# ENERGY FUNCTIONALS FOR MEDICAL IMAGE SEGMENTATION: CHOICES AND CONSEQUENCES

by

Chris McIntosh

B.Sc. (Hons.), Simon Fraser University, 2005

A Thesis submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

in the
School of Computing Science
Faculty of Applied Sciences


SIMON FRASER UNIVERSITY
Fall 2011

## APPROVAL

**Name:** Chris McIntosh

**Degree:** Doctor of Philosophy

**Title of Thesis:** Energy Functionals for Medical Image Segmentation: Choices and Consequences

**Examining Committee:** Dr. Mark Drew
Chair

---

Dr. Ghassan Hamarneh, Senior Supervisor

---

Dr. Brian Funt, Supervisor

---

Dr. Greg Mori, SFU Examiner

---

Dr. Anthony Yezzi, External Examiner,
Professor of Electrical and Computer Engineering,
Georgia Tech

**Date Approved:** 30 November 2012

# Declaration of
# Partial Copyright Licence

# Abstract

Medical imaging continues to permeate the practice of medicine, but automated yet accurate segmentation and labeling of anatomical structures continues to be a major obstacle to computerized medical image analysis. Though there exists numerous approaches for medical image segmentation, one in particular has gained increasing popularity: energy minimization-based techniques, and the large set of methods encompassed therein. With these techniques an energy function must be chosen, segmentations must be initialized, weights for competing terms of the energy functional must be tuned, and the resulting functional minimized. There are a lot of choices involved, and their consequences are not always clear. In this thesis we explore the different consequences of these choices, and provide novel methods to overcome two of the more significant problems encountered: local minima, and parameter settings.

**Keywords:** medical image analysis; segmentation; deformable models; energy minimization

*To my family. To my friends.*
*To the men and women who've sacrificed for our freedoms.*

*"Everybody's got their problems.*
*Everybody says the same thing to you.*
*It's just a matter how you solve them."*

— *Deryck Whibley*, Sum 41, *2002*

*"Know the rules well, so you can break them effectively."*

— *Dalai Lama XIV*

*"If I have seen further it is by standing on the shoulders of giants."*

— *Sir Isacc Newton*, Letter to Robert Hooke, *1675*

# Acknowledgments

As a culmination of many years and many discussions this dissertation would not have been possible without the help and hard work of many. I would like to thank SFU professors Drs. Ghassan Hamarneh, Brian Funt, Mark Drew and Greg Mori for inspiring me to pursue research in computer vision and medical image analysis. Each of them has helped improve my writing, shape my research knowledge through many valuable courses, and been a constant example of what we can all accomplish if we put our minds to it. You have all been role models, and for that I will always be grateful.

In addition to my professors, I met numerous students over my years at SFU who deserve acknowledgment: Ryan Bigelow in particular, my roommate for 5 years and close friend, whose own hard work served as a strong example; Andy Rova, my friend and group partner, without whom computer vision would have been infinitely less fun; Dr. Aaron Ward, my labmate and friend, who always leads by example; and the many other students whose own drive served as inspiration.

Thank you to Dr. Roger Tam, Assistant Professor in the Department of Radiology at the University of British Columbia, and member of the MS/MRI Research Group. Roger has not not only been a superb clinical collaborator, providing all of the corpus callosum data and expert segmentations used in Chapter 3, but has provided excellent advise and feedback on talks, papers, and my career plans. Thank you to Dr. Martha Shenton of the Harvard Medical School for providing the MRI data used in Chapter 4, and Peter Plett for assisting with the early stages of the Genetic Algorithm code development. Thank you to Dr. Anthony Yezzi of Georgia Tech for his valuable questions and feedback while serving as the external examiner for this dissertation. Thank you to Dr. Stella Atkins for her constant encouragement and excellent career advice.

I would also like to thank the Canadian Institute for Health Research, the National

Sciences and Engineering Research Council of Canada, and the Michael Smith Foundation for Health Research for graciously funding my research. Without governmental funding agencies we would not have the opportunity to think, to ponder, and to study.

Lastly, but most of all, I would like to thank my friends and family to whom this dissertation is dedicated: Thank you to my parents, Don and Wenda, without your love, support, and teachings I would never have made it here. Thank you to my brother Kyle for all you taught me. Thank you to Asami for your understanding and support; and thank you to my grandparents, Dora, Bertha, and Erwin, your hard work and sacrifices were given so my generation would know a better life—I dedicate my research to that spirit and your memory.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

Medical image segmentation (MIS) has important applications in computer-aided diagnosis, statistical shape analysis, and medical image visualization. Several classifications of segmentation techniques exist, including edge, pixel, and region-based techniques, in addition to clustering, and graph-theoretic approaches [171, 128, 143, 121]. However, the unreliability of traditional, purely pixel-based methods in the face of shape variation and noise has caused recent trends [121] to focus on incorporating prior knowledge about the location, intensity, and shape of the target anatomy [57]. One type of approach that has been of particular interest to meeting these requirements is that of energy minimization methods due to their inherent ability to allow multiple competing goals to be considered.

However, there are lot of unanswered questions when adapting energy-minimization–based MIS methods to novel problems. Of utmost importance for a given MIS task are these key questions:

1. What energy function should be used?

2. How should the energy function be optimized?

The main contribution of this work is both an exploration of, and a proposed solution to, these questions. The hypothesis is simple, and may seem intuitive, but sometimes the most obvious and intuitive solutions work. Rather than arbitrarily picking an energy function, training data should be used to determine the energy function, and to setup the optimizer. Once the energy function is setup, the optimizer that makes the least comprise in accuracy should be selected. We propose that learned energy functionals will outperform

their handpicked counterparts, and that optimizing complex energy functionals directly will provide superior accuracy to taking convex approximations to them.

This dissertation will be split into four main chapters. Chapter 2 will provide a depth of background information on MIS using energy minimization, including a review of key works, and some discussions on example problems. This introductory chapter will also outline the issues that begin to arise in adapting state-of-the-art energy-minimization–based MIS methods to real problems, namely the two questions posed above. The second chapter outlines the energy function weighting problem in greater detail, proposes a solution, and validates that indeed a significant improvement in performance is gained by learning the weights for energy functionals. We have previously published the main contributions of this chapter in the following publications [96, 98, 100]. The third chapter provides a more in-depth look at the optimization problem itself, and compares the convex approximations being rapidly adopted in MIS to our alternative solution that uses Genetic Algorithms (GA) and non-convex shape spaces to remain more faithful to the desired energy function. This chapter is published as [93, 97, 99]. Each chapter will compare our proposed solutions to state-of-the-art alternatives, and discuss key advantages and weaknesses. We end with a chapter of conclusions and future research directions in MIS.

The more detailed contributions of Chapter 3 are as follows. We explore the need for varying the energy functional across images, demonstrating significant improvements in accuracy over methods that use a fixed-set of weights for all novel images. We analytically calculate the optimal energy functional weights of competing energy terms for training $(segmentation, image)$ pairs, and use that training data to infer weights for novel images utilizing learned image manifolds. Our formulation is general, and can can be directly applied to any energy minimizing segmentation technique to improve its segmentation and increase its ability to generalize to larger datasets (e.g. [17, 18, 76, 126]). Our method is able to fully automatically initialize and tune parameters optimally, with results presented on a large ($\mathbb{N} = 470$) data set of images, and as an added bonus it provides the automatic cropping of images prior to segmentation for a 10 to 20 times speedup.

The more detailed contributions of Chapter 4 are an exploration of the application of GA to deformable models as well as the development of a novel deformable-model based on localized shape statistics using hierarchical regional principal component analysis (HRPCA). We fit complex, medial-based shape models to medical images under arbitrary energy functionals through the use of GA. Essentially, we replace the traditionally used gradient descent

with GA, and convex, implicit, global shape statistics with non-convex, explicit, localized ones. GA address the typical initialization, local minima problems associated with traditional energy-minimization techniques. Our unique application of a medial-based deformable model with HRPCA maintains a statistically feasible shape model. HRPCA has the added benefit that it also reduces the size of the search space by constraining the evolution to deformations that are intuitive (stretch, bulge, bend) and are driven in terms of localized principal modes of variation, instead of modes of variation across the entire shape. On a data set of 50 mid-sagittal brain magnetic resonance images, our method compares favorably to leading segmentation techniques, including convex / non-convex gradient based optimizers and globally optimal graph-theoretic combinatorial optimization techniques.

In order to keep this thesis more specific, a number of related contributions and corresponding publications made during my PhD have not been included [58, 64, 94, 92, 95, 101, 59, 60, 62, 61, 102, 1]. Roughly speaking these contributions include: two co-authored papers on interactive MIS [58, 64]; four first-authored papers on artificial-life models for segmentation [92, 95, 101, 102], and a corresponding open-source artificial-life for MIS framework with over 2500 downloads; three co-authored book chapters on deformable models and artificial-life MIS techniques [59, 60, 62]; a co-authored journal paper on medical image visualization in IEEE Transactions on Medical Imaging [61]; and a co-authored paper on convex, multi-region, probabilistic MIS using the isometric log-ratio transform [1].

# Chapter 2

# Background and Motivations

This chapter surveys the field of energy minimization as it applies to MIS. In energy minimization methods of this nature, a function evaluates the goodness of a segmentation for a particular image, and the minimization of that function yields the segmentation of the image (Fig. 2.1). Though highly variant in nature, the application of energy minimization methods to MIS is commonly built on five primary building blocks: (*i*) the energy function; (*ii*) the segmentation representation; (*iii*) the image representation; (*iv*) the training data; (*v*) and the minimizer. In what follows we provide an overview of each building block, and the major works therein over the past few decades.

At a high level, energy minimizing deformable models work by deforming a user provided initial shape to fit to a target structure in a medical image. Shape-changing deformations result from the minimization, with respect to the shape, of a cost function measuring how plausible is the shape model and how well it aligns with the boundaries of the target anatomy in the image. Since the shape model itself is most commonly represented by a function, the cost function is often termed an energy functional and its gradient is derived using methods from variational calculus. The shape deformations are therefore typically simulated by solving an initial value problem using gradient descent optimization algorithms [41].

## 2.1 Definitions and Foundations

We first give an outline of the energy minimization for MIS process. We define a gray-level image $I_i$, and its corresponding segmentation (i.e. pixel lables) $S_i$, each having $N$ pixels. Then $\mathbf{I}= \{I_1, I_2, ..., I_{\mathbb{N}}\}$ and $\mathbf{S}= \{S_1, S_2, ..., S_{\mathbb{N}}\}$ are sets of images and their corresponding,

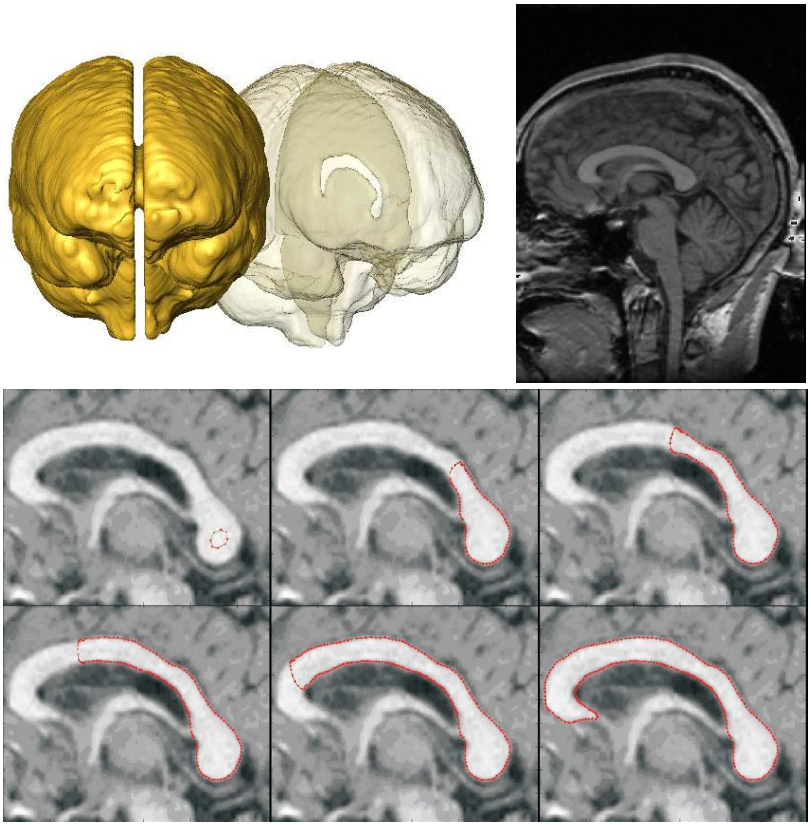


Figure 2.1: (Color Figure) A corpus callosum (CC). (top) The CC is the band of nerve fiber tissue connecting the left and right hemispheres of the brain. (bottom) An energy minimization segmentation process. A shape model with progressively lower energy (left-to-right), showing a minimum of the energy function in bottom right.

ground-truth segmentations. In a slight abuse of the notation, we occasionally omit the subscript $i$ from $I_i$ and $S_i$ for clarity, and instead use $I$ and $S$.

The first step in any energy minimization problem is the identification of the form of the energy function. In the next section we will briefly group some popular energy terms into three main categories: boundary, region, and shape. Boundary terms are concerned primarily with the object boundary, region with the region inside or outside the object, and shape terms with the shape of the object. Here we use these groupings to build a general energy functional. It may be convex or non-convex, as can the shape space over which it will be minimized. A general form is $E(S, I, \mathbf{w}) = w_1 \times boundary(S, I) + w_2 \times region(S, I) + w_3 \times shapePrior(S)$, with free parameter $\mathbf{w} = [w_1, w_2, w_3]$. We note that the boundary and

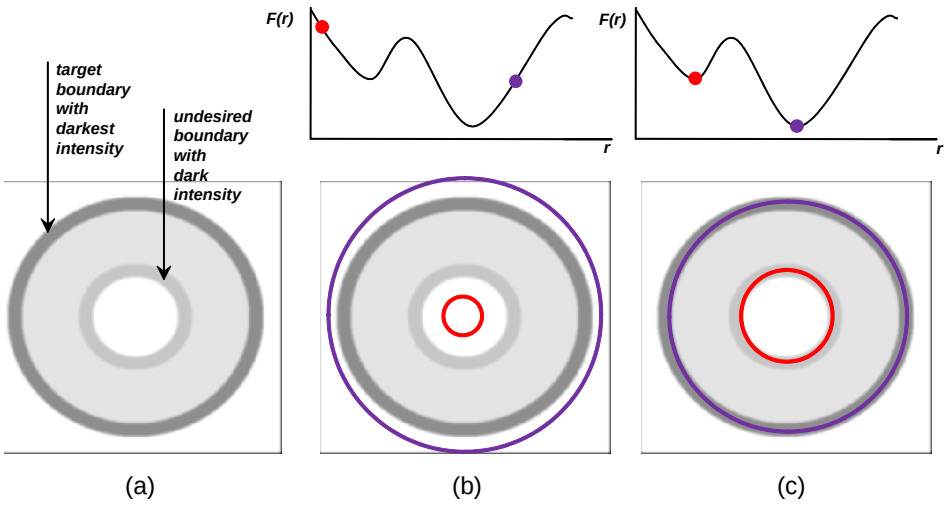



Figure 2.2: (Color Figure) Synthetic example of single parameter deformable model with local minima. The circular deformable model's only parameter is its radius $r$. The energy function $F(r)$ is minimal for the circle with darkest average intensity. The input image is shown in (a) with the darkest outmost boundary representing the global minima. In (b) two deformable models are initialized. In (c) after gradient descent, each model has moved to the nearest minima.

region terms are often referred to as external terms, since they involve things external to the shape model, while the shape prior is deemed an internal term. More generally, we can write: $E(S, I, \mathbf{w}) = w_1 J_1(S, I) + w_2 J_2(S, I) + ... + w_n J_n(S, I)$, where $J_i$ is the $i$th energy term, and $\mathbf{w} = [w_1, ..., w_n]$ are the weights. Depending on the value of $\mathbf{w}$, minima of $E$ tend towards best minimizing the boundary, region, or shape terms. We note here that depending on the nature of $S$, $E$ may be called an energy function or an energy functional, with the latter indicating $S$ itself is a function.

The segmentation problem is to solve:

$$S^* = \arg\min_S E(S|I, \mathbf{w}), \tag{2.1}$$

which involves choosing a $\mathbf{w}$ and, depending on the nature of the energy functional, may also require training appearance and/or shape priors, and setting an initialization. A gradient descent-based solver is typically used but combinatorial approaches have also been explored for discretized versions of the problem [11] (see Section 2.1.5 for details). In either case, non-convexity, or supermodularity for combinatorial problems, can be quite problematic. There is no guarantee that another solution does not exist which better minimizes the energy,

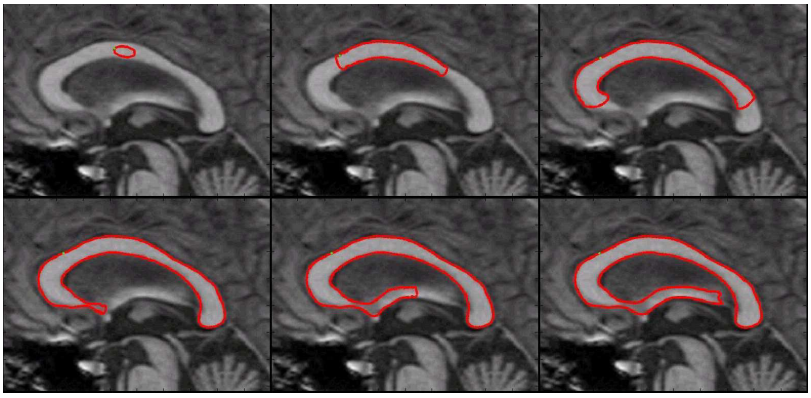


Figure 2.3: (Color Figure) An energy minimization segmentation process. A shape model with progressively lower energy (left-to-right), showing a minima of the energy function in bottom right. Notice the leaking into neighboring structures that occurs as a result of weak edge strength.

and thus is potentially a better segmentation. Ideally both functional and shape spaces are convex, guaranteeing globally optimal solutions. In Chapter 4 we explore how best to treat non-convex functions.

We now begin our more in depth discussion of the different components of the energy minimization for MIS process.

### 2.1.1 Energy Function

One of the earliest developed, and perhaps most recognized, examples of energy minimization methods being applied to image segmentation is that of deformable models. Deformable models for MIS gained popularity since the 1987 introduction of *Snakes* by Terzopoulos *et al.* [154, 71]. However, with only smoothness and a simple image gradient-based boundary term, these models were highly susceptible to weak edges, noise, and entrapment in local minima (Fig. 2.3).

Since Snakes paved the way, there have since been numerous papers attempting to increase accuracy by contributing novel energy terms, each designed to address a particular class of images, or alleviate a particular problem with the original terms. As there have been far too many proposals to survey them all, here we focus on the key terms, many of which themselves have spawned numerous new approaches to MIS or changed the way we think about the problem at hand. As much of the initial development focused on external energy

terms, namely on the boundary and region terms that deal with the image processing itself, we begin our discussion there.

One of the most fundamental problems noted with snakes relates to their boundary capture range. If placed near a strong edge in an image, the contour would quickly *snap* to the edge, but if initialized only a few pixels away the influence of the edge would not reach the contour. There are, in fact, numerous causes of this problem relating to not only the external energy terms[23, 17, 169], but also the way the segmentation was original represented (an explicit contour - see Section 2.1.2 for details), and the minimization technique being used.

An early attempt to rectify the aforementioned problem was by adding a deflation or inflation force to the contour that would attempt to shrink/grow it towards edges [23]. Rather than rely on a constant force, gradient vector flows extend the influence of edges off into homogeneous regions, thus increasing the capture range [169]. Geodesic active contours (GAC) were similarly developed by both Caselles *et al.* in [17] and Yezzi *et al.* in [170]. The approach of Caselles *et al.* formulated a deformable model optimization problem as that of finding the optimal path in a Riemannian space. Termed geodesic active contours (GAC), these popular deformable models work by minimizing curve-length, where length is measured as the geodesic distance on a Riemannian manifold defined via an edge indicator function. The shortest curve is then, by definition, the curve along the edges of the image, and thus GAC automatically shrink the curve to the edges. GAC have become the canonical example of boundary-based deformable models.

However, what about objects whose boundaries are not so crisp, due to their inherent nature or noise (Fig. 2.3)? In these cases the intensity statistics of the area both inside and outside the contour can be used to attempt to divide the image into maximally separated regions. The approaches of Chan *et al.* and of Yezzi *et al.* [18, 157] were similarly developed. Both are modeled after the Mumford Shah functional wherein images are approximated by piecewise-smooth functions [108]. The approach of Chan *et al.* is referred to as active contours without edges (ACWE), and has become a popular example of energy minimization for image segmentation. However, in their initial formulations both methods approximate images by piecewise-constant functions, i.e. objects are assumed to have a constant intensity. When objects are noisy, or their intensity changes gradually, a piecewise-smooth approximation is better suited, and thus an extension to piecewise-smooth functions was developed [19].

When both boundary- and region-derived statistics are not enough, shape-based terms

are used. Shape terms provide resilience to false boundaries by heavily penalizing the implausible shape configurations that the false boundaries imply. The most basic terms attempt to achieve boundary smoothness by minimizing curve length, segmentation area, or curvature. More advanced terms compare the shape of the segmentation to some prior model of the shape in an effort to minimize their dissimilarity [26, 28, 29, 30, 163, 76, 32, 33, 118, 36, 42, 127, 35, 160] (see Section 2.1.4 for discussions).

There has even been work trying to combine edge-terms, piecewise-constant terms, and shape terms into a single formulation [15]. However, with so many competing assumptions about the object and the image inherent in each formulation, some level of trade-off must be established in the resulting functional (i.e. the weights, $\mathbf{w}$, must be set).

If not appropriately set, the weights $\mathbf{w}$ can cause significant error. In fact, our results demonstrate that optimizing the weights has dramatic effects; reducing error in large data sets by as much as 60% [96]. However, optimizing the weights by hand for even a single image can be a long and tedious task, with no real guarantee of obtaining the correct segmentation. As such, there has been a number of works that seek to automatically set the weights [96, 98, 152, 2, 150, 44, 73]. We defer a more detailed treatment of this topic to Chapter 3.

### 2.1.2 Segmentation Representation

As expected, all functions must have domains and so in turn energy functionals must have domains. That domain is the space of possible segmentations of the image, or shapes. There are many different ways to represent the underlying segmentation, and that choice in turn impacts the image and shape terms that can be readily evaluated. For example, a shape using closed contours pairs most readily with region statistics. Here we summarize some of the most prevalent shape representations in use.

Naturally, we start with the representation originally detailed by Terzopolous *et al.*, namely the explicit contour model [154, 71]. Here the contour is defined as an explicit, parameterized function of its arc length, thus integrating along the arc-length integrates along the contour and allows for the evaluation of both internal and external terms (see Appendix B for details). These methods were extended to represent not only contours, but surfaces and volumes [104, 37, 146, 22, 24, 165, 89]. Other explicit models were introduced relating to spring-mass systems, where each boundary point is a mass, connected to other masses via springs [87, 106]. There can be other types of masses, namely internal nodes

and medial nodes [125]. However, these aforementioned works are constrained to a fixed topology, which posed problems in many applications.

There have been two major directions of work to address topological adaptability. The first direction focused on novel ways to automatically re-parameterize the contour or surface enabling the evolution into complex geometries [90, 86, 88, 91]. Initially developed in 2D, *T-Snakes*, work by subdividing the image domain into a Freudenthal triangulation [90, 86]. At each deformation step the intersections of the contours with the triangular grids are found, and a set of rules is followed to determine if the contour point should be split or merged. This work was later extended to 3D, with the introduction of *T-Surfaces* [88, 91]. Delingette simultaneously developed a somewhat related approach where simplex meshes are used to model the shape, and topological changes are performed "semi-automatically with an automatic detection of topological problems, but a manual validation of all operations" [38]. However, in both cases the explicit re-parameterization of the contour or surface can be costly and may not generalize well to even higher dimensions.

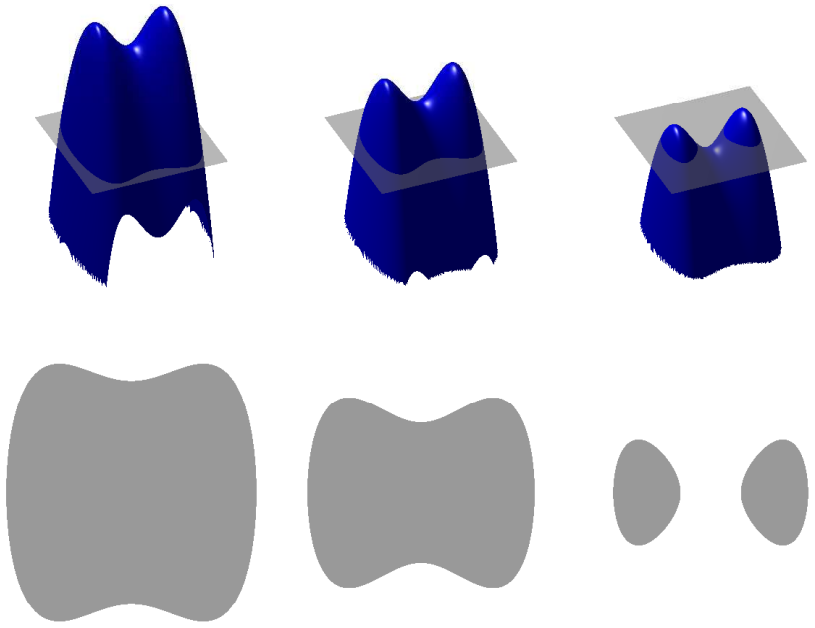


Figure 2.4: An exemplar signed distance function as a shape representation. (top) A SDF undergoing a simple deformation from left-to-right. (bottom) The zero level sets of the SDFs, displaying the segmentation each SDF represents. Notice how the topology automatically changes without any re-parameterization.

In contrast, implicit contours were built from the ground-up to handle changes in topology [116, 136, 17, 115]. In these representations the boundary is implied by a given function, instead of explicitly defined. Implicit representations are built around the signed-distance-function (SDF), where the object boundary is defined as the zero level set of the function. Integrating over the domain of the SDF implicitly integrates over the contour and thus allows for the evaluation of both internal and external terms, as before. Changes in topology are handled internally by the shape representation and require no re-parameterization of the model (Fig. 2.4). Similarly, other functions can be used to implicitly represent shapes including characteristic functions [156], and probability maps [35].

More recently, graph methods have emerged where the segmentation is represented by the assignment of labels to a graph [140, 3, 9, 4, 107, 13, 11, 12, 52]. Pixels are represented as vertices in the graphs, and edges between pixel-vertices represent a connectivity neighborhood. Energy functionals, often called cost functions in graph papers, are expressed as sums over the vertices and their neighborhoods, that vary as a function of how each vertex is labeled.

Finally, our discussion thus far has been limited to that of a single object class, but it is also possible to represent multiple object classes at one time, using so-called multi-class shape representations that are built on both implicit [174, 133, 119, 120, 159] and graph-based representations [13, 53]. Implicit shape models are adapted to multiple classes by defining multiple implicit functions, where the combinations and intersections of the functions denotes which class is being represented. Graph methods, however, are extended by increasing the number of possible labels for each vertex.

In the end, the choice of representation is ultimately determined by the desired segmentation task. As such, it is worth noting that specialized models can exist, where a shape representation is designed specifically for one type of anatomy. A popular example is that of tubular-branching objects, namely vessels, where cylindrical models can be built and deformed [114, 92].

### 2.1.3 Image Representation

As might be expected, how the image is represented, will impact how the terms of the energy functional can be evaluated on it. In fact, early on there were many contributions demonstrating how existing techniques could be modified to fit higher dimensional data, vector-valued data [138, 134, 20, 7], tensor data [161, 131], texture-heavy images [119, 120],

or even temporal data, as opposed to the static, 2D, greyscale images early algorithms were presented on. Furthermore, numerous methods have been adapted to handle the intricacies of specific medical imaging modalities including magnetic resonance images, and ultrasound. As an example, we talk briefly about some of the issues inherent to angiography in Section 2.2.

### 2.1.4 Training Data

A substantial amount of knowledge is often available about anatomical structures of interest-characteristic shape, position, orientation, symmetry, relationship to neighboring structures, associated landmarks, etc. and about plausible image intensity characteristics, subject to natural biological variability or the presence of pathology. Once collected, the training data typically comes into the energy functional in the place of shape priors [28], but appearance priors [30] have also been developed (See [66] for a complete survey). As shape priors have been a particular area of interest in the field, here we discuss some of the key works.

In many applications, prior knowledge about object shape variability is available or can be obtained by studying a training set of shape examples. This knowledge restricts the space of allowable deformations to a learned shape-space that approximates the space of anatomically feasible shapes [26, 28, 29, 30, 163, 76, 32, 33, 118, 36, 42, 127, 35, 160]. One of the most notable works in this area is that of Cootes *et al.*, where they introduced and refined active shape models (ASM) [26, 27, 28]. In ASM, principal component analysis (PCA) is calculated over a set of landmark points extracted from training shapes. The resulting principal components are used to construct a point distribution model (PDM) and an allowable shape domain (ASD). In a natural extension to their previous work, Cootes *et al.* modify their method to include image intensity statistics [30]. Staib and Duncan constrained the deformable models in Fourier space by conforming to probability distributions of the parameters of an elliptic Fourier decomposition of the boundary [147]. Statistical prior shape knowledge was also incorporated in implicit, level-set–based deformable models. Leventon *et al.* introduced statistical shape priors by using PCA to capture the main modes of variation of the level set representation [76]. However, as Pohl *et al.* point out, level-sets do not form a vector space and hence more accurate shape statistics could be captured by transforming the shapes into a vector space using the logarithm of odds before performing PCA [127].

Though simpler to optimize than their non-linear counterparts, linear models of shape

deformation may not always adequately represent the variance observed in real data. Linear shape models assume the data lies on a linear manifold, but shapes often lie on nonlinear manifolds where the manifold's properties are not accurately captured by linear statistics[45]. Picture fitting an ellipse to an 'S'-like shape space. In order to include the entire letter extraneous white-space (non-valid shapes) must also be included. Nonlinear shape models have been introduced to address this problem [145, 29, 32, 33, 45, 36, 42, 31].

However, we argue that the problem with linear statistics, as described above, is not necessarily due to the application of a linear model to nonlinear data but rather because of the implicit nature in which the statistics are applied. Chapter 4 will provide greater discussion and details on this problem, and demonstrate our proposed solution.

### 2.1.5 Minimizer

Once the image and shape representations have been set, and the problem formulated. All that is left is to solve the minimization process, and obtain the resulting segmentation. Though it may sound simple enough, this area has been a major focus of criticism of energy minimization based methods over the years, and as such has recently become one of the most focused areas for research. Specifically, in their original inception, many of the aforementioned methods were plagued by problems of local minima and sensitivity to initialization. Here we describe the changes and revelations in the field on this topic over the past decade. For a complete review the interested reader is referred to the following representative, but far from comprehensive, list [74, 112, 54, 34] for details.

Energy-functional minimization can be carried out in a variety of ways. One solution is to perform explicit differentiation under the Euler-Lagrange equations, where each new application with a modified energy functional must be accompanied by one such derivation [154, 71, 18, 17] (see Appendix A for details). The result is a set of equations, which, if satisfied, guarantee a stationary point of the energy functional. The solution is then obtained through a gradient descent process where the change in the shape model (with respect to an artificial time variable) is equated to the Euler-Lagrange equatuion, i.e. the deformable models comes to rest when the equations are satisfied [154, 71]. There are, however, two main drawbacks with this approach. Firstly, performing gradient descent in the presence of image noise can lead to instability in the deformations over time [148]. Secondly, as the number of dependent variables (shape, location, scale, orientation, etc.) increases so does the complexity of the search space, which often increases the number of local minima and

requires the calculation of an increasingly large number of derivatives.

On the issue of stability, there has been work by Sundaramoorthi *et al.* on reformulating the gradient flow using Sobolev-type inner products, that induce favorable regularity prosperities into the flow, thus bringing smoother deformations over time [148]. Recently Bar and Sapiro introduce a Newton-type method built on more generalized norms than the $L^2$ norm, obtained by equating the Euler-Lagrange constraint to an artificial time variable [8]. They include the Sobolev norm, and demonstrate improved results over $L^2$ norms, with faster convergence and more accuracy in the presence of noise. Unless used on convex problems, however, these methods are still prone to local minima.

If local minima do not suffice global minima must be sought, and as such there have been a number of recent approaches to obtaining the global optima of energy functionals [3, 9, 4, 107, 11, 12, 52, 112, 5, 77, 14, 35]. There are three primary directions towards this goal: min-paths, min-cuts, and convex approximations.

Min-path techniques are formulated on the basis of Dijkstra's algorithm for finding the shortest path in an undirected graph with nonnegative edge weights. They were first presented for 2D interactive segmentation [3, 9], extended to 3D [3], and later specialized to 4D for vessel segmentation [168, 77].

Graph cuts were demonstrated as a global minimization technique for a popular energy-functional [17], as a special case of computing a geodesic on a Riemannian space whose metric is computed from the image [11]. However, graph cuts have been shown to apply only to a restricted class of energy functionals that are submodular [74], and their solutions are discrete approximations to the continuous formulations whose accuracy is dependent on the resolution of the approximating graph [11]. Naturally, as that resolution increases so does their running time. Random walkers were developed in a similar nature, solving image segmentation as a graph problem wherein the global optimum is obtained to a particular cost function [52]. In fact, graph cuts and random walkers have been shown to be specific instantiations of a single framework [142].

Another line of work has come from the relaxation of the underlying shape model from a non-convex space to a convex one; thereby defining convex energy functionals which can then be minimized instead of their non-convex counterparts. This convex relaxation work, which began in 2004 with a simple restricted class of functionals [112], was later extended to a broader class in [14], and then a similar work appeared in 2008 with the addition of a shape prior [35]. However, restrictions still exist in that the functionals and the shape

spaces they are optimized over must be convex when defined over the relaxed space, and that the relaxed shape-space must itself be convex.

Though not guaranteed to find global optima, GA have also been applied to the minimization of energy functionals [6, 7, 82, 155, 43, 110]. At a high-level, GA work by performing many simultaneous local searches, each individually optimizing the energy functional via a random walk. At the end of the process, the search that yielded the lowest value for the energy functional is adopted as the segmentation. However, as this topic is more pertinent to Chapter 4, we defer our specific review of these methods to that chapter.

A somewhat related direction is that of multiobjective optimization, where each term is simultaneously optimized rather than optimizing a linear sum of the terms (3.1). Hence, in multiobjective optimization no weights are provided to combine the competing terms of the functional, instead a solution is sought for which no term can be improved without sacrificing another [25]. We defer a more detailed discussion of the works in this area to Chapter 3, as the concepts relate heavily to the topic of weight optimization.

### 2.1.6 Related Methods

Though our discussion thus far has been focused on energy minimization methods for image segmentation that follow the prescribed building blocks, there are a few bodies of related research that follow a different path. Here we briefly detail two of them.

#### Relation to Bayesian Methods

Throughout this chapter we have examined numerous methods built on energy functionals of the form: $E(S|I,\mathbf{w}) = w_1 \times internal(S) + w_2 \times external(S,I)$. However, an interesting parallel to probabilistic approaches can be observed with a few simple assumptions and manipulations of this general form. Here we demonstrate that this model is actually equivalent to performing image segmentation via Bayesian inference. First, we restate the segmentation problem as:

$$S^* = \arg\max_S P\left(S|I,\mathbf{w}\right), \tag{2.2}$$

where

$$P\left(S|I,\mathbf{w}\right) = \frac{P\left(I|S,\mathbf{w}\right)P\left(S|\mathbf{w}\right)}{P\left(I|\mathbf{w}\right)}. \tag{2.3}$$

Maximizing (2.2) is equivalent to minimizing its negative logarithm:

$$S^* = \arg\min_S \left(-\log P\left(S|I, \mathbf{w}\right)\right), \tag{2.4}$$

where $= -\log P\left(I|S, \mathbf{w}\right) - \log P\left(S|\mathbf{w}\right)$, and the denominator has been removed as it has no consequence on the minimization. Now suppose we model our probabilities as:

$$\begin{aligned} P\left(S|\mathbf{w}\right) &= e^{-w_1 \times internal(S)} \\ P\left(S|I, \mathbf{w}\right) &= e^{-w_2 \times external(S,I)} \end{aligned} \tag{2.5}$$

then substituting back into (2.4) yields:

$S^* = \arg\min_S \left(-\log P\left(S|I, \mathbf{w}\right)\right) = w_1 \times internal(S) + w_2 \times external(S, I)$, as before. From here it becomes possible to examine many of the approaches previously cited, and see what independence assumptions they are making from a Bayesian standpoint.

#### Relation to Segmentation via Registration

Thus far we have discussed methods where the dependent function of the energy functional is one describing the shape of the current segmentation. However, a related field exists where the dependent function is instead a transformation describing how one or more images are related. This is the field known as image registration, and it involves finding a mapping from one image to another, identifying which pixels in a source image map to which pixels in a target image. Image registration can be used for segmentation when the mapping from a novel image is found to a training image with a known segmentation, since that mapping effectively labels the novel image. As this field is far to large and diverse to cover here, we refer the interested reader to [83], and more recently [175], for a complete survey. We do note, however, that many of the same issues discussed in this dissertation, i.e. global versus local optima, and setting the weights for the energy functional, are also important problems in registration.

## 2.2 MIS via Energy Minimization

As already noted in this chapter, energy minimization methods have been applied to a wide variety of MIS problems. Two popular application domains are those of cardiac images, and vascular images. As an example of how energy minimization can be applied to MIS, here

we briefly discuss some key works relating to vascular segmentation. Though a complete survey of energy minimization for MIS does not exist, the interested reader is referred to [87] for a related survey of MIS using deformable models.

One structure of particular interest in the diagnosis and understanding of many diseases is vasculature [16]. Vessel segmentation remains an interesting application area of energy minimization methods because of its unique challenges. Firstly, the topology is complex, and as such many of the already mentioned topology adaptive shape models were first demonstrated in their application to vessel segmentation [91]. Secondly, the vessels are often of very low contrast motivating advances in image terms [47, 167, 168, 158]. For example, Vasilevskiy and Siddiqi built flux maximizing geometric flows based on the observation that the gradient vector field near a vessel should be orthogonal to the vessel [158]. They define a flux maximizing geometric flow as one for which the inward normals of the underlying curve align with the direction of the gradient vector field. Near vessels the gradient vector field points inward toward the vessel centerline, and thus maximizing the flux will align the boundary of the segmentation to the boundary of the vessel. The last major challenge in vessel segmentation is that the vessels can be very thin, pushing the boundaries of numerical stability in many techniques and motivating new methods with increased stability to thin structures [80, 148]. For example, Lorigo *et al.* modify GAC to deform along the medial-axis of a tubular shape, as opposed to its surface [80]. For a complete survey of vessel segmentation techniques the reader is referred to [75].

## 2.3 Discussion

Having briefly touched on each of the five primary building blocks of energy minimization methods, we conclude with a discussion of the main issues concerning their usage for MIS. Namely, issues relating to how to build the energy functional, represent the segmentation, deal with the different imaging modalities themselves, train priors, and finally, how to minimize the resulting system. We begin by touch briefly on issues relating to validating the method, as that is a fundamental step in the research process, that we have not yet discussed.

Every MIS method must be validated. There are two main approaches to validate a MIS method: expert segmentations, and synthetic data; each of which has its own inherit advantages and drawbacks. Expert segmentations can be time consuming and costly to obtain.

Furthermore, expert segmentations suffer from both inter- and intra-operator variability because multiple experts, or even the same expert on different days, can obtain differing segmentations of the same object. Warfield *et al.* attempt to address problems with inter- and intra-operator variability through an expectation-maximization algorithm that weights different segmentations according to a variety of measurements and rules [164]. The key advantage to expert segmentations of real medical data is that the data is real and hence it demonstrates the applicability of the method to the problem at hand. Synthetic data can be created by either physical or computational phantoms. Physical phantoms are those physically constructed and then imaged in some manner, whereas computational phantoms are simulated using mathematical models designed to replicate human anatomy under specific imaging protocols [21]. The main drawback in both cases is realism: segmenting a phantom well does not necessarily mean real data will also be segmented with high accuracy. The main advantage of synthetic data is certainty about the ground truth segmentations. Whether real data or synthetic data, a measure of dissimilarity between the automatic segmentation and the ground truth must eventually be computed. Standard approaches for evaluating segmentation results given ground-truth segmentation include the Hausdorff distance, Dice coefficient [39], the Jaccard index [69], and the Tanimoto coefficient [151]. In this dissertation we make use synthetic data for illustration purposes, and real data for validation. For consistency, we will present all results using the Jaccard index [69], but defer presentation of the formula until it is needed.

Choosing the right energy functional for the given task is a crucial first step. Ideally, one hopes for strictly convex functions with their global minima lying at the correct segmentations for a given set of images, but this ideal scenario is rarely the case. The main challenges here are determining what energy terms could make a good functional, and how to weight them in a such way as to best segment the images. We expand on this topic in Chapter 3.

There are often two main concerns when choosing the segmentation representation. Firstly, will it allow for the training of appropriate shape priors, and secondly, will it introduce problems in the minimization stage. Some shape representations do not form vector spaces, and thus performing statistics on them is difficult. Level sets (SDFs) are one such example, where a specific method for computing statistics over the resulting shape space was needed [127]. Level set-based representations can also cause problems in the minimization stage, as they do not form a convex-space, and thus introduced non-convexity into the

minimization problem [35]. However, they remain a popular technique due to the sub-pixel accuracy, and automatic handling of topological changes.

Each imaging modality has its own inherent problem associated with it relating to different types of noise and spatial or temporal resolution of the structure of interest. As already exemplified with a brief treatment of blood vessel segmentation methods, each application can require customized energy functionals, segmentation representations, and optimization strategies.

When training priors, one is often forced to balance two competing goals: having the priors accurate versus having the resulting energy functional solvable. Linear statistics do not often fit the training data well, but lead to easily solved energy functionals, whereas non-linear statistics fit the data but are difficult to work with. The problem with linear statistics, however, is often not necessarily due to the application of a linear model to non-linear data but rather due to the global aspect of the shape statistics themselves, and due to the shape representation to which they are applied. Global shape statistics are those that model the shape variation of the entire shape simultaneously, i.e. each shape is a single point in some high-dimensional space, and the statistics (linear or not) describe some restricted set of that space. Many deformable model approaches adopt a boundary based representation. As a consequence the statistics are calculated using boundary models of the shape instead of models representing the interior and skeletal topology of the structures, leading to a loss of accuracy (see Chapter 4 for details).

The primary problem in minimization remains how to globally optimize an increasingly large set of energy functionals, as opposed to the restricted sets seen thus far [74, 14, 35]. Interestingly enough, even with global minima, existing functionals are not proving accurate enough, motivating the search for new energy terms or ways of building functionals that can better respect the image variability.

In short, developing MIS methods remains a daunting task from start to finish with numerous areas for research relating to each of the five building blocks commonly found in energy minimization techniques. However, with the ever growing popularity of medical imaging for diagnosis and disease understanding, developing robust, automatic techniques for MIS remains an important goal. In order to reach that goal we must continue to make breakthroughs in terms of accuracy and reliability of MIS methods. To do that, we need to consider where the bulk of the error and performance variability is coming from, so as to best focus our efforts there.

This dissertation does just that, and in doing so proposes novel solutions to two general MIS problems. In Chapter 3 we demonstrate that the weights can be responsible for more than 60% of the segmentation error, and provide a method to tune the weights automatically. Then in Chapter 4 we consider the impact of the aforementioned issues in building a shape-space, and the impact of choosing convex approximations to non-convex functionals and shape spaces.

# Chapter 3

# Optimal Weights

Medical imaging continues to permeate the practice of medicine, but automated yet accurate segmentation and labeling of anatomical structures continues to be a major obstacle to computerized medical image analysis. Though there exists numerous approaches for MIS, one in particular has gained increasing popularity: energy minimization-based techniques, and the large set of methods encompassed therein. With these techniques segmentations must be initialized, weights for competing terms of an energy functional must be tuned, and the resulting functional minimized. There is a substantial amount of guesswork involved. We seek to reduce this guesswork by analytically determining the optimal parameters, and automatically initializing the optimization procedure.[1].

## 3.1 Introduction

In the previous chapter we described an increasingly large number of energy terms that are being developed. As a consequence of having so many different energy terms, a growing problem exists in choosing the right terms for a given application, and in particular choosing the relative importance, or weight, of each term. In this work we demonstrate the significant impact choosing the correct weights can have on MIS (Fig. 3.1).

The primary goal of this work is to address two highly coupled problems in energy-minimization–based MIS:

1. **Choosing the energy terms to use:** Ideally, the best energy terms are those

[1]This work builds upon and extends that presented in [96, 98].

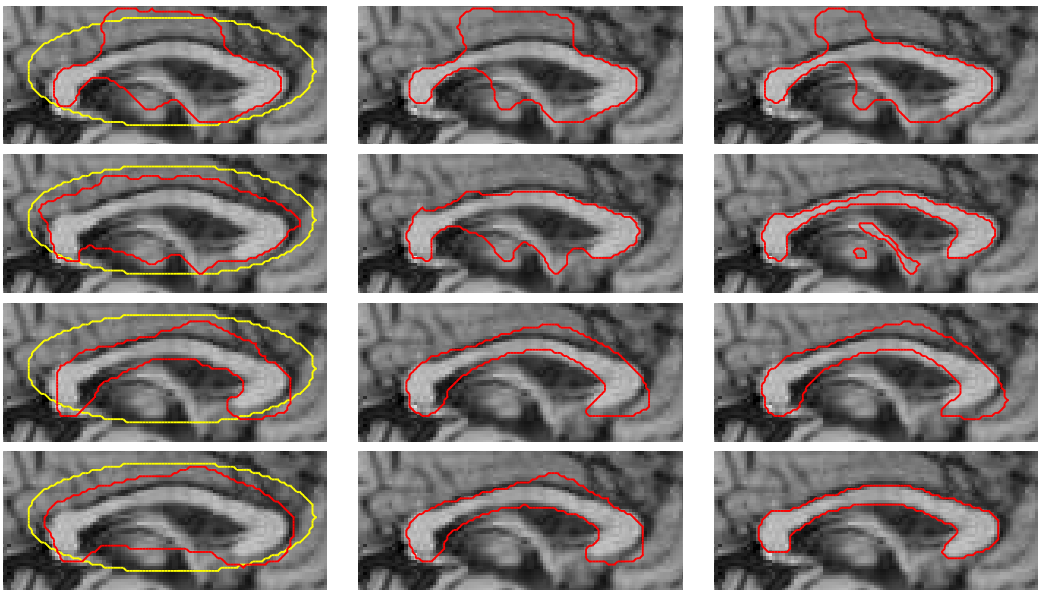


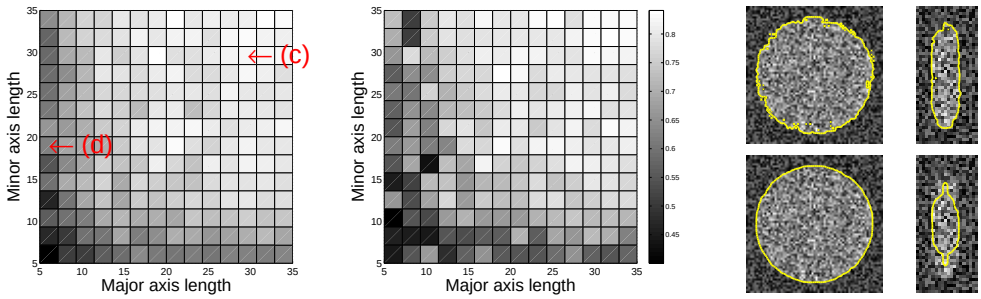
Figure 3.1: (Color Figure) An energy minimization segmentation process (red) with two different sets of weights yielding very different accuracy. Each row employs a single set of weights, with energy minimized from left-to-right. Initializations are shown in yellow. From top-to-bottom, excessive boundary term weight, region term weight, and shape prior weight. Bottom row demonstrates the optimal weights (Section 3.3, (3.4)), yielding the correct segmentation. In all but the last row, the segmentation progresses erroneously, misleading the user that different energy terms, or perhaps a different initialization or solver, is needed. In reality all that was needed was a re-weighting of the energy terms (bottom row).

where the ground truth segmentation of each image is the term's global optimum. Traditionally, choosing the correct terms is a laborious task, performed manually, but one that greatly impacts the methods accuracy. For example, a term may be found to perform well for one image, and then poorly for a second image. A second term may work poorly for the first image, but well for the second. In other words, one term may not suffice for all images. Furthermore, though a term may individually perform poorly, it may work well in combination with other terms, necessitating a setting of relative importance between terms.

2. **Choosing the relative importance of the energy terms:** Two seemingly erroneous terms may be combined to provide a vastly better energy functional than either term by itself, but there is a seemingly infinite set of possible weights to evaluate. As before, the discovered weights may work well for one image, but poor for another.

Leading to the question, is a single energy-functional sufficient for an entire set of medical images? Or, is it better to automatically adapt the energy-functional to each image, by modifying the weights on a per-image basis.

We propose that both problems can be addressed by automatically tuning the weights using ground truth segmentations as training data. Rather than having a user painstakingly try and select energy terms, and subsequently tune their weights, we outline a method to automate the process on either an image-set, or a per-image basis (i.e. one set of weights for a set of images, or a set of optimal weights for each novel image individually). In what follows we elaborate on the finer points of the weight tuning problem.




(a) (b) (c) (d)

Figure 3.2: Segmenting ellipses with varying lengths of major and minor axes in noisy images, using a simplified energy: $energy(shape, image) = w \times region(shape, image) + (1 - w) \times shapePrior(shape)$. (a) optimal values of $w$ for each noisy ellipse image, *given* its known segmentation, as calculated by our algorithm (Section 3.3). Circles are along the diagonal. Brighter pixels imply higher values of $w$. (b) optimal values of $w$ for each image, *without* knowing its segmentation, as calculated by our algorithm (Section 3.4). Note the similarity of the brightness pattern (weights) in (b) to that in (a). (a) and (b) span the intrinsic 2D space (corresponding to change in major and minor axes) of the image manifold learned from 225 ellipses. The locations of the example images in (c) and (d) are indicated with small arrows in (a). (c) Segmentation of a noisy circle image with $w = 0.347$ (optimal parameter from (d)) (top) and with its optimal parameter $w = 0.826$ (bottom). (d) segmentation of a noisy ellipse image with its optimal parameter $w = 0.347$ (top) and with $w = 0.826$ (optimal parameter from (c)) (bottom). Note the need for different weights to properly segment the different images.

### 3.1.1 The Building Blocks of an Energy Functional

Selecting from different combinations of boundary, region, and shape terms, we can construct an energy functional for MIS. For example, in building a strong objective function for

MIS it might be advantageous to combine boundary-terms, region terms, and shape terms into a single formulation [15]. However, combining them requires determining the relative importance (weights) of each term. At a high level the resulting functional would resemble:

$$\begin{aligned} energy(shape, image) = \quad & w_1 \times boundary(shape, image) \\ & + w_2 \times region(shape, image) \\ & + w_3 \times shapePrior(shape), \end{aligned} \tag{3.1}$$

where $\mathbf{w} = [w_1, w_2, w_3]$ are scalar weights. Classically, the weights, $\mathbf{w}$, are manually set by the user (empirically or by trial and error) or via an offline brute force (within reasonable ranges) in an effort to balance the trade-off between the terms. For even the familiar user, the weight setting process can be laborious with little intuitive understanding of how much to change a specific weight, or exactly how the resulting segmentation will be impacted. Instead, we seek to analytically determine the weights for an image, thus removing the need for user involvement and adding greater automation and accuracy to the MIS process.

Using one or more training examples we build a method to automatically determine the weights that best optimize a set criterion that analytically describes the best energy functional. At a high level, we solve for the weights that encourage the ground truth segmentation to be a minimum of the energy functional, while discouraging other incorrect segmentations from being minima themselves. Furthermore, where sufficient training data is available we present a technique that can infer the optimal weights, along with specialized shape and intensity priors, for a novel image on a per-image basis. In what follows we elaborate more on what it means to have different weights for different images.

### 3.1.2 Is a Single Energy Functional Sufficient?

Classically, the underlying assumption has been that once the function, parameters[2], initialization, and solver are in place, they can be fixed across all images. For example, consider the weights $\mathbf{w} \in [0, 1]^3$ in the simplified formulation (3.1). Segmentation is performed by finding the *shape* that minimizes the *energy* for a given *image*. As $w_3$ approaches zero the model will be attracted solely to image features. As $w_1, w_2$ approach zero the objective

[2]We identify two types of common parameters found in energy functionals: the weights $\mathbf{w}$ which balance competing terms of the functional; and the parameters that describe the prior shape and appearance models extracted from training data.

function will favor smooth shapes with no regard to the image data. However, what if the characteristics of the images vary across the data set?

For noisy images smoothness must be emphasized, while for highly curved structures smoothness must be relaxed. Different patients have different anatomy, and they may be imaged at different times with different noise. Therefore, is it realistic for a fixed set of weights to address the variability present in meaningful medical image data sets? We argue that it is not. Fig. 3.2 shows how even for a simple data set different images can have different optimal weights, and weights that work for one image may not work for another. The training data set used to provide the shape and appearance priors is also typically fixed over the data set. A better approach is to adjust the training data, and hence the priors, to reflect the desired segmentation. Suppose, for example, the shape prior is a Gaussian distribution modeling a set of training shapes, but the shape distribution is a non-linear space with locally linear segments, like for example, 'S'-shaped. Modifying the training set to include only a few likely shapes is advantageous as it accurately models the local shape variation. This is akin to fitting 5 different Gaussians to an 'S'-like shape space, each modeling a locally straight part of the curve, instead of trying to fit a single Gaussian that would include a lot of erroneous shapes (white space). Similarly, the initialization should also be image-specific since the correct segmentation of that image is likely the best initialization. Though different images need different parameters, we purport that similar images will require similar parameters (See Section 3.4 for details).

Recent research has begun to address the idea that groups of similar images embedded in $\mathbb{R}^N$, where $N$ is the number of image pixels, lie on manifolds [137]. However, there has been little work on performing segmentation in a way that respects the image manifold, i.e. method parameters reflect the data's possibly non-linear variability and the energy functional is custom tailored such that its minima lie on the corresponding points of the segmentation manifold to the input images (Fig. 3.3).

The aforementioned problem of selecting parameters and initializing models is a serious one that remains unsolved. This work addresses the problem. The goal is to produce energy functionals with global, or at least local, minima for every (*ground truth segmentation*,*image*) pair; since anything less yields a suboptimal solution for some of these pairs (Fig. 3.3). Our approach is to expressly modify the energy functional with every new image to be segmented by varying its parameters.

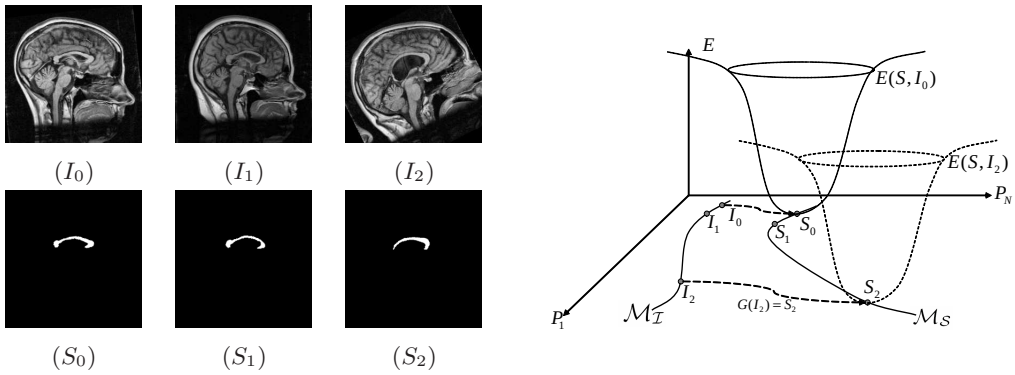



Figure 3.3: Ideal energy functionals $E(shape, image)$. Let graylevel image $I$ and segmentation $S$ be represented as images with $N$ pixels ($P_1$ to $P_N$). Then $\{I_1, I_2, ..., I_{\mathbb{N}}\} \in \mathbf{I}$, and $\{S_1, S_2, ..., S_{\mathbb{N}}\} \in \mathbf{S}$, are sets of images and corresponding segmentations where each element is a point in $\mathbb{R}^N$. Samples from $\mathbf{I}$ and $\mathbf{S}$ lie on manifolds $\mathcal{M}_{\mathbf{I}}$ and $\mathcal{M}_{\mathbf{S}}$, respectively. Three ($S$,$I$) pairs (left panel) positioned on their respective manifolds (right panel). Segmenting an image is represented by a smooth mapping $G : \mathbb{R}^N \rightarrow \mathbb{R}^N$, since for each $I_i$ there exists a corresponding segmentation $S_i$. Similar images, $I_0$ and $I_1$, are nearby on $\mathcal{M}_{\mathbf{I}}$ and require similar segmentations, $S_0$ and $S_1$. The ideal energy functional $E(S, I_0)$, is a convex functional minimized at the segmentation $S_0$ on $\mathcal{M}_{\mathbf{S}}$.

### 3.1.3 Contributions and Related Work

Based on the previous discussions our hypotheses are three fold. First, we believe that changing the weights will have a dramatic effect on the accuracy of a given method. Perhaps not a bold hypothesis, but an analysis of the effect of weights is often not presented in MIS papers. Second, we believe that learned parameters will outperform manual parameters. Again, a reasonable assertion, but various works continue to set parameters by hand, by trial and error, or by brute force. Third, where sufficient training data is available, inferring the weights on a per-image–basis is superior to asserting that a single set of weights will work for all novel image data of the same class of imaged objects. Our contribution is thus an exploration of these three hypotheses, and a novel method that answers these questions.

To the best of our knowledge there are only a few papers on a related topic. In [48], Gennert and Yuille find weights that cause the segmentation (i.e. the minimum of the energy functional) to have maximal energy, which does not address the need for ground truth segmentations to be minima of the resulting functional. Though we are focusing on continuous functionals, a related field of approaches has come up in the study of combinatorial problems. The first set is based on recent advances in maximum margin estimation, wherein

the parameters of the objective function are sought such that the ground truth structures (in our case segmentations) have the least possible amount of energy in comparison to erroneous structures (segmentations) [152, 2, 150, 44]. As a natural contrast to max-margin estimation, maximum-likelihood estimators have also been explored for weight optimization [113]. However, in addition to being limited to combinatorial objective functions rather than continuous ones, these methods [152, 2, 150, 44, 113] propose a fixed set of parameters for novel samples (in our case images), whereas we provide a method using geodesic interpolation to infer the optimal parameters on a per-image basis. In other words, these works assume that a single set of parameters works for an entire test set. As demonstrated in Fig. 3.2 this is often not the case in image segmentation. Greatly improved results can be obtained by adapting the parameters to the individual images, as we propose.

Other methods seek to learn objective function weights, but instead of graph-theoretic models, focus on active shape models (ASM) [173, 172, 166]. Zhang *et al.* develop a method for learning a conditional density function over model parameters, which can then be locally optimized to yield the desired segmentation [173, 172]. They propose different approaches in [173] and [172], based on regression and ranking methods of learning, respectively. Wimmer *et al.* also learn functions over the model parameters using regression methods [166]. However, similar to the graph-theoretic models above, these approaches propose a single set of weights for all novel images, as opposed to learning parameters on a per-image basis. Furthermore, they are designed specifically for ASM, and may not be generally applicable to implicit deformable models like level sets [136, 17, 18, 157] as the dimensionality of the energy-functional is dramatically higher in the later.

Another related direction was introduced recently in [73]. Even though this method is developed to find suitable weights on a per-image basis, it is restricted to combinatorial objective functions and is also quite different than ours in design. Given a range of weights, Kolmogorov *et al.*'s method simultaneously solves the objective function for a set of weights that bound how the weights influence the solution. Each solution is then treated as a potential segmentation. They propose a number of heuristics, including user intervention, to select the best segmentation from a set of potential ones. In contrast our method seeks to analytically determine the optimal weights for training data, and then automatically infers those weights for novel images, thus automating the process. Lastly, Koikkalainen *et al.* [72] use a nearest neighbors approach to initialize the segmentation procedure but do not define optimal parameters for the model, nor do they make use of manifold learning to calculate

distances, perform interpolation, or obtain optimal parameters for the model.

As noted in Chapter 2, a somewhat related direction is that of multiobjective optimization, where each term is simultaneously optimized rather than optimizing a linear sum of the terms (3.1). Hence, in multiobjective optimization no weights are provided to combine the competing terms of the functional, instead a solution is sought for which no term can be improved without sacrificing another. That solution is known as a *Pareto optimal* solution [25]. The set of Pareto optimal solutions for a given problem is known as the *Pareto front*, and there is no preference amongst them unless a ranking is provided between the objectives. Nakib *et al.* recently used multiobjective optimization to determine the parameters for a thresholding algorithm for image segmentation [109]. Hanning *et al.* present an approach using a piecewise approximation of the image similar to the Mumford Shah model [108] using multiobjective optimization to decide the trade off between the number of segments and the image approximation error [65]. However, as none of these works has weights to modify, there is no inherent way to encourage the ground truth to be a minima of the resulting functional. It is also interesting to note that, for convex functions, if the ground truth lies on the Pareto front, then by definition a set of weights must exist that causes the optima to be the ground truth for the linear sum of terms formulation (3.1) [49]. Though multiobjective optimization removes the problem of setting the weights, in doing so it also removes the ability to tune the energy functional for a given image.

Common to all of the aforementioned works is the acknowledgment that ad-hoc parameter tuning is a serious problem in image segmentation, and that learning weights is superior to hand-tuning. However, a study and corresponding method, on varying the weights on a per-image basis has not been presented. In this work, we present such a study and propose a method that allows weight learning on a per-image basis.

The main contributions of this work are, thus, as follows. To the best of our knowledge ours is the first work to: (1) analytically derive the optimal energy functional weights of competing energy terms for training $(segmentation, image)$ pairs; (2) explore the need for varying the energy functional in energy-minimizing segmentation techniques; (3) calculate optimal parameters of novel images utilizing learned image manifolds; (4) provide a general formulation that can be directly applied to any energy minimizing segmentation technique (e.g. [17, 18, 76, 126]) to improve its segmentation and increase its ability to generalize to larger datasets; (5) to fully automatically initialize and tune parameters optimally, with results presented on a large ($\mathbb{N} = 470$) data set of images; (6) to provide the automatic

cropping of images prior to segmentation for a 10 to 20 times speedup (more details on this contribution will be presented in section Section 3.4).

In what follows we describe the theory behind our method (Section 3.2), and how to analytically describe the optimal weights of a functional (Section 3.3). We then detail how to apply our method to novel images (Section 3.4). Finally, we provide results in Section 3.5, and discuss our results and future work (Section 3.6).

## 3.2 MIS: Background and Foundations

As previously noted in Section 2.1.5, the Euler Lagrange provides the necessary condition, under an $L_2$ norm assumption, for the first variation of $E$ to equal 0, i.e. a stationary point. However, different methods use different norm assumptions, or even first derivatives in place of first variations when $E$ is simply a function. To remain abstract of which method for differentiation is chosen, we're going to write the functional derivative, which is a generalization of the gradient, of $E$ with respect to $S$ as $E_S(S, I, \mathbf{w}) = w_1T_1(S, I) + w_2T_2(S, I) + ... + w_nT_n(S, I)$, where $T_i$ is the functional derivative of term $J_i$ (think gradient, but for functionals). This could be calculated, for example, by computing the Euler Lagrange equations for $E$, provided one assumes the space of functions, $S$, to be locally Euclidean (i.e. for small changes in $S$, an $L_2$ norm is an accurate distance metric).

If not appropriately set, the weights, $\mathbf{w}$, can cause significant error. In fact, our results show that optimizing the weights can have dramatic effects, reducing error in large data sets by as much as 60% (Section 3.5.2). However, optimizing the weights by hand for even a single image can be a long and tedious task, with no real guarantee of obtaining the correct segmentation.

Instead of *guessing* the optimal weights, suppose we write a function $\gamma(\mathbf{w}|I_j, S_j)$ evaluating how well weight $\mathbf{w}$ works for a given image-segmentation pair $(I_j, S_j)$ such that a parameter is deemed better when it causes $S^*$ in (2.1) to approach $S_j$, i.e. causes the minimum of $E$ to be the correct segmentation. Given $S_j$, we could then calculate the ideal weights for a particular image $I_j$ by solving $\mathbf{w}^* = \arg\min_{\mathbf{w}} \gamma(\mathbf{w}|I_j, S_j)$. It is important that $\gamma$ itself be convex or globally solvable in $\mathbf{w}$. If $\gamma$ was not globally solvable, uncertainty would remain in that some other $\mathbf{w}^*$ may better minimize $\gamma$, and thus better segment the image.

Of course, the above definition of $\mathbf{w}^*$ only works for a single image-segmentation pair $(I_j, S_j)$, where the ground-truth segmentation is already known. What we propose is to learn

a function, $\mathbf{w}(I) : \mathbb{R}^N \rightarrow \mathbb{R}^n$, i.e. a mapping from images to weights, which will enable us to estimate the optimal weights $\mathbf{w}^*$ for a novel image. For notational convenience, we define the optimal weights for $I_j$ as $\mathbf{w}^*(I_j)$, and the estimated weights from our learned function as $\mathbf{w}(I_j)$. Where $\mathbf{w}$ appears without an argument, $I$, it is simply a vector of weights. We will also write $w_1(I_j)$ when referring to only the $1_{st}$ weight, and so on. Defining $\mathbf{w}$ as a function of $I$ enable us to write a new form of energy functional:

$$\begin{aligned} E(S, I) = \quad & w_1(I)J_1(S, I) + w_2(I)J_2(S, I) + ... \\ & + w_n(I)J_n(S, I). \end{aligned} \tag{3.2}$$

## 3.3 Optimal Energy Functional Weights

For each $(I_j, S_j)$, where $I_j \in \mathbf{I}$ and $S_j \in \mathbf{S}$, the task is to find the optimal values for the free weights $\mathbf{w}^*(I_j)$, so that we can later learn the mapping $\mathbf{w}(I)$. This section explores the notion of 'optimal'.

One computationally intractable approach for finding $\mathbf{w}^*(I_j)$ is to try all possible weight combinations, run the segmentation method, and then select the weights with the least segmentation error.

A better approach is to find the weights $\mathbf{w}^*$ that minimize the magnitude of the derivative, in our case $E_S$, of the energy functional at the correct segmentation $S_j$. Doing so encourages $S_j$ to be a minimum of $E$ (i.e. $E_S(S_j|I_j, \mathbf{w}^*) = 0$). We measure the magnitude as $\|E_S(S_j|I_j, \mathbf{w})\|^2$, and seek the weights $\mathbf{w}$ that cause $\|E_S(S_j|I_j, \mathbf{w})\|^2 \rightarrow 0$.

In addition to encouraging $S_j$ to be a minimum of $E$, we must also discourage incorrect segmentations from being minima as well. Therefore we minimize the magnitude $|E_S(S_j|I_j, \mathbf{w})|^2$ for $S_j$ while maximizing it for all other possible shapes (in a direction toward the optimal solution). For a given shape $S_i$ a vector $(S_i - S_j)$ in $\mathbb{R}^N$ represents the direction towards $S_j$. Since $E_S(S_i|I_j, \mathbf{w})$ is a vector in $\mathbb{R}^N$ dictating in what direction, and by what amount, the solution will change at the point $S_i$, a normalized dot-product (projection-like approach) will measure how much in the right direction $E_S(S_i|I_j, \mathbf{w})$ points.

Combining these two goals (minimize derivative for the ground truth, maximizing it for erroneous segmentations), for an energy functional with a form like that in (3.2), we can

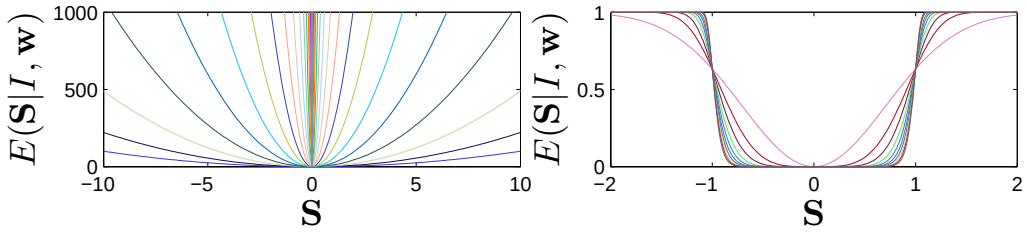



Figure 3.4: Varying the shape of energy functionals. Left: Various functionals that show an increasing extent to which the gradient at neighboring shapes points towards the correct segmentation, represented by $S = 0$. Right: Various functionals that show how a set of neighboring shapes can also become minima, a degenerate case.

define $\mathbf{w}^*(I_j)$ as:

$$\begin{aligned} \mathbf{w}^*(I_j) = \quad & \arg\min_{\mathbf{w}} \|E_S(S_j|I_j,\mathbf{w})\|^2 \\ & -\lambda \sum_{i\in\mathcal{N}_\mathbf{S}} F_{\mathcal{N}_\mathbf{S}}(S_j,S_i) E_S(S_i|I_j,\mathbf{w}) \cdot \tfrac{(S_i-S_j)}{|S_i-S_j|} \\ s.t. \quad & \mathbf{w} \geq \mathbf{0} \\ & \mathbf{w}'\mathbf{1} = 1, \end{aligned} \tag{3.3}$$

where $\mathcal{N}_\mathbf{S}$ denotes a set of nearby (or similar) shapes in the domain of $E$, and $F_{\mathcal{N}_\mathbf{S}}$ is used to weight closer segmentations according to their proximity. The neighborhood $\mathcal{N}_\mathbf{S}$ is used instead of the entire shape space to reduce computational complexity, and can be estimate by manifold learning (see Section 3.4). The second term, dubbed the neighborhood term, is negative and has $\|E_S(S_i|I_j,\mathbf{w})\|$ omitted from the normalized dot-product to reward large steps in the correct direction. Solving (3.3), yields the optimal weights for image $I_j$. As (3.3) can be written as a convex, quadratic, solving for $\mathbf{w}^*$ is a fast, convex optimization problem (See Appendix C for details).

However, notice that there is a weighting, $\lambda$, between competing terms of (3.3). A balance must be struck between: $(i)$ the degree by which $S_j$ is a minimum of $E$; and $(ii)$ the degree by which the derivative at neighboring points in the shape space points towards $S_j$ (Fig. 3.4-left). Make $\lambda$ too small and $S_j$ might be a minimum, but so might the entire neighborhood (Fig. 3.4-right). Make $\lambda$ too large and the neighborhood will point in the right direction, but $S_j$ might no longer be a minimum.

One way to avoid $\lambda$ is to shift the second term into a constraint rather than a weighted term. In order to do so, we make use of recent research into convex functionals for image

segmentation, where a convex energy functional, $E$, is minimized over a convex shape space represented as probability maps, i.e. $S(x) \in [0,1]$ for all points $x$ in the image domain $\Omega$, to yield a convex segmentation problem [35]. When $E$ is convex our only concern appears to be making $S_j$ as much a minimum as possible, since there are no local minima. As a result we can replace the neighborhood term by a *constraint rather than a cost term* since the degree to which the neighbors point towards $S_j$ does not change whether or not $E_S(\mathbf{w}|S^j, I_j) = 0$ (i.e. we must avoid forcing the gradient in the neighborhood to point at $S_j$ at the cost of making $E_S(\mathbf{w}|S^j, I_j) \neq 0$). Thus instead of (3.3) we re-define $\mathbf{w}^*(I_j)$ as

$$\begin{aligned} \mathbf{w}^*(I_j) = \quad & \arg\min_{\mathbf{w}} \|E_S(S_j|I_j, \mathbf{w})\|^2 \\ s.t. \quad & \sum_{i \in \mathcal{N}_\mathbf{S}} F_{\mathcal{N}_\mathbf{S}}(S_j, S_i) E_S(S_i|I_j, \mathbf{w}) \cdot \tfrac{(S_i - S_j)}{|S_i - S_j|} \geq 1 \\ & \mathbf{w} \geq \mathbf{0} \\ & \mathbf{w}'\mathbf{1} = 1. \end{aligned} \tag{3.4}$$

The result is a convex function in $\mathbf{w}$ under a linear constraint. A convex function under a linear constraint can be solved via convex optimization, and thus the optimal $\mathbf{w}(I_j)$ is guaranteed. It is entirely possible for an energy functional to be unable to satisfy the constraint regardless of $\mathbf{w}$, although we did not encounter this situation for even a single result. In cases where the constraint cannot be satisfied a slack variable can be readily introduced while maintaining convexity. In non-convex methods (3.3) would be used instead, as the magnitude of the neighborhood term is important, as opposed to simply ensuring it remains greater-than-or-equal-to one. Note that written in these forms, (3.4) is not equivalent to (3.3) with $\lambda = \inf$, because the later would completely ignore minimizing $\|E_S(S_j|I_j, \mathbf{w})\|^2$ in favor of minor improvements in the neighborhood term. Slack variables could be used to write a single equation for both formulations, but the formulation would be less direct for point of presentation.

Though not obvious at first glance, the segmentation, $S^*$, from energy functionals like (3.1) is a maximum a posteriori (MAP) solution [149, 87], which is a regularized form of maximum-likelihood estimation. The proposed methods of weight inference are designed to minimize the derivative of the energy functional at the ground truth and thus make it the MAP estimate. It is, therefore, natural to compare our model to a max-margin model [150].

We now briefly outline the max-margin approach outlined in [150], as we will provide experimental comparisons in Section 3.5. Intuitively, the method seeks to find the weights

that cause the correct segmentation to have lower energy than any other segmentation, thus forcing it to be a minima of $E$. Mathematically, this means to cause $S_j$ to be a minimum of $E(S_j|I_j, \mathbf{w})$ by constraining $E(S_i|I_j, \mathbf{w}) - E(S_j|I_j, \mathbf{w}) \geq \epsilon$ for all possible shapes $S_i \neq S_j$, while making $\epsilon$ as large as possible. The quantity $E(S_i|I_j, \mathbf{w}) - E(S_j|I_j, \mathbf{w}) \geq \epsilon$, is known as the margin. The maximization problem can instead be written as a minimization problem as follows:

$$\begin{aligned} \mathbf{w}^*(I_j) = \arg\min_{\mathbf{w}} & \frac{1}{2}\|\mathbf{w}\|^2 + C\left(\frac{1}{|\Psi|}\right)\sum_{i}^{|\Psi|}\xi_i \\ s.t. \quad & \mathbf{E}_i\mathbf{w}' \geqslant \Delta(S_j, S_i) - \xi_i \quad \forall S_i \in \Psi \\ & \boldsymbol{\xi}, \mathbf{w} \geq \mathbf{0} \end{aligned} \tag{3.5}$$

where

$$\mathbf{E}_i = E(S_i, I_j) - E(S_j, I_j) = \{J_p(S_i, I_j) - J_p(S_j, I_j)\}_{p=1}^{N}, \tag{3.6}$$

$\Psi$ is the set of all possible segmentations, $|\Psi|$ is the cardinality of the set, $\boldsymbol{\xi} = \{\xi_i\}_{i=1}^{|\Psi|}$ is a set of slack variables that allow some segmentations to have negative margin, $\Delta(S_j, S_i)$ is a distance metric measuring the difference between two segmentations, and $C$ is a free parameter controlling the trade off between maximizing the margin and penalizing the slack. However, it is not feasible to compute the set of all possible shapes, and thus an approximation algorithm for the constraints must be used (see [150] for details). It is important to note that in order to compute the approximate constraint set, an iterative procedure is used, whereunder the image must be segmented using the current guess at $\mathbf{w}^*$, yielding either $S^* = S_j$ or $S^* \neq S_j$. If $S^* \neq S_j$, then $S^*$ becomes a member of $\Psi$, if $S^* = S_j$ the iterative procedure terminates. The constraints, if satisfied, avoid the degenerate case (Fig. 3.4-right), but the method does not directly seek to make $S_j$ a minimum of $E$. The problem with (3.5) is subtle, but simple. By introducing slack variables, $\boldsymbol{\xi}$, a balancing act is introduced like in (3.3).

## 3.4 Segmenting Novel Images

Thus far we have considered calculating the optimal weights for images and energy functionals with known segmentations (using (3.3) or (3.4)). Here we turn our attention to

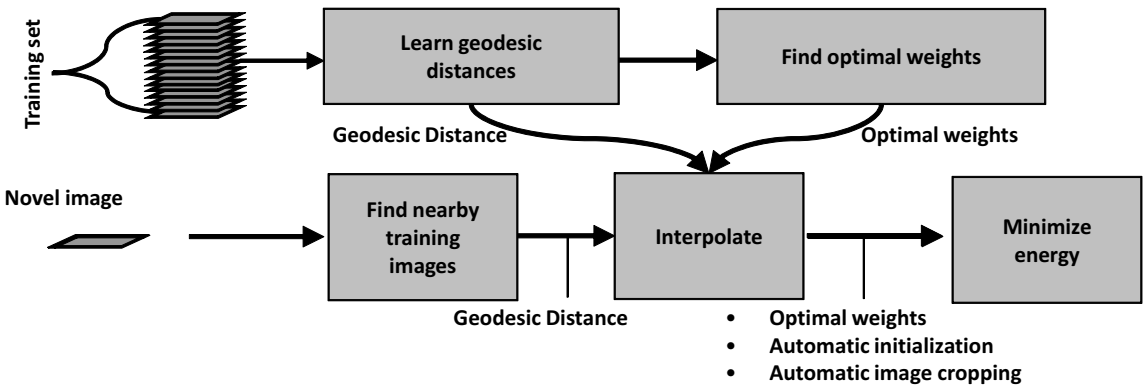



Figure 3.5: Overview of our proposed weight optimization method.

optimizing the weights for novel images.

The first approach is simple. Given our set of training image-segmentation pairs we will have $\mathbb{N}$ samples of $\mathbf{w}^*(I)$, we can average the resulting weights and use the obtained average for any novel image. If we artificially generate the neighborhood, $\mathcal{N}_\mathbf{S}$, using [56], and assume a uniform distance function, $F_{\mathcal{N}_\mathbf{S}}$, then this approach requires no manifold learning of any kind, but it does not address the problem of different images being better segmented under different weights.

Our second approach is to interpolate over a learned manifold to find values at novel images. In order to interpolate, we need a metric for measuring distances between images, see below. The set of images with the shortest distances constitutes the neighboring images, $\mathcal{N}_\mathbf{I}$. The corresponding ground truth segmentations are denoted $\mathcal{N}_\mathbf{S}$.

We assume $\mathbf{I}$ is smooth over its domain, the space of a particular class of images (e.g. MRI brain scans of normal adults), and that the mapping from images to segmentations is smooth. In other words, we assume that similar images have similar parameters, and similar segmentations. As such, we use a normalized Gaussian kernel, defined over the image distances, to interpolate both the parameters and initializations. For shape and appearance priors, we limit the training data to $\mathcal{N}_\mathbf{I}$ and $\mathcal{N}_\mathbf{S}$, since we are more confident that the correct shape and appearance information is similar to the training data lying in those neighborhoods. The process is summarized in Fig. 3.5.

Manifold learning methods are a special class of nonlinear dimensionality reduction techniques that enable the calculation of geodesic distances between data points [153]. We make use of these techniques to calculate distances between both images and segmentations.

Distances between neighboring segmentations allow us to define

$$F_{\mathcal{N}_\mathbf{S}}(S_j, S_i) = 1 - \frac{g\left(S_i, S_j\right)}{\sum\limits_{t \in \mathcal{N}_\mathbf{S}} g\left(S_t, S_j\right)} \tag{3.7}$$

where $g(S_i, S_j)$ is the learned geodesic distance between shape $S_i$ and shape $S_j$. The function $F_{\mathcal{N}_\mathbf{S}}(S_j, S_i)$ then acts as a weighting for the given neighborhood $\mathcal{N}_\mathbf{S}$, normalized to sum to one (i.e. the weight decreases as a function of distance from the center of the neighborhood).

An added benefit of finding the location of the novel image on the image manifold is that an approximate segmentation can be obtained by interpolating over $\mathcal{N}_\mathbf{S}$. Doing so yields a good initialization for the energy-minimization process, and also allows a region of interest to be automatically cropped out of the image, thereby reducing the image size and increasing run speed. There is, however, a caveat to this discussion. Either the images need to be affine registered, or the manifold needs to be trained sufficiently enough to include degrees of freedom corresponding to variations in pose. For our results in this chapter we chose to affine register the images because the off-the-shelf energy functionals we adopt, specifically the shape priors, do not adequately reflect pose variations [35, 36]. Another option would be to forgo auto-cropping, use a pose-invariant manifold learning method, say one built on scale and rotation invariant Gabor filters [144], and then use a pose-invariant segmentation method, like the one we develop in Chapter 4.

Though numerous points of uncertainty have been addressed by our method, there are a few remaining free parameters: (*i*) the dimensionality of the segmentation manifold; (*ii*) the dimensionality of the image manifold; (*iii*) the input image-distance metric used as input to the manifold learning algorithm; (*iv*) the manifold learning algorithm to be used; and finally (*v*) the interpolation function used to determine parameters for novel images, as a function of their distance to similar images. As these choices are somewhat application dependent, we include a specification of their values in the results section.

## 3.5 Results

We ran two primary sets of experiments. The first set, deemed the **Training** experiments, explores which $\gamma(\mathbf{w})$, as defined in section 3.3, is the best approach for obtaining the weights. The second set, the **Testing** experiments, is designed to demonstrate the impact that inferring the optimal weight on a per-image-basis has on segmenting highly variable data sets, in

comparison to state-of-the-art approaches of using a fixed set of weights for an entire data set.

All error results in each set are computed using the Jaccard distance (4.7) [69]. Where run-times are presented, they are on a 2 GHz AMD Opteron, with method implementations in MATLAB® (2010a, The MathWorks, Natick, MA).

Unless otherwise noted, all results, including those for competing methods, use: the same gradient descent optimizer for minimizing $E(S|I, \mathbf{w})$; the same training data; as well as our automatic cropping and initialization features. These steps ensure that our advantages over competing methods are due to our weight optimization, as opposed to differences in optimization, training data, or other features of our manifold-learning. As noted, the proposed (3.3), includes a free parameter $\lambda$ as does the max-margin approach, (3.5), [150]. In order to ensure proper comparison between training and testing experiments, these parameters were also fixed for all experiments. Both parameters were tuned by inspection over the range $(0, \infty)$, starting at approximately 0 and increasing until a decline in training performance was obtained. For (3.3), $\lambda = 0.0333$, and for (3.5), $C = 1$. The max-margin method, (3.5), also requires a distance metric, $\Delta(S_j, S_i)$, which we set to the Hamming loss function as suggested in [150].

### 3.5.1 Training Experiments

In order to explore which $\gamma(\mathbf{w})$ is the best approach for obtaining the weights, we perform our experiments on two medical data sets. After we conclude which $\gamma(\mathbf{w})$ is optimal, we demonstrate the impact weight optimization can have on MIS. Note that for all experiments in this section the ground truth segmentation is assumed known, hence why we call these training experiments, they measure each weight optimization technique's ability to find the best weights possible for the training data. In the next section complete leave-one-out validation will be performed. Note, however, that for each image, both the original image $I_j$, the ground truth, $S_j$ are *excluded* from the shape and intensity training, respectively.

The first set is comprised of 470 $256 \times 256$ affine registered mid-sagittal MRI, with corresponding expert-segmented corpora callosa (CC) (Fig. 3.3). The CC is a bundle of white matter fibre whose cross sectional area and shape correlate with neurological diseases such as multiple sclerosis and schizophrenia [84]. The vast amount of data acquired in clinical studies of these diseases motivates the necessity of a reliable CC segmentation algorithm.

The second is a set of 14 $256 \times 199$ affine registered, axial MRI with corresponding expert-segmented left- and right- lateral ventricles. Like the CC, the lateral ventricles play an important role in brain function, and are a desirable structure for analysis in many disease studies. Note, for space and ease of presentation, and unless otherwise noted, all results will be presented as $\varepsilon = x, y$ where $x$ are results for the CC data, and $y$ the lateral ventricle data, since the same experiments are being run on both data sets. For clarity, tabular results are presented separately.

In order to simplify the comparisons between different $\gamma(\mathbf{w})$ functions, we focus our experiments in this section on convex functions. By focusing on convex functions we eliminate any issues relating to initializations, and instead focus on which method best fits the training data. In particular, we adopt the approach of [35], where the shape $S$ is described using a PCA shape space with parameters, $\alpha$ weighting the main-modes of variation. See [35] for implementation details. In order to learn a PCA shape space for the ventricle data, we use the method of [56] which synthesizes additional training shapes based on exemplar segmentations. Synthetic shapes created from $S_j$ are removed from the training set for all experiments on image $I_j$.

Our energy functional takes form as a boundary term, a regional term, and a shape regularization term:

$$\begin{aligned} E(\alpha) = \int_{\Omega} & \left(w_1(I)h(x)\,|\nabla S(x)| + \right. \\ & \left. w_2(I)(f(x)S(x) + g(x)(1 - S(x)))\right) dx + w_3(I)\alpha^T \Sigma^{-1} \alpha \end{aligned} \tag{3.8}$$

where $\alpha$ are the PCA shape parameters describing $S$, $f = -log(P_{obj}(I))$, $g = -log(P_{bk}(I))$, for object and background intensity prior distributions $P_{obj}, P_{bk}$, $h = \frac{1}{1+|\nabla \mathrm{I}|}$, and $\Sigma^{-1}$ characterizes the allowable shape distribution (see [35] for details). Notice that the energy takes the general form $energy(shape, image) = w_1 \times boundary(shape, image) + w_2 \times region(shape, image) + w_3 \times shapePrior(shape)$, as before. The boundary term is perhaps one of the most canonical boundary terms, taken from geodesic active counters (GAC) [17]. The region term is also a popular term based on the Mumford-Shah function [108], adapted from [18]. In-order to compute the image neighborhood we use $k$-ISOMAP with $k = 10$, and reduce the image space to a 5-manifold for each data set, chosen for being the elbow both scree plots. For PCA on the shape spaces we retain $K = 15$ modes of variation for the CC data and $K = 140$ for the ventricle data, chosen for being the elbows of the corresponding scree plots. Note that the two data sets are processed completely independently with their

own manifolds, shape spaces, intensity priors, etc.

Table 3.1: Error comparison between tested methods for CC images.

| | **Method** | **mean** $\varepsilon$ | **min** $\varepsilon$ | **max** $\varepsilon$ | **std. dev.** $\varepsilon$ |
|---|---|---|---|---|---|
| Training | Szummer *et al.* (3.5) [150] | 0.143 | 0.049 | 0.590 | 0.080 |
| | Proposed (3.3) | 0.119 | 0.047 | 0.429 | 0.050 |
| | Proposed (3.4) | 0.116 | 0.047 | 0.429 | 0.049 |
| Testing | Szummer *et al.* [150], fixed weights | 0.421 | 0.185 | 0.980 | 0.158 |
| | Proposed (3.4), inferred weights | 0.121 | 0.049 | 0.465 | 0.056 |

Table 3.2: Error comparison between tested methods for ventricle images.

| | **Method** | **mean** $\varepsilon$ | **min** $\varepsilon$ | **max** $\varepsilon$ | **std. dev.** $\varepsilon$ |
|---|---|---|---|---|---|
| Training | Szummer *et al.* (3.5) [150] | 0.424 | 0.244 | 0.796 | 0.150 |
| | Proposed (3.3) | 0.276 | 0.184 | 0.514 | 0.084 |
| | Proposed (3.4) | 0.277 | 0.190 | 0.516 | 0.082 |
| Testing | Szummer *et al.* [150], fixed weights | 0.370 | 0.207 | 0.774 | 0.210 |
| | Proposed (3.4), inferred weights | 0.278 | 0.195 | 0.529 | 0.084 |

Before running any experiments, we account for error relating to the decomposition of the shape spaces into a linear combination of eigenvectors by computing the error for the reconstructed ground truth segmentations. Using the Jaccard distance metric, $\varepsilon$, for $K = 15$ and $K = 140$, we found that $\varepsilon = 0.1136, 0.0329$ respectively, averaged independently across the data sets (i.e. $\varepsilon = 0.1136$ for the CC data, and $\varepsilon = 0.0329$ for the ventricle-data). This is effectively a lower-bound on how accurate our results with this segmentation method can be.

Our experiments in this section are two-fold: we first measure how well the proposed, (3.4), and the max-margin, (3.5), [150] perform for an idealized energy functional; second, we measure performance for a realistic energy functional. Note that all experiments in this section include the image to be segmented in the training data, and segment each test image using the weights learned exactly for that particular test image. Though this may seem peculiar, the point of this section is to measure how well eqs. (3.3), (3.4), and

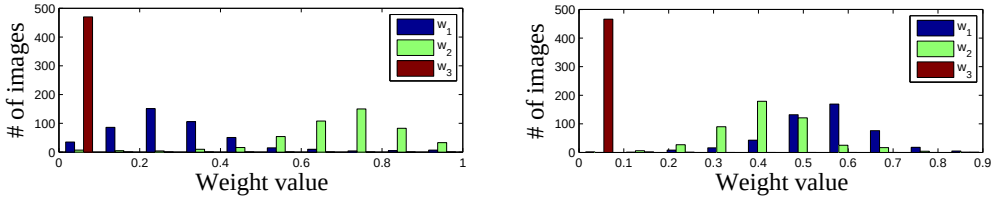



Figure 3.6: (Color Figure) Histograms of learned weights for the CC-images, using (3.4) (left), and the max-margin approach, (3.5) [150] (right). Note in both methods there is a great deal of variability, suggesting that a single set of weights will not perform well for a large data set. Note that in general our method weights the region information ($w_2$) higher, whereas the max-margin approach weights the boundary term ($w_1$) higher.

(3.5) perform, not how well a given energy functional, or method of inferring weights for novel images performs. In other words, these experiments directly compare each equation's ability to cause $S^*$ to approach $S_j$ for each $(I_j, S_j)$, where $I_j \in \mathbf{I}$ and $S_j \in \mathbf{S}$. We call these experiments *'residual error'* experiments, because they represent the error left over after everything has been optimally setup for the energy functional.

To measure how well eqs. (3.3), (3.4), and (3.5) can balance an ideal energy functional (one for which the minimum can be positioned exactly at the correct segmentation using the optimal weights) we calculate the error of energy functional (3.8) plus a fourth term whose unique global minimum is always the correct segmentation for the given image (i.e. we gave the functional weights $\mathbf{w}$ the power to achieve 100% accuracy by adding a rigged, strictly convex term with an optimum at exactly the ground truth for each image). Using this rigged term, and eqs. (3.3) and (3.4) we obtained an average error of $\varepsilon = 0.1136, 0.0329$; exactly the lower bound noted earlier, validating that our method can achieve the full potential of a given functional. Using (3.5), we again obtain an average error of $\varepsilon = 0.1136, 0.0329$. Examining the resulting weights vectors, we note that from both (3.3) and (3.4), we obtain $\mathbf{w}^* = [0\,0\,0\,1]$, whereas (3.5) yields $\mathbf{w}^* = [0.0161\,0.0067\,0.0050\,0.9723\,]$. The minor difference in weight vectors can be attributed to (3.5) opting to further increase the margin by including portions (non-zero weights) of other terms. However, examining the run-times, we see that our proposed method, eqs. (3.3) and (3.4), are much faster to optimize than the max-margin technique, (3.5), as the latter is an iterative procedure requiring the engery functional to be optimized at each step, while our method requires only the solution to a single convex-quadratic. Specifically, (3.5) took on average, $5.85, 3.21$ iterations-per-image to converge, resulting in an average runtime $92.8, 65.7$ seconds, versus one iteration and $0.512, 0.284$

seconds for both eqs. (3.3) and (3.4). Of course, a faster segmetnation method would speed up (3.5), but at $18.4, 20.47$ seconds-per-image the method is reasonably fast, and even at 1-second-per-image (3.5) would still take $5.85, 3.21$ seconds, much slower than $0.512, 0.284$ seconds-per-image, especially when training large data sets.

Finally, to measure how well eqs. (3.3), (3.4), and (3.5), can balance a real energy functional we calculate the error of energy functional (3.8) using each equation. Using energy functional (3.8) with (3.3) we obtained an error of $0.119, 0.276$, whereas with (3.4) we obtained $\varepsilon = 0.116, 0.277$. Using (3.5) $\varepsilon = 0.143, 0.424$; an increase in error of about 23% and 53%, respectively. Results are summarized in tables 3.1 and 3.2, for the CC data and the ventricle data, respectively. Though the difference between eqs. (3.3) and (3.4) is minor, (3.3) could suffer from poor choices of $\lambda$, and so where applicable, i.e. when using convex energy functionals, we focus on (3.4). As there is only a negligible change in run-times from the previous experiment, we do not re-include those results here.

To further examine the results, we present the histograms of the learned weights for the CC data using eqs. (3.3) and (3.5) (Fig. 3.6). There are two interesting results to note. Keep in mind, all weights are kept positive, to ensure convexity of (3.8), and normalized to sum to one, since all scalar multiplies of $\mathbf{w}$ imply the same segmentation. Firstly, both methods kept the weight on the shape prior relatively low, but they disagreed on the weights for the region and boundary terms ($w_1$, $w_2$, respectively). Our proposed method, (3.4), prefered to give the bulk of the weight to the region term, while the max-margin approach [150], (3.5), weighted them more equally. Secondly, both methods resulted in a high degree of variability in the weights, implying that a single $\mathbf{w}$ will not perform well for an entire data set, but rather should be varied across the data set. We further explore this in the next set of experiments.

Returning to the topic of picking the best equation to optimize the weights, from these results we concluded that (3.4) is the faster, more accurate choice for weight optimization. Finally note that, eqs. (3.3) and (3.4) can be used to choose both the terms and the weights of a given energy functional, since setting a term's weight to zero is equivalent to not choosing it.

### 3.5.2 Testing Experiments

In the last section we validated our proposed weight optimization equations in comparison to the max-margin equation proposed in [150], regardless of the inference method used to

obtain weights for novel images. In this section we validate the amount of influence optimal weights can have on the accuracy of an energy functional when optimized for each image, and validate our proposed method of weight inference on a per-image-basis. We conduct four main experiments.

Before detailing our experiments we note that as this chapter is about the application of manifolds to segmentation, issues related to learning the manifold will not be addressed in this work. Our results demonstrate that these manifolds can indeed be learned, even for small data sets. We will also show that the manifold learning itself contributed only a small fraction of the error, and thus focus for improving results should be on the energy functions themselves not the manifold learning process.

Our first experiment tests our hypothesis that similar images will have similar weights, and that those weights can indeed be learned. We tested our method on a synthetic data set of 225 noisy ellipse images, where only the major and minor axes lengths of the ellipses are varied, i.e. a 2-manifold. Keeping the variability low allows direct visualization of the resulting manifolds. To learn the manifolds we used a MATLAB implementation of k-ISOMAP [153] from http://isomap.stanford.edu/. Sample images are shown in Fig. 3.2 (c) and (d), while (a) shows how the smoothness weight $w$ varies gradually as a function of the manifold, with more eccentric ellipses having the least smoothness requirement. For this example we set $k = 7$ in k-ISOMAP, and reduce the image space to a 2-manifold. Adopting the Chan-Vesse energy functional of [18] and using leave-one-out validation, we compare between two scenarios: ($i$) fixed initialization with the average ellipse and fixed parameters obtained by averaging the set of optimal parameters; and ($ii$) adaptively chosen optimal parameters and initializations as functions of the coordinates of the novel image on $\mathcal{M}_{\mathbf{I}}$. We found a 47.1% reduction in average error using optimal parameters over fixed (0.190 vs 0.358). This means that almost 50% of the error was due to erroneous parameters and/or initializations.

In what follows we demonstrate our method on real data, using the previously outlined CC and ventricle data sets. In our first experiment we use energy functional (3.8), but apply leave-one-out validation. We test two techniques, weight inference for novel images using the proposed method (3.4), and weight inference using the max-margin technique proposed in [150]. We have already demonstrated that on a per-image basis the proposed method outperforms max-margin during training, but here we test the differences between our proposed approach of making $\mathbf{w}$ a function of $I$, and the standard approach of fixing

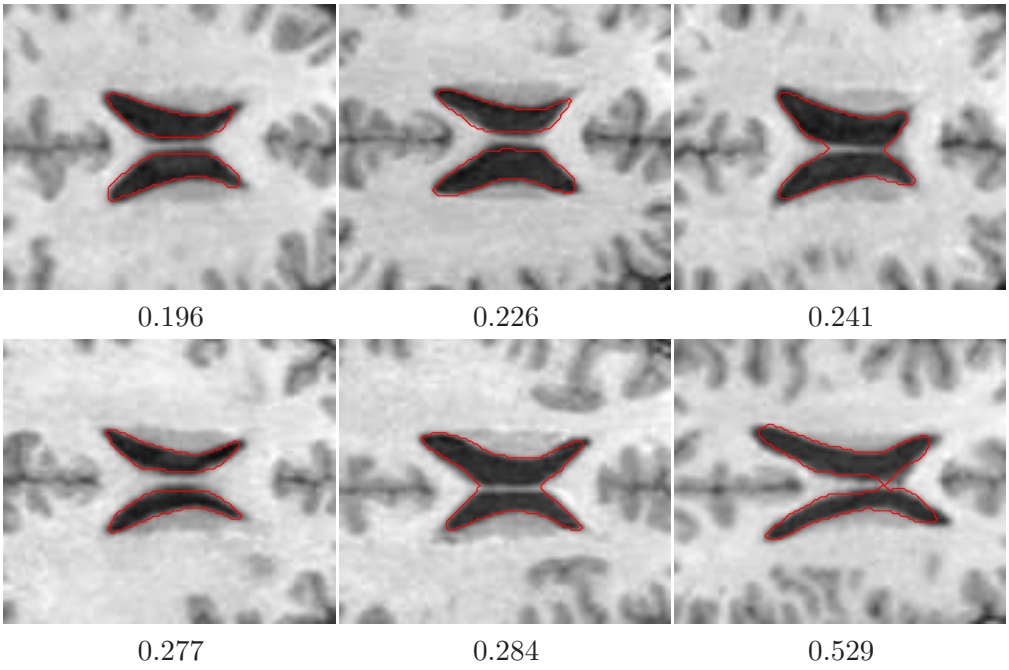



Figure 3.7: (Color Figure) Ventricle segmentation results for 6 of the images. (red) Automatic, optimal parameter segmentations for proposed method, (3.4), demonstrating the full range of error, measured in $\varepsilon$.

$\mathbf{w}$ constant for all $I$ (as is used in [150]). When weights are fixed for all images, i.e. do not vary as a function of the images, we term them *fixed* weights.

We summarize the different results between our proposed method and [150] in tables 3.1 and 3.2, for the CC data and the ventricle data, respectively. The main note of interest is that the accuracy of our proposed method in testing outperforms max-margin in both training and testing for both data sets. As expected, given the weight histograms, our proposed method of adapting the weights on a per-image basis significantly outperforms the fixed weight technique reducing average error by more than 70% for the CC images, and 24% the ventricle images; a similar performance increase to what was obtained with the synthetic ellipse examples. We suspect the smaller performance difference was obtained with the ventricle data due to the smaller nature of that data set, making it easier to fit a single set of weights to the entire data set. A set of ventricle segmentations for our proposed method are shown in Fig. 3.7, notice that even in the worst case the segmentation is not

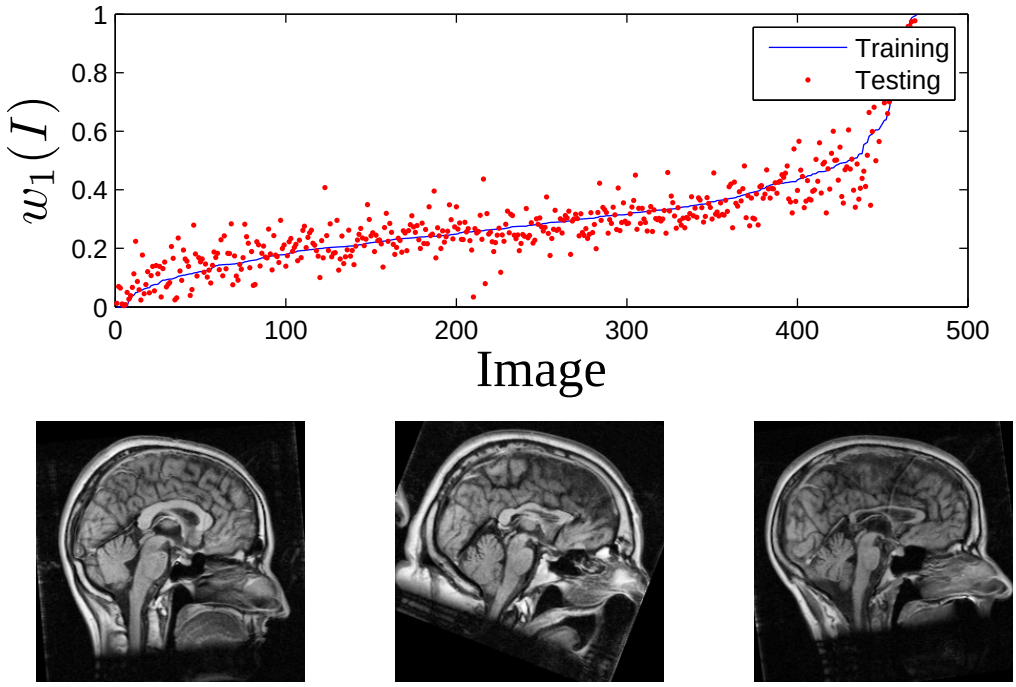



Figure 3.8: Visualization of the learned weighting function $w_1(I)$. (Top) $w_1(I)$ as a function of the image, sorted by training values of $w_1(I)$ in blue. Values of $w_1(I)$ obtained during testing are shown in red. (Bottom) Three equally spaced samples at $w_1(0)$, $w_1(235)$, $w_1(470)$. Notice the increasingly fuzzy boundary as the images progress from left-to-right, where the leftmost image has mostly a boundary-based energy-functional, and the rightmost image has been assigned a predominantly region-based energy-functional.

widely erratic. We also provide a visualization of the learned function $\mathbf{w}(I)$ for the CC images in Fig. 3.8. We have sorted the images by their $w_1$ values obtained during the training experiments in order to make a 1-D image manifold that we can then plot the $w_1(I)$ testing-function over. The weights learned during testing, plotted as red-dots, are very similar to the training function, demonstrating that our manifold learning was indeed able to adequately locate neighborhoods of similar images for each of the novel images. Furthermore, examining the set of images in the bottom panel of the figure, which are evenly sampled from along $w_1(I)$, we notice an intuitive result. As the boundary of the CC becomes increasingly blurred, the boundary weight, $w_1$, decreases.

In order to analyze how well our method performs on generalizing to novel data we also test leave-235-out (i.e. training on the remaining 235), leave-370-out (i.e. training on the

Table 3.3: Error comparison between training sizes for CC data.

| | **Method** | training size | **mean** $\varepsilon$ | **min** $\varepsilon$ | **max** $\varepsilon$ | **std. dev.** $\varepsilon$ |
|---|---|---|---|---|---|---|
| Testing | Proposed (3.4) | 50 | 0.231 | 0.225 | 0.235 | 0.005 |
| | Proposed (3.4) | 100 | 0.186 | 0.179 | 0.189 | 0.004 |
| | Proposed (3.4) | 235 | 0.129 | 0.124 | 0.135 | 0.005 |

remaining 100), and leave-420-out cross-validation(i.e. training on the remaining 50) for the CC data. For each training/testing size, 5 different training/testing permutations of the data set are tested, and the results are averaged across the 5 permutations. Results are included in table 3.3, and statistics are calculated across the mean error in each of the 5 permutations. Interestingly, there is only a minor change in accuracy for leave-235-out testing when compared to the leave-one-out testing experiments in table 3.1 ($\varepsilon = 0.129$ with leave-235-out versus $\varepsilon = 0.121$ with leave-one-out). Naturally, more complicated weight distributions would require more training data to accurately model. Furthermore, though the error with leave-420-out validation is higher, it still outperforms the fixed weight max-margin approach [150], despite the later being trained on 469 images (leave-one-out). In other words, even in the presence of lower training set sizes, inferring the weights for each novel images out performs using a fixed $\mathbf{w}$ for all images.

Before moving onto a new energy-functional, we also make a note about the localized shape spaces resulting from our manifold-learning process. As mentioned earlier, once a novel image is located on the image manifold, the shape prior can be constructed using the segmentations of nearby images, as opposed to the entire training set. Doing so effectively creates a localized shape space, that in the case of energy-functional (3.8), is convex. When sufficient training data is available, this is actually a very powerful added benefit of our technique, reducing shape space error from $\varepsilon = 0.187$ to 0.114, averaged across all 470 CC images (i.e. a reduction of 39%).

Lastly, we perform experiments with non-convex methods. We define the energy functional as a weighted combination of terms from [17, 18, 36] with its functional derivative as in the form of:

$$\begin{aligned} E_\Phi = \quad & w_1(I)g(c+\kappa)\left|\nabla\Phi\right| + w_2(I)\nabla\Phi\cdot\nabla g \\ & + w_3(I)\left(H(\Phi)\left|I-c_1\right|^2 + (1-H(\Phi))\left|I-c_2\right|^2\right) \\ & + w_4(I)\nabla_\Phi E_{shape}^{nonlinear}. \end{aligned} \tag{3.9}$$

where $\Phi$ is the level set representation of $S$, $H(\Phi)$ is the Heaviside function, and $c_1, c_2$ are average intensities as defined in [18]. We set $c = -2.0$, since $w_1$ will scale accordingly regardless of its value, and use $g$ as defined in [17]. To demonstrate our method's performance with different energy functionals, two leave-one-out validation experiments were performed: $(i)$ optimal parameters vs hand-tuned parameters, no outside intensity term ( $H(\Phi)\left|I-c_1\right|^2$), and no auto-cropping (Fig. 3.9 left); $(ii)$ optimal parameters vs averaged optimal parameters, the complete function and auto-cropping enabled (Fig. 3.9 right). As this functional is a level-set formulation [136], it is significantly higher dimension than that of [35], so we also use this example to test the impact of our auto-cropping on both accuracy and speed. With auto-cropping enabled segmentation took approximately 25 seconds per image, as opposed to 3 minutes. As with our previous experiments, significant reductions in average error are obtained via per-image weight learning: 0.20 vs 0.54 in experiment $(i)$, and 0.14 vs 0.17 in $(ii)$; yielding improvements of 63% and 18%, respectively.

## 3.6 Discussion

Choosing the right energy functional is a crucial first step in building a MIS technique. Ideally, one hopes for strictly convex functions with their global minimum lying at the correct segmentations for a given set of images, but this ideal scenario is rarely the case. The main challenges here are determining what energy terms could make a good functional, and how to weight them in such a way as to best segment the images. In some sense the decision comes down boundary versus region versus shape prior, as those are the broadest classification of different terms.

In a typical energy minimization approach the parameter setting, initialization, and minimization phases are well known to be problematic. When a method fails, there are often unanswered questions: what if a different initialization was used, what about a different minimizer, what if different weights were used between competing terms of the energy functional? Hence, the source of erroneous segmentations cannot be easily pinpointed to the energy functional as opposed to one of the many unknowns of the segmentation process.

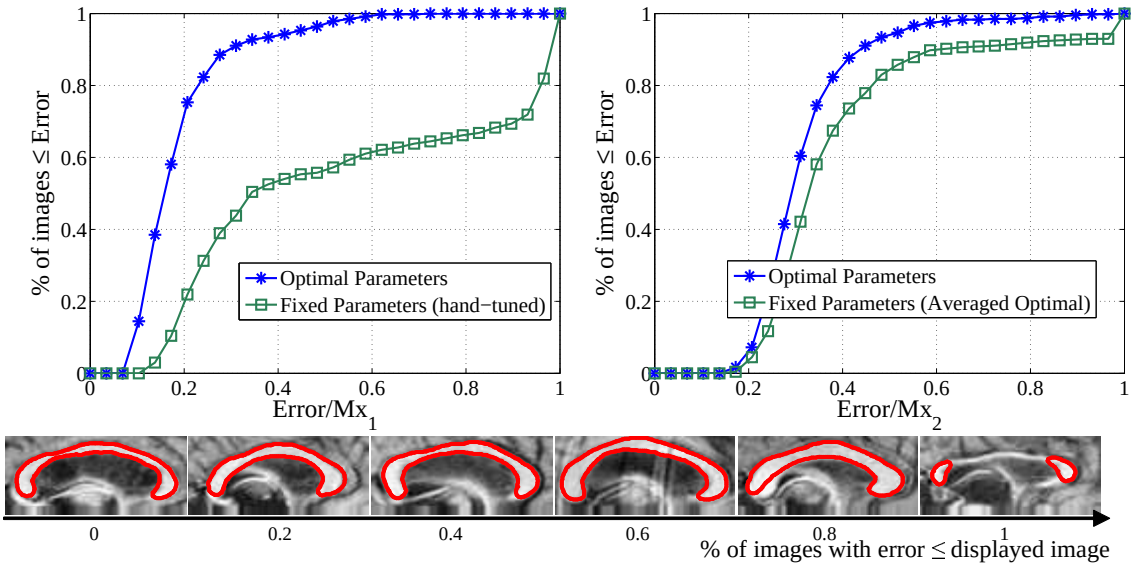



Figure 3.9: (Color Figure) CC segmentation results. (Top) Error plots where $\mathrm{Mx}_1$ and $\mathrm{Mx}_2$ are the maximum values of the optimal parameter error result for experiment $(i)$ and $(ii)$, respectively. (Bottom) Automatic, optimal parameter segmentations demonstrating the full range of error.

The goal is to reduce and ultimately eliminate these points of uncertainty.

On the practical side, we motivated and outlined a new fully-automatic technique for MIA, which addresses the parameter-setting, and initialization problems. Our approach requires training, mimicking the reliance of humans on example images to learn, and incorporating into the learning new images as they are segmented. Our results in the testing experiments clearly demonstrate our method's ability to segment a large number of unseen images, which we foresee carrying over to other datasets and objective functions. On the theoretical side, we explored the consequences of using fixed parameters for energy-minimizing MIA techniques, defined the notion of optimal weights that favor convex energy functionals, and demonstrated a relation between image manifolds and energy-based segmentation.

There are two additional interesting observations to make from our results. Firstly, for the CC images in both training and testing our proposed method's accuracy approaches the lower bound imposed by PCA of 0.1136, whereas for the ventricle data, it does not. An improved shape prior must be obtained for the CC data in order to improve accuracy. This is because the segmentation method being tested was an energy functional defined explicitly over the weights of the modes of variation of the shape prior, i.e. it could not segment objects

outside of the PCA shape space. A shape space with greater respect for the anatomy would have lower error since it could more accurately address novel CC images. As previously noted the PCA shape space error for the ventricle data was smaller, at 0.0329. The low value implies that the shape space was better than that used for the CC data, but this does not consider the amount of erroneous segmentations each space can represent. Still, the results for the ventricle data could also be improved by a better shape prior. However, instead of a shape prior that allows for greater anatomical variability, the ventricle data requires a shape space that greater restricts the erroneous segmentations. Secondly, based on this discussion one would have expected the non-convex, non-linear shape-based energy functional, (3.9), to outperform its convex counterpart as the convex space is likely to include many erroneous segmentations due the non-linear variability of anatomy. While our implementation of [35] is convex on a per-image-basis, the training data, and hence the shape space, is customized to each image, what we referred to as a localized shape space, and is therefore non-convex in context of the set of all images. Had we used a fixed shape-space for all images the error would have been at least 0.187, which is higher than the results obtained with (3.9). In Chapter 4 we define terminology to better discuss these issues, and propose a shape representation that allows for better shape statistics. Our results in Chapter 4 compare favorably to [35], the PCA-based shape representation used for the convex results in this chapter with energy functional (3.8).

In summary, we have demonstrated the merits of using optimal parameters for MIS, but raised new questions as to how exactly to find those parameters. Our proposed method outperformed the ad-hoc hand-tunning that is currently used in most research settings. However, it is worth noting that in some cases different images will not require different weights. In the end it comes down to the stability of the features in the image set. If the strength, and therefore relative importance, of a feature decreases across the set of images, then naturally the weight of the corresponding energy term needs to vary as well.

Our future work focuses on extending and testing these methods in 3D MIS, as well as multi-region segmentation approaches where each region can have different weights [1].

# Chapter 4

# GA-HRPCA

We explore the application of GA to deformable models through the proposition of a novel method for MIS that combines GA with non-convex, localized, medial-based shape statistics. We replace the more typical gradient descent optimizer used in deformable models with GA, and the convex, implicit, global shape statistics with non-convex, explicit, localized ones. Specifically, we propose GA to reduce typical deformable model weaknesses pertaining to model initialization, pose estimation and local minima, through the simultaneous evolution of a large number of models. Furthermore, we constrain the evolution, and thus reduce the size of the search-space, by using statistically-based deformable models whose deformations are intuitive (stretch, bulge, bend) and are driven in terms of localized principal modes of variation, instead of modes of variation across the entire shape that often fail to capture localized shape changes. Although GA are not guaranteed to achieve the global optima, our method compares favorably to the prevalent optimization techniques, convex / non-convex gradient-based optimizers and to globally optimal graph-theoretic combinatorial optimization techniques, when applied to the task of corpus callosum segmentation in 50 mid-sagittal brain magnetic resonance images[1].

## 4.1 Introduction

The primary goal of this work is to address two open and related problems in deformable-model–based MIS:

[1] ©2011 IEEE. Reprinted, with permission, from [99]. This work builds on that presented in [55] and extends [93, 97].

1. **Shape model fidelity**: The ideal shape model must capture the anatomical variability of the target anatomy without representing invalid shapes. However, the better the shape model the harder it becomes to formalize and optimize in a segmentation framework. For example, a convex shape-model is readily formulated and optimized, but inaccurately fits a non-convex shape-space; non-convex shape-spaces are more accurate, but are harder to formalize and optimize. By the same token, localized, medial-axis–based shape statistics better respect anatomical variability than global statistics[124], but their inherent non-convexity makes them challenging to optimize. There is a trade-off: gain optimizablity, lose fidelity.

2. **Optimizability**: As the objective function of a deformable model more faithfully encodes the desired properties of the target anatomy, the energy landscape becomes more complicated and problems with initializations and local minima arise. For example, adding pose to even convex shape-model–based deformable models is well known to cause non-convexity [35]. At the expense of the desired faithfulness, recent trends have focused on ensuring global optima by simplifying the objective function through the use of convex, or submodular, approximations and shape spaces [11, 12, 52, 112, 14, 35]. Thus trading fidelity for optimizability. Where this sacrifice is not practical, a different solution is required. Genetic algorithms (GA) are a powerful alternative optimization-technique that place very little requirement on the objective function, but to the best of our knowledge they remain entirely unexplored in their application to localized shape-based image segmentation. There are, however, some existing applications of GA to deformable models (See Section 4.1.3 for details).

In what follows, we elaborate on these open problems, and propose our solution that combines GA with localized shape training. There is a trade-off in 1) and 2) above. Shape model fidelity is desirable, but lowers optimizability. Optimizability is desirable, but lowers fidelity. We will use GA to mitigate the trade-off that is usually required between between 1) and 2) above, allowing complex shape models while reducing problems with initializations and still being able to attain solutions close to the global optimum.

### 4.1.1 Shape Model Fidelity

As briefly discussed in Section 2.1.4, linear models of shape deformation may not always adequately represent the variance observed in real data. Linear shape models assume the

data lies on a linear manifold, but picture fitting an ellipse to an 'S'-like shape space. Nonlinear statistics are often used [145, 29, 32, 33, 45, 36, 42, 31] in order to improve shape model fidelity (i.e. reduce the number of erroneously represented shapes), but we propose a different solution.

We argue that the shape model fidelity problem is not necessarily due to the application of a linear model to nonlinear data but rather because of the implicit nature in which the statistics are applied. By *implicit* we mean the statistics attempt to model variation in the shape, rather than variation in the parameters governing the deformations themselves. Note that we are not referring to the shape representation being implicit or explicit, but instead whether the deformations are implicitly or explicitly studied. Implicit shape statistics result from the majority of previous deformable model approaches adopting a boundary-based shape representation, aside from a few exceptions [124]. As a consequence, the statistics are calculated using boundary models of the shape instead of models representing the interior and skeletal topology of the structures. Studying the underlying structural changes of a shape allows deformations that were previously non-convex to be decomposed into linear models. We refer to these as *explicit* shape statistics since they are calculated over the very parameters responsible for varying the object's shape. Consider an object represented by a single pixel. Different images of the object show the pixel moving around in a circle. A non-linear function is required to describe the pixels motion and hence no linear statistics can capture the motion adequately as long as it is the object's $x, y$ position being studied. However, once decomposed into a function of sin and cos, the underlying parameter that controls the objects variability is linear in its variation, and hence linear statistics will have greatly improved shape model fidelity. The same argument carries forward, albeit more complexly, to a more complex object. A simple bending of a shape's medial-axis is a linear deformation under the appropriate representation, as it's simply a rotation of some of the medial-nodes (Fig. 4.1). However, the bending is a highly non-convex deformation once embedded in the image domain, as either an implicit shape [127] or an explicit-boundary based model [28].

In addition to being implicit in nature, the statistics are often global in nature, whereby *global* we mean the statistics attempt to model variation in the entire shape. In other words, each shape is a single point in some high-dimensional space, and the statistics, linear or not, describe some restricted set of that space. Global shape statistics are unable to restrict deformations to particular locations of the shape model or anatomy. Furthermore,

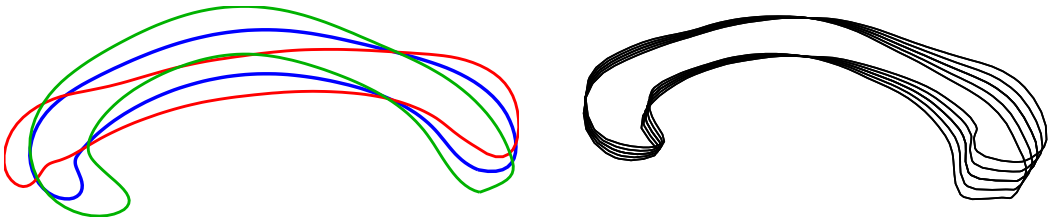
as most dimensionality reduction techniques are designed to preserve variance, small changes in shape are discarded to preserve larger ones. Though small in scale, those changes can represent true signal changes (i.e. not due to noise) in a localized region of a shape [162, 111]. Global deformations can not adapt to localized variations in shape, which are often of high interest, e.g. regions of pathology.

Decomposing implicitly complex deformations, like bending an object, into simpler linear parameters is made possible through localized shape statistics defined over a medial-based shape representation, which allows deformations to be further quantified into specific types (bending, bulging, stretching) that can be linearly measured. Specifically, medial-axis–based 2D shape representations enable such deformations by describing the object's shapes in terms of an axis positioned in the middle of the object, along with thickness values assigned to each point on the axis that imply the object's boundary. Rotating part of the medial axis produces bending deformations, increasing the thickness along part of the axis creates bulging, and lengthening the medial-axis produces a stretching of the object.

Medial-based shape representations have been emerging as a powerful alternative to boundary-based techniques [10, 40, 55, 58, 123, 141, 135]. In addition to decomposing seemingly complex deformations into their linear variables, medial-axis representations also avoid having to define absolute positions of boundary points. With medial-axis representations, deformations like bending the object are performed directly on the medial-axis, which implicitly moves the boundary and ensures the affected boundary points move together. By comparison, with a boundary representation, one has to carefully move all the connected points to ensure boundary continuity. Statistics on the nonlinear manifolds of medial representations have also been proposed [45]. However, there has been little exploration of building deformable models using these shape representations and their corresponding shape statistics [124, 93]. In this work, we explore this very notion by building a medial-based deformable model with deformations governed by explicit, localized shape statistics.

### 4.1.2 Optimizability

As reviewed in Chapter 2, energy-functional minimization can be carried out in a variety of ways through various forms of optimization. One solution is to perform explicit differentiation under the Euler-Lagrange equations, where each new image segmentation application that requires a modified energy functional must be accompanied by one such derivation.



(a) (b)

Figure 4.1: (Color Figure) Shape deformations of the boundary of the corpus callosum (CC) under its global modes of variation using a boundary-based model (a) and a medial-axis–based model (b). In (a) the mean shape is shown in blue. The red and green contours indicate a positive and negative deformation of the CC along the first principal component. Notice the global and linear statistics in (a) enforce a very un-natural deformation of the shape compared to the deformations in (b).

The result is a set of conditions, which guarantee a stationary point of the energy functional. The solution is then obtained through a gradient descent process where the change of the shape model (with respect to an artificial time variable) is equated to the Euler-Lagrange equations. When applied to typical deformable models, which are represented by non-convex energy functionals, the optimization process still requires initialization at some optimal target location of an image, with some shape, orientation, and scale. Together, the location, orientation, and perhaps the scale of the shape model are referred to as the pose. Depending on the initialization, a different segmentation is typically obtained, and different initializations can therefore produce different amounts of segmentation error (Fig. 4.2). It is, as such, common for deformable models to be described as having problems with local minima. One proposed solution is to use convex energy functionals, which by definition have no local minima.

Simply put, though they can be globally optimized, convex energies and shape spaces may not be expressive enough to capture the characteristics of the segmentation problem at hand as accurately as their non-convex counterparts. In other words, they may not faithfully encode the desired properties of the target anatomy. This is especially true for the shape spaces, as our results demonstrate later in the chapter (Section 4.3). Similarly, not all energy functionals and shape spaces can be represented convexly. There is thus a trade-off: gain global optimality but likely lose accuracy. As a consequence, these works do not provide a definitive solution to the problem. Another solution present in the optimization

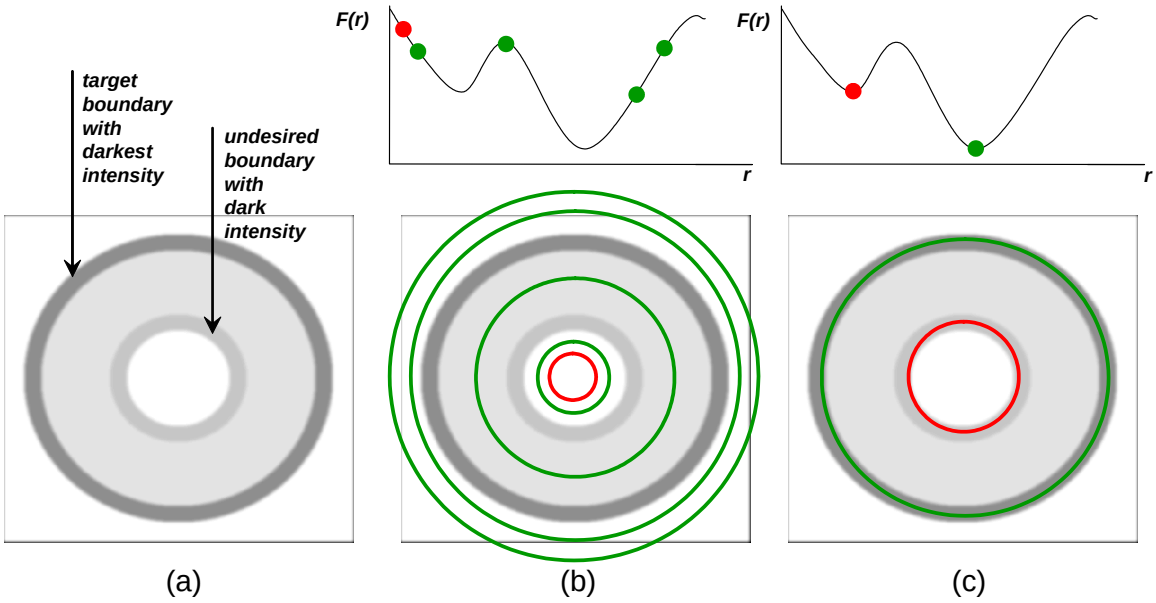



Figure 4.2: (Color Figure) Building on the synthetic example of a single parameter deformable model with local minima in Fig. 2.2. The circular deformable model's only parameter is its radius $r$. The energy function $F(r)$ is minimal for the circle with darkest average intensity. The input image is shown in (a) with the darkest outmost boundary representing the global minima. (b) A traditional deformable model is initialized at a single point (red) while GA-based statistical deformable models are initialized at multiple locations (green) and perform mutations on $r$. (c) The GA converges to the global minimum (darkest), while the gradient descent based deformable model gets stuck in a lighter local minimum.

literature when initialization and local minima are problematic is GA, a method which is largely unexplored in relation to shape-model–based deformable models.

To re-emphasize the argument of the previous paragraph we pose the following question: Is it worth sacrificing the complexity of the energy functional, and thereby reducing its ability to address image variability, in exchange for global optimality, or can more complex functions and shape spaces be utilized with an approximate optimization method and achieve superior accuracy? Our hypothesis is that while convex approximations to non-convex functions and shape spaces can be globally optimized they sacrifice fidelity to the data and, in doing so, are less accurate than non-convex spaces and functions optimized using GA. We propose a novel method that uses GA to overcome the initialization and local-minima entrapment typically encountered in non-convex problems. While we are not guaranteed global optima,

our results show that using non-convex shape spaces our method can outperform convex shape spaces under the same energy function (Section 4.3), i.e. the energy function remains fixed, while only the shape constraints and solvers are varied.

### 4.1.3 Contributions and Related Work

To summarize the previous sections, in order to be applicable to MIS, deformable models require learned, intuitive shape models, and a level of robustness to initialization. It is well known that implicit shape statistics, though easily represented and applied to deformable models, provide unintuitive deformations of shape and often fail to adequately represent non-convex, localized deformations [124]. Localized shape statistics are more intuitive to clinicians, and better represent local variations in shape, but the proposed methods remain largely unapplied to deformable models. Robustness to initialization can be handled through convex optimization methods, but there are drawbacks in that not all energy functions are convex, convex functionals may not faithfully represent the segmentation problem, and convex shape-spaces may not accurately summarize the variability of the target anatomical structure. Therefore, what is desirable is a method that generalizes to a larger class of functionals (convex or not) while allowing the exploration of the search space in a manner that still converges towards an optimal solution. The method must allow the exploration to be carried out from a variety of initial locations, and enable it to be done in a way that reflects the non-convex, localized variations of shape in terms of bends, bulges and stretches. We propose one such method.

In this chapter, we explore the application of GA to deformable models as well as propose a novel deformable-model based on localized shape statistics using HRPCA (as defined in Chapter 1). Our main contribution in this chapter is a novel method for fitting complex, medial-based shape models to medical images under arbitrary energy functionals. In a narrower sense, our new segmentation method replaces gradient descent with GA, and convex, implicit, global shape statistics with non-convex, explicit, localized ones. GA address the typical initialization, local minima problems associated with traditional energy-minimization techniques. Our unique application of a medial-based deformable model with HRPCA maintains a statistically feasible shape model. It also reduces the size of the search space by constraining the evolution to deformations that are intuitive (stretch, bulge, bend) and are driven in terms of localized principal modes of variation, instead of modes of variation across the entire shape. Though not expressly guaranteed to find the global optimum,

our method compares favorably to leading segmentation techniques, including convex / non-convex gradient based optimizers and globally optimal graph-theoretic combinatorial optimization techniques, on a data set of 50 mid-sagittal brain magnetic resonance images.

Though prior works have used GA for MIS, here we limit our discussion to those pertaining to deformable models. For a comprehensive survey of the use of GA for MIS, the reader is referred to [85]. A few methods have used GA to minimize traditional deformable models [6, 7, 82, 155, 43, 110]. Ballerini extends the classical active contour models [154] by using GA to directly minimize the standard energy functional [6]. Members of the GA population are hypothetical shape configurations, represented by their explicit contour locations. The method was later extended to color images by using one image term per color channel [7]. MacEachern and Manku presented a similar method using a binary representation of the contour [82]. Similarly, Tohka presented simplex meshes paired with image-based energies, minimized via a hybrid GA-greedy approach, and applied the technique to the segmentation of 3D medical images [155]. Fan *et al.* also develop a GA method for an explicit active contour, but describe their method using Fourier descriptors and employ parallel GAs to speed up minimization [43]. A different shape representation, known as topological active nets, is used by Ibáñez *et al.* to enable the segmentation of objects with unknown topologies, or even multiple objects in the same scene[110]. However, aside from simple boundary smoothness constraints all of these methods are based on classical active contour models or their variants without incorporating prior shape knowledge, making them prone to latching to erroneous edges and ill-equipped to handle gaps in object boundaries. In [67], GA were used with statistically-based ASMs, where the parameter space consists of possible ranges of values for the pose and shape parameters of the model. The objective function to be maximized reflects the similarity between the gray levels related to the object in the search stage and those found from training. Additional works use convex, implicit, global shape statistics assuming a Gaussian distribution around a mean shape [50, 132, 103]. Mignotte and Meunier [103] incorporate prior shape information by defining the mean as a circular deformable template, while [132] uses a PDM for occluded shape reconstruction, and [50] uses a level set shape representation and a learned mean from training data. Although these techniques apply GA to produce generations of plausible populations of shapes, the statistically-based deformations are global (over the whole shape) and their convexity may not offer the required flexibility to accommodate for shape variations that are restricted to particular locations, nor are they intuitively defined (bulge, bend, stretch) deformations.

In summary, though GA have been used to optimize both classical, and shape-prior–based deformable models, they have not been applied to deformable model methods incorporating localized, medial-based shape statistics. In Section 4.3, we show superior accuracy to a globally optimized method that makes use of convex, implicit, global PCA-based shape statistics [35].

Other works have set out to automatically fit medial-based shape models to image data [81, 79]. These approaches build on the m-reps shape representation of Pizer *et al.* [122], in which shapes are represented using a global-to-local scale approach as objects, "figures" (anatomical subregions or parts), and "atoms" (individual medial nodes). Once initialized by registration to manually specified points, models are fit to the data using a hierarchical local-search method, i.e. starting first with each object, then each figure, and finally each atom. At each scale, models are fit to the image data using an iterative conditional modes (ICM) strategy across the scale's entities (objects, figures, or atoms), with the conjugate gradient method being used to optimize each entry (object, figure, or atom) individually [70]. Our HRPCA-based approach complements the aforementioned works [81, 79] by computing statistics across a spectrum of shape scales and locations; ranging from an individual medial node (or atom), and adding increasingly more atoms until covering a whole anatomical part (or figure). This is achieved through our HRPCA in which multi-location, multi-scale HRPCA statistics are gathered for different deformation profiles. This offers greater deformation flexibility (i.e. at multiple locations, sizes, and deformation types) during the model-to-image fitting and, at the same time, remains faithful to the underlying variability in the training data. Inspired by the approach in [81, 79], we also seek to fit our shape model to the image in a global to local manner. However, in our case we employ genetic algorithms in order to alleviate the initialization requirement of the gradient descent-like ICM method, which is known to be heavily sensitive to its initial configuration [78] (Section 9.3.1, page 247). Furthermore, we present comparison to numerous convex and non-convex methods for MIS, exploring the aforementioned questions relating to energy function and shape space fidelity versus optimizability (Section 4.3).

To the best of our knowledge, none of the existing image segmentation techniques evolve a population of shapes using intuitive, spatially constrained, and plausible deformations, nor have they enabled such deformations under general energy functionals that are customizable for the problem domain.

## 4.2 Methods

In this work, we build a deformable model that uses a medial shape representation to provide a way to synthesize intuitive deformations and localized shape statistics (HRPCA). Our use of HRPCA captures regional medial-based shape information, thereby providing controlled localized fitting of specific anatomical regions of a shape (Fig. 4.1-b). We solve our resulting optimization problem using GA, demonstrating a generic way to fit complex shape models to medical images.

In order for the descriptions that follow to lend more readily to standard descriptions of GA, from here on in the chapter we consider the problem of fitting an HRPCA-based deformable model to an image as a fitness maximization problem, rather than an energy functional minimization problem. This is done without loss of generality.

We begin with an overview of how we use HRPCA to build localized shape statistics, where some of the key differences between what we have referred to as global and local statistics should become clearer. We also provide a short introduction to GA for the unfamiliar reader. We then use a motivating example to describe how GA can be used for deformable models, before describing our use of GA in detail.

### 4.2.1 Statistically-Constrained Localized and Intuitive Deformations using HRPCA

We use the multi-scale (hierarchical) and multi-location (regional) PCA method introduced in [55] on a training set of medial shape profiles. We will first give an overview of medial shape profiles and then proceed to describe how HRPCA is applied in this work.

Medial-axis–based 2D shape representations describe the object's shapes in terms of an axis positioned in the middle of the object along with thickness values assigned to each point on the axis that imply the shape of the boundary. We therefore describe the shape as a mapping from $\mathbb{R}$ to $\mathbb{R}^4$, the domain of which is a parameter $m$ that traverses the medial axis. We use a single primary medial axis, though secondary medial axes or subfigures would be needed to represent more complex structures. The range of the mapping consists of 4 scalar values for each $m$, forming medial profiles. These are a length profile $L(m)$, an orientation profile $R(m)$, a left (with respect to the medial axis) thickness profile $T^l(m)$, and a right thickness profile $T^r(m)$, where $m$ is now discretized, $m = 1, 2, ..., N$, $N$ is the number of medial nodes, and nodes 1 and $N$ are the terminal nodes. For all examples

here and all results presented later, $N = 100$. The length profile represents the distances between consecutive pairs of medial nodes. The orientation profile represents the angles between segments connecting consecutive pairs of medial nodes. Finally, the thickness profiles represent the distances between medial nodes and their corresponding boundary points on both sides of the medial axis (Fig. 4.3), and will be different only when the shape is asymmetric. In other words, corresponding boundary points $x_m^l$ and $x_m^r$ lie at distances $T^l(m)$ and $T^r(m)$, respectively, along a direction normal to the medial axis at $x_m$ (Fig. 4.4). As the medial-axis is discretized into $N$ nodes, implying $N-1$ segments, the normal to node $x_m$ is measured from the medial-segment connecting $x_m$ and $x_{m-1}$ (see right-angles in Fig. 4.4-right). Clearly, $x_1$ is an exception and in that case the medial-segment connecting $x_1$ and $x_2$ is used. Other more elaborate medial representation may be adopted as well (e.g. m-reps [122]).

In order to build energy functions using HRPCA-based shape models, the boundary of the CC shape must be reconstructed from the four medial profiles and a set of affine parameters that describe the objects pose and scale. The affine parameters are: the position, $(t_x, t_y)$, describing the $(x, y)$-coordinate of a known reference node; the scale, $(s_x, s_y)$, describing the scale of the object; the base angle, $\theta$, describing the angular offset of the model from the x-axis. To reconstruct the object's shape given its set of medial profiles, we calculate the positions of the medial and boundary nodes from a known reference node at location $x_1 = (t_x, t_y)$. The next node at position $x_2 = x_1 + v_1$ is determined using an offset vector $v$ whose angle is specified by the orientation profile plus the base angle $\theta$, and whose length is specified by the stretch profile scaled by $(s_x, s_y)$. The corresponding boundary nodes $x_2^l$ and $x_2^r$ (Fig. 4.4-right) are then orthogonal to the medial axis, at a distance specified by the thickness profile scaled by $(s_x, s_y)$. This process is repeated recursively, generating $x_3 = x_2 + v_2$, and so on. For complete details see [55]. Finally, with the medial profiles like those shown in Fig. 4.3 as an input, we can reconstruct the CC structure in Fig. 4.4. Using polygon rasterization the reconstructed set of nodes can then be used to create a binary image, of the same size as an input image, $I$, with ones representing pixels on or within the shapes border, and zeros representing pixels not belonging to the shape [46].

The profiles are thus rotation- and translation-invariant, and capture intuitive measures of shape: length, orientation, and thickness. Altering the different profiles produces intuitive, controlled deformations: stretching, bending, and bulging, respectively. For example, setting $\theta = 0$ and $\{R(m) = 0\}_{m=1}^{N}$ would produce a straight-line medial axis, parallel to the

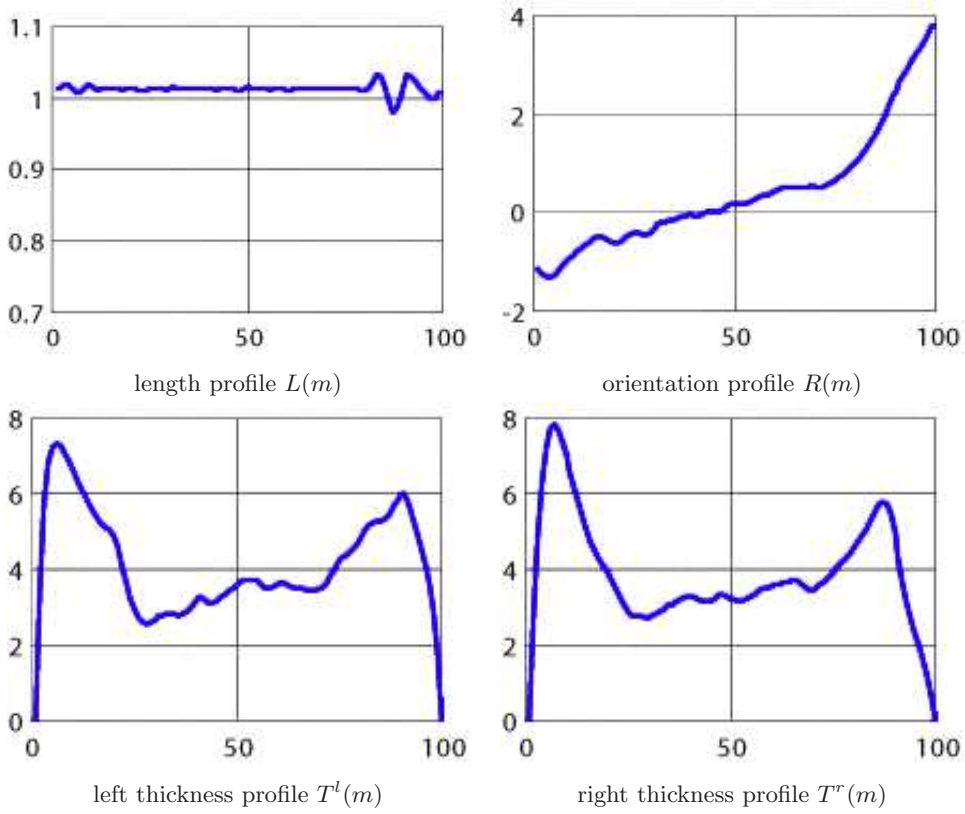



Figure 4.3: Example medial shape profiles used to reconstruct a CC shape model. The values of the medial profiles are plotted along the y-axis with pixels for length and thickness profiles, and radians for the orientation profile. All plots are a function of the medial node number $m = 1, 2, ..., 100$, indicated along the x-axis. Adapted from [55].

x-axis of the image. Setting $\{R(m) = \pi/2\}_{m=N/2}^{N}$ would create a right-angle bend at the mid-point of the medial-axis. Increasing $\theta$ would simply rotate the object around the reference point. Of course, setting $R(m)$ randomly may be undesirable and produce unwanted shapes. Next we describe how to use training data to calculate statistics for each profile in order to restrict the deformations to plausible shapes.

In HRPCA the principal component analysis is a function of the location, $l$, scale, $s$, and type of shape profile, $d$ (i.e. length $L(m)$; orientation $R(m)$; left thickness $T^l(m)$; or right thickness $T^r(m)$ profiles) (Fig. 4.5). For clarity of presentation, we define $p_d(m)$ as the $d^{th}$ shape profile (e.g. $p_1(m) = L(m)$). Hence for each location, scale, and shape profile type, we obtain an average medial sub-profile, $\bar{p}_{dls}$, the main modes of variation, $\mathbf{M}_{dls}$, and the amount of variation each mode explains. The subscript "$dls$" indicates that the

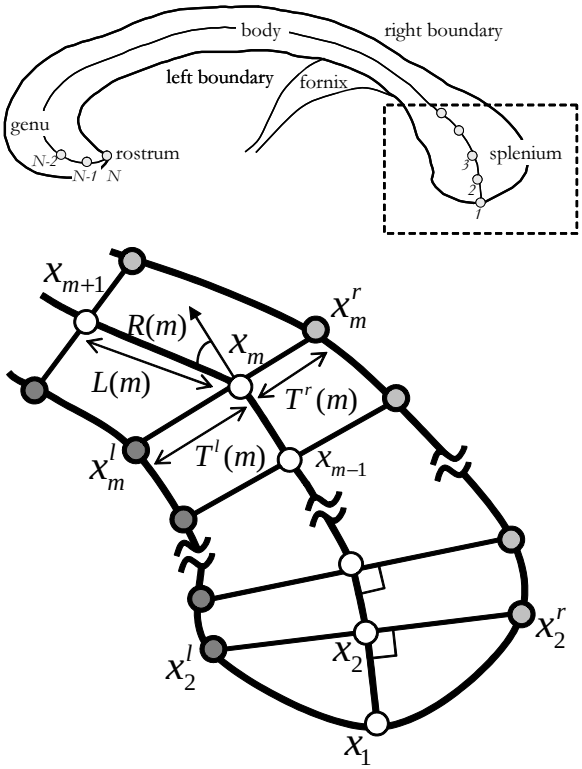



Figure 4.4: Representing the CC using medial profiles. (Top) Anatomically labeled CC shape reconstruction resulting from the medial profiles in figure 4.3. (Bottom) The details of the medial profile representation for the splenium region (dashed). Medial nodes are shown in white, whereas the left and right boundary nodes are shown in dark and light gray, respectively. $x_m$, $x_m^l$ and $x_m^r$ are the $m^{th}$ medial, left boundary and right boundary nodes, respectively. $L(m)$, $R(m)$, $T^l(m)$ and $T^r(m)$ are the length, orientation, left and right thickness profile values, respectively. Adapted from [55].

statistics are made specific to a location $l$ (e.g. to the left of the shape), scale or extent $s$ (e.g. affecting one tenth of the shape, half of it, or all of it), and the deformation profile type $d$ (e.g. bending or bulging). Note, of course, that at each location, $l$, a maximum scale, $s$, of $N - l + 1$ can be selected. For example, $p_{L,1,5} = \bar{p}_{1,1,5} + \mathbf{M}_{1,1,5}\alpha_{1,1,5}$ would calculate a new $L(m) = p_{1,1,5}$, for $m = 1...5$, since only those 5 positions are affected. Example modes of variation for a global bending deformations are shown in Fig. 4.1-b, i.e. $d = 2$, $l = 1$ and $s = 100$. For each location, $l$, with scale $s$, there are $s$ values, say thickness values, for the $T^r$ profile and, as such, for each profile, location, and scale, $\mathbf{M}_{dls}$ is an $s \times s$ covariance matrix for the $s$ modes of variation of length $s$. Consequently, we can now generate a statistically

feasible stretch, bend, or bulge deformation at a specific location and scale in terms of the corresponding main modes of variation [55].

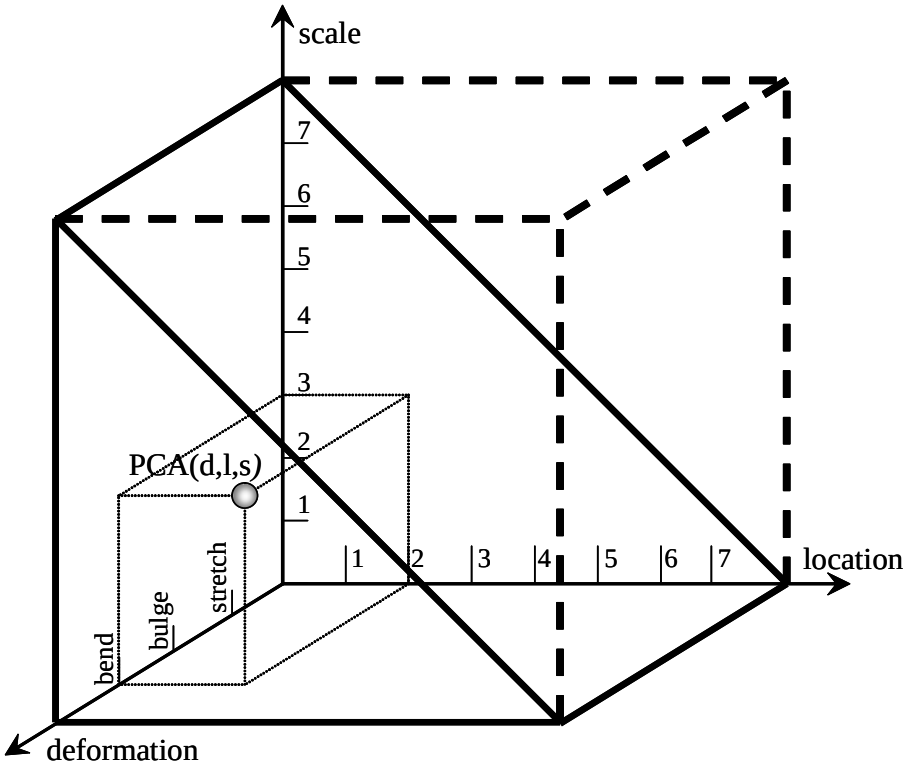
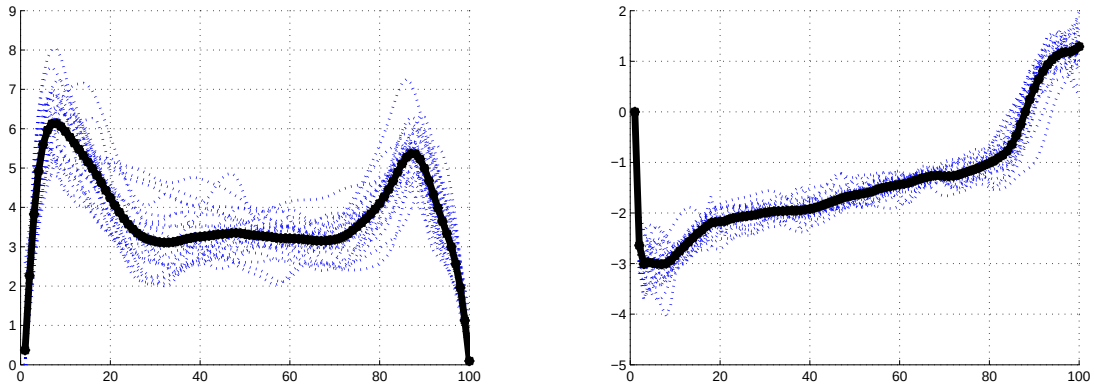



Figure 4.5: HRPCA is a function of the deformation (d), location (l), and scale (s). Since all scales are not valid for all locations, only a subset of the full space (the right angle prism drawn with solid black) is valid (see Section 4.2.4). Adapted from [55].

More specifically, a new statistically feasible profile, made specific to a particular deformation type, location and scale, is calculated by perturbing the mean training profiles by a weighted sum of the covariance of the corresponding training profiles using the following model of medial profile variations,

$$p_{dls} = \bar{p}_{dls} + \mathbf{M}_{dls}\alpha_{dls} \tag{4.1}$$

where $p_{dls}$ is the resulting shape profile, $d$ is the deformation profile type ($L(m)$, $R(m)$, $T^l(m)$, or $T^r(m)$), $l$ and $s$ are the specified location and scale values, $\bar{p}_{dls}$ is the average medial profile for the specified type ($L(m)$, $R(m)$, $T^l(m)$, or $T^r(m)$), the columns of $\mathbf{M}_{dls}$ encode the main variation modes for the specified profile type, and $\alpha_{dls}$ are weights of the



Thickness $T^l(m)$
(a)

Orientation $R(m)$
(b)

Figure 4.6: Overlay plot of (a) thickness profiles, measured in pixels, and (b) orientation profiles, measured in radians. Results are from a training set of 50 CC shapes, with both profiles plotted as a function of the medial-node number, $m = 1...100$. Mean profiles appear in bold. Adapted from [55].

variation modes and are typically limited to $\pm 3$ standard deviations.

Note that for any shape profile type multiple variation modes can be activated by setting the corresponding weighting factors to non-zero values. Since each variation mode acts at a certain location and scale, we obtain for each profile

$$p_d = \bar{p}_d + \sum_{l=1}^{N} \sum_{s=1}^{N-l+1} \mathbf{M}_{dls} \alpha_{dls}. \tag{4.2}$$

The upper limit of the inner summation of (4.2) indicates that all locations have PCA performed at them for all scales as long as the combination of locations and scales is valid.

In summary, varying the weights of one or more of the variation modes alters the length, orientation, or thickness profiles and generates, upon shape reconstruction, statistically feasible stretch, bend, or bulge deformations at specific locations and scales.

Each scale and location provides a unique measurement for the analysis of a shape. Examples of the global (maximal scale) thickness and orientation samples with their corresponding averages are show in Fig. 4.6. These examples also serve to illustrate how visualization of these profiles can provide better understanding of the variations present in the data by identifying contributions of different types of deformations along the anatomical

shape. Notice, for example, that towards the center of the CC (value range 30-50 on the horizontal axis of the plots) the bulk of the variation lies in the thickness profile, something not readily interpreted from results with classical PCA approaches (i.e. global PCA on boundary, or implicit, shape models, as opposed to localized PCA on a medial-axis based shape model). With classical PCA approaches one would capture the variation in the CC's shape, but could not readily interpret it as being mostly a change in thickness instead of a bending or stretching of the object, nor could one readily identify exactly where the variance occurs on the model.

We note, however, that although each scale and location provides a unique measurement for analysis, there is a degree of redundancy in the representation when building a statistical model of deformations. Each profile is a real-valued vector of length $N$, and thus lies in $\mathbb{R}^N$. Any orthogonal set of $N$ vectors is a basis for $\mathbb{R}^N$, and by defintion any set larger than $N$ must have vectors that are linear combinations of other members of the set. As such: (1) Any large-scale deformation can be represented as the union of multiple smaller-scale deformations and (2) the entire $N \times 4 = 400$ dimensional space of deformations can be captured via a global, explicit PCA (over the whole shape, i.e. by setting $l = 1$ and $s = N$), assuming all the PCA modes are retained in the model and no limits are placed on the amount of variation (typically measured as multiples of standard deviations) along each mode. In other words, with enough modes of variation, any single scale of deformations can capture the entire shape space.

By design, global PCA seeks a subset of vectors that maximally represents the variability in the data, assuming the data follows a Gaussian distribution. A significant drawback of this global approach is that varying the weight of a single variation mode generally causes all the landmark positions to change (Fig. 4.1). In other words, although the global model produces feasible shape deformations only, a desirable trait, it generally produces global deformations over the entire object, making it difficult to guarantee fitting one part of the shape without unfitting another part. In contrast, setting $s = 37$, for example, produces 64 possible locations (from location $l = 1$ to $l = N - s + 1 = 64$). Each location has 37 modes of variation for a total of $64 \times 37 = 2,368$ variables. Clearly only $N = 100$ of these vectors can be orthogonal, but no two vectors need point in the same direction. Hence each vector can represent a unique and statistically valid localized deformation of the shape model for a particular anatomical region.

These statistically valid and localized deformations can be used effectively during the

segmentation process, as will be better detailed in the next section. For example, suppose our shape model fits perfectly over all the corpus callosum (CC) except for a misalignment around the splenium region of the CC (Fig. 4.4). In this scenario, the shape of the model must be adjusted only around the splenium without affecting any of the correctly aligned parts. This is simply accomplished by setting $l = 1$ and $s = 37$ to localize the deformations around the splenium part only and ensure plausible deformations respecting the variations observed in the training set in that region only. A compact yet global PCA ($s = 100$) is unlikely to include a mode of variation that directly accomplishes this task. Instead some complex combinations of several global modes must be used to ensure the desired localized effect.

The trade off is clear, finer scale deformations mean more variables that will need to be optimized, but the statistics will better represent localized changes in shape. Larger scale deformations have fewer variables but sacrifice localized changes to preserve global ones, and thus it is difficult to fit one part of the model without unfitting another part. We can exploit the advantages of both options during our segmentation procedure by starting with larger scale deformations, and then moving to increasingly finer scale deformations (Section 4.2.4 for details). By starting with large scale deformations we can find the best global fit of the model to the image data without wasting time on localized changes that represent a small portion of the overall error. Once the global model is fit, localized deformations can refine anatomical subregions and thus give a better fit to localized variability in the data (results of this can be seen in Fig. 4.10).

The preceding discussion outlined how to learn our shape model through examples. Additional details on HRPCA are available in [55] for the interested reader. Next we describe how to build a fitness function that can be used to characterize how well the shape model has been fit to the image. The fitting procedure itself, performed using GA, is described in Section 4.2.4. Essentially we need to discover how to deform the shape (what pose parameters and modes of variation, $\alpha_{dls}$ to activate) in order to fit the shape to the image. The GA optimization explores the space of these deformations and finds the ones that best fit the model to the data (those that score highly according to the objective function).

### 4.2.2 Building a Fitness Function

The fitness function is used for ranking the population in order to determine which members should reproduce and which should survive. It is akin to the energy functional in typical

deformable models, in that it is the objective function being optimized. As noted in Section 4.1, the use of GA allows us to easily adapt the function to any given task including both convex and non-convex prior shape and image-based knowledge; something globally-optimal energy-minimizing models do not generally allow for [74]. Rather than focusing on easily described and optimized terms, fitness functions can be constructed that directly reflect the characteristics thought to be exhibited by the target structure. The general form of our energy function is then a weighted combination of the desired terms. For example, we can adopt the fitness function $Fit_1(i)$, where $i$ is a potential segmentation, to consider the mean and standard deviation of the image intensity enclosed by the shape's boundary, the average edge strength along said boundary, and the learned size of the CC:

$$\begin{aligned} Fit_1(i) &= w_1 \exp\left(\frac{-F_1(i)}{2\sigma_1^2}\right) + w_2\left(1 - \exp\left(\frac{-F_2(i)}{2\sigma_2}\right)\right) + \\ &\quad w_3\left(1 - \exp\left(\frac{-F_3(i)}{2\sigma_3}\right)\right) + w_4 \exp\left(\frac{-F_4(i)}{2\sigma_4}\right) \end{aligned} \tag{4.3}$$

where

$$\begin{aligned} F_1(i) &= \left(\mu - \sum_{x=1}^{\mathcal{N}} q(x)\right)^2 & F_2(i) &= \frac{\sum_{x=1}^{\mathcal{N}} \|\nabla I(x)\|\|\nabla q(x)\|}{\sum_{x=1}^{\mathcal{N}} \|\nabla q(x)\|} \\ F_3(i) &= \frac{\sum_{x=1}^{\mathcal{N}} I(x)q(x)}{\sum_{x=1}^{\mathcal{N}} q(x)} & F_4(i) &= \sqrt{\frac{\sum_{x=1}^{\mathcal{N}} (I(x) - F_3(i))}{\sum_{x=1}^{\mathcal{N}} q(x)}} \end{aligned} \tag{4.4}$$

$x_i ... x_{\mathcal{N}}$ are the indices of the $\mathcal{N}$ pixels, $q(x)$ is a binary image of the reconstructed shape (Section 4.2.1), and $\mu$ and $\sigma_1$ are the average and standard deviation, respectively, of the sizes of the CCs learned from the training set. $I$ is the image, and $\sigma_2, \sigma_3, \sigma_4$ are the learned averages in edge strength, mean image intensity, and image intensity standard deviation. Hence, $F_1(i)$ represents the conformance to the learned area of shape $i$, $F_2(i)$ the average gradient magnitude along the shape's boundary, $F_3(i)$ the mean image intensity enclosed by the shape's boundary, and $F_4(i)$ the standard deviation of the intensity enclosed by the shape's boundary. Finally, $w_1, w_2, w_3$ and $w_4$ are scalar weights controlling the importance of each term in the segmentation process.

Alternatively, we can adopt a more traditional energy functional commonly used in the

deformable model community:

$$\begin{aligned} Fit_2(i) &= -\sum_{x=1}^{\mathcal{N}} \left(w_1 \left(f(x)q(x) + g(x)(1-q(x))\right) + \right. \\ &\quad \left. w_2 h(x) \left|\nabla q(x)\right|\right) + w_3 F_1(i) \end{aligned} \tag{4.5}$$

where $f = -log(P_{obj}(I))$, $g = -log(P_{bk}(I))$, for object and background likelihoods $P_{obj}$ and $P_{bk}$, $h = \frac{1}{1+|\nabla \mathrm{I}|^p}$, and $F_1$ is the shape term as defined in (4.4) (see [35] for details). This functional has recently been shown to be convex when defined over convex shape-spaces [35], and can therefore be globally optimized in that scenario. Using this functional is therefore a good way to investigate the fidelity trade-off between convex and non-convex shape-spaces, since we can compare the accuracy of the guaranteed global optima under convex shape-spaces to the potentially global optima found using the proposed non-convex shape-space. Since with GA we seek to define a maximization problem, the function of [35] has been negated, thus making it a concave optimization problem over a non-convex shape-space.

Depending on the image segmentation task at hand, one functional will likely perform better than the other. We simply illustrate here that novel, as well as traditional, functionals can be optimized with GA.

### 4.2.3 Genetic Algorithms

GA are an optimization technique that models our understanding of evolution. In essence, a number of simultaneous potential solutions (the population) each having an encoded state (the chromosome) perform a random walk (mutations) around the search space, while forming new solutions from combinations of existing solutions (crossover) and, thus, adjusting and refocusing the efforts of the search on regions of the search space with exceptionally "fit" members (good solutions). A few important choices are made during any application of GA: how to encode the population (binary, integer, decimal, etc.), how to mutate the population (mutate all genes, some genes, etc.), how to select the parents for crossovers (roulette wheel, tournament selection, etc.), how to perform those crossovers (uniform, single-point, multi-point, etc.), and finally what fitness function to use for evaluation. Though these choices seem complex, we found through experimentation that different genetic operators yielded negligible impact on our results. The interested reader is referred to [105] for an in-depth review of GA.

**Algorithm 1** GA for HRPCA

**Require:** A set of images, $\mathbf{I}$, an HRPCA matrix, $\mathbf{M}$, a fitness function $Fit$, and GA variables:

Population size, $PopSize = 32$.
Number of crossovers-per–evolution, $CrossNum = 9$.
Maximum number of standard deviations to deform, $MxStdDev = 2$.
Lower- and Upper-bounds on the position of the reference node, $x_1$, $(Min_x, Min_y)$ and $(Max_x, Max_y)$. Always set such that the extent of the mean shape model lies within the image bounds.
Number of modes of variation to consider (for each $d, l, s$), $k = 5$.
Population, $\mathcal{P}$, with members denoted $\mathcal{P}(i)$ and base-node position for $\mathcal{P}(i)$: $(Min_x, Min_y) \leq (\mathcal{P}(i).t_x, \mathcal{P}(i).t_y) \leq (Max_x, Max_y)$; model orientation angle $0 \leq \mathcal{P}(i).\theta \leq \pi$; scale $(0.6, 0.6) \leq (\mathcal{P}(i).s_x, \mathcal{P}(i).s_y) \leq (1.4, 1.4)$; and HRPCA weights $\mathcal{P}(i).\alpha$.
Number of iterations allowed without improvement, $ChangeIter = 40$.
Maximum number of iterations allowed, $MaxIter = 500$.
The best population member seen so far, *best*, with fitness $best.fit$.

**procedure** GA-HRPCA
    Randomly initialize $\mathcal{P}$, from a uniform distribution with $\mathcal{P}.\alpha = 0$.
    $\mathcal{P}$,*best* = UpdateBest($\mathcal{P}$,*best*,0)
    *cnt* = 1
    $\mathcal{P}$,*best*,*cnt* = Evolve($\mathcal{P}(i).\alpha$=0,*cnt*) ▷ Search only for top candidates in position, orientation, and scale.
    $\mathcal{P}$,*best*,*cnt* = Evolve($s = N$,*cnt*) ▷ Perform global deformations. There are $4 \times k \times (N - s + 1) + 5 = 25$ free variables at this stage.
    $\mathcal{P}$,*best*,*cnt* = Evolve($s = 37$,*cnt*) ▷ Perform medium scale deformations. There are $4 \times k \times (N - s + 1) + 5 = 1285$ free variables at this stage.
    $\mathcal{P}$,*best*,*cnt* = Evolve($s = 18$,*cnt*) ▷ Perform fine scale deformations. There are now $4 \times k \times (N - s + 1) + 5 = 1665$ free variables at this stage.
    $\mathcal{P}$,*best*,*cnt* = Evolve($s = 11$,*cnt*) ▷ Perform finer scale deformations. There are now $4 \times k \times (N - s + 1) + 5 = 1805$ free variables at this stage.
    **return** *best*
**end procedure**

Continued in **Algorithm 2**.

**Algorithm 2** GA for HRPCA continued.

**procedure** UPDATEBEST($\mathcal{P}$,*best*,*changed*)
  Sort $\mathcal{P}(i)$ in ascending order according to $Fit(i)$.
  **if** $Fit(PopSize) \geq best.fit$ **then**
    $best = \mathcal{P}(PopSize)$ and $best.fit = Fit(PopSize)$
    $changed = 0$
  **else**
    $\mathcal{P}(1) = best$ ▷ Replace the worst member of the population with the best.
    $changed = changed + 1$ ▷ Keep track of how many iterations since fitness last improved.
  **end if**
  **return** $\mathcal{P}$,*best*,*changed*
**end procedure**

**procedure** EVOLVE($Constraints$,*cnt*) ▷ While $best.fit$ is improving and $cnt \leq MaxIter$, perform crossovers and mutations.
  $changed = 0$
  **while** $(changed \leq ChangeIter)$ AND $(cnt \leq MaxIter)$ **do**
    **for** $i = 1$ to $CrossNum$ **do**
      Select parents $(a, b)$ with probability of selection $P_{selection}(a)$ and $P_{selection}(b)$
      Replace $\mathcal{P}(i)$ with crossover on $\mathcal{P}(a), \mathcal{P}(b)$. ▷ Replaces the least-fit individuals.
    **end for**
    $\mathcal{P}$,*best*,*changed* = UpdateBest($\mathcal{P}$,*best*,*changed*)
    **for** $i = 1$ to $PopSize$ **do**
      Mutate $\mathcal{P}(i)$, only modifying $\mathcal{P}(i).\alpha$ with scale, $s$, described by $Constraints$.
    **end for**
    $\mathcal{P}$,*best*,*changed* = UpdateBest($\mathcal{P}$,*best*,*changed*)
    $cnt = cnt + 1$
  **end while**
  **return** $\mathcal{P}$,*best*,*cnt*
**end procedure**

### 4.2.4 Genetic Algorithms for HRPCA

An overview of our GA is given in Alg. 1 including a summary of all the free parameters and the associated settings that were used for all results presented in Section 4.3. Fixed values were set by observing results on the training data. For example, the maximum number of iterations, MaxIter, was set by noting that the algorithm always converged within 500 iterations.

In the subsequent sections, we describe the representation of individuals, the encoding of the model into chromosomes (deformation weights) to be optimized, the method of mutating (deforming) the model, our selection and crossover methods, and our fitness function (energy function to minimize).

#### Population Representation

In MIS using GA, the individuals forming the population represent potential shapes of the target structure, each having some level of accuracy measured by the fitness function (Section 4.2.2). We require a shape representation that allows us to describe and control the shape deformations intuitively and in terms of our calculated shape statistics. Consequently, we represent each individual by its associated stretching, bending and thickness profiles along with its global orientation, base-node position, and scale (Section 4.2.1).

#### Encoding the Shape Parameters of the Population for GA

We use chromosomes to represent the set of all the weights of the principal components as obtained from the HRPCA, where each gene represents a weight (as a floating point number) for a particular deformation, location, scale and mode of variation (Fig. 4.7-top). In total, there are at most $\sum_1^d \sum_{l=1}^{N} \sum_{s=1}^{N-1+1} s = d * (1/6)(2N + 3N^2 + N^3)$ weights available for mutation since, for each of the four deformations, $d = 4$, we have $N = 100$ different locations, but for each location, $l$, we can only have up to $N - l + 1$ scales, $s$, each of which has at most $s$ weights for the $s$ principal components (Fig. 4.5). In our application, this would add up to $686,800$ dependent variables for our model, which motivates the need for GA with a coarse-to-fine approach to search the highly multivariate space. The large number of variables is based on the assumption that for each deformation type, location, and scale all modes of variation are used. However, in practice we found that only a maximum of 5 modes of variation for each deformation type at each location and scale are needed to account for

at least 90% of the variance, with at most $s$ modes for $s \leq 5$. As previously discussed, there is a level of redundancy in the HRPCA space whereby any large scale deformation can be achieved by a combination of smaller scale deformations. Although smaller scale deformations have more valid locations and therefore more variables to optimize, they are desirable in that they better represent smaller, localized changes in the model. To avoid optimizing a needlessly high number of variables early on, we slowly decrease the scale of the deformation in a coarse-to-fine manner, and never optimize over more than one scale at a time (see *Mutations and Crossovers* section for details).

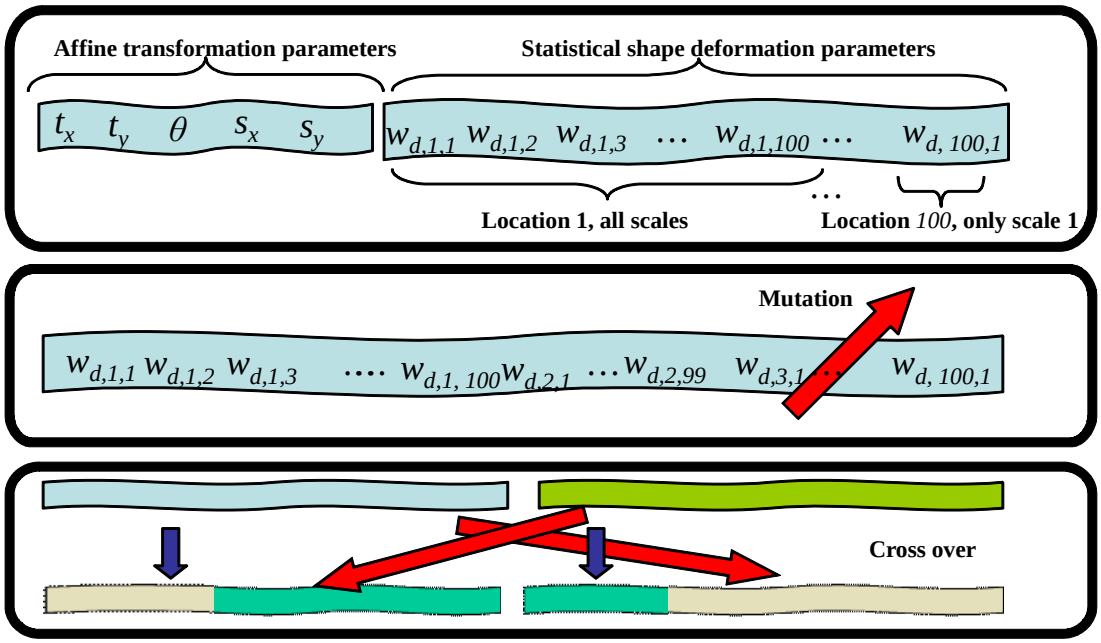



Figure 4.7: Segmenting an anatomical structure amounts to finding the optimal set of shape parameters. (Top) In our GA implementation (Section 4.2.4), we represent each shape as a chromosome with genes encoding affine and statistical shape deformation parameters. Affine transformation parameters are those encoding the global rotation, scale, and position of the shape. Statistical shape deformation parameters are those representing weights for a particular deformation $d$, location $l$, and scale $s$. As noted in Fig. 4.5, Section 4.2.4, endmost locations have only a restricted scale. (Middle) Mutation (Section 4.2.4) is performed by altering some weights of the HRPCA. (Bottom) Crossover amounts to swapping a set of weights between two individuals.

### Mutations and Crossovers

As previously discussed, GA use mutations to walk randomly around the search space and crossovers to initialize the search in new locations that result from combinations of pairs of current solutions.

With regards to mutations, in order to steer the evolution in a coarse to fine approach and, thereby, facilitate faster initial convergence, we employ a coarse-to-fine approach. Initially we constrain the mutations to the affine transformation parameters: base-node position (translation values) $(t_x, t_y)$, model orientation angle $\theta$, and scale values $(s_x, s_y)$ (Fig. 4.7-top). Since our initial shape is the mean CC (obtained by setting all weights, $\alpha_{dls}$ in (4.2), to zero), it can be expected to provide a reasonably strong fitness value when an acceptable position, orientation, and scale are set. In essence, we eliminate the possibility of getting a low score for a good location, scale and orientation, simply because of a bad random shape mutation.

With an adequate location, scale, and orientation obtained, we allow the mutations to begin including shape deformations (Fig. 4.7-middle). Dynamic mutation of a single gene amounts to altering the corresponding weight by sampling it from a uniform random distribution under the constraint that the total weight lies within $\pm 3$ standard deviations of the corresponding mode of variation (square root of the explained variance obtained in HRPCA). Modifying a weight will change the medial profiles and hence the reconstructed shape boundary (Section 4.2.1).

During the initial phases of the evolution, every member of the population undergoes a random deformation with a global scale, and at random amplitudes set in multiples of the corresponding standard deviations, thus resulting in a new shape. The initial constraint to global deformations is well-suited for our statistical deformations as localized deformations (say bulging the splenium in Fig. 4.4) will not help until an acceptable global fit is obtained. Consequently, after no noticeable change in population fitness occurs, we allow the deformations to begin varying in both position and scale to include at first larger deformations (those corresponding to an entire anatomical region and, hence, a primary area of variation), and then smaller deformations which amount to small variations in local regions (Fig. 4.8). This process is outlined in algorithm Alg. 1, where the *Evolve* function is sequentially called with decreasing scale.

In essence, GA use crossover to combine the information from two existing *"parents"*

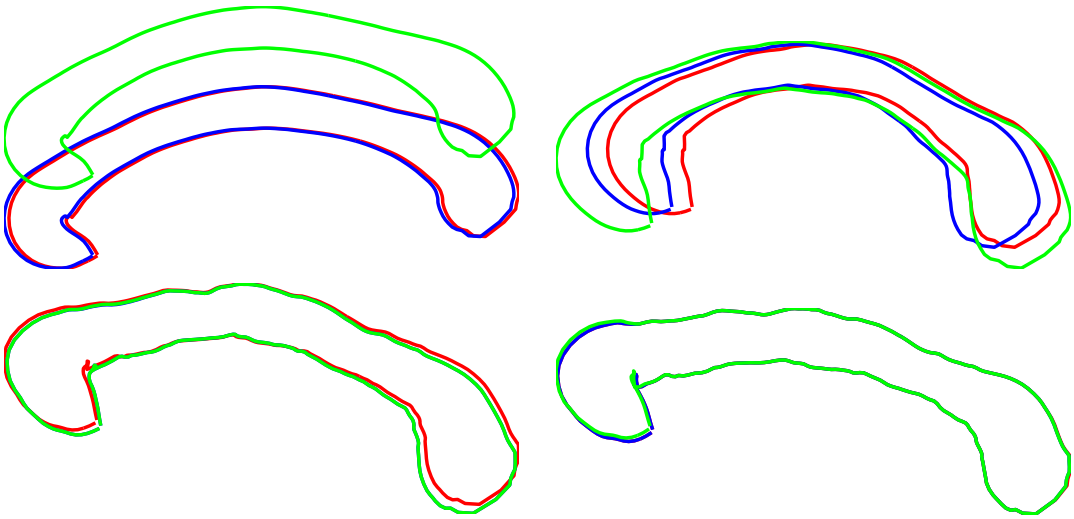


Figure 4.8: (Color Figure) Examples of three sequentially performed mutations on four different members of the population. The scale of the mutation is decreased, progressing from left-to-right top-to-bottom, from affine, to global scale, to medium scale, and finally small scale. Note that the magnitude of each deformation is entirely random, and only the range over which the deformation operates is decreasing. In all images, the red shape was mutated into the green, and then green was mutated into blue.

into a single *"offspring"* that contains genes from each parent. We used uniform crossover, which makes an independent random decision for each gene whereunder both parents have an equal probability of making the contribution (Fig. 4.7-bottom and Fig. 4.9).

### Selection

Genetic algorithms require a method of determining which members of a generation will reproduce, and which will survive. We use roulette wheel selection to randomly select members for reproduction. Our goal is to ensure that the fittest members of the population have the highest probability of selection in proportion to how much their fitness dominates the rest of populous. Each member $i$ of a population $\mathcal{P}$ has a probability of selection equal

Figure 4.9: (Color Figure) Example crossover operations. Each image shows an overlay plot of the two parents and the resulting offspring. The figure shows parents in red and blue, with offspring in green. Note how the offspring carries shape properties from both parents and thus appears somewhat 'in between' the two parents.

to

$$P_{selection}\left(i\right) = \frac{\left(Fit\left(i\right) - \min\limits_{k\in\mathcal{P}} Fit\left(k\right)\right)}{\sum\limits_{j\in\mathcal{P}}\left(Fit\left(j\right) - \min\limits_{k\in\mathcal{P}} Fit\left(k\right)\right)} \tag{4.6}$$

where $Fit(i)$ is the fitness function of the member $i$ (Section 4.2.2), and $\min\limits_{k\in\mathcal{P}} Fit\left(k\right)$ is the minimum fitness value of the population. The normalization (subtracting of the minimum fitness value and dividing by the summation) in (4.6) is performed to ensure valid probabilities for even negative valued fitness functions. We experimented with different methods for determining probabilities from negative fitness functions, and noted only negligible impact on our results. We also employ an *"elitist"* strategy under-which the best member of the population is always maintained, and the weakest of the set are replaced by the new individuals resulting from the crossover operation. A set number of crossover operations are performed during each iteration of the genetic algorithm. We empirically set the number of crossover operations per iteration to 9 for all results, with different experimental values yielding negligible impact.

## 4.3 Results and Discussions

We demonstrate our work through its application to CC segmentation in mid-sagittal brain magnetic resonance images (MRI). The CC is a bundle of white matter fibre connecting the left and right brain hemispheres, and whose cross sectional area and shape correlate with neurological diseases such as multiple sclerosis and schizophrenia [84, 117, 129]. The vast amount of data acquired in clinical studies of these diseases motivates the necessity of a reliable CC segmentation algorithm.

We conduct all of our experiments on a set of 50 mid-sagittal magnetic resonance images [139], presenting both qualitative and quantitative results of the fully-automatic segmentation of the CC using our GA driven, statistically-constrained deformable models. All experiments are performed using leave-one-out validation for training the shape and intensity priors. In order to discount pose problems for some of the methods tested, all of the data has been affinely registered. We do, however, still search over pose in our GA-HRPCA method. As our tested energy functions are all affine invariant, it should be clear that including or not including the affine registration step for novel images does not change the results for GA-HRPCA. The affine registration step does, however, allow us to obtain globally optimal solutions to the method in [35], which we will use for comparison.

For clarity, we divide our experiments into two sets: those dealing with the strict validation of GA-HRPCA in its ability to segment the CC; and those exploring the aforementioned trade-off between energy function and shape space fidelity, and convexity. All error results in each set are computed using the Jaccard distance:

$$\varepsilon = 1 - (|Auto \cap Manual|)/|Auto \cup manual|, \tag{4.7}$$

where $\varepsilon = 0$ represents the ideal segmentation, $|z|$ denotes the area of $z$, and $Auto$ and $Manual$ denote the automatic segmentation and the manual expert delineation, respectively [69].

### 4.3.1 Experiment Set A: Validating GA-HRPCA

We run a total of 500 evolution iterations per image, which takes about 8.5 minutes per image on a 2 GHz AMD Opteron, implemented in MATLAB© (2010a, The MathWorks, Natick, MA). The most expensive operation is the crossover operation since, as previously mentioned, there are a larger number of dependent variables that must be copied between

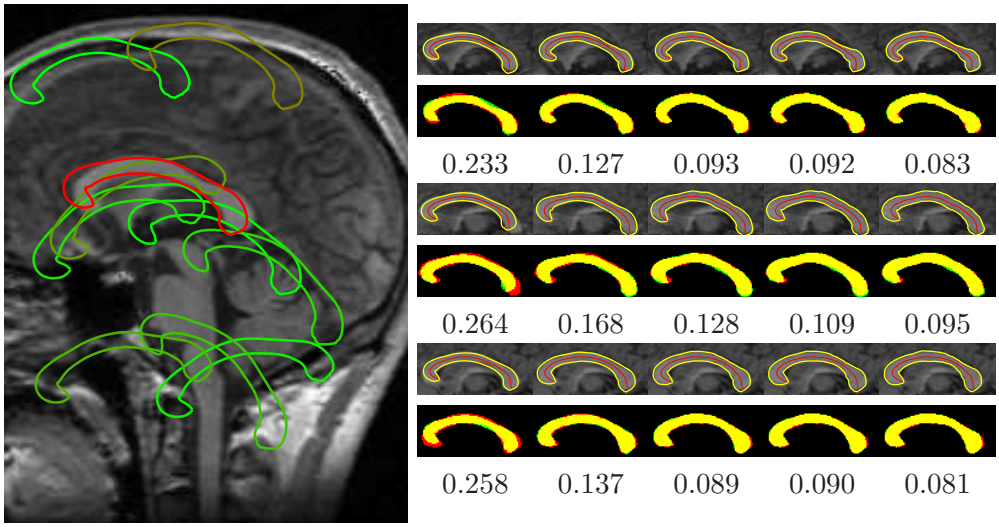



Figure 4.10: (Color Figure) Example segmentations using GA-HRPCA. (left) Finding the pose for one example image, showing ten evenly sampled members from the population, in terms of fitness. Red indicates the fittest model moving linearly to green as the worst of the ten. (right) Three example segmentation (six rows) results progressing left to right, showing fittest individual after population initialization (left), global deformations (next two), and local deformations (last two). Odd image rows show a cropped plot overlaid on the MRI, with the medial axis highlighted in red. Cropping is only for visualization purposes, and all results are produced on full images. Even image rows show color images with the red channel set equal to the ground truth segmentation, the green channel the GA-HRPCA segmentation, and all zeros for the blue channel, i.e. green for true-negatives, yellow for true-positives, red for false-negatives, and green for false-positives. Below each example segmentation, error $\varepsilon$ is shown.

the models. As already noted, in order to speed up convergence, we employ a coarse-to-fine approach where deformations progress from larger to increasingly smaller scales. Again, the basic idea is that initially there is little need for fine, small scale model adjustments since the model may not be in the right position at all. This process is depicted in Figs. 4.10 and 4.11. Notice that although the finer scale deformations represent a small amount of the global shape variance, they allow a significant reduction in error by reducing the localized errors in the fitting that remained after the global fit (second image in each row in Fig. 4.10).

We conducted three main experiments with GA-HRPCA. Since GA are randomized algorithms and may not always produce the same result, we run the algorithm 24 times

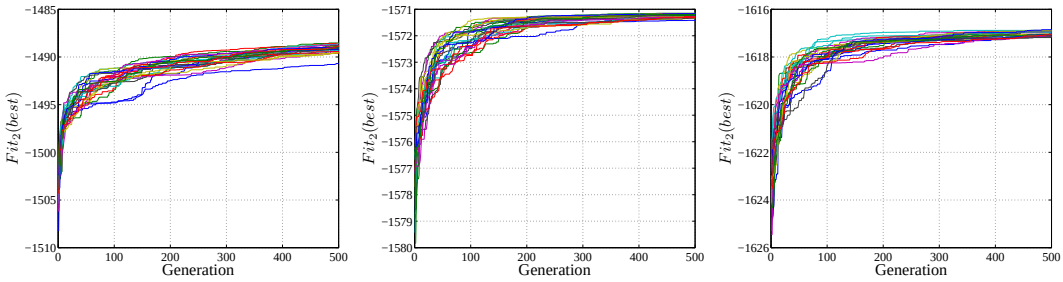



Figure 4.11: Plot of fitness of the best individual versus generation number for the three sample MRI images from Fig. 4.10, for 24 different runs per image. The slope of the curve demonstrates a rapid increase in fitness as the CC is located and the shape model heavily deformed (first 100 generations) and a more gradual increase as the shape model is manipulated to an increasingly fine degree.

per image. Our first experiment was performed using $Fit_1$, (4.3). We learn the optimal $\mathbf{w}$ for $Fit_1$ and $Fit_2$ using the max-margin technique [150]. We employed the max-margin technique in these experiments as we intentionally wanted to use a fixed set of weights for all images to better facilite the cross-comparisons of all the tested methods. The optimal weights for $Fit_1$ where $w_1 = 0.0090$, $w_2 = 0.1101$, $w_3 = 0.8809$, $w_4 = 0$. Note that $F_1, ..., F_4$ are strictly positive for positive valued images, and therefore each term in $Fit_1$ is normalized between 0 and 1. Setting $w_4 = 0$ means that through training we learned $F_4$ was misleading the fitness function, and was unable to accurately predict the standard deviation of the CC. Our second experiment was performed using $Fit_2$ in (4.5), where the optimal weights for $Fit_2$ were: $w_1 = 0.0145$, $w_2 = 0.9707$, $w_3 = 0.0148$. We note that for $Fit_2$, $w_2 \gg w_1$ does not imply that one term has vastly more influence than the other because their scales are not normalized. From here on we refer to this experiment with $Fit_2$ as $Fit_2 - A$. Our third experiment was also performed using $Fit_2$, but we used the fittest result from all 24 runs for each image. In essence, we increase the population size of the GA by a factor of 24. We refer to this experiment with $Fit_2$ as $Fit_2 - B$. The same experiment was conducted using $Fit_1$ without any noticeable improvement, and we defer discussion until the end of this section. To give a visual impression of the segmentation error we present some results of $Fit_2 - B$ in Fig. 4.12, evenly sampled from least to most error. We note that even the worst result, far-right, is still a very plausible CC segmentation.

For point of interest we also conducted a fourth experiment to validate that our GA-HRPCA shape model can fit each ground truth segmentation without error. Simply put,

we segmented the ground truth segmentations directly, using $Fit = -\varepsilon$. The average error was indeed $\varepsilon \approx 0$, validating that the GA-HRPCA model can accurately segment CC data, given an ideal energy function.

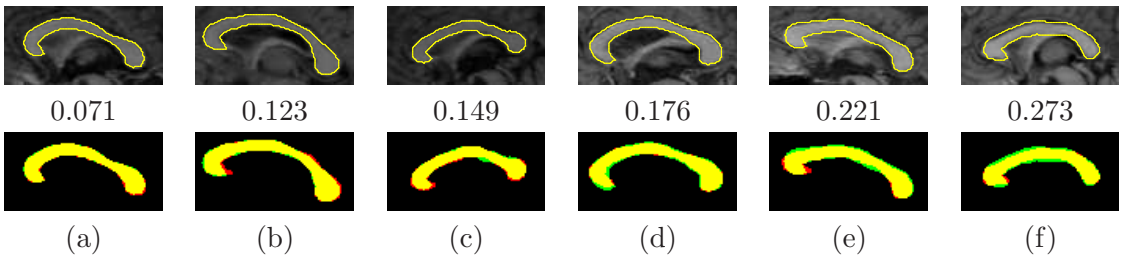



Figure 4.12: (Color Figure) Example segmentations performed using GA-HRPCA $Fit_2 - B$. Six different images are shown, progressing from least-to-most segmentation error under evenly spaced samples. Note how even the worst segmentation is still a plausible CC segmentation. Top row shows segmentations overlaid on cropped MRI, with $\varepsilon$ underneath. Bottom row shows color images with the red channel set equal to the ground truth segmentation, the green channel the GA-HRPCA segmentation, and all zeros for the blue channel, i.e. green for true-negatives, yellow for true-positives, red for false-negatives, and green for false-positives.

Results for our three main experiments are shown in table 4.1. Since GA are randomized, and we segment each image 24 times, we first compute expected error values for each image by averaging the error results from all 24 runs. The 50 resulting expected error values can then be used to compute statistics in expected error across images, as is done in table 4.1. In table 4.2, the standard deviation (std. dev.) is calculated across all 24 runs, and then statistics (mean, max and std. dev.) are calculated over the 50 images to demonstrate that even though GA are a randomized method, the average std. dev. is still low (i.e. for a given image, there is little change in its error between runs). These numbers cannot be computed for $Fit_2 - B$ because there is effectively only one run of that method. Notice that the std. dev. is about twice as large across images than within each image, for example 0.060 vs 0.022 for $Fit_1$. This indicates that the variation in error as a result of GA-HPRCA being a randomized algorithm is small in comparison to the variation caused by how accurately the energy function itself reflects the desired image properties. From that we conclude that in order to improve segmentation results a new energy function should be sought as opposed to attempting to improve the optimizer or, in our case, *tweak* the GA. Note that

we specifically say energy function here, not shape space, because we have already shown the HRPCA shape model can fit any of the ground truth CCs ($Fit = -\varepsilon$).

There is, however, an outlying case where the std. dev. across numerous runs of the GA for a specific image is larger than normal for $Fit_2 - A$ (at 0.118 it lies at over 4 std. dev. from the mean of 0.023). This explains the significant reduction in maximal error obtained when moving from $Fit_2 - A$ to $Fit_2 - B$. The increase in accuracy indicates that some local optima were obtained in $Fit_2 - A$, but can be surpassed with a more exhaustive search. This raises the question of whether we are better off with guaranteed global optima from convex functions, or possibly local optima from our GA-HRPCA method. In the next section we conduct a few experiments with global and non-global methods to explore this question.

It is interesting to note that the minimal error increases from $Fit_2 - A$ to $Fit_2 - B$. This is because in one case improving fitness leads to increased segmentation error, $\varepsilon$ , implying the fitness does not directly correlate with accuracy. Though it may seem an odd result, this is not uncommon for both convex or non-convex functions used in practice for segmentation.

Comparing our two potential fitness functions, $Fit_1$ and $Fit_2$, in tables 4.1 and 4.2 shows that $Fit_1$ obtained superior results to $Fit_2 - A$ and $Fit_2 - B$. Interestingly, $Fit_1$ did not noticeably benefit from a more exhaustive search. We believe $Fit_1$'s better performance stems from its simpler assumption of an exponential relationship for intensity having fewer parameters to estimate and, therefore, being more resilient to noise than the potentially over-fit learned likelihoods in $Fit_2$. The maximal std. dev. and std. dev. of std. dev. are also much lower for $Fit_1$, and this explains why a similar reduction in error was not observed for $Fit_1$ when conducting a more exhaustive search.

Table 4.1: Error results for our proposed GA-HRPCA. Results are first averaged across all runs of each image, and then statistics are computed across those values for 50 images.

| **Method** | **mean** $\varepsilon$ | **median** $\varepsilon$ | **min** $\varepsilon$ | **max** $\varepsilon$ | **std. dev.** $\varepsilon$ |
|---|---|---|---|---|---|
| GA-HRPCA $Fit_1$ | 0.176 | 0.161 | 0.081 | 0.360 | 0.060 |
| GA-HRPCA $Fit_2 - A$ | 0.211 | 0.197 | 0.093 | 0.495 | 0.071 |
| GA-HRPCA $Fit_2 - B$ | 0.202 | 0.192 | 0.116 | 0.347 | 0.055 |
| GA-HRPCA $Fit = -\varepsilon$ | 0.031 | 0.029 | 0.018 | 0.052 | 0.008 |

Table 4.2: Error results for our proposed GA-HRPCA with std. dev. calculated across runs for each image, and then statistics computed across images.

| **Method** | **mean std. dev.** $\varepsilon$ | **max std. dev.** $\varepsilon$ | **std. dev. of std. dev.** $\varepsilon$ |
|---|---|---|---|
| GA-HRPCA $Fit_1$ | 0.022 | 0.073 | 0.012 |
| GA-HRPCA $Fit_2 - A$ | 0.023 | 0.118 | 0.023 |
| GA-HRPCA $Fit = -\varepsilon$ | 0.006 | 0.010 | 0.002 |

### 4.3.2 Experiment Set B: Fidelity versus Convexity

We segment the same set of 50 CC images under a few popular frameworks:

1. Geodesic Active Contours (GAC) [17]: Gradient descent performed on a non-convex level-set–based deformable model with no shape prior. The energy functional itself is convex, and is modeled by the second term in (4.5), $h(x)\,|\nabla q(x)|$. Non-convexity is induced by the level-set shape representation (i.e. non-convex domain).

2. Geocuts [11]: A globally optimal solution to the GAC energy using graph cuts.

3. Random Walker (RW) [52]: A submodular (convex) graph-based formulation with a globally optimal solution and no shape information.

4. Cremers 2008 [35]: A globally optimal segmentation method using $Fit_2$, (4.5), with a global shape prior[2] implemented via PCA (i.e. convex, implicit, global statistics instead of non-convex, explicit, localized ones).

We note that although Geocuts and RW do not have explicit shape priors they still impose a specific form of shape constraints; the seed points function as hard constraints in both methods that shrink the set of possible segmentations and thus form an allowable shape space, albeit a user created one. It is important to note that as Geocuts and RW are interactive methods, and our method is fully-automatic, we are not validating against them, but rather using them to explore the results of minimizing convex functions with varying degrees of shape constraints. For exploration, we choose to use random seeds drawn from

[2]This changes the third term slightly, but both will end up being Gaussian priors centered on the mean shape, with learned variation from PCA.

the ground truth foreground and background, and we report the percentage (%) of pixels used as seeds in each of our experiments (Fig. 4.13). For example, a 1% seeding means that, on average, one seed is placed in every $10 \times 10$ pixel region of the image. The user could draw better seeds, i.e. fewer seeds that would also produce less segmentation error. However, this is essentially providing a greater constraint (fewer seeds in number, higher in tightness, i.e. more restricted shape space). We also note that although Geocuts and RW do not have weights between competing terms, they each have an inherit parameter. The Geocuts method has the free parameter $p$ in $h = \frac{1}{1+|\nabla \mathrm{I}|^p}$, and RW has the $\beta$ parameter which controls the mapping from intensity differences to edge weights (see [52] for details). We ran each method over a large set of parameter values. Thus for each method, we have optimized its free parameters to minimize differences in accuracy stemming from a lack of parameter tuning, and instead highlight the best results that each method can obtain. To that same end, we initialize GAC with the ground truth segmentation, thereby placing it under optimal conditions. For comparison purposes that will become clear shortly, we also minimize the GAC energy function using GA-HRPCA, i.e. $Fit_2$ with $w_1 = 0$, $w_2 = 1$, $w_3 = 0$, we denote this method as GAC–GA-HRPCA.

Our reasons for choosing the above methods are to compare and contrast the consequences of each competing method's trade-offs with the trade-offs present in GA-HRPCA. These trade-offs are highlighted in table 4.3, where the optimizability, shape flexibility, and shape fidelity are measured. Note that for Geocuts and RW we report multiple entries (rows) each using a different seeding percentage. This enables us to explore how the energy functions behave under increasing amounts of constraint.

Shape flexibility measures the method's ability to fit any valid CC. To measure flexibility, we run each method on the ground truth segmentations directly, and measure the average error using $\varepsilon$. Low values indicate low error, and therefore high flexibility in that the method had no problem segmenting all 50 valid CCs. High values here are due to the shape space being too restrictive to accommodate the shape variability in the target CC. As one expects, methods without shape restrictions (GAC, Geocuts and RW) have perfect flexibility, since they can segment any shape.

Shape fidelity, measures the methods ability to constrain the shape to only valid members of the anatomy, and is estimated as error, under measurement $\varepsilon$, in segmenting novel images. We say estimated because, as reported, shape fidelity implicitly includes energy function error, and is thus an approximation. However, when comparing fidelity between equal

energy functions under different constraints, it is clear that changes in accuracy must be entirely a result of the solver or the constraints, as the energy function itself remains fixed. A perfect score is $\varepsilon = 0$ in both shape columns.

To balance the comparison, we also include the level of manual constraints required, as otherwise fully-manual segmentation would be the best possible method according to the table. Manual constraints are rated on a scale from 0 to 10, 0 being no manual intervention, and 10 being the most of any method surveyed here. These ratings correspond directly to the percentage of the image that needed to be seeded for each image, as seeding required direct knowledge of the ground truth segmentation. Notice that for manual methods, increasing fidelity requires increasing manual intervention. In automatic methods, increasing flexibility and fidelity comes at a loss of optimality, as non-convex shape-spaces are required. We provide a more detailed review of these results below.

Table 4.3: Fidelity vs Optimizability. See text for details.

| **Method** | **optimizability** | **shape flexibility** $\varepsilon$ | **shape fidelity** $\varepsilon$ | **manual constraints** | **automatic constraints** |
|---|---|---|---|---|---|
| RW [52] 10% seeding | Global | 0 | 0.132 | 10 | N |
| Geocuts [11] 10% seeding | Global | 0 | 0.160 | 10 | N |
| RW [52] 5% seeding | Global | 0 | 0.165 | 5 | N |
| Geocuts [11] 5% seeding | Global | 0 | 0.182 | 5 | N |
| RW [52] 0% seeding | Global | 0 | 0.844 | 0 | N |
| Geocuts [11] 0% seeding | Global | 0 | 0.375 | 0 | N |
| GAC [17] | Local | 0 | 0.478 | 0 | N |
| GAC–GA-HRPCA | Local | 0.03 | 0.270 | 0 | Y |
| Cremers 2008 [35] | Global | 0.14 | 0.376 | 0 | Y |
| GA-HRPCA $Fit_1$ | Local | 0.03 | 0.176 | 0 | Y |

Table 4.4: Error comparison between tested methods for 50 images.

| **Method** | **mean** $\varepsilon$ | **median** $\varepsilon$ | **min** $\varepsilon$ | **max** $\varepsilon$ | **std. dev.** $\varepsilon$ |
|---|---|---|---|---|---|
| GAC [17] | 0.478 | 0.433 | 0.094 | 1.000 | 0.297 |
| Geocuts [11] 0% seeding | 0.375 | 0.370 | 0.094 | 0.981 | 0.232 |
| Geocuts [11] 5% seeding | 0.182 | 0.171 | 0.089 | 0.346 | 0.061 |
| Geocuts [11] 10% seeding | 0.160 | 0.149 | 0.086 | 0.309 | 0.057 |
| RW [52] 0% seeding | 0.844 | 0.933 | 0.288 | 0.970 | 0.170 |
| RW [52] 5% seeding | 0.165 | 0.152 | 0.106 | 0.407 | 0.055 |
| RW [52] 10% seeding | 0.132 | 0.117 | 0.070 | 0.343 | 0.057 |
| Cremers 2008 [35] | 0.376 | 0.319 | 0.158 | 0.895 | 0.186 |
| GA-HRPCA $Fit_1$ | 0.176 | 0.161 | 0.081 | 0.360 | 0.060 |
| GA-HRPCA $Fit_2 - A$ | 0.211 | 0.197 | 0.093 | 0.495 | 0.071 |
| GA-HRPCA $Fit_2 - B$ | 0.202 | 0.192 | 0.116 | 0.347 | 0.055 |

Examining the fidelity column of table 4.3 in particular, we notice that the convex methods score poorly unless they are heavily constrained with manual intervention (i.e. high seeding percentage for Geocuts and RW). The high error of GAC, in combination with the fact that we use the ground truth as the initialization, implies directly that a simple convex functional, under few optimization constraints, yields a very high segmentation error. Simply put, there are too many erroneous segmentations with lower energy than the ground truth. This is, of course, expected. Notice, however, that with minimal constraints Geocuts yields almost as high a segmentation error as GAC. Again, this is not surprising; the global optima includes many non-CC edges, while the local optima can get stuck before reaching increasingly erroneous segmentations. Had we initialized GAC randomly instead of using the ground truth segmentation, its error would have been far higher. Adding more constraints to Geocuts demonstrates that under appropriate constraints a global optima to a convex function can yield good results. However, as presented in Fig. 4.13, the constraints need to get rather tight before erroneous segmentations are eliminated. Naturally, a region-based energy function, as is used in RW, suffers the same fate. We have thus far reached a rather expected result, global optima to simple energy functionals can yield good segmentations, but they require significant constraints (the seeding percentage has to be high). As manual constraints are undesirable, automatic constraints are preferable.

Tight shape constraints can be automated through the use of shape prior information, and statistically shape-constrained energy functionals can be solved, as with Cremers 2008. However, in order to introduce a shape constraint that is convex, i.e. can be globally optimized, the constraint is made simple (global-in-scale PCA). It is especially interesting to compare the results from Cremers 2008 with GA-HRPCA. The Cremers 2008 row of results in table 4.3 compares the same energy function, with the same weights, $\mathbf{w}$, to our GA-HRPCA method in $Fit_2 - A$ and $Fit_2 - B$. We believe the Cremers 2008 error is high because not only is the shape space unable to address the variability present in the data (i.e. low shape flexibility), but numerous non-CC shapes are still allowable and have lower energy than other, better segmentations (i.e. low shape fidelity). Similar discussions apply when comparing Geocuts to GAC and GAC–GA-HRPCA, since all of the methods use the same energy functional. Under appropriate constraints GAC has low error, and therefore high fidelity, and our method is able to automatically enforce those constraints. We also note that re-running Cremers 2008 with an increasing number of PCA variation modes had negligible impact on the results, because while more modes increases shape space flexibility, it also reduces fidelity. More specifically, with more modes of variation the convex, implicit, global shape space can include the ground truth segmentation but can also include other more erroneous segmentations, stemming from the fact that the true underlying shape-space is not Gaussian (imagine fitting an ellipse to the letter 'S', to include the entire letter much of the space outside the letter must also be included).

There is, thus, a trade-off present in all choices. Neither locally optimal nor globally optimal methods without shape information seem well-suited to fully-automatic segmentation of the CC. Shape information can be added, and the solution kept global, but the shape information must be kept convex, which may not accurately fit the desired shape space, without simultaneously allowing erroneous shapes. Alternatively, with HRPCA a more expressive shape constraint can be automatically built, but it cannot be solved optimally and thus necessitates GA in order to minimize sensitivity to initialization. A comprehensive set of results is presented in table 4.4.

Notice that the standard deviation between images for each tested method, last column of table 4.4, is in general much greater than the standard deviation across different runs of the same image for our GA-HRPCA method (table 4.1, column 7, **mean std. dev.** $\varepsilon$). This implies that one can expect a greater variation in accuracy by applying a given method to a new image, than from re-running our GA on a specific image.

## 4.4 Conclusions

We have developed a novel segmentation technique by exploring the main concerns associated with both traditional and statistically-based deformable models. By using GA to address the initialization and local minima problems associated with traditional energy-minimization techniques, we propose an alternative to convex optimization methods [52, 35, 11, 112, 14]. In doing so, we trade guaranteed global optimality for a shape space that more accurately fits the true allowable shape domain (localized, medial-based shape statistics), i.e. higher flexibility and fidelity. Using the same energy functional, we showed a significant improvement in accuracy over the globally optimized, convex, global, implicit PCA shape space, Cremers 2008 method. Consequently, in answer to our question posed in the introduction: Is it worth sacrificing the complexity of the energy functional, and thereby reducing its ability to address image variability, in exchange for global optimality, or can more complex functions and shape spaces be utilized with an approximate optimization method and achieve superior accuracy? In our application the sacrifice is not worth it. It is better to trade convexity with guaranteed global optimality for a more expressive functional and shape-space with GA-based optimization that achieves close to optimal solutions. We therefore believe research into finding energy functions and shape spaces with high fidelity, with a de-emphasis on convexity, is a very important area of future research. While it is true that global solutions to any such functions would be desired, our results demonstrate that convexly approximating the non-convex shape-space in the application of CC segmentation causes a higher error than does approximating the global optima of the non-convex space via GA. We imagine the difference will only increase for structures with more complex shape deformations between members of the population, as these structures will form increasingly non-convex shape spaces.

Furthermore, as it's well known that pose introduces an inherent non-convexity in shape-model–based deformable models, for example [35], we provide a method that directly handles the pose problem. In essence, the basic idea with GA is to avoid choosing one pose over another until there is enough evidence suggesting one pose is vastly superior. For example, suppose the energy function had two local optima with different pose but similar shape; a binary image containing a target shape, and a second similar structure with a different pose. Eq. (4.6) ensures that our method would keep approximately half of its population at each pose initially, since population members under each pose would have equal probability

of replacement. Our GA would then start shifting the bulk of the population towards whichever object better satisfies the fitness function once shape deformations were allowed. This is exemplified by Fig. 4.10, as numerous false pose configurations have similar response to the fitness function. However, the pose that corresponds with the CC has a much higher degree of fitness and so it quickly overtakes the focus of the search.

We have demonstrated how GA can be combined with constrained shape deformations to effectively explore the search space of a complex energy functional, thereby incorporating prior-knowledge into the solution while retaining multiple simultaneous searches of the space. In essence, we have constrained the members of the GA population to lie within the allowable shape domain thus greatly reducing the search space traditionally associated with deformable models, but without sacrificing fidelity to the target anatomical shape space through global PCA and the inherent convex approximations.

Though other works have used GA to drive deformable models [6, 7, 67, 82, 103, 43, 50], to the best of our knowledge, no works have combined GA with statistical-based deformations in a way that yields intuitively constrained deformations (bends, bulges and stretches) that render the results more interpretable by clinicians and enable regional statistical analysis [55].

Our method can handle a variety of classical energy terms [35, 18, 17]. Other terms can be easily added related to texture, color or other image features. There is no requisite that the terms be convex or even differentiable, they simply need to properly reflect the characteristics of the appearance of the target object in the image data. Not requiring terms to be differentiable is an interesting side benefit of using GA.

Our method is also extendable to other segmentation problems. Specifically, given a training set of shapes for a different anatomical structure, one can perform skeletonisation [51] followed by medial profile extraction and, subsequently, HRPCA. More complex topologies than the single medial-axis structure used in our experiments would be an interesting and natural extension to our method. The largest challenge would be learning the underlying medial-structure, but once learned we believe there would be very little modification required to the GA-HRPCA algorithm. One potential modification would be to fit the primary medial-axis first, then extend down to the branches in a coarse-to-fine approach similar that used to explore different scales in algorithm 1.

We are working on extending these ideas to 3D where the genes of the population become the 3D shape representation's parameters [63]. However, it remains to be explored

how much added complexity must be introduced to address complicated 3D objects with possibly varying topology using a medial representation, and whether a medial based shape representation will remain the ideal choice for an evolutionary computing strategy with localized deformations in 3D. Another interesting avenue is the extension to multilabel segmentation, where GA could be used to model numerous, interacting, simultaneously evolving populations, each representing a different anatomical structure.

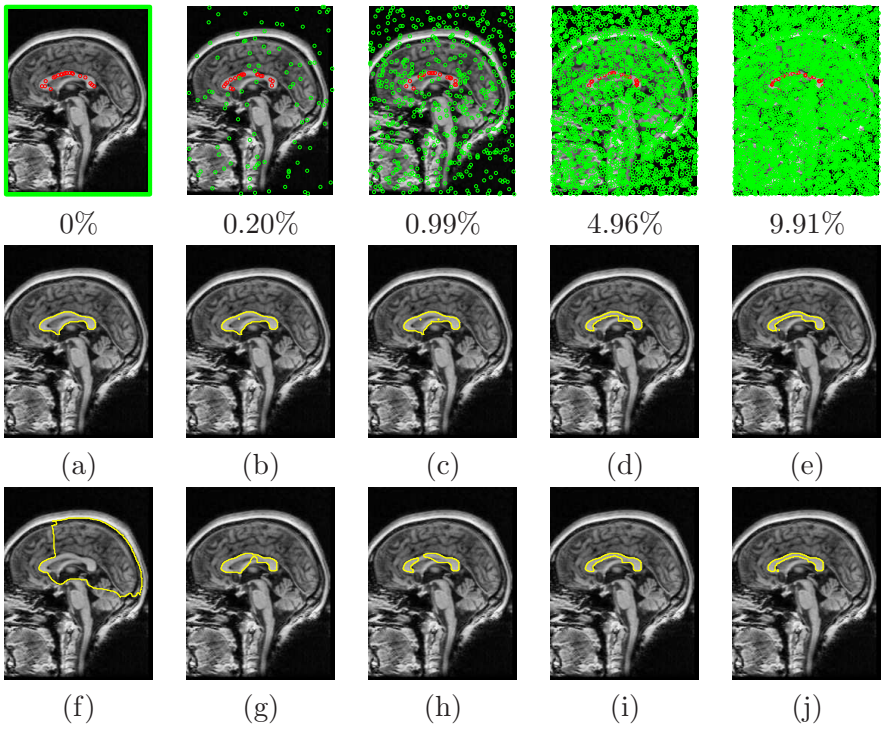



Figure 4.13: (Color Figure) Visualization of Geocuts and RW seeding percentages. Images demonstrate the increasing degree of constraint, in the form of additional seed points, required to prevent these simple functionals from producing highly erroneous segmentations. Top row, seeds with seeding percentage listed (See text for details). Middle row (a-d), Geocuts segmentations for corresponding seed percentages. Bottom row (e-h), RW segmentations. Notice leakage into fornix dip (below the CC) in both methods, due to lack of shape information. As the constraints get tighter, the possible set of segmentations shrinks to exclude the fornix.

# Chapter 5

# Conclusions and Future Work

MIS remains a challenging problem, but one that if solved will bring a number of improved treatment options for patients, and better analysis tools to study diseases. With the end goal being clinically usable segmentation tools, it is important to target the research problems that are responsible for the largest portions of segmentation inaccuracy and unreliability.

This dissertation demonstrated that the weights in energy-minimization–based MIS methods play a large role in accuracy, as does taking convex approximations to non-convex shape spaces. Our results motivated the use of non-convex energy functionals solved via manifold-based initializations, or more generally GA, and customized to the data via weight optimization. Specifically, we have demonstrated the need for varying the energy functional in energy-minimizing segmentation techniques, showing that the large degree in image variability is better addressed by varying the energy functional weights and parameters, than by leaving them fixed. To that end, we detailed a technique to analytically derive the optimal energy functional weights of competing energy terms for training $(segmentation, image)$ pairs, and then infer optimal parameters of novel images utilizing learned image manifolds. The presented methods provide a general formulation that can be directly applied to any energy minimizing segmentation technique (e.g. [17, 18, 76, 126]) to improve its segmentation and increase its ability to generalize to larger datasets. An added benefit was to be able to fully automatically initialize the segmentation process.

We have also demonstrated the application of GA to deformable models, building a novel deformable-model based on localized shape statistics using HRPCA. Our method enables the fitting of complex, medial-based shape models to medical images under arbitrary energy functionals. Our results demonstrate that approximately solving a non-convex functional

yields better results than solving a convex approximation.

We believe future work should turn towards better energy functionals. Specifically, energy functionals incorporating increasing degrees of learned, contextual knowledge [130, 68]. Learning and testing these functionals will require large data sets, and we believe that the MIS community needs to work towards the creation of sets containing thousands of images. These sets need to be anonymized and made publicly available to enable better validation of competing MIS methods.

# Appendix A

# Euler Lagrange

Starting with a function $\Phi(x)$, we build an energy functional:

$$E\left(\Phi\right) = \int L\left(x, \Phi\left(x\right), \Phi_x\left(x\right), \Phi_{xx}\left(x\right)\right)\mathrm{d}x \tag{A.1}$$

where $\Phi_x$ is the first derivative of $\Phi$ with respect to $x$, and $\Phi_{xx}$ the second. The functional $E$ describes a desirable feature of $\Phi$, in that it takes a high value when $\Phi$ is doing poorly, and a low value when $\Phi$ is doing well. We know from Fermat's theorem that $E$ obtains its extremum at a stationary point, i.e. where the derivative is zero. To describe a stationary point one typically writes:

$$\lim_{\tau \to 0} \frac{E(\Phi + \tau\omega) - E(\Phi)}{\tau} = 0 \tag{A.2}$$

where $\omega$ is a scalar, but here $\omega$ is a function, and so this equation is known as the *first variation*, as opposed to fist derivative, of the functional. Developed in the 1750s, the Euler-Lagrange equation plays a fundamental role in variational calculus, defining a necessary condition under which the first variation tends to zero, and therefore the functional reaches a stationary point [41]. For a functional $E$ as above, the general form of the Euler-Lagrange differential equation is:

$$\frac{\partial L}{\partial \Phi} - \frac{\partial}{\partial x}\left(\frac{\partial L}{\partial \Phi_x}\right) + \frac{\partial}{\partial x^2}\left(\frac{\partial L}{\partial \Phi_{xx}}\right) = 0 \tag{A.3}$$

See appendix B for an example of its application to energy minimization methods for MIS.

# Appendix B

# Snakes: Details and Derivations

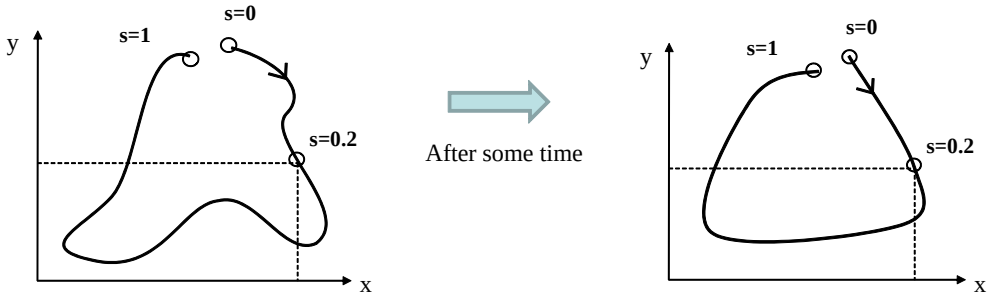



Figure B.1: A parameterized contour undergoing a simple deformation (left-to-right).

Classical deformable shape models [154], are represented as a 2D parametric contour $\mathbf{v}(s) = (x(s), y(s))$, where $s \in [0, 1]$ traverses the contour (Figure B.1), and $\mathbf{v}$ is deformed to fit to image data by minimizing an energy term $\xi$,

$$\xi(\mathbf{v}(s)) = \alpha(\mathbf{v}(s)) + \beta(\mathbf{v}(s)). \tag{B.1}$$

that depends on both the shape of the contour and the image data $I(x, y)$ reflected via the internal and external energy terms, $\alpha(\mathbf{v}(s))$ and $\beta(\mathbf{v}(s))$, respectively.

The internal energy term is given as

$$\alpha(\mathbf{v}(s)) = \int_0^1 \frac{1}{2}\left(w_1 \|\mathbf{v}_s\|^2 + w_2 \|\mathbf{v}_{ss}\|^2\right) \mathrm{d}s. \tag{B.2}$$

Whereas the external energy term is given as

$$\beta\left(\mathbf{v}\left(s\right)\right)=\int_{0}^{1}w_{3}\left(s\right)P\left(\mathbf{v}\left(s\right)\right)\mathrm{d}s. \tag{B.3}$$

The weighting functions $w_1$ and $w_2$ control the tension and flexibility of the contour, respectively, and $w_3$ controls the influence of image data. $w_i$'s can depend on $s$ but are typically set to different constants. For the contour to be attracted to image features, the function $P\left(\mathbf{v}\left(s\right)\right)$ is designed such that it has minima where the features have maxima. For example, for the contour to be attracted to high intensity changes (high gradients) we can choose

$$P\left(\mathbf{v}\left(s\right)\right)=P\left(x\left(s\right),y\left(s\right)\right)=-\left\|\nabla\left[G_{\sigma}*I\left(x\left(s\right),y\left(s\right)\right)\right]\right\| \tag{B.4}$$

where $G_\sigma * I$ denotes the image convolved with a smoothing (e.g. Gaussian) filter with a parameter $\sigma$ controlling the extent of the smoothing (e.g. variance of Gaussian).

The contour $\mathbf{v}$ that minimizes the energy $\xi$ must, according to the calculus of variations [41], satisfy the vector-valued partial differential (Euler-Lagrange) equation:

$$\frac{\partial f}{\partial\mathbf{v}}-\frac{\partial}{\partial s}\left(\frac{\partial f}{\partial\mathbf{v}_{s}}\right)+\frac{\partial^{2}}{\partial s^{2}}\left(\frac{\partial f}{\partial\mathbf{v}_{ss}}\right)=0 \tag{B.5}$$

where $f=\frac{1}{2}\left(w_{1}\left\|\mathbf{v}_{s}\right\|^{2}+w_{2}\left\|\mathbf{v}_{ss}\right\|^{2}\right)-w_{3}\left\|\nabla(G_{\sigma}*I(\mathbf{v}))\right\|$. The first step in applying the Euler Lagrange is to determine the partial derivatives as follows:

$$\begin{aligned}\frac{\partial f}{\partial\mathbf{v}}&=-w_{3}\nabla\left\|\nabla(G_{\sigma}*I(\mathbf{v}))\right\|\\ \frac{\partial}{\partial s}\left(\frac{\partial f}{\partial\mathbf{v}_{s}}\right)&=w_{1}\mathbf{v}_{ss}\\ \frac{\partial^{2}}{\partial s^{2}}\left(\frac{\partial f}{\partial\mathbf{v}_{ss}}\right)&=w_{2}\mathbf{v}_{ssss}\end{aligned} \tag{B.6}$$

then substituting back into eq. (B.5) we get the final set of equations:

$$-w_{1}\mathbf{v}_{ss}+w_{2}\mathbf{v}_{ssss}-w_{3}\nabla\left\|\nabla(G_{\sigma}*I)\right\|=0 \tag{B.7}$$

where we make note that $\mathbf{v}\left(s\right)=\left(x\left(s\right),y\left(s\right)\right)$, and so the above can be expanded to:

$$\begin{aligned}-w_{1}x_{ss}+w_{2}x_{ssss}-w_{3}\frac{\partial\left\|\nabla(G_{\sigma}*I)\right\|}{\partial x}&=0\\ -w_{1}y_{ss}+w_{2}y_{ssss}-w_{3}\frac{\partial\left\|\nabla(G_{\sigma}*I)\right\|}{\partial y}&=0\end{aligned} \tag{B.8}$$

In order to solve the system of equations we introduce an artificial time step, by equating the above equations to $-\frac{\partial \mathbf{v}}{\partial t}$. This yields a first-order iterative optimization method, though as outlined in this report other choices for optimization methods exist.

# Appendix C

# Weight Optimization Convexity

We now show the convexity of (3.3) in $\mathbf{w}$ as follows. Letting $C_i = F_{\mathcal{N}_\mathbf{S}}(S_j, S_i)(S_i - S_j)\,|S_i - S_j|^{-1}$, and expanding the norm and the dot-product in (3.3) to expressly sum over the image domain, $\Omega$, gives:

$$\arg\min \sum_{\mathbf{x}\in\Omega} \left[ \left( \sum_{p=1}^{n} w_p T_p(S_j) \right)^2 - \sum_{i\in\mathcal{N}_\mathbf{S}} \left( C_i \sum_{p=1}^{n} w_p T_p(S_i) \right) \right]. \tag{C.1}$$

Factoring like terms over pixels yields:

$$\arg\min \left[ \left( \sum_{p=1}^{n} \left( w_p \sum_{\mathbf{x}\in\Omega} T_p(S_j) \right) \right)^2 - \sum_{i\in\mathcal{N}_\mathbf{S}} \left( C_i \sum_{p=1}^{n} \left( w_p \sum_{\mathbf{x}\in\Omega} T_p(S_i) \right) \right) \right]. \tag{C.2}$$

Now define two $1 \times n$ vectors: $\mathbf{T}$ with entries $\mathbf{T}_p = \sum_{\mathbf{x}\in\Omega} T_p(S_j)$; and $\mathbf{V}$ with entires $\mathbf{V}_p = \sum_{\mathbf{x}\in\Omega} \sum_{i\in\mathcal{N}_\mathbf{S}} (C_i T_p(S_i))$. By collecting like terms and writing (3.3) in vectorized form, we can formulate a quadratic programming problem:

$$\arg\min \left( \mathbf{w}^T \mathbf{T}^\mathrm{T} \mathbf{T} \mathbf{w} - \mathbf{V}\mathbf{w} \right), \tag{C.3}$$

where $\mathbf{T}^T\mathbf{T}$ is by definition positive semi-definite; ensuring the optimization problem is convex.

# Bibliography


[1] Shawn Andrews, Chris McIntosh, and Ghassan Hamarneh. Convex multi-region probabilistic segmentation with shape prior in the isometric logratio transformation space. In *IEEE International Conference on Computer Vision (IEEE ICCV)*, 2011. 3, 47

[2] D. Anguelov, B. Taskarf, V. Chatalbashev, D. Koller, D. Gupta, G. Heitz, and A. Ng. Discriminative learning of Markov random fields for segmentation of 3D scan data. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 169 – 176 vol. 2, jun. 2005. 9, 27

[3] Alexandre X. Falc ao and Jayaram K. Udupa. Segmentation of 3D objects using live wire. In Kenneth M. Hanson, editor, *Medical Imaging 1997: Image Processing*, volume 3034, pages 228–235. SPIE, 1997. 11, 14

[4] Alexandre X. Falc ao, Jayaram K. Udupa, Supun Samarasekera, Shoba Sharma, Bruce Elliot Hirsch, and Roberto de A. Lotufo. User-steered image segmentation paradigms: Live wire and live lane. *Graphical Models and Image Processing*, 60(4):233 – 260, 1998. 11, 14

[5] Ben Appleton and Hugues Talbot. Globally minimal surfaces by continuous maximal flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:106–118, 2006. 14

[6] L. Ballerini. Genetic snakes for medical image segmentation. In *Proceedings of the SPIE Conference on Mathematical Modeling and Estimation Techniques in Computer Vision*, volume 3457, pages 284–295, 1998. 15, 55, 85

[7] Lucia Ballerini. Genetic snakes for color images segmentation. In Egbert Boers, editor, *Applications of Evolutionary Computing*, volume 2037 of *Lecture Notes in Computer Science*, pages 268–277. Springer Berlin / Heidelberg, 2001. 11, 15, 55, 85

[8] Leah Bar and Guillermo Sapiro. Generalized Newton-type methods for energy formulations in image processing. *SIAM Journal of Imaging Science*, 2(2):508–531, 2009. 14

[9] William A. Barrett and Eric N. Mortensen. Interactive live-wire boundary extraction. *Medical Image Analysis*, 1(4):331 – 341, 1997. 11, 14

[10] H. Blum. Biological shape and visual science. *Theoretical Biology*, 38:205–287, 1973. 51

[11] Y. Boykov and V. Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, volume 1, pages 26 –33, oct. 2003. 6, 11, 14, 49, 79, 81, 82, 84

[12] Yuri Boykov and Gareth Funka-Lea. Graph cuts and efficient N-D image segmentation. *International Journal of Computer Vision (IJCV)*, 70(2):109–131, 2006. 11, 14, 49

[13] Y.Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 105–112, 2001. 11

[14] Xavier Bresson, Selim Esedoḡlu, Pierre Vandergheynst, Jean-Philippe Thiran, and Stanley Osher. Fast global minimization of the active contour/snake model. *Journal of Mathematical Imaging and Vision*, 28(2):151–167, jun. 2007. 14, 19, 49, 84

[15] Xavier Bresson, Pierre Vandergheynst, and Jean-Philippe Thiran. A variational model for object segmentation using boundary information and shape prior driven by the Mumford-Shah functional. *International Journal of Computer Vision*, 68(2):145–162, 2006. 9, 24

[16] Elizabeth Bullitt, Guido Gerig, Stephen Aylward, Sarang Joshi, Keith Smith, Matthew Ewend, and Weili Lin. Vascular attributes and malignant brain tumors. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003*, volume 2878 of *Lecture Notes in Computer Science*, pages 671–679. Springer Berlin / Heidelberg, 2003. 17

[17] Vicent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22:61–79, 1997. 2, 8, 11, 13, 14, 27, 28, 37, 44, 45, 79, 81, 82, 85, 88

[18] T.F. Chan and L.A. Vese. Active contours without edges. *Image Processing, IEEE Transactions on*, 10(2):266 –277, feb. 2001. 2, 8, 13, 27, 28, 37, 41, 44, 45, 85, 88

[19] Tony Chan, Mark Moelich, and Berta Sandberg. Some recent developments in variational image segmentation. In *Image Processing Based on Partial Differential Equations*, pages 175–210. Springer, 2007. 8

[20] Tony F. Chan, B. Yezrielev Sandberg, and Luminita A. Vese. Active contours without edges for vector-valued images. *Journal of Visual Communication and Image Representation*, 11(2):130 – 141, 2000. 11

[21] Chris A. Cocosco, Vasken Kollokian, Remi K.-S. Kwan, G. Bruce Pike, and Alan C. Evans. Brainweb: Online interface to a 3D MRI simulated brain database. *NeuroImage*, 5:425, 1997. 18

[22] I. Cohen, L.D. Cohen, and N.J. Ayache. Using deformable surfaces to segment 3-d images and infer differential structures. *CVGIP*, 56(2):242–263, sept. 1992. 9

[23] Laurent D. Cohen. On active contour models and balloons. *Computer Vision Graphics and Image Processing: Image Understanding*, 53(2):211–218, 1991. 8

[24] L.D. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-d and 3-d images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(11):1131–1147, nov. 1993. 9

[25] Yann Collette and Patrick Siarry. *Multiobjective Optimization Principles and Case Studies*. Decision Engineering. Springer, 2002. 15, 28

[26] T. Cootes, C. Taylor, D. Cooper, and J. Graham. Training models of shape from sets of examples. In *British Machine Vision Conference*, pages 9–18. Springer-Verlag, 1992. 9, 12

[27] T. Cootes, C. Taylor, A. Hill, and J. Halsam. The use of active shape models for locating structures in medical images. In *Proceedings of the 13th International Conference on Information Processing in Medical Imaging*, pages 33–47. Springer-Verlag, 1993. 12

[28] T. F. Cootes, D. Cooper, C. J. Taylor, and J. Graham. Active shape models- their training and application. *Computer Vision and Image Understanding*, 61:38–59, 1995. 9, 12, 50

[29] T.F. Cootes and C.J. Taylor. A mixture model for representing shape variation. In *British Machine Vision Conference*, pages 110–119. Springer-Verlag, 1997. 9, 12, 13, 50

[30] Timothy F. Cootes, Gareth J. Edwards, and Christopher J. Taylor. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):681–685, 2001. 9, 12

[31] D. Cremers. Nonlinear dynamical shape priors for level set segmentation. *Journal of Scientific Computing*, 35(2-3):132–143, jun. 2008. 13, 50

[32] D. Cremers, T. Kohlberger, and C. Schnörr. Nonlinear shape statistics via kernel spaces. In B. Radig and S. Florczyk, editors, *Pattern Recognition (Proc. DAGM)*, volume 2191 of *LNCS*, pages 269–276, Munich, Germany, sept. 2001. Springer. 9, 12, 13, 50

[33] D. Cremers, T. Kohlberger, and C. Schnörr. Nonlinear shape statistics in Mumford–Shah based segmentation. In A. Heyden et al., editors, *Computer Vision - ECCV 2002*, volume 2351 of *LNCS*, pages 93–108, Copenhagen, may 2002. Springer. 9, 12, 13, 50

[34] D. Cremers, T. Pock, K. Kolev, and A. Chambolle. Convex relaxation techniques for segmentation, stereo and multiview reconstruction. In *Advances in Markov Random Fields for Vision and Image Processing*. MIT Press, 2011. 13

[35] D. Cremers, F.R. Schmidt, and F. Barthel. Shape priors in variational image segmentation: Convexity, lipschitz continuity and globally optimal solutions. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1 –6, jun. 2008. 9, 11, 12, 14, 19, 32, 35, 37, 45, 47, 49, 56, 66, 74, 79, 81, 82, 84, 85

[36] S. Dambreville, Y. Rathi, and A. Tannen. Shape-based approach to robust image segmentation using kernel PCA. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 977 – 984, jun. 2006. 9, 12, 13, 35, 44, 50

[37] H. Delingette, M. Hebert, and K. Ikeuchi. Shape representation and image segmentation using deformable surfaces. *Image Vision Computing*, 10(3):132–144, 1992. 9

[38] Hervé Delingette. General object reconstruction based on simplex meshes. *International Journal of Computer Vision*, 32:111–146, 1999. 10

[39] Lee R. Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945. 18

[40] P. Dimitrov, J.N. Damon, and K. Siddiqi. Flux invariants for shape. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–835 – I–841, jun. 2003. 51

[41] L.E. Elsgolc. *Calculus of Variations*. Pergamon Press Ltd., 1962. 4, 90, 92

[42] P. Etyngier, F. Segonne, and R. Keriven. Shape priors using manifold learning techniques. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1 –8, oct. 2007. 9, 12, 13, 50

[43] Yong Fan, Tianzi Jiang, and D.J. Evans. Volumetric segmentation of brain images using parallel genetic algorithms. *Medical Imaging, IEEE Transactions on*, 21(8):904 –909, aug. 2002. 15, 55, 85

[44] Thomas Finley and Thorsten Joachims. Training structural svms when exact inference is intractable. In *ICML '08: Proceedings of the 25th international conference on Machine learning*, pages 304–311, New York, NY, USA, 2008. ACM. 9, 27

[45] P.T. Fletcher, Conglin Lu, S.M. Pizer, and Sarang Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23(8):995–1005, aug. 2004. 13, 50, 51

[46] James D. Foley, Andries van Dam, Steven K. Feiner, and John Hughes. *Computer Graphics: Principles and Practice in C*. Addison-Wesley, 2 edition, aug. 1995. 58

[47] A. Frangi, W. Niessen, R. Hoogeveen, T. van Walsum, and M. Viergever. Quantitation of vessel morphology from 3D mra. In Chris Taylor and Alain Colchester, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI'99*, volume 1679 of *Lecture Notes in Computer Science*, pages 358–367. Springer Berlin / Heidelberg, 1999. 17

[48] M.A. Gennert and A.L. Yuille. Determining the optimal weights in multiple objective function optimization. *Computer Vision*, 5:87–89, 1998. 26

[49] Arthur M. Geoffrion. Proper Efficiency and the Theory of Vector Maximization. *Journal of Mathematics, Analysis and Applications*, 22(3):618–630, June 1968. 28

[50] Payel Ghosh and Melanie Mitchell. Segmentation of medical images using a genetic algorithm. In *Proceedings of the 8th annual conference on Genetic and evolutionary computation*, GECCO '06, pages 1171–1178, New York, NY, USA, 2006. ACM. 55, 85

[51] P. Golland, W. Eric, and L. Grimson. Fixed topology skeletons. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 10 –17, 2000. 85

[52] Leo Grady. Random walks for image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 28(11):1768–1783, nov. 2006. 11, 14, 49, 79, 80, 81, 82, 84

[53] Leo Grady and Gareth Funka-Lea. Multi-label image segmentation for medical applications based on graph-theoretic electrical potentials. *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, pages 230–245, 2004. 11

[54] Leo Grady and Jonathan R. Polimeni. *Discrete Calculus: Applied Analysis on Graphs for Computational Science*. Springer, 2010. 13

[55] Ghassan Hamarneh, R. Abu-Gharbieh, and Tim McInerney. Medial profiles for modeling and statistical analysis of shape. *International Journal of Shape Modeling*, 10(2):187–209, 2004. 48, 51, 57, 58, 59, 60, 61, 62, 64, 85

[56] Ghassan Hamarneh, Preet Jassi, and Lisa Tang. Simulation of ground-truth validation data via physically- and statistically-based warps. In *Lecture Notes in Computer Science, Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 459–467, 2008. 34, 37

[57] Ghassan Hamarneh, Tim McInerney, and Demetri Terzopoulos. Deformable organisms for automatic medical image analysis. In Wiro Niessen and Max Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2001*, volume 2208 of *Lecture Notes in Computer Science*, pages 66–76. Springer Berlin / Heidelberg, 2001. 1

[58] Ghassan Hamarneh and Chris McIntosh. Physics-based deformable organisms for medical image analysis. *Proceedings of SPIE Medical Imaging: Image Processing*, 5747:326–335, 2005. 3, 51

[59] Ghassan Hamarneh and Chris McIntosh. Deformable organisms for medical image analysis (chapter 12). *Deformable Models: Biomedical and Clinical Applications*, pages 387–443, 2007. 3

[60] Ghassan Hamarneh and Chris McIntosh. Physically and statistically based deformable models for medical image analysis (chapter 11). *Deformable Models: Biomedical and Clinical Applications*, pages 335–386, 2007. 3

[61] Ghassan Hamarneh, Chris McIntosh, and Mark Drew. Perception-based visualization of manifold-valued medical images using distance-preserving dimensionality reduction. *IEEE Transactions on Medical Imaging*, 30(7):1314–1327, 2011. 3

[62] Ghassan Hamarneh, Chris McIntosh, Tim McInerney, and Demetri Terzopoulos. Deformable organisms: An artificial life framework for automated medical image analysis (chapter 15). *Computational Intelligence In Medical Imaging: Techniques and Applications*, pages 433–474, 2009. 3

[63] Ghassan Hamarneh, Aaron Ward, and Richard Frank. Quantification and visualization of localized and intuitive shape variability using a novel medial-based shape representation. In *IEEE International Symposiom on Biomedical Imaging (IEEE ISBI)*, pages 1232–1235, 2007. 85

[64] Ghassan Hamarneh, Johnson Yang, Chris McIntosh, and Morgan Langille. 3d live-wire-based semi-automatic segmentation of medical images. In *SPIE Medical Imaging*, volume 5747, pages 1597–1603, 2005. 3

[65] Tobias Hanning, René Schöne, and Georg Pisinger. Vector image segmentation by piecewise continuous approximation. *Journal of Mathematical Imaging and Vision*, 25:5–23, 2006. 10.1007/s10851-005-4385-5. 28

[66] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3D medical image segmentation: A review. *Medical Image Analysis*, 13(4):543 – 563, 2009. 12

[67] A. Hill and C.J. Taylor. Model-based image interpretation using genetic algorithms. *Image and Vision Computing*, 10(5):295–300, 1992. 55, 85

[68] H. Ishikawa. Transformation of general binary mrf minimization to the first-order case. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(6):1234 –1249, june 2011. 89

[69] Paul Jaccard. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin del la Société Vaudoise des Sciences Naturelles*, 37:547–579, 1901. 18, 36, 74

[70] Siddiqi Kaleem and Pizer Stephen. *Medial Representations: Mathematics, Algorithms and Applications*. Springer Publishing Company, Incorporated, 2008. 56

[71] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, 1987. 7, 9, 13

[72] Juha Koikkalainen. Model library for deformable model-based segmentation of 3-d brain mr-images. In Takeyoshi Dohi and Ron Kikinis, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2002*, volume 2488 of *Lecture Notes in Computer Science*, pages 540–547. Springer Berlin / Heidelberg, 2002. 10.1007/3-540-45786-0_67. 27

[73] V. Kolmogorov, Y. Boykov, and C. Rother. Applications of parametric maxflow in computer vision. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, oct. 2007. 9, 27

[74] V. Kolmogorov and R. Zabin. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):147–159, feb. 2004. 13, 14, 19, 65

[75] David Lesage, Elsa D. Angelini, Isabelle Bloch, and Gareth Funka-Lea. A review of 3D vessel lumen segmentation techniques: Models, features and extraction schemes. *Medical Image Analysis*, 13(6):819 – 845, 2009. Includes Special Section on Computational Biomechanics for Medicine. 17

[76] M.E. Leventon, W.E.L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 1, pages 316 –323, 2000. 2, 9, 12, 28, 88

[77] Hua Li and A. Yezzi. Vessels as 4d curves: Global minimal 4d paths to extract 3d tubular surfaces. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW '06. Conference on*, pages 82–82, June 2006. 14

[78] Stan Z. Li. *Markov random field modeling in image analysis*. Advances in pattern recognition. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 3 edition, 2009. 56

[79] Xiaoxiao Liu, Ja-Yeon Jeong, J.H. Levy, R.R. Saboo, E.L. Chaney, and S.M. Pizer. A large-to-fine-scale shape prior for probabilistic segmentations using a deformable

m-rep. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, pages 1 –8, jun. 2008. 56

[80] L. M. Lorigo, O. D. Faugeras, W. E. L. Grimson, R. Keriven, R. Kikinis, A. Nabavi, and C. F. Westin. Curves: Curve evolution for vessel segmentation. *Medical Image Analysis*, 5(3):195 – 206, 2001. 17

[81] Conglin Lu, Stephen Pizer, Sarang Joshi, and Ja-Yeon Jeong. Statistical multi-object shape models. *International Journal of Computer Vision*, 75:387–404, 2007. 10.1007/s11263-007-0045-0. 56

[82] L.A. MacEachern and T. Manku. Genetic algorithms for active contour optimization. In *Circuits and Systems, 1998. ISCAS '98. Proceedings of the 1998 IEEE International Symposium on*, volume 4, pages 229 –232, may 1998. 15, 55, 85

[83] J.B.Antoine Maintz and Max A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1 – 36, 1998. 16

[84] E. Marszal, E. Jamrow, J. Pilch, and et. al. Agenesis of corpus callosum: Clinical description and etiology. *Journal of Child Neurology*, 15:401–405, 2000. 36, 74

[85] U. Maulik. Medical image segmentation using genetic algorithms. *Information Technology in Biomedicine, IEEE Transactions on*, 13(2):166 –173, mar. 2009. 55

[86] T. McInerney and D. Terzopoulos. Topologically adaptable snakes. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 840–845, jun. 1995. 10

[87] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: A survey. *Medical Image Analysis*, 1(2):91–108, 1996. 9, 17, 32

[88] T. McInerney and D. Terzopoulos. Topology adaptive deformable surfaces for medical image volume segmentation. *Medical Imaging, IEEE Transactions on*, 18(10):840–850, oct. 1999. 10

[89] Tim McInerney and Demetri Terzopoulos. A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4d image analysis. *Computerized Medical Imaging and Graphics*, 19(1):69 – 83, 1995. 9

[90] Tim McInerney and Demetri Terzopoulos. Medical image segmentation using topologically adaptable snakes. In *Proceedings of the First International Conference on Computer Vision, Virtual Reality and Robotics in Medicine*, pages 92–101, London, UK, 1995. Springer-Verlag. 10

[91] Tim McInerney and Demetri Terzopoulos. T-snakes: Topology adaptive snakes. *Medical Image Analysis*, 4(2):73 – 91, 2000. 10, 17

[92] C. McIntosh and G. Hamarneh. Vessel crawlers: 3D physically-based deformable organisms for vasculature segmentation and analysis. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 1084 – 1091, jun. 2006. 3, 11

[93] C. McIntosh and G. Hamarnerh. Genetic algorithm driven statistically deformed models for medical image segmentation. In *ACM Workshop on Medical Applications of Genetic and Evolutionary Computation Workshop (MedGEC), in conjunction with the Genetic and Evolutionary Computation Conference (GECCO)*, 2006. 2, 48, 51

[94] Chris McIntosh and Ghassan Hamarneh. I-do: A deformable organisms framework for itk. *Insight Journal*, Special Issue: MICCAI 2006 Open Science Workshop:1–14, 2006. 3

[95] Chris McIntosh and Ghassan Hamarneh. Spinal crawlers: Deformable organisms for spinal cord segmentation and analysis. In *Lecture Notes in Computer Science, Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 808–815, 2006. 3

[96] Chris McIntosh and Ghassan Hamarneh. Is a single energy functional sufficient? adaptive energy functionals and automatic initialization. In Nicholas Ayache, Sébastien Ourselin, and Anthony Maeder, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2007*, volume 4792 of *Lecture Notes in Computer Science*, pages 503–510. Springer Berlin / Heidelberg, 2007. 10.1007/978-3-540-75759-7_61. 2, 9, 21

[97] Chris McIntosh and Ghassan Hamarneh. Evolutionary deformable models for medical image segmentation: A genetic algorithm approach to optimizing learned, intuitive, and localized medial based shape deformation (chapter 4.3). *Genetic and Evolutionary Computation: Medical Applications*, pages 47–67, 2009. 2, 48

[98] Chris McIntosh and Ghassan Hamarneh. Optimal weights for convex functionals in medical image segmentation. In *International Symposium on Visual Computing: Special Track on Optimization for Vision, Graphics and Medical Imaging: Theory and Applications (ISVC OVGMI)*, volume 5875-I, pages 1079–1088, 2009. 2, 9, 21

[99] Chris McIntosh and Ghassan Hamarneh. Medial-based deformable models in non-convex shape-spaces for medical image segmentation using genetic algorithms. *IEEE Transactions on Medical Imaging*, pages (In press, 18 pages), 2011. 2, 48

[100] Chris McIntosh and Ghassan Hamarneh. Optimal parameters in medical image segmentation. *IEEE Transactions on Medical Imaging*, vol(no):(submitted, 19 pages), 2011. 2

[101] Chris McIntosh, Ghassan Hamarneh, and Greg Mori. Human limb delineation and joint position recovery using localized boundary models. In *IEEE Workshop on Motion and Video Computing (IEEE WMVC)*, pages 31–38, 2007. 3

[102] Chris McIntosh, Ghassan Hamarneh, Matt Toom, and Roger Tam. Spinal cord segmentation for volume estimation in healthy and multiple sclerosis subjects using crawlers and minimal paths. In *IEEE Conference on Healthcare Informatics, Imaging and Systems Biology (IEEE HISB)*, pages 25–31, 2011. 3

[103] M. Mignotte and J. Meunier. Deformable template and distribution mixture-based data modeling for the endocardial contour tracking in an echographic sequence. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, page 230, 1999. 55, 85

[104] James V. Miller, David E. Breen, William E. Lorensen, Robert M. O'Bara, and Michael J. Wozny. Geometrically deformed models: a method for extracting closed geometric models form volume data. *SIGGRAPH Comput. Graph.*, 25(4):217–226, 1991. 9

[105] Melanie Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, Cambridge, MA, 1996. 66

[106] J. Montagnat, H. Delingette, and N. Ayache. A review of deformable surfaces: topology, geometry and deformation. In *Image and Vision Computing*, volume 19(14), pages 1023–1040, 2001. 9

[107] Eric N. Mortensen and William A. Barrett. Interactive segmentation with intelligent scissors. *Graphical Models and Image Processing*, 60(5):349 – 384, 1998. 11, 14

[108] David Mumford and Jyant Shah. Optimal approximation by piecewise smooth functions and associated variational problems. *Communications on Pure Applied Mathematics*, 42:577–685, 1989. 8, 28, 37

[109] A. Nakib, H. Oulhadj, and P. Siarry. Image thresholding based on pareto multiobjective optimization. *Engineering Applications of Artificial Intelligence*, In Press, Corrected Proof:–, 2009. 28

[110] O. Ibá nez, N. Barreira, J. Santos, and M.G. Penedo. Genetic approaches for topological active nets optimization. *Pattern Recognition*, 42(5):907 – 917, 2009. 15, 55

[111] Bernard Ng, Rafeef Abugharbieh, and Ghassan Hamarneh. Group mrf for fMRI activation detection. In *IEEE Computer Vision and Pattern Recognition (IEEE CVPR)*, pages 2887–2894, 2010. 51

[112] Mila Nikolova, Selim Esedoglu, and Tony F. Chan. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM Journal on Applied Mathematics*, 66(5):1632–1648, 2006. 13, 14, 49, 84

[113] Sebastian Nowozin, Peter Gehler, and Christoph Lampert. On parameter learning in CRF-based approaches to object class image segmentation. In Kostas Daniilidis, Petros Maragos, and Nikos Paragios, editors, *Computer Vision - ECCV 2010*, volume

6316 of *Lecture Notes in Computer Science*, pages 98–111. Springer Berlin / Heidelberg, 2010. 27

[114] T. O'Donnell, T.E. Boult, Xi-Sheng Fang, and A. Gupta. The extruded generalized cylinder: a deformable model for object recovery. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*, pages 174 –181, June 1994. 11

[115] S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging Vision and Graphics*. Springer Verlag, 2003. 11

[116] Stanley Osher and James A. Sethian. Fronts propagating with curvature dependent speed: Algorithms based on hamilton-jacobi formulations. *JOURNAL OF COMPUTATIONAL PHYSICS*, 79(1):12–49, 1988. 11

[117] J. Pantel, J. Schroder, M. Jauss, and et. al. Topography of callosal atrophy reflects distribution of regional cerebral volume reduction in alzheimer's disease. *Psychiatry Research*, 90:180–192, 1999. 74

[118] N. Paragios, M. Taron, X. Huang, M. Rousson, and D. Metaxas. On the representation of shapes using implicit functions. In Hamid Krim and Anthony Yezzi, editors, *Statistics and Analysis of Shapes*, Modeling and Simulation in Science, Engineering and Technology, pages 167–199. Birkhäuser Boston, 2006. 9, 12

[119] Nikos Paragios and Rachid Deriche. Coupled geodesic active regions for image segmentation: A level set approach. In David Vernon, editor, *Computer Vision - ECCV 2000*, volume 1843 of *Lecture Notes in Computer Science*, pages 224–240. Springer Berlin / Heidelberg, 2000. 11

[120] Nikos Paragios and Rachid Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *International Journal of Computer Vision*, 46:223–247, 2002. 11

[121] D. L. Pham, C. Xu, and J. L. Prince. A survey of current methods in medical image segmentation. *In Annual Review of Biomedical Engineering*, 2:315–338, 2000. 1

[122] SM Pizer, DS Fritsch, PA Yushkevich, VE Johnson, and EL Chaney. Segmentation, registration, and measurement of shape variation via image object shape. *IEEE Transactions on Medical Imaging*, 18(10):851–865, 1999. 56, 58

[123] S.M. Pizer, G. Gerig, S. Joshi, and S.R. Aylward. Multiscale medial shape-based analysis of image objects. *Proceedings of the IEEE*, 91(10):1670 – 1679, oct. 2003. 51

[124] Stephen M. Pizer. Guest editorial—medial and medical: A good match for image analysis. *International Journal of Computer Vision*, 55(2-3):79–84, 2003. 49, 50, 51, 54

[125] Stephen M. Pizer, P. Thomas Fletcher, Sarang C. Joshi, Andrew Thall, James Z. Chen, Yonatan Fridman, Daniel S. Fritsch, A. Graham Gash, John M. Glotzer, Michael R. Jiroutek, Conglin Lu, Keith E. Muller, Gregg Tracton, Paul A. Yushkevich, and Edward L. Chaney. Deformable <i>m-reps</i> for 3d medical image segmentation. *International Journal of Computer Vision*, 55(2-3):85–106, 2003. 10

[126] C. Pluempitiwiriyawej, J.M.F. Moura, Yi-Jen Lin Wu, and Chien Ho. STACS: new active contour scheme for cardiac mr image segmentation. *Medical Imaging, IEEE Transactions on*, 24(5):593 –603, may 2005. 2, 28, 88

[127] Kilian M. Pohl, John Fisher, Sylvain Bouix, Martha Shenton, Robert W. McCarley, W. Eric L. Grimson, Ron Kikinis, and William M. Wells. Using the logarithm of odds to define a vector space on probabilistic atlases. *Medical Image Analysis*, 11(5):465 – 477, 2007. Special Issue on the Ninth International Conference on Medical Image Computing and Computer-Assisted Interventions - MICCAI 2006. 9, 12, 18, 50

[128] R. A. Robb. *Biomedical Imaging, Visualization, and Analysis*. Wiley-Liss Inc., 2000. 1

[129] R. Rosenthal and L.B. Bigelow. Quantitative brain measurements in chronic schizophrenia. *British Journal of Psychiatry*, 121:259–64, 1972. 74

[130] C. Rother, P. Kohli, Wei Feng, and Jiaya Jia. Minimizing sparse higher order energy functions of discrete variables. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1382 –1389, june 2009. 89

[131] Mikaël Rousson, Christophe Lenglet, and Rachid Deriche. Level set and region based surface propagation for diffusion tensor MRI segmentation. In Milan Sonka, Ioannis A. Kakadiaris, and Jan Kybic, editors, *Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis*, volume 3117 of *Lecture Notes in Computer Science*, pages 123–134. Springer Berlin / Heidelberg, 2004. 10.1007/978-3-540-27816-0_11. 11

[132] C. F. Ruff, S. W. Hughes, and D. J. Hawkes. Volume estimation from sparse planar images using deformable models. *Image and Vision Computing*, 17(8):559 – 565, 1999. 55

[133] Christophe Samson, Laure Blanc-Féraud, Gilles Aubert, and Josiane Zerubia. A level set model for image classification. *International Journal of Computer Vision*, 40:187–197, 2000. 10.1023/A:1008183109594. 11

[134] G. Sapiro. Vector-valued active contours. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR '96, 1996 IEEE Computer Society Conference on*, pages 680 –685, June 1996. 11

[135] T.B. Sebastian, P.N. Klein, and B.B. Kimia. Recognition of shapes by editing their shock graphs. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(5):550 –571, may 2004. 51

[136] J.A. Sethian. *Level Set Methods: Evolving Interfaces in Geometry, Fluid Mechanics, Computer Vision and Materials Science*. Cambridge University Press, 1996. 11, 27, 45

[137] H. S. Seung and D.D. Lee. The manifold ways of perception. *Science*, 290:2268–69, 2000. 25

[138] J. Shah. Curve evolution and segmentation functionals: application to color images. In *Image Processing, 1996. Proceedings., International Conference on*, volume 1, pages 461 –464 vol.1, September 1996. 11

[139] M. Shenton, R. Kikinis, F. Jolesz, S. Pollak, M. LeMay, C. Wible, H. Hokama, J. Martin, D. Metcalf, M. Coleman, and et al. Abnormalities in the left temporal lobe and thought disorder in schizophrenia: A quantitative magnetic resonance imaging study. In *New England J. Medicine*, volume 327, pages 604–612, 1992. 74

[140] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:731, 1997. 11

[141] K. Siddiqi, S. Bouix, A. Tannenbaum, and S. Zucker. Hamilton-Jacobi skeletons. *International Journal of Computer Vision*, 48(3):215–231, 2002. 51

[142] A.K. Sinop and L. Grady. A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1 –8, oct. 2007. 14

[143] M. Sonka and J. Fitzpatrick. *Handbook of Medical Imaging, Volume 2: Medical Image Processing and Analysis*. SPIE-International Society for Optical Engine, 2000. 1

[144] Richard Souvenir and Robert Pless. Image distance functions for manifold learning. *Image and Vision Computing*, 25(3):365 – 373, 2007. Articulated and Non-rigid motion. 35

[145] PD Sozou, TF Cootes, CJ Taylor, and EC Di Mauro. Non-linear point distribution modelling using a multi-layer perceptron. In *British Machine Vision Conference*, pages 107–116, 1995. 13, 50

[146] Lawrence H. Staib and James S. Duncan. Deformable fourier models for surface finding in 3D images. In *Proc. Second Conf. on Visualization in Biomedical Computing (VBC-92)*, volume 1808, pages 90–104. SPIE, 1992. 9

[147] L.H. Staib and J.S. Duncan. Boundary finding with parametrically deformable models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(11):1061 –1075, November 1992. 12

[148] Ganesh Sundaramoorthi, Anthony Yezzi, and Andrea C. Mennucci. Sobolev active contours. *International Journal of Computer Vision*, 73(3):345–366, 2007. 13, 14, 17

[149] Richard Szeliski. Bayesian modeling of uncertainty in low-level vision. *International Journal of Computer Vision (IJCV)*, 5:271–301, 1990. 32

[150] Martin Szummer, Pushmeet Kohli, and Derek Hoiem. Learning CRFs using graph cuts. In David Forsyth, Philip Torr, and Andrew Zisserman, editors, *Computer Vision - ECCV 2008*, volume 5303 of *Lecture Notes in Computer Science*, pages 582–595. Springer Berlin / Heidelberg, 2008. 9, 27, 32, 33, 36, 38, 39, 40, 41, 42, 44, 76

[151] T.T. Tanimoto. Technical report, IBM Internal Report, nov. 1957. 18

[152] Ben Taskar, Vassil Chatalbashev, Daphne Koller, and Carlos Guestrin. Learning structured prediction models: a large margin approach. In *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pages 896–903, New York, NY, USA, 2005. ACM. 9, 27

[153] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–23, 2000. 34, 41

[154] D. Terzopoulos. On matching deformable models to images. *Topical Meeting on Machine Vision, Technical Digest Series*, 12:160–167, 1987. 7, 9, 13, 55, 91

[155] J. Tohka. Global optimization of deformable surface meshes based on genetic algorithms. In *Image Analysis and Processing, 2001. Proceedings. 11th International Conference on*, pages 459 –464, sep 2001. 15, 55

[156] A. Tsai, W. Wells, C. Tempany, E. Grimson, and A. Willsky. Mutual information in coupled multi-shape model for medical image segmentation. *Medical Image Analysis*, 8(4):429 – 445, 2004. 11

[157] A. Tsai, Jr. Yezzi, A., and A.S. Willsky. Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification. *Image Processing, IEEE Transactions on*, 10(8):1169 –1186, August 2001. 8, 27

[158] A. Vasilevskiy and K. Siddiqi. Flux maximizing geometric flows. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(12):1565 – 1578, December 2002. 17

[159] Luminita A. Vese and Tony F. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *International Journal of Computer Vision*, 50:271–293, 2001. 11

[160] Nhat Vu and B.S. Manjunath. Shape prior segmentation of multiple objects with graph cuts. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1 –8, jun. 2008. 9, 12

[161] Zhizhou Wang and Baba C. Vemuri. Tensor field segmentation using region based active contour model. In *Computer Vision - ECCV 2000*, volume 3024 of *Lecture Notes in Computer Science*, pages 304–315. Springer Berlin / Heidelberg, 2004. 11

[162] Aaron Ward and Ghassan Hamarneh. Gmat: The groupwise medial axis transform for fuzzy skeletonization and intelligent pruning. *IEEE Transactions Pattern Analysis and Machine Intelligence (IEEE TPAMI)*, 32(6):1084 – 1096, 2010. 51

[163] S.K Warfield, Michael Kaus, Ferenc A Jolesz, and Ron Kikinis. Adaptive, template moderated, spatially varying statistical classification. *Medical Image Analysis*, 4(1):43 – 55, 2000. 9, 12

[164] S.K. Warfield, K.H. Zou, and W.M. Wells. Simultaneous truth and performance level estimation (staple): an algorithm for the validation of image segmentation. *Medical Imaging, IEEE Transactions on*, 23(7):903–921, jul. 2004. 18

[165] Ross T. Whitaker. Volumetric deformable models: active blobs. *Visualization in Biomedical Computing 1994*, 2359(1):122–134, 1994. 9

[166] M. Wimmer, F. Stulp, S. Pietzsch, and B. Radig. Learning local objective functions for robust face model fitting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(8):1357 –1370, aug. 2008. 27

[167] O. Wink, W. Niessen, B. Verdonck, and M. Viergever. Vessel axis determination using wave front propagation analysis. In Wiro Niessen and Max Viergever, editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2001*, volume 2208 of *Lecture Notes in Computer Science*, pages 845–853. Springer Berlin / Heidelberg, 2001. 10.1007/3-540-45468-3_101. 17

[168] O. Wink, W. J. Niessen, and M.A. Viergever. Multiscale vessel tracking. *IEEE Transactions on Medical Imaging*, 23(1):130–133, 2004. 14, 17

[169] Chenyang Xu and J.L. Prince. Snakes, shapes, and gradient vector flow. *IEEE Transactions on Image Processing*, 7(3):359–369, 1998. 8

[170] Jr. Yezzi, A., 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, april 1997. 8

[171] T. S. Yoo. *Insight into Images: Principles and Practice for Segmentation, Registration, and Image Analysis*. A k Peters Ltd., Wellesey, MA, USA, 2004. 1

[172] Jingdan Zhang, Shaohua Zhou, Dorin Comaniciu, and Leonard McMillan. Discriminative learning for deformable shape segmentation: A comparative study. In David Forsyth, Philip Torr, and Andrew Zisserman, editors, *Computer Vision - ECCV 2008*, volume 5302 of *Lecture Notes in Computer Science*, pages 711–724. Springer Berlin / Heidelberg, 2008. 27

[173] Jingdan Zhang, S.K. Zhou, D. Comaniciu, and L. McMillan. Conditional density learning via regression with application to deformable shape segmentation. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1 –8, june 2008. 27

[174] Hong-Kai Zhao, T. Chan, B. Merriman, and S. Osher. A variational level set approach to multiphase motion. *Journal of Computational Physics*, 127(1):179–195, 1996. 11

[175] Barbara Zitová and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977 – 1000, 2003. 16