SHAPE AND APPEARANCE INFORMATION INTEGRATION IN MEDICAL IMAGE ANALYSIS AND COMPUTER VISION

BY XIAOLEI HUANG

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

New Brunswick, New Jersey

May, 2006

ABSTRACT OF THE DISSERTATION

Shape and Appearance Information Integration in Medical Image Analysis and Computer Vision

by Xiaolei Huang

Dissertation Director: Prof. Dimitris N. Metaxas

In many fundamental problems in medical image analysis and computer vision, object shape and appearance are two important sources of information that when integrated, provide more reliable image interpretation. In this thesis, I propose a new perspective on how to represent the shape, appearance, and deformations of an object of interest in images. Under this new perspective, I develop novel and efficient algorithms that coherently integrate shape and appearance information to augment traditional shape-only deformable models for robust segmentation, to better perform shape alignment and registration in arbitrary dimension, to register images of single or multiple modalities in a joint shape and appearance feature space, and to build a new framework for learning statistical shape and appearance models that requires significantly less human effort than the well-known Active Shape and Appearance Models. I also present several applications of this research on topics including heart wall motion analysis in tagged MRI images, and tracking/learning/transfer of 3-D facial expressions.

Acknowledgements

I am grateful to Professor Dimitris Metaxas for his advice, encouragement, trust, and support over my Ph.D. years. He has been an excellent advisor and always directed me toward doing fundamental research that will make a difference. He also exposed me to infinite possibilities by doing research in Medical Image Analysis, Computer Vision and Graphics. None of the work in this thesis would have happened without him.

I would like to thank the other members of my doctoral committee: Prof. Leon Axel, Prof. James Duncan, Prof. Casimir Kulikowski and Prof. Vladimir Pavlovic, for their advice, help and valuable suggestions regarding this thesis. It is a privilege for me to have each of them serve in my committee.

I also thank Prof. Nikos Paragios, and Dr. Chenyang Xu. They have been wonderful mentors to me, and contributed many ideas and insights to the work that I present in this thesis.

Thanks also go to many professors at Rutgers that have helped and supported me through the long process, especially to Prof. Doug Decarlo, Prof. Peter Meer, Prof. Sven Dickinson, Prof. Matthew Stone, Prof. Manish Singh, Prof. Ahmed Elgammal, Prof. Eric Allender, and Prof. William Steiger.

Special thanks to my friends and colleagues from the Center for Computational Biomedicine Imaging and Modeling (CBIM), and the VILLAGE lab. I benefited a lot from their friendship and help, and they made my years at Rutgers a real pleasure.

Finally, thanks go to many of my collaborators without whose help I would not have achieved as much. Special thanks to Prof. Leon Axel at New York University, Dr. Yiyong Sun at Siemens Corporate Research, Prof. Dimitris Samaras and Prof. Peisen Huang at the State University of New York - Stony Brook, for providing crucial data that are used in my doctoral research.

Dedication

To my husband, Gang Tan, and to my parents, Qingke Huang and Fengzhu Yang. For their love and support I work and I enjoy.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Shape refers to geometry, or any spatial attributes of an object as defined by its boundary and outline. In a general sense, shape also represents the spatial arrangement or composition of perceptual structures, geometric features or patterns on an object.

Appearance refers to visible aspects of an object, including color, gray-level intensity distribution, texture, visual patterns, appearance of constituent parts, among others.

Since images are functions not only of object shape but also appearance properties, the two sources of information are often used complementarily in order to develop robust solutions for many computer vision and medical image analysis problems. In this thesis, we focus on two important problems: segmentation and registration.

1.1 Segmentation

Segmentation is the partition of an image domain I into several constituent subsets. Since there are many possible partitions, the "right" one is often pursued in the context of prior world knowledge. This prior knowledge can be low level, such as coherence criteria on the brightness, color, texture or motion within each subset. Or equally important is the knowledge in mid-level and high-level, such as statistics on the shape and appearance of objects that appear in the image. During the past decade, researchers have realized that it is difficult to solve the segmentation problem robustly using low-level image processing alone, because of the common presence of image noise, cluttered objects, nonuniform object texture, variations in lighting, and various other artifacts in natural or medical images. To address these difficulties, model-based methods have been extensively studied and widely used, with considerable success because of their ability to integrate high-level knowledge about object shape and appearance properties with low-level image processing.

The models being used can be either *deformable models* [62, 106, 21, 14, 70], or *statistical shape and appearance models* [26, 25, 65]. *Deformable models* are curves or surfaces that deform under the influence of internal smoothness and external image forces to delineate object boundary. Compared to low-level edge detection methods, deformable models have the advantage of estimating boundary with smooth curves or surfaces that bridge over boundary gaps. However, traditional shape-only deformable models that use edge information alone are often sensitive to image noise, spurious edges, and produce results that are highly dependent on model initialization. *Statistical shape and appearance models* are learned *a priori* from examples to capture variations in the shape and appearance of an object of interest in images. When applied to segmentation, the models deform toward object boundary but with constraints to deform only in ways characteristic of the object they represent. These statistical models encode high-level knowledge in a more specific manner and are often more robust for image interpretation; yet they require more efforts because they need collection/annotation of training data, alignment/registration of training examples, and learning of statistics for positive (and negative) examples using generative (or discriminative) classifiers.

In both types of models, integrating region statistics constraints (i.e. appearance statistics) into boundary (i.e. shape) based models has been central toward more robust, well-behaved models in boundary extraction and segmentation. Along the line of *deformable models*, region-based strategies have been proposed to dynamically estimate intensity/texture statistics of the region inside a deformable model using parametric (e.g. Gaussian, Mixture-of-Gaussian) or nonparametric methods, and to derive model deformations that ensure the statistical coherence inside the model. These strategies are usually formulated in energy minimization [126, 83, 122, 53] or stochastic optimization frameworks [90, 18] to derive region-based forces on the deformable model, which is also under the influence of image gradient (i.e. edge) forces. This way, both region and edge forces work complementarily to aid the model overcome local minima due to spurious edges, and to prevent the model from leaking at boundary gaps. Along the line of *statistical prior models*, using statistical shape models to guide image search produces reliable segmentation results in noisy, cluttered images [26, 65]. A generalization to statistical appearance models uses also the interior region information [25], and enables registration of a target object with the learned prior model. Being complementary to each other,

the integration of statistical shape and appearance models results in a powerful image analysis paradigm.

1.2 Registration

Registration is the process of establishing point-by-point correspondences between images of a scene or shapes of an object captured from different view points. For that purpose, parameters of global transformation and/or local deformation models are to be recovered to geometrically transform a *moving* image/shape to achieve high spatial correspondence with a *fixed* image/shape. Example global transformation models include rigid, similarity, affine; and example local deformation models include displacement vector field, thin plate spline, free form deformations, and so on. The registration problem has been widely studied because it is important in various computer vision and medical image analysis applications, such as object recognition, tracking, image fusion, change detection, and stereo depth perception.

While both shape and image registration are important and have been studied intensively independent of each other, registration in a joint feature space that considers shape and image (here image refers to appearance, intensity, texture, etc.) simultaneously is interesting and in many applications necessary. In the application of learning *statistical shape and appearance models*, training examples, which are image regions covered by instances of a target object, need to be registered before meaningful statistical features could be extracted from corresponding elements. Many existing methods [26, 28, 65, 34] focus on establishing correspondences between basic elements of the boundary shapes; a few [25] further apply interpolation to propagate the shape registration field into the areas inside. However, registering training examples based on boundary shape alone may fail for some objects of interest such as those with symmetric (e.g. near circular) shapes but varying interior appearance. Joint registration using shape and appearance uses all the information in the image region covered by the target object; it provides additional deformation constraints for the large area inside the object, hence leads to more robust and accurate correspondences. In the application of registering two images of a scene, hybrid methods [104, 49, 63, 56] that integrate geometric features (such as landmark

points, edge curves/surfaces) with intensity values from the full image content, are gaining popularity. These methods combine the advantages of shape-based and intensity-based registration methods, and they are more flexible in that their frameworks allow edges or other salient shape structures to be weighted higher than average pixels during registration.

1.3 Investigations on Shape and Appearance Information Integration

As shape and appearance are two complementary sources of information that when integrated, provide the most reliable results, how to achieve this integration is nontrivial and has been a problem that arose in the context of model-based segmentation, registration, tracking, among others.

In model-based segmentation, the main difficulty lies in the fact that traditionally, shape and appearance have very different representations. Shapes are usually represented by point sets or parametric curves or surface, while appearances are captured using statistical features (such as mean, variance, distribution) of an image region's gray level intensity or texture. Shape parameters usually form a vector which describes the boundary elements, while appearance/region parameters represent accumulative statistics that do not have a vector structure. This large difference in their representation spaces makes it difficult to unify shape and appearance in one optimization process. As a result, shape (or boundary) and appearance (or region) information are often accounted in separate optimization processes, and boundary and region parameters are updated iteratively. For instance, in the literature of deformable model based segmentation, region analysis strategies were proposed [90, 126, 59, 18] to augment the "snake" (active contour) models for segmentation. In [126], a generalized energy function that combines aspects of snakes/balloons and region growing is proposed and the minimization of the criterion is guaranteed to converge to a local minimum. This formulation approximates the region intensity statistics using parameters of a Gaussian distribution, while the model shape is represented by a parametric spline curve. The differences in representation prevent the use of gradient descent methods to update both region parameters and shape parameters in a unified optimization process; hence the two sets of parameters are not updated simultaneously in [126], rather they are estimated in separate steps and the energy function has to be minimized in an iterative way.

In other hybrid segmentation frameworks such as those proposed by [18, 59], a region based segmentation module is used to get a rough binary mask of the object of interest. Then this rough estimation of the object can be used to initialize a deformable model, which will deform to fit edge features in the image using gradient information. However, in these frameworks, the region-based and edge-based modules are still separate energy minimization processes, so that the integration is still imperfect and errors from one module can hardly be corrected by the other.

For registration, hybrid methods [104, 49, 63, 56] are gaining increasing attention because they integrate geometric features, which are from shape information, with intensity values from the full image content, which are appearance information. Most of these hybrid methods focus on incorporating geometric feature constraints into the intensity-based energy functionals to achieve smoother and faster optimization. Example geometric features include object boundary contours, landmark points, edges, curves, and/or surface patches [13, 71]; and popular intensity-based energy functionals are sum of squared differences, mutual information, cross correlation, etc. This type of integration is very effective for global registration, because the global transformation models such as rigid and affine are applicable to both shape and image, hence both feature-based and intensity-based criteria can be defined in one optimization framework with respect to the common transformation parameters. The integration becomes problematic however during local non-rigid registration. The main difficulty lies in the differences between shape and intensity representations, and many popular non-rigid deformation models are not applicable to both shape and intensity. For instance, the optical flow like local deformation model, which defines a deformation vector at every pixel, is widely used to register intensity images; it is not suitable for registering shapes, however, because it does not guarantee preserving the topology and coherence of a shape after deformation even with advanced regularization and smoothness constraints (e.g. a closed shape can be deformed to an open structure, or vice versa) [85]. Another popular non-rigid deformation model, the Thin Plate Spline (TPS), can be used to derive a dense deformation field given correspondences between two sets of sparse landmark points [11]. Although TPS can represent deformations for both boundary shape and interior image region, it requires explicitly finding correspondences (between points, regions, etc.), which is a hard problem that lacks very robust solutions.

1.4 Thesis Outline and Contributions

In this thesis, we propose a new perspective on how to represent the shape, appearance, and deformations of an object of interest in images. Under this perspective, we are able to develop new algorithms that integrate shape and appearance information in unified frameworks and solve effortlessly many problems that we discussed in previous integration approaches.

The main contributions of this thesis are:

1. This thesis identifies problem domains in which the integration of shape and appearance information helps improve robustness and efficiency.

2. It reviews previous work on shape and appearance integration in various problem domains and analyzes their limitations.

3. It proposes a new perspective on how to integrate shape and appearance information more naturally through choosing proper shape, appearance, and deformation representations.

4. It introduces a new class of deformable models, Metamorphs, which naturally integrate boundary shape and region appearance statistics for robust model-based segmentation.

5. It presents a global-to-local registration framework which can be applied to registration in the shape space, in the intensity space, as well as in the joint shape and intensity feature space. The correspondences established by the registration framework are used in learning statistical shape models, statistical shape and appearance models, and the learned prior models guide robust image search and object segmentation.

6. The thesis also presents an image registration algorithm that integrates shape context and image intensity through finding good correspondences between salient "region" features.

7. The algorithms introduced in this thesis were implemented in Matlab, C, or C++, and various prototype systems are demonstrated on real-world applications such as segmenting the heart in MRI images, segmenting lesions in breast or prostate ultrasound images, registering 3D face range scans, tracking 3D facial expression, registering images of single or multiple modalities, among others.

This thesis is organized as follows. Chapter 2 introduces the new perspective of representations. Shape is represented implicitly as distance map ”images”. Object appearance is captured using nonparametric kernel-based density estimation of intensity/texture distributions. An Incremental Free Form Deformations (IFFD) model is proposed to be a unified deformation model for both shape and appearance. Using these representations, we are able to develop novel algorithms for model-based segmentation, registration and visual learning that tightly couple shape and appearance and achieve more robust results.

Chapter 3 introduces Metamorphs, a new class of deformable models that dynamically integrate model-interior region statistics with edge information during model-based segmentation. The new models use the shape, appearance and deformation representations in chapter 2, and by doing so, they can naturally integrate region (appearance) and edge (shape) information to derive model deformations in a unified variational framework. Furthermore, a Metamorphs model has an on-learning aspect that constraints the model deformations such that the interior statistics of the model after each deformation is consistent with the statistics learned from the past history of model interiors. The extension of Metamorphs segmentation to images with large-scale textures is also presented.

Chapter 4 presents new algorithms for learning coupled prior shape and appearance models based on representations in chapter 2. First, a new global-to-local shape registration algorithm is introduced. It can be used to establish continuous, smooth and one-to-one correspondences between shapes in arbitrary dimension; in particular, it can register boundary shapes of training examples and establish correspondences between them in order to learn a statistical shape model. Second, as a natural extension of the shape registration algorithm, a new joint registration algorithm is introduced to register images (or training examples) in a joint shape and intensity feature space. It establishes correspondences for shapes and interior textures simultaneously by maximizing mutual information in both shape and intensity spaces. Third, the dense correspondences are used to build a coupled shape and appearance statistical model, then the model is applied to robust segmentation and image interpretation.

The algorithms in chapter 3 and 4 integrate shape and appearance in unified energy minimization frameworks and solve system parameters through gradient-descent optimization. Chapter 5 introduces a hybrid image registration algorithm on the other side of the spectrum. The

basic idea is to first detect feature correspondences, then solve transformation parameters in a least-squares sense and in closed-form. The integration of shape and appearance is realized through feature detection and correspondence finding. Rather than using traditional geometric features such as curvature extreme points, curves/surface patches, our method detects salient "region" features, each of which has an associated scale and whose interior intensities (appearance) can be matched using similarity measures such as mutual information. Shape information is incorporated by considering geometric configuration constraints between the region features during correspondence finding. The geometric configuration constraints are enforced in an Expectation-Maximization framework to find a joint correspondence between multiple pairs of region features that result in a consistent transformation; other feature pairs, which either are outlier matches or degrade matching performance, are effectively pruned.

Several real-world applications of the algorithms proposed in this thesis are presented in Chapter 6. In one application, the Metamorphs deformable models introduced in Chapter 3 are applied to heart wall motion tracking in noisy, tagged MRI images of the heart. In another application, the Global-to-local shape registration algorithm introduced in Chapter 4 is used for high-resolution 3D facial expression tracking, learning, transfer and synthesis.

Finally conclusions are drawn, and future research directions are outlined in chapter 7.

Chapter 2

Representations for Shape and Appearance Integration: On Shape, Appearance and Deformations

This chapter introduces representations for shape, appearance and deformation that will facilitate the integration of shape and appearance information in various computer vision and medical image analysis applications.

2.1 Background

In the literature, both shape and appearance representations have been intensively studied. Different representations are used to suit the need of different applications.

2.1.1 Review on Shape Representations

Shape refers to geometry, or any spatial attributes of an object as defined by its boundary and outline. A common classification of shape representations has three categories. First, an explicit shape representation describes the set of points that belong to an object boundary explicitly. Examples are point clouds (or point sets) [7, 19], binary voxel grids, or Octree [116]. Second, a parametric representation encodes important shape information using a few parameters by finding appropriate mathematically-complete mapping functions. Examples are parametric curves/surfaces [29, 74] such as triangle meshes and NURBS, harmonic representations such as fourier descriptors [106], medial axes [101], among others. Third, an implicit representation embeds a shape in a higher dimensional space and describes it as an iso-surface of a spatial field function in the embedding space. Examples are the level set shape representations, in which a shape is represented as the zero level set of a higher dimensional distance function [81, 85, 64].

Comparing the explicit, parametric and implicit shape representations, each has its own merits and drawbacks.

- The *explicit* point clouds representation is intuitive and generic since it can easily represent shapes in 2D and 3D with arbitrary topology. However, its known limitation when used for shape matching and registration is that, it strongly depends on the sampling rule which affects the number of shape elements, their distribution, etc. For instance, given two shapes to be registered, each represented by a point cloud, the two point sets may not be sampled at corresponding locations due to low-resolution or improper sampling, and this can lead to inherent inconsistencies in the two point sets, thus cast problems when point correspondences are pursued between the two shapes [20].

- Unlike point clouds, the *parametric* curves/surfaces shape representation supports valid correspondences in a continuous domain. Its main disadvantage is in the difficulties to parameterize shapes in high dimensions and/or with complex topology. Fourier descriptors [106] and medial axis [101] are two other *parametric* shape representations that are excellent when measuring the dissimilarity between shapes, but they do not support a vector description of shape boundary elements so that they are not suited for registration when dense correspondences need to be established between shape boundary elements.

- The implicit shape representation is gaining increasing attention recently, both in shape registration [85] and in statistical shape modeling [64]. It is attractive in that it is a generic representation that handles naturally shapes of arbitrary dimension and arbitrary topology. This is because it represents shapes using the distance map ”images” derived from their distance transforms and it does not require parameterization of the shapes. The representation is also stable and robust to shape perturbations and noise, as shown by the formal proofs given in [128]. The main concern associated with the implicit shape representation is in its computational complexity since it embeds a shape in a higher dimensional space. Another concern is in its topology freedom when preserving shape topology is desired after deformation.

2.1.2 Review on Appearance Representations

A representation of object appearance in images encodes brightness variations caused by 3D shape, surface reflectance properties, sensor parameters, illumination conditions, etc. There are

two types of appearance representations. The first type encodes the appearance of an object in one image of interest or in an image sequence. In applications such as segmentation, registration and tracking, this type of representation is commonly used online when there is no prior knowledge available about what the object looks like, hence low-level coherence properties on the intensity, color, texture, or motion of the object are assumed. The second type of representation encodes the appearance and appearance variations among a large set of images which are all for the same class of object and are all correlated to some degree. Parametric representations have been dominant which compress this large image set and map it to a low dimensional manifold [79]. Applications of the second type of representation include visual learning, recognition and prior-model based segmentation, in which mid-level and high-level knowledge are learned off-line to specify the appearance statistics of an object of interest in images.

Online Appearance Representation

Without any prior knowledge, many appearance representations of the first type make assumptions on the intensity distribution of all pixels inside an object. Gaussian, Mixture-of-Gaussian models are common assumptions on the distribution. If a Gaussian distribution is assumed, the intensities of an object are parameterized by a mean intensity μ and a variance σ^2:

$$\mathbf{P}(i|O) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(i-\mu)^2}{2\sigma^2}} \tag{2.1}$$

where $i = 0, ...255$ denotes an intensity value, and O represents pixel intensities inside the object. If a Mixture-of-Gaussian assumption is made, the object intensities are parameterized by the means and variances of several component functions, each being a Gaussian:

$$\mathbf{P}(i|O) = \sum_{n=1}^{K} \frac{1}{\sqrt{2\pi\sigma_n^2}} e^{\frac{-(i-\mu_n)^2}{2\sigma_n^2}} \tag{2.2}$$

where K components are assumed and for each component $n = 1, ..., K$, the mean intensity is μ_n, and the variance is σ_n^2.

Other than parametric models such as Gaussian and Mixture-of-Gaussian, nonparametric kernel-based density estimation (also known as the Parzen windows technique in pattern recognition literature [32]) is a popular nonparametric statistical method [99]. It represents a generalization of the Mixture-of-Gaussian model, where it does not make assumptions about the number of modes in a distribution, rather it treats every single sample as a Gaussian kernel and integrates over these small Gaussian kernels to derive the overall nonparametric estimation of the probability density function (p.d.f.). Recently this nonparametric technique has been applied to imaging and computer vision, most notably in modeling the varying background in video sequences [35], and in approximating multi-modal intensity density functions of color images [23]. In this thesis, we use this nonparametric representation to approximate the p.d.f. of a deformable model's interior intensities. Detailed formulation of the representation is given in Section 2.3. One important advantage of the nonparametric representation is that it can be approximated directly from pixel intensities inside a deformable model, requiring no parameter estimation. Furthermore, when the model deforms, its interior (pixels) changes, hence the nonparametric intensity p.d.f. gets updated automatically.

Although the nonparametric p.d.f approximation can represent object (or model) interior appearance with complex multi-modal intensity distributions, it mostly deals with pixel-wise intensity statistics, and does not account for the spatial correlation (i.e. context information) between neighboring pixels, nor the scale and pattern of the basic building blocks such as texture elements that constitute the object (or model) interior. This limits the nonparametric representation's applicability in scenarios where the object has texture with large-scale periodic patterns, or when the background has very similar intensity distribution with the object, but very different texture because the scale and pattern of the texture elements are different. To address this limitation, we introduce a hierarchy featuring three levels of texture representation. The bottom level takes the same form as the nonparametric intensity p.d.f. representation. The middle level determines a local "best" scale for the model-interior texture element, by minimizing the symmetrized Kullback-Leibler Divergence (KLD) between the intensity p.d.f. within the local scale and the overall model interior p.d.f. The third level further considers the spatial correlation between pixels within the local scale, by constructing a small gabor filter bank targeted at segmenting the specific texture patterns that appear inside the model. Detailed

formulation for the hierarchy of texture representations is given in Section 2.3.

Offline Learning of Appearance Representation

Unlike the first type of appearance representations, which capture an unknown object's intensity/texture statistics during run time, the second type of appearance representations are statistical models learned off-line to encode the appearance and appearance variations among a large set of images which are all for the same class of known object. Given a new image that is known to contain the object, the learned prior model can be used to guide the detection and robust segmentation of the object.

Because of variations in shape, surface reflectance properties, illumination and other conditions, the appearance of an object can vary significantly in the large set of training images; yet because these images are all of the same object, they are often correlated to a large degree. So one dominant approach for learning statistical prior models has been using generative parametric methods to compress the large training image set to a low dimensional representation of object appearance. The most intensively studied method for this purpose are linear dimension-reduction methods such as Principal Component Analysis (PCA) [25, 79], and nonlinear dimension-reduction methods such as Kernel PCA [89], isometric feature mapping (Isomap) [108], and local linear embedding (LLE) [91, 36]. Another approach that is popular in object detection is to use supervised or semi-supervised learning methods to train a classifier that is able to differentiate between positive and negative examples of an object. Commonly used learning methods along this line are AdaBoost [113, 43], Support Vector Machines (SVM) [82], Neural Networks [92], MAP decision rules [98], and Co-training[66].

2.1.3 Representations for Integrating Shape and Appearance

In this thesis, we propose a new perspective of representations that are well suited for integrating shape and appearance information. The proposed representations are naturally coupled and allow us to develop novel segmentation and registration algorithms that integrate shape and appearance and achieve more robust results.

Within the proposed framework, we represent each shape using the implicit representation. In this way, shapes are implicitly represented as "images" in the space of distance transforms

where shapes correspond to the zero level set of the distance functions. The level set values in the shape embedding space is analogous to the intensity values in the intensity space. As a result, we can encode the shape (or edge) information in one "image", and intensity (or appearance) information in another. The two kinds of images are images of different modality, but they are parallel to each other, and both can be manipulated using pixels on a common regular grid. The shape and intensity spaces can therefore be conveniently unified this way.

To capture the intensity/texture statistics of an object, we use the nonparametric kernel-based approximation (with scale and pattern detection for textured images). With this representation, the statistics of object (or deformable model interior) appearance do not require extra parameters; the probability density functions can be approximated directly given pixels inside the object (or model), and they are updated automatically when the object (or model) deforms.

Another key component of the framework is an Incremental Free Form Deformations (IFFD) representation, which is proposed to serve as a unified deformation model for both shape and appearance. IFFD is an extension of Free Form Deformations (FFD) [102, 94, 54], which is a popular approach for modeling deformations in graphics, animation and rendering. It is a space warping technique, which consists of embedding an object inside a space, and deforming the object through deforming the space. It couples naturally with the implicit shape representation, where it deforms a shape via deforming its embedding "image" space. It can also deform an intensity image by deforming a regular control lattice overlaid on the image. As a result, IFFD is the most attractive in the unified shape and appearance space, where its deformation can be derived based on information from both the implicit shape "image" and the explicit intensity image. Other than FFD, the optical flow like local deformation field and the Thin Plate Splines are two other popular non-rigid deformation techniques. However, when dealing with both shape and intensity, they have their limitations which are discussed in Section 1.3.

In the remainder of this chapter, we present in detail the formulations for the shape, appearance, and deformation representations in our framework.

2.2 Implicit Shape Representation as Distance Map "Images"

To facilitate the integration of shape and intensity, we use the implicit shape representation in our framework. The Euclidean distance transform is used to embed a shape of interest as the zero level set of a distance function in the higher dimensional volumetric space. The distance transform is invariant to translation and rotation, and one can predict the effect of scale variations on such representation [85]. Hence it is a powerful selection to implicitly represent a shape in arbitrary dimension.

We first consider the 2D case. Let $\Phi : \Omega \rightarrow R^+$ be a Lipschitz function that refers to the distance transform of a shape $\mathcal{M}$. By definition Ω is bounded since it refers to the image domain. The shape defines a partition of the image domain Ω: the region that is enclosed by $\mathcal{M}$, $[\mathcal{R}_{\mathcal{M}}]$, and the background region $[\Omega - \mathcal{R}_{\mathcal{M}}]$. Given these definitions, the following implicit shape representation is considered:

$$\Phi_{\mathcal{M}}(x,y) = \begin{cases} 0, & (x,y) \in \mathcal{M} \\ +\mathcal{D}((x,y),\mathcal{M}), & (x,y) \in \mathcal{R}_{\mathcal{M}} \\ -\mathcal{D}((x,y),\mathcal{M}), & (x,y) \in [\Omega - \mathcal{R}_{\mathcal{M}}] \end{cases} \tag{2.3}$$

where $\mathcal{D}((x,y),\mathcal{M})$ refers to the minimum Euclidean distance between the image pixel location $\mathbf{x} = (x,y)$ and the shape $\mathcal{M}$. If $\mathcal{M}$ is an open structure, the un-signed distance transform is used instead. Some pictorial examples of the implicit shape representation can be found in [Fig. (2.1)].

The representation is similarly defined in 3D. Let $\Phi : \Omega \rightarrow R^+$ be a Lipschitz function that refers to the distance transform of a 3D shape $\mathcal{M}$. Ω refers to the 3D volumetric image domain. Then the implicit shape representation in 3D is defined as:

$$\Phi_{\mathcal{M}}(x,y,z) = \begin{cases} 0, & (x,y,z) \in \mathcal{M} \\ +\mathcal{D}((x,y,z),\mathcal{M}), & (x,y,z) \in \mathcal{R}_{\mathcal{M}} \\ -\mathcal{D}((x,y,z),\mathcal{M}), & (x,y,z) \in [\Omega - \mathcal{R}_{\mathcal{M}}] \end{cases}$$

where $\mathcal{D}((x,y,z),\mathcal{M})$ refers to the minimum Euclidean distance between the 3D image voxel

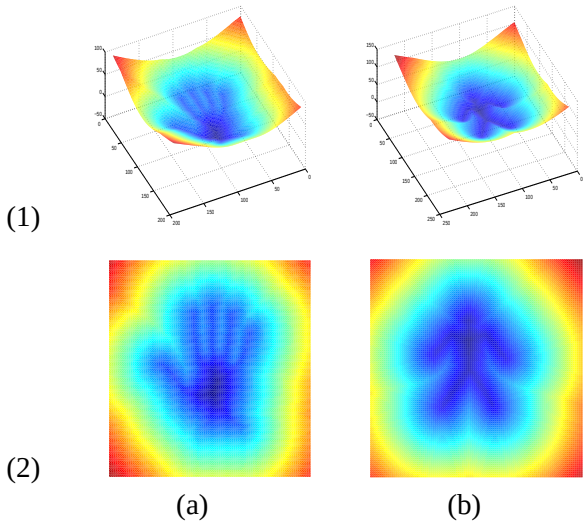


Figure 2.1: Implicit Representations of 2D Shapes: (a) hand, (b) dude. (1) visualization of the higher dimensional embedding space. (2) projection of the implicit representation as a 2D image.

location $\mathbf{x} = (x, y, z)$ and the shape $\mathcal{M}$.

By representing shapes as distance map "images", the implicit shape representation facilitates the integration of shape and intensity information. This representation also provides a feature space in which objective functions that are optimized using a gradient descent method can be conveniently used. One can prove that the gradient of the embedding distance function is a unit vector in the normal direction of the shape; and the representation satisfies a sufficient condition for the convergence of gradient descent methods, which requires continuous first derivatives. Furthermore, in registration problems, the use of the implicit representation provides additional support to the registration process around the shape boundaries and facilitates the imposition of smoothness constraints, since one would like to align the original structures as well as their clones that are positioned coherently in the image/volume plane. Finally, one can refer to recent work [128] that demonstrates the robustness and stability of such representation to noise and shape perturbations.

There are two concerns associated with the implicit shape representation that need to be addressed. The main concern is in its computational complexity since it has one dimension higher than the original shape. In our work, this efficiency problem is addressed by using only a narrow band around the shape in the embedding space as the sample domain for segmentation

or registration purposes. This significantly speeds up the execution, while producing comparable results to that using the full image domain. Another concern is the preservation of shape topology after deformation because the order information between neighboring shape elements are lost in the embedding space. In our framework, this problem is solved implicitly because we use Incremental Free Form Deformations (IFFD) to represent deformations. The IFFD, with cubic B-spline basis functions as the choice of interpolating spline, guarantees C^1 continuity at control points and C^2 continuity everywhere else during deformation. These properties ensure continuous, smooth deformations that preserve the topology and coherence of shapes in their implicit representation.

2.3 Nonparametric Appearance Representations

In our framework, we use the nonparametric appearance representation to approximate the interior intensity or texture statistics of an object or a deformable model. In this way, we do not need any explicit appearance parameters, since the nonparametric representation can be derived from pixel intensities directly. Coupled with the parameter-free implicit shape representation, we can achieve the integration of shape and appearance without separate shape or appearance parameters, and both sources of information can be represented on a pixel-wise basis in the image domain.

We propose three levels of nonparametric appearance representation, which cover the continuum from gray-level intensity to textures with large-scale periodic patterns.

2.3.1 Bottom Level: Nonparametric Kernel-based Intensity p.d.f. Estimation

The first level of nonparametric representation considers accumulative pixel intensity statistics. Suppose a model (deformable model or statistical model), $\Phi_{\mathcal{M}}$, is placed on an image I, the image region bounded by the model is $\mathcal{R}_{\mathcal{M}}$, then the nonparametric kernel-based intensity p.d.f. estimated using a Gaussian kernel is:

$$\mathbf{P}(i\big|\Phi_{\mathcal{M}}) = \frac{1}{V(\mathcal{R}_{\mathcal{M}})} \iint_{\mathcal{R}_{\mathcal{M}}} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(i-I(\mathbf{y}))^2}{2\sigma^2}} d\mathbf{y} \tag{2.4}$$

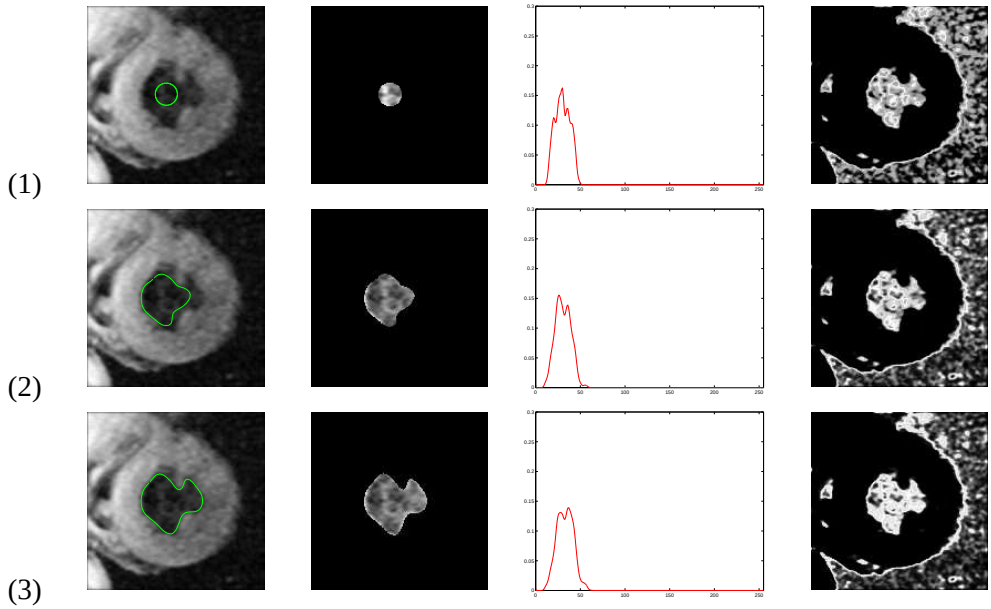


Figure 2.2: The Left Ventricle Endocardium segmentation, demonstrating Metamorphs texture representation. (1) Initial model. (2) Intermediate result after 4 iterations. (3) Final converged result after 10 iterations. (a) The evolving model drawn on original image. (b) Interior of the evolving model. (c) The intensity p.d.f. of the model interior. (d) The image intensity probability map according to the p.d.f. of the model interior.

where $i = 0, ..., 255$ denotes the pixel intensity values, $V(\mathcal{R}_{\mathcal{M}})$ denotes the volume of $\mathcal{R}_{\mathcal{M}}$, $\mathbf{y}$ represents pixels in the region $\mathcal{R}_{\mathcal{M}}$, and σ is a constant specifying the width of the Gaussian kernel (we set $\sigma = 4$ pixels in all our experiments).

One example of this nonparametric density estimation for a deformable model can be seen in Fig. 2.2. The zero level set of the evolving models $\Phi_{\mathcal{M}}$ are drawn on top of the original image in Fig. 2.2(a). The model interior regions $\mathcal{R}_{\mathcal{M}}$ are cropped and shown in Fig. 2.2(b). Given the model interiors, their nonparametric intensity p.d.f.s $\mathbf{P}(i\big|\Phi_{\mathcal{M}})$ are shown in Fig. 2.2(c), where the horizontal axis denotes the intensity values $i = 0, ...255$, and the vertical axis denotes the probability values $P \in [0, 1]$.

Over the entire image I, for any pixel $\mathbf{x}$, with intensity value i, we can also evaluate the probability of this pixel's intensity according to the model interior intensity p.d.f., using Eq. 2.4. This way, we can compute a probability (or likelihood) map as shown in Fig. 2.2(d).

2.3.2 Middle Level: Nonparametric Statistics of Texons

In order to take into account the spatial correlation between neighboring pixels, the middle level appearance representation first determines a "best" natural scale for the texture elements that are basic building blocks of the model interior (or object) texture. (We call such texture elements "texons", following the naming convention in [60, 69, 127].) Then the nonparametric statistics of the texons are estimated.

Best Local Scale for Model Interior Texons

We compute the "best' scale of the texons using a detector based on comparing the texon-interior intensity p.d.f. with the whole model-interior (or object) intensity p.d.f., and we determine the scale of the texon as the smallsest scale that provide a texon p.d.f. that is sufficiently close to the overall model interior p.d.f.

Suppose a model, $\Phi_{\mathcal{M}}$, is placed on image I, and the image region bounded by the model is $\mathcal{R}_{\mathcal{M}}$, we use the nonparametric kernel-based method (Eq. 2.4) to approximate the p.d.f. of the model-interior intensity, $\mathbf{P}(i\big|\Phi_{\mathcal{M}})$. We also denote $\mathbf{P}(i\big|\Phi_{\mathcal{M}})$ as p_m.

Similarly, the intensity p.d.f. for a local texon can also be defined as in Eq. 2.4, the only difference being that the integration is over pixels inside the texon. Let us denote a texon of scale s centered at a pixel $\mathbf{x}$ by $T(\mathbf{x}, s)$, and its interior intensity p.d.f. by $p_{T(\mathbf{x},s)}$. Then $p_{T(\mathbf{x},s)}$ is defined by:

$$p_{T(\mathbf{x},s)} = \mathbf{P}(i\big|T(\mathbf{x}, s)) = \frac{1}{V(T(\mathbf{x}, s))} \iint_{T(\mathbf{x},s)} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(i-I(\mathbf{y}))^2}{2\sigma^2}} d\mathbf{y} \tag{2.5}$$

To measure the dissimilarity between two probability density functions, we adopt an information-theoretic distance measure, the Kullback-Leibler (K-L) Divergence [1]. Since the K-L divergence is asymmetric, we instead use one of its symmetrized relative – the Chernoff Information. The Chernoff Information between p_1 and p_2 is defined by:

$$C(p_2\|p_1) = \max_{0\leq t\leq 1} \; -\log \mu(t)$$

where $\mu(t) = \int [p_1(i)]^{1-t}[p_2(i)]^t di$. A special case of Chernoff "distance" is the Bhattachayya

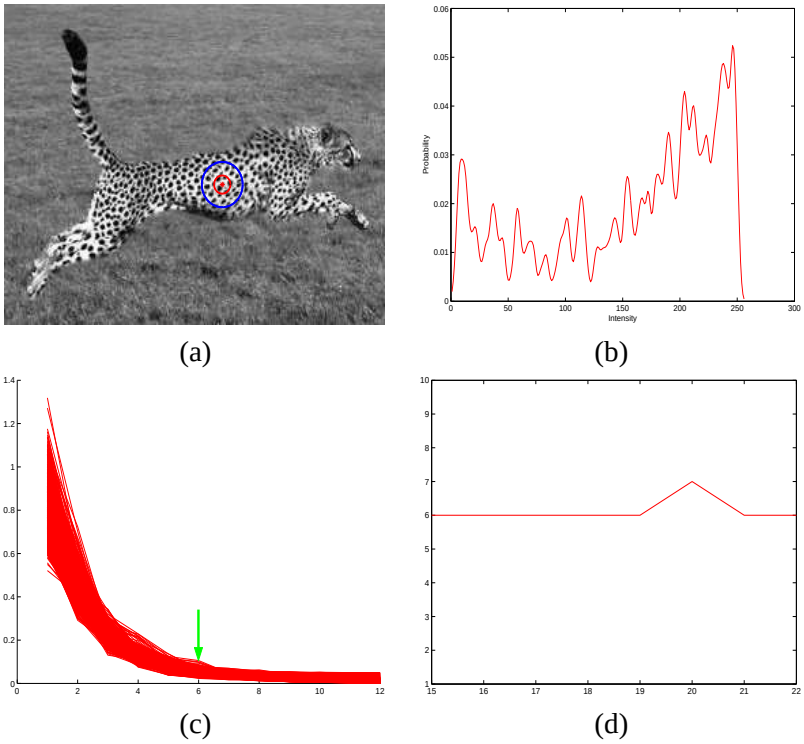


Figure 2.3: (a) Cheeta image. Outer large circle shows a model initialized inside the object of interest, inner small circle shows the determined best scale for model-interior texons. (b) Overall model interior p.d.f. (c) Y axis: K-L distance between texon p.d.f. and overall model-interior p.d.f.; X axis: changing scale (i.e. radius) of the texon under evaluation. Each curve represents a texon centered at a different pixel inside the model. (d) The best scale determined remains stable as we change the size of the model.

"distance", in which t is chosen to be $\frac{1}{2}$, i.e., the Bhattachayya "distance" between p_1 and p_2 is:

$$B(p_2\|p_1) = -\log \mu(\frac{1}{2}) \tag{2.6}$$

In order to facilitate notation, we write:

$$\rho(p_2\|p_1) = \mu(\frac{1}{2}) = \int [p_1(i)]^{\frac{1}{2}} [p_2(i)]^{\frac{1}{2}} di \tag{2.7}$$

Clearly, when the value for ρ ranges from one to zero, the value for B goes from zero to infinity.

In summary, the steps in determining the scale of texons inside a model (or an object) are as follows.

1. Approximate the intensity p.d.f. of the overall model interior (Eq. 2.4). Denote this p.d.f. as p_m.

The p.d.f. for a cheetah example based on an initial model (indicated by the large circle) in Fig. 2.3(a) is displayed in Fig. 2.3(b).

2. Choose a best scale $\hat{s}$ for the model interior texons among all possible scales between $1...S$ [1].

   To determine the best scale $\hat{s}$, we compute the Bhattachayya distance between the model-interior p.d.f., p_m, and the texon intensity p.d.fs, $p_{T(\mathbf{x},s)}$, for all pixels $\mathbf{x}$ inside the model and for all scales $s = 1...S$. Fig. 2.3(c) visualizes the functional relationship between such distances and the scale in a graph. In the graph, each curve represents the "distance-scale" function for texons centered at a different pixel. From the graph, we can see that, as the scale increases, the Bhattachayya distance decreases asymptotically at all pixels, and all curves finally converge at a small value. This behavior proves the validity of the usage of this symmetrized K-L distance measure, and it also exposes to us a way to determine the natural scale of the model interior texons – the scale corresponding to the point of inflection on the Distance-Scale function curves. Since we get a scale value for every pixel inside the model this way, we use a robust estimator, the median estimator, to choose the best scale $\hat{s}$ as the median of the inflection-point scales chosen for all these pixels. On Fig. 2.3(a), we indicate the best scale computed this way by the inner small circle.

Based on our experiments, this "best" natural scale for model-interior texons determined using the method above is invariant to the size of the model. Fig. 2.3(d) shows the functional relation between the best scale chosen vs. the size of the model for the cheetah example. We can see from the curve that the best scale remains stable as the size of the model changes. This behavior is also observed in many other examples that we tested.

Nonparametric Statistics of Texons

Once we have determined the scale $\hat{s}$ for model-interior texons, we can estimate the nonparametric statistics of the texons. We use the kernel-based approximation to capture the probability

[1] Here we assume that the model interior contains at least one texon, and the largest test scale S is smaller than the size of the model.

density function (p.d.f.) of the Bhattachayya distance values (see Eq. 2.6), which are computed as the distances between the p.d.fs of texons of scale $\hat{s}$ and the overall model-interior p.d.f. Suppose a model, $\Phi_{\mathcal{M}}$, is placed on image I and bounds an image region, $\mathcal{R}_{\mathcal{M}}$, then the non-parametric kernel-based Bhattachayya distance value p.d.f. estimated using a Gaussian kernel is:

$$\mathbf{P}(T(\mathbf{x},\hat{s})\big|\Phi_{\mathcal{M}}) = \frac{1}{V(\mathcal{R}_{\mathcal{M}})}\iint_{\mathcal{R}_{\mathcal{M}}} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(B(p_{T(\mathbf{x},s)}\|p_m)-B(p_{T(\mathbf{y},s)}\|p_m))^2}{2\sigma^2}} d\mathbf{y} \qquad (2.8)$$

where

- $V(\mathcal{R}_{\mathcal{M}})$ is the volume of the model-interior region $\mathcal{R}_{\mathcal{M}}$,
- $T(\mathbf{x},\hat{s})$ represents a texon centered at any pixel $\mathbf{x}$ in image I and with scale $\hat{s}$,
- $p_{T(x,\hat{s})}$ is the intensity p.d.f. of the texon $T(\mathbf{x},\hat{s})$,
- p_m is the overall model interior intensity p.d.f.,
- $B(p_{T(\mathbf{x},s)}\|p_m$ is the Bhattachayya distance between the texon p.d.f. and the overall model-interior p.d.f.,
- $T(\mathbf{y},\hat{s})$ represents any texon centered at a pixel y inside the model and with scale $\hat{s}$,
- $p_{T(y,\hat{s})}$ is the intensity p.d.f. of the texon $T(\mathbf{y},\hat{s})$, and
- $B(p_{T(\mathbf{y},s)}\|p_m$ is the Bhattachayya distance between the texon p.d.f. and the overall model-interior p.d.f., and
- σ is a constant specifying the width of the Gaussian kernel.

Using the above Eq. 2.8, we can compute the probability of any texon of scale $\hat{s}$ being consistent with the model-interior texture, thus being part of the object on which the model is initialize. On the cheetah image (Fig. 2.3(a)), the texture probability map computed using this middle level representation is shown in Fig. 2.5(c).

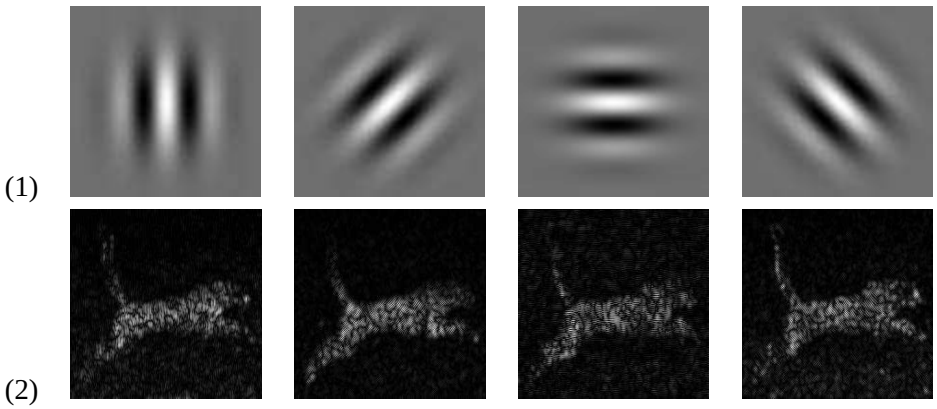


Figure 2.4: (1) Gabor filters in a small bank with constant frequency and shape, and varying orientation. (2) Responses of the cheetah image to Gabor filters in (1).

2.3.3 Top Level: Nonparametric Statistics of Gabor Filter Responses

In the middle level representation, one limitation of using the nonparametric intensity p.d.f. to approximate texon-interior statistics is that, the information on pixel order and spatial correlation between pixels within a texon is lost. For instance, if we take a texon inside the object, randomly re-permute all pixels within it to generate a new texon, then copy this new texon to locations surrounding the object, then the computation in Eq. 2.8 would have trouble differentiating these two kinds of texons, even though they appear different.

To address this problem, the third level texture representation further considers the spatial correlation between pixels within a texon, by applying a small number of gabor filters [27] to the model interior and learning statistics of the Gabor responses.

A small Gabor filter bank with $N(N = 4)$ Gabor filters are shown in Fig. 2.4(1). The filters have constant frequency and shape, but with varying orientations. The frequency and Gaussian-envelop shape of the filters are computed based on the pre-determined scale, $\hat{s}$, of model-interior texons. For each of the N Gabor filters, we get a response image $R_n, n = 1, ..., N$, as shown in Fig. 2.4(2). Then instead of on the original image, we compute the nonparametric texon statistics on each response image, using the middle level representation equation Eq. 2.8. Let us denote the probability density function acquired on R_n as $\mathbf{P}_n(T(\mathbf{x}, \hat{s})\big|\Phi_{\mathcal{M}})$, then the top-level texon statistics based on all Gabor filter responses is defined by:

$$\mathbf{P}(T(\mathbf{x}, \hat{s})\big|\Phi_{\mathcal{M}}) = \prod_{n=1}^{N} \mathbf{P}_n(T(\mathbf{x}, \hat{s})\big|\Phi_{\mathcal{M}}) \tag{2.9}$$

This relation can be easily derived given that the responses from different Gabor-filter bases are conditionally independent of each other.

On the cheetah image (Fig. 2.3(a)), the texture probability map computed using this top level representation is shown in Fig. 2.5(d).

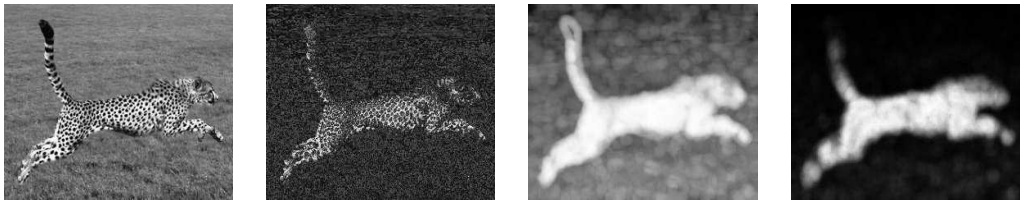
2.3.4 Choosing the Right Level of Appearance Representation

Given the three levels of nonparametric appearance representation, it is often domain-specific as to which level is the right level to use in segmentation problems. From the bottom to the top level, the texture representation gets increasingly more specific so that textures that can not be differentiated by a lower level representation can be differentiated by a higher level representation. On the other hand, as the representation gets more and more specific, an object with gradually varying and non-uniform texture patterns may be partitioned into several parts.

A useful parameter that can assist the choice is the scale of the texons, $\hat{s}$, since this scale reveals to some extent the characteristics of the model-interior texture. If $\hat{s}$ is very small (e.g. the radius of texons is less than 3 pixels wide), the model-interior texture is mostly homogeneous with some level of noise, hence it is not necessary to further consider the spatial correlation between pixels, and the bottom level nonparametric representation (Eq. 2.4) is sufficient and it is the most efficient. An example of this case can be seen from Fig. 2.2. On the other hand, if $\hat{s}$ is rather large, we predict that the model-interior texture consists of periodic mosaics of large-scale patterns, then the bottom level representation may not be appropriate, as can be seen from Fig. 2.5(b). In this case, either the middle level (Fig. 2.5(c)) or the top level (Fig. 2.5(d)) representations can be used, depending on the accuracy and performance requirements for segmentation.

2.4 Dynamics: Deformation Representation for Both Shape and Appearance

In many vision and imaging problems, a key component is the dynamics or deformations that drive one shape or image to another. In model-based segmentation, deformations are to be solved to deform a model to fit an object's boundary. In registration, deformations are to be solved to establish correspondences between two shapes or two images.



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

Figure 2.5: (a) the original cheetah image, (b) likelihood map computed from the bottom level representation using intensity statistics, (c) likelihood map computed from the middle level representation using texon statistics, (d) likelihood map computed from the top level representation using gabor filter response statistics.

In our pursuit of integrating shape and appearance, we found the Free Form Deformations (FFD) is a powerful tool to be a unified deformation model for both shapes and appearances. FFD is a space warping technique that represent the deformations of a space. In our unified shape and intensity feature space, there are two "images" associated with one underlying image domain: one image encodes the shape information using the implicit shape distance map "image", and the other is the intensity image that encodes the appearance information. Hence by deforming the space that corresponds to the common underlying image domain, both shape and appearance are deforming. In the inverse problem, the FFD deformations (or deformations of the space) can be derived from energy functions that are based on information from both the shape and intensity images.

The essence of FFD is to deform an object by manipulating a regular control lattice P overlaid on its volumetric embedding space. The deformation of the control lattice consists of displacements of all the control points in the lattice, and from these sparse displacements, a dense deformation field for every pixel in the embedding space can be acquired through interpolation using an interpolating basis function, such as Bezier spline or B-spline functions. One illustrative example is shown in [Fig. (2.6)]. A circular model [Fig. (2.6).1.a] is implicitly embedded as the zero level set of a distance function [Fig. (2.6).1.b]. A regular control lattice (drawn in green) is overlaid on this embedding space. When the embedding space deforms due to the deformation of the FFD control lattice as shown in [Fig. (2.6).b], the model undergoes an expansion in its object-centered coordinate system. [Fig. (2.6).c] shows another example of free-form deformation given a particular FFD control lattice deformation.

In this thesis, we propose an extension of the FFD technique, which we call the Incremental

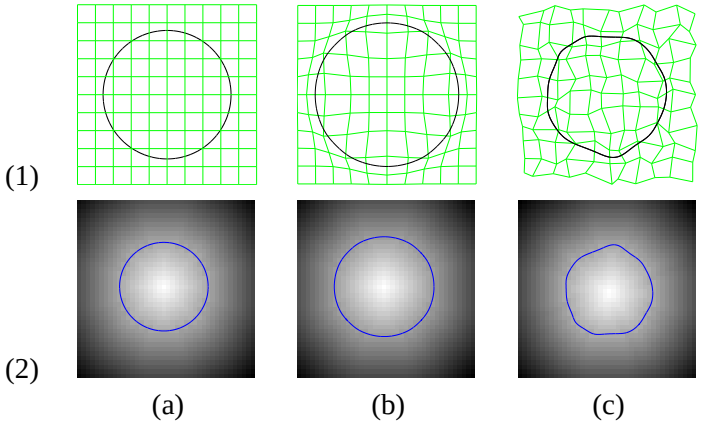


Figure 2.6: Model shape deformations based on FFD. (1) The deforming models. (2) The implicit representations for the model shapes. (a) The initial shape. (b) Example FFD control lattice deformation to expand the shape. (c) Another FFD control lattice deformation to deform the shape in a free-form manner.

Free Form Deformations (IFFD) [54]. IFFD uses the Cubic B-spline functions as the interpolating basis functions. In this way, it enforces smoothness constraints implicitly, guaranteeing C^1 continuity at control points and C^2 continuity everywhere else. As a result, the recovered deformation field is smooth, continuous, preserves shape topology/coherence, and guarantees a one-to-one mapping. The formulation of IFFD is presented below.

2.4.1 IFFD Deformation Formulation

Let us consider a lattice of control points

$$P = \{P_{m,n}\} = \{(P^x_{m,n}, P^y_{m,n})\};\ m = 1, ..., M, \quad n = 1, ..., N \tag{2.10}$$

overlaid to a region $\Gamma_c = \{\mathbf{x}\} = \{(x, y)|1 \leq x \leq X, 1 \leq y \leq Y\}$ in the embedding space that encloses a model (or an object). Let us denote its initial regular configuration with no deformation as P^0 (e.g., [Fig. (2.6).1]), and the deforming configuration as $P = P^0 + \delta P$. Then the IFFD parameters are the deformation improvements of the control points in both x and y directions:

$$\mathbf{\Theta} = \delta P = \{(\delta P^x_{m,n}, \delta P^y_{m,n})\};\ (m, n) \in [1, M] \times [1, N] \tag{2.11}$$

Suppose the control lattice deforms from P^0 to P, the deformed position of any pixel $\mathbf{x} = (x, y)$ in the embedding space is defined by a tensor product of cubic B-splines:

$$D(\mathbf{x}) = \sum_{k=0}^{3}\sum_{l=0}^{3} B_k(u)B_l(v)P_{i+k,j+l} \tag{2.12}$$

where $i = \lfloor \frac{x}{X} \cdot (M-1) \rfloor + 1,\ j = \lfloor \frac{y}{Y} \cdot (N-1) \rfloor + 1$. This is the familiar definition for cubic B-spline based interpolation, and the terms in the formula refer to:

1. $P_{i+k,j+l}$, $(k, l) \in [0, 3] \times [0, 3]$ are the coordinates of the sixteen control points in the neighborhood of pixel $\mathbf{x}$.

2. $B_k(u)$ represents the k^{th} basis function of cubic B-spline:

$$\begin{aligned} B_0(u) &= (1-u)^3/6 \\ B_1(u) &= (3u^3 - 6u^2 + 4)/6 \\ B_2(u) &= (-3u^3 + 3u^2 + 3u + 1)/6 \\ B_3(u) &= u^3/6 \end{aligned}$$

   with $u = \frac{x}{X} \cdot M - \lfloor \frac{x}{X} \cdot M \rfloor$.

   $B_l(v)$ is similarly defined, with $v = \frac{y}{Y} \cdot N - \lfloor \frac{y}{Y} \cdot N \rfloor$.

According to our IFFD formulation $P = P^0 + \delta P$, we can re-write [Eq. 2.12] in terms of the IFFD parameters $\mathbf{\Theta} = \delta P$:

$$\begin{aligned} D(\mathbf{\Theta}; \mathbf{x}) &= \sum_{k=0}^{3}\sum_{l=0}^{3} B_k(u)B_l(v)(P^0_{i+k,j+l} + \delta P_{i+k,j+l}) \\ &= \sum_{k=0}^{3}\sum_{l=0}^{3} B_k(u)B_l(v)P^0_{i+k,j+l} + \sum_{k=0}^{3}\sum_{l=0}^{3} B_k(u)B_l(v)\delta P_{i+k,j+l} \end{aligned} \tag{2.13}$$

Based on the linear precision property of B-splines, a B-spline curve through collinear control points is itself linear, hence the initial regular configuration of control lattice P^0 generates

the un-deformed space, i.e., for any pixel $\mathbf{x}$ in the sampling domain, we have:

$$\mathbf{x} = \sum_{k=0}^{3}\sum_{l=0}^{3} B_k(u)B_l(v)P^0_{i+k,j+l} \tag{2.14}$$

where i, j are derived the same way as in [Eq. 2.12].

Now combining [Eq. 2.13] and [Eq. 2.14], we have:

$$\begin{aligned} D(\mathbf{\Theta};\mathbf{x}) &= \mathbf{x} + \delta D(\mathbf{\Theta};\mathbf{x}) \\ &= \mathbf{x} + \sum_{k=0}^{3}\sum_{l=0}^{3} B_k(u)B_l(v)\delta P_{i+k,j+l} \end{aligned} \tag{2.15}$$

Compared to the traditional FFD, the IFFD formulation above simplifies the integration of smoothness constraints, and accounts for an efficient multi-level approach (see section 4.2.4) to deal with both large and small non-rigid deformations.

2.5 Summary

In summary, this chapter presents the shape, appearance and deformation representations in our framework. In the next few chapters, we will introduce novel algorithms for model-based segmentation, registration and visual learning that use these representations to naturally integrate shape and appearance information to achieve more robust results.

Chapter 3

Metamorphs: Deformable Shape and Appearance Models

Integrating region statistics constraints into traditional shape-based deformable models has been a centerpiece of the efforts toward more robust, well-behaved models in boundary extraction and segmentation. Most of the work in this area consists of methods that loosely couple edge and region information either in *parametric* deformable models or in *geometric* deformable models. In this chapter, we propose a new class of deformable models, Metamorphs, which integrate model shape and interior appearance more naturally and address the limitations in previous integration efforts. The main features of Metamorphs are the implicit shape representation, the nonparametric approximation of model-interior intensity statistics, and the incremental free form Deformations (IFFD) introduced in Chapter 2. With these representations, a unified variational framework can be formulated to derive the Metamorphs model dynamics when the models are applied to segmentation. The framework consists of both edge and region energy terms and both types of terms are differentiable with respect to a common set of IFFD parameters. A Metamorphs model can be initialized far-away from the boundary and efficiently converge to an optimal solution. The main driving forces are an anisotropic balloon force derived from region-based energy terms, and edge-based forces derived from edge-based energy terms. During model deformation, the forces are updated dynamically and the deformations are constrained to guarantee consistent model-interior intensity statistics. The Metamorphs formulation also allows natural merging and competition of multiple models. To segment objects with large-scale textures, the texture representation with texon scale analysis introduced in chapter 2 is used, and forces are derived from nonparametric texture statistics to drive model deformations. We demonstrate the robustness of Metamorphs models using both natural and medical images that have high noise levels and complex textures.

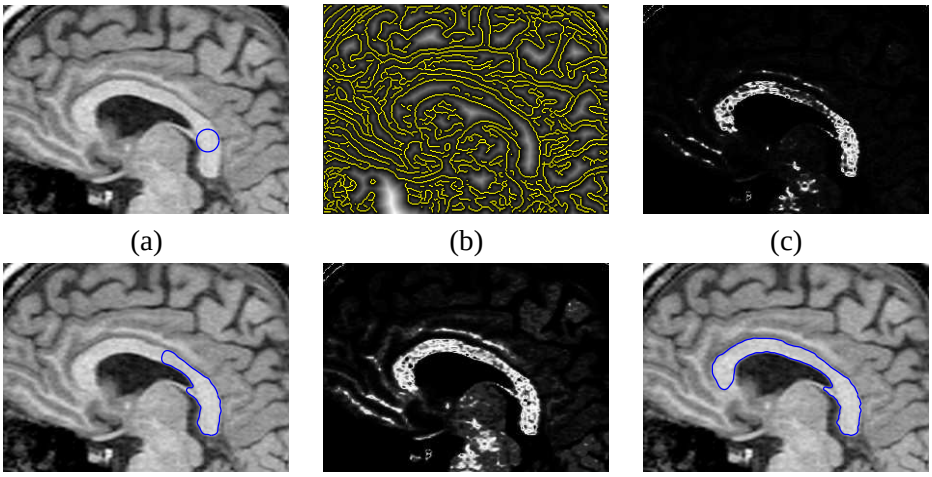


Figure 3.1: Metamorphs segmentation of brain structure. (a) A MRI image of the brain, with the initialized circular model drawn on top. (b) Edges detected using canny edge detector. (c) The intensity likelihood map computed according to the probability density function of the initial model interior. (d) Intermediate evolving model after 15 iterations. (e) The intensity likelihood map according to the intermediate model's interior statistics. (f) Final converged model after 38 iterations.

3.1 Introduction

Automated image segmentation is a fundamental problem in computer vision and medical image analysis. It remains difficult to solve the problem robustly however, due to the common presence of cluttered objects, object texture, image noise, variations in lighting, and various other artifacts in natural or medical images. To address these difficulties, deformable model-based methods have been extensively studied and widely used, with considerable success because of their ability to integrate high-level knowledge with low-level image processing.

Deformable models [62, 106, 21, 14, 70] are curves or surfaces that deform under the influence of internal smoothness and external image forces to delineate object boundary. Compared to local edge-based methods, deformable models have the advantage of estimating boundary with smooth curves or surfaces that bridge over boundary gaps. However, they may get stuck in local minima during evolution, when there is high image noise or spurious structures inside/around the object boundary.

In this chapter, we propose a new class of deformable models which integrate region intensity or texture information with boundary or edge information to achieve more robust segmentation. We term the new models "Metamorphs".

The basic framework of applying a Metamorphs model to boundary extraction is depicted in Fig. 3.1. The goal is to find the boundary of the corpus callosum brain structure on a MRI image of the brain. First, a simple-shape (e.g. circular) model is initialized centered around a seed patch inside the corpus callosum (see the blue circle in Fig. 3.1(a)). The model then deforms toward image edges as well as toward the boundary of a region that has similar intensity statistics as the model interior [1]. Fig. 3.1(b) shows the edges detected using a canny edge detector; note that the edge detector with automatically-determined thresholds gives result that have spurious edges and boundary gaps. To counter the effect of noise in edge detection, a region of interest (ROI) that has similar intensity statistics with the model interior is approximated. We first estimate the model-interior probability density function (p.d.f.) of intensity, then a likelihood map is computed which specifies the likelihood of a pixel's intensity according to the model-interior p.d.f. Fig. 3.1(c) shows the likelihood map computed based on the initial model interior; and we threshold the likelihood map to get the ROI. The evolution of the model is then derived using a gradient descent method from a unified variational framework that consists of energy terms defined on both edges and the ROI boundary. Fig. 3.1(d) shows the model after 15 iterations of deformation. As the model deforms, the model interior and its intensity statistics change, and the new model-interior statistics leads to the update of the likelihood map and the update of the ROI boundary for the model to deform toward. This online adaptive learning process empowers the model to find the boundary of objects with non-uniform appearances more robustly. Fig. 3.1(e) shows the updated likelihood map given the evolved model in Fig. 3.1(d). Finally, the model converges taking a balance between the edge and region influences, and the result is shown in Fig. 3.1(f).

The key property of Metamorphs is in that they naturally couple edge information with region statistics. By doing so, the new models generalize two major classes of deformable models in the literature: the *parametric* models and the *geometric* models, which are traditionally shape-based, and take into account only edge or image gradient information. In the remainder of the Introduction section, we will briefly review the *parametric* and *geometric* models, as well as previous efforts to incorporate region statistics into these models. We then discuss in more detail the novel aspects and contributions of Metamorphs.

[1] the model interior refers to the area in the image that is enclosed by the current deformable model.

3.1.1 Shape-based Deformable Models

Various snakes and deformable models proposed in the literature can be largely classified into two categories. The first class is the parametric (explicit) models that explicitly represent deformable curves and surfaces in their parametric form during the segmentation process. Examples are "Snakes" (or Active Contour Models) [62] and their extensions in both 2D and 3D [106, 21, 74, 72, 118]. The other class of deformable models is the geometric (implicit) deformable models [14, 70, 121, 15]. These models represent curves and surfaces implicitly as the level set of a higher-dimensional scalar function [103, 80], and the model evolution is based on the theory of curve evolution, with speed function specifically designed to incorporate image information. Comparing the two classes of deformable models, the parametric models have a compact representation and allow fast convergence, while the geometric models can handle naturally topological changes.

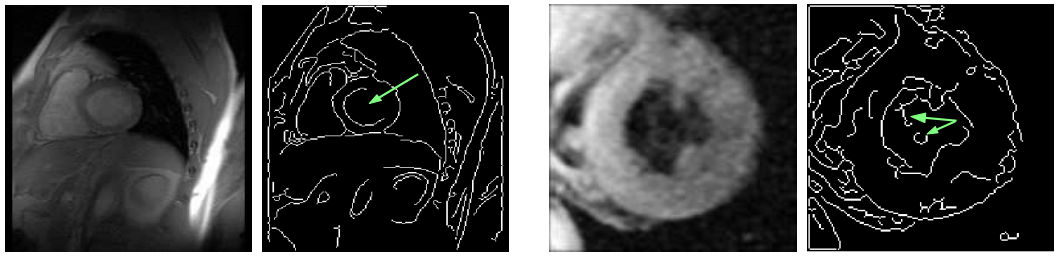
Although the parametric and geometric deformable models differ both in their formulations and in their implementations, both classes traditionally use primarily edge (or image gradient) information to derive external image forces to drive a shape-based model. In parametric models, a typical formulation [62] for the energy term deriving the external image forces is as follows:

$$E_{ext}(\mathcal{C}) = -\int_0^1 \left|\nabla \hat{I}\big(\mathcal{C}(s)\big)\right|^2 ds \tag{3.1}$$

Here $\mathcal{C}$ represents the parametric curve model parameterized by curve length s, $\hat{I} = G_\sigma * I$ is the image I after smoothing with a Gaussian kernel of standard deviation σ, and $\nabla \hat{I}(C)$ is the image gradient along the curve. Basically by minimizing this energy term, the accumulative image gradient along the curve is maximized, which means that the parametric model is attracted toward strong edges that correspond to pixels with local-maxima image gradient values.

In geometric models, a typical formulation [14] for the objective function that drives the front propagation of the level set function is:

$$E(\mathcal{C}) = \int_0^1 g\big(\left|\nabla \hat{I}(\mathcal{C}(s))\right|\big)\left|\mathcal{C}'(s)\right| ds, \quad \text{where} \quad g\big(\left|\nabla \hat{I}\right|\big) = \frac{1}{1+\left|\nabla \hat{I}\right|^2} \tag{3.2}$$



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

Figure 3.2: Potential problems in using a shape-only deformable model that moves under the influence of external forces derived from edge or image gradient information. The problematic areas are pointed by arrows.

Here $\mathcal{C}$ represents the front (i.e. zero level set) curve of the evolving level set function. To minimize the objective function, the front curve deforms along its normal direction $\mathcal{C}''(s)$, and its speed is controlled by the speed function $g(|\nabla \hat{I}|)$. Given the form of the speed function (Eq. (3.2)), one can see that, $g(|\nabla \hat{I}|)$ is defined based on image gradient $\nabla \hat{I}$, and it is positive in homogeneous areas and zero at ideal edges. Hence the curve evolves at a nearly constant speed across homogeneous regions and stops at strong edges.

The reliance on edge information in both types of traditional deformable models, however, makes the models sensitive to noise, spurious edges, and highly dependent on the model initialization. For instance, Fig. 3.2 shows several situations where deformable models deforming based on edges may have trouble converging to the optimal solutions. In Fig. 3.2(a-b), because of the weak image gradient along one side of the Left Ventricle, there is a large gap on the detected edges. Even though a parametric model can bridge over a small gap, it is likely to leak through a gap of this scale, and be attracted to the strong edges of the Epicardium instead. A geometric model is even less effective in countering the effect of boundary gaps, thus more likely to leak through the gap and converge to a wrong solution. In Fig. 3.2(c-d), there are some high image gradient areas and spurious edges detected inside the object of interest. In this case, a parametric model started small inside the object may get stuck in local minima and fails to reach the desired boundary; on the other hand, a topologically-free geometric model will keep evolving toward the boundary, yet the resulting front curve may contain small holes or islands, which correspond to the spurious structures instead of the true boundary.

3.1.2 Integrating Region Statistics Constraints

In order to address the limitations in shape-only deformable models, and develop more robust models for boundary extraction, there have been significant efforts to integrate region information into both parametric and geometric deformable models.

Along the line of parametric models, region analysis strategies have been proposed [90, 126, 59, 18] to augment the "snake" (active contour) models. In [90], a region-based energy criterion for active contours is introduced by including photometric energy terms that assume the local partition of the image into an object region and a background region. The optimization of the integrated energy function is mostly heuristic however, and it accounts for internal and external energies in separate steps. In [126], a generalized energy function that combines aspects of snakes/balloons and region growing is proposed and the minimization of the criterion is guaranteed to converge to a local minimum. Yet this formulation still does not address the problem of unifying shape and intensity, because it approximates the region intensity statistics using parameters of a Gaussian distribution, while the model shape is represented by a parametric spline curve. This large difference in representation prevented the use of gradient descent methods to update both region parameters and shape parameters in a unified optimization process. As a result, the two sets of parameters are not updated simultaneously in [126], rather they are estimated in separate steps and the energy function has to be minimized in an iterative way. In other hybrid segmentation frameworks such as those proposed by [18, 59], a region based segmentation module is used to get a rough binary mask of the object of interest. Then this rough estimation of the object can be used to initialize a deformable model, which will deform to fit edge features in the image using gradient information. In these frameworks, the region-based and edge-based modules are still separate energy minimization processes, so that the integration is still imperfect and errors from one module can hardly be corrected by the other.

Along the line of geometric models, the integration of region and edge information [122, 97, 112, 83] has been mostly based on solving reduced cases of the minimal partition problem in the Mumford and Shah model for segmentation [78]. In the Mumford-Shah model, an optimal piecewise smooth function is pursued to approximate an observed image, such that the function

varies smoothly within each region, and rapidly or discontinuously across the boundaries of different regions. The solution represents a partition of the image into several regions. A typical formulation of the framework is as follows:

$$F^{MS}(u, C) = \int_{\Omega} (u - u_0)^2 dxdy + a \int_{\Omega \setminus C} |\nabla u|^2 dxdy + b|C| \tag{3.3}$$

Here u_0 is the observed, possibly noisy image, and u is the pursued "optimal" piecewise smooth approximation of u_0. Ω represents the image domain, ∇u is the gradient of u, and C are the boundary curves that approximate the edges in u_0. One can see that the first term of the function minimizes the difference between u and u_0, the second term pursues the smoothness within each region (i.e. outside the set C), and the third term constraints the boundary curves C to be smooth and have the shortest distance. Although this framework nicely incorporates gradient and region criteria into a single energy function, no practical globally-optimal solution for the function is available, most notably because of the mathematical difficulties documented e.g. in [78]. In the recent few years, progress has been made and solutions for several reduced cases of the Mumford-Shah functional and their implementations have been proposed in the level set framework. One approach presented in [122] is able to segment images that consist of two or three types of regions, each characterizable by a given statistics such as the mean intensity and variance. The approach is implemented in a curve evolution framework and is able to cluster pixels in an image based on both geometric and statistical constraints. Nevertheless the algorithm requires known *a priori* the number of segments in the image and its performance depends upon the discriminating power of the chosen set of statistics (i.e. the means and variances). Another approach in [97] applies the multi-phase level set representation to piece-wise constant segmentation based on a reduced model of Mumford and Shah. It is considered as solving a classification problem because it assumes the mean intensities of classes are known *a priori*, and only the set of boundaries are unknown. In the works presented by [15, 112], piece-wise constant and piece-wise smooth approximations of the Mumford-Shah functional are derived for two-phase (i.e. two regions) [15] or multiphase (i.e. multiple regions) [112] cases in a variational level set framework. The optimization of the framework is based on an

iterative algorithm that approximates the region mean intensities and level-set shape in separate steps. The approach is still not a guaranteed global minimizer of the functional, hence its numerical results depend on the initialization of the model curves, and it may converge to a local minimum only. *Geodesic Active Region* [83] is another frame partition framework which integrates edge and region based modules. The algorithm consists of two stages: a modeling stage that constructs a likelihood map of edge pixels and approximates region/class statistics using mixture-of-Gaussian components, and a segmentation stage that uses level set techniques to solve for a set of smooth curves that are attracted to edge pixels and partition regions that have the expected properties of the associated classes. The first stage of the algorithm requires some off-line learning, and there is no online adaptive learning aspects in the method. From the above approaches, one can see that they all solve the frame partition problem, and assume piece-wise constant, piece-wise smooth, Gaussian, or Mixture-of-Gaussian intensity distributions within each partitioned region. However, the pursuit of a partition of the entire image may cause some difficulties when dealing with busy images that contain many objects and clutter. Their assumptions on the region intensity distributions also limit their effectiveness in finding boundaries of objects whose interiors have textured appearance, and/or complex multi-modal intensity distributions.

3.1.3 The Metamorphs Model

Having reviewed previous works on incorporating region constraints into shape-based deformable models, we propose a new class of deformable models, which we call “Metamorphs”. Metamorphs efficiently address several limitations in previous integration efforts, by combining the best features of *parametric* and *geometric* models, and introducing novel formulations that unify the representations for shape and intensity and derive the model deformations from both edge and region information in a unified variational framework. The shape of a Metamorphs model is implicitly embedded in a higher-dimensional space of distance transforms, thus represented by a distance map “image”. This implicit shape representation has been introduced in Chapter 2.2, and we use it in Metamorphs to specify model geometry so that no explicit shape parameters are needed. To capture the intensity or texture statistics of the model-interior region, Metamorphs use the nonparametric kernel-based density approximation, which is introduced

in Chapter 2.3. Unlike Gaussian or Mixture-of-Gaussian representations [126, 83], the nonparametric statistics does not have explicit parameters and it gets updated automatically as the model interior changes due to model deformation. The only set of parameters for Metamorphs are those specifying model deformations, which are parameterized by the cubic B-spline based Incremental Free Form Deformations (IFFD) [54]. The IFFD model is an extension of the Free Form Deformations (FFD) models [2, 38, 6], and it is introduced in Chapter 2.4.

In this chapter, we introduce the model dynamics of Metamorphs when they are used for boundary finding in images. We formulate both edge and region energy terms that are differentiable with respect to the common set of model-deformation parameters. The overall energy function is then optimized by a gradient-descent based method to deform the model toward object boundary. During model evolution, a Metamorphs model has the online-learning aspect which will constrain the model deformations such that the interior statistics of the model after each deformation is consistent with the statistics learned from the past history of the model interiors. The edge and region energy terms will have complementary effects and they will aid the model to overcome local minima due to small spurious edges inside the object, to prevent the model from leaking at boundary gaps, and to enable the segmentation of objects with intensity inhomogeneity and multi-modal interior statistics. An anisotropic balloon-force term is also conveniently defined according to region constraints, which leads to two-way forces that efficiently grow or shrink the model toward true object boundary. Furthermore, if multiple models are initialized, they are allowed to evolve simultaneously, and upon collision they can naturally either merge or compete based on whether their interior statistics are sufficiently close. In the case that the object to be segmented has large-scale texture, the nonparametric texture statistics representations introduced in Chapter 2.3.2 and 2.3.3 can be used to derive region-based forces to deform Metamorphs models toward the object boundary.

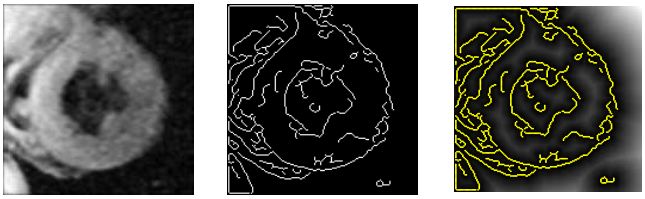
The Metamorphs framework has some similar components to works described by Region Competition (RC) [126], Geodesic Active Contours (GAC) [14], and level-set based methods [112, 83]. Our similarity to RC is in that both approaches do not solve a frame partition problem, rather they initialize the segmentation by putting multiple seed models across the image; and both have a Bayes energy term that aims to maximize intensity likelihood given current estimate of the region statistics. However, the differences between the two are numerous: our

work uses nonparametric statistics while RC uses Gaussian parameters; our work represents model shape implicitly which enables natural extension to higher dimensions and topology changes, while RC requires explicit parameterization of the model curves/surfaces; we are able to reduce two sets of parameters (one for shape, one for intensity) down to one (for model deformations) so that a unified gradient-descent based optimization scheme can be derived, while RC keeps the two sets of parameters separate, hence adapts an iterative greedy algorithm that finds a local minimum of its energy functional by updating the two parameters sets in alternating steps. The similar component between our work and GAC is in the design of the balloon forces. We use a similar multiplicative form as in Eq. (3.2) to incorporate the speed function that controls the speed of Metamorphs model evolution due to balloon forces. But our model representation and speed function definition are very different from that in GAC: instead of defining the speed function based on image gradient as in GAC, we derive the speed function from region constraints; instead of isotropic, nearly constant speed in homogeneous regions, our speed function is anisotropic and its value is proportional to a model point's distance from the region boundary. Compared to other level-set approaches [112, 83], although Metamorphs also use the implicit shape representation, the model evolution is more efficient and more robust to noise and spurious structures, because their deformations are parameterized by FFD rather than implemented in the curve evolution framework. The nonparametric intensity statistics in Metamorphs is also more generic than those using Gaussian or Mixture-of-Gaussian.

The remainder of this chapter is organized as follows. In section section 3.2, we introduce the energy functional and optimization schemes for Metamorphs when they are applied to boundary finding in images. In section 3.3, we present experimental results using both intensity and textured images. We conclude with discussions in section 3.4.

3.2 Boundary Finding with Metamorphs Model

In this section, we present the variational framework in which a Metamorphs model can be used to find the boundary of an object of interest. Given a model initialized inside the object of interest, we apply the scale analysis introduced in Chapter 2.3.2 to determine the scale of the



(a) (b) (c)

Figure 3.3: The effect of small spurious edges on the "shape image". (a) An MR image of the heart; the object of interest is the endocardium of the left ventricle. (b) The edge map of the image. (c) The derived "shape image" (un-signed distance transform of the edge map), with edge points drawn on top. Note the effect of the small spurious edges on the "shape image" inside the object.

model-interior texture, $\hat{s}$. If the scale is very small, we treat the image as an intensity image, and use the bottom level nonparametric intensity p.d.f. representation for model-interior region statistics (see chapter 2.3.1); in this case, the Metamorphs model dynamics are derived from both edge-based energy terms and region-based energy terms, whose formulations will be given in section 3.2.1. On the other hand, if the texon scale $\hat{s}$ is rather large (see Chapter 2.3.4), we predict that the model-interior texture consists of periodic mosaics of large-scale patterns; in this case, we formulate, in section 3.2.2, the energy terms that are derived from either the middle level (see chapter 2.3.2) or the top level (see chapter 2.3.3) region texture statistics.

3.2.1 Boundary Finding in Intensity Images

In intensity images, the motion of a Metamorphs model toward object boundary is driven by two types of energy terms derived from the image: the edge data terms E_E, and the region data terms E_R. So the overall energy functional E is defined by:

$$E = E_E + kE_R \tag{3.4}$$

where k is a constant balancing the contribution of the two types of terms. In our formulation, we are able to omit the model smoothness term in traditional parametric or level-set based deformable models, since this smoothness is implicit by using the Incremental Free Form Deformations. Next, we derive the edge and region data terms respectively.

The Edge Data Terms

A Metamorphs model is attracted to edge features with high image gradient values. We encode the edge information of an image using a "shape image", Φ, which is derived from the unsigned distance transform of the edge map of the image. The edge map is always computed using a standard Canny Edge Detector implementation with default parameter settings. In Fig. 3.3(c), we can see the "shape image" of an example MR heart image.

To evolve a Metamorphs model toward image edges, we define two edge data terms – an interior term E_{E_i} and a boundary term E_{E_b}:

$$E_E = E_{E_i} + aE_{E_b} \tag{3.5}$$

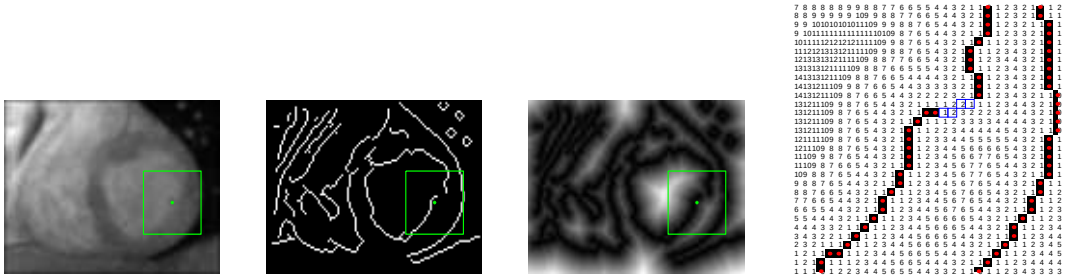
In the interior edge-based term, we aim to minimize the Sum-of-Squared-Differences between the implicit shape representation values in the model interior and the underlying "shape image" values at corresponding deformed positions. This can be written as:

$$E_{E_i} = \frac{1}{V(\mathcal{R}_\mathcal{M})} \iint_{\mathcal{R}_\mathcal{M}} \big(\Phi_\mathcal{M}(\mathbf{x}) - \Phi(D(\mathbf{\Theta};\mathbf{x}))\big)^2 d\mathbf{x} \tag{3.6}$$

In the above equation 3.6, the definitions for the following terms can be recalled from Chapter 2:

- $\Phi_\mathcal{M}$ refers to the implicit representation for model shape $\mathcal{M}$,
- $\mathcal{R}_\mathcal{M}$ refers to the model-interior region (i.e. the region that is enclosed by $\mathcal{M}$ in the image domain),
- $V(\mathcal{R}_\mathcal{M})$ is the volume of the model-interior region, and
- $D(\mathbf{\Theta};\mathbf{x})$ refers to the deformed position of the pixel § given IFFD control lattice configuration specified by the IFFD parameters $\mathbf{\Theta} = \delta P = \{(\delta P^x_{m,n}, \delta P^y_{m,n})\};\ (m,n) \in [1,M] \times [1,N]$.

During optimization, this term will deform the model along the gradient direction of the underlying "shape image". Thus it will expand or shrink the model accordingly, serving as a two-way balloon force implicitly and making the attraction range of the model large.



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

Figure 3.4: At a small gap in the edges, the boundary data term constraints the model to go along a path that coincides with the smooth shortest path connecting the two open ends of the gap. (a) Original Image. (b) The edge map, note the small gap inside the green square region. (c) The "shape image". (d) Zoom-in view of the region inside the green square. The numbers are the "shape image" values at each pixel location. The red dots are edge points, the small blue squares indicate a path favored by the boundary term for a Metamorphs model.

The previous interior term is good in attracting the model toward boundary structures from far away. However, when there are small spurious edges detected within an object, the "shape image" inside the object could differ from the model shape representation in the surrounding areas of those small edges. One such example can be seen in Fig. 3.3(a-c). To make the model deformation more robust to such situations, we consider a separate boundary term, which allows higher weights for pixels in a narrow band around the model boundary $\partial\mathcal{R}_{\mathcal{M}}$.

$$E_{E_b} = \frac{1}{V(\partial\mathcal{R}_{\mathcal{M}})} \iint_{\partial\mathcal{R}_{\mathcal{M}}} \big(\Phi(D(\boldsymbol{\Theta};\mathbf{x}))\big)^2 d\mathbf{x} \tag{3.7}$$

Intuitively, this term will encourage deformations that map the model boundary to image edge locations where the underlying "shape image" distance values are as small (or as close to zero) as possible. In the energy functional combining the interior and boundary edge data terms [Eq. (3.5)], by setting the value of the constant $a > 1$, those model boundary pixels get higher weights.

One additional advantage of the boundary term is that, at an edge with small gaps, this term will constrain the model to go along the "geodesic" path on the "shape image", which coincides with the smooth shortest path connecting the two open ends of a gap. This behavior can be seen from Fig. 3.4. Note that at a small gap of the edge map, the boundary term favors a path with the smallest accumulative distance values to the edge points.

The Region Data Terms

One of the most attractive aspects of the Metamorphs deformable models is that their interior intensity statistics are learned dynamically, and their deformations are influenced by forces derived from this dynamic region information. This region information is very important to help the models out of local minima, and converge to the true object boundaries. In Fig. 3.3, the spurious edges both inside and around the object boundary degrade the reliability of the "shape image" and the edge data terms. Yet the intensity probability map computed based on the interior intensity statistics of the model, as shown in Fig. 2.2(d), gives pretty clear indication on where the rough boundary of the object is. In another MR heart image shown in Fig. 3.5(1.a), a large portion of the object (Endocardium) boundary is missing during computation of the edge map using the default canny edge detector settings [Fig. 3.5(1.b)]. Relying solely on the "shape image" [Fig. 3.5(1.c)] and the edge data terms, a model would have leaked through the large gap and mistakenly converged to the outer epicardium boundary. In this situation, the intensity probability maps [Fig. 3.5(2-4.d)] computed based on the intensity statistics of the model-interior region become the key to optimal model convergence.

In our framework, we define two region data terms – a "Region Of Interest" (ROI) based balloon term E_{R_l} and a Maximum Likelihood term E_{R_m}, so the overall region-based energy function E_R is:

$$E_R = E_{R_l} + bE_{R_m} \tag{3.8}$$

We determine the "Region Of Interest" (ROI) as the largest possible region in the image that has a consistent intensity distribution as the model interior. The purpose of the ROI-based balloon term is to efficiently evolve the model toward the boundary (i.e. perimeter) of the ROI.

Given a model $\mathcal{M}$ on image I [Fig. 3.6(a)], we first compute the image intensity probability map P_I [Fig. 3.6(b)], based on the model interior intensity statistics (see Eq. 2.4 in section 2.3.1). Then a threshold (typically the mean probability over the entire image domain) is applied on P_I to produce a binary image BP_I. More specifically, those pixels that have probabilities higher than the threshold in P_I are given the value 1 in BP_I, and all other pixels are set to the value 0 in BP_I. We then apply binary image analysis on BP_I to extract the connected component that overlaps the model. Small holes in this connected component are filled using

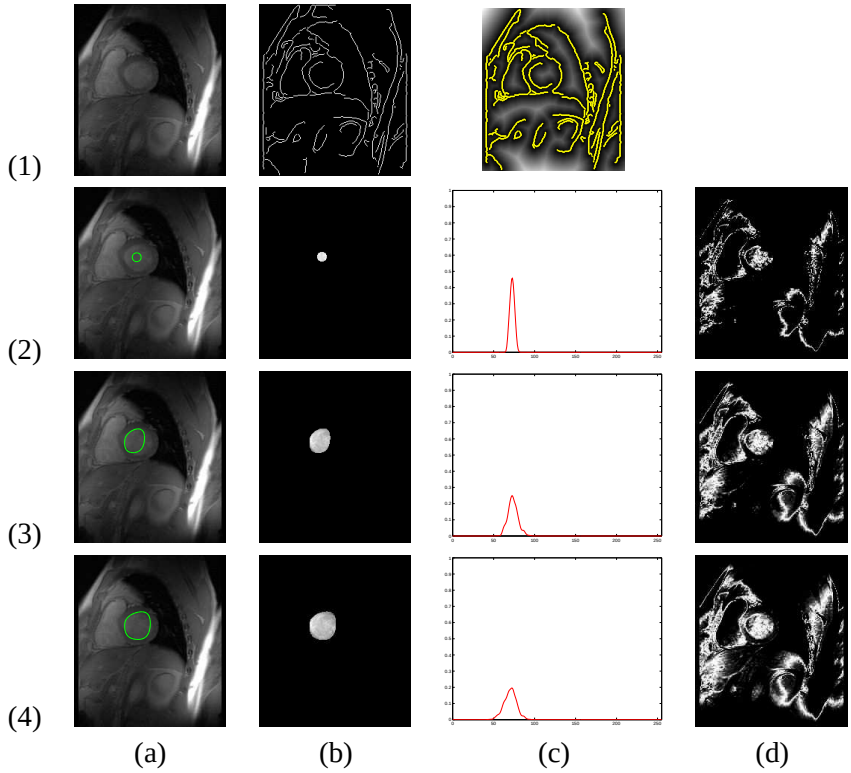


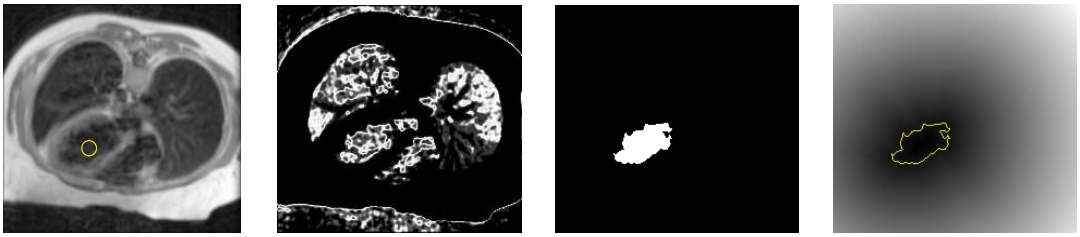
Figure 3.5: Segmentation of the Endocardium of the Left Ventricle in a MR image with a large portion of the object boundary edge missing. (1.a) the original image. (1.b) the edge map. (1.c) the "shape image". (2) initial model. (3) intermediate model. (4) converged model. (a) zero level set of the current model drawn on the image. (b) model interiors. (c) the interior intensity p.d.f.s. (d) intensity probability maps.

morphological operations, and finally we take this connected component as the current ROI. Suppose the binary mask of this ROI is BI_r [Fig. 3.6(c)], we encode its boundary information by computing the "shape image" of BI_r, which is the un-signed distance transform of the region boundary [Fig. 3.6(d)]. Denote this "shape image" as Φ_r, the ROI-based balloon term is defined as follows:

$$E_{R_l} = \frac{1}{V(\partial\mathcal{R}_\mathcal{M})} \iint_{\partial\mathcal{R}_\mathcal{M}} \Phi_r(\mathbf{x}) \big(\Phi_\mathcal{M}(D(\mathbf{\Theta}; \mathbf{x}))\big) d\mathbf{x} \tag{3.9}$$

where $\partial\mathcal{R}_\mathcal{M}$ refers to the model affinity (i.e. a narrow band around the zero level set of the model).

There are two components in the above ROI term, and they are combined multiplicatively. The key to understand the first component, $\Phi_\mathcal{M}(D(\mathbf{\Theta}; \mathbf{x}))$, is to take note that this model shape representation has negative values outside the model, zero value on the model, and positive



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

Figure 3.6: Deriving the ROI based region data term. (a) The model shown on the original image. (b) The intensity probability map computed based on the model interior statistics. (c) The ROI derived from the probability map after thresholding. (d) The "shape image" encoding boundary information of the ROI.

values inside the model (see Eq. 2.3). Hence by this component alone, the model would expand and grow like a balloon so as to minimize the value of the energy term. The second component in the energy term, Φ_r, is the ROI "shape image" and encodes the distance value of each pixel from the ROI region boundary. It serves as a weighting (or modulation) factor for the first component so that the speed of model evolution is proportional to the distance of the model from the ROI boundary. That is, the model moves fast when it is far away from the boundary and the underlying $\Phi_r(\mathbf{x})$ values are large in the model affinity; it slows down as it approaches the boundary, and stops at the boundary. This property of adaptively changing speed leads to improved model evolution behavior.

Within the overall energy minimization framework, this ROI-based balloon term is the most effective in countering the effect of small spurious edges inside the object of interest (e.g. in Fig. 3.3 and Fig. 3.12), since the ROI boundary is derived from region information alone. The adaptively changing balloon forces generated by the term also expedite model convergence and improve convergence accuracy, especially when the shape of the object is elongated, or has salient protrusions or concavities.

The previous ROI term is efficient to deform the model toward object boundary when the model is still far away. When the model gets close to the boundary, however, the ROI derived may become less reliable due to gradual intensity changes in the boundary areas. To achieve better convergence, we design another Maximum Likelihood (ML) region-based data term that constrains the model to deform toward areas where the pixel probabilities of belonging to the model interior intensity distribution are high. This ML term is formulated by maximizing the

log-likelihood of pixel intensities in a narrow band around the model after deformation:

$$\begin{aligned} E_{R_m} &= -\frac{1}{V(\partial\mathcal{R}_{\mathcal{M}})}\iint_{\partial\mathcal{R}_{\mathcal{M}}} log\mathbf{P}(I(D(\mathbf{\Theta};\mathbf{x}))\big|\Phi_{\mathcal{M}})d\mathbf{x} \\ &= -\frac{1}{V(\partial\mathcal{R}_{\mathcal{M}})}\iint_{\partial\mathcal{R}_{\mathcal{M}}}\Big[log\frac{1}{V(\mathcal{R}_{\mathcal{M}})} \\ &\quad +log\frac{1}{\sqrt{2\pi}\sigma}+log\iint_{\mathcal{R}_{\mathcal{M}}} e^{\frac{-(I(D(\mathbf{\Theta};\mathbf{x}))-I(\mathbf{y}))^2}{2\sigma^2}}d\mathbf{y}\Big]d\mathbf{x} \end{aligned} \tag{3.10}$$

During model evolution, when the model is still far away from the object boundary, this ML term generates very little force to influence the model deformation. When the model gets close to the boundary, however, the ML term generates significant forces to prevent the model from leaking through large gaps (e.g. in Fig. 3.5), and help the model to converge to the true object boundary.

Dynamic Evolution of the Model

In our formulations above, both edge data terms and region data terms are differentiable with respect to the model deformation IFFD parameters $\mathbf{\Theta}$, thus a unified gradient-descent based parameter updating scheme can be derived using both edge and region information. Based on the energy term definitions, one can derive the following evolution equation for each element $\mathbf{\Theta}_i$ in the deformation parameters $\mathbf{\Theta}$:

$$\frac{\partial E}{\partial\mathbf{\Theta}_i} = \big(\frac{\partial E_{E_i}}{\partial\mathbf{\Theta}_i}+a\frac{\partial E_{E_b}}{\partial\mathbf{\Theta}_i}\big)+k\big(\frac{\partial E_{R_l}}{\partial\mathbf{\Theta}_i}+b\frac{\partial E_{R_m}}{\partial\mathbf{\Theta}_i}\big) \tag{3.11}$$

- The motion due to the edge data terms are:

$$\frac{\partial E_{E_i}}{\partial\mathbf{\Theta}_i} = \frac{1}{V(\mathcal{R}_{\mathcal{M}})}\iint_{\mathcal{R}_{\mathcal{M}}} 2\big(\Phi_{\mathcal{M}}(\mathbf{x})-\Phi(D(\mathbf{\Theta};\mathbf{x}))\big)\cdot\big(-\nabla\Phi(D(\mathbf{\Theta};\mathbf{x}))\cdot\frac{\partial}{\partial\mathbf{\Theta}_i}D(\mathbf{\Theta};\mathbf{x})\big)d\mathbf{x}$$

$$\frac{\partial E_{E_b}}{\partial\mathbf{\Theta}_i} = \frac{1}{V(\partial\mathcal{R}_{\mathcal{M}})}\iint_{\partial\mathcal{R}_{\mathcal{M}}} 2\Phi(D(\mathbf{\Theta};\mathbf{x}))\cdot\big(\nabla\Phi(D(\mathbf{\Theta};\mathbf{x}))\cdot\frac{\partial}{\partial\mathbf{\Theta}_i}D(\mathbf{\Theta};\mathbf{x})\big)d\mathbf{x}$$

- And the motion due to the region data terms are:

$$\frac{\partial E_{R_l}}{\partial \mathbf{\Theta}_i} = \frac{1}{V(\partial \mathcal{R}_\mathcal{M})} \iint_{\partial \mathcal{R}_\mathcal{M}} \Phi_r(\mathbf{x}) \big(\nabla \Phi_\mathcal{M}(D(\mathbf{\Theta}; \mathbf{x})) \cdot \frac{\partial}{\partial \mathbf{\Theta}_i} D(\mathbf{\Theta}; \mathbf{x}) \big) d\mathbf{x}$$

$$\frac{\partial E_{R_m}}{\partial \mathbf{\Theta}_i} = -\frac{1}{V(\partial \mathcal{R}_\mathcal{M})} \iint_{\partial \mathcal{R}_\mathcal{M}} \Big[\big(\iint_{\mathcal{R}_\mathcal{M}} e^{\frac{-(I(D(\mathbf{\Theta};\mathbf{x})) - I(\mathbf{y}))^2}{2\sigma^2}} d\mathbf{y} \big)^{-1} \iint_{\mathcal{R}_\mathcal{M}} e^{\frac{-(I(D(\mathbf{\Theta};\mathbf{x})) - I(\mathbf{y}))^2}{2\sigma^2}} \cdot (-\frac{(I(D(\mathbf{\Theta};\mathbf{x})) - I(\mathbf{y}))}{\sigma^2} \cdot \big(\nabla I(D(\mathbf{\Theta};\mathbf{x})) \cdot \frac{\partial}{\partial \mathbf{\Theta}_i} D(\mathbf{\Theta};\mathbf{x}) \big)) d\mathbf{y} \Big] d\mathbf{x}$$

In the above formulas, the partial derivatives with respect to the IFFD deformation parameters, $\frac{\partial}{\partial \mathbf{\Theta}_i} D(\mathbf{\Theta}; \mathbf{x})$, can be easily derived from the deformation formula for $D(\mathbf{\Theta}; \mathbf{x})$ [Eq. (2.15)]. Details are given in the Appendix.

The Model Fitting Algorithm

Having defined the energy terms, the overall model fitting algorithm consists of the following steps:

1. Initialize the deformation parameters $\mathbf{\Theta}$ to be $\mathbf{\Theta}^0$, which indicates no deformation.
2. Compute $\frac{\partial E}{\partial \mathbf{\Theta}_i}$ for each element $\mathbf{\Theta}_i$ in the deformation parameters $\mathbf{\Theta}$.
3. Update the parameters $\mathbf{\Theta}'_i = \mathbf{\Theta}_i - \lambda \cdot \frac{\partial E}{\partial \mathbf{\Theta}_i}$. λ is the gradient descent step size.
4. Using the new parameters, compute the new model $\mathcal{M}' = D(\mathbf{\Theta}'; \mathcal{M})$.
5. Update the model. Let $\mathcal{M} = \mathcal{M}'$, re-compute the implicit shape representation $\Phi_\mathcal{M}$, and the new partitions of the image domain by the new model: $[\mathcal{R}_\mathcal{M}]$, $[\Omega - \mathcal{R}_\mathcal{M}]$, and $[\partial \mathcal{R}_\mathcal{M}]$. Also re-initialize a regular FFD control lattice to cover the new model, and update the ROI "shape image" ϕ_r based on the new model interior.
6. Repeat steps 1-5 until convergence.

In the algorithm, after each iteration, both model shape and model-interior intensity statistics get updated, and deformation parameters get re-initialized for the new model. This allows continuous, both large-scale and small-scale deformations for the model to converge to the energy minimum.

One important advantage of our framework is that, as the model evolves, the model interior changes, hence the model-interior intensity statistics get updated and the new statistics are used for further model evolution. This online learning property makes our model evolution framework a dynamic region growing process, which is more adaptable to segment objects with non-homogeneous interior intensities and more robust to noise and small islands inside an object. As examples, the evolving model-interior p.d.f.s and their corresponding image intensity likelihood maps can be seen from Fig. 2.2, Fig. 3.5, and Fig. 3.12.

3.2.2 Boundary Finding in Textured Images

Recall from Chapter 2.3 that there are three levels of nonparametric appearance representation in Metamorphs. The texon scale $\hat{s}$ is the key parameter to determine which level of appearance representation to use (see Chapter 2.3.4) when applying Metamorphs to a new image for segmentation. If the image's texon scale is small, we use the bottom level of representation, and the boundary finding framework has been described above in Section 3.2.1. In this section, we present the boundary finding framework for textured images whose natural texon scale is large. We consider either the middle level (see chapter 2.3.2) or the top level (see chapter 2.3.3) region texture statistics. In either case, a probability (or likelihood) map is computed (see Eq. 2.8 and Eq. 2.9), which represents the likelihood of the texon surrounding every pixel in the image having a consistent statistics with the model-interior. For instance, on the cheetah image example (see Fig. 3.7(a)) that we used in Chapter 2, the likelihood map computed using the top-level texon statistics is shown in Fig. 3.7(b)[2]. Let us denote this likelihood map $\mathbf{P}(T(\mathbf{x}, \hat{s})\big|\Phi_{\mathcal{M}})$ as L_I. We then improve the likelihood map by taking into account context information using Markov Random Fields (MRF) based belief propagation, and formulate the variational framework for Metamorphs-based boundary finding in textured images. We will describe the belief propagation step and the variational framework in detail below.

[2]This is the same as Fig. 2.5(d).

(a) (b) (c)

Figure 3.7: (a) the original cheetah image. Initial Model: blue circle; Texon scale: red circle, (b) likelihood map computed based on the top-level texon statistics, (c) updated likelihood map after applying BP based MRF.

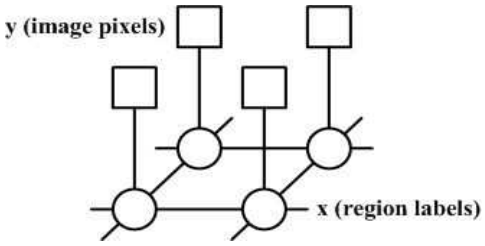


Figure 3.8: The MRF Graphical Model.

Contextual Confirmation through Belief Propagation

The likelihood map L_I, computed by Eq. 2.8 using the middle level texon statistics or by Eq. 2.9 using the top level statistics, quantifies the probability of every local texon belonging to part of the texture region of interest. However, all measurements are still local, and no context information between neighboring texons is accounted for. Markov Random Field (MRF) models are often used to capture dependencies between neighboring cliques (e.g. pixels, texons, etc.), and can be applied on the likelihood map to reduce noise and improve neighborhood consistency.

Given a typical graphical-model illustration for MRF, as shown in Fig. 3.8, the graph has two kinds of nodes: hidden nodes (circles in Fig. 3.8, representing region labels) and observable nodes (squares in Fig. 3.8, representing image pixels). Edges in the graph depict relationships between the nodes.

Let n be the number of the hidden/observable states (i.e., the number of pixels in the image). A configuration of the hidden layer is:

$$\mathbf{h} = (h_1, ..., h_n), h_i \in V, i = 1, ..., n \tag{3.12}$$

where V is a set of region labels, such as $V = 0, 1$, where the value 0 indicates different texture from the model interior, and the value 1 indicates same texture as the model interior.

Similarly, a configuration of the observable layer is:

$$\mathbf{o} = (o_1, ..., o_n), o_i \in D, i = 1, ..., n \tag{3.13}$$

where D is a set of pixel values, e.g., the original likelihood values in the map L_I. The relationship between the hidden states and the observable states (also known as local evidence) can be represented as the compatibility function:

$$\phi(h_i, o_i) = P(o_i|h_i) \tag{3.14}$$

Similarly, the relationship between the neighboring hidden states can be represented as the second compatibility function:

$$\psi(h_i, h_j) = P(h_i, h_j) \tag{3.15}$$

Now the inference problem can be viewed as a problem of estimating the MAP solution of the MRF model:

$$\mathbf{h}_{MAP} = argmax_{\mathbf{h}} P(\mathbf{h}|\mathbf{o}) \tag{3.16}$$

where

$$P(\mathbf{h}|\mathbf{o}) \propto P(\mathbf{o}|\mathbf{h})P(\mathbf{h}) \propto \prod_i \phi(h_i, o_i) \prod_{(i,j)} \psi(x_i, x_j) \tag{3.17}$$

The exact MAP inference in MRF models is computationally infeasible, and we use an approximation technique based on the Belief Propagation (BP) algorithm, which is an inference method proposed by [86] to efficiently estimate Bayesian beliefs in the network by iteratively passing messages between neighbors. We assume the likelihood values in each region follow a Gaussian distribution:

$$\phi(h_i, o_i) = \frac{1}{\sqrt{2\pi\sigma_{x_i}^2}} exp\big(- \frac{(o_i - \mu_{x_i})^2}{2\sigma_{x_i}^2}\big) \tag{3.18}$$

and the compatibility function between neighboring hidden states is represented by:

$$\psi(o_i, o_j) = \frac{1}{Z} exp\big(\frac{\delta(o_i - o_j)}{\sigma^2}\big) \tag{3.19}$$

where $\delta(x) = 1$ if $x = 0$; $\delta(x) = 0$ if $x \neq 0$, σ controls the degree of similarity between neighboring hidden states, and Z is a normalization constant.

After this step of MRF contextual confirmation, the resulting new likelihood map is denoted by L_I^c. One example demonstrating the effect of this step can be seen in Fig. 3.7(c). In our experiments, we use the $\{0, 1\}$ region labels as the hidden states, hence by thresholding at 0.5, we can differentiate regions that have similar texture with the model-interior from other background regions.

Deformable Model Dynamics

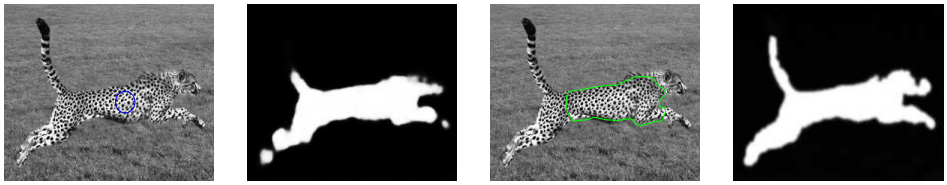
In order to evolve the deformable model toward the boundary of the texture region of interest, we derive the model dynamics in a variational framework by defining an efficient energy term that leads to both external texture/image forces and internal balloon forces.

Given the likelihood map L_I^c computed based on the current model-interior texture statistics, we define an energy term that produces forces to evolve the model toward the textured object boundary as follows:

$$E_{texture} = \frac{1}{V(\partial\mathcal{R}_\mathcal{M})} \iint_{\partial\mathcal{R}_\mathcal{M}} \mathcal{L}_I^c(\mathbf{x})\big(\Phi_\mathcal{M}(D(\mathbf{\Theta}; \mathbf{x}))\big) d\mathbf{x} \tag{3.20}$$

where $\Phi_\mathcal{M}$ is the implicit representation of the current model (Eq. 2.3), $\partial\mathcal{R}_\mathcal{M}$ refers to the model affinity (i.e. a narrow band around the zero level set of the model), $V(\partial\mathcal{R}_\mathcal{M})$ refers to the volume of model affinity region $\partial\mathcal{R}_\mathcal{M}$, and $D(\mathbf{\Theta}; \mathbf{x})$ is the IFFD definition for the position of a sample pixel $\mathbf{x}$ after deformation (Eq. 2.15).

Note that the energy term above (Eq. 3.20) has a similar form as the anisotropic region-based balloon force defined by Eq. 3.9. There are two components in the energy term and they are combined multiplicatively. The component $\Phi_\mathcal{M}(D(\mathbf{\Theta}; \mathbf{x}))$ makes the model expand and grow along its normal direction like a balloon, and the speed of expansion is weighted



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

Figure 3.9: (a) Initial model. (b) Likelihood map (after MRF) based on initial model. (c) An intermediate model. (d) Likelihood map re-computed based on the intermediate model.

(modulated) by the other likelihood component $\mathcal{L}_I^c(\mathbf{x})$. Hence the model grows toward object boundary with anisotropic speed modulated by the underlying likelihood map value, and it stops at object boundary where the likelihood values in $\mathcal{L}_I^c$ decrease to zero. The energy term generates forces that lead the model to efficiently converge to textured object boundary, even when the object shape has salient protrusions and concavities.

The energy term defined in Eq. 3.20 is differentiable with respect to the model deformation parameters $\mathbf{\Theta}$, hence a unified gradient-descent based parameter updating scheme can be derived:

$$\frac{\partial E_{texture}}{\partial \mathbf{\Theta}_i} = \frac{1}{V(\partial \mathcal{R}_{\mathcal{M}})} \iint_{\partial \mathcal{R}_{\mathcal{M}}} \mathcal{L}_I^c(\mathbf{x}) \big(\nabla \Phi_{\mathcal{M}}(D(\mathbf{\Theta}; \mathbf{x})) \cdot \frac{\partial}{\partial \mathbf{\Theta}_i} D(\mathbf{\Theta}; \mathbf{x}) \big) d\mathbf{x} \tag{3.21}$$

In the above formula, the partial derivatives with respect to the deformation (FFD) parameters, $\frac{\partial}{\partial \mathbf{\Theta}_i} D(\mathbf{\Theta}; \mathbf{x})$, can be easily derived from the model deformation formula Eq. 2.15, and the details are given in the Appendix.

One important advantage of this model-based texture segmentation framework is that, as the model evolves, the model interior changes, hence the model-interior texture statistics get updated and the new statistics are used for further model evolution. This online learning property enables our deformable model framework to segment objects with non-uniform texture patterns to some extent. In Fig. 3.9, we show the evolution in the likelihood map as the model evolves from an initial circular model to an intermediate model.

3.2.3 Multiple Model Initialization and Merging

When multiple models are initialized in an image, each model evolves based on its own dynamics. To allow merging and competition of the multiple models, a collision detection step is

applied after every few iterations to check whether the interiors of more than one models overlap. Although collision detection requires complicated algorithms in parametric deformable models [46], it is straightforward in Metamorphs because of the implicit model shape representation. Suppose the implicit representations for two models being tested are: $\Phi_{\mathcal{M}_a}(\mathbf{x})$ and $\Phi_{\mathcal{M}_b}(\mathbf{x})$. According to the definition of implicit shape representation (Eq. 2.3), $\Phi_{\mathcal{M}_a}$ and $\Phi_{\mathcal{M}_b}$ have positive values for pixels inside the model, negative values outside, and zero on the model. So to detect collision, we test every pixel $\mathbf{x}$ that has positive value in $\Phi_{\mathcal{M}_a}$. If for any such pixel $\mathbf{x}$ ($\Phi_{\mathcal{M}_a}(\mathbf{x}) > 0$), $\Phi_{\mathcal{M}_b}(\mathbf{x})$ is also positive, then a collision is detected, because $\mathbf{x}$ is inside both model $\mathcal{M}_a$ and model $\mathcal{M}_b$. Upon completion of each collision detection step, all models that collide are checked to see whether their interior intensity statistics are close; the colliding models are merged only if their statistics are sufficiently close.

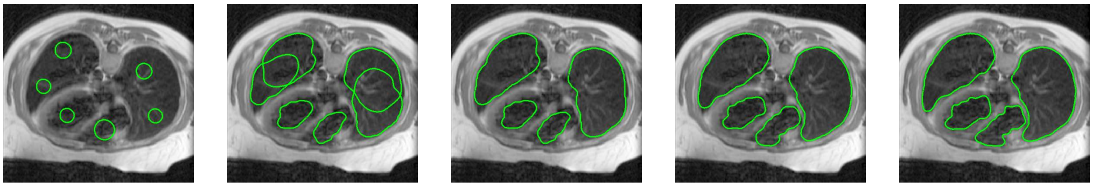
Suppose a collision is detected between model A and model B. Since the model interior appearances are represented using nonparametric p.d.f.s, the Kullback-Leibler Divergence can be used to measure the dissimilarity between two p.d.f.s. Suppose the intensity p.d.f. for model A is p_A and the p.d.f for model B is p_B, then the Kullback-Leibler Divergence between the two distributions is defined by:

$$D_{p_A \| p_B} = \int_U p_A(i) log \frac{p_A(i)}{p_B(i)} di \tag{3.22}$$

where U denotes the set of all intensity values. If this K-L distance is sufficiently small, the algorithm decides the statistics of the two models are sufficiently close, and the two models will be merged; otherwise, the two models will keep evolving on their own.

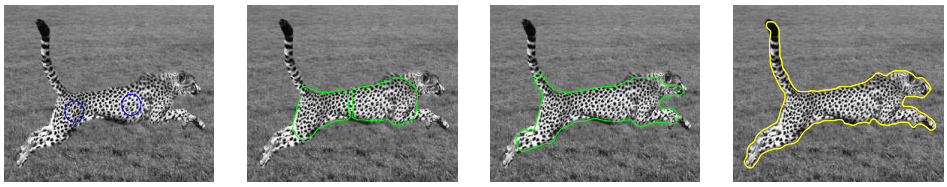
If two models in collision are to be merged, the new model's implicit representation can be easily derived from the representations of the two models before merging. Suppose the implicit representations for the two models to be merged are: $\Phi_{\mathcal{M}_a}(\mathbf{x})$ and $\Phi_{\mathcal{M}_b}(\mathbf{x})$. Then the implicit representation for the merged model will simply be: $\Phi_{\mathcal{M}}(\mathbf{x}) = \max\big(\Phi_{\mathcal{M}_a}(\mathbf{x}), \Phi_{\mathcal{M}_b}(\mathbf{x})\big)$. Thereafter this new model's interior statistics are updated and it evolves in place of the two old models.

Fig. 3.10 shows an image of the chest where we initialize multiple models inside the objects of interest including the left and right lungs, and the left and right ventricles. The models



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

Figure 3.10: (a) Multiple initialized models. (b) Result after the models evolve on their own for 5 iterations. (c) Collision detection, and merging after passing the statistics tests. (d) Result after 5 more iterations. (e) Converged models after 16 iterations.



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

Figure 3.11: (a) Two initial models. (b) Two models evolving on their own before merging. (c) The two models are merged into one new model upon collision and the new model continues evolving. (d) The final converged model.

first evolve on their own, and if any two models collide, they merge into one new model if their interior intensity statistics are sufficiently close. The converged models are shown in Fig. 3.10(d) to demonstrate the segmentation result.

Another model topology change example on the cheetah texture image is shown in Fig. 3.11, where two models are initialized and they first evolve on their own, then merge into one new model upon collision.

3.3 Experimental Results

In this section, we demonstrate segmentation results on both intensity and texture images using the Metamorphs framework.

3.3.1 Boundary Finding in Intensity Images

Some boundary finding examples on intensity images using Metamorphs have been shown in Fig. 2.2, Fig. 3.5, and Fig. 3.10. In Fig. 3.12, we show another example in which we segment the left ventricle of the heart in a noisy tagged MRI image. We use the intensity image

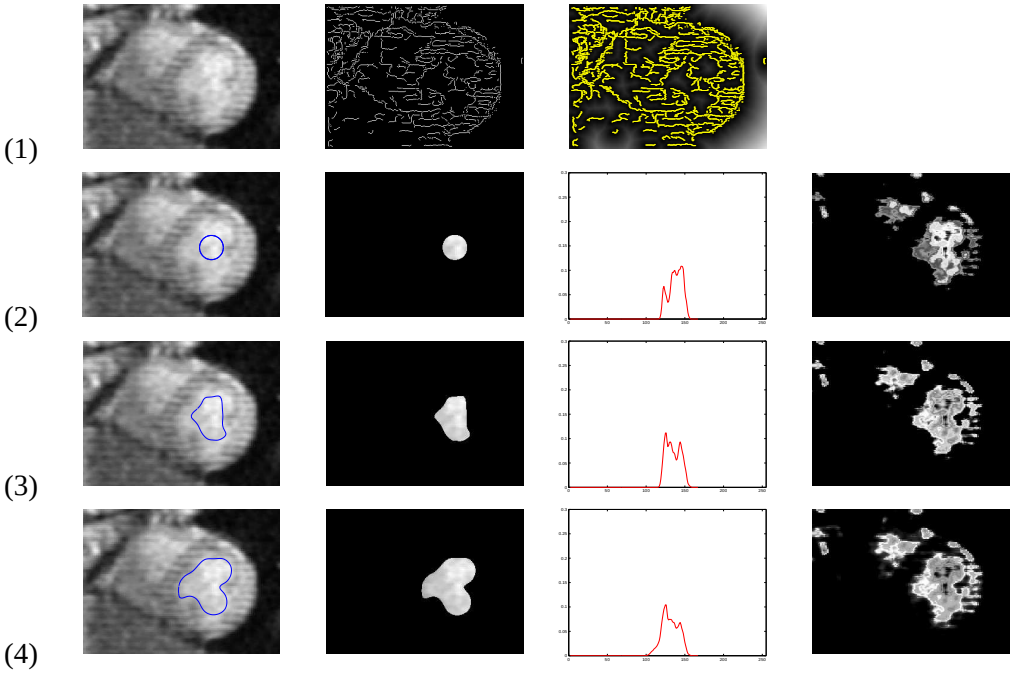


Figure 3.12: Tagged MR heart image example. (1.a) Original image. (1.b) Edge map. (1.c) "shape image" derived from the edge map. (2) Initial model. (3) Intermediate result. (4) Converged model (after 12 iterations). (2-4)(a) The evolving model. (2-4)(b) Model interior. (2-4)(c) Model interior intensity p.d.f. (2-4)(d) Intensity probability map according to the p.d.f. in (c).

framework for this image because the interior of the object of interest, left ventricle, is non-textured, although the tagging lines have obvious texture patterns. Note that, due to the tagging lines and intensity inhomogeneity, the detected edges of the object are fragmented, and there are spurious edges inside the region. In this case, the integration of edge and region information was critical in helping the model out of local minima.

We also apply our algorithm to ultrasound breast images to test its ability to deal with objects whose interior intensity distribution has multiple modes, or has high noise and speckle patterns. Fig. 3.13 shows two such examples, and the goal is to use Metamorphs models to find boundaries of the breast lesions. Because of the nature of the ultrasound images, there is no clear contrast edges that separate a lesion from its surrounding normal tissue. The criterion in spotting and locating lesions is usually that the lesion areas are denser hence appear darker than its surroundings. One can see from Fig. 3.13(c) that the nonparametric kernel-based method can represent pretty well the differences in the intensity statistics of ultrasound speckle patterns, hence providing important information about where the lesion boundaries are.

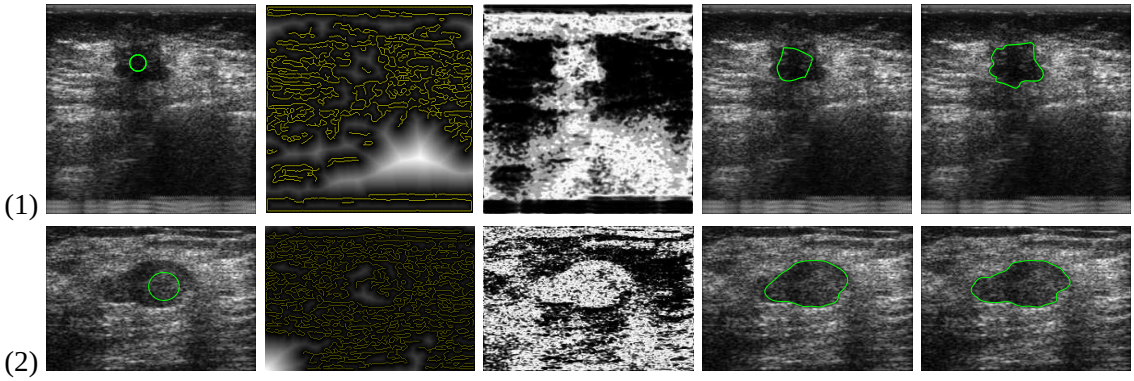


Figure 3.13: Segmenting lesions in ultrasound breast images. (a) The original ultrasound image, with the initial model drawn on top, (b) The shape image derived from the edge map, (c) Intensity likelihood map, (d) Intermediate model after 4 iterations for example (1), and 13 iterations for example (2), (e) Final converged model after 11 iterations for (1) , and 20 iterations for (2).

Figure 3.14: Boundary finding in the pepper image. (a) Original image, with initial models drawn on top. (b) The shape image derived from the edge map. (c) Intermediate result showing the models after 10 iterations. (d) Final converged models after 14 iterations. (e) The three pepper segments enclosed by the three converged models.

Other than the medical images, we tested our algorithm on natural images in which occlusion, specularity, shadow, reflection, and other conditions are common. Fig. 3.14 shows the segmentation result using a pepper image. Several circular models are initialized, and their interiors capture the intensity variations on the three foreground peppers due to lighting, shadow and specularity. The models evolve, merge and finally converge to the result shown in Fig. 3.14(d-e). During model evolution, the two models initialized on the elongated pepper quickly merge after a few iterations because their interior intensities are very close, while the two models initialized on the dark-colored pepper do not merge until the top model evolves and includes part of the specular region (see Fig. 3.14(c)) because only then the two model interiors have sufficiently close statistics (Eq. 3.22).

Fig. 3.15 demonstrates the experiment on a picture of people. Several circular models are

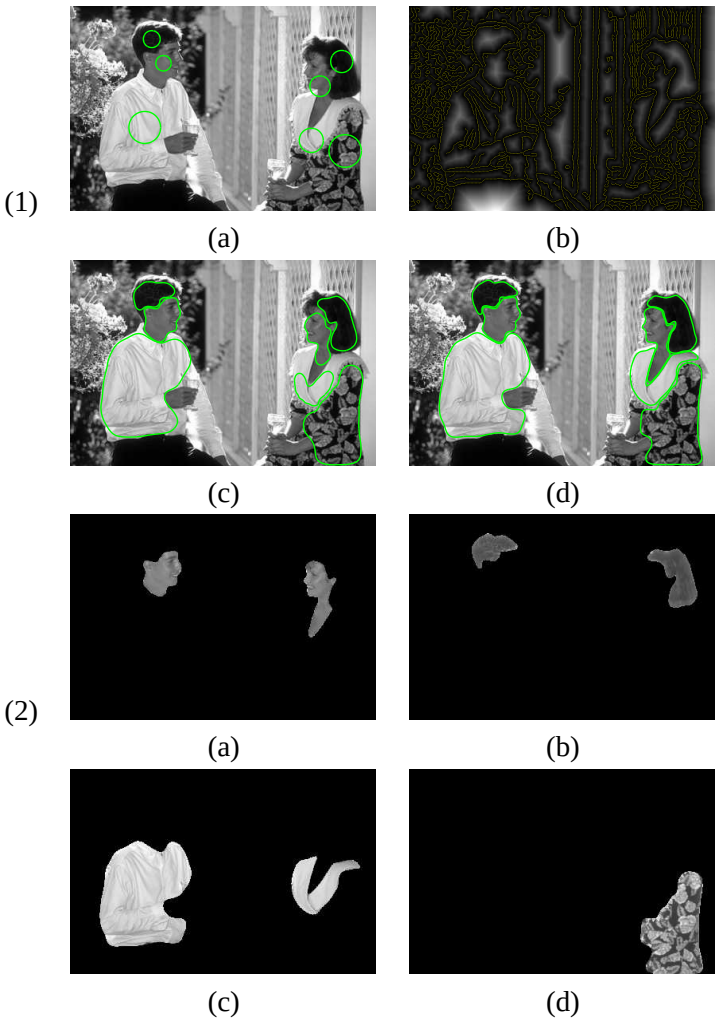


Figure 3.15: Boundary finding in a picture of people. (1) the evolution of models; (2) finally segmented patches. (1.a) Original image, with initial models drawn on top. (1.b) The shape image derived from the edge map, with edge points drawn on top. (1.c) Intermediate result showing the models after 8 iterations. (1.d) Final converged models after 22 iterations. (2.a) Skin color patches that correspond to faces. (2.b) Patches that correspond to hair. (2.c) Patches that correspond to white shirt. (2.d) Patches that correspond to the women's textured dress.

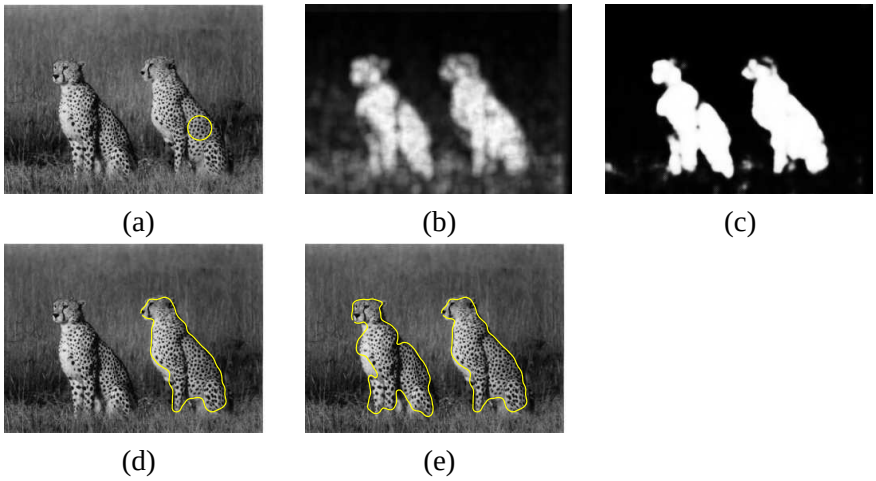
initialized on the face, hair, clothes of the two people. The converged models are shown in Fig. 3.15(1.d). On the faces, small interior structures such as the eyes and eyebrows did not stop the models from converging to the face boundaries. The texture of the women's dress consists of large-scale patterns and multiple colors; by initializing a model whose interior captures the color changes within the texture, the model accurately converges to the texture boundary without getting stuck at interior edges produced by the changing colors. The model on the man's white shirt stopped before including the left arm because of the strong appearance and edge boundary generated by the glass and bright sunlight. In the second row, Fig. 3.15(2) shows the segments enclosed by the converged models. In Fig. 3.15(2.a-c), the segments that are displayed together have similar intensity statistics according to the Kullback-Leibler Divergence criterion in Eq. 3.22.

3.3.2 Boundary Finding of Textured objects in Texture Images

One texture segmentation result using the cheetah image first shown in Fig. 2.3 can be seen in Fig. 3.11. We also run our algorithm on a variety of other images with textures of different patterns and scales. Figures 3.16-3.17 show typical segmentation results. In all the cases, we initialize several seed points inside the textured regions of interest, then a texture-consistency likelihood map is computed based on each model interior, the models evolve on their own dynamics, and those models with similar texture statistics are allowed to merge upon collision. The likelihood map for each model is re-computed after every 5 iterations of model evolution since the model interior statistics change as the model deforms.

Fig. 3.16 is an experiment run on an image containing two cheetahs. The likelihood maps computed based on the initial model are shown, and the converged model finds the boundary for one of the cheetahs. By initializing another model in another high-likelihood area, we are able to get the boundary for the other cheetah (Fig. 3.16(e)).

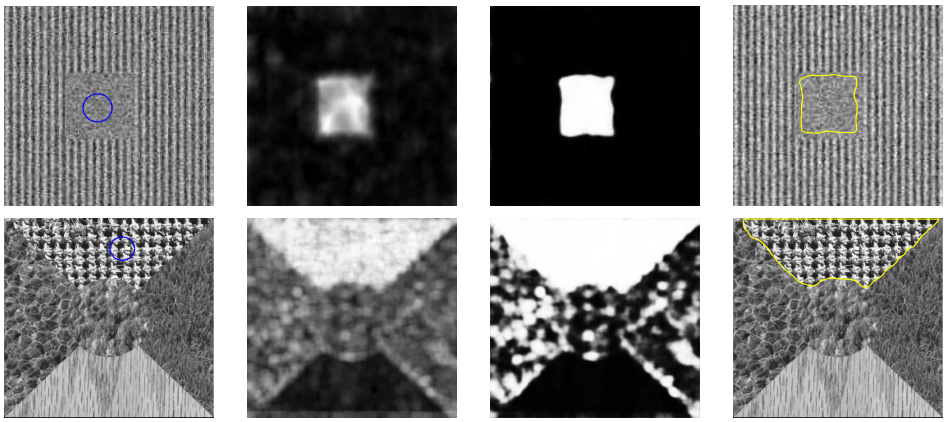
In Fig. 3.17, we demonstrate our algorithm using two synthetic images. The image on the top row has a small-scale homogeneous region in the center, and large-scale periodic line patterns in the background. The line pattern is generated using a sinusoidal signal. To test the robustness of the method to noise, we randomly added high level of Gaussian noise to the entire image. The segmentation result shows that our method can deal with both small-scale



(a) (b) (c)

(d) (e)

Figure 3.16: (a) Original image with initial model. (b) Likelihood map based on Gabor response statistics. (c) Likelihood map after Belief Propagation. (d) The converged model. (e) Both cheetah boundaries detected after initializing another model in the other high-likelihood area.



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

Figure 3.17: (a) Original images. (b) Likelihood maps based on model-interior texture statistics. (c) Likelihood maps after BP. (d) The converged models at texture boundary.

and large-scale texture patterns, and has good differentiation power even in the presence of high noise levels. On the bottom row, we show the performance on a synthetic texture mosaic image. The image consists of five texture regions of similar intensity distribution, and we demonstrate the likelihood map and segmentation of one of the regions. We are able to segment the other four regions in the mosaic using the same method.

3.3.3 Performance and Parameters

The Metamorphs model evolution is computationally efficient. For all the examples shown, the segmentation process takes less than $200ms$ to converge on a 2GHz PC station. Several reasons contribute to this. First, the IFFD parameterization of model deformations significantly reduces the number of local deformation parameters, while guaranteeing the model's smoothness properties. Second, most computation only involves pixels that are either within a narrow band surrounding the model or inside the model. The only whole-image computation, which is the intensity probability map, is done efficiently using the Fast Gauss Transform [35] in linear time.

The resolution of the IFFD control lattice used to cover the model, $M \times N$ in Eq. 2.10, is initially set to be 10×10 for all examples, and it is dynamically adjusted during model evolution to allow for both global and highly local deformations. More specifically, we increase the resolution along both x and y directions by one after each iteration. Based on the properties of IFFD, the higher resolution is the IFFD control lattice, the higher curvature the model can capture. By increasing the resolution dynamically, the model can quickly evolve toward the boundary with global smoothness constraints in the beginning, then once it is near the boundary, the increasing resolution of the lattice enables it to fit into the detailed convexities and concavities of the object boundary. The capability of the Metamorphs model to capture high curvature features on boundaries is demonstrated through Fig. 2.2, Fig. 3.10 (e.g. tips of lung), and Fig. 3.15 (e.g. ears, clothing-generated corners).

The three weight factors that balance the contributions from different energy terms, k, a and b in Eq. (3.11), are estimated automatically for each Metamorphs model based on its surrounding image information. The weighting factor between the edge terms and the region terms, k, is determined by a confidence measure, C_e, of the computed edge map. To decide this confidence value, we compute the "region of interest" (see section 3.2.1) after initializing a model, then C_e is determined by the complexity of image gradient or edge map within this ROI. The confidence value is low if there are high gradients and edges inside the region; the value is high otherwise. Then we set the value for the weighting factor $k = \frac{1}{C_e}$. The other two weighting factors a and b are set to be: $a > 1, b > 1$. This is because we always assign

higher weights to data terms that make the model converge when it is near the boundary, i.e. the boundary term E_{E_b} and the Maximum Likelihood term E_{R_m}. In all our experiments, this automatic weight-factor estimation scheme gives good and stable performance.

For different examples, there are only two parameters that need to be adjusted by human intervention. The first one is the gradient descent step size λ (see section 3.2.1). The reason to allow users to manually adjust λ is to avoid too big step sizes that blow up the model after the first iteration, as well as to avoid too small step sizes that make model convergence too slow. For the examples shown, typical values for λ are between 20 and 40. The second parameter that can be adjusted by users is the threshold δ for determining model convergence. The model convergence threshold δ is measured in terms of the l^2-norm of the IFFD parameter vector $|\mathbf{\Theta}|$. Recall from Section 2.4.1 that $\mathbf{\Theta}$ represents the displacements of IFFD control points, then $|\mathbf{\Theta}| = \left|[\delta P^x_{1,1}\delta P^y_{1,1}...\delta P^x_{m,n}\delta P^y_{m,n}...\delta P^x_{M,N}\delta P^y_{M,N}]^{\mathrm{T}}\right|$ measures the magnitude of the overall control lattice displacement. That is, if the derived movement of IFFD control points (Eq. 3.11) between two consecutive iterations is very small (smaller than the threshold), we consider the model has converged. The reason to manually adjust the model convergence threshold is to allow the user to control the model when occasionally it is necessary to make a compromise between preserving sharp corners on the boundary and avoiding model leakage into another neighboring object with similar intensity statistics. Setting the threshold to a very small value will enable the model to gradually fit into high-curvature corners; and setting the threshold to a reasonably small value can generally prevent the model from leaking through boundary gaps that survive both in edges (as gaps) and in regions (as narrow bridges). For the examples shown, typical values for δ are between 0.2 and 1.0 pixels.

In the case that multiple models are initialized on an image, the same set of parameters are used for all models. In the beginning, all models are active, and during evolution, if any model converges, its status is changed to inactive. The algorithm runs until all models converge and are no longer active.

3.3.4 Performance Comparison with Other Boundary Finding Methods

In this section, we compare experimentally Metamorphs with some other boundary finding methods in the literature including several well-known snake (deformable) models and region-statistics based segmentation methods. The comparison results are demonstrated on two gray-level image examples (one chest MRI image and one breast ultrasound image) in Fig. 3.18 and Fig. 3.19.

Comparison with Other Snake Models

Compared to various snake (deformable) models in the literature, the main contribution of the Metamorphs model lies in its novel way of integrating edge and region information for robust boundary finding. The advantage of Metamorphs is most significant when models are initialized far away from the object boundary. Using circular model initializations in Fig. 3.18(1.a) and Fig. 3.18(2.a) for the chest MR and the breast ultrasound images respectively, we first attempt to recover the object boundary using a snake model with balloon forces [21]. [3] The external potential field for the balloon model is computed based on the gray-level edge strength map (see Fig. 3.18(1.b) and 3.18(2.b)). We manually adjusted the weight factor between the external force and the inflation balloon force, but found it hard to strike a balance between the two. Fig. 3.18(1.c) and 3.18(2.c) show the balloon snake model after 120 iterations for the MR image and after 150 iterations for the ultrasound image respectively. The balloon model is starting to fail at this time, because some part of the model has already surpassed the boundary at places that have relatively weak image-gradient magnitudes while some other parts still haven't reached the true boundary. Since a balloon model applies explicitly either inflation or deflation forces that only support one-way evolution, once the model mistakenly goes over the true boundary, it can not recover from the mistake.

Next, we test the performance of the Gradient Vector Flow (GVF) snakes [118] on the images. The GVF snakes were proposed to make snake models have better convergence behavior

[3] We do not show comparison results with the traditional snake model [62] because such a model will shrink to a point under its internal forces given the far-away initializations in our experiment.

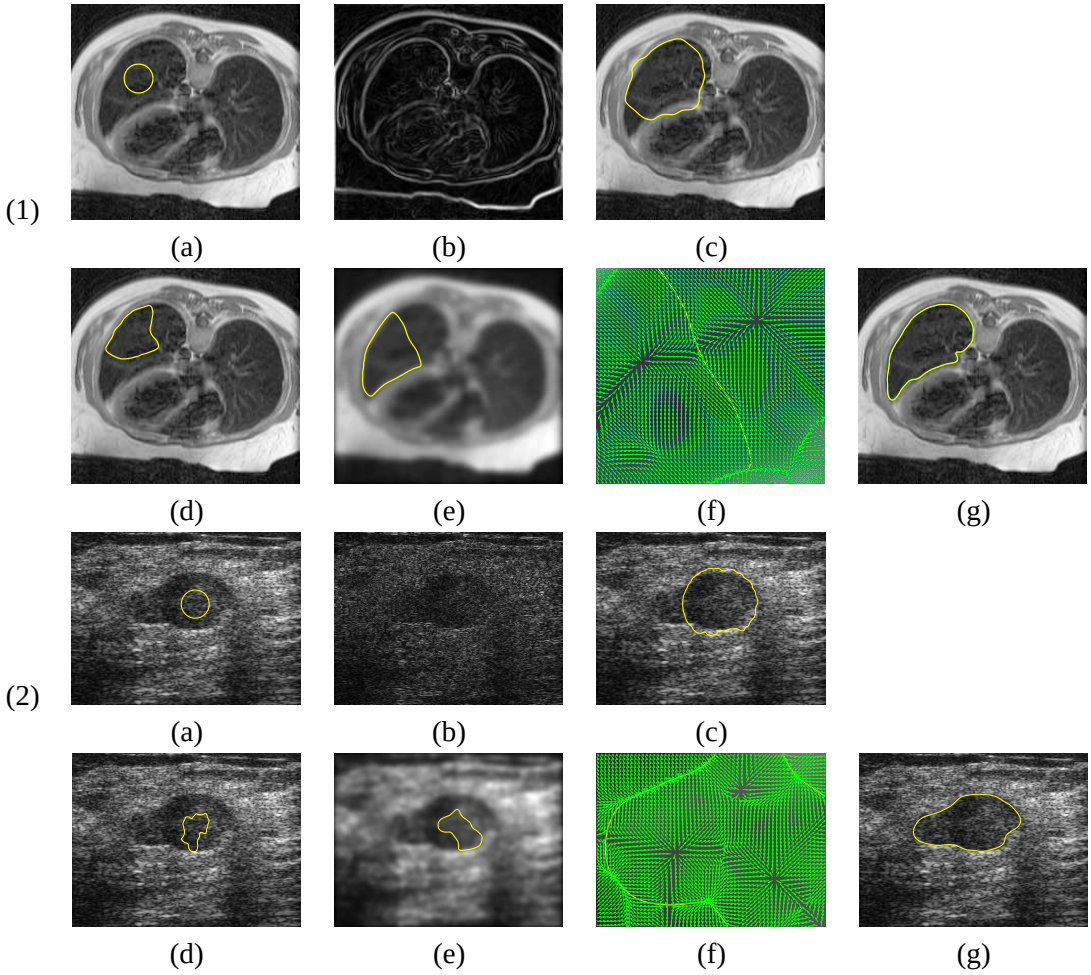


Figure 3.18: Comparison between various snake models in the literature and Metamorphs. (1) Results on a chest image. (2) Results on segmenting a breast lesion in an ultrasound image. (a) original image with initial model drawn on top. (b) gray-level edge map. (c) result using snake model with balloon forces. (d) result using GVF snake on original image. (e) result using GVF snake on smoothed image after applying Gaussian smoothing. (f) underlying GVF potential field that caused GVF snake local minima. (g) result using Metamorphs without image smoothing.

into concavities and be less sensitive to initialization. For the chest MR image, without smoothing the original image and when initialized as in Fig. 3.18(1.a), a GVF snake converged after 120 iterations and the result is shown in Fig. 3.18(1.d). Note that the GVF snake is not very robust to image noise and spurious edges and it got stuck in local minima. To reduce noise, we further applied Gaussian smoothing to the original image and run a GVF snake using the same initialization on the smoothed image. The converged GVF snake after 150 iterations is shown on the smoothed image in Fig. 3.18(1.e). This time the underlying GVF potential field that caused the local minima on the upper portion of the snake model is shown in Fig. 3.18(1.f).

We tried the same tasks on the breast ultrasound image, with a GVF model initialized as shown in Fig. 3.18(2.a). On the original image without smoothing, the GVF snake converged after 80 iterations and the result is shown in Fig. 3.18(2.d). After applying Gaussian Smoothing, we run a GVF snake on the smoothed image, and the converged model after 100 iterations is shown in Fig. 3.18(2.e). The model is stuck in local minima, and part of the underlying GVF potential field that contributes to the local minima can be seen in Fig. 3.18(2.f). The local minima problem with GVF is related to the essence of the GVF potential field, which is a laplacian diffusion of the gray-level edge map's gradient vectors. So when strong image gradients (edges) or small islands remain inside the object even after smoothing, the GVF snake gets attracted to them and gets stuck in local minima. This behavior is also typical of all other deformable models that rely on image gradient or edge information alone.

Finally, for comparison, we show the result from Metamorphs using the same initializations. The parameter setting for the Metamorphs model on the chest MR image is the same as that in Fig. 3.10, and the parameter setting on the breast ultrasound image is the same as that in Fig. 3.13(2). The models are run on the original images without smoothing. The Metamorphs model on the chest MR image reached convergence after 24 iterations (see Fig. 3.18(1.g)), and the model on the breast ultrasound image reached convergence after only 18 iterations (see Fig. 3.18(2.g)). From the results, one can see that while other snake models fail to segment accurately due to local variations in image gradient caused by image noise or object texture, the Metamorphs model is fast and robust in convergence because it naturally integrates both region statistics and image gradient information.

Comparison with Region-based Segmentation Methods

Popular region-based segmentation methods such as region growing, Markov Random Fields, and graph cuts have the advantage that they group pixels whose intensities follow consistent statistics, hence they are less sensitive to localized image noise. However they often generate irregular region boundaries, small holes inside regions of interest, and they may mistakenly link separate regions with similar statistics by a narrow bridge. Metamorphs has the advantages of region-based methods because of its nonparametric region statistics test and region-based energy terms. Meanwhile, Metamorphs generates smooth region boundaries, avoids small

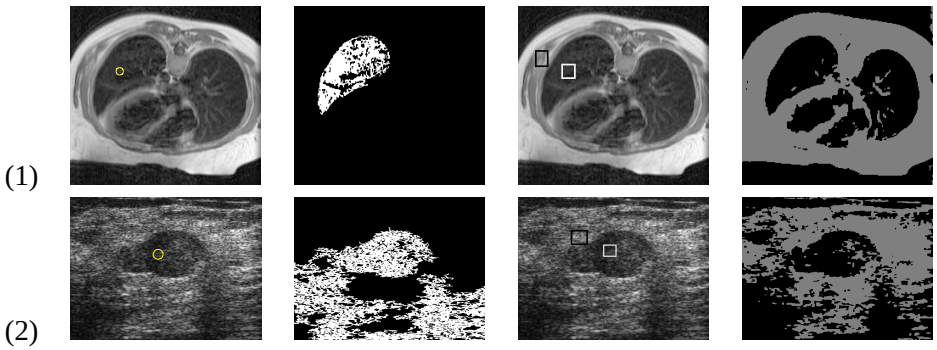


Figure 3.19: Comparing segmentation results from Region Growing (RG) and Markov Random Fields (MRF) with that from Metamorphs. (a) seed patches for RG are enclosed by yellow circles. (b) RG segmentation results. (c) two-class initialization for MRF: object class sample patches are enclosed by white rectangles, and background class sample patches are enclosed by black rectangles. (d) MRF segmentation results using the algorithm described in [8]. Object class is rendered in black, and background class is rendered in gray.

holes and narrow bridges, because of its internal smoothness constraints as a hybrid deformable model based approach.

In Fig. 3.19, we compare the segmentation results from region growing and Markov Random Fields with that from Metamorphs. First, the region growing (RG) algorithm we implement is based on nonparametric region statistics. Starting with a small seed patch, RG grows the region by gradually adding in surrounding connected pixels as long as the pixels follow the nonparametric region statistics [4]. When the growing stops, the result for the MR image is shown in Fig. 3.19(1.b) and the result for the ultrasound image is shown in Fig. 3.19(2.b). One can see that although for the MR image, RG finds roughly the object region, the region has many holes and the boundary is irregular. RG failed to find the lesion boundary on the ultrasound image due to ultrasound speckle patterns and leaking into the background. Second, the Markov Random Fields (MRF) implementation we use is based on the supervised Bayesian MRF image classification algorithm described by [8]. We specified the images consisting of two classes: the object class (the class sample patches are enclosed by white rectangles in Fig. 3.19(1.c) and Fig. 3.19(2.c)), and the background class (the class sample patches are enclosed by black rectangles in Fig. 3.19(1.c) and Fig. 3.19(2.c)). Given the class sample patches, the

[4] We consider a pixel follows the statistics if its intensity likelihood is greater than a threshold according to Eq. 2.4.

algorithm computes the intensity mean and variance for each class and applies MRF to improve classification. The MRF segmentation result after 266 iterations for the chest MR image is shown in Fig. 3.19(1.d), and the result after 346 iterations for the breast ultrasound image is shown in Fig. 3.19(2.d). One can see that the MRF segmentation is good for the MR image, although it still generates irregular boundary and small holes/islands. The MRF segmentation failed on the ultrasound image since it did not separate the lesion object from part of the background that has similar statistics and it generated small holes/islands inside the object. For comparison, the clean and smooth object boundaries found by our Metamorphs model-based method can be seen in Fig. 3.18(1.g) and Fig. 3.18(2.g).

3.4 Discussions

We have presented a new class of deformable models, Metamorphs, which possess both boundary shape and interior appearance statistics. We propose variational frameworks in which Metamorphs models can be applied to boundary finding in both intensity and texture images. During boundary finding, both edge and region information are coupled coherently to drive the deformation of the model toward object boundary.

The main contributions of the work lie in several aspects. First, the Metamorphs models represent a generalization of previous *parametric* and *geometric* deformable models, to take into account model-interior region information, while being computationally efficient. Second, the proposed framework does not require learning statistical shape and appearance models *a priori*, but the model deformations are constrained such that interior statistics of the model after deformation are consistent with the statistics learned adaptively from the past history of the model interiors. Third, compared to other works that integrate edge and region information for segmentation, our framework is more natural in that it does not have separate parameters to represent model shape and model-interior appearance statistics. The only set of parameters in our framework is the IFFD parameters that specify model deformation, and when the model deforms, its implicit shape representation and its interior nonparametric intensity or texture statistics get updated automatically. When used for boundary finding, the Metamorphs dynamics can be derived from edge and region energy terms that are both differentiable with respect to

the IFFD deformation parameters in a common variational framework. Lastly, our algorithm supports model topology changes and multiple models can merge upon collision if their interior statistics are sufficiently close.

The Metamorphs framework can also be extended to work on color images and 3D images. For color images, edges can be computed using color edge detectors (e.g., [96]); to derive the region terms, model-interior intensity statistics can be considered in three independent channels (Red, Green and Blue), and the overall color likelihood map can be computed by taking the multiplication of likelihood maps from the three channels. In 3D, the implicit model shape representation and kernel-based intensity p.d.f. estimation remain the same. The IFFD use regular control lattices in 3D and a 3D tensor product of B-spline polynomials. For model dynamics, the definitions for all edge terms and region terms remain valid in 3D. Compared to traditional deformable models, Metamorphs have several advantages in higher dimensions. Compared to the parametric models, Metamorphs use implicit distance functions to represent model shapes, hence eliminate the need for explicit parameterization of high-dimensional shapes; compared to the geometric models, Metamorphs use IFFD which possess implicit smoothness constraints and have far fewer parameters than an evolving front surface, hence enable more efficient computation.

In the current implementation of Metamorphs, we assume user-guided model initialization. That is, the user initializes one or several circular models within the objects of interest by clicking two points for each model: the first point is the centroid, and the distance between the first and the second point specifies the radius. Our method is robust to poor initializations where a model covers part of the background, as long as the majority of the model interior has consistent texture with the object of interest. This is because all energy terms in our framework generate two-way forces, and the model will be pulled back toward object boundary if it is initialized outside. We could also potentially automate the initialization process through supervised learning of the texture statistics of the objects.

When presented with different initialization conditions for the same task, the converged Metamorphs models do have small differences, but the variations are mostly within a small range around the ground truth. Such variations can be seen from Fig. 3.10(e) vs. Fig. 3.18(1.g), and Fig. 3.13(2.e) vs. Fig. 3.18(2.g). Since it is hard even for humans to reach consensus on

a unique ground truth for most segmentation tasks, we believe a good strategy is to take the average of several converged models that resulted from different initializations, when high segmentation accuracy is desired.

Although we assume user-defined seed points to start the simple-shape initial models, our method can be directly applied to full-field image segmentation by starting multiple initial models on a regular lattice covering the image. The topology freedom of the models enables evolving models with similar statistics to merge, and finally the image is partitioned into regions of homogeneous textures.

Another interesting future direction is to explore dynamically changing edge maps instead of one static edge map pre-computed using canny edge detector. This is analogous to the existing capability of our framework to learn and exploit the dynamically changing model-interior intensity statistics, which distinguishes the method from other region based segmentation techniques such as region growing.

To use Metamorphs models to segment objects with holes, the basic idea is to segment layer by layer. That is, we can first segment the inner-most layer, then for the next outer layer, we can use the inner-layer boundary as initialization, and exclude the inner-layer interior when computing the new model-interior statistics. Coupling Metamorphs with statistical prior models such as those learned through Active Shape and Appearance Models could also help.

Other than segmentation, the Metamorphs framework can also be applied to many other applications such as tracking, shape reconstruction, etc. For tracking in a video sequence, we can use the converged model from a previous frame as the initialization for the next frame, then the Metamorphs dynamics are the same as that in segmentation to guide the model convergence on the new frame. Useful tracking techniques such as Kalman filtering can also be integrated with the Metamorphs in similar manners to their integration with the Snakes. For shape reconstruction, the forces driving a Metamorphs model can be derived from the distance between the model and the sparse point set, and the free form deformations enable the model to smoothly interpolate between the sparse points.

Appendix

We can analytically derive the partial derivatives $\frac{\partial}{\partial \mathbf{\Theta}_i} D(\mathbf{\Theta}; \mathbf{x})$ for the incremental FFD parameters in $\mathbf{\Theta}$:

$$\delta F_{m,n} = (\delta F^x_{m,n}, \delta F^y_{m,n});\ m = 1, ..., M, \quad n = 1, ..., N$$

Without loss of generality, one can consider the $(m, n)th$ control point and its deformations in both directions. Then, from the definition for the deformations $D(\mathbf{\Theta}; \mathbf{x})$, the following relations hold:

$$\frac{\partial\ \delta D(\mathbf{\Theta}; \mathbf{x})}{\partial \delta F^x_{m,n}} = \begin{cases} \begin{bmatrix} B_{m-i}(u)\ B_{n-j}(v) \\ 0 \end{bmatrix}, & 0 \leq m - i, n - j \leq 3 \\ \\ \mathbf{0}, & otherwise \end{cases}$$

$$\frac{\partial\ \delta D(\mathbf{\Theta}; \mathbf{x})}{\partial \delta F^y_{m,n}} = \begin{cases} \begin{bmatrix} 0 \\ B_{m-i}(u)\ B_{n-j}(v) \end{bmatrix}, & 0 \leq m - i, n - j \leq 3 \\ \\ \mathbf{0}, & otherwise \end{cases}$$

Chapter 4

Learning Coupled Shape and Appearance Prior Models for Segmentation

This chapter presents a novel framework for learning coupled shape and appearance prior models using the implicit shape representation and free form deformations introduced in Chapter 2. First, a new global-to-local shape registration algorithm is developed. It can be used to establish continuous, smooth and one-to-one correspondences between shapes in arbitrary dimension. In particular, we use it to register boundary shapes of training examples and establish correspondences between them in order to learn a statistical shape model. Second, as a natural extension of the shape registration algorithm, a new joint registration algorithm is introduced to register images (or training examples) in a joint shape and intensity feature space. It establishes correspondences for shapes and interior textures simultaneously by maximizing mutual information in both shape and intensity spaces. Third, the dense correspondences established by the registration algorithms are used to build a coupled shape and appearance statistical model, then the model is applied to robust prior-model guided segmentation and image interpretation.

4.1 Introduction

Learning shape and appearance prior representations for an object of interest has been central to many model-based medical image analysis and computer vision algorithms. Using shape models to guide image search produces reliable segmentation results in noisy, cluttered images. A generalization to statistical appearance models uses also the interior region information, and enables registration of a target object with the learned prior model. Being complementary to each other, the integration of statistical shape and appearance models results in a powerful image analysis paradigm.

Although numerous methods in the literature have been proposed to learn shape and appearance prior models, most are hampered by the automated alignment and registration problem of training examples. In the seminal work of Active Shape and Appearance Models (ASM [26] and AAM [25]), models are built from analysing the shape and appearance variabilities across a set of labelled training examples. Typically landmark points are carefully chosen and manually placed on all examples by experts to assure good correspondences. This assumption leads to a natural framework for alignment and statistical modeling, yet it also makes the training process time-consuming. Yang & Duncan [120] proposed a shape-appearance joint prior model for Bayesian image segmentation. Their work does not deal with registration of the training examples, however, and assumes the training data are already aligned.

A number of automated shape registration and model building methods have been proposed [28], [34], [50], [12], [20]. These approaches either establish correspondences between geometric features, such as critical points of high curvature [50]; or find the "best" corresponding parametrization model by optimizing some criterion, such as minimizing accumulated Euclidean Distance [34], [20], Minimum Description Length [28], or Spline Bending Energy [12], [20]. Both geometric feature based and explicit parameterization based registration methods are not suitable for incorporating region intensity information. In [64], the implicit shape representation using level sets is considered, and shape registration algorithms using this representation have been proposed [84, 54].

Non-rigid registration is a popular approach to build statistical atlas and to model the appearance variations [42, 93, 17]. The basic idea is to establish dense correspondences between textures through non-rigid registration. However, few of the existing methods along this line are able to take into account shape information or to be coupled with shape registration.

In this chapter, we introduce a new algorithm for global-to-local shape registration based on the implicit shape representation, mutual information and free form deformations. The shape registration algorithm is then applied to register training examples of an object of interest to learn a statistical shape model. The registration framework is easily extensible to a unified shape and intensity feature space, hence we can achieve registration in the joint shape and intensity space by maximizing mutual information between both shape and intensity. The dense correspondences between training shape and texture enable us to learn a coupled shape and

appearance prior model, which is then used for robust model-based segmentation. In the remainder of the chapter, we first introduce the shape registration and statistical shape modeling algorithms, we then introduce registration and learning in the joint shape and intensity spaces and model-based segmentation using the coupled prior model.

4.2 Global-to-local Shape Registration in Implicit Spaces

4.2.1 Previous Work on Shape Registration

Shape registration is critical to various imaging and vision applications [111]. Global registration, also known as shape alignment, aims to recover a global transformation that brings the pose of a source shape as close as possible to that of a target shape. The alignment has extensive uses in recognition, indexing and retrieval, and tracking. To further account for important local deformations, non-rigid local registration is needed to establish dense correspondences between the basic elements of shapes, such as points, curvature, etc. Medical imaging is a domain that requires local registration such as in building statistical models for internal organs [24], and intra-subject or atlas registration of 2D/3D anatomical structures.

There has been a lot of previous research on the shape registration problem [85, 125, 10], as well as on similar problems such as shape matching [7, 24, 29, 100], and point set matching [19]. The algorithms proposed differ in the following three main aspects.

1. **Shape Representation** is the selection of an appropriate representation for the shapes of interest. Clouds of points [7, 19], parametric curves/surfaces [29, 74], fourier descriptors [106], medial axes [101], and more recently, implicit distance functions [85, 64] are often considered.

2. **Transformation** refers to the selected global, local, or hierarchical (global-to-local) transformation model, which is used to transform the source shape to match with the target shape. Global transformation models apply to an entire shape; and examples are rigid, similarity, affine and perspective. Local transformation models can represent pixel-wise deformations that deform a shape locally and non-rigidly; and examples include optical flow [85, 16], Thin Plate Splines (TPS) [7, 19], and space deformation techniques such

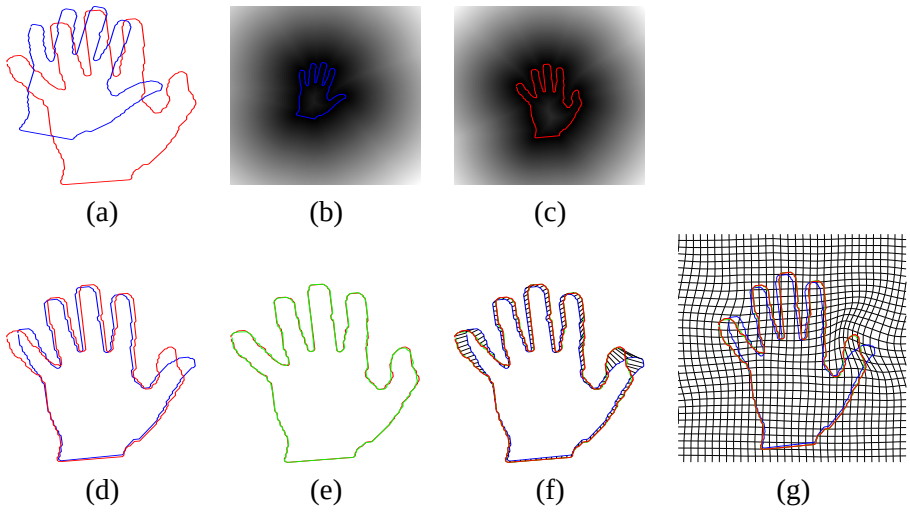


Figure 4.1: (a) Initial condition(source shape in blue, target shape in red). (b) The implicit source shape representation using a distance map; points on the shape overlap the zero level set (as drawn in color). (c) The implicit target shape representation. (d) Global alignment using Mutual Information; only the aligned shapes (zero level sets of the implicit representations) are shown. (e) Result after local non-rigid registration using IFFD; the transformed source shape (in green) is shown overlaid on the target shape (in red). (f) Established correspondences using IFFD. (g) The embedding space deformation to achieve local registration.

as Free Form Deformations (FFD) [102, 94]. Hierarchical models are also popular since they cover the entire transformation domain using both global and local transformations.

3. **Registration Criterion** is the approach used to recover the optimal transformation parameters given a shape representation and a transformation model. One can classify existing approaches into two sub-categories. The first is to establish explicit geometric feature correspondences and then estimate the transformation parameters using the correspondences [7, 44, 58]. The second is to recover the optimal transformation parameters through optimization of energy functionals [125, 10, 85, 19].

4.2.2 Overview of Our Shape Registration Algorithm

We propose a hierarchical shape registration method using the implicit distance function shape representation in a variational framework. Our overall approach is depicted in [Fig. (4.1)]. The shapes of interest are represented in an implicit form (see Ch. 2), embedded in the space of distance functions of a Euclidean metric [Fig. (4.1).b-c]. Global alignment using an arbitrary motion model is achieved by maximizing mutual information [22, 114, 107] between two

shape-embedding distance functions [Fig. (4.1).d]. For local non-rigid registration, parameters of the cubic B-spline based Incremental Free Form Deformation model (IFFD) are recovered by minimizing the sum-of-squared-differences between the two globally aligned distance functions [Fig. (4.1).e]. The resulting registration field [Fig. (4.1).g] preserves shape topology, is smooth, continuous and gives dense one-to-one correspondences between the source and the target shapes [Fig. (4.1).f].

4.2.3 Global Registration by Maximizing Mutual Information

As discussed in Chapter 2, the implicit shape representation used in our framework is inherently translation/rotation invariant [85]. When a shape undergoes scale variations, the intensity values of its associated distance map (i.e., its implicit representation) scale accordingly. Therefore the registration of distance maps of a shape in various scales is analogous to matching images in multiple modalities that refer to the same underlying scene elements. Mutual information, an information-theoretic criterion for measuring the global statistical dependency of its two input random variables, has been shown in the literature [87] to be able to address such matching objective. The integration of mutual information and the implicit representation gives rise to a global alignment framework that is invariant to translation, rotation, scaling, and accommodates transformations in arbitrary dimensions.

In order to facilitate notation let us denote the source shape representation $\Phi_{\mathcal{D}}$ as f [Fig. (4.1).b], and the target shape representation $\Phi_{\mathcal{S}}$ as g [Fig. (4.1).c]. Both f and g are intensity "images" where the intensity values refer to the distance values to the underlying shapes respectively. In the most general case, let us consider a sample domain Ω in the image domain of the source representation f [1], then global registration is equivalent to recovering the parameters $\mathbf{\Theta} = (\theta_1, \theta_2, ..., \theta_N)$ of a parametric transformation A, such that the mutual information between $f_\Omega = f(\Omega)$ and $g^A_\Omega = g\big(A(\mathbf{\Theta}; \Omega)\big)$ is maximized. The definition for such mutual information is:

$$MI(f_\Omega, g^A_\Omega) = \mathcal{H}\left[p^{f_\Omega}(l_1)\right] + \mathcal{H}\left[p^{g^A_\Omega}(l_2)\right] - \mathcal{H}\left[p^{f_\Omega, g^A_\Omega}(l_1, l_2)\right] \tag{4.1}$$

[1] In practice, this sample domain contains the collection of pixels in a narrow band around the zero level set.

The terms in the above formula are: (i) l_1 and l_2 denote the intensity (distance value) random variables in the domains f_Ω and g_Ω^A respectively; (ii) $\mathcal{H}$ represents the differential entropy; (iii) p^{f_Ω} is the intensity probability density function (p.d.f.) in the source sample domain f_Ω; (iv) $p^{g_\Omega^A}$ is the intensity p.d.f. in the projected target domain g_Ω^A; and (v) p^{f_Ω,g_Ω^A} is their joint distribution.

This mutual information measures the general dependence between the target distance function and the transformed source distance function. It consists of three components: (i) the entropy of the source, $\mathcal{H}\left[p^{f_\Omega}(l_1)\right]$, (ii) the entropy of the projection of the source on the target given the transformation, $\mathcal{H}\left[p^{g_\Omega^A}(l_2)\right]$, and (iii) the joint entropy between the source and its projection on the target, $\mathcal{H}\left[p^{f_\Omega,g_\Omega^A}(l_1,l_2)\right]$. Maximizing this mutual information quantity encourages transformations where f_Ω statistically correlate with g_Ω^A.

We can further expand the formula in [Eq. 4.1] using the definition for differential entropy:

$$\mathcal{H}\left[p^{f_\Omega}(l_1)\right] = -\int_{\mathcal{R}^1} p^{f_\Omega}(l_1) log p^{f_\Omega}(l_1) dl_1 = -\iint_{\mathcal{R}^2} p^{f_\Omega,g_\Omega^A}(l_1,l_2) log p^{f_\Omega}(l_1) dl_1 dl_2 \quad (4.2)$$

$$\mathcal{H}\left[p^{g_\Omega^A}(l_2)\right] = -\int_{\mathcal{R}^1} p^{g_\Omega^A}(l_2) log p^{g_\Omega^A}(l_2) dl_2 = -\iint_{\mathcal{R}^2} p^{f_\Omega,g_\Omega^A}(l_1,l_2) log p^{g_\Omega^A}(l_2) dl_1 dl_2 \quad (4.3)$$

$$\mathcal{H}\left[p^{f_\Omega,g_\Omega^A}(l_1,l_2)\right] = -\iint_{\mathcal{R}^2} p^{f_\Omega,g_\Omega^A}(l_1,l_2) log p^{f_\Omega,g_\Omega^A}(l_1,l_2) dl_1 dl_2 \quad (4.4)$$

Combining [Eqs. 4.1, 4.2, 4.3 and 4.4], one can derive the criterion to perform global alignment using an arbitrary transformation model A with parameters $\mathbf{\Theta}$, by maximizing mutual information, which is equivalent to minimizing the following energy functional:

$$E_{Global}(A(\mathbf{\Theta})) = -MI(f_\Omega, g_\Omega^A) = -\iint_{\mathcal{R}^2} p^{f_\Omega,g_\Omega^A}(l_1,l_2) log \frac{p^{f_\Omega,g_\Omega^A}(l_1,l_2)}{p^{f_\Omega}(l_1) p^{g_\Omega^A}(l_2)} dl_1 dl_2 \quad (4.5)$$

The probability density functions in the energy functional are approximated using a nonparametric, differentiable Gaussian Kernel-based Density Estimation model. Using this model, the marginal probability density functions are:

$$p^{f_\Omega}(l_1) = \frac{1}{V(\Omega)} \iint_\Omega G(l_1 - \underbrace{f(\mathbf{x})}_{\alpha}) d\mathbf{x} \quad (4.6)$$

$$p^{g_\Omega^A}(l_2) = \frac{1}{V(\Omega)} \iint_\Omega G(l_2 - \underbrace{g(A(\mathbf{\Theta};\mathbf{x}))}_{\beta}) d\mathbf{x} \quad (4.7)$$

Where $\mathbf{x} = (x, y)$ refer to pixels in the sample domain Ω, $V(\Omega)$ represents the volume of the sample domain, and $G(a)$ represents the value of a one dimensional zero-mean Gaussian kernel at location a:

$$G(a) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{a^2}{2\sigma^2}}$$

where σ is a small constant controlling the kernel width (we set $\sigma = 4$ in all experiments).

Similarly, we can derive the expression for the joint probability density function using a two dimensional zero-mean Gaussian kernel:

$$p^{f_\Omega, g^A_\Omega}(l_1, l_2) = \frac{1}{V(\Omega)} \iint_\Omega G(l_1 - \underbrace{f(\mathbf{x})}_{\alpha}, l_2 - \underbrace{g(A(\mathbf{\Theta}; \mathbf{x}))}_{\beta}) d\mathbf{x} \tag{4.8}$$

where the 2D kernel $G(a, b)$ is given by:

$$G(a, b) = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}(\frac{a^2}{\sigma_1^2} + \frac{b^2}{\sigma_2^2})}$$

and σ_1, σ_2 are the constants specifying the kernel widths in 2D.

The calculus of variations with a gradient descent method can now be used to minimize the cost function E_{Global} and recover the transformation parameters $\theta_i, i = 1, ..., N$. The parameter evolution equations are derived as follows:

$$\begin{aligned} \frac{\partial E_{Global}}{\partial \theta_i} = \quad & -\tfrac{1}{V(\Omega)} \iint_\Omega \Bigg[\iint_{\mathcal{R}^2} \Big(1 + log \tfrac{p^{f_\Omega, g^A_\Omega}(l_1, l_2)}{p^{f_\Omega}(l_1) p^{g^A_\Omega}(l_2)}\Big) \Big(- G_\beta(l_1 - \alpha, l_2 - \beta)\Big) dl_1 dl_2 \Bigg] \\ & \big(\nabla g(A(\mathbf{\Theta}; \mathbf{x})) \cdot \tfrac{\partial}{\partial \theta_i} A(\mathbf{\Theta}; \mathbf{x})\big) d\mathbf{x} \end{aligned} \tag{4.9}$$

By substituting the general parameters $\mathbf{\Theta}$ with specific transformation parameters, the method supports registration using any global transformation model for shapes in 2D/3D.

Examples of such global alignment for 2D shapes using the similarity transformation model are given in [Fig. (4.2).1-4], and 2D examples using the affine transformation are shown in [Fig. (4.2).4]. The 3D registration examples will be shown in Chapter 4.2.5.

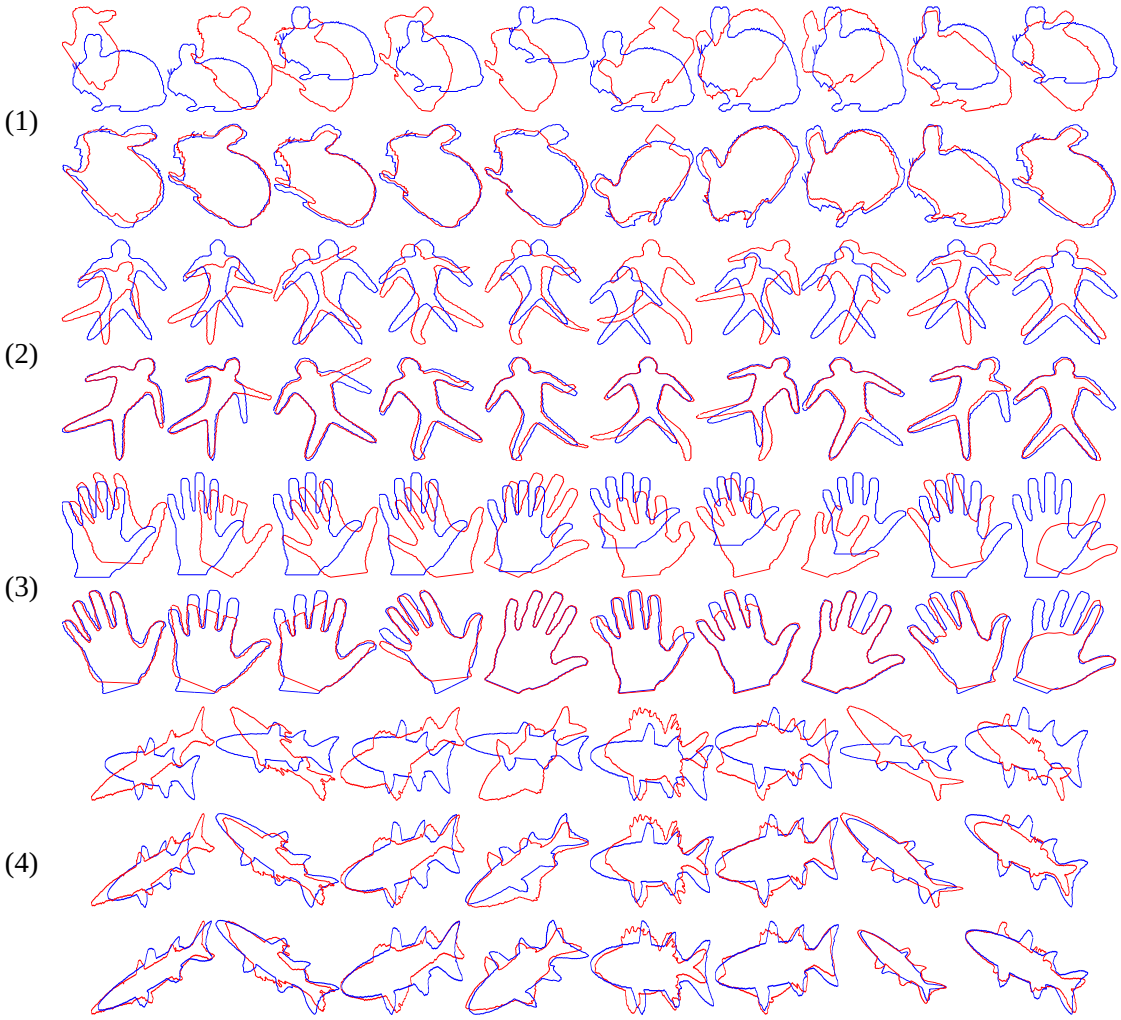
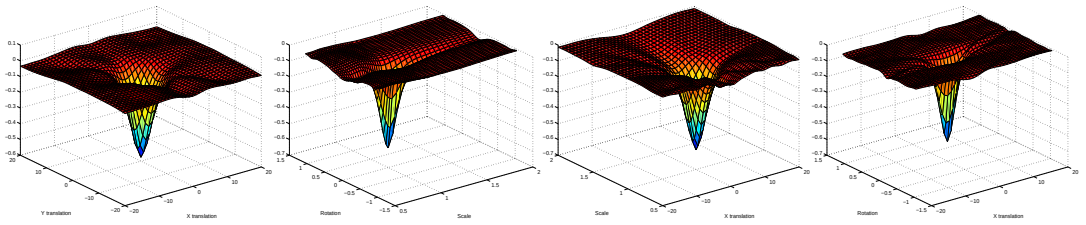


Figure 4.2: Global registration examples. (1) Bunny, (2) Dude, (3) Hand, (4) Fish. (odd rows) Initial conditions (source in blue vs. target in red), (even rows) Alignment result using the similarity transformation model, (last row) Alignment result using the Affine transformation. Each column corresponds to a different trial. Only the zero level sets of the registered distance functions are shown in contour form.



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

Figure 4.3: Empirical validation of global registration (a) Translations in x, y directions unknown, (b) Scale and rotation unknown, (c)Translation in x and scale unknown, (d) Translation in x and rotation unknown.

Empirical Evaluation of the Global Criterion

Gradient descent optimization techniques often suffer from being sensitive to the initial conditions. The form of the objective function is a good indicator regarding the efficiency and stability of an optimization framework.

In order to perform a study on the performance of our global registration technique, we take the 2D similarity transformation model with four parameters: translations in x and y directions respectively, isotropic scale factor and the 2D rotation angle. Then we constrain the unknown parameter space in two dimensions, and empirically evaluate the form of the global registration objective function. For an example "dude" shape [Fig. (4.2).2], we have studied the following four cases: (1) translations in x, y directions are unknown [Fig. (4.3).a], (2) scale and rotation are unknown [Fig. (4.3).b], (3) translation in x and scale are unknown [Fig. (4.3).c], and (4) translation in x and rotation are unknown [Fig. (4.3).d]. In each case, we quantized the search space using a uniform sampling rule (100 elements) for all unknown parameters. Translations in (x, y) were in the range of $[-20, 20] \times [-20, 20]$, scale was in $[0.5, 2.0]$ and rotation in $\left[-\frac{\pi}{3}, \frac{\pi}{3}\right]$. Then, one can estimate the projections of the objective function in the space of two unknown parameters, by considering all possible combinations derived from the sampling strategy (the other two parameters are fixed). The resulting projections of the functional, as shown in [Fig. (4.3).a-d], have some nice properties: they are smooth and exhibit a single global minimum. Hence the objective function has a convex form for all combinations that involve two unknown registration variables and this is a good indicator for a well-behaved optimization criterion with smooth convergence properties.

4.2.4 Free Form Local Registration and Correspondences

Global registration can be an acceptable solution to a large number of computer vision applications. Medical imaging is an area where quite often global motion is not a valid answer when solving the dense registration and correspondences problem [39]. Local deformations are a complementary component to the global registration model. However, dense local motion (warping fields) estimation is an ill-posed problem since the number of variables to be recovered is often larger than the number of available constraints. Smoothness as well as other forms of constraints were employed to cope with this limitation.

In our shape registration framework, a global transformation A is recovered using the mutual information criterion. One can use such transformation to transform the source shape $\mathcal{D}$ to a new shape $\hat{\mathcal{D}} = A(\mathcal{D})$. Then, on top of this global registration result, local registration is equivalent to recovering a pixel-wise local deformation field that creates correspondences between the implicit representation $[\Phi_{\mathcal{S}}]$ of the target shape $\mathcal{S}$ and the implicit representation $[\Phi_{\hat{\mathcal{D}}}]$ of the transformed source shape $\hat{\mathcal{D}}$. Such a local deformation field $D(\mathbf{x})$ can be represented using the Incremental Free Form Deformations (IFFD) model introduced in Chapter 2 (See Eq. 2.15), since IFFD can implicitly enforce smoothness constraints, it preserves shape topology and guarantees a one-to-one correspondence between two shapes.

Local Registration Optimization Criterion and Gradient Descent

Considering the Incremental Free Form Deformations (IFFD) formulation (see Ch. 2), dense registration is achieved by incrementally evolving a control lattice P according to a deformation improvement $[\delta P]$, and the inference problem is solved by minimizing a Sum-of-Squared-Differences criterion, with respect to the control lattice deformation improvements, which are the parameters of IFFD: $\mathbf{\Theta} = \delta P = \{(\delta P^x_{m,n}, \delta P^y_{m,n})\};\ (m,n) \in [1,M] \times [1,N]$. That is, local registration is equivalent to finding the control lattice deformation δP such that when applied to the embedding space of the source shape, the deformed source shape coincides with the target shape. Since the structures to be locally registered in our framework are the distance transform of the target shape - $[\Phi_{\mathcal{S}}]$, and the distance transform of its globally aligned source

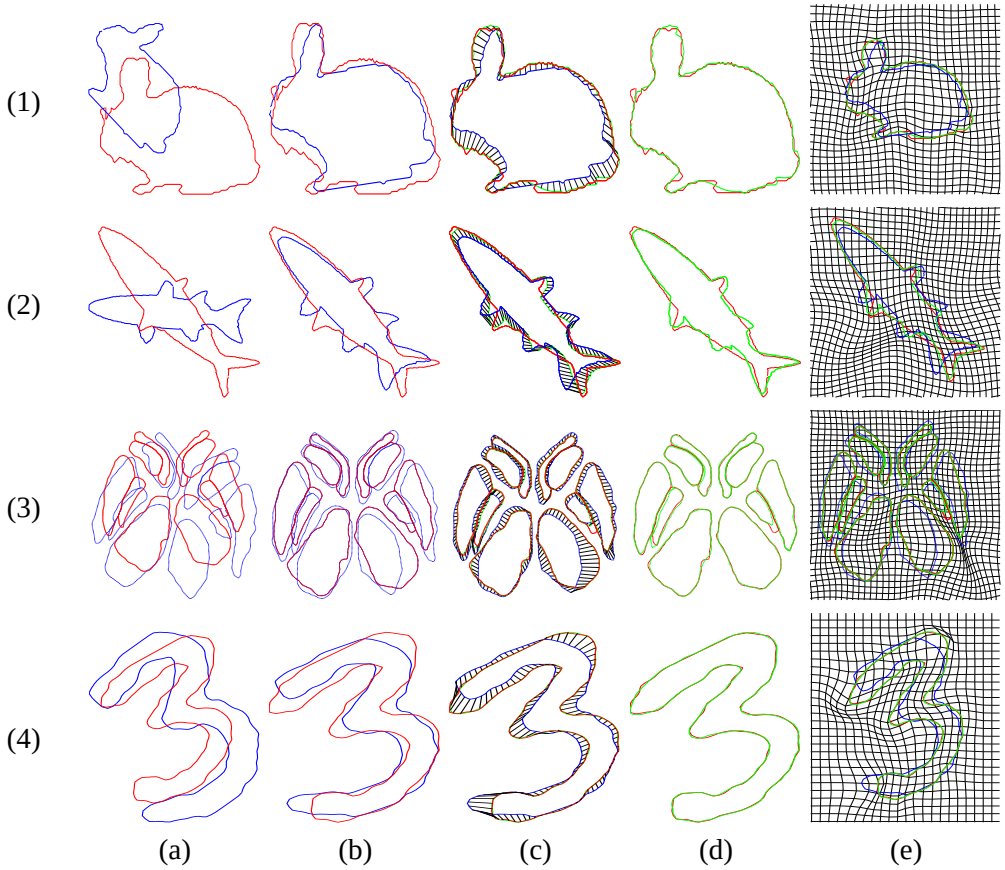


Figure 4.4: Incremental B-spline FFD local registration. (1) Bunny, (2) Fish, (3) Brain Structure, (4) Digit 3. (a) Initial conditions (source shape in blue, target shape in red), (b) Result after global registration, (c) Established correspondences after local registration; only the zero level set (i.e., shape) correspondences are shown, (d) Locally deformed source shape (in green) overlaid on the target (in red), (e) Final IFFD control lattice configuration depicting the space warping to achieve local registration.

shape - $[\Phi_{\hat{\mathcal{D}}}]$, the Sum-of-Squared-Differences (SSD) criterion can be considered as the data-driven term to recover the deformation field $D(\mathbf{\Theta};\mathbf{x}))$:

$$E_{data}(\mathbf{\Theta}) = \iint_{\Omega} \left(\Phi_{\hat{\mathcal{D}}}(\mathbf{x}) - \Phi_{\mathcal{S}}(D(\mathbf{\Theta};\mathbf{x}))\right)^2 d\mathbf{x} \tag{4.10}$$

In order to further preserve the regularity of the recovered registration flow field, one can consider an additional smoothness term on the local deformation field δL. We consider a computationally efficient smoothness term:

$$E_{smoothness}(\mathbf{\Theta}) = \iint_{\Omega} \left(\left|\left| \frac{\partial \delta D(\mathbf{\Theta};\mathbf{x})}{\partial x} \right|\right|^2 + \left|\left| \frac{\partial \delta D(\mathbf{\Theta};\mathbf{x})}{\partial y} \right|\right|^2 \right) d\mathbf{x} \tag{4.11}$$

Such a smoothness term is based on an error norm with known limitations. One can replace this smoothness component with more elaborate norms. Within the proposed framework, an implicit smoothness constraint is also imposed by the B-Spline FFD, which guarantees C^1 continuity at control points and C^2 continuity everywhere else. Therefore there is no need for introducing complex and computationally expensive regularization components.

The data-driven term [Eq. 4.2.4] and the smoothness term [Eq. 4.2.4] can now be integrated into one energy functional to recover the IFFD parameters:

$$E_{Local}(\mathbf{\Theta}) = \iint_{\Omega} \left(\Phi_{\hat{\mathcal{D}}}(\mathbf{x}) - \Phi_{\mathcal{S}}(D(\mathbf{\Theta};\mathbf{x}))\right)^2 d\mathbf{x} + \alpha \iint_{\Omega} \left(\left|\left| \frac{\partial \delta D(\mathbf{\Theta};\mathbf{x})}{\partial x} \right|\right|^2 + \left|\left| \frac{\partial \delta D(\mathbf{\Theta};\mathbf{x})}{\partial y} \right|\right|^2 \right) d\mathbf{x} \tag{4.12}$$

where α is the constant balancing the contribution of the two terms. In our experiments, the typical values for α are in the range of $1 \sim 5$; smaller values lead to faster convergence, while larger values result in smoother deformation fields. The one-to-one mapping property is guaranteed regardless of the α value.

The calculus of variations and a gradient descent method can be used to optimize the local registration objective function [Eq. 4.12]. One can obtain the following evolution equation for each parameter θ_i in the IFFD control lattice deformation parameters $\mathbf{\Theta}$:

$$\begin{aligned} \frac{\partial}{\partial \theta_i} E_{Local}(\mathbf{\Theta}) = & -2 \iint_{\Omega} \left(\Phi_{\hat{\mathcal{D}}}(\mathbf{x}) - \Phi_{\mathcal{S}}(D(\mathbf{\Theta};\mathbf{x}))\right) \left(\nabla \Phi_{\mathcal{S}}(D(\mathbf{\Theta};\mathbf{x})) \cdot \frac{\partial}{\partial \theta_i} \delta D(\mathbf{\Theta};\mathbf{x})\right) d\mathbf{x} \\ & + 2\alpha \iint_{\Omega} \frac{\partial}{\partial x} \delta D(\mathbf{\Theta};\mathbf{x}) \cdot \frac{\partial}{\partial \theta_i} \left(\frac{\partial}{\partial x} \delta D(\mathbf{\Theta};\mathbf{x}) \right) + \frac{\partial}{\partial y} \delta D(\mathbf{\Theta};\mathbf{x}) \cdot \frac{\partial}{\partial \theta_i} \left(\frac{\partial}{\partial y} \delta D(\mathbf{\Theta};\mathbf{x}) \right) d\mathbf{x} \end{aligned} \tag{4.13}$$

The partial derivatives in the above formula can be easily derived from the model deformation equation in [Eq. 2.15]. Details are given in the Appendix.

Once the optimal IFFD parameters and the local registration field $\hat{L}$ are derived, dense one-to-one correspondences can be established between each point $\mathbf{x} = (x, y)$ on the source structure, with its deformed position $\hat{L}(\mathbf{x})$ on the target structure. These correspondences include not only the correspondences for those pixels located on the zero level set, which are points on the source and target shapes, but also correspondences between nearby level sets which are clones of the original shapes coherently positioned in the embedding image/volume space.

The performance of the proposed local registration paradigm is demonstrated for various 2D examples shown in [Figs. (4.4, 4.1)]. And 3D examples will be presented in Chapter 4.2.5 (see [Fig. (4.6), as well as in Chapter 6.2.4 (see [Fig. (6.8)]).

Multi-resolution Incremental Free Form Deformations (IFFD)

To account for both large-scale and highly local non-rigid deformations, we can use an efficient multi-level implementation of the IFFD framework, as shown in [Fig. (4.5)]. To this end, multi-resolution control lattices are used according to a coarse-to-fine strategy. A coarser level control lattice is applied first to account for relatively global non-rigid deformations; then the space deformation resulting from the coarse level registration is used to initialize the configuration of a finer resolution control lattice, and at this finer level, the local registration process continues to deal with highly local deformations and achieve better matching between the deformed source shape and the target. Generally speaking, the hierarchy of control lattices can have arbitrary number of levels, but typically $2 \sim 3$ levels are sufficient. The layout of the control lattices in the hierarchy can be computed efficiently using a progressive B-spline subdivision algorithm [41]. At each level, we can solve for the incremental deformation of the control lattice using the scheme presented in section 4.2.4. In the end, the overall deformation field is defined by the incremental deformations from all levels. In particular, the total deformation $\delta D(\mathbf{x})$ for a pixel $\mathbf{x}$ in a hierarchy of r levels is:

$$\delta D(\mathbf{x}) = \sum_{k=1}^{r} \delta D^k(\Theta^k; \mathbf{x}) \tag{4.14}$$

where $\delta D^k(\Theta^k; \mathbf{x})$ refers to the deformation improvement at this pixel due to the incremental deformation Θ^k of the kth level control lattice.

4.2.5 Shape Registration in 3D

The proposed Global-to-local shape registration algorithm can be naturally extended to 3D. For global registration, parameters of a 3D transformation model can be solved by maximizing mutual information in the 3D sample domain. For local registration, free form deformations can be defined by the 3D tensor product of B-spline polynomials, and the SSD energy functional is

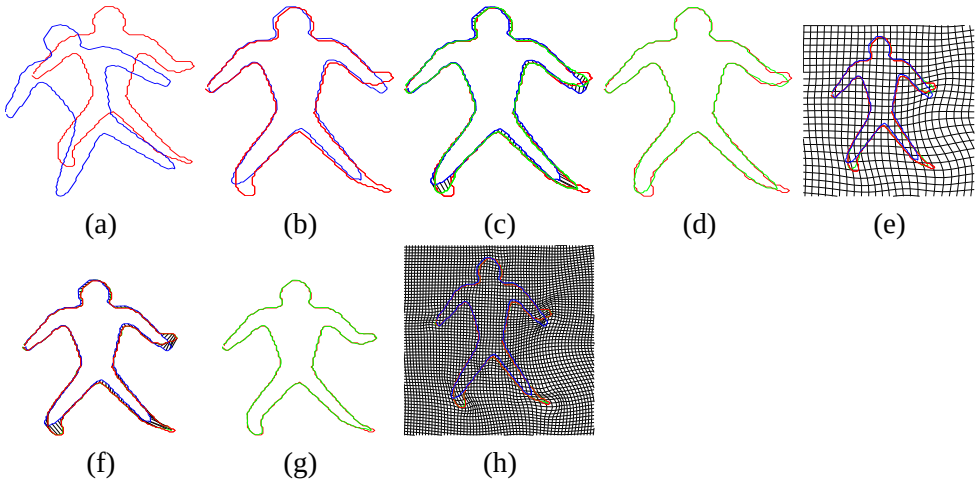


Figure 4.5: Multi-level Incremental FFD for local registration. (a) Initial Condition, (b) After global registration, (c) Established correspondences using a coarse resolution IFFD control lattice for local registration, (d) Coarse resolution matching result, (e) Coarse resolution control lattice (space) deformation, (f) Refined correspondences by a finer resolution IFFD control lattice, (g) Finer resolution matching result, (h) Finer resolution control lattice (space) deformation.

defined in the 3D volumetric domain. Geometric feature constraints can be specified in 3D as well to increase registration accuracy. The detailed 3D formulation and some 3D registration examples are given in Chapter 6.2.4.

Here we show one example of the 3D registration framework in [Fig. (4.6)] for registering a pair of 3D face range scans. The global transformation model consists of translation, scaling, and quaternion-based rotation [Fig. (4.6).1]. The local incremental FFD model uses control lattices in the 3D space and a 3D tensor product of B-spline polynomials. Qualitatively the result after global-to-local registration can be seen from two views: the front view [Fig. (4.6).2(front)], and the side view [Fig. (4.6).2(side)]. Quantitatively, the sum-of-squared-differences matching error (Eq. 4.2.4) after global registration was 8.3. The IFFD based local registration used three resolutions of control lattices in a coarse-to-fine manner and ran 20 iterations for each resolution. After the coarsest-level (10*10 lattice) IFFD local registration, the matching error was reduced to 3.4; after the middle level (20*20 lattice), the matching error was reduced to 1.8; and after the finest level (40*40 lattice), the matching error was reduced to 1.2. The total time spent for global and multi-level local registration was 4.6 minutes.

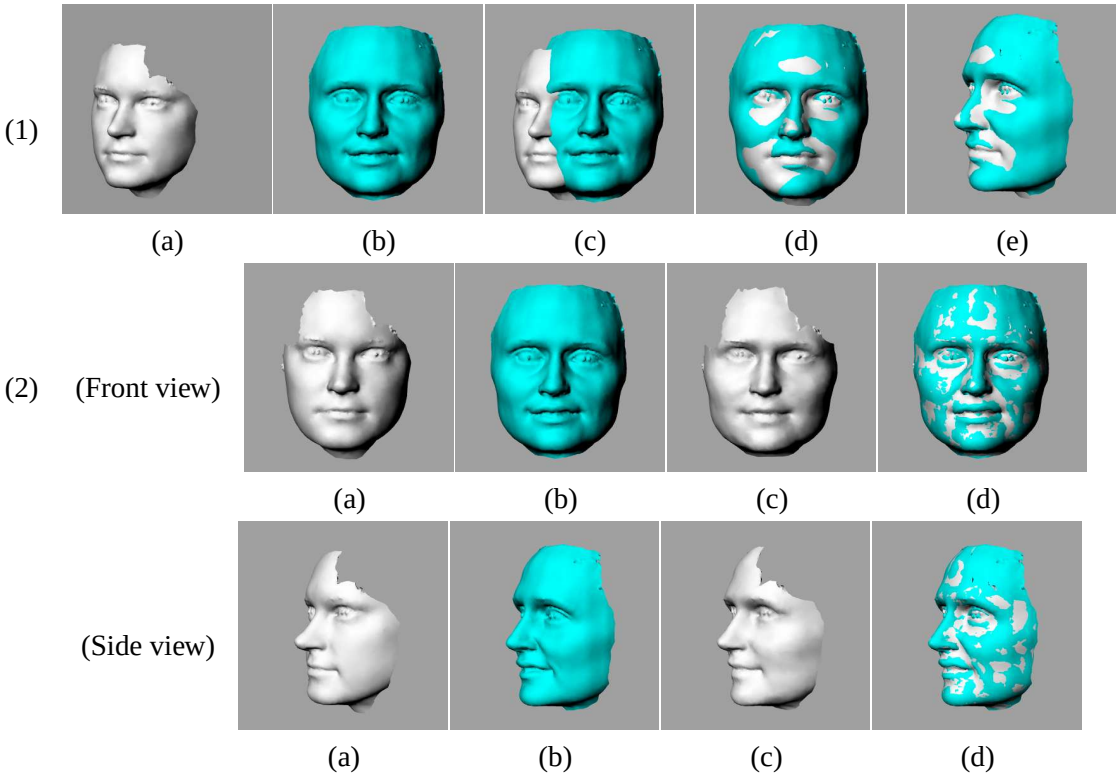


Figure 4.6: Global-to-local registration for open 3D structures (both source and target shapes are from face range scan data). (1) Global registration using the 3D similarity transformation model: (a) source shape; (b) target shape; (c) initial pose of the source relative to the target; (d & e) globally transformed source shown overlaid on the target - front view (d) and side view (e). (2) Local registration using IFFD: (Front view & Side view): (a) source shape after rigid transformation; (b) target shape; (c) locally deformed source shape after IFFD registration; (d) locally deformed source shape shown overlaid on the target.

4.3 Statistical Organ Shape Modeling and Prior Shape Model Guided Segmentation

The shape registration algorithm introduced in the previous section 4.2 can be applied to statistical shape modeling of objects of interest, because learning a compact representation that can capture the shape variations in an object of interest requires establishing dense correspondences across a set of training examples.

As an example, we show the statistical modelling of systolic left ventricle (LV) shapes from ultrasonic images, using 40 pairs of hand-drawn LV contours. We first apply global rigid registration to align all contours to the same target, as shown in [Fig. (4.7)]. Local registration based on free form deformations is then used to non-rigidly register all these contours to the

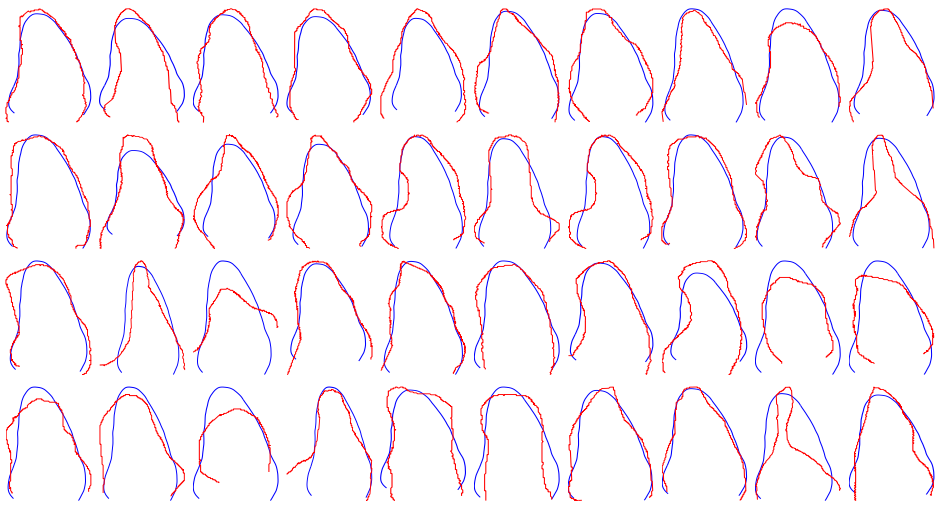


Figure 4.7: Rigid Registration for User-Determined Ground Truth (Systole) shapes of the Left Ventricle from Ultrasonic Images (multiple views). (blue) target mean shape, (red) registered source shape.

common target (see grid deformations in [Fig. (4.8)]). In order to establish dense one-to-one correspondences between all the aligned contours, we pick a set of sample points on the common target and compute their correspondences on each training contour based on the local registration result (see [Fig. (4.9)] for established local correspondences).

Using the established correspondences, the Principal Component Analysis (PCA) technique can be applied to build a Point Distribution Model (PDM) [24] to capture the statistics of the corresponding elements across the training examples. Assuming $\phi_{i=1\ldots n}$ are n column vectors, each representing the point coordinates from one example in the training set [Fig. (4.7)]. A zero mean assumption can be made for each vector $\{\phi_i\}$ by estimating the mean vector $\overline{\phi}$ (i.e. mean shape) and subtracting the mean vector from the training samples $\{\phi_i\}$. Given the set of training samples and the mean vector, one can define the covariance matrix as follows:

$$\left[\hat{\Sigma} = E\{\phi_i \phi_i^T\}\right]$$

It is well known that the principal orthogonal directions of maximum variation for $\{\phi_i\}$ are the eigenvectors of $\hat{\Sigma}$. One can replace the $\hat{\Sigma}$ with the sample covariance matrix that is given by $\left[\phi_M^T \phi_M\right]$. ϕ_M is the matrix formed by concatenating the set of examples $\phi_{i=1\ldots n}$. Then, the eigenvectors of $\hat{\Sigma}$ can be computed through the singular value decomposition (SVD) of

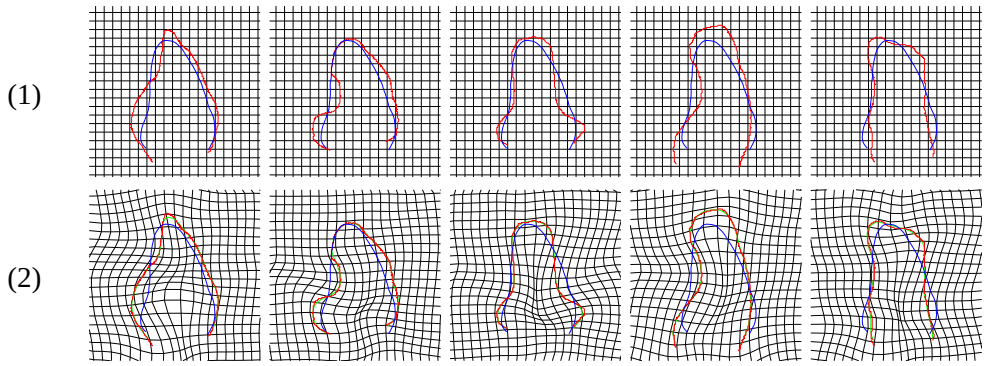


Figure 4.8: Local Non-rigid registration using Incremental FFD. (1) initial undeformed grid overlaid on global rigid registration result (blue - mean reference shape), (2) deformed grid to map the reference shape to various training shapes. Each column corresponds to a different trial.

$\phi_M = \mathbf{U}\Sigma\mathbf{V}^T$. The eigenvectors of the covariance matrix $\hat{\Sigma}$ are the columns of metric $\mathbf{U}$ while the elements of the diagonal matrix Σ are the eigenvalues which refer to the variance of the data in the direction of the basis vectors. The magnitude of the eigenvectors can be used to determine the number of basis vectors (m) to keep in order to retain the largest amount of variation within the training data while reducing those dimensions with very small variation. For any new example ϕ, assuming

$$\phi = \overline{\phi} + \Sigma_{j=1}^{m} b_j \, U_j$$

where $\overline{\phi}$ is the mean shape, m is the number of retained modes of variation, U_j are these modes (eigenvectors), and b_j are linear weight factors that combine the modes and reconstruct the new example, we can compute the linear weight factors

$$\mathbf{b} = \mathbf{U}^{+}(\phi - \overline{\phi})$$

where $\mathbf{b}$ is the vector consisting of $b_j, j = 1, ..., m$, and $\mathbf{U}^{+}$ is the pseudo-inverse of the eigenvector matrix $\mathbf{U}$. Typically, the new example is considered one instance of the learned object of interest, if all weight factors $b_j, j = 1, ...m$, are within the allowable range of variation defined by the eigenvalues.

On the ultrasonic systolic left ventricle shape example, the computed principal components for the statistical model can be seen in [Fig. (4.10)]. The model captures the variations in

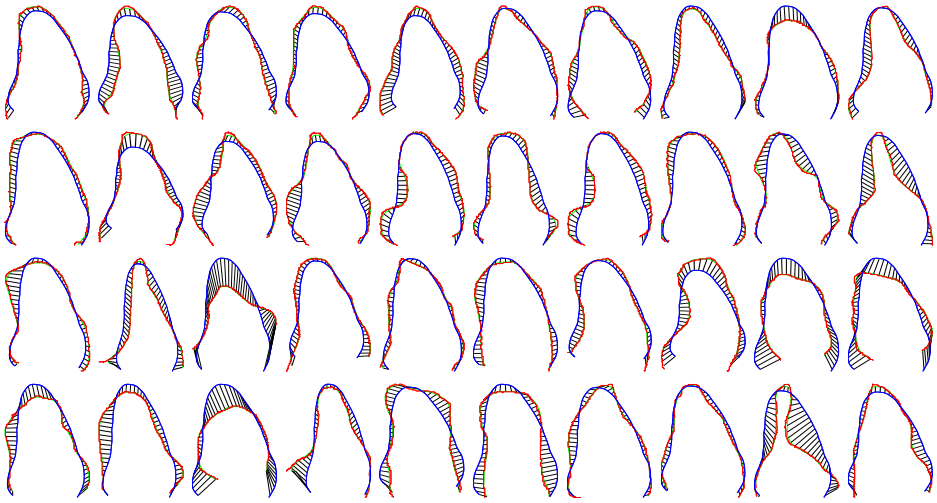


Figure 4.9: Established correspondences using IFFD. (red) source shapes after global transformations, (blue) target mean shape, (dark lines) correspondences for a fixed set of points on the mean shape.

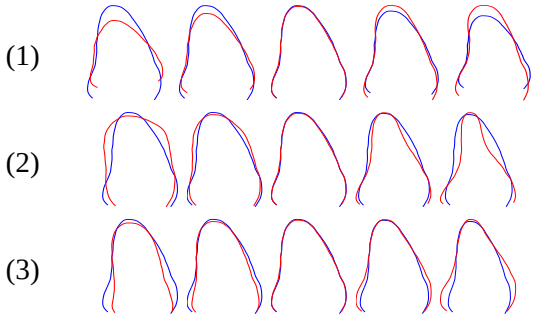


Figure 4.10: PCA modelling for the systolic Left Ventricle shapes using the established local correspondences. (1) first mode, (2) second mode, (3) third mode; For each mode, from left to right shows the mode changing from $-2\sqrt{\lambda_i}$ to $2\sqrt{\lambda_i}$.

the training set well, and generates new shapes that are consistent with the training examples. This also justifies to some extent the validity of the established correspondences using our registration algorithm.

The learned statistical shape model can be used to guide image search and segmentation on unseen images in a way similar to the Active Shape Models [26]. Using the statistical shape model of the left ventricle learned above (see [Fig. (4.10)]), we show some example segmentation results on ultrasound images of the heart during the systolic phase in Fig. 4.11.

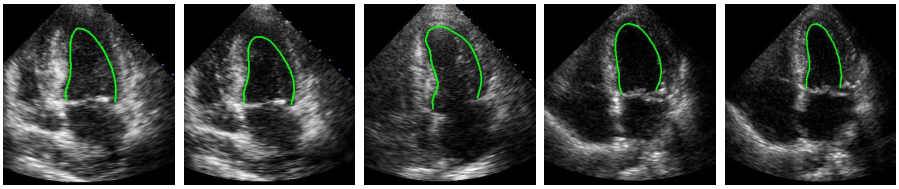
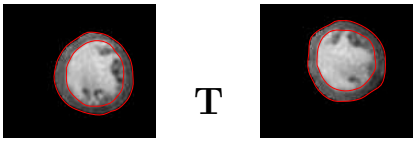


Figure 4.11: Statistical Shape Model guided segmentation of left ventricle shapes in echocardiograms (cardiac ultrasound images) during the systolic phase.



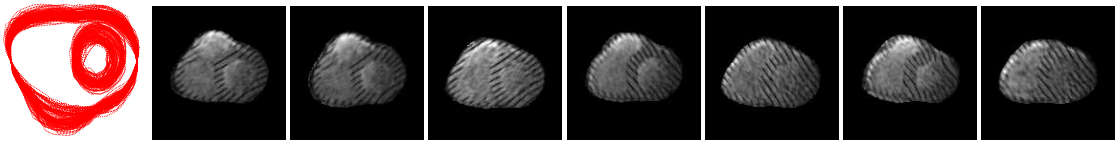
(a) (b)

Figure 4.12: Illustrating potential problem in shape-only registration. (a) Example one with contour shapes shown, (b) Example two, linked to Example one through an unknown transformation $\mathbf{T}$.

4.4 Learning Coupled Shape and Appearance Models and Model-based segmentation

Using statistical shape models to guide image search help produce reliable segmentation results in noisy, cluttered images. However, methods considering boundary shape alone may fail in robust registration of training examples for some objects of interest. [Fig. (4.12)] illustrates one example demonstrating potential problems in shape-only registration. The training examples are the left ventricle examples collected from a set of MRI images of the heart. Due to the near-to circular Left Ventricle shapes in the two different training examples in [Fig. (4.12)], it is unfeasible for any automated shape-only alignment algorithm to approximate the transformation (e.g. rotation angle) between the two. In this case, the joint registration in both shape and intensity spaces between training examples is imminent to address these limitations. Furthermore, joint registration using both shape and texture provides additional deformation constraints for the large area inside the object of interest.

The registration and learning framework introduced in section 4.2 and section 4.3 can be easily extended to a joint shape and intensity feature space. we present next the extended learning framework and demonstrate it to build a coupled shape and appearance prior model for the left ventricle and whole heart in short-axis cardiac tagged MR images. We then use the



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 4.13: The globally-aligned training examples. (a) All aligned contours overlaid together. (b-h) Some examples of the globally aligned textures. Note that due to tagging lines in the heart wall and RV topology irregularity, we consider the whole-heart shape but texture only inside the LV.

prior model to guide segmentation of the heart chambers in noisy, cluttered images, and we show quantitative validation on the segmentation results by comparing to expert solutions.

4.4.1 Unified Shape and Intensity Feature Space

Within our proposed framework, we represent each shape using a Euclidean distance map. In this way, shapes are implicitly represented as "images" in the space of distance transforms where shapes correspond to the zero level set of the distance functions. The level set values in the shape embedding space is analogous to the intensity values in the intensity (appearance) space. As a result, for each training example, we have two "images" of different modality, one representing its shape and another representing its intensity (grey-level appearance). The shape and intensity spaces are conveniently unified this way.

We use the Mutual Information as the similarity criterion to be optimized. Such information-theoretic criterion has been successfully used for dealing with images of multi-modalities. Suppose A and B are two training examples, and the similarity transformation between them is T. Let us denote the level set value random variables in the shape space for example A as X_S^A and the shape variable for example B given the transformation is $X_S^{T(B)}$, and denote their intensity random variables in the intensity space as X_I^A and $X_I^{T(B)}$ respectively. The Mutual Information between the two examples in the joint shape and intensity spaces given the transformation can be defined as:

$$\begin{aligned} \mathcal{M}_J(A, T(B)) &= \mathcal{M}_S(A, T(B)) + \alpha \mathcal{M}_I(A, T(B)) \\ &= \mathcal{H}(X_S^A) + \mathcal{H}(X_S^{T(B)}) - \mathcal{H}(X_S^A, X_S^{T(B)}) + \\ &\quad \alpha \big[\mathcal{H}(X_S^A) + \mathcal{H}(X_S^{T(B)}) - \mathcal{H}(X_S^A, X_S^{T(B)}) \big] \end{aligned} \tag{4.15}$$

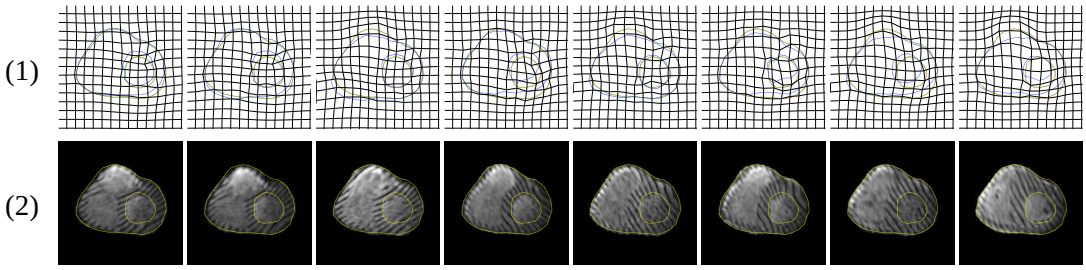


Figure 4.14: Demonstraing local FFD registration between training examples. (1) Each training shape (in blue) deforms to match a target mean atlas (in red). The deformed training shapes are shown in green. The FFD control lattice deformations are also shown. (2) The registered textures. Note that each training texture is non-rigidly deformed based on FFD and registered to a mean texture atlas. All textures cover a same area in the common reference frame. Dense pixel-wise correspondences are established.

where $\mathcal{H}$ represents the differential entropy and α is a constant balancing the contributions of shape and intensity in measuring the similarity. To solve for the transformation parameters T, we maximize the above mutual information similarity measure in Eq. 4.15. The optimization is done through a gradient-descent based method in a similar way to the optimization of Eq. 4.1. In Fig. 4.13, we show the aligned examples for an articulated whole heart model collected from our training set of tagged MRI images. Here we randomly picked one example as the atlas, and aligned all other examples to it by maximizing mutual information in the joint shape and intensity spaces.

4.4.2 Local Registration using IFFD and Mutual Information

After global alignment, the next step towards building a statistical shape and appearance model is to solve the dense correspondences problem. We extend the nonrigid shape registration algorithm in section 4.2.4 to the unified shape and intensity space, thus achieving simultaneous registration on both shapes and textures of the training examples. This joint registration provides additional constraints on the deformation field for the large area inside the object.

The Incremental Free Form Deformations (IFFD) is used to model the local deformations, and dense registration is achieved by evolving a control lattice P according to a deformation improvement $[\delta P]$.

To non-rigidly register the atlas A and a rigidly aligned training example B, we consider a sample domain Ω in the common reference frame. The mutual information criterion defined in

both shape and intensity spaces can be considered to recover the deformation field $\delta D(\mathbf{\Theta};\mathbf{x})$ that registers A and B:

$$E(\delta D(\mathbf{\Theta})) = \mathcal{M}_S\big(B(\Omega), A(D(\mathbf{\Theta};\Omega))\big) + \alpha\mathcal{M}_I\big(B(\Omega), A(D(\mathbf{\Theta};\Omega))\big) \tag{4.16}$$

In the equation, $D(\mathbf{\Theta};\Omega)$ represents the deformed domain of the initial sample domain Ω, i.e. $D(\mathbf{\Theta};\mathbf{x}) = \mathbf{x} + \delta D(\mathbf{\Theta};\mathbf{x})$, for any $\mathbf{x} \in \Omega$.

A gradient descent optimization technique is then used to optimize the mutual information cost function, and to recover the parameters of the dense, smooth, one-to-one registration field δD. Then correspondences can be established between each point $\mathbf{x} = (x, y)$ on example B, with its deformed position $\hat{D}(\mathbf{x})$ on the atlas A. The correspondences are valid on both the "shape" images and the intensity images. We show the results using this local registration algorithm in Fig. (4.14).

4.4.3 Statistical Modeling of Shape and Appearance

By registration in the joint shape and intensity space, we are able to recover the deformation fields that establish correspondence between both training shapes and textures. We apply Principle Component Analysis (PCA) on the deformed FFD control lattices to capture variations in shape. The feature vectors are the coordinates of the FFD control lattice points in x and y directions in the common reference frame. We also use PCA on the registered object interior textures to capture variations in intensity. Here the feature vectors are the image pixel intensities from each registered texture.

Fig. (4.15) illustrates the mean atlas and three primary modes of variation for both the shape deformation fields (Fig. (4.15).1) and intensities (Fig. (4.15).2). The shape model shown uses the articulated heart model with epicardium and LV endocardium, and the texture model is for the LV interior texture only (due to the presence of tagging lines in heart walls and RV irregularity).

4.4.4 Coupled Prior based Segmentation

Given an unseen image, we perform segmentation by registering the learned prior model with the image based on both shape and texture.

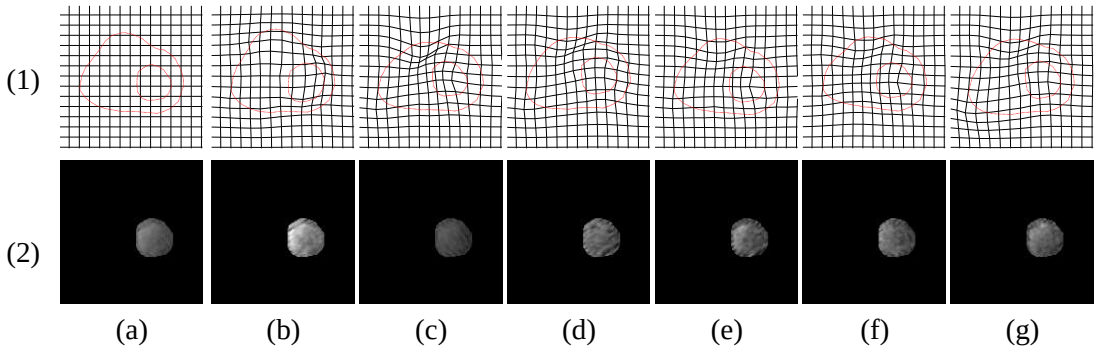


Figure 4.15: PCA modeling on the FFD control lattice deformations to capture variations in shape, and on registered textures to capture the variations in appearance. (1.a) The mean FFD control lattice configuration and mean shape. (1.b-c) Varying first mode of FFD deformations: -2σ reconstruction in (b) and 2σ in (c). (1.d-e) Second mode of FFD deformations. (1.f-g) Third mode of FFD deformations. (2.a) The mean LV texture (based on pixel-wise correspondences). (2.b-c) Varying first mode of LV texture. (2.d-e) Second mode of LV texture. (2.f-g) Third mode of LV texture.

In the image, we encode the gradient information of the image using a "shape image", which is derived from the un-signed distance transform of the edge map of the image. Then we register the learned prior shape and appearance model with the image in both shape and intensity spaces. The energy functional is the same as Equation 4.16, except that here B consists of the new intensity image and the derived "shape" image. Another difference from the learning process is that, during optimization, instead of using directly the recovered FFD parameter increments to deform the prior model, we back-project the parameter increments to the PCA-based feature space, and magnitudes of the allowed actual parameter changes are constrained to have a 2σ upper bound. This scheme is similar to that used in Active Shape and Appearance Models.

Using the statistical model learned as shown in Fig. 4.15, we conduct automated segmentation via statistically constrained registration in both shape and intensity spaces on two novel sequences of 4D spatial-temporal tagged MR images of the heart. Each of the 4D test sequence has 24 spatial locations and 18 time points, which gives us over 400 2D test images. Example segmentation results on the two datasets are shown in Fig. 4.16. During segmentation, the learned prior model is registered to the images based on the shape models of the epicardium and LV, and texture model of the LV. We also approximate the position and shape of the right ventricle, but we do not show the RV segmentation here since we did not learn a prior model for the RV.

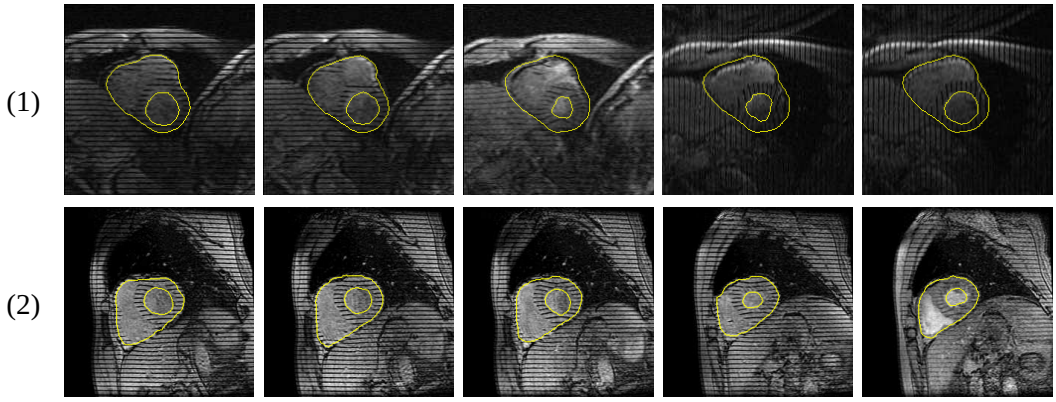


Figure 4.16: Coupled prior based segmentation results on two novel tagged MR image sequences. (1) Example segmentation results on novel sequence 1. (2) Example results on novel sequence 2.

Quantitative validation is performed by comparing the automated segmentation results with expert solutions. Denote the expert segmentation results in the images as ℓ_{true}, and the results from our method as ℓ_{prior}. We define the false negative fraction (FNF) to indicate the fraction of tissue that is included in the true segmentation but missed by our method: $FNF = \frac{|\ell_{true} - \ell_{prior}|}{|\ell_{true}|}$. The false positive fraction (FPF) indicates the amount of tissue falsely identified by our method as a fraction of the total amount of tissue in the true segmentation: $FPF = \frac{|\ell_{prior} - \ell_{true}|}{|\ell_{true}|}$. And the positive fraction (TPF) describes the fraction of the total amount of tissue in the true segmentation result that is overlapped with our method: $TPF = \frac{|\ell_{true} \cap \ell_{prior}|}{|\ell_{true}|}$. On the novel tagged MR sequence 1, our segmentation results produce the following average statistics: $FNF = 2.4\%, FPF = 5.1\%, TPF = 97.9\%$. On the novel sequence 2, the average statistics are: $FNF = 2.9\%, FPF = 5.5\%, TPF = 96.2\%$.

4.5 Summary

In this chapter we have proposed a novel, generic algorithm for learning coupled prior shape and appearance models. The main contributions of this chapter are three folds. First, we propose a new global-to-local shape registration algorithm that is able to establish continuous, smooth and one-to-one correspondences. The algorithm can be used to register training shapes and establish correspondences between them in order to learn a statistical shape model. Second, to learn coupled shape and appearance models, we propose to work in a unified shape and

intensity feature space. Third, a global-to-local registration algorithm based on the implicit shape representation, FFD and mutual information performs registration both between shapes and between textures simultaneously in order to acquire accurate correspondences to build a coupled shape and appearance model.

Chapter 5

Hybrid Image Registration based on Configural Matching of Scale-Invariant Salient Region Features

In previous chapters, shape and appearance information are integrated in unified energy-minimization frameworks and system parameters are solved through gradient-descent optimization. In this chapter we explore the integration of shape and intensity information on the other side of the spectrum. We introduce a novel method for aligning images under arbitrary poses, based on finding correspondences between image "region" features. Rather than using traditional geometric features such as curvature extreme points, curves/surface patches, our hybrid method detects salient "region" features, each of which has an associated scale and whose interior intensities (appearance) can be matched using robust similarity measures such as mutual information. Shape information is incorporated by considering geometric configuration constraints between the region features during correspondence finding. The geometric configuration constraints are enforced in an Expectation-Maximization framework to find a joint correspondence between multiple pairs of region features that result in a consistent transformation; other feature pairs, which either are outlier matches or degrade matching performance, are effectively pruned.

5.1 Introduction

Image registration aims to spatially align one image to another. For that purpose, parameters of a global transformation model, such as rigid, affine or projective, are to be recovered to geometrically transform a *moving* image to achieve high spatial correspondence with a *fixed* image. The problem has been studied in various contexts due to its significance in a wide range of areas, including medical image fusion, remote sensing, recognition, tracking, mosaicing, and so on.

Existing methods for image registration can largely be classified into three categories: feature-based methods, intensity-based methods, and hybrid methods that integrate the previous two. Traditional feature-based methods use sparse geometric features such as points [109], curves, and/or surface patches [13, 71], and their correspondences to compute an optimal transformation. These methods are relatively fast. However, the main critiques of this type of methods in the literature are the robustness of feature extraction, the accuracy of feature correspondences, and the frequent need of user interaction. Intensity-based registration methods [114, 22] operate directly on the intensity values from the full image content, without prior feature extraction. These methods have attracted much attention in recent years since they can be made fully automatic and can be used for multi-modality image matching by utilizing appropriate similarity measures. However, these methods tend to have high computational cost due to the need for optimization on complex, non-convex energy functions. In addition, they require the poses of two input images be close enough to converge to a local optimum. Furthermore, they often perform poorly when partial matching is required. Recently, several hybrid methods are proposed that integrate the merits of both feature-based and intensity-based methods [104, 49, 63]. Most of them focus on incorporating user provided or automatically extracted geometric feature constraints into the intensity-based energy functionals to achieve smoother and faster optimization.

Despite the vast efforts, however, several hard problems in registration still remain. First, dealing with structure appearing/dissappearing between two images is still challenging. For instance, tumor growth/shrinkage in medical images acquired in the clinical tracking of treatment, trees/shadows or construction in aerial images taken at different times, and occlusion in other natural images often lead to significant differences in local image appearance (see Figs. 5.1, 5.7). Second, it is still difficult to match images acquired by sensors of different modalities in general, since different sensors, such as MRI, CT or PET, may produce very dissimilar images of the same scene. The relationship between the intensities of the matching pixels is often complex and not known *a priori*. Image noise and intensity inhomogeneity also add to this complexity. Last, but not least, given two input images under arbitrary poses, recovering the globally optimal transformation efficiently is a hard problem due to the large parameter

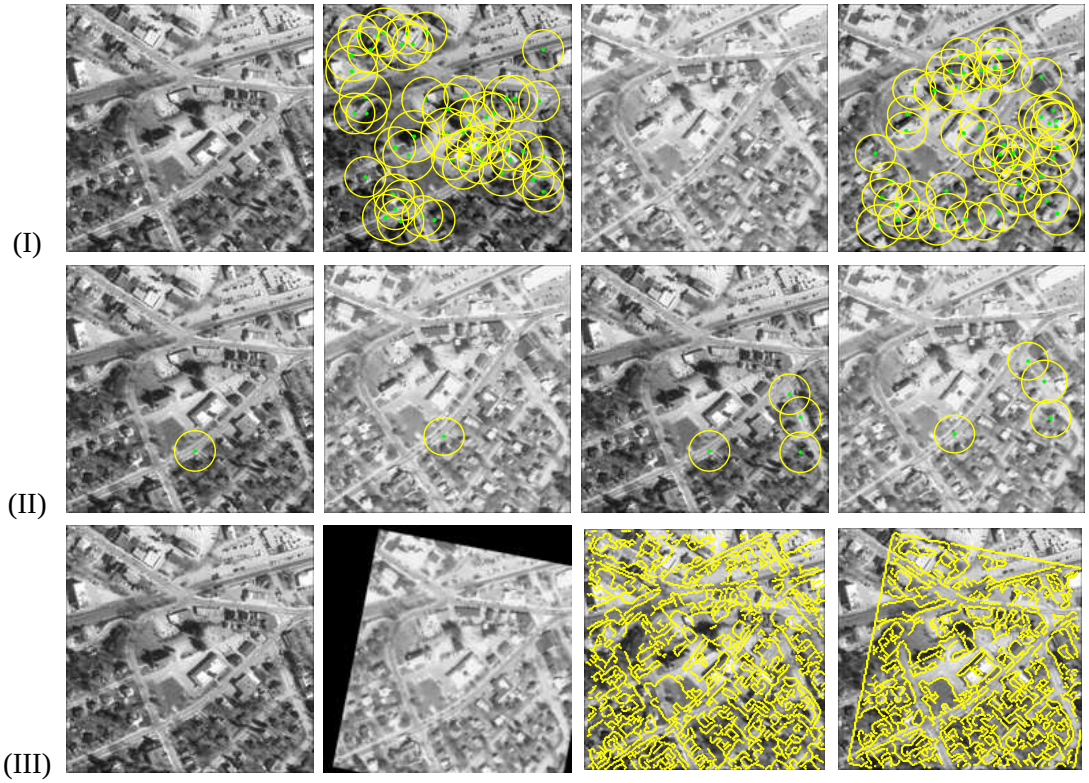


Figure 5.1: The registration method based on matching scale-invariant salient region features. (I.a) The fixed image I_f. (I.b) Salient region features (shown as yellow circles) detected on I_f. (I.c) The moving image I_m. (I.d) Salient region features detected on I_m. (II.a-b) The first corresponding feature pair chosen. (II.c-d) The corresponding feature pairs chosen by the algorithm upon convergence. (III.a-b) Registration result: (III.a) the fixed image I_f, and (III.b) the transformed moving image I_t based on the transformation parameters recovered using the chosen feature correspondences. (III.c-d) Comparison of the edge superimposed maps: (III.c) edges (in yellow) from the original moving image I_m superimposed on fixed image I_f, and (III.d) edges from the transformed moving image I_t superimposed on fixed image I_f.

search space. To tackle these problems, the integration of both feature-based and intensity-based methods is very attractive since they are of complementary nature. While intensity-based methods are superior in multi-modal image matching and have better robustness to image noise and inhomogeneity, the feature-based methods are more natural to handle the structure appearing/dissappearing problem, occlusion, and partial matching as well as to align images despite of their initial poses.

In this chapter, we propose a new hybrid image registration method that integrates shape and intensity information by matching a small number of scale-invariant salient region features. Rather than using traditional geometric features such as curvature extrema points, curves/surface

patches, the image alignment in our approach is driven directly by image intensities within automatically extracted salient regions. The overall approach is depicted in Fig. 5.1. First, on both the fixed and moving images, salient region features are selected, using an entropy-based detector, as those areas (each associated with a best scale) with the highest local saliency in both spatial and scale spaces (see Fig. 5.1, I.a-d). Then a *region component matching* (RCPM) step is used to determine the likelihood of each hypothesized fixed-moving pairing of two region features. The likelihood of each pairing is measured by the normalized mutual information between the two regions. The result of this step is a total ordering of the likelihoods of all hypotheses about individual feature matches. Due to image noise or intensity changes, the top matches from this result often contain an unpredictable portion of outliers (i.e., mismatches), whose effects can only be partially alleviated by the use of robust estimation techniques. In the literature, the global one-to-one correspondence constraint [7, 19] has been widely used. However, in the presence of unmatchable features or in the situation of partial matching, this global constraint is neither sufficient nor valid. To address these limitations, we emphasize the importance of the geometric configural constraints in preserving the global consistency of individual matches. Utilizing the top individual feature correspondence candidates from the RCPM step, we further design a *region configural matching* (RCFM) step in which we detect a joint correspondence between multiple pairs of salient region features (see Fig. 5.1, II.c-d). The strict geometric constraints imposed by the joint correspondence make the algorithm very effective in pruning false feature matches. The combinatorial complexity associated with detecting joint correspondences is addressed in an efficient manner by using one feature pair correspondence as a minimal base (see Fig. 5.1, II.a-b), then incrementally add to the base new feature pairs using an Expectation-Maximization algorithm. The likelihood of each hypothesized joint correspondence is always measured based on the global "alignedness" between the fixed image and the transformed moving image, given the transformation computed from the hypothesis. This allows convergence to the globally optimal transformation parameters. Various experiments on registering aerial images and medical images of single and multiple modalities demonstrate the effectiveness of the proposed method both quantitatively and qualitatively.

5.1.1 Previous Work

The proposed image registration method is largely inspired by the pioneering works from the object recognition literature [40, 47, 61]. From their works, we learned two important aspects that would be beneficial when used in image registration. The first aspect is the use of scale-invariant *region* features. In [40], objects are modeled as flexible constellations of regions (parts) in order to learn and recognize object class models. An entropy-based feature detector [61] is used to select region features that have complex intensity distributions and are stable in both spatial and scale spaces. When adapting the idea of region features to solve image registration problems, suitable and robust similarity measures need to be defined between region intensity values, to deal with multi-modal matching, image noise, and intensity inhomogeneity. The second aspect is the importance of geometric configural constraints in robust feature matching. In [47], the role of geometric constraints in object recognition is studied in depth using edge and other geometric features, and an *interpretation tree* (IT) algorithm is developed to search for globally consistent feature correspondences. In this chapter, we present a new method of implementing the geometric configural matching. Compared to the interpretation tree search algorithms whose best-case and worst-case complexities can be significantly different, our method has a very predictable low computational cost and has the best-case and worst-case complexities on the same order.

The remainder of the chapter is organized as follows. In section 5.2, we describe the salient region feature detector. In section 5.3, we present our region feature based image registration algorithm. Experimental results on both aerial and medical images are demonstrated in section 5.4. We summarize with discussion in section 5.5.

5.2 Scale-Invariant Salient Region Features

The line of research on feature-based image matching has long been restrained by the question: what features to use? An interesting feature selection criterion was proposed for tracking under occlusion and disocclusion situations in [105]. The criterion states that the right features for

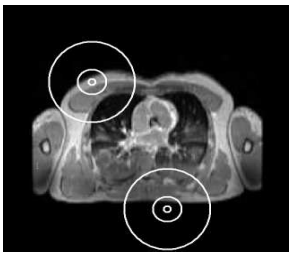


Figure 5.2: Demonstrating our belief that every point in the image can be made unique if a proper scale of its neighborhood is selected to calculate the feature. (Inner most circle) Locally at a small scale, the point neighborhood appear homogeneous. (Middle circle) At a larger scale, the point neighborhood begins to appear unique. (Large Circle) At a scale that is large enough, every point appears unique based on the characteristics of its neighborhood.

tracking are exactly those that make the tracker work best. Applying similar reasoning, we believe that good features for image registration should be those that are "unique" or "rare". The uniqueness or rarity of a feature we refer here is in the context of correspondence, i.e., given a feature from one image, whether the likelihood of having multiple corresponding features on the matching image is low, not in the context of its uniqueness or rarity in occurrence within the same image. For example, in the use of point features for image matching, the traditional intuition and argument is that pixels in homogeneous regions (similarly, points with low curvatures on curve or surface segments) tend to be ambiguous in correspondence and should either not be chosen as the preferred feature point or weighted less important during the matching process [110, 3]. We argue, however, that these popular beliefs are only correct in a relative sense and that the "uniqueness" of a feature is closely related to its associated scale. At a smaller scale, edge points, corner points, or points with high curvature appear to be more unique than others. At a medium or larger scale, points in homogeneous regions or with low curvature begin to appear unique as well. Medial axis points of a shape or a homogeneous region are examples of these type of points that are unique at the scale they are associated with. We believe that every point regardless of their local characteristics (edgeness, cornerness, medialness, curvature, etc.) in the image can be made unique if a proper scale and its neighborhood is selected to calculate the feature[1]. One pictorial example of this point of view is demonstrated in Fig. 5.2.

Thus motivated, we seek to use scale-invariant region features as the basis for our proposed

[1]Note that we are not the first to exploit this observation for image registration. For instance, in [104], promising results have been obtained recently for non-rigid brain image registration using an attribute vector of geometric moment invariants at different scales.

registration method. In [61], a salient region feature detector is proposed. The salient regions are found using an entropy-based detector, which aims to select regions with highest local saliency in both spatial and scale spaces. For each pixel $\mathbf{x}$ on an image, a probability density function (PDF) $p(s, \mathbf{x})$ is computed from the intensities in a circular region of certain scale described by a radius s centered at $\mathbf{x}$. The local differential entropy of the region is defined by:

$$\mathcal{H}(s, \mathbf{x}) = -\int_{\mathcal{R}} p_i(s, \mathbf{x}) \log_2 p_i(s, \mathbf{x}) di$$

where i takes on values in the set of possible intensity values. The best scale $S_{\mathbf{x}}$ for the region centered at $\mathbf{x}$ is selected as the one that maximizes the local entropy: $S_{\mathbf{x}} = \text{argmax}_s \mathcal{H}(s, \mathbf{x})$. Then the saliency value, $\mathcal{A}(S_{\mathbf{x}}, \mathbf{x})$, for the region with the best scale is defined by the extrema entropy value, weighted by the best scale and a differential self-similarity measure in the scale space:

$$\mathcal{A}(S_{\mathbf{x}}, \mathbf{x}) = \mathcal{H}(S_{\mathbf{x}}, \mathbf{x}) \cdot S_{\mathbf{x}} \cdot \int_{\mathcal{R}} \left\| \frac{\partial}{\partial s} p_i(s, \mathbf{x})\big|_{S_{\mathbf{x}}} \right\| di$$

Since the saliency metric is applicable over both spatial and scale spaces, the saliency values of region features at different locations and scales are comparable.

For the proposed registration method, we apply the following steps to pick a low number N ($N < 100$ for all our experiments) of salient region features (each defined by its center and the best scale):

- For each pixel location $\mathbf{x}$, compute the best scale $S_{\mathbf{x}}$ of the region centered at it, and its saliency value $\mathcal{A}(S_{\mathbf{x}}, \mathbf{x})$.
- Identify the pixels with local maxima in saliency values. Then the salient regions of interest are those that are centered at these pixels and have the best scales.
- Among the local maxima salient regions, pick the N most salient ones as region features for the image.

One of the main advantages of the salient region features is that they are theoretically invariant to rotation, translation and scale. We also quantitatively validate the invariance properties in section 5.4.1. Some examples on the extracted salient regions are shown in Fig. 5.1(I.a-d) and

in Fig. 5.4(II.a-d).

5.3 The Salient Region based Registration Algorithm

Once we have extracted the salient region features from both the fixed and moving images, the alignment of the two images is achieved by finding a robust joint correspondence between multiple pairs of region features. This joint correspondence is then used to estimate the parameters of a desired transformation model. In this chapter, we consider the 2D similarity transformation. This transformation can be described by four parameters: $(t_x, t_y, \sigma, \theta)$, where t_x, t_y are the translation along x and y directions respectively, σ is the isotropic scaling factor, and θ is the rotation angle.

Several notations are introduced as follows:

- I_f is the fixed image, I_m is the moving image, and I_t is the transformed moving image. We aim to recover the parameters of a similarity transformation that geometrically transforms the moving image to be aligned with the fixed image.
- Suppose N_f salient region features are detected on I_f, and N_m features on I_m.
- $C_{i,j}$ denotes the hypothesized correspondence between the ith region feature on I_f and the jth feature on I_m. Here $(i, j) \in [1, N_f] \times [1, N_m]$.
- $C_{i_1,j_1} \cap C_{i_2,j_2} \cap ... \cap C_{i_k,j_k} ...$ denotes a hypothesized joint correspondence between multiple region feature pairs: i_1th region on I_f corresponds to j_1th region on I_m, i_kth region on I_f corresponds to j_kth region on I_m, etc.

5.3.1 Region Component Matching (RCPM)

In the RCPM step, we measure the likelihood of each hypothesized correspondence between a region feature from I_f and a region feature from I_m, respectively. That is to say, we want to measure the likelihood $\mathcal{L}_{local}(C_{i,j})$ for each individual feature correspondence hypothesis $C_{i,j}$. We can then acquire a total ordering of these hypotheses according to their likelihoods.

We define the likelihood to be proportional to the similarity between the interior intensities of the two salient regions involved. Let us denote the ith region on I_f as A, and the jth region

on I_m as B. Before measuring their intensity similarity, we first normalize their scales by supersampling (using bicubic interpolation) the smaller region to match the scale of the larger region. This also leads to scale-invariant matching. The translation invariance is intrinsic by aligning the two region centers. To further achieve rotation invariance, we sample the parameter space for rotation sparsely [2], and use the largest similarity value over all possible angles as the similarity between the two regions. The similarity measure we use is a normalized form of mutual information, the Entropy Correlation Coefficient (ECC) [68]. Such metric has been proven robust in the literature in dealing with multi-modal image matching, image noise and intensity inhomogeneity.

Formally, the likelihood of a correspondence hypothesis $C_{i,j}$ is defined as:

$$\mathcal{L}_{local}(C_{i,j}) = \max_{\theta} ECC(A, B^{\theta})$$

where B^{θ} is the scale-normalized region B after rotating angle θ. The Entropy Correlation Coefficient (ECC) between the two regions is defined by:

$$ECC(A, B^{\theta}) = 2 - \frac{2\mathcal{H}(A, B^{\theta})}{\mathcal{H}(A) + \mathcal{H}(B^{\theta})}$$

where $\mathcal{H}$ indicates the joint or marginal differential entropy of the intensity value random variables of the two regions. Given two inputs u and v, the value of $ECC(u, v)$ has the following properties: $ECC(u, v)$ is scaled to $(0, 1)$, such that 0 indicates full independence and 1 complete dependence between the two inputs. Furthermore, $ECC(u, v)$ increases almost linearly when the relationship between u and v varies from full independence to complete dependence, which makes it an attractive measure of the likelihood that u corresponds to v.

Using this ECC definition, the likelihood values of all feature correspondence hypotheses $C_{i,j}$, where $(i, j) \in [1, N_f] \times [1, N_m]$, are comparable regardless of the scales of the region features. Thus we are able to sort these hypotheses in the order of descending likelihood. We then choose the top M such hypotheses to be used in the next configural matching step to extract a globally consistent joint correspondence. Here we make the assumption that there will

[2] Typically, the rotation angles are sampled uniformly between $[-\pi, \pi)$ at an interval $\pi/36$.

be at least 2 valid feature correspondences among the top M candidates. From our extensive experiments, we found this assumption to be fairly reasonable with typical values of M between $20 \sim 40$.

The RCPM step also generates useful information regarding the transformation to align the two images, based on the purely local region-based matching. For instance, given a high likelihood correspondence between a region A on I_f and a region B on I_m, we can estimate the scaling factor by: $\sigma = \frac{s_A}{s_B}$, where s_A and s_B are the scales of the two regions respectively. The rotation angle can be estimated as: $\theta = \text{argmax}_{\theta'} ECC(A, B^{\theta'})$. And the translation can also be estimated by the displacement between the center of the region A and the center of the region B after rotation and scaling. These estimates are associated with the related feature correspondence hypothesis, to provide the initial estimate for the transformation in the next region configural matching step.

As a result of the RCPM step, we have a total ordering of the individual feature correspondence hypotheses. In addition, based on the top M hypothesis $C_{i,j}$, we have a transformation parameter estimate: $(t_x, t_y, \sigma, \theta)^{C_{i,j}}$. As an example, we show the top 5 region feature correspondence hypotheses for the pair of aerial images in Fig. 5.3. From the results, one can see that the RCPM step is able to extract good individual feature matches based on local region intensity-based matching. In the next configural matching step, we will demonstrate the use of geometric constraints to pick out the true correspondences (e.g., Fig. 5.3, I-III, V) and prune the outliers (e.g., Fig. 5.3, IV).

Note that the RCPM step has the most complexity of our entire registration method, since it has $N_f \times N_m$ hypothesis testings, and a total ordering of their likelihoods are pursued. However, because the number of region features, N_f and N_m, are low, the algorithm is still computationally efficient.

5.3.2 Region Configural Matching (RCFM)

In the RCFM step, we aim to detect a joint correspondence $C_{i_1,j_1} \cap C_{i_2,j_2} \cap ... \cap C_{i_k,j_k} ...$ between multiple pairs of region features, which results in the maximum likelihood in terms of global image "alignedness". The intuition behind the configural matching is that, while false matches are very likely to arise when we search for individual local feature correspondences,

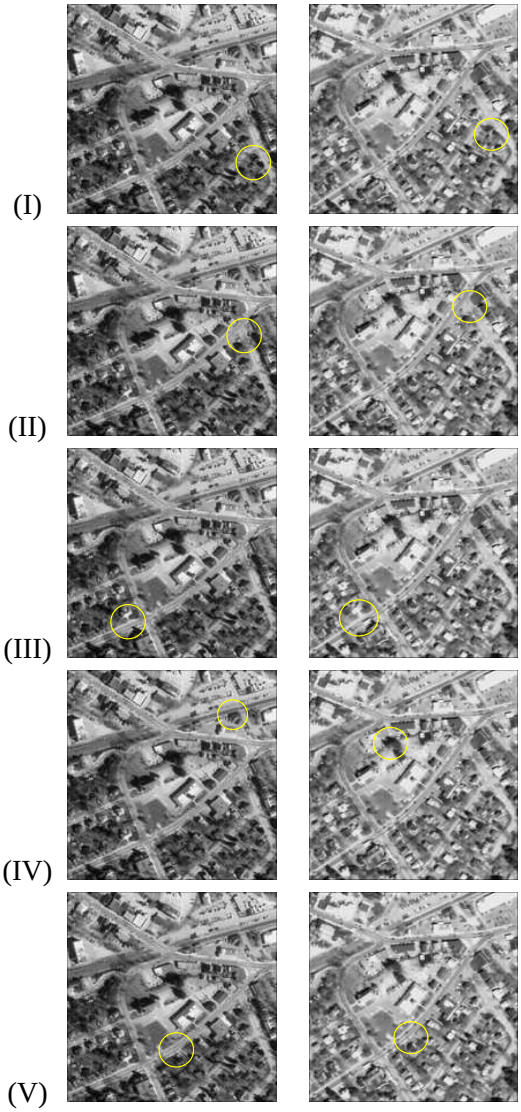


Figure 5.3: The top five candidate region feature correspondences computed by the region component matching (RCPM) step. The result is shown for the pair of aerial images in Fig. 5.1.

the likelihood of a global geometrically consistent joint correspondence between multiple feature pairs being false is very low due to the strict geometric configuration constraints imposed by the joint correspondence.

We measure the likelihood of a hypothesized joint correspondence with n feature pairs using the ECC measure between the overlapping portions of the fixed image I_f and the transformed moving image $T_n(I_m)$. Here the transformation $I_t = T_n$ is estimated from all feature

pairs contained by current hypothesis. This can be written as:

$$\mathcal{L}_{global}(C_{i_1,j_1} \cap C_{i_2,j_2} \cap ... \cap C_{i_n,j_n}) = ECC(T_n(I_m), I_f), \\ (i_k, j_k) \in [1, N_f] \times [1, N_m], k = 1, ..., n \tag{5.1}$$

This likelihood measures the global image "alignedness" under current hypothesis. In the end, we want to find a joint correspondence that has the maximum likelihood, while containing adequate number of feature pairs (typically a few) to recover the parameters of a similarity transformation.

To address the combinatorial complexity in detecting the joint correspondence, we first compute a minimal correspondence base of l feature pairs and get an initial estimation of the transformation. As shown in section 5.3.1, one correspondence between a pair of region features is sufficient to derive a transformation estimate, i.e., $l = 1$. To choose this first correspondence, we measure $\mathcal{L}_{global}(C_{i,j})$ for each individual feature match among the top M hypothesized correspondences resulted from the RCPM step. Using Equation 5.1, the parameters of T_l are $(t_x, t_y, \sigma, \theta)^{C_{i,j}}$ when measuring the likelihood of $C_{i,j}$. Then the first feature pair in the minimal correspondence base is the correspondence yielding the maximum likelihood, i.e.,

$$C_{i_1,j_1} = \underset{C_{i,j}}{\operatorname{argmax}}\, \mathcal{L}_{global}(C_{i,j})$$

To allow converging to a globally optimal solution, we further use a generalized Expectation-Maximization (EM) algorithm to incrementally add in new feature pairs to the joint correspondence base, while refining the center locations of the corresponding features. The generalized EM algorithm is described as follows:

1. Let current joint correspondence be $C = (C_{i_1,j_1} \cap ...C_{i_l,j_l})$. Locally refine the region feature centers in C in sub-pixel accuracy to achieve better matching, and use the refined corresponding region centers to estimate a current transformation T.

2. **E-step:** For each feature pair $C_{i,j}$ that is in the top M individual matches, but not in the current joint correspondence C, estimate the likelihood of this feature pair being a valid correspondence in terms of global consistency as $\mathcal{L}_{global}(C \cap C_{i,j})$, $C_{i,j} \notin C$.

3. **M-step:** Choose the new feature correspondence $C_{\hat{i},\hat{j}}$ that has the maximum likelihood. We also require the addition of $C_{\hat{i},\hat{j}}$ increasing the global image "alignedness".

   If $\mathcal{L}_{global}(C \cap C_{\hat{i},\hat{j}}) > \mathcal{L}_{global}(C)$

   Then

   a Let the new joint correspondence be $C = (C \cap C_{\hat{i},\hat{j}})$.

   b Locally refine the centers of the region features in the joint correspondence in sub-pixel accuracy to achieve better matching.

   c Re-compute the transformation T using the new joint correspondence.

   d Repeat EM steps 2-3.

   Else Output current transformation T as the converged transformation to align the fixed image I_f and the moving image I_m.

For the aerial image example, the feature pair shown in Fig. 5.1(II.a-b) is chosen to be the first feature pair in the minimal correspondence base. Note that this first pair in the RCFM step is not necessarily the same as the top feature pair resulted from the RCPM step, since different criteria are used to determine the likelihoods for ranking purposes. In fact, the first pair chosen by the RCFM step in Fig. 5.1(II.a-b) is the 5th feature pair in the RCPM step (see Fig. 5.3, V), because the transformation estimated from this feature pair gives rise to the maximum global image "alignedness". All the feature pairs in the final converged joint correspondence are shown in Fig. 5.1(II.c-d). Based on these correspondences, a similarity transformation is recovered and the registered image pair (i.e., the fixed image and the transformed moving image) is shown in Fig. 5.1(III.a-b).

Having at most M iterations, our RCFM step is very efficient. Two key points contribute to this efficiency: First, pick a minimal correspondence base with only one feature pair; Second, use the EM algorithm to add in new feature pairs incrementally, thus enabling the converged joint correspondence to include as many good feature pairs as possible, while keeping a minimal complexity.

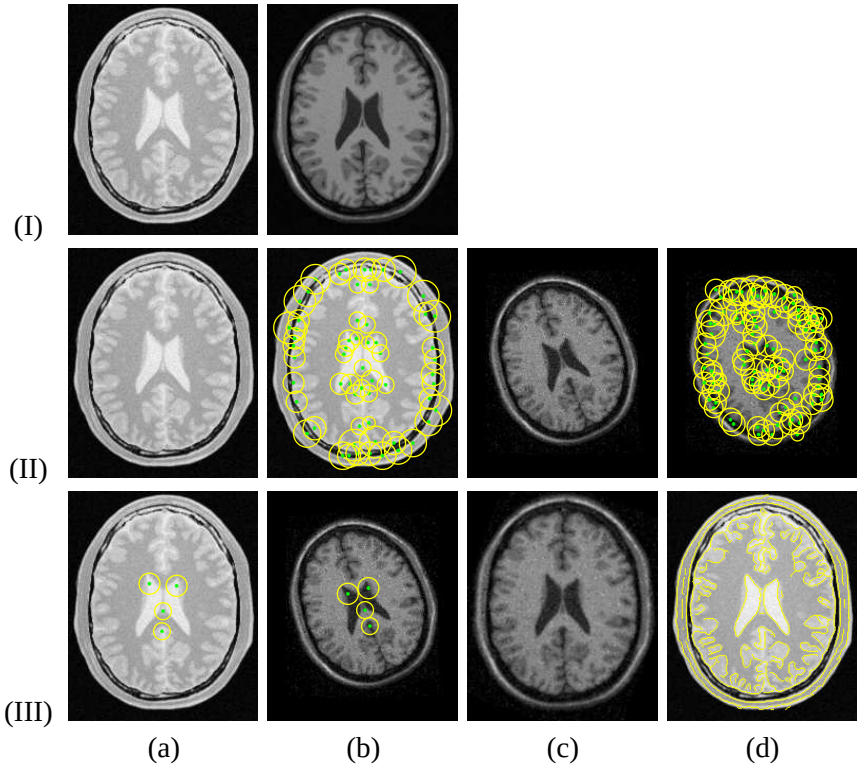


Figure 5.4: Registration on the pair of brain images used in the simulation experiment. (I.a) Original PD-weighted MR brain image. (I.b) Original T1-weighted MR brain image. (II.a) The fixed image I_f. (II.b) Salient region features on I_f. (II.c) The moving image I_m. (II.d) Salient region features on I_m. (III.a-b) The feature pairs in the joint correspondence chosen by the algorithm upon convergence. (III.c) The transformed moving image I_t. (III.d) The edge superimposed map after registration: edges from I_t (in red) superimposed on fixed image I_f.

5.4 Experiments

In this section, we present both the quantitative and the qualitative results of applying our image registration method on several simulated and real images.

5.4.1 Quantitative Results on Simulated Moving Images

In order to quantitatively validate the robustness, accuracy, and efficiency of the proposed method, we conduct a series of controlled experiments using a pair of brain images with the moving image simulated from a known transform. The first image is a PD (proton density) weighted MR brain image (see Fig. 5.4, I.a), and the second image is a T1 weighted MR brain image (see Fig. 5.4, I.b). The two images are originally registered, and the size of the images is 217×181.

In our first controlled experiment, we study the invariance properties of our method to scaling, rotation, and translation. We use the PD image as the fixed image, then simulate different moving images by artificially transform the T1 image with controlled parameters. The parameters are chosen according to the following four cases:

1. Case 1 studies the invariance to scaling. To this end, we fix the translation ($t_x = 0, t_y = 0$) and rotation ($\theta = 0$), but vary the scale factor σ in the range $[0.5, 1.5]$.

2. Case 2 studies the invariance to rotation. We fix the translation ($t_x = 0, t_y = 0$) and scaling factor ($\sigma = 1$), but vary the rotation angle in the range $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

3. Case 3 studies the invariance to translation. Here only the translation parameters t_x, t_y are varied in the range $[-50, 50]$.

4. Case 4 studies the combined effect of the transformation parameters by varying all parameters simultaneously: t_x, t_y in the range $[-50, 50]$, σ in the range $[0.5, 1.5]$, and θ in the range $[-\frac{\pi}{2}, \frac{\pi}{2}]$.

In each case, we generate 50 simulated moving images. Then we apply our registration algorithm to register the fixed image with each simulated moving image respectively. Since we know the ground truth transformation that was used to simulate each moving image, we can compare these ground truth with the recovered transformation parameters by our method. Three statistical performance measures are computed from the study and the results are listed in Table 5.1. The first measure is the *percentage of correctness* (correctness). In a registration trial, if the recovered transformation is sufficiently close to the ground truth [3], this trial results in a correct registration, otherwise, it is taken as a false registration case. The second measure is the *average error* (error). This measure gives the average error (i.e., difference) of the recovered transformation parameters from the ground truth. It reflects the accuracy and convergence property of our registration method. The last measure is the *average execution time* (time) for one trial of registering a pair of fixed and moving images. Note that our method is currently

[3]We consider the recovered transformation correct if its difference from the ground truth is less than a predefined error threshold. Typically, we set the threshold as follows: scale error less than 0.05, rotation angle error less than 5 degrees, translation error in x direction less than $D_x/50$, and translation error in y direction less than $D_y/50$, where D_x, D_y are the dimensions of the image along x and y directions, respectively.

	correctness	error	time
Case 1	98%	(0.9, 1.1, 0.027, 0.0)	138 s
Case 2	100%	(0.5, 0.6, 0.009, 1.5)	155 s
Case 3	100%	(0.2, 0.4, 0.000, 0.0)	155 s
Case 4	94%	(1.4, 1.7, 0.031, 2.1)	150 s

Table 5.1: Quantitative validation of the invariance properties of the method. For each case, the percentage of correct registration (correctness), the average error in recovered transformation parameters (error), and the average execution time for one trial (time) are given. The given errors are in the format: $(t_x, t_y, \sigma, \theta)$, where translation errors t_x and t_y are in pixels, rotation angle errors are in degrees, and the scaling errors are given relative to the original image scale. The times are given in seconds.

implemented in Matlab with several functions written in C++ and that all the experiments are conducted on a 2GHz PC workstation.

range of λ	correctness	error	time
[5, 10]	100%	(0.3, 0.6, 0.007, 0.4)	142 s
[10, 20]	97%	(0.7, 0.9, 0.006, 1.2)	142 s
[20, 30]	90%	(0.9, 1.3, 0.009, 2.4)	144 s

Table 5.2: Quantitative simulation study of the performance of the method when images are corrupted by different levels of Gaussian noise. Three different cases are shown in three rows. The cases differ by the range of the standard deviation λ of the Gaussian noise added. For each case, three statistical measures are given in the same format as in Table 5.1.

In the second controlled experiment, we study the robustness of the method to image noise. We use the original PD image as the fixed image, then generate test moving images by adding different levels of Gaussian noise to the original T1 image, and transforming the noise corrupted images according to random transformations. The Gaussian noise we add has zero mean with standard deviation λ. In Table 5.2, we show the three performance measures for three test cases. The three cases differ by the range of the standard deviation of the Gaussian noise added. (All possible values for the standard deviation are between $[0, 255]$). For each case, 30 noise corrupted T1 images are generated and randomly transformed, where the transformation parameters vary in the same ranges as in the first controlled experiment. From the results, one can see that, the method is quite robust to high levels of noise. This is partly due to the stability of the entropy-based region feature detector and the robustness of the intensity-based Entropy Correlation Coefficient (ECC) similarity measure. It is also due to the fact that our algorithm

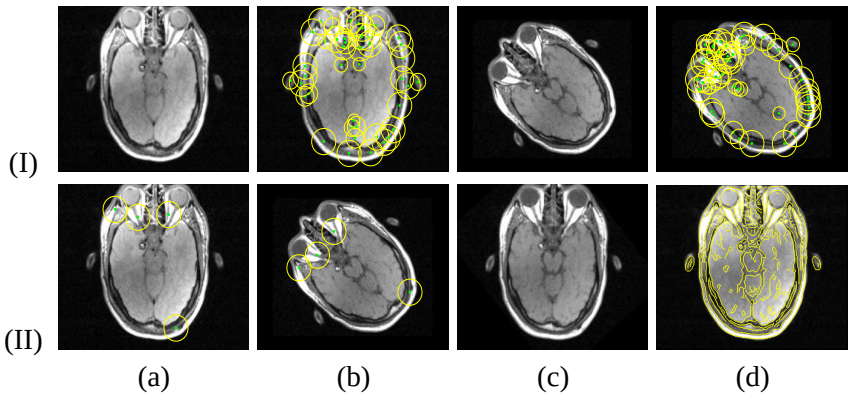


Figure 5.5: Registering a pair of real brain images from the Vanderbilt Database. (I.a) The fixed image. (I.b) Salient region features detected on the fixed image. (I.c). The moving image. (I.d) Salient region features on the moving image. (II.a-b) The corresponding feature pairs chosen by the algorithm upon convergence. (II.c) The transformed moving image. (II.d) The edge superimposed map after registration: edges (in yellow) from the transformed moving image superimposed on the fixed image.

requires only a small number of good matched features to register the images. One pictorial example selected among all simulated experiments is shown in Fig. 5.4(II-III). In this example, the moving image I_m (see Fig. 5.4, II.c) is generated by adding Gaussian noise with zero mean, standard deviation 25 to the original T1-weighted image, then scaling down the image by 20%, and rotating by 20 degrees.

5.4.2 Qualitative Results on Real images

Experiments with the simulated moving images in the previous section provide a quantitative study on the performance of our registration method. Real world images often have significant levels of noise and intensity inhomogeneity. Furthermore, between the pair of images to be registered, structures may appear or disappear, and intensities for the same structure may change. We have shown the result of our algorithm on a pair of real aerial images in Fig. 5.1. In this section, we apply our method to more real world medical images from several domains. These results demonstrate the effectiveness of our method on image registration problems that could be difficult to be solved using either pure intensity-based or pure feature-based methods.

Figure 5.5 shows the result of registering two real brain images. This pair of images is from the Vanderbilt Database [117]. Note that the algorithm successfully picks up several distinctive region features, and is able to recover the large rotation between the two images.

Another example on registering two MR chest images is shown in Fig. 5.6. This pair

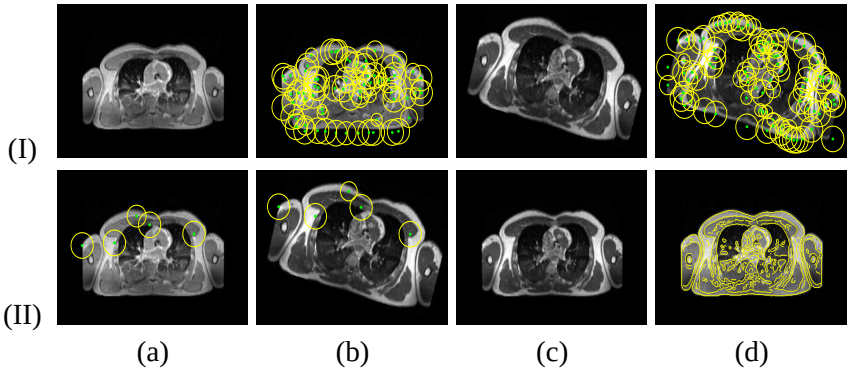


Figure 5.6: Registering a pair of chest MR images from the Visible Human project database. The layout of the images is the same as those in Fig. 5.5.

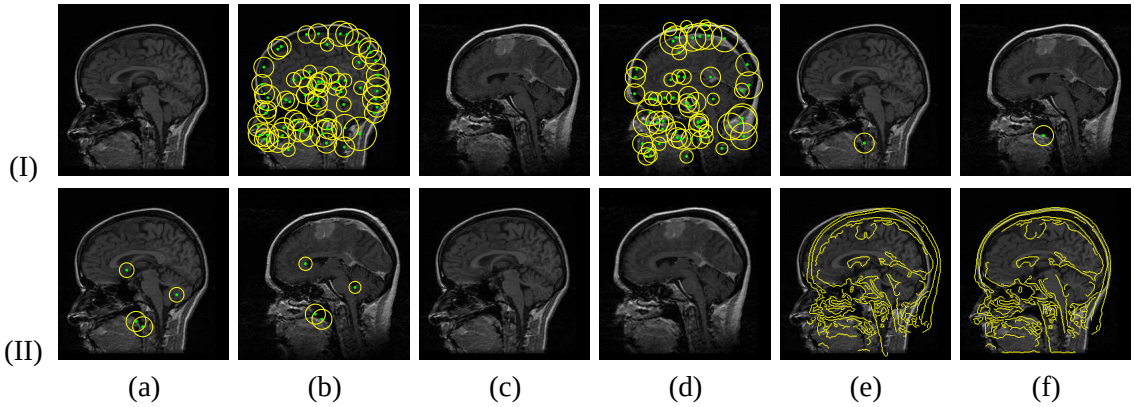


Figure 5.7: Registering brain images with tumor. (I.a) The fixed image. (I.b) Salient region features detected on the fixed image. (I.c). The moving image. (I.d) Salient region features on the moving image. (I.e-f) The first corresponding feature pair chosen. (II.a-b) The corresponding feature pairs chosen by the algorithm upon convergence. (II.c-d) The registration result: (II.c) the fixed image, and (II.d) the transformed moving image. (II.e-f) Comparison of the edge superimposed maps: (II.e) edges from the original moving image superimposed on the fixed image, and (II.f) edges from the transformed moving image superimposed on the fixed image.

of images is from the Visible Human Project database. The fixed image is a T1-weighted MR image, and the moving image is a PD-weighted MR image. Despite the different tissue intensity characteristics between the two images, the salient region feature pairs chosen by the method to recover the transformation parameters correspond very well both in scale and location (see Fig. 5.6, II.a-b).

To demonstrate the performance of our algorithm on images with appearing and disappearing structures, we use a pair of brain images, with one of which contains a tumor. The two images are from two different subjects, and the tumor in one of the images changes its appearance significantly. The results produced by our method are shown in Fig. 5.7. Here

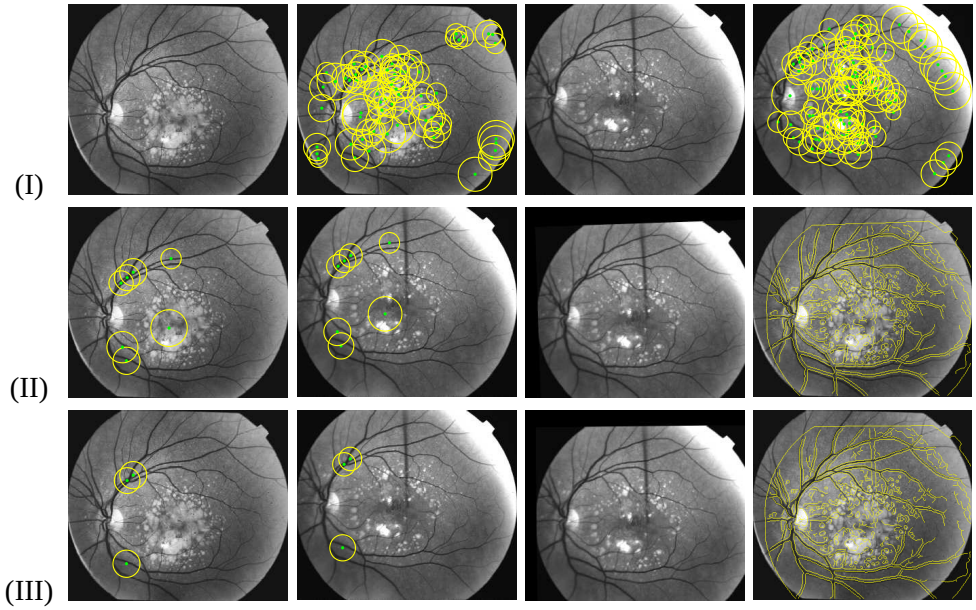


Figure 5.8: Registering two curved human retinal images. (I.a) The fixed image. (I.b) Salient region features on the fixed image. (I.c). The moving image. (I.d) Salient region features on the moving image. (II.a-b) The hand picked feature pairs that seem to correspond well. (II.c) The transformed moving image using the seven hand-picked feature correspondences. (II.d) Edges of the transformed moving image (in yellow) superimposed on the fixed image. (III.a-b) The corresponding feature pairs automatically chosen by the algorithm upon convergence. (III.c) The transformed moving image. (III.d) Edges of the transformed moving image superimposed on the fixed image.

the feature-based aspect of our algorithm enables it to focus on regions of similar appearance within a natural scale, thus being robust to the appearance and disappearance of local structures.

Last, but not least, we show the effectiveness of the proposed method on robust partial matching and mosaicing applications. We use a pair of curved human retinal images, as in [13]. The results are shown in Fig. 5.8. In this experiment, we also demonstrate the importance of the EM procedure in incrementally selecting good feature correspondences that increase the matching similarity and guaranteeing convergence. In Fig. 5.8, row II, we handpicked the feature pairs that seem to correspond to each other well. This results in seven feature pairs, and we transform the moving image using the transformation recovered by these feature pairs (see Fig. 5.8, II.c-e). In the last row III of Fig. 5.8, we show the feature correspondences automatically chosen by the method. There are only three best feature pairs chosen, and the transformation result can be seen in Fig. 5.8(III.c-e). Comparing the edge superimposed map in Fig. 5.8(II.d) and that in Fig. 5.8(III.d), one can see that the three feature pairs chosen by the method in fact produce better transformation than using all the seven handpicked feature

pairs. The comparison can be seen more clearly from the two zoom-in views of the edge superimposed maps: Fig. 5.8(II.e) *vs.* Fig. 5.8(III.e).

5.5 Discussion and Summary

In our current implementation of the geometric configural constraints, it is worth noting that the measure for the "goodness" of a candidate feature correspondence is based on its likelihood value and whether its addition will increase the global image "alignedness". On one hand, this permits us to efficiently recover the few best feature correspondences and to detect a convergence without explicitly setting hard thresholds. On the other hand, the strictness of the constraint also eliminates feature pairs that essentially correspond to each other individually (e.g. some handpicked good feature pairs in Fig. 5.8, row II, which are not chosen by the algorithm), but could deteriorate the overall global image alignment once added to the joint correspondence.

To summarize, in this chapter we have presented a novel image registration method based on the region component and configural matching using scale-invariant salient region features. The proposed method possesses characteristics of both feature-based and intensity-based methods. While the overall framework is based on finding correspondences between features, all the feature correspondence likelihoods and decisions are made according to intensity-based similarity measures between region features and images. The method is efficient in that it recovers a transformation using sparse salient region feature correspondences. It is also very robust because it exploits strict global geometric constraints when finding a joint correspondence between multiple feature pairs. The algorithm can be extended to deal with more complicated transformation models such as affine, projective transformations as well as non-rigid deformations in both 2D and 3D.

Chapter 6

Applications in Medical Image Analysis, Computer Vision, and Graphics

The deformable models, segmentation, registration and visual learning algorithms introduced in previous chapters can be applied to a wide range of real-world applications. In this chapter, we present two such applications. In the first application, the Metamorphs deformable models introduced in Chapter 3 are applied to heart wall segmentation and motion tracking in noisy, tagged MRI images of the heart. In the second application, we extend the Global-to-local shape registration algorithm introduced in Chapter 4 to 3D and use it for high-resolution facial expression tracking. The dense correspondences established by the tracking and registration algorithms are then applied to facial expression learning, synthesis and re-targeting.

6.1 Metamorphs Deformable Models and Tunable Gabor Filters for Robust Segmentation of 4D Cardiac Tagged MRI Images

Cardiovascular diseases are the main cause of death in the western countries. Many heart diseases, such as ischemia and RV hypertrophy are thought to correlate strongly to the shape and motion of the heart. Tagged cardiac magnetic resonance imaging (MRI) is a well-known technique for non-invasively imaging the regional cardiac function by enabling visualization of detailed myocardium motion and deformation. It has great potential to help early diagnosis and analysis of many cardiovascular diseases.

One example of tagged MRI imaging is the Spatially non-selective Spatial Modulation of Magnetization (SPAMM) technique [5, 4]. In SPAMM, a set of equally spaced parallel tagging planes are generated within the myocardium to leave identifiable landmark bands at end-diastole through spatial modulation of magnetization. The imaging planes are perpendicular to the tagging planes, so that the tags appear as parallel dark stripes in MRI images at end-diastole.

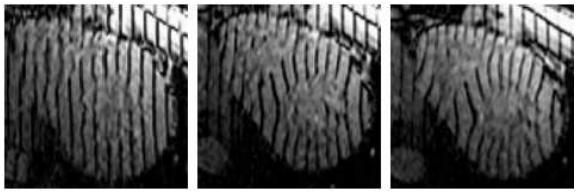


Figure 6.1: Example tagged MRI images of heart in a cardiac cycle, in short axis view.

As the underlying myocardium deforms during the cardiac cycle in vivo, the tags deform with it accordingly. Since the tags move with the myocardium, they can record myocardium motion in the direction that is normal to the tagging stripes, making it possible to establish dense point correspondences over time. Some example tagged MRI images of the heart in short axis view can be seen from Fig. 6.1.

A set of spatio-temporal (4D) tagged MRI images of the heart provides qualitative and quantitative information about the morphology, kinematic function and material properties of the heart. However, before this technique can be used in routine clinical evaluation, we need to solve several image analysis tasks. The first task is to extract the heart wall boundaries and the tagging stripes. The second task is to analyze the Left Ventricle (LV) and Right Ventricle (RV) shapes and their motion. Third, given LV and RV motion information, we can perform myocardium strain analysis. And finally, intracavitary blood flow can be modelled and analyzed. Among these image analysis tasks, several researchers have noted that one of the rate-limiting steps which prevents tagged MR from clinical use is the robust extraction and tracking of cardiac contours and myocardium tags. Although there have been intensive research efforts on automated LV and RV segmentation in tagged MRI images [76, 77, 2], efficient and robust segmentation still remains a difficult task due to the common presence of image noise, cluttered objects, object texture, as well as the complexities added by the tagging lines.

In this section, we present a robust method for segmenting and tracking of cardiac contours and myocardium tags in 4D tagged MRI images of the heart via spatio-temporal propagation. The method consists of two main techniques: the Metamorphs segmentation for robust boundary extraction, and the tunable Gabor filter bank [88] for tagging lines enhancement, removal and myocardium tracking. The Metamorphs segmentation algorithm has been introduced in Chapter 3. Next we describe the tunable Gabor filter bank technique and its application to

tagging line enhancement, removal and myocardium tracking. We then present the integration of Metamorphs with Gabor filter bank methods to provide a robust and efficient framework for contour segmentation and tag tracking in tagged MR images, with minimal human intervention. Based on the framework, a prototype system is built and experimental results from the system will be shown.

6.1.1 The Tunable Gabor Filter Bank Technique for Tagged MRI Analysis

Basic Definitions

Gabor filters have been widely used in image processing applications, such as texture segmentation [33] and edge detection [73]. The 2D Gabor filter was first introduced by Daugman [27]. It is basically a 2D Gaussian multiplied by a complex 2D sinusoid, as shown below:

$$h(x, y) = g(x', y') \cdot s(x, y) \tag{6.1}$$

In the above equation (6.1), $s(x, y)$ is a complex 2D sinusoid function, i.e.,

$$s(x, y) = \exp^{-j2\pi(Ux+Vy)} \tag{6.2}$$

where (U, V) are the 2D frequencies of the complex sinusoid, and the orientation of the sinusoid in the frequency domain is given by:

$$\phi = \arctan(V/U) \tag{6.3}$$

Also in Eq. (6.1), $g(x', y')$ is a 2D Gaussian,

$$g(x', y') = \frac{1}{2\pi\sigma_{x'}\sigma_{y'}} \exp^{-\frac{1}{2}\left((\frac{x'}{\sigma_{x'}})^2+(\frac{y'}{\sigma_{y'}})^2\right)} \tag{6.4}$$

where $x' = x\cos(\theta) + y\sin(\theta), y' = -x\sin(\theta) + y\cos(\theta)$ represent the 2D Gaussian spatial coordinates after rotating by an angle θ, and $\sigma_{x'}, \sigma_{y'}$ are the standard deviations of the Gaussian

which give the approximate spatial extent of the 2D Gaussian envelop. The 2D Gaussian envelop need not be symmetric, hence $\sigma_{x'}, \sigma_{y'}$ may not be equal. In our application, an asymmetric Gaussian envelop fits the tagging line pattern better: we can set the σ along the direction that is perpendicular to the tagging lines to be larger than that in the direction parallel to the tagging lines. In practice, we experimentally set the Gaussian envelop to be an ellipsoid whose long axis is 4 times as long as the short axis, and the σ's in Eq. (6.4) are set to be:

$$\sigma_{x'} = \frac{1}{\sqrt{(U^2 + V^2)}}$$
$$\sigma_{y'} = \frac{1}{4\sqrt{(U^2 + V^2)}}$$

The orientation of the Gaussian envelop θ need not be the same as the orientation of the sinusoid ϕ. But for the purpose of normalization, we set these two angle to be the same in our application. That is, the sinusoid is always perpendicular to the long axis of the Gaussian envelop.

Gabor Filtering of 4D Tagged MRI Images

At time 1 of the tagged MR imaging process, when the tagging lines are initially straight and equally spaced, in the spectral domain of the input tagged MR image, there exist several isolated harmonic peaks representing its frequency characteristics. And the first harmonic peak represents the main patterns of the image, which are the un-deformed tagging lines. We set the frequency parameters (U, V) of the Gabor filter to capture the un-deformed tag pattern by automatically finding the coordinates of the image's first harmonic peaks in the spectral domain. Over time, the tagged MR images are taken during a heart beat cycle, and the tagging lines move along with the underlying myocardium, hence the spacings and orientations of the tagging lines change accordingly. These changes in the spatial domain lead to corresponding changes in the frequency domain. So the new frequencies U' and V' are tuned as follows:

$$U' = \mathcal{R}(U + iV) \cdot m \cdot \exp^{(i\Delta\phi + \omega)}$$
$$V' = \mathcal{I}(U + iV) \cdot m \cdot \exp^{(i\Delta\phi + \omega)}$$

Figure 6.2: Extracted tagging lines after convolution with the tunable Gabor filter bank, for the MR image in Fig. 6.1. The myocardium contours are drawn for better readability.

where $m, \Delta\phi$ and ω are the magnitude, angle, and phase modulation, respectively in the frequency domain. $\mathcal{R}$ denotes the real part, and $\mathcal{I}$ denotes the imaginary part of the complex equation. By modulating m, we can change the gabor filter to capture changes in tagging line spacing; by modulating $\Delta\phi$, we can change the gabor filter to capture changes in tagging line orientation; and by modulating the phase angle ω, we can change the gabor filter to capture changes in the relative position of the current pixel with respect to the nearby tagging line.

Tunable Gabor Filter Bank for Tagging Line Enhancement and Removal

To process the 4D tagged MR images, we modify the parameters $m, \Delta\phi$, and ω of the initial Gabor filter derived at time 1 to fit the deformed tag patterns over time in a cardiac cycle. The initial un-tuned Gabor filter and the modified Gabor filters make up a tunable Gabor filter bank. By convolving the input tagged MRI images with an m and $\Delta\phi$ tunable Gabor filter bank, we are able to extract out the pixels that are on the tagging lines, regardless of the different tag-spacings and orientations. Examples of the extracted tagging lines after this step can be seen from Fig. 6.2.

A tunable gabor filter bank also allows the "removal" of tagging lines by filling in areas that are between or near the deformed tagging lines. This is achieved by the modulation of all three parameters $m, \Delta\phi$, and phase angle ω. In particular, since the modulation of phase angle ω represents position shifting of the current pixel with respect to the nearest tagging line, tuning ω makes the enhanced region shift away from the tagging lines. By tuning all three parameters and always returning the highest response from the Gabor filter bank, we get high response not only on or near tagging lines, but also in the regions between the tagging lines. Hence we get similar high response in areas both on and between tagging lines, and achieve the effect of tagging line "removal". This tag "removal" step is especially useful after time

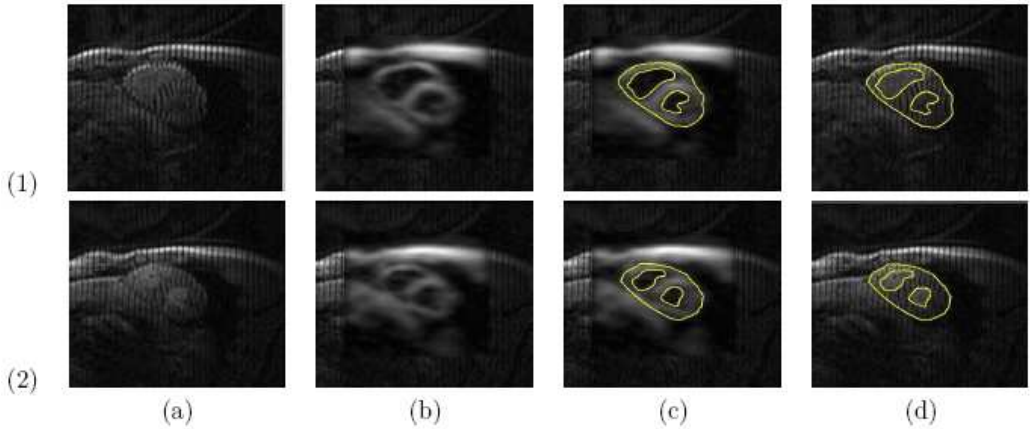


Figure 6.3: De-tagged images at mid-systolic phase and Metamorphs segmentation of LV/RV/epicardium boundaries. (1) segmentation at mid-systolic time 7, slice position 7. (2) segmentation at time 7, slice position 10. (a) original image. (b) image with tags removed by gabor filtering. (c) cardiac contours segmented by Metamorphs on detagged images. (d) contours projected on the original image.

1 (end-diastole), because over time, the tag patterns in the blood are flushed out very soon, and this tag removal method can then enhance the blood-myocardium contrast and facilitate myocardium segmentation. As shown in Fig. 6.3, the de-tagged images in a mid-systolic phase make the LV/RV/epicardium boundary segmentation much easier, we can then use Metamorphs deformable models to reliably segment out the boundaries at this mid-systolic phase before propagating the segmented contours to other time points.

Myocardium Tracking

At each pixel in an input image, we apply the tunable Gabor filter bank and find out the optimal filter parameters, m, $\Delta\phi$ and ω that maximize the Gabor filter response at that pixel. This gives us three parameter maps when considering all pixels in the image, and from the parameter maps, we can acquire information about image region properties. Fig. 6.4 shows the three optimal parameter maps for an example input tagged MR image. We can clearly see the differences in tagging line spacing (m), orientation ($\Delta\phi$) and position shift from tagging line (ω) from the parameter maps. The flat-gray areas in the parameter maps correspond to regions in the image that have maximum gabor filter response values below a threshold, which are also regions without tagging line patterns.

From the m parameter map and the ω parameter map, we can acquire a tissue point's relative

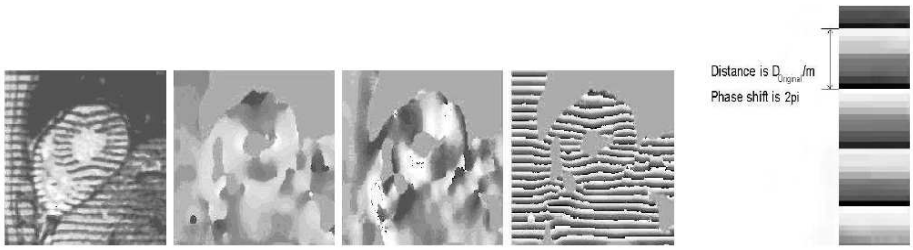


Figure 6.4: Optimal parameter values that give the maximum gabor filter response in a gabor filter bank. The first image is the original input MR image. The second image is the optimal spacing m map; the bright color indicates small spacing, and dark color indicates large spacing. The third image is the optimal orientation $\Delta\phi$ map; the bright color means the orientation of the tagging line is from lower left to upper right, and dark color means the orientation is from lower right to upper left. The fourth image is the phase ω map; the color varies from dark to bright as the phase angle varies from $-\pi$ to $+\pi$. The last figure illustrates the relationship between tag spacing and phase shift.

distance with respect to the nearest tagging line. That is, at a certain pixel, the distance between this pixel and the nearby tagging line center is determined by:

$$D = D_{original} \cdot \omega/(2\pi \cdot m) \tag{6.5}$$

where $D_{original}$ is the original spacing between two un-deformed tagging lines. If the deformation of a certain material point between two consecutive time points in a tagged MR sequence is not bigger than half of the spacing between two nearby tagging lines, which is true in almost all the tagged MR sequences because of the high imaging speed, the change in the ω maps coupled with the change in the m maps can be used to estimate the displacement of the underlying material point:

$$\begin{aligned} \Delta D &= D_{original} \cdot \Delta\omega/(2\pi \cdot m) \\ &= D_{x_{original}} \cdot \Delta\omega_x/(2\pi \cdot m_x) + D_{y_{original}} \cdot \Delta\omega_y/(2\pi \cdot m_y) \end{aligned} \tag{6.6}$$

For a typical short axis (SA) view tagged MRI sequence, we have two sets of data whose tagging lines are initially perpendicular to each other. Thus we can use Eq. (6.6) to calculate the deformations in the two perpendicular directions from the two datasets respectively. In this way, we get the complete deformation information about the myocardium in the cardiac cycle, and achieve myocardium tracking.

6.1.2 Segmentation and Tracking Framework with Experimental Results

We integrate Metamorphs segmentation (Chapter 3) with the Tunable Gabor Filter technique (Section 6.1.1) to construct a 4D spatio-temporal integrated tagged MRI image analysis system. By using the two techniques in a complementary manner, exploiting specific domain knowledge about the heart anatomy and temporal characteristics of the tagged MR images, we can achieve efficient, robust segmentation with minimal user interaction. The algorithm consists of the following main steps.

1. Tag removal for images at the mid-systolic phase. Given a 4D spatio-temporal tagged MR image dataset of the heart, we start by filtering using a tunable Gabor filter bank on images of a 3D volume that corresponds to a particular time point in the middle of the systolic phase, which we term the 'center time'. For a typical dataset in which the systolic phase is divided into 13 time intervals, we apply Gabor filtering on images in the 3D volume acquired at time 7, when tag patterns in the endocardium are flushed out by blood but tag lines in the myocardium are clearly visible.

2. Metamorphs segmentation using the de-tagged images. Given the de-tagged Gabor response images at time 7 (e.g., see Fig. 6.3), we use Metamorphs to segment the cardiac contours including the epicardium, the LV and RV endocardium. As shown in Chapter 3, the Metamorphs deformable models can be initialized far-away from the object boundary and efficiently converge to an optimal solution. Hence to segment the LV and RV endocardium, the user only needs to initialize a circular model by clicking one point (the seed point) inside each object of interest (LV or RV), then the surrounding region intensity statistics and the gradient information automatically drive the model to converge to the endocardium boundaries. We then automatically initialize a Metamorphs model for the epicardial contour by merging the endocardial contours and expanding the interior volume according to myocardium thickness statistics. The model is then allowed to evolve and converge to the epicardium boundary on the original image [1].

3. Spatial propagation at the mid-systolic center time. At the mid-systolic center time,

[1] During our experiments, we find the epicardium boundary is better estimated using Metamorphs on the original image rather than on the de-tagged image, because the epicardium boundary is often blurred with neighboring structures on the de-tagged image.

we perform segmentation at several key frames which correspond to transition frames in terms of heart topology, then let the segmented contours propagate to their nearby frames in space. In short axis cardiac MR images, from the apex to the base, the topology of the 2D cardiac contours goes through the following variations: (1) one epicardium, (2) one epicardium and one LV endocardium (in some cases of the RV hypertrophy patients, one epicardium and one RV endocardium are also possible), (3) one epicardium, one LV endocardium and one RV endocardium, and (4) one epicardium, one LV endocardium and two RV endocardium. The key frames that we segment using Metamorphs include one center frame from images with the third topology and three transition frames to other topologies. This spatial propagation actually provides a quick initialization method (rather than manually clicking the seed points as mentioned in step 2) for all the rest of the non-key frames in the 3D volume.

4. Boundary tracking using tunable Gabor filters over time. Once we have segmented the cardiac contours at time 7, we keep tracking the motion of the myocardium and the segmented contours over time using the gabor filter technique (see Section 6.1.1). This temporal propagation of the cardiac contours significantly reduces computation time, since it enables us to do supervised segmentation at only one time point, then fully automated segmentation of the complete 4D dataset can be achieved. It also improves segmentation accuracy because we capture the overall trend in heart deformation more accurately by taking into account the temporal connection between segmented boundaries.

5. User interaction to improve boundary contours. In practice, we provide the option to further refine the automatically segmented boundaries by manual correction. Doctors are provided with easy-to-use manual correction tools so that they can modify the contours until satisfied.

6. Tagging lines tracking within the heart wall (myocardium). The tagging lines are straight lines at time 0 (preparing for image acquisition). They are equally spaced at an interval of $1/\sqrt{U^2+V^2}$. Starting from time 1, we keep tracking the tagging lines only within the myocardium by considering only the heart wall regions bounded by the segmented cardiac contours. The model for tagging lines is basically a set of initially parallel *Snakes* [62] which deform over time under the influence of external forces that come from the dark lines in the original images and the enhanced tagging lines in the tag-enhanced images.

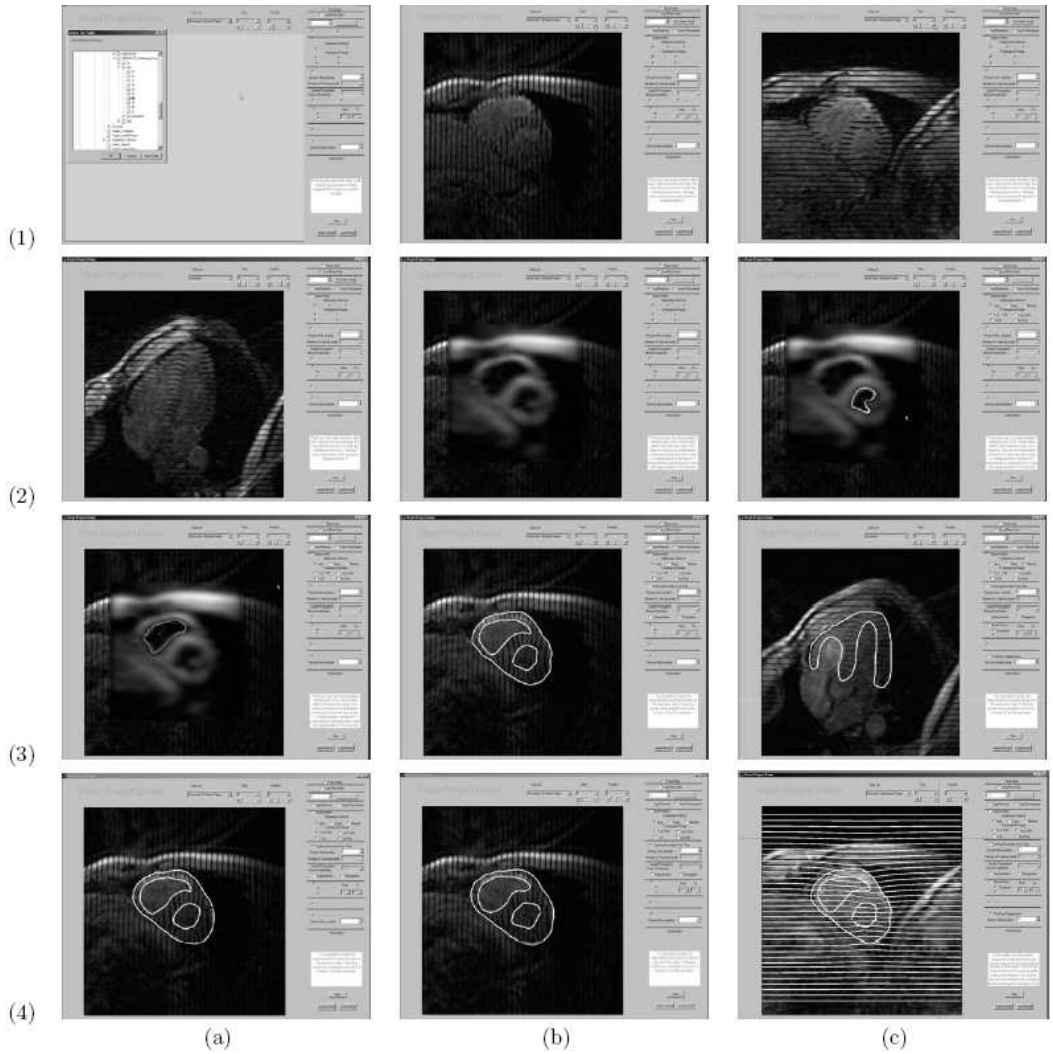


Figure 6.5: Screen snapshots of the 4D tagged MR analysis system. (1-a) reading in the SA and LA volumes. (1-b,1-c,2-a) examining the data sets. (2-b) de-tagged image at the center time. (2-c,3-a) Metamorphs segmentation on the de-tagged images. (3-b,3-c) segmentation results at the center time. The papillary muscle is excluded from the myocardium by manual interaction. (4-a,4-b) temporal propagation of the segmented contours. (4-c) tagging lines tracking.

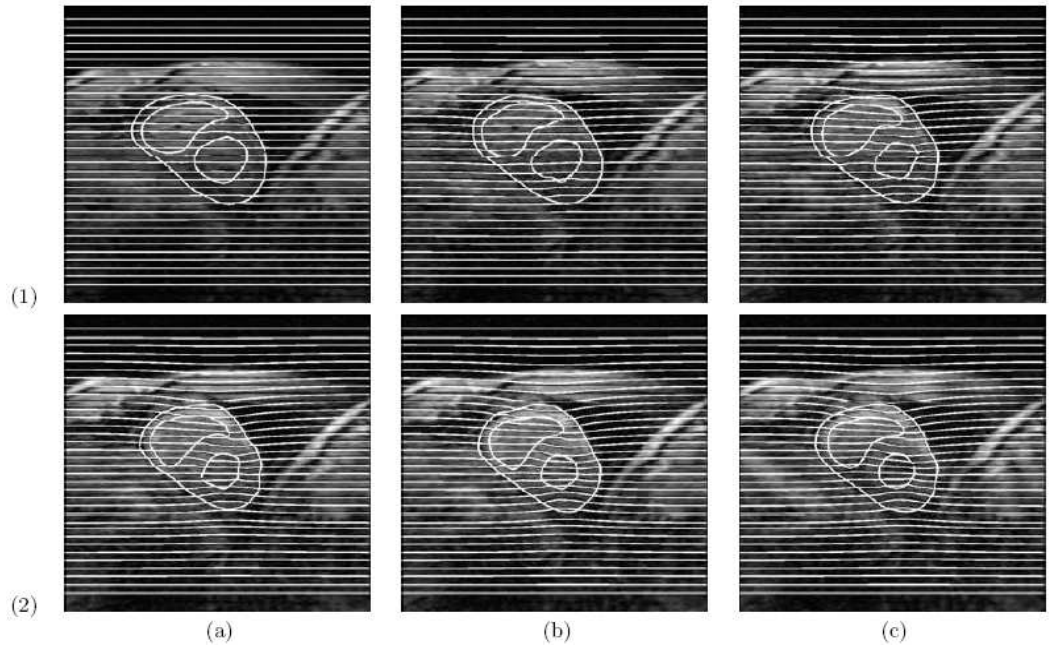


Figure 6.6: Contours and tagging lines segmented and tracked by the tagged MR analysis system. The results are for a short axis (SA) horizontal-tagged data set at a center location at times 1, 3, 5, 7, 9, and 11.

A prototype of our 4D spatial-temporal tagged MRI image analysis system is implemented in the Matlab GUI environment. The user first loads in the raw MRI data of the short axis and long axis volumes first (see Fig. 6.5(1-a)). Then the user is allowed to examine the whole data sets, which consist of two short axis and one long axis volumes, and determine the slice index of the center time (Fig. 6.5(1-b, 1-c, 2-a)). The Gabor filtering based tag removal step is done on the 3D volume at the center time (Fig. 6.5(2-b)). Then the user can choose the key frames in this 3D volume, and run Metamorphs segmentation on the key frames (Fig. 6.5(2-c, 3-a, 3-b, 3-c)). The segmented LV endocardium, RV endocardium and Epicardium contours are then propagated spatially (optional) and temporally (Fig. 6.5(4-a, 4-b)). In practice the spatial propagation step is optional because for most clinical analysis, segmentation and tracking on one typical slice is enough unless a full 4D model is required. The users can always use manual interaction to correct the contours during the whole segmentation and propagation process. After the heart boundaries are segmented in all time points, the tagging line *Snakes* models can be automatically initialized, and tagging lines within the myocardium heart wall are tracked automatically through *Snakes* deformations (see Fig. 6.5(4-c)). Also in Figure 6.6, we show a typical set of contour segmentation and tagging line tracking results generated by our system.

6.2 High-resolution 3D Facial Expression Tracking

In this section, we present a second real-world application that uses the global-to-local shape registration and learning algorithms that we introduced in Chapter 4.

6.2.1 Introduction

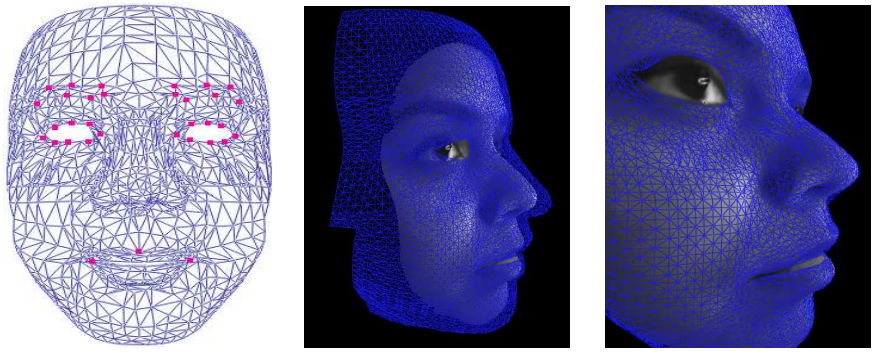
Synthesis and re-targeting of facial expressions is central to facial animation and often involves significant manual work in order to achieve realistic expressions. Recent technological advances have made possible the acquisition of high resolution dynamics 3D shape data that capture very accurate geometry at speeds that exceed video frame rate. such ranging techniques include structured light [52, 51, 95], and spacetime stereo [124, 30]. The high quality that such data provides is very attractive in analysis of facial expressions. However, the dense data samples returned by these 3D face scan techniques are not registered in object space and hence there is no guarantee of intra-frame correspondences. Thus to use such data for temporal study of the subtle dynamics in expressions and for expression synthesis and re-targeting, efficient facial expression tracking and shape registration algorithms are needed in order to establish correspondences between data in different frames for the same face as well as between different faces. Facial expression tracking and face geometry registration are also essential for a variety of other applications such as facial expression recognition, classification, detection of emotional states, among others.

In the literature, there are few tracking algorithms proposed for the high-resolution 3D facial expression data. Most existing facial motion or expression tracking algorithms utilize image data from 2D video sequences [31, 37, 45, 67, 119], and focus on the accurate tracking of a low number of facial features such as points located around the brows, eyes, nose, mouth, etc. While the movements of these feature points in an expression can often be used effectively in classification, they are hardly sufficient in most recognition applications, since many distinct characteristics of a person's expression lie in the subtle details such as wrinkles and furrows. In video sequences, it can be very difficult to capture these details due to the lost of information in projection, lighting, shadow, and other conditions. Among the few 3D methods that establish intra-frame correspondences between face range scans, some depend on markers that are

attached on the face performing an expression [48] or depend on facial feature correspondences manually selected by users [123]; others use 3D shape registration algorithms such as [9, 125] to establish correspondences based on facial geometry. However, when used for expression tracking, both marker-based and geometry-based methods lack a proper modelling of the motion style in an expression, which results from the combined effect of global facial motion that is caused by muscle action, and subtler expression details such as wrinkles that are generated by highly local skin deformations.

In the following subsections, we introduce a novel hierarchical tracking framework for 3D dynamic expression data, which can both track global facial motion and fit to expression details, providing a tight coupling between global and local deformations. The high quality moving face range scans we use are acquired using the system described by [52]. In order to track facial features and establish dense intra-frame correspondences, we use a multi-resolution deformable face model. On the coarse level we use a mesh with one thousand (1K) nodes that was first developed for robust face tracking in low quality 2D images [45] and extend it to deal with 3D range data. This method is fast, and the deformation parameters for each facial motion are few and intuitive. However it cannot capture the large amount of local deformations and so we use it for a coarse-level tracking. The local deformations and details in expressions are captured in a higher level fitting process. For each frame of the range scan, the resulting mesh from the coarse-level tracking is used to initialize a subdivided fine mesh with sixteen thousand (16K) nodes. This fine mesh is registered to the frame using the 3D extension of the local non-rigid shape registration algorithm that we introduced in Chapter 4 and in [57, 55]. This algorithm integrates an implicit shape representation [81] and the cubic B-spline based Free Form Deformations (FFD) model [102, 94], and generates a registration/deformation field that is smooth, continuous and gives dense one-to-one correspondences.

Using our hierarchical framework, we did tracking experiments on dynamic facial scan of seven different subjects, and conducted both qualitative and quantitative validation on the tracking accuracy. The results are very promising, showing the potential of our algorithm to serve as an efficient way to parameterize high resolution 3D dynamic expression data in order to make it easy to use while preserving the accuracy and visual quality that such data guarantees.



(a) (b) (c)

Figure 6.7: (a) The generic face model with manually selected feature points. (b) The face model and the face scan data are roughly aligned. (c) The result of the initial fitting to a 3D face scan data.

6.2.2 System Overview: A Hierarchical Tracking Framework

To track the facial motion in an expression, we use a multi-resolution deformable face model. The face model has two resolutions: a coarse-level mesh with 1K nodes and a fine-level mesh with 16K nodes.

We use the 16K node mesh for the initial fitting between the face model and an actor's face scan before performing an expression (i.e., the first frame). Figure 6.7 demonstrates this initial fitting process. First, the face model and the 3D scan data are roughly aligned by hand (Figure 6.7(b)). Then the 3D extension of the local FFD based non-rigid shape registration algorithm is used to register the face model with this range scan, achieving a complete surface match (Figure 6.7(c)). Also in order to constrain the initial dense correspondences established by the registration algorithm, we define a small set of feature points on the face model (typically around 30, see Figure 6.7(a) as an example), then manually select their correspondences on the range data. These feature correspondences are incorporated as hard constraints during the optimization process of the registration algorithm (see Section 6.2.4 for details). As a result, the initial correspondences established, especially between facial features such as tip of the nose and corners of the eyes, are very reliable.

After the initial fitting, a hierarchical scheme is adopted to track the intra-frame deformations in an expression. In the coarse level, we use the 1K node face model and extend the deformable tracking system in [45] to track 3D dynamic range scans. In order to fit to expression details, for each frame of the range data, we use the coarse level tracking result to initialize

the subdivided 16K node mesh in a higher level. Then this 16K node refined mesh is registered to the frame using the same variational non-rigid shape registration algorithm used for initial fitting. This hierarchical tracking/fitting protocol provides a tight coupling between global and local deformations, and results in efficient and very detailed fitting to the 3D face scan data (see Figure 6.8 for examples).

Next we describe in detail the global tracking and local non-rigid 3D shape registration algorithms.

6.2.3 Global Deformation

According to [75], we can express the position $\mathbf{p}(t)$ of a point on the 3D model as the sum of a reference model $\mathbf{s}(t)$ and a displacement $\mathbf{d}(t)$, i.e.

$$\mathbf{p} = \mathbf{s} + \mathbf{d} \tag{6.7}$$

The shape, position, and orientation of the reference model $\mathbf{s}$ can also change. We define the reference shape as

$$\mathbf{s} = \mathbf{T}(\mathbf{q}; \mathbf{e}) \tag{6.8}$$

where $\mathbf{s}$ is the result of a geometric primitive $\mathbf{e}$ undergoing the *global deformation* $\mathbf{T}$. $\mathbf{T}$ depends on a set of n control parameters $\mathbf{q} = (q_1, q_2, \ldots, q_n)^T$ [45]. Some of these parameters affect the general position of the object (global rotation and translation), some affect the shape (like a global scaling), and some affect only parts of the object. Since they have a common mathematical formulation as defined in Eq. (6.8), we do not need any distinction between these parameters.

Assuming the global deformations $\mathbf{T}$ is differentiable [75], for every point s_i on the surface of the reference model $\mathbf{s}$, the derivative of s_i with respect to q_i is the Jacobian J_i:

$$J_i = \begin{bmatrix} \vdots & & \vdots \\ \frac{\partial s_i}{q_1} & \cdots & \frac{\partial s_i}{q_n} \\ \vdots & & \vdots \end{bmatrix} \tag{6.9}$$

where each column l of the Jacobian J_i is the gradient of s_i with respect to the parameter q_l. To keep updating the parameters $\mathbf{q}$ to track the global deformations in an expression, we use the following dynamic system updating formula:

$$\dot{\vec{q}} = \vec{f_g} + F_{internal}(\vec{q}) \tag{6.10}$$

where $\vec{f_g}$ is a n-dimensional displacement called *generalized force* and $F_{internal}(\vec{q})$ is the result of the internal forces of the model (e.g. elasticity). For each frame, the following steps are done iteratively to derive the parameter values at equilibrium:

1. Calculate the 3D external force $\vec{f_i}$ on each point s_i of the deformable model's surface. This force is derived from the 3D displacements between the model point and its closest data point on the face scan.

2. Calculate the generalized force

$$\vec{f_g} = \sum_i \vec{f_{g_i}} = \sum_i J_i^T \vec{f_i} \tag{6.11}$$

3. Compute the generalized internal force $F_{internal}(\vec{q})$.

4. Calculate the derivative $\dot{\vec{q}}$ as defined in Eq.(6.10).

5. Do an Euler integration step:

$$\vec{q} = \vec{q} + \lambda \dot{\vec{q}} \tag{6.12}$$

   where λ is the learning rate.

6. Repeat step 1 to 5, until $\dot{\vec{q}}$ is close to zero.

In our system implementation of the algorithm, we first use a Iterative Closest Point (ICP) method [9] to rigidly align the model and face scans, taking advantage of the dense 3D scan data. Then the face model is divided into several deformable regions whose shape and motion are represented by a few control parameters. Typically, for a smiling expression the face model is divided into 10 small regions with a total of 17 parameters. The changes in these parameters during global tracking are derived from the dynamic system updating scheme described in this

section. Because of the small parameter set, the global tracking step is very fast, though it can not capture detailed local deformations.

6.2.4 Local Deformations

To further recover the local deformations $\mathbf{d}(t)$, as in Eq. (6.7), we use the 3D extension of the local FFD based shape registration algorithms introduced in Chapter 4. Since the algorithm in Chapter 4 was presented for the 2D case, we give the formulation in 3D below. In the formulation, we refer to the deforming 3D face mesh model as the source surface, and the face range scan as the target surface. We use the $16K$ fine-level mesh model for the local registration.

The Implicit Shape Representation

Both the 3D face model and the range images, which are surfaces, are implicitly represented in a higher dimensional volumetric space. Given a surface S, the Euclidean distance transform is used to embed this surface as the zero level set of a distance function Φ_S defined in the embedding space Ω:

$$\Phi_S(x,y,z) = \begin{cases} 0 & , \quad (x,y,z) \in S \\ D((x,y),S) & , \quad (x,y,z) \in [\Omega - S] \end{cases} \tag{6.13}$$

where $D((x,y,z),S)$ refers to the min Euclidean distance between the grid location (x,y,z) and the shape S. In shape registration, such a representation facilitates the imposition of constraints on smoothness and coherent correspondence, since one would align the original surfaces as well as their clones that are positioned coherently in the volume plane.

IFFD local registration

To achieve local registration between a source surface S and a target surface D, we aim to recover a deformation field that creates correspondences between the implicit representations Φ_S and Φ_D. We model such a local deformation field $L(\mathbf{x}), \mathbf{x} = (x,y,z)$, using the incremental Free Form Deformations (IFFD) in 3D.

Let us consider a 3D lattice of control points,

$$P_{m,n,o} = (P^x_{m,n,o}, P^y_{m,n,o}, P^z_{m,n,o}) \tag{6.14}$$

where $(m, n, o) \in [1, M] \times [1, N] \times [1, O]$, overlaid to a region $\Gamma = \{\mathbf{x}\} = \{(x, y, z)|1 \leq x \leq X, 1 \leq y \leq Y, 1 \leq z \leq Z\}$ in the embedding space that encloses the source surface S. Suppose the initial configuration of the control lattice P^0 is regular, and the deforming control lattice is $P = P^0 + \delta P$, then in our approach, the local deformation parameters are the IFFD parameters, which are the deformations of the control points in all directions:

$$\Theta = \{(\delta P^x_{m,n,o}, \delta P^y_{m,n,o}, \delta P^z_{m,n,o})\} \tag{6.15}$$

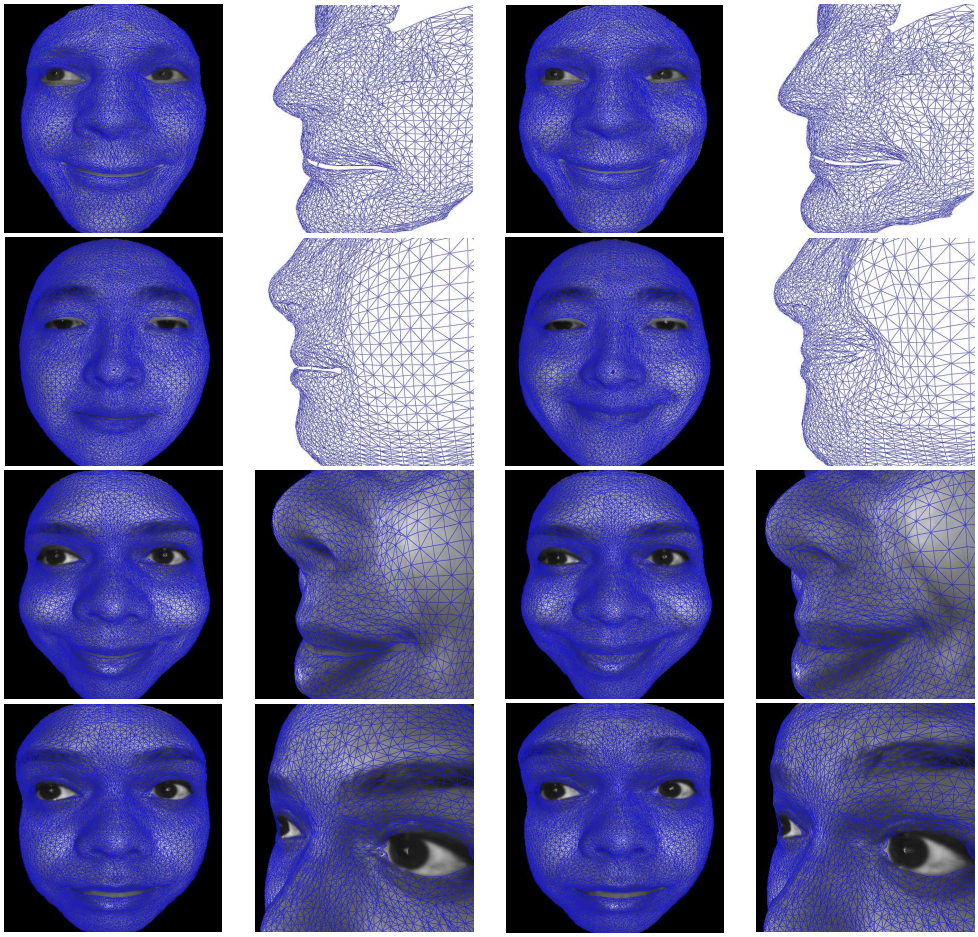
where $(m, n, o) \in [1, M] \times [1, N] \times [1, O]$. Under these specifications, the deformed position of a point $\mathbf{x} = (x, y, z)$ in the sample domain [2] given the deformation of the control lattice from P^0 to P, is defined in terms of a tensor product of Cubic B-splines:

$$\begin{aligned} L(\Theta; \mathbf{x}) &= \mathbf{x} + \delta L(\Theta; \mathbf{x}) \\ &= \textstyle\sum_{q=0}^{3} \sum_{l=0}^{3} \sum_{r=0}^{3} [B_q(u) B_l(v) B_r(w) \\ &\quad (P^0_{i+q,j+l,k+r} + \delta P_{i+q,j+l,k+r})] \\ i &= \lfloor \tfrac{x}{X} \cdot (M-1) \rfloor + 1 \\ j &= \lfloor \tfrac{y}{Y} \cdot (N-1) \rfloor + 1 \\ k &= \lfloor \tfrac{z}{Z} \cdot (O-1) \rfloor + 1 \end{aligned} \tag{6.16}$$

The terms of the deformation component refer to:

- $\delta P_{i+q,j+l,k+r}$, $(q, l, r) \in [0, 3] \times [0, 3] \times [0, 3]$ are the deformations of pixel $\mathbf{x}$'s 64 adjacent control points.
- $B_q(u)$ is the q^{th}, $B_l(v)$ is the l^{th} and $B_r(w)$ is the r^{th} basis function of a Cubic B-spline.
- $\delta L(\mathbf{x}) = \sum_{q=0}^{3} \sum_{l=0}^{3} \sum_{r=0}^{3} B_q(u) B_l(v) B_r(w)\ \delta P_{i+q,j+l,k+r}$ is the incremental deformation for pixel $\mathbf{x}$.

[2] We use a narrow band around the zero level set surface as the sample domain to ensure efficiency.



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

Figure 6.8: [Top Row]: Snapshots of the *smile* expression of subject 1. [Second Row]: The *smile* expression of subject 2. [Third Row]: The *smile* expression of subject 3. [Bottom Row]: The *Raising eyebrow* expression of subject 3. [Column a]: Front view of frame 1. [Column b]: Close-up view of Column a (without range scan - for showing details; with range scan - for showing correspondences). [Column c]: Front view of frame 2. [Column d]: Close-up view of Column c.

Having defined the form of the local deformation field $L(\mathbf{x})$ with respect to the IFFD parameters $\Theta = \delta P$, local registration is now equivalent to finding the control lattice deformation δP such that the deformed source surface coincides with the target surface. The Sum-of-Squared-Differences (SSD) criterion is used as a data-driven term to recover the parameters:

$$E_{data}(\Theta) = \iint_{\Omega} \left(\Phi_D(\mathbf{x}) - \Phi_S(L(\Theta;\mathbf{x}))\right)^2 d\mathbf{x} \tag{6.17}$$

In order to further preserve the regularity of the recovered registration flow, one can consider a smoothness term on the local deformation field δL. We consider a computationally efficient

smoothness term:

$$E_{smooth}(\Theta) = \iint_{\Omega} \left(\left|\left| \frac{\partial}{\partial x}\delta L(\Theta;\mathbf{x}) \right|\right|^2 + \left|\left| \frac{\partial}{\partial y}\delta L(\Theta;\mathbf{x}) \right|\right|^2 + \left|\left| \frac{\partial}{\partial z}\delta L(\Theta;\mathbf{x}) \right|\right|^2 \right) d\mathbf{x} \tag{6.18}$$

An additional implicit smoothness constraint is also imposed by the B-Spline FFD, which guarantees C^1 continuity at control points and C^2 continuity everywhere else.

We can also enhance the accuracy of the tracking system by imposing correspondence constraints on certain feature points such as tip of the nose and corners of the eyes. Assuming we have n_c features, and for each of them, there is a pair of corresponding points, $\mathbf{x}_{si}$ on the source surface S and $\mathbf{x}_{di}$ on the target surface D, where $i = 1, \ldots, n_c$. Then the feature correspondence constraints can be expressed as

$$E_{feature}(\Theta) = \sum_i (L(\Theta;\mathbf{x}_{si}) - \mathbf{x}_{di})^2; i \in [1, n_c] \tag{6.19}$$

The data-driven term, the smoothness constraint term, and the feature correspondence constraint term can be integrated in a single objective function,

$$\begin{aligned} E(\Theta) &= E_{data}(\Theta) + \alpha E_{smooth}(\Theta) + \beta E_{feature}(\Theta) \\ &= \iint_{\Omega} \left(\Phi_D(\mathbf{x}) - \Phi_S(L(\Theta;\mathbf{x}))\right)^2 d\mathbf{x} \\ &+ \alpha \iint_{\Omega} \left(\left|\left| \frac{\partial}{\partial x}\delta L(\Theta;\mathbf{x}) \right|\right|^2 + \left|\left| \frac{\partial}{\partial y}\delta L(\Theta;\mathbf{x}) \right|\right|^2 \right. \\ &\quad \left. + \left|\left| \frac{\partial}{\partial z}\delta L(\Theta;\mathbf{x}) \right|\right|^2 \right) d\mathbf{x} \\ &+ \beta \textstyle\sum_i (L(\Theta;\mathbf{x}_{si}) - \mathbf{x}_{di})^2 \end{aligned} \tag{6.20}$$

where $i \in [1, n_c]$, n_c is the number of feature points, and α and β are the constants balancing the contributions from different terms. Using the calculus of variation and a gradient descent

method, such an objective function can be optimized to recover the deformation parameters Θ,

$$\begin{aligned}
\frac{\partial E(\Theta)}{\partial \theta_i} &= -2 \iint_{\Omega} \big[\big(\Phi_{\hat{D}}(\mathbf{x}) - \Phi_S(L(\Theta;\mathbf{x})) \big) \\
&\quad \cdot \nabla \Phi_S(L(\Theta;\mathbf{x})) \cdot \tfrac{\partial \delta L(\Theta;\mathbf{x})}{\partial \theta_i} \big] d\mathbf{x} \\
&+ 2\alpha \iint_{\Omega} \big[\tfrac{\partial}{\partial x} \delta L(\Theta;\mathbf{x}) \tfrac{\partial}{\partial \theta_i} \big(\tfrac{\partial}{\partial x} \delta L(\Theta;\mathbf{x}) \big) \\
&\quad + \tfrac{\partial}{\partial y} \delta L(\Theta;\mathbf{x}) \tfrac{\partial}{\partial \theta_i} \big(\tfrac{\partial}{\partial y} \delta L(\Theta;\mathbf{x}) \big) \\
&\quad + \tfrac{\partial}{\partial z} \delta L(\Theta;\mathbf{x}) \tfrac{\partial}{\partial \theta_i} \big(\tfrac{\partial}{\partial z} \delta L(\Theta;\mathbf{x}) \big) \big] d\mathbf{x} \\
&+ 2\beta \textstyle\sum_i \big[(L(\Theta;\mathbf{x}_{si}) - \mathbf{x}_{di}) \tfrac{\partial}{\partial \theta_i} \big(L(\Theta;\mathbf{x}_{si}) \big) \big]
\end{aligned} \tag{6.21}$$

and consequently the local registration field $L(\Theta;\mathbf{x})$. Correspondences thus can be established between points $\mathbf{x} = (x, y, z)$ on the source surface and the points $L(\mathbf{x})$ on the target surface. And the displacements between the corresponding points consist of the local deformations $\mathbf{d}(t)$ in Eq. (6.7).

6.2.5 Facial Expression Tracking Experimental Results

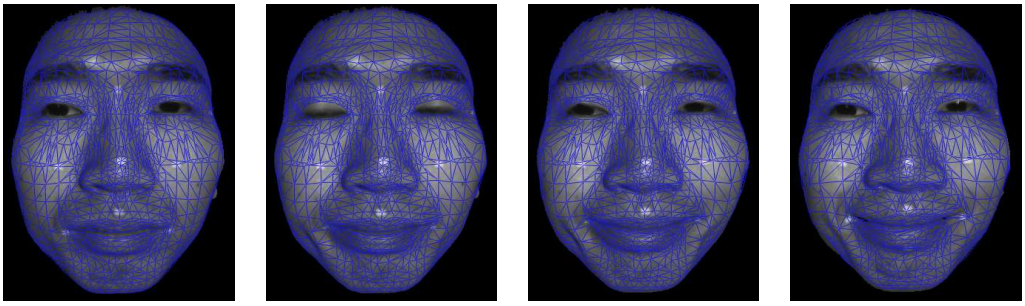
We conducted tracking experiments using the dynamic facial scans of 7 different subjects. These data are acquired using the 3D high resolution shape acquisition system described by [52]. For each subject, data for two different expressions are collected: the *smile* expression and the *raising eyebrow* expression. On each data sequence, we first register its first frame with the face model (at fine level $16K$ nodes), then we keep tracking the intra-frame deformations using the tightly coupled global and local tracking algorithm. This hierarchical tracking protocol results in efficient and very detailed fitting to the 3D face scan data. Example tracking results are shown in Figure 6.8. The fine details in an expression captured using our method is demonstrated in Figure 6.8(a-b), and the high accuracy of the intra-frame correspondences established during tracking is demonstrated in Figure 6.8(c-d).

Our system is implemented using C++ under the Linux environment. All our experiments run at interactive rate on a Pentium Xeon 3GHz dual processor platform.

For qualitative evaluation of the model point tracking results, we compare the texture of the original scan data with the synthesized texture based on the tracking results. We generate synthetic texture for the $16K$-node face control mesh that tracks the range scans, by applying

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

Figure 6.9: [Top Row]: Comparison between original texture of a subject's colored range scans and synthesized texture of the tracking face control mesh, for the *raising eyebrow* expression. [Second row]: Comparison for the *smile* expression. (a) Snapshot 1 from the original scan data. (b) Snapshot 1 from the synthesized rendering of the tracking result. (c) Snapshot 2 from the original scan data. (d) Snapshot 2 from the synthesized rendering of the tracking result.



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

Figure 6.10: Selected tracking results of a 'smile' sequence, with 50 frames in total. The resulting meshes are illustrated in blue color and white dots are attached markers for verification purposes only. (a) frame 1, (b) frame 5, (c) frame 10, and (d) frame 37.

the texture retrieved from the registered first frame to the remaining frames of each sequence, adding in shading and shadow effects considering the change in facial geometry during the expression. The comparison results on frames from example tracking sequences are shown in Figure 6.9. Ideally, we should have compared the reflectance maps of the original scans with the synthesized reflectance maps. But our comparison of direct luminance-based textures provides sufficient evidence on the validity of the tracking result.

To further conduct quantitative validation on the accuracy of the tracking algorithm, we perform a number of experiments on 3D facial expression sequences with attached markers. The markers are for validation purpose only. In order to be detected successfully, the size of markers is around 4mm by 4mm. An example tracking result is demonstrated in Figure

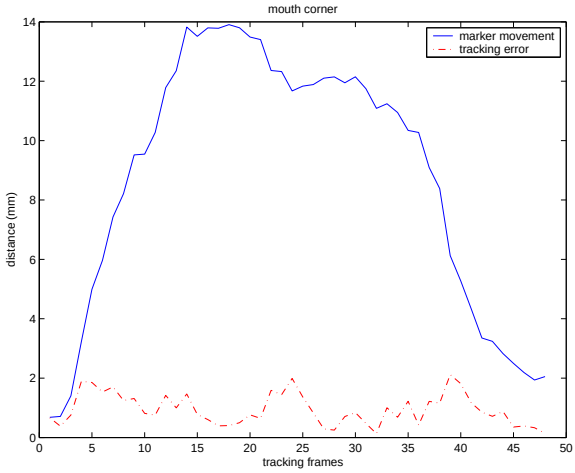


Figure 6.11: Tracking error of the marker on a mouth corner.

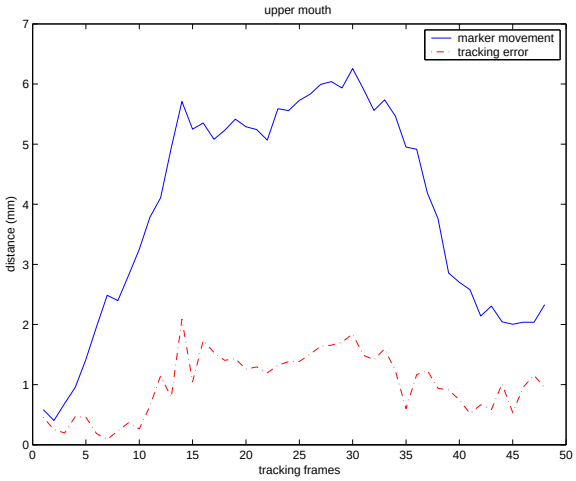


Figure 6.12: Tracking error of the marker on the upper mouth.

6.10, where the blue meshes are resulting coarse level control mesh, and the white dots are the attached markers for verification purposes only. Figures 6.11-6.14 show the algorithm's tracking error estimations on the mouth corner, upper mouth, cheek, and nose tip respectively. As we can see, in most cases the tracking error is around 1mm. This error is very low given that the resolution of the 3D range scan data is around $0.5mm$.

6.2.6 Discussion

The hierarchical facial expression tracking system we implemented can be used to efficiently parameterize the large amount of high-resolution 3D dynamic range scan data, by dealing with

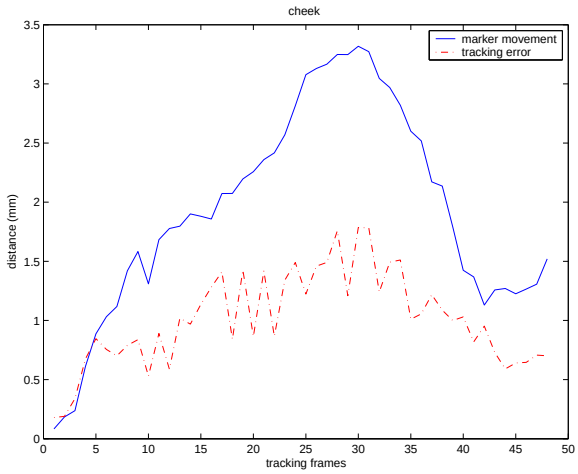


Figure 6.13: Tracking error of the marker on a cheek.

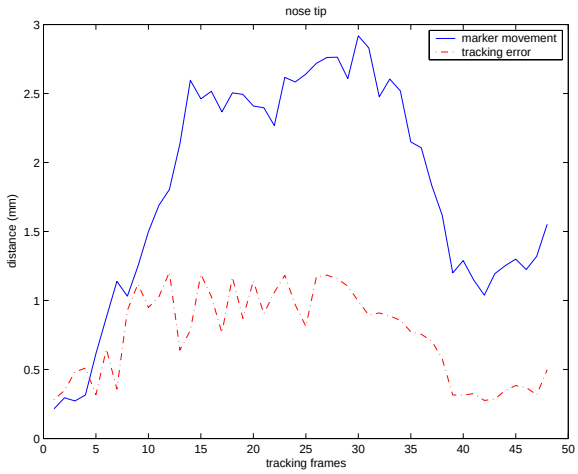


Figure 6.14: Tracking error of the marker on the nose tip.

both large-scale deformations and free-form style fine-details existing in the facial expressions. The accuracy and resolution of our method allow us to capture and track subtle expression details and hence to use the tracking parameters for motion analysis and expression recognition.

The dense intra-frame correspondences established by the system can also be used to learn the mapping between dynamic facial motion and expression style, hence enabling the synthesis of new expressions and re-targeting of one person's expression style onto another person's facial geometry. Our work published in [115] explores this idea by using a nonlinear dimensionality reduction framework, the Local Linear Embedding (LLE) [91], to learn the most discriminating characteristics of an individual's expression as that person's "expression style". Then new expressions can be synthesized, either as dynamic morning between individuals or

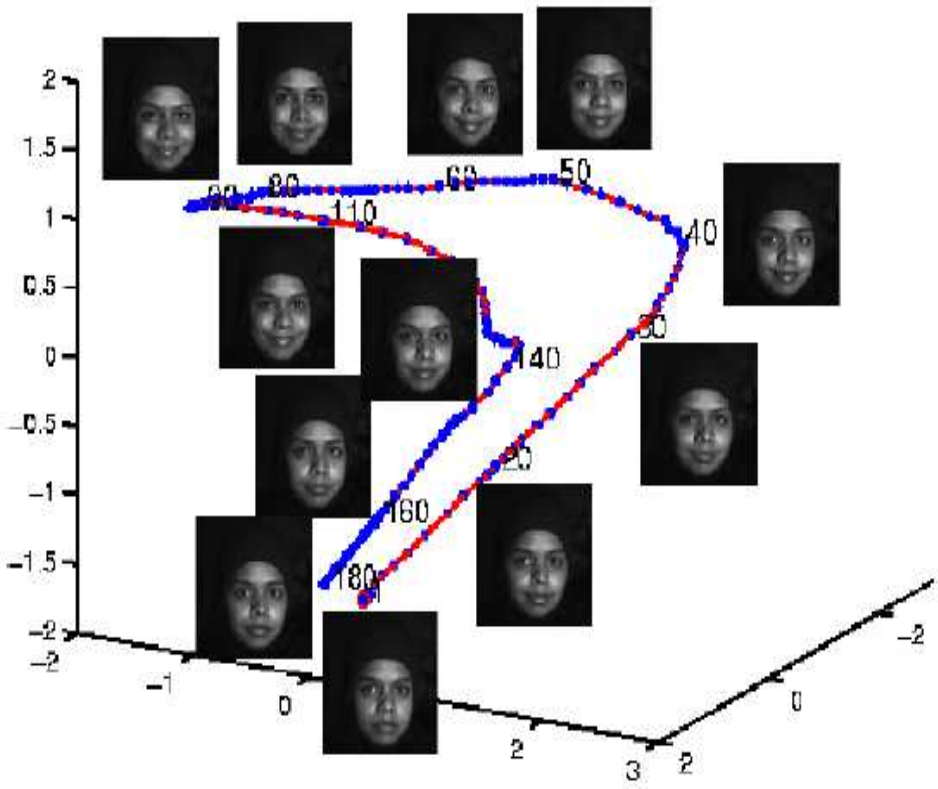


Figure 6.15: Low dimensional representation of a "smile" expression. An embedding of the smile motion by LLE shows that the smile motion can be well embedded in a one dimensional manifold located in the 3-D Euclidean space. Manifold points for similar faces are located nearby in the manifold.

as expression transfer (re-targeting) from a source face to a target face. One example of the learned low dimensional manifold for a "smile" expression is shown in Fig. 6.15. And an example of expression synthesis and re-targeting is shown in Fig. 6.16. More details of the learning, synthesis and re-targeting methods can be found in [115].

6.3 Summary

In this chapter, we presented one application of the Metamorphs deformable models introduced in Chapter 2 to 4D tagged MRI images analysis, and another application of the 3D local IFFD shape registration algorithm to high-resolution 3D facial expression tracking. There are a lot of other applications of the algorithms proposed in this thesis. For instance, Metamorphs segmentation can be applied to shape reconstruction, object tracking, and organ (e.g. heart) modelling. The shape and image registration algorithms can be used for motion correction, establishing correspondences between diagnosis imaging scans at different time points. And other than the

Figure 6.16: (First Column) Subject 1. (Second Column) Subject 2. (Third Column) Subject 1 with synthetic smile transferred from Subject 2. (Fourth Column) Detail of the synthesized smile.

heart and face, the algorithms can be applied to images of the brain, chest, and many other types of images.

Chapter 7

Conclusions

This thesis proposed shape, appearance and deformation representations that are suited for shape and appearance information integration to better solve computer vision and medical image analysis problems. Using the proposed representations, several novel algorithms are developed for robust model-based segmentation, shape and/or image registration, coupled shape and appearance prior model learning, and the potential of the algorithms are demonstrated through extensive experiments and real-world applications.

Integrating information from multiple sources to solve computer vision and medical image analysis problems more robustly is an important and challenging research topic. This thesis aimed to contribute to integrating shape and appearance information, which are two key aspects of an image or an object of interest. The combination of implicit shape representation, nonparametric image/object appearance representation, and space-warping Free Form Deformations provided a fresh framework in which deformable models and statistical shape and appearance models can be defined and learned naturally in the joint shape and intensity spaces. There are future work to pursue however toward the goal of information integration for robust image interpretation in medical image analysis and computer vision.

First, the learning technique used in this thesis for statistical shape and appearance models is Principal Component Analysis (PCA). There are other nonlinear generative models, such as Local Linear Embedding (LLE) and nonlinear Independent Component Analysis (ICA), to be investigated since they may be more appropriate to model the shape and appearance variations. Another direction is to learn discriminate classifiers using shape features and/or appearance features and combine the classifiers based on our confidence level on each source of information.

Second, in medical image analysis, it is often important to register images of multiple

modalities so that complementary information from them can be fused and visualized on a single platform for the purposes of diagnosis and discovery. One type of multi-modal image registration is particularly attractive, which is the registration between functional (e.g. PET, fMRI) and structural (e.g. CT, MRI) images. The image registration algorithms proposed in this thesis need to evaluated on registering functional and structural images, and new algorithms can be developed to exploit the constraints in this type of image registration.

Third, the Metamorphs deformable models generalize traditional shape-only deformable models and represent a class of deformable models that have both boundary shape and interior region statistics. It will be intriguing to extend Metamorphs to also include model material properties for the purposes of organ modeling and surgery simulation.

References

[1] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *J. Roy. Stat. Soc.*, 28:131–142, 1966.

[2] A. A. Amini, Y. Chen, M. Elayyadi, and P. Radeva. Tag surface reconstruction and tracking of myocardial beads from SPAMM-MRI with parametric B-spline surfaces. *IEEE Trans. on Medical Imaging*, 20(2):94–103, 2001.

[3] B. Avants and J. Gee. Comparison and evaluation of retrospective intermodality image registration techniques. In *Proc. of second Int'l Workshop on Biomedical Image Registration, PA, USA*, 2003.

[4] L. Axel and L. Dougherty. Heart wall motion: Improved method of spatial modulation of magnetization for MR imaging. *Radiology*, 272:349–350, 1989.

[5] L. Axel and L. Dougherty. MR imaging of motion with spatial modulation of magnetization. *Radiology*, 171(3):841–845, 1989.

[6] E. Bardinet, L. D. Cohen, and N. Ayache. A parametric deformable model to fit unstructured 3D data. *Computer Vision and Image Understanding*, 71(1):39–54, 1998.

[7] S. Belongie, J. Malik, and J. Puzicha. Matching Shapes. In *Proc. of IEEE Int'l Conf. on Computer Vision*, pages 456–461, 2001.

[8] M. Berthod, Z. Kato, S. Yu, and J. Zerubia. Bayesian image classification using Markov Random Fields. *Image and Vision Computing*, 14:285–295, 1996.

[9] P. Besl and N. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), 1992.

[10] P.J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[11] F. L. Bookstein. Principal wraps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, June 1989.

[12] F. L. Bookstein. Landmark methods for forms without landmarks: Morphometrics of grouping differences in outline shape. *Medical Image Analysis*, 1(3):225–243, 1997.

[13] A. Can, C.V. Stewart, B. Roysam, and H.L. Tanenbaum. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(3):347–364, 2002.

[14] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. In *Proc. of IEEE Int'l Conf. on Computer Vision*, pages 694–699, 1995.

[15] T. Chan and L. Vese. Active contours without edges. *IEEE Trans. on Image Processing*, 10(2):266–277, 2001.

[16] C. Chefd'Hotel, G. Hermosillo, and O. Faugeras. A Variational Approach to Multi-Modal Image Matching. In *Proc. of IEEE Workshop in Variational and Level Set Methods*, pages 21–28, Vancouver, Canada, 2001.

[17] M. Chen, T. Kanade, D. Pomerleau, and J. Schneider. 3D Deformable Registration of Medical Images using a Statistical Atlas. In *Proc. of Int'l Conf. on Medical Imaging Copmuting and Computer-Assisted Intervention*, pages 621–630, 1999.

[18] T. Chen and D.N. Metaxas. Image segmentation based on the integration of markov random fields and deformable models. In *Proc. of Int'l Conf. on Medical Imaging Copmuting and Computer-Assisted Intervention*, pages 256–265, 2000.

[19] H. Chui and A. Rangarajan. A New Algorithm for Non-Rigid Point Matching. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages II: 44–51, 2000.

[20] H. Chui and A. Rangarajan. Learning an atlas from unlabeled point-sets. In *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 179–186, Dec. 2001.

[21] L. D. Cohen and I. Cohen. Finite-element methods for active contour models and balloons for 2-D and 3-D images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 15:1131–1147, 1993.

[22] A. Collignon, F. Maes, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multimodality image registration using information theory. In *Proc. of Int'l Conf. on Information Processing in Medical Imaging*, pages 263–274, 1995.

[23] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

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

[25] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. In *Proc. of European Conf. on Computer Vision*, volume 2, pages 484–498, 1998.

[26] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models - their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[27] J. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, 2(7):1160–1169, 1985.

[28] R. H. Davies, T. F. Cootes, J. C. Waterton, and C. J. Taylor. An Efficient Method for Constructing Optimal Statistical Shape Models. In *Proc. of Int'l Conf. on Medical Imaging Copmuting and Computer-Assisted Intervention*, pages 57–65, 2001.

[29] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. 3D Statistical Shape Models using Direct Optimization of Description Length. In *Proc. of European Conf. on Computer Vision*, pages 3–20, 2002.

[30] J. Davis, R. Ramamoorthi, and S. Rusinkiewicz. Spacetime stereo: A unifying framework for depth from triangulation. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 359– 366, June 2003.

[31] D. DeCarlo and D. Metaxas. Optical flow constraints on deformable models with applications to face tracking. *International Journal of Computer Vision*, 38(2):99–127, 2000.

[32] R. O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis*. Wiley, 1973.

[33] D. Dunn, W. E. Higgins, and J. Wakeley. Texture segmentation using 2D Gabor elementary functions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 16:130–149, 1994.

[34] N. Duta, A. K. Jain, and M.-P. Dubuisson-Jolly. Learning 2d shape models. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume II, pages 8–14, 1999.

[35] A. Elgammal, R. Duraiswami, and L. S. Davis. Efficient kernel density estimation using the fast gauss transform with applications to color modeling and tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(11):1499–1504, 2003.

[36] A. Elgammal and C. S. Lee. Separating style and content on a nonlinear manifold. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 478–485, 2004.

[37] L.A. Essa and A. Pentland. A vision system for observing and extracting facial action parameters. In *CVPR'94*, pages 76–83, 1994.

[38] P. Faloutsos, M. van de Panne, and D. Terzopoulos. Dynamic Free-Form Deformations for Animation Synthesis. *IEEE Trans. Visualization and Computer Graphics*, 3:201–214, 1997.

[39] J. Feldmar and N. Ayache. Rigid, Affine and Locally Affine Registration of Free-Form Surfaces. *ijcv*, 18:99–119, 1996.

[40] R. Fergus, P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages II:264–271, 2003.

[41] D. Forsey and R. Bartels. Hierachical B-spline Refinement. *ACM Transactions on Computer Graphics*, 22:205–212, 1988.

[42] A. F. Frangi, D. Rueckert, J. A. Schnabel, and W. J. Niessen. Automatic 3D ASM Construction via Atlas-based Landmarking and Volumetric Elastic Registration. In *Proc. of Int'l Conf. on Information Processing in Medical Imaging*, pages 78–91, 2001.

[43] Y. Freund and R. E. Schapire. A decision-theoretic generalization of the on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119–139, 1997.

[44] Y. Gdalyahu and D. Weinshall. Flexible Syntactic Matching of Curves and its Application to Automatic Hierarchical Classification of Silhouettes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(12):1312–1328, 1999.

[45] Siome Klein Goldenstein, Christian Vogler, and Dimitris Metaxas. Statistical cue integration in dag deformable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):801–813, 2003.

[46] N. Govindaraju, D. Knott, N. Jain, I. Kabul, R. Tamstorf, R. Gayle, M. Lin, and D. Manocha. Interactive Collision Detection between Deformable Models using Chromatic Decomposition. *ACM Trans. on Graphics (TOG), Proc. of ACM SIGGRAPH 2005*, 24(3):991–999, 2005.

[47] E. Grimson. *Object Recognition by Computer: the Role of Geometric Constraints*. MIT Press, Cambridge, MA, 1990.

[48] Brian Guenter, Cindy Grimm, Daniel Wood, Henrique Malvar, and Fredric Pighin. Making faces. In *SIGGRAPH'98*, pages 55–66, 1998.

[49] T. Hartkens, D.L. Hill, A.D. Castellano-Smith, D.J. Hawkes, C.R. Maurer, A.J. Martin, W.A. Hall, H. Liu, and C.L. Truwit. Using points and surfaces to improve voxel-based nonrigid registration. In *Proc. of Int'l Conf. on Medical Imaging Copmuting and Computer-Assisted Intervention*, pages II:565–572, 2002.

[50] A. Hill, C. J. Taylor, and A. D. Brett. A Framework for Automatic Landmark Identification using a New Method of Non-rigid Correspondences. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(3):241–251, 2000.

[51] P. S. Huang, Q. Hu, F. Jin, and F. P. Chiang. Color-encoded digital fringe projection technique for high-speed three-dimensional surface contouring. *Optical Engineering*, 38(6):1065–1071, 1999.

[52] P. S. Huang, C. Zhang, and F. P. Chiang. High-speed 3-d shape measurement based on digital fringe projection. *Opt. Eng.*, 42(1):163–168, 2003.

[53] X. Huang, D. Metaxas, and T. Chen. Metamorphs: Deformable shape and texture models. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 496–503, 2004.

[54] X. Huang, N. Paragios, and D. Metaxas. Establishing local correspondences towards compact representations of anatomical structures. In *Proc. of Int'l Conf. on Medical Imaging Copmuting and Computer-Assisted Intervention*, pages 926–934, 2003.

[55] X. Huang, N. Paragios, and D. Metaxas. Shape registration in implicit spaces using information theory and free form deformations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2006.

[56] X. Huang, Y. Sun, D. Metaxas, F. Sauer, and C. Xu. Hybrid image registration based on configural matching of scale-invariant salient region features. In *Second IEEE Workshop on Image and Video Registration, in conjunction with CVPR'04*, July 2004.

[57] X. Huang, S. Zhang, Y. Wang, D. Metaxas, and D. Samaras. A hierarchical framework for high resolution facial expression tracking. In *Third IEEE Workshop on Articulated and Nonrigid Motion, in conjunction with CVPR'04*, July 2004.

[58] A. E. Johnson and M. Hebert. Recognizing Objects by Matching Oriented Points. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 684–689, 1997.

[59] T.N. Jones and D.N. Metaxas. Automated 3D segmentation using deformable models and fuzzy affinity. In *Proc. of Int'l Conf. on Information Processing in Medical Imaging*, pages 113–126, 1997.

[60] B. Julesz. Texons, the elements of texture perception, and their interactions. *Nature*, 290(5802):91–97, 1981.

[61] T. Kadir and M. Brady. Saliency, scale and image description. *International Journal of Computer Vision*, 45(2):83–105, 2001.

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

[63] Y. Keller and A. Averbuch. Implicit similarity: A new approach to multi-sensor image registration. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages II:543–548, 2003.

[64] M. Leventon, E. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 1316–1323, 2000.

[65] M. E. Leventon, E. L. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 1316–1323, 2000.

[66] A. Levin, P. Viola, and Y. Freund. Unsupervised improvement of visual detectors using co-training. In *Proc. of IEEE Int'l Conf. on Computer Vision*, pages 626–633, 2003.

[67] J.J. Lien, T.K. Kanade, A.Z. Zlochower, J.F. Cohn, and C.C. Li. Subtly different facial expression recognition and expression intensity estimation. In *CVPR'98*, pages 853–859, 1998.

[68] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multi-modality image registration by maximization of mutual information. *IEEE Trans. on Medical Imaging*, 16(2):187–198, 1997.

[69] J. Malik, S. Belongie, Leung T., and J. Shi. Contour and texture analysis for image segmentation. *Int'l Journal of Computer Vision*, 43(1):7–27, 2001.

[70] R. Malladi, J. Sethian, and B.C. Vemuri. Shape modeling with front propagation: A level set approach. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17(2):158–175, 1995.

[71] C.R. Maurer, R.J. Maciunas, and J.M. Fitzpatrick. Registration of head CT images to physical space using a weighted combination of points and surfaces. *IEEE Trans. on Medical Imaging*, 17(5):753–761, 1998.

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

[73] R. Mehrotra, K. R. Namuduri, and N. Ranganathan. Gabor filter-based edge detection. *Journal of Pattern Recognition*, 25:1479–1493, 1992.

[74] D. Metaxas. *Physics-Based Deformable Models*. Kluwer Academic Publishers, 1996.

[75] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(6):580–591, 1993.

[76] A. Montillo, D. N. Metaxas, and Axel L. Automated segmentation of the left and right ventricles in 4D cardiac SPAMM images. In *Proc. of Int'l Conf. on Medical Imaging Copmuting and Computer-Assisted Intervention*, pages 620–633, 2002.

[77] A. Montillo, D. N. Metaxas, and Axel L. Automated model-based segmentation of the left and right ventricles in tagged cardiac MRI. In *Proc. of Int'l Conf. on Medical Imaging Copmuting and Computer-Assisted Intervention*, pages 507–515, 2003.

[78] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989.

[79] S. Nayar, H. Murase, and S. Nene. Parametric appearance representation. *Early Visual Learning*, 1996.

[80] S. Osher and N. Paragios. *(Editors). Geometric Level Set Method in Imaging, Vision and Graphics*. Spring Verlag, 2002.

[81] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed : Algorithms based on the Hamilton-Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.

[82] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: an application to face detection. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 130–136, 1997.

[83] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *Int'l Journal of Computer Vision*, 46(3):223–247, 2002.

[84] N. Paragios, M. Rousson, and V. Ramesh. Matching Distance Functions: A Shape-to-Area Variational Approach for Global-to-Local Registration. In *Proc. of European Conf. on Computer Vision*, pages II:775–790, 2002.

[85] N. Paragios, M. Rousson, and V. Ramesh. Non-Rigid Registration Using Distance Functions. *Computer Vision and Image Understanding*, 89:142–165, 2003.

[86] J. Pearl. *Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufman Publishers, 1988.

[87] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual Information Based Registration of Medical Images: A Survey. *IEEE Trans. on Medical Imaging*, 22(8):986–1004, 2003.

[88] Z. Qian, A. Montillo, D. Metaxas, and L. Axel. Segmenting Cardiac MRI Tagging Lines using Gabor Filter Banks. In *25th Int'l Conf. of the IEEE Engineering in Medicine and Biology Society (EMBS)*, pages 630–633, 2003.

[89] S. Romdhani, S. Gong, and A. Psarrou. A multi-view nonlinear active shape model using kernal pca. In *Proc. of British Machine Vision Conference*, volume 2, pages 483–492, 1999.

[90] R. Ronfard. Region-based strategies for active contour models. *Int'l Journal of Computer Vision*, 13(2):229–251, 1994.

[91] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[92] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20:22–38, 1998.

[93] D. Rueckert, A. F. Frangi, and J. A. Schnabel. Automatic construction of 3d statistical deformable models using non-rigid registration. In *Proc. of Int'l Conf. on Medical Imaging Copmuting and Computer-Assisted Intervention*, pages 77–84, 2001.

[94] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid Registration Using Free-Form Deformations: Application to Breast MR Images. *IEEE Transactions on Medical Imaging*, 8:712–721, 1999.

[95] S. Rusinkiewicz, O. Hall-Holt, and Levoy Marc. Real-time 3d model acquisition. In *SIGGRAPH'02*, volume 1281 of *3D acquisition and image based rendering*, pages 438 – 446, 2002.

[96] M. A. Ruzon and C. Tomasi. Color edge detection with the compass operator. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 160–166, 1999.

[97] C. Samson, L. Blanc-Feraud, G. Aubert, and J. Zerubia. A level set model for image classification. *Int'l Journal of Computer Vision*, 40(3):187–198, 2000.

[98] H. Schneiderman and T. Kanade. A statistical method for 3D object detection applied to faces and cars. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 746–751, 2000.

[99] D. W. Scott. *Multivariate Density Estimation*. Wiley-Interscience, 1992.

[100] T. Sebastian, P. Klein, and B. Kimia. Alignment-based recognition of shape outlines. *Lection Notes in Computer Science*, 2059:606–618, 2001.

[101] T. Sebastian, P. Klein, and B. Kimia. Recognition of Shapes by Editting Shock Graphs. In *Proc. of IEEE Int'l Conf. on Computer Vision*, pages 755–762, 2001.

[102] T. W. Sederberg and S. R. Parry. Free-form deformation of solid geometric models. In *Proc. of the 13th Annual Conf. on Computer Graphics*, pages 151–160, 1986.

[103] J. A. Sethian. *Level Set Methods: Evolving interface in Geometry, Fluid Mechanics, Computer Vision and Material Science (First Edition)*. Cambridge University Press, 1996.

[104] D. Shen and C. Davatzikos. HAMMER: Hierarchical attribute matching mechanism for elastic registration. *IEEE Trans. on Medical Imaging*, 21(11):1421–1439, 2002.

[105] J. Shi and C. Tomasi. Good features to track. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 593–600, 1994.

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

[107] C. Studholme, D. L. G. Hill, and D. J. Hawkes. An Overlap Invariant Entropy Measure of 3D Medical Image Alignment. *Pattern Recognition*, 32(1):71–86, 1999.

[108] J. Tenenbaum. Mapping a manifold of perceptual observations. In *Advances in Neural Information Processing*, volume 10, pages 682–688, 1998.

[109] J.-P. Thirion. New feature points based on geometric invariants for 3D image registration. *International Journal of Computer Vision*, 18(2):121–137, May 1996.

[110] S.J. Timoner. *Compact Representations for Fast Nonrigid Registration of Medical Images*. Ch.6, Ph.D. dissertation, AI Tech Report 2003-015, MIT, 2003.

[111] R. Veltkamp and M. Hagedoorn. State-of-the-art in Shape Matching. Technical Report UU-CS-1999-27, Utrecht University, 1999.

[112] L. A. Vese and T. F. Chan. A multiphase level set framework for image segmentation using the Mumford and Shah model. *Int'l Journal of Computer Vision*, 50(3):271–293, 2002.

[113] P. Viola and M. Jones. Robust real-time face detection. *Int'l Journal of Computer Vision*, 57(2):137–154, 2004.

[114] P. Viola and W. Wells. Aligment by Maximization of Mutual Information. In *Proc. of IEEE Int'l Conf. on Computer Vision*, pages 16–23, 1995.

[115] Y. Wang, X. Huang, C. S. Lee, S. Zhang, Z. Li, D. Samaras, D. Metaxas, A. Elgammal, and P. Huang. High resolution acquisition, learning and transfer of dynamic 3-d facial expressions. *Computer Graphics Forum (Proc. of Eurographics)*, 23(3):677–686, 2004.

[116] J. Weng and N. Ahuja. Octrees of objects in arbitrary motion: Representation and efficiency. *Computer Vision, Graphics and Image Processing*, 39(2):167–185, 1987.

[117] J. West, J. Fitzpatrick, M. Wang, B. Dawant, C. Maurer, R. Kessler, and R. Maciunas. Comparison and evaluation of retrospective intermodality image registration techniques. In *Proc. of the SPIE Conf. on Medical Imaging, vol. 2710*, pages 332–347, 1996.

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

[119] Y. Yacoob and L. Davis. Computing spatio-temporal representations of human faces. In *CVPR'94*, pages 70–75, 1994.

[120] J. Yang and J. S. Duncan. 3D image segmentation of deformable objects with shape-appearance joint prior models. In *Proc. of Int'l Conf. on Medical Imaging Copmuting and Computer-Assisted Intervention*, pages 573–580, 2003.

[121] A. Yezzi, S. Kichenssamy, A. Kumar, P. Olver, and A. Tannebaum. A geometric snake model for segmentation of medical imagery. *IEEE Trans. on Medical Imaging*, 16(2):199–209, 1997.

[122] A. J. Yezzi, A. Tsai, and A. Willsky. A statistical approach to snakes for bimodal and trimodal imagery. In *Proc. of IEEE Int'l Conf. on Computer Vision*, volume 2, pages 898–903, 1999.

[123] Jun yong Noh and Ulrich Neumann. Expression cloning. In *SIGGRAPH*, pages 277–288, 2001.

[124] Li Zhang, Brian Curless, and Steven M. Seitz. Spacetime stereo: Shape recovery for dynamic scenes. In *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pages 367–374, June 2003.

[125] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, 13(2):119–152, 1994.

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

[127] S. C. Zhu, C. E. Guo, Y. Z. Wang, and Z. J. Xu. What are textons? *Int'l Journal of Computer Vision*, 62(1):121–143, 2005.

[128] W. Zhu and T. Chan. Stability for Shape Comparison Model. Technical Report 0308, UCLA-CAM, 2003.

Curriculum Vita

Xiaolei Huang

1999 B. S.E. in Computer Science, Tsinghua University, Beijing, China.

2001 M.Sc. in Computer Science, Rutgers - The State University of New Jersey, New Brunswick, NJ, USA.

2006 Ph. D. in Computer Science, Rutgers - The State University of New Jersey, New Brunswick, NJ, USA.

1999-2002 Teaching Assistant, Department of Computer Science, Rutgers - The State University of New Jersey, New Brunswick, NJ, USA.

2002-2005 Research Assistant, Division of Computer and Information Sciences, Rutgers - The State University of New Jersey, New Brunswick, NJ, USA.

2005-2006 Staff Scientist, Computer Aided Diagnosis and Therapy Solutions Group, Siemens Medical Solutions, Malvern, PA, USA.

Publications

Journal Papers and Book Chapters

[TPAMI'06] Xiaolei Huang, Nikos Paragios, Dimitris Metaxas, "Shape Registration in Implicit Spaces using Information Theory and Free Form Deformations", to appear in *IEEE Trans. on Pattern Analysis and Machine Intelligence* (PAMI), 2006.

[GraphicsForum'04] Yang Wang, Xiaolei Huang, Chan-Su Lee, Song Zhang, Zhiguo Li, Dimitris Samaras, Dimitris Metaxas, Ahmed Elgammal, and Peisen Huang, "High Resolution Acquisition, Learning and Transfer of Dynamic 3-D Facial Expressions," In *Computer Graphics Forum*, 23(3):677-686, 2004. (Also presented at Eurographics 2004.)

[CVbookChapter'05] Dimitris Metaxas, Xiaolei Huang, Ting Chen, "Integrating Shape and Texture in Deformable Models: From Hybrid Methods to Metamorphs," In *Mathematical Models in Computer Vision: The Handbook*, N. Paragios, Y. Chen and O. Faugeras (Editors), Springler, 2005.

[ShapeBookChapter'05] Nikos Paragios, Maxin Taron, Xiaolei Huang, Mikael Rousson, and Dimitris Metaxas, "On the Representation of Shapes Using Implicit Functions," In *Statistics and Analysis of Shapes*, H. Krim and A. Yezzi (Editors), Springer Verlag, 2005.

Selected Conference Papers

[ECCV'06] Lior Wolf, Xiaolei Huang, Ian Martin, and Dimitris Metaxas, "Patch-based Texture Edges and Segmentation," to appear in *Proc. of the 9th European Conf. on Computer Vision*, ECCV, 2006.

[ISBI'06] Junzhou Huang, Xiaolei Huang, Dimitris Metaxas, and Debarata Banerjee, "3D Tumor Shape Reconstruction from 2D Bioluminescence Images," to appear in *Proc. of the IEEE Int'l Symposium on Biomedical Imaging: From Nano to Macro*, ISBI, 2006.

[MICCAI'05] Weijun He, Xiaolei Huang, Dimitris Metaxas, and Xiaoyou Ying, "Efficient Learning by Combining Confidence-rated Classifiers to Incorporate Unlabeled Medical Data," In *Proc. of the 8th Annual Int'l Conf. on Medical Image Computing and Computer Assisted Intervention*, MICCAI'05, LNCS-3749, pp. 745-752, 2005.

[CVPR'04] Xiaolei Huang, Dimitris Metaxas, and Ting Chen, "Metamorphs: Deformable Shape and Texture Models," In *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, CVPR'04 (1), pp. 496-503, 2004.

[MICCAI'04] Xiaolei Huang, Zhiguo Li, and Dimitris Metaxas, "Learning Coupled Prior Shape and Appearance Models for Segmentation," In *Proc. of the 7th Annual Int'l Conf. on Medical Image Computing and Computer Assisted Intervention*, MICCAI'04 (1), LNCS-3216, pp. 60-69, 2004.

[MICCAI'03] Xiaolei Huang, Nikos Paragios, and Dimitris Metaxas, "Establishing Local Correspondences towards Compact Representations of Anatomical Structures," In *Proc. of the 6th Annual Int'l Conf. on Medical Image Computing and Computer Assisted Intervention*, MICCAI'03 (2), LNCS-2879, pp. 926-934, 2003.

[CVPR'03] Manish Singh, and Xiaolei Huang, "Computing Layered Surface Representations: An Algorithm for Detecting and Separating Transparent Overlays," In *Proc. of the IEEE Computer Society Conf. on Computer Vision and Pattern Recognition*, CVPR'03 (1), pp. 11-18, 2003.

Other Conference and Workshop Papers

[EMMCVPR'05] Xiaolei Huang, Zhen Qian, Rui Huang, and Dimitris Metaxas, "Deformable-model based Textured Object Segmentation," In *Proc. of the 4th Int'l Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition* (EMMCVPR), LNCS-3757, pp. 119-135, 2005.

[SPIE-PW'06] Xiaolei Huang, Dimitris Metaxas, Lata G. Menon, Philipp Mayer-Kuckuk, Joseph R. Bertino, and Debabrata Banerjee, "Recovering 3D Tumor Locations from 2D Bioluminescence Images and Registration with CT Images," In *Proc. of the Conf. on Multimodal Biomedical Imaging, Biomedical Optics 2006*, as part of SPIE Photonics West Symposium.

[SPIE-MI'05] Zhen Qian, Xiaolei Huang, Dimitris Metaxas, Ting Chen, Leon Axel, "Robust Segmentation of 4D Cardiac MRI-tagged Images via Spatio-temporal Propagation," In *Proc. Of SPIE, Medical Imaging: Physiology, Function and Structure from Medical Images*, Vol. 5746, pp. 580-591, 2005.

[IVR'04] Xiaolei Huang, Yiyong Sun, Dimitris Metaxas, Frank Sauer, and Chenyang Xu, "Hybrid Image Registration based on Configural Matching of Scale-Invariant Salient Region Features," In *Second IEEE Workshop on Image and Video Registration*, IVR'04, in conjunction with CVPR'04, Washington D.C., July, 2004.

[ANM'04] Xiaolei Huang, Song Zhang, Yang Wang, Dimitris Metaxas, and Dimitris Samaras, "A Hierarchical Framework for High Resolution Facial Expression Tracking," In *Third IEEE Workshop on Articulated and Nonrigid Motion*, ANM'04, in conjunction with CVPR'04, Washington D.C., June, 2004.

[WSEAS'04] Dimitris Metaxas, Ting Chen, Xiaolei Huang, and Leon Axel, "Cardiac Segmentation from MRI-Tagged and CT Images," In *Proc. of the 8th WSEAS International Conf. on Computers, special session on Imaging and Image Processing of Dynamic Processes in biology and medicine*, also in WSEAS Transactions on Computers, 2004.

[ICDIA'02] Chenyang Xu, Xiaolei Huang, Arun Krishnan, and Sanjiv Samant, "An Automated Image-based Method for Multi-Leaf Collimator Positioning Verification in Intensity Modulated Radiation Therapy," In *Proc. of the International Conf. on Diagnostic Imaging and Analysis*, ICDIA'02, Shanghai, China, 2002.