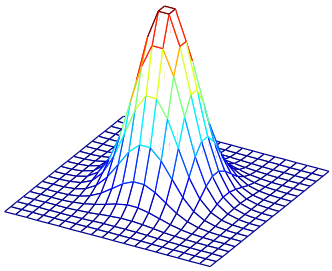
Examples of (2.14) are plotted in Figure 2.1.



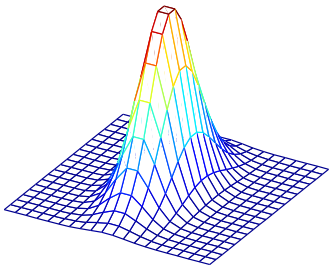
(a) In 1 dimension, with varying Σ



(b) In 1 dimension, with varying μ



(c) In 2 dimensions, with isotropic Σ



(d) In 2 dimensions, with anisotropic Σ

Figure 2.1: Gaussian probability density functions.

Suppose $\underline{v}$ is a constant n-vector and A is a constant $m \times n$ matrix. The effects of linear transformations of $\underline{x}$ are as follows.

$$\underline{x} + \underline{v} \sim N\left(\underline{\mu} + \underline{v}, \Sigma\right) \tag{2.15}$$

$$A\underline{x} \sim N\left(A\underline{\mu}, A\Sigma A^{\mathrm{T}}\right) \tag{2.16}$$

The function $-\log\left(p\left(\underline{x}\right)\right)$ is readily derived from (2.14) and is closely related to the

Mahalanobis distance, which is denoted by $d_{\text{mahal}}\left(\underline{x}\right)$ and is defined in (2.18).

$$
\begin{aligned}
-\log\left(p\left(\underline{x}\right)\right) &= \log\left(\left(2\pi\right)^{\frac{n}{2}}\sqrt{\det\Sigma}\right) + \frac{1}{2}\left(\underline{x}-\underline{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\underline{x}-\underline{\mu}\right) && (2.17)\\
d_{\text{mahal}}\left(\underline{x}\right) &= \sqrt{\left(\underline{x}-\underline{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\underline{x}-\underline{\mu}\right)} && (2.18)\\
\left(d_{\text{mahal}}\left(\underline{x}\right)\right)^2 &= \left(\underline{x}-\underline{\mu}\right)^{\mathrm{T}}\Sigma^{-1}\left(\underline{x}-\underline{\mu}\right) && (2.19)\\
&= 2\left(-\log\left(p\left(\underline{x}\right)\right)\right) - c && (2.20)
\end{aligned}
$$

c in (2.20) refers to the term in (2.17) that does not involve $\underline{x}$.

The Mahalanobis distance is a statistical tool for measuring the distance from a sample to a population mean. It accounts for the anisotropy and orientation of the population's distribution. The Mahalanobis distance is easy to calculate and has a straightforward interpretation in applications of principal component analysis (PCA).

2.3.1 Principal Component Analysis

PCA can be used as a tool for learning a probability distribution from a set of training exemplars. When the data are assumed to be drawn from a multivariate Gaussian distribution, PCA projects the data onto a set of orthogonal bases such that the projection coefficients are standard (zero mean, unit variance) independent univariate Gaussian random variables.

This formulation is useful in a generative setting, as independent samples from a standard univariate Gaussian distribution can be combined to create new data instances. It is also useful in a discriminative setting where the Mahalanobis distance is a natural measure on the data.

Suppose

$$
\underline{\underline{x}} = \left(\begin{array}{ccc} \underline{x}_1 & \ldots & \underline{x}_k \end{array}\right) \qquad (2.21)
$$

is a set of k samples and that each sample $\underline{x}_i$ is written as an n element column vector. The full set of data is written as a $n \times k$ matrix.

The first step in PCA is to center the data and compute its $n \times n$ covariance matrix.

$$\begin{aligned} \text{Let } \underline{\mu} &= \underset{i=1..k}{\text{mean}} \left(\underline{x}_i \right) && (2.22) \\ \underline{x}'_i &\leftarrow \underline{x}_i - \underline{\mu} && (2.23) \\ \Sigma &= \frac{1}{k-1} \left(\underline{\underline{x}}' \right) \left(\underline{\underline{x}}' \right)^{\mathrm{T}} && (2.24) \end{aligned}$$

The covariance matrix is known to be symmetric:

$$\Sigma_{i,j} = \Sigma_{j,i} \; \forall i, j, \tag{2.25}$$

and positive semi-definite:

$$\underline{a}^{\mathrm{T}} \Sigma \underline{a} \geq 0 \; \forall \underline{a}. \tag{2.26}$$

From these two properties it follows that singular value decomposition on Σ will yield its eigenvectors and eigenvalues.

$$\begin{aligned} \Sigma &= V \Lambda V^{\mathrm{T}} \text{ , where} && (2.27) \\ \Lambda_{i,i} &= \lambda_i && (2.28) \\ \Lambda_{i,j \neq i} &= 0 && (2.29) \\ \Sigma V_{\cdot,j} &= \lambda_j V_{\cdot,j} && (2.30) \\ V^{\mathrm{T}} V &= \mathrm{Id}_n && (2.31) \end{aligned}$$

The matrix Λ in (2.27) is the diagonal matrix of the eigenvalues of Σ. The matrix Id_n in (2.31) is the $n \times n$ identity matrix. As a convention, the eigenvalues of Σ are sorted in descending order.

$$i < j \rightarrow \lambda_i \geq \lambda_j \tag{2.32}$$

A sample $\underline{x} \sim N\left(\underline{\mu}, \Sigma\right)$ can be rewritten as follows.

$$\underline{x} = \underline{\mu} + \left(\underline{x} - \underline{\mu}\right) \tag{2.33}$$

$$\underline{x} = \underline{\mu} + \sum_{i=1}^{n} \left(\left(\underline{x} - \underline{\mu}\right) \cdot V_{\cdot,i}\right) V_{\cdot,i} \tag{2.34}$$

$$\underline{x} = \underline{\mu} + \sum_{i=1}^{n} \frac{\left(\left(\underline{x} - \underline{\mu}\right) \cdot V_{\cdot,i}\right)}{\sqrt{\lambda_i}} \sqrt{\lambda_i} V_{\cdot,i} \tag{2.35}$$

Each coefficient of the form $\frac{\left(\left(\underline{x} - \underline{\mu}\right) \cdot V_{\cdot,i}\right)}{\sqrt{\lambda_i}}$ in (2.35) can be shown to be an independent and identically distributed (IID) standard normal random variable by (2.15) and (2.16).

$$\begin{aligned}
\left(\underline{x} - \underline{\mu}\right) &\sim N\left(\mu - \mu, \Sigma\right) && (2.36)\\
&\sim N\left(0_n, \Sigma\right) && (2.37)\\
\frac{V_{\cdot,i}}{\sqrt{\lambda_i}}^{\mathrm{T}} \left(\underline{x} - \underline{\mu}\right) &\sim N\left(0, \frac{V_{\cdot,i}}{\sqrt{\lambda_i}}^{\mathrm{T}} \Sigma \frac{V_{\cdot,i}}{\sqrt{\lambda_i}}\right) && (2.38)\\
&\sim N\left(0, \frac{1}{\lambda_i} V_{\cdot,i}^{\mathrm{T}} \Sigma V_{\cdot,i}\right) && (2.39)\\
&\sim N\left(0, \frac{1}{\lambda_i} V_{\cdot,i}^{\mathrm{T}} V \Lambda V^{\mathrm{T}} V_{\cdot,i}\right) && (2.40)
\end{aligned}$$

$$\begin{aligned}
&\sim N\left(0, \frac{1}{\lambda_i} \lambda_i\right) && (2.41)\\
&\sim N\left(0, 1\right) && (2.42)
\end{aligned}$$

This leads to the following expression for $\left(d_{\text{mahal}}\left(\underline{x}\right)\right)^2$:

$$\left(d_{\text{mahal}}\left(\underline{x}\right)\right)^2 = \sum_{i=1}^{n} \frac{\left(\left(\underline{x} - \underline{\mu}\right) \cdot V_{\cdot,i}\right)^2}{\lambda_i} \,. \tag{2.43}$$

Equation (2.35) also leads to an algorithm for generating new data instances according to the learned distribution. Let $\underline{\alpha}$ be an n tuple of IID standard normals. A new data instance $\underline{a}$ can

be generated by

$$\underline{a} \quad = \quad \underline{\mu} + \sum_{i=1}^{n} \alpha_i \sqrt{\lambda_i} V_{\cdot,i}. \tag{2.44}$$

Its squared Mahalanobis distance is merely

$$\left(d_{\text{mahal}}\left(\underline{a}\right)\right)^2 \quad = \quad \sum_{i=1}^{n} {\alpha_i}^2 \ . \tag{2.45}$$

PCA For Dimensionality Reduction and Robust Estimation

Equations (2.35) and (2.44) assume that all n eigenvalues of Σ are non-zero and are retained by the PCA. In practice this is typically an unworkable approximation.

Many applications in medical image analysis suffer from the high dimension, low sample size (HDLSS) problem. Because the cost of training exemplars is frequently high and because data representations for shape models and images can require hundreds, millions, or more features, the number of samples $k \ll n$. In this case $\text{rank}\,(\Sigma) \leq (k-1)$. Thus, Σ has at most $(k-1)$ non-zero eigenvalues. Even when HDLSS is not a problem, correlations in the data may mean that it actually lies in an $m < n$ dimensional subspace.

When only the first m eigenvalues of Σ are non-zero, rewriting (2.35) as

$$\underline{x} = \quad \underline{\mu} + \textstyle\sum_{i=1}^{m} \frac{\left(\left(\underline{x}-\underline{\mu}\right)\cdot V_{\cdot,i}\right)}{\sqrt{\lambda_i}} \sqrt{\lambda_i} V_{\cdot,i} \tag{2.46}$$

reduces the dimensionality of the data from n to m, as each sample can represented by its m coefficients of the non-zero eigenvectors.

Projection of a sample onto the first m eigenvectors:

$$\underline{x} \approx \quad \text{Proj}_m\left(\underline{x}\right) = \underline{\mu} + \textstyle\sum_{i=1}^{m} \frac{\left(\left(\underline{x}-\underline{\mu}\right)\cdot V_{\cdot,i}\right)}{\sqrt{\lambda_i}} \sqrt{\lambda_i} V_{\cdot,i} \tag{2.47}$$

is a reasonable approximation of the data when $0 < \lambda_{m+1} < \epsilon$ for some small ϵ. Projections of the centered training data onto the eigenvectors of Σ have the following properties that justify the use of PCA for dimensionality reduction.

$$V_{\cdot,1} = \arg\max_{\underline{v}} \text{var}\left(\underline{x}' \cdot \underline{v}\right) \tag{2.48}$$

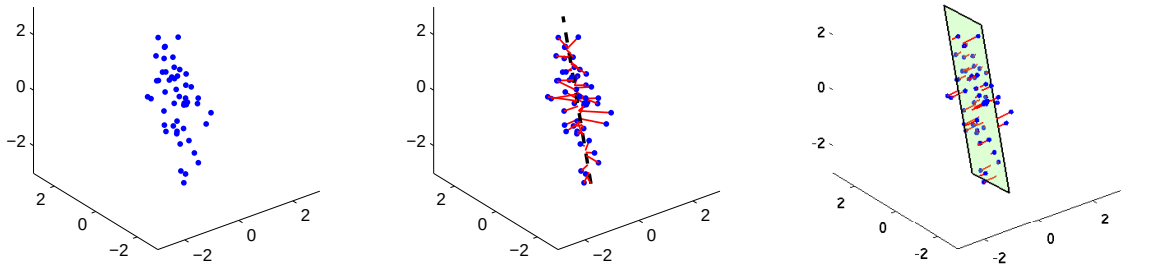
$$V_{\cdot,1} = \arg\min_{\underline{v}} \text{var}\left(\underline{x}' - \left(\underline{x}' \cdot \underline{v}\right)\underline{v}\right) \tag{2.49}$$

$$V_{\cdot,m} = \arg\max_{\underline{v}} \text{var}\left(\left(\underline{x}' - \text{Proj}_{m-1}\left(\underline{x}'\right)\right) \cdot \underline{v}\right) \tag{2.50}$$

$$V_{\cdot,m} = \arg\min_{\underline{v}} \text{var}\left(\underline{x}' - \text{Proj}_{m-1}\left(\underline{x}'\right) - \left(\underline{x}' \cdot \underline{v}\right)\underline{v}\right) \tag{2.51}$$

$V_{\cdot,1}$ is the direction that maximizes the variance of the projections of the centered data and minimizes the variance of the residuals. Likewise, the eigenvectors corresponding to the m largest eigenvalues span the m-dimensional space that maximizes the variance of the projections of the centered data and minimizes the variance of the residuals. Thus, $\underline{\mu}$ plus the span of $V_{\cdot,1}...V_{\cdot,m}$ is the best m dimensional linear space for approximating the original training data.

The relationship between the data $\underline{\underline{x}}$ and the eigenvectors of Σ is illustrated in Figure 2.2.



(a) Points in $\mathbb{R}^3$

(b) Projection onto the first principal component

(c) Projection onto the first two principal components

Figure 2.2: Principal component analysis of a toy data set. Figure 2.2(a) shows a scatter plot of points in 3 dimensions. Figure 2.2(b) adds a heavy line indicating the first principal component: the line formed by $\mu + \alpha V_1$. The projections of the points onto this line are shown with a thin red line. Figure 2.2(c) shows the points projected onto the first two principal components: the plane formed by $\mu + \alpha V_1 + \beta V_2$.

Not only does PCA maximize the variance of $\text{Proj}_m\left(\underline{x}\right)$ and minimize the variance of

$\underline{x} - \text{Proj}_m\left(\underline{x}\right)$, but these variances are known.

$$\text{var}\left(\text{Proj}_m\left(\underline{x}\right)\right) = \sum_{i=1}^{m}\left(\lambda_i^2\right) \tag{2.52}$$

$$\text{var}\left(\underline{x} - \text{Proj}_m\left(\underline{x}\right)\right) = \sum_{i=m+1}^{n}\left(\lambda_i^2\right) \tag{2.53}$$

There are two commonly used strategies for choosing the number of eigenmodes to retain when using PCA for dimensionality reduction. One may choose an explicit threshold, ϵ, that all retained eigenvalues must be above.

$$m = \arg\max_{m*}\left(\lambda_{m*} \geq \epsilon\right) \tag{2.54}$$

Or one may choose to retain a variable number eigenmodes so long as a fixed percentage γ of the total variance is retained.

$$m = \arg\min_{m*}\left(\frac{\sum_{i=1}^{m*}\lambda_i}{\sum_{i=1}^{n}\lambda_i} \geq \gamma\right) \tag{2.55}$$

There are computational and statistical benefits of using a reduced-dimensional PCA to learn the probability distributions optimized during a segmentation (2.11, 2.12). One computational benefit of dimensionality reduction is that the storage requirement for the distribution scales with the number of eigenvectors that are retained. Another computational benefit is that the Mahalanobis distance for a data object can be used as a surrogate for its log probability. The cost of computing the Mahalanobis distance scales with m. This is true whether the data object needs to be projected into the PCA space or the data is generated from the PCA by expanding a given set of coefficients (2.44). In this second case, the segmentation consists of optimizing over the set of coefficients. The size of the search space and the cost of the optimization increase with m as well.

A statistical benefit of PCA dimensionality reduction is that when the number of data samples is small, the larger eigenvalues and their corresponding eigenvectors are estimated more stably than the smaller ones [59]. In the case of medical image segmentation, the PCA is

typically performed on some artifact of a carefully drawn manual segmentation. These manual segmentations are very expensive, so it has become the norm to work with small datasets. Discarding the smallest eigenvectors from the PCA improves the overall robustness of the estimation of the probability distribution.

Because of this property of robustness against a limited sample size and because it allows computationally efficient methods for approximating the log probabilities needed to drive a segmentation, PCA is a popular tool for estimating the probability distributions used during object-based segmentation. Examples of its use will be given in Section 2.5. In some of these examples, the Mahalanobis distance is not used explicitly as a log probability calculation. Instead, the projection of a candidate object into the PCA space, and perhaps a truncation of the coefficients, will be used to constrain the optimization to lie in the learned shape space.

2.4 Medial Geometry

The segmentation method that I present in Section 2.5.3 uses medial descriptors as its shape representation. These medial descriptors provide a compact representation for an object $\underline{\underline{m}}$ and provide a coordinate system on the interior of $\underline{\underline{m}}$.

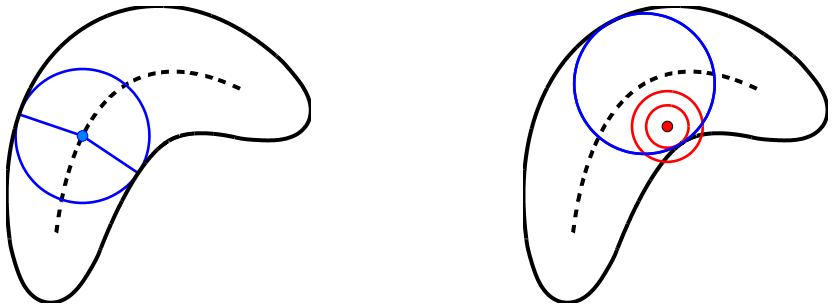
The medial axis of an object $\underline{\underline{m}}$ was defined by Blum [5] as the locus of centers of maximal inscribed balls. Throughout this Section, $\underline{p}$ will denote a point on the medial axis of $\underline{\underline{m}}$ and $\mathcal{B}\left(\underline{p}\right)$ will denote the maximal inscribed ball of $\underline{\underline{m}}$ that is centered at $\underline{p}$. By definition, an inscribed ball must lie on the interior of the object. To be maximal, it must not be contained in any other inscribed ball:

$$\mathcal{B}\left(\underline{p}\right) \quad \subset \quad \underline{\underline{m}} \tag{2.56}$$

$$\mathcal{B}\left(\underline{p}\right) \subset \mathcal{B}' \quad \rightarrow \quad \mathcal{B}' \not\subset \underline{\underline{m}}\,, \forall \text{ balls } \mathcal{B}'. \tag{2.57}$$

An example object and its medial axis are shown in Figure 2.3

Giblin and Kimia [29] classify points on the medial axis based on the contact between the maximal inscribed ball centered at that point and the boundary of the object. Suppose that the first k derivatives of the function that measures the squared Euclidean distance from the



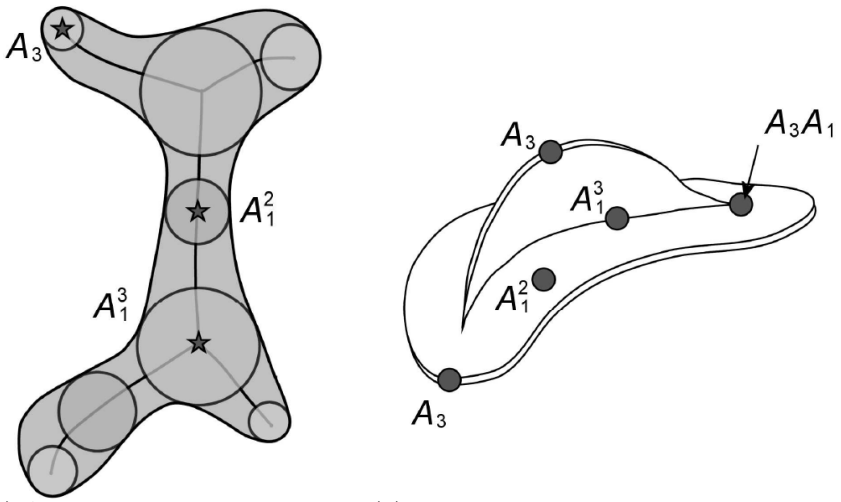
(a) The highlighted point on the medial axis is the center of a maximal inscribed ball. The ball is drawn with the two spoke vectors which point to the contact points on the object boundary.

(b) A point that is not on the medial axis is highlighted. The larger ball centered at this point is not inscribed as it crosses into the exterior or the object. The smaller ball is inscribed in the object but is not maximal as it is a subset of another inscribed ball, which is also shown.

Figure 2.3: Medial geometry. An object boundary (black line) is shown with its medial axis (dashed line). A single point on the medial axis is highlighted in Figure 2.3(a). A point that is not on the medial axis is highlighted in Figure 2.3(b).

boundary of $\underline{\underline{m}}$ to $\underline{p}$ vanish at the point of contact and that the $(k+1)^{\text{th}}$ derivative is non-zero there. This contact is said to be of order k, and $\underline{p}$ is of type A_k. When $\mathcal{B}\left(\underline{p}\right)$ makes order k contact with m points on the boundary of $\underline{\underline{m}}$, $\underline{p}$ is said to be of type A_k^m. The catalog of generically occurring points is quite small, and Giblin and Kimia have fully enumerated it.

The medial axis of a generic object in two dimensions will generically consist primarily of one or more curves of A_1^2 points. If more than one such curve is present, the intersection of three curve segments at a point of A_1^3 contact will generically occur. An A_1^3 point can be understood as the coincidence of A_1^2 points from each of the three segments coinciding. Each of these A_1^2 points shares one of its contact points with the medial point from one of the other segments and shares its other contact point with the medial point from the third segment. Each end of a curve of A_1^2 points will generically occur either at an A_1^3 junction or at an A_3 point. At the A_3 point at the end of a segment of the medial axis, the two contact points collapse onto a single point that is an extremum of curvature on $\underline{\underline{m}}$, and $\underline{p}$ is the center of curvature of $\underline{\underline{m}}$ for that point. Figure 2.4(a) shows a two dimensional object and its medial axis and illustrates the generically occurring contact types.



(a) A two dimensional object and its medial axis.

(b) The medial axis of a three dimensional object.

Figure 2.4: Medial geometry. Generic contact points on medial axes are shown. These figures have been reproduced from [93] with permission from the author.

The medial axis of a generic object in $\mathbb{R}^3$ will generically consist of sheets of A_1^2 points as well as special curves at the boundaries and intersections of the sheets and special points at the intersections of the curves. Similarly to the two dimensional cases, three sheets of A_1^2 points can join in a curve of A_1^3 points. Otherwise a sheet of A_1^2 points will generically be bounded by a curve of A_3 points. The medial locus for a three dimensional object can generically include two contact types that are not seen for two dimensional objects. An A_1^3 junction curve can intersect an A_3 end curve at an A_3A_1 point. This happens when a sheet is attached to two others along part of its boundary, the A_3A_1 point is known as a "fin creation" point as it resembles a fin emerging from the body of a shark. Four A_1^3 curves bounding six A_1^2 sheets can generically intersect at a A_1^4 point. An example of the medial locus for a generic three dimensional object and its contact types can be seen in Figure 2.4(b).

The catalog of contact-types is the basis for a strategy for understanding the object at a coarse scale. We can decompose $\underline{\underline{m}} \subset \mathbb{R}^3$ into components that correspond to distinct A_1^2 sheets on its medial axis. These sheets correspond to coarse scale structures in $\underline{\underline{m}}$. The connective A_1^3 curves allow us to understand the relationships between these structures. Similarly $\underline{\underline{m}} \subset \mathbb{R}^2$ can be decomposed into components that correspond to A_1^2 curves and understood by the

connections between them. I will use the notation M to refer to a curve or sheet of A_1^2 points.

The A_1^2 nature of points on M is instrumental in understanding how the medial axis can be found from the volume of an object. Suppose $\underline{p}$ is of type A_1^2. $\underline{p}$ is equidistant to two distinct nearest points on the boundary of $\underline{\underline{m}}$ and must lie on its Voronoi diagram [45]. One method for finding the medial axis of $\underline{\underline{m}}$ is to prune the Voronoi diagram for its boundary [80].

Consider the function $d_{\underline{\underline{m}}} : \Omega \rightarrow \mathbb{R}^+$, where $d_{\underline{\underline{m}}}\left(\underline{x}\right)$ is defined as the Euclidean distance from $\underline{x}$ to the nearest point on the boundary of $\underline{\underline{m}}$. The value of $d_{\underline{\underline{m}}}\left(\underline{p}\right)$ is equivalent to the radius of $\mathcal{B}\left(\underline{p}\right)$. Since $\mathcal{B}\left(\underline{p}\right)$ is tangent to $\underline{\underline{m}}$ at the point of contact, they share the same normal direction. This direction can be found as the gradient of the distance map, *i.e.,* $\nabla d_{\underline{\underline{m}}}$. The fact that $\mathcal{B}\left(\underline{p}\right)$ makes contact with two nearest points on the boundary of $\underline{\underline{m}}$ gives rise to the grass-fire [68] method for finding the medial axis. In this analogy suppose that the volume of $\underline{\underline{m}}$ is full of evenly distributed prairie grass and that the grass at all boundary points is ignited simultaneously. The fire will burn inwards, along $\nabla d_{\underline{\underline{m}}}$, at a constant speed. The places where two fires quench each other are points on the medial locus. The time it takes for the fire to reach such a point is equivalent to the radius of the maximal inscribed ball at that point.

If the center and radius of each of the maximal inscribed balls is known, the boundary of $\underline{\underline{m}}$ can be reconstructed from an envelope around the balls. Each tuple of $\left(\underline{p}, d_{\underline{\underline{m}}}\left(\underline{p}\right)\right)$ is known as a order 0 medial atom. Although all information about $\underline{\underline{m}}$ is captured in its order 0 medial atoms, it is often convenient to have an explicit model of where $\mathcal{B}\left(\underline{p}\right)$ makes contact with the boundary of $\underline{\underline{m}}$. An order 1 medial atom extends the order 0 medial atom with a pair of vectors from $\underline{p}$ to the points of contact. The collection of all order 1 medial atoms for $\underline{\underline{m}}$ are known as the *m-rep* for that object.

When the medial axis of an object in $\mathbb{R}^2$ is decomposed into curves of A_1^2 points, the medial atoms on each curve can be understood as continuous functions of a single parameter u, *e.g.,* normalized arc length along the curve. In this case the m-rep for the object is the set of medial atom functions from each A_1^2 curve. Typical notation is to define the order 0 medial atoms with two continuous functions $\underline{X} : \{u\} \rightarrow \mathbb{R}^2$ and $r : \{u\} \rightarrow \mathbb{R}^+$ that give the central position and radius for each maximal inscribed ball. The order 1 medial atoms require two additional

continuous functions $\underline{U}^{+1} : \{u\} \rightarrow \mathbb{S}^1$ and $\underline{U}^{-1} : \{u\} \rightarrow \mathbb{S}^1$ that each give a unit vector in the direction of a point of contact. That is, the contact points are at $\underline{X}(u) + r(u)\,\underline{U}^{+1}(u)$ and $\underline{X}(u) + r(u)\,\underline{U}^{-1}(u)$. Similarly the m-rep for an object in $\mathbb{R}^3$ can be understood as the set of functions that produce medial atoms for each sheet of A_1^2 points on the medial locus. Each sheet can be indexed by a pair of parameters denoted by the tuple $\underline{u}$, and its medial atoms can be understood as the following continuous functions: $\underline{X} : \left\{\underline{u}\right\} \rightarrow \mathbb{R}^3$, $r : \left\{\underline{u}\right\} \rightarrow \mathbb{R}^+$, $\underline{U}^{+1} : \left\{\underline{u}\right\} \rightarrow \mathbb{S}^2$, and $\underline{U}^{-1} : \left\{\underline{u}\right\} \rightarrow \mathbb{S}^2$.

Damon [22] defined a generalization of a medial sheet known as a skeletal structure for an object. The skeletal structure consists of two-sided sets of points and their boundaries (*i.e.,* curves and their end-points in $\mathbb{R}^2$ or sheets and their end-curves in $\mathbb{R}^3$). An outward vector field that points to the boundary of the object is defined on each side of each set of points. Each of these vector fields is understood as the product of a scalar function r and a unit vector field $\underline{U}$. It is clear that an m-rep that uses order 1 medial atoms fits the definition of a skeletal structure.

Thus far I have only discussed the transformation from a boundary representation of a shape into a medial representations. It is interesting to think about transformations in the other direction. One could define an arbitrary skeletal structure and think about the boundary it implies. Damon invented the radial shape operator S_{rad} to link the differential geometry of a skeletal structure with the differential geometry of its implied boundary.

S_{rad} is calculated locally for one side of the skeletal structure and separately for the other. When computing S_{rad} for a medial sheet M, $\underline{U}$ can be assumed to consistently be chosen as either $\underline{U}^{+1}$ or $\underline{U}^{-1}$. The algorithm for computing S_{rad} at for a medial atom on M follows.

Suppose $\underline{v}$ is a vector that is tangent to the medial sheet at $\underline{p}$: $\underline{v} \in T_{\underline{p}}M$. S_{rad} is an operator that takes $\underline{v}$ to the vector in $T_{\underline{p}}$ that is opposite the projection along $\underline{U}$ into $T_{\underline{p}}$ of the directional derivative of $\underline{U}$ with respect to $\underline{v}$:

$$S_{rad}\left(\underline{v}\right) \;=\; -\mathrm{proj}_{\underline{U}}\left(\frac{\partial \underline{U}}{\partial \underline{v}}\right) \tag{2.58}$$

Suppose X' is the boundary point that corresponds to the medial atom at $\underline{X}\left(\underline{u}\right)$.

$$X' \;=\; \underline{X}\left(\underline{u}\right) + r\left(\underline{u}\right)\underline{U}\left(\underline{u}\right) \tag{2.59}$$

When S_{rad} at $\underline{X}\left(\underline{u}\right)$ is written in coordinates as a matrix, it can be related to the differential geometry shape operator S at X' as follows.

$$S \;=\; \left(\mathrm{Id} - r\left(\underline{u}\right)S_{rad}\right)^{-1} S_{rad}\ , \tag{2.60}$$

where Id is the identity matrix. From (2.60), Damon has shown that the eigenvalues of S: $\{\kappa_1, \kappa_2\}$, *i.e.,* the principal curvatures of $\underline{\underline{m}}$ at X' have the following relationship with the eigenvalues of S_{rad}: $\{\kappa_{r1}, \kappa_{r2}\}$.

$$\kappa_i \;=\; \frac{\kappa_{ri}}{1 - r\left(\underline{u}\right)\kappa_{ri}}\ i \in \{1, 2\} \tag{2.61}$$

Because $\underline{\underline{m}}$ will have a fold in it at X' when $\kappa_i = \infty$ and because κ_{ri} is a continuous function for which 0 is a valid value, Damon is able to derive the compatibility condition (2.62) for legal boundaries from (2.61).

$$\kappa_{ri} \;<\; \frac{1}{r\left(\underline{u}\right)} \tag{2.62}$$

S_{rad} is well defined for the A_1^2 points on a medial sheet. Damon defined an analogous operator S_E, known as the radial edge operator that is valid at the A_3 points on the boundary of a medial sheet. S_E can be used to define compatibility conditions along that boundary.

It has been shown that an object $\underline{\underline{m}}$ that satisfies (2.62) everywhere along each medial sheet and thus has no folds in its surface will have the property that none of its medial atoms have crossing spoke vectors inside $\underline{\underline{m}}$. It follows that any point $\underline{x}$ on the interior of $\underline{\underline{m}}$ can be assigned unique coordinates in the following way. The first coordinate is used to identify a figure, *i.e.,* to select a sheet of A_1^2 points from the medial axis of $\underline{\underline{m}}$. The next coordinates, $\underline{u}$ are used to select a medial atom from that sheet. The final coordinates, $\phi \in \{-1, +1\}$ and

$\tau : 0 \leq \tau \leq 1$ identify a spoke vector and a fractional radius such that for some figure

$$\underline{x} \;=\; \underline{X}\left(\underline{u}\right) + \tau r\left(\underline{u}\right)\underline{U}^{\phi}\left(\underline{u}\right) . \tag{2.63}$$

This object-relative coordinate system for the entire volume of the object is one of the important strengths of m-reps.

A weakness of the m-rep is that the number of medial sheets and the connections between medial sheets can be unstable with respect to relatively small surface deformations. Moreover, there may be medial atoms on a branching medial axis that have little information about the object boundary and only seem to exist to allow the medial axis to continue to a more salient location. Blum labels such medial atom as "ligature". The sampled m-rep, which henceforth will be abbreviated to m-rep[1], offers many of the benefits of the continuous medial axis, such as the existence of Damon's radial and edge shape operators and an object-relative coordinate system, but is explicitly designed to avoid the instability of the number and connectivity of medial sheets and to provide a compact representation where all atoms make salient contributions to the object boundary.

The m-rep consists of sheets of sampled order 1 medial atoms that are arranged on a curve for objects with (nearly) circular cross-section and in a lattice for more general objects. In the m-rep atoms at the end of each curve or lattice are not strictly medial atoms. Let $\underline{u}'$ denote the $\underline{u}$ coordinate of an end atom. A_3 edge points are approximated by a medial atom at an A_1^2 location that is augmented by another scalar function, $\eta : \left\{\underline{u}'\right\} \rightarrow \mathbb{R}^+$. η is known as the elongation factor, and it implies another position on the boundary of $\underline{\underline{m}}$ at

$$\underline{X}'\left(\underline{u}'\right) \;=\; \underline{X}\left(\underline{u}'\right) + \frac{\eta\left(\underline{u}'\right) r\left(\underline{u}'\right)}{2}\left(\underline{U}^{+1}\left(\underline{u}'\right) + \underline{U}^{-1}\left(\underline{u}'\right)\right) \tag{2.64}$$

The outward normal direction to $\underline{\underline{m}}$ at $\underline{X}'\left(\underline{u}'\right)$ is given by the spoke bisector $\frac{1}{2}\left(\underline{U}^{+1}\left(\underline{u}'\right) + \underline{U}^{-1}\left(\underline{u}'\right)\right)$.

The point $\underline{X}'\left(\underline{u}\right)$ is an extremum of curvature on the boundary of $\underline{\underline{m}}$, as is the place where an A_3 point on a continuous medial locus makes contact with its boundary. It is a sample of

[1]The entity described by my previous definition of m-rep will henceforth be written as "parameterized m-rep."

the crest curve on the boundary of $\underline{\underline{m}}$. The curve formed by interpolating between the three spoke-ends of an end-atom explains the portion of the boundary that would be implied by medial atoms that lie between that sampled medial atom and the A_1^3 point on the blum medial locus. Although these end atoms are not strictly medial, they do meet Damon's criteria for skeletal structures.

Han [37] has developed a method for interpolation between the relatively sparsely sampled medial atoms of an m-rep. This interpolation allows a parameterized m-rep to be recovered from the samples of an m-rep. This is useful for understanding the object-relative coordinate system for the entire object volume.

When an m-rep is formed from more than one sheet, the linkage is specified so that atoms of one medial sheet, known as the subfigure, are placed at specified boundary positions on another figure, known as the host. A set of parameters define a new surface that blends between the boundaries implied by the host and subfigure. The blending algorithm and its parameters are described in greater detail in Han's doctoral dissertation [35]. In short, the connection can be understood as a generalization of a skeletal structure in which the two sides of the hubs on the skeletal structure pull away from each other. A one-sided skeletal curve splits off from the medial axis of the host figure and approximates a ligature section of the subfigure. This one-sided skeletal curve ends at a place where it meets the traditional medial axis for the salient portion of the subfigure and another one-sided skeletal curve that approximates the "other" side of the medial axis in the ligature region. Han has been able to calculate S_{rad} and evaluate the compatibility condition for the one-sided skeletal structures, and thus he has been able to enforce geometric propriety in the blend region.

By construction, m-reps avoid changes to their meta-parameters: the number of medial lattices or curves, the connections between them, and the sampling of each. The m-rep instance for a new shape is produced by adjusting the atoms on a template m-rep to make its boundary best match the boundary of the target object. This depends of having access to an appropriate template m-rep.

Styner [78] has developed a method for identifying the meta-parameters of an m-rep for a new population. A default m-rep can be produced with those meta-parameters, and through

a series of automatic [36] or manual edits, it can be made to match the target object.

Because of these advantages: its compact representation, the object-relative coordinate system for the object volume, the geometric benefits of Damon's radial and edge shape operators, and the stability of the subfigure sheets, m-reps are attractive for use in object-based segmentation. The details of m-rep-based segmentation are given in Section 2.5.3.

2.5 Popular Object- and Atlas- Based Segmentation Methods

I survey three popular object-based segmentation techniques: geodesic active contours [92, 12], the active shape model (ASM) and point distribution model (PDM) [19, 43, 88], and the deformable m-rep [65, 47, 50], in Sections 2.5.1 - 2.5.3 respectively. For each of these methods I describe the shape representation and its geometric constraints as well as the optimization method and its cost function or probability distribution(s).

The ASM and the m-rep can each be used to define a statistical deformable shape model (SDSM) for an object class. SDSMs are commonly evaluated based on their generalization ability: a measure of how well the SDSM can model a new instance of the same object class and based on their specificity: a measure of the extent to which objects generated by SDSM belong to the intended object class. Precise definitions of generalization ability and sensitivity are given by Styner *et al.* [81].

I discuss a fourth segmentation method: registration of an atlas image [16, 72, 41, 27], in Section 2.5.4. This method is not typically considered object-based, but there are opportunities for interaction between object-based and atlas-based segmentations that make it relevant to this dissertation.

2.5.1 Snakes, Geodesic Active Contours

Kass *et al.* introduce *snakes* as energy minimizing spline [42] that is attracted to salient edges and contours in an image. In their notation $\mathbf{v}(s) = (x(s), y(s))$ represents a curve $\mathbf{v}$ in 2D that is parameterized by its fractional arc length s. The energy functional that they seek

to minimize is:

$$E^*_{\text{snake}} = \int_0^1 E_{\text{int}}\left(\mathbf{v}\left(s\right)\right) + E_{\text{image}}\left(\mathbf{v}\left(s\right)\right) + E_{\text{con}}\left(\mathbf{v}\left(s\right)\right) ds \quad . \tag{2.65}$$

The three terms in this energy functional E_{int}, E_{image}, and E_{con} represent energies due to the internal bending of the spline, due to image forces, and due to external constraints respectively. The spline is evolved to a configuration that minimizes the energy functional for a given image.

The external constraint energy E_{con} in (2.65) provides a way for a user to steer the snake. A spring force can be used to penalize the snake for moving a specified point away from a designated location, and a repulsion force can be used to penalize the snake for moving towards a designated location. When these external constraints are specified, the snake produces a semi-automatic segmentation. The other two energies in (2.65), the internal spline state E_{int} and the image energy E_{image}, require no human input, and when they are used without E_{con} the snake produces a fully automatic segmentation.

During the 20 years since [42] was published, the snakes methodology has evolved into a widely used technology known as *geodesic active contours* [92, 12]. A change in the shape representation is one of the most important improvements of these newer methods over the original snakes.

In the geodesic active contours framework the shape model $\underline{\underline{m}}_L$ used to segment L in an n-dimensional image is implicitly represented by the zero level set of a scalar-valued function on the imaging domain:

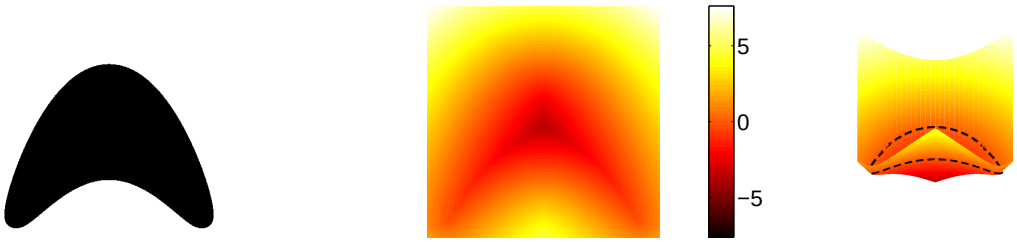
$$\phi : \Omega \rightarrow \mathbb{R}, \tag{2.66}$$

rather than by the explicit parameterization used in the original snakes. Typically the signed distance map (SDM) is used as ϕ [75]. Let $\underline{x} \in \Omega$ and let $d\left(\underline{x}\right)$ denote the distance from $\underline{x}$ to

the nearest point on the boundary implied by ϕ .

$$\phi\left(\underline{x}\right) = \begin{cases} -d\left(\underline{x}\right), & \underline{x} \text{ is on the interior of the object} \\ d\left(\underline{x}\right), & \underline{x} \text{ is on the exterior of the object} \end{cases} \tag{2.67}$$

Negative distances in this map are associated with the interior of the object, positive distances are associated with the exterior, and the zero level set implies the boundary of the object. An example of this shape representation can be seen in Figure 2.5.



(a) The binary map for an object in $\mathbb{R}^2$

(b) The signed distance map for this object

(c) The signed distance map interpreted as a height function in $\mathbb{R}^3$

Figure 2.5: Implicit shape representations.

This implicit shape representation has several useful properties. It allows the topology of $\underline{\underline{m}}_L$to change during the segmentation process. Such a change would not be possible with the original snake, or indeed with the ASM [19] or m-rep [36] shape representations discussed in Sections 2.5.2 and 2.5.3. This change in topology is desirable when the segmentation target is likely to have a different topology than the shape template, for example, when the segmentation target is in the presence of pathology. However, a change in topology is undesirable when the target has a known, fixed topology. Bai et. al [3] have demonstrated an extension of this method that is constrained to a fixed topology, but their constraint is computationally expensive.

Another useful property of the implicit shape representation is that it is computationally efficient to determine whether any given voxel lies on the interior or the exterior of the object.

All one needs to do is query the sign of $\phi\left(\underline{x}\right)$.

$$S_L = \left\{\underline{x} : \phi\left(\underline{x}\right) \leq 0\right\} \tag{2.68}$$

This allows one to design (2.65) so that the E_{image} term encourages different intensity distributions for the interior and exterior of the object. Let μ and ν denote the average image intensity in the interior and exterior of the object respectively.

$$\mu = \underset{\underline{x} \in S_L}{\text{mean}}\left(\underline{I}\left(\underline{x}\right)\right) \tag{2.69}$$

$$\nu = \underset{\underline{x} \notin S_L}{\text{mean}}\left(\underline{I}\left(\underline{x}\right)\right) \tag{2.70}$$

The functional popularized by Chan and Vese [14],

$$E_{\text{image}} = \int_{\underline{x} \in S_L}\left(\underline{I}\left(\underline{x}\right) - \mu\right)^2 + \int_{\underline{x} \ni S_L}\left(\underline{I}\left(\underline{x}\right) - \nu\right)^2, \tag{2.71}$$

encourages this separation of regions.

During the segmentation process, (2.65) is minimized, typically by following a gradient descent procedure for which the gradient formulas are known *a priori*. This allows the entire segmentation to be performed quickly, and these level-set methods are frequently used in real-time applications driven by sonography or fluoroscopy.

As the shape deforms during the segmentation process, the evolution of ϕ can be limited to translation along its normals. To understand this, it is useful to consider that any translation in a tangent plane to the surface would constitute a reparameterization of surface. Because the correspondences produced by this shape representation are based on voxels in Ω , rather than some anatomical property of the target object or some other geometric property of ϕ, allowing such a reparameterization would have no effect on the implied shape.

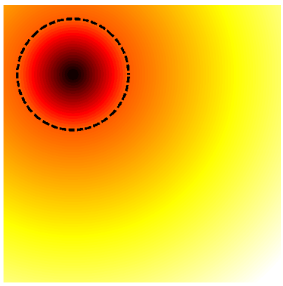
This optimization typically can only be guaranteed to produce a segmentation result that is a local minimizer of the energy function. In general the desired segmentation result, *i.e.,* the global minimum of (2.65), the segmentation needs to be initialized suitably close to that answer.

The tunneling descent method by Tao and Tagare [83] is able to relax that requirement by imposing new restrictions on the image data. They make the assumption that the image is such that one can tell when the segmentation boundary is entirely within the object and when it is entirely outside of the object. They seed the optimization with an initial shape on the interior of the target object. During the each iteration of the optimization the object is required to grow. They escape from local minima by growing the object in the way that causes the smallest increase in the cost function. This optimization procedure runs until the entire segmentation boundary is outside of the target object. At that point they choose the object with the lowest cost from the set of local minima that they had encountered, and they return that object as the segmentation of the image.

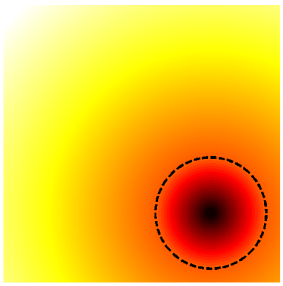
Two relative weaknesses of the level-set methods are exposed when they are used to study populations of objects. As was mentioned earlier, the implicit shape representation produces voxel-based correspondences. This is a weakness with regard to morphometric studies that require anatomical or geometric correspondences. Although an explicit shape representation may provide these correspondences directly, further processing, *e.g.,* by the method proposed by Cates *et al.* [13], is needed to build these correspondences for implicitly represented objects. The other major weakness of the level-set methods is that there does not exist a principled means to learn a probability distribution on the shapes themselves.

Several research groups [86, 46, 69] have proposed using PCA on a vector formed by the concatenation of all values from the SDM. A major problem is that SDMs lie in a nonlinear space. This problem is illustrated in Figure 2.6. The figure shows the boundaries and the SDMs from two example objects in $\mathbb{R}^2$. The average of those two SDMs is not another SDM. In fact, because the two example objects are non-overlapping and well separated, this average function does not even have a zero level-set. In general new vectors produced by (2.44) will not correspond to new SDMs, but they will however describe functions $\Omega \to \mathbb{R}^2$. In practice, the training SDMs could be aligned so that a new instance produced by (2.44) will have a zero level-set that would be interpreted as the new object boundary.

Another problem with using PCA on SDMs is that it is a memory-intensive task. In particular, when working with objects in $\mathbb{R}^3$, the number of voxels n in each SDM can be quite



(a) Example SDM with a ball in the upper left corner



(b) Example SDM with a ball in the lower right corner

(c) The average of these two SDMs

Figure 2.6: The linear average of two SDMs need not be a SDM itself. The SDMs for two balls are shown in Figure 2.6(a) and Figure 2.6(b) with a dashed line through the zero level-set. The two balls are non-overlapping. The linear average of these two SDMs is shown in Figure 2.6(c). This SDM does not imply a shape: it has no zero level-set.

large. For example, an SDM that occupies 256^3 voxels, each containing a double-precision (8 byte) distance, requires 128 megabytes of storage. When working with a population of k objects, keeping all the SDMs in memory becomes problematic for relatively small k. A $256^3 \times 256^3$ covariance matrix, with 8 byte elements requires 2 petabytes of storage, making it completely impractical with commodity hardware that is available today.

There is a trick that allows the eigenvalues and eigenvectors of a large covariance matrix to be computed without ever constructing the covariance matrix itself. Here I will use the notation of Section 2.3.1, assuming $\underline{\underline{x}}'$ is the $n \times k$ matrix formed by repeated application of (2.23) for $i \in \{1..k\}$ and Σ is the unknown $n \times n$ covariance matrix. We can compute a $k \times k$ symmetric positive semi-definite matrix Σ':

$$\Sigma' = \frac{1}{k-1} \left(\underline{\underline{x}}'\right)^{\mathrm{T}} \left(\underline{\underline{x}}'\right) \tag{2.72}$$

and its eigenvalues $\Lambda' = \{\lambda'_i\}$ and eigenvectors V'. It follows that for $i \in \{1..k\}$

$$\Sigma \underline{\underline{x}}' V'_i = \underline{\underline{x}}' \Sigma' V'_i \tag{2.73}$$

$$= \lambda'_i \underline{\underline{x}}' V'_i \tag{2.74}$$

Thus, each eigenvector of Σ can be found by premultiplying an eigenvector of Σ' by $\underline{\underline{x}}'$ and then normalizing the result to a unit-vector. Its eigenvalue will be the corresponding eigenvalue of Σ'.

This method allows us to perform PCA on a set of SDMs without having to compute the $n \times n$ covariance matrix. However, doing so remains a computationally difficult task due to the requirements that we store the $n \times k$ matrix $\underline{\underline{x}}'$ and multiply vectors by it. Moreover, our concern than the SDMs do not lie in a vector space has not been addressed.

Each of the next two segmentation methods that I present, ASMs and m-reps, use a shape representation that leads to a more principled SDSM. Both shape representations are compact: each shape is typically represented by a small number of samples. Even when working with objects from high resolution three dimensional images, storing an $n \times n$ matrix in memory and computing its eigenvalues can be done on commodity hardware. In Section 2.5.2, I will show that ASMs use a linear shape representation to which PCA can be directly applied. M-reps lie in a non-linear space, but as I will explain in Section 2.5.3, they can be mapped into a linear space where PCA is applicable.

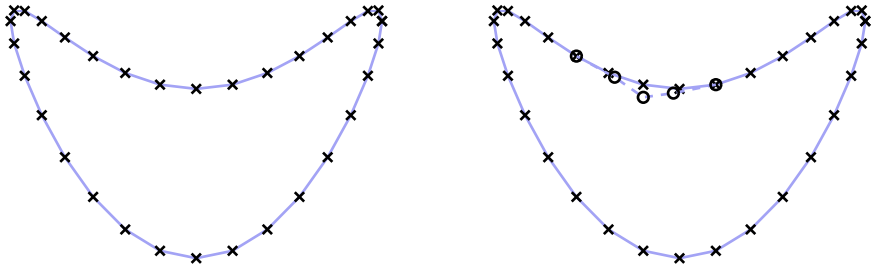
2.5.2 Active Shape Models

Active shape models (ASMs) [18, 19] are an object-based segmentation method based on an SDSM. In the classical ASM, shape representation is provided by a point distribution model (PDM). The PDM is a dense set of landmarks, typically on the boundary of S_L. The PDM is written as a tuple of these landmark coordinates. Let k denote the number of landmarks.

$$\underline{\underline{m}}_L = \begin{pmatrix} x_1 & y_1 & [z_1] & \ldots & x_k & y_k & [z_k] \end{pmatrix}^{\mathrm{T}} \tag{2.75}$$

A sample PDM is shown in Figure 2.7(a).

To build an ASM, one must first produce PDMs from a population of shapes. It is assumed that the number of boundary samples k from each PDM is fixed and that the landmarks are placed with some meaningful correspondence. That is, the i^{th} landmark, where $i \in \{1..k\}$, in each PDM instance is placed in a location that is equivalent across the population according to



(a) Point distribution model. Landmarks are indicated with an 'x'

(b) A perturbation is introduced. Modified landmarks are indicated with an 'o'

Figure 2.7: A point distribution model for a shape similar to the implicitly described object shown in Figure 2.5 is given in Figure 2.7(a). A perturbation of this object is shown in Figure 2.7(b)

some anatomic or geometric criteria. The ASM is trained by applying PCA to the collection of PDM tuples and truncating to a relatively small number of eigenmodes of variation.

The generalization ability and specificity of the ASM depend on the correspondences between landmarks across the training PDM instances. While it may be possible to manually place landmarks with adequate correspondences into two-dimensional images, it is a difficult and time consuming task. In three dimensions, this task is nearly impossible. To that end, Davies *et al.* [23, 24] provide an information theoretic method to produce PDMs with minimal description length for objects with spherical topology. Their method produces a homeomorphic map from a triangulated mesh of the object's surface onto a sphere and then minimizes the description length for the entire population of objects by manipulating the spherical coordinates of each landmark through further homeomorphic maps of the sphere. Cates *et al.* [13] provide a method for automatic landmark placement that is not dependent on the spherical topology of the target objects. This method maximizes the differential entropy of a set of particles on the surface of each object, to achieve adequate uniformity of spacing on the surface, while minimizing the differential entropy of each set of corresponding particles across the training population, to tighten the probability distribution on the corresponding particles. A PDM for each object is produced with landmarks positioned at the final particle locations.

The alignment of the training exemplars is another topic that affects the generalization ability of an ASM. I will refer to the portion of the changes between two objects that can be

described a global similarity transformation as "pose" changes the remainder of changes as "shape" changes. Because pose changes are global, they tend to produce more variation in the training data than do shape changes, which are often understood locally. Thus, pose changes are likely to dominate the largest eigenmodes of variation when the training exemplars are unaligned. In the HDLSS setting there may not be enough remaining eigenmodes to adequately describe shape changes in unaligned objects.

Alignment based on the Procrustes distance [7, 31], which measures the shape change between a pair of shapes, can be used to prevent pose changes from dominating the eigenmodes of variation for an ASM. The Procrustes distance between a pair of PDM instances, $\underline{\underline{m}}_1$ and $\underline{\underline{m}}_2$, is defined as

$$d_{\text{pro}}\left(\underline{\underline{m}}_1, \underline{\underline{m}}_2\right) = \min_{\mathcal{T}} \left\| \underline{\underline{m}}_1 - \mathcal{T}\left[\underline{\underline{m}}_2\right] \right\| , \tag{2.76}$$

where $\mathcal{T}$ is a similarity transformation, $\mathcal{T}[\cdot]$ denotes the application of $\mathcal{T}$ to the PDM, and $\|\cdot\|$ denotes a Euclidean distance.

Unless $\underline{\underline{m}}_1$ and $\underline{\underline{m}}_2$ have been scale normalized, the Procrustes distance is asymmetric: $d_{\text{pro}}\left(\underline{\underline{m}}_1, \underline{\underline{m}}_2\right) \neq d_{\text{pro}}\left(\underline{\underline{m}}_2, \underline{\underline{m}}_1\right)$ [74]. Procrustes alignment [56] was designed to be robust against the order of the objects and requires a symmetric distance function. In the first steps of the alignment, the each object $\underline{\underline{m}}_i$ is normalized to $\underline{\underline{m}}_i{}'$ that is centered at the origin and has unit total variance of its landmark positions. The rotation that minimizes (2.76) for these normalized shapes can be used to produce a new shape instance $\underline{\underline{m}}_2^{\dagger}$ by aligning $\underline{\underline{m}}_2{}'$ to $\underline{\underline{m}}_1{}'$.

$$\begin{aligned} \underline{\underline{m}}_2^{\dagger} &= \mathcal{T}^{\dagger}\left[\underline{\underline{m}}_2{}'\right] \text{ where,} \\ \mathcal{T}^{\dagger} &= \arg\min_{\mathcal{T}} \left\| \underline{\underline{m}}_1{}' - \mathcal{T}\left[\underline{\underline{m}}_2{}'\right] \right\| \end{aligned} \tag{2.77}$$

and $\mathcal{T}^{\dagger}$ is constrained to be a rotation. This can be generalized to a set of objects by aligning each of the normalized objects to their mean.

Procrustes alignment allows the training of an SDSM that captures variations in shape without being corrupted by variations in pose. However, this can lead to complications when the model is used to segment a new image. Because the shape statistics do not allow pose

changes, it may not be possible to recover if the SDSM is mis-initialized into the target image. That is, if the similarity transformation from the aligned objects into the target image is mis-estimated, the model cannot deform in a way that corrects that error.

In Chapter 3, I will propose an application-specific optimization problem that can reduce the risk of mis-initialization when an SDSM that does not allow pose changes is used for bladder segmentation. An alternative to the separate initialization step that I propose is to train the SDSM on shapes that include some small changes in pose which will be present in, but will not dominate, the eigenmodes of deformation. For example, Székely *et al.* [82], use an ASM to model variation in hippocampal structure. They trained their ASM in a coordinate system where the entire brain was aligned across images. Small pose changes between hippocampi exist in this coordinate system and were captured by the ASM, but the brain-scale alignment prevents the SDSM from being dominated by large changes in pose.

Once the issues of correspondence and alignment are addressed, the ASM can be trained via PCA to produce a shape space (2.47). This shape space is then used for segmentation.

Segmentation via ASMs is posed as a local optimization problem. The optimization begins with the mean PDM used as the initial shape estimate. During each iteration of optimization, the ASM deforms by allowing each landmark to move in the direction normal to the shape boundary at that position, although the means of calculating this normal direction is not explicitly specified in the classic ASM papers [18, 19]. The distance that each landmark moves in this normal direction is found by solving a one-dimensional optimization problem to minimize a local image-match function of the form (2.11). After each iteration of optimization, the deformed PDM is projected back into the truncated PCA space (2.47)

Each PDM instance has a graph structure that is used to reconstruct the full boundary of S_L from the landmarks. ASM requires this structure to be fixed across the training population and preserves this structure for the new objects it produces. This differs from the level-set methods, which allow changes in topology. A weakness of ASM relative to the level-set methods is that ASM cannot prevent folding of the object, *i.e.,* the mapping between the volumes of two objects in the same ASM shape space can be one-to-many, while this is never the case with the level-set methods.

Because ASM segmentation produces a result that lies in the PCA shape space, the ASM makes use of a stronger geometric prior than the original level-set methods [12, 92]. The ASM was designed to generate new shapes that are the same types of objects as the training exemplars, while the level-set methods were originally designed merely to produce objects with smooth boundaries.

PCA on PDMs is principled in the sense that $\mathbb{R}^n$ is a vector space, so linear statistics may be applied. However, when the ASM is trained on models in a Procrustes aligned space, the PCA shape space can include objects that are not in the Procrustes space, *i.e.,* such that the total variance in the landmark positions is not precisely 1.0. Moreover, it is often convenient to describe geometric entities with oriented points rather than by points alone. This is the approach taken by m-reps, and in Section 2.5.3 I will describe how that shape representation allows natural shape variations to be describe efficiently. This distinction between points in linear space and oriented points in a nonlinear space will be important in Chapter 5 when I introduce the rotational-flows method for interpolating pairs of oriented points.

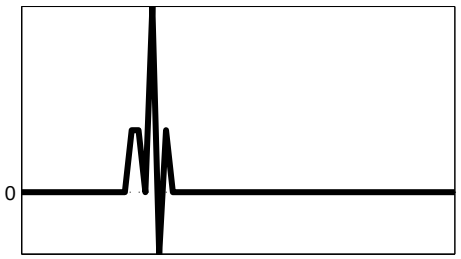
Improvements to both the image-match function and the shape representation have been proposed since the classical ASM was introduced.

Cootes *et al.* [18] designed the ASM optimization to be general with respect to the image-match function. They suggest that in simple cases edge strength can be used, but they also propose a more general image-match function computed as a Mahalanobis distance on a vector of normalized samples of the derivative of the image along the surface normal at each landmark. Since then, other image-match methods, including the following have been proposed. Kelemen *et al.* [43], compute a Mahalanobis distance on a vector of intensity values sampled along the surface normal at each landmark. Van Ginneken *et al.* [88] propose building a tuple of image features at each landmark position during training of the ASM and then applying machine learning techniques to identify an optimal subset of these features for use at segmentation time. Their image-match function minimizes the $L_{\max}$ distance from such a tuple to precomputed exemplar tuples for that landmark.

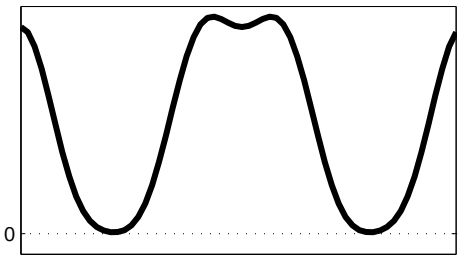
As an alternative to using PDMs as the underlying shape representation, ASMs have been extended to use parametric shape representations such as Fourier coefficients [82], spherical

harmonic (SPHARM) coefficients [43], and spherical wavelet coefficients [61]. An advantage that these parameterized shape representations hold over PDMs is that the parameterized model can be used to establish correspondences from an arbitrary set of boundary landmarks, but the PDM requires the landmark positions to correspond across training images.

Although the parametric shape representations relax the requirement for landmark correspondences, they also sacrifice the localizability of surface perturbations. PDMs, as well as SDMs, are able to locally describe small changes to the object boundary. An example of a perturbed PDM is illustrated in Figure 2.7(b). In this example the perturbation can be described by changes to just a few landmarks of the PDM. Similarly, it could be described by changes in a small neighborhood in an SDM. With the parametric representations, such a change may be described by changes to a set of coefficients related to a variety of frequencies or scales, as is illustrated in Figure 2.8. However, because global changes tend to be responsible for more variance in a shape population than are local changes and because global and local changes are merged in principal modes of variation, the localizability of surface perturbations in a PDM is unlikely to survive PCA training.



(a) The abscissa indexes into a PDM vector of the form $(x_1, y_1, ..., x_n, y_n)$. (2.75). The ordinate indicates the magnitude of the change at each coordinate between the objects shown in Figure 2.7.



(b) The abscissa indexes into a vector of Fourier coefficients. The ordinate inicates the magnitude of the change of that coefficient between the objects shown in Figure 2.7. Equivalently, the Fourier transform of the vector shown in Figure 2.8(a).

Figure 2.8: The changes between the objects shown in Figure 2.7. When the objects are represented in the format of (2.75), the changes are understood locally, as seen in Figure 2.8(a). When the objects are represented by a vector of Fourier coefficients, the changes occur across the spectrum, as seen in Figure 2.8(b).

The spherical wavelet representation has been touted as a means to circumvent the HDLSS problem. Spherical wavelets coefficients are calculated at a variety of localities and scales.

Nain *et al.* [61] propose computing PCA independently for various bands of coefficients at each scale. They claim that given k training exemplars they are able to learn up to $(k-1)$ modes of variation per band, when classical ASM can learn at most $(k-1)$ total modes of variation. They produce these additional modes of variation by ignoring correlations across the bands. They may be able to improve the generalization ability of their SDSM by sacrificing its specificity.

Segmentation via m-reps will be discussed in Section 2.5.3. Because the m-rep shape representation combines information about position and orientation, non-linear statistics are needed to describe a population of m-rep objects. These shape statistics are capable of describing non-linear transformations such as bending and twisting in a single eigenmode of deformation.

2.5.3 Discrete m-reps

Discrete m-reps, introduced in Section 2.4, are a shape representation that approximates the medial axis of an object. M-rep based segmentation is formulated as an optimization that in special cases can be understood as a MAP (2.7) problem. To simplify notation, I will only discuss the case where the target object in $\mathbb{R}^3$ can be described by a single lattice of medial atoms. However this basic method has been extended to handle objects with a nearly circular cross section that can be more stably described by a curve of medial atoms [73], objects with branching medial axes that are described by the host/subfigure relationship described in Section 2.4 [38], and multi-object complexes [67, 54].

Because the boundary of an m-rep is implied by the configuration of its medial atoms, precise manual editing of the boundary of $\underline{\underline{m}}$ can be a non-intuitive task. I will discuss Merck and Han's method for producing an m-rep that is well fit to an object modeled with some other shape representation [57, 36]. Then I will discuss Fletcher's procedure for learning a shape prior $p\left(\underline{\underline{m}}\right)$ [26] and Broadhurst's procedure for learning an image likelihood function $p\left(\underline{I}|\underline{\underline{m}}\right)$ [9] from a set of training m-reps. This section will conclude with a discussion of how these probability distributions are used at segmentation time.

Fitting training models

In order to learn $p\left(\underline{\underline{m}}\right)$ we need a set of training m-reps $\left\{\underline{\underline{m}}_i\right\}$ where i is an index over the training cases. My collaborators and I have developed a method [57, 36] that produces a geometrically legal m-rep instance $\underline{\underline{m}}_i$ fit to a reference segmentation of the corresponding gray scale image $\underline{I}_i$ in a training set. This method transforms each reference segmentation into an SDM and then applies an iterative procedure that deforms a sequence of template m-reps to match the zero level set of the SDM.

In the first iteration, a template m-rep with the desired configuration of medial atoms on its lattice is deformed to match each of the reference images. Preliminary m-reps $\underline{\underline{m}}_i^1$ are found during the following optimization over the space of all m-reps.

$$\underline{\underline{m}}_i^1 \quad = \quad \arg\min_{\underline{\underline{m}}} \left(\alpha d_{\mathrm{I}}\left(\underline{\underline{m}}\right) + \beta d_{\mathrm{LM}}\left(\underline{\underline{m}}\right) + \delta c_{\mathrm{srad}}\left(\underline{\underline{m}}\right) + \gamma c_{\mathrm{reg}}\left(\underline{\underline{m}}\right)\right) \ , \tag{2.78}$$

where $\alpha, \beta, \delta,$ and γ are non-negative weights and the functions d_{I}, d_{LM}, c_{srad}, and c_{reg} are defined as follows.

The function d_{I} measures the Euclidean distance from the boundary of the m-rep to the boundary of the reference segmentation. The boundary of the m-rep is found by enumerating the set of spoke-ends of the medial atoms according to (2.59) and (2.64). Because the set of medial atoms is sparse, an interpolation of the spoke-ends [84] or of the medial atoms [37] is done to produce a dense sampling of the boundary. d_{I} is implemented as the sum over this dense set of boundary points of the square of the Euclidean distance to the nearest point on the surface of the reference segmentations. By favoring a small value of d_{I}, the optimization (2.78) is likely to produce an m-rep with approximately the same boundary as the reference segmentation.

There is a potential ambiguity at this stage in the fitting process as there may be different m-rep instances that have approximately the same interpolated boundary. To relieve this ambiguity, the user is allowed to manually specify landmarks on the object boundary that link the image volume and object relative coordinate system. A landmark is given fixed object-

relative coordinates $\left(\underline{u}, \phi\right)$ for the entire training population; however, the user is responsible for assigning coordinates coordinates in Ω relative to the target image. The function d_{LM} is implemented as the sum over the set of these landmarks of the square of the Euclidean distance from the specified spoke end to the designated position in Ω . Because the optimization (2.78) favors small values of d_{LM}, the m-reps fit to a set of training images are likely to be in correspondence at the designated landmark coordinates.

The function c_{reg} also helps to establish correspondences in the object relative coordinate system and to resolve the ambiguity in the fitting stage. This function is a penalty for m-rep instances with irregularly spaced medial atoms. c_{reg} is implemented by summing over each of the medial atoms the square of its distance to the mean of its neighbors. The computations of the distances between medial atoms and of the mean of a set of medial atoms are done according to methods developed by Fletcher *et al.* [26]. These methods are summarized in the discussion of statistics on m-reps that will follow. The optimization (2.78) favors small values of c_{reg} and is likely to generate m-reps with evenly spaced medial atoms. Consequentially, the correspondences at designated landmark positions established by the d_{LM} term of (2.78) are likely to be extrapolated to all of the medial atoms.

The function c_{srad} is used to encourage (2.78) to produce m-reps with legal, *i.e.,* non-folding, surfaces. It is implemented as the sum over each medial atom of a continuous function of the eigenvalues of S_{rad} at that medial atom. This function is nearly zero for a medial atom that satisfies Damon's compatibility condition (2.62) but has a large positive value for a medial atom that violates the compatibility condition. Because (2.78) favors small values of c_{srad} it is likely to produce m-reps with geometrically legal boundaries.

The optimization (2.78) is performed at two scales. First, a coarse scale optimization is seeded with a template m-rep $\underline{\underline{m}}^0$. Candidate models $\underline{\underline{m}}$ are produced by applying a global similarity transformation to $\underline{\underline{m}}^0$. The parameters of this similarity transformation, *i.e.,* scale, angles of rotation, and translation, are allowed to vary during this stage of the optimization. The second stage of the optimization occurs at a finer scale, and is seeded with the minimizer of (2.78) from the coarse scale. At the fine scale, the medial atoms are selected one at a time and the parameters of the selected medial atom: the position, radius, spoke directions, and, if

applicable, the elongation factor are allowed to vary. This stage of the optimization converges after just a few iterations per medial atom.

After the fine scale optimization described above has completed, each of the training m-reps $\left\{\underline{\underline{m}}_i^1\right\}$ should be well fit to its reference segmentation, have a legal boundary, and be in correspondence with the other training m-reps via the object relative coordinate system. However, outlier cases will occasionally occur and are temporarily discarded. The successfully fit training m-reps are pooled, and using the methods that will be described shortly, a preliminary mean shape model $\underline{\underline{\mu}}^1$ and shape prior $p^1\left(\underline{\underline{m}}\right)$ are computed for use as an SDSM.

The next iteration of the optimization begins with the following coarse scale optimization seeded at $\underline{\underline{\mu}}^1$.

$$\underline{\underline{m}}_i^2 \quad = \quad \arg\min_{\underline{\underline{m}}} \left(\alpha d_{\mathrm{I}}\left(\underline{\underline{m}}\right) + \beta d_{\mathrm{mahal}}\left(\underline{\underline{m}}\right) + \delta c_{\mathrm{srad}}\left(\underline{\underline{m}}\right)\right) \tag{2.79}$$

This optimization is constrained so that each candidate m-rep $\underline{\underline{m}}$ can be generated by the SDSM $p^1\left(\underline{\underline{m}}\right)$. The functions d_{I} and c_{srad} have the same meaning as before, and d_{mahal} measures the Mahalanobis distance (2.18) from $\underline{\underline{m}}$ to the preliminary mean $\underline{\underline{\mu}}^1$ with respect to the probability distribution $p^1\left(\underline{\underline{m}}\right)$.

The d_{mahal} term in (2.79) in some sense replaces the d_{LM} and c_{reg} terms from (2.78). Since $p^1\left(\underline{\underline{m}}\right)$ was trained on m-reps with regularly spaced atoms and correspondences via the object relative coordinate system, any shape instance admitted with high probability by $p^1\left(\underline{\underline{m}}\right)$, *i.e.,* one with a low d_{mahal}, is expected to also have these properties.

The third iteration for each training case is seeded at $\underline{\underline{m}}_i^2$ and is intended to allow fine scale refinements that could not be achieved by the SDSM based fitting (2.79).

$$\underline{\underline{m}}_i^3 \quad = \quad \arg\min_{\underline{\underline{m}}} \left(\alpha d_{\mathrm{I}}\left(\underline{\underline{m}}\right) + \delta c_{\mathrm{srad}}\left(\underline{\underline{m}}\right) + \gamma c_{\mathrm{reg}}\left(\underline{\underline{m}}\right)\right) \tag{2.80}$$

The functions d_{I}, c_{srad}, and c_{reg} have the same meaning as in (2.78), but the regularity weight γ is typically relaxed for (2.80).

The set $\left\{\underline{\underline{m}}_i^3\right\}$ are typically well fit to the reference segmentations with regularly spaced medial atoms and correspondences in the object relative coordinate system and hence are suit-

able exemplars for training the probability distributions used at segmentation time. However, if the user is dissatisfied, they could use $\left\{\underline{\underline{m}}_i^3\right\}$ or hand edited m-reps in place of $\left\{\underline{\underline{m}}_i^1\right\}$ to retrain preliminary probability distributions for use in another iteration of (2.79) and (2.80).

Learning a prior distribution

Fletcher *et al.* have shown [26] that each medial atom of an m-rep lattice lies in a curved space and that linear statistics such the arithmetic mean and PCA are not appropriate in such a space. This section summarizes Fletcher's method for learning a shape prior $p\left(\underline{\underline{m}}\right)$ from a training population of m-reps. Aligning the training m-reps can tighten the shape prior and will be discussed later in this section. As a simplification, I will only address medial atoms found on the interior of the lattice, and I will omit the qualifier "on the interior of the lattice" for the remainder of this section. A straightforward extension that treats the elongation factor η like the radius r is all that is needed to apply these methods to the medial atoms on the boundary of the lattice.

Each medial atom lies in the space

$$\mathcal{M}(1) \quad = \quad \mathbb{R}^3 \times \mathbb{R}^+ \times \mathbb{S}^2 \times \mathbb{S}^2 \tag{2.81}$$

and a collection of n such medial atoms lies on the direct product space

$$\mathcal{M}(n) \quad = \quad \prod_{i=1}^{n} \mathcal{M}(1). \tag{2.82}$$

The spaces $\mathcal{M}(1)$ and $\mathcal{M}(n)$ are each Riemannian manifolds. The metric on $\mathcal{M}(1)$ (2.91) is explained below; the metric on $\mathcal{M}(n)$ is understood as the sum of the square metric distances between corresponding medial atoms.

$\mathcal{M}(1)$ can be understood as a symmetric space. The operator $\circ : \mathcal{M}(1) \rightarrow \mathcal{M}(1)$ is defined below (2.86) as is an identity element $1_{\mathcal{M}(1)} \in \mathcal{M}(1)$ (2.87) such that

$$\begin{aligned} 1_{\mathcal{M}(1)} \circ \underline{a}_1 \quad &= \quad \underline{a}_1 \circ 1_{\mathcal{M}(1)} \\ &= \quad \underline{a}_1 : \forall \underline{a}_1 \in \mathcal{M}(1). \end{aligned} \tag{2.83}$$

Let the function $\text{Rotation}_{\underline{v}_1 \to \underline{v}_2}$ be defined for $\underline{v}_1, \underline{v}_2 \in \mathbb{S}^2$ as rotation about the axis mutually orthogonal to $\underline{v}_1$ and $\underline{v}_2$ by the shortest angle such that $\text{Rotation}_{\underline{v}_1 \to \underline{v}_2}\left(\underline{v}_1\right) = \underline{v}_2$. Because the axis of rotation between $\underline{v}_1$ and $\pm\underline{v}_1$ is ambiguous, the function is defined as follows for those special cases.

$$\underset{\underline{v}_1 \to \underline{v}_1}{\text{Rotation}}\left(\underline{v}_2\right) = \underline{v}_2 \tag{2.84}$$

$$\underset{\underline{v}_1 \to -\underline{v}_1}{\text{Rotation}}\left(\underline{v}_2\right) = -\underline{v}_2 \tag{2.85}$$

Suppose $\left\{\underline{a}_i = \left(\underline{X}_i, r_i, \underline{U}_i^{+1}, \underline{U}_i^{-1}\right) : i \in \{1,2\}\right\} \subset \mathcal{M}(1)$.

$$\underline{a}_1 \circ \underline{a}_2 = \left(\underline{X}_1 + \underline{X}_2, r_1 r_2, \underset{(0,0,1) \to \underline{U}_1^{+1}}{\text{Rotation}}\left(\underline{U}_2^{+1}\right), \underset{(0,0,1) \to \underline{U}_1^{-1}}{\text{Rotation}}\left(\underline{U}_2^{-1}\right)\right), \tag{2.86}$$

$$1_{\mathcal{M}(1)} = \left((0,0,0), 1, (0,0,1), (0,0,1)\right) \tag{2.87}$$

It can be shown that each $\underline{a}_1 \in \mathcal{M}(1)$ has an inverse $\underline{a}_1{}^{-1} \in \mathcal{M}(1)$ such that

$$\begin{aligned} \underline{a}_1 \circ \underline{a}_1{}^{-1} &= \underline{a}_1{}^{-1} \circ \underline{a}_1 \\ &= 1_{\mathcal{M}(1)}. \end{aligned} \tag{2.88}$$

Fletcher designed his metric, which I will denote $\langle \cdot, \cdot \rangle_{\mathcal{M}(1)}$, to be left-invariant with respect to $\circ$:

$$\left\langle \underline{a}_1, \underline{a}_2 \right\rangle_{\mathcal{M}(1)} = \left\langle \underline{a}_3 \circ \underline{a}_1, \underline{a}_3 \circ \underline{a}_2 \right\rangle_{\mathcal{M}(1)} \; \forall \underline{a}_3 \in \mathcal{M}(1). \tag{2.89}$$

From (2.89) and the definition of $\left\langle 1_{\mathcal{M}(1)}, \cdot \right\rangle_{\mathcal{M}(1)}$ given below (2.90) we can derive $\langle \cdot, \cdot \rangle_{\mathcal{M}(1)}$ in (2.91).

$$\begin{aligned} \left\langle 1_{\mathcal{M}(1)}, \underline{a}_i \right\rangle_{\mathcal{M}(1)} &= \|\underline{X}_i\|^2 + \bar{r}\left(\log r_i\right)^2 + \bar{r}\left(\arccos\left((0,0,1) \cdot \underline{U}_i^{+1}\right)\right)^2 \\ &\quad + \bar{r}\left(\arccos\left((0,0,1) \cdot \underline{U}_i^{-1}\right)\right)^2 \end{aligned} \tag{2.90}$$

$$\left\langle \underline{a}_1, \underline{a}_2 \right\rangle_{\mathcal{M}(1)} = \left\langle 1_{\mathcal{M}(1)}, \underline{a}_1{}^{-1} \circ \underline{a}_2 \right\rangle_{\mathcal{M}(1)} \tag{2.91}$$

In (2.90), $\bar{r}$ is a constant that is needed because the hub position is specified in physical units, *e.g.,* mm, but the other components of $\mathcal{M}(1)$. It is defined with respect to a set of medial atoms (for example, corresponding medial atoms from a population of objects) as the geometric mean radius in that set. Suppose that the changes from the object boundary implied by $\underline{a}_1$ to the object boundaries implied by $\underline{a}_2 \circ \underline{a}_1$ and $\underline{a}_3 \circ \underline{a}_1$ are equal in magnitude. Fletcher's definition of $\bar{r}$ was intended to make $\left\langle \underline{a}_1, \underline{a}_2 \circ \underline{a}_1 \right\rangle_{\mathcal{M}(1)}$ commensurate with $\left\langle \underline{a}_1, \underline{a}_3 \circ \underline{a}_1 \right\rangle_{\mathcal{M}(1)}$. Pizer and Jeong have recently realized that this design goal is better fulfilled by the following definition of $\left\langle 1_{\mathcal{M}(1)}, \cdot \right\rangle_{\mathcal{M}(1)}$.

$$\begin{aligned} \left\langle 1_{\mathcal{M}(1)}, \underline{a}_i \right\rangle_{\mathcal{M}(1)} &= \|\underline{X}_i\|^2 + \bar{r}\,(\log r_i)^2 + \frac{\bar{r}}{2}\left(\arccos\left((0,0,1)\cdot \underline{U}_i^{+1}\right)\right)^2 \\ &\quad + \frac{\bar{r}}{2}\left(\arccos\left((0,0,1)\cdot \underline{U}_i^{-1}\right)\right)^2 \end{aligned} \tag{2.92}$$

However, the results presented in the latter chapters of this dissertation were all produced using Fletcher's definition of this metric (2.90).

Suppose $\underline{\underline{m}}_1, \underline{\underline{m}}_2 \in \mathcal{M}(n)$ and that for $i \in \{1,2\}$ and $j \in \{1..n\}$, the notation $\underline{a}_i^j \in \mathcal{M}(1)$ is used to refer to the j^{th} medial atom in object i. We define $\langle \cdot, \cdot \rangle_{\mathcal{M}(n)}$ as follows.

$$\left\langle \underline{\underline{m}}_1, \underline{\underline{m}}_2 \right\rangle_{\mathcal{M}(n)} = \sum_{j=1}^{n} \left\langle \underline{a}_1^j, \underline{a}_2^j \right\rangle_{\mathcal{M}(1)} \tag{2.93}$$

Fletcher applies this definition of distance on the manifold $\mathcal{M}(n)$ (2.93) to Fréchet's generalization of the mean. In particular, given a set of k m-reps $\left\{ \underline{\underline{m}}_i : i \in \{1..k\} \right\} \subset \mathcal{M}(n)$, their Fréchet mean $\underline{\underline{\mu}}$ is defined as

$$\underline{\underline{\mu}} = \arg\min_{\underline{\underline{m}}} \sum_{i=1}^{k} \left\langle \underline{\underline{m}}, \underline{\underline{m}}_i \right\rangle_{\mathcal{M}(n)} \tag{2.94}$$

Because $\mathcal{M}(n)$ is a Riemannian manifold, it can be approximated locally by a linear space. Fletcher's principal geodesic analysis (PGA) is based on computing eigenmodes of variation in the tangent hyperplane to the manifold at $\underline{\underline{\mu}}$, and converting the line along each eigenvector in the tangent space into a geodesic curve of the manifold.

Let $\underline{\underline{m}} \in \mathcal{M}(n)$ denote an m-rep and $\left\{\underline{a}_i : i \in \{1..n\}\right\} \subset \mathcal{M}(1)$ its medial atoms. The tangent hyperplane at $\underline{\underline{m}}$ can be understood as the direct product of the tangent hyperplanes at each atom:

$$T_{\underline{\underline{m}}}\mathcal{M}(n) \;=\; \prod_{i=1}^{n} T_{\underline{a}_i}\mathcal{M}(1) \tag{2.95}$$

Points can be mapped from the manifold to the tangent plane via the Riemannian Log map, and vice-versa via the Riemannian Exp map. Fletcher's definitions of $\mathrm{Log}_{1_{\mathcal{M}(1)}} : \mathcal{M}(1) \rightarrow T_{1_{\mathcal{M}(1)}}\mathcal{M}(1)$ and $\mathrm{Exp}_{1_{\mathcal{M}(1)}} : T_{1_{\mathcal{M}(1)}}\mathcal{M}(1) \rightarrow \mathcal{M}(1)$ will be given below. Then we can use the left-invariance of $\mathcal{M}(1)$ to define the maps anywhere on $\mathcal{M}(1)$ and the direct product (2.95) to define the maps on $\mathcal{M}(n)$.

Let $\mathrm{Log}_{(0,0,1)} : \mathbb{S}^2 \rightarrow T_{(0,0,1)}\mathbb{S}^2$ denote the Riemannian Log map at $(0,0,1)$ on the unit sphere. Let $\underline{v} = (v_1, v_2, v_3) \in \mathbb{S}^2$ denote any point on the sphere other than the antipodal point $(0,0,-1)$, which does not have a unique representation in $T_{(0,0,1)}\mathbb{S}^2$.

$$\mathrm{Log}_{(0,0,1)}\left(\underline{v}\right) \;=\; \begin{cases} \frac{\arccos(v_3)}{\sin(\arccos(v_3))}\,(v_1, v_2)\;, & \|v_3\| \neq 1 \\ (0,0)\;, & \underline{v} = (0,0,1) \end{cases} \tag{2.96}$$

Let $\underline{a} = \left(\underline{X}, r, \underline{U}^{+1}, \underline{U}^{-1}\right) \in \mathcal{M}(1)$. The definition of the Riemannian Log map at $1_{\mathcal{M}(1)}$ follows.

$$\mathrm{Log}_{1_{\mathcal{M}(1)}}\left(\underline{a}\right) \;=\; \left(\underline{X}, \bar{r}\log r, \bar{r}\mathrm{Log}_{(0,0,1)}\left(\underline{U}^{+1}\right), \bar{r}\mathrm{Log}_{(0,0,1)}\left(\underline{U}^{-1}\right)\right)\;, \tag{2.97}$$

where $\bar{r}$ is constant defined in (2.90) and is used to put the various components of $\mathrm{Log}_{1_{\mathcal{M}(1)}}\left(\underline{a}\right)$ into units commensurate with the hub position.

Let $\mathrm{Exp}_{(0,0,1)} : T_{(0,0,1)}\mathbb{S}^2 \rightarrow \mathbb{S}^2$ denote the Riemannian Exp map on the plane tangent to the sphere at $(0,0,1)$. Let $\underline{w} = (w_1, w_2) \in T_{(0,0,1)}\mathbb{S}^2$ denote a point in this tangent plane.

$$\mathrm{Exp}_{(0,0,1)}\left(\underline{w}\right) \;=\; \begin{cases} \left(w_1 \cdot \frac{\sin\|\underline{w}\|}{\|\underline{w}\|}, w_2 \cdot \frac{\sin\|\underline{w}\|}{\|\underline{w}\|}, \cos\|\underline{w}\|\right)\;, & \|\underline{w}\| > 0 \\ (0,0,1)\;, & \underline{w} = (0,0) \end{cases} \tag{2.98}$$

Let $\underline{\alpha} = \left(\underline{X}, \rho, \underline{V}^{+1}, \underline{V}^{-1}\right) \in T_{1_{\mathcal{M}(1)}}\mathcal{M}(1)$. The definition of the Riemannian Exp map at $1_{\mathcal{M}(1)}$ follows.

$$\mathrm{Exp}_{1_{\mathcal{M}(1)}}\left(\underline{\alpha}\right) = \left(\underline{X}, \exp\frac{\rho}{\bar{r}}, \mathrm{Exp}_{(0,0,1)}\left(\frac{\underline{V}^{+1}}{\bar{r}}\right), \mathrm{Exp}_{(0,0,1)}\left(\frac{\underline{V}^{-1}}{\bar{r}}\right)\right) \tag{2.99}$$

Fletcher's algorithm for PGA follows. We start with a set of k m-reps $\left\{\underline{\underline{m}}_i : i \in \{1..k\}\right\} \subset \mathcal{M}(n)$ and compute their Fréchet mean $\underline{\underline{\mu}}$ (2.94). Using the Riemannian Log map we project each $\underline{\underline{m}}_i$ into $T_{\underline{\underline{\mu}}}\mathcal{M}(n)$. By the left invariance of $\langle\cdot,\cdot\rangle_{\mathcal{M}(n)}$ we can define these tangent space projections as $\left\{\underline{\underline{\omega}}_i\right\}$ as follows.

$$\underline{\underline{\omega}}_i = \mathrm{Log}_{\underline{\underline{\mu}}}\left(\underline{\underline{m}}_i\right) \tag{2.100}$$

$$= \mathrm{Log}_{1_{\mathcal{M}(n)}}\left(\underline{\underline{\mu}}^{-1} \circ \underline{\underline{m}}_i\right) \tag{2.101}$$

Each tangent space projection $\underline{\underline{\omega}}_i$ can be understood as a vector in $\mathbb{R}^{8n}$. Dimensionality reducing PCA can be applied to the $8n \times k$ matrix $\begin{pmatrix} \underline{\underline{\omega}}_1 & \ldots & \underline{\underline{\omega}}_k \end{pmatrix}$.

As was the case with ASM, the alignment of training m-reps will affect the types of shape changes seen in the SDSM learned from them. Fletcher [26] defines a generalization of Procrustes alignment that is based on the metric distance in $\mathcal{M}(n)$ and is known as the geodesic Procrustes alignment. The algorithm for geodesic Procrustes alignment of a set of k m-reps $\left\{\underline{\underline{m}}_i\right\}$ follows. A similarity transformation $\mathcal{T}_i$ is created for each $\underline{\underline{m}}_i$ and is initialized to be the translation that puts the center of gravity of $\underline{\underline{m}}_i$ at the origin. Geodesic Procrustes alignment minimizes

$$\sum_{i=1}^{k}\sum_{j=1}^{i}\left\langle \mathcal{T}_i\left[\underline{\underline{m}}_i\right], \mathcal{T}_j\left[\underline{\underline{m}}_j\right]\right\rangle_{\mathcal{M}(n)} \tag{2.102}$$

by adjusting the rotation and scale parameters of each $\mathcal{T}_i$:

$$\mathcal{T}_i = \arg\min_{\mathcal{T}}\left\langle \mathcal{T}\left[\underline{\underline{m}}_i\right], \underline{\underline{\mu}}_i\right\rangle_{\mathcal{M}(n)}, \tag{2.103}$$

where $\underline{\underline{\mu}}_i$ denotes the Fréchet mean of $\left\{\underline{\underline{m}}_j : j \neq i\right\}$. The procedure applies the optimization (2.103) to the m-reps one at a time until (2.102) converges. Note that because each object

is being aligned to the others, the optimization cannot decrease (2.102) by setting the scale factor of any $\mathcal{T}_i$ to be arbitrarily small.

When PGA is applied to m-reps, PCA is applied to their projections into the tangent hyperplane at their Fréchet mean. Another m-rep $\underline{\underline{m}}$ can be projected into this tangent PCA space by aligning it to $\underline{\underline{\mu}}$ if necessary, then applying the Log map (2.100) at $\underline{\underline{\mu}}$, and then applying the PCA projection (2.47). This projection can be used to measure the Mahalanobis distance $d_{\text{mahal}}\left(\underline{\underline{m}}\right)$ from $\underline{\underline{\mu}}$ to $\underline{\underline{m}}$.

This PCA allows a new object $\underline{\underline{\omega}} \in T_{\underline{\underline{\mu}}}\mathcal{M}(n)$ to be generated via (2.44). The Riemannian Exp map can lift $\underline{\underline{\omega}}$ up to an m-rep $\underline{\underline{m}}$ on the manifold $\mathcal{M}(n)$.

$$\underline{\underline{m}} = \text{Exp}_{\underline{\underline{\mu}}}\left(\underline{\underline{\omega}}\right) \tag{2.104}$$

$$= \underline{\underline{\mu}} \circ \text{Exp}_{1_{\mathcal{M}(1)}}\left(\underline{\underline{\omega}}\right) \tag{2.105}$$

If the SDSM was learned from aligned m-reps, an initializing similarity transformation needs to be applied to the m-rep produced by (2.104).

PGA on m-reps can be understood as PCA in the tangent space. A single eigenmode of variation in the PGA space can describe naturally occurring non-linear transformations such as local bending, twisting, and magnification that are linear in the m-rep tangent space. These transformations could not be represented by a single eigenmode of variation a linear shape representation, *i.e.,* a PDM.

In this dissertation, I will focus on the segmentation of a single object that can be described by n medial atoms configured on a single medial lattice. A shape prior learned via PGA on a set of training m-reps that each lie in $\mathcal{M}(n)$ is adequate for my purposes. However, more complicated objects and object-complexes have been modeled by m-reps. Han *et al.* [38] generalized PGA so that it could be applied to an object that is described by an m-rep consisting of multiple lattices of medial atoms connected in host/subfigure relationships. Suppose such an object could be described by a host in $\mathcal{M}(n_1)$ and a subfigure in $\mathcal{M}(n_2)$. Han proposes applying PGA hierarchically, first to the host and subfigure simultaneously as an entity in $\mathcal{M}(n_1) \times \mathcal{M}(n_2) = \mathcal{M}(n_1 + n_2)$ and then sequentially to residuals on the host in

$\mathcal{M}(n_1)$ and the subfigure in $\mathcal{M}(n_2)$.

Lu *et al.* [54] are interested in a complex of objects, each of which can be described by an m-rep with a single lattice of medial atoms. They also apply a hierarchical form of PGA, first to the complex of objects in $\prod_i \mathcal{M}(n_i) = \mathcal{M}(\sum_i n_i)$, then in each $\mathcal{M}(n_i)$ on object-scale residues relative to the entire ensemble, and then finally in $\mathcal{M}(1)$ on residues from each atom relative to other nearby atoms within the individual objects. Related work by Pizer *et al.* [67] uses a hierarchical PGA on a complex of objects and can model interobject relationships by augmenting the object-scale representation of one object with a few medial atoms from neighboring objects. Further work by Pizer *et al.* [66] combines the aforementioned hierarchical PGA with a hierarchy of image intensity statistics. The image intensity statistics at each scale are learned according to the theory that I summarize in the next section. Liu *et al.*, yours truly included, [51] use a hierarchical PGA on a single figure m-rep in $\mathcal{M}(n)$ followed by fine scale PGA on individual atom residues in $\mathcal{M}(1)$ relative to the object-scale prediction of that atom. This scheme has been used to produce higher quality segmentations than a similar scheme based on object-scale PGA alone.

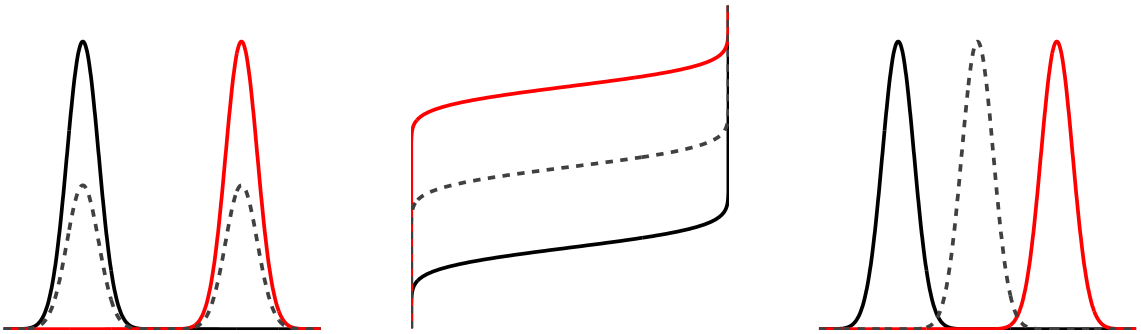
PGA is used to learn $p\left(\underline{\underline{m}}\right)$ for m-rep-based segmentation. Indeed, object-scale PGA is how the preliminary SDSM used to fit m-reps to reference segmentations (2.79) was learned.

Training an image likelihood distribution

In object-based segmentation, the image likelihood distribution $p\left(\underline{I}|\underline{\underline{m}}\right)$ gives the likelihood of the object $\underline{\underline{m}}$, given the image intensity patterns $\underline{I}$ that surround it. Broadhurst's method based on regional intensity quantile functions (RIQF) [9] is a principled method for learning $p\left(\underline{I}|\underline{\underline{m}}\right)$ from a set of training exemplars. In this section, I will summarize the theory of RIQFs and describe how they can can be applied to m-rep-based segmentation.

It is natural to represent the image in the neighborhood around $\underline{\underline{m}}$ as a histogram of voxel values. Given a set of training cases one would want to apply PCA to the set of histograms to learn $p\left(\underline{I}|\underline{\underline{m}}\right)$. However, when the intensity distributions from each training image all belong to the same family of distributions and vary only in location and scale, *i.e.,* in mean and variance, these changes are non-linear in the histogram representation. When the same

distribution is represented as a quantile function (QF), *i.e.,* by the inverse of its cumulative density function (CDF), these changes are linear as can been seen in Figure 2.9. Broadhurst has shown that regional intensity distributions from medical images frequently do vary only in location and scale, and thus, he argues for learning $p\left(\underline{I}|\underline{\underline{m}}\right)$ via PCA in the quantile function space.



(a) Histograms of samples from two Gaussian distributions that differ only in their means (red and black solid lines) and the arithmetic mean of the histograms (dashed line). The mean histogram is clearly non-Gaussian.

(b) Quantile functions representations for the same two Gaussian distributions (solid lines) and the arithmetic mean of the quantile functions (dashed line).

(c) The three quantile functions have been converted to histograms. The histogram from the mean quantile function (dashed line) is a Gaussian. Its mean is the average of the means of the original pair of Gaussians.

Figure 2.9: Linear averaging of probability distributions. The same distributions are represented as histograms in Figure 2.9(a) and as quantile functions in Figure 2.9(b). The quantile functions are drawn as histograms in Figure 2.9(c). The average quantile function averages the location of the original distributions, while the average of the histogram blends the original distributions.

Rather than learn $p\left(\underline{I}|\underline{\underline{m}}\right)$ from a single QF per training image, Broadhurst suggests decomposing the image into object relative regions. At a coarse scale the regions could correspond to the interior and exterior of $\underline{\underline{m}}$, similar to the regions defined by Chan and Vese [14]. At a fine scale, the interior and exterior of $\underline{\underline{m}}$ could each be decomposed into multiple regions. Suppose $\underline{I}$ is decomposed into m regions, $\left\{ \underline{I}^1, \ ..., \ \underline{I}^m \right\}$, the likelihood function can be understood as

$$p\left(\underline{I}|\underline{\underline{m}}\right) \quad = \quad p\left(\underline{I}^1, ..., \underline{I}^m|\underline{\underline{m}}\right) . \tag{2.106}$$

By assuming that intensities in each region are independent, (2.106) can be rewritten as follows:

$$p\left(\underline{\underline{I}}|\underline{\underline{m}}\right) = \prod_{i=1}^{m} p\left(\underline{\underline{I}}^i|\underline{\underline{m}}\right) \tag{2.107}$$

Each $p\left(\underline{\underline{I}}^i|\underline{\underline{m}}\right)$ in (2.107) is learned from PCA on RIQFs: QFs of image intensity from the region $\underline{\underline{I}}^i$.

The sampling of each RIQF can be customized via a series of parameters. Broadhurst associates each image region with a contiguous set of boundary points from the object $\underline{\underline{m}}$ and the outward normal direction at each of the boundary points. The first customizable attribute of an image region specifies whether it lies on the interior or exterior of $\underline{\underline{m}}$. The second customization defines the size of the region by excluding voxels that are more than a maximum distance from the nearest point from the set of boundary samples associated with the region.

Broadhurst uses weighted QFs (analogous to weighted histograms) in each region, so that voxels near the object boundary have heavier weights than far away points. One of the boundary points is given the special designation of "anchor point" for the region. Voxels near the designated anchor point have heavier weights than voxels near the edge of the region. The fall-off rates of each of these weighting schemes is customizable.

The final set of customizations has to do with special intensity values in the image region. For example, in CT images air and bowel gas have distinctly low intensities and dense bones have distinctly high intensities. These extreme values can corrupt an RIQF in a soft-tissue region. The corruption may cause the variation in corresponding RIQFs across a population of images to be manifest in ways other than changes in location and scale. This is especially a problem with bowel gas, since unless explicit measures are taken at image acquisition time, the quantity of bowel gas in a region can be unstable across images. To deal with this problem, Broadhurst allows threshold values to be set so that gas and bone voxels do not contribute to an RIQF. Depending on the configuration of the RIQF, the numbers of gas and bone voxels in the region can each be modeled as a univariate Gaussian, or the gas and bone voxels in the region can simply be discarded.

The configuration parameters need to be set in an application specific manner. Once that decision has been made, the training images can be divided into regions, RIQFs can be sampled, and if appropriate, gas and bone intensity distributions can be learned for each region. Then by the independence assumption, PCA can be applied to the QFs and to the gas and bone intensity counts from each region, in a region by region manner, to learn $p\left(\underline{I}|\underline{\underline{m}}\right)$ (2.107)

RIQFs have been trained relative to m-reps and have been used for m-rep-based segmentation. The m-rep spoke-ends give a coarse sampling of points on the object boundary, and from Section 2.4 we know that the spoke-vectors are normal to the object boundary at the spoke-ends. Point-normal interpolation of the spoke-ends [84] has been used to define a dense point-normal field on the boundary from which RIQFs could be sampled. The medial atoms themselves could be interpolated [37] as another means of generating this dense sampling of point-normals on the object boundary.

The boundary of an m-rep can be specified with sub-voxel precision, but the training m-rep only approximates the boundary of the voxel-precise reference segmentation to which it was fit. Because of this approximation, a voxel near the boundary of the m-rep may lie on one side (interior or exterior) of the m-rep, but actually be on the opposite side of the boundary of the reference segmentation. The reference segmentation can be used as a mask to prevent such a voxel from contributing to the wrong RIQF. This procedure leads to a more robust estimation of the intensity distribution in each region, but requires a correction when these RIQFs are used to segment a new image, for which a reference segmentation is unavailable.

M-rep-relative RIQFs have been trained at two scales: at the scale of an entire objects and at a smaller scale for a neighborhood near a spoke end. The object-scale RIQFs define two global regions: one for the interior of the object and one for the exterior. In this setting the choice of anchor point is arbitrary, and the weighting factors are set so that all voxels on the boundary contribute equally to the QF but the weight falls off for voxels further away from the boundary. The RIQFs at smaller scale define an interior and an exterior region per m-rep spoke-end. The spoke end is designated as the anchor point of both of those regions, and voxels along a dense collection of point-normals near that spoke end make a weighted contribution to the RIQF. Here the correspondences established by the m-rep object-relative

coordinate system are leveraged to construct corresponding regions across the population of images. Many objects can be characterized by a homogeneous image intensity pattern in the interior of the object and a mixture of intensities on the exterior. The image intensity distributions around such an object can be modeled by RIQFs for a single object-scale interior region and multiple smaller scale regions on the exterior of the object.

Segmentation via m-reps

Segmentation via m-reps is posed as an energy minimization problem. The segmentation $\underline{\underline{m}}_L$ is found via the optimization

$$\underline{\underline{m}}_L = \arg\min_{\underline{\underline{m}}} \left(\alpha \sum_{i=1}^{k} \left(d_{\text{mahal}} \left(\underline{I}^i | \underline{\underline{m}} \right)^2 \right) + \beta d_{\text{mahal}} \left(\underline{\underline{m}} \right)^2 \right), \tag{2.108}$$

where each $\underline{I}^i$ term refers to an image region relative to the candidate object $\underline{\underline{m}}$ and α and β are non-negative weights that control the relative influence of the image-relative and purely geometrical terms of the cost function.

An image likelihood distribution $p\left(\underline{I}|\underline{\underline{m}}\right)$ is learned via PCA on RIQFs, and a shape prior $p\left(\underline{\underline{m}}\right)$ is learned via PGA on m-reps. Because the intensity distribution in each region is assumed to be independent of the intensities in the other regions and because $p\left(\underline{I}|\underline{\underline{m}}\right)$ and $p\left(\underline{\underline{m}}\right)$ are both assumed to be Gaussian distributions, (2.108) can be understood as

$$\underline{\underline{m}}_L = \arg\min_{\underline{\underline{m}}} \left(-\alpha \sum_{i=1}^{k} \log \left(p \left(\underline{I}^i | \underline{\underline{m}} \right) \right) - \beta \log \left(p \left(\underline{\underline{m}} \right) \right) \right) \tag{2.109}$$

$$= \arg\min_{\underline{\underline{m}}} \left(-\alpha \log \left(p \left(\underline{I}|\underline{\underline{m}} \right) \right) - \beta \log \left(p \left(\underline{\underline{m}} \right) \right) \right) \tag{2.110}$$

Each candidate m-rep $\underline{\underline{m}}$ is produced by first generating a vector of PGA coefficients $\underline{\theta}$. Let $\underline{\underline{\mu}} \in \mathcal{M}(n)$ denote the Fréchet mean m-rep computed during PGA training. The vector $\underline{\theta}$ is used in (2.44) to generate $\underline{\underline{\omega}} \in T_{\underline{\underline{\mu}}}\mathcal{M}(n)$. The candidate m-rep $\underline{\underline{m}}$ is defined to be $\text{Exp}_{\underline{\underline{\mu}}}\left(\underline{\underline{\omega}}\right)$.

This method for constructing the candidate m-rep offers the following benefits. The vector of PGA coefficients has fewer parameters than the m-rep itself, and thus this method allows a more efficient optimization than if the m-rep parameters were free to change independently.

Moreover, the PGA shape space constrains the optimization to only consider object instances that have a credible shape given the training population.

The $d_{\text{mahal}}\left(\underline{\underline{m}}\right)^2$ term of (2.108) is calculated as the sum of the square coefficients $\underline{\theta}$ (2.45). Each of the $d_{\text{mahal}}\left(\underline{I}^i|\underline{\underline{m}}\right)$ terms is calculated by sampling the RIQFs in the region $\underline{I}^i$ relative to $\underline{\underline{m}}$ and then applying (2.43) in the PCA space of the training RIQFs from $\underline{I}^i$. Thus, the objective function can be understood as a function of $\underline{\theta}$ and (2.108) can be minimized via standard numerical optimization techniques in this linear parameter space.

Each Mahalanobis distance in (2.108) can be understood as a random variable drawn from a χ^2 distribution that has as many degrees of freedom as there were eigenmodes retained by the PCA or PGA at training time. Strictly speaking, the $d_{\text{mahal}}\left(\underline{I}^i|\underline{\underline{m}}\right)^2$ terms may have additional degrees of freedom related to the gas and bone intensity modeling discussed in the previous section.

Well known properties of χ^2 distributions are that if $X \sim \chi^2_f$ and $Y \sim \chi^2_g$ are independent χ^2 random variables with f and g degrees of freedom respectively, $X + Y \sim \chi^2_{f+g}$, and the variances of X and Y are $2f$ and $2g$ respectively. With respect to (2.108), the image match term $\sum_{i=1}^{k}\left(d_{\text{mahal}}\left(\underline{I}^i|\underline{\underline{m}}\right)^2\right)$ can be understood as a χ^2 random variable. This image match is typically drawn from a distribution with greater degrees of freedom and hence greater variance than the distribution from which $d_{\text{mahal}}\left(\underline{\underline{m}}\right)^2$ is drawn.

If the weights α and β in (2.108) were set to be equal, m-rep-based segmentation could be understood as a MAP (2.7) problem. However, because in practice there is so much more variance in the image match terms than in the shape-prior terms, noise in the image match function could offset a steep increase in $d_{\text{mahal}}\left(\underline{\underline{m}}\right)^2$ and justify a nonsensical minimum of the objective function. "Pablo", the reference implementation of software used for m-rep-based segmentation, assigns values to the weights α and β that normalize the two χ^2 random variables corresponding to the image likelihood and the shape prior to each have unit variance. This prevents the domination of the objective function by the the image-match term.

Segmentation via m-reps is attractive for several reasons. It utilizes a compact shape representation that allows the robust estimation of $p\left(\underline{\underline{m}}\right)$ from relatively few samples, *i.e.,* in the HDLSS setting. It also produces an object relative coordinate system that allows $p\left(\underline{I}|\underline{\underline{m}}\right)$

to be defined at a scale that offers the benefits of being more local than global regions, but coarser and likely more robust than using one voxel per region. The availability of $p\left(\underline{\underline{m}}\right)$ and $p\left(\underline{I}|\underline{\underline{m}}\right)$ allows segmentation to be posed in a principled manner (2.108).

M-rep-based segmentation frequently produces high quality results. However, the image match terms of (2.108) are typically non-convex for real images, and such non-convexity produces a risk that the optimization can be trapped in a local minimum that does not correspond with a clinically usable segmentation. In Chapter 3, I present an initialization strategy that has been shown in practice to enable an m-rep-based method for bladder segmentation to begin its optimization suitably close to the global minimum of (2.108) and thus to produce a clinically useful segmentation more frequently than the previously used naive alternative. In Chapter 4, I present a novel method based on PCA of RIQFs that evaluates the credibility of a segmentation and has been shown to be highly discriminative when applied to segmentations of the bladder and prostate from CT.

2.5.4 Atlas-Based Segmentation

Thus far, I have discussed the following object-based segmentation methods: level-set methods, ASM, and m-reps. These methods all assume that the geometry and image intensity patterns of a target structure can be modeled and then used during an optimization process to segment a novel image. Atlas-based segmentation methods can be applicable even when those assumptions do not hold.

Atlas-based segmentation methods assume the existence of a reference image, which I will call the atlas image and will denote by $\underline{I}^0 : \Omega^0 \rightarrow V$, and its segmentation, which I will call the atlas and will denote by $S^0 \subset \Omega^0$. Given a target image $\underline{I} : \Omega \rightarrow V$, a transformation $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ such that

$$f\left(\Omega^0\right) \approx \Omega \tag{2.111}$$

is found through a process known as registration. The atlas-based segmentation of $\underline{I}$ is $f\left(S^0\right)$.

An atlas-based segmentation method can be characterized by how the atlas is constructed and by how the atlas is registered to a target case. In this dissertation I will not address

the issue of atlas construction. However, because the shape interpolation method that I will present in Chapter 5 is relevant to the problem of registration, I will briefly discuss popular image registration methods. A registration method can be classified along two axes: the class of functions $f(\cdot)$ that the method can produce and what is meant by $\approx$ in (2.111).

A rigid registration method is constrained so that f produces a rigid transformation of $\mathbb{R}^3$. For any pair of points $\left\{\underline{p}_1, \underline{p}_2\right\} \subset \mathbb{R}^3$, a rigid transformation f will satisfy the following equation.

$$\left\| f\left(\underline{p}_1\right) - f\left(\underline{p}_2\right) \right\| = \left\| \underline{p}_1 - \underline{p}_2 \right\| \tag{2.112}$$

Rigid registrations are appropriate when the poses of the target object and atlas may be different, but their shapes are essentially the same. For example, in Chapter 3, I will discuss an application of rigidly registered CT images of the pelvis of one subject who was scanned on multiple days. Because one could reasonably model the subject's pelvic bones during the time period in which the images were acquired as rigid bodies, a rigid registration of these images can be used to normalize the pose of the patient in the images.

Non-rigid registration is needed when the target object is allowed to have a different shape than the atlas. Non-rigid registration is commonly needed when the atlas is not personalized to the target patient, or when the segmented structures in the atlas are prone to deformation. Affine registration methods allow the rigidity assumption (2.112) to be falsified, but constrain f to be an affine transformation. That is, f preserves collinearity, and for each $\underline{p} \in \mathbb{R}^3$, $f\left(\underline{p}\right)$ can be understood as

$$f\left(\underline{p}\right) = \mathbf{A}\underline{p} + \underline{v}\,, \tag{2.113}$$

where $\mathbf{A}$ is a constant 3×3 matrix and $\underline{v}$ is a constant vector in $\mathbb{R}^3$.

Affine registration tends to be computationally tractable since an affine function f can be described by just 12 parameters. There are applications where an affine registration can adequately map the atlas onto the target image. However, in some applications a more general registration function is needed.

Of particular interest are registration methods that that relax the affinity assumption (2.113) but can guarantee that f will be a diffeomorphism [16]. That is, f will be a continuous,

invertible function that has a continuous inverse. The diffeomorphic property is desirable as it prevents the formation of folds and tears in $f\left(\Omega^0\right)$. Diffeomorphic registration can be computationally expensive, as f is typically understood as a time-series of velocity fields whose composition transform Ω^0 into Ω . This formulation allows individual points in the Ω^0 to flow along curved paths to their destinations in Ω in a manner than prevents the formation of folds of the space. However, the manipulation of these velocities field tends to be memory-intensive.

Spline-based methods, such as those based on thin-plate splines [33, 6] and those based on free form deformation of a grid of b-spline control points [72] are also popular forms of registration. These methods allow for local deformations that falsify the affinity assumption (2.113). In their classical forms these methods do not guarantee diffeomorphic results, but their implementation and memory requirements can be simpler that the diffeomorphic registration methods.

Figure 2.10 illustrates registration of a pair of toy data sets via rigid, affine, and thin-plate spline methods.

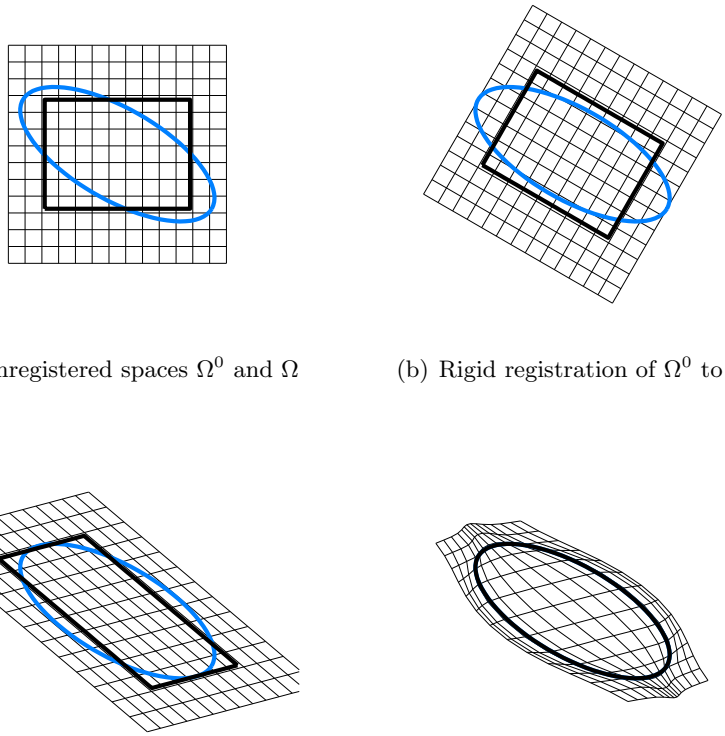
Registration is typically posed as an energy minimization problem. The approximation (2.111) can be understood to mean an energy functional $C\left(\right)$ has been minimized by the registration function f. The energy functional $C\left(\cdot\right)$ may be defined by corresponding landmarks $\left\{\underline{p}_i^0\right\} \in \Omega^0$ and $\left\{\underline{p}_i \in \Omega\right\}$, where the index i implies the correspondence.

Because rigid and affine functions have such small degrees of freedom and because they are frequently used to register landmark sets that are larger than the number of free parameters, rigid and affine registrations of landmark data can be performed by minimizing the following functional.

$$C\left(f\right) = \sum_i \left\| f\left(\underline{p}_i^0\right) - \underline{p}_i \right\|^2 \tag{2.114}$$

Registration methods that allow local deformations typically have sufficiently large degrees of freedom to exactly match any set of corresponding landmarks, *i.e.,* to minimize (2.114) to 0. Consequently, such registrations are performed via minimization of functionals that are more complicated functional than (2.114).

For example, thin-plate splines minimize an energy functional that is defined based on cor-



(a) Unregistered spaces Ω^0 and Ω

(b) Rigid registration of Ω^0 to Ω

(c) Affine registration of Ω^0 to Ω

(d) Thin-plate spline registration of Ω^0 to Ω

Figure 2.10: Examples of the different classes of registration methods. The rectangle (dark line) is being registered to the ellipse (bright line). The effect of the registration function f on a uniformly spaced grid in the coordinate sytem of the original rectangle is shown (thin line).

responding landmark sets. The thin-plate spline will provide an exact match at the landmark points, and is defined elsewhere in Ω^0 to minimize the bending energy of a set virtual metal plates. The bending of the plates defines a set of height functions. In R^3 these functions can be denoted by $\{h_x, h_y, h_z\}$ and the registration function f can be understood as

$$f\left(\underline{p}^0\right) = \underline{p}^0 + \left(h_x\left(\underline{p}^0\right), h_y\left(\underline{p}^0\right), h_z\left(\underline{p}^0\right)\right) \tag{2.115}$$

Corresponding landmarks were used by Joshi *et al.* [41] to drive a diffeomorphic registration. In this method, the volumes Ω^0 and Ω are modeled as fluids, and the registration

finds an optimal fluid flow to match the landmarks in Ω^0 to the landmarks in Ω . As was the case with thin-plate splines, there are too many degrees of freedom for minimization of (2.114) alone to be meaningful. Instead the cost functional that is minimized is of the form

$$C(\cdot) = C_{\text{data}}(\cdot) + \lambda C_{\text{regularity}}(\cdot) , \tag{2.116}$$

where the data term $C_{\text{data}}(f)$ is of the form (2.114), $C_{\text{regularity}}(f)$ regularizes the registration function f based on a metric on diffeomorphisms themselves, and λ is a weighting function.

The placement of these corresponding landmarks is equivalent to the challenging problem of producing PDMs that was discussed in Section 2.5.2. Vaillant and Glaunes [87] use the concept of currents from differential geometry to define a metric on shapes without the need for explicit correspondences. Glaunes and Joshi [30] define an energy functional that can be minimized to produce a shape-driven diffeomorphic warp that is of the form (2.116). The data term in their energy functional is based on the metric on shapes. Their regularization term is based on the metric on diffeomorphisms.

When a landmark-driven or shape-driven registration maps an atlas onto a target image, it provides a semi-automatic segmentation. Atlas-based segmentations can fit into a fully automatic framework by using image intensity information rather than landmark positions to drive the registration. In the classic work on demons [85] image based registration is driven by control points which diffuse either along an optical flow [39] direction or along the direction normal to an iso-intensity contour in the image. More generally, image-driven registration can be understood as an energy minimization problem where the energy functional is of the form (2.116). I will discuss common forms for the data term $C_{\text{data}}(\cdot)$ below. The regularization term $C_{\text{regularity}}(\cdot)$ typically depends on the type of registration function f for which the cost functional is being minimized.

A popular data term used during image-driven registration is the sum of squares of intensity differences.

$$C_{\text{data}}(f) = \sum_{\underline{p} \in \Omega^0} \left\| \underline{I}^0\left(\underline{p}\right) - \underline{I}\left(f\left(\underline{p}\right)\right) \right\|^2 \tag{2.117}$$

This data term only makes sense if the intensities of $\underline{I}^0$ and $\underline{I}$ are commensurate. For example, Foskey *et al.* [27] have used it in the context of CT images of the male pelvis for a fluid-flow based diffeomorphic registration of a patient-specific atlas image to additional CT images of that patient. CT images have normalized intensities, so the assumption of commensurate intensities can be reasonable even if the images are of different patients or were acquired from different scanners.

MR images do not have normalized units, so a data term based on the probability distributions of $\underline{I}^0\left(\underline{p}\right)$ and $\underline{I}\left(f\left(\underline{p}\right)\right)$ is typically more appropriate than (2.117). Chung *et al.* [17] discuss three popular data terms that fit into this category. These functionals are based on MLE, mutual information, and the Kullback-Leibler distance respectively and are applicable not only to uncalibrated MR image registration, but also to the more general problem of multi-modal image registration. Zollei *et al.* [94] show how those three data terms can be understood in a common framework for multi-modal registration.

Advantages of atlas-based segmentation methods are that they can be fully automatic, a population of segmentations can be understood through correspondences in the coordinate system of the atlas, and when a sufficiently general registration method is used it can locally describe fine scale changes to a target object. Disadvantages of atlas-based segmentation methods are the memory intensive nature of the more general registration methods and the corresponding difficulty in learning a shape prior from a set of general registration functions. Arsigny *et al.* have proposed a log-Euclidean framework [2] that allows PCA to be applied to a subclass of diffeomorphic functions, but the statistical analysis of general registration functions is still challenging because not all registrations can be modeled in that framework and because PCA is not robust to small training sample sizes given the high dimensional representation of each registration function.

In Chapters 3, 4, and 5 I will describe experiments performed on m-rep-based segmentation. The principled manner for learning a shape prior via PGA and the object-relative coordinate system provided by the m-rep framework make it well suited for my studies on initialization from non-intensity data (Chapter 3) and identifying non-credible regions in a segmentation (Chapter 4). My interpolation scheme (Chapter 5) simultaneously exploits posi-

tion and orientation information, precisely the data that is captured by medial atoms. There are, however, interesting connections between my method for shape interpolation and the landmark-based registration schemes I have discussed above. The configuration of landmarks during intermediate timepoints of a fluid-flow registration are themselves a shape interpolation. Moreover, samples from interpolated shapes produced by my method can be understood as landmarks to drive a registration process, for example to extrapolate the segmentation of a medially-described object to the entire image volume.

2.6 Evaluation of Segmentation

In Section 2.5, I discussed several object- and atlas-based segmentation methods. Now, I will assume the existence of a medical image $\underline{I} \subset \Omega$ and its segmentation $S \subset \Omega$, and I will address the question of how to evaluate S.

In clinical practice this evaluation can be understood by a simple binary decision: did the segmentation provide the end user with an acceptable representation of the subject's anatomy. Since the cost of failing this test is the potential mistreatment of a patient, a segmentation method developed in a research setting should not be put to clinical use until it has been demonstrated to be effective. To evaluate segmentations one needs quantitative measures to use as proxies for the clinical relevance of a segmentation.

The quantitative methods for segmentation evaluation that I will discuss assume that ground truth is known. That is, the segmentation S is intended to identify voxels that are known *a priori* to truly form the set $S^{\text{truth}} \subset \Omega$. In research settings, such as a phantom study, where the geometry of the imaging target is well understood, or a cadaver-study, where virtual slices through an image of the subject can be correlated with histological slices of the subject itself, S^{truth} can be known. When S^{truth} cannot be known, it may be approximated by a reference segmentation produced by a method other than the one that is being tested. The STAPLE [89] algorithm can be used to combine multiple segmentations of the same image into an estimate of ground truth.

When S^{truth} is known, it is often referred to as the "gold standard". Estimates of truth

produced from STAPLE have been dubbed the "bronze standard". A single manual segmentation typically receives a less flattering moniker such as the "wood standard" or the "dirt standard". The evaluations that I report in the latter chapters of this dissertation are relative to the wood standard estimate of S^{truth}.

These evaluations are typically posed as distances between ∂S: the boundary of S and $\partial S^{\text{truth}}$: the boundary of S^{truth} or as overlap between the volumes S and S^{truth}.

The symmetric Hausdorff distance defined below (2.121) is a commonly used measure of surface distance between S and S^{truth}. For any $s \in \partial S$ let

$$d\left(s, S^{\text{truth}}\right) = \min_{s' \in \partial S^{\text{truth}}} \left\| s - s' \right\| . \tag{2.118}$$

For any $s' \in \partial S^{\text{truth}}$ let $d\left(s', S\right)$ be defined similarly. The Hausdorff distances from S to S^{truth} and from S^{truth} to S are defined below.

$$d_{\text{H}}\left(S, S^{\text{truth}}\right) = \max_{s \in \partial S} d\left(s, S^{\text{truth}}\right) \tag{2.119}$$

$$d_{\text{H}}\left(S^{\text{truth}}, S\right) = \max_{s' \in \partial S^{\text{truth}}} d\left(s', S\right) \tag{2.120}$$

The symmetric Hausdorff distance (frequently incorrectly abbreviated to Hausdorff distance) is defined as follows.

$$d_{\mathbf{H}}\left(S, S^{\text{truth}}\right) = \max\left(d_{\text{H}}\left(S, S^{\text{truth}}\right), d_{\text{H}}\left(S^{\text{truth}}, S\right)\right) \tag{2.121}$$

The fact that the symmetric Hausdorff distance is the maximum distance from any point on either ∂S or $\partial S^{\text{truth}}$ to the nearest point on the other boundary is both an advantage and a disadvantage. When $d_{\mathbf{H}}\left(S, S^{\text{truth}}\right)$ is small, S is likely to provide a good estimate of S^{truth} everywhere. However, $d_{\mathbf{H}}\left(S, S^{\text{truth}}\right)$ is not robust to outliers. A large symmetric Hausdorff may indicate a segmentation that is consistently bad, or it may indicate a segmentation that is wrong only in a very small neighborhood. When $d_{\mathbf{H}}\left(S, S^{\text{truth}}\right)$ is the only reported measurement, there is no way to distinguish between those two cases.

Other commonly reported surface distances are more robust to outliers, but this robustness

comes at the cost of perhaps reporting a small distance for a case with significant errors in a neighborhood. The average surface distance:

$$d_{\text{mean}}\left(S, S^{\text{truth}}\right) = \frac{\text{mean}_{s\in\partial S}\left(d\left(s, S^{\text{truth}}\right)\right) + \text{mean}_{s'\in\partial S^{\text{truth}}}\left(d\left(s', S\right)\right)}{2} \tag{2.122}$$

and the 90% worst distance defined below (2.125) are commonly used to evaluate a segmentation.

$$d_{0.90,\rightarrow}\left(S, S^{\text{truth}}\right) = \inf_k\left(\frac{\left\|\left\{s : s\in\partial S, d\left(s, S^{\text{truth}}\right)\leq k\right\}\right\|}{\|\{s : s\in\partial S\}\|} \geq 0.90\right) \tag{2.123}$$

$$d_{0.90,\leftarrow}\left(S, S^{\text{truth}}\right) = \inf_k\left(\frac{\left\|\left\{s' : s'\in\partial S^{\text{truth}}, d\left(s', S\right)\leq k\right\}\right\|}{\left\|\left\{s' : s'\in\partial S^{\text{truth}}\right\}\right\|} \geq 0.90\right) \tag{2.124}$$

$$d_{0.90}\left(S, S^{\text{truth}}\right) = \frac{d_{0.90\rightarrow}\left(S, S^{\text{truth}}\right) + d_{0.90\leftarrow}\left(S, S^{\text{truth}}\right)}{2} \tag{2.125}$$

Note that (2.123 - 2.125) can be generalized to any threshold level. The 90^{th} percentile is interesting because it provides a robust estimate of worst case performance.

Segmentation performance is also frequently reported with a measurements of the volume overlap between S and S^{truth}. Such a measure reports a score ranging from 0 for no overlap to 1 for complete agreement. Because S^{truth} is known to be correct, it may be intuitive to define volume overlap asymmetrically:

$$\text{Overlap}\left(S, S^{\text{truth}}\right) = \frac{\left\|S\cap S^{\text{truth}}\right\|}{\left\|S^{\text{truth}}\right\|}. \tag{2.126}$$

This measure is flawed as a sufficiently large S will appear to be in perfect agreement with S^{truth}.

Symmetric measures of volume overlap are used to avoid this problem. Two commonly reported symmetric measures of volume overlap are the Jaccard score (2.127) [40] and the Dice similarity coefficient (2.128) [25].

$$\mathrm{JAC}\left(S, S^{\mathrm{truth}}\right) = \frac{\left\|S \cap S^{\mathrm{truth}}\right\|}{\left\|S \cup S^{\mathrm{truth}}\right\|} \tag{2.127}$$

$$\begin{aligned}\mathrm{DSC}\left(S, S^{\mathrm{truth}}\right) &= \frac{\left\|S \cap S^{\mathrm{truth}}\right\|}{\frac{1}{2}\left(\|S\| + \left\|S^{\mathrm{truth}}\right\|\right)} \\ &= \frac{2\left\|S \cap S^{\mathrm{truth}}\right\|}{\|S\| + \left\|S^{\mathrm{truth}}\right\|}\end{aligned} \tag{2.128}$$

These measures were originally designed to compare sets of biological traits in different species. When the measures are applied to labels in discrete images, there are two possible interpretations. The following interpretation allows for an arbitrarily large overestimation of the agreement between S and S^{truth}. Let $\left\|S \cap S^{\mathrm{truth}}\right\|$ be understood as the number of voxels which have the same label according to the segmentation and the ground truth and let $\|S\| = \left\|S^{\mathrm{truth}}\right\| = \left\|S \cup S^{\mathrm{truth}}\right\| = \|\Omega\|$. In this scenario the Jaccard score and Dice coefficient are equivalent, and the overlap measurement can be artificially inflated by padding Ω with additional voxels that S and S^{truth} will agree are not part of the structure of interest. For this reason, it is better to interpret $\left\|S \cap S^{\mathrm{truth}}\right\|$ as a count of voxels that S and S^{truth} agree are in the structure of interest and to interpret $\left\|S \cup S^{\mathrm{truth}}\right\|$ as a count of voxels that at least one of S and S^{truth} believe are in the structure of interest.

The Jaccard score and Dice coefficient can be expressed as functions of each other, so there is no need to report both measures.

$$\begin{aligned}
\mathrm{DSC}\left(S, S^{\mathrm{truth}}\right) &= \frac{2\left\|S \cap S^{\mathrm{truth}}\right\|}{\|S\| + \left\|S^{\mathrm{truth}}\right\|} && (2.129)\\
&= \frac{2\left\|S \cap S^{\mathrm{truth}}\right\|}{\left\|S \cup S^{\mathrm{truth}}\right\| + \left\|S \cap S^{\mathrm{truth}}\right\|} && (2.130)\\
&= \frac{2\left\|S \cap S^{\mathrm{truth}}\right\|}{\left\|S \cup S^{\mathrm{truth}}\right\|} \cdot \frac{\left\|S \cup S^{\mathrm{truth}}\right\|}{\left\|S \cup S^{\mathrm{truth}}\right\| + \left\|S \cap S^{\mathrm{truth}}\right\|} && (2.131)\\
&= 2\frac{\mathrm{JAC}\left(S, S^{\mathrm{truth}}\right)}{1 + \mathrm{JAC}\left(S, S^{\mathrm{truth}}\right)} && (2.132)
\end{aligned}$$

The work by Crum *et al.*[21] provides an interesting perspective on the measures of segmentation quality. It generalizes definitions of volume overlap and surface distance so that they can be applied to "soft" segmentations where each voxel is assigned a probability of belonging to the target structure. That work also presents measures of surface distance in a common framework with measures of volume overlap, by defining distance as the minimal amount of dilation of the volume needed to achieve a desired overlap.

The m-rep-based segmentations I will discuss in Chapter 3 are "hard". Each voxel is classified as strictly inside or strictly outside the target object, based on its relationship with the implied boundary of the m-rep. I will report $d_{\mathrm{mean}}\left(S, S^{\mathrm{truth}}\right)$, $d_{0.90}\left(S, S^{\mathrm{truth}}\right)$, and $\mathrm{DSC}\left(S, S^{\mathrm{truth}}\right)$ for these segmentations relative to a wood-standard estimate of ground truth.

Chapter 3

Semiautomatic Segmentation Initialization via a Sparse Set of Contours

Statistical deformable shape models (SDSM) used for object-based segmentation are sensitive to the alignment of the training exemplars. When the training models have been inadequately aligned, the SDSM will be dominated by global pose changes. Object-based segmentation from such an SDSM is likely to get trapped in a minimum of the cost function that produces an unsatisfactory result because the shape space is not rich enough to describe the target object and the geometric probability distribution is too broad. When the training models have been tightly aligned, *i.e.,* via Procrustes alignment or a generalization thereof, the resulting shape space and probability distribution tend to have greater generalization ability and specificity and are better able to describe the shape of the target object.

There are two challenges in using such an SDSM to segment a target image. The first challenge is that a transformation that aligns the SDSM to the (unknown) target object is needed because the variability in the SDSM itself will not be able to correct an erroneous pose estimate. The second challenge arises because image match functions are typically non-convex. When the segmentation is initialized with an accurate estimate of the pose of the target object but no specific knowledge of the shape of the target object, it may produce an unacceptable result in a local minimum of the image match function. Thus, the second challenge is to estimate the shape of the target object prior to its segmentation.

In this chapter, I will introduce a function that measures the agreement between a set of

manually speficied points on the boundary of an object and points on the boundary of an SDSM instance. This cost function does not require *a priori* knowledge of exact correspondences between the two sets of points.

This function can be minimized over the space of similarity transformations, in order to align a population of training objects. Alternating minimization of this function over the space of similarity transformations and over the space of deformations allowed by the SDSM allows the segmentation of a target image to be initialized by an object that estimates both the pose and the shape of the target object. Such a segmentation is less likely to be trapped in a local minimum of the image match function than is a segmentation that is initialized without an estimate of the shape of the target object.

I will demonstrate the initialization of bladder segmentation by applying this alternating method to two m-rep-based SDSMs. The first SDSM was learned from training models that were aligned at the scale of the entire pelvic region. The second SDSM was learned from training models that were aligned at the scale of the bladder by similarity transformations that minimize the same cost function that is minimized via the alternating method to initialize the segmentation of a target image.

The segmentation initializations produced by applying the alternating method to these SDSMs were quantitatively better segmentations than the previous method, which was initialized by estimating only the pose of the target object, based on a rigid alignment of the target image at the scale of the entire pelvic region. Frequently, the initializations that my method produced were competitive with the final segmentation result from the previous method and with other segmentation results reported in the literature.

This chapter is organized as follows. I describe the driving medical problem of segmenting the bladder for adaptive radiotherapy treatment planning in Section 3.1. I specify the cost function that is minimized by my methods for alignment of training bladders and semiautomatic initialization of bladder segmentation in Section 3.2. The two m-rep-based SDSMs that I have used with this method are described in Sections 3.3 and 3.4, with results of the initializations and subsequent segmentations given for the two SDSMs in Sections 3.3.1 and 3.4.1, respectively. This chapter concludes in Section 3.5 with discussions of how the user-

input burden of this method could be relieved, of opportunities and limitations of applying this method to other objects in the male pelvis, and of how this method could be extended to produce an interactive segmentation tool.

3.1 Adaptive Radiotherapy of the Male Pelvis

A cancer patient being treated with external beam radiotherapy (RT) will typically be the subject of a planning process, during which a customized course of RT will be defined. After this custom plan has been produced, the patient will undergo treatment.

RT treatment planning is an optimization problem. An optimal plan is one that maximizes the probability of controlling the tumor while minimizing the risk of damaging other sensitive tissues near the tumor. These probabilities are understood by what is known as dose-volume histogram (DVH) analysis. The treatment target and healthy structures near it that are at risk of being irradiated are segmented. Computer simulations of treatments are run, and for each of the segmented structures a DVH, a cumulative historgam of dose delivered to tissue volume, is produced. The prescribed treatment plan will be one that deposits sufficient dose into the segmented target region to control the tumor, while not exceeding a threshold for dose deposited in the segmented regions to be spared.

Because a serious risk of RT is damage caused by the irradiation of healthy tissue and because healthy tissue tends to recover from mild irradiation more quickly than tumor does, it is common for a patient to be treated by a fraction of the prescribed radiation dose on each of a number of treatment dates. Fractionated RT tends to effectively control the disease and presents less risk of permanently harming nearby healthy tissue than does a single massive dose. However, in conventional fractionated RT the same treatment plan is used on each treatment day.

Because the patient's soft tissues undergo various deformations between treatment fractions, it is desirable to periodically update the treatment plan in response to where the treatment doses have actually accumulated or in response to the patient's configuration at the precise moment when the next treatment will begin. In order to understand how dose has

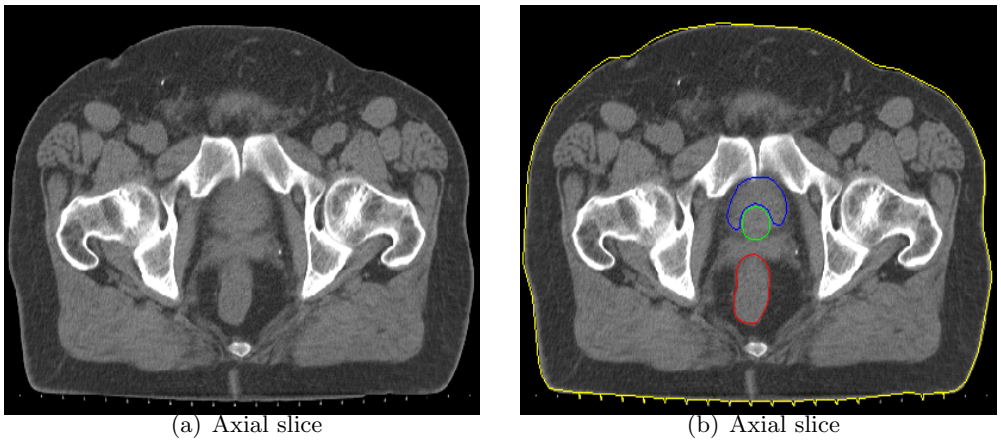
accumulated in the patient, segmented images from the previous treatment days are needed. Similarly, in order to update the treatment plan prior to a treatment, a segmented image from that treatment day is needed. These techniques for adjusting the treatment plan in response to segmented treatment-day images are known as adaptive radiotherapy (ART).

The experiments described is this chapter relate to ART treatment of prostate cancer. In these experiments, CT images of the patient's pelvic region are available from approximately 15 of the treatment days. The clinical protocol for the treatment of these patients called for the manual segmentation of the prostate, bladder, and rectum. The entire prostate gland is considered to be the target volume for this treatment, and the bladder and rectum are considered healthy tissues that need to be spared from radiation. My research is part of an ongoing project to develop automated methods to produce these segmentations.

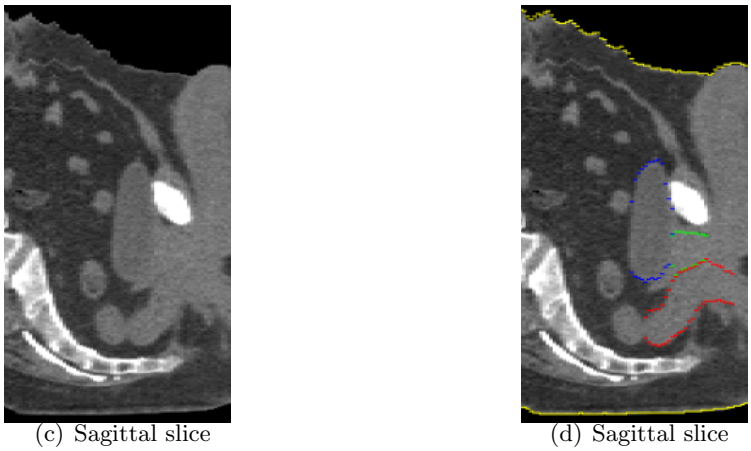
All of the images were acquired with the patient in the same position. The patient lies supine on the table as it passes through the CT scanner. An example image and its manual segmentation are shown in Figure 3.1. The coordinate system of the image can be understood as follows. Axial slices through the image are normal to the inferior/superior axis of the patient (Figure 3.1(a)). The z coordinate refers to axial slice number. Sagittal slices through the image are normal to the left/right axis of the patient (Figure 3.1(c)) and are indexed by the x coordinate. Coronal slices through the image are normal to the anterior/posterior axis of the patient (Figure 3.1(e)) and are indexed by the y coordinate.

The axial slices of each image have higher resolution than sagittal or coronal slices. Each voxel occupies an area of approximately 1mm $\times$ 1mm in the (x, y) plane and protrudes approximately 3mm in the axial direction. For several reasons, including this issue of resolution and a historical preference, the medical experts who contributed to this ART project prefer to produce manual segmentations by drawing contours relative to axial image slices. These experts primarily use sagittal and coronal slices of the image to aid in their understanding in 3D as they draw in the axial slices.

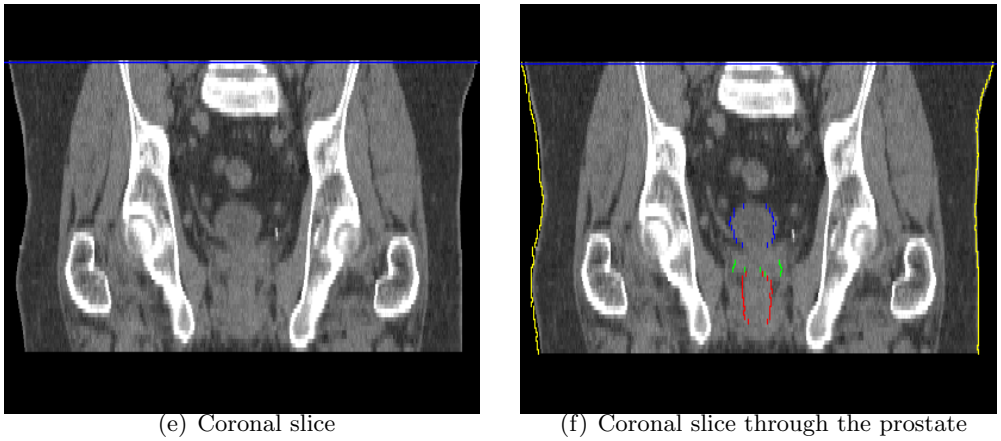
ART requires each of the treatment (-day) images to be segmented. It is challenging to construct an SDSM for use in object-based segmentation of treatment-day images that takes advantage of the knowledge that each treatment image is of the same patient. Training a



(a) Axial slice (b) Axial slice



(c) Sagittal slice (d) Sagittal slice



(e) Coronal slice (f) Coronal slice through the prostate

Figure 3.1: Slices through a CT image of the male pelvis are shown. Segmentations of the bladder (blue), prostate (green), rectum (red), and skin (yellow) are shown in Figures 3.1(b), 3.1(d) and 3.1(f).

robust SDSM with data from the current patient only is not clinically realistic because of an HDLSS problem. There is not enough training data available to produce an SDSM on an early

treatment day.

The SDSM described in Section 3.3 is trained from the target patient only. While it has been useful for benchmarking initialization and segmentation algorithms, it is not a clinically realistic model for segmenting early treatment-day images because it uses information from the later treatment days to produce the probability distributions used to segment the earlier treatment days. An SDSM produced by pooling a single training case from each of multiple other patients can avoid the HDLSS problem but will model between-patient variation rather than the within-patient variation relevant to segmenting ART treatment-day images.

The SDSM described in Section 3.4 is produced by pooling carefully aligned residues of other patients' treatment day objects from their patient-specific means. This creates a model of within-patient variation trained from multiple patients. The covariance structure from these residues is combined with a mean-to-date estimate of the object in the target patient, producing a robust patient-specific SDSM that describes within-patient variation even for early treatment days. It is not, however, appropriate for segmenting the first treatment image.

3.2 Objective Function for Semi-automatic Initialization Without Explicit Correspondences

An SDSM that is trained on aligned exemplars will typically have a tighter distribution (*i.e.,* less total variance in its eigenmodes of variation) and consequently will typically exhibit greater generalization ability and specificity than if it was trained on unaligned exemplars. At segmentation time though, these benefits can only be realized if the SDSM is adequately initialized into the target image. That is, if the unknown segmentation of the target image can be aligned to the SDSM and the shape of the SDSM is sufficiently close to the shape of the unknown segmentation.

In this section, I present an optimization that produces an SDSM instance that is well matched to manually specified points on a subset of the boundary of the unknown segmentation. This SDSM instance approximates the pose and the shape of the unknown segmentation and is a suitable initialization for SDSM-based segmentation.

The SDSMs that I will describe in Sections 3.3 and 3.4 were trained using PGA [26] on m-reps, using the procedure described in Section 2.5.3. Each SDSM is characterized by a Fréchet mean m-rep $\underline{\underline{\mu}} \in \mathcal{M}(n)$ and a set of eigenmodes of variation $\{V_j\} \subset T_{\underline{\underline{\mu}}}\mathcal{M}(n)$. To simplify notation, I will assume that each eigenvector V_j has been scaled so that its norm is $\sqrt{\lambda_j}$ (the square root of the corresponding eigenvalue). A new shape instance in the space in which the aligned training models lie can be generated by the formula

$$\underline{\underline{m}}^{\text{aligned}} = \text{Exp}_{\underline{\underline{\mu}}}\left(\sum_j \theta_j V_j\right), \tag{3.1}$$

where θ_j is a tuple of PGA coefficient in units of standard deviation. Given a similarity transformation $\mathcal{T}$ that maps the alignment space into the target image space, a new shape instance in the target image space can be generated by the following formula:

$$\underline{\underline{m}} = \mathcal{T}\left[\text{Exp}_{\underline{\underline{\mu}}}\left(\sum_j \theta_j V_j\right)\right]. \tag{3.2}$$

The initialization method that I have developed minimizes a cost function $C(\cdot) : \mathcal{M}(n) \rightarrow [0, \infty)$ that measures the disagreement between a bladder instance of the form (3.2) and manually selected points on designated axial slices of the target image. An m-rep that minimizes this cost approximates the segmentation of the target image. This m-rep is a reasonable initialization for the m-rep-based segmentation of the image.

The cost function $C(\cdot)$ measures the distances from each of the manually specified points in the target image to the nearest point in a corresponding subset of the SDSM implied boundary. I call the subset of the SDSM boundary in which the nearest point to a manually specified point may live a "τ-band". Because the user does not need to explicitly model the correspondences between the points in the image and points on the SDSM boundary, the τ-bands produce loose correspondences. Minimizing $C(\cdot)$, and thus the distances from the manually selected points to the corresponding τ-bands on the SDSM boundary, causes the SDSM to approximate the size, shape, and pose of the unknown segmentation. The method of establishing these correspondences via τ-bands and the cost function $C(\cdot)$ are defined below.

This initialization differs from Procrustes-based methods in the following ways. This method does not require explicit correspondences between the manually specified bladder points and samples from the boundary of bladder model. Specifying the bladder points used during this initialization is a simple task for the user and does not require a complex understanding of the shape in 3D, while specifying landmarks for a Procrustes-based alignment is a difficult and time-consuming task. One reason why this manual task is so simple is that the points used by this initialization come from a relatively small subset of the boundary that lies on just three axial image slices. Because this set of points does not cover the full boundary of the target object, there cannot be a 1-1 and onto mapping from the (densely sampled) boundary of the SDSM instance to the manually specified points. Thus a Procrustes alignment of the full SDSM boundary with the manually specified points is not possible. Although it is implemented differently, this aspect of my method is in the same spirit as Rangarajan's "softassign" Procrustes method [70], which given two sets of landmarks points with unknown correspondences splits each set into inliers and outliers and aligns the two inlier sets. Another difference is that initialization methods based on Procrustes alignment only estimate the pose of the object, but the method presented here estimates both the pose and the shape of the bladder in the target image.

The following notation will aid in the presentation of $C(\cdot)$. Let $x : \mathbb{R}^3 \to \mathbb{R}$, $y : \mathbb{R}^3 \to \mathbb{R}$, and $z : \mathbb{R}^3 \to \mathbb{R}$ be functions that access the coordinates of a point in $\mathbb{R}^3$.

$$x(\alpha, \cdot, \cdot) = \alpha \tag{3.3}$$

$$y(\cdot, \alpha, \cdot) = \alpha \tag{3.4}$$

$$z(\cdot, \cdot, \alpha) = \alpha \tag{3.5}$$

Let $\partial \underline{\underline{m}} \subset \mathbb{R}^3$ denote a set of points sampled from the boundary of the m-rep instance $\underline{\underline{m}}$. Let the extreme z (axial) coordinates of $\underline{\underline{m}}$ be known as $z^0\left(\underline{\underline{m}}\right)$ and $z^1\left(\underline{\underline{m}}\right)$ defined below.

$$z^0\left(\underline{\underline{m}}\right) = \min_{\underline{x} \in \partial \underline{\underline{m}}} z\left(\underline{x}\right) \tag{3.6}$$

$$z^1\left(\underline{\underline{m}}\right) = \max_{\underline{x} \in \partial \underline{\underline{m}}} z\left(\underline{x}\right) \tag{3.7}$$

Let $\mathcal{P} \subset \mathbb{R}^3$ denote a set of manually specified points in the target image that the user believes to lie on the boundary of the bladder. The user is required to specify the extreme axial coordinates of the bladder in the target image. The minimum and maximum axial coordinates are denoted by $z^0_{\mathcal{P}}$ and $z^1_{\mathcal{P}}$, respectively.

Let τ be a user specified threshold that is used to establish the loose correspondence between $\mathcal{P}$ and $\underline{\underline{m}}$. For fixed $\mathcal{P}$, $z^0_{\mathcal{P}}$, and $z^1_{\mathcal{P}}$ as defined above, I say that points $\underline{p} \in \mathcal{P}$ and $\underline{x} \in \partial\underline{\underline{m}}$ lie in the same "τ-band" if the z-coordinate of $\underline{p}$ relative to the known axial span from $z^0_{\mathcal{P}}$ to $z^1_{\mathcal{P}}$ is within a specified tolerance τ of the z-coordinate of $\underline{x}$ relative to the axial span of $\underline{\underline{m}}$. I define the τ-band distance from $\underline{p}$ to $\underline{x}$ to be the Euclidean distance between them if the points are in loose correspondence by virtue of lying in the same τ-band and to be infinitely large if such a correspondence cannot be established.

$$d_\tau\left(\underline{p}, \underline{x}\right) = \begin{cases} \left\|\underline{p} - \underline{x}\right\| & , \left(\frac{z\left(\underline{p}\right) - z^0_{\mathcal{P}}}{z^1_{\mathcal{P}} - z^0_{\mathcal{P}}} - \frac{z\left(\underline{x}\right) - z^0\left(\underline{\underline{m}}\right)}{z^1\left(\underline{\underline{m}}\right) - z^0\left(\underline{\underline{m}}\right)}\right)^2 \leq \tau^2 \\ \infty & , \text{otherwise} \end{cases} \tag{3.8}$$

The τ-band distance from a point $\underline{p}$ to an m-rep $\underline{\underline{m}}$ is defined to be the minimum of the τ-band distance (3.8) from $\underline{p}$ to any point on the boundary of $\underline{\underline{m}}$.

$$d_\tau\left(\underline{p}, \underline{\underline{m}}\right) = \min_{\underline{x} \in \partial \underline{\underline{m}}} d_\tau\left(\underline{p}, \underline{x}\right) \tag{3.9}$$

I will refer to the point $\underline{x}$ that minimizes (3.9) as the "near-point" to $\underline{p}$. The near-point can be understood to be in correspondence with $\underline{p}$. I will also use the notation that $\underline{p}$ is matched by its near-point.

The τ-band distance between $\mathcal{P}$ and $\underline{\underline{m}}$ is defined to be the square root of the sum of

squares of (3.9) from each point in $\mathcal{P}$ to $\underline{\underline{m}}$.

$$d_\tau\left(\mathcal{P}, \underline{\underline{m}}\right) = \sqrt{\sum_{\underline{p} \in \mathcal{P}} d_\tau\left(\underline{p}, \underline{\underline{m}}\right)^2} \tag{3.10}$$

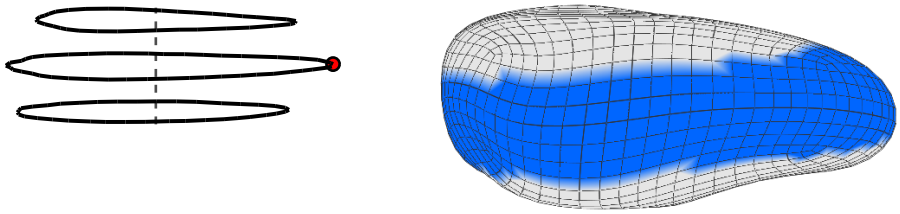
The τ-band distance from a fixed $\mathcal{P}$ (3.10) is the aforementioned cost function that measures the agreement of an m-rep instance to manually specified points. The m-rep used for initialization of the segmentation is known as $\underline{\underline{m}}^{\text{init}}$ and can be found via the following optimization.

$$\underline{\underline{m}}^{\text{init}} = \arg\min_{\underline{\underline{m}}} d_\tau\left(\mathcal{P}, \underline{\underline{m}}\right) \tag{3.11}$$

The remainder of this section is organized as follows. I will first discuss the parameter τ. Then I will discuss how $\mathcal{P}$ is specified and the space over which (3.11) is minimized.

The parameter τ creates a loose correspondence between $\mathcal{P}$ and $\underline{\underline{m}}$. As the user specifies points in $\mathcal{P}$, he is not required to perform the challenging task of precisely mapping a landmark in image coordinates into the object relative coordinate system for an unknown m-rep instance. Instead the τ-band distance insures that a point specified near the inferior end of the bladder in the image data will be matched by a point near the inferior end of the m-rep. Likewise, a point specified near the superior end of the bladder image will be matched by a point near the superior end of the m-rep. Figure 3.2 shows a manually specified point relative to $\mathcal{P}$, and the band of points on an m-rep boundary to which it could potentially be matched.

The value of the parameter τ should be in the range $0 \leq \tau \leq 1$. When $\tau = 1$, (3.8) simplifies to $d_\tau\left(\underline{p}, \underline{x}\right) = \left\|\underline{p} - \underline{x}\right\|$ and any point on the boundary of the m-rep can match to any manually specified point in (3.10). This is undesirable because the minimizer of (3.11) could be an m-rep that is much larger than the bladder in the image data and has a subregion that is well matched to all of $\mathcal{P}$. Indeed, that occurred in outlier cases during my preliminary experiments. When $\tau = 0$, (3.9) requires each point in $\mathcal{P}$ to match an m-rep boundary point at precisely the same relative position along the z-axis. This setting is undesirable because it requires an extremely dense sampling of $\partial\underline{\underline{m}}$ in (3.9) and consequently will make that optimization quite slow. Moreover, this setting is too restrictive. A non-trivial τ value allows more flexibility in



(a) Hand drawn contours of the bladder from three designated axial slices of the image (solid line). A point of interest is indicated in red. The axial extent of the bladder is indicated by the dashed line.

(b) The band of points (in blue) on the boundary of the m-rep that are within an acceptable threshold of having the same axial coordinate relative to the entire m-rep as the marked point in Figure 3.2(a) has relative to the manually specified axial extent of the bladder in image

Figure 3.2: A point specified in a single manual slice of the image (Figure 3.2(a)) and the band of points on the m-rep surface (Figure 3.2(b)) that are candidates to be the near-point to the manually specified point.

where the near-point for each manually specified point can lie. In practice I have found that when $\tau = 0.3$, (3.11) tends to produce a high quality initialization.

Note that (3.10) is an asymmetric function. The optimization (3.11) favors an m-rep such that each point in $\mathcal{P}$ is close to its near-point, but there is no penalty for points on the m-rep boundary that are far away from $\mathcal{P}$. This allows $\mathcal{P}$ to be specified incompletely. As long as there is enough information in $\mathcal{P}$ to describe the shape and pose of the bladder, the user does not need to exhaustively identify each point on the boundary of the bladder.

In practice I have found that defining $\mathcal{P}$ to be the vertices of piecewise linear contours on the boundary of the bladder on three designated axial image slices is sufficient. The three slices are defined accordingly. First the user is responsible for identifying the axial span of the bladder. Suppose the user determines that the bladder is visible on slices $\left\{z_{\mathcal{P}}^0..z_{\mathcal{P}}^1\right\}$. The three axial images slices that the user is asked to manually segment are $z_{\mathcal{P}}^0 + 1$, $z_{\mathcal{P}}^1 - 1$, and $\left\lfloor \left(z_{\mathcal{P}}^0 + z_{\mathcal{P}}^1\right) /2 \right\rfloor$.

Because $\mathcal{P}$ has been specified to describe the shape and pose of the target bladder, a bladder-shaped object that has the members of $\mathcal{P}$ on its boundary is likely to be a close

approximation of the target bladder. An m-rep $\underline{\underline{m}}$ with a low τ-band distance to $\mathcal{P}$, with $\tau = 0.3$, typically approximates the bladder in the target image. I find such a model by the following procedure which implements (3.11) by alternating between refining estimates of the pose and shape of the bladder m-rep.

Let $\underline{\underline{\mu}} \in \mathcal{M}(n)$ and $\{V_j\} \subset T_{\underline{\underline{\mu}}}\mathcal{M}(n)$ denote a Fréchet mean m-rep and eigenmodes of variation learned via PGA. I do not make any assumptions about how the training m-reps are aligned before PGA is applied. Let $\mathcal{T}$ denote a similarity transform and let $\underline{\theta}$ denote a set of PGA coefficients. No specific knowledge is needed to produce initial estimates of $\mathcal{T}$ and $\underline{\theta}$.

$$\mathcal{T}^{(0)} = \mathrm{Id}_3 \tag{3.12}$$

$$\underline{\theta}^{(0)} = \{0\,, ...,\, 0\} \tag{3.13}$$

The following iterative algorithm finds $\mathcal{T}$ and $\underline{\theta}$ that deform $\underline{\underline{\mu}}$ so that it matches the pose and shape of the bladder in the target image.

$$\mathcal{T}^{(k)} = \arg\min_{\mathcal{T}} d_\tau \left(\mathcal{P}, \mathcal{T} \left[\mathrm{Exp}_{\underline{\underline{\mu}}} \left(\sum_j \theta_j^{(k-1)} V_j \right) \right] \right) \tag{3.14}$$

$$\underline{\theta}^{(k)} = \arg\min_{\underline{\theta}} d_\tau \left(\mathcal{P}, \mathcal{T}^{(k)} \left[\mathrm{Exp}_{\underline{\underline{\mu}}} \left(\sum_j \theta_j V_j \right) \right] \right) \tag{3.15}$$

Each iteration of (3.14) is seeded with $\mathcal{T} = \mathcal{T}^{(k-1)}$. Likewise, each iteration of (3.15) is seeded with $\underline{\theta} = \underline{\theta}^{(k-1)}$. The overall optimization stops when the k^{th} iterations of (3.14) and (3.15) combined do not improve the objective function. That is, given some convergence criteria ϵ, the optimization stops when

$$d_\tau \left(\mathcal{P}, \mathcal{T}^{(k-1)} \left[\mathrm{Exp}_{\underline{\underline{\mu}}} \left(\sum_j \theta_j^{(k-1)} V_j \right) \right] \right) - d_\tau \left(\mathcal{P}, \mathcal{T}^{(k)} \left[\mathrm{Exp}_{\underline{\underline{\mu}}} \left(\sum_j \theta_j^{(k)} V_j \right) \right] \right) \leq \epsilon \tag{3.16}$$

Assuming (3.14) and (3.15) have converged by the k^{th} iteration, the initialization m-rep

can be defined as

$$\underline{\underline{m}}^{\text{init}} = \mathcal{T}^{(k)}\left[\text{Exp}_{\underline{\underline{\mu}}}\left(\sum_j \theta_j^{(k)} V_j\right)\right] \tag{3.17}$$

Let $\mathcal{T}^{\text{init}} = \mathcal{T}^{(k)}$ be known as the initialization similarity transform. Let $\underline{\underline{\mu}}^{\text{init}}$ be the m-rep instance defined as follows.

$$\begin{aligned}\underline{\underline{\mu}}^{\text{init}} &= \left(\mathcal{T}^{\text{init}}\right)^{-1}\left[\underline{\underline{m}}^{\text{init}}\right] \\ &= \text{Exp}_{\underline{\underline{\mu}}}\left(\sum_j \theta_j^{(k)} V_j\right)\end{aligned} \tag{3.18}$$

The initialization m-rep is used to initialize automatic segmentation of the form (2.108) where the shape is constrained to m-reps of the form

$$\underline{\underline{m}} = \mathcal{T}^{\text{init}}\left[\text{Exp}_{\underline{\underline{\mu}}^{\text{init}}}\left(\sum_j \theta_j V_j\right)\right], \tag{3.19}$$

where the PGA coefficients $\underline{\theta}$ are allowed to vary.

Note that (3.19) only approximates PGA as the eigenvectors $\{V_j\}$ were defined in $T_{\underline{\underline{\mu}}}\mathcal{M}(n)$. This poses a theoretical problem as $\{V_j\}$ need not be the directions of maximal variance when the training m-reps are projected into $T_{\left(\underline{\underline{\mu}}^{\text{init}}\right)}\mathcal{M}(n)$. Indeed, when $\text{Exp}_{\underline{\underline{\mu}}}$ is applied to those eigenvectors, it uses an $\bar{r}$ term that is defined relative to the training m-reps that were used to produce $\underline{\underline{\mu}}$. $\text{Exp}_{\underline{\underline{\mu}}^{\text{init}}}$ is implemented in (3.19) to use that same $\bar{r}$ value, even though parallel transport of geodesics on $\mathcal{M}(n)$ from $\underline{\underline{\mu}}$ to $\underline{\underline{\mu}}^{\text{init}}$ may require $\bar{r}$ to change. Empirically, this approximated PGA works quite well for bladder segmentation as will be shown in the experimental results in Sections 3.3 and 3.4.

3.3 SDSM with pelvic-scale alignment

This section describes an SDSM that has been useful for benchmarking initialization and segmentation algorithms in the context of segmenting the bladder from CT for ART. It is not a clinically realistic model for segmenting early treatment-day images because it includes

training exemplars from later treatment days.

Given a set of $k \approx 15$ treatment day CT images of the pelvis $\left\{ \underline{I}_i : i \in \{1..k\} \right\}$ of a single patient and reference manual segmentations of the bladder in those images, the following method was used to train an SDSM. M-reps $\left\{ \underline{\underline{m}}_i^{\text{fit}} : i \in \{1..k\} \right\} \subset \mathcal{M}(n)$ were fit to reference segmentations via the method described by Han and Merck [36, 57].

The pelvic bones in each image were automatically segmented via an intensity threshold in a region of interest. The pelvic bones are visible in Figure 3.1 as the high intensity structures surrounding the bladder and prostate. The shape of the pelvic bones is stable over the course of ART treatments, and thus the pose of the pelvic bones is used to align $\left\{ \underline{\underline{m}}_i^{\text{fit}} \right\}$. A rigid transformation $\mathcal{T}_i^{\text{bone}}$ is learned by rigidly registering the pelvic bones in $\underline{I}_1$ to the pelvic bones in $\underline{I}_i$.

The transformations $\left\{ \mathcal{T}_i^{\text{bone}} \right\}$ can be used to correct some pose changes among $\left\{ \underline{\underline{m}}_i^{\text{fit}} \right\}$, *i.e.,* pose changes due to how the patient is positioned on the table. However, these transformations cannot account for how the pose of the bladder changes relative to the pelvic bones, *i.e.,* as it fills and empties or as a consequence of changes in the rectum. The overlap among $\left\{ \underline{\underline{m}}_i^{\text{fit}} \right\}$ and among $\left\{ \left(\mathcal{T}_i^{\text{bone}} \right)^{-1} \left[\underline{\underline{m}}_i^{\text{fit}} \right] \right\}$ is shown in Figure 3.3.

SDSMs that can be used to segment the bladder in each $\underline{I}_i$ were trained by leaving out the data from that day. That is, PGA on $\left\{ \left(\mathcal{T}_j^{\text{bone}} \right)^{-1} \left[\underline{\underline{m}}_j^{\text{fit}} \right] : j \neq i \right\}$ was used to learn each shape prior $p\left(\underline{\underline{m}}_i \right)$, and the associated mean model $\underline{\underline{\mu}}_i$ and eigenmodes of deformation in $T_{\underline{\underline{\mu}}_i} \mathcal{M}(n)$. Intensity statistics in the form of m-rep spoke-end scale RIQFs were recorded for each $\underline{I}_i$ relative to $\underline{\underline{m}}_i^{\text{fit}}$. Each likelihood distribution $p\left(\underline{I}_i | \underline{\underline{m}} \right)$ was learned from the RIQFs from days $j \neq i$.

These SDSMs were used in experiments to compare initialization according to (3.14, 3.15, and 3.17) and segmentation of the form (2.108) constrained to a shape space of the form (3.19) with a naive alternative, where

$$\underline{\underline{m}}^{\text{init}} = \mathcal{T}_i^{\text{bone}} \left[\underline{\underline{\mu}}_i \right] \tag{3.20}$$

and segmentation is strictly according to (2.108).

3.3.1 Results

Section 2.6 described three metrics for segmentation evaluation: Dice similarity coefficient (2.128), average surface distance (2.122), and 90% worst surface distance (2.125). The contour-based bladder initializations and subsequent segmentations and the naive alternatives described in Section 3.3 for 78 total images from 5 ART patients were compared to manually produced reference segmentations according to these metrics. Results are plotted in Figures 3.4(a) (Dice coefficient), 3.5(a) (average surface distance), and 3.6(a) (90% worst surface distance). The abscissa of each plot indexes the images, sorted from the worst for each metric to the best. The ordinate of each plot indicates the value of the metric for that case. A summary of these plots is given in Table 3.1. Each entry in this table indicates the mean and standard deviation of a particular metric from a particular experiment.

Experiment	DSC	Distance (mean)	Distance (0.90)
Naive Initialization	0.811 ± 0.088	3.46 ± 1.62	7.49 ± 3.60
Naive Segmentation	0.928 ± 0.024	$\mathbf{1.26 \pm 0.37}$	$\mathbf{2.71 \pm 0.97}$
Contour-based Initialization	0.927 ± 0.024	1.32 ± 0.50	2.83 ± 1.23
Contour-based Segmentation	$\mathbf{0.930 \pm 0.023}$	$\mathbf{1.25 \pm 0.45}$	$\mathbf{2.68 \pm 1.01}$

Table 3.1: Evaluation of the bladder initializations and segmentations of 78 ART images using the methods reported in Section 3.3. Each measurement is reported as the mean score ± one standard deviation. "DSC" is the Dice similarity coefficient (2.128) with a manually produced reference segmentation. "Distance (mean)" is the average surface distance (2.122) in mm to that reference segmentation. "Distance (0.90)" is the 90% worst surface distance (2.125) in mm to that reference segmentation.

The distributions of volume overlap (Dice coefficient) with the reference segmentation for naive segmentations, contour-based initialization and contour-based segmentations are quite similar. The means and standard deviations of these three distributions are virtually indistinguishable. Based on the Dice coefficient distribution, each of these three bladder populations shows better agreement with the reference segmentations than the set of bladders produced by the naive initialization method.

Table 3.2 provides more information about the distribution of Dice coefficients for the different bladder populations. It lists the best and worst Dice coefficients as well as the three quartiles of each distribution. Because the most significant change in Dice coefficient between

Experiment → Quartile ↓	Naive Initialization	Naive Segmentation	Contour-based Initialization	Contour-based Segmentation
Worst Case	0.564	0.814	0.809	**0.834**
Lower	0.757	**0.921**	0.920	**0.921**
Median	0.838	0.934	0.934	**0.935**
Upper	0.888	**0.944**	0.942	0.943
Best Case	0.925	0.958	0.955	**0.959**

Table 3.2: The Dice similarity coefficient (2.128) of overlap with a manually produced reference segmentation broken down by quartiles.

the naive segmentations, contour-based initializations, and contour-based segmentations is in worst case performance, the worst quarter (*i.e.,* 20 cases) of each distribution is plotted in Figure 3.4(b). In this plot we can see that the contour-based segmentations have better overlap with the reference segmentations than do the naive segmentations for six of the seven worst cases. The naive segmentations have slightly better overlap in most of the remainder of the lower quartile, and then the contour-based segmentations have slightly better overlap for the majority of the remainder of the distribution.

Table 3.3 and Figure 3.5(b) provide a similar summary of the distribution of average surface distance to the reference segmentation for each of the bladder populations. On average the naive segmentation and contour-based segmentation perform the best, although the contour-based initialization is nearly as good and is distinctly better than the naive initialization. Worst case analysis shows that the naive segmentation is superior to the contour-based segmentation in the three worst cases, the contour-based segmentation and has a shorter average surface distance to the reference segmentation than the naive segmentation for the four next worst cases, and then the distributions are essentially equivalent. After those first seven worst cases, the contour-based initialization performs as well as the two segmentation methods. In the best cases the contour-based segmentations are again distinctly superior. Indeed, the best contour-based initializations have lower average surface distance to the corresponding reference segmentations than do the best naive segmentations.

The 90% worst surface distances reported in Table 3.4 and Figure 3.6(b) show essentially the same pattern as was seen for the average surface distance distributions. Although the

Experiment → Quartile ↓	Naive Initialization	Naive Segmentation	Contour-based Initialization	Contour-based Segmentation
Worst Case	7.67	**2.95**	4.30	3.68
Upper	4.13	**1.37**	1.45	1.43
Median	3.04	1.18	1.24	**1.17**
Lower	2.28	1.03	1.02	**1.00**
Best Case	1.20	0.84	0.71	**0.69**

Table 3.3: The average surface distance (2.122) in mm to a manually produced reference segmentation broken down by quartiles.

naive-segmentations produce the best results for the three worst cases, the contour-based segmentations show a significant improvement in the next few cases, before the distributions even out. Again the contour-based initializations tend to perform as well as the segmentation methods after those worst cases. As was the case with the report of average surface distance, the best cases produced by contour-based initialization and segmentation are superior to the best naive segmentations.

Experiment → Quartile ↓	Naive Initialization	Naive Segmentation	Contour-based Initialization	Contour-based Segmentation
Worst Case	20.28	**7.25**	10.06	7.58
Upper	9.29	**2.88**	3.03	3.07
Median	6.58	**2.43**	2.60	2.46
Lower	4.92	2.15	2.15	**2.07**
Best Case	2.59	1.76	1.49	**1.42**

Table 3.4: The 90% worst surface distance (2.125) in mm to a manually produced reference segmentation broken down by quartiles.

From these results I draw the following conclusions. The contour-based initialization method produces bladders that typically provide the same level of agreement with reference segmentations as do naively produced segmentations, and they do so without the use of a patient-specific image likelihood distribution. In the most difficult cases, image data is important for refining the initialization. In a case where the contour-based initialization does not provide as strong agreement with the reference segmentation as the naive segmentation does, a segmentation seeded with that initialization result will typically be competitive with the naive segmentation.

3.4 SDSM with bladder-scale alignment

This section describes a clinically realistic SDSM that can be used to initialize and segment the bladder from CT for ART. The SDSM models within-patient shape changes, but is trained using data from other patients so the early treatment days do not pose a HDLSS problem.

Let $\left\{\left\{\underline{I}_i^j : i \in \{1..k_j\}\right\} : j \in \{1..q\}\right\}$ denote sets of CT images of the male pelvis that will be used to train the SDSM. The sets are indexed by $j \in \{1..q\}$ corresponding to q different patients. Treatment day images of patient j are indexed by $i \in \{1..k_j\}$. I assume that reference segmentations are available for each of these training images and that m-reps $\underline{\underline{m}}_i^{j,\text{fit}} \in \mathcal{M}(n)$ are fit to the reference segmentations. The m-reps will be aligned first within-patient and then globally, across patients, before they are pooled to train the SDSM.

Let each $\mathcal{T}_i^{j,\text{bone}}$ denote the rigid transformation that registers $\underline{I}_1^j$ to $\underline{I}_i^j$ based on the pose of the pelvic bones. This registration was mentioned in Section 3.3 and can be performed automatically. Let $\underline{\underline{m}}_i^{j,\text{bone}}$ (3.21) denote the training m-rep, after bone-based alignment, for patient j on his i^{th} day of treatment. Let $\underline{\underline{\mu}}^{j,\text{bone}}$ denote the Fréchet mean bone-aligned bladder m-rep for patient j.

$$\underline{\underline{m}}_i^{j,\text{bone}} = \left(\mathcal{T}_i^{j,\text{bone}}\right)^{-1}\left[\underline{\underline{m}}_i^{j,\text{fit}}\right] \tag{3.21}$$

$$\underline{\underline{\mu}}^{j,\text{bone}} = \arg\min_{\underline{\underline{m}}} \sum_{i=1}^{k_j} \left\langle \underline{\underline{m}}_i^{j,\text{bone}}, \underline{\underline{m}} \right\rangle_{\mathcal{M}(n)} \tag{3.22}$$

An across-patient Fréchet mean m-rep $\underline{\underline{\mu}}^{\text{global}}$ and patient-specific (treatment-day independent) transformations are used for further alignment and are defined as follows.

$$\underline{\underline{\mu}}^{\text{global}} = \arg\min_{\underline{\underline{m}}} \sum_{j=1}^{q} \left\langle \underline{\underline{\mu}}^{j,\text{bone}}, \underline{\underline{m}} \right\rangle_{\mathcal{M}(n)} \tag{3.23}$$

$$\mathcal{T}^{j,\text{global}} = \arg\min_{\mathcal{T}} \left\langle \mathcal{T}\left[\underline{\underline{\mu}}^{\text{global}}\right], \underline{\underline{\mu}}^{j,\text{bone}} \right\rangle_{\mathcal{M}(n)}, \tag{3.24}$$

where the transformation $\mathcal{T}^{j,\text{global}}$ is found via optimization (3.24) over the space of similarity transformations.

Because bone-aligned bladders can vary in pose, a further within-patient bladder alignment

is performed before the across-patient similarity transformation is applied. Let $\mathcal{P}_i^j$ denote the set of points on the bladder of $\underline{I}_i^j$ on three designated slices as specified in Section 3.2. These points are used to find a similarity transformation $\mathcal{T}_i^{j,\text{bladder}}$ that refines the within-patient bladder alignments.

$$\begin{aligned}
\mathcal{T}_i^{j,\text{bladder}} &= \arg\min_{\mathcal{T}} d_\tau \left(\mathcal{T}_i^{j,\text{bone}} \left[\mathcal{T} \left[\underline{\underline{\mu}}^{j,\text{bone}} \right] \right], \mathcal{P}_i^j \right) && (3.25) \\
\underline{\underline{m}}_i^{j,\text{bladder}} &= \left(\mathcal{T}_i^{j,\text{bone}} \left[\mathcal{T}_i^{j,\text{bladder}} \right] \right)^{-1} \left[\underline{\underline{m}}_i^{j,\text{fit}} \right] \\
&= \left(\mathcal{T}_i^{j,\text{bladder}} \right)^{-1} \left[\underline{\underline{m}}_i^{j,\text{bone}} \right] && (3.26) \\
&\approx \underline{\underline{\mu}}^{j,\text{bone}} && (3.27)
\end{aligned}$$

I use the symbol "≈" in (3.27) to signify that $\underline{\underline{m}}_i^{j,\text{bladder}}$ and $\underline{\underline{\mu}}^{j,\text{bone}}$ have approximately the same pose. Let

$$\underline{\underline{m}}_i^{j,\text{global}} = \left(\mathcal{T}^{j,\text{global}} \right)^{-1} \left[\underline{\underline{m}}_i^{j,\text{bladder}} \right] \tag{3.28}$$

From (3.23) and (3.24) it follows that each $\underline{\underline{m}}_i^{j,\text{global}} \approx \underline{\underline{\mu}}^{\text{global}}$. Thus, the pool of m-reps $\left\{ \underline{\underline{m}}_i^{j,\text{global}} \right\}$, where i and j are allowed to vary, are more tightly aligned than $\left\{ \underline{\underline{m}}_i^{j,\text{bladder}} \right\}$ or $\left\{ \underline{\underline{m}}_i^{j,\text{bone}} \right\}$, as is illustrated in Figure 3.7. PGA on $\left\{ \underline{\underline{m}}_i^{j,\text{global}} \right\}$ should theoretically capture only shape changes.

An SDSM trained by PGA on $\left\{ \underline{\underline{m}}_i^{j,\text{global}} \right\}$ is not ideally suited for the segmentation of ART treatment-day images. This SDSM would capture between-patient shape variation, which is likely to be different from the day-to-day changes in a within-patient series of ART images. Moreover, the between-patient SDSM would be biased based on the number of training images, k_j that were available for each patient.

Based on the assumption that all patients undergo the same types of day-to-day shape changes, I use the following method to train an SDSM on within-patent, day-to-day shape variation. Recall that the action of medial atoms upon each other

$$\circ : \mathcal{M}(1) \times \mathcal{M}(1) \to \mathcal{M}(1)$$

was defined in (2.86) and can be extended via a direct product to

$$\circ : \mathcal{M}(n) \times \mathcal{M}(n) \to \mathcal{M}(n)$$

Let

$$\underline{\underline{\mu}}^{j,\text{global}} = \left(\mathcal{T}^{j,\text{global}}\right)^{-1} \left[\underline{\underline{\mu}}^{j,\text{bladder}}\right] \tag{3.29}$$

The inverse action of this patient-specific mean, captures within-patient shape change in the space to which the multi-patient m-reps were aligned. Let

$$\Delta\underline{\underline{m}}_i^{j,\text{global}} = \left(\underline{\underline{\mu}}^{j,\text{global}}\right)^{-1} \circ \underline{\underline{m}}_i^{j,\text{global}} \tag{3.30}$$

PGA on the set of m-rep residues $\left\{\Delta\underline{\underline{m}}_i^{j,\text{global}}\right\}$ where i and j vary produces an SDSM of within-patient shape variability.

Let $\Delta\underline{\underline{\mu}}^{\text{global}}$ be the Fréchet mean m-rep residue. PGA on $\left\{\Delta\underline{\underline{m}}_i^{j,\text{global}}\right\}$ produces a set of eigenvectors $\{V_w\} \subset T_{\left(\Delta\underline{\underline{\mu}}^{\text{global}}\right)}\mathcal{M}(n)$. Because each m-rep residue is being treated as a sample from a zero-mean Gaussian distribution, I make the approximation that

$$\Delta\underline{\underline{\mu}}^{\text{global}} = \text{Id}_{\mathcal{M}(n)} \tag{3.31}$$

Thus given a tuple of PGA coefficients $\underline{\theta}$, a new m-rep residue $\Delta\underline{\underline{m}}$ can be generated as follows.

$$\Delta\underline{\underline{m}} = \text{Exp}_{\text{Id}_{\mathcal{M}(n)}}\left(\sum_w \theta_w V_w\right) \tag{3.32}$$

An m-rep bladder instance $\underline{\underline{m}}^{\text{global}} \approx \underline{\underline{\mu}}^{\text{global}}$ can then be deformed as follows.

$$\begin{aligned} \underline{\underline{m}}^{\text{def,global}} &= \underline{\underline{m}}^{\text{global}} \circ \text{Exp}_{\text{Id}_{\mathcal{M}(n)}}\left(\sum_w \theta_w V_w\right) \\ &= \text{Exp}_{\underline{\underline{m}}^{\text{global}}}\left(\sum_w \theta_w V_w\right) \end{aligned} \tag{3.33}$$

To use these eigenmodes of within-patient shape change to segment the i'^{th} treatment day

image for target patient j' (*i.e.,* not one of the training patients), the base model $\underline{\underline{m}}^{\text{global}}$ in (3.33) should capture patient-specific information about the shape of the target patient's bladder. This base model is defined (3.35) below.

Let $\underline{\underline{\mu}}_{i'}^{j,\text{bone}}$ denote the patient-specific mean-to-date bladder shape using data available just prior to the segmentation of $\underline{I}_{i'}^{j'}$. $\underline{\underline{\mu}}_{i'}^{j',\text{bone}}$ is computed according to (3.21) and (3.22) as the Fréchet mean of the bladder segmentations from images $\left\{\underline{I}_{i}^{j'} : i \in \{1..(i'-1)\}\right\}$, after each of those segmentations have been rigidly registered to $\underline{I}_{1}^{j'}$. This mean-to-date m-rep can be understood as the expected value of the target patient's bladder shape (in the coordinate system of $\underline{I}_{1}^{j'}$) after treatment days $\{1..(i'-1)\}$. It is reasonable to assume that the previous days' images have been segmented before the i'^{th} treatment day, however the need for this patient-specific data prevents this method from being used to segment the first image of a new patient. The construction of an appropriate SDSM for that purpose is outside the scope of this dissertation.

The bone-aligned mean-to-date m-rep $\underline{\underline{\mu}}_{i'}^{j',\text{bone}}$ can then be aligned to the global training mean in a manner similar to (3.24). That alignment produces $\underline{\underline{\mu}}_{i'}^{j',\text{global}}$: the base model for use in (3.33).

$$\mathcal{T}_{i'}^{j',\text{global}} = \arg\min_{\mathcal{T}} \left\langle \mathcal{T}\left[\underline{\underline{\mu}}^{\text{global}}\right], \underline{\underline{\mu}}_{i'}^{j',\text{bone}} \right\rangle_{\mathcal{M}(n)} \tag{3.34}$$

$$\underline{\underline{\mu}}_{i'}^{j',\text{global}} = \left(\mathcal{T}_{i'}^{j',\text{global}}\right)^{-1}\left[\underline{\underline{\mu}}_{i'}^{j',\text{bone}}\right] \tag{3.35}$$

The SDSM that combines the patient-specific mean-to-date shape $\underline{\underline{\mu}}_{i'}^{j',\text{global}}$ with the within-patient shape variation learned via PGA on $\left\{\Delta\underline{\underline{m}}_{i}^{j,\text{global}}\right\}$ can be used in (3.14, 3.15, and 3.17) to initialize the segmentation of $\underline{I}_{i'}^{j'}$. That initialization m-rep can be used to constrain a segmentation via (2.108) to a shape space of the form (3.19). The results of bladder initialization and segmentations in 73 ART images are discussed in Section 3.4.1. These segmentations were produced using image-likelihood distributions learned via a variation of PCA on RIQFs that combines the mean-to-date QF in each region from the target patient with within-patient variability learned through PCA on the residues of QFs in the corresponding regions of the training case relative to their patient-specific mean QFs.

3.4.1 Results

Bladder initialization and segmentation of 73 CT images $\left\{\left\{\underline{I}_i^j : i \in 2..k_j\right\} : j \in 1..5\right\}$ was performed according to the method described in Section 3.4. As in Section 3.3.1, these m-rep models were evaluated through measurements of their Dice coefficient (2.128) with reference to manual segmentations and their average surface distance (2.122) and 90% worst surface distance (2.125) to those reference segmentations. These evaluations are summarized in Table 3.5.

Experiment	DSC	Distance (mean)	Distance (0.90)
Naive Initialization	0.815 ± 0.088	3.35 ± 1.59	7.30 ± 3.59
Naive Segmentation	$\mathbf{0.929 \pm 0.023}$	$\mathbf{1.24 \pm 0.36}$	$\mathbf{2.65 \pm 0.95}$
Contour-based Initialization	0.921 ± 0.023	1.39 ± 0.36	2.96 ± 0.84
Contour-based Segmentation	0.924 ± 0.020	1.34 ± 0.36	2.83 ± 0.83

Table 3.5: Evaluation of the bladder initializations and segmentations reported in Section 3.4. Each measurement is reported as the mean score ± one standard deviation. "DSC" is the Dice similarity coefficient (2.128) with a manually produced reference segmentation. "Distance (mean)" is the average surface distance (2.122) in mm to that reference segmentation. "Distance (0.90)" is the 90% worst surface distance (2.125) in mm to that reference segmentation.

Table 3.5 also reports evaluations of naive initialization and segmentations, that were produced according to the method described in Section 3.3. Note that the naive initialization and segmentation results are not identical to the naive results reported in Section 3.3.1. The naive results reported here exclude the measurements for the first treatment day of each patient because no contour-based results exist for those images.

Table 3.5 shows that the naive segmentation method produces the best results according to each of the metrics. This is not surprising, since the naive method makes use of target-patient-specific information that cannot be known on the treatment days. The shape and image intensity patterns from future treatment days are used to build the naive models. The contour-based methods reported here use only information that could be known at treatment time. The contour-based methods perform nearly as well as the naive segmentation method, and the contour-based initialization does so without needing to learn an image likelihood distribution.

The distribution of measured Dice coefficients is presented in Figure 3.8 and Table 3.6. Although the contour-based segmentation method on average performs slightly worse than the naive alternative, it does produce greater overlap with the reference segmentations in the worst cases.

Experiment → Quartile ↓	Naive Initialization	Naive Segmentation	Contour-based Initialization	Contour-based Segmentation
Worst Case	0.564	0.814	0.813	**0.858**
Lower	0.764	**0.921**	0.910	0.910
Median	0.842	**0.934**	0.923	0.928
Upper	0.889	**0.945**	0.937	0.936
Best Case	0.925	0.958	0.957	**0.963**

Table 3.6: The Dice similarity coefficient (2.128) of overlap with a manually produced reference segmentation broken down by quartiles.

Further analysis of the average surface distances from these m-reps to the reference segmentations is presented in Table 3.7. This metric confirms that the contour-based segmentations are nearly as good as the naive segmentations, although there is less improvement in the worst cases than was seen during the other evaluation. The distributions of average surface distance measurements are shown in Figure 3.9.

Experiment → Quartile ↓	Naive Initialization	Naive Segmentation	Contour-based Initialization	Contour-based Segmentation
Worst Case	7.67	2.95	**2.53**	2.64
Upper	4.01	**1.33**	1.55	1.49
Median	2.79	**1.15**	1.30	1.27
Lower	2.09	**1.03**	1.18	1.11
Best Case	1.20	0.84	0.80	**0.74**

Table 3.7: The average surface distance (2.122) in mm to a manually produced reference segmentation broken down by quartiles.

90% worst surface distance measures are reported in Table 3.8 and Figure 3.10. The evaluation using this measurement is comparable to the evaluation based on the Dice coefficient. Although the contour-based segmentations are on average slightly worse than the naive segmentations, the contour-based methods do improve the worst case performance.

Experiment → Quartile ↓	Naive Initialization	Naive Segmentation	Contour-based Initialization	Contour-based Segmentation
Worst Case	20.28	7.25	**5.76**	6.50
Upper	8.79	**2.85**	3.33	3.17
Median	5.89	**2.39**	2.73	2.69
Lower	4.84	**2.15**	2.39	2.31
Best Case	2.59	1.76	1.60	**1.51**

Table 3.8: The 90% worst surface distance (2.125) in mm to a manually produced reference segmentation broken down by quartiles.

3.5 Discussion

I have introduced a method for the initialization of SDSM-based segmentation of the bladder in CT images of ART patients. This method is based on minimizing the τ-band distance (3.10) from a set of manually specified points $\mathcal{P}$ on three axial slices of the image to the boundary of the SDSM instance by iteratively refining estimates of the pose (3.14) and the shape (3.15) of the bladder in the target image. In Section 3.3, I discussed this method in the context of a clinically unrealistic SDSM and image likelihood distribution that were trained on images and segmentations of the patient's bladder from many other ART treatment days. The results of that experiment were that the initializations agreed with reference segmentations nearly as well as did image-driven, SDSM-based segmentations that were initialized via rigid registration of the pelvic bones. Image-driven segmentations seeded by these initializations were typically superior to the segmentations produced by the alternative method. In Section 3.4, I discussed a clinically realistic SDSM that combined the best estimate at treatment time of the shape of the target patient's bladder with a model of within-patient bladder-shape variation that was trained from other patients. Although the probability distributions used for initialization and segmentation in this experiment included less patient-specific information than was used in the previous experiment, the results were nearly as strong.

The initialization results I have reported are comparable to segmentation results reported in the literature. Recently Costa *et al.* [20] reported measures on segmentations of the bladder from CT equivalent to an Dice coefficient of 0.882, and Pasquier *et al.* [63] reported segmentations of the bladder from multi-modal MR images with an average Jaccard score (2.127) of

0.88. That Jaccard score is equivalent to a Dice coefficient of 0.936. The average Dice coefficient for clinically realistic contour-based initializations that I reported in Section 3.4 was 0.921. That score is nearly as high as the value reported by Pasquier *et al.* and significantly higher than the value reported by Costa *et al.*. Moreover, both of those segmentation methods were designed to leverage image intensity data, while the contour-based initialization was designed to be a preprocessing step prior to an intensity-driven segmentation.

This initialization method relies on the user's ability to identify $\mathcal{P}$ on the boundary of the bladder on three designated axial image slices. In the experiments reported here, $\mathcal{P}$ typically consisted of 110 boundary points. An area of future research is to better understand how the number and location of points to be specified in order for the method to be able to estimate the pose and shape of the bladder.

In a preliminary experiment, I have seen that redefining $\mathcal{P}$ so that it only contains points on the designated axial slices that lie on a strong edge in that axial slice (*i.e.,* only using points that the user can easily identify) does not significantly affect the initialization quality. The quartiles of the Dice coefficient for initialization based on a clinically realistic SDSM using this subset of $\mathcal{P}$ and using the original $\mathcal{P}$ (as reported in Section 3.4.1) is reported in Table 3.9. This experiment was performed using an earlier version of the software than that which produced the results in Section 3.4.1, so changes may be due to other factors besides the modification of $\mathcal{P}$. However these results do suggest that rigorous contouring on each of the designated slices is probably not needed to initialize the bladder.

	Worst Case	Lower Quartile	Median	Upper Quartile	Best Case
Original $\mathcal{P}$	0.813	**0.910**	**0.923**	**0.937**	**0.957**
$\mathcal{P}$ on edges only	**0.845**	0.900	0.920	0.932	0.956

Table 3.9: Dice coefficient for contour-based initializations with manually produced reference segmentations. The top row reports the result from Table 3.6. The bottom row reports the result after $\mathcal{P}$ has been pruned to only include points that lie on a strong edge in the image.

Another area of future work is in generalizing this method so that it can initialize SDSM-based segmentation of other organs in the male pelvis. Applying this method directly to

rectum segmentation has not been fruitful because so much of the shape variation in the rectum occurs outside of the designated axial slices. It is likely that $\mathcal{P}$ would have to be redefined in order for this method to successfully initialize rectum segmentation. Preliminary results of prostate initialization via this method [15] have been quite encouraging. It is likely that there is sufficient prostate-shape variation on three axial image slices to allow the pose and shape of the gland to be recovered by this initialization method.

Although between-patient shape variation is typically greater than within-patient shape variation, preliminary tests of bladder and prostate initialization using this method with SDSMs trained on between-patient variation [15] have been successful. Indeed, my colleague Edward Chaney reports that prostate segmentation initializations produced using a SDSM that models between-patient variation and with $\mathcal{P}$defined by four to eight boundary points from each of three designated axial slices have been successful. Initializations from an SDSM that captures between-patient variation will likely be used during the segmentation of the first treatment-day image for a new patient. The resulting segmentation, possibly after manually corrections, could then be used as an estimate of the mean-to-date shape prior to the second treatment day. The initialization and segmentation of that image could then be performed using a SDSM based on the the tighter distribution of within-patient shape variation.

Because $\underline{\underline{m}}^{\text{init}}$, the bladder instance produced by minimizing the τ-band distance to manually specified points, tends to be such a good approximation of the pose and shape of the bladder in the target image, this method could provide the basis for an interactive segmentation routine. Tools that allow the user to edit $\mathcal{P}$ by adding points (possibly on other image slices), adjusting the within-slice location of a point, and removing points would enable the interactive refinement of the result. The method for evaluating and signaling the local credibility of a segmentation that I will present in Chapter 4 could provide important feedback and guidance for the user during this interaction.

(a) Axial slice, unaligned

(b) Sagittal slice, unaligned

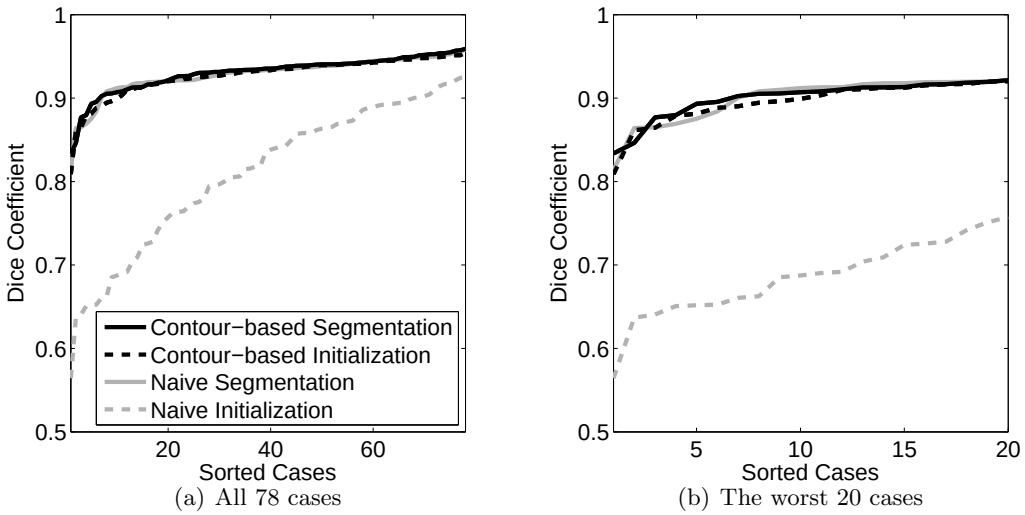
(c) Coronal slice, unaligned

(d) Axial slice, bone-based alignment

(e) Sagittal slice, bone-based alignment

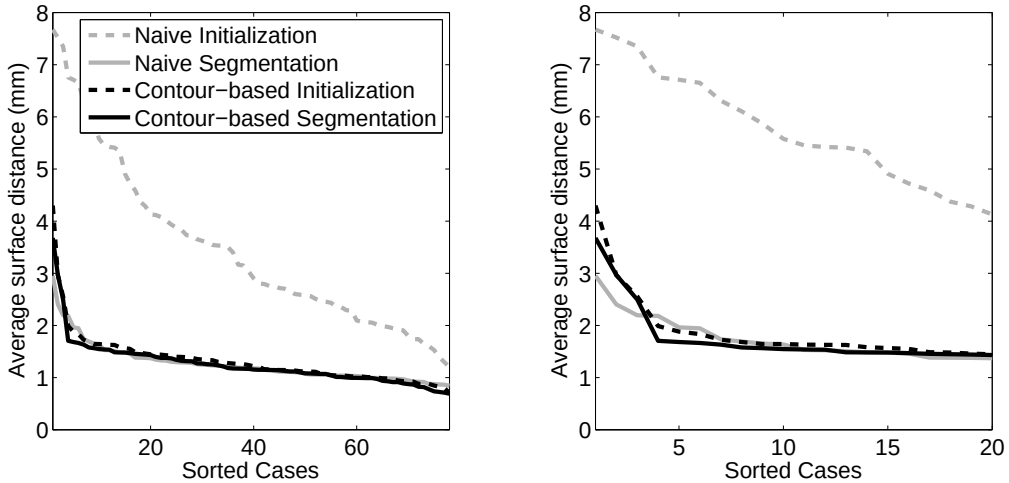
(f) Coronal slice, bone-based alignment

Figure 3.3: The overlap of m-reps $\left\{\underline{\underline{m}}_i^{\mathbf{fit}}\right\}$ fit to the manual segmentations of the bladder in 14 CT images of a single patient's pelvis is shown. The number of images for which a given voxel is on the interior of the bladder is indicated by the darkness of that voxel. Figures 3.3(a), 3.3(b), and 3.3(c) show the overlap of the unaligned m-reps. The lack of concentrated black regions means that PGA on these m-reps would learn significant pose and shape changes. Figures 3.3(d), 3.3(e), and 3.3(f) show the overlap of m-reps of the form $\left\{\left(\mathcal{T}_i^{\mathbf{bone}}\right)^{-1}\left[\underline{\underline{m}}_i^{\mathbf{fit}}\right]\right\}$. These m-reps have been aligned via a rigid transformation based on the pose of the pelvic bones. Although there are some pose changes between images, the most significant variation in this population is due to shape changes.



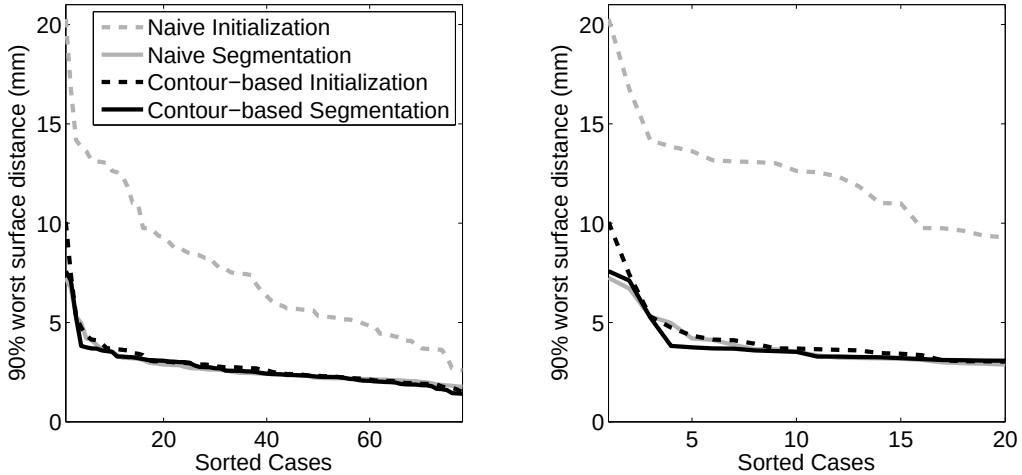
(a) All 78 cases (b) The worst 20 cases

Figure 3.4: The Dice similarity coefficient (2.128) of manually produced reference segmentations and bladder initializations and segmentations (Figure 3.4(a)). The cases in the lower quartile have the worst overlap with the corresponding reference segmentations and are shown in Figure 3.4(b)



(a) All 78 cases (b) The worst 20 cases

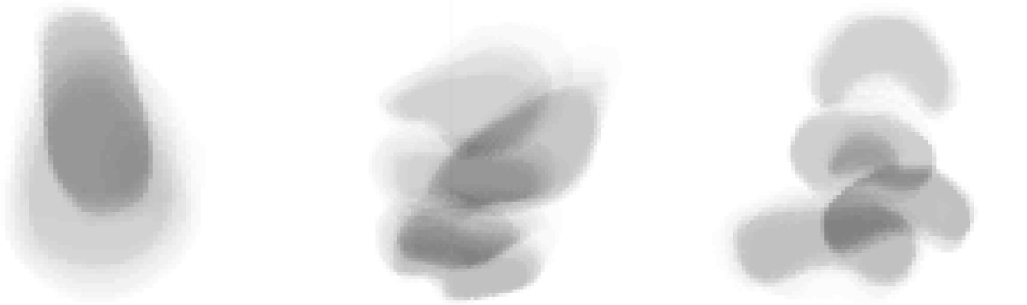
Figure 3.5: The average surface distance (2.122) between manually produced reference segmentations and bladder initializations and segmentations (Figure 3.5(a)). The cases in the upper quartile are the farthest from the corresponding reference segmentations and are shown in Figure 3.5(b)



(a) All 78 cases

(b) The worst 20 cases

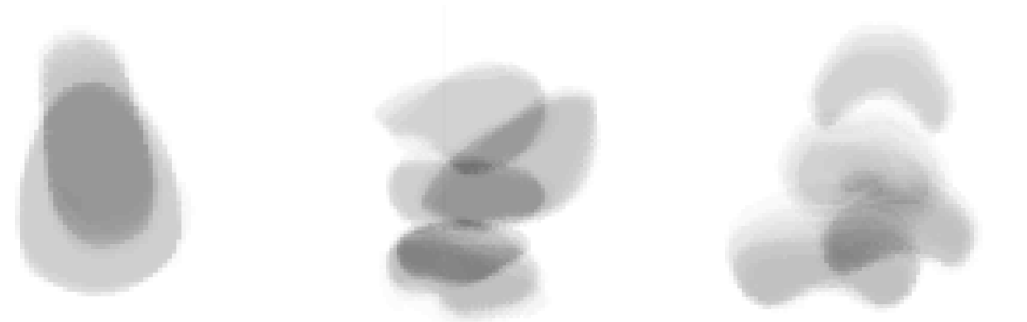
Figure 3.6: The 90% worst surface distance (2.125) between manually produced reference segmentations and bladder initializations and segmentations (Figure 3.6(a)). The cases in the upper quartile are the farthest from the corresponding reference segmentations and are shown in Figure 3.6(b)



(a) Axial slice, bone-based alignment

(b) Sagittal slice, bone-based alignment

(c) Coronal slice, bone-based alignment



(d) Axial slice, bladder-based alignment

(e) Sagittal slice, bladder-based alignment

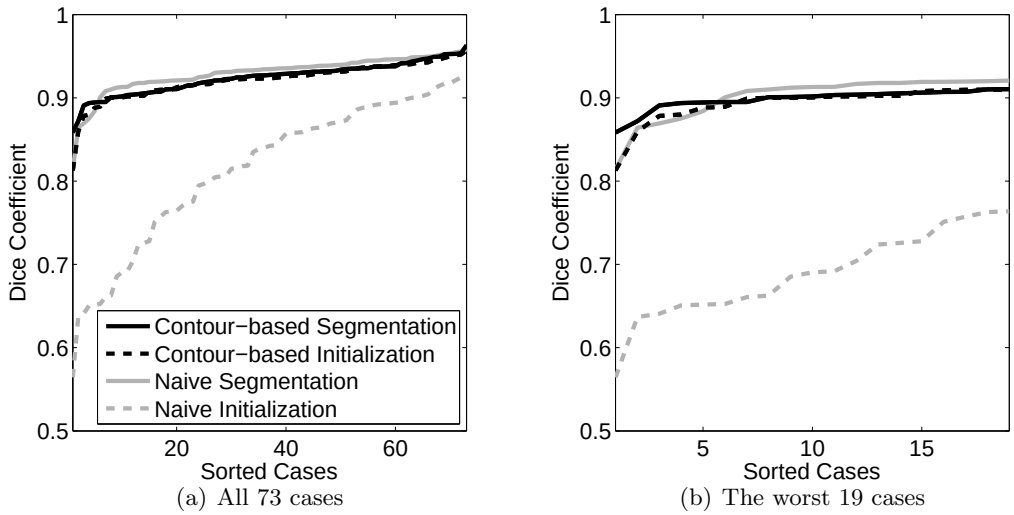
(f) Coronal slice, bladder-based alignment

(g) Axial slice, multi-patient alignment

(h) Sagittal slice, multi-patient alignment

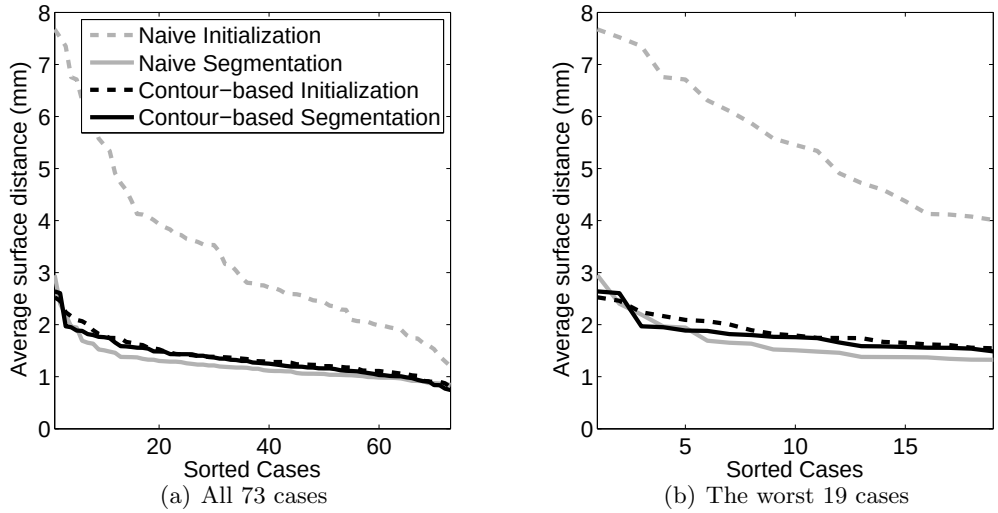
(i) Coronal slice, multi-patient alignment

Figure 3.7: The overlap of m-reps fit to the manual segmentations of the bladder in 78 CT images of a the male-pelvis from 5 ART patients is shown. The number of images for which a given voxel is on the interior of the bladder is indicated by the darkness of that voxel. Figures 3.7(a), 3.7(b), and 3.7(c) show the overlap of m-reps $\left\{\underline{\underline{m}}_i^{j,\mathbf{bone}}\right\}$ (3.21). Clusters corresponding to the five patients can be seen in these figures, and the inhomogeneity within each cluster indicates that further within-patient alignment is possible. Figures 3.7(d), 3.7(e), and 3.7(f) show the overlap of m-reps $\left\{\underline{\underline{m}}_i^{j,\mathbf{bladder}}\right\}$ (3.26). Clusters corresponding to the five patients can still be seen, but there is greater overlap within each cluster. Figures 3.7(g), 3.7(h), and 3.7(i) show the overlap of m-reps $\left\{\underline{\underline{m}}_i^{j,\mathbf{global}}\right\}$ (3.28) These bladders are well aligned to each other.



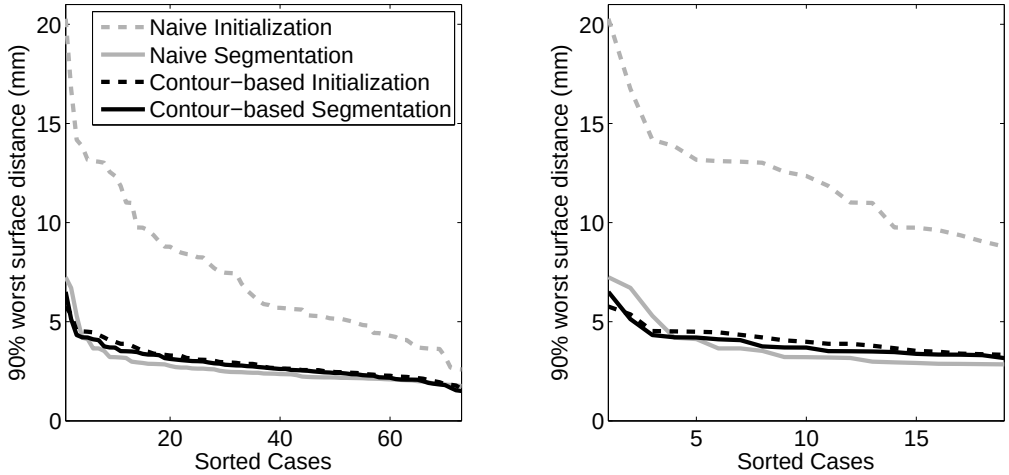
(a) All 73 cases (b) The worst 19 cases

Figure 3.8: The Dice similarity coefficient (2.128) of manually produced reference segmentations and bladder initializations and segmentations (Figure 3.8(a)). The cases in the lower quartile have the worst overlap with the corresponding reference segmentations and are shown in Figure 3.8(b)



(a) All 73 cases (b) The worst 19 cases

Figure 3.9: The average surface distance (2.122) between manually produced reference segmentations and bladder initializations and segmentations (Figure 3.9(a)). The cases in the upper quartile are the farthest from the corresponding reference segmentations and are shown in Figure 3.9(b)



(a) All 73 cases

(b) The worst 19 cases

Figure 3.10: The 90% worst surface distance (2.125) between manually produced reference segmentations and bladder initializations and segmentations (Figure 3.10(a)). The cases in the upper quartile are the farthest from the corresponding reference segmentations and are shown in Figure 3.10(b)

Chapter 4

Estimating the Local Credibility of a Segmentation

A preliminary version of this chapter was published as [48].

Limits on human time and attention restrict the number of images that can be manually segmented. The cost of manual segmentations scales linearly with the number of images to be analyzed and the number of relevant slices per image. However, an automated system can be scaled and distributed in order to segment an arbitrarily large data set. Another setting in which manual segmentation of a medical image is not a viable option is when the underlying data changes rapidly. During the time a careful human rater spends segmenting the structures of interest in the image, the data can change to the point where the image is no longer an accurate portrait of the subject, and the segmentation of image data is irrelevant to the medical problem. An automated system may be able to provide a sufficiently rapid response to produce a segmentation during the short interval in which the image can be used.

An automated system that frequently produces clinically acceptable segmentations brings with it a jeopardy that a segmentation failure will go undetected and will be put to use. For the same reasons of cost that manual segmentations cannot be performed in these situations, fully manual validation of automatic segmentations is not a viable option. In the situation where the underlying biology is rapidly changing, the validity of the image would expire before the validity of the image segmentation could be established. Because of these limitations, I

propose a novel, automated method for rapidly detecting and signaling the location of non-credible regions on a segmentation boundary.

My method for identifying non-credible regions requires a geometry to image match ("image match") function that can be evaluated region by region at the segmentation surface. Let $\underline{I}$ denote an image and let $\underline{\underline{m}}$ denote some parameterization of the segmentation of the structure of interest in the image data. The image match $f(\underline{\underline{m}}, \underline{I})$ serves as a proxy for the distance from the segmentation surface to the unknown surface representing the ground truth for the object in the image. The method requires that $\underline{I}$ can be decomposed into object-relative regions, indexed by i, $\left\{\underline{I}^i\right\}$ and that $f(\underline{\underline{m}}, \underline{I})$ can be decomposed into local terms $f^i\left(\underline{\underline{m}}, \underline{I}^i\right)$ for each region. Given two neighboring regions $i \neq j$, I make the simplifying assumption that $f^i\left(\underline{\underline{m}}, \underline{I}^i\right)$ and $f^i\left(\underline{\underline{m}}, \underline{I}^j\right)$ are independent. This is likely an unworkable assumption due to spatial correlation of image intensity patterns. Each local image match term indicates the goodness of fit for a subregion of the segmented object. Statistical analysis on the distribution of the value of each local image match term over a set of well fit training cases produces tests to identify non-credible regions on a segmentation surface.

In a region where the local image match has an unusually poor value, the surface distance to the unknown truth is expected to be large, so the segmentation should be considered non-credible. Let us assume, without loss of generality, that the value of the local image match function increases as the localized goodness of fit decreases. This is consistent with the notation used in Section 2.5, where an image match function was defined as a penalty that has a low cost for an SDSM instance that matches the image data.

Given a new image $\hat{\underline{I}}$ and its segmentation $\hat{\underline{\underline{m}}}$, I want to identify the set of regions:

$$\mathcal{N} \quad = \quad \left\{i : P_{\underline{\underline{m}}, \underline{I}}\left[f^i\left(\underline{\underline{m}}, \underline{I}^i\right) \geq f^i\left(\hat{\underline{\underline{m}}}, \hat{\underline{I}}^i\right)\right] < \rho\right\} \tag{4.1}$$

for some critical value ρ. $\mathcal{N}$ is the set of non-credible regions on the segmentation surface.

Many of the shape representations discussed in Section 2.5 allow the regional decomposition of a target image and of an image match function. This can be seen in the following examples:

1. The Chan-Vese functional [14] is an image match that is understood relative to two

regions corresponding to the interior and exterior of an object and is frequently used in segmentations based on level-set methods (Section 2.5.1).

2. An image region can be defined as a neighborhood around a PDM landmark, and the ASM objective function can be understood as the composition of local terms for each region (Section 2.5.2).

3. The m-rep object-relative coordinate system can decompose an object into regions, and PCA on RIQFs [9] defines regional image match functions (Section 2.5.3).

Given a shape model that yields object-relative coordinates and an image match function amenable to local decomposition, (4.1) requires the probability distribution for each local image match term to be known in order to identify outliers. Certain classes of local image match functions allow principled assumptions about this distribution to be made.

When $f^i(\cdot,\cdot)$ is the log-likelihood of a multivariate Gaussian, its value can be mapped onto the square of a Mahalanobis distance (2.20). The squared Mahalanobis distance can be understood as the sum of squared IID standard normal variables (2.43) and thus as a χ^2 random variable [11]. This is the case when the image match function is produced via PCA on a training set that can be assumed to follow a multivariate Gaussian distribution. The image match based on RIQFs is trained in this manner and has been used in the results presented in this chapter.

Given my ability to detect non-credible regions using (4.1), I must now decide how my system should proceed when such a region is detected. Three options for handling this situation are the following:

1. Alert the user to the locations of $\mathcal{N}$ and defer to the user's judgment on how to react

2. Notify the user and force the user to interactively resegment the object in $\mathcal{N}$

3. Automatically adjust the segmentation.

Currently, I disregard the third option, noting that in the special case of segmentation by optimizing the fit of a deformable shape model [19, 65] an optimization that yields an outlier

value of the image match function may indicate a segmentation failure but does not prescribe a method to correct it. In Section 4.4, I will discuss a special case where automatic correction of a non-credible segmentation may be possible.

I therefore recommend that when a non-credible region is detected the system should notify the user and rely on the user's expert judgment to resolve the situation. In some applications the user may choose to run a semi-automatic segmentation editor, such as the one proposed by Grady [34], in order to correct the segmentation in the non-credible regions. In other applications it may be sufficient for the user to manually evaluate and correct the segmentation.

The benefit of using this method to detect and signal non-credible regions is that manual validation resources are only used in the cases and locations where a segmentation error is likely to have occurred. The prohibitively expensive step of manually verifying each segmentation in its entirety has been eliminated.

The rest of this chapter is organized as follows. I describe an experiment in which this method was applied to m-rep-based segmentations of the bladder and prostate in Section 4.1. Visualizations of the non-credible regions that were identified during this experiment are presented in Section 4.2. In Section 4.3, I present an evaluation of the detected non-credible regions. This chapter concludes with a discussion of these results and of future work.

4.1 Detecting Non-credible Regions

In this section, I describe the materials used during the application of this method to detect non-credible regions in m-rep-based bladder and prostate segmentations. The image data, m-reps, and RIQF-based image match used during this experiment are specified.

The data set used during the experiments reported here consists of 80 CT images of the pelvic region of 5 patients receiving ART to treat prostate cancer. Each patient's treatment has been fractionated over a series of dates, and a new image was acquired prior to many treatment sessions. The bladder and prostate in each image were segmented by Bayesian optimization over m-rep deformable models [64]. These segmentations required us to train a

geometric prior distribution and an image match function.

Let $\underline{I}_t^p$ denote the t^{th} image of patient p. Let $\underline{\underline{m}}_t^{p,k}$ denote the m-rep segmentation of the bladder ($k = 0$) and the prostate ($k = 1$) in $\underline{I}_t^p$. A manual segmentation of each image was performed. These segmentations of patient p on all days other than the target day were used to train the shape prior and the image match. Let $\underline{\underline{m}}_t^{p,0,\text{fit}}$ denote a training m-rep fit to the manual segmentation of the bladder in $\underline{I}_t^p$. Each $\underline{\underline{m}}_t^{p,0,\text{fit}}$ was constructed, via the method of Han and Merck [36, 57], so that the m-rep coordinate system preserved anatomical correspondences. Each bladder segmentation $\underline{\underline{m}}_t^{p,0}$ was produced using a shape prior learned by Principal Geodesic Analysis [26] over the set of m-reps $\left\{\underline{\underline{m}}_j^{p,0,\text{fit}} : j \neq t\right\}$, holding p constant, and using an image match trained on object-scale RIQFs for the same set of models. Each prostate segmentation $\underline{\underline{m}}_t^{p,1}$ was produced by applying the same procedure to the prostate training data.

Although the bladder segmentation methodology used in these experiments is similar to the naive method reported in Section 3.3, the probability distributions used were not precisely equivalent. These training m-reps $\left\{\underline{\underline{m}}_j^{p,o,\text{fit}}\right\}$ were produced by a preliminary implementation of the method of Han and Merck [36, 57] that assigned different values to the weights used in (2.78) and (2.79), than were used in the training for the experiments reported in Section 3.3. After the training used in this experiment was completed, the distance function d_{I} was improved to better estimate the distance from the boundary of the m-rep to reference segmentation in cases when that distance was likely to be significantly different from the distance in the other direction, *i.e.,* from the reference segmentation to the m-rep. These changes in the training tools led to differences between the training m-reps, which in turn affected SDSM training and RIQF region definitions. Furthermore, although the segmentations used in these experiments were produced using object-scale RIQFs, the test of non-credibility uses smaller scale RIQFs that are described below.

The local image match functions used in the non-credibility test were trained on regions defined by the m-rep coordinate system. Each image region $\underline{I}^i$ is defined to be the neighborhood of a sampled m-rep spoke end, which serves as the anchor point for the region and is identified by its object-relative coordinates (u, v, ϕ). Each neighborhood is defined to extend 1 cm along

the surface normal, and each voxel makes a Gaussian weighted contribution to the QF based on it's distance from the object boundary. The peak of this Gaussian is at the object boundary, and its standard deviation is defined to be 1/3 cm. Each region is further constrained so that only voxels for which the nearest surface point is within 2.5 cm of the region's anchor point are allowed to contribute to the QF.

Because the training m-reps preserve anatomical correspondence, local RIQF statistics on a corresponding region across the training population for an object can be learned. PCA on the training cases yields a mean μ^i and n eigenmodes of variation $\left\{\lambda_j^i, v_j^i\right\}$ for each RIQF. The local image match function used to assess the local credibility of this region in each target case, $\underline{\underline{m}}_t^{p,k}$ can be understood as a Mahalanobis distance of an RIQF Q^i sampled from the target image, according to the probability distribution learned from the PCA.

$$f^i\left(\underline{\underline{m}}, \underline{I}^i\right) = \sum_{j=1}^{n} \frac{\left(\left(Q^i - \mu^i\right) \cdot v_j^i\right)^2}{\lambda_j^i} + \frac{\|r\|^2}{\lambda_r^i} \tag{4.2}$$

The first term is the Mahalanobis distance to the intensity quantiles observed in the target case in the PCA space truncated to n eigenmodes of variation. This term can be understood as a χ^2 random variable with n degrees of freedom. The second term accounts for the residue outside of this PCA space; this residue is assumed to follow an isotropic Gaussian probability distribution.

$$r = \left\| Q^i - \mu^i - \sum_{j=1}^{n} \left(\left(Q^i - \mu^i\right) \cdot v_j^i\right) v_j^i \right\|^2 \tag{4.3}$$

r is weighted by the standard deviation $\sqrt{\lambda_r^i}$ in the training cases that is unaccounted for in the truncated PCA space, so the second term in (4.2) can be assumed to be the square of standard normal random variable, or equivalently as a χ^2 random variable with one degree of freedom. Because the sum of independent χ^2 random variables is also distributed as χ^2, $f^i\left(\underline{\underline{m}}, \underline{I}^i\right)$ can be understood as a ${\chi^2}_{n+1}$ random variable.

This image match has several properties that are desirable for detecting non-credible regions in image segmentations. I have observed positive correlation between image match value and

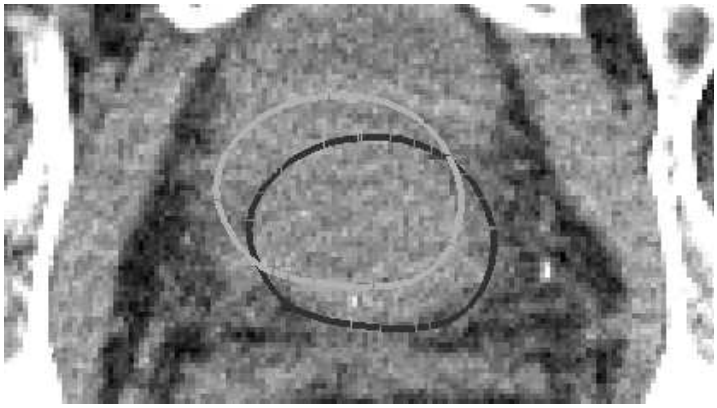


Figure 4.1: An axial slice of a CT image illustrating why the exterior image match is used to detect non-credibility in prostate segmentations. The two segmentations shown in the image are quite different, yet have similar interior intensity patterns. It is the local exterior intensity pattern that allows one to distinguish between the acceptable segmentation (dark contour) and the erroneous segmentation (bright contour).

segmentation error distance, suggesting that the outliers of local image match are likely to be regions where a localized failure has occurred. This image match models the variability in image intensity for a population of objects and is appropriate for evaluating the segmentation of new image of the same class. Because the value of the image match function follows the χ^2 distribution, a critical value of that distribution can be used as the threshold ρ in (4.1).

The local RIQFs used to train the credibility test for bladder segmentations use three eigenmodes of variation to describe the image intensity quantiles from the exterior of the object and two eigenmodes to describe the interior of the object. Two additional degrees of freedom, one each for the exterior and interior, are needed to account for the residue outside of these eigenmodes. Voxels that were segmented as gas or bone via intensity thresholding were excluded from the RIQFs and did not contribute to the match function. Thus, the distribution of local image match values can be approximated as $f^i(\cdot,\cdot) \sim {\chi^2}_7$. The choice of a threshold value f such that $P\left[f^i(\cdot,\cdot) > f\right] < \rho$ for any value ρ can be made via the known CDF for the ${\chi^2}_7$ distribution. Any region of a bladder segmentation where the image match exceeds f is considered to be non-credible.

Only the exterior RIQFs were used to evaluate the local credibility of prostate segmentations because of the lack of image contrast between the bladder and prostate. Figure 4.1

shows two possible segmentations of a prostate on an axial slice of a CT image. One of these segmentations is correct; the other is the result of a gross shift towards the bladder. Because the image intensity patterns for bladder and prostate are similar, the interior histograms for these two segmentations are roughly equivalent, as is their values of interior image match. A trained expert's ability to detect the segmentation error is based on the intensity patterns at the exterior of the prostate, in the region away from the bladder. These regional intensity distributions are quite different, and the exterior RIQF image match will be able to distinguish between them. Exterior RIQFs with three eigenmodes of variability are used to detect non-credible regions of the prostate. The additional degree of freedom for residue outside of the PCA space allows the distribution of image match values in a region to be approximated by $f^i(\cdot,\cdot) \sim {\chi^2}_4$. A threshold value can be chosen from the CDF of that probability distribution, and any image match above that threshold can be interpreted as evidence that the segmentation in that region is non-credible.

4.2 Visualizing Non-credible Regions

My method evaluates a localized image match function, region by region, along the surface of a segmentation. Those regions where the image match value exceeds a threshold, chosen based on the probability distribution for that regional image match, are considered to have a non-credible segmentation. I prepare a visualization as in Figure 4.2 that allows the user to understand the location of suspected segmentation errors.

This visualization is analogous to those proposed by Niessen [62] to communicate validation results. In the left pane of Figure 4.2, the non-credible region is indicated by a red patch on the surface of the segmented bladder. In the right pane, the non-credible region is indicated by a red segment on the contour where the segmented bladder intersects an axial image slice. Based on the visualization, the user is responsible for evaluating the segmentation in these regions, and correcting it as necessary.

In the bladder segmentation shown in Figure 4.2, the non-credible region appears to be over-segmented, *i.e.,* the segmentation captures some background voxels in addition to all of

the bladder voxels in that region. The user would presumably correct this segmentation by bringing the highlighted boundary closer to the high contrast edge in the image.

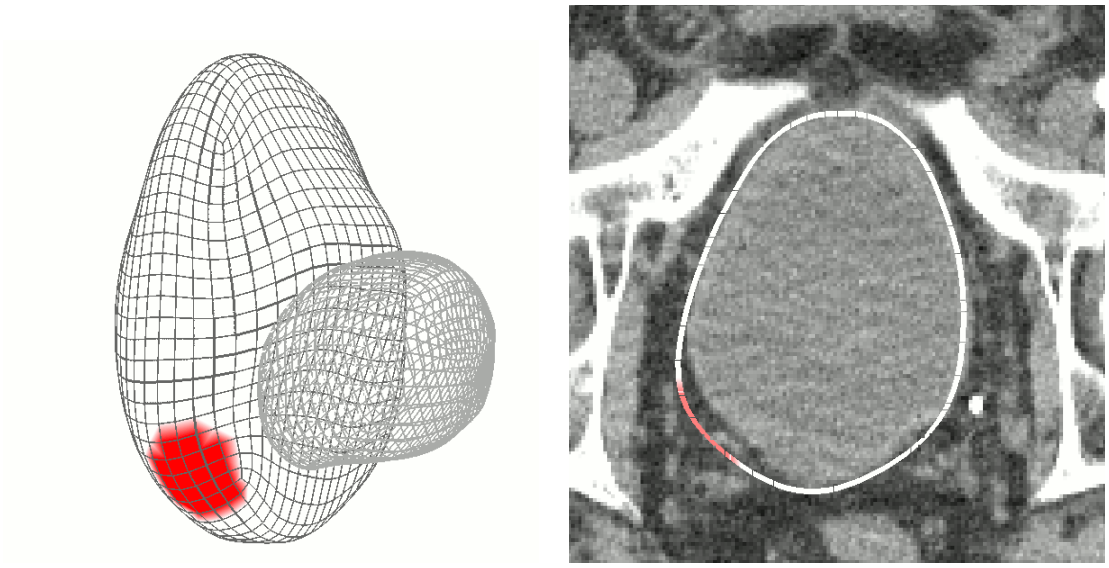


Figure 4.2: (Left) A display indicating a non-credible region (red tiles) on a bladder segmentation surface. For orientation purposes, the prostate segmentation surface is shown with unshaded tiles. (Right) The intersection of the same surface with an axial slice of the CT image. In the regions where the segmentation was determined to be credible (white contour) it is well fit to the image data. A significant error is visible in the non-credible (red contour) region.

A second example of this visualization is shown in Figure 4.3. In the left pane we see the prostate surface with non-credible regions signaled by red shading. In the right pane we see the intersection of this surface with a sagittal slice of the CT image. Manual validation and correction of the prostate segmentation in the non-credible region is recommended. Due to the inherent challenge in manually segmenting the prostate, it is not clear what action should be taken in this non-credible region.

Figure 4.4 shows a result from an early study evaluating the local credibility of kidney segmentations. CT reconstruction errors produced the bright circular artifacts in this image. The segmentation regions near these artifacts were all flagged as non-credible, even though in this case no correction to the segmentation is needed. Still, there is value in having the user verify the segmentation in these regions and in alerting the user to the imaging failure.

The visualizations shown thus far (Figures 4.2, 4.3, and 4.4) each use a binary map (credible

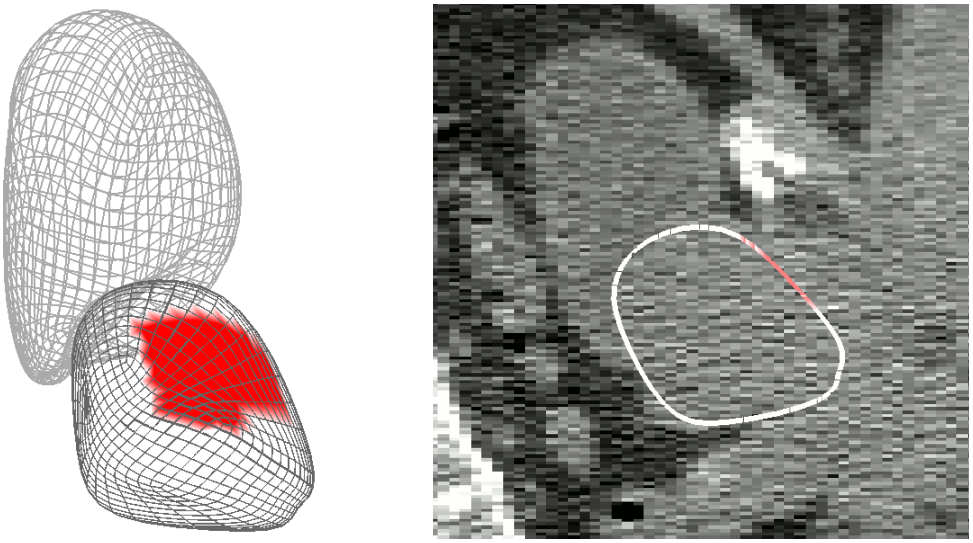


Figure 4.3: (Left) A display indicating a non-credible region (red tiles) on a prostate segmentation surface. For orientation purposes, the bladder segmentation is shown with unshaded tiles. (Right) The intersection of the prostate surface with a sagittal slice of the CT image. Non-credible regions are indicated by the red contour.

or not) to indicate the location of non-credible regions. A visualization that shows multiple levels of segmentation credibility allows the user to prioritize his validation tasks. Figure 4.5 shows the regional image match scores for the bladder segmentation shown in Figure 4.2, ranging from the most credible (dark) to the least credible (bright).

4.3 Validation

To validate my test for local non-credibility I wished to show that the regions I detected as having outlier values of the local image match function are the same regions where local segmentation failures have occurred. Although I do not know the ground truth for the images in my study, I do have access to the wood standard: a single manual segmentation for each of these images. These reference segmentations are used to define local segmentation failure as occurring when the distance from the representative point for a region (*i.e.,* the m-rep spoke end that serves as the anchor point) to the nearest point on the manual segmentation surface exceeds a limit. That is when $d^i(\underline{\underline{m}}, \underline{I}) > \epsilon$.

The scale of local segmentation errors that I can detect with my test is related to the

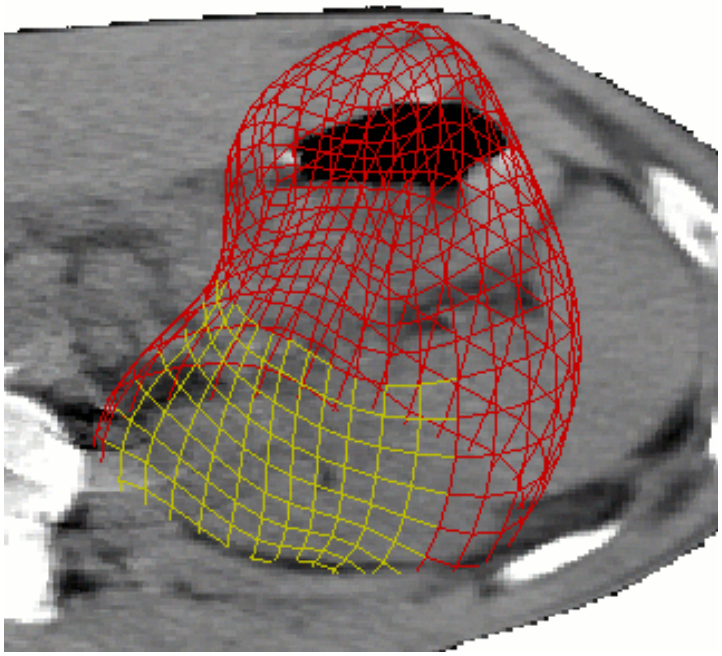


Figure 4.4: A miscalibrated detector in the scanner led to reconstruction errors that produced the concentric circles in the image. The regions on the surface of a kidney segmentation that intersects these imaging errors are labeled as non-credible (yellow). In this case, the segmentation in these regions with outlier intensity patterns is acceptable, but there is still value in alerting the clinician to the imaging errors.

quality of fits used to train $f^i(\cdot,\cdot)$. Surely the image match cannot be sensitive to errors of the same order of magnitude that were present at training time. For all but two outlier cases of the 80 bladder training m-reps in $\left\{\underline{\underline{m}}_t^{p,0,\text{fit}}\right\}$, the maximum distance from any m-rep spoke-end to the reference segmentation was less than 4.0 mm. The maximum surface error for any of the training m-reps for the prostate, $\left\{\underline{\underline{m}}_t^{p,1,\text{fit}}\right\}$, was less than 4.5 mm. Based on these measurements the test of non-credibility is expected to be sensitive to segmentation errors where $d^i(\underline{\underline{m}}, \underline{I}) > \epsilon > 4.5$ mm. Empirically, I found that the test performed well with error defined as $\epsilon = 5.5$ mm for the bladder and $\epsilon = 6.0$ mm for the prostate.

The segmentation of the bladders and prostates in all 80 CT images were globally quite good. The average over all cases of the average surface distance to the reference segmentation for the bladder was 1.4 mm with the maximum over all cases being 2.7 mm. [1] For the prostate

[1]Since these experiments were performed, the methodology and tools for m-rep-based segmentations have improved. The clinically realistic bladder initializations reported in Section 3.4 also had on average a 1.4 mm

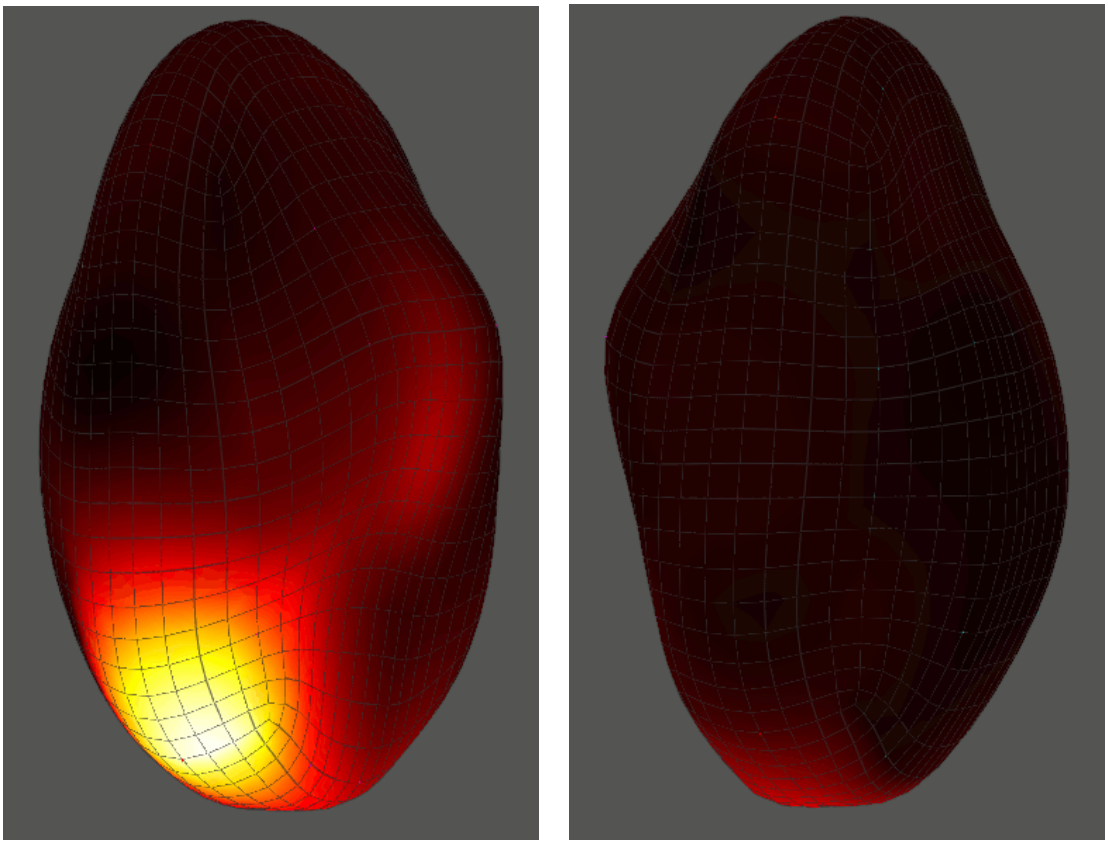


Figure 4.5: A display indicating the relative credibility of regions on the surface of the bladder segmentation that was shown in Figure 4.2. The surface is colored according to a heat-scale on the local image match value. The least credible regions are shown in "hot" colors (white, yellow). The most credible regions are shown in "cold" colors (black, dark red). The same bladder is shown twice in this figure. The left pane gives a view that shows the indentation into the bladder, near the prostate. The right pane shows a view of the opposite side of the bladder.

the average of the average surface distances was 1.2 mm with a maximum of 4.0 mm.

The distribution of local errors in the bladder segmentations was as follows. In 64 of the 80 segmentations no local error was ever greater than $\epsilon = 5.5$ mm. In the remaining 16 cases, 61 out of 1248 possible regions had a significant local error. The spatial distribution of these regions can be seen in the left pane of Figure 4.6. The errors are the most densely concentrated at the low contrast area where the prostate indents into the bladder and on the opposite side

average surface distance to the reference segmentation with the maximum of 2.5 mm. The segmentations discussed in that chapter had on average a 1.3 mm average surface distance to the reference segmentation with the maximum of 2.6 mm. Those results were produced with less patient-specific information than was used during these experiments.

of the medial sheet from that location.

There were fewer cases with local errors in the prostate segmentations than there were for the bladder. 76 of the 80 segmentations had no local error greater than $\epsilon = 6.0$ mm. In the remaining 4 cases, 38 out of 296 possible regions had a significant local error. The spatial distribution of these regions can be seen in the right pane of Figure 4.6. There was a large cluster of errors on the face of the prostate that is adjacent to the bladder. There were also smaller clusters of errors on the opposite faces.

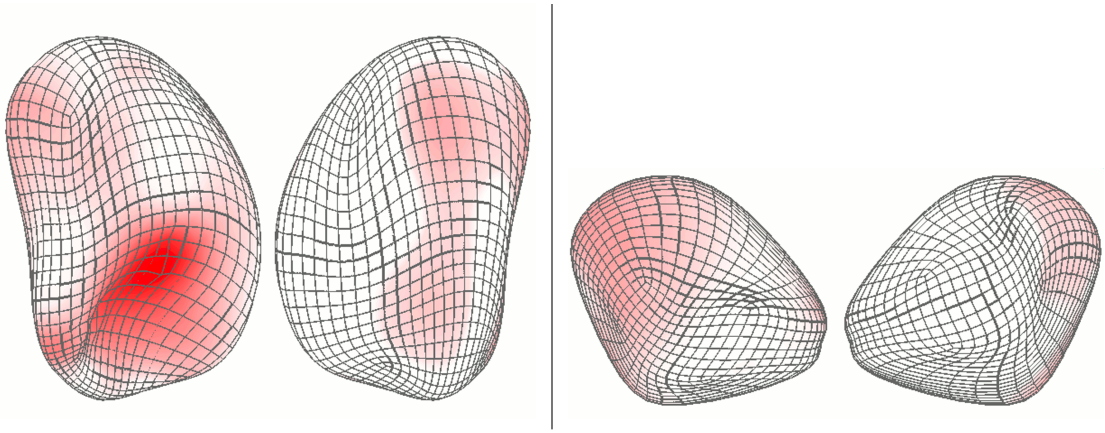


Figure 4.6: M-rep implied surfaces for a bladder (left) and prostate (right). Darker red shading signals regions where errors occurred more frequently. The most frequent errors occurred in the low contrast image region near the boundary between the bladder and the prostate.

The cases and regions with gross localized errors tend to have larger values of the local RIQF image match function than those where the local error is below the ϵ threshold. This is illustrated in Figure 4.7 and summarized below. For the median 50% of acceptably segmented bladder regions, $7.19 \leq f^i(\underline{\underline{m}}, \underline{I}^i) \leq 22.92$, but $36.76 \leq f^i(\underline{\underline{m}}, \underline{I}^i) \leq 125.00$ for the median 50% of grossly erroneous bladder regions. For the median 50% of acceptably segmented prostate regions, $2.48 \leq f^i(\underline{\underline{m}}, \underline{I}^i) \leq 7.05$, but $11.35 \leq f^i(\underline{\underline{m}}, \underline{I}^i) \leq 31.59$ for the median 50% of grossly erroneous prostate regions.

4.3.1 ROC Analysis

These definitions of error have been used to validate of the test of non-credibility. For an image match threshold of f, the test result at a region on a segmentation surface can be

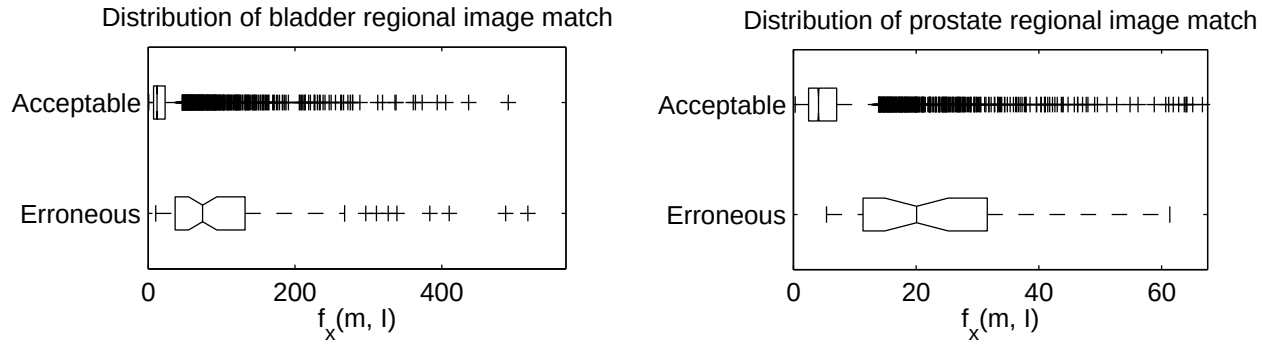


Figure 4.7: Distribution of the local RIQF image match value for the bladder (left) and the prostate (right) segmentations shown using box plots. The image match values for regions where $d^i(\underline{\underline{m}}, \underline{I}) \geq \epsilon$ tends to be larger than for those regions where $d^i(\underline{\underline{m}}, \underline{I}) < \epsilon$. Because each image match value is modeled as a χ^2 random variable, the skewness of the distributions shown in the plots is expected.

classified in one of four ways:

	$f^i\left(\underline{\underline{m}}, \underline{I}^i\right) \geq f$	$f^i\left(\underline{\underline{m}}, \underline{I}^i\right) < f$
$d^i\left(\underline{\underline{m}}, \underline{I}\right) \geq \epsilon$	True-Positive	False-Negative
$d^i\left(\underline{\underline{m}}, \underline{I}\right) < \epsilon$	False-Positive	True-Negative

For a fixed error definition ϵ, the true-positive rate (TPR) and false-positive rate (FPR) can be understood as functions of the test threshold f that characterize the performance of the test.

$$TPR(f) = \frac{\left\| f^i\left(\underline{\underline{m}}, \underline{I}^i\right) \geq f \wedge d^i\left(\underline{\underline{m}}, \underline{I}\right) \geq \epsilon \right\|}{\left\| d^i\left(\underline{\underline{m}}, \underline{I}\right) \geq \epsilon \right\|} \tag{4.4}$$

$$FPR(f) = \frac{\left\| f^i\left(\underline{\underline{m}}, \underline{I}^i\right) \geq f \wedge d^i\left(\underline{\underline{m}}, \underline{I}\right) < \epsilon \right\|}{\left\| d^i\left(\underline{\underline{m}}, \underline{I}\right) < \epsilon \right\|} \tag{4.5}$$

The true-positive rate of a test is also known as its "sensitivity" and is similar in spirit to the generalization ability of an SDSM. Just as an SDSM with high generalization ability tends to be able to model new shape instances from the target population, a test with high sensitivity tends to be positive for new data from the target class. The complement of the false-positive rate of a test, *i.e.,* $1 - FPR$, is also known as its "specificity" and is similar

in spirit to the specificity of an SDSM. Just as an SDSM with high specificity tends to only generate shapes that belong to the target population, a test with high specificity tends to only be positive for data that truly belong to the target class.

A plot of the sensitivity of a family of tests as a monotonically increasing function of $1-$specificity, *i.e.,* TPR as a function of FPR, is known as a receiver operating characteristic (ROC) curve and is useful for understanding the behavior of the tests. It is trivial to design a test such that $TPR = FPR = 1$ (this test is positive for every possible input) or $TPR = FPR = 0$ (the test is negative for every possible input). The ROC curve for a test that randomly decides when to return positive should lie on the $TPR = FPR$ line. The ROC curve for an effective family of tests will rise sharply, offering high sensitivity even in highly specific configurations.

An ROC curve is often summarized by a measure of the area under the curve (AUC). This value is bounded by $0 \leq \text{AUC} \leq 1$, with a high AUC indicating that fewer false-positives need be tolerated in order to achieve a high true-positive rate. Another interpretation of AUC is that it is the probability that when presented with two data instances, one that truly belongs to the class the test is intended to identify and one that does not, the test will have a greater response to the truly positive instance.

I refer the reader to Metz *et al.* [58] for a more complete discussion of ROC analysis. This work includes a statistical model for the conditional distribution of the test result given the class of its input and algorithms for robustly estimating AUC for an ROC curve.

Returning to the test of segmentation non-credibility, varying the image threshold f produces an ROC curve (Figure 4.8). The ROC curve illustrates the tradeoff between the false-positives and false-negatives of the test of non-credibility, and allows the user to select an image match threshold for which the tradeoff is clinically acceptable. Its AUC can be understood as:

$$AUC = P\left[f^i\left(\underline{\underline{m}}_1, \underline{I}_1^i\right) > f^i\left(\underline{\underline{m}}_2, \underline{I}_2^i\right) \mid \left(d^i\left(\underline{\underline{m}}_1, \underline{I}_1\right) \geq \epsilon\right) \wedge \left(d^i\left(\underline{\underline{m}}_2, \underline{I}_2\right) < \epsilon\right)\right] \quad (4.6)$$

This gives the probability that given corresponding regions from two randomly selected segmentations, one with a gross error in that region and one that is acceptably close to the

reference segmentation, the test will correctly identify the erroneous region. That is, the erroneous region will have the greater image match score.

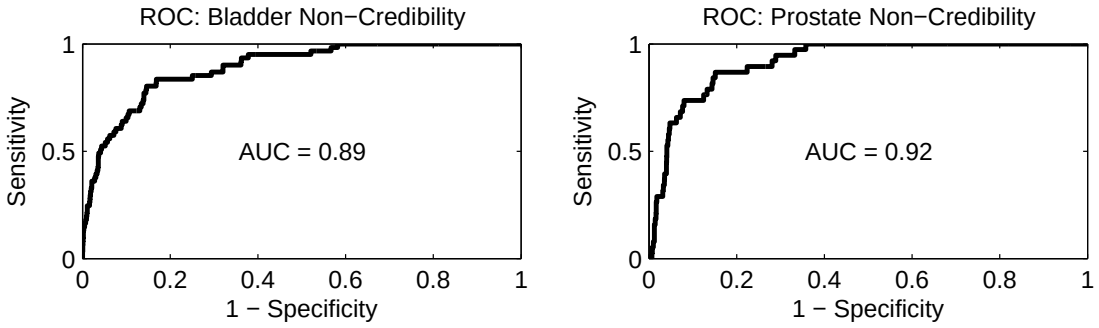


Figure 4.8: ROC curves characterizing the performance of the test for non-credible regions in bladder and prostate segmentations. For both organs the $AUC \approx 0.9$ indicating that the test successfully distinguishes between segmentation regions that are acceptably close to the reference segmentation and those that are too great of a distance away.

Figure 4.8 shows the ROC curve for the tests of regional non-credibility of bladder and prostate segmentations. AUC is 0.89 for the bladder and 0.92 for the prostate, showing that in nearly 90% of cases, the test correctly distinguishes between a region with a significant local error and a region that is correctly segmented. This ROC curve is used to decide the threshold level at which to report non-credibility. This choice is a tradeoff between the expected true-positive and false-positive rate of the test and needs to be made with application specific rules in mind. The illustrations in this chapter (Figures 4.2 and 4.3) were produced using threshold values that minimize the number of false-positive reports.

4.4 Discussion and Conclusions

I have presented a novel method for automatically detecting and signaling the location of non-credible regions on a segmentation surface. This method is based on detecting statistical outliers of a local image match function. I have applied this method to m-rep based segmentations of the bladder and prostate in 80 CT images. I validated these results with ROC analysis relating the local image match value with the distance from a representative point for each region to a reference segmentation of the same image. The AUCs for these ROC curves were 0.89 for the bladder and 0.92 for the prostate, indicating that the test is successful

at distinguishing between those regions which are acceptably segmented and those that are grossly erroneous.

The experiments reported here used a different set of image match functions to detect non-credible regions than were used to produce the segmentations. That the method for detecting non-credible regions in segmentations can be decoupled from the segmentation methodology allows it to be more widely applied. It suggests a strategy for comparing segmentation methods. The differences between two segmentation methods can be understood by the sets of regions for which one method produces a credible segmentation but the other does not. It also suggests a strategy for comparing image match functions, the differences between two such image match functions can be understood in terms of regions with disputed credibility.

The test of regional credibility for the bladder that was presented in Section 4.1 interprets the sum of the interior and exterior image matches as a single χ^2 random variable. This assumes that the interior and exterior matches are independent of each other. Given a correct segmentation, this is a fair assumption to make. Since the test of segmentation credibility is intended to be applied to segmentations that are not known to be correct, this assumption may be too strong. Evaluating the credibility of the interior and exterior regions separately is likely to identify more suspected non-credible regions. One can imagine a case similar to the one shown in Figure 4.2, but with less over-segmentation. In such a hypothetical case, the exterior RIQF could be sufficiently close to its mean to offset the deviation in the interior RIQF, *i.e.,* the sum of those image matches may be an inlier according to its χ^2 distribution even though the interior match is an outlier according to its distribution. By evaluating the credibility of the interior and the exterior regions separately, the problem in that interior region could still potentially be detected.

Although this test was originally designed to aid in manual validation of segmentations, it may also provide useful feedback in an automatic segmentation process. In the bladder segmentation shown in Figure 4.2, the RIQF for the interior region near the highlighted part of the boundary is a statistical outlier. However, a quantile function of the portion of that region that is close to the boundary is likely to be an inlier according to the probability distribution on QFs from the corresponding exterior region. If that is the case, it is evidence that the

segmentation should be corrected by local erosion in that region. Similarly, an image that has been locally under-segmented may have an outlier exterior QF, a portion of which is an inlier according to the corresponding interior RIQF distribution, in which case the segmentation should be corrected by local dilation in that region. An icon that suggests the appropriate correction could be included in an interactive segmentation editor. However, since these errors could be detected by software, there is hope that they could be corrected automatically.

Automatic corrections could be incorporated into a local refinement stage that allows an m-rep-based segmentation to escape the PGA shape space. I believe errors, such as the one shown in Figure 4.2, occur because the PGA shape space lacks generalization ability. In that case, local refinements, either based on atom-scale PGA [51, 54] or based on non-probabilistic adjustments to the parameters of a medial atom [64], are likely necessary to correct the errors. During a refinement of this sort of segmentation error, the probability distribution on QFs from the opposite side of the boundary could be used as an *a contrario* [60] distribution and the objective function formulated to explicitly avoid an interior that looks like exterior (or vice-versa). Of course, this technique is only valid in regions where the interior and exterior intensity distributions are expected to be different.

The best strategy for utilizing an *a contrario* RIQF distribution is a topic for future research. One possibility is to optimize over the region-depth on the side of the boundary corresponding to the outlier image match, in order to estimate the distance to the true object boundary. In other words, the RIQF sampling depth that produces the most likely QF according to the *a contrario* distribution could be understood as a measure of the local segmentation error in that region. An alternative would be to slide the segmentation boundary along its normal, in search of a position that produces a RIQF that is likely according to its distribution and unlikely according to the *a contrario* distribution. A third possibility is to leverage recent work by Stough [77] on mixing operations on QFs. The sampled QF for a candidate segmentation could be understood as a mixture of the distribution of QFs for that region, the *a contrario* distribution and noise. The objective function would be designed to favor smaller contributions from the *a contrario* distribution. Although these strategies have not yet been tested, I expect them to be useful in failure cases like the one shown in Figure 4.2.

The RIQF sampling depth, the associate weighting scheme for the contribution of voxels to a QF, and the radius of the projection of a region onto the segmentation surface are three parameters that were introduced in Section 2.5.3 and that affect the probability distributions learned from RIQFs. These parameters also affect the performance of the test for regional credibility. The stability of a QF increases with the number of voxels that contribute to it, and thus with the sampling depth and radius. If too few voxels contribute the QF, its instability will lead to greater variability in the population of corresponding QFs. This variability may allow an RIQF sampled from an erroneously segmented target image to be accepted as an inlier. However, the QF is intended to model homogeneous tissue intensities, and if the region is allowed to expand too much, the QF will be corrupted by other tissue classes, and the distribution of corresponding QFs across images cannot be assumed to be stable. Indeed, Stough [77] has shown that RIQFs defined at a medial-spoke scale are better suited for bladder and prostate segmentations than RIQFs defined at an object-scale. Defining the regions to be too large can also lead to an RIQF sampled from an erroneously segmented target image being accepted as an inlier. Thus, tuning the RIQF parameters requires a bit of finesse, so that QFs stably estimate the appropriate intensity distributions within and across images.

The test validation reported in this chapter relies on three simplifying assumptions that warrant further work. First, the manual segmentation of each image serves as the wood standard ground truth for each image. Given what is known about the intra- and inter- expert segmentation variability, a better, bronze standard estimate of the ground truth could be made by combining multiple manual segmentations as with the STAPLE [89] algorithm. An improved ground truth estimate would allow for a better measurement of localized segmentation errors. The second simplifying assumption is related to the measurement of localized segmentation errors. The reported results were based on the distance from a single representative point per region to the manual segmentation. An improvement would be to incorporate the distribution of distances from points within the region. The third simplifying assumption is that the image match at neighboring regions is independent. Given that images are spatially correlated, this assumption is not likely to hold.

There are two significant consequences to the correlation of image data that future work

could address. One stream of future work is to incorporate neighbor relationships in the validation strategy. New rules for labeling a region could be developed that account for the status of nearby regions. For example, when adjacent regions both have a large image match but only one has a significant error distance, it might be appropriate to consider both regions as true-positives. The other stream of future work is related to enhancements of the RIQF image match, to model conditional distributions of intensity quantiles, and to allow regions to merge, split, and slide along the segmentation boundary [77].

Chapter 5

Interpolation of Oriented Geometric Objects via Rotational Flows

A preliminary version of this chapter was published as [49].

There are many interesting applications for surface interpolation. Perhaps the most direct application is to generate animations that visualize the differences between objects. In the general realm of shape analysis, a surface interpolation scheme can be used to measure the difference between objects, which in turn can be used to align objects or calculate a mean. In the more specialized field of medical image analysis, the paths that boundary points follow during the interpolation can be used to drive a registration process which then enables morphometry and atlas-based segmentation. It is these latter problems, registration and atlas-based segmentation, that have inspired this research.

The graphics community has been interested in creating visually pleasing morphs between objects. For example, the work by Breen and Whitaker [8] can be used to interpolate between implicitly represented shapes. The implicit shape representation allows for changes in topology. Although this feature is valued in the context of general purpose animation, in our target applications it is undesirable.

Interpolation of landmark positions has been performed via large deformation diffeomorphisms [41] found with respect to a Riemannian metric. These interpolations benefit from the diffeomorphic property. Each interpolated shape is guaranteed not to fold or tear. However, the interpolation is not guaranteed to be shape maintaining, and the computation cost of such

an interpolation tends to be high.

Recent work by Vaillant and Glaunes [87] and by Glaunes and Joshi [30] has used the concept of currents from differential geometry to define a metric on shapes. This metric is used to drive a global optimization that produces a diffeomorphic warp under the large deformation framework [41]. This warp can then be used to bring the reference shape forward, through a series of interpolating shapes, and into its target configuration. However, the computation needed to produce this warp can be expensive.

Pizer *et al.* [65] advocate the use of medial shape representations (m-reps, see Section 2.5.3) because of their ability to describe nonlinear shape changes such as bending, twisting, and widening. Fletcher *et al.* [26] showed that each medial atom in an m-rep object lies on the manifold $\mathcal{M}(1)$, and that geodesics on this manifold can be defined using a standard Riemannian metric (2.91). A pair of m-rep objects can be interpolated by interpolating each pair of corresponding medial atoms along the geodesic between them. However, as a medial atom follows such a geodesic, its hub moves along a linear path in $\mathbb{R}^3$ independent of the changes to its orientation and size. Due to the straight-line nature of these paths, a pair of medial atoms could collide during the interpolation, creating an undesired fold in the image space. Even if no atoms collide during the interpolation, an interpolated surface may not appear to belong to the intended object class.

In this work I propose a novel interpolation for shapes that can be described by a set of oriented samples (vertices). This method that uses the orientation information to move the vertices along curved paths. The path for each vertex is computed from information that is local to that vertex. These paths are easy to compute and easy to understand. Although the method is not guaranteed to produce a diffeomorphism, it does tend to produce visually satisfying interpolations that under certain circumstances have shape-preserving properties.

Methods for interpolating point-normal pairs in a plane, point-frame pairs in three dimensions, and m-rep shape models are specified in Section 5.1. Proofs of the shape maintaining properties of the two-dimensional method are given in Section 5.1.1. Interpolations produced by the methods on synthetic data in two and three dimensions and on m-rep models of a lung during the respiratory cycle are shown in Section 5.2. Evaluations that show the three-

dimensional method tends to produce interpolations that are more shape preserving than linear interpolation are discussed in Section 5.2.2. A method for computing a mean from a shape population such that the mean of a pair of shapes is the midpoint of the rotational-flows interpolation between them, is introduced in Section 5.3 and demonstrated on m-rep models of bladders. This chapter concludes with a summary of these results and a discussion of future work.

5.1 Methods

Throughout this chapter, I will use the following notation. Let $\mathbf{x} = (x, y) \in \mathbb{R}^2$ denote a point in the plane. Let $\mathbf{E} = \{\mathbf{e_1}, \mathbf{e_2}\} \in \mathbf{SO}(2)$ denote an orthogonal basis for a frame in two dimensions. The reference curve, $S[0] = \{(\mathbf{x}_{i,0}, \mathbf{E}_{i,0})\}$ is a discrete set of points, indexed by i, with each oriented by its frame. The target curve, $S[1]$ is defined similarly, and a correspondence is assumed between $(\mathbf{x}_{i,0}, \mathbf{E}_{i,0})$ and $(\mathbf{x}_{i,1}, \mathbf{E}_{i,1})$. A similar definition applies in three dimensions, with $\mathbf{x} \in \mathbb{R}^3$ and $\mathbf{E} \in \mathbf{SO}(3)$ discrete samples of a surface.

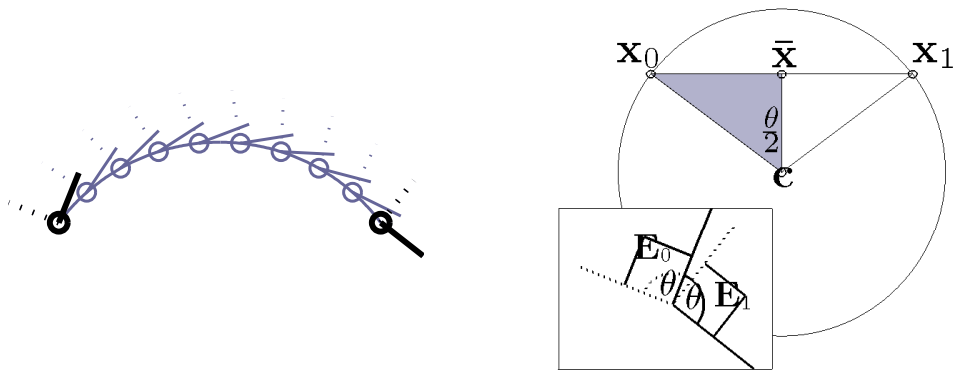
The rotational-flows interpolation for a set of oriented points indexed by i in $N \in \{2, 3\}$ dimensions will define a set of functions $f_i : (0, 1) \rightarrow \mathbb{R}^N \times \mathbf{SO}(N)$ with the property that $f_i(0) = (\mathbf{x}_{i,0},\ \mathbf{E}_{i,0})$, $f_i(1) = (\mathbf{x}_{i,1},\ \mathbf{E}_{i,1})$, and $f_i(t)$ smoothly interpolates along the time axis t.

As a notational shortcut, I will discuss a single function f which interpolates between point-frame pairs $(\mathbf{x}_0, \mathbf{E}_0)$ and $(\mathbf{x}_1, \mathbf{E}_1)$. The full interpolation is given by applying the set of f_i to the corresponding data.

5.1.1 Interpolation in two dimensions

The method of rotational flows interpolates between a pair of oriented points by sweeping out a circular arc between $\mathbf{x}_0$ and $\mathbf{x}_1$ that covers an angle equal to the angular distance between the frames $\mathbf{E}_0$ and $\mathbf{E}_1$. An example of such an interpolation can be seen in Figure 5.1(a).

Let $\bar{\mathbf{x}} = \frac{1}{2}(\mathbf{x}_0 + \mathbf{x}_1)$, $\Delta\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_0$, and let $\mathbf{x}_\perp$ be a unit vector in the direction found by rotating $\Delta\mathbf{x}$ counterclockwise by 90 degrees. The geometry of the circular path defined by the



(a) Examples of planar points: $\mathbf{x}_0$ and $\mathbf{x}_1$ and the corresponding frames: $\mathbf{E}_0$ and $\mathbf{E}_1$. The solid line indicates the $\mathbf{e}_{0,t}$ basis direction, and the dotted line indicates $\mathbf{e}_{1,t}$. The interpolation provided by my method is indicated by a lighter color. Note that the frame rotates concurrently with the point.

(b) The geometry of rotational-flows interpolation in two dimensions. The rotation required to bring $\mathbf{E}_0$ onto $\mathbf{E}_1$ (inset) defines the angle θ swept out by the interpolating path from $\mathbf{x}_0$ to $\mathbf{x}_1$. The path lies on a circle whose geometry can be derived from the positions of the points and the angle θ.

Figure 5.1: Rotational-flows interpolation in two dimensions.

method is illustrated in Figure 5.1(b). The center $\mathbf{c}$ is equidistant to $\mathbf{x}_0$ and $\mathbf{x}_1$, so it must lie on their perpendicular bisector: the line of the form $\bar{\mathbf{x}} + \alpha\mathbf{x}_\perp$, parameterized by α. The angle $\theta \in (-\pi, \pi]$ is defined as the angle between $\mathbf{e}_{0,0}$ and $\mathbf{e}_{0,1}$ with a positive sign indicating counterclockwise rotation. The method identifies a center of rotation, $\mathbf{c}$ such that the angle subtended by $\mathbf{x}_0, \mathbf{c}, \text{and } \mathbf{x}_1$ is θ. From the geometry in Figure 5.1(b), it follows that

$$\mathbf{c} = \bar{\mathbf{x}} + \frac{1}{2}\cot\frac{\theta}{2}\|\Delta\mathbf{x}\|\mathbf{x}_\perp \tag{5.1}$$

$$f(t) = (\mathbf{c} + \mathbf{R}_{t\theta}(\mathbf{x}_0 - \mathbf{c}), \mathbf{R}_{t\theta}\mathbf{E}_0) \tag{5.2}$$

where $\mathbf{R}_{t\theta}$ denotes rotation by $t\theta$ degrees.

The denominator in (5.1) is 0 when $\theta = 0$. The limiting behavior defines the interpolation in this case. When $\theta = 0$, $\mathbf{E}_0 = \mathbf{E}_1$, so it is natural to think of $f(t)$ as not having a rotational component. Considering $\lim_{\theta\to 0}$ (5.1), $\mathbf{c}$ is infinitely far from the data points, and the circular

arc connecting them is a straight line:

$$\lim_{\theta \to 0} f(t) \quad = \quad (\mathbf{x}_0 + t(\mathbf{x}_1 - \mathbf{x}_0), \mathbf{E}) \ . \tag{5.3}$$

Properties of this interpolation scheme are discussed in the remainder of this section.

Invariance to similarity transformations

Let $(\mathbf{x}_0, \mathbf{E}_0)$ and $(\mathbf{x}_1, \mathbf{E}_1)$ be a pair of oriented points whose rotational-flows interpolation is found by $f(t) = (\mathbf{x}_t, \mathbf{E}_t)$. Suppose that $\mathbf{R}$ defines a rotation, σ a uniform scaling and τ a translation. Let $g(\cdot)$ denote the rotational-flows interpolation from $(\sigma\mathbf{R}\mathbf{x}_0 + \tau, \mathbf{R}\mathbf{E}_0)$ to $(\sigma\mathbf{R}\mathbf{x}_1 + \tau, \mathbf{R}\mathbf{E}_1)$. It can be shown that

$$g(t) \quad = \quad (\sigma\mathbf{R} f(t) + \tau, \mathbf{R}\mathbf{E}_t) \tag{5.4}$$

Resumption of an interpolation

Let $f(\cdot)$ be the rotational-flows interpolation such that $f(0) = (\mathbf{x}_0, \mathbf{E}_0)$ and $f(1) = (\mathbf{x}_1, \mathbf{E}_1)$. For any $t : 0 < t < 1$ let $g_t(\cdot)$ be the rotational-flows interpolation between the interpolated oriented point $f(t)$ and $(\mathbf{x}_1, \mathbf{E}_1)$. It is a simple proof to show that $g_t(s) = f(t + s(1-t))$ and thus each interpolated point $g_t(s)$ lies on the path swept out by $f(\cdot)$. Note that $\mathbf{c}$ is equidistant to $\mathbf{x}_t$ and $\mathbf{x}_1$ so it lies on their perpendicular bisector. The angle $\angle(\mathbf{x}_t; \mathbf{c}; \mathbf{x}_1)$ is $(1-t) \cdot \theta$. This is precisely the amount of rotation needed to bring $\mathbf{E}_t$ to $\mathbf{E}_1$. This property is illustrated in Figure 5.2

Interpolation of a similar curves

Suppose that the curve $S[0]$ is similar to $S[1]$. The local interpolations defined by (5.2) will produce interpolated objects $S[t]$ that are also similar to $S[0]$.

It is an easily proven consequence of (5.3), that if $S[1]$ can be produced by a similarity transformation of $S[0]$ that has no rotational component, the rotational-flows interpolation will interpolate that transformation such that every $S[t]$ is related to $S[0]$ by a similarity

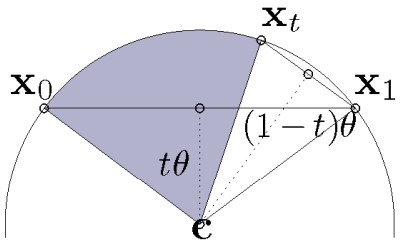


Figure 5.2: The point $\mathbf{x}_t$ is reached after interpolating t fraction of the way from $\mathbf{x}_0$ to $\mathbf{x}_1$. Interpolation from $\mathbf{x}_t$ to $\mathbf{x}_1$ continues along the original circular arc.

transformation of that form. Theorem 3 shows that this behavior also holds when $S[0]$ and $S[1]$ are related by a similarity transformation with a non-trivial rotation component.

Theorem 1 (Rotational centers for two similar sets) *Let $C = \{\mathbf{c}_i\}$ be the set of rotational centers defined in (5.1). Suppose that $S[1]$ can be produced by applying a similarity transform to $S[0]$ with the following components: rotation by an angle ϕ, uniform scaling by a factor σ, and translation by the vector $(\Delta x, \Delta y)$. When $\sigma \neq 1$, C is similar to $S[0]$. When $\sigma = 1$, C collapses to a single point.*

Let i, j index two distinct points $\mathbf{x}_{i,0}$, $\mathbf{x}_{j,0}$ in $S[0]$ and their corresponding centers of rotation $\mathbf{c}_i$, $\mathbf{c}_j$. The ratio of the distance between the centers of rotation to the distance between the points is given in (5.5) and is constant with respect to i and j. When $\sigma = 1$, $\|\mathbf{c}_i - \mathbf{c}_j\|^2 = 0, \forall i, j$, and C consists of multiple copies of a single point. When the scale factor $\sigma \neq 1$, $S[0]$ and C are similar.

$$\frac{\|\mathbf{c}_i - \mathbf{c}_j\|^2}{\|\mathbf{x}_{i,0} - \mathbf{x}_{j,0}\|^2} = \frac{(\sigma-1)^2}{4\left(\sin\frac{\phi}{2}\right)^2} \tag{5.5}$$

$$\angle(\mathbf{x}_{i,0} - \mathbf{x}_{j,0}; \mathbf{0}; \mathbf{c}_i - \mathbf{c}_j) = \frac{\pi+\phi}{2} \tag{5.6}$$

$S[0]$ can be produced by applying a similarity transformation with rotational component $-\phi$ and scale factor σ^{-1} to $S[1]$. Let C' denote the set of centers of rotation for the rotational flows interpolation from $S[1]$ to $S[0]$. The scale of C' can be shown to be equivalent to the scale of C by substitution into (5.5).

$$\begin{aligned}
\frac{\|\mathbf{c}_i - \mathbf{c}_j\|^2}{\|\mathbf{x}_{i,1} - \mathbf{x}_{j,1}\|^2} &= \frac{\left(\sigma^{-1}-1\right)^2}{4\left(\sin\frac{-\phi}{2}\right)^2} \\
&= \frac{\left(\frac{1-\sigma}{\sigma}\right)^2}{4\left(\sin\frac{\phi}{2}\right)^2} \\
&= \frac{1}{\sigma^2}\frac{(\sigma-1)^2}{4\left(\sin\frac{\phi}{2}\right)^2} \\
&= \frac{\|\mathbf{x}_{i,0} - \mathbf{x}_{j,0}\|^2}{\|\mathbf{x}_{i,1} - \mathbf{x}_{j,1}\|^2} \cdot \frac{\|\mathbf{c}_i - \mathbf{c}_j\|^2}{\|\mathbf{x}_{i,0} - \mathbf{x}_{j,0}\|^2}
\end{aligned} \tag{5.7, 5.8}$$

Similarly, the orientation of C' can be shown to be the same as the orientation of C by substitution into (5.6).

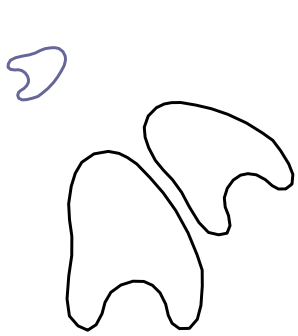
$$\begin{aligned}
\angle(\mathbf{x}_{i,1} - \mathbf{x}_{j,1}; \mathbf{0}; \mathbf{c}_i - \mathbf{c}_j) &= \frac{\pi-\phi}{2} \\
&= \frac{\pi+\phi}{2} - \phi \\
&= \angle(\mathbf{x}_{i,0} - \mathbf{x}_{j,0}; \mathbf{0}; \mathbf{c}_i - \mathbf{c}_j) \\
&\quad -\angle(\mathbf{x}_{i,0} - \mathbf{x}_{j,0}; \mathbf{0}; \mathbf{x}_{i,1} - \mathbf{x}_{j,1})
\end{aligned} \tag{5.9}$$

$$\angle(\mathbf{x}_{i,0} - \mathbf{x}_{j,0}; \mathbf{0}; \mathbf{x}_{i,1} - \mathbf{x}_{j,1}) + \angle(\mathbf{x}_{i,1} - \mathbf{x}_{j,1}; \mathbf{0}; \mathbf{c}_i - \mathbf{c}_j) = \angle(\mathbf{x}_{i,0} - \mathbf{x}_{j,0}; \mathbf{0}; \mathbf{c}_i - \mathbf{c}_j) \tag{5.10}$$

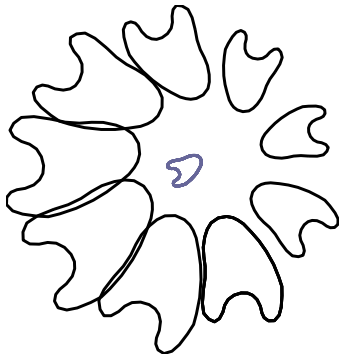
Assume without loss of generality that $\sigma > 1$. An explicit formula for $\mathbf{c}_i$ is given in (5.11), where $\sigma' = (\sigma-1) / \left(2\sin\frac{\phi}{2}\right)$ and $\phi' = (\pi+\phi)/2$.

$$\begin{aligned}\mathbf{c}_i &= \sigma' \mathbf{R}_{\phi'} \mathbf{x}_{i,0} \\ &\quad + \frac{1}{2}\left(\Delta x\,(\sigma+1) + \Delta y(\sigma-1)\cot\frac{\phi}{2}, \Delta y\,(\sigma+1) - \Delta x\left(\sigma-1)\cot\frac{\phi}{2}\right)\right) \quad (5.11)\end{aligned}$$

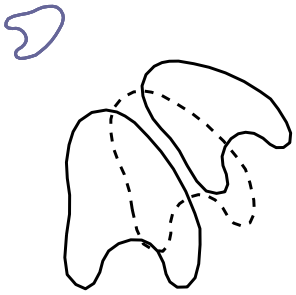
Figure 5.3(a) shows an example of a pair of similar shapes and the centers of rotation identified by my method. Here the curve $S\,[0]$ was taken from a manual contour of a bladder on its mid-axial slice in CT. $S\,[1]$ was produced by applying a known similarity transform to $S\,[0]$.



(a) $S\,[0]$ and $S\,[1]$ (in black) and their centers of rotation C (in purple).



(b) A set of curves (black) produced by rotating the points in $S\,[0]$ about the corresponding point in C (purple).



(c) The interpolated curve $S\,[0.5]$ (dashed line) for this example.

Figure 5.3: The locus of centers of rotation implied by a pair of similar objects is itself similar to those objects.

When $\phi = 0$, the denominator of (5.5) is 0. This is consistent with (5.3). Because there is no rotational component to such a similarity transform, the method provides a straight-line interpolation which can be interpreted as rotation about points that are infinitely far away from the original data and from each other.

Theorem 2 (Rotation about a similar set) *Let $\{\mu_\cdot\}$ denote a set of points in $\mathbb{R}^2$. Let $\{\nu_\cdot\}$ denote a set of similar points such that for each i, ν_i is formed by rotating μ_i by an angle of ϕ about the origin, scaling the result by σ, and then translating by $(\Delta x, \Delta y)$. Now let μ_i^ψ be found by rotating μ_i an angle of ψ about ν_i. The scaling factor from $\{\mu_\cdot\}$ to $\left\{\mu_\cdot^\psi\right\}$ is given by (5.12).*

$$\begin{aligned} \frac{\|\mu_i^\psi - \mu_j^\psi\|^2}{\|\mu_i - \mu_j\|^2} &= 1 + 2\sigma^2 - \\ & \quad 2\sigma\left(\cos(\phi) - \cos(\phi - \psi) + \sigma\cos(\psi)\right) \end{aligned} \tag{5.12}$$

Equation (5.12) is a general statement: the new shape formed by rotating one shape about another similar shape, in the manner defined above, is also similar to the original shape. When σ is 0, (5.12) simplifies to a well known truth: rotating an object about a fixed point is a rigid transformation.

Figure 5.3(b) shows the result of various rotations of a curve about a similar curve.

Theorem 3 (Rotational flows between similar objects) *Suppose again that $S[0]$ and $S[1]$ are related by a similarity transform with rotational component ϕ, scaling factor σ, and an arbitrary translational factor. Any $S[t]$ is similar to $S[0]$. The square of scaling factor from $S[0]$ to $S[t]$ is given by (5.13).*

$$\begin{aligned} \frac{\|\mathbf{x}_{i,t} - \mathbf{x}_{j,t}\|^2}{\|\mathbf{x}_{i,0} - \mathbf{x}_{j,0}\|^2} &= 1 + \frac{(\sigma-1)^2}{2\left(\sin\frac{\phi}{2}\right)^2} \\ & \quad -\frac{\sigma-1}{\sin\frac{\phi}{2}}\left(\sin\left(\frac{\phi}{2} - t\phi\right) - \sin\left(\frac{\phi}{2}\right) + \frac{\sigma-1}{2\sin\frac{\phi}{2}}(\cos t\phi)\right) \end{aligned} \tag{5.13}$$

Equation (5.13) follows from substituting (5.5, 5.6) into (5.12) and from (5.2) which states that $\angle(\mathbf{x}_{i,0}; \mathbf{c}_i; \mathbf{x}_{i,t}) = t\phi$. The rotational-flows interpolation between two similar curves is shape maintaining. An example of this property can be seen in Figure 5.3(c).

The special case of (5.13) when $\sigma = 1$ shows that the rotational-flows interpolation between two congruent curves is shape and size maintaining. When $\sigma > 1$, the derivative of (5.13) with respect to t is strictly positive over the interval $t \in [0, 1]$. Likewise, when $0 < \sigma < 1$, $\frac{d}{dt}$ (5.13) is strictly negative over that interval. Although (5.12) allows the set of points to collapse onto a single point, the monotonicity of scale when (5.13) is restricted to $t \in [0, 1]$ guarantees that the interpolation between similar, but non-trivial, shapes will not pass through the degenerate configuration.

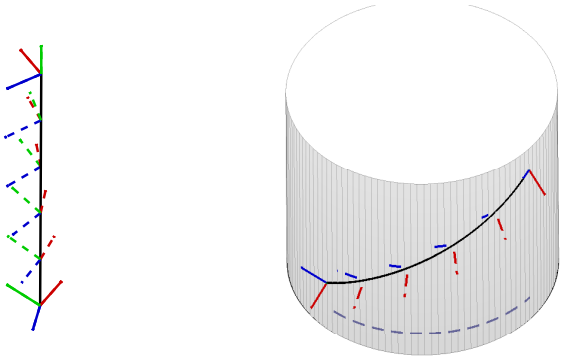
5.1.2 Interpolation in three dimensions

Let $(\mathbf{x}_0, \mathbf{E}_0)$ and $(\mathbf{x}_1, \mathbf{E}_1)$ be corresponding oriented points in $\mathbb{R}^3$. There exists a unit vector $\mathbf{v}$ that defines the axis of rotation from $\mathbf{E}_0$ to $\mathbf{E}_1$. Let the magnitude of this rotation be denoted by $0 \leq \phi \leq \pi$. The three dimensional rotational-flows interpolation between these points is designed to rotate $\mathbf{x}_0$ by ϕ about the axis $\mathbf{v}$.

Since $\mathbf{x}_0$ and $\mathbf{x}_1$ may lie in different planes that are orthogonal to $\mathbf{v}$, the desired rotational path between them may not exist. Projecting $\mathbf{x}_0$ and $\mathbf{x}_1$ along $\mathbf{v}$ into a common plane allows a path to be found by the two dimensional rotational-flows algorithm (5.2). This path is combined with translation along $\mathbf{v}$. Let $v_0 = (\mathbf{x}_0 \cdot \mathbf{v})$ and $\mathbf{x'}_0 = \mathbf{x}_0 - v_0\mathbf{v}$. Let v_1 and $\mathbf{x'}_1$ be defined in a similar fashion. Let $f_2(t) = (\mathbf{x'}_t, \cdot)$ be the two-dimensional rotational-flows interpolation between these projected points. The three-dimensional rotational-flows interpolation is given by (5.14) and is illustrated in Fig. 5.4.

$$f(t) \quad = \quad \left(\mathbf{x'}_t + (v_0 + t(v_1 - v_0))\,\mathbf{v}, \mathbf{R}_{\mathbf{v},t\phi}\mathbf{E}_0\right) \tag{5.14}$$

This method can be understood as a generalization of Rossignac's method for rigid body interpolation [71]. Each sample $(\mathbf{x}, \mathbf{E})$ is analogous to an independent rigid body. An interpolated surface $S[t]$ is produced by allowing each sample to follow its own screw-like path.



(a) Rotation of the frame $\mathbf{E}_0$ (top) into $\mathbf{E}_1$ (bottom

(b) The rotational flows interpolation from $(\mathbf{x}_0, \mathbf{E}_0)$ (left) to $(\mathbf{x}_1, \mathbf{E}_1)$ (*right*) combines the two dimensional rotational flows interpolations of projected points $\mathbf{x'}_0$ to $\mathbf{x'}_1$ with translation along the axis of rotation.

Figure 5.4: Three-dimensional rotational-flows interpolation. $\mathbf{x}_t$ lies on a helix whose axis is the axis of rotation between $\mathbf{E}_0$ and $\mathbf{E}_1$.

When the surfaces $S[0]$ and $S[1]$ are similar, each $S[t]$ produced by rotational-flows interpolation is related to $S[0]$ by a transformation that is linear with respect to the vertex positions in homogeneous coordinates. The interpolated shapes are not necessarily similar to $S[0]$ and $S[1]$ since scaling within the $\mathbf{v} = 0$ plane per (5.13) and scaling due to the translation along the the $\mathbf{v}$ axis occur with different scale factors. However, in Section 5.2.2 I argue that this interpolation is more shape preserving than a straight-line interpolation.

5.1.3 Interpolation of m-rep shape models

The discrete m-rep [65] provides a sampled medial representation for a shape in three dimensions. I refer the reader to the summaries of medial geometry in Section 2.4 and m-reps in Section 2.5.3. This representation has been used for a variety of applications in medical image analysis including shape modeling [36, 57], image segmentation [64], and statistical shape analysis [32].

Each shape instance (referred to as an object or an m-rep for convenience) consists of a 2-D lattice of medial atoms, indexed by two parameters: u and v. Each medial atom in $\mathcal{M}(1)$ is composed of a hub position: $\underline{X}(u,v)$, a radius: $r(u,v)$, and two spoke vectors: $\underline{U}^{+1}(u,v)$

and $\underline{U}^{-1}(u,v)$. The medial atoms on the edges of the lattice have an additional additional parameter $\eta(u,v)$ that controls the length of an implied spoke that is in the direction of $\left(\underline{U}^{+1}(u,v) \; + \; \underline{U}^{-1}(u,v)\right)$.

Fletcher [26] defined a metric on $\mathcal{M}(1)$ (2.91), but it measures changes in hub position independently from changes in spoke orientation. Consequently the geodesics under this metric are characterized by linear hub paths. Interpolation of medial atoms along such a geodesic produces a straight-line interpolation of the hubs.

I will now describe rotational-flows interpolation between corresponding medial atoms (*i.e.,* those with the same (u,v) coordinates) on different instances of the same shape. As a notational shortcut the (u,v) indices will be omitted. The orientation of the medial axis at $\underline{X}$ is used to define the following frame that is used to interpolate the hubs using the method described in Section 5.1.2.

The $\mathbf{e}_1$ basis is chosen to be in the direction of $\underline{U}^{+1} - \underline{U}^{-1}$ because the difference between the spoke vectors is known to be normal to the medial axis. The $\mathbf{e}_2$ basis is in the direction of $\partial \underline{X}/\partial u - \left(\partial \underline{X}/\partial u \cdot \mathbf{e}_1\right)\mathbf{e}_1$. Because the directional derivative $\partial \underline{X}/\partial u$ is estimated from the discretely sampled hub positions, the Gram-Schmidt process [90] is used to ensure $\mathbf{e}_2$ truly lies in the plane orthogonal to $\mathbf{e}_1$. The remaining basis is defined by $\mathbf{e}_3 = \mathbf{e}_1 \times \mathbf{e}_2$.

The radius and elongation of the m-rep are interpolated geometrically: $r_t = r_0 \left(r_1/r_0\right)^t$ and $\eta_t = \eta_0 \left(\eta_1/\eta_0\right)^t$. The spoke vectors are interpolated geodesically after accounting for rotation of the hubs. Let $\underline{U}'^{+1}$ represent the $\underline{U}^{+1}$ spoke in frame relative coordinates:

$$\underline{U}'^{+1} = \left(\underline{U}^{+1} \cdot \mathbf{e}_1, \underline{U}^{+1} \cdot \mathbf{e}_2, \underline{U}^{+1} \cdot \mathbf{e}_3\right)^T.$$

Let $\underline{U}'^{+1}_0$ and $\underline{U}'^{+1}_1$ represent those vectors at times 0 and 1 respectively. The axis $\mathbf{w}$ and angle ϕ are defined as:

$$\phi = \cos^{-1}\left(\underline{U}'^{+1}_0 \cdot \underline{U}'^{+1}_1\right) \tag{5.15}$$

$$\mathbf{w} = \left(\underline{U}'^{+1}_0 \times \underline{U}'^{+1}_1\right) \tag{5.16}$$

The spoke is interpolated geodesically in frame coordinates,

$$\underline{U}'^{+1}_{t} = \mathbf{R}_{\mathbf{w},t\phi}\underline{U}'^{+1}_{0} \tag{5.17}$$

$$\underline{U}^{+1}_{t} = \sum_{i=1}^{3}\left(\underline{U}'^{+1}_{t}\cdot\mathbf{e}_{i,t}\right)\mathbf{e}_{i,t} \tag{5.18}$$

The other spoke, $\underline{U}^{-1}$ is interpolated in the same way. One desirable property of this interpolation is that it preserves the normality of the spoke difference to the medial sheet. This difference is collinear with $\mathbf{e}_1$, the direction normal to the interpolated medial axis.

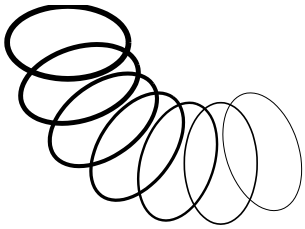
5.2 Results

5.2.1 Interpolations of planar curves

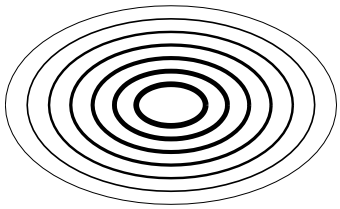
Rotational-flows interpolations of planar curves are shown in Figures 5.5 - 5.7. In each of these examples, $S[0]$ is drawn with the thinnest line. The line thickens at each time step with $S[1]$ drawn with the heaviest line.

Figure 5.5 shows rotational-flows interpolations between pairs of similar ellipses. Figure 5.5(a) shows the interpolation between congruent ellipses. Their size and shape are maintained during the interpolation. The reference and target curves shown in Figure 5.5(b) differ only in scale. None of the interpolated shapes have been rotated or translated. The reference and target curves in 5.5(c) differ by a full similarity transformation. Note that the centers of rotation are similar to the original and interpolated shapes. Figure 5.5(d) shows the interpolation between congruent ellipses after their correspondence has deliberately been corrupted. The interpolated shapes are no longer ellipses.

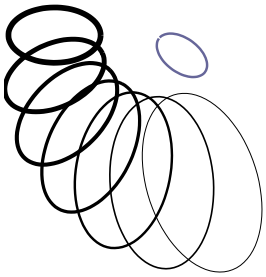
Figure 5.6 shows rotational-flows interpolation between shapes that are not similar. The interpolation in Figure 5.6(a) is between two ellipses with different eccentricity. In Figure 5.6(b) the interpolation is between an ellipse and a bent curve. In these sequences the interpolated curves are visually satisfying. Figure 5.6(c) shows an indentation forming during the interpolation between ellipses. This problem is related to the correspondences between $S[0]$ and $S[1]$. Figure 5.6(d) shows the resulting interpolation after $S[1]$ has been resampled to



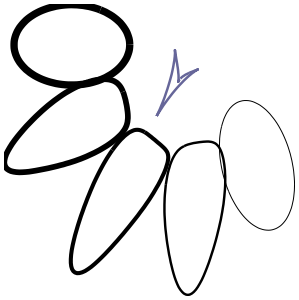
(a) Between congruent ellipses



(b) Between ellipses that vary only in scale



(c) Between similar ellipses

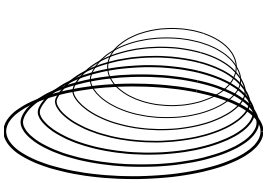


(d) Between congruent ellipses, but with poor correspondence

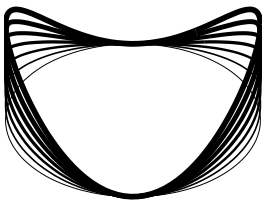
Figure 5.5: Rotational-flows interpolation between similar shapes in 2D. In each example the time axis is denoted by line thickness.

produce better correspondences, as will be discussed below.

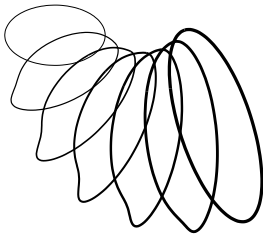
Figure 5.7 shows an interpolation where a concave region becomes convex. The curves $S[0]$ and $S[1]$ were formed by applying a bending transformation to an ellipse. Figure 5.7(a) is unsatisfactory because the interpolated shape $S[0.5]$ has two protrusions at each end: one from the downward bend and one from the upward bend. Figure 5.7(b) shows the rotational-flows interpolation using the standard ellipse as a key frame. This use of an intermediate key frame is a well known technique in the graphics literature. Another corrective technique from the graphics literature is to prealign shapes prior to the morph. It has been my experience that a translational alignment between shapes can lead to a more satisfying interpolation. In Figure 5.7(c), the shapes have been prealigned and the interpolation is forced to pass through the standard ellipse. The method is sensitive to the correspondences between the oriented points



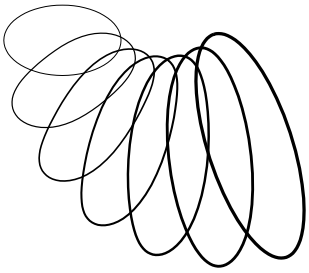
(a) Between translated and non-uniformly scaled ellipses



(b) Between unbent and bent curves



(c) Between affinely related ellipses. Here the interpolations deviate from the desired shape space of ellipses



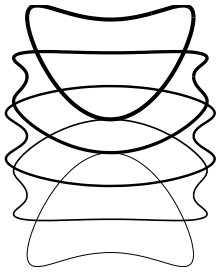
(d) Correspondences from Figure 5.6(c) have been redefined to produce a more satisfying interpolation

Figure 5.6: Rotational-flows interpolation between non-similar shapes in 2D

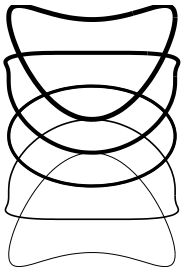
in $S[0]$ and $S[1]$. Figure 5.7(d) shows rotational-flows interpolation on the original curves, after correspondences have been improved so that the curves are understood as rotated copies of each other.

In each set of examples (Figures 5.5 - 5.7) the correspondences between samples of the reference and target curves affected the quality of the interpolation.

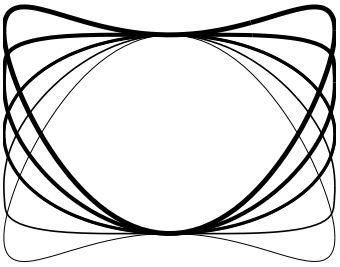
Figures 5.5(a) and 5.5(d) show rotational flows interpolations between curves with equivalent boundaries. In Figure 5.5(a), the source curve $S[0]$ is of the form $\mathbf{R}_\theta(\alpha\cos\phi, \beta\sin\phi)$ and the reference curve $S[1]$ is of the form $(\alpha\cos\phi, \beta\sin\phi)$, where α, β, and θ are fixed, and samples are taken at matching ϕ values. The $\mathbf{e}_1$ basis vector at each sample is chosen to be the outward normal to each curve. Thus, the angle swept out by each rotational-flows path is exactly θ. In Figure 5.5(d), the reference curve is modified to be of the form $(\alpha\cos(\phi+\Delta\phi), \beta\sin(\phi+\Delta\phi))$.



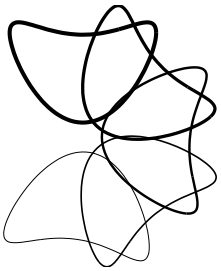
(a) The original interpolation, note the pair of protrusions at each end of $S[0.5]$.



(b) The interpolation has been forced to pass through an ellipse at $t = 0.5$. A single protrusion still exists at the ends of $S[0.25]$ and $S[0.75]$.



(c) The shapes have been prealigned and $S[0.5]$ is forced to be the ellipse. Each $S[t]$ appears to be a bent version of the ellipse.



(d) Correspondences have been redefined so that the interpolation is a shape-preserving rotation.

Figure 5.7: Interpolation between shapes where a concave region becomes convex.

The $\mathbf{e}_1$ basis is assigned to be the outward normal to this reparameterized curve, and consequently the samples near the vertex of each ellipse are interpolated along paths that sweep out larger angles than the paths of samples from lower curvature regions. The change in correspondence and the associated variability in the frame rotations are responsible for the degradation in interpolation quality from Figure 5.5(a) to Figure 5.5(d).

Figures 5.6(c) and 5.6(d) show another example where a change in correspondences affects rotational-flows interpolation. The reference curve in Figure 5.6(c) is of the form $(\alpha_0 \cos\phi, \beta_0 \sin\phi)$ and the target curve is of the form $\mathbf{R}_\theta\,(\alpha_1 \cos\phi, \beta_1 \sin\phi)$. For both curves, the $\mathbf{e}_1$ basis is chosen to be the outward normal at each sample. α_0, β_0, θ, α_1, and β_1 are all held constant, and correspondence between samples is established via matching values of ϕ. Because the

eccentricity of the ellipse $S[0]$ differs from that of $S[1]$, the rotational change from $\mathbf{e}_{1,0}$ to $\mathbf{e}_{1,1}$ varies with ϕ. The same reference curve samples are used in Figure 5.6(d); however, the target curve is reparameterized as $\mathbf{R}_\theta(\alpha_1 \cos\phi_1, \beta_1 \sin\phi_1)$. Corresponding samples are taken so that

$$(\beta_0 \cos\phi, \alpha_0 \sin\phi) = (\beta_1 \cos\phi_1, \alpha_1 \sin\phi_1) \tag{5.19}$$

Under this sampling scheme, the rotational change from $\mathbf{e}_{1,0}$ to $\mathbf{e}_{1,1}$ is exactly θ for each pair of corresponding points. As was the case with the previous example, establishing correspondences that minimize the variance in the frame rotations allows that method of rotational flows to produce a more satisfying interpolation.

Figure 5.7 shows a variety of interpolations between a pair of bent curves. The correspondences in Figures 5.7(a) - 5.7(c) were established through equally spaced samples according to polar coordinates. The interpolations shown in Figures 5.7(b) and 5.7(c) improve on the one shown in Figure 5.7(a), by explicitly specifying the shape of $S[0.5]$ and by translationally aligning the reference and target curves. These improvements bring the interpolation closer to an ideal interpolation that would "unbend" the reference curve into an ellipse and then continue to bend it upward into the target curve. Figure 5.7(d) shows a different interpolation between these curves produced after the target curve has been resampled to be understood as a rotated copy of the reference curve. As was the case with the examples shown in Figures 5.5(a) and 5.6(d), the rotational change at all pairs of corresponding points is constant, and the resulting interpolation is visually satisfying.

When the rotational changes between the reference and target curves are constant across the samples, the method of rotational flows has, in the cases I have studied, produced pleasing interpolations. An area of future work that I will discuss in Chapter 6 is to study further the relationship between changes in the local rotations between nearby points, and to develop a theory that can predict the quality of the interpolation, given these changes. However, it should be noted that frames could be trivially constructed so that there is no change in orientation at any vertex. In this case, rotational-flows interpolation would produce the exact same paths as linear interpolation. Thus it appears that merely minimizing the variation in the

rotational changes is not sufficient to ensure the quality of the interpolation. The magnitude of the rotational changes will likely need to be commensurate with the actual rotation between objects.

A similar piece of future work would study if modifying the method so that the angle swept out by each circular arc corresponds to a (weighted) average rotational change would improve the quality of the interpolations produced. Although it should be noted that such a modification would sacrifice the property that the current method can be evaluated independently for each pair of corresponding vertices.

5.2.2 Shape preservation during three-dimensional interpolation

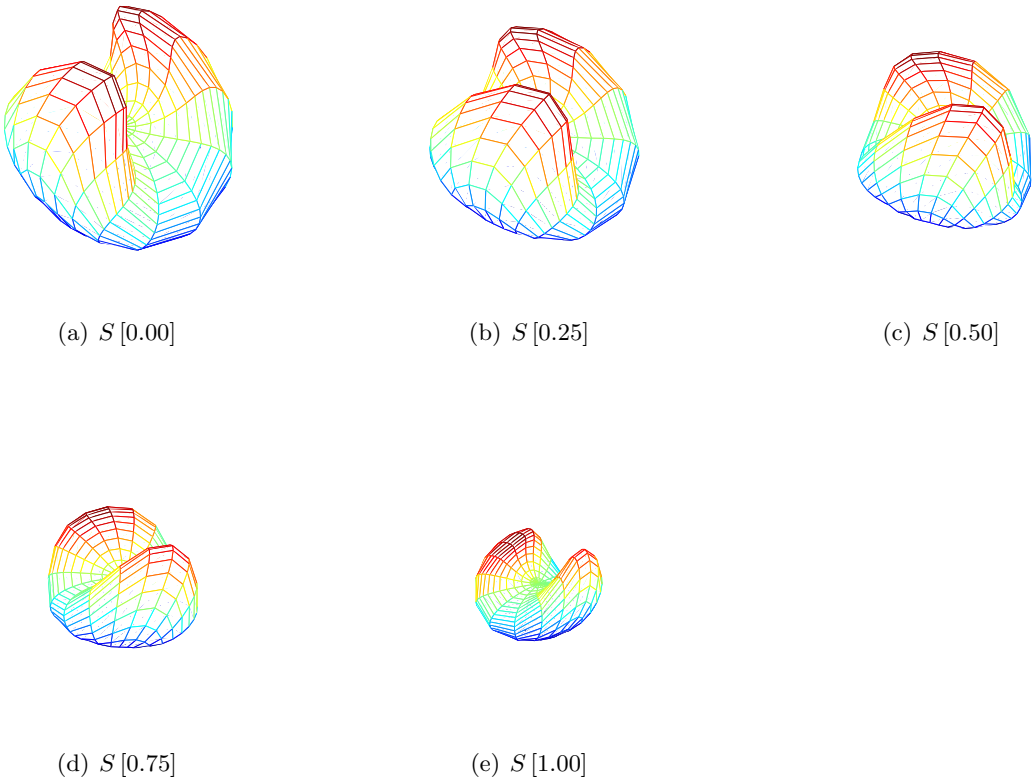
Figure 5.8 shows the rotational-flows interpolation between two similar surfaces. As predicted, $S[0]$ is not similar to $S[t]$ for almost all $0 < t < 1$. However, the interpolated $S[t]$'s produced by rotational flows are more like $S[0]$ than the intermediate shapes produced by a linear interpolation.

Let $S[0]$ and $S[1]$ be similar shapes, and let $\mathbf{M}$ be the 4×4 matrix such that when the points in $S[0]$ and $S[1]$ are expressed in homogeneous coordinates, $S[1] = \mathbf{M}S[0]$. Let $\mathbf{M}_3$ denote the upper 3×3 block of $\mathbf{M}$. I use the fractional anisotropy (FA) [4] of $\mathbf{M}_3$, to measure the degree to which the transformation preserves shape. Let, $\lambda_1 \geq \lambda_2 \geq \lambda_3$ be the square roots of the eigenvalues of $\mathbf{M}_3\mathbf{M}_3{}^T$. Let $\bar{\lambda} = \frac{1}{3}\sum_{i=1}^{3} \lambda_i$. FA is defined as

$$fa(\lambda_1, \lambda_2, \lambda_3) = \sqrt{\frac{\frac{1}{2}\sum_i \left(\lambda_i - \bar{\lambda}\right)^2}{\frac{1}{3}\sum_i \lambda_i^2}} \tag{5.20}$$

and can be understood as the standard deviation of the eigenmodes of the transformation divided by the root mean square eigenmode. The value of $fa(\cdot)$ ranges from 0 for a shape-preserving transformation to 1 for an affine transformation that collapses the shape onto a single axis.

Using $S[0]$ and $S[1]$ as defined in Figure 5.8, I compared the fractional anisotropy of the transformation from $S[0]$ to $S[t]$ as t varied during straight-line and rotational-flows interpolations. Figure 5.9 shows that the rotational-flows interpolation between these similar shapes



(a) $S\,[0.00]$ (b) $S\,[0.25]$ (c) $S\,[0.50]$

(d) $S\,[0.75]$ (e) $S\,[1.00]$

Figure 5.8: Rotational-flows interpolation in three dimensions between similar shapes.

is more shape preserving than straight-line interpolation.

To further understand the performance of rotational-flows interpolation versus straight-line interpolation, I ran the following experiments. Let $\mathcal{T}$ denote the similarity transformation such that $\mathcal{T}\,[S\,[0]] = S\,[1]$ in the previous example. I generated a family of similarity transformations $\mathcal{T}_\phi$ by setting the angle of rotation in the transformation to be ϕ while holding the other parameters constant. I then generated new target surfaces $S_\phi\,[1] = \mathcal{T}_\phi\,[S\,[0]]$. For each ϕ I measured the maximum FA of the transformation from $S\,[0]$ to $S_\phi\,[t]$ over $t \in [0,1]$ for rotational-flows and straight-line interpolations. These FA values are plotted in Figure 5.10(a).

I performed a similar experiment in which only the scale parameter of the similarity transformation between $S\,[0]$ and $S\,[1]$ varied, and the other similarity transformation parameters were held constant. The results of these experiments are plotted in Figure 5.10(b).

In both experiments, the rotational-flows interpolation was more shape preserving than

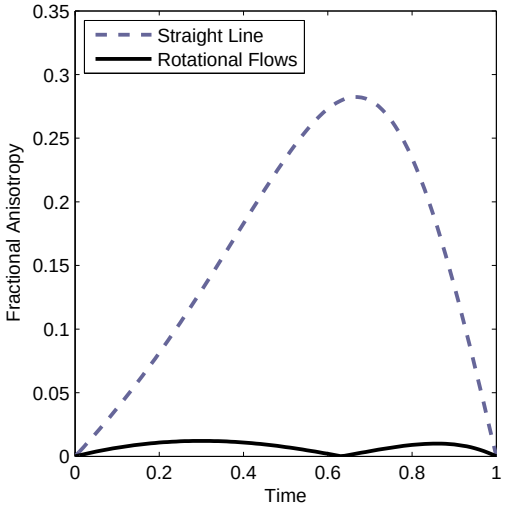


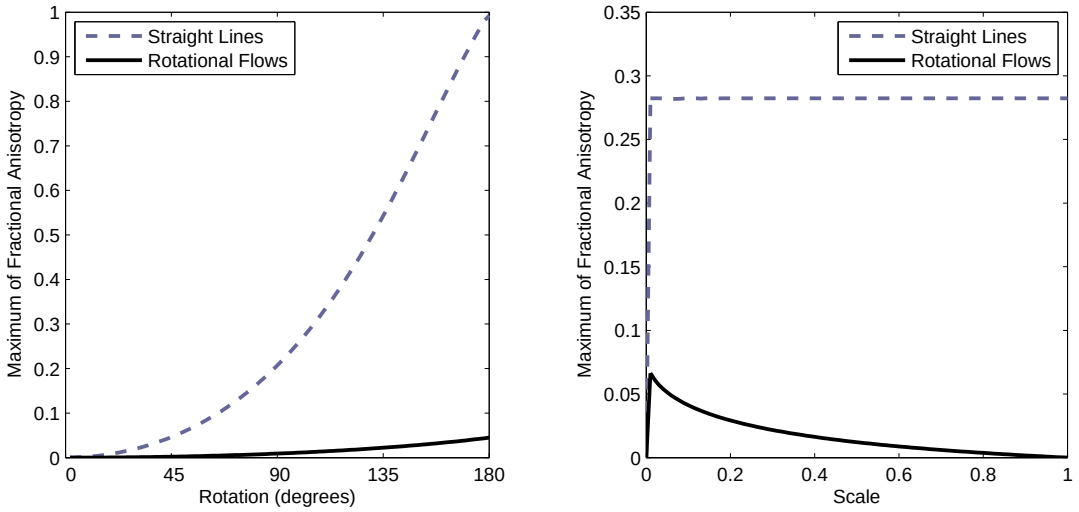
Figure 5.9: Fractional anisotropy of the linear transformations from $S[0]$ to $S[t]$ for shapes produced by the linear and rotational-flows methods interpolations of the reference and target curves shown in Figure 5.8.

straight-line interpolations. Moreover, the level of fractional anisotropy was quite low for all scales and rotation angles were considered.

In the general case, when $S[0]$ and $S[1]$ are not similar, the three-dimensional rotational-flows interpolation frequently produces intermediate shapes that appear reasonable. An example of such is shown in Figure 5.11. Although there is no linear transformation from $S[0]$ to $S[t]$ in this example, the fractional anisotropy of the least squares estimates of such a transformation has its maximum at $t = 1$. As is shown in Figure 5.12, this compares favorably with interpolation by linear paths. In that case, there is a spurious maximum of fractional anisotropy for an interpolated shape.

5.2.3 An example using m-reps

M-rep models were fit to a time series of lung segmentations using the method of Han and Merck [36, 57]. Figure 5.13 shows the rotational-flows interpolation from the model that corresponds with peak inspiration to the model that corresponds with peak expiration. Informally, the interpolated surfaces exhibit the behavior we expect from a lung during exhalation. The lung is decreasing in volume, and the most visible motion is where the inferior portion



(a) As angle of rotation between $S[0]$ and $S[1]$ varies, $\max_{0<t<1}$ FA of the transformation from $S[0]$ to $S[t]$.

(b) As the scale factor between $S[0]$ and $S[1]$ varies, $\max_{0<t<1}$ FA of the transformation from $S[0]$ to $S[t]$.

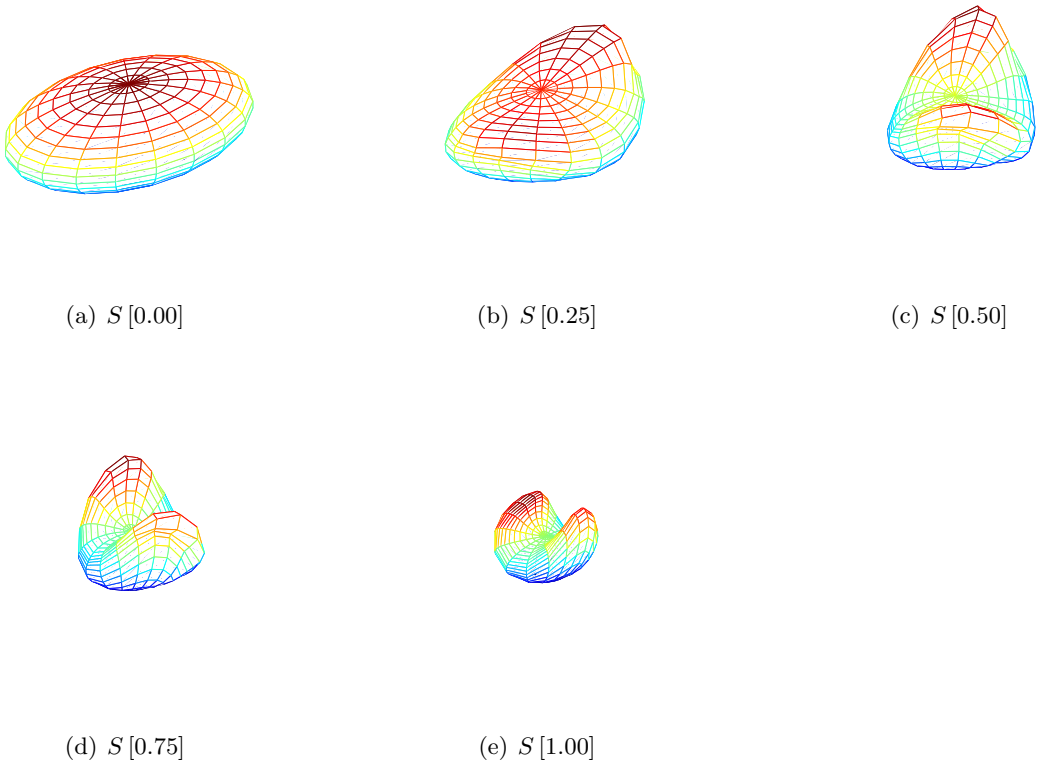
Figure 5.10: Comparing fractional anisotropy for transformations produced by rotational flows with those produced by straight-line interpolation.

rises as the diaphragm relaxes. However, in this example, the intermediate lungs produced by interpolating each medial atom along a geodesic in $\mathcal{M}(1)$ exhibit the same behavior.

5.3 Computation of a Mean

Rotational-flows interpolation provides a way to understand the differences between a pair of objects. In this section, I discuss preliminary work that uses the interpolation method to construct a mean that summarizes a population of objects. This mean is defined such that when it is constructed for a pair of objects, it is equivalent to $S[0.5]$: the midpoint of the rotational-flows interpolation between the objects.

One possible way to construct such a mean is to define a metric on $\mathbb{R}^N \times \mathbf{SO}(N)$ that has the property that the geodesics under the metric are equivalent to the paths produced by rotational-flows interpolation. Such a metric could be used in the Fréchet framework to produce a mean. Because of the way the rotational-flows method couples information about position and orientation, it is my opinion that such a metric will be difficult to invent, but I



(a) $S\,[0.00]$ (b) $S\,[0.25]$ (c) $S\,[0.50]$

(d) $S\,[0.75]$ (e) $S\,[1.00]$

Figure 5.11: Rotational-flows interpolation in three dimensions between non-similar shapes.

will discuss the possiblity further in Chapter 6.

Without such a metric the Fréchet mean cannot be computed. Instead, I propose a method for computing a mean shape that is analogous to a method for computing the arithmetic mean of a set of numbers. Let $\{x_1, ..., x_k\} \subset \mathbb{R}$ be a set of k numbers. Their arithmetic mean $\bar{x}$ is found by the familiar formula:

$$\bar{x} \;=\; \frac{1}{k}\sum_{i=1}^{k} x_i\,. \tag{5.21}$$

Basic algebraic manipulations allow (5.21) to be rewritten as follows:

$$\bar{x} \;=\; \frac{1}{k-1}\sum_{j=2}^{k}\left(\frac{x_1}{k} + \frac{(k-1)\,x_j}{k}\right) \tag{5.22}$$

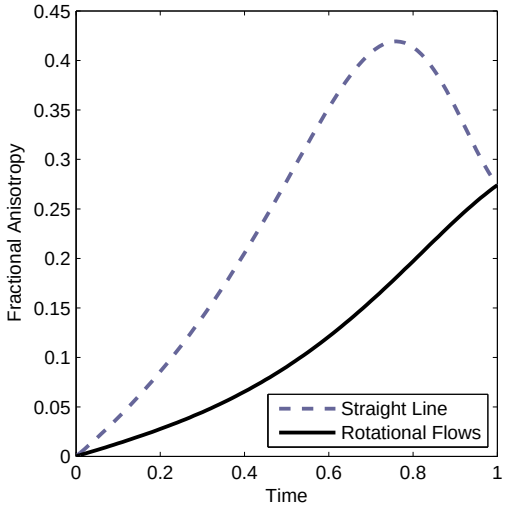
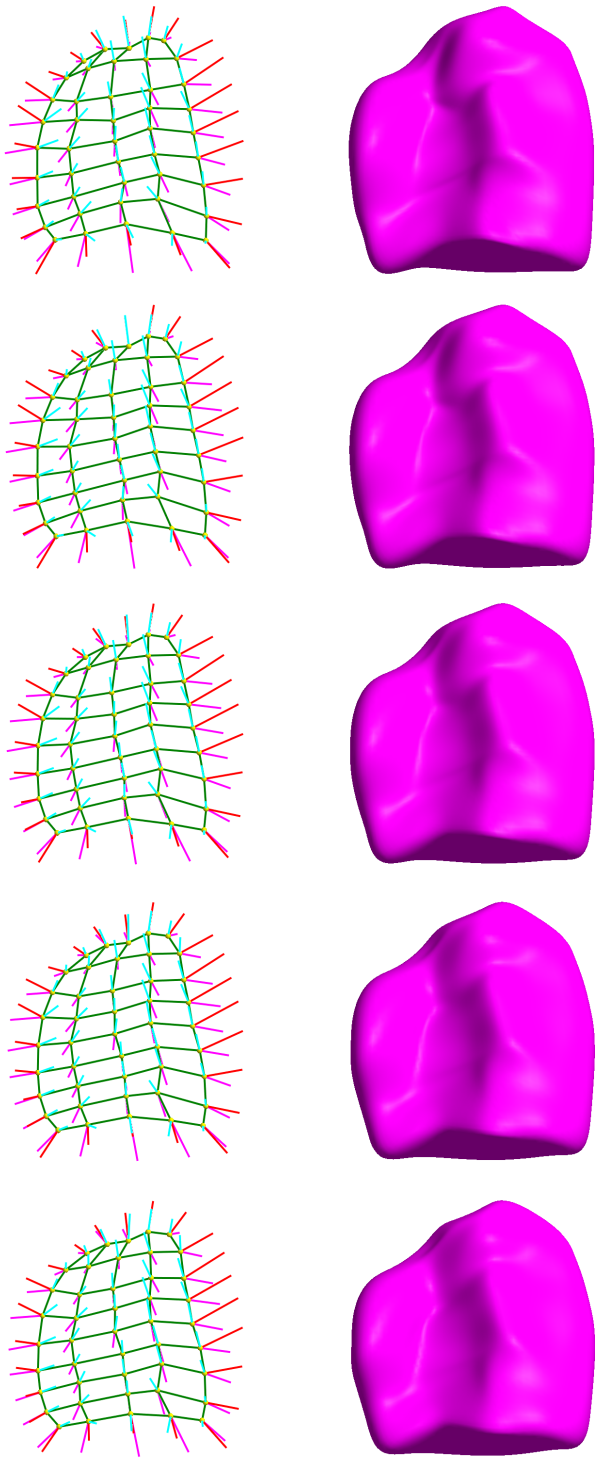


Figure 5.12: Fractional anisotropy of the linear transformations from $S[0]$ to $S[t]$ for shapes produced by the linear and rotational-flows methods interpolations of the reference and target curves shown in Figure 5.11.

Equation (5.22) states that the arithmetic mean of the k originial numbers is equivalent to the arithmetic mean of $k-1$ convex combinations, of the form $(x_1/k + x_j(k-1)/k)$, from the original set. Each such convex combination of numbers can be understood a linear interpolation $S[(k-1)/k]$, where $S[0] = x_1$ and $S[1] = x_j$. It is in a similar manner that I compute a mean shape by producing convex combinations of shapes through rotational-flows interpolations.

Starting with a set of k sampled, oriented shapes, I designate a single special instance (x_1) that is chosen to be $S[0]$ for $k-1$ interpolations. Each of the other shapes serves as $S[1]$ for one such interpolation. The set $\{S[(k-1)/k]\}$ produced by these interpolations is a $(k-1)$-element set that has the same mean as the original set. Reapplying the method to this set then produces a $(k-2)$-element set which also has the same mean. After $k-1$ iterations, and the computation of $\left(n^2-n\right)/2$ convex combinations, the method will converge to a single shape that is treated as the mean.

This procedure is illustrated in Figure 5.14. The original data, shown in the upper left pane, consists of m-rep models of the bladder from five patients. Each convex combination shown in the subsequent panes was produced via the interpolation method described in Section 5.1.3.



(a) Medial atoms of the m-rep (b) Implied surface boundary

Figure 5.13: Rotational-flows interpolation of m-rep lung models. $S\,[0]$ (Top) corresponds with peak inspiration. $S\,[1]$ (Bottom) corresponds with peak expiration.

Bladder	Relative Volume
1	0.847
2	0.954
3	1.098
4	0.992
5	1.136

Table 5.1: Relative volumes of the five bladders shown in Figure 5.14. The geometric mean of these volumes is 1.000.

This method for computing a mean is not precisely equivalent to (5.22). Because addition is associative and commutative, (5.22) produces the same answer regardless of the ordering of the data. This is not the case for rotational-flows interpolation. However, when the orientation of the samples are close together, *i.e.,* all instances of the same basis vector lie in a hemisphere, the $\mathbf{SO}(N)$ component of rotational-flows interpolation behaves like addition and is robust against the ordering of the data.

Recalling that the rotational-flows interpolation of a pair of points with equivalent frames is merely the linear interpolation of the points, the following procedure can be used to produce a mean shape that is independent of the ordering of the input shapes. First I compute the mean via convex combination for all possible orderings of the data, *i.e.,* $k!$ times. Then, assuming any frame orientation is stable across all $k!$ means, I find the mean of the $k!$ means by linearly averaging the vertex positions across the $k!$ means and combining those positions with the stably estimated frames.

An example of such a mean is shown in Figure 5.15. When compared to the Fréchet mean bladder produced using the Riemannian metric on medial atoms and the mean bladder produced using the Euclidean metric on PDM representations of the bladders, the mean produced by convex combinations of rotational flows had a volume that was more consistent with the input data. This can be seen in Tables 5.1 and 5.2.

5.4 Discussion

I have proposed a novel method for shape interpolation that moves oriented points along easily computed and understood rotational paths that combine changes in position and ori-

Mean Bladder	Relative Volume
Convex Combination of Rotational Flows	1.010
Fréchet via Riemannian Metric on M-reps	0.952
Euclidean on PDMs	0.904

Table 5.2: Means of the five bladders shown in Figure 5.14 were computed via the following three methods: 1) via convex combination using rotational-flows, 2) in the Fréchet framework using a Riemannian metric on the space of m-reps, and 3) by converting each of the original m-reps into a PDM, and then computing the average of each landmark position (using the Euclidean metric). The relative volume of the mean produced using rotational-flows interpolations (shown in Figure 5.15) is closest to 1.000; it is the most consistent with the volumes of the input data (see Table 5.1).

entation. This interpolation offers superior shape maintenance when compared with linear interpolation of the points without regard for their orientation. I have demonstrated this method with interpolations of synthetic data sets in two and three dimensions and with interpolations of medial shape models derived from patient data.

The method cannot be guaranteed to produce diffeomorphic transformations. Future work includes developing a better understanding of conditions on the target and reference objects that lead to two interpolated points occupying the same position at some intermediate time. Similarly I need to develop a method to automatically modify the interpolation to overcome such a failure. Once these failures and their recovery are better understood, I will be able to explore extrapolating the paths from our interpolation to deform an entire image volume.

A preliminary version of this extrapolation has been done using thin-plate splines [33, 6] in the large deformation diffeomorphism framework [41]. Diffeomorphisms are closed under composition: given two diffeomorphism, $\phi_1 : \mathbb{R}^N \rightarrow \mathbb{R}^N$ and $\phi_2 : \mathbb{R}^N \rightarrow \mathbb{R}^N$, $\phi_1 \circ \phi_2$ is also a diffeomorphism. Moreover, when the change in landmark positions is sufficiently small, thin-plate splines will produce a diffeomorphism. Thus, the idea is to use rotational-flows interpolation to determine the position of boundary points at s time-steps. Thin-plates splines will extrapolate the change from each $S[j/s]$ to $S[(j+1)/s] : j \in \{0..(s-1)\}$ to produce a set of transformations $\{\phi_j\}$. Assuming the time-steps are adequately small, each of these transformations can be assumed to be diffeomorphic. The composition of these transformations: $\prod_{j=0}^{s-1} \phi_j$ can then be assumed to be a diffeomorphic transformation that extrapolates the change from $S[0]$ to $S[1]$ to the entire volume.

The method described above uses rotational-flows interpolation rather than geodesic interpolation based on a Riemmanian metric, as was originally proposed by Joshi and Miller [41], to determine landmark positions. That method is guaranteed to produce a diffeomorphic transformation of the space, but it may not produce a shape maintaining interpolation. Figure 5.16 is a sketch of case where the method based on the Riemannian metric, with initial paths determined by gradients of a binary map, is expected to produce an interpolation that is not shape maintaining.

A second line of future work involves computing statistics from a population of objects, based on the interpolating transformations between them. Section 5.3 defined a mean based on convex combinations that were implemented as rotational-flows interpolations. However, for a set of n objects, $O\left(n^2\right)$ interpolations are needed to compute a mean, and $O\left(n!\right)$ interpolations are needed to compute a mean that is not biased by the ordering of the input data.

The computational complexity of computing each (biased) mean via convex combinations can easily be reduced. The arithmetic mean of k numbers (5.21) can be reformulated as (5.23), as is used for online averaging.

$$\bar{x} = \frac{k-1}{k}\left(\frac{1}{k-1}\sum_{j=1}^{k-1} x_j\right) + \frac{1}{k}x_k \tag{5.23}$$

This formulation of the mean requires only $k-1$ convex combinations to be taken. x_1 is treated as an initial estimate of the mean. This estimate is refined by combination with x_2; that estimate is further refined by combination with x_3 and so on.

However using this formulation to compute a mean shape does not address the larger problem: the bias introduced by the ordering of the data and the $O\left(k!\right)$ algorithm to remove that bias. Future work includes reducing that computational complexity, as well as defining higher order statistics such as variance. Both of those tasks may well depend on defining a metric on $\mathbb{R}^N \times \mathbf{SO}\left(N\right)$ that can be used in the Fréchet framework.

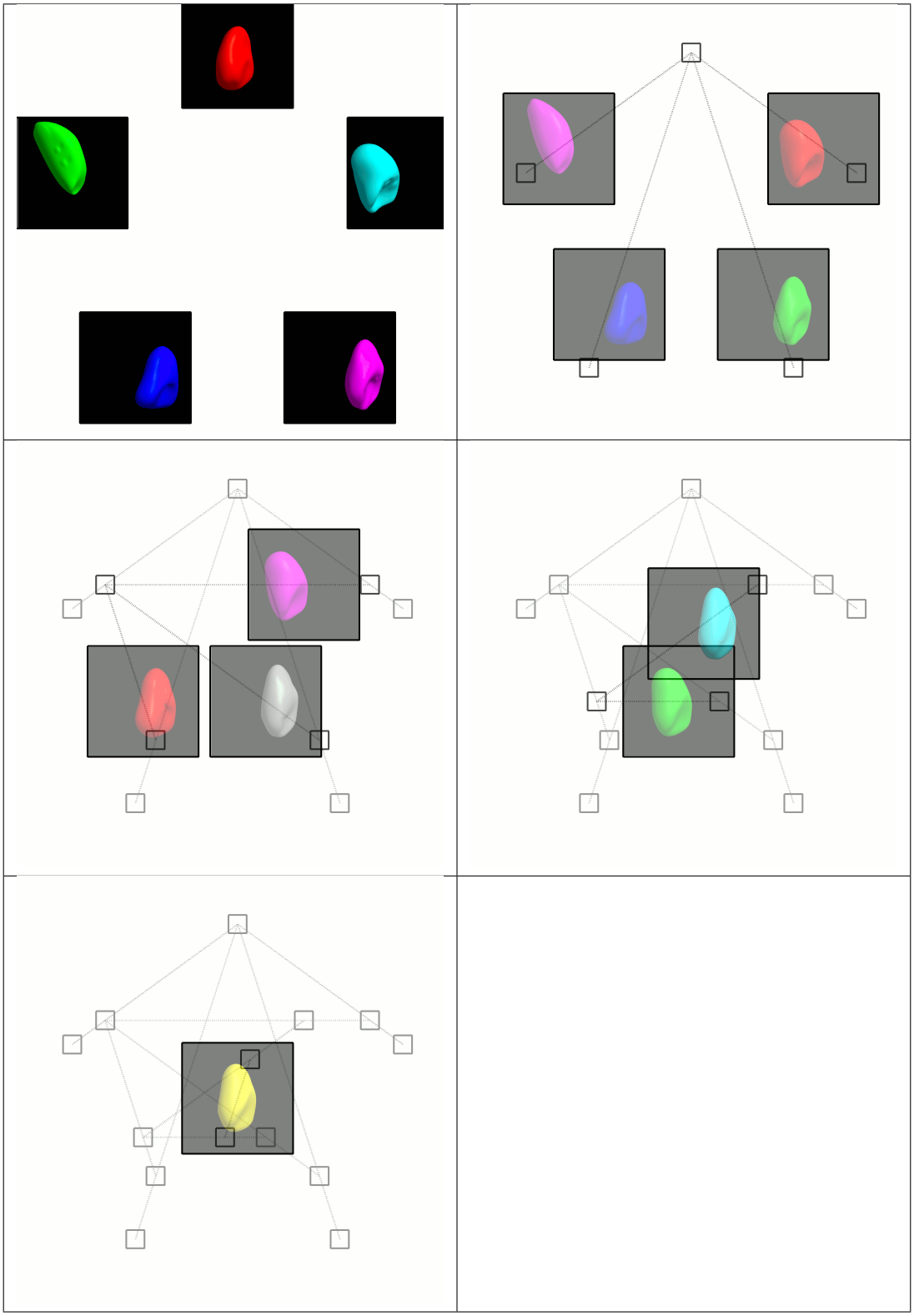


Figure 5.14: Computation of a mean bladder via convex combinations produced by rotational-flows interpolations.

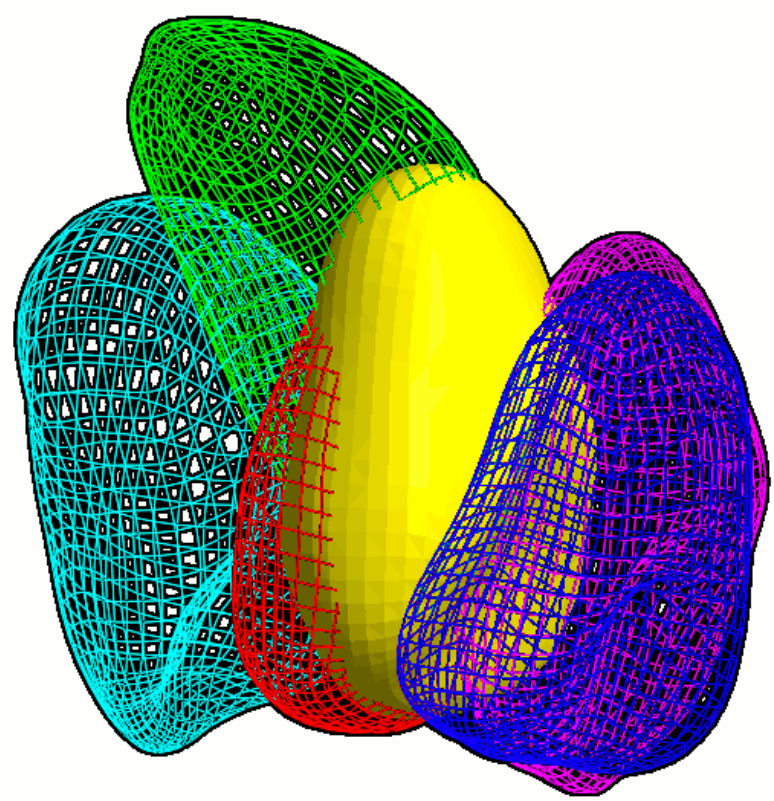


Figure 5.15: The mean bladder (yellow, solid) produced from the data set shown in Figure 5.14 is shown relative to the input bladder m-reps (colored, meshed)

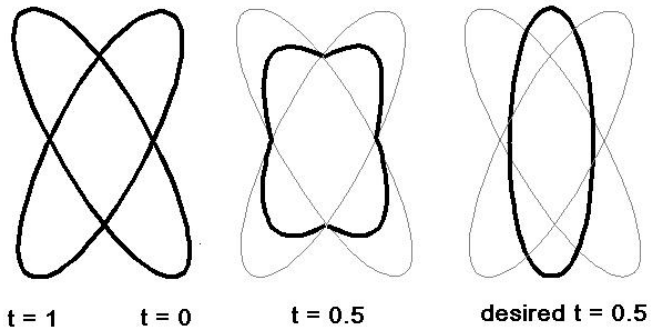


Figure 5.16: A sketch of what a Riemannian interpolation between ellipses at different orientations may look like when the initial landmark paths are determined by gradients of the binary image. The method described in [41] is believed to produce a bone-shaped object halfway between the two ellipses. An ellipse at a different orientation is the desired interpolated object and would be produced by the method of rotational flows.

Chapter 6

Discussion and Conclusions

In this chapter, I revisit the contributions of this dissertation, and I discuss avenues of future research.

6.1 Summary of Contributions

1. *Evidence that given a shape prior with adequately high generalizability and specificity, a dense object boundary can be recovered from a sparse sampling of boundary points.*

   Bladder SDSMs with high generalization ability and specificity were described in Sections 3.3 and 3.4. I used these SDSMs to drive the optimization formulated as Equations 3.14, 3.15, and 3.17 in order to recover a dense object boundary from samples on just three axial image slices. The average and worst case surface distance measurements from contour-based initializations to reference segmentations of the same images that were reported in Sections 3.3.1 and 3.4.1 show that each object boundary was indeed effectively recovered from its sparse set of boundary samples.

   Bladder initialization is just one example of how a strong shape prior can be used to recover a description of an object from a relatively small number of samples on its boundary. In unpublished work on a test-bed of synthetic shapes constructed by applying three independent transformations of random magnitude to an ellipsoid, I found that a reasonable approximation of a shape could frequently be found in a PGA shape space by minimizing the τ-band distance (eqn. 3.10) to a set $\mathcal{P}$ of five randomly selected points on

the boundary of the shape. That the shape could be recovered from such a small $\mathcal{P}$ is not surprising given that there were only three degrees of freedom in the shape population. In another piece of unpublished work, my colleague Edward Chaney produced qualitatively good approximations of the shapes and poses of prostates by minimizing the τ-band distance to a set $\mathcal{P}$ that contained four boundary points on each three designated axial image slices.

Because the variation seen in many populations of anatomical objects cannot be described so simply, further work is needed to develop a theory that allows a user to understand the optimal way to define $\mathcal{P}$ given a shape prior. Indeed, attempting to define "optimal" in this context could be the subject of future work. In a simple task where the entire boundary of the target object is known *a priori*, the optimality of $\mathcal{P}$ is likely to be a tradeoff between the cardinality of the set and the quality of the recovered shape. However, when the boundary of the target object is unknown, the difficulty of locating each element of $\mathcal{P}$ must also be considered.

2. *Strategies for how a user can simply produce a boundary sampling that allows recovery of a shape model for the bladder from easily identifiable points found on relatively few CT image slices.*

   The τ-band distance (eqn. 3.10) discussed in Section 3.2 can be used to create loose correspondences between points known to be on the boundary of the target object and points on the boundary of an SDSM as it deforms. The results given in Sections 3.3.1 and 3.4.1 were produced by minimizing the τ-band distance from points $\mathcal{P}$ that lie on three designated axial slices: the penultimate inferior and superior slices and the mid-axial slice on which the bladder lies to the SDSM boundary.

   In Section 3.5, I discuss a study in which only the subset of those points that lie on strong within-slice edges of the image were used. These points could be more easily identified and specified by the user, and this truncation of the input set did not significantly affect initialization performance. My colleague Edward Chaney has performed experiments where $\mathcal{P}$ contains boundary points from the mid-axial, mid-coronal, and mid-sagittal

slices through the bladder [15], rather than from the three axial slices previously used. Qualitatively, he found that those initialization models were frequently good. His study will be revisited in Section 6.2.

The use of bladder shapes recovered via such an optimization to initialize a segmentation process is discussed with Contribution 3. As was mentioned in the discussion of Contribution 1, a strategy for defining the optimal $\mathcal{P}$ is the subject of future work. However, the definition of $\mathcal{P}$ that I have used in the experiments discussed in Chapter 3 has been effective for recovering bladder shapes and has been seen to be qualitatively effective for recovering prostate shapes as well.

3. *Applications of a bladder shape model recovered from boundary points on three axial image slices to initialize or refine a semi-automatic segmentation.*

   When segmentation is performed via minimization of a non-convex cost function, there is a risk that the process will fail if the optimization stops at a local minimum of the cost function that produces an unacceptable segmentation. In the context of m-rep-based bladder segmentation, the m-rep instance that minimizes the τ-band distance to manually selected points chosen from the boundary of the bladder in a target image is typically close enough to the desired segmentation to mitigate this risk. This can be seen in the results given in Sections 3.3.1 and 3.4.1.

   The initialization m-reps that were produced by minimizing the τ-band distance to manually specified points on three axial slices of the image were shown to be close to manually produced reference segmentations. In subsequent m-rep-based segmentations, the results of which were also reported in Sections 3.3.1 and 3.4.1, the non-convex image-match function never led the optimizer to an unreasonable answer. In outlier cases where the initialization m-rep had significant disagreement with the reference segmentation, the image-driven segmentation tended to correct those errors. In the frequent cases where the initialization m-rep agreed with the reference segmentation, the image-driven segmentation also agreed with the reference segmentation. That is, the optimization was not drawn to an undesirable minimum of the cost function.

Minimization of the τ-band distance from an m-rep to a manually specified point set $\mathcal{P}$ can be used to produce semi-automatic segmentations in the following ways. As was discussed in Section 3.5, allowing the user to rerun the optimization after augmenting $\mathcal{P}$ with additional points in regions where the initialization m-rep and the reference segmentation disagree is a semi-automatic segmentation method that does not explicitly process image intensities. Edward Chaney has encountered prostate segmentation scenarios where a poorly estimated probability distribution of RIQFs allows the segmentation to be drawn away from a reasonable initialization [15]. In these cases he has found that adding the τ-band distance from a candidate m-rep to $\mathcal{P}$ as an additional term in the segmentation cost function prevents the segmentation from wandering too far off. This additional penalty term turns the automatic m-rep-based segmentation algorithm into a semi-automatic one. As before, allowing the user to run this process multiple times and to edit $\mathcal{P}$ between iterations is likely to be an effective semi-automatic segmentation method.

4. *A novel tool that identifies statistical outliers of a localized geometry-to-image-match function as regions where a segmentation is not credible.*

   In Chapter 4, I described a statistical test that can be applied to a regional image match function relative to a segmentation of a medical image. Assuming that the image match can be used as a proxy for the distance to the boundary of the unknown true segmentation of the image, the proposed segmentation is likely to be non-credible in regions where the image match takes on an improbably poor value.

   In Section 4.1, this test is fully specified for a segmentation that can be represented as an m-rep and for a local image match function based on RIQFs in regions anchored by m-rep spoke ends. The non-credible regions can be visualized according to a hard threshold on the image match value, as in Figures 4.2 and 4.3 or with a color scheme that indicates the degree of non-credibility, as in Figure 4.5. These visualizations allow the user to allocate their validation resources to the regions of the segmentation that are most likely to need further review.

Allowing the user to prioritize their segmentation validation efforts addresses a problem where technological advancements might otherwise be lost to a human bottleneck. Improvements in CT scanner technology are allowing images to be acquired with smaller slice thickness and interslice spacing than in previous generations, and this trend is likely to continue. Consequently, the number of image slices in which a fixed-size target object appears has been increasing from generation to generation. This is increasing the cost of manual, slice by slice validation of the segmentation of an image. There is a risk that automatic segmentations of high quality, high resolution CT images would not be clinically useful because of this validation cost. The method I have discussed allows users to plan their validation efforts so that the amount of time allocated to each region (or slice) reflects the likelihood of a segmentation error having occurred in that region.

5. *A novel method for shape interpolation that synchronizes local changes of position and orientation. This method is driven by "rotational flows" about "rotational centers"*

   I introduce rotational-flows interpolation in Chapter 5. The two-dimensional method is defined in Section 5.1.1. This interpolation causes an oriented point to follow a path along a circular arc from its position at time 0 to its position at time 1. The angle swept out by this arc is equivalent to the change in orientation from time 0 to time 1. In this sense the interpolation can be understood as rotating the point about its rotational center.

   A three-dimensional variation of the method is defined in Section 5.1.2. This interpolation causes a point-frame pair to follow a helical path from its position at time 0 to its position at time 1. This helix is a geodesic curve on the surface of the cylinder whose axis is the axis of the smallest rotation between the frames at times 0 and 1 and whose radius is defined by the two-dimensional rotational-flows interpolation between the projections of the points into a designated plane.

   The three-dimensional method is further extended in Section 5.1.3 so that it can be applied to m-rep models of anatomical structures.

   The two-dimensional method has been proven to be shape maintaining when interpo-

lating between geometrically similar objects. Moreover, when it interpolates between congruent objects, it is both shape and size maintaining. In the more general case, it tends to produce visually satisfying interpolations between curves.

It is frequently the case that when the three-dimensional method is applied to similar objects, it produces an intermediate shape that is not similar to them. However, in experiments described in Section 5.2.2 interpolations produced by this method were shown to be more shape preserving than interpolations on geodesics of Riemannian metrics on point-tuples, *i.e.,* generalized linear interpolations.

A serious weakness of linear interpolation is that the interpolation between unaligned data, in particular between objects with fairly similar shape but with large differences in orientation, may produce degenerate intermediate objects that essentially collapse into a single point. The interpolation methods based on rotational flows are more tolerant of changes in orientation and are unlikely to fail in such cases.

Rotational-flows interpolation is likely to have widespread applicability enabling users to understand populations of objects. Applications of this method include visualization of shape changes, due to population differences or a process internal to a single subject, summarizing a population of objects, as will be discussed under Contribution 6, and the long-term goal of the research that led to this dissertation: producing landmark paths so that deformation between objects in an atlas and target image can be extrapolated to produce a deformation from the atlas volume to the entire target image.

My colleague Rohit Saboo has developed a tool that produces and composes thin-plate spline interpolations of landmark paths [73]. This tool can be used to create the aforementioned warp from an atlas to a target image. Such a warp is capable of transferring the segmentations of other targets near the atlas objects to the target image.

Rohit Saboo has generated landmarks paths for a set of structures in the head and neck based on geodesic interpolation of m-rep models, and he has applied his tool to extrapolate those paths to produce a warp from an atlas image to a target image. He and I have seen that when rotational-flows interpolation is used to produce landmark

paths for the same objects, his tool produces a comparable, yet slightly different, warp. In these particular target structures, the geodesic interpolation is reasonable. However, if this method were applied to structures that exhibit greater changes in orientation, the geodesic interpolation would be less satisfactory. In such a case, the benefits of using rotational-flows interpolation to produce the landmark paths would be more apparent.

6. *A method for computing the mean of a shape population based on pairwise interpolations via the method of rotational flows.*

   A generalization of the calculation of an arithmetic mean that can be understood as a series of convex combinations of the data is presented in Section 5.3. To compute this mean for a shape population, I implement convex combination as a rotational-flows interpolation between a pair of shapes. This method was demonstrated on a population of five m-rep bladder models. The size of the mean computed by this method is more consistent with the population than the size of means produced via the Fréchet formulation using a metric on medial atoms and via an arithmetic mean of PDM landmark coordinates.

   A drawback to this method is that it is expensive to compute. Given a population of N shapes, the method requires $O\left(N^2\right)$ convex combinations to produce a mean. Such a mean will be biased according to the ordering of the shapes in the population. A method for combining $O\left(N!\right)$ such means into an unbiased estimate exists, but it is only suitable for small N. In Section 5.4, I suggested an improved method that constructs a biased mean from $O\left(N\right)$ convex combinations. Producing an unbiased estimate would still require the computation of $O\left(N!\right)$ biased means.

   As I noted in the discussion of Contribution 5, the method of rotational flows is more tolerant of changes in orientation than is the method of linear interpolation. Similarly, the method for computing a mean that I have proposed is likely to produce a reasonable output if the population of input shapes is unaligned, whereas another mean, *i.e.,* one based on a Riemannian metric, produces a degenerate mean.

   This allows a variety of analyses to be applied to unaligned shape data. Well known

algorithms that rely on the computation of a mean, such as k-means clustering [55], PCA, and PGA, just to name three, could be reimplemented to use this mean. These modified algorithms might then produce an insightful analysis of unaligned shape data for which the previous options were either to pre-align the data, discarding information by doing so, or to accept a failed analysis of the unaligned data. In many cases implementation of these modified algorithms would benefit from the invention of a distance metric such that the paths produced by rotational-flows interpolation are geodesics under the metric. This metric does not yet exist, and its design is an area for future work.

6.1.1 Thesis statement revisited

Thesis: *Atlas-based segmentation of a dense region consisting of multiple objects and the spaces between them can be performed effectively given high quality segmentations of the objects in the target space and a means to transfer those segmentations to the entire space. Initializing an object-based segmentation by matching the object to a sparse set of boundary contours can produce results that are appropriate for this purpose. Estimating the credibility of the segmentation by detecting statistical outliers of the image match in local regions relative to the object allows the clinician to be confident in the segmentation results. Rotational-flows interpolations from atlas object instances to credible segmentations produce a set of landmark paths that can be used in the large-scale diffeomorphism framework to map the full set of labels on the atlas to the target image.*

There are medical problems for which object-based segmentation methods produce relevant solutions. These problems are often characterized by one or more segmentation targets in situations where both the geometry and image intensity patterns of each target object can be modeled. The benefits of using an object-based segmentation method typically include improvements in the reproducibility of segmentations when compared to manually produced segmentations, the ability for a medical expert who would otherwise need to produce a manual segmentation to refocus his time and attention to other aspects of patient care, and the ability to process large and/or time-sensitive image sets for which manual segmentation may be infeasible.

As the technologies used for object-based segmentation have matured, previously intractable problems in medical image analysis have received the attention of researchers. One such problem, referred to in my thesis statement, is *segmentation of a dense region consisting of multiple objects and the spaces between them.* The challenge in this problem is that the geometry and image intensities of targets in the interstitial spaces cannot necessarily be modeled.

I have proposed a multistage solution to this problem. The first stage of the solution requires the user to produce *high quality segmentations* of a subset of the targets, and the second stage is for the user validate those segmentations. The third stage is *to transfer* the object segmentations *to the entire space.*

In my thesis statement, I refer to the need to produce accurate object-based segmentations in the following sentence: *Initializing an object-based segmentation by matching the object to a sparse set of boundary contours can produce results that are appropriate for this purpose.* Contributions 1 - 3 are related to my method for initializing SDSM-based segmentation by minimizing the τ-band distance from the SDSM to relatively few manually specified points on the object boundary. This method has produced high quality bladder segmentation initializations that were themselves comparable in quality to bladder segmentation results reported in the literature and that led to segmentations with less risk due to local minima problems than was associated with the previous initialization method.

I address the problem of validating such segmentations in the following sentence: *Estimating the credibility of the segmentation by detecting statistical outliers of the image match in local regions relative to the object allows the clinician to be confident in the segmentation results.* Contribution 4 is a test that identifies segmentation regions where a local image match function takes on an improbably poor value and a method for visualizing the locations of such regions. If a segmentation has a localized failure, that failure is likely to have occurred in such a region. I advise users with limited time to allocate to segmentation evaluation to focus their validation efforts in suspected non-credible regions.

Given the τ-band distance for segmentation initialization and the test of regional segmentation credibility, a user can produce and validate high quality object-based bladder segmentations. With some application-specific engineering work, I believe the initialization method

described in Chapter 3 could be applied to many other targets of object-based segmentation.

After producing and validating high quality segmentations of target objects, the user is ready to extrapolate the changes from atlas to segmented objects in order to understand changes to targets in the interstitial spaces. *Rotational-flows interpolations from atlas object instances to credible segmentations produce a set of landmark paths that can be used in the large-scale diffeomorphism framework to map the full set of labels on the atlas to the target image.* Contribution 5 is a method for interpolating from an atlas object to the segmentation of that object in a target image. Such an interpolation produces a set of landmark paths that describe the deformations of that target object. The combined set of landmark paths that described the deformations from the atlas to the target image can be extrapolated to produce a deformation from the atlas volume to the target image volume that preserves the segmentations of the previously segmented objects. This deformation can then be used to bring the atlas segmentations of the other nearby structures of interest into the target volume.

6.2 Future Work

The work I have presented in Chapters 3 - 5 and Section 6.1 provides opportunities for future research. Those opportunities are revisited in this section.

6.2.1 Semiautomatic Segmentation Initialization via a Sparse Set of Contours

In Chapter 3, I introduced a method for initializing the object-based segmentation of the bladder from a CT image. The method initializes the segmentation from the bladder instance that minimizes its τ-band distance to a set of manually specified points $\mathcal{P}$ on the boundary of the bladder in the penultimate-inferior, penultimate-superior, and mid-axial image slices. This method has been demonstrated on m-rep-based SDSMs that model the day to day shape variation of a patient undergoing ART.

The specification of $\mathcal{P}$ is a topic for future work. Requiring the user to contour the three designated image slices imposes far less of a burden than requiring a full manual segmentation.

This choice of $\mathcal{P}$ has worked well in practice, but there is no theoretical proof of its optimality. The definition of an optimal $\mathcal{P}$ needed to recover the shape of an object is the subject of future research. Indeed, the definition of optimality is open for debate, as this parameter selection is likely to be a multicriterion optimization.

The computational complexity of the initialization is one such criterion; it scales with $|\mathcal{P}|$. A lower bound on the cardinality of the size of the set is likely to be the number of degrees of freedom in the shape space used for initialization. However, because shape spaces are often non-linear and because there may be interactions between parameters, a larger number of points may be needed.

The user-burden, measured as the wall-time that elapses while the user is specifying the points in $\mathcal{P}$, is another criterion to be considered. This burden will likely depend on image intensity patterns as well as $|\mathcal{P}|$. $\mathcal{P}$ should ideally consist of points that the user can easily find.

Another criterion to consider is the desired level of agreement between the initialization object and the unknown segmentation of the target image. The initialization is intended to be good enough to mitigate the risk of segmentation failing by producing a locally optimal, but unacceptable result. However, given two possible initializations, both of which cause the segmentation process to produce the desired result, one that minimizes both the user-burden and the computational cost should be preferred even if the other is closer to the desired answer.

After an optimality criterion for $\mathcal{P}$ has been defined, the process of defining user guidelines for finding an optimal $\mathcal{P}$ is still a challenging problem. Exhaustively searching the powerset of a dense set of boundary points would be prohibitively expensive. Thus, a heuristic for designing $\mathcal{P}$ is likely needed.

Because keeping $|\mathcal{P}|$ small is a desirable, each specified boundary point should provide a lot of information about the target object but not be strongly correlated with the other specified boundary points. Recall that the method of Cates *et al.*[13] produces corresponding landmarks for a population of objects such that the position of each landmark is is tightly distributed across the population but spread out with respect to the particular object instance. One can imagine applying that method either to a set of training shapes, or to a set of shapes

generated from an SDSM, and then defining $\mathcal{P}$ to be those landmark positions. A study of the agreement of those training shapes with approximations that minimize the τ-band distance to $\mathcal{P}$ produced by Cates's method, as a function of $|\mathcal{P}|$ would help to illustrate the tradeoff between the simplicity of $\mathcal{P}$ and accuracy of the recovered shape. Such a study would be useful for learning a lower bound on $|\mathcal{P}|$ for a target object, but the actual landmark positions assigned by Cates's method might be discarded since it is difficult for a user to map those 3D points into a target image.

A theory of how to define $\mathcal{P}$ would simplify the challenge of extending this method to other target objects or to other SDSMs. Because such a theory does not yet exist, *ad hoc* methods have been applied to these tasks. In an unpublished study on the rectum, my colleagues and I found that defining $\mathcal{P}$ as the penultimate-inferior, penultimate-superior, and mid-axial image slices did not capture global shape information about the rectum, but adding samples from the mid-sagittal image slice was helpful. Edward Chaney [15] has also been interested in exploring other uses for this method. He has found that the penultimate-inferior, penultimate-superior, and mid-axial image slices typically contain enough information about the global shape of the prostate to allow my method to be applied to that gland. He has also found that a drastically reduced $\mathcal{P}$, consisting of four well spaced samples from each of those contours is also effective.

In experiments on bladder initialization using an SDSM that captures between-patient shape variation, Chaney [15] has found that defining $\mathcal{P}$ to contain samples from the mid-axial, mid-sagittal, and mid-coronal slices has been effective. In those experiments, he uses a τ-band distance where τ is defined to be 1, *i.e.,* any point on the candidate object surface can map to any point in $\mathcal{P}$. In the near future, I plan to develop a generalization of τ-band distance that is more appropriate for data, such as this, that is not aligned in axial contours.

Another area for future research that I identified in Chapter 3 regards the use of an initialization model at segmentation time. Recall the following notation that was used in Section 3.2. The initialization model $\underline{\underline{m}}^{\text{init}}$ exists in the coordinate system of the target image. $\underline{\underline{\mu}}^{\text{init}}$ is a shape model that is geometrically similar to $\underline{\underline{m}}^{\text{init}}$ and that was found via PGA deformation of a mean model $\underline{\underline{\mu}}$. That is, $\underline{\underline{\mu}}^{\text{init}}$ was produced via the Riemmanian map $\text{Exp}_{\underline{\underline{\mu}}}$ of some vector in a truncated PCA subspace of the tangent hyper-plane to the m-rep manifold at the

mean model $\left(T_{\underline{\underline{\mu}}}\mathcal{M}(n)\right)$, and $\underline{\underline{m}}^{\text{init}} = \mathcal{T}\left[\underline{\underline{\mu}}^{\text{init}}\right]$ where $\mathcal{T}$ denotes the initializing similarity transformation found via Equation 3.14.

Equation 3.19 approximates PGA by generating a new shape via a linear combination of eigenvectors in the tangent plane to the initialization model, *i.e.,* $T_{\left(\underline{\underline{\mu}}^{\text{init}}\right)}\mathcal{M}(n)$, when those eigenvectors were actually principal geodesics in the tangent plane to the training mean, *i.e.,* $T_{\underline{\underline{\mu}}}\mathcal{M}(n)$. One means of correcting this approximation is to develop the appropriate parallel transport to transform the principal geodesics at $\underline{\underline{\mu}}$ into geodesics at $\underline{\underline{\mu}}^{\text{init}}$. An alternative would be to use PGA to learn the probability distribution on m-rep differences that describe how initialization models $\left\{\underline{\underline{m}}^{\text{init}}\right\}$ differ from training models $\left\{\underline{\underline{m}}^{\text{fit}}\right\}$ fit to reference segmentations of corresponding images. Because $\underline{\underline{m}}^{\text{init}}$ typically approximates well the segmentation of the target image, residuals of the form $\Delta\underline{\underline{m}} = \left(\underline{\underline{m}}^{\text{init}}\right)^{-1} \circ \underline{\underline{m}}^{\text{fit}}$ would presumably be quite small. Consequently PGA on these residuals would form a tighter distribution than the PGA used during initialization. This would mitigate the risk of the image-driven segmentation "wandering away" from the nearly-correct initialization model.

6.2.2 Estimating the local credibility of a segmentation

In Chapter 4, I introduced a test that identifies regions on a segmentation boundary where a local image match has an improbably poor score. The segmentation in these regions is suspected of being non-credible. I validated this test by via ROC analysis comparing the regions it detected with regions for which the distance from a representative point to the nearest point on the boundary of a reference segmentation exceeded a threshold value.

In this section, I will discuss areas of future work relevant to the test of non-credibility itself. Then I will discuss future work related to the validation of the test of non-credibility and related to the larger open problem of establishing non-uniform tolerances for segmentations.

One area of future work is to extend the test of non-credibility so that it can characterize a certain class of errors. In cases where there is a strong difference between the interior and exterior intensity patterns near a region on the boundary of an object, a non-credible region one side of the boundary might be better understood by taking into account the expected intensity patterns from the opposite side of the boundary. The intensity patterns from the opposite side

of the boundary can be used to learn an *a contrario* distribution, *i.e.,* probabilities on what the intensities on the original side of the boundary should not be.

A local over-segmentation, *i.e.,* a segmentation region where the boundary is actually drawn exterior to the object, may manifest itself as follows. The exterior of that region has the desired intensity pattern, but the interior of the region, which actually contains a number of voxels that truly belong to the exterior, would have an intensity pattern that is a mixture of what is expected and what should be excluded based on the *a contrario* distribution. A similar argument could be made for a local under-segmentation, after substituting interior for exterior and vice versa.

It may be possible to slide a representative point in a region along the normal to the segmentation boundary, in search of a position that improves the interior image match without substantially degrading the exterior image match. With respect to image match functions based on RIQFs, there are other possible ways to learn the magnitude of a such a local over-segmentation that would be worth exploring. It may also be possible to factor the sampled QF from the interior region into components from the interior and exterior distributions. The magnitude of the segmentation error would be derived from the fraction of the mixture due to exterior intensities. However, it is not clear that such a factorization is possible, especially considering the weighted contributions that voxels make to RIQFs. Noting that the truly-exterior voxels in this interior region are all closer to the segmentation boundary than are any truly-interior voxels, another possible way to learn the magnitude of a local over-segmentation is to reduce the sampling depth of the interior region to the largest value such that the corresponding QF is an inlier of the *a contrario* distribution. That sampling depth is an approximation of the error magnitude. This method may require the *a contrario* distribution to be learned at a variety of scales, so that each truncated QF is compared to a distribution of RIQFs at the same depth.

Another area of future work related to the test of regional non-credibility is to revisit the assumption I make in Chapter 4 that the regional image matches are independent of each other. Given that local correlations in image data exist and given that the decomposition of segmentations into regions in the experiments I have discussed has been somewhat arbitrary, it

is likely that this independence assumption does not hold. Joshua Stough's recent discoveries regarding the clustering of these regions [77] may lead to a more complete understanding of inter-region relationships. Modifying the test of credibility to be aware these relationships may prove fruitful.

Stough's region clustering methods combine small regions, at the scale of medial spoke-ends, to form larger regions that share similar intensity distributions. Benefits of this clustering are that intensity statistics for large regions are more stably estimated than for smaller regions because larger regions contain more voxels than smaller regions, and that the variation across images of intensity patterns in a large homogeneous region, as would be produced by clustering, is better modeled via PCA of RIQFs than the variation in arbitrary large regions which are likely to contain mixtures of other distributions.

Errors of the test of non-credibility could be explained by regions that are too small because the small region size limits the number of voxels contributing to a region, and thus makes it difficult to estimate that distribution. Consequently false-positives of the test of non-credibility could be expected since a noisy estimate of the QF for a target region would be compared with the training population for that region. Noise in the training RIQFs may cause false-negatives as large variance in the training data could cause the QF for a mis-segmented region to have an acceptable image match score.

Overly large regions could also corrupt the test of non-credibility. If the region is large, the distribution of its voxels could be robustly estimated. However, that distribution is likely to be a mixture of various components, so variation across a population of such distributions is unlikely to be well modeled by PCA on RIQFs. Thus, the image match for a target region would be difficult to interpret.

Stough's clustering of correlated local regions has the potential to identify somewhat large regions for which the intensity distribution can be robustly estimated. The variation in such distributions is likely to be describable via PCA on RIQFs. Thus, the image match score is likely to truly reflect the dissimilarity between a target QF and the training population, so an outlier should be suspected of non-credibility.

The correlations of the intensity patterns across regions also provide an opportunity for

further research on the validation of the test of non-credibility. The validation reported in Chapter 4 defined ground truth in terms of the surface distance from the anchor point of a region to the nearest point in a reference segmentation. A simple extension to this method would be to compute an expected distance from many boundary points in the region to the reference segmentation. This contribution of each distance to the expectation could be weighted, just as is the contribution of each voxel to the RIQF. A more sophisticated extension might also take into account the correlations between regions. For example, a region with an outlier image match that is close to the reference segmentation, but adjacent to a region with a significant error might not be counted as a false positive of the test.

Future work should also revisit the validation of the test of non-credibility. The experiments described in Chapter 4 took place under limited circumstances. Additional studies are needed to determine the extent to which this method could be applied to other target objects, and the extent to which it is robust against the image match parameters. Until broader studies are performed, I cannot rule out the possibility that the success of the method is the result of some lucky property of the limited dataset used in the original study.

An open problem in medical image analysis is how to establish tolerances on a segmentation that allow the magnitude of likely segmentation errors to be understood in the absence of ground truth. With respect to the problem of ART treatment planning for a prostate cancer patient, uniform margins are added to the segmented volumes to ensure an adequate dose reaches the prostate. However, the size of these margins is typically a prescribed part of the clinical protocol and is not influenced by the segmentation. Consequently, the margin around the prostate is expected to include some other tissues; those tissues are at risk of being damaged by the high radiation dose they will receive.

The work I presented in Chapter 4, in particular the test for identifying non-credible regions and the visualization of regions where m-rep-based bladder and prostate segmentations have produced gross errors (Figure 4.6), could lead to tighter, non-uniform margins around segmented volumes. Intuitively, a segmentation region in which local mis-segmentation historically has not occurred and for which a regional image match is credible should not require as large of a margin as a region for which segmentation is known to be difficult and the regional

image match is non-credible.

Thus, an area of future work is to determine local tolerances for a segmentation. Based on preliminary work in this direction, I believe that a solution that computes the expected local segmentation error is possible. This expectation could be computed via some appropriate regression on data specific to the target image, such as regional image matches for a region and its neighbors, as well as on population data such as a prior distribution of errors in that region. A more sophisticated approach to determining local segmentation tolerances would be to produce a probability distribution on possible errors at a specified point on the segmentation boundary, rather than the a single expected error. Forming such a distribution is another topic for future work.

6.2.3 Interpolation of Oriented Geometric Objects via Rotational Flows

In Chapter 5, I introduced novel methods for shape interpolation. For two dimensional data, the method moves oriented points along circular arcs so that the angle swept out by an arc matches the corresponding change in orientation. For three dimensional data, the method moves point-frame pairs along helical paths. The angle swept out by such a helix matches the shortest rotation between the corresponding frames and has its cylinder axis in the same direction as the axis of that shortest rotation.

Interpolations produced by these methods are frequently visually satisfying and are often more shape-maintaining than interpolations produced by moving corresponding points along the line segment that separates them. However, in certain circumstances the interpolation between two legal shapes, *i.e.,* two shapes that neither tear nor fold, can produce a new shape that does have a fold. An example of such an interpolation is shown in Figure 6.1.

One interpretation of illegal interpolations, such as the one shown in Figure 6.1, is that for some small timestep Δt, the transformation from $S\left[\hat{t}\right]$ to $S\left[\hat{t}+\Delta t\right]$ is diffeomorphic in the time range $0 < \hat{t} \leq t < 1$, but the transformation from $S\left[t\right]$ to $S\left[t+\Delta t\right]$ is non-diffeomorphic. This limits the applicability of method for the target problem of constructing landmark paths from which diffeomorphic transformations of the volume can be computed.

For many anatomical objects, including the bladders and lungs that I studied in Chapter

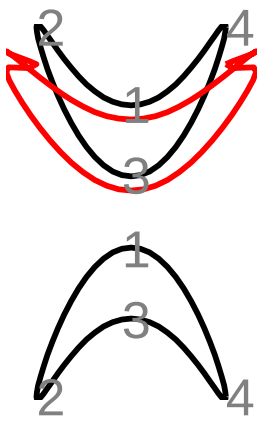


Figure 6.1: Example of a non-diffeomorphic interpolation produced by the method of rotational flows. Corresponding points on the curve that is bent upwards ($S[0]$) and the curve that is bent downwards ($S[1]$) are labeled 1, 2, 3 and 4. The interpolated curve $S[0.25]$ is shown in red. This interpolated curve features two self intersections, as well as several cusps.

5, the method will produce legal interpolations. However, an important piece of future work is to understand the geometric criteria that are necessary determine whether an interpolation will be legal or illegal.

Illegalities likely form when the changes in orientation from $t = 0$ to $t = 1$ for two vertices that are close to each other on $S[0]$ are significantly different. This is likely to occur at points in a neighborhood that contains a maximum of curvature on either $S[0]$ or $S[1]$. Further work is needed to understand the criteria that allow an illegality to form.

Once the formation of illegal interpolations is understood, a modification of the method that prevents such illegalities will be necessary. I suspect that smoothing the set of changes in orientation, *i.e.,* via some local averaging of nearby changes, will be one technique for producing a legal interpolation when the original method would fail. In Section 5.2.1, I showed several examples where unsatisfactory, but legal, interpolations were improved when the shapes were reparameterized to have constant changes in orientation. It is my conjecture that a local smoothing of orientational changes without reparameterizing the shapes would also produce satisfactory interpolations. Another open problem is the determination of optimal parameters, *i.e.,* support and contribution weights, for such a smoothing.

A separate area for further research involves quantification of changes between objects.

The current method allows the changes between shapes to be understood qualitatively, *i.e.,* animation of the rotational-flows interpolation between a pair of shapes visualizes the differences between them. A metric on point-frame pairs, such that rotational-flows interpolations are geodesic paths, would allow quantitative analysis of shape changes.

It is my impression that such a metric will not be found for $\mathbb{R}^N \times SO(N)$ because the distance from one point-frame to another depends on the trajectory the first point-frame is following. I suspect that by embedding $\mathbb{R}^N \times SO(N)$ in a higher dimensional space that also includes trajectory may allow for an appropriate metric to be discovered. Techniques for computing this metric when the initial trajectory is unknown, would require further work.

BIBLIOGRAPHY

[1] `http://www.chezlark.com/pictures/Fracture.jpg`. 1

[2] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. A log-Euclidean framework for statistics on diffeomorphisms. 66

[3] Y. Bai, X. Han, and J. Prince. Octree Grid Topology Preserving Geometric Deformable Model for Three-Dimensional Medical Image Segmentation. In *Proceedings of Information Processing in Medical Imaging (IPMI)*, pages 556–568, 2007. 35

[4] P. J. Basser and C. Pierpaoli. Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor mri. *Journal of Magnetic Resonance, Series B*, 111:209–219, 1996. 142

[5] H. Blum. A transformation for extracting new descriptors of shape. *Models for the Perception of Speech and Visual Form*, pages 362–380, 1967. 25

[6] F. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989. 63, 150

[7] F. Bookstein. *Morphometric tools for landmark data*. Cambridge University Press Cambridge, 1991. 41

[8] D. Breen and R. Whitaker. A level-set approach for the metamorphosis of solid models. *Visualization and Computer Graphics, IEEE Transactions on*, 7(2):173–192, Apr-Jun 2001. 125

[9] R. Broadhurst, J. Stough, S. Pizer, and E. Chaney. A statistical appearance model based on intensity quantiles histograms. *ISBI*, 2006. 45, 55, 107

[10] J. Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 8(6):679–698, 1986. 11

[11] G. Casella and R. Berger. *Statistical inference*. Duxbury Press Belmont, Calif, 1990. 107

[12] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic Active Contours. *International Journal of Computer Vision*, 22(1):61–79, 1997. 15, 33, 34, 43

[13] J. Cates, P. Fletcher, M. Styner, M. Shenton, and R. Whitaker. Shape Modeling and Analysis with Entropy-Based Particle Systems. In *Proceedings of Information Processing in Medical Imaging (IPMI)*, pages 333–345, 2007. 37, 40, 165

[14] T. Chan and L. Vese. Active contours without edges. *Image Processing, IEEE Transactions on*, 10(2):266–277, 2001. 36, 56, 106

[15] E. L. Chaney, 2008. personal correspondence. 98, 157, 158, 166

[16] G. Christensen, R. Rabbitt, and M. Miller. Deformable templates using large deformation kinematics. *Image Processing, IEEE Transactions on*, 5(10):1435–1447, 1996. 33, 62

[17] A. Chung, W. Wells III, A. Norbash, and W. Grimson. Multi-modal image registration by minimising kullback-leibler distance. *MICCAI02 Medical Image Computing and Computer-Assisted Intervention*, pages 525–532. 66

[18] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam. Use of active shape models for locating structures in medical images. *Image and Vision Computing*, 12(6):355–365, 1994. 15, 39, 42, 43

[19] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61:38–59, January 1995. 11, 33, 35, 39, 42, 107

[20] M. Costa, H. Delingette, and N. Ayache. Automatic Segmentation of the Bladder Using Deformable Models. *Biomedical Imaging: From Nano to Macro, 2007. ISBI 2007. 4th IEEE International Symposium on*, pages 904–907, 2007. 96

[21] W. R. Crum, O. Camara, and D. L. G. Hill. Generalized overlap measures for evaluation and validation in medical image analysis. *Medical Imaging, IEEE Transactions on*, 25:1451–1461, 2006. 71

[22] J. Damon. Geometry and medial structure. In K. Siddiqi and S. Pizer, editors, *Medial Representations: Mathematics, Algorithms, and Applications*, chapter 3. Kluwer, 2007. 29

[23] R. Davies, C. Twining, T. Cootes, J. Waterton, and C. Taylor. 3D statistical shape models using direct optimisation of description length. *7th European Conference on Computer Vision*, 3:3–21, 2002. 40

[24] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. A minimum description length approach to statistical shape modeling. *IEEE Trans. Med. Imaging*, 21(5):525–537, 2002. 40

[25] L. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302, 1945. 69

[26] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *Medical Imaging, IEEE Transactions on*, 23:995–1005, 2004. 11, 45, 47, 49, 53, 79, 109, 126, 136

[27] M. Foskey, B. Davis, L. Goyal, S. Chang, E. Chaney, N. Strehl, S. Tomei, J. Rosenman, and S. Joshi. Large deformation three-dimensional image registration in image-guided radiation therapy. *Physics in Medicine and Biology*, 50(24):5869–5892, 2005. 33, 66

[28] D. Gabor et al. Theory of Communication. 1946. 11

[29] P. J. Giblin and B. B. Kimia. Local forms and transitions of the medial axis. In K. Siddiqi and S. Pizer, editors, *Medial Representations: Mathematics, Algorithms, and Applications*, chapter 2. Kluwer, 2007. 25

[30] J. Glaunes and S. Joshi. Template estimation form unlabeled point set data and surfaces for computational anatomy. In *Mathematical Foundations of Computational Anatomy: Geometrical and Statistical Methods for Modelling Biological Shape Variability*, pages 58–65, October 2006. 65, 126

[31] C. Goodall. Procrustes Methods in the Statistical Analysis of Shape. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(2):285–339, 1991. 41

[32] K. Gorczowski, M. Styner, J.-Y. Jeong, J. Marron, J. Piven, H. Hazlett, S. Pizer, and G. Gerig. Statistical shape analysis of multi-object complexes. *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–8, 17-22 June 2007. 135

[33] A. Goshtasby. Registration of images with geometric distortions. *Geoscience and Remote Sensing, IEEE Transactions on*, 26(1):60–64, 1988. 63, 150

[34] L. Grady and G. Funka-Lea. An energy minimization approach to the data driven editing of presegmented images/volumes. In *MICCAI*, pages 888–895, 2006. 108

[35] Q. Han. *Proper Shape Representation of Single- and Multi-Figure Anatomical Objects*. PhD thesis, The University of North Carolina. 32

[36] Q. Han, D. Merck, J. Levy, C. Villarruel, J. Damon, E. Chaney, and S. Pizer. Geometrically proper models in statistical training. In *Proceedings of Information Processing in Medical Imaging*, pages 751–762, 2007. 11, 33, 35, 45, 46, 86, 109, 135, 144

[37] Q. Han, S. Pizer, and J. Damon. Interpolation in discrete single figure medial objects. *IEEE CVPR Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2006)*, page 85, 2006. 32, 46, 58

[38] Q. Han, S. Pizer, D. Merck, S. Joshi, and J. Jeong. Multi-figure Anatomical Objects for Shape Statistics. In *Proceedings of Information Processing in Medical Imaging (IPMI)*. Springer, 2005. 45, 54

[39] B. Horn and B. Schunck. Determining Optical Flow. *Artificial Intelligence*, 17:185–203, 1981. 65

[40] P. Jaccard. THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE. 1. *New Phytologist*, 11(2):37–50, 1912. 69

[41] S. Joshi and M. Miller. Landmark matching via large deformation diffeomorphisms. *Image Processing, IEEE Transactions on*, 9(8):1357–1370, 2000. 33, 64, 125, 126, 150, 151, 153

[42] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321 – 331, 1988. 11, 15, 33, 34

[43] A. Kelemen, G. Szekely, and G. Gerig. Elastic model-based segmentation of 3-D neuroradiological data sets. *Medical Imaging, IEEE Transactions on*, 18(10):828–839, 1999. 15, 33, 43, 44

[44] C. Ketting, M. Austin-Seymour, I. Kalet, J. Unger, S. Hummel, and J. Jacky. Consistency of three-dimensional planning target volumes across physicians and institutions. *Int J Radiat Oncol Biol Phys*, 37(2):445–53, 1997. 13

[45] D. Lee. Two-Dimensional Voronoi Diagrams in the L p-Metric. *Journal of the ACM (JACM)*, 27(4):604–618, 1980. 28

[46] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proceedings of Computer Vision and Pattern Recognition*, pages 316–323, 2000. 37

[47] J. Levy, R. Broadhurst, J. Jeong, X. Liu, J. Stough, G. Tracton, S. Pizer, and E. Chaney. Prostate and Bladder Segmentation Using a Statistically Trainable Model. *International Journal of Radiation Oncology, Biology, Physics*, 69(3S):353–353, 2007. 33

[48] J. Levy, R. Broadhurst, S. Ray, E. Chaney, and S. Pizer. Signaling local non-credibility in an automatic segmentation pipeline. In J. P. W. Pluim and J. M. Reinhardt, editors, *Proceedings of SPIE*, volume 6512, page 65123Q. SPIE, February 2007. 105

[49] J. H. Levy, M. Foskey, and S. M. Pizer. Rotational flows for interpolation between sampled surfaces. *IEEE CVPR Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2008)*, 2008. 125

[50] J. H. Levy, K. Gorczowski, X. Liu, S. M. Pizer, and M. Styner. Caudate segmentation using deformable m-reps. In T. Heimann, M. Styner, and B. van Ginneken, editors, *MICCAI Workshop on 3D Segmentation in the Clinic: a grand challenge*, pages 47–55, October 2007. 33

[51] X. Liu, J.-Y. Jeong, J. H. Levy, R. R. Saboo, E. L. Chaney, and S. M. Pizer. A large-to-fine shape prior for probabilistic segmentations using a deformable m-rep. *IEEE CVPR Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2008)*, page tbd, 2008. 55, 122

[52] J. Logue, C. Sharrock, R. Cowan, G. Read, J. Marrs, and D. Mott. Clinical variability of target volume description in conformal radiotherapy planning. *Int J Radiat Oncol Biol Phys*, 41(4):929–31, 1998. 13

[53] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 11

[54] C. Lu, S. Pizer, S. Joshi, and J. Jeong. Statistical Multi-Object Shape Models. *International Journal of Computer Vision*, 75(3):387–404, 2007. 45, 55, 122

[55] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967. 162

[56] K. Mardia and I. Dryden. *Statistical Shape Analysis*. John Wiley & Sons, Ltd, 1998. 41

[57] D. Merck, G. Tracton, R. Saboo, J. Levy, E. Chaney, S. Pizer, and S. Joshi. Training models of anatomic shape variability. *Medical Physics*, To appear: 2008. 45, 46, 86, 109, 135, 144

[58] C. E. Metz and X. Pan. 'proper' binormal roc curves: theory and maximum-likelihood estimation. *J. Math. Psychol.*, 43(1):1–33, 1999. 119

[59] K. E. Muller, 2006. personal correspondence. 24

[60] P. Musé, F. Sur, F. Cao, Y. Gousseau, and J. Morel. An A Contrario Decision Method for Shape Element Recognition. *International Journal of Computer Vision*, 69(3):295–315, 2006. 122

[61] D. Nain, S. Haker, A. Bobick, and A. Tannenbaum. Multiscale 3-D Shape Representation and Segmentation Using Spherical Wavelets. *Medical Imaging, IEEE Transactions on*, 26(4):598–618, 2007. 44, 45

[62] W. J. Niessen, C. J. Bouma, K. L. Vincken, and M. A. Viergever. Error metrics for quantitative evaluation of medical image segmentation. In R. Klette, H. S. Stiehl, M. A. Viergever, and K. L. Vincken, editors, *Theoretical Foundations of Computer Vision*, pages 275–284. Kluwer, 1998. 112

[63] D. Pasquier, T. Lacornerie, M. Vermandel, J. Rousseau, E. Lartigau, and N. Betrouni. Automatic Segmentation of Pelvic Structures From Magnetic Resonance Images for Prostate Cancer Radiotherapy. *International Journal of Radiation Oncology, Biology, Physics*, 68(2):592–600, 2007. 96

[64] S. Pizer, P. Fletcher, S. Joshi, A. Gash, J. Stough, A. Thall, G. Tracton, and E. Chaney. A method & software for segmentation of anatomic object ensembles by deformable m-reps. *Medical Physics*, 32(5):1335–1345, May 2005. 108, 122, 135

[65] S. Pizer, T. Fletcher, Y. Fridman, D. Fritsch, A. Gash, J. Glotzer, S. Joshi, A. Thall, G. Tracton, P. Yushkevich, and E. Chaney. Deformable m-reps for 3d medical image segmentation. *International Journal of Computer Vision*, 55(2):85–106, November-December 2003. 15, 33, 107, 126, 135

[66] S. Pizer, J. Jeong, R. Broadhurst, S. Ho, and J. Stough. Deep Structure of Images in Populations via Geometric Models in Populations. *this volume, International Workshop on Deep Structure, Singularities and Computer Vision (DSSCV)*, 2005. 55

[67] S. Pizer, J. Jeong, C. Lu, K. Muller, and S. Joshi. Estimating the statistics of multi-object anatomic geometry using inter-object relationships. *Proc. Workshop on Deep Structure, Singularities and Computer Vision, Springer LNCS*, 2005. 45, 55

[68] S. Pizer, K. Siddiqi, G. Székely, J. Damon, and S. Zucker. Multiscale Medial Loci and Their Properties. *International Journal of Computer Vision*, 55(2):155–179, 2003. 28

[69] K. Pohl, J. Fisher, M. Shenton, R. W. McCarley, W. Grimson, R. Kikinis, and W. Wells. Logarithm odds maps for shape representation. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention*, Lecture Notes in Computer Science, pages 955–963, 2006. 37

[70] A. Rangarajan, H. Chui, and F. Bookstein. The Softassign Procrustes Matching Algorithm. *Information Processing in Medical Imaging*, pages 29–42, 1997. 80

[71] J. Rossignac and J. Kim. Computing and visualizing pose-interpolating 3D motions. *Computer-Aided Design*, 33(4):279–291, 2001. 134

[72] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *Medical Imaging, IEEE Transactions on*, 18(8):712–721, 1999. 33, 63

[73] R. R. Saboo, 2008. personal correspondence. 45, 160

[74] G. Seber. *Multivariate Observations*. Wiley, 1984. 41

[75] J. Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Materials Science*. Cambridge University Press, 1999. 34

[76] W. Smith, C. Lewis, G. Bauman, G. Rodrigues, D. DSouza, R. Ash, D. Ho, V. Venkatesan, D. Downey, and A. Fenster. Prostate volume contouring: A 3D analysis of segmentation using 3DTRUS, CT, and MR. *International Journal of Radiation Oncology, Biology, Physics*, 67(4):1238–1247, 2007. 13

[77] J. Stough. *Object Relative Regional Appearance for Deformable Model Segmentation*. PhD thesis, The University of North Carolina, 2008. 122, 123, 124, 169

[78] M. Styner. *Combined Boundary-Medial Shape Description of Variable Biological Shapes*. PhD thesis, Ph. D. thesis, Department of Computer Science, University of North Carolina at Chapel Hill, 2001. 32

[79] M. Styner, H. Charles, J. Park, and G. Gerig. Multi-site validation of image analysis methods-Assessing intra and inter-site variability. *SPIE Medical Imaging*, 4684:278–286, 2002. 13

[80] M. Styner and G. Gerig. Three-dimensional medial shape representation incorporating object variability. *Proc. of Computer Vision and Pattern Recognition CVPR*, pages 651–656, 2001. 28

[81] M. Styner, K. Rajamani, L. Nolte, G. Zsemlye, G. Szekely, C. Taylor, and R. Davies. Evaluation of 3D correspondence methods for model building. *Information Processing in Medical Imaging (IPMI)*, pages 63–75, 2003. 33

[82] G. Szekely, A. Kelemen, C. Brechbuhler, and G. Gerig. Segmentation of 2-D and 3-D objects from MRI volume data using constrained elastic deformations of flexible Fourier contour and surface models. *Medical Image Analysis*, 1(1):19–34, 1996. 15, 42, 43

[83] Z. Tao and H. Tagare. Tunneling descent for map active contours in ultrasound segmentation. *Medical Image Analysis*, 11(3):266–281, 2007. 37

[84] A. Thall. *Deformable Solid Modeling via Medial Sampling and Displacement Subdivision*. PhD thesis, The University of North Carolina, 2004. 46, 58

[85] J. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998. 65

[86] A. Tsai, A. Yezzi Jr, W. Wells, C. Tempany, D. Tucker, A. Fan, W. Grimson, and A. Willsky. A shape-based approach to the segmentation of medical imagery using level sets. *Medical Imaging, IEEE Transactions on*, 22(2):137–154, 2003. 37

[87] M. Vaillant and J. Glaunes. Surface Matching via Currents. *Proceedings of Information Processing in Medical Imaging (IPMI)*, pages 381–392, 2005. 65, 126

[88] B. van Ginneken, A. Frangi, J. Staal, B. ter Haar Romeny, and M. Viergever. Active shape model segmentation with optimal features. *Medical Imaging, IEEE Transactions on*, 21(8):924–933, 2002. 33, 43

[89] S. K. Warfield, K. H. Zou, and W. M. W. III. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *Medical Imaging, IEEE Transactions on*, 23(7):903–921, 2004. 67, 123

[90] E. W. Weisstein. "Gram-Schmidt Orthonormalization." From MathWorld — A Wolfram Web Resource, `http://mathworld.wolfram.com/Gram-SchmidtOrthonormalization.html`. 136

[91] W. Wells III, W. Grimson, R. Kikinis, and F. Jolesz. Adaptive segmentation of MRI data. *Medical Imaging, IEEE Transactions on*, 15(4):429–442, 1996. 14

[92] A. Yezzi Jr, S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum. A geometric snake model for segmentation of medical imagery. *Medical Imaging, IEEE Transactions on*, 16(2):199–209, 1997. 33, 34, 43

[93] P. Yushkevich. *Statistical Shape Characterization Using the Medial Representation*. PhD thesis, University of North Carolina, 2003. 27

[94] L. Zollei, J. Fisher III, and W. Wells III. A Unified Statistical and Information Theoretic Framework for Multi-modal Image Registration. *Proceedings of Information Processing in Medical Imaging (IPMI)*, pages 366–377, 2003. 66