2-26-2018 3:00 PM

Ultrasound-Augmented Laparoscopy

Uditha Lakmal Jayarathne
The University of Western Ontario

Supervisor
Terry M. Peters
The University of Western Ontario

Graduate Program in Biomedical Engineering
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

Abstract

Laparoscopic surgery is perhaps the most common minimally invasive procedure for many diseases in the abdomen and thorax. Since the laparoscopic camera provides only the surface view of the internal organs, in many procedures, surgeons use laparoscopic ultrasound (LUS) to visualize deep-seated surgical targets. Conventionally, the 2D LUS image is visualized in a display spatially separate from that which displays the laparoscopic video. Therefore, reasoning about the geometry of hidden targets requires mentally solving the spatial alignment and resolving the modality differences, which are cognitively very challenging. Moreover, the mental representation of hidden targets in space acquired through such cognitive mediation may be error prone, and cause incorrect actions to be performed.

To remedy this, advanced visualization strategies are required where the US information is visualized in the context of the laparoscopic video. To realize such visualization schemes, efficient computational methods are required (i) to accurately align the US image coordinate system with that centred in the camera, (ii) to accurately represent 3D information of hidden targets from a series of 2D US images, and (iii) to blend the US information with the surface image provided by the camera such that the surgeons perceive the geometry of hidden targets accurately.

In this thesis, a complete pipeline is described to visualize hidden targets, imaged by a LUS probe, in 3D in the context of the laparoscopic video. A novel method to register US images with a coordinate system centred in the camera is detailed with an experimental investigation into accuracy bounds in representing an imaged target in this coordinate system. This method eliminates the requirement for extrinsic tracking devices in the operating room (OR), significantly reducing both the financial and logistical overhead. An improved method to blend US information with the surface view provided by the camera is also presented with an experimental investigation into the accuracy of perception of the target locations in space.

The work presented here, together with concurrent development in related fields, will enable image-guidance in laparoscopic soft-tissue surgery. The suggested improvements will

increase both the efficiency and the safety of many minimally invasive abdominal procedures.

Keywords: Laparoscopic Surgery, Laparoscopic Ultrasound, Visualization, Pose Estimation, Ultrasound Reconstruction, Direct Volume Rendering, Psychophysical Evaluation

Acknowledgements

During the past six years, from the day I started my graduate work at the Robarts Research Institute to the day I finished compiling this thesis, I received enormous support from a large group of people. Many of them contributed by providing financial support, involving in discussions, or by arranging travel to various destinations throughout the world.

I would like to start by acknowledging my supervisor, Dr. Terry Peters, for giving me the opportunity to pursue my graduate studies in his work-class research laboratory. He encourages every student to think outside the box and solve clinically impactful problems. Terry makes sure that we have access to the best research facilities required to undertake world class research. I appreciate Terry's trust in me to represent his lab in many international conferences, and I will never forget the education I received from him. In no specific order, I would also like to thank Dr. Elvis Chen, John Moore, Dr. Ali Khan, and Dr. Xiongbiao Luo, Terry's research staff, for their support, either in the development of software tools, setting up experiments, or lengthy hours of theoretical discussions. Special thanks should go to Dr. Roy Eagleson, Dr. Hanif Ladak, and Dr. Stephen Pautler for steering me in the right direction being in my advisory committee. I also express my gratitude to Christine Ellwood, and the administrative staff at the Robarts Research Institute for their assistance in all the paperwork.

Since I first came to London, many people helped me with many logistical issues. My sincere gratitude goes to Janette Wallace, Jackie Williams, and many people in the Sri Lankan community living in London for making this city feels like home. Also, special recognition is given to my academic siblings Dr. Kamyar Abhari, Dr. John Baxter, Dr. Jonathan McLeod, (Dr.-to-be) Golafsoun Ameri, and Adam Ranking for their enormous support and many hours of useful discussions.

Finally, I would like to thank my family. Even though I am thousands of miles away from home, endless love and support from my parents, my sister and brother always helped me to stay focussed in my academic goals even under most stressful situations. Most of all, I thank my beautiful wife, Yashmi, for her endless love. Her support in realizing my academic dreams was tremendous. She was always by my side in many difficult situations, and I will never forget her commitment during the past three years of our married life.

This work was supported financially by the Canadian Institute of Health Research (CIHR), the Natural Science and Engineering Research Council (NSERC), the Canadian Foundation for Innovation (CFI), and the University of Western Ontario.

Dedicated to my parents, Yashmi, and specially to our future ...

"Believe nothing, unless it agrees with your own reason"
- Lord Buddha

Contents

List of Figures

List of Tables

List of Abbreviations

2D	Two Dimensional (typicaly (x,y) domain)
3D	Three Dimension (typically (x,y,z) domain)
DoF	Degrees of Freedom
MIS	Minimally Invasive Surgery
US	Ultrasound
IGS	Image Guided Surgey
CT	Computed Tomography
MRI	Magnetic Resonance Imaging
OR	Operating Room
DVR	Direct Volume Rendering
HMD	Head Mounted Display
SPECT	Single-photon Emision Tomography
CCD	Charge-coupled Device
HD	High Definition
LUS	Laparoscopic Ultrasound
VATS	Video-assisted Thoracic Surgery
PnP	Perspective-n-Point
ICP	Iterative Closest Point
FLE	Fiducial Localization Error
EKF	Extended Kalman Filter
GMM	Gaussian Mixture Model
UT	Unscented Transform
SRUKF	Square-Root Kalman Filter
RMSE	Room Mean Squared Error
TRE	Target Registration Error
VTK	Visualization Took Kit

Chapter 1

Introduction

Today, surgery is the only curative treatment for many life threatening diseases. During a surgical procedure, the surgeon navigates to the proximity of the diseased region in an organ, and executes a curative surgical action (excision, suturing, thermally destroying diseased tissue etc.). To help access the target region manually, visualize the pathology with the surgeons' direct vision and apply the curative treatment, traditionally, surgeries are performed with large incisions, hence the term open-surgery. These incisions not only cause significant hemorrhage, pain, and increased healing time, they elevate the risk of post-operative complications like infections and hernia, which may require secondary interventions.

To minimize undesired side-effects in open surgery, minimally invasive surgical (MIS) approaches were introduced over time to treat many diseases in the brain, heart, and the abdomen. In contrast to the large incisions in open surgery, the MIS procedures are performed through small incisions that are optimal for the surgical task. Therefore, hemorrhage, risk for infections and post-operative complications can be drastically reduced. Even though MIS approaches bring significant benefits to the patient, they require advanced surgical skills to be performed safely: Small incisions obstruct the surgeons direct vision while the direct access to the pathology is restricted. Therefore, the surgeons require reliable, indirect means to visualize, and access the organs of interest. Traditionally, real-time imaging modalities such as optical, ultrasound (US), and fluoroscopy replace the surgeon's direct vision while catheters or miniature instruments that are manipulated from outside the patient's body are used to perform surgical actions indirectly. For instance, in minimally invasive surgery to repair aotic/mitral value in the heart, fluoroscopy and transesophageal echocardiography, a real-time US stream obtained from a probe inserted through the esophagus, provides real-time imaging of the surgical site. Surgical actions are performed at the distal end of a catheter [1], or miniaturized special instrument [2]. Conventionally, the real-time images are viewed in a display placed at a distance from the patient. Therefore, additional cognitive efforts are necessary to fuse information from the images, and coordinate the required action at the distal ends of the tools. In addition, images provided by some modalities are difficult to interpret. For example, US images often contain speckle noise, and are very difficult to interpret without *a-priori* knowledge about the imaged anatomy. Fluoroscopy lacks soft-tissue contrast, hence, *a-priori* knowledge may be necessary to prevent accidentally hitting a critical structure. Thus, fusing anatomical knowledge with the images to interpret them, relating the images to the patient's anatomy, and coordinating the actions at the distal tips of the instruments is an essential skill the surgeons need to master in order to safely perform surgeries using MIS approaches.

1.1 Image-guided Minimally Invasive Surgery

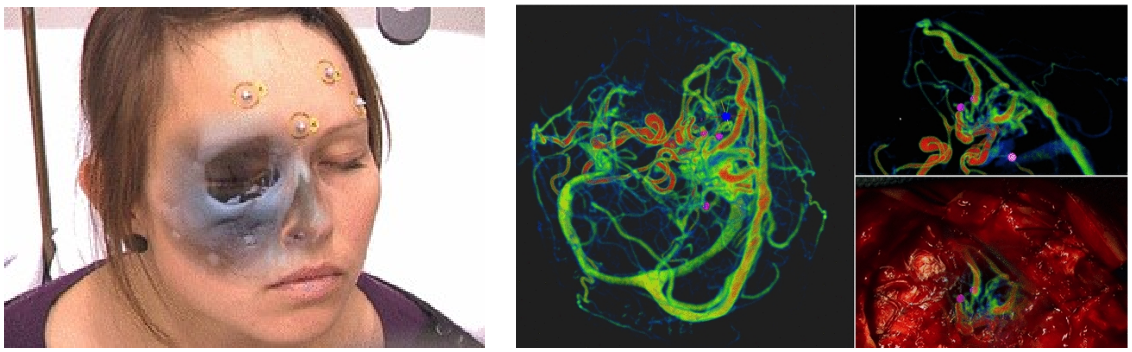
To make the surgery safe by minimizing the required cognitive efforts, image-guided surgical (IGS) system were introduced to MIS approaches. These systems employ technologies to track the position and orientation of surgical tools and imaging devices relative to the patient, allowing tools and images to be registered to the patient, and displayed in a common frame of reference. Since the images and the tools are registered to the patient, surgeons can visualize where they are in the image space while they operate. In addition to real-time imaging modalities, pre-operative images such as x-ray computed tomography (CT) and magnetic resonance imaging (MRI) can also be registered to the patient either using external fiducials, or using anatomical landmarks [3]. In this manner, pre-operative plans may be brought into the surgical suite enabling easy translation of the plans into actual surgery. With pre-operative images, surgical tools, and other imaging sources registered with the patient, surgeons now have wealth of information to compensate for the lost direct vision in MIS.

IGS systems have improved the safety of many minimally invasive procedures. They have become an essential part of the operating room (OR), particularly in neuro and orthopaedic surgeries. The relatively static, rigid nature of the anatomical regions of interest makes the registration processes simple and easy, which is one of the reasons behind wide adaptation of these systems by neuro and orthopaedic surgical communities. When the environment is deformable, the registrations can be updated by registering the intra-operative images with the pre-operative images using deformable registration methods [4]. However, such techniques are typically computationally expensive, and hence may not be suitable for realtime applications. Nevertheless, several attempts to extend the use of IGS systems to highly deformable, dynamic tissue environments can be found in the literature [5, 6].

1.2 Visualization in Image-guided Surgery

Visualizing multiple images registered to a common coordinate system is a challenging task. To avoid unnecessary visual clutter in the visual field that may degrade the surgeon's performance, only the most relevant information should be presented to the surgeon at the appropriate time during the procedure. To mitigate this situation, advanced visualization techniques have been developed as outlined by Seilhorst et al. [7], making IGS more effective and safe.

Most IGS systems use naïve rendering methods to visualize registered pre-operative and intra-operative image together with 3D models of the surgical tools in a common coordinate system [2, 10, 11]. Typically such images are visualized in a monoscopic display placed 1-2 feet away from the patient. Volumetric images are rendered in this coordinate system using surface rendering [12], or direct volume rendering (DVR) methods [13]. Rather than rendering the



(a) (b)

Figure 1.1: (a) A CT image registered to the patient is rendered in a video see-through HMD [8] © IEEE 2007, (b) Volume rendering of blood vessels visualized in the context of an image captured by a camera affixed to the surgical microscope in brain surgery [9] © Springer 2015

whole volume, some systems employ manual, interactive or fully-automatic [14] segmentation methods to extract a regions of interest, and render. Several attempts to render information in stereoscopic displays can be found in the literature [15], with the intention of improving depth perception, and thereby improving the accuracy of the guided surgical task. A major drawback in these systems is that significant cognitive efforts are necessary to relate the visualization to the action site. The cognitive processes involved in this may result in erroneous surgical actions to be performed that may end up in undesired consequences.

To minimize errors due to dissociated perception from action, images registered to the patient are displayed at the correct spatial location through head-mounted displays (HMD) (Fig. 1.1(a)) [8, 16–18]. Since the images are rendered in the patient frame of reference, the perceptual system is coupled with the action frame of reference, making the actions more intuitive, hence improvement in sugical performance can be expected. In neurosurgery, the surgical microscope itself can be modified to display rendered images in the context of the real anatomy [9, 19–21] eliminating the need for an HMD (Fig. 1.1(b)). Whichever display technology used, the virtual objects should be blended properly with the real scene to guarantee accurate perception of the location of the rendered targets. As discussed in section 1.8, inaccurate mental representations acquired from the rendered targets may result in undesirable, life threatening consequences.

1.3 Assessment of IGS Systems

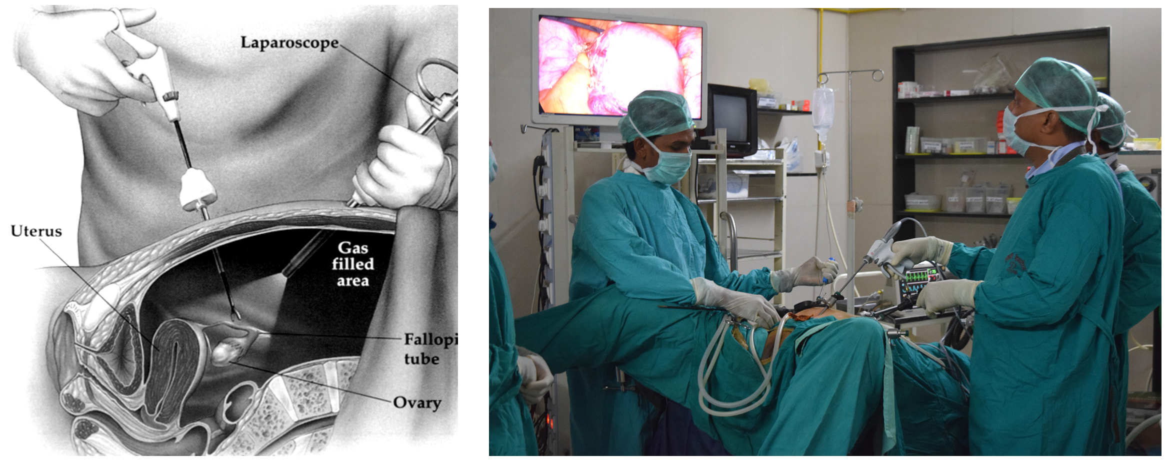
The performance of an IGS system depends not only on the technical aspects of the system, but also on how the surgeon interacts with the system and the patient. Therefore, assessing an IGS system performance is a complex procedure. Jannin and Korb [22] define six levels of IGS system assessment depending on the assessed property of the system. At the lowest level, technical parameters of the system such as the accuracy, precision, latency etc. are assessed under laboratory conditions. At level two, the therapeutic/diagnostic reliability of the system is evaluated under simulated laboratory environments such as phantom studies. At level three, surgical performance is assessed to determine the efficacy of the system in the clinic. Levels 4-6 assess patient outcomes, economic aspects, and social, legal, and ethical aspects based on data gathered routinely in the clinic at multiple centers. Thus, thorough evaluation of IGS systems at all these levels requires studies over a lengthy period involving patients at multiple clinical facilities.

Before an IGS technology is tested for performance in surgery involving patients, its technical parameters and reliability should be evaluated under laboratory conditions. Technical parameters of the system can be assessed objectively fairly easily using highly accurate measuring systems. However, assessment of the reliability of a system is a complex process since the experiments should have control over fairly large parameter space. To maintain high degree of control over the experiment, the surgical scenario may be simulated at a cost of losing realism. On the other hand, experiments can be conducted in more realistic surgical setting, but at a cost of loosing the control over the experiment. It is important to assess the system under different testing environments to gain a thorough understanding about the system.

Quantifying complex human-machine interactions is another complication in level 2 IGS system assessment. Phantom-based psychophysical studies can be designed to objectively study human factors, but the results may vary among different subject groups. For instance, experienced surgeons may be better/worse at certain surgical tasks compared to resident surgeons, and produce completely different results. Therefore, subject demographics should be taken into consideration during experimental design. There are certain human factors that cannot be quantitatively measured. Subjective assessments using a ranking system such as the NASA Task Load Index (NASA TLX) [23] may help in these cases.

1.4 Image Guidance in Soft-tissue Surgery

Unlike in neurosurgery and orthopaedic surgery where the surgical environment is relatively static, image-guidance in soft-tissue surgery is very challenging. Due to the highly deformable and dynamic nature of the environment, pre-operative images registered to the patient's anatomy, do not provide reliable means to localize, and target hidden surgical structures. As an alternative, surgeons use complimentary intra-operative imaging modalities (optical, ultrasound, intraoperative single-photon emission computed tomography (SPECT) imaging with miniature gamma probes[24]) to help them acquire a detailed understanding about the geometry (spatial location, 3D form etc.) of the surgical targets and the surrounding. Such a detailed understanding of the surgical environment is crucial for accurate execution of therapeutic actions. However, the conventional method of visualizing these intra-operative imaging modalities on separate displays is far from intuitive and may result in erroneous mental representations of the task. As a result, surgical errors may occur that have life-threatening consequences. This dissertation attempts to identify technical limitations in fusing two real-time imaging modalities, namely intra-operative US and optical imaging, and propose clinically viable solutions that have the potential to improve surgical performance. While it focuses mainly on laparoscopic interventions, the concepts developed can be easily adapted to any intervention involving ultrasound and optical imaging.



(a) (b)

Figure 1.2: (a) a cross section of the abdomen during a laparoscopic intervention. The surgeon holds the laparoscopic camera while he/she operates on the patient using a laparoscopic instrument. Typically the abdominal cavity is filled with CO_2 to increase the working space, (b) conventionally, the laparoscopic video, monocular in this case, is displayed in a monitor that is placed away from the patient. The surgeon has to operate on the patient while his eyes are focused on the display, thus decoupling his/her action from the perception.

1.5 Laparoscopic Surgery

Since its first use in humans over a century ago [25], laparoscopy has received enormous attention as an effective minimally invasive surgical approach. Unlike open-surgery, where the surgeon accesses the organ of interest through a wide incision made on the patients' abdomen or thorax, in laparoscopy, the surgery is performed through small incisions typically on the order of $5 - 15$mm across. These incisions, known as ports, provide restricted access to the patient's body cavity. While a fiber-optic camera with a light source, inserted through one of the ports, provides visualization of the surgical site, the surgery is performed through long, slender instruments (Fig. 1.2(a)). The video from the camera is conventionally displayed in a monitor placed in front of the surgeon, but $1 - 2$ meters away from the patient(Fig. 1.2(b)). This video is typically monoscopic, but stereoscopic cameras and display systems have been introduced, and are used in a few surgical centers worldwide. Prior to the invention of the CCD chip, the analog video captured by the camera was displayed on an analog monitor. However, with the advancement of video technology, most ORs today are equipped with digital laparoscopic cameras and high definition (HD) video displays. This not only improves the visual fidelity of the captured anatomy, but also allows manipulation, enhancement, and storage of the captured videos for improved surgical guidance, teaching and training purposes.

Compared to the open surgical approach for many procedures [26–29], laparoscopy offers significantly lower pain, haemorrhage, and fewer post-operative complications, but with comparable clinical outcomes to conventional surgical procedures. Despite these advantages however, it presents several significant challenges to the surgeon. The laparoscopic video, displayed in a monitor remote from the patient, dissociates surgeon's perception from his actions. In addition, the monoscopic display used even in modern ORs today, offers limited depth perception. As a result, surgeons must master skills to infer depth from monocular depth cues, and execute actions at a location that is spatially disassociated from their perception. Moreover, surgical actions performed with long, slender instruments not only limit the range of motion, but also requires the mastery of difficult motor skills to compensate for the *fulcrum effect*[30]. These instruments significantly reduce tactile sensation, making it very difficult to determine the forces exerted at the tip of the instrument accurately. Nevertheless following years of training, surgeons learn to operate under these conditions and perform complex procedures on a daily basis.

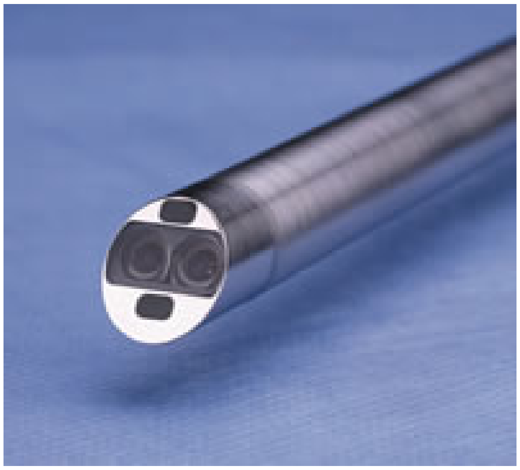
1.6 Robot-Assisted Laparoscopic Surgery

Robot-assisted laparoscopic surgical systems such as the *daVinci* (Intuitive Surgic Inc.) robotic system (Fig. 1.3) have been introduced with the intention of eliminating ergonomical issues with conventional laparoscopy. These systems replace the monoscopic laparoscopic camera with a stereoscopic one (Fig. 1.4(a)), and display the captured video in a stereoscopic display at the surgeon's console(Fig. 1.4(b)) where the surgeon sits comfortably rather than standing at the patient at an awkward posture. This allows the surgeon to perform very long procedures

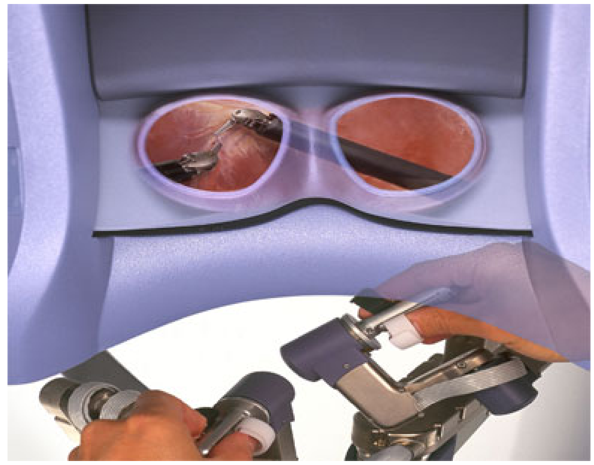
Figure 1.3: *daVinci* surgical robotic system. The surgeon sits at the console where his gestures are captured and converted to commands that control the slave-robot at the patient site. At the console the surgeon gets stereoscopic visual feedback. Typically these system have four robotic arms: one to hold the stereoscopic laparoscopic camera, two arms controlling surgical instruments, and one auxiliary arm. Image courtesy of Intuitive Surgical Inc.

without fatigue impairing his/her performance. The surgeon's hand gestures are captured by the mechanical gesture tracking system (Fig. 1.4(b)) at the surgeon's console, allowing the manipulators at the distal ends of the robotic arms to be driven interactively. Thus, these systems effectively replicate the surgeon's gestures at the wristed laparoscopic instruments with the ability to scale the magnitude of their motion as desired.

One major advantage the robotic surgical systems offer over the conventional laparoscopic techniques is the improved dexterous manipulation of the end-effectors. This allows the surgeons to operate inside highly constrained spaces, such as the pelvic cavity, just as intuitively as they would perform in an open surgery with their hands inside the patients body. In addition, the flexible surgical tools controlled by sophisticated algorithms in some of these systems (e.g. SinglePort by Intuitive Surgical, SPORT system by Titan medical, Surgibot by TransEntrix) allow the surgeon to reach locations inside the patient's body through a single incision and with minimal damage to the surrounding anatomy[31–33]. The high definition stereoscopic vision system in these tele-operated systems enable improved depth perception, permitting complex



(a)



(b)

Figure 1.4: (a) stereoscopic laparoscopic camera used in the *daVinci* surgical system. It has a baseline of 5-6*mm*. A fiber-optic light source illuminates the surgical site, (b) stereoscopic display and the gesture capturing hardware at the surgeon's console of a *daVinci* surgical system.

surgical tasks to be performed with significantly fewer errors compared to the performance with monoscopic viewing[34]. Thus, these systems assist both experienced and novice surgeons to perform otherwise very difficult surgeries safely and efficiently.

1.7 Intra-operative Ultrasound in Laparoscopic Surgery

The laparoscopic camera can capture only the surface view of the internal organs. When the surgeons need to visualize hidden critical structures, they often use laparoscopic ultrasound (LUS), which employs an ultrasound probe designed specifically for laparoscopic applications. These probes are usually cylindrical in shape, and have a linear or a curvilinear array of transducers at the distal end (Fig.1.5(a)). The transducer arrays are typically one dimensional, and provide a 2D cross-sectional image. Although several attempts to provide real-time volumetric imaging with 2D transducer arrays can be found in the literature[35], these technologies have not yet made their way to the OR. The distal end of these probes can be rigid, or articulated manually using the controllers at the handle (Fig. 1.5(a)). Clinically, probes with articulated, rather than rigid, tips are preferred since they offer better maneuverability. Recently, miniature

(a) (b)

Figure 1.5: (a) laparoscopic ultrasound probe with an articulated tip. The linear transducer array provides a 2D US image while the manual controller at the handle enables articulation of the tip to reach otherwise difficult space, (b) drop-in (or pick-up) US probe with a curvilinear transducer array, designed specifically for robot-assisted surgical procedures. The probe has a grooved ridge that fits a laparoscopic/robotic grasper, enabling it to be picked up and manipulated. Image courtesy of BK Medical Systems Inc.

ultrasound probes that can be picked up and manipulated by a grasper tool were introduced with application to robot-assisted surgery (Fig. 1.5(b)). Since these probes provide improved maneuverability and surgeon autonomy, they are preferred over conventional versions in many robot-assisted surgical procedures[36].

The 2D ultrasound image captured by a laparoscopic ultrasound probe is typically visualized in a display separate to that displaying the laparoscopic video. During conventional laparoscopic surgery, this second display is usually on the ultrasound machine itself, while in robot-assisted surgery with the *daVinci* system the image is presented in a separate display panel known as the TilePro™ (Fig. 1.6). In either case, the ultrasound image and the laparoscopic video are spatially dissociated. In addition, these two modalities provide two distinct types of information: laparoscopy, being a projective imaging modality, provides a perspective projection of the scene, while the ultrasound image provides a tomographic view into the tissue. Therefore, reasoning about the geometry of the hidden surgical targets requires mentally solving the spatial alignment problem and the resolving the modality differences which are cognitively very challenging[37, 38].

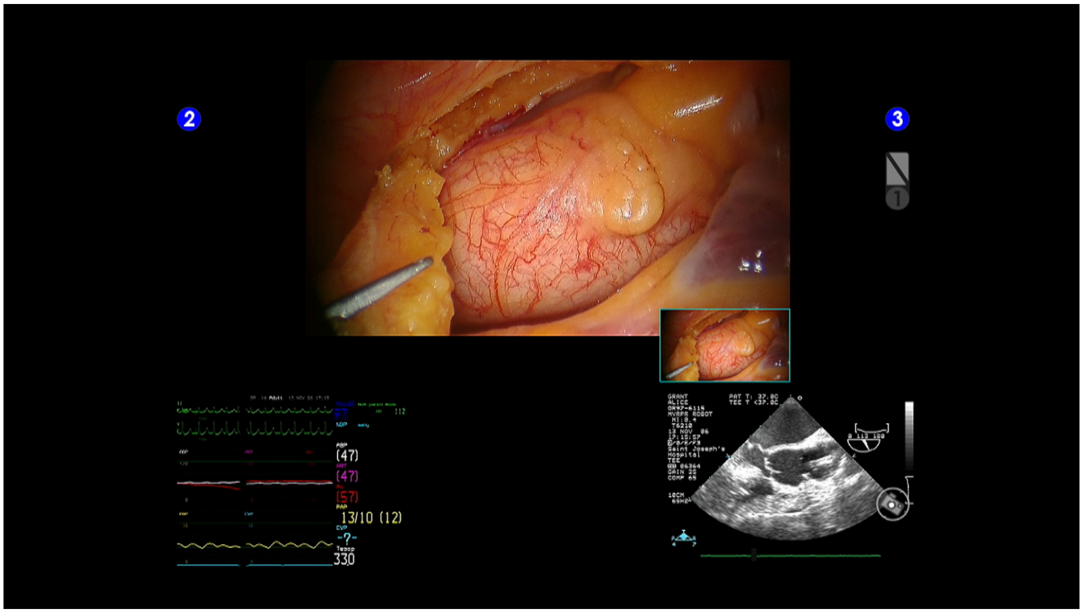


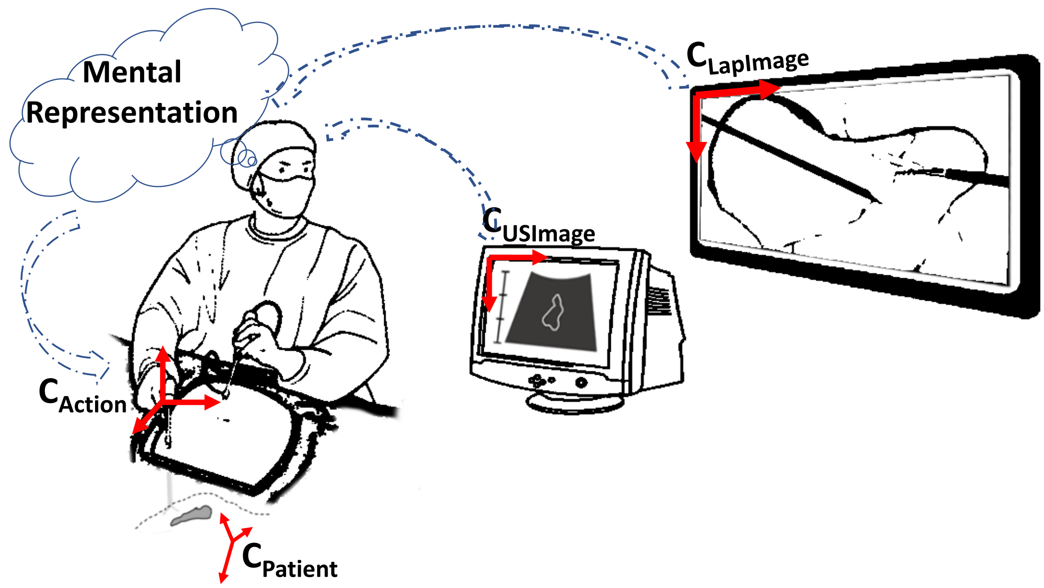
Figure 1.6: The TilePro™ display (bottom) in *daVinci* surgeon's console, with the primary display providing the camera view (top). This secondary display allows the surgeons to visualize pre-operative/intra-operative images, or physiological measurements during an intervention. Image courtesy of Intuitive Surgical Inc.

Planning and execution of many critical surgical tasks benefit from the use of ultrasound during laparoscopic interventions. In laparoscopic and robot-assisted partial nephrectomy, tumor margins and resection plans for both *endophytic* tumours (that have grown into the tissue) and *exophytic* tumours (that have grown outward beyond the organ surface) can be determined with a higher degree of certainty with laparoscopic ultrasound [33]. Moreover, ultrasound helps to improve the understanding of the adjacent structures that should be avoided to prevent complications [39]. During tumor resection tasks for small nodules in the lung with the minimally invasive video assisted thoracic surgical (VATS) approach, the use of intraoperative ultrasound has the potential to improve the localization of small, non-visible and non-palpable nodules[40]. The use of US also significantly reduces the risk of conversion of VATS procedures into more invasive thoracotomies[41]. In laparoscopic cholecystectomy, the use of laparoscopic ultrasound has been shown to be an effective means of delineating the bile-duct anatomy during difficult situations[42] while damage to the bile-duct is significantly reduced[43]. Even though the simultaneous use of these two modalities enable numerous advantages, the conventional method of mental fusion of information may result in surgical errors, due to the error-prone cognitive processes involved. In oncologic surgery, these errors could occur in the form of a positive resection margin, or hitting a major surgical structure, in laparoscopic cholecystectomy, bile-duct damage, or severe bleeding. Typically, these intraoperative complications are life threatening.

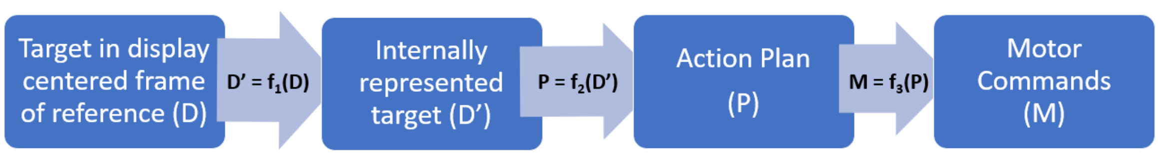
1.8 Cognitive Processes in Ultrasound-guided Action

A human reaching action guided by the visual system can be described by a sequence of transformations [44] (Fig. 1.7(b)): (1) the transformation that construct a mental representation of the target in space with visual inputs, (2) the transformation that produces an action plan with specific movement parameters based on the mental representation, and (3) the transformation that converts the action plan to motor commands to execute the action. Under transformation (1), image-centered target representation D is transformed to an internal spatial representation D' with respect to the perceptual frame of reference. For actions guided by the direct vision, this transformation causes no alternation to the input information because the target is already in the perceptual frame of reference. This internal representation is then mapped to the action centered frame of reference, and an action plan consisting of specific parameters is determined under transformation (2). The mapping (3) uses these parameters, and transforms them to specific motor commands that execute the specified action. Any of these mappings is subject to error while accumulation of error across various stages is also a possibility.

In laparoscopic surgery, the surgeons do not have direct visualization of the action site; the laparoscopic camera captures the surgical scene and displays a magnified view on a monitor that is displaced from the action site. Thus, the reference frame for surgical actions is no longer coupled to the reference frame for perception. In addition, the action site is only accessible through long slender tools, or through robotic instruments, hence the hand-centered frame of



(a)



(b)

Figure 1.7: (a) Schematic of laparoscopic surgery where laparoscopic ultrasound is used to visualize subcutaneous structures: An internal representation is constructed from the image information read from the two images, based on which surgical actions are planned and executed, (b) Cognitive transformations in visually-guided action: Target represented in image reference frame is transformed to an internal representation. Based on this representation, an action plan is derived based on which appropriate motor commands are derived

reference, where action is naturally centered, is decoupled from end-effector centered frame of reference (Fig. 1.7(a)). However, the visual feedback of the movement of the end-effector allows learning of the mapping from the hand-centered reference frame to the end effector reference frame, yet any change in the camera pose necessitates a re-learning. According to perceptual-motor learning theory, such learning, at least at their early stages, requires cognitive mediation [45, 46], and the performance may vary depending on the level of the surgeons experience [47].

When a surgeon uses an US image visualized on a monitor that is spatially dissociated from that displays the laparoscopic video to guide a surgical action (Fig. 1.7(a)), additional mental transformations are required to construct an accurate mental representation of the target in the perceptual reference frame. This mental transformation involves at least mental translation, and mental scaling to account for the scaling difference between two displays. In practice, however, mental rotations are always involved, requiring significant taxing cognitive resources, and can easily cause excessive cognitive load [48, 49]. Given limited cognitive resources, as well as timing constrains associated with certain surgical actions [50], these cognitive processes may be vulnerable to error. Such errors directly affect the accuracy of the mental representations, and may propagate to the action plan, and eventually to the surgical action. Erroneous surgical actions could have life threatening consequences, hence avoiding them is of utmost importance.

In addition to the errors in the mental representation, errors in the other parts of the transformation chain depicted in (Fig. 1.7(a)) could lead to surgical errors as well. A major source of error could be the transformation from the hand-centered frame of reference to the end-effector centered frame of reference. Similar to the cognitively mediated spatial representation of the targets, the learning of this mapping requires access to limited cognitive resources. Their scarcity, and time constraints may result in learning an erroneous mapping that may lead to undesirable consequences. However, with extensive training and with the ergonomically designed surgical robotic systems, the occurrence of such error can be reduced drastically.

1.9 Perceiving 3D Form from Cross-sectional Images

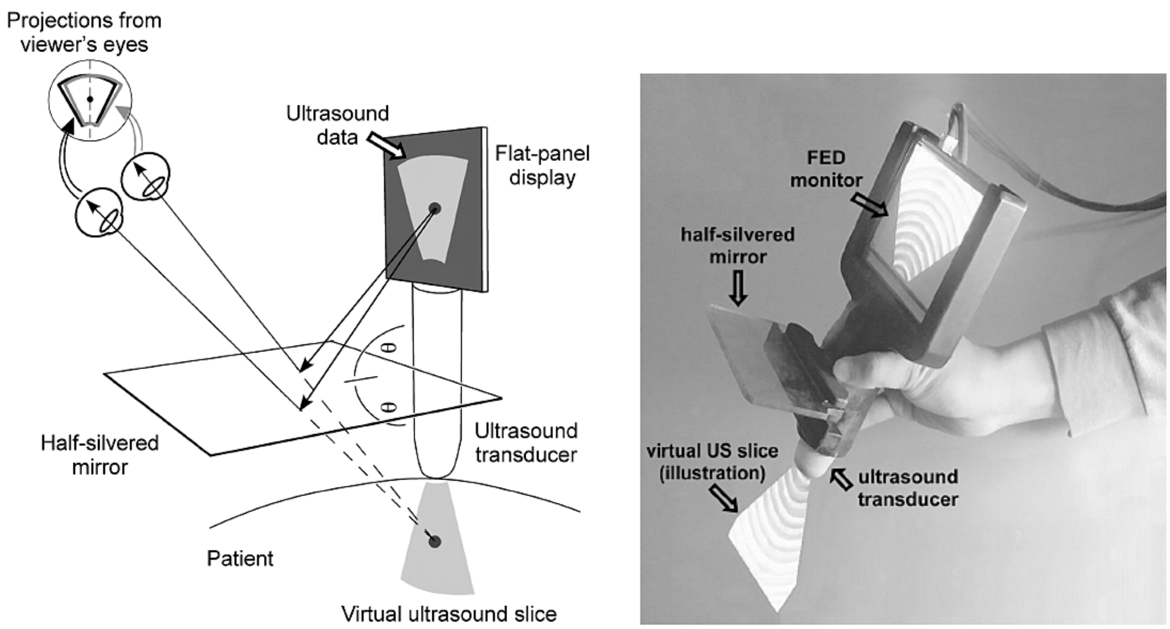
The success of many laparoscopic tasks guided by laparoscopic US depends not only on the accurate understanding of where the target is located in space, but also on the understanding of the 3D form of the target and its surrounding. Prior to the intervention, the surgeon typically constructs a mental representation of the 3D form of the targets from pre-operative CT or MRI images. These images are either visualized by conventional methods such as scrolling through a stack of tomographic images, or by using volume visualization method such as ray tracing [51]. However, due to the highly deformable nature of the organs (e.g. kidney, liver, lung etc.) and the surrounding soft-tissue environment in the thoraco-abdominal cavity, this pre-operatively constructed mental representation may not accurately represent the reality during a laparoscopic surgery.

An intraoperative imaging modality such as US can help acquire an updated representation of the 3D form of the targets. Since the current technology, at least as far as it is employed in laparoscopy, is limited to 1D-array transducers that produce 2D US images, revealing the 3D geometry of hidden targets in laparoscopy requires moving the probe over the region of interest and viewing the 2D images that vary with the probe motion. From the directly perceived 2D images, the surgeon constructs a mental 3D representation of the target and its surroundings. This process of visualizing 3D objects from a series of 2D cross-sectional images is analogous to the aperture viewing phenomenon known as *anorthoscopic perception* that has been widely studied in cognitive science[52]. During experiments designed to study the underlying cognitive processes involved in this phenomenon, subjects visualize a larger picture through a small slit that provides a limited field of view. Either the picture or the slit moves relative to the other, exposing parts of the picture to the subjects as it moves. By stitching the piecemeal images, the subjects can perceive the whole picture, possibly with minor distortions[53, 54]. Several hypotheses attempt to describe the underlying cognitive mechanisms in anorthoscopic perception. One of these - the post-retinal storage hypothesis[55], posits that the piecemeal information is stored in working memory and then combined to construct the whole image. This theory requires some means of localizing each visible aperture in a common spatial frame of reference[56]. In a configuration where the aperture moves relative to the static image, which is analogous to US probe moving over the target region of interest, the spatial location of successive visual inputs with respect to a common frame of reference is used. Kinesthetic cues may also be used if the observer controls the movement[57]. Once the visual apertures are localized, they are combined to construct a representation of the entire image.

When US, visualized in a separate display, is used to understand the 3D form of hidden targets during a laparoscopic procedure, the common frame of reference for localization is that centered on the laparoscopic display. Therefore, the image structures localized in the US display frame of reference must be mentally transformed to the frame of reference of the laparoscopic display. Such mental transformation, as described above, requires additional cognitive processes, and access to scarce cognitive resources. Surgeons may obtain very limited kinesthetic cues in a laparoscopic environment due to indirect manipulation of the probe. However, it is often the case that a surgical assistant manipulates the LUS probe while the surgeon controls other instruments. When this happens, no kinesthetic cues contribute in the localization of the US images, and the process depends entirely on the visual cues. Once the images are localized, the piecemeal representations require storage in the working memory before they are aggregated to construct a complete mental representation. However, the limited cognitive resources, and time constraints associated with certain laparoscopic surgical tasks may render these cognitive processes error-prone, resulting in erroneous mental representations. The accuracy of such cognitively mediated representations may also depend on the observers' spatial ability [58, 59]. To minimize these errors, and reduce the risks of erroneous surgical actions that have life threatening consequences, improvements in visualization methods for this type of image fusion are necessary.

1.10 In Situ Visualization of US

The conventional strategy of visualizing US information, known as *ex situ* visualization because the visualization occurs off site, invokes several cognitive processes to construct a mental representation of both the spatial location and the 3D form of a hidden target, as discussed in Sections 1.8 and 1.9. An alternative approach where the visualization happens at the imaging origin is known as *in situ* visualization. With this mode of visualization, expensive cognitive



(a) (b)

Figure 1.8: (a) schematic diagram of the SonicFlashlight device. The half-silvered mirror places a virtual US image at the correct spatial location allowing (in situ) visualization, (b) a realization of the schematic illustrated in (a). © 2005 IEEE

processes that transform information across different frames of reference are eliminated since the perception and action frames of reference are coupled. Perception-action coupling is very important for performing visually-guided action intuitively with a high degree of accuracy.

In laparoscopic surgery, perception and action are already disconnected. However, with extensive learning, the surgeon can maintain a good coupling between the visual perception and the surgical action at the end-effector of the laparoscopic tool. In robot-assisted surgery this coupling may be achieved relatively easily as a consequence of the ergonomic design of these systems. Assuming that a such coupling can be achieved, if US information can also be visualized in the frame of reference of the laparoscopic video display, US guided actions could be more accurate, since the mental representations in such a visualization strategy do not depend on cognitive processes that are vulnerable to error. Such a visualization could be termed *hybrid in situ* visualization.

Several studies have been conducted in the past to investigate the advantages of *in situ* US visualization. Wu et al. [60, 61] demonstrated that *in situ* visualization of US helps the observers localize completely hidden targets significantly more accurately compared to the *ex situ* visualization technique in their study using a hand-held apparatus called the Sonic Flashlight that allows the US image to be visualized at the imaged location (Fig. 1.8). Further experiments showed that the *in situ* visualization technique significantly improves the understanding of the 3D form of hidden objects compared to the *ex situ* visualization strategy[56, 62]. They attribute these improvements to the accurate internal representations the observers acquire through perception, in contrast to those constructed through cognitive mediation. These studies focused on non-laparoscopic use of US where the subjects held the probe manually while direct vision to the imaging/action site was allowed. In laparoscopic surgery however, since surgeons visualize the surgical site through a monocular/stereoscopic laparoscopic camera, their visual perception is drastically different from direct vision. Unlike in these studies, surgeons' perception is dissociated from their actions, introducing a new set of variables to consider during experimental design. In addition, in laparoscopic surgery, the surgeon may not be holding the probe himself, hence, unlike in these experiments, kinesthetic cues may not be available to help localize US images. Given these conditions in laparoscopy, further experiments may be required to investigate the the efficacy of *hybrid in situ* visualization of US during laparoscopic interventions.

1.11 Motivation: Ultrasound-augmented Laparoscopy

The hypothesis that the mental representations acquired through perception is more accurate than the cognitively mediated ones may apply to LUS as well. However, the efficiency of such representations in laparoscopy depends on two major factors: (1) the accuracy with which US information can be spatially registered with the laparoscopic video frame of reference, and (2) the method of presentation of the US information in the context of the laparoscopic video.

To display US image information in this manner to enable internal representations through

(a) (b) (c)

Figure 1.9: (a) laparoscopic camera image (monocular) showing the imaging tip of a LUS probe lying on an abdominal organ, (b) 2D US image for the probe depicted in (a), and (c) LUS image presented in the context of the laparoscopic camera image. Note that even the image is rendered at the correct scale and pose, it is perceived to float above the rest of the scene

perception, the transformation that maps each US pixel to the frame of reference centered in the laparoscopic camera should be first determined by some external means. In the conventional, *ex situ* visualization paradigm, this transformation is determined mentally with the involvement of expensive cognitive processes. Given the rigidity of the US transducer, this transform can be modeled by six degrees of freedom (6DoF; 3DoF in rotation, and 3DoF in translation) rigid body transformation in 3D space. One can easily estimate this transformation by employing an extrinsic tracking system (magnetic[63], optical[64], magneto-optic hybrid [65] etc.) to track both the camera and the probe in a common frame of reference. However, such extrinsic methods add financial and logistical overhead to the existing work-flow. In addition, the accuracy of some of these tracking systems may be affected by the presence of ferromagnetic materials in the OR. Therefore, an alternative means of tracking that is robust to the variables in a laparoscopic environment is required. Such a method should be able to compute the 6DoF transformation at least at the camera frame rate, which ranges from 24-60 frames per second (fps), to enable smooth operation. The method should be sufficiently accurate since it directly affects where the US image information is presented in the laparoscopic video frame of reference. Eventually, its accuracy will determine where the observer perceives the spatial location of a clinically important target.

Once registered to the laparoscopic video frame of reference, US image information should be presented to the observer in a manner that allows him/her to perceive the spatial location and the 3D form of a hidden target without ambiguity. A popular method used to achieve this is to texture map the US image to a 2D plane that is placed at the correct spatial position, orientation, and scale[63, 64, 66]. The laparoscopic camera image is set as the background texture to provide the context (Fig. 1.9(c)). This method of presentation conveys ambiguous depth cues resulting in perception that the US image is in front of the rest of the scene. Moreover, the interpretation of the content in the US image in this presentation becomes difficult at certain probe poses. In many laparoscopic applications, the camera is situated directly above the LUS probe due to the standard placement of laparoscopic ports. In such situations, the US image plane is nearly perpendicular to the camera imaging plane, making it difficult to interpret its content if this mode of presentation is used. Moreover, the overlay of a single image still requires cognitively involved mental integration of images for 3D form perception, even though *hybrid in situ* visualization helps in localizing the 2D US images in the frame of reference of the laparoscopic video. Reconstructing a 3D image by compounding 2D US images as the surgeon moves the probe over an organ, and presenting 3D US image information instead of 2D images, may be less cognitively taxing since it avoids mental integration. With an appropriate method of blending US information with the camera image, such a method may help surgeons perceive the spatial location as well as the 3D form of a hidden surgical target without cognitive mediation. Surgical actions that rely on mental representations acquired through such a visualization could be significantly less susceptible to error.

1.12 Hypothesis and Research Questions

Based on the primitive hypothesis that mental representations of a hidden target in space acquired through perception is more accurate than those that are cognitively mediated, a slightly different hypothesis can be derived with specific application to laparoscopic interventions;

compared to the conventional *ex situ* visualization, *hybrid in situ* visualization of US improves the surgeon's ability to perceive the spatial location and the 3D form of hidden surgical targets in laparoscopic interventions. However, as mentioned in the previous section, the improvements may depend on the accuracy of spatial registration of laparoscopic video and the LUS video, and the method of presentation of the registered LUS information. Given these factors, the following research questions arise, which should be adequately answered to guide the development of an effective *hybrid in situ* visualization strategy.

1. How do we spatially register LUS image the laparoscopic camera image with minimal overhead to the existing OR work-flow? What are the error-bounds in such a registration?

2. How do we reveal 3D information of hidden surgical targets from 2D US without cognitively involved visualization approach? What errors should we expect in such a computational approach?

3. How do we present spatially registered US information in the context of the laparoscopic video such that the surgeons perceive the spatial location and the 3D form without ambiguity?

The answers to these questions pave the path to an effective visualization pipeline with which we can test the above mentioned hypothesis that is specific to laparoscopic interventions. The testing of this secondary hypothesis may strengthen (or weaken) the primary hypothesis which it is based on, and will help us understand visually-guided human action better.

1.13 Thesis Outline

The mission of this thesis is to answer the research questions raised in Section 1.12, with the underlying hypothesis that *hybrid in situ* visualization of LUS improves the perception of the spatial location and the 3D form of hidden surgical targets. The search for answers to these research questions resulted in the development of a set of algorithms to spatially register LUS with the laparoscopic video stream, and effectively visualize LUS information in the context of the laparoscopic video, that are superior to the state-of-the-art. These algorithms are described in detail in Chapter 2 through 4. Moreover, in Chapter 4, a set of experiments is described that involve experienced US users, and attempts to test the hypothesis mentioned in Section 1.12. The results of these experiments supports the hypothesis, but further experiments may be required to prove their validity in the clinic.

1.13.1 Robust, Intrinsic Tracking of a Laparoscopic Ultrasound Probe

Chapter 2 describes a novel method to spatially register the LUS information with real-time laparoscopic video, that involves a special fiducial pattern attached to the curved back surface of the LUS probe. Compared to other methods based on extrinsic tracking methods, such as magnetic tracking, this method adds minimal overhead to the existing OR work-flow. The algorithm uses monocular laparoscopic images to estimate the 6DoF pose of the probe with respect to the camera, and by using priors on the pose, and on the probe motion, it runs in video frame-rates. It is robust to partial occlusion of the fiducial pattern, and demonstrates sub-milimeter target registration errors against an optical tracking-based reference. The chapter also describes how the proposed method can be extended to stereo/multi-view camera images with application to robot-assisted laparoscopic interventions.

1.13.2 Accuracy in Freehand 3D US Reconstruction with Robust Visual Tracking

In Chapter 3, computational methods to compound 2D US images into a 3D volume that are suitable to laparoscopic interventions is described. The GPU-implemented algorithm described in this chapter runs at very high frame-rates, making it suitable for *hybrid in situ* visualization applications. By imaging a tissue-mimicking phantom with structures resembling anatomical targets, it is demonstrated that the 3D US volumes resulted from this method are geometrically accurate. Both quantitative and qualitative results are presented.

1.13.3 Visualizing Ultrasound In the Context of Laparoscopy

The methods described in Chapters 2 and 3 allows 3D US information to be brought to the frame of reference of the laparoscopic video. Methods to present this information in the context of the laparoscopic video are detailed in chapter 4. The naïve method of blending US information with the surface view provided by the laparoscopic video results in ambiguous depth cues. To alleviate this issue, an intuitive technique is described in this section. The efficacy of the proposed method is determined in a series of experiments involving experienced US users in an environment that mimic a laparoscopic interventions. The results of the experiments reveal the strengths and weaknesses of the proposed method, and highlight the importance of incorporating depth cues in an ultrasound-augmented laparoscopic display.

Bibliography

[1] J. J. Gallagher, R. H. Svenson, J. H. Kasell, L. D. German, G. H. Bardy, A. Broughton, and G. Critelli, "Catheter technique for closed-chest ablation of the atrioventricular conduction system," *New England Journal of Medicine*, vol. 306, no. 4, pp. 194–200, 1982. PMID: 7054682.

[2] J. T. Moore, M. W. A. Chu, B. Kiaii, D. Bainbridge, G. Guiraudon, C. Wedlake, M. Currie, M. Rajchl, R. V. Patel, and T. M. Peters, "A navigation platform for guidance of beating heart transapical mitral valve repair," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 1034–1040, April 2013.

[3] R. R. Shamir, L. Joskowicz, S. Spektor, and Y. Shoshan, "Localization and registration accuracy in image guided neurosurgery: a clinical study," *International Journal of Computer Assisted Radiology and Surgery*, vol. 4, p. 45, Oct 2008.

[4] N. Hata, T. Dohi, S. Warfield, W. Wells, R. Kikinis, and F. A. Jolesz, "Multimodality deformable registration of pre- and intraoperative images for mri-guided brain surgery," in *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings* (W. M. Wells, A. Colchester, and S. Delp, eds.), pp. 1067–1074, Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.

[5] A. N. Sridhar, A. Hughes-Hallett, E. K. Mayer, P. J. Pratt, P. J. Edwards, G.-Z. Yang, A. W. Darzi, and J. A. Vale, "Image-guided robotic interventions for prostate cancer," *Nature Reviews Urology*, vol. 10, pp. 452–462, June 2013.

[6] C. Schneider, S. Thompson, J. Totz, Y. Song, K. Gurusamy, S. Ourselin, D. Stoyanov, M. Clarkson, D. Hawkes, and B. Davidson, "18. a pilot study evaluating the overlay display method for image guidance in laparoscopic liver surgery," *European Journal of Surgical Oncology (EJSO)*, vol. 42, no. 9, p. S72, 2016.

[7] T. Sielhorst, M. Feuerstein, and N. Navab, "Advanced medical displays: A literature review of augmented reality," *Journal of Display Technology*, vol. 4, pp. 451–467, December 2008.

[8] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab, "Contextual anatomic mimesis hybrid in-situ visualization method for improving multi-sensory depth perception in medical augmented reality," in *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 129–138, November 2007.

[9] M. Kersten-Oertel, I. Gerard, S. Drouin, K. Mok, D. Sirhan, D. S. Sinclair, and D. L. Collins, "Augmented reality in neurovascular surgery: feasibility and first uses in the operating room," *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, pp. 1823–1836, November 2015.

[10] J. Moore, C. Clarke, D. Bainbridge, C. Wedlake, A. Wiles, D. Pace, and T. Peters, "Image guidance for spinal facet injections using tracked ultrasound," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009: 12th International Conference, London, UK, September 20-24, 2009, Proceedings, Part I* (G.-Z. Yang, D. Hawkes, D. Rueckert, A. Noble, and C. Taylor, eds.), pp. 516–523, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009.

[11] G. Ameri, J. S. H. Baxter, D. Bainbridge, T. M. Peters, and E. C. S. Chen, "Mixed reality ultrasound guidance system: a case study in system development and a cautionary tale," *International Journal of Computer Assisted Radiology and Surgery*, August 2017.

[12] J. Beyer, M. Hadwiger, S. Wolfsberger, and K. Bhler, "High-quality multimodal volume rendering for preoperative planning of neurosurgical interventions," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, pp. 1696–1703, November 2007.

[13] I. Viola, A. Kanitsar, and M. E. Groller, "Importance-driven volume rendering," in *Proceedings of the Conference on Visualization '04*, VIS '04, (Washington, DC, USA), pp. 139–146, IEEE Computer Society, 2004.

[14] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation," *Annual Review of Biomedical Engineering*, vol. 2, no. 1, pp. 315–337, 2000. PMID: 11701515.

[15] T. A. N. Hernes, S. Ommedal, T. Lie, F. Lindseth, T. Langø, and G. Unsgaard, "Stereoscopic Navigation-Controlled Display of Preoperative MRI and Intraoperative 3D Ultrasound in Planning and Guidance of Neurosurgery: New Technology for Minimally Invasive Image-Guided Surgery Approaches," *min - Minimally Invasive Neurosurgery*, vol. 46, pp. 129–137, June 2003.

[16] H. Fuchs, M. A. Livingston, R. Raskar, D. Colucci, K. Keller, A. State, J. R. Crawford, P. Rademacher, S. H. Drake, and A. A. Meyer, "Augmented reality visualization for laparoscopic surgery," in *Medical Image Computing and Computer-Assisted Intervention — MICCAI'98: First International Conference Cambridge, MA, USA, October 11–13, 1998 Proceedings* (W. M. Wells, A. Colchester, and S. Delp, eds.), pp. 934–943, Berlin, Heidelberg: Springer Berlin Heidelberg, 1998.

[17] K. Abhari, J. S. H. Baxter, E. C. S. Chen, A. R. Khan, T. M. Peters, S. de Ribaupierre, and R. Eagleson, "Training for planning tumour resection: Augmented reality and human factors," *IEEE Transactions on Biomedical Engineering*, vol. 62, pp. 1466–1477, June 2015.

[18] I. Kuhlemann, M. Kleemann, P. Jauer, A. Aschweikard, and F. Ernst, "Towards x-ray free endovascular interventions using hololens for on-line holographic visualisation," *Healthcare Technology Letters*, vol. 4, pp. 184–187(3), October 2017.

[19] P. J. Edwards, A. P. King, C. R. Maurer, D. A. D. Cunha, D. J. Hawkes, D. L. G. Hill, R. P. Gaston, M. R. Fenlon, A. Jusczyzck, A. J. Strong, C. L. Chandler, and M. J. Gleeson, "Design and evaluation of a system for microscope-assisted guided interventions (magi)," *IEEE Transactions on Medical Imaging*, vol. 19, pp. 1082–1093, November 2000.

[20] M. Kersten-Oertel, I. Gerard, S. Drouin, K. Mok, D. Sirhan, D. Sinclair, and D. L. Collins, "Augmented reality in neurovascular surgery: First experiences," in *Augmented Environments for Computer-Assisted Interventions: 9th International Workshop, AE-CAI 2014, Held in Conjunction with MICCAI 2014, Boston, MA, USA, September 14, 2014. Proceedings* (C. A. Linte, Z. Yaniv, P. Fallavollita, P. Abolmaesumi, and D. R. Holmes, eds.), pp. 80–89, Cham: Springer International Publishing, 2014.

[21] M. Kersten-Oertel, I. J. Gerard, S. Drouin, K. Mok, D. Sirhan, D. S. Sinclair, and D. L. Collins, "Augmented reality for specific neurovascular surgical tasks," in *Augmented Environments for Computer-Assisted Interventions: 10th International Workshop, AE-CAI 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 9, 2015. Proceedings* (C. A. Linte, Z. Yaniv, and P. Fallavollita, eds.), pp. 92–103, Cham: Springer International Publishing, 2015.

[22] P. Jannin and W. Korb, "Assessment of image-guided interventions," in *Image-Guided Interventions: Technology and Applications* (T. Peters and K. Cleary, eds.), pp. 531–549, Boston, MA: Springer US, 2008.

[23] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Advances in Psy.*, pp. 139–183, 1988.

[24] B. Fuerst, J. Sprung, F. Pinto, B. Frisch, T. Wendler, H. Simon, L. Mengus, N. S. van den Berg, H. G. van der Poel, F. W. B. van Leeuwen, and N. Navab, "First Robotic SPECT for Minimally Invasive Sentinel Lymph Node Mapping," *IEEE Transactions on Medical Imaging*, vol. 35, pp. 830–838, March 2016.

[25] M. Hatzinger, S. T. Kwon, S. Langbein, S. Kamp, A. Häcker, and P. Alken, "Hans Christian Jacobaeus: Inventor of human laparoscopy and thoracoscopy.," *Journal of endourology*, vol. 20, pp. 848–50, November 2006.

[26] H. J. Bonjer, C. L. Deijen, G. A. Abis, M. A. Cuesta, M. H. G. M. Van Der Pas, E. S. M. De Lange-De Klerk, A. M. Lacy, W. A. Bemelman, J. Andersson, E. Angenete, J. Rosenberg, A. Fuerst, and E. Haglind, "A Randomized Trial of Laparoscopic versus Open Surgery for Rectal Cancer," *New England Journal of Medicine*, vol. 14372, no. 2, pp. 1324–32, 2015.

[27] I. S. Gill, L. R. Kavoussi, B. R. Lane, M. L. Blute, D. Babineau, J. R. Colombo, I. Frank, S. Permpongkosol, C. J. Weight, J. H. Kaouk, M. W. Kattan, and A. C. Novick, "Comparison of 1,800 Laparoscopic and Open Partial Nephrectomies for Single Renal Tumors," *The Journal of Urology*, vol. 178, no. 1, pp. 41–46, 2007.

[28] M. Lesurtel, D. Cherqui, A. Laurent, C. Tayar, and P. L. Fagniez, "Laparoscopic versus open left lateral hepatic lobectomy: a case-control study," *Journal of the American College of Surgeons*, vol. 196, no. 2, pp. 236–242, 2003.

[29] W. J. Scott, M. S. Allen, G. Darling, B. Meyers, P. A. Decker, J. B. Putnam, R. W. Mckenna, R. J. Landrenau, D. R. Jones, R. I. Inculet, and R. A. Malthaner, "Video-assisted thoracic surgery versus open lobectomy for lung cancer: A secondary analysis of data from the American College of Surgeons Oncology Group Z0030 randomized clinical trial," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 139, no. 4, pp. 976–983, 2010.

[30] A. G. Gallagher, N. McClure, J. McGuigan, K. Ritchie, and N. P. Sheehy, "An Ergonomic Analysis of the Fulcrum Effect in the Acquisition of Endoscopic Skills," *Endoscopy*, vol. 30, pp. 617–620, September 1998.

[31] J. H. Kaouk, G.-P. Haber, R. Autorino, S. Crouzet, A. Ouzzane, V. Flamand, and A. Villers, "A Novel Robotic System for Single-port Urologic Surgery: First Clinical Investigation," *European Urology*, vol. 66, pp. 1033–1043, December 2014.

[32] F. C. Holsinger, "A flexible, single-arm robotic surgical system for transoral resection of the tonsil and lateral pharyngeal wall: Next-generation robotic head and neck surgery," *The Laryngoscope*, vol. 126, pp. 864–869, April 2016.

[33] J. S. Lam, J. Bergman, A. Breda, and P. G. Schulam, "Importance of surgical margins in the management of renal cell carcinoma.," *Nature clinical practice. Urology*, vol. 5, pp. 308–17, June 2008.

[34] Y. Munz, K. Moorthy, A. Dosis, J. D. 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*, vol. 18, pp. 611–616, April 2004.

[35] E. D. Light, S. F. Idriss, K. F. Sullivan, P. D. Wolf, and S. W. Smith, "Real-Time 3D Laparoscopic Ultrasonography," *Ultrasonic Imaging*, vol. 27, pp. 129–144, July 2005.

[36] B. F. Kaczmarek, S. Sukumar, R. K. Kumar, N. Desa, K. Jost, M. Diaz, M. Menon, and C. G. Rogers, "Comparison of Robotic and Laparoscopic Ultrasound Probes for Robotic Partial Nephrectomy," *Journal of Endourology*, vol. 27, pp. 1137–1140, September 2013.

[37] J. Shepard, R. N. and Metzler, "Mental Rotation of Three-Dimensional Objects," *Science*, vol. 171, pp. 701–703, 1971.

[38] A. Bundesen, C. and Larsen, "Visual Transformation of Size," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 1, pp. 214–220, 1975.

[39] G. Janetschek, P. Daffner, R. Peschel, and G. Bartsch, " Laparoscopic Nephron Sparing Surgery for Small Renal Cell Carcinoma," *The Journal of Urology*, vol. 159, pp. 1152–1155, April 1998.

[40] H. Wada, T. Anayama, K. Hirohashi, T. Nakajima, T. Kato, T. K. Waddell, S. Keshavjee, I. Yoshino, and K. Yasufuku, "Thoracoscopic ultrasonography for localization of subcentimetre lung nodules.," *European journal of cardio-thoracic surgery : official journal of the European Association for Cardio-thoracic Surgery*, vol. 49, pp. 690–7, February 2016.

[41] M. Khereba, P. Ferraro, A. Duranceau, J. Martin, E. Goudie, V. Thiffault, and M. Liberman, "Thoracoscopic localization of intraparenchymal pulmonary nodules using direct intracavitary thoracoscopic ultrasonography prevents conversion of VATS procedures to thoracotomy in selected patients.," *The Journal of thoracic and cardiovascular surgery*, vol. 144, pp. 1160–5, November 2012.

[42] K. A. Perry, J. A. Myers, and D. J. Deziel, "Laparoscopic ultrasound as the primary method for bile duct imaging during cholecystectomy," *Surgical Endoscopy*, vol. 22, pp. 208–213, January 2008.

[43] J. Machi, J. O. Johnson, D. J. Deziel, N. J. Soper, E. Berber, A. Siperstein, M. Hata, A. Patel, K. Singh, and M. E. Arregui, "The routine use of laparoscopic ultrasound decreases bile duct injury: a multicenter study," *Surgical Endoscopy*, vol. 23, pp. 384–388, February 2009.

[44] B. Wu, R. L. Klatzky, and G. Stetten, "Learning to Reach to Locations Encoded from Imaging Displays.," *Spatial cognition and computation*, vol. 8, pp. 333–356, October 2008.

[45] P. M. Fitts, "Perceptual-Motor Skill Learning," in *Catagories of Human Learning*, pp. 381–391, Academic Press Inc., 1964.

[46] R. L. Klatzky, B. Wu, D. Shelton, and G. Stetten, "Effectiveness of augmented-reality visualization versus cognitive mediation for learning actions in near space," *ACM Transactions on Applied Perception*, vol. 5, pp. 1–23, January 2008.

[47] M. Wilson, J. McGrath, S. Vine, J. Brewer, D. Defriend, and R. Masters, "Psychomotor control in a virtual laparoscopic surgery training environment: gaze control parameters differentiate novices from experts.," *Surgical endoscopy*, vol. 24, pp. 2458–64, October 2010.

[48] J.-S. Hyun and S. J. Luck, "Visual working memory as the substrate for mental rotation.," *Psychonomic bulletin & review*, vol. 14, pp. 154–8, February 2007.

[49] D. J. Prime and P. Jolicoeur, "Mental Rotation Requires Visual Short-term Memory: Evidence from Human Electric Cortical Activity," *Journal of Cognitive Neuroscience*, vol. 22, pp. 2437–2446, November 2010.

[50] R. H. Thompson, B. R. Lane, C. M. Lohse, B. C. Leibovich, A. Fergany, I. Frank, I. S. Gill, M. L. Blute, and S. C. Campbell, "Every Minute Counts When the Renal Hilum Is Clamped During Partial Nephrectomy," *European Urology*, vol. 58, no. 3, pp. 340–345, 2010.

[51] P. S. Calhoun, B. S. Kuszyk, D. G. Heath, J. C. Carley, and E. K. Fishman, "Three-dimensional Volume Rendering of Spiral CT Data: Theory and Method," *RadioGraphics*, vol. 19, pp. 745–764, May 1999.

[52] F. Zöllner, "Ueber eine neue Art anorthoskopischer Zerrbilder," *Annalen der Physik und Chemie*, vol. 193, pp. 477–484, January 1862.

[53] E. M. Palmer, P. J. Kellman, and T. F. Shipley, "A theory of dynamic occluded and illusory object perception.," *Journal of Experimental Psychology: General*, vol. 135, pp. 513–541, November 2006.

[54] R. Fendrich, J. W. Rieger, and H.-J. Heinze, "The effect of retinal stabilization on anorthoscopic percepts under free-viewing conditions," *Vision Research*, vol. 45, pp. 567–582, March 2005.

[55] J. S. Girgus, L. H. Gellman, and J. Hochberg, "The effect of spatial order on piecemeal shape recognition: A developmental study," *Perception & Psychophysics*, vol. 28, pp. 133–138, March 1980.

[56] B. Wu, R. L. Klatzky, and G. D. Stetten, "Mental visualization of objects from cross-sectional images.," *Cognition*, vol. 123, pp. 33–49, April 2012.

[57] J. M. Loomis, R. L. Klatzky, and S. J. Lederman, "Similarity of Tactual and Visual Picture Recognition with Limited Field of View," *Perception*, vol. 20, pp. 167–177, April 1991.

[58] M. Hegarty, M. Keehner, C. Cohen, D. R. Montello, and Y. Lippa, "The Role of Spatial Cognition in Medicine: Applications for Selecting and Training Professionals.," in *Applied spatial cognition: From research to cognitive technology.*, pp. 285–315, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers, 2007.

[59] M. Lanca, "Three-Dimensional Representations of Contour Maps," *Contemporary Educational Psychology*, vol. 23, pp. 22–41, Janurary 1998.

[60] B. Wu, R. L. Klatzky, D. Shelton, and G. D. Stetten, "Psychophysical evaluation of in-situ ultrasound visualization.," *IEEE transactions on visualization and computer graphics*, vol. 11, no. 6, pp. 684–93, 2005.

[61] R. L. Klatzky, B. Wu, and G. Stetten, "Spatial Representations From Perception and Cognitive Mediation: The Case of Ultrasound.," *Current directions in psychological science*, vol. 17, pp. 359–364, December 2008.

[62] B. Wu, R. L. Klatzky, and G. Stetten, "Visualizing 3D objects from 2D cross sectional images displayed in-situ versus ex-situ.," *Journal of experimental psychology. Applied*, vol. 16, pp. 45–59, March 2010.

[63] C. L. Cheung, C. Wedlake, J. Moore, S. E. Pautler, and T. M. Peters, "Fused video and ultrasound images for minimally invasive partial nephrectomy: A phantom study," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010: 13th International Conference, Beijing, China, September 20-24, 2010, Proceedings, Part III*

(T. Jiang, N. Navab, J. P. W. Pluim, and M. A. Viergever, eds.), pp. 408–415, Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

[64] X. Kang, M. Azizian, E. Wilson, K. Wu, A. D. Martin, T. D. Kane, C. A. Peters, K. Cleary, and R. Shekhar, "Stereoscopic augmented reality for laparoscopic surgery.," *Surg. Endosc.*, pp. 1–9, 2014.

[65] M. Feuerstein, T. Reichl, J. Vogel, A. Schneider, H. Feussner, and N. Navab, "Magneto-optic tracking of a flexible laparoscopic ultrasound transducer for laparoscope augmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007: 10th International Conference, Brisbane, Australia, October 29 - November 2, 2007, Proceedings, Part I* (N. Ayache, S. Ourselin, and A. Maeder, eds.), pp. 458–466, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[66] P. Pratt, A. Di Marco, C. Payne, A. Darzi, and G.-Z. Yang, "Intraoperative ultrasound guidance for transanal endoscopic microsurgery," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part I*, pp. 463–470, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

Chapter 2

Robust, Intrinsic Tracking of a Laparoscopic Ultrasound Probe

This chapter is adapted from the papers,

- Jayarathne U.L., McLeod A.J., Peters T.M., Chen E.C.S. (2013) Robust Intraoperative US Probe Tracking Using a Monocular Endoscopic Camera. In: Mori K., Sakuma I., Sato Y., Barillot C., Navab N. (eds) Medical Image Computing and Computer-Assisted Intervention MICCAI 2013. MICCAI 2013. Lecture Notes in Computer Science, vol 8151. Springer, Berlin, Heidelberg

- Jayarathne U.L., Luo X., Chen E.C.S., Peters T.M. (2015) Simultaneous Estimation of Feature Correspondence and Stereo Object Pose with Application to Ultrasound Augmented Robotic Laparoscopy. In: Linte C., Yaniv Z., Fallavollita P. (eds) Augmented Environments for Computer-Assisted Interventions. MICCAI 2015. Lecture Notes in Computer Science, vol 9365. Springer, Cham

- Uditha L. Jayarathne, Elvis C.S Chen, John Moore, Terry M. Peters, "Robust, Intrinsic Tracking of a Laparoscopic Ultrasound Probe for Ultrasound-augmented Laparoscopy", IEEE Transactions on Medical Imaging, (*submitted*)

2.1 Introduction

In many laparoscopic procedures, surgeons use laparoscopic ultrasound (LUS) to visualize surgical targets hidden deep inside organs. For example, in laparoscopic resection tasks of many *endophytic* tumors, LUS is used to determine the resection margins, and to gain better understanding about the 3D form of the tumor and its surrounding [1–3]. During these procedures, the 2D ultrasound (US) image is conventionally displayed separately from the laparoscopic video. During conventional laparoscopic surgery, this second display is usually on the ultrasound machine itself, while in robot-assisted surgery with the *daVinci* system the US image is presented in a separate display panel known as the TilePro™. In either case, the ultrasound image and the laparoscopic video are spatially dissociated. In addition, these two modalities provide two distinct types of information: laparoscopy, being a projective imaging modality, provides a perspective projection of the scene, while the ultrasound image provides a tomographic view into the tissue. Therefore, reasoning about the geometry of the hidden surgical targets requires mentally solving the spatial alignment problem and the resolving the modality differences, which is cognitively very challenging [4, 5]. Mental representations of the hidden surgical targets in space acquired through such cognitive mediation are error-prone [6], and may cause incorrect surgical actions to be performed. *Hybrid in situ* visualization where US information is displayed in the frame of reference of the laparoscopic video enables mental representations of the targets through perception [7] which are more accurate. Such visualizations require the US image to be mapped to the camera coordinate system which involves a six degrees of freedom (6DoF) rigid transformation to be solved at camera frame rate.

2.1.1 Related Work

Many attempts to solve the rigid transformation that maps US image to a camera centered coordinate system can be found in the literature. A popular approach is to employ an *extrinsic* tracking system to track both the laparoscopic camera, and the ultrasound probe in a common coordinate system. Magnetic [8, 9], optical [10], or combined magnetic and optical tracking [11] are commonly used methods with application to conventional laparoscopy. However, optical tracking alone cannot be used to track the articulated imaging tip of commonly used LUS probes, since the line-of-sight constraints require the retro-reflective spheres to be attached on the probe handle. With application to robot-assisted surgery, one could compute this transform based on kinematic tracking of the camera and the robotic manipulator that holds the US probe [12]. However, such approaches require the US probe to be rigidly attached to the robotic arm, so that the transformation of the imaging tip with respect to the kinematic coordinate system can be computed by aggregating the pose of the end-effector with the constant transformation from the end-effector to the imaging tip. This constant transformation can be determined *apriori* through a calibration process. Despite its popularity, extrinsic-tracking-based solutions adds financial and logistical overhead to the existing operating room (OR) work-flow. In addition, the accuracy of such solutions may be affected by error accumulation in long transformation chains involved, or by the factors in the OR that affect the robustness of tracking. In particular, the accuracy of magnetic-tracking-based solutions may be affected by the presence of ferromagnetic materials in most ORs.

An attractive alternative to extrinsic-tracking-based solutions is image-based, *intrinsic* methods. During US guided tasks, the imaging tip of the LUS probe is always visible in the camera image. Therefore, image-based techniques can be developed to estimate the 6DoF pose of the LUS probe in the camera coordinate system. Leven et. al. [12] evaluated the feasibility of image-based pose tracking with stereoscopic video by attaching a colour marker to the back surface of a LUS probe. A spiralled-stripe on the marker enabled the estimation of the axial rotation. However, the reported tracking accuracy is not adequate for clinical applications, which could be partly attributed to high uncertainty in feature segmentation and localization. A slightly different approach, where an alternative, highly localized fiducial pattern is used, has demonstrated clinically acceptable accuracy bounds [13–17]. The fiducials, localized at the intersection of two high contrast edges or at the center of a circle, provides a point-cloud in 3D space, and a set of 2D coordinates in the image space. The 3D point cloud is defined with respect to a coordinate system defined arbitrarily on the pattern itself. If the correspondence between the 3D point-cloud and the 2D point-set localized in the image-space is known, the 6DoF pose can be estimated as a solution to the *perspective-n-point* (PnP) problem which can be solved accurately and efficiently [18–20]. In stereo-laparoscopy, where two views of the fiducial pattern are available simultaneously, the corresponding points localized in each image can be triangulated to obtain a 3D point-cloud in the camera coordinate system. The pose is then determined by registering this point-cloud with the fiducial point cloud by using an algorithm such as *iterative closest point* (ICP) [21] that solves for the 3D-3D correspondence, and the rigid transformation simultaneously.

Methods that rely on highly localized point-fiducials effectively solve two computational problems; (a) the 3D-to-2D point correspondence problem, and (b) the pose estimation problem. The use of a planar and structured fiducial geometry enables easy computation of a solution to the correspondence problem. For this reason the best performing methods use fiducial patterns arranged in a planar-grid [13–15, 17]. Since the fiducial geometry is known and simple, the 2D coordinate of an undetected fiducial can be interpolated to improve the robustness of tracking. However, most of the clinically used LUS probes are cylindrical in shape, making the planar fiducial pattern difficult to use. In addition, a planar pattern has a very limited tracking range. In a recent paper, Zhang et al. [16] described a method in which the fiducial pattern could cover the entire curved, back surface of a clinically used LUS probe. However, the method makes several assumptions that may not be true in practical surgery. In particular, it is not clear how the proposed method handles fiducial occlusions (either due to physical occlusion, or the feature detector being insensitive), as well as outliers, all of which are unavoidable in practice.

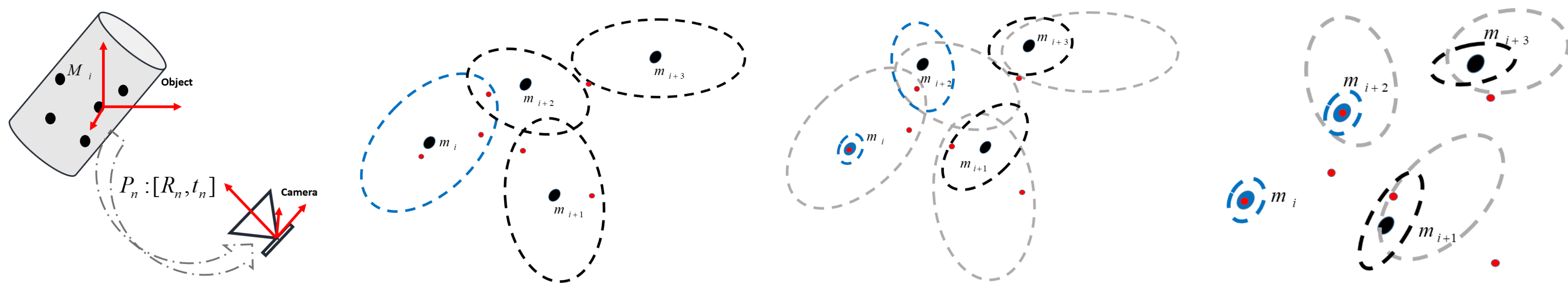
2.1.2 Contributions

In this chapter, I describe an image-based, intrinsic method to compute the pose of a fiducial pattern with fiducials randomly distributed over the curved back surface of a clinical LUS probe. The method allows some fiducials to be occluded while outliers, detected as a result of an imperfect feature detector, are properly handled. Instead of solving the correspondence problem and the pose problem sequentially, they are solved jointly using a Kalman Filter-based framework [22], that allows easy integration of *priors* on the pose based on the topology of the pose space, and the probe motion. Incorporation of strong priors based on probe motion allows the algorithm to converge to a solution quickly, enabling real-time performance. Moreover, the uncertainty of the solution is also reported, which is very useful in acquiring a comprehensive understanding about the tracking performance. The chapter includes a detailed description of a monocular image-based method as well as its extension to stereo/multi-view imaging models, an empirical investigation into the efficacy of the proposed methods, and a detailed discussion including limitations and insight into the future.

2.2 Methods

Since most clinical LUS probes do not provide adequate visual cues to be tracked in 6DoF, a special fiducial pattern is attached to the curved back surface of the probe. One could attach any textured pattern to the probe, and employ state-of-the-art detectors such as SIFT [23]/SURF [24] to localize interest points in image space. This will allow the correspondence problem to be solved explicitly using the descriptors computed by SIFT/SURF for each of the detected interest points. However, as demonstrated by Zeisl et al. [25], such features may introduce significant Fiducial Localization Error (FLE) that may propagate to the pose estimates. Therefore, to minimize the FLE, the fiducials in the pattern used in this work are localized at the intersection of two high contrast edges.

In the following subsections I describe the mathematical details of the pose estimation algorithm. This discussion assumes that the the laparoscopic camera image is free from geometric distortions while the camera is approximated by the *pinhole* model with its intrinsic matrix $\mathbf{A}$ known through an appropriate calibration method. When multiple views are involved, the pose $\mathbf{P}_c$ of one camera relative to the other is assumed to be known, in addition to the intrinsic matrix of each camera.



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

Figure 2.1: (a) Fiducial locations are defined with respect to a local coordinate system of the probe. Given these 3D points and their image space locations, the objective is to estimate the pose $\mathbf{P}_n$ that maps the 3D points to a coordinate system centered on the camera, (b) Given the prior on the pose, the image space locations of the model points (in black) is determined with their image space covariance (dashed ellipses). The output of the feature detector is depicted by red dots. A putative match is sought in the search region in blue, and the pose and the its covariance is updated, (c) Due to the established correspondence hypothesis, the image space covariance shrinks. Note the reduction in the size of elliptical regions compared to their previous size (in gray). The image space covariance for the matched model point has now shrunk to measurement noise (in blue), (d) The image space covariance is further reduced after two correspondence hypotheses. One last correspondence hypothesis is made at this stage before the correctness of the hypotheses is evaluated using the Eq. (2.6)

2.2.1 Preliminaries

Let us assume that each fiducial marker (*model poitns*) in the pattern in 3D space, is determined with respect to a local coordinate system by some means, and that the coordinate of the i^{th} point is given by $\mathbf{M}_i$. Let the cardinality of the *model points set* be N. When the fiducial pattern is visible in the n^{th} frame of a camera, an image processing routine (*feature detector*) is applied to the image to localize the fiducial points in the camera image. Let $\mathbf{u}_j$ be the 2D coordinate of the j^{th} fiducial localized in the image space. Occlusions of some fiducials, and outliers in the image space due to imperfections of the feature detector are allowed, hence, the cardinality of the two point-sets may not be equal. The rigid pose $\mathbf{P}_n$ of the fiducial pattern with respect to the camera in the n^{th} frame (Fig. 2.1(a)), is parameterized as a 6D vector $\mathbf{P}_n = [r_1, r_2, r_3, t_x, t_y, t_z]^T$ with three parameters representing rotations[1], and three parameters representing translation in 3D space. Given the model points $\mathbf{M}$, feature detector response $\mathbf{u}$, and the camera intrinsics $\mathbf{A}$, the objective is to estimate $\mathbf{P}_n$.

2.2.2 Simultaneous Pose and Correspondence from a Single View

Let us begin by assuming a *Gaussian prior* on the pose represented by its mean pose $\mathbf{P}_n$, and its 6x6 covariance matrix $\mathbf{\Sigma}_n^p$ for the n^{th} camera frame. Section 2.2.4 details how these priors are computed in this tracking framework. Given this prior, the image space location $\mathbf{m}_i$ of model point $\mathbf{M}_i$ is given by the Eq. (2.1) while the corresponding image space covariance $\mathbf{\Sigma}_i$ is given by Eq. (2.2),

$$\mathbf{m}_i = Proj(\mathbf{P}_n, \mathbf{M}_i) \tag{2.1}$$

$$\mathbf{\Sigma}_i = J(\mathbf{M}_i)\mathbf{\Sigma}_n^p J(\mathbf{M}_i)^T + \mathbf{R} \tag{2.2}$$

where $Proj(\mathbf{P}_n, \mathbf{M}_i)$ is the operator that projects the i^{th} model point with pose $\mathbf{P}_n$, $J(.)$ is its Jacobian, and $\mathbf{R}$ is a 2x2 diagonal matrix representing isotropic measurement uncertainty. $\mathbf{\Sigma}_i$

[1] exponential maps are used to parameterize rotations

defines a search region for a putative match, and I only consider 2D points such that

$$(\mathbf{m}_i - \mathbf{u}_j)^T \mathbf{\Sigma}_i (\mathbf{m}_i - \mathbf{u}_j) \leq \Psi \tag{2.3}$$

with $\Psi = 3$ giving 99% confidence in matching. Thus, the search space for a match for $\mathbf{M}_i$ reduces to an elliptical region (Fig. 2.1(b)). Inside the search region for putative matches, several feature points may be found, each of which is equally probable to be a correct match. Hypothesizing that one such point $\mathbf{u}_j$ is the correct match, the pose and its covariance are updated using Extended Kalman Filter (EKF) equations, Eq. (2.4) and Eq. (2.5).

$$\mathbf{P}'_n = \mathbf{P}_n + \mathbf{K}(\mathbf{u}_j - \mathbf{m}_i); \tag{2.4}$$

$$\mathbf{\Sigma}'^p_n = (\mathbf{I} - \mathbf{K}J(\mathbf{P}_n))\mathbf{\Sigma}_n, \tag{2.5}$$

where, $\mathbf{K}$ is the optimal Kalman gain and $\mathbf{I}$ is an identity matrix.

With the updated pose and its covariance, the image space locations and their covariances for model points are computed again using Eqs. (2.1) and (2.2). As illustrated in Fig. 2.1(c), the image space covariance has been reduced as a result of the established correspondence. Note that the image space covariance of the previously considered model point has now shrunk to the level of measurement noise. At this stage, a match is searched for a different fiducial in its corresponding search region. One could select this fiducial such that the covariance is maximally reduced [26, 27]. However, in this work, the fiducial to be considered next is randomly selected. For a putative match, the pose and its covariance is updated again using Eqs. (2.4) and (2.5) to yield further reduced image space covariance (Fig. 2.1(d)). A match is searched in another search region, and the pose and its covariance is updated again. After three such pose updates, the image space covariance does not evolve significantly. At this point, all the fiducial points are projected into the image space and the euclidean distance to the closest 2D point is computed. If no feature is found within a predefined distance threshold D for a projected fiducial, the particular fiducial is considered to be occluded. The error in Eq. 2.6 is then computed to test the correctness of the hypothesized correspondence triplet,

$$Error(\mathbf{P}) = \sum_{(\mathbf{M},\mathbf{u})\in Matches} ||\mathbf{u} - Proj(\mathbf{P}, \mathbf{M})|| + \lambda|NotDetected| \tag{2.6}$$

where $||.||$ represent the euclidean norm, $|NotDetected|$ is the cardinality of undetected fiducials, and λ is a tunable parameter.

If the error in Eq. (2.6) drops below a pre-defined threshold, the hypothesized correspondence triplet is consistent with image data, and hence is considered to be correct. Therefore the pose can be further refined to yield a more accurate estimate. Note that at this point, correct 3D-2D point correspondence is known, hence, more accurate PnP solvers [28] can be used for pose refinement. However, in this work, the refined final pose is achieved by applying an EKF update using Eqs. (2.4) and (2.5) for each model-feature correspondence. If the error threshold cannot be met, the algorithm backtracks and evaluates the error with the next probable correspondence candidate.

To manage correspondence hypotheses during the search for the pose that minimizes Eq. (2.6), a *tree* data structure traversed in *depth first* order is used. Each node in the tree holds a 3D-2D correspondence pair, the current pose and its covariance. To better understand the traversal process, consider the following scenario illustrated in Fig. (2.2). Suppose that we begin with a Gaussian prior on the pose given by its mean $\mathbf{P}_n^0$, and covariance $\mathbf{\Sigma}_n^0$. At the first depth level, *child nodes* hold values for $\{\mathbf{P}_n^0, \mathbf{\Sigma}_n^0, \mathbf{M}_i\}$. Let us assume that model point $\mathbf{M}_1$ is chosen, and there are k putative matches. The hypothesis that the 2D point $\mathbf{u}_1$ is the correct match results in a child node that holds values for $\{\mathbf{P}_n^1, \mathbf{\Sigma}_n^1, \mathbf{M}_1, \mathbf{u}_1\}$ with $\mathbf{P}_n^1$ and $\mathbf{\Sigma}_n^1$ being the updated pose and its covariance (Fig. 2.2). Let us now assume that we select the search space for $\mathbf{M}_3$, and that $\mathbf{u}_5$ is hypothesized to be the correct match among p putative matches. This results in the child node containing values for $\{\mathbf{P}_n^2, \mathbf{\Sigma}_n^2, \mathbf{M}_3, \mathbf{u}_5\}$, where $\mathbf{P}_n^2$, and $\mathbf{\Sigma}_n^2$ represents

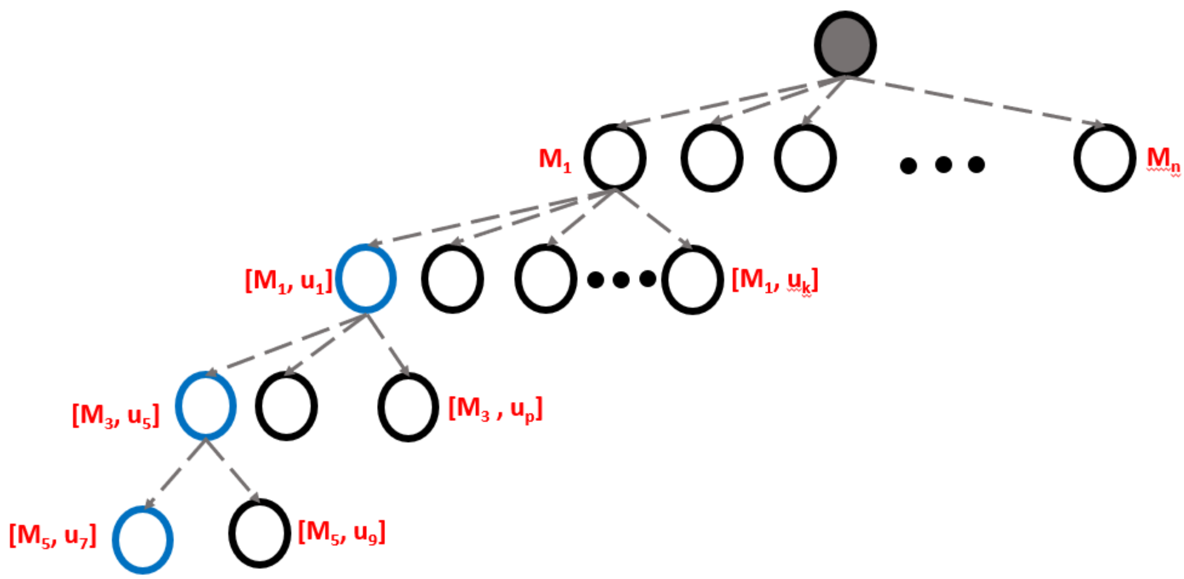


Figure 2.2: The correspondence hypotheses are managed in a tree data structure traversed in depth-first order. Each child node contains a correspondence pair, the current pose and its covariance. Indicated in blue are the nodes that have been visited in the discussed scenario.

the updated pose and its uncertainty at this node. Another hypothesis that model point $\mathbf{M}_5$ pairs with 2D point $\mathbf{u}_7$ spawns a child node that stores $\{\mathbf{P}_n^3, \mathbf{\Sigma}_n^3, \mathbf{M}_5, \mathbf{u}_7\}$ with updated pose and covariance. At this point, the correctness of the three hypotheses is evaluated by checking whether the error in Eq.(2.6) drops below the pre-defined threshold, the value of which is determined empirically. If the threshold is not met, the algorithm backtracks and hypothesizes a different pairing spawning a new child. This process is continued until the error threshold is met.

2.2.3 Handling Not-Detected Points

After three consecutive correspondence hypotheses, all the fiducials are projected into the image space, and the fiducials that do not find a 2D point within a pre-defined distance radius are considered to be undetected. However, the fiducials involved in the hypotheses triplet itself might not have been detected in the first place. Such situations are handled by modeling the probability of multiple consecutive not detected points as a process of sampling without replacement [22]: Given the number of model points N, the cardinality of the not detected points Nd, the probability of picking r not detected fiducials is,

$$Pr = Nd!(N-r)!/(Nd-r)!N! \tag{2.7}$$

Since the number of undetected fiducials at a given frame is not known *a priori*, the value of Nd is set such that at least 70% of these are visible in a given frame. When this probability drops below a pre-defined value, 5% in the experiments, the three fiducials involved are considered to be detected.

2.2.4 Priors

The joint pose and correspondence estimation framework described above begins with an initial prior on the pose and its covariance, and converges to a solution that is consistent with image data. This prior for the first frame is provided by a Gaussian Mixture distribution [22] that models the 6D pose space for the LUS probe. By assuming minimum and the maximum orientation and the translation constraints on the probe motion, pose samples are simulated, and a Gaussian Mixture model (GMM), a weighted sum of 6D Gaussian components given by Eq. (2.8), is learned [29] offline to represent the 6D pose space.

$$p(\mathbf{P}) = \sum_{i=1}^{K} w_i \mathcal{N}(\mathbf{P}_i, \mathbf{\Sigma}_i) \tag{2.8}$$

Alternatively, one could employ an extrinsic tracking system to acquire adequate number of pose samples to learn the GMM. During initialization, starting from the Gaussian component with the highest weight, each component is sequentially selected to provide the mean pose and the covariance for algorithm initialization. If a pose is not found for a particular Gaussian component, that with the next highest weight is selected. Once the algorithm converges to a solution, the initial pose and the covariance for the next frame is predicted based on the Kalman filter state prediction step. An identity process model given by Eq. (2.9) and Eq. (2.10) is used, where the subscript n stands for the frame number, and $\mathbf{Q}$ is the process noise.

$$\mathbf{P}_{n+1} = \mathbf{P}_n \tag{2.9}$$

$$\mathbf{\Sigma}_{n+1} = \mathbf{\Sigma}_n + \mathbf{Q} \tag{2.10}$$

These motion-based priors are much stronger than those based on the GMM, hence, the algorithm to converges rapidly.

2.2.5 Extension to Multiple Views

The joint correspondence and pose estimation framework described above can be extended to multiple views with minor modifications. A notation similar to that used in section 2.2.2 is employed to describe the algorithmic details here. Let $\mathbf{m}_i^c$ be a vector of the form $[\mathbf{m}_i^1, ..., \mathbf{m}_i^C]^T$ representing the locations of the model point $\mathbf{M}_i$ on C different camera images, and $\mathbf{\Sigma}_i^c$ its covariance. Unlike the linearization approach used in the previous EKF-based method (Eq. (2.1) and Eq. (2.2)), the *Unscented Transform* (UT) [30] is used to determine $\mathbf{m}_i^c$ and $\mathbf{\Sigma}_i^c$. The UT avoids complex Jacobian computation in the EKF-based approach, and its superiority in handling non-linearities in the multi-view imaging model results in better numerical stability. The UT uses $2L + 1$ *sigma points* $\mathbf{X}$, a set of samples in the input space acquired through a deterministic sampling approach, that are propagated through the non-linear transformation to compute the mean and the covariance in the transformed space. Here L is the dimension of the state vector, six in this work. The sigma points are given in the Eq. (2.11),

$$\mathbf{X} = [\mathbf{P}_n \quad \mathbf{P}_n + \gamma\mathbf{S}_n \quad \mathbf{P}_n - \gamma\mathbf{S}_n] \tag{2.11}$$

where $\mathbf{S}_n = chol\{\mathbf{\Sigma}_n^p\}$ with $chol\{.\}$ representing the matrix square-root operator via Cholesky factorization, and $\gamma = \sqrt{L+\lambda}$ with $\lambda = \alpha^2(L+\kappa) - L$ where α and κ are scaling parameters.

In all the experiments α and κ is set to a very a small value (0.0001). Image space location and covariance of the model point $\mathbf{M}_i$ is given by Eq. (2.12) through Eq. (2.14).

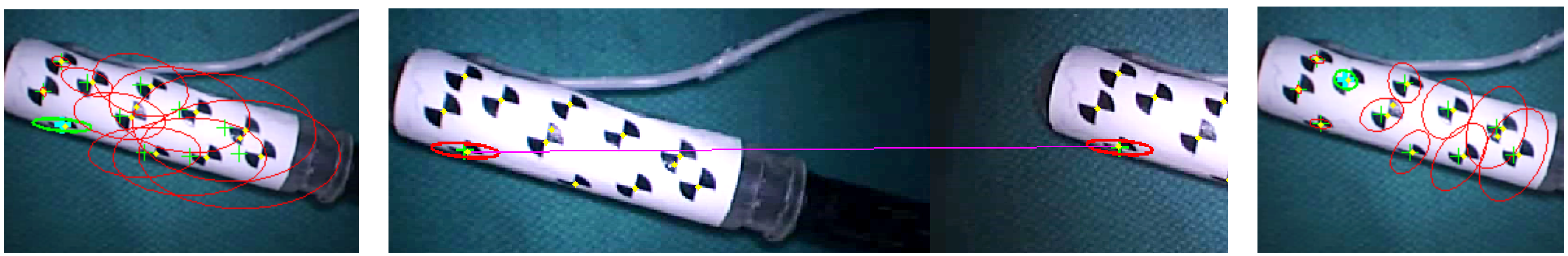
$$\mathbf{m}_i^c = \sum_{l=0}^{l=2L} W_l^{(m)} Proj^c(\mathbf{X}_l, \mathbf{M}_i); \tag{2.12}$$

$$\mathbf{\Sigma}_i^c = qr\{\left[\sqrt{W_1^{(c)}}[Proj^c(\mathbf{X}_{1:2L}, \mathbf{M}_i) - \mathbf{m}_i^c] \quad \sqrt{\mathbf{R}}\right]\}; \tag{2.13}$$

$$\mathbf{\Sigma}_i^c = cholupdate\{\mathbf{\Sigma_i^c}, Proj^c(\mathbf{X}_0, \mathbf{M}_i), W_0^{(c)}\}, \tag{2.14}$$

where $\mathbf{X}_l$ is the l^{th} sigma point and the operator $Proj^c(\mathbf{X}_l, .)$ returns the vector of the image space locations in the C^{th} camera for the sigma point $\mathbf{X}_l$. $\{W_i\}$ is a set of scalar weights such that $W_0^{(m)} = \lambda/(L+\lambda)$, $W_0^{(c)} = \lambda/(L+\lambda) + (1-\alpha^2+\beta)$, $W_i^{(m)} = W_i^{(c)} = 1/(2(L+\lambda))$ with $i = 1, ..., 2L$. Assuming all the distributions are Gaussian, the scalar parameter β is set to 2 in this work. $qr\{.\}$ is a short-hand notation for the QR Decomposition where the upper triangular matrix is returned, and $cholupdate\{.\}$ represents the Cholesky update process which is necessary to account for negative values of W_0.

Once the image space location and covariance are determined for the initial pose $\mathbf{P}_n$ and its covariance $\mathbf{\Sigma}_n$, an elliptic search region (Fig. 2.3(a)) is chosen and putative matches are sought, similar to the approach described in Section 2.2.2. Since each model point $\mathbf{M}_i$ is imaged by multiple cameras, putative matches must be searched in corresponding elliptical regions across multiple images. If multiple 2D points are found in the elliptical search regions for a given model point, the search space for the pose could grow exponentially since each combination is equally likely. To prevent this from happening, the geometry of multiple views is exploited: For each feature point lying inside the elliptical search region in the k^{th} camera, Samson distance [31] given by Eq. (2.15) is computed to each point in the $k+1^{th}$ camera lying inside the corresponding search ellipse. Here the 3x3 matrix $\mathbf{F}$ is the fundamental matrix between the two cameras considered, determined through a calibration process.



(a) (b) (c)

Figure 2.3: An instance of the algorithm. In (a) and (c) only the left image is shown due to space limitations: (a) Model points are projected to the image space (green + marker) with their covariance (in red ellipses while the bold green ellipse represents the one currently being evaluated) (b) Within the ellipse currently being evaluated, the correspondence between the detected features (yellow dots) across images is determined based on minimum Sampsons distance (c) image space uncertainty drops after the SRUKF update. This is indicated by the shrinkage of the search ellipses.

$$d_{sampson} = \frac{(\mathbf{m}_i^{k+1^T} \mathbf{F} \mathbf{m}_i^k)^2}{||\mathbf{F}\mathbf{m}_i^k||^2 + ||\mathbf{F}^T \mathbf{m}_i^{k+1}||^2} \tag{2.15}$$

By pairing feature points that are less than a predefined Sampsons distance threshold (Fig. 2.3(b)), correspondence hypotheses that are geometrically inconsistent are eliminated. A geometrically consistent point-pair forms a vector $\mathbf{u}_j$, which is hypothesized to correspond to the model point $\mathbf{M}_i$. Based on this hypothesis, the pose and its covariance is updated using the Square-Root Kalman Filter (SRUKF) state update equations given by Eq. (2.16) through Eq. (2.18),

$$\mathbf{P}'_n = \mathbf{P}_n + \mathbf{K}_s(\mathbf{m}_i^c - \mathbf{u}_j^c) \tag{2.16}$$

$$\mathbf{U} = \mathbf{K}_s \mathbf{\Sigma}_i^c \tag{2.17}$$

$$\mathbf{S}'_n = cholupdate\{\mathbf{S}_n, \mathbf{U}, -1\} \tag{2.18}$$

where $\mathbf{K}_s$ is the Kalman Gain. Note that $\mathbf{S}'_n$ is the square-root of the updated covariance of the pose, from which $\mathbf{\Sigma}'_n$ is derived from it.

This update reduces the covariance of the pose further constraining the search regions across images, as illustrated by the shrinkage of the search ellipses in Fig. 2.3(c). Model points are projected back again with the updated pose and the covariance, and a putative match is searched as described above. After three such updates the evolution of the pose (and the covariance) becomes negligible, and at this point, the model points are projected to each image and the error in Eq. (2.19) is computed to evaluate the validity of the three hypothesized correspondences.

$$Error(\mathbf{P}) = \sum_{k=1}^{k=C} \sum_{(\mathbf{M},\mathbf{u}^k)\in Matches} ||\mathbf{u}^k - Proj_k(\mathbf{P},\mathbf{M})|| + \lambda|NotDetected| \tag{2.19}$$

Here $\mathbf{u}^k$ is the 2D point closest to the model point projected to the k^{th} camera, $Proj_k(.,.)$ is the projection operator for the k^{th} image, and C is the number of cameras. If this error drops below a pre-defined threshold, the updated pose is further refined by applying an SRUKF-update (Eq. (2.16) through Eq. (2.18)) for each model-feature correspondence to yield a more accurate estimate. Otherwise, the algorithm continues with a different correspondence hypothesis similar to the strategy described in Section 2.2.2. Similar to the EKF-based method described for the monocular case, the search space is managed using a tree data structure traversed in depth-first fashion, and non-detected points are handled in a similar manner. Priors described in Section 2.2.4 are used to initialize the pose and the covariance for a given frame.

2.3 Experiments and Results

In this section, I detail the experiments conducted followed by the results to quantitatively and qualitatively demonstrate the efficacy of the methods developed for a monoscopic camera as well as for a stereoscopic camera.

2.3.1 Experimental Setup

An Olympus stereo-laparoscopic camera from the *daVinci S* (Intuitive Surgical Inc., USA) surgical system was used in all the conducted experiments, with the left camera channel used for monocular-image-based tracking experiments. Videos from from left and the right cameras were captured at 640x480 resolution, and the geometric calibration for each was performed by employing a checkerobard calibration pattern[32]. In addition, the fundamental matrix for the stereo-pair was derived with a method based on least median of squares [33].

To qualitatively demonstrate the efficacy of the image-based tracking framework with application to US-augmented laparoscopy, a clinical LUS probe (LAP9-4/38, Ultrasonix, Canada) was used with a fiducial pattern attached to the curved back surface of its imaging tip (Fig. 3.1(a)). The 3D position of each fiducial was determined accurately with a measurement microscope (STM6-LM, Olympus, Japan), and a local coordinate system was defined on the fiducial space with one fiducial arbitrarily chosen as the origin. To evaluate the accuracy of the pose estimates, a mock-probe was 3D printed . A fiducial pattern similar to that attached to the LUS probe was attached to the curved back-surface of the mock-probe, and the locations of the fiducials were determined by using the measurement microscope. The mock-probe was rigidly placed at the geometric center of an optically tracked assembly (*validation tool*) (Fig.3.1 (b)). When the LUS probe, or the validation tool was visible in camera images, the 2D image locations of the fiducials were determined by employing an efficient algorithm[34] in each left and right camera frame with its output refined to sub-pixel accuracy.

An optical tracking system (OTS) (Spectra, NDI, Canada) was used to provide the reference poses by tracking a stereoscopic endoscope and the validation tool. To be able to compare the image-based estimates with the ground truth, the transformation from the dynamic reference body (DRB) attached to the laparoscopic camera to the coordinate system centered at the optical origin of the left camera, known as the hand-eye calibration, was estimated by using a globally optimal algorithm [35]. The transformation between the fiducial pattern and the optically tracked DRB on the validation tool was determined by a robust method that minimizes the least squared error [36]. Videos from the camera were saved, and the pose of the probe or the validation tool was estimated for each video frame offline using image-based methods detailed in Sections 2.2.2 and 2.2.5, implemented in Matlab. All the computations were performed in a computer with Intel Core i7 (3.4GHz) 64-bit CPU and 16 GB of RAM running Microsoft Windows 7.

(a) (b) (c)

Figure 2.4: a) A fiducial pattern is attached to the curved back surface of the imaging tip of a clinically used LUS probe, (b) Optically tracked validation tool with a mock-probe rigidly placed at the geometric center of the retro-reflective spheres, (c) The PVC filled Z-phantom is tracked using a planar checkerboard pattern attached to one of its flat surfaces. The probe is tracked by using the method described in Section 2.2.2

2.3.2 Comparison to the Optical Tracking-based Reference

The validation tool was moved in the 3D space in front of the camera while the endoscope and the validation tool were tracked by the optical tracking system. The stereoscopic video was saved together with the OTS-based pose of the validation tool with respect to the DRB on the camera. Pose estimates of the mock-probe were then computed using three image-based methods: monocular image-based method described in Section 2.2.2 termed MONO, multi-view image-based method detailed in Section 2.2.5 applied to stereoscopic images termed STEREO, and the stereoscopic triangulation-based method termed TR. Conventionally, the TR method works by first establishing feature correspondence across the two views and triangulating [37] to obtain a 3D point-cloud in the camera coordinate system. This point-cloud is then rigidly registered to the model-points using an algorithm like ICP that solves for both 3D-3D point correspondence and the rigid body transformation. However, the ICP algorithm is sensitive to initialization, and may converge to the wrong solution in the presence of occlusion and clutter in the triangulated point-set. Therefore, to avoid any biases in the TR implementation, the 3D-2D point correspondences solved by the STEREO method were used, and recovered the pose parameters by registering the point-sets using the standard *orthogonal procrustes* analysis [38].

2.3.3 US-Video Overlay

To demonstrate the suitability of the proposed intrinsic tracking methods for the intended laparoscopic application (i.e *in-situ* US visualization), a phantom-based experiment was conducted. To this end, the constant transformation between the US image-based coordinate system to the fiducial marker-based coordinate system, known as US calibration transformation, is required in addition to the pose of the fiducial pattern. This transformation was determined using a conventional method that employs a Z-phantom [39] adapted to this application. The Z-phantom was filled with polyvinyl chloride-plastisol (PVC) compound, which serves as a clear tissue-mimicking medium (Fig. 3.1(c)). A rectangular checkerboard pattern was rigidly attached to a planar surface of the phantom to enable image-based 6DoF tracking of the phantom using an iterative technique based on Levenberg-Marquard optimization [33].

Seventeen US/video image pairs were acquired for calibration. For each video frame, the pose of the LUS probe with respect to the camera was determined using the monocular image-based method described in Section 2.2.2, and the pose is transformed to checkerboard pattern-based coordinate system on the phantom. In each US image, the two fiducials, corresponding to the cross line of the Z-phantom were manually identified to obtain a total of 34 homologous points in both the US image and the 3D space. The calibration was accomplished using the Orthogonal Procrustes algorithm which resulted in mean Fiducial Registration Error (FRE) of 2.2 ± 0.3mm.

Once US calibration is determined, it was used to overlay US images in video together with the poses of the LUS probe. A block of PVC through which a surgical needle was inserted was imaged. The alignment of the US visible portion with the portion visible in the laparoscopic image qualitatively validates the pose estimates computed by the proposed method. To demonstrate the ability of the proposed method to estimate the pose under fiducial occlusion, several fiducials were covered with a surgical instrument.

2.3.4 Results

Over 650 frames, 95% confidence interval (CI) of the root mean squared error (RMSE) of the translation was observed to be 2.0 ± 0.1mm, 1.6 ± 0.1mm and 2.4 ± 0.1mm for MONO, STEREO and TR respectively. Assuming that the translation errors are Gaussian, independently and identically distributed (i.i.d), a paired t-test revealed that the translational error resulted from the STERO method is significantly lower compared to the MONO method ($p < 0.001$), and the TR method ($p < 0.001$). In addition, the TR method resulted in the largest translational error ($p < 0.001$). The angular errors were analyzed as errors of rotation around the axes in the camera coordinate system (Fig 2.5). With the same assumption as above, paired t-tests were carried out for rotational errors as well. The TR method demonstrated significantly larger rotational error about the camera x-axis (95% CI [0.17 ± 0.01], compared to [0.02 ± 0.002] in STEREO,

$p < 0.001$), while the errors from the other two methods were statistically indistinguishable ($p > 0.6$). In estimating the rotation about the camera y-axis, TR method demonstrated the worst performance (95% CI [5.45 ± 0.35], $p < 0.001$), compared to the MONO (95% CI [0.82 ± 0.08]), and the STEREO method (95% CI [0.63 ± 0.05]). Even though the estimates resulting from the STEREO and MONO methods are practically indistinguishable, the STEREO method resulted in statistically better estimates ($p < 0.001$). Once again, the worst rotational errors about the z-axis came from the TR method (95% CI [0.84 ± 0.05], $p < 0.001$). Statistically, the MONO method resulted in better estimates (95% CI [0.36 ± 0.02, $p < 0.001$) compared to the STEREO method (95% CI [0.42 ± 0.03]).

Considering 3D points corresponding to each US pixel as targets, the Target Registration Error (TRE) was computed for MONO, STEREO and TR methods, using the estimated US calibration matrix to transform US pixel locations to the coordinate system defined on the fiducial pattern. TRE measures the true registration error at targets that the surgeon is interested in. In this application, 3D locations imaged by the LUS probe are the targets of interest. The mean TRE maps for each method are shown in Fig. 2.6. The TR method demonstrates the worst TRE (mean ranging from 1.6 – 4.9mm), while the TRE results in STEREO demonstrates the best (mean ranging from 0.9 – 1.1mm). Even though minor, an improvement in the mean TRE is observed in STEREO compared to that in MONO (compare range of mean TRE in MONO, 1.1 – 1.3mm, to that in STEREO).

To investigate the computational performance of the proposed methods, the running time of MONO and STEREO method in the Matlab implementation was measured. MONO method required 44ms for initialization from the GMM-based prior, while the STEREO method took 196.2ms. Once initialized, MONO method required 15.6ms on average to compute the pose per camera frame while the STEREO method required 37.9ms.

Fig. 2.7 qualitatively demonstrates the performance of the proposed methods. As illustrated in Fig. 2.7(a), the STEREO method estimates the pose, and correspondence (2D-2D correspondence cross images, and 3D-2D model-to-image correspondence) even under occlusion of some fiducials by a surgical tool. In the augmented image in Fig. 2.7(b), note that the needle seen in US image is in alignment with the ends of the needle protruding from the phantom, visually validating the estimated pose. Also, note that the pose has been estimated even under partial feature occlusion. It is important to notice that the overlaid US image in Fig. 2.7(b) is perceived to be floating above the rest of the scene, even though it is rendered at the correct spatial location. Remedies to the issue of naïve rendering are discussed in the following chapters.

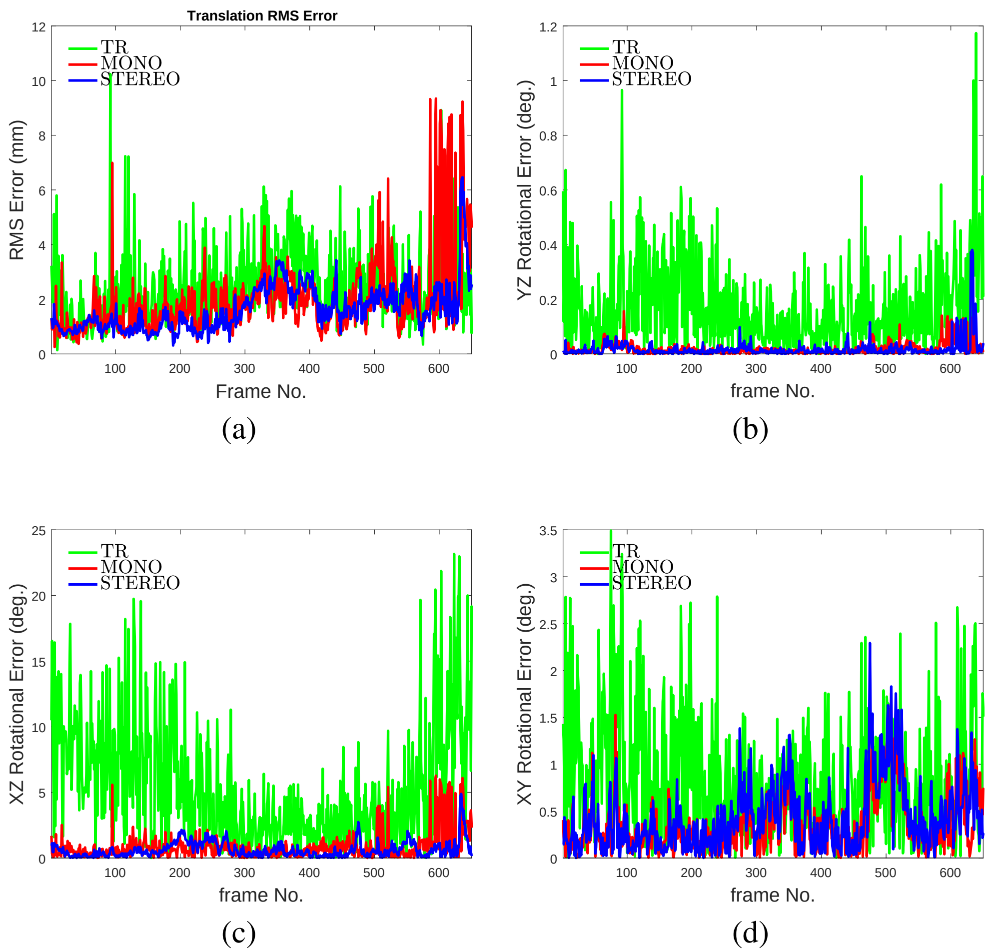
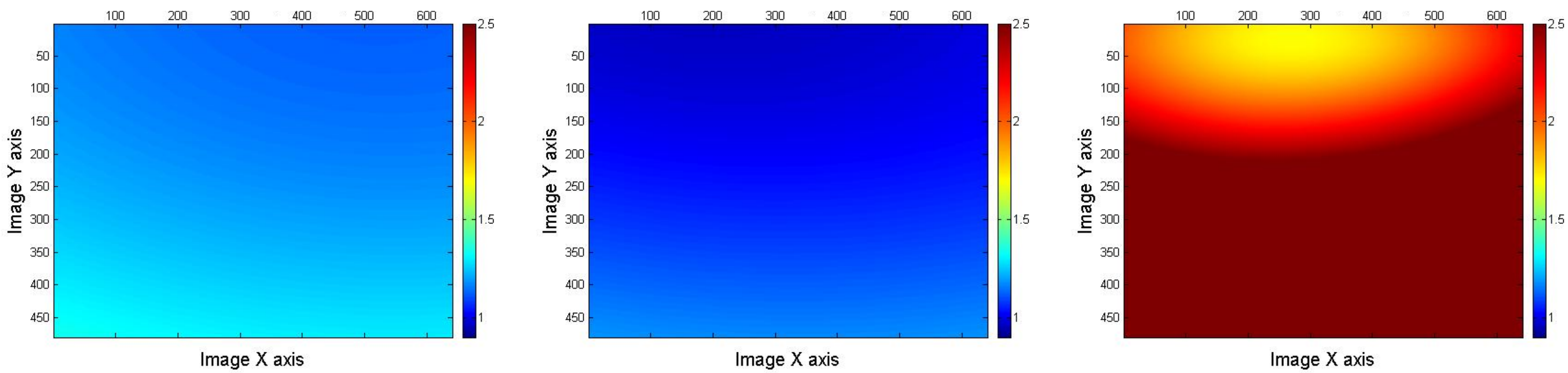
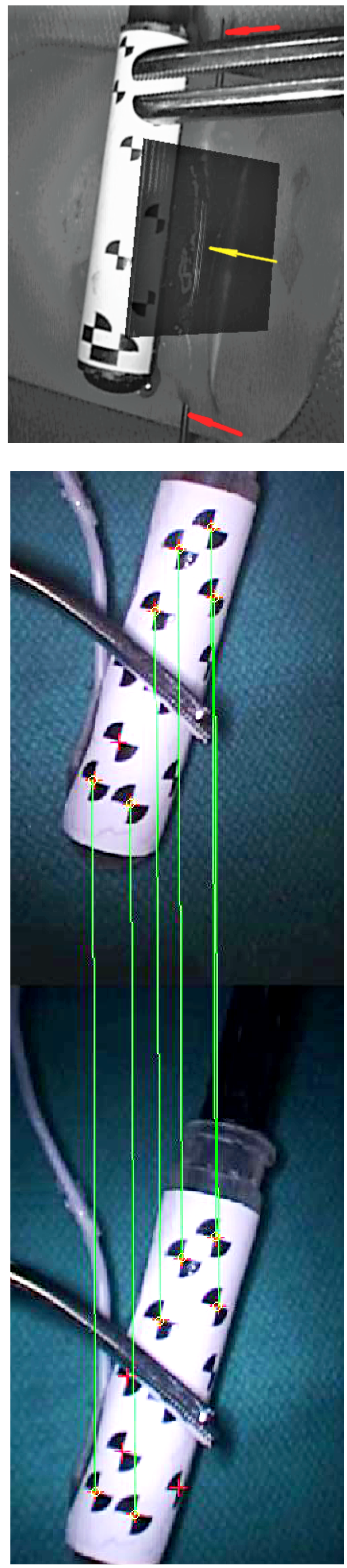


Figure 2.5: Error in pose estimates, RMSE in translation (a), absolute error in rotation in YZ-plane (b), absolute error in rotation in the XZ-plane (c), and the absolute error in rotation about the XY-plane (d). All the errors are computed with respect to the OTS measurements. Errors in TR, MONO, and STEREO methods are indicated in green, red, and blue respectively



(a) (b) (c)

Figure 2.6: The mean TRE maps for (a) MONO, (b) STEREO, and (c) TR methods. The X and Y axes correspond to that of the US image while the origin of the fiducial pattern is close to the top-left corner. Errors are in millimeters. Note that the scale in the error bar is the same in each image



(a) (b)

Figure 2.7: (a) Stereo image showing that the proposed method has accurately determined the left-right feature correspondence even in the presence of clutter and partial feature occlusion of the fiducials, (b) US image is overlaid in the laparoscopic image using the pose estimates computed by the method detailed in the Section 2.2.5. Note that the US visible part of the needle (pointed by yellow arrow) is in alignment with its protruding ends visible in the laparoscopic image (pointed by red arrows) visually validating the computed pose estimates and the US calibration. The pose has been estimated under partial occlusion of the fiducial pattern by the surgical instrument

2.4 Discussion

Based on the experiments with the validation tool, the MONO and STEREO methods both produce significantly more accurate estimates compared to the TR method. The uncertainty in stereo triangulation could be culprit for the increased error in the TR method. In this method, the uncertainty in localization propagates to uncertainty in the triangulated points, and eventually results higher uncertainty in the computed estimates. In the Kalman filter-based approaches in both MONO and STEREO methods, each model-feature correspondence pair updates the pose vector directly without an intermediate step such as triangulation. Therefore, the accuracy of the computed pose by these two methods depends only on the fiducial localization uncertainty in the image space, and the nonlinear propagation of the uncertainty between the pose space and the measurement space. In the MONO method, this propagation is achieved by using the Jacobian of the nonlinear transformation, while in the STEREO the unscented transform is involved. The Jacobian approximation captures the nonlinearity only at first order accuracy (Taylor series expansion) while the unscented transform demonstrates third order accuracy for Gaussian distributions [40]. This difference in accuracy could be one reason why the STEREO method produces significantly better estimations, particularly in translation. Another reason for the improved performances of the STEREO method could be the wealth of depth information contained in the multiple-views.

The estimates resulting from the MONO and the STEREO methods are significantly less jittery compared to those from the TR method. One can use a separate Kalman filter to smooth the output from the TR method, but that will add computational cost on top of the pose computation. The embedded Kalman filter in the MONO and the STEREO methods enables smooth evolution of the estimates over time within the pose estimation framework itself without adding extra computational cost. Compared to the STEREO method, the MONO approach demonstrates better run-time performance, which could be attributed to the Jacobian-based linearization of the nonlinear transformation which can be computed efficiently. The unscented transform increases computational cost, thereby increasing the the computational cost of the

STEREO method. However, the STEREO method still demonstrates video rate (24fps) running time even with the unoptimized Matlab implementation. I believe that with optimized C/C++ -based, perhaps on parallel hardware, higher frame rates can be achieved.

Despite the improved performance, the proposed methods have some limitations that may be resolved with further research. As indicated in the results section, the GMM-prior-based initialization is quite expensive. This is revealed by the proposed methods spending significant time during initialization compared to subsequent estimation. Compared to the GMM representation of the pose space, the probe-motion-based prior used in subsequent frames is very strong. This results in tight search ellipses that significantly reduce the search space for the correspondences. Incorporating more prior knowledge into the GMM is one method of reducing the computational cost during the initialization. A parallel implementation where each Gaussian component is evaluated simultaneously could be another method of reducing the computational burden. In addition, the errors computed in the quantitative experiment using the validation tool could have contributions from both the hand-eye calibration, and the calibration from the transformation between the DRB on the validation tool and the fiducial pattern. However, this error contribution equally affects the three pose estimation methods compared. Therefore, the conclusions made from this experiment are still valid. Nevertheless, future research could investigate methods to minimize this error contribution to investigate the absolute error bounds of the proposed methods.

2.5 Conclusions

In this chapter, an image-based method is proposed to estimate the pose of a laparoscopic ultrasound probe for *in situ* visualization of ultrasound in laparoscopic surgery. It does not require an extrinsic spatial tracking system while the only requirement is to attach a fiducial pattern to the curved back surface of the probe. The probe pose and the image-to-fiducial correspondence is estimated simultaneously using the a Kalman filter-based framework. The framework is generalized to multiple views with application to stereo/robot-assisted laparoscopy, but with the expense of increased computational cost. With reference to an optical tracking-based reference, the proposed multi-view pose estimation framework achieves sub-millimeter TREs for targets imaged by the US probe, while the monocular image-based method demonstrates slightly degraded accuracy. Both of the methods achieve more accurate estimates compared to the conventional stereo-triangulation based pose estimation strategy that commercial optical tracking systems are based on. Future research should look into methods of further improving the running time of the proposed methods, particularly the multi-view-based method, and further improving the accuracy of the estimates by fusing multiple visual cues (points, edges, depth, etc.).

Bibliography

[1] J. S. Lam, J. Bergman, A. Breda, and P. G. Schulam, "Importance of surgical margins in the management of renal cell carcinoma.," *Nature clinical practice. Urology*, vol. 5, pp. 308–17, June 2008.

[2] G. Janetschek, P. Daffner, R. Peschel, and G. Bartsch, " Laparoscopic Nephron Sparing Surgery for Small Renal Cell Carcinoma," *The Journal of Urology*, vol. 159, pp. 1152–1155, April 1998.

[3] H. Wada, T. Anayama, K. Hirohashi, T. Nakajima, T. Kato, T. K. Waddell, S. Keshavjee, I. Yoshino, and K. Yasufuku, "Thoracoscopic ultrasonography for localization of subcentimetre lung nodules.," *European journal of cardio-thoracic surgery*, vol. 49, pp. 690–7, February 2016.

[4] J. Shepard, R. N. and Metzler, "Mental Rotation of Three-Dimensional Objects," *Science*, vol. 171, pp. 701—-703, 1971.

[5] A. Bundesen, C. and Larsen, "Visual Transformation of Size," *J Exp Psychol Hum Percept Perform*, vol. 1, pp. 214–220, 1975.

[6] R. L. Klatzky, B. Wu, and G. Stetten, "Spatial Representations From Perception and Cognitive Mediation: The Case of Ultrasound.," *Current directions in psychological science*, vol. 17, pp. 359–364, Dec. 2008.

[7] B. Wu, R. L. Klatzky, D. Shelton, and G. D. Stetten, "Psychophysical evaluation of in-situ ultrasound visualization.," *IEEE transactions on visualization and computer graphics*, vol. 11, no. 6, pp. 684–93, 2005.

[8] C. L. Cheung, C. Wedlake, J. Moore, S. E. Pautler, and T. M. Peters, "Fused video and ultrasound images for minimally invasive partial nephrectomy: A phantom study," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010: 13th International Conference, Beijing, China, September 20-24, 2010, Proceedings, Part III* (T. Jiang, N. Navab, J. P. W. Pluim, and M. A. Viergever, eds.), pp. 408–415, Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

[9] X. Liu, S. Kang, W. Plishker, G. Zaki, T. D. Kane, and R. Shekhar, "Laparoscopic stereoscopic augmented reality: toward a clinically viable electromagnetic tracking solution.," *Journal of medical imaging (Bellingham, Wash.)*, vol. 3, p. 045001, October 2016.

[10] X. Kang, M. Azizian, E. Wilson, K. Wu, A. D. Martin, T. D. Kane, C. a. Peters, K. Cleary, and R. Shekhar, "Stereoscopic augmented reality for laparoscopic surgery.," *Surgical Endoscopy*, pp. 1–9, 2014.

[11] M. Feuerstein, T. Reichl, J. Vogel, A. Schneider, H. Feussner, and N. Navab, "Magneto-optic tracking of a flexible laparoscopic ultrasound transducer for laparoscope augmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007: 10th International Conference, Brisbane, Australia, October 29 - November 2, 2007, Proceedings, Part I* (N. Ayache, S. Ourselin, and A. Maeder, eds.), pp. 458–466, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[12] J. Leven, D. Burschka, R. Kumar, G. Zhang, S. Blumenkranz, X. D. Dai, M. Awad, G. D. Hager, M. Marohn, M. Choti, C. Hasser, and R. H. Taylor, "Davinci canvas: A telerobotic surgical system with integrated, robot-assisted, laparoscopic ultrasound capability," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005: 8th International Conference, Palm Springs, CA, USA, October 26-29, 2005, Proceedings, Part I* (J. S. Duncan and G. Gerig, eds.), pp. 811–818, Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[13] P. Pratt, A. Di Marco, C. Payne, A. Darzi, and G.-Z. Yang, "Intraoperative ultrasound guidance for transanal endoscopic microsurgery," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part I* (N. Ayache, H. Delingette, P. Golland, and K. Mori, eds.), pp. 463–470, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[14] P. Pratt, A. Jaeger, A. Hughes-Hallett, E. Mayer, J. Vale, A. Darzi, T. Peters, and G.-Z. Yang, "Robust ultrasound probe tracking: initial clinical experiences during robot-assisted partial nephrectomy," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–9, 2015.

[15] J. Wang, E. Kobayashi, and I. Sakuma, "Coarse-to-fine dot array marker detection with accurate edge localization for stereo visual tracking," *Biomedical Signal Processing and Control*, vol. 15, pp. 49 – 59, 2015.

[16] L. Zhang, M. Ye, P.-L. Chan, and G.-Z. Yang, "Real-time surgical tool tracking and pose estimation using a hybrid cylindrical marker," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, pp. 921–930, June 2017.

[17] R. Singala, P. Edgcumbe, P. Pratt, C. Nguan, and R. Rohling, "Intra-operative ultrasound-based augmented reality guidance for laparoscopic surgery," *Healthcare Technology Letters*, August 2017.

[18] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An Accurate O(n) Solution to the PnP Problem," *International Journal of Computer Vision*, vol. 81, pp. 155–166, February 2009.

[19] S. Li, C. Xu, and M. Xie, “A Robust O(n) Solution to the Perspective-n-Point Problem,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 1444–1450, July 2012.

[20] Y. Zheng, Y. Kuang, S. Sugimoto, K. Astrom, and M. Okutomi, “Revisiting the PnP Problem: A Fast, General and Optimal Solution,” in *2013 IEEE International Conference on Computer Vision*, pp. 2344–2351, IEEE, December 2013.

[21] P. Besl and H. McKay, “A method for registration of 3-D shapes,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[22] F. Moreno-Noguer, V. Lepetit, and P. Fua, “Pose Priors for Simultaneously Solving Alignment and Correspondence,” in *European Conf. on Computer Vision*, 2008.

[23] D. G. Lowe and D. G., “Distinctive Image Features from Scale-Invariant Keypoints,” *International Journal of Computer Vision*, vol. 60, pp. 91–110, November 2004.

[24] H. Bay, T. Tuytelaars, and L. V. Gool, “SURF : Speeded Up Robust Features,” pp. 404–417, 2006.

[25] B. Zeisl, P. F. Georgel, F. Schweiger, E. Steinbach, and N. Navab, “Estimation of Location Uncertainty for Scale Invariant Feature Points,” in *Procedings of British Machine Vision Conference 2009*, pp. 1–12, 2009.

[26] M. Chli and A. J. Davison, “Active Matching,” in *Computer Vision – ECCV 2008: 10th European Conference on Computer Vision*, pp. 72–85, Springer, Berlin, Heidelberg, October 2008.

[27] A. Handa, M. Chli, H. Strasdat, and A. J. Davison, “Scalable active matching,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1546–1553, IEEE, June 2010.

[28] C.-P. Lu, G. D. Hager, and E. Mjolsness, “Fast and globally convergent pose estimation from video images,” *IEEE Trans. on PAMI*, vol. 22, no. 6, pp. 610–622, 2000.

[29] M. Figueiredo and A. Jain, “Unsupervised learning of finite mixture models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 381–396, 2002.

[30] R. Van der Merwe and E. Wan, “The square-root unscented Kalman filter for state and parameter-estimation,” in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*, pp. 3461–3464, 2001.

[31] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge University Press, 2003.

[32] Z. Zhang and S. Member, “A Flexible New Technique for Camera Calibration,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 11, pp. 1330–1334, 2000.

[33] G. B. Adrian Kaehler, Bradski, Gary, *Learning OpenCV: Computer Vision with the Opencv Library*. 2008.

[34] S. Bennett and J. Lasenby, "ChESS¢: Quick and robust detection of chess-board features," *Computer Vision and Image Understanding*, vol. 118, pp. 197–210, 2014.

[35] J. Heller, M. Havlena, and T. Pajdla, "A branch-and-bound algorithm for globally optimal hand-eye calibration," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1608–1615, 2012.

[36] M. Shah, "Comparing two sets of corresponding six degree of freedom data," *Computer Vision and Image Understanding*, pp. 1355–1362, 2011.

[37] R. I. Hartley and P. Sturm, "Triangulation," *Computer Vision and Image Understanding*, vol. 68, pp. 146–157, November 1997.

[38] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 376–380, April 1991.

[39] D. G. Gobbi, R. M. Comeau, B. K. Lee, and T. M. Peters, "Correlation of pre-operative MRI and intra-operative 3D ultrasound to measure brain tissue shift," in *Medical Imaging*, Proc. SPIE 3982, pp. 77–84, 2000.

[40] R. Van Der Merwe and E. Wan, "The unscented kalman filter for nonlinear estimation," in *IEEE Adaptive Systems for Signal Processing, Communications, and Control Symposium*, pp. 153–158, 2000.

Chapter 3

Accuracy in Freehand 3D-US Reconstruction with Robust Visual Tracking

This chapter is adapted from the papers,

- Uditha L. Jayarathne, Elvis C. S. Chen, John Moore, Terry M. Peters, ”Freehand 3D-US reconstruction with robust visual tracking with application to ultrasound-augmented laparoscopy,” Proc. SPIE 9786, Medical Imaging 2016: Image-Guided Procedures, Robotic Interventions, and Modeling, 978617 (18 March 2016)
- Uditha L. Jayarathne, John Moore, Elvis C.S. Chen, and Terry M. Peters, “Visualizing Ultrasound In the Context of Laparoscopy”, IEEE Transactions on Visualization and Computer Graphics, (*submitted*)

3.1 Introduction

The robust visual tracking method described in the previous chapter enables US images to be transformed to the coordinate system centered in the laparoscopic camera. The next challenge is to present this transformed US information to the surgeon in the context of the surface image provided by the camera, such that the surgeon perceives both the spatial location, and the 3D form of the imaged targets accurately. A popular approach to this visualization problem is to texture map the image to a 2D plane scaled appropriately and placed at the correct spatial pose in front of a virtual camera, with the intrinsics of the virtual camera matching those of the real camera. The surface image is set as the background to provide the context. As discussed in detail in the next chapter, this method of visualization has several issues: (1) the content in the augmented-US image is difficult to interpret at certain probe poses, and (2) perception of 3D form of the imaged targets require mental integration of 2D images across space and time. One strategy to eliminate these issues is to reconstruct a 3D US volume as the probe moves over the region of interest, and to render the volume at the correct spatial pose in the camera coordinate system. The 3D rendering of the volume eliminates the vantage point-based issue, and off-loads expensive cognitive processes minimizing room for error. However, this method requires a very fast reconstruction technique to enable real-time performance. In addition, the reconstruction technique should be accurate, as erroneous reconstructions may result in inaccurate spatial perception leading to surgical errors.

Many 3D US reconstruction techniques exist in the literature, where a predefined voxel grid is filled by US pixels in 2D US images tracked in a common coordinate system. Depending on the manner whereby the voxels in the structured grid are filled, these methods are categorized into three groups: (1) pixel-based, (2) voxel-based, and (3) hybrid methods. In pixel-based methods, each US pixel is transformed to the voxel space using tracking information, and is assigned to the nearest voxel, as in the pixel-nearest-neighbor method [1, 2], or smeared to nearby voxels through a 3D kernel [3, 4]. These methods are sometimes known as *forward methods*. One major issue with pixel-base approaches is that some voxels may never get filled, hence, a *hole-filling* method is required to obtain a high quality reconstruction [1]. Unfortunately, hole-filling is an expensive process, which hinders the running time of the algorithm. Voxel-based methods operate by assigning the nearest US pixel for a given voxel in the grid without any weighting, as in voxel-nearest-neighbor method [5, 6], or with weighing-based on perpendicular distance to the corresponding US image [7, 8]. This method fills all the voxels, hence, the hole-filling step is generally not required. However, this approach requires that all the US images to be available before the reconstruction begins, which renders it unsuitable for real-time applications. Hybrid methods attempt to exploit the best characteristics of these two methods, transforming a fixed number of US images to the voxel coordinate system, and employing a voxel-based approach to fill a region of interest (ROI) in the volume. In this manner, high quality reconstruction can be achieved quickly without employing an expensive hole-filling step.

This chapter describes a hybrid freehand US reconstruction technique that can be implemented in the GPU to achieve real-time performance. It achieves high quality reconstructions without the requirement for a hole-filling step. While any tracking method, either extrinsic or intrinsic, can be used to track US images in 6DoF, in this work the robust, image-based tracking method described in detail in Chapter 2 is employed. The experiments detailed in this chapter help validate the visual tracking method itself, while enabling an investigation of gross errors introduced by the system in representing a hidden target in 3D.

3.2 Methods

In this section, the technical details of the hybrid US reconstruction method are described in detail. This is followed by a description of a phantom-based experiments conducted to demonstrate the geometric accuracy of reconstruction. Throughout the experiments, a clinical laparoscopic probe is employed with a fiducial pattern attached to the curved back surface to enable tracking. The robust, monocular visual tracking method detailed in Chapter 2, is employed to track the probe in 6DoF in the laparoscopic camera coordinate system. The quality of the US reconstruction with visual tracking is demonstrated by scanning a PVA-C phantom that includes lumens and some embedded spherical fiducials. These fiducials allow registration of the US volume with a micro-CT scan of the phantom, enabling the alignment of the lumens localized in the US volume to be compared to that localized in the CT.

3.2.1 Robust Intrinsic Tracking

To track the LUS probe in 6DoF, the robust intrinsic tracking method described in chapter 2 was employed. A fiducial pattern was attached to the curved back surface of a clinical laparoscopic US probe (LAP9-4/38, Ultrasonix, Canada)(Fig. 3.1(a)), and the 3D locations of the fiducials, refered to as *model-points*, were determined accurately using a measurement microscope (STM6-LM, Olympus, Japan), and a coordinate system was defined with respect to an arbitrarily chosen fiducial point. Video frames, captured from the left channel of a stereo laparoscopic camera used in a *da Vinci S* surgical system (Intuitive Surgical Inc., USA) at a resolution 640x480 at 30 frames per second, were streamed to a portable computer via modules included in the PLUS software library [9]. The video image was corrected for barrel and tangential distortions estimated using the OpenCV software library [10]. In each video frame, the 2D location of each fiducial is determined by an efficient method [11] followed by sub-pixel refinement. Model-points, their image-space locations, and the intrinsics determined by a generic camera calibration method [12] were the input to the monoscopic-image-based pose estimation algorithm.

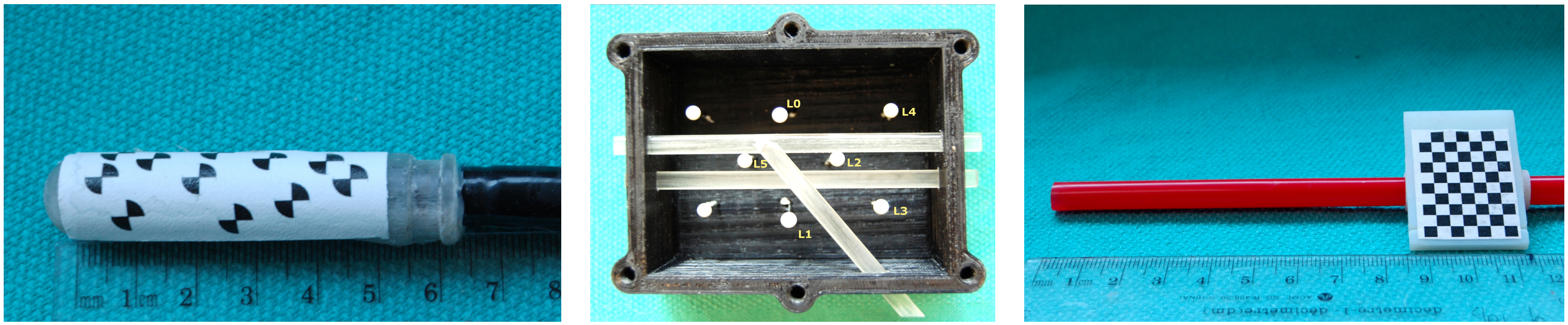
Given the 3D fiducial locations (model-points) $\mathbf{m}_i, i = 1, 2, ..., M$, and detected 2D locations $\mathbf{u}_k, k = 1, 2, ..., N$, the algorithm estimates the model-to-feature correspondence and the 6DoF pose $\mathbf{P}$ of the probe jointly by minimizing the objective in Eq. 3.1 using an Extended Kalman Filter (EKF)-based algorithm;

$$Error(\mathbf{P}) = \sum_{(\mathbf{m},\mathbf{u})\in Matches} ||\mathbf{u} - Proj(\mathbf{P};\mathbf{m})|| + \tau|, NotDetected| \tag{3.1}$$

where $Proj(\mathbf{P}, \mathbf{m}_i)$ is the operator that projects the model point $\mathbf{m}_i$ with the pose p, $|NotDetected|$ is the cardinality of the undetected model points, while γ is a tuning parameter. The reader is referred to the chapter 2 for full details of the algorithm.

The pose that minimizes Eq. 3.1 was further refined by applying an EKF state update step for each matched correspondence. Based on the refined pose, the prior for the subsequent frame was determined by a standard EKF prediction step with an identity process model.

The algorithm was implemented in Matlab and applied to the captured video to obtain the pose estimates offline.



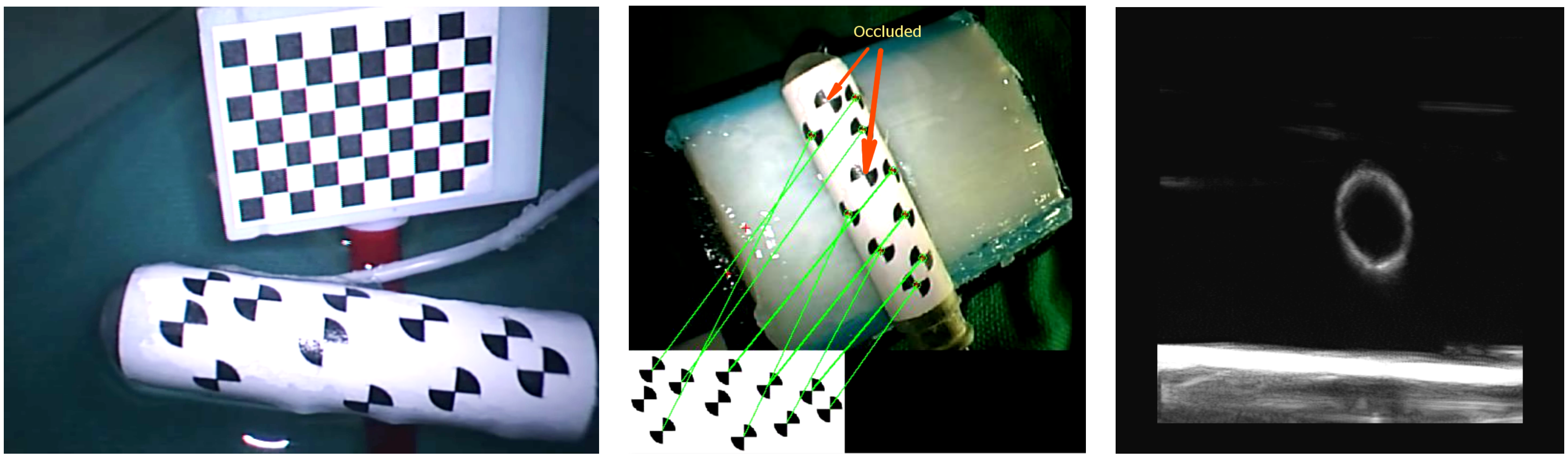
(a) (b) (c)

Figure 3.1: The experimental setup: (a) A fiducial pattern attached to the curved back surface of the laparoscopic probe to enable image-based tracking, (b) The phantom-case prior to adding polyvinyl alcohol(PVA). Rapid prototyped tubes were removed after adding PVA to create open lumens. Plastic-bead fiducials are numbered (L0, L1, etc.) for referencing purposes. (c) Checkerboard pattern attached on the flat surface of the US calibration tool to enable tracking of the straw in the camera coordinate system

3.2.2 US Calibration

The method described in section 3.2.1 allows the probe to be tracked relative to a coordinate system in the camera. To track the US images in this coordinate system, the constant transformation from the US-image-based coordinate system to that based on the tracking fiducials is required. To determine this transform, a novel method that casts the problem as a point-to-line registration problem was employed [13, 14].

A hollow plastic straw was used as the calibration object, whose 3D orientation can be tracked relative to the camera coordinate system using a checkerboard pattern attached to the flat surface of its plastic holder (Fig. 3.1(a) and Fig. 3.2(a)). The holder was designed and 3D printed such that two points on the central axis of the plastic straw are known with respect to a coordinate system on the checkerboard pattern. One could also use a tracked needle, but the hollow straw enables highly accurate localization in the US image resulting in accurate estimates [15]. While the LUS probe was fixed and stationary, the calibration object was moved such that the plastic straw could be imaged at different parts of the US image upto a maximum depth of 4cms. The video image was saved at standard definition (640x480) resolution together with the US image captured at 645x595 resolution. For each camera image, the image-based method described in section 3.2.1 provided the pose of the probe while the pose of the checkerboard pattern was determined by an accurate, iterative algorithm [16] implemented in the OpenCV software library. In each US frame, the centroid of the image of the straw (Fig. 3.2(c)) was manually determined. These centroids, together with the orientation of the plastic straw were fed into the Anisotropic-Scaled ICP(AS-ICP) algorithm (Appendix A), where the point-to-line registration transform in form of anisotropic scales, followed by rotation and translation, is the US calibration we seek.



(a) (b) (c)

Figure 3.2: US calibration using tracked hollow plastic straw: (a) Camera image illustrating an instance in data acquisition for US calibration. The plastic straw and the US probe is tracked by using the fiducial patterns attached. (b) The pose estimation algorithm has accurately determined the 3D-2D feature correspondence even if some 2D features are occluded while some clutter is present. (c) Captured US image used to determine the US calibration transform. Note the clear elliptical image of the straw phantom centroid of which can be easily determined easily and accurately

3.2.3 Realtime 3D Ultrasound Reconstruction

With the pose of the US images determined by the concatenation of the transforms determined by the methods described in sections 3.2.1 and 3.2.2, they are stitched to a 3D volume by using the hybrid reconstruction technique described by Ludvigsen [17]. In this method, the US images are approximated by 2D planes, with their poses known with respect to a coordinate system centered in the camera. The rectangular extent of the 3D volume is determined *a-priori* using a scout scan that can be acquired quickly even for complex anatomy, since image quality is inconsequential. In practice, the region of interest (ROI) determined by scout is small (up to about 7 cms) for target sizes associated with the intended laparoscopic applications.

The W most recent 2D US images with their corresponding poses are accumulated into a buffer. Using the pose of the US image, we transform three points, $\mathbf{p}_0^i$, $\mathbf{p}_1^i$ and $\mathbf{p}_2^i$, corresponding to three corners in the i^{th} image, to obtain their coordinates $\mathbf{P}_0^i, \mathbf{P}_1^i$ and $\mathbf{P}_2^i$ in the coordinate system centered in the camera. Triplets of these points define a unique set of planes in 3D (Fig. 3.3(a)) given by,

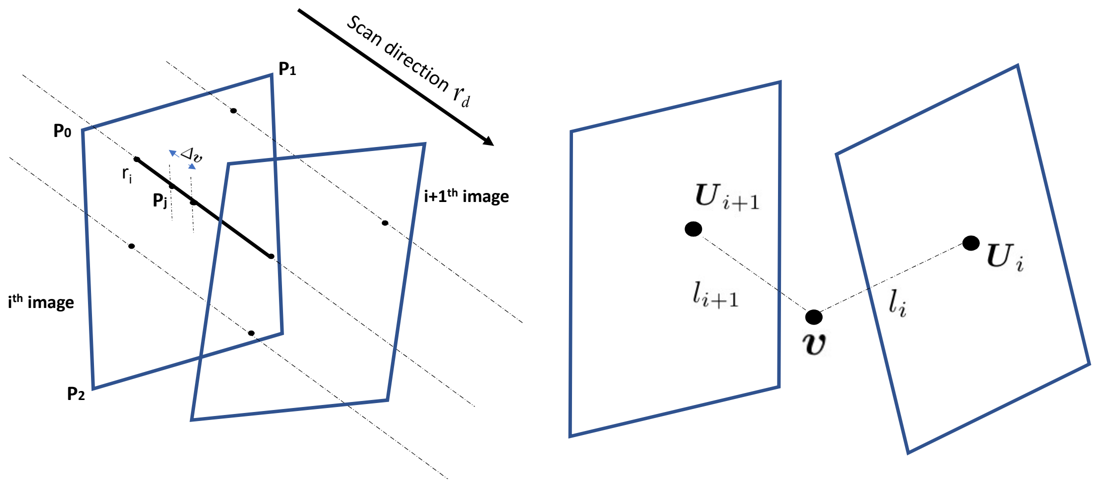
$$a_i X + b_i Y + c_i Z + d_i = 0 \tag{3.2}$$

where $a_i = n_x, b_i = n_y, c_i = n_z, d_i = -\mathbf{n}_i \cdot \mathbf{P}_0$, and $\mathbf{n}_i$ is the normal vector to the i^{th} plane. Given the planar approximation of US images, the points between two adjacent image planes lie on rays, starting at pixel point $\mathbf{r}_i$ on the i^{th} image plane in the direction $\mathbf{r}_d$ of the other image, with their coordinates are given by

$$\mathbf{P}_{ij} = (\mathbf{r}_i + t\mathbf{r}_d)/\Delta\mathbf{v} \tag{3.3}$$

where the scalar $t = -((a_i, b_i, c_i) \cdot \mathbf{r}_0 + d)/((a_i, b_i, c_i) \cdot \mathbf{r}_d)$ and $\Delta\mathbf{v}$ is the voxel spacing in the output volume.

The intensity of a voxel with coordinates given by equation (3.3) receives contribution from the closest pixels in W adjacent images. A distance weighted orthogonal projection scheme [7](Fig. 3.3 (b)) is used to determine intensity contributions: Let $\mathbf{U}_i$ be be the pixel location in the i^{th} image closest to the voxel $\mathbf{P}_{ij}$, and let l_i be the distance. Let I_i be the pixel intensity at $\mathbf{U}_i$



(a) (b)

Figure 3.3: (a) 2D US images are represented by their planer equations using three points, and (b) the intensity value of a voxel between US scans is determined based on the distance weighted orthogonal projection method

determined using bilinear interpolation involving four surrounding pixels. The intensity of the voxel $\mathbf{P}_{ij}$ is the distance weighted intensity average given by,

$$\mathbf{I} = \frac{\sum l_i I_i}{\sum l_i} \tag{3.4}$$

where the summation is over W adjacent images.

It is often the case that the same voxel is visited multiple times during scanning. In such situations, we alpha-blend the existing intensity value of the voxel with the new one, eliminating the need for a memory intensive accumulation buffer.

Implementation

To enable easy integration with the rendering pipeline detailed in chapter 4, and to achieve real-time frame-rates, the 3D reconstruction algorithm is implemented in C/C++ as a VTK filter with GPU acceleration. In the following experiments, a window of four US images is used ($W = 4$) to improve the reconstruction quality. The isotropic output voxel resolution was set to be 0.5 *mm*. When the voxels are visited multiple times, alpha-blending was used with the alpha value set to 0.7.

3.2.4 Experiments

The geometric accuracy of the reconstruction method was demonstrated by imaging a phantom that mimics soft-tissue. The phantom, a 80mm x 50mm x 50mm block of polyvinyl alcohol cryogel (PVA-C), underwent one freeze-thaw cycle, and was equipped with two embedded lumens, one 20mm from the base, the other 30mm. The latter lumen bifurcated near the center of the block, at a 50 degree angle. Eight plastic beads mounted on pins, were placed randomly around the lumens (Fig. 3.1(b)), to be used as fiducial markers to register a CT-image of the phantom to the US volume. An ellipsoidal geometry was used for the lumens (2mm high, 5mm wide), to represent collapsed bronchi or blood vessels under probe pressure. These lumens were water-filled during imaging. A 5mm thick boundary was added to the inside walls and the base to improve US image quality by damping ultrasound reflection artifacts. Finally, a micro-CT of the phantom was obtained at 0.154mm isotroic resolution to provide the reference for structures visible in the US image.

The phantom was imaged by the LUS probe while its motion was captured by the laparoscopic camera at a resolution 640x480 at 30 frames per second. Corresponding US frames were saved at the same frame rate. The 6DoF pose of the US probe was determined for each frame offline by the robust, image-based tracking method described in section 3.2.1, and with concatenation with the US calibration, the pose of the US captured US images were determined relative the camera coordinate system. The poses and the US images were the input to the freehand US reconstruction method implemented in OpenCL as detailed in section 3.2.3 to run in the GPU. A computer with an Intel Core i7 processor, 8 GBs of RAM, a GTX 1070 GPU running Microsoft Windows 10 was used for the experiments.

In the reconstructed US volume, the distance between the plastic beads localized manually in the 3D US volume were compared to the reference provided by the micro-CT to assess geometric distortion during reconstruction. These fiducials were then used to align the 3D US volume with the micro-CT volume through a rigid transform, using the Landmark Registration module in 3D Slicer (`www.slicer.org`). The lumens were manually segmented in both volumes (3D US and micro-CT) and the centerlines were determined by using a skeletonization algorithm [18] followed by manual adjustments. The voxels in the 3D US centerlines were then transformed to micro-CT coordinate system using the rigid transformation, and the closest Euclidean distance to micro-CT centerline voxels was determined.

3.3 Results

A sample laparoscopic camera image is shown in Fig. 3.2(b), with the correspondence accurately solved by the image-based pose estimation method. Note that some fiducials have not been detected while some outliers have been picked by the feature detector. The pose estimation algorithm has nevertheless determined the correct correspondence, and the pose, regardless of these imperfections in the feature detector. The US calibration using the poses determined by this method resulted in Fiducial Registration Error (FRE) of 0.57mm for 20 US frames corresponding to different poses of the plastic straw.

A Direct Volume Rendering (DVR) of the reconstructed 3D US volume using a one-dimensional transfer function is shown in Fig. 4.5(a). Note that the lumens are clearly visible while the strong US reflections from the plastic beads can be easily identified in this rendering. The 15 distance measurements computed between the plastic beads localized in the 3D US volume differs from those measured in the micro-CT volume by 2.27mm on average (Table 3.1). The FRE of the rigid registration between the US volume and the micro-CT volume was calculated to be 1.38*mm*. Fig. 4.5(b) qualitatively represents the accuracy of this transformation. Note how the lumen walls in the US image align with the corresponding micro-CT structures.

Fig. 3.6 shows the Euclidean distance error between the centerlines of the tubular structures after registration. This is equivalent to the target registration error (TRE), which measures the true registration error, hence is clinically very important. The color in every point in the lumen wall in Fig. 3.6 indicates the error associated with the closest centerline voxel. Note the increase in the error as the probe moves further away from the camera. The mean and the standard deviation of the distance between the two centerlines were 1.52*mm* and 1.14*mm* respectively (Table 3.1) while its maximum was 4.82*mm*. The histogram of centerline alignment errors is also shown in Fig. 3.4.

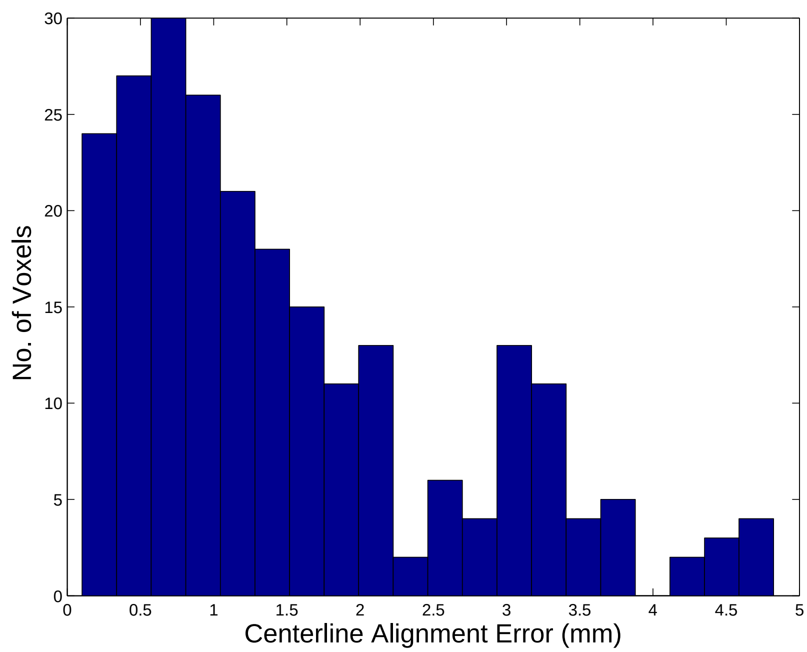
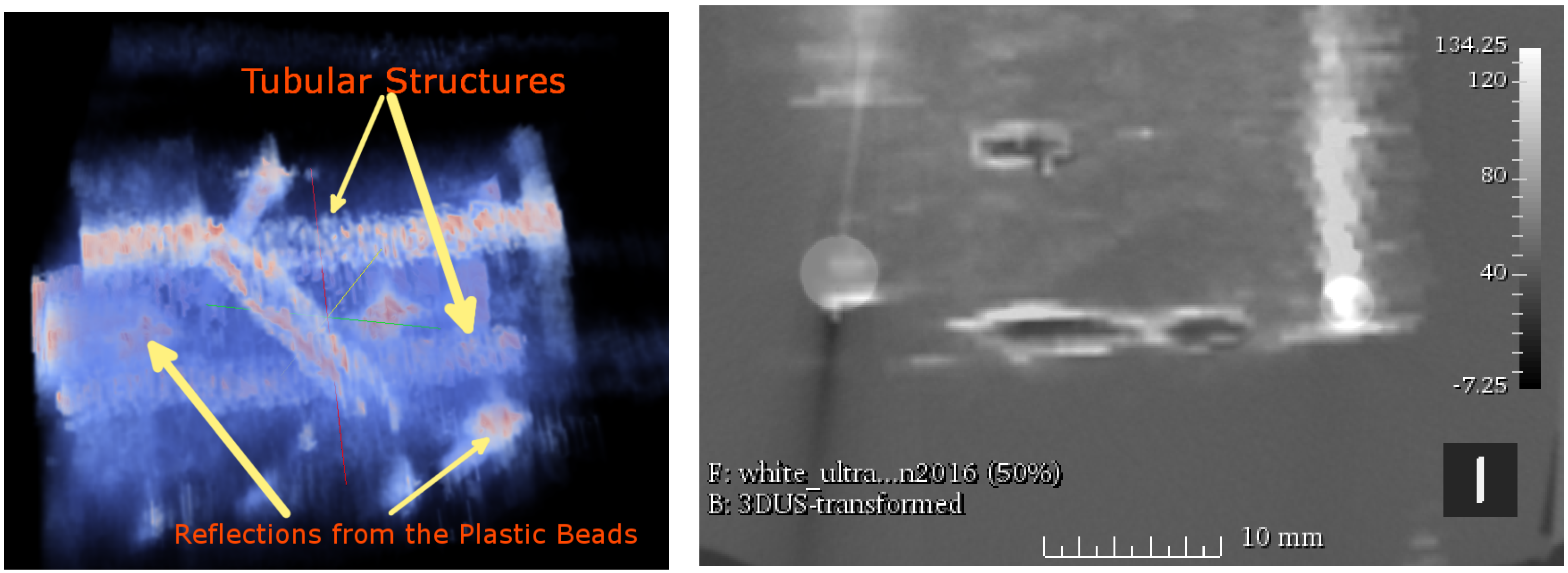


Figure 3.4: Histogram of distances (in milimeters) between the centerlines of the tubular structures extracted from the micro-CT scan and the registered 3D US volume

	N	Mean(mm)	STD(mm)	95% Confidence Intervals (mm)
Fiducial Distance Error	15	2.27	1.09	[1.67, 2.87]
Centerline Distance Error	239	1.52	1.14	[1.37, 1.66]

Table 3.1: Fiducial distance error and the centerline distance error after the registration



(a) (b)

Figure 3.5: Geometric accuracy of the freehand US reconstruction: (a) Direct Volume Rendered (DVR) US volume of the phantom. The tubular structures and the reflections from the plastic-bead fiducials are clearly visible, (b) Image showing an overlay of transformed US data (shown in white in the fore-ground) on micro-CT data (shown in gray in the background)

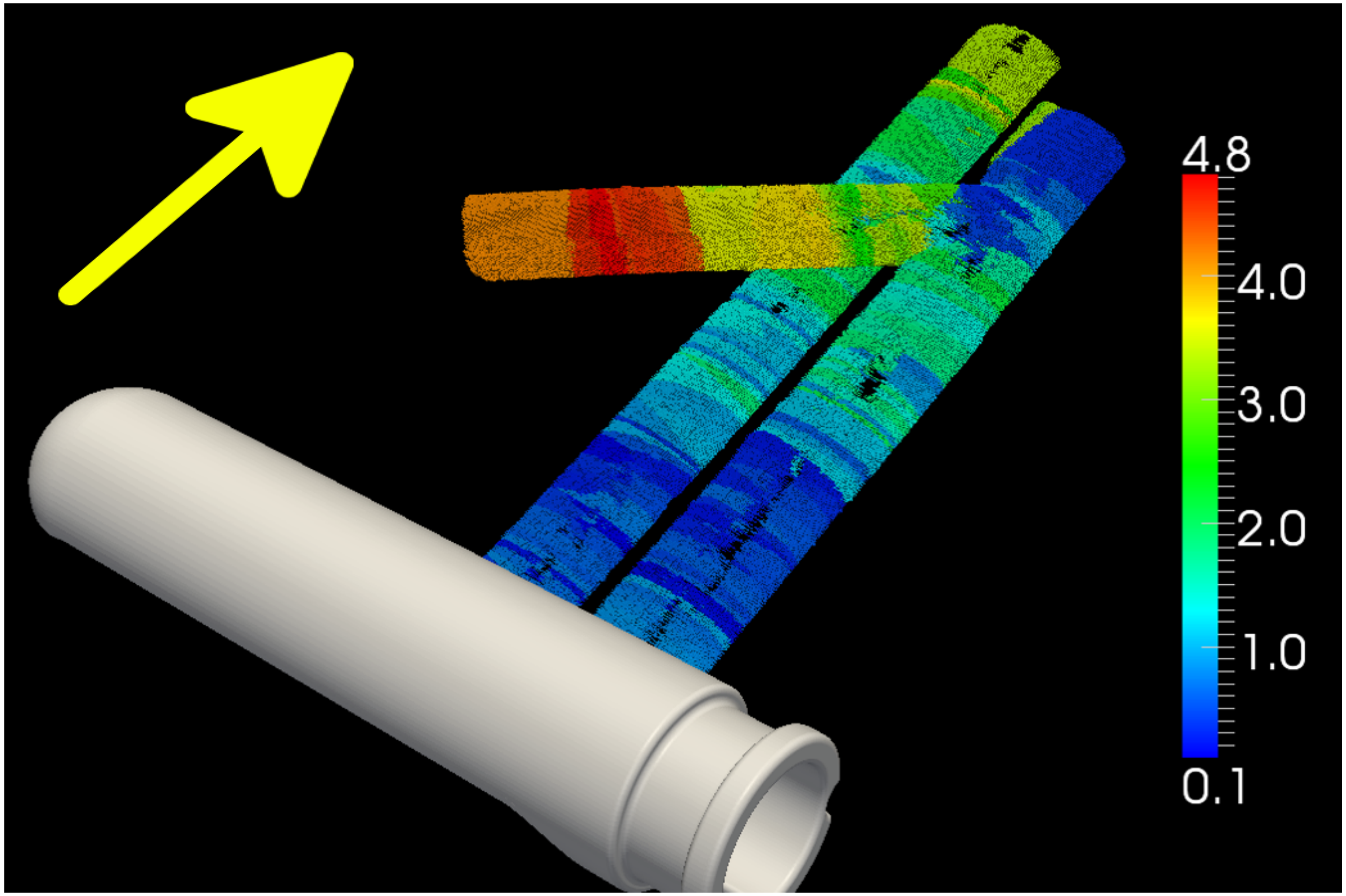


Figure 3.6: Centerline alignment error projected to the nearest voxel on the lumen wall: The scan direction is indicated by the arrow. Note the increase in the alignment error as the probe move further away from camera

3.4 Discussion

This chapter describes the investigation of the accuracy of freehand US reconstruction with the robust, visual tracking method detailed in chapter 2 using a PVA phantom that mimic anatomical structures. A 3D US volume reconstructed with a GPU-accelerated hybrid algorithm is registered to a micro-CT scan of the phantom via manually localized fiducials. The distances between these fiducials are calculated in each volume to obtain a quantitative measure on the geometric accuracy of the reconstruction. After registration, the distance between the centerlines of the tubular structures of the phantom is computed providing another quantitative measurement on the reconstruction quality. In addition, an overlay visualization of US and micro-CT images, and DVR visualization of the reconstructed US volume obtained by applying the estimated registration provides qualitative measures.

In addition to the errors in probe tracking and calibration, a significant contributor to the TRE is manual localization of plastic bead fiducials in the US volume. Strong US reflections from these fiducials render pin-pointing their location difficult and error-prone, as reflected in the reported FRE of the US and micro-CT registration. In addition, the deformation of the phantom due to probe pressure could also have an effect on the TRE.

The error map in Fig. 3.6 has an increasing trend as the probe moves away from the camera (see Fig. 3.1(b) for details on how the phantom was placed in front of the camera), and this tendency could be due to the drop in accuracy of pose estimates along the positive Z-axis of the camera and due to inaccurately modeled lens distortions. Significant improvements to pose estimates can be obtained with multi-view cameras as described in chapter 2, while errors due to inaccurately modeled lens distortion can be reduced by employing non-parametric distortion model [19].

The reconstruction method detailed in this chapter, stitches 2D US images into a 3D volume in real-time. Since the US images are tracked relative to the camera coordinate system, the pose of the reconstructed volume is intrinsically known in this coordinate system. Therefore, the volume can be rendered in the camera coordinate system enabling visualization of hidden anatomy in 3D in the context of the laparoscopic camera image. Technical details of achieve such visualization, and a study of human factors associated are dealt within the following chapter.

3.5 Conclusions

In this chapter, geometric accuracy of freehand, real-time 3D US reconstruction with visual tracking is investigated with a phantom study. The results suggest that errors up to four millimeters can be expected with the system, while an increase an error could be expected when the probe moves away from the camera center. In addition, deformations due to the probe pressure may have a significant contribution. The experiment helps in estimating the average gross error the system introduces in representing a hidden target in 3D when the image-based method proposed in Chapter 2 is used for localization.

Bibliography

[1] D. G. Gobbi and T. M. Peters, "Interactive intra-operative 3d ultrasound reconstruction and visualization," in *Medical Image Computing and Computer-Assisted Intervention — MICCAI 2002: 5th International Conference Tokyo, Japan, September 25–28, 2002 Proceedings, Part II* (T. Dohi and R. Kikinis, eds.), pp. 156–163, Berlin, Heidelberg: Springer Berlin Heidelberg, 2002.

[2] T. R. Nelson and D. H. Pretorius, "Three-dimensional ultrasound imaging," *Ultrasound in Medicine and Biology*, vol. 24, no. 9, pp. 1243 – 1270, 1998.

[3] R. Rohling, A. Gee, and L. Berman, "A comparison of freehand three-dimensional ultrasound reconstruction techniques," *Medical Image Analysis*, vol. 3, no. 4, pp. 339 – 359, 1999.

[4] R. S. Jos-Estpar, M. Martn-Fernndez, P. Caballero-Martnez, C. Alberola-Lpez, and J. Ruiz-Alzola, "A theoretical framework to three-dimensional ultrasound reconstruction from irregularly sampled data," *Ultrasound in Medicine and Biology*, vol. 29, no. 2, pp. 255 – 269, 2003.

[5] S. Sherebrin, A. Fenster, R. N. Rankin, and D. Spence, "Freehand three-dimensional ultrasound: implementation and applications," in *Proceedings of SPIE Medical Imaging*, vol. 2708, pp. 2708 – 2708 – 8, 1996.

[6] R. W. Prager, A. Gee, and L. Berman, "Stradx: real-time acquisition and visualization of freehand three-dimensional ultrasound," *Medical Image Analysis*, vol. 3, no. 2, pp. 129 – 140, 1999.

[7] J. W. Trobaugh, D. J. Trobaugh, and W. D. Richard, "Three-dimensional imaging with stereotactic ultrasonography," *Comp. Med. Imag. and Graph.*, pp. 315–323, 1994.

[8] P. Coupé, P. Hellier, N. Azzabou, and C. Barillot, "3d freehand ultrasound reconstruction based on probe trajectory," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005: 8th International Conference, Palm Springs, CA, USA, October 26-29, 2005, Proceedings, Part I* (J. S. Duncan and G. Gerig, eds.), pp. 597–604, Berlin, Heidelberg: Springer Berlin Heidelberg, 2005.

[9] A. Lasso, T. Heffter, A. Rankin, C. Pinter, T. Ungi, and G. Fichtinger, "PLUS: Open-Source Toolkit for Ultrasound-Guided Intervention Systems," *IEEE Trans. on Biomed. Engineering*, pp. 2527–2537, 2014.

[10] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[11] S. Bennett and J. Lasenby, "Chess: Quick and robust detection of chess-board features," *Computer Vision and Image Understanding*, vol. 118, pp. 197–210, 2014.

[12] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 11, pp. 1330–1334, 2000.

[13] E. C. Chen, A. J. McLeod, J. S. Baxter, and T. M. Peters, "An iterative closest point framework for ultrasound calibration," in *MICCAI 2015 Workshop on Augmented Environments for Computer-assisted Intervension*, pp. 69–79, 2015.

[14] E. C. S. Chen, A. J. McLeod, J. S. H. Baxter, and T. M. Peters, "Registration of 3d shapes under anisotropic scaling," *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, pp. 867–878, Jun 2015.

[15] G. Ameri, A. J. McLeod, J. S. H. Baxter, E. C. S. Chen, and T. M. Peters, "Line fiducial material and thickness considerations for ultrasound calibration," in *Proceedings of SPIE Medical Imaging*, pp. 941529–941529–9, 2015.

[16] C.-P. Lu, G. D. Hager, and E. Mjolsness, "Fast and globally convergent pose estimation from video images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 610–622, 2000.

[17] H. Ludvigsen, "Real-Time GPU-Based 3D Ultrasound Reconstruction and Visualization," Master's thesis, Norwegion University of Science and Technology, 2010.

[18] R. L. K. Ta-Chih Lee and C.-N. Chu, "Building skeleton models via 3-D medial surface/axis thinning algorithms.," *Computer Vision, Graphics, and Image Processing*, pp. 462–478, 1994.

[19] R. Hartley and S. B. Kang, "Parameter-Free Radial Distortion Correction with Center of Distortion Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1309–21, 2007.

Chapter 4

Visualizing Ultrasound In the Context of Laparoscopy

This chapter is adapted from the papers,

- Jayarathne U.L., Moore J., Chen E.C.S., Pautler S.E., Peters T.M. (2017) Real-Time 3D Ultrasound Reconstruction and Visualization in the Context of Laparoscopy. In: Descoteaux M., Maier-Hein L., Franz A., Jannin P., Collins D., Duchesne S. (eds) Medical Image Computing and Computer-Assisted Intervention MICCAI 2017. MICCAI 2017. Lecture Notes in Computer Science, vol 10434. Springer, Cham

- Uditha L. Jayarathne, John Moore, Elvis C.S. Chen, and Terry M. Peters, "Visualizing Ultrasound In the Context of Laparoscopy", IEEE Transactions on Visualization and Computer Graphics, (*submitted*)

4.1 Introduction

Chapters 2 and 3 discuss methodologies to bring 2D US into the camera coordinate system, and capture the 3D geometry of hidden targets by stitching 2D US images into a volume. In addition, these chapters discuss the expected accuracy bounds in presenting a hidden target in 3D. Methods to render this information in the context of the laparoscopic video, and how different rendering methods affect the perception of the location of the target are discussed in this chapter. As formalized in the Introduction, I use the term *hybrid in situ* to describe visualizations where US information is rendered in the context of the laparoscopic image.

An effective *hybrid in situ* visualization strategy requires efficient solutions to two problems: (1) mapping US image pixel information from the US image coordinate system to that based on the laparoscopic camera, and (2) presenting US image information in the context of the laparoscopic video such that the surgeon perceives the spatial location and the 3D form of imaged targets accurately. State-of-the art methods to solve the problem (1), and a novel image-based method that eliminates the need for extrinsic tracking systems is detailed in chapter 2. Interestingly, very few attempts to solve the visualization problem (2), can be found in the literature. The focus of this chapter is to propose a practical solution to this problem, and assess it performance through a series of perceptual experiments.

4.2 Related Work

The most common strategy to render US data transformed to the laparoscopic camera frame of reference is to texture map the US image to a 2D plane placed at the correct spatial position, orientation, and scale [1–4] in the camera frame of reference. The laparoscopic camera image is set as the background texture to provide the context. The textured US image plane is then blended with the background by adjusting its transparency to avoid complete occlusion of the scene by the US image. This method of rendering conveys ambiguous depth cues resulting in perception that the US image is in front of the rest of the scene. To remedy this, Hughes-Hallett et al. [5] overlaid the US image on an inner surface of a cube that moves with the probe, allowing the surgeons to appreciate depth more easily. However, the efficacy of this approach has not been investigated experimentally. In addition, the interpretation of the content in the US image in these schemes becomes difficult at certain probe poses, particularly when the rendered image plane makes a nearly perpendicular angle with the camera imaging plane. In many laparoscopic applications including laparoscopic partial adrenalectomy [6] and in thoracoscopic localization of pulmonary nodules using LUS in VATS [7], such situations are common due to the standard placement of laparoscopic ports. An overlaid single image may not convey much information in these situations.

Very few attempts have been made to render 3D US information reconstructed from 2D US images in the context of the laparoscopic video. In a recent article, Oh et al. [8] presented a method where an US volume is reconstructed by mechanically moving the probe, and the maximum intensity projection (MIP) of the volume is overlaid in the laparoscopic video. The mechanical assembly proposed in this work adds logistical overhead in practical surgery. Moreover, the volume reconstruction takes several seconds which may not be practical in abdominal surgery involving moving soft-tissues. In addition, the *alpha blended* volume appears to convey inaccurate depth cues.

4.2.1 Contributions

In this chapter, a visualization strategy to eliminate issues in the existing *hybrid in situ* visualization techniques is presented. This allows the surgeon to interactively visualize hidden targets in 3D, in contrast to the cognitively demanding method of 3D form perception through mental integration of 2D cross-section images cross space and time [9]. In contrast to similar US visualization techniques [10], the proposed method reconstructs a full 3D US volume of the scanned anatomy in real-time. This volume can be used to register pre-operative images, allowing pre-operative plans to be brought into the surgical scene if required. The GPU accelerated implementation of the proposed method runs in real-time, and is fully compatible with the Visualization Toolkit[1] (VTK). In addition, a perceptual study involving experienced US users is conducted to investigate efficacy of the proposed visualization scheme, and compare its performance to the conventional method of visualizing US in a separate display.

[1] http://www.vtk.org/

4.3 Methodology

In this section, technical details of the visualization scheme are described. The pose of the US probe with respect to a camera-centered coordinate system is assumed to be known by an intrinsic or an extrinsic tracking technology. In addition, the constant transformation between the US image coordinate system to that centered in the tracking sensor/marker, known as the US calibration transformation, is assumed to be known through an appropriate calibration method. The pose of the probe coupled with the US calibration transform allows the US image to be mapped to the coordinate system centered in the camera. When the surgeon moves the probe over tissue to explore the underlying 3D anatomy, the tracked 2D US images are stitched into a 3D volume in the camera coordinate system. The 3D reconstruction algorithm is implemented in the GPU to enable real-time performance. The reconstructed volume is then rendered by using an efficient ray-casting algorithm, which is also implemented on the GPU to accelerate computations. The rendered image of the 3D US volume is then blended with the laparoscopic camera image providing the surface view of the organs, using two different blending methods; one *naïve alpha blending* method, the other a distance dependent opacity function. This transparency function is also modulated by high frequency information in the surface image. These blending schemes are described in detail in Section 4.3.2. The proposed visualization scheme is compared to the conventional method of viewing US, by conducting a perceptual study involving experienced US users using the setup detailed in Section 4.3.3.

4.3.1 Real-time 3D US Reconstruction

The visualization pipeline employs the high quality 3D ultrasound reconstruction algorithm detailed in Chapter 3, implemented in the GPU to achieve real-time performance. The US images are approximated by 2D planes, with their pose known with respect to a coordinate system centered in the camera. The rectangular extent of the 3D volume is determined *a-priori* using a scout scan that can be acquired quickly, even for complex anatomy, since image quality is inconsequential. In practice, the region of interest (ROI) determined by the scout is small (up to about 7 cms) for target sizes associated with the intended laparoscopic applications.

For each four US frames captured, the voxel grid is updated using a distance weighted orthogonal projection scheme [11]. Thus, as the surgeon scans an organ of interest, a 3D US volume is incrementally filled. It is often the case that the same voxel is visited multiple times during scanning. In such situations, the existing intensity value of the voxel is alpha-blended with the new one, eliminating the need for a memory intensive accumulation buffer. For a detailed description of the reconstruction method, the reader is referred to the Chapter 3.

4.3.2 Blending the US Volume with the Camera Image

As the surgeon move the probe on the target region, the 2D US images are captured and a 3D volume is reconstructed. For technical details of the algorithm the reader is refered to the Chapter 3. The objective is to visualize the targets of interest captured by this volume in the context of the laparoscopic image. Since the pose of each 2D US image is determined with respect the camera coordinate system, the pose of the reconstructed volume is intrinsically determined in this coordinate system. For every US image captured, the 3D volume is updated, and rendered using a GPU-accelerated, efficient ray-casting method [12]. A simple one-dimensional transfer function is used in this work while higher-dimensional transfer functions can be easily integrated. The volume is rendered in front of a virtual camera whose intrinsic parameters match with those of the laparoscopic camera. The resulting 2D image is blended with the laparoscopic camera image resulting in an US-augmented image (Fig. 4.2(d)).

Two blending schemes are used. In the first, *naïve blending scheme*, the output of ray-casting is blended with the camera image with a constant alpha value for every pixel. In this scheme, the output pixel value $\mathbf{I}_{rgb}$ is given by,

$$\mathbf{I}_{rgb} = \alpha \mathbf{I}^{v}_{rgb} + (1-\alpha)\mathbf{I}^{c}_{rgb} \tag{4.1}$$

where $\mathbf{I}^{v}_{rgb}$ and $\mathbf{I}^{c}_{rgb}$ are intensity in the rendered volume, and the camera image respectively.

In the second scheme, refered to as *keyhole blending scheme*, the rendered volume is blended with the camera image through a circular opacity window, or *keyhole*. Inside the keyhole the opacity changes as a function of the Euclidean distance from the centre (Fig.4.1), while the opacity outside the keyhole is saturated making the surface image completely opaque. The edge map of the camera image, obtained with the *sobel* operator, modulates the opacity inside the window to yield the opacity function depicted in Fig. 4.2(c). The final image is the alpha-blending (Eq. 4.1) of the rendered volume (Fig. 4.2(a)) and the camera image (Fig. 4.2(b)), with alpha value determined by the opacity function in Fig. 4.2(c). This blending scheme results in the image shown in Fig. 4.2(d). Note the improved perception of depth ordering as a result of edge features inside the keyhole. In the literature, similar methods have been described by several other authors [13–15] to blend volumetric images (CT/MRI) with a camera image.

Implementation

For every camera and US image captured, the US volume is rendered to the texture memory of the GPU, to which the camera image is also copied. Edge response and opacity computations are performed by fragment shader programs implemented in C/C++ as part of a render-pass in VTK, and the output of the shader programs is rendered to a monoscopic/stereoscopic display. The alpha value in the *naïve blending* scheme is set to be 0.7 in all the following experiments.

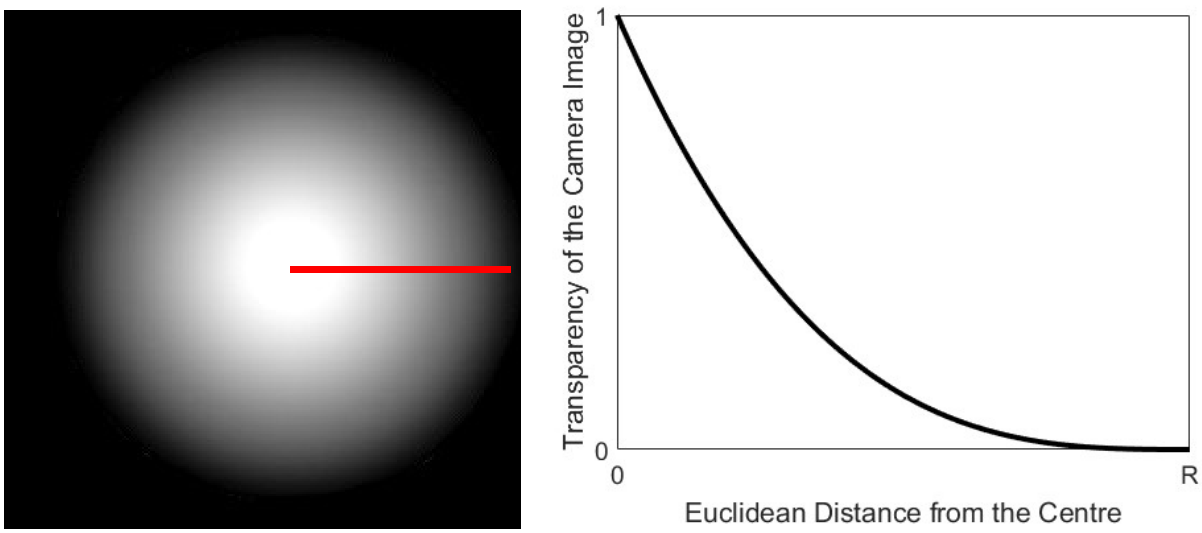


Figure 4.1: Distance dependent transparency function inside a circular region. Function values for pixels through the red line are shown on the right. Note that the full transparency corresponds to the pixel in the center

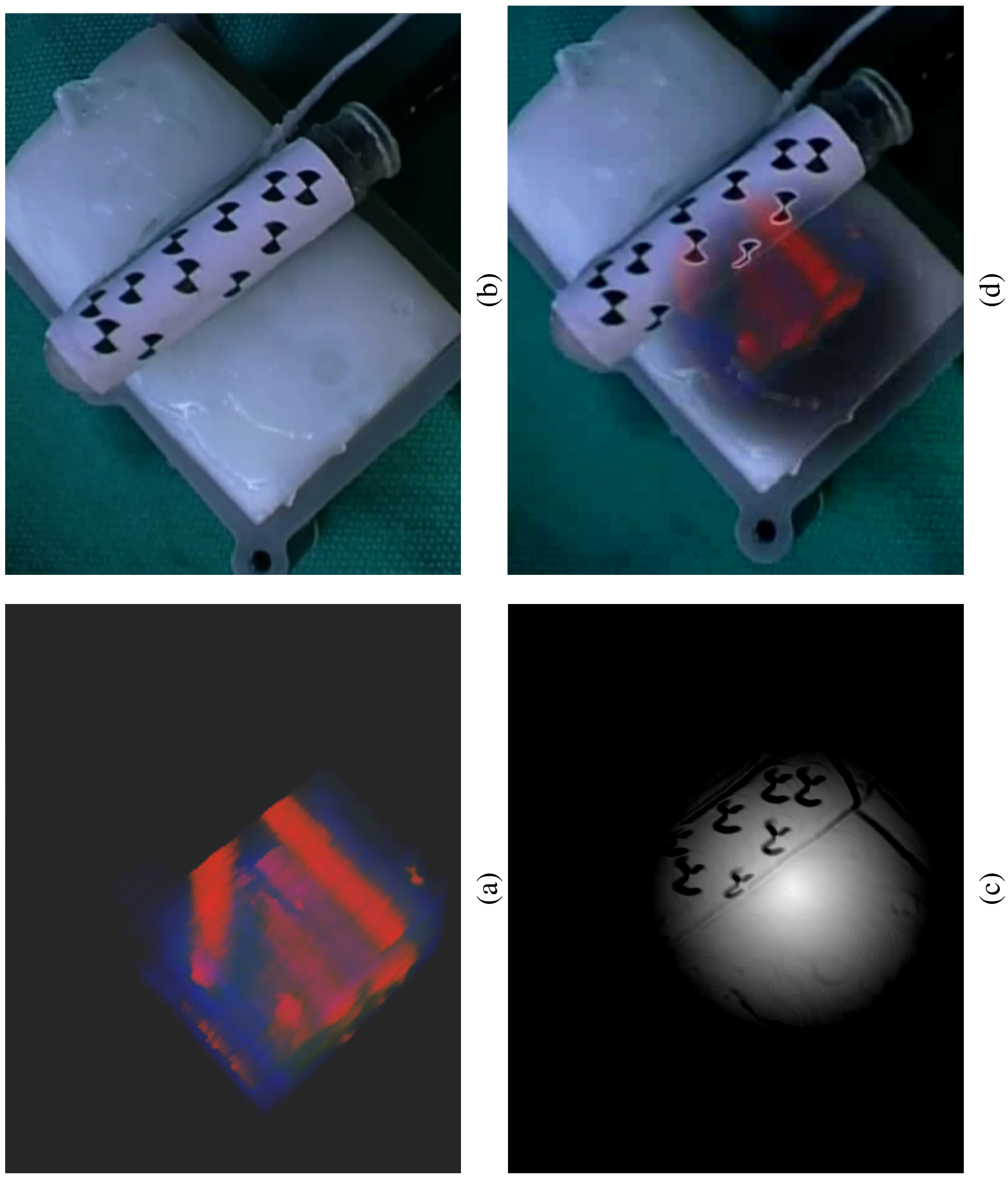


Figure 4.2: *keyhole blending* scheme: The rendered US volume (a), is blended with the surface image provided by the laparoscopic camera (b), using the transparency function in (c) to obtain the blended image in (d). In (c) white represents complete translucency

4.3.3 Experimental Setup

The efficacy of the *hybrid in situ* visualization scheme, detailed in Section 4.3.2 is studied, by conducting a perceptual study involving phantoms. The experimental configuration comprised of a stereoscopic laparoscopic camera (Olympus, Japan) providing the surface image, an ultrasound machine with a hand-held linear probe (Ultrasonix, Analogic, Canada) providing tomographic image, an optical tracking system (Vicra, Northern Digital Inc., Canada) tracking the imaging devices and tools, a passive stereoscopic display system (VisionSense, PA, USA) enabling both monoscopic and stereoscopic visualization, and a portable computer providing the required computational power (Fig. 4.3). Retro-reflective fiducials were attached to the camera, linear US probe, phantoms, and a pointing-tool to enable tracking of their pose in six degrees of freedom (6DoF) by the optical tracking system. The constant transformation ${}^{c}\mathbf{T}_{DRBc}$ from the fiducial coordinate system to that centered in the camera, known as hand-eye calibration, was determined accurately using a Procrustean Perspective-n-Point solution [16]. In addition, US calibration ${}^{DRBu}\mathbf{T}_{us}$, the transformation from the fiducial coordinate system on the probe to the US image, was determined using an algorithm that cast the problem as a point-to-line registration [17]. These calibrations, together with the poses read from the optical tracking system, allow the pose of the US image ${}^{c}\mathbf{T}_{us}$ to be computed in the camera coordinate system,

$$ {}^{c}\mathbf{T}_{us} = {}^{c}\mathbf{T}_{DRBc}.\left[{}^{w}\mathbf{T}_{DRBc}\right]^{-1}.{}^{w}\mathbf{T}_{DRBu}.{}^{DRBu}\mathbf{T}_{us} \tag{4.2} $$

where ${}^{w}\mathbf{T}_{DRBc}$, and ${}^{w}\mathbf{T}_{DRBu}$ are 4x4 matrices representing the pose of the dynamic reference body (DRB) on the camera, and the US probe respectively. The experiments were conducted with the subject seated (Fig. 4.3) while a surgical drape blocks their direct line-of-sight to the action site. The display and the US machine were placed in front of the subject, mimicking the setup in surgery. The laparoscopic images, captured at 800x600 resolution, together with the US image were streamed using the PLUS software library [18] to the portable computer with an Intel Core i7 processor, 32 GB RAM and a Quadro K5000 GPU, running Microsoft

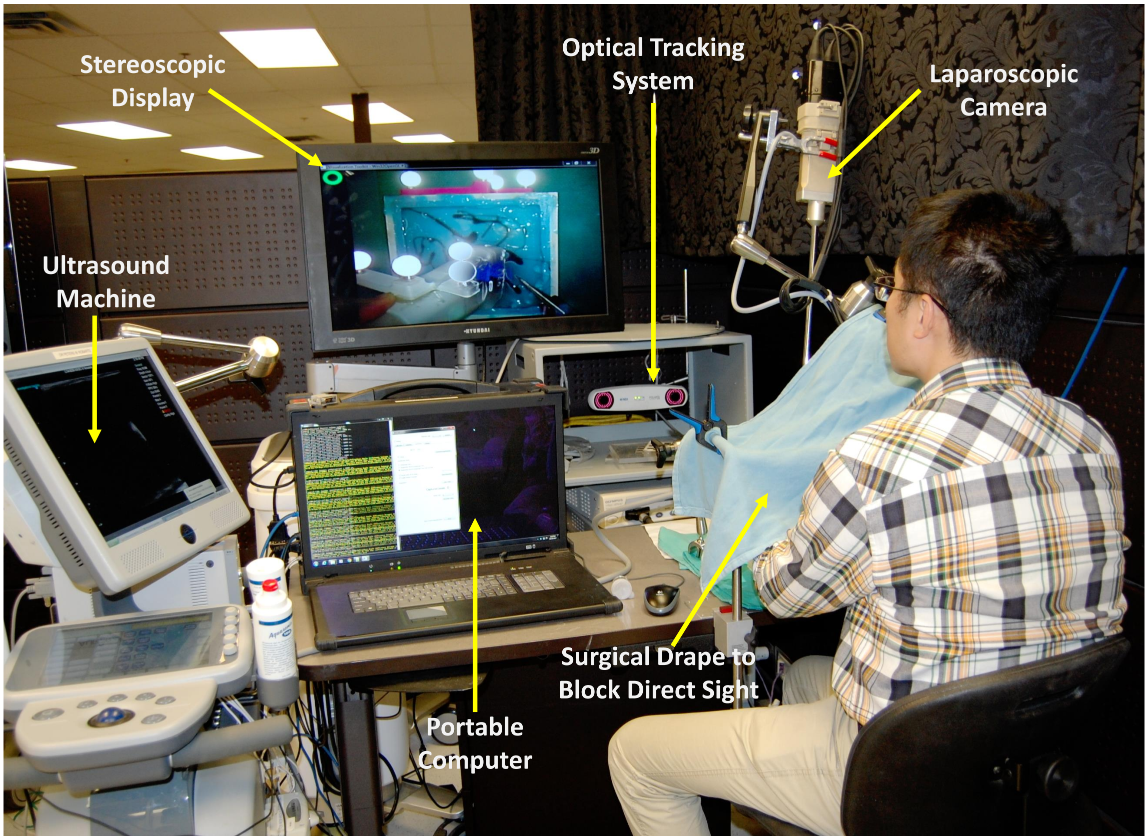


Figure 4.3: The experimental setup: The laparoscopic and the ultrasound videos are captured by the portable computer, processed and the output is displayed in the passive stereoscopic monitor in front of the subject. This monitor enables both monoscopic and stereoscopic visualization while polarizing glasses are required for stereoscopic viewing. The subjects direct sight was occluded by a surgical drape to avoid any biases. An optical tracking system was used to track the pose of the US probe, camera, and the phantom in a common coordinate system allowing US to be rendered in the context of the laparoscopic image, and enabling perceived target location to be compared to the CT-based reference

Windows 7. The portable computer composed the final image, with or without US overlay, to be rendered in the display in monoscopic/stereoscopic mode depending on the visualization mode.

Six identical box phantoms, inner space measuring 10 *cm* x 10 *cm* x 5 *cm* (LxWxH), were 3D printed using an Ultimaker 2e[2] 3D printer. In each box, eight 6.35 *mm* of diameter hemispherical divots, used for landmark based registration, surrounded the outer walls, with a DRB

[2] www.ultimaker.com

mounted to one wall to enable 3D tracking. Each of the six boxes held three silicone spheres, 6.2 *mm* in diameter, mounted on thin shafts and placed such that their relative locations roughly form an equilateral triangle at the center. The spheres were placed at three different depth levels, approximately 5 *mm*, 15 *mm* and 25 *mm* from the surface, with their ordering randomized across phantoms to avoid learning effects. The inner walls of the boxes were coated with 4 *mm* of Mold Star 16 FAST silicone[3], to dampen US reflections (Fig. 4.4(a)). Three of the boxes were then filled with polyvinyl alcohol cryogel (PVA-C) for ultrasound imaging, while the other three were left open to be used to assess subjects' base-line localization performance. A 1 *mm* thick layer of silicone (green), textured with black silicone was placed on top of the PVA-C (Fig.4.4 (b-c)). The silicone was added to prevent water evaporation from the PVA-C while the black texture provided surface features. Finally, a CT image of each phantom was obtained at 0.42 *mm* x 0.42 *mm* x 0.83 *mm* resolution and was registered to the tracking DRB using the divots on the walls. Using this registration to transform 2D/3D US localized targets to CT space, the mean target registration error (TRE) of the system was measured to be 1.35 ± 0.07 *mm* with 2D US images, and 0.99 ± 0.17 *mm* in 3D US volumes.

4.4 Experiments and Results

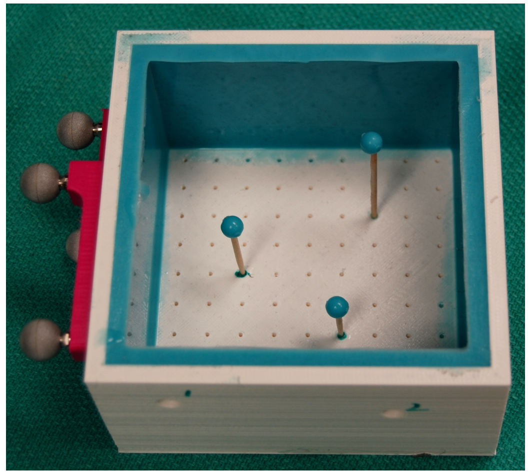
In the experiments, the subjects' ability to localize a hidden target using ultrasound in both monoscopic and stereoscopic laparoscopy was assessed. Seven experienced US users, four male and three female, consented under a protocol approved by the Western University Research Ethics Board, participated in the study using the setup described in Section 4.3.3. All subjects had normal or corrected-to-normal vision with stereo acuity better than 40" of arc assessed on the stereoscopic display [19].

Four modes of ultrasound visualization were tested per viewing condition (stereoscopic / monoscopic): (1) conventional method with US displayed in the monitor on the US machine,

[3] www.smooth-on.com

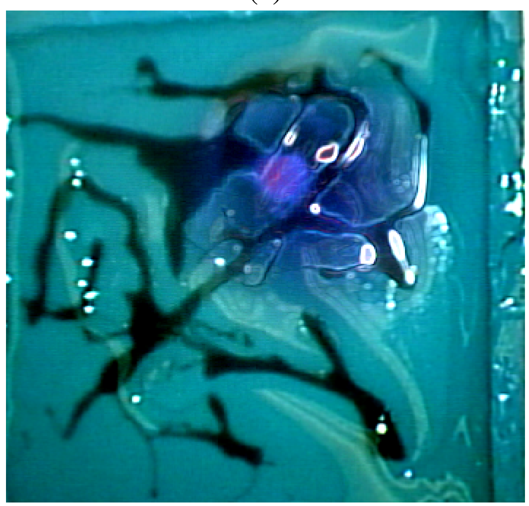
(2) *hybrid in situ* visualization approach with *naïve blending* (Fig.4.4 (b)), (3) *hybrid in situ* visualization with the *keyhole blending* (Fig.4.4 (c)), and (4) hybrid approach where the *keyhole blending* approach is used while the subjects were allowed to read the depth of the target from the US image on the machine. During the experiment, the subjects interactively visualized the targets in the PVA-C filled phantoms using US in different visualizaiton modes. The perceived position was then indicated by the triangulation-by-pointing method [20], where the subjects pointed an optically tracked tool to the centroid of the target at three different poses. During an experimental trial, three pointer poses relative to the phantom were recorded, and the error between the perceived location, determined by triangulation [21], and the reference location given by the CT was computed (Fig. 4.4(d)). The task duration was also saved to be used in the analysis following the experiment.

With each visualization mode, subjects localized the targets in all three PVA-filled phantoms, each subject providing three data points per target depth per visualization mode. The order of presentation of the phantoms and the order of visualization were counterbalanced across different subjects to minimize any bias. Following the US-based localization experiments, the subjects localized targets under direct laparoscopic viewing using phantoms that had the PVA-C medium removed, allowing their base-line performance to be assesed. The stereoscopic viewing experiments were performed at the same time as the experiments with the monoscopic viewing condition, with their order couterbalanced across different subjects to eliminate any bias. Finally, the users provided their subjective opinion on the difficulty of the task by using the NASA TLX ranking system [22].

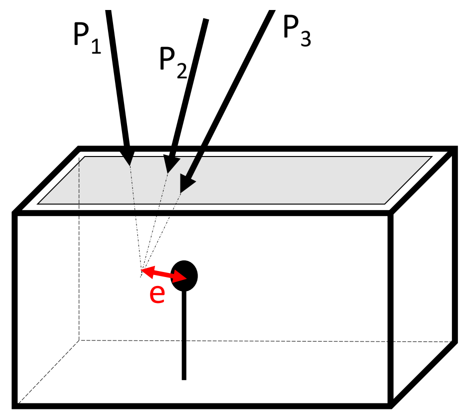


(a)

(b)



(c)



(d)

Figure 4.4: (a) Phantom box with the optical tracking spheres attached (outer left wall): Three silicone spheres, mounted on this shafts, are placed such that their relative locations roughly form a triangle at the center. Spheres are placed at three different depth levels from the surface. Inner walls of the box was coated with a thick layer of silicone to dampen US reflections, (b) 3D US volume blended with the laparoscopic image with *naïve blending* method, (c) *keyhole blending* scheme is used to blend the US volume with the camera image, (d) the perceived target location was determined by triangulating three poses (P_1, P_2, and P_3) of an optically tracked pointer. The error (**e**) between the perceived point that the reference location provided by the registered CT image is computed and used in the analysis

4.4.1 Experiment I: Monoscopic Viewing

In this experiment, the subjects' ability to localize a hidden target using ultrasound in monocular laparoscopy was assessed. The accuracy of localization, as well as the time required to complete the task, was measured.

Experiment

A 3 (depth) x 5 (visualization mode: conventional, *hybrid in situ* with *naïve blending*, *hybrid in situ* with *keyhole blending*, hybrid, and direct laparoscopic visualization), 2-way experimental design was implemented with three replications, resulting in 45 trials per subject. The target localization in clear phantoms under direct laparoscopic viewing was performed at the end to avoid any bias. Other visualization modes, and phantom testing order were counterbalanced across subjects, to avoid bias. When a target was localized, subjects indicated its perceived location by pointing to its centroid at three different poses. For each target, three poses together with the localization time were saved for post experimental analysis. After testing each visualization mode (except the direct laparoscopic visualization mode), the subjective ranking on the difficulty of the localization task was indicated by the subjects in the NASA TLX scale.

Results

A 2-way repeated measures ANOVA was applied to the depth component of the localization measurements to reveal that both the visualization mode and the target depth have effects on depth perception (main effect (visualization): $F(2.79, 55.84) = 8.793, p < 0.001$, main effect (depth): $F(2, 40) = 6.354, p = 0.006$). Moreover, an interaction between the visualization mode and the target depth was observed (depth x visualization: $F(4.60, 92.01) = 3.763, p = 0.005$). The Mauchly's test for sphericity indicated that the assumption for sphericity, the condition where the variance of the differences between all combinations of related levels are equal, has been violated with two effects (visualization mode and its interaction with target depth). Violation of the assumption for sphericity increases the Type I error rate, hence, in these cases, the Greenhouse-Geisser correction was applied to achieve more valid critical F value. To reveal the effect of the mode of visualization at each depth level, three 1-way repeated measure ANOVAs were applied, with Greenhouse-Gaisser correction applied whenever the sphericity assumption has been violated. At the shallowest depth level, the mode of visualization indicated no effect on depth perception ($F(2.47, 49.45) = 0.199, p = 0.862$). However, at the depth level 15 mm from the phantom surface, the visualization mode was revealed to have an effect ($F(2.85, 57.05) = 3.151, p = 0.034$). Post-hoc analysis with Sidak correction further showed that the *hybrid* visualization method resulted in perception of significantly more depth compared to the *naïve blending*-based method and the direct laparoscopic localization method (Fig. 4.5(a)). The visualization mode had an effect on depth perception at depth level 25 mm from the surface ($F(2.91, 58.14) = 8.781, p < 0.001$). Post-hoc analysis with Sidak correction further revealed that with the *hybrid* visualization method subjects perceived more depth compared to the other modes of visualization ($p < 0.04$). Interestingly, this method resulted in slight overestimation of depth (Fig. 4.5(a)).

Fig. 4.5(b) shows the perceived x-y locations of the hidden targets localized using different modes of visualization with the monoscopic viewing conditions. The direct laparoscopic localization method demonstrated a small bias in both x and y directions, compared to all the other modes of visualization (main effect (x - visualization): $F(2.33, 46.55) = 15.952, p < 0.001$, target depth: $F(2, 40) = 1.185, p = 0.316$, target depth x visualization: $F(3.20, 63.93) = 1.573, p = 0.202$; main effect (y - visualization): $F(2.89, 57.70) = 7.287, p < 0.001$, target depth: $F(1.42, 28.35) = 24.527, p < 0.001$, target depth x visualization $F(4.22, 84.38) = 4.538, p = 0.002$). Greenhouse-Geisser correction was applied whenever the sphericity assumption was violated

Fig. 4.5(c) summarizes the results for task duration for each visualization mode. A 2-way repeated measure ANOVA was applied to the data to reveal that the visualization mode has an effect on the task duration (main effect(visualization mode): $F(1.96, 39.13) = 5.031, p = 0.012$, main effect (depth): $F(2, 40) = 0.597, p = 0.555$, target depth x visualization: $F(3.21, 64.27) =$

1.243, $p = 0.058$). Post-hoc analysis with Sidak correction revealed that the conventional method required significantly more time compared to the *keyhole blending*-based method ($p = 0.045$), and the *hybrid* method ($p = 0.034$).

Based on the NASA TLX-based reporting, summarized in Fig. 4.6(a), subjects experienced on average lowest mental demand, physical demand, temporal demand, effort and frustration with the *keyhole blending*-based method while their subjective evaluation on performance was maximum with the *hybrid* technique. Statistically the *hybrid* method was not different from the *keyhole blending*-based method with respect to all other parameters. The conventional method required, on average, highest mental, physical and temporal demand, effort and result in high frustration. Moreover, with the conventional method, subjects felt significantly low confidence on their performance compared to the *hybrid* method.

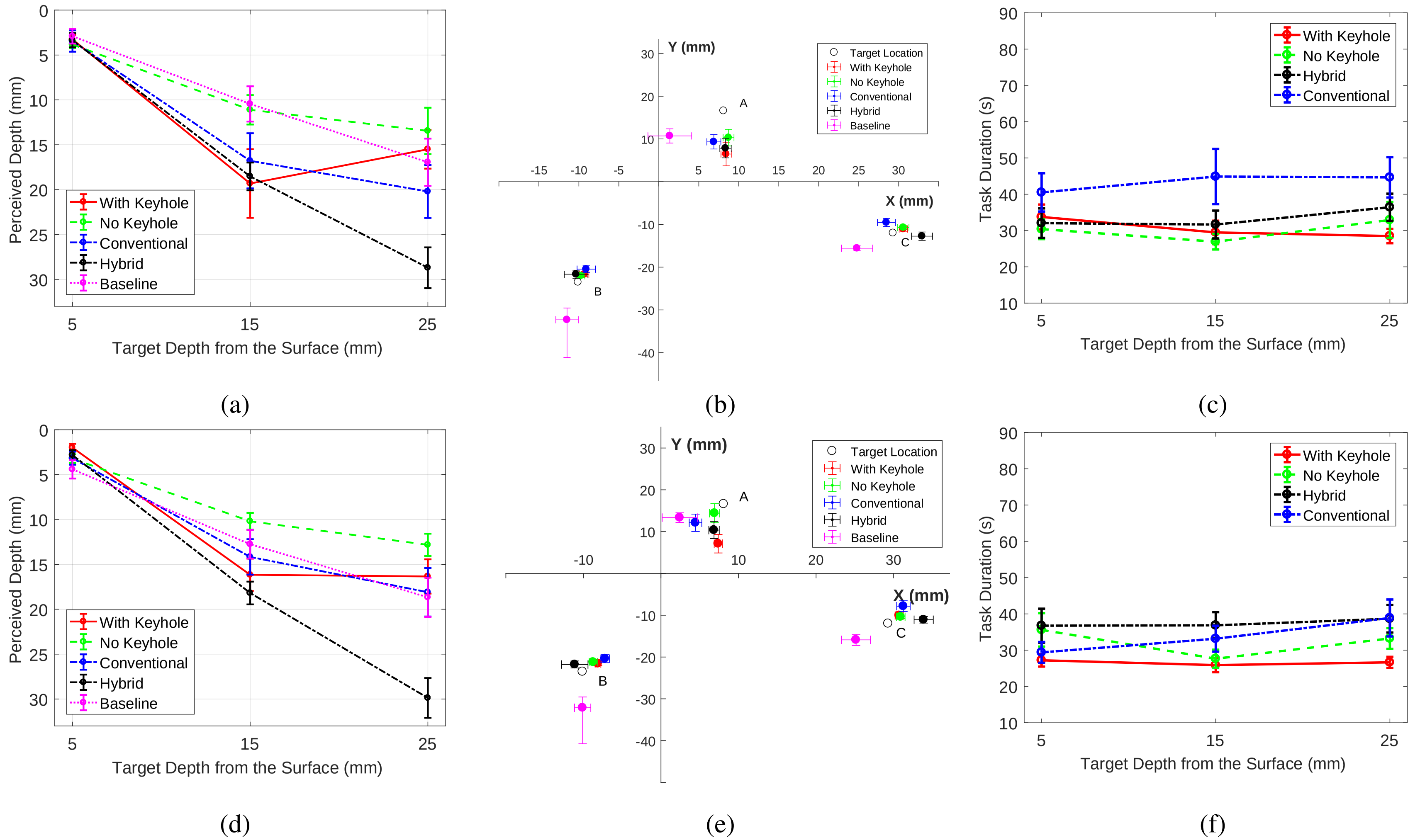


Figure 4.5: Results of the perceptual study (n = 21): depth perception in (a) monoscopic viewing and (d) stereoscopic viewing, perception of x-y location in (b) monoscopic viewing and (e) stereoscopic viewing. The depth from the phantom surface to target A, B and C are 5, 15, 25 mm respectively, task duration in (c) monoscopic viewing and (f) stereoscopic viewing. Graphs are best viewed in color

4.4.2 Experiment II: Stereoscopic Viewing

The objective of this experiment was to assess the subjects ability to localize hidden targets using ultrasound in stereo-laparoscopy with a passive stereoscopic display (Fig. 4.3). The stereoscopic depth cue *binocular disparity* is available to the viewers with a passive stereoscopic display, and as a result, improved depth perception can be expected. The experimental conditions in this experiment matched those in Experiment I allowing comparison of the results obtained with this experiment with those obtained from the Experiment I, permitting assessment of the effect of binocular disparity in a *hybrid in situ* US visualization environment.

Experiment

The same experimental procedure with a 3(depth) x 5(visualization mode), 2-way experimental design as in Experiment I was followed, except that the laparoscopic image with / without US-overlay was rendered in stereo. During the experiment, subjects wore passive polarizing glasses to enable stereoscopic visualization. Similar to the Experiment I, three pointer poses and task duration were recorded for analysis, with the subjective opinion on task difficulty reported in the NASA TLX scale.

Results

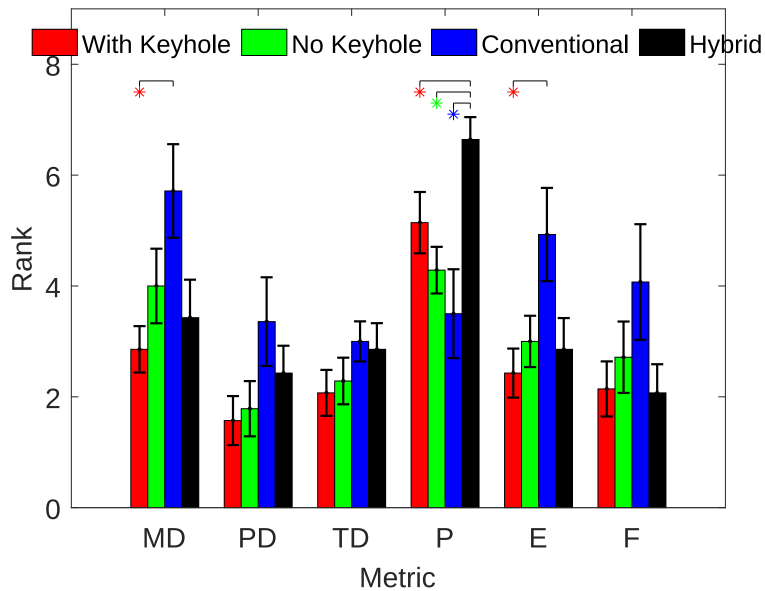
A 2-way repeated measures ANOVA was applied to the perceived depth data to check for any effects. Both visualization and the target depth were revealed to have an effect on depth perception (main effect (visualization): $F(F(4,80) = 15.790, p < 0.001)$, main effect (target depth): $F(2,40) = 7.661, p = 0.002)$. In addition, an interaction between the visualization mode and the target depth was observed (visualization x target depth: $F(4.60, 92.05) = 5.187, p < 0.001)$. The Mauchly's test indicated that the sphericity assumption with the interaction effect has been violated, hence, Greenhouse-Geisser correction was applied. Three 1-way repeated measures ANOVAs were applied to the data to investigate the effect of visualization at each depth level. At the shallowest depth level no effect was observed (main effect (visualization):

$F(2.50, 49.78) = 1.711, p = 0.184$). At 15 mm from the surface, the visualization mode had an effect of the perceived depth (main effect (visualization): $F(4, 80) = 4.923, p = 0.001$). Post-hoc analysis with the Sidak correction revealed that with the *hybrid* method, subjects perceived significantly more depth compared to the *naïve blending*-based method ($p < 0.001$), and the direct laparoscopic viewing method ($p = 0.005$). The *keyhole blending* method resulted in the perception of significantly more depth compared to the *naïve blending* method ($p = 0.02$) at this depth level. At more profound depth levels (25 mm from the surface), the visualization mode has an effect on depth perception (main effect (visualization): $F(4, 80) = 12.234, p < 0.001$), with, based on post-hoc analysis with Sidak correction, the *hybrid* method outperforming other modes of visualization ($p < 0.026$). Similar to the experiment I, a slight overestimation in depth was observed with this method (Fig. 4.5(d)).

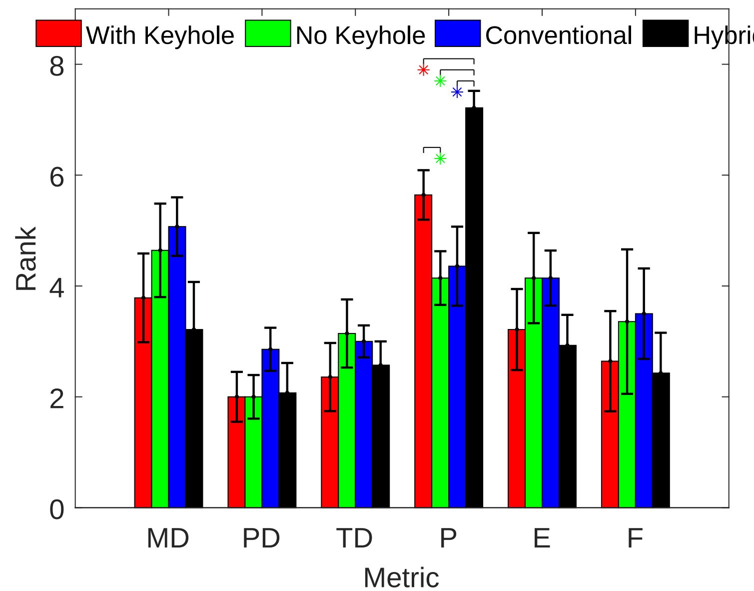
A 2-way repeated measure ANOVA was applied to both x and y directional components of localization errors. Both the visualization and the target depth had effects on localization accuracy while interactions were also observed (main effect (x - visualization): $F(2.50, 50.15) = 12.184, p < 0.001$, main effect (x - depth): $F(2, 40) = 10.061, p < 0.001$, visualization x target depth: $F(2.97, 59.36) = 3.067, p = 0.035$; main effect (y - visualization): $F(2.65, 53.03) = 9.210, p < 0.001$, main effect (y - depth): $F(1.40, 27.91) = 13.619, p < 0.001$, visualization x target depth: $F(4.48, 89.60) = 5.961, p < 0.001$). Greenhouse-Geisser correction was applied whenever the sphericity assumption was violated. Post-hoc analysis revealed that the direct laparoscopic localization has a significant bias in both x and y directions (Fig. 4.5(e)).

In Fig. 4.5(f) task duration results for the study with stereoscopic viewing is summarized. A 2-way repeated measures ANOVA was applied to this data to determine that the visualization mode has an effect on the task duration irrespective of the target depth (main effect (visualization): $F(3, 60) = 4.849, p = 0.004$, main effect (target depth): $F(2, 40) = 1.903, p = 0.162$, visualization x target depth: $F(3.5, 70) = 2.492, p = 0.058$). Post-hoc analysis revealed that the *hybrid method* required significantly more time compared to the *keyhole blending*-based method under stereoscopic viewing conditions.

Based on subjective ranking on the NASA TLX scale (Fig. 4.6(b)), on average, the *hybrid* method required lowest mental demand and effort. The subjects demonstrated significantly high levels of confidence in task performance with this approach compared to the other methods. In addition, the *keyhole blending*-based method was preferred over *naïve blending*-based method when it comes to task performance.



(a)



(b)

Figure 4.6: Aggregated (n=7) subjective ranking based on NASA-TLX for (a) monoscopic viewing, and (b) stereoscopic viewing. Statistical significance is indicated by an asterisk with corresponding color. MD - Mental Demand, PD - Physical Demand, TD - Temporal Demand, P - Performance, E - Effort, F - Frustration

4.5 Discussion

Results of Experiment I reveal the effects of the US visualization method on the perception of the location of hidden targets under monoscopic laparoscopy. The subjects perceived slightly more depth when the *keyhole blending*-based method was used compared to the *naïve blending*-based methods, but the improvement appear to plateau with increasing depth. The high frequency surface features inside the trasparent region in the *keyhole blending* approach cue the subjects about depth ordering. These cues, together with the distance dependent transparency function, result in improved depth perception. Interestingly, the conventional method of US visualization outperforms these two overlay visualization methods in terms of depth perception with deeper targets. Subjects can always read the depth of the target from the depth scale associated with the US image with the conventional approach, whereas in the overlay visualization methods they are limited by the depth cues provided by the monoscopic display. The depth scale result in more accurate internal representations compared to that acquired from the two overlay visualization methods. However, the accuracy comes at the expense of higher cognitive effort, as indicated by significantly higher task duration and subjective ranking associated with the conventional visualization method. Moreover, the conventional approach tends to underestimate the distance with increasing depth [21]. The *hybrid* approach, where the subjects combine the depth scale with the overlay, finds a good compromise between localization accuracy and task duration. The trends in subjective rankings confirms that *hybrid* visualization approach outperforms the other modes of visualization, particularly at more profound depth levels.

Trends similar to those observed in monoscopic visualization were observed with stereoscopic visualization. The *naïve blending*-based method resulted in worst depth perception while the *keyhole blending*-based method and the conventional visualization method resulted in performance similar to what one would demonstrate with direct laparoscopic localization. However, larger depths were significantly under estimated. Again, the *hybrid* method resulted in improved depth perception even though the depth is over estimated slightly. Interestingly however, this method required the longest time to complete the task. This could be due to the interplay of two cognitively expensive processes; stereopsis computation, and fusing depth information read from the US depth scale with the percept acquired from the overlay. Nevertheless, the trends in subjective ranking suggests that the *hybrid* visualization method provides significantly high confidence in task performance with comparatively low cognitive effort.

In both stereoscopic and monoscopic visualization experiments, the direct laparoscopic localization method resulted in biases along both x and y directions. When US is used for localization, the probe is placed directly above the target providing strong cues about its x-y location. This cue does not exist in direct laparoscopic localization technique, hence the bias. Wu et al. [21] observed similar behavior with their experiments with the Sonic FlashLight device.

In the experiments a significant effect of stereopsis on depth perception was not observed. This could be either due to the limited depth resolution in the passive stereoscopic display system employed, or a result of vergence-accommodation mismatch: Consider the following equation that relates the disparity δ created by two points in space [23],

$$\delta \approx I\Delta D/D^2 \tag{4.3}$$

where I is the interpupillary distance, and D is the viewing distance. To estimate ΔD, D needs to be estimated from the vergence angle, but in principle it can be estimated by the accommodation system as well. For a target rendered beyond the display screen where the subjects eyes are fixated, D will be underestimated if the accommodation system has significant contribution for the estimation of D. As a result, ΔD will be underestimated. With a series of psychophysical experiments Watt et. al [24] demonstrated that this is in fact the case. We may be observing the effects of this phenomenon in the experiments with the stereoscopic viewing.

A major hurdle one needs to overcome in translating the proposed *hybrid in situ* visualization pipeline detailed in this chapter to the clinic is the design of the opacity transfer function.

While the current implementation uses a one-dimensional transfer function for simplicity, it provides supports for higher dimensional transfer functions. Future research will evaluate the learning of complex transfer functions to reveal clinically important targets surrounded by soft-tissue. In addition, future studies should study the 3D form perception with the proposed visualization scheme.

4.6 Conclusion

In this chapter, a method to visualize hidden surgical targets in 3D using laparoscopic ultrasound in laparoscopic surgery was presented. In this approach, as the surgeon scans an organ of interest, the 2D US are compounded into a 3D volume in real-time, and rendered in the context of the laparoscopic image at the correct spatial pose, hence referred to as *hybrid in situ* visualization. Different methods of blending the rendered volume, and the laparoscopic image were considered, and using psychophysical experiments involving experienced US users, compared to the conventional method of visualizing US in a separate display. The results of the experiments suggest that *hybrid in situ* visualization with a *keyhole blending* scheme, where the volume is blended using a custom opacity function featuring high frequency surface features, results in the perception of the location of hidden targets with reduced cognitive efforts, both in monoscopic and stereoscopic laparoscopy. By combining numerical depth value of the imaged target with this visualization scheme resulting in a *hybrid* approach, improvements in depth perception could be achieved, but with a reasonable compromise in cognitive efforts. Future studies are required to reveal the efficacy of the proposed methods in terms of perception of 3D form of hidden targets.

Bibliography

[1] M. Feuerstein, T. Reichl, J. Vogel, A. Schneider, H. Feussner, and N. Navab, “Magneto-optic tracking of a flexible laparoscopic ultrasound transducer for laparoscope augmentation,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007: 10th International Conference, Brisbane, Australia, October 29 - November 2, 2007, Proceedings, Part I* (N. Ayache, S. Ourselin, and A. Maeder, eds.), pp. 458–466, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[2] C. L. Cheung, C. Wedlake, J. Moore, S. E. Pautler, and T. M. Peters, “Fused video and ultrasound images for minimally invasive partial nephrectomy: A phantom study,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010: 13th International Conference, Beijing, China, September 20-24, 2010, Proceedings, Part III* (T. Jiang, N. Navab, J. P. W. Pluim, and M. A. Viergever, eds.), pp. 408–415, Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.

[3] P. Pratt, A. Di Marco, C. Payne, A. Darzi, and G.-Z. Yang, “Intraoperative ultrasound guidance for transanal endoscopic microsurgery,” in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part I* (N. Ayache, H. Delingette, P. Golland, and K. Mori, eds.), pp. 463–470, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[4] X. Liu, S. Kang, W. Plishker, G. Zaki, T. D. Kane, and R. Shekhar, “Laparoscopic stereoscopic augmented reality: toward a clinically viable electromagnetic tracking solution.,” *Journal of medical imaging (Bellingham, Wash.)*, vol. 3, p. 045001, October 2016.

[5] A. Hughes-Hallett, P. Pratt, E. Mayer, A. Di Marco, G.-Z. Yang, J. Vale, and A. Darzi, “Intraoperative ultrasound overlay in robot-assisted partial nephrectomy: first clinical experience.,” *Eur. Urol.*, pp. 671–2, 2014.

[6] S. E. Pautler, P. L. Choyke, C. P. Pavlovich, K. Daryanani, and M. M. Walther, “Intraoperative ultrasound aids in dissection during laparoscopic partial adrenalectomy.,” *J. Urol.*, pp. 1352–5, 2002.

[7] M. Khereba, P. Ferraro, A. Duranceau, J. Martin, E. Goudie, V. Thiffault, and M. Liberman, “Thoracoscopic localization of intraparenchymal pulmonary nodules using direct intracavitary thoracoscopic ultrasonography prevents conversion of VATS procedures to thoracotomy in selected patients.,” *The Journal of thoracic and cardiovascular surgery*, vol. 144, pp. 1160–5, November 2012.

[8] J. Oh, X. Kang, E. Wilson, C. a. Peters, T. D. Kane, and R. Shekhar, "Stereoscopic augmented reality using ultrasound volume rendering for laparoscopic surgery in children," in *Proc. SPIE 9036, Medical Imaging 2014: Image-Guided Procedures, Robotic Interventions, and Modeling* (Z. R. Yaniv and D. R. Holmes, eds.), vol. 9036, p. 90360Y, March 2014.

[9] B. Wu, R. L. Klatzky, and G. Stetten, "Visualizing 3D objects from 2D cross sectional images displayed in-situ versus ex-situ.," *Journal of Exp. Psy.*, pp. 45–59, 2010.

[10] W. Garrett, H. Fuchs, M. Whitton, and a. State, "Real-time incremental visualization of dynamic ultrasound volumes using parallel BSP trees," in *Proc. IEEE Vis. '96*, pp. 235–240,, 1996.

[11] J. W. Trobaugh, D. J. Trobaugh, and W. D. Richard, "Three-dimensional imaging with stereotactic ultrasonography," *Comp. Med. Imag. and Graph.*, pp. 315–323, 1994.

[12] J. S. H. Baxter, T. M. Peters, and E. C. S. Chen, "A unified framework for voxel classification and triangulation," in *Proc. SPIE Med. Imaging* (K. H. Wong and D. R. Holmes III, eds.), vol. 7964, p. 796436, International Society for Optics and Photonics, March 2011.

[13] C. Bichlmeier, F. Wimmer, S. M. Heining, and N. Navab, "Contextual Anatomic Mimesis Hybrid In-Situ Visualization Method for Improving Multi-Sensory Depth Perception in Medical Augmented Reality," in *Proc. IEEE Symp. on Mixed and Augmented Reality*, pp. 1–10, 2007.

[14] M. Lerotic, A. J. Chung, G. Mylonas, and G.-Z. Yang, "pq-space based non-photorealistic rendering for augmented reality," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2007: 10th International Conference, Brisbane, Australia, October 29 - November 2, 2007, Proceedings, Part II* (N. Ayache, S. Ourselin, and A. Maeder, eds.), pp. 102–109, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[15] M. Kersten-Oertel, I. Gerard, S. Drouin, K. Mok, D. Sirhan, D. S. Sinclair, and D. L. Collins, "Augmented reality in neurovascular surgery: feasibility and first uses in the operating room," *International Journal of Computer Assisted Radiology and Surgery*, vol. 10, pp. 1823–1836, November 2015.

[16] I. Morgan, U. Jayarathne, A. Rankin, T. M. Peters, and E. C. S. Chen, "Hand-eye calibration for surgical cameras: a procrustean perspective-n-point solution," *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, pp. 1141–1149, July 2017.

[17] G. Ameri, A. J. McLeod, J. S. H. Baxter, E. C. S. Chen, and T. M. Peters, "Line fiducial material and thickness considerations for ultrasound calibration," in *Proc. SPIE Med. Imaging*, 2015.

[18] A. Lasso, T. Heffter, A. Rankin, C. Pinter, T. Ungi, and G. Fichtinger, "PLUS: Open-Source Toolkit for Ultrasound-Guided Intervention Systems," *IEEE Trans. on Biomed. Engineering*, pp. 2527–2537, 2014.

[19] D. Gadia, G. Garipoli, C. Bonanomi, L. Albani, and A. Rizzi, "Assessing stereo blindness and stereo acuity on digital displays," *Displays*, vol. 35, no. 4, pp. 206–212, 2014.

[20] S. S. Fukusima, J. M. Loomis, and J. A. Da Silva, "Visual perception of egocentric distance as assessed by triangulation.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 23, no. 1, pp. 86–100, 1997.

[21] B. Wu, R. L. Klatzky, D. Shelton, and G. D. Stetten, "Psychophysical evaluation of in-situ ultrasound visualization.," *IEEE transactions on visualization and computer graphics*, vol. 11, no. 6, pp. 684–93, 2005.

[22] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," *Advances in Psy.*, pp. 139–183, 1988.

[23] I. P. Howard and B. J. Rogers, *Seeing in depth*. I. Porteous, 2002.

[24] S. J. Watt, K. Akeley, M. O. Ernst, M. S. Banks, K. K., and C. D., "Focus cues affect perceived depth," *Journal of Vision*, vol. 5, p. 7, December 2005.

Chapter 5

Conclusion

The objective of this thesis was to develop an effective visualization method to render LUS information in the context of the laparoscopic video to enable accurate perception of the spatial location of surgically important hidden targets. In achieving this objective, answers to several research questions were sought, and Chapters 2 through 4 summarize the mission in search for answers for these questions.

In Chapter 2, an efficient method to register 2D US images with a coordinate system centered in the laparoscopic camera was investigated. The method eliminates the need for extrinsic tracking systems, resulting in potentially lower-cost devices, and smaller operating room footprints for the devices. However, it requires a minor cosmetic modifications to the probe to provide sufficient features to enable full 6DoF pose estimation. The computational methods detailed in this chapter estimate the pose of the probe in real-time by using information in monoscopic or stereoscopic laparoscopic camera images. The chapter also studied the accuracy of the estimated pose parameters using an optical tracking system as the reference. In addition, error in registering a target imaged by the US image to the camera-centered coordinate system by using the proposed intrinsic method was also investigated. Overall, the chapter attempted to answer the question of how LUS images are registered to the laparoscopic camera frame of reference with minimal overhead to the existing OR work-flow, and what error-bounds should be expected.

Chapter 3 seeks answers to the question of how 3D information of hidden targets can be captured by a series of 2D US images, and what error-bounds should be expected in such a process. The hybrid US reconstruction algorithm introduced in this chapter can be implemented on the GPU to enable reconstruction of a 3D US volume by incrementally stitching 2D US images as the LUS probe moves over an organ of interest. The geometric accuracy of the reconstructed volume was investigated by imaging a PVA-C phantom, and registering the reconstructed volume to its micro-CT. The results of this study reveal the gross error the system introduces when a hidden target is represented in 3D in the camera-centered coordinate system.

Methods to render hidden surgical targets in 3D in the context of the laparoscopic video was investigated in Chapter 4. A 3D US volume can be reconstructed by incrementally stitching 2D US images in the camera coordinate system by using the methods detailed in Chapters 2 and 3. This volume can be blended with the surface image provided by the laparoscopic camera to yield a more intuitive US-augmented video by using the techniques detailed in this chapter. Chapter 4 also discussed the results of a perceptual study conducted involving experienced US users to investigate the accuracy of localizing a hidden target using different visualization methods, both under monoscopic and stereoscopic viewing conditions. Overall, the chapter attempted to answer the question of how the spatially registered US information should be presented, such that the surgeons perceives the spatial location and the 3D form of hidden targets accurately.

The thesis was based on the hypothesis that, compared to the conventional *ex situ* visualization, *hybrid in situ* visualization of US improves the surgeons ability to perceive the spatial location and the 3D form of hidden surgical targets in conventional and robot-assisted laparoscopic surgery. The perceptual study detailed in Chapter 4 allows this hypothesis to be tested for the particular visualization method, detailed in Chapters 2, 3 and 4, in a simulated surgical environment. Thus, the system has been assessed up to the level two in the IGS system assessment scheme described by Jannin and Korb [1]. The results of the experiments supports the hypothesis on which this thesis is based under simulated surgical environment. Further studies may be necessary to test this hypothesis in more realistic surgical environments involving both novice and experienced laparoscopic surgeons.

5.1 Concurrent Development

During the time this thesis was prepared, there were several development by other researchers in related fields. The work by Pratt et al. [2, 3] on image-based tracking, and visualization of US in robot-assisted surgery is one such example. The authors use a planar fiducial pattern attached to a custom-made micro-US probe that is tracked in 6DoF in real-time based on the information in the laparoscopic images. They developed an intuitive method to visualize 2D US images in the context of the laparoscopic images, and demonstrated their efficacy in clinical settings. In addition, they utilized their tracking method in autonomous soft-tissue scanning applications intended for autonomous tumor resections [4], and motion-compensated US reconstructions [5].

A major challenge in visualizing 2D US images in the context of the laparoscopic images, is that for certain probe poses, particularly when the US image is nearly perpendicular to the camera imaging plane, the content in the augmented-image is difficult to interpret. One solution to this problem is to synthesize an appropriate view of the scene using the appearance and the structure of the surgical scene captured by the camera. Such views must be generated at real-time frame-rates to be useful in practice. The structure of the scene can be computed in real-time either by using passive stereo techniques, or by structured-light techniques. Chang et al. [6] developed a real-time passive stereo method based on variational optimization to reconstruct the structure of the surgical scene. The GPU-accelerated method was reported to demonstrate near-real-time performance, but further development may be necessary to improve the method to be useful in the view synthesis problem. A simple and practical structured light system that can be tracked by an image-based technique, was recently developed by Edgecumbe et al. [7]. The method allows a surface point-cloud to be constructed in the camera coordinate system, which can be used to approximate the structure of the surgical scene. However, further development may be necessary to assess the efficacy of this method in the clinical environment.

5.2 A Look Into the Future

At the end of each chapter in this thesis, I suggested future research directions for each method described. However, the most valuable future research direction will be the assessment of performance and the limitations of the proposed methods in clinical settings. Without thorough assessment, proving the clinical importance of the developed tools will be difficult. The thesis investigated the perception of only the spatial location of the hidden targets. Therefore, further studies that involve both novice and experienced surgeons are necessary to investigate whether *hybrid in situ* visualization improves the surgeons perception of 3D form of hidden target and the surrounding critical structures.

5.2.1 Future Image-guided Soft-tissue Surgery

The techniques developed in this thesis, together with the developments in related fields, will enable image-guidance in soft-tissue surgery in future: Improved image-based tracking methods, such as that described in Chapter 2, will allow intraoperative imaging probes such as LUS and hand-held nuclear imaging probes, and surgical tools to be tracked with respect to the camera coordinate system. If the camera is also tracked by using a sparse [8–10], or a dense localization technique [11], intra-operative images, and surgical tools can be related to the patient coordinate system. Assistance from robotic systems may be required to compensate for tissue motion in these registrations. Once all the information sources are registered to a common coordinate system, several research question will arise:

- How should the registered information be rendered such that intended surgical actions can be performed optimally?
- How accurately can the particular surgical task can be performed with the help of the the guidance system?
- Does the system improve the performance irrespective of the surgeons experience?

To answer these question, properly designed experiments involving both novice and experienced surgeons will be required, and the results will determine the value of the developed technology in the clinic.

5.2.2 Fusing Pre-operative and Intra-operative Imaging

Even though intraoperative imaging can provide a real-time visual feedback into the soft-tissue structures, in certain situations the quality of such images may limit their usefulness in surgery. For instance, in lung resection surgeries for small nodules using the VATS approach, LUS is often used to localize the nodules and determine resection margins [12]. Typically, the lung is fully collapsed to enable US penetration, but due to trapped air the quality of US is often affected, rendering nodule localization task difficult. As a result, many VATS procedures are converted into more invasive thoracotomies so that the nodules can be localized by palpation. One way to solve the quality issue in intraoperative images is to register them with a pre-operative image of the anatomy, so that the pre-operative images can complement those obtained intra-operatively. However, large deformation of the anatomy from its pre-operative to intra-operative state may render the registration problem extremely difficult. In addition, the algorithms to solve this registration problem should run at adequate speed to make sure such techniques are useful in practical surgery. Future research should focus on such registration problems. In addition to improving the image quality, such registration algorithms will enable pre-operative plans to be brought into the surgical scene to increase both the efficiency and safety of the surgical procedure.

Bibliography

[1] P. Jannin and W. Korb, "Assessment of image-guided interventions," in *Image-Guided Interventions: Technology and Applications* (T. Peters and K. Cleary, eds.), pp. 531–549, Boston, MA: Springer US, 2008.

[2] P. Pratt, A. Di Marco, C. Payne, A. Darzi, and G.-Z. Yang, "Intraoperative ultrasound guidance for transanal endoscopic microsurgery," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012: 15th International Conference, Nice, France, October 1-5, 2012, Proceedings, Part I* (N. Ayache, H. Delingette, P. Golland, and K. Mori, eds.), pp. 463–470, Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[3] P. Pratt, A. Jaeger, A. Hughes-Hallett, E. Mayer, J. Vale, A. Darzi, T. Peters, and G.-Z. Yang, "Robust ultrasound probe tracking: initial clinical experiences during robot-assisted partial nephrectomy," *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–9, 2015.

[4] P. Pratt, A. Hughes-Hallett, L. Zhang, N. Patel, E. Mayer, A. Darzi, and G.-Z. Yang, "Autonomous ultrasound-guided tissue dissection," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part I* (N. Navab, J. Hornegger, W. M. Wells, and A. Frangi, eds.), pp. 249–257, Cham: Springer International Publishing, 2015.

[5] L. Zhang, M. Ye, S. Giannarou, P. Pratt, and G.-Z. Yang, "Motion-compensated autonomous scanning for tumour localisation using intraoperative ultrasound," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II* (M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne, eds.), pp. 619–627, Cham: Springer International Publishing, 2017.

[6] P.-L. Chang, D. Stoyanov, A. J. Davison, and P. E. Edwards, "Real-time dense stereo reconstruction using convex optimisation with a cost-volume for image-guided robotic surgery," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013: 16th International Conference, Nagoya, Japan, September 22-26, 2013, Proceedings, Part I* (K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab, eds.), pp. 42–49, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.

[7] P. Edgcumbe, P. Pratt, G.-Z. Yang, C. Nguan, and R. Rohling, "Pico lantern: Surface reconstruction and augmented reality in laparoscopic surgery using a pick-up laser projector," *Medical Image Analysis*, vol. 25, no. 1, pp. 95 – 102, 2015.

[8] P. Mountney, D. Stoyanov, A. Davison, and G.-Z. Yang, "Simultaneous stereoscope localization and soft-tissue mapping for minimal invasive surgery," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006: 9th International Conference, Copenhagen, Denmark, October 1-6, 2006. Proceedings, Part I* (R. Larsen, M. Nielsen, and J. Sporring, eds.), pp. 347–354, Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[9] P. Mountney and G.-Z. Yang, "Soft tissue tracking for minimally invasive surgery: Learning local deformation online," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008: 11th International Conference, New York, NY, USA, September 6-10, 2008, Proceedings, Part II* (D. Metaxas, L. Axel, G. Fichtinger, and G. Székely, eds.), pp. 364–372, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[10] P. Mountney, D. Stoyanov, and G. Z. Yang, "Three-dimensional tissue deformation recovery and tracking," *IEEE Signal Processing Magazine*, vol. 27, pp. 14–24, July 2010.

[11] P.-L. Chang, A. Handa, A. J. Davison, D. Stoyanov, and P. E. Edwards, "Robust real-time visual odometry for stereo endoscopy using dense quadrifocal tracking," in *Information Processing in Computer-Assisted Interventions: 5th International Conference, IPCAI 2014, Fukuoka, Japan, June 28, 2014. Proceedings* (D. Stoyanov, D. L. Collins, I. Sakuma, P. Abolmaesumi, and P. Jannin, eds.), pp. 11–20, Cham: Springer International Publishing, 2014.

[12] M. Khereba, P. Ferraro, A. Duranceau, J. Martin, E. Goudie, V. Thiffault, and M. Liberman, "Thoracoscopic localization of intraparenchymal pulmonary nodules using direct intracavitary thoracoscopic ultrasonography prevents conversion of vats procedures to thoracotomy in selected patients," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 144, no. 5, pp. 1160 – 1166, 2012.

Appendices

Appendix A

Anisotropically Scaled ICP

Let **X**, and **Y** be two point sets with their one-to-one correspondence known. The Anisotropically Scalled Iterative Closest Point (ASICP), a variant of the well known Iterative Closest Point (ICP) algorithm [1], estimates the rigid transformation that registers the two point sets by minimizing,

$$FRE = \frac{1}{N_p} \sum_{i=1}^{N_p} ||(\mathbf{RSX}_i + \mathbf{t}) - \mathbf{Y}_i|| \tag{A.1}$$

where $\mathbf{R}$ is a rotational matrix, $\mathbf{S} = diag\{s1, s2, s3\}$ is a diagonal scaling matrix, N_p is the number of point pairs, and $||.||$ is the Eclidean norm. An algorithm to solve this minimization problem is give below.

Demean the data to obtain $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$, and $\mathbf{B} = \hat{\mathbf{Y}}^T.\hat{\mathbf{X}}$;

Normalize the row of $\hat{\mathbf{X}}$ such that $\hat{\mathbf{X}}.\hat{\mathbf{X}}^T = 1$;

while $\Delta FRE > threshold$ **do**

  $[\mathbf{U}, \mathbf{\Lambda}, \mathbf{V}] = svd(\mathbf{B}.diag(diag(\mathbf{R}^T.\mathbf{B}))$

  $\mathbf{R} = \mathbf{U}.diag(1, 1, det(\mathbf{U}.\mathbf{V})).\mathbf{V}^T$

  compute new FRE

end

$\mathbf{S} = diag(\mathbf{B}^T.\mathbf{R})$

Algorithm 1: Anisotropically Scaled ICP (AICP) Algorithm

Bibliography

[1] Besl, Paul J.; N.D. McKay, *A Method for Registration of 3-D Shapes*, IEEE Trans. on Pattern Analysis and Machine Intelligence. Los Alamitos, CA, USA: IEEE Computer Society, 4 (2): 239256, 1992.

Appendix B

Copyright Releases

For the chapters that were adapted from previously published articles, permission was sought to reproduce the content in this thesis. In the following, I have attached a copy of the copyright agreement that explicitly allow reproduction, or a copy of the license or letter of permission to reproduce the material

B.1 Releases for Material in Chapter 2

This Agreement between Mr. Uditha Jayarathne ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4257871475613
License date	Dec 28, 2017
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	Robust Intraoperative US Probe Tracking Using a Monocular Endoscopic Camera
Licensed Content Author	Uditha L. Jayarathne, A. Jonathan McLeod, Terry M. Peters et al
Licensed Content Date	Jan 1, 2013
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	yes
Title	Ultrasound-augmented Laparoscopy
Instructor name	Terry M. Peters, PhD
Institution name	Western University, Canada
Expected presentation date	Feb 2018
Portions	I will be re-using the full content in this article.
Requestor Location	Mr. Uditha Jayarathne Imaging Research Laboratories Robarts Research Institute P.O. Box 5015 100 Perth Drive London, ON N5Y2N5 Canada Attn: Mr. Uditha Jayarathne
Billing Type	Invoice
Billing Address	Mr. Uditha Jayarathne Imaging Research Laboratories Robarts Research Institute P.O. Box 5015 100 Perth Drive London, ON N5Y2N5 Canada Attn: Mr. Uditha Jayarathne
Total	0.00 CAD
Terms and Conditions	

1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

   If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Where **print only** permission has been granted for a fee, separate permission must be obtained for any additional electronic re-use.

3. Permission granted **free of charge** for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

4. A licence for 'post on a website' is valid for 12 months from the licence date. This licence does not cover use of full text articles on websites.

5. Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

6. Permission granted for books and journals is granted for the lifetime of the first edition and does not apply to second and subsequent editions (except where the first edition permission was granted free of charge or for signatories to the STM Permissions Guidelines http://www.stm-assoc.org/copyright-legal-affairs/permissions/permissions-guidelines/), and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence.

7. Rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

8. The Licensor's permission must be acknowledged next to the licensed material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

9. Use of the material for incidental promotional use, minor editing privileges (this does not include cropping, adapting, omitting material or any other changes that affect the meaning, intention or moral rights of the author) and copies for the disabled are permitted under this licence.

10. Minor adaptations of single figures (changes of format, colour and style) do not require the Licensor's approval. However, the adaptation should be credited as shown in Appendix below.

Appendix — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g.

Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj. [JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.0

This Agreement between Mr. Uditha Jayarathne ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4257880141224
License date	Dec 28, 2017
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	Simultaneous Estimation of Feature Correspondence and Stereo Object Pose with Application to Ultrasound Augmented Robotic Laparoscopy
Licensed Content Author	Uditha L. Jayarathne, Xiongbiao Luo, Elvis C. S. Chen et al
Licensed Content Date	Jan 1, 2015
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	yes
Title	Ultrasound-augmented Laparoscopy
Instructor name	Terry M. Peters, PhD
Institution name	Western University, Canada
Expected presentation date	Feb 2018
Portions	I will be using the full content in this article in my thesis.
Requestor Location	Mr. Uditha Jayarathne Imaging Research Laboratories Robarts Research Institute P.O. Box 5015 100 Perth Drive London, ON N5Y2N5 Canada Attn: Mr. Uditha Jayarathne
Billing Type	Invoice
Billing Address	Mr. Uditha Jayarathne Imaging Research Laboratories Robarts Research Institute P.O. Box 5015 100 Perth Drive London, ON N5Y2N5 Canada Attn: Mr. Uditha Jayarathne
Total	0.00 CAD
Terms and Conditions	

1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

   If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Where **print only** permission has been granted for a fee, separate permission must be obtained for any additional electronic re-use.

3. Permission granted **free of charge** for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

4. A licence for 'post on a website' is valid for 12 months from the licence date. This licence does not cover use of full text articles on websites.

5. Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

6. Permission granted for books and journals is granted for the lifetime of the first edition and does not apply to second and subsequent editions (except where the first edition permission was granted free of charge or for signatories to the STM Permissions Guidelines http://www.stm-assoc.org/copyright-legal-affairs/permissions/permissions-guidelines/), and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence.

7. Rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

8. The Licensor's permission must be acknowledged next to the licensed material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

9. Use of the material for incidental promotional use, minor editing privileges (this does not include cropping, adapting, omitting material or any other changes that affect the meaning, intention or moral rights of the author) and copies for the disabled are permitted under this licence.

10. Minor adaptations of single figures (changes of format, colour and style) do not require the Licensor's approval. However, the adaptation should be credited as shown in Appendix below.

Appendix — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g.

Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

<u>Note: For any republication from the British Journal of Cancer, the following credit line style applies:</u>

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj. [JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.0

B.2 Releases for Material in Chapter 3 and 4

This Agreement between Mr. Uditha Jayarathne ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4257890728336
License date	Dec 28, 2017
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Springer eBook
Licensed Content Title	Real-Time 3D Ultrasound Reconstruction and Visualization in the Context of Laparoscopy
Licensed Content Author	Uditha L. Jayarathne, John Moore, Elvis C. S. Chen et al
Licensed Content Date	Jan 1, 2017
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	full article/chapter
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	yes
Title	Ultrasound-augmented Laparoscopy
Instructor name	Terry M. Peters, PhD
Institution name	Western University, Canada
Expected presentation date	Feb 2018
Portions	Figures, and some content will be re-used in my PhD thesis.
Requestor Location	Mr. Uditha Jayarathne Imaging Research Laboratories Robarts Research Institute P.O. Box 5015 100 Perth Drive London, ON N5Y2N5 Canada Attn: Mr. Uditha Jayarathne
Billing Type	Invoice
Billing Address	Mr. Uditha Jayarathne Imaging Research Laboratories Robarts Research Institute P.O. Box 5015 100 Perth Drive London, ON N5Y2N5 Canada Attn: Mr. Uditha Jayarathne
Total	0.00 CAD
Terms and Conditions	

1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

   If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Where **print only** permission has been granted for a fee, separate permission must be obtained for any additional electronic re-use.

3. Permission granted **free of charge** for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

4. A licence for 'post on a website' is valid for 12 months from the licence date. This licence does not cover use of full text articles on websites.

5. Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

6. Permission granted for books and journals is granted for the lifetime of the first edition and does not apply to second and subsequent editions (except where the first edition permission was granted free of charge or for signatories to the STM Permissions Guidelines http://www.stm-assoc.org/copyright-legal-affairs/permissions/permissions-guidelines/), and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence.

7. Rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

8. The Licensor's permission must be acknowledged next to the licensed material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

9. Use of the material for incidental promotional use, minor editing privileges (this does not include cropping, adapting, omitting material or any other changes that affect the meaning, intention or moral rights of the author) and copies for the disabled are permitted under this licence.

10. Minor adaptations of single figures (changes of format, colour and style) do not require the Licensor's approval. However, the adaptation should be credited as shown in Appendix below.

Appendix — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g.

Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT]** (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj. [JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.0

From: Nicole Harris
Sent: December 29, 2017 11:41 AM
To: Uditha Jayarathne
Subject: RE: Attention: Reprint Permission

Dear Mr. Jayarathne,

Thank you for seeking permission from SPIE to reprint material from our publications. Publisher's permission is hereby granted under the following conditions:

(1) you obtain permission of one of the authors;

(2) the material to be used has appeared in our publication without credit or acknowledgment to another source; and

(3) you credit the original SPIE publication. Include the authors' names, title of paper, volume title, SPIE volume number, and year of publication in your credit statement.

Sincerely,

Nicole Harris
Administrative Editor, SPIE Publications
1000 20th St.
Bellingham, WA 98225

SPIE is the international society for optics and photonics. http://SPIE.org

SPIE.

From: Uditha Jayarathne
Sent: Thursday, December 28, 2017 4:34 PM
To: reprint_permission
Subject: Attention: Reprint Permission

Dear sir/madam,

I am requesting permission to re-use the images, tables, and content in my previously published paper titled 'Freehand 3D US Reconstruction with Robust Visual Tracking with Application to Ultrasound-augmented Laparoscopy' in my PhD thesis. Please refer to the following information

Title and authors: 'Freehand 3D US Reconstruction with Robust Visual Tracking with Application to Ultrasound-augmented Laparoscopy' – by Uditha L. Jayarathne, Elvis C.S Chen, John Moore, Terry M. Peters
Volume, issue and page numbers: Proc. SPIE 9786, Medical Imaging 2016: Image-Guided Procedures, Robotic Interventions, and Modeling, 978617 (18 March 2016)
What will be reproduced: Images, tables, and content
Where will the material be republished: The material will be a part of my PhD thesis titled Ultrasound-augmented Laparoscopy. This will be published by the University of Western Ontario.

I would appreciate if permission could be granted for the aforementioned article at your earliest convenience.

Vita

Name:	Uditha L. Jayarathne
Post-secondary Education and Degrees:	University of Moratuwa Moratuwa, Sri Lanka B.Sc. Electronic and Telecomm. Engineering, 2006 - 2010
Honours and Awards:	Graduate Scholarship NSERC CREATE Program 2011 - 2013
	Western Graduate Research Scholarship University of Western Ontario 2011 - 2016
	Mahapola Merit Scholarship Ministry of Higher Education 2006 - 2010
Related Work Experience:	Graduate Research Assistant University of Western Ontario 2011 - 2017
	Graduate Teaching Assistant University of Western Ontario 2011 - 2016

Associate Research Engineer
Zone24x7 Inc., Koswatte, Sri Lanka
2010 - 2011

Book Chapters

Uditha L. Jayarathne, "Ultrasound-augmented Laparoscopy: Technology and Human Factors", in Augmented Reality in Medicine. Terry M. Peters, Ziv Yaniv, Christian Linte, Jackie Williams (Editors), CRC Press (*in revision*)

Journal Articles

Uditha L. Jayarathne, John Moore, Elvis C.S. Chen, and Terry M. Peters, "Visualizing Ultrasound In the Context of Laparoscopy", IEEE Transactions on Visualization and Computer Graphics, (*submitted*)

Uditha L. Jayarathne, Elvis C.S Chen, Terry M. Peters, "Robust, Intrinsic Tracking of a Laparoscopic Ultrasound Probe for Ultrasound-augmented Laparoscopy", IEEE Transactions on Medical Imaging, (*submitted*)

Elvis C.S Chen, Isabella Morgan, **Uditha L. Jayarathne**, Burtan Ma, Terry M. Peters,"Hand-Eye Calibration Using a Target Registration Error Model", Healthcare Technology Letters/International IET Healthecare Technology Letters, 2017

Isabella Morgan, **Uditha L. Jayarathne**, Adam Rankin, Terry M. Peters, and Elvis C.S. Chen, "Hand-eye Calibration for Surgical Cameras: A Procrustean Perspective-n-Point Solution", International Journal of Computer Assisted Radiology and Surgery, pp 1–9, 2017

Ed Ginzel, and **Uditha L. Jayarathne**, "Computer Tomography of Ultrasonic Pulses", The Online Journal of Nondestructive Testing & Ultrasonics, NDT.net, 2017

Peer-reviewed Conference Proceedings

Uditha L. Jayarathne, John Moore, Elvis C.S Chen, Stephen Pautler, and Terry Peters, "Real-time 3D US Reconstruction and Visualization in the Context of Laparoscopy", Lecture Notes in Computer Science, Medical Image Computing and Computer Assisted Interventions (MICCAI), 2017 (8 pages, **poster presentation**)

Xiongbiao Luo, A. Jonathan McLeod, **Uditha L. Jayarathne**, and Terry M. Peters, "Multi-scale Retinex Aggregation to Enable Robust Dense Stereo Correspondence", Proceedings of IEEE International Conference on 3D Vision (3DV), 2015 (9 pages, **poster Presentation**)

Xiongbiao Luo, **Uditha L. Jayarathne**, Stephen E. Pautler, and Terry M. Peters, "Binocular

Endoscopic 3D Scene Reconstruction Using Color and Gradient-boosted Aggregation Stereo Matching for Robotic Surgery", International Conference on Image and Graphics, 2015, (13 pages, **podium Presentation**)

Xiongbiao Luo, **Uditha L. Jayarathne**, A. Jonathan McLeod, and Kensaku Mori, "Enhanced Differential Evolution to Combine Optical Mouse Sensor with Image Structural Patches for Robust Endoscopic Navigation", Lecture Notes in Computer Science, Medical Image Computing and Computer Assisted Interventions (MICCAI), 2014 (8 pages, **poster presentation**)

Uditha L. Jayarathne, A. Jonathan McLeod, Terry M. Peters, and Elvis C.S Chen, "Robust Intraoperative US Probe Tracking with a Monocular Endoscopic Camera", Lecture Notes in Computer Science, Medical Image Computing and Computer Assisted Interventions (MICCAI) 8151, 363-370, 2013 (8 pages, **poster presentation**)

Conference Proceedings

Uditha L. Jayarathne, Elvis C.S. Chen, John Moore, and Terry Peters, "Freehand 3D-US Reconstruction with Robust Visual Tracking with Application to Ultrasound-augmented Laparoscopy", Proceeding of SPIE Medical Imaging, 2016 (6 pages, **podium presentation**)

Xiongbiao Luo, **Uditha L. Jayarathne**, A. Jonathan McLeod, Stephen E. Pautler, Christopher M. Schlacta, and Terry M. Peters, "Uncalibrated Stereo Rectification and Disparity Range Stabilization: A Comparison of Different Feature Detectors", Proceedings of SPIE Medical Imaging, 2016 (8 pages, **podium presentation**)

Xiongbiao Luo, A. Jonathan McLeod, **Uditha L. Jayarathne**, Stephen E. Pautler, Christopher M. Schlacta, and Terry Peters, "Towards Disparity Joint Upsampling for Robust Stereo-endoscopic Scene Reconstruction in Robotic Prostatectomy", Proceedings of SPIE Medical Imaging, 2016 (10 pages, **podium presentation**)

Uditha L. Jayarathne, Xiongbiao Luo, Elvis C.S. Chen, and Terry M. Peters, "Simultaneous Estimation of Feature Correspondence and Stereo Object Pose with Application to Ultrasound-augmented Robotic Laparoscopy", Proceedings of International Workshop on Augmented Environments for Computer-assisted Interventions, 2015 (11 pages, **podium presentation**)

A. Jonathan McLeod, John S. H. Baxter, **Uditha L. Jayarathne**, Stephen Pautler, Terry Peters, and Xiongbiao Luo, "Stereoscopic Motion Magnification in Minimally Invasive Robotic Prostatectomy", Proceedings of International Workshop on Computer-assisted and Robotic Endoscopy, 2015 (11 pages, **poster presentation**)

Xiongbiao Luo, **Uditha L. Jayarathne**, and Terry M. Peters, "Boosted Hybrid EM-Video Endoscopic Navigation Using Organ Centerline Constraint and Structural Measure Under Tissue Deformation", Proceedings of Hamlyn Symposium on Medical Robotics, 2014

Elvis C.S. Chen, A. Jonathan McLeod, **Uditha L. Jayarathne**, and Terry M. Peters, "Solving for Free-hand Real-time 3D Ultrasound Calibration with Anisotropic Orthogonal Procrustes

Analysis", SPIE Medical Imaging, 2014 (**poster presentation**)

Golafsoun Ameri, John S. Baxter, A. Jonathan McLeod, **Uditha L. Jayarathne**, Elvis C.S. Chen, and Terry M. Peters, "Synthetic Aperture Imaging in Ultrasound Calibration", SPIE Medical Imaging, 2014 (**podium presentation**)

Research Abstracts

Uditha L. Jayarathne, Elvis C.S. Chen, John Moore, and Terry M. Peters,"Visualizing US In-situ in Laparoscopic Interventions", **poster presentation** at Symposium of Imaging Network of Ontario (ImNO), 2017 London, Canada

Uditha L. Jayarathne, Xingbiao Luo, Elvis C.S. Chen, and Terry M. Peters,"Simultaneous Estimation of Feature Correspondence and Stereo Object Pose with Application to Ultrasound-augmented Robotic Laparoscopy", **poster presentation** at Symposium of Imaging Network of Ontario (ImNO), 2016 Toronto, Canada

Uditha L. Jayarathne, A. Jonathan McLeod, Terry M. Peters, and Elvis C.S. Chen, "Robust Intraoperative Ultrasound Probe Tracking Using a Monocular Endoscopic Camera ", **podium presentation** at Symposium of Imaging Network of Ontario (ImNO), 2014 Toronto, Canada

Uditha L. Jayarathne, A. Jonathan McLeod, Terry M. Peters, and Elvis C.S. Chen, "Robust Intraoperative Ultrasound Probe Tracking for Ultrasound Augmented Laparoscopy", **poster presentation** at Imaging Discovery Day, 2013 London, Canada

Uditha L. Jayarathne, Chen E.C., Peters T.M., "Visual Tracking for Laparoscopic Video and Ultrasound Augmentation", **poster presentation** at the Symposium of Imaging Network of Ontario (ImNO), 2013 Toronto, Canada