SURFACE REGISTRATION FOR PHARYNGEAL RADIATION TREATMENT PLANNING

Qingyu Zhao

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2017

Approved by:

Stephen Pizer

Marc Niethammer

Ron Alterovitz

J.S. Marron

Julian Rosenman

ABSTRACT

QINGYU ZHAO: Surface Registration for Pharyngeal Radiation Treatment Planning
(Under the direction of Stephen Pizer)

Endoscopy is an in-body examination procedure that enables direct visualization of tumor spread on tissue surfaces. In the context of radiation treatment planning for throat cancer, there have been attempts to fuse this endoscopic information into the planning CT space for better tumor localization. One way to achieve this CT/Endoscope fusion is to first reconstruct a full 3D surface model from the endoscopic video and then register that surface into the CT space. These two steps both require an algorithm that can accurately register two or more surfaces. In this dissertation, I present a surface registration method I have developed, called Thin Shell Demons (TSD), for achieving the two goals mentioned above.

There are two key aspects in TSD: geometry and mechanics. First, I develop a novel surface geometric feature descriptor based on multi-scale curvatures that can accurately capture local shape information. I show that the descriptor can be effectively used in TSD and other surface registration frameworks, such as spectral graph matching. Second, I adopt a physical thin shell model in TSD to produce realistic surface deformation in the registration process. I also extend this physical model for orthotropic thin shells and propose a probabilistic framework to learn orthotropic stiffness parameters from a group of known deformations. The anisotropic stiffness learning opens up a new perspective to shape analysis and allows more accurate surface deformation and registration in the TSD framework. Finally, I show that TSD can also be extended into a novel groupwise registration framework.

The advantages of Thin Shell Demons allow us to build a complete 3D model of the throat, called an endoscopogram, from a group of single-frame-based reconstructions. It

also allows us to register an endoscopogram to a CT segmentation surface, thereby allowing information transfer for treatment planning.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my principal advisor Prof. Stephen Pizer for the continuous support of my Ph.D study and related research over the five years. He is truly a great mentor to me. His guidance has helped me a lot in both scientific research and in personal life. I could not have imagined having a better advisor and mentor for my Ph.D study. I would also like to thank my other two academic advisors, Prof. Marc Niethammer and Ron Alterovitz, for their patience, motivation and guidance all along the way. I would like to thank Prof. Julian Rosenman for providing this wonderful opportunity in working in the field of radiation oncology. I would like to thank Prof. Steve Marron for the meaningful discussions and serving as my committee member.

A special thanks to my family. Words cannot express how grateful I am to my parents and my wife for all of the sacrifices that they've made on my behalf. I wouldn't be able to accomplish my work without their love and support. Finally, I thank all of my friends and those who have helped here at UNC.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

TSD	Thin Shell Demons
MRF	Markov Random Field
EM	Expectation-Maximization
FEM	Finite Element Method
HKS	Heat Kernel Signatures
CT	Computed Tomography
CGAL	Computational and Geometry Algorithms Library
SfS	Shape-from-Shading
SfMS	Shape-from-Motion-and-Shading
SfM	Structure-from-Motion
OD	Overlapping Distance
MAP	Maximum-a-Posteriori
LDDMM	Large Deformation Diffeomorphic Metric Mapping
SPD	Symmetric Positive Definite

CHAPTER 1

Introduction

1.1 Pharyngeal Radiation Treatment Planning

Modern radiation therapy treatment planning relies on imaging modalities such as computed tomography (CT) to determine tumor location and spread. CT scans are preferred since it is an x-ray based technique that can capture normal-tissue/tumor characteristics (electron density) as seen by the treatment beams used in radiation oncology and be used directly for dose calculation (Pereira et al., 2014). CT images also provide anatomical information necessary for tumor localization and verification of the treatment plans. However, for head and neck cancer, CT may not fully reveal tumor locations due to limited spatial and intensity resolution. Specifically, in the case of pharyngeal cancer, evidence has been shown that the cancer begins in the flat, squamous cells that make up the lining surface of the anatomical structures in the nasopharynx (Wei and Kwong, 2010). This fact makes it more difficult for tumor detection and localization in CT because CT inherently does not image tissue boundaries (Zhang et al., 2014).

Therefore, for pharyngeal cancer, an additional kind of medical imaging modality, namely endoscopy, is obtained at treatment planning time. A nasopharyngoscopy is a procedure wherein the physician inserts a flexible scope tube through the patient's nose, advances it into the back of the throat, and takes videos of the pharyngeal and laryngeal regions. Endoscopic videos provide direct optical visualization of the pharyngeal surface and provide information, such as a tumor's texture, color and superficial (mucosal) spread, that cannot be seen on CT. However, despite these advantages, the use of endoscopy for treatment planning is significantly limited by the fact that

the 2D frames from the endoscopic video do not explicitly provide 3D spatial information, thus prohibiting direct 3D tumor localization. Moreover, reviewing the video is time-consuming, and the optical views do not provide the full geometric conformation of the throat.

Hence, a data fusion between CT and endoscopy would be helpful in producing improved tumor localization and likely better treatment plans; this has motivated doing the registration between the two imaging modalities. In the next section, I will briefly describe the overall pipeline for the fusion between a 3D CT image and an endoscopic video, and I will point out the need for surface registration that is used in this pipeline.

1.2 CT/Endoscopy Fusion Pipeline

Given a 3D head-and-neck CT scan of a patient and a nasopharyngoscopic video of the same patient, the goal of the fusion is to transfer the texture and location information extracted from the endoscopic video into the CT space.

The general idea is to register an endoscopy-reconstructed 3D surface model with a CT segmentation surface. Fig. 1.1 shows the overall pipeline. The pipeline can be generally divided into the two following steps.

i. A 3D surface model, which we call an *endoscopogram*, is built from the endoscopic video using many 2D video frames. To do this, a partial 3D surface model is reconstructed for each

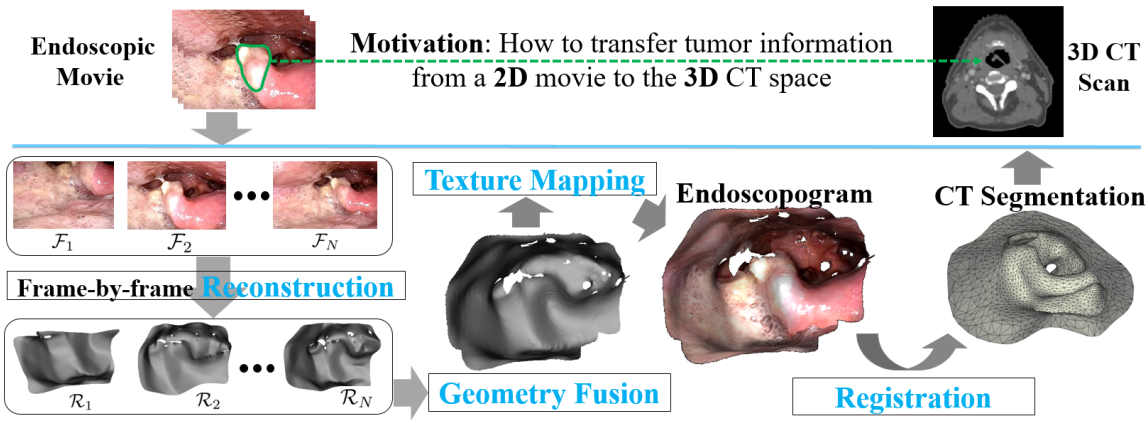


Figure 1.1: CT/endoscopy fusion pipeline.

individual 2D frame. Then multiple single-frame-based partial 3D reconstructions are registered together to achieve a complete pharyngeal surface model (endoscopogram).

ii. The endoscopogram is registered to the 3D pharyngeal surface segmented out from the CT scan. The computed deformation between the two surfaces allows the transfer of tumor location and texture information from the endoscopogram to the CT space.

The above two steps require accurate surface registration. Surface registration is a long-studied topic (Audette et al., 2000) in the medical image computing field, and many promising methods have been proposed for various applications (Tam et al., 2013). Unfortunately, none of them suits the CT/endoscope fusion pipeline. The next section elaborates on some special challenges we have to face in our surface registration problem.

1.3 Challenges of Surface Registration for the Fusion

Surface Disparity. As mentioned above, the endoscopogram is built from multiple single-frame-based reconstructions. The challenge here for the surface registration is that all individual single-frame-based reconstructions are only partially overlapping with each other due to the constantly changing camera viewpoint across frames. Moreover, the individual reconstructions may have missing data (holes) due to camera occlusion. Despite the missing data situation, there is also a significant disparity between the endoscopogram and the CT surface. Because of the partial volume effect on CT and tissue collapse caused by pharyngeal deformations, there are surface patches in either the CT surface or the endoscopogram that do not have counterparts. This fact requires us to have a surface registration method that can handle partial data, different surface topology and other disparity problems. However, many surface registration algorithms are purely based on matching intrinsic surface geometry, thereby not suitable in our application.

Large Number of Surfaces. The endoscopogram is fused from hundreds of single-frame reconstructions. The naive incremental stitching method successively uses pairwise surface registration to add a new single-frame reconstruction to the current result, but this introduces serious

bias onto early frames and propagates errors quickly to later frames. A better choice is to use a groupwise surface registration method that can simultaneously register all the frames to a single target. Current groupwise methods rely on having a known template or iteratively estimating the mean surface. However, neither approach is feasible in our application due to the only partially overlapping nature of the single-frame reconstructions.

Large Nonrigid Deformation. Another challenge in building the endoscopogram is that one needs to take into account the fact that all single-frame-based reconstructions may be slightly deformed since the tissue may have deformed between 2D frame acquisitions. Such non-rigid tissue deformations, often caused by the swallowing motion of the patient during endoscopy, are physical processes governed by pharyngeal muscles. Other physical deformations in the pharynx come from the patient's posture change between the CT scan time and the endoscopy time. During a CT scan, the patient is lying down on a curved table, whereas during endoscopy the patient is sitting straight up. Different gravity effects may cause significant shape changes in the pharynx, yielding large deformations between the endoscopogram surface and the CT segmentation surface. Finally, the presence of the endoscope can irritate the throat, leading to drastic pharyngeal motion. This mixture of different sources of deformations poses serious challenges to most current surface registration methods.

Physical Reality. Following the above argument, we prefer producing physically plausible deformations in the registration process. Unfortunately, most surface registration algorithms do not consider the physical properties of the surface material. Moreover, the apparent pharyngeal surface deformation is affected by the underlying muscles and bones. An accurate physical model therefore requires comprehensive anatomical knowledge of the entire head-and-neck region, including material stiffness parameters of different tissue types. However, patient-specific measurements of these parameters are not obtainable in our application, which makes realistic physical modeling extremely challenging.

Alternatively, many statistical frameworks have been applied to produce "realistic" deformations that are predicted from a set of training deformation examples. However, this statistical

training/prediction procedure relies on the pure geometric appearance of the training examples, and physical reality is rarely considered in this line of thinking.

Initial Surface Alignment. The endoscopogram surface and the CT surface live in two unrelated Euclidean spaces respectively: the camera space and the CT imaging system coordinate space. They therefore differ by a similarity transform. We need to first determine that transformation before solving for the aforementioned non-rigid deformation. However, an endoscopogram only portrays that portion of the pharyngeal surface that is present in the endoscopic video, so the initial alignment (transformation, rotation, scaling) between this partial surface and a complete CT segmentation surface is made difficult. Simple alignment methods, such as Iterative Closest Point (ICP), will not suffice because of the additional large non-rigid deformations.

1.4 A Brief Outline of the Proposed Methods

With all the aforementioned challenges in mind, in this dissertation I investigate the surface registration problem under the context of CT/Endoscope fusion from three major perspectives: geometry, physics and statistics. With the consideration of all three aspects, my proposed method, Thin Shell Demons, can efficiently solve the surface registration problem in our application.

1.4.1 Geometric Feature Descriptor

In our application, surface registration is better understood as matching/aligning corresponding anatomical structures between the two surfaces. Since different anatomical structures present distinct geometric appearance, I have designed a novel geometric feature descriptor that can capture rich curvature information of local shapes. By matching the descriptors between the two surfaces, we can achieve reasonably good initial correspondences. Moreover, this feature descriptor can also be seamlessly built into a widely used surface matching framework, named spectral graph matching. I show that using the descriptor can achieve better matching results compared to the original spectral graph matching. In addition, this feature descriptor can also be integrated with the

following Demons-based surface registration method to produce explicit deformation vector fields. I show that the feature matching strategy for surface registration can effectively handle partially overlapping surfaces with missing data and topology change.

1.4.2 Thin Shell Physical Model

We seek to produce physically plausible deformations from one surface to the other. This may require complex 3D physical modeling of the pharyngeal mechanics. To simply the problem, I propose to approximate 3D mechanics using the thin shell model, which concentrates only on the physical properties of 2D surface geometry. Results have shown that the thin shell model can produce reasonable registration results in our application while having a low computational cost. Based on the thin shell physical model, I propose a pairwise surface registration method, called Thin Shell Demons (TSD), that can be used to register the endoscopogram to the CT surface. In general, TSD constructs some virtual attraction forces between the two surfaces using geometric feature matching and computes a physically plausible deformation under the attraction. The method is robust against partial data and surface topology change.

1.4.3 Surface Disparity Estimation

The above TSD method only computes surface deformations without explicitly estimating the disparity between the CT surface and the endoscopogram. However, an explicit quantification of the disparity regions can be beneficial in handling more complicated situations, such as the partial volume effect and tissue collapse. To that end, I propose a joint estimation framework that can simultaneously estimate deformations and incompatible regions between surfaces. On the one hand, the estimated disparity renders feature matching more robust between surfaces. On the other hand, it can also be used to assign a certainty measurement to the endoscopy-to-CT tumor transfer for radiation treatment planning.

1.4.4 Groupwise TSD

By mimicking the idea of N-body interaction, I extend TSD into a groupwise framework, where the attraction forces are computed among N surfaces and can gradually deform all the surfaces towards an implicit mean shape. The advantages of groupwise TSD over other groupwise frameworks include its template-free nature and the ability to handle partially overlapping surfaces with missing data and different topology. Groupwise TSD is well suited for registering many single-frame-based reconstructions into a unified piece of geometry (endoscopogram).

1.4.5 Anisotropic Elasticity Learning

The thin shell model used in TSD assumes the surface to be a homogeneous elastic material, but we realize that human tissues are mostly inhomogenous and anisotropic, which means the tissue elasticity varies at different locations and along different directions. A direct stiffness test is challenging in our application, but the endoscopic video of each patient contains a lot of throat motion for that specific patient, which can be helpful in inferring possible anisotropic tissue elasticity. I propose a probabilistic framework to learn local anisotropic elasticity parameters using deformations extracted from the video. For simplicity the original thin shell model is extended into the orthotropic case, where 4 stiffness parameters are computed for each local region via an optimization over a Markov Random Field constructed on the surface. The resulting elasticity parameters are smoothly varying across the surface and are shown to produce more realistic deformations compared to the isotropic and homogeneous model. Furthermore, I show that by using an Expectation-Maximization algorithm we can jointly estimate surface deformations and elasticity parameters in a groupwise registration setting.

1.5 Thesis and Contributions

Thesis: *Geometric information and physical modeling can improve surface registration results in the context of CT/endoscopy fusion. The anisotropic parameters in the physical model can be inferred probabilistically from a set of observed material deformations.*

The contributions of this dissertation are as follows:

1. A novel geometric feature descriptor has been designed for discrete triangle meshes. This high-dimensional descriptor can capture multiscale curvature information of local shapes and provide the opportunity for advanced feature matching for matching pharyngeal shapes.

2. A geometric-feature-based spectral graph matching method has been proposed to compute dense correspondences between two partial surfaces. The method makes use of the aforementioned feature descriptor to produce initial correspondences for an improved joint spectral graph matching.

3. A surface registration method, called Thin Shell Demons, has been proposed to efficiently deal with partial surfaces with different topology. The method uses both geometric feature matching and a thin shell deformation model to produce physically plausible surface deformations.

4. A joint estimation framework based on Monte-Carlo Expectation-Maximization has been proposed to simultaneously estimate deformations and incompatible regions between the two surfaces.

5. A groupwise surface registration framework based on TSD has been proposed to robustly register many partially overlapping surfaces with missing patches. The framework mimics the N-body interaction principle and is template-free to deal with the challenge of mean surface estimation.

6. An anisotropic thin shell model has been studied, and a discretization of the model has been designed for triangle meshes. A probabilistic framework has been proposed to infer anisotropic elasticity parameters from a set of training deformations.

7. Based on the above elasticity learning framework, a joint estimation framework has been proposed to simultaneously estimate surface deformations in a groupwise registration setting and the elasticity parameters.

Besides the above methodological contributions, I have also accomplished the following engineering contributions:

1. A similarity transform fitting method has been designed to robustly compute an initial alignment between the partial endoscopogram surface and the complete CT surface.

2. A texture fusion algorithm has been designed to consistently stitch texture patches from multiple frames into a unified endoscopogram texture.

3. An overall pipeline for building an endoscopogram has been designed to effectively integrate reconstruction, geometry registration and texture fusion.

With the above scientific and engineering contributions, tumor transfer from an endoscopic movie to the CT space is enabled. To conclude, this dissertation develops necessary image computing techniques for making potential improvements in radiation treatment planning of pharyngeal cancer.

1.6 Overview of Chapters

The remainder of this dissertation is organized in the following chapters: Chapter 2 reviews necessary mathematical, mechanical and registration algorithm backgrounds for understanding this dissertation. Chapter 3 describes the novel geometric feature descriptor and its use in a joint spectral graph matching method. Chapter 4 describes the proposed Thin Shell Demons surface registration method. Chapter 5 describes the endoscopogram reconstruction pipeline with a focus on the groupwise extension of TSD. Chapter 6 describes two frameworks related to anisotropic elasticity estimation. Chapter 7 revisits the TSD algorithm with a focus on the CT/endoscopogram registration problem and describes the joint registration/disparity estimation framework.

CHAPTER 2

Background

This chapter presents some background required in this dissertation. In particular, mathematical background of differential geometry is briefly reviewed in Section 2.1. Background for surface registration algorithms is discussed in Section 2.2. Finally, some basic physics for 2D/3D elastic materials is reviewed in Section 2.3.

2.1 Geometry of Surfaces

This dissertation has to deal with surfaces represented by noisy discrete triangle meshes. However, I will first review properties of continuous smooth curved surfaces, which are mathematically defined as 2-dimensional *differentiable manifolds*. This will facilitate analysis of practical real-life problems, and we can in turn draw many useful insights from some of the well-developed theories. In this section, I will briefly review some mathematical definitions and properties of differentiable manifolds. Without loss of generality, we assume the manifold is 2-dimensional (surfaces), but the theories can be applied to higher dimensional manifolds.

2.1.1 Differentiable Manifold

The key entities investigated in this dissertation are smooth curved surfaces, such as endoscopic reconstruction surfaces and CT segmentation surfaces. Such surfaces are more often mathematically understood as *differentiable manifolds*. A **Differentiable Manifold** is a specific type of *topological manifold*. The local surface is constrained to smoothly resemble a 2D Euclidean space, a 2D plane.

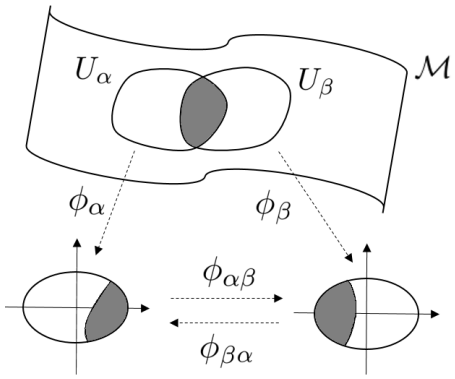


Figure 2.1: Charts on a 2D manifold. Two surface patches U_α, U_β are respectively parameterized via ϕ_α, ϕ_β.

Formally, a differentiable manifold is a topological manifold equipped with an equivalence class of *atlases* whose transition maps are all differentiable.

An **atlas** on a topological space $\mathcal{M}$ is a collection of pairs $\{(U_\alpha, \phi_\alpha)\}$ called *charts*, where the U_α are open sets that cover $\mathcal{M}$, and for each index α,

$$\phi_\alpha : U_\alpha \to \mathbb{R}^2 \tag{2.1}$$

is a homeomorphism of U_α onto an open subset of 2-dimensional real space. In other words, a chart serves as a 2D parameterization of the local surface onto $\mathbb{R}^2$. The **transition maps** of the atlas are the functions

$$\phi_{\alpha\beta} = \phi_\beta \circ \phi_\alpha|_{\phi_\alpha(U_\alpha \cap U_\beta)} : \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\alpha \cap U_\beta). \tag{2.2}$$

If the atlas is C^1 (all $\phi_\alpha, \phi_{\alpha\beta}$ are C^1 differentiable), it is called a differentiable structure of a differentiable manifold. Intuitively, this definition (Fig. 2.1) reflects the notion of "patching together pieces of flat spaces to make a manifold". Different atlases (patchings) may produce "the same" manifold.

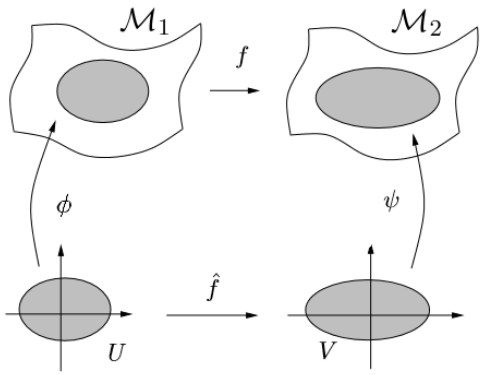


Figure 2.2: Local representation of a map between manifolds. f is a differentiable map between $\mathcal{M}_1$ and $\mathcal{M}_2$. $\hat{f}$ is the corresponding map represented on local charts.

2.1.2 Differentiable Maps

In this dissertation I am interested not only in the surface properties themselves but also in the matching process between two surfaces. The notion of matching, also referred to as registration, between two surfaces is mathematically understood as a mapping function that associates each point on the first surface to a point on the second surface. (The background of surface registration algorithms is given in Section 2.2.) Formally, let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two 2D differentiable manifolds. A map $f : \mathcal{M}_1 \to \mathcal{M}_2$ is said to be differentiable at a point $p \in \mathcal{M}_1$ if there exist parameterizations (U, ϕ) of $\mathcal{M}_1$ at p (i.e., $p \in \phi(U)$) and (V, ψ) of $\mathcal{M}_2$ at $f(p)$, with $f(\phi(U)) \subset \psi(V)$, such that the map

$$\hat{f} := \psi^{-1} \circ f \circ \phi : U \subset \mathbb{R}^2 \to \mathbb{R}^2 \tag{2.3}$$

is smooth. As Fig. 2.2 shows, this definition decomposes a global mapping function into local mapping functions parameterized on local pieces (charts).

As coordinate changes are smooth, this definition is independent of the parameterizations chosen at $f(p)$ and p. A differentiable map $f : \mathcal{M}_1 \to \mathcal{M}_2$ between two manifolds is called a **diffeomorphism** if it is bijective and its inverse $f^{-1} : \mathcal{M}_1 \to \mathcal{M}_2$ is also differentiable.

2.1.3 Tangent Spaces

The two previous subsections should have given a sense that the trick to differential geometry is to parameterize everything on local coordinate systems. For example, as I will study in later chapters, a mapping between two surfaces is sometimes interchangeable with the notion of a deformation field. It is a known fact that the global deformation of a surface can be represented by a collection of local stretching parameters and local bending parameters. It turns out that tangent spaces are the most commonly used local flat spaces for such local parameterizations.

Recall from elementary vector calculus that a vector $\mathbf{v} \in \mathbb{R}^3$ is said to be *tangent* to a surface $\mathcal{M}$ at a point $p \in \mathcal{M}_2$ if there exists a differentiable curve $c : (-\epsilon, \epsilon)$ such that $c(0) = p$ and $\dot{c}(0) = \mathbf{v}$. The set $T_p\mathcal{M}$ of all these vectors is a 2-dimensional vector space, called the tangent space to $\mathcal{M}$ at p, which can be identified with the plane in $\mathbb{R}^3$ that is tangent to $\mathcal{M}$ at p.

Now we can parameterize a local manifold patch onto the tangent plane. Choosing a parameterization $\phi : U \subset T_p\mathcal{M} \to \mathcal{M}$ around p, the curve c on the manifold can be mapped back to the tangent plane as $\hat{c}(t)$. Its coordinates on the tangent plane are given by

$$\hat{c}(t) := (\phi^{-1} \circ c)(t) = (x^1(t), x^2(t))). \tag{2.4}$$

Taking its derivative at point p, we can write

$$\dot{c}(0) = \dot{x}^1(0)(\frac{\delta}{\delta x^1})_p + \dot{x}^2(0)(\frac{\delta}{\delta x^2})_p, \tag{2.5}$$

where $(\frac{\delta}{\delta x^i})_p$ denotes the directional derivative operator associated to the vector tangent to the curve c_i at p. It can be proven that the tangent plane $T_p\mathcal{M}$ is a 2D plane spanned by $\{(\frac{\delta}{\delta x^1})_p, (\frac{\delta}{\delta x^2})_p\}$

We now can talk about a differential map between two manifolds in terms of the collection of local transformations defined on the tangent plane. Let $f : \mathcal{M}_1 \to \mathcal{M}_2$ be a differentiable map

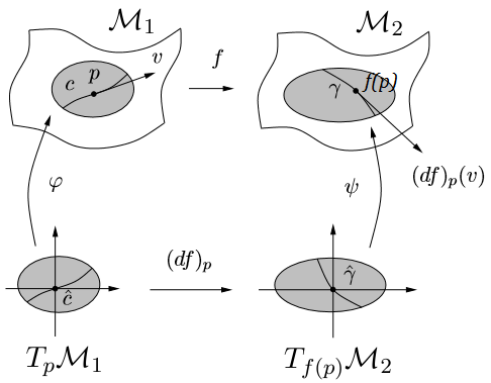


Figure 2.3: The derivative of a differentiable map f is a linear transformation $(df)_p$ that maps a tangent vector $v \in T_p\mathcal{M}_1$ to $(df)_p(\mathbf{v}) \in T_f(p)\mathcal{M}_2$.

between smooth manifolds. For $p \in \mathcal{M}_1$, the derivative of f at p is the map

$$(df)_p := T_p\mathcal{M}_1 \to T_{f(p)}\mathcal{M}_2. \tag{2.6}$$

It can be proven that $(df)_p$ is a *linear* transformation. In other words, as shown in Fig. 2.3, $(df)_p$ transforms a tangent vector $\mathbf{v} \in T_p\mathcal{M}_1$ into another tangent vector $(df)_p(\mathbf{v}) \in T_p\mathcal{M}_1$.

2.1.4 Riemannian Manifolds

As mentioned in the last subsection, the deformation of a surface can be classified into two categories: stretching and bending. Between those two, the stretching deformation will always alter the distance measurement between two points on the surface; this is an important phenomenon in this study. In the standard 3D Euclidean space the metric properties of distances are determined by the canonical Cartesian coordinates. In a general differentiable manifold, however, there are no such preferred coordinates to define distances, angles and volumes. We must rely on a new structure, a special tensor field called the *Riemannian metric*.

A **Riemannian metric** on a manifold $\mathcal{M}$ is a function that smoothly assigns to each point $p \in \mathcal{M}$ a symmetric positive definite 2-tensor g on the tangent space $T_p\mathcal{M}$. The tensor g is called a

metric tensor; it defines how distances are measured locally. If $x : V \to \mathbb{R}^2$ is a local chart, we have

$$g = \sum_{i,j=1}^{n} g_{ij} dx^i \otimes dx^j, \tag{2.7}$$

in V, where dx^i is the 1-norm associated with $(\frac{\delta}{\delta x^i})$ and

$$g_{ij} = \langle (\frac{\delta}{\delta x^i}), (\frac{\delta}{\delta x^j}) \rangle. \tag{2.8}$$

A smooth manifold $\mathcal{M}$ equipped with a Riemannian metric g is called a **Riemannian manifold**, and is denoted by $(\mathcal{M}, g)$.

A Riemannian metric allows us to compute the length $||v|| =< v, v >^{\frac{1}{2}} = \sqrt{g_p(v, v)}$ of any vector $\mathbf{v} \in T_p\mathcal{M}$. Therefore we can measure the length of a curve $c : [a, b] \to \mathcal{M}$ by integrating along its path

$$l(c) = \int_a^b ||\dot{c}(t)|| dt. \tag{2.9}$$

If two manifolds are equipped with identical Riemannian metrics, they are regarded as *isometric*. Formally, let $(\mathcal{M}_1, g)$ and $(\mathcal{M}_2, h)$ be two Riemannian manifolds equipped with metrics g and h respectively. A diffeomorphism $f : \mathcal{M}_1 \to \mathcal{M}_2$ is said to be an isometry if $f \circ h = g$, i.e., the metrics are identical after the mapping. This also means that f doesn't change distance measurement between the two manifolds, i.e., $d(p, q) = d(f(p), f(q))$, where $d(\cdot, \cdot)$ is the distance function between two points on the manifold.

The algebraic form of the metric tensor g is different under different parameterizations. For a point $p \in \mathcal{M}$, a special parametrization $\phi_p : U \to \mathbb{R}^2$ is called canonical if the metric tensor g_p is the identity matrix. Now let us consider a diffeomorphic map $f : \mathcal{M}_1 \to \mathcal{M}_2$ between two Riemannian manifolds. For a pair of correspondences $p \to q$, $p \in \mathcal{M}_1$ and $q = f(p) \in \mathcal{M}_2$, let g_p and g_q be their respective metric tensors under canonical parameterization. We have the following transformation:

$$g_p = J_p^T g_q J_p, \tag{2.10}$$

where J_p is the Jacobian of $(df)_p$ (derivative of f at p, which is a linear transformation defined on $T_p\mathcal{M}_1$). In particular, the matrix $J_p^T J_p$ characterizes how an infinitesimal circle is deformed into an infinitesimal ellipse on the tangent plane. In physics, this matrix is understood as a tensor

$$\epsilon = J_p^T J_p - I, \tag{2.11}$$

which is known as the *Cauchy-Green tangential strain tensor*. It characterizes how a surface is locally stretched.

2.1.5 Levi-Civita Connection

In Chapter 5, in order to study the anisotropic directional information for elastic materials, the notion of smoothness of a vector field has to be defined on a manifold. In the Euclidean space, one way to measure the "smoothness" of vector fields is through the notion of directional derivatives. To be more specific, if X and Y are two vector fields in Euclidean space, we can define the directional derivative $\nabla_X Y$ of Y along X. Parallel transport of a vector $\mathbf{v}$ along X is expressed as $\nabla_X \mathbf{v} = 0$, which means the direction of $\mathbf{v}$ keeps unchanged during the transport. This definition, however, uses the existence of Cartesian coordinates, which no longer holds in a general manifold. To overcome this difficulty we must introduce more structures.

The **Levi-Civita Connection** generalizes the idea of taking directional derivatives to manifolds. Intuitively, the parallel transport of a vector from one tangent plane to another on a manifold is done by first "unfolding" the local manifold to a flat space, then transporting the vector in a Euclidean fashion and finally "folding" the flat space back to a manifold. Formally, the derivative vector $(\nabla_X Y)_p \in T_p\mathcal{M}$, known as the covariant derivative of Y along X, is computed as

$$\nabla_X Y = \sum_{i=1}^{n} (X \cdot Y^i + \sum_{j,k=1}^{n} \Gamma_{jk}^i X^j Y^k) \frac{\delta}{\delta x^i}, \tag{2.12}$$

where the form X^i indicates the i^{th} coordinate of the vector field X, and Γ^i_{jk} are the *Christoffel symbols* for the Levi-Civita connection:

$$\Gamma^i_{jk} = \frac{1}{2}\sum_{l=1}^{n} g^{il}(\frac{\delta g_{kl}}{\delta x_j} + \frac{\delta g_{jl}}{\delta x_k} - \frac{\delta g_{jk}}{\delta x_l}) \tag{2.13}$$

where $g^{ij} = (g^{-1}_{ij})$.

With this definition of the Levi-Civita connection, we can talk about taking the **covariant derivative** of a vector field along a differentiable curve c. Let $V(t)$ be a vector field on a differentiable curve c such that $V(t) \in T_{c(t)}\mathcal{M}$. Its covariant derivative along c is given by

$$\frac{DV}{dt}(t) := \nabla_{\dot{c}(t)}V = (\nabla_X Y)_{c(t)}. \tag{2.14}$$

Then a vector field V defined along a curve c is said to be *parallel* along c if $\frac{DV}{dt}(t) = 0$.

With the definition of directional derivative, we can also generalize the notion of a "straight line" to manifolds. To that end, a curve c is called a **geodesic** if $\dot{c}$ is parallel along c, i.e., if $\frac{D\dot{c}}{dt}(t) = 0$. Intuitively, geodesics serve as straight lines in Riemannian geometry. They are locally distance-minimizing paths; The *geodesic distance* $d(p, q)$ between two points p and q of $\mathcal{M}$ is defined as the infimum of the length taken over all continuous, piecewise continuously differentiable curves.

2.1.6 Curvature

Finally, the notion of surface "curvedness" is constantly used throughout this dissertation. In the registration task, in order to build correspondences between two surfaces, similarly curved surface patches have to be first detected. Curvatures are well-defined for spatial curves. However, as discussed in Section 2.1.3, for a point $p \in \mathcal{M}$, any tangent vector $\mathbf{v} \in T_p\mathcal{M}$ can be associated with a spatial curve $c : (-\epsilon, \epsilon)$ such that $\dot{c}(0) = \mathbf{v}$, which means that the curvatures are different at p along different tangent vector directions.

It turns out that the curvature at a point p can be fully encoded by the *second fundamental form* $\mathrm{II}(\mathbf{u}, \mathbf{v})$, algebraically represented by a 2×2 symmetric matrix:

$$[\mathrm{II}_{ij}] = \begin{bmatrix} \omega^2\langle \mathbf{f}_1 \rangle & \omega^1\langle \mathbf{f}_1 \rangle \\ \omega^2\langle \mathbf{f}_2 \rangle & \omega^1\langle \mathbf{f}_2 \rangle \end{bmatrix} \tag{2.15}$$

where $\{\mathbf{f}_1, \mathbf{f}_2\}$ stands for two orthogonal coordinate directions spanning the tangent plane, and $\omega^i\langle \mathbf{f}_j \rangle$ is the curvature component that quantifies the turn of the normal about $\mathbf{f}_i$ when moving along $\mathbf{f}_j$.

With this definition, $\mathrm{II}(\mathbf{v}, \mathbf{v})$ gives the curvature along direction $\mathbf{v}$ as $\mathbf{v}^T[\mathrm{II}_{ij}]\mathbf{v}$. The **principal curvatures** at p, denoted k_1 and k_2, are the maximum and minimum values of this curvature, and they are computed as the eigenvalues of $[\mathrm{II}_{ij}]$. The corresponding eigenvectors are the **principal directions**, vectors along which the principal curvatures are achieved. Note that both the algebraic form of $[\mathrm{II}_{ij}]$ and the coordinates of principal directions are dependent on the choice of local parameterization (orthogonal basis).

The *Gaussian curvature* at p is defined as the product of the two principal curvatures: $K = k_1 k_2$, and the *mean curvature* is defined as the mean of the two principal curvatures: $H = (k_1 + k_2)/2$. Gaussian curvature and mean curvature are both very important curvature measurements, as Gaussian curvature indicates the local surface type (hyperbolic, flat, elliptic) and mean curvature indicates local curvedness in average.

2.2 Surface Registration

The goal of surface registration (shape matching or alignment) is to find point-to-point correspondences between two or multiple geometric surfaces. This problem is a key algorithmic component in various tasks, such as 3D scan alignment and statistical medical shape analysis. Surface registration methods can be generally classified into two categories: rigid and non-rigid registration. Recently most research attention has been drawn to the non-rigid case because in many applications a rigid

transformation can not sufficiently registration deformable objects. In fact, non-rigid registration problems are more ill-posed and thus require more complicated models. Therefore, this section only gives the background of existing non-rigid surface registration methods.

2.2.1 Matching-Based Methods.

One major category of surface registration methods relies on explicitly identifying corresponding points between surfaces. Some early methods design hand-crafted feature descriptors (also called fingerprints, feature signatures) for surface geometry and perform feature matching to produce a set of corresponding points (Sun et al., 2009; Gatzke et al., 2005; Zaharescu et al., 2009). These methods usually can credibly find a set of correspondences at places that yield distinct and matchable feature descriptors. However, feature matching by itself may produce outliers, which can lead to an illegal or non-smooth global mapping between surfaces. Therefore, even though feature descriptors play an important role in the registration process, further mechanisms have to be incorporated to guarantee the correctness and smoothness of the global mapping. Higher-order graph constraints (Zeng et al., 2010) or higher-order Markov Random Fields (MRF) (Zeng et al., 2013) have been proposed, but these algorithms are NP-hard in nature and thus only feasible in small scale problems. In Chapter 3 I will introduce a new geometric feature descriptor that can help the registration and will show that it can be combined with other methods to produce smooth mappings.

2.2.2 Embedding-Based Methods.

Another family of surface registration methods is based on *surface embedding*. The core of such methods is to embed the two surfaces into some other domain, where the matching can be naively solved. Usually, these methods are feasible only when the two surfaces to be registered satisfy certain geometric constraints. For example, when the two surfaces are isometric (Section 2.1.2), Multidimensional Scaling (MDS) can be adopted (Elad and Kimmel, 2003; Bronstein et al., 2006). The goal of the classical MDS (Borg and Groenen, 2005) embedding is to place the surface vertices in a low-dimensional Euclidean space such that the pairwise Euclidean distances among vertices in

the embedded domain are as close to the original pairwise geodesic distances as possible. This is usually done by first constructing an N-by-N affinity matrix storing pairwise geodesic distances for a surface with N vertices. Then an eigendecomposition is carried out on that affinity matrix (sometimes a transformed affinity matrix) to yield the optimal Euclidean embedding of the N vertices. When both surfaces are embedded in the same Euclidean space, a simple nearest-neighbor matching based on the embedded Euclidean coordinates is performed to yield the final matching results.

MDS along with the idea of eigendecomposition on affinity matrices is indeed a powerful tool for surface matching. However, MDS requires a dense affinity matrix for preserving geodesic distances among all pairs of vertices. From Section 2.1.4, we can see that isometry can also be defined via identical Riemannian metrics. Based on this notion, the spectral matching method has shown its success by using the Laplace-Beltrami operator instead of the dense affinity matrix to perform eigendecomposition (Zigelman et al., 2006; Balci et al., 2007; Reuter, 2010). Closely related to the Riemannian metric, the Laplace-Beltrami operator is defined only from local surface geometry, so its discrete version has a sparse matrix structure, allowing fast computation for surfaces with hundreds of thousands of vertices. Other variants along this line of thinking are based on graph Laplacian (Lombaert et al., 2011; Mateus et al., 2008) and Laplace-Beltrami functional spaces (Pokrass et al., 2013; Ovsjanikov et al., 2012). Generally speaking, all these methods rely on surfaces having the same intrinsic geometry (isometry) and can only handle bending deformations between surfaces. Moreover, they certainly can not handle more complicated issues such as surface topology change or the existence of missing surface patches (partial data).

Conformal mapping (Gu et al., 2004) and the Möbius transformation (Lipman and Funkhouser, 2009) can be used to deal with near-isometry situations by relaxing the geodesic-preserving constraint to the angle-preserving constraint. Quasi-conformal mapping (Lam et al., 2014; Zeng and Gu, 2011) has also been proposed to handle the most general diffeomorphic situation. However, these methods mostly require the surface to be genus-zero in order to be embedded into a common

disc/sphere domain, and they have also not been shown to be effective in dealing with partial data and complicated topology change.

To summarize, embedding-based methods can be surprisingly robust and fast as long as certain geometric constraints are satisfied. However, topology change and missing data are the two critical issues that challenge the feasibility of such methods in our application. In Section 3.3.3, I will discuss a variant of the spectral matching method that can moderately handle different intrinsic geometry.

2.2.3 Deformation-Based Methods.

As discussed in the previous section, embedding-based methods are inherently not suitable for dealing with different intrinsic geometry. A better way to understand our registration problem is to seek a deformation that carries one surface closer to the other. This will bypass the many geometric constraints needed in the embedding-based methods. Following this notion, the registration is formulated as an optimization over the set of possible deformations to minimize a function of two energy terms: data mismatch and deformation regularity. The first term, also known as data fidelity, encourages the optimized deformation field to carry one surface as close to the other as possible. The second term, also known as regularization, prefers to produce "realistic" surface deformations. Many approaches have been proposed to explore different formulations of these two terms.

Some works in registering 3D range scans (Pauly et al., 2005; Li et al., 2008) have been using the so-called closest-point rule, which refers to the strategy that each point on a surface is driven to the closest point on the other surface. Clearly, this strategy suffers with large deformations and inaccurate point positions. LDDMM (Bauer and Bruveris, 2011) and Currents (Vaillant and Glaunès, 2005) have provided another elegant mathematical framework that produces diffeomorphic deformations between surfaces by comparing their normal fields. However, this framework still can not handle missing surface parts.

Thirion's Demons registration algorithm (Thirion, 1998) has provides an alternate way for finding the deformation between object contours. The original concept in the first Demons regis-

tration paper (Thirion, 1998) states that a moving image can be attracted to a static image by the so-called virtual forces. (Pennec et al., 1999) further formulated the Demons method as a viscous fluid registration model that is solved in a greedy way. Yeo et. al. extended the Demons idea to spherical surfaces (Yeo et al., 2010). Dedner et. al. (Dedner et al., 2007) proposed a Demons framework for general surface registration using an implicit surface representation and transformed the surface registration problem into a 3D *image* registration problem. The Demons forces were determined from both intensity and intensity-level-surface curvature information, which was shown to be superior to the closest-point rule. Recently, Iglesias et. al. (Iglesias et al., 2013) developed a similar framework with implicit surface representation and a physically meaningful smoothness term by regarding the surface as a layer of an elastic material. However, the implicit presentation in these works requires embedding a surface into the ambient space using a signed-distance level-set function. However, voxelizing the 3D space can highly increase the data dimension and lacks the triangulation flexibility of surface meshes. Moreover, the signed-distance function itself is hard to track under large deformations.

Nevertheless, the Demons idea is still appealing because it has few assumptions about deformation/surface properties; it does not require surface completeness and identical topology. Based on this observation, I will introduce in Chapter 4 a physics-based surface registration method called Thin Shell Demons.

2.3 Thin Shell Mechanics

The deformation of the pharyngeal tissues are physical processes caused by surrounding muscles and the forces on these tissues from the endoscope. A realistic physical model is then needed in a registration framework to produce such deformations. In our case, an endoscopic movie only sees the pharyngeal surface and the only CT information relevant to registration is that surface. Therefore, I propose to only model surface deformations. In particular, I adopt the thin shell physical

model, which has been shown to be effective in the graphics literature (Gingold et al., 2004) to animate deformations of curved surface structures.

Thin shells are thin flexible structures with a high ratio of width to thickness. While their well-known counterparts, thin plates, relax to a flat shape when unstressed, thin shells are characterized by a curved undeformed configuration. To be more specific, the term shell is applied to bodies bounded by two curved surfaces, where the distance between the surfaces is small in comparison with other body dimensions (Fig. 2.4). The locus of points that lie at equal distances from these two curved surfaces defines the middle surface of the shell.

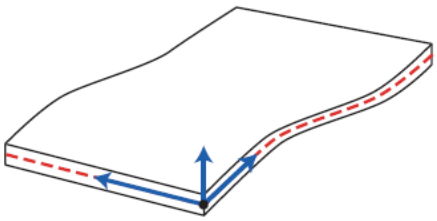


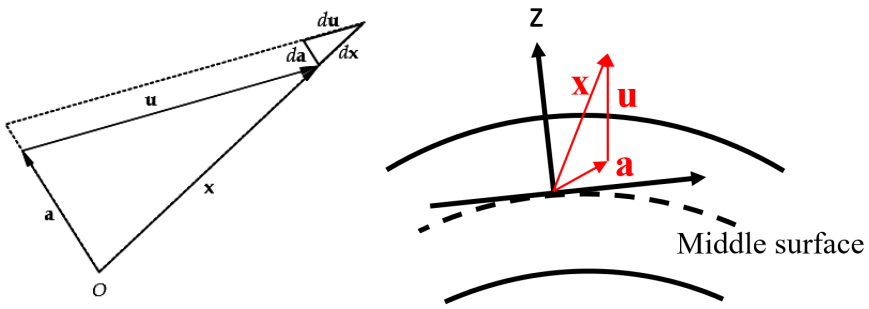
Figure 2.4: A thin shell structure.

For the purpose of analysis, a shell may be considered as a three-dimensional body, and the theory of linear elasticity may then be applied. However, this will generally be complicated and computationally demanding. In the theory of shells, an alternative simplified method is therefore employed. Based on some hypotheses made in the later sections, the 3D mechanics of a shell may be reduced to the analysis of its middle surface only. This also serves our purpose, which is to concentrate on the surface deformation instead of the whole 3D body.

For notational convenience in the following, we constrain subscripts denoted by Greek letters to have values in $\{1, 2\}$, and those denoted by Roman letters to have values in $\{1, 2, 3\}$.

2.3.1 Strain-Displacement Relation for Thin Shells

General 3D Strain. Strain refers the relative displacement of particles in an infinitesimal area of an object. Suppose an elastic deformation transforms a vector $\mathbf{a} = [a_1, a_2, a_3]$ into $\mathbf{x} = [x_1, x_2, x_3]$.



(a) (b)

Figure 2.5: (a) General situation: in a infinitesimal local region, a vector **a** is transformed into **x**. (b) Thin shell: the third coordinate direction (z-direction) overlaps with the normal direction.

In engineering, the strain **E** is a measurement of relative displacement $ds^2 - ds_0^2$, where ds_0 is the increment of initial length and ds is the increment of current length. Fig. 2.5 suggests the following relation

$$x_i = a_i + u_i,\ dx_i = da_i + du_i,\ ds = ||d\mathbf{a}||,\ ds_0 = ||d\mathbf{x}||, \tag{2.16}$$

where $\mathbf{u} = [u_1, u_2, u_3]$ is the displacement vector. Then it is easy to get

$$ds^2 - ds_0^2 = (u_{j,i} + u_{i,j} + u_{i,k}u_{j,k})da_i da_j = 2\varepsilon_{ij} da_i da_j, \tag{2.17}$$

where Einstein summation is used in the above equation over i,j and k. In Eq. 2.17 the notation $\bullet_{,i}$ represents the partial derivative with respect to the i^{th} coordinate. From this equation the Green-Lagrangian strain tensor is defined to be of the following form

$$\mathbf{E} = [\varepsilon_{ij}] = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{bmatrix} \tag{2.18}$$

where $\varepsilon_{ij} = u_{j,i} + u_{i,j} + u_{i,k}u_{j,k}$. The notion of strain allows the measurement of local elastic deformation. In fact, we can see that given an arbitrary vector **v**, $\mathbf{v}^T\mathbf{E}\mathbf{v}$ gives the squared length change along the **v** direction.

Thin Shell Strain Approximation. We place the coordinate system on a thin shell in such a way that the third coordinate direction (z-direction) overlaps with the normal and the first two coordinate directions span the tangent plane. We further assume the Love-Kirchhoff hypothesis is satisfied: the in-plane displacement is a linear function of the z-coordinate (out-of-plane coordinate): $u_\alpha = u^0_\alpha - z u_{3,\alpha}$, where u^0_α is the in-plane displacement (stretching) of the middle surface. With this hypothesis, Eq. 2.18 can be recast in the form (Ventsel and Krauthammer, 2001):

$$\varepsilon_{\alpha\beta} = \epsilon_{\alpha\beta} + z\kappa_{\alpha\beta} \tag{2.19}$$

where $\epsilon_{\alpha\beta} = \frac{1}{2}(u^0_{\alpha,\beta} + u^0_{\beta,\alpha})$ is the 2D strain tensor of the middle surface, which is also known as the Cauchy-Green tangential strain tensor (Eq. 2.11). $\kappa_{\alpha\beta} = \mathrm{II}'_{\alpha\beta} - \mathrm{II}_{\alpha\beta}$ is the shape operator (curvature) difference induced by the bending of the middle surface.

It has also been shown in (Ventsel and Krauthammer, 2001) that the strain tensor's out-of-plane terms $E_{3\alpha}, E_{\beta 3}, E_{33}$ can be negligible in the thin shell situation. In this way, the 3D strain of a thin shell can be reasonably approximated using the 2D stretching and bending strain of its middle surface.

2.3.2 Thin Shell Deformation Energy

Coupled with strain is a physical quantity called stress, which expresses the internal forces that neighboring particles of an elastic material exert on each other. The existence of stress will induce potential energy, which I call thin shell deformation energy in our case. The measurement of such deformation energy plays an important role in my work. As will be discussed later, we prefer in many situations a low energy configuration of material deformations.

Hooke's Law. In the modern theory of elasticity, the relation between strain and stress of an elastic material is summarized by Hooke's Law:

$$[\sigma_{11}, \sigma_{22}, \sigma_{12}]^T = C[\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{12}]^T, \tag{2.20}$$

where $\sigma_{\alpha\beta}$ and $\varepsilon_{\alpha\beta}$ are the in-plane stress and strain, and C is a 3×3 positive definite matrix, called a *stiffness matrix*, characterizing *local* elasticity.

In the most common isotropic case, the elastic property of a local point is the same along all directions, and the stiffness matrix has the following simplified form:

$$C = \frac{E}{1-\nu^2} \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & (1-\nu)/2 \end{bmatrix} \tag{2.21}$$

where E is Young's modulus, which represents the material's stiffness, and ν is the Poisson's ratio, which represents the material's compressibility.

For the isotropic case, it can be further shown that the local deformation energy of a point on a thin shell can be classified into two catergories: membrane (stretching) and bending. With Eq. 2.19, the membrane energy can be computed as

$$W_{membrane} = \frac{Eh}{2(1-\nu^2)}((1-\nu)\text{Tr}(\epsilon^2) + \nu(\text{Tr}\epsilon)^2), \tag{2.22}$$

where $\text{Tr}(\cdot)$ stands for the trace operation and h is the shell thickness. The bending energy has a similar form:

$$W_{bending} = \frac{Eh^3}{24(1-\nu^2)}((1-\nu)\text{Tr}(\kappa^2) + \nu(\text{Tr}\kappa)^2). \tag{2.23}$$

Then the total local energy $W = W_{membrane} + W_{bending}$. Note that the above energy computation is only related to a single local point on the shell. To get the total deformation energy of an entire shell, one has to integrate W over the area of the shell:

$$W_{total} = \int_S W dS \tag{2.24}$$

Orthotropy. In the theory of elasticity mechanics, the term *anisotropy*, as opposed to isotropy, implies different elastic properties in different directions. In that case, the stiffness matrix C can be

an arbitrary 3×3 positive definite matrix with 6 free parameters. As a matter of fact, human tissue is mostly anisotropic (Kroon and Holzapfel, 2008). Therefore, the study of anisotropic elasticity becomes essential in realistic physical modeling.

In particular, we focus on a special type of anisotropic material, called *orthotropic* material. For an orthotropic shell, the anisotropy of a point on the shell is symmetric w.r.t. two orthogonal axes, known as the *natural axes*, on the tangent plane of the middle surface. This leads to a stiffness matrix in the following form when $\sigma_{\alpha\beta}$ and $\varepsilon_{\alpha\beta}$ are parameterized under this natural-axes coordinate system:

$$C = \begin{bmatrix} c_1 & c_2 & 0 \\ c_2 & c_3 & 0 \\ 0 & 0 & c_4 \end{bmatrix} = \frac{1}{1-\nu_{12}\nu_{21}} \begin{bmatrix} E_1 & \nu_{21}E_1 & 0 \\ \nu_{12}E_y & E_2 & 0 \\ 0 & 0 & 2G_{12}(1-\nu_{12}\nu_{21}) \end{bmatrix}, \tag{2.25}$$

where the E_α are the Young's moduli along the natural axes, the $\nu_{\alpha\beta}$ are the Poisson's ratios, and G_{12} is the shear modulus.

Similar to the isotropic case, the local energy is the sum of the membrane and bending energy:

$$W = \frac{Eh}{2(1-\nu^2)}\epsilon^T C \epsilon + \frac{Eh^3}{24(1-\nu^2)}\kappa^T C \kappa, \tag{2.26}$$

where $\epsilon = [\epsilon_{11}, \epsilon_{22}, \epsilon_{12}]^T$ and $\kappa = [\kappa_{11}, \kappa_{22}, \kappa_{12}]^T$. Note that Eq. 2.26 only holds when $\epsilon_{\alpha\beta}$ and $\kappa_{\alpha\beta}$ are parameterized under the natural-axes coordinate system. If they are parameterized under an arbitrary orthogonal frame instead of rotated into the natural axes, a further rotation matrix has to be incorporated in Eq. 2.26. Details on this will be introduced in Chapter 6.

CHAPTER 3

Geometric-Feature-Based Spectral Graph Matching

The goal of surface registration is to seek a 3D deformation field that can align *corresponding regions* in the two surfaces. Since a surface represents the boundary of an object of interest, these *corresponding regions* are usually identified through distinguishable geometric shapes of local regions of the object. For example, in both the endoscopogram and CT surface, the epiglottis region appears as a convex curvy ridge, where the curvature along the sagittal direction is larger than that along the coronal direction. The pharyngeal wall in both surfaces is approximately cylindrical, where the curvature along the axial direction is very small. Therefore, the capability of describing local shapes is a fundamental building block of a successful surface registration method that seeks to match similar geometric structures.

Image-based feature descriptors (Lowe, 1999; Bay et al., 2006) have been successfully applied in many computer vision tasks, but they generally cannot be directly applied for triangle meshes. Spin images (Johnson and Hebert, 1999) and Heat Kernel Signatures (HKS) (Sun et al., 2009) are the two most widely used surface features that can capture local shape information. However, spin images are variant to large shape deformations because the feature descriptor is computed in the 3D ambient space around a local point. HKS is computed purely based on intrinsic geometry, i.e., geodesic properties between every two points in a patch, so it cannot handle non-isometric deformations. In the following section, I will present a novel geometric feature descriptor that can capture rich curvature information of local shapes, and based on that descriptor I introduce a new

spectral graph matching method that can produce accurate dense correspondences between two surfaces.

3.1 Geometric Feature Descriptor

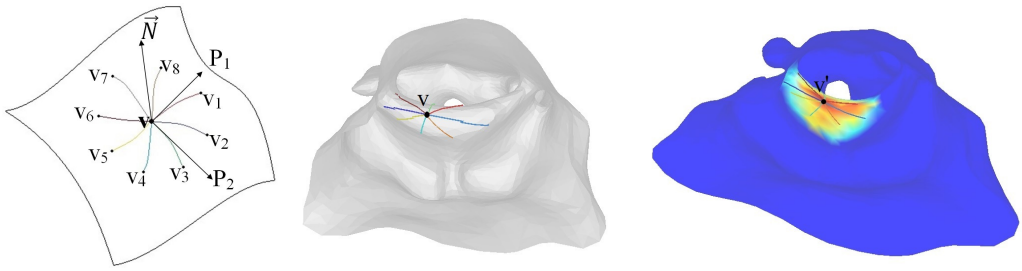
As discussed in Section 2.1, a surface is mathematically defined as a differentiable manifold $\mathcal{M}$. Without loss of generality, we assume the two principal directions and the associated two principal curvatures can be defined almost anywhere in $\mathcal{M}$, with finite singular points, known as umbilics, where the curvature is the same along all directions (Koenderink, 1990).

We design a novel feature descriptor f to create a *geometric signature* $f(v)$ for a given point $v \in \mathcal{M}$. Since local shape can be described by curvatures measured at different scales in a local region, the feature descriptor f is designed to collect curvature information on both its own location and a number of surrounding locations. As shown in Fig. 3.1a, for a given point v, we find 8 surrounding points $\{v_i \mid i = 1...8\}$ by going along 8 equally angularly spaced geodesic directions $\{\mathbf{g}_i \mid i = 1...8\}$ from v by a certain distance d. The choice of the value of d will be discussed in Section 3.4. The two orthogonal geodesic directions $\mathbf{g}_1$ and $\mathbf{g}_3$ are respectively the two principal directions $\mathbf{p}_1$ and $\mathbf{p}_2$. In order to capture curvatures at different scales, the descriptor is defined based on the curvature and normal information collected at the 9 points in the local patch: $f(v) = (\mathbf{C}, \mathbf{S}, \Delta\mathbf{N}, \Delta\mathbf{N}_{1,5}, \Delta\mathbf{N}_{3,7})$. The signature $f(v)$ is detailed as follows.

The two principal curvatures k_1, k_2 can uniquely determine the local shape, namely the amount of surface bends in all directions. On top of that, Koenderink (Koenderink, 1990) defines another two curvature measures c, s, which are more informative in differentiating different type of shapes and are derivable from the two principal curvatures k_1, k_2 in the following way:

$$c = \sqrt{\frac{k_1^2 + k_2^2}{2}} \tag{3.1}$$

$$s = -\frac{2}{\pi} \arctan[(k_1 + k_2)/(k_1 - k_2)] \tag{3.2}$$



(a) A local surface patch (b) $\mathcal{M}_1$ (c) $\mathcal{M}_2$

Figure 3.1: (a) Local geometry from which the geometric signature $f(v)$ is computed. (b) A vertex v, indicated as the cross point, is selected in $\mathcal{M}_1$. The feature descriptor (the same one as shown in part (a)) is deployed on v to get the geometric feature vector $f(v)$. (c) The value of the i^{th} row in the confidence score array Δ is plotted on $\mathcal{M}_2$ (red indicates large value). $v' \in \mathcal{M}_2$ is regarded as the most similar point to $v \in \mathcal{M}$.

Intuitively, c indicates the level of curvedness, and s indexes shape types, i.e., being convex, concave, hyperbolic or parabolic. These two curvatures are computed at v and the 8 surrounding points $\{v_i \mid i = 1...8\}$ to describe local curvatures, i.e., $\mathbf{C} = c \cup \{c_i \mid i = 1...8\}$, $\mathbf{S} = s \cup \{s_i \mid i = 1...8\}$. Larger scale measures of curvature between each of the surrounding points and the center are computed as the magnitudes of the normal direction difference between each of the end points and the center point, i.e., $\Delta\mathbf{N} = \{||\mathbf{n} - \mathbf{n}_i||_2 \mid i = 1...8\}$. Finally, normal direction differences between two extreme endpoint pairs (v_1, v_5) and (v_3,v_7) are computed to describe the general shape structure, i.e., $\Delta\mathbf{N}_{1,5} = ||\mathbf{n}_1 - \mathbf{n}_5||$, $\Delta\mathbf{N}_{3,7} = ||\mathbf{n}_3 - \mathbf{n}_7||$.

Discretization. In the discrete setting, a surface is represented as a triangle mesh $\mathcal{M}$. With some abuse of notation, the feature descriptor is applied to each vertex v in the mesh. In this work, geometric properties like normal directions, principal directions and curvatures are estimated for all the vertices using the *Computational and Geometry Algorithms Library* (CGAL). The 8 geodesic directions are sampled on the tangent plane, and geodesic marching is performed by the discrete Levi-Civita parallel transport (Crane et al., 2010). Since in the discrete case the end point of a geodesic path can end up being an arbitrary point on a triangle, whereas curvatures can only be estimated at triangle vertices, I choose the nearest vertex to the end point of the geodesic path to be $\{v_i \mid i = 1...8\}$. Fig. 3.1b shows a case where the descriptor is deployed on a vertex of a triangle mesh.

3.2 Feature Matching

With the novel multiscale curvature feature descriptor f, we can perform traditional feature matching techniques to find corresponding regions between two surfaces. Here I propose a *feature distance* measurement for comparing the similarity between two feature vectors. A naive way is to adopt the L_2 metric on the feature vector space, and the distance between the two feature vectors of v and u is simply defined as $||f(v) - f(u)||_2$.

This definition naturally assumes that the order of the 8 surrounding points is consistent between v and u. However, the two principal directions are only uniquely defined up to a 180-degree rotation, which means that both $\{\mathbf{p}_1, \mathbf{p}_2\}$ and $\{-\mathbf{p}_1, -\mathbf{p}_2\}$ are valid principal direction pairs. In practice, this poses a problem that the principal directions might experience a sign change from place to place. Taking consideration of this, the distance is then defined as

$$\min\{||f(v) - f(u)||, ||f(v) - f^*(u)||\}, \tag{3.3}$$

where f^* represents the feature vector computed after rotating the principal directions by 180 degrees.

In our application, when the two surfaces are rigidly aligned, corresponding anatomical regions shouldn't be spatially too far apart. Therefore, a soft Euclidean threshold τ is added into the feature distance measurement, yielding

$$\delta(v, u) = \min\{||f(v) - f(u)||, ||f(v) - f^*(u)||\} + \alpha(1 + e^{-(||x_v - x_u|| - \tau)})^{-1}, \tag{3.4}$$

where the second part is a sigmoid function penalizing a too large Euclidean distance $||x_v - x_u||$ between two corresponding vertices and where α is a weighting factor. With this feature distance measurement, we can construct an $N_1 \times N_2$ distance 2D-array $\mathbf{D}$ between two surfaces $\mathcal{M}_1, \mathcal{M}_2$ with N_1 and N_2 vertices respectively.

Normalizing the Distance Array. Here an additional step is to normalize the distance array to be a *confidence array*. When used for computing credible initial correspondences, the distance measurement is not commensurable across different regions. For example, the matching between two flat regions is highly ambiguous, but the feature distance is likely to be zero due to the small local curvatures. Distinguishable curved shapes are less ambiguous to match, but these places tend to yield larger feature distances. To summarize, in the original distance array **D**, small feature distances don't correlate well with good matchings. To deal with this problem, I propose the following normalization technique.

The idea is to consider a confidence score that measures how likely $\{v_i \in \mathcal{M}_1, u_j \in \mathcal{M}_2\}$ is a pair in correspondence. We consider both-way corresponding likelihoods measured by $\delta^1_{i,j}$ and $\delta^2_{i,j}$ respectively. $\delta^1_{i,j}$ is defined as the likelihood of $u_j \in \mathcal{M}_2$ being the most similar vertex of $v_i \in \mathcal{M}_1$, compared to all other vertices in $\mathcal{M}_2$. A simple way to compute this quantity is to normalize the i^{th} row of **D** to the range of $[0, 1]$.

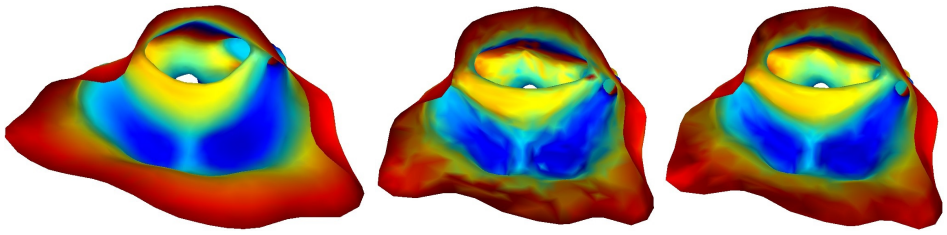
$$\delta^1_{i,j} = 1 - (\delta(i,j) - \min_k \delta(i,k))/(\max_k \delta(i,k) - \min_k \delta(i,k)). \tag{3.5}$$

$\delta^2_{i,j}$ is defined and computed vice versa:

$$\delta^2_{i,j} = 1 - (\delta(i,j) - \min_k \delta(k,j))/(\max_k \delta(k,j) - \min_k \delta(k,j)) \tag{3.6}$$

Because the two likelihoods are now at the same scale, the confidence score $\Delta_{i,j}$ is computed by taking the sum of $\delta^1_{i,j}$ and $\delta^2_{i,j}$. All the confidence scores will form a $N_1 \times N_2$ confidence score array Δ. As shown in Fig. 3.1b and Fig. 3.1c, for the vertex v_i selected in $\mathcal{M}_1$, the ith row in Δ is color-coded in $\mathcal{M}_2$. The vertex with the largest value is selected as the corresponding point. The overall dense correspondences based on this strategy are color-coded as shown in Fig. 3.2ab.

Selecting Most Credible Correspondences. After constructing the Δ array, I use an iterative max-row-column approach described in (Lipman and Funkhouser, 2009) to construct an optimal set of t initial correspondences. In each iteration, I select the largest non-zero element $\Delta_{i,j}$ in the array



(a) $\mathcal{M}_1$ (b) $\mathcal{M}_2$ (c) $\mathcal{M}_2$

Figure 3.2: (a) $\mathcal{M}_1$ is uniformly colored. The overall correspondences are indicated by the corresponding color in $\mathcal{M}_2$. (b) Correspondences derived from direct feature matching via confidence scores. (c) Correspondences derived from geometric-feature-based spectral graph matching.

Δ and add (v_i, v_j) to the initial correspondence set. To avert non-one-to-one correspondences, we zero out the ith row and jth column of Δ after each selection. I repeat this procedure t times to select the t most credible correspondences.

3.3 Geometric-Feature-Based Spectral Graph Matching

My first idea is to investigate the use of spectral graph matching in our application of pharyngeal surface registration. Spectral graph theory (Chung, 1997) has been applied to the surface matching problem in various applications (Lombaert et al., 2011). One advantage of the spectral method is that spectral representations of shapes are in fact invariant in isometry; that is, geodesic-preserving deformations of a shape preserve its spectral representation. This allows shapes of the same intrinsic geometry to be perfectly matched in the spectral domain. However, perturbations in shape isometry can cause inconsistencies in spectral representations and thereby harm the matching accuracy. Lombaert et al. (Lombaert et al., 2013) proposed a joint spectral graph matching method that can handle this inconsistency problem. In this section, I will review the method background and show that the matching accuracy can be further improved by incorporating the previously proposed geometric feature matching technique. Finally, I will discuss the disadvantage of the conventional spectral matching method in registering surfaces with different intrinsic geometry and show that the problem can be moderately solved by my proposed the improvement.

3.3.1 Spectral Graph Matching on an Association Graph

Two graphs $\mathcal{G}_1 = \{\mathbf{V}_1, \mathbf{E}_1\}$ and $\mathcal{G}_2 = \{\mathbf{V}_2, \mathbf{E}_2\}$ can be constructed from the two surfaces $\mathcal{M}_1$ and $\mathcal{M}_2$ with the vertices and edges of their triangle meshes. An *association graph* $\mathcal{G} = \{\mathbf{V}, \mathbf{E}\}$ is built by connecting $\mathcal{G}_1$ and $\mathcal{G}_2$ with a set of initial links. $\mathcal{G}$ is equipped with an $|N_1 + N_2| \times |N_1 + N_2|$ affinity array $\mathbf{W}$, where the affinities $w_{i,j}$ between vertex i and vertex j are measured using the Euclidean distance between two vertices in the original 3D space for both intra-surface links and inter-surface links; that is, $w_{i,j} = ||x_i - x_j||^{-2}$ if $\exists e_{i,j} \in E$. With $\mathbf{V}$ being the ordered pair $(\mathbf{V}_1, \mathbf{V}_2)$, $\mathbf{W}$ takes the form of

$$\begin{pmatrix} \mathbf{W}_1 & \mathbf{W}_{21} \\ \mathbf{W}_{12} & \mathbf{W}_2 \end{pmatrix} \tag{3.7}$$

where $\mathbf{W}_1, \mathbf{W}_2$ are intra-surface affinities and $\mathbf{W}_{12}$ and $\mathbf{W}_{21}$ are inter-surface affinities. The *graph Laplacian* operator $\mathbf{L}$ is defined as $\mathbf{L} = \mathbf{D} - \mathbf{W}$, where $\mathbf{D}$ is a diagonal array with $d_i = \sum_j w_{i,j}$.

The spectral decomposition of $\mathbf{L}$ provides an orthogonal set of eigenvectors $[\mathbf{u}^1, \mathbf{u}^2, ..., \mathbf{u}^{|N_1+N_2|}]$ with corresponding non-$decreasing$ eigenvalues. The first eigenvector is always constant with an associated zero eigenvalue; it is not useful for our matching purpose. Then the *spectral embedding* of the graph into a k-dimensional Euclidean space (spectral domain) is given by $[\mathbf{u}^2, \mathbf{u}^3, ..., \mathbf{u}^{k+1}]$. Formally, we define $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_k]$ as an $|N_1 + N_2| \times k$ array. Then the first k eigenmodes with non-zero eigenvalues $[\mathbf{u}^2, \mathbf{u}^3, ..., \mathbf{u}^{k+1}]$ is the solution to the following minimization problem:

$$\arg \min_{\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_k} \sum_{i,j=1}^{|N_1+N_2|} w_{i,j} \parallel f^{(i)} - f^{(j)} \parallel^2, \text{ with } \mathbf{F}^T\mathbf{F} = I \tag{3.8}$$

where $f^{(i)}$ is the i^{th} row of $\mathbf{F}$, representing the embedded Euclidean coordiantes (spectral representation) of the i^{th} vertex. Intuitively, the k eigenmodes define an embedding into a k-dimensional Euclidean space that tries to respect the edge lengths of the graph. In other words, the distance between neighboring vertices in the embedded domain, computed by $||f^{(i)} - f^{(j)}||$, is close to that in the original 3D space $||x_i - x_j||$.

Moreover, each eigenvector $\mathbf{u}^i$, known as the i^{th} vibration mode of graph $\mathcal{G}$, can be separated into two functions: $\mathbf{u}_1^i$, the first N_1 values of $\mathbf{u}^i$, representing the ith vibration mode of $\mathcal{G}_1$, and $\mathbf{u}_2^i$, the last N_2 values of $\mathbf{u}^i$, representing the ith vibration mode of $\mathcal{G}_2$. In fact, the aforementioned spectral embedding of the association graph $\mathcal{G}$ simultaneously embeds both graphs $\mathcal{G}_1$,$\mathcal{G}_2$ into the same spectral domain. As mentioned in (Lombaert et al., 2013), this association-graph-based spectral embedding provides several advantages over separate spectral embeddings. In particular, it solves the eigenvector permutation problem, namely that the order of eigenvectors often changes when they are associated with similar eigenvalues. Here, the inter-surface links of the association graph ensure that the combined eigenvector $\mathbf{u}^i$ includes consistent vibration modes $(\mathbf{u}_1^i, \mathbf{u}_2^i)$ from both graphs.

After knowing the spectral coordinates for all vertices, the final matching between the two surfaces is accomplished by a nearest-neighbor search in the k-dimensional spectral domain. Note that this method only yields a dense matching as a result but does not explicitly produce deformation fields between surfaces.

3.3.2 Geometric-Feature-Based Affinity Matrix

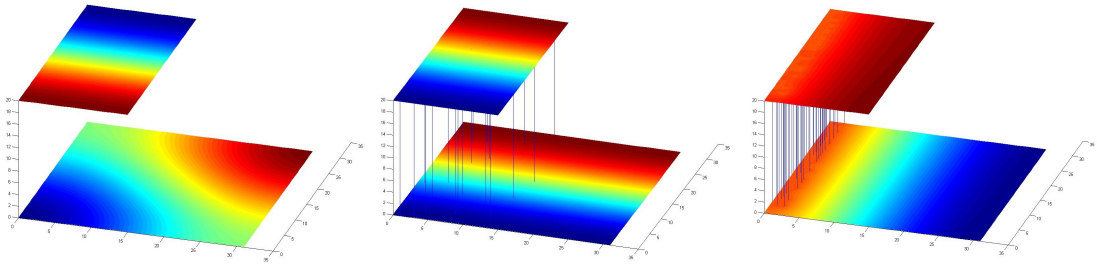
The inter-surface affinity in the Lombaert paper was defined according to the Euclidean distance between vertices, which is conceptually unnatural, because in most large deformation situations, two corresponding vertices might have a large Euclidean distance, ending up with a small affinity, even though there is a clear evidence showing the correspondence is correct and should have a high affinity. Therefore, I propose to compute the inter-surface affinity based on the confidence score of the initial correspondences derived by geometric feature matching. With t initial correspondences selected via the algorithm in Section 3.2, the affinity array W is now defined as

$$w_{i,j} = \begin{cases} ||x_i - x_j||^{-2} & \text{if } v_i, v_j \text{ are in same the surface,} \\ \Delta_{i,j} & \text{if } (v_i, v_j) \text{ is in the initial correspondence set,} \\ 0 & \text{otherwise.} \end{cases} \tag{3.9}$$

The final matching result is shown in Fig. 3.2c. As we can see, the correspondences are smoother than from the feature matching directly.

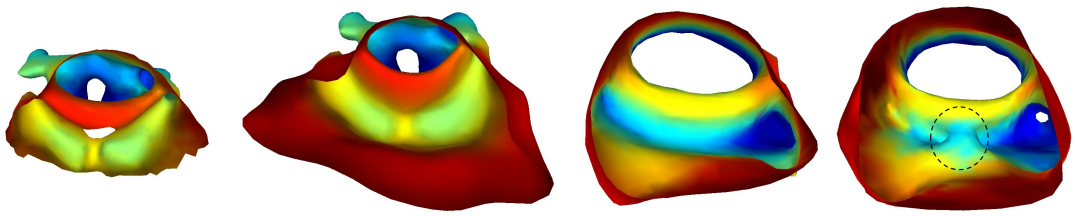
3.3.3 Different Intrinsic Geometry

In our application, the two surfaces have different intrinsic geometry, such as different boundary locations and holes. Conventional separated spectral decompositions (Lombaert et al., 2011) in this situation will yield two totally different sets of eigenmodes. Just think of the simplest partial surface problem in Fig. 3.3a, in which one surface is a half of the other one. The first eigenmodes have distinct patterns, because surfaces with different sizes have different vibration modes. However, if only 5% of all the initial correspondences are assigned between the two surfaces, as shown in Fig. 3.3b, the first eigenmodes become consistent with each other. Intuitively, a joint vibration can be achieved by associating the partial surface onto the other one using the initial links, so that the partial surface is forced to vibrate together with the other. Moreover, it is already obvious in the objective function (Eq. 3.8) that the energy is minimized when both intra-surface and inter-surface affinities are preserved in the spectral domain, which means corresponding vertices have similar embedded coordinates, as well as vibration properties.



(a) (b) (c)

Figure 3.3: (a) Separate spectral decompositions are respectively applied to two surfaces. The top surface is half of the bottom complete surface. The first eigenmode of each eigendecomposition is respectively color-coded on that surface. (b) A joint spectral decomposition is applied on an association graph connected by a sparse set of initial links. The color-coding shows the first joint eigenmode. (c) A joint spectral decomposition is applied on an association graph with initial links only on one side. **Conclusion**: only Fig. b shows the desired eigenmode for the matching between the two surfaces.



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

Figure 3.4: The color-coded correspondences (a,b) between a complete surface and a partial surface with a hole and truncation. (c,d) between surfaces with a bridge (circled region).

I found that the initial links have to be scattered all over the surface, but not necessarily densely. For example, the first vibration modes are as shown in Fig. 3.3c if the initial links are only on one side of the surface. Intuitively, two pieces of paper won't be stuck together if there is only one side glued together.

Therefore, it is essential to find a credible set of initial links. As will be shown in Section 3.4, conventional spectral matching is not able to provide correct correspondences (initial links). However, the novel geometric feature descriptor has the advantage of providing robust initial links regardless of whether the surface being partial or not, because the correspondences are derived only using local geometric features. For the same reason, in most situations where the partial surface has holes in it, the joint vibration can still be achieved. Figs. 3.4ab show the final matching result for a partial surface with a hole and a truncated boundary.

This geometric-feature based spectral method can also handle some other simple topology changes. However, in many cases, regions with complicated topology changes usually yield inconsistent geometric features, which makes the initial correspondences unstable. For example, as shown in Figs. 3.4cd, there is a bridge connecting the epiglottis and the pharyngeal wall, and the matching there is not reasonable.

3.4 Results

I tested different versions of the spectral graph matching method on 12 surface pairs created from 6 patients' CT data. The pharyngeal surface from the pharynx down to the vocal cord was

automatically segmented from each patient's CT image. Each surface has 2000-6000 vertices, with an approximately 2cm$\times$3cm elliptical cross section. We manually applied synthetic deformations to the surfaces with the help of a medical physicist to make them realistic, ending up with 12 surface pairs, two for each patient. The synthetic deformation includes the distortion and contraction of the pharyngeal wall and the closing and opening of the laryngeal region and of the epiglottis. We define the *vertex registration error* for each vertex as the Euclidean distance between its estimated corresponding vertex and its ground truth corresponding vertex. The *surface registration error* for each surface pair is defined as the average vertex registration error over all vertices.

3.4.1 Synthetic Data

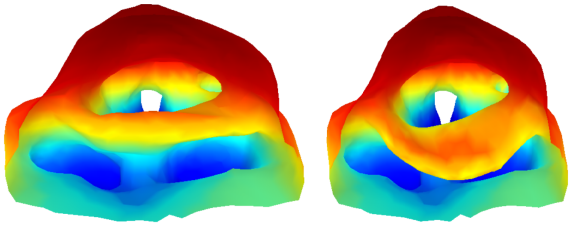
Complete Surface Registration. In the first experiment, the two surfaces in a surface pair are both complete. In other words, each vertex on a triangle mesh has a corresponding vertex on the other mesh. In this complete surface registration scenario, I tested 3 different options of the spectral graph matching method. In the first option, I used the joint spectral graph matching method proposed in (Lombaert et al., 2013); it first uses separate spectral graph matching to compute initial correspondences, followed by a joint spectral graph matching using Euclidean-distance-based inter-surface affinities. In the second option, I used a hybrid method between (Lombaert et al., 2013) and my proposed method. To be specific, the initial links were estimated by geometric feature matching (Section 3.3.2). However, the inter-surface affinity associated with these initial links were still computed as their Euclidean distances (length). The third option is the proposed method; the inter-surface affinity being the feature matching confidence score.

The average surface registration errors for the 12 surface pairs under the three matching options are shown in the right column of Table 3.1. The initial error was computed by the closest-point rule; that is, for each vertex on a triangle mesh, the closest vertex on the other mesh is regarded as its correspondence. This error metric relates to the amount of misalignment of surfaces before any registration. We can see that for the complete surface registration scenario the conventional spectral graph matching method (Option 1) can already produce reasonable results. Our feature-matching-

based initial links (Option 2) by itself can improve the accuracy by around 25%. The best results come from the combination of both of my proposed improvements (Option 3): feature-matching-based initial links and confidence-score-based inter-surface affinity. Fig. 3.5 shows a surface pair with color-coded correspondence estimation.

Table 3.1: Error for complete surface registration

	Average surface registration error (mm)
Initial Error	3.09±1.73
1. Conventional Spectral Graph Matching	1.83± 2.37
2. Feature + Euclidean Distance	1.38±2.55
3. Feature + Confidence Score	0.67±0.96



(a) (b)

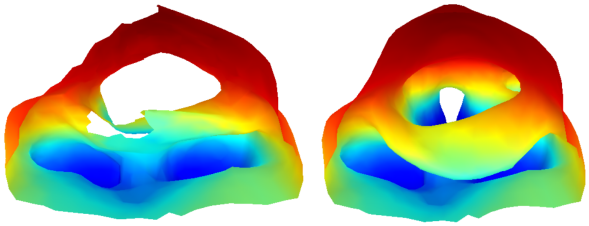
Figure 3.5: (a) A CT segmentation surface. (b) A synthetically deformed surface created from the surface in Figure a. The deformation includes the opening of the epiglottis and the contraction of the pharyngeal wall.

Partial Surface Registration. In the partial surface matching scenario, I picked one surface from each pair and manually created holes in large deformation regions, such as the epiglottis, and truncated the surface outside of a different locus. This simulates the missing data situation in real endoscopic reconstruction surfaces. The average surface registration errors are shown in the right column of Table 3.1. We can see that the conventional spectral graph matching method (Option 1) won't work for partial surfaces due its deficiency in dealing with different intrinsic geometry (Section 3.3.3). In contrast, the proposed feature-matching-based initial links and confidence-score-based inter-surface affinity can still produce reasonable results. In addition to Fig. 3.4, Fig. 3.6 shows one more case of partial surface registration.

The Bridging Situation. It has already been shown in Fig. 3.4d that topology change caused by the bridge between the tip of epiglottis and the pharyngeal wall can yield inaccurate matching results.

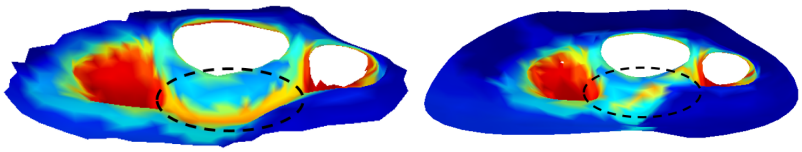
Table 3.2: Error for partial surface registration

	Average surface registration error (mm)
Initial Error	3.48±1.79
1. Conventional Spectral Graph Matching	3.26±6.71
2. Feature + Euclidean Distance	1.90±2.15
3. Feature + Confidence Score	1.15±1.36



(a) (b)

Figure 3.6: The same surface pair as shown in Fig. 3.5 except that some holes and truncation were added to (a).



(a) (b)

Figure 3.7: (a) A CT segmentation surface. (b) A bridge is manually created between the epiglottis tip and the pharyngeal wall. The circled regions indicate inaccurate matching.

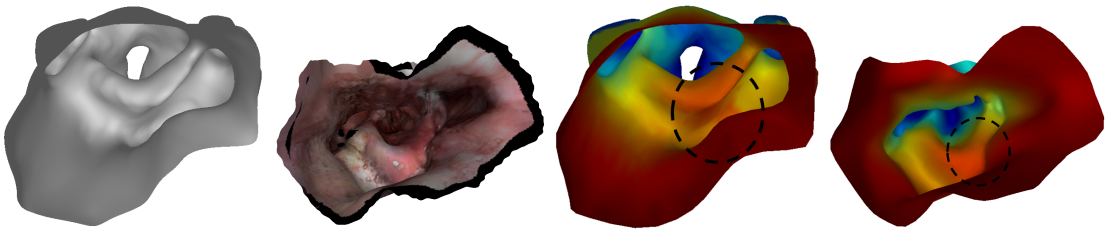
The reason why such topology change is more difficult to handle than the partial surface situation is that the bridging severely violates the isometry assumption; the geodesic distance between the epiglottis tip and the pharyngeal wall drastically changes after the bridging occurs; this also significantly changes the surface's vibration modes, thereby leading to unmatchable eigenvectors. Fig. 3.7 further shows a case where the matching fails at the bridging region.

Optimal Algorithm Parameters. I studied the choice of different parameter values. All the parameters were chosen to produce the lowest average surface registration error on a validation dataset, which contains two other surface pairs (different from the testing surface pairs). As a result, 15 eigenmodes were used to perform the final matching; the size of the initial correspondence

set was chosen as half the number of vertices; the geodesic path distance d was set to 4mm; the Euclidean distance threshold τ was set to 1cm.

3.4.2 Real Endoscopic Reconstruction Surfaces

I applied the improved spectral graph matching method to the registration between CT surfaces and endoscopic reconstruction surfaces. The algorithm parameters were chosen the same as the ones used in the synthetic data test. The endoscopic reconstruction surface in Fig. 3.8, which we call an *endoscopogram*, was produced by the pipeline introduced in Chapter 5. The endoscopogram was first rigidly aligned to the CT surface using an initial alignment method that will be discussed in Section 7.1. The matching result is shown in Fig. 3.8cd. We can see that the matching is generally reasonable except for the bridging region (circled in Fig. 3.8cd) where the tip of epiglottis touches the pharyngeal wall. This real data experiment further shows that the method cannot deal with severe topology change.



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

Figure 3.8: (a) A pharyngeal CT segmentation surface. (b) An endoscopic video reconstruction surface. (c,d) Color-coded correspondences between the CT surface (c) and the reconstruction (d). The circled regions indicate the bridging situation.

3.5 Conclusion

I have presented an improved spectral graph matching method for pharyngeal surface registration. A novel geometric feature descriptor along with a feature matching algorithm has been proposed to estimate a set of initial correspondences between two surfaces. An association graph based on

such initial correspondences produces more accurate joint spectral graph matching. I showed the method's potential to handle partial surface matching and discussed its disadvantages when dealing with complicated topology change. The results suggest that this approach might be applicable to other surface registrations with large deformations, holes and truncations but might fail in the situation of complicated topology change.

CHAPTER 4

Thin Shell Demons

A useful insight from last chapter is that the spectral-graph-matching-based method is inherently not suitable for handling complicated missing surface patches and topology change. A more reasonable solution is to seek a deformation that carries one surface closer to the other. As mentioned in Section 2.2.3, Thirion's Demons registration algorithm (Thirion, 1998) provides the initial motivation to our work. In Thirion's algorithm, a moving image can be attracted and gradually deformed to a static image by the so-called *virtual forces*. Unlike many other registration methods, it has very few underlying assumptions about deformation properties or surface completeness/topology, such as isometric invariance (Lombaert et al., 2013), local angle preservation (Lam et al., 2014), or identical compact support (Vaillant and Glaunès, 2005). Following the Demons algorithm, my idea is to regard one surface as being static and the other as being deformable. The deformable surface can be gradually attracted and deformed to the static one. When the two surfaces finally overlay with each other, the overlaid regions are regarded as corresponding regions. Unlike the spectral method, where surface intrinsic geometry is being matched in the spectral domain, the Demons-based method directly seeks a deformation field to align objects in the original 3D space, and thereby is little affected by surfaces with different topology. With this motivation in mind, we have to answer three questions.

How to define virtual forces? Dedner et. al. (Dedner et al., 2007) in their work have shown that curvature-derived demons force produces superior results than the traditional intensity-derived demons force. Moreover, curvature measures are even more important in our application. The key reason is that there are many easily identifiable geometric structures in the pharynx, including the

epiglottis, the arytenoids and the larynx. The pharyngeal wall has a less distinctive shape but still can produce matchable curvature patterns. In addition, our 3D reconstruction method (Price et al., 2016) is likely to produce more credible local geometry than global point positions. Therefore, I conclude that curvature information has to be considered in defining the virtual attraction forces.

What is an appropriate surface deformation model? Next, we have to understand how a surface can be deformed under the attraction forces. Since pharyngeal deformation is a physical process caused by surrounding muscles, we want our computed deformations to be physically plausible. A common physical modeling technique is the 3D Finite Element Model (FEM). In order to use FEM, one has to have extensive prior knowledge about the entire head-and-neck region, including the structural relationship among different anatomies and their physical properties. To this end, early work (Ki et al., 2006) has adopted a very coarse approximation to divide the throat into several cross-sectional regions with different stiffness, but this can hardly characterize different tissue types, such as the pharyngeal wall, the epiglottis, vocal cords, arytenoids, etc. Since in our application an endoscopic movie only sees the pharyngeal surface and the only CT information relevant to registration is that surface, I adopt a physical thin shell model to only simulate deformations for *surface* structures. The thin shell model, an essentially 2D model, is simpler to analyze than the 3D finite element method. Even though it cannot fully characterize tissue mechanics, the lack of 3D anatomical information for constructing a 3D FEM model makes the thin shell a reasonable choice that can provide physical reality to a certain extent.

Which surface is static/deformable? We have to determine the attraction direction between the two surfaces, namely which surface being static (attracting) and which surface being deformable (attracted). We realize that both the CT surface and the reconstruction surface may have bridges. Fig. 3.8 has already shown a case where the reconstruction algorithm fails to differentiate between the tip of epiglottis and the pharyngeal wall. Bridges are also created in the CT surface because fine geometric structures are blurred out in the CT image under the partial volume effect. However, the CT surface is more complete and smoother, whereas the reconstruction surface usually has a lot of

holes and noisy geometry. This fact favors the CT surface to be the deformable one because the physical deformation energy can be more reliably computed for smooth and complete surfaces.

To combine the above ideas, I propose a surface registration framework, *Thin Shell Demons* (TSD), that can leverage both geometry and mechanics. The general idea is to deform the surface to match geometrically similar structures between the two surfaces in a physically realistic way. To be specific, the CT surface is regarded as an elastic thin shell and is deformed using Demons forces determined by the geometric feature matching method introduced in Chapter 3. In addition, I propose to incorporate the thin shell model with cross-object structural links to preserve structures that are non-local w.r.t. distance along the shell.

4.1 Thin Shell Deformation Model

As discussed in Section 2.1.2, deformation-based registration methods rely on having a regularization term to produce smooth deformation fields. In TSD the regularization is formulated as the thin shell deformation energy; that is, we encourage low deformation energy configurations of the shell. This section introduces how to compute such energy for a triangle mesh and further proposes a novel structural energy to the existing thin shell model.

As reviewed in Section 2.3, the thin shell model has been commonly used to produce realistic deformations for surface structures. The deformation of a shell is regarded as the integration of local deformations in infinitesimal regions (Eq. 2.24). Local deformations can be further classified into stretching and bending. With some abuse of notation, we use $\mathcal{M}$ to denote both a thin shell and its middle surface. In the discrete case, a shell is represented as a triangle mesh.

4.1.1 Stretching Strain

Local stretching characterizes the area/length change in an infinitesimal region of the middle surface $\mathcal{M}$. It is quantified by the Cauchy-Green tangential strain tensor ϵ. Now, let us regard the deformation of $\mathcal{M}$ as a diffeomorphic map f between its original configuration $\mathcal{M}$ and its deformed

configuration $\mathcal{M}'$. Mathematically speaking, the tensor ϵ associated with a local point p is defined by

$$\epsilon = J_p^T J_p - I, \tag{4.1}$$

where I is an identity matrix and J_p is the Jacobian of tangential linear transformation $(df)_p$ (see also Section 2.1.4). Intuitively, it describes how an infinitesimal circle is mapped to an infinitesimal ellipse.

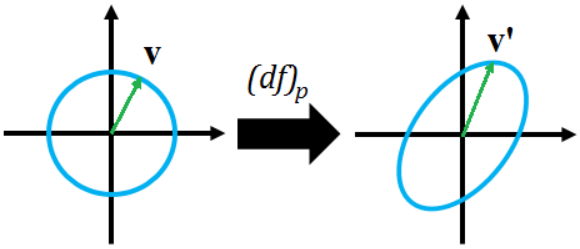


Figure 4.1: The Cauchy-Green strain tensor ϵ quantifies how an infinitesimal circle is deformed into an infinitesimal ellipse.

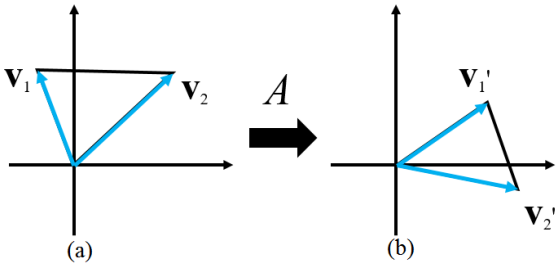


Figure 4.2: (a) A triangle before deformation. The two edges are represented as $\mathbf{v}_1, \mathbf{v}_2$ under an arbitrary 2D local coordinate system. (b) After deformation, the two edges are represented as $\mathbf{v}_1', \mathbf{v}_2'$ under another arbitrary 2D local coordinate system.

In a triangle mesh, we regard each individual triangle as the tangent plane of a local region. The tensor ϵ is estimated for each triangle using the method in (Zeng et al., 2013). Consider a triangle in its original configuration as shown in Fig. 4.2a. With an arbitrary 2D local coordinate system on the triangle plane, the two vectors associated with the two edges are parameterized by 2D vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ respectively. After the deformation the triangle ends up with the shape in Fig. 4.2b. With another 2D local coordinate system, the two deformed edges are parameterized by $\mathbf{v}_1'$ and $\mathbf{v}_2'$. Then

it is obvious that the transformation between the vectors is linear: $\mathbf{v}_1' = A\mathbf{v}_1$ and $\mathbf{v}_2' = A\mathbf{v}_2$. The linear transformation matrix A can be computed by $A = [\mathbf{v}_1', \mathbf{v}_2'][\mathbf{v}_1, \mathbf{v}_2]^{-1}$.

In fact, any other vector $\mathbf{v}$ in the undeformed configuration (Fig. 4.2a) will be mapped to a new vector in Fig. 4.2b via the same linear transformation: $\mathbf{v}' = A\mathbf{v}$. We can see that A is an algebraic realization of J_p under such a discretization. Compared to Eq. 4.1, the Cauchy-Green strain tensor for a triangle is naturally computed by

$$\epsilon = (A^T A - I). \tag{4.2}$$

Finally, the local stretching energy $W_{membrane}$ associated with a single triangle can be computed based on the stretching tensor ϵ (see also definition in Eq. 2.22):

$$W_{membrane} = \frac{Eh}{2(1-\nu^2)}((1-\nu)\text{Tr}(\epsilon^2) + \nu(\text{Tr}\epsilon)^2), \tag{4.3}$$

where E is the Young's modulus and ν is the Poisson's ratio. (Here we assume the shell thickness $h = 1$. As will be shown later, this thickness parameter will be considered by additional scaling factors.)

4.1.2 Bending Strain

The bending of a shell will induce local curvature change. This is quantified by a 2×2 bending strain matrix κ, which is computed by the shape operator change. (Normally in the literature the Greek character κ denotes curvature. Here we use it to denote curvature change.) To be consistent with the discrete definition of the stretching tensor, we compute κ for each triangle.

Essentially, we need a way to compute discrete shape operators for triangles. I adopted the method proposed in (Grinspun et al., 2006). Consider a triangle with its 3 neighboring triangles as shown in Fig. 4.3. θ_i $(i = 1, 2, 3)$ measures the dihedral angle between adjacent triangles, l_i is the length of edge i, t_i is a unit tangent vector orthogonal to edge i, and A is the triangle area. Then the

discrete shape operator is defined as

$$\mathrm{II} = \sum_{i=1,2,3} \frac{\theta_i}{2Al_i} t_i \otimes t_i, \tag{4.4}$$

where the summation is over three triangle edges and $\otimes$ is the outer product. Recall that a shape operator encodes curvature along every direction around a local point; that is, for any tangent vector $\mathbf{v}$, $\mathbf{v}^T \mathrm{II} \mathbf{v}$ gives the curvature along that direction. Therefore, by the definition of Eq. 4.4,

$$\mathbf{v}^T \mathrm{II} \mathbf{v} = \sum_{i=1,2,3} \frac{\theta_i}{2Al_i} \|\mathbf{v}^T t_i\|^2. \tag{4.5}$$

Intuitively, the above equation states that the curvature along any direction is computed as a weighted sum of the three curvatures (the dihedral angle θ_i) across the three edges.

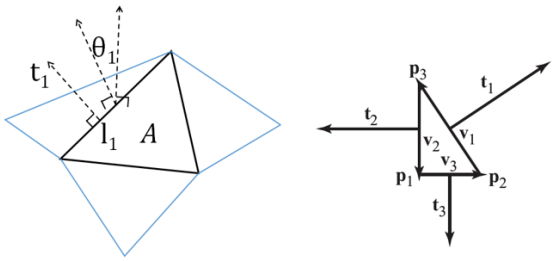


Figure 4.3: A basis stencil used for computing the discrete shape operator. θ_i $(i = 1, 2, 3)$ measures the dihedral angle between adjacent triangles, l_i is the length of edge i, t_i is a unit tangent vector orthogonal to edge i, and A is the triangle area.

The definition of discrete shape operators may seem straightforward, but the difference between two shape operators cannot be easily defined. Consider a surface $\mathcal{M}_1$ and its deformed version $\mathcal{M}_2$. From Section 2.1.2, a point $p \in \mathcal{M}_1$ is mapped to $f(p) \in \mathcal{M}_2$ via the mapping function f. The two shape operators II_p and $\mathrm{II}_{f(p)}$ are respectively defined on two different tangent planes $T_p\mathcal{M}_1$ and $T_{f(p)}\mathcal{M}_2$, so they are not directly subtractable. One has to transform $\mathrm{II}_{f(p)}$ from $T_{f(p)}\mathcal{M}_2$ back to $T_p\mathcal{M}_1$ via the inverse of the tangential linear transformation $(df)_p$. This is called a "pull-back" operation. In the discrete case, we can first compute the shape operators II and II' for the same

triangle before and after deformation respectively. Then the shape operator change is

$$\kappa = \mathrm{II} - A^T \mathrm{II}' A, \tag{4.6}$$

where A is the algebraic representation of the tangential linear transformation matrix. Finally, the local bending energy $W_{membrane}$ associated with a single triangle can be computed based on the bending strain κ (see also definition in Eq. 2.23):

$$W_{bending} = \frac{Eh^3}{24(1-\nu^2)}((1-\nu)\mathrm{Tr}(\kappa^2) + \nu(\mathrm{Tr}\kappa)^2). \tag{4.7}$$

4.1.3 Structural Energy

In our problem the surface is a boundary representation of solid tissues that show shape integrity at certain regions. For example, the epiglottis does not change its thickness. However, the thin shell model only concentrates on local energy computations without explicitly modeling shape/position relationships between distant regions. To handle this, I propose to add *structural links* into the thin shell model. Cross-object structural links L_c are connected between two subregions of the

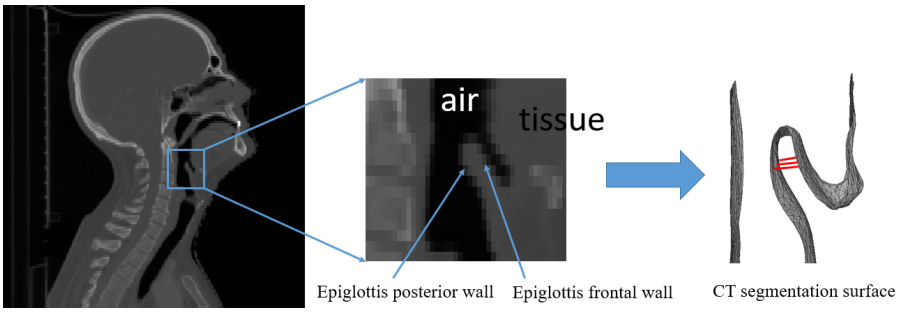


Figure 4.4: Five structural links (red) are manually placed between the frontal and posterior wall of the epiglottis. The structural energy is based on the links' length change to preserve the epiglottal thickness.

surface. As shown in Fig. 4.4, five links are manually placed between the frontal and posterior wall of the epiglottis to preserve the epiglottal thickness. The related structural energy is defined as $W_{L_c} = \sum_i (\Delta|L_c^i|)^2$, where $\Delta|L_c^i|$ is the length change of the i^{th} cross-object link.

Having introduced all three kinds of energies, the deformation energy of a shell is the integration of local stretching and bending energies over all triangles, plus the structural energy:

$$W_{shell} = \int_S (\lambda_{bend} W_{bending} + \lambda_{mem} W_{membrane}) dS + \lambda_l W_{L_c}, \tag{4.8}$$

where $\lambda_{bend}, \lambda_{mem}, \lambda_l$ are non-negative weighting factors among different energy terms.

4.2 Thin Shell Demons Algorithm

In Thirion's Demons algorithm, the notion is that the object boundary in a static image contains a set of so-called demons that can produce virtual forces to attract the object in a moving image until the two images are close. This idea is naturally extendable to surface registration. However, several problems arise when we move from images to surfaces.

1. Image-based demons forces are derived from image gradient information, whereas in surface registration no gradient can be defined in the ambient space.

2. Image similarity can be defined by intensity difference, whereas we use geometry similarity for surface registration.

3. In order to produce a smooth deformation, the image-based Demons algorithm makes use of Gaussian smoothing or differential operators (∇, ∇^2) defined on the deformation function, whereas we use the thin shell model for producing smooth surface deformations.

Here I introduce Thin Shell Demons for surface registration. We assume $\mathcal{M}_1$ is static and $\mathcal{M}_2$ is deformable. The vertices on the triangle mesh of $\mathcal{M}_1$ are considered as demons that can produce geometry-based forces to attract and deform $\mathcal{M}_2$. We then solve for the surface deformation of $\mathcal{M}_2$ according to the augmented thin shell deformation model. In this way, geometrically similar structures between the two surfaces are attracted closer in a physically plausible way. Similar to the image-based Demons method, the algorithm can be formulated as the iterative process shown in Algorithm 1.

Algorithm 1 Thin Shell Demons

1: Precompute geometric features for the static surface $\mathcal{M}_1$.
2: $k = 0$, $\mathcal{M}_2^{(0)} = \mathcal{M}_2$.
3: Compute geometric features for the deformable surface $\mathcal{M}_2^{(k)}$, where $\mathcal{M}_2^{(k)}$ is the deformable surface in the k^{th} iteration.
4: Compute demons forces $\{f\}$ between $\mathcal{M}_1$ and $\mathcal{M}_2^{(k)}$ based on feature matching.
5: Compute the deformation Φ_k from $\mathcal{M}_2^{(k)}$ to $\mathcal{M}_1$ under $\{f\}$. Then $\mathcal{M}_2^{(k+1)} = \mathcal{M}_2^{(k)} \circ \Phi_k$
6: Go back to Step 3 until convergence.

4.2.1 Geometric-Feature-Based Demons Force

In order to effectively match the shapes of the two surfaces, we hope the demons forces can attract similar geometric structures towards each other. Therefore, we use the feature descriptor introduced in Chapter 2 to produce demons forces. Assume that some initial rigid alignment of $\mathcal{M}_1$ and $\mathcal{M}_2^{(k)}$ has already been performed and that the feature similarity between $u \in \mathcal{M}_1$ and $v \in \mathcal{M}_2^{(k)}$ is computed as $\delta(v, u)$ (Eq. 3.4). For each vertex $v \in \mathcal{M}_2^{(k)}$, we find the most geometrically similar vertex $m(v) \in \mathcal{M}_1$ as its matching point, such that

$$\delta(v, m(v)) = \min\{\delta(v, u) | u \in \mathcal{M}_1\}. \tag{4.9}$$

Then the force vector $f(v) = (m(v) - v)$ defines a virtual force applied on v induced by a demon $m(v) \in \mathcal{M}_1$. In other words, $f(v)$ is a vector pointing from v to $m(v)$. Forces computed in this way may contain outliers, so we associate each force vector $f(v)$ with a confidence score $c(v) = e^{-\delta(v,m(v))}$ based on the feature similarity, which indicates how likely the force vector is accurately estimated.

4.2.2 Computing Deformations

Given the set of demons forces $\{f\}$, we need to compute the deformation Φ_k that brings $\mathcal{M}_2^{(k)}$ closer to $\mathcal{M}_1$. Many works (Li et al., 2008; Chang and Zwicker, 2008) have explored a purely geometry-motivated treatment of the problem, but in our scenario within-patient surface deformation

is induced by an underlying physical process. Therefore, we regard $\mathcal{M}_2$ as an elastic thin shell equipped with thin shell energies introduced in Section 4.1. Then the problem becomes to solve the thin shell deformation induced by the external force $\{f\}$.

In order to incorporate the confidence score information of $\{f\}$, I propose an *optimization-based* approach for Thin Shell Demons. The deformation Φ_k can be approximated by minimizing the objective function

$$E(\Phi) = \int_M c(v)\|\Phi(v) - f(v)\|^2 dv + W_{shell}(\Phi). \tag{4.10}$$

The first part penalizes the inconsistency between the deformation vector and the force vector applied on a point and uses the confidence score to weight the penalization. The second part minimizes the augmented thin shell deformation energy. The minimizer of Eq. 4.10 is taken as the most physical plausible deformation that can align similar geometric structures between the two surfaces.

Implementation. The deformation Φ is parameterized by a global deformation vector field defined on the vertices; that is, $\Phi(v)$ is a 3D vector indicating the movement of vertex v. The stretching and bending strains of a triangle can be derived using such deformation vectors in a local neighborhood. If $\mathcal{M}_2$ has N vertices, Eq. 4.10 is a high-dimensional function with $3N$ unknown variables, which makes the estimation of its Hessian matrix very complicated. Therefore, I solve this optimization using the L-BFGS (Limited-memory BFGS) method, which is an optimization algorithm in the family of quasi-Newton methods. L-BFGS approximates the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm using a memory-efficient Hessian approximation. It is often used for solving high-dimensional problems where the Hessian matrix of the objective function is hard to explicitly compute.

4.3 Thin Shell Energy Approximation

4.3.1 Isometric Approximation to the Bending Strain

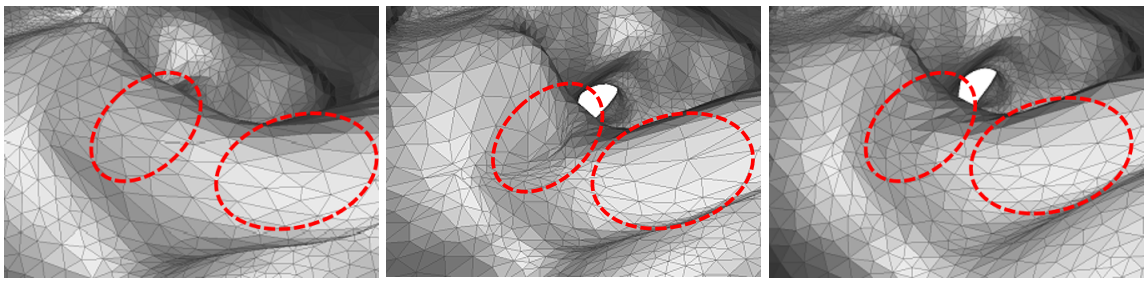
Eq. 4.6 involves first the computation of the tangential linear transformation A and then the computation of shape operators II and II$'$. The composition of the two terms makes the formulation of Eq. 4.6 so complex that even its gradient is not easily derivable. An optional approximation here is to neglect the tangential transformation in the computation of the bending strain. While the stretching component is separately considered by $W_{membrane}$, we assume that the surface is undergoing an isometric deformation (pure bending and no stretching) when we compute the bending strain. In other words, we assume a triangle doesn't change its shape during deformation, which means the linear transformation matrix A in Eq. 4.6 can be approximated to be an identity matrix. Therefore, the bending strain can be simplified as

$$\kappa = \sum_{i=1,2,3} \frac{\Delta\theta_i}{2Al_i} t_i \otimes t_i, \tag{4.11}$$

where A, l_i, t_i are regarded as constants and only need to be computed once for the undeformed configuration. In this way, we only need to compute the gradient for the $\Delta\theta_i$ term.

4.3.2 Quadratic Approximation

The above approximation facilitates gradient computation, but both stretching and bending strains are still highly non-linear, which slows down the convergence of the optimization and also increases the likelihood of converging to local minima. An alternative approach is to adopt quadratic models for computing stretching and bending energies. (Pauly et al., 2005) proposed to use the length change of triangle edges to model stretching. (Bergou et al., 2006) proposed a quadratic model for computing per-edge-based bending energy. With these two approaches, the entire objective function (Eq. 4.10) can be written as a quadratic function, and the optimization becomes the problem of



(a) (b) (c)

Figure 4.5: (a) Original tessellation before registration (b) Deformed tessellation after registration using edge-based stencils (Pauly et al., 2005; Bergou et al., 2006) (c) Deformed tessellation after registration using triangle-based stencils (Section 4.1.1 and 4.1.2)

solving a sparse linear system. In my experiments I found that quadratic models can give up to 250-times speedup (from several minutes to half a second) in the optimization stage. Despite the advantage of speedup, a downside of adopting quadratic models is that edge-based stencils for computing local stretching and bending may overlook the shape integrity of triangles. Fig. 4.5 shows two versions of the deformed surfaces after registration. We can see that triangle-based stencils well preserve the tessellation, whereas edge-based stencils over-distort the triangle shapes.

4.4 Results

For the TSD algorithm used in this subsection, I adopted the bending strain approximation discussed in Section 4.3.1 for convenient gradient computation. However, since the registration between CT and endoscopic surfaces happens during the radiation treatment planning stage and is not time-critical, I didn't use the quadratic approximation discussed in Section 4.3.2 to speed the registration up. In turn using the full energy term will yield more accurate registration results in the following two synthetic tests.

4.4.1 Synthetic Deformation

Data. I first tested TSD with the same synthetic data as in Chapter 3. To remind the readers, we had 24 surface pairs: 12 for complete surface registration and 12 for partial surface registration. Each

surface pair includes a CT segmentation surface and its synthetically deformed version. Again, I used the *surface registration error* as the metric.

TSD Parameters. I used the same validation set as in Chapter 3 to determine optimal algorithm parameter values. The Young's modulus E and Poisson's ratio ν were set to 2 and 0.25 respectively. The weighting parameters $\lambda_{bend}, \lambda_{mem}$ were chosen to be 1 and 10. As will be explained later, I left out the structural energy term in the synthetic test. All the parameters related to geometric feature construction and matching were chosen to have the same values as in Chapter 3.

Comparison. I compared Thin Shell Demons with three other strategies suggested in other works to indicate the importance of the choices made in TSD. The first method is my improved spectral graph matching method discussed in Chapter 3. The second method is a deformation-based method with thin shell energy as deformation regularization, but it uses the closest point search suggested in (Li et al., 2008) to drive the deformation instead of geometric feature matching. The third method uses geometric feature matching, but it only uses the edge-based stretching energy as the deformation regularization (Pauly et al., 2005). Since the above three methods are fully automatic, I left out the manually constructed structural links in the synthetic test to make the comparison fair. Also, the weighting parameters within the energy of the last two methods were re-tuned using the same validation dataset to produce the best results.

Fig. 4.6 shows the registration error of the 12 complete surface pairs using the 4 methods. We can see that in several cases the edge-based stretching energy can produce slightly better results than Thin Shell Demons. This is because without any bending energy regularization it allows more flexible local deformation to adapt to the target surface. However, as mentioned in Fig. 4.5, the trade-off is that it may end up with more irregular triangle shapes. Our thin shell energy does a better job at preserving the triangulation pattern during deformation, which is important for texture mapping, one of the objectives in the later fusion application. A typical synthetic deformation and the registration result are shown in Fig. 4.8.

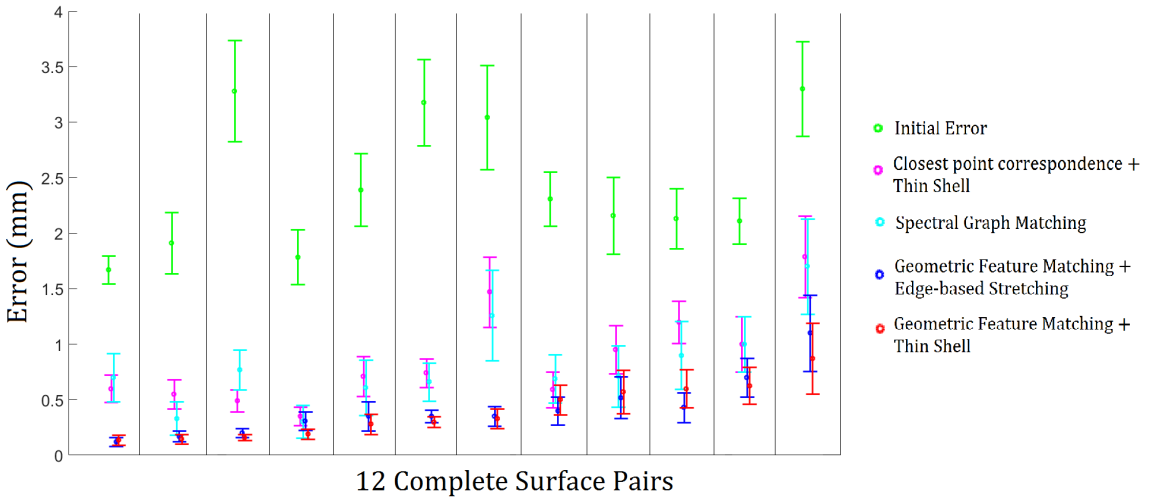


Figure 4.6: Registration error for 12 complete surface pairs. The center-point of each vertical bar is the surface registration error, which is defined as the average of *vertex registration error*; the length of the error bar denotes the standard deviation of the *vertex registration error*.

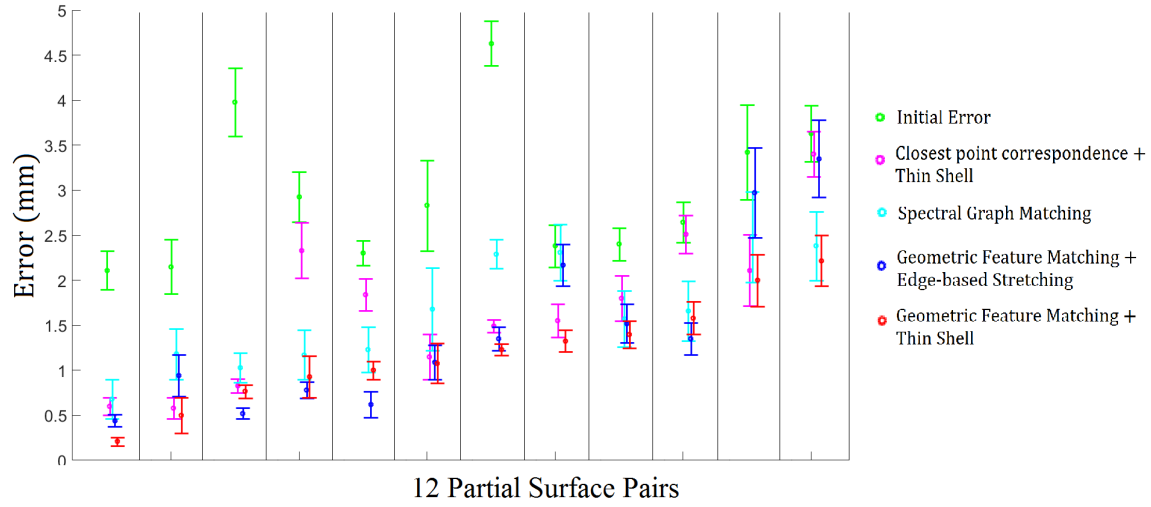


Figure 4.7: Registration error for 12 partial surface pairs.

4.4.2 The Phantom

I also tested TSD on data derived from a 3D-printed phantom. The procedure for producing the phantom is shown in Fig. 4.9. First, a binary segmentation was performed to separate the airway

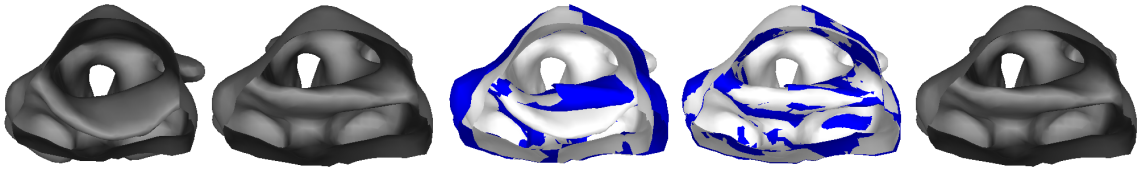


Figure 4.8: From left to right: deformable surface before registration; static surface; initial overlay between the two surfaces before registration.; final overlay between the two surfaces after registration; deformable surface after registration.

from soft tissues of the throat in a patient's 3D CT image. Then we 3D-printed a cuboid volume of that binary segmentation with the solid part corresponding to the soft tissues. We used the 3D-printed model as a real-size static phantom of the throat. We took endoscopy and CT of the phantom model and produced its reconstruction and CT segmentation surfaces. Finally, we performed TSD to register those two surfaces.

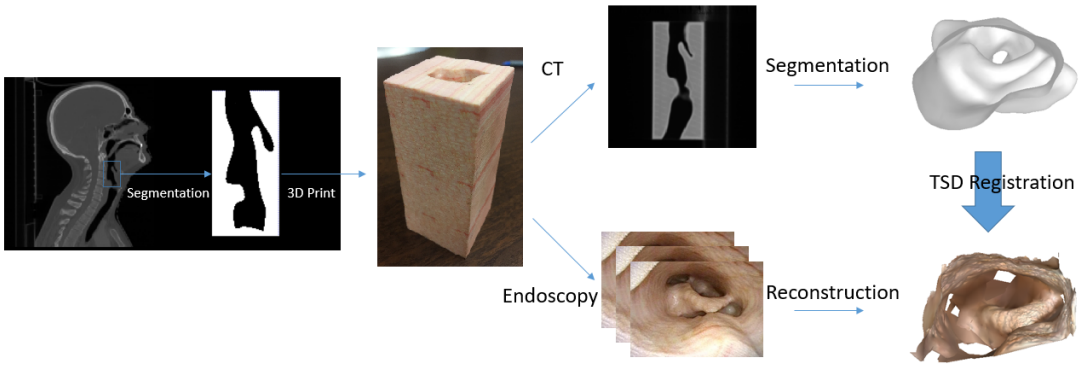


Figure 4.9: Experimental setup for the phantom case: a real-size static 3D-printed model was produced from a CT segmentation of a real patient. We took endoscopy and CT of the phantom model and produced its reconstruction and CT segmentation. TSD was used to find the deformation caused by reconstruction artifacts.

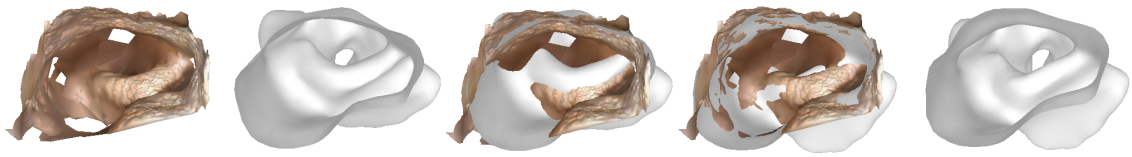
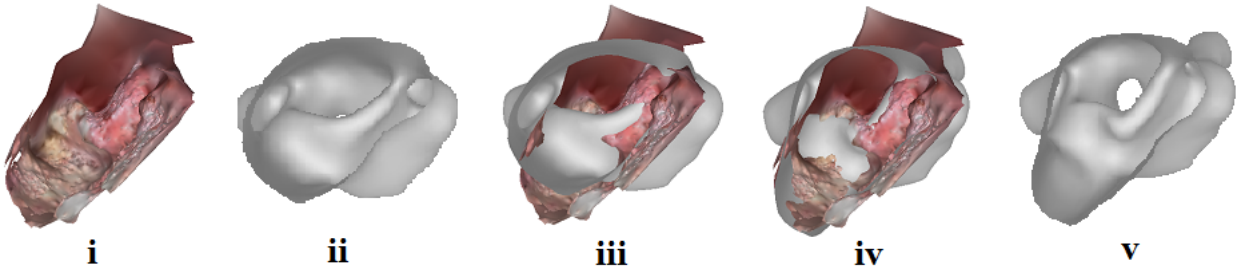


Figure 4.10: TSD on the phantom case. From left to right: Endoscopic reconstruction; CT surface before registration; Initial overlay; Final overlay; Registered CT.

Since the phantom is static, the initial deviation between the two surfaces is mainly caused by reconstruction artifacts. We can see from Fig. 4.10 that TSD can also handle such non-physical deformations.

4.4.3 Real Patient Data



(a) The reconstruction surface in this example is derived from ColMap (Schönberger and Frahm, 2016)

(b) The reconstruction surface in this example is derived from a single frame using Shape-from-Shading

Figure 4.11: TSD on the two real patient cases. From left to right: Endoscopic reconstruction; CT surface before registration; Initial overlay; Final overlay; Registered CT.

In real patient cases, we don't have ground truth correspondences between CT and endoscopic reconstruction surfaces. Here I show only some qualitative results for real patient data. In this chapter, I first show TSD registration results with some early reconstruction surfaces derived from some other existing techniques (Schönberger and Frahm, 2016; Ahmed and Farag, 2006). After introducing the groupwise TSD and endoscopogram reconstruction pipeline in Chapter 6, I will show more results with the latest reconstruction surfaces produced by our own method. Fig. 4.11 shows two cases where the CT surfaces were registered to the endoscopic reconstruction surfaces produced from patients' endoscopy. The endoscopic surface in Fig. 4.11a was produced by CoLMap (Schönberger and Frahm, 2016), and the endoscopic surface in Fig. 4.11b was produced by Shape-

from-Shading (Ahmed and Farag, 2006). In both cases, the reconstruction surfaces were first rigidly aligned to the CT surfaces using an initial alignment method that will be discussed in Section 7.1. The cross-object structural links were added in this test. We can see that after registration (Fig. 4.11iv) the two surfaces are well overlaid with each other compared to the initial overlay results (Fig. 4.11iii). The final deformed CT surfaces (Fig. 4.11v) geometrically resemble the endoscopic reconstructions.

Structural Energy. I found that without the cross-subject structural links the quality of the deformation in the epiglottis was highly compromised in real data registration. This is because the geometry near the epiglottal area in the reconstruction is less accurate, thus generating more outliers in the force field $\{f\}$ that will easily distort the epiglottis. For the same registration in Fig. 4.11b, the influence of the structural links on the deformed shape of the epiglottis is shown in Fig. 4.12. Without using the links (Fig. 4.12c), the epiglottis was unreasonably distorted with respect to its original configuration (Fig. 4.12a).

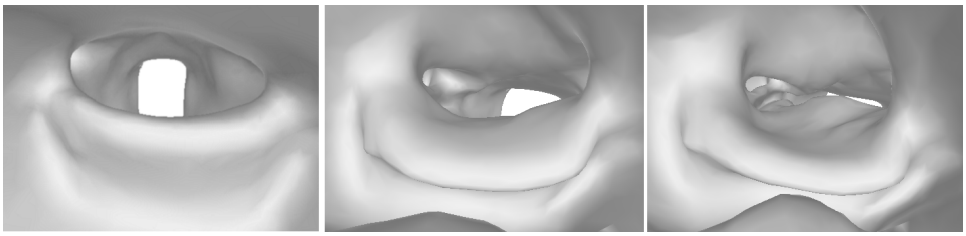


Figure 4.12: (a) Original epiglottis (b) Deformed epiglottis with structural links. (c) Deformed epiglottis without structural links.

4.5 Discussion and Conclusion

I have proposed a physics-motivated surface registration framework: Thin Shell Demons. I adopted the "attraction forces" idea from Thirion's Demons image registration algorithm to our surface registration problem. The virtual attraction forces are computed by geometric feature matching. I also adopted a thin shell physical model to produce physically plausible surface deformations. Novel cross-object links were proposed to preserve non-local shape structures. Based on an optimization

formulation TSD can attract similar geometric structures between two surfaces in a physically realistic fashion. Experiments have shown that the combination of geometric feature matching and the thin shell model provides superior registration in the context of CT/endoscopy fusion.

Despite the promising results, there are several potential issues of TSD that I need to address in the future:

Physical Reality. The thin shell model is a simplified physical model only considering tangential strains (Ventsel and Krauthammer, 2001), so that its energy can be effectively optimized, but how much the pharyngeal surface would follow this model remains a question. The real surface deformation is mainly caused by the underlying muscles' contraction. Therefore, it won't be fully characterized by the mechanics of a shell. As mentioned in the beginning of this Chapter, an alternative way to model the mechanics of the throat is to consider not only the surface but also the surrounding muscles by 3D finite element method (Xu et al., 2009). However, that requires complex anatomical knowledge of the throat, which is not available in our application. Considering different trade-offs, I chose the thin shell model to produce physically reasonable deformations. It can handle larger scale problems by reducing the problem from 3D to 2D, but I also equipped the traditional model with structural links, which could be thought of as additional 3D information that 3D FEM can provide.

Structural Links. Currently, the cross-object structural links are only placed into CT surfaces. It is conceivable to also construct such links in the endoscopic reconstructions. Nevertheless, they have to be manually constructed in the current stage, and it remains an open question whether the number and the position of these links can be automatically determined.

Statistical Parameter Learning. In our model we assume uniform elasticity parameters across the entire surface, which is obviously not realistic. The stiffness of the throat is not only spatially dependent but also anisotropic. For example, the pharyngeal wall mostly stretches and contracts in the lateral direction; the epiglottis is extremely bendable along the pharynx. To improve physical reality, one solution is to statistically learn the stiffness parameters at different places from multiple known pharyngeal deformations. This idea will be explored in Chapter 6.

CHAPTER 5

Groupwise TSD and Its Use in Endoscopogram Construction

TSD is essentially a pairwise surface registration framework, where a deformable surface can be attracted to a static surface. This framework perfectly suits the purpose of CT-segmentation/endoscopy-reconstruction surface registration. However, in the endoscopy reconstruction process itself, there is a *geometry fusion* step that requires registering multiple partial reconstruction surfaces into a single and complete surface. The pairwise TSD cannot sufficiently handle such a groupwise registration problem.

In this chapter, I will propose an extension of Thin Shell Demons, namely the groupwise TSD, that can be used for geometry fusion in the endoscopy 3D reconstruction pipeline. I will first briefly describe the entire pipeline and motivate the need for geometry fusion. I will then introduce the key components of the pipeline, with the concentration on geometry fusion. Finally, I will show validation results on the groupwise TSD and show some of the final endoscopy reconstruction results.

3D Reconstruction from Endoscopic Images

Most existing reconstruction techniques heavily rely on two fundamental algorithms: Shape-from-Shading (SfS) and Structure-from-Motion (SfM). For example, a pilot study (Meisner et al., 2013) has been applied in laryngeal endoscopic reconstruction using semi-automatic SfM. Even though the proposed system in that study shows some promising results, it is well known that SfM can only recover a sparse point cloud without any detail geometry. Therefore, SfS is further

incorporated in some other works (Tokgozoglu et al., 2012; Hong et al., 2014; Wu et al., 2010) to achieve detail geometry reconstruction using local shading information. The use of the above two algorithms has achieved remarkable success in applications like colonoscopy inspection (Hong et al., 2014), laparoscopic surgery (Maier-Hein et al., 2013) and orthopedic surgeries (Wu et al., 2010). However, it is still the case that most existing techniques can only handle (a) single-image reconstruction (Hong et al., 2014), (b) non-deformable scenes (Wu et al., 2010) and (c) Lambertian surfaces (Tokgozoglu et al., 2012).

The novel endoscopic reconstruction pipeline proposed by me and my colleagues aims to reconstruct a longer video sequence of a deformable scene with complex light reflectance. This is achieved by a single-frame-reconstruction stage followed by a cross-frame fusion stage. Fig. 5.1 shows the pipeline for reconstructing a surface model, which we call an endoscopogram, from an endoscopic video sequence. The reconstruction has three major steps:

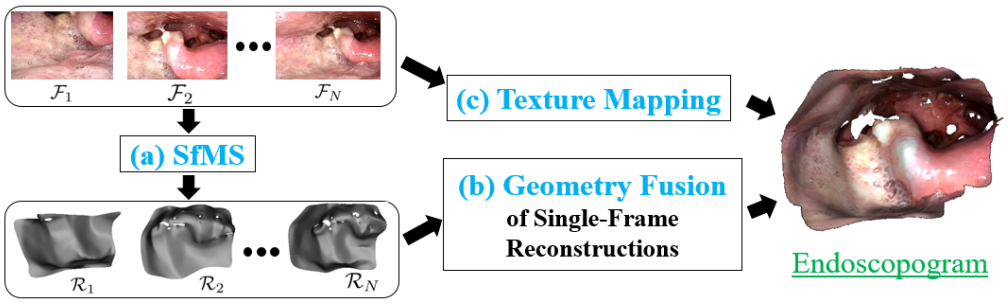


Figure 5.1: The construction of an endoscopogram involves three major steps: (a) Structure-from-Motion-and-Shading (SfMS); (b) geometry fusion; (c) texture fusion.

(a). A single-frame reconstruction surface is produced for each individual frame image using Shape-from-Motion-and-Shading (SfMS).

(b). A geometry fusion step is carried out to geometrically fuse all the single-frame reconstructions into a single surface.

(c). The surface derived in Step (b) is textured using all the raw frames to produce the final endoscopogram.

In the rest of this chapter, I will briefly introduce the techniques used in Step (a) and Step (b). I will focus on discussing Step (b), which is one of the major methodological contributions of this dissertation.

5.1 Shape-from-Motion-and-Shading

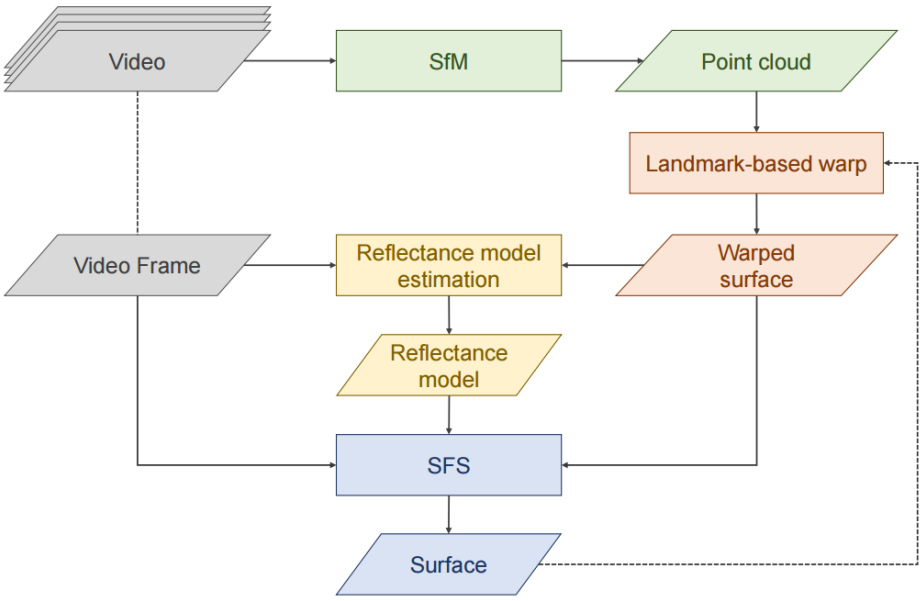


Figure 5.2: Structure-from-Motion-and-Shading diagram (figure from (Price et al., 2016)).

In this subsection I will briefly introduce the first step in our pipeline, *Shape-from-Motion-and-Shading* (SfMS). Details of this method are given in (Price et al., 2016). The diagram of SfMS is shown in Fig. 5.2. In SfMS, two key algorithms, Structure-from-Motion (SfM) and Shape-from-Shading (SfS), are combined in an effective way to produce 3D reconstructions. Given a series of video frames, SfM (Schönberger and Frahm, 2016) is used to simultaneously estimate camera motions across frames and sparse geometry (a 3D point cloud) of the 3D scene. SfM provides reasonably accurate sparse point positions, but it does not yield dense geometry. The other algorithm, Shape-from-Shading (SfS) (Ahmed and Farag, 2006), is a monocular method of depth estimation that, given a single image viewing a scene, recreates depth for every single pixel in that image. The formulation of SfS is closely related to local surface geometry but lacks global

geometry constraints. Moreover, SfS estimation relies on an accurate light reflectance model, which is usually very difficult to acquire in unknown environments. Therefore, the idea of SfMS is to use SfM-derived sparse points as global constraints and to iteratively estimate the reflectance model parameters and the SfM-regularized SfS depth map for individual frames.

The input to SfMS is a video sequence of hundreds of consecutive frames $\{\mathcal{F}_i \mid i = 1...N\}$. SfMS generates for each frame $\mathcal{F}_i$ its own individual reconstruction surface $\mathcal{R}_i$. (Strictly speaking, $\mathcal{R}_i$ is a depth image with a single depth value for each pixel. For post-processing convenience, I transform $\mathcal{R}_i$ to a triangle mesh.) Each surface $\mathcal{R}_i$ is transformed into the world space using the camera position parameters estimated from SfM. Since each image $\mathcal{F}_i$ only visualizes a part of the throat, each $\mathcal{R}_i$ is only a partial reconstruction. Therefore, there is an apparent need to fuse multiple single-frame reconstructions $\{\mathcal{R}_i\}$ into a single geometry $\mathcal{R}$. This fusion step involves both texture and geometry. We will first discuss geometry fusion in the next section.

5.2 Geometry Fusion

Geometry fusion is a challenging task in our application. Due to the constantly changing camera viewpoint, all individual reconstruction surfaces $\{\mathcal{R}_i\}$ are only partially overlapping with each other. Moreover, since the tissue may have deformed between 2D frame acquisitions, there will be non-rigid deformations among those individual reconstructions. Finally, mesh facets that are nearly tangent to the camera viewing ray are regarded as occlusion regions and are removed, which leaves many holes in the surfaces. To summarize, the end result of this is that the surfaces $\{\mathcal{R}_i\}$ have missing patches and non-rigid deformations and are only partially overlapping with each other.

The goal of geometry fusion is to non-rigidly deform all these surfaces into a consistent geometric configuration, thus unifying the tissue deformation and minimizing reconstruction inconsistency among different frames.

The most naive solution would be to apply the pairwise TSD discussed in the previous chapter: choose a reconstruction surface of one particular frame as the reference and incrementally register

other surfaces from nearby frames to that reference. However, it has been shown in many cases (Lyu et al., 2015; Joshi et al., 2004; Geng et al., 2009) that this incremental construction suffers from the so called "reference selection bias". Any sort of fusion error made in the early stage can be easily propagated to later frames.

The above argument has motivated the use of groupwise surface registration, where all surfaces are treated equally and registered simultaneously. Existing groupwise surface registration methods often rely on having or iteratively estimating a mean geometry (template) (Durrleman et al., 2009, 2012; Lyttelton et al., 2007; Ramachandran et al., 2014; van de Giessen et al., 2012). However, in our situation, the topology change and partially overlapping data renders initial template geometry estimation almost impossible. Therefore, a groupwise registration method that does not require a template or a mean surface is desired. To that end, template-free methods have been studied for certain applications (Balci et al., 2007; Lyu et al., 2015), such as 2D images and spherical surfaces, but it has not been shown that such methods can be generalized to arbitrary surfaces. The joint spectral graph framework (Arslan et al., 2015; Lombaert et al., 2013) provides an alternative way for matching a group of surfaces without estimating the mean, but as discussed in Chapter 3, these methods are naturally not suitable for handling surfaces with different intrinsic geometry.

5.2.1 The N-body Scenario

My insight is that the notion of virtual attraction forces is still valid for every surface pair among the N surfaces; a surface can attract and be attracted by all other surfaces simultaneously. Thus, we propose a groupwise deformation scenario as an analog to the N-body problem: N surfaces are deformed under the influence of their mutual forces. The deformation of a single surface can be computed based on the overall attraction forces provided from all the other surfaces. With the physical thin shell model, each single surface's deformation can be topology-preserving and not influenced by its partialness. This groupwise attraction strategy can bypass the need of the mean geometry estimation and can still deform all surfaces into a single geometric configuration. With

this notion in mind, we now have to define (1) mutual forces among N surfaces; (2) an evolution strategy to deform the N surfaces.

Mutual Forces. In order to derive mutual forces, correspondences should be credibly computed among N partial surfaces. Again, the attraction forces between a surface pair are constructed as the vectors connecting estimated corresponding points between the two surfaces. As discussed in Chapter 3, geometric feature matching can identify a set of geometrically similar points between two surfaces. Additionally, in the context of registering endoscopic frame reconstruction surfaces, each surface $\mathcal{R}_i$ has an underlying texture image $\mathcal{F}_i$. Thus, we can also compute texture correspondences between two frames using standard computer vision techniques (Hartley and Zisserman, 2004). In particular, given two frames, SIFT (Lowe, 1999) matching points are first detected using an exhaustive search. The resulting matching points may contain outliers (false correspondences). Therefore, assuming the scene undergoes a homographic transformation between the two frames, we perform a RANSAC (Fischler and Bolles, 1981) algorithm to remove outlier matchings. Since RANSAC is less robust with too few data items, we perform the above matching procedure only between temporally close frame pairs, i.e. $\{\mathcal{F}_i, \mathcal{F}_j \mid 0 < |i - j| \leq T\}$, to yield a larger number of matching results. I heuristically set $T = 0.5s$ to guarantee that two frames within that time range can be robustly matched. Finally, the remaining inlier SIFT matchings on the two frames can be directly transformed to 3D vertex correspondences on the two surfaces via the SfSM reconstruction procedure.

In the end, any given vertex $v \in \mathcal{R}_i$ will have M corresponding vertices in the other surfaces $\{\mathcal{R}_j \mid j \neq i\}$. As mentioned above, these M correspondences are detected by either SIFT matching or geometric feature matching. ($M = 0$ means we have not been able to detect any of its correspondences.) Accordingly, we construct M attraction force vectors $\{f^\beta(v) = u^\beta - v \mid \beta = 1...M\}$, where u^β is the β^{th} correspondence of v in some other surface. To remind the readers, the $\{f\}$ vectors we call forces are not in any sense real physical forces; they are simply vectors connecting corresponding vertices between surfaces. Moreover, we associate such force vectors with confidence

scores $\{c^\beta(v)\}$. They are defined by

$$c^\beta(v) = \begin{cases} \delta(u^\beta, v) & \text{if } \{u^\beta, v\} \text{ is a geometric correspondence,} \\ 1 & \text{if } \{u^\beta, v\} \text{ is a SIFT (texture) correspondence,} \end{cases} \tag{5.1}$$

where δ is the geometric feature distance defined in Chapter 3, Eq. 3.4. Since we only consider inlier SIFT matchings using RANSAC (Hartley and Zisserman, 2004), the confidence score for texture correspondences is a constant 1. We then define the overall force exerted on v as a kernel-weighted average based on confidence scores:

$$\bar{f}(v) = \frac{\sum_{\beta=1}^{M} c^\beta(v) f^\beta(v)}{\sum_{\beta=1}^{M} c^\beta(v)}. \tag{5.2}$$

The associated overall confidence score is

$$\bar{c}(v) = \frac{1}{M} \sum_{\beta=1}^{M} c^\beta(v). \tag{5.3}$$

Deformation Strategy. With the attraction forces defined for all vertices from all surfaces, we want to evolve the N surfaces accordingly to such force interaction. We denote $\{\bar{f}_i\}$ as the set of attraction forces exerted on $\mathcal{R}_i$. Since $\bar{f}_i$ already summarize forces induced by all other surfaces, the deformation of $\mathcal{R}_i$ can be solved independently. Again, I use an optimization-based approach to solve for the deformation field Φ_i of $\mathcal{R}_i$:

$$E(\Phi_i) = \int_M \bar{c}(v)(\Phi_i(v) - \bar{f}_i(v))^2 dv + E_{shell}(\Phi_i). \tag{5.4}$$

Compared to Eq. 4.10, we can tell that the only difference is that the pairwise forces are replaced by the overall attraction forces.

Therefore, a groupwise deformation scenario is to evolve the N surfaces by iteratively estimating their mutual attraction forces and solving for their deformations. However, a potential hazard of this

algorithm is that without a common target template the N surfaces could oscillate, especially in the early stage when the force magnitudes are large and tend to overshoot the deformation. To this end, I observe that the thin shell energy regularization weights $\lambda_{bend}, \lambda_{mem}$ control the deformation flexibility (see Eq. 4.8 for a precise definition of $\lambda_{bend}, \lambda_{mem}$). Thus, to avoid oscillation, we start with large regularization weights and gradually tune them down through the iterations. To summarize, I design the following groupwise registration algorithm:

Algorithm 2 N-body Groupwise Surface Registration

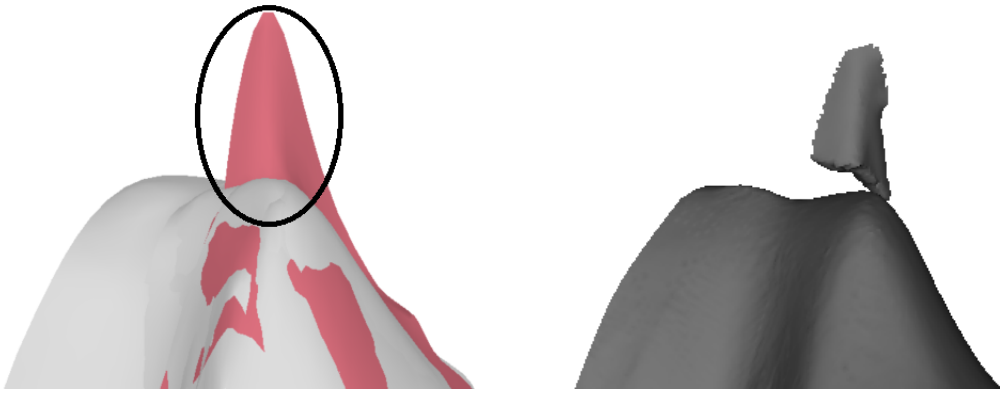
1: Start with large regularization weights: $\lambda_{bend}^{(0)}, \lambda_{mem}^{(0)}$
2: In iteration k, compute $\{f_i\}$ for each surface $\mathcal{R}_i^{(k)}$ using the current N surface configurations $\{\mathcal{R}_i^{(k)} \mid i = 1...N\}$
3: Optimize Eq. 5.4 independently for each surface to obtain $\mathcal{R}_i^{(k+1)}$
4: $\lambda_{bend}^{(k+1)} = \sigma * \lambda_{bend}^{(k)}$, $\lambda_{mem}^{(k+1)} = \sigma * \lambda_{mem}^{(k)}$, with $\sigma < 1$
5: Go to step 2 until convergence.

In fact, there is another motivation for gradually downweighting the thin shell energy weights. As the iterations proceed, the surfaces will get closer, and the attraction force computation will become more accurate with fewer outliers, so less regularization will be needed. However, a too small σ will quickly shrink the regularization weights to zero. Therefore, I choose $\sigma = 0.95$ and only run 10 of the above iterations.

In the end, all partial surfaces $\{\mathcal{R}_i\}$ are deformed into a mutually consistent shape configuration. Letting $\{\mathcal{R}_i'\}$ denote this set of registered surfaces, the final step is to create a single surface geometry $\mathcal{R}$ from the registered surfaces $\{\mathcal{R}_i'\}$. This can be easily done via a simple fusion method proposed in (Curless and Levoy, 1996). This method first creates signed distance functions from all the surfaces. Then an average signed distance function is computed. Finally, the zero-level-set surface is extracted from the average signed distance function as the fused geometry. Therefore, as long as all the surfaces are well aligned, the average signed distance function can be robustly computed to generate the final fusion.

5.2.2 Outlier Geometry Removal

The final step of geometry fusion (Curless and Levoy, 1996) assumes the registered surfaces $\{\mathcal{R}'_i\}$ are geometrically consistent with each other. However, this fusion step can be seriously harmed by the outlier geometry created by SfMS. Outlier geometries are local surface parts that are wrongfully estimated by SfMS under bad lighting conditions (insufficient lighting, saturation, or specularity). These sub-surfaces are drastically different from all other surfaces and thereby are carried over by the deformation process to $\{\mathcal{R}'_i\}$ (Fig. 5.3a). This artifact severely violates the aforementioned geometric consistency assumption in the final fusion step; that is, outlier geometry yields a signed distance function that is inconsistent with all other functions, so the average signed distance function is heavily polluted. Fig. 5.3 shows a typical case, where five registered surfaces are overlaid together. The pink surface has a piece of outlier geometry (circled in black) that doesn't coincide with the other four surfaces. A direct fusion by (Curless and Levoy, 1996) will create an undesirable result (Fig. 5.3b).

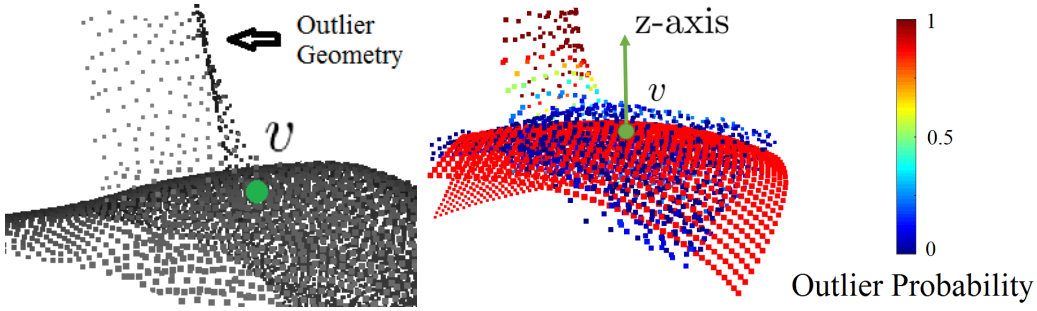


(a) (b)

Figure 5.3: (a) 5 registered surfaces are overlaid together with the pink surface having a piece of outlier geometry (circled in black). (b) A direct geometry fusion with the presence of outlier geometry creates an unreasonable result.

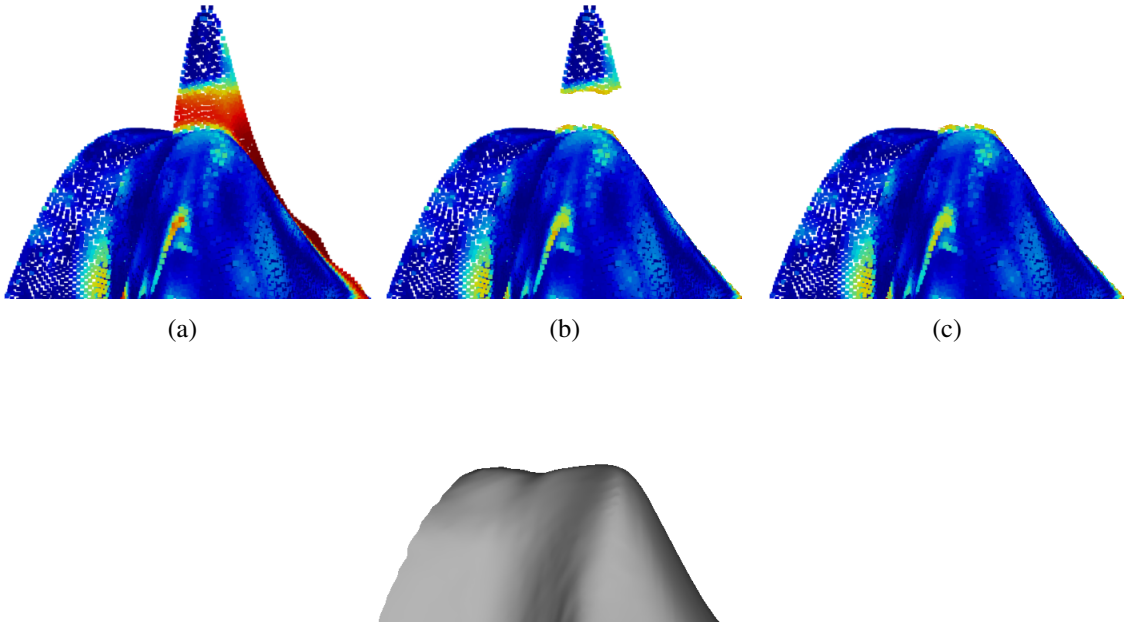
My observation is that outlier geometry changes a local surface's topology by creating wrong branches, which violates many basic differential geometry properties. We know that the local surface around a point in a smooth 2-manifold can be approximately presented by a quadratic

Monge Patch $h : U \rightarrow \mathbb{R}^3$, where U defines a 2D open set in the tangent plane, and h is a quadratic function (Gray, 1997). In simple words, a valid local surface patch without any outlier geometry resembles the shape of a quadratic function. My idea is that if we fit a local quadratic surface at a branching place using some robust regression methods (Andersen, 2008), the surface points on the wrong branch of outlier geometry will be counted as outliers.



(a) (b)

Figure 5.4: (a) Local point cloud $\mathcal{N}(v)$ around vertex v. (b) Robust quadratic fitting (red grid) to normalized $\mathcal{N}(v)$. The outlier scores of $\mathcal{N}(v)$ are indicated by the color-coding.



(a) (b) (c)

(d)

Figure 5.5: (a) Color-coded outlier scores $\mathcal{W}$ of all vertices in $\mathcal{L}$. (b) The remaining point cloud after thresholding the ourlier scores $\mathcal{W}$. (c) The largest remaining component. (d) Fused surface created from the largest component of the remaining point cloud after outlier geometry removal.

We define the 3D point cloud $\mathcal{L}$ as the ensemble of all vertices in all registered surfaces $\{\mathcal{R}'_i\}$. The goal is to classify for each vertex $v \in \mathcal{L}$ whether it is an outlier vertex (a vertex in outlier geometries) based on its outlier score $\mathcal{W}(v)$. The idea is to fit a quadratic surface to each local region of $\mathcal{L}$ to produce outlier scores for vertices in that local region. To be specific, for a given vertex v, $\mathcal{N}(v)$ is defined as the set of vertices in its neighborhood, i.e., $\mathcal{N}(v) = \{u \mid u \in \mathcal{L}, ||u - v|| < \tau\}$. We first transform the local point cloud $\mathcal{N}(v)$ by taking v as the center of origin and the normal direction of v as the z-axis. Then, we use Iteratively Reweighted Least Squares (Green, 1984) to fit a quadratic polynomial to the normalized $\mathcal{N}(v)$ (Fig. 5.4b). The method produces an outlier score for every vertex $u \in \mathcal{N}(v)$, which is then accumulated into $\mathcal{W}(u)$ (Fig. 5.5a). Next, we remove the outlier vertices by thresholding the accumulative outlier scores $\mathcal{W}$ (Fig. 5.5b). This thresholding procedure will leave the remaining point could as several detached components (5.5b), among which the component with most vertices (5.5c) is regarded as the true point cloud and is used to produce the final single geometry $\mathcal{R}$ (Curless and Levoy, 1996) (Fig. 5.5d).

5.3 Texture Fusion

After geometry fusion a single surface $\mathcal{R}$ is created from the set of registered surfaces $\{\mathcal{R}'_i\}$. The remaining task is to determine the texture of $\mathcal{R}$. Remember that each single-frame reconstruction $\mathcal{R}_i$, as well as its registered version $\mathcal{R}'_i$, is associated with a 2D texture image $\mathcal{F}_i$. The goal of texture mapping then becomes to assign a color to each vertex in $\mathcal{R}$ based on the set of texture images $\{\mathcal{F}_i\}$. Fig. 5.6 gives a high-level concept of texture mapping. For each vertex $v \in \mathcal{R}$, we find its corresponding vertex in one of the registered single-frame reconstructions $\mathcal{R}'_i$ and trace back its color in the corresponding frame $\mathcal{F}_i$.

The challenge here is that a region (vertex) in $\mathcal{R}$ usually has corresponding vertices in multiple single-frame reconstructions because that region is observed in multiple video frames. This poses the question of determining which particular frame to draw its color from. It turns out that this procedure can be formulated as a labeling problem. The label of a vertex $v \in \mathcal{R}$, denoted as

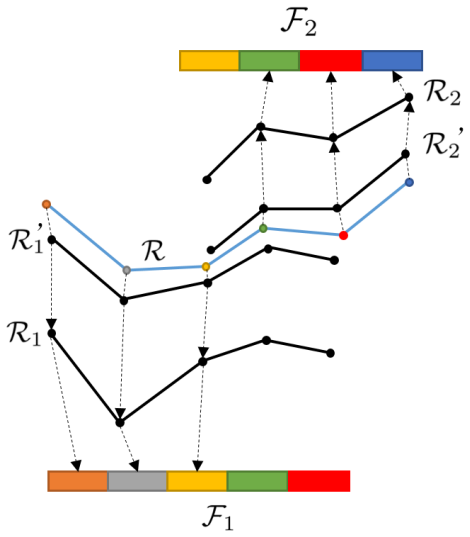


Figure 5.6: Each vertex of the fused surface $\mathcal{R}$ finds its corresponding vertex in a registered single-frame reconstruction $\mathcal{R}_i'$ and traces back the color in the corresponding texture image $\mathcal{F}_i$. Some vertices (green, yellow, red) in $\mathcal{R}$ have ambiguity in choosing the corresponding frame because they have corresponding vertices in both $\mathcal{R}_1'$ and $\mathcal{R}_2'$.

$l(v) \in \{1...N\}$, indicates that we choose v's corresponding vertex in $\mathcal{R}_{l(v)}'$ as the color source. Our goal now is to find an optimal label field for all vertices in $\mathcal{R}$. One important factor in doing this is that we want to encourage regional label consistency. This is because too frequent label switching between neighboring locations can lead to noisy artifacts due to illuminace change across frames.

Here I design a Markov Random Field (MRF) to optimize the label field. We denote v^k as the k^{th} vertex in $\mathcal{R}$ and $d(\cdot,\cdot)$ as the Euclidean distance between two points. We define $\mathcal{C}_i(v^k) \in \mathcal{R}_i'$ as the closest point of v^k in $\mathcal{R}_i'$, i.e.,

$$d(v^k, \mathcal{C}_i(v^k)) = \min\{d(v^k, u) \mid u \in \mathcal{R}_i'\}. \tag{5.5}$$

The color of v^k is indicated by a label $l(v^k) \in \{1...N\}$. Specifically, $l(v^k) = i$ means that the color of v^k is pulled from $\mathcal{F}_i$ by tracing the corresponding pixel of $\mathcal{C}_i(v^k)$ in $\mathcal{F}_i$. The label field $\{l\}$ can be computed by minimizing the following MRF objective function:

$$E(l) = \sum_k S(l^k) + \lambda_{smooth} \sum_k \sum_{k' \in \mathcal{N}(v^k)} (1 - \boldsymbol{\delta}(l^k, l^{k'})). \tag{5.6}$$

In the above equation, the scoring function $S(\cdot)$ is defined for each individual vertex's label. It prefers pulling color from the closest non-boundary points in $\{\mathcal{R}_i'\}$.

$$S(l^k) = \begin{cases} d(v^k, \mathcal{C}_i(v^k)) & \text{if } \mathcal{C}_i(v^k) \notin \partial\mathcal{R}_i, \\ +\infty & \text{if } \mathcal{C}_i(v^k) \in \partial\mathcal{R}_i. \end{cases} \tag{5.7}$$

The second term of Eq. 5.6 encourages regional label consistency. $\mathcal{N}(v^k)$ is defined as the neighborhood around v^k (see also Section 5.4 for its precise definition). $\boldsymbol{\delta}(\cdot,\cdot)$ is the Kronecker delta function:

$$\boldsymbol{\delta}(i,j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{5.8}$$

Finally, λ_{smooth} plays the role of weighting the two terms. The MRF function is minimized by the Iterated Conditional Modes method (Besag, 1986).

5.4 Results

Since my major scientific contribution in this chapter is the geometry fusion algorithm discussed in Section 5.2, I will mainly validate that part in the following. In addition to the thin shell energy weights $\lambda_{bend}, \lambda_{mem}$, another important model parameter is the annealing parameter σ in Algorithm 2. I selected these parameters by tuning on a test patient's data (separate from the datasets presented here). They were chosen as $\lambda_{bend} = \lambda_{mem} = 1$ and $\sigma = 0.95$.

Synthetic Data

Similar to the previous two chapters, I first validated our groupwise registration algorithm using synthetic data. The same dataset of the 6 patients' CT surfaces was used in this test. To simulate the group of partially overlapping single-frame reconstructions, I generated for each patient's CT surface 20 partial surfaces by taking depth maps from different camera positions in the CT space. Then I applied synthetic deformations to each of the 20 surfaces. The registration error was

measured as the average *surface registration error* (see Section 3.4 for its precise definition) over all pairs of surfaces. Since the CT surface is textureless, I only incorporated geometric correspondences (Eq. 5.1) in computing the mutual attraction forces.

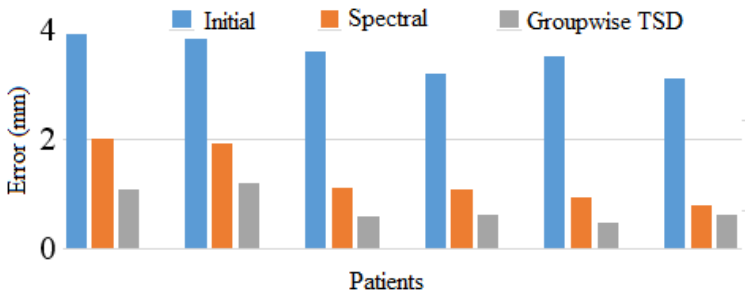


Figure 5.7: Error plots of synthetic data for 6 patients.

Due to the unique template-free nature of groupwise TSD and the aforementioned difficulty in estimating the mean surface, I only compared groupwise TSD with another method that can match many partial surfaces without estimating the mean. The method being compared with is based on the proposed spectral graph matching method in Chapter 3, with an extension from pairwise surface matching to the groupwise situation. To be specific, geometric-feature-based inter-surface links were added between every pair of surfaces, and the joint spectral graph matching was carried out on an affinity graph made of N surfaces. Fig. 5.7 shows the comparison on the 6 patients' data. My proposed groupwise TSD significantly reduced error and performed better than the spectral-graph-based method.

Phantom Data

I further tested groupwise TSD with the phantom data. The production of the phantom model was described in Section 4.4.2. The endoscopic video and 3D CT scan were collected for the model. I produced SfMS reconstructions for 600 frames in the video, among which 20 reconstructions were uniformly selected for geometry fusion. They were first downsampled to approximately 2500 vertices. Then they were aligned to the CT space via a similarity transform, whose transformation parameters were derived by a Monte Carlo method (Murphy, 2012). Since the phantom is rigid, there is supposed to be no deformation among the single-frame reconstructions, and they should

all perfectly match the CT surface. Hence, the groupwise registration plays the role of unifying inconsistent SfMS estimations. No outlier geometry trimming was performed in this test due to the good lighting condition in the phantom environment.

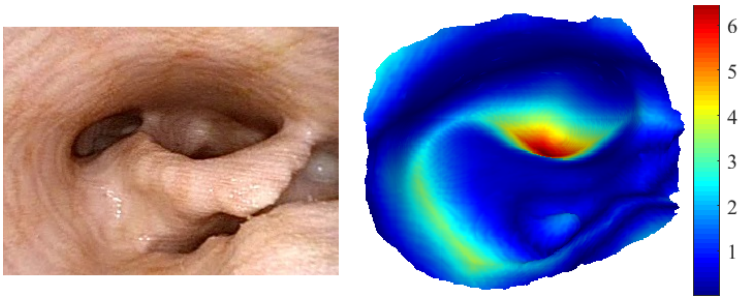


Figure 5.8: A phantom endoscopic video frame (left) and the fused geometry (right) with color-coded deviation (in millemeters) from the ground truth CT surface.

The registration error was measured slightly differently in this test. Since we have the CT surface as a ground truth geometry fusion result, we define a vertex's deviation as its distance to the nearest point in the CT surface. The average deviation of all vertices is **1.24mm** for the raw SfMS reconstructions and is **0.94mm** for the fused geometry, which shows that the registration can help filter out inaccurate SfMS estimations. Moreover, I found that more than 20 surfaces for geometry fusion will not further increase accuracy but will be computationally slower. Fig. 5.8 shows that the fused geometry resembles the ground truth CT surface in general. However, we can see that the error is significantly smaller in the near-camera regions and larger in the more distant regions. This is potentially caused by the flawed lighting model that is used in SfMS. Details of this reconstruction artifact are discussed in (Wang et al., 2017).

Real Patient Endoscopograms

I produced endoscopograms for 8 video sequences extracted from 4 patients' endoscopies. Each video sequence has around 150 frames (3-4 seconds long), among which 20 frames were uniformly sampled for geometry fusion. Outlier geometry trimming was used since lighting conditions were often poor. Since no registration ground truth is available in this test, I used the *overlap distance* (OD) introduced in (Huber and Hebert, 2003) as the accuracy metric. OD measures the average

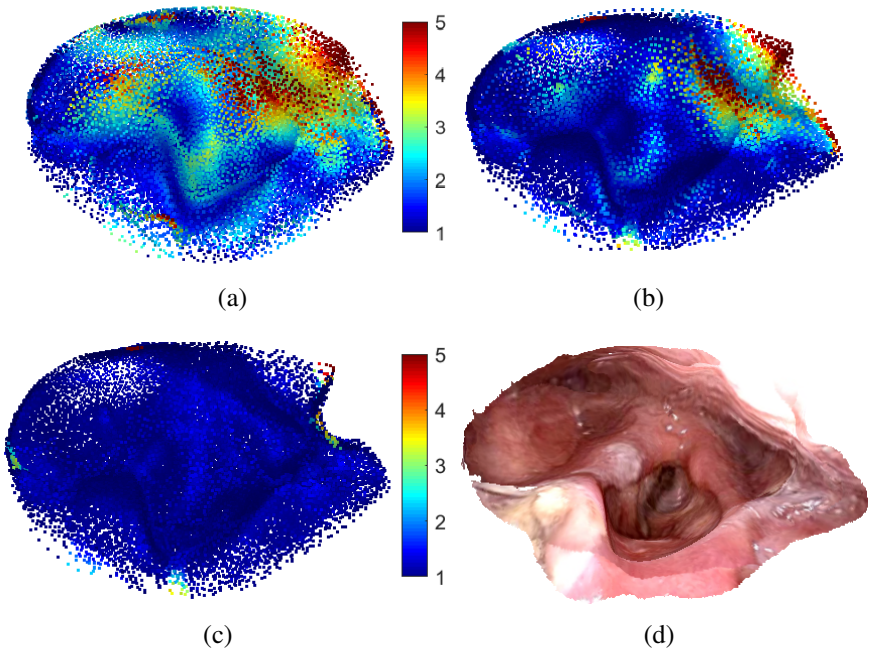


Figure 5.9: OD plot on the point cloud of 20 surfaces (a) before registration; (b) after registration; (c) after outlier geometry removal. (d) The final endoscopogram.

surface deviation between all pairs of overlapping regions. A zero OD means that the common regions among the N surfaces overlay with each other perfectly.

Before any registration, the initial OD of the 8 geometry fusion cases has an average of **1.6±0.13mm**. After groupwise TSD, the average OD is **0.58±0.05mm**. Finally after outlier geometry removal, the average OD is **0.24±0.09mm**. Fig. 5.9 shows the OD plot for one of the cases. Fig. 5.10 shows several other produced endoscopograms.

Figure 5.10: Four endoscopograms produced by the entire pipeline.

5.5 Discussion

I have described a pipeline for producing an endoscopogram from a video sequence. An essential step in the pipeline is the geometry fusion. I have proposed a novel groupwise surface registration algorithm and an outlier-geometry trimming algorithm. I have demonstrated via synthetic and phantom tests that my proposed N-body scenario is robust for registering partially-overlapping surfaces with missing data. Finally, I produced endoscopograms for real patient endsocopic videos.

Future Work

1. A current limitation of the pipeline is that the video sequence is at most 3-4 seconds long for robust SfM estimation. Future work involves fusing multiple endoscopograms from different sub-sequences of the video sequence.

2. An interesting finding is that the groupwise attraction formulation in Section 5.2.1 can be potentially interpreted from an entropy-minimization perspective. My observation is that if we assume a vertex, together with its corresponding vertices, follows a 3D Gaussian distribution, Eq. 5.2 actually defines the entropy-minimizing position for that vertex when all its correspondences' positions are fixed. This observation is enlightening because it can potentially bridge the gap between my method with other existing entropy-based groupwise registration methods (Cates et al., 2007). However, details of this idea need to be further examined.

3. A remaining problem is the convergence property of the groupwise registration; that is, whether the surface deformations are guaranteed to converge. In Section 5.2.1, I have formulated the registration in analog to the physical N-body problem. In that original problem, it has been proven that the N bodies can converge to a stable solution in the limit of time when there is none of certain types of singularities. However, the proof of the convergence in my proposed approach is more complicated in two ways:

(a) The forces are simply vectors connecting correspondences rather than real gravity forces. The vectors are also weighted by the confidence scores in the objective function.

(b) The correspondences change over iterations with geometric feature matching updates.

For the above two reasons, it is not obvious that the deformations are guaranteed to converge. However, one can imagine that when the surfaces are close enough, the geometric feature matching is likely to produce constant correspondences. In that situation, due to the entropy-minimization nature discussed above, each set of corresponding points will converge to a single location.

Even though the convergence is not theoretically guaranteed, I have not observed any divergence situation in practice. In fact, the current setting of the algorithm is to stop after a fixed number of iterations, so the final convergence property becomes less critical.

CHAPTER 6

Anisotropic Elasticity Estimation

A key factor in Thin Shell Demons is that it uses elastic thin shell deformation energy as the deformation regularization term. In fact, the idea of adopting elastic models dates back to (Bajcsy and Kovačič, 1989) and has been extensively studied since then. Even though methods based on some other regularization formulations (Christensen et al., 1996; Thirion, 1998; Beg et al., 2005) have also produced reasonable registration results, the elasticity-model-based regularization is particularly appealing because in many medical applications anatomical deformations are indeed elastic processes caused by muscles or other forces.

The key to realistic physical modeling is to apply proper elasticity parameters. In the previous two chapters, there are the only two physical parameters, namely the Young's modulus E and Poisson's ratio ν, and these have been determined empirically and set to be identical everywhere on the shell. This is equivalent to having two underlying assumptions in the thin shell model:

(1) *Isotropy*: for any local region in a shell, its elastic property is the same along any direction.

(2) *Homogeneity*: the elastic property is the same everywhere on the shell.

Eq. 2.21 exhibits the fact that $\{E, \nu\}$ are the only elastic parameters for a homogeneous and isotropic shell.

In fact, the homogeneity and isotropy assumption has become prevalent in most existing registration algorithms, where only a single regularization weighting parameter is used for the entire domain. However, studies have shown that human tissue elasticity is both *inhomogeneous* (e.g., different tissue types show different stiffness) (Miga, 2003) and *anisotropic* (e.g., different stiffness

along and across the tissue fiber direction) (Kroon and Holzapfel, 2008). In our nasopharynx case, we also observe such phenomena. For example, in a swallowing motion,

(a) The epiglottis primarily bends in the longitudinal direction and is much stiffer along the transverse direction (anisotropy);

(b) The pharyngeal wall contracts and expands primarily in the transverse direction (anisotropy);

(c) The arytenoid cartilages are stiffer than the soft tissues (imhomogeneity), etc.

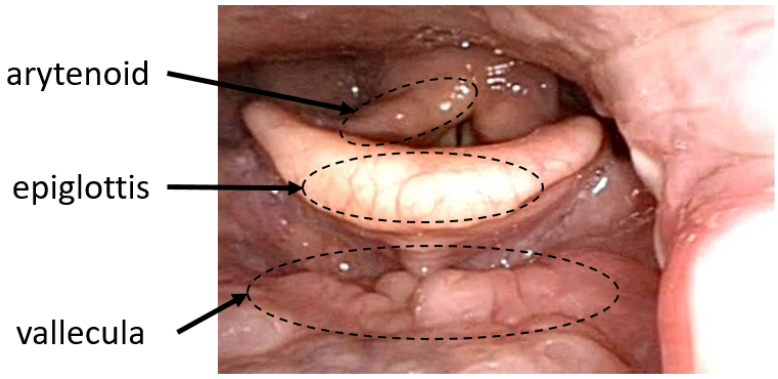


Figure 6.1: Anatomy of the pharynx: the three tissue types shown above have different anisotropic elasticity properties.

These facts render the simple two-parameter thin shell model unrealistic.

The above argument motivates research interest in studying spatially varying tissue elasticity, not only for registration, but also for simulation (Schneider et al., 2009) and pathology analysis (Yang et al., 2016). However, most approaches for non-uniform elasticity estimation have to use a sophisticated mechanical system equipped with a force generation/measurement capability, which is not available in our registration setting. In our application, the only accessible data that can reveal tissue elastic properties are the tissue deformations revealed in an endoscopic movie sequence. Therefore, it is desirable to have a method that can estimate spatially varying anisotropic elasticity parameters only using a set of known material deformations.

In this chapter, I will introduce a method for deriving elasticity parameters from deformations that have happened within a patient (e.g., the above mentioned tissue deformation in a movie sequence). I first propose to incorporate the thin shell model with *orthotropic* elasticity (Section 2.3.2). Next I propose the following two statistical frameworks:

(1) *Elasticity Estimation*: I show that with some proper prior assumptions, spatially inhomogeneous and orthotropic elasticity parameters can be estimated from a set of known shell (surface) deformations via a novel *maximum-a-posteriori* (MAP) optimization. This framework will be discussed in Section 6.3.

(2) *Registration and Elasticity Joint Estimation*: With the first framework, I show that we can jointly estimate the shell elasticity parameters and shell deformations in a groupwise registration setting. Such a joint estimation provides significantly better registration accuracy than the old groupwise TSD method, where only an isotropic homogeneous thin shell model is used. This joint estimation framework will be discussed in Section 6.4.

In the geometry fusion step of the endoscopogram construction pipeline, the above two frameworks can retrieve insightful tissue elastic properties from the deformations among the group of single-frame reconstructions, and they can in turn improve the geometry fusion result.

6.1 Related Work

Closely related to our work is a research branch known as *spatially-varying registration*, the idea of which is to let regularization strength be dependent on location. This can be modeled by spatially-varying diffusion (Freiman et al., 2011), non-stationary Gaussian processes (Gerig et al., 2014) or applying a non-stationary metric in the LDDMM setting (Vialard and Risser, 2014). Despite their theoretical appeal, those methods explore the problem mostly from the computational aspect and lack physical motivation, and they also don't handle the anisotropic situation. The notion of spatially-varying registration has been also used in elastic models (Davatzikos, 1997; Alterovitz et al., 2006), but the elasticity parameters are only associated with a small number of known segmented tissue types.

Automatic elasticity estimation has been studied in different medical applications (Misra et al., 2010; Kroon and Holzapfel, 2008). With tissue displacements and external forces taken as known values, the elasticity can be computed directly as an inverse problem of the Finite Element Method

(FEM). Elastography is another widely used non-invasive procedure for determining local elastic properties, but it either requires a force exertion/measurement device or a vibration actuation mechanism (Green et al., 2013), which is often not available in other imaging modalities.

Therefore, *modality-independent* and purely *image-based* approaches are desired and have been under investigation for several years. Miga et. al. (Miga, 2003) introduced registration-based elastography to estimate tissue stiffness of an object given two images of it undergoing an elastic deformation. Risholm et. al. (Risholm et al., 2011) extended this approach by forming a probabilistic model over the registration parameters and inhomogeneous isotropic elasticity parameters. While my proposed framework is related to theirs, mine is different by incorporating anisotropy and by applying the model to surface data.

In a broader context, *Statistical Shape Analysis* seeks a statistical distribution or a low dimensional subspace, called a shape space, for describing a given set of shapes (or shape deformations). Most existing approaches (Bauer et al., 2014; Schulz et al., 2015) construct the shape space by constraining the shape's global appearance, such as deformation vector fields, point positions or normal directions. Our framework provides an alternative perspective in the sense that it seeks the underlying physical reason that can best explain the given shape deformations.

6.2 Orthotropic Thin Shell

In material science and solid mechanics, orthotropy is referred to as a special kind of anisotropy. In the most general 3D situation, orthotropic materials have material properties that differ along three mutually-orthogonal axes of rotational symmetry. In the thin shell situation, when out-of-tangent-plane strain is neglected, the elastic model is reduced to 2D; that is, for each local point on a shell, its elastic properties are symmetric with respect to two orthogonal axes (tangent vectors), known as natural axes. Orthotropy has shown its effectiveness in modeling fibrous tissues in the situations where the stiffness is usually different in a direction parallel to the fibers than in the transverse direction (Kroon and Holzapfel, 2008; Schneider et al., 2009). This model is also a reasonable

choice in our application, because most throat anatomies show different deformation flexibility along two different directions (the epiglottis and pharyngeal wall usually show different amount of motion along transverse and longitudinal directions). Moreover, unlike arbitrary anisotropy, the orthotropic model has fewer model parameters to fit and thus tends to be more stable in producing simulated deformations (Li and Barbic, 2015).

Details of thin shell mechanics were introduced in section 2.3. To remind the readers, local elasticity of a point on a shell is characterized by a 3×3 positive definite matrix C, called a stiffness matrix. The C matrix can be an arbitrary positive definite matrix for general anisotropy. In the special case of orthotropy, if we parameterize all the quantities using a local coordinate frame made by the natural axes, the stiffness matrix can be written as the following simplified form with many zero entries:

$$C = \begin{bmatrix} c_1 & c_2 & 0 \\ c_2 & c_3 & 0 \\ 0 & 0 & c_4 \end{bmatrix} = \frac{1}{1-\nu_{xy}\nu_{yx}} \begin{bmatrix} E_x & \nu_{vu}E_x & 0 \\ \nu_{xy}E_y & E_y & 0 \\ 0 & 0 & 2G_{xy}(1-\nu_{xy}\nu_{yx}) \end{bmatrix}, \tag{6.1}$$

where $\{E_x, E_y\}$ are the Young's moduli along the natural axes, $\{\nu_{xy}, \nu_{yx}\}$ are the Poisson's ratios, and G_{xy} is the shear modulus. In the following text, I will use C to specifically denote such a simplified matrix, called the *canonical orthotropic stiffness matrix*. Such matrices live in the positive definite matrix space, denoted as $\mathbf{SPD}_C$.

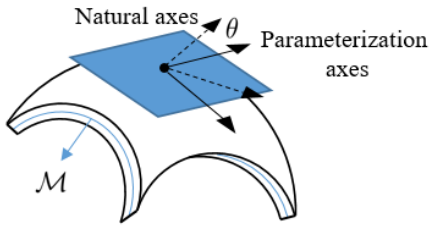


Figure 6.2: A thin shell model: for a local point, the elastic properties on the tangent plane (blue) are symmetric with respect to two natural axes. The local strains may be parameterized by any other orthogonal frame. The angle, θ, between the two frames is known as the canonical angle.

With the stiffness matrix, the *local* thin shell deformation energy is measured by the W function:

$$W(\varphi, \kappa, C) = \lambda_{mem}\varphi^T C\varphi + \lambda_{bend}\kappa^T C\kappa, \tag{6.2}$$

where $\varphi = [\varphi_{xx}, \varphi_{yy}, \varphi_{xy}]^T$ is the tangential Cauchy-Green strain tensor characterizing local stretching, $\kappa = [\kappa_{xx}, \kappa_{yy}, \kappa_{xy}]^T$ is the shape operator difference characterizing local bending (local curvature change), and $\{\lambda_{mem}, \lambda_{bend}\}$ are the global mixing weights determined by shell thickness (the same definition as in Chapter 4).

Rotation of Frame. When σ and ε are parameterized under an arbitrary orthogonal frame (Fig. 6.2) instead of rotated into the natural axes, we have the following relationship:

$$\begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} \\ \varepsilon_{xy} & \varepsilon_{yy} \end{bmatrix} = \begin{bmatrix} cos(\theta) & sin(\theta) \\ -sin(\theta) & cos(\theta) \end{bmatrix} \begin{bmatrix} \varepsilon'_{xx} & \varepsilon'_{xy} \\ \varepsilon'_{xy} & \varepsilon'_{yy} \end{bmatrix} \begin{bmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{bmatrix}, \tag{6.3}$$

where $\{\varepsilon'_{\alpha\beta}\}$ is the strain tensor parameterized under an arbitrary orthogonal frame and θ is the rotation angle between the two frames, known as the *canonical angle*. The same rotational relationship applies for σ. Combining with Eq. 2.20, the stiffness matrix under an arbitrary frame is $C' = R^{-1}CR$, where

$$R = \begin{bmatrix} cos(\theta)^2 & sin(\theta)^2 & 2cos(\theta)sin(\theta) \\ sin(\theta)^2 & cos(\theta)^2 & -2cos(\theta)sin(\theta) \\ -cos(\theta)sin(\theta) & cos(\theta)sin(\theta) & cos(\theta)^2 - sin(\theta)^2 \end{bmatrix}. \tag{6.4}$$

In other words, the combination of C and θ uniquely determines the *local* orthotropic stiffness matrix parameterized under an arbitrary frame. The associated W deformation energy function becomes

$$W(\varphi, \kappa, C, \theta) = \lambda_{mem}\varphi^T C'\varphi + \lambda_{bend}\kappa^T C'\kappa, \tag{6.5}$$

6.3 Elasticity Estimation via MAP

In the registration setting described in Chapters 4 & 5, both the material deformations and material elasticity are unknown variables. In this section, we first assume the deformations are given and investigate only elasticity estimation. The joint estimation of deformation and elasticity will be discussed later in the next section.

We assume some observed tissue deformations are the realization of tissue elasticity of a single patient. Then a common way to estimate elasticity parameters is to solve an inverse problem given such deformations and given external force measurements (how much force is applied to the material to induce the deformation). However, when only the material deformations are available, the parameter estimation can be highly ill-posed. Therefore, I opt for energy-based models that are commonly used in statistical mechanics (Jaynes, 1957). In these models high probability states are associated with low energy configurations. Here I propose a *Physical-Energy-Based Markov Random Field* (MRF) model. The idea is to find a set of elasticity parameters to "best explain" the observed deformations from a probabilistic point of view, which indicates that the associated total deformation energy has to be minimized.

6.3.1 Problem Statement

Given a reference shell $\mathcal{M}$ and a set of N observed deformations $\mathbf{\Phi} = \{\Phi^\alpha : \mathcal{M} \to \mathbb{R}^3 | \alpha = 1...N\}$, our goal is to find the canonical orthotropic stiffness matrix function $\mathbf{C} : \mathcal{M} \to \mathbf{SPD}_C$ and the canonical angle function $\boldsymbol{\theta} : \mathcal{M} \to \mathbf{S}^1$, namely the orthotropic elasticity parameters of every location on the shell. To simplify the problem, the continuous shell $\mathcal{M}$ is first discretized to a triangle-mesh $\{\mathcal{T}^j | j = 1...M\}$ with M triangles. Then each triangle $\mathcal{T}^j$ is associated with its own local elasticity parameters (C^j, θ^j). Each deformation Φ^α is reparameterized locally on the tangent planes (triangles); that is, for each deformation Φ^α, each triangle $\mathcal{T}^j$ has its own local stretching strain $\varphi^{\alpha j}$ and bending strain $\kappa^{\alpha j}$. In other words, the group of deformations $\{\Phi^\alpha | \alpha = 1...N\}$ are represented by the set of local strains $\{\varphi^{\alpha j}, \kappa^{\alpha j} | \alpha = 1...N, j = 1...M\}$. Finally, the goal is

to estimate the set of elasticity parameters $(\mathbf{C}, \boldsymbol{\theta}) = \{C^j, \theta^j | j = 1...M\}$ for all triangles from the given set of strains $\{\varphi^{\alpha j}, \kappa^{\alpha j} | \alpha = 1...N, j = 1...M\}$.

6.3.2 The MRF Model

My idea is that the local elasticity parameters for a triangle should yield small deformation energy for the associated local strains. Meanwhile, we assume the parameters should vary smoothly across the shell. Therefore, I build an MRF model on the dual graph of the triangle mesh as shown in Fig. 6.3. Each node on the graph represents the elasticity parameters to be estimated for the underlying triangle. We want to find $(\mathbf{C}, \boldsymbol{\theta})$ to maximize the posterior distribution $p(\mathbf{C}, \boldsymbol{\theta} | \boldsymbol{\Phi})$:

$$p(\mathbf{C}, \boldsymbol{\theta} | \boldsymbol{\Phi}) \propto p(\boldsymbol{\Phi} | \mathbf{C}, \boldsymbol{\theta}) p(\mathbf{C}, \boldsymbol{\theta}). \tag{6.6}$$

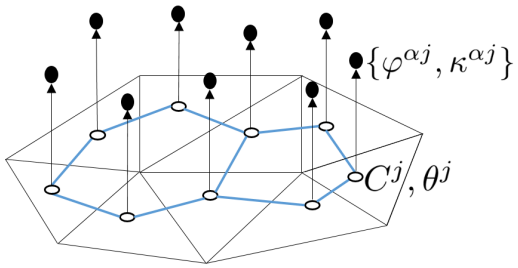


Figure 6.3: A Gaussian MRF model with nodes (white) defined on the dual graph (blue) of a triangle mesh. Node j (triangle $\mathcal{T}^j$) is associated with unknown variables (C^j, θ^j) and a set of observed variables $\{\varphi^{\alpha j}, \kappa^{\alpha j} | \alpha = 1...N\}$

Likelihood. The likelihood $p(\boldsymbol{\Phi} | \mathbf{C}, \boldsymbol{\theta})$ is associated with the total deformation energy of all observed deformations. Assuming local deformation energy follows independent Boltzmann distributions, I design the likelihood as the following:

$$p(\boldsymbol{\Phi} | \mathbf{C}, \boldsymbol{\theta}) = \prod_{\alpha} \prod_{j} p(\varphi^{\alpha j}, \kappa^{\alpha j} | C^j, \theta^j) \propto \prod_{\alpha} \prod_{j} \exp(-W(\varphi^{\alpha j}, \kappa^{\alpha j}, C^j, \theta^j)) \tag{6.7}$$

We can see that after taking the negative logarithm of the above equation, this is equivalent to the discrete version of Eq. 2.24, i.e., the total energy being the integration of local energies over all triangles.

Prior. Now consider the prior distribution of $p(\mathbf{C}, \boldsymbol{\theta})$, the second term of Eq. 6.6. In Bayesian frameworks, having proper prior assumptions plays an important role in handling noisy data observations. Here I make two types of heuristic prior assumptions: each node has its own per-node prior function ψ_j, and each edge has an edge potential function ψ_{ij}. Then the overall prior distribution is designed in the form:

$$p(\mathbf{C}, \boldsymbol{\theta}) \propto \prod_j \psi_j(C^j) \prod_{i,j} \psi_{ij}(C^i, \theta^i, C^j, \theta^j). \tag{6.8}$$

Per-Edge Prior. The ψ_{ij} function models the spatial smoothness nature of the shell's orthotropic property by penalizing the difference of stiffness matrices between neighboring nodes and the smoothness of the natural-axes direction field. We assume $\boldsymbol{C}$ and θ are independent and design

$$\psi_{ij}(C^i, \theta^i, C^j, \theta^j) \propto \exp(-d(C^i, C^j)^2) \cdot \exp(-(p_{ij}(\theta^i) - \theta^j)^2). \tag{6.9}$$

In the first factor on the right-hand side, $d(\cdot, \cdot)$ is a proper distance metric for $\mathbf{SPD}_C$, which encourages neighboring stiffness matrices to be similar. I use the Log-Euclidean metric (Arsigny et al., 2006) for its computational convenience. Some other commonly used metrics include the Affine-invariant (Pennec et al., 2006) and the Scaling-rotation metric (Jung et al., 2015).

In the second factor on the right-hand side, $p(\cdot)$ is the Levi-Civita parallel transport operation (Crane et al., 2010) that transports the vector associated with θ_i on $\mathcal{T}^i$ to the neighboring triangle $\mathcal{T}^j$. Fig. 6.4 illustrates this discrete parallel transport on neighboring triangles. In that figure, we use the x-y frame to denote the two natural axes. The natural axes are transported from $\mathcal{T}^i$ to $\mathcal{T}^j$ to be compared with the ones in $\mathcal{T}^j$. The angle of rotation, $(p_{ij}(\theta^i) - \theta^j)$, is measured by either the x-axis rotation or the y-axis rotation. In doing this, we assume that the correspondence of the two axis directions is well defined between neighboring locations; that is, in the above figure, the

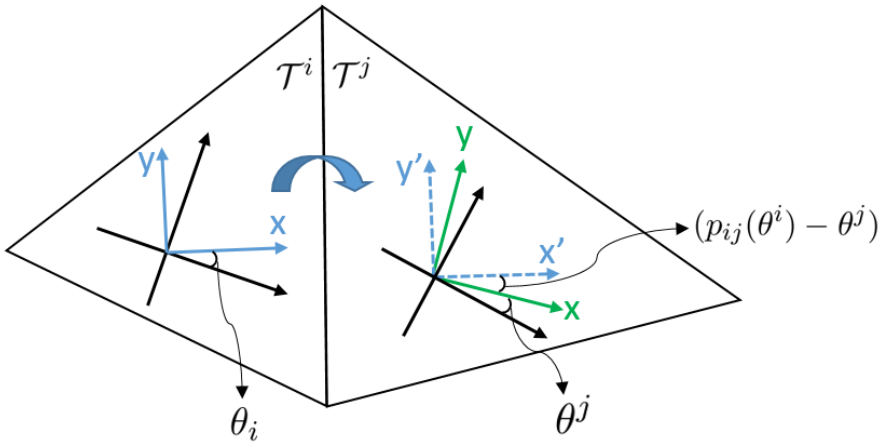


Figure 6.4: Illustration of discrete parallel transport between two neighboring triangles. The black frames are the parametrization frames, and the colored frames are the natural axes. The natural axes of $\mathcal{T}^i$ (blue) are transported into $\mathcal{T}^j$ and are compared with the natural axes of $\mathcal{T}^j$ (green).

blue x-axis corresponds to the green x-axis. However, this assumption ignores the existence of singularities, i.e., points that have isotropic elasticity. The mechanism in Fig. 6.4 does not apply to those singular points because they don't have preferred natural axes with the same stiffness along all directions (Crane et al., 2010). In this dissertation, I leave this situation as an open problem and will further study it in the future. In our application, we only deal with surfaces with boundary. Such surfaces theoretically don't have to have singularities. Therefore, we assume there is no singular point, so Eq. 6.4 can be applied everywhere in the surface.

Per-Node Prior. A common challenge for image-based elasticity estimation without force measurements is the scale ambiguity. For example, observing a large deformation can either indicate a small material stiffness or a large force being applied. Thus, a direct energy minimization of Eq. 6.7 without any scale constraint will lead to a trivial solution (zero-stiffness). To avoid this, existing methods (Yang et al., 2016; Risholm et al., 2011) usually set a heuristic scale based on prior knowledge and recover elasticity parameters only relative to that given scale. In my work, I formulate this scale prior as the per-node prior function ψ_j. To be specific, I assume anisotropy is distributed in a Gaussian sense with the isotropic case being the mean situation. Therefore, given an

isotropic prior Young's modulus E and Poisson's ratio ν, I design

$$\psi_j(C^j) \propto exp(-d(C^j, \bar{C})^2), \tag{6.10}$$

where $\bar{C}$ is the commonly used isotropic stiffness matrix derived from E and ν. In this way, the estimated orthotropic parameters are only relative to the given isotropic $\bar{C}$.

MAP Analysis. Finally, we optimize the negative log posterior in the following form:

$$\begin{aligned} -\log p(\mathbf{C}, \boldsymbol{\theta}|\boldsymbol{\Phi}) = &\sum_{i,j}[\lambda_1 d(C^i, C^j)^2 + \lambda_2(p_{ij}(\theta^i) - \theta^j)^2] + \\ &\sum_{\alpha}\sum_{j} W(\epsilon^{\alpha j}, \kappa^{\alpha j}, C^j, \theta^j) + \sum_j \lambda_3 d(C^j, \bar{C})^2, \end{aligned} \tag{6.11}$$

where the λ parameters are the global weights. Because the stiffness matrices are represented in the Log-Euclidean form (Arsigny et al., 2006), the objective function is differentiable everywhere with respect to $(\mathbf{C}, \boldsymbol{\theta})$ and can be minimized via any gradient descent method.

6.3.3 A Toy Example.

In this subsection, I illustrate the use of the above elasticity estimation framework via a simple toy example. We consider a bar-shaped surface shown in Fig. 6.5a. The ground truth elasticity parameters on the surface, including both the orthotropic canonical stiffness matrices and natural axes directions, were manually designed. Fig. 6.5b visualizes the two Young's moduli along the two natural axes directions respectively. The bar is more elastic at the center (inhomogeneity) and more elastic along the vertical direction (orthotropy). The other elasticity parameters were set to satisfy $\nu_{xy}\nu_{yx} = 0.25^2$ and $G_{xy} = 2kPa$. I simulated 20 synthetic deformations to the bar (Fig. 6.5d) by first fixing its two ends at random positions as boundary constraints and then optimizing Eq. 6.7 to solve for the deformations using ground truth elasticity. The resulting synthetic deformations were taken as the group of observed deformations.

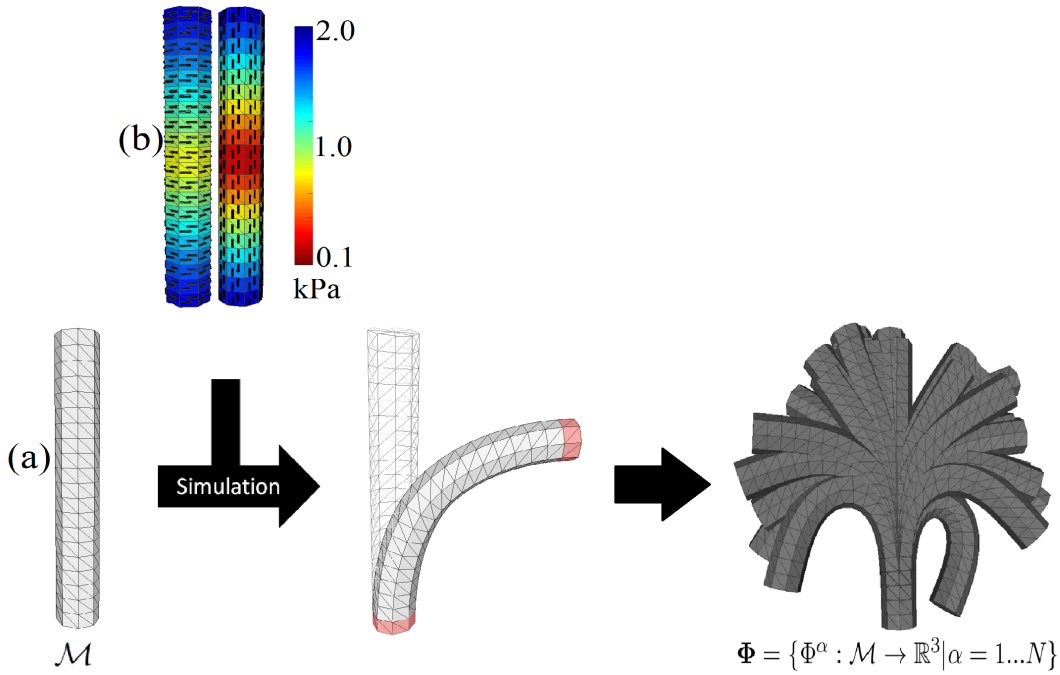


Figure 6.5: (a) A reference bar-shaped surface. (b) The two ground truth Young's moduli are respectively color-coded across the surface. Red regions indicate smaller Young's moduli (more elastic). Each local Young's modulus is associated with a natural axis direction (black vector fields) (c) A surface deformation can be derived by first fixing its two ends at designated positions and by optimizing Eq. 7 using ground truth elasticity. (d) The group of simulated deformations $\mathbf{\Phi}$.

Now we can perform the elasticity estimation using Eq. 6.11. The weighting parameters $\{\lambda_1 = 1, \lambda_2 = 10, \lambda_3 = 0.1\}$ were chosen empirically to best fit this toy problem. I found that the natural axes first have to be accurate to yield meaningful anisotropy, so I set a larger λ_2 to regularize the vector field. Other model parameters $\{\lambda_{mem} = 80, \lambda_{bend} = 10\}$ and the isotropic prior $\{E = 2\text{kPa}, \nu = 0.25\}$ were chosen the same as before.

I mainly investigated the estimation accuracy of the three most important parameters in the orthotropic model: the canonical angle gives major anisotropic directions, and the two Young's moduli characterize the corresponding stiffness along the two directions. Fig. 6.6a shows the two estimated Young's moduli and the estimated natural axes directions. The average canonical angle error is 0.74 degree, which shows we can successfully estimate the natural axes directions. Fig. 6.6b shows that with the estimated orthotropic elasticity parameters the simulated deformation is more accurate in the sense that the center-elastic part has a larger bending effect than the one simulated

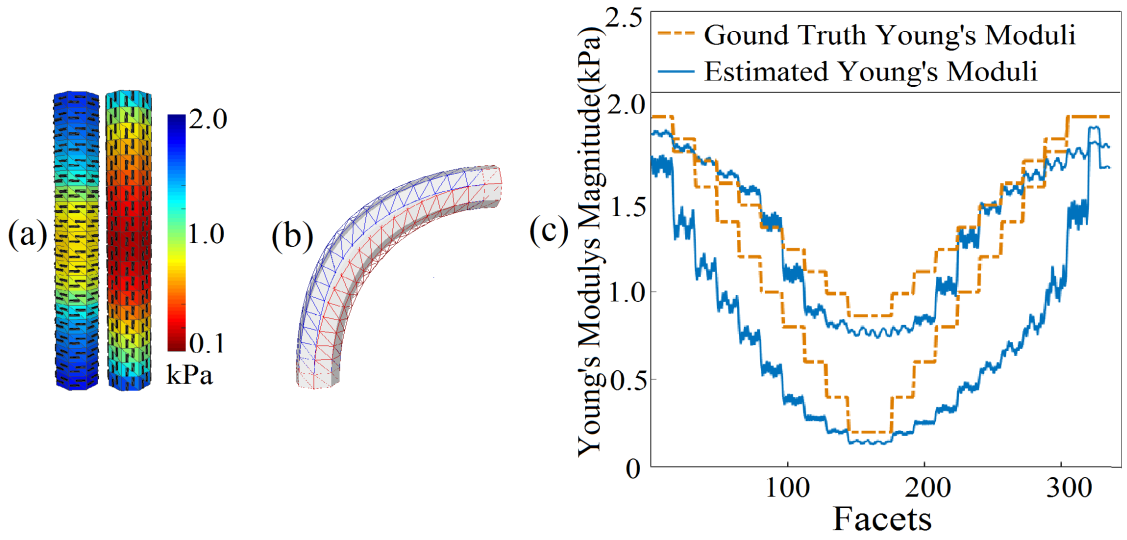


Figure 6.6: (a) Estimated Young's moduli and the associated estimates of natural axes. (b) Simulated surface deformations derived from ground truth elasticity (blue wireframe), estimated elasticity (gray surface) and isotropic elasticity (red frame). (c) The two ground truth Young's moduli (the orange curves) and the two estimated Young's moduli (the blue curves) on all faces.

from isotropic elasticity under the same boundary constraints. Finally, Fig. 6.6c quantitatively shows the estimation error of the two Young's moduli on all triangles. It is worth mentioning that the magnitude of the estimation is dependent on the given scale of $\bar{C}$, which means we will see vertical fluctuation of blues curves given different choices of $\bar{C}$.

6.4 Joint Estimation of Registration and Elasticity

In the groupwise registration analysis in Chapter 5, both the material deformations and elasticity parameters are unknown variables. Formally, we want to investigate the joint probability of the group of deformations $\mathbf{\Phi}$ and elasticity parameters $(\mathbf{C}, \boldsymbol{\theta})$, given a reference shell $\mathcal{M}$ and its many deformed versions $\{\mathcal{M}^{\alpha} | i = \alpha ... N\}$. A common approach is to treat one set of variables, e.g., $(\mathbf{C}, \boldsymbol{\theta})$, as latent variables and perform an *Expectation-Maximization* algorithm to estimate the *marginal posterior*

$$p(\mathbf{\Phi}|\boldsymbol{\mathcal{M}}) = \int_{\mathbf{C},\boldsymbol{\theta}} p(\mathbf{\Phi}, \mathbf{C}, \boldsymbol{\theta}|\boldsymbol{\mathcal{M}}), \tag{6.12}$$

where $\boldsymbol{\mathcal{M}}$ denotes the combination of the given data $\mathcal{M}$ and $\{\mathcal{M}^{i} | i = 1 ... N\}$.

The EM algorithm (Murphy, 2012) seeks to find the mode of the above marginal posterior by iteratively applying these two steps:

Expectation step (E step): Calculate the expected value of the log marginal posterior function, with respect to the conditional distribution on latent variables $(\mathbf{C}, \boldsymbol{\theta})$ under the current estimate of the deformations $\boldsymbol{\Phi}^{(k)}$. Construct a so-called Q-function in the following form:

$$Q(\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}) = \mathrm{E}_{\mathbf{C},\boldsymbol{\theta}|\boldsymbol{\mathcal{M}},\boldsymbol{\Phi}^{(k)}}[\log p(\boldsymbol{\Phi}, \mathbf{C}, \boldsymbol{\theta}|\boldsymbol{\mathcal{M}})] \tag{6.13}$$

Maximization step (M step): Update the deformations by maximizing the Q-function:

$$\boldsymbol{\mathcal{D}}^{(\boldsymbol{k+1})} = \arg\max_{\boldsymbol{\Phi}} Q(\boldsymbol{\Phi}|\boldsymbol{\Phi}^{(k)}) \tag{6.14}$$

The challenge in this EM algorithm is that the integral required in the expectation operation of the E-step is intractable, so we cannot differentiate the objective function in the M-step. Many existing works (Risholm et al., 2011; Zhang et al., 2013) proposed to first draw a large number of samples from the conditional distribution on $(\mathbf{C}, \boldsymbol{\theta})$ and then to use the sample average to approximate the expectation operation. However, this Monte-Carlo-based approximation leads to challenges in reliable and efficient sampling: for example, Gibbs sampling is computationally expensive, and determining an appropriate burn-in period poses serious theoretical challenges. Therefore, in this work, I opt to use the "mode approximation" trick to approximate the conditional distribution on $(\mathbf{C}, \boldsymbol{\theta})$. More precisely, I adopt the following approximation:

$$p(\mathbf{C}, \boldsymbol{\theta}|\boldsymbol{\mathcal{M}}, \boldsymbol{\Phi}^{(\boldsymbol{k})}) \approx \delta(\mathbf{C}^*, \boldsymbol{\theta}^*), \tag{6.15}$$

where

$$\mathbf{C}^*, \boldsymbol{\theta}^* = \arg\max_{\mathbf{C},\boldsymbol{\theta}} \log p(\mathbf{C}, \boldsymbol{\theta}|\boldsymbol{\mathcal{M}}, \boldsymbol{\Phi}^{(\boldsymbol{k})}). \tag{6.16}$$

In other words, the probability density over the entire distribution is aggregated to the position of its mode. In this way, the Q function in the E-step can be approximated by

$$Q(\mathbf{\Phi}|\mathbf{\Phi}^{(k)}) \approx \hat{Q}(\mathbf{\Phi}|\mathbf{\Phi}^{(k)}) = \log p(\mathbf{\Phi}, \mathbf{C}^*, \boldsymbol{\theta}^*|\boldsymbol{\mathcal{M}}), \tag{6.17}$$

and the M-step then becomes another MAP analysis on $\mathbf{\Phi}$:

$$\begin{aligned} \boldsymbol{\mathcal{D}}^{(\boldsymbol{k+1})} &= \arg\max_{\mathbf{\Phi}} \hat{Q}(\mathbf{\Phi}|\mathbf{\Phi}^{(k)}) && (6.18)\\ &= \arg\max_{\mathbf{\Phi}} [\log p(\mathbf{\Phi}|\mathbf{C}^*, \boldsymbol{\theta}^*, \boldsymbol{\mathcal{M}}) + \log p(\mathbf{C}^*, \boldsymbol{\theta}^*)] && (6.19)\\ &= \arg\max_{\mathbf{\Phi}} [\log p(\mathbf{\Phi}|\mathbf{C}^*, \boldsymbol{\theta}^*, \boldsymbol{\mathcal{M}})] && (6.20) \end{aligned}$$

The above approximations lead to a variant of the EM algorithm, which known as *alternating optimization* or *hard EM* (Murphy, 2012). When applied in our application, we get our *registration/elasticity joint estimation* framework:

1. *Input*: a reference shell $\mathcal{M}$ and a set of deformed shells $\{\mathcal{M}^\alpha|\alpha = 1...N\}$. Initialize the elasticity parameters to be $\bar{C}$ everywhere.

2. In iteration (k), with the current estimate of $(\mathbf{C}^{(k)}, \boldsymbol{\theta}^{(k)})$, perform MAP to update $\mathbf{\Phi}$:

$$\mathbf{\Phi}^{(k)} = \arg\min_{\mathbf{\Phi}} [-\log p(\mathbf{\Phi}|\mathbf{C}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\mathcal{M}})] \tag{6.21}$$

   This step is essentially to perform surface registrations between $\mathcal{M}$ and $\{\mathcal{M}^\alpha|\alpha = 1...N\}$ given the current estimate of the inhomogeneous and orthotropic elasticity. We assume this groupwise registration can be decomposed into a group of independent pairwise registrations, which means

$$p(\mathbf{\Phi}|\mathbf{C}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\mathcal{M}}) = \prod_{\alpha=1}^{N} p(\Phi^\alpha|\mathbf{C}^{(k)}, \boldsymbol{\theta}^{(k)}, \mathcal{M}, \mathcal{M}^\alpha) \tag{6.22}$$

Therefore, we can update the deformation Φ^α of each registration independently. The pairwise registration can be accomplished using Thin Shell Demons algorithm with the orthotropic energy as deformation regularization.

3. Given the current estimate of $\mathbf{\Phi}^{(k)}$, perform the elasticity estimation framework introduced in the previous section,

$$(\mathbf{C}^{(k+1)}, \boldsymbol{\theta}^{(k+1)}) = \arg\min_{\mathbf{C},\boldsymbol{\theta}} [-\log p(\mathbf{C}, \boldsymbol{\theta} | \mathbf{\Phi}^{(k)}, \mathcal{M})], \tag{6.23}$$

to update the elasticity estimation.

4. Iterate Steps 2 and 3 until convergence.

6.5 Experiments

Synthetic Head-and-Neck Data. I tested the two frameworks, namely the elasticity estimation framework and the joint estimation framework, with synthetic deformations on 5 real patients' head-and-neck CT data. For each patient, the CT segmentation surface was used as the reference surface. The construction of simulated surface deformations was similar to the toy example: for each patient's CT surface,

(1) I manually assigned ground truth orthotropic elasticity parameters and natural axes directions to the reference surface to reflect known anatomical facts: the epiglottis being stiffer than the vallecula and the pharyngeal wall being more elastic cross-sectionally (Fig. 6.8).

(2) I simulated 20 synthetic deformations from 20 manually constructed boundary conditions. These deformations include the expansion/compression of the pharyngeal wall and the opening/-closing of the epiglottis.

Fig. 6.7 shows a CT surface and one of its deformed versions. To test the elasticity estimation framework alone, I estimated the elasticity parameters of the reference surface directly from the

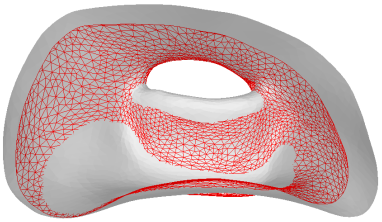


Figure 6.7: A reference surface (gray surface) and one of its synthetic deformations (red wireframe).

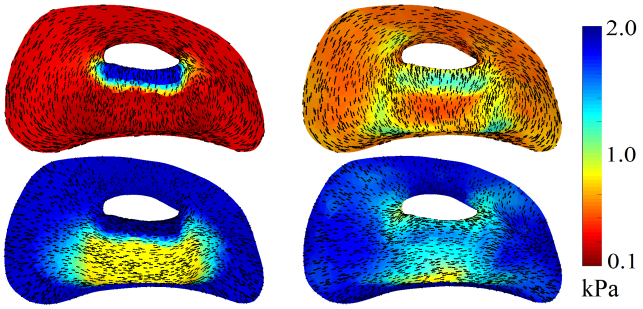


Figure 6.8: (a) Ground truth Young's moduli along the two natural axes. The epiglottis (blue region in the top figure) is set to be stiffer than the vallecula (yellow region in the bottom figure). (b) Estimated elasticity using ground truth deformations.

above 20 simulated deformations. I tuned λ_1 down to 0.1 to avoid overly smoothing the estimation. All the other algorithm parameters were kept the same as in the toy example.

Fig. 6.8 shows the results for one case. We can tell that the general pattern of the two Young's moduli and the natural axes can be reasonably recovered, but the scale difference with the ground truth suggests the proposed method only recovers parameters up to a scale relative to the prior isotropic elasticity, which is consistent with our expectation in Section 6.3.2. Moreover, the elasticity-smoothness term in Eq. 6.7 tends to yield blurred estimation. Due to these artifacts, the average error over all facets for the two Young's moduli are 0.41kPa and 0.38kPa respectively. The average canonical angle error is 12 degrees.

To test joint estimation accuracy, I performed one iteration of the framework introduced in Section 6.4. The 20 simulated deformations were used as ground truth deformations for the registration. To be more specific, I first performed 20 independent registrations between the reference CT surface and the 20 simulated surfaces using isotropic elasticity, followed by an

elasticity estimation using the 20 resulting deformations. This gives estimated orthotropic elasticity on the reference surface for second round registrations. Fig. 6.9 gives accuracy measurement under different registration options. The error is computed as the average surface registration errors over all 5 cases. The abscissa (x-axis) in Fig. 6.9 denotes the iterations within the Thin Shell Demons registration, not to be confused with the overall joint estimation iteration.

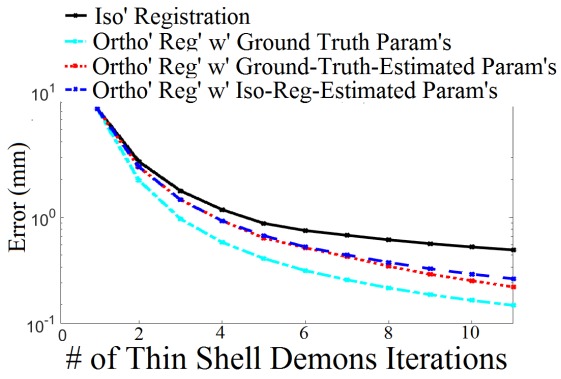


Figure 6.9: Registration accuracy over registration iterations under different options. The 2^{nd}-round orthotropic registration (blue curve) performs better than the 1^{st}-round isotropic registration (black). Meanwhile, it is only slightly worse than the results derived using ground truth elasticity parameters. This means further iterations won't improve the accuracy too much.

The four curves give registration accuracy derived from four registration options.

(1) The cyan curve is the registration accuracy using ground-truth orthotropic parameters. However, this is unachievable in situations where we can not get access to perfect elasticity parameters.

(2) The red curve is the registration accuracy using orthotropic parameters estimated from ground-truth (simulated) deformations. Since deformations are also unknown quantities in the joint estimation setting, we can never be able to recover more accurate elasticity parameters with the proposed elasticity estimation framework. In other words, this curve gives the accuracy upper-bound for the joint estimation framework.

(3) The black curve is associated with the traditional isotropic registration. The isotropic elasticity parameters were set to be $\bar{C}$.

(4) The 2^{nd}-round orthotropic registration (blue curve) performs better than the 1^{st}-round isotropic registration (black). Meanwhile, it is only slightly worse than the red curve, which means that further elasticity update won't improve the results too much.

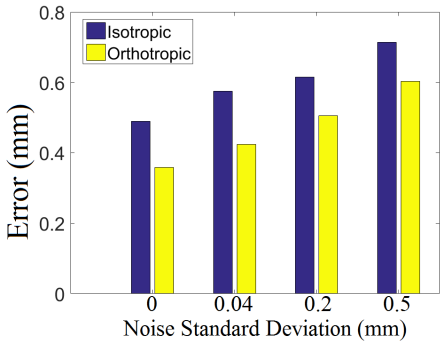


Figure 6.10: The 2^{nd}-round orthotropic registration performs better than the isotropic registration under different levels of noise.

I also tested the robustness of this joint estimation framework under the effect of noise. Different levels of white Gaussian noise were added to all vertices. Fig. 6.10 shows that the 2^{nd}-round orthotropic registration performs better than the isotropic registration in all 4 cases.

Real Endoscopic Data. I further used the elasticity framework to investigate the pharyngeal deformations contained in live nasopharyngoscopy. An endoscopic video provides direct visualization of a patient's pharyngeal surface and usually captures its rich swallowing motion. Elasticity estimation on this frame-by-frame surface deformation can help us better understand tissue characteristics and facilitate further analysis, such as the registration between the endoscopy and CT of the same patient for radiation treatment planning.

With the pipeline introduced in Chapter 5, I first reconstructed the endoscopogram surface from the video as the reference surface $\mathcal{M}$. Using groupwise TSD, the endoscopogram $\mathcal{M}$ was fused from N individual single-frame reconstructions $\{\mathcal{M}^\alpha | \alpha = 1...N\}$ produced by Shape-from-Motion-and-Shading. Groupwise TSD also yielded the set of deformations $\mathbf{\Phi} = \{\Phi^\alpha | \alpha = 1...N\}$ from the endoscopogram $\mathcal{M}$ to each single-frame reconstruction. (Actually, groupwise TSD produces a deformation field from each individual surface to the fused surface, but here I simply took its inverse deformation field as Φ^α.) Finally, I applied the elasticity estimation framework to estimate

orthotropic parameters from Φ and $\mathcal{M}$. The algorithm parameters used in this experiment were kept the same as before.

I tested on two patients' endoscopic video data. For each video sequence I sampled 20 individual frames focusing on the laryngeal region to produce the endoscopogram. Fig. 6.11 shows the results are consistent with throat anatomy: the epiglottis and the arytenoid cartilage be stiffer than the laryngeal region, the larynx being more elastic along the patient axial direction.

There are two potential uses of the estimated orthotropic parameters.

(1) The orthotropic parameters in the endoscopogram can be mapped back to each individual single-frame reconstruction to perform a 2^{nd}-round groupwise TSD with orthotropic energy function. This is exactly the joint estimation framework discussed in 6.4, with the pairwise TSD replaced by groupwise TSD. However, the mapping of orthotropic parameters between surfaces is beyond the scope of this dissertation.

(2) The orthotropic parameters provide an improved regularization term for the registration to the CT surface.

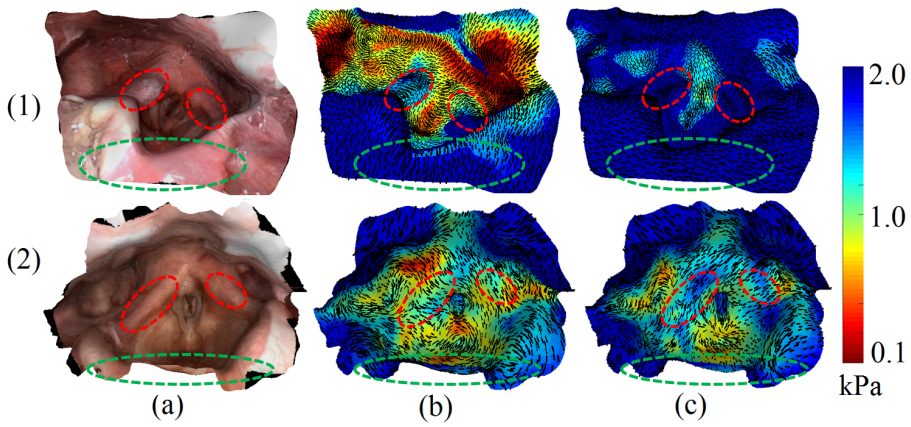


Figure 6.11: (a) 3D endoscopogram surfaces reconstructed from video. Red circles indicate the arytenoid cartilage. Green circles indicate the epiglottis. (b)(c) Estimated Young's moduli and the associated natural axes.

6.6 Discussion

I have introduced a statistical framework to estimate inhomogeneous and anisotropic elasticity parameters of a thin shell structure from a set of its known deformations; An MAP analysis on a novel MRF-based probability distribution can automatically recover both the orthotropic stiffness matrix and natural axes directions of every location on the shell. I have shown that this framework can be further used as a part in a joint registration and elasticity estimation framework. Both the elasticity estimation framework and the joint estimation framework can be helpful in studying within-patient deformations of anatomical surfaces. Despite the promising results shown in the experiments, future work still has to address the following concerns:

1. In many situations anatomical surfaces are deformed by the underlying muscles, so it is not appropriate to simply model the surface as a shell structure. The framework should be generalized to the 3D volume situation.

2. I assume the observed deformations are the realization of tissue elasticity of a same patient. The framework is not applicable for cross-patient deformations, or any other non-physical deformations.

3. The physical energy minimization part in Eq. 7 tends to produce stiffness that is less than the prior. However, this may not be an issue in registration because only relative elasticity is needed for spatially-varying registration.

4. In some applications, it is possible to also assign priors to natural axes directions. For example, bone structures usually have different stiffness along the longitudinal and transverse directions.

5. Umbilic points (isotropic regions) are not explicitly handled in this work.

6. Model parameters' selection needs to be further studied.

7. In several real endoscopic cases, real tissue deformations were dominated by other non-physical deformations (e.g., reconstruction errors). The applicability of the proposed framework in this situation needs to be further examined.

CHAPTER 7

CT/Endoscopy Surface Registration Revisited

The goal of CT/endoscopy fusion is to transfer the tumor-related information from the endoscopic video space to the 3D CT image space. As introduced in Chapter 1, the final step of such data fusion is to register the endoscopic reconstruction surface with the CT segmentation surface.

Chapters 3 & 4 have respectively introduced two potential methods for this surface registration task. In those two chapters, I have shown reasonable registration results with some early endoscopy reconstructions derived from (Ahmed and Farag, 2006) and (Schönberger and Frahm, 2016). We concluded that the quality of those reconstructions were poor, and I introduced in Chapter 5 our improved reconstruction surface, named the endoscopogram. To work with our new reconstruction data, we should revisit the CT/endoscopy registration problem while taking special properties of endoscopograms into account.

Challenges. There are two major challenges we have to face in registering CT surfaces with endoscopograms:

(1) The CT segmentation surfaces exists in the 3D space defined by the imaging system, and their coordinates have an absolute scale measured in millimeters. The endocopogram, however, is derived from Structure-from-Motion, which uses a reference camera coordinate system and only recovers geometry up to a relative scale. Therefore, assuming that the endoscopogram is perfectly reconstructed up to scale, there is a similarity transformation (rotation, translation, scaling) between the CT space and the endoscopogram space. Such a transformation is only related to the two imaging spaces and has nothing to do with patients' tissue deformations. Furthermore, as will be shown later, the depth-scale error in in the endoscopogram requires two more scaling parameters

in addition to the basic similarity transformation. Therefore, such parameters are essential for an accurate computation of the later non-rigid tissue deformations.

(2) The throat conformation in the CT image might differ significantly from that in the endoscopic video. We observe in many cases that the throat is in a much more open position in CT than in endoscopy. This is primarily because during the endoscopy time the presence of the endoscope causes drastic swallowing motion of the patient. Another source of throat conformation change comes from the patient's posture change between the CT scan time and the endoscopy time. During a CT scan, the patient is supine on a curved table, whereas during endoscopy the patient is sitting straight up. Different gravity effects may cause significant shape change in the entire pharyngeal region, thereby yielding large deformations between the endoscopogram surface and the CT segmentation surface. Therefore, we should take gravity into consideration in the registration process.

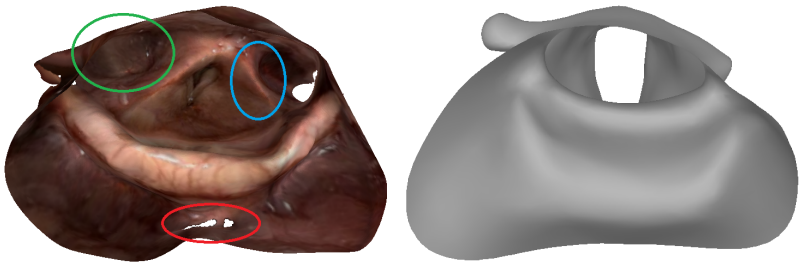
(3) I have briefly mentioned the missing patch and topology change situation in previous chapters. Now let's comprehensively summarize different sources of disparity between CT surfaces and endoscopograms.

i Due to camera occlusion, some regions are not observed in the endoscopic video, leaving holes in the endoscopogram. Therefore, there are surface patches in the CT surface that do not have counterparts in the endoscopogram (Fig. 7.1 red).

ii The low-resolution CT image can not resolve fine anatomical structures due to the partial volume effect, whereas in the high-resolution endoscopic frames, those structures can be identified and reconstructed. In this situation, there are structures in the endoscopogram that do not have counterparts in the CT surface.

iii The aforementioned patient posture change at the CT scan time can lead to large deformations (usually large expansion of the pharynx), which can cause the tissues to collapse together. Combined with the partial volume effect, such collapsed regions can not be easily segmented

from the image. Depending on different situations, there will be missing patches either in the CT surface or the endoscopogram surface.



(a) (b)

Figure 7.1: Different kinds of disparity between an endoscopogram and a CT surface. Red: missing patches in the endoscopogram due to occlusion. Blue: missing parts in the CT surface due to the partial volume effect. Green: missing parts in the CT surface due to tissue collapsing.

To conclude, we call all the above situations the *disparity* between the two surfaces. Both surfaces have *incompatible regions*, which are surface patches that do not have corresponding parts in the other surface. We should take such disparity into consideration in the registration process.

Contribution of This Chapter. In this chapter, I propose a quick engineering solution to handle the first two challenges. For the disparity situation, Chapters 3 & 4 have discussed methods that are moderately robust to missing patches (mostly represented as holes and truncations) without explicitly estimating them. However, an explicit quantification of the disparity in the registration process can provide the following two advantages:

(1) Having the disparity estimation, we can forbid feature matching in the incompatible regions, thereby rendering the registration more robust. In addition to holes, this can also handle more complex disparity situations, such as the tissue collapsing situation and the partial volume artifact, both of which can not be sufficiently handled by the previous methods.

(2) In transferring the tumor location from the endoscopogram to the CT surface, we can relate the disparity estimation with the certainty of the tumor transfer; that is, we are more certain about the transfer in the compatible regions because these regions can be explicitly registered. We can still predict deformations for the incompatible regions, but we are less certain about the tumor transfer there.

With the above motivations in mind, I propose a joint estimation framework that can simultaneously estimate registration (deformations between two surfaces) and disparity (incompatible regions). The rest of the chapter will first briefly introduce several engineering solutions to the CT/endoscopogram initial alignment problem and then elaborate on the novel joint estimation framework.

7.1 Initial Alignment

This section proposes an engineering solution to initially align the endoscopogram into the CT space. I particularly choose this endoscopogram-to-CT alignment direction because the later transferred tumor will be in the CT space and thus will be directly usable for radiation treatment planning. This initial alignment is designed to account for (a) the similarity transformation between the CT imaging surface and the endoscopy camera space; (b) the general shape change of the throat caused by the gravity.

A standard similarity transformation contains a 3D translation vector, a 3D rotation and a scaling factor. However, I observe that a single scaling factor will not suffice in our situation. Therefore, I propose to fit three factors for 3 coordinate axes respectively. The resulting initial alignment therefore contains a translation, a rotation and three scalings. The reasons will be explained in the following steps:

(1) **Setting the z-axis.** The coordinate system for the CT surface is well defined; the z-axis direction corresponds to the patient's longitudinal direction (along the throat). We assume the longitudinal direction (z-axis) for the endoscopogram can be manually designated. After setting the z-axes of the two surfaces, the remaining transformation parameters are a 3D translation vector, three scaling factors and a rotation about the z-axis.

(2) **Scaling and shifting of the z-axis.** (Wang et al., 2017) have shown that the depth estimation of the endoscopogram has a flawed scale, so I propose to fit a scale factor separately for the z-axis to approximately correct for the depth scale in the endoscopogram. We assume the z-coordinates of

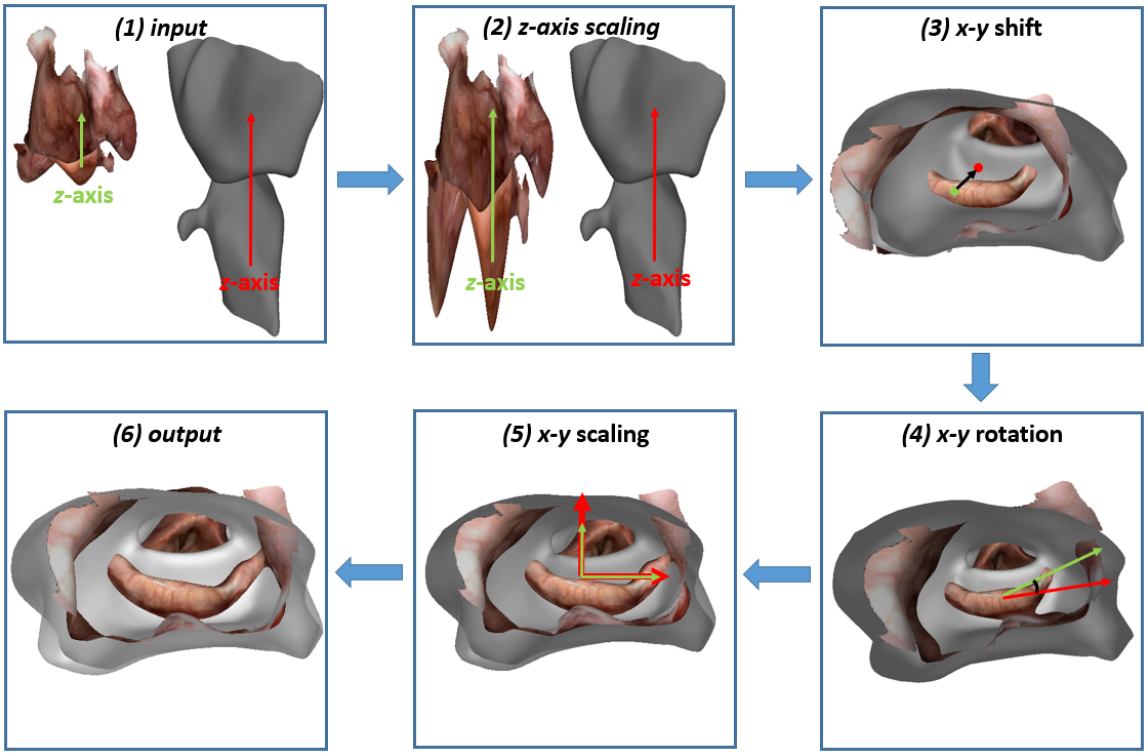


Figure 7.2: The initial alignment pipeline.

two anatomical landmarks, namely the tip of epiglottis and the anterior commissure, are available for both surfaces. Then the z-coordinates of the endoscopogram can be linearly scaled and shifted such that the z coordinates of the two landmarks can match with those of the CT surface.

(3) **Shifting in the x-y plane.** The endoscopogram surface is shifted on the x-y plane, such that the mean (x, y) coordinates match with those of the CT surface.

(4) **Rotation about the z axis.** A 2×2 second-moment matrix for each surface is derived from the (x, y) coordinates of all vertices from that surface. An eigendecompostion on the second-moment matrix yields two orthogonal eigenvectors, indicating the two major elliptical directions on the x-y plane. A rotation is then applied to the endoscopogram to match its two eigendirections to the ones of the CT surface.

(5) **Scaling of the x-y axes.** We assume the general shape change in the throat conformation induced by the gravity can be approximated by an elliptical deformation on the x-y plane. Since the major elliptical directions are already aligned in the previous step, we only need to scale the

endoscopogram along the x-y axes respectively to match the corresponding eigenvalues; that is, the two scaling factors on the $x-y$ plane are

$$\{\sqrt{\frac{u_{ct}^{x}}{u_{endo}^{x}}}, \sqrt{\frac{u_{ct}^{y}}{u_{endo}^{y}}}\} \tag{7.1}$$

where the u variables are the eigenvalues; for example, u_{endo}^{x} is the x-axis eigenvalue derived from the second moment matrix of the endoscopogram.

Finally, Fig. 7.2 shows all the initial alignment steps. The output is a transformed endoscopogram, which will be non-rigidly registered to the CT surface.

7.2 Surface Registration with Disparity Estimation

As mentioned in the beginning of this chapter, disparity between the two surfaces will lead to *incompatible regions*, surface patches that do not have correspondences in the other surface. This section proposes a novel joint estimation framework that can simultaneously estimate registration and incompatible regions.

To quantify the incompatible regions, we define two indicator functions $\mathcal{I}_1, \mathcal{I}_2$ for the two surfaces S_1, S_2. An indicator function is a binary random field defined on vertices. The function value (0 or 1) indicates whether a vertex has a correspondence in the other surface; that is, $\mathcal{I}_1(x) = 0$ means $S_1(x)$ does not have a correspondence in S_2. To make the entire formulation symmetric, I define correspondingly two deformation fields Φ_1, Φ_2 for the two surfaces. With the indicator functions being latent variables, the goal of registration is to determine the two deformation fields Φ_1 and Φ_2 that can align the *compatible regions* between S_1 and S_2. In other words, we only want to register surface patches where the indicator function equals 1. In doing this, we also want to determine the latent indicator functions, namely the incompatible regions.

7.2.1 MAP via EM

In this chapter, I use $\alpha = \{1, 2\}$ to denote the combination of a quantity for both surfaces, i.e., $S_\alpha = \{S_1, S_2\}$, etc.

Similar to the joint estimation framework discussed in Section 6.4, I propose to use the EM algorithm to find the mode of the posterior on the deformation fields. This subsection describes the high-level recipe for the EM procedure. Details of different terms in the following formulation will be further explained in the next subsection. First, the posterior we want to optimize is

$$p(\Phi_\alpha | S_\alpha) = \sum_{\mathcal{I}_\alpha} p(\Phi_\alpha, \mathcal{I}_\alpha | S_\alpha). \tag{7.2}$$

This posterior models the probability of having Φ_α as the registration results given the surfaces. It can be marginalized over the latent indicator functions. Since in our application an indicator function is a finite-dimensional binary field, the integral associated with the marginalization becomes a discrete summation. To find the mode of the above posterior, the EM algorithm first constructs a Q-function, which is the expected value of the log posterior with respect to the conditional distribution on $\mathcal{I}_\alpha$ under the current estimate of the deformations Φ_α:

$$Q(\Phi_\alpha | \Phi_\alpha^{(k)}) = \mathrm{E}_{\mathcal{I}_\alpha | S_\alpha, \Phi_\alpha^{(k)}} [\log p(\Phi_\alpha, \mathcal{I}_\alpha | S_\alpha)]. \tag{7.3}$$

When we assume the deformations Φ_α and $\mathcal{I}_\alpha$ are independent, the Q-function can be rewritten as

$$Q(\Phi_\alpha | \Phi_\alpha^{(k)}) = \mathrm{E}_{\mathcal{I}_\alpha | S_\alpha, \Phi_\alpha^{(k)}} [\log p(S_\alpha | \Phi_\alpha, \mathcal{I}_\alpha) + \log p(\Phi_\alpha) + \log p(\mathcal{I}_\alpha)] - \log p(S_\alpha). \tag{7.4}$$

Then we iterate between the following two steps:

E-step. The E-step computes the conditional distribution on $\mathcal{I}_\alpha$, given the current estimate of the deformations Φ_α:

$$p(\mathcal{I}_\alpha | S_\alpha, \Phi_\alpha^{(k)}) = \frac{1}{Z_0} p(S_\alpha | \Phi_\alpha^{(k)}, \mathcal{I}_\alpha) p(\mathcal{I}_\alpha), \tag{7.5}$$

where Z_0 is the partition function that is used to normalize the distribution. (In the following text, I will always use Z_0 to denote the partition function.)

M-step. Using the above conditional distribution, the M-step maximizes the Q-function (Eq. 7.4) to update the deformations:

$$\begin{aligned}
\Phi_\alpha^{(k+1)} &= \arg\max_{\Phi_\alpha} Q(\Phi_\alpha|\Phi_\alpha^{(k)}) && (7.6)\\
&= \arg\max_{\Phi_\alpha} \mathrm{E}_{\mathcal{I}_\alpha|S_\alpha,\Phi_\alpha^{(k)}}[\log p(S_\alpha|\Phi_\alpha,\mathcal{I}_\alpha) + \log p(\Phi_\alpha) + \log p(\mathcal{I}_\alpha)] && (7.7)\\
&= \arg\max_{\Phi_\alpha} \{\mathrm{E}_{\mathcal{I}_\alpha|S_\alpha,\Phi_\alpha^{(k)}}[\log p(S_\alpha|\Phi_\alpha,\mathcal{I}_\alpha)] + \log p(\Phi_\alpha)\}. && (7.8)
\end{aligned}$$

7.2.2 Likelihoods and Priors

The above EM formulation involves the computation of several probability distributions.

i. Likelihood $p(S_\alpha|\Phi_\alpha,\mathcal{I}_\alpha)$: This likelihood term is required in both the E-step (Eq. 7.5) and the M-step (Eq. 7.8). In traditional image registration (Zitová and Flusser, 2003), the likelihood term models how well an image deformation aligns a moving image with a target image. Similarly, our likelihood term also models data matching. In particular, it models how well the deformations align the compatible regions between the two surfaces. In other words, given the deformations, the two deformed surfaces $S_1' = S_1 \circ \Phi_1$ and $S_2' = S_2 \circ \Phi_2$ should be close in the compatible regions.

Therefore, for a point $x \in S_1'$, where $S_1'(x)$ gives its 3D coordinates, we define $\mathcal{C}_1(x)$ as its closest point in S_2'. Then the squared distance $\|S_1'(x) - \mathcal{C}_1(x)\|^2$ gives the quality of alignment for that point. We define $\mathcal{C}_2$ vice versa. Finally, the likelihood is constructed as a Boltzmann distribution:

$$p(S_\alpha|\Phi_\alpha,\mathcal{I}_\alpha) = \frac{1}{Z_0}\exp(-\gamma L(S_\alpha,\Phi_\alpha,\mathcal{I}_\alpha)), \tag{7.9}$$

$$L(S_\alpha,\Phi_\alpha,\mathcal{I}_\alpha) = \sum_{x\in S_1}(\mathcal{I}_1(x)\cdot\|S_1'(x)-\mathcal{C}_1(x)\|^2) + \sum_{x\in S_2}(\mathcal{I}_2(x)\cdot\|S_2'(x)-\mathcal{C}_2(x)\|^2). \tag{7.10}$$

In other words, we only penalize misalignment in the compatible regions.

In principle, the definition of this likelihood term should be identical for both E-step and M-step. In the E-step (Eq. 7.5), the deformations Φ_α are taken as fixed values, and the indicator functions

$\mathcal{I}_\alpha$ are the unknowns, so the closest point operations only need to be computed once. In the M-step, however, the deformations Φ_α are unknown variables to be optimized. This yields a challenge that the likelihood term of Eq. 7.10, particularly the closest point operators $\mathcal{C}_\alpha$, is not differentiable with respect to Φ_α, thereby preventing the use of any gradient-descent-based optimization method.

To this end, I observe that the likelihood term defined in Eq. 7.10 is closely related to the likelihood term (Eq. 4.10 and Eq. 5.4) in the previous TSD formulation; instead of using the closest point for measuring the quality of alignment, e.g., $\|S_1'(x) - \mathcal{C}_1(x)\|^2$, Eq. 5.4 uses the geometric feature matching result, namely $\|\Phi_1(x) - f_1(x)\|^2$. Recall that in the original groupwise TSD formulation $f_1(x)$ denotes the attraction force vector exerted on $x \in S_1$ derived by feature matching. The computation of f can be predetermined based on the given surfaces S_α in their original configurations and thus is independent of Φ_α. In the current situation, given the indicator functions $\mathcal{I}_\alpha$, the feature matching should be performed only between the compatible regions, and the associated resulting force vector should be denoted as $f_1(x; \mathcal{I}_\alpha)$. Therefore, the likelihood based on geometry feature matching is

$$\hat{L}(S_\alpha, \Phi_\alpha, \mathcal{I}_\alpha) = \sum_{x \in S_1} (\mathcal{I}_1(x) \cdot \|\Phi_1(x) - f_1(x; \mathcal{I}_\alpha)\|^2) + \sum_{x \in S_2} (\mathcal{I}_2(x) \cdot \|\Phi_2(x) - f_2(x; \mathcal{I}_\alpha)\|^2). \quad (7.11)$$

The advantage of the above likelihood over the likelihood in Eq. 7.10 is that it can be easily differentiated with respect to Φ_α: given a set of indicator functions $\mathcal{I}_\alpha$, the above likelihood becomes a quadratic function only on Φ_α. In other words, as long as $\mathcal{I}_\alpha$ are fixed, the attraction forces f_α will be determined.

By comparing the above two likelihoods, my insight is that both likelihood definitions essentially model the deviation between a vertex and its correspondence. In the E-step, with the fixed deformations Φ_α estimated from the previous M-step, the two surfaces are already overlaid with each other. Therefore, closest points more accurately determine correspondences. On the other hand, the M-step is essentially re-performing a registration to compute the surface deformations Φ_α, so correspondences derived from geometric feature matching are more reliable. Therefore, I

propose to use Eq. 7.10 as the likelihood in the E-step and to use the Eq. 7.11 as the likelihood in the M-step (Eq. 7.8).

ii. Prior on Indicator Functions $p(\mathcal{I}_\alpha)$

In computing the conditional distribution $p(\mathcal{I}_\alpha|S_\alpha, \Phi_\alpha^{(k)})$, $p(\mathcal{I}_\alpha)$ models our prior knowledge of the indicator functions. Similar works (Chitphakdithai and Duncan, 2010; Pohl et al., 2006) have been assuming that the prior distribution of indicator functions is spatially independent and factorizable:

$$p(\mathcal{I}) = \prod_{x \in S} p(\mathcal{I}(x)), \tag{7.12}$$

where $p(\mathcal{I}(x))$ defines a Bernoulli distribution on a single vertex x. The Bernoulli distribution is the probability distribution of a binary random variable, which takes the value 1 with probability q and the value 0 with probability $1 - q$, i.e.,

$$p(\mathcal{I}(x)) = \begin{cases} q & \text{if } \mathcal{I}(x) = 0, \\ 1 - q & \text{if } \mathcal{I}(x) = 1. \end{cases} \tag{7.13}$$

The spatial independence assumption suggests the indicator function value (0 or 1) on a vertex is not related to the function values of neighboring locations. The advantage of such an assumption is that the expectation operation in the objective function (Eq. 7.8) becomes tractable. However, for the same reason, the estimated indicator function is usually noisy, i.e., has frequent transitions between 0s and 1s between neighboring locations. Therefore, it is desirable to have spatially smooth estimations. With this motivation in mind, I formulate the prior as a Boltzman distribution in the following form:

$$p(\mathcal{I}) = \frac{1}{Z_0}\exp(-\beta H(\mathcal{I}) - \lambda B(\mathcal{I})), \tag{7.14}$$

$$H(\mathcal{I}) = \sum_x (1 - p(\mathcal{I}(x))), \tag{7.15}$$

$$B(\mathcal{I}) = \sum_{x,y} (1 - \delta(\mathcal{I}(x), \mathcal{I}(y))). \tag{7.16}$$

In the above equations, the $H(\mathcal{I})$ function encourages the indicator function value of each individual vertex to follow its own Bernoulli distribution. One has the flexibility in defining $p(\mathcal{I}(x))$ according to the application. For example, a reasonable Bernoulli prior for the endoscopogram would encourage the function value of high curvature places to be zero because those places are likely to be absent in the CT surface due to the partial volume effect. As a first trial, I simply applied the same Bernoulli distribution to all the vertices.

In the $B(\mathcal{I})$ function (Eq. 7.16), the Kronecker-Delta function ($\delta(i, j)$) encourages the function values to be consistent between neighboring vertices. This $B(\mathcal{I})$ function can help producing smooth indicator functions, but it also adds complexity to the overall conditional distribution (Eq. 7.5) and makes the expectation operation in the M-step (Eq. 7.8) not tractable anymore. In the next subsection, I will discuss two approximations to handle such intractability.

ii. Prior on Deformations $p(\Phi_\alpha)$: The prior of deformation fields is again formulated as the thin shell energy. This term is independent of $\mathcal{I}_\alpha$, which means the deformation regularization is applied to the entire surface domain. Even though we care mostly about the registration in the compatible regions, this prior term can still predict the most probable deformation for the remaining surface parts. Therefore, tumor transfer can still be achieved even if the tumor is located in the incompatible regions.

7.2.3 Two Approximations in EM

Consider the conditional distribution

$$p(\mathcal{I}_\alpha|S_\alpha, \Phi_\alpha^{(k)}) = \frac{1}{Z_0}\exp(-\gamma L(S_\alpha, \Phi_\alpha, \mathcal{I}_\alpha) - \beta H(\mathcal{I}) - \lambda B(\mathcal{I})). \tag{7.17}$$

Due to the spatial smoothness term B, this distribution cannot be factorized onto individual vertices, so the expectation operation in the objective function of Eq. 7.8 is not tractable. Now, I will introduce two approximations to handle such intractability.

Mode Approximation (Hard EM)

Adopting the same mode approximation trick discussed in Section 6.4, I use the mode of $p(\mathcal{I}_\alpha|S_\alpha, \Phi_\alpha^{(k)})$ to represent the whole probability distribution (see details of the mode approximation idea in Section 6.4). This yields the following registration/disparity joint estimation framework:

1. *Input*: The two indicator functions $\mathcal{I}_\alpha^{(0)}$ are set to 1 everywhere on the two surfaces.

2. *M-step*: In iteration k, with the current estimate of $\mathcal{I}_\alpha^{(k)}$, the M-step becomes

$$\begin{aligned}
\Phi_\alpha^{(k+1)} &= \arg\max_{\Phi_\alpha}\{\mathrm{E}_{\mathcal{I}_\alpha|S_\alpha,\Phi_\alpha^{(k)}}[\log p(S_\alpha|\Phi_\alpha, \mathcal{I}_\alpha)] + \log p(\Phi_\alpha)\} && (7.18)\\
&\approx \arg\max_{\Phi_\alpha}[\log p(S_\alpha|\Phi_\alpha, \mathcal{I}_\alpha^{(k)}) + \log p(\Phi_\alpha)] && (7.19)\\
&\approx \arg\max_{\Phi_\alpha}[-\gamma\hat{L}(S_\alpha, \Phi_\alpha, \mathcal{I}_\alpha^{(k)}) + \log p(\Phi_\alpha)] && (7.20)
\end{aligned}$$

The above objective function is equivalent to a groupwise TSD registration between the two surfaces. As explained in the last subsection, the negative log likelihood $\hat{L}$ is essentially a data matching term derived from attraction forces between the compatible regions. In other words, we only want to register the compatible regions indicated by $\mathcal{I}_\alpha^{(k)}$.

3. *E-step*: With the new update of deformations $\Phi_\alpha^{(k+1)}$, find the mode of $p(\mathcal{I}_\alpha|S_\alpha, \Phi_\alpha^{(k+1)})$:

$$\mathcal{I}_\alpha^{(k+1)} = \arg\max_{\mathcal{I}_\alpha} \log p(\mathcal{I}_\alpha|S_\alpha, \Phi_\alpha^{(k+1)}) \tag{7.21}$$

$$\approx \arg\max_{\mathcal{I}_\alpha} [-\gamma L(S_\alpha, \Phi_\alpha^{(k+1)}, \mathcal{I}_\alpha) - \beta H(\mathcal{I}_\alpha) - \lambda B(\mathcal{I}_\alpha)]. \tag{7.22}$$

The above optimization can be solved by Iterated Conditional Modes (Besag, 1986).

4. Iterate to Step 2 until convergence.

In the end, besides the estimated deformations, the indicator functions $\mathcal{I}_\alpha^{(k)}$ in the last E-step are taken as the final estimation of the latent variables. The advantage of using a mode approximation is that it can achieve fast expectation computation by considering only the mode of a distribution. However, for the same reason, the mode cannot adequately characterize the entire probability distribution and cannot model the distribution uncertainty. These facts can lead to inferior expectation approximation. For example, (Allassonnière et al., 2007) have shown that the mode approximation scheme for diffeomorphic atlas building performs poorly under image noise. Therefore, I will discuss another expectation approximation method next.

Monte-Carlo Approximation (Monte-Carlo EM)

The Monte-Carlo approximation has been extensively used in the EM algorithm for intractable expectations. The idea is that in the E-step, we first sample a large number of samples, usually by a Markov-chain Monte-Carlo (MCMC) sampler, from the conditional distribution of Eq. 7.17. Then in the M-step, the expectation can be effectively approximated by the sample average. It can be proven that with sufficient samples the sample average converges to the true expectation. This variant of EM is known as Monte-Carlo EM. It is worth mentioning that the Monte-Carlo approximation is more preferable in this situation than in the previous registration/elasticity joint estimation framework because

(a) We only need to estimate a binary variable for each vertex, whereas we need to estimate 5 continuous variables per vertex in the previous elasticity/registration joint estimation framework. The sampling space is much smaller in this case.

(b) A Markov-chain Monte-Carlo (MCMC) sampler for continuous random fields requires an effective proposal distribution, which is theoretically challenging to obtain. Moreover, elasticity parameters live on the manifold of positive definite matrices, and MCMC samplers are generally not well developed for manifold data. However, MCMC for binary random fields can be effectively carried out with the "single-flip" trick that is commonly used in sampling the Ising model (Newman and Barkema, 1999) (flipping the state of a random single vertex at each time).

To apply Monte-Carlo EM in our situation, we get the following registration/disparity joint estimation framework:

1. *Input*: The two indicator functions $\mathcal{I}_\alpha^{(0)}$ are initially set to 1 everywhere on the two surfaces. The deformations $\Phi_\alpha^{(0)}$ are initially computed by

$$\Phi_\alpha^{(0)} = \arg\max_{\Phi_\alpha} \{-\gamma \hat{L}(S_\alpha, \Phi_\alpha, \mathcal{I}_\alpha^{(0)}) + \log p(\Phi_\alpha)\}. \tag{7.23}$$

   The above initialization step essentially computes a groupwise TSD registration between the two surfaces without considering disparity.

2. *E-step*: In iteration k, with the current estimate of deformations $\Phi_\alpha^{(k)}$, we sample a large number of samples from $p(\mathcal{I}_\alpha | S_\alpha, \Phi_\alpha^{(k)})$. In the end, we get $\{\mathcal{I}_\alpha^1, ..., \mathcal{I}_\alpha^N\}$, where N is the number of samples. Precisely speaking, each sample $\mathcal{I}_\alpha^i$ contains two indicator functions defined for the two surfaces. The Metropolis-Hastings algorithm (Newman and Barkema, 1999) is a popular method for sampling from a distribution in the form of Eq. 7.17.

3. *M-step*: With the N samples, the M-step becomes

$$\begin{aligned}
\Phi_\alpha^{(k+1)} &= \arg\max_{\Phi_\alpha} \{\mathrm{E}_{\mathcal{I}_\alpha | S_\alpha, \Phi_\alpha^{(k)}} [\log p(S_\alpha | \Phi_\alpha, \mathcal{I}_\alpha)] + \log p(\Phi_\alpha)\} && (7.24) \\
&\approx \arg\max_{\Phi_\alpha} \{\frac{1}{N} \sum_{i=1}^{N} [\log p(S_\alpha | \Phi_\alpha, \mathcal{I}_\alpha^i)] + \log p(\Phi_\alpha)\} && (7.25) \\
&\approx \arg\max_{\Phi_\alpha} [\frac{-\gamma}{N} \sum_{i=1}^{N} \hat{L}(S_\alpha, \Phi_\alpha, \mathcal{I}_\alpha^i) + \log p(\Phi_\alpha)] && (7.26)
\end{aligned}$$

The first term in Eq. 7.26 is the average likelihood over all sampled indicator functions.

4. Iterate to Step 2 until convergence.

Unlike the previous mode-approximation, in the last E-step, we do not use any single-point estimation of $p(\mathcal{I}_\alpha|S_\alpha, \Phi_\alpha^{(k)})$. Instead, the sample mean $\frac{1}{N}\sum_{i=1}^{N}\mathcal{I}_\alpha^i$ is used as the expectation of the indicator functions.

7.3 Tumor Transfer

The ultimate purpose of the CT/endoscopogram registration is tumor transfer. To be specific, a tumor is first circled on the endoscopogram surface. Then its location is carried over to the CT space using the deformation fields computed by the proposed registration method. Now we assume that S_1 is the endoscopogram and S_2 is the CT surface, and that we have computed the deformations Φ_1 and Φ_2 between the two surfaces. In other words, the two deformed surfaces $S_1' = S_1 \circ \Phi_1$ and $S_2' = S_2 \circ \Phi_2$ are well aligned in the compatible regions. In order to transfer the location of a point $x \in S_1$ from S_1 to S_2, we essentially need to compute the deformation field $\Phi_1 \circ \Phi_2^{-1}$. In other words, the location $S_1(x)$ is first carried over by a forward deformation Φ_1 onto $S_1'(x)$. Since S_1' and S_2' are already registered, the location of $S_1'(x)$ is further carried over by the inverse deformation Φ_2^{-1} onto S_2.

One problem in doing this is that the inverse deformation field Φ_2^{-1} is defined on the discrete vertex domain of S_2', thereby not directly applicable to S_1'. To that end, I simply use a kernel regression (Bierens, 1994) to interpolate the deformation field Φ_2^{-1} into the domain of S_1'.

A special case is that the circled tumor is located in the incompatible regions of the endoscopogram. Even though the deformation field is computed for the entire endoscopogram domain, we are less certain about the deformation in the incompatible regions because those regions are not explicitly registered to any part of the CT surface, and the deformation estimation in those regions is only based on the thin shell regularization. For the clinical purpose, a potential solution is to associate the transferred location with a "certainty" or "tolerance" measurement based on the

indicator function values. This part of work is still under investigation and will be integrated into the software under development for the upcoming clinical test.

7.4 Results

In this section, I will first show the strength of the proposed joint estimation framework with synthetic tests and then demonstrate its use in real patient cases.

Model parameters of the conditional distribution.

We first need to determine the q parameter of the Bernoulli distribution (Eq. 7.13). It controls how likely a vertex is classified to be an incompatible vertex without using any knowledge of the surface registration quality.

Another three important model parameters include the three weighting factors γ, β, λ in the conditional distribution $p(\mathcal{I}_\alpha|S_\alpha, \Phi_\alpha^{(k)})$. Respectively they weight the likelihood, the per-node Bernoulli prior and the smoothness prior. In both hard EM (mode-approximation) and Monte-Carlo EM, the relative scale among the three parameters is critical because it reflects our beliefs about the indicator functions; as examples, should the prior dominate the likelihood, or how smooth do we want our binary field to be.

Besides the relative scale of the three parameters, their overall magnitudes also play an important role. A simple rearrangement of Eq. 7.17 yields

$$p(\mathcal{I}_\alpha|S_\alpha, \Phi_\alpha^{(k)}) = \frac{1}{Z_0}\exp(-s(\frac{\gamma}{s}L + \frac{\beta}{s}H + \frac{\lambda}{s}B)). \tag{7.27}$$

We can see that sum of the three parameters, $s = \gamma + \beta + \lambda$, determines the concentration of the probability density function. In other words, s reflects how much uncertainty is built into the above conditional distribution.

All the above parameter-selection criteria are conceptually reasonable but quantitatively vague, so it is complicated to find the optimal model parameters. The computational complexity of Monte-Carlo EM also renders it impossible to exhaustively traverse the model parameter space.

In this research, I have not had time to further explore the parameter-selection part. However, an observation is that under various parameter settings all the following quantitative results show a similar relative pattern. Therefore, using the more time-efficient mode-approximation approach, I tuned the parameters to achieve the best registration accuracy on a separate synthetic surface pair. In the following experiments, I set the parameters to be $q = 0.9, \gamma = 1, \beta = 5, \lambda = 2$.

7.4.1 Synthetic Tests

This subsection presents synthetic experimental results. Four surface pairs were respectively created from four patients' CT data. For each CT segmentation surface, a synthetic deformation was manually applied based on the same principles discussed in previous chapters. Then incompatible regions were manually placed in both the original surface and the deformed surface. The simulated disparity included the following:

(a) Removal of high curvature regions to simulate the partial volume effect;

(b) Bridging anatomical structures to simulate tissue collapse;

(c) Creating truncations and holes to simulate camera occlusion.

Since both the deformation and the incompatible regions were manually constructed, the ground truth for both entities was available. Here I tested three registration options: (a) registration without disparity estimation (groupwise TSD), (b) joint estimation via mode approximation (Hard EM) and (c) joint estimation via Monte-Carlo approximation (Monte-Carlo EM). To remind the readers, a straight groupwise TSD (option (a)) directly estimates the two deformation fields. The two joint estimation approaches (option (b) and (c)) iterate between groupwise TSD and the modeling of the conditional distribution of Eq. 7.17 via mode-approximation or Monte-Carlo approximation. I will show in the following that the two joint estimation approaches are superior.

Registration Accuracy

The following table gives the registration accuracy of the three registration options. For the two joint estimation approaches, I performed 10 iterations of the EM algorithm. The registration results

Table 7.1: Surface registration error (mm)

Registration Options	Patient 1	Patient 2	Patient 3	Patient 4
Initial Error	3.64±1.42	3.4±1.8	5.0±1.5	4.8±1.44
Groupwise TSD	0.67±0.42	0.93±0.39	1.52±0.85	1.38±0.98
Hard EM	0.59±0.4	0.75±0.45	0.85±0.46	1.05±0.64
Monte-Carlo EM	**0.54±0.36**	**0.63±0.4**	**0.71±0.43**	**0.95±0.51**

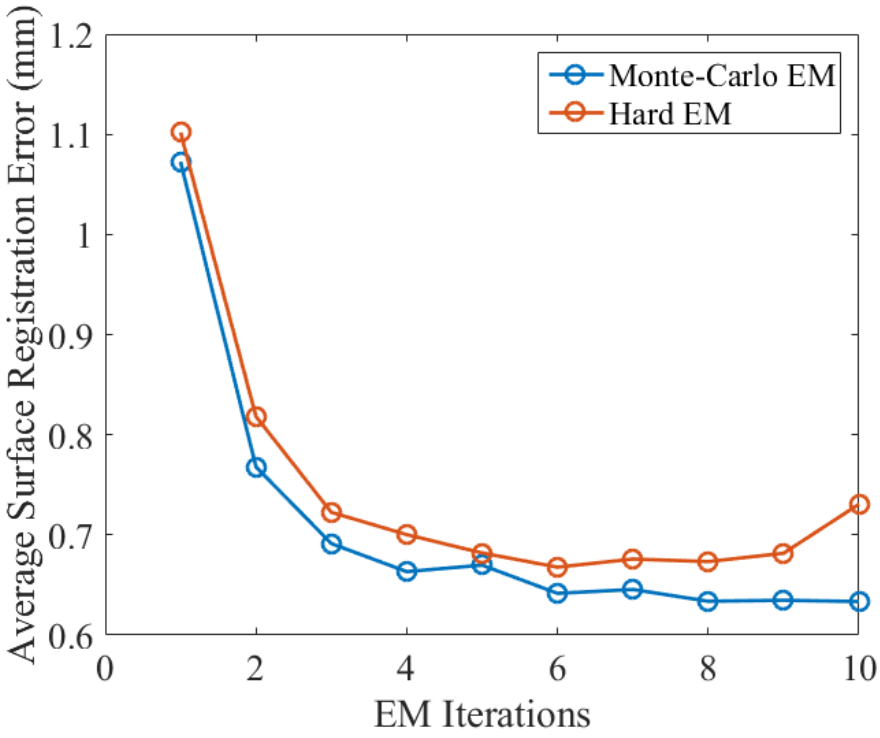


Figure 7.3: Convergence curves of the average surface registration error.

of the last iteration are given in the following table (the last two rows). The error was measured as the *surface registration error*.

We can see that registration/disparity joint estimation achieved much more accurate registration results than groupwise TSD alone. Meanwhile, for the two joint estimation approaches, Monte-Carlo EM performs slightly better than hard EM. As explained before, this accuracy improvement comes from the fact that the Monte-Carlo approach models the uncertainty of the conditional distribution of Eq. 7.17 and better approximates the expectation by performing a sample average operation instead of considering just the mode.

For the two EM variants, I further plotted the surface registration error averaged over the four patients. The blue curve in Fig. 7.3 shows that the error measurement of the Monte-Carlo EM approach converges nicely through EM iterations. However, the slight rise of the orange curve after iteration 6 indicates that the hard EM approach might not converge to the ground truth.

Disparity Estimation Accuracy

For the two joint estimation approaches, I have also investigated their performance in estimating the incompatible regions. For the hard EM approach, the mode of the conditional distribution Eq. 7.17 in the last EM iteration is computed as the final estimation of the latent variable $\mathcal{I}_\alpha$. For the Monte-Carlo EM approach, the sample mean of Eq. 7.17 in the last EM iteration is first computed and then the final binary field is derived by thresholding the sample mean at 0.5. The error is measured as the total number of wrong estimations on the binary variables. Assuming the $\mathcal{I}_\alpha^*$ give the ground truth binary fields,

$$\text{error} = \sum_x |\mathcal{I}_1(x) - \mathcal{I}_1^*(x)| + \sum_x |\mathcal{I}_2(x) - \mathcal{I}_2^*(x)|. \tag{7.28}$$

Table 7.2: Disparity estimation accuracy of the two joint estimation approaches

	Patient 1	Patient 2	Patient 3	Patient 4
Number of vertices	2088	3119	4721	5012
Number of incompatible vertices	823	580	1199	1236
Hard EM	143	125	243	239
Monte-Carlo EM	**111**	**113**	**206**	**195**

We can clearly from Table 7.2 see that Monte-Carlo EM is better at modeling the latent variables by the sampling procedure. Fig. 7.4 further shows that the disparity estimation error averaged over the 4 patients reduces through EM iterations.

Fig. 7.5 shows a set of qualitative results for a synthetic surface pair. The final estimations of the two binary fields derived from hard EM and Monte-Carlo EM are respectively shown in Fig. 7.5ii and Fig. 7.5iii. In particular, we can see from *surface 2* that the Monte-Carlo-based estimation is smoother and more accurate (reflected in the circled regions). Furthermore, the sampling procedure

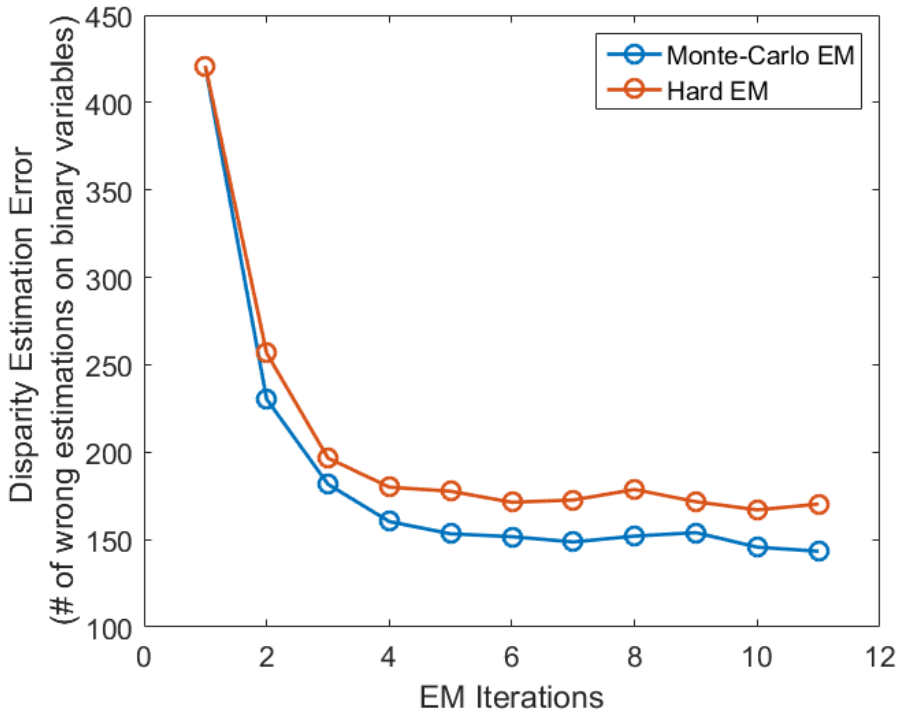
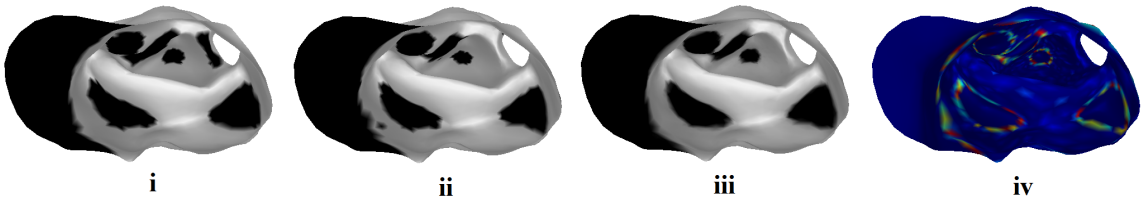
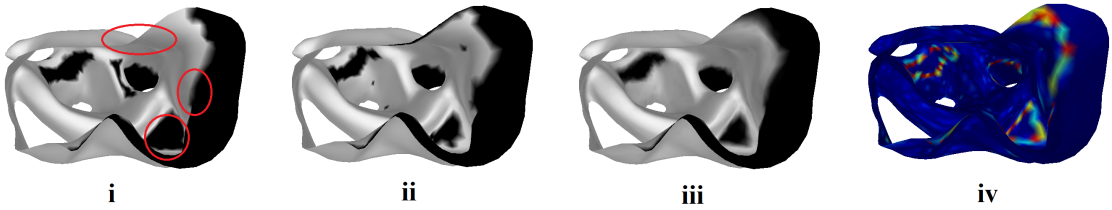


Figure 7.4: Convergence curves of the average disparity estimation error.



(a) Surface 1:



(b) Surface 2:

Figure 7.5: The upper row and the bottom row respectively show two surfaces to be registered in one synthetic case. From left to right: ground truth indicator functions, mode-derived indicator functions, Monte-Carlo-derived indicator functions, uncertainty maps.

of the latent binary fields in Monte-Carlo EM yields a byproduct, which is the *sample covariance*.

The *sample covariance* of a binary random field forms a covariance matrix. The diagonal entries of

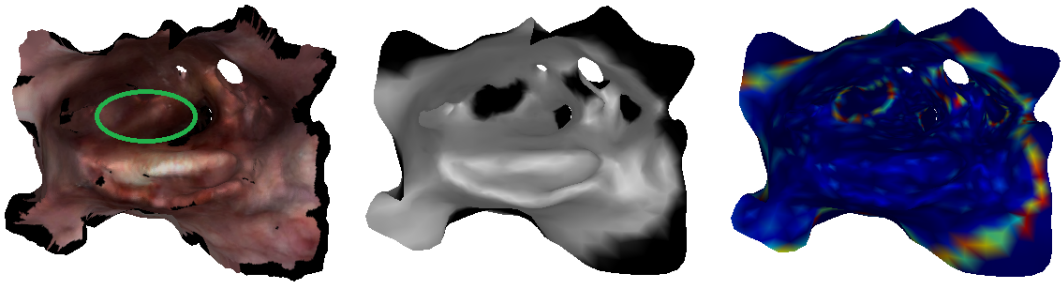
that matrix are the variances (uncertainty) of the indicator function values on each vertex. When plotted on a surface, this gives the so-called uncertainty map. Fig. 7.5iv shows a pair of uncertainty maps associated with the two surfaces. We can see that the largest uncertainty exists along the boundary between the 0-regions and 1-regions. In other words, we are not certain about the places where the transitions between 0 and 1 happen.

One significant drawback of Monte-Carlo EM is that computing the sample average of the likelihood is extremely slow because it involves N times of geometric feature matching for the N samples. It took around 4 hours on a i7-3770k@3.5GHz CPU CPU to run 10 EM iterations with each surface having around 2000 vertices, whereas the hard EM approach only took around 5 minutes. Therefore, depending on the application, one needs to determine whether to sacrifice computational efficiency or registration accuracy in choosing between hard EM and Monte-Carlo EM. In our case, radiation treatment planning does not require real-time computation, so in the following real patient experiments, Monte-Carlo EM was always used to produce registration results.

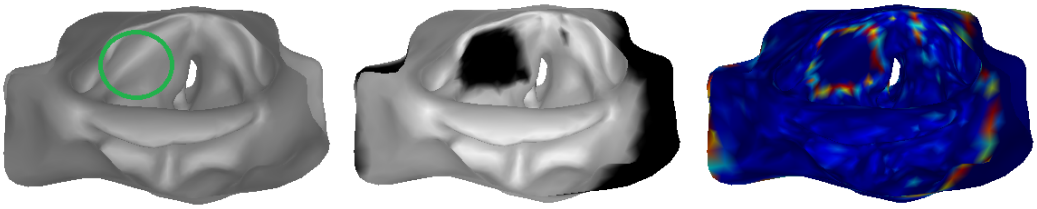
7.4.2 Real Patient Data

CT/endoscopogram registration was carried out for 4 real patients. For each patient, the endoscopogram was first aligned to the CT space using the method introduced in Sect 7.1. Then Monte-Carlo EM was performed to estimate the non-rigid deformations. The resulting deformations of the 4 cases are ready to be used for a upcoming clinical test. Results of the clinical test will be reported when available. Here, I show some qualitative results for one of four patients in Fig. 7.6 and Fig. 7.7. This case is special in the sense that the left arytenoid in the endoscopogram was not successfully reconstructed. This fact is reflected by the incompatible regions (circled in green in Fig. 7.7) estimated in both the CT surface and the endoscopogram.

Figure 7.6: left: an endoscopogram. Middle: a stripe pattern is plotted on the endoscopogram surface. Right: The stripe pattern is transferred to the CT surface using the resulting deformations $\Phi_1 \circ \Phi_2^{-1}$. The cyan regions correspond to incompatible regions in the CT surface that do not have counterparts in the endoscopogram.



(a) Endoscopogram (the missing left arytenoid is circled in green)



(b) CT segmentation

Figure 7.7: (a) Upper row: an endoscopogram of a patient. (b) Bottom row: a CT segmentation surface of the same patient. Left: Original endoscopogram and CT surfaces. Middle: the estimated indicator functions. Right: the two uncertainty maps. Conclusion: in this case, the major incompatible regions were estimated to be on the arytenoids, especially the left one. This is because the arytenoids in the endoscopogram were badly reconstructed.

7.5 Discussion

In this chapter, I have particularly dealt with the registration between a CT surface and an endoscopogram for the purpose of tumor transfer. First, an initial alignment strategy has been proposed to incorporate a translation, a rotation and three scaling factors to align the endoscopogram to the CT space. Then different sources of disparity between the two surfaces have been summarized.

To handle disparity, I have extended groupwise TSD into a joint estimation framework, where both registration and incompatible regions can be estimated. Based on hard EM and Monte-Carlo EM, I have developed two different variants of the joint estimation approach. I have shown with synthetic data that the two joint estimation approaches performs better than groupwise TSD alone. In particular, Monte-Carlo EM can produce superior accuracy in both registration and disparity estimation. However, the major drawback of Monte-Carlo EM is that it usually takes hours of computation.

Remaining Problems

(1) Usually in an EM framework, the prior on the latent variables can significantly affect the final estimation. In this dissertation, I have simply applied a non-informative prior (a uniform Bernoulli distribution) on the binary indicator functions. One can imagine curvature-related priors can be assigned to the endoscopogram indicator function to model the partial volume effect. Moreover, I have observed that certain anatomical regions (the tip of epiglottis and the pyriform sinus) are more likely to collapse under deformations. Such observations can also be formulated as a prior on the incompatible regions.

(2) One theoretical problem is that the formulation of the likelihood is different in the E-step and M-step. Future work should investigate how much this will affect the estimation.

(3) The inefficiency of Monte-Carlo EM comes from repetitive feature matchings for all sampled indicator functions. The power of parallel computing needs to be explored to the maximum extent. There are two major computational steps in Monte-Carlo EM that can be paralleled: (a) The feature matching procedure itself involves extensive but independent pairwise vertex matchings. (b) Feature matchings for different sampled indicator functions are entirely independent and parallelizable.

(4) Currently the effect of gravity is modeled by the proposed ad-hoc solution, namely the different scaling factors in the x-y plane. A more complex anatomical model of the head-and-neck region should be explored to simulate the deformations caused by gravity.

(5) Model parameter selection needs to be further investigated. Some existing methods (Zhang et al., 2013; Awate and Radhakrishnan, 2015) have formulated the model parameters also as latent variables and estimated them automatically. Such methods can be considered in our case.

(6) From the 4 synthetic cases, we can observe slight improvements from Monte-Carlo EM. However, it is not clear whether such improvements are statistically significant. Further statistical hypothesis tests should be carried out.

CHAPTER 8

Conclusion and Discussion

This chapter reviews the contributions of this dissertation in Section 8.1. This is followed by a discussion of several important issues related to this dissertation in Section 8.2. Section 8.3 lists some remaining theoretical and technical questions that need to be solved in the future.

8.1 Summary of Contributions

This section revisits the thesis and claims laid out in Chapter 1 and presented through Chapters 3-7. Each contribution is restated along with a discussion of how it was applied in the application of CT/Endoscopy fusion.

1. *A novel geometric feature descriptor has been designed to capture rich curvature information around a local patch.*

   Anatomical structures in the throat have distinguishable shapes. A reliable registration of pharyngeal surfaces should be able to recognize similarity between such geometric structures. A geometric feature descriptor has been proposed in Chapter 3 to collect multi-scale curvatures in a local region. Along with some other informative geometric entities, such as principal and normal directions, the proposed descriptor can successfully generate distinct geometric signatures at different throat regions. A simple synthetic example has shown that a direct pairwise feature matching strategy can already reasonably estimate a small set of matching points between surfaces. With these matchings, the following two proposed registration

methods, namely the joint spectral graph matching method and Thin Shell Demons, have achieved successful registration results.

2. *An improved spectral-graph-matching method has been proposed for matching partial surfaces with complete surfaces.*

   Based on a recently proposed joint spectral graph matching method, an improvement has been developed in Section 3.3 by incorporating the above geometric feature matching points. The matchings were added as inter-surface links of a joint affinity graph. The improved method has been tested with the registration between a CT segmentation and an endoscopic reconstruction. Experiments have been performed on 24 synthetic surface pairs and 3 real patients' endoscopic reconstructions and CT surfaces. Results have shown that geometric-feature-based links outperform the traditional Euclidean-distance-based links.

   Spectral graph matching for surfaces with different intrinsic geometry was an untouched area. My results have first indicated that the improved joint spectral matching is robust for partial surfaces with holes and truncations but that the method still fails under severe topology changes, such as bridging.

3. *A novel physics-motivated registration method, named Thin Shell Demons, has been proposed. The method incorporates both geometric feature information and a thin shell physical model to produce realistic deformations between surfaces.*

   There is usually a significant intrinsic geometry difference between CT and endoscopic surfaces, caused by the partial volume effect, camera occlusion and tissue collapsing. This fact suggests that spectral methods are inherently not applicable. A deformation-based registration method should be adopted instead. Chapter 4 introduced a novel surface registration method, Thin Shell Demons (TSD), that directly seeks a deformation between two surfaces. Energy based on the thin shell model has been adopted as a deformation regularization term for producing physically plausible surface deformations. The above feature matching results were formulated as virtual forces that attract geometrically similar regions. I have further

proposed to add novel structural links in the traditional thin shell model to preserve non-local shape structures. When applied to the same 24 synthetic surface pairs, results have suggested that TSD can produce superior results compared to the spectral graph matching method.

TSD alone has been primarily used for registering a patient's endoscopogram to the CT surface of that patient. To that end, results have shown that TSD can work with reconstruction surfaces derived from various algorithms. Moreover, I have also shown that the proposed structural links can help preserve the thickness of the epiglottis during the registration. As will also be mentioned later, an extension of the pairwise TSD has been used in registering many single-frame reconstructions for endoscopogram construction.

4. *A novel groupwise surface registration algorithm has been proposed to register multiple partially overlapping surfaces. The method has been applied for registering cross-frame reconstructions in the task of endoscopogram reconstruction.*

   A pipeline for reconstructing a surface model from an endoscopic video sequence has been developed in Chapter 5. Particularly, a novel groupwise registration method has been proposed as an extension of the pairwise TSD. Groupwise TSD can simultaneously register a group of partially overlapping surfaces without having or estimating the mean/template surface. This is achieved by formulating an N-body attraction scenario in which each surface's deformation is determined by the virtual attraction forces from all other surfaces. Experiments on synthetic partial surfaces from 6 patients' CT data have achieved an average surface registration error less than 1mm. Groupwise TSD has been successfully applied to 8 real endoscopic video sequences for fusing the geometry of multiple single-frame reconstructions into a single surface.

5. *An elasticity estimation framework has been developed to automatically learn orthotropic elasticity from a group of known deformations. I have further shown that a joint estimation approach can simultaneously learn orthotropic elasticity parameters and estimate deformations in a groupwise registration setting.*

A deformation-based registration approach relies on having a reasonable regularization term. In this dissertation, a thin shell model is adopted as regularization for producing physically realistic deformations. I have argued that a common homogeneous and isotropic thin shell cannot sufficiently model tissue mechanics. Chapter 6 has introduced a novel statistical elasticity estimation framework that can automatically estimate inhomogeneous and anisotropic elasticity parameters from a group of given deformations. In particular, an orthotropic elasticity model has been adopted to model different tissue stiffness along two orthogonal natural axes directions. I have shown how both the elasticity (stiffness) parameters and the natural axes directions can be estimated using an MAP analysis. A toy example and synthetic tests based on 5 patients' CT surfaces have shown its effectiveness in recovering elasticity parameters up to a given scale. Tests on two patients' endoscopic data focusing on the laryngeal region have shown that the elasticity estimation results are consistent with throat anatomy.

With the above elasticity estimation, a joint registration/elasticity-estimation framework has been further developed to iteratively estimate deformations and elasticity parameters. Such a framework can be applied to learn tissue elastic properties from frame-to-frame tissue deformations contained in the endoscopic video. The learned elasticity has been shown in turn to improve the registration between the endoscopogram and the CT surface.

6. *A joint registration/disparity estimation approach has been proposed to deal with the CT/endoscopogram registration.*

   Chapter 7 summarized three major causes for the significant difference between the CT and endoscopogram surfaces. I have argued that such disparity needs to be explicitly modeled during the TSD registration process. A hard EM approach and a Monte-Carlo EM algorithm have been introduced to jointly estimate registration and disparity. Experiments on 4 synthetic cases have shown that the two proposed approaches can reasonably recover disparity regions. Meanwhile, the registration accuracy is also improved with the consideration of disparity.

7. *Several engineering steps have been accomplished in completing the pipeline for tumor transfer between endoscopy and CT. These engineering contributions include the following, all detailed in Chapter 7:*

   i. An initial alignment method has been designed to compute the translation, rotation and three scaling factors between an endoscopogram and a CT surface. This step allows further a non-rigid registration between the two surfaces.

   ii. A texture fusion algorithm has been designed to put texture on an endoscopogram using many 2D frames' texture. The algorithm uses an MRF to pull color from a nearby vertex of a registered single-frame reconstruction and also to encourage color consistency between neighboring vertices.

   iii. An overall pipeline for building an endoscopogram has been designed to effectively integrate reconstruction, geometry fusion and texture fusion.

8. *A clinical study in the throat region has been initiated based on the proposed technical solutions.*

   The above scientific and engineering contributions have enabled tumor transfer from endoscopy to CT. As a result, a clinical study is about to start in order to validate whether endoscopy-derived tumor locational information can indeed help making effective radiation treatment plans. Therefore, besides the contributions made in the image computing field, this dissertation plays a critical role in making potential improvements in the current practice of pharyngeal radiation treatment planning.

   The successful use of my methods in the pharynx case has led to a preliminary study in the colon, where 3D reconstructions are derived from colonoscopic videos for detecting uninspected regions and colonoscope retargeting.

On the basis of the above contributions and their successful use in CT/endoscopy fusion, I have established the following thesis:

Thesis: *Geometric information and physical modeling can improve surface registration results in the context of CT/endoscopy fusion. The anisotropic parameters in the physical model can be inferred probabilistically from a set of given material deformations and can in turn improve the registration.*

8.2 Discussion and Future Work

This section further discusses several important topics involved in this dissertation and some potential future research directions.

1. *Surface registration vs. volume registration.*

   In the task of CT/endoscopy fusion, the CT image is a 3D volume image, whereas the endoscopic video only contains information on the tissue surface. To match the two modalities, I chose to discard the 3D information in the CT image and to only use its segmentation surface. This leads to a surface-to-surface registration problem. However, it is also conceivable to construct a 3D volume image, e.g., a signed distance image, from the endoscopogram surface and perform a volume-to-volume image registration. In fact, several works (Iglesias et al., 2013; Dedner et al., 2007) have been following this line of thinking. In particular, some narrow-band formulations have been adopted to register 3D level-set images associated with two surfaces. This dissertation lacks any serious consideration of the volume-based approach. Nevertheless, it is conceivably challenging as well to handle missing data in a volume-based registration; for example, how to create a 3D signed-distance image from a partial endoscopogram surface in order to be matched to the CT volume image.

2. *Applicability of spectral methods on different intrinsic geometry.*

   To the best of my knowledge, spectral methods for different intrinsic geometry had been an uncharted territory. This dissertation first shows that some simple situations (holes and truncations) can be handled by the improved spectral graph matching method. To further

handle more complex situations, it is conceivable to incorporate the disparity estimation and to modify the surfaces, such as breaking bridges or patching holes, to make them have similar intrinsic geometry. However, the motivation for further exploration in this direction is arguable, because the essence and advantage of spectral methods lie in the nature of isometry. For handling different intrinsic geometry, one should seek a deformation-based method instead.

3. *Physical reality of the thin shell model.*

   The question has been mentioned several times in the dissertation: to what extent can the proposed physical model reflect real tissue mechanics? An accurate physical modeling of human tissues is still an open research problem. In my work, the thin shell model can only approximate the mechanics of surface-like structures, and it is obviously a flawed model for pharyngeal muscles. To this end, my opinion is that the merit of the thin shell model should be judged by the specific application; in our case, the question to ask is whether the thin shell model can produce registration that is sufficiently good for radiation treatment planning (errors within 3mm). To that end, we are about to start a clinical research, in which we will validate whether the tumor transfer results derived from my registration algorithm will help tumor localization and thus help radiation treatment planning. The outcome of that clinical research will further indicate the usability of the thin shell model.

4. *The need for an anatomical model.*

   As mentioned before, an alternative to the thin shell model is the 3D finite element model, but the challenge lies in the fact that 3D FEM requires a comprehensive anatomical model for pharyngeal reconstruction and registration, namely the 3D anatomical composition in the head-and-neck region and the structural relationship between anatomies. This dissertation has not been devoted in constructing that model. However, further exploration in that direction is likely to improve our understanding of the pharyngeal shape change caused by gravity, the swallowing motion, pharyngeal pathology, etc.

A 4D CT recorded during a swallowing motion is another data source that can potentially be used to study the statistical deformation space and the elasticity parameters in the pharynx. However, the clinical motivation of using 4D CTs is not clear because current practice in radiation treatment planning only uses conventional CT.

5. *Applicability of probabilistic elasticity estimation.*

   The use of probabilistic elasticity estimation is still premature in various applications. The major problem is the ambiguity caused by the lack of force measurements. Therefore, in a probabilistic Bayesian formulation, elasticity estimation is not likely to succeed unless at least one of the following conditions is satisfied:

   (a) When the data is limited, e.g., with only a reference object and a deformed object available, elasticity estimation can easily overfit to that particular deformation. In that case, a comprehensive prior knowledge of anatomical elasticity should be used to avoid overfitting and to handle noise. In statistics, this kind of prior is known as an informative prior or a subjective prior, a prior that is not dominated by the likelihood and that has an impact on the posterior distribution.

   (b) In contrast, a prior distribution is noninformative or objective if the prior is ”flat” relative to the likelihood function and has minimal impact on the posterior distribution. As suggested by its name, many statisticians favor noninformative priors because they appear to be more objective. In my proposed elasticity framework, the uniform isotropic elasticity prior serves as a noninformative prior because it only determines the estimation scale, and the estimated elasticity patterns are mostly related to the likelihood. Therefore, to work with a noninformative prior, it is favorable to have a large number of observed tissue deformations to yield a meaningful likelihood; for example, the frame-by-frame endoscopic reconstructions provide rich deformations that can accurately reveal tissue elasticity.

Some other works (Yang et al., 2016; Lee et al., 2012) can also estimate elasticity from limited data by reducing the number of unknown variables. However, these methods are less interesting for our spatially-varying registration purpose.

6. *Use and extension of probabilistic elasticity estimation.*

   In this dissertation, the proposed elasticity estimation framework is mainly used as a regularization term in registration. However, it is also conceivable to be used as a prior for image segmentation. Other preliminary experiments have been performed with the intention of using the estimated elasticity for cancer grading or tumor detection. In a broader context, pharyngeal elasticity estimation is of interest to physiological and biomedical researchers; the estimated elasticity can provide insights in studying swallowing motion.

   One limitation of the current elasticity estimation framework is that it can only be applied to single-patient deformations, but the framework can be potentially generalized to a cross-patient situation by either performing elasticity parameter transfer or deformation transfer between two patients' surfaces. Such an extension is coupled with serious theoretical challenges. For example, different patients' surfaces may have different topology; also, surface deformations are not Euclidean quantities and cannot be easily transported.

7. *Probabilistic elasticity estimation vs. shape-appearance statistical models.*

   The proposed thin shell elasticity estimation provides a new perspective to traditional shape analysis. Instead of fitting a probability distribution on shape appearances (Loncaric, 1998; Nitzken et al., 2014), the proposed method seeks the underlying physical reasons for shape deformations. Therefore, it is conceivably better at "extrapolating" deformations. For example, in our application, endoscopy-derived elasticity parameters can be also used to produce realistic endoscopogram-to-CT deformations, whereas in shape-appearance statistical models the probability distribution specifically fitted to endoscopic throat conformations cannot be used for other kinds of deformations.

However, shape-appearance statistical models are more generalizable because they can be also applied to study cross-patient shape variance or general shape variance not induced by elastic deformations. However, as mentioned in *item 5*, even though probabilistic elasticity estimation has the potential to be extended into the cross-patient case, it is still limited to elastic deformations.

8. *Atlas-based approaches*

   Atlas-based or template-based approaches have been widely used in image segmentation (Iglesias and Sabuncu, 2014) and registration (Sotiras et al., 2013). The idea is that information of interest, such as anatomical segmentations, can be transferred or adapted to the new data item from one or multiple given templates (single-atlas or multi-atlas). In our case, it is conceivable to apply an atlas-based approach for endoscopogram reconstruction. Assuming anatomical conformations are similar across patients, the reconstruction will be based on patient-specific endoscopic video as well as anatomical structural information transferred from an endoscopogram atlas or multiple atlases. Alternatively, we can adopt a CT-model-based reconstruction approach, in which a patient's endoscopic reconstruction is guided by the CT surface of that patient. Of course, in doing this, one has to consider the resolution difference and large anatomical deformations between CT and endoscopy, but it is likely that the CT-model-based approach can yield more accurate global geometry than the current approach. One might also consider constructing an elasticity atlas (or multiple atlases) for probabilistic elasticity estimation; that is, instead of using a noninformative prior, the elasticity will be derived by considering both patient-specific deformations and atlas-based priors.

9. *Clinical extension*

   The methods developed in this dissertation have been primarily used in the context of nasopharyngoscopy. However, the methods are generalizable enough to be applied to the reconstruction tasks of other endoscopies, such as gastrointestinal endoscopy, laparoscopy, etc. Although there are many other reconstruction algorithms published for handling those

tasks, my novel registration methodology suggests that it is specifically preferable in the situations where there are notable tissue deformations across frames.

In particular, the work in this dissertation has stimulated a study in colonoscopy reconstruction. The goal of colonoscopy reconstruction includes efficient visualization and review, identifying uninspected colon patches, camera retargeting, polyp detection, etc. Preliminary results have shown that similar techniques (Wang et al., 2017) based on groupwise TSD can reconstruct short segments of the colon. An additional need in the colon case is that the reconstruction has to be achieved in real time. Since there is no speed requirement in the context of radiation treatment planning, I have not explored the parallel computing aspect to the maximum extent. Currently, pairwise TSD runs within 2 seconds on an Intel i7-3770k@3.5GHz CPU. Groupwise TSD among N surfaces is computationally equivalent to N independent and parallelizable pairwise TSD registrations, which means the running time will further scale up when the number of surfaces exceeds the number of CPU processors. In addition, there are certain algorithmic components, such as curvature computation for every vertex, that can be deployed to a many-core GPU. Therefore, a future direction is to design a mixed CPU/GPU parallelization framework to effectively handle large number of surfaces.

8.3 Remaining Technical Issues

Besides the high-level future work mentioned in Section 8.2 ,this section revisits some remaining theoretical and technical problems that need to be solved in the future.

1. *Model parameter selection.*

   In this dissertation, we accomplished model parameter selection primarily by using a validation set containing several synthetic data cases (usually less than 3). This can lead to serious overfitting. Automatic model parameter selection should be studied as an alternative approach.

2. *The construction of structural links.*

   Chapter 4 has shown that structural links can help preserve shapes of non-local structures. However, these links have to be manually placed. It would be interesting to statistically learn the optimal locations for placing these links.

3. *The formulation of elasticity priors.*

   In the elasticity estimation framework, the isotropic stiffness matrix prior is conceptually regarded as the mean of all anisotropic stiffness matrices. However, the physical energy term in the objective function tends to produce stiffness that is always smaller than that isotropic prior. A potential solution is to constrain that prior to be the mean of all estimated anisotropic matrices during the optimization process.

4. *Singularities in orthotropic elasticity estimation.*

   The natural axes directions essentially form a frame field on the surface domain. Differential geometry theory tells us that in a close surface there must be singular places where the frame cannot be uniquely defined. In our case, these singularities correspond to isotropic locations. This dissertation has not explicitly modeled such singularities because our surfaces are always open, but we need to solve this hidden trouble in the future.

5. *Transferring orthotropic elasticity between surfaces.*

   The last subsection has explained the need for transferring elasticity parameters between surfaces. To the best of my knowledge, orthotropic elasticity transfer between two triangle meshes has not been studied. Assuming the two surfaces are well aligned, one needs to solve an interpolation problem with certain spatial smoothness constraints.

6. *Temporal consistency across single-frame reconstructions.*

   All the single-frame reconstructions can be ordered in time. Since temporally adjacent frames have little deformation in between, this temporal order can be used as an additional regularization term in both SfMS reconstruction and geometry fusion. Based on this motivation, (Wang

et al., 2017) has proposed a fusion-guided SfMS approach that incorporates cross-frame consistency, but the temporal order is still not explicitly modeled. Future work should place a temporal regularizer on both reconstruction depth values and lighting model parameters.

REFERENCES

Ahmed, A. H. and Farag, A. A. (2006). A new formulation for shape from shading for non-lambertian surfaces. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2*, pages 1817–1824.

Allassonnière, S., Amit, Y., and Trouvé, A. (2007). Towards a coherent statistical framework for dense deformable template estimation. *Journal of the Royal Statistical Society Series B*, 69(1):3–29.

Alterovitz, R., Goldberg, K., Pouliot, J., Hsu, I.-C. J., Kim, Y., Noworolski, S. M., and Kurhanewicz, J. (2006). Registration of MR prostate images with biomechanical modeling and nonlinear parameter estimation. *Med. Phys.*, 33(2):446–454.

Andersen, R. (2008). Modern methods for robust regression. *Sage University Paper Series on Quantitative Applications in the Social Sciences*, pages 07–152.

Arsigny, V., Fillard, P., Pennec, X., and Ayache, N. (2006). Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Phys Med Biol*, 56(2):411–421.

Arslan, S., Parisot, S., and Rueckert, D. (2015). Joint spectral decomposition for the parcellation of the human cerebral cortex using resting-state fmri. In Ourselin, S., Alexander, D. C., Westin, C.-F., and Cardoso, M. J., editors, *IPMI 2015*, volume 9123 of *LNCS*, pages 85–97. Springer International Publishing.

Audette, M. A., Ferrie, F. P., and Peters, T. M. (2000). An algorithmic overview of surface registration techniques for medical imaging. *Medical Image Analysis*, 4(3):201–217.

Awate, S. P. and Radhakrishnan, T. (2015). Colocalization estimation using graphical modeling and variational bayesian expectation maximization: Towards a parameter-free approach. In *Information Processing in Medical Imaging: 24th International Conference, IPMI 2015, Sabhal Mor Ostaig, Isle of Skye, UK, June 28 - July 3, 2015, Proceedings*, pages 3–16. Springer International Publishing.

Bajcsy, R. and Kovačič, S. (1989). Multiresolution elastic matching. *Computer Vision, Graphics, and Image Processing*, 46(1):1–21.

Balci, S. K., Golland, P., Shenton, M., and Wells, W. M. (2007). Free-form b-spline deformation model for groupwise registration. In *MICCAI 2007*, pages 23–30.

Bauer, M. and Bruveris, M. (2011). A new Riemannian setting for surface registration. In *3nd MICCAI Workshop on Mathematical Foundations of Computational Anatomy*, pages 182–194.

Bauer, M., Bruveris, M., and Michor, P. W. (2014). Overview of the geometries of shape spaces and diffeomorphism groups. *J Math Imaging Vis*, 50(1):60–97.

Bay, H., Tuytelaars, T., and Gool, L. V. (2006). Surf: Speeded up robust features. In *ECCV 2006*, pages 404–417.

Beg, M., Miller, M., Trouvé, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61(2):139–157.

Bergou, M., Wardetzky, M., Harmon, D., and Grinspun, D. Z. E. (2006). A quadratic bending model for inextensible surfaces. In *SGP '06 Proceedings of the fourth Eurographics symposium on Geometry processing*, pages 227–230.

Besag, J. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society*, 48(3):48–259.

Bierens, H. J. (1994). *"The NadarayaWatson kernel regression function estimator". Topics in Advanced Econometrics.* New York: Cambridge University Press.

Borg, I. and Groenen, P. (2005). *Modern Multidimensional Scaling: theory and applications*. New York: Springer-Verlag, second edition.

Bronstein, A. M., Bronstein, M. M., and Kimmel, R. (2006). Generalized multidimensional scaling: A framework for isometry-invariant partial surface matching. *Proceedings of the National Academy of Sciences*, 103(5):1168–1172.

Cates, J., Fletcher, P. T., Styner, M., Shenton, M., and Whitaker, R. (2007). Shape modeling and analysis with entropy-based particle systems. In *Information Processing in Medical Imaging 2007*, pages 333–345.

Chang, W. and Zwicker, M. (2008). Automatic registration for articulated shapes. *Computer Graphics Forum (Proceedings of SGP 2008)*, 27(5):1459–1468.

Chitphakdithai, N. and Duncan, J. S. (2010). Non-rigid registration with missing correspondences in preoperative and postresection brain images. In *MICCAI 2010*, volume 6361, pages 367–374.

Christensen, G., Rabbitt, R., and Miller, M. (1996). Deformable templates using large deformation kinematics. *Computer Vision, Graphics, and Image Processing*, 5(10):1435–1447.

Chung, F. (1997). Spectral graph theory. *AMS*.

Crane, K., Desbrun, M., and Schröder, P. (2010). Trivial connections on discrete surfaces. *Computer Graphics Forum (SGP)*, 29(5):1525–1533.

Curless, B. and Levoy, M. (1996). A volumetric method for building complex models from range images. In *SIGGRAPH 1996*, pages 303–312.

Davatzikos, C. (1997). Spatial transformation and registration of brain images using elastically deformable models. *Computer Methods in Applied Mechanics and Engineering*, 66(2):207–222.

Dedner, A., Lüthi, M., Albrecht, T., and Vetter, T. (2007). Curvature guided level set registration using adaptive finite elements. *Pattern Rcognition*, 4713:527–536.

Durrleman, S., Pennec, X., Trouvé, A., and Ayache, N. (2009). Statistical models of sets of curves and surfaces based on currents. *Med. Image Anal.*, 13(5):793–808.

Durrleman, S., Prastawa, M., Korenberg, J., Joshi, S., Trouv, A., and Gerig, G. (2012). Topology preserving atlas construction from shape data without correspondence using sparse parameters. In Ayache, N., Delingette, H., Golland, P., and Mori, K., editors, *MICCAI 2012*, volume 7512 of *LNCS*, pages 223–230. Springer Berlin Heidelberg.

Elad, A. and Kimmel, R. (2003). On bending invariant signatures for surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(10):1285–1295.

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

Freiman, M., Voss, S. D., and Warfield, S. K. (2011). Demons registration with local affine adaptive regularization: application to registration of abdominal structures. In *IEEE International Symposium on Biomedical Imaging 2011*, page 1219C1222.

Gatzke, T., Grimm, C., Garland, M., and Zelinka, S. (2005). Curvature maps for local shape comparison. In *Shape Modeling and Applications 2005*.

Geng, X., Christensen, G. E., Gu, H., Ross, T. J., , and Yang, Y. (2009). Implicit reference-based group-wise image registration and its application to structural and functional mri. *Neuroimage*, 47(4):1341–1351.

Gerig, T., Shahim, K., Reyes, M., Vetter, T., and Lüthi, M. (2014). Spatially varying registration using gaussian processes. In *MICCAI 2014*, pages 627–635.

Gingold, Y., Secord, A., Han, J. Y., Grinspun, E., and Zorin, D. (2004). A discrete model for inelastic deformation of thin shells. Technical report, Courant Institute of Mathematical Sciences.

Gray, A. (1997). *"A Monge Patch", Modern Differential Geometry of Curves and Surfaces with Mathematica*. CRC Press.

Green, M., Geng, G., Qin, E., Sinkus, R., Gandevia, S., and Bilston, L. (2013). Measuring anisotropic muscle stiffness properties using elastography. *NMR Biomed*, 26(11):1387–1394.

Green, P. (1984). Iteratively reweighted least squares for maximum likelihood estimation, and some robust and resistant alternatives. *Journal of the Royal Statistical Society*, 46(2):149–192.

Grinspun, E., Gingold, Y., Reisman, J., and Zorin, D. (2006). Computing discrete shape operators on general meshes. *Comput. Graph. Forum*, 25(3):547–556.

Gu, X., Wang, Y., Chan, T., Thompson, P., and S.T., Y. (2004). Genus zero surface conformal mapping and its application to brain surface mapping. *IEEE Trans Med Imaging.*, 23(8):949–58.

Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition.

Hong, D., Tavanapong, W., Wong, J., Oh, J., and de Groen, P. C. (2014). 3D reconstruction of virtual colon structures from colonoscopy images. *Computerized Medical Imaging and Graphics*, 38(1):22–23.

Huber, D. F. and Hebert, M. (2003). Fully automatic registration of multiple 3D data sets. *Image and Vision Computing*, 21(7):637–650.

Iglesias, J. A., Berkels, B., Rumpf, M., and Scherzer, O. (2013). A thin shell approach to the registration of implicit surfaces. In *Vision, Modeling, and Visualization 2013*, pages 89–96.

Iglesias, J. E. and Sabuncu, M. R. (2014). Multi-atlas segmentation of biomedical images: A survey. *CoRR*, abs/1412.3421.

Jaynes, E. (1957). Information theory and statistical mechanics. *Physical Review*, 106(4).

Johnson, A. E. and Hebert, M. (1999). Using spin images for efficient object recognition in cluttered 3d scenes. *IEEE Transactions on PAMI*, 21(5):433–449.

Joshi, S., Davis, B., Jomier, M., and Gerig, G. (2004). nbiased diffeomorphic atlas construction for computational anatomy. *NeuroImage*, 23(8):949–58.

Jung, S., Schwartzman, A., and Groisser, D. (2015). Scaling-rotation distance and interpolation of symmetric positive-definite matrices. *SIAM J. Matrix Analysis Applications*, 36(3):1180–1201.

Ki, S. M., McCuIIoch, T. M., Bae, H., and Kim, S. J. (2006). Biomechanical model for muscular dysfunction ofthe human pharynx using finite element analysis. *Annals of Otology, Rhinology and Laryngology*, 115(11):864–870.

Koenderink, J. (1990). *Solid Shape*. The MIT Press.

Kroon, M. and Holzapfel, G. A. (2008). Estimation of the distributions of anisotropic, elastic properties and wall stresses of saccular cerebral aneurysms by inverse analysis. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 464(2092):807–825.

Lam, K. C., Gu, X., and Lui, L. M. (2014). Genus-one surface registration via teichmüller extremal mapping. In *MICCAI 2014*, volume 8675, pages 25–32.

Lee, H.-P., Foskey, M., Niethammer, M., Krajcevski, P., and Lin, M. C. (2012). Simulation-based joint estimation of body deformation and elasticity parameters for medical image analysis. *IEEE Trans. Med. Imaging*, 31(11):2156–2168.

Li, H., Sumner, R. W., and Pauly, M. (2008). Global correspondence optimization for non-rigid registration of depth scans. *Computer Graphics Forum, SGP 08*, 27(5).

Li, Y. and Barbic, J. (2015). Stable anisotropic materials. *IEEE Trans. Vis. Comput. Graph*, 21(10):1129–37.

Lipman, Y. and Funkhouser, T. (2009). Möbius voting for surface correspondence. *ACM Transactions on Graphics*, 28(72).

Lombaert, H., Grady, L., Polimeni, J., and Cheriet, F. (2011). Fast brain matching with spectral correspondence. In Szekely, G. and Hahn, H. K., editors, *IPMI 2011*, volume 6801 of *LNCS*, pages 660–673. Springer Berlin / Heidelberg.

Lombaert, H., Sporring, J., and Siddiqi, K. (2013). Fast brain matching with spectral correspondence. In *IPMI 2013*, volume 7917 of *LNCS*, pages 376–389. Springer Berlin / Heidelberg.

Loncaric, S. (1998). A survey of shape analysis techniques. *Pattern Recognition*, 31(8):983 – 1001.

Lowe, D. (1999). Object recognition from local scale-invariant feature. In *ICCV 1999*, pages 1150–1157.

Lyttelton, O., Boucher, M., Robbins, S., and Evans, A. (2007). An unbiased iterative group registration template for cortical surface analysis. *Neuroimage*, 34:1535–1544.

Lyu, I., Kim, S. H., Seong, J.-K., Yoo, S. W., Evans, A., Shi, Y., Sanchez, M., Niethammer, M., and Styner, M. A. (2015). Robust estimation of group-wise cortical correspondence with an application to macaque and human neuroimaging studies. In *Front Neurosci. 2015*.

Maier-Hein, L., Mountney, P., Bartoli, A., Elhawary, H., Elson, D., Groch, A., Kolb, A., Rodrigues, M., Sorger, J., Speidel, S., and Stoyanov, D. (2013). Optical techniques for 3d surface reconstruction in computer-assisted laparoscopic surgery. *Medical Image Analysis*, 17(8):974–996.

Mateus, D., Horaud, R., Knossow, D., Cuzzolin, F., and Boyer, E. (2008). Articulated shape matching using laplacian eigenfunctions and unsupervised point registration. In *CVPR 2008*.

Meisner, E., Hager, G., Ishman, S., Brown, D., Tunkel, D., and Ishii, M. (2013). Anatomical reconstructions of pediatric airways from endoscopic images: A pilot study of the accuracy of quantitative endoscopy. *The Laryngoscope*, 123(11):2880–2887.

Miga, M. (2003). A new approach to elastography using mutual information and finite elements. *Physics in Medicine and Biology*, 48(4).

Misra, S., Ramesh, R., and Okamura, A. (2010). Modelling of non-linear elastic tissues for surgical simulation. *Computer Methods in Applied Mechanics and Engineering*, 13(6):811–818.

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press.

Newman, M. E. J. and Barkema, G. T. (1999). *Monte Carlo methods in statistical physics*. Clarendon Press.

Nitzken, M. J., Casanova, M. F., Gimelfarb, G., Inanc, T., Zurada, J. M., and El-Baz, A. (2014). Shape analysis of the human brain: A brief survey. *IEEE Journal of Biomedical and Health Informatics*, 18(4).

Ovsjanikov, M., Ben-Chen, M., Solomon, J., Butscher, A., and Guibas, L. (2012). Functional maps: A flexible representation of maps between shapes. *ACM Transactions on Graphics (TOG) - SIGGRAPH 2012 Conference Proceedings*, 31(4).

Pauly, M., Mitra, N., Giesen, J., Guibas, L., and Gross, M. (2005). Example-based 3D scan completion. In *Symposium on Geometry Processing 2005*, pages 23–32.

Pennec, X., Cachier, P., and Ayache, N. (1999). Understanding the "demon's algorithm": 3d non-rigid registration by gradient descent. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI'99: Second International Conference, Cambridge, UK, September 19-22, 1999. Proceedings*, pages 597–605.

Pennec, X., Fillard, P., and Ayache, N. (2006). A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66(1):41–66.

Pereira, G. C., Traughber, M., and Jr., R. F. M. (2014). The role of imaging in radiation therapy planning: Past, present, and future. *BioMed Research International*.

Pohl, K., Fisher, J., Grimson, W., Kikinis, R., and Wells, W. (2006). A bayesian model for joint segmentation and registration. *Neuroimage*, 31(1):228–239.

Pokrass, J., Bronstein, A. M., Bronstein, M. M., Sprechmann, P., and Sapiro, G. (2013). Sparse modeling of intrinsic correspondences. *Eurographics Computer Graphics Forum*, 32(2):459–468.

Price, T., Zhao, Q., Rosenman, J., Pizer, S., and Frahm, J.-M. (2016). Shape from motion and shading in uncontrolled environments. In *Under submission, To appear at http://midag.cs.unc.edu/*.

Ramachandran, S., Calcagno, C., Mani, V., Robson, P. M., and Fayad, Z. A. (2014). Registration of dynamic contrast-enhanced mri of the common carotid artery using a fixed-frame template-based squared-difference method. *J Magn Reson Imaging*, 39(4).

Reuter, M. (2010). Hierarchical shape segmentation and registration via topological features of laplace-beltrami eigenfunctions. *International Journal of Computer Vision*, 89(2):287–308.

Risholm, P., Ross, J., Washko, G., and Wells, W. (2011). Probabilistic elastography: estimating lung elasticity. In *IPMI 2011*, pages 699–710.

Schneider, R., Faust, G., Hindenlang, U., and Helwig, P. (2009). Inhomogeneous, orthotropic material model for the cortical structure of long bones modeled on the basis of clinical CT or density data. *Computer Methods in Applied Mechanics and Engineering*, 198(27-29):2167–2174.

Schönberger, J. L. and Frahm, J.-M. (2016). Structure-from-motion revisited. In *IEEE Conference on Computer Vision and Pattern Recognition 2016*.

Schulz, J., Pizer, S. M., Marron, J., and Godtliebsen, F. (2015). Nonlinear hypothesis testing of geometrical object properties of shapes applied to hippocampi. *Journal of Mathematical Imaging and Vision*, 54(1):15–34.

Sotiras, A., Davatzikos, C., and Paragios, N. (2013). Deformable medical image registration: A survey. *Medical Imaging, IEEE Transactions on*, 32(7):1153–1190.

Sun, J., Ovsjanikov, M., and Guibas, L. (2009). A concise and provably informative multi-scale signature based on heat diffusion. *Eurographics Symposium on Geometry Processing 2009*, 28(5):1383–1392.

Tam, G., Cheng, Z., Lai, Y., Langbein, F., Liu, Y., Marshall, D., Martin, R., Sun, X., and Rosin, P. (2013). Registration of 3d point clouds and meshes: a survey from rigid to nonrigid. *IEEE Trans Vis Comput Graph*, 19(7):1199–1217.

Thirion, J. (1998). Image matching as a diffusion process: an analogy with Maxwells demons. *Medical Image Analysis*, 2(3):243–260.

Tokgozoglu, H. N., Meisner, E. M., Kazhdan, M., and Hager, G. D. (2012). Color-based hybrid reconstruction for endoscopy. In *CVPR Workshops 2012*, pages 8–15. IEEE Computer Society.

Vaillant, M. and Glaunès, J. (2005). Surface matching via currents. In *IPMI 2005*, volume 3565, pages 381–392.

van de Giessen, M., Vos, F. M., Grimbergen, C. A., van Vliet, L. J., and Streekstra, G. J. (2012). An efficient and robust algorithm for parallel groupwise registration of bone surfaces. In *MICCAI 2012*, pages 164–171.

Ventsel, E. and Krauthammer, T. (2001). *Thin Plates and Shells: Theory: Analysis, and Applications*. CRC Press.

Vialard, F.-X. and Risser, L. (2014). Spatially-varying metric learning for diffeomorphic image registration: A variational framework. In *MICCAI 2014*, pages 227–234.

Wang, R., Price, T., Zhao, Q., Frahm, J.-M., Rosenman, J. G., and Pizer, S. (2017). Improving 3d surface reconstruction from endoscopic video via fusion and refined reflectance modeling. In *SPIE Medical Imaging*.

Wei, W. I. and Kwong, D. L. W. (2010). Current management strategy of nasopharyngeal carcinoma. *Clin. Exp. Otorhinolaryngol.*, 3(1):1–12.

Wu, C., Narasimhan, S. G., and Jaramaz, B. (2010). A multi-image shape-from-shading framework for near-lighting perspective endoscopes. *International Journal of Computer Vision*, 86(2):211–228.

Xu, C., Brennick, M. J., Dougherty, L., and Wootton, D. M. (2009). Modeling upper airway collapse by a finite element model with regional tissue properties. *Med. Eng. Phys.*, 31(10):1343–48.

Yang, S., Jojic, V., Lian, J., Chen, R., Zhu, H., and Lin, M. C. (2016). Classification of prostate cancer grades and t-stages based on tissue elasticity using medical image analysis. In *MICCAI 2016*, pages 627–635.

Yeo, B. T., Sabuncu, M. R., Vercauteren, T., Ayache, N., Fischl, B., and Golland, P. (2010). Spherical demons: fast diffeomorphic landmark-free surface registration. *IEEE Transactions on Medical Imaging*, 29(3):650–668.

Zaharescu, A., Boyer, E., Varanasi, K., and Horaud, R. (2009). Surface feature detection and description with applications to mesh matching. In *Computer Vision and Pattern Recognition 2009*, pages 373–380.

Zeng, W. and Gu, X. D. (2011). Registration for 3D surfaces with large deformations using quasi-conformal curvature flow. In *CVPR 2011*, pages 2457–2464.

Zeng, Y., Wang, C., Gu, X., Samaras, D., and Paragios, N. (2013). A generic deformation model for dense non-rigid surface registration: A higher-order MRF-based approach. In *IEEE International Conference on Computer Vision 2013*, pages 3360–3367.

Zeng, Y., Wang, C., Wang, Y., Gu, X., Samaras, D., and Paragios, N. (2010). Dense non-rigid surface registration using high-order graph matching. In *Computer Vision and Pattern Recognition 2010*, pages 382–389.

Zhang, J., Tian, D., Lin, R., zhou, G., Peng, G., and Sua, M. (2014). Phase-contrast x-ray ct imaging of esophagus and esophageal carcinoma. *Sci. Rep.*

Zhang, M., Singh, N., and Fletcher, P. T. (2013). Bayesian estimation of regularization and atlas building in diffeomorphic image registration. In *IPMI 2013*.

Zigelman, G., Kimmel, R., and Kiryati, N. (2006). Laplace-beltrami eigenfunctions towards an algorithm that "understands" geometry. *SMI '06 Proceedings of the IEEE International Conference on Shape Modeling and Applications 2006*, page 13.

Zitová, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000.