Imperial College London
Department of Computing

Intraoperative Imaging and Tool Tracking for Robotic Surgery

Lin Zhang

2017

Supervised by Professor Guang-Zhong Yang

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Computing of Imperial College London
and the Diploma of Imperial College London

Declaration

This thesis presents original work of the titled author.

That which is not is referenced accordingly.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives license. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

Minimally invasive laparoscopic interventions, despite being widely used within different disciplines of surgical oncology, are restricted by several factors including limited contact sensing and haptic feedback, a loss of stereoscopic depth perception, indirect camera navigation and reduced dexterity. The development of robotic systems in recent years has aimed to address these limitations by adding force-sensing capabilities to the instruments, providing a stereo camera with a robotic holder, and increasing the degree of freedom in the instruments. However, most of the systems only allow teleoperation where the robot directly duplicates the surgeon's motions without any level of autonomy. Despite their increasing application, many robotic systems have been designed without taking full advantage of new imaging techniques that can assist surgical diagnosis and planning via multi-scale anatomical and pathological information. Existing studies on the integration of imaging techniques into intraoperative guidance have demonstrated promising outcomes in terms of improved surgical precision, reduced tumour margins, better nerve-sparing results and an improved consistency of treatment. This research proposes a framework for the automated scanning of various imaging modalities including endomicroscopy, optical coherence tomography and ultrasound. The scanning results are fused with the endoscopic view to provide a clear and intuitive visualisation for the surgeon, aimed at assisting surgical planning. To facilitate the automation of the scanning task via visual servoing, real-time vision-based tracking methods for surgical instruments have been developed. Furthermore, dynamic surgical environments, such as tissue motion, are considered in the autonomous scanning framework,

by estimating the motion characteristics and adapting the motion during the scanning task. The results provide valuable insights into improving surgical diagnosis and navigation for image-guided robotic surgery in the form of autonomous robotics, and motivate the development of autonomous systems for surgical tasks, while reducing the cognitive workload of the surgeon.

Acknowledgements

First and foremost, I would like to express my sincerest gratitude to my supervisor, Professor Guang-Zhong Yang, director of the Hamlyn Centre for Robotic Surgery. He has been supportive throughout four years of my PhD study with his encouragement and innovative ideas. His pioneering vision on robotic surgery has been motivating me when developing an image-guided application for robotic surgery. The Hamlyn Centre not only offers me fabulous lab environment and advanced pieces of equipment but also an opportunity to work with a group of smart and self-motivated researchers.

I would like to thank my fellow doctoral students for their suggestion, cooperation and of course friendship. Especially, I would like to thank Dr Menglong Ye and Dr Petros Giataganas who immensely supported me during the most difficult time of my PhD study. I also want to thank Yang Hu for his selfless help and valuable discussion on my research. To Maria Grammatikopoulou, I thoroughly enjoyed working with you on the nanorobot project.

My sincere thanks also go to Dr Stamatia Giannarou, Dr Michael Hughes, Dr Philip Pratt and Dr Su-Lin Lee, who taught me a lot during collaboration for papers.

My graduate experience would not have been half as fun without my office colleagues and friends. Thank you Dr Alessandro Vandini, Dr Hedyeh Rafii-Tari, Khushi Vyas, Ming Zhao, and Mafalda Camera.

I also want to thank Dr Jindong Liu, Dr Jianzhong Shang, Miss Raphaele Raupp, Gauthier Gras, Mali Shen, who always welcome me and provide as much help as they can when I work at the South Kensington campus.

Last but not the least, I would like to thank my family members who always trust me regardless what I do. Most importantly, I must thank my father, Zongjian, who supports me spiritually throughout my PhD study and my life in general.

Contents

Acknowledgements **5**

1. Introduction **26**
1.1. Research Objectives 28
1.2. Thesis Overview 29
1.3. Publications from this thesis 34

2. Literature Review on Image-Guided Navigation and Surgical Task Automation **36**
2.1. Introduction 36
2.2. Imaging Modalities and Fusion Techniques in MIS . . . 38
2.2.1. Imaging Modalities and Applications 40
2.2.2. Imaging Fusion and Visualisation 47
2.3. Tool Tracking and Navigation Techniques in MIS . . . 50
2.4. Surgical Robotic Platforms 53
2.4.1. Commercial Systems 54
2.4.2. Research Platforms 57
2.5. Robotic Assistance for Surgical Task Automation . . . 59
2.5.1. Classification of Autonomy in Robotic Surgery . 59
2.5.2. Benefits of Surgical Task Automation 62
2.5.3. Developments of Surgical Task Automation . . 64
2.5.4. Challenges of Surgical Task Automation 66
2.6. Conclusion 68

3. Marker based Surgical Instrument Tracking **69**
3.1. Introduction 69
3.2. Design of Hybrid Marker 71

3.3. Efficient Detection of Hybrid Features 72
3.4. Identification of the Hybrid Marker 75
3.5. Marker Tracking using Optical Flow and Homography 76
3.6. Pose Estimation and Ambiguity Elimination 77
3.7. Experiments and Results 79
3.7.1. Hardware Set-up 79
3.7.2. Work Space Analysis 80
3.7.3. Detection Rates 82
3.7.4. Pose Estimation Error 85
3.7.5. Example Application for Ultrasound Overlay . 86
3.8. Discussion and Conclusion 86

4. Model based Instrument Tracking for Robotic Surgery 88
4.1. Introduction 88
4.2. Part-based Online Templates for Tool Detection 90
4.3. Tool Part Verification via 2D Geometrical Context . . 94
4.4. From 2D to 3D Tool Pose Estimation 95
4.5. Experiments and Results 96
4.6. Discussion and Conclusion 99

5. Autonomous Endomicroscopic Scanning and Fusion on Surface 101
5.1. Introduction 101
5.2. Hardware and Software Overview 103
5.3. Trajectory Planning Based on 3D Reconstruction . . . 106
5.4. Pose Estimation with Kalman Filter 108
5.5. Closed-Loop Scanning with Visual Servoing 110
5.6. Endoscopic 3D Fusion and Visualisation 111
5.7. Experiments and Results 113
5.7.1. Stereo Validation 113
5.7.2. Hand-eye calibration accuracy 113
5.7.3. Accuracy of Visual Servoing 115
5.7.4. Phantom and *Ex Vivo* Experiment 116

5.8. Discussion and Conclusion 118

6. Autonomous Endomicroscopic Scanning and Fusion with Depth Information 121

6.1. Introduction 121
6.2. Overview of the Scanning Framework and Workflow . . 123
6.3. Closed-Loop Scanning using pCLE and OCT images . 128
6.3.1. Servoing using probe-based Confocal Laser Endomicroscopy (pCLE) Images 129
6.3.2. Servoing using Optical Coherence Tomography (OCT) Images 130
6.4. Multiscale Fusion and Visualisation 131
6.5. Experiments and Results 133
6.5.1. Accuracy of pCLE Visual Servoing 133
6.5.2. Robustness to Unexpected Motion 134
6.5.3. Validation of OCT Distance Servoing 135
6.5.4. OCT Volume Rendering and Fusion 136
6.6. Discussion and Conclusion 137

7. Motion Compensated Ultrasound Scanning 145

7.1. Introduction 145
7.2. Motion Estimation and Online Update 147
7.2.1. Tissue Motion Estimation 147
7.2.2. Online Motion Update 150
7.3. Motion-Compensated Ultrasound Scanning 151
7.4. Tumour Segmentation and Visualisation 152
7.5. Validation and Experimental Results 153
7.5.1. Experimental Setup 153
7.5.2. Motion Estimation and Stabilisation 154
7.5.3. Tumour Scanning under Tissue Motion 155
7.6. Discussion and Conclusion 157

8. Conclusion and Future Research Directions 158

8.1. Achievements and Contributions of this Thesis 158

8.2. Future Research Directions 161
8.3. Conclusion 163

A. Autonomous Ultrasound Scanning for Intraoperative Tumour Localisation and Diagnosis **165**
A.1. Materials and Methods 166
A.2. Results 168
A.3. Discussion 170

B. Vision-guided Learning by Demonstration for Adaptive Surgical Robot Control **171**
B.1. Materials and Methods 172
B.2. Results 174
B.3. Discussion 176

C. Image Re-Use Permissions **177**

Bibliography **178**

List of Tables

2.1. Comparison between imaging modalities for Minimally Invasive Surgery (MIS) in terms of their penetration depth, resolution and acquisition rate. 40

3.1. Comparison of visual marker based tracking methods for intraoperative ultrasound probe 70
3.2. Work space for the proposed marker on various tool size 80
3.3. Detection rates of different markers in different work distances 84
3.4. Pose estimation error for the proposed marker on different tool size 84

4.1. Translation and rotation errors (mean±std) on Seqs.1-6. Tracking accuracies with run-time speed in Hz (in brackets) compared to [141] on their dataset (Seqs.7-12). 98

5.1. Stereo Reconstruction Accuracy 113
5.2. Error of Hand-eye calibration 114
5.3. Camera Visual Servoing Error (Mean ± Std) 116

6.1. Evaluation of Endomicroscopic Servoing 135

7.1. Estimation results of parameter on three different motion profiles 153
7.2. Quantitative results of tumour scanning under three different motion profiles. 156

List of Figures

2.1. A demonstration of intraoperative navigation using a preoperative Computed Tomography (CT)/Magnetic Resonance Imaging (MRI) model for image-guided robotic partial nephrectomy ©2012 Springer [6]. (a) Kidney model with tumours built from CT/MRI scan; (b) An augmented endoscopic view of the tumour using inverse realism method [7]; (c) Excision boundary for the tumour dissection; (d) Overlay the boundary in the endoscopic view. 39

2.2. A demonstration of intraoperative ultrasound overlay in robot-assisted partial nephrectomy ©2014 Elsevier [9]. The ultrasound image is superimposed on the endoscopic view (a) with a cutaway and 1-mm ruler (b). . 41

2.3. Laparoscopic white light images (a,c) of metastatic peritoneal nodules and the corresponding narrow band images (b,d) ©2017 Springer[17]. 42

2.4. Examples of using fluorescence for tumour and vessel identification during da Vinci® partial nephrectomy ©2017 Intuitive Surgical. Inc.: assessment of parenchymal perfusion during tumour excision (a-b); identification of vessels in renal hilum (c-d). An example of laparoscopic Augmented Reality (AR) using fluorescence for identification of ischemic area in a bowel loop (e-f) ©2017 Elsevier[21]. 43

2.5. An example application of the endomicroscope in a transoral robotic surgery ©2015 John Wiley & Sons[25]. (a) The endomicroscopy probe is inserted through a da Vinci® Introducer, (b,c) the endoscopic imaging of the oropharynx with corresponding microendoscopic images and hematoxylin-eosin stained histopathologic images from the oropharynx of squamous cell carcinoma (b) and benign mucosa (c). 45
2.6. (a) A view of *in vivo* OCT probe on prostate gland adjacent to cavernous nerve in a rat model, and OCT imaging and the corresponding histopathology of rat cavernous nerve (b) and prostate glandular tissue ©2008 Elsevier[37]. 46
2.7. (a) Laparoscopic OCT for human ovarian cancer diagnostics, (b) OCT image of ovarian stroma in endometriosis with increased vasculature and (c) corresponding histopathology ©2009 Elsevier[38]. 46
2.8. 3D visualisation of intravascular ultrasound images and fluorescence lifetime images. The fluorescence lifetime values are mapped on to the 3D lumen surface. Figure by [24] used under CC BY/Adapted from original. . . . 48
2.9. Fusion of ultrasound and endomicroscopy ©2015 IEEE[50]. 48
2.10. Overlay of FLIM scanning results on the endoscopic view (a) in a da Vinci® robotic surgery (courtecy of Jakob Unger, UC Davis, USA), where a map of FLIM (b-c) is generated via scanning for tissue characterisation. 49
2.11. A white light image (a) and the corresponding fluorescence lifetime image (b) of a coronary artery section. (c) Fusion of the fluorescence lifetime image with OCT image. Reprinted with permission from [51], Optical Society of America. 50

2.12. An *in vivo* study of the intraoperative CT registration using fluorescent markers for robotic surgery. The white light image of the fluorescent markers attached to a porcine kidney (a) and the corrsponding fluorescence image recorded with the da Vinci® Firefly™ system (b). The 3D CT data (c) is registered to the endoscopic view using the fluorescent markers (d) ©2017 Springer[52]. . 51
2.13. The da Vinci® Xi surgical robot consists of (from left to right): surgeon console, vision cart and patient-side cart ©2017 Intuitive Surgical, Inc. 54
2.14. Two surgical robots developed by TransEnterix: Senhance™ system (a) and SurgiBot™ system (b-c) ©2017 TransEnterix, Inc. 55
2.15. The Flex® system includes an operational tower with a touch screen graphical user interface, a framework for stabilisation of compatible flexible instruments and the flexible endoscope. The tip of the flexible endoscope contains 6 LEDs for illumination, a high-definition camera and two working channels ©2015 John Wiley & Sons[83]. 56
2.16. An example experimental setup for the da Vinci® robot with dVRK controllers and an intraoperative ultrasound imaging machine. 57
2.17. Surgical robotic platform for research. (a) RAVEN-II Surgical Robot ©2013 IEEE [97]; (b) KUKA LBR Med robot ©KUKA Aktiengesellschaft; (c) DLR MiroSurge surgical robot ©2009 IEEE [98]. 58
2.18. A diagram showing autonomy level of surgical robot defined in literatures. 62
2.19. (a) Human-machine collaborative framework for surgical pin and suture task ©2011 IEEE[107]. (b) Human-robot collaboration on inclusion segmentation experiment with different control modes ©2016 IEEE [106]. . 64

2.20. Automated surgical task using da Vinci® Research Kit (dVRK). (a) Soft tissue debridement task ©2015 IEEE [89]; (b) Suturing task ©2016 IEEE [90]; (c) endomicroscopic scanning task ©2017 IEEE [109]. 65

3.1. (a) A top view of the proposed marker for cylindrical object tracking. The hybrid marker consists of circular-dot patterns and chessboard vertices. (b) A side view of a cylinder showing related parameters to define a marker's local coordinate frame. 71
3.2. Algorithm workflow of the marker detection algorithm. The binary images are generated using threshold values between 70 and 100 with an interval of 10. The detected dots are circled in red, and the chessboard vertices are labelled in yellow. The first dot in the pattern is circled in blue. 73
3.3. A sequence of images showing how the tracking algorithm reacts when the marker is partially occluded by fake blood. The green outline of the marker indicates successful detection while yellow outline means detection failure and tracking component is used. 76
3.4. Qualitative results of the tracking framework for the proposed marker on various types of surgical tool. A 5mm monopolar cautery (top row), a da Vinci® large needle driver (middle row) and a laparoscopic ultrasound probe (bottom row) is tracked, respectively. The coordinate axis in red, green and blue shows that the pose has been correctly estimated. 77

3.5. An illustration of ambiguous pose removal using chessboard vertices. The red and blue circles indicate projections of the model chessboard vertices using two candidate poses. The two quadrangles represent the projected outline of an ultrasound image on the image. In this example, we can easily see that the pose corresponds to the red is correct. 79
3.6. 3D printed rigid body with various size for validation (*left*) and definition of the marker's local coordinate system (*right*). 80
3.7. An housing adaptor that can hold a planar ultrasound probe. With the housing, the planar probe can be tracked by using the proposed marker. 81
3.8. A snapshot of the experiment for detection rate comparison. For each distance group, the same movement has been reproduced to test the detection performance of different markers. The circular-dot marker cannot be detected while the marker is in a very slant view. In the far distance group, we show a scenario where the marker is rotated around its axial axis, so only the other side is visible. 82
3.9. Ultrasound image is fused with laparoscopic image using an inverse realism technique [7]. 86

4.1. Illustration of transformations used in this chapter. . . 90
4.2. Virtual rendering example of the large needle drivers (a) and the corresponding camera view (b). The predefined keypoint locations are labeled in (a). The red skeletons show the projection based on the tool kinematics, and green skeletons show the tracking results of the proposed framework (b). 91

4.3. (a) Extracted gradient orientations from virtual rendering where the the orientations are quantised and colour-coded as shown in the pie chart. (b) The corresponding camera view with the tracking results. 92
4.4. (a) An example of part-based templates; (b) Quantised gradient orientations from a camera image; (c) Part-based template matching results of tool parts; (d) and (e) Geometrical context verification; (f) Inlier detections obtained after verification. 93
4.5. Detection rate results of our online template matching and GradBoost [139] on two single-tool tracking sequences. 97
4.6. The overall rotation angle errors (mean±std) along each axis of the end-effector on Seqs.1-6. 97
4.7. Qualitative results. (a-c) phantom data (Seqs.1-3); (d) *ex vivo* ovine data (Seq.4); (e) and (g) *ex vivo* porcine data (Seqs.9 and 12); (f) *in vivo* porcine data (Seq.11). Red lines indicate the tool kinematics, and green lines indicate the tracking results of the proposed framework with 2D detections in coloured dots. 98

5.1. An overview of our proposed system framework for autonomous endomicroscopic scanning and 3D mosaicing. 104
5.2. A description of coordinate systems (frames) and related transformations. B: robot base frame; E: end-effector frame; M: marker frame; D: probe tip frame; C: camera frame; I: mosaic point cloud frame. 105
5.3. Illustration of two types of trajectory pattern for pCLE scanning: spiral (left) and raster (right). Three parameters that define the trajectory are labelled with red double-arrow line. 106

5.4. Snapshots of the phantom and *ex vivo* experiments. (a) and (c) Experimental setups; (b) and (d) A snapshot of the surgical view captured by the laparoscope. The KeyDot® marker was overlaid with the streaming microsopic images for AR. 112
5.5. As the marker moves away from the starting reference pose, the predicted pattern (projection on image is shown in red dots) deviates from the detected one. 114
5.6. Demonstration of mosaic results on a printed grid phantom with black holes (outlined by the dashed rectangle). Three trials are identified by three different colours and a red guide axis is placed for assisting comparison. . . 117
5.7. Comparison of spiral and raster trajectories. The actual trajectories measured on the mosaic images are shown in red dotted line. 118
5.8. Exemplar 3D fusion results. (a) and (b) online multi-region fusion results from the *ex vivo* gastric mucosa experiment; (c) and (d) Online multi-region fusion results from phantom experiment. 119

6.1. An overview of our experimental setup for autonomous optical biopsy probe scanning and multi-scale fusion. . 124
6.2. The data flow streaming from the different imaging modalities is processed for visualisation and servoing purposes. From a pair of stereo images, a surface of the scene is reconstructed as a point cloud. By stitching pCLE images, a mosaic image can be created, and a 3D volume can be built from OCT images. These results are fused into a unified window for multi-scale visualisation. . . . 126

6.3. Illustration of the coordinate frames and dimension of the hardware; (a) shows a da Vinci® tool grasping the pick-up adapter holding the pCLE and OCT probes. The dimension of the adapter and probes is labelled; (b) shows the definition of the coordinate frames used in the visual servoing and fusion. The marker and pCLE probe coordinate frames are denoted as {**M**} and {**P**} respectively. The volume and mosaic image coordinate frames are denoted as {**V**} and {**I**} respectively; (c) shows an augmented laparoscopic view where a pCLE image is overlaid on the adaptor. 128
6.4. Resulting mosaics from pCLE servoing on a printed grid phantom. Results from trial 2, 4, 7 and 10 are shown here. In all trials, the mosaic image shows the grid pattern clearly. An example of misaligned square due to the inaccuracy in the visual servoing is indicated in trial 2. 132
6.5. Illustration of how pCLE servoing improves mosaicing results. The left column shows the scanning trajectory based on kinematic reading of the robot's end-effector and the right column shows the corresponding mosaic results on the same phantom. The first row is the result from using only the kinematic model, while the second row uses closed-loop scanning. The effect of unexpected motion on the mosaic image is shown in the third row. The unexpected motion is shown in the red dashed circle in (f) before the closed-loop servoing recovers and returns the probe to the desired trajectory relative to the tissue surface. 139

6.6. Mosaic results obtained using the framework under unexpected motion with various velocities. The 'arm' connecting the mosaic over the planned region and the additionally scanned region (caused by the motion) is enlarged to show that it is continuous, suggesting that the probe correctly returned to its planned trajectory on the tissue surface. 140
6.7. Demonstration of the correlation between the distance to the tissue surface measured in the OCT image (bottom row) and the quality of pCLE images (top row). As the surface moves closer to the probe, from left to right in the figure, the probe-to-tissue distance measured by the OCT probe becomes smaller while the pCLE image quality improves. Note that the pCLE images are autocontrasted. 141
6.8. The desired OCT distance shown in red is assumed to be constant during servoing. As the phantom starts moving, the deviation between the desired potion and the current position measured by the OCT channel initially increases and then decreases towards zero as the system corrects for the motion. Finally, the deviation converges to zero when the phantom stops. 1 pixel in the OCT image corresponds to approximately 10 μm in real space. 142
6.9. Demonstration of pCLE scanning with and without OCT distance compensation on a curved surface. Without the OCT distance servoing, the mosaic could only continue for a short distance before the probe lost contact, the pCLE image was degraded and the lateral visual servoing failed. When OCT distance servoing was turned on, a mosaic was generated over the whole trajectory. . 143

6.10. A volume reconstruction of a surgical needle with a thickness of 0.3 mm. The red rectangle indicates the segment that has been reconstructed in the microscopic volume shown on the left. 144
6.11. An example of multiscale fusion. A macro scale stereo reconstruction is shown (a). To link with the micro scale, the mosaic image is overlaid on the stereo reconstruction (b). To visualise the micro scale images, a pCLE mosaic and an OCT volume are shown. 144

7.1. (a) Coordinate frames defined in the proposed system. {**B**}: robot base; {**E**}: robot end-effector; {**M**}: marker attached on the probe; {**D**}: transducer; {**C**}: stereo camera; {**S**}: tissue surface; {**U**}: ultrasound image (b) Hardware Setup; 147
7.2. Flowchart of the proposed framework. MC: Motion Compensated; VS: Visual Servoing. 148
7.3. Illustration of motion estimation using Principle Component Analysis (PCA). The tissue motion is tracked in the camera coordinate (a) and projected in the PCA space (b). The first principle component, which corresponds to the largest motion variation, is learned by fitting a respiration model to it. 149
7.4. Small errors in the period of the learned motion characteristics can result in large phase drift during the scanning task. The online updated tissue motion can keep up with the ground truth while the learned motion gradually deviates from the actual tissue motion 150
7.5. Quantitative results of ultrasound stabilisation on phantoms (a) and *ex vivo* tissue (b) with and without Motion Compensation (MC). 153

7.6. Qualitative results of ultrasound stabilisation. The top and bottom rows are the trials without and with motion compensation, respectively. Three motion profiles were tested on a phantom (a) and *ex vivo* tissue (b-c), with directions along the X-, Y- and Z-axis of the motorised device. 154
7.7. Comparison of the estimated and ground truth motion for different motion characteristics. 155
7.8. Qualitative results of two example autonomous scanning tasks on a phantom (a-d) and *ex vivo* tissue (e-h). The obtained tumours and 3D tissue surfaces are visualised in (d) and (h). See the supplementary video via `https://youtu.be/EsomHSwIz2g`. 156

8.1. (a) An example use of a navigation aid (called DART) for intraoperative tumour dissection Figure by [178] used under CC BY/Cropped from original. (b) The procedure has two phases: the surgeon first perform ultrasound scanning after the DART is placed near the tumour, then AR can be performed by tracking the DART during tumour dissection ©2016 Springer [179] 162

A.1. Coordinate frames defined in this work: surface frame $\{\mathbf{S}\}$, marker frame $\{\mathbf{M}\}$, end-effector frame $\{\mathbf{E}\}$, robot base frame $\{\mathbf{B}\}$, ultrasound image frame $\{\mathbf{U}\}$ and camera frame $\{\mathbf{C}\}$. 167
A.2. A 3D visual comparison between the CT ground truth (a) and reconstructed models of proposed framework (b-c). The 10mm and 15mm tumours are coloured in red and green respectively. The green points in b-c show planned trajectory for the ultrasound scanning, where the yellow arrows indicate the scanning direction. . . . 168

A.3. Top row: the reconstructed tumour is overlaid on laparoscopic images to provide an augmented view for tumour localisation and diagnosis. Bottom row: ultrasound image is overlaid according to current position of the probe. Left and right column show examples of the 10mm and 15mm tumour respectively. 169

B.1. Experimental setup of the da Vinci® system with the dVRK (a), needle tracking for moving the needle from initial position (b) to desired target insertion point (c), wound detection under different camera views (d, e). . 172

B.2. The running suture task (a) and the learned motion for "Subtask 1" (b) as reproduced from multiple demonstrations using the HSMM. The ellipses correspond to the covariances for each state of the HSMM. 174

B.3. Examples of the needle driver tip trajectory for automated initialisation of the needle at the desired insertion point (orange) followed by automated playback of the learned motion for the first subtask (blue) under different initial conditions in the robot base coordinate system [mm]. 175

List of Acronyms

AR	Augmented Reality
CAS	Computer-Assisted Surgery
CLE	Confocal Laser Endomicroscopy
CT	Computed Tomography
DOF	Degree-of-Freedom
DVRK	da Vinci® Research Kit
EKF	Extended Kalman Filter
EM	electromagnetic
EPNP	Efficient Perspective-n-Points
FOV	Field-of-View
IGS	Image-Guided Surgery
IME	Interquartile Mean Error
LUS	Laparoscopic ultrasonography
MC	Motion Compensation
MIS	Minimally Invasive Surgery
MRI	Magnetic Resonance Imaging
NCC	Normalised Cross-Correlation

OCT	Optical Coherence Tomography
PCA	Principle Component Analysis
PCLE	probe-based Confocal Laser Endomicroscopy
PNP	Perspective-n-Points
PSM	Patient Side Manipulator
QGO	Quantised Gradient Orientations
RMIS	Robot-assisted Minimally Invasive Surgery
RMSE	Root Mean Square Error
ROS	Robot Operating System
VS	Visual Servoing
VTK	Visualisation Tool Kit

1. Introduction

It was not until the mid-19th century that surgery was considered a valuable treatment for human diseases or injuries due to the previous lack of proper pain control through anaesthesia. The discovery of anaesthesia allowed more intricate operations in the internal regions of the human body to take place. However, more dangerous post-operative infections were caused by inadequate antiseptic techniques during surgery. The introduction of sterile surgery and the use of anaesthetic laid the foundation for modern invasive surgical techniques. Wilhelm Roentgen discovered X-rays, which could be used as a medical diagnostic tool allowing the internal structure of the human body to be seen and examined without the need to make an incision. Since the 20^{th} century, the improvements in medical imaging technology have promoted the development of Image-Guided Surgery (IGS) in which surgeons are able to examine patient's anatomy by using CT, MRI, positron emission tomography and ultrasound scans. Over the past fifty years, these imaging modalities have been extensively used for preoperative surgical planning, as well as in intraoperative navigation in the context of the wider field of Computer-Assisted Surgery (CAS). Similar to how GPS assists a driver in navigating a city, CAS guides a surgeon around the patient's anatomy during a procedure. The two key components of CAS are the registration of the patient's anatomical model alongside the intraoperative scene and the real-time tracking of the related surgical instruments and anatomy. The challenge is minimising the errors in the registration and tracking process which could lead to an unreliable map for surgical guidance. In recent years, there has been increasing use of intraoperative imaging modali-

ties for CAS which includes ultrasound scans, Confocal Laser Endomicroscopy (CLE) and OCT. To accommodate the application of these intraoperative imaging modalities, instrument tracking is essential in order to provide the correct information to the surgeon.

With recent advances in medical imaging, surgeons are able to perform diagnoses or surgeries inside a patient's body without making large incisions by using an endoscope. This enables MIS where special surgical instruments are inserted via a few small incisions. There are several benefits of MIS to the patient compared to open surgery. These include reduced haemorrhaging, less pain, a shorter hospital stay and less chance of postoperative complications. Despite improved patient outcomes, MIS is more difficult from the surgeon's perspective due to loss of dexterity, poor depth perception and the non-intuitive motor skill caused by the fulcrum effect.

To address the above issues, robotic systems have been introduced into operating theatres as smart instruments with high precision and improved dexterity. One of the most successful commercial systems for Robot-assisted Minimally Invasive Surgery (RMIS) is the da Vinci® surgical robot (Intuitive, CA). The da Vinci® robot allows the surgeon to perform surgery from a seated position at a master console with a magnified three-dimensional (3D) image of the surgical field. The movements of the surgeon's hands are duplicated on the instruments with jointed-wrist design through a robotic arm inside the patient's body. It has been applied to cardiac, colorectal, gynecological, thoracic and urologic surgery. However, the system is expensive, which is the main criticism of the system.

Surgical robots can easily incorporate medical imaging techniques for navigation and guidance during RMIS. An example of this has been presented in [1] where preoperative imaging data is registered to an intraoperative endoscopic view in which the surgeon can see subsurface information via AR techniques. The capabilities of the surgical robot have been largely extended due to the use of medical imaging techniques. However, the maximum potential of the robot cannot be

reached without taking full advantage of artificial intelligence. Surgical robots can offer the possibility of automating the whole or parts of the procedures which require high accuracy and repetitive actions while reducing the cognitive workload of the surgeon. The potential challenges of automating surgical tasks include the dynamic environment with deforming tissue, the lack of force sensing between an instrument and the tissue, and variable surgical plans for different patients. Furthermore, ethical and legal issues may prevent the use of autonomous surgery. Robust algorithms and hardware designed to automate surgical tasks by using medical imaging and computer vision techniques are the crucial components of autonomous surgery.

1.1. Research Objectives

Surveys such as that conducted in [2] have shown that fatigue can negatively impact a surgeon's performance in terms of accuracy, leading to tissue damage, increased leak rates, non-goal-directed motion and operating time. These findings highlight the benefits of automating surgical tasks or subtasks that are tedious and repetitive or that require high precision.

Numerous studies [3, 4] have demonstrated that systems with low-level shared control between the surgeon and the robot, such as the active constraint technique [5], show promise over freehand in terms of improved accuracy and shorter completion time. With regards to a control framework that automates surgical tasks using a robot, visual servoing, which uses visual information to control the robot, is one of the most effective approaches as video of the target scene that is captured by the laparoscope is always available. In order to build an autonomous framework for executing surgical tasks, the framework is required to have several essential components: reliable and real-time tracking of surgical instruments, accurate 3D scene recovery and a visual servoing method that is robust to errors in the system calibration and is able to deal with a dynamic environment such as moving tissue.

This thesis has four major aims. Firstly, it aims to develop a vision-based instrument tracking algorithm that reliably runs in real-time, facilitating the automation of surgical tasks via visual servoing. Secondly, this thesis proposes an autonomous scanning framework for optical biopsy using pCLE and OCT, which are held by a pick-up mount that is grasped by the da Vinci® robot. In this scanning framework, the 3D surface of the surgical field is recovered so that an optimal trajectory for the scanning task can be planned accordingly. The surgeon does not directly control the robot but is responsible for the selection of the scanning region, the inspection of the planned trajectory, as well as the supervision of the scanning motion executed by the robot. Thirdly, several visualisation methods for different imaging modalities are investigated which include the overlay of pCLE images in the endoscopic view; a unified visualisation consists of both macro and micro views of the images in different scales. Fourthly, the autonomous scanning framework is applied to intraoperative ultrasounds with an extension where the dynamic environment is considered in order to deal with the moving tissue caused by respiration.

1.2. Thesis Overview

The outline of the thesis is as follows:

Chapter 2 provides an introduction to the medical imaging modalities and their applications in surgical diagnostic and planning, as well as the fusion and visualisation of different imaging modalities. It then reviews surgical tool tracking techniques that can be utilised for navigation and automation in robotic surgery. Next, this chapter briefly reviews surgical robotic platforms used for laparoscopic surgery. In the last part, the robotic assistance for surgical task automation is discussed. In particular, this chapter tries to define different levels of autonomy of surgical robotics based on the existing literature. The benefits of automating surgical tasks in MIS are described followed by

several sample applications. Furthermore, this chapter presents the challenges faced in developing autonomous surgical robots, taking into consideration potential ethical and legal concerns.

Chapter 3 proposes a novel marker design to facilitate tracking of cylindrical instruments. The marker consists of three patterns of dots and corners which can be detected efficiently in endoscopic images. The corresponding tracking framework for the proposed marker improves its reliability compared to previous methods. Thanks to the marker design, the visibility of the marker is significantly improved, providing a large working space. In addition to the detection process, the temporal information of the marker is utilised in the tracking in case of partial occlusion or reduced visibility. The estimation method for the instrument pose eliminates the ambiguity of the pose by utilising additional chessboard corners on the marker. To test the applicability and the usability of the proposed tracking framework, it is modified for different tool sizes (6 mm to 14 mm in diameter). Furthermore, validation results show that the tracking framework provides low pose estimation error while tracking the instrument in real-time. An example of the application of instrument tracking is that ultrasound images are being overlaid on the endoscopic images using tracked ultrasound probe positions to provide an augmented view for intraoperative surgical navigation in MIS.

Chapter 4 proposes a model-based real-time tool tracking framework for robotic surgery. A CAD model of the robotic tool is used to generate the virtual contour templates based on the robot's kinematics. The tool detection is realised by applying a template matching method on the individual parts of the tools, allowing for articulated tool tracking. The outlier template matches of the tool parts are rejected based on the 2D geometrical context. The 3D pose is then estimated by using an Extended Kalman Filter (EKF). Experiments are conducted on phantom, *ex vivo* and *in vivo* video data, and the results show that the proposed method outperforms the state-of-the-art methods.

Chapter 5 presents an autonomous framework using the da Vinci® robot for endomicroscopic mosaicing and the online fusion of the reconstructed 3D endoscopic view with microscopic images. The framework utilises both laparoscopic images and pCLE images to close the control loop to achieve the level of precision required for high-quality mosaic generation, which relies on a sufficient overlap between consecutive image frames. The framework also provides a comprehensive visualisation and fusion of the multi-scale images by overlaying the pCLE images onto the surgeon's view, while generating 2D mosaics in real-time, and fusing the mosaics into a 3D stereo reconstruction of the surgical scene. Furthermore, different trajectory patterns are compared via a phantom experiment, and the spiral pattern which leads to less robot backlash is chosen for the following experiments, including the ones presented in the following chapter.

Chapter 6 extends the autonomous framework by including OCT into the scanning task for tissue micro-structure assessment. The framework can now be used for the automated 2D scanning of the pCLE/OCT probes, providing tissue surveillance of a large area through image stitching. Closed-loop control of the robot on the pCLE image plane (in the lateral position) can result in the loss of probe contact with the tissue. Therefore, the OCT images are used to control the robot to maintain the desired distance along the axial direction. The OCT images are used to create a 3D volume of the tissue structure which can be fused with a stereo reconstruction and a mosaic image in the fusion framework. This provides the surgeon with a multi-scale 3D view of the operating site.

Chapter 7 adopts the autonomous framework for intraoperative ultrasound scanning and extends it to consider the dynamic environment, such as moving tissue caused by respiration. The tissue motion is assumed to be periodic so that the underlying tissue motion characteristics can be estimated before the scanning by utilising an effective vision-based approach. Next, an optimal scanning trajectory is gener-

ated by considering both the tissue surface shape and the ultrasound transducer dimensions. The motion model is then incorporated into the visual servoing framework. Furthermore, the motion model is updated online to compensate for errors in the model. The proposed system is validated with both phantom and *ex vivo* experiments.

Chapter 8 summarises the key findings and contributions of this thesis and addresses the emerging trends and future research directions for robot autonomy in surgery.

The key contributions of this thesis include:

- Designing a novel hybrid marker and the corresponding tracking framework that enables tracking of cylindrical instruments in real-time and allows large working space.
- Developing a model-based tracking algorithm for robotic instrument that has real-time performance, by incorporating both kinematic and imaging information to estimate pose of the articulated robotic instrument.
- Proposing an autonomous framework for pCLE/OCT scanning that allows for the generation of large mosaics with minimal intervention from the surgeon.
- Presenting a visual servoing technique to control the da Vinci® robot to move smoothly for the scanning task, by utilising multiple imaging modalities including the laparoscope, pCLE and OCT.
- Providing the surgeon with an intuitive visualisation of multi-scale image information in order to support the surgical diagnosis and the decision-making, by developing an online multi-scale fusion scheme for robotic surgery.

- Presenting a trajectory planning method for the ultrasound scanning task that ensures efficient coverage and the optimal probe orientation as well as a good tissue contact, by considering surface curvature and the dimension of the ultrasound transducer.
- Developing an extension of the autonomous framework for intraoperative ultrasound scanning of the tumour under tissue motion, by estimating the motion model using vision-based techniques and updating the model online to correct the errors.

1.3. Publications from this thesis

The work presented in this thesis has resulted in a number of peer-reviewed full conference papers and journal publications:

- **Zhang, L.**, Ye, M., Giannarou, S., Pratt, P., Yang, G.-Z.: Motion-Compensated Autonomous Scanning for Tumour Localisation using Intraoperative Ultrasound. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 619–627 (2017).
- **Zhang, L.**, Ye, M., Chan, C., Yang, G.-Z.: Real-Time Surgical Tool Tracking and Pose Estimation using a Hybrid Cylindrical Marker. *International Journal of Computer Assisted Radiology and Surgery*. 12(6), 921–930. (2017).
- **Zhang, L.**, Ye, M., Giataganas, P., Hughes, M., Bradu, A., Podoleanu, A., Yang, G.-Z.: From Macro to Micro: Autonomous Multiscale Image Fusion for Robotic Surgery. *IEEE Robotics and Automation Magazine*. 24(2), 63–72 (2017).
- **Zhang, L.**, Ye, M., Giataganas, P., Hughes, M., Yang, G.-Z.: Autonomous scanning for endomicroscopic mosaicing and 3d fusion. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 3587–3593 (2017).
- **Zhang, L.**, Ye, M., Chan, C., Yang, G.-Z.: Autonomous Ultrasound Scanning for Intraoperative Tumour Localisation and Diagnosis. In: Hamlyn Symposium. pp. 5–6 (2016).
- Ye, M., **Zhang, L.**, Giannarou, S. and Yang, G.-Z.: Real-time 3D Tracking of Articulated Tools for Robotic Surgery. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) pp. 386–394 (2016).

- Pratt, P., Hughes-Hallett, A., **Zhang, L.**, Patel, N., Mayer, E., Darzi, A. and Yang, G.-Z.: Autonomous ultrasound-guided tissue dissection. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 249–257 (2015).

- Rafii-Tari, H., Vandini, A., **Zhang, L.,** Hughes-Hallett, A., Yang, G.-Z.: Vision-guided Learning by Demonstration for Adaptive Surgical Robot Control. In: Hamlyn Symposium. pp. 39–40 (2015).

Additional journal and conference paper related to computer vision and robotics that have been co-authored:

- Zhang, Z., **Zhang, L.**, Yang, G.-Z.: A computationally efficient method for hand–eye calibration. *International Journal of Computer Assisted Radiology and Surgery*. pp. 1–13 (2017).

- Grammatikopoulou, M., **Zhang, L.**, Yang, G.-Z.: Depth Estimation of Optically Transparent Laser-Driven Microrobots. In: IEEE International Conference on Intelligent Robots and Systems (IROS). [In Press] (2017).

- Ye, M., Johns, E., Handa, A., **Zhang, L.**, Pratt, P., Yang, G.-Z.: Self-Supervised Siamese Learning on Stereo Image Pairs for Depth Estimation in Robotic Surgery. In: Hamlyn Symposium. pp. 27–28 (2017).

2. Literature Review on Image-Guided Navigation and Surgical Task Automation

2.1. Introduction

The adoption of MIS is largely due to its benefits of small keyhole incisions on patients. The small incisions lead to significantly less blood loss compared to open surgery, and therefore a faster recovery and shorter hospital stay. However, it is physically more difficult for the surgeon who has to look at a distant monitor, which provides a 2D view of the operating site. Furthermore, the laparoscope is being held by an assistant without the surgeon having direct control. The assistant can suffer from fatigue during long operations which results in an unsteady field of vision. In addition to the vision issues, the instruments used in laparoscopic surgery introduce a fulcrum effect due to the incision being made in the abdominal wall. The fulcrum effect requires the surgeon to move the instruments reversely, which requires a prolonged and sustained learning curve.

The disadvantages of MIS can be addressed by surgical robots such as the da Vinci® system where a 3D stereo display and instruments with wrist movement are provided. The surgeon is able to sit in an ergonomically comfortable pose and rest the arms on a padded bar.

The instruments are controlled by the surgeon directly using a pair of master manipulators. The surgeon's hand motions are scaled and duplicated on the instrument with tremor filtration. Hence, the surgical robot extends the surgeon's ability as accurate and small movements are achievable, allowing for more complex surgeries to be conducted with a minimally invasive approach. The robotic manipulators are under the direct control of the surgeon in a teleoperation scheme where no intelligence is available. Existing research has established that some automated surgical tasks have comparable or better outcome in terms of precision and speed compared to the surgeon completing the tasks manually. The automation of tedious and repetitive tasks can reduce the fatigue of the surgeon during long operations, which avoids potential safety issues.

Advances in imaging technology (such as CT, MRI, ultrasound and endomicroscopy) have had a significant improvement in disease diagnosis and surgical planning. Both preoperative and intraoperative imaging modalities can be used in RMIS to allow surgeons to see inside the body and underneath the tissue so that complications can be minimised. Several studies have shown that the image guidance for robotic surgery can be achieved by using the patient model from a preoperative CT scan and an intraoperative ultrasound. Recent developments in imaging technology have revealed that the molecular and cellular imaging modalities such as endomicroscopy and OCT can be used in RMIS where metabolical and physiological information is available for surgeons in real-time. Another benefit of autonomous surgery is the real-time fusion of different imaging information from the macro scale to the micro scale, which allows surgeons to obtain a more complete understanding of the condition.

Autonomous surgery allows for different imaging information to be incorporated and surgical tasks to be completed with robot control. This chapter also explores the different levels of autonomy in robotic surgery, followed by the advantages of an autonomous surgical system. Both commercial products and the research surgical robotic platforms

are introduced. Lastly, the current challenges in developing a clinically acceptable system with automated features are discussed.

2.2. Imaging Modalities and Fusion Techniques in MIS

Image-guided techniques are fast becoming an essential component of MIS to address some of its limitations including the lack of tactile feedback to the surgeon and a reduced 3D perception. A major challenge of MIS is that the localisation of the targeted anatomical structures, such as non-superficial tumours and critical nerves, is almost impossible with the human eye if no haptic feedback is provided. Without the exact location and the dimension of these targets, the decision-making during surgery becomes a very challenging task. Some studies have presented the efficacy of using preoperative imaging for surgical guidance and navigation. However, it is not sufficient to only rely on preoperative images because the shifting of abdominal organs under the pneumoperitoneum may cause differences between the preoperative images and the surgical field. To address this issue, preoperative imaging data can be registered with intraoperative imaging data to provide updated real-time navigation for laparoscopic procedures. While conventional imaging modalities, such as X-rays/CT, MRI, and ultrasound scans, have been routinely used in IGS, other optical imaging techniques that allow for cellular imaging and the evaluation of tissue micro-scale structure have been emerged for surgical oncology. New imaging techniques, which provide real-time structural or pathological information of the tissue, can assist in tumour margin delineation during the procedure. Furthermore, imaging data from multiple modalities can be fused to provide multi-scale comprehensive tissue information.

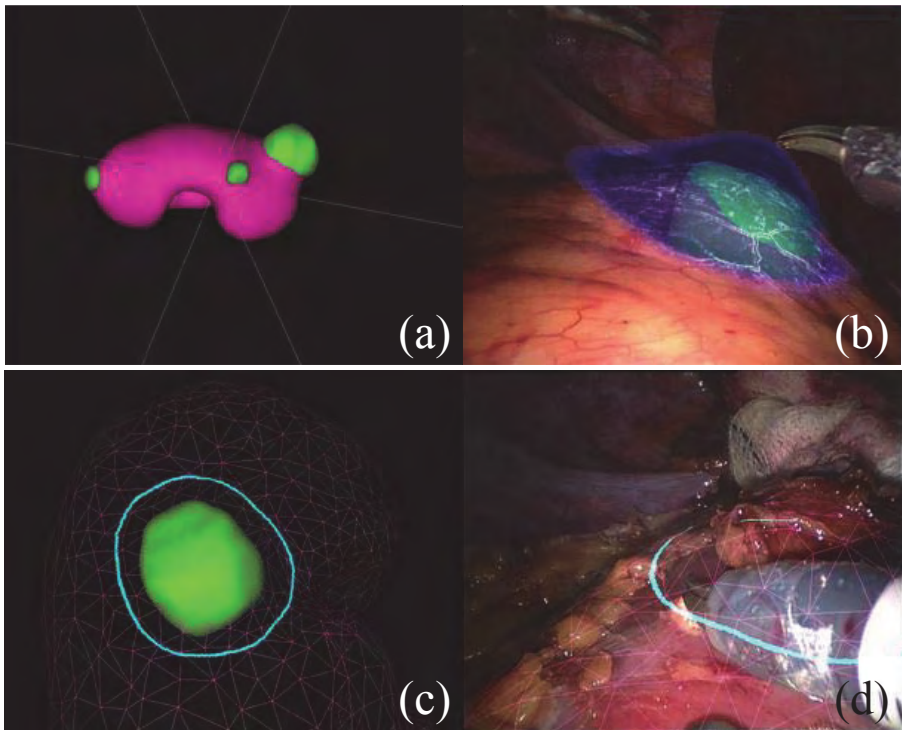


Figure 2.1.: A demonstration of intraoperative navigation using a pre-operative CT/MRI model for image-guided robotic partial nephrectomy ©2012 Springer [6]. (a) Kidney model with tumours built from CT/MRI scan; (b) An augmented endoscopic view of the tumour using inverse realism method [7]; (c) Excision boundary for the tumour dissection; (d) Overlay the boundary in the endoscopic view.

Table 2.1.: Comparison between imaging modalities for MIS in terms of their penetration depth, resolution and acquisition rate.

Technique	Image depth	Resolution	Acquisition rate
CT	–	∼1 mm	video rate
MRI	–	∼1 mm	up to 30 Hz
Ultrasound	10 to 100 mm	20 to 100 µm	video rate
Fluorescence	∼10 mm	camera resolution	video rate
NBI	1 to 2 mm	camera resolution	video rate
CLE	∼100 µm	∼1 µm	10 to 120 Hz
OCT	up to 3mm	∼ 10 µm	up to 120 Hz

2.2.1. Imaging Modalities and Applications

CT and MRI are two primary imaging modalities used for IGS in which the image data is often used preoperatively to create an anatomical atlas for surgical planning and navigation. While CT provides accurate and relatively fast imaging at the expense of the exposure to radiation, MRI provides high-resolution imaging without the radiation issue but with much lower acquisition speeds. Both modalities are utilised for intraoperative navigation by using the preoperative patient model that is registered with the endoscopic view, as shown in Fig. 2.1. Furthermore, intraoperative CT and MRI imaging techniques have been developed in relation to MIS for locating abnormal tissue and avoiding critical structures. The intraoperative imaging can be registered with the endoscopic view to assisting navigation and decision-making [8] by registering the CT image data with the endoscopic view manually or via marker-based methods. However, the shortages of CT, such as its cumbersome and slow registration and the exposure to radiation, restrict its intraoperative applications. Although intraoperative MRI does not have the radiation issue, it has very a slow update rate, and it requires specialised instruments for intraoperative usage. Nevertheless, MRI-compatible instruments and robotic devices are being actively developed and translated into clinical practices.

Laparoscopic ultrasonography (LUS) is a type of medical sonogra-

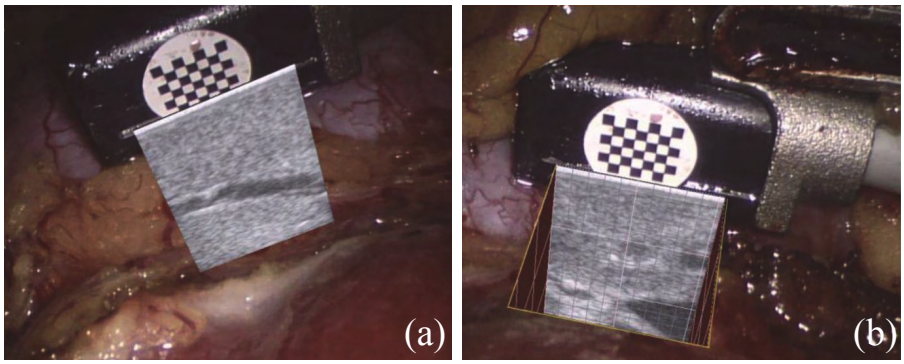


Figure 2.2.: A demonstration of intraoperative ultrasound overlay in robot-assisted partial nephrectomy ©2014 Elsevier [9]. The ultrasound image is superimposed on the endoscopic view (a) with a cutaway and 1-mm ruler (b).

phy in which a specialised transducer is required. The transducer must be able to fit through a 10 mm laparoscopic trocar. Existing studies have suggested that LUS is beneficial for MIS by providing staging information of the tumours below the surface of organs. LUS has been commonly used in laparoscopic liver surgery for tumour identification and localisation [10, 11], laparoscopic cholecystectomy for bile duct screening [12], and pancreas surgery for cancer assessment [13]. Recently, LUS has also been used to facilitate intraoperative tumour localisation during partial nephrectomy through the utilisation of advanced visualisation techniques such as AR [14, 9], as shown in Fig. 2.2. Robotic ultrasound probes have been developed which can be grasped by a robotic instrument to improve the manipulability of the probe [15]. Furthermore, intraoperative ultrasound imaging can be integrated with preoperative image data to develop a reliable navigation system [16] for MIS by registering the preoperative patient model with the ultrasound volume.

Narrow band imaging (NBI) is a non-invasive optical imaging technique that uses only two wavelengths of light to visualise the mucosa and blood vessels. NBI offers high-contrast images that highlight the mucosal surface layer and the haemoglobin-rich areas such as blood

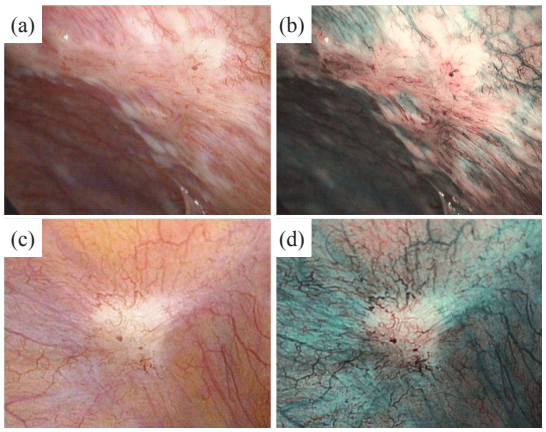


Figure 2.3.: Laparoscopic white light images (a,c) of metastatic peritoneal nodules and the corresponding narrow band images (b,d) ©2017 Springer[17].

vessels. As shown in Fig. 2.3, NBI images can assist surgeons in defining resection margins for a cancerous tissue by emphasising features such as capillary and mucosal patterns. In transoral robotic surgery (TORS), NBI is used to detect early superficial pharyngeal cancers [18] which tends to be difficult when using a standard endoscope. Additionally, NBI is utilised in laparoscopic surgery for gastric cancer [17], bile leak [19] and ovarian cancer [20].

Near-infrared fluorescence (NIRF) imaging is an innovative imaging technique that uses contrast agents with fluorescent characteristics in the near-infrared spectrum. The most widely-used contrast agent is indocyanine green (ICG), although there are other clinically-approved fluorescent dyes such as methylene blue. The da Vinci robot also has fluorescence imaging capability, as shown in Fig. 2.4(a-d). Numerous studies have demonstrated that NIRF imaging is a safe and effective method to localise vessels and tumours from the surrounding tissue in many types of surgery [22, 23] including partial nephrectomy, cholecystectomy, colorectal surgery, ureteroureterostomy, lymph node dissection, partial adrenalectomy and thymectomy. Another emerging technique is fluorescence lifetime imaging (FLIM) [24] which produces

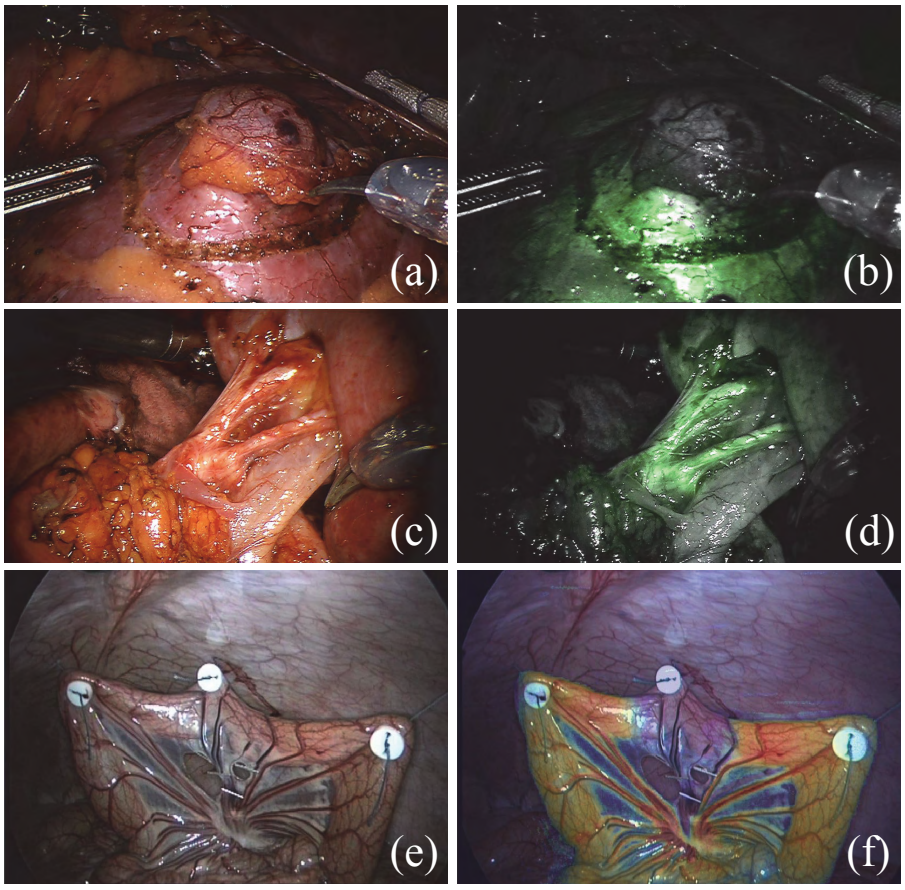


Figure 2.4.: Examples of using fluorescence for tumour and vessel identification during da Vinci® partial nephrectomy ©2017 Intuitive Surgical. Inc.: assessment of parenchymal perfusion during tumour excision (a-b); identification of vessels in renal hilum (c-d). An example of laparoscopic AR using fluorescence for identification of ischemic area in a bowel loop (e-f) ©2017 Elsevier[21].

an image based on the differences in the fluorescence decay of tissues. FLIM is able to provide intraoperative evaluation and the potential diagnosis of tumours in various procedures.

While conventional imaging modalities are essential for surgical oncology, early disease diagnosis requires micro-scale real-time imaging. CLE is an endoscopic imaging modality that allows for a real-time optical biopsy by providing on-site histologic information on a sub-cellular level. The tissue being examined is illuminated by a low-power laser beam and the fluorescence of light reflected from the tissue is detected through a small aperture. The aperture is placed to align with the conjugate focal plane of the tissue so that only the returning light that is emitted from a desired focal plane is detected. While CLE was initially used for gastrointestinal procedures, the applications of CLE have been rapidly expanding. In particular, CLE is being utilised to assist surgical diagnosis during MIS, including intra-biliary imaging [26, 27], intra-abdominal imaging [28, 29, 30], oral cavity [31] and larynx [32] imaging and cervical imaging [33], as well as neurosurgery imaging [34, 35, 36]. As shown in Fig. 2.5, CLE has been successfully applied to transoral robotic surgery for distinguishing between carcinomas and benign mucosa.

OCT is a diagnostic imaging technology that allows for the real-time, high-resolution, cross-sectional imaging of biological tissues *in vivo*. OCT imaging is analogous to ultrasound, but instead of using acoustical waves, OCT emits infrared light and measures the light backscattered from the patient's tissue. By combining the backscattered light and a reference signal, a high-resolution spatial image of the tissue microstructure can be produced. An example of OCT images and the corresponding histopathological images is shown in Fig. 2.6. OCT was initially developed for non-invasive ocular imaging as infrared light can penetrate different ocular media, which can be distinguished in OCT images [39]. Despite numerous applications in ophthalmology [40], OCT is also being used emerging in other surgical fields. OCT is utilised in laparoscopic surgery to acquire real-time diagnostic images

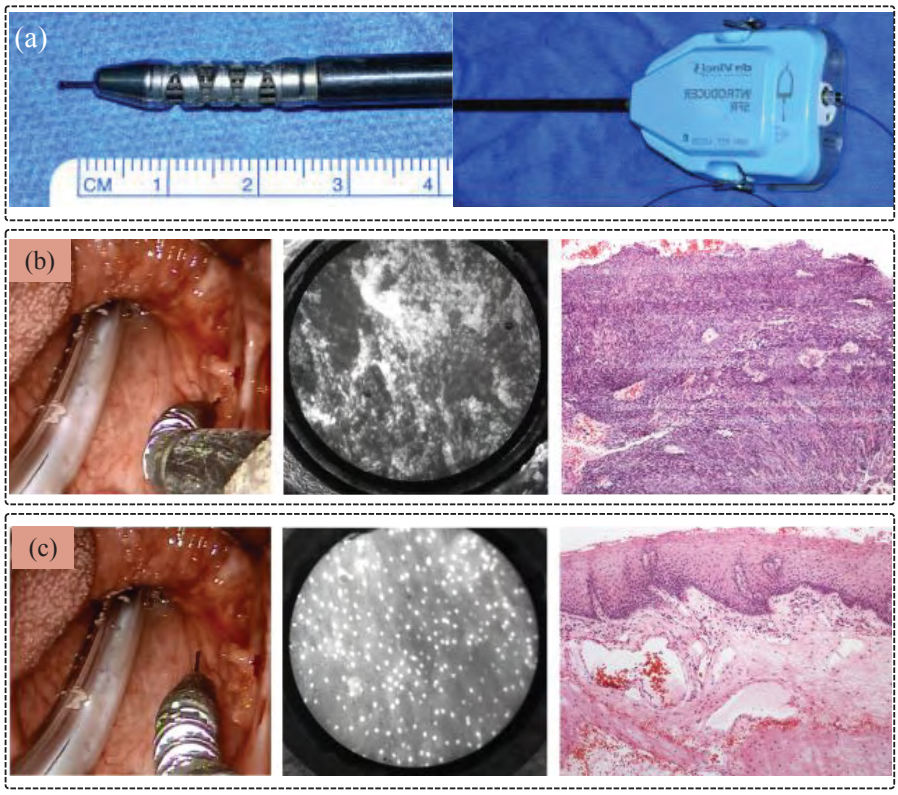


Figure 2.5.: An example application of the endomicroscope in a transoral robotic surgery ©2015 John Wiley & Sons[25]. (a) The endomicroscopy probe is inserted through a da Vinci® Introducer, (b,c) the endoscopic imaging of the oropharynx with corresponding microendoscopic images and hematoxylin-eosin stained histopathologic images from the oropharynx of squamous cell carcinoma (b) and benign mucosa (c).

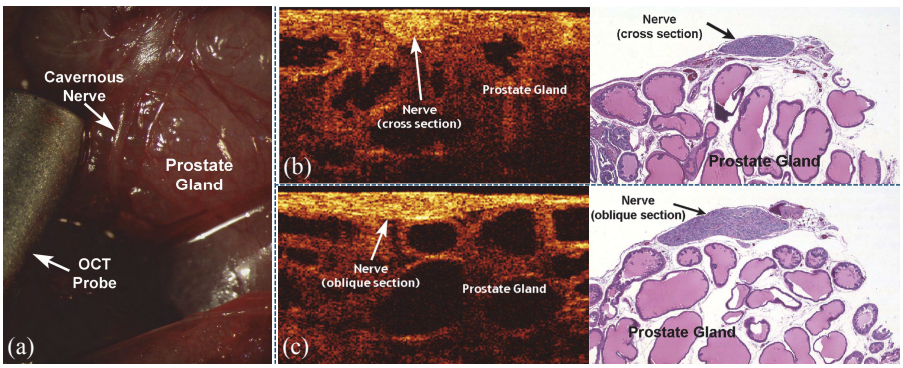


Figure 2.6.: (a) A view of *in vivo* OCT probe on prostate gland adjacent to cavernous nerve in a rat model, and OCT imaging and the corresponding histopathology of rat cavernous nerve (b) and prostate glandular tissue ©2008 Elsevier[37].

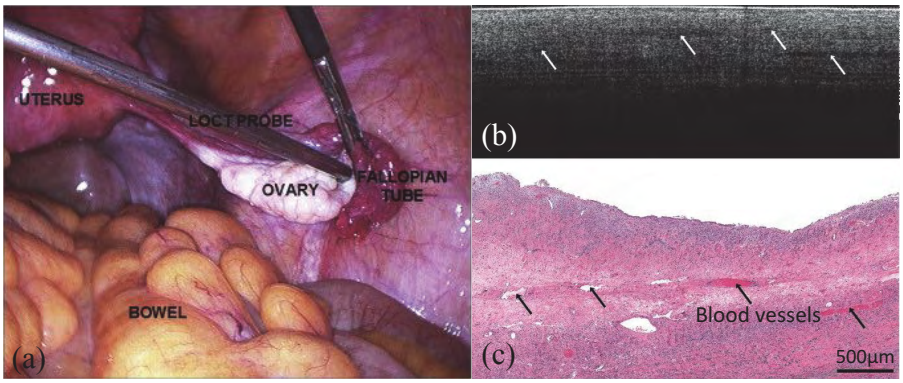


Figure 2.7.: (a) Laparoscopic OCT for human ovarian cancer diagnostics, (b) OCT image of ovarian stroma in endometriosis with increased vasculature and (c) corresponding histopathology ©2009 Elsevier[38].

for prostatectomy [41, 42], partial nephrectomy [43] and the identification of ovarian tissue [38], which is shown in Fig. 2.7. Furthermore, OCT is applied to other procedures including: spinal surgery [44], neural surgery [45], gastrointestinal surgery [46] and breast surgery [47], as well as intravascular surgery [48]. For a detailed introduction to OCT surgical applications, the reader can refer to a comprehensive review in [49].

2.2.2. Imaging Fusion and Visualisation

Intraoperative surgical planning and decision-making requires a comprehensive understanding of the positional and structural information of the targeted tumour and the surrounding critical anatomic structures such as nerves and blood vessels. While conventional intraoperative imaging modalities such as CT and MRI provide high-resolution images of the anatomy and the physiological processes of the body, emerging imaging techniques such as CLE and OCT assist in early cancer diagnosis by offering micro-scale information about the tissue. The fusion of the information or images produced by different imaging modalities can enhance the surgical capabilities of the surgeon. Using the advanced visualisation of the information from fused imaging, the surgeon can not only localise the tumour but also distinguish between normal and cancerous tissue rapidly.

Despite the extensive use of intraoperative ultrasounds, which facilitates tumour and vital anatomy localisation without radiation, many other intraoperative imaging modalities can be fused with ultrasound. FLIM can be combined with intravascular ultrasounds [24] to provide a concurrent visualisation of human coronary vessel composition, structure and pathology, which is shown in Fig. 2.8. A recent study has integrated the FLIM into the robotic surgery workflow for tissue characterisation, as shown in Fig. 2.10, by overlaying the FLIM scanning results on the endoscopic view. Furthermore, CLE can be fused with ultrasound by using a customised robotic device [50], as shown in

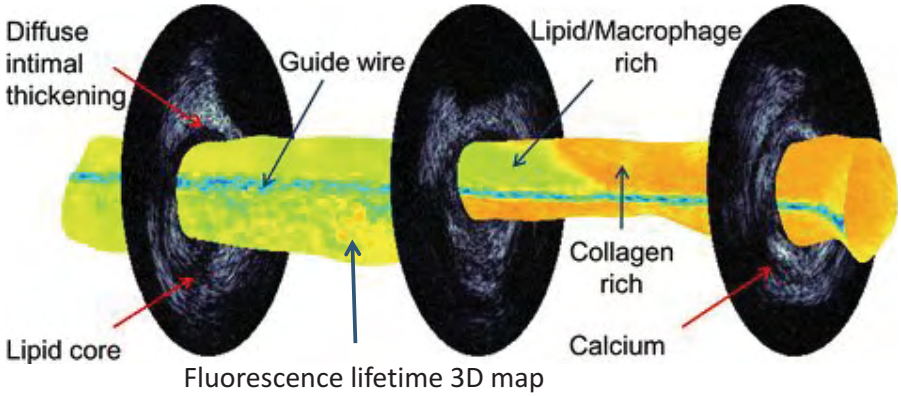


Figure 2.8.: 3D visualisation of intravascular ultrasound images and fluorescence lifetime images. The fluorescence lifetime values are mapped on to the 3D lumen surface. Figure by [24] used under CC BY/Adapted from original.

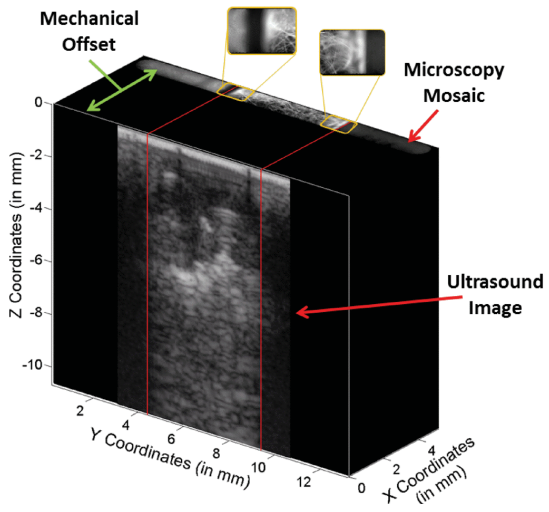


Figure 2.9.: Fusion of ultrasound and endomicroscopy ©2015 IEEE[50].

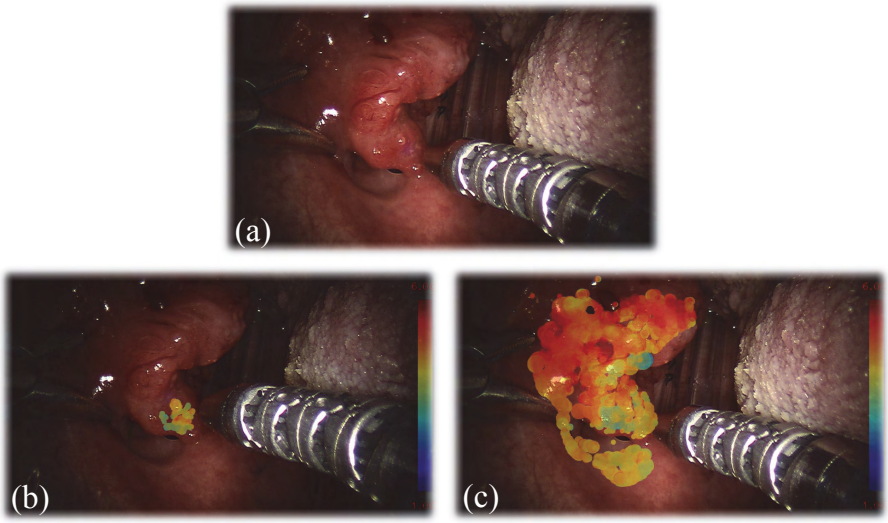


Figure 2.10.: Overlay of FLIM scanning results on the endoscopic view (a) in a da Vinci® robotic surgery (courtecy of Jakob Unger, UC Davis, USA), where a map of FLIM (b-c) is generated via scanning for tissue characterisation.

Fig. 2.9. The ultrasound imaging technique is enhanced by being fused with imaging modalities that provide complementary information.

Another imaging modality that is fused frequently is the fluorescence imaging. Recently, [51] proposed a fibre-based imaging platform that provides simultaneous FLIM and OCT images using a double-clad fibre. The result of the FLIM and OCT fusion is shown in Fig. 2.11. Additionally, the fluorescence imaging technique can be used to provide markers for intraoperative CT model registration and AR visualisation in laparoscopy [53]. This idea has been applied to the robotic surgery, as shown in Fig. 2.12. The fluorescent markers have been shown to be robust in the presence of smoke, blood and tissue.

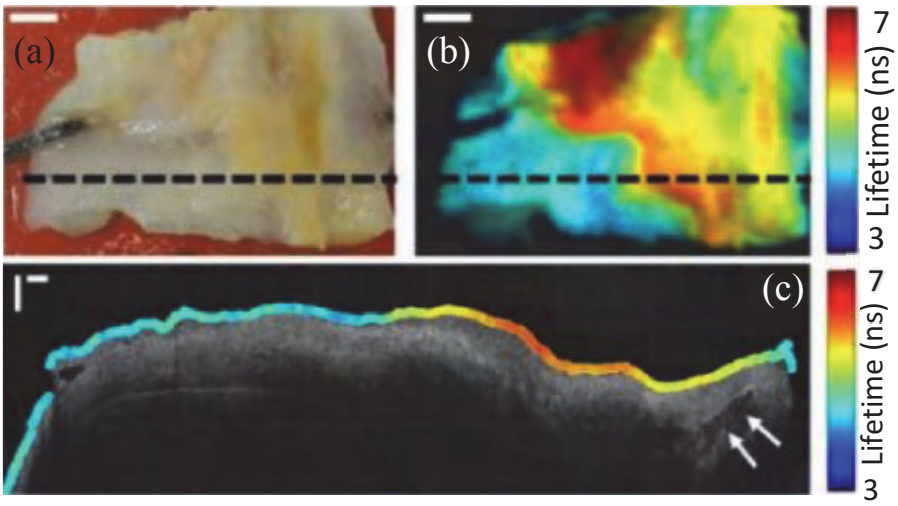


Figure 2.11.: A white light image (a) and the corresponding fluorescence lifetime image (b) of a coronary artery section. (c) Fusion of the fluorescence lifetime image with OCT image. Reprinted with permission from [51], Optical Society of America.

2.3. Tool Tracking and Navigation Techniques in MIS

The localisation and tracking of surgical instruments is an important factor in IGS where the relative position between the instruments and the targeted tissue is used for guidance and navigation. Numerous studies have been conducted on the tracking of surgical instruments, which is an essential component of an image-guided navigation system. The most widely-used and reliable tracking methods that are applied extensively use external devices such as electromagnetic (EM) and optical trackers. Despite the various applications of EM trackers for endovascular procedures [54, 55], they are also used in laparoscopic [56] and robotic [57] surgery. For a comprehensive review of EM trackers in medical applications, please refer to [58]. Regarding optical trackers, several applications have emerged for CAS [59, 60, 61]. In addition to the EM and optical trackers, an ultrasound-based tracking method

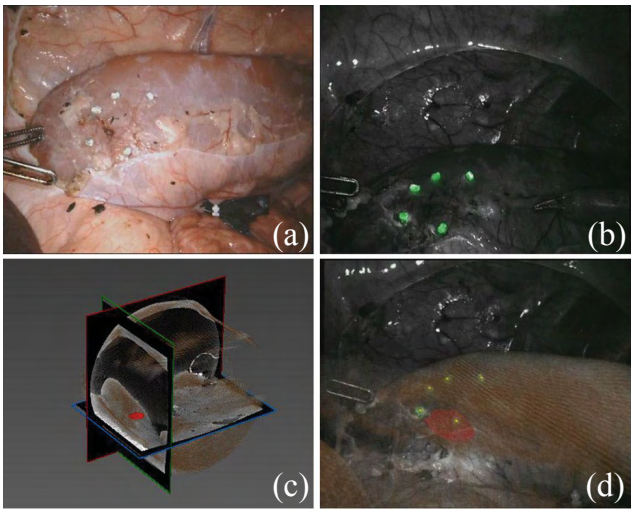


Figure 2.12.: An *in vivo* study of the intraoperative CT registration using fluorescent markers for robotic surgery. The white light image of the fluorescent markers attached to a porcine kidney (a) and the corrsponding fluorescence image recorded with the da Vinci® Firefly™ system (b). The 3D CT data (c) is registered to the endoscopic view using the fluorescent markers (d) ©2017 Springer[52].

has been proposed [62], but the low tracking frame rate restricts its practical application for IGS.

While external devices can be used for instrument tracking, vision-based tracking techniques do not require extensive hardware integration as they process endoscopic images directly. Several marker-based methods have been successfully developed and applied in clinical setups successfully by the use of fiducial markers consisting of chessboard vertices, [63], circular dots [64], hybrid features and specialised features [65].

Furthermore, purely image-based methods have been developed to eliminate the need for modification of the instrument design. Regarding the image-based methods which do not utilise a marker, a detailed review of the techniques can be found in [66]. To improve the robustness and accuracy of the tracking approaches, there are several studies that have attempted to combine different tracking devices by using a template model with kinematics [67, 68], and EM with an optical tracker [69]. Recently, deep learning techniques has been applied for recognition and pose estimation of surgical tools [70] where promising results are shown.

To fully exploit the potential and the advantage of imaging and tracking techniques in the context of IGS, AR is a useful visualisation technique for planning and guiding laparoscopic surgery. Laparoscopic AR has the potential to address the visual limitations of laparoscopy including the lack of depth perception, a small camera view and poor hand-eye coordination. As mentioned in [71], a classic AR system provides accurate AR in the endoscopic view by registering the patient's model created from the preoperative or intraoperative images with the current surgical scene. Furthermore, emerging techniques in laparoscopic AR, such as fluorescence-based AR and self-augmentation, are reliable in providing accurate visualisation of the biological information of tissue [21].

The implementation and integration of the above components for a complete image-guided navigation system can be a challenging task due

to the diverse hardware and software compatibility. To address this issue, several open-source development platforms for image-guided navigation systems are available. Regarding medical imaging processing, analysis and visualisation, there is the Visualisation Tool Kit (VTK) [72], the medical imaging interaction toolkit [73], as well as the slicer [74]. For the tracking of instruments and image-guided navigation, the image-guided surgery toolkit [75] and CustusX [76] may be the best options. These open-source platforms have greatly contributed to the rapid development of IGS.

2.4. Surgical Robotic Platforms

The surgical robotic platform is a comparatively new field of work in MIS which is also a recent trend in surgery. Over the past decade, there has been an increasing interest and clinical experience in the development of surgical robotic platforms. These platforms deliver potential advantages including an increased precision and flexibility of motion, an enhanced depth perception and the elimination of tremor, while also offering a comfortable surgical set-up by allowing the surgeon to operate the robot in a ergonomic position. The emerging robotic platforms are becoming more flexible, smaller incisions required and enhanced in terms of sensing ability. They also have more ergonomic interfaces, more of a compact footprint and enhanced portability, as well as being more intelligent and autonomous. As a result, surgical robots have take advantages over both human surgeons and machines to achieve better surgical outcomes. While there are many existing surgical robotic systems available for different types of surgeries, this chapter only focuses on those which are capable of laparoscopic surgery, which is the main theme of this thesis. In the following sections, the existing commercial systems are presented, followed by an overview of the current research on the development of robotic surgical techniques.



Figure 2.13.: The da Vinci® Xi surgical robot consists of (from left to right): surgeon console, vision cart and patient-side cart ©2017 Intuitive Surgical, Inc.

2.4.1. Commercial Systems

The most widely-used surgical robot on the market to date is the da Vinci® surgical system, developed by Intuitive Surgical (Sunnyvale, CA, USA). The da Vinci® robot has a typical master-slave setup which allows the surgeon to sit at a console to perform the procedures, by manipulating the surgical tools held by the robot at the patient's side. The most recent da Vinci® Xi system has improved the anatomic access compared to its predecessor, by providing easier docking, a larger motion range and an ability to attach the laparoscope to any arm, while reducing the arm size for improved portability. As the robot offers better visualisation and greater dexterity compared to conventional laparoscopic approaches, it has been adopted for various applications.

The development of alternatives to the da Vinci® robot is ongoing as the costs of the robot and its consumables are too high, preventing its promotion in developing countries. Recently, a surgical robot developed in China called Micro Hand S, has been successfully used for cholecystectomies and appendectomies, and preliminary results have

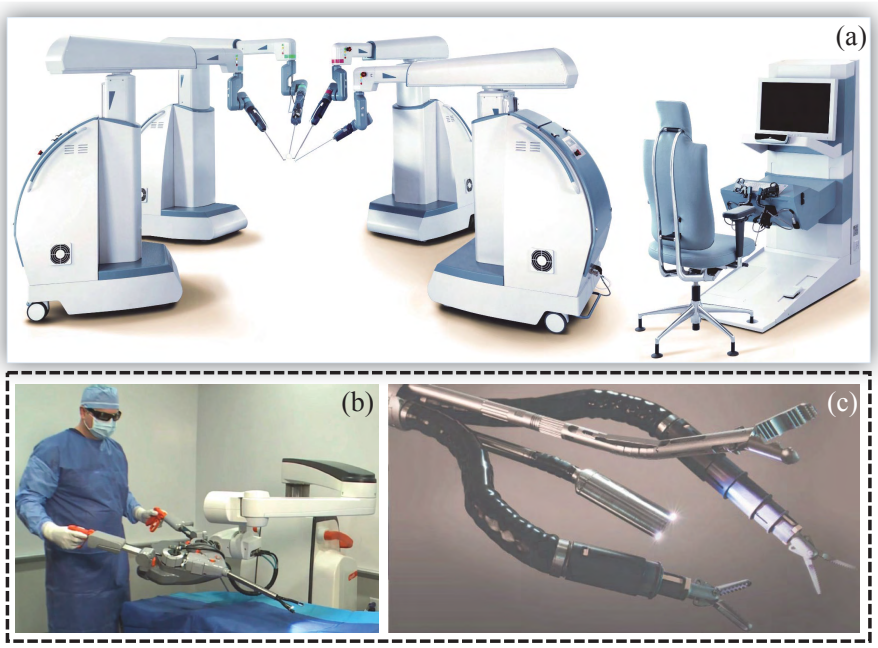


Figure 2.14.: Two surgical robots developed by TransEnterix: Senhance™ system (a) and SurgiBot™ system (b-c) ©2017 TransEnterix, Inc.

been reported in [77]. Similar to the da Vinci® robot, the Micro Hand S robot uses the master-slave setup with a console and slave arms. The main advantage of the Micro Hand S is the price, which is expected to be as low as one-tenth of the cost of the da Vinci® robot [77]. Recently, a Korean surgical system called 'Revo-i', which is similar to the da Vinci® robot, has completed a preclinical study for cholecystectomy [78]. However, substantial studies and clinical trials are required to confirm the safety and the efficacy of these systems.

Another commercial robot for laparoscopic surgery is the Senhance Surgical Robotic System (TransEnterix, NC, USA), which offers force sensing for haptic feedback, a 3D monitor for depth visualisation, an eye-tracking device for camera control and a set of reusable instruments. The system has been used in urology [79], gynaecology [80, 81] and thoracic procedures. Despite the functionalities provided by the



Figure 2.15.: The Flex® system includes an operational tower with a touch screen graphical user interface, a framework for stabilisation of compatible flexible instruments and the flexible endoscope. The tip of the flexible endoscope contains 6 LEDs for illumination, a high-definition camera and two working channels ©2015 John Wiley & Sons[83].

Senhance robot, its patient side arms are bulky which take up a lot of space, and it is difficult to move around. Another useful robot for single-port surgery, produced by the same company, is the 'SurgiBot' system [82] which offers reduced incisions, scalable motion and instrument clutching to allow the surgeon to adopt the most ergonomic pose. However, the system has recently been rejected for the FDA clearance, and there are no further updates on the system at the time of writing.

The most recent robotic surgical systems tend to have a reduced footprint and a more compact design, allowing for single-port access or even access through natural orifices without incisions being made into the skin. An example of this is the FDA-approved system Flex Robotic System (Medrobotics Corp., MA, USA), which has been used in transoral [84, 85] and colorectal [86] surgery. The advances in surgical robots enhance surgeons' capabilities by providing the access to complex and narrow anatomical locations with adequate flexibility and dexterity. Navigating such systems can be a challenging task for a surgeon so image guidance which fuses multiple modalities may offer improved results.

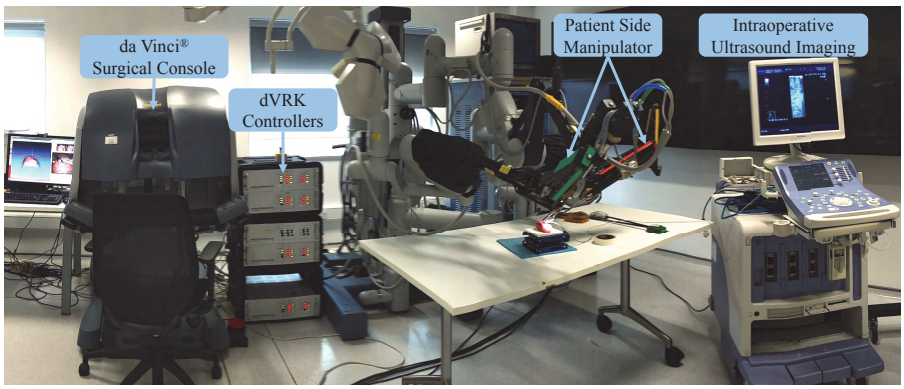


Figure 2.16.: An example experimental setup for the da Vinci® robot with dVRK controllers and an intraoperative ultrasound imaging machine.

2.4.2. Research Platforms

The development of robotic surgical techniques, including task automation, novel surgical instruments, surgical guidance and navigation, is an active research field where common research platforms can support and accelerate such developments by providing the essential hardware and software components. The most common platforms allow for the fast implementation of novel ideas and concepts, as well as rapid instrument prototyping without effort needed to build a surgical robot. Furthermore, common hardware and shared software frameworks can facilitate collaboration between research groups. There are several research platforms available including the dVRK, the RAVEN surgical robot, the DLR MicroSurge robotic system, the Medineering Positioning Arm (Medineering, Munich, Germany) and the KUKA LBR Med robot, as well as other customisable surgical robotic platforms.

The dVRK is a set of controllers with firmware and software for robot communication, developed by the Johns Hopkins University in collaboration with Intuitive Surgical. Currently, dVRK supports part of the first-generation da Vinci® robotic system, which includes master tool manipulators, Patient Side Manipulators (PSMs) and endoscopic camera manipulators. A Robot Operating System (ROS) interface for

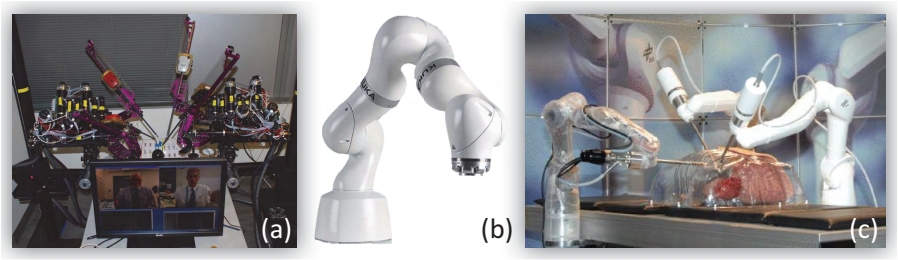


Figure 2.17.: Surgical robotic platform for research. (a) RAVEN-II Surgical Robot ©2013 IEEE [97]; (b) KUKA LBR Med robot ©KUKA Aktiengesellschaft; (c) DLR MiroSurge surgical robot ©2009 IEEE [98].

joint space and Cartesian control is provided, alongside a teleoperation program. The dVRK platform has attracted a lot of interest, and it has facilitated the research of novel surgical instruments [87, 88], human-robot collaboration, surgical task automation [89, 90, 91] and multiple sensors fusion [87, 91]. Similar to the dVRK, RAVEN is also an open surgical robotic platform consisting of two cable-driven laparoscopic arms which are teleoperated with two Geomagic Touch haptic devices (Geomagic, NC, USA). Several studies have also been conducted using the RAVEN platform, including the cooperative control of surgical tasks [92, 93], the automation of surgical debridement [94] and tumour ablation [95] tasks, and surgical robot training [96].

The MiroSurge system developed by the German Aerospace Centre (DLR) is another research platform for laparoscopic surgical applications. Each MiroSurge system [99] has three to five MIRO robot arms at the patient's side, and two actuated haptic master manipulators, Sigma 7 (Force Dimension, Switzerland). The MIRO robot is a versatile and lightweight manipulator that can integrate specialised instruments and multimodal sensors. The MIRO arm can be coupled with a MICA instrument [100], which enables extra dexterity and force-sensing capability and is customised for specific surgical tasks. There are other robotic arms that have been built for medical applications such as the the KUKA LBR Med robot (KUKA, Germany)

and the Medineering positioning arm (Medineering, Germany). They have been specifically designed for medical scenarios by considering the safety issues that can occur during the execution of tasks.

2.5. Robotic Assistance for Surgical Task Automation

2.5.1. Classification of Autonomy in Robotic Surgery

The scope of autonomy in surgery depends on the type and degree of automation for all computerised technical systems that are used in surgery. These systems include imaging acquisition and processing modalities, surgical action planning and decision-making software, and robotic arms for the execution of the surgical action. Existing studies have proposed several classification methods for autonomy in surgery. Wolf et el. [101] classify robotic surgical devices into three groups: passive, semi-active and active. In this definition, the passive devices are the surgical navigation systems that are used for the position tracking of relevant devices and the patient's anatomy, such as the optical tracking device OptoTrak (Northern Digital Inc., Canada). The semi-active devices aim to guide or increase the surgeon's control and accuracy in the use of the operating instruments. The scaling and tremor filtration function provided by the da Vinci® surgical system can be seen as a semi-active feature. With regards to active robotic systems, the device actively operates on a patient's anatomy based on a predefined trajectory. The first surgical robotic system that fell into this category was the Robodoc system (Integrated Surgical Systems, CA), which was used to mill predefined bone areas for total knee arthroplasty. This is only a broad classification in which the robotic systems are considered to be mechanical devices with pre-programmed functions.

A more comprehensive classification is proposed by [102] in which surgeons and automated systems are seen as two agents in the surgical process. The degree of the automation of a given system is determined by how many surgical functions are actively supported by the system. Manzey et al. [102] divides surgical functions into three groups: (a) acquisition and analysis of information; (b) decision-making and plans for action; (c) execution of the surgical action. The surgeon or the automated system can play either active or passive roles depending on their involvement in the particular function group. The autonomy levels are ordered from '0' to '5' where '0' represents no automation and '5' represents high level of autonomy.

In the absence of automation, robotic systems are used in a master-slave mode where the operation performed by the surgeon on the master, is duplicated exactly by the slave. The motion scaling and filtration feature in the da Vinci® system is not considered to be automation. In the case of a low degree of automation, a system is required to play an active role in the acquisition and analysis of information. For example, a surgical navigation system that provides support for intraoperative data acquisition and processes may leave the decision-making and task execution to the surgeon. As the degree of automation increases, the autonomous system becomes move active in the surgical functions (such as decision-making and the execution of surgical actions). With regards to high-degree automation, the surgeon takes a passive role in all three surgical function groups, although the surgeon is still required to monitor the automated functions and intervene if the automation fails. This classification method for automated surgical systems not only includes the surgical planning and action execution functions involved in CAS but also considers monitoring and image analysis as part of automation.

To accommodate the increasing level of autonomy in surgical robotics, [94] uses the term "supervised autonomy" to define a type of system that is capable of automating surgical sub-tasks under the surgeon's supervision. Recently, Yip et al. [103] has adopted this term in a pa-

per on their proposed classification method for robotic surgery. The automation levels defined in the paper are: direct control, shared control, supervised autonomy, and full autonomy. Direct control means that there is no automation in the system and surgeons are in complete control of surgical tasks by teleoperating the robot. In shared control, the motion of the surgical tool is controlled by the surgeon and the robot simultaneously. An example of this is the active constraints algorithms(also known as the virtual fixture) [5] which are implemented in surgical robots to prevent the surgeon from moving outside a predefined safe region. As the autonomy level increases, the robot executes surgical tasks with increasing degrees of automation under the surgeon's supervision, which is supervised autonomy. In this case, the surgeon observes the operation and is responsible for the high-level decision-making, while the robot executes the motions. Regarding full autonomy, the robotic system completely replaces the surgeon during the operation.

In addition to the above classification, Yang et al. [104] propose a framework for medical robotics with six levels of autonomy. The automation levels are ordered from '0' to '5' which are, respectively, no autonomy, robot assistance, task autonomy, conditional autonomy, high autonomy and full autonomy. By following this framework, Yip et al. [103] classify the levels of autonomy similarly, although in this work supervised autonomy was not differentiated into task autonomy and conditional autonomy. Task autonomy refers to the automation of specific tasks initiated by the surgeon. In the case of conditional autonomy, multiple task strategies are available for the surgeon to select based on the current condition of the operation. Both task and conditional autonomy are essential components of high and full autonomy. Here, full autonomy means that no human surgeon is needed while the previous classifications still require a surgeon to oversee the operation.

To provide an overview of the above classification methods for autonomy in surgery, Fig. 2.18 depicts the autonomy levels as a spectrum in which similar colours represent similar autonomy levels. In the figure,

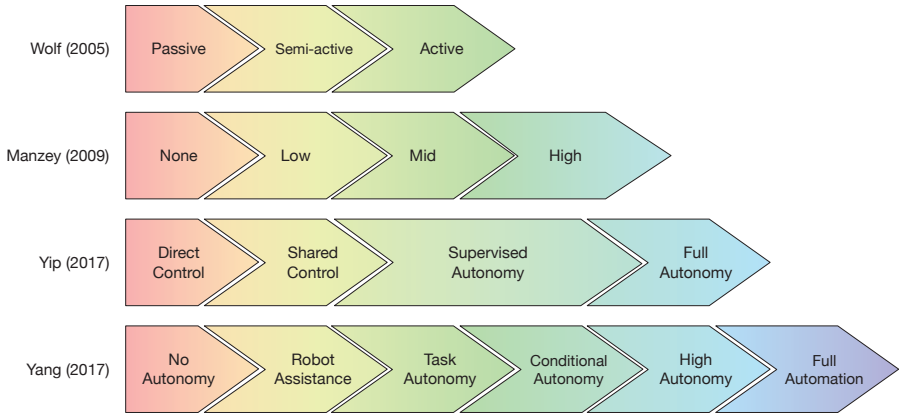


Figure 2.18.: A diagram showing autonomy level of surgical robot defined in literatures.

all of the classification methods start from the left with no automation and to the right with increasing levels of autonomy. As the developments in the automation of surgical tasks emerges, a clear pathways from the current assistance-type automation to high-level and even full automation will be available.

2.5.2. Benefits of Surgical Task Automation

As the level of autonomy increases, there are many advantages in automating surgery. A major advantage of a robotic system is that it provides high precision and accuracy while performing repetitive tasks tirelessly and stably. Given that robotic systems have computational and mechanical advantages while a human surgeon is adept at high-level contextual decision-making, it is important to take advantage of both robots and humans. By carrying out tedious and cognitively-demanding tasks, the robotic system reduces the surgeon's fatigue and allows the surgeon to focus on the clinical aspects of a procedure, such as the decision-making, surgical workflow and risk management [105]. Many studies [106] show that collaboration between human surgeons and a robotic system can lead to improved accuracy and shorter task-completion times. A classic example is the active

constraint application, where the robot restricts the paths or positions of instruments so that damages to critical anatomical structures (such as nerves and blood vessels) can be avoided. Furthermore, robotic systems provide increased manipulability and dexterity compared to laparoscopic surgery by using wristed instruments.

Robotic systems also have several advantages in terms of utilising medical imaging such as CT, MRI and ultrasound scans. Anatomical models built by preoperative imaging modality can be easily imported to the robotic software for surgical planning and navigation. Furthermore, intraoperative imaging modalities can be seen as robotic sensors which adapts to the changing environment based on the imaging information, as well as other sensors and commands. When automating surgical tasks, the data gathered from different sources needs to be used effectively in order to take the optimal action based on the current status of the procedure. Another advantage of a robotic system in IGS is that the surgeon's exposure to the radiation from the X-ray fluoroscopy exposed to surgeons is significantly reduced as the surgery can be performed in a different room.

Another benefit of automating surgery using a robotic system is to provide consistent and standardised treatments, contributing to better efficiency and surgical outcomes. Despite the target anatomy being different from patient to patient, the action of a particular task should be similar for each patient. An example of this is that the quality of a wound closure depends on the surgeon's skill of suturing, whereas automating the task provides consistent results regardless of the surgeon's skill level.

In addition to the advantages in terms of surgical outcomes and ergonomic considerations, automation in surgery is also beneficial for hospital administration as it shortens hospital stays. Furthermore, the skilled surgical staff that would be required for low-level tasks, such as holding the camera, can be redistributed to other areas of the hospital.

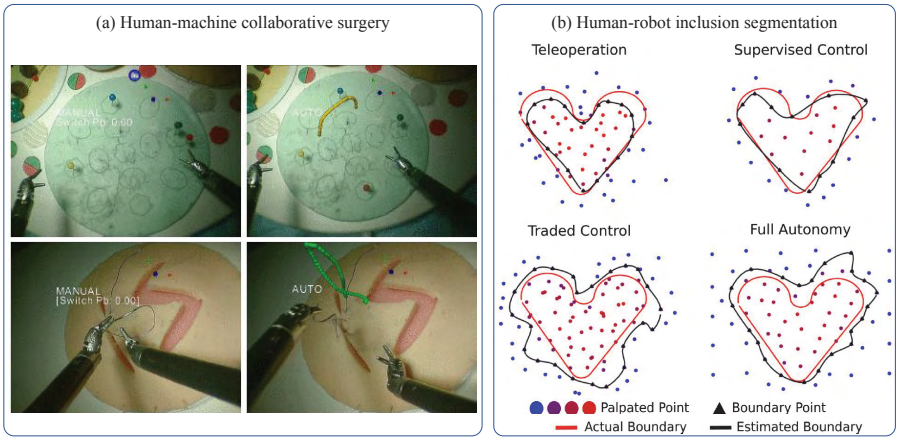


Figure 2.19.: (a) Human-machine collaborative framework for surgical pin and suture task ©2011 IEEE[107]. (b) Human-robot collaboration on inclusion segmentation experiment with different control modes ©2016 IEEE [106].

2.5.3. Developments of Surgical Task Automation

Collaborative Framework for Surgical Task

The strengths of both the human surgeon and the robot can be fully exploited in robotic surgery, by utilising a collaborative framework. In such collaborative framework, surgical tasks can be partially performed by the robot autonomously with surgeon's control or supervision and the rest by the surgeon manually. Previous studies have used the Learning from Demonstration (LfD) technique [107, 92] to teach the robot to recognise and perform surgical tasks. As shown in Fig. 2.19(a), a task can be divided into several subtasks where the workload can be distributed between a human operator and a robot. A recent [106] evaluates the performance of different human-robot collaboration models, which can be shown in Fig. 2.19(b). The results indicate that the robot can identifies boundaries with higher sensitivity and less variance than human operators, though the human operators can complete the task faster.

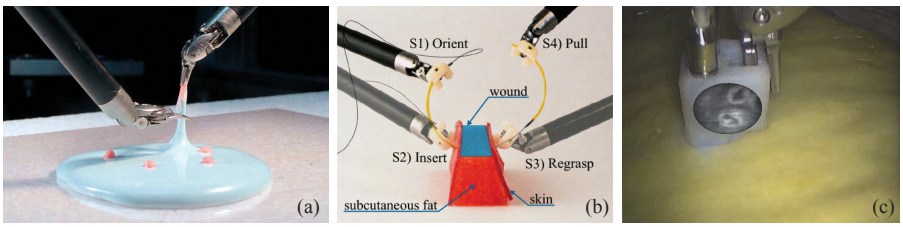


Figure 2.20.: Automated surgical task using dVRK. (a) Soft tissue debridement task ©2015 IEEE [89]; (b) Suturing task ©2016 IEEE [90]; (c) endomicroscopic scanning task ©2017 IEEE [109].

Automation Framework for Surgical Task

Several studies have increased the levels of autonomy in robotic surgery by proposing frameworks for surgical task automation [94, 89, 95, 90, 91]. Various methods have been utilised to automate the surgical tasks. The methods based on visual servoing use imaging feedback of the laparoscope to control the robot [94, 91]. Other methods utilise learning based method by using the learning from demonstration technique [89]. Recently, the deep reinforcement learning technique has been applied on dVRK for multilateral surgical pattern cutting [108]. As shown in Fig. 2.20, different types of surgical tasks can be automated by using the dVRK platform.

Motion Compensation

Compensation of physiological motion is an active research topic in the context of robotic surgery. The physiological motion induced by respiration or heartbeat can be very disturbing for the surgeon during surgical procedures that require high accuracy. To overcome this problem, numerous robotic instruments or devices have been developed for the surgical operations that are affected by the physiological motion significantly. Research groups have tried to mitigate this issue in cardiac surgery [110, 111, 112], by developing a motorised instrument with ultrasound guidance [110], presenting a predictive control

algorithm based on biological signals like the electrocardiogram [111], and applying a predictive control method on a surgical robot [112]. Although most of the motion compensation methods or systems use cardiac surgery as their exemplar application, the methods can be applied to other procedures that contain periodic respiratory motion.

Despite the fact that external sensors have been successfully used to capture the physiological motion for the compensation task, visual tracking methods based on endoscopic images have been developed for tissue deformation recovery and tracking [113, 114, 115]. These approaches [114, 115] use feature-based methods to extract and track salient features on the target organ surface such as blood vessels, fat tissue or other distinguishable structures for estimating the organ motion. The main benefit of the feature-based methods is the low computational requirement so that real-time performance can be achieved. However, their accuracy highly depends on the availability of the salient features in the surgical scene. To address this issue, previous study [113] has proposed to use the Thin-Plate Splines for modelling the tissue deformations. The method [113] allows tracking regions with poor texture by choosing a set of flexible control points. Furthermore, the feature-based tracking methods have been extended by considering robust online learning and classification schemes [116].

2.5.4. Challenges of Surgical Task Automation

The tissue motion of the patient caused by respiration and their beating heart makes the surgical scene a challenging environment. Also, the tissue and organ deformation caused by pneumoperitoneum or patient movement may invalidate preoperative plans that are created based on the patient model reconstructed from preoperative imaging, such as CT and MRI scans. Consideration of such a dynamic environment when developing an automated system can potentially provide a more reliable and robust system for real surgical scenarios. Existing studies [117] have discussed facilitating minimal intraoperative movement by

fixing the patient with pins or frames. However, this is very difficult to accomplish in laparoscopic surgery due to the nature of the soft tissue involved. Another approach is to track the changing environment so that the robot can adapt itself to the movement. Therefore, imaging techniques can be used to estimate the tissue motion intraoperatively to provide guidance and background information for the automation of surgical tasks in a dynamic environment. Furthermore, tissue manipulation such as ablation, cutting, dissection and suturing could lead to topological tissue changes requiring an update of the patient model. Other visual occlusions in surgery such as instruments, bleeding and electrocautery smoke may also result in difficulties for systems that rely on the direct vision of endoscopic cameras. Hence, multiple imaging modalities may solve this problem by using intraoperative ultrasound, CT and MRI scans for surgical navigation. Nevertheless, automation of surgical tasks requires real-time imaging performance, which is still a challenging problem for some modalities.

While routine and repetitive surgical tasks are relatively easy to automate, it is usually difficult to identify such well-defined tasks. For example, there is no consensus among surgeons on the optimal management of positive surgical margins during partial nephrectomy [118]. Therefore, it would be a useful feature if autonomous systems consider task-specific specifications or constraints when executing of routine surgical tasks.

In addition to the technical challenges, which may be solved as the technology improves, ethical and legal issues have become a major concern when implementing surgical task automation in clinical practice. As the robotic instruments and devices have an increased level of automation, the related regulatory, ethical and legal issues require careful consideration [104]. For example, it is important to understand how to prevent the surgeons losing their skills or developing a possible over-reliance on an automated systems which have greater levels of automation. Furthermore, it is not clear who should take responsibility among the human surgeons, the hospital, the device manufacturer and

the software developers is the usage of an automated system results in surgical errors that harm the patient. These questions have to be answered before an automated system can be used in the operating room.

2.6. Conclusion

Robotic techniques have revolutionised the field of MIS over the last two decades. Robotic devices have extended the capability of surgeons by providing higher positioning accuracy, improved dexterity and steady tool movement. The field has also benefited from medical imaging techniques through the use of the imaging modalities in the different stages of the surgery for diagnosis, planning, navigation and result confirmation. The automation of specific surgical tasks has become possible by combining the imaging and robotic techniques.

The automation of tedious and repetitive tasks or tasks that require high precision can potentially benefit RMIS. Open surgical robot platforms have facilitated the development and the research of automated systems by providing common hardware and shared software frameworks. The most common approach to realising automation is visual servoing, which is a method that utilises imaging information for task planning, navigation and execution.

There are many challenges to overcome when building an automated system. The technical difficulties include a changing environment, visual occlusions and the standardisation of procedures. Furthermore, the ethical and legal issues must be addressed properly when introducing an automated system from a laboratory setup to the operating room. The current research of automated surgical systems mainly focuses on task-specific automation and human-robot cooperation, which is feasible, safer and is more likely to be accepted by surgeons. Fully automated surgery is not desirable or acceptable until all of the technological and ethical issues are addressed.

3. Marker based Surgical Instrument Tracking †

3.1. Introduction

To identify the anatomical structures such as vessels and tumours during MIS, intraoperative ultrasound has been used [9]. However, a well-known drawback of using intraoperative ultrasound is that surgeons need to interpret the ultrasound images and perceptually relate these images to the operative scene captured by a laparoscope [119]. This is particularly challenging, as the ultrasound images and laparoscopic videos are displayed separately on different screens. To reduce the mental workload of a surgeon, a previous study presented in [9] has introduced an intuitive visualisation method, which is achieved by registering the 2D ultrasound images into the surgical scene. This approach enables surgeons to observe details from both ultrasound and laparoscope in a single display. To provide consistent image overlays, the ultrasound probe needs to be tracked during operation.

Electromagnetic [120], optical [121] and hybrid [69] approaches have been used to track the surgical instruments with reasonable accuracy. Nevertheless, a direct use of laparoscopic images for tracking instrument in MIS is more practical than other methods where extra tracking devices [69] are used. The introduction of additional tracking devices

†Content from this chapter was published as:
Real-Time Surgical Tool Tracking and Pose Estimation using a Hybrid Cylindrical Marker. Zhang, L., Ye, M., Chan, C., Yang, G.-Z.. International Journal of Computer Assisted Radiology and Surgery. 12(6), 921-930. (2017)

Table 3.1.: Comparison of visual marker based tracking methods for intraoperative ultrasound probe

Method	Probe shape	Pattern type	Robotic/Freehand	Real-time
[124]	Planar	Chessboard	Freehand	Yes
[125]	Planar	Chessboard	Robotic	No
[64]	Planar	Circular-dot	Robotic	Yes
[63]	Cylindrical	Chessboard	Freehand	No
Ours	Cylindrical	Hybrid	Both	Yes

into the theatre would not only occupy the valuable space in the operating field, but also suffer from intrinsic limitations of these tracking devices such as ferromagnetic interference and line-of-sight issues. Previous methods have extracted 2D locations of a planar [64] or cylindrical [63] marker that is rigidly attached to the probe. In Table 3.1, a list of marker-based tracking methods for intraoperative ultrasound probes is presented. Among these methods, pose estimation is treated as a Perspective-n-Points (PnP) problem [122]. However, pose estimation of a planar marker may provide two ambiguous solutions where the incorrect one should be eliminated [123].

In this chapter, a new marker design is proposed for cylindrical surgical tools. The marker consists of three patterns of circular dots and lines of chessboard vertices, as shown in Fig. 3.1. The circular-dot pattern is used to estimate the pose of the tool while the chessboard vertices are used to eliminate the ambiguity in pose estimation. To further improve detection rates, a tracking component considering temporal information is employed to cater for failures in marker detection. The proposed tracking framework is evaluated with various surgical tools including ultrasound probes and robotic instruments. Detailed validation is provided and comparison results demonstrate our framework outperforms other alternative approaches, both in detection rates, work space coverage and pose estimation accuracies.

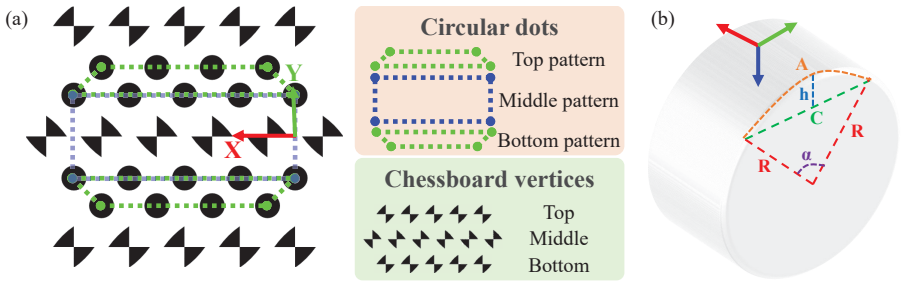


Figure 3.1.: (a) A top view of the proposed marker for cylindrical object tracking. The hybrid marker consists of circular-dot patterns and chessboard vertices. (b) A side view of a cylinder showing related parameters to define a marker's local coordinate frame.

3.2. Design of Hybrid Marker

This chapter proposes a novel hybrid marker including both circular dots and chessboard vertices for a cylindrical object as shown in Fig. 3.1(a). The circular dots can be labelled into three groups denoted as top, middle and bottom patterns according to their relative location. Each pattern has two lines of circular dots which are used for pose estimation. Top and bottom patterns are vertically asymmetric and the middle pattern is symmetric such that adjacent patterns can be easily differentiated. In addition, we placed three lines of chessboard vertices between the circular-dot patterns. The middle vertices have a 90 degrees orientation shift relative to the top/bottom vertices. These vertices are used to remove ambiguous pose estimation in section 3.6. The marker can be horizontally placed on a cylindrical object such that lines on the pattern are parallel to the axial axis of a cylindrical tool. We define a local coordinate frame of a cylinder, as shown

in Fig. 3.1(b), where several parameters can be defined:

$$\begin{aligned} A &= \alpha R \\ C &= 2R \sin \frac{\alpha}{2} \\ h &= R(1 - \cos \frac{\alpha}{2}). \end{aligned} \tag{3.1}$$

Here A is the arc length when the angle and radius is α and R, respectively. C is the chord length and h is the corresponding segment height.

Given a marker position $p = [x, y]^T$ located in a 2D coordinate frame as indicated in Fig. 3.1(a), its corresponding 3D position $P = [X, Y, Z]^T$ in the local coordinate frame of the cylinder can be defined by using equation 3.1:

$$\begin{aligned} X &= x \\ Y &= R \sin \frac{y}{R} \\ Z &= \begin{cases} R - \sqrt{R^2 - Y^2} & \text{if } y \leq \frac{\pi R}{2} \\ R + \sqrt{R^2 - Y^2} & \text{if } y > \frac{\pi R}{2} \end{cases}. \end{aligned} \tag{3.2}$$

These points will be used as model points for the marker pose estimation in section 3.6.

3.3. Efficient Detection of Hybrid Features

A workflow of the detection algorithm for the proposed marker is shown in Fig. 3.2. At each iteration, a grayscale image is obtained from a colour image and used for circular-dot detection. The grayscale image is then used to generate multiples binary images, as shown in Fig. 3.2, based on thresholding. The use of multiple binary images is effective for

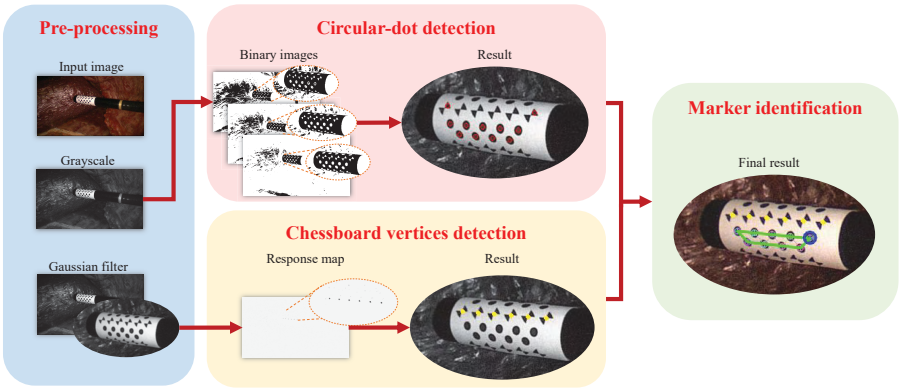


Figure 3.2.: Algorithm workflow of the marker detection algorithm. The binary images are generated using threshold values between 70 and 100 with an interval of 10. The detected dots are circled in red, and the chessboard vertices are labelled in yellow. The first dot in the pattern is circled in blue.

discarding blobs that are false positives, as they would not consistently appear when varying the threshold values. The threshold values should be determined according to the lighting condition of the scene. In our experiments, we used a set of thresholds (70-100) with an interval of 10, which is a good trade-off between speed and accuracies. To locate circular dots, we apply contour tracing on the binary images to extract circular blobs. In this chapter, the method proposed in [126] is utlised in our implementation* to extract the circular blobs.

Several criteria were used to filter spurious blobs which do not belong to the marker. A blob is removed if its area is not in line with a desired value which can be defined based on the possible minimal and maximal size of a blob in the image. In addition, the convexity of a blob is also used because the perspective projection of a circular-dot on an image results in a convex shape, whereby non-convex blobs can be removed. We define the convexity as the ratio of the blob's area to the area of the blob's convex hull. Another adopted criterion for removing polygonal

*an implementation of OpenCV is adopted [127]

blobs is the circularity which is defined as:

$$O = \frac{4 \cdot \pi \cdot E}{P^2}, \tag{3.3}$$

where E and P are the area and perimeter of the blob, respectively. Lastly, a blob belongs to the marker should appear at least N times in the binary images, where N is called repeatability. The best numbers for area, convexity, circularity, repeatability are 50-2000, 0.85, 0.75 and 2.

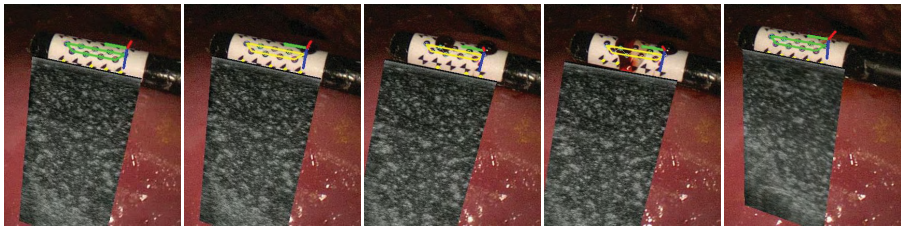
The chessboard vertices are detected using an efficient detector [128] designed for chessboard corner feature. As shown in Fig. 3.2, a Gaussian filter is applied to the grayscale image to remove speckles and noises that might degrade the performance of the chessboard detector. For each pixel in the filtered image, a ring of 16 pixels around the pixel are sampled at a constant radius with equal angular spacing. A response map is calculated using the sampled pixels, which is shown in Fig. 3.2, and a chessboard feature belongs to the marker if its value in the response map is greater than a threshold value. We follow the method in [128] to calculate the response map. Specifically, an overall response for each pixel equals to a sum response subtracting a difference response and mean response. The sum response calculates the summation of opposite samples. The difference response is defined as the difference of opposite samples. The mean difference is an absolute difference of a local intensity mean and a larger spatial mean. To further remove spurious features, which tend to have low responses, non-maximum suppression [129] is applied to retrieve the location of peak responses. In addition to offering the position information, the chessboard detector also provides the orientation of a vertex in the image coordinate using 1 of 8 labels. Given the orientation information, a spurious feature can be removed as its orientation label will not in accord with the desired one which will be described in section 3.4.

Furthermore, triangle forming the chessboard vertices may be detected as a dot. Therefore, the detected chessboard features are also

used to reject the false detection of a circular dot.

3.4. Identification of the Hybrid Marker

To estimate the marker pose, the correspondence information of the detected dots and the model points is required. We have used the following cues. Firstly, dots that are close to each other are assigned to the same cluster. The cluster that has the largest number of dots is identified as the marker. Secondly, four endpoints locating at corners are found by using the convex hull of the cluster and the corners should have sharp angles. The endpoints are sorted based on the fact that two long edges are parallel, and then the model circular-dot pattern can be transformed to the image using correspondences of the sorted endpoints. Thirdly, the identity of each detected dot is found by using the closest point in the transformed model pattern. Since the top and bottom patterns are the same but vertically flipped, if an asymmetric pattern is detected, we need to decide which pattern it belongs to. Typically, surgical instrument movements are constrained by the incision point, so it is assumed that the pattern cannot be flipped horizontally. Since the circular dots in the detected pattern are identified, if the first dot of the pattern (as depicted in Fig. 3.2) is on the left side of the last dot, the pattern is a top pattern, otherwise, it is a bottom pattern. Finally, the orientation of the symmetric (middle) pattern can be determined similarly by looking at the relative position of the dots. By having the labeled circular-dot pattern, the chessboard vertices can be identified as well. The relative orientation between the chessboard vertices and the circular-dot pattern is fixed, therefore the desired orientation of the vertices can be deduced from the orientation of the circular-dot pattern.



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

Figure 3.3.: A sequence of images showing how the tracking algorithm reacts when the marker is partially occluded by fake blood. The green outline of the marker indicates successful detection while yellow outline means detection failure and tracking component is used.

3.5. Marker Tracking using Optical Flow and Homography

In practice, occlusion can occur due to rotational motion of the probe as well as blood stains on the marker, all of which would result in missing dots during marker detection. To address this problem, we compute a homography $\mathtt{H}$, which is a 3×3 matrix with 8 Degree-of-Freedom (DoF), representing perspective transformation between two planes. The circular-dot pattern is defined in a reference coordinate frame where position of the dots is denoted as p_r. During tracking, the projection of the circular dots on the camera image plane is denoted p_m. The homography links them via: $p_m = \mathtt{H}p_r$.

If the detection component of the framework is not able to detect the marker, a sparse optical flow method [130] is applied, considering the temporal information, to find the current position of visible dots. The dots are used to estimate the homography $\mathtt{H}$. This homography can be estimated using only four pairs of non-collinear points, thus robust to occlusion. The position of a missing dot can be calculated from its correspondence in the reference coordinate frame by using the homography.

An illustration of how the tracking component deals with the oc-

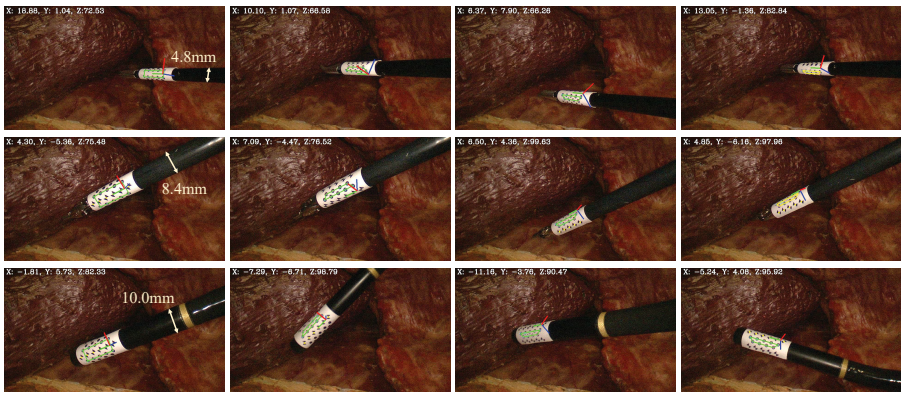


Figure 3.4.: Qualitative results of the tracking framework for the proposed marker on various types of surgical tool. A 5mm monopolar cautery (top row), a da Vinci® large needle driver (middle row) and a laparoscopic ultrasound probe (bottom row) is tracked, respectively. The coordinate axis in red, green and blue shows that the pose has been correctly estimated.

clusion is shown in Fig. 3.3. The marker is detected correctly at the beginning. As blood is stained onto the probe, which is shown in Fig. 3.3(b)-(d), the detection method cannot extract the marker's location and the tracking method is used. The tracking method, on the other hand, is able to locate the marker correctly until a major area of the pattern has been occluded, as shown in Fig. 3.3(d). In this case, the tracking results start drifting which leads to wrong pose estimation. In Fig. 3.3(e), detection of the marker is recovered when the blood stain is cleaned. For self-occlusion, as shown in Fig. 3.4, the algorithm can deal with large rotational motion where the marker is partially visible.

3.6. Pose Estimation and Ambiguity Elimination

Given the model points defined in the marker's local coordinate frame and their corresponding tracked projections on an image of a calibrated

camera, the marker pose can be estimated via PnP methods. However, a planar model has the rotation ambiguity which corresponds to a reflection of the points about the plane whose normal points along a line-of-sight that pass through the points [123]. The pose estimation problem becomes ambiguous when the projection of the model is close to affine, which typically occurs if the pattern is either small or far from the camera. We employed [123], which is particularly useful as it provides two solutions that can correctly align the projections using the estimated pose with the detected correspondences. One of the solutions should be eliminated when ambiguity occurs. Normally, the correct solution will result in a smaller re-projection error, which represents the difference between the projections and tracked dots. In each image, he re-projection errors from both solutions are compared and if both give small errors close to the zero the ambiguity occurs. To resolve this issue, we use points from a different plane where the wrong solution gives a large re-projection error.

An ambiguous case is shown in Fig. 3.5 in which two marker poses are estimated using the detected circular dots indicated in green. If we project the circular-dot pattern using the two estimated poses, then both projections will be aligned with the tracked pattern. Therefore, we take advantage of the chessboard vertices on the hybrid marker which are not located in the same plane as the circular-dot pattern. In Fig. 3.5, projected vertices using the two poses are shown in red and blue respectively. They are compared to the current vertices detection, which is drawn as yellow dots, and the closer one (smaller re-projection error) is chosen as the correct pose. In this example, the pose corresponds to the red projections is correct. To verify this, we place an ultrasound image locating at the transducer of the probe and then project the outline of the ultrasound image on the camera image using the estimated poses. As shown in Fig 3.5, it is clear that the red projected outline is the correct one.

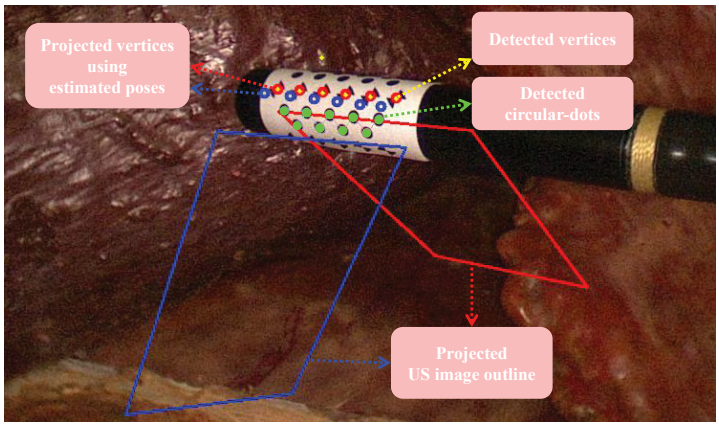


Figure 3.5.: An illustration of ambiguous pose removal using chessboard vertices. The red and blue circles indicate projections of the model chessboard vertices using two candidate poses. The two quadrangles represent the projected outline of an ultrasound image on the image. In this example, we can easily see that the pose corresponds to the red is correct.

3.7. Experiments and Results

3.7.1. Hardware Set-up

Images of surgical view are captured using a 10mm diameter monocular laparoscope (Karl Storz GmbH, Germany) whose intrinsic parameters are calibrated using [131]. We tested the proposed framework on two ultrasound transducers: (a) UST-533 linear array microsurgery probe; (b) UST-5550 linear array 2-way laparoscopic probe. Both transducers are driven by a Prosound α-10 machine (Hitachi Aloka Medical Ltd, Japan). Laparoscopic and ultrasound images are streamed to a computer (3.4 GHz CPU, 16G RAM) using DVI2USB converters (Epiphan System Inc, Canada).

To validate the robustness of the tracking framework on the proposed cylindrical marker, five cylindrical rigid bodies with various diameters from 6 to 14 mm are designed and printed, as shown in Fig. 3.6. Four

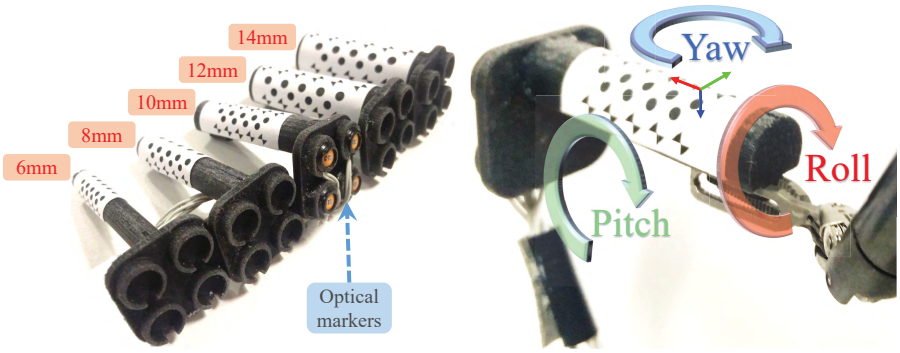


Figure 3.6.: 3D printed rigid body with various size for validation (*left*) and definition of the marker's local coordinate system (*right*).

Table 3.2.: Work space for the proposed marker on various tool size

Diameter (mm)	Distance to camera (mm)	Roll (deg)	Pitch (deg)	Yaw (deg)
6	30 to 125	±68	±56	±81
8	35 to 160	±76	±71	±83
10	40 to 185	±82	±76	±85
12	60 to 200	±85	±78	±83
14	60 to 210	±89	±78	±80

active optical sensors are placed on the body which is tracked by an Optotrak Certus system (Northern Digital Inc, Canada) that achieves 0.1 mm accuracy. The rigid body can be grasped and moved by a da Vinci® Cadiere forceps instrument (Intuitive Surgical, USA). As the UST-533 probe does not have a cylindrical shape, a housing adaptor is printed, as shown in Fig. 3.7, so that it can be tracked by using the proposed marker.

3.7.2. Work Space Analysis

To identify the work space of the proposed marker and associated detection algorithm, we recorded the maximal distance and rotation of the tool, above which the marker cannot be detected. We investigated

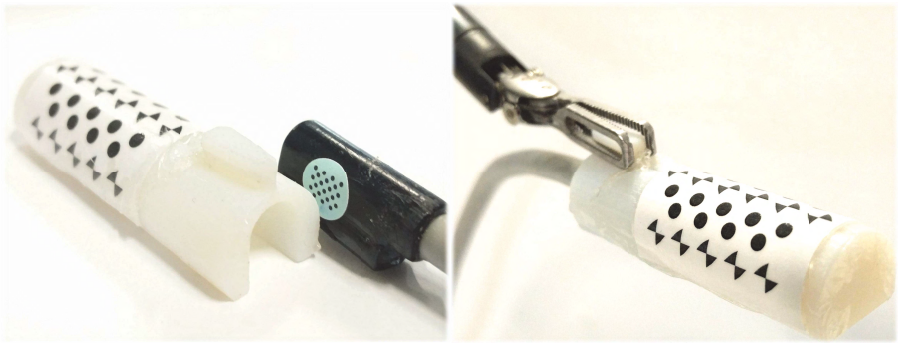


Figure 3.7.: An housing adaptor that can hold a planar ultrasound probe. With the housing, the planar probe can be tracked by using the proposed marker.

the translational motion along camera's optical axis and the rotational motion around the local axis (roll, pitch and yaw) of the validation rigid body as depicted in Fig. 3.6. The dimension of the marker has been modified based on the size configuration in order to maximise the marker's coverage on the curved surface. For testing the rotational motion, the marker was placed about half of its maximal work distance to obtain better visibility. We repeated the experiment on all five markers with different sizes whose results are shown in Table 3.2. A positive correlation can be found between the marker size and distance to the camera. As the marker becomes larger, the maximal work distance is accordingly extended due to better visibility. We observed that a reasonable work space for most minimally-invasive procedures ranges from 50 to 200 mm. The rotation around the roll axis has also been extended up to ± 89 degrees when the tool diameter increases to 14 mm. Regarding rotation around the pitch and yaw axis, the marker is capable of dealing with rotation more than ± 70 and ± 80 degrees, respectively. In summary, these results show that the proposed marker can be applied to instruments with different size while providing a large work space.

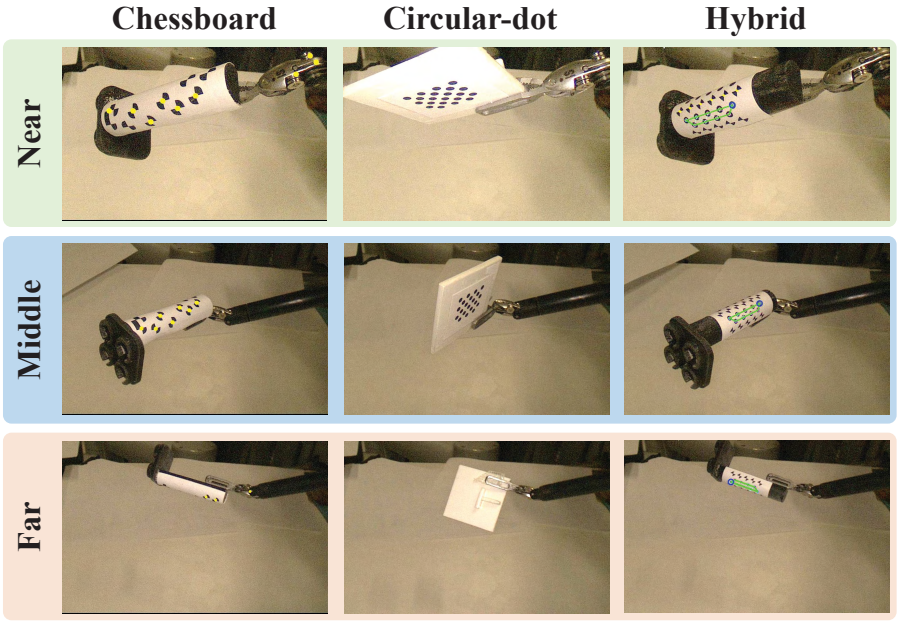


Figure 3.8.: A snapshot of the experiment for detection rate comparison. For each distance group, the same movement has been reproduced to test the detection performance of different markers. The circular-dot marker cannot be detected while the marker is in a very slant view. In the far distance group, we show a scenario where the marker is rotated around its axial axis, so only the other side is visible.

3.7.3. Detection Rates

As reported in [64], the detection rate of a circular-dot feature is higher than a chessboard feature. In order to evaluate the detection rate of the proposed marker, which consists of both circular-dot and chessboard vertices, we compared with two previous marker designs with one containing only chessboard vertices [63] and the other one containing only circular-dot [64]. The proposed hybrid and the chessboard only markers were placed on the 10mm diameter cylindrical rigid body. The circular-dot marker can only be placed on a planar surface due to the limitation of its detection algorithm. The size of the circular-dot

marker was set to be similar to our hybrid marker for a fair comparison. The experiment was divided into three groups based on different distances to the camera: near (50-100 mm), middle (100-150 mm) and far (150-200 mm), as shown in Fig. 3.8. For each group, we run three trials for the three types of markers using the same trajectory which mainly consists of rotational motion around the roll, pitch and yaw axis. The illumination level was fixed in the same distance group. The trajectory is firstly recorded from a manual demonstration using the da Vinci® robot with dVRK controllers [132]. During the trial, the recorded trajectory is played back and the detection results are saved. For the chessboard vertices marker, a detection result is considered as a success if at least four features are found as required by [63]. For the circular-dot marker, we only apply the detection part without tracking so that the performance is similar to [64]. In this case, a frame is accepted if the number of detected blobs equals to the number of dots in the marker. For circular-dot in the hybrid marker, one is considered to be a success if either of the top, middle or bottom pattern is detected.

The quantitative results, as shown in Table 3.3, confirm that the hybrid marker with proposed tracking framework is superior to previous methods in terms of marker detection rate. In particular, the circular-dot pattern(s) in the hybrid marker were nearly 100% detected in different distances. The planar circular-dot marker has much less detection rate due to self-occlusion caused by the rotational motion. Regarding the chessboard feature, the results based on the hybrid marker are comparable to the method introduced in [63] for near and middle distance. However, our chessboard feature detection method can be deteriorated when the marker is far away. This is because more rigorous criteria were used to ensure the correspondence between the detected chessboard features and model points for accurate pose estimation. For example, if less than four chessboard features with the same orientation are detected, the result is regarded as a failure.

Table 3.3.: Detection rates of different markers in different work distances

Distance	Chessboard vertices [63]	Circular-dot [64]	Hybrid chessboard (ours)	Hybrid dots (ours)
Near (50-100 mm)	87.2%	53.2%	98.3%	99.7%
Middle (100-150 mm)	86.8%	61.5%	84.7%	100%
Far (150-200 mm)	79.5%	75.7%	11.0%	99.8%

Table 3.4.: Pose estimation error for the proposed marker on different tool size

Diameter (mm)	Trans. Mean Error ± STD (mm)	Rot. Mean Error ± STD (deg)
6	1.43 ± 1.09	0.55 ± 0.38
8	4.15 ± 1.53	1.70 ± 0.47
10	2.80 ± 2.02	0.72 ± 0.52
12	2.53 ± 1.40	0.69 ± 0.33
14	3.79 ± 1.70	0.90 ± 0.45

3.7.4. Pose Estimation Error

In order to quantitatively validate the pose estimation error, we used the Optotrak Certus system to obtain the ground truth. To this end, two registrations are required: optical sensors to the hybrid marker $\mathbf{T}_S^M$ and optical system to laparoscope $\mathbf{T}_O^L$. In the following of this thesis, $\mathbf{T}_1^2$ is denoted as a transformation from coordinate $\{\mathbf{1}\}$ to $\{\mathbf{2}\}$. The problem can be treated as a $AX = YB$ problem which is defined as:

$$\mathbf{T}_O^S \cdot \mathbf{T}_S^M = \mathbf{T}_O^L \cdot \mathbf{T}_L^M, \tag{3.4}$$

where $\mathbf{T}_O^S$ and $\mathbf{T}_L^M$ is optical sensors pose in the optical system coordinate frame and marker pose in laparoscope coordinate frame, respectively. A total of 10 $\mathbf{T}_O^S$ and $\mathbf{T}_L^M$ were acquired, so that $\mathbf{T}_S^M$ and $\mathbf{T}_O^L$ can be calculated using [133].

For experimental validation, we ran five trials for the markers with different diameters. For each trial, a total of 50 measurements were made near half of the maximum work distance (as defined in Table 3.2). For each measurement, a relative pose between the estimated and ground truth pose is calculated as:

$$\mathbf{T}_M^{M^\star} = \mathbf{T}_L^{M^{-1}}, \tag{3.5}$$

where $\mathbf{T}_S^{M^\star} = \mathbf{T}_S^M$ is the calibrated transformation between the optical sensors and hybrid marker. The pose estimation error, as shown in Table 3.4, is the translation and rotation component of $\mathbf{T}_M^{M^\star}$. It is worth noting that this error combines the registration error of $\mathbf{T}_S^M$ and $\mathbf{T}_O^L$ in equation 3.4. The results in Table 3.4 show that the pose estimation error is slightly smaller than results presented in [63] where an error of 4.4±3.3 mm was reported. It is worth noting that the 8 mm tool performed worse than the 6 mm one due to the marker design. For the 6 mm marker, the top and bottom chessboard vertices are closer than the 8 mm marker, so they can be detected more frequently, as

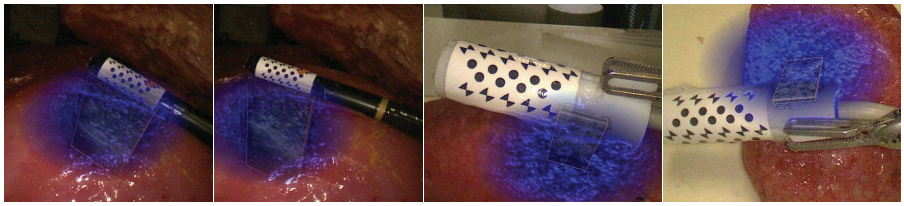


Figure 3.9.: Ultrasound image is fused with laparoscopic image using an inverse realism technique [7].

shown in Fig. 3.6. To improve detectability of the 8mm marker, the pattern design can be adjusted so that at least one line of chessboard vertices is visible.

3.7.5. Example Application for Ultrasound Overlay

Here, an example application of the proposed tracking approach for surgical AR is demonstrated. To this end, we combine our tracking framework with an AR approach introduced in [7]. As shown in Fig. 3.9, the proposed framework provides accurate tracking, which enables seamless integration intraoperatively with see-through vision. The speed including detection and tracking is 20 frames per second on 960x540 images. The visualisation with inverse realism can be processed in real-time [134].

3.8. Discussion and Conclusion

The proposed tracking framework allows for large rotational and translational motion whilst providing superior detection rate and pose estimation accuracy. These are shown in Tables 3.2, 3.3 and 3.4. In comparison with the previous methods [64, 63], the proposed framework improves the detection rate by considering the temporal information between consecutive images. Further, the proposed method achieves competitive pose estimation accuracy compare to [63] while allowing

real-time processing which enhances its usability for clinical applications. The current framework can cope with self- and partial occlusion as demonstrated in Fig. 3.3. It is worth noting that the effect of smoke is not yet considered. This will be included in our future work. We have also shown that the proposed marker can be applied to different types of surgical instruments as shown in Fig. 3.4.

Currently, the camera calibration is carried out off-line and the parameters are assumed to be fixed during tracking. A useful improvement of the framework is to include simultaneous camera calibration as the marker is tracked. This will enhance the applicability of the framework in clinical use where camera calibration may be troublesome when camera focus has to be changed.

In conclusion, a novel hybrid marker made of circular dots and chessboard vertices has been proposed, along with an algorithm for tracking surgical tools. The circular-dot pattern is used for pose estimation while chessboard vertices are used to address the issue of ambiguous pose estimation. In addition, temporal information is considered to deal with partial occlusion. Extensive experiments have been conducted on tracking a cylindrical ultrasound probe with different scenarios. Both qualitative and quantitative results confirm that our framework outperforms the current state-of-the-art approaches in terms of detection rates and pose accuracies, whilst also provides a large work space. Furthermore, we also show the results of our framework on different surgical tools with varying sizes, and the results verify the practical value of our framework in minimally invasive procedures.

4. Model based Instrument Tracking for Robotic Surgery † §

4.1. Introduction

Instrument tracking is an enabling technique in the context of CAS. Allowing the continuous localisation of surgical instruments, it is a prerequisite for providing instrument control during autonomous execution of surgical tasks. In addition, augmented display of *in vivo* imaging modalities, such as ultrasound, requires consistent tool tracking during operation. To this end, approaches have been proposed to obtain the position and orientation of the surgical tool by using external tracking devices such as electromagnetic sensor [58], optical tracker [121] or hybrid system [69]. However, the introduction of additional tracking devices into the theatre would not only occupy the valuable space in the operating field, but also suffer from intrinsic limitations of these tracking devices such as ferromagnetic interference and line-of-

†Content from this chapter was published as:
Real-time 3D Tracking of Articulated Tools for Robotic Surgery. Ye, M., Zhang, L., Giannarou, S. and Yang, G.-Z.. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) pp. 386-394 (2016).

§This work has been conducted in collaboration with Dr Menglong Ye. The author of this thesis developed a software for the virtual tool rendering and data communication with the da Vinci® robot, and participated all CAD model processing, experiments, data processing and labelling, and validation. Dr Menglong Ye developed the components for template matching, 2D part detection and verification, and 3D pose estimation, as well as method validation.

sight issues. Therefore, a direct use of laparoscopic images for tracking instruments in MIS is more practical than other methods. Beside the marker-based methods introduced in Chapter 3, model-based markerless methods [66] have been developed to detect and track surgical instruments.

Recent advances in surgical robots have significantly improved the dexterity of the surgeons, along with enhanced 3D vision and motion scaling. Surgical robots such as the da Vinci® (Intuitive Surgical, Inc. CA) platform, can allow the augmentation of preoperative data to enhance the intraoperative surgical guidance. In robotic surgery, tracking of surgical tools is an important task for applications such as safe tool-tissue interaction and surgical skills assessment.

In the last decade, many approaches for surgical tool tracking have been proposed. The majority of these methods have focused on the tracking of laparoscopic rigid tools, including using template matching [135] and combining colour-segmentation with prior geometrical tool models [136]. In [137], the 3D poses of rigid robotic tools were estimated by combining random forests with level-sets segmentation. More recently, tracking of articulated tools has also attracted a lot of interest. For example, Pezzementi et al. [138] tracked articulated tools based on an offline synthetic model using colour and texture features. The CAD model of a robotic tool was used by Reiter et al. [67] to generate virtual templates using the robot kinematics. However, thousands of templates were created by configuring the original tool kinematics, leading to time-demanding rendering and template matching. In [139], boosted trees were used to learn predefined parts of surgical tools. Similarly, regression forests have been employed in [140] to estimate the 2D pose of articulated tools. In [141], the 3D locations of robotic tools estimated with offline trained random forests, were fused with robot kinematics to recover the 3D poses of the tools. Whilst there has been significant progress on surgical tool detection and tracking, none of the existing approaches have thus far achieved real-time 3D tracking of articulated robotic tools.

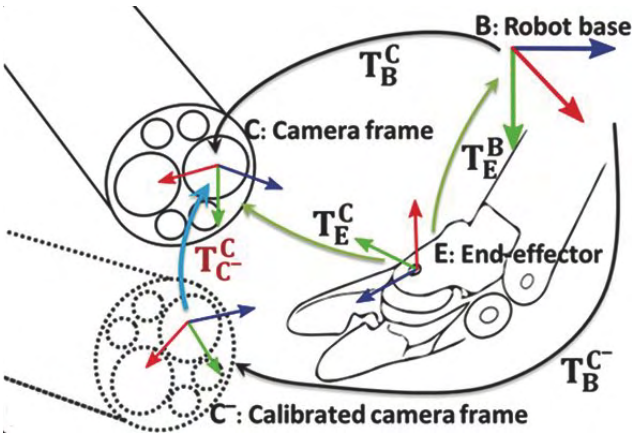


Figure 4.1.: Illustration of transformations used in this chapter.

This chapter presents a framework for real-time 3D tracking of articulated tools in robotic surgery. Similar to [67], CAD models have been used to generate virtual tools and their contour templates are extracted online, based on the kinematic readings of the robot. The tool detection on the real camera image is performed via matching the individual parts of the tools rather than the whole instrument. This enables the proposed method to deal with the changing pose of the tools due to articulated motion. Another novel aspect of the proposed framework is the robust verification approach based on 2D geometrical context, which is used to reject outlier template matches of the tool parts. The inlier 2D detections are then used for 3D pose estimation via the EKF. Experiments have been conducted on phantom, *ex vivo* and *in vivo* video data, and the results verify that the proposed approach outperforms the state-of-the-art.

4.2. Part-based Online Templates for Tool Detection

The proposed framework includes three main components. The first component is a virtual tool renderer that generates part-based tem-

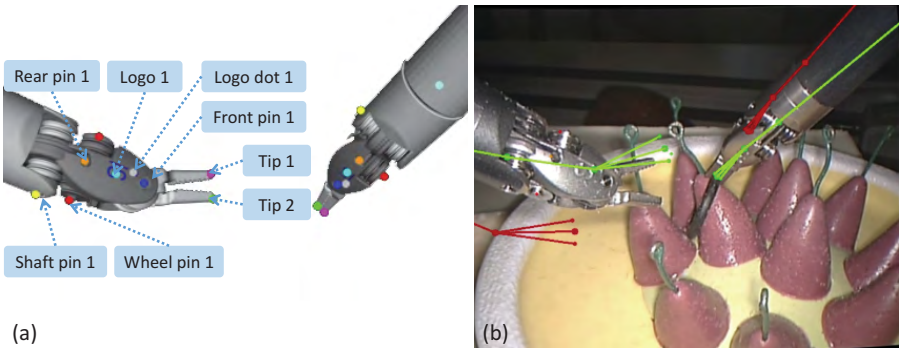


Figure 4.2.: Virtual rendering example of the large needle drivers (a) and the corresponding camera view (b). The predefined keypoint locations are labeled in (a). The red skeletons show the projection based on the tool kinematics, and green skeletons show the tracking results of the proposed framework (b).

plates online. After template matching, the second component performs verification to extract the inlier 2D detections. These 2D detections are finally fused with kinematic data for 3D tool pose estimation. The proposed framework is implemented on the da Vinci® robot. The robot kinematics are retrieved using the dVRK [132].

In this work, to deal with the changing pose of articulated surgical tools, the tool detection has been performed by matching individual parts of the tools, rather than the entire instrument, similar to [139]. To avoid the limitations of offline training, the proposed method generates the part models on-the-fly such that the changing appearance of tool parts can be dynamically adapted.

To generate the part-based models online, the CAD model of the tool and the robot kinematics have been used to render the tool in a virtual environment. The pose of a tool in the robot base frame B can be denoted as the transformation $\mathbf{T}_E^B$, where E is the end-effector coordinate frame shown in Fig. 4.1. $\mathbf{T}_E^B$ can be retrieved from dVRK (kinematics) to provide the 3D coordinates of the tool in B. Thus, to set the virtual view to be the same as the laparoscopic view, a standard hand-eye calibration [142] is used to estimate the transformation

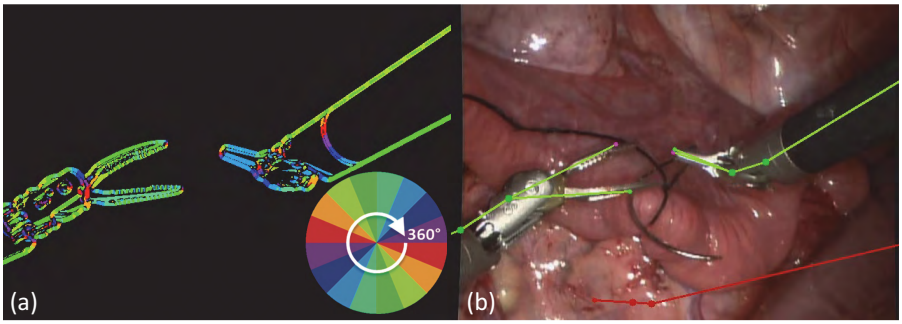


Figure 4.3.: (a) Extracted gradient orientations from virtual rendering where the the orientations are quantised and colour-coded as shown in the pie chart. (b) The corresponding camera view with the tracking results.

$\mathbf{T}_B^C$ from B to the camera coordinate frame C. However, errors in the calibration can affect the accuracy of $\mathbf{T}_B^C$, resulting in a 3D pose offset between the virtual tool and the real tool in C. In this regard, the transformation found from the calibration is presented as $\mathbf{T}_B^{C^-}$, where C^- is the camera coordinate frame that includes the accumulated calibration errors. Therefore, a correction transformation denoted as $\mathbf{T}_{C^-}^C$ can be introduced to compensate for the calibration errors.

Here, n=14 keypoints $\mathbf{P}^B = \left\{p_i^B\right\}_{i=1}^n$ are defined on the tool, and the large needle driver is taken as an example. The keypoints include the points shown in Fig. 4.2 and those on the symmetric side of the tool. These keypoints represent the skeleton of the tool, which also apply to other da Vinci® tools. At time t, an image I_t can be obtained from the laparoscopic camera. The keypoints can be projected in I_t with the camera intrinsic matrix $\mathbf{K}$ via

$$\mathbf{P}_t^I = \frac{1}{s}\mathbf{K}\mathbf{T}_{C^-}^C\mathbf{T}_B^{C^-}\mathbf{P}_t^B. \tag{4.1}$$

Here, s is the scaling factor that normalises the depth to the image plane.

To represent the appearance of the tool parts, the Quantised Gradient Orientations (QGO) approach [143] has been used (see Fig. 4.3).

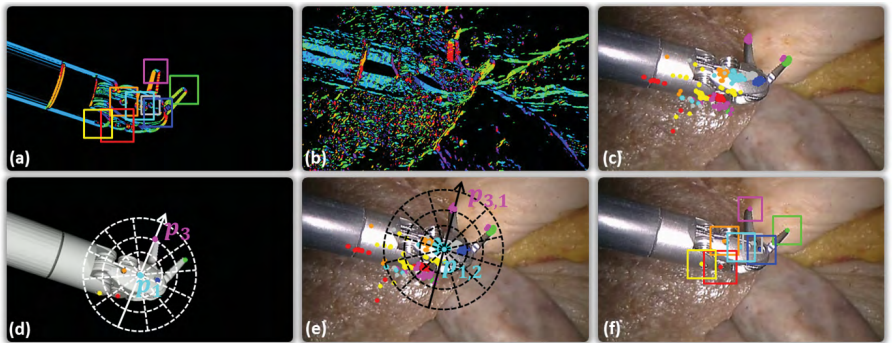


Figure 4.4.: (a) An example of part-based templates; (b) Quantised gradient orientations from a camera image; (c) Part-based template matching results of tool parts; (d) and (e) Geometrical context verification; (f) Inlier detections obtained after verification.

Bounding boxes are created to represent part-based models and centred at the keypoints in the virtual view (see Fig.4.4(a)). The box size for each part is adjusted according to the z coordinate of the keypoint with respect to the virtual camera centre. QGO templates are then extracted inside these bounding boxes. As QGO represents the contour information of the tool, it is robust to cluttered scenes and illumination changes. In addition, a QGO template is represented as a binary code by quantisation, thus template matching can be performed efficiently.

Note that not all of the defined parts are visible in the virtual view, as some of them may be occluded. Therefore, the templates are only extracted for those m parts that facing the camera. To find the correspondences of the tool parts between the virtual and real images, QGO is also computed on the real image (see Fig.4.4(b)) and template matching is then performed for each part via sliding windows. Exemplar template matching results are shown in Fig.4.4(c).

4.3. Tool Part Verification via 2D Geometrical Context

To further extract the best location of the tool parts, a consensus-based verification approach [116] is included. This approach analyses the geometrical context of the correspondences in a PROgressive SAmple Consensus (PROSAC) scheme [144]. For the visible keypoints $\{p_i\}_{i=1}^{m}$ in the virtual view, their 2D correspondences are denoted in the real camera image as $\{p_{i,j}\}_{i=1,j=1}^{m,k}$, where $\{p_{i,j}\}_{j=1}^{k}$ represent the top k correspondences of p_i sorted by QGO similarities.

For each iteration in PROSAC, the proposed method selects two point pairs from $\{p_{i,j}\}_{i=1,j=1}^{m,k}$ in a sorted descending order. These two pairs represent the correspondences for two different parts, e.g., pair of p_1 and $p_{1,2}$, and pair of p_3 and $p_{3,1}$. The two pairs are then used to verify the geometrical context of the tool parts. As shown in Fig.4.4(d) and (e), two polar grids are used to indicate the geometrical context of the virtual view and the camera image. The origins of the grids are defined as p_1 and $p_{1,2}$, respectively. The major axis of the grids can be defined as the vectors from p_1 to p_3 and $p_{1,2}$ to $p_{3,1}$, respectively. The scale difference between the two grids is found by comparing $d(p_1, p_3)$ and $d(p_{1,2}, p_{3,1})$, where $d(\cdot, \cdot)$ is the euclidean distance. The angular and radial bin sizes can be defined as 30 degrees and 10 pixels (allowing moderate out-of-plane rotation), respectively. With these, two polar grids can be created and placed on the virtual and camera images. A point pair is determined as an inlier if the two points are located in the same zone in the polar grids. Therefore, if the number of inliers is larger than a predefined value, the geometrical context of the tools in the virtual and the real camera images are considered as matched. Otherwise, the above verification is repeated until it reaches the maximum number (100) of iterations. After verification, the inlier point matches are used to estimate the correction transformation $\mathbf{T}_{C^-}^{C}$.

4.4. From 2D to 3D Tool Pose Estimation

This section describes how to combine the 2D detections with 3D kinematic data to estimate $\mathbf{T}^{C}_{C^{-}}$. Here the transformation matrix is represented as a vector of translations and rotation angles for each axis: $\mathbf{x} = [t_x, t_y, t_z, \alpha, \beta, \gamma]^T$. The n observations (corresponding to the tool parts defined in Section 4.2) can be denoted as $\mathbf{z} = [u_1, v_1, \ldots, u_n, v_n]^T$, where u and v are their locations in the camera image. To estimate $\mathbf{x}$ on-the-fly, the EKF has been adopted to find $\mathbf{x}(k)$ given the observations $\mathbf{y}(k)$ at time k. The process model is defined as $\mathbf{x}(k) = \mathbf{I}\mathbf{x}(k-1) + \mathbf{w}(k)$, where $\mathbf{w}(k)$ is the process noise at time k, and $\mathbf{I}$ is the transition function defined as the identity matrix. The measurement model is defined as $\mathbf{y}(k) = \mathbf{h}(\mathbf{x}(k)) + \mathbf{v}(k)$, with $\mathbf{v}(k)$ being the noise. $\mathbf{h}(\cdot)$ is the nonlinear function with respect to $[t_x, t_y, t_z, \alpha, \beta, \gamma]^T$:

$$\mathbf{h}(\mathbf{x}(k)) = \frac{1}{s}\mathbf{K}\mathbf{f}(\mathbf{x}(k))\mathbf{T}^{C^{-}}_{B}\mathbf{P}^{B}_{t}, \tag{4.2}$$

which is derived according to Eq.4.1. Note here, $\mathbf{f}(\cdot)$ is the function that composes the translation and euler angles (in $\mathbf{x}(k)$) into the 4×4 homogenous transformation matrix $\mathbf{T}^{C}_{C^{-}}$. As Eq.4.2 is a nonlinear function, the Jacobian matrix $\mathbf{J}$ of $\mathbf{h}(\cdot)$ is derived regarding each element in $\mathbf{x}(k)$.

For iteration k, the predicted state $\mathbf{x}^{-}(k)$ is calculated and used to predict the measurement $\mathbf{y}^{-}(k)$, and also to calculate $\mathbf{J}(k)$. In addition, $\mathbf{y}(k)$ is obtained from the inlier detections (Section 4.3), which is used, along with $\mathbf{J}(k)$ and $\mathbf{x}^{-}(k)$, to derive the corrected state $\mathbf{x}^{+}(k)$ which contains the corrected angles and translations. These are finally used to compose the transformation $\mathbf{T}^{C}_{C^{-}}$ at time k, and thus the 3D pose of the tool in C is obtained as $\mathbf{T}^{C}_{E} = \mathbf{T}^{C}_{C^{-}}\mathbf{T}^{C^{-}}_{B}\mathbf{T}^{B}_{E}$. Note that if no 2D detections are available at time k, the previous $\mathbf{T}^{C}_{C^{-}}$ is then used.

At the beginning of the tracking process, an estimate ${}^{0}\mathbf{T}^{C}_{C^{-}}$ is required to initialise EKF, and correct the virtual view to be as close as possible to the real view. Therefore, template matching is performed in multiple scales and rotations for initialisation, however, only one

template is needed for matching of each tool part after initialisation. The Efficient Perspective-n-Points (EPnP) algorithm [145] is applied to estimate ${}^{0}\mathbf{T}^{C}_{C-}$ based on the 2D-3D correspondences of the tool parts matched between the virtual and real views and their 3D positions from kinematic data.

The proposed framework can be easily extended to track multiple tools. This only requires to generate part-based templates for all the tools in the same graphic rendering and follow the proposed framework. As template matching is performed in binarised templates, the computational speed is not deteriorated.

4.5. Experiments and Results

The proposed framework has been implemented on an HP workstation with an Intel Xeon E5-2643v3 CPU. Stereo videos are captured at 25Hz. In the `C++` implementation, the workloads of the part-based rendering and image processing are allocated into two CPU running threads, enabling real-time performance on the proposed framework. The rendering part is implemented based on VTK and OpenGL, of which the speed is fixed as 25Hz. As the proposed framework only requires monocular images for 3D pose estimation, only the images from the left camera were processed. For image size 720x576, the processing speed is $\approx$ 29Hz (without any GPU programming). The threshold of the inlier number in the geometrical context verification is empirically defined as 4. For initialisation, template matching is performed with additional scale ratios of 0.8 and 1.2, and rotations of $\pm$15 degrees, which does not deteriorate the run-time speed due to template binarisation. The proposed method was compared to the tracking approaches for articulated tools including [139] and [141].

For demonstrating the effectiveness of the online part-based templates for tool detection, the proposed approach is compared to the method described in [139], which is based on boosted trees for 2D tool part detection. For ease of training data generation, a subset of

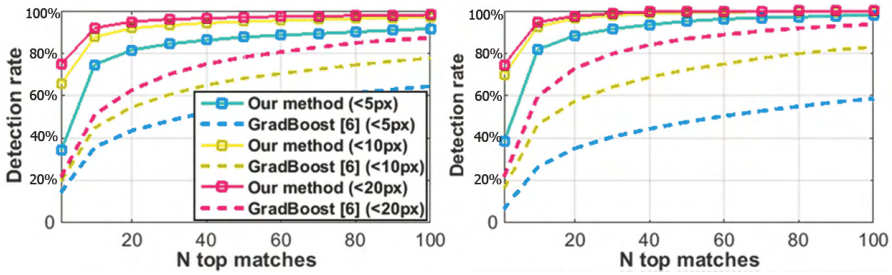


Figure 4.5.: Detection rate results of our online template matching and GradBoost [139] on two single-tool tracking sequences.

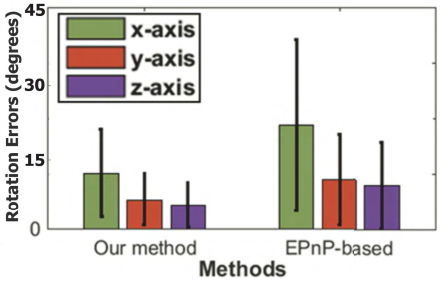


Figure 4.6.: The overall rotation angle errors (mean±std) along each axis of the end-effector on Seqs.1-6.

the tool parts was evaluated in this comparison, namely the front pin, logo, and rear pin. The classifier was trained with 6000 samples for each part. Since [139] applies to single tool tracking only, the trained classifier along with the proposed approach were tested on two single-tool sequences (1677 and 1732 images), where ground truth data was manually labelled. A part detection is determined to be correct if the distance of its centre and ground truth is smaller than a threshold. To evaluate the results with different accuracy requirements, the threshold was therefore sequentially set to 5, 10, and 20 pixels. The detection rates of the methods were calculated among the top N detections. As shown in Fig.4.5 the proposed method significantly outperforms [139] in all accuracy requirements. This is because the templates used in the proposed method are generated adaptively online.

To validate the accuracy of the 3D pose estimation, the centre lo-

Table 4.1.: Translation and rotation errors (mean±std) on Seqs.1-6. Tracking accuracies with run-time speed in Hz (in brackets) compared to [141] on their dataset (Seqs.7-12).

3D Pose Error					**Tracking Accuracy**		
Seq.	**Our method**		EPnP-based		Seq.	**Our method**	[141]
	Trans.(mm)	Rot.(rads.)	Trans.(mm)	Rot.(rads.)			
1	$\mathbf{1.31 \pm 1.15}$	$\mathbf{0.11 \pm 0.08}$	3.10 ± 3.89	0.12 ± 0.09	7	**97.79%(27)**	97.12%(1)
2	$\mathbf{1.50 \pm 1.12}$	$\mathbf{0.12 \pm 0.07}$	6.69 ± 8.33	0.24 ± 0.19	8	**99.45%(27)**	96.88%(1)
3	$\mathbf{3.14 \pm 1.96}$	$\mathbf{0.12 \pm 0.08}$	8.03 ± 8.46	0.23 ± 0.20	9	**99.25%(28)**	98.04%(1)
4	$\mathbf{4.04 \pm 2.21}$	$\mathbf{0.19 \pm 0.15}$	5.02 ± 5.41	0.29 ± 0.18	10	96.84%**(28)**	**97.75%**(1)
5	$\mathbf{3.07 \pm 2.02}$	$\mathbf{0.14 \pm 0.11}$	5.47 ± 5.63	0.26 ± 0.20	11	96.57%**(36)**	**98.76%**(2)
6	$\mathbf{3.24 \pm 2.70}$	$\mathbf{0.12 \pm 0.05}$	4.03 ± 3.87	0.23 ± 0.13	12	**98.70%(25)**	97.25%(1)
Overall	$\mathbf{2.83 \pm 2.19}$	$\mathbf{0.13 \pm 0.10}$	5.51 ± 6.45	0.24 ± 0.18	Overall	**97.83%**	97.81%

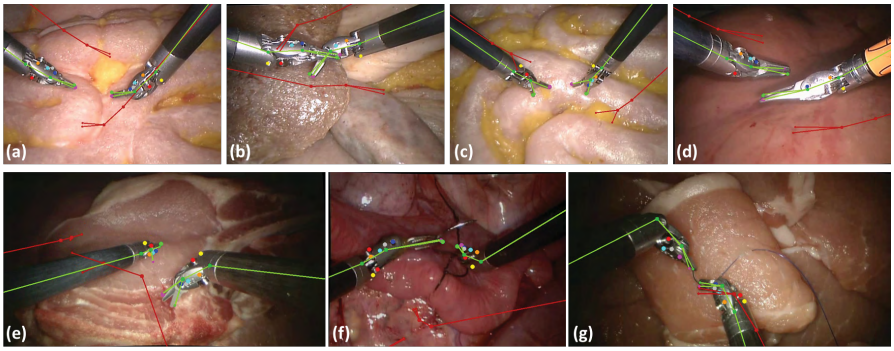


Figure 4.7.: Qualitative results. (a-c) phantom data (Seqs.1-3); (d) *ex vivo* ovine data (Seq.4); (e) and (g) *ex vivo* porcine data (Seqs.9 and 12); (f) *in vivo* porcine data (Seq.11). Red lines indicate the tool kinematics, and green lines indicate the tracking results of the proposed framework with 2D detections in coloured dots.

cations of the tool parts are manually labelled on both left and right camera images on phantom (Seqs.1-3) and *ex vivo* (Seqs.4-6) video data to generate the 3D ground truth. The tool pose errors are then obtained as the relative pose between the estimated pose and the ground truth. The proposed approach was also compared to the 3D poses estimated performing EPnP for every image where the tool parts are detected. However, EPnP generated unstable results and had inferior performance to the proposed approach as shown in Table 4.1 and Fig.4.6.

The proposed framework has also been compared to the method described in [141]. As their code is not publically available, the proposed framework is run on the same *ex vivo* (Seqs.7-10,12) and *in vivo* data (Seq.11) used in [141]. Example results are shown in Fig.4.7(e-g). For achieving a fair comparison, the tracking accuracy is evaluated as explained in their work, and presented both our results and theirs reported in the paper in Table 4.1. Although the proposed framework achieved slightly better accuracies than their approach, our processing speed is significantly faster, ranging from 25–36Hz, while theirs is approximately 1–2Hz as reported in [141]. As shown in Figs.4.7(b) and (d), the proposed method is robust to occlusion due to tool intersections and specularities, thanks to the fusion of 2D part detections and kinematics. In addition, the proposed framework is able to provide accurate tracking even when $\mathbf{T}_B^{C^-}$ becomes invalid after the laparoscope has moved (Fig.4.7(c), Seq.3). This is because $\mathbf{T}_{C^-}^{C}$ is estimated online using the 2D part detections.

4.6. Discussion and Conclusion

This chapter has proposed a real-time framework for 3D tracking of articulated tools in robotic surgery. Online part-based templates are generated using the tool CAD models and robot kinematics, such that efficient 2D detection can then be performed in the camera image. For rejecting outliers, a robust verification method based on 2D geometrical context is included. The inlier 2D detections are finally fused with robot kinematics for 3D pose estimation. The proposed framework can run in real-time for multi-tool tracking, thus can be used for imposing dynamic active constraints and motion analysis. The results on phantom, *ex vivo* and *in vivo* experiments demonstrate that the proposed approach can achieve accurate 3D tracking, and outperform the current state-of-the-art.

The tool localisation accuracy of the proposed method may be insufficient for the automation of some surgical tasks that require fine

movements. Furthermore, the visual occlusion of the tool, such as specular highlight and the blood stain on the tool, can result in significant errors in the pose estimation. Therefore, both accuracy and robustness of the method have to be improved before it can be used for visual servoing of the tool. In the experiments, we found that errors in the pose estimation mainly occur along the camera depth direction. This is probably caused by the use of a single camera instead of the stereo camera provided by the da Vinci robot. Therefore, a potential improvement can be made by utilised both cameras to reduce the error along the depth direction.

5. Autonomous Endomicroscopic Scanning and Fusion on Surface †

5.1. Introduction

In the previous chapters, real-time tool tracking methods are proposed which enable surgical task automation by using visual servoing. In this chapter, we will investigate how autonomous scanning combined with robotic guidance can be used to provide detailed tissue characterisation. As mentioned before, robotic MIS systems with stereo vision can provide better manipulation and depth perception. In particular, the da Vinci® robot (Intuitive Surgical Inc., CA) is a successful surgical platform which is widely used for treatment of gynaecological and urological cancer. The da Vinci® robot is purely a master-slave system that provides tele-operation, with no automation or artificial intelligence features. In this chapter, the potential of the da Vinci® system for surgical automation is investigated by using the da Vinci Research Kit.

Whilst human guidance is essential for MIS, a recent review [146] has concluded that surgical automation can be used for some important steps of the operation, which could provide new augmented abilities

†Content from this chapter was published as:
Zhang, L., Ye, M., Giataganas, P., Hughes, M., Yang, G.-Z.: Autonomous scanning for endomicroscopic mosaicing and 3d fusion. In: IEEE International Conference on Robotics and Automation (ICRA). pp. 3587-3593 (2017).

for the surgeon, allowing them to perform tasks more accurately and intuitively. To this end, studies have been conducted, for instance, to maintain consistent motion for ultrasound elastography [147]. In [107], a human-machine collaborative framework was proposed to improve surgeons' performance by semi-automating surgical subtasks. Further studies have also investigated how to perform automatic surgical tissue dissection [148].

Another target for automation is the use of pCLE for generating *in situ* microscopic images of tissue. pCLE has been used extensively in the gastrointestinal tract and abdominal organs (e.g. [149, 150]), mainly to discriminate between normal and cancerous tissue. However, with a Field-of-View (FoV) of typically 0.25-1 mm, it is difficult to characterise larger areas of tissue or to return to previously sampled points [151]. Video mosaicking to synthesise a larger FoV requires the operator to maintain optimal probe pose and tissue-contact while performing slow and controlled scanning motions with sub-millimetre accuracy, which is a difficult manual task. Robotised instruments could therefore potentially allow larger, more consistent mosaics to be created. pCLE has previously been used both with an industrial robot and in bespoke devices [152, 153, 5], while [50] presented simultaneous robotic pCLE and ultrasound scanning. Integration with the da Vinci® [25, 154], has also been investigated, but weak-depth perception and poor ergonomics mean that it is still challenging to scan consistently while maintaining probe orientation and continuous surface contact. Therefore, the development of an autonomous pCLE scanning system would benefit its use in surgical operations by reducing the cognitive load on users and improving the scanning accuracy over larger 3D surfaces.

This chapter presents a 6 DoF visual servoing method based on the da Vinci® robot with the dVRK. With the proposed approach, the robot is able to achieve smooth motion with sub-millimetre accuracy, and it has been demonstrated that an autonomous large-area pCLE scan can be performed by the robot over a user-defined area. In addi-

tion, to facilitate intraoperative tissue diagnosis and identification, a 3D visualisation method is proposed to fuse the 3D tissue surface with multiple microscopic mosaics on-the-fly. This 3D fusion approach is designed to provide the surgeon with an intuitive and effective real-time visualisation of multi-scale imaging information, which supports surgical diagnosis and planning. Our framework has been tested on phantom and *ex vivo* tissue experiments and the results demonstrate its potential clinical value.

5.2. Hardware and Software Overview

The hardware components of the framework, highlighted in red in Fig. 5.1, consist of a PSM of a da Vinci® robot with dVRK controllers, a stereo laparoscope, and a custom endomicroscope system. The dVRK controllers are connected to a host PC via a IEEE 1394 firewire interface in a daisy chain topology. The stereoscopic system provides SD (720x576) video streaming for both left and right channels at 25 Hz. In the host PC, the stereo stream is captured by a Kona 4 PCIe frame grabber (AJA Video System).

The endomicroscope system is an in-house laser line-scanning fluorescence system, described fully in [155], coupled to a Cellvizio UHD Probe (Mauna Kea Technologies, France). The probe, which consists of a 30,000 core fibre imaging bundle and a micro-lens, provides a FoV of 240 μm and a fibre-sampling limited resolution of approximately 2.4 μm. The line-scanning system provides optical sectioning, meaning that the endomicroscope collects light only from an in-focus plane approximately 20 μm in depth. Processed microscopic images are scaled to 300x300 pixels and served to the controller at 80 Hz via a TCP/IP connection. In addition to the host PC, a second PC with modest configuration is dedicated to 2D mosaicing and 3D visualisation which runs at 15 Hz and 1 Hz, respectively.

The software component of the proposed framework runs on the ROS, highlighted in green in Fig. 5.1. The stereo images captured

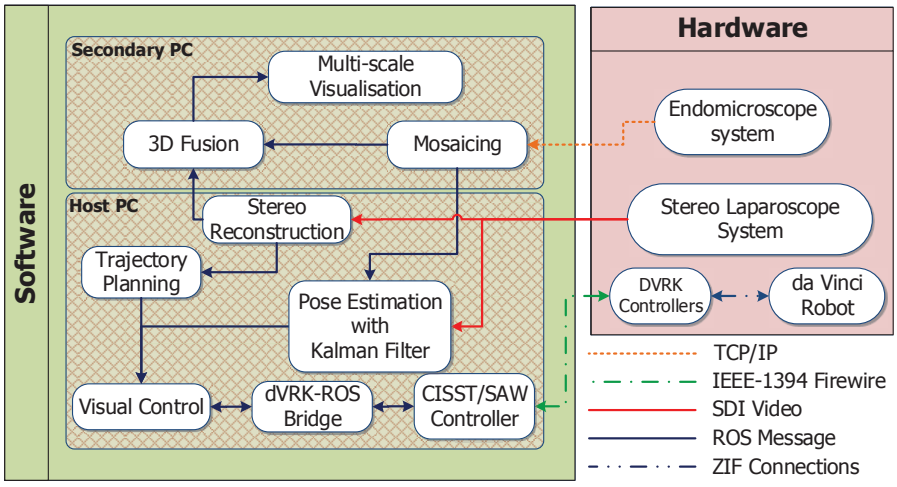


Figure 5.1.: An overview of our proposed system framework for autonomous endomicroscopic scanning and 3D mosaicing.

from the camera are used for 3D tissue surface reconstruction via a stereo matching method. The reconstructed 3D surface is used for two purposes: (1) to plan a scanning trajectory based on its positional and normal information; (2) to provide 3D fusion with the 2D image mosaics. The video stream from the laparoscope system is also used for estimating the pose of the pCLE probe.

As shown in Fig. 5.2, the end-effector of the robot grasps an adapter that holds the pCLE probe. A marker (KeyDot®, Key Surgical, Minnesota) is attached to the adapter for pose estimation of the probe. The stereo system is calibrated using the method presented in [131] to obtain the intrinsic and extrinsic camera parameters such that the 3D depths of the tissue can be recovered via stereo matching. The trajectory planning consists of two parts: a global plan to guide the probe to the desired location of the scan (identified by the user), and a local plan to generate a mosaic.

A visual control component closes the loop by comparing the current and desired probe poses and commanding the robot to minimise their difference, causing the probe to follow the planned trajectory. The robot's end-effector Cartesian pose is read and set via a dVRK-ROS

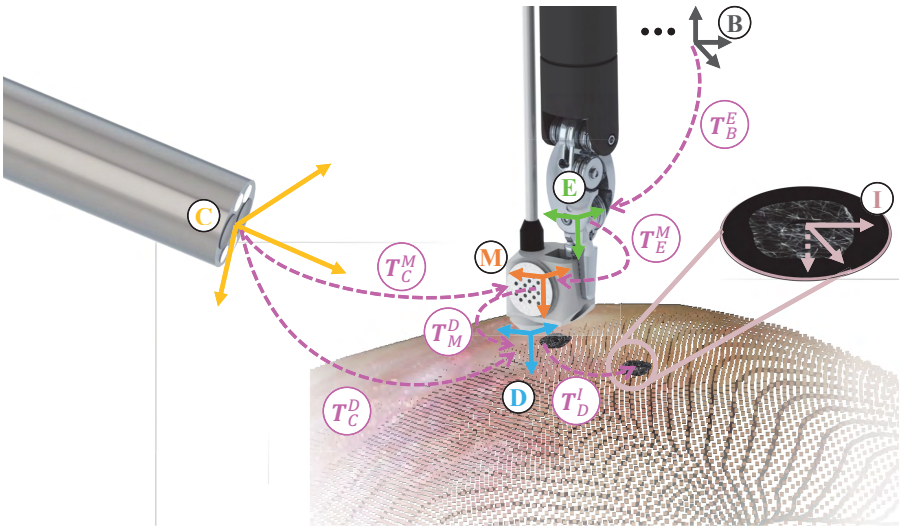


Figure 5.2.: A description of coordinate systems (frames) and related transformations. B: robot base frame; E: end-effector frame; M: marker frame; D: probe tip frame; C: camera frame; I: mosaic point cloud frame.

component which is connected to a low level PID controller implemented by the SAW package using the cisst library [132].

pCLE images are stitched together in real-time using pairwise registration of frames by template matching via normalised cross-correlation. A position estimate obtained from the registration process is fused with the pose estimation from the camera using a Kalman filter, to provide probe position information used for closed-loop control during the scanning. The image mosaic is also fused with the reconstructed surface on-the-fly to provide both macro- and micro-views of the scanned region. By providing accurate and robust pose estimation, this visual servoing approach permits smooth and accurate probe scanning, allowing the formation of contiguous image mosaics.

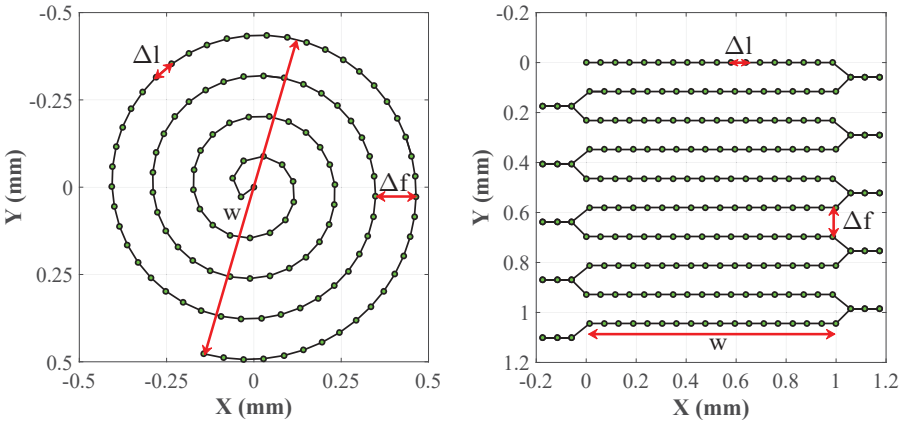


Figure 5.3.: Illustration of two types of trajectory pattern for pCLE scanning: spiral (left) and raster (right). Three parameters that define the trajectory are labelled with red double-arrow line.

5.3. Trajectory Planning Based on 3D Reconstruction

In this work, the scanning trajectory is planned to adapt to the 3D profile of the tissue surface. The stereo images are rectified and undistorted by using the camera intrinsic and extrinsic parameters. The Efficient LArge-scale Stereo (ELAS) [156] approach has been applied to the stereo images for 3D surface reconstruction. To determine the normal to a point on the surface, the problem is approximated by the problem of estimating the normal of a plane tangent to the surface, similar to [157].

The scanning trajectory, which is a sequence of probe tip poses in the camera frame ($\mathbf{T}_C^D$), is planned in two stages: the global trajectory planning, which moves the probe to the starting pose for scanning, and the local trajectory planning, which performs high resolution scanning from the starting pose. To generate the probe tip poses in the global trajectory, the proposed method interpolates between the current probe tip pose at a standby position, and the desired starting pose

for scanning. The desired starting pose ensures that the endomicroscope is perpendicular to the tissue surface (using the surface normal from the 3D reconstruction) which is important for high quality microscopic image retrieval.

For the local scanning trajectory, it can be assumed that scanning surface is locally planar, which is reasonable as each individual scanning region is usually small (about 2x2mm) and most tissues are compliant. The local scanning trajectory can be planned on a 2D plane, with the aim of generating a contiguous mosaic over the desired area of tissue. Two types of scanning patterns have been investigated, namely the raster and spiral trajectories. Example patterns are shown in Fig. 5.3. A set of common parameters are used to specify the trajectory generation for both patterns. The diameter of the scanned region is defined by w which is typically 1-2 mm in our experiment. In order to reduce the likelihood of gaps in the mosaic image, but without excessive over-sampling, the overlap parameter Δf is set to half of the FoV of the endomicroscope, which is 240 μm. the proposed method also defines a point spacing parameter Δl to set the interval between points in the trajectory.

For the raster pattern, each trajectory point (x_k, y_k) is defined as:

$$\begin{aligned} x_k &= \begin{cases} k \cdot \Delta l, & k \in \left[0, \frac{w}{\Delta l}\right], k \text{ in odd row}, \\ w - (k \cdot \Delta l), & k \in \left[0, \frac{w}{\Delta l}\right], k \text{ in even row}, \end{cases} \\ y_k &= k \cdot \Delta f, \qquad k \in \left[0, \frac{w}{\Delta f}\right]. \end{aligned} \tag{5.1}$$

As the probe maintains direct contact with the tissue at all times, minor tissue deformation would occur. Abrupt changes of direction (particularly moving between lines in the raster pattern) results in loading and unloading of the tissue [158]. To minimise this tissue deformation, an extra loading-unloading phase between two scanning lines is added, where the probe advances further than the limit of the planned trajectory, as shown in the Fig. 5.3. This also helps compensating for the backlash of the robot which occurs during large directional changes.

For the spiral pattern the k-th trajectory point is given by:

$$\begin{cases} x_k = b \cdot \theta_k \cdot \cos\theta_k \\ y_k = b \cdot \theta_k \cdot \sin\theta_k \end{cases} \tag{5.2}$$

where b and θ_k are calculated as:

$$\begin{cases} b = \frac{\Delta f}{2\pi}, \\ l_{sp} = \frac{b}{2} \cdot (\frac{w}{2b})^2, \\ \theta_k = \sqrt{\frac{2k \cdot \Delta l}{b}}, k \in \left[0, \frac{l_{sp}}{\Delta l}\right]. \end{cases} \tag{5.3}$$

Here, Δf is the distance between two successive spiral loops and l_{sp} is the total length of the spiral trajectory.

5.4. Pose Estimation with Kalman Filter

Two types of imaging information are used for probe pose estimation: the laparoscopic camera image and the endomicroscope image. For probe pose estimation using the camera, a marker with asymmetric dots is attached to the adapter (see Fig. 5.2). The vision-based detection-tracking method proposed in chapter 3 is used for marker recognition. The detection component is based on blob detection similar to [159], which is applied on every image in the sequence. In order to improve the marker recognition rate, a tracking component based on pyramidal optical flow [130] is included to track the dots of the marker in time. Once the marker pattern is recognised in an image, the efficient perspective-n-points algorithm [145] is applied to calculate the marker pose based on 2D-3D correspondences. The probe pose can be calculated as: $\mathbf{T}_C^D = \mathbf{T}_C^M \cdot \mathbf{T}_M^D$, where $\mathbf{T}_M^D$ is a fixed transformation measured beforehand.

When using endomicroscope images for positional estimation of the probe, only translational motion on the image plane is considered. In order to filter the noisy camera measurements and fuse information

from both imaging modalities, a Kalman filter is applied to obtain a final estimate of the probe probe. Here, a discrete-time linear stationary signal model is considered:

$$\mathbf{x}(k) = \mathbf{A}\mathbf{x}(k-1) + \mathbf{w}(k) \tag{5.4}$$

where $\mathbf{x}(k) \in \mathbb{R}^n$ is the state vector. $\mathbf{w}(k) \in \mathbb{R}^n$ is a sequence of process noise which is assumed to be drawn from a zero mean multivariate normal distribution whose covariance matrix is denoted by $\mathbf{Q}(k)$. In this work, the state vector $\mathbf{x}(k) = [t_x, t_y, t_z, \dot{t_x}, \dot{t_y}, \dot{t_z}, \alpha, \beta, \gamma, \dot{\alpha}, \dot{\beta}, \dot{\gamma}]^\intercal$ represents the position and orientation (in Euler angle form) of the probe together with their first derivatives (translational and angular velocity). The state transition matrix is defined as: $\mathbf{A} = \left(\begin{smallmatrix} \mathbf{A}_T & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_T \end{smallmatrix}\right)$ in which $\mathbf{A}_T = \left(\begin{smallmatrix} \mathbf{I}_3 & \Delta t \cdot \mathbf{I}_3 \\ \mathbf{0} & \mathbf{I}_3 \end{smallmatrix}\right)$. The time interval between samples is denoted as Δt.

The relationship between the measurements from sensors and the states is defined as:

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{v}(k) \tag{5.5}$$

where $\mathbf{y}(k) \in \mathbb{R}^l$ is the measurement vector, and $\mathbf{v}(k) \in \mathbb{R}^l$ is a sequence of observation noise which is assumed to be zero mean Gaussian white noise with covariance $\mathbf{R}(k)$. Since two sensors are available, the measurement matrix C is defined separately for camera and pCLE:

$$\begin{cases} \mathbf{C}_c = \left(\begin{smallmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \end{smallmatrix}\right), \\ \mathbf{C}_p = \left(\begin{smallmatrix} \mathbf{I}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{smallmatrix}\right). \end{cases} \tag{5.6}$$

The error covariance of states is updated in the prediction stage as:

$$\mathbf{P}(k|k-1) = \mathbf{A}\mathbf{P}(k-1|k-1)\mathbf{A}^\intercal + \mathbf{Q}(k-1). \tag{5.7}$$

The Kalman gain for the camera and pCLE is given by:

$$\begin{cases} \mathbf{K}_c(k) = \mathbf{P}(k|k-1)\mathbf{C}_c^\intercal(\mathbf{C}_c\mathbf{P}(k|k-1)\mathbf{C}_c^\intercal + \mathbf{R}_c(k))^{-1}, \\ \mathbf{K}_p(k) = \mathbf{P}(k|k-1)\mathbf{C}_p^\intercal(\mathbf{C}_p\mathbf{P}(k|k-1)\mathbf{C}_p^\intercal + \mathbf{R}_p(k))^{-1}, \end{cases} \tag{5.8}$$

The measurement residual is given by:

$$\mathbf{r}(k) = \mathbf{y}(k) - \mathbf{C}\mathbf{x}(k-1). \tag{5.9}$$

The correction is then performed using:

$$\begin{cases} \mathbf{x}(k) = \mathbf{x}(k|k-1) + \frac{1}{2}(\mathbf{K}_c(k)\mathbf{r}_c(k) + \mathbf{K}_p(k)\mathbf{r}_p(k)), \\ \mathbf{P}(k) = (\mathbf{I} - \frac{1}{2}(\mathbf{K}_c(k)\mathbf{C}_c + \mathbf{K}_p(k)\mathbf{C}_p))\mathbf{P}(k|k-1). \end{cases} \tag{5.10}$$

During the execution of a global trajectory to the scanning region, there is no pCLE image information (the probe is not in contact with the tissue) and thus only the camera data is fed into the Kalman filter. When the probe arrives at the scanning region, the states of the Kalman filter are reinitialised as the servoing is switched from the camera to the mosaic image coordinates.

5.5. Closed-Loop Scanning with Visual Servoing

The proposed framework computes the hand-tool transformation $\mathbf{T}_E^M$ between the end-effector and marker on the adaptor via a standard hand-eye calibration approach [142]. The advantage of using the hand-tool transformation is that it will not be affected when the robot and camera are repositioned, making it more practical during surgical procedures. Although $\mathbf{T}_E^M$ would change slightly if the adapter is released and re-grasped, it is found that the resulting small error (a few millimetres) in the hand-tool transformation is acceptable for the visual servoing control.

Given a current estimated probe pose $\mathbf{T}_C^D$ and a desired pose $\mathbf{T}_C^{D^\star}$ in the camera frame (in the planned trajectory), a new command pose can be calculated for the robot to minimise the difference. The relative probe pose, denoted as $\mathbf{T}_D^{D^\star}$, is given by: $\mathbf{T}_D^{D^\star} = \left(\mathbf{T}_C^D\right)^{-1} \cdot \mathbf{T}_C^{D^\star}$, where $\mathbf{T}_C^D$ and $\mathbf{T}_C^{D^\star}$ are obtained from the Kalman filter and the planned trajectory, respectively. Since the local trajectory is in the mosaic image coordinates, a transformation $\mathbf{T}_D^I = \begin{pmatrix} \mathbf{R}_D^I & \mathbf{0} \\ 0 & 1 \end{pmatrix}$ is required to convert the trajectory from 2D to 3D. $\mathbf{R}_D^I$ is essentially a 2D rotation between the probe and the mosaic, which can be calibrated by driving the robot along a line.

With this, one can calculate the robot command $\mathbf{T}_B^E$ that minimises the relative probe pose:

$$\mathbf{T}_B^{E^\star} = \mathbf{T}_B^E \cdot \mathbf{T}_E^M \cdot \mathbf{T}_M^D \cdot \mathbf{T}_D^{D^\star} \cdot \left(\mathbf{T}_{M^\star}^{D^\star}\right)^{-1} \cdot \left(\mathbf{T}_{E^\star}^{M^\star}\right)^{-1} \tag{5.11}$$

The vision-based correction is performed on every pose along the planned trajectory. To produce a smooth scanning motion, the trajectory between two adjacent poses is interpolated and executed using the robot kinematic model.

5.6. Endoscopic 3D Fusion and Visualisation

Using the 3D reconstruction of the scanning region and the probe pose in camera coordinates, the registration of the mosaic images with the tissue surface can be done. This enables us to create a multi-scale map that links macro- and micro-views, which are obtained from the stereo camera and the pCLE, respectively. For pCLE mosaicing, an approach similar to standard real-time techniques discussed in [160] is utilised, using normalised cross-correlation to estimate the relative shift between each pair of consecutive frames. To avoid overwriting information from previous frames, and minimise image boundary effects,

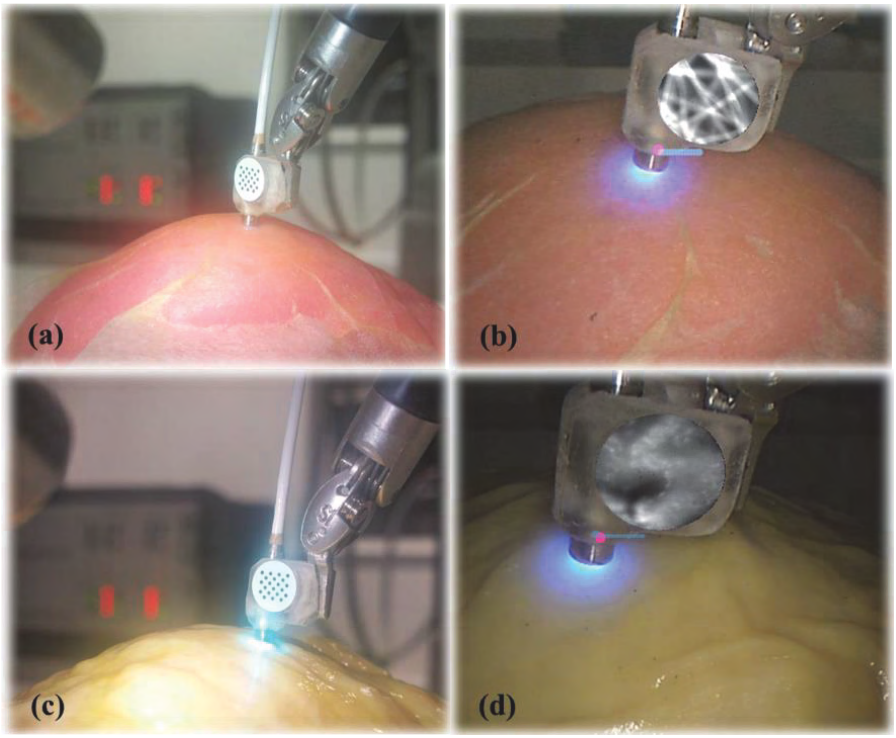


Figure 5.4.: Snapshots of the phantom and *ex vivo* experiments. (a) and (c) Experimental setups; (b) and (d) A snapshot of the surgical view captured by the laparoscope. The KeyDot® marker was overlaid with the streaming microsopic images for AR.

blending is applied. The 2D mosaic is then converted to a 3D mosaic point cloud where each point represents a pixel in the mosaic.

In order to register the mosaic point cloud with the surface, $\mathbf{T}_I^C$ is required. As illustrated in Fig. 5.2 it can be calculated as: $\mathbf{T}_C^I = \mathbf{T}_C^{M_0} \cdot \mathbf{T}_{M_0}^D \cdot \mathbf{T}_D^I$, where M_0 represents the coordinate frame of the marker at the beginning of the local scan. With $\mathbf{T}_I^C$ for each scanning region, the proposed method is able to register mosaics from multiple region scans onto the same 3D surface. The 3D mosaic point cloud is updated on-the-fly when the microscopic images are mosaicked. An illustration of this multi-scale 3D fusion can be seen in Fig. 5.8.

This chapter describes three ways to visualise the scanning and mo-

Table 5.1.: Stereo Reconstruction Accuracy

		ELAS [156]	**COCV** [161]
Cardiac1	MAE (mm)	0.89 ± 0.70	1.24± 0.89
	RMS (mm)	1.31 ± 0.98	1.85 ± 0.82
Cardiac2	MAE (mm)	1.21 ± 1.56	1.47 ± 1.23
	RMS (mm)	1.77 ± 2.16	2.66 ± 1.47

saicing on-the-fly: (1) an augmented view from the camera that allows both the robot and target tissue to be seen by overlaying the endomicroscope images onto the tracking marker on the adaptor, as shown in Fig. 5.4(b) and (d); (2) a real-time, continuously updated 2D mosaic map; (3) an online 3D fusion of mosaics onto the reconstructed surface. These methods provide a clear and intuitive visualisation for the surgeon, aimed at assisting surgical planning.

5.7. Experiments and Results

5.7.1. Stereo Validation

For assessing the accuracy of ELAS stereo reconstruction, a comparison of the method with the state-of-the-art [161] is provided using the Hamlyn cardiac datasets (`http://hamlyn.doc.ic.ac.uk/vision/`). The results are provided in Table 5.1, and present that the ELAS method provides competitive accuracies.

5.7.2. Hand-eye calibration accuracy

Due to the kinematic error of a tendon driven manipulator such as the da Vinci® robot, the hand-eye calibration would only be accurate within a space where the calibration data has been taken. Given the calibrated hand-tool transformation $\mathbf{T}_E^M$ and an initial marker pose $\mathbf{T}_C^{M_0}$, the proposed method can predict a marker pose using robot for-

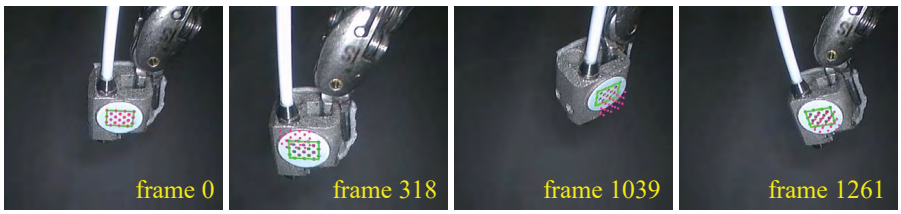


Figure 5.5.: As the marker moves away from the starting reference pose, the predicted pattern (projection on image is shown in red dots) deviates from the detected one.

Table 5.2.: Error of Hand-eye calibration

	Translation (mm)	Rotation (deg)
MAE	1.33 ± 0.78	0.64 ± 0.37
RMSE	1.54	0.74

ward kinematics:

$$\hat{\mathbf{T}}_C^M = \mathbf{T}_C^{M_0} \cdot \mathbf{T}_{M_0}^{E_0} \cdot \mathbf{T}_{E_0}^B \cdot \mathbf{T}_B^E \cdot \mathbf{T}_E^M. \tag{5.12}$$

The ground truth of the marker pose $\mathbf{T}_C^M$ was obtained from manual annotation. The error between the ground truth and the pose estimated by the hand-eye model is calculated by using Eq. 5.12.

The error pose is computed as: $\Delta\mathbf{T}_{\hat{M}}^M = \left(\hat{\mathbf{T}}_C^M\right)^{-1} \cdot \mathbf{T}_C^M$. Starting from the initial pose $\mathbf{T}_C^{M_0}$, as the marker moves within the workspace for the scanning task, two values are calculated: translational error and rotational error. The translational error is the Euclidean norm of the translational part of $\Delta\mathbf{T}_{\hat{M}}^M$ while the rotational error is the mean of the Euler angle of the rotational part. As presented in Table 5.2, the translational and rotational errors are approximately 1.5 mm and 0.7 degrees respectively, which is sufficient to support sub-millimetre-scale accuracy for the visual servoing.

5.7.3. Accuracy of Visual Servoing

To quantitatively validate the camera visual servoing approach, the experiments used the Optotrak Certus system (Northern Digital Inc, Canada) which achieves 0.1 mm accuracy. Optical sensors were attached to a custom probe adapter, so that a rigid transformation could be found between the optical sensors and the marker frame. Ten trials were run, with each consisting of two runs: trajectory-following with Visual Servoing (VS), and Kinematic Motion (KM) only. The planned trajectories, which are defined in the camera coordinates, were used as the ground truth, and the optically-tracked results of two runs were transformed into the camera coordinate for comparison. To this end, it is assumed that the starting pose of the two runs should align with the first pose in the ground truth trajectory. For each pair of measured and ground truth poses at the same time instance, the pose difference is considered as the error. For translational errors, the Euclidean distance is computed, and for rotational errors, the average of the Euler angles from the x, y and z axes are used. The measured translational errors and rotational errors are summarised in Table 5.3. Both translational and rotational accuracies using the camera visual servoing are improved over kinematic-only control. The motion using the camera visual servoing was able to achieve on average approximately 0.21 mm for the translational error and 1.23 degrees for the rotational error.

Since the camera-based visual servoing is used to move the probe to an initial position for scanning, the repeatability can be qualitatively evaluated by comparing the differences among multiple mosaics acquired at the same nominal position. Three trials were run, in which the probe was moved to the same position before scanning and mosaicing over a printed grid phantom. As shown in Fig. 5.6, the mosaic images from different trials (identified by the colour) are well-aligned at their centres, both horizontally and vertically, and the camera visual servoing provides good repeatability with only about 50 to 150 μm deviation. This shows good correspondence with the quantitative

Table 5.3.: Camera Visual Servoing Error (Mean ± Std)

	Translation Error (mm)		Rotation Error (degrees)	
Trial	**VS**	**KM**	**VS**	**KM**
Trial 1	0.238 ± 0.06	0.863 ± 0.26	0.89 ± 0.55	4.51 ± 1.45
Trial 2	0.222 ± 0.08	0.360 ± 0.21	1.21 ± 0.53	3.57 ± 0.79
Trial 3	0.189 ± 0.08	0.507 ± 0.16	1.19 ± 0.55	1.96 ± 1.00
Trial 4	0.185 ± 0.07	0.588 ± 0.37	1.54 ± 0.46	5.54 ± 1.27
Trial 5	0.172 ± 0.06	0.462 ± 0.27	1.44 ± 0.55	5.62 ± 1.11
Trial 6	0.226 ± 0.07	0.915 ± 0.40	1.57 ± 0.49	5.22 ± 1.08
Trial 7	0.247 ± 0.06	0.909 ± 0.37	1.11 ± 0.48	4.05 ± 0.87
Trial 8	0.303 ± 0.09	0.542 ± 0.17	1.25 ± 0.50	0.79 ± 0.52
Trial 9	0.165 ± 0.10	0.484 ± 0.13	0.84 ± 0.45	3.39 ± 0.80
Trial 10	0.164 ± 0.07	0.666 ± 0.13	1.21 ± 0.57	1.44 ± 1.04
Total	**0.211 ± 0.07**	**0.630 ± 0.25**	**1.226 ± 0.51**	**3.61 ± 0.99**

accuracy estimated above (0.211 mm).

The mosaic images in Fig. 5.6 also provide us with a measure of the accuracy of the camera and endomicroscope combined servoing. This is performed by comparing the original layout of the grid pattern with the obtained mosaicked grid, and in our experiment, the error is approximately 60 μm, including the uncertainty introduced by the laser printing of the pattern.

5.7.4. Phantom and *Ex Vivo* Experiment

The experiment steps can be summarised as follow:

1. A stereo reconstruction of current surgical scene is generated.
2. A user places the probe at a standby location.
3. A user selects a target scanning region with the global and local trajectories being generated accordingly.
4. The robot moves the probe along the global trajectory with servoing from the laparoscopic camera images.
5. Probe arrives at the starting position for the mosaic scan over the desired region; robot drives probe to follow local trajectory using both microscopic and laparoscopic images for servoing.

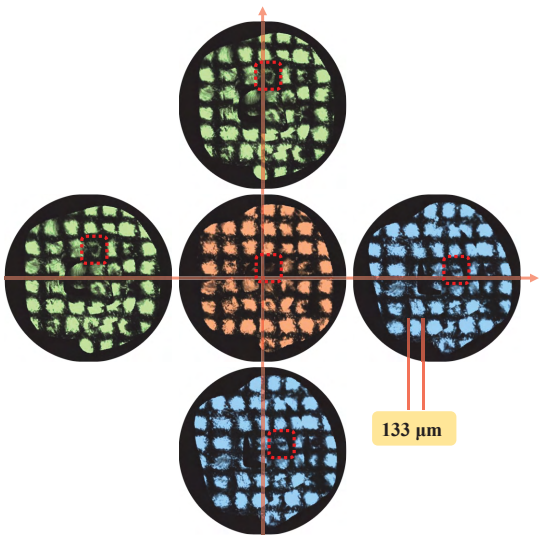


Figure 5.6.: Demonstration of mosaic results on a printed grid phantom with black holes (outlined by the dashed rectangle). Three trials are identified by three different colours and a red guide axis is placed for assisting comparison.

6. After finishing the scan, the probe returns to the standby location, ready to scan the next region.

To validate the approach, autonomous scans were performed on a custom-made PVA cryogel kidney phantom. In order to produce high quality microscopic images, lens tissue paper stained with the topical fluorescent contrast agent acriflavine, was placed on top of the surface. A snapshot of the experiment is provided in Fig. 5.4(a-b). Scanning tasks were then performed on the phantom surface using both spiral and raster trajectories. An example of the 2D mosaics obtained is shown in Fig. 5.7. The actual trajectory line drawn on the mosaic image shows that the robot suffers from the raster motion due to backlash, as the raster pattern consists of abrupt changes of motion direction. The spiral pattern is considered to be superior as the motion direction changes gradually.

The 3D fusion results are presented in Fig. 5.8 which demonstrates that our approach is able to perform multi-region scanning. These

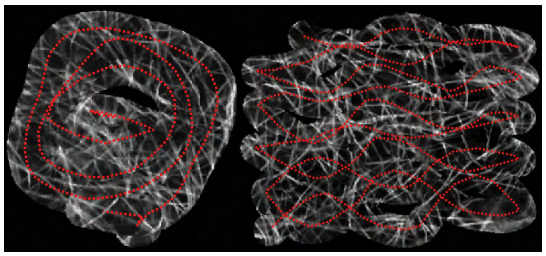


Figure 5.7.: Comparison of spiral and raster trajectories. The actual trajectories measured on the mosaic images are shown in red dotted line.

results have shown our proposed framework is able to provide continuous image mosaicing, benefiting from the smooth and accurate motion generated by the visual servoing approach. Experiments were also performed using *ex vivo* animal tissue (see Fig. 5.4(c-d)). The gastric mucosa of a porcine stomach tissue sample was again stained with the topical acriflavine contrast agent. Examples of the 3D fusion results are provided in Fig. 5.8. Both the phantom and *ex vivo* experiments have demonstrated that our framework is capable of performing autonomous pCLE scanning and generating high quality image mosaics. Again, due to the advantages of global surface reconstruction and visual servo control, the proposed framework is able to scan multiple regions on the same surface. This also provides a global-local mapping where the user can not only see each scanning location from macroscale but also see a zoomed-in mosaic image.

5.8. Discussion and Conclusion

The proposed framework presents good repeatability and accuracy, with all trials demonstrating consistently low errors, sufficient to allow pCLE scanning tasks to be completed. The mosaicing results from both phantom and *ex vivo* experiments, shown in Fig. 5.8, demonstrate that the system can scan multiple regions and obtaining good quality mosaics. The results are a significant improvement over the

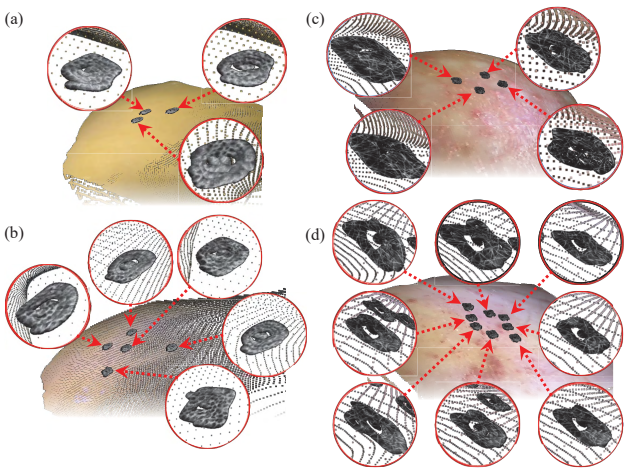


Figure 5.8.: Exemplar 3D fusion results. (a) and (b) online multi-region fusion results from the *ex vivo* gastric mucosa experiment; (c) and (d) Online multi-region fusion results from phantom experiment.

basic linear mosaics produced by manually operated pCLE, suggesting enhanced clinical value for intraoperative pathological analysis and surgical planning.

Whilst one might expect the accuracy of stereo reconstruction to be insufficient for pCLE scanning, it was found that most of scanning tasks could be completed successfully. However, it was observed during the stereo validation that the recovered surface tends to have a constant offset despite good shape matching to the ground truth. To address this erroneous offset, the probe tip to marker transformation $\mathbf{T}_D^M$ was adjusted to ensure continuous contact with the surface. This solution is not ideal and would cause problems when the reconstruction error becomes too large ($> 2.0mm$), as it would cause the probe to either lose contact with the surface or apply too much pressure, leading to excessive deformation. It is also observed that the image quality tends to decrease as the contact force increases. A comprehensive solution to this problem, would be to perform additional visual servoing based on the quality of microscopic images. With this, the image quality could

then be used to detect loss of contact or excessive pressure, and adjust the robot position accordingly [162].

In this chapter, an autonomous framework using the da Vinci® robot is proposed for endomicroscopic mosaicing and online fusion of 3D reconstructed macroscopic images with microscopic images. Both laparoscopic camera and pCLE images have been utilised to achieve continuous and smooth scanning. In addition, the proposed framework provides a comprehensive visualisation scheme for users, which includes: (1) an overlay of microscopic images into the surgical view for AR; (2) an online updating 2D mosaic image; and (3) 3D fusion of the mosaic images with the tissue surface on-the-fly. These visualisation options have a range of potential benefits for intraoperative tissue analysis and surgical planning, which will be explored in future work. The proposed framework has been tested on phantoms and *ex vivo* tissue, and quantitative results obtained from an external tracking system have shown that our framework can achieve accuracy better than the FoV of the endomicroscope, thus permitting smooth and accurate microscopic scanning.

6. Autonomous Endomicroscopic Scanning and Fusion with Depth Information †

6.1. Introduction

Recent studies [146] have suggested that automation of some surgical subtasks, particularly those that are tedious and repetitive or require high precision, can be beneficial in improving accuracy and reducing the cognitive load of the surgeon. For example, several studies have investigated automation of surgical suturing subtasks, including using a suturing tool under fluorescence guidance [163], and other studies have explored areas such as autonomous tissue dissection [148]. In Chapter 5, an surgical automation framework has been presented for pCLE scanning tasks that required high accuracy and smooth motion.

As mentioned in Chapter 5, individual pCLE images need to be sticked together in order to form large mosaics which are more useful for surgical diagnosis. Several mosaicing algorithms have been developed for this purpose, including both real-time [164] and more robust offline processing [165] approaches. However, the operator is required

†Content from this chapter was published as:
From Macro to Micro: Autonomous Multiscale Image Fusion for Robotic Surgery. Zhang, L., Ye, M., Giataganas, P., Hughes, M., Bradu, A., Podoleanu, A., Yang, G.-Z.. IEEE Robotics and Automation Magazine. 24(2), 63-72 (2017).

to maintain optimal probe orientation while performing smooth and controlled scanning motions with sub-millimetre accuracy. For pCLE in particular, which requires direct tissue contact, the probe-tissue force is critical [154], both to ensure good quality images and to limit the amount of tissue deformation occurring during scanning. Mosaicing under manual control is therefore extremely challenging.

To address this issue, several studies have attempted to robotise pCLE probes with the aim of achieving consistent imaging over large areas of tissue. These developments have included dedicated robotic mechanisms for 2D scanning [50, 153, 166] and for maintaining the desired probe-tissue contact force [154]. Integration with existing robotic systems, such as the da Vinci® [154, 25], has also been investigated. However, even with robotic assistance, limited depth perception and poor ergonomics mean that it can still be difficult for clinicians to perform a continuous and smooth scan during tele-operated surgery; maintaining the probe in a specific orientation and ensuring continuous contact with the surface remains difficult and tedious.

OCT differs from pCLE in that it provides cross-sectional images of the tissue's structure. The nominal depth range is typically several millimetres, although penetration into scattering tissue is limited to 1-2 mm in practice. OCT has achieved significant clinical success in ophthalmology, and endoscopic OCT has extended the range of applications to areas such as the gastrointestinal tract [167]. It has also been used in laparoscopic and robotic prostatectomy [40] for characterising neurovascular bundles along the prostate in real-time. As for pCLE, it is necessary to mosaic OCT images if a large FoV is required, and this shares many of the difficulties with pCLE mosaicing. However, since OCT is a cross-sectional technique that does not require tissue contact, it can also be used to determine the distance between the probe tip and the tissue. Several studies have suggested the potential benefits of combining cross-sectional OCT images and fluorescence microscopy (e.g., [168]), although most of the designs for combined probes require sacrificing the resolution of the fluorescence channel. Here, by mount-

ing the two probes side-by-side, the proposed framework combines the advantages of both techniques to provide complementary diagnostic information – high-resolution surface information and lower resolution sub-surface information.

The autonomous scanning system presented in Chapter 5 has the potential to support the use of optical biopsy in surgical operations. In this chapter the scanning motion is visually servoed not only using pCLE images (in the direction transverse to the tissue surface) but also by OCT images (in the axial direction). With this approach, the robot is able to ensure that optimal contact force is maintained between the pCLE probe and the tissue, assisting with smooth motion (avoiding stick-slip) and helping to form continuous mosaics.

Furthermore, to facilitate intraoperative tissue diagnosis and identification, a 3D visualisation method is proposed to fuse the reconstruction of the 3D tissue surface with 2D endomicroscopy mosaics and 3D OCT volumes on-the-fly, an improvement over the work in Chapter 5. This 3D fusion approach is designed to provide the surgeon with intuitive real-time visualisation of multi-scale image information, supporting surgical diagnosis and decision-making. The framework has been validated with a series of phantom and *ex vivo* tissue experiments, with the results demonstrating the potential clinical value of the approach.

6.2. Overview of the Scanning Framework and Workflow

The system links custom pCLE and OCT systems with a stereo laparoscope and a PSM of a da Vinci® surgical robot with dVRK controllers. The dVRK controllers (da Vinci Research Kit, Medical Motion) allow the conventional master console of the da Vinci® to be bypassed and replaced by the autonomous control system. They are connected to a host PC via an IEEE 1394 firewire interface in a daisy chain topology.

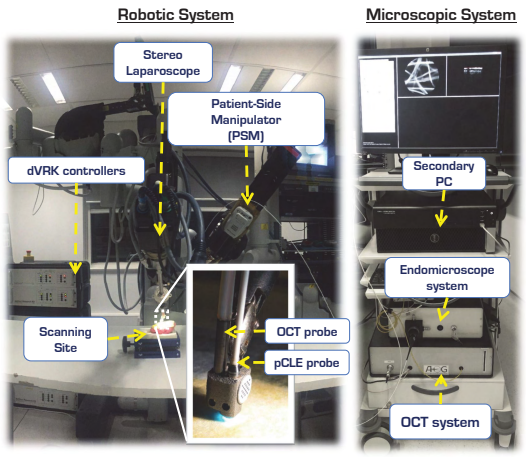


Figure 6.1.: An overview of our experimental setup for autonomous optical biopsy probe scanning and multi-scale fusion.

The stereoscopic system provides standard-definition (720x576) video streaming (for both left and right channels) at 25 Hz, which is captured by the host PC using a Kona 4 PCIe frame grabber (AJA Video System). An example of the system setup can be seen in Fig. 6.1. The robotic system consists of a set of dVRK controllers and both the pCLE and OCT probes are grasped by a da Vinci® PSM. The microscopic system consists of an CLE system, an OCT system and a PC used to capture and process pCLE and OCT images.

The pCLE and OCT systems are developed with in-house designs, both consisting of a flexible fibre optic probe connected to an external base unit. The distal tips of the fibre probes are held by a pick-up mount which can be grasped by da Vinci® instruments. The pCLE system is a high-speed line-scanning endomicroscope [155] coupled to a Cellvizio UHD fibre probe (Mauna Kea Technologies). It excites tissue at a wavelength of 488 nm and collects fluorescence emission above 500 nm. It is used with typical fluorescent stains which can be excited at 488 nm, such as acriflavine. The probe, which consists of a 30,000-core fibre imaging bundle and a micro-lens, provides a FoV of 240 μm and a fibre-sampling limited resolution of approximately 2.4

μm. Since each fibre core within the bundle transmits a 'pixel' of the image, no scanning mechanism is needed at the distal tip of the probe; it is entirely passive. The external line-scanning system illuminates and images a single line of the tissue (via the bundle) at a time, a technique which ensures light is predominantly collected only from an in-focus plane approximately 9 μm in depth. The line-scan design (rather than point-scanning as in conventional confocal microscopes) sacrifices some axial resolution in exchange for a maximum frame rate of 120 fps, which is beneficial for this application as it allows for higher speed scanning.

The OCT system uses a swept source laser (Axsun Technologies, Billerica, MA) with a central wavelength of 1300 nm. The laser sweeps over a wavelength of 110 nm at a frequency of 100 kHz, with each sweep allowing acquisition of a single axial line (an A scan) through the tissue by the technique of spectral domain low coherence interferometry. The forward viewing fibre probe is 13 mm long (fabricated at the Institute of Applied Physics, Russia) and has a 2.7 mm diameter rigid tip. This tip incorporates copper wires and a NdFeB magnet, providing a mechanism for scanning the fibre at 40 Hz and hence generating 2D images. A GRIN micro-lens focuses the beam onto the tissue, and the assembly is covered with Teflon tubing. The FoV of the probe is approximately 1 mm.

Both optical biopsy systems are controlled via a single software interface, developed in LabVIEW (National Instruments), running on a secondary PC. Images are acquired and processed in real time, resized to 300x300 (pCLE) and 90x300 pixels (OCT), corresponding to real dimensions of 240 x 240 μm and 900 x 3000 μm respectively, and are streamed to the host PC at 40 Hz via a TCP/IP connection. The software component of the framework is implemented using the ROS across two computers, with the workload of robot motion control separated from imaging processing and 3D visualisation. The ROS uses TCP/IP to communicate with the pCLE/OCT PC and the servoing PC; this has a round trip time of about 5 ms. The 2D mosaic process-

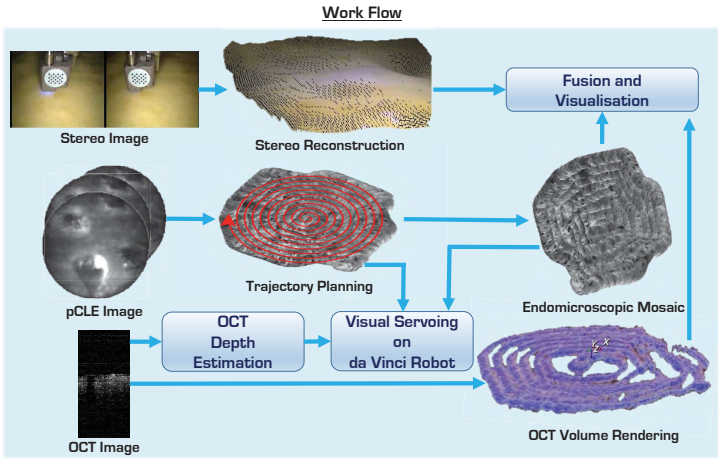


Figure 6.2.: The data flow streaming from the different imaging modalities is processed for visualisation and servoing purposes. From a pair of stereo images, a surface of the scene is reconstructed as a point cloud. By stitching pCLE images, a mosaic image can be created, and a 3D volume can be built from OCT images. These results are fused into a unified window for multi-scale visualisation.

ing runs at 10 Hz (100 ms). Therefore, for lateral motion compensation using pCLE images, the delay is approximately 110 ms. For vertical motion compensation using OCT distance estimation, the delay is approximately 40 ms.

As shown in Fig. 6.2, the stereo images captured by the camera are used for 3D tissue surface reconstruction via a stereo matching method. Images captured by the pCLE system are registered pairwise by normalised cross-correlation and combined by the dead-leaf method. Concurrently, OCT images are mapped into a 3D volume based on the current probe pose, providing a 3D volumetric reconstruction over the scanned tissue. The pCLE mosaic, OCT volume and reconstructed surface are fused together on-the-fly to provide both macro- and micro-views of the scanned region.

To generate a mosaic over a defined area, the probe is scanned over a pre-defined trajectory using the robot, with the trajectory planned

in the coordinate system of the mosaic. A visual control component closes the loop by comparing the current and desired probe poses and driving the robot to minimise their difference. This results in the desired region of tissue being imaged regardless of kinematic errors or tissue motion and deformation.

Separate from the visual servoing loop, the end-effector pose of the robot (in Cartesian space) is then read and set via a dVRK-ROS component which is connected to a low-level PID controller implemented by the SAW package using the cisst library [132].

The scanning trajectory is planned on a 2D plane with the aim of generating a mosaic over the desired area of tissue. A spiral trajectory is preferable to a raster pattern as it avoids sudden changes in direction and hence exerts less deformation on the tissue. To generate a desired spiral trajectory, the distance between two successive spiral loops (Δr) is pre-defined. This defines the constant b:

$$b = \frac{\Delta r}{2\pi}. \tag{6.1}$$

In order to reduce the likelihood of producing gaps in the mosaic image, whilst preventing excessive over-sampling, Δr is set to half of the endomicroscope's FoV (240 μm). The radius of the spiral trajectory r_{sp} represents the size of the scanned region, and the total length of the spiral trajectory l_{sp} is calculated as:

$$l_{sp} = \frac{b}{2} \cdot (\frac{r_{sp}}{b})^2. \tag{6.2}$$

Finally, the k-th trajectory point is defined as:

$$\begin{cases} x_k = b \cdot \theta_k \cdot \cos\theta_k \\ y_k = b \cdot \theta_k \cdot \sin\theta_k, \end{cases} \tag{6.3}$$

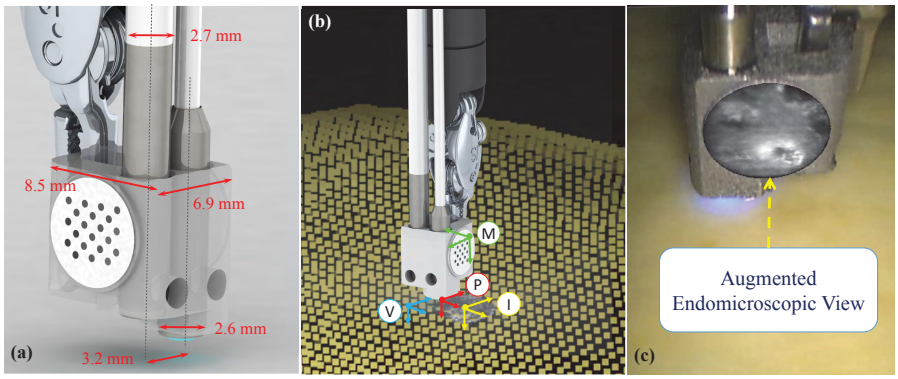


Figure 6.3.: Illustration of the coordinate frames and dimension of the hardware; (a) shows a da Vinci® tool grasping the pick-up adapter holding the pCLE and OCT probes. The dimension of the adapter and probes is labelled; (b) shows the definition of the coordinate frames used in the visual servoing and fusion. The marker and pCLE probe coordinate frames are denoted as $\{\mathbf{M}\}$ and $\{\mathbf{P}\}$ respectively. The volume and mosaic image coordinate frames are denoted as $\{\mathbf{V}\}$ and $\{\mathbf{I}\}$ respectively; (c) shows an augmented laparoscopic view where a pCLE image is overlaid on the adaptor.

where θ_k is obtained using equation 6.1 and 6.2:

$$\theta_k = \sqrt{\frac{2l_k}{b}}, l_k \in [0, l_{sp}]\,. \tag{6.4}$$

6.3. Closed-Loop Scanning using pCLE and OCT images

At each iteration of the control loop, the desired robot end-effector command is defined as:

$$\mathbf{T}_B^{E*} = \mathbf{T}_B^E \cdot \mathbf{T}_E^M \cdot \mathbf{T}_M^P \cdot \mathbf{T}_P^{P*} \cdot \mathbf{T}_{P*}^{M*} \cdot \mathbf{T}_{M*}^{E*}. \tag{6.5}$$

The definition of the different coordinate systems is illustrated in Fig. 6.3. $\mathbf{T}_B^E$ is the end-effector pose in robot base coordinates, calculated using forward kinematics. The transformation $\mathbf{T}_P^{P*}$ is the transformation between the current and desired probe positions. $\mathbf{T}_P^{P*}$ is calculated in every iteration during closed-loop control using the displacement between the current and desired probe positions. $\mathbf{T}_P^{P*}$ is defined as:

$$\mathbf{T}_P^{P*} = \begin{bmatrix} \mathbf{I} & \Delta\mathbf{t} \\ 0 & 1 \end{bmatrix}, \tag{6.6}$$

where $\Delta\mathbf{t} = \left[\Delta t_x, \Delta t_y, \Delta t_z\right]^\intercal$ is a displacement vector and $\mathbf{I}$ is a 3×3 identity matrix. In order to obtain the displacement vector, the information from either or both of the pCLE and OCT images is used. The rest of the transformations in equation 6.5 are constants that can be measured or calibrated in advance. In particular, the end-effector to marker transformation $\mathbf{T}_E^M$ is calibrated using a standard hand-eye calibration method [142]. The marker to probe transformation $\mathbf{T}_M^P$ is determined from the CAD model of the adapter. It should also be noted that $\mathbf{T}_{M*}^{E*} = (\mathbf{T}_E^M)^{-1}$ and $\mathbf{T}_{P*}^{M*} = (\mathbf{T}_M^P)^{-1}$. Continuous detection of the marker is not required during local scanning with closed-loop servoing, as it is only used to determine the transformation between the robot end-effector and probe, which does not change in time with a rigid setup. Note that the scanning surface is assumed to be a plane as each individual scanning region is usually small (about 2x2mm).

6.3.1. Servoing using pCLE Images

To allow a continuous mosaic to be generated over the desired area of tissue, the visual servoing loop is closed on the mosaic image. The pCLE mosaicing is performed using an approach similar to the standard real-time technique described in [164, 155], using normalised cross-correlation to estimate the relative shift between each pair of consecutive frames. At the beginning of each scanning procedure, the probe position in the mosaic image ${}^m p(t = 0) = (0, 0)$ is located at the centre

of the image. During scanning, an estimate of the current probe position ${}^{m}p(t)$ at time t is obtained from the mosaic (i.e. from integrated pair-wise image shifts over time). Next, by comparing ${}^{m}p(t)$ with the k-th desired probe position in the trajectory ${}^{m}p^{*}(k)$, the displacement between the current and desired position is calculated as:

$$\begin{bmatrix} \Delta t_x \\ \Delta t_y \end{bmatrix} = k_m \cdot \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \cdot \begin{bmatrix} {}^{m}p_x^{*} - {}^{m}p_x \\ {}^{m}p_y^{*} - {}^{m}p_y \end{bmatrix}, \tag{6.7}$$

where α is an 2D rotation angle between the probe and the mosaic image, and k_m is a constant that converts the displacement from pixels to real distance. To calibrate the rotation angle α, the probe is moved by the robot in a horizontal line scan using the laparoscope. The angle α between the scanned line and the desired horizontal line can be calculated.

6.3.2. Servoing using OCT Images

The distance from the OCT probe to the tissue can be estimated by detecting the top surface of the tissue in the OCT cross-sectional image, which can be seen in Fig. 6.1. The OCT probe is mounted slightly higher than the pCLE probe, so that the top surface of the tissue appears at a non-zero depth in the OCT image. Since the top surface usually maintains the most intense signal, this can be found by simple peak detection, taking the first peak above a user-defined threshold. To mitigate the influence of noise, a Kalman filter is applied to assist accurate distance estimation for servoing. Here, the proposed framework utilises a constant velocity model for the Kalman filter as the motion of robot along the depth direction is smooth. At the beginning of each scan, it is assumed that a good initial contact has been made by the user (i.e. clear pCLE images can be seen), and the current OCT distance estimate d_{oct} is recorded. This distance is then set as the desired distance d_{oct}^{*} and the robot is thus required to maintain this distance during scanning. The displacement along the z-axis is defined

as:

$$\Delta t_z = k_o \cdot (d^*_{oct} - d_{oct}). \tag{6.8}$$

Here k_o is a constant that converts the distance from pixels to real distance.

6.4. Multiscale Fusion and Visualisation

In this work, three different imaging modalities are fused together in a unified visualisation framework including both macro- and micro-views. For the macro-view, a pair of stereo images is used to reconstruct a 3D surface of the scanning region using an efficient stereo matching method [156]. The micro-view then consists of a mosaic image obtained from pCLE images overlaid onto a 3D volume reconstructed from OCT images.

To place the mosaic image with the OCT volume, the proposed framework takes account of the known lateral offset between the OCT and pCLE probes due to the design of the pick-up mount, as shown in Fig. 6.3. The size of the volume is set according to the defined size of the scanning region such that it will be large enough to contain all OCT images during scanning. The size of each voxel in the volume is set equal to the size of the OCT image pixels, such that each voxel represents approximately 10 μm in the plane of the cross-section, and 10 μm along the out-of-plane direction. A 2D-3D mapping allows each pixel in each OCT image to be mapped to a voxel in the volume. For each scan, the volume is initialised such that all voxels are set to invisible. The pose of each OCT image in the volume coordinate frame is then obtained from the current position of the probe in the mosaic image coordinate frame, taking into account the known offset between the two probes. The value of the voxel is simply set equal to the intensity of the corresponding pixel in the OCT image. If the voxel has already been set using a previous OCT image, its value is updated by averaging the new and current values.

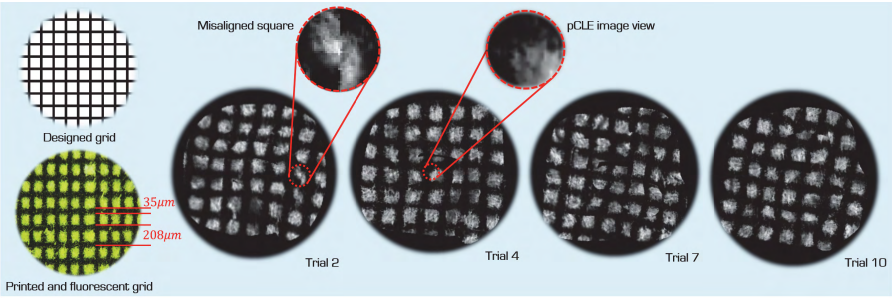


Figure 6.4.: Resulting mosaics from pCLE servoing on a printed grid phantom. Results from trial 2, 4, 7 and 10 are shown here. In all trials, the mosaic image shows the grid pattern clearly. An example of misaligned square due to the inaccuracy in the visual servoing is indicated in trial 2.

The mosaic image is then scaled according to its real size and placed in the same visualisation framework as the OCT volume. To register the OCT volume and pCLE mosaics with the macro surface reconstruction, the proposed framework makes use of a circular-dot marker attached to the adaptor Fig. 6.3(a). The definition of different coordinate frames is defined in Fig. 6.3(b). When a scan is started, the initial pose of the marker in camera coordinates $\mathbf{T}_C^M$ is recorded. Since there is a known transformation between the marker and the probe $\mathbf{T}_M^P$ and a known offset $\mathbf{T}_V^I$ between the two probes, the registration matrices $\mathbf{T}_C^I$ and $\mathbf{T}_C^V$ can be calculated as:

$$\begin{cases} \mathbf{T}_C^I = \mathbf{T}_C^M \cdot \mathbf{T}_M^P \cdot \mathbf{T}_P^I \\ \mathbf{T}_C^V = \mathbf{T}_C^I \cdot \mathbf{T}_I^V. \end{cases} \tag{6.9}$$

It should be noted that $\mathbf{T}_P^I = \mathbf{I}$ as the probe is located at the centre of the mosaic image when the scan starts. With the transformation $\mathbf{T}_C^I$ and $\mathbf{T}_C^V$, the mosaic image and volume can be registered with the surface reconstruction.

6.5. Experiments and Results

6.5.1. Accuracy of pCLE Visual Servoing

To evaluate the accuracy and consistency of the pCLE visual servoing, a test phantom containing a known grid pattern is manufactured. The phantom was printed by a laser printer on a sheet of paper and coated by a fluorescent marker, making it visible in the pCLE channel. Every square in the grid pattern had a line thickness of 35 μm and width of 208 μm, as shown in Fig. 6.4. To avoid the effect of printing inaccuracies, which can be seen in Fig. 6.4 where the printed grid pattern is not the same as the design, a benchtop microscope was used to take several measurements of the width of individual squares. The average of the measurements was then taken as the ground truth width of a square. For the results shown below, the measured ground truth width of a square was 213.24 μm (compared to an intended width of 208 μm).

To test the consistency of the visual servoing, 10 scans were performed over the same phantom, generating 10 mosaic images. To maintain the independence of the trials, the phantom was repositioned between each run. In each mosaic image, the width of every square is measured and compared with the ground truth value. In total, 36 measurements were taken along the horizontal and vertical direction for each image. Due to the intrinsic inaccuracy of the servoing, squares in the pattern can be misaligned as shown in Fig. 6.4. The fraction of misaligned squares is presented in Table 6.1. The Root Mean Square Error (RMSE) and Interquartile Mean Error (IME) were then calculated for each trial, with the IME ignoring the effect of outliers. These results are shown in Table 6.1 where the total RMSE is 14.8 μm and the total IME is 10.0 μm. If this is compared to the FoV of the probe (240 μm), it is clear that both qualitative and quantitative results show that the proposed pCLE servoing method can achieve a high accuracy and good consistency using the da Vinci® robot, for the case of a non-deforming sample.

Tendon-driven robots such as the da Vinci® system, are often considered incapable of performing tasks that require such high precision due to backlash and kinematic inaccuracy. To demonstrate how the proposed method improves the robot's capability for pCLE scanning, scanning results with and without the pCLE servoing are compared. As shown in Figs. 6.5(a) and (b), although the kinematic model appears to show that the robot should follow the desired trajectory correctly, the actual mosaic image does not show the corresponding motion, indicating that there is an error in the kinematic model. With pCLE servoing, comparison of the trajectory and resulting mosaic image demonstrates that the controller corrects the kinematic error in order to follow the desired trajectory in the space of the mosaic image. The diameter of the spiral trajectory from the kinematic model is about 0.93 mm and from the pCLE servoing on the printed pattern and porcine tissue is 1.41 mm and 1.85 mm, respectively.

The framework was then tested on *ex vivo* porcine stomach tissue stained with acriflavine. Unlike the phantom experiments, deformation is now caused by probe interaction with the tissue during scanning. In Fig. 6.5(g), the kinematic trajectory indicates that the robot attempts to compensate for this deformation, since the deviation is 0.4 mm more than for the non-deforming phantom shown in Fig. 6.5(c). It can also be seen that the scanning trajectory of an ideal spiral is typically 'squared-off' as shown in Figs. 6.5(c) and (g). The most likely causes of this are backlash in the tendon-driven robot, and the limited bandwidth of the visual servoing control of the robot position.

6.5.2. Robustness to Unexpected Motion

The system is also somewhat robust to unexpected motion and deformation of the scanning target. To show this qualitatively, the scanning and mosaicing process was interrupted by moving the phantom manually. The first row of Figs. 6.5(e) and (f) shows that the servoing is able to recover following the unexpected phantom motion. To confirm

Table 6.1.: Evaluation of Endomicroscopic Servoing

	Misalignment (%)	RMSE (µm)	IME (µm)
Trial 1	17	14.1	8.3
Trial 2	10	15.0	9.1
Trial 3	7	14.6	11.1
Trial 4	7	15.1	8.1
Trial 5	8	12.9	9.7
Trial 6	3	12.9	9.4
Trial 7	17	15.4	9.9
Trial 8	16	12.5	10.0
Trial 9	16	16.2	12.2
Trial 10	16	18.1	13.0
Total	11	14.8	10.0

the capability of the system under unexpected motion of the tissue, the tissue was placed on a motorised translation stage which was moved laterally with different velocities during the robotic scanning. The mosaic results are shown in Fig. 6.6. The maximal velocity of the lateral motion that would not cause failure (discontinuities in the mosaic image) was 0.5 mm/s. When the velocity was increased further to 0.63 mm/s, there was an obvious discontinuity when the system attempted to correct the motion and continue the mosaicing. Improving the system to deal with faster motion would require an increase in the mosaic processing frame rate, which is currently set to 10 frames per second. However, this is not limited in practice by the mosaic algorithm but by the use of a tendon-driven robot that is not capable of fast and accurate motion. If the mosaic frame rate was increased without an increase in velocity, the mosaic algorithm would fail as the positional shift between frames would be too small to detect accurately.

6.5.3. Validation of OCT Distance Servoing

In order to obtain high quality pCLE images, as shown in Fig. 6.7, the probe should be placed in gentle contact with the tissue at all times.

In the probe holder, the OCT probe is positioned with an axial offset, slightly higher than the pCLE probe. Optimal pCLE images should therefore be obtained when the distance to the tissue, given by the position of the top surface in the OCT cross-sectional image, is at a fixed value around this distance (the exact distance required is determined experimentally). To evaluate how effective the OCT distance servoing is at maintaining continuous tissue contact, and hence good image quality, a translation stage was used to move a phantom in a cyclic motion with a constant linear velocity along the axial direction of the OCT probe (both towards and away from the probe). Trials were performed with different linear velocities ranging from 16 to 80 μm/s.

The results in Fig. 6.8 show how well the robot maintains a desired distance relative to the moving phantom under different peak velocities (16, 32 and 80 μm/s). When the phantom moves, the deviation between the current and desired distances can be generally kept within 4 pixels (approximately 40 μm). When the phantom stops, the current distance successfully converges to the desired one.

To determine whether the OCT distance servoing improves the robustness of the pCLE scanning, a spiral scanning task on a curved surface phantom was tested with and without the distance servoing. Fig. 6.9, shows an example where the mosaicing cannot continue without OCT distance servoing as the pCLE probe loses contact with the phantom and the image becomes blurred. In contrast, as shown in the zoomed region of Fig. 6.9, when contact begins to be lost, the robot is successfully able to recover.

6.5.4. OCT Volume Rendering and Fusion

To validate the OCT volume rendering, a surgical needle with a thickness of 0.3 mm was scanned. In this experiment, the OCT distance servoing was disabled so that only pCLE images were used for the control. As shown in Fig. 6.10, a segment of the needle is scanned and

reconstructed in an OCT volume. From the volume, it can be clearly seen that the needle is placed on a flat tissue surface. To visualise both macro and micro information, a multiscale image fusion is shown in Fig. 6.11. From the stereo reconstruction in macro scale, the fusion framework can zoom in to the scanned region in micro scale where a mosaic image and an OCT volume are presented.

6.6. Discussion and Conclusion

This chapter presents an automated scanning framework for pCLE and OCT optical biopsy using the da Vinci® surgical robot. It is capable of generating large area mosaics of both pCLE and OCT images. A crucial feature is that pCLE images are used to close the control loop, and mosaicing results from both static and deforming phantoms demonstrated that this effectively compensates kinematic errors. Furthermore, by using OCT images to maintain a constant distance to the tissue, and hence ensure consistent contact between the pCLE probe and the tissue, the system is able to compensate for target motion along the axial direction. This visual servoing allows for the correction of errors due to tissue deformation, robot positioning and grasping of the pick-up probe. The accuracy of this correction is better than the FoV of the pCLE probe, resulting in continuous 2D mosaics without gaps or discontinuities, which represent a common problem for open loop control. Finally, the resulting high-resolution tissue maps at micro-scale can be fused in real-time with a stereo reconstruction at macro-scale from the laparoscopic image-feed, providing the surgeon with a multi-scale 3D view of the operating site. This augmented visualisation provides a range of potential benefits for intraoperative tissue characterisation and surgical planning. Future work will investigate complementary servoing using the laparoscopic image, both for robot motion to the desired scan site, and to improve robustness to loss of contact (and hence loss of pCLE images for servoing), and an investigation of the potential of the system to handle patient motion.

A study comparing the results to master-slave controlled scanning will confirm the practical clinical benefits of the system in providing the surgeon with real-time intraoperative cellular-scale tissue analysis.

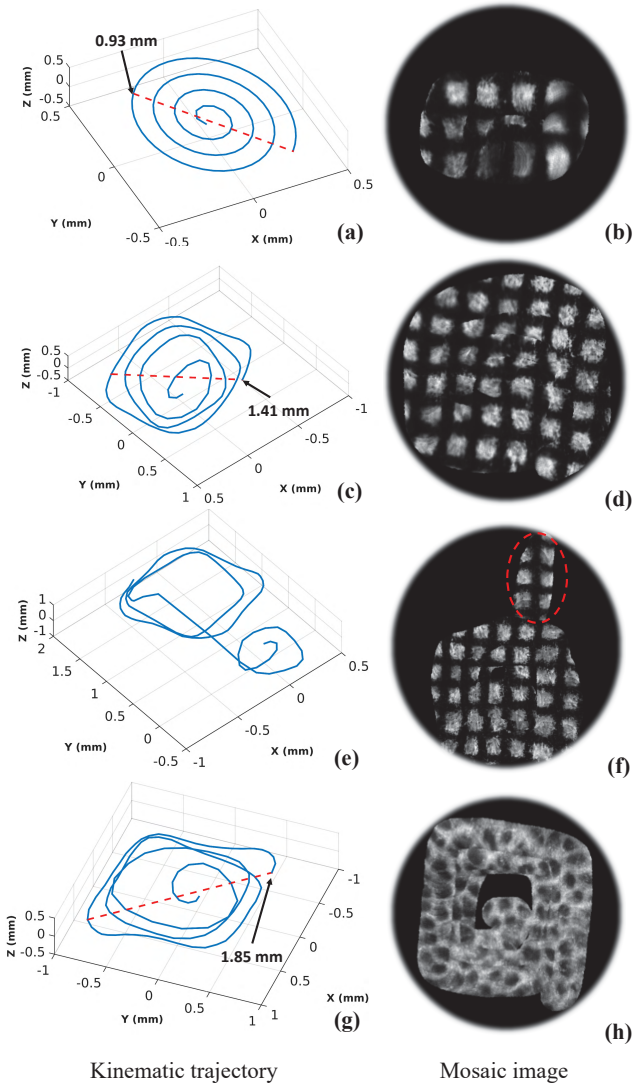


Figure 6.5.: Illustration of how pCLE servoing improves mosaicing results. The left column shows the scanning trajectory based on kinematic reading of the robot's end-effector and the right column shows the corresponding mosaic results on the same phantom. The first row is the result from using only the kinematic model, while the second row uses closed-loop scanning. The effect of unexpected motion on the mosaic image is shown in the third row. The unexpected motion is shown in the red dashed circle in (f) before the closed-loop servoing recovers and returns the probe to the desired trajectory relative to the tissue surface.

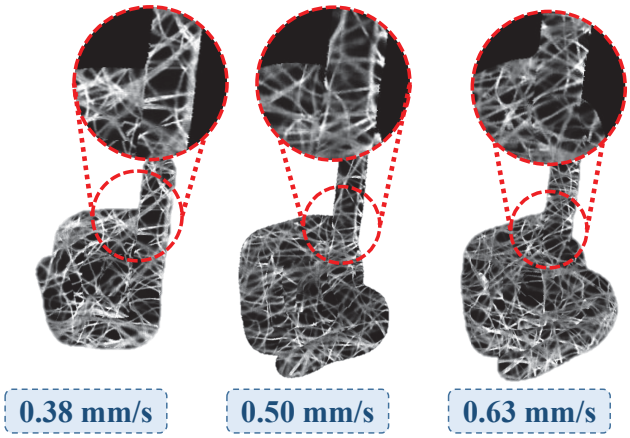


Figure 6.6.: Mosaic results obtained using the framework under unexpected motion with various velocities. The 'arm' connecting the mosaic over the planned region and the additionally scanned region (caused by the motion) is enlarged to show that it is continuous, suggesting that the probe correctly returned to its planned trajectory on the tissue surface.

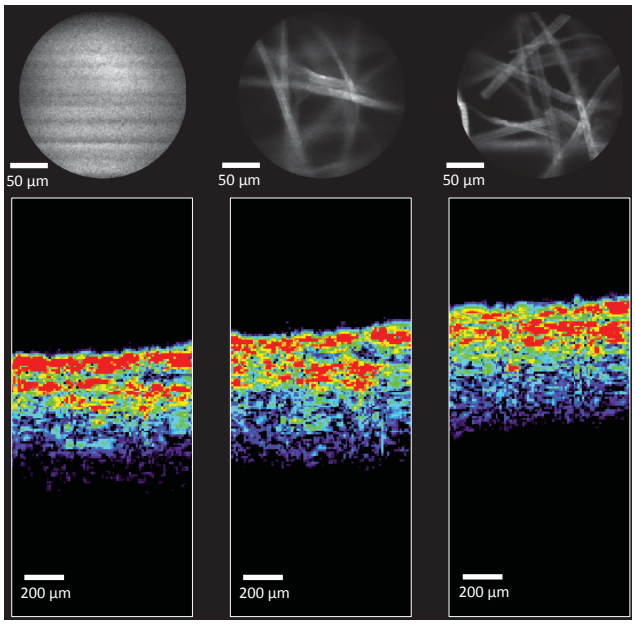


Figure 6.7.: Demonstration of the correlation between the distance to the tissue surface measured in the OCT image (bottom row) and the quality of pCLE images (top row). As the surface moves closer to the probe, from left to right in the figure, the probe-to-tissue distance measured by the OCT probe becomes smaller while the pCLE image quality improves. Note that the pCLE images are autocontrasted.

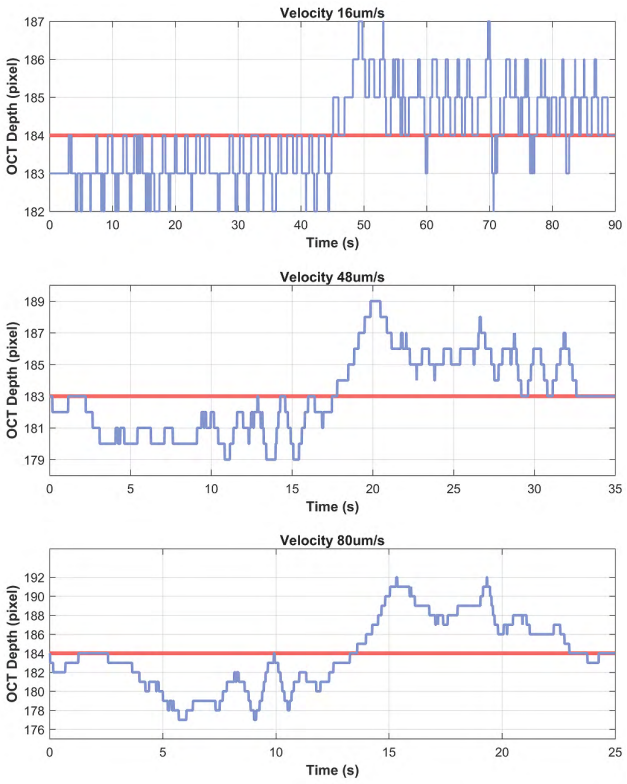


Figure 6.8.: The desired OCT distance shown in red is assumed to be constant during servoing. As the phantom starts moving, the deviation between the desired potion and the current position measured by the OCT channel initially increases and then decreases towards zero as the system corrects for the motion. Finally, the deviation converges to zero when the phantom stops. 1 pixel in the OCT image corresponds to approximately 10 μm in real space.

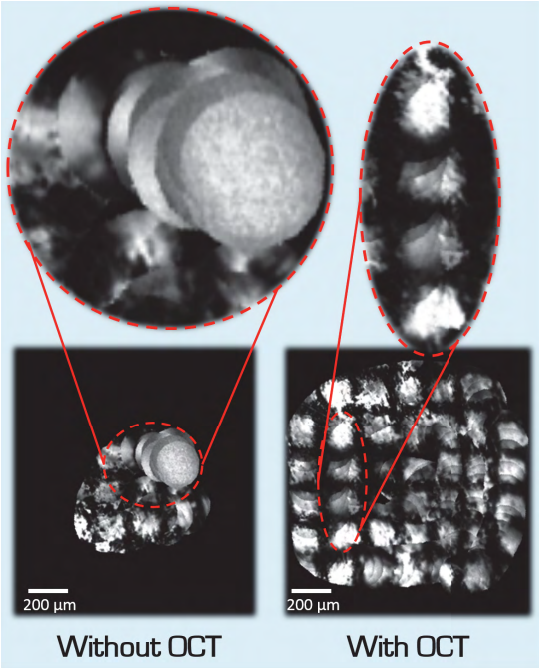


Figure 6.9.: Demonstration of pCLE scanning with and without OCT distance compensation on a curved surface. Without the OCT distance servoing, the mosaic could only continue for a short distance before the probe lost contact, the pCLE image was degraded and the lateral visual servoing failed. When OCT distance servoing was turned on, a mosaic was generated over the whole trajectory.

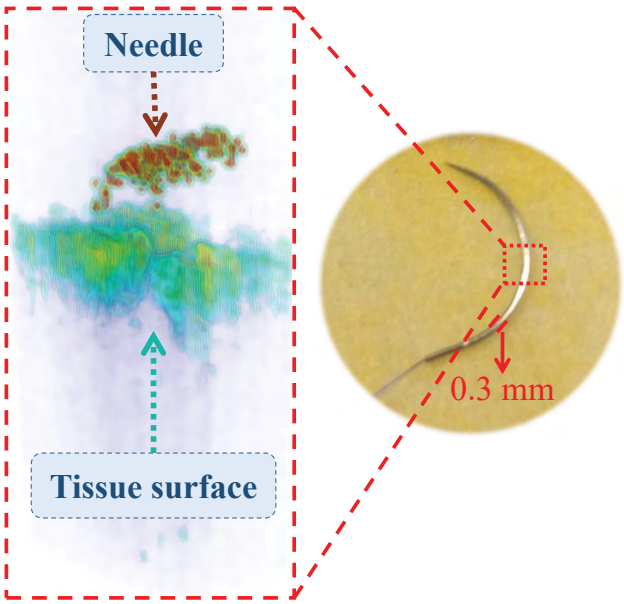


Figure 6.10.: A volume reconstruction of a surgical needle with a thickness of 0.3 mm. The red rectangle indicates the segment that has been reconstructed in the microscopic volume shown on the left.

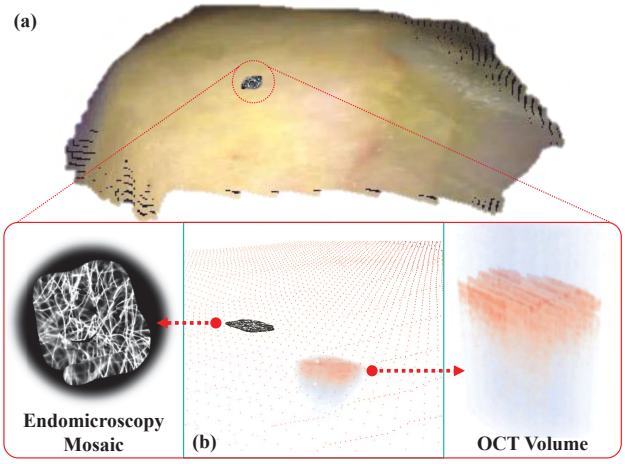


Figure 6.11.: An example of multiscale fusion. A macro scale stereo reconstruction is shown (a). To link with the micro scale, the mosaic image is overlaid on the stereo reconstruction (b). To visualise the micro scale images, a pCLE mosaic and an OCT volume are shown.

7. Motion Compensated Ultrasound Scanning †

7.1. Introduction

In the past decade, intraoperative ultrasound has been increasingly used in both minimally invasive interventions and robot-assisted laparoscopy to assist the localisation and targeting of pathological tissue during tumour dissection. It enables intraoperative visualisation of the tissue without involving ionizing radiation and facilitates the surgeon to define optimal dissection planes. In addition, injury to critical anatomical structures can be avoided, as localisation of margins for tumour dissection is improved. However, manually controlled ultrasound scanning can significantly increase the cognitive load of surgeons, due to the need for maintaining optimal scanning orientation and consistent contact with the tissue, and simultaneously interpreting the ultrasound images during scanning.

In Chapter 5 and 6, automation of repetitive scanning task that requires high precision has been presented. Intraoperative ultrasound can also benefit from the surgical automation in robotic surgery. In [119], an auxiliary robotic arm holding an ultrasound probe has been used to follow the motion of a tele-operated arm such that image guidance can be provided to the surgeon during tumour resection. A 6-

†Content from this chapter was published as:
Motion-Compensated Autonomous Scanning for Tumour Localisation using Intraoperative Ultrasound. Zhang, L., Ye, M., Giannarou, S., Pratt, P., Yang, G.-Z.. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 619-627 (2017).

DoF ultrasound visual servoing approach has been proposed in [169] to control the motion of a robotic system equipped with an ultrasound probe based on intensity information from ultrasound B-mode images. Surgical suturing subtasks have been automated in [163] by using a dedicated suturing tool under fluorescence guidance.

An automatic tumour dissection framework using intraoperative ultrasound has been proposed in [148] but it has been designed for planar tissue surfaces. In [109], a framework for autonomous scanning with an endomicroscopy probe over complex tissue surfaces has been proposed to perform microscopic image mosaicking. The above autonomous systems have been designed to deal with static surgical environments which limit their applicability in real surgical scenarios. Hence, a 6-DoF autonomous scanning approach that adapts to tissue motion has yet to be proposed. Tissue motion recovery is important for autonomous execution of surgical tasks. In [170], a respiratory model of tissue motion has been extracted for simultaneously estimating camera motion and tissue deformation. Recently, a probabilistic tracking and surface mapping method was proposed for free-form tissue deformation recovery [171].

In this chapter, a novel system for autonomous ultrasound scanning under tissue motion is presented for intraoperative tumour localisation. The system is aimed for kidney and liver surgery where tissue motion is mainly caused due to respiration. An efficient approach is proposed to plan in advance the optimal trajectory for the ultrasound scanning, considering the surface curvature as well as ultrasound transducer dimensions for efficient coverage, which improves the ultrasound image quality and reduces scanning time. For motion compensation, an effective vision-based approach is proposed to learn the tissue motion. The learned motion model is updated online and incorporated into the scanning movement of the ultrasound probe. A 3D tumour model is reconstructed from a sequence of ultrasound images obtained after scanning, which provides an intuitive visualisation of tumour location under the tissue surface. The proposed system has been validated with

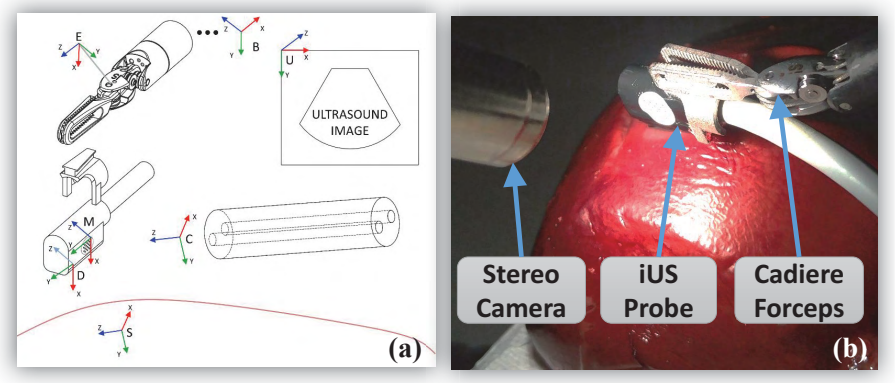


Figure 7.1.: (a) Coordinate frames defined in the proposed system. {**B**}: robot base; {**E**}: robot end-effector; {**M**}: marker attached on the probe; {**D**}: transducer; {**C**}: stereo camera; {**S**}: tissue surface; {**U**}: ultrasound image (b) Hardware Setup;

phantom and *ex vivo* experiments using ground truth motion data and CT ground truth. The presented results verify that our system allows the FoV of the ultrasound to follow the tissue motion, which simplifies the tumour localisation task for the surgeon.

7.2. Motion Estimation and Online Update

7.2.1. Tissue Motion Estimation

For motion compensation, the tissue motion can be learned prior to scanning, before the ultrasound probe enters the FoV of the stereo camera, thereby eliminating occlusion of the scene. At time t, a 3D surface point $\mathbf{P}_t$ is estimated by matching its 2D projections $\mathbf{p}_t^l$ and $\mathbf{p}_t^r$ on the left and right images of the camera, respectively, using the stereo matching method proposed in [172].

To recover the global motion of the tissue, the relative motion $\Delta\mathbf{P}_t = \mathbf{P}_0 - \mathbf{P}_t$ of a set of points on the tissue surface is calculated. The first

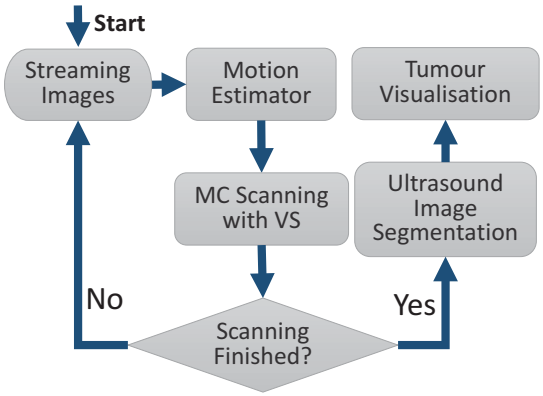


Figure 7.2.: Flowchart of the proposed framework. MC: Motion Compensated; VS: Visual Servoing.

stereo image pair at time $t = 0$ is denoted as the reference and all the following images are compared to it. The optical flow [173] is estimated between the reference image and the image at time t to find the correspondence pair $\{\mathbf{p}_0^l, \mathbf{p}_t^l\}$. For accurate motion estimation, a grid of points is sampled from the target scanning region, which are tracked as above. To eliminate outlier correspondences, forward-and-backward tracking is performed and the inconsistent points are discarded in the motion estimation. The tissue motion $\Delta\mathbf{P}_t$ is estimated as the median of the motion of the consistent grid points.

In order to learn the underlying motion characteristics, PCA is used to extract the major motion and project it from the camera coordinate space to the PCA space. The PCA component with the largest eigenvalue corresponds to the respiratory motion.

In this chapter, a typical respiratory model with asymmetric period similar to [170] is employed. The model can be represented as:

$$z(t) = z_0 - b\cos^{2n}(\frac{\pi t}{\tau} - \phi) \tag{7.1}$$

where z_0 is the exhale position of the respiration, b is the amplitude, τ is the period of the respiration, ϕ is the phase and n describes the degree of asymmetry (the larger the n, the longer the expiration). In this

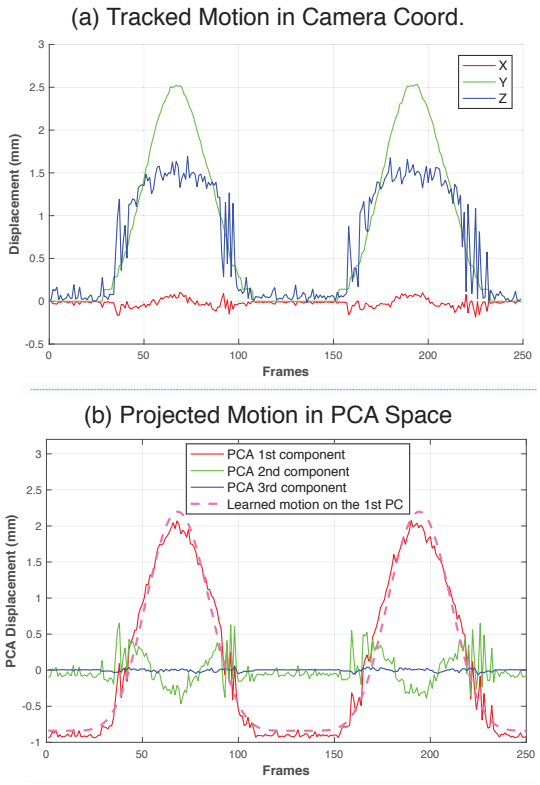


Figure 7.3.: Illustration of motion estimation using PCA. The tissue motion is tracked in the camera coordinate (a) and projected in the PCA space (b). The first principle component, which corresponds to the largest motion variation, is learned by fitting a respiration model to it.

work, $n = 3$ which represents a typical respiratory motion model [174]. The parameters of the model are estimated such that the model fits to the major motion in the PCA space. This is treated as a non-linear least-squares problem which can be solved by using the Levenberg-Marquardt optimisation method. Unlike [170], where a fixed phase $\phi = 0$ has been assumed, all four parameters (z_0, b, τ and ϕ) are estimated in this chapter, observing two motion cycles.

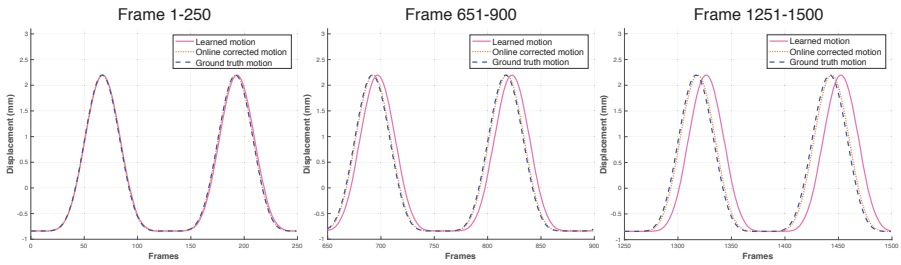


Figure 7.4.: Small errors in the period of the learned motion characteristics can result in large phase drift during the scanning task. The online updated tissue motion can keep up with the ground truth while the learned motion gradually deviates from the actual tissue motion

7.2.2. Online Motion Update

Small errors in the period of the learned motion (τ) can result in the latency between the robot and tissue motion during scanning. As shown in Fig. 7.4, the phase of the learned model will be gradually shift away from the actual tissue motion. To correct this phase shift, the tissue motion model is also estimated online during ultrasound scanning. During scanning, the user can define a region on the tissue for online motion estimation. The same 3D reconstruction result from [172] is used, and perform optical flow tracking on the selected tissue region. With the tracked 2D points, tissue movements can be estimated by solving a PnP problem. Online motion estimation is performed using the tissue movements in the last two cycles, and updates only the parameter $\hat{\phi}$ during scanning, as it is found that phase is an effective parameter for adjusting the latency between the robot and tissue motion.

7.3. Motion-Compensated Ultrasound Scanning

To generate a desired trajectory for ultrasound scanning, a target region on a reference image is defined by the user. A starting point is selected and a zig-zag trajectory which covers the required area is planned on the 2D reference image. From stereo matching, the corresponding 3D positions of these 2D trajectory points can be acquired. The distance between each line of the zig-zag trajectory is defined as half width of the transducer in order to scan the region efficiently. At each 3D point along the trajectory, surface normal is calculated using its neighbouring points. The coordinate systems are defined in Fig. 7.1(a). The scanning trajectory is planned with respect to the pose of the marker (circular-dot pattern) attached on the ultrasound transducer. It is defined as a sequence of desired marker poses ($\mathbf{T}_C^{M*}$) in the camera coordinate system such that the transducer can be aligned with the surface normal at every point of the trajectory. Tissue deformation caused by the contact of the probe is kept to minimum since the trajectory is planned based on the tissue surface.

In this work, a position-based visual servoing method is employed to control the robot in the Cartesian space. More specifically, a desired pose for the robot end-effector is calculated in each control loop as:

$$\mathbf{T}_B^{E*} = \mathbf{T}_B^E \cdot \mathbf{T}_E^M \cdot \widetilde{\mathbf{T}}_M^{M*} \cdot \hat{\mathbf{T}}_M^{M*} \cdot \mathbf{T}_{M*}^{E*}, \tag{7.2}$$

where $\mathbf{T}_B^E$ is the current end-effector pose in the robot base coordinate system. $\mathbf{T}_E^M = (\mathbf{T}_{M*}^{E*})^{-1}$ is a constant transformation between the end-effector and the marker on the probe that can be measured. $\mathbf{T}_B^{E*}$ relies only on $\mathbf{T}_M^{M*}$ and any change in $\mathbf{T}_E^M$ will not affect $\mathbf{T}_B^{E*}$. The desired relative movement of the marker depends on the motion caused by the scanning motion as well as the motion for compensating for respiration, which are denoted as $\widetilde{\mathbf{T}}_M^{M*}$ and $\hat{\mathbf{T}}_M^{M*}$, respectively. The desired relative

marker pose for scanning is calculated as:

$$\widetilde{\mathbf{T}}_M^{M*} = \hat{\mathbf{T}}_M^{M*} \cdot (\mathbf{T}_C^M)^{-1} \cdot \widetilde{\mathbf{T}}_C^{M*}. \tag{7.3}$$

Here, $\widetilde{\mathbf{T}}_C^{M*}$ and $\mathbf{T}_C^M$ are the desired marker pose and the detected marker pose (in the camera coordinate system) in the reference and current frame, respectively. The desired relative marker pose for motion compensation is calculated as:

$$\hat{\mathbf{T}}_M^{M*} = \mathbf{T}_M^D \cdot \mathbf{T}_D^S \cdot \mathbf{T}_S^{S*} \cdot \mathbf{T}_{S*}^{D*} \cdot \mathbf{T}_{D*}^{M*}, \tag{7.4}$$

where $\mathbf{T}_M^D = \mathbf{T}_{M*}^{D*}$ is a constant offset from the marker to the transducer that can be measured. $\mathbf{T}_D^S = \mathbf{T}_{S*}^{D*}$ is a rotation between the transducer and surface point which has the same orientation as the camera coordinate. $\mathbf{T}_S^{S*} = \left(\begin{smallmatrix} \mathbf{I}_3 & \Delta\mathbf{x} \\ 0 & 1 \end{smallmatrix}\right)$ is the transformation from the current to the next surface position which can be calculated from the learned motion model using the PCA transformation. The tissue position $\mathbf{x}_t$ relative to the reference frame at time t can be estimated via: $\mathbf{x}_t = \mathbf{w}^{-1} \cdot [z(t), 0, 0]^\intercal$, in which $\mathbf{w}$ is the PCA eigenvectors describing the transformation from the 3D to the PCA space. Then $\Delta\mathbf{x} = \mathbf{x}_{t+1} - \mathbf{x}_t$.

7.4. Tumour Segmentation and Visualisation

To create a 3D mesh of the tumour for intraoperative visualisation, the ultrasound images captured during scanning are segmented. Any segmentation method can be integrated into the system (manual segmentation is used in this chapter). The tumour segmented boundary is represented as a number of points where each point is connected to its neighbours across ultrasound images so that a 3D mesh can be reconstructed. The segmented boundary points are transformed from the 2D image coordinate system to the 3D camera coordinate system

Table 7.1.: Estimation results of parameter on three different motion profiles

	Profile 1		Profile 2		Profile 3	
	τ (**Frames**)	b (**mm**)	τ (**Frames**)	b (**mm**)	τ (**Frames**)	b (**mm**)
Phantom Estimated	75.2±0.3	3.0±0.1	125.1±0.4	3.0±0.1	124.9±1.0	5.0±0.1
Phantom Ground Truth	75.0	3.0	125.0	3.0	125.0	5.0
Ex Vivo Estimated	75.1±0.2	3.2±0.2	125.2±0.3	3.1±0.3	125.1±0.4	4.8±0.4
Ex Vivo Ground Truth	75.0	3.0	125.0	3.0	125.0	5.0

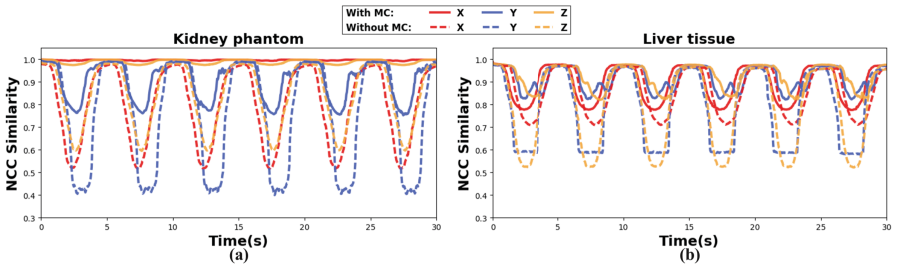


Figure 7.5.: Quantitative results of ultrasound stabilisation on phantoms (a) and *ex vivo* tissue (b) with and without MC.

C. To this end, an ultrasound coordinate system U is defined in which each pixel of the ultrasound image can be represented as a 3D point P^U. The segmented boundary points are then transformed to the coordinate C via: $P^C = \mathbf{T}_M^C \cdot \mathbf{T}_D^M \cdot \mathbf{T}_U^D \cdot P^U$ where $\mathbf{T}_U^D$ can be calibrated by following [175]. The created tumour mesh is registered in C as shown in Fig. 7.8(d) and (h).

7.5. Validation and Experimental Results

7.5.1. Experimental Setup

The proposed autonomous ultrasound scanning system includes an Aloka ultrasound machine with an UST-533 linear transducer, a standard da Vinci® surgical robot, and a workstation for real-time pro-

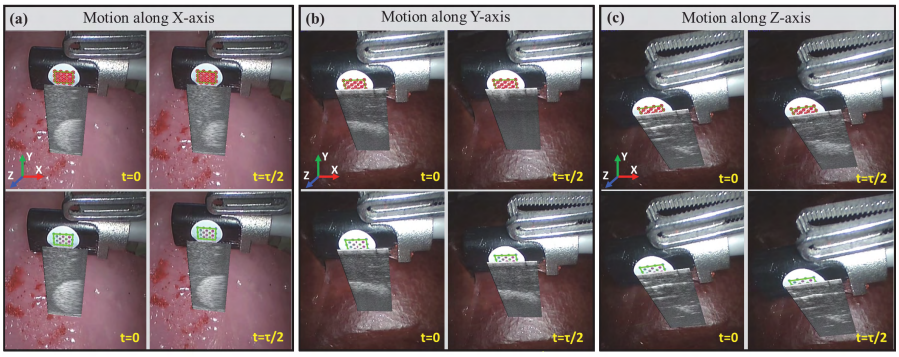


Figure 7.6.: Qualitative results of ultrasound stabilisation. The top and bottom rows are the trials without and with motion compensation, respectively. Three motion profiles were tested on a phantom (a) and *ex vivo* tissue (b-c), with directions along the X-, Y- and Z-axis of the motorised device.

cessing and robot control. The robot includes a sterescopic camera streaming images at 25Hz, and a patient site manipulator which is controlled via the dVRK. As shown in Fig. 7.1(b), a circular-dot pattern (KeySurgical Inc., USA) has been attached to the ultrasound probe to facilitate tool tracking and pose estimation [148] at 25Hz. The motion estimation module (including optical flow and parameter estimation) runs at ≈30Hz. To evaluate our proposed framework, experiments were performed on two kidney phantoms (made of polyvinyl alcohol) and an *ex vivo* tissue (ovine liver). For *ex vivo* experiments, tumours are simulated using olives which are embedded inside the tissue.

7.5.2. Motion Estimation and Stabilisation

To simulate respiratory motion, a motorised device is used to host the targets for scanning, such that periodic motion can be generated. Three motion profiles with different periods ($\tau \in \{75, 125\}$) and amplitudes ($b \in \{3, 5\}$) are simulated. For each motion profile, 9 trials were conducted to estimate motion parameters τ and b. These trials contain 3 trials each for motion along X, Y, and Z axis (w.r.t. motorised

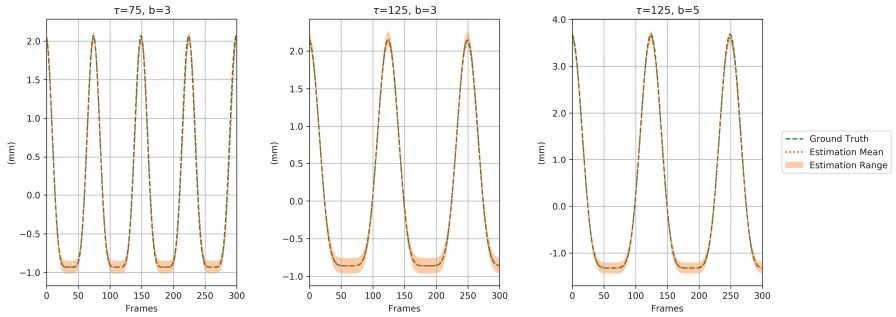


Figure 7.7.: Comparison of the estimated and ground truth motion for different motion characteristics.

device coordinate system). Table 7.1 presents the mean and standard deviation values of these parameters. The estimated and ground truth motion is shown in Fig. 7.7. It can be seen that our motion estimation provides accurate results compared to the ground truth models (defined on the motorised device).

The accuracy of our system can also be demonstrated using "ultrasound stabilisation". As shown in Fig. 7.6, the robot is controlled to adapt to certain surface points on the tissue in the presence of tissue motion. The difference in obtained ultrasound images show the effect of motion estimation. The stabilisation accuracy is quantified using Normalised Cross-Correlation (NCC) calculated between each image and the reference image obtained at exhale position. A high NCC measure indicates a good quality of motion estimation and compensation. Fig. 7.5 presents the NCC measures calculated with and without motion compensation on three different motion profiles. It is clear that by introducing motion estimation and compensation, the NCC measures can be significantly improved.

7.5.3. Tumour Scanning under Tissue Motion

To evaluate the performance of tumour scanning under tissue motion, further experiments are conducted on phantom and *ex vivo* tissue. For each target, a user defines a point-of-interest for scanning and a region

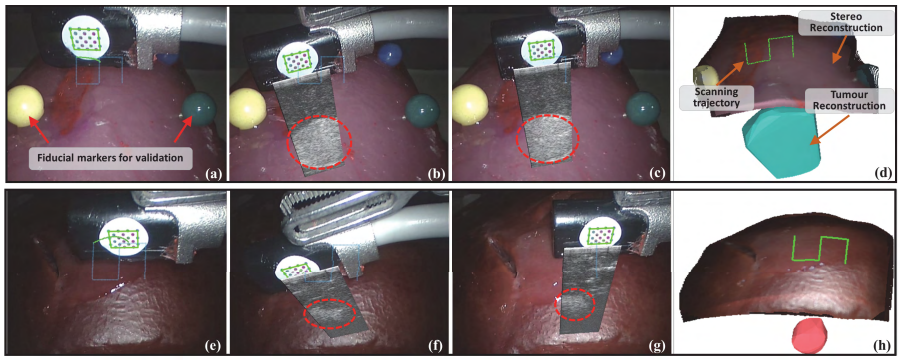


Figure 7.8.: Qualitative results of two example autonomous scanning tasks on a phantom (a-d) and *ex vivo* tissue (e-h). The obtained tumours and 3D tissue surfaces are visualised in (d) and (h). See the supplementary video via `https://youtu.be/EsomHSwIz2g`.

Table 7.2.: Quantitative results of tumour scanning under three different motion profiles.

	Phantom 1			Phantom 2		
	Profile 1	Profile 2	Profile 3	Profile 1	Profile 2	Profile 3
Location Error (mm)	3.85	2.33	3.81	2.78	2.57	3.38
Diameter Error (mm)	0.69	0.61	0.33	0.84	0.61	0.77

for motion estimation. These initialisation steps take few seconds via a customised GUI. Our framework then performs autonomous scanning with motion compensation, followed by ultrasound image segmentation and tumour visualisation. The qualitative results are presented in Fig. 7.8. To validate the reliability of scanning, the locations and diameter of the obtained tumours are compared to the ground truth (via CT scanning). As shown in Table 7.2, the proposed framework presents accuracy results amounting to 0.64 mm (size) and 3.12 mm (location), which are close to the clinically accepted tumour dissection margin errors (5-7 mm) reported in [175]. It is worth noting that, the location errors include the registration error between the CT and camera coordinate systems. Quantitative results are not provided for the *ex vivo* data due to the difficulty in obtaining ground truth structures

as in our case the embedded tumour is not fixed between scans.

7.6. Discussion and Conclusion

This chapter presents a motion-compensated autonomous system for intraoperative ultrasound scanning during robotic surgery. The proposed system performs vision-based tissue motion estimation with online update to recover periodic respiratory motion. The estimated motion parameters are then combined with 6-DoF visual servoing for accurate continuous scanning under tissue motion. The obtained ultrasound images are segmented such that tumours can be localised and visualised to the surgeons. Both the quantitative and qualitative results from phantom and *ex vivo* experiments demonstrate the capability of our system for motion-compensated autonomous scanning. The proposed system allows the FoV of the ultrasound to follow the tissue motion, which not only simplifies scanning tasks for surgeons but can also permit observation of temporal changes of the underlying tissue such as blood flow and contrast uptake.

8. Conclusion and Future Research Directions

8.1. Achievements and Contributions of this Thesis

In recent years, there have been considerable interest in the use of robotic devices and imaging technology for laparoscopic surgery. Surgical robots offer significantly greater dexterity than conventional laparoscopic tools while providing high mechanical accuracy. Medical imaging techniques such as CT and MRI scans provide important subsurface information and volumetric models of a patient's anatomy, which are useful for preoperative planning and diagnosis. As the preoperative model may differ from the patient's true anatomy during surgery, intraoperative imaging can be utilised for real-time diagnosis and guidance. For example, laparoscopic ultrasounds have been shown to be a useful tool for the localisation of tumour boundaries and the identification of critical anatomical structures such as blood vessels and nerves. Moreover, new imaging modalities such as CLE and OCT enable a real-time optical biopsy by providing on-site histologic information on a sub-cellular level. While these modalities provide a potentially useful tool for early disease diagnosis during surgical oncology, it is difficult for them to characterise larger areas of tissue due to their small FoV. For this reason, video mosaicking which synthesises a larger area is needed. However, this is a difficult manual task (even when performed using a robot) which requires the operator to maintain

an optimal probe pose and tissue contact while performing slow and controlled scanning motions with sub-millimetre accuracy. The goal is to develop an autonomous scanning framework using a surgical robot such as the da Vinci® system by making use of the visual information and the mechanical accuracy of the robot. Furthermore, the scanning results can be presented via an intuitive and effective real-time visualisation of multi-scale imaging information, which assists in surgical diagnosis and planning.

The review in Chapter 2 provides an overview of the main developments in image-guided robotic surgery and the automation of surgical tasks according to different autonomy levels. Previous definitions of the different levels of autonomy for surgical robotics are reviewed and discussed. The existing imaging modalities for surgical diagnosis and guidance and their applications in RMIS are introduced. An overview of the existing commercial and research-based robotic platforms for laparoscopic surgery is provided. The research developments in human-robot collaboration and the automation of surgical tasks are reviewed. Moreover, a discussion of the technical challenges, as well as the ethical and legal issues related to the development of surgical autonomy, is provided.

The marker-based instrument tracking method presented in Chapter 3 facilitates the automation of surgical tasks via a monocular endoscopic camera. The proposed hybrid marker and the associated tracking algorithm allows for the localisation of cylindrical surgical instruments. Experimental results show that the tool tracking framework can achieve a large workspace while offering comparable accuracy compared to the existing state-of-the-art methods. The tracking framework is successfully utilised in laparoscopic ultrasound guidance by overlaying the ultrasound images on the endoscopic view.

The model-based instrument tracking method described in Chapter 4 has been designed for robotic instruments with articulated joints. The proposed tracking method performs template matching of the contour features on each part of the robotic instrument by projecting the

CAD model based on the robot's forward kinematics. The outlier template matches are rejected by a robust verification method based on 2D geometrical context. The inlier detections are used for 3D pose estimation via the EKF. The experiments on the phantom, *in vivo* and *ex vivo* data, confirm that the proposed approach outperforms the state-of-the-art methods while providing real-time performance.

Chapter 5 introduces the autonomous framework for endomicroscopic scanning and fusion based on the da Vinci® robot. The framework provides autonomous scanning and 2D mosaicing over a user-defined region of interest. The probe carried by the da Vinci® robot can achieve motion with sub-millimetre accuracy so that the generation of large mosaics with minimal intervention from the surgeon is possible. The tracking method proposed in the Chapter 3 is used for the tracking of the probe which guarantees a smooth and accurate scanning motion. Furthermore, the intuitive visualisation and fusion of the multi-scale images offer significant possibilities to enhance surgical procedures by providing the operator with cellular-scale information over a large area.

A novel real-time multi-scale fusion scheme for robotic surgery is proposed in Chapter 6, based on extending the scanning framework from Chapter 5. By incorporating live pCLE images and OCT images for lateral and axial position control, respectively, precise probe scanning can be achieved. The use of OCT images ensures consistent contact between the pCLE probe and the tissue, which was a shortcoming of the scanning framework presented in Chapter 5. The system generates a tissue map consisting of a 2D pCLE mosaic for surface information and 3D OCT volume for subsurface structure. The resulting microscopic tissue maps can be fused real-time with a stereo reconstruction from the laparoscopic video, providing the surgeon with a multi-scale 3D view of the operating site.

A system for automated ultrasound tumour scanning on moving tissue is presented in Chapter 7, by estimating the motion model of the tissue and then incorporating the motion into the visual servoing

framework. The tissue motion characteristics can be learned through an effective vision-based approach and updated online based on the 3D tracking of the tissue. The scanning trajectory can be customised based on the tissue shape and the dimension of the transducer for efficient coverage. The robot can maintain the ultrasound probe in a relatively static position in the presence of tissue motion while moving the probe to follow the planned scanning trajectory. To assist in tumour localisation, the tumour model built from the ultrasound images can be visualised along with the 3D tissue surface.

8.2. Future Research Directions

While the research in this thesis focuses on the automation of surgical tasks with minimal intervention from the surgeon, shared control between the human surgeon and the robot is seen as a promising solution to enhance the surgeon's capability in complex surgical tasks. For example, the autonomous ultrasound scanning in Chapter 7 can be modified to allow the surgeon to have control of the ultrasound probe movement. In a shared control setup, the robot stabilises the probe to follow the tissue motion so that the surgeon can conduct out the scanning without having to adjust for the tissue motion. Furthermore, a performance assessment of the different control modes, similar to the user study presented in [106], would be useful. Existing studies [176, 177] have tried to address the shared control component for heart motion compensation by assuming that the motion profile is known. [119] presents an ultrasound guidance system for robot-assisted laparoscopic radical prostatectomy by registering a transrectal ultrasound robot with the da Vinci® robot so that the ultrasound probe can follow the movement of the da Vinci® instrument which is manipulated by the surgeon. This suggests that the auxiliary arm of the da Vinci® robot, which is supported by the dVRK, can be used as a form of smart assistance while the surgeon performs surgical tasks.

One of the most difficult technical problems for surgical navigation

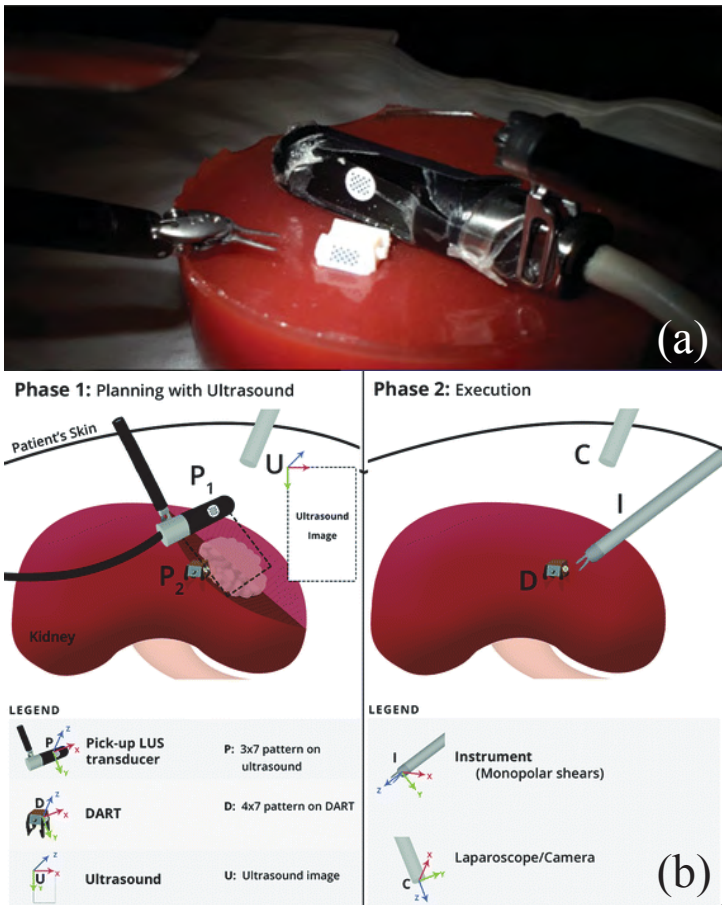


Figure 8.1.: (a) An example use of a navigation aid (called DART) for intraoperative tumour dissection Figure by [178] used under CC BY/Cropped from original. (b) The procedure has two phases: the surgeon first perform ultrasound scanning after the DART is placed near the tumour, then AR can be performed by tracking the DART during tumour dissection ©2016 Springer [179]

and automating surgical tasks is the estimation of soft tissue motion, which should include both global movement and local deformation. The framework proposed in Chapter 6 only considers a only small amount of local deformation, and the system presented in Chapter 7 only considers the global periodic motion. Hence, an accurate and reliable motion tracking algorithm of tissue with real-time performance is required. The problem becomes more challenging when the tissue changes dramatically due to instrument interaction such as grasping, cutting and dissection. A potential solution to this problem is the biomechanical model [180], which can be utilised to estimate soft tissue deformation. With regards to tumour dissection, such as a partial nephrectomy, [178] presents a 3D-printed navigation aid that can be inserted into the cortex of the kidney, where it remains approximately fixed relative to the tumour during the dissection. The estimation of tissue motion can be simplified by tracking the marker attached to the navigation aid using the method proposed in Chapter 3. Furthermore, automating the tissue dissection procedure by using the navigation aid is feasible.

Despite the marker-based instrument tracking method proposed in Chapter 3 facilitating the visual servoing of the surgical instruments, the requirements of the marker limit its usability in some applications in which fixing a marker an instrument is difficult. The model-based instrument tracking method proposed in Chapter4 does not rely on any marker, but the pose estimation accuracy is inadequate. Therefore, further improvement in the instrument tracking methods is desirable as this will extend the type of surgical task that can be automated. Recent advances in deep learning may be a solution to this problem as promising results have been presented in [70].

8.3. Conclusion

Although minimally invasive laparoscopic surgery has been widely used in different disciplines of surgical oncology, it is limited by a number

of factors, including limited contact and tactile feedback, the loss of stereo depth perception, indirect camera navigation and reduced dexterity. In recent years, the appearance of surgical robots has addressed these limitations by integrating built-in force sensors into the surgical instruments, providing a 3D camera that is held by a robotic arm and increasing the freedom of the surgical instruments. However, most systems can only replicate the surgeon's movements without any automatic function. Additionally, many robotic platforms are not designed to take advantage of new imaging technologies that can provide multi-scale anatomical and pathological information for improved surgical diagnosis and planning. Existing research on the application of imaging techniques for intraoperative guidance has yielded positive results by improving surgical accuracy, reducing tumour margins and sparing critical nerves. This study presents a framework for the automatic scanning of various imaging modalities. The scanning results can be combined with the endoscopic view and visualised in a clear and intuitive way to facilitate surgical diagnosis and navigation. In order to automate the scanning task through visual servoing, this research also develops two real-time visual tracking methods for surgical instruments. Furthermore, the autonomous scanning framework can cope with a dynamic surgical environment, such as tissue motion, by adding the estimated motion characteristics into the scanning motion. The results presented in this thesis can encourage the development of autonomous systems for surgical tasks, while reducing the cognitive workload of the surgeon, and providing insights into improving surgical diagnosis and navigation for image-guided robotic surgery in the form of autonomous robotics.

A. Autonomous Ultrasound Scanning for Intraoperative Tumour Localisation and Diagnosis †

Intraoperative ultrasound (iUS) has become a popular tool in robot-assisted laparoscopic surgery to facilitate in situ pathology localisation. However, manually controlled ultrasound scanning can cause significant cognitive loads to surgeons, due to the need for maintaining optimal scanning orientation and consistent contact with the tissue, as well as covering a wide area for tumour detection and surveillance. In [181], an iUS probe has been controlled by an auxiliary robotic arm to provide guidance to the surgeon for tumour resection. Although an autonomous tumour dissection framework using the iUS has been proposed in [148] on planar tissue surfaces, a 6-DoF autonomous scanning approach that adapts to arbitrary tissue surfaces has yet to be presented. This work proposes an automatic iUS scanning framework using the dVRK. Our framework consists of trajectory planning based on 3D surface reconstruction, an automatic scanning via a 6-DoF vi-

†Content from this chapter was published as:
Autonomous Ultrasound Scanning for Intraoperative Tumour Localisation and Diagnosis. Zhang, L., Ye, M., Chan, C., Yang, G.-Z.. In: Hamlyn Symposium 2016. pp. 5-6 (2016)

sual servoing guidance, tumour segmentation and reconstruction using the ultrasound images.

A.1. Materials and Methods

This work made use of a da Vinci® PSM via the dVRK [132] that provides kinematic control of the robot. A stereo laparoscope system provides SD (720x576) video streaming for both left and right channels at 25 Hz. A UST-533 linear array ultrasound probe (Aloka Medical, Japan) has been mounted in a custom-made clip that can be picked-up by the robot using Cadiere forceps. A marker (KeyDot®, Key Surgical Inc, USA) was attached on the probe to assist visual tracking and pose estimation. A customised PVA cryogel kidney phantom was utilised with similar elastic and chromatic properties to human tissue.

The tissue surface is recovered from a disparity map using a stereo matching method [156]. The region for scanning is chosen by the user, and the scanning trajectory is planned adaptively to the 3D surface at the beginning of the task. To calculate the local poses (along the trajectory), which place the probe perpendicular to the surface, a local coordinate frame $\{\mathbf{S}\}$ on each surface point is defined using surface gradients and normals. To transform the trajectory from frame $\{\mathbf{S}\}$ to frame $\{\mathbf{M}\}$ (the marker frame), a matrix $\mathbf{T}_M^S$ is used and defined based on the calibrated transducer's position. These local poses also enforce that the centre of the transducer is aligned with the surface point.

For every step of visual control, a desired marker pose in frame $\{\mathbf{C}\}$ is calculated as: $\mathbf{T}_C^{M*} = \mathbf{T}_C^{S*} \cdot \mathbf{T}_S^M$, where $\mathbf{T}_C^{S*}$ is current desired transformation from frame $\{\mathbf{C}\}$ to $\{\mathbf{S}\}$. As shown in Fig. A.1, a desired robot command $\mathbf{T}_B^{E*}$ is calculated based on the current and desired marker pose, $\mathbf{T}_C^M$ and $\mathbf{T}_C^{M*}$, in a camera frame $\{\mathbf{C}\}$ via: $\mathbf{T}_B^{E*} = \mathbf{T}_B^{E*} \cdot \mathbf{T}_E^M \cdot \mathbf{T}_M^C \cdot \mathbf{T}_C^{M*} \cdot \mathbf{T}_M^E$. The transformation between end-effector frame $\{\mathbf{E}\}$ and marker frame $\{\mathbf{M}\}$ is a constant which can be either measured or calibrated using a standard hand-eye calibration method. For every

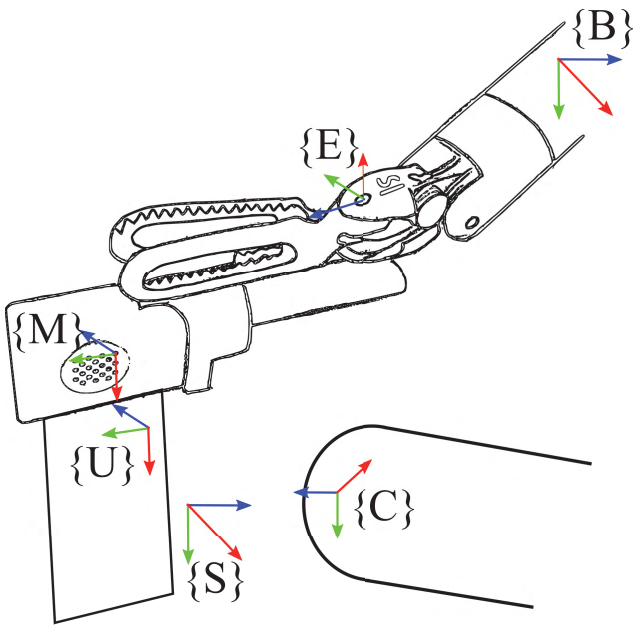


Figure A.1.: Coordinate frames defined in this work: surface frame {**S**}, marker frame {**M**}, end-effector frame {**E**}, robot base frame {**B**}, ultrasound image frame {**U**} and camera frame {**C**}.

image in a video sequence, the marker is detected using a circular-grid detection method followed by a pose estimation step [145] that calculates the marker's pose $\mathbf{T}_C^M$ in frame {**C**}. In order to improve the marker detection rate, this work has included a tracking component based on pyramidal optical flow [130] to track the circular grid along time.

To segment ultrasound images that contain tumours, speckle noises are removed by applying a Butterworth second order filter followed by an active contour [182] segmentation given a manually chosen seed region. The boundary and holes of the segmented tumour region are smoothed and filled by a set of morphological operations. The boundary is represented as a number of points where each point has correspondence across ultrasound images. In order to reconstruct a 3D model of the tumour using the ultrasound images, the segmented

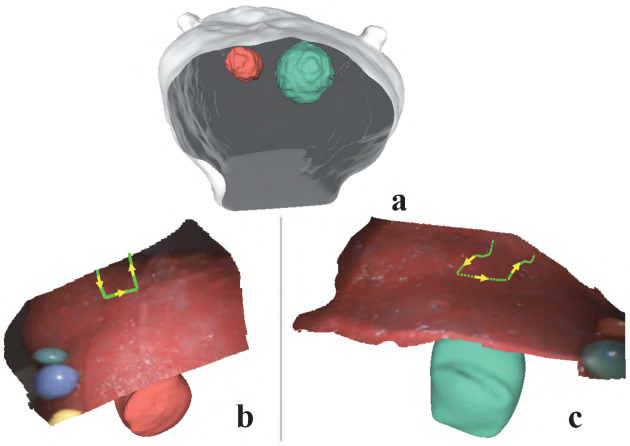


Figure A.2.: A 3D visual comparison between the CT ground truth (a) and reconstructed models of proposed framework (b-c). The 10mm and 15mm tumours are coloured in red and green respectively. The green points in b-c show planned trajectory for the ultrasound scanning, where the yellow arrows indicate the scanning direction.

boundary points need to be transformed from a 2D image coordinate frame to the 3D camera frame $\{\mathbf{C}\}$. To this end, an ultrasound coordinate frame $\{\mathbf{U}\}$ is defined in which each pixel of the ultrasound image can be represented as a 3D point $\mathbf{p}^U$. For N segmented boundary points $\mathbf{p}^U(i)$ are then transformed to frame $\{\mathbf{C}\}$ via: $\mathbf{p}^C(i) = \mathbf{T}_C^M \cdot \mathbf{T}_M^U \cdot \mathbf{p}^U(i), i \in [1, N]$, where $\mathbf{T}_M^U$ can be found via calibration and measurement. A 3D model of the tumour can then be created by connecting all the corresponding boundary points.

A.2. Results

To validate the overall accuracy of the proposed framework, two spherical tumours (diameter: 10mm and 15mm) are implanted in the kidney phantom. The ground truth (position and size) of the tumours is obtained from a CT model of the phantom. Fiducial markers are placed on the phantom in order to assist validation. Each tumour is scanned

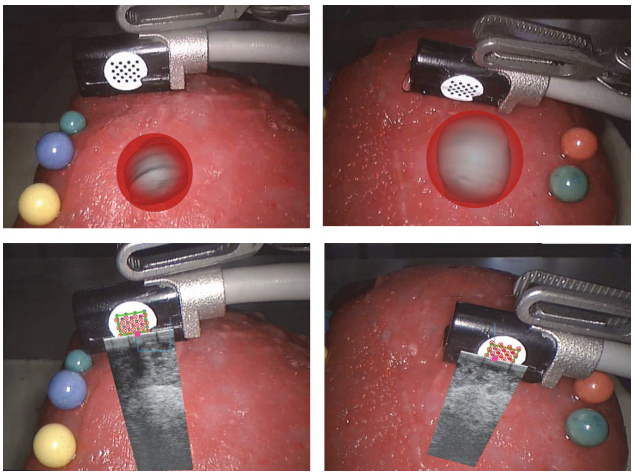


Figure A.3.: Top row: the reconstructed tumour is overlaid on laparoscopic images to provide an augmented view for tumour localisation and diagnosis. Bottom row: ultrasound image is overlaid according to current position of the probe. Left and right column show examples of the 10mm and 15mm tumour respectively.

autonomously while ultrasound images with pose information $\mathbf{T}_C^U$ are recorded. As shown in Fig. A.2, two tumour models are reconstructed from the ultrasound images using the described method. To compare the reconstructed models with the ground truth, spheres are fit to the CT and reconstructed models, such that the centre position and diameter can be calculated. The errors of centre position are 2.70mm and 2.80mm while the errors of diameter are 0.57mm and 0.43mm for 10mm and 15mm tumours, respectively.

Here two AR methods of visualisation are provided to assist tumour localisation and diagnosis. As shown in the top row of Fig. A.3, reconstructed tumour models are overlaid on the laparoscopic images, displaying the location and size of the tumour relative to the phantom. In order to have better perception of depth and shape of the tumour, inverse realism [7] technique is used. The second visualisation is shown in the bottom row where an ultrasound image is projected and drawn on the laparoscopic image in real-time. The advantage is to provide

surgeons both laparoscopic and ultrasound view simultaneously during autonomous scan without distraction. For a 15x10mm region, average scanning time is approximately 90 seconds, which is mainly affected by the scanning path planned at the beginning of the task.

A.3. Discussion

This work proposed a supervised autonomous ultrasound scanning framework for the robot-assisted laparoscopic surgery. It provides smooth scanning motion and maintains optimal orientation using the reconstructed surface. By using stereo vision for motion planning, robust marker detection and tracking, the robot is able to conduct a smooth ultrasound scanning on a non-planar tissue surface autonomously. Different visualisation techniques have been used to present the scanning results of tumours. Both qualitative and quantitative results have demonstrated the feasibility of the framework to detect tumour size and position accurately. Future work will focus on considering tissue deformation and testing the proposed framework in *in vivo* user studies and further improving the surgical AR visualisation.

B. Vision-guided Learning by Demonstration for Adaptive Surgical Robot Control †

Autonomous robot-assisted execution of repetitive surgical tasks has the potential to reduce fatigue, whilst improving surgical navigation, procedure times and workspace usage. The increased use of telesurgical robots and development of open-source platforms offers new opportunities for autonomous navigation which has led to a growing interest in applying the learning-by-demonstration (LbD) framework used in robotics towards complete automation of repetitive tasks [183, 107]. While vision information has been used towards tool tracking and target detection for autonomous task execution [94], the combination of LbD with vision-based 3D landmark recognition and tool tracking towards adaptive trajectory transfer has not been explored. This work proposes a LbD framework for automating surgical robotic execution of subtasks under stereo vision guidance, using the dVRK. The approach consists of learning optimum motion trajectories for different subtasks of a suturing task, and using stereo vision for 3D anatomical landmark recognition and needle tracking, to allow adap-

†Content from this chapter was published as:
Vision-guided Learning by Demonstration for Adaptive Surgical Robot Control. Rafii-Tari, H., Vandini, A., Zhang, L., Hughes-Hallett, A., Yang, G.-Z.. In: Hamlyn Symposium. pp. 39-40 (2016)

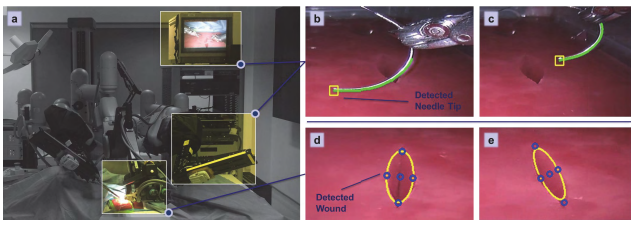


Figure B.1.: Experimental setup of the da Vinci® system with the dVRK (a), needle tracking for moving the needle from initial position (b) to desired target insertion point (c), wound detection under different camera views (d, e).

tation of learned trajectories to different clinical scenarios with moving camera views.

B.1. Materials and Methods

This work made use of the dVRK provided by Intuitive Surgical with the software implementation based on the cisst/SAW libraries [132]. Two large needle driver tools were used at the patient side manipulator (Fig. B.1). A custom-made PVA cryogel phantom (providing similar elastic and tensile properties as normal tissue) with customised built-in wounds (width=6mm) was made for these experiments. Motion models were learned from multiple experienced demonstrations (n=6,>10 simulator/surgical procedures) performing a running suture with three needle insertions. The suturing task is considered to consist of a set of subtasks to be modelled individually and executed by the robot automatically in a sequential order. However, the focus of this study is "**Subtask 1**": needle insertion through the wound with the right instrument", which for this task can be considered the more dexterous subtask. This subtask is preceded by an initialisation step involving moving the needle to the correct insertion point by the right instrument.

Each demonstration consisting of the 3D Cartesian positions and the rotations (represented as 4D quaternions) of the needle driver tip

was manually segmented and trajectories from multiple demonstrations were temporally aligned using a combination of dynamic time warping and spherical linear interpolation. For learning the movement patterns and reproducing trajectories that capture the main features of the training data, this work makes use of Hidden Semi-markov Models (HSMM) [184], a form of explicit-duration Hidden Markov Model (HMM), where the self-transition probability of the standard HMM is replaced with a parametric model of the state duration, thereby providing a better balance between temporal and spatial constraints than a standard HMM. As a result, a smooth generalised form of the motion for the desired subtask was learned.

The proposed vision algorithm is used for two main purposes: 1) wound detection for extracting the geometry and orientation of the wound under different anatomical variations/camera views 2) needle tracking in order to provide real-time 3D positioning of its tip.

In order to move the needle from an arbitrary initial position to desired target insertion point (Fig. B.1), after which automatic execution of the learned trajectories for Subtask 1 can be performed, the inverse kinematics of the robot is used. The transformation between the coordinate system at the base of the robot and at the stereo endoscopic camera was calculated offline through hand-eye calibration.

Initially, the wound is detected in the stereo images through a semi-automatic algorithm that asks the user to provide two input points and geometrically models the wound with an ellipse, while a feature image based on the analysis of the eigenvalues of the Hessian matrix is computed to enhance the borders of the wound. Once the 3D pose of the wound is estimated through triangulation, the needle is driven towards the desired target insertion point using the robot's inverse kinematics. Needle tracking is achieved through an algorithm that detects small straight segments on a feature image and only considers segments close to the detected needle in previous frames and with similar orientation. The segments are combined to create a continuous curve that represents the needle shape in the current frame. Using these results the

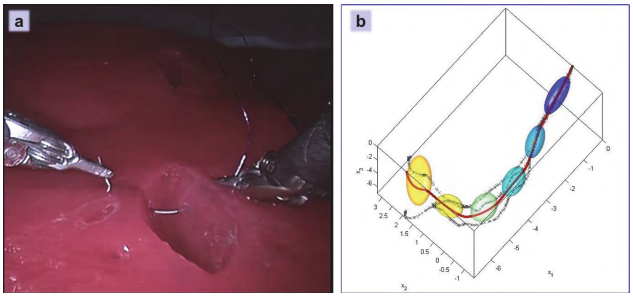


Figure B.2.: The running suture task (a) and the learned motion for "Subtask 1" (b) as reproduced from multiple demonstrations using the HSMM. The ellipses correspond to the covariances for each state of the HSMM.

3D needle tip position is estimated through triangulation.

B.2. Results

Fig. B.2 shows the running suture task (a) as well as the HSMM structure for Subtask 1, as learned from multiple experienced demonstrations, and the smooth generalised version of the motion extracted (b), with different colours corresponding to the different states of the HSMM. The optimum number of states was selected using the Bayesian information criterion.

The proposed method has been validated by automating the needle initialisation towards the target insertion point, using the vision system, followed by automated playback of the learned trajectory for Subtask 1. The whole execution was considered successful if the needle tip passed through the wound, and exited on the other side, enough to be grasped by the left instrument. The automation has been tested with the wound under different camera views, ranging from a viewing angle of -10 degrees to +10 degrees with respect to the camera, as well as different initial poses and distances of the needle with respect to the wound. The algorithm vas validated multiple times (n=7) under these different initial conditions. In each case the wound detection

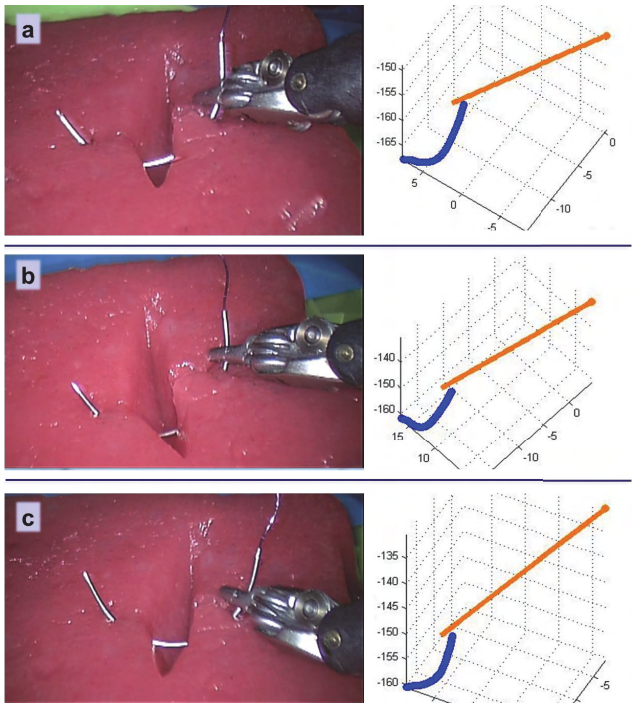


Figure B.3.: Examples of the needle driver tip trajectory for automated initialisation of the needle at the desired insertion point (orange) followed by automated playback of the learned motion for the first subtask (blue) under different initial conditions in the robot base coordinate system [mm].

algorithm estimates the 3D pose of the anatomy which was used to drive the needle towards the desired target insertion point. Fig. B.3 shows examples of the needle driver tip trajectory for automated initialisation of the needle at the desired insertion point in the robot base coordinate system (depicted in orange), followed by automated playback of the learned motion for Subtask 1 (depicted in blue). The initialisation step is sensitive to the accuracy of the wound detection, the robot kinematics, and the hand-eye calibration. The results depict successful and smooth execution of the desired motion under different initial conditions.

B.3. Discussion

This work proposes a LbD approach, using the dVRK, for learning and automating motion trajectories for different subtasks of a suturing task under vision guidance. By using stereo vision for 3D perception, anatomical landmark detection and needle tracking, the approach addresses different clinical scenarios under varying initial conditions and camera views, thereby allowing for subject-specific variability and tissue deformation. Future work will further extend the framework for adapting the learned trajectories to more complex anatomical variations by combining the needle tracking with visual servoing, to further overcome the inherent errors in the needle initialisation.

C. Image Re-Use Permissions

IEEE Images

All images replicated in this thesis from IEEE publications have been referenced according to the following statement from the IEEE:

"The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, the following statement may be used as a permission grant:
In the case of illustrations or tabular material, we require that the copyright line ©[Year of original publication] IEEE appear prominently with each reprinted figure and/or table".

Springer Images

The images replicated in this thesis from Springer publications have been referenced according to the following format: (©[year of publication] reproduced from [citation to original article], with permission from Springer). Individual licenses for the reused images have been obtained from Springer.

- Fig. 2.1, Springer licence number: 4214670900675
- Fig. 2.3, Springer licence number: 4198370405291
- Fig. 2.12, Springer licence number: 4198241079228
- Fig. 8.1, Springer licence number: 4214661409634

Elsevier Images

The images replicated from Elsevier have been referenced according to the following format: ©[year of publication] Elsevier [citation to original article]. Individual licenses for each of the reused images have been obtained from Elsevier.

- Fig. 2.2, Elsevier licence number: 4214670197372
- Fig. 2.4, Elsevier licence number: 4214670032816
- Fig. 2.6, Elsevier licence number: 4196761358649
- Fig. 2.7, Elsevier licence number: 4197031506739
- Fig. 2.14, Elsevier licence number: 4214670809943

John Wiley and Sons Images

The images reused in this thesis from John Wiley and Sons have been referenced according to the following format: ©[year of publication] John Wiley & Sons [citation to original article]. Individual licenses for each of the reused images have been obtained from John Wiley and Sons.

- Fig. 2.5, John Wiley and Sons licence number: 4196671385959
- Fig. 2.15, John Wiley and Sons licence number: 4214670427650

Images from Individual Researchers

- Fig. 2.10 is obtained from Jakob Unger at University of California, Davis.

Bibliography

[1] Liu, W. P., Reaugamornrat, S., Deguet, A., Sorger, J. M., Siewerdsen, J. H., Richmon, J., and Taylor, R. H., "Toward intraoperative image-guided transoral robotic surgery," *Journal of robotic surgery*, vol. 7, no. 3, pp. 217–225, (2013).

[2] Sturm, L., Dawson, D., Vaughan, R., Hewett, P., Hill, A. G., Graham, J. C., and Maddern, G. J., "Effects of fatigue on surgeon performance and surgical outcomes: a systematic review," *ANZ journal of surgery*, vol. 81, no. 7-8, pp. 502–509, (2011).

[3] Pezzementi, Z., Okamura, A. M., and Hager, G. D., "Dynamic guidance with pseudoadmittance virtual fixtures," in *Robotics and Automation, 2007 IEEE International Conference on*, pp. 1761–1767, IEEE, (2007).

[4] Li, M. and Taylor, R. H., "Performance of surgical robots with automatically generated spatial virtual fixtures," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, pp. 217–222, IEEE, (2005).

[5] Davies, B., Jakopec, M., Harris, S. J., Baena, F. R. Y., Barrett, A., Evangelidis, A., Gomes, P., Henckel, J., and Cobb, J., "Active-constraint robotics for surgery," *Proceedings of the IEEE*, vol. 94, no. 9, pp. 1696–1704, (2006).

[6] Pratt, P., Mayer, E., Vale, J., Cohen, D., Edwards, E., Darzi, A., and Yang, G.-Z., "An effective visualisation and registration system for image-guided robotic partial nephrectomy," *Journal of Robotic Surgery*, vol. 6, no. 1, pp. 23–31, (2012).

[7] Lerotic, M., Chung, A. J., Mylonas, G., and Yang, G.-Z., "Pq-space based non-photorealistic rendering for augmented reality," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 102–109, Springer, (2007).

[8] Bernhardt, S., Nicolau, S. A., Agnus, V., Soler, L., Doignon, C., and Marescaux, J., "Automatic localization of endoscope in intraoperative ct image: a simple approach to augmented reality guidance in laparoscopic surgery," *Medical image analysis*, vol. 30, pp. 130–143, (2016).

[9] Hughes-Hallett, A., Pratt, P., Mayer, E., Di Marco, A., Yang, G.-Z., Vale, J., and Darzi, A., "Intraoperative ultrasound overlay in robot-assist partial nephrectomy: first clinical experience," *European urology*, vol. 3, no. 65, pp. 671–672, (2014).

[10] John, T. G., Greig, J. D., Crosbie, J. L., Miles, W., and Garden, O. J., "Superior staging of liver tumors with laparoscopy and laparoscopic ultrasound.," *Annals of surgery*, vol. 220, no. 6, p. 711, (1994).

[11] Barbot, D. J., Marks, J. H., Feld, R. I., Liu, J.-B. B., and Rosato, F. E., "Improved staging of liver tumors using laparoscopic intraoperative ultrasound," *Journal of surgical oncology*, vol. 64, no. 1, pp. 63–67, (1997).

[12] Wu, J. S., Dunnegan, D. L., and Soper, N. J., "The utility of intracorporeal ultrasonography for screening of the bile duct during laparoscopic cholecystectomy," *Journal of Gastrointestinal Surgery*, vol. 2, no. 1, pp. 50–60, (1998).

[13] Catheline, J., Turner, R., Rizk, N., Barrat, C., and Champault, G., "The use of diagnostic laparoscopy supported by laparoscopic ultrasonography in the assessment of pancreatic cancer," *Surgical endoscopy*, vol. 13, no. 3, pp. 239–245, (1999).

[14] Su, L.-M., Vagvolgyi, B. P., Agarwal, R., Reiley, C. E., Taylor, R. H., and Hager, G. D., "Augmented reality during robot-assisted laparoscopic partial nephrectomy: toward real-time 3d-ct to stereoscopic video registration," *Urology*, vol. 73, no. 4, pp. 896–900, (2009).

[15] Kaczmarek, B. F., Sukumar, S., Kumar, R. K., Desa, N., Jost, K., Diaz, M., Menon, M., and Rogers, C. G., "Comparison of robotic and laparoscopic ultrasound probes for robotic partial nephrectomy," *Journal of endourology*, vol. 27, no. 9, pp. 1137–1140, (2013).

[16] Langø, T., Vijayan, S., Rethy, A., Våpenstad, C., Solberg, O. V., Mårvik, R., Johnsen, G., and Hernes, T. N., "Navigated laparoscopic ultrasound in abdominal soft tissue surgery: technological overview and perspectives," *International journal of computer assisted radiology and surgery*, vol. 7, no. 4, pp. 585–599, (2012).

[17] Kikuchi, H., Kamiya, K., Hiramatsu, Y., Miyazaki, S., Yamamoto, M., Ohta, M., Baba, S., and Konno, H., "Laparoscopic narrow-band imaging for the diagnosis of peritoneal metastasis in gastric cancer," *Annals of surgical oncology*, vol. 21, no. 12, pp. 3954–3962, (2014).

[18] Tateya, I., Ishikawa, S., Morita, S., Ito, H., Sakamoto, T., Murayama, T., Kishimoto, Y., Hayashi, T., Funakoshi, M., Hirano, S., and others,, "Magnifying endoscopy with narrow band imaging to determine the extent of resection in transoral robotic surgery of oropharyngeal cancer," *Case reports in otolaryngology*, vol. 2014, (2014).

[19] Diana, M., Usmaan, H., Legnèr, A., Yu-Yin, L., D'Urso, A., Halvax, P., Nagao, Y., Pessaux, P., and Marescaux, J., "Novel laparoscopic narrow band imaging for real-time detection of bile

leak during hepatectomy: proof of the concept in a porcine model," *Surgical endoscopy*, vol. 30, no. 7, pp. 3128–3132, (2016).

[20] Gagliardi, M. L., Polito, S., Fagotti, A., Fanfani, F., and Scambia, G., "Narrow-band imaging in laparoscopic management of recurrent platinum sensitive ovarian cancer," *Journal of minimally invasive gynecology*, vol. 20, no. 1, pp. 10–12, (2013).

[21] Bernhardt, S., Nicolau, S. A., Soler, L., and Doignon, C., "The status of augmented reality in laparoscopic surgery as of 2016," *Medical image analysis*, vol. 37, pp. 66–90, (2017).

[22] Schols, R. M., Connell, N. J., and Stassen, L. P., "Near-infrared fluorescence imaging for real-time intraoperative anatomical guidance in minimally invasive surgery: a systematic review of the literature," *World journal of surgery*, vol. 39, no. 5, pp. 1069–1079, (2015).

[23] Tobis, S., Knopf, J., Silvers, C., Yao, J., Rashid, H., Wu, G., and Golijanin, D., "Near infrared fluorescence imaging with robotic assisted laparoscopic partial nephrectomy: initial clinical experience for renal cortical tumors," *The Journal of urology*, vol. 186, no. 1, pp. 47–52, (2011).

[24] Fatakdawala, H., Gorpas, D., Bishop, J. W., Bec, J., Ma, D., Southard, J. A., Margulies, K. B., and Marcu, L., "Fluorescence lifetime imaging combined with conventional intravascular ultrasound for enhanced assessment of atherosclerotic plaques: an ex vivo study in human coronary arteries," *Journal of cardiovascular translational research*, vol. 8, no. 4, pp. 253–263, (2015).

[25] Patsias, A., Giraldez-Rodriguez, L., Polydorides, A. D., Richards-Kortum, R., Anandasabapathy, S., Quang, T., Sikora, A. G., and Miles, B., "Feasibility of transoral robotic-assisted high-resolution microendoscopic imaging of oropharyngeal squamous cell carcinoma," *Head & neck*, vol. 37, no. 8, (2015).

[26] Meining, A., Shah, R., Slivka, A., Pleskow, D., Chuttani, R., Stevens, P., Becker, V., and Chen, Y., "Classification of probe-based confocal laser endomicroscopy findings in pancreaticobiliary strictures," *Endoscopy*, vol. 44, no. 03, pp. 251–257, (2012).

[27] Shieh, F. K., Drumm, H., Nathanson, M. H., and Jamidar, P. A., "High-definition confocal endomicroscopy of the common bile duct," *Journal of clinical gastroenterology*, vol. 46, no. 5, p. 401, (2012).

[28] Delius, S. v., Feussner, H., Wilhelm, D., Karagianni, A., Henke, J., Schmid, R., and Meining, A., "Transgastric in vivo histology in the peritoneal cavity using miniprobe-based confocal fluorescence microscopy in an acute porcine model," *Endoscopy*, vol. 39, no. 05, pp. 407–411, (2007).

[29] Kantsevoy, S. V., Kiesslich, R., Dray, X., Magno, P., Jagannath, S. B., Buscaglia, J. M., Goetz, M., Dunbar, K. B., Assumpcao, L. R., Canto, M. I., and others„ "Translumenal intraperitoneal confocal laser endomicroscopy: a pilot study in a porcine model," *Gastrointestinal Endoscopy*, vol. 67, no. 5, p. AB117, (2008).

[30] Schneider, C., Johnson, S. P., Walker-Samuel, S., Gurusamy, K., Clarkson, M. J., Thompson, S., Song, Y., Totz, J., Cook, R. J., Desjardins, A. E., and others„ "Utilizing confocal laser endomicroscopy for evaluating the adequacy of laparoscopic liver ablation," *Lasers in surgery and medicine*, vol. 48, no. 3, pp. 299–310, (2016).

[31] Haxel, B. R., Goetz, M., Kiesslich, R., and Gosepath, J., "Confocal endomicroscopy: a novel application for imaging of oral and oropharyngeal mucosa in human," *European Archives of Oto-Rhino-Laryngology*, vol. 267, no. 3, pp. 443–448, (2010).

[32] Pogorzelski, B., Hanenkamp, U., Goetz, M., Kiesslich, R., and Gosepath, J., "Systematic intraoperative application of confocal

endomicroscopy for early detection and resection of squamous cell carcinoma of the head and neck: a preliminary report," *Archives of Otolaryngology–Head & Neck Surgery*, vol. 138, no. 4, pp. 404–411, (2012).

[33] Tan, J., Quinn, M., Pyman, J., Delaney, P., and McLaren, W., "Detection of cervical intraepithelial neoplasia in vivo using confocal endomicroscopy," *BJOG: An International Journal of Obstetrics & Gynaecology*, vol. 116, no. 12, pp. 1663–1670, (2009).

[34] Foersch, S., Heimann, A., Ayyad, A., Spoden, G. A., Florin, L., Mpoukouvalas, K., Kiesslich, R., Kempski, O., Goetz, M., and Charalampaki, P., "Confocal laser endomicroscopy for diagnosis and histomorphologic imaging of brain tumors in vivo," *PLoS One*, vol. 7, no. 7, p. e41760, (2012).

[35] Eschbacher, J., Martirosyan, N. L., Nakaji, P., Sanai, N., Preul, M. C., Smith, K. A., Coons, S. W., and Spetzler, R. F., "In vivo intraoperative confocal microscopy for real-time histopathological imaging of brain tumors," *Journal of neurosurgery*, vol. 116, no. 4, pp. 854–860, (2012).

[36] Martirosyan, N. L., Cavalcanti, D. D., Eschbacher, J. M., Delaney, P. M., Scheck, A. C., Abdelwahab, M. G., Nakaji, P., Spetzler, R. F., and Preul, M. C., "Use of in vivo near-infrared laser confocal endomicroscopy with indocyanine green to detect the boundary of infiltrative tumor," *Journal of neurosurgery*, vol. 115, no. 6, pp. 1131–1138, (2011).

[37] Rais-Bahrami, S., Levinson, A. W., Fried, N. M., Lagoda, G. A., Hristov, A., Chuang, Y., Burnett, A. L., and Su, L.-M., "Optical coherence tomography of cavernous nerves: a step toward real-time intraoperative imaging during nerve-sparing radical prostatectomy," *Urology*, vol. 72, no. 1, pp. 198–204, (2008).

[38] Hariri, L. P., Bonnema, G. T., Schmidt, K., Winkler, A. M., Korde, V., Hatch, K. D., Davis, J. R., Brewer, M. A., and Barton, J. K., "Laparoscopic optical coherence tomography imaging of human ovarian cancer," *Gynecologic oncology*, vol. 114, no. 2, pp. 188–194, (2009).

[39] Al-Mujaini, A., Wali, U. K., and Azeem, S., "Optical coherence tomography: clinical applications in medical practice," *Oman medical journal*, vol. 28, no. 2, p. 86, (2013).

[40] Goel, R. K. and Kaouk, J. H., "Optical coherence tomography: the past, present and future," *Journal of Robotic Surgery*, vol. 1, no. 3, pp. 179–184, (2007).

[41] Boppart, S. A., Herrmann, J. M., Pitris, C., Stamper, D. L., Brezinski, M. E., and Fujimoto, J. G., "Real-time optical coherence tomography for minimally invasive imaging of prostate ablation," *Computer Aided Surgery*, vol. 6, no. 2, pp. 94–103, (2001).

[42] Aron, M., Kaouk, J. H., Hegarty, N. J., Colombo, Jr, J. R., Haber, G.-P., Chung, B. I., Zhou, M., and Gill, I. S., "Second prize: Preliminary experience with the niris™ optical coherence tomography system during laparoscopic and robotic prostatectomy," *Journal of endourology*, vol. 21, no. 8, pp. 814–818, (2007).

[43] Chung, B., Tresser, N., Kareta, K., and others„ "Niris optical coherence tomography system. principles of operation and applications in partial nephrectomy," in *Abstract of the 5th international kidney cancer symposium, Chicago*, (2006).

[44] Beaudette, K., Strupler, M., Driscoll, M., Maciejko, R., Aubin, C., and Boudoux, C., "Towards a handheld probe based on optical coherence tomography for minimally invasive spine surgeries," *Stud Health Technol Inform*, vol. 158, pp. 49–54, (2010).

[45] Boppart, S. A., "Optical coherence tomography: technology and applications for neuroimaging," *Psychophysiology*, vol. 40, no. 4, pp. 529–541, (2003).

[46] Gora, M. J., Sauk, J. S., Carruth, R. W., Gallagher, K. A., Suter, M. J., Nishioka, N. S., Kava, L. E., Rosenberg, M., Bouma, B. E., and Tearney, G. J., "Tethered capsule endomicroscopy enables less invasive imaging of gastrointestinal tract microstructure," *Nature medicine*, vol. 19, no. 2, pp. 238–240, (2013).

[47] Boppart, S. A., Luo, W., Marks, D. L., and Singletary, K. W., "Optical coherence tomography: feasibility for basic research and image-guided surgery of breast cancer," *Breast cancer research and treatment*, vol. 84, no. 2, pp. 85–97, (2004).

[48] Yoo, H., Kim, J. W., Shishkov, M., Namati, E., Morse, T., Shubochkin, R., McCarthy, J. R., Ntziachristos, V., Bouma, B. E., Jaffer, F. A., and others„ "Intra-arterial catheter for simultaneous microstructural and molecular imaging in vivo," *Nature medicine*, vol. 17, no. 12, pp. 1680–1684, (2011).

[49] Carrasco-Zevallos, O. M., Viehland, C., Keller, B., Draelos, M., Kuo, A. N., Toth, C. A., and Izatt, J. A., "Review of intraoperative optical coherence tomography: technology and applications," *Biomedical Optics Express*, vol. 8, no. 3, pp. 1607–1637, (2017).

[50] Dwyer, G. and others„ "A miniaturised robotic probe for real-time intraoperative fusion of ultrasound and endomicroscopy," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pp. 1196–1201, IEEE, (2015).

[51] Sherlock, B. E., Phipps, J. E., Bec, J., and Marcu, L., "Simultaneous, label-free, multispectral fluorescence lifetime imaging and optical coherence tomography using a double-clad fiber," *Optics Letters*, vol. 42, no. 19, pp. 3753–3756, (2017).

[52] Stenau, E., Simpfendorfer, T., Azizian, M., Liu, W. P., Reiter, K. A., Bergman, D., Mohareri, O., Sorger, J., Teber, D., and Maier-Hein, L., "In vivo robustness analyses of intraoperative registration with fluorescent markers," in *Annual Conference of the International Society for Computer Aided Surgery*, (2017).

[53] Wild, E., Teber, D., Schmid, D., Simpfendörfer, T., Müller, M., Baranski, A.-C., Kenngott, H., Kopka, K., and Maier-Hein, L., "Robust augmented reality guidance with fluorescent markers in laparoscopic surgery," *International journal of computer assisted radiology and surgery*, vol. 11, no. 6, pp. 899–907, (2016).

[54] Barratt, D. C., Davies, A. H., Hughes, A. D., Thom, S. A., and Humphries, K. N., "Optimisation and evaluation of an electromagnetic tracking device for high-accuracy three-dimensional ultrasound imaging of the carotid arteries," *Ultrasound in medicine & biology*, vol. 27, no. 7, pp. 957–968, (2001).

[55] Condino, S., Ferrari, V., Freschi, C., Alberti, A., Berchiolli, R., Mosca, F., and Ferrari, M., "Electromagnetic navigation platform for endovascular surgery: how to develop sensorized catheters and guidewires," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 8, no. 3, pp. 300–310, (2012).

[56] Lin, A., Trejos, A., Mohan, S., Bassan, H., Kashigar, A., Patel, R., and Malthaner, R., "Electromagnetic navigation improves minimally invasive robot-assisted lung brachytherapy," *Computer Aided Surgery*, vol. 13, no. 2, pp. 114–123, (2008).

[57] Christie, S., "Electromagnetic navigational bronchoscopy and robotic-assisted thoracic surgery," *AORN journal*, vol. 99, no. 6, pp. 750–763, (2014).

[58] Franz, A. M., Haidegger, T., Birkfellner, W., Cleary, K., Peters, T. M., and Maier-Hein, L., "Electromagnetic tracking in

medicine—a review of technology, validation, and applications," *IEEE transactions on medical imaging*, vol. 33, no. 8, pp. 1702–1725, (2014).

[59] Konishi, K., Nakamoto, M., Kakeji, Y., Tanoue, K., Kawanaka, H., Yamaguchi, S., Ieiri, S., Sato, Y., Maehara, Y., Tamura, S., and others„ "A real-time navigation system for laparoscopic surgery based on three-dimensional ultrasound using magneto-optic hybrid tracking configuration," *International Journal of Computer Assisted Radiology and Surgery*, vol. 2, no. 1, pp. 1–10, (2007).

[60] Ukimura, O. and Gill, I. S., "Imaging-assisted endoscopic surgery: Cleveland clinic experience," *Journal of Endourology*, vol. 22, no. 4, pp. 803–810, (2008).

[61] Elfring, R., de la Fuente, M., and Radermacher, K., "Assessment of optical localizer accuracy for computer aided surgery systems," *Computer Aided Surgery*, vol. 15, no. 1-3, pp. 1–12, (2010).

[62] Stoll, J., Novotny, P., Howe, R., and Dupont, P., "Real-time 3d ultrasound-based servoing of a surgical instrument," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pp. 613–618, IEEE, (2006).

[63] Jayarathne, U. L., McLeod, A. J., Peters, T. M., and Chen, E. C., "Robust intraoperative us probe tracking using a monocular endoscopic camera," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 363–370, Springer, (2013).

[64] Pratt, P., Jaeger, A., Hughes-Hallett, A., Mayer, E., Vale, J., Darzi, A., Peters, T., and Yang, G.-Z., "Robust ultrasound probe tracking: initial clinical experiences during robot-assisted partial nephrectomy," *International journal of computer assisted radiology and surgery*, vol. 10, no. 12, pp. 1905–1913, (2015).

[65] Zhao, T., Zhao, W., Halabe, D. J., Hoffman, B. D., and Nowlin, W. C., "Fiducial marker design and detection for locating surgical instrument in images," (Dec. 27 2016). US Patent 9,526,587.

[66] Bouget, D., Allan, M., Stoyanov, D., and Jannin, P., "Vision-based and marker-less surgical tool detection and tracking: a review of the literature," *Medical image analysis*, vol. 35, pp. 633–654, (2017).

[67] Reiter, A., Allen, P. K., and Zhao, T., "Articulated surgical tool detection using virtually-rendered templates," in *Computer Assisted Radiology and Surgery (CARS)*, (2012).

[68] Ye, M., Zhang, L., Giannarou, S., and Yang, G.-Z., "Real-time 3d tracking of articulated tools for robotic surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 386–394, Springer, (2016).

[69] Feuerstein, M., Reichl, T., Vogel, J., Schneider, A., Feussner, H., and Navab, N., "Magneto-optic tracking of a flexible laparoscopic ultrasound transducer for laparoscope augmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 458–466, Springer, (2007).

[70] Kurmann, T., Neila, P. M., Du, X., Fua, P., Stoyanov, D., Wolf, S., and Sznitman, R., "Simultaneous recognition and pose estimation of instruments in minimally invasive surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 505–513, Springer, (2017).

[71] Nicolau, S., Soler, L., Mutter, D., and Marescaux, J., "Augmented reality in laparoscopic surgical oncology," *Surgical oncology*, vol. 20, no. 3, pp. 189–201, (2011).

[72] Schroeder, W. J., Lorensen, B., and Martin, K., *The visualization toolkit: an object-oriented approach to 3D graphics*. Kitware, (2004).

[73] Nolden, M., Zelzer, S., Seitel, A., Wald, D., Muller, M., Franz, A. M., Maleike, D., Fangerau, M., Baumhauer, M., Maier-Hein, L., and others„ "The medical imaging interaction toolkit: challenges and advances," *International journal of computer assisted radiology and surgery*, vol. 8, no. 4, pp. 607–620, (2013).

[74] Kikinis, R., Pieper, S. D., and Vosburgh, K. G., "3d slicer: a platform for subject-specific image analysis, visualization, and clinical support," in *Intraoperative imaging and image-guided therapy*, pp. 277–289, Springer, (2014).

[75] Gary, K., Ibanez, L., Aylward, S., Gobbi, D., Blake, M. B., and Cleary, K., "Igstk: an open source software toolkit for image-guided surgery," *Computer*, vol. 39, no. 4, pp. 46–53, (2006).

[76] Askeland, C., Solberg, O. V., Bakeng, J. B. L., Reinertsen, I., Tangen, G. A., Hofstad, E. F., Iversen, D. H., Våpenstad, C., Selbekk, T., Langø, T., and others„ "Custusx: an open-source research platform for image-guided therapy," *International journal of computer assisted radiology and surgery*, vol. 11, no. 4, pp. 505–519, (2016).

[77] Yi, B., Wang, G., Li, J., Jiang, J., Son, Z., Su, H., Zhu, S., and Wang, S., "Domestically produced chinese minimally invasive surgical robot system "micro hand s" is applied to clinical surgery preliminarily in china," *Surgical endoscopy*, vol. 31, no. 1, pp. 487–493, (2017).

[78] Lim, J. H., Lee, W. J., Park, D. W., Yea, H. J., Kim, S. H., and Kang, C. M., "Robotic cholecystectomy using revo-i model msr-5000, the newly developed korean robotic surgical system: a

preclinical study," *Surgical endoscopy*, vol. 31, no. 8, pp. 3391–3397, (2017).

[79] Bozzini, G., Gidaro, S., and Taverna, G., "Robot-assisted laparoscopic partial nephrectomy with the alf–x robot on pig models," *European urology*, vol. 69, no. 2, pp. 376–377, (2016).

[80] Fanfani, F., Monterossi, G., Fagotti, A., Rossitto, C., Alletti, S. G., Costantini, B., Gallotta, V., Selvaggi, L., Restaino, S., and Scambia, G., "The new robotic telelap alf-x in gynecological surgery: single-center experience," *Surgical endoscopy*, vol. 30, no. 1, pp. 215–221, (2016).

[81] Gueli Alletti, S., Rossitto, C., Cianci, S., Perrone, E., Pizzacalla, S., Monterossi, G., Vizzielli, G., Gidaro, S., and Scambia, G., "The senhance™ surgical robotic system ("senhance") for total hysterectomy in obese patients: a pilot study," *Journal of Robotic Surgery*, pp. 1–6, (2017).

[82] Walker, A. S. and Steele, S. R., "The future of robotic instruments in colon and rectal surgery," in *Seminars in Colon and Rectal Surgery*, vol. 27, pp. 144–149, Elsevier, (2016).

[83] Schuler, P. J., Duvvuri, U., Friedrich, D. T., Rotter, N., Scheithauer, M. O., and Hoffmann, T. K., "First use of a computer-assisted operator-controlled flexible endoscope for transoral surgery," *The Laryngoscope*, vol. 125, no. 3, pp. 645–648, (2015).

[84] Mandapathil, M., Duvvuri, U., Güldner, C., Teymoortash, A., Lawson, G., and Werner, J. A., "Transoral surgery for oropharyngeal tumors using the medrobotics® flex® system–a case report," *International journal of surgery case reports*, vol. 10, pp. 173–175, (2015).

[85] Remacle, M., Prasad, V., Lawson, G., Plisson, L., Bachy, V., and Van der Vorst, S., "Transoral robotic surgery (tors) with the

medrobotics flex™ system: first surgical application on humans," *European Archives of Oto-Rhino-Laryngology*, vol. 272, no. 6, pp. 1451–1455, (2015).

[86] "Medrobotics® corporation announces first in human colorectal procedure."

[87] Naidu, A. S., Naish, M. D., and Patel, R. V., "A breakthrough in tumor localization: Combining tactile sensing and ultrasound to improve tumor localization in robotics-assisted minimally invasive surgery," *IEEE Robotics & Automation Magazine*, vol. 24, no. 2, pp. 54–62, (2017).

[88] Francis, P., Eastwood, K. W., Bodani, V., Price, K., Upadhyaya, K., Podolsky, D., Azimian, H., Looi, T., and Drake, J., "Miniaturized instruments for the da vinci research kit," *IEEE Robotics & Automation Magazine*, vol. 1070, no. 9932/17, (2017).

[89] Murali, A., Sen, S., Kehoe, B., Garg, A., McFarland, S., Patil, S., Boyd, W. D., Lim, S., Abbeel, P., and Goldberg, K., "Learning by observation for surgical subtasks: Multilateral cutting of 3d viscoelastic and 2d orthotropic tissue phantoms," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pp. 1202–1209, IEEE, (2015).

[90] Sen, S., Garg, A., Gealy, D. V., McKinley, S., Jen, Y., and Goldberg, K., "Automating multi-throw multilateral surgical suturing with a mechanical needle guide and sequential convex optimization," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*, pp. 4178–4185, IEEE, (2016).

[91] Zhang, L., Ye, M., Giataganas, P., Hughes, M., Bradu, A., Podoleanu, A., and Yang, G.-Z., "From macro to micro: Autonomous multiscale image fusion for robotic surgery," *IEEE Robotics & Automation Magazine*, (2017).

[92] Power, M., Rafii-Tari, H., Bergeles, C., Vitiello, V., and Yang, G.-Z., "A cooperative control framework for haptic guidance of bimanual surgical tasks based on learning from demonstration," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pp. 5330–5337, IEEE, (2015).

[93] Berthet-Rayne, P., Power, M., King, H., and Yang, G.-Z., "Hubot: A three state human-robot collaborative framework for bimanual surgical tasks based on learned models," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*, pp. 715–722, IEEE, (2016).

[94] Kehoe, B., Kahn, G., Mahler, J., Kim, J., Lee, A., Lee, A., Nakagawa, K., Patil, S., Boyd, W. D., Abbeel, P., and others„ "Autonomous multilateral debridement with the raven surgical robot," in *Robotics and Automation (ICRA), 2014 IEEE International Conference on*, pp. 1432–1439, IEEE, (2014).

[95] Hu, D., Gong, Y., Hannaford, B., and Seibel, E. J., "Semi-autonomous simulated brain tumor ablation with ravenii surgical robot using behavior tree," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pp. 3868–3875, IEEE, (2015).

[96] Glassman, D., White, L., Lewis, A., King, H. H., Clarke, A., Glassman, T., Comstock, B., Hannaford, B., and Lendvay, T. S., "Raven surgical robot training in preparation for da vinci® use: A randomized prospective trial.," in *MMVR*, pp. 135–141, (2014).

[97] Hannaford, B., Rosen, J., Friedman, D. W., King, H., Roan, P., Cheng, L., Glozman, D., Ma, J., Kosari, S. N., and White, L., "Raven-ii: an open platform for surgical robotics research," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 954–959, (2013).

[98] Tobergte, A., Konietschke, R., and Hirzinger, G., "Planning and control of a teleoperation system for research in minimally invasive robotic surgery," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pp. 4225–4232, IEEE, (2009).

[99] Hagn, U., Konietschke, R., Tobergte, A., Nickl, M., Jörg, S., Kübler, B., Passig, G., Gröger, M., Fröhlich, F., Seibold, U., and others„ "Dlr mirosurge: a versatile system for research in endoscopic telesurgery," *International journal of computer assisted radiology and surgery*, vol. 5, no. 2, pp. 183–193, (2010).

[100] Thielmann, S., Seibold, U., Haslinger, R., Passig, G., Bahls, T., Jörg, S., Nickl, M., Nothhelfer, A., Hagn, U., and Hirzinger, G., "Mica-a new generation of versatile instruments in robotic surgery," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pp. 871–878, IEEE, (2010).

[101] Wolf, A., Jaramaz, B., Lisien, B., and DiGioia, A., "Mbars: mini bone-attached robotic system for joint arthroplasty," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 1, no. 2, pp. 101–121, (2005).

[102] Manzey, D., Strauss, G., Trantakis, C., Lueth, T., Roettger, S., Bahner-Heyne, J., Dietz, A., and Meixensberger, J., "Automation in surgery: a systematic approach," *Surg Technol Int*, vol. 18, pp. 37–45, (2009).

[103] Yip, M. and Das, N., "Robot autonomy for surgery," *arXiv preprint arXiv:1707.03080*, (2017).

[104] Yang, G.-Z., Cambias, J., Cleary, K., Daimler, E., Drake, J., Dupont, P. E., Hata, N., Kazanzides, P., Martel, S., Patel, R. V., and others„ "Medical robotics—regulatory, ethical, and le-

gal considerations for increasing levels of autonomy," *Sci. Robot.*, vol. 2, p. eaam8638, (2017).

[105] Dogangil, G., Davies, B., and Rodriguez y Baena, F., "A review of medical robotics for minimally invasive soft tissue surgery," *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, vol. 224, no. 5, pp. 653–679, (2010).

[106] Kaplan, K. E., Nichols, K. A., and Okamura, A. M., "Toward human-robot collaboration in surgery: performance assessment of human and robotic agents in an inclusion segmentation task," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*, pp. 723–729, IEEE, (2016).

[107] Padoy, N. and Hager, G. D., "Human-machine collaborative surgery using learned models," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pp. 5285–5292, IEEE, (2011).

[108] Thananjeyan, B., Garg, A., Krishnan, S., Chen, C., Miller, L., and Goldberg, K., "Multilateral surgical pattern cutting in 2d orthotropic gauze with deep reinforcement learning policies for tensioning," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pp. 2371–2378, IEEE, (2017).

[109] Zhang, L., Ye, M., Giataganas, P., Hughes, M., and Yang, G.-Z., "Autonomous scanning for endomicroscopic mosaicing and 3d fusion," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, pp. 3587–3593, IEEE, (2017).

[110] Yuen, S. G., Kettler, D. T., Novotny, P. M., Plowes, R. D., and Howe, R. D., "Robotic motion compensation for beating heart intracardiac surgery," *The International journal of robotics research*, vol. 28, no. 10, pp. 1355–1372, (2009).

[111] Bebek, O. and Cavusoglu, M. C., "Predictive control algorithms using biological signals for active relative motion canceling in robotic assisted heart surgery," in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pp. 237–244, IEEE, (2006).

[112] Gangloff, J., Ginhoux, R., de Mathelin, M., Soler, L., and Marescaux, J., "Model predictive control for compensation of cyclic organ motions in teleoperated laparoscopic surgery," *IEEE Transactions on Control Systems Technology*, vol. 14, no. 2, pp. 235–246, (2006).

[113] Richa, R., Poignet, P., and Liu, C., "Efficient 3d tracking for motion compensation in beating heart surgery," *Medical image computing and computer-assisted intervention–MICCAI 2008*, pp. 684–691, (2008).

[114] Mountney, P., Stoyanov, D., and Yang, G.-Z., "Three-dimensional tissue deformation recovery and tracking," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 14–24, (2010).

[115] Yip, M. C., Lowe, D. G., Salcudean, S. E., Rohling, R. N., and Nguan, C. Y., "Tissue tracking and registration for image-guided surgery," *IEEE transactions on medical imaging*, vol. 31, no. 11, pp. 2169–2182, (2012).

[116] Ye, M., Giannarou, S., Meining, A., and Yang, G.-Z., "Online tracking and retargeting with applications to optical biopsy in gastrointestinal endoscopic examinations," *Medical image analysis*, vol. 30, pp. 144–157, (2016).

[117] Taylor, R. H., Mittelstadt, B. D., Paul, H. A., Hanson, W., Kazanzides, P., Zuhars, J. F., Williamson, B., Musits, B. L., Glassman, E., and Bargar, W. L., "An image-directed robotic system for precise orthopaedic surgery," *IEEE Transactions on Robotics and Automation*, vol. 10, no. 3, pp. 261–275, (1994).

[118] Antic, T. and Taxy, J. B., "Partial nephrectomy for renal tumors: lack of correlation between margin status and local recurrence," *American journal of clinical pathology*, vol. 143, no. 5, pp. 645–651, (2015).

[119] Mohareri, O., Ischia, J., Black, P. C., Schneider, C., Lobo, J., Goldenberg, L., and Salcudean, S. E., "Intraoperative registered transrectal ultrasound guidance for robot-assisted laparoscopic radical prostatectomy," *The Journal of urology*, vol. 193, no. 1, pp. 302–312, (2015).

[120] Liu, X., Kang, S., Plishker, W., Zaki, G., Kane, T. D., and Shekhar, R., "Laparoscopic stereoscopic augmented reality: toward a clinically viable electromagnetic tracking solution," *Journal of Medical Imaging*, vol. 3, no. 4, pp. 045001–045001, (2016).

[121] Leven, J., Burschka, D., Kumar, R., Zhang, G., Blumenkranz, S., Dai, X. D., Awad, M., Hager, G. D., Marohn, M., Choti, M., and others„ "Davinci canvas: a telerobotic surgical system with integrated, robot-assisted, laparoscopic ultrasound capability," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 811–818, Springer, (2005).

[122] Wu, Y. and Hu, Z., "Pnp problem revisited," *Journal of Mathematical Imaging and Vision*, vol. 24, no. 1, pp. 131–141, (2006).

[123] Collins, T. and Bartoli, A., "Infinitesimal plane-based pose estimation," *International Journal of Computer Vision*, vol. 109, no. 3, pp. 252–286, (2014).

[124] Pratt, P., Di Marco, A., Payne, C., Darzi, A., and Yang, G.-Z., "Intraoperative ultrasound guidance for transanal endoscopic microsurgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 463–470, Springer, (2012).

[125] Edgcumbe, P., Nguan, C., and Rohling, R., "Calibration and stereo tracking of a laparoscopic ultrasound transducer for augmented reality in surgery," in *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions*, pp. 258–267, Springer, (2013).

[126] Suzuki, S. and Abe, K., "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32–46, (1985).

[127] Bradski, G. and Kaehler, A., *Learning OpenCV: Computer vision with the OpenCV library.* " O'Reilly Media, Inc.", (2008).

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

[129] Neubeck, A. and Van Gool, L., "Efficient non-maximum suppression," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3, pp. 850–855, IEEE, (2006).

[130] Bouguet, J.-Y., "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," *Intel Corporation*, vol. 5, pp. 1–10, (2001).

[131] Zhang, Z., "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, (2000).

[132] Kazanzidesf, P., Chen, Z., Deguet, A., Fischer, G. S., Taylor, R. H., and DiMaio, S. P., "An open-source research kit for the da vinci® surgical system," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6434–6439, IEEE, (2014).

[133] Shah, M., "Comparing two sets of corresponding six degree of freedom data," *Computer Vision and Image Understanding*, vol. 115, no. 10, pp. 1355–1362, (2011).

[134] Gras, G., Marcus, H. J., Payne, C. J., Pratt, P., and Yang, G.-Z., "Visual force feedback for hand-held microsurgical instruments," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 480–487, Springer, (2015).

[135] Sznitman, R., Ali, K., Richa, R., Taylor, R. H., Hager, G. D., and Fua, P., "Data-driven visual tracking in retinal microsurgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 568–575, Springer, (2012).

[136] Wolf, R., Duchateau, J., Cinquin, P., and Voros, S., "3d tracking of laparoscopic instruments using statistical and geometric modeling," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 203–210, Springer, (2011).

[137] Allan, M., Chang, P.-L., Ourselin, S., Hawkes, D. J., Sridhar, A., Kelly, J., and Stoyanov, D., "Image based surgical instrument pose estimation with multi-class labelling and optical flow," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 331–338, Springer, (2015).

[138] Pezzementi, Z., Voros, S., and Hager, G. D., "Articulated object tracking by rendering consistent appearance parts," in *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pp. 3940–3947, IEEE, (2009).

[139] Sznitman, R., Becker, C., and Fua, P., "Fast part-based classification for instrument detection in minimally invasive surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 692–699, Springer, (2014).

[140] Rieke, N., Tan, D. J., Alsheakhali, M., Tombari, F., di San Filippo, C. A., Belagiannis, V., Eslami, A., and Navab, N., "Surgical tool tracking and pose estimation in retinal microsurgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 266–273, Springer, (2015).

[141] Reiter, A., Allen, P. K., and Zhao, T., "Appearance learning for 3d tracking of robotic surgical tools," *The International Journal of Robotics Research*, vol. 33, no. 2, pp. 342–356, (2014).

[142] Tsai, R. Y. and Lenz, R. K., "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," *Robotics and Automation, IEEE Transactions on*, vol. 5, no. 3, pp. 345–358, (1989).

[143] Hinterstoisser, S., Cagniart, C., Ilic, S., Sturm, P., Navab, N., Fua, P., and Lepetit, V., "Gradient response maps for real-time detection of textureless objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 876–888, (2012).

[144] Chum, O. and Matas, J., "Matching with prosac-progressive sample consensus," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1, pp. 220–226, IEEE, (2005).

[145] Lepetit, V. and others„ "Epnp: An accurate o (n) solution to the pnp problem," *International journal of computer vision*, vol. 81, no. 2, pp. 155–166, (2009).

[146] Moustris, G. and others„ "Evolution of autonomous and semi-autonomous robotic surgical systems: a review of the literature," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 7, no. 4, pp. 375–392, (2011).

[147] Billings, S. and others„ "System for robot-assisted real-time laparoscopic ultrasound elastography," in *SPIE Medical Imaging*,

pp. 83161W–83161W, International Society for Optics and Photonics, (2012).

[148] Pratt, P. and others„ "Autonomous ultrasound-guided tissue dissection," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, pp. 249–257, Springer, (2015).

[149] Meining, A. and others„ "Direct visualization of indeterminate pancreaticobiliary strictures with probe-based confocal laser endomicroscopy: a multicenter experience," *Gastrointestinal endoscopy*, vol. 74, no. 5, pp. 961–968, (2011).

[150] Wallace, M. B. and others„ "Preliminary accuracy and interobserver agreement for the detection of intraepithelial neoplasia in barrett's esophagus with probe-based confocal laser endomicroscopy," *Gastrointestinal endoscopy*, vol. 72, no. 1, pp. 19–24, (2010).

[151] Ye, M., Giannarou, S., Patel, N., Teare, J., and Yang, G.-Z., "Pathological site retargeting under tissue deformation using geometrical association and tracking," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 67–74, Springer, (2013).

[152] Latt, W. T. and others„ "A hand-held instrument to maintain steady tissue contact during probe-based confocal laser endomicroscopy," *IEEE transactions on biomedical engineering*, vol. 58, no. 9, pp. 2694–2703, (2011).

[153] Rosa, B. and others„ "Building large mosaics of confocal edomicroscopic images using visual servoing," *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 4, pp. 1041–1049, (2013).

[154] Giataganas, P. and others„ "Force adaptive robotically assisted endomicroscopy for intraoperative tumour identification," *International journal of computer assisted radiology and surgery*, pp. 1–8, (2015).

[155] Hughes, M. and Yang, G.-Z., "Line-scanning fiber bundle endomicroscopy with a virtual detector slit," *Biomedical Optics Express*, vol. 7, no. 6, pp. 2257–2268, (2016).

[156] Geiger, A. and others„ "Efficient large-scale stereo matching," in *Computer Vision–ACCV 2010*, pp. 25–38, Springer, (2011).

[157] Rusu, R. B., *Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments*. PhD thesis, Technische Universität München, (2009).

[158] Erden, M. S. and others„ "Understanding soft-tissue behavior for application to microlaparoscopic surface scan," *IEEE transactions on biomedical engineering*, vol. 60, no. 4, pp. 1059–1068, (2013).

[159] Pratt, P. and others„ "Intraoperative ultrasound guidance for transanal endoscopic microsurgery," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 463–470, Springer, (2012).

[160] Hughes, M. and Yang, G.-Z., "High speed, line-scanning, fiber bundle fluorescence confocal endomicroscopy for improved mosaicking," *Biomedical optics express*, vol. 6, no. 4, pp. 1241–1252, (2015).

[161] Chang, P.-L. and others„ "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*, pp. 42–49, Springer, (2013).

[162] Varghese, R. and others„ "A framework for sensorless and autonomous probe-tissue contact management in robotic endomicroscopic scanning," in *Robotics and Automation (ICRA), 2017 IEEE International Conference on*, IEEE, (2017, In press).

[163] Shademan, A., Decker, R. S., Opfermann, J. D., Leonard, S., Krieger, A., and Kim, P. C., "Supervised autonomous robotic soft tissue surgery," *Science translational medicine*, vol. 8, no. 337, pp. 337ra64–337ra64, (2016).

[164] Vercauteren, T., Meining, A., Lacombe, F., and Perchant, A., "Real time autonomous video image registration for endomicroscopy: fighting the compromises," *Proc. SPIE*, vol. 6861, pp. 68610C–68610C–8, (2008).

[165] Vercauteren, T., Perchant, A., Malandain, G., Pennec, X., and Ayache, N., "Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy," *Medical Image Analysis*, vol. 10, no. 5, pp. 673–692, (2006).

[166] Zuo, S. and others„ "Towards intraoperative breast endomicroscopy with a novel surface scanning device," *Biomedical Engineering, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, (2015).

[167] Rollins, A. M., Ung-Arunyawee, R., Chak, A., Wong, R. C., Kobayashi, K., Sivak, M. V., and Izatt, J. A., "Real-time in vivo imaging of human gastrointestinal ultrastructure by use of endoscopic optical coherence tomography with a novel efficient interferometer design," *Optics letters*, vol. 24, no. 19, pp. 1358–1360, (1999).

[168] Lorenser, D., Quirk, B. C., Auger, M., Madore, W.-J., Kirk, R. W., Godbout, N., Sampson, D. D., Boudoux, C., and McLaughlin, R. A., "Dual-modality needle probe for combined fluorescence imaging and three-dimensional optical coherence tomography," *Optics Letters*, vol. 38, no. 3, pp. 266–268, (2013).

[169] Nadeau, C. and Krupa, A., "Intensity-based ultrasound visual servoing: Modeling and validation with 2-d and 3-d probes," *IEEE Trans. Robot.*, vol. 29, no. 4, pp. 1003–1015, (2013).

[170] Mountney, P. and Yang, G.-Z., "Motion compensated SLAM for image guided surgery," in *MICCAI*, pp. 496–504, Springer, (2010).

[171] Giannarou, S., Ye, M., Gras, G., Leibrandt, K., Marcus, H. J., and Yang, G.-Z., "Vision-based deformation recovery for intraoperative force estimation of tool–tissue interaction for neurosurgery," *Int. J. Comp. Assist. Radiol. Surg.*, vol. 11, no. 6, pp. 929–936, (2016).

[172] Yang, Q., Wang, L., and Ahuja, N., "A constant-space belief propagation algorithm for stereo matching," in *CVPR*, pp. 1458–1465, (2010).

[173] Farnebäck, G., "Two-frame motion estimation based on polynomial expansion," in *SCIA*, pp. 363–370, Springer, (2003).

[174] Lujan, A. E., Balter, J. M., and Ten Haken, R. K., "A method for incorporating organ motion due to breathing into 3d dose calculations in the liver: sensitivity to variations in motion," *Med. Phys.*, vol. 30, no. 10, pp. 2643–2649, (2003).

[175] Hughes-Hallett, A., Mayer, E. K., Marcus, H. J., Cundy, T. P., Pratt, P. J., Darzi, A. W., and Vale, J. A., "Augmented reality partial nephrectomy: examining the current status and future perspectives," *Urology*, vol. 83, no. 2, pp. 266–273, (2014).

[176] Ruszkowski, A. and others„ "On the feasibility of heart motion compensation on the da vinci® surgical robot for coronary artery bypass surgery: Implementation and user studies," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pp. 4432–4439, IEEE, (2015).

[177] Ruszkowski, A., Schneider, C., Mohareri, O., and Salcudean, S., "Bimanual teleoperation with heart motion compensation on the

da vinci® research kit: Implementation and preliminary experiments," in *Robotics and Automation (ICRA), 2016 IEEE International Conference on*, pp. 4101–4108, IEEE, (2016).

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

[179] Edgcumbe, P., Singla, R., Pratt, P., Schneider, C., Nguan, C., and Rohling, R., "Augmented reality imaging for robot-assisted partial nephrectomy surgery," in *International Conference on Medical Imaging and Virtual Reality*, pp. 139–150, Springer, (2016).

[180] Salcudean, S. E., Sahebjavaher, R. S., Goksel, O., Baghani, A., Mahdavi, S. S., Nir, G., Sinkus, R., and Moradi, M., "Biomechanical modeling of the prostate for procedure guidance and simulation," in *Soft tissue biomechanical modeling for computer assisted surgery*, pp. 169–198, Springer, (2012).

[181] Mohareri, O. and Salcudean, S., "da vinci® auxiliary arm as a robotic surgical assistant for semi-autonomous ultrasound guidance during robot-assisted laparoscopic surgery," in *Proceedings of the 7th Hamlyn Symposium on Medical Robotics*, pp. 45–46, (2014).

[182] Chan, T. F. and Vese, L. A., "Active contours without edges," *IEEE Transactions on image processing*, vol. 10, no. 2, pp. 266–277, (2001).

[183] Van Den Berg, J., Miller, S., Duckworth, D., Hu, H., Wan, A., Fu, X.-Y., Goldberg, K., and Abbeel, P., "Superhuman performance of surgical tasks by robots using iterative learning from human-guided demonstrations," in *Robotics and Automa-*

tion (ICRA), 2010 IEEE International Conference on, pp. 2074–2081, IEEE, (2010).

[184] Calinon, S., Pistillo, A., and Caldwell, D. G., "Encoding the time and space constraints of a task in explicit-duration hidden markov model," in *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pp. 3413–3418, IEEE, (2011).