GEOMETRIC STATISTICALLY BASED METHODS FOR THE SEGMENTATION AND REGISTRATION OF MEDICAL IMAGERY

A Thesis
Presented to
The Academic Faculty

by

Yi Gao

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Department of Biomedical Engineering

Georgia Institute of Technology
May 2011

GEOMETRIC STATISTICALLY BASED METHODS FOR THE SEGMENTATION AND REGISTRATION OF MEDICAL IMAGERY

Approved by:

Professor Allen Tannenbaum, Advisor
School of Electrical and Computer Engineering and Department of Biomedical Engineering
Georgia Institute of Technology

Professor Don P. Giddens, Advisor
School of Biomedical Engineering
Georgia Institute of Technology

Professor Xiaoping Hu
School of Biomedical Engineering
Emory University

Professor Arthur Stillman
Department of Radiology
Emory University

Professor Anthony J. Yezzi
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: 23 November 2010

献给 父亲母亲

献给 妻子

To my parents,

To my wife

ACKNOWLEDGEMENTS

I would like to thank those who have been supporting me and helping me throughout my doctorate years.

First, I would like to thank my advisers, Professor Allen Tannenbaum and Professor Don Giddens. From you I learned not only knowledge and techniques, but also how to be a great researcher. It is and will always be my great honor to be student of you.

I also would like to thank my thesis committee, Professor Xiaoping Hu, Professor Arthur Stillman and Professor Anthony Yezzi for your helpful comments and suggestions, which improved my thesis in many aspects.

Thank you to everyone in our great group: Yan, Behnood, Jehoon, Peter, Ivan, Jacob, Delphine, Eric, Marc, Yogesh, John, Sam, Xavier, Ponnappan, Shawn, Jimi, Gallagher, Rome, Tauseef, Vandana, Eli, Oleg, Jin, Binjian, Stephanie, Massimiliano and Sungho. Thank you for you help all through the five years. And it's always been a pleasure to spend time in the lab, attend group meetings, and share the knowledge and happiness with you.

I would like to thank my intern supervisors, Gozde Unal, Greg Slabaugh and Fang Tong at Siemens Corporate Research, for being great mentors to me and guiding me through the wonderful internship experiences.

I would also like to thank my adviser in the Math department, Professor Christopher Heil. From you I learned a lot about the rigorous and beauty of math.

Thanks to my friends here and in China, for your friendship and you supporting.

To my dear wife Xiaxia, to my parents, for everything.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Medical image analysis aims at developing techniques to extract information from medical images. Among its many sub-fields, image registration and segmentation are two important topics. In this report, we present four pieces of work, addressing different problems as well as coupling them into a unified framework of shape based image segmentation. Specifically:

1. We link the image registration with the point set registration, and propose a globally optimal diffeomorphic registration technique for point set registration.
2. We propose an image segmentation technique which incorporates the robust statistics of the image and the multiple contour evolution. Therefore, the method is able to simultaneously extract multiple targets from the image.
3. By combining the image registration, statistical learning, and image segmentation, we perform a shape based method which not only utilizes the image information but also the shape knowledge.
4. A multi-scale shape representation based on the wavelet transformation is proposed. In particular, the shape is represented by wavelet coefficients in a hierarchical way in order to decompose the shape variance in multiple scales. Furthermore, the statistical shape learning and shape based segmentation is performed under such multi-scale shape representation framework.

CHAPTER I

JUSTIFICATION OF THE RESEARCH TOPIC

Medical image analysis is a very active research area in the recent decades. In particular, the image segmentation and registration are topics drawn lots of attentions. In deed, one of the main objective of medical image segmentation is to assist the doctors in identifying objects of interest in the multi-modality medical imagery. On the other hand, registration solves the problem of aligning the common target in two or more images into one common coordinate system. The two seemingly distinct areas are in fact closely related. In fact, if we register an anatomy atlas to a new image, equivalently we provide a segmentation of the organ/tissues of the new image. Likewise, when the target in two or more images are segmented out, registering them then becomes a much easier problem. Hence, the strategy we take in this proposal is as follows: we first study the two topics separately and then combine the two in a coupled framework.

More explicitly, the point set registration method is discussed in Chapter 2. Then, we turn to address the image segmentation algorithm in Chapter 3. Furthermore, the registration and segmentation is coupled in a shape based segmentation framework in Chapter 4. Finally, the shape is represented in a multi-scale way using wavelet transformation in Chapter 5, and the statistical shape learning and shape based segmentation are further carried out in such multi-scale framework.

More explicitly, the first chapters addresses the point set registration problem. We study the point set registration and propose a global optimal diffeomorphic registration scheme for point sets. The global optimality is achieved by considering the registration as a state estimation problem and solved under the particle filtering framework. Then in Chapter3, the topic turns to image segmentation. Specifically, we develop a method to perform multi-target segmentation. The semi-automatic algorithm is initiated by user provided multi-labeled map, which gives several “seeds” in each expected objects. Then, the characteristics of the targets are learned from the corresponding seeds, which is further used to drive the multiple active contours evolve to the desired boundary positions. Moreover, the interactions among the contours are incorporated to guarantee the mutual-exclusive of them. This scheme effectively reduces the contour leakage in low contrast region. Nevertheless, due to the fact that this method only utilizes the low-level image information (in contrast to high level shape prior), in some cases the results are not sufficiently satisfying.

Hence, Chapter 4 unifies the image registration and segmentation in a coupled framework. That is, the shape of the target object is incorporated in the segmentation process. To this end, the algorithm first learns the shapes prior from a set of training shapes and performs the shape-based segmentation thereafter. Before performing statistical learning of the shape, registration is necessary to remove the pose variance, such that all the variances captured by the subsequent statistical learning is due to shape variances. To perform the registration, techniques proposed in Chapters 2 are used. Then, the training shape space is constructed by the principle component analysis and further shape based segmentation is carried out in this space to utilize both image information as well as the prior shape knowledge.

Unfortunately, the shape representation above suffers from the problem that the small/local scale shape variances are sometimes overwhelmed by the large/global scale shape variances, due to the fact that the former ones usually have larger energy. However, in many cases the small “bump” on a smooth surface is an important indicator of pathology. To address this issue, in Chapter 5 we propose a multi-scale shape representation, using the wavelet transformation. Therefore, the shape is represented and further learned in multiple scales. Based on that, the multi-scale shape based segmentation is utilized in the extraction of hippocampus from the MR images.

Before providing the detail of the methods, we review some of the key results in the related areas.

1.1 Literature review for image registration

The basic objective of image registration can be roughly described as to find a suitable transformation, applying on one of the images, to make it similar to the other. Different researchers, in various application scenarios, take different ways to address the above phrases of “suitable transformation” and “similar”. To measure the similarity between two images, some researchers choose to detect the features in the image first, and then measure the similarity of the two images by the alignment of the two sets of the detected features [17, 64, 62, 51]. Since the features capture the important information in the images thus alignment of them is a sensible answer to the alignment of the two images. Besides, the feature based methods have the advantages that features are sparse representation of the image and thus the computational complexity is not heavy. Moreover, the user interaction can be incorporated through providing the features such as by clicking on the corresponding points in two images. However, it relies heavily on the detection of the features: manually assigning features are time consuming while automatic feature detection is in fact an image segmentation problem where the general solution is not well established. Hence instead of explicitly extracting features before registration, some researchers directly measure the difference between the two images. A straight forward way is to use the L^p norm, or the cross-correlation of the two images [58, 49]. Such model has the limitation that it assumes the two images be of the same modality. To loosen this assumption, authors in [65, 61, 36] used mutual information and authors in [20] used the gradient of the intensity. However, this category of similarity measurements require the main objects to be registered in the two images to be largely overlapped. Moreover, in the cases where one image contains only a portion of the other (partial structure registration), the minimum of the similarity functional does not reflect the visually suitable alignment.

After the similarity measurement is chosen, various transformation models are employed to maximize it. The transformation model is largely dependent on the underlying application. Rigid transformation, though simple, is a suitable choice for intra-subject registration [36, 23]. Similarity [26] and affine transformations are usually considered to be on the "boundary" of human perception about the concept of shape: two objects differing by a similarity or an affine transform may be thought of as having the same shape, while transformations with more degree-of-freedom (DOF) is considered to be of another shape [43, 14]. Registration under transformations with more DOF than affine include those parametrized by B-Spline [52], Thin Plate Spline [4] or NURBS [63]. Furthermore, in [57, 22], the transformation is not parametrized but treated as a vector field whose values are defined pointwisely.

Various optimization schemes are employed to optimize the cost function defined by the similarity measurement over the transformation. Usually the gradient based local optimization schemes are used for transformation model with a few parameters such as rigid, affine or splines cases. The iterative methods are often used for cases where the transformation/deformation is pointwisely defined [57, 22].

It is important to note that almost all of the above schemes only handle the local registration task. Indeed, the global image registration problem is only briefly addressed in [26, 1]. On the other hand, this notion of "global" registration is usually performed using a point set framework. So we plan to address the image registration problem by utilizing the techniques in the point set registration field. To this end, we first review the literature in the point set registration fields.

1.2 Literature review for point set registration

Similar to image registration, point set registration addresses the problem of finding the best matching between two or more point sets. In contrast to image registration, however, point-set registration usually requires an extra step of explicitly finding the correspondences between points in the two sets. Then the transformation parameters can be estimated. This is the main theme of the well-known Iterative Closest Point (ICP) algorithm introduced in [3]. However, the basic ICP approach is widely known to have issues with local minima. To address this issue, Fitzgibbon [16] introduced a robust variant by optimizing the cost functional via the Levenberg-Marquardt algorithm. Moreover, Chui and Rangarajan employed an annealing scheme to broaden the convergence range and reduce the influence of outlier [9]. However the temperature in the annealing scheme needs to be carefully chosen to balance the convergence range and algorithm stability. In addition, point-set registration can be alternatively be viewed as a parameter estimation task, whereby the transformation parameters are considered to be random variables. To this end, the authors of [35, 40, 54] employ filtering techniques to estimate the distribution of *rigid* transformation parameters.

1.3 Literature review for image segmentation

Image segmentation is usually recognized as a process of locating the region of interested in an image. In situations that the contrast between the target and its surrounding regions are high, simply applying thresholding will generate satisfying result [49]. However such simple cases do not occur very much in reality. One common practice consists of user initialization with one or several clicks (often called “seeds”) in the target, and the algorithm then takes over to extract the desired object. A simple but intuitive example using such strategy is the region growing method [56]. Although the formalism is simple and straightforward, it reflects the two key roles of the user initialization: *Position*: the positions of the initial seeds indicate the estimated position of the target; *Feature*: the image information in a given neighborhood of the seeds should be employed to learn the necessary characteristics of the desired object as well as to drive the segmentation. Nevertheless, original region growing only depends on the image intensity, and thus is many times not suitable for noisy and textured imagery. Furthermore, the segmentation boundary is not guaranteed to be as smooth as many times required. To address the first problem, Pichon *et al.* used robust statistics for better modeling of the image features at the locations of the seeds, and a fast marching algorithm to grow the segmentation contour [47]. In order to utilize both image information and contour geometry, Kass *et al.* proposed the active contour method to evolve a contour in a variational manner [28]. However, the original active contour scheme is known to be not easy to be extended to higher dimension as well as difficult to handle the topological change of the contour. To this end, the level set method is utilized to implicitly represent the contour [42]. Under the level set framework, a widely used strategy is to evolve the contour according to the image gradient information [37, 30, 6, 69]. However, the image gradient information is prone to noise perturbation. Hence, authors in [70, 8, 39] utilize the regional image information such as mean, variance, and probability density function to characterize the target feature as well as to drive the contour evolution. Recently, a local regional information driven contour evolution is proposed in [33], which is both robust to image noise and unrelated image information remote from the contour.

Moreover, another desired feature for segmentation is the ability to simultaneously extract multiple objects. This can be quite advantageous in medical image analysis, where several related targets all need to be captured. However, most active contour algorithms are tailored to handle only one target at a time. Thus, the given algorithm needs to be executed sequentially several times in order to obtain the required multiple objects. However, since the individual segmentation processes do not interact with each other, it is difficult to guarantee mutual exclusiveness among contours. To address that, multiple object segmentation has been discussed in several papers [73, 5, 60, 19, 72, 55]. In these works, the algorithms require the contours to be mutually exclusive (not overlapping). In addition, they also assume that the union of the regions bounded by the contours must be equal to the entire image domain.

Nevertheless, the image information alone is often times not enough for algorithms to find the desired boundary. To this end, we need to incorporate the prior knowledge of the shape of the target. A well-known shape based segmentation method, active shape model (ASM), is proposed in [10]. In that work, a set of training shapes are learned before the segmentation process. In the training set, each shape, i.e. 2D contour, is represented by a parametrized curve. For each point on the curve, the corresponding points on other training shapes are known in advance, and their positions are modeled by a Gaussian distribution. The authors apply the principle component analysis to learn the shape variances. Furthermore, given a new image, the contour is evolving constrained by the learned shape information. Moreover, authors in [34, 59] uses the level set function to represent the contour and extend the shape based method to 3D cases.

CHAPTER II

GLOBAL OPTIMAL DIFFEOMORPHIC REGISTRATION FOR POINT SETS

In this chapter, we present a hierarchical method to perform globally optimal diffeomorphic registration, To register the “moving point set” to the “fixed point set”. Accordingly, we first approach the registration through a filtering scheme. Then the point set is hierarchically decomposed, and the registration is performed in finer scales to address the local deformations. Furthermore, the transformations at local levels are then fused to form a smooth and invertible diffeomorphic transformation.

2.1 Preliminary

In order to achieve better performance in point set registration, we adopted the idea by viewing the registration as a state estimation process. Hence, we first provide the algorithm for state estimation: the particle filtering method.

2.1.1 Particle Filtering (PF)

Particle filtering is a sequential Monte Carlo method [53] that provides a sequential estimate of the distribution of the state variable of a dynamical system. We denote the state variable at time t as $\boldsymbol{x}_t$ and the observation as $\boldsymbol{y}_t$. The objective then is to estimate the distribution of $\boldsymbol{x}_t$ based on all the observations made until time t, $\boldsymbol{y}_{1,2,\ldots,t}(=: \boldsymbol{y}_{1:t})$, namely $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$. With this goal, the process and observation models are given as:

$$\begin{cases} \boldsymbol{x}_{t+1} = f(\boldsymbol{x}_t, u_t) & (1) \\ \boldsymbol{y}_t = g(\boldsymbol{x}_t, v_t) & (2) \end{cases}$$

where f and g may be nonlinear functions while u_t and v_t are the process and observation noises, respectively. We assume u_t and v_t are independent and are both independent in time. Further, we assume the distribution of the initial state $p(\boldsymbol{x}_0)$ is known.

The recursive estimation of $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ consists of two steps, namely the *prediction* and *update* steps. Assuming $p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})$ is available, the prediction step gives the prior PDF of $\boldsymbol{x}_t$ at time t as:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})d\boldsymbol{x}_{t-1} \tag{3}$$

where

$$p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \int \delta(\boldsymbol{x}_t - f(\boldsymbol{x}_{t-1}, u))p(u)du. \tag{4}$$

Here δ denotes the Dirac function.

At time t after the observation $\boldsymbol{y}_t$ is available, it can then be used to update the estimation to obtain the posterior PDF:

$$p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t}) = \frac{p(\boldsymbol{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})}{p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1})} \tag{5}$$

where

$$p(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}) = \int \delta(\boldsymbol{y}_t - h(\boldsymbol{x}_t, v))p(v)dv \tag{6}$$

In cases where f and g are non-linear, the analytical result of $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t})$ is rarely available. Therefore, one can expect a numerical approximation of the PDF. To this end, particle filter employs the Bayesian recursion under the Monte Carlo framework.

Firstly, samples (particles) are obtained from the initial prior distribution $p(\boldsymbol{x}_0)$ and they are denoted as $\{\boldsymbol{x}_0(i) : i = 1, \ldots, N\}$. Secondly, we assume the particles $\{\boldsymbol{x}_{t-1}(i) : i = 1, \ldots, N\}$ approximating the density $p(\boldsymbol{x}_{t-1}|\boldsymbol{y}_{1:t-1})$ are available, then the prior distribution of $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})$ is computed. Specifically, the particles $\{\boldsymbol{x}_t^*(i) : i = 1, \ldots, N\}$ approximating the prior density $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1})$ are computed as $\boldsymbol{x}_t^*(i) = f(\boldsymbol{x}_{t-1}(i), u(i))$ where $u(i)$ are realizations of the process noise. Thirdly, with the arrival of $\boldsymbol{y}_t$, the likelihood of each $\boldsymbol{x}_t^*$ is computed as:

$$q_i = \frac{p(\boldsymbol{y}_t|\boldsymbol{x}_t^*(i))}{\sum_{j=1}^{N} p(\boldsymbol{y}_t|\boldsymbol{x}_t^*(j))} \tag{7}$$

for $i = 1, \ldots, N$ and $q_0 = 0$. Lastly, the posterior particles $\{\boldsymbol{x}_t(i) : i = 1, \ldots, N\}$ are obtained by sampling from $\{\boldsymbol{x}_t^*(i) : i = 1, \ldots, N\}$ such that $P(\boldsymbol{x}_t(j) = \boldsymbol{x}_t^*(i)) = q_i, \forall j$. This is achieved by generating N uniform distributed (on $(0, 1]$) random variables w_i's and assigning $\boldsymbol{x}_t(i) = x_t^*(M)$ where M satisfies

$$\sum_{j=0}^{M-1} q_j < w_j \leq \sum_{j=0}^{M} q_j \tag{8}$$

2.2 Proposed method

In this section, we first present a global optimal affine registration between two point sets by viewing the registration as a state estimation process and solved by particle filtering. In addition, the resulting moving point set is further decomposed if it does not meet certain accuracy criteria, and registration is performed at finer scale to correct the local deformation. Such decomposition-registration cycle continues until satisfying result is achieved. Finally, under a Lie-group/Lie-algebra framework, a diffeomorphic transformation is constructed as the final result.

2.2.1 Affine Registration using PF

Denote the fixed point set as $\mathscr{P} = \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_M\}$ and the moving point set, which is to be registered to the fixed point set, as $\mathscr{Q} = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N\}$. It is noted that $\boldsymbol{q}_i, \boldsymbol{p}_i \in \mathbb{R}^3 \times \{1\}$ are represented in homogeneous coordinates. To register $\mathscr{Q}$ to $\mathscr{P}$, a global optimal affine registration procedure is carried out. For this purpose, we define the cost function:

$$C(\boldsymbol{B}) := \frac{1}{N} \sum_{i=1}^{N} \|\boldsymbol{B}\boldsymbol{q}_i - Cl(\boldsymbol{B}\boldsymbol{q}_i)\| + \frac{\lambda}{\det^2(\boldsymbol{B}(1:3, 1:3))}. \tag{9}$$

In equation (9), we have the transformation matrix

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{t} \\ \boldsymbol{0}_{1\times 3} & 1 \end{pmatrix} \tag{10}$$

where $\boldsymbol{A} \in \mathbb{R}^{3\times 3}$, $\det(\boldsymbol{A}) \neq 0$ is the affine transformation matrix, and $\boldsymbol{t} \in \mathbb{R}^3$ is the translation vector. In addition, $Cl : \mathbb{R}^3 \to \mathscr{P}$ maps a point in $\mathbb{R}^3$ to its closest point in $\mathscr{P}$. Note that the second term in equation (9), with weighting $\lambda > 0$, penalizes $\det(\boldsymbol{A})$ from getting too close to zero. Without this, the cost function may have trivial minimizers.

Concerning numerical details, a KD-tree data structure may be utilized to achieve a fast ($O(\log M)$) search [29]. The minimization of the registration cost function E in (9) is a 12 dimensional nonlinear unconstrained optimization problem. The gradient of E with respect to the affine matrix $\boldsymbol{A}$ and translation $\boldsymbol{t}$ are computed. Then the BFGS algorithm, one of the most popular quasi-Newton methods [41], is employed to obtain a fast (super-linear) convergence.

Nevertheless, the above affine ICP is a local minimization process and is known to be sensitive to initialization and local minima. To overcome such drawbacks and achieve global optimality, we formulate the affine registration as a parameter estimation task, and solve it using particle filters.

In affine registration, the state space $\mathscr{S}$ is 12 dimensional where the first 9 dimensions are for the affine matrix and the last 3 dimensions are for translation. Denoting the state vector as $\boldsymbol{x} \in \mathscr{S}$, the process model takes the form:

$$\boldsymbol{x}_{t+1} = \mathcal{R}(\boldsymbol{x}_t + \boldsymbol{u}), \tag{11}$$

where the operator $\mathcal{R} : \mathscr{S} \to \mathscr{S}$ takes $\boldsymbol{x}_t$ as the initial configuration, proceeds with a few steps of the deterministic affine registration, and returns the resulting parameter estimated as $\boldsymbol{x}_{t+1}$.

The observation model is:

$$\boldsymbol{y}_t = \mathcal{E}(\boldsymbol{x}_t) + \boldsymbol{v} \tag{12}$$

where the operator $\mathcal{E} : \mathscr{S} \to \mathbb{R}^+$ gives the cost function under the state $\boldsymbol{x}_t$. The process and observation noise signals are denoted by $\boldsymbol{u}$ and $\boldsymbol{v}$, respectively. It can be seen from above that both the process and the observation models are highly non-linear.

In addition to the process and observation models, the prior distribution $p(\boldsymbol{x}_0)$ is required. Usually, without *a priori* information of the solution, $p(\boldsymbol{x}_0)$ is assumed to be Gaussian or uniformly distributed. Altogether, the complete global optimal affine ICP can be described in Algorithm 1. This algorithm will be referred to as affine-PF-ICP in the subsequent discussion.

Algorithm 1 Affine Registration by Particle Filtering

```
Sample from $p(\boldsymbol{x}_0)$ to get $\{\boldsymbol{x}_0(i) : i = 1, \ldots, N\}$
for $i = 1, 2, ..., t$ do
    Obtain the prior samples $\{\boldsymbol{x}_t^*(i)\}$ by the process model.
    Evaluate the likelihood $q_i$'s using (7)
    Resample to get posterior $\{\boldsymbol{x}_t(i)\}$ using (8)
end for
```

2.2.2 Transformation decomposition

The global affine registration presented in the previous section yields an affine transformation $\tilde{\boldsymbol{B}} \in \mathbb{R}^{4\times4}$. Although such affine registration is optimal in the global sense, and is capable of handling the cases where two point sets differ by even an 180° rotation (around one or more axes), it is not able to provide satisfying results where the local deformations exist. To measure the residual error at each point, define

$$\boldsymbol{e}_i := \boldsymbol{p}_j - \tilde{\boldsymbol{B}}\boldsymbol{q}_i = Cl(\tilde{\boldsymbol{B}}\boldsymbol{q}_i) - \tilde{\boldsymbol{B}}\boldsymbol{q}_i \tag{13}$$

to be the error vector at each registered point, where $\boldsymbol{q}_j$ is the closest point in $\mathscr{P}$ to $\tilde{\boldsymbol{B}}\boldsymbol{q}_i$. Accordingly, we define the average error magnitude:

$$E := \frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{e}_i\|^2, \tag{14}$$

which is consistent with the cost function in equation (9) but without the penalty term.

Such a quantity is used to determine if the registration requirement is satisfied. If not, the moving point set is decomposed in a way described below. To proceed, each point in the registered moving point set is augmented using its error vector defined above to form a 8-dimensional feature vector:

$$\boldsymbol{f}_i := [\eta \boldsymbol{q}_i; (1-\eta)\boldsymbol{e}_i] \in \mathbb{R}^8 \tag{15}$$

It is noted that since the 4-th component of $\boldsymbol{q}_i$ and $\boldsymbol{e}_i$ are respectively 1 and 0, $\boldsymbol{f}_i$ is effectively in $\mathbb{R}^6$. Consequently, the feature vector combines the original space information of the point set as well as the registration error. In practice, η is chosen such that $\eta \boldsymbol{q}_i$ and $(1-\eta)\boldsymbol{e}_i$ are of the same magnitude. In addition, a (dis-)similarity matrix $W \in \mathbb{R}^{N\times N}$ is defined as $W(i,j) := \exp(-\|\boldsymbol{f}_i - \boldsymbol{f}_j\|^2)$.

It is noted that the construction of the matrix W does not take $O(N^2)$ temporally or spatially. In fact, in the feature space a KD-tree is constructed. Consequently, for each feature vector, only its neighbors in a certain range are selected to construct the W matrix in order to make it sparse. To decompose the point set into two groups such that the "cross-group similarity" is minimized, the normalized cut algorithm [55] is adopted. To this end, we further define the diagonal matrix D as $D_{i,i} = \sum_j W(i,j)$ and an eigen-decomposition problem:

$$D^{-\frac{1}{2}}(D-W)D^{-\frac{1}{2}}\boldsymbol{x} = \lambda \boldsymbol{x} \tag{16}$$

is solved using the Lanczos algorithm [18]. Since the left-hand-side matrix in equation (16) is symmetric positive definite, all the eigen values are positive real numbers. Among them, the eigenvector corresponding to the second smallest eigenvalue is picked. A standard k-means clustering or simply thresholding will decompose $\mathscr{Q}$ into two subsets, denoted as U_1 and U_2.

With the point set decomposed, U_1 and U_2 are registered separately to $\mathscr{P}$. In particular, we note that the cost function in equation (9) is non-symmetric between $\mathscr{P}$ and $\mathscr{Q}$: it only traverses the moving point set. This design naturally provides the capability of registering a partial structure to an entire one, which is the property we expect when registering U_1 and U_2 to $\mathscr{P}$. Subsequently, their registration results are evaluated to determine if further decomposition of U_1 and/or U_2 is necessary. This process goes on and finally, the decomposition-registration cycle stops when the criteria is met in all the point subsets. As shown in Theorem 1, convergence is guaranteed.

Theorem 1 (Convergence). *Given any positive error criteria $\epsilon > 0$, the above hierarchical scheme will find the decomposition $\{U_i\}$ of $\mathscr{Q}$ and the corresponding transformations $\{T_i\}$ such that*

$$\frac{1}{|\mathscr{Q}|} \sum_{i} \sum_{q \in U_i} \|Cl(T_i q) - T_i q\|^2 < \epsilon \tag{17}$$

It is noted that the above theorem states the ideal case. In practice, however, the existence of the second term in equation (9) will make the theorem work approximately. Finally, The procedure in this section can be summarized as follows:

It may be argued that the nonlinearity such hierarchical decomposition can achieve may be not as accurate as those methods using splines, NURBS, or even point-wise deformation fields. However, at least theoretically the decomposition can be performed to point-wise accuracy. Moreover, we observe that in many situations, the data set can be grouped into a few components, where an single affine transformation is descriptive enough for each group, i.e., point-wise accuracy is not always necessary. For instance, in the example shown in Section 2.3.2, an affine transformation is enough

Algorithm 2 Registration & Decomposition($\mathscr{P}$, $\mathscr{Q}$)

```
Require: Fixed Point set $\mathscr{P}$, moving point set $\mathscr{Q}$
Register $\mathscr{Q}$ to $\mathscr{P}$ as in Algorithm 1
if Error is above criteria then
    Decompose $\mathscr{Q}$ into $\mathscr{U}_1 \cup \mathscr{U}_2$
    for $i = \{1, 2\}$ do
        Registration & Decomposition($\mathscr{P}$, $\mathscr{U}_i$)
    end for
else
    return point subsets of $\mathscr{Q}$ and their registration parameters.
end if
```

for each section of finger between two joints. Therefore, the proposed method provides a nice balance between nonlinearity and global optimality.

2.2.3 Fusion of affine transformations

In the previous section, the moving point set is decomposed and registered separately to the fixed point set. Although doing so more accurately addresses the local deformation and reduces the registration error at different scale, it generates more than two affine transformations, each of which applying to certain portion of the moving point set. On the other hand, in the final result, what one expects is a coherent transformation applied on the whole moving point set. Hence we need to “stitch” the pieces together.

More generally, any transformation is locally approximated to the first order by an affine transformation. However, in practice it is not only computationally intensive to calculate the local transformation at each point, but also not necessary because in many cases nearby points undergo similar affine transformation. Therefore, the domain can be decomposed into a few components, each of which is transformed by an affine transformation, and the transformation of the whole domain is the combination of them. However, doing so poses the problem of how to construct the global transformation from each of those components.

More rigorously, in the 3D Euclidean space $\mathbb{R}^3$, there are N affine transformations parametrized by: (A_i, T_i) with $A_i \in \mathbb{R}^{3\times3}, i = 1, \ldots, N$ and $T_i \in \mathbb{R}^{3\times1}, i = 1, \ldots, N$. In addition, each affine transformation corresponds to a weighting function $\omega_i : \mathbb{R}^3 \to [0, 1], i = 1, \ldots, N$, where high $\omega_i(\boldsymbol{x})$ indicates the i-th transformation (A_i, T_i) plays a dominant role in the region closed to $\boldsymbol{x}$. To fuse (A_i, T_i)'s into a single (nonlinear) transformation A while respecting their corresponding ω_i's, it was shown in [2] that a linear combination in the Euclidean space given by $F(\boldsymbol{x}) := \sum_i \omega_i(\boldsymbol{x})(A_i\boldsymbol{x} + T_i)$ will develop singularities.

To address this issue, form the affine transformations a Lie group, and the "infinitesimal" transformations associated with them defines the corresponding Lie algebra [2]. Consequently, instead of performing the above straightforward weighted combination of the affine transformations in the Lie group, one seeks the alternative in the Lie algebra, and then the weighted average is mapped back as the final transformation. It is shown in [2] that such a transformation is a diffeomorphism. More explicitly, we assume the hierarchical decomposition registration scheme results in a list of point subsets: $\{U_i \subset \mathscr{Q} : i = 1, \ldots, K\}$ of the moving point set $\mathscr{Q}$, and their corresponding affine transformations: $\{B_i \in \mathbb{R}^{4\times4} : i = 1, \ldots, K\}$. To fuse them, a weighting function $\omega_i : \mathbb{R}^3 \to [0, 1]$ is designed for each transformation B_i (and the corresponding U_i) as:

$$\omega_i(\boldsymbol{x}) = e^{-\|\boldsymbol{x} - Cl_i(\boldsymbol{x})\|^2}, \tag{18}$$

where the function $Cl_i : \mathbb{R} \to U_i$ returns the closest point in the point set U_i to a given position in $\mathbb{R}^3$. It can be seen that ω_i indicates each transformation B_i's spatial extension: the weight of the transformation B_i is high around the point set U_i and is low at where far from U_i. Referring to the above discussion, the transformation matrix B_i is in the Lie group (whose direct weighted sum, $\sum_i \omega_i B_i$, develops singularities). Next, the infinitesimal motion L_i in the Lie algebra is computed by the matrix logarithm [2]. Then, the weighting function $\omega(\cdot)$ is evaluated at $\boldsymbol{x}$ and the weighted average infinitesimal transformation in the Lie algebra is obtained as:

$$L(\boldsymbol{x}) := \sum_i \omega_i(\boldsymbol{x}) L_i. \tag{19}$$

Finally, the matrix exponential $B(\boldsymbol{x})$ of $L(\boldsymbol{x})$ is computed and the image of $\boldsymbol{x}$ is $B(\boldsymbol{x})\cdot\boldsymbol{x}$. As shown in [2], the transformation $B(\boldsymbol{x}) : \boldsymbol{x} \in \mathbb{R}^3 \times \{1\}$ is a diffeomorphism.

2.3 Experiments

In this section, we provide various qualitative and quantitative examples to show the performance and robustness of the proposed method.

2.3.1 Affine ICP using particle filtering

To demonstrate the particle filtering affine ICP in Section 2.2.1 successfully overcomes the local minima problem and achieves global optimality, we use an example where the ground truth is known *a priori*. In Figure 1, the blue point set is transformed to another affine pose as shown by the red one, and the objective is to register the red point set to the blue one.

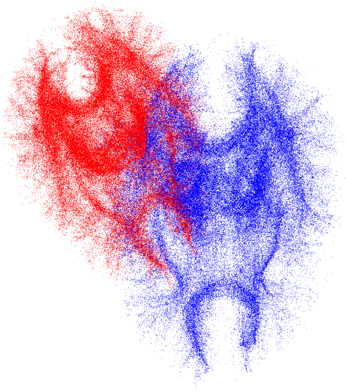


Figure 1: One point set in two different affine poses

For this purpose, minimizing the affine-ICP cost function defined in equation (9) gives the result shown in Figure 2. The blue points are again from the fixed point set while the red ones are from the registered moving point set. Due to the local minima, the poses of the two point sets are not well registered. On the other hand,

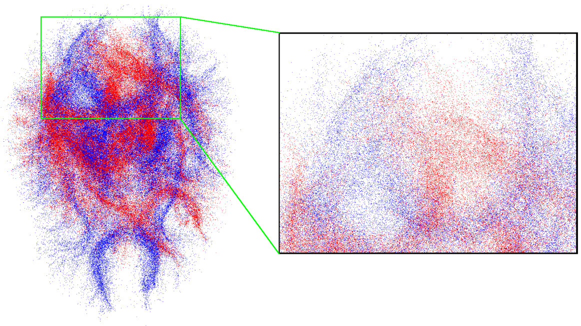


Figure 2: Registration using affine ICP

the affine-PF-ICP correctly restores the original pose as shown in Figure 3.

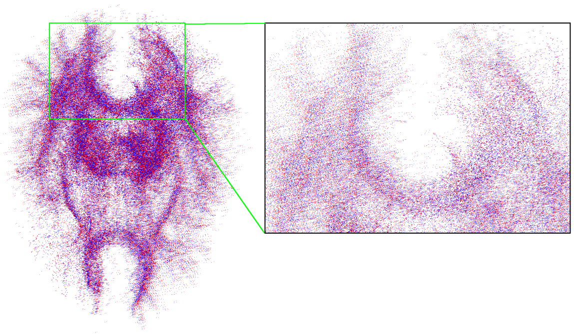


Figure 3: Registration using Particle Filtering

To further analyze the performance of the affine-PF-ICP quantitatively, the following experiments were conducted: The blue point set in Figure 1 is again fixed, and is also employed as the moving point set, i.e., it is to be registered to itself. The starting affine parameters (12 dimensional) are randomly chosen. Hence, the ground truth of a good registration is that the final affine matrix converges to the identity matrix, and that the translation vector converges to a zero vector. Written in a 12-dimensional parameter vector, this ground truth parameter is $(1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0)$. To initialize the registration, a random initial affine matrix is constructed by adding to the identity matrix a random matrix, each of whose elements is normally distributed around 0 with standard deviation of $\sigma_A \in \{0.05i;\ i = 1, \dots, 6\}$ with i being the trial index. Similarly, the initial translation is a 3D normal random vector centered at $\mathbf{0}$ with standard deviation $\sigma_T \in \{0.25i \times \text{range};\ i = 1, \dots, 6\}$, where the "range" is the range of the fixed point set. In the statistical analysis, 40 sets of initial parameters are generated for each trial. The l^2 distance between the initial parameters and the ground truth parameter, as well as the l^2 distances between the final parameters and the ground truth parameter are recorded. These are shown in Figure 4 in semi-log plot for the vertical axis.

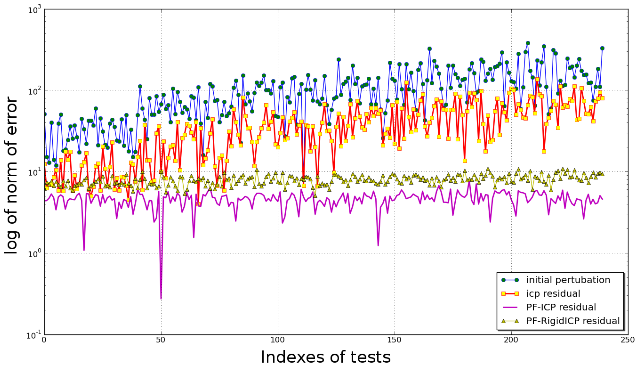


Figure 4: Error comparison in semi-log-y plot. Top curve: l^2 distances from the starting parameters to the ground truth. Second curve: l^2 distance from the final parameters of the affine-ICP to the ground truth. Third curve: l^2 distance from the final parameters of the rigid-PF-ICP registration, as proposed in [54], to the ground truth. Bottom curve: l^2 distance from the final parameters of the proposed affine-PF-ICP to the ground truth.

2.3.2 A walk-through example

After illustrating the (linear) affine-PF-ICP, we use this example to demonstrate the whole process of the (nonlinear) hierarchical diffeomorphic registration algorithm. Figure 5 shows two sets of points. (For better appreciation of their relative positions, we also show the surface. However, **only** the points are used in the algorithm.) The fixed point set is colored in cyan and it is extracted from the public available data set #72 on `shapes.aimatshape.net`.[1] The moving point set, in yellow, is number #759[2]. Both point sets are the index and middle fingers from different left hands. It can also be observed that apart from the affine pose difference between the two point sets, there are local deformations, e.g., the fingers of #72 are bent, while those of #759 are straighter.

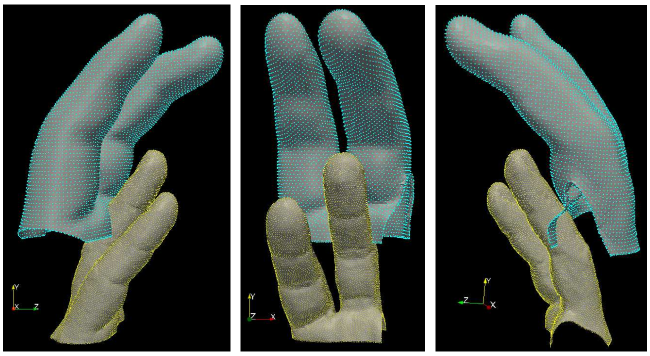

(a) (b) (c)

Figure 5: Two finger point sets in their original poses.

The result of applying the affine-PF-ICP is shown in Figure 6 where the fixed point set is in cyan and the registered moving point set is in yellow. It can be observed

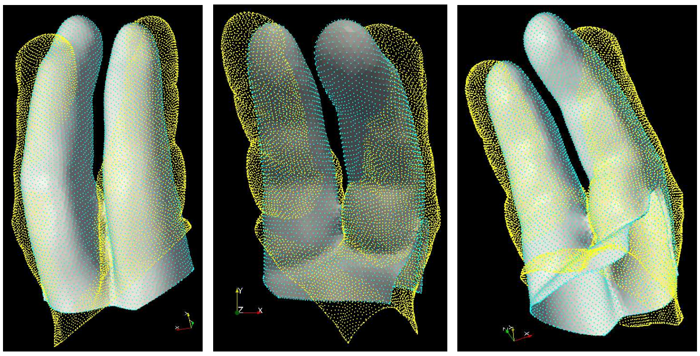

(a) (b) (c)

Figure 6: Finger point set after affine-PF-ICP registration

that a single affine transformation, though being globally optimal, is not enough to

---

[1] http://shapes.aimatshape.net/view.php?id=72

[2] http://shapes.aimatshape.net/viewgroup.php?id=759

align the two point sets. In fact, visual inspection reveals that the angles between the two fingers in the two point sets are different, as well as there are finer deformations. Applying the decomposition scheme in Section 2.2.2, the registered moving point set is separated into two, as shown in Figure 7.

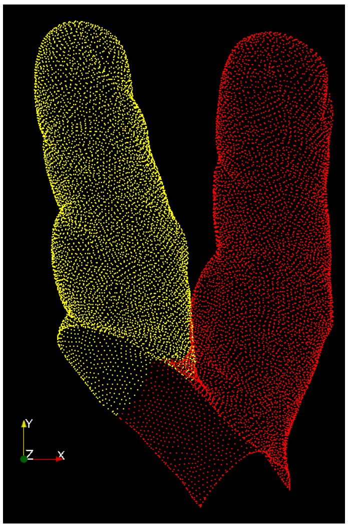
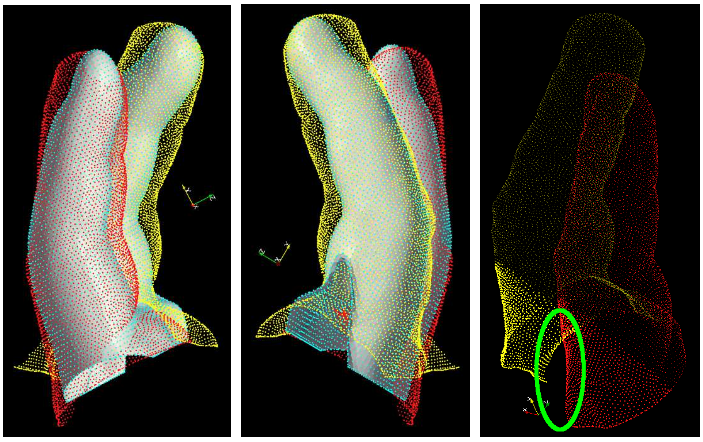


Figure 7: Decompose the moving point set into two.

Each point subset is then registered to the fixed point set separately. Due to the asymmetry in the registration cost function defined in equation (9), registering a partial structure to the whole structure is naturally handled. Indeed, comparing Figure 8(a) to Figure 6, we see the angular difference problem is resolved. However, not surprisingly it also results in tearing, which is highlighted in Figure 8(b) and 8(c).

The moving point set is further decomposed hierarchically. After registration at a finer scale, the computed transformations are fused into a single diffeomorphic registration, shown in Figure 9.

As can be observed in enclosed region highlighted by the green circle of Figure 9(d) and 9(e), the multiple transformations are nicely “stitched” together at the regions highlighted by the green ellipse, and the sub-point sets are nicely registered to the



(a) (b) (c)

Figure 8: Register the two sub-sets to the fixed point set separately. Comparing subplot 8(a) and 8(b) with Figure 6(a) and Figure 6(c) will see the angle difference being solved. However, subplot 8(c) highlights the tearing in region enclosed by the circle brought about by registering separately.

fixed point set (Figure 9(a)9(b)9(c)). Interestingly, in this example the decomposition is consistent with the anatomical structure of the fingers. However, in general the purpose of decomposition is for finer scale registration and does not necessarily correspond to anatomical decomposition.

Quantitatively, we plot the reduction of the registration error in Figure 10. The vertical axis is the error defined in equation (14) (The fixed point set has the range in: $[-0.26, 0.13] \times [0.26, 0.85] \times [-0.24, 0.31]$.) In the figure the dashed line with the downward triangle marker plots the log of error of just affine-PF-ICP without decomposition. It can be seen that the error stops decreasing after a certain number of iterations. However, with the decomposition, the error could be further reduced by registration at finer level, shown as the “square” line. After the second decomposition, the error further drops, as indicated by the “upper-triangle” line, and the “star” line after the third decomposition. Furthermore, it is noted that when there are several components, the largest error is plotted. Although the current result after three decomposition is satisfying, further decomposition can be carried on if desired.

2.3.3 Tract registration

In this set of experiments we further test the algorithm on larger and more complex point sets. Here each point set represents the neuro-tracts of the human white matter. Different functional regions in the cerebral cortex (gray matter) are connected by those tracts. Moreover, clinical correlations have been found between the shape of the tracts and certain brain diseases such as schizophrenia [32]. Hence, under the scenario of statistical analysis of fiber tracts of different subjects, it is crucial to register them into a common coordinate system. Furthermore, the tracts can be categorized into different groups, each of which contains tracts connecting similar functional regions of the gray matter. However, on the one hand it is very time consuming for a neuroscientist to hand group the tracts, and on the other hand the fully automatic tract classification is a challenging problem in its own right. Thus, in this experiment one set of fiber tracts has been grouped into several groups manually and is used as the registration template, to which all the others are registered to and clustered accordingly.

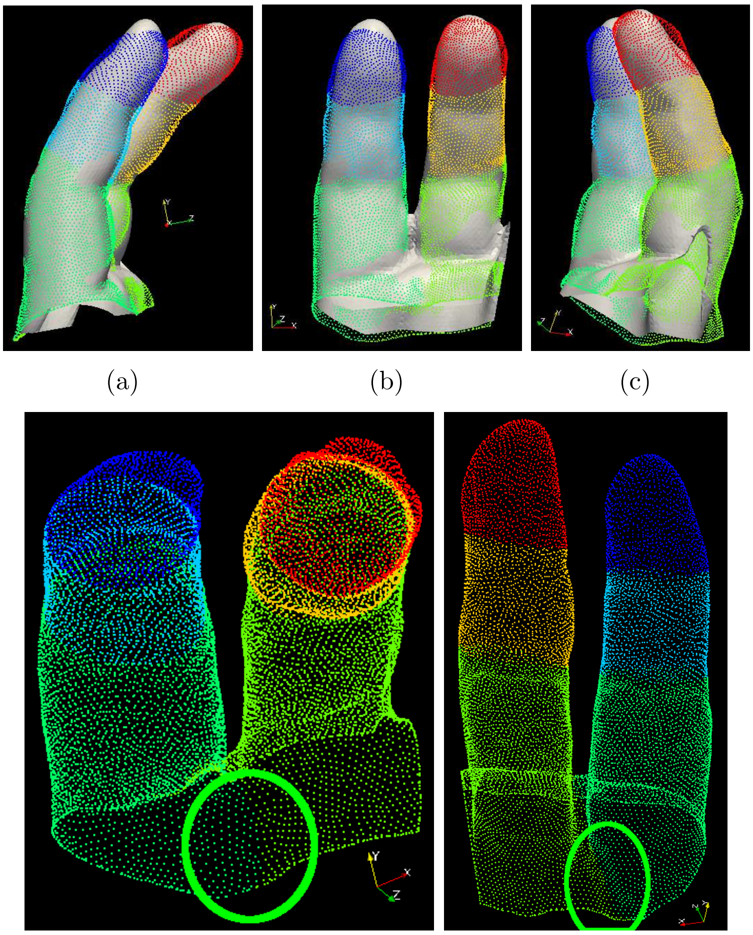
Figure 11 shows the original pose and position of the clustered template fiber tracts (point set), and one moving point set (un-clustered). It can be seen that in addition to the size, the two point sets are flipped, which makes the previous local-optimization based algorithms not well-suited to the registration task at hand.

Figures 12(a), 12(c), and 12(e) show the results of registering the moving point set to the template, using the affine-PF-ICP method. Since this figure only shows the performance of the registration, for clarity of display, the template point set is plotted in cyan. The registered moving point set is in yellow. In can be seen that although the affine transformation is global optimal, the fine part (noted in the box) is not well captured. Utilizing the fact that the template point set has been clustered into several anatomically significant groups, we are able to further refine the registration by decomposing the moving point set. In fact, the above global optimal affine registration, though being not accurate at the fine scale, does provide a reasonable initial result. Based on that, the moving point set is clustered into groups using the closest point criterion. That is, for each point in the affine-PF-ICP registered moving point set, its class is decided by that of the closest point in the template point set. Therefore, each moving point subset is registered to its corresponding template sub-point set by the method described in Section 2.2.1.

In this test, the template point set has 908894 points and has been clustered into 10 tract bundles, each of which is registered to its corresponding bundle in the fixed point set. Based on the new registration result, the moving point set is re-grouped to achieve a better result. The grouping/registration cycles are iterated and finally, the 10 paired point subset registration results are fused to form a global diffeomorphic transformation, which is shown in Figures 12(b), 12(d), and 12(f). In particular, note that those regions not having been well registered previously are handled quite nicely.

Furthermore, in this experiment the fixed point set has been clustered. Thus the registration scheme simultaneously provides a clustering for the moving point set. Using different colors for different clusters, they are shown in Figure 13.

The method is further tested on more tracts data sets. Each point set has a similar initial position and pose configuration as detailed above. In each of the Figures 14, 15, and 16, the registration results are overlapped onto the same fixed point set on the right side and the clustering outcomes of the moving data sets are shown on the right. Moreover, we perform the quantitative analysis on the fiber tracts data sets using the proposed method, affine-PF-ICP, and the method in [54]. For each final result, the error (computed using equation (14)) is plotted in Figure 17. The fixed point set has the range: $[-70.3, 66.7] \times [-87, 91.8] \times [-66.673]$.



(a) (b) (c)

(d) (e)

Figure 9: Polyaffine registration result. Different colors in subplots 9(a), 9(b) and 9(c) indicate different groups formed by the hierarchical decomposition. Compare them with Figure 8(a) and 8(b) to find the local deformations get well registered. Also compare subplot 9(d) and 9(e), highlighted in the circle, with Figure 8(c) to see the "smooth stitching".

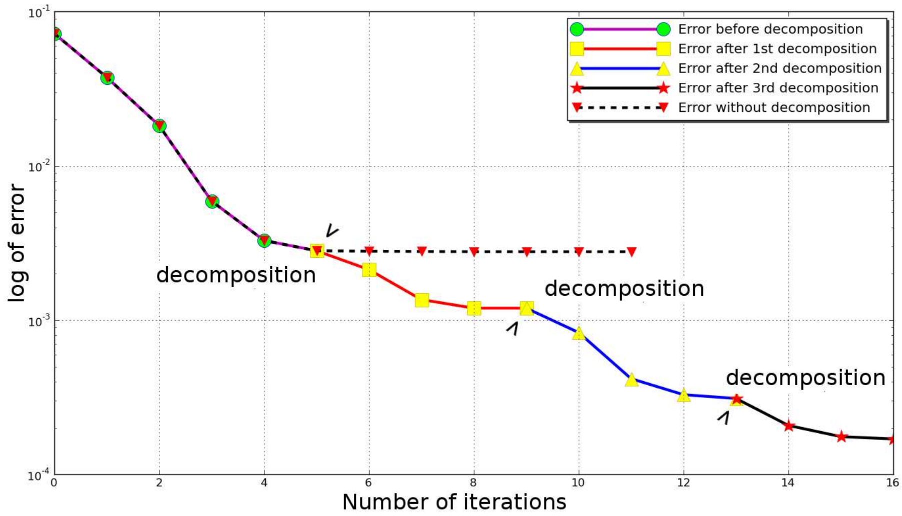
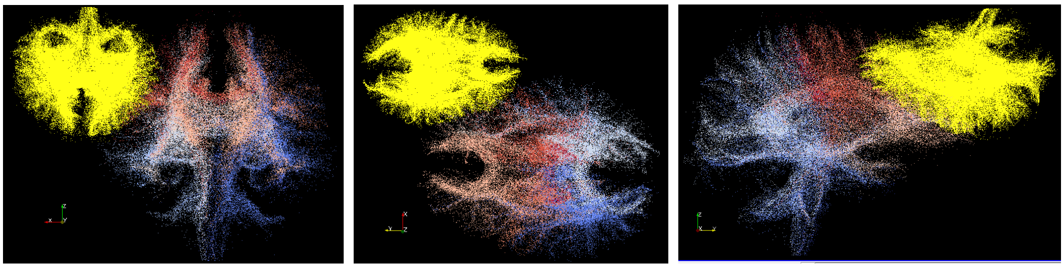
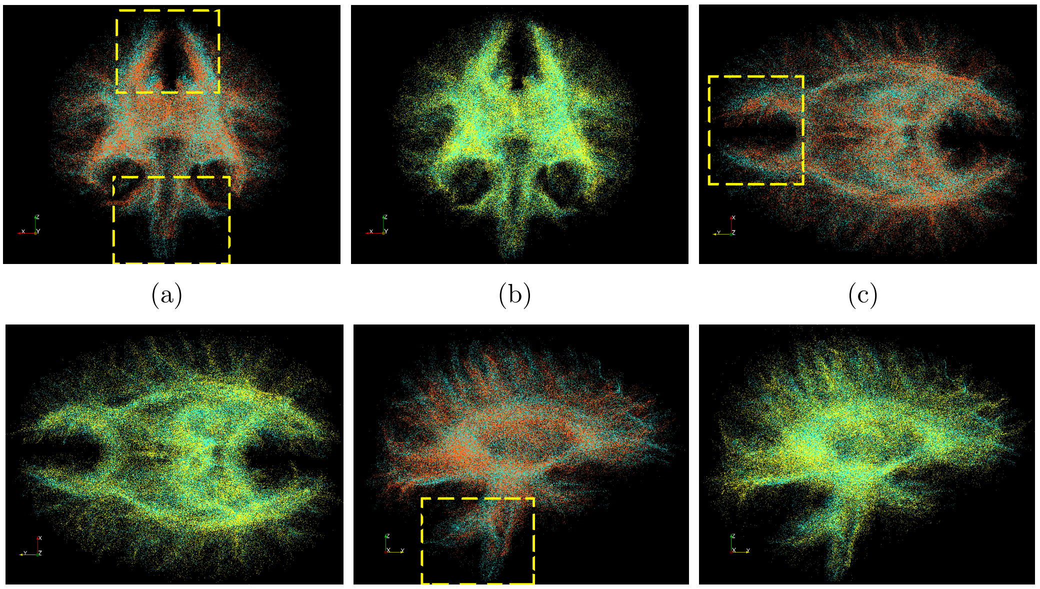


Figure 10: Error reduction through the hierarchical decomposition-registration. Dash line indicates error without decomposition (single affine-PF-ICP) so it saturated after certain number of iterations. With decomposition, indicated by little arrow, the error is reduced.



(a) Anterior view (b) Superior view (c) Right view

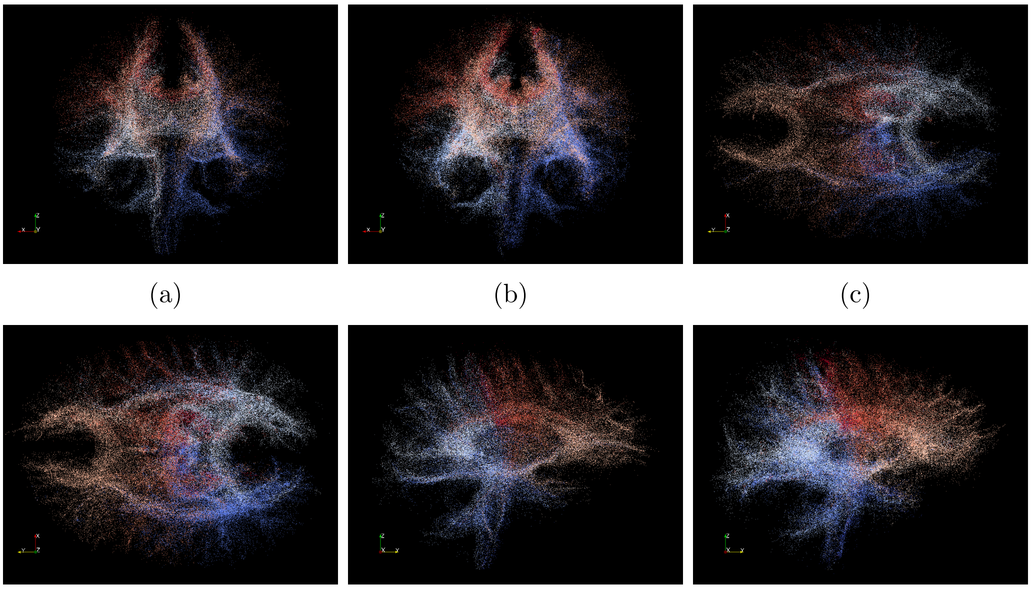
Figure 11: Clustered template point set (colored) and un-clustered moving point set (yellow), in their original pose and position.



(a) (b) (c)

(d) (e) (f)

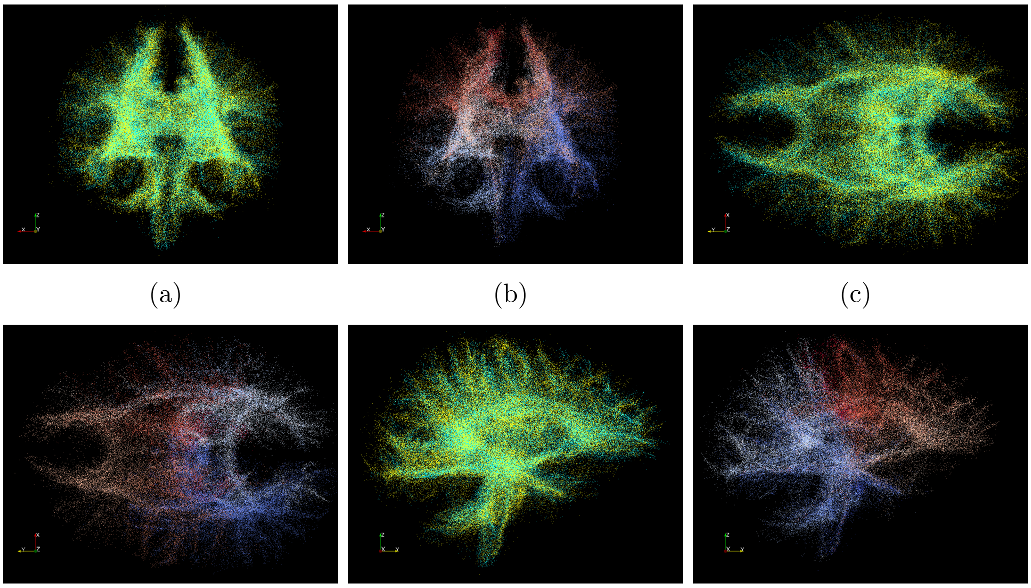
Figure 12: In all sub-plots, the template point set is in cyan. Left column: affine-PF-ICP registered moving point set (orange). See the mis-matching in the box. Right column: hierarchical decomposition-registration result (yellow). Moving point set has 1528792 points. Running time 30min.



(a) (b) (c)

(d) (e) (f)

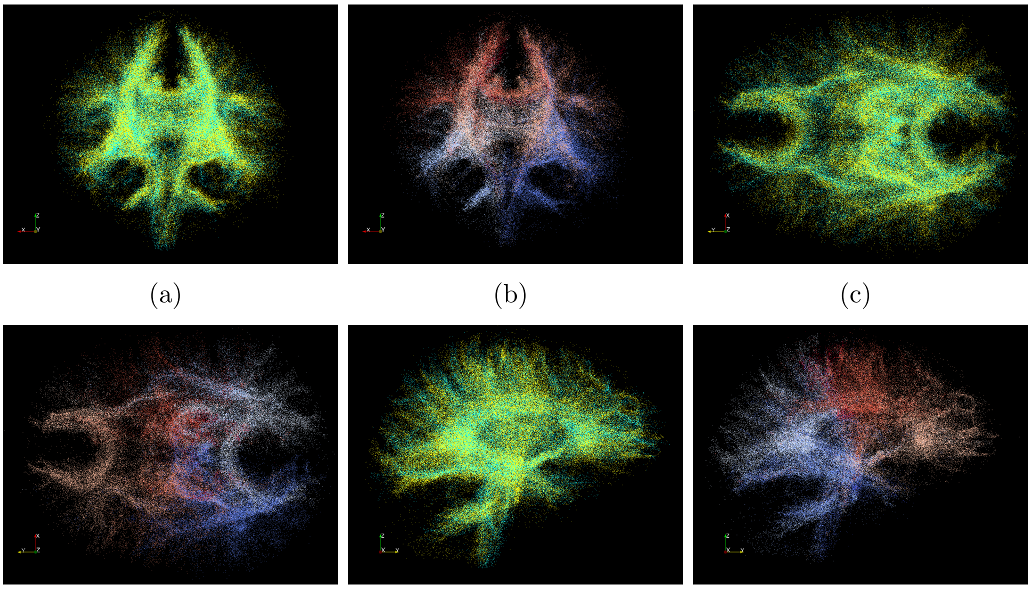
Figure 13: Left column: colors indicate clusters in template point set. Right column: moving point set got clustered while registered.



(a) (b) (c)

(d) (e) (f)

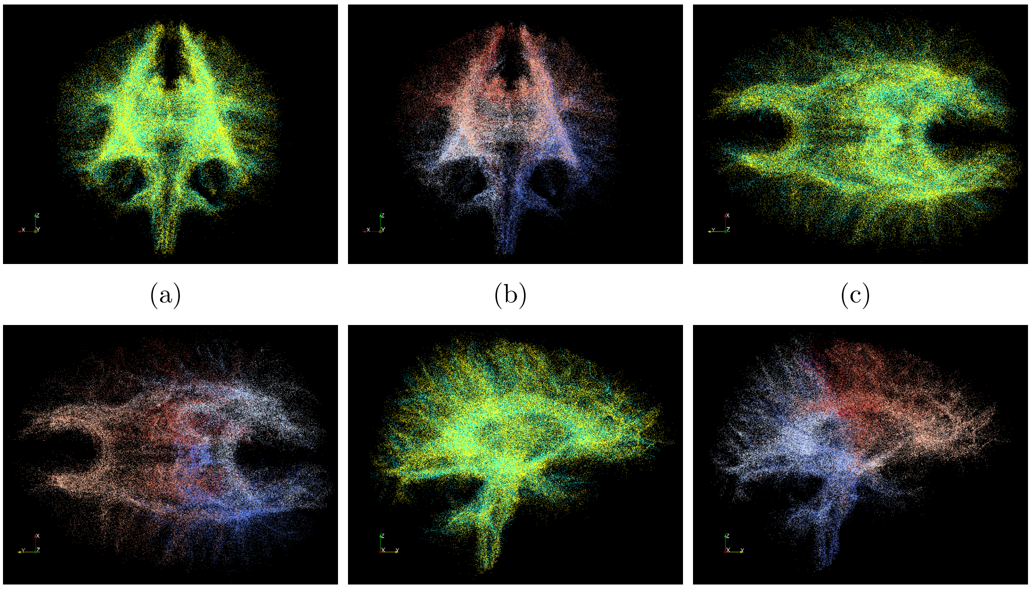
Figure 14: Left column: the registration result. The fixed point set is in cyan and the registered moving point set is in yellow color. Right column: registering and clustering results for the second moving point set. Moving point set has 1176057 points.



(a) (b) (c)

(d) (e) (f)

Figure 15: Left column: the registration result. The fixed point set is in cyan and the registered moving point set is in yellow color. Right column: registering and clustering results for the third moving point set. Moving point set has 1376690 points.


(a) (b) (c)

(d) (e) (f)

Figure 16: Left column: the registration result. The fixed point set is in cyan and the registered moving point set is in yellow color. Right column: registering and clustering results for the fourth moving point set. Moving point set has 1614770 points.

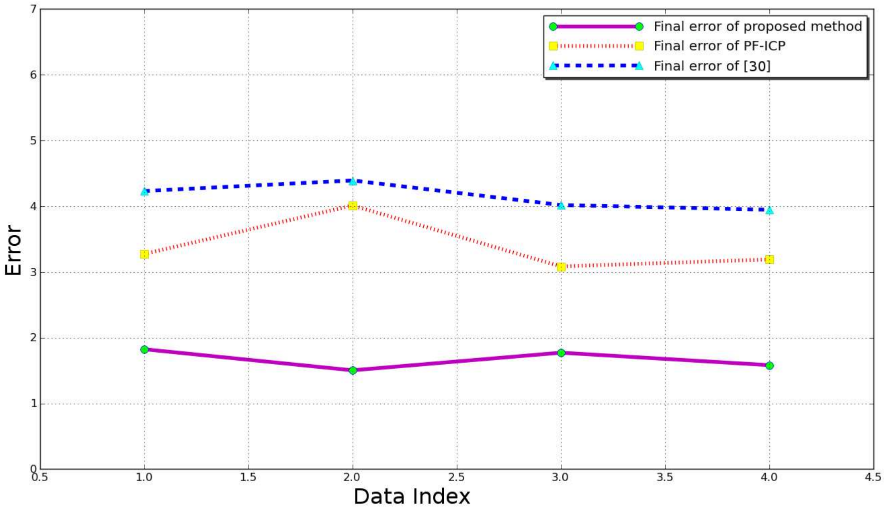


Figure 17: Comparing the final error on the four tracts data using different methods.

CHAPTER III

MULTI-OBJECT SEGMENTATION USING LOCAL ROBUST STATISTICS AND CONTOUR INTERACTION

With the final objective of combining registration and segmentation in a coupled framework of shape-based segmentation method, in this chapter, we turn to the topic of segmentation. In particular, we propose a semi-automatic image segmentation method to extract multiple object from a given image. To this end, the algorithm starts with user providing several "seeds" in different target objects. Then, the image features are extracted and learned according to those features around the seeds. Such learned information drives the segmentation contour evolution and the multiple contours interact with each other to guarantee the mutual exclusive condition. Finally, the contours converge to equilibrium position and provide the results. In Section 3.1, we present the details of the overall method, including target characteristics adaptive learning, individual contour evolution, and contour interaction. Next, we show some experiments and tests performed on CT and MR data in Section 3.2.

3.1 Proposed Method

If we consider the segmentation process in our own visual system, we observe that when human is recognizing the objects in a scene, several basic steps take place in sequence [44]. We will illustrate this via an example. Suppose that we want to trace out the boundary of both the liver and the right kidney in medical imagery. First, prior anatomy knowledge drives our attention to the right abdominal region. Second, we focus at an area where we believe to be most "liver-like," and learn the liver characteristics in this particular image. With such knowledge, we then move our focus to enclose more tissue that looks similar to those representative regions. Usually, such similarity ends when we reach a remote area. In particular, at the boundary where the liver touches the right kidney, the decision is difficult. Under such a situation, we apply a similar procedure to the kidney, and we come back to the same ambiguous region. However, this time with the information from both sides (liver and kidney), internally we perform a competition: we compare the current voxel with both the liver and the kidney to decide which boundary should advance, so the other should retreat. Finally, the boundaries of liver and kidney are placed at the balanced locations of the competition.

The segmentation scheme presented in this paper is a mathematical model for the above process. It is a semi-automatic method because the first step above is achieved by the user providing a label map indicating different targets by different labels. Each subsequent step is handled by an automatic algorithm and is detailed in what follows below.

3.1.1 Online feature learning

Denote the image to be segmented as $I : \Omega \to \mathbb{R}$ where $\Omega \subset \mathbb{R}^d$ is an open set and $d \in \{2, 3\}$. Likewise, the user provided label map is denoted as $L : \Omega \to \mathbb{N} \cup \{0\}$ where 0 indicates background and non-zero positive integers indicate the target object labels. For ease of discussion, in this paper, we assume the distinct labels to be consecutively ranging from 0 to N, an arbitrary positive integer. Moreover, the labeled region can be defined by several “clicks”, and does not have to be close to the desired boundary. Next, voxels with the non-zero labels are categorized into different “seed groups” as $G_i = \{\boldsymbol{x} \in \Omega : L(\boldsymbol{x}) = i\}$.

In order to fully utilize the information given by the label map, we note that the seed group not only indicates the location of the target, but also provides some sample voxels contained in it. Hence, instead of making general assumptions on the target characteristics such as brighter/darker than surrounding area, we can learn them in an online fashion. Often times, the image intensity alone is not descriptive enough. Hence, a feature vector is extracted at each voxel, forming a feature image $\boldsymbol{f} : \Omega \to \mathbb{R}^{D_f}$. Subsequently, the segmentation is performed in the feature space. There are many choices for the feature vector such as wavelet coefficients, Fourier descriptors, Hessian matrix, etc. In this paper, we choose local robust statistics [24, 47] because they are not sensitive to image noise, and may be computed quickly.

To this end, for each voxel $\boldsymbol{x}$ in the image, we define the feature vector $\boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^{D_f}$ by combining several robust statistics derived in a neighborhood $B(\boldsymbol{x}) \subset \Omega$ around $\boldsymbol{x}$. More explicitly, we denote $MED(\boldsymbol{x})$ as the intensity median within $B(\boldsymbol{x})$. In addition, the local intensity range is also an important characteristic, but is sensitive the noise. To address this issue, the distance between the first and third quartiles, namely the inter-quartile range ($IRQ(\boldsymbol{x})$), is calculated as the second feature. Furthermore, the local intensity variance is a good candidate but again it is sensitive to outliers. In contrast, the median absolute deviation (MAD) is much more robust and is computed as $MAD(\boldsymbol{x}) := median_{\boldsymbol{y} \in B(\boldsymbol{x})}(I(\boldsymbol{y}) - MED(\boldsymbol{x}))$. Consequently, we define the feature vector $\boldsymbol{f}(\boldsymbol{x})$ as:

$$\boldsymbol{f}(\boldsymbol{x}) = (MED(\boldsymbol{x}), IRQ(\boldsymbol{x}), MAD(\boldsymbol{x}))^T \in \mathbb{R}^3 \tag{20}$$

With the space of feature vectors thus defined, seed groups are now characterized by the probability density function of the feature vectors estimated by:

$$p_i(\boldsymbol{f}) = \frac{1}{|G_i|} \sum_{\boldsymbol{x} \in G_i} K_\eta(\boldsymbol{f} - \boldsymbol{f}(\boldsymbol{x})) \tag{21}$$

where K is the kernel function. In this work, we use the Gaussian kernel. Its variance is chosen to be η times the MAD of the seed group. η is preset to be 0.1, and we have found that this works for all the cases tested.

3.1.2 Contour evolution

To simplify the notation, we present the contour evolution in 2D. However it is noted that the method can be easily extended to 3D. In fact, all the experiments in Section 3.2 are in 3D. First, we denote the family of evolving closed contours as $C_i : [0,1] \times \mathbb{R}^+ \to \mathbb{R}^2$. Without interactions among contours (interaction is addressed in Section 3.1.3 below), each contour evolves independently in order to minimize the energy functional:

$$E_i(C_i) := \int_{\boldsymbol{x}\ in\ C_i} (p^c - p_i(\boldsymbol{x}))d\boldsymbol{x} + \lambda \int_{C_i} ds \tag{22}$$

where p^c is the cut-off probability density used to prevent the contour leakage [68]. Likewise, $\lambda > 0$ is the smoothness factor. Computing the first variation of E_i and we obtain the flow of C_i:

$$\frac{\partial C_i(q,t)}{\partial t} = [p^c - p_i(C_i(q,t)) + \lambda \kappa_i(q,t)]\, \boldsymbol{N}_i(q,t) \tag{23}$$

in which $\boldsymbol{N}_i$ is the inward unit normal vector field on C_i and κ_i is the curvature of the contour.

3.1.3 Contour interaction

Although the p^c term in equation (22) helps to prevent contour leakage, in many cases the result is not sufficiently satisfying. Indeed, it often results in the problem that certain regions are over-segmented, while some others are under-segmented. The leakage issue, i.e., making decisions in a transitional region, is sometimes a difficult task even for the human visual system. However, one particular strategy the visual system takes, is to approach the decision boundary from both sides by competition, rather than preventing the leakage from a single direction. To this end, we enable the interaction amongst the previously individually evolving contours using standard principles from Newtonian mechanics. First, we regard the right hand side of equation (23) as the force applied on the infinitesimal curve segment at the position $C_i(q,t) =: \boldsymbol{p} \in \mathbb{R}^2$. Now with the interaction among curves, such a curve segment will also experience forces from other curves:

$$F_i^{ext}(\boldsymbol{p}) = -\sum_{j \neq i} \int_{C_j} e^{|\boldsymbol{p} - C_j(w,t)|} (p_j(\boldsymbol{p}) - p^c) \boldsymbol{N}_j(\boldsymbol{p}) dw. \tag{24}$$

Accordingly, the curve flow equation for C_i is now updated as:

$$\frac{\partial C_i(q,t)}{\partial t} = \left[p_i(C_i(q,t)) - p^c - \lambda \kappa_i(q,t)\right] \boldsymbol{N}_i(q,t) + F_i^{ext}(C_i(q,t)). \tag{25}$$

The exponential term controls the "influence range" of the force. When curves are far away, this term reduces the F_i^{ext} effectively to zero. Moreover, using the "sparse field level set" implementation [66], the computation of F_i^{ext} is very efficient. In general, the contour evolution scenario is as follows: At the outset, the contours do not touch each other because the seeds are sparsely scattered in the domain. Thus each F_i^{ext} is approximately 0 and each contour evolves individually. As the evolution proceeds, the contours get closer and the mutual interactions begin to take place. Moreover, they will compete and finally rest at balanced (equilibrium) positions. Throughout the whole process, the contours are governed by the action/reaction principle from mechanics, and will never overlap with each other, which is a necessary feature for multi-object segmentation.

3.2 Experiments and Results

Numerically, the contour evolution is implemented using the sparse field level set method for fast computation and flexibility in contour topology [66]. Moreover, in computing the robust statistics, the neighborhood size $B(\boldsymbol{x})$ is fixed at $3 \times 3 \times 3$. This value was used throughout all of our tests. Similarly, the p^c, λ in equations (25) are respectively fixed at 0.1, 0.3 for all of the tests. In what follows, we demonstrate the application of the proposed method in two scenarios: T1 weighted MR brain imagery (vervet and human), and CT abdominal data. The results are compared quantitatively with manual segmentations.

3.2.1 Vervet brain segmentation

We first test on a T1 weighted MR images of the brain of vervets. In order to highlight the leakage problem as well as how the proposed multi-object scheme solves this problem, initially, only the white matter is segmented. As shown in Figure 18(a), the contour leakage gives a final result that contains not only white matter but also part of cerebellum. However, using the proposed method to segment several related objects gives the result shown in Figure 18(b). It can be seen that the final labeling of the cerebellum, shown in white, not only fully captures the cerebellum region, but also effectively prevents leakage from intruding into the white matter. Furthermore, we show the 3D views of the multiple segmented objects: white matter, cerebellum, and ventricle. To highlight the region where the contour interaction between the white matter and cerebellum helps prevent leakage, we show the view from both posterior and inferior. It can be observed that there is no intersection between the contours. In particular, the cerebellum contour nicely "pushes" the white matter contour out, and so prevents leakage into the cerebellum.

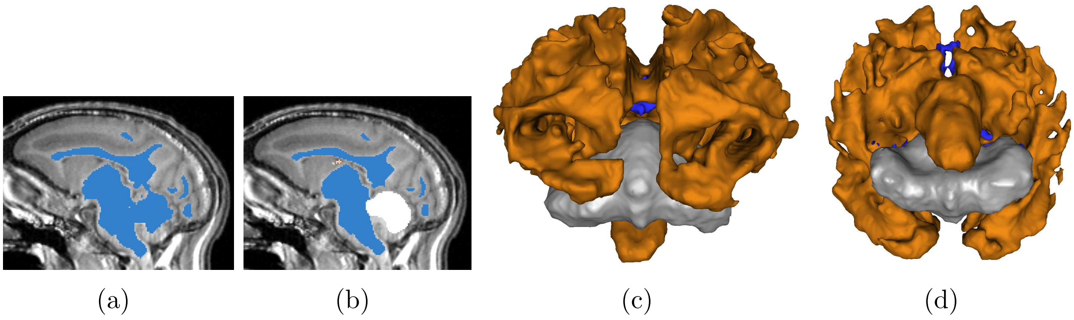

(a) (b) (c) (d)

Figure 18: In Subplot 18(a), we only segment one object (white matter). However, the contour leaks into part of cerebellum and part of brain stem. In 18(b), when segmenting several objects simultaneously, the white label for cerebellum effectively prevents the leakage. 3D plots include posterior 18(c) and inferior 18(d) views. It can be observed that there is no intersection between the surfaces.

3.2.2 Quantitative analysis for ventricle and caudate nucleus

In this second experiment, we extract both the ventricle and the caudate nucleus from MR images and present the results both qualitatively and quantitatively. In the experiment, the caudate nucleus is a difficult object to extract due to the poor contrast with its surrounding tissues. In fact, if we only place seeds in the caudate, we get the result shown in Figure 19(a) where the large leakage is circled. On the other hand, if we also places some seeds around caudate, we also capture some portion of white matter as shown in Figure 19(b) in almond color. Simultaneously, the caudate shape is kept intact and no leakage occurs. The almond part can be discarded because the caudate is the only object of interested and the final result is shown in Figure 19(c).

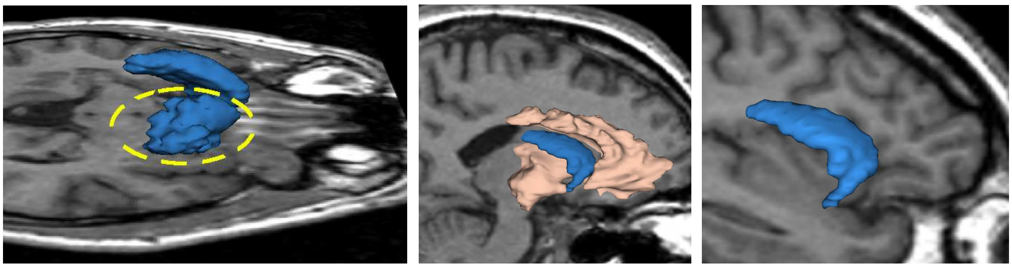
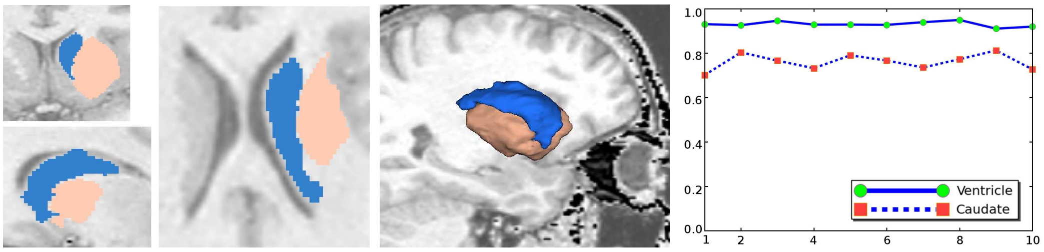

(a) (b) (c)

Figure 19: If only place seeds in caudate we get segmentation in Subplot 19(a) where the leakage is circled in yellow (viewing from superior-right). After putting some auxiliary seeds in the surrounding tissue we get results in the sagittal view in 19(b) where the caudate shape is kept intact. Discarding the auxiliary region and the caudate is shown alone in 19(c). (Sagittal view from right.)

Performing the same scheme on another subject gives the results in Figure 20(a) and 20(b) where we show both the segmentation and the original image. In addition to the caudate, the method is also applied on ventricle which is an easier segmentation task. In total, we performed 10 tests on different subjects. The Dice coefficients are computed against expert segmentations, and are plotted in Figure 20(c).

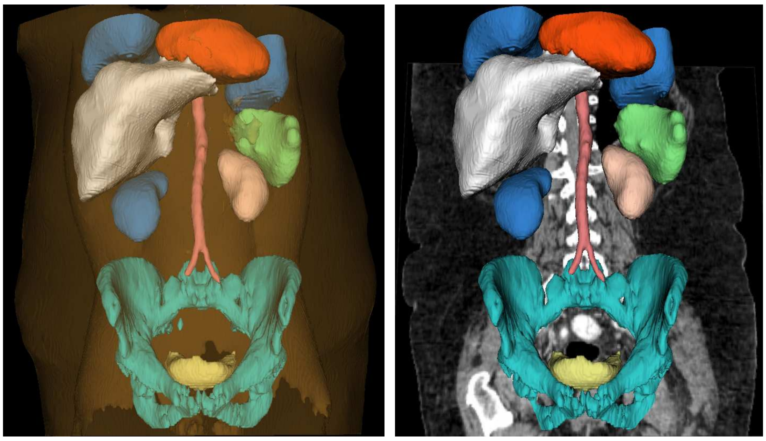


(a) (b) (c)

Figure 20: Subplot 20(a) and 20(b) overlay the segmentation results on the original images. The almond region is again auxiliary for preventing leakage. Subplot 20(c) shows the Dice coefficients of segmenting 10 ventricles and caudates, comparing with expert segmentation.

3.2.3 Abdominal organ segmentation

The proposed algorithm is general purposed and can be used for many different tasks. Indeed, although the previous examples only utilize the multi-object segmentation capability for leakage prevention, in the last experiment, 11 different organs/tissues are extracted from an abdominal CT image. The size of the image is $512 \times 512 \times 204$ and the running time on a machine with 3.0GHz Intel Core 2 Quad CPU and 8G memory is about 8 minutes. The result is shown in Figure 21.



(a) (b)

Figure 21: Segmentation of heart, two lungs, liver, two kidneys, spleen, abdominal aorta, pelvis, bladder, skin/muscle/fat. The subplot 21(b) removes skin/muscle/fat but overlays the original image.

3.3 Discussion

In this note, we proposed a general-purpose image segmentation scheme for medical data. In particular, the image features are extracted using certain local robust statistics as the segmentation criterion. Subsequently, the object characteristics are learned from the user initialization which is further used to guide the active contour evolution in a variational framework. Furthermore, we incorporate the interactions between the contours into the evolution motivated by simple principles from mechanics. This not only effectively reduces the contour leakage, but also results in a multi-object segmentation scheme without assuming that the union of the segmentation regions is the entire the whole domain.

Future work includes exploring more choices for the image features, such as Fourier/wavelet descriptors. Furthermore, we will incorporate shape priors for the multiple targets. Combined with the contour interaction, this is expected to further improve our results.

CHAPTER IV

IMAGE REGISTRATION AND SEGMENTATION IN A COUPLED SHAPE BASED MODEL

In some situations, the data driven image segmentation alone is not sufficient to guarantee the outcome of the segmentation. Indeed, the noise as well as the low contrast still remain to be difficult problems. To further improve the results, we need the shape prior as a guiding information in the segmentation process. Therefore, in this chapter we combine the image registration, statistical learning, and image segmentation in a coupled shape-based framework.

4.1 Proposed method and experiments

In Section 4.1.1, we demonstrate how to represent images as point-sets, and register images in the point-set framework under a particle filtering framework. After that, the shape prior is constructed in Section 4.1.2. Next, in Section 4.1.3, the shape prior is combined with the local image statistics to perform the segmentation. We note that experiments and results will be given as each method is presented.

4.1.1 Prostate Shape Registration

The way we represent images is crucial and so we summarize here some of the more common representations. We define an (intensity) image to be a non-negative function $f : \Omega \to \mathbb{R}^+$, where Ω is some compact domain in $\mathbb{R}^d$. In this work $d = 2$ or 3. Numerically, a given image may be represented as a discrete function defined on a uniform grid, where a value is associated with each spatial sampling location, namely the image intensity. We will refer to this as the discrete function representation (DFR). Using the probability density function (PDF), below we will define another representation called the point-set representation (PSR) for images. The central idea is to represent the image as a set of random samples rather than as a discrete function. We will show that such a representation is fully equivalent to DFR, and so no information is lost. However, PSR handles some of the difficult registration issues that are normally associated when the images are represented using DFR.

4.1.1.1 Point-set Representations of Images

We now assume that our image is represented by a continuous function f on the compact image domain Ω. Set $\|f\|_1 := \int_\Omega f(\boldsymbol{x})d\boldsymbol{x}$ and $\tilde{f}(\boldsymbol{x}) := f(\boldsymbol{x})/\|f\|_1$, we then have:

$$\begin{cases} \tilde{f}(\boldsymbol{x}) \geq 0, \quad \forall \boldsymbol{x} \in \Omega & (26) \\ \displaystyle\int_\Omega \tilde{f}(\boldsymbol{x})d\boldsymbol{x} = 1 & (27) \end{cases}$$

This simple normalization allows us to treat $\tilde{f}$ as a PDF defined on Ω. In doing so, we can then represent the image f by drawing samples from this distribution. Indeed, we adopt the rejection sampling algorithm of [12] in order to obtain M samples from the above distribution, giving the set of points $\mathscr{P} = \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_M : \boldsymbol{p}_i = (p_i^x, p_i^y, p_i^z) \in \Omega\}$.

Unlike DFR, where a real (or integer) number is associated with each spatial position, the image is now purely represented by 2D or 3D points. Consequently, the higher intensity regions in the image are now represented as the denser points in the corresponding point set.

It is easy to show that the DFR representation can be easily obtained, modulo a normalization factor, from the PSR. Indeed, this is actually the PDF estimation problem [15]. Given the PSR $\mathscr{P} = \{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_M : \boldsymbol{p}_i = (p_i^x, p_i^y, p_i^z) \in \Omega\}$, the DFR can be approximated as:

$$\tilde{f}_M(\boldsymbol{x}) := \frac{1}{M}\sum_i^M K_\sigma(\boldsymbol{x} - \boldsymbol{p}_i) \qquad (28)$$

where $K_\sigma(\boldsymbol{x})$ is a kernel function and σ is its bandwidth. As $M \to \infty$ and $\sigma \to 0$, we have $\tilde{f}_M(\boldsymbol{x}) \to \tilde{f}(\boldsymbol{x})$ [15].

4.1.1.2 Affine Image Registration under PSR

Given the images $f, g : \Omega \to \mathbb{R}^+$ we can obtain the corresponding PSR's as described in the previous section. We denote the point-set for g as $\mathscr{Q} = \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N : \boldsymbol{q}_i = (q_i^x, q_i^y, q_i^z) \in \Omega\}$. Then, we register the two images f and g by aligning their corresponding point-sets $\mathscr{P}$ and $\mathscr{Q}$ using the following:

$$E(\boldsymbol{A}, \boldsymbol{t}) := \frac{1}{N} \sum_{i=1}^{N} \|\boldsymbol{A}\boldsymbol{q}_i + \boldsymbol{t} - Cl(\boldsymbol{A}\boldsymbol{q}_i + \boldsymbol{t})\| + \frac{\lambda}{\det^2(\boldsymbol{A})} \tag{29}$$

where $\boldsymbol{A} \in \mathbb{R}^{3\times3}$, $\det(\boldsymbol{A}) \neq 0$ is the affine transformation matrix, $\boldsymbol{t} \in \mathbb{R}^3$ is the translation vector, and $Cl : \mathbb{R}^3 \to \mathscr{P}$ maps a point in $\mathbb{R}^3$ to its closest point in $\mathscr{P}$. Note that the second term in (29) with weighting $\lambda > 0$, penalizes $\det(\boldsymbol{A})$ from getting close to zero. Indeed, this is because the optimization process wants to register the two point sets by minimizing the cost E with respect to A and t. However, without the second term, one cheap but incorrect way to minimize E is to set A to be the zero matrix. Then the moving point set, after multiplied by A, will degenerate to a single point $(0, 0, 0)$, and one can then set t to the coordinate of any point in the fixed point set for a "perfect match". In such a scenario, the cost function E will be 0, but apparently it is a false result. Hence, to prevent such a degeneration, we penalize the trend of the determinant of A going to 0 in E with the second term. In this case, if the determinant A becomes close to zero, this term will increase to prevent such situation.

Concerning numerical details, a KD-tree data structure may be utilized to achieve a fast ($O(\log M)$) search [29]. The minimization of the registration cost functional E in (29) is a 12 dimensional nonlinear unconstrained optimization problem (6 dimensions for each 2D image). The gradient of E with respect to the affine matrix $\boldsymbol{A}$ and translation $\boldsymbol{t}$ are computed. Then the BFGS algorithm, one of the most popular quasi-Newton methods [41], is employed to obtain a fast (super-linear) convergence. The resulting (locally) optimal affine transformation parameters are applied to the image g.

One of the advantages of using PSR is that it naturally handles the case of large translations between the two given images. Specifically, many registration schemes under the DFR align the images by minimizing certain cost functionals. However, the cost functional usually takes an integral or summation form evaluated on the domain of overlap of the two images. Hence, a small cost functional value may be due to the effect of the overlapping domain being small, rather than the two images being well registered. This situation is often observed when two images differ by a large translation. In addition, under such a configuration the fastest way to reduce the cost functional is to further shrink the overlapping domain area by increasing the translation. However, this will only degrade the registration result. This is a fundamental drawback of using DFR. In contrast, without the concept of the fixed image domain, the registration in PSR naturally handles the above difficulty. In particular, when the translation is large, the gradient always tends to reduce this distance. This will be demonstrated in the experimental section.

Another commonly known problem with DFR is the long computation time. Under DFR, traversing of the domain grids may be quite time consuming. In contrast, PSR sparsely represents the image by far fewer points (comparing to the number of the grids in DFR). Hence, the solution time is significantly reduced by more than two orders of magnitude (see Table 1 below).

4.1.1.3 Prostate Shape Registration via Particle Filtering

Though PSR has certain advantages, it is still a local optimization procedure. Specifically, although large translations are effectively handled, large rotations are not. Unfortunately, a large rotation is common in prostate registration where the supine and prone views of the prostate need to be registered. However, in such supine/prone registration cases, we have the prior knowledge that the optimal rotation would be either close to 0° or 180°. Ideally, we would like to naturally incorporate this *a priori* information in a global registration setting. Thus, we treat the registration problem as system parameter estimation task where the twelve transformation parameters constitute the state variables of a dynamic system. Such estimation can be solved under the particle filtering framework. Moreover, using particle filtering, the *a priori* information can be easily combined. Therefore, we adopt the method in Section 2.2.1 to achieve the global optimal registration for the prostate shapes (under PSR).

In Section 4.1.1.1 two salient properties of using PSR in registration were discussed. It is further argued there that such advantages enable one to use particle filtering to achieve global registration under PSR. Firstly, to pursue global optimization, essentially all methods (e.g., simulated annealing, genetic algorithms, and even particle filtering) contain the idea of stochastically exploring a large part of the state space. However, under DFR those states corresponding to long translations will result in small or even zero cost functional values and therefore will be erroneously accepted. This fundamentally excludes the applicability of particle filtering to DFR-based (local) registration schemes. Contrastingly, the cost functional (29) behaves consistently, and can be nicely fitted to the particle filtering framework. Secondly, the global scheme is computationally more costly. Thus, the local step in the more global scheme should be computationally efficient. Here again, PSR fits well with this requirement.

4.1.1.4 Registration Experiments and Results

We provide experiments to demonstrate (1) the behavior of different registration cost functionals, (2) robustness of the proposed method to initialization, (3) supine/prone prostate registration, and (4) computational efficiency. It has been noticed that the number of the points in the set, which is derived via PSR, is a parameter for the proposed method. However, in all of the experiments performed, we use 500 points to represent a moderate size 2D image and 5000 points for a 3D image. Nevertheless, we have observed that the algorithm is fairly robust to the choice of number of points.

Cost functional behavior In the first set of experiments, we compare the region of convergence for several widely used cost functionals to that of our proposed method. These include mean square error (MSE) in [59] as well as the scheme based upon mutual information (MI) [65, 61]. This is achieved by first translating a 2D image, as seen in Fig. 22(a), in the x-y plane. Then, one can interpret the energy as a function of x-y displacement (Fig. 22(b)-22(d)).

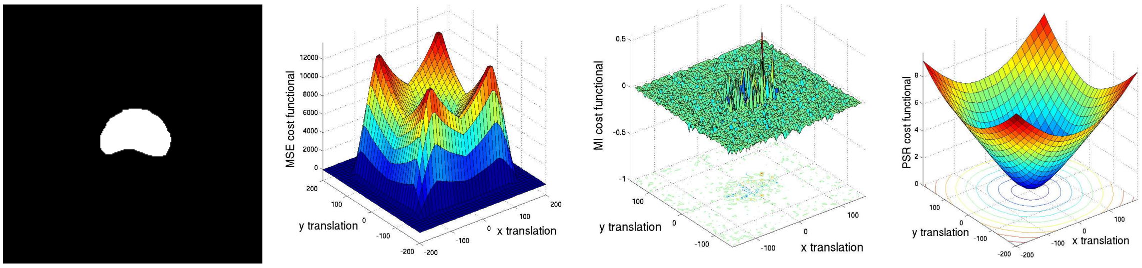
Ideally, the cost functional should have a minimum at $(0, 0)$ and smoothly increase as the translations increase. Fig. 22(b) is the plot of the MSE with respect to various translations. The valley in the middle can be regarded as the "region of convergence." That is, if the initial translation parameter is within this region, the MSE registration algorithm will gradually drive the parameter to converge to the ground truth. Not only is the region of convergence small (relative to that of PSR), more importantly, the cost functional values drop to zero when the transitions are large. As described in Section 4.1.1.3, such a phenomenon makes MSE an inadequate cost functional for "stochastic probing" based global image registration.

Fig. 22(c) is obtained in a similar fashion except that (negative) mutual information (MI) cost functional is now employed. Moreover, to mitigate the error resulting from the implementation, we opt to use the MI registration scheme found in the InsightToolkit [25][1]. Apart from being noisy, the cost functional is flat for most of the regions. Such behaviors of the cost functionals make the registration process very sensitive to the initialization and the optimization step size.

As can be seen in Fig. 22(d), the minimum of the proposed cost functional is

---

[1] The image is first normalized to $[-1, 1]$ and smoothed with a Gaussian filter with variance 10.0. Then, the image is translated and the MI is measured.



(a) Testing prostate binary image (b) MSE (c) MI (d) PSR

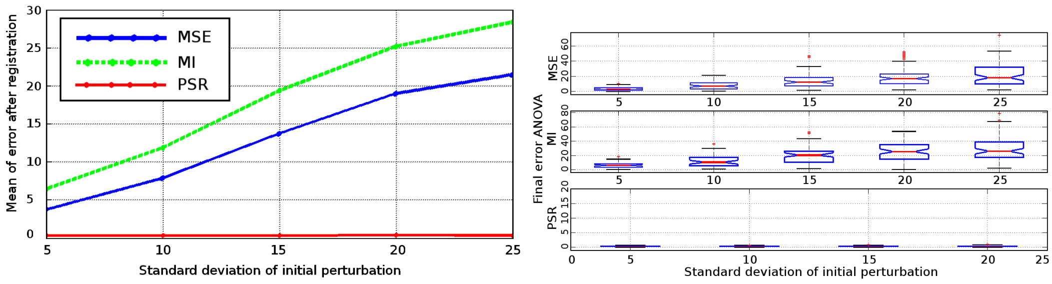
Figure 22: Plot of different cost functional values with respect to various 2D translations.

located at the correct position and smoothly grows monotonically outward as translation increases. This result can be mainly attributed to the representation provided under PSR, whereas MSE and MI registration approaches suffer the inherent drawback of representing an image on a fixed grid. Hence, no matter how large the translation is, the registration process is able to drive it to the correct place.

Robustness to Initialization Previous section gives more of a visual demonstration of the behavior of the various cost functionals. It provides the intuition but has certain limitations: Due to the dimensionality, the image is restricted to 2D and the transformation is restricted to 2D translation. However the real optimization is in 6D affine transformation space for 2D image, and 12D affine transformation space for 3D images, which are difficult to display. To evaluate the robustness in there, in this section we perform real test for 2D images. Specifically, two identical images are registered by starting from a random position in the registration parameter space. Therefore, the ground truth for the registration is the identity matrix and the zero translation vector. Again, three types of registration are compared, namely affine image registration using both MSE and MI as well as our proposed method. All three methods are run until convergence. Furthermore, to evaluate the performance of the registration, the affine matrix and the translation vector are concatenated together to form a 6 dimensional state space vector. The registration result is then compared with the ground truth by the vector l^2 metric, denoted as the “recovery error.” We note that the initial affine matrix and the translation vector are perturbed separately. This is because usually the perturbation for translation is one or two degrees of magnitude larger than that of the affine matrix. So if they are perturbed together, the recovery error will be dominated by its translation components.

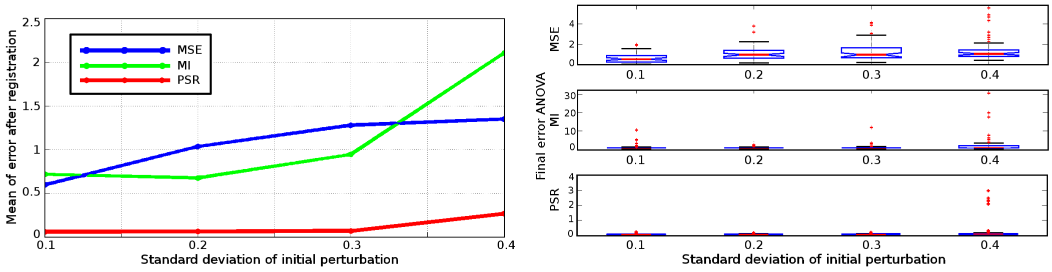
Thus, the robustness to the initial translation is tested first. The initial translation vector is set to a Gauss random variable with the standard deviation (STD) ranging in $\{5, 10, \ldots, 25\}$. For each STD, 100 realizations are generated as the initial translation, and the initial affine matrix is set to the identity matrix. After the registration converges, 100 recovery errors are recorded for one type of registration. Fig. 23(a) shows the means of the recovery errors at different STD levels. The horizontal axis shows the STD of the initial translations vector, while the vertical axis is the mean of the recovery error. It can be observed that when the initial perturbation becomes larger, the MSE and MI registration recovery errors grow larger. At the same time, the PSR always register the two images. Such a result is consistent with the cost functional analysis in the previous section. In addition to the means, Fig. 23(b) shows the spread in the recovery errors. Specifically, the notches indicate the medians of each set of 100 recovery errors while the box encloses those recovery errors within one quartile. This plot further demonstrates that when using MSE- or MI-based approaches, the median performance is not only poor, but the stability is also unsatisfactory.

The experiment is conducted similarly for the initial affine matrix. The initial translation vector is now set to zero while the affine matrix is a perturbed identity matrix. Specifically, each element of the identity matrix is added with a Gauss random variable with the STD ranging in $\{0.1, 0.2, 0.3, 0.4\}$. Similarly, 100 tests are performed for each STD for all three types of registration schemes and the recovery errors are recorded. Fig. 24 shows that as the perturbation on the initial state gets larger, the



(a) (b)

Figure 23: Recovery error analysis for initial translation perturbation. Details given in text.



(a) (b)

Figure 24: Recovery error analysis for initial affine matrix perturbation. Details given in text.

mean recovery errors of all the three registration schemes grow. Interestingly, not only does the mean recovery errors of the PSR scheme grow the least, but their variances are also the smallest.

Supine-Prone prostate registration One challenging problem in prostate registration is to register the supine and prone prostates. Fig. 25 shows one case of the supine/prone prostates in the axial, sagittal and coronal views. The moving prostate 3D image (blue) is overlaid on the fixed image (white).

First, MSE image affine registration is used to register the two images and the result is shown in Fig. 26. The moving image (red) is stretched to align with the fixed image. However, the local registration scheme could not detect the global optimal configuration 180° away. Therefore, it provides an erroneous result where different sides of prostates are aligned.

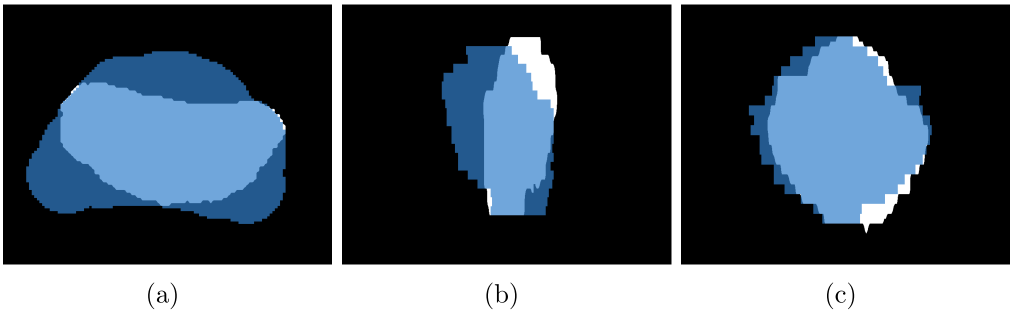

(a) (b) (c)

Figure 25: Supine/prone prostates, before registration. Subplots show the axial, sagittal, and coronal views.

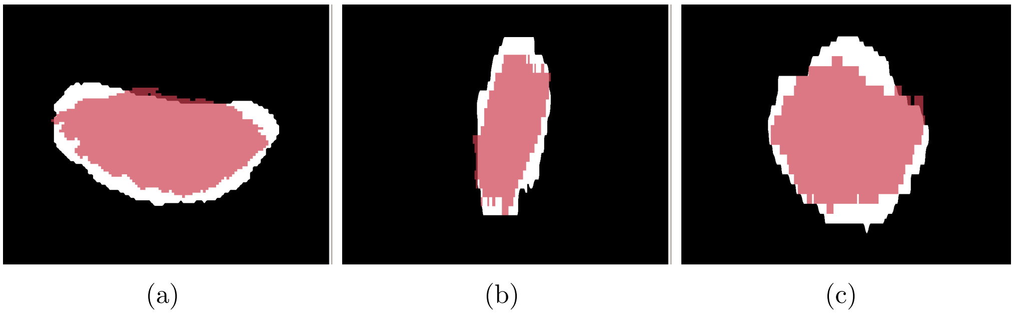

(a) (b) (c)

Figure 26: Supine/prone prostates registration using image affine registration with respect to MSE cost function. Subplots show the axial, sagittal, and coronal views.

The registration using the proposed global affine registration scheme under PSR is then conducted. Under the particle filtering framework, the prior knowledge can be incorporated into the construction of the prior distribution $p(\boldsymbol{x}_0)$ to reflect the fact that the two images may (or may not) differ by 180° around the z-axis. In other words, under the probability framework this can be interpreted by the fact that the optimal rotation around the z-axis, θ, has a higher probability of taking the values near 0° and 180°. Thus, its prior distribution is defined to be:

$$p(\theta_0) = \frac{1}{2\sqrt{2\pi}\sigma} \left(e^{-\frac{\theta^2}{2\sigma^2}} + e^{-\frac{(\theta-\pi)^2}{2\sigma^2}} \right), \tag{30}$$

whose plot is shown in Fig. 27. In cases where such prior knowledge is not available, a uniform distribution is a common choice.

The results generated by the proposed algorithm are shown in Fig. 28. The registered moving image is denoted by the red color and it can be observed that the

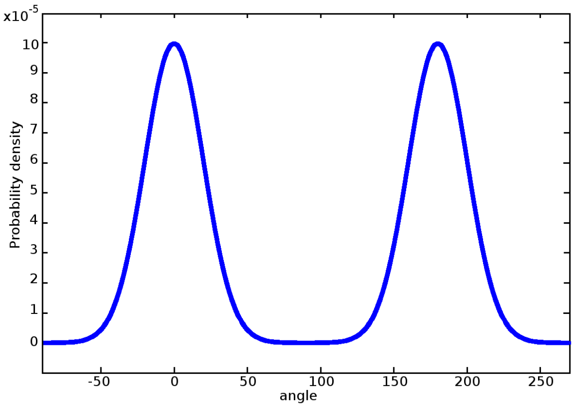


Figure 27: Prior for rotation

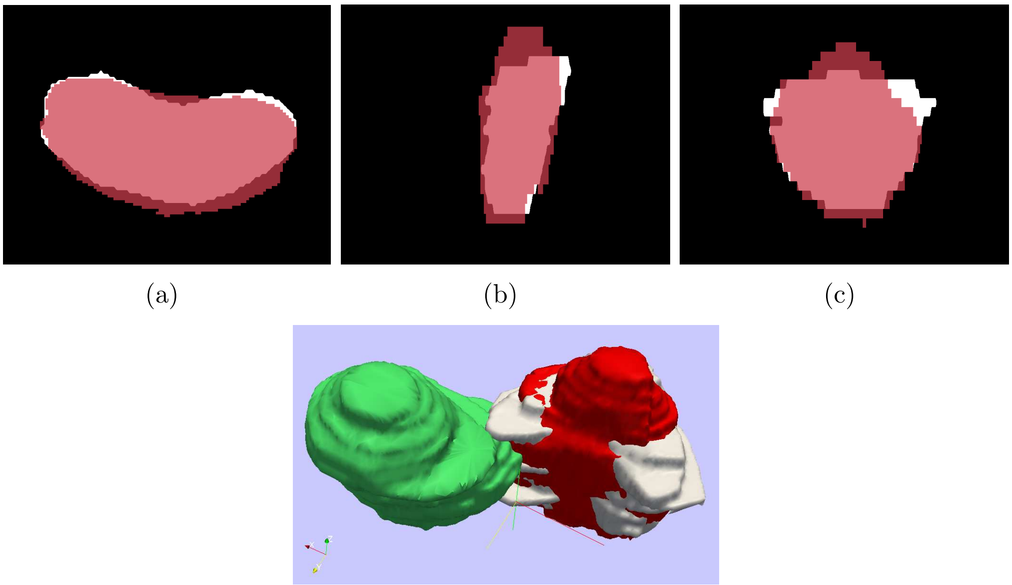
large rotation is correctly recovered. We can better appreciate the global registration results using the 3D view as seen in Fig. 28(d). The fixed prostate is again in white, the moving image (before registration) is in green, and its registered version is in red.

Furthermore, it is noted that although particle filtering is a general state estimation framework and theoretically registration using MSE/MI also seems to fit in such framework. However, for our purposes, it may not be the best choice. This is mainly due to two reasons: First, as plotted in Fig. 22, MSE and MI cost functionals depend on the overlapping of the image sample grids. Hence, to explore the optimal solution, particle filtering samples at those regions corresponding to (false) low cost functional values, such as the large remote areas in Fig. 22(b) and 22(c). Secondly, the computation time of using MSE/MI with particle filtering is very long especially for 3D case. This further prohibits the combination of MSE/MI with particle filtering.

In addition to a typical case for demonstration here, the global registration scheme has been applied to 112 data sets[2] to test the robustness of the supine/prone prostate registration. For this purpose, we arbitrarily pick one image as the template and align all the other 111 images with the template. As seen in Fig. 29, the variation

---

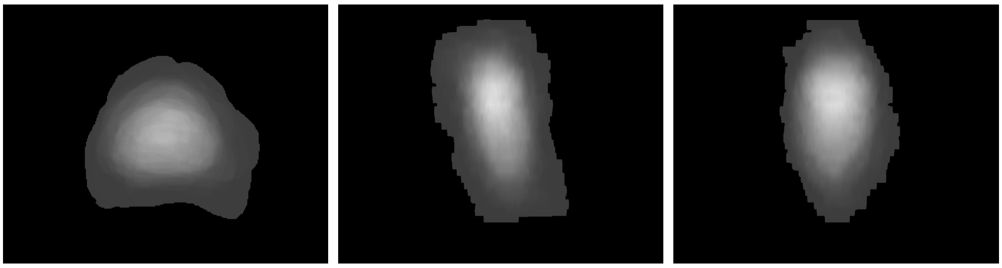
[2] public available at `http://prostatemrimagedatabase.com/index.html`



(a) (b) (c)

(d)

Figure 28: Supine/prone prostates registration using PSR affine registration and particle filtering. Subplots show the axial, sagittal, coronal, and 3D views.

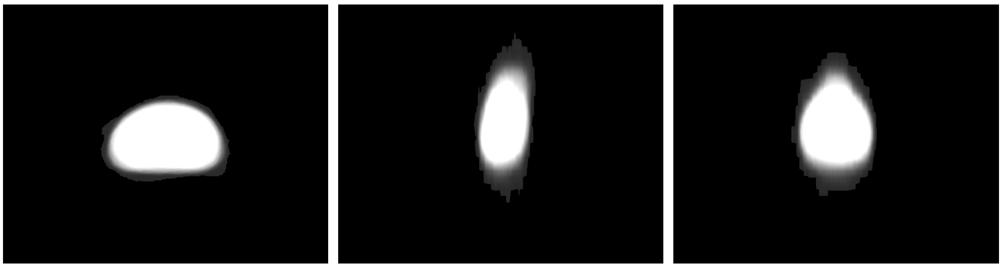


(a) Axial (b) Sagittal (c) Coronal

Figure 29: Overlay all the prostate shapes before registration.

in pose and shape among the 112 images are very large *before registration*. After the proposed supine/prone prostate registration method is run to convergence in all of the 111 tests, the registered images are summed and we arrive at the results presented in Fig. 30. It can be seen that the pose and shape variations are drastically reduced.

Increased Efficiency Registering image via point-sets is much more efficient, especially in 3D. The mean times of registering the 112 prostates using different methods are shown in Table 1. The efficiency is due to the sparsity of the PSR. For instance,



(a) Axial (b) Sagittal (c) Coronal

Figure 30: Overlay all the prostate shapes after registration.

Table 1: Running times comparison. Image size is 256 × 256 × 26. MSE and MI registration codes are implemented as in [25]. 5000 points are used for PSR.

Image	MSE	MI	Proposed Method
3D prostate	433.5 sec	610.0 sec	2.5 sec

in the prostate case shown in Fig. 28, 5000 points are used to represent the image, whereas 1703936 image voxels are contained in the whole image. Hence, the computation is very significantly reduced. All the methodologies are implemented in C++ with a Pentium®3.20GHz CPU with 4G RAM.

4.1.2 Shape Prior Construction

With the training shapes registered, the shape prior is constructed for the subsequent segmentation. Before learning, an appropriate shape representation is important. Interestingly, although binary/label maps are widely used in literature, they violate the usual Gaussian assumption in the PCA framework. This can attributed to the fact that the intensities of the binary/label map can only be 0 or 1, which are not likely to constitute a Gaussian distribution. The signed distance function (SDF) is also commonly used. However, the SDF representation generally has large values far from the zero level set. Therefore, during the learning step, the variations may overwhelm those around the zero contour causing inconsistencies in shape learning.

4.1.2.1 Shape representation using hyperbolic tangent

In this work, we modify the SDF via a transformation of the form $s(\boldsymbol{x}) = \mathscr{T}(\mathrm{SDF}(\mathbf{x}))$ to provide a better representation for shapes. More precisely, we want to choose the mapping $\mathscr{T}$ that preserves the zero level set of the SDF and eliminates the large values/variances far from the zero level set.

Sigmoid functions are good candidates for this purpose. Thus, a natural choice is to apply the hyperbolic tangent to the SDF to represent shapes:

$$s(\boldsymbol{x}) = \tanh(\mathrm{SDF}(\boldsymbol{x})) = \frac{e^{2\mathrm{SDF}(\boldsymbol{x})} - 1}{e^{2\mathrm{SDF}(\boldsymbol{x})} + 1}. \tag{31}$$

Note that $\tanh(0) = 0$ and $\lim\limits_{t \to \pm\infty} \tanh(t) = \pm 1$ so we preserve the zero level set as the object boundary, but eliminate the variance far from the boundary. This will benefit the learning phase. Moreover, since $\tanh'(x)|_{x=0} = 1$, we have that $\tanh(x) \approx x$ when $x \approx 0$. Hence, around the zero level set, $s(\boldsymbol{x})$ is close to the SDF. This representation of shape will be referred to as the *T-SDF* (T for tanh) in what follows.

Denote the N manually segmented training images, which have been previously aligned by the method described above, as $I_i : \Omega \to \{0, 1\}$ where $\Omega \subset \mathbb{R}^3$. The T-SDF representations of the registered images are then computed as $s_i = \tanh(\mathrm{SDF}(I))$: $i = 1, \dots, N$.

4.1.2.2 Shape learning

The standard PCA is adopted to learn the shapes. The mean shape is obtained as:

$$\bar{s}(\boldsymbol{x}) = \frac{1}{N} \sum_{i=1}^{N} s_i(\boldsymbol{x}). \tag{32}$$

Then, the mean shape is subtracted from each shape, i.e. $\tilde{s}_i = s_i - \bar{s}$. Since each $\tilde{s}_i$ is a 3D volume, we can concatenate the rows to form a long vector η_i. Then the covariant matrix is formed as

$$C = \frac{1}{N} \left(\eta_1, \dots, \eta_N\right) \left(\eta_1, \dots, \eta_N\right)^T, \tag{33}$$

and the singular value decomposition gives:

$$C = G\Lambda G^T \tag{34}$$

where Λ is a diagonal matrix containing the eigenvalues and the columns of G store the eigenvectors. Note that these are reshaped to the original image size and are denoted as $g_i(\boldsymbol{x})$. Usually, only the eigenshapes corresponding to the first L eigenvalues are kept while the others (with smaller eigenvalues) are ignored. Hence, the shape prior is a space spanned by $\{g_i : 1, \ldots, L\}$. In the subsequent segmentation, the shape is constrained to lie within this space.

We note that besides the shapes of the prostate, the mean and the variance of the image intensity within the prostate could also be learned if the training shapes and their corresponding original images are both available.

4.1.3 Shape-Based Prostate Segmentation

In this section, we describe our segmentation strategy for MR prostate data. Briefly, given the image to be segmented, it is preprocessed under a Bayesian framework to highlight the region of interest. Then a variational scheme based on local regional information is used to extract the prostate from the posterior image. We now give the details.

4.1.3.1 Bayesian preprocessing

Given an image $J(\mathbf{x})$, the likelihood $l_J(\mathbf{x})$ is computed as

$$l_J(\mathbf{x}) := \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(J(\mathbf{x}) - \mu)^2}{2\sigma^2}\right), \tag{35}$$

where the μ and σ are the mean and standard deviation of the object intensity, respectively. Both μ and σ may either be provided by the user or learned during the learning process. To compute the posterior we still need the prior term. While uniform priors are often used in previous works [21, 67] where the posterior is in fact the (normalized) likelihood, we show in this work that with a careful construction of the prior, the posterior is more convincing and makes the segmentation easier.

To this end, we propose to use the image content based directional distance in the prior. This means that both the image content and the distance to the object center are considered when calculating the distance. Specifically, given the image $J(\mathbf{x})$, we first construct the metric [46, 48, 38] $d : \Omega \times A \to \mathbb{R}^+$ by:

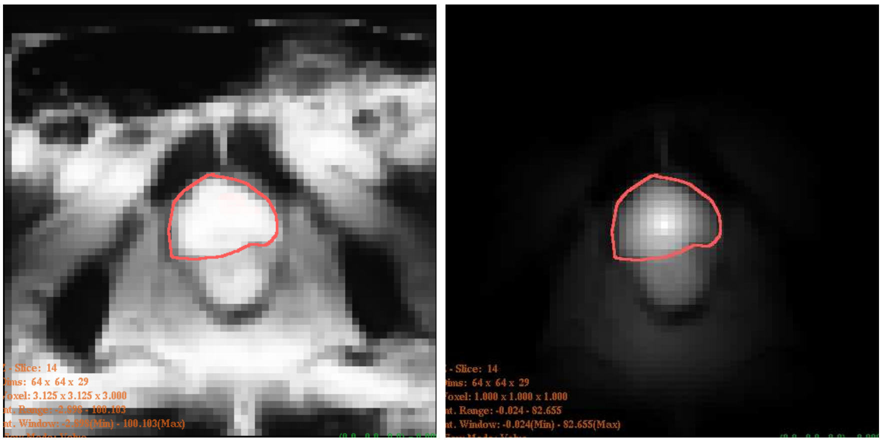
$$d(\mathbf{x}, \phi) = |J(\mathbf{x} - J(\mathbf{x} + \phi))|\,, \tag{36}$$

where $\phi \in A = \{(0, 0, \pm1), (0, \pm1, 0), (\pm1, 0, 0)\}$. Denote the estimated center of the object by $\mathbf{p}_c$. Similarly to μ above, $\mathbf{p}_c$ may be learned or assigned. This enables us to compute a directional distance map (DDM) ψ by solving the Hamilton-Jacobi-Bellman equation:

$$\begin{cases} 0 = \inf\limits_{\phi \in A} \{\ d(\mathbf{x}, \phi) + \langle \phi, \nabla\psi(\mathbf{x}) \rangle \}\,, \\ \psi(\mathbf{p}_c) = 0. \end{cases} \tag{37}$$

Equation (37) may be solved efficiently by using the the fast sweeping method proposed by Kao *et al.* [27]. Thus, with ψ obtained in this manner, we can define the prior by $\textit{prior} := \exp(-\psi)$, and compute the posterior by the Bayesian rule.

Figure 31 shows the comparison of the posteriors computed using different priors. The images shown are one slice taken from the given 3D prostate volume. In both images, the red contour shows the target object. In Figure 31(a), a uniform prior is used to compute the posterior. Since the average intensity within the prostate is similar to that of the surrounding tissue, both prostate and surrounding tissue have high posterior (bright), poorly differentiate the object from the background. On the other hand, in Figure 31(b), the prior was constructed from the DDM. Here the prostate is almost the only bright region in the posterior image. From the posterior image alone we can already differentiate the prostate fairly well. Though the bright



(a) (b)

Figure 31: Posterior image of a slice: (a) using uniform prior, (b) using DDM as prior.

region below the prostate is difficult to exclude at this stage, the subsequent shape-based segmentation recognizes it as background.

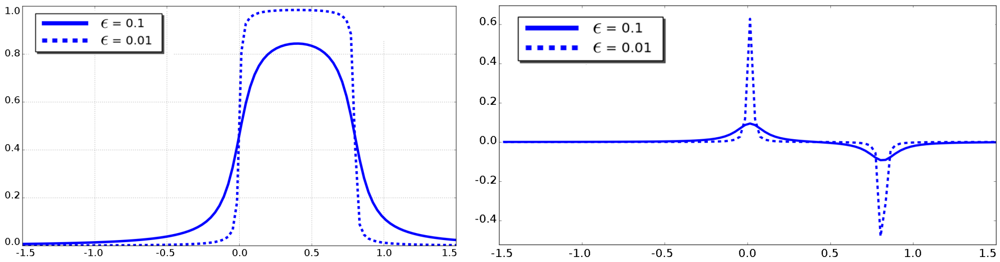
4.1.3.2 Segmentation in the posterior image

We propose a local regional information based segmentation scheme, in which the segmentation curve is driven to maximize the difference between the average posterior within a banded region inside and outside of the curve. More specifically, given the current segmentation curve, which is represented by the zero level set of the T-SDF, it is driven to enclose the desired object in the posterior image $K(\boldsymbol{x})$ by minimizing the following cost functional:

$$\begin{aligned} E &:= -\frac{1}{2}(u-v)^2 \\ &= -\frac{1}{2}\left(\frac{\int_\Omega \mathcal{H}_B(h(\boldsymbol{x}))K(\boldsymbol{x})d\boldsymbol{x}}{\int_\Omega \mathcal{H}_B(h(\boldsymbol{x}))d\boldsymbol{x}}\right. \\ &\qquad \left. - \frac{\int_\Omega \mathcal{H}_B(-h(\boldsymbol{x}))K(\boldsymbol{x})d\boldsymbol{x}}{\int_\Omega \mathcal{H}_B(-h(\boldsymbol{x}))d\boldsymbol{x}}\right)^2, \end{aligned} \tag{38}$$

where $h(\boldsymbol{x})$ is the T-SDF representation of the curve and is defined as:

$$h(\boldsymbol{x}) = \bar{s}(A\boldsymbol{x}+T) + \sum_i^L \omega_i g_i(A\boldsymbol{x}+T). \tag{39}$$



(a) Banded Heaviside function

(b) Derivative of the Banded Heaviside function

Figure 32: Banded Heaviside function and its derivative with different ϵ coefficient values.

Here, A and T denote the affine matrix and translation vector, respectively. Moreover, in equation (38), the $\mathcal{H}_B : \mathbb{R} \rightarrow \mathbb{R}$ is the "banded Heaviside function" defined as:

$$\mathcal{H}_B(t) = \frac{1}{\pi}\left(\arctan\left(\frac{t}{\epsilon}\right) - \arctan\left(\frac{t-B}{\epsilon}\right)\right). \tag{40}$$

The banded Heaviside function and its derivative are plotted in Fig. 32.

The banded Heaviside function realizes the localized property of the cost functional. Therefore, u is the mean of posterior in a banded region inside the object. The width of the band is determined by $B \in (0, 1]$. ($B = 0.8$ is a choice that worked well for all our tests.) Similarly, v is the mean of posterior in a banded region outside the object. Comparing to [71, 7], this cost functional is more robust to the influence remote to the curve. Moreover, because the value range of the T-SDF representation is $(-1, 1)$, as $B \rightarrow 1$, E in (38) converges to the global cost functional as in [71, 7].

By minimizing the segmentation cost functional with respect to T, A and the ω_i's, the optimal contour and transformation are found. To achieve this, the gradient of E is calculated and this finite dimensional (12+L dimensions in 3D) nonlinear optimization problem is solved using the BFGS method for fast convergence [41].

4.1.3.3 Segmentation Results

33 MRI prostate data sets were collected from different subjects on a 3.0 Tesla Philips machine from Queen's University, Kingston, Ontario, Canada [31]. The image volume

grid size is $256 \times 256 \times 27$ with $0.51mm$ in-plane resolution and $3mm$ slice thickness. The boundaries were manually traced out, and 20 cases were utilized for learning while the other 13 for testing. For comparison, we also trained and tested the algorithm proposed in [59] with the same set of training data and testing data, respectively. Specifically, among the three choices for the segmentation energy in [59], the Chan-Vese model is chosen.

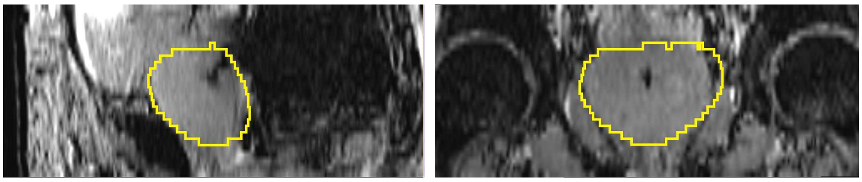
Fig. 33 shows the segmentation result for one patient in the testing data set. Note that we have chosen the number of principle modes to be $L = 6$. The center of the target object, i.e. the $\mathbf{p}_c$ in (37), is given by a click in the prostate region. The mean and standard deviation for (35) are 300 and 100, respectively.

Fig. 34 shows the segmentation result for a second patient in the testing data set. In the experiment, all the parameters are the same as the previous case except the mean and the standard deviation are 150 and 50, respectively. (In this figure, it appears that the intensity within the prostate is similar to that of Fig. 33, but this is the result of the window/level being adjusted for better visual appearance.)

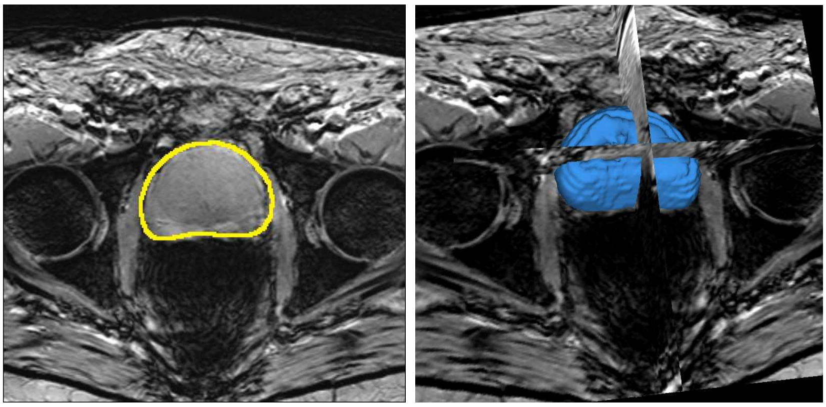
In the experiment of Fig. 35, the mean and the standard deviation are 500 and 200, respectively. Although the bladder in this image is extremely bright, the proposed method correctly captured the position and the shape of the prostate.

Fig. 36 shows another example where the prostate shape is different from the previous ones in that this one is more spherical. Still the method yields a visually excellent segmentation and this indicates that the learned shape prior does have the capacity to represent different shapes. Note, the mean and the standard deviation in this case are 200 and 100, respectively.

Furthermore, instead of showing the screen shots of all 13 testing results, we also compute the Dice coefficient which quantitatively compares each segmentation result (by the proposed method as well as the method given in [59]) with the corresponding manual drawing. The two sets of coefficients are plotted in Figure 37. It



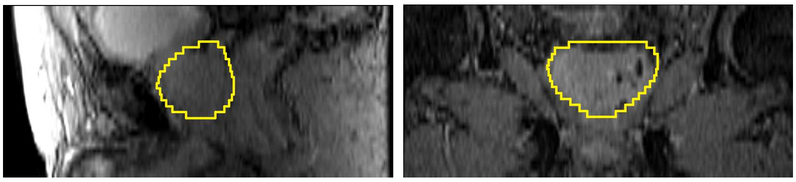
(a) Sagittal (b) Coronal



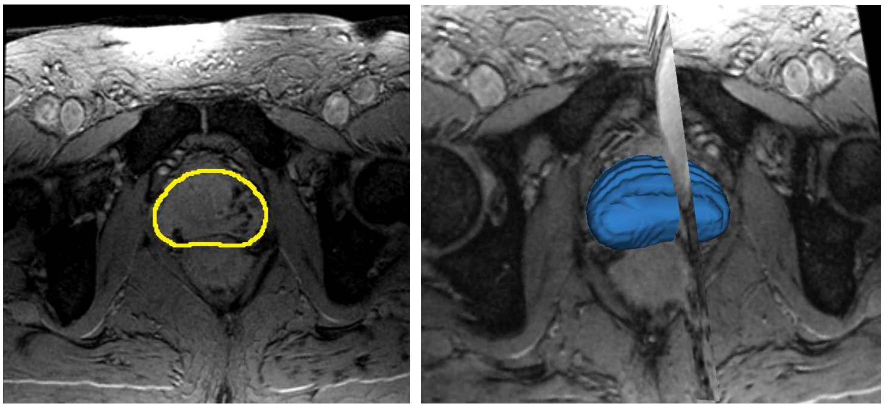
(c) Axial (d) 3D

Figure 33: Segmentation results for Patient 1

can be seen that the proposed method provides satisfying results overall. Note that the key reason that the method of [59] produces low Dice coefficients for data sets #3, #11, #12, and #1 is because it employs the Chan-Vese model which assumes the image to be bi-modal. However, those four images are T2 weighted image in which the bladder region is the brightest. Therefore, it extracts the brightest region and misses the prostate region. On the other hand, the proposed method in this paper only looks at the locally prominent features and hence is more robust to the influence in the remote regions.

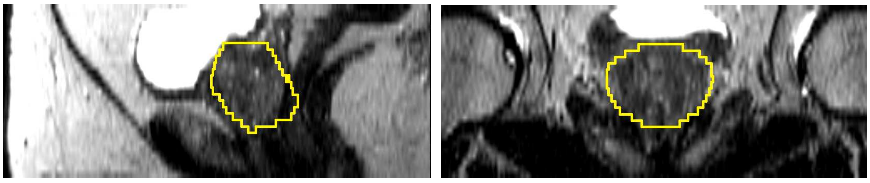


(a) Sagittal (b) Coronal

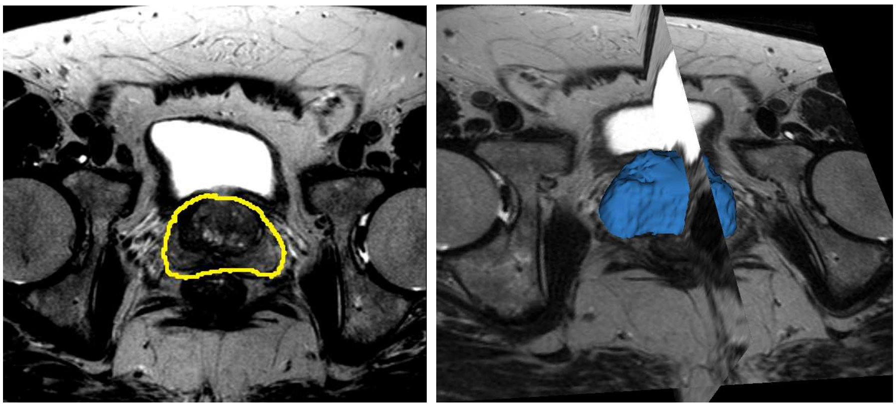


(c) Axial (d) 3D

Figure 34: Segmentation results for Patient 2

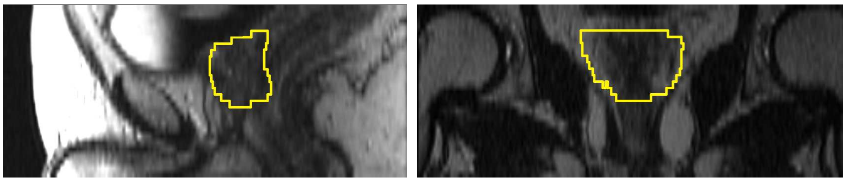


(a) Sagittal (b) Coronal

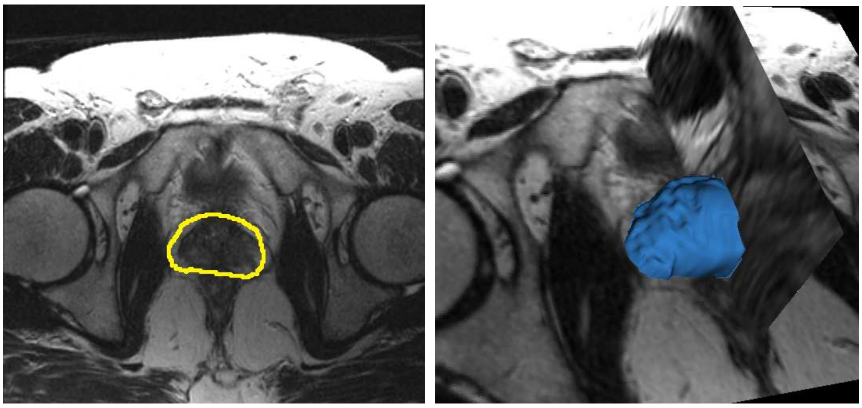


(c) Axial (d) 3D

Figure 35: Segmentation results for Patient 3



(a) Sagittal (b) Coronal



(c) Axial (d) 3D

Figure 36: Segmentation results for Patient 4

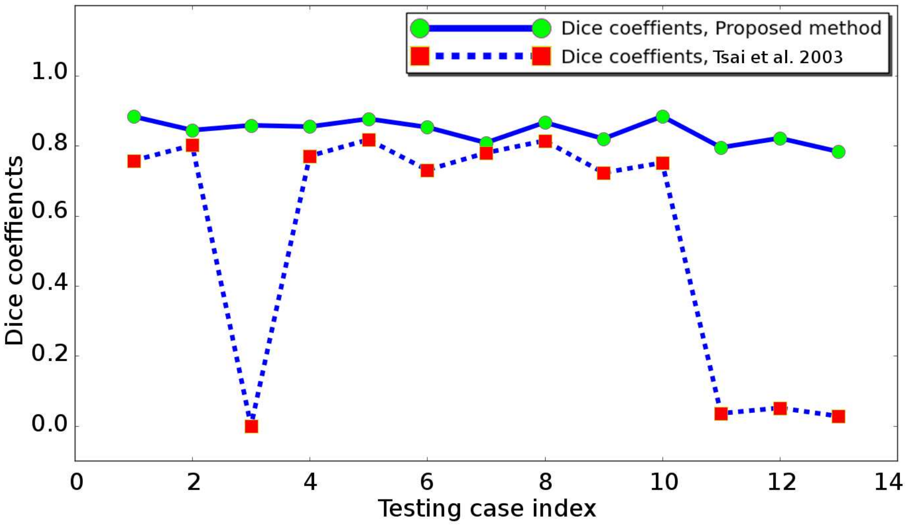


Figure 37: Dice coefficients of the proposed method and the method in [59]. The 3rd, 11th, 12th, and 13th testing images are T2 images where the bladder is very bright. Hence the Chan-Vese model used in [59] extracts the bladder region instead of prostate.

CHAPTER V

MULTI-RESOLUTION SHAPE REPRESENTATION AND SHAPE BASED SEGMENTATION USING WAVELETS WITH APPLICATIONS TO RADIOTHERAPY PLANNING

The method proposed in Chapter 4 has certain disadvantage that the number of eigen-shapes are limited by the number of training set, which is usually not large. Furthermore, the eigen values corresponding to the eigen-shapes indicate the magnitude of the variances. Hence, the local/small scale shape variance is likely to be overwhelmed by a more global/large scale shape variance. However, in many cases the small "bump" on a smooth surface is an important indicator of pathology. To address this issue, we propose a multi-scale representation of shape. Furthermore, the shape is learned in multiple scales, resulting in eigen-vectors in multiple scales. Therefore, not only this provide more abundant eigen-modes, but also the local/small shape variance will not get overwhelmed by global/large ones.

Furthermore, such segmentation technique can be applied to the scenario of radiotherapy planning. Indeed, during the past half-century, the cornerstone of treatment for brain metastases has been whole brain irradiation (WBI). WBI has multiple salutary effects including rapid relief of neurological signs and symptoms as well as enhanced local control. Unfortunately, WBI may also engender side effects including memory deficits and decrements in quality of life. Since memory control is thought to be mediated by the hippocampus, attention has been turned to whole brain radiotherapeutic techniques that allow sparing of the hippocampus. In order to be able to minimize dose deposition within the hippocampus, clinicians must be able to confidently identify that structure. However, manually tracing out the hippocampus for each patient is time consuming and subject to the individual bias. To this end, an automated method can be very useful for such a task.

5.1 Proposed Method

In this section, we provide the details of using the proposed multi-scale shape based segmentation technique to extract the hippocampus region from the MR brain imagery. In general, the method can be divided into an off-line learning step and an on-line segmentation step. In the learning step, a set of MR brain images (called training images) along with their segmented label maps for the hippocampus (and the amygdala) are provided for the statistical learning algorithm. The reason for including amygdala is because the visual boundary between hippocampus and amygdala is hardly seen in the images at hand. Hence, in order to obtain an accurate hippocampus segmentation, we first treat the hippocampus-amygdala as a single complex in the shape learning. After the shape of the complex is learned, such shape information is then utilized in extracting the complex from a new MR brain image (called the raw image). Next, the hippocampus and the amygdala are separated to form the final result. The details of each step will now be provided.

5.1.1 Shape learning

A set of N training MR images: $J_1, J_2, \ldots, J_N : \mathbb{R}^3 \to \mathbb{R}$ is provided for the algorithm, along with the corresponding label maps: $U_1, U_2, \ldots, U_N : \mathbb{R}^3 \to \{0, 1, 2\}$ in which the hippocampus and amygdala are labeled by 1 and 2, respectively. For the reasons stated above, the labels of the amygdala and hippocampus are first combined as a single label. That is, we define $\tilde{U}_i(\boldsymbol{x}) = 1$ if $U_i(\boldsymbol{x}) = 1$ or 2. Then, the shapes of the hippocampus-amygdala complex (HAC) are aligned so that the variances captured by the subsequent statistical learning step are only due to shape, not pose. The aligned shapes are denoted by $V_1, \ldots, V_N$. Next, in order to aid the statistical learning and the subsequent segmentation, we represent the aligned shapes as *sparse field level set* (SFLS) function [66], and denote them by $L_1, \ldots, L_N$.

At this point, we are able to formulate the multi-scale shape representation for the HAC's. In general, given a SFLS function L, its wavelet decomposition is written as:

$$W(\boldsymbol{u}) := DWT[L(\boldsymbol{x})] \tag{41}$$

where we assume the domain dimensions of L are all power of 2. Different wavelet bases can be chosen. In fact, in this study we tested the Daubechies wavelets $D4$, $D6$, and $D8$, but the final segmentation results were not significantly affected by the choice. Moreover, the wavelet coefficients are shrunk using the soft-thresholding proposed in [13] to remove the noise on the shape. By slight abuse of notation, the denoised wavelet coefficients are still denoted by W, which are stored in an 3D volumetric array the same size as L. Next, in order to represent the shape in multiple scales, the coefficients are decomposed into multiple bands corresponding to various scales. As is illustrated in Figure 38, the coefficients in each difference scale are concatenated to form a long column vector, called band, and all coefficients are rearranged into different bands.

Formally, we write this operation as:

$$\mathcal{R}[W] = \{B_1, B_2, \ldots, B_K\} \tag{42}$$

where K is the total number of bands. For example, in Figure 38 there are $K = 13$ bands in total. It is noted that later we will need to "inversely rearrange" the wavelet coefficients from the bands and this process is denoted as:

$$\mathcal{R}^{-1}\left[\{B_1, B_2, \ldots, B_K\}\right] = W \tag{43}$$

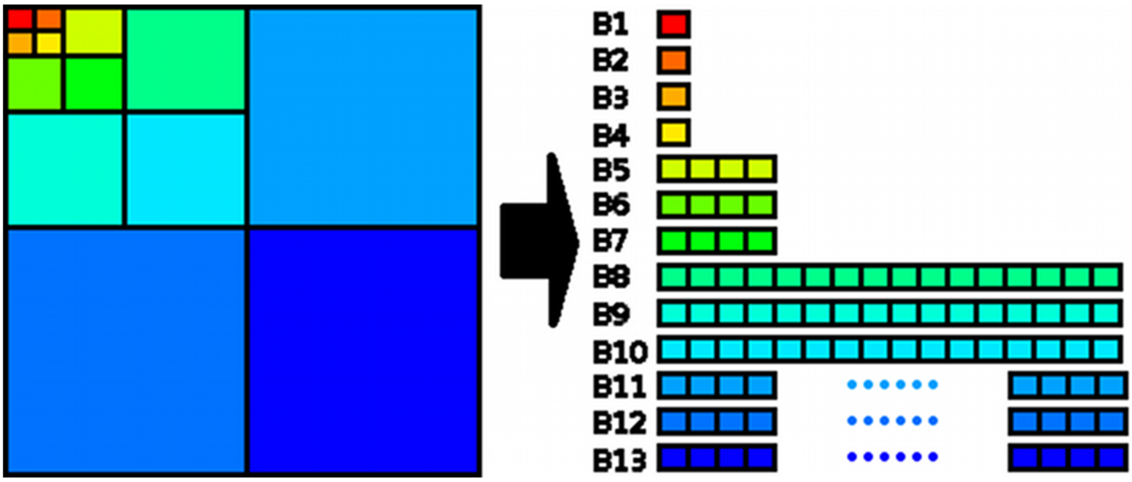


Figure 38: Decomposition of wavelet coefficients into various bands. The band structure of wavelet coefficients for an image having size 16 × 16. The coefficients in difference scales are rearranged into different vectors (bands).

Up to this point, the boundaries of the N hippocampus-amygdala complexes are now represented by N *band structures*:

$$\mathcal{B}^{(i)} = \left\{B_1^{(i)}, B_2^{(i)}, \ldots, B_K^{(i)}\right\}, i = 1, \ldots, N \tag{44}$$

A nice property of such multi-scale representation derives from the fact that the shape variance in different spatial locations and frequency scales are now separated in various bands.

In order to statistically learn the shape variances existing in the training shapes, we apply principle component analysis (PCA) to each band in order to compute the statistical model of the shape. More explicitly, the mean band structure is computed as:

$$\begin{aligned}\bar{\mathcal{B}} &= \left\{\bar{B}_1, \bar{B}_2, \ldots, \bar{B}_K\right\} && (45)\\ &= \left\{\frac{1}{N}\sum_{i=1}^{N} B_1^{(i)}, \frac{1}{N}\sum_{i=1}^{N} B_2^{(i)}, \ldots, \frac{1}{N}\sum_{i=1}^{N} B_K^{(i)}\right\} && (46)\end{aligned}$$

The mean shape can be considered as the shape without any variations, regardless of the scale of the variance. Because of this, it is expected that the mean shape computed this way should be consistent with the mean shape from a non-multi-scale shape representation. In fact, it is easy to show that such mean shape, defined as:

$$M := DWT^{-1}\left[\mathcal{R}^{-1}\left[\bar{\mathcal{B}}\right]\right] \tag{47}$$

is the same as the one computed in [34, 59].

While the proposed shape representation does not affect the mean shape, the multi-scale property is beneficial when the shape variances are analyzed. In order to utilize such property, PCA is performed in each band to capture the shape variance. Formally, the mean of each band is first removed as $\tilde{B}_k^{(i)} = B_k^{(i)} - \bar{B}_k; i = 1, 2, \ldots, N$. The covariance matrix for the k-th band is then formed as:

$$C_k := \frac{1}{N}\sum_{i=1}^{N} \tilde{B}_k^{(i)}\left(\tilde{B}_k^{(i)}\right)^{\top} \in \mathbb{R}^{l_k \times l_k} \tag{48}$$

where $l_k = \left|B_k^{(1)}\right|$ denotes the length of the k-th band. Furthermore, the eigen-decomposition of C_k gives:

$$C_k = Q_k \Lambda_k Q_k^{\top} = \left(\boldsymbol{q}_1^k, \ldots, \boldsymbol{q}_{p_k}^k\right)\begin{pmatrix} \lambda_1^k & 0 & \ddots \\ 0 & \ddots & 0 \\ \ddots & 0 & \lambda_{p_k}^k \end{pmatrix} Q_k^{\top} \tag{49}$$

where $p_k := \min(N-1, l_k)$ and the vectors $\boldsymbol{q}_j^k; j = 1, \ldots, p_k$ consist of the bases for this band. In contrast to the PCA scheme used in [11, 34, 59], where for N training shapes there are at most $N-1$ bases for the whole learned shape space, here we have bases in each band to record the shape variance at the given scale. Specifically, since the band length is varying, in the k-th band, the number of bases does not exceed p_k. For an 2D image with size 16×16 shown in Figure 38, it can be seen that starting from the band B_{11}, the band lengths $l_k, (k \geq 11)$ are greater or equal than 64. So for a moderate size of training shapes, except for the first few bands, we have $p_k = N-1$. Hence the total number of bases is: $\sum_{k=1}^{N} p_k \approx K(N-1)$. In comparison to $N-1$, this greatly enhances the capability of capturing the shape variances at multiple scales.

In order to construct the allowable shape space which restricts the subsequent segmentation, we assume the eigenvalues obtained in Equation (48) are ordered decreasingly. Furthermore, the first Y eigenmodes, $\boldsymbol{q}_1^k, \dots, \boldsymbol{q}_Y^k$ are picked to record the shape variance in this band (scale). In practice, we fixed Y at $\min(6, l_k)$. Moreover, for every $k = 1, \dots, K$, the shape space in that band, $\mathfrak{S}_k$ is defined as:

$$\mathfrak{S}_k := \left\{ \left(\boldsymbol{q}_1^k, \dots, \boldsymbol{q}_Y^k\right) \begin{pmatrix} \sqrt{\lambda_1^k} & 0 & \ddots \\ 0 & \ddots & 0 \\ \ddots & 0 & \sqrt{\lambda_Y^k} \end{pmatrix} \begin{pmatrix} a_1^k \\ \vdots \\ a_Y^k \end{pmatrix} ; a_i \in [-6, 6],\ \forall i \right\} \tag{50}$$

The above formula defines a hyper- rectangular shaped region centered at the mean in each band, with axes of length 6 times the standard deviation. Equivalently, this imposes a uniformly distributed prior for the shapes. In most of the literature, however, the Gaussian distribution is used. We should note here that in order to maximize the shape likelihood in the Gaussian case, there is always a tendency of evolving towards the mean shape which eliminates the individual shape variance. This is not always a desired property. On the other hand, the uniform distribution does preserve the shape variances. Finally, we note that the “shape space” above is in fact topologically closed but not a mathematical vector space, because it is not closed under the addition.

5.1.2 Shape based segmentation

In this section, we describe our multi-scale shape representation based segmentation strategy with its application on extracting the hippocampus from MR images. Briefly, given the image to be segmented, first we use a registration method to estimate the target location in the new image. Next, a data driven segmentation scheme is applied to extract the initial estimate of the shape. Thirdly, the multi-scale shape constraints are used to regularize the previously obtained shape in the learned shape space. The second and the third steps are iterated until arriving at a steady state, which provides the final segmentation. We now give the details.

5.1.2.1 *Shape initialization*

Given a new image $I : \mathbb{R}^3 \to \mathbb{R}$, the objective of the present section is to provide an initial shape, which is to be used as the starting point for the subsequent segmentation process presented in the next section. To this end, an atlas-based segmentation approach is adopted [50]. That is, all of the training images are deformably registered to the new image by maximizing mutual information [65, 52]. By using mutual information, the proposed method is capable of handling images of multiple modalities. With the optimal deformable transformations being computed and denoted by $T_1, \ldots, T_N$, the initial shape $P(\boldsymbol{x})$ is then defined using the "majority rule":

$$P(\boldsymbol{x}) = \begin{cases} 1, & \text{if } \sum_{i=1}^{N} \tilde{U}_i \left(T_i(\boldsymbol{x}) \right) \geq \frac{N}{2} \\ 0, & \text{otherwise} \end{cases} \tag{51}$$

This is to be used as the initialization for the subsequent segmentation for the new image I.

5.1.2.2 *Data driven segmentation*

Given the initial shape P, the proposed segmentation method is an alternating data-driven and multi-scale shape-based process that starts with a data-driven (not shape guided) algorithm. Specifically, we use the energy functional defined in [33]: Denote the initial contour, the level set representation of the initial shape $P(\boldsymbol{x})$ in the previous section, as ϕ. Then, the contour evolves to minimize the energy functional defined as:

$$E(\phi) := \alpha \int \delta\phi(\boldsymbol{x}) |\nabla\phi(\boldsymbol{x})| d\boldsymbol{x} + \int \delta\phi(\boldsymbol{x}) \int \chi(x, y) \left(\mathcal{H}(\phi(\boldsymbol{y}))(I(\boldsymbol{y}) - u_x) \right. \tag{52}$$

$$\left. + (1 - \mathcal{H}(\phi(\boldsymbol{y})))(I(\boldsymbol{y}) - v_x) \right) d\boldsymbol{y} d\boldsymbol{x} \tag{53}$$

In Equation (53), $\mathcal{H}$ is the smoothed Heaviside function defined in [33], and δ is the derivative of $\mathcal{H}$. Moreover, the χ is the characteristic function for the r-neighborhood around $\boldsymbol{x}$, and α is a positive weighting parameter. One can compute the first variation of E with respect to ϕ and we obtain the corresponding gradient descent flow for ϕ; see [33] for all the details. The evolution of ϕ is only performed for a few steps (not until convergence), and the resulting ϕ is then fed into the shape-filtering processing described in the next section.

5.1.2.3 *Shape filtering with shape prior*

Since physical imagery is never perfect (noise, etc.), the data driven segmentation ϕ performed in the previous section usually suffers from under-segmentation, leakage, or a combination of both. To solve this problem, the extracted shape in the previous step is processed through a "shape filter" so as to fall into the admissible shape space. First of all, the current contour ϕ is registered to the mean shape M in Equation (47) by minimizing the following energy over all $T \in bmT$, the space of 3D similarity transformations:

$$E_M(T) := \int \left(\phi(T(\boldsymbol{x})) - M(\boldsymbol{x})\right)^2 d\boldsymbol{x} \tag{54}$$

The optimal similarity transformation is denoted by T^* and the registered contour $\phi(T^*(\boldsymbol{x}))$ is denoted by $\psi(\boldsymbol{x})$. Moreover, $\psi(\boldsymbol{x})$ is sampled on the same discrete domain as $M(\boldsymbol{x})$, so that they have the same dimensions. Then, the wavelet transform is applied to $\psi(\boldsymbol{x})$ to obtain its band structure as:

$$\mathfrak{B}^{\psi} = \left\{B_1^{\psi}, B_2^{\psi}, \ldots, B_K^{\psi}\right\} \tag{55}$$

In a manner similar to that of the learning part of the algorithm, the mean bands are subtracted from $\mathfrak{B}^{\psi}$ to obtain $\tilde{\mathfrak{B}}^{\psi}$ as:

$$\tilde{\mathfrak{B}}^{\psi} = \left\{B_k^{\psi} - \bar{B}_k; k = 1, \ldots, K\right\} =: \left\{\tilde{B}_k^{\psi}; k = 1, \ldots, K\right\} \tag{56}$$

Next, in each of the K bands, the band $\tilde{B}_k^{\psi}$ is projected to the learned band space. Without loss of generality, we can take the k-th band. Accordingly, we project $\tilde{B}_k^{\psi}$ to the bases vectors in the band $\{\boldsymbol{q}_j^k; j = 1, \dots, p_k\}$ and we obtain the coefficients:

$$A_k = \left(\frac{\tilde{B}_k^{\psi} \cdot \boldsymbol{q}_j^k}{\sqrt{\lambda_j^k}}; j = 1, \dots, p_k \right)^{\top} =: (a_1, a_2, \dots, a_{p_k})^{\top} \in \mathbb{R}^{p_k} \tag{57}$$

In order to restrict the given novel space to the learned shape space, the novel shape is mapped to its closest shape in the allowable shape space learned in the ***Shape Learning*** part of the scheme. Accordingly, we adjust each A_k as:

$$\tilde{A}_k := \begin{cases} A_k, & if \|A_k\| \leq 6 \\ \frac{6A_k}{\|A_k\|}, & if \|A_k\| > 6 \end{cases} \tag{58}$$

After that, the adjusted band structure $\hat{\mathfrak{B}}^{\psi}$ is obtained as:

$$\hat{\mathfrak{B}}^{\psi} := \left\{ (\boldsymbol{q}_1^k, \dots, \boldsymbol{q}_Y^k) \begin{pmatrix} \sqrt{\lambda_1^k} & 0 & \ddots \\ 0 & \ddots & 0 \\ \ddots & 0 & \sqrt{\lambda_Y^k} \end{pmatrix} \cdot \tilde{A}_k + \bar{B}_k; \forall k \right\} \tag{59}$$

and the filtered shape $\tilde{\psi}$ is then obtained as:

$$\tilde{\psi} = DWT^{-1} \left[\mathcal{R}^{-1} \left[\hat{\mathfrak{B}}^{\psi} \right] \right] \tag{60}$$

Finally, the filtered shape is transformed back to the new image domain using the inverse of the similarity transformation T^* recorded above. Thus $\tilde{\psi}((T^*)^{-1}(\boldsymbol{x}))$ is then used as the initial segmentation contour for the next-round data driven segmentation. These two steps alternate until convergence. Usually, it only takes two or three iterations for convergence on the data sets we used.

5.1.2.4 Separation of the hippocampus and amygdala

The previous segmentation provides a binary image $H(\boldsymbol{x}) : \mathbb{R}^3 \rightarrow \{0, 1\}$ representing the hippocampus-amygdala complex, not the hippocampus. In order to form the final segmentation of the hippocampus, the complex needs to be separated. Such separation is achieved in this section using a scheme similar to described in ***Shape***

initialization section above. More explicitly, the binary label images $\tilde{U}_i(\boldsymbol{x})$'s are deformably registered to the newly obtained binary image $H(\boldsymbol{x})$. The optimal deformable transformations are denoted by $\tau_1, \ldots, \tau_N : \mathbb{R}^3 \to \mathbb{R}^3$. Then, the final label image $F(\boldsymbol{x})$ is defined as:

$$F(\boldsymbol{x}) := \begin{cases} 1, & if \sum_{i=1}^{N} U_i(\tau_i(\boldsymbol{x})) \in [N, 2N] \\ 0, & if \sum_{i=1}^{N} U_i(\tau_i(\boldsymbol{x})) \in [0, N) \end{cases} \tag{61}$$

5.2 Experiments and results

In this section, we report the results of 24 spoiled gradient-recalled (SPGR) coronal-plane volumetric T1-weighted images acquired in the Brigham and Women's Hospital, Harvard Medical School. The ages of the subjects range from 28 to 55 with mean 43.6 and standard deviation 7. The image resolution is $0.938 \times 0.938 \times 1.5mm^3$. Fortunately, we have all the images segmented by radiologist at the Brigham and Women's Hospital. In order to make full use of all the data, we choose the leave-one-out strategy to perform the shape learning and segmentation. That is, 23 out of 24 images as well as their segmented shapes are used for learning and the image left is segmented and the result is compared with its expert hand segmentation. The segmentation is fully-automatic, so that no human interaction is needed. On an Intel 3.0GHz quad-core machine with 8G memory, segmenting one volumetric image takes about 1 hour.

5.2.1 Hippocampus and Amygdala segmentation results

One important component of the proposed method is to first segment the hippocampus-amygdala complex as a whole, and then separate the two components. In doing so, we successfully solve the segmentation difficulty caused by the fact that the image intensity patterns in amygdala and hippocampus are extremely similar, if not identical. Therefore, in Figure 39 we present the segmented surfaces for both hippocampus and amygdala.

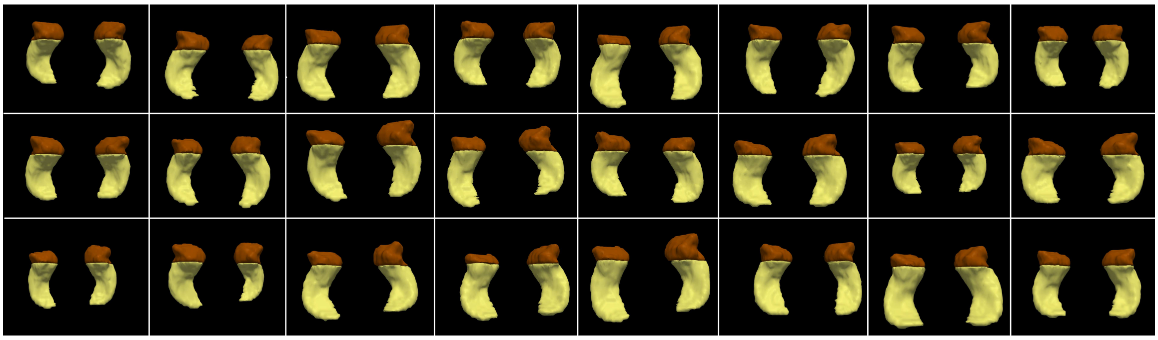


Figure 39: The segmented surfaces of both hippocampus and amygdala for all the 24 subjects.

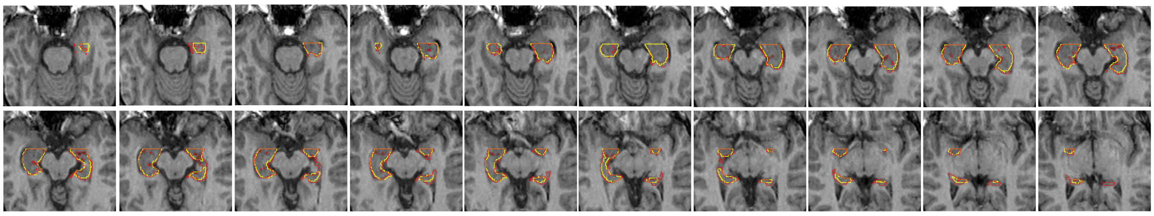


Figure 40: Axial slices of a segmentation result. The yellow contours are generated by the proposed automatic segmentation algorithm, while the red contours are drawn manually by an radiologist in BWH.

Moreover, it is noted that further qualitative and quantitative analysis in the subsequent sections are targeted only on the hippocampus since that is the main objective of present work.

5.2.2 Segmentation results shown by slices

From here on, we only perform the analysis on the hippocampus. In Figure 40 we present the segmentation of one randomly chosen volume.

In this result, the hippocampus boundary generated by the proposed automatic algorithm is colored in yellow whereas the manual boundary, drawn by an expert radiologist at the Brigham and Women's Hospital, is colored in red and used as the reference. As can be observed from the comparison, the two sets of contours match very well in all the slices. Due to space constraints, only one case is shown here. However, it is noted that all the 24 experiments work consistently, and the detailed

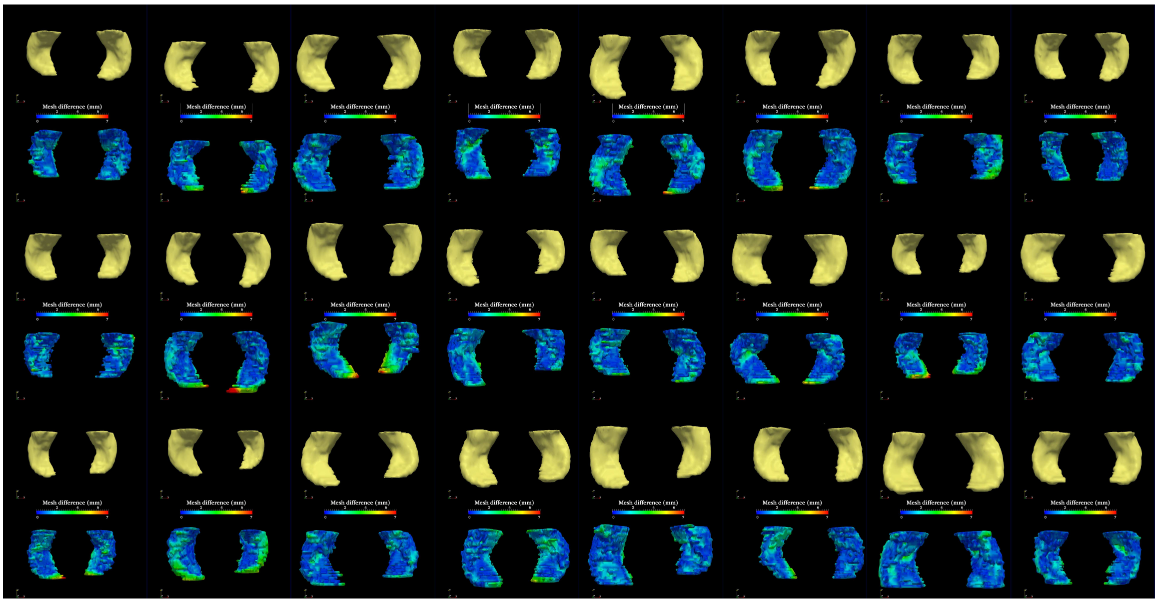


Figure 41: The surfaces of all the 24 testing hippocampi. The yellow ones are generated by the proposed method and the ones below are the corresponding manual results. On the manual surfaces, the color shows the difference between the automatic and manual surfaces.

statistical analysis is given below.

5.2.3 Distance on mesh

In order to demonstrate the performance of the algorithm in 3D, in this section we show the surfaces of the segmented hippocampi. In Figure 41, all the 24 hippocampal surfaces are shown.

From Figure 41, we can observe that the manual delineated hippocampal surfaces have many discontinuities. This is mainly due to the fact that the contours are traced out slice-by-slice. Thus, although in that process the image information in the adjacent slices is utilized, nevertheless, the resulting surfaces still have a jagged appearance. On the other hand, the proposed 3D segmentation method achieves better continuity and smoothness in all the three spatial dimensions, and the jagged appearance is avoided.

Moreover, in order to quantify the differences between the two sets of results, the difference maps are computed: Given any point on the manual surface, the difference map value is the closest distance to the automatically generated surface. Altogether, they form a scalar function defined on the manual surface. The difference maps are illustrated as the colors on the manual surfaces: Given any location on the surface, the red/blue color means that there exists a relatively large/small difference between the surface generated by expert and the algorithm, respectively. As can be seen in Figure 41, except at few isolated regions, the automatic results are very consistent with the expert manual results.

5.2.4 Further quantitative analysis

To further analyze the results quantitatively, similar to [45], the following parameters were measured: (1) left/right hippocampus volume; (2) volume ratio (automatic/manual) for the left and right hippocampus (ideal value is 1); (3) volume overlap ratio (VOR), i.e., the ratio of the volume of intersection to the volume of union (ideal value is 1); and (4) correctly delineated volume (CDV), i.e., the ratio of the volume of intersection to the manual volume (ideal value is 100%). In addition to these, the Hausdorff distance (HD), which represents the largest distance between the corresponding automatic and manual segmentation surfaces, is also computed. All the results are given in Table 2. It is seen that the algorithm performed correctly and consistently in all the cases.

5.3 Discussion

In the present work, we described a multi-scale shape based framework to automatically extract the hippocampus from MR brain imagery. First, we proposed a multi-scale representation for the shape using the wavelet transform. In doing so, we were able to take advantage of the multi-scale property of the wavelet representation. In particular, in learning the shapes, a small-scale shape variance was not overwhelmed by large ones. Moreover, in the segmentation step, we combined the

Table 2: Quantitative analysis for the segmentation results.

		Volume Manual(cm^3)	Volume Auto(cm^3)	Volume ratio[1]	VOR[2]	HD[3]	CDV[4]
Left	Mean	3.80	3.93	1.04	0.81	4.89	96%
Left	STD	0.64	0.64	0.10	0.048	1.31	4.3%
Right	Mean	3.59	3.78	1.05	0.82	4.97	88%
Right	STD	0.57	0.62	0.11	0.045	1.27	6.4%

[1] Volume ratio is computed as automatic/manual

[2] VOR: Volume overlap ratio is computed as the ratio of the volume of intersection to the volume of union. Optimal value is 1.0

[3] HD: Hausdorff distance is defined as the largest distance between the corresponding automatic and manual segmentation surfaces. Optimal value is 0.0

[4] CDV: Correctly delineated volume is defined as the ratio of the volume of intersection to the manual volume. Optimal value is 100%

atlas-based scheme and the active contour methods to provide an automatic and robust segmentation algorithm to extract the hippocampus. Finally, the algorithm was tested on 24 T1 MR brain images and the results were analyzed.

Further work includes investigating the number of training shapes needed for learning as well as the atlas-based initialization, in order to achieve optimal efficiency and accuracy. Moreover, in the data driven segmentation step, the local image mean intensity is used as the criteria in separating the hippocampus with its surrounding tissue. However, since the training images are available, one promising alternative would be learning the image information of hippocampus region in the training images, and use such information to drive the segmentation.

This methodology can now be incorporated into existing treatment planning approaches. In this manner, the objective of assessing the merits of whole brain irradiation with hippocampal avoidance can be pursued. With time, the neuro-oncologic community will be able to determine whether control of microscopic disease can be achieved without significant compromise of neurocognitive function.

APPENDIX A

PROOF OF THEOREM 1

Proof. The proof consists of two parts:

First we show the average error magnitude decreases monotonically at every scale. In fact, assume the affine transformation B optimally register point set $\mathscr{Q}$ to $\mathscr{P}$, i.e.

$$B = \arg\min_B \frac{1}{N} \sum_{q \in \mathscr{Q}} \|Cl(Bq) - Bq\|^2 \tag{62}$$

Furthermore, $\mathscr{Q}$ is decomposed to U_1 and U_2, and each of them is registered to $\mathscr{P}$ by starting from B. Then by the optimization process discussed in Section 2.2.1, it is guaranteed that the resulting two affine transformations B_1 and B_2 will yield smaller error:

$$\begin{aligned} &\sum_{q \in U_1} \|Cl(B_1 q) - B_1 q\|^2 + \sum_{q \in U_2} \|Cl(B_2 q) - B_2 q\|^2 \\ &\leq \sum_{q \in \mathscr{Q}} \|Cl(Bq) - Bq\|^2 \end{aligned} \tag{63}$$

Second, we claim the average error magnitude approaches zero. In fact, as the decomposition-registration proceeds, the following limit case can always be reached in finite time: $\mathscr{Q}$ is decomposed into N sub-point-sets, each of which contains one point. Then there exists N affine (in fact translation is enough) transformations such that all the errors go to zero. □

REFERENCES

[1] Ait-Aoudia, S. and Mahiou, R., “Medical Image Registration by Simulated Annealing and genetic algorithms,” *Geometric Modeling and Imaging–New Trends*, pp. 145–148, 2007.

[2] Arsigny, V., Commowick, O., Ayache, N., and Pennec, X., “A fast and log-Euclidean polyaffine framework for locally linear registration,” *J. of Math. Imaging and Vision (JMIV)*, vol. 33, no. 2, pp. 222–238, 2009.

[3] Besl, P. J. and McKay, N. D., “A Method for Registration of 3-D Shapes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[4] Bookstein, F., “Linear methods for nonlinear maps: Procrustes fits, thin-plate splines, and the biometric analysis of shape variability,” *Brain Warping*, pp. 157–181, 1998.

[5] Brox, T. and Weickert, J., “Level set segmentation with multiple regions,” *IEEE TIP*, vol. 15, no. 10, p. 3213, 2006.

[6] Caselles, V., Kimmel, R., and Sapiro, G., “Geodesic active contours,” *International journal of computer vision*, vol. 22, no. 1, pp. 61–79, 1997.

[7] Chan, T., Sandberg, B., and Vese, L., “Active contours without edges for vector-valued images,” *Journal of Visual Communication and Image Representation*, vol. 11, no. 2, pp. 130–141, 2000.

[8] Chan, T. and Vese, L., “Active contours without edges,” *IEEE TIP*, vol. 10, no. 2, pp. 266–277, 2001.

[9] Chui, H. and Rangarajan, A., “A new point matching algorithm for non-rigid registration,” *Computer Vision and Image Understanding*, vol. 89, no. 2-3, pp. 114–141, 2003.

[10] Cootes, T., Taylor, C., Cooper, D., Graham, J., and others, “Active shape models-their training and application,” *Computer vision and image understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[11] Cootes, T., Taylor, C., Cooper, D., Graham, J., and others, “Active Shape Models-Their Training and Application,” *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995.

[12] Devroye, L., *Non-uniform random variate generation.* Springer-Verlag New York, 1986.

[13] Donoho, D. and others, "De-noising by soft-thresholding," *IEEE transactions on information theory*, vol. 41, no. 3, pp. 613–627, 1995.

[14] Dryden, I. and Mardia, K., *Statistical shape analysis.* Wiley New York, 1998.

[15] Duda, R., Hart, P., and Stork, D., *Pattern Classification.* Wiley New York, 2001.

[16] Fitzgibbon, A. W., "Robust registration of 2D and 3D point sets," *Image Vision and Computing*, vol. 21, no. 13-14, pp. 1145–1153, 2003.

[17] Flusser, J. and Suk, T., "A moment-based approach to registration of images with affinegeometric distortion," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 2, pp. 382–387, 1994.

[18] Golub, G. and Van Loan, C., *Matrix computations.* Johns Hopkins Univ Pr, 1996.

[19] Grady, L., "Random walks for image segmentation," *IEEE TPAMI*, vol. 28, no. 11, p. 1768, 2006.

[20] Haber, E. and Modersitzki, J., "Intensity gradient based registration and fusion of multi-modal images," *Methods of Information in Medicine*, vol. 46, no. 3, pp. 292–299, 2007.

[21] Haker, S., Sapiro, G., and Tannenbaum, A., "Knowledge-based Segmentation of SAR Data with Learned Priors," *IEEE Transactions on Image Processing*, vol. 9, no. 2, pp. 299–301, 2000.

[22] Haker, S., Zhu, L., Tannenbaum, A., and Angenent, S., "Optimal mass transport for registration and warping," *International Journal of Computer Vision*, vol. 60, no. 3, pp. 225–240, 2004.

[23] Hill, D., Hawkes, D., Crossman, J., Gleeson, M., Cox, T., Bracey, E., Strong, A., and Graves, P., "Registration of MR and CT Images for Skull Base Surgery Using Point-like Anatomical Features," *British journal of radiology*, vol. 64, no. 767, pp. 1030–1035, 1991.

[24] Huber, P. and Ronchetti, E., *Robust statistics.* Wiley-Blackwell, 2009.

[25] Ibanez, L. and others, *The ITK software guide.* Kitware, 2003.

[26] Jenkinson, M. and Smith, S., "A global optimisation method for robust affine registration of brain images," *Medical Image Analysis*, vol. 5, no. 2, pp. 143–156, 2001.

[27] Kao, C., Osher, S., and Qian, J., "Lax-Friedrichs Sweeping Scheme for Static Hamilton-Jacobi Equations," *Journal of Computational Physics*, vol. 196, no. 1, pp. 367–391, 2004.

[28] Kass, M., Witkin, A., and Terzopoulos, D., "Snakes: Active contour models," *International journal of computer vision*, vol. 1, no. 4, pp. 321–331, 1988.

[29] Kennel, M., "KDTREE 2: Fortran 95 and C++ software to efficiently search for near neighbors in a multi-dimensional Euclidean space," *Arxiv preprint physics/0408067*, 2004.

[30] Kichenassamy, S., Kumar, A., Olver, P., Tannenbaum, A., and Yezzi, A., "Gradient flows and geometric active contour models," in *Proceedings of the Fifth International Conference on Computer Vision*, p. 810, IEEE Computer Society Washington, DC, USA, 1995.

[31] Krieger, A., Susil, R., Menard, C., Coleman, J., Fichtinger, G., Atalar, E., and Whitcomb, L., "Design of a novel MRI compatible manipulator for image guided prostate interventions," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 2, pp. 306–313, 2005.

[32] Kubicki, M., Park, H., Westin, C., Nestor, P., Mulkern, R., Maier, S., Niznikiewicz, M., Connor, E., Levitt, J., Frumin, M., and others, "DTI and MTR abnormalities in schizophrenia: analysis of white matter integrity," *Neuroimage*, vol. 26, no. 4, pp. 1109–1118, 2005.

[33] Lankton, S. and Tannenbaum, A., "Localizing region-based active contours," *IEEE Transactions on Image Processing*, vol. 17, no. 11, pp. 2029–2039, 2008.

[34] Leventon, M., Grimson, W., and Faugeras, O., "Statistical Shape Influence in Geodesic Active Contours," in *IEEE CVPR*, vol. 1, 2000.

[35] Ma, B. and Ellis, R., "Surface-based registration with a particle filter," *Lecture Notes in Computer Science*, pp. 566–573, 2004.

[36] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P., "Multimodality image registration by maximization of mutualinformation," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.

[37] Malladi, R., Sethian, J., and Vemuri, B., "Shape modeling with front propagation: A level set approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 158–175, 1995.

[38] Melonakos, J. and Pichon, E. and Angenent, S. and Tannenbaum, A., "Finsler Active Contours," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 412–423, 2008.

[39] Michailovich, O., Rathi, Y., and Tannenbaum, A., "Image segmentation using active contours driven by the bhattacharyya gradient flow," *IEEE Transactions on Image Processing*, vol. 16, no. 11, pp. 2787–2801, 2007.

[40] Moghari, M. and Abolmaesumi, P., "Point-based rigid-body registration using an unscented kalman filter," *IEEE Transactions on Medical Imaging*, vol. 26, no. 12, pp. 1708–1728, 2007.

[41] Nocedal, J. and Wright, S., *Numerical Optimization.* Springer, 1999.

[42] Osher, S. and Sethian, J., "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations," *Journal of computational physics*, vol. 79, no. 1, pp. 12–49, 1988.

[43] Palmer, S., *Vision science: Photons to phenomenology.* MIT press, 1999.

[44] Palmer, S., *Vision science: Photons to phenomenology.* MIT press, 1999.

[45] Pasquier, D., Lacornerie, T., Vermandel, M., Rousseau, J., Lartigau, E., and Betrouni, N., "Automatic segmentation of pelvic structures from magnetic resonance images for prostate cancer radiotherapy," *International journal of radiation oncology, biology, physics*, vol. 68, no. 2, pp. 592–600, 2007.

[46] Pichon, E., *Novel Methods for Multidimensional Image Segmentation.* PhD thesis, Georgia Institute of Technology, 2005.

[47] Pichon, E., Tannenbaum, A., and Kikinis, R., "A statistically based flow for image segmentation," *Medical image analysis*, vol. 8, no. 3, pp. 267–274, 2004.

[48] Pichon, E., Westin, C., and Tannenbaum, A., "A Hamilton-Jacobi-Bellman Approach to High Angular Resolution Diffusion Tractography," in *MICCAI*, pp. 180–187, 2005.

[49] Pratt, W. and others, *Digital image processing.* Wiley-Interscience, 2007.

[50] Rohlfing, T., Brandt, R., Menzel, R., Russakoff, D., and Maurer, C., "Quo vadis, atlas-based segmentation?," *Handbook of Biomedical Image Analysis*, pp. 435–486, 2005.

[51] Rohr, K., Stiehl, H., Sprengel, R., Beil, W., Buzug, T., Weese, J., and Kuhn, M., "Point-based elastic registration of medical image data using approximating thin-plate splines," *Lecture Notes in Computer Science*, vol. 1131, pp. 297–306, 1996.

[52] Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., and Hawkes, D., "Nonrigid Registration Using Free-form Deformations: Application Tobreast MR Images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.

[53] Salmond, D., Gordon, N., and Smith, A., "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," in *IEE Proc. F, Radar and signal processing*, vol. 140, pp. 107–113, 1993.

[54] Sandhu, R., Dambreville, S., and Tannenbaum, A., "Particle filtering for registration of 2D and 3D point sets with stochastic dynamics," in *IEEE CVPR*, pp. 1–8, 2008.

[55] Shi, J. and Malik, J., "Normalized cuts and image segmentation," *IEEE TPAMI*, vol. 22, no. 8, pp. 888–905, 2000.

[56] Sonka, M., Hlavac, V., and Boyle, R., *Image Processing, Analysis, and Machine Vision Second Edition.* International Thomson, 1999.

[57] Thirion, J., "Image Matching as a Diffusion Process: an Analogy with Maxwell's Demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, 1998.

[58] Toga, A., *Brain warping.* Academic Press, 1999.

[59] Tsai, A., Yezzi Jr, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W., and Willsky, A., "A Shape-based Approach to The Segmentation of Medical Imagery Using Level Sets," *IEEE Transactions on Medical Imaging*, vol. 22, no. 2, pp. 137–154, 2003.

[60] Vese, L. and Chan, T., "A multiphase level set framework for image segmentation using the Mumford and Shah model," *IJCV*, vol. 50, no. 3, pp. 271–293, 2002.

[61] Viola, P. and Wells III, W., "Alignment by maximization of mutual information," *International Journal of Computer Vision*, vol. 24, no. 2, pp. 137–154, 1997.

[62] Vujovic, N. and Brzakovic, D., "Establishing the correspondence between control points in pairs of mammographic images," *IEEE Transactions on Image Processing*, vol. 6, no. 10, pp. 1388–1399, 1997.

[63] Wang, J. and Jiang, T., "Nonrigid Registration of Brain MRI Using NURBS," *Pattern Recognition Letters*, vol. 28, no. 2, pp. 214–223, 2007.

[64] Wang, W. and Chen, Y., "Image registration by control points pairing using the invariant properties of line segments," *Pattern Recognition Letters*, vol. 18, no. 3, pp. 269–281, 1997.

[65] Wells, W., Viola, P., Atsumi, H., Nakajima, S., and Kikinis, R., "Multimodal volume registration by maximization of mutual information," *Medical Image Analysis*, vol. 1, no. 1, pp. 35–51, 1996.

[66] Whitaker, R., "A level-set approach to 3D reconstruction from range data," *IJCV*, vol. 29, no. 3, p. 231, 1998.

[67] Yang, Y., Stillman, A., Tannenbaum, A., and Giddens, D., "Automatic Segmentation of Coronary Arteries Using Bayesian Driven Implicit Surfaces," *IEEE ISBI*, 2007.

[68] Yang, Y., Tannenbaum, A., Giddens, D., and Coulter, W., "Knowledge-based 3D segmentation and reconstruction of coronary arteries using CT images," in *IEEE EMBS*, pp. 1664–1666, 2004.

[69] Yezzi Jr, A., Kichenassamy, S., Kumar, A., Olver, P., and Tannenbaum, A., "A geometric snake model for segmentation of medical imagery," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 199–209, 1997.

[70] Yezzi Jr, A., Tsai, A., and Willsky, A., "A Statistical Approach to Snakes for Bimodal and Trimodal Imagery," in *Proceedings of the International Conference on Computer Vision-Volume 2-Volume 2*, IEEE Computer Society Washington, DC, USA, 1999.

[71] Yezzi Jr, A., Tsai, A., and Willsky, A., "A Statistical Approach to Snakes for Bimodal and Trimodal Imagery," in *ICCV*, p. 898, 1999.

[72] Zhao, H., Chan, T., Merriman, B., and Osher, S., "A Variational Level Set Approach To Multiphase Motion," *Journal of computational physics*, vol. 127, no. 1, pp. 179–195, 1996.

[73] Zhu, S. and Yuille, A., "Region competition: Unifying snakes, region growing, and bayes/MDL for multiband image segmentation," *IEEE TPAMI*, vol. 18, no. 9, pp. 884–900, 1996.

VITA

Yi Gao was born in Xi'an, Shaanxi Province in China on May 24th, 1981. He received a Bachelor of Science in Biomedical Engineering in 2003 from Tsinghua University, Beijing, China. After he received a Master of Science in Biomedical Engineering in 2005 in the same university, he started his doctorate research at Georgia Institute of Technology. At Georgia Tech, he also received a Master degree of Science in Mathematics in 2008. He has been a graduate research assistant in the Minerva research lab and the Cardiovascular lab at Georgia Tech since 2005 and his doctorate research has focused on computer vision and medical image analysis.