3D Reconstruction of Anatomical Structures from Endoscopic Images

Chenyu Wu

CMU-RI-TR-10-04

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Robotics

The Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

01/2010

Thesis Committee
Branislav Jaramaz (Co-chair)
Srinivasa Narasimhan (Co-chair)
Yanxi Liu
Ko Nishino (Drexel University)

This thesis is dedicated to my father who taught me to use what I have learned to help people, It is also dedicated to my mother who taught me that if I make a wish and work hard, it will come true. Finally this thesis is dedicated to my grandmothers who taught me to be brave and patient and to my beloved grandfather who moved to heaven last month.

Abstract

Endoscopy is attracting increasing attention for its role in minimally invasive, computer-assisted and tele-surgery. Analyzing images from endoscopes to obtain meaningful information about anatomical structures such as their 3D shapes, deformations and appearances, is crucial to such surgical applications. However, 3D reconstruction of bones from endoscopic images is challenging due to the small field of view of the endoscope, large image distortion, featureless surfaces and occlusion by blood and particles. In this thesis, a novel methodology is developed for accurate 3D bone reconstruction from endoscopic images, by exploiting and enhancing computer vision techniques such as shape from shading, tracking and statistical modeling.

We first designed a complete calibration scheme to estimate both geometric and photometric parameters including the rotation angle, light intensity and light sources' spatial distribution. This is crucial to our further analysis of endoscopic images. A solution is presented to reconstruct the Lambertian surface of bones using a sequence of overlapped endoscopic images, where only partial boundaries are visible in each image. We extend the classical shape-from-shading approach to deal with perspective projection and near point light sources that are not co-located with the camera center. Then, by tracking the endoscope, the complete occluding boundary of the bone is obtained by aligning the partial boundaries from different images. A complete and consistent shape is obtained by simultaneously growing the surface normals and depths in all views. Finally, in order to deal with over-smoothness and occlusions, we employ a statistical atlas to constrain and refine the multi-view shape from shading. A two-level framework is also developed for efficient atlas construction.

Acknowledgments

First and foremost I am heartily thankful to my advisor, Branislav Jaramaz, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject. This thesis would not have been possible without his insight into this field and continuous advice on problem formulations. He has made available his support in a number of ways for my research, work and life, which has become a constant source of inspiration and encouragement.

I owe my deepest gratitude to my co-advisor, Srinivasa Narasimhan. This thesis would not have been completed without his commitment and diligent efforts which not only influenced the content of the thesis but also shaped my research methodology. His dedication to research has deeply influenced and inspired me. The knowledge and skills that I have learned from him will continue to guide me in the future.

I would like to thank the other members of my thesis committee, Yanxi Liu and Ko Nishino for their insightful discussion and comments on my thesis. Their expertise and guidance have helped me tremendously on my research. I would like to thank George Stetten and Doug James for their valuable advice, and Suzanne Lyons Muth for her help during my graduate program. I would like to thank Leon Gu, Qifa Ke and Sanjeev J. Koppal for their helpful discussion on my thesis work. I would like to thank Patricia Murtha, James E. Moody, and other ICAOS members for their help on collecting data. I would like to thank Sichen Sun and Jie Liu for their help on the final thesis editing.

I am indebted to Wenyi Zhao and Christopher J. Hasser from Intuitive Surgical, Inc. for a precious experience gained. I am also indebted to my officemates and fellow students at CMU for making my time in graduate school productive and rewarding.

I would like to show my gratitude to Jing Xiao and Tadashi Shiozaki from Epson Research and Development, Inc. for all their support in my thesis writing.

Finally, I owe everything to my dear parents. It is only through their love, encourage and dedication that I have accomplished my Ph.D.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One of the main goals of contemporary surgery is to enable minimally invasive procedures. During surgery, an endoscope consisting of a camera and one or more light sources is inserted through a small incision into the body to acquire images for analysis. In addition to visualizing the interior of the anatomy, its role can be significantly enhanced by tracking the endoscope in conjunction with surgical navigation systems or robotic surgical systems.

There are many kinds of endoscopes: rigid or flexible, forward viewing or oblique viewing. The oblique viewing endoscope is the one most commonly used in orthopaedic surgery. Fig. 1.1 shows an oblique endoscope illuminating and observing an artificial spine. This endoscope has two light sources placed at its tip for illumination.

1.1 Challenges from Endoscopic Images

Because the tip of the endoscope is typically a few millimeters away from the bone surface, we are only able to observe a small portion of the bone. As a result, it can be difficult even for a skilled surgeon to deduce actual bone shape from a single 2D endoscopic image. Thus, there is an immediate need for explicit computer reconstruction of bone shapes from endoscopic images. There are many advantages to reconstructing 3D shape from endoscopic images, such as real-time, radiation free and low cost in-vivo visualization.

In the past several years researchers have explored many 3D reconstruction approaches [48]. For example, endoscopic images are overlaid on models from CT and MRI to localize the endoscopic views in a larger anatomic context [51, 13, 21, 86], where an expensive CT or MRI and a complex registration procedure are required. Some researchers use spe-

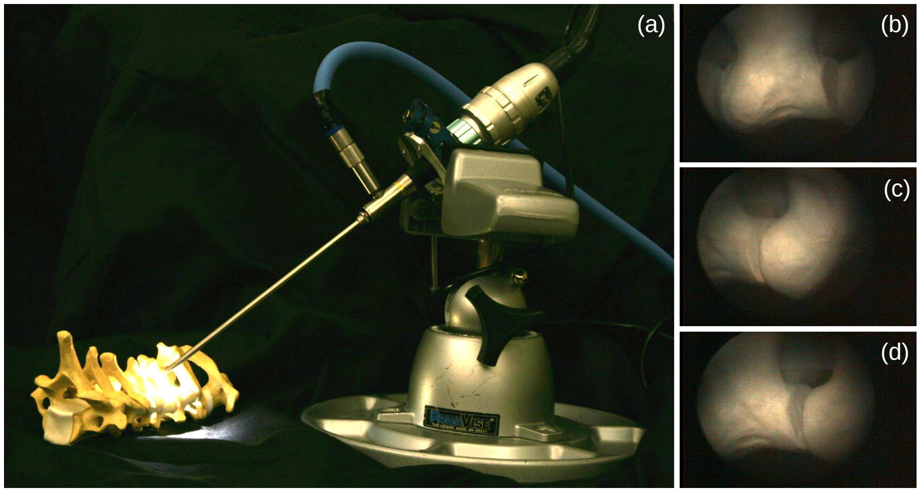


Figure 1.1: Illustration of an endoscope (a) and endoscopic images of an artificial spine (b,c,d). (It is difficult to perceive the shape of the object in the images because of small field of view, distortion, lack of texture, partial boundaries, and shading caused by near-field lighting.)

cial devices such as structured light [31, 45, 37]. This method requires the installation of a specially designed high speed laser projector at the tip of the endoscope, which is not commonly used in the operation rooms. For soft tissues with good features, standard computer vision technologies can be easily applied to track and reconstruct the 3D information.

Classical approaches, including shape-from-shading [68, 111, 29, 66, 95, 110, 39] and shape-from-motion [19, 66, 67, 97, 7, 88, 108] cannot be directly applied to featureless bone images. Besides, due to the small distances between the sources, the camera and the scene, the endoscopic images are very different from the images of natural scenes under distant lighting, such as from the sun or the sky. Moreover, during minimally invasive surgery, many substances (such as blood, pieces of bone, tissue) may obscure the bone surface. This can mislead any local constraints used in shape recovery. In addition, environmental noise including fog, tools interaction, specularity, etc. need to be considered as well.

1.2 Assumptions

This thesis takes the first step towards meeting the above challenges. We build a simplified setup for ex-vivo experiments under the following assumptions:

1. Artificial bones with Lambertian reflectance are used in our experiments.
2. A monocular oblique endoscope is used for capturing images.
3. A surgical navigation system is used for tracking the endoscope.
4. Neither random occlusions nor environmental noise are taken into account.
5. The focus of the endoscope is fixed during experiments.

1.3 Goals and Contributions

Our goal for the ex-vivo setup is to reconstruct a large field of view of the bone using all the information available during surgery. We will combine the shading information of the bone images and the motion information from the tracking system to achieve this goal. Specifically, this thesis makes the following contributions:

1. We developed a complete calibration scheme to obtain the parameters of the photometry and geometry of the endoscope. The photometry parameters include the camera response function, light source intensity and spatial distribution. The geometric parameters include intrinsic, extrinsic and rotation parameters. The rotation between the camera head and the scope cylinder creates an additional degree of freedom which makes calibrating oblique endoscopes more difficult. All the calibrated parameters are used in our reconstruction algorithm.
2. We modeled the appearance of the bone under perspective projection and near-point lighting, without assuming that the light sources are located at the center of the projection, to reconstruct the shape from a single featureless bone image. Our method extends the classical shape from shading algorithm to include near lighting and perspective projection and recovers the scene depth Z explicitly in addition to recovering the surface normals.
3. We presented a multi-image shape-from-shading (MISFS) framework to reconstruct a big part of the bone from multiple images by tracking the endoscope. We computed the global boundary constraints by aligning local contours in the world coordinates. One should notice that our global shape-from-shading framework does not really merge shape-from-shading and shape-from-motion in the classical sense of the terms. We performed shape-from-shading for all images simultaneously and use global boundary constraints in each iteration.

4. We developed an semi-automatic two-level registration procedure to construct the statistical shape prior for bone structures from population. This method segments, generates and aligns the 3D surface model from CT scanned images simultaneously. This statistical shape prior, also known as statistical atlas, is used to improve the MISFS algorithm in a global way.

5. We presented a two-step algorithm by combining bottom-up reconstruction and top-down refining. In the bottom-up step, we search for an atlas shape that most closely matches the initial erroneous shape resulting from the MISFS algorithm. The surface normal and depth of the atlas shape are then used as initialization and geometric constraints instead of the smoothness constraint in the MISFS. The bottom-up method exploits the atlas to improve the MISFS by solving the over-smoothness and partial occlusion problems. The top-down step is inspired by the generative model method. We compare the original endoscopic images with the synthesized images from the atlas shape obtained in the bottom-up step. The difference is used to refine the atlas shape. The top-down refinement increases shape details and accuracy.

1.4 Outline of Dissertation

In Ch. 2, we will review literature, concerning current solutions for 3D estimation in computer vision for both natural scenes and internal organs. We then introduce a complete calibration procedure for calibrating endoscopes' geometry and photometry in Ch. 3. In Ch. 4 we present a solution to the reconstruction of bone from multiple images. In Ch. 5 we will introduce the statistical atlas and in Ch. 6, we present a two-step algorithm to further improve reconstruction. Summarization and discussion are in Ch. 7.

Chapter 2

State of the Art

The field of surgery has been revolutionized over the past twenty years as a result of the introduction of endoscopic surgery, new imaging devices, computer assisted navigation systems and robotic technologies in the operating room. Compared to traditional surgery, these technologies provide numerous advantages such as higher accuracy, smaller incision, minimal invasiveness, less pain and blood loss, and faster recovery for patients. With the invention of the endoscope, surgeons are able to view interior anatomical structures through a tiny incision, instead of through their naked eyes. In addition, a real-time endoscopic image can be sent back to the computer for analysis. By using the position tracking technique, surgical tools can be localized in real-time. In some specific surgeries, for example, an orthopaedic surgery, e.g. bones and implants can also be tracked. These tracked objects, along with endoscopic images, and previous augmented 3D models from CT or MRI scans can be displayed on the computer simultaneously, in order to provide a virtual guide for the surgeons.

2.1 Computer Assisted Diagnosis and Surgery

Since the invention of the first 2D radiography in 1895, more and more imaging modalities have been developed and applied in the medical field. For instance, the 2D ultrasound was introduced in several countries since the 1940s [23, 104] and the 3D CT was introduced by British engineer Godfrey Hounsfield in the 1970s [87]. With these modalities, doctors are able to examine interior anatomical structures without invasive operations. After the 3D MRI was introduced into the market in the 1980s [2], doctors have been able to gather more information from the images due to the MRI's strength in observing soft tissues. Later in

1987, the first 3D ultrasound scanner was developed at Duke University [24] to capture images of the heart from outside the body. As technology enabled smaller ultrasound arrays, researchers started to design probes that could fit within catheters threaded through blood vessels for imaging the heart and its vasculature from the inside.

Conventional CT and MRI scans produce cross section ”slices” of the body that are viewed sequentially by radiologists who must imagine or extrapolate from these views the actual 3D anatomy. By using computer algorithms these cross sections can be rendered as the direct 3D representations of human anatomy. Virtual endoscopy, a 3D representation based navigation system for diagnosis was therefore proposed [77, 78, 99, 100, 62, 40, 82]). Virtual endoscopy creates a 3D virtual environment of the human anatomy based on CT or MRI images. A virtual camera is placed in such a virtual environment and is navigated around. Then a series of virtual images can be synthesized and analyzed by computers. This technique provides a way to make surgery plannable. However, there is still no texture information involved in the system. How to obtain a textured image during surgery? The answer is to capture real images via endoscopes. An endoscope has lens system and light sources at the tip that could be inserted into the body. With a near-field lighting, images of the interior structures are captured and sent back to the computer. Such an instrument supports a better visualization for surgeons. For example, these endoscopic images can be fused into the previous 3D CT models [51, 22].

2.2 Endoscopy

To view clearly through a tiny hole of the body requires a certain amount of lighting. To record what has been viewed requires a good lens system to transmit the light rays back to the image recording device. Endoscopy meets the above requirements. A surgical procedure using endoscopes is referred as minimally invasive surgery, also known as endoscopy.

The history of endoscopy is long. Early in 1805, a German doctor, Philip Bozzini tried to use a candle to illuminate the interior structures through a hole on an animal. Other pioneers cited in medical journals include French surgeon Antoine Jean Desormeaux (1815-1894), American otolaryngologist Chevalier Jackson (1865-1958), Polish physician Jan Mikulicz-Radecki (1850-1905), German urologist Maximilian Nitze (1848-1906), German gastroenterologist Rudolph Schindler (1888-1968), and German gynecologist Kurst Semm (1927-2003). The first ”endoscopy” in the world was conducted by a military surgeon, William Beaumont, in 1922. He used a flashlight as a light source to examine a wounded soldier’s abdomen through a tiny hole. At that time, however, the light source was outside the body. Since the 1920s, electric light source has been utilized since it can be transmitted through a rigid tube although not very deeply. After 1956, when the optical fiber was invented, viewing of interior structures became practical.

Early endoscopes were designed to use as an eyepiece to view through the incision but not for recording images. Because the tip of the instrument is very small (the diameter of the tip is around 10-15mm), the field of view is restricted. In 1929, the oblique viewing endoscope was designed in Germany by Kalk, who enlarged the field of view by rotating the scope cylinder. No image recording devices were available for endoscopes until 1954, when a viewfinder rod lens system was developed to capture images. And then, Soulas (France) developed the first televised bronchoscope (1956) and closed circuit television program (1959). CCD chip camera appeared in 1984 and high definition CCD has been adopted by Olympus for endoscopes since 2003.

Since 1970s, nurses have been required to have laparoscopy training, and then routine laparoscopy surgery started to operate in 1984. The state-of-the-art systems run the gamut from the possibility of disposable rigid endoscopes (Optiscope Technologies, Ltd. (Katzrin, Israel)), spectrally-encoded endoscopy (Massachusetts General Hospital's Wellman Center for Photomedicine (Boston)), Minute On-Chip Sensor MT9V021 (Micron Technology, Inc. (Meridian, Idaho and Glasgow, UK)), wireless ingestible imaging capsules (Tarun Mullick et el. Olympus Corp., Given Imaging Ltd. (1989-2006)), real-time structured light endoscopy (University of North Carolina at Chapel Hill, 2003), to stereoscopic endoscopy (da Vinci Surgical System manufactured by Intuitive Surgical Inc.) to breakthrough three-dimensional imaging.

2.3 2D to 3D endoscopy

Early video cameras for endoscopic surgeries are monocular and allow only two-dimensional visualization. The lack of depth perception significantly reduces a surgeon's ability to determine the precise size and location of anatomical structures and limits his/her capacity to maneuver, diagnose and operate efficiently.

Visionsense Corporation (Orangeburg, N.Y. and Petah-Tikva, Israel) developed a stereoscopic sensor that provides the surgeon with real-time, high-resolution, natural stereoscopic vision. The proprietary single sensor is based on multidisciplinary technologies combined with sophisticated image processing algorithms. Since 2000, many other companies such as Olympus started to manufacture stereoscopic endoscopes, which have been widely used in surgical systems such as the da Vinci's system (Intuitive Surgical Inc.). two optical elements are put within a 12-millimeter diameter endoscope. One of them is for the left eye and the other one is for the right eye, and both of them supply endoscopic images to a separate video camera. They are combined together to provide a true three-dimensional visualization. Those image are then projected onto left and right monitors where the surgeon fuse a 3D perspective by watching both, simultaneously. However, projecting stereo images to the left and right monitors cannot provide exact depth information for computer analysis. The researchers in the University of North Carolina at Chapel Hill later developed a real-time structured light depth extraction for endoscopy [31, 50]. A high speed projector projecting structured light into interior structures was designed to be inserted to the body and a high speed camera was used to sample the structured light patterns and outputs digital signals to computers. 3D ultrasound probe [24] was also designed to acquire real-time images of interior structures during endoscopy. With the development of the image processing techniques, software based solutions to 3D endoscopy becomes promising.

2.4 3D Vision for Regular Scenes

There is a long and rich history of 3D reconstruction in computer vision [42, 114, 74, 73, 35]. For natural scenes captured by regular cameras, which contains lots of shape features such as corners, edges, it is easy to locate the features and find correspondences between adjacent views. If the camera is fixed, then we can mosaic images together to obtain a panorama. Otherwise if the camera is moving, we can apply stereo or structure from motion to reconstruct the 3D information of the scene. A brief comparison between the different methods in listed below:

1. Multi-view Stereo: Rich texture features are required for establishing correspondences. Given both correspondences and calibrated camera motions, depth can be recovered. This method assumes that the lighting is distant and fixed. In other words, the texture features do not depend on the lighting conditions.

2. Shape-from-motion: Rich texture features are required for establishing correspondences. Given only correspondences, motion and shape structure can be computed by factorization. Same as the above approach, this method assumes that the texture features do not depend on the lighting conditions.

3. Shape-from-shading: Constant surface albedos are assumed in this method. Different methods have been developed to deal with distant or near-field lighting, and orthogonal or perspective projection camera models. Initial values on occluding boundaries are required as a starting point for this approach.

4. Shape-from-silhouette: Given camera motions and image silhouettes, surfaces of 3D object can be reconstructed. This method requires multiple cameras working simultaneously to achieve high resolution reconstruction.

5. Photometric stereo: This method reconstructs the surface normal by using several images of the same surface taken from the same viewpoint but under illumination from different directions. In each case there is a well-defined light source direction from which to measure the surface orientation. Therefore, the change of the intensities in the images depends on both local surface orientations and illumination directions. Varying lightings are the key to this solution.

6. Shape-from-structured light: By projecting structured patterns onto the surface, the correspondences can be easily established between different views. This approach requires special devices to create structured patterns such as laser projectors.

Most of the conventional methods assume distant lighting and require rich features and correspondences, or expensive experimental setups (e.g. structured light). So it is still challenging to obtain an affordable solution for 3D reconstruction from endoscopic images with the existing methods.

2.5 3D Vision for Surgery

Endoscopic images of anatomical structures are very different from what we have seen in the regular scenes. Small field of view, big distortion, varying illumination, featureless bones, non-lambertian and deformable soft tissues, sub-surface scattering, fog, tool interactions, etc., make the reconstruction procedure very hard.

Researchers have tried many existing methods for 3D visualization from endoscopic images [48]. For example, overlaying endoscopic images with 3D models obtained from CT or MRI scans [51, 13, 21, 86], which requires pre computing the 3D model and complicated registration procedure. Registration and referencing [102, 103] between surgical tools and 3D CT models for endoscopy have been studied. Structured light [31, 45, 37] is a promising approach due to its high accuracy, but it needs special devices such as high speed projector and cameras. Photometric stereo [105, 54, 55] requires images captured under different lighting conditions and is not applicable to experiments with a conventional endoscope with fixed lighting directions.

In orthopaedic surgery, given bone surfaces have few identifiable features, surface shading is the primary cue for shape. Shape-from-shading has a long and rich history in computer vision and biological perception [114, 25], however, most of the work focused on distant lighting, orthographic projection and Lambertian surfaces [42, 56, 114]. Few papers considered the more realistic scenarios. Real materials can be non-Lambertian model [60]. With a pin-hole camera, the projection is perspective instead of orthographic [71, 59, 36, 68, 81, 95]. For some special imaging devices, such as endoscopes, the light sources are not located at infinity [68, 76], thus $1/r^2$ attenuation term of the illumination need to be taken into account [76]. There are three papers that have established the equations for the perspective projection [75, 94, 17], and also there are a variety of mathematical tools that have been utilized to solve the perspective shape from shading (PSFS) problem [53, 17, 95, 76, 96]. The most relevant work about shape-from-shading under a near point light source and perspective projection for an endoscope was addressed in [68, 29, 66, 95], and they all assume that the light source and camera center are co-located. However, this assumption is inaccurate in the setting of our work, because the scene is 5-10 mm away from the endoscope and the distance between the source and the camera center is 3.5mm, which is in the same order of magnitude as the distance to the scene.

Due to the small field of view of the endoscope, only a partial shape can be obtained from a single image. By capturing image sequences while the endoscope is moving, it is possible to cover a larger part of the shape, which is easier to perceive [83]. There is a long and rich history of shape-from-motion in computer vision (both calibrated and uncalibrated) [73, 74, 35], and in the context of the anatomical reconstruction, the heart coronary [66, 67], tissue phantom [88], and other internal organs [97, 7, 108] were reconstructed using motion cues (image features and correspondences). However, it is difficult to establish correspondences for textureless (or featureless) bones.

Neither shape-from-shading nor shape-from-motion can individually solve the problem of bone reconstruction from endoscopic images. Our key idea in this thesis is to utilize the strengths of both to develop a multi-view shape-from-shading approach, and to introduce a shape prior to deal with over-smoothness and partial occlusions.

Chapter 3

Calibration of Endoscope's Geometry and Photometry

Camera geometric calibration, an important step in endoscope related applications, mostly based on Tsai's model [98], has been addressed in several work [21, 115, 84, 109]. However, except for [109], most of those methods dealt with the forward-viewing endoscope, in which the viewing direction is aligned with the axis of the endoscope. Due to the size of the small incision, the range of movement for such a tool is restricted and the field of view is very small. In order to view sideways, an oblique scope was designed with a tilted viewing direction such that a wider viewing field can be reached by rotating the scope cylinder. Fig. 3.1 shows an oblique-viewing endoscope. A new degree of freedom, a rotation, happens between the scope cylinder and the camera head. Yamaguchi et al. first modeled and calibrated the oblique scopes [109]. They formulated the rotation of the scope cylinder as an additional extrinsic parameter. They also used two extra transformations to compensate the rotation θ of the lens system and the stillness of the camera head. Yamaguchi et al's camera model successfully compensates the rotation effect but their method requires five additional parameters, and the model is complex. In this chapter we introduce an alternative approach to simplify the calibration. We attach an optical marker to the scope cylinder instead of the camera head, with a newly designed coupler (as Fig. 3.1(b) illustrates). We come up with a simpler model with only one additional intrinsic parameter.

Camera photometric calibration, another important process in illumination related applications, is performed to find the relationship between the image irradiance and the image intensity for the camera. This relationship is called the camera response function. Traditional photometric calibration recovers only the camera response function by changing the camera's exposure time. However, compared to regular cameras, it is hard to

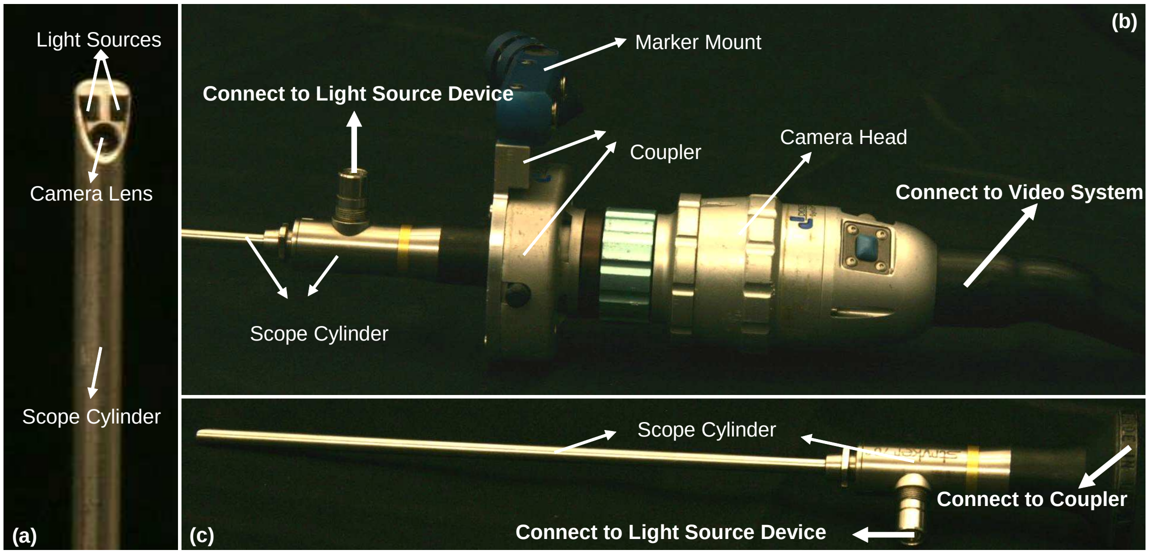


Figure 3.1: An oblique endoscope consists of a scope cylinder with a lens and point light sources at the tip (the tip has a tilt from the scope cylinder), a camera head that captures video images, and a light source device that supports the illumination. Scope cylinder is connected to the camera head via a coupler. This connection is flexible such that you can rotate either the scope cylinder or the camera head separately, or rotate them together.

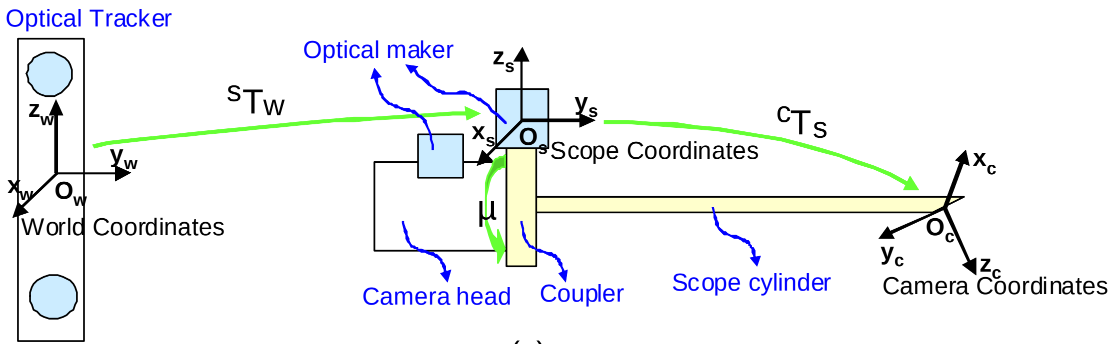
control the exposure time for endoscopes. Given the near field sources, the light spatial distribution can be anisotropy. And the light intensity can be changed during surgery. We present a method to calibrate all the unknown parameters simultaneously.

3.1 Geometric Calibration

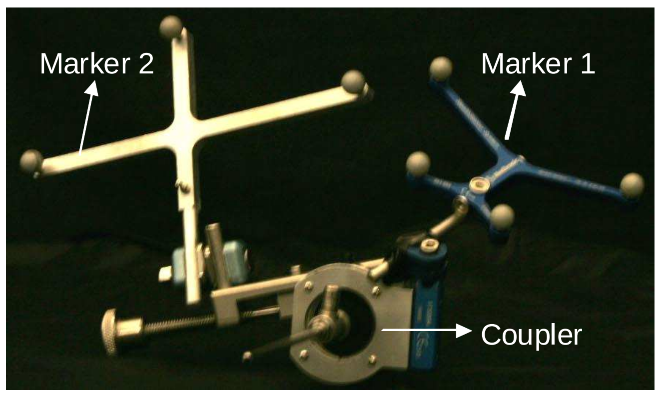
An orthopedic oblique endoscope is equipped with a single camera and one or more point light sources at the tip of the scope. In our work, we use two oblique endoscopes for testing. One of them is shown in Fig. 3.1 and the other is in Fig. 3.5.

Table 3.1: Table of Notation in Sec. 3.1

O_w	-	origin of world coordinates
O_s	-	origin of scope coordinates
O_c	-	origin of camera coordinates
${}^{m}T_w, {}^{s}T_w$	-	transformation from world to marker/scope
${}^{c}T_s, {}^{c}T_m(\theta)$	-	transformation from scope/marker to camera
λ	-	arbitrary scale factor
$\vec{P}_w$	-	3D point in world coordinates
$\vec{\mu}_i$	-	2D image pixel with rotation θ
$\vec{\mu}'_i$	-	2D image pixel without rotation θ
A_c	-	camera intrinsic matrix
θ	-	rotation angle
$\vec{l}_s$	-	axis of scope cylinder
$\vec{l}_h$	-	z-axis of lens system
$T_R(\theta; \vec{l}_s)$	-	rotation around $\vec{l}_s$
$T_R(-\theta; \vec{l}_h(\theta))$	-	inverse rotation around $\vec{l}_h(\theta)$
cc	-	camera principal point
$R(\theta)$	-	rotation function of θ
O_1	-	origin of Marker 1's coordinates
O_2	-	origin of Marker 2's coordinates
$\vec{P}_r$	-	3D point in Marker 2's coordinates
$\vec{P}_r^A, \vec{P}_r^B, \vec{P}_i$	-	$\vec{P}_r$ in Marker 1's coordinates at position A, B, i
$\vec{P}_r^B$	-	$\vec{P}_r$ in Marker 1's coordinates at position B
$\vec{O}$	-	center of trajectory of Marker 2 in Marker 1's coordinates
${}^{o_1}T_{o_w}{}^{i}, {}^{o_2}T_{o_w}{}^{i}$	-	transformation from world to O_1, O_2 for Marker 2's position i
M	-	number of corner points in calibration image



(a)



(b)

Figure 3.2: The geometric model of endoscope in conjunction with a tracking system. A new coupler (see Fig. 3.1 (b))is designed for mounting an optical marker to the scope cylinder, which ensures that the transformation from the scope(marker) coordinates O_s to the lens system (camera) coordinates O_c is fixed. The world coordinates O_w is defined by the optical tracker. Two optical markers are attached to the coupler and the camera head separately for the use of calculating the rotation angle θ in between.

3.1.1 Modeling Oblique-viewing Endoscope

Yamaguchi et al's camera model is based on Tsai's model:

$$\begin{aligned} \lambda\vec{\mu}_i &= A\,{}^cT_m(\theta)\,{}^mT_w\vec{P}_w \\ {}^cT_m(\theta) &= T_R(-\theta;\vec{l}_h(\theta))T_R(\theta;\vec{l}_s)\,{}^cT_m(0) \end{aligned} \tag{3.1}$$

In their model, λ is an arbitrary scale factor, $\vec{P}_w$ is a 3D point in the world coordinates, $\vec{\mu}_i$ is the corresponding 2D image pixel. A_c is the camera intrinsic matrix. ${}^mT_w$ is the rigid transformation from the world coordinates to the optical marker coordinates, ${}^cT_m(\theta)$ is the rigid transformation from the marker (camera head) to the camera coordinates. ${}^cT_m(\theta)$ is dependent on the rotation angle θ. Considering the marker coordinates (camera head) as a reference, only the lens system rotates while the camera head, i.e., the image plane, remains fixed irrespective to the rotation. The authors formulated such a transformation due to the rotation by decomposing the one physical rotation into two mathematical rotations. $T_R(\theta; \vec{l}_s)$ is the rotation of both scope cylinder and the camera head (image plane) around the axis of cylinder $\vec{l}_s$. $T_R(-\theta; \vec{l}_h(\theta))$ is the inverse rotation of the image plane around the z-axis of lens system $\vec{l}_h$. Both $\vec{l}_s$ and $\vec{l}_h$ have two unknown parameters. Although this model works well, it is complex.

As Fig. 3.2 (b) shows, we attach an optical marker to the scope cylinder instead. Our model is still an extension of Tsai's model but simpler. The geometric model illustrated in Fig. 3.2 (a) is given by:

$$\begin{aligned} \lambda \vec{\mu}'_i &= A_c \cdot {}^cT_s \cdot {}^sT_w \cdot \vec{P}_w \\ \vec{\mu}_i &= R(\theta) \cdot (\vec{\mu}'_i - cc) + cc \end{aligned} \tag{3.2}$$

where $\vec{P}_w$ is a 3D point in the world coordinates, $\vec{\mu}'_i$ is the corresponding 2D image pixel without any rotation, $\vec{\mu}_i$ is the image pixel with the rotation θ. ${}^sT_w$ is the rigid transformation from the world coordinates to the optical marker coordinates (scope cylinder). ${}^cT_s$ is the rigid transformation from the marker (scope cylinder) to the camera coordinates and independent on θ. cc is the principal point. $R(\theta)$ represents the rotation of the image plane around cc by θ. The intrinsic matrix A_c and the external matrix ${}^cT_m$ are calibrated by using Zhang's method [115]. ${}^mT_w$ is directly obtained from the tracking system. We only need to estimate one additional degree of freedom, the rotation angle. A comparison between Yamaguchi et al's model and ours can be found in Appendix II.

Rotation angle can be estimated by using a rotary encoder, as Yamaguchi et al [109] did. When it is not available, rotation angles can be estimated by using two optical markers: one attached to the scope cylinder and the other to the rod (camera head).

3.1.2 Estimation of Rotation Angle Using Two Optical Markers

Let the marker attached to the scope cylinder be the Marker 1 and the marker to the camera head be the Marker 2 (see Fig. B.1 (b)). As Fig. 3.3 shows, when we rotate the camera head around the scope cylinder from the position "A" to "B" by θ, the point $\vec{P}_r$ in the

Marker 2's coordinates O_2 will move along a circular trajectory with respect to the point $\vec{O}$ on the axis of the scope cylinder, which is in the Marker 1's coordinates O_1. We estimate the center $\vec{O}$ of the circle first and compute θ by

$$\theta = \arccos \frac{\|O\vec{P}_r^A\|^2 + \|O\vec{P}_r^B\|^2 - \|P_r^A\vec{P}_r^B\|^2}{2\|O\vec{P}_r^A\| \cdot \|O\vec{P}_r^B\|} \tag{3.3}$$

The center of the circle can be represented in terms of the transformations from the world coordinates O_w to the Marker 1's coordinates O_1, and the Marker 2's coordinates O_2. At least 3 different positions in the Marker 2's coordinates (O_2) (with different θ) are necessary.

3.1.3 Estimation of the center of circle in 3D

We rotate the camera head around the cylinder to acquire 3 different positions with respect to the Marker 2. Let the transformation matrix from the world coordinates O_w to both Marker 1's coordinates O_1 and Marker 2's coordinates O_2 for the position i be (${}^{o_1}T_{o_w}{}^i$, ${}^{o_2}T_{o_w}{}^i$) ($i = 1, 2, 3$). Given any point $\vec{P_r}$ in O_2, we compute the position $\vec{P_i}$ in O_1 corresponding to different rotations as:

$$\vec{P_i} = {}^{o_1}T_{o_w}{}^i \cdot ({}^{o_2}T_{o_w}{}^i)^{-1} \cdot \vec{P_r}, i = 1, 2, 3. \tag{3.4}$$

$\vec{O}$ is the center of the circumcircle of the triangle ($\Delta\vec{P_1}\vec{P_2}\vec{P_3}$).

Let $\vec{R_1} = \vec{P_1} - \vec{P_3}$, $\vec{R_2} = \vec{P_2} - \vec{P_3}$, the normal of the triangle is $\vec{n}_\Delta = \vec{R_1} \times \vec{R_2}$. The perpendicular bisector $\vec{L_1}$ of $\vec{R_1}$ and $\vec{L_2}$ of $\vec{R_2}$ can be computed as:

$$\begin{aligned} \vec{L_1} &= \vec{P_3} + \vec{R_1}/2 + \lambda_1 \cdot \vec{n}_\Delta \times \vec{R_1} \\ \vec{L_2} &= \vec{P_3} + \vec{R_2}/2 + \lambda_2 \cdot \vec{n}_\Delta \times \vec{R_2} \end{aligned} \tag{3.5}$$

where λ_1 and λ_2 are parameters of the line $\vec{L_1}$ and $\vec{L_2}$. The intersection of the two lines represents the center of the circle. From Eq. 3.5 we can derive the center of the circle by

$$\vec{O} = \frac{(\vec{R_2} - \vec{R_1}) \cdot \vec{R_1}/2}{|\ \vec{R_1} \times \vec{R_2}\ |^2} \cdot (\vec{R_1} \times \vec{R_2}) \times \vec{R_2} + \vec{R_2}/2 + \vec{P_3} \tag{3.6}$$

It can be easily proved that $\vec{O}$ does not depend on the selection of $\vec{P_r}$. Since at least 3 different positions are necessary, we rotate the camera head around the scope cylinder by

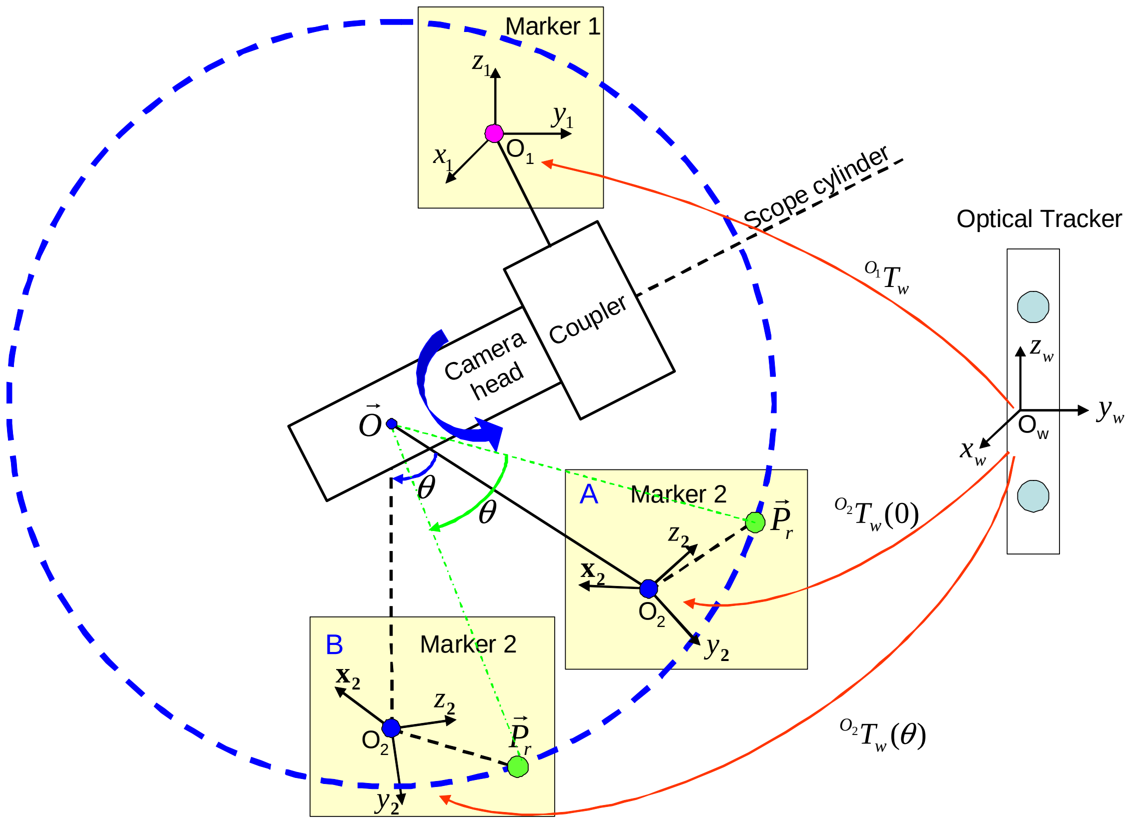


Figure 3.3: Illustration of the relationship between the rotation angle θ and two markers' coordinates. The Marker 1 is attached to the scope cylinder and the Marker 2 is attached to the camera head. "A" indicates the position of the Marker 2 when $\theta = 0$ and "B" indicates the position of the Marker 2 with a rotation θ. For any point $\vec{P}_r$ in the Marker 2's coordinates, its trace with the rotation of the camera head follows a circular trajectory in the Marker 1's coordinates (It moves from the position $\vec{P}_r^A$ to $\vec{P}_r^B$). This circle is also in the plane perpendicular to the axis of the scope cylinder. $\vec{O}$ is the center of the circle.

N different angles. We apply a RANSAC algorithm to estimate $\vec{O}$ using random positions, and finally select $\vec{O}^*$ corresponding to the smallest variance as the center of the circle. The pseudo code of RANSAC is listed in Table 4.3. It can also be proved that θ does not depend on the selection of P_r as well. A similar RANSAC algorithm, as Table 3.3 shows, is used to compute θ. Fig. 3.4 shows estimated rotation angles using the RANSAC algorithm for

Table 3.2: Pseudo RANSAC code for estimating the center of the circle

Loop k=1:K (K=2000)
Generate a random point P_r from 3D space
Generate random number x,y,z between [1,N]
Compute P_x,P_y,P_z using Eq. 3.4
Compute O_k using Eq. 3.6
Compute $\mid \overline{O_k P_j} \mid, j \in [1, N], j \neq x, y, z$
Compute v_k
Save O_k, v_k
End loop
Return $O_q, q = arg_k min(v_k)$

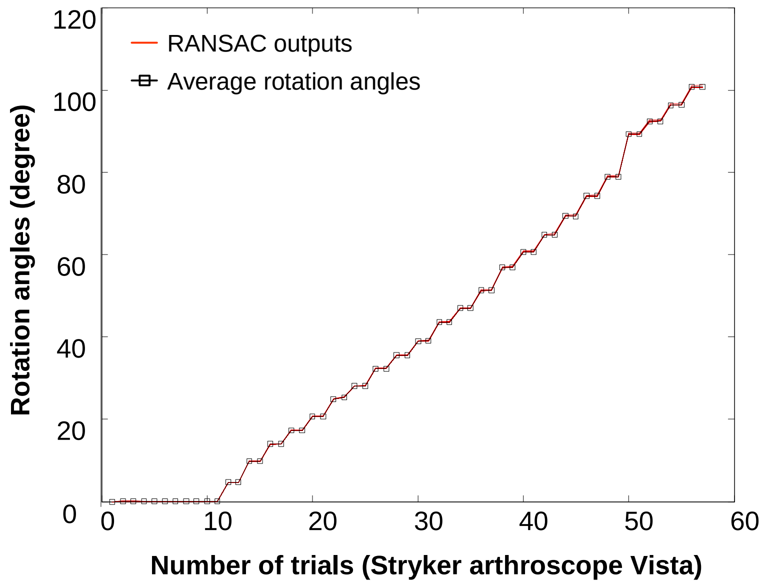
Table 3.3: Pseudo RANSAC code for estimating the rotation angle

Loop k=1:K (K=1000)
Generate a random point P_r from 3D space
Compute P_A and P_B using Eq. 6.2
Compute θ_k using Eq. 3.3
End loop
Return $\theta = \frac{1}{K} \sum_k \theta_k$

two different endoscopes. The red curves are estimated angles from different RANSAC iterations, and the black curve is the average. We can see that the variance of the estimation is very small (less than 0.2 degree).

3.1.4 Experimental Results

We tested our algorithm using two different systems. We first tested it in our lab. We used a Stryker 344-71 arthroscope Vista (70 degree, 4mm) oblique-viewing endoscope, a DYONICS DyoCam^{TM} 750 video camera, a DYONICS DYOBRITE 3000 light source, and Polaris (Northern Digital Inc., Ontario, Canada) optical tracker. Next we tested it in



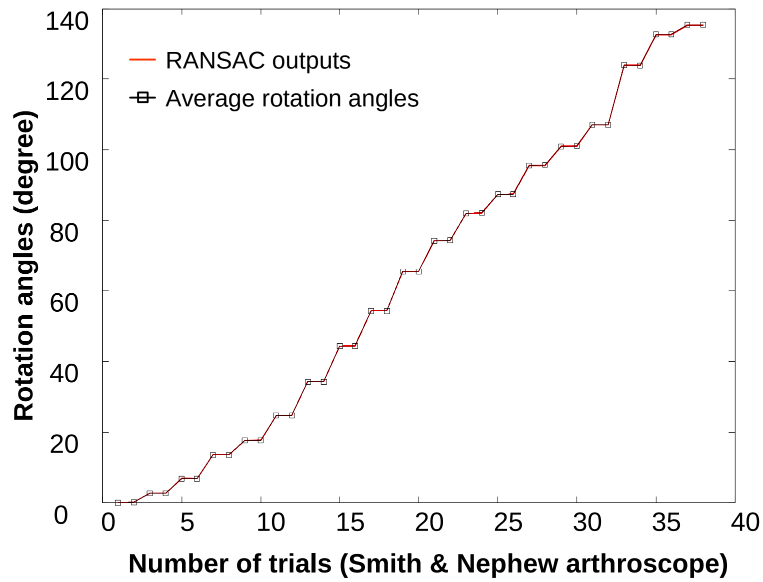


Figure 3.4: Estimated rotation angles for two endoscopes. In each trial, we rotated the camera head with respect to the scope cylinder and captured an image. We captured a few images for the initial position. After that we took two images for each rotation angle. The red curves are estimated rotation angles from different RANSAC iterations. The black curve is the average.

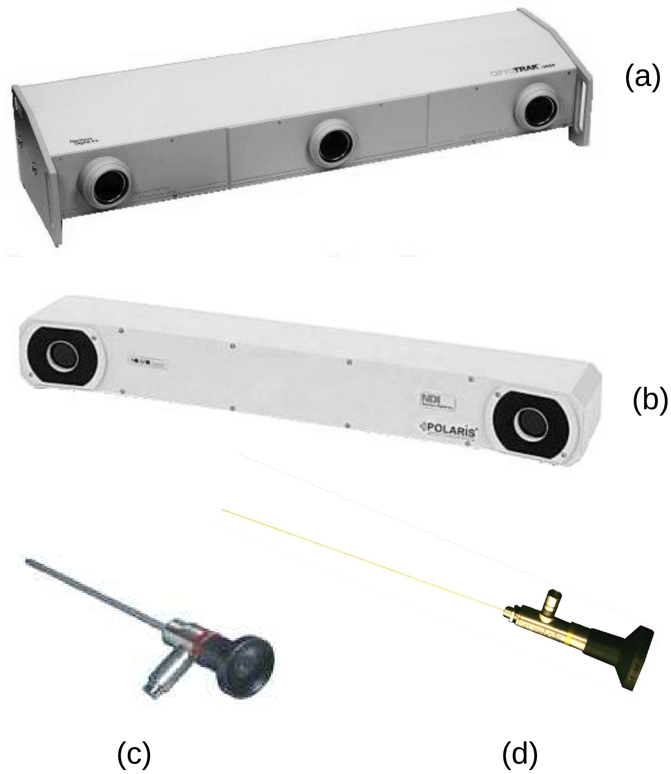
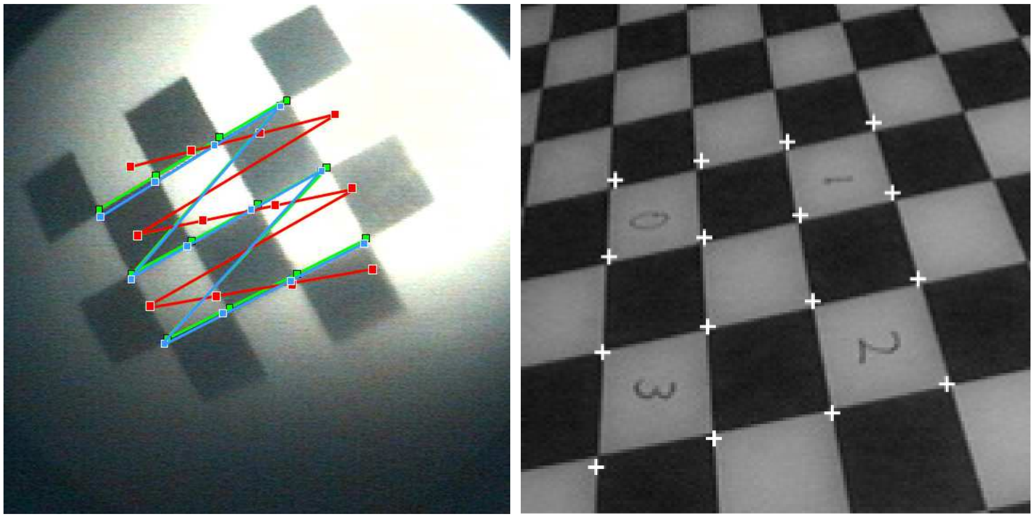


Figure 3.5: Optical trackers and endoscopes used in the experiments. (a) OPTOTRAK optical tracker (Northern Digital Inc., Ontario, Canada). (b) Polaris optical tracker (Northern Digital Inc., Ontario, Canada). (c) Smith & Nephew video arthroscope - autoclavable SN-OH 272589 (30 degree, 4mm). (d) Stryker 344-71 arthroscope Vista (70 degree, 4mm).

the operating room. We used a Smith & Nephew video arthroscope - autoclavable SN-OH 272589 (30 degree, 4mm), a DYONICS video camera and light source, OPTOTRAK (Northern Digital Inc., Ontario, Canada) optical tracker. Fig. 3.5 shows the different endoscopes and optical trackers.

The endoscope was first fixed and the calibration pattern was rotated on the table for capturing images. A set of images were captured without a rotation between the scope cylinder and the camera head. They were used to estimate both the intrinsic matrix A (including the focal length and radial distortion coefficients) and extrinsic matrix ${}^{c}T_{s}$ using Zhang's method [115] (implemented using OpenCV functions). After that, while rotating



(a) (b)

Figure 3.6: (a) Illustration of the back projection with and without a rotation compensation. Green points are ground truth - 2D corner pixels on the image of the calibration pattern. Red points are back projection of the 3D world positions of the corners using the first equation of Eq. 3.2, which has no rotation compensation. Blue points are back projection using both equations of Eq. 3.2. Since the rotation is included in the camera model, the back projected pixels are much closer to the ground truth than the red points. (b) An image used in Yamaguchi et al. [109]'s paper. This image has a higher resolution, better lighting and less distortion than ours.

the camera head with respect to the scope cylinder, another set of images were captured and the center of the circle can be computed by using Eq. 3.6. Next, we fixed the calibration pattern, with two optical markers attached to the scope cylinder and the camera head, we captured the third set of images by applying normal operations to the endoscope (moving the whole scope body or rotating the camera head with respect to the scope cylinder (or more natural description: rotating the scope cylinder with respect to the camera head)). This set of images were used to evaluating the calibration. The initial position of the camera head was considered as the reference position A illustrated in Fig. 3.3. Fig. 3.6 illustrates the back projection of 3D corners of the calibration pattern with (blue) and without (red) a rotation compensation. Green points are ground truth. For each rotation angle of the endoscope, we calculated the average back projection error for this angle

$$\epsilon(\theta) = \frac{1}{M}\sum_{i=1}^{M} \mid \vec{\mu}_i - \vec{\mu}(\vec{P_w^i}, \theta) \mid \tag{3.7}$$

where $\vec{P}_i$ is a 3D point in the world coordinates, $\vec{\mu}_i$ is the corresponding 2D image pixel, $\vec{\mu}(\vec{P_w^i}, \theta)$ is the back projected 2D image pixel of $\vec{P}_i$ computed by using Eq. 3.2. M is the number of corners on the calibration pattern. We used different grid patterns (3x4, 4x5, 5x6, 6x7. The size of each checker is 2mm x 2mm). In order to obtain enough light on the grid pattern, the endoscope needs to be placed very close to the target (usually 5-15mm). Since smaller grids cannot capture the radial distortion and the bigger grids will exceed the field of view, the 5x6 grid was selected to provide the best results.

We did many trials by moving and rotating the endoscope randomly and estimated θ simultaneously. The averaged back projection error with respect to the different rotation angles are shown in Fig. 3.7. Fig. 3.7 (a) shows the result using the Stryker 344-71 arthroscope Vista (70 degree, 4mm) and Polaris optical tracker. Fig. 3.7 (b) shows the result using the Smith $\&$ Nephew video arthroscope - autoclavable SN-OH 272589 (30 degree, 4mm) and OPTOTRAK optical tracker. The red curve represents the back projection error without taking into account of the rotation angle, and the blue curve shows the error considering the rotation angle. The results show that including the rotation angle into the camera model significantly improves the accuracy of the calibration.

Fig. 3.7 shows that different endoscopes have different accuracy. The reason is that endoscopes have different magnification and optical trackers have different accuracy (according to the manufacturer, RMS error is 0.1mm for OPTOTRAK and 0.3mm for Polaris). Yamaguchi et al. [109] used an OTV-S5C laparoscope (Olympus Optical Co. Ltd., Tokyo, Japan) and Polaris optical tracker. They achieved a high accuracy of less than 5mm back projection error when the rotation angle is within 140 degrees. Our results show that we can achieve the same level accuracy when the rotation angle is within 75 degrees. Beyond this range, due to the bigger magnification, larger radial distortion and poorer lighting (a comparison between images used in Yamaguchi et al.'s experiments and ours is shown in Fig. 3.6), the back projection error is increased to 13mm when the rotation angle is 100 degrees. Given endoscopes with the same level of quality, we should be able to achieve the same level of accuracy, with a simpler setup and procedure.

3.2 Photometric Calibration

Until now, photometric calibration for the endoscope is seldomly addressed in literature. We propose a method to compute the radiometric response function of the endoscope, and also simultaneously calibrate the directional/spatial intensity distribution of the near light sources (the camera and the sources are "on/off" at the same time) and light intensities. Our method is inspired by Litvinov et al.'s work [61].

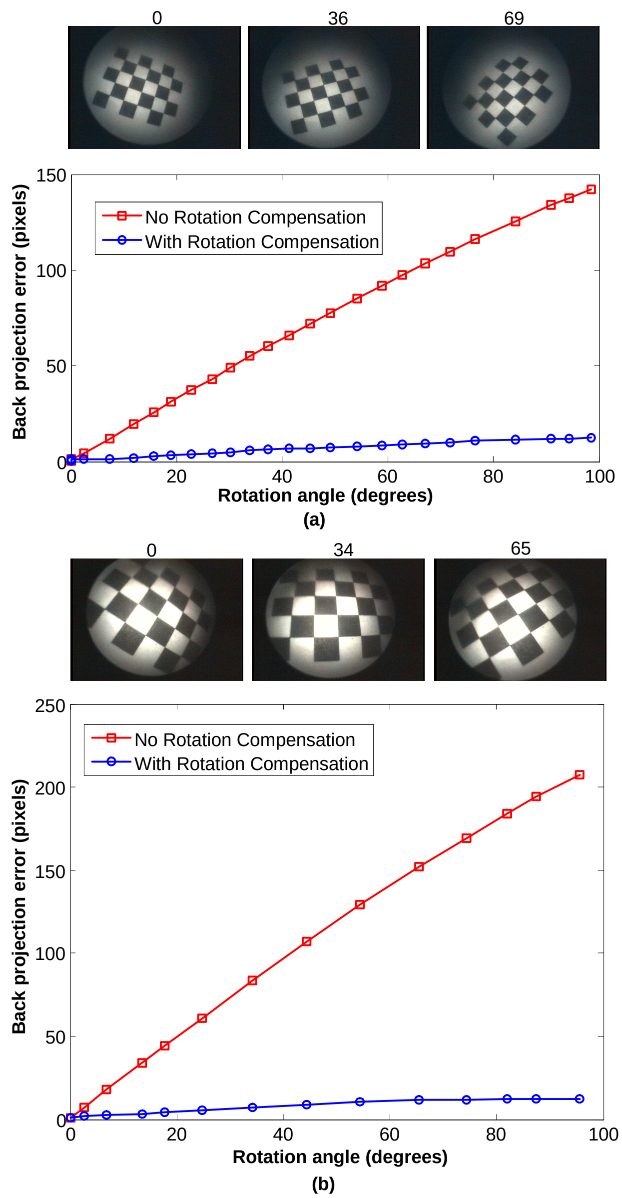


Figure 3.7: Back projection errors with respect to the rotation angles for two systems. (a) Stryker 344-71 arthroscope Vista and Polaris optical tracker in our lab. (b) Smith & Nephew video arthroscope and OPTOTRAK optical tracker in the operating room. Images in the top row of (a) and (b) correspond to different rotation angles (the number is shown on the top of each image). The red curves in (a) and (b) represent the errors without a rotation compensation. The blue curves in (a) and (b) are errors with a rotation compensation.

Table 3.4: Table of Notation in Sec. 3.2

R	-	scene radiance
I_0, I_{0j}	-	light intensity
ρ, ρ_i	-	surface albedo
$\vec{n}$	-	surface normal
s_1, s_2	-	light sources
l_1, l_2	-	light rays
r_1, r_2	-	distance from light source to the surface point
P	-	surface point
$I(\cdot)$	-	image irradiance function
$\tilde{x}, \tilde{y}$	-	image coordinates
x, y, z	-	world coordinates
$H(\cdot)$	-	camera response function
$v(\cdot)$	-	intensity value
$M(\cdot)$	-	source spatial distribution function
$\tilde{M}(\cdot)$	-	modified source spatial distribution function
$h(\cdot), eta_i, \gamma_j, \tilde{m}$	-	log representation of $H^{-1}, \rho_i, I_{0j}, \tilde{M}$

3.2.1 Image Irradiance Equation

Assuming the bone surface is Lambertian, the scene radiance can be computed according to Lambertian cosine law:

$$R(x, y, z) = I_0 \rho (\frac{\vec{n} \cdot \vec{l_1}}{r_1^2} + \frac{\vec{n} \cdot \vec{l_2}}{r_2^2}) \tag{3.8}$$

where I_0 is the light intensity of $\vec{s_1}$ and $\vec{s_2}$, ρ is the surface albedo. We use the unit vector $\vec{n}$ to represent the surface normal at $\vec{P}$, $\vec{l_1}$ and $\vec{l_2}$ are two light rays incident at $\vec{P}$, and r_1 and r_2 are the distance from each light source to the surface. (x, y, z) indicates the 3D location of the scene point P (see details in Ch. 4).

On the other hand, The image irradiance $I(\tilde{x}, \tilde{y})$ is related to the image intensity, also known as gray level v, through the camera response function $H(\cdot)$:

$$I(\tilde{x},\tilde{y}) = \frac{H^{-1}[v(\tilde{x},\tilde{y})]}{M(\tilde{x},\tilde{y})} \tag{3.9}$$

where $M(\tilde{x},\tilde{y})$ represents the anisotropy of the source spatial distribution. The two sources are identical and they are oriented in the same way so we assume that their spatial distribution functions $M(\tilde{x},\tilde{y})$ are equal. From Eqs 3.8 and 3.9 we have:

$$\begin{aligned} H^{-1}[v(\tilde{x},\tilde{y})] &= \rho \cdot I_0 \cdot \tilde{M}(\tilde{x},\tilde{y}) \\ where \quad \tilde{M}(\tilde{x},\tilde{y}) &= M(\tilde{x},\tilde{y}) \cdot (\frac{\vec{n}\cdot\vec{l_1}}{r_1^3} + \frac{\vec{n}\cdot\vec{l_2}}{r_2^3}) \end{aligned} \tag{3.10}$$

During the calibration, we used a Macbeth color chart with known albedos for each patch. We captured a set of images by varying the source intensity for each patch. We applied $\log$ to both sides of Eq. 3.10 and obtained a linear system:

$$h[v_i^j(\tilde{x},\tilde{y})] = \eta_i + \gamma_j + \tilde{m}(\tilde{x},\tilde{y}) \tag{3.11}$$

where i indicates the surface albedo and j indexes the light intensity. $h[v_i^j(\tilde{x},\tilde{y})] = \log\{H^{-1}[v_i^j(\tilde{x},\tilde{y})]\}$, $\eta_i = \log(\rho_i)$, $\gamma_j = \log(I_{0j})$ and $\tilde{m}(\tilde{x},\tilde{y}) = \log[\tilde{M}(\tilde{x},\tilde{y})]$. The unknowns ($h(\cdot)$, γ_j, $\tilde{m}(\tilde{x},\tilde{y})$) can be estimated by solving this linear system of equations. The cosine term is then estimated by physically measuring the distance to the chart from the scope tip and finally $M(\tilde{x},\tilde{y})$ is recovered.

3.2.2 Solution to $h(\cdot)$

Given the fixed light intensity γ_j and pixel value $v(\tilde{x},\tilde{y})$ but two different albedos η_{i_1} and η_{i_2}, we have

$$\begin{cases} h[v_{i_1}^j(\tilde{x},\tilde{y})] - \gamma_j - \tilde{m}(\tilde{x},\tilde{y}) - \eta_{i_1} = 0 \\ h[v_{i_2}^j(\tilde{x},\tilde{y})] - \gamma_j - \tilde{m}(\tilde{x},\tilde{y}) - \eta_{i_2} = 0 \end{cases} \tag{3.12}$$

Subtract the first line from the second line of Eq. 3.12 we obtain

$$h[v_{i_2}^j(\tilde{x},\tilde{y})] - h[v_{i_1}^j\tilde{x},\tilde{y})] = \eta_{i_2} - \eta_{i_1} \tag{3.13}$$

We selected different pixels in the same image (albedo) or different images (albedos) to make as many equations as Eq. 3.13, as long as we fixed the light intensity for each pair of albedos. Since $v_{i_1}^j(\tilde{x},\tilde{y})$ changes from 0 to 255(image intensity), we only need 256 such equations and stack them as:

$$\begin{bmatrix} \cdots & & 1^* & & -1^* & \cdots \\ \cdots & -1^\# & & 1^\# & & \cdots \\ & \cdots & & & \cdots & \end{bmatrix} \cdot \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(255) \end{bmatrix} = \begin{bmatrix} \eta_{i_2} - \eta_{i_1} \\ \eta_{i_4} - \eta i_3 \\ \vdots \end{bmatrix} \tag{3.14}$$

where 1^* and -1^* correspond to the column $v_{i_2}^j(\tilde{x},\tilde{y})+1$ and $v_{i_1}^j(\tilde{x},\tilde{y})+1$, respectively. $1^\#$ and $-1^\#$ correspond to the column $v_{i_4}^j(\tilde{x},\tilde{y})+1$ and $v_{i_3}^j(\tilde{x},\tilde{y})+1$, respectively. Therefore $h(v)$ is solved from Eq. 3.14 and $H(\cdot) = \{\exp[h(v)]\}^{-1}$.

3.2.3 Solution to γ_j

Given the fixed albedo η_i and pixel value $v(\tilde{x},\tilde{y})$ but two different light intensities γ_{j_1} and γ_{j_2} we have

$$\begin{cases} h[v_i^{j_1}(\tilde{x},\tilde{y})] - \gamma_{j_1} - \tilde{m}(\tilde{x},\tilde{y}) - \eta_i = 0 \\ h[v_i^{j_2}(\tilde{x},\tilde{y})] - \gamma_{j_2} - \tilde{m}(\tilde{x},\tilde{y}) - \eta_i = 0 \end{cases} \tag{3.15}$$

Subtract the first line from the second line of Eq. 3.15:

$$h[v_i^{j_2}(\tilde{x},\tilde{y})] - h[v_i^{j_1}(\tilde{x},\tilde{y})] = \gamma_{j_2} - \gamma_{j_1} \tag{3.16}$$

We use the minimum light intensity γ_1 as a reference, for other light intensities $\gamma_j, j = 2,\cdots,N_{light}$, we have

$$\gamma_j = \gamma_1 + \{h[v_i^j(\tilde{x},\tilde{y})] - h[v_i^1(\tilde{x},\tilde{y})]\} \tag{3.17}$$

With the estimated $h[v(\tilde{x},\tilde{y})]$ and by changing the albedos and pixels, we compute the average for each γ_j as below and $\bar{I_0}_j = \exp(\gamma_1)\cdot\exp(\bar{\gamma}_j)$.

$$\bar{\gamma}_j = \gamma_1 + \frac{1}{N_{albedo}} \cdot \frac{1}{N_{pixels}} \sum_i^{N_{albedo}} \sum_{\tilde{x},\tilde{y}}^{N_{pixels}} \{h[v_i^j(\tilde{x},\tilde{y})] - h[v_i^1(\tilde{x},\tilde{y})]\} \tag{3.18}$$

3.2.4 Solution to $\tilde{m}(\tilde{x},\tilde{y})$

Again, Given the fixed albedo η_i and light intensity γ_j but two different pixels $(\tilde{x}_p,\tilde{y}_p)$ and $(\tilde{x}_q,\tilde{y}_q)$ we have

$$\begin{cases} h[v_i^j(\tilde{x}_p,\tilde{y}_p)] - \gamma_j - \tilde{m}(\tilde{x}_p,\tilde{y}_p) - \eta_i = 0 \\ h[v_i^j(\tilde{x}_q,\tilde{y}_q)] - \gamma_j - \tilde{m}(\tilde{x}_q,\tilde{y}_q) - \eta_i = 0 \end{cases} \tag{3.19}$$

Subtract the first line from the second line of Eq. 3.19:

$$h[v_i^j(\tilde{x}_q, \tilde{y}_q)] - h[v_i^j(\tilde{x}_p, \tilde{y}_p)] = \tilde{m}(\tilde{x}_q, \tilde{y}_q) - \tilde{m}(\tilde{x}_p, \tilde{y}_p) \tag{3.20}$$

For each γ_j and η_i, instead of using $C^2_{N_{pixels}}$ different pairs of pixels, we choose only $N_{pixels} = 720 \times 480$ pairs and stack the equations as below:

$$\underbrace{\begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & & \cdot & \\ & & & & \cdot \\ -1 & & & & 1 \end{bmatrix}}_{\Phi} \cdot \begin{bmatrix} \tilde{m}(\tilde{x}_1, \tilde{y}_1) \\ \tilde{m}(\tilde{x}_2, \tilde{y}_2) \\ \vdots \\ \tilde{m}(\tilde{x}_{N_{pixels}}, \tilde{y}_{N_{pixels}}) \end{bmatrix} = \begin{bmatrix} h[v_i^j(\tilde{x}_1, \tilde{y}_1)] - h[v_i^j(\tilde{x}_2, \tilde{y}_2)] \\ h[v_i^j(\tilde{x}_2, \tilde{y}_2)] - h[v_i^j(\tilde{x}_3, \tilde{y}_3)] \\ \vdots \\ h[v_i^j(\tilde{x}_{N_{pixels}}, \tilde{y}_{N_{pixels}})] - h[v_i^j(\tilde{x}_1, \tilde{y}_1)] \end{bmatrix} \tag{3.21}$$

It's not practical to solve Eq. 3.21 using SVD directly since matrix Φ requires a huge amount of memory. Instead we have noticed that the matrix Φ in Eq. 3.21 is a special N_{pixels} by N_{pixels} matrix therefore we calculate the inverse directly by using Gauss-Jordan Elimination:

$$\Phi^{-1} = \begin{bmatrix} 0 & & & & 0 & -1 \\ -1 & 0 & & & 0 & -1 \\ -1 & -1 & 0 & & 0 & -1 \\ \vdots & \vdots & \vdots & \cdots & 0 & -1 \\ -1 & -1 & -1 & \cdots & -1 & -1 \end{bmatrix} \tag{3.22}$$

Finally we can efficiently compute each element of $\tilde{m}(\tilde{x}, \tilde{y})$ independently and again, $\tilde{M}(\tilde{x}, \tilde{y}) = \exp[\tilde{m}(\tilde{x}, \tilde{y})]$.

3.2.5 Experimental Results

A series of images of color chart were used for photometric calibration. We tested 6 different levels of image intensities. In Fig. 3.8, (a),(b) and (c) show the camera response function in Red,Green,Blue channels. (d) shows the recovered light intensity in different levels and compared to the ground truth. The smaller number in x-axis corresponds to the higher intensity. We noticed a bit variance when light intensity is high due to saturation. (e) shows the original image and (f) shows $\tilde{m}$. (g) shows the cosine term $\frac{(n)\cdot \mathbf{l_1}}{r_1^2} + \frac{(n)\cdot \mathbf{l_2}}{r_2^2}$ and (h) shows the spatial distribution function $m(x, y)$.

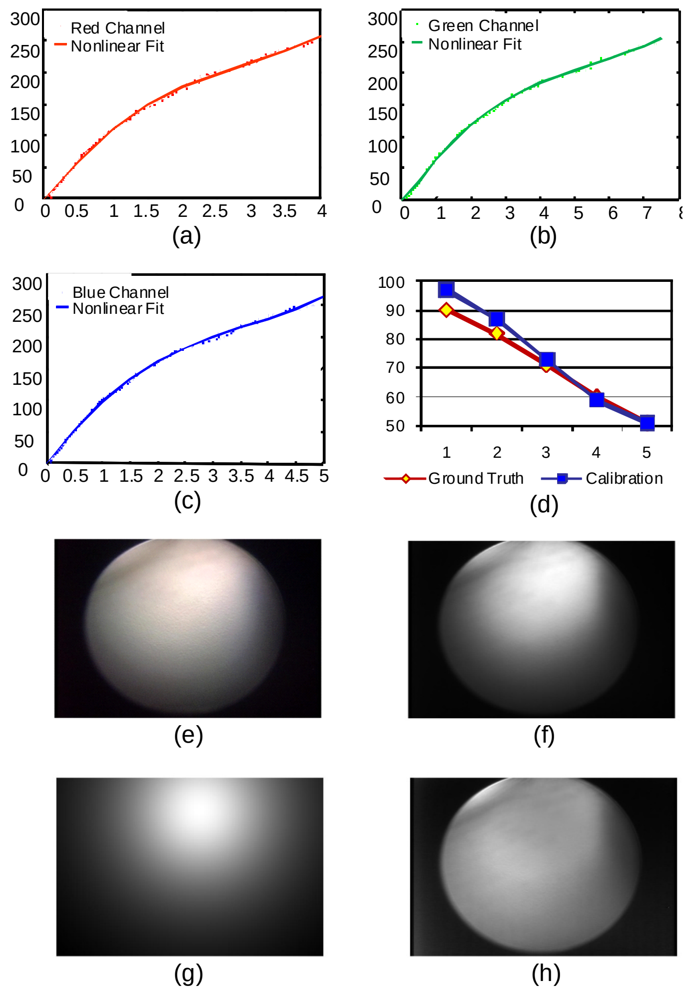


Figure 3.8: Results of photometric calibration. (a) camera response function in Red channel. Red dots represents the data points and magenta line represents the nonlinear fit. (b) camera response function in Green channel. Green dots represents the data points and magenta line represents the nonlinear fit. (c) camera response function in Blue channel. Blue dots represents the data points and magenta line represents the nonlinear fit. (d) Calibrated light intensity in different levels (blue) and ground truth (green). We use the level 6 as a reference and plot level 1-5. A small level corresponding to a high light intensity. A bit variation in the range of high intensities may be caused by saturation. (e) Original image of color chart. (f) $\tilde{m}$. (g) cosine term $\frac{(n)\cdot \mathbf{l_1}}{r_1^2} + \frac{(n)\cdot \mathbf{l_2}}{r_2^2}$. (h) spatial distribution function $m(x, y)$.

3.3 Conclusions and Discussion

We have developed a comprehensive calibration procedure for estimating both geometric and photometric parameters for oblique endoscopes. Our geometric calibration method simplifies the previous work given a newly designed coupler attached to the scope cylinder. It is easy to implement and practical to apply with the standard operating room equipments such as the navigation system. The only drawback of this method is that it requires the two markers to be visible to the optical tracker all the time, otherwise the method will fail.

According to our knowledge, photometric calibration for endoscopes has been less studied. Most of related work did not rely on the physical model of light sources, or they restricted the changing of light sources during the operation. A few of recent work applied shape-from-shading to endoscopic images based on a simplified light source model without calibrating endoscopes. However, in order to reconstruct an accurate shape from endoscopic images, the knowledge of light sources is necessary and important.

Both geometrical and photometrical parameters are very useful for 3D visualization and reconstruction of anatomical structures such as bones from endoscopic images. We will use calibrated endoscopes for artificial spine reconstruction (see Chaps. 4 and 6). The results demonstrate that using calibrated endoscopes can achieve good reconstruction result, which is promising for the real surgical applications.

Chapter 4

3D Reconstruction for Bones from Endoscopic Images

To reconstruct the 3D shape of bones from endoscopic images, we formulate the problem as a near-lighting shape-from-shading with a pinhole camera (perspective projection) and present a solution to reconstruct the Lambertian surface of bones using a sequence of overlapped endoscopic images, with partial boundaries in each image. First we formulate the shape-from-shading problem to deal with perspective projection and near point light sources that are not co-located with the camera center. Secondly, we propose a multi-image framework which aligns partial shapes obtained from different images in the world coordinates by tracking the endoscope. An iterative closest point (ICP) algorithm is used to improve the matching and recover complete occluding boundaries of the bones. Finally, a complete and consistent shape is obtained by simultaneously re-growing the surface normals and depths in all views. We demonstrate the accuracy of our technique by running simulations and conducting experiments with artificial bones.

4.1 Shape-from-Shading under Near Point Sources and Perspective Projection

With the special design of endoscopes, we formulate the shape-from-shading under near point lighting and a perspective projection, where the light sources are not located at the projection center. As Fig. 4.1 shows, given two point light sources $\vec{s_1}$ and $\vec{s_2}$, we compute the scene radiance emitted by the surface point $\vec{P}$ according to the Lambertian cosine law

Table 4.1: Table of Notation in Sec. 4.1

R	-	scene radiance
I_0	-	light intensity
ρ	-	surface albedo
$\vec{n}$	-	surface normal
s_1, s_2	-	light sources
l_1, l_2	-	light rays
r_1, r_2	-	distance from light source to surface point
P	-	surface point
$I(\cdot)$	-	image irradiance function
$\tilde{x}, \tilde{y}$	-	image coordinates
x, y, z	-	world coordinates
a, b	-	endoscope parameters
F	-	focal length
p, q	-	partial derivatives of z w.r.t. image coordinates $\tilde{x}, \tilde{y}$
$\mathbf{e}(\cdot)$	-	total error function
$\mathbf{e_i}(\cdot)$	-	irradiance error function
$\mathbf{e_s}(\cdot)$	-	smoothness constraint
λ	-	Lagrange multiplier
(k, l)	-	image indices
$z_{k,l}, p_{k,l}, q_{k,l}$	-	value of z, p, q at pixel (k, l)
n	-	iteration number
$\bar{z_{k,l}}^n, \bar{p_{k,l}}^n, \bar{q_{k,l}}^n$	-	average of neighborhood in the n^{th} iteration
$\psi(\cdot)$	-	local robust regularizer constraint
$\rho_\sigma(\cdot)$	-	robust error kernel function
ξ	-	symbol for p, q, z
$\xi_{\tilde{x}}, z_{\tilde{y}}$	-	partial derivatives of ξ w.r.t. image coordinates $\tilde{x}, \tilde{y}$
$\hat{\xi_{k,l}}^m$	-	local robust regularizer term in the n^{th} iteration

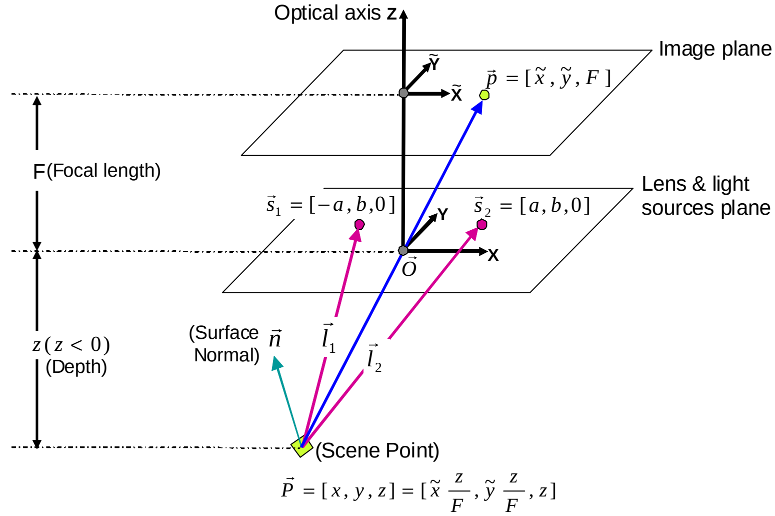


Figure 4.1: A perspective projection model for an endoscope imaging system with two near point light sources: $\vec{O}$ is the camera projection center. $\vec{s_1}$ and $\vec{s_2}$ are two light sources. We assume that the plane consisting of $\vec{O}$, $\vec{s_1}$ and $\vec{s_2}$ is parallel to the image plane. The camera coordinate system ($\mathbf{X}-\mathbf{Y}-\mathbf{Z}$) is centered at $\vec{O}$ and $\mathbf{Z}$-axis is parallel to the optical axis and is pointing toward the image plane. $\mathbf{X}$-axis and $\mathbf{Y}$-axis are parallel to the image plane. F is the focal length. a and b are two parameters related to the position of the light sources. Given a scene point $\vec{P} = (x, y, z)$, the projected image pixel is $\vec{p} = (\tilde{x}, \tilde{y}, F)$, where $(\tilde{x}, \tilde{y})$ are image coordinates. Assuming a Lambertian surface, the surface illumination therefore depends on the surface albedo, light source intensity and fall-off, and the angle between the normal and light rays.

and inverse square distance fall-off law of isotropic point sources [43]:

$$R = I_0\rho(\frac{\cos\theta_1}{r_1^2} + \frac{\cos\theta_2}{r_2^2}) \tag{4.1}$$

where I_0 is the light intensity of $\vec{s_1}$ and $\vec{s_2}$, ρ is the surface albedo. We use the unit vector $\vec{n}$ to represent the surface normal at $\vec{P}$. $\vec{l_1}$ and $\vec{l_2}$ are two light rays incident at $\vec{P}$, and r_1 and r_2 are the distances from each light source to the surface. We have

$$\begin{aligned} \cos\theta_1 &= \frac{\vec{n}\cdot\vec{l_1}}{\|\vec{l_1}\|}, \quad \vec{l_1} = \vec{s_1} - \vec{P}, \quad r_1 = \|\vec{l_1}\| \\ \cos\theta_2 &= \frac{\vec{n}\cdot\vec{l_2}}{\|\vec{l_2}\|}, \quad \vec{l_2} = \vec{s_2} - \vec{P}, \quad r_2 = \|\vec{l_2}\| \end{aligned} \tag{4.2}$$

According to [42], the surface normal $\vec{n}$ can be represented in terms of the partial derivatives of depth z with respect to x and y ((x, y, z) are camera coordinates):

$$\vec{n} = [-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1]/\sqrt{(\frac{\partial z}{\partial x})^2 + (\frac{\partial z}{\partial y})^2 + 1} \tag{4.3}$$

Note that for the orthographic projection $\partial z/\partial \tilde{x}$ and $\partial z/\partial \tilde{y}$ are used to represent surface normals, where $(\tilde{x}, \tilde{y})$ are image coordinates. Under the perspective projection we have

$$\begin{aligned} x &= \tilde{x}\frac{z}{F} \\ y &= \tilde{y}\frac{z}{F} \end{aligned} \tag{4.4}$$

where F is the focal length and $z < 0$. We take the derivatives of both sides of Eq. 4.4 w.r.t x and y and obtain

$$\begin{aligned} \frac{\partial z}{\partial x} &= \frac{1}{\tilde{x}}(F - \frac{z}{\partial x/\partial \tilde{x}}) \\ \frac{\partial z}{\partial y} &= \frac{1}{\tilde{y}}(F - \frac{z}{\partial y/\partial \tilde{y}}) \end{aligned} \tag{4.5}$$

We also take the derivatives of both sides of Eq. 4.4 w.r.t $\tilde{x}$ and $\tilde{y}$ and obtain

$$\begin{aligned} \frac{\partial x}{\partial \tilde{x}} &= \frac{1}{F}(z + \tilde{x}\frac{\partial z}{\partial \tilde{x}}) \\ \frac{\partial y}{\partial \tilde{y}} &= \frac{1}{F}(z + \tilde{y}\frac{\partial z}{\partial \tilde{y}}) \end{aligned} \tag{4.6}$$

Let $p = \partial z/\partial \tilde{x}$ and $q = \partial z/\partial \tilde{y}$, from Eqs 4.5 and 4.6 we have

$$\begin{aligned} \frac{\partial z}{\partial x} &= \frac{Fp}{z + \tilde{x}p} \\ \frac{\partial z}{\partial y} &= \frac{Fq}{z + \tilde{y}q} \end{aligned} \tag{4.7}$$

Given two light sources $s_1 = [-a, b, 0]$ and $s_2 = [a, b, 0]$ (calibration of a, b and F were discussed in Ch. 3), we can explicitly write light source vectors as follows:

$$\begin{aligned} \vec{l_1} &= [-a - \tilde{x}\frac{z}{F}, b - \tilde{y}\frac{z}{F}, -z] \\ \vec{l_2} &= [a - \tilde{x}\frac{z}{F}, b - \tilde{y}\frac{z}{F}, -z] \end{aligned} \tag{4.8}$$

Combining Eqs 4.1, 4.2, 4.3, 4.7 and 4.8, we obtain the reflectance map R as a function of $\tilde{x}, \tilde{y}, z, p, q$:

$$R(\tilde{x},\tilde{y},z,p,q) = I_0\rho\left(\frac{\vec{n}(\tilde{x},\tilde{y},z,p,q)\cdot\vec{l_1}(\tilde{x},\tilde{y},z)}{r_1(\tilde{x},\tilde{y},z)^3} + \frac{\vec{n}(\tilde{x},\tilde{y},z,p,q)\cdot\vec{l_2}(\tilde{x},\tilde{y},z)}{r_2(\tilde{x},\tilde{y},z)^3}\right)$$

$$= I_0\rho\left(\frac{p(z+\tilde{y}q)(a+\tilde{x}\frac{z}{F}) - q(z+\tilde{x}p)(b-\tilde{y}\frac{z}{F}) - \frac{z}{F}(z+\tilde{x}p)(z+\tilde{y}q)}{\sqrt{p^2(z+\tilde{y}q)^2+q^2(z+\tilde{x}p)^2+\frac{(z+\tilde{x}p)^2(z+\tilde{y}q)^2}{F^2}}\cdot\left(\sqrt{(-a-\tilde{x}\frac{z}{F})^2+(b-\tilde{y}\frac{z}{F})^2+(z)^2}\right)^3}\right.$$

$$\left.+\frac{-p(z+\tilde{y}q)(a-\tilde{x}\frac{z}{F}) - q(z+\tilde{x}p)(b-\tilde{y}\frac{z}{F}) - \frac{z}{F}(z+\tilde{x}p)(z+\tilde{y}q)}{\sqrt{p^2(z+\tilde{y}q)^2+q^2(z+\tilde{x}p)^2+\frac{(z+\tilde{x}p)^2(z+\tilde{y}q)^2}{F^2}}\cdot\left(\sqrt{(a-\tilde{x}\frac{z}{F})^2+(b-\tilde{y}\frac{z}{F})^2+(z)^2}\right)^3}\right) \tag{4.9}$$

Eq. 4.9 is similar to Eq. 2 in [95] and Eq.2 in [75], but we consider the $1/r^2$ as the fall-off since the light sources are very close to the bone surface. Eq. 4.9 can be easily extended to multiple point sources as below (M is the number of sources):

$$R(\tilde{x},\tilde{y},z,p,q) = I_0\rho\sum_{i}^{M}\left(\frac{\vec{n}(\tilde{x},\tilde{y},z,p,q)\cdot\vec{l_i}(\tilde{x},\tilde{y},z)}{r_i(\tilde{x},\tilde{y},z)^3}\right) \tag{4.10}$$

4.1.1 Solving Image Irradiance Equation

Given the image irradiance function $I(\tilde{x},\tilde{y})$, the image irradiance equation [42] is

$$R(\tilde{x},\tilde{y},z,p,q) = I(\tilde{x},\tilde{y}) \tag{4.11}$$

According to [25], different mathematical methods have been proposed to solve the equation for PSFS, including methods of resolution of PDEs [95, 96, 53, 76] and methods using optimization [17]. The optimization methods based on the variational approaches work well in most of general cases and does not need to set values to the singular and local minimum points [25], so we adopt the optimization method and plug it in the multi-image framework. Specifically, we solve Eq. 4.11 by minimizing the error between the image irradiance $I(\tilde{x},\tilde{y})$ and the reflectance map $R(\tilde{x},\tilde{y},z,p,q)$ given in Eq. 4.9. Image irradiance is obtained from the image intensity given the calibrated camera response function and spatial distribution (see details in Sec. 3.2). Different from the previous optimization methods [44], the depth z is explicitly included in R. We compute the error as:

$$\mathbf{e}(z,p,q) = \lambda \mathbf{e_i}(z,p,q) + (1-\lambda)\mathbf{e_s}(z,p,q) \tag{4.12}$$

and

$$\begin{aligned} \mathbf{e_i}(z,p,q) &= \int\int_{image}[I(\tilde{x},\tilde{y}) - R(\tilde{x},\tilde{y},z,p,q)]^2 d\tilde{x}d\tilde{y} \\ \mathbf{e_s}(z,p,q) &= \int\int_{image}[(z_{\tilde{x}}^2 + z_{\tilde{y}}^2) + (p_{\tilde{x}}^2 + p_{\tilde{y}}^2) + (q_{\tilde{x}}^2 + q_{\tilde{y}}^2)]d\tilde{x}d\tilde{y} \end{aligned} \tag{4.13}$$

where $\mathbf{e_i}(z,p,q)$ is the irradiance error and $\mathbf{e_s}(z,p,q)$ is the smoothness constraint for z, p and q. λ is a Lagrange multiplier. We find the solution to (z,p,q) by minimizing the error $\mathbf{e}(z,p,q)$ [58]:

$$[z^*,p^*,q^*] = \arg_{z,p,q} min(\lambda \mathbf{e_i} + (1-\lambda)\mathbf{e_s}) \tag{4.14}$$

Similar to [41], we discretize Eq. 4.12 and obtain:

$$\mathbf{e}(z_{k,l},p_{k,l},q_{k,l}) = \sum_k \sum_l (\lambda \mathbf{e_i}_{k,l} + (1-\lambda)\mathbf{e_s}_{k,l}) \tag{4.15}$$

The solution to Eq. 4.15 is to find $[p_{k,l}^*, q_{k,l}^*, z_{k,l}^*]$ that minimize $\mathbf{e}$:

$$\frac{\partial \mathbf{e}}{\partial p_{k,l}}|_{p_{k,l}^*} = 0, \frac{\partial \mathbf{e}}{\partial q_{k,l}}|_{q_{k,l}^*} = 0, \frac{\partial \mathbf{e}}{\partial z_{k,l}}|_{z_{k,l}^*} = 0 \tag{4.16}$$

Combining Eq. 4.15 to 4.16 yields update functions for $p_{k,l}$, $q_{k,l}$ and $z_{k,l}$ in each iteration [41]:

$$\begin{aligned} p_{k,l}^{n+1} &= \bar{p}_{k,l}^n + \frac{\lambda}{4(1-\lambda)}[I_{k,l} - R(k,l,\bar{z}_{k,l}^n,\bar{p}_{k,l}^n,\bar{q}_{k,l}^n)]\frac{\partial R}{\partial p_{k,l}}|_{\bar{p}_{k,l}^n} \\ q_{k,l}^{n+1} &= \bar{q}_{k,l}^n + \frac{\lambda}{4(1-\lambda)}[I_{k,l} - R(k,l,\bar{z}_{k,l}^n,\bar{p}_{k,l}^n,\bar{q}_{k,l}^n)]\frac{\partial R}{\partial q_{k,l}}|_{\bar{q}_{k,l}^n} \\ z_{k,l}^{n+1} &= \bar{z}_{k,l}^n + \frac{\lambda}{4(1-\lambda)}[I_{k,l} - R(k,l,\bar{z}_{k,l}^n,\bar{p}_{k,l}^n,\bar{q}_{k,l}^n)]\frac{\partial R}{\partial z_{k,l}}|_{\bar{z}_{k,l}^n} \end{aligned} \tag{4.17}$$

$\bar{p}_{k,l}^n$, $\bar{q}_{k,l}^n$ and $\bar{z}_{k,l}^n$ are local 8-neighborhood average around the pixel (k,l). Detailed derivation can be found in Appendix I. The value for the $(n+1)^{th}$ iteration can be estimated from the n^{th} iteration. During the iteration, the Lagrange multiplier λ is gradually increased such that the smoothness constraint is reduced as well. In our experiment, setting of Lagrance multiplier is empirical. We set it to 0.005 at the beginning and increase it by 0.02 whenever the error in Eq. 4.15 is reduced by 1%.

Given the initial values of p, q and z on the boundary, we compute a numerical solution to the bone shape. Variational methods usually rely on good initial guesses, so we

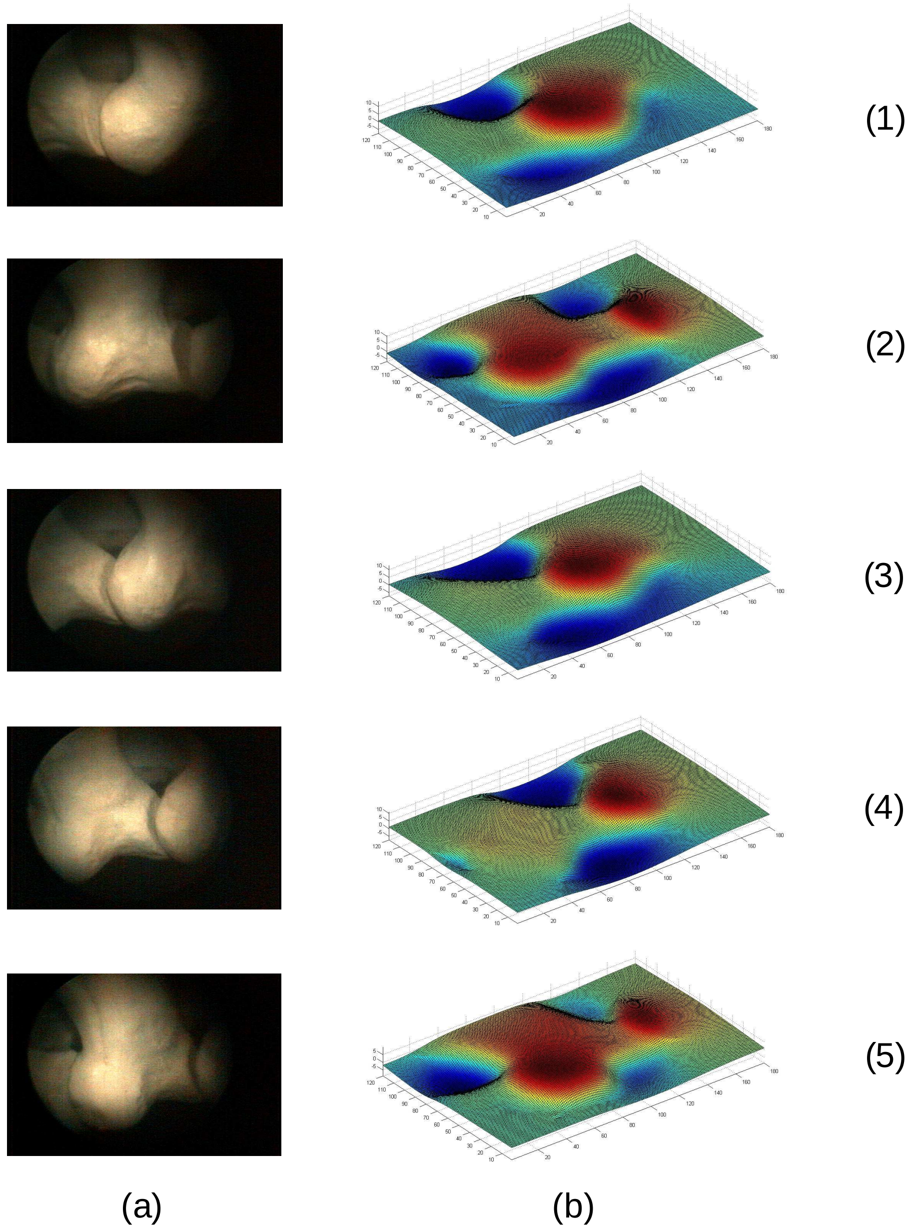


Figure 4.2: Results of shape from shading from a single image. (a) input image. (b) shape from shading. (1)-(5) are captured from different viewpoints.

manually label the boundaries of each images. After that, z is set to -1 for each pixel in the image. p and q are set to 0 for pixels that are not on the occluding boundaries. p and q at each pixel on the occluding boundaries are computed as the cross product of viewing vector (starting from the optical center) and the edge vector. Because it is difficult to estimate the initial values of z, we reset z values by integrating the normals using the method in [30], after several iterations when the error in Eq. 4.15 is reduced by 10%. We keep performing such an adjustment until the algorithm converges.

One of the drawbacks of this method is its tendency that overly smoothes out local discontinuities because only smoothness constraints are enforced. Therefore we introduce a local robust regularizer based on surface geometry [107], which considers that the reconstructed surface should be smooth, except where the probability of a discontinuity is high. The robust regularizer constraint function is defined as

$$\psi(\xi|\sigma) = \rho_\sigma(\|\frac{\partial \xi}{\partial \tilde{x}}\|) + \rho_\sigma(\|\frac{\partial \xi}{\partial \tilde{y}}\|), \;\; \rho_\sigma(\omega) = \frac{\sigma}{\pi} \log \cosh\left(\frac{\pi\omega}{\sigma}\right) \tag{4.18}$$

where $\xi \in \{\eta | p, q, z\}$. The robust error kernel function $\rho_\sigma(\omega)$, the sigmoidal derivative M-estimator (a continuous version of Huber's estimator), is proved to possess the best properties for handling surface discontinuities [107]. By discretizing Eq. 4.18 and applying the calculus of variations, we obtain the update term for ξ in terms of the robust regularizer constraint.

$$\begin{aligned} \hat{\xi}^n_{k,l} &= \frac{(\xi^n_{k+1,l} + \xi^n_{k-1,l})}{\|\xi^n_{\tilde{x}}\|} \tanh\left(\frac{\pi}{\sigma}\|\xi^n_{\tilde{x}}\|\right) + \frac{\pi}{\sigma}\left[\|\xi^n_{\tilde{x}}\| sech^2\left(\frac{\pi}{\sigma}\|\xi^n_{\tilde{x}}\|\right) - \tanh\left(\frac{\pi}{\sigma}\|\xi^n_{\tilde{x}}\|\right)\right] \\ &+ \frac{(\xi^n_{k,l+1} + \xi^n_{k,l-1})}{\|\xi^n_{\tilde{y}}\|} \tanh\left(\frac{\pi}{\sigma}\|\xi^n_{\tilde{y}}\|\right) + \frac{\pi}{\sigma}\left[\|\xi^n_{\tilde{y}}\| sech^2\left(\frac{\pi}{\sigma}\|\xi^n_{\tilde{y}}\|\right) - \tanh\left(\frac{\pi}{\sigma}\|\xi^n_{\tilde{y}}\|\right)\right] \end{aligned} \tag{4.19}$$

where $\xi_{\tilde{x}}$ and $\xi_{\tilde{y}}$ are average of the forward and backward difference with respect to $\tilde{x}$ and $\tilde{y}$. i.e. $\xi_{\tilde{x}} = [(\xi_{k+1,l} - \xi_{k,l}) + (\xi_{k,l} - \xi_{k-1,l})]/2$, $\xi_{\tilde{y}} = [(\xi_{k,l+1} - \xi_{k,l}) + (\xi_{k,l} - \xi_{k,l-1})]/2$. Note that Eq. 4.19 is in essential a scalar version of Eq. 26 in [107]. We will use $\hat{\xi}^n_{k,l}$ instead of the average $\bar{\xi}^n_{k,l}$ in Eq. 4.17.

By these means, some local discontinuities can be partially recovered. However, the recovered shape is still partial with limited information due to the small field of view.

Table 4.2: Table of Notation in Sec. 4.2

S^i	-	3D point set reconstructed from image i in camera coordinates
N_i	-	number of points in shape S^i
${}^{w}T_{m_i}$	-	transformation from marker coordinates to world coordinates
M	-	transformation from marker coordinates to camera coordinates
$\hat{S}^i_w, \hat{S}^{i-1}_w$	-	transformed 3D point set in world coordinates
${}^{i-1}\hat{S}^i_w$	-	aligned 3D point set with respect to adjacent point set $\hat{S}^{i-1}_w$
${}^{i-1}T_i$	-	rigid transformation from the i^{th} position to the $i-1^{th}$ position
x, y, z	-	world coordinates
p, q	-	partial derivatives of z w.r.t. image coordinates $\tilde{x}, \tilde{y}$

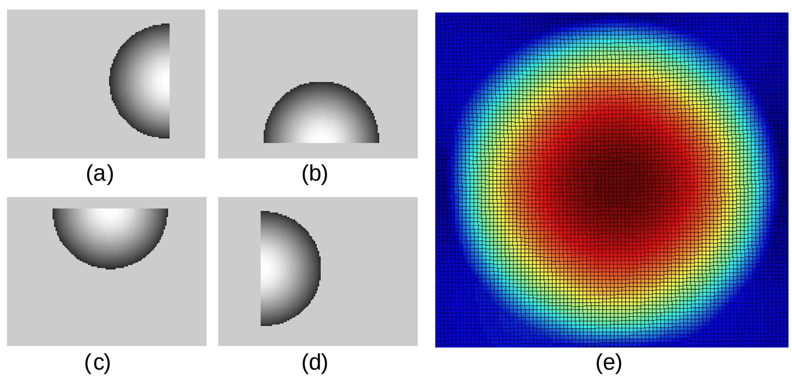


Figure 4.3: Simulation results of shape from shading from multiple views. (a)-(d) Synthesized images of different parts of a sphere. (e) Reconstructed sphere.

4.2 Global Shape-from-Shading using Multiple Partial Views

As shown in Fig. 4.2, most images include a small fraction of visible contours. How do we estimate the complete shape reliably? An intuitive solution is to merge individual shapes recovered from different views as shown in Fig. 4.3.

However, each image is defined in a view dependent local coordinates. We must first transform them to the world coordinates before we merge individual shapes. We solve this problem by using a navigation tracking system. We attach optical markers to the endoscope (as showed in Fig. 3.2) and track them to obtain transformations between the local and the world coordinates for each image.

Consider a 3D-point set $[S^i]_{3\times N_i}$ reconstructed from the image i, where N_i is the number of points on the shape S^i. If all the shapes are reconstructed correctly, without any errors from the tracking system and the calibration process, they should be perfectly aligned in the world coordinates. Let ${}^wT_{m_i}$ denote the transformation from the marker at the position capturing the i^{th} image to the world coordinates, and M denote the transformation from the marker coordinates to the camera coordinates (image distortion has been corrected in advance). Thus, any 3D-point set S^i in the local camera coordinates can be transformed to the world coordinates $\hat{S}^i_w$ by:

$$\hat{S}^i_w = {}^w T_{m_i} \cdot [M^{-1} \cdot S^i; \mathbf{1}_{1\times N_i}] \tag{4.20}$$

where $\hat{S}^i_w$ is represented using homogenous coordinates in the world reference frame. As shown in Fig. 4.4, each shape from single image is transformed to the world coordinates. However, due to the tracking error (less than 0.3mm) and the calibration error (less than 6 pixels), individual shapes $\hat{S}^i_w$ are not initially well matched with each other (e.g. Fig. 4.4 (c) and (d)). Since the bone is rigid, we use ICP [4] to improve the matching between different views by ensuring overlapped fields of view in adjacent images. Then, for each shape $\hat{S}^i_w$, ICP results in an aligned shape ${}^{i-1}\hat{S}^i_w$ with respect to the adjacent shape $\hat{S}^{i-1}_w$:

$${}^{i-1}\hat{S}^i_w = {}^{i-1} T_i * \hat{S}^i_w \tag{4.21}$$

where ${}^{i-1}T_i$ is the rigid transformation computed from ICP.

Using ICP, we can align all the shapes from one view with respect to the shape in the reference view, i.e. the shape $\hat{S}^1_w$ in the first image. Combining Eq. 4.20 and 4.21, we have

$${}^1\hat{S}^i_w = \prod_{j=1}^{i-1} {}^jT_{j+1} \cdot {}^w T_{m_i} \cdot [M^{-1} \cdot S^i; \mathbf{1}_{1\times N_i}] \tag{4.22}$$

where ${}^1\hat{S}^i_w$ is the aligned shape of $\hat{S}^i_w$ with respect to the reference shape $\hat{S}^1_w$, as shown in Fig. 4.5 (a) and (b).

Once the contours are aligned, we ”re-grow” the shape in the local views from the global boundaries. We compute the average on boundaries in Eg. 4.17 using the neighbors in the world coordinates. The key idea is to keep the values on the global boundary updated based on the shape developed in the local view, and therefore SFS in the local views are guided towards convergence under global boundary constraints.

We restart the SFS in the local coordinates simultaneously for all images. After initializing (z, p, q) in each image, we initialize values on the corresponding global boundary.

Table 4.3: Global Shape-from-Shading: Key Steps

*** Loop k=1:number of images**
  1) Compute SFS from single image using Eq. 4.17
  2) Track endoscope motions
  3) Transform local views to the world coordinates using Eq. 4.20
  4) Align each shape to the previous shape using ICP using Eq. 4.21
  End loop
*** Compute global boundaries using Eq. 4.22**
*** Compute global constraints (z, p, q) in the world coordinates**
*** Loop until converge**
  1)**Loop k=1:number of images**
    a) Update (z, p, q) in the local views according to the global constraints
    b) Compute new (z, p, q) by using Eq. 4.17
    End loop
  2)Update (z, p, q) in the world coordinates
  End loop

If there are two or more points from different views that are mapped to the same point in the world coordinates, (z, p, q) at those points are computed using the average. In each iteration, we grow each image and update the values on the global boundary. Compared to the single image shape-from-shading (see Fig. 4.4 (a)), we obtain more information from global boundaries (see Fig. 4.5 (c)). At the end of each iteration, we update the local (z, p, q) given the global constraints. We proceed this until the algorithm converges. With the global boundary constraints, our algorithm converges quickly although not in real-time (around 5 minutes for 18 images in Matlab with P4 2.4G CPU). Table 4.3 lists the key steps in the multi-view shape from shading.

We use a laser range scanner to build a ground truth shape of the spine. The maximum, minimum, mean and RMS errors are 2.8mm, 0.0mm, 1.24mm and 1.45mm respectively. This level of accuracy is practical for surgery. More views of the reconstructed surface compared to the ground truth are shown in Fig. 4.6. We also compare our results with the results under the orthographic projection assumption in Fig. 4.7. Compared to the real shape captured by a regular camera, we find that the shape under the orthographic projection is over grown near the boundaries.

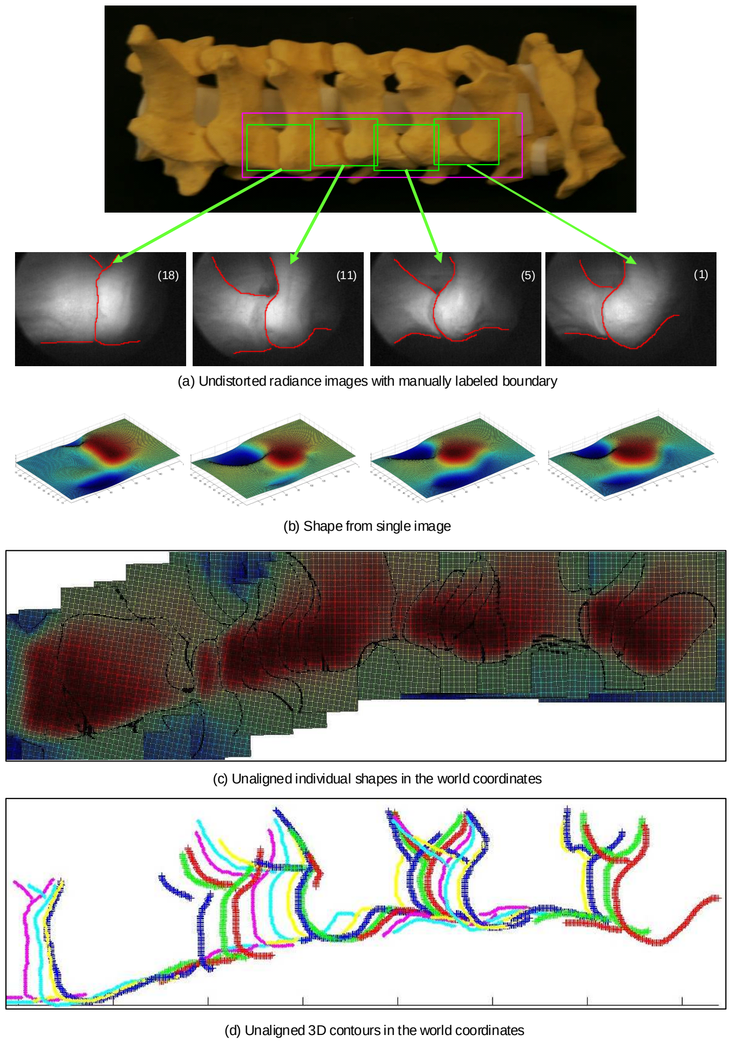


Figure 4.4: Problems of directly merging the individual shapes in the world coordinates. 18 images are captured by moving the endoscope horizontally (only translation). Four of them are shown as an example. (a) After removing the distortion and illumination effects, the boundaries in each image are labelled by hand, and the initial (p, q) are computed automatically on the boundaries. (b) Shape from each single image are reconstructed using the method described in Section 4.1. (c) Mis-aligned shapes in the world coordinates. (d) Mis-aligned 3D contours in the world coordinates.

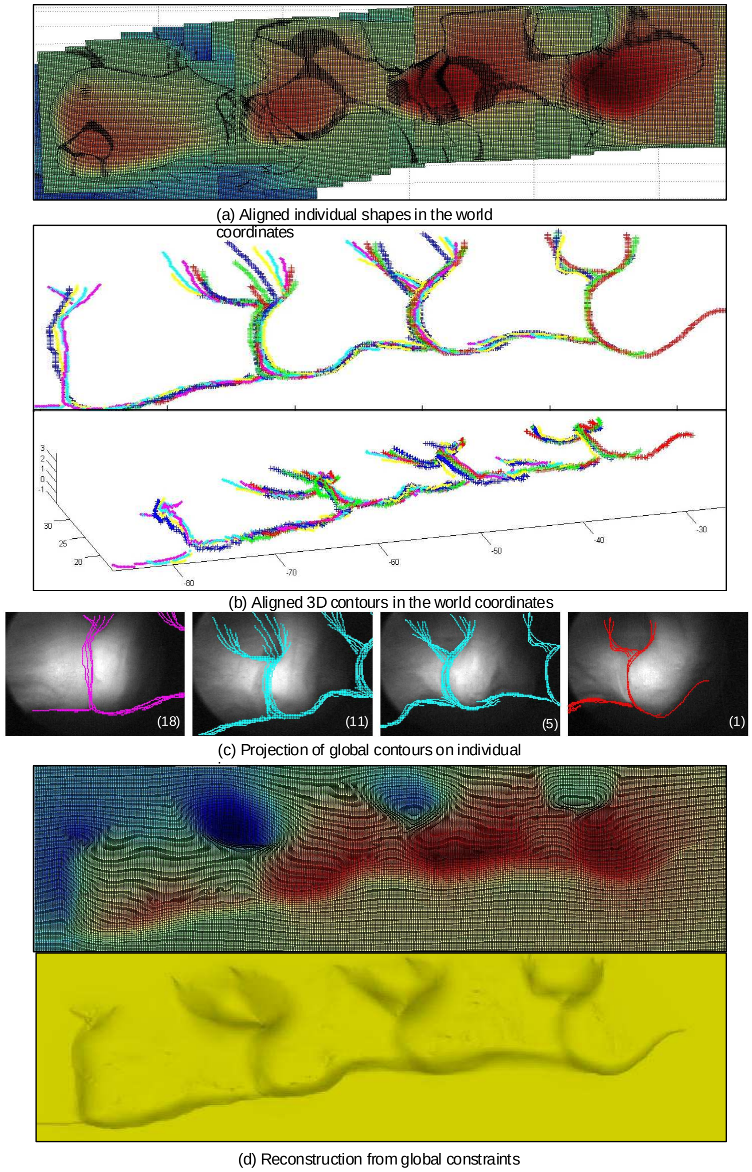


Figure 4.5: Illustration of the multi-image shape-from-shading. (a) Aligned shape in the world coordinates. (b) Aligned 3D contours in the world coordinates. (c) Projection of the global constraints onto each image. (d) Final shape reconstructed using the method described in Section 4.2.

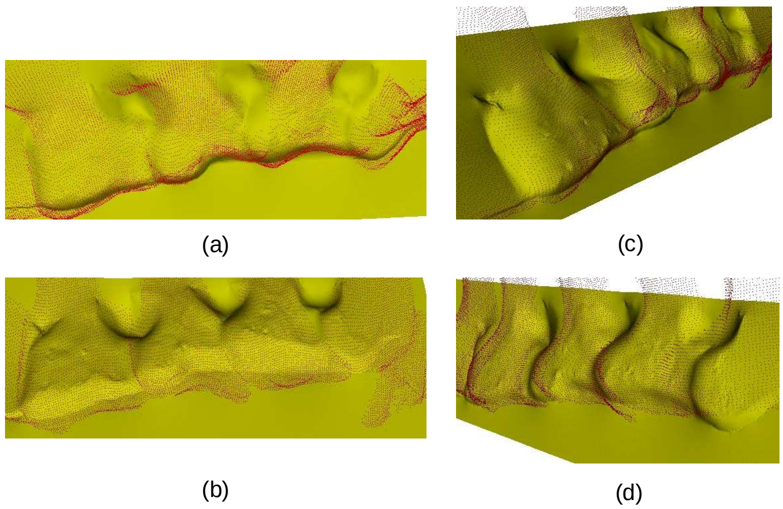


Figure 4.6: Different views of the reconstructed surface (yellow) compared to the ground truth (red). (a) top view (b) bottom view (c) left view (d) right view.

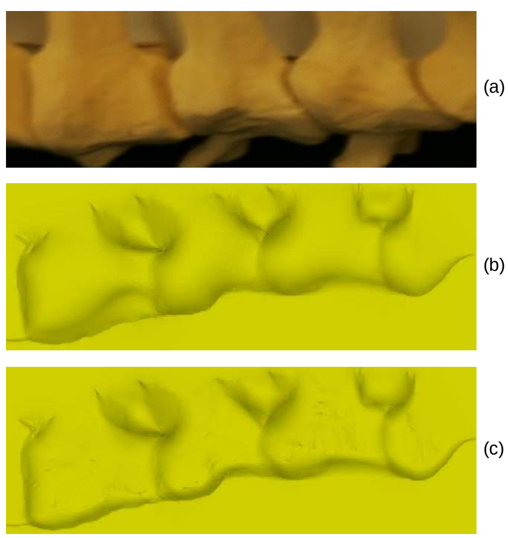


Figure 4.7: (a) Real shape captured by a regular camera. (b) Reconstructed shape under orthographic projection. (c) Reconstructed shape under perspective projection.

4.3 Conclusions and Discussion

Shape-from-shading and shape-from-motion have been successfully applied in many computer vision applications, but both have difficulties for 3D reconstruction from orthopedic endoscopy due to the featureless bone surface and partial occluding boundaries in a small field of view. In this work, we propose a method to combine the strengths of both approaches to solve our problem: we reformulate the shape-from-shading for endoscopes under near point light sources and perspective projection, and develop a multi-image shape-from-shading framework. To deal with the tracking and calibration errors, we use ICP to improve the alignment of partial shapes and contours in the world coordinates. As a result, we reconstruct the shape of a larger bone area and provide useful visualization for surgical navigation during minimally invasive procedures.

A couple of issues are worthy of further discussion. We build a near-field lighting and perspective shape-from-shading (NLPSFS) model without assuming that the light sources are located at the optical center. Our contribution is to present a new model for medical endoscopes, instead of a new solution to any existing models. We adopt the variational optimization method since it is effective for most general cases and no pre-defined values are involved for the singular and local minimum points. Since our multi-image approach is very general, other PSFS solutions such as Prados et al.'s [76] and Tankus et al.'s approach [95] can be employed to solve the NLPSFS problem as well. We plan to compare different SFS methods in the future work. In this thesis we focus on the formulation of the NLPSFS model and multi-image framework.

As mentioned in the previous section, in the real operating environment, the bone surfaces are usually 5-10 mm away from the endoscope. The light sources and camera separation distance is 3.5mm in our case. These two numbers are in the same order of magnitude as the distance to the scene. The approximation becomes inaccurate if we ignore the parameters a and b in Eq. 4.9. However, it is a reasonable approximation when the endoscope is far away from the scene. In our work we attempt to build a general and accurate model to handle more realistic cases, thus we have calibrated all parameters beforehand.

One advantage of our method is that the computation of single image SFS (see the first step of Table 1) can be parallelized since the error function in Eq. 4.15 uses only one single image. Note in the last step of Table 1, the computation of single image SFS cannot be parallelized because the global constraints need updates for each iteration.

By using ICP in our multi-image framework, we are able to deal with tracking and calibration errors. Moreover, our algorithm can tolerant some shape errors caused by the single image SFS in the first step of Table 1. Therefore we use an early stop strategy during the single image SFS by terminating the iterations before the error in Eq. 4.15 reaches zero (we set the threshold to 15% of the initial error). With this strategy, we significantly speed up the algorithm. This is one of the reasons why the result from a single image tends to be flat. Other reasons include partial occluding boundaries and the smoothness constraints.

The results from multi-image SFS (MISFS) still lack high frequency shape details due to over-smoothing. Besides, it cannot deal with occlusions due to blood and tissues [57] in surgery, and inaccurate shapes caused by endoscope rotations. To overcome above issues, we will introduce a global shape prior in the following chapters. In Ch. 5, a statistical shape prior, also known as a statistical atlas, is introduced. An efficient framework of constructing the atlas is described as well. In Ch. 6, we modify the bottom-up MISFS by enforcing the statistical atlas as a new global constraint. We also develop a top-down evaluation refinement to improve reconstruction. Please see details in the following chapters.

Chapter 5

Construction of Statistical Shape Prior for Bones from Population

The MISFS framework proposed in Ch. 4 results in shapes that are often over-smoothed, which makes it hard to apply the method to intricate structures such as spine vertebrae. This is because shape-from-shading is a fundamentally under-constrained problem and requires additional constraints to help solve it. In the absence of any other knowledge, constraints enforcing smoothness are used, often leading to shapes that are overly smoothed and lack detail. In addition, we assume that images are captured using pure translation of the endoscope. The rotation of the endoscope can result in wrong correspondences of the occluding boundaries across different views in turn leading to inaccurate reconstruction. Moreover, many substances (such as blood, pieces of bone, tissue) in surgery may obscure the bone surface and mislead any local constraints used in reconstruction.

Therefore, a global model is needed to recover the shape detail and correct the inaccurate reconstruction due to the endoscope rotations and occlusions. For example Hong et al. take the advantage of the tubular nature of the colon to design their reconstruction algorithm [39]. In this chapter, we introduce a statistical shape prior, also known as statistical atlas, for bone structures. In Ch. 6 we will present a new algorithm using this shape prior to improve the MISFS method.

Statistical atlases are used as references to interpret CT/MRI images [64], and to represent the shape or appearance of human anatomical structures [91, 90]. Some atlases are based on physical properties, such as elastic models [6, 3],"snakes" [47], geometric splines [27], and finite elements based models [70]. Others are modeled from the statistical perspective. Szeliski [92] introduced the statistical atlas to analyze the shape variation between patients. In order to analyze the shape and appearance variation, principal component analysis (PCA) is widely used [14, 65, 80]. The two most common statistical atlases are the shape atlas [16] and the appearance atlas [15]. The shape atlas uses only geometric information, such as landmarks, surfaces (boundaries of 3D objects) or crest lines [90]. The appearance atlas uses both geometric features and intensity of pixels or voxels.

As a key to build atlases, 3D registration has been studied for years in computer vision, but it still remains a critical problem in the medical image field due to the geometrical complexity of anatomical shapes, and computational complexity caused by the enormous size of volume data. It has numerous clinical applications such as statistical atlas construction for group study and statistical parameters analysis [12, 38], mapping anatomical atlases to individual patient images for disease analysis [10, 28, 46] and image segmentation [79, 89].

Depending on the type of the transformation being involved, registration can be rigid or non-rigid. If the shape has no change between the two images, registration should be rigid. For example, the intra-subject inter-modality (same patient; different imaging system) registration aligns images that are captured at the same time. However, when we take into account the time, i.e., when two images are captured at different time, such as in intra-subject registrations, most of them are non-rigid due to the shape variation of the anatomical structures caused by swelling, bone fractures, tumor growth changes, intestinal movements etc. In addition, inter-subject (different patients) registrations are usually non-rigid due to the local anatomical difference between patients. So far non-rigid (also known as deformable) registration is still a challenging problem [18].

Non-rigid registration is used to find a non-rigid transformation from one 3D surface to the reference surface by minimizing the distance between two surfaces. In general, a non-rigid transformation is represented by a global rigid or affine transformation plus a local non-linear deformation, which can be represented by radial basis functions (RBF) [113], octree-spline [93], thin-plate spline [11, 12], geometric splines [27], finite elements [70], or free form B-spline [80], etc. To evaluate the results of registration, different similarity measurements will be selected according to different image features and imaging modalities. For example, sum of squared distance (SSD) is usually used for geometric features [5], but correlation coefficients (CC) [52], Ratio Image Uniformity (RIU) [106], and mutual information (MI) [101] are used for intensity features. Registration can be simplified given known correspondences, for example, using markers [63]. Nevertheless, markers are not allowed to be used or even available in many scenarios. Alternate estimation of correspondences and transformations are widely used for both rigid cases [5] and non-rigid cases [12, 11, 33]. Moreover, with the increase of the data size and geometrical complexity, multi-resolution strategy has been adopted to the registration framework [46, 26, 85]. Sparse matrices are also used to handle the computational complexity [69].

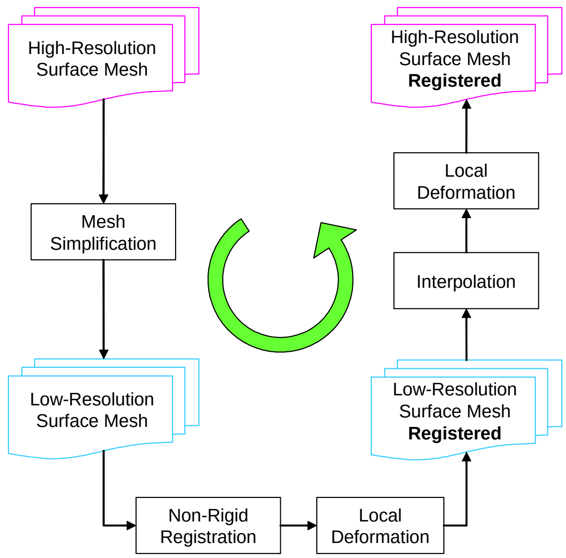


Figure 5.1: Two-level non-rigid registration framework.

In this chapter, we developed a two-level approach inspired by Chui and Rangarajan's thin-plate spline based algorithm [11] and the previous multi-resolution work [85]. Since Chui and Rangarajan's algorithm [11] is not able to handle more than 2000 3D points [69], we broke down the registration into a two-level process to deal with both computational and geometrical complexity. We first applied Chui and Rangarajan's algorithm [11] to the simplified low-resolution surfaces. To improve efficiency, instead of successively matching each resolution from coarse to fine, we directly propagated the correspondences from low resolution to high resolution by interpolation. A local refine procedure was introduced for both low-resolution and high-resolution surfaces to improve matching. Finally we applied PCA to the aligned surfaces to construct the femur atlas. Fig. 5.1 illustrates the flowchart of our two-level framework.

5.1 Two-Level Framework

5.1.1 Mesh Simplification

Garland's quadric error metrics (QEM) based mesh simplification [32] method was used to compute low-resolution surfaces. QEM is based on iterative contraction of vertex pairs. The cost of contraction is represented by a quadric error and the whole process is an iter-

Table 5.1: Table of Notation in Sec. 5.1

X^L, Y^L	-	low resolution surface
X^H, Y^H	-	high resolution surface
x_i^L, y_i^L	-	vertex on low resolution surface
x_j^H, y_j^H	-	vertex on high resolution surface
$\mathbf{c_1}$	-	coefficients for radial basis functions
$\mathbf{c_2}, \mathbf{c_3}$	-	coefficients for affine transformation
$\varphi(\cdot)$	-	symmetric radial basis function
$\mathcal{X}, \mathcal{Y}, \mathcal{Z}$	-	coordinates axes
N, N_{ref}^{low}	-	number of vertices on low resolution reference surface X^L
M, N_{ref}^{high}	-	number of vertices on high resolution reference surface X^H
S_1, S_2, S_3	-	neighboring triangles sharing the same vertex y_i
y_i^1, y_i^2, y_i^3	-	projection of x_i onto S_1, S_2, S_3
d_1, d_2, d_3	-	distance from x_i to S_1, S_2, S_3
$\mathbf{v}_i$	-	surface vector
K	-	number of training surfaces
κ	-	mean surface vector
Ψ	-	covariance matrix
$\eta_{\mathbf{i}}$	-	PCA coefficient vector
$\mathbf{U}_i$	-	eigenmatrix
$\mathbf{U}_i^S$	-	first S columns of eigenmatrix
$d(\cdot)$	-	point to surface distance
$d_m(\cdot)$	-	surface mean error
$d_{RMS}(\cdot)$	-	surface RMS error
$\bar{T}$	-	processing time

ative minimization of the quadric error. Since QEM provides a fast, simple way to guide the entire process with relatively minor storage costs, the simplification step is extremely fast.

An important parameter involved in the simplification process is the number of vertices in the low-resolution surfaces. To maintain the tradeoff between the accuracy and efficiency, a reasonable number is selected based on a series of leave-one-out experiments (See details in Sec. 5.1.6).

5.1.2 Low-Resolution Non-Rigid Registration

We applied Chui and Rangarajan's non-rigid registration method [11] to the simplified surfaces. With this method, fuzzy correspondences and a smoother optimization process can be achieved. A dual update strategy conjuncted with a deterministic annealing technique is adopted to estimate the correspondences and transformation alternately. The non-rigid transformation is parameterized using thin-plate splines to generate a smooth spatial mapping.

5.1.3 Low-Resolution to High-Resolution Interpolation

To improve the efficiency, instead of successively matching each resolution from coarse to fine, we directly propagate the correspondences from low resolution (X^L and Y^L) to high resolution (X^H and Y^H) by interpolation. The surface interpolation method is a derivative of methods known jointly as "moving least squares" [112]. Radial basis functions (RBF), finite element, multivariate spline such as thin-plate spline (2D bivariate spline) and tri-harmonic thin-plate spline, are popular techniques used for surface interpolation. Carr et al. [9] applied multivariate splines method into radial basis functions by using splines as kernel functions. In this work we chose Gaussian kernel due to its simple mathematical representation and less restrictions on nodes [9]. More specifically, we used a linear affine function plus a series of radial basis functions (RBFs) to construct the interpolation function:

$$\mathbf{y}_i^L = g(\mathbf{x}_i^L) = \underbrace{\mathbf{c_1} \cdot [\varphi(\|\mathbf{x}_i^L, \mathbf{x}_1^L\|), \cdots, \varphi(\|\mathbf{x}_i^L, \mathbf{x}_N^L\|)]' + \mathbf{c_2} + \mathbf{c_3} \cdot \mathbf{x}_i^L}_{g(\mathbf{x}_i^L)} \tag{5.1}$$

where $\mathbf{x}_i^L$ is a vertex on the low-resolution surface X^L, whose correspondence on the low-resolution surface Y^L is $\mathbf{y}_i^L$, $i = 1, 2, \cdots, N$ (N is the number of vertices on X^L). $\mathbf{x}_i^L$ and $\mathbf{y}_i^L$ are both 3×1 vectors with three coordinates. $\mathbf{c_1}$ is a $3 \times N$ coefficient matrix of radial basis functions. $\varphi(\|\mathbf{x}_i^L, \mathbf{x}_j^L\|)$ is a symmetric radial basis function. We selected a Gaussian kernel $\varphi(\mathbf{u_i}, \mathbf{u_j}) = \exp(-\|\mathbf{u_i} - \mathbf{u_j}\|/0.5)$, as suggested by [72]. $\mathbf{c_2}$ and $\mathbf{c_3}$ are coefficients for the affine transformation. $\mathbf{c_2}$ is a 3×1 vector and $\mathbf{c_3}$ is a 3×3 matrix. Given

N correspondences, there are N equations for each axis ($\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$):

$$\underbrace{\begin{bmatrix} \varphi(\mathbf{x}_1^L, \mathbf{x}_1^L) & \cdots & \varphi(\mathbf{x}_1^L, \mathbf{x}_N^L) & 1 & \mathbf{x}_1^{L^T} \\ & & \vdots & & \\ \varphi(\mathbf{x}_N^L, \mathbf{x}_1^L) & \cdots & \varphi(\mathbf{x}_N^L, \mathbf{x}_N^L) & 1 & \mathbf{x}_N^{L^T} \end{bmatrix}}_{\mathbf{P}^k} \cdot \underbrace{\begin{bmatrix} \mathbf{c_1^k}^{\mathbf{T}} \\ \mathbf{c_2^k}^{\mathbf{T}} \\ \mathbf{c_3^k}^{\mathbf{T}} \end{bmatrix}}_{\mathbf{c}^k} = \underbrace{\begin{bmatrix} \mathbf{y}_1^{L^k} \\ \vdots \\ \mathbf{y}_N^{L^k} \end{bmatrix}}_{\mathbf{yv}^k} \tag{5.2}$$

where $\mathbf{c_1}^k$, $\mathbf{c_2}^k$ and $\mathbf{c_3}^k$ denote the k^{th} row of $\mathbf{c_1}$, $\mathbf{c_2}$ and $\mathbf{c_3}$, respectively. $\mathbf{y}_i^{L^k}$ denotes the k^{th} row of $\mathbf{y}_i^L$, k can be 1,2 or 3, corresponding to the axis $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$. There are $3N$ equations in total:

$$\mathbf{P} = [\mathbf{P}^1, \mathbf{P}^2, \mathbf{P}^3]', \mathbf{c} = [\mathbf{c}^1, \mathbf{c}^2, \mathbf{c}^3]', \mathbf{y} = [\mathbf{y}^1, \mathbf{y}^2, \mathbf{y}^3]' \tag{5.3}$$

$\mathbf{P}$ is a $3N \times (N+4)$ matrix. In order to ensure smooth interpolation, additional orthogonality constraints $\sum_i \mathbf{x}_i^{L^T} \mathbf{c_{1,i}} = 0$ [8] were added to Eq. 5.2, where $\mathbf{c_{1,i}}$ denotes the i^{th} column of $\mathbf{c_1}$:

$$\underbrace{\begin{bmatrix} & & \mathbf{P} & & \\ \mathbf{x}_1^L & \mathbf{x}_2^L & \cdots & \mathbf{x}_N^L & \mathbf{0}_{4\times4} \end{bmatrix}}_{\mathbf{Q}} \cdot \ \mathbf{c} \ = \underbrace{\begin{bmatrix} \mathbf{y} \\ \mathbf{0}_{4\times1} \end{bmatrix}}_{\mathbf{w}} \tag{5.4}$$

The least-squares solution for this linear system, $\mathbf{Qc} = \mathbf{w}$, is given by $\mathbf{c} = (\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{w}$.

Finally, the correspondence of a vertex $\mathbf{x}_j^H$ in the high-resolution surface X^H is computed by Eq. 5.1: $\mathbf{y}_j^H = g(\mathbf{x}_j^H)$, for $j = 1, \cdots, M$ (M is the number of vertices on X^H).

5.1.4 Refining Registration

For both low-resolution and high-resolution surfaces, we applied a refining procedure to improve accuracy. The refining procedure includes both local and global steps.

Local Refining

The registration results can be further improved by minimizing the point-to-surface distances. This idea is illustrated by Fig. 5.2. $\mathbf{x}_i$ is a vertex on the deformed surface X, whose corresponding vertex on the surface Y is $\mathbf{y}_i$. The neighboring triangles which share the same vertex $\mathbf{y}_i$ are S_1,S_2 and S_3. We can compute the distance from $\mathbf{x}_i$ to each neighboring triangle (the distance computed from the vertex $\mathbf{x}_i$ to the plane where the triangle

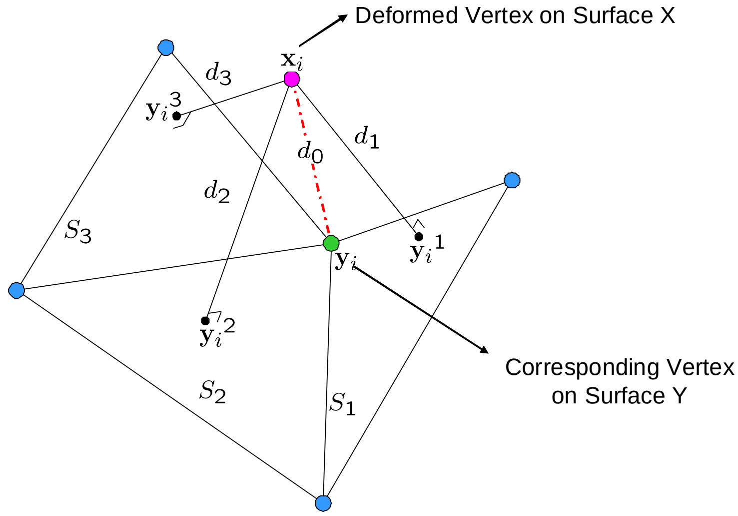


Figure 5.2: Illustration of the local refine procedure between the vertex and surface. $\mathbf{x}_i$ is a vertex on the deformed surface X, whose correspondence on the surface Y is $\mathbf{y}_i$. The projection of $\mathbf{x}_i$ to each triangle S_t sharing $\mathbf{y}_i$ is denoted by ${\mathbf{y}_i}^t$. d_t is the distance between $\mathbf{x}_i$ and ${\mathbf{y}_i}^t$. $(t = 1, 2, 3, \cdots)$

lies), i.e. d_1, d_2 and d_3. If any of them is smaller than $d_0 = \|\mathbf{x}_i - \mathbf{y}_i\|$, we will use the corresponding projected surface point to replace $\mathbf{y}_i$ to achieve a smaller surface distance.

For cases where different vertices on the surface X correspond to the same surface point on Y , we will assign this corresponding surface point to the vertex on X with the smallest distance and mark it unavailable to other vertices on X.

Global Refining

For training surfaces with sparse points, we cannot directly apply the local refining procedure since no high-resolution surface can be used as a target shape, for example, the spine data. Compared to bones such as femur and knee, spine vertebrae have complex shapes which make it hard to automatically segment from CT images (using for example, marching cubes algorithm). On the other hand it is tedious to manually label over ten thousand points to obtain high-resolution surfaces. Given some shape information, Kaus et al. [49] use learned deformable models to generate aligned 3D meshes. In our case without knowing any shape prior, we propose a semi-automatic strategy by taking advantage of the two-level framework to generate high-resolution surfaces from sparse labeled points. Given a reasonable amount of manual labeling, for instance 250 points per model, we are able to combine surface segmentation and registration in the same procedure, and reduce manual work to a minimal.

The sparse points set is first used as the low resolution input for the two-level framework. In the final refining step, instead of the local refining, we propose a global refining which uses the sparse labeled points as a target shape. We apply a RBF based warping to refine the high-resolution surface towards the original labeled points. We employ the same RBF functions used for the low-resolution to high-resolution interpolation (see Sec. 5.1.3) but different sigma for the RBF kernel. Here we choose σ as 250. The larger σ represents stronger global deformation.

5.1.5 Atlas Construction

Given aligned high-resolution surfaces, rigid pose alignment is applied to eliminate the effect of imaging poses [34] prior to atlas construction. Suppose we have K aligned surfaces and each surface is represented by a $3M \times 1$ vector $\mathbf{v}_i (i = 1, \cdots, K)$, where M is the number of vertices on each surface and each vertex has 3 components along $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ axes. We compute the mean vector κ and covariance matrix Ψ, and then apply PCA to find the low dimensional representation of the data:

$$\kappa = \frac{1}{K} \sum \mathbf{v},\ \Psi = \frac{1}{M-1} [\mathbf{v}^1 - \kappa, \cdots, \mathbf{v}^K - \kappa] \cdot [\mathbf{v}^1 - \kappa, \cdots, \mathbf{v}^K - \kappa]' \tag{5.5}$$

$$\Psi = \mathbf{U} \Lambda \mathbf{U}' \tag{5.6}$$

Therefore, a compact representation of any surface model is given by a mean vector plus a linear combination of principal components (modes):

$$\mathbf{v}_i = \kappa + \mathbf{U} \boldsymbol{\eta_i} \tag{5.7}$$

where $\boldsymbol{\eta_i}$ is a $K \times 1$ coefficient vector obtained by projecting $\mathbf{v}_i$ onto each principal axis. New surface models, not included in the data set, can be generated by manipulating the coefficient $\boldsymbol{\eta_i}$.

Without losing generality, femur and spine data have been used to demonstrate our method. Given a big number of femur data, we are able to study the parameters used to improve the algorithm. The spine atlas is later used to enhance the MISFS method (see details in Ch. 6).

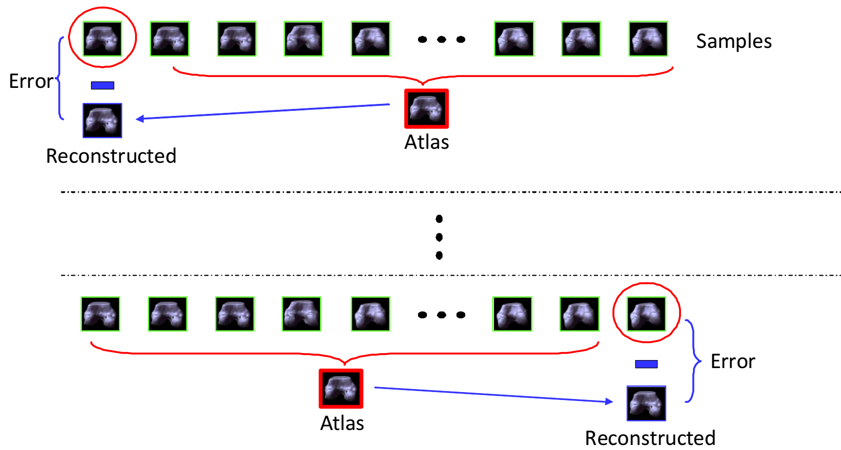


Figure 5.3: Leave one out experiment.

5.1.6 Selection of Simplification Parameters

In the two-level framework, the resolution of the simplified surfaces affects the final result. The fewer the number of points, the faster low-resolution registration is achieved but results in high-resolution registration are less accurate. To maintain both accuracy and efficiency, an appropriate number is selected based on a series of leave-one-out experiments. For some extreme cases with symmetry shapes, the fewer points may result in a perfect match in the low resolution yet a misalignment in the high-resolution. Thereby not only the resolution parameter but also some landmarks are needed to handle such ambiguity. In our work, since shapes of the femur and spine vertebra are complex enough, we do not consider such extreme situations. Thus we do not require any landmarks in our method.

Let N_{ref}^{low} denote the number of vertices in the low-resolution surfaces. As Fig. 5.3 shows,

for each surface $\mathbf{v}_i (i = 1, \cdots, K), K = 87$, we used other $K - 1$ surfaces to construct the atlas using Eq. 5.5 and 5.6. Let $\mathbf{U}_i^S$ denote the first S columns of the principal component matrix $\mathbf{U}_i$, which consists of 95% of the shape variation. Then $\mathbf{v}_i (i = 1, \cdots, K)$ can be reconstructed by this atlas:

$$\widetilde{\mathbf{v}}_i = \kappa + \mathbf{U}_i^S \mathbf{U}_i^{S^T} (\mathbf{v}_i - \kappa) \tag{5.8}$$

We compared surface distance between the original surface v_i and the reconstructed surface $\widetilde{v}_i$ by computing the mean error and root mean square error. We repeated this procedure for each surface and computed the averaged mean error and RMS error based on K

leave-one-out experiments:

$$\overline{d^t_m} = \frac{1}{K}\sum_i d_m(v_i, \widetilde{v_i}),\ \overline{d^t_{RMS}} = \frac{1}{K}\sum_i d_{RMS}(v_i, \widetilde{v_i}) \quad (5.9)$$

By tuning the number N^{low}_{ref}, we compared the error $\overline{d^t_m}$ and $\overline{d^t_m}$, and processing time $\overline{T^t}$ for the two-level registration as well. Fig. 5.4 shows when $N^{low}_{ref} \geq 0.2\% N^{high}_{ref}$, $\overline{d^t_m}$ will be less than 1mm, which is a practical number for clinical applications. Fig. 5.5 shows when $N^{low}_{ref} \geq 0.3\% N^{high}_{ref}$, the averaged processing time of the two-level registration will exceed 5 mins (2.4GHz Pentium PC with 1GB RAM). So $N^{low}_{ref} = 0.2\% N^{high}_{ref}$ was finally used in our algorithm to build atlases and estimate the error distributions.

5.2 Experiments and Results

5.2.1 Evaluation of Two-level Registration

We use the bottom portion of the femur as an example given its important relationship with the knee. Fig. 5.6 shows an example that how the surface distance is decreased in each step of the two-level framework, and visual results are shown in Figs. 5.7-5.12. In this example we have two high-resolution surfaces X^H (21130 vertices, 42256 triangles, 65.8 mm in z-axis) and Y^H (26652 vertices, 53300 triangles, 105.9 mm in z-axis) (Patient X is a 79 years old female, her femur length is 472.6 mm; Patient Y is a 53 years old female, his femur length is 477.6 mm). We first computed point-to-surface distance from X^H to Y^H [1]:

$$d(p, Y^H) = \min_{p' \in Y^H} \|p - p'\|_2, p \in X^H \quad (5.10)$$

where $\|\cdot\|_2$ is Euclidean norm. The HSV color (HSV stands for hue, saturation, value) of each vertex on X^H denotes the distance $d(p, Y^H)$. We also computed the mean error $d_m(X^H, Y^H)$ and root mean square (RMS) error $d_{RMS}(X^H, Y^H)$ between X^H and Y^H based on Eq. 5.10:

$$\begin{aligned} d_m(X^H, Y^H) &= \tfrac{1}{|X^H|}\textstyle\sum_{p\in X^H} d(p, Y^H) dX^H \\ d_{RMS}(X^H, Y^H) &= \sqrt{\tfrac{1}{|X^H|}\textstyle\sum_{p\in X^H} d(p, Y^H)^2 dX^H} \end{aligned} \quad (5.11)$$

1. Fig. 5.7 shows the high-resolution surfaces X^H and Y^H. With respect to the bounding box diagonal of Y^H (158.5mm), the mean error is 6.49% and root mean square error is 7.70%.

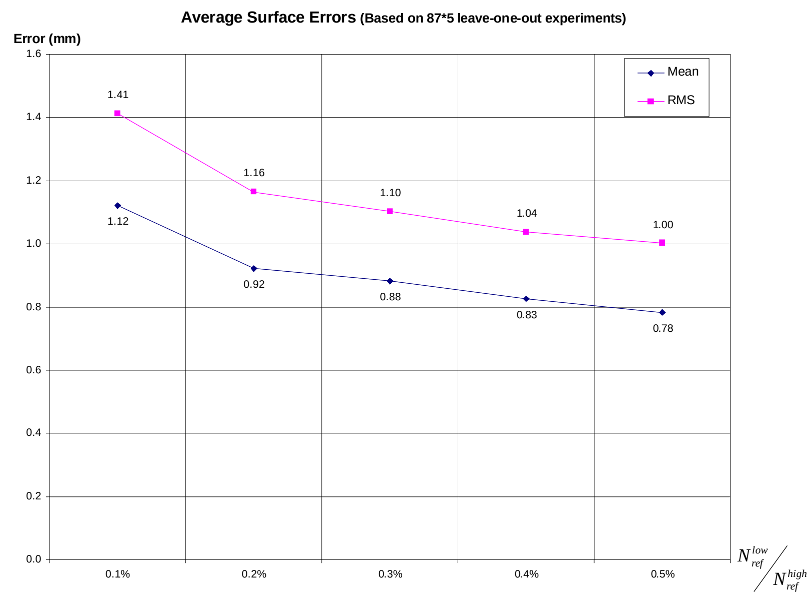


Figure 5.4: Illustration of the reconstruction errors $\overline{d_m}$ and $\overline{d_{RMS}}$ in terms of $N_{ref}^{low}/N_{ref}^{high}$.

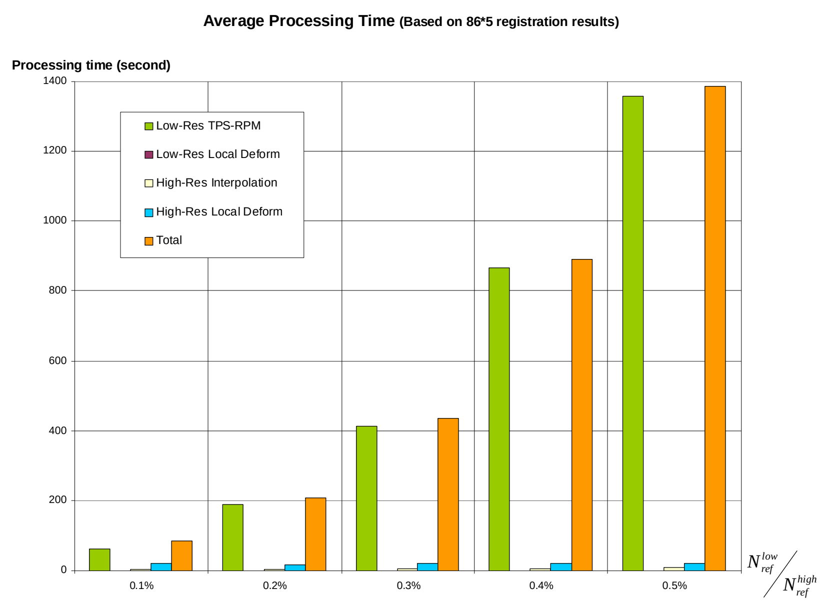


Figure 5.5: Illustration of the processing time $\overline{T}$ in terms of $N_{ref}^{low}/N_{ref}^{high}$.

Figure 5.6: Illustration of how the surface distance is decreased in each step of the two-level framework.

2. Fig. 5.8 shows the low-resolution surfaces X^L (169 vertices, 334 triangles) and Y^L (213 vertices, 422 triangles) after simplification. With respect to the bounding box diagonal of Y^L (158.3mm), the mean error is 6.53% and root mean square error is 7.74%.

3. Fig. 5.9 shows the deformed low-resolution surfaces $X^{L(1)}$ and Y^L after applying Chui and Rangarajan's non-rigid registration [11] to X^L. With respect to the bounding box diagonal of Y^L, the mean error is 1.68% and root mean square error is 2.13%. The surface distance has been significantly decreased by Chui and Rangarajan's non-rigid registration method [11].

4. Fig. 5.10 shows the deformed low-resolution surfaces $X^{L(2)}$ and Y^L after applying a local refining process to $X^{L(1)}$. With respect to the bounding box diagonal of Y^L, the mean error is 0.68% and root mean square error is 1.42%, which demonstrates that local point-to-surface refinement can decrease surface distance further.

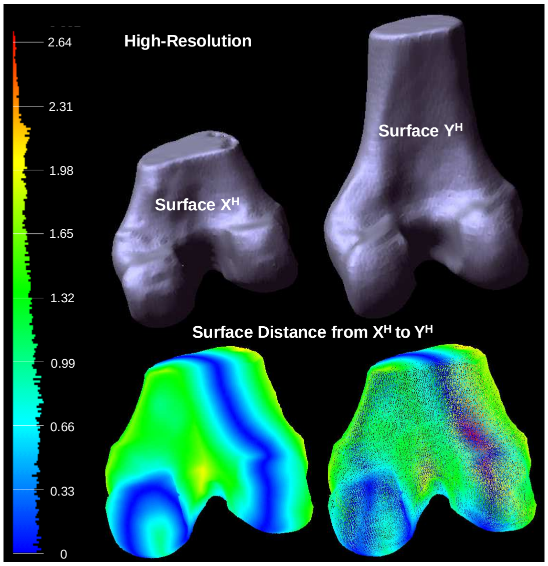


Figure 5.7: Illustration of the high-resolution surfaces X^H and Y^H. Point-to-surface distances from X^H to Y^H are illustrated by a HSV color map: the color of each vertex on X^H represents the distance $d(p, Y^H), p \in X^H$. (Both surface (left) and mesh (right) are showed in the bottom row.)

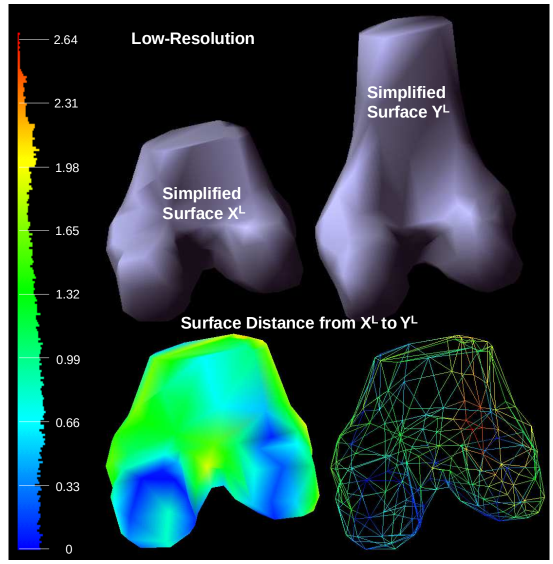


Figure 5.8: Illustration of the low-resolution surfaces X^L and Y^L after applying the mesh simplification (see Sec. 5.1.1) to both X^H and Y^H. Point-to-surface distances from X^L to Y^L are illustrated by a HSV color map: the color of each vertex on X^L represents the distance $d(p, Y^L), p \in X^L$. (Both surface (left) and mesh (right) display modes are showed in the bottom row.)

5. Fig. 5.11 shows the interpolated high-resolution surfaces $X^{H(1)}$ and Y^H after applying the interpolation to X^H. With respect to the bounding box diagonal of Y^H, the mean error is 1.65% and root mean square error is 2.10%. The reason why the surface distance slightly increases by interpolation is: only 0.80% of vertices on $X^{H(1)}$ have correspondences computed from low-resolution registration, other correspondences were obtained by interpolation.

6. Fig. 5.12 shows the deformed high-resolution surfaces $X^{H(2)}$ and Y^H after applying a local refining process to $X^{H(1)}$. With respect to the bounding box diagonal of Y^H, the mean error is 0.28% and root mean square error is 1.26%, which once again demonstrates that local point-to-surface refining procedure is helpful for decreasing surface distance.

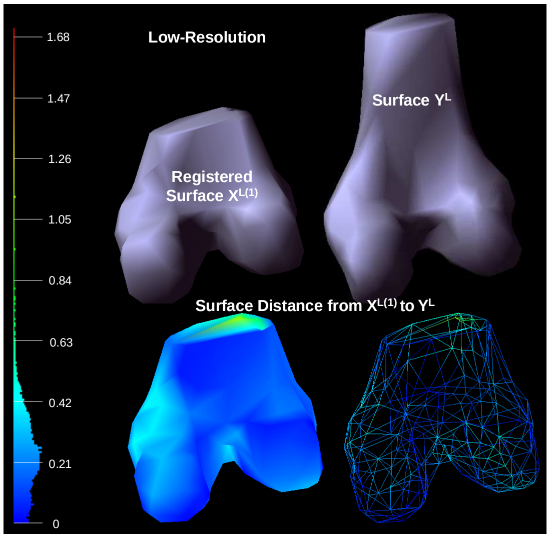


Figure 5.9: Illustration of the deformed low-resolution surfaces $X^{L(1)}$ and Y^L after applying Chui and Rangarajan's non-rigid registration method (see Sec. 5.1.2) to X^L. Point-to-surface distances from $X^{L(1)}$ to Y^L are illustrated by a HSV color map: the color of each vertex on $X^{L(1)}$ represents the distance $d(p, Y^L), p \in X^{L(1)}$. (Both surface (left) and mesh (right) display modes are showed in the bottom row.)

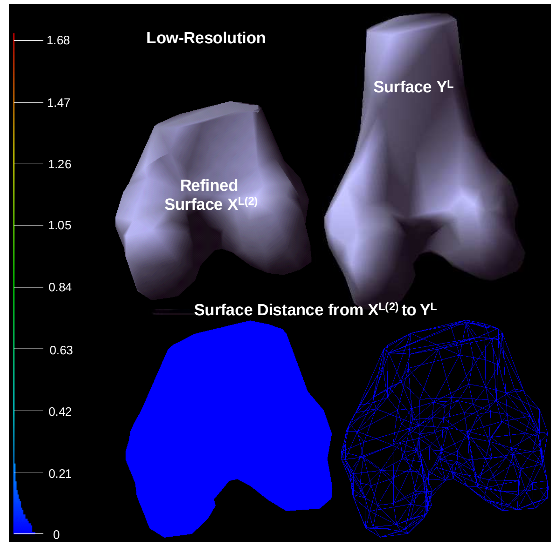


Figure 5.10: Illustration of the deformed low-resolution surfaces $X^{L(2)}$ and Y^L after applying the refinement (see Sec. 5.1.4) to $X^{L(1)}$. Point-to-surface distances from $X^{L(2)}$ to Y^L are illustrated by a HSV color map: the color of each vertex on $X^{L(2)}$ represents the distance $d(p, Y^L), p \in X^{L(2)}$. (Both surface (left) and mesh (right) display modes are showed in the bottom row.)

5.2.2 Comparison to Other Registration Methods based on Femur Data

We compared our registration algorithm with the classical iterative closet point (ICP) method since it is widely used in many medical image registration problems. Given 87 training surfaces, the distribution of RMS error and mean error of the two approaches are showed in Fig. 5.13. For RMS error, they are centered between 0.6-1.0mm (the maximum is 1.4mm) in our method, but in ICP the range is between 3.1-6.5mm (the maximum is 8.5mm). For the mean error, they are centered between 0.5-0.7mm (the maximum is 0.9mm) in our method, but in ICP the range is 2.5-6.5mm (the maximum is 7.5mm). It is obvious that the two-level non-rigid registration has a better performance than ICP does.

Since Chui and Rangarajan's thin-plate spine (TPS) based method [11] cannot deal with the complexity of our data, we cannot directly compare two methods. Alternatively,

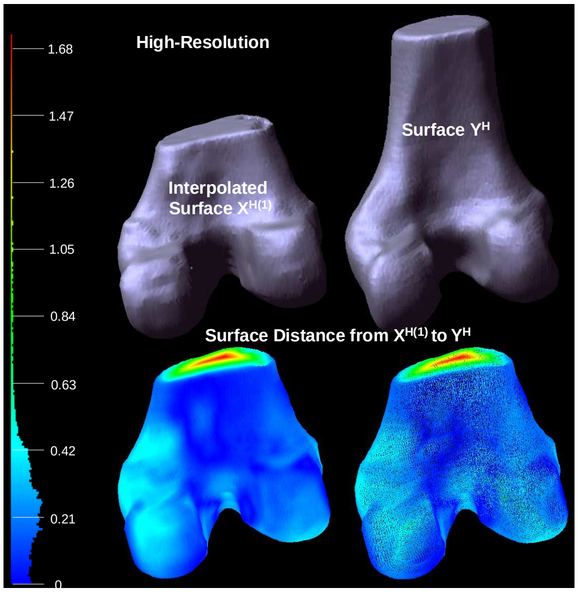


Figure 5.11: Illustration of the deformed high-resolution surfaces $X^{H(1)}$ and Y^H after applying interpolation (see Sec. 5.1.3) to X^H. Point-to-surface distances from $X^{H(1)}$ to Y^H is illustrated by a HSV color map: the color of each vertex on $X^{H(1)}$ represents the distance $d(p, Y^H), p \in X^{H(1)}$. (Both surface (left) and mesh (right) display modes are showed in the bottom row.)

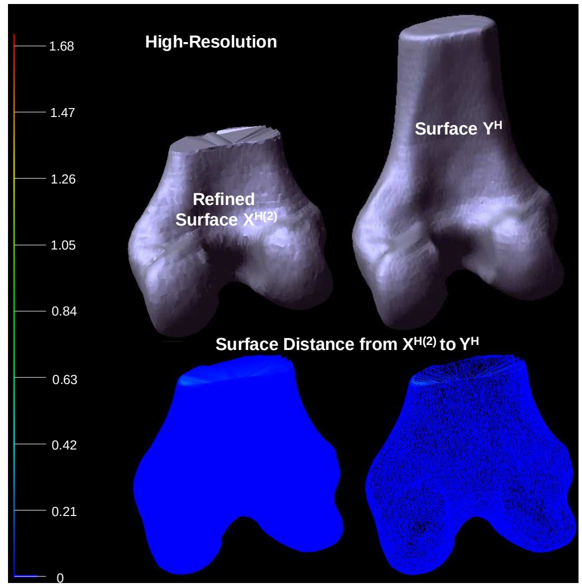


Figure 5.12: Illustration of the deformed high-resolution surfaces $X^{H(2)}$ and Y^H after applying the refinement (see Sec. 5.1.4) to $X^{H(1)}$. Point-to-surface distances from $X^{H(2)}$ to Y^H is illustrated by a HSV color map: the color of each vertex on $X^{H(2)}$ represents the distance $d(p, Y^H), p \in X^{H(2)}$. (Both surface (left) and mesh (right) display modes are showed in the bottom row.)

we have noticed from Fig. 5.5 that our method only needs 5 minutes or less to match any size of surfaces with less than 200,000 vertices. However, Chui and Rangarajan's method costs 5 minutes for 350 vertices, 10 minutes for 460 vertices, 20 minutes for 610 vertices, etc. It is obvious that our algorithm significantly improves the efficiency. To maintain a good accuracy for practical applications, a best simplification parameter was selected based on a series of leave-one-out experiments (see details in Sec. 5.1.6).

5.2.3 Femur Atlas

We have collected CT scans of 87 different patients all with one healthy femur. The scans were acquired at the West Pennsylvania Hospital (Pittsburgh, PA) and UPMC Shadyside Hospital (Pittsburgh, PA) over a period of 10 years, and annonymized before use. There

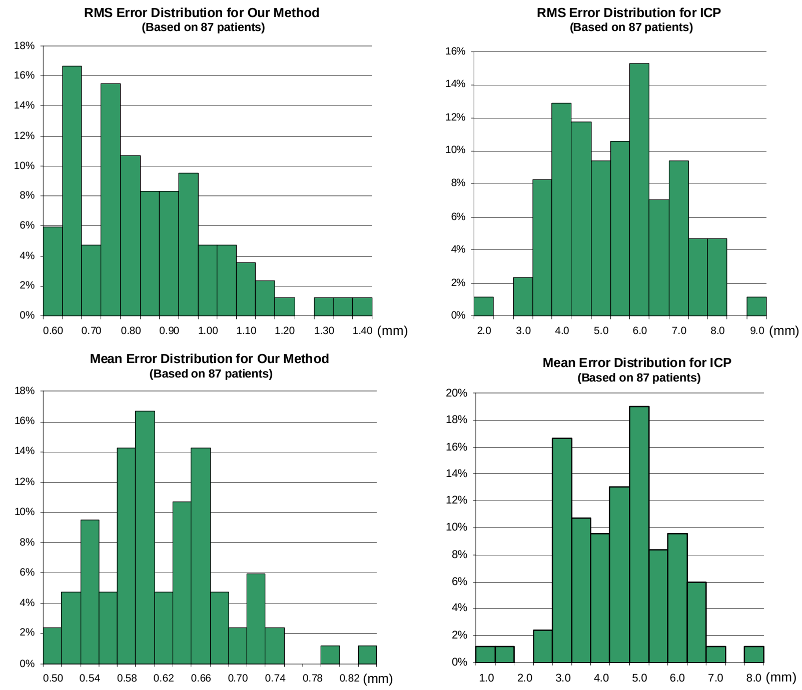


Figure 5.13: RMS and Mean error distributions ICP and our method.

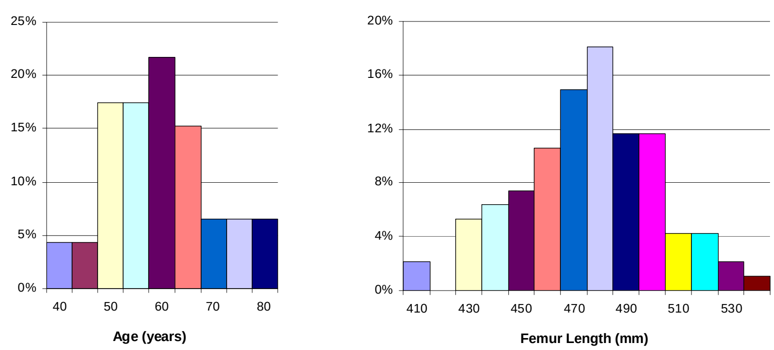


Figure 5.14: Data distributions in terms of age (left) and femur length (right).

are 53 males and 34 females. 43 are left femurs and 44 are right ones. The patients' age ranges from 39 to 78 and their femur length ranges from 400mm to 540mm (Fig. 5.14 shows the data distribution in terms of the age and femur length). The CT volumes were segmented to provide triangulated surface models using Marching Cube (MC) algorithm. All surface models were smoothed by the method proposed in [20]. The scans were ac-

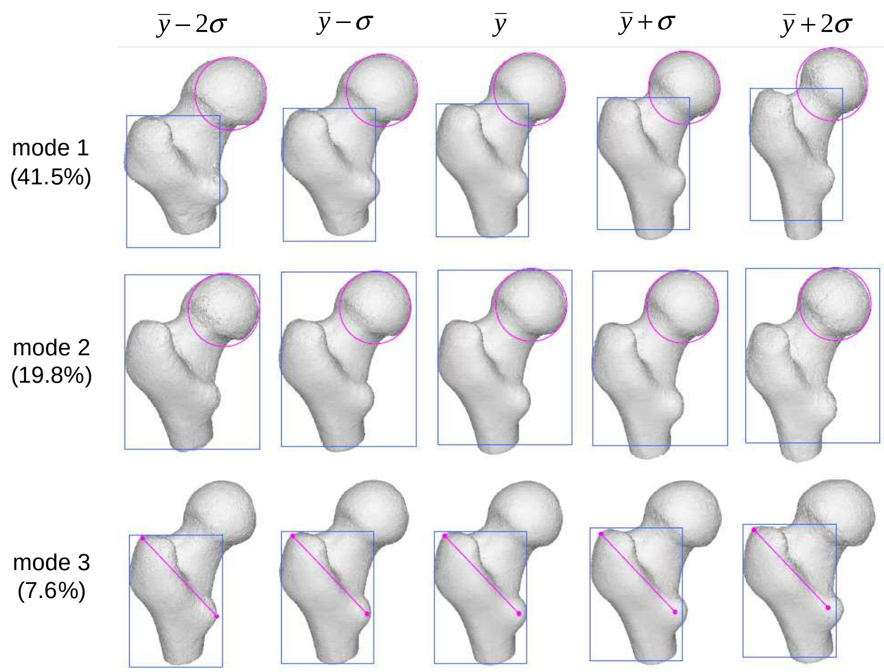


Figure 5.15: Femoral head atlas: illustration of shape variation in the first, second and third mode. The first mode encodes 41.50% of shape variation, the second and third modes encode 19.78% and 7.64% of shape variation. The color shapes highlight the local shape variation.

quired under a protocol for computer assisted total hip replacement, and contained only the proximal portion of the femur including the femoral head and the distal portion including the condyles. Thereby each femur model has two seperate surfaces containing the femoral head and condyles, respectively. Given the particularity of the femur data, we need to do a pre-alignment to get rid of the effect caused by the missing portions. Please see details in Appendix C. Experiments show that the rate of convergence has been improved from 78% to 95.2% by applying such an alignment.

Fig. 5.15 shows shape variation in the first three modes of the femoral head atlas. The color shapes (pink and blue) are used to highlight the local variation. For example, the size of the femoral head (denoted by the pink circle) gradually reduces along the first mode, but remains almost the same along the second mode. The greater trochanter (denoted by the blue area) becomes longer but narrower along the first mode. It remains the same length but becomes wider along the third mode.

Fig. 5.16 shows shape variation in the first three modes of the condyles atlas. We noticed that the top of the condyles (denoted by the blue area) changes a lot along each

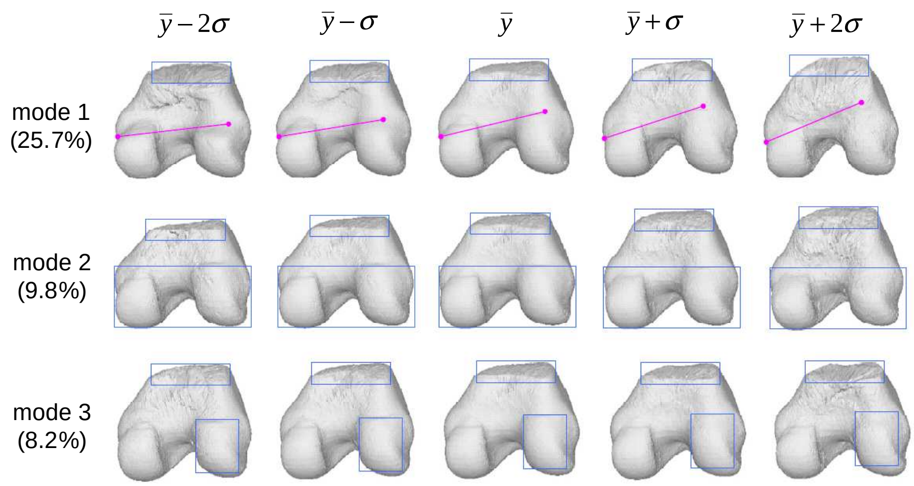


Figure 5.16: Condyles atlas: illustration of shape variation in the first, second and third mode. The first mode encodes 25.70% of shape variation, the second and third modes encode 9.75% and 8.21% of shape variation. The color shapes highlight the local shape variation.

mode. This is because each scan was cut differently. The top area can only be estimated from the reference surface without any refining from the original surfaces. The lateral epicondyle (denoted by the blue point on the right) becomes larger along the first mode, compared to the medial epicondyle (denoted by the blue point on the left). It is similar to the third mode but the change is smaller. The condyles become longer but narrower along the second mode.

Figs. 5.17-5.18 show shape variation in the first two modes of the entire femur atlas. For the first mode, the femoral head and greater trochanter follow the similar change as Fig. 5.15 shows, and the condyles become longer and narrower as Fig. 5.16 shows. The atlas of the entire femur behaves in a similar way as the individual portion does for the first mode. For the second mode, the greater trochanter becomes shorter but it remains the same as in Fig. 5.15. The condyles become much longer but they keep the same length as in Fig. 5.16. It means that the second dominant mode in the individual atlas may not be the second dominant mode in the conjunction case.

Fig. 5.19 shows the shape variation encoded in each mode of three atlases. To keep 95% of shape variation we only need 26 modes for the femoral head atlas, 49 modes for the condyles atlas and 20 modes for the entire femur atlas. The reason why the condyles atlas needs more modes to keep most of shape variation is because the training surfaces of the bottom portion of the femur change a lot in shape (i.e. Each surface contains different length of femur shaft).

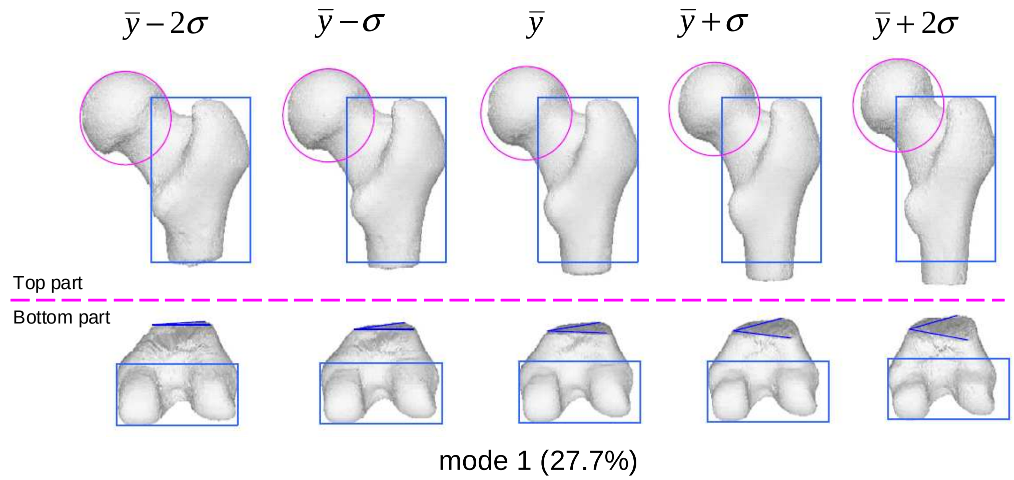


Figure 5.17: Femur atlas: 27.67% of shape variation is encoded in the first mode. The color shapes highlight the local shape variation.

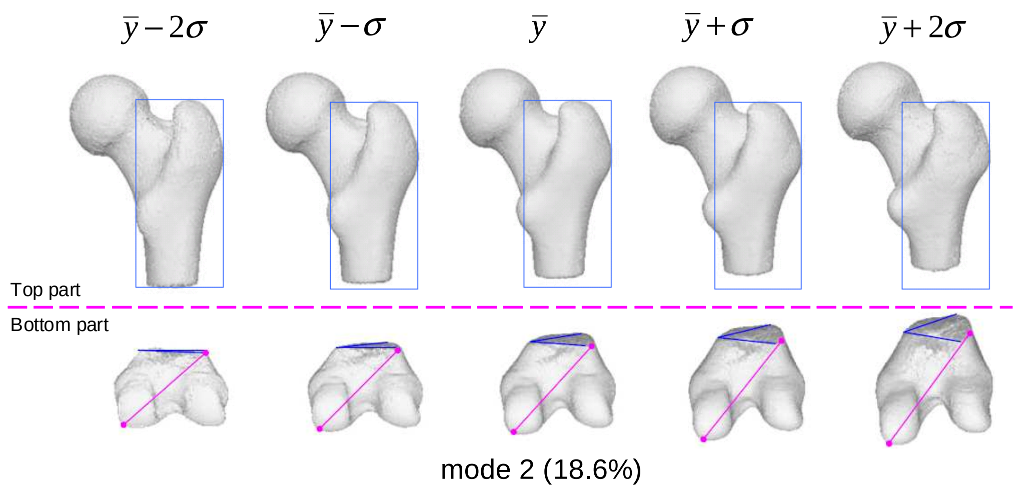


Figure 5.18: Femur atlas: 18.63% of shape variation is encoded in the second mode. The color shapes highlight the local shape variation.

5.2.4 Spine Vertebra Atlas

In our endoscopic reconstruction experiments we use lumbar vertebrae. Lumbar vertebrae are the largest segments of the movable part of the vertebral column. We select all five lumbar vertebrae (L1-L5) and the last thoracic vertebra (T12) for our study.

We acquired CT scans of three cardavers at the University of Pittsburgh Medical Center

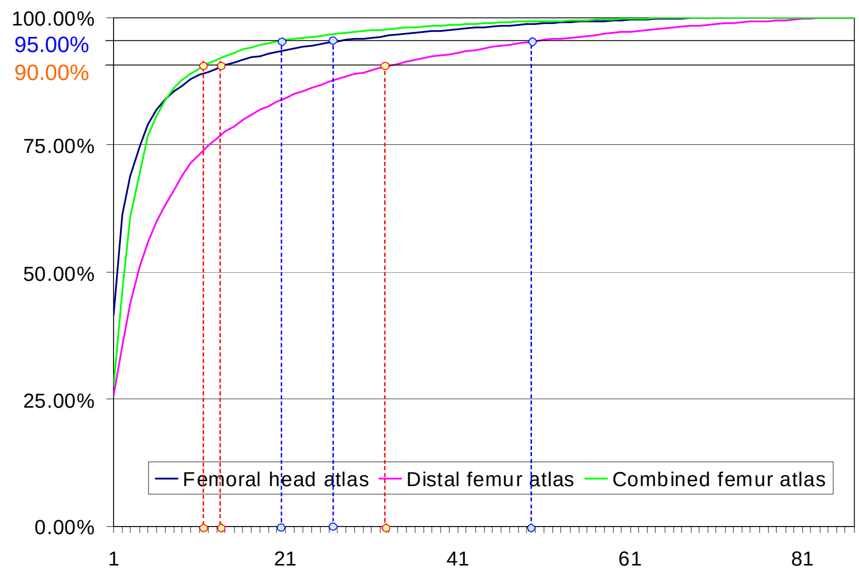


Figure 5.19: Illustration of the shape variation encoded in modes of the femoral head atlas(blue), condyles atlas(pink) and entire femur atlas(green).

(UPMC) and collected the data in the DICOM format (GE Medical Systems; DICOM format; LightSpeed QX/I; SliceThickness=1.2500; Width = 182; Height = 177; BitDepth = 16; PixelSpacing = [2x1 double]; Protocol Name = 8.1 Pelvis for Fracture). To enlarge the dataset, we have also scanned three human-size artificial spines by the same protocol. We used the solid white plastic sawbones model (Pacific Research Laboratories, Inc., # 1352-31) made of rigid plastic. The shape of the sawbones models are cast from different natural real specimens (See Appendix D for our collection of spine images).

We have collected a high-resolution 3D surface of vertebra generated by triangulating 40682 points which were carefully selected by hand. This step takes several hours but only need to be done once. We trimmed each spine CT images into small files with individual vertebra. We manually selected about 250 surface points (of the same resolution as the simplified reference surface) for each DICOM file. It took 10-15 minutes to label each vertebra. Basically we selected around 8-9 points on each image slice (12 slices from the coronal view, 12 slides from the sagittal view and 6 slices from the transverse view). Examples about selecting points are shown in Appendix E. The selection is assumed to be consistent since it was conducted by the same person based on the study of the structure of the spine vertebrae and DICOM images. However, no specific anatomic landmarks need to be labeled, it is sufficient as long as points are selected on the bone surface.

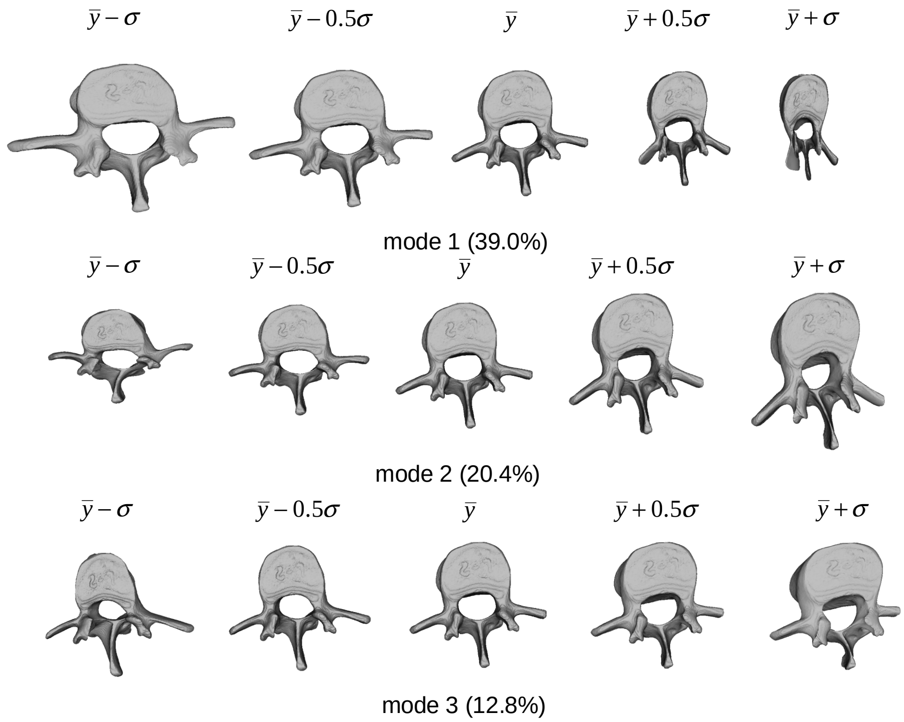


Figure 5.20: Spine vertebra atlas: 39.0% of shape variations is encoded in the first mode. The second and third modes encode 20.4% and 12.8% of shape variations.

Fig. 5.20 shows shape variations in the first three modes of the spine vertebra atlas. We can see that the vertebra becomes narrower but remains the same length among the first mode. The vertebral body (centrum) becomes larger along both the second and third modes. The spinous process becomes longer along the second mode but shorter along the third mode. It is twisted in the same way from the left to the right. We have noticed that the sampled surfaces given the bigger standard deviation (e.g. 1σ) have odd shapes. This may be caused by the insufficient number of training data. Fig. 5.21 shows shape variations encoded in each mode of the spine vertebra atlas. To keep 95% of shape variation we only need 10 modes.

5.3 Conclusions and Discussion

The semi-automatic procedure requires some manual input, e.g. selection of 40682 points for a reference model. But this procedure is a one-time task and can be conducted off line.

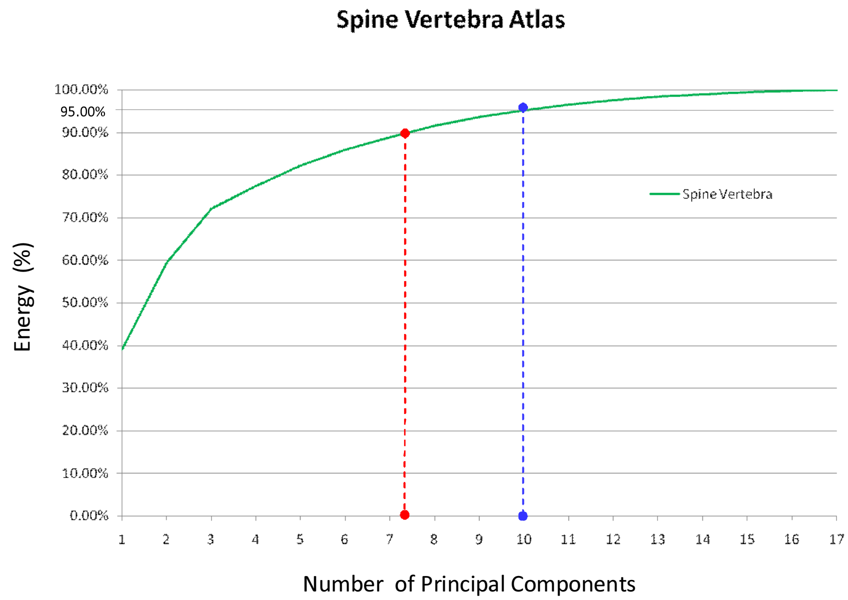


Figure 5.21: Illustration of shape variations encoded in each mode of the spine vertebra atlas.

It also requires selecting 250 points for each study image without any prior information about the shape. It normally takes 10-15 minutes, which is much less than several hours which could be used for complete manual segmentation of high resolution model. In the future e will consider using the atlas as a prior to enhance the performance and reduce manual work.

The salient points may be removed during the surface simplification and then mis-matching may occur. To handle this problem, we need to find a proper resolution for the simplified surface. The lower the faster but the more mis-matching could be established. That is why we apply the leave-one-out experiments to find a appropriate parameter for the resolution. Besides, we proposed a refining process by using the original data (either the high resolution surface or sparse labeled points) as a reference to reduce the mis-matching caused by simplification.

Experiments demonstrate that our two-level framework significantly improves efficiency of registration without decreasing accuracy of atlases. Shape variation learned from the training samples can be used for clinical studies and diagnosis. The shape atlas can also be used as a statistical shape prior for bone reconstruction from other imaging modalities such as endoscopes. We will discuss more details in Ch. 6.

Chapter 6

Improvement of Bone Reconstruction from Two-step Algorithm

As we have discussed in Ch.5, to solve over smoothness, inaccurate reconstruction and partial occlusions, a shape prior is needed to constrain the MISFS algorithm. We propose a two-step algorithm in this chapter to improve the reconstruction. A global shape constraint is enforced in the MISFS algorithm (see Sec. 6.1). The shape constraint is computed by aligning the inaccurate shape from the MISFS and the statistical shape atlas. Such a procedure is iterated as well as the shape constraints and inaccurate shape are updated in each iteration until the algorithm converges. To further improve reconstruction especially on the discontinued boundaries, we evaluate the reconstructed shape by comparing to the original endoscopic images (see Sec. 6.2). By maximizing the likelihood of gradient images synthesized from the reconstructed shape, a better reconstruction and more accurate shape is obtained.

6.1 Bottom-up Image based MISFS constrained by Statistical Shape Atlas

In this section we will discuss how to constrain the MISFS with a statistical atlas. The procedure is illustrated in Fig. 6.1. The pre-processing MISFS results in an initial shape S_1, given initial normals on global contours. S_1 is then aligned with the statistical atlas and the most closely matched atlas shape S_1^a is computed from sparse-to-dense reconstruction. S_1^a is used to update the initial normals and depths for the MISFS. The two steps, alignment

Table 6.1: Table of Notation in Sec. 6.1

S_1, S_2, S_k, S_K	-	MISFS result after the $1^{st}, 2^{nd}, k^{th}, K^{th}$ iteration
S_{mean}	-	mean shape
s, r, t, s^*, r^*, t^*	-	transformation parameters
N	-	number of surface vertices
S_1'	-	aligned partial shape after applying rigid transformation to S_1
E	-	eigenmatrix
E_1^{sub}	-	sub matrix of E
E_1^{sub+}	-	pseudo inverse of E_1^{sub}
S_{mean}'	-	sub vector of S_{mean}
S_1''	-	atlas shape corresponding to S_1'
S_1^a	-	aligned atlas shape after applying inverse rigid transformation to S_1''
n_{ax}, n_{ay}, n_{az}	-	three components of surface normal at surface point on S_1^a
x, y, z	-	world coordinates
$\tilde{x}, \tilde{y}$	-	image coordinates
F	-	focal length
z_a	-	depth on S_1^a
p_a, q_a	-	partial derivatives of z_a w.r.t. image coordinates $\tilde{x}, \tilde{y}$
$\mathbf{e_a}$	-	atlas constraint
λ_1, λ_2	-	Lagrange multipliers
k, l	-	image indices
m	-	iteration number

and MISFS are iterated until the best reconstruction is achieved. By these means, better knowledge of bones of interest are taken into account in the MISFS therefore a more accurate shape can be recovered.

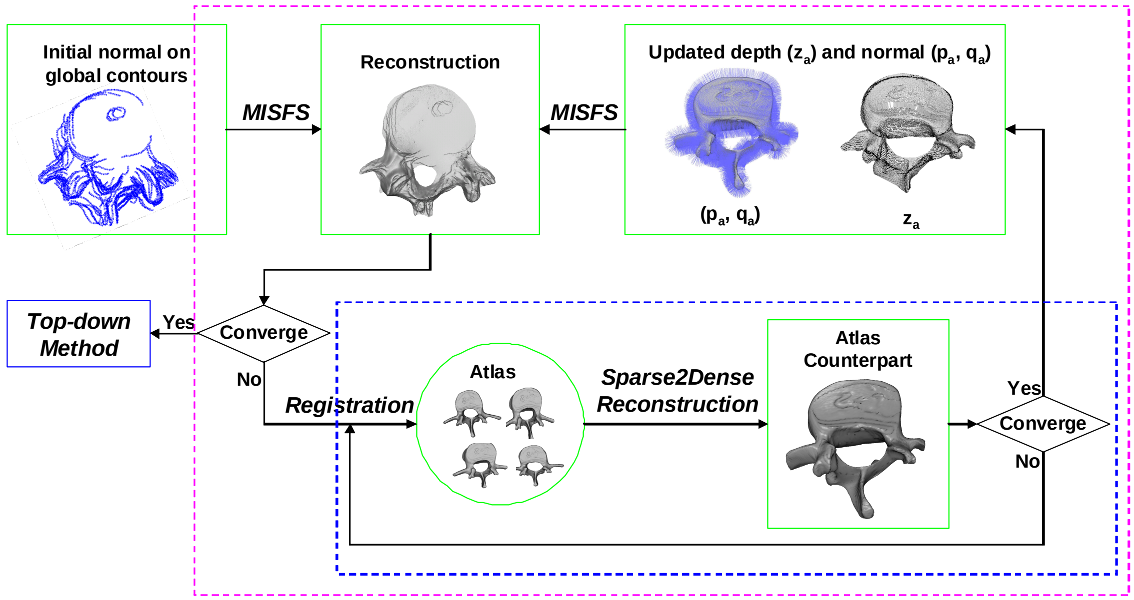


Figure 6.1: Illustration of the bottom-up MISFS: the pre-processing MISFS results in an initial shape S_1, given initial normals on global contours. S_1 is then aligned with the statistical atlas and the most closely matched atlas shape S_1^a is computed from sparse-to-dense reconstruction. S_1^a is used to update the initial normals (p_a, q_a) and depths z_a for the MISFS. The two steps, alignment and MISFS are iterated until the best reconstruction is achieved.

6.1.1 Aligning MISFS Shape with Atlas

Due to the limited range of motion for the endoscope and limited bone exposure during surgery, it is impossible to capture images of bones from 360 degrees and reconstruct a complete shape. In other words, S_1 represents a part of the atlas shape.

We first apply a scaled-ICP algorithm [5] to align S_1 with the atlas mean shape S_{mean}. We find the scaling factor s^* by comparing the model size in the 3D Eigenspace, and then compute the rotation and translation by minimizing the surface distance. i.e.

$$(r^*, t^*) = \arg\min \sum_{i}^{N} \{[r|t](s^* \cdot S_1(i)) - S_{mean}(near(i))\}^2 \tag{6.1}$$

where $near(i)$ is the index of the closest point of $S_1(i)$ on the mean shape. The initial starting position is roughly estimated based on the field of view of the reconstructed shape. We apply the transformation (r^*, t^*, s^*) to S_1 and obtain the aligned shape S_1' at the atlas pose:

$$S_1' = [r^*|t^*](s^* \cdot S_1) \tag{6.2}$$

Given the aligned partial shape S_1', a complete atlas shape is generated from sparse-to-dense reconstruction. In other words, we first project S_1' into a subspace of the atlas space to retrieve coefficients for each mode. These coefficients are used to reconstruct a complete shape from the original atlas space. Let E denote the eigenvector matrix of the atlas, and E_1^{sub} denote a sub matrix of E. Extracted from E, each row of E_1^{sub} corresponds to every vertex on the partial shape S_1'. Since E_1^{sub} is not an orthogonal matrix anymore, we use the pseudo inverse $(E_1^{sub})^+$ instead of the matrix transpose to compute the coefficients and then reconstruct the shape:

$$S_1'' = E \cdot (E_1^{sub})^+ \cdot (S_1' - S_{mean}') + S_{mean} \tag{6.3}$$

where S_{mean}' is the partial mean shape corresponding to S_1'.

The complete shape S_1'' is reconstructed from 10 modes of the atlas. Since we use the mean shape S_{mean} as the target shape for the initial alignment, wrong correspondences may exist. Thus we use S_1'' as a new target shape and iterative the alignment and reconstruction several times until S_1' and the atlas are well aligned. Finally we transform S_1'' to the original pose and scale, by applying the inverse transformation (Eq. 6.2):

$$S_1^a = \left(\left[r^{*T} | - t^*\right] S_1''\right) / s^* \tag{6.4}$$

where the complete shape S_1^a is considered as the most closest atlas shape for S_1.

6.1.2 Constraining MISFS by Atlas

In this section we use the surface geometry information of S_1^a to constrain the MISFS. Given each surface point P_i on S_1^a, the surface normal $\vec{n} = [n_{ax}, n_{ay}, n_{az}]$ is computed as the average of the surface normals from adjacent triangles. Recall Eq. 4.3 and 4.7, we have

$$\begin{gathered} \vec{n} = [-\frac{\partial z}{\partial x}, -\frac{\partial z}{\partial y}, 1] / \sqrt{(\frac{\partial z}{\partial x})^2 + (\frac{\partial z}{\partial y})^2 + 1} \\ \frac{\partial z}{\partial x} = \frac{F p_a}{z_a + \tilde{x} p_a}, \quad \frac{\partial z}{\partial y} = \frac{F q_a}{z_a + \tilde{y} q_a} \end{gathered} \tag{6.5}$$

where $p_a = \partial z / \partial \tilde{x}$ and $q_a = \partial z / \partial \tilde{y}$, $(\tilde{x}, \tilde{y})$ are image coordinates, F is the focal length and z_a is the depth at P_i. We compute the prior (p_a, q_a) from S_1^a using Eq. 6.5

$$p_a = -\frac{(n_{ax}/n_{az}) \cdot z_a}{(n_{ax}/n_{az}) \cdot \tilde{x} + F}, \quad q_a = -\frac{(n_{ay}/n_{az}) \cdot z_a}{(n_{ay}/n_{az}) \cdot \tilde{y} + F} \tag{6.6}$$

The atlas constraint $\mathbf{e_a}$ is defined by minimizing the errors of surface normals and depths compared to the aligned atlas shape S_1^a:

$$\mathbf{e_a} = \int\int_{aligned\ vertices} [(p-p_a)^2 + (q-q_a)^2 + (z-z_a)^2] d\tilde{x} d\tilde{y} \tag{6.7}$$

Recall Eq. 4.12 and Eq. 4.19, the entire error is represented by two terms, the image irradiance error $\mathbf{e_i}(z,p,q)$ and the local geometry constraint $\mathbf{e_s}(z,p,q)$ (smoothness constraint is used in Ch. 4). Now we add $\mathbf{e_a}$ to the error function as another soft constraint:

$$\mathbf{e}(z,p,q) = \lambda_1 \mathbf{e_i}(z,p,q) + \lambda_2 \mathbf{e_a}(z,p,q) + (1-\lambda_1-\lambda_2)\mathbf{e_s}(z,p,q) \tag{6.8}$$

By discretizing Eq. 6.8 we obtain:

$$\mathbf{e}(z_{k,l}, p_{k,l}, q_{k,l}) = \sum_k \sum_l [\lambda_1 \mathbf{e_i}_{k,l} + \lambda_2 \mathbf{e_a}_{k,l} + (1-\lambda_1-\lambda_2)\mathbf{e_s}_{k,l}] \tag{6.9}$$

where

$$\mathbf{e_a}_{k,l} = (p_{k,l} - p_{a k,l})^2 + (q_{k,l} - q_{a k,l})^2 + (z_{k,l} - z_{a k,l})^2 \tag{6.10}$$

The derivative of $\mathbf{e_a}$ with respect to $\eta_{k,l}$ (η represents p, q or z) are:

$$\frac{\partial \mathbf{e_a}}{\partial \eta_{k,l}} = 2(\eta_{k,l} - \eta_{a k,l}) \tag{6.11}$$

Similarly to Ch. 4 we have

$$\frac{\partial \mathbf{e_s}}{\partial \eta_{k,l}} = 8(\eta_{k,l} - \bar{\eta}_{k,l}), \quad \frac{\partial \mathbf{e_i}}{\partial \eta_{k,l}} = -2(I_{k,l} - R_{k,l})\frac{\partial R}{\partial \eta_{k,l}} \tag{6.12}$$

In Eq. 6.12, the smoothness term $\bar{\eta}_{k,l}$ is replaced by the local robust regularizer $\hat{\eta}_{k,l}$. We minimize Eq. 6.9 and obtain update functions for $\eta_{k,l}$ (η represents p, q or z):

$$\begin{aligned} \eta_{k,l}^{m+1} &= \frac{\lambda_1}{4(1-\lambda_1-\lambda_2)+\lambda_2}[I_{k,l} - R(k,l,\bar{z}_{k,l}^m, \bar{p}_{k,l}^m, \bar{q}_{k,l}^m)]\frac{\partial R}{\partial \eta_{k,l}}|_{\bar{\eta}_{k,l}^m} \\ &+ \frac{\lambda_2}{4(1-\lambda_1-\lambda_2)+\lambda_2}\eta_{a k,l} + \frac{1-\lambda_1-\lambda_2}{(1-\lambda_1-\lambda_2)+\lambda_2/4}\hat{\eta}_{k,l}^m \end{aligned} \tag{6.13}$$

Constrained by the statistical atlas, the new MISFS reconstructs a new shape S_2. We repeat the alignment and reconstruction discussed in the previous section to S_2, whose most closest atlas shape S_2^a is used to update the atlas constraint for MISFS. The above two steps are iterated to generate the shape S_3, S_4, ..., until S_K when the surface distance $\|S_K - S_{K-1}\| < \epsilon$. The most closest atlas shape of S_K is denoted by S_K^a and corresponding coefficients as c_K^a.

Table 6.2: Table of Notation in Sec. 6.2

S^a_K	-	atlas shape corresponding to MISFS result after the K^{th} iteration
S^a_F	-	refined atlas shape
c^a_K	-	coefficients corresponding to S^a_K
E	-	eigenmatrix
N	-	number of vertices on atlas shape
A	-	intrinsic matrix of endoscope
M_u	-	extrinsic matrix of endoscope at pose u
U	-	number of endoscope poses
v^u_i	-	2D image coordinates of vertex i at pose u
p^u_i	-	3D homogeneous coordinates of vertex i at pose u
$\phi(\cdot)$	-	normalization function
$G(\cdot)$	-	image gradient operator
G_o	-	gradient of original images
G_a	-	gradient of synthesized image from S^a_K
$f(\cdot)$	-	minimization function
J	-	Jacobian
$G_{o\tilde{x}}, G_{o\tilde{y}}$	-	partial derivative of G_o w.r.t. image coordinates $\tilde{x}, \tilde{y}$
$G_{a\tilde{x}}, G_{a\tilde{y}}$	-	partial derivative of G_a w.r.t. image coordinates $\tilde{x}, \tilde{y}$
S^a_F	-	refined atlas shape
c^a_K	-	coefficients corresponding to S^a_F

6.2 Top-down Refinement by Maximizing Likelihood of Image Gradients

As Fig. 6.2 illustrates, to ensure the accuracy of reconstruction, especially on the surface discontinuities, we examine the shape S^a_K by comparing the image gradients. We synthesize a series of images from S^a_K at different poses and compare them with the original endoscopic images. Since the imaging modalities are different between the synthesized images from the atlas shape and the original endoscopic images, the image appearance has a lot of difference. We thus compare the difference through contour gradients which

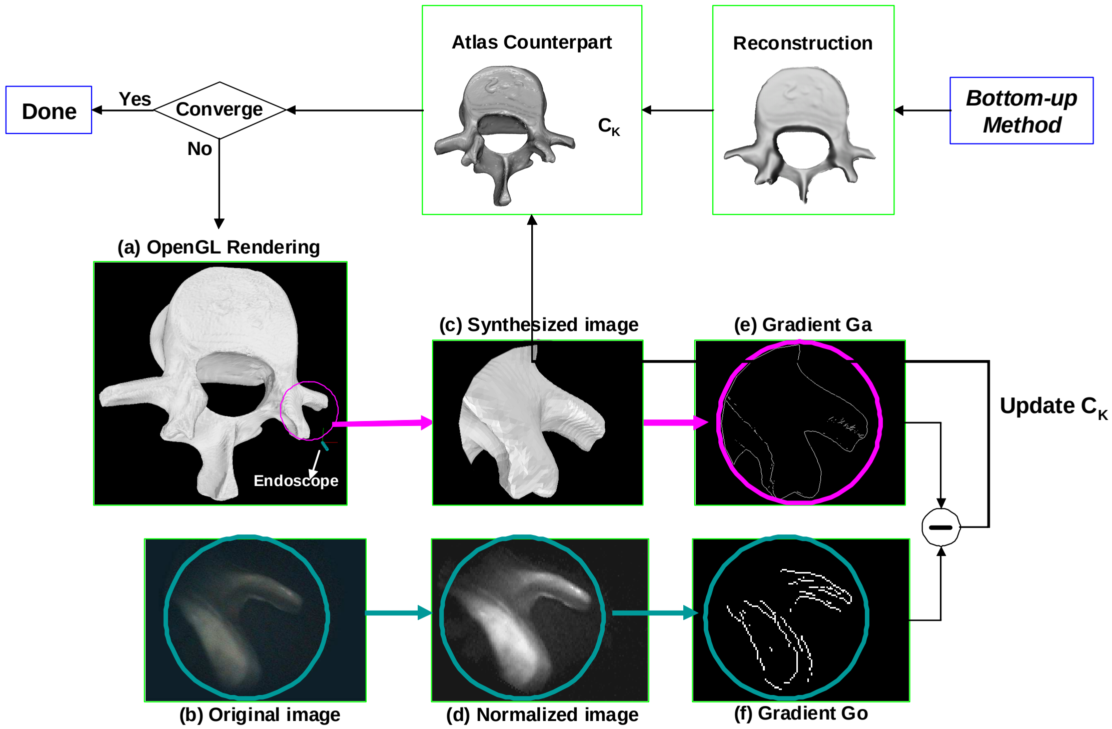


Figure 6.2: Illustration of top-down method: (a) OpenGL rendering S_K^a. (b) Corresponding endoscopic image. (c) Synthesized endoscopic image from the same camera pose. (d) Normalized endoscopic image. (e) Gradient of synthesized image G_a. (f) Gradient of normalized image G_o.

are most robust to lighting variations and strongly respond to surface discontinuities. By manipulating the coefficients c_K^a, the refined atlas shape is obtained and noted as S_F^a.

Let E denote the $3N \times 10$ eigenvectors matrix of the atlas, $E_i(3(i-1)+1:3i,:)c_K^a$ represents the vertex $i(i = 1, 2, \cdots, N)$ reconstructed from the atlas. Given the calibrated intrinsic and extrinsic matrices of the endoscope A_{3x3} and M_{3x4}^u (at pose u, $u = 1, \cdots, U$), the corresponding image pixel vector is $v_i^u = \phi(p_i^u), p_i^u = AM^u[E_i c_K^a; 1]$. $\phi(p) = [p_x/p_z; p_y/p_z]$ normalize the 3D homogeneous coordinates to 2D image coordinates. Let $G(\cdot)$ denote image gradient operator, our goal is to maximize the likelihood of the image gradients, i.e., minimize $\delta G = G_o - G_a$, where G_o computes the gradient of normalized endoscopic image (the image is normalized by removing the geometric distortion and illumination effect), G_a computes the gradient of OpenGL rendered image from

S_K^a (see Fig. 6.2). Let

$$f(c_K^a) = \frac{1}{2} \sum_u^U \sum_i^N [G_o(v_i^u) - G_a(v_i^u)]^2 \tag{6.14}$$

Using Taylor expansion, we obtain $\delta c_K^a = -J^{-1} f(c_K^a)$, where

$$\begin{aligned} J = \frac{\partial f}{\partial c_K^a} &= \sum_u^U \sum_i^N [\frac{\partial G_o}{\partial v_i^u} - \frac{\partial G_a}{\partial v_i^u}]^T \cdot \frac{\partial v_i^u}{\partial c_K^a} \\ \frac{\partial v_i^u}{\partial c_K^a} &= \frac{\partial \phi}{\partial p_i^u} \cdot \frac{\partial p_i^u}{\partial c_K^a} \end{aligned} \tag{6.15}$$

Since

$$\begin{aligned} \frac{\partial G_o}{\partial v_i^u} - \frac{\partial G_a}{\partial v_i^u} &= \begin{bmatrix} G_{o\tilde{x}}(v_i^u) - G_{a\tilde{x}}(v_i^u) \\ G_{o\tilde{y}}(v_i^u) - G_{a\tilde{y}}(v_i^u) \end{bmatrix} \\ \frac{\partial \phi}{\partial p_i^u} &= \begin{bmatrix} \frac{1}{p_i^u(z)} & 0 & \frac{-p_i^u(x)}{(p_i^u(z))^2} \\ 0 & \frac{1}{p_i^u(z)} & \frac{-p_i^u(y)}{(p_i^u(z))^2} \end{bmatrix} \\ \frac{\partial p_i^u}{\partial c_K^a} &= AM^u E_i \end{aligned} \tag{6.16}$$

Jacobian J can be rewritten as:

$$J = \sum_u^U \sum_i^N \begin{bmatrix} G_{o\tilde{x}}(v_i^u) - G_{a\tilde{x}}(v_i^u) \\ G_{o\tilde{y}}(v_i^u) - G_{a\tilde{y}}(v_i^u) \end{bmatrix}^T \cdot \begin{bmatrix} \frac{1}{p_i^u(z)} & 0 & \frac{-p_i^u(x)}{(p_i^u(z))^2} \\ 0 & \frac{1}{p_i^u(z)} & \frac{-p_i^u(y)}{(p_i^u(z))^2} \end{bmatrix} \cdot AM^u E_i \tag{6.17}$$

$p_i^u(x), p_i^u(y), p_i^u(z)$ are three components of the homogenous coordinates p_i^u, which is the function of c_K^a. In each iteration, we use previous c_K^a to compute p_i^u and v_i^u, and then the Jacobian J of Eq. 6.14. Thus c_K^a can be updated by Eq. 6.17. The final coefficients c_F^a are used to generate the refined shape S_F^a from the atlas.

6.3 Experiments and Results

6.3.1 Simulations

We first conducted several simulation trials on synthesized data. Given the statistical atlas, we manipulated the first ten coefficients and generated 10 different shapes with various poses shown in Fig. 6.3. For each virtual shape model, we rendered it in 3D graphics context and moved a virtual endoscope around to collect a set of synthesized images (see Fig. 6.3). The virtual endoscope has the fixed camera parameters thus the calibration error was not considered, neither did the tracking error. We directly called the OpenGL function to obtain the view-port vector, model-view and projection matrices. We used a bright spot light source which enables a good image quality, so we placed the virtual endoscope a bit far away from the 3D model and still captured good images. By these means we were able to image more area of the model in a single image and used much less images to reconstruct one side of the view. We collected 6-10 images for each synthesized shape.

We conducted our two-step approach on each set of synthesized images and reconstructed corresponding shapes. We compared the reconstructed shapes with the original ones in Fig. 6.4. The simulation results show clear improvement by introducing the atlas constraint, so does the refinement. Since the synthesized shapes are sampled from the atlas space, the ideal reconstruction error for S_F^a should be close to zero if the correct correspondences can be achieved. Without calibration and tracking errors, only the arbitrary camera motions will cause the misalignment and increase the reconstruction error. Fig. 6.4 shows the averaged RMS error for S_K, S_K^a and S_F^a is 0.76 mm, 0.54 mm and 0.42 mm, respectively. These numbers provide a rough lower bound for accuracy of our algorithm.

6.3.2 Artificial Lumbar Models

Real cases are more complicated due to the errors introduced by camera calibration and tracking system. And the real vertebra shape may not belong to the atlas population. The image quality also affect the contour extraction and shape-from-shading algorithm. We use the artificial lumbar vertebrae L4 and L5 to test our algorithm. For each model we captured 4 image sets from different range of views: top, bottom, left side and right side (Note that we could never have access to some views in real surgery, for example, the top view). In each image set, we collected a set of images. For instance, we captured 145 endoscopic images at 29 positions from the top view of the vertebra L4 (for each position we took 5 images and used the averaged image for reconstruction). Fig. 6.5 shows the artificial vertebra and averaged endoscopic images. The colored circles illustrate

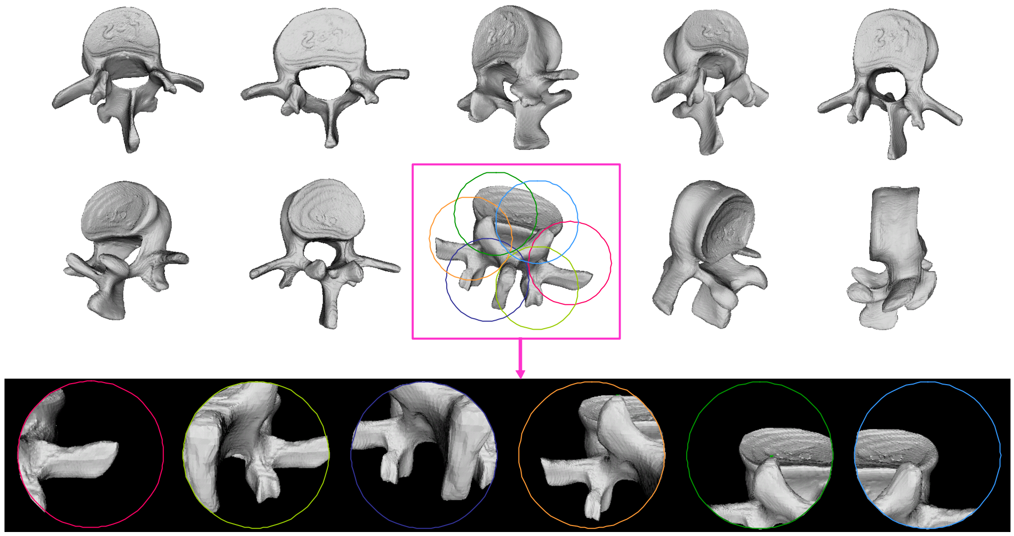


Figure 6.3: Illustration of 10 synthesized shapes from the atlas for simulation experiments. For each shape, we moved the virtual endoscope to collet a set of images, for example, the bottom images were collected from the 8^{th} shape.

the endoscope poses with different rotations and translations. The displayed images are geometric rectified and normalized for illumination. In the preprocessing MISFS with local robust regularizer, we chose $\lambda_1 = 0.001$ and increased it by 0.02 whenever the error in Eq. 6.8 decreased by 1%. Fig. 6.1 shows the resulted shape S_1. We can see some discontinuities are kept. Fig. 6.1 also shows the aligned atlas shape S_1^a, the updated surface normals (p_a, q_a) and the depths z_a, which were used for the second round MISFS with the atlas constraint. From the second round MISFS, we selected $\lambda_1 = 0.01$ (increased by 0.01) and $\lambda_2 = 0.1$ (increase by 0.05).

We compare the MISFS reconstruction S_K, bottom-up result S_K^a and top-down result S_F^a with the laser scanned ground truth (L4: 72.6 x 77.1 x 38.2 mm, scanned from 13 view points; L5: 92.9 x 73.4 x 37.8 mm, scanned from 16 view points). The RMS errors over 8 data sets are shown in Figs. 6.6, and visual results are shown in Figs. 6.7-6.8. We can see a significant improvement from the two-step method, although we have more complicated shapes, larger range of motions including rotations (In Ch. 4 we only handle the translation). Since we captured images from one side of the spine vertebra model, the MISFS result S_K can only recover the shape information for this side. Although S_K^a and S_F^a represent a complete atlas shape, most of vertices on S_K^a and S_F^a have no correspondences on S_K. Those points do not have any support from images thus correspond bigger errors. That's why the overall RMS errors are slightly worse for the atlas shape S_K^a and S_F^a: the

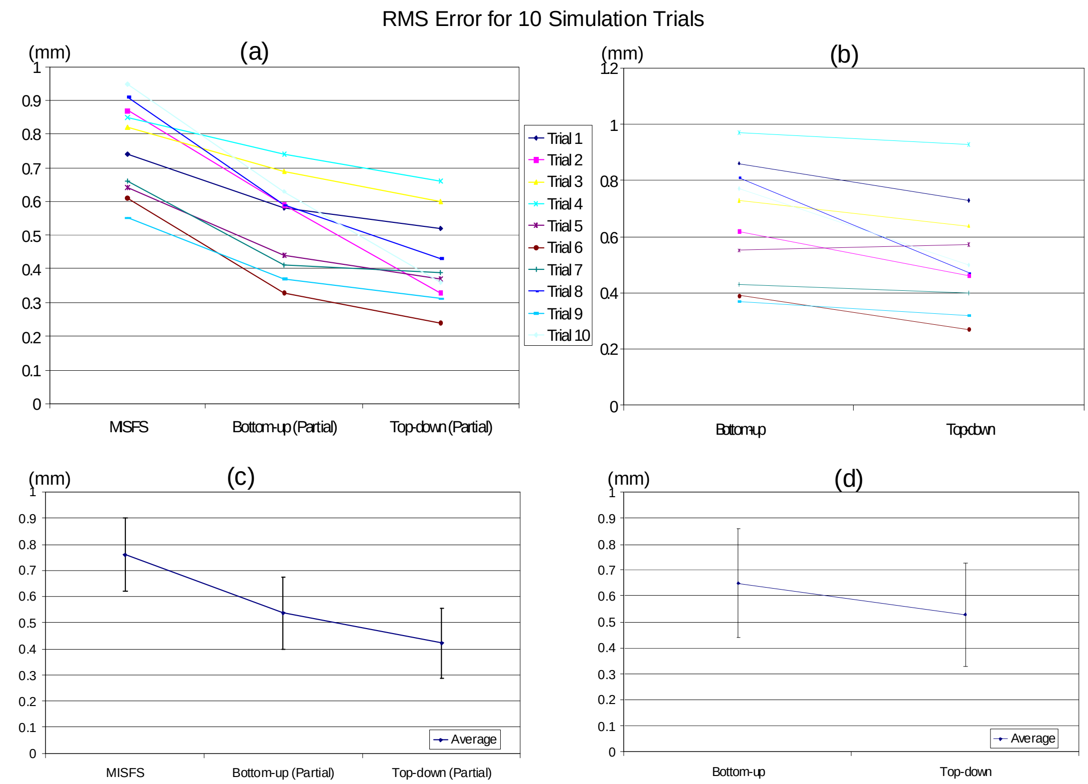


Figure 6.4: Reconstruction errors for ten synthesized shapes. The RMS errors are calculated by comparing three shapes (MISFS, bottom-up and top-down) of each synthesized images given the synthesized shape. Since we captured images from one side of synthesized shape, the reconstructed shape from MISFS can only recover the shape information for this side. Although the bottom-up result and top-down result represent complete atlas shapes, most of vertices on these two atlas shapes have no correspondences on the MISFS result. Those points do not have any support from images thus correspond bigger errors. That's why the overall RMS errors are slightly worse for the bottom-up and top-down method, as shown in (b). However, when we look at the partial RMS errors by only considering points having correspondences in the images, we can see a big improvement from the two-step algorithm in (a). (c) and (d) show the average and standard deviation over 10 trials.

averaged RMS error over eight data sets is 2.07 mm and 1.95 mm respectively. However, when we look at the RMS errors by only considering points having correspondences in the images, we can find better accuracy: the averaged RMS error over eight data sets for S_K, S_K^a and S_F^a is 1.60 mm, 1.36 mm and 1.17 mm, respectively. The standard deviation is 0.45 mm, 0.25 mm and 0.21 mm, respectively.

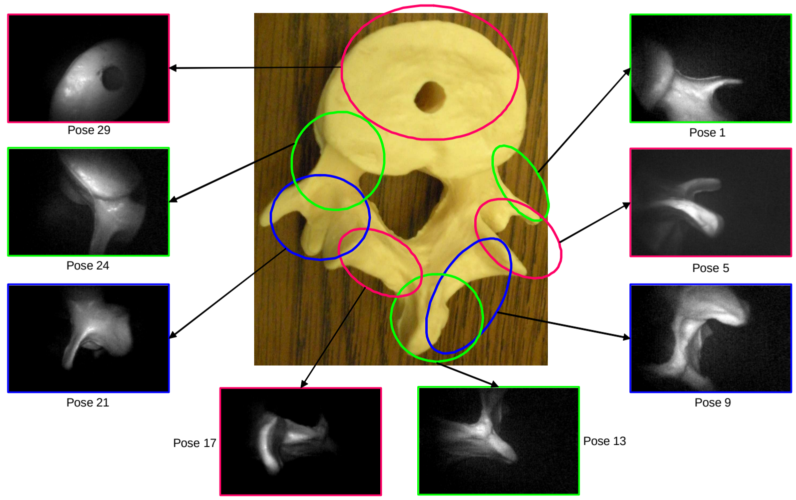


Figure 6.5: Experimental setup for lumbar vertebrae reconstruction. The artificial lumbar vertebra L4 is showed in the middle. Surrounding images are captured by an oblique endoscope from different poses involving rotations and translations, which are illustrated by colored circles. All the images are geometric rectified and normalized for illumination.

From the result of the top view of L4, we can see large errors in the vertebra body (the top area of the shape), which is due to the ambiguity during the individual shape registration, since the captured images from the pose 26 to 29 look very similar. It is the limitation of the bottom-up appearance based method. A better representation of this area can be found in top-down result. The bottom view of L5 shows an occlusion example, where some portions of the vertebra body were not imaged. The MISFS result cannot recover the occluded part but the corresponding atlas shape provides an estimation of the missing portions.

We have noticed large local errors in the mamillary process and superior articular process of L4 (the area corresponding to the pose 5 and 17 in Fig. 6.5) and L5, and in the spinous process and transverse process of L5. We have also noticed RMS errors of L5 is slightly bigger than L4. It is due to the limited number of training samples to recover the various shape variation, especially for L5. We use Th12 and L1-L5 to learn the statistical atlas. Among those vertebrae, L5 is much more different from others: its body is much deeper in front than behind, its spinous process is smaller, the interval between the inferior articular process is wider, and its transverse process are thicker. The result can be improved by collecting more training shapes in the future.

For the right view of L4 and top view of L5, the refining step decreases the partial surface distance but increases the overall difference (see Figs. 6.7-6.8). It is because the

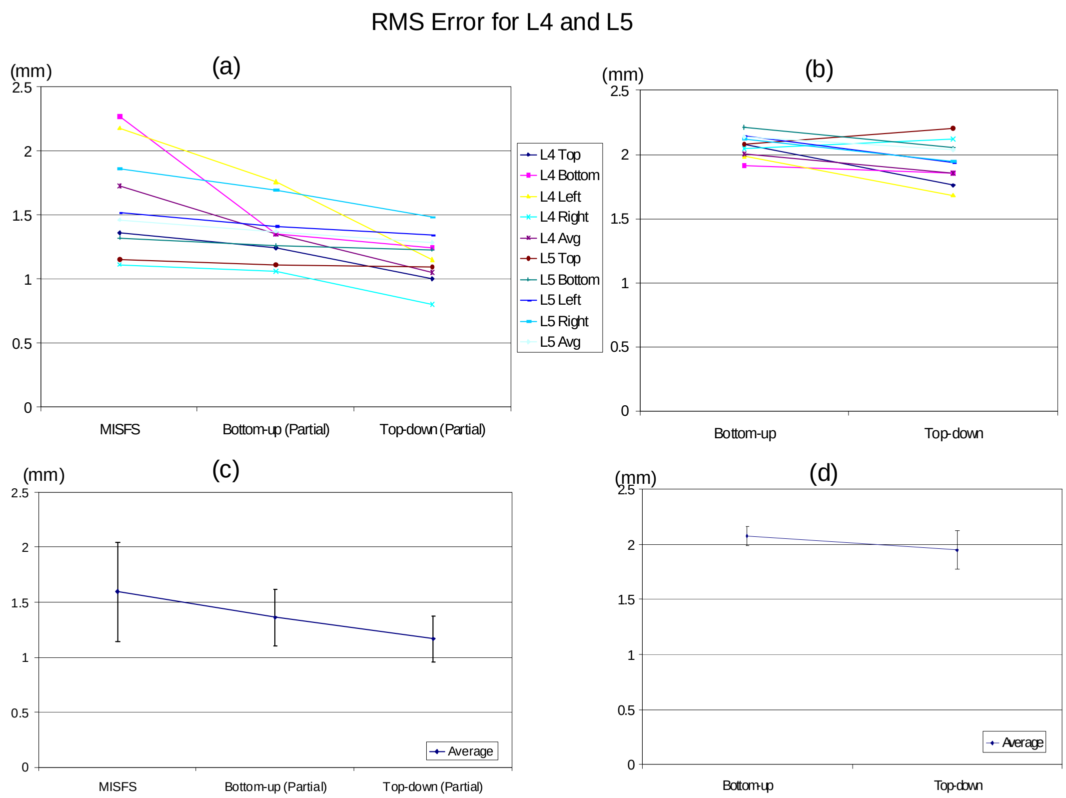


Figure 6.6: Reconstruction errors for all data sets of L4 and L5. The RMS errors are calculated by comparing three shapes (MISFS, bottom-up and top-down) of each synthesized images given the synthesized shape. Since we captured images from one side of synthesized shape, the reconstructed shape from MISFS can only recover the shape information for this side. Although the bottom-up result and top-down result represent complete atlas shapes, most of vertices on these two atlas shapes have no correspondences on the MISFS result. Those points do not have any support from images thus correspond bigger errors. That's why the overall RMS errors are slightly worse for the bottom-up and top-down method, as shown in (b). However, when we look at the partial RMS errors by only considering points having correspondences in the images, we can see a big improvement from the two-step algorithm in (a). (c) and (d) show the average and standard deviation over 8 data sets.

top-down method is basically a gradient decent procedure which may converge to a local minimum. Images from side views cannot give enough constraint for the entire shape, that's why the reconstruction errors for the left and right views are slightly bigger than the other two views. Note that lack of images from other sides of the vertebra leads misalignment and inaccurate reconstruction for invisible points.

6.4 Conclusions and Discussion

By introducing the statistical shape atlas as a new global constraint for the MISFS, we are able to deal with local discontinuities, inaccurate reconstruction and partial occlusions. We modified the bottom-up image based MISFS by introducing a global shape atlas. We also develop a top-down refining procedure to evaluate and improve the reconstruction by minimizing the difference of image gradients.

Note that since we use artificial spine vertebrae in our experiments, some views in Fig. 6.5 could never be accessed in real surgery. Although in practice we can only get a limited exposure of the shape and the errors for parts from the other sides of view may be significant, the occlusion occurs from the current imaging view can be successfully recovered by the atlas.

We have noticed that some samples from the atlas look non-natural. It is due to the lack of sufficient training data. Even with such a roughly estimated atlas, we are able to see an improvement. We believe more data collected in the future will improve our result further. So far our method has been tested in Matlab for ex vivo data only. It takes about 15 minutes for a data set with 30 images (PC:2.4GCPU/2GRAM). The computation can be paralleled to achieve much faster performance. Due to the quality of the endoscope we used in experiments, automatic contour extrapolation failed in many images. Manual labeling of boundaries is still required for our current system. Aligning the MISFS shape and the atlas also needs an initial manual alignment to achieve a good matching. The user interaction thus hinders the real-time implementation and in-vivo testing.

Above issues will be studied and addressed in our future work. We hope our system will soon be ready for in-vivo testing.

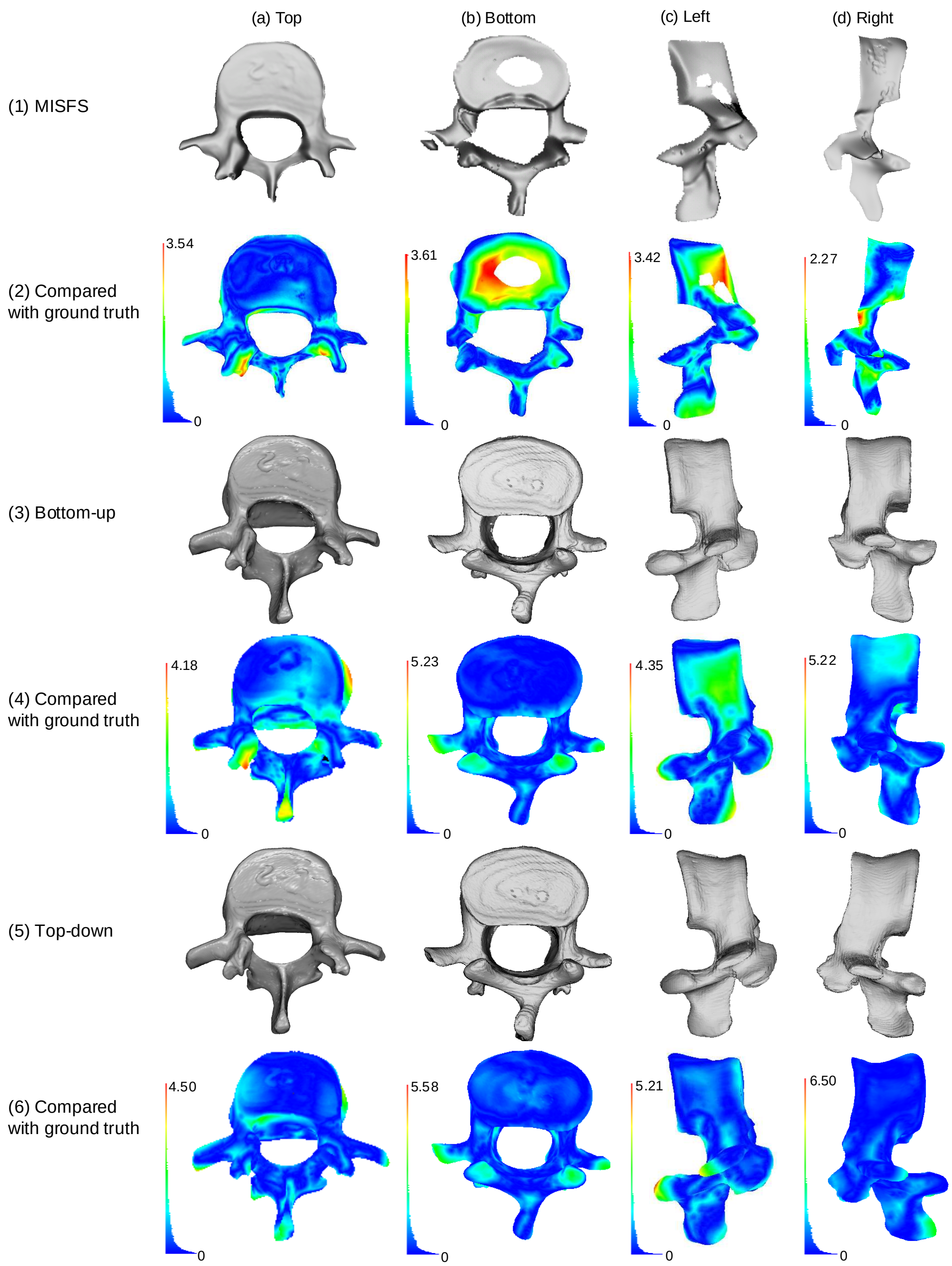


Figure 6.7: Reconstruction results for different views of L4. For each data set, we show the MISFS, bottom-up and top-down results, also show the Housdorff surface distance from each result to the ground truth. The surface distance is displayed using a JET color map. Each distance map is associated with a color bar displayed on the left (with the maximum and minimum distance (mm) labeled on the side).

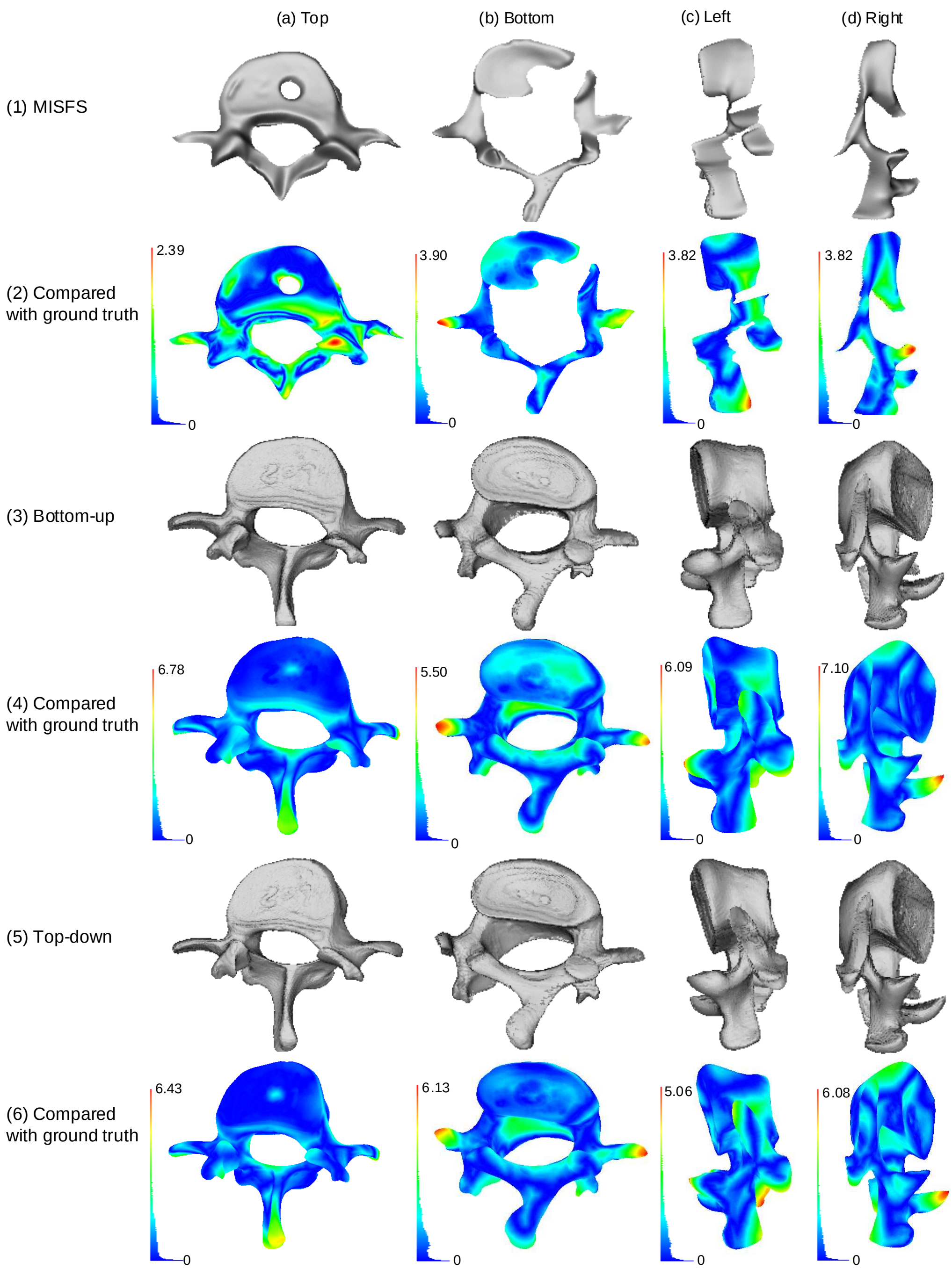


Figure 6.8: Reconstruction results for different views of L5. For each data set, we show the MISFS, bottom-up and top-down results, also show the Housdorff surface distance from each result to the ground truth. The surface distance is displayed using a JET color map. Each distance map is associated with a color bar displayed on the left (with the maximum and minimum distance (mm) labeled on the side).

Chapter 7

Conclusions and Future Work

Endoscope is a key tool used for minimally invasive surgery, and endoscopy-related image analysis attracts increasing attention. In this thesis, we have taken the first steps to develop a methodology for reconstruction of bone shape from multiple endoscopic images. The research experiments were performed in a simplified ex-vivo setup. Our contributions are listed below:

1. We developed a novel methodology to calibrate oblique-viewing endoscopes both geometrically and photometrically.
2. We formulated a novel scene radiance model for the endoscope imaging system under near point lighting and perspective projection, without assuming that the light sources are located at the projection center, to reconstruct the shape from a single image.
3. We developed a multi-image shape-from-shading algorithm by tracking the endoscope in the world coordinates and aligning partial shapes obtained from different images. Then the global contours are used to constrain the shape-from-shading across different images simultaneously. Finally, a consistent shape is reconstructed by re-growing the surface normals and depths in all views.
4. In order to deal with over-smoothness and ambiguity caused by the smoothness constraints, rotation of the endoscope and partial occlusions, we developed a two-step approach using the statistical atlas as a global shape constraint.
5. In order to fulfill the two-step reconstruction algorithm, we develop a semi-automatic two-level registration procedure to construct the statistical atlas for bone structures.

All our experiments are based on a clear ex-vivo setup. Without calibrating and tracking errors, the simulations show an average of 0.3 mm accuracy, which can be considered as an error generated by the algorithm. When the calibrating and tracking errors are taken into account, the system error is increased to 0.8 mm. This error can be further enlarged in in-vivo experiments with other environmental noise and when our assumptions outlined in Ch. 1 are not met. How much accuracy is needed depends on the surgical application. For example, bone cutting and drilling may require much higher accuracy then the 3D visualization for diagnosis and inspection.

Our current system still requires some manual input such as contour labeling or manual adjustment for initial atlas matching to improve alignment. In addition, the current algorithm does not run in real-time. We hope to speed up the system to real-time performance by rewriting the Matlab portion of the code in C++, using GPU acceleration and applying parallel computing to shape-from-shading in different images. The endoscope used in this study is relatively old and the images are low resolution video-capture images. In the future, we plan to use high-resolution endoscope with better image quality, such that we can extract contours automatically and more accurately.

As long as we can handle the above two issues, we are ready to move to the next step, the in-vivo testing. In the in-vivo testing phase, we will evaluate the effects of environmental noise (tools, fog, specularity, subsurface scattering, etc.), random occlusions (blood, pieces of bones, tissues, ligaments, etc.), changing of focus of the scope, non-Lambertian surface, and so on.

In the two-step algorithm, the top-down step uses the original images to improve the surface details and reconstruction accuracy by comparing the contour gradients images. However, the experimental results did not show significant improvement, which can be caused by the limited atlas modes we used. Improving the atlas quality by collecting a larger database of images will be critical for improved performance of this step. As of now, we just subtract two gradient images to compute the difference, which can be improved in the future by computing the distance transformed between two sets of contours, for example, using ICP.

In some cases, the target bone shape may contain an anomaly that can not be reconstructed from the atlas, the reconstruction error from the two-step algorithm might then be increased, since the atlas is derived from healthy population. Therefore, if the reconstruction error exceeds a threshold, we should go back to MISFS and use the original images for final reconstruction in that region.

Another possible solution for improving the local shape detail is to learn multiple local partial atlases. In other words, the whole shape will be segmented into several important parts and each part will be refined by the local shape variation. The problem with this method is how to segment the parts and determining how many parts are enough.

Lambertian surface is assumed in our setup. Non-Lambertian surface will make the scene radiation function more complicated. Although Lambertian is a good rough assumption for the artificial bones, how to select a good albedo still requires lots of studies and evaluations. In this thesis we manually chose a constant albedo for all different artificial bones. Additional study is needed to determine whether the assumptions used in this work can be extended to a realistic arthroscopy environment.

As far as we know, human bones show significant symmetry although it is not perfect. We should be able to use this cue to help the reconstruction, as an additional global shape constraint. Usually asymmetry is related to some anatomical abnormalities.

Our training samples for atlas construction are not sufficient, which causes the unnatural shape variation generated from the atlas space. A more theoretical way to judge if the number of training sample is enough is to use leave-one-out experiments. Our femur atlas construction is evaluated by the leave-one-out experiments which show sufficient training samples and thus smoothing and natural shape variation.

If multiple bone parts exist in the same image, the automatic contour extraction algorithm may extract contours from different bone parts and assign it to the same bone, which will cause inaccurate reconstruction, the discontinuity between two bone parts will be removed and modified. To handle such a case user interaction seems necessary at this point.

Finally, we hope our research will eventually benefit future surgical processes by making them more reliable and accurate.

Appendix A

Derivation of Eq. 4.17

In this appendix we show the derivation for Eq. 4.17. To minimize the error $\mathbf{e}$, we take the derivative of $\mathbf{e}$ w.r.t. $p_{k,l}$, where (k,l) is the position of an image pixel:

$$\begin{aligned}\frac{\partial \mathbf{e}}{\partial p_{k,l}} &= \frac{\partial}{\partial p_{k,l}}[(1-\lambda)\mathbf{e_s} + \lambda \mathbf{e_i}] \\ &= (1-\lambda)\frac{\partial \mathbf{e_s}}{\partial p_{k,l}} + \lambda \frac{\partial \mathbf{e_i}}{\partial p_{k,l}}\end{aligned} \tag{A.1}$$

for each pixel (k,l) we have

$$\begin{gathered}\mathbf{e_s}_{k,l} = (p_{k+1,l} - p_{k,l})^2 + (p_{k,l+1} - p_{k,l})^2 + (q_{k+1,l} - q_{k,l})^2 \\ +(q_{k,l+1} - q_{k,l})^2 + (z_{k+1,l} - z_{k,l})^2 + (z_{k,l+1} - z_{k,l})^2 \\ \mathbf{e_i}_{k,l} = [I_{k,l} - R(k,l,z_{k,l},p_{k,l},q_{k,l})]^2\end{gathered} \tag{A.2}$$

Given smoothness constraints:

$$\begin{aligned}\frac{\partial \mathbf{e_s}}{\partial p_{k,l}} = \frac{\partial \mathbf{e_s}_{k,l}}{\partial p_{k,l}} + \frac{\partial \mathbf{e_s}_{k-1,l}}{\partial p_{k,l}} + \frac{\partial \mathbf{e_s}_{k,l-1}}{\partial p_{k,l}} \\ &= 2((p_{k,l} - p_{k+1,l}) + (p_{k,l} - p_{k,l+1})) + 2(p_{k,l} - p_{k-1,l}) + 2(p_{k,l} - p_{k,l-1}) \\ &= 8(p_{k,l} - \frac{1}{4}(p_{k+1,l} + p_{k,l+1} + p_{k-1,l} + p_{k,l-1})) \\ &= 8(p_{k,l} - \bar{p}_{k,l})\end{aligned} \tag{A.3}$$

where $\bar{p}_{k,l}$ represents an 8-neighbor average:

$$\bar{p}^n_{k,l} = \frac{1}{5}(p^n_{k,l-1} + p^n_{k,l+1} + p^n_{k+1,l} + p^n_{k-1,l}) + \frac{1}{20}(p^n_{k-1,l-1} + p^n_{k-1,l+1} + p^n_{k+1,l-1} + p^n_{k+1,l+1}) \tag{A.4}$$

Given brightness error:

$$\frac{\partial \mathbf{e_i}}{\partial p_{k,l}} = \frac{\partial \mathbf{e_i}_{k,l}}{\partial p_{k,l}} = -2(I_{k,l} - R(k,l,z_{k,l},p_{k,l},q_{k,l}))\frac{\partial R}{\partial p_{k,l}} \tag{A.5}$$

From Eq. A.3 and A.5 we have

$$\frac{\partial \mathbf{e}}{\partial p_{k,l}} = 8(1-\lambda)(p_{k,l} - \bar{p}_{k,l}) - 2\lambda(I_{k,l} - R_{k,l})\frac{\partial R}{\partial p_{k,l}} \tag{A.6}$$

Setting Eq. A.6 to zero yields:

$$p_{k,l} = \bar{p}_{k,l} + \frac{\lambda}{4(1-\lambda)}(I_{k,l} - R(k,l,z_{k,l},p_{k,l},q_{k,l}))\frac{\partial R}{\partial p_{k,l}} \tag{A.7}$$

In order to make the algorithm robust, we use the average $\bar{p}$, $\bar{q}$ and $\bar{z}$ when computing the reflectance map $R(\tilde{x},\tilde{y},z,p,q)$ in Eq. A.7. We apply the same procedure to derive the update functions for q and z as below:

$$\begin{aligned}
p_{k,l} &= \bar{p}_{k,l} + \frac{\lambda}{4(1-\lambda)}(I_{k,l} - R(k,l,\bar{z}_{k,l},\bar{p}_{k,l},\bar{q}_{k,l}))\frac{\partial R}{\partial p_{k,l}} \mid_{\bar{p}_{k,l}} \\
q_{k,l} &= \bar{q}_{k,l} + \frac{\lambda}{4(1-\lambda)}(I_{k,l} - R(k,l,\bar{z}_{k,l},\bar{p}_{k,l},\bar{q}_{k,l}))\frac{\partial R}{\partial q_{k,l}} \mid_{\bar{q}_{k,l}} \\
z_{k,l} &= \bar{z}_{k,l} + \frac{\lambda}{4(1-\lambda)}(I_{k,l} - R(k,l,\bar{z}_{k,l},\bar{p}_{k,l},\bar{q}_{k,l}))\frac{\partial R}{\partial z_{k,l}} \mid_{\bar{z}_{k,l}}
\end{aligned} \tag{A.8}$$

Appendix B

A comparison between Yamaguchi et al's model and ours

In Fig. B.1, Yamaguchi et al's use the camera head as a reference coordinates in their hand-eye calibration system. Since surgeons rotate the scope cylinder with respect to the camera head in order to view sideways, it is a natural way to consider the camera head as a reference. However it makes the cameral model very complex. To think it in an opposite way, no matter how surgeons rotate the scope cylinder, if the reference coordinates is on the cylinder, the lens system is fixed with respect to the cylinder but the camera head rotates by θ. Thus the external parameters are not affected by the rotation anymore. Since the image plane is in the camera head, the rotation only affects the image plane. Our method is therefore developed based on above observation. Yamaguchi et al.'s model needs five additional parameters but we only need one. They use two optical markers and one rotary encoder, and we only need two optical markers.

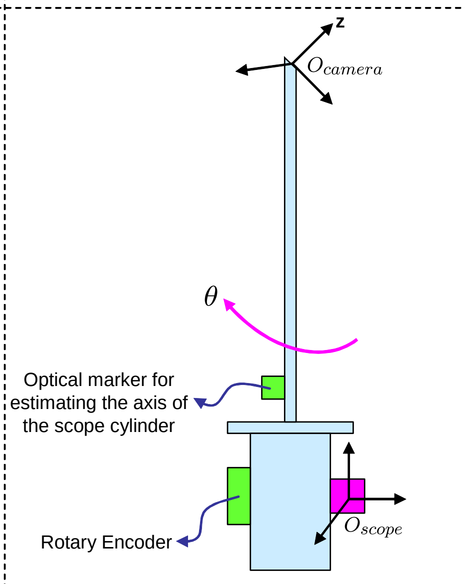
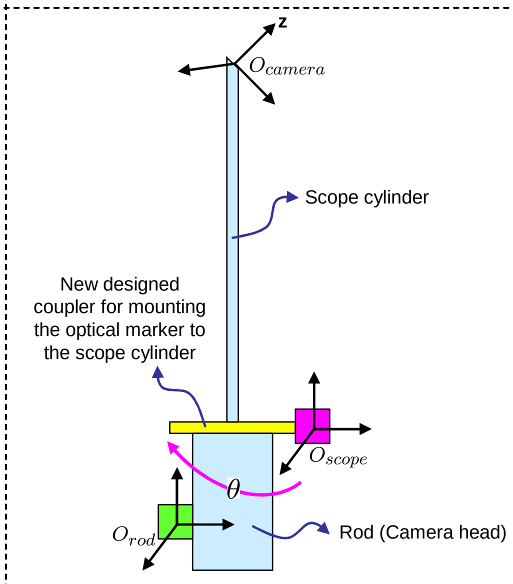
	Ymaguchi et al.'s system	Our system
System		
Reference part	Rod (Camera head)	Scope cylinder
Rotated part	Scope cylinder	Rod (Camera head)
Fact	1. Lens system is rotated around the scope cylinder by θ 2. Image plane is fixed	1. Image plane is rotated around the principal point by θ 2. Lens system is fixed
Extra Transformation	1. Rotate the scope cylinder around its axis by θ 2. Inversely rotate the image plane around z-axis of the lens system by θ	1. Rotate the image plane around the principal point by θ
Unknown parameters	1. θ 2. Axis of the scope cylinder 3. Axis of the lens system	1. θ 2. Principal point

Figure B.1: A comparison between Yamaguchi et al.'s system and ours. In Yamaguchi et al.'s system, the camera head is tracked such that the transformation from the marker to the lens system is not fixed but depends on the rotation angle θ. Let the marker coordinates be the reference, the lens system is rotated around the scope cylinder by θ, but the image plane (in the camera head) remains still. They use two additional transformations to describe the effect of the rotation, which results in a complex model. Moreover, they need to calibrate the axis of both the scope cylinder and the lens system by using another optical maker attached to the scope cylinder. Based on our observation, it is possible to simplify this model if we fix the transformation between the marker and the lens system. We design a coupler that enables the mounting of an optical marker onto the scope cylinder. We let the marker coordinates be the reference, thus the lens system is fixed. The rotation only affects the image plane while the camera head is rotating around the cylinder (reference). And the image plane only rotates around the principal point. As a result, we come up with a simple model (see details in the text).

Appendix C

Initial Alignment for Femur Data

As Fig. C.1 shows, the bottom portion of the femur is used as an example. The surface Y has more femur shaft, but less shaft remains on the the surface X. If we directly align both centers, which is the same as the previous work [11] did, experiments show that the registration process will be very slow and may not converge in several cases. The reason is that a portion of the surface Y (highlighted in blue in Fig. C.1) has no counterpart on the surface X. In order to improve upon the process, we estimate the pseudo center of Y instead of the true center. After that the pose of two surfaces are estimated and aligned.

The height of X is used to estimate the pseudo height of Y. Assuming the axis $\mathcal{Z}$ is along the scan direction from the knee to hip, we selected points bounded by the pseudo height of Y (denoted by black in Fig. C.1) to estimate the pseudo center $\kappa_{Y'}$ and the covariance matrices for the point set $\{\mathbf{p}_X\}$ and $\{\mathbf{p}_{Y'}\}$:

$$\begin{aligned} \kappa_X &= \tfrac{1}{N_X} \sum \mathbf{p}_X \\ \kappa_{Y'} &= \tfrac{1}{N_{Y'}} \sum \mathbf{p}_{Y'} \end{aligned} \tag{C.1}$$

$$\begin{aligned} \Psi_X &= \tfrac{1}{N_X - 1}[\mathbf{p}_X^1 - \kappa_X, \cdots, \mathbf{p}_X^{N_X} - \kappa_X] \cdot [\mathbf{p}_X^1 - \kappa_X, \cdots, \mathbf{p}_X^{N_X} - \kappa_X]' \\ \Psi_{Y'} &= \tfrac{1}{N_{Y'} - 1}[\mathbf{p}_{Y'}^1 - \kappa_{Y'}, \cdots, \mathbf{p}_{Y'}^{N_{Y'}} - \kappa_{Y'}] \cdot [\mathbf{p}_{Y'}^1 - \kappa_{Y'}, \cdots, \mathbf{p}_{Y'}^{N_{Y'}} - \kappa_{Y'}]' \end{aligned} \tag{C.2}$$

where $N_{Y'}$ is the number of points $\{\mathbf{p}_{Y'}\}$ in Y which satisfy $(z_Y - \min z_Y) < (\max z_X - \min z_X)$. We can solve for the principle axes by decomposing the covariance matrix using the moment analysis:

$$\Psi_X = \mathbf{U}_X \Lambda_X \mathbf{U}_X{}',\ \Psi_{Y'} = \mathbf{U}_{Y'} \Lambda_{Y'} \mathbf{U}_{Y'}{}' \tag{C.3}$$

Each column of $\mathbf{U}_X$ represents a principle axis of the point set $\{\mathbf{p}_X\}$, and $\mathbf{U}_{Y'}$ for $\{\mathbf{p}_{Y'}\}$. As Fig. C.1 shows, three axes represent the pose of the point set: red for $\{\mathbf{p}_X\}$ and green

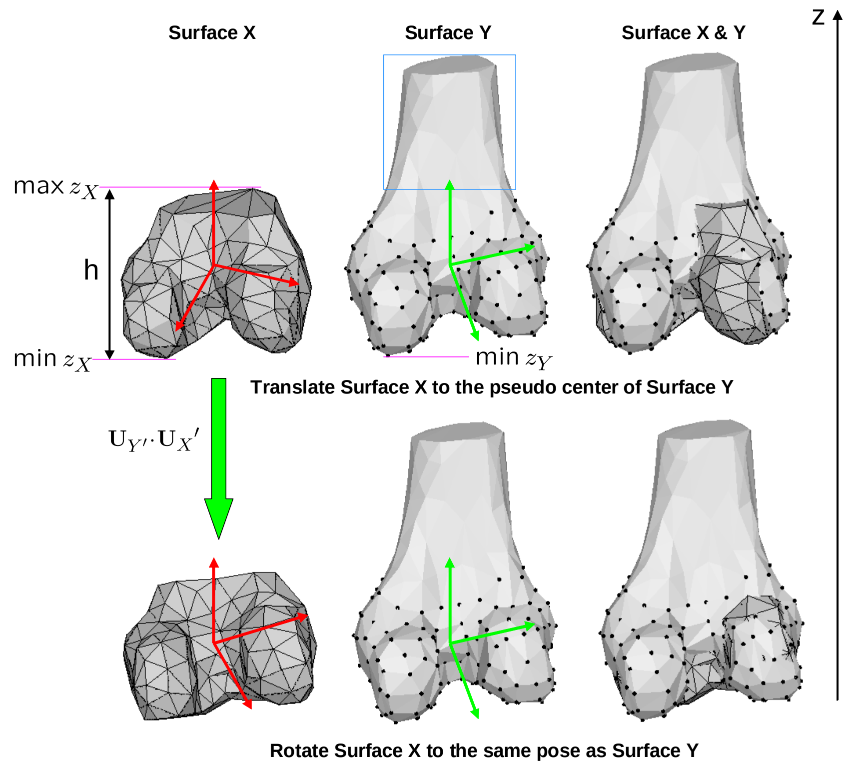


Figure C.1: Initial alignment between the surface X and Y. The first row compares the translated X to Y. Black points on Y are used to compute the pseudo center. The second row compares the translated and rotated X to Y. Red axes represent the pose of X, green axes for Y.

for $\{\mathbf{p}_{Y'}\}$. The rotation from coordinates of X to Y' is given by $\mathbf{U}_{Y'} \cdot \mathbf{U}_X{}'$. We apply a rigid transformation $[\mathbf{U}_{Y'} \cdot \mathbf{U}_X{}' | (\kappa_{Y'} - \kappa_X)]$ to the point set $\{\mathbf{p}_X\}$ and the two point are aligned.

Appendix D

Spine Vertebra Training Images

Figs. D.1-D.6. show CT scans of three artificial spines (labeled by “$JanetSpine$”, “$Cadaver\_One$”, “$Cadaver\_2$”) and three real spines (labeled by “$Icaos\_Spine$”, “Sam_Spiney”, and “A_To_G”).

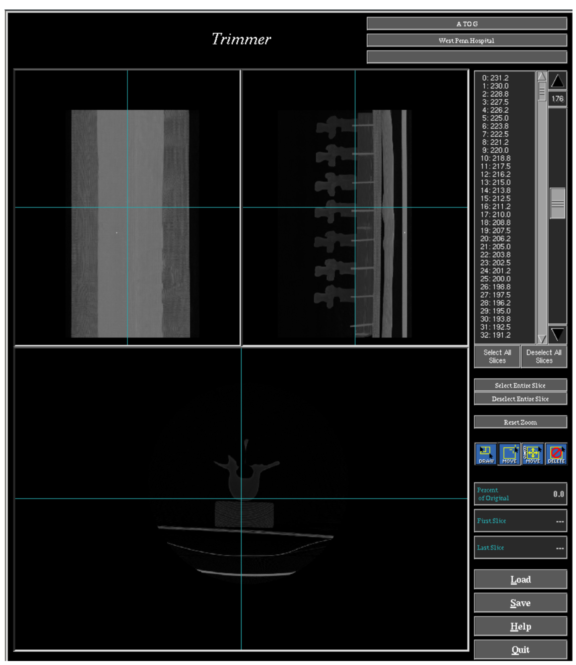


Figure D.1: Spine CT image: A_To_G, which is an artificial spine with human size.

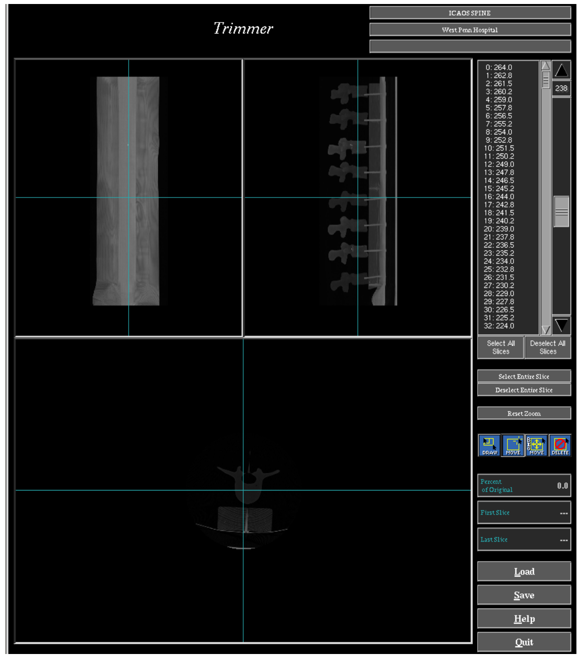


Figure D.2: Spine CT image: $Icaos$, which is an artificial spine with human size.

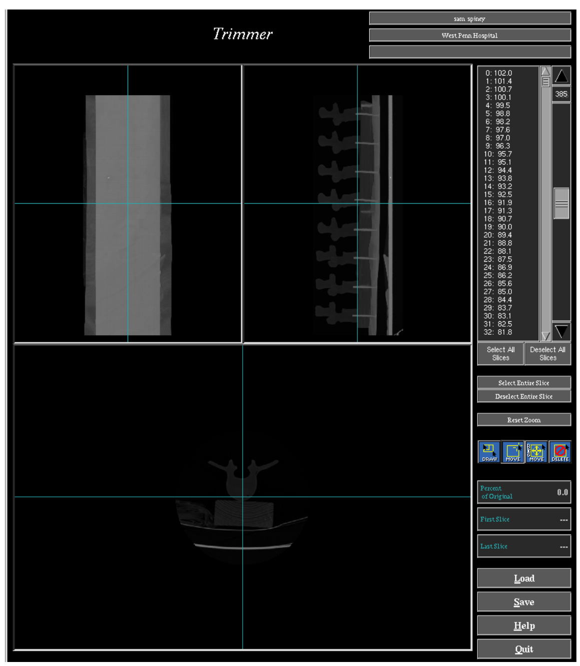
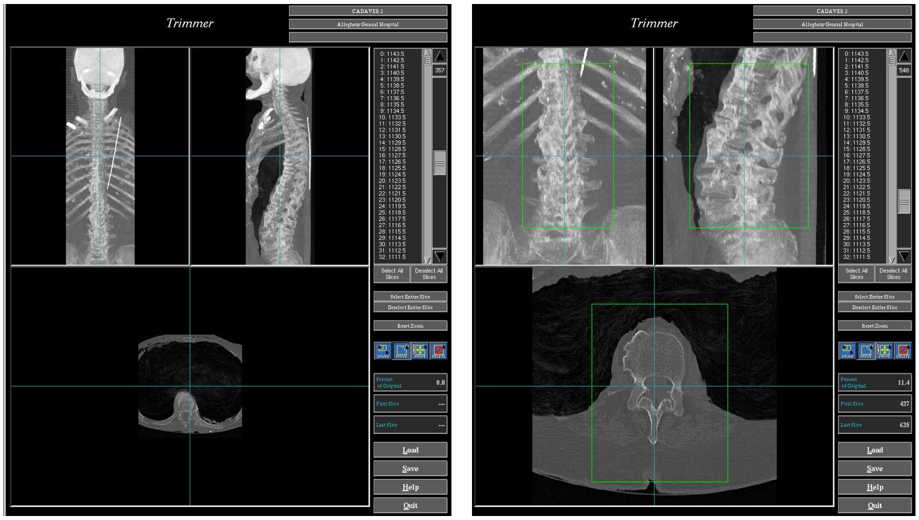
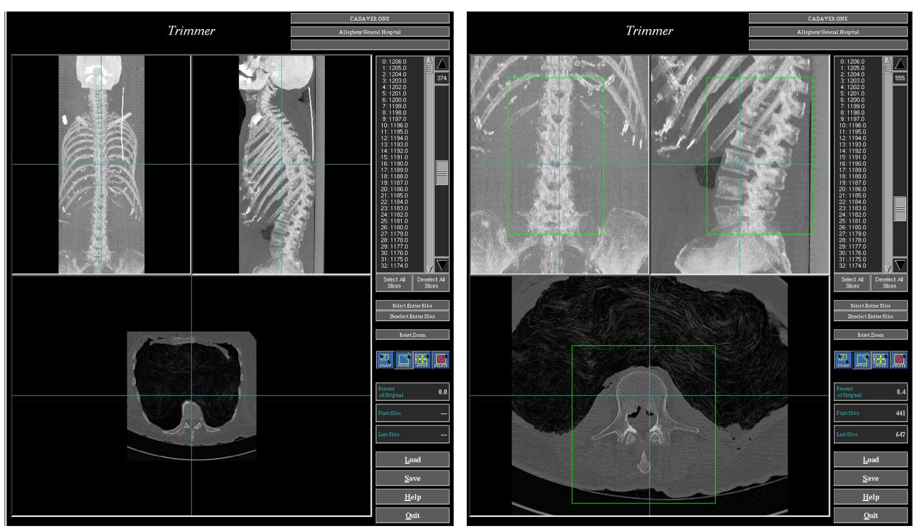


Figure D.3: Spine CT image: Sam, which is an artificial spine with human size.



(a) (b)

Figure D.4: Spine CT image: $Cadaver\_2$, which is from a cadaver.



(a) (b)

Figure D.5: Spine CT image: $Cadaver\_One$, which is from a cadaver.

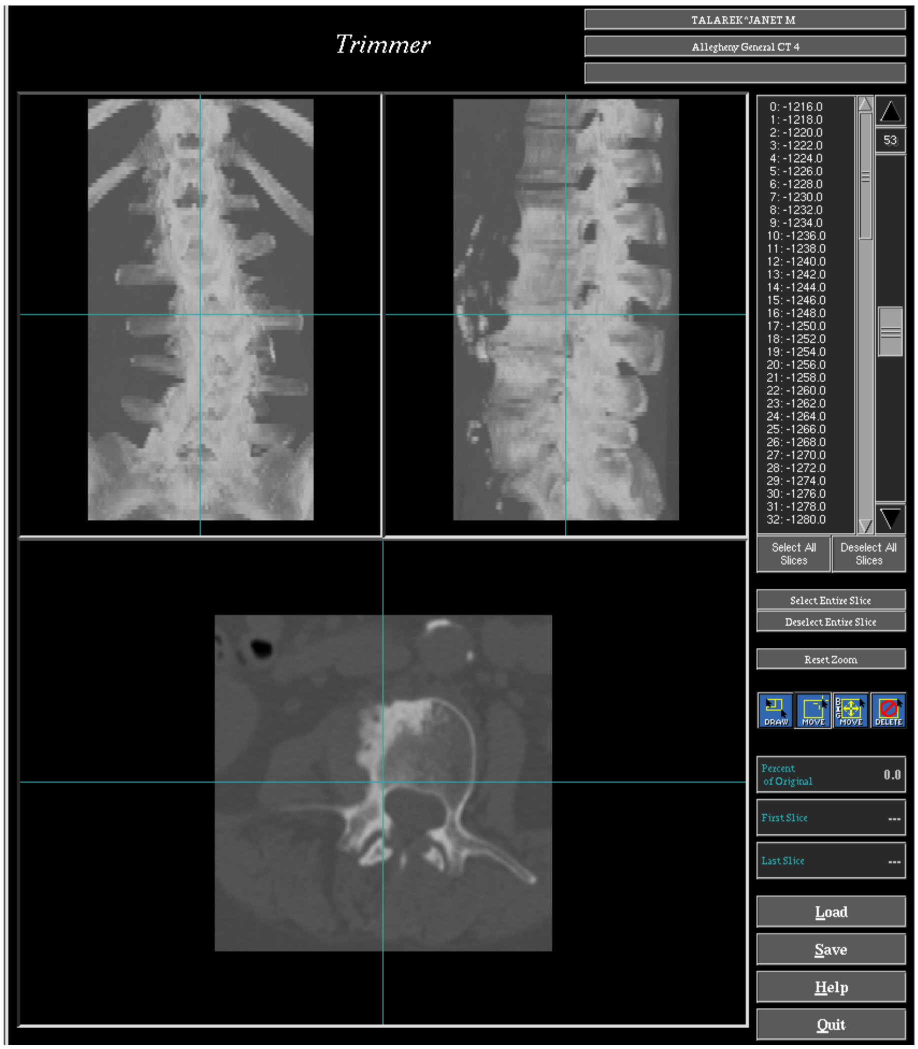


Figure D.6: Spine CT image: *Janet*, which is from a patient.

Appendix E

Instruction for Selecting Points from Spine CT Images

For people who have no anatomical background of spine and vertebrae, it is difficult for them to select appropriate surface points on individual vertebrae from CT images. In this appendix we use several examples to illustrate how to select points from different views. Fig. E.1 shows an example from the coronal view. Fig. E.2 shows an example from the sagittal view. Fig. E.3 shows an example from the transverse view.

Since Th12 is a bit different from lumbar vertebrae, we need to be careful when labeling Th12. Fig. E.4 shows the difference between Th12 and L2 from the transverse view. (a) shows L2, (b) and (c) show Th12. The area highlighted in red in (c) does not belong to Th12 but is often mis-assigned to Th12.

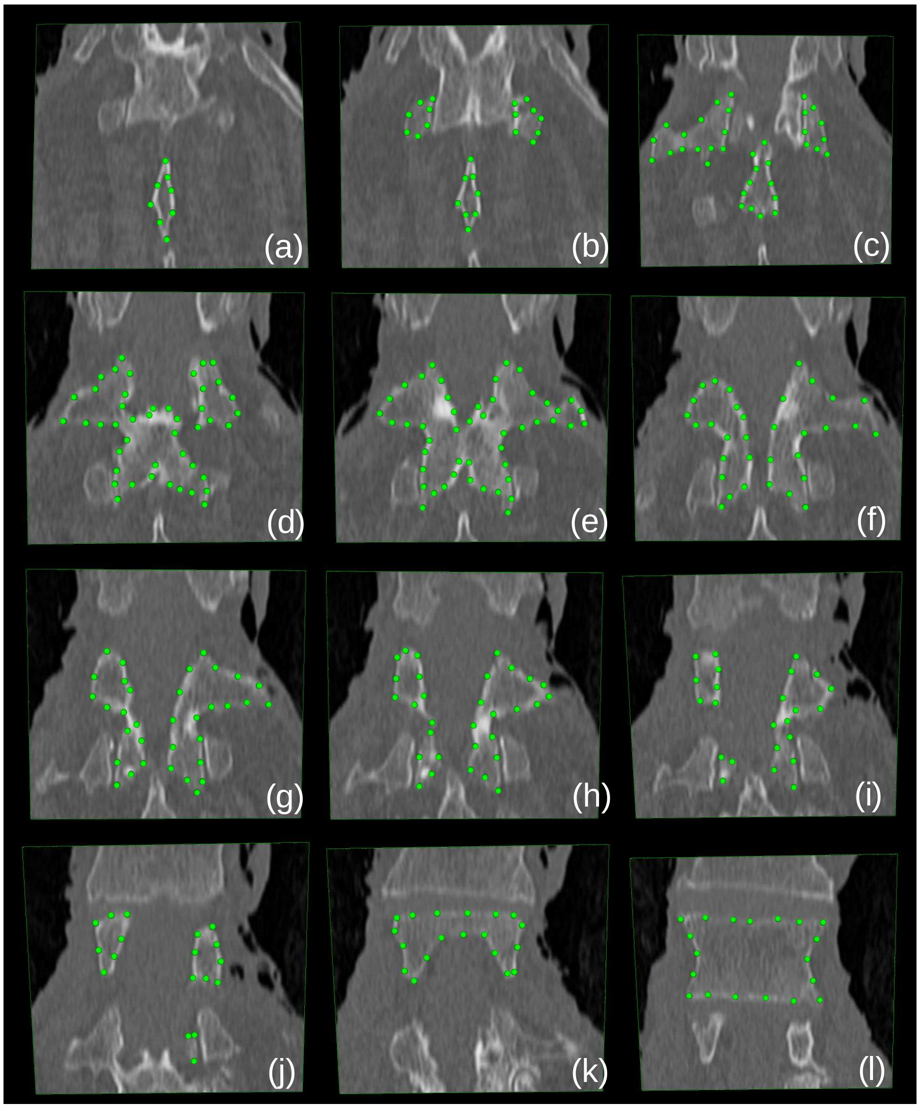


Figure E.1: Examples for selecting from the coronal view. (a)-(f) show different positions from the back to the front.

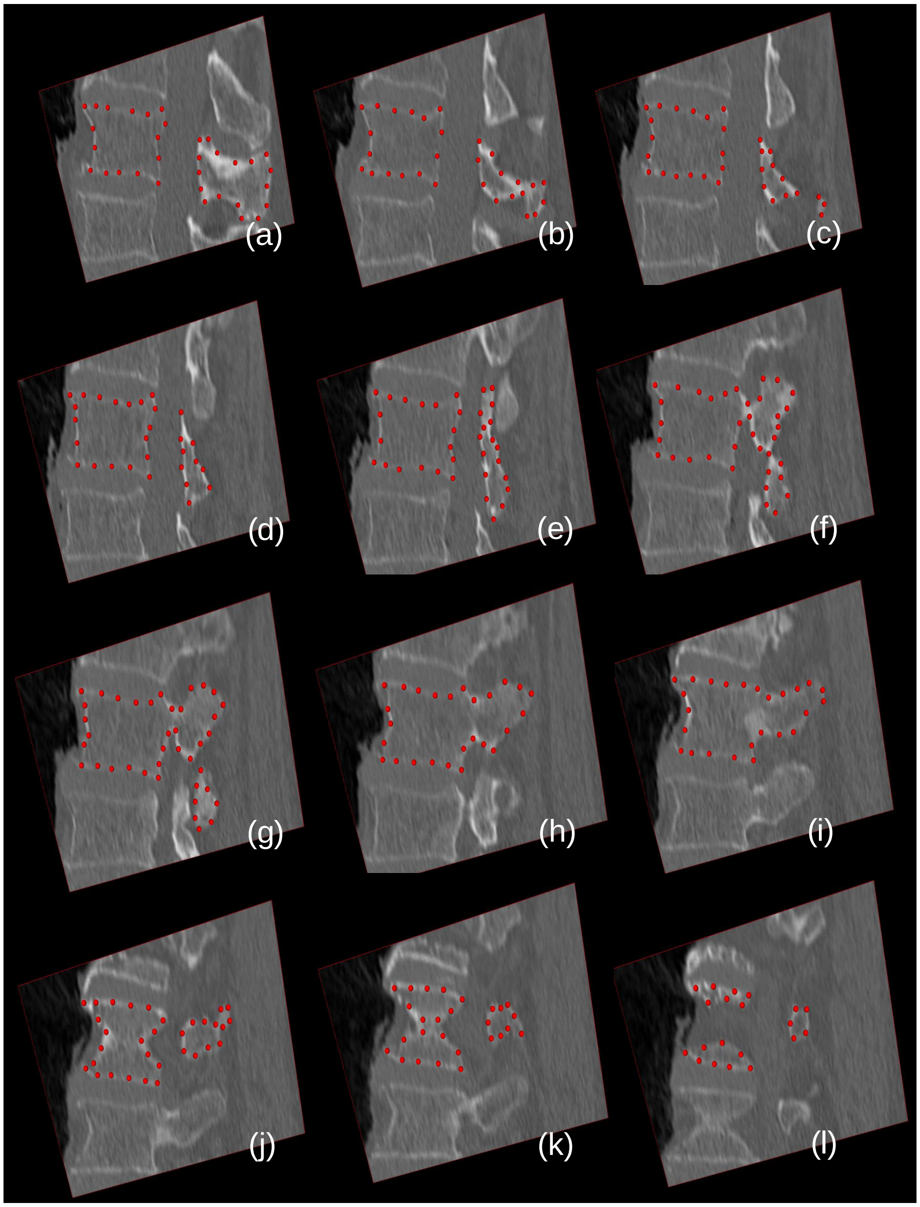


Figure E.2: Examples for selecting from the sagittal view. (a)-(l) show different positions from the middle to the left.

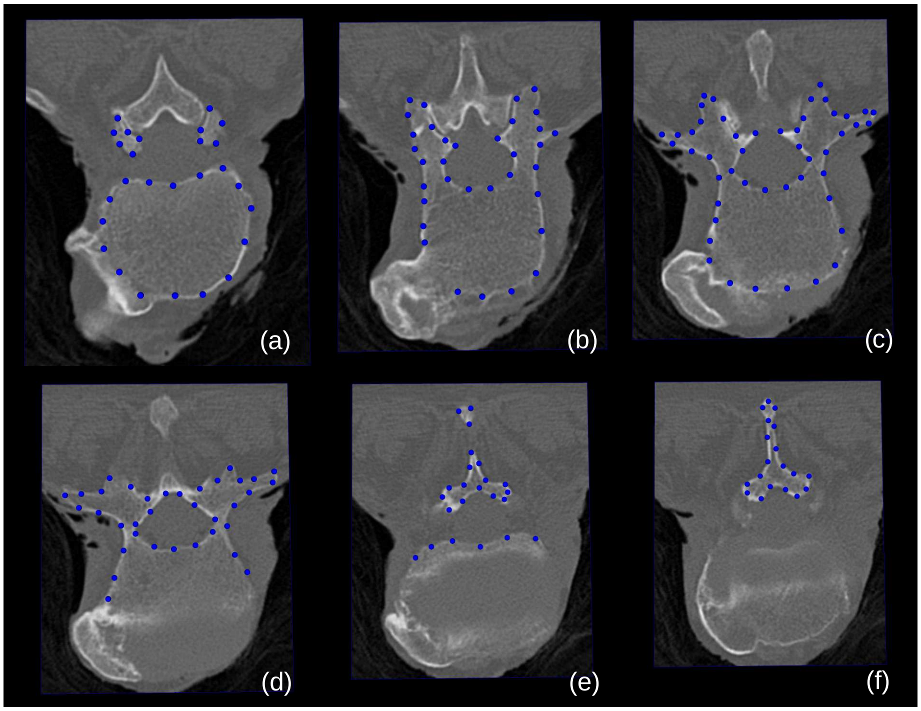


Figure E.3: Examples for selecting from the transverse view. (a)-(l) show different positions from the top to the bottom.

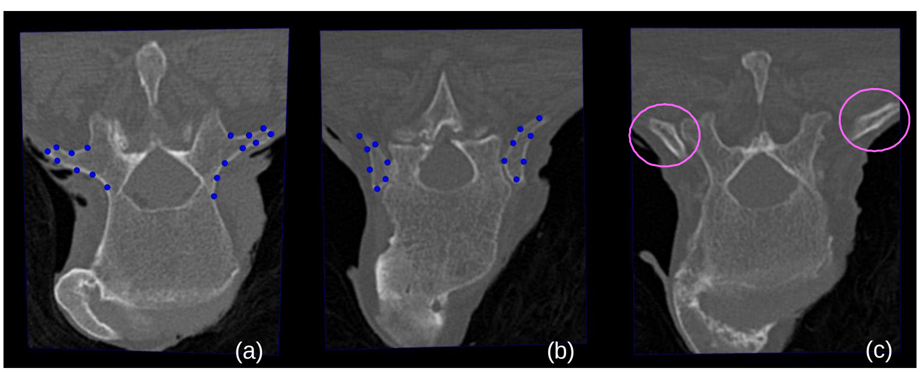


Figure E.4: An illustration of the the difference between Th12 L2 vertebra from the transverse view. (a) shows L2, (b) and (c) show Th12. The area highlighted in red in (c) does not belong to Th12 but is often mis-assigned to Th12.

Bibliography

[1] N. Aspert, D. Santa-Cruz, and T. Ebrahimi. Mesh: measuring errors between surfaces using the hausdorff distance. *Proceedings of IEEE International Conference on Multimedia and Expo (ICME'02)*, 1:705–708, 2002.

[2] A. Aziz and M. Uetani. Introduction to mri - history. *http://www.med.nagasaki-u.ac.jp/radiolgy/MRIhistory.html*, 2002.

[3] R. Bajcsy and S. Kovacic. Multiresolution elastic matching. *Proceedings of IPPR Conference on Computer Vision, Graphics and Image Processing (CVGIP'89)*, 46:1–21, 1989.

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

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

[6] C. Broit. *Optimal registration of deformed images*. PhD thesis, Computer and Information Science Department, Univ. of Pennyslvania.

[7] D. Burschka, M. Li, R. Taylor, and G.D. Hager. Scale-invariant registration of monocular stereo images to 3d surface models. *Proceedings of Intelligent Robots and Systems (IROS'04)*, 3:2581–2586, 2004.

[8] J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. Mccallum, and T. R. Evans. Reconstruction and representation of 3d objects with radial basis functions. In *Proceedings of ACM Annual Conference on Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH'01)*, pages 67–76, 2001.

[9] J. C. Carr, W. R. Fright, and R. K. Beatson. Surface interpolation with radial basis functions for medical imaging. *IEEE Transactions on Medical Imaging*, 16(1):96–107, 1997.

[10] M. Chen. *3-D Deformable registration using a statistical atlas with applications in medicine*. PhD thesis, Robotics Institute, School of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, 1999.

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

[12] H. Chui, A. Rangarajan, J. Zhang, and C. M. Leonard. Unsupervised learning of an atlas from unlabeled point-sets. *IEEE Transactions on Patter Analysis and Machine Intelligence*, 2.

[13] M. J. Clarkson, D. Rueckert, A. P. King, P. J. Edwards, D. L. G. Hill, and D. J. Hawkes. Registration of video images to tomographic images by optimising mutual information using texture mapping. *Proceedings of the Second International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'99)*, LNCS 1679:579–588, 1999.

[14] T.F. Cootes, C. Beeston, G.J. Edwards, and C.J. Taylor. A unified framework for atlas matching using active appearance models. *IEEE Transactions on Patter Analysis and Machine Intelligence*, pages 322–333, 1999.

[15] T.F. Cootes and C.J. Taylor. Statistical models of appearance for medical image analysis and computer vision. *Proceedings of SPIE Image Processing*, 4322:238–248, 2001.

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

[17] F. Courteille, A. Crouzil, Jean-Denis Durou, and P. Gurdjos. Towards shape from shading under realistic photographic conditions. *Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)*, 2:277–280, 2004.

[18] B.M. Dawant. Non-rigid registration of medical images: purpose and methods, a short survey. *Proceedings of IEEE International Symposium on Biomedical Imaging*, pages 465–468, 2002.

[19] K. Deguchi, T. Sasano, H. Arai, and H. Yoshikawa. 3d shape reconstruction from endosope iamge sequences by the factorization method. *IEICE Transactions on Information and Systems*, E79-D(9):1329–36, 1996.

[20] M. Desbrun, M. Meyer, P. Schroder, and A.H. Barr. Implicit fairing of irregular meshes using diffusion and curvature flow. *Proceedings of ACM Annual Conference on Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH'99)*, pages 317–324, 1999.

[21] D. Dey, D. G. Gobbi, P. J. Slomka, K. J. Surry, and T. M. Peters. Automatic fusion of freehand endoscopic brain images to three-dimensional surfaces: creating stereoscopic panoramas. *IEEE Transactions on Medical Imaging*, 21(1):23–30, 2002.

[22] D. Dey, D. G. Gobbi, P. J. Slomka, K. J. M. Surry, and T. M. Peters. Automatic fusion of freehand endoscopic brain images to three-dimensional surfaces: Creating stereoscopic panoramas. *IEEE Transactions on Medical Imaging*, 21(1):23–30, 2002.

[23] I. Donald, J. MacVicar, and T. G. Brown. Investigation of abdominal masses by pulsed ultrasound. *The Lancet*, 1:1188–95, 1958.

[24] N.C. Durham. 3d ultrasound device poised to advance minimally invasive surgery. *http://www.pratt.duke.edu/news/?id=417*, 2006.

[25] Jean-Denis Durou, Maurizio Falcone, and Manuela Sagona. Numerical methods for shape-from-shading: A new survey with benchmarks. *Computer Vision and Image Understanding*, 109(1):22–43, 2008.

[26] L. M. Ellingsen and J. L. Prince. Deformable registration of ct pelvis images using mjolnir. *Proceedings of SPIE Conference on Medical Imaging*, 6144:329–337, 2006.

[27] G Farin. *Curves and surfaces for CAGD*. New York, 1993.

[28] M. Fleute, S. Lavallée, and L. Desbat. Integrated approach for matching statistical shape models with intra-operative 2d and 3d data. *Proceedings of the Fifth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'02)*, 2:364–372, 2002.

[29] C. H. Q. Forster and C. L. Tozzi. Toward 3d reconstruction of endoscope images using shape from shading. *Proceedings of the 13th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPHI'00)*, pages 90–96, 2000.

[30] R. T. Frankot and R. Chellappa. A method for enforcing integrability in shape from shading algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):439–451, 1988.

[31] H. Fuchs, M. A. Livingston, R. Raskar, D. Colucci, K. Keller, A. State, J. R. Crawford, P. Rademacher, S. H. Drake, and A. A. Meyer. Augmented reality visualization for laparoscopic surgery. *Proceedings of the First International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'98)*, LNCS 1496:934–943, 1998.

[32] M. Garland and P.S. Heckbert. Surface simplification using quadric error metrics. *Proceedings of ACM Annual Conference on Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH'97)*, pages 209–216, 1997.

[33] J. Glaunes, A. Trouvé, and L. Younes. Diffeomorphic matching of distributions: a new approach for unlabelled point-sets and sub-manifolds matching. *Proceedings of International Conference on Computer Vision and Pattern Recognition (CVPR'04)*, 2:712–718, 2004.

[34] D. Goryn and S. Hein. On the estimation of rigid body rotation from noisy data. *IEEE Transactions on Patter Analysis and Machine Intelligence*, 17(2):1219–1220, 1995.

[35] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2nd edition, 2004.

[36] J. K. Hasegawa and C. L. Tozzi. Shape from shading with perspective projection and camera calibration. *Computers and Graphics*, 20(3):351–364, 1996.

[37] M. Hayashibe, N. Suzuki, and Y. Nakamura. Laser-scan endoscope system for intraoperative geometry acquisition and surgical robot safety management. *Medical Image Analysis*, 10:509–519, 2006.

[38] D.L.G. Hill, P.G. Batchelor, M. Holden, and D.J. Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46(3):1–45, 2001.

[39] Dongho Hong, Wallapak Tavanaponq, Johnny Wong, JungHwan Oh, and Piet C. de Groen. 3d reconstruction of colon segments from colonoscopy images. *Proceedings of the Ninth IEEE International Conference on Bioinformatics and Bioengineering*, pages 53–60, June 2009.

[40] L. Hong, A. Kaufamn, Y-C. Wei, A. Viswambharan, M. Wax, and Z. Liang. 3d virtual colonoscopy. *IEEE Symposium on Frontiers in Biomedical Visualization*, pages 26–32, 1995.

[41] B. K. P. Horn and M. J. Brooks. The variational approach to shape from shading. *Computer Vision, Graphics, and Image Processing (CVGIP'86)*, 33(2):174–208, 1986.

[42] B. K. P. Horn and M. J. Brooks. *Shape from Shading*. The MIT Press, Cambridge, Massachusetts, 1989.

[43] B. K. P. Horn and R. W. Sjoberg. Calculating the reflectance map. *Applied Optics*, 18(11):1770–1779, 1979.

[44] K. Ikeuchi and B. K. P. Horn. Numerical shape from shading and occluding boundaries. *Artificial Intelligence*, 17(1-3):141–184, 1981.

[45] A. K. S. Iyengar, H. Sugimoto, D. B. Smith, and M. S. Sacks. Dynamic in vitro quantification of bioprosthetic heart valve leaflet motion using structured light projection. *Annals of Biomedical Engineering*, 29(11):963–973, 2001.

[46] S. Jaume, B. Macq, and S. K. Warfield. Labeling the brain surface using a deformable multiresolution mesh. *Proceedings of the Fifth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'02)*, pages 451–458, 2002.

[47] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal on Computer Vision*, 1(4):321–331, 1988.

[48] A. Kaufman and J. Wang. *Visualization in Medicine and Life Sciences*, chapter 3D Surface Reconstruction from Endoscopic Videos, pages 61–74. Mathematics and Visualization. Springer Berlin Heidelberg, 2007.

[49] M. R. Kaus, V. Pekar, C. Lorenz, R. Truyen, S. Lobregt, and J. Weese. Automated 3-d pdm construction from segmented images using deformable models. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(8):1005–1013, 2003.

[50] K. P. Keller, J. D. Ackerman, M. H. Rosenthal, H. Fuchs, and A. State. Methods and systems for real-time structured light depth extraction and endoscope using real-time structured light depth extraction. *http://www.patentstorm.us/patents/6503195-fulltext.html*, US Patent Issued on January 7, 2003.

[51] R. Kikinis, G. P. Langham, and F. A. Jolesz. Surgical planning using computer-assisted three-dimensional reconstructions. *Computer Integrated Surgery*, pages 147–154, 1995.

[52] J. Kim and J. A. Fessler. Intensity-based image registration using robust correlation coefficients. *IEEE Transactions on Patter Analysis and Machine Intelligence*, 23(11):1430–1444, 2004.

[53] Ron Kimmel and James A. Sethian. Optimal algorithm for shape from shading and path planning. *Journal of Mathematical Imaging and Vision*, 14(3):237–244, 2001.

[54] R. Kozera. Existence and uniqueness in photometric stereo. *Applied Maths and Computation*, 44(1):1–103, 1991.

[55] R. Kozera. On shape recovery from two shading patterns. *International Journal of Pattern Recognition and Artificial Intelligence*, 6(4):673–698, 1992.

[56] R. Kozera. An overview of the shape from shading problem. *Machine Graphics and Vision*, 7(1):291–312, 1998.

[57] R. Kozera and L. Noakes. Noise reduction in photometric stereo with non-distant light sources. *Proceedings of the International Conference on Computer Vision and Graphics (ICCVG'04)*, 32:103–110, 2004.

[58] Y. G. Leclerc and A. F. Bobick. the direct computation of height from shading. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'91)*, pages 552–558, 1991.

[59] K. M. Lee and C.-C. J. Kuo. Shape from shading with perspective projection. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 59(2):202–212, 1994.

[60] K. M. Lee and C.-C. J. Kuo. Shape from shading with a generalized reflectance map model. *Computer Vision and Image Understanding*, 67(2):143–160, 1997.

[61] A. Litvinov and Y. Y. Schechner. Addressing radiometric nonidealities: a unified framework. *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2:52–59, 2005.

[62] W. E. Lorensen, F. A. Jolesz, and R. Kikinis. The exploration of cross-sectional data with a virtual endoscope. *Interactive Technology and the New Paradigm for Healthcare*, pages 221–230, 1995.

[63] C. R. Maurer, J. M. Fitzpatrick, M. Y. Wang, R. L. Galloway, R. J. Maciunas, and G. G. Allen. Registration of head volume images using implatable fiducial markers. *IEEE Transactions on Patter Analysis and Machine Intelligence*, 16:447–462, 1997.

[64] T. McInerney and D. Terzopoulos. Deformable models in medical image analysis: a survey. *Medical Image Analysis*, 1(2):91–10, 1996.

[65] S. Meller and W.A. Kalender. Building a statistical shape model of the pevis. *International Congress Series*, 1268:561–566, 2004.

[66] F. Mourgues, F. Devernay, G. Malandain, and È. Coste-Manière. 3d reconstruction of the operating field for image overlay in 3d-endoscopic surgery. *Proceedings of the IEEE and ACM International Symposium on Augmented Reality (ISAR'01)*, pages 191–192, 2001.

[67] F. Mourgues, F. Devernay, and G. Malandain nd E. Coste-Manière. 3d reconstruction of the operating field for image overlay in 3d-endoscopic surgery. *Proceedings of IEEE and ACM International Symposium on Augmented Reality (ISAR'01)*, pages 191–194, 2001.

[68] T. Okatani and K. Deguchi. Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center. *Computer Vision and Image Understanding*, 66(2):119–131, 1997.

[69] X. Papademetris, A.P. Jackowski, R.T. Schultz, L.H. Staib, and J.S. Duncan. Nonrigid brain registration using extended robust point matching for composite multisubject fmri analysis. *Proceedings of the Sixth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'03)*, 2879:788–795, 2003.

[70] J. Park, D. Mataxas, A. Young, and L. Axel. Deformable models with parameter functions for cardiac motion analysis from tagged mri data. *IEEE Transactions on Medical Imaging*, 15:278–289, 1996.

[71] M. A. Penna. A shape from shading analysis for a single perspective image of a polyhedron. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):545–554, 1989.

[72] F. Pighin, J. Hecker, D. Lischnski, R. Szeliski, and D.H. Salesin. Synthesizing realistic facial expressions from photographs. *Proceedings of ACM Annual Conference on Special Interest Group on Graphics and Interactive Techniques (SIGGRAPH'98)*, pages 75–84, 1998.

[73] C. J. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(3):206–218, 1997.

[74] M. Pollefeys, R. Koch, and L. V. Gool. Self-calibration and metric reconstruction inspite of varying and unknown intrinsic camera parameters. *International Journal of Computer Vision*, 32(1):7–25, 1999.

[75] E. Prados and O. Faugeras. “perspective shape from shading” and viscosity solution. *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV’03)*, 2:826–831, 2003.

[76] E. Prados and O. Faugeras. Shape from shading: a well-posed problem? *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, 2:870–877, 2005.

[77] R. A. Robb and C. Barillot. Interactive display and analysis of 3-d medical images. *IEEE Transactions on Medical Imaging*, 8(3):217–226, 1989.

[78] R. A. Robb and D. P. Hanson. The analyze software system for visualization and analysis in surgery simulation. *Computer Integrated Surgery*, 8(3):217–226, 1989.

[79] T. Rohlfing, D.B. Russakoff, and C.R. Maurer. Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. *IEEE Transactions on Medical Imaging*, 23(8):983–994, 2004.

[80] D. Rueckert, A.F. Frangi, and J.A. Schnabel. Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(8):1014–1025, 2003.

[81] D. Samaras and D. N. Metaxas. Coupled lighting direction and shape estimation from single images. *Proceedings of the International Conference on Computer Vision (ICCV’99)*, 2:868–874, 1999.

[82] R. M. Satava and R. A. Robb. Virtual endocopy: Application of 3-d visualization to medical diagnosis. *Presence*, 6(2):179–197, 1997.

[83] S. Seshamani, W. Lau, and G. Hager. Real-time endoscopic mosaicking. *Proceedings of the Ninth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI’06)*, LNCS 4190:355–363, 2006.

[84] R. Shahidi, M. R. Bax, C. R. Maurer, J. A. Johnson, E. P. Wilkinson, B. Wang, J. B. West, M. J. Citardi, K. H. Manwaring, and R. Khadem. Implementation, calibration and accuracy testing of an image-enhanced endoscopy system. *IEEE Transactions on Medical Imaging*, 21(12):1524–1535, 2002.

[85] C. Shen, D. Davatzikos. Hammer: hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):1421–1439, 2002.

[86] H. Shinichi, M. Daichi, N. Toshiya, T. Norimichi, and M. Yoichi. Reconstruction of 3d organ image using endoscope with magneto-position-sensor. *IEIC Technical Report (Institute of Electronics, Information and Communication Engineers, Japan)*, 106(145):13–18, 2006.

[87] J. Simon and T Richard. Introduction to ct physics. *Cross-Sectional Imaging Made Easy "Made Easy" Series, Chapter 1*, 2004.

[88] D. Stoyanov, A. Darzi, and G. Z. Yang. A practical approach towards accurate dense 3d depth recovery for robotic laparoscopic surgery. *Computer Aided Surgery*, 10(4):199–208, 2005.

[89] M. Straka, A.L. Cruz, L.I. Dimitrov, M. Sramek, D. Fleischmann, and E. Groller. Bone segmentation in ct-angiography data using a probabilistic atlas. *Proceedings of Vision, Modeling, Visualization (VMV'03)*, pages 505–512, 2003.

[90] G. Subsol, J.P. Thirion, and Ayache N. A general scheme for automatically building of anatomical atlas: application to a skull atlas. *INRIA, France 2586*, May 1995.

[91] Ayache N. Subsol G., Thirion J.P. First steps towards automatic building of anatomical atlas. *INRIA, France 2216*, March 1994.

[92] R. Szeliski. Bayesian modeling of uncertainty in low-level vision. *International Journal on Computer Vision*, 5:271–301, 1990.

[93] R. Szeliski and S. Lavallee. Matching 3-d anatomical surfaces with non-rigid deformations using octree-splines. *International Journal on Computer Vision*, 18:171–186, 1996.

[94] A. Tankus, N. Sochen, and Y. Yeshurun. A new perspective [on] shape-from-shading. *Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03)*, 2:862–869, 2003.

[95] Ariel Tankus, Nir Sochen, and Yehezkel Yeshurun. Reconstruction of medical images by perspective shape-from-shading. *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04)*, 3:778 – 781, 2004.

[96] Ariel Tankus, Nir Sochen, and Yehezkel Yeshurun. Shape-from-shading under perspective projection. *International Journal of Computer Vision*, 63(1):21–43, 2005.

[97] T. Thormahlen, H. Broszio, and P. N. Meier. Three-dimensional endoscopy. *http://citeseer.ist.psu.edu/thormaehlen02threedimensional.html*, 2002.

[98] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, 3:323–344, 1987.

[99] D. J. Vining and D. W. Gelfand. Noninvasive colonoscopy using helical ct scanning, 3d reconstruction, and virtual reality (abstr.). *SGR Scientific Program*, 70, 1994.

[100] D. J. Vining, R. Y. Shifrin, and Grishaw E. K. Virtual colonoscopy. *Radiology*, 193:446, 1994.

[101] W. M. Wells, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.

[102] C. Wengert, P. C. Cattin, J. M. Duff, and G. Székely. Markerless endoscopic registration and referencing. *Proceedings of the Ninth International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'06)*, 4190:816–823, October 2006.

[103] C. Wengert, J. M. Duff, C. Baur, G. Székely, and P. C. Cattin. Fiducial-free endoscopic vertebra referencing. *International Meeting for Computer Assisted Orthopaedic Surgery (CAOS 2007)*, June 2007.

[104] J. Woo. A short history of the development of ultrasound in obstetrics and gynecology. *Obstetric Ultrasound http://www.ob-ultrasound.net/history1.html*, 2006.

[105] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980.

[106] R. P. Woods, J. C. Mazziotta, and S. R. Cherry. Mri-pet registration with automated algorithm. *Journal of Computer Assisted Tomography*, 17:536–546, 1994.

[107] Philip L. Worthington and Edwin R. Hancock. New constraints on data-closeness and needle map consistency for shape-from-shading. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(12):1250–1267, 1999.

[108] Chia-Hsiang Wu. *3D Reconstruction of Human Inner Structure from Video by Using Geometric Constraints*. PhD thesis, Computer Science and Information Engineering Department, National Cheng Kung University, Taiwan, 2007.

[109] T. Yamaguchi, M. Nakamoto, Y. Sato, K. Konishi, M. Hashizume, N. Sugano, H. Yoshikawa, and S. Tamura. Development of a camera model and calibration procedure for oblique-viewing endoscopes. *Computer Aided Surgery*, 9(5):203–214, 2004.

[110] Lei Yang and Jiu qiang Han. 3d shape reconstruction of medical images using a perspective shape-from-shading method. *Measurement Science and Technology*, 19(065502), 2008.

[111] S. Y. Yeung, H. T. Tsui, and A. Yim. Global shape from shading for an endoscope image. *Proceedings of the Second International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'99)*, pages 328–32, 1999.

[112] G. Yngve and G. Turk. Robust creation of implicit surfaces from polygonal meshes. *IEEE Transactions on Vizualization and Computer Graphics*, 8(4):346–359, 2002.

[113] H. Yu. *Automatic rigid and deformable medical image registration*. PhD thesis, Mechanical Engineering Department, Worcester Polytechnic Institute, Worcester, MA.

[114] R. Zhang, P. S. Tsai, J. E. Cryer, and M. Shah. Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690–706, 1999.

[115] Z. Zhang. A flexible new technique for camera calibration. Technical Report MSR-TR-98-71, Microsoft Research, 1998.