Improving the Accuracy of 3D Reconstruction in Robotic Vision Applications

Yuanzheng Gong

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

University of Washington

2016

Reading Committee:

Eric J. Seibel, Chair

Blake Hannaford

Duane Storti

Program Authorized to Offer Degree:
Mechanical Engineering

University of Washington

Abstract

Improving the Accuracy of 3D Reconstruction
in Robotic Vision Applications

Yuanzheng Gong

Chair of the Supervisory Committee:
Research Professor Eric J. Seibel
Mechanical Engineering

Three-dimensional (3D) surface reconstruction is a process for retrieving the 3D shape and appearance of real objects or scenes. The generated 3D point clouds can be used in many fields, including entertainment, measurement, design, reverse engineering, homeland security and *etc.* Over the past decades, 3D reconstruction has been widely used in clinical diagnosis and surgical treatment of diseases, such as X-ray, ultrasound, computed tomography (CT) and magnetic resonance imaging (MRI). However, all of these technologies are radiography-based volumetric 3D reconstruction instead of 3D surface reconstruction. With the growing need of tissue texture information for clinical purposes, the 3D reconstruction based on endoscopic images plays a more vital role than ever, especially in tumor diagnosis and surveillance of esophagus, lung, stomach, bladder and etc.

In this work, new algorithms were developed to solve specific 3D reconstruction problems in biomedical applications. To reconstruct the 3D internal surface of a human organ, such as bladder or stomach, a sequence of 2D endoscopic images were captured by rotating and moving the scope around inside of the organs. This 3D reconstruction solely based on images is called Structure-from-Motion (SfM). To overcome the problems of insufficient features in medical images and short camera baselines, the camera poses were initially estimated by constraining the surface on a spherical shape at the first step. The more realistic organ surface was then reconstructed by releasing the spherical constraints. Extra features were

built to handle multiple scanning videos and recover the physical scale of the 3D surface with reference lesion target.

To reduce the human error and surgical operating time in removing the tumor/cancer in brain, a semi-automated surgical robotic system with 3D vision is being developed. By providing an accurate 3D surface model of the surgical field based on a RGB (red, green and blue) camera attached to the surgical tool, the robot could perform tedious operation of residual tumor tissue removal automatically. Camera position and orientation were also known throughout the surgery from the robotic system. This 3D reconstruction with known camera parameters is called Multi-view Stereo. Due to the mechanical limitation of robotic system, the camera pose parameters were with certain errors (tolerance). To utilize these inaccurate but bound constrained variables, Bound Constrained Bundle Adjustment (BCBA) algorithm was developed based on gradient projection to generate accurate 3D model efficiently.

Besides biomedical applications, 3D computer vision is emerging in traditional industries, such as manufacture and quality control applications. To build a potential in-line 3D metrology tool for internal threads in automobile engine blocks, two 3D reconstruction algorithms were developed with forward-view and side-view cameras, respectively. Axial-stereo vision algorithm was proposed to create dense 3D point cloud of internal surface based on two forward-view images that are aligned on the optical axis. Feature-based panoramic 3D registration algorithm was developed to register different side-view image-generated 3D surface patches together, by taking advantage of the robustness and accuracy of SIFT features. Each side-view patch of the repeated geometry of a threaded hole was reconstructed by multi-view stereo. Comparing with traditional 3D point clouds registration algorithm Iterative Closest Point (ICP), our algorithm has the advantages of high-efficiency and high-accuracy, especially for the registration of repetitive geometries.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Eric J. Seibel, for his continuous support, understanding and encouragement of my Ph.D. study and related research. Before joined HPL, he told me in the department open house day, paraphrased, "Life is short, I want to do something meaningful to the human society." At that moment, I made a decision to join his group. And I made the right choice. Throughout my entire Ph.D. life, his guidance helped me in all the time of research and writing of academic papers. I could not have imagined having a better advisor and mentor for my Ph.D. study. Beyond research subject, Eric also encouraged me to grow professionally as a scientist and engineer by shaping my research methodology. He gave me many opportunities to speak at international conferences and different public events. He put students' needs above his own. I will never forget the days we came to the lab at 6:30 am every Tuesday in 2015, giving weekly report to a company that funded us a highly-intensive 6-month industry project; I will never forget those sleepless nights we were working together before deadlines; I will never forget his help on revising my paper drafts with lots of patience in these 4 years.

I would like to thank the rest of my committee: Professor Blake Hannaford, Professor Linda Shapiro and Professor Duane Storti for their encouragement and insightful comments which widen my research from various perspectives.

My sincere thanks also goes to my fellow collaborator Danying Hu for her help, encouragement and friendship, and also other friends in UW BioRobotics lab, Yangming Li and Nava Aghdasi. Also I thank my friends Haixin Liu, Romi, Long Gui, Hao Mao, Sijing Shao, Evelyn, Hiromi, Muneaki, Zhou Yang, Yue Shi, Yanlin Peng, Seibel's family and other friends I have made and people who have impacted my life.

I am completely indebted to my family: my parents, brother and sisters, for their unending love, infinite generosity and understanding, for supporting me spiritually throughout my life in general. Without their support this achievement would not have been possible.

DEDICATION

TO MY PARENTS

Mr. Kaifu Gong and Mrs. Xiaofen He

Chapter 1

INTRODUCTION

1.1 3D Reconstruction

3D reconstruction is a process to retrieve the geometry and appearance of real objects or scene. The generated 3D point clouds can be utilized for visualization, game, music, movie, measurement, design, reverse engineering, human health, homeland security and *etc* [48, 148].

3D surface reconstruction of an object can be achieved by two main categories: active and passive [12]. The active approaches interfere with the object either mechanically (contact) or radiometrically (non-contact) [83]. An example of contact active 3D reconstruction method is using depth gauge to measure the distance of each surface point of the object, and collect these 3D data into a point cloud, such as Coordinate Measurement Machine (CMM) [14]. The biggest advantage of CMM is high precision and accuracy. But it comes with several drawbacks, such as high price, low portability and slow acquisition speed; also it may damage the sample since it is tactile measurement. Non-contact active methods are mainly using optical techniques. Time of flight (ToF) [13, 3, 116] is one non-contact method that can retrieve the depth between light source and object by measuring the time that light flies. It is usually used in laser radar or laser tracker system for medium-long range measurement, covering range of centimeters level up to kilometers. Comparing with other 3D reconstruction techniques, ToF is very fast, robust, but with low depth resolution. Another common active non-contact approach is based on structured-light [112, 36, 178]. By projecting structured-light of known pattern onto a 3D surface, such as grid, fringe, or speckle patterns, a distorted pattern image is captured from other perspective than where the projection light source is. The dense 3D information can be recovered by analyzing how the pattern is distorted.

Passive methods require multiple overlapped images of normally illuminated object with-

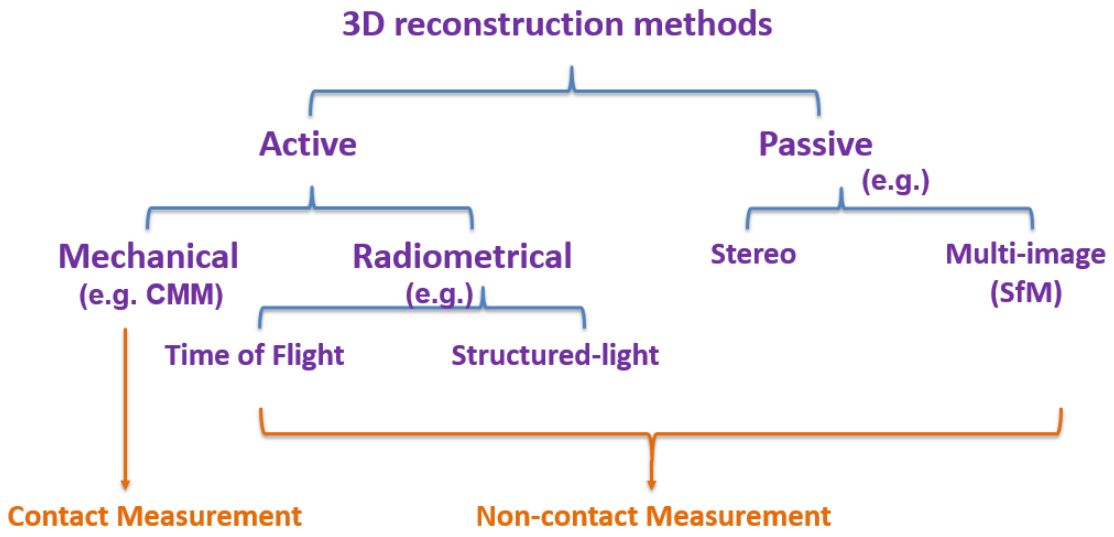


Figure 1.1: The categorization of the common 3D surface reconstruction methods.

out interfering or requirement of any artificial light. Passive methods are also can be considered as non-contact 3D reconstruction methods. One passive way to acquire the 3D information of an object is to perform stereo vision [123], which is motivated from human vision. A pair of well-calibrated cameras, locating side-by-side, captures the left and right images of the same object with image overlap. The 3D scene can be reconstructed with triangulation by finding the corresponding points in the stereo image pairs. Besides this traditional side-by-side stereo, axial-stereo vision system is developed to reconstruct the 3D model of internal surface by placing the two cameras along the optical axis [44]. Another typical passive 3D reconstruction method is feature-based reconstruction from multiple images, such as structure from motion (SfM) [155, 63]. The data acquisition speed can be as high as the camera reaches; thus, is easy to realize real-time 3-D data acquisition with low cost. However, hinging on identifying the corresponding pairs between two camera images, its accuracy is usually not high if the object surface does not have very strong reflectivity variations. Similar to stereo vision, SfM can be used to reconstruct the 3D scene by triangulation of the correspondences. While the difference is that the camera parameters could be unknown but recovered later with bundle adjustment [140, 64, 15]. To better ar-

range the categorization of the common 3D surface reconstruction methods, the approaches mentioned were listed in Figure 1.1. This Ph.D proposal is focusing on the medical and industrial applications of passive 3D reconstruction with robot assistance.

1.2 Scanning fiber endoscope (SFE)

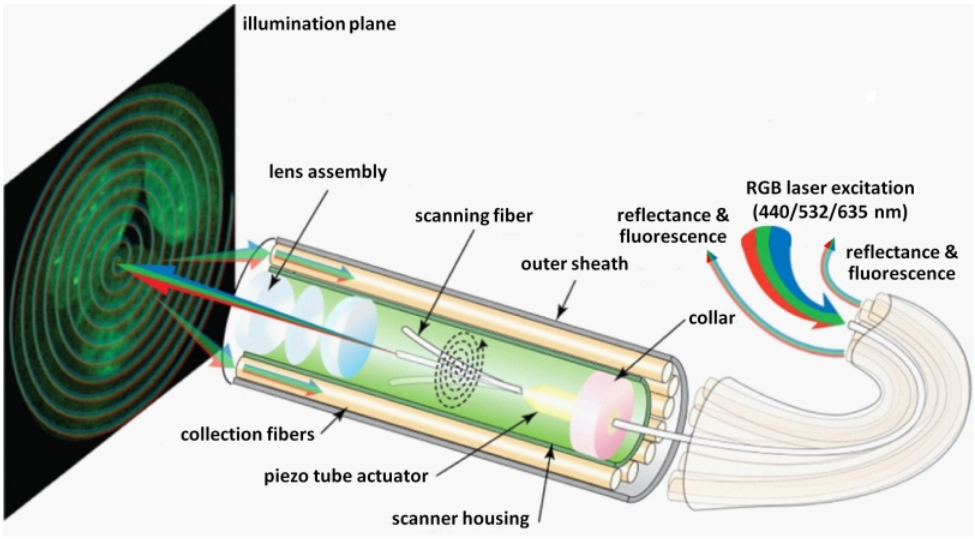


Figure 1.2: The schematic diagram of Scanning Fiber Endoscope (SFE). A white optical beam consisting of RGB laser light is being scanned by a single mode optical fiber in a spiral pattern and focused onto an object at low optical power by 2 or 3 lenses. The backscattered and fluorescence light is collected by a ring of multimode optical fibers that are wrapped around the 1-mm diameter fiber scanner assembly. Color video stream is generated with $>$ 500 line resolution at 30 Hz frame rate.

The SFE attains unconventionally high resolution for the flexible endoscope of only 1.2-mm in diameter (Fig. 1.2). Instead of using a single optical fiber or CCD sensor element for each pixel in the endoscope image, the SFE generates an image from scanning a single beam of red (R), green (G), and blue (B) laser light from a sub-millimeter fiber scanner. The fiber scanner consists of the cantilevered optical fiber driven near its mechanical resonance of lateral vibration (12 KHz) by a 0.4-mm tubular piezoelectric actuator. This spiral-scanned illumination fiber is single mode for deep-red to blue laser wavelengths. By collecting the laser reflectance backscattered from the tissue, full-color reflectance images

are formed; and collecting wavelength-shifted fluorescence emission, multiple channels of wide-field fluorescence images are formed.

To create video rate imaging, the collected light is sampled at 25 MHz on 4 channels concurrently to form > 500-line images at 30 frames per second [70]. By collecting light non-confocally, such as through a ring of multimode optical fibers (0.66 numerical aperture) surrounding the lensed fiber scanner, a 2-mm diameter scope is constructed with 5-10 nM fluorescence dye detection sensitivity to sodium fluorescein and Cy5.5 at a typical working distance of 2 cm [167]. Simultaneous reflectance and fluorescence imaging has been demonstrated using two techniques:

- technique 1: By replacing one reflectance channel, (e.g. green) for fluorescence, while providing color reflectance imaging from the other two reflectance channels (e.g. red and blue), fluorescence guided procedures can be conducted in vitro [165, 166, 122, 121].

- technique 2: A new multispectral fluorescence SFE was constructed with RGB lasers exciting three separate fluorescence channels, and one reflectance channel provides grayscale reflectance imaging for fluorescence guided biopsy [167].

Furthermore, this grayscale reflectance channel can be used to provide distance compensation to the fluorescence imaging channels, which allows a more quantitative fluorescence imaging feature of the SFE [165]. The multiple fluorescence channels also allows autofluorescence to be measured at every pixel and eliminated from the fluorescence image, providing a more quantitative fluorescence image, especially at the low dye concentrations that will be encountered at the margins of fluorescence-guided brain surgery [167].

1.3 Motivation and objective

1.3.1 3D reconstruction for automated bladder surveillance system with multi-segmented flexible robot

Bladder cancer accounts for estimated 72,000 new cases and 15,000 deaths annually and is the fourth most common cancer among men and the eighth most common caused of

cancer-related death in the United States [127]. The high rate of local recurrence of bladder cancer highlights the importance of long term follow-up surveillance [26, 69, 30]. Among different diagnostic methods, cystoscopy is considered as the gold standard for detecting the recurrence of bladder cancer [144]. It can achieve bladder cancer surveillance and lower urinary track pathology diagnosis by providing direct visual inspection on the bladder's inner surface. The most common follow-up strategy includes regular cystoscopy every 3 to 6 months in the first five years following bladder tumor removal and annual assessment thereafter [124].

The current cystoscopy has several disadvantages: 1) In the conventional clinical procedure, the cystoscope is inserted through the urethra and glided into the bladder. The patients may feel very uncomfortable or even painful, especially for man. A cystoscope with smaller outer diameter is preferred for the bladder surveillance; 2) The field of view (FOV) of current cystoscope is narrow so that the urologist can only see a small portion of the bladder inner surface. It is very hard for them to picture what the entire bladder looks like, which may cause an incomplete scan or missed regions being observed by the clinician; 3) the performance of the cystoscopic surveillance depends on the ability and experience of the urologist [159]. Thus, the performance of cystoscopy can be improved with rapid and complete inspection by reconstructing the 3D bladder virtual model with a tiny, flexible, automated biomedical robot.

1.3.2 Accurate 3D reconstruction of brain tumor cavity for image-guided surgical robotic system

Malignant brain tumors affect over 50,000 adults in US each year, and the rate has been increasing for the past decades. One of the most common primary brain tumors is categorized as glioblastomas [74]. The patients with these aggressive tumors typically have only at 2% chance to survive for five years [98]. Complete surgical resection of tumor tissue remains one of the most important factors for survival in patients with cancer [139]. Tumor resection in the brain is exceptionally difficult because leaving residual tumor leads to recurrence and decreased survival, and removing healthy brain tissue can impair patients memory, mobility

and other vital functions. Unfortunately, tumor margins are often indistinct under direct (optical) visualization. Although MRI and CT imaging can locate tumors, their use intra-operatively during surgery is very difficult and costly [115, 97]. Ultrasound imaging has low resolution, requires tissue contact, and lacks tissue-specific contrast [99, 25]. Optical imaging provides high-resolution non-contact imaging that is familiar with the surgeon, but penetration depth is limited. To enhance optical contrast during intraoperative tumor removal, fluorescence contrast is often used to label brain tumor margins at high resolution and contrast [74, 93, 167]. This technique is especially useful when brain tumors and their margins are exposed to the surgeon so the inherent shallow penetration depth (in white brain matter) of the visible (400 - 700 nm) and NIR light (700 - 1000 nm) [169] is not limiting.

Thus a surgical robotics system of high degree of dexterity, accurate navigation and highly precise resection is an ideal candidate for image-guided removal of fluorescently-labeled brain tumor cells. Moreover, an accurate multi-modal 3D reconstruction of the brain tumor surgical field is required to provide accurate 3D coordinates of the tumor cells to the surgical robot system for the proposed future application of robotically-assisted, fluorescence guided removal of residual tumor tissue.

1.3.3 Robot-assisted 3D optical metrology of tiny internal threads hole for automobile industry

The 3D metrology of internal surfaces is very important in manufacturing. Nowadays, the automotive manufactures incur high costs to ensure the quality of the internal threaded holes within the engine block. The quality of these machined holes are crucial to keep engine sealed and prevent catastrophic failure over the many years of operation, and very important for the specific company's record on safety, performance, and longevity of their products. To this day, the major automobile manufacturers evaluate the quality of threads in the engine block by using manual contact gauges [56]. The most common method for inspecting threaded parts by mechanical contact gauges is the Go/No-Go gauge, a ring or plug that checks for assembly. However, it comes with many disadvantages:

- 1) The evaluation of the internal threads is mainly depends on the operators' skills.
- 2) This contact measurement causes undesirable mechanical wear on the threads in the engine block.
- 3) The gauge requires maintenance and periodic certification calibration due the contact with threads.
- 4) The Go/No-Go characteristic of standard gauge cannot provide quantitative data.
- 5) Since there are more than 60 million cars produced each year, the inspect the engine block threads by manual work takes lots of labors, time and effort. In summary, the cost is very high.

Other 3D metrology methods contain Coordinate Measurement Machine (CMM), structure-light, x-ray and etc. But they are not feasible to achieve the 3D metrology of the internal surface, especially for small internal threads on an assembly line of large engine blocks. To better control the quality of the threads in the engine block, an accurate 3D optical dense reconstruction is desired for small internal complex surface, which can be achieved by utilizing ultra-thin camera and highly-accurate industry robot.

1.3.4 Solution to the biomedical and industrial applications

All the objectives mentioned above require several common things. First, a camera with ultrathin outer diameter and multi-mode (reflectance and fluorescence) is required for limited work condition of these applications, such as the space limitation of of cystoscopic detection of bladder small tumors, guidance of brain surgeries, and quantitative inspection of threaded holes. Secondly, the image acquisition system is driven by robot, such as the automated flexible cysctoscope robot, image-guided surgical robotics and industrial robots. Furthermore, the applications demand accurate 3D reconstruction, especially for the threads measurement.

To meet these requirements, the SFE is an ideal candidate for imaging due to its ultra-thin size and multi-modal function. Different robotic systems are applied to lead the endoscope for scanning and inspection. We collaborate with three robotics groups to accomplish these tasks. Three 3D reconstruction algorithms were developed : 1) bound constrained

Structure-from-Motion for biomedical applications, 2) axial-stereo internal surface measurement and 3) side-view multiview stereo for the threads metrology.

1.4 Dissertation overview

A general overview of each chapter is given below:

In Chapter 2, by collaborating with Smart Medical Devices Lab, Qatar University (Qatar), we propose a versatile, wire-driven, multi-segment automated robot for steering the SFE to follow optimal, pre-programmed scan trajectories. The multiple degrees-of-freedom (DOF) of the robot provide the advantage of maintaining the camera (SFE) perpendicular to the bladder wall with a safety distance. A 3D panoramic reconstruction view of the bladder is generated by custom software to improve comprehensive surveillance and clinical efficiency. This software can also handle multiple videos and generate physical-scale 3D model with reference target.

In Chapter 3, a synthetic phantom of debulked tumor from brain is fabricated having spots of fluorescence representing residual tumor. 3D surface map of this surgical field is produced by moving the SFE over the phantom by micro-positioning stage during concurrent reflectance and fluorescence imaging. The accurate poses information of the SFE are retrieved from micro-positioning stage to simplify the 3D reconstruction algorithm: bundle adjustment. By utilizing the pre-known neighboring frames, an improved algorithm with 3$\times$ efficiency succeeds to generated 3D multi-modal reconstruction of the synthetic phantom.

In Chapter 4, the miniature flexible endoscope SFE is affixed to the arm of RAVEN II surgical robot providing programmable motion with feedback control using stereo surveillance cameras. An accurate 3D multi-modal reconstruction of debulked brain tumor cavity with pre-known geometry is generated in an efficient way with 3$\times$ efficiency comparing with previous algorithm. Algorithm efficiency and accuracy is important to the robotically-assisted surgery since the 3D reconstruction is repeated intraoperatively during margin clean-up.

In Chapter 5, by matching the features from SFE video frames to the 3D virtual model that has been built, the camera pose (position and orientation) is recovered in near real-time by implementing a constrained bundle adjustment algorithm. Navigational error during the

approach to fluorescence target (residual tumor) is determined by comparing the calculated camera pose to the measured camera pose using a micro-positioning stage. Error analysis in this proof-of-concept study produced 3 mm distance error and 2.5 degree of orientation error on average. This work demonstrates feasibility of micro-camera 3D guidance of a robotic surgical tool.

In Chapter 6, a bound constrained bundle adjustment (BCBA) algorithm is proposed and developed to take advantage of the known but inaccurate information of the parameters of camera and object points. The theoretical development is provided with emphasis on the convergence property of the proposed BCBA algorithm. Moreover, the experimental result demonstrated its feasibility and reliability for a practical problem. Comparing with conventional BA algorithm that considers a parameter either as known fixed-value or an unconstrained variable, BCBA achieves less optimization iterations, less computation time, smaller ICP error and more reliable parameter estimation. Besides this, BCBA also achieved a good estimation of other parameters that have no constraints.

In Chapter 7, two images are captured by SFE as it moves along the optical axis by a short step within a small internal thread hole (7mm as minor diameter). The accurate position and orientation data of the two camera poses are provided by micro-positioning stage. By utilizing the proposed axial-stereo vision algorithm, a dense 3D model is generated with high-resolution, which demonstrated the feasibility of our proposed axial-stereo metrology approach.

In chapter 8, a side-view approach was proposed by using side-viewing camera to achieve 3D metrology of internal threads, especially for thread peak and valley. By using the similar experiment setup as the axial-stereo, a side-viewing camera is moving linearly from the opening of the hole to the bottom. A sequence of 2D side-view images are collected to generate a point cloud (called a "quadrant" representing the partial point cloud of the hole at a specific perspective) by multiview stereo algorithm. By rotating the threaded hole for complete scan of the entire internal surface, multiple quadrants are created. The conventional way for 3D point cloud registration ICP doesn't work for this case, since these quadrants are very similar in geometry. By using our proposed feature-based 3D panoramic registration, a complete 3D model of the internal threads is well generated. The comparison

between our model and X-ray data demonstrates the feasibility and potential of our 3D optical measurement method as an inline metrology approach.

In Chapter 9, a conclusion about this dissertation research is drawn. The future work in 3D metrology is proposed and discussed to solve the 3D metrology problem for the measurement of internal threads of engine block. By expend the idea of axial-stereo vision, multiview axial-stereo vision can be a new approach to achieve complete and dense 3D reconstruction. Considering these algorithms are sensitive to the accuracy of the industrial robot that steers the endoscope within the tiny thread hole, different image acquisition algorithm will be utilized to reduce the image distortion. For 3D biomedical imaging, several future work are proposed to make the current work more robust and efficient, such as dense 3D point cloud generation and the fusion of Kalman filter and bundle adjustment.

Chapter 2

3D PANORAMA FOR AUTOMATED BLADDER SURVEILLANCE SYSTEM

Cystoscope is an invaluable tool for bladder cancer surveillance and lower urinary track pathology diagnosis. The performance of cystoscopy can be improved by automating the procedure to ensure rapid and complete inspection, having a predicted pathway for imaging. This study proposes a versatile, wire-driven, multi-segment mechanism for steering an imaging probe to follow optimal, pre-programmed scan trajectories. A novel flexible imaging probe, SFE, of 1.2 mm in diameter generates high quality reflectance or fluorescent images. A 3D panoramic reconstruction view of the bladder is generated by custom software to confirm comprehensive surveillance and improve clinical efficiency. The multiple degrees-of-freedom (DOF) of the mechanism provide the advantage of maintaining the camera perpendicular to the bladder wall with a safety distance. Within a synthetic bladder phantom made from patient-specific geometry information, the apparatus is able to generate a 3D panorama of about 60% of the bladder's inner surface, while accepting 50% image overlap between adjacent images. The novel imaging capability of the SFE, combined with a robotized multi-segment steering mechanism and image stitching software proposes potential advancements to the multimodal 3D panorama of the bladder.

2.1 Background and objective

Bladder cancer accounts for estimated 72,000 new cases and 15,000 deaths annually and is the fourth most common cancer among men and the eighth most common caused of cancer-related death in the United States [127]. The high rate of local recurrence of bladder cancer highlights the importance of long term follow-up surveillance [26, 69, 30]. Among different diagnostic methods, cystoscopy is considered as the gold standard for detecting the recurrence of bladder cancer [144]. The most common follow-up strategy includes regular cystoscopy every 3 to 6 months in the first five years following bladder tumor removal and

annual assessment thereafter [124]. However, the performance of the cystoscopic surveillance depends on the ability and experience of the urologist [159].

The cystoscope is an invaluable tool in identifying pathology of the lower urinary tract by providing direct visual inspection on the bladder's inner surface. The outcomes of cystoscopy can be improved by automatizing the scan procedure to reduce the workload of the urologist and to provide consistent results. In the previous efforts to develop an automated cystoscope prototype, shape-memory alloy actuators were used for active distal end steering of bladder cancer surveillance system [173, 172]. The actuators were distributed along the longitudinal shaft of the imaging probe to generate unilateral bending motions. Combined with translational motions and rotational motions about the longitudinal shaft, the mechanism manipulated the imaging probe to scan the whole inner surface of a synthetic bladder phantom. However, heat dissipation from the actuator was a potential thermal hazard to the organ. Furthermore, because the scan motions by translation-bending-rotation pattern is less capable of steering the imaging camera to follow arbitrary trajectory, this limits the types of geometric surface on which good imaging results can be obtained [16].

2.2 Multi-segmented flexible robot

This study use a mult-segmented flexible robot that is developed by *Smart Medical Device Lab, Qatar University, Qatar*. This robot contains a wire-driven steering mechanism with multiple segments connected by hinge joints for versatile bending motions in automated cystoscopy, see Fig. 2.1 and Fig. 2.2 The use of hinge joints can avoid the afore mentioned problem of ball/spherical joints so that the steering motions can be controlled accurately by wires. The feature of multiple degrees-of-freedom of the structure is necessary for generating perpendicular views at the optimal imaging distance on the surface of an irregular-shaped bladder.

In order to unburden clinicians with a faster and easier diagnostic method accompanying the proposed automated cystoscope, 3D panoramic views of the *in vitro* phantom bladder's inner surface are built from cystoscopic videos. Customized image stitching software has been developed to achieve this purpose. Several image stitching techniques have been introduced in numerous medical fields, including dental model [76], esophagus imaging [118],

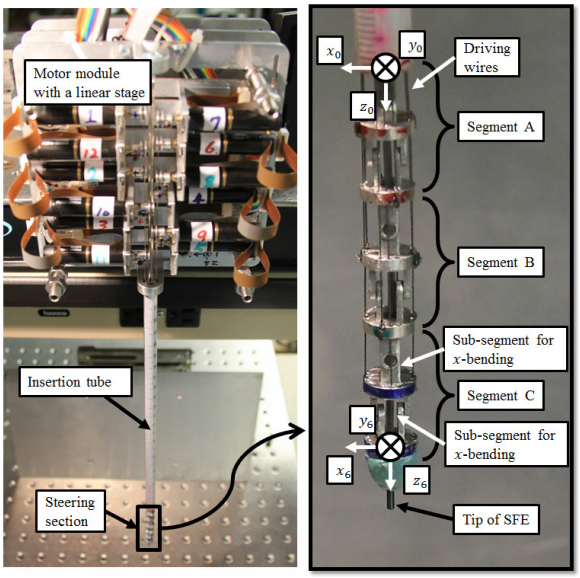


Figure 2.1: Prototype of automated cystoscope (vertically installed). Left: The prototype consists of a motor module, an insertion tube, and a flexible steerable section. Right: The structure details of the multi-segment steering section.

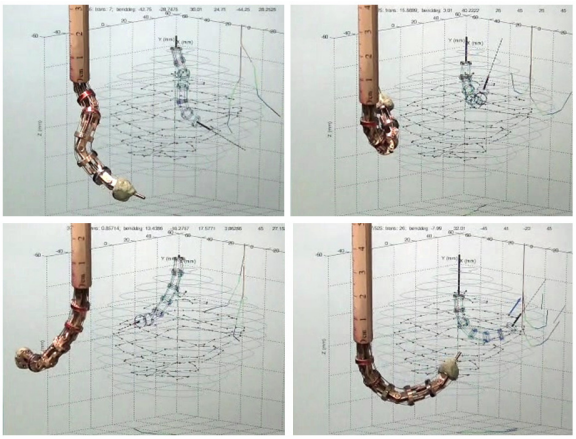


Figure 2.2: Steering motions of prototype of automated cystoscope. Captured motions during steering in air. The steering section is controlled according to the kinematic simulation shown in the background.

catadioptric endometrial imaging [125], bladder endoscope images [10], etc. Soper *et al.* developed algorithms for building 2D mosaic image and 3D model of a urothelium surface from a surveillance video stream [130]. The biggest advantage of the algorithms is that the reconstruction were performed without supplemental foreknowledge of the particular endoscope used.

In this study, a much improved 3D reconstruction software was developed to generate a physical-sized 3D bladder's inner surface, instead of the unified dimensional model in [130]. This feature is useful for locating the physical coordinates of the bladder lesion in the surveillance procedure and to provide quantitative characteristics such as size and morphology.

2.3 Experiment and 3D panorama result

A computed tomography (CT) cystogram of a female patient's bladder was used to construct a three dimensional plaster model. The same point cloud information was also used for planning scan trajectories and kinematics simulations. In building a synthetic bladder phantom, multiple thin layers of industrial liquid latex (RL-451-80, Silpak Inc., Pomona, CA, USA) were brushed on the plaster model, as shown in Fig. 2.3(a). When the latex layers were dried out, red and blue vessel features were drawn on the inner surface of the bladder phantom. Fig. 2.3(b) shows the bladder phantom inside-out, revealing the vessel features. Four star-shaped fluorescent targets of 6 mm in outer diameter were made with Fluoro-dye-in polymer (Fig. 2.3(c)) and were located over the vessel features to simulate the biomarkers for guiding biopsy and therapy [166].

2.3.1 Setup

Fig. 2.4 shows the setup for the full scan experiments. The cystoscope prototype using the proposed steering mechanism was installed vertically with the insertion tube pointing downward. The bladder phantom was supported with its internal urethral orifice facing upward so that the cystoscope can be inserted into the bladder with translational motions. A rigid PDMS concave holder sharing the same shape as the bladder phantom was used to support the lower half of the bladder phantom so that the natural shape of the phantom

(a) (b) (c)

Figure 2.3: The fabrication of bladder phantom with artificial vessels features and fluorescent markers. (a) Multiple thin layers of latex were brushed on a plaster bladder model to form the bladder phantom. (b) Vessel features drawn with red and blue ink. (c) One of the star-shaped fluorescent biomarkers.

can be preserved. The relative position and orientation between the cystoscope and the bladder phantom were set according to simulation configurations.

A SFE passes through the central cavity of the steering mechanism, with its tip sticking out (about 10 mm) and its optical axis aligned with the longitudinal axis of the mechanism. The SFE system is a flexible miniature laser imaging probe with a outer diameter of ϕ1.2 mm [120, 70]. It has a forward-view with field-of-view larger than 50° and a depth-of-focus of 2∼50 mm. The SFE video provides circular image frames containing an effective image area of 600 pixels in diameter. The SFE is multimodal. Its video output can be full-color images when using a scanning white optical beam with red, green, and blue laser light sources, as shown in Fig. 2.5(a). By deactivating the green laser source, the standard green channel only records the fluorescence emission signals. The red and blue channels were still used for reflectance imaging [166], as shown in Fig. 2.5(b).

2.3.2 Trajectory and SFE Video

One of the fundamental requirements for the mechanism in steering its tip to follow a scan trajectory is keeping a safe distance from the bladder's inner surface. In addition, it is

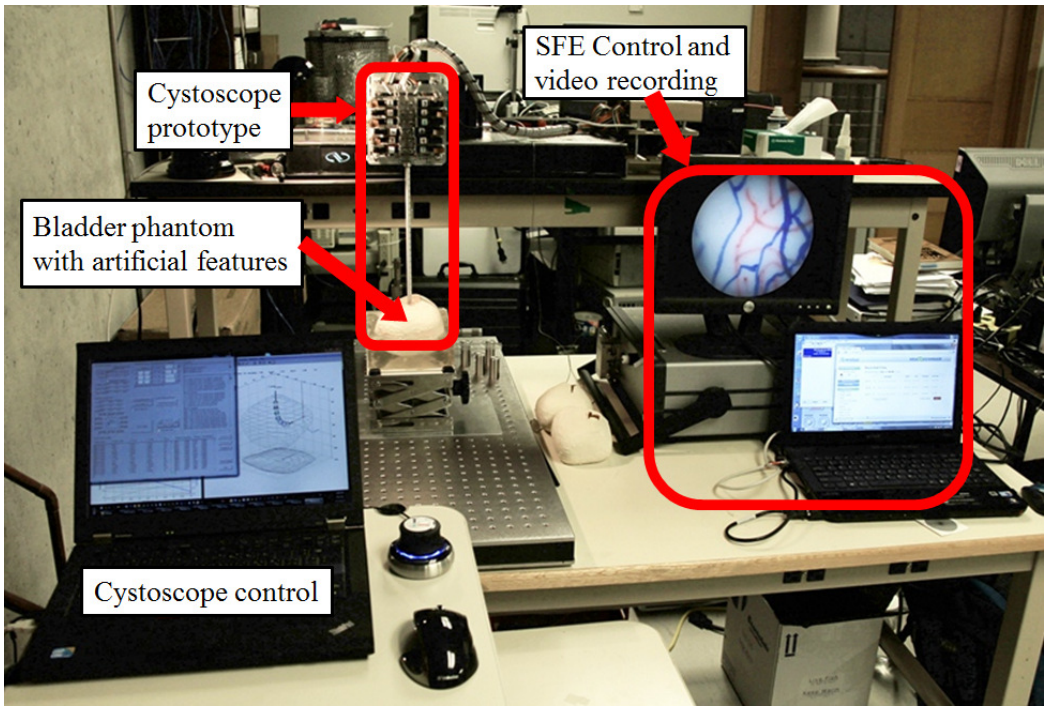
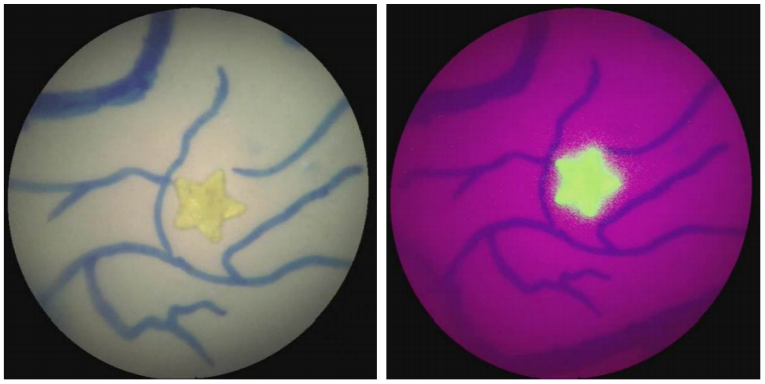


Figure 2.4: Experiment setup for automated full bladder scan.



(a) Reflectance image. (b) Flourescent image.

Figure 2.5: Different modes of SFE video outputs. (a) Reflectance image (b) Fluorescent image.

preferred to have perpendicular views on the bladder surface for less image distortion and easier visual diagnosis because the cystoscopic video is also used for direct visual inspections. Therefore, surface normals at the imaging locations are used as the target vectors in inverse kinematics.

Due to insufficient detailed features on the bladder surface, large image overlaps were required to assure correct image stitching [130]. Therefore, the density of imaging locations on the bladder surface needs to be sufficiently high in the direction in which the scan motion occurs and in the corresponding lateral direction. In this study, all imaging locations are located in one spiral trajectory. There are several types of trajectory, i.e, longitudinal, latitudinal and spiral [171]. For the same number of imaging locations, the spiral trajectories provide the shortest trajectory lengths. Fig. 2.6 shows the target trajectory for scanning the whole bladder's inner surface with a safety distance of 15 mm. The pitch of the spiral trajectory is determined to ensure 50% overlap between the images in the lateral direction of the trajectory. There are 529 imaging points on the trajectory. When the safety distance is set at 30 mm, larger area can be covered in the image frame. Therefore larger pitch can be used and the number of imaging points is reduced to 199.

With the specified target spiral trajectory, steering configurations for all imaging locations are calculated by the searching-based inverse kinematics solving method. Due to the limited workspace of the mechanism, not every imaging location can be matched to a configuration with zero position/orientation error. For example, the position (in mm) and orientation (as a unit vector) of the camera for the first imaging location are $[0.0183, 0.0131, 70.0000]^T$ and $[0.0005, 0.0005, 1.0000]^T$, respectively. They are well matched by the results of inverse kinematics, $[0.0236, 0.0183, 70.0000]^T$ and $[0.0005, 0.0005, 1.0000]^T$, respectively. However, because the required steering configurations are out of the workspace of the system, a configuration with minimum weighted error is used instead. When the ideal steering configuration is out of the mechanism's workspace, the resultant imaging locations shift from the target ones, resulting in a feasible scan trajectory that is different from the target one, as shown in Fig. 2.6.

The scan experiment was performed on the empty bladder phantom. The scan lasted for about 10 minutes with the safety distances being set at 15 mm. The SFE video outputs

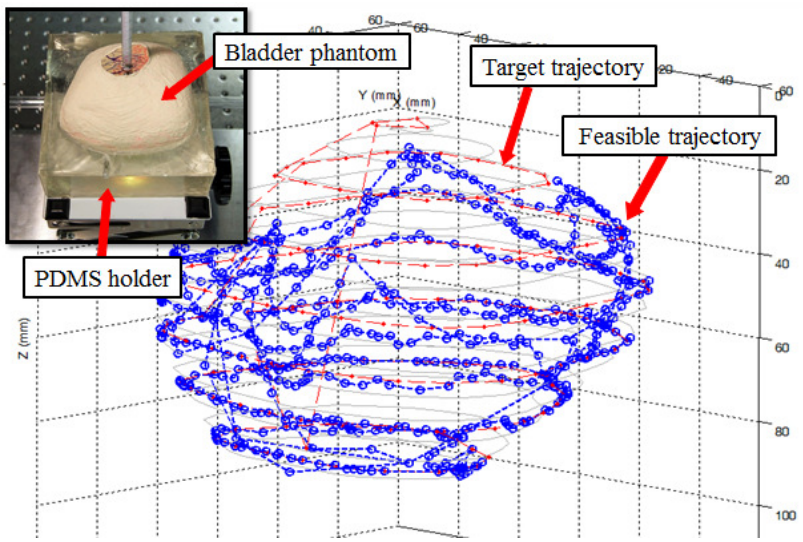


Figure 2.6: Generation of a scan trajectory. The target trajectory (dot markers) is generated to cover the whole bladder's inner surface, with a safety distance of 15 mm and 50% of image overlap. The feasible trajectory (circle markers) is actual scan trajectory based on inverse kinematics solutions.

were recorded as AVI files at 30 fps by a VGA recorder (Epiphan System Inc., Ottawa, Canada), generating 19,753 SFE video frames.

2.3.3 Image Stitching & 3D Reconstruction

Image stitching is widely used in medical imaging field to extend the field-of-view of the camera with limited view angle [104, 80]. Image registration is an important step in 2D image stitching, in which scale-invariant feature transform (SIFT) algorithm was applied to find feature points from each SFE image frame. Random Sample Consensus (RANSAC) was used to refine the homography matrix between images and also determine matching frames which share image overlaps [163]. The homography matrix is a 3×3 matrix describing the mapping between two images. It can be determined by four pairs of corresponding points. All the images sharing overlaps can be registered and stitched together to build a large 2D panoramic image. The error of the homography estimation mainly came from the RANSAC algorithm. In each iteration in RANSAC, four pairs of corresponding features were randomly

chosen to solve the homography matrix, while other corresponding features were classified as inliers or outliers based on the homography values. The homography of the iteration that contains the largest number of inliers was selected. So the error of homography estimation depends on how many iterations running in the RANSAC algorithm. In our study, the iteration times was chosen high enough to ensure that the probability of accurate estimation was at least 0.99 [181].

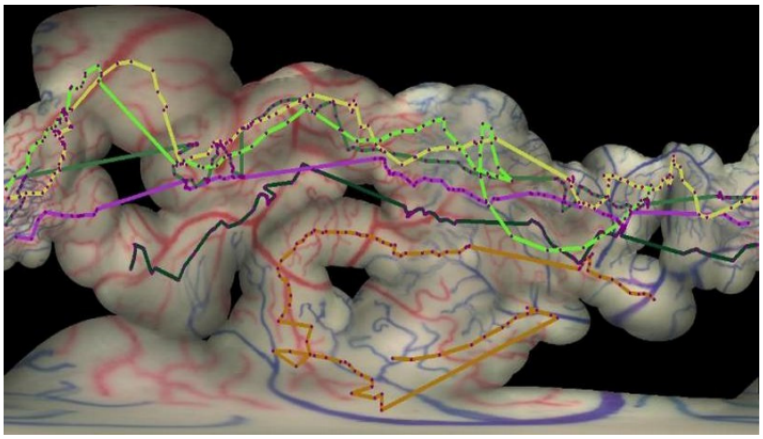
Based on the surface stitching algorithm developed in [130] that can stitch image frames into a 2D mosaic and create a unified 3D model, major improvements in the algorithm had been made as follows.

1. Different sets of parameters were applied for different input images. In this study, two set of parameters in SIFT feature detection algorithm were used to process the videos obtained at 15 mm and 30 mm safety distance, respectively.

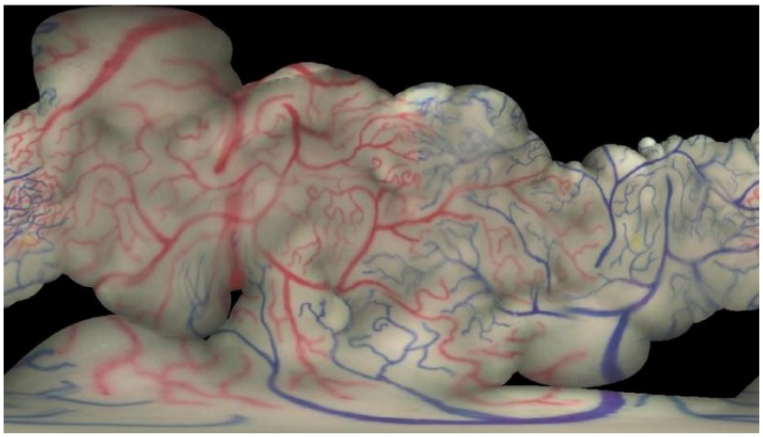
2. In the previous algorithm, the matching frames were found by searching the neighbor frames in the scanning direction as well as in the lateral direction. In this study, the deviated trajectories reduce image overlaps, especially between images in the lateral direction. Therefore, the bladder inner surface was reconstructed as numerous image portions instead of a complete one. By combining the two sets of videos, there are more images to provide the missing matching information of the gaps between these image portions and a complete 3D reconstruction can be obtained.

3. Extra feature was added into the algorithm that the position of each frame was tracked in the 2D panoramic image generation procedure. The centers of adjacent frames were lined up to show the trajectory of SFE for better visualization.

4. The improved reconstruction algorithm can generate a model of physical size, instead of relative size as in previous algorithm. This improvement provides the physical location of lesions in 3D space and also enable the quantification analysis between the 3D reconstruction and actual phantom in the future to develop a more reliable reconstruction algorithm.



(a)



(b)

Figure 2.7: The reconstruction of 2D panoramic view. (a) The result of 2D image stitching. The scan trajectories sections are shown as colored solid lines with the image centers marked as dots in the lines. (b) The improved 2D panoramic stitching result with two scans.

The stitched 2D image from all SFE video frames obtained at 15 mm safety distance is shown in Fig. 2.7(a). The scan trajectory was shown as dotted curves of colors that were different from the red and blue vessel features. Every curve of the same color represents a circular section of the spiral scan trajectories. Six trajectory sections can be identified in the figure.

The blank areas at the upper part of the 2D image indicate the uncovered regions near the internal urethral orifice due to the limited range of the bending angle ($\pm 45^\circ$ for the proposed mechanism). Several small blank areas exist within the stitched area due to errors in control and in positioning the bladder phantom.

The areas near urethral orifice can be imaged with a greater extent of bending of the steering section at increased bending angles. The missed portions of the stitched area can be eliminated by (1) higher precision in positioning the bladder phantom with respect to the mechanism as in simulation, and (2) trajectories planning with increased image overlap. Combining the cystoscopic videos of multiple scans improves the robustness of image mosaicing and 3D reconstruction. For example, in this case, the missing image information can be provided by another scan with the same image overlap percentage but an increased safety distance of 30 mm (i.e., further from the bladder surface). By re-running the software over the combined two videos with two sets of parameters, a more complete 2D panoramic image was generated with 10% increase in coverage, as shown Fig. 2.7(b).

The improved algorithm reconstructed a 3D point cloud of all the feature points captured in the scan videos. Spline fitting was then applied to generate a mesh of the whole surface [130]. More precise 3D surface can be generated from videos covering more surface areas. To quantify the reconstruction result, Iterative Closest Points (ICP) was employed to align the reconstructed bladder model to the CT-scanned model [11, 19]. Fig. 2.8(a) shows the comparison between the CT-scanned model and the reconstructed model. The average distance between the models is 5.5 mm (about 5% of the bladder outer diameter).

The physical-scale 3D panorama of the bladder's inner surface was built by warping the 2D image on the reconstructed model of the bladder using the florescent stars as precise size references, as shown in Fig. 2.8(b). The sizes of the features on the image will be precise if the geometry model of the bladder is accurate, which can be achieved using computed

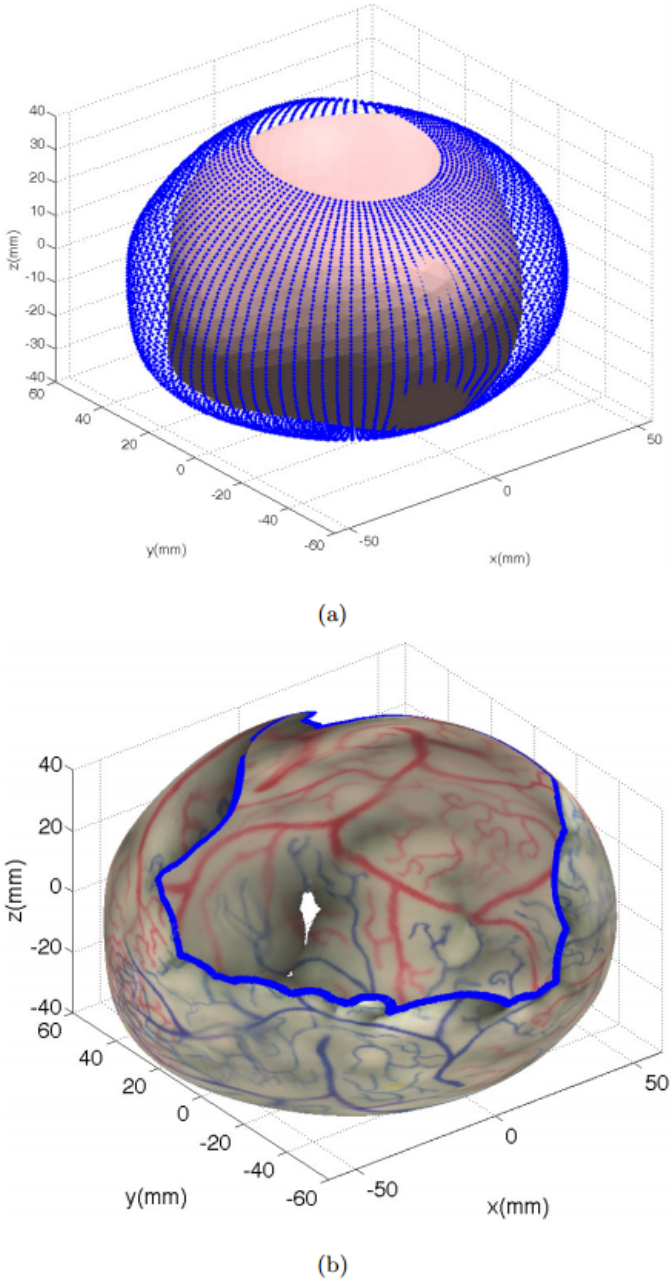


Figure 2.8: The reconstruction of a physical-scale 3D bladder phantom.(a) The comparison between the CT-scanned model (surface) and the model reconstructed by the improved algorithm (point cloud). (b) The 3D panorama of bladder's inner surface built by warping the 2D image on the reconstructed model of the bladder.

tomography technology.

2.4 Discussions

Cystoscope is an invaluable tool in common urology procedures. The automation of the cystoscopy may eliminate the need for direct clinician oversight and reduce the duration of the procedures. The results of cystoscopies in the form of digital 3D panoramic images can be used as the medical record documentation of the patient. Automated cystoscope has the potential for telemedicine based on the flexibility on the time and location to perform the examinations. Future studies can focus on the economic evaluation model comparing the costs and consequences of using conventional bladder cancer surveillance versus the proposed automated cystoscopic system.

With the limitations of the current proposed design, there are several issues need to be discussed and highlighted for future research.

2.4.1 Bending angle limitation

The main challenge in designing the steering mechanism is to reduce the overall outer diameter. Compromises had been made in choosing the outer diameter with consideration of the available machining technology, the size of the bladder, and the diameter of the urethra. To achieve perpendicular views on the area of the bladder's inner surface near the internal urethral orifice, it requires the proposed steering mechanism to be capable of backward-bending, which requires bending angles of at least $\pm 60^\circ$ per sub-segment. However, for the proposed mechanism with a wire-driven hinged chain, the safety maximum bending angle is about $\pm 45^\circ$ in bidirectional rotations. In case the bending angle is larger than 45°, it is likely that wire tensions cannot generate any restoring moment about the hinge joint axis.

It is possible to achieve bidirectional fully backward-bending, i.e. 180°, in both x- and y- directions by simply adding one more sub-segment for the bending in each direction. However, without increasing the difficulties in mechanical manufacturing, the outer diameter of the steering section needs to be increased by 1 mm to accommodate the additional 4 guiding holes for the wires that control the newly added sub-segments. The length of each sub-segment also increases with the outer diameter to 10.5 mm, resulting in a total length

of the steering section of 84 mm. It requires extra efforts to manipulate such a long steering section inside the bladder while satisfying the requirement for a safety distance between the tip and the bladder inner surface. For example, in the case of using the 180° bending feature, the physician may need to steer the sub-segments that have been just inserted inside the bladder so that the steering section takes the shape of a 3D curve rather than a straight tube during insertion.

If there are less application restrictions, such as in the case of laparoscopy, it is very promising to use the steering mechanism with more steering units. Given the safety requirement and the spatial restriction from the urethra and the bladder, a better solution could be to make the steering section as a continuum robot that is made of superelastic material, where the hinge joints are replaced by flexure hinges to provide larger capacity of bending motions.

2.4.2 Combining multiple scans

With the intrinsic ±45° bending angle limitation of the proposed mechanism (thus 125° bending backward), about 60% of the bladder's inner surface can be covered and reconstructed, as shown in Fig. 2.7. For those imaging points that are close to the boundary of the configuration space, the safety distance requirement is ensured to be met at the cost of increased orientation errors, which introduces reduced acute view angles and increased distances between the actual imaging locations and the target ones. With actual view angles, the results are images with degenerated quality. In some cases, the obtained images are not suitable for image matching and stitching due to 1) uneven illumination, i.e., those bladder surface closer to the scope looks extremely bright in the image while those areas far away is appeared to be very dark and noisy; and 2) image matching issue, i.e., the low-quality images provide less or unreliable SIFT features.

The feature-based image stitching algorithm does not require every scan to follow the same trajectory. Therefore, the information of missing areas in one scan can be supplemented from other follow-up scans. By combining the video frames of multiple scans and re-computing over the combined image sets, a more complete 3D bladder model can be

reconstructed.

2.4.3 Exploring the 3D Panorama

The reconstructed 3D panorama views can be saved for later examinations. Currently the 3D panorama is explored in a simple MATLAB graphic user interface in a similar way of exploring a virtual globe. The panorama view is virtually grasped and manipulated via the knob of a 3D mouse. A dedicated program can be developed for the ease of exploration. The program can be equipped with additional functions, such as retrieving the original cystoscopic video clips that cover the selected areas of interest.

2.5 Conclusions

Cystoscopy is the gold standard procedure in identifying pathology in the lower urinary tract and is commonly used in postoperative bladder cancer surveillance. This study proposed the design and development of a wire-driven multi-segment steering mechanism along with customized 3D panorama software for automating the cystoscopic procedure in order to achieve better image quality and novel user interface. The mechanism provided multiple degrees-of-freedom to provide perpendicular views on most areas of the irregular-shaped surface of the phantom bladder while satisfying the safety distance criteria. Spiral scan trajectories were used for fast scan the whole bladder inner surface. The robust image stitching algorithm could build the 3D panoramic view of the scanned areas based on the recorded video(s). Multiple videos of different scans on the same bladder were able to provide improved stitching results.

Chapter 3

MAPPING SURGICAL FIELDS BY USING A MULTIMODAL SCOPE OF KNOWN POSES

Endoscopic visualization in brain tumor removal is challenging because tumor tissue is often visually indistinguishable from healthy tissue. Fluorescence imaging can improve tumor delineation, though this impairs reflectance-based visualization of gross anatomical features. In this study, the SFE is used to accurately navigate and resect tumors. By affixing the SFE to a robotic arm providing programmable motion, a multimodal surface map of the surgical field can be generated by 3D computer vision algorithm.

To test this system, synthetic phantoms of debulked tumor from brain are fabricated having spots of fluorescence representing residual tumor. 3D surface maps of this surgical field are produced by moving the SFE over the phantom during concurrent reflectance and fluorescence imaging (30Hz video). SIFT-based feature matching between reflectance images is implemented to select a subset of key frames, which are reconstructed in 3D by bundle adjustment. The resultant reconstruction yields a multimodal 3D map of the tumor region that can improve visualization and robotic path planning.

Efficiency of creating these maps is important as they are generated multiple times during tumor margin clean-up. By using pre-programmed vector motions of the robot arm holding the SFE, the computer vision algorithms are optimized for efficiency by reducing search times. Preliminary results indicate that the time for creating these 3D multimodal maps of the surgical field can be reduced to one third by using known trajectories of the surgical robot moving the image-guided tool.

3.1 Background and object

Complete surgical resection of tumor tissue remains one of the most important factors for survival in patients with cancer. Tumor resection in the brain is exceptionally difficult because leaving residual tumor tissue leads to decreased survival and removing normal

healthy brain tissue leads to life-long neurological deficits. Unfortunately, tumor margins are often indistinct under direct (optical) visualization. To aid intraoperative tumor removal, the brain tumor can be labeled with fluorescence contrast [131, 145]. Ideally, the surgeon is provided with both reflectance (color) and fluorescence wide-field images. These qualities are provided from a flexible miniature multimodal laser-scanning scope, SFE [70].

The SFE we used in this study is only 1.2mm in diameter with a 9mm rigid tip length, while forward-view images at moderately wide field of view at 70 degrees, high resolution (608 x 608 which is unprecedented in 1-mm scopes), and depth of focus of 4-40 mm that matches conventional flexible cystoscopes. By deactivating the green laser source, the standard green channel only record the fluorescence emission signals. The red and blue channels were still used for simultaneous reflectance imaging. Recently, multiple fluorescence targets can be imaged quantitatively using a phantom of a human esophagus with known concentrations of fluorescence dye using the multimodal SFE [165, 166].

In this study, the miniature imaging system SFE was attached directly to one robotic arm of RAVEN [79]. RAVEN is a surgical robot with two 7 degree-of-freedom (DOF) manipulators, which provide a very high degree of dexterity, precise cutting, and accurate navigation with surgeon-assisted visual feedback. It is an ideal candidate for robotically assisted surgery. In this study, programmable motion of the robot arm was performed, and simultaneous positions and orientations were provided through RAVEN system.

The goal of this collaborative research is to develop an image-guided surgical robotics system by utilizing 3D surface mosaic software [130] with known SFE positions and orientations provided by RAVEN. This improved software would generate a physical scale 3D model of residual tumor that is fluorescently labeled after the majority of the tumor has been removed (debulked), and would also accelerate the speed of 3D model reconstruction. This study explores the robotic use of the SFE for 3D intraoperatively forming panoramic maps of the surgical field that will be used by the surgeon to plan the final stages of residual tumor removal.

3.2 3D multimodal reconstruction algorithm

In our previous work, a surface mosaic software was utilized to produce 3D panoramic surface and 2D map of bladder based on a bladder surveillance video by using computer vision techniques [130]. The reconstruction software is written in MATLAB. The procedure utilized scale-invariant feature transform (SIFT) to find out feature points, and random sample consensus (RANSAC) to figure out best matching features, and bundle adjustment to create a 3D panoramic surface by simultaneously refining the positions and orientations of camera (SFE) and the 3D coordinates of features. Finally, a thin plate spline algorithm was applied to generate a smooth 3D surface by fitting these feature points. This reconstruction algorithm solely based on endoscope images can adequately build a 3D model, but it is time-consuming.

Our hypothesis is that knowing camera position and orientation the 3D reconstruction can be generated in less time and with more accurate results compared to our previous approach [130] based solely based on features within the hollow organ. By attaching SFE directly to micro-position stage, the positions and orientation of SFE can be retrieved. To test the concept, we set the distances of each neighbor camera positions as equal and the orientation as constant during the experiment. By applying these information of camera into the previous 3D surface mosaic software, the reconstruction speed can be accelerated and improved. Firstly, by knowing each position of cameras, the frame pairs with overlap (matching frame) can be easily discovered. Thus, it costs less time than the conventional method by searching through every frame for finding the matching frames. Secondly, in the previous algorithm, bundle adjustment is a huge optimization problem of unknown parameters: pose of cameras and the 3D coordinates of features. Since the motion of camera is known in this study, bundle adjustment would be simplified as a smaller optimization of only 3D coordinates of features. The a priori knowledge of matching frame pairs and simplified optimization problem would reduce the computation time. Moreover, the known camera information may result in a physical scale 3D model, instead of a relative shape with the previous software approach [130].

To maximize the advantage of known camera position and orientation, a new feature

was added to the previous software [130] to make the reconstruction more efficient and more robust. By knowing the real movement of camera, the positions of feature points in the neighbor frames can be predicted. Thus, for a specific feature point, we just need to search a small region around its prediction position in the neighbor frame to find its matching feature. This approach reduces the computation time in the step of finding the best matching features, since we do not need to search throughout the whole image. And this constrained searching region can reduce the interference of similar features that locate in some other regions of image, which can discover more matching features pairs and at the same time enhances their robustness and accuracy.

The other advantage of this reconstruction algorithm based on feature prediction is to improve the efficiency of bundle adjustment. As mentioned previously, the bundle adjustment would be simplified with known cameras poses. Furthermore, the more robust matching features provide a more accurate initial estimate of the reconstructed shape, which can decrease the number of iterations performed during bundle adjustment. Both of the advantages of feature prediction reduce the computation time dramatically and make the algorithm more efficient.

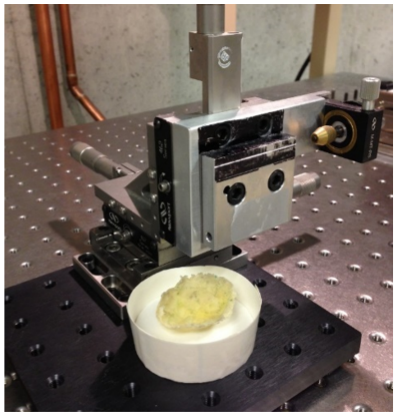
3.3 3D mosaicking on micro-positioning stage

Prior to experimentation with the RAVEN, the 3D surface mosaic software was tested with accurate known camera position and orientation providing by a micro-positioning stage. This experiment demonstrates the feasibility and reliability of the new reconstruction algorithm. All the codes are Matlab implemented, and running on a workstation Dell Precision M4700 with 2.7 GHz Intel i7-3740QM CPUs, 20.0 GB memory in a 64-bit Window operating system.

To mimic the debulked tumor in brain, a cup-shaped latex phantom of 3cm x 4cm was made, consisting of painted synthetic tissue and multiple small spots of fluorescence that represent residual tumor at the margins, see Fig. 3.1(a). By illumination with a 444 nm laser (blue), these fluorescence spots emit green light.

The phantom is imaged across the open cavity at known camera positions and orientations that represent the programmed route that the robotic arm will traverse over the

(a)



(b)

Figure 3.1: The experiment on micro-positioning stage. (a) A cup-shaped latex phantom to represent brain tissue after tumor debulking. Multiple fluorescence spots were randomly scattered on the surface to represent residual tumor. (b) The experiment setup on optics table. The micro-positioning stage would hold SFE to scan the tumor phantom with known camera poses and motions.

surgical field. These locations and orientations are chosen to provide at least 50% overlap of the visual field so that all images can be stitched together to form a panoramic map. Since the fluorescence signal *in vivo* is expected to be weak from the residual tumor tissue, the SFE is expected to remain stationary at these preprogrammed locations of the robotic arm to increase signal to noise. In the initial experiment, a micrometer micro-positioning stage was used to provide accurate positions and orientations of cameras, Fig. 3.1(b). A miniature SFE was attached to the micro-positioning stage with fixed orientation pointing perpendicularly downwards. 30 camera positions were chosen, which evenly distributed in a 5×6 grid. The distance among each neighbor camera positions is 5 mm.

The standard SFE RGB reflectance imaging was performed and recorded. White light combining with red (635 nm), green (532 nm) and blue (444 nm), was projected onto the surgical field, and the return RGB signals are then simultaneously captured and amplified by three photomultiplier tubes (PMTs), Fig. 3.2(a). Since the fluorescence spots have emission spectrum of green light under blue laser excitation, the standard green channel only records the fluorescence emission signals after the green laser source was deactivated. In this study,

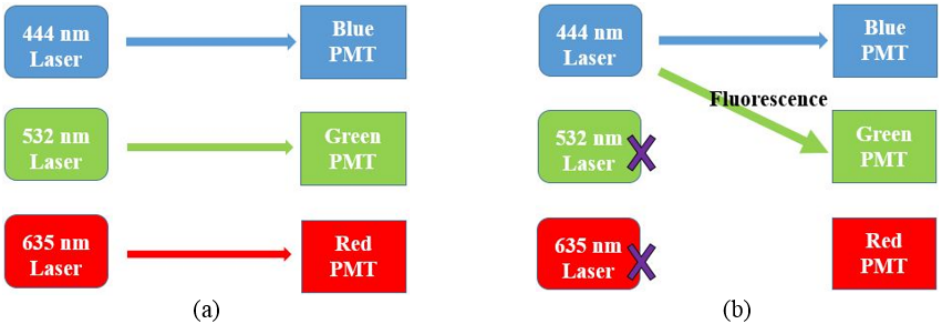


Figure 3.2: Reflectance and multimodal (reflectance + fluorescence) imaging diagrams for the SFE using a green fluorescence biomarker. (a) Standard SFE RGB imaging. Each RGB laser source is filtered and amplified using a color-specific PMT channel; (b) SFE fluorescence imaging, with green and red laser sources inactive. Reflectance images are formed using blue laser illumination only, while fluorescence is detected within the green color channel.

we turned off the red and green laser source, the captured fluorescence image on the green channel and reflectance on the blue channel, Fig. 3.2(b).

There are many more features in the reflectance images versus the fluorescence images, Fig. 3.3(a, b). The fluorescence emission signal could be all black when all tumor tissue is removed, so reflectance images are relied upon for reconstructing the surgical field. Multimodal 3D image can be generated by merging reflectance with a green fluorescence overlay, Fig. 3.3(c). By first removing field distortions and calibrating the SFE based on the measured field of view (70-degree cone angle), accurate mapping of the phantoms dimensions are possible.

In the 3D reconstruction software, numerous features of each frame were detected and characterized by SIFT. This is implemented by using the SIFT algorithm as part of the VLFeat toolbox (www.vlfeat.org). In any pair of frames I and II, Fig. 3.4, the feature-space distances were calculated for each feature in frame I to every single feature in frame II, and found out the features in frame II with the nearest and the second nearest distances. To keep good (real) matches, a further criterion are preserved that the ratio of nearest and second nearest distances should be less than a threshold. We set 0.8 as the threshold in this study.

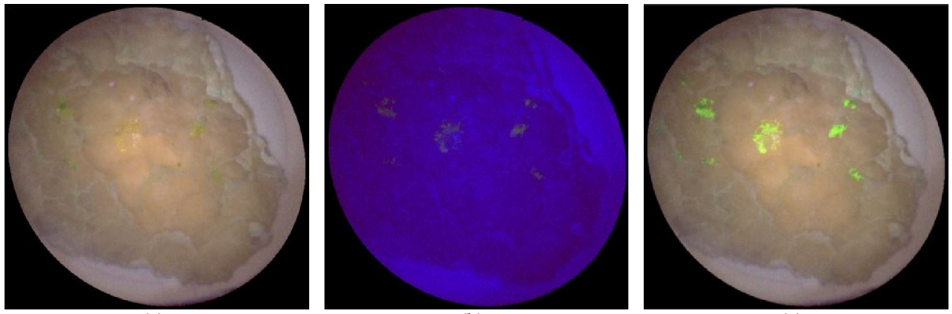
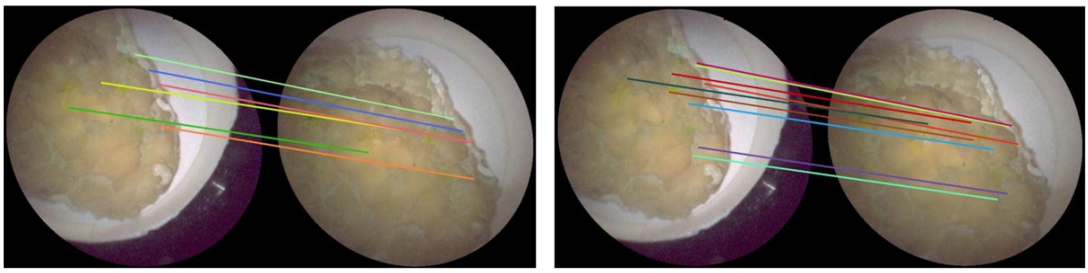


Figure 3.3: The reflectance and fluorescence SFE image of tumor phantom. (a) The standard reflectance image of SFE; (b) the fluorescence imaging by deactivating red and green laser source. The phantom segments with fluorol dye, which represent tumor residual, emit fluorescence signals (green) under blue laser activation; (d) Multimodal 3D image that is generated by merging color reflectance + fluorescence.

In the previous algorithm solely based on feature, we had to search throughout every features in frame II to find matches. By providing accurate position and orientation of the cameras, a reasonable estimate of the homography between image pairs can be made. Thus, the positions of matching features in frame II can be roughly calculated. A search window was set around the prediction position so that the real matching feature can be detected in a more robust and more efficient way comparing with the previous algorithm.

In addition to the improvement of computation efficiency, our new algorithm can find more matches than the previous one. This was advantageous to the robustness of the reconstruction, especially for insufficient image overlap, less than 50%. A single image may contains several similar features. In another word, they do not have much difference in feature space. Thus the radio of the nearest and second nearest distance may not satisfy the radio criterion. By search a small region around the features prediction position, we can reduce the interference from similar features that locate in other parts of the image. In this way, our new algorithm can keep more matches than the previous work, Fig. 3.4.

Given the known positions of the endoscope for each frame, the relatively large problem of simultaneously resolving both scene and camera parameters, known generally as structure from motion, simplifies greatly. Also narrowing down the search region provides larger



(a) (b)

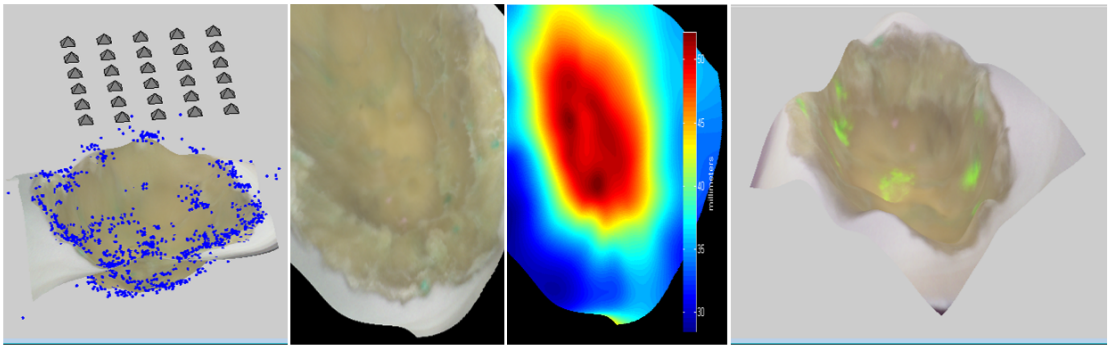
Figure 3.4: Matching features in a pair of frames I and II were detected with/without feature position prediction and lined up with colorful lines. (a) Less matching pairs were found by the previous algorithm. (b) More matches were detected with the prediction of feature points.

number of matches. This improvement provides a more accurate initial estimate of the reconstructed shape, reducing the number of iterations performed during bundle adjustment.

The 3D panoramic mapping procedure of the debulked fluorescence tumor from brain tissue is illustrated in Fig. 3.5(a)-(c). The first in the series represents the reconstructed 3D map of the phantom with 30 positions 5x6 of the SFE camera above, and the features as blue points used for mosaicking the video frames together into the panorama. The second pair of images Fig. 3.5(b), shows a portion of the 3D panoramic image in reflectance color contrast only, and the matching pseudo-color depth map at lower lateral spatial resolution, but absolute depth scale. The third image is 3D surface reconstruction of the entire phantom with multimodal color contrast, having bright green representing the fluorescence residual tumor. This 3D reconstruction process takes approximately two and half minutes for processing all 30 SFE multimodal images. This computational time can be reduced to around 50 seconds in our proposed method by reducing the searching range during the SIFT and decreasing bundle adjustment iteration.

During the reconstruction process the matching fluorescence image is carried with the reflectance image to generate a multimodal 3D map of the surgical field, Fig. 3.5(c). However, the expected time required for 3D reconstruction is considered long for regularly mapping of the surgical field in an iterative fashion. The scenario envisioned is that a 3D map is created before each attempt to remove all residual tumor tissue, which is repeated at increasingly

fluorescence detector sensitivity until the final map displays no residual fluorescence from the tumor paint.



(a) (b) (c)

Figure 3.5: The 3D reconstruction of debulked tumor phantom with fluorescence spots. (a) Image sampling grid 5x6 of SFE above phantom with blue points as the features and small triangles as camera locations. (b) Reconstructed a map from the SFE reflectance images and the corresponding depth map with color scale. (c) Multimodal 3D reconstruction from the 30 SFE locations above the phantom showing residual tumor.

3.4 3D mosaicking on robot RAVEN

The good reconstruction result shows the feasibility of our new algorithm with feature position prediction on micro-positioning stage. Then similar experiment would performed on RAVEN platform which also can provides position and orientation. RAVEN is a surgical robot with two 7 degree-of-freedom (DOF) manipulators, providing a very high degree of dexterity, which is a good candidate for robotically assisted surgery. Beside of remote control by surgeons, RAVEN can operate with command code. In this study, programmable motion of the RAVEN arm was performed.

In the experiment on RAVEN, a new phantom was made with industrial liquid latex (RL-451-80, Silpak Inc., Pomona, California, USA), as the structural material of the phantom. To provide high density and different contrast mechanisms, vessel features were drawn inside of the phantom with red and blue dyes, Fig. 3.6(a). SFE was attached to RAVEN arm and pointing downwards on the tumor phantom, Fig. 3.6(b). In RAVENs code, motion

was generated along a grid of 56 positions (7x8) with equal distance among the neighbor positions. Since RAVENs long robot arm rotates around a fix original point, and the designed motion of SFE has small horizontal translation, so we ignore the even smaller change of orientation of camera.

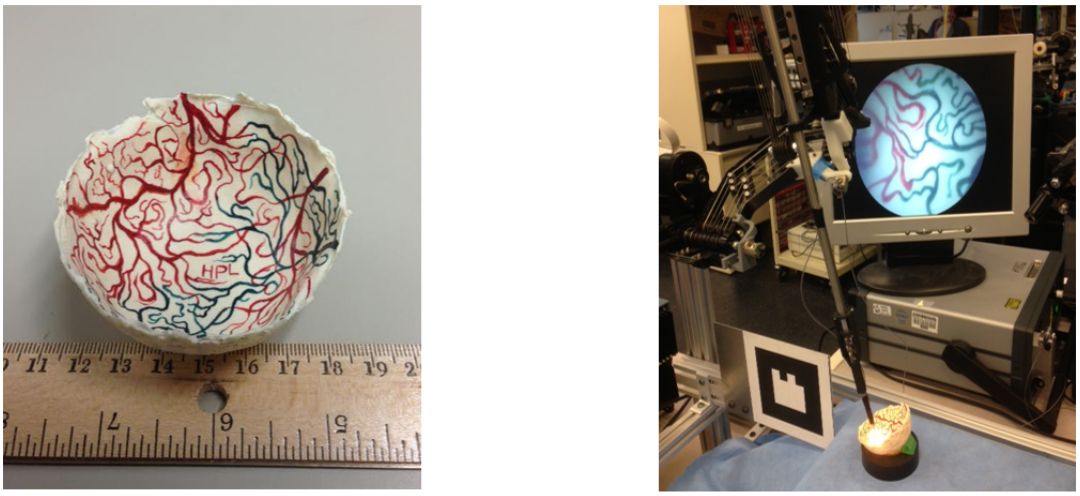


Figure 3.6: The experiment with RAVEN system. (a) A cup-shaped latex phantom to represent brain tissue after tumor debulking. (b) The experiment setup on optics table. The RAVEN arm would hold SFE to scan the tumor phantom with known programmable motion.

It is difficult for RAVEN to provide high accurate position and orientation, since the system was designed for constant visual feedback control by the surgeon. Because each joint in the RAVEN robot arm does not have encoders. The states of the RAVEN robot arm is calculated from encoders mounted on the motors, but the joints [60]. Multiple joints are coupled and connected to motors by a single, long cable. Thus, the measurements on motor do not reflect the real joint angles due to mechanical slack or stretch in cable. As a result, the real movement of RAVEN is different with the programmed motion that we designed in the command code.

By applying the same software with the SFE image captured on the RAVEN system, we have a very different result as shown in Fig. 3.7(a). The reconstructed geometry is clearly

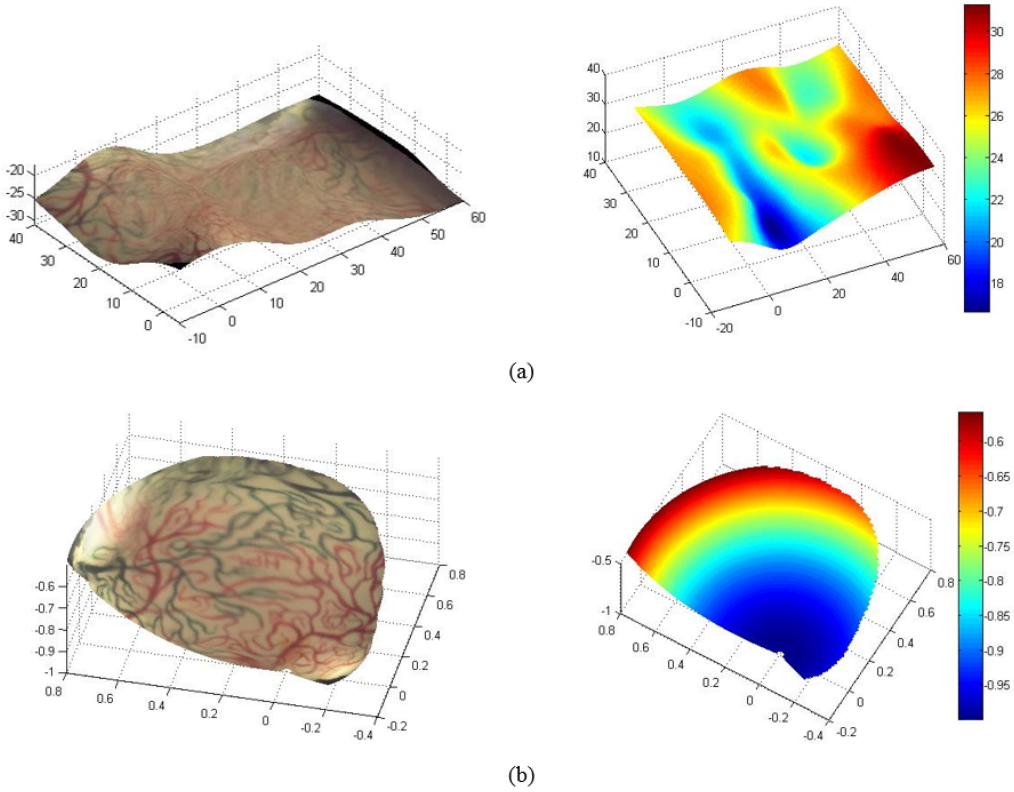


Figure 3.7: The 3D reconstruction of debulked tumor cavity phantom with dense vessel features and no residual fluorescence tumor. (a) Reconstructed 3D surface and depth map based on known camera positions and orientations; (b) Reconstructed 3D surface and depth map solely based on features.

different from the phantom shape. In addition, the features on the 3D mosaic surface looks very blurry. The reason is the algorithm based on known camera motion needs highly accurate information. Slight change of orientation appears to result in totally different reconstructed shape, and also the inaccuracy of camera position leads to fuzzy features in the 3D reconstruction. By applying the previous software solely based on features, the reconstructed model is very close to the phantom shape, and features are more clearly resolved.

This 3D reconstruction process based on known camera takes about 230 seconds for processing all the 56 images, while it takes around 800 seconds for the previous algorithm.

The new approach is much faster method by reducing the searching range during the SIFT and decreasing bundle adjustment iteration, but highly sensitive to the accuracy of camera position and orientation. The previous algorithm does not rely on knowing camera position and orientation, but is about 3.5 slower and requires a large amount of features.

3.5 Summary and discussion

Comparing with the previous 3D surface mosaic algorithm [130], this new reconstruction algorithm proposed in this study is: 1) more efficient, since smaller search region is required to find the matches and better initial shape estimation results in iteration reduced in bundle adjustment; 2) more robust, due to smaller search region reduces the interference of similar features. More matches can be found, which lead to a better initial estimation to bundle adjustment; 3) doesnt require large amount of matching features; 4) yields a multimodal 3D map of the tumor region that can improve visualization and robotic path planning; and 5) physical scale achieved instead of only relative dimensions. This new addition to our surface mosaic software is useful for locating the physical coordinates of the bladder lesion in the surveillance procedure and to provide quantitative characteristics such as size and morphology.

Importantly, the new algorithm requires highly accurate camera position and orientation. In the experiment using a micro-positioning stage, we have good geometry reconstructed at speed of 3 times faster than previous algorithm, Fig. 6. However, the 3D reconstruction with inaccurate camera position and orientation values from the RAVEN data failed to reproduce a representative panorama of the phantom, Fig. 3.7(a). The new algorithm is so sensitive to the position and orientation of camera that variability and inaccuracy of this a priori data creates unusable 3D reconstructions.

To better utilize the information of known camera, especially inaccurate information, a combination algorithm is demanded. In that case, we still can narrow down the search region and speed up the algorithm with good initial shape estimation. But instead of using the known inaccurate camera position and orientation directly, bundle adjustment should be an optimization problem of unknown scene and constrained camera parameters. This a priori knowledge of constraint of camera can be determined by the accuracy of RAVEN

system. Moreover, this future 3D reconstruction approach, which relies on both feature and known but inaccurate camera poses, can correct the inaccurate camera position and orientation by strong feature restraint.

Chapter 4

3D RECONSTRUCTION OF SURGICAL FIELD USING CALIBRATED MEDICAL ROBOT

Brain tumor margin removal is challenging because diseased tissue is often visually indistinguishable from healthy tissue. Leaving residual tumor leads to decreased survival, and removing normal tissue causes life-long neurological deficits. Thus a surgical robotics system of high degree of dexterity, accurate navigation and highly precise resection is an ideal candidate for image-guided removal of fluorescently-labeled brain tumor cells. As the following work of Chapter 3, SFE was also used in this study to collect multimodal images. This miniature flexible endoscope was affixed to the arm of RAVEN II surgical robot providing programmable motion with feedback control using stereo-pair surveillance cameras. To verify accuracy of the 3D reconstructed surgical field, a multimodal physical-sized model of debulked brain tumor was used to obtain the 3D locations of residual tumor for robotic path planning to remove fluorescent cells. Such reconstruction is repeated intraoperatively during margin clean-up so that algorithm efficiency and accuracy is important to the robotically-assisted surgery. Experimental results indicate that the time for creating this 3D surface can be reduced to one third by using known trajectories of robot arm, and the error from the reconstructed phantom is within 0.67 mm in average compared to the model design.

4.1 Biomedical imaging for tumor

Complete surgical resection of tumor tissue remains one of the most important factors for survival in patients with cancer [139]. Tumor resection in the brain is exceptionally difficult because leaving residual tumor leads to recurrence and decreased survival, and removing healthy brain tissue can impair patients memory, mobility and other vital functions. Unfortunately, tumor margins are often indistinct under direct (optical) visualization. Although MRI and CT imaging can locate tumors, their use intraoperatively during surgery is very

difficult and costly [115, 97]. Ultrasound imaging has low resolution, requires tissue contact, and lacks tissue-specific contrast [99, 25]. Optical imaging provides high-resolution non-contact imaging that is familiar with the surgeon, but penetration depth is limited. To enhance optical contrast during intraoperative tumor removal, fluorescence contrast is often used to label brain tumor margins at high resolution and contrast [74, 93, 167]. This technique is especially useful when brain tumors and their margins are exposed to the surgeon so the inherent shallow penetration depth (in white brain matter) of the visible (0.5-0.83 mm) and NIR light (0.9-1.0 mm) [169] is not limiting. The mechanisms for providing fluorescence contrast can range from increased leakiness of capillaries feeding a tumor [105, 29] to photodynamic therapy sensitizers accumulating in tumor tissue [72], to molecularly-specific fluorescence probes [145, 132].

The devices used to image fluorescently-labeled brain tumors range from large surgical microscopes [108], to LED flashing illumination with large camera detection [38], to hand-held cameras systems [152], to goggles worn by the surgeon [75]. The main criteria is having wide-field video imaging that can provide both reflectance and tumor-specific fluorescence image contrast. Another key criterion is the ability to detect at very high sensitivity, typically down to nanoMolar concentrations of fluorescence dye. One more key criterion is the capability to use biocompatible dyes for the translation to human clinical procedures, which limit the choice of visible-NIR dyes to fluorescein and simple derivatives, 5-ALA and other chemical precursors to PpIX [141], methylene blue, and derivatives of ICG [93]. In the future, a criterion will be flexibility and small size to enable brain surgeries through less-invasive keyhole openings in the skull [139]. All of these criteria are met by our laser-based, flexible endoscope, SFE [70].

4.2 Robotic Image Guided Interventions

With the characteristics of flexibility and ultra-thin size, the SFE is advantageous as a guidance scope in robotic Image-Guided Interventions (IGI), which allows for less pain and smaller incisions, quicker recovery and fewer side effects [157]. Since the SFE is so small, lightweight, and flexible, the scope can be attached to a larger surgical tool which can be used for delicate microsurgeries. Thus, the SFE affords a new concept in IGI which

is placing a forward-viewing eye directly on a tool rather than the traditional method of putting the tools through a much larger endoscope for surgical procedures such as NOTES [50, 8] (Natural Orifice Transluminal Endoscopic Surgery). Flexible endoscopes can be used to inspect the internal surgical surface through a keyhole [82] but the drawback of conventional endoscopic surgery is lack of depth information [5], which becomes a highly significant barrier to robotically-assisted brain surgery. This impaired depth perception issue is overcome in the central field of view (FOV) by stereopsis, such as within daVinci® surgical robot series [89, 142, 143]. But the rigid laparoscopes on commercial daVinci® systems are restrictive to straight lines of sight and relatively large surgical fields. Another disadvantage of endoscopic surgery is the inability of providing the surgeon an overview of the entire surgical field in 3D due to limited FOV of the endoscope. Surgeons and their assistants must redraw a mental 3D mosaic of the surgical field as they move the endoscope within the larger surgical field [158]. Thus a 3D panorama reconstruction software of the surgical field is demanded, which also can facilitate the autonomous surgery robot to better control the process of image-guided tasks, such as tumor margin removal [156].

When applied to surgical robotics, the position of a robot tool can be tracked and recorded within the robots motion control software [92, 65, 88]. By placing SFE on a surgical tool directly, the poses of SFEs (camera) can be obtained from robot system. Unlike the previous studies that solely based on images data, this 3D reconstruction can be generated in a more efficient and accurate way by simplifying the procedure of bundle adjustment [160, 43].

In this study, the SFE imaging system was attached directly to one robotic arm of RAVE II [79] as an eye on the tool to guide the removal of residual tumor during simulated robotically-assisted brain surgery. The RAVEN II surgical robot is an open source hardware and software robotic system for research on Robotic Surgical Assistants (RSAs). RAVEN was designed for use in either laparoscopy or open surgery. The robot consists of two arms, which can be tele-operated separately by the user via a remote master. Each arm has 7 degrees of freedom (DOF) providing a very high degree of dexterity, precise cutting and accurate navigation with surgeon-assisted visual feedback [51].

Creating this 3D reconstruction of the surgical field in real-time is advantageous to

robotically-assisted brain surgery due to tissue deformation and shifting that occurs during surgical resection [126] caused by gravity, cerebrospinal fluid, hyperosmotic drugs and etc. [90, 91] and of course tissue removal.

4.3 Methodology

In the preliminary study [43], the experiment on micro-positioning stage provided the feasibility of our proposed algorithm development for reconstructing the surgical field in 3D. This work was extended onto the RAVEN II surgical robotic research platform to validate the feasibility of our proposed image-guided surgical robot system that generates and later relies upon the 3D surface reconstruction of the surgical field. The purpose of this work is to validate the accuracy of this 3D reconstruction model using calibrated trajectories of the robotically-moved endoscope to the original 3D phantom. Thus in this study, an accurate-shape, near-realistic phantom was pre-designed and 3D printed to mimic the debulked brain tumor without matching optical properties of the brain tissue other than surface color and target fluorescence.

The RAVEN has been designed for only visual guidance control and not autonomous computer control. The RAVEN uses a cable driven mechanism. Thus, programmable positioning errors are unpreventable and time-variant, so that the direct positioning control of the robot under the autonomous motion mode is inherently imprecise [60]. This unreliability of the camera pose information from the RAVEN system readings leads to the unsuccessful 3D reconstruction [43]. This study shows that an external optical tracking system for the RAVEN provides the needed camera pose information so the two major goals of this project can be attained. These two goals are demonstrating (1) accurate 3D virtual reconstruction of the geometry of the surgical field phantom with fluorescence features, and (2) algorithmic approach to providing these reconstructions at near-real time for repeated implementation in a robotic-assisted surgical procedure.

4.3.1 3D Reconstruction Algorithm with Feature Position Prediction

In previous studies, a surface mosaic software was developed to produce 3D internal panoramic surface solely based on a bladder surveillance video [130, 170]. By attaching SFE directly

to an accurate RAVEN robot arm, the positions and orientation of SFE can be read out from RAVEN system. The efficiency and accuracy of 3D reconstruction can be improved by applying these values of SFE pose as pre-knowledge [160]. To take advantage of programmable surgical robot RAVEN, we set the distances of each neighboring SFE position as equal in this study to boost the efficiency of reconstruction algorithm.

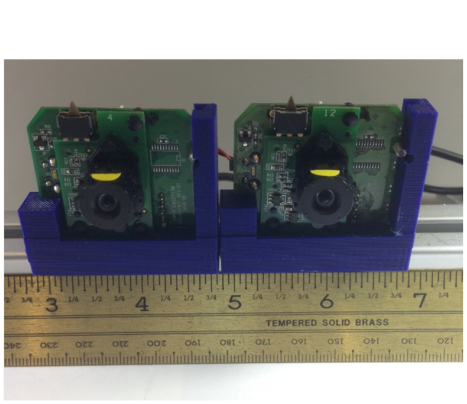
With equidistant SFE positions, the homography of one pair of neighbor SFE frames can be estimated and then would be applied to other pairs of neighbor frames to approximately predict the positions of feature points, we call it *feature position prediction*. Thus, instead of searching whole image to find the matching feature, we reduce our search to a small region around the prediction position, which makes the algorithm more efficient.

4.3.2 Stereo System on RAVEN II Platform

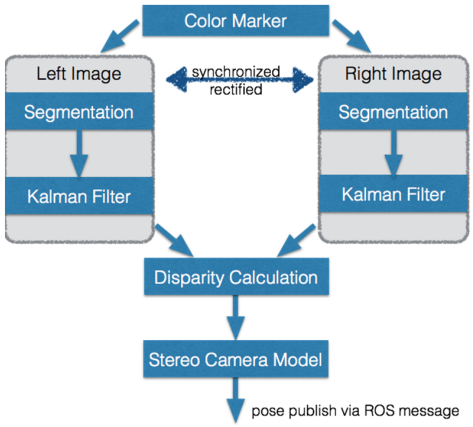
Numerous tests of RAVEN accuracy were carried out by sending commands of different pure translations in x, y, z direction, and comparing the deviation between command and real movement. The conclusion was drawn that the accuracy of RAVEN was changing based on different cases, such as the operation time and initial location of RAVEN arm.

To address this inaccuracy issue, one possible solution includes adding external tracking system to obtain the actual pose information of the robot end-effector. Previous tracking systems are optical based systems, which are sensitive to the lightening condition. For example, Kehoe et al. [60] used two high resolution industrial cameras for the construction of the stereo tracking system in their autonomous multilateral debridement task on the RAVEN surgical research platform. Mahler et al. [81] used PhaseSpace LED-based motion capture that consists of 4 cameras to record the true pose of the robot end-effector. These trackers can provide accurate pose information of the robot arm but nonetheless are expensive. Alternative solution to improve the control precision of a cable-driven robot is to install encoders to each joint directly in joint space. In this case, the encoders reflect the true joint angles, but the robot could become very bulky due to installation limitations and it might be difficult for sterilization due to electric cable wiring.

In our study, a custom optical tracking system was utilized to track color markers in a



(a)



(b)

Figure 4.1: The stereo system. (a) The hardware of stereo system with two webcams mounting on an aluminum bar. (b) The workflow of the stereo tracking algorithm, Kalman filter was utilized for robust tracking.

pair of stereo images, which were provided by two low-cost Logitech QuickCam Communicate Delux webcams. Both webcams were snapped into a 3D-printed base that can be mounted on an aluminum bar with adjustable base distance as shown in Fig. 4.1(a). The optical axes of the two webcams were aligned approximately parallel in this setup to provide a reasonable field of view based on the robot working area. The standard USB Video Class (UVC) driver was used for image streaming. The synchronization of the two webcams was implemented on the software level using the approximate time ROS message filter. The openCV library was utilized for image processing to segment the markers from image. For the purpose of robust tracking, a Kalman filter with four internal states was implemented for de-noising in each image frame. The workflow of the stereo tracking system is illustrated in Fig. 4.1(b). The average performance of the stereo tracking system using a 4mm x 4mm color marker in a about 20mm x 15mm workspace is listed in Table 4.1. The accuracy of x and y direction is 0.2 mm, which is higher than the accuracy in z direction.

4.4 Experiment

In the preliminary study [43], the 3D reconstruction of debulked brain tumor phantom was generated with known SFE (camera) position and orientation on micro-positioning stage. The reconstruction algorithm with feature position prediction, took 1/3 computation time comparing with conventional one. The general shape of the 3D virtual reconstruction appears similar to the phantom, demonstrating the feasibility and reliability of new algorithm, but there was no dimensional comparative analysis between the real phantom and virtual 3D model. To accurately navigate robot and resect tumors, quantification was analyzed in this study with ICP algorithm.

All the codes were MATLAB implemented, and running on a workstation Dell Precision M4700 with 2.7 GHz Intel i7-3740 QM CPUs, 20.0 GB memory in a 64-bit Window 7 operating system.

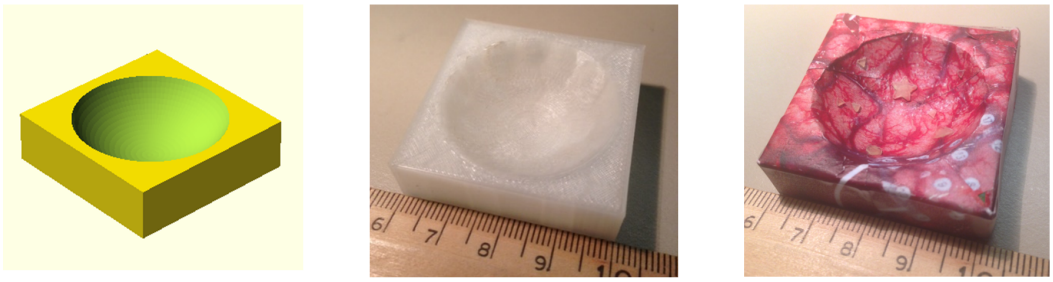
4.4.1 Phantom Fabrication

A cup-shaped latex phantom (3cm x 4cm) was made manually to mimic the debulked tumor cavity in preliminary study [43]. It is representative with irregular shape and multiple scattered weak fluorescence spots to imitate the practical surgical field. Different with the previous phantom, a new phantom was made to validate and quantify the reconstruction result with CAD designed geometry and provide more realistic coloration.

In this study, a 3D model of spherical dome was designed using OpenSCAD with maximum radius of 15 mm and depth of 10 mm, Fig. 4.2(a). This designed model was saved as a STereoLithography (.stl) file and read by the 3D printer MakerBot Replicator (MakerBot Industries, Brooklyn, New York). The phantom Fig. 4.2(b) was printed using PLA (Poly-

Table 4.1: The performance evaluation of the custom optical tracking system

Camera Resolution	1280 pixels
Tracking Frequency	7 Hz
Accuracy	x axis: ± 0.2mm; y axis: ± 0.2mm; z axis: ± 0.5mm



(a) (b) (c)

Figure 4.2: Phantom fabrication. (a) The CAD-designed phantom model of a spherical dome with radius 15mm and depth 10mm. (b) The 3D-printed model of debulked brain tumor cavity with 4.0-cm square PLA material. (c) The phantom with near-realistic surgical features and multiple scattered fluorescence targets representing the tumor residual tissue.

lactic Acid) material. In order to make a phantom for dimensional tests that also allowed multiple image contrast enhancement on the surface of the 3D surface, a near-realistic phantom was made without adding the complexity of light penetration deeply into the substrate. Specifically, colored paper from a photograph of actual brain surgery[1] (featured with cortex, blood vessels, and even electrocorticography numbered electrodes) was glued over the PLA substrate, Fig. 4.2(c). This limited the laser light penetration to approximately the thickness of the colored paper, reducing uncertainty in our 3D measurement.

Also for quantification purposes, eight fluorescence targets of known shapes and sizes were scattered on the surface of spherical dome phantom, representing the residual tumor at the margins. These targets are fabricated and cropped from a dye-in-polymer material that is formed by dissolving the Fluorol dye (Exciton Inc., Dayton, Ohio) powder into polyurethane resin (AquaClear Resin, ArtMolds, Summit, New Jersey) [165]. By illumination with a 444 nm laser (blue), these fluorescence targets emit green light but are nearly clear and colorless in RGB reflectance imaging.

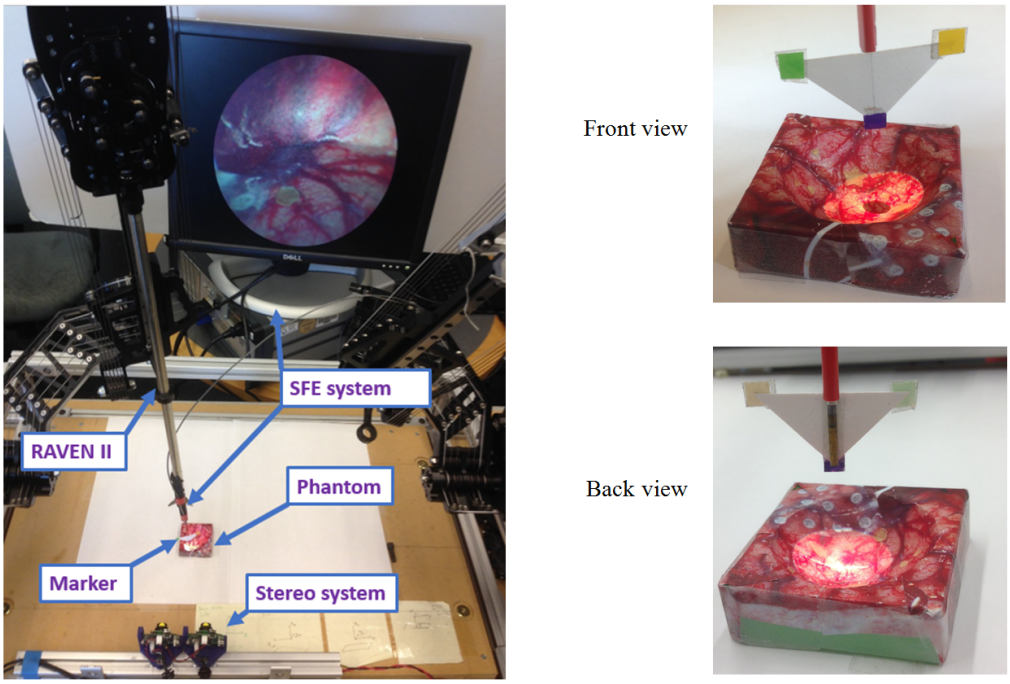


Figure 4.3: The 3D reconstruction experiment with SFE and RAVEN II system. (a) The experiment setup. RAVEN arm would hold SFE to scan the tumor phantom along planned trajectory with stereo system as close loop feedback. (b) Front and back view of the layout among SFE, marker plate and spherical-dome-shaped phantom; the SFE was aligned with the triangle height to simplify the transformation matrix between SFE frame and marker frame.

4.4.2 Experiment setup and procedure

In the experiment, SFE was affixed onto one robotic arm and pointing downwards the tumor phantom that placed on the horizontal workspace of RAVEN, as shown in Fig. 4.3(a). A triangular marker plate with three distinguished color dots in each corner was attached to SFE probe, Fig. 4.3(b). The triangular height of the marker plate was approximately aligned with the optical axis of the SFE to simplify the transform estimation between the marker and the SFE. The end-effector of the robot arm was programmed to automatically travel along a 2D 5x6 grid map, paralleling to the workspace. All grid nodes were equally spaced

[1]See the website: http: /dxline.info/diseases/brain-surgery#prettyPhoto.

in distance of 5mm. The SFE has depth of focus of 4-40mm that matches conventional flexible cystoscopies [122]. Bright image with plenty fine details can be captured when SFE is close to the object; relatively dark image with larger scene is obtained as SFE moves further away from the object. In this experiment, the front end of SFE is located about 15mm away from the tumor phantom, where we can capture images with enough features and also guarantee more than 50% image overlap among neighbor frames for reconstruction algorithm.

With stereo system as close-loop feedback, RAVEN arm can move along the grid nodes very accurately. At each grid node, a reflectance and fluorescence image was taken respectively. Since the fluorescence signal in vivo is expected to be weak from the residual tumor tissue, the SFE is expected to remain stationary at these preprogrammed locations of the robotic arm to increase signal to noise by video-frame averaging. In the meantime, the locations of the colorful markers were recorded for three seconds. Average value of each location was calculated to minimize the noise issue of stereo tracking system, so that better estimation of the SFE pose (position and orientation) can be achieved.

4.4.3 Multimodal Imaging

As florescence imaging can enhance tumor delineation, a flexible, wide-field calibrated multimodal SFE of 1.6 mm outer diameter was utilized to capture both reflectance and fluorescence images in real-time at high-resolution. Moreover, the reflectance contrast of the vascularity can be increased with enhanced spectral imaging (ESI) mode of the SFE scope by enhancing the blue channel over the red channel, which shares the same concept as narrow band imaging [96] that is widely used to improve the visibility of blood vessels. The 1.6-mm SFE has larger multimode optical fibers surrounding the fiber scanner to improve fluorescence sensitivity compared to the 1.2-mm SFE. As shown in Fig. 4.4(a, b, c) there are many more features in the reflectance and ESI images versus the fluorescence images. A grayscale threshold was set to highlight the residual tumor based on pixel grayscale value in the fluorescence image. The white points/regions represent tumor in Fig. 4.4(d), which may look noisy due to high PMT gain to capture weak florescence signal. The tumor region

was filtered and segmented by image morphology technique of image opening and closing, shown in Fig. 4.4(e). Multimodal image can be generated by merging reflectance with a green fluorescence overlay, Fig. 4.4(f).

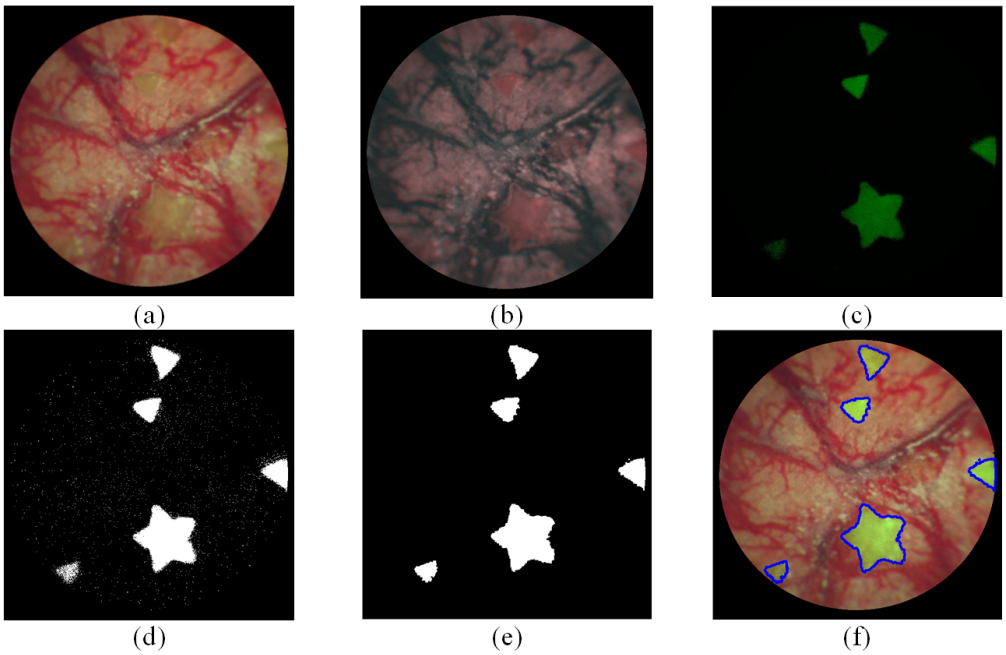


Figure 4.4: The reflectance and fluorescence SFE image of tumor phantom. (a) The standard reflectance imaging of SFE; (b) The enhanced spectral imaging of SFE; (c) The fluorescence imaging by deactivating red and green laser source. The florescence targets represent tumor residual, emitting fluorescence signals (green) under blue laser activation; (d) The white points or regions represent tumor cells by filtering with a specific grayscale threshold; (e) The segmentation of tumor regions with image morphology techniques; (f) Multimodal 2D image that is generated by merging color reflectance and fluorescence. The blue curvatures represent the outlines of tumors.

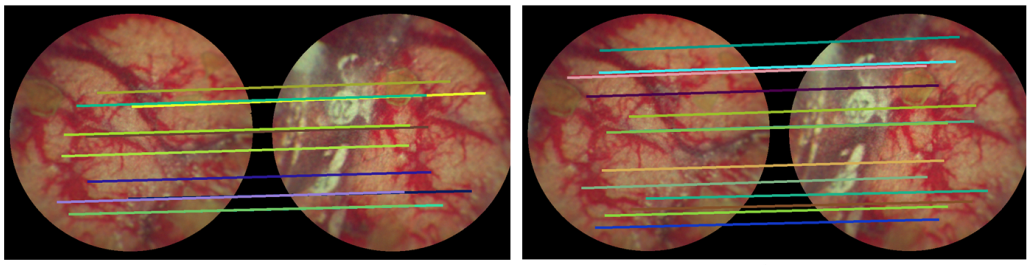
4.4.4 3D reconstruction

In the first step of 3D reconstruction, SIFT features were detected and characterized with open source VLFeat toolbox[2]. In each pair of frames I and II, shown in Fig. 4.5, the feature-space distances were calculated for each feature of frame I to every single feature of frame

[2]see the website of vlfeat at www.vlfeat.org

II in conventional algorithm, and found out the features of frame II with the nearest and the second nearest distances. To keep good (real) matches, the ratio of nearest and second nearest distances was preserved to be less than a threshold [130]. We set 0.8 as the threshold in this study to eliminate unreliable matching.

With accurate position and orientation of the SFE, the homography between image pairs can be estimated. Thus, the positions of matching features can be roughly predicted. A search window was set around the prediction position so that the real matching feature can be detected in a more robust and more efficient way comparing with the previous algorithm. A search window with radium of 100 pixels was applied in our study.



(a) (b)

Figure 4.5: Matching features in a pair of frames I and II were detected with/without feature position prediction and lined up with colorful lines. (a) Less matching pairs were found by the conventional algorithm. (b) More matches were detected with the prediction of feature points.

In addition to the improvement of computation efficiency, our new algorithm was advantageous to the robustness of the reconstruction, especially for insufficient image overlap (<50%). Considering an image contains several similar features, the ratio of the nearest and second nearest distance may not satisfy the ratio threshold. By searching a small region around the feature prediction position, we can reduce the interference from similar features that are located in other parts of the image. In this way, our new algorithm can keep more matches than the previous work [130]. Figure 4.5 shows the matching feature pairs with/without feature position prediction approach. Taking the two images in Figure 4.5 as an example, 216 pairs of matching points were found by conventional method, Fig. 4.5(a), while 278 pairs were detected with feature position prediction, Fig 7(b). One twentieth of

the matching points were randomly picked and lined up for better illustration in Fig. 4.5.

Bundle adjustment was implemented after matching the feature points. By giving the known positions and orientation of the endoscope for each frame, bundle adjustment simplifies greatly by only solving the 3D coordinates of each feature. And more robust matching features were provided by narrowing down the search region. These improvements make the 3D reconstruction algorithm more efficient and robust.

4.5 Result

The 3D panoramic mapping procedure of the spherical-dome fluorescence tumor phantom is illustrated in Fig. 4.6. Fig. 4.6(a) represents the reconstructed 3D map of the phantom with 30 positions 5x6 of the SFE (camera) above, and the features as blue points used for mosaicking the images together into the panorama. Fig. 4.6(b) shows the overview of 3D panoramic image in reflectance color contrast only, in which the eight tumor targets are difficult to see. Fig. 4.6(c) is the corresponding depth map of reconstruction in pseudocolor. Fig. 4.6(d) is 3D surface reconstruction of the entire phantom with multimodal color contrast, having bright green targets representing the fluorescence residual tumor. Image morphology method was utilized to segment the tumors, which were encircled with blue curves and marked with series number labels. Eight tumors were detected and labeled in series, coincident with the total number of florescence targets used in phantom fabrication. The shape and size of the tumor targets were also well recovered in 3D virtual reconstruction. Taking the star tumor (#4) as an example, the average length of the edges was 6.2 mm in reconstruction. The error was 3.3% comparing with the actual fluorescent star (6 mm edge) in the phantom. The 3D virtual overview with residual green tumor may be visualized in both reflectance (showing complete vascularity) and multimodal view with pseudocolor replacing reflectance contrast at tumor locations by slowly flashing between visualization modes to maintain highest information content according to the surgeons requirements.

To quantify the reconstruction result, ICP [11] was employed to align and compare the reconstructed virtual model to the CAD designed model (ground truth model) [19], Fig. 4.6(e). The CAD model is illustrated by the blue dots as a reference point cloud, while the reconstructed one was translated and rotated to best match the CAD design by mini-

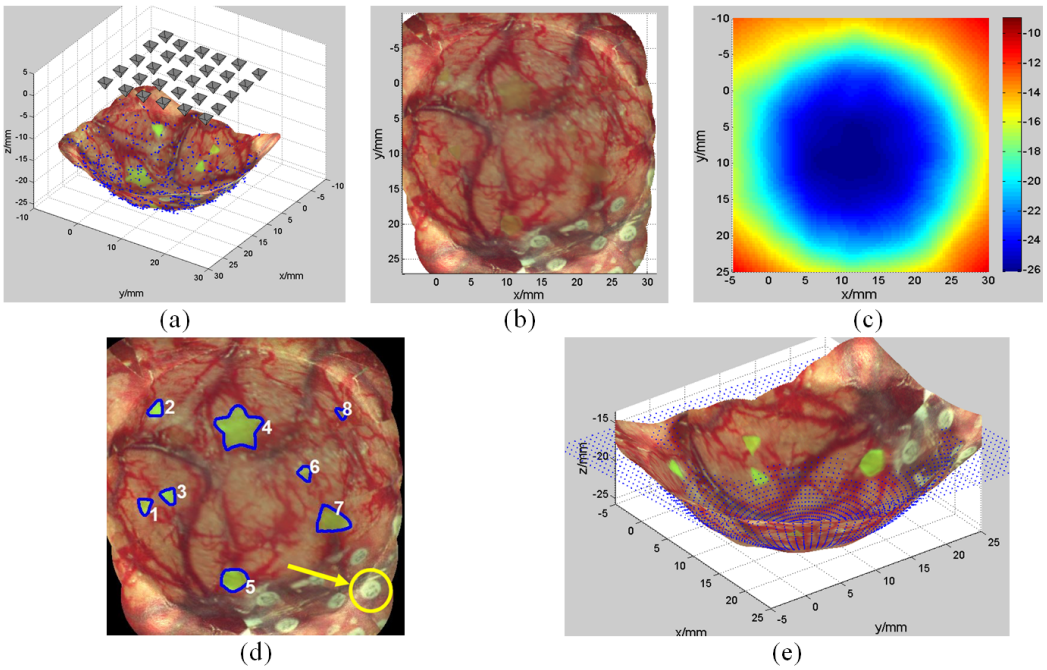
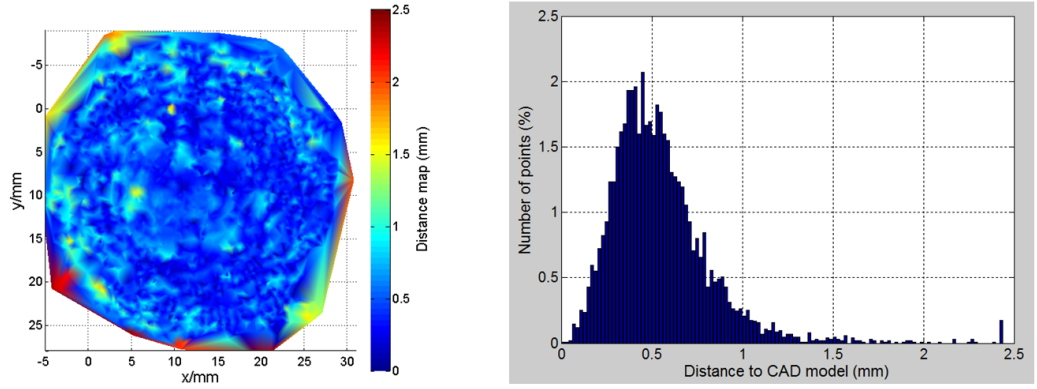


Figure 4.6: The 3D reconstruction of spherical-dome tumor cavity phantom with surgical image and residual fluorescence tumor targets. (a) Image sampling grid 5x6 of SFE above phantom with blue points as the features and small pyramids as SFE locations. (b) A reconstructed map from the SFE reflectance images. (c) The corresponding depth map with pseudo-color scale. (d) Overview of multimodal 3D reconstruction showing residual tumor with outlines and labels (#13 is circled with arrow). All of eight tumor targets were detected. (e) The comparison of reconstructed 3D model to the CAD designed model, shown as blue point cloud.

mizing the mean squared error (MSE) cost function of the distance from each reconstructed feature point to the closest point in reference point cloud. To demonstrate that the ICP result provides a reliable alignment, eight different positions and orientations around the CAD model were chosen as the initial poses of the reconstructed model [103]. The deviation between 3D virtual reconstruction and CAD model was calculated after alignment. The obtained ICP error values were in the range of 0.6736-0.6762 mm with average of 0.6742 mm and standard deviation of 0.874 μm, which indicated the fine alignment does not rely on the initial poses. The mosaic image was a little blurred comparing with the original SFE image, as shown in Fig. 4.6(d). But the numbered electrode patches are still readable in

the 3D mosaicked image, such as #13 patch with yellow arrow in Fig. 4.6(d).

To better quantify the error of reconstructed and CAD model, one alignment was chosen for further error analysis. The error distribution map is shown in Fig. 4.7(a). Most of the error values were between $[0 - 1]$ mm and located in the central region. Most of the high-value error (>1.0 mm) located along the phantom flat periphery due to insufficient density of features. These features of large error can be easily eliminated based on their spatial locations. Fig. 4.7(b) showed the histogram of error distribution which indicated the average error is about 0.6mm. The error values in the range of 0-1mm occupied 96.3% of all the data.



(a) (b)

Figure 4.7: Error distribution between reconstructed and CAD model. (a) The error distribution map overview with main low-value error points locate in the spherical dome area. (b) The error histogram showing the average error is about 0.6 mm.

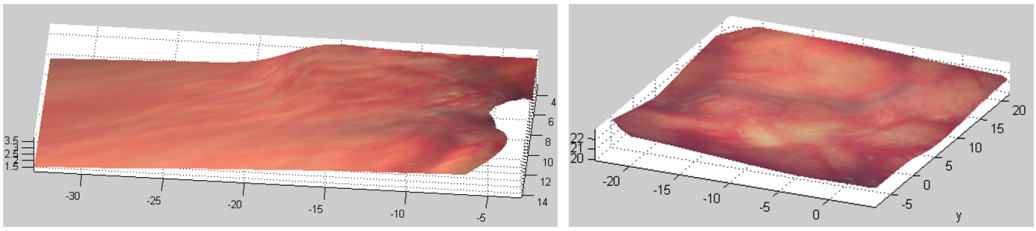
This new 3D reconstruction process based on known SFE (camera) poses takes about 280 seconds for processing all the 30 reflectance images, while it takes around 980 seconds for the conventional algorithm. The new approach is 3.5 times faster by reducing the searching range during the SIFT and decreasing bundle adjustment iteration. However, the drawback is that this new algorithm is highly sensitive to the accuracy of camera position and orientation. The conventional algorithm does not rely on knowing camera position and orientation [130], but is much slower ($> 3\times$) and requires a large amount of features in all

the images [43].

4.6 Discussion

Comparing with the previous 3D surface mosaic algorithm, this new reconstruction algorithm proposed in this study: 1) is more efficient, as >3× efficiency was achieved since the matching features can be found more quickly within a smaller search region and the iteration can be reduced in bundle adjustment because of better initial shape estimation; 2) is more robust, due to smaller search region reduces the interference of similar features in the same image. More matches can be found, which lead to a better initial estimation to bundle adjustment; 3) yields a multimodal 3D map of the tumor region that can improve visualization and robotic path planning; and 4) generates physical-scale 3D model instead of a relative dimensions, with tumor targets of actual shape and size recovered. This real size reconstruction is useful to locate the physical coordinates of the brain tumor residual in the removal procedure and to provide quantitative characteristics such as size and morphology. But the reconstruction algorithm is very sensitive to the input of SFE pose data. Inaccurate SFE position and orientation may cause a clearly different geometry away from the phantom [130]. To clearly demonstrate this dimensional error when the stereo surveillance system is not providing accurate camera pose additional tests with flat square paper with our surgery scene was performed, as shown in Fig. 4.8. The reconstruction with and without accurate data varies dramatically as the square planar paper scene is highly distorted in Figure 4.8(a). The motion control of the robot arm with stereo vision augmentation facilitates the reconstruction process significantly, Fig. 4.8(b).

Precise control of the cable-driven surgical robot under the autonomous mode is challenging due to the nonlinear factors such as mechanism backlash, cable stretch under tension, and the imperfection of the available kinematic model. One possible solution includes the application of an external tracking system to get accurate pose data of robot arm as demonstrated in this work. Another approach is to mount encoders to each joint directly to measure the real joint angles, but the robot could become very costly and bulky. In our study, an economical and effective stereo tracking system was built to improve the positioning precision of the SFE and provide high accurate pose data to the 3D reconstruction algorithm.



(a) (b)

Figure 4.8: The 3D reconstruction of a flat printed paper of a surgical scene at low resolution. (a) The 3D reconstruction with the data from RAVEN autonomous mode, showing unreasonable shape 40x10x3 mm and no recognizable features in a highly distorted mosaic image. (b) The 3D reconstruction with the data from stereo surveillance system, exhibiting as a square planar surface and recognizable features such as vessels.

The accuracy of stereo system depends on the webcam image resolution, distance between stereo to markers and lightening condition. A homogeneous light environment decreases the tracking noise greatly which stabilizes the robot control quickly.

Experiments showed the environmental lighting plays a more significant role than the camera resolution in marker tracking, which will be an issue for practical surgery since lighting is not homogeneous. However we can take advantage of other information to enhance the robustness, such as the shape and size of the markers. Moreover, another drawback of this external stereo tracking system is that it may not be feasible within an enclosed surgical space, such as minimally invasive surgery (MIS) with flexible tool shafts.

Florescence imaging is also influenced by the environment lighting since the emission signal is very weak. We darkened the environment and raised the PMT detection sensitivity to capture the florescence signals in this study. However, in the future the SFE may be used in an enclosed keyhole surgical field (MIS) which can be sheltered from interference of environmental lighting, like daVinci Si^{TM} robot with FireFly^{TM} florescence imaging system.

The combination of multimodal SFE, stereo assisted RAVEN robot, 3D reconstruction with feature position prediction and a spherical-dome shape phantom containing florescence targets, an accurate 3D virtual model representing debulked brain tumor cavity can be generated with $>$ 3x efficiency with all the tumor targets outlined and labeled. For the future

work, a more advanced reconstruction algorithm is required to eliminate the feature blurring in the stitched image, which is caused by imprecise calibration of SFE internal parameters, such as focal length, field of view and distortion. As we know, any measurement system has its own accuracy, which is typically a range, such as 0.2 mm for our stereo system. Instead of using the known SFE position and orientation as fixed values directly, bundle adjustment needs to be improved to optimize the problem of box-constrained parameters (3D scene, SFE pose, focal length, field of view, distortion and etc.). In this way, we still can narrow down the search region and speed up the algorithm with good initial shape estimation, from which a more accurate and precise 3D reconstruction model with less blurred features can be generated.

Future work will also require a more realistic phantom that can mimic the optical properties and even allow for material resection to mimic image-guided surgery. This initial phantom was simplified with the main property of dimensional accuracy of the 3D-printed PLA model and glued photographic paper and fluorescence surface targets for multimodal contrast enhancement because this work focused on validating the dimensional accuracy of our 3D reconstruction algorithm to the surface of this phantom. Future phantoms will require more translucent materials that provide multimodal contrast such as dyed gels containing scattering particles [49] The versatile SFE can support multiple modes of imaging, such as reflectance using enhanced spectral imaging which enhances contrast of the blood vessels and vascularized tissues [70], and wide-field fluorescence imaging of near-infrared dye molecules that are conjugated to molecular specific probes that bind to brain tumor cells [167]. Additional SFE imaging modes that are molecular-specific can also be included in a future phantom or *ex vivo* tissue model, such as autofluorescence imaging [176] and time-resolved fluorescence lifetime imaging and spectroscopy [119] and coherent Raman imaging and spectroscopy [109].

Finally, a systematic analysis of error between reconstructed and CAD-designed models needs to be performed in the future, such as geometry, size and location of each tumor target. Moreover, the computation procedure of new algorithm was not fast enough (> 200 sec for 30 frames although the efficiency rose to $> 3\times$), since tissue deformation and shift occurs during surgical resection. Also the 3D reconstruction of surgical field needs to perform

in an iterative fashion to remove all clearly visualized residual tumor tissue. Thus, faster computation can reduce the surgery time and result in a more precise virtual model. In the future, the algorithm can be boosted by written into C++ and partially implementing with parallel computing platform CUDA.

Chapter 5

TOWARD REAL-TIME 3D IMAGE-GUIDED SURGERY

The challenge is to accurately guide the surgical tool within the 3D surgical field for robotically-assisted operations such as tumor margin removal from a debulked brain tumor cavity. The proposed technique is 3D image-guided surgical navigation based on matching intraoperative video frames to a 3D virtual model of the surgical field. A small (1.6-mm diameter) laser-scanning endoscopic camera was attached to a mock minimally-invasive surgical tool that was manipulated toward a region of interest (residual tumor) within a phantom of a debulked brain tumor. Video frames from the endoscope provided features that were matched to the 3D virtual model, which were recorded earlier by raster scanning over the surgical field. Camera pose (position and orientation) is recovered by implementing a constrained bundle adjustment algorithm. Navigational error during the approach to fluorescence target (residual tumor) is determined by comparing the calculated camera pose to the measured camera pose using a micro-positioning stage. From these preliminary results, computation efficiency of the algorithm in Matlab code is near real-time (2.5 sec for each estimation of pose). The computation efficiency can be improved by implementation in C++. Error analysis produced 3 mm distance error and 2.5 degree of orientation error on average. The sources of these errors come from 1) inaccuracy of the 3D virtual model, generated on a calibrated RAVEN robotic platform with stereo tracking; 2) inaccuracy of endoscope intrinsic parameters, such as focal length; and 3) any endoscopic image distortion from scanning irregularities. This work demonstrates feasibility of micro-camera 3D guidance of a robotic surgical tool.

5.1 3D image-guided surgery

In the previous study (Chapter 3 and Chapter 4), the SFE is attached to a calibrated robotic arm and scanning across a phantom of a cavity from debulked brain tumor with

fluorescence targets representing residual tumor tissue, accurate 3D virtual reconstructions of the surgical field can be generated [42].

Once the fluorescence target is located, the SFE-guided surgical tool tip needs to be safely navigated against the target for the residual tumor clean-up operation. The challenges are providing accurate and efficient (time and space) 3D-image guidance that can run in tandem with a surgeons visual feedback control. The tedious process of cleaning up residual tumor cells by their fluorescence glow may be automated in the future, such as the repeated suturing tasks than have been automated in robotic-assisted surgeries. Figure 5.1 shows the scheme of 3D image-guided surgery. The 3D virtual model of the debulked tumor cavity was generated earlier (Chapter 4). As the surgical tool approaching to the region of interest, the pose of the scope can be recovered by registering the 2D live endoscopic image to the pre-known 3D virtual model. So the pose of the tool can be also recovered since the spatial relation of tool and scope is fixed. Thus, an optimal surgical path of the tool can be designed to better accomplish specific surgical tasks. And also the path can be re-adjusted along the way towards to the residual tumor, which is shown with series numbers 1-3 in Figure 5.1.

Advantages of providing a 3D image-guided approach to navigate a surgical tool in a delicate minimally-invasive surgery like residual brain tumor clean-up are listed: 1) Optimize the surgery: Comparing to conventional 2D endoscopic-guided surgery, a 3D virtual model of surgical field provides depth information, which can be used to assist the surgeon to plan the surgical procedure, or an autonomous medical robot to generate an optimal surgical/therapy path. 2) Secure the surgery: Depth information in the 3D image-guided surgery can define a safe working space, so that the damage of the tissue or the endoscope can be avoided as the surgical tool is moving around in the surgical field. 3) Assist the key-hole surgery: An external stereo optical system [42] can be utilized to track the pose of surgical tool and/or endoscope in open surgery, but this is not feasible within enclosed surgical fields, such as key-hole surgeries. The proposed 3D image-guided technique in this study can recover the position and orientation information of the endoscope (eye on robot arm) in near real-time. 4) Refine the 3D virtual model of surgical field by adding more and updated endoscopic images: the initial 3D model was generated by raster scanning over the surgical field. New higher-resolution detail can be added to refine the 3D model as the tool is moving towards

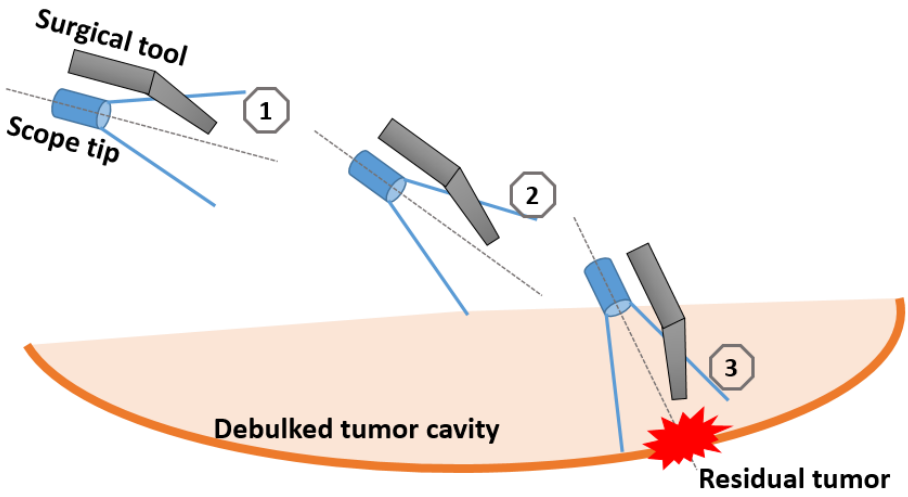


Figure 5.1: The scheme of image-guided surgery based on a 3D virtual model. The pose of scope/tool can be recovered in near real-time by registering 2D endoscopic video frame to the 3D virtual surgical field model. The optimal path of surgical tool can be designed and re-adjusted towards the residual tumor, showing with series numbers 1 to 3.

to the tumor. 5) Adjust the motion of surgical tool: As the pose of the tool in 3D virtual model frame is known more rapidly in the future, the precise control of the surgical tool can be achieved by taking the recovered pose information as a close-loop feedback.

The estimation of the pose of SFE by registering a 2D image to a 3D virtual model can be achieved in two stages. The first stage is to find the matching features between them. SIFT (scale-invariant feature transform) and RANSAC (random sample consensus) algorithm was utilized to find the matching points from a video frame from the live endoscopic video, and the stitched image of 3D model. The second stage is to refine the pose parameters by these matching points with an optimization approach. Instead of solving 3D scene and camera parameters simultaneously, a constrained bundle adjustment was used to solely recover pose information in our study since the 3D scene and its 2D projection were known.

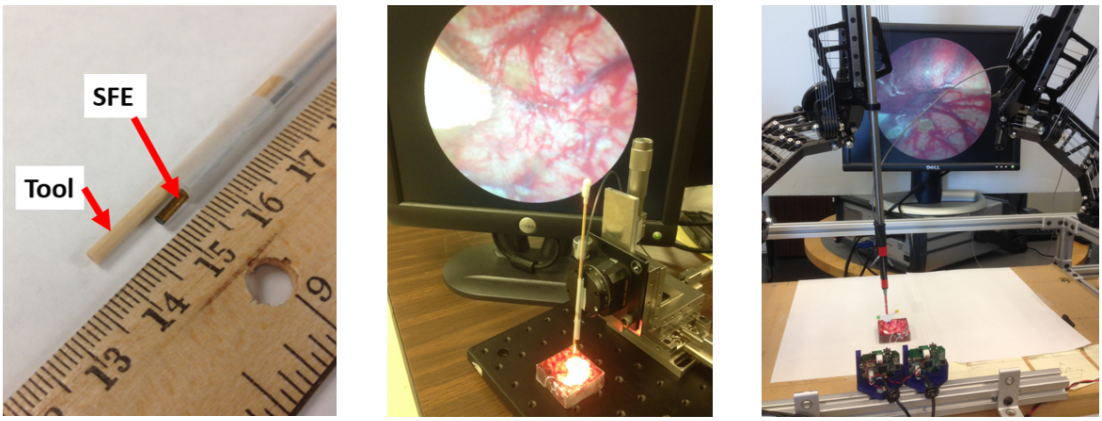


Figure 5.2: The experiment setup. (a) A multimodal SFE of 1.6mm outer diameter was affixed along a surgical tool. (b) The tool with SFE was attached on a micro-positioning stage. Tool tip shown in video. (c) The tool with SFE was attached on RAVEN II surgical robot to create 3D model.

5.2 Experiment and result

In this proof-of-concept study, a 1.6mm multimodal SFE was affixed along with the mock surgical tool (Figure 5.2a) that was attached on a micro-positioning stage, see Figure 5.2b, which can provide accurate position and orientation information as the ground truth to analyze the feasibility of the algorithm. The same 3D printed phantom in previous study [42] was used to mimic the surgical field with realistic surgical features and multiple scattered fluorescence targets. This concept will be extended to surgical robot RAVEN II for more application-specific development, such as testing autonomous surgical procedures, see Figure 5.2c. Accurate multimodal 3D model of surgical field was previously generated on RAVEN surgical robot platform [42] that is used directly in this study as prior knowledge. The 3D model shows that fluorescence markers were detected and highlighted in the real-size 3D phantom, representing the residual tumor after the debulking procedure, see the green targets in Figure 5.3a, from which an optimal path of autonomous surgery can be designed.

In this study, the proof-of-concept experiment was setup on a micro-positioning stage as

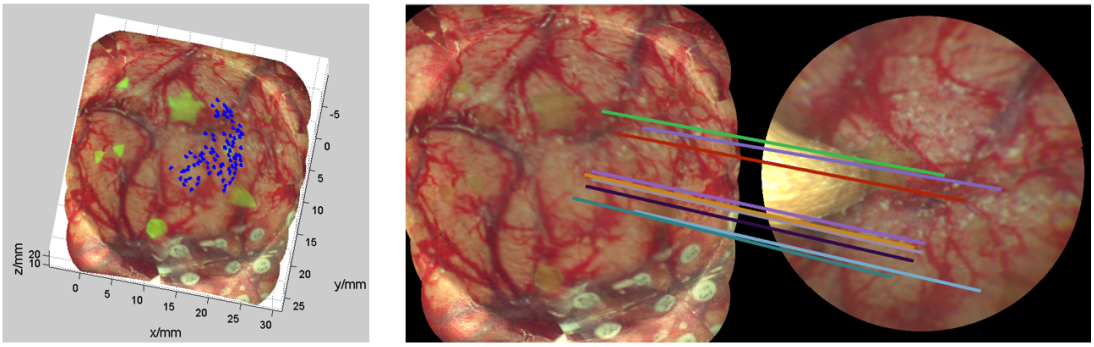


Figure 5.3: The feature matching of 2D endoscopic image with 3D virtual model. (a) The 3D virtual model of surgical field with matching features to SFE video frame labeled in blue. (b) Matching features between 3D reconstructed virtual model of phantom (left) and a single SFE video frame (right), with feature alignment shown by the multiple colored lines.

shown in Figure 5.2B to analyze the feasibility of our proposed method. Since the tool was extended along the optical axis of the SFE, some field of view of SFE was blocked by the tool, see Figures 5.2B and Figure 5.3B. Five arbitrary poses of the tool/SFE were selected and read from the micro-positioning stage as ground truth. Five reflectance images were captured to register with the pre-reconstructed 3D virtual model. SIFT and RANSAC was utilized to find the matching points. Taking the image in Figure 5.3 as an example, 92 pairs of matching points were found even with some field of view being blocked by the tool; shown as a blue dots in Figure 5.3A, and 1/10th of matching points randomly selected for visualization check in Figure 5.3B.

The estimation of the five poses was achieved successfully with our constrained bundle adjustment algorithm. All software was implemented in Matlab, running on a workstation Dell Precision M4700 with 2.7 GHz Intel i7-3740QM CPUs, 20.0 GB memory in a 64-bit Window operating system. The difference of pose resulted in various computation details, such as number of matching points, computation time, and distance and orientation error to the ground truth, see Table 5.1.

The difference of number of matching points is caused by the difference of perspective and also the blocked field of view in the endoscopic images. The computation time was

Table 5.1: The experiments results of five different poses of SFE

	No. of matching points	Computation time	Distance Error	Orientation Error
Pose 1	54	2.31 sec	3.12 mm	1.47°
Pose 2	48	2.45 sec	3.08 mm	1.52°
Pose 3	92	2.67 sec	3.11 mm	2.64°
Pose 4	68	2.66 sec	3.06 mm	4.47°
Pose 5	84	2.27 sec	3.00 mm	2.10°
Avg.	(NA)	(NA)	3.12 mm	2.44°
Std.	(NA)	(NA)	3.12 mm	1.23°

relatively stable, and majority of computation time was devoted to finding the matching points between captured video frame and reconstructed 3D model. Experiments showed the constrained bundle adjustment took only 0.3 sec on average. The distance between the estimated position and measured one was 3.09 mm on average with 0.043 mm as the standard deviation based these five trials. The orientation error was 2.44 degrees on average with standard deviation of 1.23 degrees. These errors are also consistent with approaching a target in three steps as illustrated in Figure 1, while more work is required to determine if higher resolution images from a closer endoscope can reduce approach errors while updating the 3D model in near real-time.

5.3 Discussion and conclusion

Recent work in accurately aligning the neurosurgical 3D space interoperatively after the brain shifts during skull opening comes from the laboratory of Dr. Michael Miga (Simpson et al., 2014). Using a phantom of brain surgery, the 3D positional target registration error was recorded for 5 repeated measures using laser range scanning and conoscopic holography. The results of 2.1 mm +/- 0.2 mm error for laser range scanning and 1.9 mm +/- 0.4 mm error for the conoscopic holography compare favorably with our preliminary results. Although both

3D measurement systems are laser-based and provide greater accuracy than our approach, their large size restricts their use to open surgery. In our case, the laser-based micro-camera is even smaller in size than most surgical tools, which allows introduction through less invasive keyholes and registration from within these semi-enclosed cavities. A future goal of surgical robotics is the reduction of size and advancement of keyhole applications (Marcus et al., 2013), which is the desired application of this new approach.

Recovered the position and orientation in near real-time using a single, ultrathin, flexible, and multimodal endoscope (attached to a surgical tool that blocks a portion of the visual field) in the 3D virtual model space of simulated surgical field. Future work is improving the speed and accuracy of the real-time SFE video registration to the 3D virtual model using phantoms that allow resection and the RAVEN surgical robot. This work builds toward automating residual tumor cleanup in robotically-assisted keyhole surgeries.

Chapter 6

BOUND CONSTRAINED BUNDLE ADJUSTMENT FOR RELIABLE 3D RECONSTRUCTION

Bundle adjustment (BA) is a common estimation algorithm that is widely used in machine vision as the last step in a feature-based 3D reconstruction algorithm. BA is essentially a non-convex non-linear least-square problem that can simultaneously solve the 3D coordinates of all the feature points describing the scene geometry, as well as the parameters of the camera. The conventional BA takes a parameter either as a fixed-value or as an unconstrained variable based on whether the parameter is known or not. In cases where the known parameters are inaccurate but constrained in a range, conventional BA produces in an incorrect 3D reconstruction by using these parameters as fixed values. Alternately, these inaccurate parameters can be treated as unknown variables, but this does not take advantage of the information in a constrained variable, and the resulting reconstruction can be erroneous since the BA optimization halts at a dramatically incorrect local minimum due to its non-convexity. In many practical applications that require 3D reconstruction, there are inaccurate parameters available before BA, such as a measurement with a range of uncertainty or a bounded estimate. Thus to better utilize these pre-known, constrained, but inaccurate parameters, a bound constrained bundle adjustment (BCBA) algorithm is proposed, developed and tested in this study. The experimental results demonstrate BCBA can achieve a more reliable, rapid, and accurate 3D reconstruction than conventional bundle adjustment.

6.1 Bundle adjustment and bound constraints

3D reconstruction is a process to retrieve the geometry and appearance of real objects or scene, which can be achieved by two main categories: active and passive methods [12]. A common way in active methods is to use artificial light source, such as projecting structured-light of a known pattern onto an object and then recovering the depth map from the

reflectance image [36, 47, 180]. Passive methods require multiple overlapped images of normally illuminated object without interfering any artificial light, such as stereo vision [123, 44] and structure-from-motion [1, 162, 27]. Typically in structure-from-motion, the multiple camera images from relative motion of camera and object are used for feature-based 3D reconstruction of the object or scene.

Bundle Adjustment (BA) is an optimization technique, simultaneously refining the camera parameters (focal length, center pixel, distortion, position or/and orientation), as well as the 3D coordinates of all the feature points describing the object [140]. The feature point is a specific structure in the image data, such as a corner. BA is often used as the last step in feature-based 3D reconstruction, following the prior steps of feature detection and feature matching. The cost function of BA is to minimize the norm of reprojection error of each feature point in every image. Reprojection error is the pixel distance between real position of each observed 2D feature and the calculated reprojection of its reconstructed 3D corresponding point, based on the current estimation of camera and object parameters [140]. This optimization can be expressed as a large-scale, non-convex, non-linear and real-valued least square problem [86]. The non-convexity of BA causes the existence of multiple minima so that it is difficult to find the global minimum.

Constraints are logical conditions that bound the estimation with allowable error, reflecting the real-world restriction of tolerancing. This bound constraint could be used as direct prior knowledge in the 3D reconstruction algorithm. In the most practical cases, the general constraints of the camera and object parameters are retrievable. Such as, the camera should be in the hollow body of the reconstructed model for the case of defect detection for the pipeline internal surface (location constraints); the shape of football is oval-like but planar (shape constraints); the size of a cell should be in $\sim \mu$m, instead of $\sim$cm (size constraints) and other constraints that are obvious in real world applicaiton. Even those 3D reconstruction without any case-based information, there are also some constraints that can be utilized which have been ignored: such as the focal length of a camera is positive; all the object points should be in front of camera instead of the backside; the field of view (FOV) angle is positive but less than 180 degree. Moreover for the cases that use a well-calibrated camera, the camera parameters are treated as known values without variation.

Actually these calibrated parameters are not exact but within a range of uncertainty [110]. Taking these parameters as fixed-value may cause inaccurate 3D reconstruction. Consider the uncertainty can be restrained by lower and upper values, called bound constraint, these constraints can be utilized as prior knowledge to improve the 3D reconstruction algorithm and generate more reliable and accurate model.

Previous work on constraining BA to achieve a more robust and realistic solution has been limited, and did not include the more general approach of having a bound constraint condition. Wong et al. constrained the distance between camera and object for all the frames as the same value, which was fixed to a single parameter to reduce the number of unknown variables in the BA algorithm [160]. With the similar concept, Albl et al. constrained the camera position on a circular trajectory for the cases with a revolving camera [2]. These two constrained BA approaches can both achieve more accurate 3D reconstructed with higher efficiency. However they require that either object or camera rotates around a fixed axis while the other one is stationary, which limits the generalizability of application for these two constrained BA algorithms [160, 2]. For the 3D reconstruction of planar structures such as floors and walls, planarity constraints were applied to produce an accurate model [179, 135]. But again these two applications are limited to the particular constraint of planarity. For the case of 3D reconstruction of internal surface of organ from medical images containing insufficient features, Soper et al. constrained the geometric model of bladder as a spherical shape in the initial step and then released this shape constraints for subsequent model refining [130]. This proper initialization reduces the influence of non-convexity to find the correct minimum and generates an accurate 3D reconstruction. It is essentially an approach to achieve the global minimum by choosing a good initial guess that is closer to the global minimum. However it does not guarantee that this initial guess is good enough to eliminate the deleterious effects of a local minima nearby.

All the constrained BA algorithms mentioned above are not generally applicable to beyond the strict requirements imposed by the specific applications. They achieved higher accuracy and efficiency by essentially reducing the number of variables for each of their specific cases.

Conventional BA algorithm either takes a parameter as known (fixed-value), or as an

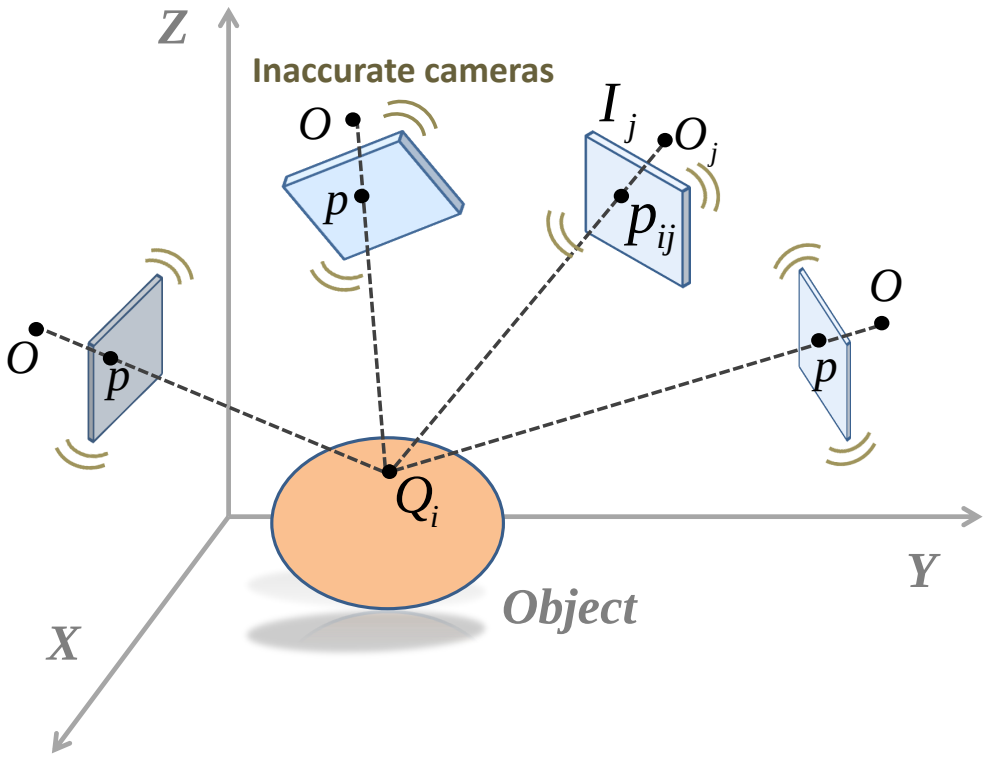


Figure 6.1: The feature-based 3D reconstruction with inaccurate camera parameters.

unknown variable (unconstrained). The former case may result in an incorrect 3D reconstruction with known but inaccurate parameters [43]. The latter case may cause erroneous estimation of camera and object parameters since the optimization falls down to a dramatically wrong local minimum. In this study, a brand new approach is proposed to solve this 3D reconstruction problem with bound constrained parameters. We call it Bound Constrained Bundle Adjustment (BCBA) algorithm. The biggest significance of this new algorithm is to achieve a reliable and accurate 3D reconstruction efficiently by taking advantage of known but inaccurate parameters, such as the intrinsic or extrinsic parameters of camera, or the general shape of object. Fig. 6.1 shows the schematic diagram of an applicable case of BCBA. A sequence of images were captured by inaccurate cameras around a static object. The term "inaccurate" represents those cameras that the intrinsic parameters, positions

or/and orientations are not exact known but in a range.

The motivation of solving this problem is to reconstruct an accurate 3D model of surgical field for autonomous image-guided surgery robotic system [43]. In our case, an endoscope (camera) was moved around by an ill-calibrated robot arm to capture images of surgical field. The camera position and orientation could be obtained approximately from robotic system, which was not accurate, but in a measurement range. Using the inaccurate information directly with conventional BA would result in an erroneous 3D model, which can be solved by utilizing the suitable bounds with BCBA.

In this study, each step of the conventional BA algorithm is presented, and the modifications made for the use of bound constrained parameters are listed and discussed. Following the proposed constrained algorithm, an experiment is presented by applying the the conventional BA algorithm that is compared to the proposed BCBA algorithm.

6.2 Bound constrained bundle adjustment algorithm (BCBA)

Our BCBA algorithm proposed in this letter is an extension of the traditional BA, which is essentially a non-linear real-valued least square problem. *Levenberg-Marquardt* (LM) algorithm is widely used as standard tool to solve BA due to its efficient damping strategy and ease implementation [77]. Instead of directly tracking nonlinear problems, LM algorithm iteratively solves a sequence of *linear* least square problems. From the optimization perspective, it can be viewed as a combination of *Gauss-Newton method* and *steepest descent method*: the closer the current iteration point approaches the local minimum, the more like the Gauss-Newton method the LM algorithm behaves. This adjusting strategy allows LM to achieve a better tradeoff between stability of the steepest descent method and faster convergence of the Gauss-Newton method. Our BCBA algorithm inherits this advantage from LM algorithm.

Given an object with a set of n 3D feature points $Q_{3xn} = \{Q_1, Q_2, ..., Q_n\}$ in world coordinate, a sequence of m images $\{I_1, I_2, ..., I_m\}$ were captured by a camera at different locations and orientations. Let R_j represent the rotation matrix and t_j denote the translation vector of the camera at I_j. Then the 3D position $P_{ij} = [X_{ij}, Y_{ij}, Z_{ij}]^T$of a feature point Q_i with respect to I_j camera coordinate is given as $P_{ij} = R_j Q_i + t_j$. The calculated pixel

location of its projection on I_j is $p_{ij} = [x_{ij}, y_{ij}]^T$, see Fig. 6.1, which can be calculated by

$$\omega \begin{bmatrix} p_{ij} \\ 1 \end{bmatrix} = K \begin{bmatrix} R_j & t_j \end{bmatrix} \begin{bmatrix} Q_{ij} \\ 1 \end{bmatrix}. \tag{6.1}$$

where $\omega = Z_{ij}$, and K is the camera calibration matrix constaining the parameters of focal lengths (f_x, f_y), center pixel (c_x, c_y) and first- and second- order radial distortion coefficients (k_1, k_2). Since we were using a single camera in this study, all the calibration matrices *K's* are the same.

Let the position of the observed feature be $\hat{p}_{ij} = [\hat{x}_{ij}, \hat{y}_{ij}]^T$, then the reprojection error $r = p - \hat{p}$ was calculated and rearranged as a column vector of size *M*. The projection error r is essentially a function of $N = 3n + 6m + 6$ variables: the 3D coordinate of each feature point $(3n)$, the camera position and orientation $(6m)$, and the identical camera intrinsic parameters (6). Denote all N variables as s. As mentioned before, information of the lower bound and upper bound on s is often available. We would like to exploit this additional bound information to better minimize the total reprojection error.

Given the bound constraints of parameters, the BCBA optimization problem can be written as

$$\begin{aligned} &\text{Objective} && \min \ f(s) = \frac{1}{2} r(s)^T r(s) \\ &\text{Subject to} && l \leq s \leq u; \qquad s \in \Re^N \end{aligned} \tag{6.2}$$

In most cases of feature-based 3D reconstruction, N is potentially equal to several thousands or larger. l and u are the same size of s, and represents the lower and upper bound vectors for the constraint of s. For those variables with no bound defined, we set $-\infty$ and $+\infty$ as its lower and upper bound, respectively.

6.2.1 Bundle adjustment

The optimization problem Eq. (6.2) without the constraints is the conventional BA problem. Its objective function $f(s)$ is non-linear, which can be approximated by a local linear

model[140], such as a quadratic Taylor series expansion of $f(s)$ below for a small step δ,

$$\begin{aligned} f(s+\delta) &\approx f(s) + g(s)^T\delta + \frac{1}{2}\delta^T H(s)\delta \\ \text{with} \qquad g(s) &\equiv \frac{df}{ds}(s) \qquad H(s) \equiv \frac{d^2 f}{ds^2}(s) \end{aligned} \tag{6.3}$$

where $g(s)$ and $H(s)$ is the gradient vector and Hessian matrix of $f(s)$, respectively. This approximation model of Eq. (6.3) is a simple quadratic with a unique global minimum that can be calculated explicitly using linear algebra.

The Gauss-Newton method is built to solve the step δ in the following linear system repeatedly

$$\bar{H}(s)\delta = -g(s) \tag{6.4}$$

where the $N \times N$ matrix $\bar{H}(s)$ is the Gauss-Newton approximation of the Hessian matrix $H(s)$ of $f(s)$. s is updated at each iteration as $s \leftarrow s+\delta$, if the cost $f(s+\delta) < f(s)$. Let $J(s)$ denote the $M \times N$ Jacobian matrix of $r(s)$, given by $J_{ij} = \partial r_i / \partial s_j$. Then the gradient $g(s)$ can be computed by $g(s) = J(s)^T r(s)$. The $\bar{H}(s)$ can be calculated by $\bar{H}(s) = J(s)^T J(s)$ by ignoring the second derivative terms. It is easy to see $\bar{H}(s)$ is symmetric matrix. For now, assume that $\bar{H}(s)$ is positive definite.

To solve step δ in Gauss-Newton method of Eq. (6.4), the inverse matrix of $\bar{H}(s)$ needs to be calculated. Since N could be very large, computing the inverse matrix of $\bar{H}(s)$ is often prohibitively expensive. Therefore, it is desired to reduce the size of the linear system first. Among all potential methods, *Schur complement* method is typical used in BA algorithm to split the original linear equation into two smaller linear systems by Gaussian elimination.

Denote the parameter column vector $s = [s_c, s_p]$ where s_c and s_p are the camera and point parameters, respectively. Similarly, the subscripts "c" and "p" are used to denote camera and point part for other notations: $J(s)$, $g(s)$ and δ. So $\bar{H}(s)$ can be expressed as a block matrix as follows:

$$\bar{H}(s) = \begin{bmatrix} \bar{H}_{cc} & \bar{H}_{cp} \\ \bar{H}_{pc} & \bar{H}_{pp} \end{bmatrix} = \begin{bmatrix} J_c^T J_c & J_c^T J_p \\ J_p^T J_c & J_p^T J_p \end{bmatrix}. \tag{6.5}$$

Since $\bar{H}(s)$ is positive definite, $\bar{H}_{cc}$ and $\bar{H}_{pp}$ are then positive definite. Eq. (6.4) also can be

rewritten as:

$$\begin{bmatrix} \bar{H}_{cc} & \bar{H}_{cp} \\ \bar{H}_{pc} & \bar{H}_{pp} \end{bmatrix} \begin{bmatrix} \delta_c \\ \delta_p \end{bmatrix} = - \begin{bmatrix} g_c \\ g_p \end{bmatrix}. \tag{6.6}$$

By separating the camera and point parameters using *Schur complement*, the solution $\delta = [\delta_c, \delta_p]$ of Eq. (6.4) can be calculated in the form of reduced camera system:

$$\begin{aligned} (\bar{H}_{cc} - \bar{H}_{cp}\bar{H}_{pp}^{-1}\bar{H}_{pc})\delta_c &= (\bar{H}_{cp}\bar{H}_{pp}^{-1})g_p - g_c \\ \delta_p &= \bar{H}_{pp}^{-1}(-g_p - \bar{H}_{pc}\delta_c) \end{aligned} \tag{6.7}$$

Note that the reason why δ_c is chosen to be computed first is that the number of camera parameters is usually smaller than the number of points. Once $\delta = [\delta_c, \delta_p]$ is solved, the parameter vector will be updated if it leads to a smaller cost $f(s)$.

The LM algorithm can be considered as the interpolation between Gauss-Newton method and steepest descent method. The steepest descent method is given by $D'\delta = -g(s)$, D' is a diagonal matrix of positive constant value, size of $N \times N$. Thus, LM can be written as

$$B\delta = -g(s), \qquad \text{with} \quad B = \bar{H}(s) + \lambda D \tag{6.8}$$

D is the diagonal matrix of $\bar{H}$. The damping factor λ is a non-negative value, adjusted at each iteration. Bigger λ brings the algorithm closer to steepest descent method, while the algorithm behaves more like a Gauss-Newton method as λ becomes smaller.

Similarly, to calculate the step at each iteration of LM algorithm, *Schur complement* is applied on Eq. (6.8) by splitting B as four sub-matrices corresponding to camera and point parameters. δ can be calculated by Eq. (6.7). Without considering the bound constraints in the optimization problem of Eq. (6.2), $s \leftarrow s + \delta$ will be accepted if the updated s vector leads to a smaller $f(s)$ value.

The non-negative damping factor λ is adjusted in each iteration. The initial value of λ in the first iteration is pre-defined as λ_0. If the current updated s leads cost reduction, a smaller λ can be used to bring the algorithm closer to Gauss-Newton method exhibiting faster convergence. Whereas if an iteration could not provide reduction in the residual, larger λ is chosen, giving a step closer to the gradient descent direction that is slow but guaranteed to converge [77]. Once the λ exceeds the pre-defined threshold λ_{max}, which is

usually very large, that means the local minimum is found. Conventional bundle adjustment is achieved by repeating the update of s and λ of Eq. (6.8) until reaching the local minimum.

6.2.2 Bound constrained bundle adjustment

BA can refine the camera and 3D structure parameters simultaneously in an efficient and stable way by utilizing LM algorithm. However, it considers parameters either as known data (fixed-value), or unknown variable (unconstrained). The conventional BA could not take advantage of the known but inaccurate information, which leads to bound constraints and is commonly available in practical applications. We now address this problem by modifying the BA algorithm mentioned above. Before the discussion, some preliminary tools are needed.

Firstly, we define the projection of the parameter vector s onto the feasible set with $[l, u]$ bound as a function:

$$proj(s) = min\{\, max\{\, l, s\}, \; u\}. \tag{6.9}$$

$max\{a, b\}$ is a vector whose ith entry as $max\{a_i, b_i\}$; and $min\{a, b\}$ is a vector whose ith entry as $min\{a_i, b_i\}$.

Secondly, an active set $A(s)$ for $s \in [l, u]$ is defined. $A(s)$ is a set that contains parameters at which either upper or lower bound is tight and the update direction drives the parameter outward away the bounds. Consider the following cases:

- Case 1: $s_i \in (l_i, u_i)$

  The computation of g_i is not constrained by the bounds, the same for $J(s_i)$. The calculation of δ can be treated as an unconstrained problem.

- Case 2: $s_i < l_i$ or $s_i > u_i$

  The s_i value will be projected onto the bounds of the feasible set by Eq. (6.9).

- Case 3: $s_i = l_i$

If $g_i > 0$, that means the updated s_i is supposed to be smaller than l_i, which will be constrained by the bounds; if $g_i \leq 0$, that means the updated s_i will be larger than l_i, falling in the feasible set. Thus it is not affected by the constraints.

- Case 4: $s_i = u_i$

  We have the opposite observation as case 3, the constraints are active if $g_i < 0$, but inactive if $g_i \geq 0$.

Finally the active set for the bound constrained problem $s \in [l, u]$ can be given below. The complementary set of $A(s)$ is called inactive set and denoted by $I(s)$.

$$A(s) = \left\{ \quad i \quad \middle| \begin{array}{ll} s_i = l_i & \text{and} \quad g_i > 0 \\ \text{or} & \\ s_i = u_i & \text{and} \quad g_i < 0 \end{array} \right\} \tag{6.10}$$

Gradient projection method [17] is then applied in this algorithm, which is a common method for solving constrained optimization problems. The projected gradient of $f(s)$ in the feasible set is a vector of size N, denoted by $\hat{g}(s)$:

$$\hat{g}_i = \begin{cases} g_i, & i \in I(s) \\ 0, & \text{otherwise} \end{cases} \tag{6.11}$$

The projected gradient is the same as the unconstrained gradient for those parameters that constraints are inactive. For the parameters that constraints are active, the projected gradient will be enforced to be zero to keep the parameters staying at the bounds.

One more component that we need to modify in Eq. (6.8) for the BCBA algorithm is the approximation of Hessian matrix $\bar{H}(s)$, which we define as *reduced Hessian* matrix. The reduced Hessian is designed as follows: since the parameters in the active set $A(s)$ are fixed then $\partial f / \partial s_i = 0$; and the projected gradient of active parameters are set to zero, we zero out all rows and columns of the Hessian matrix $\bar{H}$ that are corresponding to the active parameters, that is, to set $\bar{H}_{ij} = 0$ if either i or j is in $A(s)$. However, we will add the diagonal entries $\bar{H}_{ii},\ i \in A(s)$ back so that all diagonal entries of $\bar{H}$ are retained in the

reduced Hessian. It can be easily verified that this not change the gradient update but will later make it easier to control the positive definiteness of the reduced Hessian matrix. Finally, we define reduced Hessian matrix $\hat{H}$ to be

$$\hat{H}_{ij} = \begin{cases} \bar{H}_{ij} & \text{if } i \in I(s) \text{ and } j \in I(s) \\ \bar{H}_{ii} & \text{if } i \in A(s) \\ 0, & \text{otherwise} \end{cases} \tag{6.12}$$

Once we have the reduced Hessian matrix and projected gradient vector, Eq. (6.8) in LM algorithm is modified as below for the bound constraints.

$$\hat{B}\delta = -\hat{g}(s), \qquad \text{with} \quad \hat{B} = \hat{H}(s) + \lambda\hat{D} \tag{6.13}$$

$\hat{D}$ is the diagonal matrix of $\hat{H}(s)$, which is also equal to the diagonal matrix D. Similar to the process of Eq. (6.4), *Schur complement* is also used to calculate the step δ in Eq. (6.13). The update of $s \leftarrow proj(s + \delta)$ is accepted once it leads to a reduction in the cost function $f(s)$.

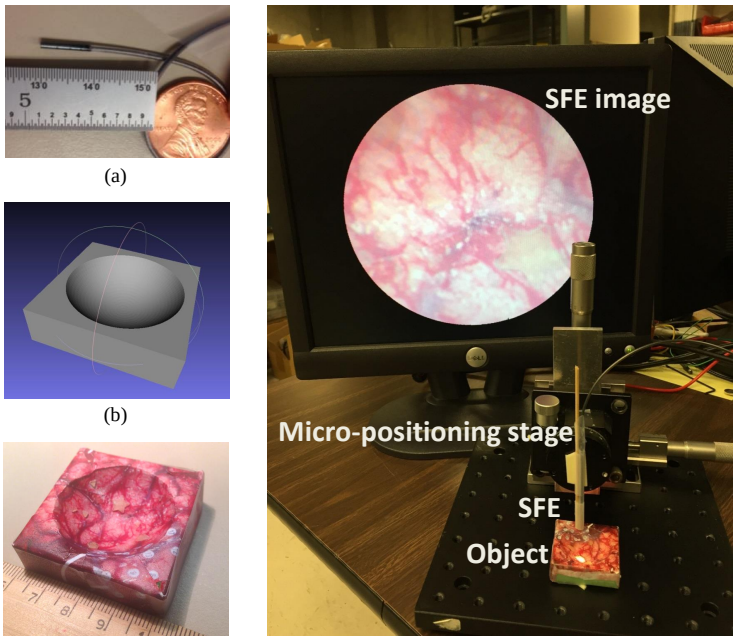
The same as BA algorithm, the non-negative damping factor λ in BCBA is also adjusted in each iteration. Smaller λ value is chosen if current iteration leads less cost residual; larger λ value can be used to guarantee the convergence to local minimum under bound constraints.

6.2.3 Convergence of BCBA

Because we have assumed that $\bar{H}(s) = J(s)^T J(s)$ is positive definite, it is easy to obtain that $\hat{H}(s)$ is also a positive-definite matrix based on the modification in Eq. (6.12). $\hat{D}$ is the diagonal matrix of $\hat{H}(s)$, it is also positive definite. Thus $\hat{B}$ is positive-definite in Eq. (6.13), regardless how the damping factor λ changes. Notice the right side of Eq. (6.13) is actually the steepest descent direction. Hence, our proposed algorithm can generate a sequence of parameter vectors (s) in each iteration that converges to a local minimum.

6.3 3D reconstruction test of BCBA

To demonstrate the feasibility of this constrained algorithm, an experiment was performed to compare the 3D reconstruction results of conventional BA and BCBA. In this study, a



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

Figure 6.2: The experiment setup. (a) the SFE of 1.6 mm outer diameter with camera frame rate at 30 Hz; (b) the CAD design of the object. It is a spherical dome with maximum radius of 17.5 mm and depth of 10 mm; (c) the 3D printed phantom with near-realistic surgical features; (d) the experiment setup on a micro-positioning stage, which can provide high accurate data of the camera positions with accuracy of 0.01 mm.

SFE [70] was used to capture a sequence of images above a surgery phantom (the object) of known geometry. The SFE is a tiny flexible endoscope of only 1.6 mm outer diameter, see Fig. 6.2(a), which is feasible to Minimal Invasive Surgery (MIS). It can achieve high resoltion, high speed full color image with wide FOV. To validate and compare two 3D reconstruction algorithms, an object with pre-known geometry was required. In this study, a 3D model was designed and 3D-printed out to mimic the surgical field [42], see Fig. 6.2(b,c). This reconstruction of surgical field aims to guide the autonomous surgical robot by obtaining the accurate 3D coordinate of residual tumor within the surgical field [41].

In this experiment, the SFE was attached on a micro-positioning stage, where the position of the camera was obtained. The orientation of the camera was set perpendicularly downwards and kept the same in the entire experiment. Fig. 6.2 shows the SFE, CAD

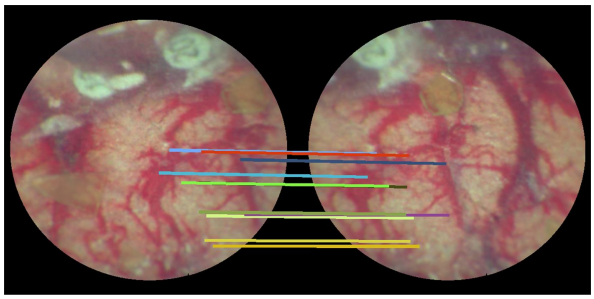


Figure 6.3: Matching features in a pair of frames with image overlap were lined up in color. One tenth of the matching features were randomly chosen for better visualization.

design, printed phantom and also the experiment setup. The distance between the SFE and the top surface of phantom is about 15 mm. Note, the BCBA algorithm is independent of the choice of camera and object. It can be applied as long as the the bound constraints exist.

In the experiment, 16 images (size of 608 x 608 pixels) were captured above the phantom using SFE. The position of SFE was read from micro-positioning stage with high accuracy, which can be considered as ground truth. To simulate known but inaccurate information, a set of noise of uniformed distribution on the interval [-0.25, 0.25] mm was added to X, Y, Z coordinate of the camera position, which can be treated as the pre-known bound constraints when using the BCBA alogrithm. Other constraints were also pre-known: a rough camera calibration obtained the FOV was in the range of [50, 56] degrees; the surface geometry of the object is about 10 mm in depth; and the distance of the object to camera is 15 mm. Thus, the depth of object points should be in the range of [15, 25] mm. However, for this experiment, we only use the constraints of camera position for BCBA. Other pre-known parameters were used to evaluate accuracy and reliability of the 3D reconstruction result.

All the codes were MATLAB implemented, and running on a workstation Dell Precision M4700 with 2.7 GHz Intel i7-3740QM CPUs, 20.0 GB memory in a 64-bit Window operating system.

In the same manner as conventional BA, features were detected in each image and matched with its correspondences in other images. In this study, scale-invariant feature

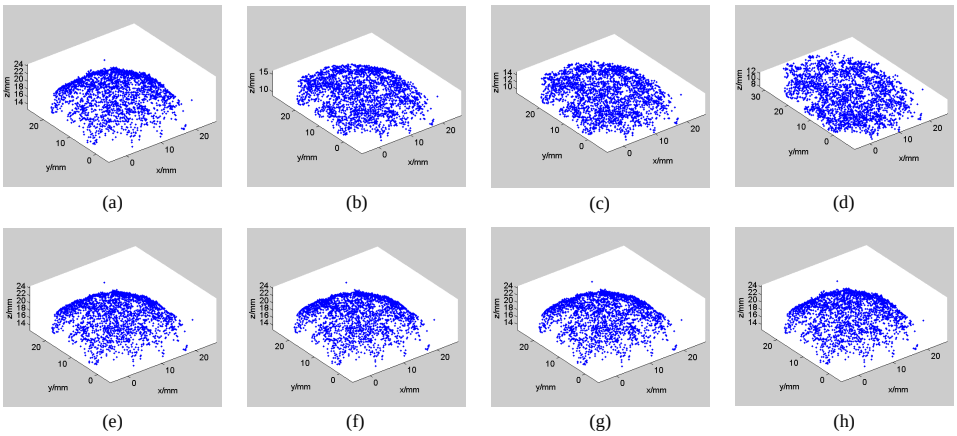


Figure 6.4: The procedures of BA and BCBA, respectively. (a-d) shows the reconstructed 3D point cloud by BA was shrinking to a flat one. (e-h) shows minor adjustment of the 3D points happened in the BCBA optimization process due to the bound constraints.

transform (SIFT) algorithm [78] was applied to find feature points from each SFE image. Fig. 6.3 shows the two endoscopic frames with image overlap, the matching features were lined up with color lines. This experiment generated 2,250 feature points and 5,689 point-to-image reprojection in total from 16 frames.

After adding the same amount of random uniformed noise ($\in [-0.25, 0.25]$ mm) to the accurate camera position, both BA and BCBA algorithms were performed. The process of BA showed the reconstructed 3D model shrinking from an initial spherical-like shape to a flat surface in the final optimization iteration, see Fig. 6.4(a-d). While the process of BCBA showed the reconstructed points bounded in a range and quickly stablized to a spherical shape, Fig. 6.4(e-h). We can notice that the reconstructed point clouds at the first iterations of BA and BCBA are the same, Fig. 6.4(a,e), since the initial guess of parameter vector was identical for both algorithms. Fig. 6.4(d,h) is the final reconstructed 3D points of BA and BCBA, respectively.

Fig. 6.5(a) show the BA resulted in a flat surface that is very different to our real object of a spherical concave surface. In addition, the reconstructed 3D points locate around depth of 10 mm, which is dramatically deviated from the experiment setup of 15 mm distance. Whereas the BCBA algorithm achieves a much better 3D reconstruction with most points

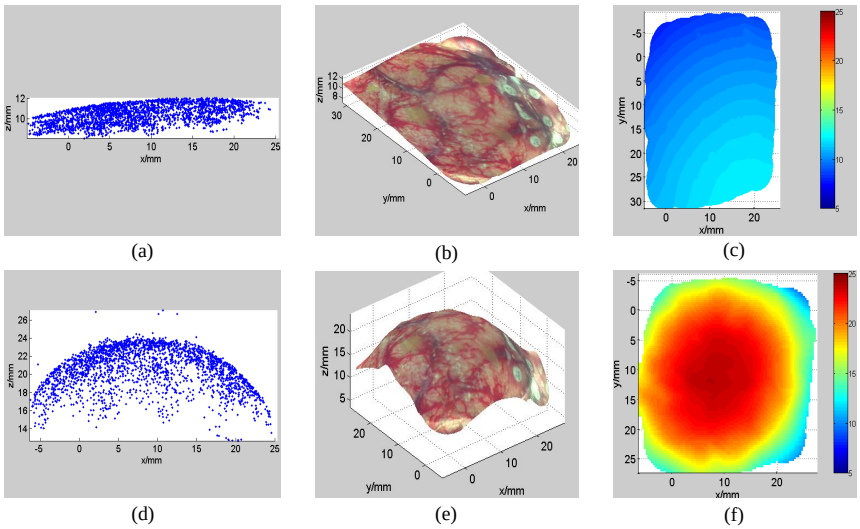


Figure 6.5: The comparison of 3D reconstruction results between BA and BCBA. (a-c) and (d-f) shows the reconstructed 3D point clouds, 3D surface and the depth map of BA and BCBA, respectively.

distributing as a spherical concave dome shape, see Fig. 6.5(d). Furthermore, 98.85% of the reconstructed object points are located in the range of [14, 24] mm, which is overall 1 mm shift away from the experimentally measured value. Once the 3D point cloud was obtained, a thin spline algorithm was applied to generate a smooth 3D surface by fitting these feature points [130]. Fig. 6.5(b,e) and (c,f) show the reconstructed surface and also the depth map of BA and BCBA results, respectively.

To quantify the reconstruction result, ICP algorithm [11] was employed to align and compare the reconstructed virtual models to the CAD designed model (ground truth, see Fig. 6.2(b)). Fig. 6.6 gives the comparison of ICP error analysis of BA and BCBA. The alignment of two point clouds were shown in Fig. 6.6(a,d), the blue point cloud represent the CAD designed model. Qualitatively, the BA algorithm produces a radically different 3D surface that does not match the CAD model; while the BCBA result matches the CAD model very well. The distance maps between the aligned models were shown in Fig. 6.6(b,e), and also the histgrams of the distance in Fig. 6.6(c,f). The quantitative error of BA's result was very large, which was diminished greatly by using bound constrained algorithm.

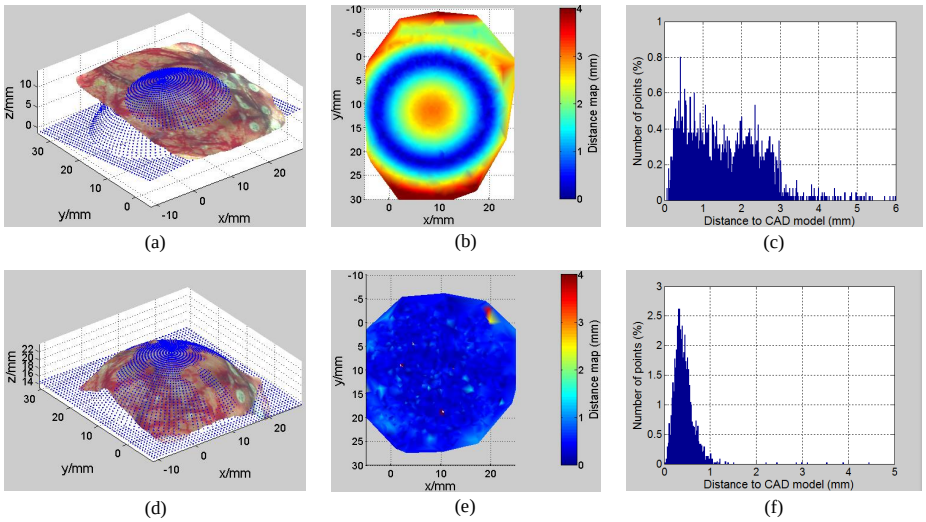


Figure 6.6: The comparison of ICP error analysis between BA and BCBA. (a, d) the 3D alignment of CAD point cloud and reconstructed surface for BA and BCBA, respectively; (b, e) the error map of BA and BCBA, respectively; (c, f) the histogram of ICP error distribution of BA and BCBA, respectively.

To further understand the significance of our proposed BCBA alogrithm, a table was generated with other comparison details of BA and BCBA. Table 6.1 shows the number of iteration, time spent, the minimum error achieved, the estimated parameters of camera (FOV), the estimated depth of 3D points Q_z and also the ICP error. With the bound constraints of camera position, the BCBA optimization can achieve to the local minimum faster than BA. Also the accurate 3D reconstruction result of BCBA has much smaller ICP error. Moveover, only the pre-known constraints of camera position were used in this study, the estimated FOV and Q_z can be compared with the pre-known values to demonstrate the reliabilty of our constrained algorithm. For example, the estimated FOV using BA was 165.39°, which is far away from the calibration value $\in [50°,\ 56°]$. In stark contrast, the BCBA algorithm estimates a FOV of 50.14° which is within our experimentally determined range (Table 6.1).

Table 6.1 shows BA can result in smaller cost value of the objective function Eq. (6.2). The reason is that the corresponding local minimum locates out of the bounds. BA can

Table 6.1: The experiment result comparison of BA and BCBA algorithm.

	BA	BCBA
Bounds used	(no constraints)	camera position
No. of iteration	125	10
Computation time	56.96 sec	8.85 sec
ICP error	1.8482 mm	0.5031 mm
Final Cost value	1.948	2.293
Estimated FOV	165.39°	50.14°
Estimated Q_z	$\in$ [8.0 mm, 12.1 mm]	$\in$ [12.6 mm, 27.0 mm]

achieve the local minimum while BCBA was restrained by the bound constraints, which results in larger cost value. Note that all the value in Table 6.2 might change based on different random noise added to the accurate camera position data.

In order to validate the reliability of BCBA algorithm, the same comparison shown above was performed for an additional ten times with different sets of uniformed noise ($\in [-0.25,\ 0.25]$ mm). Since the accuracy and efficiency are the most important criterias for the potential applications of this proposed algorithm, we listed the comparison of ICP error between BA and BCBA in Table 6.2, as well as the computation time. The obtained ICP error values of BA were in the range of $[1.8776,\ 1.9072]$ mm with an average of 1.8903 mm and a standard deviation of 0.011 mm; while for BCBA results, ICP error values were within $[0.4521,\ 0.5343]$ mm with an average of 0.4955 mm and a standard deviation of 0.0306 mm. These data show that BCBA can generate more reliable and accurate 3D models than BA with very small deviation for the ten data sets simulating the real-world measurements. While the computation time of either BA or BCBA varied a lot for these ten cases. BA spent $[67.34,\ 49.27]$ sec, with average of 60.61 sec and deviation of 6.88 sec; whereas BCBA spent much less time for the computation, which was in the range of $[3.56,\ 30.61]$ sec, with average of 14.40 sec and deviatin of 9.56 sec.

From the comparison results of these ten tests in Table 6.2, we see the ICP error was

Table 6.2: The comparisons of ICP error and computation time of the BA and BCBA algorithms with different noise.

	ICP error		Computation time	
	BA (mm)	BCBA (mm)	BA (sec)	BCBA (sec)
case 1	1.9072	0.5070	63.86	6.52
case 2	1.8787	0.4915	50.65	21.95
case 3	1.8998	0.5283	63.46	3.63
case 4	1.8830	0.4663	67.34	19.56
case 5	1.8890	0.4725	65.35	7.80
case 6	1.8776	0.5343	49.27	10.34
case 7	1.8931	0.5148	63.91	14.54
case 8	1.8801	0.4623	64.58	30.61
case 9	1.8888	0.4521	52.57	3.56
case 10	1.9060	0.5264	65.15	25.44
avg.	1.8903	0.4955	60.61	14.40
std.	0.0110	0.0306	6.88	9.56

very stable with a small deviation. Overall, the accuracy of BCBA was around 3.8× higher than the conventional BA algorithm. The time spend for both BA and BCBA was less stable, being highly dependent on the noise added. On average, the BCBA'efficiency was roughly 4× higher than BA's for these ten tests.

6.4 Discussion and Conclusion

This study proposed a bound constrained BA algorithm to take advantage of the known but inaccurate information of the parameters of camera and object points. The theoretical development wass provided with emphasis on the convergence property of the proposed BCBA algorithm. Moreover, the experimental result demonstrated its feasibility and reliability for a practical problem. Comparing with conventional BA algorithm that considers a parameter either as known fixed-value or an unconstrained variable, BCBA achieves less optimization iterations, less computation time, smaller ICP error and more reliable parameter estimation. For the specific tests in this study, the accuracy and efficiency of BCBA was 3.8× and about 4× higher than the conventional BA algorithm, respectively. Besides this, BCBA also achieved a good estimation of other parameters that have no constraints, such as the estimation of camera FOV in Table 6.1.

As the extension work of our previous research [43, 42], these advantages of BCBA can be utilized for intraoperatively reconstructing accurate 3D virtual models of the surgical field to extract the tumor coordinates, which requires highly efficient processing. These advantages also can be utilized to provide reliable estimation of camera poses for 3D image-guided surgery [41] and MIS. Not limited in biomedical research, the BCBA algorithm also may be utilized in other fields, such as a common application of the 3D reconstruction of Google street view [62]. The position of street-view-car is provided by GPS; it is inaccurate but the error range is pre-known. Such bound constraints can be ultilized to improve the accuracy and efficiency of the 3D reconstruction of the street scene, similar to the experiment performed in this study with bound constrained camera positions.

Previous approaches to constraining the BA algorithm cannot be generalized to the examples of cameras involved in robotic MIS and 3D street-view reconstruction. Several constrained BA algorithms could improve the reconstruction accuracy and efficiency [160,

2, 179, 135], but their applications were limited in only specific cases, such as planery object reconstruction or self-rotating object or camera. Soper et al.'s method achieved reliable 3D reconstruciton, but limited to the spherical-like internal surface for a specific medical application [130]. In contrast to these constained algorithms, our proposed BCBA is applicable to these specific cases as well as more general cases as long as the bound constraints of parameters are known.

However, notice that the constrained algorithms [160, 2, 179, 135, 130] may work better than BCBA for their particular cases since they used constraints among different parameters, such as the X, Y, Z parameters of object are subject to specific mathematics equations of plane, circle or sphere. But they assumed the objects are perfect without tolerance, which cannot be true for real world applications. To improve the 3D reconstruction for such cases, the furture work is to improve BCBA with geometric constraints. For example, for the 3D reconstruction of bladder phantoms [130], we can replace the sphericity constraint to BA with a spherical-like shape constraint that contains lower/upper bounds once the mimimum and maximum radius of the bladder is known based on human anatomical data.

The same as conventional bundle adjustment, BCBA is also subject to halting at a local minimum. Even within the bounds, different initial guesses of parameter vector s produce different optimization results, as shown in Table 6.2. Nonetheless in BCBA, the pre-known bounds can prevent the reconstucted model running out to an unreasonable shape. Future work will be finding the global minimum in the feasible set, from which the computation could be dramatically reduced, compared with the unconstrained cases.

Chapter 7

AXIAL-STEREO 3D OPTICAL METROLOGY OF INTERNAL SURFACE

As the rapid progress in the development of optoelectronic components and computational power, 3D optical metrology becomes more and more popular in manufacturing and quality control due to its flexibility and high speed. However, most of the optical metrology methods are limited to external surface. This chapter proposed a new approach to measure tiny internal 3D surfaces with a small camera and axial-stereo vision algorithm. A dense, accurate point cloud of internally machined threads was generated to compare with its corresponding X-ray 3D data as ground truth, and the quantification was analyzed by ICP algorithm.

7.1 3D optical metrology

Due to the rapid development of photonic sources, sensors, optomechatronics, and computer processing, the application of 3D optical metrology has been growing for inspection of machined external surfaces in manufacturing [117]. With automation, 3D optical metrology is increasingly replacing the traditionally used metrology techniques in industry, which include mechanical gauging techniques such as the Coordinate Measurement Machine (CMM) [55], due to its characteristics of non-contact, non-destruction, high efficiency and high accuracy [175].The generated 3D point cloud can be used in quality control, inspection, virtual design, reverse engineering, simulation and planning. Optical metrology techniques can be generally categorized as active and passive [32]. The active approach is to obtain the 3D coordinate data by interfering with the object using light, such as Time-of-Flight (TOF) measurement [3, 116], optical holography [23], optical to X-ray computed tomography (CT) scanning [95], structured-light reconstruction [180, 47, 36, 48]. In contrast, passive 3D systems require only a sensor, such as a camera, to collect digital images of the light scattering from the surface of an object. Typical techniques of passive vision contains stereo, structure

from motion [162] and shape from shading [177].

Installed 3D optical metrology systems are not adequate to provide high-quality metrology of small internal surface, except X-ray CT that allows for the measurement of internal structures with sufficient atomic number difference. For larger structures, the optical metrology equipment can fit inside and is becoming more and more important in industry. An important application is crack inspection, quality control, installation, and even revamping of an existing pipeline in the oil/gas industry [71]. In this case, the pipe can fit the complicated components of the active optical metrology instruments inside. For smaller structures, there is a need to rapidly and carefully inspect machined holes in engine manufacturing, but lack optical 3D inspection technologies that can physically fit inside the structures. An emerging application is highly accurate and reproducible 3D inspection of tiny threaded holes [56]. The main function of the threads is to hold different machined parts together, such as joining the engine block to the head in the automotive industry. Thus, the quality of the threads and their measured inner profiles are crucial for the engine safety, performance, efficiency and longevity [45, 56]. There are various sizes of internally threaded holes in the automotive engine block, most of them are smaller than 1 cm diameter.

State-of-the-art techniques of optically measuring the small inner profile of internal structures have been developed more for inspection of internal human organs than for internal machined or fabricated structures. These endoscopic techniques range from conventional stereo [21], photometric stereo [100, 24], structured light scope [114, 22], time of flight [101], to 3D shape from motion [130]. In all cases, either the scopes are larger than 6-mm in diameter, or the 3D surface reconstruction is incomplete or at low resolution. Smaller optical scanning scopes based on low-coherence interferometry and called optical coherence tomography (OCT) were also developed for medical imaging [59]. To form 3D surface profiles in tubes as small as 2-10 mm, a focused laser beam is scanned radially using a spinning mirror or prism [68]. Although OCT was developed to image below the surface of translucent biological tissue, the technique can be applied to measuring opaque surface structures with lateral resolutions in the 10-30 micron range, but with limited depth of focus [128]. With higher-speed 3D imaging using Fourier domain OCT, new inline inspection applications are being developed for industry [168], although no 3D surface reconstruction of threaded holes

in shiny metal has been reported for these OCT techniques. Strong specular reflection of the shiny metal surface does lead to large artifacts in 3D reconstructions [147]. Moreover, OCT systems involve a complex (expensive) design of rapid probe rotation has been a barrier of entry into the high demands (e.g. robustness) for inline manufacturing systems.

Specifically for 3D metrology of small pipes, Yoshizawa and Wakayama [174] developed a 5-mm diameter laser-scanning probe that strikes a cone mirror to generate a disk-like beam on the inner wall of the pipe. The reflected light is then captured by camera and processed with triangulation principle to generate the 3D profile. However, this device is still large in relation to the internally threaded blind hole in the engine block that could not accommodate the working range of 15 mm to 40 mm from the probe tip [174].

7.2 Axial-stereo vision

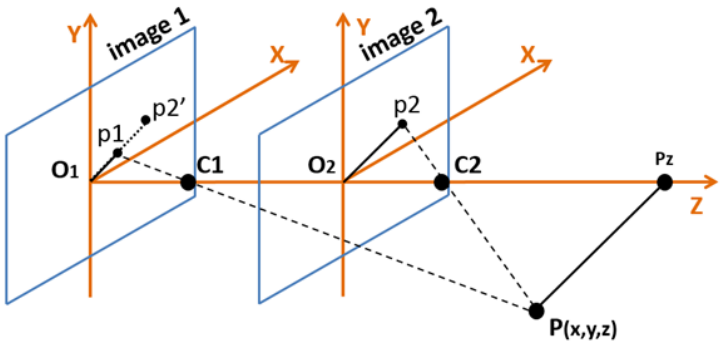
A 3D reconstruction algorithm with a SFE video of internal surface has been developed in our previous study in medical field based on SIFT features (scale-invariant feature transform) [43, 42]. It does not meet the requirement of high-quality dense reconstruction for industry application since there are much fewer features in the internally machined metal parts. In this study, a stereo vision technique was applied to generate the dense 3D model. Instead of using lateral-stereo (conventional stereopsis), axial-stereo vision [113] was established with two SFE images by moving SFE along with the axis of thread hole.

7.2.1 Mathematics model

Stereo vision is an approach to extract 3D information of scene from two digital images that captured by two cameras with known spatial positions and orientations. Conventional stereo vision system contains two side-by-side cameras, in a manner similar to human binocular vision. To satisfy the goals of reducing space, cost, and complexity of examining small restricted spaces such as a threaded hole, a single SFE (monocular vision) was utilized to establish an axial-stereo vision by moving SFE along the axis of hole to generate the 3D model, which had been barely studied [113, 37].

Figure 7.1 shows the schematics of axial-stereo vision. Two cameras are set up as their optical axis coincided, shown as Z direction in the schematics. And also the X and Y axis

of these two image plane should be kept paralleled. For a point P(x,y,z) in 3D space, its projections in two images locate at p1 and p2, respectively. p2' is the orthogonal projection of p2 on the image 1. $(O_1 - C_1)$ and $(O_2 - C_2)$ is the focal length of SFE f_{SFE}, which was known from calibration. By knowing the distance of $(O_1 - C_1)$, the depth information of P can be calculated in the form of disparity $(P2' - P1)$, which was obtained by image processing.



X, Y, Z: the coordinate system in 3D space;
O1 and O2: the center of image planes of cameras;
C1 and C2: the pinholes of cameras;
P(x,y,z): a point in 3D space;
p1 and p2: the projection of P on image 1 and image 2;

Figure 7.1: The scheme of axial-stereo vision.

From the triangular relationship shown in Fig. 7.1, some equations were satisfied as:

$$\frac{(p2 - O_2)}{(P - P_z)} = \frac{(C_2 - O_2)}{(C_2 - P_z)}. \tag{7.1}$$

$$\frac{(p1 - O_1)}{(P - P_z)} = \frac{(C_1 - O_1)}{(C_1 - P_z)} = \frac{(C_1 - O_1)}{(C_1 - P_z) + (C_1 - C_2)}. \tag{7.2}$$

From equations (1) and (2), the equation of depth of P(z)=$(C_2 - P_z)$ was obtained, which only contain unknown variable of disparity:

$$P_z = (C_2 - P_z) = \frac{(p1 - O_1)}{(p2' - p1)} \times (O_1 - O_2). \tag{7.3}$$

The X and Y coordinate of P, represented as P(x), P(y), can be calculated with known focal

length f_{SFE}:

$$P_{x,y} = \frac{p2(x,y)}{(C_2 - O_2)} \times P(z) = \frac{p2(x,y)}{f_{SFE}} \times P(z). \tag{7.4}$$

7.2.2 Disparity Computation

To compute the disparity of the projection of each point in the 3D scene, we have to find out the corresponding points in the two axial images. In previous studies [42, 43], SIFT algorithm was used to detect and match the corresponding features. But in the metrology of internally machined metal hole, there are very limited features available for dense matching. In this study, we employed block matching algorithm to find the corresponding [164]. For every pixel in image 2 (see Fig. 7.1), we extracted an n-by-n-pixel block round it and search around the same location in image 1 over both X and Y directions. The sum of absolute difference (SAD) was calculated in each block comparison. The block with the minimum SAD value is the matching one. Once we have the corresponding points, the 3D information of the scene can be reconstructed by the equations (7.3) and (7.4).

7.3 Dense reconstruction test of axial-stereo

To demonstrate the feasibility of our proposed axial-stereo vision system, the 1.2 mm OD SFE was attached to a micro-positioning stage, representing a manufacture instrument that can provide accurate position and orientation of SFE, shown in Fig. 7.2(a). The micro-positioning stage we used in this study was ULTRAlign^{TM} precision integrated crossed-roller bearing linear stage, model 462-xyz-m. The position and orientation of SFE was adjusted so that the optical axis of SFE is perpendicular downwards and coincides with the axis of thread hole of 8 mm major diameter and 7 mm minor diameter.

The SFE was moving downwards with 0.25mm interval, two images were captured with size of 608 x 608 pixels, shown in Fig. 7.3. Only the front flank of the threads was captured, with almost identical thread pattern and limited features in the images. The SFE position of the right image (Fig. 7.3(b)) was 0.25mm deeper than the one of the left image (Fig. 7.3(a)).

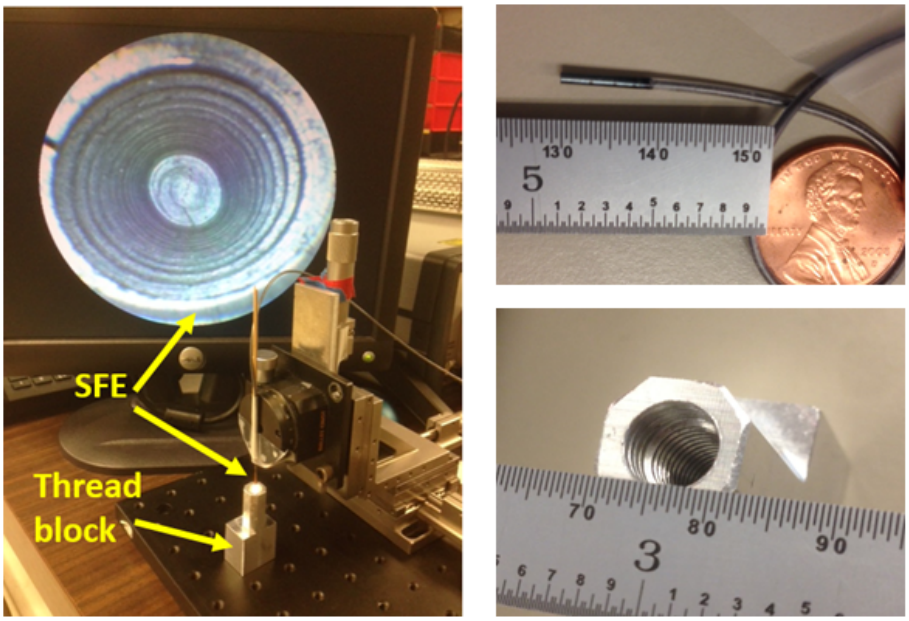


Figure 7.2: The experiment setup of axial-stereo vision system for internal 3D metrology of thread block. (a) The SFE was attached to a micro-positioning stage, aligned with the axis of thread hole. (b) the photographs of SFE (up) and thread block (bottom).

7.4 Result and Discussion

In this study, a block size of 7x7 pixels was chosen to search the corresponding points from Fig. 7.3(b) to Fig. 7.3(a). The returned disparities of the corresponding points were all integer-valued and noisy, shown in Fig. 7.4(a). A sub-pixel correction was incorporated to eliminate the integer-valued discontinuity issue [136], dynamic programming [146] was applied to eliminate the noise and imaging pyramid [137] was utilized to guide the block matching, making the algorithm running fast, shown in Fig. 7.4(b). By knowing the disparity of each pixel in the SFE images, the 3D scene would be reconstructed by equations (7.3) and (7.4). In our study, a total of 164,579 3D points were generated (Fig. 7.4(c)), and a zoom-in view is shown in Fig. 7.4(d). The triangular mesh could be generated by finding the neighbor points along the X and Y directions.

All the codes are Matlab implemented, and running on a workstation Dell Precision

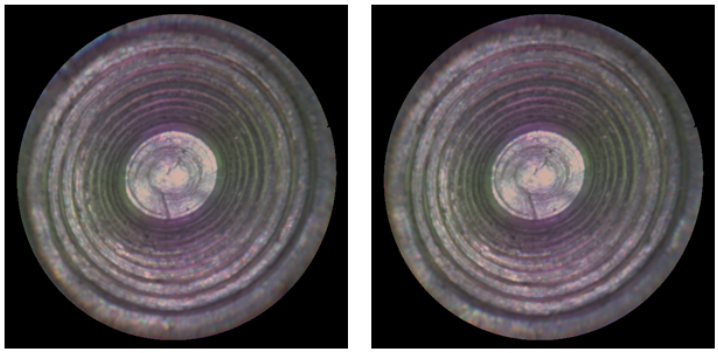


Figure 7.3: The two SFE video frames were captured with different axial depth (0.25mm) along the axis of a thread hole of 8mm as major diameter and 7mm as minor diameter.

M4700 with 2.7 GHz Intel i7-3740QM CPUs, 20.0 GB memory. It took 250 sec to generate the dense 3D model.

The 3D reconstruction result shows very clear threads with texture information since it is image-based measurement. This is a unique advantage comparing with X-ray 3D metrology, which only generates geometry information. And also the obtained point cloud can provide more quantitative quality control and monitoring of tool wear comparing with current common method of go/no-go contact gauges, which causes heavier workload, lower efficiency and higher levels of uncertainty depending on the operators skill [56].

The 3D optical metrology result contains a very noisy rim (Fig. 7.4(c)) since the outer scans of original SFE image were dark, which introduced much noise. Also this axial-stereo vision system is sensitive to specular reflection, since the block matching is based on pixel grayscale values. Moreover, only the front flank of threads were captured and reconstructed in this study, an internal mirror could be placed in front of SFE to see the backside of threads in the future study.

To quantify the reconstruction result, ICP [11] was employed to align and compare the reconstructed dense point cloud to the X-ray point cloud (ground truth model) of the same thread block. The obtained ICP error value was 0.4333mm, which was relatively large. This error may come from the inaccurate experiment setup and SFE calibration, distortion of

SFE images, the specular reflection of metal surface, and also the round error of disparity computation.

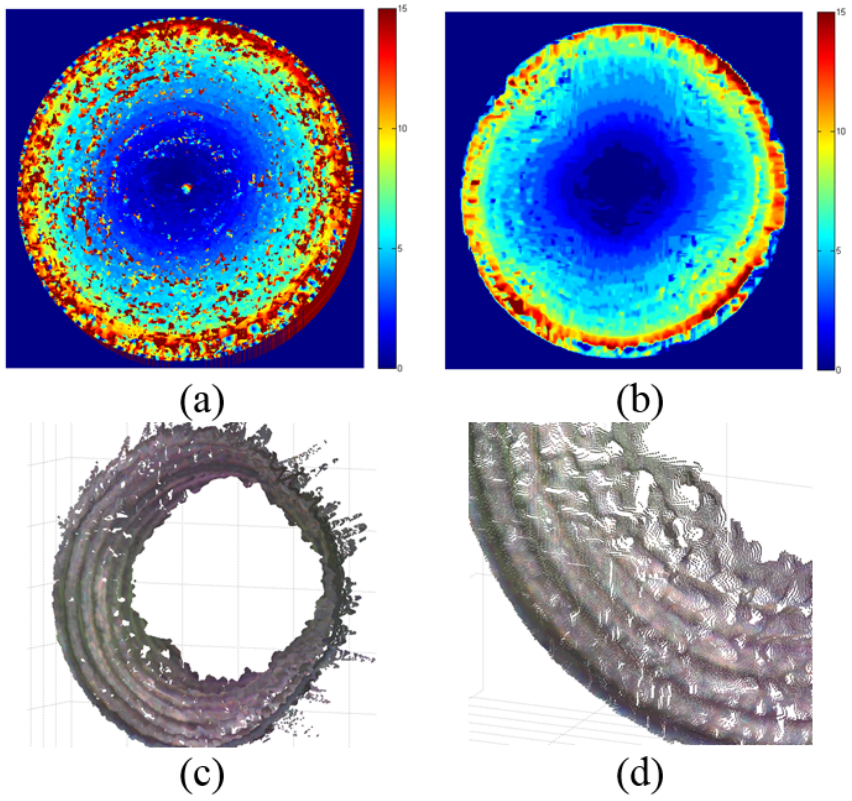


Figure 7.4: The 3D reconstruction of two SFE images that captured along the axis of an automotive thread hole. (a) The disparity map of basic block matching. (b) The disparity map of block matching with imaging pyramid, sub-pixel correction and dynamic programming. (c) The 3D dense reconstruction of internal threads. (d) The zoom-in view of the 3D reconstruction showing four clear threads.

By utilizing the axial-stereo vision algorithm, the SFE motion along the axis of a small internal thread hole (7mm as minor diameter) can generate a high-resolution 3D model, which demonstrated the feasibility of our proposed axial-stereo metrology method.

Chapter 8

3D IMAGING OF INNER SURFACE USING FEATURE-BASED 3D PANORAMIC REGISTRATION

In our previous study [45], an approach was proposed to measure the internal 3D surfaces with axial-stereo vision algorithm and a SFE of 1.2 mm outer diameter [70]. A dense point cloud was generated for a few threads from only two forward-view images along the hole axis. However, the inaccuracy was relatively large compared with the X-ray CT data used as the gold standard. In addition, the 80-degree field of view (FOV) of SFE could not be easily expanded to capture the thread profile from peak to valley in order to measure the inner (minor) and outer (major) diameters. In this chapter, we present a new 3D optical metrology approach based on machine vision from a small caliper endoscope to measure small complex internal surface profiles. This approach is formed by utilizing multiview stereo vision and a new proposed feature-based 3D panoramic registration algorithm. The 3D model of the entire internal surface (360°) is generated by registering a series of 3D point clouds, which are reconstructed separately from side-view images of different perspectives with a multiview stereo algorithm, based on the robust two-dimensional (2D) features and their 3D correspondences.

8.1 Data collection and categorization

To capture the high topographic relief of internal surface profile, such as the peaks and valleys of the internal threads, a side-view camera is used in our approach. The side-view camera provides direct vision of the inner surface of the hole, see Fig. 8.1, which ensures good and consistent image quality. However, the drawback is that we have to either rotate the camera or the object to guarantee that all the profiles of the internal surface are captured.

The procedure of scanning the entire hole containing complex internal surface can be divided into four steps, see Fig 8.1(a): step 1) placing the camera (center) along the axis of the threaded hole, facing to the side wall; step 2) with the same camera orientation,

taking a series of images as the camera moves inwards from the opening to the bottom with constant step distance. We call such sequential images as a "quadrant"; step 3) repositioning the camera at the hole opening and rotating it along the hole axis at a proper rotation degree to ensure image overlapped with the previously captured quadrant; step 4) repeating the steps 2) and 3) until finishing the entire scanning (360°) of the hole. The constant step distance and rotation angle are realized by micro-positioning and rotation stages in the experiment. In this study, the neighbor images that are collected with pure camera linear motion called an "axial neighbor", and the ones that are captured with pure camera rotation are called an "angular neighbor". During inline manufacturing, the angular and axial information (camera position and orientation) can be retrieved from industrial robotic positioning systems. By taking advantage of the known camera pose information, the 3D reconstruction of the threads is more accurate and efficient since it only needs to triangulate the 3D point cloud representing the internal surface. There are two ways to achieve the relative rotational motion between the camera and the straight hole to complete 360° scanning. Either rotate the camera within the fixed threaded hole, or to rotate the hole about the fixed camera. The choice depends on the particular objects to be measured, work space limitation and implementation convenience.

Due to the use of micro-positioning and rotation stages, the camera position and orientation data can be collected accurately and slowly. By taking advantage of the known camera pose information, the 3D reconstruction of the threads is accurate and efficient since it only needs to triangulate the 3D point cloud representing the internal surface, called multiview stereo [134]. However, processing all the images at once with state-of-art software VisualSfM [162], OpenMVG or even the customized MATLAB code [42] does not result in a satisfying 3D reconstruction. The reason is that these quadrants are captured with pure camera rotation (the baseline is zero) if the camera center is aligned with the rotation axis, Fig. 8.1(a). For this scenario, the complete 3D model can be reconstructed by stacking all the quadrants together with known rotation angles. However, we have to consider the real cases of axes misalignment in the practical applications, see Fig. 8.1(b).

There are three axes in the procedure: hole axis, camera center axis and rotation axis. The misalignment of hole axis with the other two does not affect the reconstruction result.

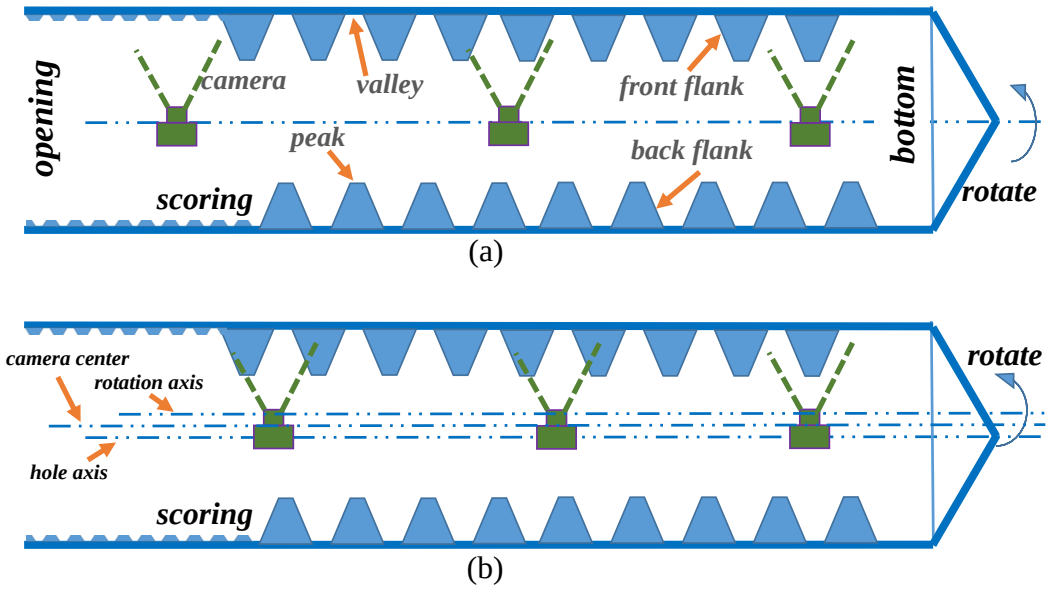
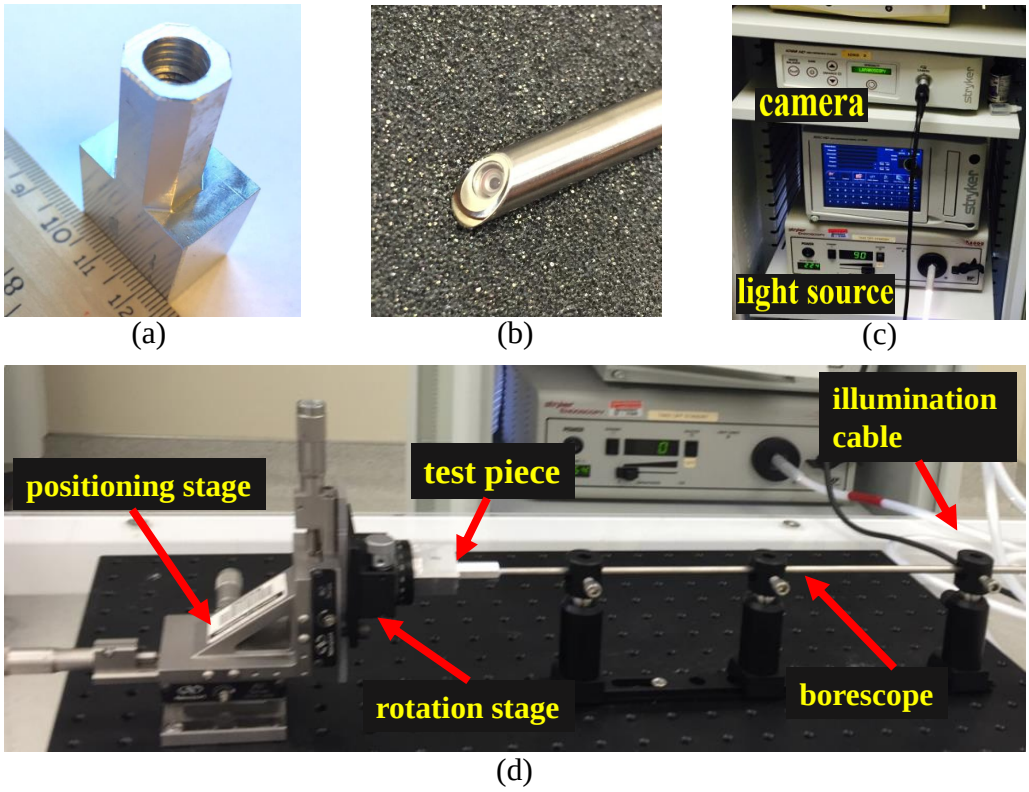


Figure 8.1: Diagram of a full axial scan of an internally threaded blind hole with 90° side-view camera.

However, the misalignment of camera center line and rotation axis could generate geometrical discontinuity once stacking quadrants together directly. In the practical cases, such misalignment is small and also unknown. Both scenarios of well-alignment and misalignment would generate significant error when triangulating the depth [35], resulting gross reconstruction error. To avoid this, we propose to use axial sequence images to reconstruct the 3D point cloud of each quadrant with a multiview stereo algorithm, and the angular neighboring frames for the 3D registration of neighbor quadrant point clouds. This is one significance of this proposed feature-based 3D registration algorithm that it can solve high depth uncertainty problem caused by little camera baseline in vision-based 3D reconstruction.

To demonstrate the feasibility of our proposed new approach to the 3D optical metrology of a machined internal structure with small internal surface, an experiment was performed to generate the 3D point cloud from within a shiny metal object and compare with current metrology techniques. In this study, an internally threaded blind hole with <7 mm minor

diameter (M8) was chosen as the test piece (object), Fig. 8.2(a). A relatively large rigid scope of 5.5 mm OD (Stryker™ scope model #502–503–045) was chosen for the ease of a stable set-up with high-resolution side-viewing capability, Fig. 8.2(b)–8.2(d). Due to the limitation of this commercial scope and external camera system, it only can achieve 45° side-view angle with 60° FOV, 1280 x 1024 pixels resolution, and 40 cm working length. Although smaller rigid and flexible scopes with forward viewing were available, there was insufficient time to create fixtures using a mirror at the distal tip that fit within the small size of this M8 hole while preserving FOV of 60°.



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

Figure 8.2: The experiment setup of 3D optical metrology of M8 threaded blind hole with commercial borescope. (a) The test piece with recessed coarse M8 threads machined in shiny aluminum metal; (b) a commercial borescope of 45° side-view angle and with resolution of 1280 x 1024 pixels; (c) the camera and illumination source of the commercial borescope; and (d) the experiment setup mounted on an optical breadboard with components of micro-positioning linear and rotation stages, test piece, borescope and customized scope holders.

To ensure stability, precision and accuracy, an optical rail was fixed on an optical breadboard, on which three identical kits of optomechanical components (rail carrier, post and holder) were set up to hold the rigid scope horizontally, see Fig. 8.2(d). Instead of rotating and translating the scope that connected to a bulky illumination system, it remained stationary during the entire experimental procedure. To achieve relative linear translation and rotation for the full axial and 360°scanning, the test piece was attached on a high-quality rotation platform, which was mounted on a micro-positioning stage (ULTRAlign™ precision integrated crossed-roller bearing linear stage, model 462-xyz-m), see Fig. 8.2(d). The reason for using micro-positioning linear and rotation stages is to get the accurate relative position and angle data between the test piece and scope, which can be retrieved from modern industrial robotic systems in future industrial applications.

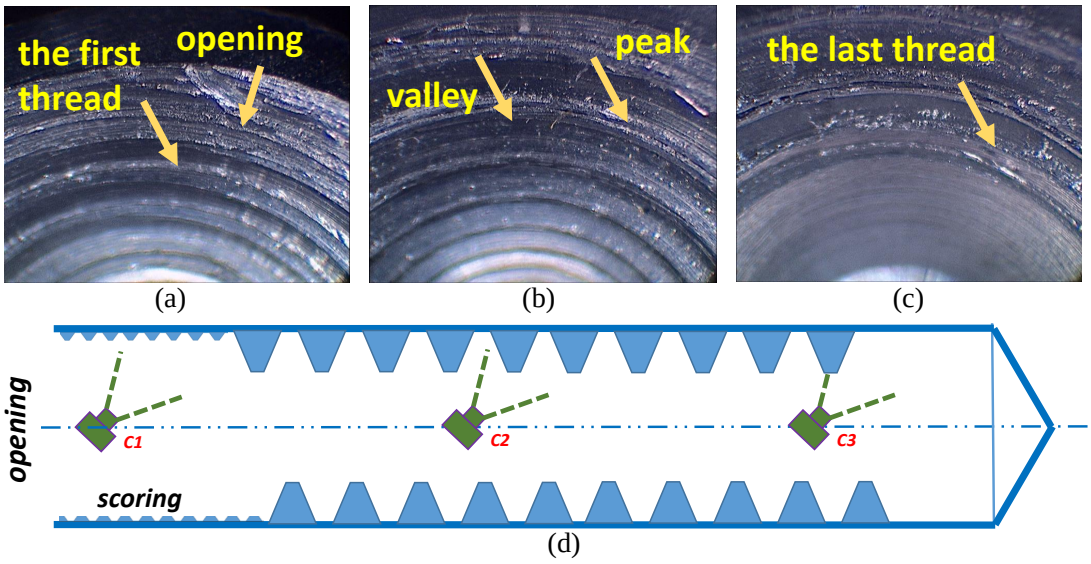


Figure 8.3: Three endoscopic image examples of the internally threaded blind M8 hole from a rigid medical scope with a 45° side-viewing angle. (a) The first frame in a quadrant with opening of the hole and the first thread labeled; (b) the thread peak and valley, shown as bright belt and dark area, respectively; (c) the last frame of a quadrant containing the last thread and the bottom of the blind hole; (d) the diagram showing the camera positions where (a)-(c) were captured, listed as *C1*, *C2* and *C3*, respectively.

The image collection starts with placing the scope at the center of the opening of the

threaded hole and pointing inwards to the hole bottom, after camera calibration with open toolbox in OpenCV. To guarantee 50% image overlap among the sequential images, the test piece was moved forwards by micro-positioning stage with known step distance of 0.1 mm in this experiment. One image was collected by the Stryker™ system at each step, Fig. 8.2(c). A sequence of axial images were recorded, which covers about 60° of the threaded hole from hole opening to the bottom. Figure 8.3 shows endoscopic images of the internal threads, with clear vision of the peak (curved bright belt) and the valley (dark area just axially beyond the peak). The similar peak-valley pattern is repeated in the sequence from Fig. 8.3(a)–8.3(c) as shown in Fig. 8.3(d). The effect of specular reflections is insignificant due to light source design of the scope and lack of consistency in the sequence of images.

After scanning one quadrant, the test piece was re-positioned so that the scope was at the center of the opening. For $\geq 50\%$ image overlap between neighbor quadrants, the test piece was rotated 30° using the rotation stage on which the scope was mounted. The same procedure above was then performed to collect a set of axial images for the second quadrant. By repeating this, 12 sets of sequence images were recorded for the 360° scanning of the entire hole.

8.2 Feature-based 3D registration for two point clouds

8.2.1 Iterative Closest Point (ICP)

3D surface reconstruction has been widely used in fields of industrial design, reverse engineering, surface defect inspection, manufacturing, virtual reality and even homeland security. Various optical 3D surface reconstruction methods have been developed, such as time-of-flight [3], structured-light [47], laser scanning [7], structure-from-motion [87], multi-view stereo vision [54] and *etc.* These 3D reconstruction methods create different 3D point clouds of various density, efficiency and accuracy. A common attribute is the generation of partial surface of the scanned object or scene in general cases due to the limited field of view of the camera/sensor. To build a complete surface, the 3D surface acquisition system needs to be moved around to capture all the parts of the object/scene from different perspectives. 3D geometrical registration of all of these point clouds, that are created from

different perspectives, into the same coordinate system is one of the most important and critical steps in 3D reconstruction. The most widely used method for registering 3D point clouds is called ICP [11].

With a proper initial rough alignment and sufficient overlapping 3D points, ICP algorithm obtains an optimal registration solution by minimizing the distance between point-to-point correspondences, known as closest point, in an iterative way [11]. The output of ICP algorithm is a 3D rigid transformation matrix (combination of rotation and translation) from source point cloud to reference cloud such that the root mean square (RMS) between correspondences is minimal. Many improved ICP algorithms have been proposed and studied since introduction of ICP [111]. Different with point-to-point approach, point-to-plane ICP minimizes the sum of the squared distance between a point in the source data and the tangent plane at its correspondence point [20]. Both of the point-to-point and point-to-plane ICP approaches require a good initial coarse alignment, which may be performed manually. Mian *et al.* developed an automatic pairwise registration of 3D point clouds by a novel tensor representation, which represents semi-local geometric structure patches of the point clouds [84]. Although ICP has become the most popular method for 3D registration, there remains a fundamental problem of ICP-based methods. They do not work well for the 3D registration of plane, cylinder and other objects with repetitive geometric structures. One solution to this problem is by attaching reference marks (RM) on the object [33]. However, the RM methods require preparation work before 3D scanning, and also they are very limited to complex and rough surfaces.

To solve this repetitive geometry issue, a feature-based 3D registration algorithm is proposed in this study to align two point clouds that are generated by vision-based 3D reconstruction, such as SfM, multiview stereo and structured-light scanning. To take advantage of the 2D texture image of the object, features are detected from the texture images and matched to find the 3D correspondences between the two point clouds, with what 3D registration is much simplified. In this study, a test is performed with multiview stereo vision data. Performance evaluation of our proposed method and ICP-based ones (point-to-point, point-to-plane and Mian's) is performed. The comparison result demonstrates that the proposed method works for the alignment of repetitive geometries and performs more

accurate, efficient and robust than ICP-based methods for this example case.

8.2.2 Feature-based 3D registration algorithm

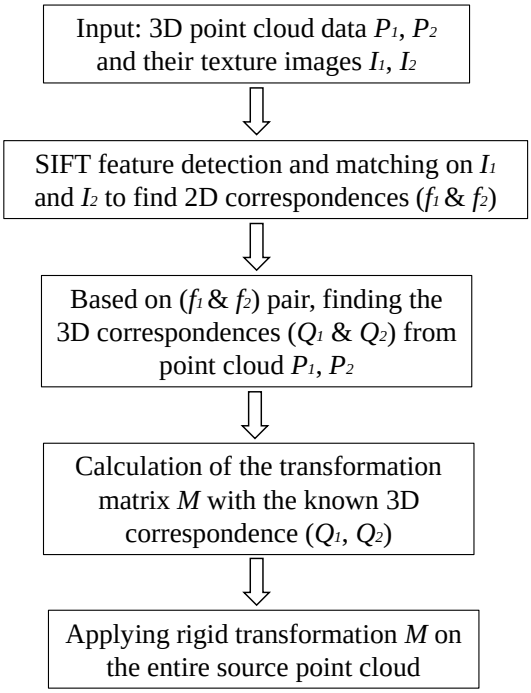


Figure 8.4: The flow chart of the proposed feature-based 3D registration algorithm

Fig. (8.4) shows the flow chart of the proposed feature-based 3D registration algorithm, which follows a similar principle as the RM method, but with the ability to automate and generalize using computer vision. Instead of manually attaching physical reference markers on the object surface in RM, our approach detects the unique "markers" (features) from the 2D texture images. By using advanced computer vision algorithms of feature detection and matching, pairs of corresponding feature points are collected. In this study, SIFT [78] detection algorithm was applied to find the unique "markers" from each texture image; RANSAC was used to select the robust matching feature pairs by calculating the Euclidean distance in SIFT descriptor space. The 3D coordinates of these correspondences can be easily retrieved from the 3D point clouds. 3D registration problem is then much simplified

by only calculating the rigid-body transformation matrix between these 3D correspondences. Unlike the RM method, our approach does not require any preparation work before scanning and is not limited to complex surface or working space.

Given a set of N axial images for the ith quadrant $I_i = \{I_{i1}, I_{i2}, ..., I_{iN}\}$ that are captured by linear motion of camera in one quadrant, feature detection and matching algorithms are applied to find the corresponding points in the image set. Let F_i represent all the features that are detected from I_i, and G_i for the features used in the 3D reconstruction. G_i is a subset of F_i, $G_i \in F_i$, since not every feature in an image has correspondences in its axial neighbors. With a priori knowledge of camera extrinsic parameters, two 3D point clouds representing these two quadrants can be generated independently by the multiview stereo vision technique, see the details in [134].

In this experiment, 155 axial images were captured for each quadrant covering the range from the opening through the bottom. Here, we utilized state-of-art software VisualSfM [162] for the processing and visualization for each quadrant, shown in Fig. 8.5, with label of the opening of threaded hole, and the start and end of the threads. The colorful straight line in Fig. 8.5 is the series of camera positions of the line of these axial images as the micro-positioning stage moved inwards.

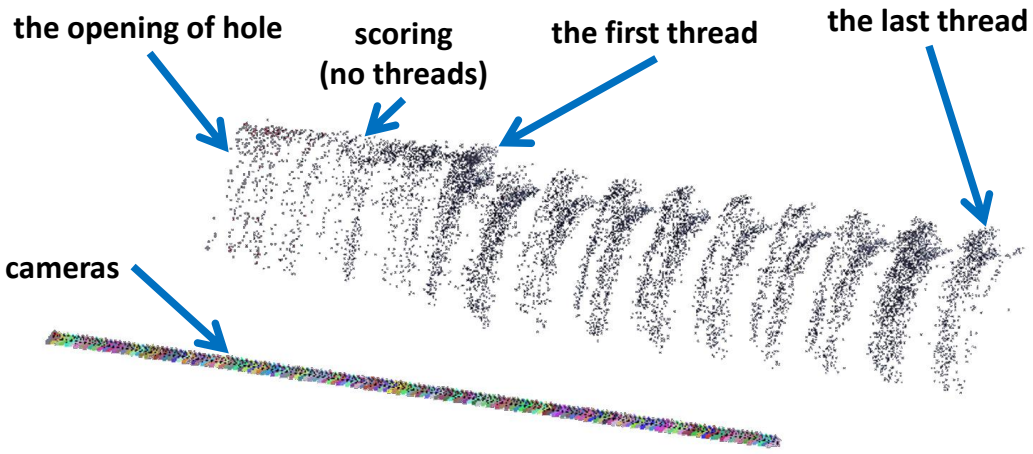


Figure 8.5: The 3D reconstruction of one quadrant of the recessed internally threaded hole from a sequence of axial images with known camera position and orientation.

To register two neighbor (ith and jth) quadrants together, we apply the same feature detection and matching algorithm on the angular neighbor pairs to find the corresponding features between these two quadrants. These correspondences are used for our proposed feature-based 3D registration that labeled as $H_{i,j}$ and $H_{j,i}$, respectively. $H_{i,j}$ represents all the features in F_i that have correspondences in F_j, so $H_{i,j} \in F_i$. Here, for the common features that are used in both 3D reconstruction and registration between ith and jth quadrants, let $P_{i,j}$ represent their 3D coordinates in the ith quadrant coordinate system. The sizes of $P_{i,j}$ and $P_{j,i}$ are the same as $3 \times m$, where m is the total number of 3D correspondences in ith and jth quadrants. To register the ith quadrant data (source) to the jth one (reference), we have

$$RP_{i,j}^{k} - t = P_{j,i}^{k}, \tag{8.1}$$

for kth pair of 3D correspondence, $k = 1, 2, ..., m$. In Eq. (8.3), $P_{i,j}$ and $P_{j,i}$ are known; R is a 3×3 rotation matrix from the coordinate system of the source data to the reference. Strictly speaking, $\det(R) = 1$. $t = [t_x, t_y, t_z]^T$ is the translation vector. By subtracting off the respective mean $\overline{P_{i,j}}$ from data $P_{i,j}$, the effect of the translation vector t is eliminated. Considering this is an over-determined problem, rotation R can be solved by:

$$min \quad \|RY_1 - Y_2\|^2 \tag{8.2}$$

where $Y_i = P_{i,j} - \overline{P_{i,j}}$. R can be obtained by taking the singular value decomposition of the covariance matrix $Y_2 Y_1^T$, and t can be calculated by $t = R\overline{P_{i,j}} - \overline{P_{j,i}}$. So far, the registration of these two point clouds is achieved by applying rotation R and translation t on the entire source point cloud.

8.2.3 Comparison of feature-based registration with ICP methods

To evaluate our proposed feature-based 3D registration algorithm, performances of our method and standard ICP-based (point-to-point, point-to-plane and Mian's) methods are compared. The visualization of the registration results are shown in Fig. 8.6. The initial alignment is shown in Fig. 8.6(a) by stacking two quadrants together directly. The obvious "discontinuity" of the two point clouds is caused by mis-alignment between camera center

and hole axis in the practical setup. Since these two point clouds were nearly identical and repetitive in geometry, all the ICP-based methods result in a complete overlapping model, see Fig. 8.6(c–e). These ICP results are completely wrong even with root mean square (RMS) error of only about 0.088 mm, see Table 8.1. In contrast, the proposed feature-based 3D registration algorithm produced a qualitatively better result, see Fig. 8.6(b), with a rough 30° rotation angle and geometrical continuity.

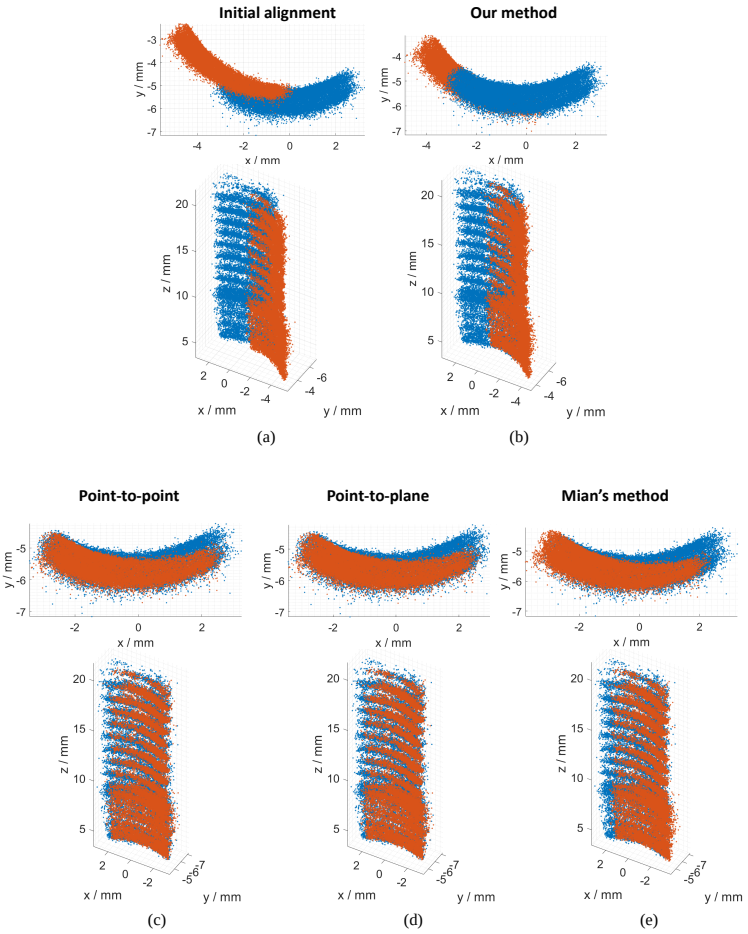


Figure 8.6: The comparison of 3D registration results by our proposed feature-based method and popular ICP approaches, showing with top and side views.

To better understand the significance of the proposed feature-based 3D registration algorithm, two tables were created with more details of the performance comparison of

Table 8.1: The comparison of our method and ICP approaches

	error of rotation angle (°)	computation time (sec)	RMS of registration (mm)
our method	2.4	0.001	0.138
point-to-point ICP	26.7	1574	0.088
point-to-plane ICP	26.0	1186	0.088
Mian's method	22.2	78	0.087

our method and ICP-based approaches. Table 8.1 shows that our method generated 2.4° error (8%) for the estimation of rotation angle (we consider 30° is the ground truth). All three ICP-based methods generated more than 22° error (74%) for the rotation angle estimation. Among ICP-based methods, Mian's performs slightly better than the other two. The performance of point-to-point and point-to-plane methods are very close, except computation time. Moreover, our method is much faster than ICPs, allowing over 50,000× improvement by only calculating a rigid transformation with known 3D correspondences.

Table 8.2.3 shows the comparison of the robustness of our method and standard ICP-based methods. Five independent tests were performed, following by the same flow chart but with different quadrants data of the internal threads. The comparison result shows that our method keeps a very low estimation error of the rotation angle within a threaded hole.

All the algorithms were implemented in MATLAB, running on a Dell Precision 5510 with 2.8 GHz Intel E3-1505M CPUs, 32.0 GB memory in a 64-bit Window operating system.

This proposed feature-based 3D registration method can be used for the registration of point clouds that are not just generated by multiview stereo 3D reconstruction, but also for other vision-based 3D reconstruction techniques, such as structured-light scanning. The texture images of the object or scene from different perspectives can be generated by three-step phase-shifting algorithm [48]. The same as our proposed approach, feature

Table 8.2: The rotation angle evaluation of our method and ICP-based approaches with multiple tests (°)

	our method	point-to-point ICP	point-to-plane ICP	Mian's method
test 1	5.51	28.69	27.98	22.46
test 2	2.44	26.70	26.02	22.16
test 3	4.05	30.27	30.53	18.71
test 4	3.24	27.45	27.45	24.09
test 5	3.10	26.18	26.39	23.29
avg.	3.67	27.86	27.86	22.14
std.	1.18	1.64	1.78	2.06

detection and matching are performed to find the 2D matching features, based on which 3D correspondences are retrieved from individual 3D model at each view. Moreover, this proposed method can be also used to improve the SfM 3D reconstruction. Different with multiview stereo, SfM doesn't require camera information. It estimates all the camera parameters and 3D scene simultaneously by solving a non-linear, non-convex optimization problem [46]. With little camera baseline and therefore high depth uncertainty, the non-convexity of SfM may generate a completely unacceptable 3D model. The strategy proposed here of classifying images as two categories, one for 3D reconstruction and the other for 3D registration, can be used to eliminate the effect of little camera baseline issue.

8.3 Feature-based 3D panoramic registration algorithm

In the above work, a feature-based 3D registration algorithm is proposed to solve the repetitive geometry issue for the point clouds that are generated by vision-based 3D reconstruction techniques. To register all the quadrants together, a panoramic registration algorithm is proposed based on Section 8.2 to achieve the registration of these highly similar quadrants from 360° scanning into an accurate 3D surface reconstruction by taking advantage of the robustness of image features and their 3D correspondences.

With the same notation in Section 8.2, we have the transformation between two neighbor

(ith and jth) quadrants as:

$$R_{i,j}(P_{i,j} - C_{i,j}) = P_{j,i}, \tag{8.3}$$

for each corresponding 3D point, where $P_{i,j}$ and $P_{j,i}$ are known, $R_{i,j}$ is a 3×3 rotation matrix, and $C_{i,j} = [C^x_{i,j}, C^y_{i,j}, C^z_{i,j}]^T$ is the linear offset of the two coordinate systems of different quadrants. It is an over-determined problem for the three variables $C_{i,j}$ in Eq. (8.3), since there are more than 3 correspondences between two neighbor quadrants with $\geq 50\%$ image overlap. Considering that the rotation angles among neighbor quadrants are identical, R was used to replace $R_{i,j}$ in Eq. (8.3) for simplification. So the 3D registration for each correspondence pair is:

$$\begin{aligned} RP^1_{i,j} - P^1_{j,i} &= RC_{i,j} \\ RP^2_{i,j} - P^2_{j,i} &= RC_{i,j} \\ &\vdots \\ RP^{m_{i,j}}_{i,j} - P^{m_{i,j}}_{j,i} &= RC_{i,j} \end{aligned} \tag{8.4}$$

where $m_{i,j}$ is the total number of the correspondence pair in ith and jth quadrants. By putting the left side of Eq. (8.4) into a column vector b with $3\times m_{i,j}$ entries, and concatenating multiple R's vertically into $\bar{R}$, then we have $\bar{R}_{i,j}C_{i,j} = b_{i,j}$. Solving C for each pair of neighbor quadrants initially appears independent from other quadrants from this equation. However, we have to consider the scenario in this study of 360° scanning, which means the last quadrant shares image overlap with the first one. Suppose there are k quadrants in the scanning, we obtain,

$$\begin{aligned} \bar{R}_{1,2}C_{1,2} &= b_{1,2} \\ \bar{R}_{2,3}C_{2,3} &= b_{2,3} \\ &\vdots \\ \bar{R}_{k-1,k}C_{k-1,k} &= b_{k-1,k} \\ \bar{R}_{k,1}C_{k,1} &= b_{k,1} \quad . \end{aligned} \tag{8.5}$$

Notice that $C_{k,1} = -C_{1,k} = -(C_{1,2} + C_{2,3} + ... + C_{k-1,k})$, Eq. (8.5) can be written into

matrix form, we have:

$$\begin{bmatrix} \bar{R}_{1,2} & & & & \\ & \bar{R}_{2,3} & & & \\ & & \ddots & & \\ & & & & \bar{R}_{k-1,k} \\ \bar{R}_{k,1} & \bar{R}_{k,1} & \bar{R}_{k,1} & \dots & \bar{R}_{k,1} \end{bmatrix} \begin{pmatrix} C_{1,2} \\ C_{2,3} \\ \vdots \\ C_{k-1,k} \end{pmatrix} = \begin{pmatrix} b_{1,2} \\ b_{2,3} \\ \vdots \\ b_{k-1,k} \\ -b_{k,1} \end{pmatrix}. \tag{8.6}$$

Let $\hat{R}$ represent the coefficient matrix in Eq. (8.6) and $\hat{b}$ for the vector of constant terms, resulting in

$$\begin{pmatrix} C_{1,2} \\ C_{2,3} \\ \vdots \\ C_{k-1,k} \end{pmatrix} = (\hat{R}^T \hat{R})^{-1} \hat{R}^T \hat{b}. \tag{8.7}$$

Since R is a full rank rotation matrix, the concatenating matrix $\bar{R}$ is full rank too. Thus, $\hat{R}$ has full rank, then $(\hat{R}^T \hat{R})^{-1}$ exists. By solving $C_{1,2}$, $C_{2,3}$, ..., $C_{k-1,k}$ from Eq. (8.7), all the quadrants can be registered together to generate the entire hole inner surface after transforming to the same coordinate system.

8.3.1 3D reconstruction result

Repeating the reconstruction of each quadrant, see Fig. 8.5, the point clouds of 12 quadrants were generated separately, which contain different total numbers of 3D points. In our experiment, there were about 12,000 points generated to represent each quadrant. In the case that the rotation axis aligned with camera centers perfectly, all the quadrants can be directly stacked together after pure rotation. However, maintaining this accurate alignment is impractical for in-line manufacturing. Trying to achieve accurate alignment manually in our setup results in the direct stack of multiple quadrants labeled in different colors in Fig. 8.7(a) and 8.7(b). To register all quadrants together, the standard approach using ICP didn't work well for this case, due to the repetitive geometries issue, see Section 8.2. Our proposed feature-based 3D panoramic registration was applied to take advantage of the high accuracy, robustness and distinguishability of SIFT features. With known rotation angle,

the calculation was simplified as illustrated in Eq. 8.7 to solve the mis-alignment problem between the rotation axis and camera centers. The final registration is an improvement over the direct stacking method, as shown in Fig. 8.7(c) and 8.7(d) with the top and front views.

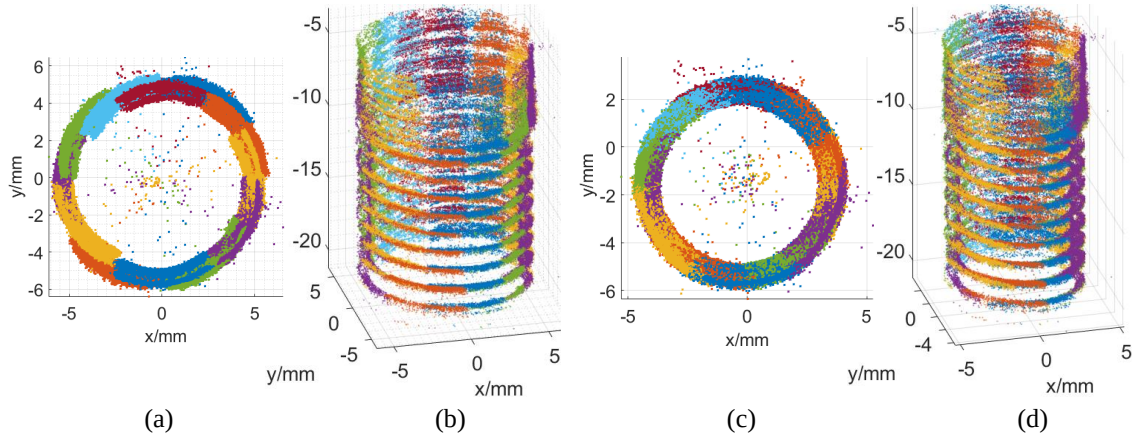


Figure 8.7: The 3D reconstruction result of entire scored (upper) and threaded portion of a blind hole with and without feature-based 3D panoramic registration algorithm. (a) and (b) show the top and front views of the reconstructed model by direct stacking without registration, respectively; (c) and (d) show the top and front views of improved 3D reconstruction with feature-based registration, respectively.

Figure 8.7(c) shows that there were a few outliers existing in the final registered 3D model, which might be caused by incorrect matching feature pairs. In this study, a statistical filter [85], which is based on the spatial distribution of each point, was applied to improve the quality of reconstructed model by identifying and trimming these outliers. Specifically, a mean distance for each point to its 30 nearest neighbors was computed in our study. With the assumption that the mean distances for all the points exhibit a Gaussian distribution with a mean and standard deviation. The points whose mean distances are 3× standard deviation larger than average were considered as outliers and removed from the final model, shown in Fig. 8.8(a) and 8.8(b).

All the software were run on a notebook Dell Precision M4700 with 20.0 GB memory and 2.7 GHz Intel i7-3740QM CPUs in a 64-bit Window operating system. The 3D reconstruction of each quadrant took about half an hour in C++; the 3D registration required

about 24 minutes to generate the entire hole; and noise reduction by statistical filter took 1.4 seconds to clean the outliers with an open source of Point Cloud Library.

8.3.2 Data comparison with X-ray CT

With 45° side-view camera, the back flank of each thread was not fully visible and not recorded. Therefore, the reconstruction of the entire hole only showed a representative model of the front flank of the threads, see Fig. 8.8(a). This lack of data may not be critical. To validate the feasibility of our approach to be a potential in-line measurement tool, several important thread parameters (minor diameter, major diameter and pitch) must be measured in the virtual 3D space and then compared with non-contact X-ray CT data. Minor diameter is defined as the minimum diameter of the threads and major diameter is the largest one; pitch is the axial distance between two neighbor threads. In this data analysis, the M8 coarse hole parameters were listed in Table 8.3 as a reference.

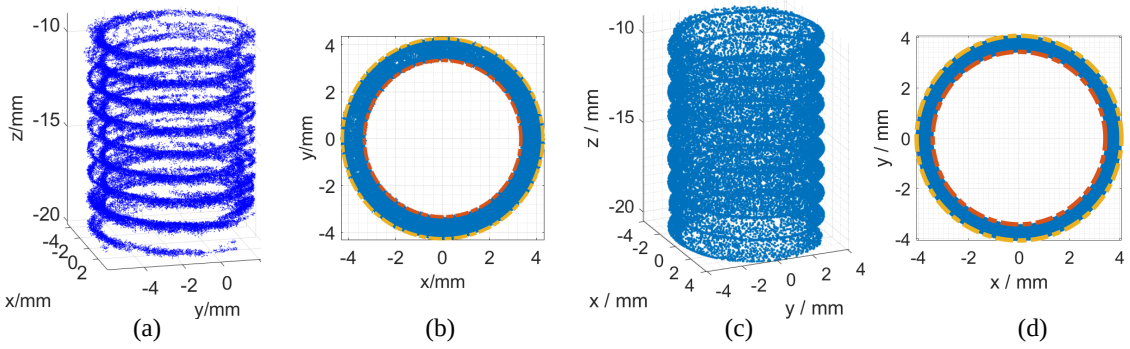


Figure 8.8: The comparison of X-ray CT and our metrology method on minor and major diameters based on 3D point clouds of internal threads. (a) our generated 3D point cloud after statistical filter; (b) the measurement of minor (red) and major (yellow) diameters by calculating the smallest and largest distances from the projected 2D points to the center; (c) the X-ray data of the same internal threads with front and back flanks; (d) the measurement of minor (red) and major (yellow) diameters by calculating the smallest and largest distances from the projected X-ray data to the center.

To measure the minor and major diameters from the 3D point cloud, only the internal threads beyond the scored bore opening is utilized for the following data analysis and comparison. Our reconstructed threads are projected to polar plane that is perpendicular

to the axis of the hole, see Fig. 8.8. The center of this projected 2D point set is calculated by fitting a circle on it. The minor radius of the hole was the smallest distance of the center to each projected point, labeled in orange; and the major radius corresponds to the largest distance, labeled in yellow, see Fig. 8.8(b). The same analysis was performed on the X-ray CT data with the result shown in Fig. 8.8(d). In the data acquisition procedure with X-ray CT, the test piece is positioned on a precision rotational stage. The X-ray light source generates a conic beam of electron that penetrates the test piece and is collected by a 2D detector as a Digital Radiograph image. During the rotation of the test piece at a constant step, a sequence of radiograph images is collected. In our study, the X-ray data collection is performed by X5000 CT scanning system (North Star Imaging Inc.). The 3D point cloud is generated by a commercial software efX-ct with its default iso value. Iso value is a number between 0 and 1, representing the intersection point where the surface will be generated based on density values within the volume. A surface is defined as a transition between two densities of materials.

Table 8.3: The measurements of minor diameter, major diameter and pitch of the threaded M8 hole with X-ray CT and our method compared to the thread specification.

	minor diameter	major diameter	pitch
X-ray CT with default iso	6.866 mm	8.145 mm	1.249 mm
our method	6.667 mm	8.539 mm	1.255 mm
M8 coarse specification	6.647 - 6.912 mm	8.000 - 8.340 mm	1.250 mm

Table 8.3 lists the measurement results of minor diameter, major diameter and pitch with X-ray CT and our method compared to the M8 specifications. All measurements of minor diameter were within the M8 specified range. Our method generated a -2.9% difference against the X-ray CT. For major diameter, the measurement result of our method generates +4.8% against with X-ray data. The measurement results of pitch of X-ray CT and our method are consistent with the manufacture standard of 1.25 mm for a M8 threaded hole, with error of -0.001 mm (-0.08%) and +0.005 mm (+0.4%), respectively.

As mentioned above, iso value is the threshold to distinguish material and non-material

within the volume. In our case, lower iso value may capture surrounding air as the metal threads; higher iso excludes materials within the threads. Therefore, different iso values generate various measurement results of minor and major diameters. Table 8.4 lists the comparison of the measurement results with different iso values. We can see that lower iso creates smaller minor and major diameters; while larger iso value "erodes" the material and generate larger diameters.

Table 8.4: The comparison of the measurement results with different iso values.

iso value	minor diameter	major diameter
2	6.467 mm	8.036 mm
4.83 (default)	6.866 mm	8.145 mm
4	6.850 mm	8.128 mm
8	6.933 mm	8.207 mm

Further analysis was performed between our method and X-ray CT for the similarity comparison of two point clouds. X-ray CT generated a dense 3D point cloud of the whole test piece including the external surface, since radiation X-ray can penetrate the entire object. The 3D point density of our reconstruction result varies across regions of thread peak, flank and valley, see Fig. 8.8(a). Because our optical reconstruction is based on 2D feature points, which distribute unevenly on the endoscopic images. More specifically, there is higher density of SIFT features on the peaks than valleys with a lack of features on the visible flanks observed at 45° side-viewing perspective. In contrast, the obtained X-ray point cloud had a more uniform 3D distribution of points on the threads surfaces of peak, valley and flank, see Fig. 8.8(c).

8.4 Discussion and conclusion

This study proposed a new approach for the 3D optical metrology of small holes with complex and repeating side-wall geometries using machine vision technique and feature-based 3D panoramic registration. Due to the size limitation of the small object, the traditional

metrology methods do not work well [55, 150, 174, 56], except X-ray CT. However, X-ray CT approach is infeasible to in-line measurement of small threaded holes in thick metal blocks. As mentioned above, rotation and linear motions are performed on the test piece, the scanning process takes about one hour for a complete and fine scan of our small test piece. It will take much more time for thick engine blocks, so that it is not feasible for practical applications. In our approach, a side-view medical scope was utilized to capture the high topographic relief of the internal surface profiles of the machined spiral thread. However, this required a complete rotation of the camera view within the blind hole. The point cloud of entire threaded hole can be generated by stacking all the reconstructed quadrants together directly with the assumption that the rotation axis passed through camera center. The existing outliers in the stacked model were then trimmed by statistical filter based on their spatial distribution. The comparison among the final reconstructed point cloud and X-ray CT data validated the feasibility of the proposed approach to be a potential method for the 3D optical in-line metrology for these small internal surface profiles in metal.

Although, a 45° side-view scope was used in the case study experiment, any wide-angle side-viewing scope could be used as long as the valley was clearly resolved with image based features to measure major diameter. By analyzing images acquired with a right-angled mirror at the distal end of a forward viewing scope, such as a rigid commercial scope or the smaller flexible SFE, this 90° side-view appears to generate a more unbiased (centered) distribution of image features at the bottom of the valleys, where both flanks are visible, see Fig. 8.9. A commercial scope of 2.9 mm OD and a 45° tilting mirror were used to capture the internal threads with 90°side-viewing angle, see Fig. 8.9(a). However, part of the scope itself was shown in the image due to the large size and small FOV. With SFE of 1.2 mm OD, internal threads were clearly shown without obstruction in Fig. 8.9(b). Both images show a more unbiased axial distribution of features with camera orientation like Fig. 8.1 versus Fig. 8.3(d) that are used in this study. In the experiment, 0.1 mm step distance was chosen as the camera moves inwards to maintain $\geq 50\%$ image overlap. Different size of the step can be utilized depending on camera orientation, FOV, and the size of hole. With larger camera FOV or larger hole, larger step distance can satisfy the image overlap requirement, resulting in fewer images that need to be collected and processed. This refinement would

dramatically decrease the computation time in both reconstruction and registration.

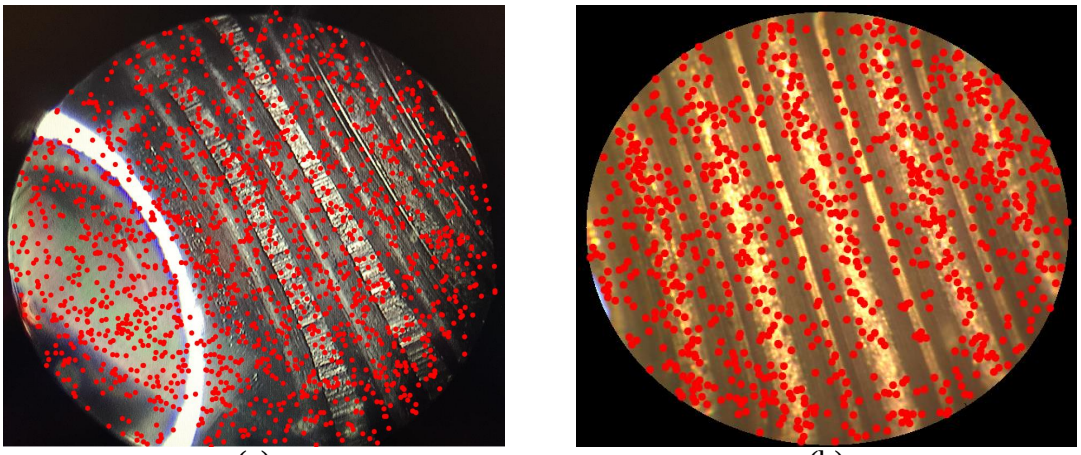


Figure 8.9: The 90° side-view images of internal threads with commercial scope and SFE, the red dots represent the detected SIFT features, which show unbiased feature distribution. (a) Endoscopic image with forward view 2.9 mm OD rigid scope and 45° tilting mirror. Part of the scope with ring illumination is shown on the left, and (b) endoscopic image with flexible 1.2 mm OD SFE. For image clarity, red-dot feature density has been reduced by 10× in (a) and 2× in (b).

In the practical application of quality control of the internal threads, Go/no-Go gauges are used as the gold standard to judge "Yes/No" for each dimensional parameter of the threaded hole. This binary decision can be automated to perform rapidly but with a major flaw from a data perspective: each parameter can only be defined to be within lower and upper bounds (Go and no-Go), instead of a specific value. The definition of the bounds highly depends on the interval of the gauge size. With the 3D point cloud that is generated by our approach (or X-ray CT), future work is to measure these parameters in the virtual space by a virtual gauge. The generated 3D point cloud can provide graded parameter values instead of a range. Moreover, point defects like pores, chips, and cracks on the internal thread surface can be detected from the dense point cloud, which cannot be achieved by Go/no-Go gauging. The advantages of using our approach to currently used metrology: 1) ability of non-contact inspection; 2) accessibility of variable data; 3) independence to operators' skills; 4) high automatability; 5) low labor and maintenance cost; 6) potential

high efficiency with high-power computer; 7) feasibility of in-line metrology; and 8) vision of the threaded surface for point, line (scratch) and spectroscopic (color) inspection. Direct vision is a big advantage of our method comparing with other technologies that are currently feasible for the 3D metrology of small internal surface: X-ray CT [95], OCT [128, 168] and laser ring beam scanning [174]. Duplicating human vision has a long term major advantage because that this can harness the incredible capabilities being generated in computer vision and machine learning using deep neural networks for Artificial Intelligence (AI) analyses [66]. Moreover, vision provides an efficient and robust feature-based tool for registration and comparison among multiple 3D reconstructions.

Future work on advancing our approach will solve several limitations listed below. 1) During the image collection, accurate camera position and rotation are demanded for the 3D reconstruction of each quadrant. By applying bound constrained bundle adjustment (BCBA) algorithm [46] with prior knowledge of the accuracy and precision of industrial robot behavior using a pre-calibrated inspection scope, our approach can be improved to handle the real-world scenario of inaccurate camera poses. BCBA can also be used to solve the registration problem if the relative rotation angles among different quadrants are inaccurate or even unknown. 2) The image collection of multiple quadrants requires multiple times for the camera to move inwards and outwards, which increases the time and complexity in practical applications. This may be solved by utilizing a cone mirror [18] placing in front of a forward-view camera. At each axial position, one frame is captured with side-view on the 360° wall of the hole. Another advantage of this design is that it avoids the 3D registration of multiple quadrants, which may result in more accurate and efficient 3D surface model formation. The drawback is that a new feature detection algorithm is required for axial images, and also a more complex calibration method is needed for camera/mirror setup. 3) Higher density point clouds are needed, especially at the surface extremes (peaks and valleys) used for dimensional metrology in this application. The state-of-the-art dense multiview stereo technique may dramatically increase the density of the final point cloud with *s-t* cut optimization and visibility and photo-consistency constraints [149].

Chapter 9

CONCLUSIONS AND FUTURE WORK

The focus of this dissertation research is 3D computer vision algorithm development for specific applications in biomedical and industrial area.

9.1 Summary of dissertation work

All the three projects in this dissertation research have some commonalities. First, all of them require 3D reconstruction based on endoscopic images; the accuracy requirements of the reconstruction results are different. The automated bladder surveillance system only needs a 3D panoramic view to assist the urologist achieve a complete scan. The 3D reconstruction of surgical field of brain tumor cavity has higher requirement of the accuracy, since the reconstructed 3D coordinates of the tumor cells are passed to the automated surgical robot (RAVEN) for the following tumor ablation path generation and *etc*; 3D metrology of internal threads require the most accurate 3D point cloud (tens of microns) for the quality control of internal threads hole in engine blocks.

The second common thing is that all 3D reconstruction approaches take advantage of the "known" camera extrinsic parameters that the robotic system provides. The multi-segmented flexible robot in the automated bladder surveillance system couldn't provide the poses of the cameras, but it performed as a spiral pattern. This information could be utilized to simplify the procedure of finding the lateral matching frames [130, 170]. RAVEN II surgical robot provides very high degree of dexterity, precise cutting, and accurate navigation with surgeon-assisted visual feedback, but the pose information of cameras that RAVEN system obtains was inaccurate due to the mechanical slack or stretch in cable. By utilizing the external stereo vision track system, an accurate 3D model was generated. The accurate 3D model also can be achieved by using the bound constrained 3D reconstruction algorithm, BCBA, as long as we known the error range of the RAVEN robot arm.

One different thing among these three applications is the 3D reconstruction algorithms. The biomedical applications (bladder and brain) found the SIFT features from multiple endoscopic images, and refine their 3D coordinated by using an optimization algorithm, bundle adjustment. This kind of reconstruction resulted sparse point cloud from these biomedical images. Whereas the 3D metrology for the surface of small internal threads hole requires not only high accuracy but also dense mapping. The endoscopic image of the threads surface contains very limited features. Block matching finds the dense correspondences by calculating minimum value of mean square error.

9.2 Future work in 3D Optical Metrology

The other future work of this dissertation research is to generate dense 3D point cloud and meshes for 3D optical metrology of small internal surface, such as <7 mm minor diameter threads hole. Nowadays, there are two main approaches for small hole metrology in industry. The most common one is to inspect the holes with mechanical contact gauges. It is simple but with many disadvantages, such as cost, mechanical wear and etc. The other common way is to use X-ray CT, but it is not feasible for in-line quality control due to its time expense.

In Chapter 7 and Chapter 8, two 3D optical metrology algorithms were developed, one is based on forward-view axial-stereo vision for dense reconstruction of a single thread; the other one is based on side-view sparse 3D reconstruction for entire hole.

To achieve **complete** and **dense** 3D scan of a hole, one approach is to process multiple axial-stereo images. Instead of using two frame, multiple images are collected as the scope moving towards to hole bottom along the hole axis, see Fig. 9.1. A single 3D point on the threads has multiple projections on these axial-stereo images. Same to conventional multi-view stereo, the surface can be retrieved by computing a cost function on a 3D voxel [123].

To generate the 3D dense reconstruction of entire threaded hole, there are several factors need to be considered:

- The experiment setup

  How to align the optical axis, axis of motion and the threads hole axis? From the

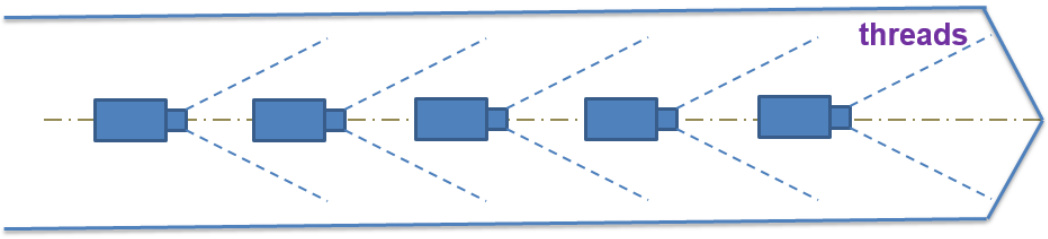


Figure 9.1: As the SFE moves along the axis of hole that is aligned with the optical axis, multiple images are captured with their pose information recorded by industry robot. Noise and mismatching problematic issue could be reduced by this multi-view stereo approach.

mathematics model that introduced in [44], it is not necessary to have these three axis aligned, but good alignment can simplify the algorithm and also improve the robustness. The reason is that with good alignment, the corresponding points in different images can be easily located. This can narrow down the search window, reduce noise and prevent the effect of similar points in other regions. This alignment issue can be solved by using three linear motion stages and two tilt adjustment stages [56].

- Image acquisition

  By continuing using the SFE, the distortion of SFE image should be corrected before 3D reconstruction. This distortion can be dramatically reduced by *non-parametric distortion correction algorithm* [9]. Fig. 9.2 shows an example[1] of the application of non-parametric distortion correction on the SFE image.

- Motion error

  As the SFE moves along the axis of threads hole, the motion may affect the recorded images. There are two reasons causing motion blur. One is because of the imaging frequency of the SFE (30fps), this is conventional motion blur which also happens on

[1] cite from Dr. Tim Soper's Ph.D dissertation [129]

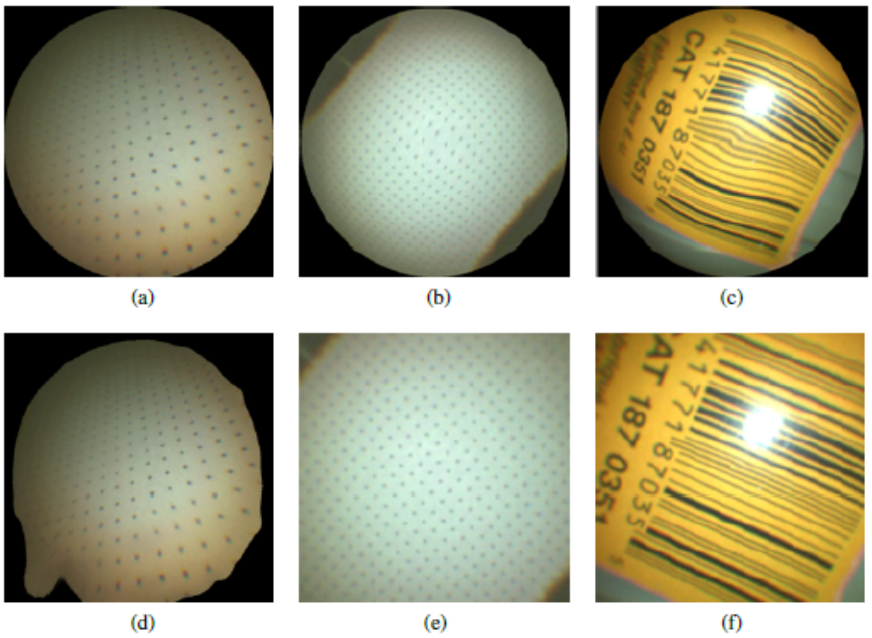


Figure 9.2: Example SFE images before and after nonparametric distortion correction. Images are paired column-wise. The first pair (a&d) are from an SFE probe that exhibits only a modest level of distortion. The remaining image pairs (b&e and c&f) are from an SFE probe with a more appreciable level of central scan distortion.

other imaging system. Many algorithms have been developed to solve this issue, such as Richardson-Lucy algorithm. The other one unique for the SFE. It is the dynamics effect of motion on the scanning fiber in the SFE. To solve the unique motion blur, we may calibrate the "the scope with motion" with the non-parametric distortion correction algorithm. The dynamic of industrial robotics system is known, based on which an experiment can be set up to simulate the motion in the real manufacturing environment with a translation motor.

- Back flank

  There is only the front flank of threads being captured by axial-stereo in the study [44]. To imaging the back flank, a small mirror can be placed at the bottom of the threads hole Fig. 9.3. The SFE image may capture the front flank only, the back flank

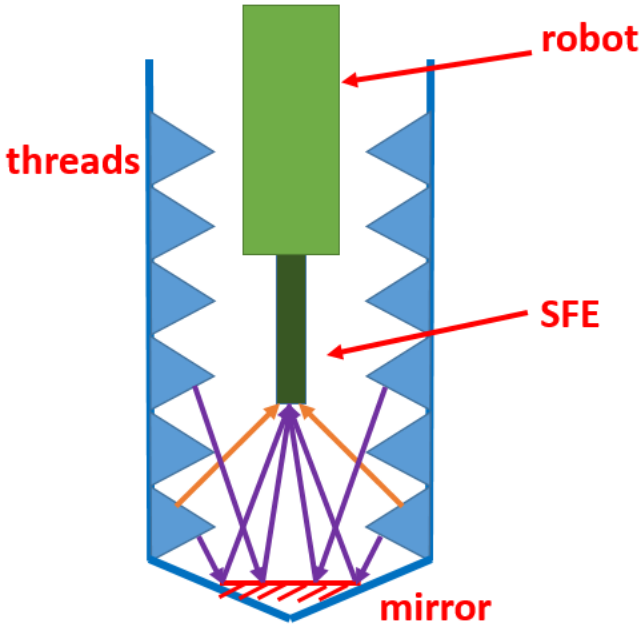


Figure 9.3: The scheme of SFE imaging of front and back flank by placing a mirror at the bottom of the threaded hole. The SFE is moved along the axis of threads hole by robot to image front and back flanks evenly. The orange represent the reflectance of front flank, as well as the purple stands for the back flank reflectance. The change of the SFE location causes the amount change of front and back flanks captured. As SFE move downwards, more back flank is captured.

only or both, depends on the location of the SFE and the FOV angle. Fig. 9.3 shows the case that the SFE captures both front and back flanks. The orange represent the light reflected by the front flank is captured, as well as the purple represents the reflectance of the back flank. Note that the alignment of optical axis, motion axis and hole axis affects the image acquisition of front and back flanks. With the mirror approach and large FOV, the entire surface of the threads can be captured as the endoscope move in and out of the threads hole.

- Computation time

  The computation time is critical to industry applications. Besides improving the algorithm itself, using different programming language can achieve much higher com-

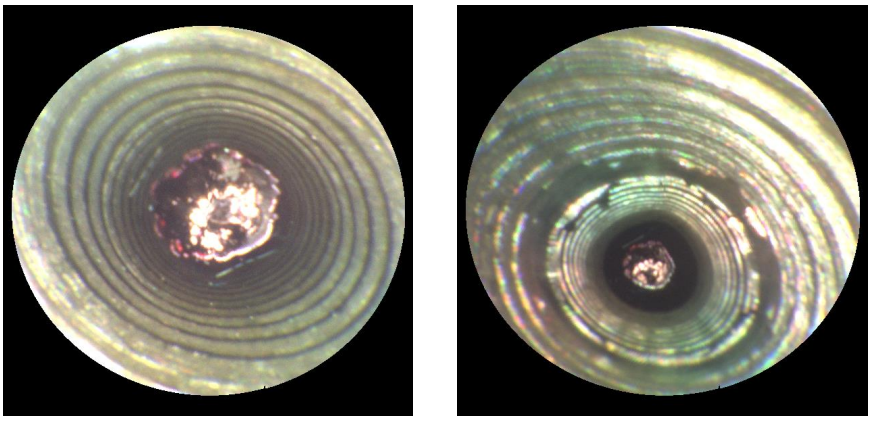


Figure 9.4: Examples of the SFE images of back flank of the threads. The left image only shows the back flank, the right one shows the back flank and also the front one.

putation efficiency. Instead of using MATLAB for the previous three projects, C/C++ software will be developed in the future work, partially implemented with CUDA for the parallel programming. CUDA has great advantage for such pixel-by-pixel computation.

9.3 Future work in 3D Image-guided Surgery

9.3.1 Dense 3D Mapping

The current 3D reconstruction of the surgical field for brain tumor removal is based on sparse features (SIFT) that were detected from endoscopic images. The 3D surface is generated by fitting spline through these generated sparse 3D points [41, 42, 58, 57]. Surface fitting may cause the reduction of geometrical details, which effects the accuracy of 3D reconstruction. To deal with this problem, a dense 3D point cloud representing the surgical field is required. The same as the future work in 3D metrology, image distortion correction is the preparation work, which can be achieved by non-parametric distortion correction algorithm.

3D dense point cloud of surgical field can be generated by multiview-stereo algorithm that can be classified as four categories [123]: 1) a cost function on a 3D volume is computed, from where a surface can be extract; 2) Evolving a surface iteratively to minimize a cost

function, such as space carving [67]; 3) 3D scene is reconstructed by consistency constraints among a set of depth maps; 4) fitting surface through sparse 3D point cloud, the one we used in the current reconstruction.

9.3.2 SLAM of the surgical robot

SLAM (simultaneous localization and mapping) is a computational problem in robotics to recover the 3D structure of the environment and also the 3D coordinate of sensor. By using Kalman Filter, the camera pose and the 3D structure can be well estimated iteratively as the robot moving around. The current camera pose can be measured by RAVEN system. The estimation of the camera pose can be converged well by Kalman filter even this measurement is inaccurate.

The advantage of implement SLAM is to improve the efficiency of the 3D reconstruction. With a good estimation of newly added camera pose, the 3D structure can be generated more efficiently than the conventional way since we have a good initial guess of the camera pose. The refined camera pose that calculated by bundle adjustment can be used to refine the Kalman filter in the next iteration.

BIBLIOGRAPHY

[1] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M Seitz, and Richard Szeliski. Building rome in a day. *Communications of the ACM*, 54(10):105–112, 2011.

[2] C. Albl and T. Pajdla. Constrained bundle adjustment for panoramic cameras. pages 118–124, 2013.

[3] Markus-Christian Amann, Thierry Bosch, Risto Myllyla, Marc Rioux, and Marc Lescure. Laser ranging: a critical review of usual techniques for distance measurement. *Optical Engineering*, 40(1):10–19, 2001.

[4] Tomer Anor, Joseph R Madsen, and Pierre Dupont. Algorithms for design of continuum robots using the concentric tubes approach: a neurosurgical example. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 667–673. IEEE, 2011.

[5] Sam Bae and Ronald Korniski. Toward a 3d endoscope for minimally invasive surgery. *SPIE Newsroom*, 2011.

[6] Andrea Bajo, Roger E Goldman, Long Wang, Dennis Fowler, and Nabil Simaan. Integration and preliminary evaluation of an insertable robotic effectors platform for single port access surgery. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 3381–3387. IEEE, 2012.

[7] Emmanuel P Baltsavias. A comparison between photogrammetry and laser scanning. *ISPRS Journal of photogrammetry and Remote Sensing*, 54(2):83–94, 1999.

[8] TH Baron. Natural orifice transluminal endoscopic surgery. *British journal of surgery*, 94(1):1, 2007.

[9] Joao P Barreto, Rahul Swaminathan, and Jose Roquette. Non parametric distortion correction in endoscopic medical images. In *3DTV Conference, 2007*, pages 1–4. IEEE, 2007.

[10] Alexander Behrens. Creating panoramic images for bladder fluorescence endoscopy. *Acta Polytechnica*, 48(3), 2008.

[11] Paul J Besl and Neil D McKay. Method for registration of 3-d shapes. In *Robotics-DL tentative*, pages 586–606. International Society for Optics and Photonics, 1992.

[12] G Bianco, A Gallo, F Bruno, and M Muzzupappa. A comparison between active and passive techniques for underwater 3d applications. *ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3816:357–363, 2011.

[13] François Blais. Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 13(1), 2004.

[14] John A Bosch. *Coordinate measuring machines and systems*. CRC Press, 1995.

[15] Duane C Brown. The bundle adjustmentprogress and prospects. *Int. Archives Photogrammetry*, 21(3):1–1, 1976.

[16] Matthew R Burkhardt, Timothy D Soper, Woon Jong Yoon, and Eric J Seibel. Controlling the trajectory of a flexible ultrathin endoscope for fully automated bladder surveillance. *IEEE/ASME Transactions on Mechatronics*, 19(1):366–373, 2014.

[17] Paul H Calamai and Jorge J Moré. Projected gradient methods for linearly constrained problems. *Mathematical programming*, 39(1):93–116, 1987.

[18] JS Chahl and MV Srinivasan. Reflective surfaces for panoramic imaging. *Applied Optics*, 36(31):8275–8285, 1997.

[19] Jie Chen, Shuning Li, and Shiaofen Fang. Quantification of tooth displacement from cone-beam computed tomography images. *American Journal of Orthodontics and Dentofacial Orthopedics*, 136(3):393–400, 2009.

[20] Yang Chen and Gérard Medioni. Object modelling by registration of multiple range images. *Image and vision computing*, 10(3):145–155, 1992.

[21] Neil T Clancy, Shobhit Arya, Ji Qi, Danail Stoyanov, George B Hanna, and Daniel S Elson. Polarised stereo endoscope and narrowband detection for minimal access surgery. *Biomedical optics express*, 5(12):4108–4117, 2014.

[22] Neil T Clancy, Danail Stoyanov, Lena Maier-Hein, Anja Groch, Guang-Zhong Yang, and Daniel S Elson. Spectrally encoded fiber-based structured lighting probe for intraoperative 3d imaging. *Biomedical optics express*, 2(11):3119–3128, 2011.

[23] Robert Collier. *Optical holography*. Elsevier, 2013.

[24] Toby Collins and Adrien Bartoli. 3d reconstruction in laparoscopy with close-range photometric stereo. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pages 634–642. Springer, 2012.

[25] Roch M Comeau, Abbas F Sadikot, Aaron Fenster, and Terry M Peters. Intraoperative ultrasound for guidance and tissue shift correction in image-guided neurosurgery. *Medical Physics*, 27(4):787–800, 2000.

[26] Michael S Cookson, Harry W Herr, Zuo-Feng Zhang, Scott Soloway, Pramod C Sogani, and William R Fair. The treated natural history of high risk superficial bladder cancer: 15-year outcome. *The Journal of urology*, 158(1):62–67, 1997.

[27] David Crandall, Andrew Owens, Noah Snavely, and Dan Huttenlocher. Discrete-continuous optimization for large-scale structure from motion. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 3001–3008. IEEE, 2011.

[28] Amir Degani, Howie Choset, Alon Wolf, and Marco A Zenati. Highly articulated robotic probe for minimally invasive surgery. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 4167–4172. IEEE, 2006.

[29] T Desmettre, JM Devoisselle, and S Mordon. Fluorescence properties and metabolic features of indocyanine green (icg) as related to angiography. *Survey of ophthalmology*, 45(1):15–27, 2000.

[30] S Donat, Amanda North, Guido Dalbagni, and Harry W Herr. Efficacy of office fulguration for recurrent low grade papillary bladder tumors less than 0.5 cm. *The Journal of urology*, 171(2):636–639, 2004.

[31] Pierre E Dupont, Jesse Lock, Brandon Itkowitz, and Evan Butler. Design and control of concentric-tube robots. *Robotics, IEEE Transactions on*, 26(2):209–225, 2010.

[32] Sabry F El-Hakim, Jean-Angelo Beraldin, and Francois Blais. Comparative evaluation of the performance of passive and active 3d vision systems. In *Digital Photogrammetry and Remote Sensing'95*, pages 14–25. International Society for Optics and Photonics, 1995.

[33] Marek Franaszek, Geraldine S Cheok, and Christoph Witzgall. Fast automatic registration of range images from 3d imaging systems using sphere targets. *Automation in Construction*, 18(3):265–274, 2009.

[34] Yasutaka Furukawa and Jean Ponce. Accurate, dense, and robust multiview stereopsis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(8):1362–1376, 2010.

[35] David Gallup, Jan-Michael Frahm, Philippos Mordohai, and Marc Pollefeys. Variable baseline/resolution stereo. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[36] Jason Geng. Structured-light 3d surface imaging: a tutorial. *Advances in Optics and Photonics*, 3(2):128–160, 2011.

[37] Katia Genovese and Carmine Pappalettere. Axial stereo-photogrammetry for 360 measurement on tubular samples. *Optics and Lasers in Engineering*, 45(5):637–650, 2007.

[38] Sylvain Gioux, Hak Soo Choi, and John V Frangioni. Image-guided surgery using invisible near-infrared light: fundamentals of clinical translation. *Molecular imaging*, 9(5):237, 2010.

[39] Roger E Goldman, Andrea Bajo, Lara S MacLachlan, Ryan Pickens, S Duke Herrell, and Nabil Simaan. Design and performance evaluation of a minimally invasive telerobotic platform for transurethral surveillance and intervention. *Biomedical Engineering, IEEE Transactions on*, 60(4):918–925, 2013.

[40] Roger E Goldman, Andrea Bajo, and Nabil Simaan. Compliant motion control for continuum robots with intrinsic actuation sensing. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 1126–1132. IEEE, 2011.

[41] Yuanzhen Gong, Danying Hu, Blake Hannaford, and Eric J Seibel. Toward real-time endoscopically-guided robotic navigation based on a 3d virtual surgical field model. In *SPIE Medical Imaging*. International Society for Optics and Photonics, 2015 (in press).

[42] Yuanzhen Gong, Danying Hu, Blake Hannaford, and Eric J Seibel. Accurate 3d virtual reconstruction of surgical field using calibrated trajectories of an image-guided medical robot. *Journal of Medical Imaging*, (in press).

[43] Yuanzhen Gong, Tomothy D Soper, Vivian W Hou, Danying Hu, Blake Hannaford, and Eric J Seibel. Mapping surgical fields by moving a laser-scanning multimodal scope attached to a robot arm. In *SPIE Medical Imaging*, pages 90362S–90362S. International Society for Optics and Photonics, 2014.

[44] Yuanzheng Gong, Richard S Johnston, C David Melville, and Eric J Seibel. Axial-stereo 3d optical metrology of internally machined parts using high-quality imaging from a scanning laser endoscope. In *International Symposium on Optomechatronic Technologies 2014*, 2014.

[45] Yuanzheng Gong, Richard S Johnston, C David Melville, and Eric J Seibel. Axial-stereo 3-d optical metrology for inner profile of pipes using a scanning laser endoscope. *International journal of optomechatronics*, 9(3):238–247, 2015.

[46] Yuanzheng Gong, De Meng, and Eric J Seibel. Bound constrained bundle adjustment for reliable 3d reconstruction. *Optics express*, 23(8):10771–10785, 2015.

[47] Yuanzheng Gong and Song Zhang. Ultrafast 3-d shape measurement with an off-the-shelf dlp projector. *Optics express*, 18(19):19743–19754, 2010.

[48] Yuanzheng Gong and Song Zhang. High-resolution, high-speed three-dimensional shape measurement using projector defocusing. *Optical Engineering*, 50(2):023603–023603, 2011.

[49] Neda Haj-Hosseini, Benjamin Kistler, and Karin Wårdell. Development and characterization of a brain tumor mimicking fluorescence phantom. In *SPIE BiOS*, pages 894505–894505. International Society for Optics and Photonics, 2014.

[50] Irfan Halim and Ali Tavakkolizadeh. Notes: The next surgical revolution? *International Journal of Surgery*, 6(4):273–276, 2008.

[51] Blake Hannaford, Jacob Rosen, Diana W Friedman, Hawkeye King, Phillip Roan, Lei Cheng, Daniel Glozman, Ji Ma, Sina Nia Kosari, and Lee White. Raven-ii: an open platform for surgical robotics research. *Biomedical Engineering, IEEE Transactions on*, 60(4):954–959, 2013.

[52] Kanako Harada, Kota Tsubouchi, Masakatsu G Fujie, and Toshio Chiba. Micro manipulators for intrauterine fetal surgery in an open mri. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, pages 502–507. IEEE, 2005.

[53] Kevin Harding. *Handbook of Optical Dimensional Metrology*. CRC Press, 2013.

[54] Richard Hartley and Andrew Zisserman. *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[55] Robert J Hocken and Paulo H Pereira. *Coordinate measuring machines and systems*. CRC Press, 2011.

[56] En Hong, Hongwei Zhang, Reuven Katz, and John S Agapiou. Non-contact inspection of internal threads of machined parts. *The International Journal of Advanced Manufacturing Technology*, 62(1-4):221–229, 2012.

[57] Danying Hu, Yuanzheng Gong, Blake Hannaford, and Eric J Seibel. Path planning for semi-automated simulated robotic neurosurgery. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 2639–2645. IEEE, 2015.

[58] Danying Hu, Yuanzheng Gong, Blake Hannaford, and Eric J Seibel. Semi-autonomous simulated brain tumor ablation with ravenii surgical robot using behavior tree. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3868–3875. IEEE, 2015.

[59] David Huang, Eric A Swanson, Charles P Lin, Joel S Schuman, William G Stinson, Warren Chang, Michael R Hee, Thomas Flotte, Kenton Gregory, Carmen A Puliafito, et al. Optical coherence tomography. *Science*, 254(5035):1178–1181, 1991.

[60] Ben Kehoe, Gregory Kahn, Jeffrey Mahler, Jonathan Kim, Alex Lee, Anna Lee, Keisuke Nakagawa, Sachin Patil, W Douglas Boyd, Pieter Abbeel, et al. Autonomous multilateral debridement with the raven surgical robot. In *International Conference on Robotics and Automation*, 2014.

[61] Chul Sung Kim, Chang-Woo Park, Bong-Seok Kim, Jae-Bok Song, and Jung-Hoon Hwang. Design of robotic surgical instrument for minimally invasive surgical robot system. In *Control, Automation and Systems (ICCAS), 2012 12th International Conference on*, pages 1720–1723. IEEE, 2012.

[62] Bryan Klingner, David Martin, and James Roseborough. Street view motion-from-structure-from-motion. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 953–960. IEEE, 2013.

[63] Jan J Koenderink, Andrea J Van Doorn, et al. Affine structure from motion. *JOSA A*, 8(2):377–385, 1991.

[64] Kurt Konolige and Motilal Agrawal. Frameslam: From bundle adjustment to real-time visual mapping. *Robotics, IEEE Transactions on*, 24(5):1066–1077, 2008.

[65] Alexandre Krupa, Jacques Gangloff, Christophe Doignon, Michel F de Mathelin, Guillaume Morel, Joël Leroy, Luc Soler, and Jacques Marescaux. Autonomous 3-d positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. *Robotics and Automation, IEEE Transactions on*, 19(5):842–853, 2003.

[66] G Senthil Kumar, U Natarajan, and SS Ananthan. Vision inspection system for the identification and classification of defects in mig welding joints. *The International Journal of Advanced Manufacturing Technology*, 61(9-12):923–933, 2012.

[67] Kiriakos N Kutulakos and Steven M Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, 2000.

[68] Guy Lamouche, Charles-Etienne Bisaillon, Marc Dufour, Bruno Gauthier, Romain Maciejko, and Jean-Pierre Monchalin. Optical coherence tomography for industrial and biomedical applications. In *Speckle06: Speckles, From Grains to Flowers*, pages 63410T–63410T. International Society for Optics and Photonics, 2006.

[69] B Leblanc, AJ Duclos, F Benard, J Côté, L Valiquette, JM Paquin, F Mauffette, R Faucher, and JP Perreault. Long-term followup of initial ta grade 1 transitional cell carcinoma of the bladder. *The Journal of urology*, 162(6):1946–1950, 1999.

[70] Cameron M Lee, Christoph J Engelbrecht, Timothy D Soper, Fritjof Helmchen, and Eric J Seibel. Scanning fiber endoscopy with highly flexible, 1 mm catheterscopes for wide-field, full-color imaging. *Journal of biophotonics*, 3(5-6):385–407, 2010.

[71] Joohyuk Lee, Hyojoo Son, Changmin Kim, and Changwan Kim. Skeleton-based 3d reconstruction of as-built pipelines from laser-scan data. *Automation in construction*, 35:199–207, 2013.

[72] Yiping Li, Roberto Rey-Dios, David W Roberts, Pablo A Valdés, and Aaron A Cohen-Gadol. Intraoperative fluorescence-guided resection of high-grade gliomas: a comparison of the present techniques and evolution of future strategies. *World neurosurgery*, 2013.

[73] Zheng Li and Ruxu Du. Design and analysis of a bio-inspired wire-driven multi-section flexible robot. *Int J Adv Robotic Sy*, 10(209), 2013.

[74] Jonathan TC Liu, Daphne Meza, and Nader Sanai. Trends in fluorescence image-guided surgery for gliomas. *Neurosurgery*, 2014.

[75] Yang Liu, Adam Q Bauer, Walter J Akers, Gail Sudlow, Kexian Liang, Duanwen Shen, Mikhail Y Berezin, Joseph P Culver, and Samuel Achilefu. Hands-free, wireless goggles for near-infrared fluorescence and real-time image-guided surgery. *Surgery*, 149(5):689–698, 2011.

[76] KE Loewke, DB Camarillo, CA Jobst, and JK Salisbury. Real-time image mosaicing for medical applications. *Studies in health technology and informatics*, 125:304–309, 2006.

[77] Manolis IA Lourakis and Antonis A Argyros. Is levenberg-marquardt the most efficient optimization algorithm for implementing bundle adjustment? In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1526–1531. IEEE, 2005.

[78] David G Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

[79] Mitchell JH Lum, Diana CW Friedman, Ganesh Sankaranarayanan, Hawkeye King, Kenneth Fodero, Rainer Leuschke, Blake Hannaford, Jacob Rosen, and Mika N Sinanan. The raven: Design and validation of a telesurgery system. *The International Journal of Robotics Research*, 28(9):1183–1197, 2009.

[80] Bin Ma, Timo Zimmermann, Manfred Rohde, Simon Winkelbach, Feng He, Werner Lindenmaier, and Kurt EJ Dittmar. Use of autostitch for automatic stitching of microscope images. *Micron*, 38(5):492–499, 2007.

[81] Jeffrey Mahler, Sanjay Krishnan, Michael Laskey, Siddarth Sen, Adithyavairavan Murali, Ben Kehoe, Sachin Patil, Jiannan Wang, Mike Franklin, Pieter Abbeel, et al. Learning accurate kinematic control of cable-driven surgical robots using data cleaning and gaussian process regression. *IEEE International Conference on Automation Science and Engineering*, pages 532–539, 2014.

[82] Hani Marcus, Dipankar Nandi, Ara Darzi, and Guang-Zhong Yang. Surgical robotics through a keyhole: from today's translational barriers to tomorrow's" disappearing" robots. *IEEE transactions on bio-medical engineering*, 60(3):674–681, 2013.

[83] Maged Mahmoud Margany and Mazlan Hashim. Three-dimensional of coastal front reconstruction using radarsat-1 sar satellite data. 2012.

[84] Ajmal S Mian, Mohammed Bennamoun, and Robyn A Owens. A novel representation and feature matching algorithm for automatic pairwise registration of range images. *International Journal of Computer Vision*, 66(1):19–40, 2006.

[85] Gary Miner, Robert Nisbet, and John Elder IV. *Handbook of statistical analysis and data mining applications*. Academic Press, 2009.

[86] Kaushik Mitra and Rama Chellappa. A scalable projective bundle adjustment algorithm using the l infinity norm. In *Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference on*, pages 79–86. IEEE, 2008.

[87] Etienne Mouragnon, Maxime Lhuillier, Michel Dhome, Fabien Dekeyser, and Patrick Sayd. Generic and real-time structure from motion using local bundle adjustment. *Image and Vision Computing*, 27(8):1178–1193, 2009.

[88] GP Moustris, SC Hiridis, KM Deliparaschos, and KM Konstantinidis. Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 7(4):375–392, 2011.

[89] Y Munz, K Moorthy, A Dosis, JD Hernandez, S Bann, F Bello, S Martin, A Darzi, and T Rockall. The benfits of stereoscopic vision in robotic-assisted performance on bench models. *Surgical Endoscopy And Other Interventional Techniques*, 18(4):611–616, 2004.

[90] Arya Nabavi, Peter McL Black, David T Gering, Carl-Fredrik Westin, Vivek Mehta, Richard S Pergolizzi Jr, Mathieu Ferrant, Simon K Warfield, Nobuhiko Hata, Richard B Schwartz, et al. Serial intraoperative magnetic resonance imaging of brain shift. *Neurosurgery*, 48(4):787–798, 2001.

[91] Haring JW Nauta. Error assessment during image guided and imaging interactive stereotactic surgery. *Computerized medical imaging and graphics*, 18(4):279–287, 1994.

[92] Carl A Nelson, Xiaoli Zhang, Bhavin C Shah, Matthew R Goede, and Dmitry Oleynikov. Multipurpose surgical robot as a laparoscope assistant. *Surgical endoscopy*, 24(7):1528–1532, 2010.

[93] Quyen T Nguyen and Roger Y Tsien. Fluorescence-guided surgery with live molecular navigation [mdash] a new cutting edge. *Nature Reviews Cancer*, 2013.

[94] Christopher Nimsky, Oliver Ganslandt, Simon Cerny, Peter Hastreiter, Günther Greiner, and Rudolf Fahlbusch. Quantification of, visualization of, and compensation for brain shift using intraoperative magnetic resonance imaging. *Neurosurgery*, 47(5):1070–1080, 2000.

[95] Alison Noble, Richard Hartley, Joseph Mundy, and James Farley. X-ray metrology for quality assurance. In *Robotics and Automation, 1994. Proceedings., 1994 IEEE International Conference on*, pages 1113–1119. IEEE, 1994.

[96] Nagaaki Ohyama, Hirohisa Machida, Kazuhiro Gono, Takashi Obi, Yasuo Hamamoto, Takao Endo, Masahiro Yamaguchi, Yasushi Sano, and Shigeaki Yoshida. Appearance of enhanced tissue features in narrow-band endoscopic imaging. *Journal of biomedical optics*, 9(3):568–577, 2004.

[97] Daniel A Orringer, Alexandra Golby, and Ferenc Jolesz. Neuronavigation in the surgical management of brain tumors: current and future trends. 2012.

[98] Quinn T Ostrom, Haley Gittleman, Paul Farah, Annie Ondracek, Yanwen Chen, Yingli Wolinsky, Nancy E Stroup, Carol Kruchko, and Jill S Barnholtz-Sloan. Cbtrus statistical report: Primary brain and central nervous system tumors diagnosed in the united states in 2006-2010. *Neuro-oncology*, 15(suppl 2):ii1–ii56, 2013.

[99] Emmanuel P Papadakis and Grant A Gordon. Ultrasonic instruments and devices. *The Journal of the Acoustical Society of America*, 109(5):1761–1763, 2001.

[100] Vicente Parot, Daryl Lim, Germán González, Giovanni Traverso, Norman S Nishioka, Benjamin J Vakoc, and Nicholas J Durr. Photometric stereo endoscopy. *Journal of biomedical optics*, 18(7):076017–076017, 2013.

[101] Jochen Penne, Kurt Höller, Michael Stürmer, Thomas Schrauder, Armin Schneider, Rainer Engelbrecht, Hubertus Feußner, Bernhard Schmauss, and Joachim Hornegger. Time-of-flight 3-d endoscopy. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, pages 467–474. Springer, 2009.

[102] T Pfeifer and W Koenig. Fiberoptics for in-line production measurement. *CIRP Annals-Manufacturing Technology*, 41(1):577–580, 1992.

[103] Kanchana Rathnayaka, Konstantin I Momot, Hansrudi Noser, Andrew Volp, Michael A Schuetz, Tony Sahama, and Beat Schmutz. Quantification of the accuracy of mri generated 3d models of long bones compared to ct generated 3d models. *Medical engineering & physics*, 34(3):357–363, 2012.

[104] Mireille Reeff, Friederike Gerhard, Philippe C Cattin, and Gábor Székely. Mosaicing of endoscopic placenta images. *GI Jahrestagung (1)*, 2006:467–474, 2006.

[105] Roberto Rey-Dios and Aaron A Cohen-Gadol. Intraoperative fluorescence for resection of hemangioblastomas. *Acta neurochirurgica*, 155(7):1287–1292, 2013.

[106] G Robinson and J Bruce C Davies. Continuum robots-a state of the art. In *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, volume 4, pages 2849–2854. IEEE, 1999.

[107] D Caleb Rucker, Bryan A Jones, and Robert J Webster. A geometrically exact model for externally loaded concentric-tube continuum robots. *Robotics, IEEE Transactions on*, 26(5):769–780, 2010.

[108] John R Ruge and Jonathan Liu. Use of 5-aminolevulinic acid for visualization and resection of a benign pediatric brain tumor: Case report. *Journal of Neurosurgery: Pediatrics*, 4(5):484–486, 2009.

[109] Brian G Saar, Richard S Johnston, Christian W Freudiger, X Sunney Xie, and Eric J Seibel. Coherent raman scanning fiber endoscopy. *Optics letters*, 36(13):2396–2398, 2011.

[110] Joaquim Salvi, Xavier Armangué, and Joan Batlle. A comparative review of camera calibrating methods with accuracy evaluation. *Pattern recognition*, 35(7):1617–1635, 2002.

[111] Joaquim Salvi, Carles Matabosch, David Fofi, and Josep Forest. A review of recent range image registration methods with accuracy evaluation. *Image and Vision computing*, 25(5):578–596, 2007.

[112] Joaquim Salvi, Jordi Pages, and Joan Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, 2004.

[113] P Sayd, M Dhome, and J-M LAVEST. Axial stereovision to recover straight cylinders. In *CESA'96 IMACS Multiconference: computational engineering in systems applications*, pages 560–564, 1996.

[114] Christoph Schmalz, Frank Forster, Anton Schick, and Elli Angelopoulou. An endoscopic 3d scanner based on structured light. *Medical Image Analysis*, 16(5):1063–1072, 2012.

[115] Michael Schulder and Peter W Carmel. Intraoperative magnetic resonance imaging: impact on brain tumor surgery. *Cancer Control*, 10(2):115–124, 2003.

[116] Sebastian Schuon, Christian Theobalt, James Davis, and Sebastian Thrun. High-quality scanning using time-of-flight depth superresolution. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, pages 1–7. IEEE, 2008.

[117] Heinrich Schwenke, Ulrich Neuschaefer-Rube, Tilo Pfeifer, and Horst Kunzmann. Optical methods for dimensional metrology in production engineering. *CIRP Annals-Manufacturing Technology*, 51(2):685–699, 2002.

[118] Eric J Seibel, Robert E Carroll, Jason A Dominitz, Richard S Johnston, C David Melville, Cameron M Lee, Steven M Seitz, and Michael B Kimmey. Tethered capsule endoscopy, a low-cost and high-performance alternative technology for the screening of esophageal cancer and barrett's esophagus. *Biomedical Engineering, IEEE Transactions on*, 55(3):1032–1042, 2008.

[119] Eric J Seibel, Javier A Jo, C David Melville, Richard S Johnston, Christopher R Naumann, and Michael D Saunders. Image-guided intervention in the human bile duct using scanning fiber endoscope system. In *SPIE BiOS*, pages 82180B–82180B. International Society for Optics and Photonics, 2012.

[120] Eric J Seibel, Richard S Johnston, and C David Melville. A full-color scanning fiber endoscope. In *Biomedical Optics 2006*, pages 608303–608303. International Society for Optics and Photonics, 2006.

[121] Eric J Seibel, C David Melville, Richard S Johnston, Yuanzheng Gong, Kathy Agnew, Seine Chiang, and Elizabeth M Swisher. In vivo laser-based imaging of the human fallopian tube for future cancer detection. In *SPIE BiOS*, pages 93040Q–93040Q. International Society for Optics and Photonics, 2015.

[122] Eric J Seibel, Timothy D Soper, Matthew R Burkhardt, Michael P Porter, and W Jong Yoon. Multimodal flexible cystoscopy for creating co-registered panoramas of the bladder urothelium. In *SPIE BiOS*, pages 82071A–82071A. International Society for Optics and Photonics, 2012.

[123] Steven M Seitz, Brian Curless, James Diebel, Daniel Scharstein, and Richard Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*, volume 1, pages 519–528. IEEE, 2006.

[124] Shomik Sengupta and Michael L Blute. The management of superficial transitional cell carcinoma of the bladder. *Urology*, 67(3):48–54, 2006.

[125] Sharmishtaa Seshamani, William Lau, and Gregory Hager. Real-time endoscopic mosaicking. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 355–363. Springer, 2006.

[126] A Simpson, Kay Sun, T Pheiffer, D Rucker, A Sills, R Thompson, and M Miga. Evaluation of conoscopic holography for estimating tumor resection cavities in model-based image-guided neurosurgery. 2014.

[127] American Cancer Society. Cancer facts & figures 2013. Technical report, Atlanta: American Cancer Society, 2013.

[128] Guiju Song and Kevin Harding. Oct for industrial applications. In *Photonics Asia*, pages 85630N–85630N. International Society for Optics and Photonics, 2012.

[129] Timothy Soper. *A navigation system for an ultrathin scanning fiber bronchoscope in the peripheral airways*. University of Washington, 2009.

[130] Timothy D Soper, Michael P Porter, and Eric J Seibel. Surface mosaics of the bladder reconstructed from endoscopic video for automated surveillance. *Biomedical Engineering, IEEE Transactions on*, 59(6):1670–1680, 2012.

[131] Liliana Soroceanu, Yancey Gillespie, MB Khazaeli, and Harald Sontheimer. Use of chlorotoxin for targeting of primary brain tumors. *Cancer Research*, 58(21):4871–4879, 1998.

[132] Mark R Stroud, Stacey J Hansen, and James M Olson. In vivo bio-imaging using chlorotoxin-based conjugates. *Current pharmaceutical design*, 17(38):4362, 2011.

[133] Robert H Sturges, Schitt Laowattana, et al. A flexible, tendon-controlled device for endoscopy. *The International journal of robotics research*, 12(2):121–131, 1993.

[134] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[135] Richard Szeliski and Philip HS Torr. Geometrically constrained structure from motion: Points on planes. In *3D Structure from Multiple Images of Large-Scale Environments*, pages 171–186. Springer, 1998.

[136] Kenji Takita, Yoshifumi Sasaki, Tatsuo Higuchi, and Koji KOBAYASHI. High-accuracy subpixel image registration based on phase-only correlation. *IEICE transactions on fundamentals of electronics, communications and computer sciences*, 86(8):1925–1934, 2003.

[137] Philippe Thevenaz, Urs E Ruttimann, and Michael Unser. A pyramid approach to subpixel registration based on intensity. *Image Processing, IEEE Transactions on*, 7(1):27–41, 1998.

[138] QB Tong, ZL Ding, JC Chen, LL Ai, and F Yuan. The research of screw thread parameter measurement based on position sensitive detector and laser. In *Journal of Physics: Conference Series*, volume 48, page 561. IOP Publishing, 2006.

[139] Jorg-Christian Tonn and Walter Stummer. Fluorescence-guided resection of malignant gliomas using 5-aminolevulinic acid: practical use, risks, and pitfalls. *Clinical neurosurgery*, 55:20–26, 2007.

[140] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustmenta modern synthesis. In *Vision algorithms: theory and practice*, pages 298–372. Springer, 2000.

[141] Pablo A Valdés, Frederic Leblond, Anthony Kim, Brent T Harris, Brian C Wilson, Xiaoyao Fan, Tor D Tosteson, Alex Hartov, Songbai Ji, Kadir Erkmen, et al. Quantitative fluorescence in intracranial tumor: implications for ala-induced ppix as an intraoperative biomarker. *Journal of neurosurgery*, 115(1):11, 2011.

[142] Maurice HPH van Beurden, Gert Van Hoey, Haralambos Hatzakis, and Wijnand A IJsselsteijn. Stereoscopic displays in medical domains: a review of perception and performance effects. In *IS&T/SPIE Electronic Imaging*, pages 72400A–72400A. International Society for Optics and Photonics, 2009.

[143] MHPH Van Beurden, WA IJsselsteijn, and JF Juola. Effectiveness of stereoscopic displays in medicine: a review. *3D Research*, 3(1):1–13, 2012.

[144] Madelon NM van der Aa, Ewout W Steyerberg, Chris Bangma, Bas WG van Rhijn, Ellen C Zwarthoff, and Theo H van der Kwast. Cystoscopy revisited as the gold standard for detecting bladder cancer recurrence: diagnostic review bias in the randomized, prospective cefub trial. *The Journal of urology*, 183(1):76–80, 2010.

[145] Mandana Veiseh, Patrik Gabikian, S-Bahram Bahrami, Omid Veiseh, Miqin Zhang, Robert C Hackman, Ali C Ravanpay, Mark R Stroud, Yumiko Kusuma, Stacey J Hansen, et al. Tumor paint: a chlorotoxin: Cy5. 5 bioconjugate for intraoperative visualization of cancer foci. *Cancer research*, 67(14):6882–6888, 2007.

[146] Olga Veksler. Stereo correspondence by dynamic programming on a tree. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 384–390. IEEE, 2005.

[147] Sébastien Vergnole, Guy Lamouche, Marc L Dufour, Bruno Gauthier, and Christian Padioleau. Artifact removal in fourier-domain oct with a rotating probe. In *Biomedical Optics (BiOS) 2008*, pages 68510G–68510G. International Society for Optics and Photonics, 2008.

[148] Adrian Catalin Voicu, Ion Gheorghe Gheorghe, Liliana Laura Badita, and Adriana Cirstoiu. 3d measuring of complex automotive parts by multiple laser scanning. *Applied Mechanics and Materials*, 371:519–523, 2013.

[149] Hoang-Hiep Vu, Patrick Labatut, Jean-Philippe Pons, and Renaud Keriven. High accuracy and visibility-consistent dense multiview stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(5):889–901, 2012.

[150] T Wakayama, H Takano, and T Yoshizawa. Development of a compact inner profile measuring instrument. In *Optics East 2007*, pages 67620D–67620D. International Society for Optics and Photonics, 2007.

[151] Toshitaka Wakayama, Kizuku Machi, and Toru Yoshizawa. Small size probe for inner profile measurement of pipes using optical fiber ring beam device. In *Photonics Asia*, pages 85630L–85630L. International Society for Optics and Photonics, 2012.

[152] Xinghua Wang, Srabani Bhaumik, Qing Li, V Paul Staudinger, and Siavash Yazdanfar. Compact instrument for fluorescence image-guided surgery. *Journal of biomedical optics*, 15(2):020509–020509, 2010.

[153] Robert J Webster and Bryan A Jones. Design and kinematic modeling of constant curvature continuum robots: A review. *The International Journal of Robotics Research*, 2010.

[154] Robet J Webster III, John P Swensen, Joseph M Romano, and Noah J Cowan. Closed-form differential kinematics for concentric-tube continuum robots with application to visual servoing. In *Experimental Robotics*, pages 485–494. Springer, 2009.

[155] Juyang Weng, Thomas S Huang, and Narendra Ahuja. *Motion and structure from image sequences*. Springer Publishing Company, Incorporated, 2012.

[156] John Whitman, Matthew P Fronheiser, Nikolas M Ivancevich, and Stephen W Smith. Autonomous surgical robotics using 3-d ultrasound guidance: Feasibility study. *Ultrasonic imaging*, 29(4):213–219, 2007.

[157] JE Wickham. The new surgery. *British medical journal (Clinical research ed.)*, 295(6613):1581, 1987.

[158] Fokko P Wieringa, Henri Bouma, Pieter T Eendebak, Jean-Paul A van Basten, Harrie P Beerlage, Geert AHJ Smits, and Jelte E Bos. Improved depth perception with three-dimensional auxiliary display and computer generated three-dimensional panoramic overviews in robot-assisted laparoscopy. *Journal of Medical Imaging*, 15001:1, 2014.

[159] Daniel Wilby, Kay Thomas, Eleanor Ray, Barnaby Chappell, and Timothy OBrien. Bladder cancer: new tur techniques. *World journal of urology*, 27(3):309–312, 2009.

[160] Kin Hong Wong and Michael Ming-Yuen Chang. 3d model reconstruction by constrained bundle adjustment. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 3, pages 902–905. IEEE, 2004.

[161] Cornell Wright, Aaron Johnson, Aaron Peck, Zachary McCord, Allison Naaktgeboren, Philip Gianfortoni, Manuel Gonzalez-Rivero, Ross Hatton, and Howie Choset. Design of a modular snake robot. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 2609–2614. IEEE, 2007.

[162] Changchang Wu. Towards linear-time incremental structure from motion. In *3DTV-Conference, 2013 International Conference on*, pages 127–134. IEEE, 2013.

[163] Fu-li Wu and Xiang-yong Fang. An improved ransac homography algorithm for feature based image mosaic. In *Proceedings of the 7th WSEAS International Conference on Signal Processing, Computational Geometry & Artificial Vision*, pages 202–207. World Scientific and Engineering Academy and Society (WSEAS), 2007.

[164] Razali Yaakob, Alihossein Aryanfar, Alfian Abdul Halin, and Nasir Sulaiman. A comparison of different block matching algorithms for motion estimation. *Procedia Technology*, 11:199–205, 2013.

[165] Chenying Yang, Vivian Hou, Leonard Y Nelson, and Eric J Seibel. Color-matched and fluorescence-labeled esophagus phantom and its applications. *Journal of biomedical optics*, 18(2):026020–026020, 2013.

[166] Chenying Yang, Vivian Hou, Leonard Y Nelson, and Eric J Seibel. Mitigating fluorescence spectral overlap in wide-field endoscopic imaging. *Journal of biomedical optics*, 18(8):086012–086012, 2013.

[167] Chenying Yang, Vivian W Hou, Emily J Girard, Leonard Y Nelson, and Eric J Seibel. Target-to-background enhancement in multispectral endoscopy with background autofluorescence mitigation for quantitative molecular imaging. *Journal of biomedical optics*, 19(7):076014–076014, 2014.

[168] Jianing Yao, Panomsak Meemon, Michael Ponting, and Jannick P Rolland. Angular scan optical coherence tomography imaging and metrology of spherical gradient refractive index preforms. *Optics express*, 23(5):6428–6443, 2015.

[169] AN Yaroslavsky, PC Schulze, IV Yaroslavsky, R Schober, F Ulrich, and HJ Schwarzmaier. Optical properties of selected native and coagulated human brain tissues in vitro in the visible and near infrared spectral range. *Physics in medicine and biology*, 47(12):2059, 2002.

[170] Xianming Ye, Yuanzheng Gong, and W. Jong Yoon. Development of multi-segment steering mechanism and 3d panorama for automated bladder surveillance system. *Mechatronics, IEEE/ASME Transactions on*, 21(2):993–1003, 2016.

[171] Jong Yoon, Tim Soper, and Eric J Seibel. Automated 3d mosaicing and scan trajectories for surveillance of bladder cancer. *Journal of Medical Devices*, 5(2):027531, 2011.

[172] W Jong Yoon, Matthew A Brown, Per G Reinhall, Sangtae Park, and Eric J Seibel. Design and preliminary study of custom laser scanning cystoscope for automated bladder surveillance. *Minimally Invasive Therapy & Allied Technologies*, 21(5):320–328, 2012.

[173] W Jong Yoon, Sangtae Park, Per G Reinhall, and Eric J Seibel. Development of an automated steering mechanism for bladder urothelium surveillance. *Journal of medical devices*, 3(1):011004, 2009.

[174] T Yoshizawa and T Wakayama. Optical inner profile measurement for pipes and tubes. *Structural Longevity*, 6(1):1–7, 2011.

[175] Toru Yoshizawa. *Handbook of optical metrology: Principles and Applications*. CRC Press, 2009.

[176] Liang Zhang, Leonard Y Nelson, and Eric J Seibel. Spectrally enhanced imaging of occlusal surfaces and artificial shallow enamel erosions with a scanning fiber endoscope. *Journal of biomedical optics*, 17(7):0760191–07601911, 2012.

[177] Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer, and Mubarak Shah. Shape-from-shading: a survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(8):690–706, 1999.

[178] Song Zhang and Peisen S Huang. High-resolution, real-time three-dimensional shape measurement. *Optical Engineering*, 45(12):123601–123601, 2006.

[179] Yongjun Zhang, Kun Hu, and Rongyong Huang. Bundle adjustment with additional constraints applied to imagery of the dunhuang wall paintings. *ISPRS Journal of Photogrammetry and Remote Sensing*, 72:113–120, 2012.

[180] Zhengyou Zhang. Microsoft kinect sensor and its effect. *MultiMedia, IEEE*, 19(2):4–10, 2012.

[181] Marco Zuliani. Ransac for dummies. *With examples using the RANSAC toolbox for Matlab and more*, 2009.

VITA

YUANZHENG (LOUIS) GONG

University of Washington ⋄ Seattle, WA 98105
e-mail: yzgong7@gmail.com

EDUCATION

University of Washington, Seattle *June 2016 (expected)*
Doctor of Philosophy in Mechanical Engineering
Thesis: Improving the Accuracy of 3D Reconstruction in Robotic Vision Applications

Iowa State University *Decemeber 2011*
Master of Engineering in Mechanical Engineering

University of Science and Technology of China *July 2009*
Bachelor of Science in Precision Machinery & Instrumentation

EXPERIENCE

University of Washington March 2013 - Present
Research Assistant *Seattle, WA*

- Developing feature-based 3D panoramic registration algorithm
- Developed axial-stereo imaging algorithm to reconstruct 3D model of internal surface
- Developed 3D image-guided surgical robot with 3D reconstruction
- Improved the algorithm of accurate 3D reconstruction for automated bladder surveillance system with different scanning videos

University of Washington September 2012 - March 2013
Teaching Assistant *Seattle, WA*

- TA of class Thermodynamics in Mechanical Engineering department

Iowa State University January 2010 - May 2011
Research Assistant *Ames, IA*

- Improved software and hardware of high-speed and superfast 3D imaging system in 3D Machine Vision Lab, Virtual Reality Application Center
- Built a Graphics User Interface for image processing and 3D geometry manipulation in C++ with Qt platform

University of Science and Technology of China July 2007 - December 2009
Undergraduate Research Assistant *Hefei, China*

- Developed a new 3D reconstruction algorithm based on confocal technique with a sequence of microscopic images (Best Undergraduate Thesis)

PATENTS (6)

- US Patent 20,150,346,115 E.J. Seibel, **Y. Gong**, F. Braun, D.L. Gourley, "3D Optical Metrology of Internal Surface," 2015
- US Patent 62/276,730US1 (Provisional) E.J. Seibel, and **Y. Gong**, "3D Surface Reconstruction of Threaded Hole Using Side-View Scope," 2016
- US Patent 62/300,665US1 (Provisional) E. J. Seibel, M. Carson, **Y. Gong**, and J. Pitton, "3D Reconstruction of Dental Caries Using the Scanning-Fiber Endoscope," 2016
- US Patent 62/240,959US1 (Provisional) E.J. Seibel, L. Nelson, M. Timoshchuk, **Y. Gong**, Z. Xu, J. McLean "Imaging of Porphyrins for Detection of Active Dental Disease", 2015
- US Patent 62/088,951US1 (Provisional) E.J. Seibel, and **Y. Gong**, "Applying Bundle Adjustment in 3D Optical Metrology," 2014
- US Patent 62/080,139US1 (Provisional) B. Hannaford, E.J. Seibel D. Hu, and **Y. Gong**, "Semi-Autonomous Brain Tumor Ablation with Fluorescently Marked Tumor," 2014

BOOK CHAPTER (1)

- S. Zhang and **Y. Gong**, "High-speed, high-resolution 3-D imaging using projector defocusing," *Depth Map and 3D Imaging Applications: Algorithms and Technologies*, Information Science Reference, 2012

JOURNAL PUBLICATION (8)

- **Y. Gong**, D. Aneja, L, Shapiro and E.J. Seibel, "3D optical measurement of small internal threads using feature-based 3D panoramic registration," *Optics Express*, 2016 (under review)
- D. Hu, **Y. Gong**, L. N. Sekhar, E.J. Seibel and B. Hannaford, "Semi-autonomous image-guided brain tumor resection using an integrated robotic system: a bench-top study," *IEEE Transactions on Robotics*, 2016 (under review)
- X. Ye, **Y. Gong** and J. Yoon, "Development of Multi-segment Steering Mechanism and 3D Panorama for Automated Bladder Surveillance System," *IEEE/ASME Transactions on Mechatronics* 21(2), 993-1003, 2016
- **Y. Gong**, D. Meng and E.J. Seibel, "Bound constrained bundle adjustment for reliable 3D reconstruction," *Optics Express*, 23(8), 10771-10785, 2015
- **Y. Gong**, R. Johnston, D. Melville and E.J. Seibel, "Axial-stereo 3D optical metrology for inner profile of pipes using a scanning laser endoscope," *International Journal of Optomechatronics*, 9(3), 238-247, 2015
- **Y. Gong**, D. Hu, E.J. Seibel and B. Hannaford, "Accurate 3D virtual reconstruction of surgical field using calibrated trajectories of an image-guided medical robot," *Journal of Medical Imaging*, 1(3), 035002, 2014.
- **Y. Gong** and S. Zhang, "High-resolution, high-speed three-dimensional shape measurement using projector defocusing," *Optical Engineering*, 50(2), 023603, 2011.

- **Y. Gong** and S. Zhang, "Ultrafast 3-D shape measurement with an off-the-shelf DLP projector," *Optics Express*, 18(19), 19743-19754, 2010. (Cover paper)

CONFERENCE PROCEEDINGS (13)

- D. Hu, **Y. Gong** B. Hannaford and E.J. Seibel, "Semi-autonomous Simulated Brain Tumor Ablation with Raven II Surgical Robot using Behavior TreePath Planning for Semi-automated Simulated Robotic Neurosurgery," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2639-2645, 2015

- D. Hu, **Y. Gong** B. Hannaford and E.J. Seibel, "Semi-autonomous Simulated Brain Tumor Ablation with Raven II Surgical Robot using Behavior Tree," *IEEE International Conference on Robotics and Automation (ICRA)*, pp 3868-3875, 2015

- **Y. Gong**, D. Hu, B. Hannaford and E.J. Seibel, "Toward real-time endoscopically-guided robotic navigation based on a 3D virtual surgical field model," *Image-Guided Procedures, Robotic Interventions, and Modeling, SPIE Medical Imaging, Proc SPIE*, 2015 (accepted)

- E.J. Seibel, C.D. Melville, R.S. Johnston, **Y. Gong**, K. Agnew, S. Chiang and E.M. Swisher, "In vivo laser-based imaging of the human fallopian tube for future cancer detection," *Endoscopic Microscopy X, Proc. SPIE Photonics West*, vol. 9304, Feb 8, 2015

- D. Hu, **Y. Gong**, B. Hannaford and E.J. Seibel, "Semi-autonomous Simulated Brain Tumor Ablation with Raven II Surgical Robot using Behavior Tree," *National Robotics Initiative Principal Investigators (NRI PI) meeting*, Washington D.C. Nov. 19-20, 2014.

- **Y. Gong**, R.S. Johnston, C.D. Melville and E.J. Seibel, "Axial-stereo 3D optical metrology of internally machined parts using high-quality imaging from a scanning laser endoscope," *International Symposium on Optomechatronic Technologies (ISOT)*, Seattle, USA, November 5-7, 2014.

- **Y. Gong**, T. D. Soper, V. W. Hou, D. Hu, B. Hannaford, and E. J. Seibel, "Mapping surgical fields by moving a laser-scanning multimodal scope attached to a robot arm," *Image-Guided Procedures, Robotic Interventions, and Modeling, SPIE Medical Imaging, Proc SPIE* vol. 9036, paper 90362S-1-8 2014.

- X. Ye, **Y. Gong** and J. Yoon, "In Vitro results of robotic cystoscope and 3D panorama for automated bladder surveillance system," *29th EUS conference*, Orlando, FL, May 17, 2014.

- **Y. Gong**, T. D. Soper, V. W. Hou, D. Hu, B. Hannaford, and E. J. Seibel, "Toward Intraoperative 3D Rendering of Surgical Field with a Laser-scanning Multimodal Scope," *National Robotics Initiative Principal Investigators (NRI PI) meeting*, Washington D.C. Oct 1-2, 2013.

- S. Zhang, Y. Wang, L. Ekstrand, Y. Xu, and **Y. Gong**, "Superfast 3-d shape measurement with binary defocusing techniques," *in International Conference on Computational & Experimental Engineering and Sciences (ICCES)*, 2011. (Invited)

- S. Zhang, **Y. Gong**, Y. Wang, J. Laughner, and I. R. Efimov, "Some recent advance on high-speed, high-resolution 3-D shape measurement using projector defocusing," *International Symposium on Optomechatronics Technologies (ISOT)*, Toronto, Canada, Oct 25-27, 2010.

- J. Laughner, **Y. Gong**, B. Filas, S. Zhang, I. R. Efimov, "Structured light imaging of epicardial mechanics," *EMBS, IEEE International Conference (EMBS10)*. Buenos Aires, Argentina,

Aug 31 - Sep 4, 2010

- **Y. Gong** and S. Zhang, "Improve 4-D shape measurement speed by using projector defocusing," *SPIE Optics & Photonics*, San Diego, CA Aug 1-5, 2010.

HONORS AND AWARDS

- GPSS Travel Grant, University of Washington, 2015
- Mechanical Engineering Fellowship, University of Washington, 2015
- Outstanding paper (#4) in 29th Engineering & Urology Society (EUS) conference, 2014
- Cover paper of Optics Express, 2010