TOWARDS ROBUST AND EFFECTIVE SHAPE PRIOR MODELING: SPARSE SHAPE COMPOSITION

BY SHAOTING ZHANG

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Computer Science

Written under the direction of

Professor Dimitris N. Metaxas

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

JANUARY, 2012

ABSTRACT OF THE DISSERTATION

Towards Robust and Effective Shape Prior Modeling:
Sparse Shape Composition

by Shaoting Zhang

Dissertation Director: Professor Dimitris N. Metaxas

Organ shape plays an important role in many clinical practices, including diagnosis, surgical planning and treatment evaluation. It is usually derived from medical images using low level appearance cues. However, due to diseases and imaging artifacts, low level appearance cues are often weak or misleading. In this situation, shape priors become critical to infer and refine the shape derived from image appearances. Effective modeling of shape priors is challenging because: 1) shape variations are complex and cannot always be modeled by parametric probability distributions; 2) a shape instance derived from image appearance cues (called an input shape) may have significant errors; and 3) local details of an input shape may be important for clinical purposes but difficult to preserve if they are not statistically significant in the training data.

In this paper we propose a novel Sparse Shape Composition model (SSC) to address these three challenges in a unified framework. With our method, a sparse set of shapes is selected from the shape repository and composed together to infer and refine an input shape. This way, the prior information is implicitly incorporated on-the-fly. Our model leverages two sparsity observations of the input shape instance: 1) the input shape can be approximately represented by a sparse linear combination of shapes in the shape repository; 2) parts of the input shape may contain large errors but such errors

are sparse. Our model is formulated as a sparse learning problem. Using $L1$ norm relaxation, it can be solved by an efficient expectation-maximization (EM) framework. Furthermore, this model is extended to effectively handle multi-resolution, local shape priors and hierarchical priors. We also propose a framework to generate high quality training data in 3D. Our framework includes geometry processing methods and shape registration algorithms.

The proposed shape prior model is extensively validated on five different medical applications: 2D lung localization in chest X-ray images, 3D liver segmentation in low-dose Computed Tomography (CT) scans, 3D segmentation of multiple rodent brain structures in Magnetic Resonance (MR) microscope, real time tracking of left ventricles in Magnetic Resonance Imaging (MRI), and high resolution CT reconstruction. Compared to state-of-the-art methods, our model exhibits better performance in all these studies.

Acknowledgements

I am grateful to Professor Dimitris Metaxas for his advice and support during my PhD study. He has been motivating and encouraging me to solve fundamental problems in Medical Image Analysis. He also gives me opportunities and freedoms to do research on diverse topics such as Computer Vision, Computer Graphics and Machine Learning. Thus I am able to propose this shape prior method and create the whole system by using techniques from all these fields. None of the work in this thesis would have happened without him.

I would like to thank other committee members: Prof. Vladimir Pavlovic, Prof. Tina Eliassi-Rad, Prof. Ron Kikinis (Harvard Medical School) for their advice, help and valuable suggestions regarding this thesis. It is an honor for me to have each of them serves in my committee.

I also thank Dr. Yiqiang Zhan (Siemens Healthcare) and Dr. Ming Yang (NEC labs), who have been wonderful mentors to me. They have contributed many ideas and helps in my research work. They also showed me the excellence of industry and research labs when I was doing summer intern with them in Pennsylvania and California.

Thanks to many professors at Rutgers that helped me in many aspects, especially to Prof. Doug Decarlo, whom I worked with as a TA of Computer Graphics, Prof. Andrew Nealen, Prof. Ahmed Elgammal, Prof. Eric Allender, and Prof. William Steiger. Also thanks to other professors and researchers, especially to Prof. Leon Axel (New York University), Prof. Xiaolei Huang (Lehigh University), Prof. Junzhou Huang (The University of Texas at Arlington) and Dr. Panayotis Thanos (Brookhaven National Lab).

Special thanks to my friends and colleagues. I benefited a lot from their friendship and help, and they made my years at Rutgers a real pleasure.

Dedication

This dissertation is dedicated to my parents: Jianguo Zhang and Jianhua Li.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Shape Prior Model for Medical Applications

Figure 1.1: Middle: whole body low-dose CT data. Left: zoom in of the liver and lung. In the marked region the boundary between the liver and the kidney can hardly observed. The appearance cue is weak because of the low contrast around the boundary. Right: zoom in of the spleen and lung. In the marked region, there is artifact induced by breath. It is part of the lung but not the spleen. Since the image information is misleading here, segmentation methods solely relying on appearance cue may accidentally include this region as spleen.

Shape is a distinctive characteristic of many human organs and plays a critical role in various medical image analysis tasks, such as segmentation, registration, tracking, etc. Although image appearance cues provide low-level evidence to derive organ shapes, the derived shape instances may be incomplete in the presence of weak (missing) appearance cues, or incorrect when misleading appearance cues are present. Fig. 1.1 shows a 3D image data from whole body CT with low dose and large slice thickness, which result in low contrast and fuzzy boundaries between organs. The boundary information
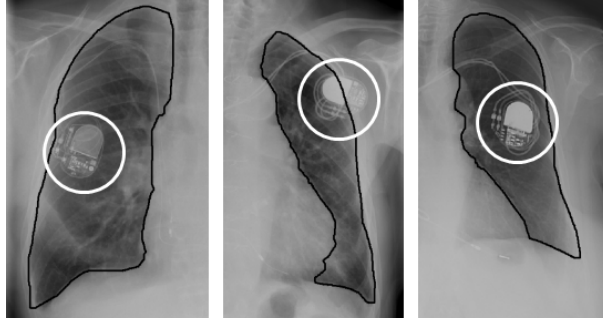
Figure 1.2: Chest X-ray images with annotated boundaries. The appearance cue is misleading because of the instruments in the marked regions.

is weak in between the liver and the kidney (Fig. 1.1 left). Furthermore, there is motion artifact induced by breath around the boundary between the spleen and the lung (Fig. 1.1 right). When segmenting the liver or spleen, segmentation methods solely relying on appearance cue may accidentally include parts of kidney or lung since the image information is misleading here. Such inaccurate segmentation may adversely affect the clinical analysis or the performance of Computer-aided diagnosis (CAD) programs. If there is a shape constraint that is independent of the appearance cues, it can prevent this over-segmentation and improve the segmentation accuracy.

Fig. 1.2 shows some challenging cases of the most widely used imaging modality, chest X-ray. Many scans are taken from X-ray, especially in developing countries. Thus such modality is extensively studied and analyzed. In the marked regions of Fig. 1.2, lung boundaries become broken due to the medical instruments. In these scenarios, segmentation methods solely relying on appearance cue may fail because of the missing or "false" boundaries. With shape prior information, the shapes correctly derived from clear boundaries is used as a complement for the information from unclear parts. Thus the erroneous segmentation can be partially corrected.

On the other hand, human organs usually contain strong shape priors as many organs have similar shape patterns. It provides opportunities to use shape models to infer and refine the organ shape in an optimal sense based on high-level shape priors. The success of these models is highly dependent on the way shape priors are modeled and on the optimization method used. One of the seminal work in this area, "Snakes" [48], models the shape prior as a general regularity term in the optimization, which assumes
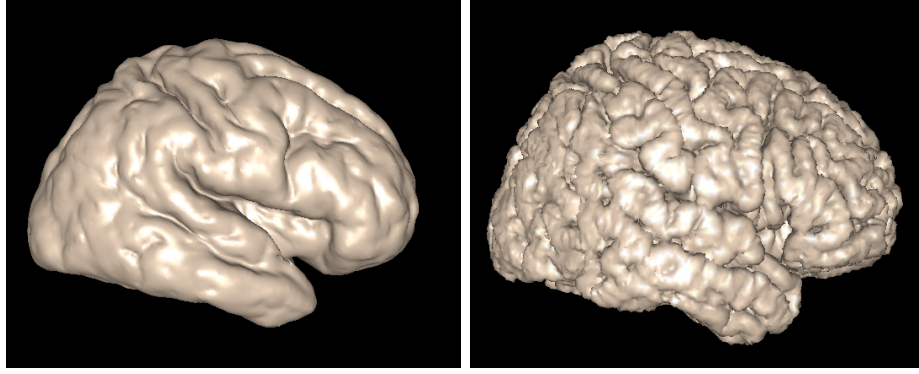
Figure 1.3: Left: The result of deformable model based segmentation using a large smoothness constraint. Most details are smoothed out. Right: The result from the same model using a small smoothness constraint. Some image noise is also included.

that the shape should deform smoothly like a membrane or a thin plate. Such regularity constraint is referred as *internal force*. The external force is derived from image appearance cues. Many variations have been proposed to improve the robustness of Snakes model. Most of them focus on handling noise and spurious image edges, since the traditional Snakes solely relying on the image gradient information can easily stuck at local minima. Region analysis strategies [103] have been incorporated in Snake-like models to improve their robustness to noise. Metamorphs method [47] was proposed to integrate shape and appearance in a unified space. The model has not only boundary information but also interior appearance, making it more robust to ambiguous boundaries and complex internal textures. Its 3D version, Active Volume Model (AVM) [75, 74], is proposed to perform volume segmentation. However, these methods still use the same shape prior as the internal force, the smoothness constraint. This type of constraint actually introduces a tradeoff between accuracy and robustness, especially when organ shapes have sharp corners or fine details. Choosing large smoothness constraint may smooth out details, while small constraint can result in erroneous segmentation as the model may include image noise or other organs. Thus it is often necessary to carefully tune parameters for different data, which is laborious. Fig. 1.3 shows the brain segmentation result of deformable models using different smoothness constraint [73]. It shows the limitation of this type of shape priors.

Subsequently, more object-specific shape priors became prevalent, where the shape
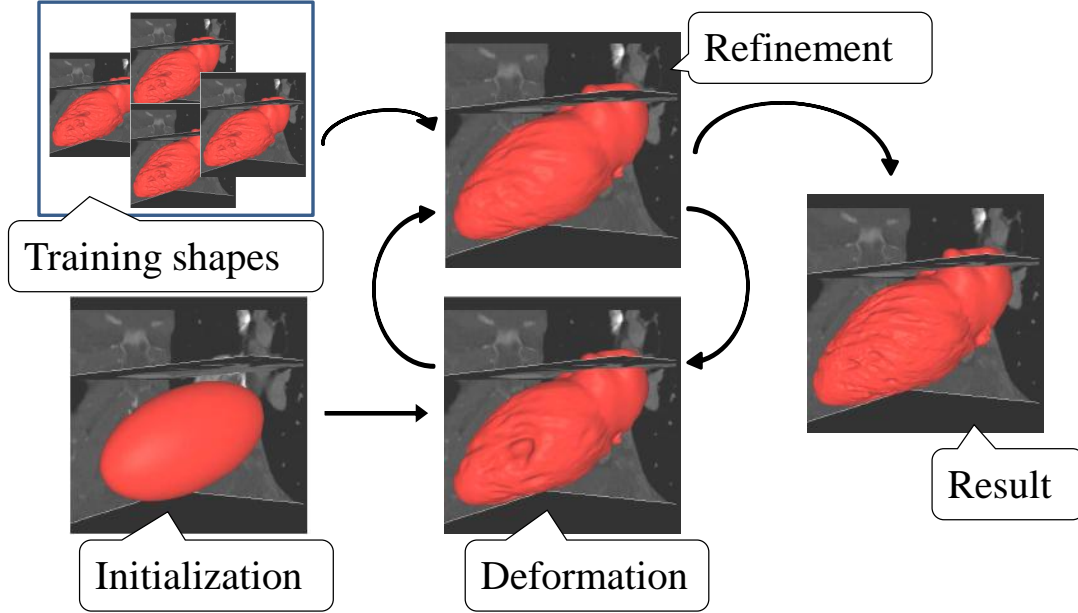
Figure 1.4: A top-down framework example of 3D cardiac segmentation. The input is a cardiac image data with a mesh initialization [75, 76]. The output is a derived shape as the segmentation result. In the processing steps, it derives an intermediate shape by deforming using appearance cues, and then refines it using shape prior constraint, such as PDM in ASM in this case. Two steps are alternately employed until accurately finding the object boundary.

priors are learned from a set of training samples, such as Active Shape Model (ASM) [12] and Level Set representation with priors [69]. ASM method has described Point Distribution Models (PDMs), which are statistical models of shape constructed from training dataset. A PDM represents an object as a set of labeled points, giving their mean positions and small set of modes of variations which describe how the object's shape can change. Principal Component Analysis (PCA) is applied to learn the PDM. Given an input shape derived from image appearance cues, it is projected to PCA space and is constrained by following PDM's major variations. Thus the refined shape has to be consistent with patterns in the training data. This type of shape priors is less sensitive to parameters than Smoothness constraint does, because it is generalized from the training data and is not specific to any of them. After this shape refinement, we can further deform this intermediate result using image appearance cues. This procedure is usually iterated several times to accurately converge to the boundary. Such approach is often referred as top-down framework (Fig. 1.4) and is widely used in computer vision
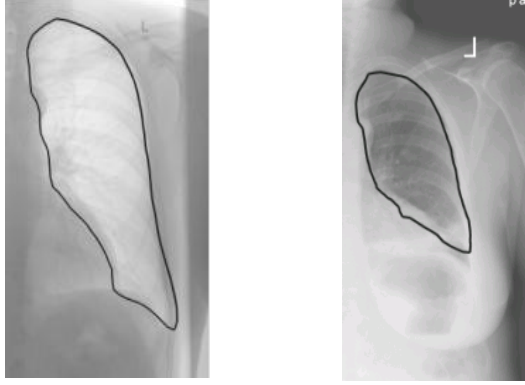
Figure 1.5: Left: A typical shape of the left lung. It is very similar to the mean shape. Right: This is also a valid shape of the left lung. However, a large part of the bottom shape is blocked by the cardiac. Thus it cannot be represented by the mean shape with some major variations.

and medical image applications. Many adaptations of this PDM type of shape priors have been proposed over the years and some of them are successfully applied in tasks of medical image analysis, especially 2D and 3D segmentation (see Sec. 2.1 for details).

1.2 Challenges of Shape Prior Modeling

PDM based shape priors are generally robust and improve the accuracy of many image analysis tasks. However, these shape prior modeling methods often confront the following three challenges. First, shape variations are usually complex and therefore difficult to model using a parametric probability distribution. The traditional ASM assumes that each shape can be represented by a mean shape plus the variations. In many real world applications, however, this assumption may not be true. Fig. 1.5 shows an application of lung segmentation. We are interested in segmenting lung regions without including the cardiac. Thus the result can be analyzed automatically by our CAD program. The left figure shows a typical shape of the left lung, which is similar to the mean shape computed from a large dataset. The right figure, however, has a special shape which rarely happens in the database. A large part of its bottom is blocked by the cardiac. Thus it cannot be represented by the mean shape with some major variations. More sophisticated approaches are needed to handle these multi-modal distribution cases.
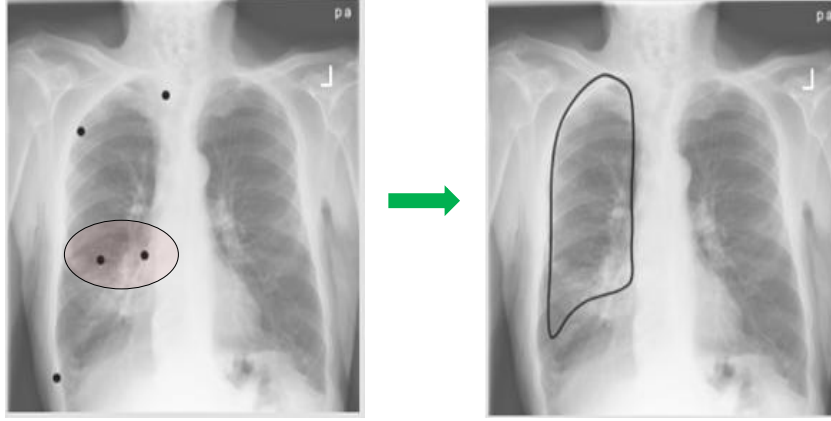
Figure 1.6: Left: Detection results with some gross errors. Black dots are anatomical landmarks automatically found by detectors. Red circle marks erroneous detections. They should be on the bottom. Right: Procrustes analysis based shape alignment using landmarks as input. The whole shape is shifted to the top.

Second, image appearance information can be highly misleading and non-Gaussian errors frequently appear in the input shape. Many shape prior models, such as ASM, can easily handle Gaussian distributed errors. However, they may fail to deal with gross errors in the input shape data. Fig. 1.6 shows the effects of detection errors in the lung segmentation application. Since some of the image information is misleading, the detector locates wrong landmarks in the middle of the lung. Using traditional types of shape priors may result in erroneous segmentation, which is shifted to upper region. Since Non-Gaussian errors severely harm the system performance when using traditional shape priors, shape models have to be robust to handle these errors.

Third, shape models should be adaptive in order to preserve local detail information in the input shape, provided such details are present in the training data, even if they are not statistically significant. For example, traditional ASM relies on PCA to learn shape statistics. It well captures major shape variations, such as size or volume changing. However, some shape details can be easily ignored or removed if they are not prevalent in the training data. Thus when refining an input data using this shape prior model, shape detail may be removed since the major variations cannot describe it. Fig. 1.7 shows an example of shape detail in the lung application. Since tip on the left bottom is not statistically significant in the training data, many shape prior methods could not preserve such information, as shown in the experiment section (Sec. 5.1).
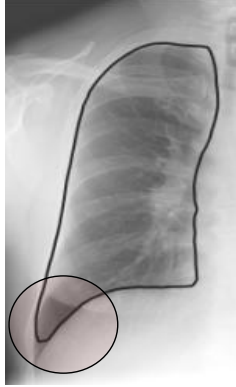
Figure 1.7: A right lung shape with detail information marked by the red circle. This shape detail may be removed during shape refinement procedure since it is not statistically significant in the training data.

While several methods have been proposed to address one or two of those challenges (Sec. 2.1), it remains that, to the best of our knowledge, none of them can tackle all three challenges simultaneously. The main purpose of this thesis is to propose a unified framework to handle all three challenges together. These cases in Fig. 1.5, Fig. 1.6 and Fig. 1.7 are further analyzed in the experiment section (Sec. 5.1).

1.3 Our Solutions – Sparse Shape Composition

In this thesis, we propose a novel Sparse Shape Composition model (SSC) [97, 98] to address the aforementioned challenges in a unified framework. Given a shape repository consists of a number of annotated shape instances, instead of explicitly learning shape priors from these shapes offline, we propose to adaptively approximate the input shape on-the-fly, by a sparse linear combination of a subset of shapes in the shape repository. Hence, the shape prior constraint is implicitly applied. Our method is inspired by recently proposed sparsity theories in the compressive sensing community, i.e., the problem of computing sparse linear representations with respect to an overcomplete dictionary of base elements [6, 17]. It has been successfully applied in many computer vision applications, such as, but not limited to, robust face recognition [84] and image restoration [59]. Yet, to our knowledge, such techniques have not been used in the context of shape priors.

There are two "sparsity" observations behind our method. First, given a large enough training dataset (repository), an instance can be approximately represented by a sparse linear combination of instances in the shape repository. Similarly, in our application each given shape is approximated by a sparse linear combination of annotated shapes. Without any assumption of a parametric distribution model (e.g., a unimodal distribution assumption in ASM), it becomes general to objects whose shape statistics can be very complex. Moreover, such a setting is able to recover detail information even if the detail of the input shape is only present in a small proportion of the shape repository and is not statistically significant. Second, the given shape information may contain gross errors, but such errors are often very sparse, e.g., there is an object occluded in the image or a point missing in the input shape. Combining these two, we formulate the shape prior task as a sparse learning problem, and efficiently solve it by an expectation-maximization (EM) type of framework. Furthermore, we explicitly model the nonlinear shape transformation in the optimization framework without assuming that the initial misalignment is small, which is different from the sparse learning method in face recognition [84].

We also introduce two extensions to improve this SSC model. First, multi-resolution and mesh partitioning are proposed to model local shape priors and improve computational efficiency [96]. As a 3D mesh may contain thousands of vertices and local shape properties, it is desirable to use divide-and-conquer approaches to solve each subproblem (i.e., local shape patch) and then integrate them together. Since each shape prior model is responsible for only one shape patch, it is relatively fast to solve and can properly deal with local information. Second, dictionary learning is employed to learn a compact codebook for acceleration purpose. When training data is large, it may not be feasible or necessary to use all of them in the same time, since it can be very slow and the information may be highly redundant. Thus, it is desirable to learn a small-size dictionary without significantly losing information.

Our method is independent of data dimensions, i.e., it works for both 2D contours and 3D meshes. However, in 3D cases, the quality of training meshes becomes crucial to the performance of both shape prior methods and segmentation algorithms. Thus

a framework is proposed to reconstruct 3D shape atlas from image data [94]. Shape registration is used to obtain one-to-one correspondence for vertices. Detail preserved smoothing is also incorporated into the shape registration framework to obtain high quality meshes without significantly losing shape detail information. These registered shapes are used as training data for the proposed method, which furthermore improve its performance and robustness.

In the experiments, we validate our method on several very diverse applications of medical image analysis, and the proposed SSC shows improved accuracy and robustness compared to some widely used approaches.

1.4   Main Contributions

The main contributions of this thesis are fourfold:

1. SSC is proposed to model shapes and implicitly incorporate the shape prior constraint effectively. It is based on sparse representations and our unified framework is able to handle non-Gaussian errors, multimodal distributions of shapes and detail information recovery, which are three main challenges of traditional shape prior modeling methods.

2. Since we do not assume that the initial misalignment of the input shape is small, this problem becomes difficult to solve as it is not convex when the transformation is rigid or similarity. We have proposed an EM type of optimization framework to efficiently solve it. We also improve the computational efficiency by introducing multi-resolution, mesh partitioning, and dictionary learning techniques.

3. The SSC model is independent of data dimensions. However, when the data is in 3D, the quality of training meshes becomes crucial to the performance of both SSC and deformable model based segmentation. We have introduced a framework to effectively compute such meshes. It includes geometry processing algorithms and shape registration method. High quality meshes are computed and one-to-one correspondence is also obtained.
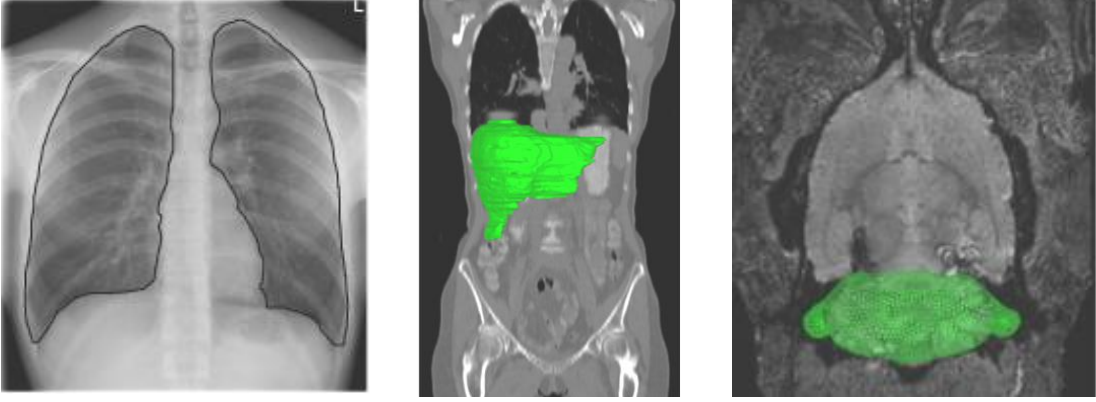
Figure 1.8: Several image modalities used in the experiments. From left to right: Chest X-ray, whole body low-dose CT, the structure visualized the cerebellum of the rat brain in MR microscope.

4. It is successfully applied to several very diverse medical applications, 2D lung localization from X-ray image, 3D liver segmentation from low-dose CT, 3D segmentation of multiple rodent brain structures from MR microscope, left ventricle tracking from cardiac MRI, and shape atlas reconstruction from high resolution CT images. Most of them are very important for clinical analysis and are challenging. Fig. 1.8 shows several image modalities used in our experiments. These extensive experiments demonstrate the superior performance of our method.

1.5   Organization

The remainder of this thesis is organized as follows. Chapter 2 reviews the relevant work both in shape prior modeling (Sec. 2.1) and also briefly introduce sparse learning methods (Sec. 2.2), as our work is to employ sparse methods to create shape priors. In the review of shape prior methods, we categorize them as per the improvements upon three challenges. Thus the advantages and limitations of existing methods can be well observed.

Chapter 3 introduces the proposed SSC model in detail, including the problem formulation and optimization framework. Several extensions have also been presented, such as modeling local shape priors using shape partitioning, and learning a compact dictionary using sparsity methods. Since the SSC framework and these extensions will

be employed to solve several medical applications, we will directly use examples from these applications as motivation or illustration.

Same as ASM, SSC model also needs training data. In 3D, high quality training data is crucial to the performance of both shape prior modeling and segmentation. Thus Chapter 4 presents a framework to obtain such 3D training data. It includes geometry processing and shape registration methods, such as detail preserved smoothing, isotropic remeshing, and discovering one-to-one correspondence.

Chapter 5 validates the proposed method by applying it to five very diverse medical applications: 1) 2D lung location from X-ray images, 2) 3D liver segmentation from low-dose CT, 3) 3D segmentation of multiple rodent brain structures from MR microscope, 4) tracking of left ventricle from cardiac MRI, and 5) cardiac atlas reconstruction from high resolution CT images. Finally we conclude and discuss the future work in Chapter 6.

Chapter 2

Relevant Work

The proposed SSC model employs sparsity constraints to robustly model shape priors. Thus in this chapter we review both shape prior modeling methods and relevant work of sparsity algorithms.

2.1 Shape Prior Modeling

As discussed earlier, many approaches have been proposed in different contexts to incorporate the shape prior constraint. In the context of medical image analysis, ASM [12] and its variations [38] are probably the most widely used approaches. Many methods have been proposed to improve the shape prior module of ASM. They mainly focus on three aspects:

2.1.1 Modeling Complex Shape Variations

A significant effort has been put on handling multimodal distribution of shapes, which cannot be represented by their mean shape and variations. A classical solution is to use a mixture of Gaussians to represent 2D shape variation [13]. Its main advantage is able to model more complex variations, including those cases where there are two or more distinct classes of shape variation. However, one problem with the method is that more examples are required to obtain reliable parameters for a mixture model than for a single Gaussian. Manifold learning techniques can also be used to learn a non-linear shape prior to alleviate this problem. For example, Etyngier et.al. [23] have introduced a new deformable model framework that integrates general non-linear shape priors using Diffusion maps. A new projection operator is presented to project it onto a manifold based on the Nystrom extension and a Delaunay partitioning of the

reduced space. Then a variational solution is provided for manifold denoising. Finally, a new energy term is designed to attract a shape towards the manifold at given constant embedding. Zhang et.al. [99] have also employed manifold learning method to overcome the limitation of ASM on statistical constraint.

However, it is still possible that shape variation is too complex to model with any parametric probability distribution. Thus shape prior models are desirable to be able to model any specific shape. Phol et.al. [66] have proposed to couple the PCA based shape modeling with a maximum a posteriori estimation problem which will be solved through an Expectation Maximization framework. This allows the system to accommodate shapes that differ somewhat from those modeled by the PCA. Thus it is not restricted to the modes of variations presented in the shape model but models patient specific abnormalities. Some methods were also proposed to decompose the shape space into multiple sub-spaces. Representative studies include patient-specific shape statistics [77, 86] or subject-specific dynamical model [104, 105, 106, 107] to constrain the deformable contours. Since shape distributions in these sub-spaces are often more compact, a particular shape might be better approximated by mean shape and its variations in a sub-space. These methods are able to simultaneously handle temporal dynamics (intra-subject variability) and inter-subject variability in order to predict the specific cardiac dynamics of a new sequence based on the shapes observed in past frames.

It is also worth mentioning that there are some successful work in the face alignment field that attacks similar problems, such as the subspace constrained mean-shift [70] and the multi-level generative model [37]. Specifically, Gu et.al. [37] have proposed a shape regularization approach by using of a multi-level generative model, and demonstrated its application in face alignment. This alignment system is capable of dealing with real-world images with a wide range of imaging conditions and appearance variations. Shape inference is also a potential solution. It constructs a surface from a set of 3D points. In [32], to infer the shape, a nearest-neighbor approach is used by finding the closest instance in a database, and the database is based on the expert's structure annotations.

These methods can model more complex shape variations than the original ASM

does. However, usually there are still some assumptions about the shape distribution.

2.1.2   Handling Non-Gaussian Errors

A large number of proposed modifications on the original ASM algorithm try to improve the robustness against outliers (erroneous landmarks/boundaries). Duta and Sonka [21] propose detecting and correcting outliers by using the variance information from the PDM. If a point is considered an outlier, it is corrected based on the position of its neighbors. Lekadir et al. [53] employ a shape metric based on the ratio of landmark distances to detect outliers. Its use of an invariant shape metric allows outlier analysis to be based only on shape information and carried out prior to the least squares minimization procedures of the ASM. Furthermore, the identified outliers are not rejected or replaced by the corresponding mean values, and a correction mechanism suggests a replacement point for each outlier before the model fitting procedure.

Other methods try to decrease outliers' influence using the weighting of residuals. Rogers and Graham [68] evaluate the use of M-estimators, image match and random sample consensus (RANSAC) [25] for this purpose. In a concluding evaluation, RANSAC was the most effective of these three methods. Nahed et al. [63] proposed to use a robust point matching algorithm [9] which rejects outliers and finds the best-fitting model. The strength of this model is that given two datasets, it is able to establish as many correspondences as possible between them while rejecting outliers. The size of the two datasets is irrelevant to the performance of the technique. Thus it is able to match point sets of arbitrary size more robustly than ASM or RPM-TPS [9] alone.

Missing landmarks/boundaries is also a special case of outliers. Yan et.al. [87] tried to use partial ASM to address this problem of missing boundary in image segmentation. They have demonstrated that missing boundary information in prostate TRUS image can be well estimated from the available partial salient contours. By using the estimated complete shape together with a robust Discrete Deformable Model, the prostate can be automatically segmented from TRUS images with significantly improved results.

2.1.3 Preserving Local Detail Information

Another difficulty is to preserve local details of the input shape when such details are also present in the training data but not statistically significant. PCA performs eigen-analysis and extracts eigenvectors with the largest eigenvalues. The discarded eigenvectors are statistically insignificant, but they may contain important local details. Some relevant work can alleviate this problem. Sparse PCA [78] obtains sparser modes and produces near-orthogonal components. Thus each mode only affects locally clustered landmarks and captures more detail information, and is able to represent anatomically meaningful variables from a data set. Some other methods divide the shape model into several independently modeled parts, such as the hierarchical approach [15]. A hierarchical representation of a shape is introduced by using its wavelet transform, followed by a PCA on the wavelet coefficients. Thus parts information can be represented by wavelet. Since the smaller parts exhibit less variation, they can be captured with fewer training samples than the variations for the full shape.

However, most discussed methods focus on solving one or two limitations. It is not trivial to handle all of them simultaneously. In our work, we address these challenges in a unified framework as outlined in Chapter 3.

2.2 Sparsity Methods

2.2.1 Sparse Representation

Sparsity methods have been widely investigated recently. It shows that a sparse signal can be recovered from a small number of its linear measurements with high probability [5, 18]. To solve these problems of sparsity priors, one can either use greedy methods such as basis pursuit (BP) [8], matching pursuit [60], orthogonal matching pursuit (OMP) [7] and stagewise OMP (stOMP) [19], or use $L1$ norm relaxation and convex optimization [5, 49, 24].

The sparsity prior has been widely used in computer vision and multimedia communities, such as, but not limited to, robust face recognition [84], image restoration [59],

background subtraction [40], MR reconstruction [58, 43], automatic image annotation [91, 30] and matting [20]. Specifically, Wright et.al. [84] have contended both theoretically and experimentally to show that sparse representation is critical for high-performance classification of high-dimensional data such as face images. It also demonstrates that the choice of features is less important than the number of features used, and occlusion and corruption can be handled uniformly and robustly with this framework. Despite its initial success in the face recognition problem, Huang et.al. [39] have demonstrated that such image sparse representation is sensitive to image plane transformations such that it cannot reconstruct the sparse representation of a geometrically transformed image. A transformation-invariant image sparse representation has been proposed to alleviate this problem in a certain level. By coupling compressive sensing and random projection manifold, this approach can efficiently recover not only sparse representation of a target image but also the image plane transformation between the target and the model images. After the success on the face recognition application, sparse representation becomes popular in some other image based applications. For example, Gao et.al. [30] have extended traditional sparse representation by considering clustering constraint to automatically annotate images. Cong et.al. [10] have extended this framework to videos, and employed sparse representation and reconstruction cost to detect abnormal events. Yet, to our knowledge, such techniques have not been used in the context of shape priors.

In this thesis, we have investigated sparse representation for shapes instead of images or videos. In this context, the main challenges are twofold: 1) shape data can have arbitrary dimensions, e.g., 2D contour or 3D mesh; 2) different from the assumption in [84], our input shape data may have large misalignment. Thus transformation parameter should also be modeled explicitly. Details are discussed in Sec. 3.

2.2.2 Dictionary Learning

When training dataset has thousands or millions of samples, it may not be feasible to use all of them due to the computational consideration. It is necessary to learn a compact dictionary to represent the original dataset by minimizing the reconstruction

errors.

Finding the compact dictionary has been extensively studied. We briefly introduce some relevant techniques. Dictionary learning typically consists of the sparse coding and codebook update. Greedy algorithms such as matching pursuit (MP) [60] and orthogonal matching pursuit (OMP) [7] can be employed for finding sparse coefficients (coding). Extensive study of these algorithms shows that if the sought solution is sparse enough, these greedy techniques can obtain the optimal solution [82]. When the sparse data has group clustering trend [45], AdaDGS [40] can be employed to further improve the performance. To update codebook, method of optimal direction (MOD) [22] and K-SVD [1] are two effective approaches. Although both of them result in similar results, we use K-SVD to learn our dictionary because of its better convergence rate.

Chapter 3

Sparse Representation Based Shape Prior Modeling

3.1 Sparse Shape Composition

3.1.1 Model Formulation

In this study, we aim to model the shape of an object using a set of existing training shape instances.

Notations and Basic Framework: Explicit parametric shape representation is employed to model a shape instance, i.e., a curve (2D) or a triangular mesh (3D) consisting of a set of vertices. To describe the ith shape in the training data, the coordinates of all its vertices are concatenated into a vector $d_i \in \mathbf{R}^n$, where n is the product of the number of vertices in each shape by the dimension. Thus the training repository can be represented as a matrix $D = [d_1, d_2, ..., d_k] \in \mathbf{R}^{n \times k}$, where k is the number of shapes. In our framework, all $d_i, i = 1, 2, 3, ..., k$ are pre-aligned using generalized Procrustes analysis [36]. $y \in \mathbf{R}^n$ is the vector of a newly-input shape which needs to be constrained or refined. Our basic framework assumes that after proper alignment, any input shape y can be approximately represented as a weighted linear combination of existing data $d_i, i = 1, 2, 3, ..., k$, and the parts which cannot be approximated are noises. We denote $x = [x_1, x_2, ..., x_k]^T \in \mathbf{R}^k$ as the coefficients or weights. Thus the value of x for the linear combination is found by minimizing the following loss function:

$$\arg\min_{x,\beta} \{\|T(y, \beta) - Dx\|_2^2\}, \tag{3.1}$$

where $T(y, \beta)$ is a global transformation operator with parameter β. It aligns the input shape y to the mean shape of existing data D. x and β are computed by solving Eq. 3.1.
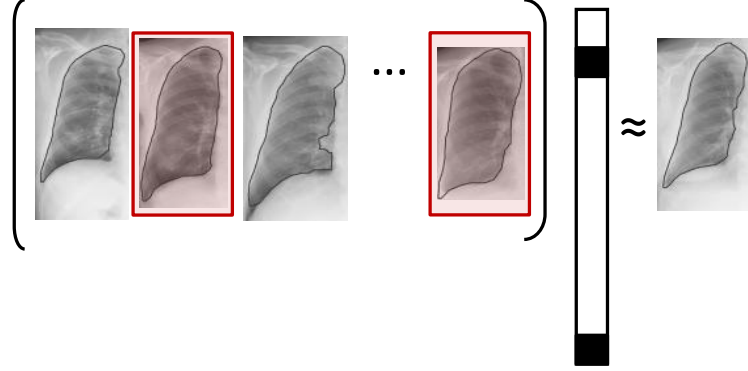
Figure 3.1: Illustration of formulations $Dx \approx T(y, \beta)$ using lung shapes as an example. Note that here D is a matrix of shapes instead of images. Vector x is sparse in this case, as only two shapes are selected. Black squares mean nonzero elements.

Thus the input shape y is constrained or refined as Dx and transformed back by the inverse of the transformation matrix using parameter β. Fig. 3.1 shows an illustration of the formulation.

Sparse Linear Combination: The limitations of Eq. 3.1 are twofold. First the data matrix D may be overcomplete ($k > n$) when the number of shapes is larger than the length of d_i. Thus the system may not have a unique solution. More constraints of the coefficient x are needed. Second, the input shape, including the noises, may be perfectly represented if any linear combination can be used. A more appropriate assumption is that the input shape can be approximately represented by a *sparse* linear combination of existing data. This way, the problem is reformulated as:

$$\underset{x, \beta}{\arg \min} \ \{\|T(y, \beta) - Dx\|_2^2\}, \tag{3.2}$$

$$s.t. \ \|x\|_0 \leq k_1$$

where $\|\cdot\|_0$ is the L^0 norm counting the nonzero entries of a vector, k_1 is the pre-defined sparsity number. Such formulation ensures that the number of nonzero elements in x is smaller than k_1. This case is illustrated in Fig. 3.1 as $k_1 = 2$. The value of k_1 depends on specific applications, and is discussed in Chapter 5.

Handling Non-Gaussian Errors: The formulation Eq. 3.2 works well for many scenarios. However, there is still one limitation in Eq. 3.2. Since the loss function is
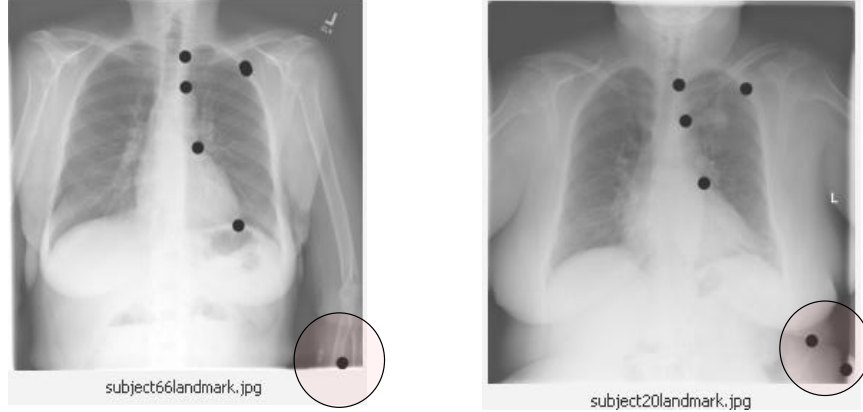
Figure 3.2: The landmark detection results contain outliers, which are marked by red circles. These outliers are usually sparse compared to the whole set of detected landmarks.

based on $L2$ norm, it assumes that the error model follows a Gaussian distribution. Thus it is sensitive to large noises or gross errors of the input shape, caused by image occlusion, points missing or detection errors (Fig. 3.2). Such problem happens frequently in many applications. In these cases, some errors can be very large, but they are relatively sparse compared to the whole data. To alleviate this problem, we explicitly model the error as a sparse vector $e \in \mathbf{R}^n$ by reformulating the problem as:

$$\arg\min_{x,e,\beta} \{\|T(y,\beta) - Dx - e\|_2^2\}, \tag{3.3}$$

$$s.t. \ \|x\|_0 \le k_1, \|e\|_0 \le k_2$$

where k_2 is the sparsity number of e. When solving Eq. 3.3, e captures sparse but large errors which are caused by occlusion or point missing. When there is no such error, the $L2$ norm loss function can deal with it well and e will be all zeros. Thus e is a good supplement which specifically handles non-Gaussian and sparse errors. Note that unlike the formulation in the robust face recognition [84], we do not assume that the misalignment is small and thus explicitly model the transformation with parameter β in Eq. 3.3.

Eq. 3.3 can alleviate the three challenges aforementioned because of the following reasons:

- General: There is no assumption of a parametric distribution model (e.g., a unimodal distribution assumption in ASM). Thus it can model complex shape statistics.

- Robust: Eq. 3.3 Explicitly modeling e with $L0$ norm constraint. Thus it can detect gross (sparse) errors.

- Lossless: It uses all training shapes. Thus it is able to recover detail information even if the detail is not statistically significant in training data.

Convex Relaxation: The constraints in Eq. 3.3 are not directly tractable because of the nonconvexity of L^0 norm. Greedy algorithms can be applied to this NP-hard L^0 norm minimization problem, but there is no guarantee to capture the global minima. In the general case, no known procedure can correctly find the sparsest solution more efficiently than exhausting all subsets of the entries for x and e. Furthermore, in practice the sparsity numbers k_1 and k_2 may change for different data in the same application. For example, some data have errors while others do not. Fortunately, recent developments in sparse representation provide a theorem to efficiently solve this kind of problems through $L1$ norm relaxation [79]. Thus Eq. 3.2 is reformulated as:

$$\arg\min_{x,\beta} \{\|T(y,\beta) - Dx\|_2^2 + \lambda_1\|x\|_1\}, \tag{3.4}$$

which is denoted as SSC(3.4), and is evaluated in the experiments section. Similarly, Eq. 3.3 is reformulated as:

$$\arg\min_{x,e,\beta} \{\|T(y,\beta) - Dx - e\|_2^2 + \lambda_1\|x\|_1 + \lambda_2\|e\|_1\}, \tag{3.5}$$

where λ_1 and λ_2 control how sparse x and e are, respectively. After relaxation, $\lambda_1\|x\|_1 + \lambda_2\|e\|_1$ is non-smooth but continuous and convex. Eq. 3.5 is our objective function of our proposed SSC. The deviation from Eq. 3.3 to Eq. 3.5 relaxes the absolute sparseness constraints of the objective function ($L0$ norm to $L1$ norm). From the shape modeling perspective, we might use more shape instances for shape composition by optimizing Eq. 3.5. However, since this deviation converts a NP hard problem to a continuous

and convex optimization problem which can be solved efficiently, it paves the way for a feasible shape composition procedure as described in Sec. 3.2.

Although our focus is on the shape prior modeling instead of sparse learning methods, it is still worth mentioning that many other methods can also achieve sparsity, such as Bayesian variable selection [16, 51, 31]. In our model, we choose $L1$ norm based sparse representation because it is a convex optimization problem, which can be effectively solved by many convex techniques and solvers. Furthermore, [17] provides theoretical proofs that the $L1$ relaxation can preserve the sparsity property of $L0$ norm constraint.

3.1.2   Model Discussions

Connections to Other Methods: It is interesting to look into Eq. 3.5 by adjusting λ_1 and λ_2 into some extreme values.

- If λ_2 is extremely large, e will be all zeros. Thus SSC is similar to methods which do not model non-Gaussian errors.

- If both λ_1 and λ_2 are large enough, e will be all zeros and x may have only one nonzero element. Thus SSC becomes the nearest neighbor method.

- If λ_2 is extremely large and λ_1 is small, a dense linear combination of shapes is used, which is able to perfectly approximate the transformed input shape. Thus SSC degenerates to the Procrustes analysis.

The insight of Eq. 3.5 indeed reveals the connections of our SSC with some other popular methods. Those methods can be regarded as special cases of SSC. In other words, SSC provides a unified framework to deal with different challenges of shape prior modeling simultaneously. SSC can also provide flexibility to meet the requirements of different applications by adjusting the sparsity of x and e.

Parameter Settings: Eq. 3.5 has two user tunable parameters λ_1 and λ_2, which are usually crucial to the performance and convergence. From a practical point of view, it is desirable that the parameters are easy to tune and not sensitive to different data

in one application. If these parameters have physical meanings, it is straightforward to adjust them. Fortunately, the parameters of our algorithm also have such a property. λ_1 controls the sparsity of x. The length of vector x is equal to the number of shapes in the shape repository. It is usually larger than 100. To generate a sparse coefficient x, a large λ_1 is necessary. λ_2 controls the sparsity of e. The length of vector e is equal to the number of vertices (multiplied by the dimension), which ranges from around 20 to $2,000$. e should not be too sparse, otherwise it cannot capture any errors. Thus λ_2 should be relatively small. Both parameters are straightforward to tune given their meanings. Furthermore, the experiments in Chapter 5 show that the same group of parameters works well for all data in one application.

Computational Efficiency: The computation efficiency of most shape prior method includes offline training and runtime testing. For example, ASM needs to apply generalized Procrustes analysis and PCA as offline learning. When new data is coming, it just needs to do matrix manipulation such as projection. The running time depends on how many modes preserved during training. The proposed SSC model only needs generalized procrustes analysis as preprocess. During testing, it needs to solve an convex optimization problem. The running time depends on the number and size of training data, and also the optimization method. Both are in realtime for a reasonable size of dataset (e.g., lung localization in Sec. 5.1). When the number or the size of training data becomes large, the optimization stage can be slow. We also present some approaches to accelerate this process, such as using local shape priors (Sec. 3.4) and learning a compact dictionary (Sec. 3.5).

3.2 Optimization

3.2.1 Optimization Framework

To solve Eq. 3.5, we need to simultaneously recover the alignment parameter β and error e. It is a typical Chicken-and-Egg problem. Furthermore, to efficiently optimize Eq. 3.5, we need to deal with the nonlinearity of $T(y, \beta)$ if the transformation is rigid or a similarity. A notable approach is to use iterative linearization and optimize all variables

---

Algorithm 1 Optimization framework to solve Eq. 3.5.

Input: Data matrix of shape repository $D \in \mathbf{R}^{n \times k}$, where k is the number of shapes, and each column is a training shape $d_i \in \mathbf{R}^n$.
Output: $y'_{refined}$.
repeat
    "E" step: β is estimated using Procrustes analysis, which is a similarity transformation and aligns the input shape y to the mean shape of D. $y' = T(y, \beta)$.
    "M" step: Eq. 3.6 is efficiently minimized use FISTA [3]. x and e are computed.
until Stop criterions
Compute $y_{refined} = Dx$.
Compute $y'_{refined} = T(y_{refined}, \gamma)$, where γ is the parameter corresponding to the inverse of the transformation matrix using β.

---

simultaneously, which was proposed and successfully applied in image alignment [64]. However, this algorithm assumes that the initial misalignment is not too large, which may not be held in our problem. Furthermore, it focuses on rigid transformation in 2D images, while we deal with nonrigid transformation in arbitrary dimensions for shapes. The efficiency of the optimization framework is also important.

Our solution is to use EM types of algorithms (or alternating minimization). Eq. 3.5 is divided into two sub-problems: 1) estimating β and computing $T(y, \beta)$, 2) efficiently minimizing this simplified linear inverse problem. In the "E" step, β is estimated using Procrustes analysis, which aligns the shape y to the mean shape. Then vector $y' = T(y, \beta)$ is obtained. In the "M" step, the following simplified problem is minimized:

$$\arg\min_{x,e} \{\|y' - Dx - e\|_2^2 + \lambda_1 \|x\|_1 + \lambda_2 \|e\|_1\}, \tag{3.6}$$

which is now a linear inverse problem. It is then efficiently solved using existing methods such as the Fast Iterative Shrinkage Thresholding Algorithm (FISTA) [3] or the proposed composite splitting algorithms in the following sections. Two procedures are iteratively employed to obtain x, e and β. Then Dx is computed as the approximated shape and is transformed to its original coordinate system. The framework is detailed in Algorithm 1.

3.2.2   Linear Inverse Problem

In this section, we introduce an efficient algorithm to solve linear inverse problems such as Eq. 3.6. Eq. 3.6 is relatively easy to solve since all constraints are based on $L1$ norm constraint. Thus FISTA [3] can be directly employed. However, it is possible that constraints consist of different types of norms, such as $L2$ and total-variation (TV) norm when modeling new priors. Thus it is desirable to improve existing algorithms. We proposed two composite splitting algorithms [42, 44, 43] to efficiently solve this generalized problem (which can also be used to solve Eq. 3.6):

$$min\{F(x) \equiv f(x) + \sum_{i=1}^{m} g_i(B_i x), x \in \mathbf{R^p}\} \tag{3.7}$$

where f is the loss function and $\{g_i\}_{i=1,...,m}$ are the prior models; f and $\{g_i\}_{i=1,...,m}$ are convex functions and $\{B_i\}_{i=1,...,m}$ are orthogonal matrices.

Here are some notations used in our algorithms:

Matrix inner product: $\langle X, Y \rangle = trace(X^H Y)$.

Gradient: $\nabla f(x)$ denotes the gradient of the function f at the point x.

The proximal map: given a continuous convex function $g(x)$ and any scalar $\rho > 0$, the proximal map associated to function g is defined as follows [2, 3]:

$$prox_\rho(g)(x) := \arg\min_u\{g(u) + \frac{1}{2\rho}\|u - x\|^2\} \tag{3.8}$$

ϵ-**optimal Solution**: Suppose x^* is an optimal solution to Eq. 3.7. $x \in \mathbf{R}^p$ is called an ϵ-optimal solution to Eq. 3.7 if $F(x) - F(x^*) \leq \epsilon$ holds.

3.2.3   Composite Splitting Algorithms

As shown in [42], if there is an efficient algorithm to solve the following problem:

$$x^k = \arg\min_x \frac{1}{2}\|x - x_g\|^2 + \sum_{i=1}^{m} g_i(B_i x) \tag{3.9}$$

The original composite regularization can then be efficiently solved by the FISTA, which obtains an ϵ-optimal solution in $\mathcal{O}(1/\sqrt{\epsilon})$ iterations. Eq. 3.9 can be considered as a denoising problem. We use composite splitting techniques to solve this problem: 1) splitting variable x into multiple variables $\{x_i\}_{i=1,\dots,m}$; 2) performing operator splitting over each of $\{x_i\}_{i=1,\dots,m}$ independently and 3) obtaining the solution x by linear combination of $\{x_i\}_{i=1,\dots,m}$. We call it Composite Splitting Denoising (CSD) method, which is outlined in Algorithm 2. Its validity is guaranteed by the following theorem:

Theorem 1. *Suppose $\{x^j\}$ the sequence generated by the CSD. If x^* is the true solution of Eq. 3.9, x^j will strongly converges to x^*.*

Algorithm 2 CSD

  Input: $\rho = 1/L$, α, β, $\{z_i^0\}_{i=1,\dots,m} = x_g$
  for $j = 1$ **to** J **do**
    for $i = 1$ **to** m **do**
      $x_i = \arg\min_x \frac{1}{2m}\|x - z_i^{j-1}\|^2 + g_i(B_i x)$
    end for
    $x^j = \frac{1}{m}\sum_{i=1}^m x_i$
    for $i = 1$ **to** m **do**
      $z_i^j = z_i^{j-1} + x^j - x_i$
    end for
  end for

Combining the CSD with ISTA [3], a new algorithm, CSA, is proposed for composite regularization, Eq. 3.7. In practice, we found that a small iteration number J in the CSD is enough for the CSA to obtain good reconstruction results. Especially, it is set as 1 in our algorithm. Numerous experimental results in the next section will show that it is good enough for real composite regularization problem.

Algorithm 3 outlines the proposed CSA. In each iteration, Algorithm 3 decomposes the original problem into m subproblems and solve them independently. For many problems in practice, these m subproblems are expected to be far easier to solve than the original joint problem. Another advantage of this algorithm is that the decomposed subproblems can be solved in parallel. Given x^{k-1}, the m subproblems to compute $\{y_i^k\}_{i=1,\dots,m}$ are solved simultaneously in Algorithm 3.

Algorithm 3 CSA

> **Input:** $\rho = 1/L$, x^0
> **repeat**
>> **for** $k = 1$ **to** K **do**
>>> **for** $i = 1$ **to** m **do**
>>>> $y_i^k = prox_\rho(g_i)\left(B_i\left(x^{k-1} - \frac{1}{L}\nabla f_i\left(x^{k-1}\right)\right)\right)$
>>> **end for**
>>> $x^k = \frac{1}{m}\sum_{i=1}^m B_i^{-1} y_i^k$
>> **end for**
> **until** Stop criterions

Algorithm 4 FCSA

> **Input:** $\rho = 1/L$, $t^1 = 1$ $x^0 = r^1$
> **repeat**
>> **for** $k = 1$ **to** K **do**
>>> **for** $i = 1$ **to** m **do**
>>>> $y_i^k = prox_\rho(g_i)\left(B_i\left(r^k - \frac{1}{L}\nabla f_i\left(r^k\right)\right)\right)$
>>> **end for**
>>> $x^k = \frac{1}{m}\sum_{i=1}^m B_i^{-1} y_i^k$
>>> $t^{k+1} = \frac{1+\sqrt{1+4(t^k)^2}}{2}$
>>> $r^{k+1} = x^k + \frac{t^k-1}{t^{k+1}}(x^k - x^{k-1})$
>> **end for**
> **until** Stop criterions

3.2.4 Fast Composite Splitting Algorithms

In this section, a fast version of CSA named as FCSA is proposed to solve Eq. 3.7, which is outlined in Algorithm 4. FCSA decomposes the difficult composite regularization problem into multiple simpler subproblems and solve them in parallel. Each subproblems can be solved by the FISTA, which requires only $O(1/\sqrt{\epsilon})$ iterations to obtain an ϵ-optimal solution.

In this algorithm, if we remove the acceleration step by setting $t^{k+1} \equiv 1$ in each iteration, we will obtain the CSA. A key feature of the FCSA is its fast convergence performance borrowed from the FISTA. FISTA can obtain an ϵ-optimal solution in $\mathcal{O}(1/\sqrt{\epsilon})$ iterations.

Another key feature of the FCSA is that the cost of each iteration is $\mathcal{O}(mp\log(p))$, as confirmed by the following observations. The step $y_i^k = prox_\rho\left(g_i\right)\left(B_i\left(r^k - \frac{1}{L}\nabla f_i\left(r^k\right)\right)\right)$ can be computed with the cost $\mathcal{O}\left(p\log(p)\right)$ for a lot of prior models g_i. The step

$x^k = \frac{1}{m} \sum_{i=1}^{m} B_i^{-1} y_i^k$ can also be computed with the cost of $\mathcal{O}(p \log(p))$. Other steps only involve adding vectors or scalars, thus cost only $\mathcal{O}(p)$ or $\mathcal{O}(1)$. Therefore, the total cost of each iteration in the FCSA is $\mathcal{O}(mp \log(p))$.

With these two key features, the FCSA efficiently solves the composite regularization Eq. 3.7 and obtains better results in terms of both the accuracy and computation complexity.

3.3   Shape Inference and Refinement

Due to imaging artifacts and diseases, appearance cues in medical images might be unreliable or misleading. On the other hand, however, strong shape priors of human anatomy provide opportunities to shape prior-based methods. To evaluate the capability of the proposed shape prior modeling, we apply it to two tasks: 1) organ localization using shape inference and landmark detection, and 2) segmentation using shape refinement and deformable model. The first task organ localization can be used as an initialization step of the segmentation framework.

3.3.1   Organ Localization using Shape Inference

The positions and orientations of the same organ vary significantly in medical image data. Quickly and accurately locating the organ is crucial to image segmentation. One approach is to find a similarity transformation matrix, and then use this matrix to align a mean shape to the organ [100]. Generally it achieves good performance. However, similarity transform only has nine degrees of freedom. Thus it may not be able to represent some specific data or shapes by transforming a mean shape (in Sec. 5.1).

To solve this problem, we propose a landmark detection and shape inference based localization method. A learning-based method is employed for landmark detection [90]. Detected landmarks can be very sparse compared to the whole shape. Furthermore, there may be gross errors or point missing from the detection results. We use SSC to infer a shape based on these detected landmarks. Compared to solely using similarity transformation to transform the mean shape, our approach has more degrees of freedom
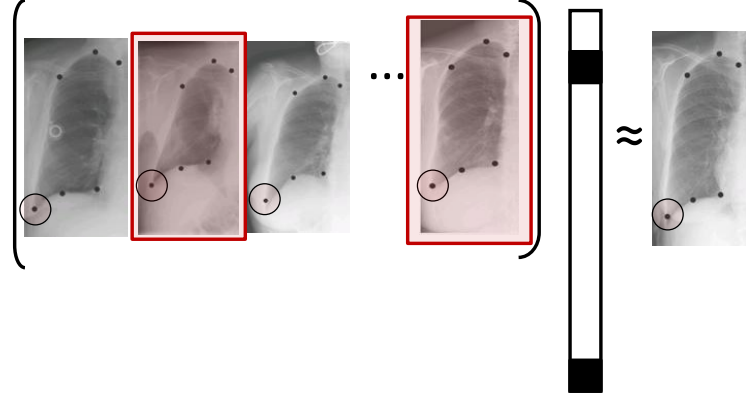
Figure 3.3: 2D lung localization using landmark detection and shape inference. A sparse set of landmarks is detected automatically. Then matrix D is resampled to represent landmark set, and the sparse coefficient x is computed. Same x is used to combine whole shapes as shape approximation.

and is able to better fit to the actual shape in the image.

In this framework, we assume that ground truths for shapes are available from training data, and the one-to-one correspondence is already obtained. We then automatically or manually choose some specific points (e.g., corner points or high curvature points) as landmarks on the shape of each data. Such training landmarks and shapes are fed into data matrices denoted as D_L and D_S, respectively. Given a testing image, its landmarks y_L are detected using a learning based method [90]. Then x and β is computed by optimizing Eq. 3.5 with D_L and y_L. Finally $D_S x$ is computed as the refined shape and transformed back to its coordinate system using inverse of the transformation matrix with parameter β. Such localization can also be used as the initialization of many segmentation algorithms. Fig. 3.3 shows one example of 2D lung localization using landmark detection and shape inference.

In Sec. 5.1, this framework is employed to locate 2D lung from X-ray image. In Sec. 5.2, this framework is used to initialize the 3D deformable model.

3.3.2 Organ Segmentation using Shape Refinement

Curve or surface based deformable models have been widely used for organ segmentation [48, 12, 85, 47, 71, 54, 108]. Many deformable models consist of two iterative
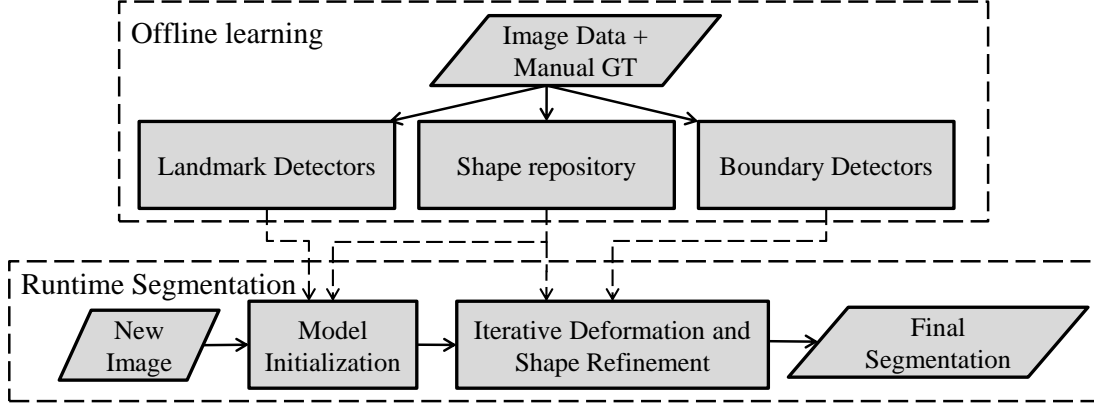
Figure 3.4: The workflow of our segmentation framework consisting the offline learning and online testing modules. Using landmark detection and shape inference, we are able to do model initialization (i.e., organ localization). After the shape model is initialized, the shape is deformed and refined alternatively to fit the image boundary.

steps, local deformation based on low level image information and global shape refinements based on high-level shape priors. For those applications that have noisy image information, high level shape information becomes especially critical. Our sparse shape prior modeling can be used for shape refinement method as a regularization step during deformation. An initialized shape is deformed following the image gradient information. During the deformation procedure, the shape refinement is employed as high level constrains to avoid getting stuck in local minima of the image information. Denote the training shape matrix as D_S, and the intermediate deformation result as y_S. Then x is computed by solving Eq. 3.5 with D_S and y_S. $D_S x$ is used as the refined shape and transformed back. In this refinement procedure, e may not have large values since the model is already roughly aligned after initialization. However, modeling e is still necessary to capture small errors not following Gaussian distribution.

In Sec. 5.2, this framework is used to segment 3D liver. The whole framework is shown in Fig. 3.4.

3.4   Extension I: Multi-resolution and Local Shape Priors

It has been widely accepted that multiresolution/hierarchical scheme should be employed to improve the efficiency and robustness of deformable segmentation [52]. In

a multiresolution scheme, only a small set of sparsely distributed vertices are used as driving vertices to estimate a rough segmentation of the initial stages. As the iterations increase, more and more vertices join the driving set to gradually reach accurate segmentation. Our sparse shape composition method naturally supports this scheme by estimating a sparse linear combination from an incomplete input. Assume $\mathbf{v}_{sub} = \mathbf{S}\mathbf{v}$ is a subset of all vertices in shape $\mathbf{v}$, where $\mathbf{S}$ is a binary diagonal matrix which indicates if the ith vertex is in the subset ($\mathbf{S}_{ii} = 1$). Eq. 3.5 can then be naturally extended as:

$$\arg\min_{\mathbf{x},\mathbf{e},\beta} \|T(\mathbf{v_{sub}}, \beta) - \mathbf{S}D\mathbf{x} - \mathbf{Se}\|_2^2, \quad s.t.\|\mathbf{x}\|_0 < k_1, \|\mathbf{e}\|_0 < k_2, \qquad (3.10)$$

Eq. 3.10 can be solved using the same EM optimization. The only difference is that the optimized $\mathbf{x}$ will be finally applied on the full space of D, such that the entire input shape is refined.

One extreme situation of Eq. 3.10 is that $\mathbf{S}$ becomes very sparse and only includes a few vertices (usually with the most distinctive appearance/geometry characteristics). In this situation, Eq. 3.10 indeed becomes the shape inference method (Sec. 3.3.1), which is the first step of our runtime segmentation system (Sec. 3.3.2). Again, by incorporating shape priors with the assumption of "sparse gross errors", our initialization method becomes robust to erroneous landmark detections due to severe diseases/imaging artifacts.

The merit of Eq. 3.10 is actually beyond the support of multiresolution deformation scheme. In practice, many 3D deformable models include many thousands of points to give an accurate description of organ shapes. The optimization of Eq. 3.6 thus has high computational complexity. In addition, since local shape statistics often lie in a more compact space than global ones, shape priors built on sub-surface are expected to improve the performance. To achieve this goal, we propose a "mesh partitioning" method, which can also be seamlessly incorporated in our sparse shape composition formula. Affinity propagation clustering [28] is employed to divide the model shape into multiple partitions. Since one-to-one correspondences are already constructed among all shapes, affinity propagation only needs to perform once for the model shape. The
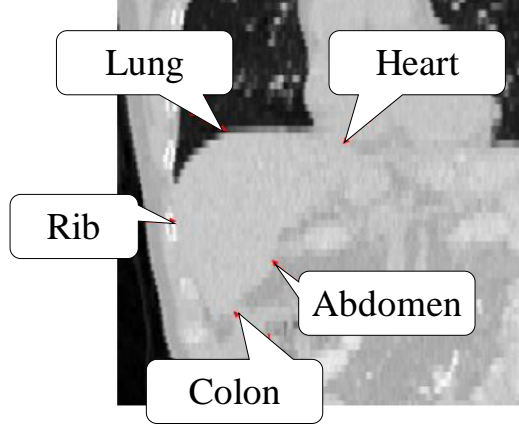
Figure 3.5: An example of liver boundaries from CT images. It includes boundaries between liver and rib, lung, heart, abdomen and colon that show heterogeneous appearance in local shapes.

similarity used in the affinity propagation is defined as the combination of the image similarity and geodesic distances between vertices [88]:

$$s(v_i, v_j) = 1 - \frac{1}{K} \sum_{k=1}^{K} \left[\alpha G\left(v_i^k, v_j^k\right) + (1 - \alpha) C\left(F\left(v_i^k\right), F\left(v_j^k\right)\right) \right] \qquad (3.11)$$

where K is the number of training subjects, v_i^k denotes the ith vertex of the kth subject. $G\left(v_i^k, v_j^k\right)$ denotes the geodesic distance between v_i^k and v_j^k. $C\left(F\left(v_i^k\right), F\left(v_j^k\right)\right)$ denotes the Euclidean distance between image feature vector calculated at v_i^k and v_j^k. The reason we also need to consider image information is that organ boundaries may have heterogeneous appearance, as shown in Fig. 3.5. Thus such partitioning also benefits the performance of boundary detectors if we train different detectors for each local shape patch. Fig. 3.6 shows an example of partition result for the liver data.

In our implementation, each divided partition is further "dilated" for several levels to produce overlaps with neighboring partitions. Finally, partitions are converted to a set of indication matrices $\mathbf{S}_1, \mathbf{S}_2, ..., \mathbf{S}_p$ used in Eq. 3.10. The optimization problem defined on the entire surface is thus decomposed to a set of sub-problems. Each partition is refined independently but the refined partitions are averaged in these overlapping regions to guarantee the smoothness of the entire surface.

The computational complexity of an existing solver (e.g., interior point method) is
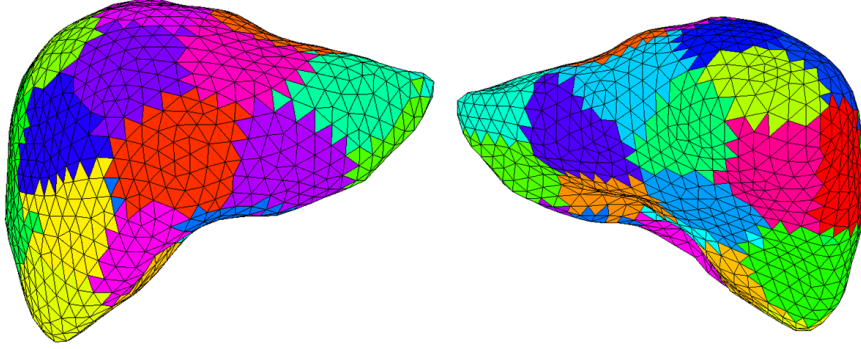
Figure 3.6: Mesh partitioning using affinity propagation. This local shape setting can learn detail information better, and is more efficient for the optimization solver.

$\mathcal{O}(N^3)$, where N is the number of vertices of the whole surface. After dividing the whole surface into p partitions with about $\frac{N}{p}$ vertices in each partition. The computational complexity is decreased to only $\frac{1}{p^2}$ of the original one, which highly improves the efficiency.

3.5 Extension II: Dictionary Learning using K-SVD

In many applications of medical image analysis, the number of training data is usually less or around hundreds. In such cases, we can just put all data into the matrix D, since Eq. 3.5 can be still efficiently solved. However, if there are thousands or millions training data, it may not be feasible to select models fast. Thus, dictionary learning technique is employed to compute a compact codebook (i.e., dictionary), which can well represent the original database.

We use K-SVD [1] to achieve this. It includes two steps, sparse coding and codebook update. Sparse coding is a greedy method which can approximate an input data by finding a sparse set of elements from the codebook. Codebook update is to generate a better dictionary given sparse coding results. Two steps are alternately employed until converge. The resulting codebook can have much less number of elements (i.e., less columns) while is still generative enough to represent input data.

In general, the objective of dictionary learning is to find dictionary D and coefficient

---

Algorithm 5 The framework of the OMP algorithm.

Input: Dictionary $D \in \mathbf{R}^{n \times k}$, input data $y_i \in \mathbf{R}^n$.
Output: coefficients $x_i \in \mathbf{R}^k$.
$\Gamma = \emptyset$.
repeat
    Select the atom which most reduces the objective

$$\arg \min_j \left\{ \min_{x'} |y_i - D_{\Gamma \cup \{j\}} x'|_2^2 \right\} \tag{3.14}$$

    Update the active set: $\Gamma \leftarrow \Gamma \cup \{j\}$.
    Update the residual using orthogonal projection

$$r \leftarrow \left(I - D_\Gamma (D_\Gamma^T D_\Gamma)^{-1} D_\Gamma^T \right) y_i \tag{3.15}$$

    Update the coefficients

$$x_\Gamma = (D_\Gamma^T D_\Gamma)^{-1} D_\Gamma^T y_i \tag{3.16}$$

until Stop criterions

X by minimizing the following equation:

$$\arg \min_{D,X} \{\|Y - DX\|_2^2\} \tag{3.12}$$

$$s.t. \quad \forall i, \|x_i\|_0 \leq L \tag{3.13}$$

Where matrix Y represents signals (all training shapes in our case), D is the unknown overcomplete dictionary, matrix X is the sparse coefficients. Denote y_i as the ith column of Y, x_i as the ith column of X, then y_i and x_i are the ith shape vector and coefficient vector respectively, with dimensionality $D \in \mathbf{R}^{n \times k}$, $y_i \in \mathbf{R}^n$ and $x_i \in \mathbf{R}^k$. It is consistent with Eq. 3.2, except that D is also unknown.

K-SVD algorithm starts from a random D and X obtained from the sparse coding stage. The sparse coding stage is based on pursuit algorithms to find the sparse coefficient x_i for each signal y_i. OMP is employed in this stage. OMP is an iterative greedy algorithm that selects at each step the dictionary element that best correlates with the residual part of the signal. Then it produces a new approximation by projecting the signal onto those elements already selected [82]. The algorithm framework of OMP is

listed in the Algorithm 5.

In the codebook update stage K-SVD employs a similar approach as K-Means to update D and X iteratively. In each iteration D and X are fixed except only one column d_i and the coefficients corresponding to d_i (ith row in X), denoted as x_T^i. The Eq. 3.12 can be rewritten as

$$\left\| Y - \sum_{j=1}^{k} d_j x_T^j \right\|_F^2 = \left\| \left(Y - \sum_{j \neq i} d_j x_T^j \right) - d_i x_T^i \right\|_F^2 \tag{3.17}$$

$$= \left\| E_i - d_i x_T^i \right\|_F^2 \tag{3.18}$$

We need to minimize the difference between E_i and $d_i x_T^i$ with fixed E_i, by finding alternative d_i and x_T^i. Since SVD finds the closest rank-1 matrix that approximates E_i, it can be used to minimize the Eq. 3.17. Assume $E_i = U \Sigma V^T$, d_i is updated as the first column of U, which is the eigenvector corresponding to the largest eigenvalue. x_T^i is updated as the first column of V multiplied by $\Sigma(1,1)$.

However, the updated x_T^i may not be sparse anymore. The solution is logical and easy. We just discard the zero entries corresponding to the old x_T^i. The detail algorithms of K-SVD are listed in the Algorithm 6.

Once D is learned from this procedure, we use it directly in the SSC model. Compared to the whole database, this compact dictionary certainly loses some information. However, K-SVD algorithm minimizes the reconstruction errors, and this compact D can highly improve the computational efficiency.

Algorithm 6 The framework of the K-SVD algorithm.

Input: dictionary $D \in \mathbf{R}^{n \times k}$, input data $y_i \in \mathbf{R}^n$ and coefficients $x_i \in \mathbf{R}^k$.
Output: D and X.
repeat
   Sparse coding:
   use OMP to compute coefficient x_i for each signal y_i, to minimize

$$\min_{x_i}\{\|y_i - Dx_i\|_2^2\} \ subject \ to \ \|x_i\|_0 \leq L \tag{3.19}$$

   Codebook update:
   for $i = 1, 2, ..., k$, update each column d_i in D and also x_T^i (ith row)
      Find the group using d_i ($x_T^i \neq 0$), denoted as ω_i
      Compute error matrix E_i as in Eq. 3.17
      Restrict E_i by choosing columns corresponding to ω_i. The resized error is denoted as E_i^R
      Apply SVD and obtain

$$E_i^R = U\Sigma V^T \tag{3.20}$$

   Update d_i as the first column of U. Update nonzero elements in x_T^i as the first column of V multiplied by $\Sigma(1, 1)$
until Stop criterions

Chapter 4

3D Shape Modeling using Mesh Quality Preserved Deformable Models

4.1 Background

The proposed SSC model is independent of data dimensions. In 3D case, shape data is usually represented as mesh, which brings more challenges than 2D contours. Since the quality of 3D training mesh is crucial to these learning based algorithms, sophisticated geometry processing methods are required to obtain them. In this chapter, we will introduce a framework to generate high quality 3D training meshes, which is usually called 3D shape atlas.

3D shape modeling of anatomies has been of particular interest and its importance has been emphasized in a number of recent studies, such as the neuroanatomical shape complex atlas and the cardiac atlas [57]. It provides a reference shape and its variances for a population of shapes. Such shape information can be useful in numerous applications such as, but not limited to, statistical analysis of the populations [12], the segmentation of the structures of interest [89], and the detection of the disease regions [81]. Besides shape modeling, image atlas has also been extensively investigated [33]. However, it is not trivial to adapt these algorithms for shape atlas. Thus we will not discuss these methods and will focus on the methods for the construction of shape model.

Shape has different representations. Cootes et al. proposed a diffeomorphic statistical shape model which analyzes the parameters of the deformation field [14]. Styner et al. used a characteristic 3D shape model dubbed m-rep to construct the atlas [80, 26]. Probably the most widely used 3D shape representation is feature point sets or landmarks from the polygon mesh. Using this representation, the mean shape and its
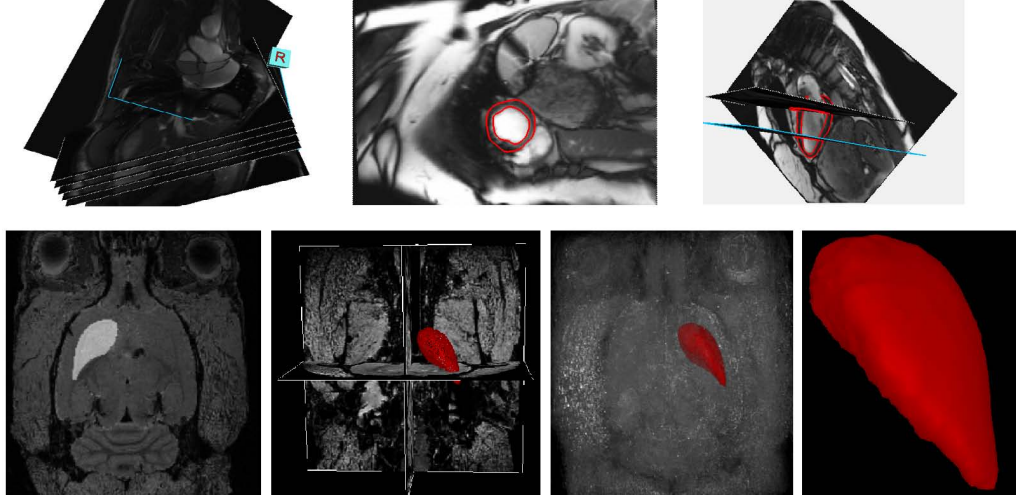
Figure 4.1: Different types of input data to create 3D shape atlas. First column: cardiac MR slices with 2D contours as the annotation. The slice thickness is large. Second column: dense MR data of rodent brain with 3D binary image as the annotation. The surface of the left striatum can be extracted from the binary image.

variances can be easily computed using generalized Procrustes analysis and Principal Component Analysis (PCA). Furthermore, such representation has been used in many segmentation algorithms, e.g., Snakes [48] and Active Shape Model (ASM) [12]. In this context, the main challenge is to robustly discover geometric correspondences for all vertices among all sample shapes. One widely used approach is to register a reference shape to the target ones. Thus the resulting shapes have nearly identical appearances as target shapes and they all share the same topology and connectivity. Adaptive focus deformable model (AFDM) [71, 72] is able to do shape registration since its attribute vectors reflect the geometric structure of the model from a local to global level. However, this method is sensitive to mesh qualities because degenerated mesh or skinny triangles can adversely affect the registration performance. Furthermore, high-quality meshes should be generated since they benefit the performance of many applications, such as statistical analysis and segmentation [75].

In this paper, we propose a unified framework [94, 29] to compute geometry correspondence for all vertices among sample shapes, and generate smooth mesh results

without significantly sacrificing shape details. A new type of energy term is incorporated into AFDM framework to preserve shape qualities during deformation. Combining this shape energy with the model energy in the original AFDM, our method is able to robustly discover one-to-one correspondence even for very complex or degenerated shapes. The whole energy function is optimized using Expectation-maximization types of algorithm. Two energy terms are minimized alternately until converge. Once the geometric correspondence and high quality meshes are generated, it is then straightforward to compute shape statistics such as mean and variance. If there are multiple structures [93], we also propose to use hierarchical modeling to effectively model them simultaneously with a limited number of samples.

Note that the formats of the input data can be diverse, such as stacks of 2D manual delineations, 3D binary images or meshes extracted from medical image data (Fig. 4.1). We assume that the inputs are a cloud of points marked on a set of sparse 2D slices (e.g., cardiac MR images in Fig. 4.1), and the slice thickness is typically several times larger than the pixel size, so that the resolution is poorer in the direction orthogonal to the slice. In the preprocess step, a 3D binary image is generated by interpolating 2D labeling. Then a surface is obtained from the 3D binary file, using the Marching Cubes algorithm. Thus different types of inputs (e.g., 3D binary image or mesh) just correspond to different starting points in the preprocess procedure.

The major contributions of this chapter are the following:

1. Propose a framework to effectively create shape atlas. It is able to generate high quality meshes without sacrificing the shape detail information. It can also robustly discover the one-to-one correspondence for very complex data (e.g., the shapes from high resolution cardiac CT images). Furthermore, this framework can handle multiple structures with a limited number of samples.

2. Using this framework, we solve several diverse and challenging tasks. Specifically, we create a high resolution cardiac shape atlas with many complex shape features such as papillary muscles and the trabeculae. We also effectively construct the atlas of multiple rodent brain structures (i.e., the cerebellum, the left and right

striatum, and the left and right hippocampus) using a small set of samples.

4.2   Algorithm Framework

As we discussed, the performance of AFDM relies on the mesh quality. With preprocess such mesh using smoothing techniques, this problem can be alleviated in a certain level. However, it is still possible that the mesh is degenerated during shape deformation. Thus it is desirable to design a unified framework to handle this problem. We define this shape registration problem as an energy minimization procedure. To do so, we defined two energy functions, *model energy* and *shape energy*. Model energy is defined to reflect the differences between the original model and the deformed model: $E_{model}(M_d, M_o)$, where M_d is the deformed mesh and M_o is the original mesh. M_d is the mesh we need to compute. Its initial value is the original mesh. Shape energy ensures that vertices are evenly distributed and shape details are roughly preserved. It is defined as $E_{shape}(M_d, L_o)$, where L_o denotes the Laplacian representation of the original shape. Laplacian representation encodes the shape detail information and smoothness constraint. Thus such energy is able to optimize shape quality without sacrificing details. We also need an *external energy* $E_{ext}(M_d, M_t)$ to move the deformable model towards target model boundaries, denoted as M_t. The energy function is defined as:

$$E = [E_{model}(M_d, M_o) + E_{ext}(M_d, M_t)] + E_{shape}(M_d, L_o) \tag{4.1}$$

Simply using model energy and external energy ($[E_{model}(M_d, M_o) + E_{ext}(M_d, M_t)]$) produces similar results as AFDM, which makes it sensitive to mesh quality. However, shape energy ensures that the mesh quality is improved during runtime. Thus this proposed energy function is more robust and can handle diverse input. The details of E_{model} and E_{shape} are introduced in the next two subsections. The external energy E_{ext} is fairly standard and is briefly introduced in the following. We first obtain a binary image I_t by using distance transform for the target mesh M_t. Then use the standard

gradient energy computed from this binary image:

$$E_{ext}(M_d, M_t) = E_{ext}(M_d, I_t) = -|\nabla I|^2, \tag{4.2}$$

where ∇ is the gradient operator.

To optimize this energy function, we use an expectation-maximization (EM) type of algorithm. During the "E" step, the model energy and external energy are minimized using similar approach as AFDM. Thus the reference shape is deformed to fit the target one, although this deformation may not be accurate due to the mesh quality. In the "M" step, the mesh quality is improved by minimizing the shape energy. As we show in Sec. 4.4, this step is formulated as a least square problem and solved efficiently. Two procedures are alternately employed to robustly register the reference model to the target model. In the following we introduce the details of two types of energies and how to solve each one.

4.3 Model Energy

Model energy is defined as the differences of attribute vectors. An attribute vector is attached to each vertex of the model, which reflects the geometric structure of the model from a local to global level. In 3D, for a particular vertex V_i, each attribute is defined as the volume of a tetrahedron on that vertex. The other three vertices form the tetrahedron are randomly chosen from the lth level neighborhood of V_i. Smaller tetrahedrons reflect the local structure near a vertex while larger tetrahedrons reflect more global information around a vertex. The attribute vector, if sufficient enough, uniquely characterizes different parts of a surface of a boundary. The volume of a tetrahedron is defined as $f_l(V_i)$. The attribute vector on a vertex is defined as:

$$F(V_i) = [f_1(V_i), f_2(V_i), ..., f_{R(V_i)}(V_i)], \tag{4.3}$$

where $R(V_i)$ is the neighborhood layers we want to use around V_i.

The model energy term reflects the differences of attribute vectors between the

original model and the deformed model:

$$E_{model}(M_d, M_o) = \sum_{i=1}^{N} \sum_{l=1}^{R(V_i)} \delta_l (f_{d,l}(V_i) - f_{o,l}(V_i))^2, \qquad (4.4)$$

where $f_{d,l}(V_i)$ and $f_{o,l}(V_i)$ are components of attribute vectors of the deformed model and model at vertex V_i, respectively. δ_l here denotes the importance of the lth neighborhood layers. $R(V_i)$ is the number of neighborhood layers around vertex V_i.

The proposed algorithm is optimized iteratively. In each iteration, a neighborhood of a vertex has been examined and the point in the neighborhood with the minimum model energy would be chosen as the new location of the vertex. The iterations continue until the energy converges. During the deformation, we suggest moving a surface segment as a whole, rather than a single vertex. This would avoid this risk of getting trapped in a local minimum, and also speed up the convergence. Let V_i be the vertex to be deformed during a particular iteration. The first to $R(V_i)$th neighborhood layers are about to move together as a surface segment. Suppose V_i is to move to $V_i + \Delta$ as a tentative position. Then the new position of each vertex $nbr_{l,m}(V_i)$, the mth vertex on lth neighborhood layer, is set to move to:

$$nbr_{l,m}(V_i) + \Delta \cdot \exp\left(-\frac{l^2}{2\delta^2}\right), \qquad (4.5)$$

where δ is a parameter determining the locality of the transformation. We make the deformation unchanged on the boundary of the surface segment, such that the continuity has been maintained.

The parameter $R(V_i)$ that determines the locality if the deformation is chosen to be large in the initial iteration, and is then gradually reduced to 1. Therefore, initially there are more vertices involved in the deformation. More global features are used in deformation. In later states, more local deformations are performed.
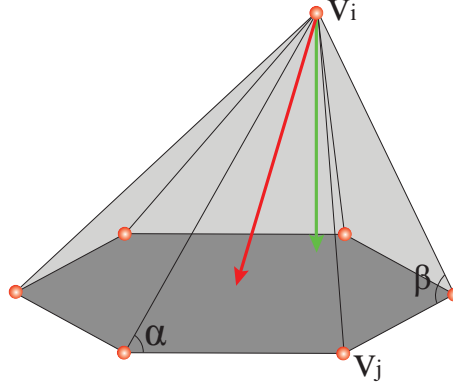
Figure 4.2: Vertex v_i and its 1-ring neighbors. The red arrow (center) is the vector obtained from the uniform weights, which points to the centroid. The green arrow (to the right) is the vector obtained from the cotangent weights ($\frac{1}{2}(\cot\alpha + \cot\beta)$), which approximates the normal.

4.4   Shape Energy

Shape energy is used to smooth the shape without losing the important details. Usually there is a tradeoff between the smoothness and keeping shape details. We have extended the Laplacian coordinate to achieve this. Let the mesh $\mathbb{M}$ of the shape be described by a pair $(\mathbb{V}, \mathbb{E})$, where $\mathbb{V} = \{v_1, ..., v_n\}$ describes the geometric positions of the vertices in $\mathbb{R}^3$ and $\mathbb{E}$ describes the connectivity. The neighborhood ring of a vertex i is the set of adjacent vertices $\mathbb{N}_i = \{j | (i, j) \in \mathbb{E}\}$ and the degree d_i of this vertex is the number of elements in $\mathbb{N}_i$. Instead of using absolute coordinates $\mathbb{V}$, the mesh geometry is described as a set of differentials $\Delta = \{\delta_i\}$. Specifically, coordinate i will be represented by the difference between v_i and the weighted average of its neighbors:

$$\delta_i = v_i - \sum_{j \in \mathbb{N}_i} w_{ij} v_j \qquad (4.6)$$

where w_{ij} is computed from cotangent weights [65] (Fig. 4.2). Assume V is the matrix representation of $\mathbb{V}$. Using a small subset $\mathbb{A} \subset \mathbb{V}$ of m anchor points, a mesh can be reconstructed from connectivity information alone. The x, y and z positions of the reconstructed object ($V_p' = [v_{1p}', ..., v_{np}']^T, p \in \{x, y, z\}$) can be solved for separately by

minimizing the quadratic energy:

$$E_{shape}(M_d, L_o) = \|M_d - L_o\| = \|L_u V_p' - \Delta\|^2 + \sum_{a \in \mathbb{A}} \|v_{ap}' - v_{ap}\|^2, \qquad (4.7)$$

where L_u is the Laplacian matrix from uniform weights, and the v_{ap} are anchor (land-mark) points. $\|L V_p' - \Delta\|^2$ tries to smooth the mesh when keeping it similar to the original shape, and $\sum_{a \in \mathbb{A}} \|v_{ap}' - v_{ap}\|^2$ keeps the anchor points unchanged. The cotangent weights approximate the normal direction, and the uniform weights point to the centroid. By minimizing the difference of these two (i.e., $L_u V'$ and Δ), the vertex is actually moved along the tangential direction. Thus the shape is smoothed without significantly losing the detail. With m anchors, (4.7) can be rewritten as a $(n+m) \times n$ overdetermined linear system $AV_p' = b$:

$$\begin{bmatrix} L \\ I_{ap} \end{bmatrix} V_p' = \begin{bmatrix} \Delta \\ V_{ap} \end{bmatrix} \qquad (4.8)$$

This is solved in the least squares sense using the method of normal equations $V_p' = (A^T A)^{-1} A^T b$. The conjugate gradient method is used in our system to efficiently solve it. The first n rows of $AV_p' = b$ are the Laplacian constraints, corresponding to $\|L V_p' - \Delta\|^2$, while the last m rows are the positional constraints, corresponding to $\sum_{a \in A} \|v_{ap}' - v_{ap}\|^2$. I_{ap} is the index matrix of V_{ap}, which maps each V_{ap}' to V_{ap}. The reconstructed shape is generally smooth, with the possible exception of small areas around anchor vertices. Different from [95], we use cotangent weights instead of uniform weights. Thus the movement along the normal direction is prevented, and shape details can be better preserved.

Although this method is able to improve mesh quality during runtime, it may still have difficulty to handle very dense and degenerated initial meshes. Furthermore, the computational efficiency can also be adversely affected by such dense meshes. Thus, we can also improve the mesh quality in the preprocess step, as shown in Fig. 4.3. Usually mesh decimation can be employed to decrease the number of vertices, and this shape energy can be used to smooth the mesh.

(a) Initial shape   (b) Mesh view   (c) Decimation

(d) Smoothing   (e) Resulting shape
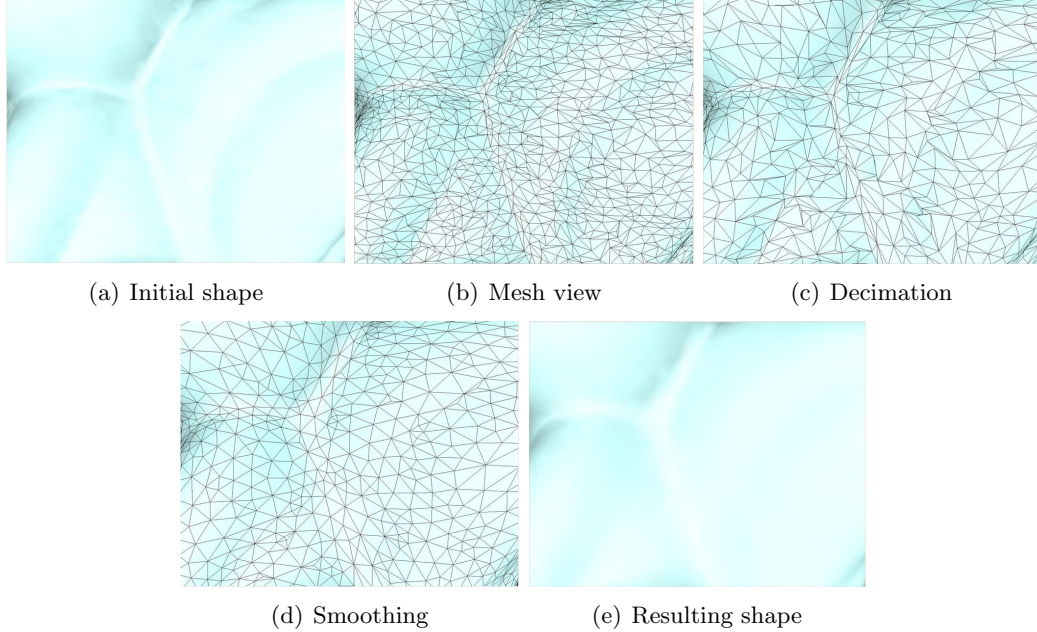
Figure 4.3: Illustration of geometry processing methods, including decimation and detail preserve smoothing. After these operations, the mesh still has similar appearance with certain level of details, while the mesh quality is highly improved.

4.5   Hierarchical Shape Statistics

The one-to-one correspondence is obtained for each vertex among all shapes after shape registration. Then the shape statistics can be computed straightforwardly using generalized Procrustes analysis and hierarchical shape prior. Given any two shapes, they can be fitted to each other using a similarity or rigid transformation. Procrustes analysis is used to find the translation, rotation and scaling components. Since there is no mean shape in the beginning, generalized Procrustes analysis arbitrarily chooses a shape to use as the reference and transforms all the rest to fit it. After that, a mean shape is computed by averaging all transformed shapes. Then, this mean shape is used as a reference shape in the next round. We repeat this procedure until the mean shape converges to a stable state. Note that normalization is necessary, as otherwise the mean shape will degenerate to a single point. After the alignment, each resulting shape is filled into a matrix as a column vector. This matrix is ready to use for the proposed SSC model (i.e., matrix D in Eq. 3.1). However, to be used by ASM, PCA is needed to get the Point Distribution Model (PDM).

To effectively model multiple structures simultaneously, we can employ shape prior hierarchically. Assume ASM type of shape priors is used. First, PCA is applied on each structure separately. Thus, shape variation of individual structure can be well discovered even with limited number of samples. Second, their relative locations with respect to the mass centroid are also modeled using PCA. This global statistics is used to place structures. Then local statistics of shapes is employed to select the important "modes" (i.e., eigenvectors corresponding to the largest eigenvalues) to cover more than 90% of the variance. Combining the mean shape and the modes, the PDM is able to summarize and describe the sample shapes concisely and accurately.

In the next section, this whole framework is employed to generate high quality training meshes for compared shape prior methods in Sec. 5. We also validate this framework using a high resolution CT reconstruction task in Sec. 5.5.

Chapter 5

Medical Applications

In this chapter, we validate our algorithms in five very diverse medical applications: 1) 2D lung localization from X-ray Image using shape inference and landmark detection; 2) 3D liver segmentation from low-dose CT using shape refinement and boundary detection; 3) 3D segmentation of rodent brain structures from MR microscope using hierarchical shape priors; 4) left ventricle tracking from cardiac MR images using 2D shape refinement; 5) reconstruction of cardiac shape atlas from high resolution CT using geometry processing and shape registration methods.

5.1 2D Lung Localization from X-ray Image

5.1.1 Clinical Background

Radiography (X-ray) is the most frequently used medical imaging modality due to its fast imaging speed and low cost. About one third radiograph exams are chest radiographs. It is used to reveal various pathologies including abnormal cardiac sizes, pneumonia shadow and mass lesions. The automation of pathology detection often requires robust and accurate lung segmentation. The major challenges of lung segmentation in radiography come from large variations of lung shapes, lung disease and pseudo-boundary close to diaphragm. In chest X-ray, the position, size and shape of lungs often provide important clinical information. Therefore, in this experiment we try to locate the left or right lung using landmark detection and shape inference. Out of 367 X-ray images (all images are from different patients), 200 are used as training data, and the rest 167 are used for testing purpose. In this study, we select training samples to

ensure a good coverage of different ages and genders (according to information from DI-COM header.) The number of training samples is determined empirically. The ground truths are binary masks of manual segmentation results. A 2D contour is extracted from each mask. To obtain the landmarks for training purpose, we manually select six specific points (e.g., corner points) on the contour, and then evenly and automatically interpolate a fixed amount of points between two neighboring landmarks along the contour. Thus a rough one-to-one correspondence is obtained for both landmarks and shapes. Since the detected landmarks may not be accurate or complete, shape prior is necessary to infer a shape from them. When applying this model, we constantly use the same parameter values for all X-ray images, i.e., $\lambda_1 = 50$ and $\lambda_2 = 0.15$.

5.1.2   Compared Methods

In this study, we compare the proposed sparsity-based shape prior modeling with other state-of-the-art methods. For a fair comparison, we intentionally embed different shape models to the same organ localization framework. (It is not fair to compare completely different end-to-end systems, e.g., our system vs. ASM system, since the performance difference, if any, cannot be solely attributed to shape prior modeling.) More specifically, the same learning-based algorithm [90] is used to detect landmarks for shape inference (organ localization). Furthermore, the shape is inferred from the detected landmarks directly without iteratively deforming and fitting to the image boundary. The reasons of this setting are twofold. First, enough times of iterative deformations may eventually bring the shape to the image boundary accurately because of the deformation strategy. It is then difficult to evaluate the performance of shape prior models. Thus we apply the inference method only once without the deformation. Second, such one-step inference process is very fast and already good enough as the input for some clinical applications, such as a CAD program. The compared methods are listed as the following:

1. PA: Procrustes Analysis is used to find a similarity transformation to fit a mean shape to detected landmarks.

2. SMS: It is the Shape Model Search module in ASM, which employs the PCA method to refine the input shape. Note that we are not using the entire ASM framework including boundary detection and iterative fitting. We focus on the key module of ASM inducing shape prior information. Since the detected landmarks are very sparse compared to the whole contour, directly fitting ASM to the small number of landmarks result in poor performance. To achieve reasonable performance, a contour is approximated by interpolating points in-between landmarks. When there is a point missing, the mean position of that point is used instead.

3. R-SMS: The shape model search step in the robust ASM [68] method uses the RANSAC framework to remove the influence of erroneous detections. When there is no outlier, R-SMS achieves similar performance as the traditional ASM.

4. SI-NN: It stands for shape inference using k nearest neighbors. It is similar to [32], which uses nearest neighbors to find the closest prototypes in the expert's structure annotations. The distance metric we used is based on the $L2$ distance between corresponding points.

5. TPS: Thin-plate-spline [4] is used to deform the mean shape to fit detected landmarks. TPS is a non-rigid and local deformation technology and has been used in robust point matching application (TPS-RPM) [9]. Since the deformation can be nonrigid and local, it is able to produce any type of shapes to fit the detection results.

6. SSC(3.4): It is the sparse learning shape method without modeling e. The result is computed by solving Eq. 3.4. Thus only sparse linear combination assumption is employed.

5.1.3   Visual Comparisons

Some representative and challenging cases are shown in Fig. 5.1, 5.2 and 5.3. In Fig. 5.1, there are some mis-detections which are considered as gross errors. The Procrustes
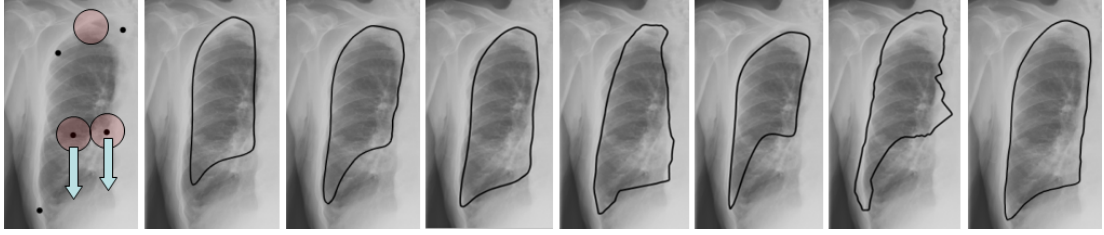
(a) Detection   (b) PA   (c) SMS   (d) R-SMS   (e) SI-NN   (f) TPS   (g) SSC(3.4)   (h) SSC

Figure 5.1: Comparisons of the right lung localization. (a) Detected landmarks are marked as black dots. There are two detection errors and one point missing (marked as circles, and the arrows point to the proper positions). (b) Similarity transformation from Procrustes Analysis. (c) Shape Model Search module in ASM, using PCA based method. (d) Shape Model Search in Robust ASM, using RANSAC to improve the robustness. (e) Shape inference method using nearest neighbors. (f) Thin-plate-spline. (g) Sparse representation without modeling e, by solving Eq. 3.4. (h) The proposed method by solving Eq. 3.5.
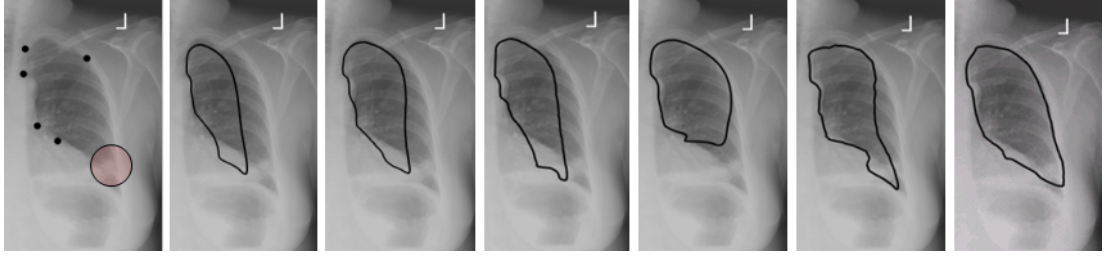
(a) Detection   (b) PA   (c) R-SMS   (d) SI-NN   (e) TPS   (f) SSC(3.4)   (g) SSC

Figure 5.2: Comparisons of the left lung localization. There is one point missing (marked by a circle), and this lung has a very special shape, which is not captured by the mean shape or its variations. Compared methods are the same as Fig. 5.1.
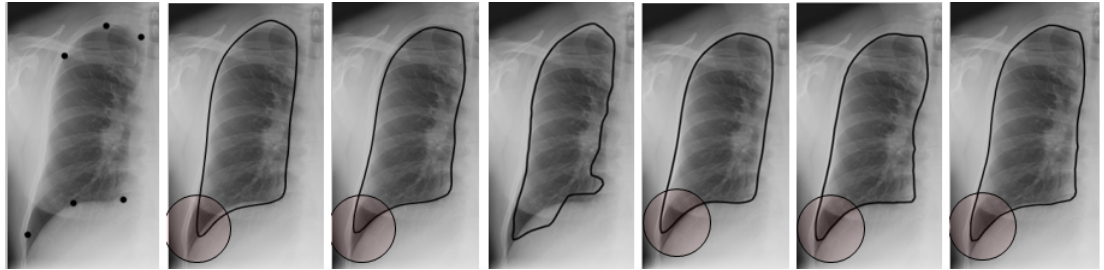
(a) Detection   (b) PA   (c) R-SMS   (d) SI-NN   (e) TPS   (f) SSC(3.4)   (g) SSC

Figure 5.3: Comparisons of the right lung localization. All six detections are roughly accurate. Thus there is no gross error. The regions marked by circles show the difference of preserved details. Compared methods are the same as Fig. 5.1.

analysis, SMS method, SI-NN algorithm and TPS cannot handle such cases. R-SMS is not sensitive to outliers and performs better. SSC(3.4) also fails to handle such non-Gaussian errors since e is not modeled. SSC can successfully capture such mis-detected points in e and generate a reasonable shape. In Fig. 5.2, the underlying shape of the lung is special and different from most other lung shapes (see the mean shape in Fig. 5.2 (b)). Furthermore, there is a missing point. Neither a transformed mean shape nor its variations can represent such shape. TPS is very flexible and able to generate special shapes. However, it fails to handle the missing point. SSC roughly captures the correct shape and generates a better result than the others. In Fig. 5.3, all six detections are correct. However, the shape's details are not preserved using the mean shape or its variations. Fig. 5.4 shows the first five modes (i.e., the five largest shape variations) of ASM. The thickness of boundaries represents major variations in that mode. Note that since the variations of the bottom left tip are not the major variation modes, ASM does not able to preserve these local shape details for this testing case in Fig. 5.3. Both SSC(3.4) and SSC discover more detail information than other methods. Thus a sparse linear combination is sufficient to recover such details even the gross error e is not modeled. Fig. 5.5 shows three SSC-selected shape components with largest weights which generate the result in Fig. 5.3. Two of them do have certain levels of detail information in the bottom left region, although they are still different from the input shape. It demonstrates that our model can discover meaningful shape components, and the final shape composition result can well approximate the testing data. Please note that the proposed model cannot "create" local shape details without the support of local appearance cues. Instead, our method aims to "preserve" shape details derived by appearance cues given these details exist in our shape repository. In particular, our method is able to preserve local shape details even when they are not statistically significant in the shape space.

Fig. 5.6 shows some results from our proposed method on challenging cases with medical instruments. Shape prior contributes to the stableness of the system. It still generates reasonable results with such misleading appearance cues. Fig. 5.7 shows a failure case of our method. Only three landmarks are detected. Using such information

(a) 1st mode   (b) 2nd mode   (c) 3rd mode   (d) 4th mode   (e) 5th mode
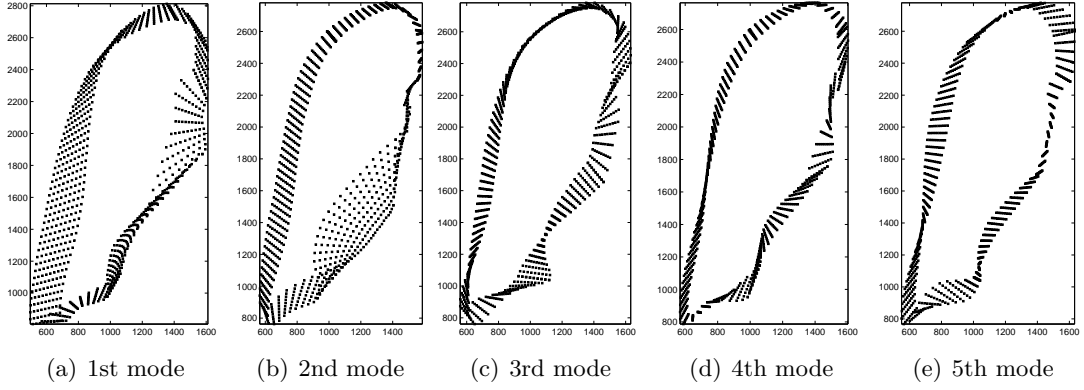
Figure 5.4: Five largest ASM modes of the right lung. In each mode, we plot shapes from -3σ to 3σ variance. The thickness of boundaries represented shape variation. For example, the major variation of the second mode is in the bottom right, which is the boundary between the right lung and the cardiac. Note that the variance of the bottom left tip is not the major variation in any mode. Thus ASM does not achieve very accurate performance for the case in Fig. 5.3.

(a) Output   (b) Shape 1   (c) Shape 2   (d) Shape 3
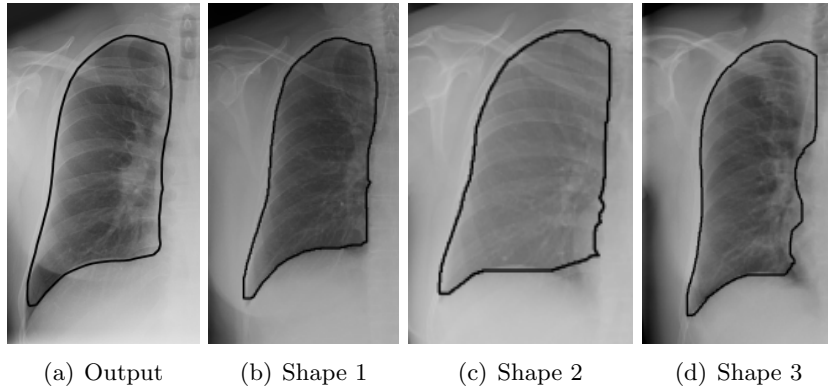
Figure 5.5: Three shape components with largest weights from our model. "Output" means the result in Fig. 5.3. The three components have weights 0.5760, 0.2156, and 0.09822, respectively.
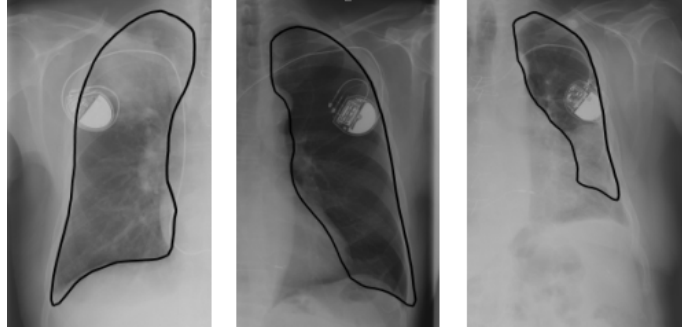
Figure 5.6: Some localization results from our proposed method on challenging cases with medical instruments. Note that the localized shape may not be exactly on the boundary, since the shape module does not use image information. However, such results are good enough for the input of CAD program or initialization of segmentation algorithms.
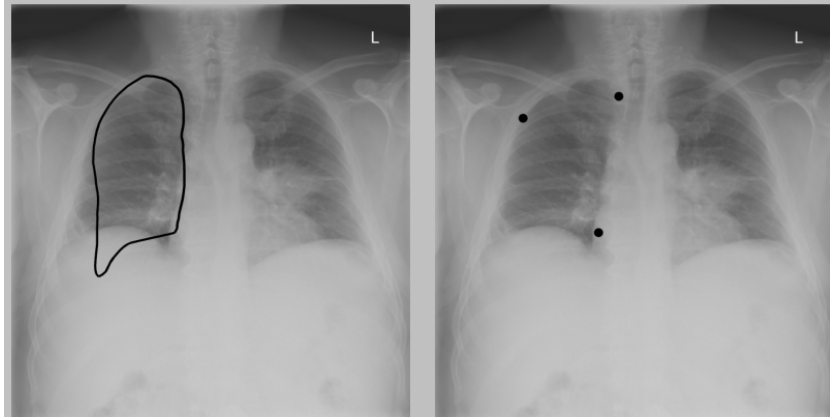
Figure 5.7: One failure case of lung localization. Only three landmarks are detected. Such information is not enough to infer a correct shape in this case.

fails to infer a reasonable shape in this case, although our method supports incomplete input. In general, less input results in a higher probability for inaccurate predictions.

5.1.4 Quantitative Comparisons

To quantitatively compare different methods, we report the mean values and standard deviations of sensitivity and specificity between binary masks in Table 5.1 and Fig. 5.8. Note that the specificity is always good in all methods. The reason is that the size of either left or right lung is relatively small compared to the whole chest X-ray image. Hence, most "true negative" can be correctly found. Thus we also report Dice Similarity Coefficient (DSC) [67], which is a good complement to the other two measurements.

	Fig. 5.1			Fig. 5.2			Fig. 5.3		
Method	P	Q	DSC	P	Q	DSC	P	Q	DSC
PA	62	99	76	50	99	64	93	99	94
SMS	66	99	78	61	99	72	93	99	95
R-SMS	81	99	88	61	99	72	93	99	95
SI-NN	81	99	87	63	98	73	87	99	90
TPS	59	99	74	75	99	79	97	98	94
SSC(3.4)	63	98	71	73	99	79	97	99	96
SSC	**87**	**99**	**91**	**92**	**99**	**91**	**98**	**99**	**96**

Table 5.1: Quantitative comparisons of seven methods. The sensitivity (P%), specificity (Q%) and Dice Similarity Coefficient (DSC%) are reported for cases in Fig. 5.1, 5.2 and 5.3.

DSC is defined as: $2 \times TP/(2 \times TP + FP + FN)$, where TP, FP and FN stand for true positive, false positive and false negative, respectively. Generally Procrustes analysis, TPS and SMS achieve good performances, especially when landmarks are correctly detected. However, they are sensitive to non-Gaussian errors. R-SMS can handle this problem because of the RANSAC method, but sometimes it fails to deal with the multimodal distribution of shapes. SI-NN is a good nonparametric method. However, it may not be able to represent shapes which do not appear in the training data. The sparse linear combination by SSC(3.4) can approximate such shape and it generally performs better than the others. Without modeling error e, this method still fails to recover a correct shape. In our proposed method, although the parameter for e is set to a relatively small value, it still contributes to the performance of the model. It performs the best in terms of sensitivity and DSC, without sacrificing the specificity. The standard deviations in Fig. 5.8 show that SSC also achieves the best stability among all compared methods.

We also conducted experiments on dictionary learning. Starting from 200 training cases, we use K-SVD to learn a 64 column dictionary. Using this dictionary for sparse shape composition, the accuracy of lung detection drops for around 1.5%, which is acceptable considering the advantages of using a much smaller data matrix.

The experiments are performed on a PC with 2.4GHz Intel Quad CPU, 8GB memory, with Python 2.5 and C++ implementations. The whole framework is fully automatic. As it benefits from the FISTA algorithm, our algorithm is very efficient. Given
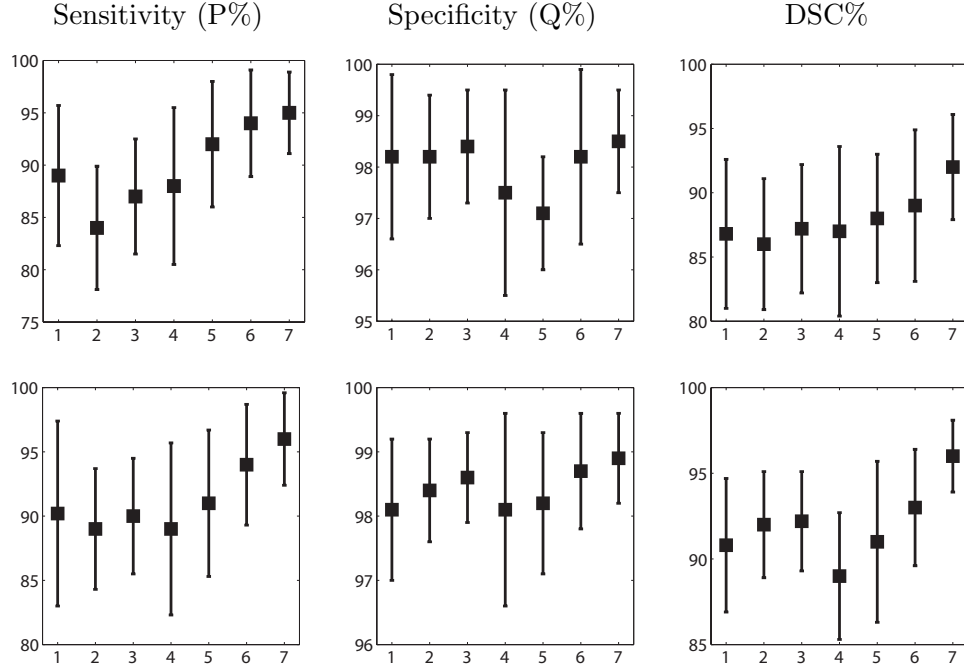
Figure 5.8: Mean values (μ) and standard deviations (σ) of P, Q and DSC from the left lung (1st row) and right lung (2nd row) of all testing data. In each figure, y-axis is the performance of P, Q or DSC. x-axis means 7 methods (from left to right) in the same order as Table 5.1. Squares denote μ, segments denote σ.

this scale of data, it takes around 0.5 second to infer a shape from the landmarks.

5.1.5 Evaluation of Parameter Sensitivity

We also conducted experiments of the parameter sensitivity of λ_2 (Fig. 5.9), which controls the sparsity of erroneous detections e. A set of parameter values is evaluated for data with or without erroneous detections. Our model consistently achieves similar performance for data without erroneous detections. In these cases, the sparse linear combination of training shapes can already approximate the input shape very well. Thus values in e are usually flat zeros in these cases, even with different values of λ_2. For data with erroneous detections, λ_2 is critical as it controls the number of nonzero elements in e. Our model still generates stable results in a reasonably wide range of parameter values, although the performance eventually goes down when λ_2 is far away from this range. This insight analysis shows that this model is not sensitive to

(a) Sensitivity (P%)

(b) Specificity (Q%)

(c) DSC%
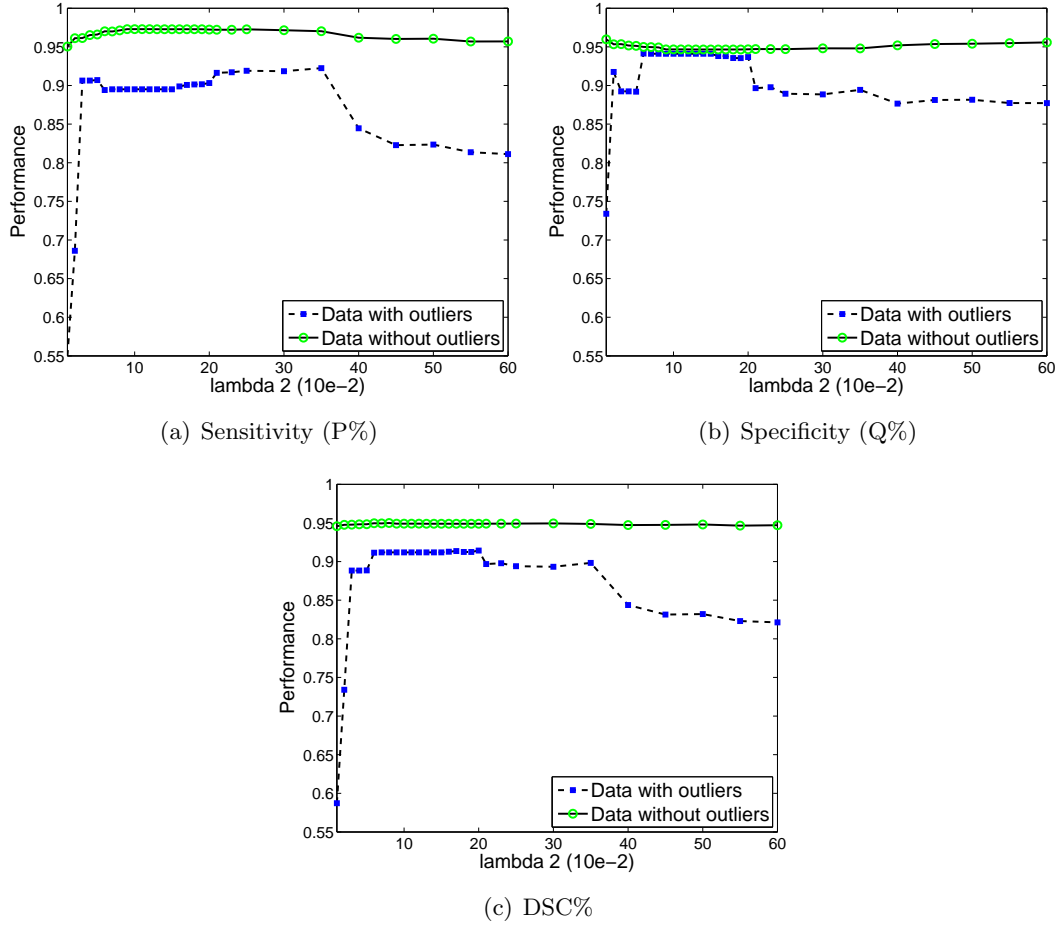
Figure 5.9: Parameter sensitivity of λ_2. A set of parameter values is tested for data with or without erroneous detections. Generally data without erroneous detections (green line) consistently achieves similar performance for different parameter values, and data with erroneous detections (blue line) also produce stable results in a wide range.

parameter values in a certain level, which makes it general for different data in the same application. Another observation is that when there is no outlier, this model usually generates slightly better results than having erroneous detections, which means that erroneous detections still adversely affect the model performance even with this sparse gross errors constraint.

5.1.6 Summaries

In this section, we have applied the proposed SSC model on the 2D lung localization problem. Three representative cases are visualized to validate the advantages of our algorithm, i.e., model complex shape variations, handle gross errors and preserve shape details. Quantitative results are also provided with a large amount of dataset. Both results show the superior of our method. Furthermore, insight experiments are designed to analyze the sensitivity of parameters. The results demonstrate that our model works well with a large range of values. Benefited from this property, SSC usually just needs to use the same parameter values for the same application.

5.2 3D Liver Segmentation from Low-dose CT

5.2.1 Clinical Background

Whole body PET-CT is an emerging medical imaging modality that combines a Positron Emission Tomography (PET) and an x-ray Computed Tomography (CT) scan. The co-registered anatomical (CT) and functional (PET) information benefits various clinical practices, especially for oncology studies. Due to the high variations of F-fluorodeoxyglucose (FDG) uptakes across different organs, the preferred way to interpret PET-CT images is in an organ-specific fashion, which requires organ segmentation. In traditional CT images, organ segmentation such as liver segmentation [56] has been extensively investigated. However, to decrease radiations to patients, CT images in PET-CT scans usually have low dose and large slice thickness, which result in low contrast and fuzzy boundaries between organs. Hence, organ segmentation in whole body PET-CT becomes more challenging than traditional CT. In this experiment we try to segment the

liver from the low-dose whole body CT, using deformable models and shape refinement [96]. The 3D ground truth of low-dose CT is manually segmented by multiple clinical experts. 67 scans are annotated. 40 are used as training data to train the landmark detector and also used to construct the data matrix D_S, the rest 27 are testing. To obtain the one-to-one correspondence for vertices among all shapes, we choose one shape as a reference and register it to all the others using adaptive-focus deformable model (AFDM) [71]. The shape has around $1,000$ vertices, and 20 are selected as landmarks. The initialization step is based on landmark detection and similar to Sec. 5.1. A surface model is then fitted into the image, and is hierarchically deformed to the image gradient information [89]. During the deformation procedure, our sparsity-based shape prior modeling is used to refine the shape, which is actually a regularization step to avoid the deformable model getting stuck in local minima of the image information. This method is fully automatic without any manual interaction.

In this study we use $\lambda_1 = 50$ and $\lambda_2 = 0.3$ for all data. For comparisons, we also employ some other shape prior modeling methods (e.g., SMS, SI-NN) to refine the intermediate deformation result after several iterations. For a fair comparison, the results for comparisons are based on the same segmentation framework, i.e., using the same landmark detection and hierarchical deformation algorithm, with different shape prior modeling.

5.2.2 Visual Comparisons

Fig. 5.10, 5.11, 5.12 and 5.13 show some visual comparisons in 3D. Fig. 5.10 compares the shape inference based initialization using the same landmark detection input. Since the image contrast of low-dose CT is very low, landmark detector may easily fail to locate correct positions. SSC is less sensitive to such errors. Its initialization result is already very close to the object boundary, which can benefit the deformation procedure. Fig. 5.11 shows the deformation results using the initialization in Fig. 5.10. The deformation modules are all the same. Shape refinement is not used in this experiment. Better initialization not only ensures fewer iterations of deformation but also produces more accurate results. SSC obtains better results compared to other methods. Fig. 5.12
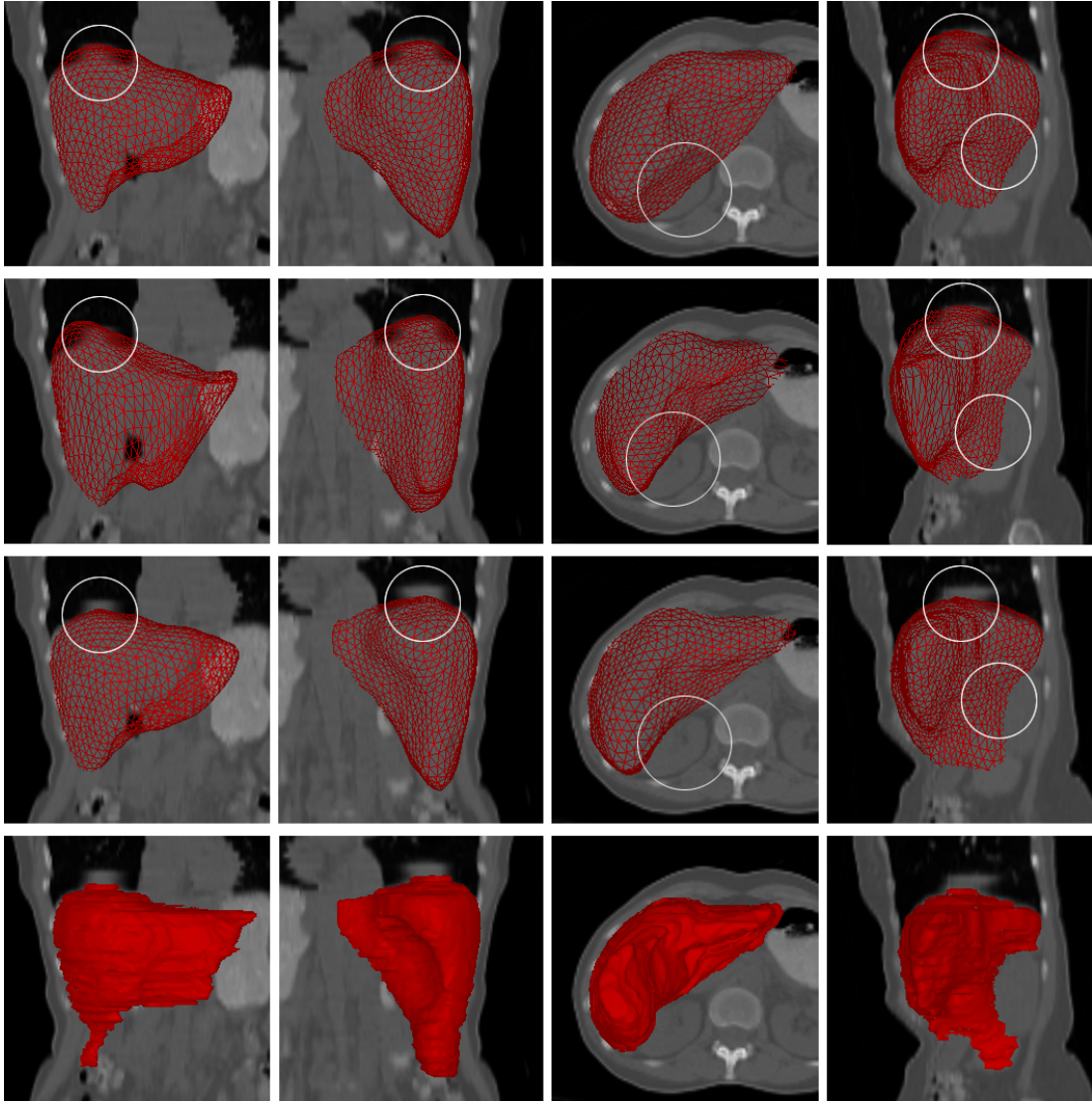
Figure 5.10: 3D initialization results for the segmentation framework. First row: using the global transformation based on Procrustes analysis. Second row: using TPS which is a local and nonrigid deformation technique. Third row: using SSC to constrain the shape. Fourth row: ground truth from manual segmentation. Its surface mesh is obtained by applying iso-surface method on the binary image. The results from global and local deformation incorrectly include part of the liver or the artifact caused by breath. The differences are marked by circles. Results from SSC are more similar to the ground truth.
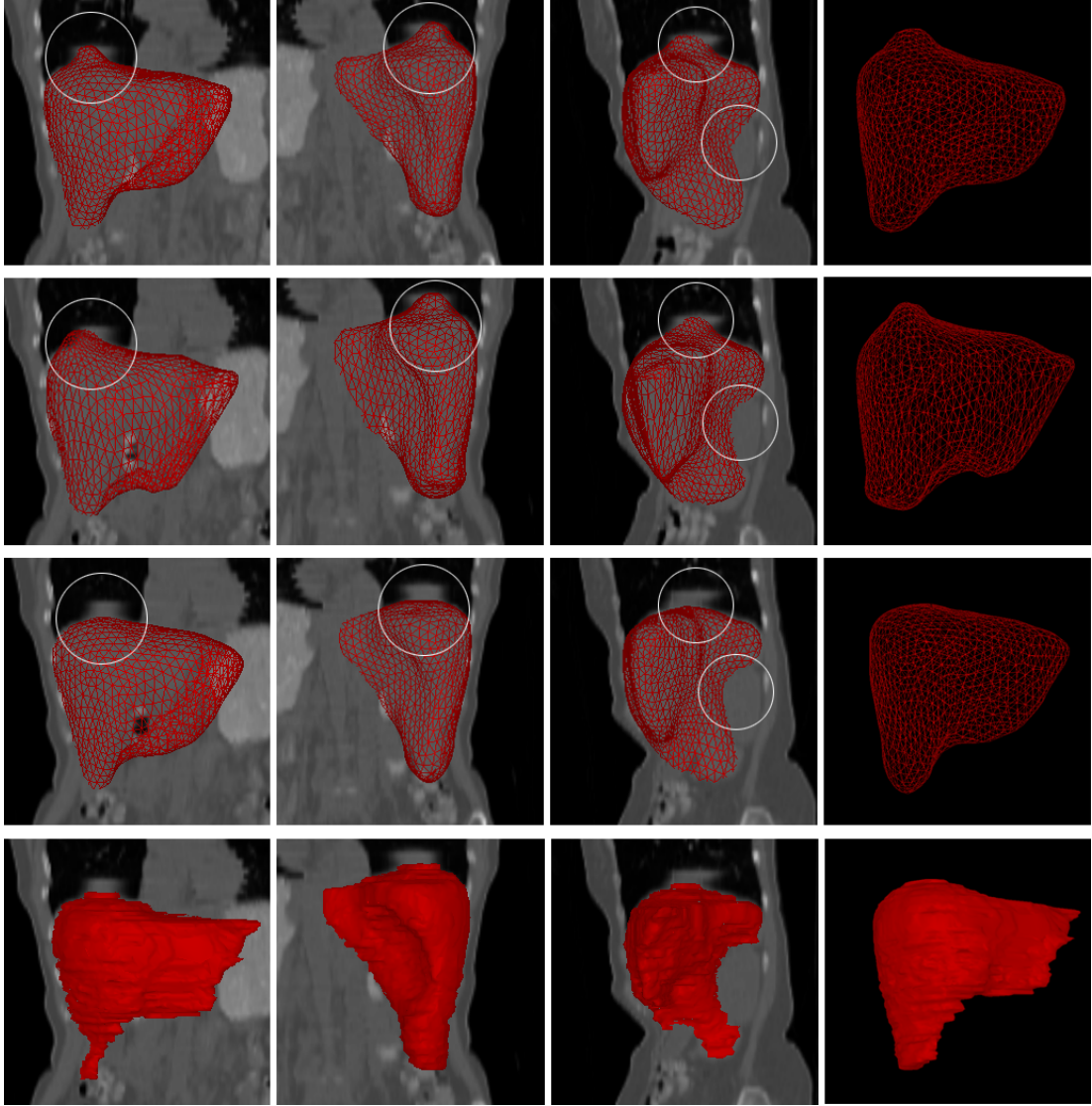
Figure 5.11: Visual comparisons of deformation results using the initialization from Fig. 5.10. Note that the deformation modules are the same. First row: using initialization from the global transformation. Second row: using initialization from TPS. Third row: using initialization from SSC. Fourth row: the same ground truth as in Fig. 5.10. Generally better initialization ensures fewer iterations of deformation and more accurate results. The differences are highlighted by circles.
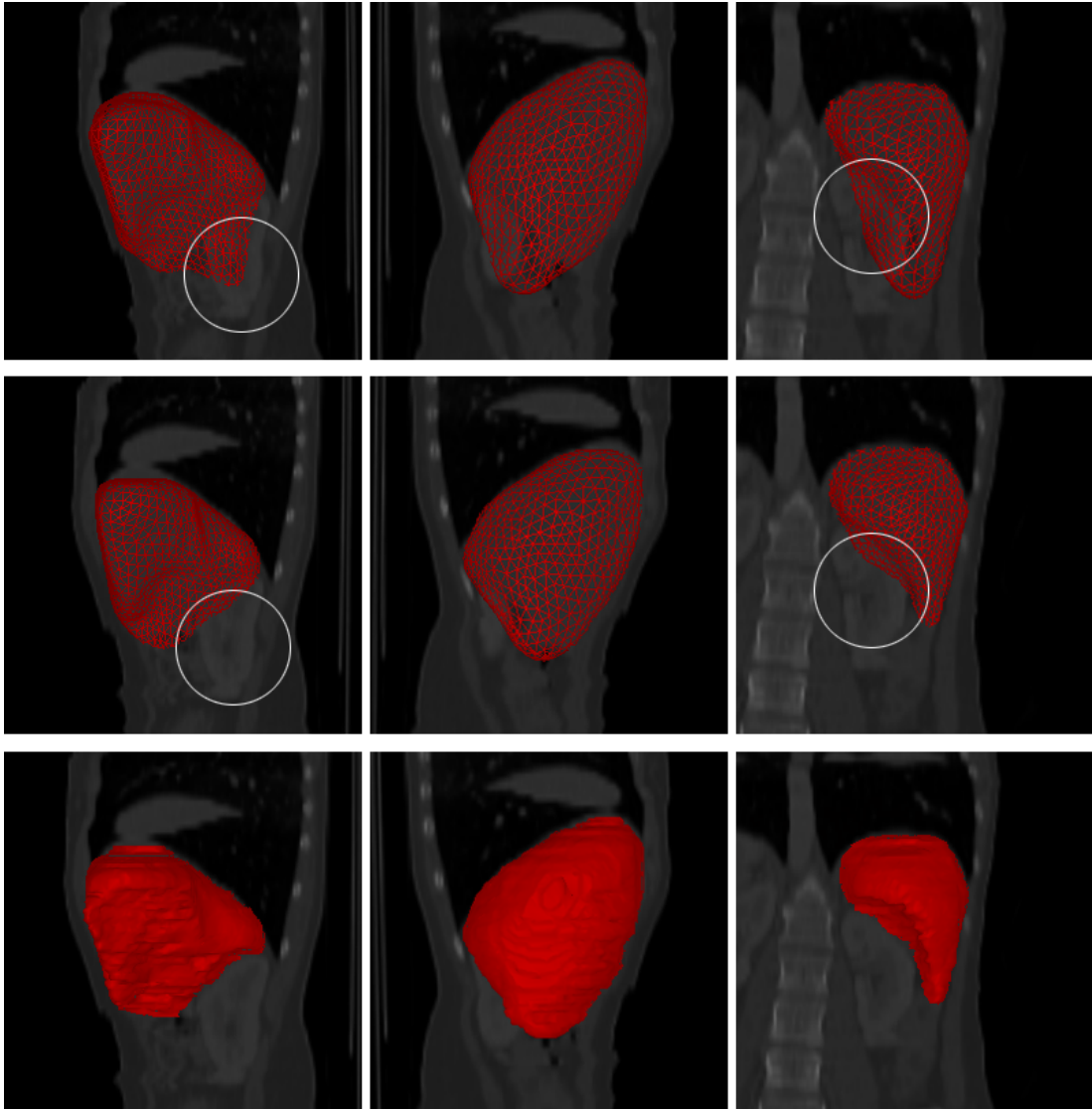
Figure 5.12: Visual comparisons of refinement results after deformation. First row: using shape model search step in SMS as the refinement. Second row: using SSC. Third row: ground truth from manual segmentation. The refined shape may not be exactly on the image boundary since this part is just a regularization step without considering any image information. The SMS result incorrectly includes a large part of the liver (marked by circles).
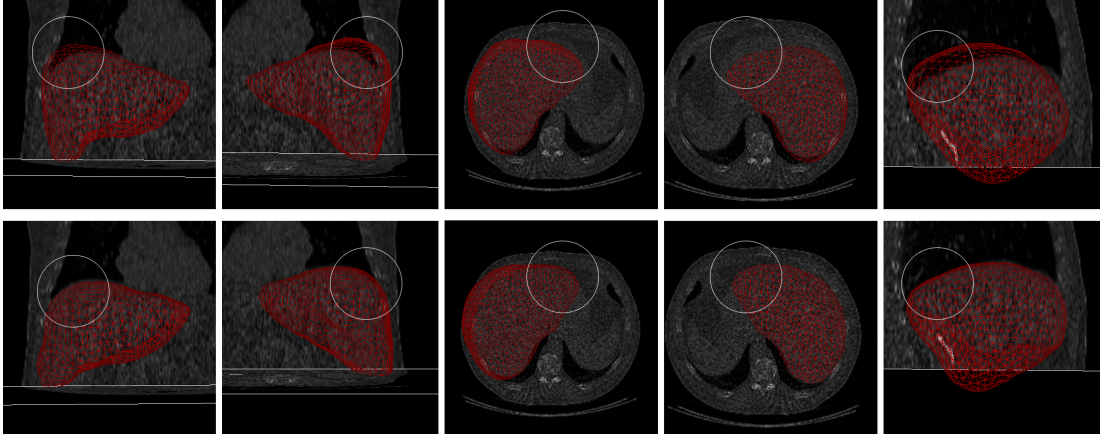
Figure 5.13: Visual comparisons of final segmentation results from highly noisy data. First row: deformation without shape constraint. Second row: using SSC as the shape constraint. Compared to the system not using shape prior, SSC is less sensitive to image noise and is more robust. The differences are marked by circles.

compares the shape refinement results after deformation. The refined shape may not be exactly on the image boundary since this part is just a regularization step without considering any image information. After the refinement, the segmentation framework still needs to perform some iterations of deformation to reach the image boundary. Certainly, a better refinement result can benefit the whole segmentation framework. The refined shape of ASM type method follows the mean shape and variations, but it incorrectly includes a large part of the kidney. SSC is more specific to this image and is more accurate. Fig. 5.13 shows the final segmentation results from a highly noisy data. Without any shape prior information, the method fails to provide an accurate result. Using proposed method, the system is more robust and less sensitive to image noise. Thus it provides more accurate results. Fig. 5.14 shows one failure case of our shape prior model. The landmark detectors generate many inaccurate predictions due to the misleading image appearance cues. In such cases, our sparse shape model may not be able to rectify mis-detections.

5.2.3 Quantitative Comparisons

Table 5.2 shows the quantitative comparisons. To evaluate 3D accuracy, we report the mean value and standard deviation of the distances between shape surfaces. SSC
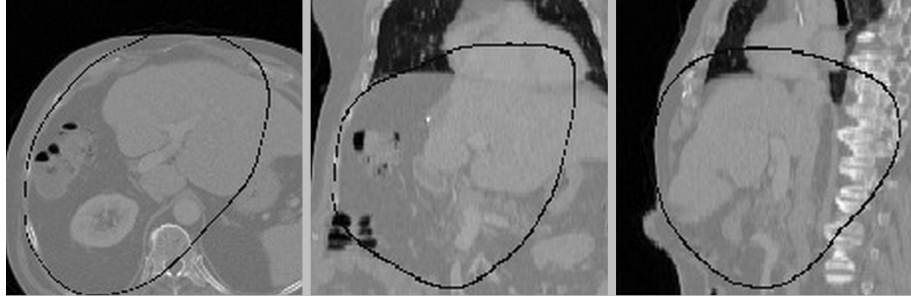
Figure 5.14: A failure case of the lung segmentation. Because of the misleading image appearance cues, the landmark detectors provide many inaccurate predictions, which adversely affect the performance of our shape prior model.

Method	Fig. 5.10	Fig. 5.12	All data
SMS	2.26 ± 1.72	1.81 ± 2.10	2.16 ± 1.68
SI-NN	4.88 ± 3.61	3.34 ± 3.78	3.82 ± 3.12
TPS	2.92 ± 2.19	5.12 ± 5.29	3.39 ± 3.16
SSC(3.4)	1.42 ± 1.02	2.39 ± 2.31	2.24 ± 1.70
SSC	$\mathbf{1.31 \pm 0.95}$	$\mathbf{1.10 \pm 0.87}$	$\mathbf{1.13 \pm 0.83}$

Table 5.2: Quantitative comparisons of the mean values and standard deviations of the distances (voxel) between surfaces.

achieves the best performance compared to other shape refinement approaches. The shape refinement step takes several minutes when applied to the whole surface directly. The whole system takes around 20 seconds (a Python implementation on a PC with 2.4GHz Intel Quad CPU) to segment liver in a $512 \times 512 \times 300$ CT volume, including data loading, initialization, deformation and shape refinement. Note that the shape refinement module not only improves the robustness of the deformable model, but also decreases the iteration times of deformation since it helps avoid local minima of image information.

5.2.4 Summaries

To summarize, we have applied SSC to segment 3D liver from low-dose whole body CT. Our segmentation framework has an initialization module and a deformation module. Both use SSC as shape constraint. Since SSC is independent of data dimension (e.g., 2D contour or 3D mesh), we simply use the same setting as in the lung application. The results show that SSC is more effective and more robust to outliers than other

compared methods. In terms of the whole segmentation framework, shape prior module can improve the robustness towards image noise.

5.3 Multiple Rodent Brain Structures from MRM

5.3.1 Clinical Background

Magnetic resonance imaging (MRI) at a spatial resolution of at least 100um in one dimension is frequently referred to as MR microscopy (MRM) and is currently available with the use of high magnetic field images. This technical achievement has permitted the detailed anatomical study of the rodent brain, which is much smaller than the human brain ($2.8cm^3$ rat brain volume vs. $1,500cm^3$ human brain volume), and requires a small voxel size ($0.001mm^3$ for rodent vs. $1mm^3$ for human brain), in order to be imaged. Rodents are often used as models of human disease not only because they frequently exhibit key features of abnormal neurological conditions but also because they are a convenient starting point for novel studies. The analysis of rodent brain image faces similar challenges to human imaging, with individual variation in size, morphology, and topology of the brain structures complicating the neuroanatomical studies. Such analysis is frequently performed by segmenting the regions of interest (ROI) in rodent brain images.

The challenge of this segmentation task is threefold. 1) The image information is sometimes incomplete or misleading. For example, there is no obvious boundary in part of the striatum, and the cerebellum contains interleaving texture (Fig. 5.15), 2) It is important to discover and/or preserve some complex local shape details, such as the paraflocculi (i.e., two protruding features of the cerebellum), 3) Given limited number of training samples, it is desirable to segment all interested structures simultaneously by effectively learning a shape atlas for multiple structures.

In this work, we use an online robust deformable models [47, 75] and hierarchical shape priors to segment multiple rodent brain structures [93]. A hierarchical modeling scheme can learn shape statistics from a small number of training data. Note that the idea of hierarchical shape prior has been investigated much in the literature [15, 27].
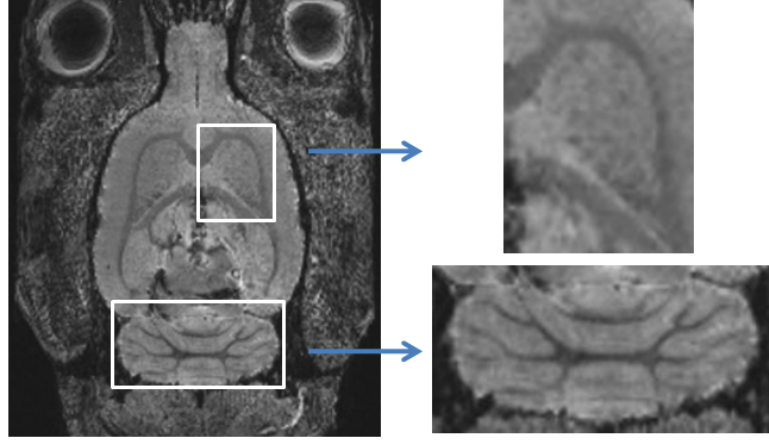
Figure 5.15: The MR image of a rat brain used in experiments. The image information is misleading because 1) part of the boundary between striatum (top) and other structures is blurred; 2) the cerebellum (bottom) has complex textures and large gradient values inside.

In the experiments, adult male Sprague-Dawley rats were transcardially perfused with 4% paraformaldehyde. Heads were stored in paraformaldehyde and scanned for magnetic resonance microscopy. The brains remained in the heads during scanning in order to avoid tissue and shape distortions during brain extraction. The heads were scanned on a 21.1T Bruker Biospin Avance scanner (Bruker Biospin Corporation, Massachusetts, USA). The protocol consisted of a 3D T2-weighted scan with echo-time (TE) $7.5ms$, repetition time (TR) $150ms$, 27.7 kHz bandwidth, field of view (FOV) of $3.4 \times 3.2 \times 3.0mm$, and voxel size $0.08mm$, isotropic. 3D annotation is manually performed on 58 volume data (overall $58 \times 512 = 29,696$ slices with resolution 512×512) by an expert neurologist. 8 are used as training data since we only need a small number of samples for hierarchical statistics. The rest 50 are used as testing. The proposed method was implemented in C++ and Python 2.6 and tested on a 2.40 GHz Intel Core2 Quad computer with 8G RAM. Fig. 5.16 shows the shape modeling results for multiple structures.

5.3.2   Segmentation Framework

In this multiple structure segmentation task, we do not use landmark and boundary detection as in Sec. 5.1 and 5.2. An online robust deformable model is employed along
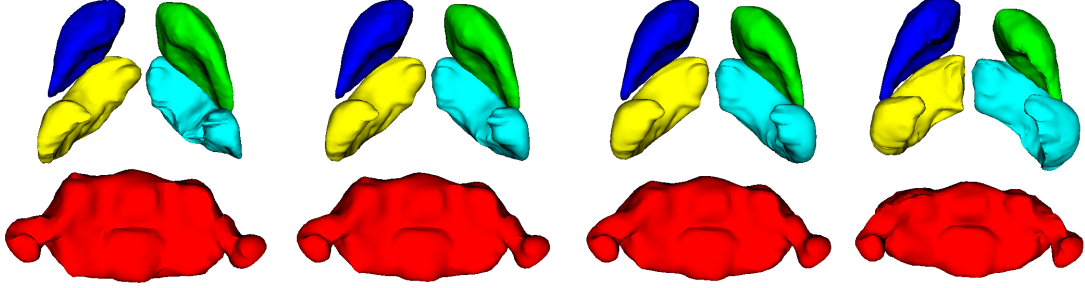
Figure 5.16: The largest shape variations of rodent brain structures, which represents the changing of the volume magnitudes. From Top to bottom: left and right striatum, left and right hippocampus, cerebellum.

with the hierarchical shape priors. Thus we introduce this segmentation framework and some relevant algorithms in this section.

Given a test 3D image, it is aligned to the reference brain. Then the mean mass centroid of all structures is used as the estimated center. The mean shape of each structure is placed according to the relative positions learned at the high level statistics. However, this initialization may not be close to the boundary of the testing data because of the variance. Thus deformable models are still needed for accurate segmentation. In order to find the boundary robustly, the initialized models are driven by both gradient and region information derived from the image. Region information alleviates the problems caused by unclear boundaries and complex textures. The overall energy function is:

$$E = E_{int} + E_{ext} = E_{int} + (E_g + k_R \cdot E_R) \tag{5.1}$$

where E_{int} is the internal (smoothness) energy, E_{ext} is the external (image) energy, E_g is the gradient term, E_R is the region term, k_R is a constant to balance the contributions of the two external energy terms. The balance between the internal and external energies is naturally controlled by the smoothness factor in the stiffness matrix by using Finite Element Method (FEM) as the deformation scheme [75].

The traditional gradient data terms usually include the gradient map, edge distance map, or a combination of both. Denote a gradient magnitude map or the distance

transform of an edge map as F_g, the gradient data term in our system is defined as:

$$E_g = \int_\Lambda F_g(\mathbf{x})d\Lambda, \ F_g = D_{edge}^2 \ or \ F_g = -\left|\nabla I\right|^2 \tag{5.2}$$

where Λ denotes the surface mesh, D_{edge} refers to the unsigned distance transform of the edge map, and ∇I represents the image gradient.

The region term encodes constraints for the model-interior appearance statistics. Considering a module using intensity statistics, the object region is predicted according to the current model-interior intensity distribution. Having both foreground object and background probabilities, we obtain a binary map that represents the predicted object region by applying the Bayesian Decision rule. Connected component analysis is then applied to the binary map to retrieve the connected component that overlaps the current model. This connected region is considered as the current ROI. Let us denote the signed distance transform of the current model's surface shape as Φ_Λ, and the signed distance transform of the ROI boundary shape as Φ_R, the region-based external energy term is defined using voxels within a narrow band around the model surface as:

$$E_R = \int_\Lambda \Phi_\Lambda(\mathbf{v})\Phi_R(\mathbf{v})d\Lambda \tag{5.3}$$

The multiplicative term provides two-way balloon forces that deform the model toward the predicted ROI boundary. This allows flexible model initializations either overlapping the object or inside the object. Using these external energy terms as image forces, the FEM model is driven to find object boundaries.

The above model may still not be able to avoid the local minimum or preserve a specific shape, especially when the texture of the image is complex. Thus a shape prior energy is added to our method to constrain and refine the shape and position of all n structures during deformation:

$$E_{prior} = \sum_{i=1}^{n} E_{local}^i + k_g \cdot E_{global} \tag{5.4}$$

where E_{local}^i is the shape constraint applied on the ith structure, and E_{global} is a constraint on the relative positions among different structures and distances with respect to the mass centroid.

In terms of the implementation, the energy optimization is achieved by using an expectation-maximization (EM) type of framework. Given intermediate segmentation results of all structures, in the "E" step, the locations of structures are adjusted as per the global statistics to prevent overlap or intersection. In the "M" step, each individual shape is constrained by aligning to the mean shape, and then mapping into PCA space to update the pose and shape parameters. Thus it guarantees that the shape deforms only into patterns consistent with the training data, which can refine the shape and prevent over-segmentation. This whole hierarchical framework is employed adaptively by defining $k_g = 1 - e^{-|\nabla \Lambda|}$, where $|\nabla \Lambda|$ is the magnitude of deformation change in surface shape. In the beginning, the shape deforms a lot so k_g is relatively large. It means that we trust more on the high level information, and put larger weights on the global statistics. After several iterations, the local statistics should be more important since the brain structure is nearby the boundary. At this time the shape deforms less so k_g is smaller. The whole segmentation framework, including training and testing, is summarized as follows:

1. Manually segment a small number of training data. Use PCA to capture shape statistics of these 3D shapes and location statistics of relative positions.

2. Given a testing data, it is aligned to the reference image. Then multiple models are placed using mean relative positions with respect to the mean mass centroid. Initialize these models, i.e., stiffness matrix and step size for FEM and the gradient magnitude or edge map.

3. Compute Φ_Λ based on the current model; predict object ROI R by applying the Bayesian Decision rule to binarizing the estimated object probability map, and compute Φ_R. Calculate the external force vector.

4. Deform the model using FEM and external forces derived from gradient and appearance information. Smoothness constraint is implicitly incorporated by FEM.
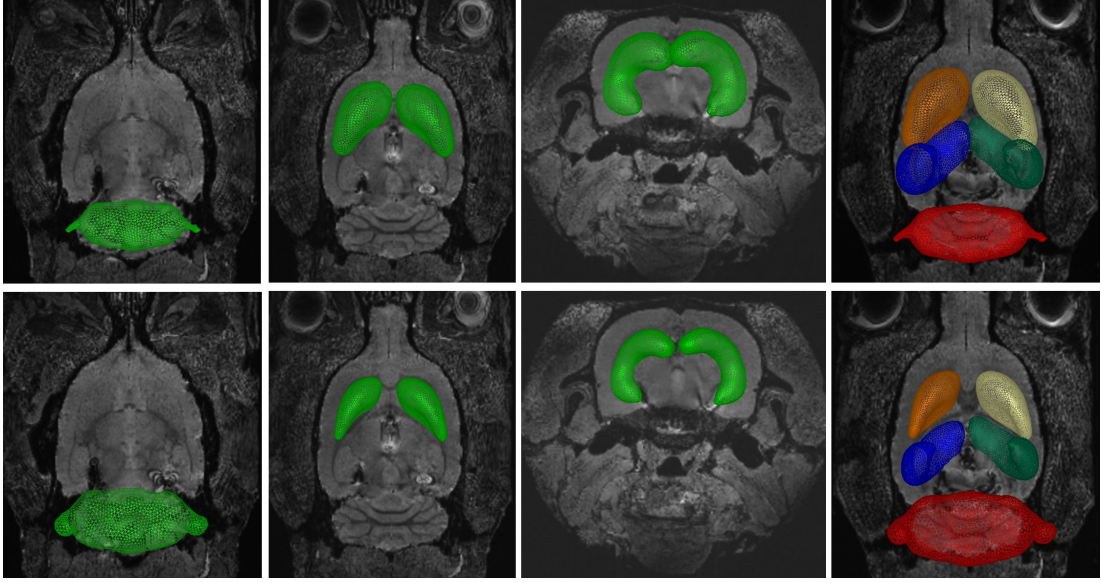
Figure 5.17: Comparison of segmentation results starting from the same initialization. The first row: traditional shape prior constraint (smoothness). The second row: using hierarchical shape prior. From left to right: the cerebellum, the left and right striatum, the left and right hippocampus, and all five structures.

5. Update the value of $k_g = 1 - e^{-|\nabla \Lambda|}$. Adjust the structure positions to prevent intersections using global prior.

6. Refine each intermediate result by local prior constraint (transform it to the mean shape, then update the pose and shape parameters consistent with the training data).

7. Repeat steps 3-6 until convergence.

Alternating the deformation module and the shape constraint module is more robust to noise and can handle more complex textures than purely using a deformation module. Due to the constraint from the shape prior, our model converges fast and robustly towards the true boundary given the mean shape of the multiple objects as initialization. One more benefit of our multiple structure model is that the size of the training data can be small, because the shape prior is built hierarchically. The prior for each structure is obtained individually, which can better discover the shape statistics than using one prior for all structures.

Structures	Prior types	Voxel distance	Volume error
	Smoothness	4.35 ± 2.17	0.22 ± 0.12
Cerebellum	Independent	1.74 ± 1.18	0.05 ± 0.02
	Hierarchical	$\mathbf{1.70 \pm 1.13}$	$\mathbf{0.04 \pm 0.02}$
	Smoothness	3.79 ± 2.05	0.51 ± 0.19
Striatum	Independent	2.93 ± 1.81	0.19 ± 0.06
	Hierarchical	$\mathbf{1.37 \pm 1.09}$	$\mathbf{0.07 \pm 0.03}$
	Smoothness	3.82 ± 2.14	0.53 ± 0.18
Hippocampus	Independent	2.69 ± 1.83	0.17 ± 0.05
	Hierarchical	$\mathbf{1.22 \pm 1.05}$	$\mathbf{0.06 \pm 0.02}$

Table 5.3: Quantitative comparisons of different prior schemes. We reported the mean and standard deviation of voxel distances between segmented surfaces and ground truth surfaces, and relative errors of volume magnitude in proportions.

5.3.3 Visual Comparisons

Fig. 5.17 shows the visual comparisons of segmentation results starting from the same initialization. In the first row, we use the robust deformable model [47] to segment each structure individually. The shape prior is the smoothness constraint. In the second row, the deformable model is combined with the hierarchical shape prior to segment all structures simultaneously. When segmenting the cerebellum, the shape prior helps to preserve the shape of paraflocculi and to avoid the local minima caused by complex textures inside. Without using shape prior, the model is attracted by large gradient values inside and the paraflocculi are smoothed out. In the striatum and hippocampus cases, there is no obvious boundary between these structures and the brain. Thus the model can easily over-segment the ROI. Using shape prior alleviates this problem. Another benefit of the hierarchical shape prior is that the spatial constraint is incorporated. The distance between two structures has to follow the location statistics. Thus structures cannot intersect with each other.

5.3.4 Quantitative Comparisons

Table 5.3 compares different priors. Note that all deformation modules are the same and based on both gradient and region information. We compared the methods of

combing the robust deformation model with a traditional smoothness shape prior [75], an independent shape prior [92] for each structure and hierarchical shape priors. The same parameters are used in all deformation modules. We reported the mean value and standard deviation of voxel distances between segmented surfaces and ground truth surfaces. Generally, the shape prior constraint improves the segmentation accuracy. In the cerebellum case, the spatial constraint in hierarchical priors only slightly improves the result, while it is much more important in the striatum and hippocampus cases. The reason is that the striatum and hippocampus are adjacent to each other. Spatial constraints can prevent the intersection of different structures, which implicitly alleviates the over-segmentation problem caused by the low contrast and ambiguous boundaries. The volume magnitude information is very important in rodent brain analysis. Thus we also reported the relative error of volume magnitude compared to the ground truth. Using the hierarchical shape prior achieves the best result, especially for the striatum and the hippocampus. This spatial constraint not only improves the segmentation accuracy for multiple structures, but also decreases the number of iterations and computation time to converge.

5.3.5   Summaries

In this section, we have applied our algorithm on rodent brain segmentation from MR microscope. Shape prior method is independent of image modalities. Thus we use similar setting as the liver segmentation task, except that hierarchical shape prior constraint is employed to ensure that brain structures cannot penetrate with each other. Such constraint is very effective when segmenting multiple objects simultaneously.

5.4   Cardiac Tracking from MRI

5.4.1   Clinical Background

Extraction of the boundary contour of a beating heart from cardiac MRI image sequences plays an important role in cardiac disease diagnosis and treatments. One of the applications is MRI-guided robotic interventions which is potentially important in

cardiac procedures such as aortic valve repair [83, 55, 62]. The major difficulty in such procedures is the path planning of the robotic needle, where it requires an accurate contour of the left myocardium on a beating heart. The algorithm should be robust, accurate and fast.

Therefore, we consider the problem of tracking the inner boundary contour of the left myocardium from cardiac MR images. If the motion of the region of interest is constrained (e.g. rigid or approximately rigid), the contour motion can be efficiently represented by a small number of parameters, e.g. the affine group. But if the "object" is arbitrarily deforming (like in our application), each contour point can move independently. Contour deformation then forms an infinite dimensional space. Direct application of motion tracking algorithms (e.g., Kalman filter, particle filters) for large dimensional problems is impractical, due to the reduction in effective particle size as dimension increases. The alternative way is to model the shape of the boundary to constraint the freedom on how the contour deforms.

Two major types of shape models are statistical shape models and deformable models. Statistical shape models are learned a priori from examples to capture variations in the shape and appearance of an object of interest in images [12, 11]. These models efficiently constraint the deformation to a specific manner thus better in image interpretation. However, it usually only takes into account local shape or appearance information, thus less robust confronting large appearance and motion changes of the beating heart. Deformable models are curves or surfaces that deform under the influence of internal smoothness and external image forces to delineate object boundary [50, 61]. They have the advantage of estimating a boundary or surface with smoothed curves but the computational cost can be expensive for temporal data. Some recent works effectively integrate both types [100] and achieve high contour segmentation accuracy, but those methods neglect the temporal connections of the image sequence, and usually require an additional learning algorithm to produce an initialized contour at each time step.

We propose a bottom-up approach to solve the contour tracking problem. On the bottom level, a collaborative trackers network provide a deformed mesh which is able

to predict an initialized contour at each time step [102]. Each tracker in the network is a particle filter tracker featuring the appearance of the tracking region as the template. The coordination of the trackers in the network is modeled by Bayesian network, which models spatio-temporal dependency relationships of multiple trackers. On the top level, a shape prior module is employed to refine the initialized contour provided by the collaborative trackers network, and eventually obtains a smoothed contour of the left endocardium over time. Our approach has the following advantages:

1. The tracking performance of the collaborative trackers network not only depends on the local appearance of individual trackers, but also depends on the performance of its neighboring trackers, namely its context. Once the shape prior is applied on the top level, the bottom-up approach integrates shape, appearance and context information in a unified framework, providing a robust solution to the contour tracking problem.

2. The collaborative trackers network featuring particle filter and Bayesian network mechanisms automatically enforce temporal consistency of the contours on subsequent images. It automatically generates an initialized contour input for the shape refinement at each time step.

3. The approach is computationally light weighted. It takes 0.17 second on average to process one image on a 2.4GHz desktop.

5.4.2 Collaborative Contour Tracking with Shape Prior

The contour of the left endocardium at time t is represented by a set of points $C_t = \{(x_1^t, y_1^t), ..., (x_N^t, y_N^t)\}$ on the 2D image plane. The goal of the approach is to accurately estimate the contour $\bar{C}_t$ at each time step. For this purpose, we first use a collaborative tracking network composing of multiple particle filter trackers to track the endocardial region (Fig. 5.18 left). Then an ASM shape prior model is applied to refine the tracking results (Fig. 5.18 right). The two parts work iteratively to achieve accurate contour tracking performance.
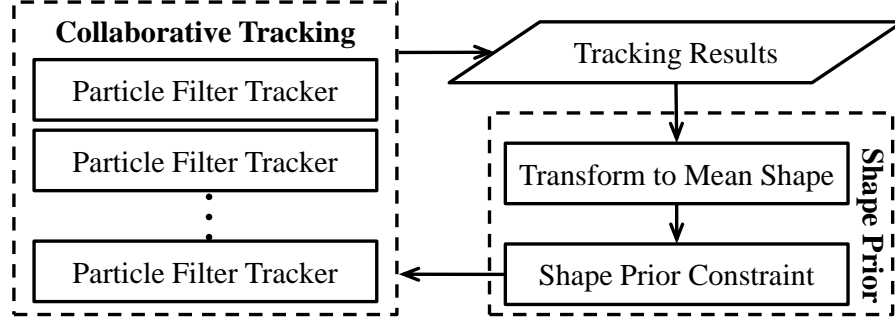
Figure 5.18: Framework of our algorithms, which includes the tracking network and shape prior module.

Collaborative Tracking: Given an initial contour region C_0, the collaborative tracking approach is able to automatically compute a coarse contour E_t at each time step, providing the base for the shape refinement algorithm. The initial contour C_0 is manually selected on the first frame. For the endocardial region, one could use a circle instead of the manual selection. We adopted Zhou et al.'s tracking method [102] as the tracking mechanism.

Tracking Network Computation: The collaborative trackers network is composed of multiple particle filter trackers $\Gamma_t = \{T_i^t, ..., T_m^t\}$ (Fig. 5.19 (a)). For each individual tracker T_i^t, we measure its tracking performance by computing its surviving probability. The adjacent trackers $\{T_j^t, ..., T_m^t\}$ provide evidence to compute its surviving probability via a Bayesian network:

$$p(W_i^t, \hat{\Theta}_i^t | G_i^{t-1}, Z_i^t) \propto p(G_i^{t-1}) \prod_k N(\hat{\theta}_i^t | \hat{\theta}_k^t, \sigma^2) \prod_k p(\hat{\theta}_k^t | z_k^t) \tag{5.5}$$

$\hat{\Theta}_i^t = \{\hat{\theta}_i^t, \hat{\theta}_j^t, ..., \hat{\theta}_m^t\}$ and $Z_i^t = \{z_i^t, z_j^t, ..., z_m^t\}$ are the estimated states and observations of T_i^t and its adjacent trackers. W_i^t represents the event that tracker T_i^t survives. G_i^{t-1} is the Bayesian network at time $t-1$, whose probability $p(G_i^{t-1})$ is known at time t. $N(\hat{\theta}_i^t | \hat{\theta}_k^t, \sigma^2)$ is the probability density of $\hat{\theta}_i^t$ on the Normal distribution centered at $\hat{\theta}_k^t$ with variance σ^2. Eq. 5.5 features three important terms: 1) $p(G_i^{t-1})$ measures the surviving probability from the previous time step. A tracker that performed well at the
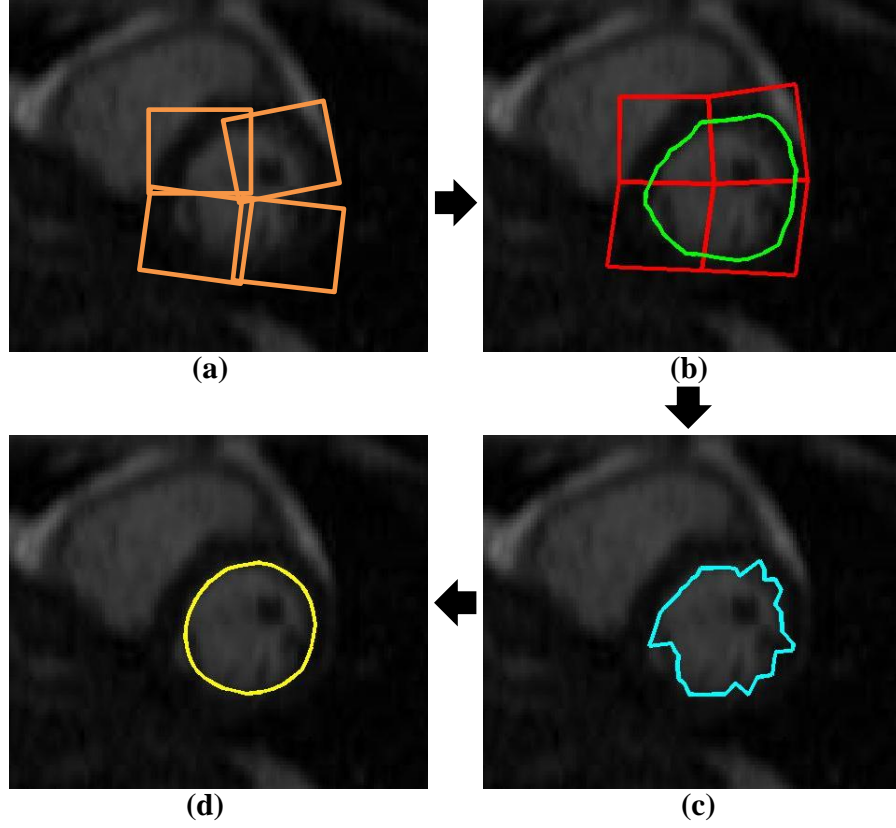
Figure 5.19: (a) Trackers' network that composes of 4 particle filter trackers; (b) Deformation mesh computed from the trackers' network (colored in red), and the deformed contour (colored in green); (c) Computed edge along the normal direction of the deformed contour; (d) Extracted shape after shape refinement.

previous time step is likely to survive in the current time step. 2) $\prod_k p(\hat{\theta}_k^t | z_k^t)$ measures the tracking performance of $T_{i,t}$ and its surrounding trackers. Individual trackers with high performance have high surviving probability. 3) $\prod_k N(\hat{\theta}_i^t | \hat{\theta}_k^t, \sigma^2)$ measures the motion pattern of the tracker network. Uniform motion patterns increase the chance of survival. Those terms enforce both spatial and temporal consistency of the trackers network, ensuring robust performance.

Deformation Mesh and Interpolated Contour (C_0 to D_t): The deformation mesh (Fig. 5.19 (b) in red) M_t is composed of a set of nodes $M_t = (c_1^t, c_2^t, ..., c_n^t)$, which are distributed over the targeted heart region. In the initialization step, the nodes coincide with the defining points of the individual trackers. Subsequently,each node's

location is decided by the linked trackers, as follows:

$$c_i^t = \frac{1}{\sum_{k=1}^{L_i^t} \beta_k^t} \sum_{k=1}^{L_i^t} s_k^t \beta_k^t, \tag{5.6}$$

where $\{s_k^t\}_{k=1}^{L_i^t}$ indicates all the corner coordinates of the linked trackers at a certain location and β_k^t is the impact factor of each linked tracker; we use the surviving probability computed in Eq. 5.5 as the impact factor here. L_i^t is the total number of linked trackers in locale i at time t. Given the contour of the previous time step D_{t-1}, we are able to linearly interpolate a deformed contour D_t according to the coordinates of this deformed mesh (Fig. 5.19 (b) in green).

Edge Detection (D_t to E_t): The contour D_t is computed by rigid linear interpolation, thus has the same shape as the initial contour C_0, and can't adapt to the edge of the endiocardial region very well. Therefore, we use an edge detection method to align D_t to the edge of the tracking region E_t. Specifically, E_t is computed by finding the largest first derivative along the normal direction of contour D_t (Fig. 5.19 (d)). However, the extracted edges always fall into local minimum. Therefore it requires a shape refinement procedure to smooth the edges and constrain the deformation. That's why we pass E_t as the input for the following shape priors model.

Refinement using Shape Priors: The above tracking method may not be able to accurately capture the boundary since it can stuck at the local minimum, especially when the texture of images is complex. Thus a shape refinement procedure is added in our method to constrain the deformation by following the patterns of a training database using proposed shape prior module. This step can prevent over and under-segmentation and provide a shape constraint. Alternatively employing the collaborative tracking and the shape prior based refinement is more robust to noises and can handle more complex textures than purely using the tracking algorithm. Our model converges fast and robust towards the boundary because of the benefit of the good-quality tracking result and the constraint of the shape prior. One more benefit of our model is that the size of the training data can be small, because the tracking algorithm already considers the temporal information and constraint, and is relatively robust to image noise in single

(a) **(b)** **(c)**
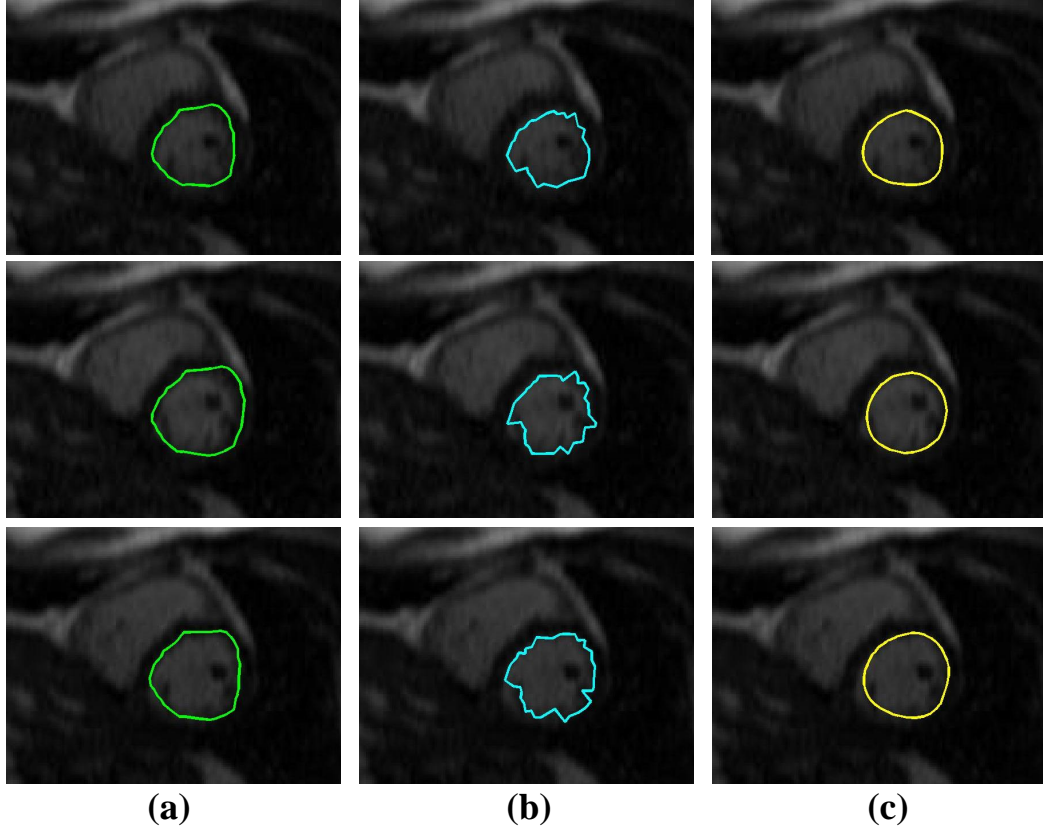
Figure 5.20: Comparative results on a CINE testing sequence with 15 frames. (a) Tracking results by the deformation mesh approach. (b) Tracking by edge detection. (c) Tracking results using the proposed method. The rows correspond to frame index 3, 9 and 15.

slice. Thus training information is only used for the shape refinement procedure, which is an auxiliary step in our model.

5.4.3   Visual and Quantitative Comparisons

The performance of the contour tracking approach is critical to MRI-guided robotic interventions. At each time step, the accurate contour of the left ventricle serves as the "boundary" to limit the freedom of the robotic needle so that it won't accidentally touch the cardiac wall. The approach should also be fast enough to minimize the computational cost during a real-time heart surgery. Therefore, we compare the proposed collaborative contour tracking approach with the state-of-art methods in terms of accuracy and speed.
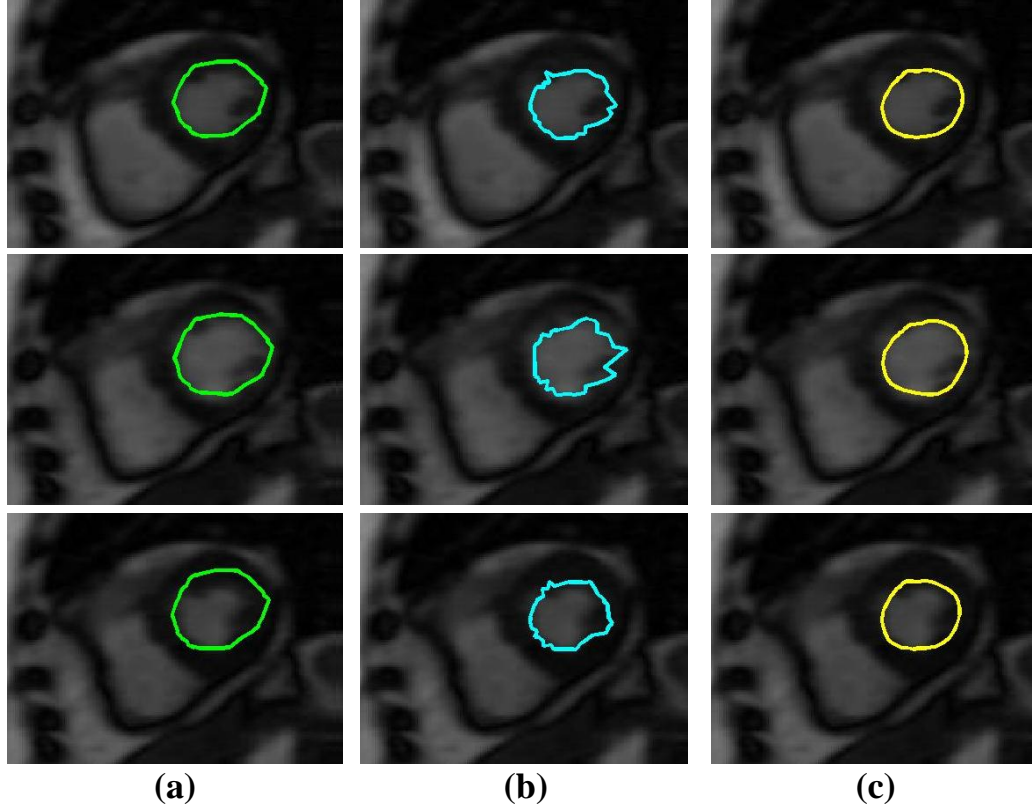
Figure 5.21: Comparative results on a real-time testing sequence with 750 frames. (a) Tracking results by the deformation mesh approach. (b) Tracking by edge detection. (c) Tracking results using the proposed method. The rows corresponds to frame index 239, 251, 409.

We performed experiments on 14 MRI sequences with 12-750 slices (frames) for each sequence and 1,728 slices (frames) in total. Data was acquired from 7 patients and by two types of MRI acquisition protocols: (a) CINE MRI sequences where patients hold the breath during MRI scanning; This type of MRI sequences contains relatively fewer slices (12-25 frames) since the patient is not able to hold the breath for long time. (b) Real-time MRI sequences where patients have the natural breathing during MRI scanning. This type of MRI sequences contains more slices (50-750 frames) and the cardiac motion pattern is more complicated because of the additional respiratory motion. From all the images, we randomly selected 150 images for training, and all the images for testing. We compare the proposed collaborative contour tracking approach with the following approaches:

- Deformation mesh approach: The collaborative tracking network [101] is able to
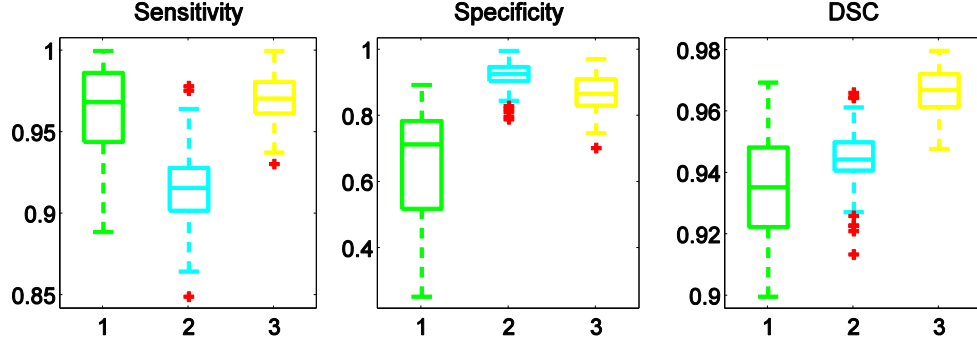
Figure 5.22: Boxplots of sensitivity, specificity and DSC of all testing data. In each subplot, y-axis is the performance of sensitivity, specificity or DSC. x-axis aligns 3 methods from left to right in the same order as in Fig. 5.20 and Fig. 5.21.

provide a deformation mesh based on individual trackers' states. At each time step, it is able to produce the current contour by interpolating the initialized contour according to the current deformation mesh.

- Edge detection approach: This approach simply computes the edge along the normal direction of the interpolated contour from the deformation mesh.

Some representative and challenging cases are shown in Fig. 5.20 and Fig. 5.21. Fig. 5.20 shows the results from a short CINE sequence. The deformable mesh approach performs well visually because the sequence includes just 15 frames with little shape variation over time. The only problem is that the shape of the contour is static no matter how the left ventricle deforms due to the linear interpolation. The edge detection approach performs worse visually because the edge detection method is easy to drop into local minima. Our proposed approach provides a smoothed shape that matches the deformed left ventricle. Fig. 5.21 shows the results from a long real-time sequence. The deformable mesh approach performs worse than in the CINE sequence because of the large cardiac motion. The deformation mesh is only able to compensate the global motion of each tracker, but not the accurate local motion of the shape. Therefore, the interpolated shape is not accurate. The edge detection approach still produces fuzzy results visually. Our proposed approach not only captures the motion variations over time, but also provides a smoothed accurate contour.

To quantitatively compare different methods, we report the boxplots of sensitivity

(P), specificity (Q) and Dice Similarity Coefficient (DSC) in Fig. 5.22. The computation was performed on all the testing data, and the ground-truth contours were annotated manually by an expert. We prefer the approach with large mean value and small variance since it shows accuracy and stableness.

In terms of sensitivity, our method is much better than the edge method and slightly better than the deformation mesh method. For the specificity, the result from edge method is also good. The reason is that this method always under-segment the ROI. Thus most of the background is correctly predicted. However, as a trade-off, its sensitivity is not as good as the others. Considering both P and Q, our method outperforms deformable mesh method and edge detection method. DSC combines both sensitivity and specificity information. Thus it is a reasonable measurement to compare accuracy. The results of DSC further validate our proposed method.

5.4.4 Summaries

In this experiment, we have improved a tracking algorithm by incorporating shape prior constraint. Efficiency and stableness is very important for tracking methods. Adding this prior module does not significantly increases the computational complexity due to the fast shape update. Furthermore, it also improves the accuracy because the shape refinement step helps the model to avoid local minima problems.

5.5 Cardiac Shape Atlas from High Resolution CT

5.5.1 Clinical Background

In this section, we evaluate the geometry processing and shape registration methods in the framework proposed in Chapter 4 by generating cardiac shape atlas from high resolution CT. As we mentioned, geometry processing and shape registration algorithms are employed to generate 3D training data for the aforementioned applications (i.e., liver segmentation and rodent brain segmentation), since mesh quality and one-to-one correspondence are crucial for the performance of shape prior modeling and segmentation. However, it is important to show that these geometry processing and shape registration

methods can effectively produce such high quality meshes without sacrificing detail information. We choose this high resolution CT reconstruction task to validate our shape atlas framework since such data contains many detail information.

Recent developments on the 320 multi-detector CT technologies have made the volumetric acquisition of 4D high resolution cardiac images in a single heart beat possible. We applied our framework to 10 cardiac CT volumes, which captures a whole cycle of cardiac contraction. The CT data were acquired on a 320-MDCT scanner using a conventional ECG-gated contrast-enhanced CT angiography protocol. The imaging protocol parameters include: prospectively triggered, single-beat, volumetric acquisition; detector width 0.5 mm, voltage 120 KV, current $200 - 550$ mA. Reconstructions were done at 10 equally distributed time frames in a cardiac cycle. The resolution of each time frame is $512 \times 512 \times 320$. Our reconstruction method successfully captured the papillary muscles and the trabeculae of the left ventricle. The challenge of this task is to handle these complex shape details. Our framework is able to provide high quality meshes without removing these details. Furthermore, it can robustly discover the one-to-one correspondence among different time frames.

5.5.2 Visual and Quantitative Comparisons
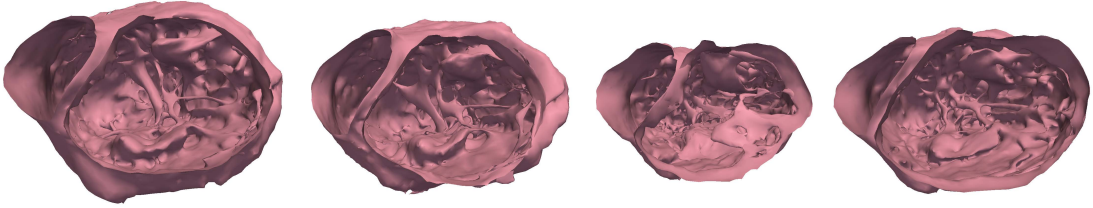
Figure 5.23: The cardiac shapes extracted from high resolution CT images. The complex shape details are captured, such as the papillary muscles and the trabeculae.

Fig. 5.23 shows the shape atlas reconstruction results from high resolution cardiac CT images, using geometry processing and shape registration methods. The three-dimensional structures, their relationship and their movement during the cardiac cycle are much more readily appreciated from the shape model than from the original volumetric image data. Fig. 5.24 visualizes the errors between our processed atlas and the
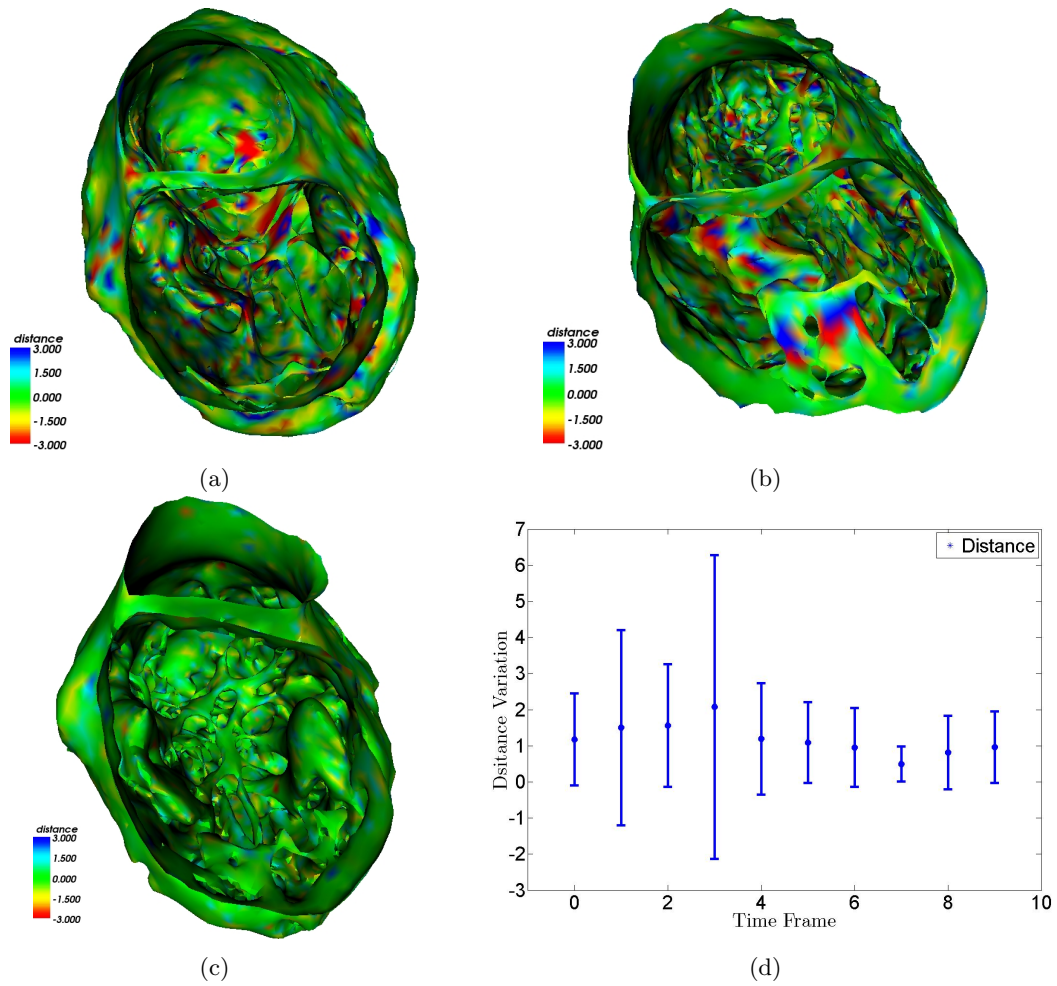
Figure 5.24: Differences of the cardiac shape to the ground truth. Green colors mean that the distances are within one voxel. Red colors mean under-segmentation while blue colors mean over-segmentation. (a)(b)(c) Different frames of the differences of the left ventricle to the ground truth. (d) Distances distribution of every frame.
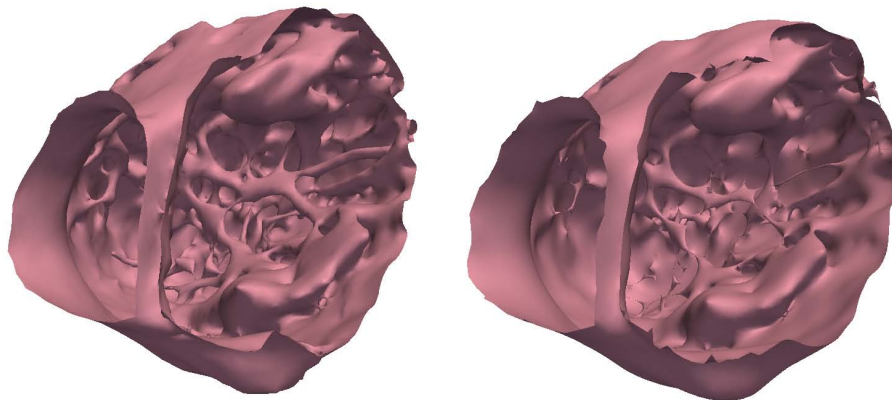
Figure 5.25: Comparison of smoothing methods. The left result is from the cotangent weights, and the right one is from the uniform weights. Cotangent weights well preserve shape details, while the uniform weights smooth out some local features.
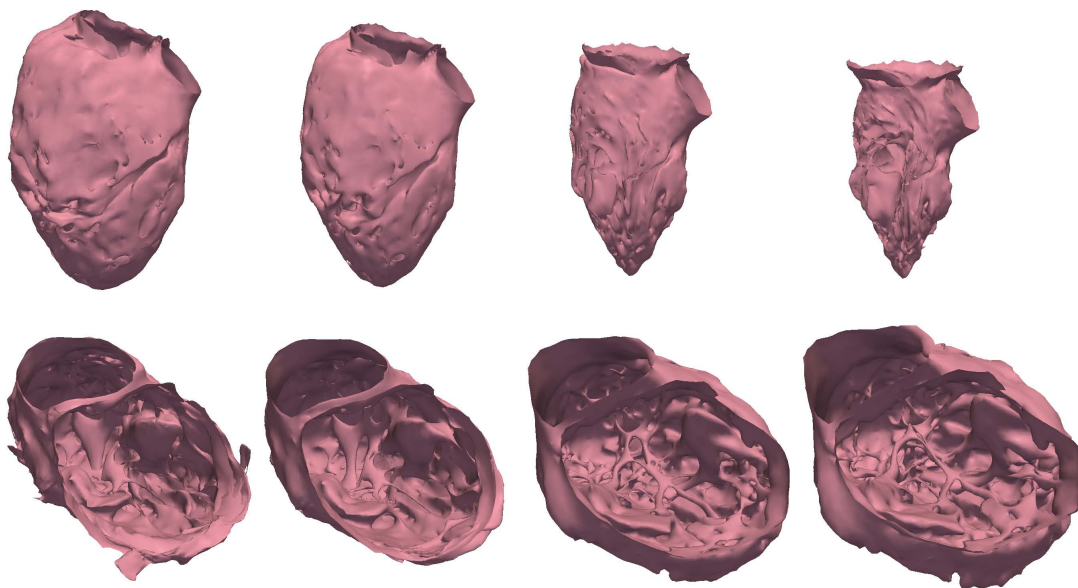
Figure 5.26: Modes with largest variances, from -3σ to 3σ. The first mode represents the changing of the volume size. The second mode is the changing of shape details such as papillary muscles.

original mesh annotated by clinical experts. The color indicates the distance from each vertex of our results to the manual segmentation. Fig. 5.24(d) shows the quantitative evaluation of the results. The mean distance from every vertex to the ground truth is about one voxel. During the diastolic frames, which are frames 4 to frame 9, there are smaller distances and less distance variations. On the other hand, during cardiac systole, because of large deformations between neighboring frames, the results have larger errors. Fig. 5.25 compares the results from different smoothing methods, i.e., using cotangent weights or uniform weights. Using cotangent weights can preserve more details than the uniform weights since the vertices are moved toward the tangential direction.

Fig. 5.26 visualizes the shape variation along the first and second principal directions. Due to the benefits of the detail preserved smoothing and one-to-one correspondence shape registration, our framework is able to robustly handle such complex cases. The first mode represents the changing of the volume magnitude, and the second mode captures the changing of shape details such as the papillary muscles and the trabeculae. Such information can be used in clinical applications to categorize the cardiac properties.

5.5.3   Summaries

To summarize, our shape atlas framework is employed as a preprocess step to generate high quality meshes and obtain one-to-one correspondence for all 3D applications. In this section we have validated the performance of this framework by applying it on a very challenge application, cardiac reconstruction from high resolution CT. The experiment results show that mesh quality can be highly improved while the shape details are also properly preserved. Thus, this framework is a reliable step for the preprocess of these shape prior modeling methods such as ASM and SSC.

5.6   Discussion

The experimental results show the following facts.

1. This implicitly incorporated shape constraint benefits the interpretation of images. Such shape-based information improves the robustness and accuracy of low-level algorithms.

2. The sparse linear combination of shape repository is able to well approximate an input shape if the input does not have gross errors. The $L1$ norm constraint of e handles the non-Gaussian residuals caused by occlusion, mis-detection or point missing.

3. Our method is able to use a subset of vertices on the contour and surface as the input. Sparse learning is applied on this subset and computes a group of coefficients. Such coefficients are then used to combine the whole contour or surfaces, which generate reasonable results. This property can seamlessly incorporate with multi-resolution scheme. Our method can naturally extend to multi-resolution without significant overhead of implementation or computational complexity.

4. When the number of training data are huge (e.g., thousands or millions), it is infeasible to simply stack all shapes into the data matrix. In this case, dictionary learning technique is employed to learn a compact dictionary, whose size is much smaller than the whole dataset. Such dictionary highly improves the computational efficiency without significantly sacrificing accuracy.

5. Mesh quality and one-to-one correspondence are two important requirements for many shape prior modeling methods such as ASM and SSC. Our shape atlas reconstruction framework generates such training meshes effectively. The mesh quality is highly improved, while the detail information is still well preserved.

6. We use it to solve several diverse medical image applications with different modalities and different types of shapes (i.e., 2D contour and 3D surface mesh). Thus SSC is independent of the application or data dimension, and does not substantially increase the computational complexity. Existing shape based segmentation systems can be easily extended by adding this proposed method as a shape refinement module.

Chapter 6

Conclusions

This thesis proposed a sparse representation based method to implicitly model shape priors. It alleviates three challenging problems in a unified framework, i.e., modeling complex shape variations, handling non-Gaussian errors and preserve local detail information of the input shape. Some extensions have been proposed to improve its performance and computational efficiency. They include multi-resolution support, modeling of local shape priors using mesh partitioning, and dictionary learning. A segmentation framework is also proposed by using this method as a shape inference module and shape refinement module. Furthermore, we have discussed an effective framework to generate high quality training meshes as the input for these shape prior methods. This framework is extensively validated on five very diverse and challenging medical applications: 2D lung localization in X-ray images, 3D liver segmentation in low-dose CT scans, 3D segmentation of multiple rodent brain structures in MR microscope, left ventricle tracking in cardiac MRI, and reconstruction of cardiac shape atlas in high resolution CT. Compared to state-of-the-art methods, our shape prior model exhibits better performance in these studies.

There are some future directions to pursue toward the goal of robust shape prior modeling. First, the spatial information is not exploited yet. Only positions of vertices are considered in this proposed model. It is also desirable to further employ the connectivity among vertices. For example, when one vertex is recognized as outliers, its neighbors may also have a high probability to be outliers. In this case, traditional sparse representation cannot properly employ such information. Structured sparsity or group sparsity [46, 41] is able to incorporate such prior constraint. Spatial information can be encoded as groups and modeled as a regularization term. Thus all vertices inside

the same group have similar probability to be outliers or not. The potential challenge is how to determine group size. Large group size can improve the robustness but also increase computational overhead. Furthermore, grouping in 3D is also time and storage consuming. Thus it is necessary to effectively model this information.

Second, the optimization framework can also be improved. Currently an EM type of method is employed to find the sparse linear combination. Theocratically there is no guarantee that such method can always find global minimum. Although our algorithm works well in many real world applications, it is still desirable to develop an optimization framework which simultaneously solves all unknown variables. A potential solution is to use iterative linearization as the one used in robust alignment for linearly correlated Images [64]. However, input for our system can have large misalignment. This may cause problems for iterative linearization algorithms.

Third, currently linear combination is used to model shape priors. It is possible that some shapes cannot be represented by this linear combination. Non-linear relations can be investigated and properly employed to flexibly model shapes.

Last but not the least, we have applied this shape prior method on many medical applications, including 2D/3D segmentation, tracking, etc. We also plan to apply it on more medical applications, and also computer vision tasks, such as, but not limited to, registration and face alignment. Particularly, we plan to work on shape analysis using statistical information [34, 35]. Such analysis can be widely used in diverse applications, such as understanding developmental and anatomical aspects of disorders when comparing patients versus normal controls and studying morphological changes caused by aging. We believe that SSC model can be potentially used in these algorithms since shape prior can improve the performance and robustness.

References

[1] M. Aharon, M. Elad, , and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. In *IEEE Transaction Signal Processing*, pages 4311–4322, 2006.

[2] A. Beck and M. Teboulle. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Transaction on Image Processing*, 18(113):2419–2434, 2009.

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[4] F. Bookstein. Principal warps: thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6), 1989.

[5] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[6] E. Candes and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Transaction on Information Theory*, 52(12):5406 –5425, 2006.

[7] S. Chen, S. A. Billings, and W. Luo. Orthogonal least squares methods and their application to non-linear system identification. In *International Journal of Control*, pages 1873–1896, 1989.

[8] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 43(1):129–159, 2001.

[9] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *Computer Vision and Image Understanding*, 89(2-3):114–141, 2003.

[10] Y. Cong, J. Yuan, and J. Liu. Sparse reconstruction cost for abnormal event detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.

[11] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In *European Conference on Computer Vision*, pages 484–498, 1998.

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

[13] T. Cootes and C. J. Taylor. A mixture model for representing shape variation. In *Image and Vision Computing*, pages 110–119, 1997.

[14] T. Cootes, C. J. Twining, K. O. Babalola, and C. J. Taylor. Diffeomorphic statistical shape models. *Image and Vision Computing*, 26:326–332, 2008.

[15] C. Davatzikos, X. Tao, and D. Shen. Hierarchical active shape models, using the wavelet transform. *IEEE Transactions on Medical Imaging*, 22(3):414–423, 2003.

[16] P. Dellaportas, J. Forster, and I. Ntzoufras. On bayesian model and variable selection using mcmc. *Statistics and Computing*, 12(1):27–36, 2002.

[17] D. L. Donoho. For most large underdetermined systems of equations, the minimal l1-norm near-solution approximates the sparsest near-solution. Technical report, Communications on Pure and Applied Mathematics, 2004.

[18] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[19] D. L. Donoho, Y. Tsaig, I. Drori, and J. luc Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. Technical report, 2006.

[20] Q. Duan, J. Cai, and J. Zheng. Fast environment matting extraction using compressive sensing.

[21] N. Duta and M. Sonka. Segmentation and interpretation of mr brain images. an improved active shape model. *IEEE Transactions on Medical Imaging*, 17(6):1049 –1062, 1998.

[22] K. Engan, S. Aase, and J. Hakon Husoy. Method of optimal directions for frame design. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 2443–2446. IEEE, 1999.

[23] P. Etyngier, F. Segonne, and R. Keriven. Shape priors using manifold learning techniques. In *International Conference on Computer Vision*, pages 1 –8, 2007.

[24] M. Figueiredo, R. Nowak, and S. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):586–597, 2007.

[25] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981.

[26] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Transactions on Medical Imaging*, 23:995–1005, 2004.

[27] A. Frangi, D. Rueckert, J. Schnabel, and W. Niessen. Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling. *IEEE Transactions on Medical Imaging*, 21(9):1151–1166, 2002.

[28] B. Frey and D. Dueck. Clustering by passing messages between data points. *Science*, 315(5814):972, 2007.

[29] M. Gao, J. Huang, S. Zhang, Z. Qian, S. Voros, D. Metaxas, and L. Axel. 4d cardiac reconstruction using high resolution ct images. *Functional Imaging and Modeling of the Heart*, 6666:153–160, 2011.

[30] S. Gao, L. Chia, and I. Tsang. Multi-layer group sparse coding–for concurrent image classification and annotation. In *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2011.

[31] E. George and R. McCulloch. Variable selection via gibbs sampling. *Journal of the American Statistical Association*, pages 881–889, 1993.

[32] B. Georgescu, X. Zhou, D. Comaniciu, and A. Gupta. Database-guided segmentation of anatomical structures with complex appearance. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 429–436, 2005.

[33] S. Gerber, T. Tasdizen, S. Joshi, and R. Whitaker. On the manifold structure of the space of brain images. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, volume 5761 of *LNCS*, pages 305–312. 2009.

[34] P. Golland, W. Grimson, M. Shenton, and R. Kikinis. Deformation analysis for shape based classification. In *Information Processing in Medical Imaging*, pages 517–530. Springer, 2001.

[35] P. Golland, W. Grimson, M. Shenton, and R. Kikinis. Detection and analysis of statistical differences in anatomical shape. *Medical Image Analysis*, 9(1):69–86, 2005.

[36] C. Goodall. Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society*, 53:285–339, 1991.

[37] L. Gu and T. Kanade. A generative shape regularization model for robust face alignment. In *European Conference on Computer Vision*, pages 413–426, 2008.

[38] T. Heimann and H.-P. Meinzer. Statistical shape models for 3D medical image segmentation: A review. *Medical Image Analysis*, 13(4):543 – 563, 2009.

[39] J. Huang, X. Huang, and D. Metaxas. Simultaneous image transformation and sparse representation recovery. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[40] J. Huang, X. Huang, and D. Metaxas. Learning with dynamic group sparsity. In *IEEE International Conference on Computer Vision*, pages 64–71. IEEE, 2009.

[41] J. Huang, X. Huang, and D. Metaxas. Learning with dynamic group sparsity. In *International Conference on Computer Vision*, pages 64 –71, 2009.

[42] J. Huang, S. Zhang, and D. Metaxas. Fast optimization for mixture prior models. In *European Conference on Computer Vision*, pages 607–620. Springer, 2010.

[43] J. Huang, S. Zhang, and D. Metaxas. Efficient mr image reconstruction for compressed mr imaging. *Medical Image Analysis*, 15(5):670 – 679, 2011.

[44] J. Huang, S. Zhang, and D. N. Metaxas. Efficient MR image reconstruction for compressed MR imaging. In *Medical Image Computing and Computer-Assisted Intervention*, pages 135–142. Springer, 2010.

[45] J. Huang and T. Zhang. The benefit of group sparsity. *The Annals of Statistics*, 38(4):1978–2004, 2010.

[46] J. Huang, T. Zhang, and D. Metaxas. Learning with structured sparsity. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 417–424. ACM, 2009.

[47] X. Huang and D. Metaxas. Metamorphs: Deformable shape and appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(8):1444 –1459, 2008.

[48] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1:321–331, 1987.

[49] S. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky. An interior-point method for large-scale l1-regularized least squares. *IEEE Journal of Selected Topics in Signal Processing*, 1(4):606–617, 2007.

[50] T. Kohlberger, D. Cremers, M. Rousson, R. Ramaraj, and G. Funka-Lea. 4D shape priors for a level set segmentation of the left myocardium in SPECT sequences. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 92–100, 2006.

[51] L. Kuo and B. Mallick. Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 65–81, 1998.

[52] G. Langs, N. Paragios, and S. Essafi. Hierarchical 3D diffusion wavelet shape priors. In *IEEE International Conference on Computer Vision*, pages 1717–1724, 2010.

[53] K. Lekadir, R. Merrifield, and G. zhong Yang. Outlier detection and handling for robust 3D active shape models search. *IEEE Transactions on Medical Imaging*, 26:212–222, 2007.

[54] C. Li, C. Xu, C. Gui, and M. D. Fox. Level set evolution without re-initialization: A new variational formulation. *IEEE Conference on Computer Vision and Pattern Recognition*, 1:430–436, 2005.

[55] M. Li, D. Mazilu, and K. Horvath. Robotic system for transapical aortic valve replacement with mri guidance. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2008*, pages 476–484, 2008.

[56] H. Ling, S. Zhou, Y. Zheng, B. Georgescu, M. Suehling, and D. Comaniciu. Hierarchical, learning-based automatic liver segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[57] M. Lorenzo-Valdes, G. I. Sanchez-Ortiz, A. G. Elkington, R. H. Mohiaddin, and D. Rueckert. Segmentation of 4D cardiac MR images using a probabilistic atlas and the em algorithm. *Medical Image Analysis*, 8(3):255 – 265, 2004.

[58] M. Lustig, D. Donoho, and J. Pauly. Sparse mri: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[59] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Non-local sparse models for image restoration. In *International Conference on Computer Vision*, pages 2272 –2279, 2009.

[60] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. In *IEEE Transaction Signal Processing*, pages 3397–3415, 1993.

[61] J. McEachen and J. Duncan. Shape-based tracking of left ventricular wall motion. *IEEE Transactions on Medical Imaging*, 16(3):270–283, 1997.

[62] E. McVeigh, M. Guttman, R. Lederman, M. Li, O. Kocaturk, T. Hunt, S. Kozlov, and K. Horvath. Real-time interactive mri-guided cardiac surgery: Aortic valve replacement using a direct apical approach. *Magnetic resonance in medicine*, 56(5):958–964, 2006.

[63] J. A. Nahed, M. pierre Jolly, and G. zhong Yang. Robust active shape models: A robust, generic and simple automatic segmentation tool. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, 2006.

[64] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma. Rasl: Robust alignment via sparse and low-rank decomposition for linearly correlated images. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2010.

[65] U. Pinkall and K. Polthier. *Computing discrete minimal surfaces and their conjugates*. Citeseer, 1993.

[66] K. Pohl, S. Warfield, R. Kikinis, W. Grimson, and W. Wells. Coupling statistical segmentation and pca shape modeling. *Medical Image Computing and Computer-Assisted Intervention*, pages 151–159, 2004.

[67] A. Popovic, M. de la Fuente, M. Engelhardt, and K. Radermacher. Statistical validation metric for accuracy assessment in medical image segmentation. *Int J Comput Assist Radiol Surg*, 2:169–181, 2007.

[68] M. Rogers and J. Graham. Robust active shape model search. In *European Conference on Computer Vision*, pages 517–530, 2002.

[69] M. Rousson and N. Paragios. Shape priors for level set representations. In *European Conference on Computer Vision*, volume 2351, pages 416–418. 2002.

[70] J. Saragih, S. Lucey, and J. Cohn. Face alignment through subspace constrained mean-shifts. In *International Conference on Computer Vision*, pages 1034 –1041, 2009.

[71] D. Shen and C. Davatzikos. An adaptive-focus deformable model using statistical and geometric information. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):906 –913, 2000.

[72] D. Shen, E. Herskovits, and C. Davatzikos. An adaptive-focus statistical shape model for segmentation and shape modeling of 3-d brain structures. *IEEE Transactions on Medical Imaging*, 20(4):257–270, 2001.

[73] T. Shen, X. Huang, H. Li, S. Zhang, and J. Huang. A 3d laplacian-driven parametric deformable model. In *International Conference on Computer Vision*, 2011.

[74] T. Shen, H. Li, and X. Huang. Active volume medical models for medical image segmentation. *IEEE Transactions on Medical Imaging*, 2011.

[75] T. Shen, H. Li, Z. Qian, and X. Huang. Active volume models for 3D medical image segmentation. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 707 –714, 2009.

[76] T. Shen, S. Zhang, J. Huang, X. Huang, and D. Metaxas. Integrating shape and texture in 3d deformable models: From metamorphs to active volume models. *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*, pages 1–31, 2011.

[77] Y. Shi, F. Qi, Z. Xue, L. Chen, K. Ito, H. Matsuo, and D. Shen. Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics. *IEEE Transactions on Medical Imaging*, 27(4):481 –494, 2008.

[78] K. Sjostrand and et.al. Sparse decomposition and modeling of anatomical shape variation. *IEEE Transactions on Medical Imaging*, 26(12):1625 –1635, 2007.

[79] J. Starck, M. Elad, and D. Donoho. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Transactions on Image Processing*, 14:1570–1582, 2004.

[80] M. Styner, G. Gerig, J. Lieberman, D. Jones, and D. Weinberger. Statistical shape analysis of neuroanatomical structures based on medial models. *Medical Image Analysis*, 7:207–220, 2003.

[81] P. M. Thompson and A. W. Toga. Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations. *Medical Image Analysis*, 1(4):271–294, September 1997.

[82] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. In *IEEE Transaction Information Theory*, pages 2231–2242, 2004.

[83] N. Tsekos, A. Khanicheh, E. Christoforou, and C. Mavroidis. Magnetic resonance-compatible robotic and mechatronics systems for image-guided interventions and rehabilitation: a review study. *Annual Review Biomedical Engineering*, 9:351–387, 2007.

[84] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(2):210 –227, 2009.

[85] C. Xu and J. Prince. Snakes, shapes and gradient vector flow. *IEEE Transactions on Image Processing*, 7:359–369, 1998.

[86] P. Yan and J. Kruecker. Incremental shape statistics learning for prostate tracking in trus. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 42–49, 2010.

[87] P. Yan, S. Xu, B. Turkbey, and J. Kruecker. Discrete deformable model guided by partial active shape model for trus image segmentation. *IEEE Transactions on Biomedical Engineering*, 57(5):1158–1166, 2010.

[88] Y. Zhan, M. Dewan, and X. S. Zhou. Cross modality deformable segmentation using hierarchical clustering and learning. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 1033–1041, 2009.

[89] Y. Zhan and D. Shen. Deformable segmentation of 3D ultrasound prostate images using statistical texture matching method. *IEEE Transactions on Medical Imaging*, 25(3):256 –272, 2006.

[90] Y. Zhan, X. Zhou, Z. Peng, and A. Krishnan. Active scheduling of organ detection and segmentation in whole-body medical images. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 313–321, 2008.

[91] S. Zhang, J. Huang, Y. Huang, Y. Yu, H. Li, and D. Metaxas. Automatic image annotation using group sparsity. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3312–3319. IEEE, 2010.

[92] S. Zhang, J. Huang, M. Uzunbas, T. Shen, F. Delis, X. Huang, N. Volkow, P. Thanos, and D. Metaxas. 3D segmentation of rodent brain structures using active volume model with shape priors. *IEEE International Symposium on Biomedical Imaging*, pages 433–436, 2011.

[93] S. Zhang, J. Huang, M. Uzunbas, T. Shen, F. Delis, X. Huang, N. Volkow, P. Thanos, and D. N. Metaxas. 3d segmentation of rodent brain structures using hierarchical shape priors and deformable models. In *International Conference on*

Medical Image Computing and Computer Assisted Intervention, volume 6893 of *LNCS*, pages 611–618, 2011.

[94] S. Zhang, M. Uzunbas, Z. Yan, M. Gao, J. Huang, D. Metaxas, and L. Axel. Construction of left ventricle 3D shape atlas from cardiac mri. *Functional Imaging and Modeling of the Heart*, 6666:88–94, 2011.

[95] S. Zhang, X. Wang, D. Metaxas, T. Chen, and L. Axel. LV surface reconstruction from sparse tMRI using laplacian surface deformation and optimization. In *IEEE International Symposium on Biomedical Imaging*, pages 698–701, 2009.

[96] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. Metaxas, and X. Zhou. Deformable segmentation via sparse shape representation. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, volume 6892 of *LNCS*, pages 451–458, 2011.

[97] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. Metaxas, and X. Zhou. Sparse shape composition: A new framework for shape prior modeling. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1025–1032, 2011.

[98] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. N. Metaxas, and X. S. Zhou. Towards robust and effective shape modeling: Sparse shape composition. *Medical Image Analysis*, 16(1):265 – 277, 2012.

[99] W. Zhang, P. Yan, and X. Li. Estimating patient-specific shape prior for medical image segmentation. In *International Symposium on Biomedical Imaging*, pages 1451–1454. IEEE, 2011.

[100] Y. Zheng, A. Barbu, B. Georgescu, M. Scheuering, and D. Comaniciu. Four-chamber heart modeling and automatic segmentation for 3D cardiac CT volumes using marginal space learning and steerable features. *IEEE Transactions on Medical Imaging*, 27:1668–1681, 2008.

[101] Y. Zhou, N. Tsekos, and I. Pavlidis. Cardiac mri intervention and diagnosis via deformable collaborative tracking. In *International Conference on Functional Imaging and Modeling of the Heart*, volume 6666 of *LNCS*, pages 188–194. 2011.

[102] Y. Zhou, E. Yeniaras, P. Tsiamyrtzis, N. Tsekos, and I. Pavlidis. Collaborative tracking for MRI-guided robotic intervention on the beating heart. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010*, pages 351–358, 2010.

[103] S. Zhu and A. Yuille. Region Competition: Unifying snakes, region growing, and Bayes/MDL for multi-band image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(9):884–900, 1996.

[104] Y. Zhu, X. Papademetris, A. Sinusas, and J. Duncan. Bidirectional segmentation of three-dimensional cardiac MR images using a subject-specific dynamical model. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 450–457, 2008.

[105] Y. Zhu, X. Papademetris, A. Sinusas, and J. Duncan. Segmentation of left ventricle from 3D cardiac MR image sequences using a subject-specific dynamical model. *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[106] Y. Zhu, X. Papademetris, A. Sinusas, and J. Duncan. A dynamical shape prior for LV segmentation from RT3D echocardiography. In *International Conference on*

Medical Image Computing and Computer Assisted Intervention, pages 206–213, 2009.

[107] Y. Zhu, X. Papademetris, A. Sinusas, and J. Duncan. Segmentation of the left ventricle from cardiac MR images using a subject-specific dynamical model. *IEEE Transactions on Medical Imaging*, 29(3):669–687, 2010.

[108] X. Zhuang, K. Rhode, R. Razavi, D. Hawkes, and S. Ourselin. A registration-based propagation framework for automatic whole heart segmentation of cardiac MRI. *IEEE Transactions on Medical Imaging*, 29(9):1612 –1625, 2010.

Chapter 7
Curriculum Vitae

Education

- **Rutgers University**, New Brunswick, NJ   2007 - 2012
  Ph.D., Computer Science (GPA 4.00/4.00)

- **Shanghai Jiao Tong University**, Shanghai, China   2005 - 2007
  M.S., Computer Software and Theory

- **Zhejiang University**, Hangzhou, China   2001 - 2005
  B.E., Software Engineering

Experience

- **Research Assistant**, Rutgers University,   2009 - 2011
- **Research Intern**, NEC Laboratories America, Inc.,   summer 2011
- **Research Intern**, Siemens Healthcare USA, Inc.,   summer 2010
- **Teaching Assistant**, Rutgers University,   2007 - 2009
- **Research Assistant**, Shanghai Jiao Tong University,   2005 - 2007

Publications

- **Journal Papers and Book Chapters**

  1. [**MedIA'12**] Shaoting Zhang, Yiqiang Zhan, Maneesh Dewan, Junzhou Huang, Dimitris Metaxas and Xiang Zhou: Towards Robust and Effective Shape Modeling: Sparse Shape Composition, *Medical Image Analysis*, Volume 16, Issue 1, Pages 265-277, January 2012.

  2. [**MedIA'11**] Junzhou Huang, Shaoting Zhang and Dimitris Metaxas: Efficient MR Image Reconstruction for Compressed MR Imaging, *Medical Image Analysis*, Volume 15, Issue 5, pp. 670-679, October 2011.

  3. [**CVIU'11**] Junzhou Huang, Shaoting Zhang and Dimitris Metaxas: Composite Splitting Algorithms for Convex Optimization, accepted by *Computer Vision and Image Understanding*.

  4. [**BookChapter'11**] Tian Shen, Shaoting Zhang, Junzhou Huang, Xiaolei Huang and Dimitris Metaxas: Integrating Shape and Texture in 3D Deformable Models: From Metamorphs to Active Volume Models, *Multi Modality State-of-the-Art Medical Image Segmentation and Registration Methodologies*, (corresponding author).

  5. [**GMOD'11**] Shaoting Zhang, Junzhou Huang and Dimitris Metaxas: Robust Mesh Editing Using Laplacian Coordinates, *Graphical Models*, volume 73, issue 1, pp.10-19, 2011.

- **Selected Conference Papers**

  1. [**ICCV'11a**] Hongsheng Li, Junzhou Huang, Shaoting Zhang, and Xiaolei Huang: Optimal Object Matching via Convexification and Composition, accepted by *13th International Conference on Computer Vision*, 2011.

  2. [**ICCV'11b**] Tian Shen, Xiaoleo Huang, Hongsheng Li, Shaoting Zhang, and Junzhou Huang: A 3D Laplacian-Driven Parametric Deformable Model, accepted by *13th International Conference on Computer Vision*, 2011.

  3. [**MICCAI'11a**] Shaoting Zhang, Junzhou Huang, Mustafa Uzunbas, Tian Shen, Foteini Delis, Xiaolei Huang, Nora Volkow, Panayotis Thanos and Dimitris N. Metaxas: 3D Segmentation of Rodent Brain Structures Using Hierarchical Shape Priors and Deformable Models, accepted by *14th Annual International Conf. on Medical Image Computing and Computer Assisted Intervention*, 2011.

  4. [**MICCAI'11b**] Shaoting Zhang, Yiqiang Zhan, Maneesh Dewan, Junzhou Huang, Dimitris Metaxas and Xiang Zhou: Deformable Segmentation via Sparse Shape Representation, accepted by *14th Annual International Conf. on Medical Image Computing and Computer Assisted Intervention*, 2011. (**MICCAI Young Scientist Award Runner-Up**)

  5. [**MICCAI'11c**] Scott Kulp, Mingchen Gao, Shaoting Zhang, Zhen Qian, Szilard Voros, Leon Axel and Dimitris Metaxas: Using High Resolution Cardiac CT Data to Model and Visualize Patient-Specific Interactions Between Trabeculae and Blood Flow, accepted by *14th Annual International Conf. on Medical Image Computing and Computer Assisted Intervention*, 2011.

  6. [**MICCAI'11d**] Ting Chen, Yiqiang Zhan, Shaoting Zhang and Maneesh Dewan: Automatic Alignment of Brain MR Scout Scans using Data-adaptive Multi-structural Model, accepted by *14th Annual International Conf. on Medical Image Computing and Computer Assisted Intervention*, 2011.

  7. [**CVPR'11**] Shaoting Zhang, Yiqiang Zhan, Maneesh Dewan, Junzhou Huang, Dimitris Metaxas, Xiang Sean Zhou: Sparse Shape Composition: A New Framework for Shape Prior Modeling, *In Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011.

  8. [**FIMH'11a**] Mingchen Gao, Junzhou Huang, Shaoting Zhang, Zhen Qian, Szilard Voros, Dimitri Metaxas and Leon Axel: 4D Cardiac Reconstruction Using High Resolution CT Images , *In the Proc. of the International Conference on Functional Imaging and Modeling of the Heart*, LNCS 6666, 2011. (**Best Paper Award**)

  9. [**FIMH'11b**] Shaoting Zhang, Mustafa Uzunbas, Zhennan Yan, Mingchen Gao, Junzhou Huang, Dimitri Metaxas and Leon Axel: Construction of Left Ventricle 3D Shape Atlas from Cardiac MRI , *In the Proc. of the International Conference on Functional Imaging and Modeling of the Heart*, LNCS 6666, 2011.

  10. [**ISBI'11**] Shaoting Zhang, Junzhou Huang, Mustafa Uzunbas, Tian Shen, Foteini Delis, Xiaolei Huang, Nora Volkow, Panayotis Thanos, Dimitris Metaxas: 3D Segmentation of Rodent Brain Structures Using Active Volume

Model With Shape Priors , *In the Proc. of the IEEE Int'l Symposium on Biomedical Imaging: From Nano to Macro*, 2011. (oral presentation)

11. [**MICCAI'10**] Junzhou Huang, Shaoting Zhang and Dimitris Metaxas: Efficient MR Image Reconstruction for Compressed MR Imaging, *In Proc. of the 13th Annual International Conference on Medical Image Computing and Computer Assisted Intervention*, LNCS 6361, pp. 135-142, 2010. (**MICCAI Young Scientist Award**)

12. [**ECCV'10**] Junzhou Huang, Shaoting Zhang and Dimitris Metaxas: Fast Optimization for Mixture Prior Models, *In Proc. of the 11th European Conference on Computer Vision*, LNCS 6313, pp. 607-620, 2010.

13. [**CVPR'10a**] Shaoting Zhang, Junzhou Huang, Yuchi Huang, Yang Yu, Hongsheng Li and Dimitris Metaxas: Automatic Image Annotation Using Group Sparsity, *In Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3312-3319, 2010. (**oral presentation, acceptance rate 4.5%**)

14. [**CVPR'10b**] Yuchi Huang, Qingshan Liu, Shaoting Zhang and Dimitris Metaxas: Image Retrieval via Probabilistic Hypergraph Ranking, *In Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3376-3383, 2010.

15. [**ISBI'10**] Shaoting Zhang, Junzhou Huang, Wei Wang, Xiaolei Huang and Dimitris Metaxas: Discriminative Sparse Representations for Cervigram Image Segmentation, *In Proc. of the IEEE Int'l Symposium on Biomedical Imaging: From Nano to Macro*, pp. 133-136, 2010.

16. [**EG'10a**] Shaoting Zhang, Andrew Nealen and Dimitris Metaxas: Skeleton Based As-Rigid-As-Possible Volume Modeling, *In Proc. of Eurographics short papers*, 2010.

17. [**EG'10b**] Peter Borosan, Reid Howard, Shaoting Zhang and Andrew Nealen: Hybrid Mesh Editing, *In Proc. of Eurographics short papers*, 2010.

18. [**IBSI'09**] Shaoting Zhang, Xiaoxu Wang, Dimitris Metaxas, Ting Chen and Leon Axel: LV Surface Reconstruction From Sparse tMRI Using Laplacian Surface Deformation And Optimization, *In Proc. of the IEEE Int'l Symposium on Biomedical Imaging: From Nano to Macro*, pp. 698-701, 2009.

19. [**MICCAI'08**] Xiaoxu Wang, Ting Chen, Shaoting Zhang, Dimitris Metaxas, and Leon Axel: LV Motion and Strain Computation from tMRI Based on Meshless Deformable Models, *In Proc. of the 13th Annual International Conference on Medical Image Computing and Computer Assisted Intervention*, LNCS 5241, pp. 636-644, 2008.

Patents

- [**Siemens**] Shaoting Zhang, Yiqiang Zhan, Maneesh Dewan, Ting Chen and Xiang Zhou: System and Methods for Robustly Learning Shape Statistics Using Sparse Representation, invention disclosure, Docket Number: 2010E20120 US.

- [**Siemens**] Ting Chen, Maneesh Dewan, Yiqiang Zhan, Shaoting Zhang and Xiang Zhou: Adaptive Multi-Structural Atlas Construction for Auto Align System, invention disclosure, Docket Number: 2010E17098 US.

- [**SJTU**] Lixu Gu and Shaoting Zhang: Real-time Collision Detection in Surgical Simulation System, filed in China, application number: 200610147640.0

- [**SJTU**] Lixu Gu, Jingsi Zhang, Shaoting Zhang and Pengfei Huang: Simulation of Cutting and Stitching for Deformable Models, filed in China, application number: 200610147641.5

- [**SJTU**] Lixu Gu, Pengfei Huang and Shaoting Zhang: Deformation Simulation Using Skeleton Based Mass-Spring Model, filed in China, application number: 200610148207.9