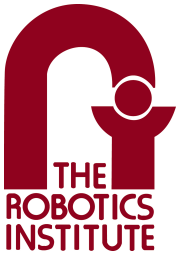
Towards Vision Guided Retinal Vein Cannulation with an Actively Stabilized Handheld Robot

Shohin Mukherjee

May 2017



The Robotics Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Cameron N. Riviere, Chair
John Galeotti
Arun Srivatsan

Submitted in partial fulfillment of the requirements for the degree of Master of Science.

CMU-RI-TR-17-20

Keywords: Surgical robotics, Computer Vision, Micron, Retinal Vessel Cannulation, Surface Reconstruction, EyeSLAM

For my parents, Dr. Sraban Mukherjee and Dr. Urmi Mukherjee, who have been my pillars of support. For my grandfather, Dr. Debabrata Banerjea, who has been my idol, and through his immense dedication and devotion, inspired me to pursue a career in science.

Abstract

In this thesis we describe work towards retinal vessel cannulation using an actively stabilized handheld robot, guided by monocular vision. Retinal vein cannulation is an incredibly delicate procedure because of the presence of physiological tremor in a surgeon's hand which is typically greater than the diameter of the retinal vessels. Because of this reason, cannulation is not an approved medical procedure, even though it has the potential to treat several retinal diseases.

Our first contribution is an improved version of EySLAM, an eyeball motion tracking algorithm, that delivers 30 Hz real-time simultaneous localization and mapping of the human retina and vasculature during intraocular surgery. We implement graph based SLAM using the incremental matrix re-ordering and sparse nonlinear incremental graph optimization algorithm: ISAM2. This work also handles loop closures and demonstrates increased robustness to quick shaky motions and drift due to uncertainities in the motion estimation. We show considerable improvement over the earlier version of the algorithm.

Our second contribution is a monocular camera based surface reconstruction technique using laser beam scanning over the retina. We then use the reconstructed surface to estimate a coordinate transform between the 2D image plane coordinates and the the global 3D tracking frame of Micron. We introduce a motion scaling framework around the transformed targets for greater precision during approach. Experiments are conducted in a wet eye phantom to show the higher accuracy of the surface reconstruction as compared to standard stereo reconstruction. Further, experiments to show the increased surgical accuracy due to motion scaling are also carried out.

Our third contribution is to develop force control during cannulation with a new tool that integrates a micro-needle, a 2D force sensor and a laser probe. We then propose a integrated framework to carry out vessel cannulation with Micron by splitting the control mode into two stages: a motion scaling mode during approach to the vessel, and a force control mode after contact.

Acknowledgments

This thesis is the result of the effort and guidance of multiple people, who invested their invaluable time, energy and support.

Research is always a collaborative effort of multiple individuals. First of all, I would like to thank my advisor Dr. Cameron Riviere for his mentorship and guidance. I am grateful to him for giving me the independence to build upon my own ideas, and also provide useful insights in my research. I would also like to thank John Galeotti for his guidance and for agreeing to serve on my committee. I am also grateful to Michael Kaess for providing his deep technical insight on the EyeSLAM work. A huge thank you to Dr. Joseph Martel for providing his surgical knowledge and letting me observing surgeries and getting surgical data. Without the technical assistance and guidance of Robert MacLachlan, research would have been extremely hard.

I would like to thank the previous and current members of the Surgical Mechatronics Lab, Sungwook Yang, Brian Becker and Shishir Gupta for helping me coming up to speed with the Micron project and keeping Micron running. A big cheer goes to my colleagues in the RI Masters program, Dhruv Saxena, Rosario Scalise, Sameer Bardapurkar, Shushman Choudhury, Xuning Yang, Merritt Jenkins, Puneet Puri, Anirudh Vemula, Vishal Dugar and many others, who helped me maintain my sanity.

Most importantly, I am extremely grateful to my parents for their unwavering support and love. Without their invaluable advice and guidance, this thesis would not have been possible.

Contents

List of Figures

List of Tables

Chapter 1

Background

Robot-assisted surgery is being increasingly adopted in a wide range of surgical procedures because it offers precise manipulation leading to repeatable and safer patient outcomes. Some of the fields where robots have been introduced are cardiovascular surgery [1–5], gastrointestinal surgery [6–8], urology [9, 10], gynecology [10–12], neurosurgery [13–15] and orthopedic surgery [16–18]. Robot-assist can be especially help in microsurgery, in which an intricate surgical procedure is performed with miniaturized instruments, typically under a microscope. These micro-metre scale procedures require very high levels of accuracy. One such class of surgeries is vitreo-retinal surgery. Vitreo-retinal surgery refers to several procedures to treat ophthalmologic diseases like retinal detachment, vitreos hemorrhage, macular pucker, diabetic retinopathy, retinal vascular occlusion (RVO), etc [1]. In this work, we propose robot-assisted vessel cannulation.

1.1 Clinical Significance

Retinal vein cannulation refers to the direct delivery of drugs into the veins of the retina. Cannulation has the potential to be useful in the treatment of diseases such as retinal vein occlusion (RVO), which occurs when a clot obstructs blood flow in a central (CRVO) or branch (BRVO) vein of the eye. RVO is the second most common retinal vascular disease after diabetic retinopathy and affects an estimated 16.4 million adults worldwide [19]. Though RVO has no proven effective treatment, a promising experimental procedure called retinal endovascular surgery (REVS) exists. REVS involves cannulating the vein and directly injecting clot-dissolving plasminogen activator (t-PA) [20, 21]. However the problem lies in the fact that the average luminal diameter of the large human retinal arteries is 120 μm and decreases to 8-15 μm in the periphery. It is extremely difficult to manipulate such fine vessels, given the presence of physiological tremor in the surgeon's hand and limited dexterity [22].

1.2 Robots in Retinal Surgery

Numerous groups have worked on different classes of robots for vitreo-retinal surgery. In [23], a telerobotic robot: Robot Assisted Micro Surgery (RAMS), for eye surgery was developed. It was

a cable-driven robot designed for high dexterity, incorporating a 6 degrees-of-freedom (DOF) master-slave system. It was limited by the complexitry of the software control and the lack of mechanical remote center of motion. In [24] a parallel manipulator capable of imposing a remote center of motion (RCM) was also developed for intravascular drug delivery and microvascular pressure measurement.

The most common master-slave robot is the da Vinci surgical system, which allows surgeons to remotely control a patient-side slave robot through manipulation of wristed instruments that are anthropomorphically-aligned to the surgical scene [25]. The system originally was designed for laproscopic surgery. In [26], a surgical micro-manipulator system that can be integrated with the da Vinci sytem to enable intraocular surgical procedures was developed. It was designed to address the shortcomings of the da Vinci system for intraocular procedures, namely the limited maneuverability and positioning of the instruments and the inadvertent tension-stresses applied to the eye during the procedure. In [27], the intraocular robotic interventional surgical system (IRISS), was introduced. It is a dedicated microsurgical platform and its feasibility in performing both anterior and posterior segment intraocular surgery was shown in porcine eyes. Nakano et al. developed a master-slave parallel robot to obtain high stability in intraocular surgery [28].

The Steady Hand Robot was developed at JHU [29]. The surgical tool was held simultaneously by the surgeon's hand and an actively controlled robot arm. The controller sensed the forces exerted by the operator on the tool and the tool on the environment and used it to provide smooth, tremor-free, precise positional control and force scaling. The design was improved in [30] to eliminate the limitations of the first prototype, in particular the parts of the mechanism nearest the patient being bulky and ergonomically inconvenient for the surgeon. The new robot: Eye Robot, had only 5 DOF: 3 translations and 2 rotations. By eliminating two local DOF (tool insertion and spin), a thin tool holder was made possible and the interaction between the robot and microscope work space was reduced. The robot was tested by performing cannulation on a biological model with the chorioallantoic membrane of a chicken embryo. In [31], the design was modified to develop the Eye Robot which incorporated both a significantly improved manipulator and an integrated micro-force sensing tool.

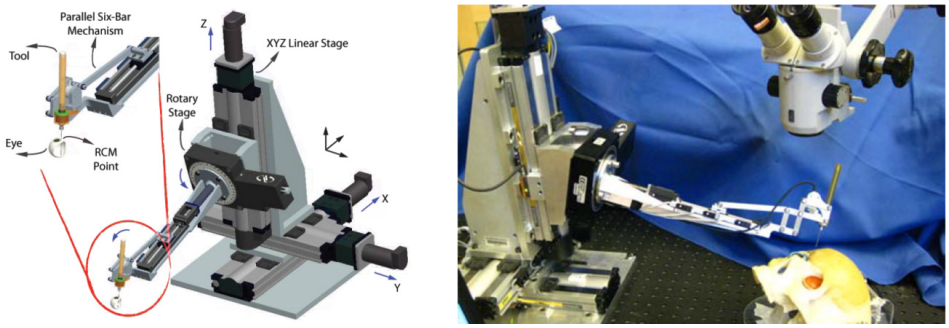


Figure 1.1: JHU Steady Hand Eye Robot [31].

1.3 Micron

Micron is an actively stabilized handheld surgical robot that compensates for the surgeon's physiological hand tremor. The first version of Micron was introduced in [32]. It had a 3-DOF manipulator based on piezoelectric stack actuators and a 6-DOF inertial sensing module. The tremor estimation and cancellation algorithm was presented in [33]. In order to reduce the weight of the handheld intrument due to the seven piezoelectirc stacks in series, a new prototype was introduced in [34], which adopted mechanical amplification. However the new design was mechanically complex and incorporated many joints and moving parts. This led to degradation in performance over time. In [35], a flexure-based parallel manipulator, actuated by three piezoelectric stacks was developed.

However it was observed that tremor frequencies way below 10 Hz need to be suppressed for use in microsurgery, which is lower than what could be accurately detected with inertial sensors. Therefore in [36] a new design of Micron was proposed. It had substantial changes. The largest change was replacing inertial sensing and open-loop control with DC-accurate optical sensing and closed-loop control. For that a custom-built optical tracking system (Apparatus to Sense Accuracy of Position ASAP) was developed, which could detect motion at frequencies lower than what could be detected by inertial sensors [37]. It also incorporated piezo-biomorph actuators on a large base to increase the range of motion.

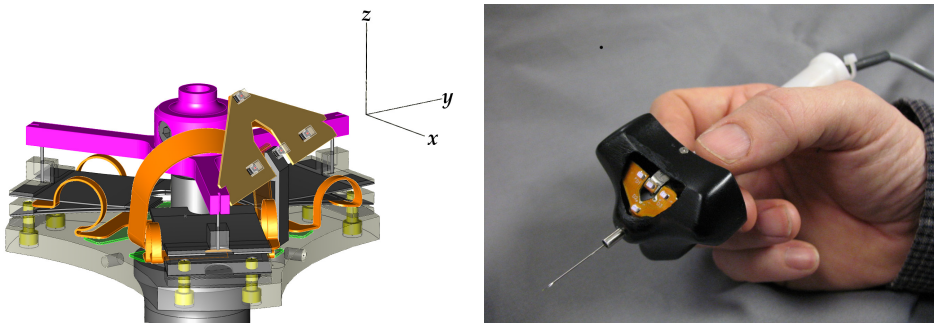


Figure 1.2: 3-D0F Micron [36].

The current version of Micron, as introduced in [38], is a 6 DOF system with a Gough-Stewart platform. The actuators are piezoelectric linear motors. The device has two sets of infrared LEDs, three of which are mounted on the moving platform and three are fixed to the handle. The LEDs are optically tracked by ASAP, which uses two position sensitive detectors to triangulate the position of each LED in space [37]. Figure 1.3 shows the exploded view of the latest version of Micron.

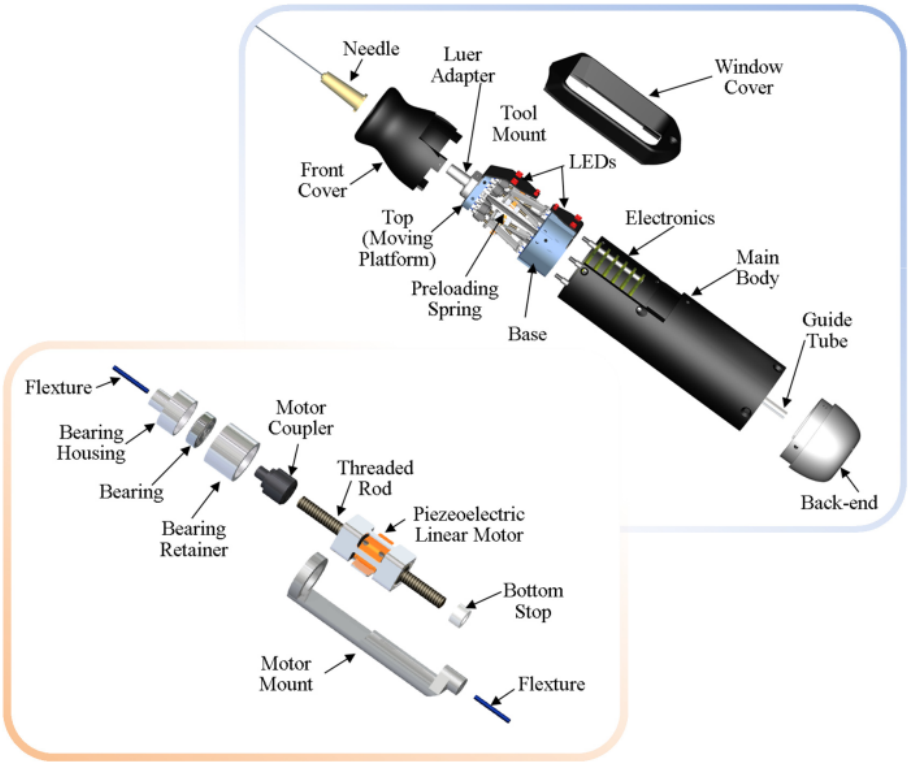


Figure 1.3: Exploded view of Micron [39].

1.4 Robot-Aided Retinal Vein Cannulation

In [40] vein cannulation using Micron was attempted. However the system suffered from several challenges due to which its application in an intact eye was not feasible. For example, the micropipette tip was painted for tracking which was not robust. However the primary shortcoming of the system was that it used stereo-vision to reconstruct the retinal surface and to convert the image space coordinates into the global ASAP coordinates. Stereo-reconstruction does not work in an intact eye during vitreo-retinal surgery because of the nonlinear optics of the eye. The optical path includes the cornea, the lens and saline solution introduced during vitrectomy. The eye is also covered with a contact lens to provide a wide-angle view during operation. All this leads to considerable optical distortion. The accuracy of stereo-reconstruction depends greatly on the accuracy of the camera calibration. Most camera calibration methods assume a classical perspective camera model in a single medium as air, an assumption which does not hold in an intact eye.

Other groups have also worked on this problem. In [41], a force-sensing micro-needle was integrated with Micron and the steady hand eye robot (SHER) in order to carry out retinal vein cannulation. The force sensor was used to detect puncture. The deviation from the puncture point was shown to be reduced by 65% in Micron and by 95% in SHER. In [42], vein cannulation in anesthetized pigs was performed using the Preceyes micromanipulator. The Preceyes micromanipulator is a teleoperated system consisting of a motion controller for hand motion in-

put by the surgeon, which is coupled with a table mounted instrument manipulator holding the surgical instrument with a clutch. In [43], a robot that can be controlled in a comanipulated or a telemanipulated manner was used to carry out retinal vein cannulation in porcine eyes. It also incorporated a puncture detection algorithm using FBG force sensors attached to the needle. In [44], a comanipulated robot was used to achieve retinal vein cannulation in an in vivo porcine retinal vein occlusion model.

1.5 Contribution Of The Thesis

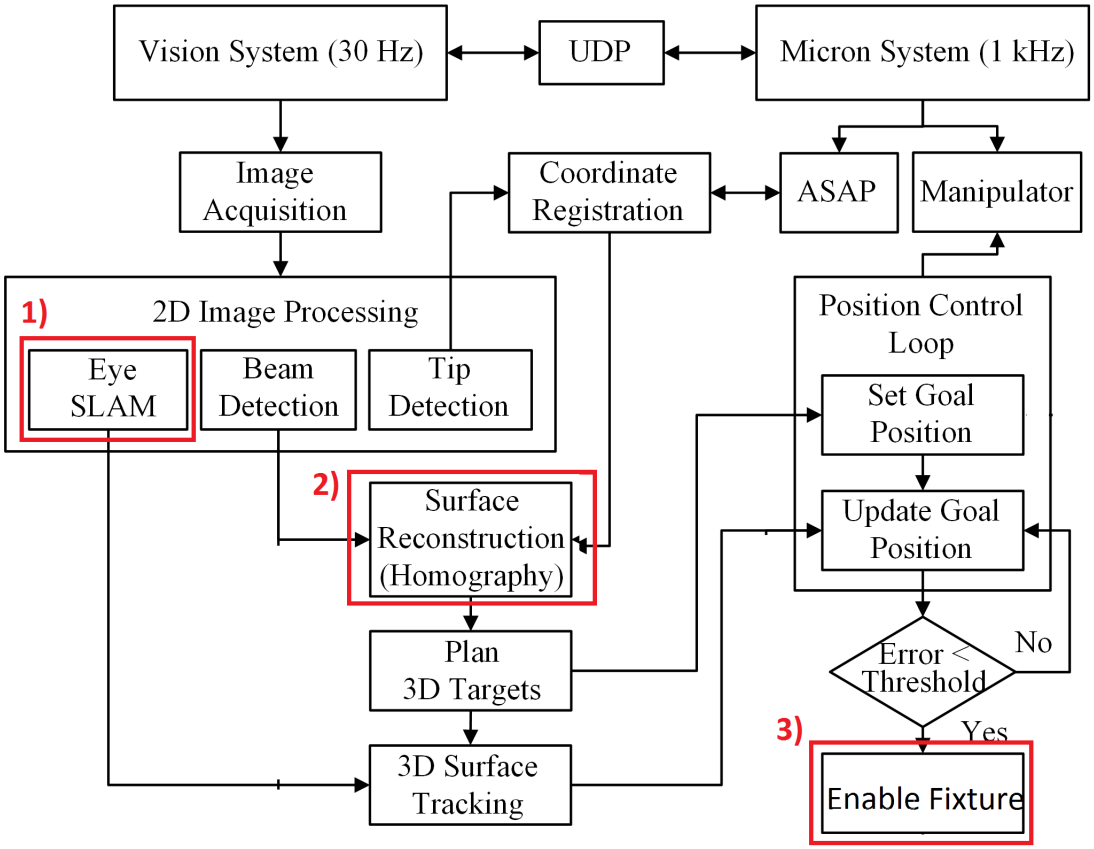


Figure 1.4: Block diagram showing the data flow of the system

This thesis contributes to the development of three sub-systems in the Micron system, which are highlighted in red in Figure 1.4.

The first chapter introduces pose-graphs based EyeSLAM, which delivers simultaneous localization and mapping (SLAM) of the human retina and vasculature during intraocular surgery. The work extends our previosu work by a graph-based SLAM formulation using the sparse nonlinear graph optimization algorithm iSAM2. The second chapter introduces homography matrix

estimation using monocular vision and application of the previously developed laser surface reconstruction to Micron guided vein cannulation. The final chapter describes our work towards achieving force control in cannulation.

Chapter 2

Pose-Graphs-Based EyeSLAM

Becker and Riviere [45] introduced EyeSLAM—a real-time vessel simultaneous localization and mapping (SLAM) algorithm, which employed the iterative closest point (ICP) algorithm for registration between a skeletonized version of the occupancy map and the current vessel observations. EyeSLAM was improved further in [46] by introducing the following two components:

- A fast correlative scan-matching method proposed by Olson et al. [47], in place of ICP.
- More robust vessel detection with better rejection of spurious detections.

However, the latest version of the algorithm employs scan matching for frame-by-frame registration, which corresponds to visual odometry in mobile robotics. Employing pure odometry for SLAM leads to significant drift since the uncertainty in the motion model builds up over time. Therefore there is a need to explore more advanced SLAM algorithms that handle loop closures to remove drift.

2.1 Problem Definition

The retinal localization and mapping problem can be defined as follows. Given a series of input video frames $I = [I_0, I_1, \cdots, I_T]$, over a discretized time period $t \in [0, 1, \cdots, T]$, we want to obtain a global map in the form of vasculature points and the corresponding camera viewpoint locations $L = [L_0, L_1, \cdots, L_T]$ of the input video frames in the map. The retina is modeled as a plane because typical retinal surgeries have high magnification with the field of view being only a few mm^2 of a $25\,\mathrm{mm}$ diameter eye. The motion model is a 3-DOF planar motion model parameterized as a 2D translation and a rotation: $[t_x, t_y, \theta]$.

2.2 Factor Graph

We represent the SLAM problem as a factor graph optimization. A factor graph is a bipartite graph $F = (\mathcal{U}, \mathcal{V}, \mathcal{E})$, where $f_i \in \mathcal{U}$ are factor nodes and $x_i \in \mathcal{V}$ are variable nodes. Edges $e_{ij} \in \mathcal{E}$ always exist between factors and variables [48]. A factor graph defines the factorization

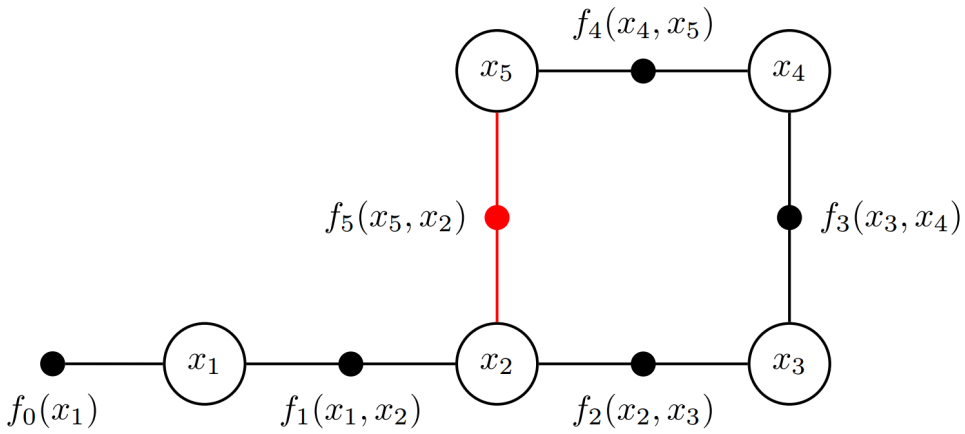


Figure 2.1: Factor graph representation [48] of EyeSLAM. The unfilled circles correspond to the variables over which we intend to infer. The incremental motion constraints are shown in black. The loop closure constraints are shown in red. $f_0(x_1)$ is a prior factor to fix a coordinate frame.

of a global function $f(X)$ as:

$$f(X) = \prod_i f_i(X_i) \tag{2.1}$$

Each factor $f_i(X_i)$ in the graph is proportional to a corresponding factor in the posterior probability function $p(X|Z)$. Each variable in the graph corresponds to the pose of the retina at a certain time instant. Solving for the poses involves performing maximum a posteriori (MAP) inference on the variables, given the information obtained from the uncertain measurements

$$X^{MAP} = \arg\max_X f(X) \tag{2.2}$$

$$= \arg\max_X \prod_i f_i(X_i) \tag{2.3}$$

In our factor graph formulation, the factors are of two types:a motion factor between the current and the previous frame, which are obtained as explained in 2.5; and loop closure factors, which are obtained as explained in section 2.6. Each of these factors is of the following form:

$$f_i(X_i) \propto \exp\left(-\frac{1}{2}||h_i(X_i) - z_i||^2_{\Sigma_i}\right) \tag{2.4}$$

This assumes measurements are corrupted by zero-mean, normally distributed Gaussian noise. Therefore, performing MAP inference involves solving the following non-linear optimization:

$$X^{MAP} = \arg\min_X \sum_i ||h_i(X_i) - z_i||^2_{\Sigma_i} \tag{2.5}$$

The nodes in our graph contain local maps of the vasculature tree that was observed in the pose that corresponds to that node. When a new image frame is obtained from the camera, first its vasculature tree is extracted and the incremental motion relative to the previous frame is estimated using scan-matching. A new node corresponding to the current frame is then added to

the factor graph with a motion factor, shown in black in Fig. 2.1. Then loop closure is checked for. If loop closure is detected, a loop closure factor is also added, which is shown in red. The graph is incrementally optimized using the iSAM2 algorithm [49].

2.3 Feature Extraction

In visual odometry there are two categories of methods to construct an optical flow field. In the first category the frames are directly matched in a dense or semi-dense fashion [50] [51]. This eliminates the need for feature extraction and exploits all the information present in the images. The second category of methods relies on feature detectors/descriptors such as SIFT [52], SURF [53], FAST [54], Harris [55] or other custom-designed detectors. However such detectors fail to find distinctive points on the textureless retina.

We use the network of blood vessels on the retina as features. They are extracted using the fast vessel detection algorithm proposed by Can et al. [56]. It is a highly efficient algorithm, suitable for real-time high definition videos. Its efficiency comes from direct processing on gray-level data without any preprocessing, and from processing only a minimally necessary fraction of pixels in an exploratory manner, avoiding low-level image-wide operations such as thresholding, edge detection, and morphological processing. To remove spurious detections of vessel-like structures such as the instrument tip of the light-pipe that is used for illumination, each potential vessel point undergoes a color test and a bloom proximity test [46]. The color test rejects pixels that are too dark or insufficiently red, while the bloom proximity test rejects vessel points that are too close to large white specular blooms in the image. As in [46], EyeSLAM is equipped with a filter that applies a mask on the dark area circling the visible retina area to mask out fringing effects caused by the microscope.

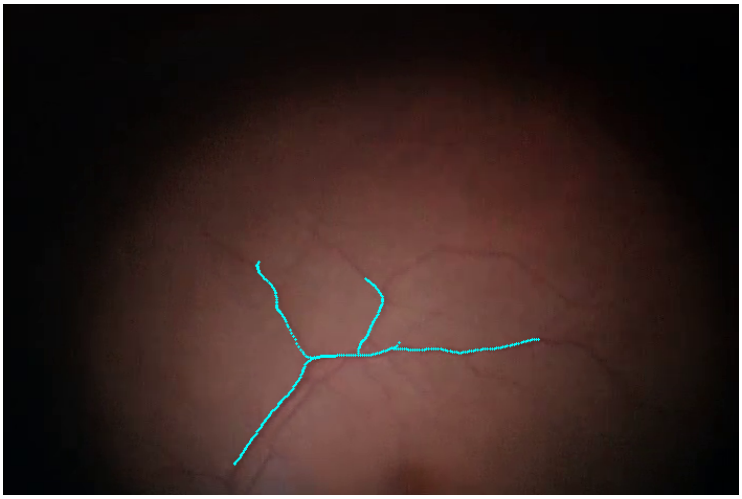


Figure 2.2: Vessel extraction in a real eye using the algorithm proposed in [56]. The algorithm detects the major vessels. The thinner vessels are harder to detect, but once they are detected in subsequent frames, they will be added to the global map.

2.4 Mapping via Occupancy Grids

Similar to work in [46], a global occupancy map holds the current best estimate of all the observed vasculature. Each pixel in the map stores a score that is proportional to the probability of it being a vessel pixel. At each time instant, the current observations Z_t are transformed to the map with the localization estimate L_t^{-1}. For each vein point Z_t^i, a fixed value is added to the map at the corresponding cell location, thereby increasing the probability of a vessel existing at that location. The occupancy map has a maximum allowed value in order to keep the scores bounded. In [46], a decay function was used to the decrease the probability of all grid cells, allowing vessels which have not been detected for a while to vanish. What this essentially means is that the part of the built map that is no longer consistent with the current observations, is discarded. Though this approach generates a smoother map by artificially removing the drift, it does not solve the underlying cause of the drift, i.e., the accumulated gross error in the localization caused by the inaccurate motion estimate. Hence even though the map appears smooth, the localization is still under gross error. The utility of employing EyeSLAM is to accurately track the eyeball motion so that it can be compensated for during control. A smooth map is preferable but not a necessity.

2.5 Motion Estimation via Scan Matching

To estimate the eyeball motion, a 3-DOF planar motion model parameterized by a 2D translation and a rotation is used. In the first formulation of the algorithm, the iterative closest point (ICP) algorithm was used for registration between a skeletonized version of the occupancy map and the current vessel observations [45]. Any rapid motion caused it to fail.

In the second formulation [46], ICP was replaced with the fast correlative scan-matching approach. The scan-matching was performed incrementally between the features observed in the current frame and a global map of features accumulated in the form of an occupancy map. It used a hierarchial approach in which a first scan on a low resolution version of the map quickly finds an approximate solution and avoids local minima. This approximate solution is then used to initialize a search on the high resolution map. To ensure speed of execution, at most 500 random vessel points are used for scan matching. A constant-velocity Kalman filter was used to smooth the localization estimation.

In this work we use a graph-based formulation to allow correction of the map upon loop closure, requiring a different strategy for scan matching. Matching against the global map results in corruption of the map caused by drift: When revisiting a previously observed part of the map, drift in the state estimate results in a duplication of structure in the map. In a graph-based solution, instead of matching against the previous map, one matches to the previous frame only, generating pair-wise constraints between pose estimates. Upon revisiting previously observed parts of the scene, one additionally matches against older, nearby frames to generate again pair-wise constraints. But this time they serve as loop closures between the current frame and an older part of the trajectory. The graph optimization then corrects the trajectory, and one can re-render a corrected global map—though that is only needed for visualization, not for the SLAM algorithm itself.

However in practice, this approach needs to be refined because noise in the vessel detection causes failure in scan matching. The features that were observed in one frame may not be observed in the immediately following frame. Therefore we propose a hybrid approach which estimates motion factors by performing registration between the current frame and a locally accumulated map; without significantly invalidating the independence assumption that a factor between two nodes should express the relationship between just the two nodes. We do this by using a decay function that reduces the probabilities that were added to the occupancy map every N^{th} frame by a scale γ, where $\gamma \in (0, 1)$. A value of $\gamma = 0$ is the same as matching consecutive frames. A value of $\gamma = 1$ is the same as matching the current frame with a globally accumulated map as was done in [46]. The lower the value of γ and the higher the value of N, the more accurate is the model, but less robust is the motion estimation. This is because the probabilities that were added to the occupancy map due to previously observed frames decay by a larger fraction and more frequently. In our implementation γ was chosen to be 0.9 and N was chosen as 10, based on preliminary experimentation.

2.6 Loop Closure Detection

If the pose estimate from incremental scan-matching is in gross error, while the camera is observing an already mapped region of the retina, the likelihood of the measurements being explained by the pose and map estimate is vanishingly small. This leads to previously visited areas getting re-mapped in the wrong global location and the error accumulates without bound [57]. Therefore loop closure factors are added to the graph when the camera re-observes a previously observed scene. In order to estimate loop closure, the local map of the current node needs to be compared to the local maps of every node that has been previously added to the group. Therefore the number of times we need to run the expensive scan-matching increases linearly with the graph size. This is too costly given the real time requirements of our application. In order to save computation time, a better approach is to compare the current node with the nodes that are inside its uncertainty ellipse. However this may still be very computationally expensive, especially if all frames are being processed, in which case there will be multiple nodes within the uncertainty ellipse. Therefore we always match with a fixed number of previously visited nodes, depending upon how many threads can run in parallel in the given machine. In order to ensure that the nodes that were recently visited are not matched with the current node, and loop constraints are added to the graph, we enforce a minimum time difference between the current frame and the frame that is matched with it for a loop constraint.

2.7 Evaluation

2.7.1 Qualitative Evaluation

We evaluated the new version of EyeSLAM on several videos of real human eye surgeries. The maps developed on 14 of these videos are shown in Fig. 2.6. These videos are of different surgical procedures including retinal membrane peeling and laser photocoagulation. They have been shot under different surgical conditions such as lighting, resolution, and tool occlusion. There is no ground truth available for real surgery videos, hence a qualitative evaluation was made. In these experiments we made sure that we process all frames offline so that we can compare the performance with the previous formulation of EyeSLAM. Hence the code was not optimized to run in real-time. Real-time performance can easily be achieved by running the algorithm on a faster processor with more cores.

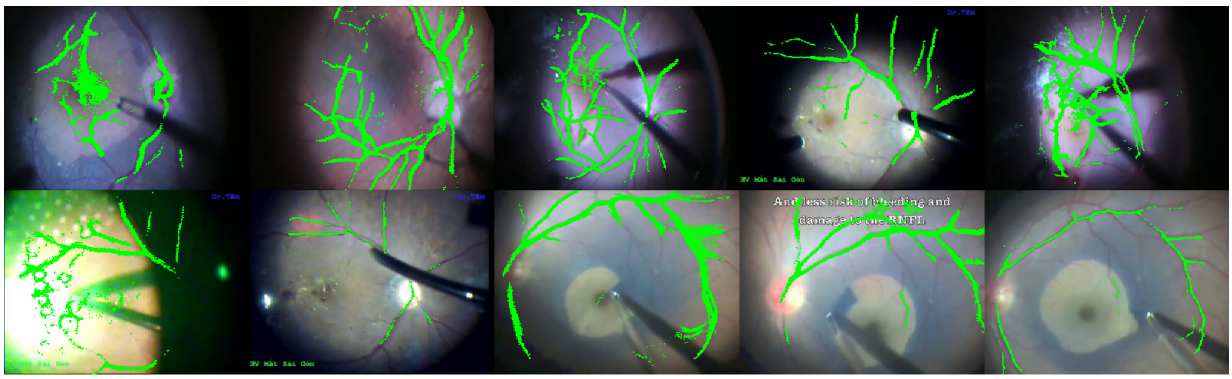
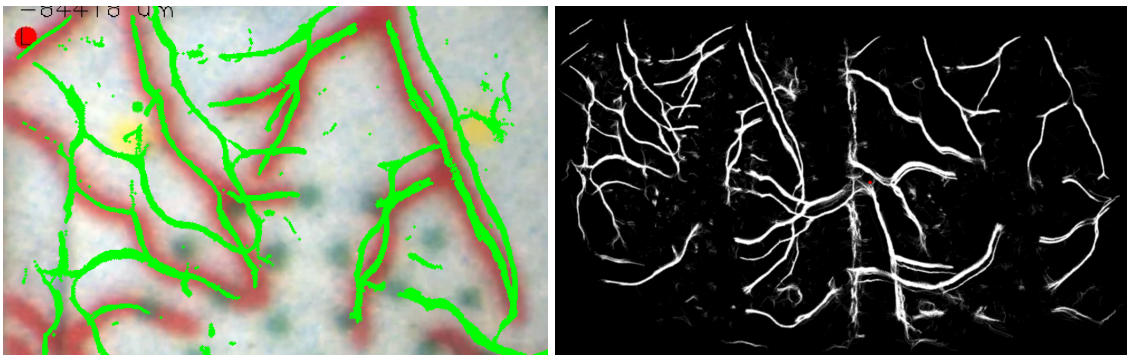
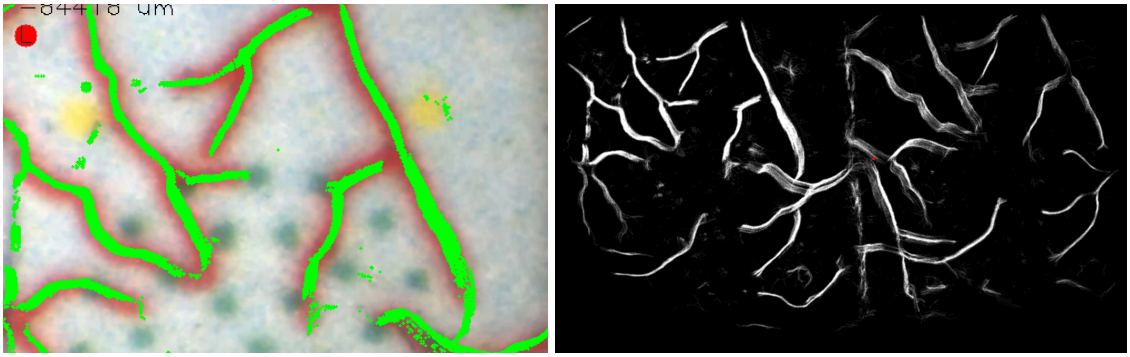


Figure 2.6: EyeSLAM was run on several video sequences of eye surgery. The videos correspond to different surgical procedures, under different lighting conditions. It can be seen that EyeSLAM manages to map most of the vasculature. It also handles tool occlusions very well.

Fig. 2.3(a) and 2.3(b) show the older EyeSLAM and the current version working on video sequence A. The colored image on the left is the map superimposed on the image frame, the image on the right is the global occupancy map. The brighter the pixel, the more likely it is for a vessel to exist at that point. The drift due to the inaccurate motion estimation is clearly visible in Fig. 2.3(a). The current EyeSLAM takes into account loop closure and therefore eliminates the drift, as expected, and as can be seen in Fig. 2.3(b). Fig. 2.4 shows the trajectory of the center of the camera's frame with respect to the global map for video sequence A. The camera executes a loop and returns close to the starting position. The loop is detected and the drift in the motion is corrected which can be seen in Fig. 2.3(b). Fig. 2.5(a) and 2.5(b) compare the older EyeSLAM and the current version working on a real surgery video. The map in fig. 2.5(b) is sharper because of less drift as compared to the map in fig. 2.5(a).



(a) Old EyeSLAM



(b) New EyeSLAM

Figure 2.3: Comparison of SLAM maps generated from old and new EyeSLAM on video sequence A. The images on the right show the global map. The images on the left show the map superimposed on the current frame. The drift is apparent in old EyeSLAM which leads to duplicated structures, which has been corrected in the new EyeSLAM.

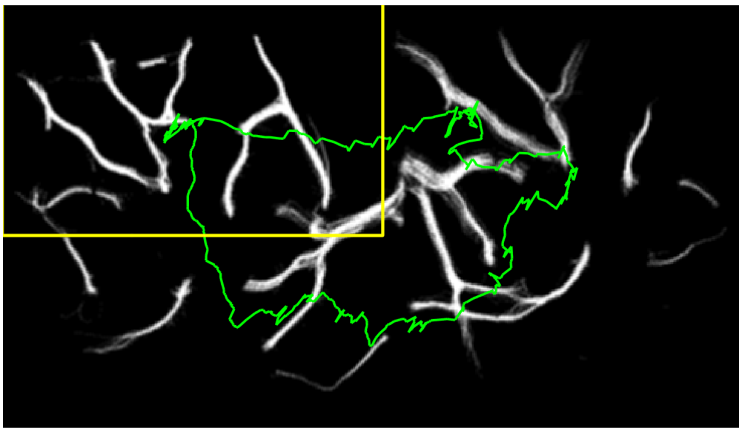
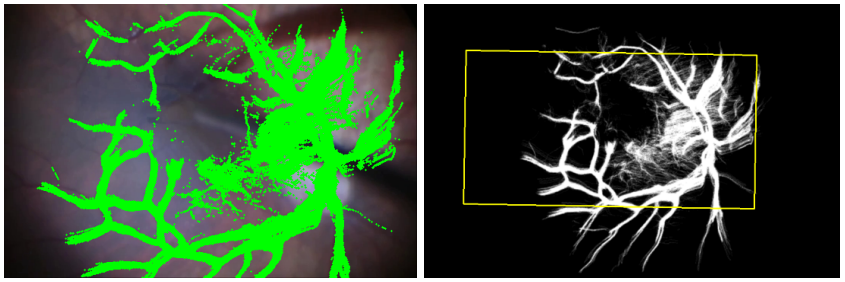
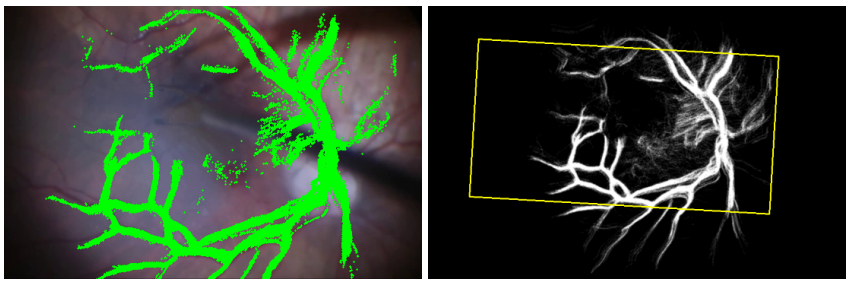


Figure 2.4: Camera trajectory (green) superimposed on the globally corrected map for video sequence A. The yellow frame shows the footprint of the last image, which completed a loop in the camera trajectory.



(a) Old EyeSLAM



(b) New EyeSLAM

Figure 2.5: Comparison of SLAM maps generated from old and new EyeSLAM on a real eye surgery.

2.7.2 Accuracy on Phantom

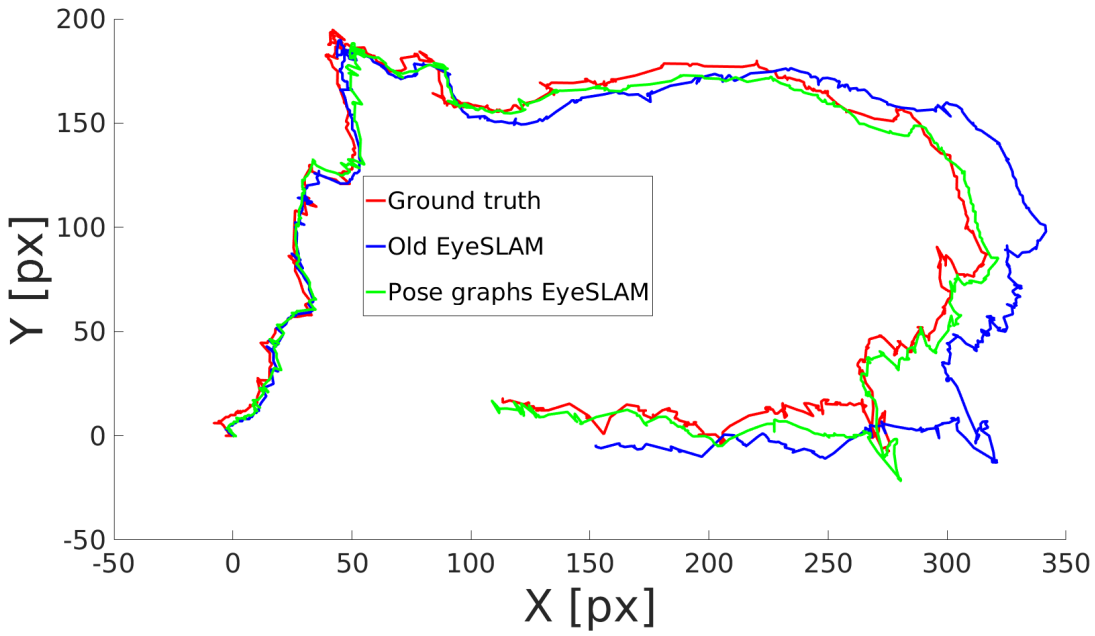
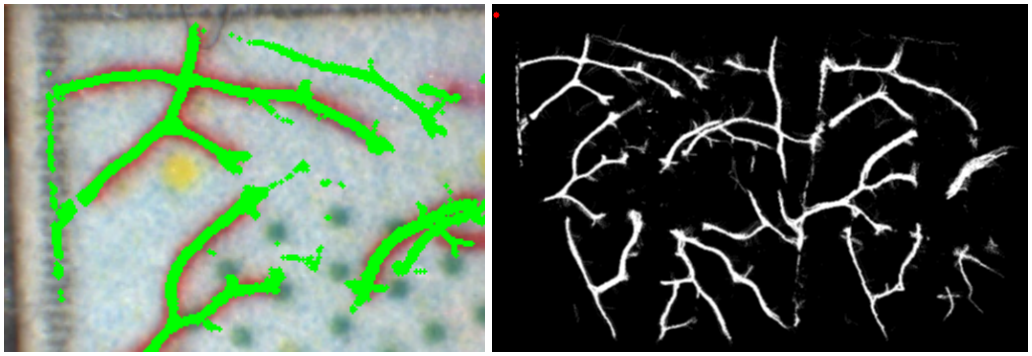
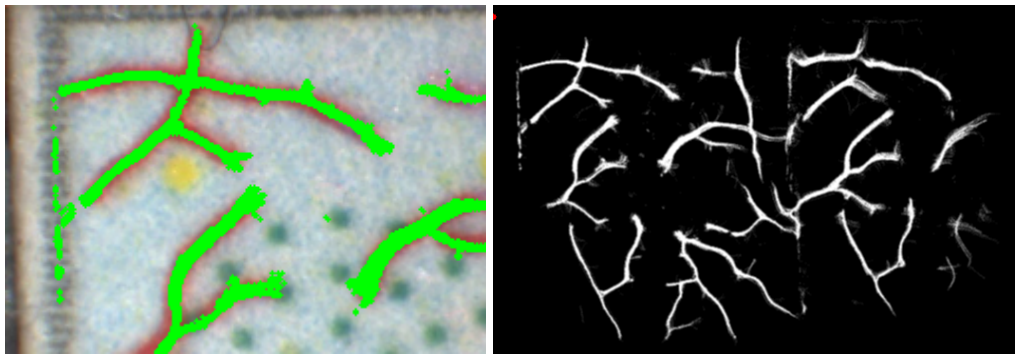


Figure 2.7: Camera trajectory for ground truth, old EyeSLAM and factor graphs based EyeSLAM.

In order to better test our hypothesis that the pose-graphs based EyeSLAM performs better, we carried out quantitative evaluation on an eye phantom. This was not possible to do on real surgery videos because no ground truth is available or can be generated for them. We ran the evaluation on phantom video sequence B. The eye phantom consisted of a printed retina, similar to that used in [46]. Ground truth was generated by tracking colored fiducials printed on the retina phantom. Fig. 2.7 compares the camera trajectory for the old EyeSLAM and the factor graphs based EyeSLAM with the ground truth trajectory.



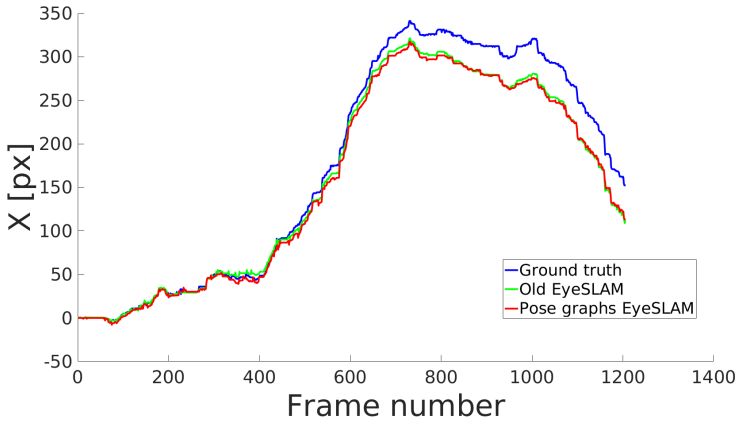
(a) Old EyeSLAM



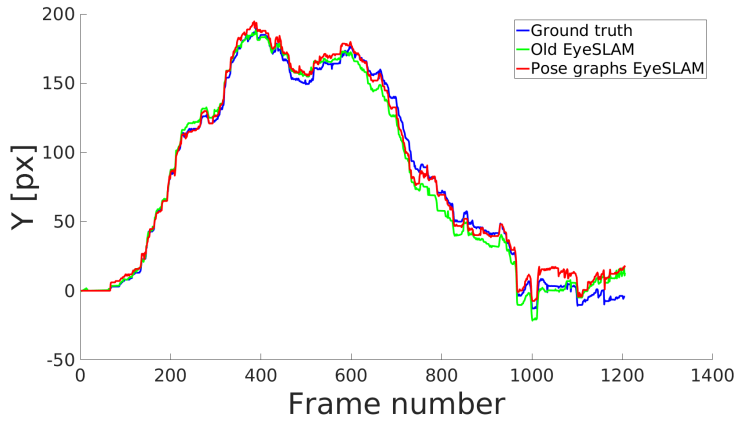
(b) New EyeSLAM

Figure 2.8: Comparison of SLAM maps generated from old and new EyeSLAM on video sequence B.

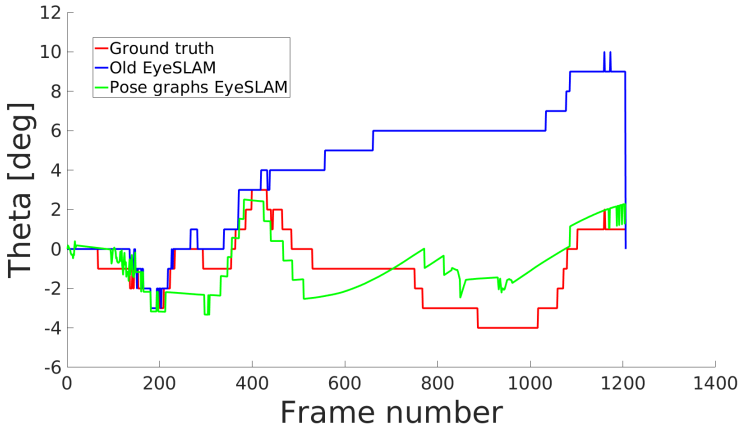
Fig. 2.8 compares the SLAM maps, where drift is apparent for the old EyeSLAM, while our implementation of factor graphs based EyeSLAM estimates camera motion much more accurately. The average error in translation was 19.6 pixels for the old EyeSLAM and 6.5 pixels for the factor graphs based EyeSLAM. From Fig. 2.7 it can be seen that the drift is apparent for the old EyeSLAM, while our implementation of pose graphs based EyeSLAM estimates camera motion much more accurately. Fig. 2.9 compares the translation in X and Y, and rotation for the two cases with the ground truth.



(a)



(b)



(c)

Figure 2.9: Comparing translation and orientation for ground truth, old EyeSLAM and factor graphs based EyeSLAM.

2.8 Discussion

In this work we introduced a factor graph-based formulation of retinal SLAM. The graph-based formulation can incorporate loop closures and therefore eliminate the drift that is introduced by noise in the frame-by-frame scan matching. We evaluated our new algorithm on several real eye surgery videos under different lighting conditions and in the presence of tool occlusion. We also evaluated our formulation of EyeSLAM on eye-phantom videos and compared the generated map with that generated from the previous formulation of EyeSLAM. The factor-graph-based EyeSLAM generated a more consistent map. We also showed reduction in average pixel error in the 2D camera motion estimation.

Future improvements should include optimization to allow the algorithm to run in real-time on high-definition videos. Furthermore, implementing 3D SLAM using stereo-vision while handling distortion could be beneficial.

Chapter 3

Retinal Surface Reconstruction

In [39], a new method for retinal surface reconstruction using monocular vision was introduced. It relied on circular scanning of the retinal surface using a surgical laser attached to Micron and geometric analysis of the projected elliptical beam pattern. The method was tested in various conditions and its feasibility was established in a realistic eye model which incorporated optical distortion by lenses. The reconstruction was then used to carry out laser photocoagulation with Micron. The 3-DOF motion of the tool tip was decoupled into the 2-DOF planar motion parallel to the retinal surface and the 1-DOF motion along the axis of the tool. The decoupled 2-DOF motion was then controlled via image-based servoing, to place the laser beam onto a target position using a monocular camera. The 1-DOF axial motion was controlled to maintain a constant standoff distance from the estimated surface. However, visual servoing is challenging in cannulation. Instead of tracking a laser spot, a needle tip needs to be tracked, which is unreliable in a real eye. There have been some tool-tracking algorithms developed [58–60]. However, the cannulation needle is very thin and accurate detection of the tool tip is hard. In this paper, we apply the new surface reconstruction technique to develop a vessel cannulation technique that can be used in an intact eye.

3.1 Tool and Eye Phantom

In order to use the surface reconstruction using laser as described in 3.3, it was essential to have a tool that has both the laser probe and the cannulation needle. The eye phantom consists of a network of blood vessels printed on a paper with water on top, to serve as saline that replaces the vitreous humor during vitrectomy. The tool and the phantom are shown in Figure 3.1.

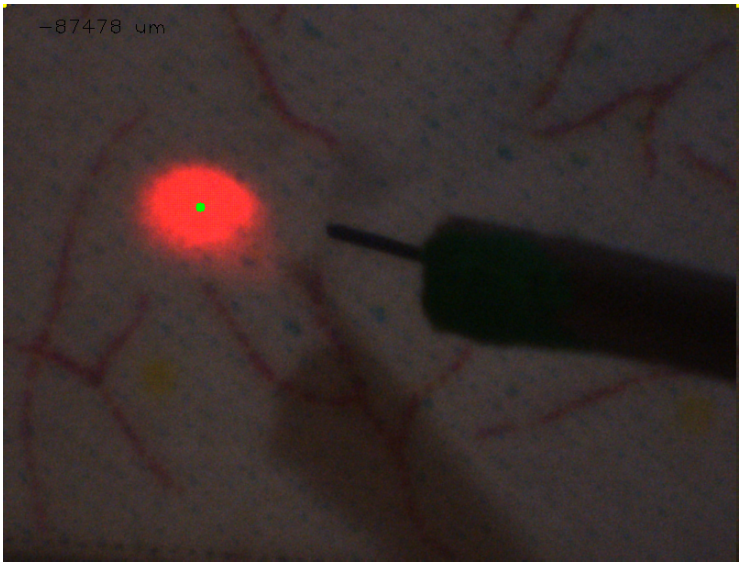


Figure 3.1: Tool and eye phantom used.

3.2 Tool Tip Calibration

The surface reconstruction described in [39] is very sensitive to the tool tip offset, which is the offset of the tool tip from the origin of Micron's coordinate system. In order to calibrate the tool tip offset, we use an approach developed in [61]. An optimizer is used to minimize the error between the projected and detected tip positions in both the cameras, while scanning the tool over the surface. The projected tip positions are obtained by multiplying the camera projection matrices with the 3D ASAP coordinates of the tool tip. The camera projection matrices are simultaneously calibrated during the optimization. The tip offset is modeled in terms of angular variations θ_x and θ_y, and the length of the tool l_0, from the origin of Micron's coordinate system. The optimization minimized the following objective function:

$$\min \sum_{i=1}^{n} [(p^i_{ProjL} - p^i_{DetL})^2 + (p^i_{ProjR} - p^i_{DetR})^2] \tag{3.1}$$

where p^i_{ProjL} and p^i_{ProjR} are the projected tip positions for the left and right cameras and p^i_{DetL} and p^i_{DetR} are the detected tip positions in the left and right cameras.

3.3 Surface Reconstruction

Since the control of Micron happens in the global coordinates of ASAP, there is a need to model the retina in the ASAP coordinates. Due to the unreliability of stereo reconstruction in an intact eye [62], Yang et al. [39] developed a monocular surface reconstruction method, which uses a surgical laser scanning its aiming beam over the retinal surface. The region of interest of the retina is approximated as a plane. A circular scan is carried out by Micron over the retina, around a pivot point. The pivot point is the RCM (remote center of motion). The RCM in vitreoretinal

surgery is the point of insertion of the tool into the eyeball. The circular pattern projects an ellipse on the plane to be estimated. The beam locations on the image plane are detected using a beam detection algorithm, which uses a sequence of image processing techniques like thresholding, morphological operations and contour detection. An ellipse is fitted through these beam points after outliers have been removed using RANSAC. The plane to be estimated can be described by tilt of the plane initially perpendicular to the axis of the cone. From Figure 3.2, θ_{plane} and d_{plane} are estimated by (3.2) and (3.5). In the following equations, m_a and m_b are the major and minor axis lengths respectively of the fitted ellipse and $\boldsymbol{v_{cone}}$ is the vector along the axis of the cone. s_{cam} is the image scale in µm/pixel.

$$\theta_{plane} = \pm \arcsin \sqrt{\frac{1 - \frac{1}{\gamma^2}}{1 + \frac{1}{\tan^2 \theta_{cone}}}} \tag{3.2}$$

$$\gamma = \frac{m_b}{m_a} \tag{3.3}$$

$$\tan \theta_{cone} = \frac{h_{RCM}}{r_{cone}} \tag{3.4}$$

$$d_{plane} = \frac{\cos \theta_{plane} (\tan^2 \theta_{cone} - \tan^2 \theta_{plane})}{\tan \theta_{cone}} m_a s_{cam} \tag{3.5}$$

$$\boldsymbol{P_{plane}} = \boldsymbol{P_{RCM}} - d_{plane} \boldsymbol{v_{cone}} \tag{3.6}$$

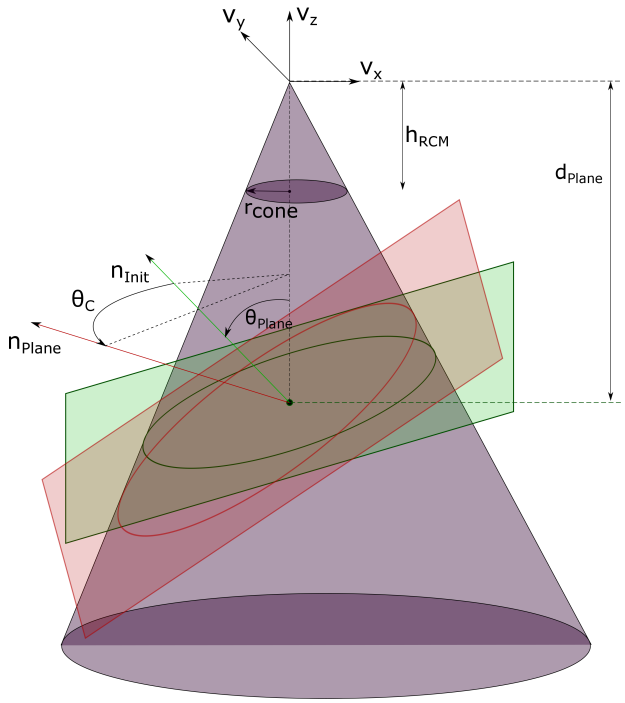


Figure 3.2: Geometric analysis and parameters of ellipse formed on the plane from the circular laser scan [39].

The angle θ_{plane} does not uniquely describe the tilted plane, since any arbitrary vector orthogonal to the axis of the cone can be taken as the axis of rotation. Thus, rotating the true normal about the axis of the cone by any arbitrary angle leads to an ellipse of the same dimensions. Here, by true normal we mean the normal of the surface we are trying to estimate.

An initial normal vector, as explained in [39], can be set by rotating the axis of the cone v_z about the y-axis of the cone coordinates by θ_{plane}.

$$\boldsymbol{n_{init}} = \boldsymbol{R}(\boldsymbol{v_y}, \theta_{plane})\boldsymbol{v_z} \tag{3.7}$$

The true surface normal can be considered as a rotation of an initial normal vector n_{init} about the axis of the cone v_z by an angle θ_c. In order to determine θ_c, an optimizer is used to match the relative orientation of the 2D beam positions in the image with the 3D beam positions obtained by finding the intersection of the ray from the RCM to tip position of Micron with the estimated plane.

Given k points on the trajectory, the objective function in (3.8) was defined as in [39].

$$F = \min_{\theta_c \in [-\pi, +\pi]} \sum_{i=1}^{k} (\theta^i_{beam} - \theta^i_{proj})^2. \tag{3.8}$$

where the phase angles of the ith data point with respect to the major axis of the ellipse obtained from the beam trajectory on the image and the one obtained from the projected points on the estimated plane are θ^i_{beam} and θ^i_{proj} respectively. Once we have the θ_c that minimizes the given

objective function, we can find the true normal by (3.9).

$$\boldsymbol{n_{plane}} = \boldsymbol{R}(\boldsymbol{v_z}, \theta_c)\boldsymbol{n_{init}} \tag{3.9}$$

The plane point P_{plane} given by (3.6) and the plane normal given by (3.9) completely define the reconstructed retinal plane.

3.4 Homography Matrix Estimation

In order to convert a cannulation target in the image space to the ASAP coordinate plane, we utilize the planar assumption of the region of interest. We then find a homography matrix between the retina and the image plane. This is done by obtaining a Direct Linear Transform (DLT) [63] between the detected 2D beam positions in the image and the corresponding expected 3D beam positions in the ASAP frame, obtained by finding the intersection of the ray from the RCM to the tip with the estimated plane from the cone beam reconstruction. These are the same correspondences that were used in the surface reconstruction. Since the distance between the microscope and the retina is large compared to the dimensions of the region of interest, the angle of incidence for all points on the region of interest is approximately the same. Therefore, a linear relationship exists between the two coordinate systems as long as the optics of the medium remain constant. If the surface is reconstructed just before the cannulation operation, this approximation can be made.

Therefore,

$$\begin{bmatrix} X_{ASAP} \\ Y_{ASAP} \end{bmatrix} = \boldsymbol{H} \begin{bmatrix} x_{camera} \\ y_{camera} \end{bmatrix} \tag{3.10}$$

$$Z_{ASAP} = (-aX_{ASAP} - bY_{ASAP} - 1)/c \tag{3.11}$$

where $[X_{ASAP}, Y_{ASAP}, Z_{ASAP}]$ is the 3D world coordinate, $[x_{camera}, y_{camera}]$ is the 2D image space pixel coordinate, $[a, b, c]$ are the geometric plane parameters and H is the homography matrix.

3.5 Surface Tracking with EyeSLAM

During vitreoretinal surgery, the surgeon manipulates the eye with a surgical tool to explore the region of interest. The patient may also move the eye. Because of this voluntary and involuntary motion of the eyeball, there is movement in the microscopic view of the retina. Therefore, any goal position defined in the image coordinates needs to be updated to account for this movement. For this purpose, Becker et. al developed EyeSLAM - a real time vessel mapping and localization algorithm [45] [46]. We use EyeSLAM to track the 2D motion described by 3 degrees of freedom ($[t_x, t_y, \theta]$) of the retinal plane and update the image-space goal position with a rigid transformation. This updated goal position is then mapped to ASAP coordinates using the homography matrix. Figure 3.3 shows EyeSLAM working in a wet eye phantom. EyeSLAM always runs in the background to track eyeball motion. The cone beam surface reconstruction and the homography matrix calculation is done once, after which motion scaling is engaged around

the target point in the ASAP coordinates obtained by converting the image space target with the computed homography matrix.

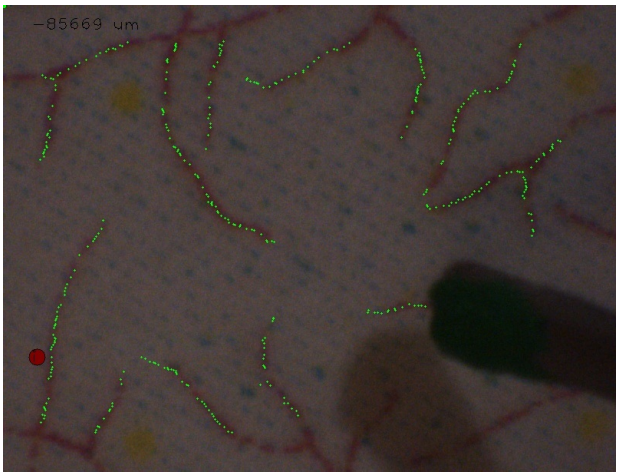


Figure 3.3: Vessel tracking with EyeSLAM in a wet eye phantom. The green dots correspond to the current location of vessels.

3.6 Motion Scaling

Once the goal position for cannulation has been established in the ASAP coordinates, we define a hemispherical region around it. When the tool tip is outside this region, Micron operates in the normal tremor-cancellation mode which employs a low-pass filter that serves to attenuate high-frequency motions due to hand tremor [64]. Inside this region, motion scaling is engaged.

$$\boldsymbol{P_T} = \boldsymbol{P_R} + \lambda(\boldsymbol{F_L}(\boldsymbol{P_N}) - \boldsymbol{P_R}) \tag{3.12}$$

In (3.12), $\boldsymbol{P_T}$ is the scaled goal position being sent to Micron's position controller, $\boldsymbol{F_L}(\boldsymbol{P_N})$ is the tremor-compensated null position, $\boldsymbol{P_R}$ is a reference point and λ is the scale factor. Two choices for the reference point are logical. One possible choice is the point of entry of the tool tip into the region. In this case, there is a smooth transition from tremor cancellation mode to motion scaling mode. However, while exiting the region, if the point of exit is not the same as the point of entry, there is a jerk during the transition. A more obvious choice is to choose the intended cannulation goal position as P_R. However, in case of a constant scaling factor, there is a jerk during transition while both entering and exiting the motion scaling region. This is because there is a sudden transition from the scaled goal position to the null position of the manipulator, which is the default goal position in the absence of motion scaling.

In [65], in order to carry out a graceful transition between motion scaling and tremor compensation, the scale was decreased exponentially over time, which also provided robustness to intermittently noisy distance measurements due to unreliable tip tracking. However, since all our control happens in the much more reliable ASAP frame, and distances are measured in the global

ASAP coordinates instead of the image coordinates, we simply change the scale factor linearly as a function of distance from the target, as given by (3.13).

$$\lambda(r) = \frac{1 - \lambda_0}{R} r + \lambda_0 \tag{3.13}$$

R is the distance at which motion scaling is engaged and r is the current distance from the target. A scale factor of 0 corresponds to a hard virtual fixture and a scale factor of 1 corresponds to normal tremor compensation and no motion scaling. λ_0 corresponds to the maximum motion scaling at $r = 0$.

3.7 Results

3.7.1 Accuracy of Coordinate Transform

In order to measure the accuracy of the surface reconstruction and the image space to ASAP coordinate transform with the homography matrix, a grid of targets was defined on the retinal plane in the image space. The Micron tip was then commanded to servo to each of these targets, while maintaining a clearance of 500 μm in the cone beam case and 1500 μm in the stereo case. It was observed that without a large enough clearance, the needle tip would hit the surface of the retina phantom, due to the extremely inaccurate stereo reconstruction. The cone beam reconstruction is much more accurate, therefore a smaller clearance was sufficient for the experiments. A tip-detection algorithm was used to detect the final tip position in the image. Since there was no ground truth available for the position of the tool tip with respect to the surface along the surface normal dimension, the error was defined in terms of 2D pixel distance between the detected tip position and the target point. Figure 3.4 shows the target points and the detected tip positions for the cone beam reconstruction and stereo reconstruction respectively in a dry eye phantom. Figure 3.6 shows the plot of pixel error with the corresponding target. Figures 3.5 and 3.7 show the same for a wet eye phantom.

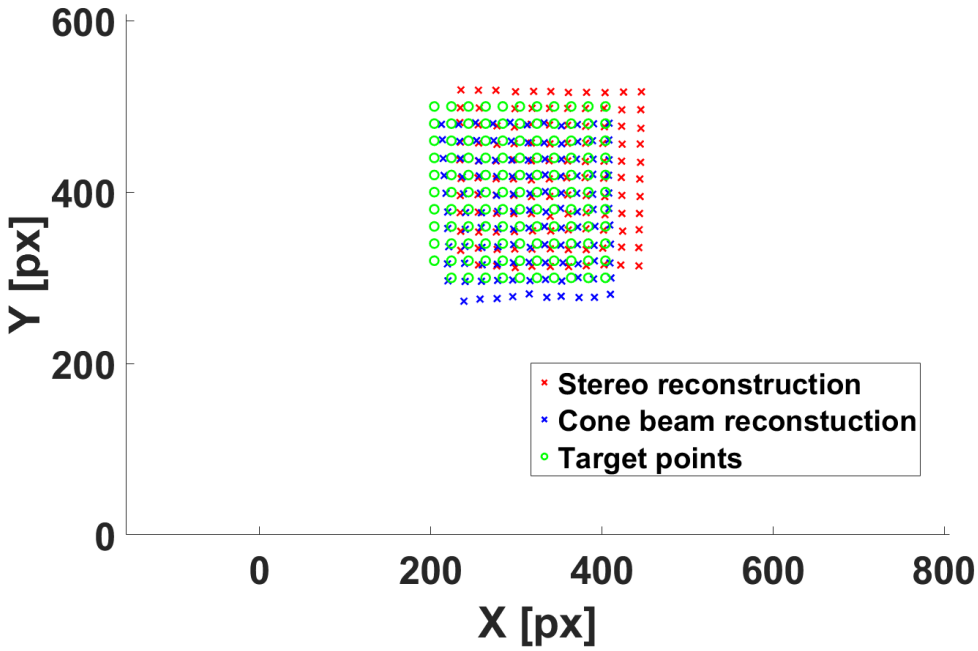


Figure 3.4: Map of target points for cannulation and the corresponding tool tip location in image coordinates for cone beam reconstruction and stereo reconstruction, in a dry eye phantom.

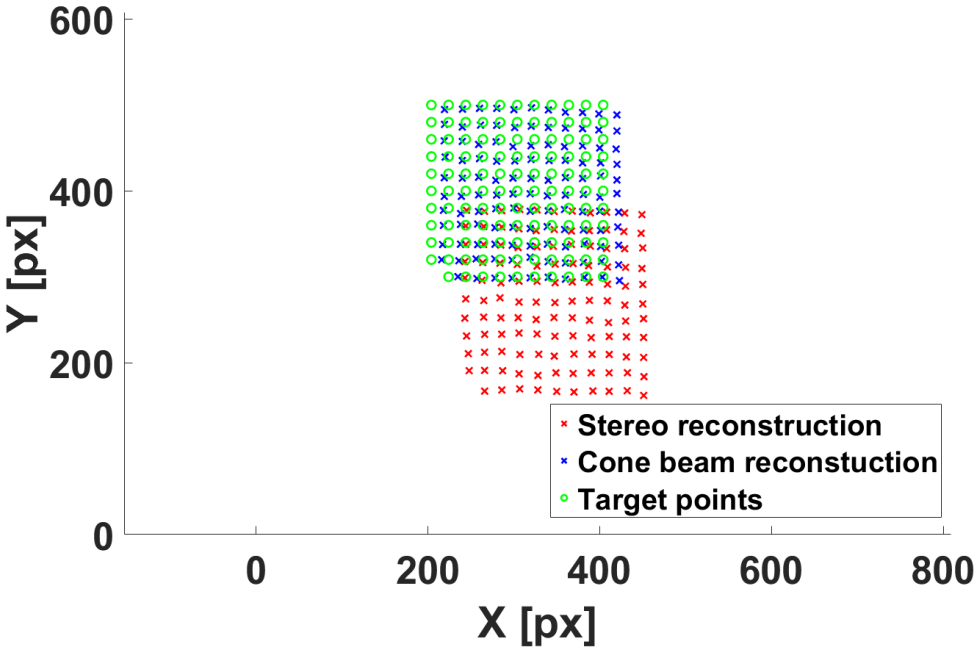


Figure 3.5: Map of target points for cannulation and the corresponding tool tip location in image coordinates for cone beam reconstruction and stereo reconstruction, in a wet eye phantom.

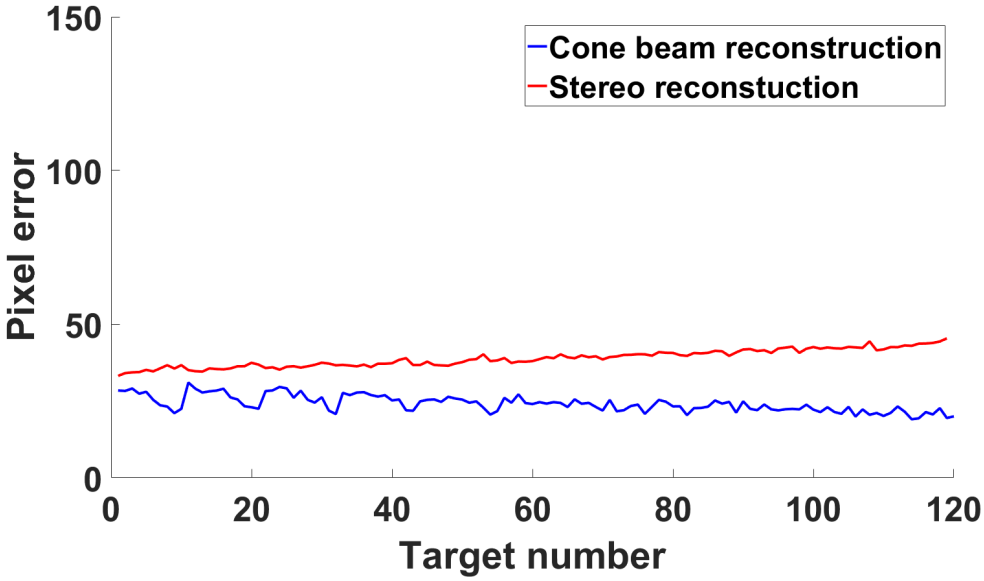


Figure 3.6: 2D Pixel error between target point and detected tip position for cone beam reconstruction and stereo reconstruction, in a dry eye phantom.

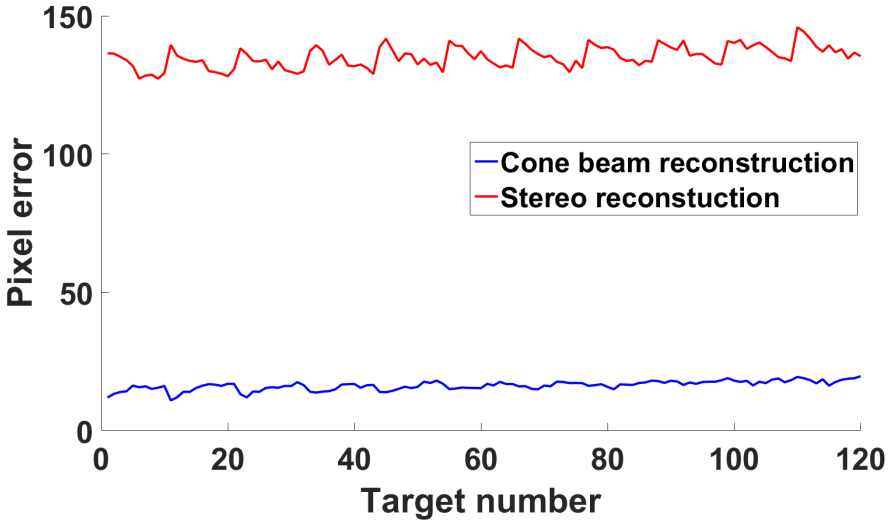


Figure 3.7: 2D Pixel error between target point and detected tip position for cone beam reconstruction and stereo reconstruction, in a wet eye phantom.

The accuracy of the surface reconstruction was compared to a standard stereo algorithm which was impelemented as follows. The projection matrices for the cameras were calibrated as in [63]. The laser probe was swept over the retinal surface while detecting the 2D tip positions $p_c^i \in \mathbb{R}^2, \forall c \in \{L, R\}$ in the left and right images, which were matched with the corresponding 3D ASAP tip positions $P_{ASAP}^i \in \mathbb{R}^3$. Thus we obtain camera projection matrices

$M_c \in \mathbb{R}^{3X4}, \forall c \in \{L, R\}$ using DLT, which are then used for triangulating a goal position in 3D.

$$\boldsymbol{p_{tip}} = \boldsymbol{M_c P_{tip}^{ASAP}} \tag{3.14}$$

Table 3.1: Average pixel error.

	Stereo	Cone beam
Air	38.9 px	24.1 px
Water	135.2 px	16.3 px

Table 3.2: Standard deviation of pixel error.

	Stereo	Cone beam
Air	2.8 px	2.6 px
Water	3.9 px	1.7 px

Table 3.1 shows the average pixel error between the target and the detected tool tip. It was observed that the average pixel error between the target and the detected tip positions was similar in a dry eye phantom for the two cases. However, in a wet eye phantom, the pixel error in the cone beam reconstruction case was about 12% of the error in the stereo reconstruction case.

3.7.2 Motion Scaling

In order to demonstrate the automatic activation of motion scaling close to the target, the approach of the vessel during cannulation was simulated in the wet eye phantom. The cannulation needle was attached to the end of the laser probe, and the tool tip calibration as explained in 3.2 was re-run to account for the needle length. The motion was scaled about the target point and the scale was linearly decreased from $\lambda_0 = 1$ to $\lambda_0 = 0.3$ as the distance between the target and the needle tip decreased from R = 4 mm to 0 mm. Figures 3.8, 3.9 and 3.10 show the X,Y and Z positions of the tool tip in the ASAP coordinates for three conditions: no tremor cancellation, normal tremor cancellation and motion scaling with tremor cancellation.

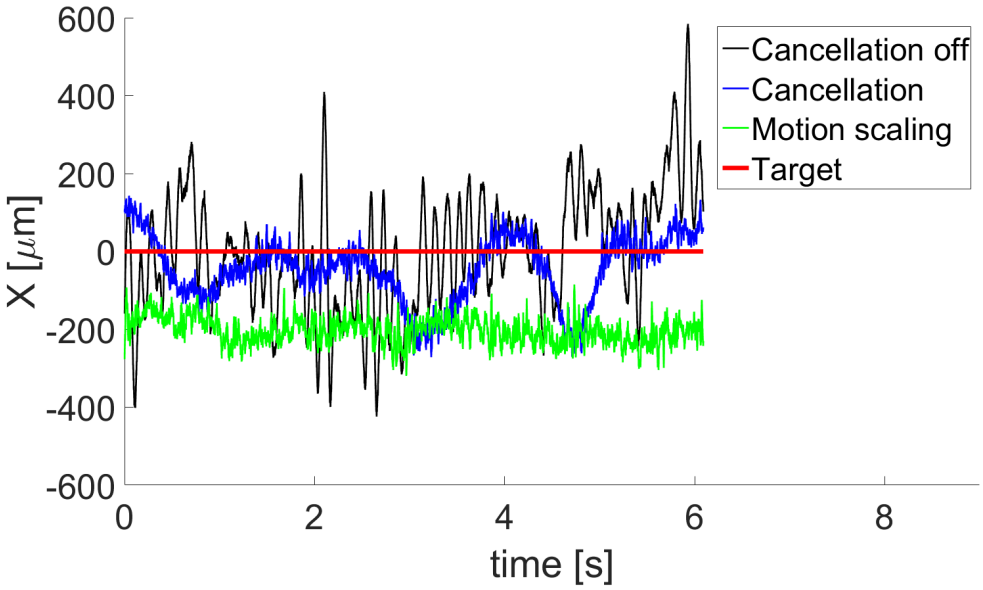


Figure 3.8: X position of the needle tip with time while holding steady at the cannulation target in the wet eye phantom, in ASAP coordinates.

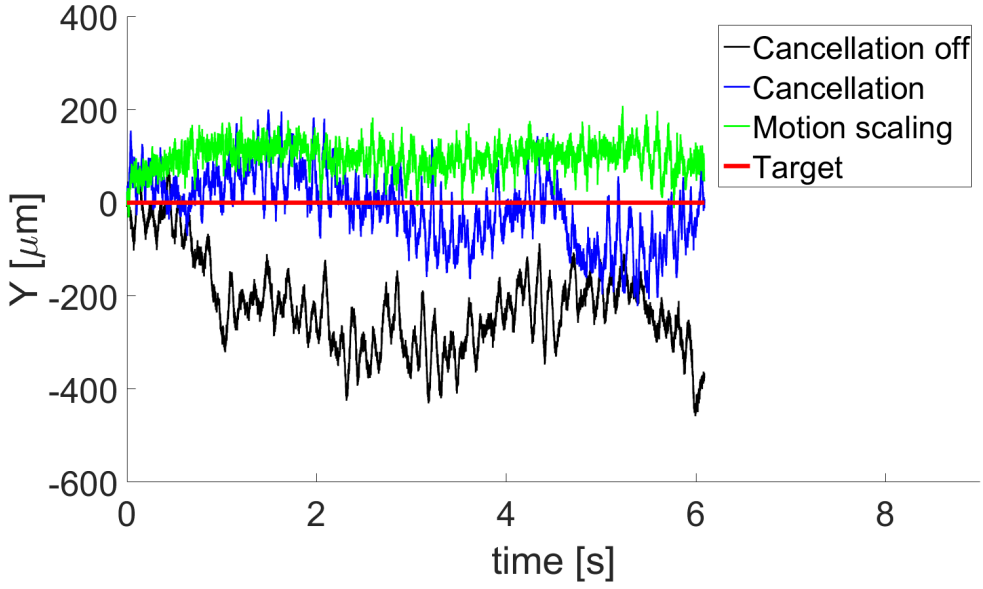


Figure 3.9: Y position of the needle tip with time while holding steady at the cannulation target in the wet eye phantom, in ASAP coordinates.

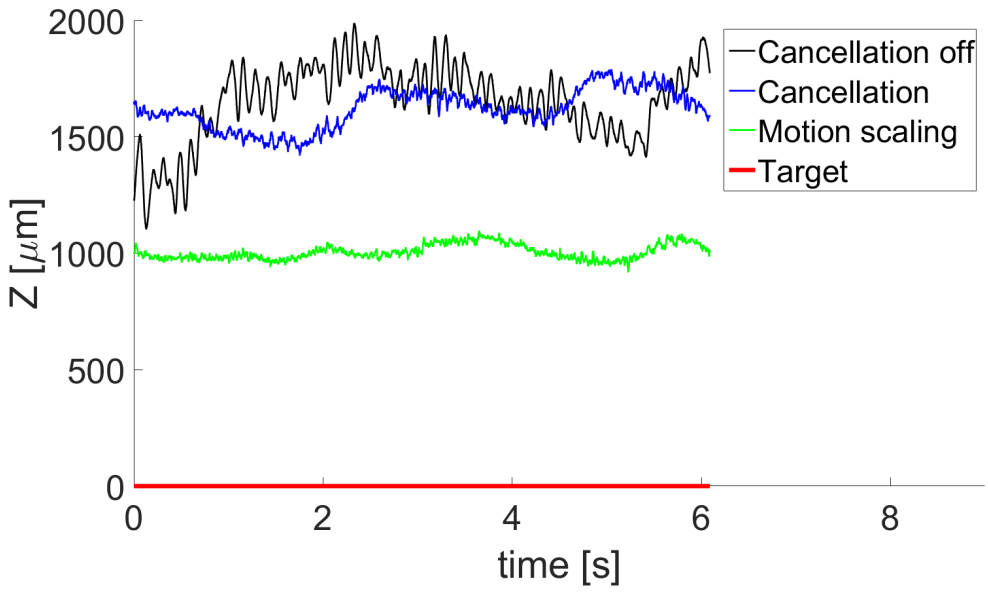


Figure 3.10: Z position of the needle tip with time while holding steady at the cannulation target in the wet eye phantom, in ASAP coordinates.

Table 3.3: Mean error and standard deviation of errors in ASAP coordinates.

	μ_{net} [mm]	σ_{net} [mm]	σ_x [mm]	σ_y [mm]	σ_z [mm]
Cancellation off	1.68	0.176	0.153	0.097	0.165
Cancellation on	1.62	0.086	0.082	0.074	0.085
Motion scaling	1.03	0.031	0.034	0.030	0.033

Table 3.3 shows the mean and standard deviation of 3D errors measured in ASAP coordinates. Error standard deviation is the most meaningful metric here. Mean error or bias is primarily a function of the operator's visual perception, and is not affected by the technique presented here, except to the extent that cancellation of higher frequency error may assist the operator to correctly perceive mean positions. With motion scaling engaged, the standard deviation in error was about 36% of the standard deviation with pure cancellation and 18% of the error with no cancellation. The duration for the experiments was chosen based on our prior experience of using Micron in simulated surgery settings [64] and has no influence on the proposed method.

3.8 Discussion

In this work, image guided retinal vessel cannulation was proposed. Challenges faced in our previous attempts at cannulation were successfully solved using monocular vision based surface reconstruction and subsequent coordinate mapping. The error in the coordinate mapping was

about 12% of the error from a standard stereo-based triangulation approach, in a realistic wet eye phantom. The automated motion-scaling methodology that relies on this coordinate mapping led to increased precision as indicated by the 64% reduction in the standard deviation of error, compared to pure tremor cancellation. As can be seen in Figures 9-11, the high-frequency oscillations due to hand tremor are clearly large when tremor compensation is turned off. It can also be seen that the low-frequency oscillations are much lesser in amplitude when motion scaling is active as compared to filter-based tremor compensation. This is expected since any high-frequency deviation which passes through the low-pass shelving filter is scaled down. Importantly, as evident from Figure 11, the error in the Z axis, which was roughly perpendicular to the plane, was less in the motion-scaling case, showing that it was easier to keep the cannulation needle closer to the target without hitting the retinal plane when motion scaling was engaged. Retinal vessels are around $100\,\mu\mathrm{m}$ in diameter. As reported in Table III., the standard deviation in error in our approach is less than $100\,\mu\mathrm{m}$. In contrast, in the unaided case, the error standard deviation is over $100\,\mu\mathrm{m}$. This indicates the advantage of using the methods developed here for retinal vessel cannulation.

Chapter 4

Force Control

In this chapter we describe our work towards force control during Micron-aided retinal vein cannulation. Unlike similar applications involving needle insertion in vessels, such as sampling of blood from the vessels of the forearm, the required forces for cannulating retinal vessels are almost imperceptible. The chorioallantoic membrane (CAM) of the chicken embryo is a valid model in vivo for the human retina [66] and has been used in retinal vein cannulation studies. Such studies have confirmed that most forces are below the human perception threshold [67]. Therefore, there is a need to enforce force constraints during cannulation so that damage to the retinal tissue can be avoided.

4.1 Integrated Cannulation Tool

In Chapter 3, we presented a laser scan based surface reconstruction technique that can be used to transform targets defined in the image coordinates to ASAP coordinates, in order to achieve motion scaling close to the target. In order to facilitate force control along with laser-based reconstruction through one tool, an integrated cannulation tool was developed by our collaborators at JHU. The tool is similar to the one developed by Gonenc et. al [68]. It consists of 3 fiber Bragg grating (FBG) sensors fixed symmetrically around the tool shaft. It can measure the transverse forces on the tip in 2D. It also incorporates a surgical laser probe and a micro-needle. The micro-needle has a beveled tip and is pre-bent relative to the tool axis, enabling a gradual and a safer approach to the retina surface. An outer tube carries the FBGs and and covers the laser probe and the micro-needle. The outer tube is retractable, and is actuated by a linear micro-motor (Squiggle-RV-1.8, , Victor, NY).

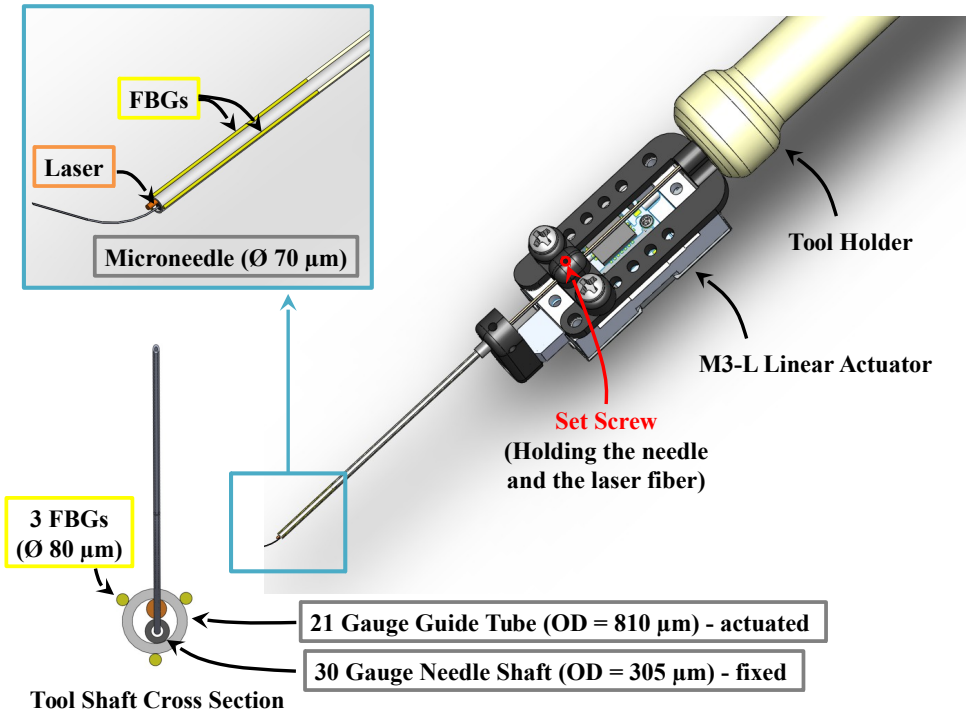


Figure 4.1: Design of the integrated cannulation tool.

Figure 4.2: Micron with the integrated tool.

4.2 Hybrid Position and Force Control

Force control during retinal membrane peeling with Micron was developed in [69]. The system is based on the classic hybrid position/force control framework proposed by Craig and Raibert [70]. Here we adapt the concept for force control during retinal vessel cannulation. This work is ongoing, and no quantitative evaluation of force controlled has not been done yet.

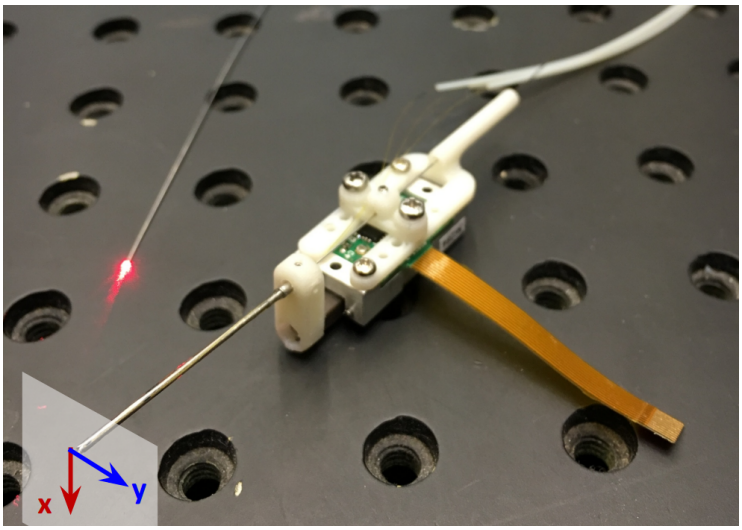


Figure 4.3: Coordinate system of the force tool.

The forces are only measured along the two transverse directions to the tool X and Y, shown in Figure 4.3. The force along the tool is set to 0. The measured force is then converted from the tool coordinates to the ASAP coordinates. Note that the third component of force is 0 in the tool coordinates but not in the ASAP coordinates. A low-pass filter is then used to reduce noise in the force measurements.

There is a soft (α) and a hard (β) threshold specified for the force control along all three directions. No force control is active for forces below α. The force filter filters the force error $e_f = F - \alpha$ according to equation 4.1, along all three directions independently.

$$e_{fc} = \begin{cases} 0, & F < \alpha \\ e_f \frac{F-\alpha}{\beta-\alpha}, & \alpha \leq F \leq \beta \\ e_f & F > \beta \end{cases} \tag{4.1}$$

The filtered force is then sent to a PID force corrector. It has gains $K_p = 1.4$, $K_d = 0.15$ and $K_i = 1.4$, with a gain of 0.005 on an exponential decay term for the integral gain K_i. The force corrector block outputs a position offset which is then added to the position goal of the position loop.

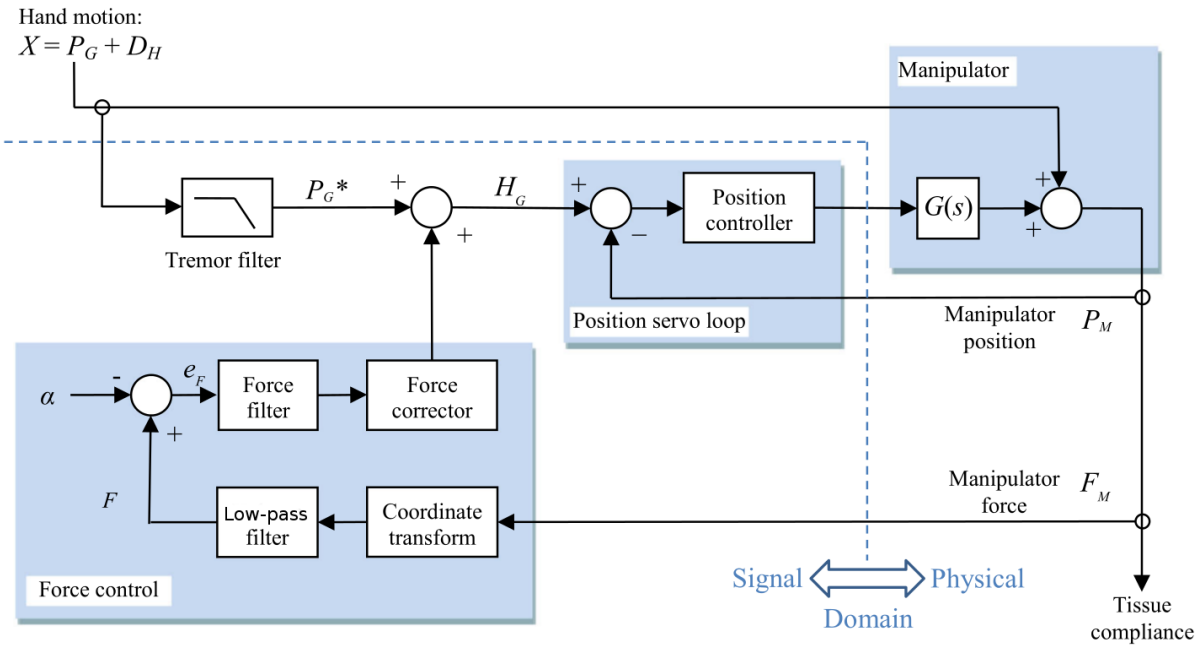


Figure 4.4: Block diagram of the hybrid position/force control loop implemented on Micron [69].

Chapter 5

Summary

In this thesis we presented a framework for Micron-aided retinal vessel cannulation. In this chapter we summarize our main contributions and give directions for future research.

5.1 Pose-graphs Based Retinal SLAM

Our first contribution is an algorithm to estimate eyeball motion by simultaneous localization and mapping of the retina. We developed a factor graph formulation of the retinal SLAM problem. We described our approach to generate pair-wise constraints between frames which lead to incremental motion factors and loop-closure constraints. The graph was then optimized using the iSAM2 algorithm. We presented our qualitative results on real eye surgery videos. We also showed quantitative improvement over our previous EyeSLAM algorithm, on video sequences of an eye phantom.

5.2 Surface Re-construction and Homography Matrix Estimation

Our second contribution is a homography matrix estimation, which converts targets defined in the image coordinates into the global ASAP coordinates. We described our monocular camera based surface reconstruction using laser scanning and the homography matrix estimation. We presented a motion scaling framework around the target and transition from and to normal tremor cancellation mode. We evaluated the accuracy of the surface reconstruction and compared it to a standard stero-camera-based surface reconstruction, showing considerable improvement. We also demonstrated the ease of using Micron with our motion scaling framework over normal tremor cancellation.

5.3 Force Control

Our third contribution is a force control framework using a tool that integrates a cannulation micro-needle, force sensors and a laser probe. We described our approach to achieving hybrid position/force control.

5.4 Future Work

Future improvements in EyeSLAM should include optimization to allow the algorithm to run in real-time on high-definition videos. Furthermore, implementing 3D SLAM using stereo-vision, while handling distortion could be beneficial.

The monocular camera based surface reconstruction can be improved by developing real-time continuous surface re-estimation to account for eyeball movement once the surface has been reconstructed. Furthermore, it would be useful if arbitrary beam patterns created by manual scanning could also allow estimation of the surface.

The hybrid force/position control should be validated with experiments in eye phantoms and animal retinal tissue ex vivo and in vivo. Force sensing and vision can be fused to develop sophisticated force fixtures. For example, depending upon the location of the needle, the soft and hard thresholds for force control can be adjusted. The force information should also be used for puncture detection which should trigger automatic control mode transitions and force fiber rebiasing.

Finally, all these subsystems need to be seamlessly integrated to achieve retinal vessel cannulation. Experiments should be carried out in live animals in order to validate the advantages of using Micron for retinal vessel cannulation.

Bibliography

[1] I. Fleming, M. Balicki, J. Koo, I. Iordachita, B. Mitchell, J. Handa, G. Hager, and R. Taylor, "Cooperative robot assistant for retinal microsurgery," *Medical Image Computing and Computer-Assisted Intervention–MICCAI*, pp. 543–550, 2008. 1

[2] F. W. Mohr, V. Falk, A. Diegeler, T. Walther, J. F. Gummert, J. Bucerius, S. Jacobs, and R. Autschbach, "Computer-enhanced robotic cardiac surgery: experience in 148 patients," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 121, no. 5, pp. 842–853, 2001. 1

[3] P. Modi, E. Rodriguez, and W. R. Chitwood, "Robot-assisted cardiac surgery," *Interactive Cardiovascular and Thoracic Surgery*, vol. 9, no. 3, pp. 500–505, 2009. 1

[4] J. F. Kuniholm, G. Buckner, W. Nifong, and M. Orrico, "Automated knot tying for fixation in minimally invasive, robot-assisted cardiac surgery," *Journal of Biomechanical Engineering*, vol. 127, no. 6, pp. 1001–1008, 2005. 1

[5] V. Falk, T. Walther, R. Autschbach, A. Diegeler, R. Battellini, and F. W. Mohr, "Robot-assisted minimally invasive solo mitral valve operation," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 115, no. 2, pp. 470–471, 1998. 1

[6] J. Himpens, G. Leman, and G. Cadiere, "Telesurgical laparoscopic cholecystectomy." *Surgical Endoscopy*, vol. 12, no. 8, p. 1091, 1998. 1

[7] C. Gutt, T. Oniu, A. Mehrabi, A. Kashfi, P. Schemmer, and M. Büchler, "Robot-assisted abdominal surgery," *British Journal of Surgery*, vol. 91, no. 11, pp. 1390–1397, 2004. 1

[8] E. J. Hanly and M. A. Talamini, "Robotic abdominal surgery," *The American Journal of Surgery*, vol. 188, no. 4, pp. 19–26, 2004. 1

[9] S. Kaul, R. Laungani, R. Sarle, H. Stricker, J. Peabody, R. Littleton, and M. Menon, "da vinci-assisted robotic partial nephrectomy: technique and results at a mean of 15 months of follow-up," *European Urology*, vol. 51, no. 1, pp. 186–192, 2007. 1

[10] A. W. Partin, J. B. Adams, R. G. Moore, and L. R. Kavoussi, "Complete robot-assisted laparoscopic urologic surgery: a preliminary report." *Journal of the American College of Surgeons*, vol. 181, no. 6, pp. 552–557, 1995. 1

[11] A. Tewari, A. Srivasatava, and M. Menon, "A prospective comparison of radical retropubic and robot-assisted prostatectomy: Experience in one institution," *BJU International*, vol. 92, no. 3, pp. 205–210, 2003. 1

[12] R. Littleton, S. Kaul, R. Sarle, R. Laungani, H. Stricker, J. Peabody, and M. Menon, "Da

vinci assisted robotic partial nephrectomy: Technique and initial experience at the vattikuti urology institute," *Surgical Laparoscopy Endoscopy & Percutaneous Techniques*, vol. 16, no. 4, pp. 285–286, 2006. 1

[13] P. R. Rizun, P. B. McBeth, D. F. Louw, and G. R. Sutherland, "Robot-assisted neurosurgery," in *Seminars in Laparoscopic Surgery*, vol. 11, no. 2, 2004, pp. 99–106. 1

[14] C. Meng, T. Wang, W. Chou, S. Luan, Y. Zhang, and Z. Tian, "Remote surgery case: robot-assisted teleneurosurgery," in *Proceedings IEEE International Conference on Robotics and Automation*, vol. 1, 2004, pp. 819–823. 1

[15] P. D. Le Roux, H. Das, S. Esquenazi, and P. J. Kelly, "Robot-assisted microsurgery: a feasibility study in the rat," *Neurosurgery*, vol. 48, no. 3, pp. 584–589, 2001. 1

[16] E. Glassman, W. A. Hanson, P. Kazanzides, B. D. Mittelstadt, B. L. Musits, H. A. Paul, and R. H. Taylor, "Image-directed robotic system for precise robotic surgery including redundant consistency checking," Feb. 4 1992, US Patent 5,086,401. 1

[17] W. L. Bargar, A. Bauer, and M. Börner, "Primary and revision total hip replacement using the robodoc (r) system." *Clinical Orthopaedics and Related Research*, vol. 354, pp. 82–91, 1998. 1

[18] R. H. Taylor, L. Joskowicz, B. Williamson, A. Guéziec, A. Kalvin, P. Kazanzides, R. Van Vorhis, J. Yao, R. Kumar, A. Bzostek *et al.*, "Computer-integrated revision total hip replacement surgery: concept and preliminary results," *Medical Image Analysis*, vol. 3, no. 3, pp. 301–319, 1999. 1

[19] S. Rogers, R. L. McIntosh, N. Cheung, L. Lim, J. J. Wang, P. Mitchell, J. W. Kowalski, H. Nguyen, T. Y. Wong, I. E. D. Consortium *et al.*, "The prevalence of retinal vein occlusion: pooled data from population studies from the United States, Europe, Asia, and Australia," *Ophthalmology*, vol. 117, no. 2, pp. 313–319, 2010. 1.1

[20] L. A. Bynoe, R. K. Hutchins, H. S. Lazarus, and M. A. Friedberg, "Retinal endovascular surgery for central retinal vein occlusion: initial experience of four surgeons," *Retina*, vol. 25, no. 5, pp. 625–632, 2005. 1.1

[21] J. N. Weiss and L. A. Bynoe, "Injection of tissue plasminogen activator into a branch retinal vein in eyes with central retinal vein occlusion," *Ophthalmology*, vol. 108, no. 12, pp. 2249–2257, 2001. 1.1

[22] S. S. Park, J. Sigelman, E. Gragoudas, and W. Tasman, "The anatomy and cell biology of the retina," *Duanes Foundations of Clinical Ophthalmology*, vol. 1, pp. 21–50, 1996. 1.1

[23] P. Schenker, H. Das, and T. Ohm, "A new robot for high dexterity microsurgery," in *Computer Vision, Virtual Reality and Robotics in Medicine*, 1995, pp. 115–122. 1.2

[24] P. S. Jensen, K. W. Grace, R. Attariwala, J. E. Colgate, and M. R. Glucksberg, "Toward robot-assisted vascular microsurgery in the retina," *Graefe's Archive for Clinical and Experimental Ophthalmology*, vol. 235, no. 11, pp. 696–701, 1997. 1.2

[25] G. H. Ballantyne and F. Moll, "The da Vinci telerobotic surgical system: the virtual operative field and telepresence surgery," *Surgical Clinics of North America*, vol. 83, no. 6, pp. 1293–1304, 2003. 1.2

[26] A. P. Mulgaonkar, J.-P. Hubschman, J.-L. Bourges, B. L. Jordan, C. Cham, J. T. Wilson, T.-C. Tsao, and M. O. Culjat, "A prototype surgical manipulator for robotic intraocular micro surgery," *Stud Health Technol Inform*, vol. 142, no. 1, pp. 215–7, 2009. 1.2

[27] E. Rahimy, J. Wilson, T. Tsao, S. Schwartz, and J. Hubschman, "Robot-assisted intraocular surgery: development of the IRISS and feasibility studies in an animal model," *Eye*, vol. 27, no. 8, pp. 972–978, 2013. 1.2

[28] T. Nakano, N. Sugita, T. Ueta, Y. Tamaki, and M. Mitsuishi, "A parallel robot to assist vitreoretinal surgery," *International Journal of Computer Assisted Radiology and Surgery*, vol. 4, no. 6, pp. 517–526, 2009. 1.2

[29] R. Taylor, P. Jensen, L. Whitcomb, A. Barnes, R. Kumar, D. Stoianovici, P. Gupta, Z. Wang, E. Dejuan, and L. Kavoussi, "A steady-hand robotic system for microsurgical augmentation," *The International Journal of Robotics Research*, vol. 18, no. 12, pp. 1201–1210, 1999. 1.2

[30] B. Mitchell, J. Koo, I. Iordachita, P. Kazanzides, A. Kapoor, J. Handa, G. Hager, and R. Taylor, "Development and application of a new steady-hand manipulator for retinal surgery," in *IEEE International Conference on Robotics and Automation*, 2007, pp. 623–629. 1.2

[31] A. Üneri, M. A. Balicki, J. Handa, P. Gehlbach, R. H. Taylor, and I. Iordachita, "New steady-hand eye robot with micro-force sensing for vitreoretinal surgery," in *Proceedings IEEE International Conference on Biomedical Robotics and Biomechatronics*. IEEE, 2010, pp. 814–819. 1.2, 1.1

[32] C. N. Riviere, W. T. Ang, and P. K. Khosla, "Toward active tremor canceling in handheld microsurgical instruments," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 5, pp. 793–800, 2003. 1.3

[33] C. N. Riviere, R. S. Rader, and N. V. Thakor, "Adaptive cancelling of physiological tremor for improved precision in microsurgery," *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 7, pp. 839–846, 1998. 1.3

[34] W. T. Ang, P. Pradeep, and C. Riviere, "Active tremor compensation in microsurgery," in *International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1. IEEE, 2004, pp. 2738–2741. 1.3

[35] D. Y. Choi and C. N. Riviere, "Flexure-based manipulator for active handheld microsurgical instrument," in *International Conference of the Engineering in Medicine and Biology Society*. IEEE, 2005, pp. 2325–2328. 1.3

[36] R. A. MacLachlan, B. C. Becker, J. C. Tabarés, G. W. Podnar, L. A. Lobes Jr, and C. N. Riviere, "Micron: an actively stabilized handheld tool for microsurgery," *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 195–212, 2012. 1.3, 1.2

[37] R. A. MacLachlan and C. N. Riviere, "High-speed microscale optical tracking using digital frequency-domain multiplexing," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 6, pp. 1991–2001, 2009. 1.3, 1.3

[38] S. Yang, R. A. MacLachlan, and C. N. Riviere, "Manipulator design and operation of a six-degree-of-freedom handheld tremor-canceling microsurgical instrument," *IEEE/ASME*

Transactions on Mechatronics, vol. 20, no. 2, pp. 761–772, 2015. 1.3

[39] S. Yang, "Handheld micromanipulator for robot-assisted microsurgery," Ph.D. dissertation, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, May 2015. 1.3, 3, 3.2, 3.3, 3.2, 3.3, 3.3

[40] B. C. Becker, S. Voros, L. A. Lobes, J. T. Handa, G. D. Hager, and C. N. Riviere, "Retinal vessel cannulation with an image-guided handheld robot," in *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2010, pp. 5420–5423. 1.4

[41] B. Gonenc, N. Tran, P. Gehlbach, R. H. Taylor, and I. Iordachita, "Robot-assisted retinal vein cannulation with force-based puncture detection: Micron vs. the steady-hand eye robot," in *IEEE International Conference of the Engineering in Medicine and Biology Society*, 2016, pp. 5107–5111. 1.4

[42] M. D. de Smet, T. C. Meenink, T. Janssens, V. Vanheukelom, G. J. Naus, M. J. Beelen, C. Meers, B. Jonckx, and J.-M. Stassen, "Robotic assisted cannulation of occluded retinal veins," *PloS one*, vol. 11, no. 9, p. e0162037, 2016. 1.4

[43] A. Gijbels, K. Willekens, L. Esteveny, P. Stalmans, D. Reynaerts, and E. Vander Poorten, "Towards a clinically applicable robotic assistance system for retinal vein cannulation," in *IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, 2016, pp. 284–291. 1.4

[44] K. Willekens, A. Gijbels, L. Schoevaerdts, L. Esteveny, T. Janssens, B. Jonckx, J. H. Feyen, C. Meers, D. Reynaerts, E. Vander Poorten *et al.*, "Robot-assisted retinal vein cannulation in an in vivo porcine retinal vein occlusion model," *Acta Ophthalmologica*, 2017. 1.4

[45] B. C. Becker and C. N. Riviere, "Real-time retinal vessel mapping and localization for intraocular surgery," in *IEEE International Conference on Robotics and Automation*, 2013, pp. 5360–5365. 2, 2.5, 3.5

[46] D. Braun, S. Yang, J. N. Martel, C. N. Riviere, and B. C. Becker, "Real-time localization and mapping of retinal vessels during intraocular microsurgery," *Int. J. Med. Robot.,* submitted 2016. 2, 2.3, 2.4, 2.5, 2.7.2, 3.5

[47] E. B. Olson, "Real-time correlative scan matching," in *IEEE International Conference on Robotics and Automation*, 2009, pp. 4387–4393. 2

[48] F. Dellaert and M. Kaess, "Factor graphs for robot perception," *Foundations and Trends in Robotics*, 2017, submitted. 2.2, 2.1

[49] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and F. Dellaert, "iSAM2: Incremental smoothing and mapping using the Bayes tree," *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 216–235, 2012. 2.2

[50] J. Engel, J. Sturm, and D. Cremers, "Semi-dense visual odometry for a monocular camera," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 1449–1456. 2.3

[51] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," in *European Conference on Computer Vision*, 2014, pp. 834–849. 2.3

[52] D. G. Lowe, "Object recognition from local scale-invariant features," in *The proceedings of*

the IEEE International Conference on Computer Vision, vol. 2, 1999, pp. 1150–1157. 2.3

[53] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008. 2.3

[54] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," *Computer vision–ECCV 2006*, pp. 430–443, 2006. 2.3

[55] C. Harris and M. Stephens, "A combined corner and edge detector." in *Alvey Vision Conference*, vol. 15, no. 50, 1988, pp. 10–5244. 2.3

[56] A. Can, H. Shen, J. N. Turner, H. L. Tanenbaum, and B. Roysam, "Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms," *IEEE Transactions on Information Technology in Biomedicine*, vol. 3, no. 2, pp. 125–138, 1999. 2.3, 2.2

[57] P. Newman and K. Ho, "SLAM-loop closing with visually salient features," in *IEEE International Conference on Robotics and Automation*, 2005, pp. 635–642. 2.6

[58] R. Richa, M. Balicki, E. Meisner, R. Sznitman, R. Taylor, and G. Hager, "Visual tracking of surgical tools for proximity detection in retinal surgery," in *International Conference on Information Processing in Computer-Assisted Interventions*. Springer, 2011, pp. 55–66. 3

[59] R. Richa, M. Balicki, R. Sznitman, E. Meisner, R. Taylor, and G. Hager, "Vision-based proximity detection in retinal surgery," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 8, pp. 2291–2301, 2012. 3

[60] R. Sznitman, R. Richa, R. H. Taylor, B. Jedynak, and G. D. Hager, "Unified detection and tracking of instruments during retinal microsurgery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 5, pp. 1263–1273, 2013. 3

[61] S. Yang, R. A. MacLachlan, and C. N. Riviere, "Toward automated intraocular laser surgery using a handheld micromanipulator," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Sept 2014, pp. 1302–1307. 3.2

[62] B. C. Becker, R. A. MacLachlan, L. A. Lobes, G. D. Hager, and C. N. Riviere, "Vision-based control of a handheld surgical micromanipulator with virtual fixtures," *IEEE Transactions on Robotics*, vol. 29, no. 3, pp. 674–683, June 2013. 3.3

[63] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003. 3.4, 3.7.1

[64] R. A. MacLachlan, B. C. Becker, J. C. Tabarés, G. W. Podnar, L. A. L. Jr., and C. N. Riviere, "Micron: An actively stabilized handheld tool for microsurgery," *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 195–212, Feb 2012. 3.6, 3.7.2

[65] B. Becker, "Vision-based control of a handheld micromanipulator for robot-assisted retinal surgery," Ph.D. dissertation, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, September 2012. 3.6

[66] T. Leng, J. M. Miller, K. V. Bilbao, D. V. Palanker, P. Huie, and M. S. Blumenkranz, "The chick chorioallantoic membrane as a model tissue for surgical retinal research and simulation," *Retina*, vol. 24, no. 3, pp. 427–434, 2004. 4

[67] O. Ergeneman, J. Pokki, V. Pocepcova, H. Hall, J. J. Abbott, and B. J. Nelson, "Characterization of puncture forces for retinal vein cannulation," *Journal of Medical Devices*, vol. 5, no. 4, p. 044504, 2011. 4

[68] B. Gonenc, R. H. Taylor, I. Iordachita, P. Gehlbach, and J. Handa, "Force-sensing microneedle for assisted retinal vein cannulation," in *IEEE SENSORS*, 2014, pp. 698–701. 4.1

[69] T. S. Wells, S. Yang, R. A. MacLachlan, L. A. Lobes, J. N. Martel, and C. N. Riviere, "Hybrid position/force control of an active handheld micromanipulator for membrane peeling," *The International Journal of Medical Robotics and Computer Assisted Surgery*, 2015. 4.2, 4.4

[70] J. J. Craig and M. H. Raibert, "A systematic method of hybrid position/force control of a manipulator," in *Proceedings IEEE Computer Society International Computer Software and Applications Conference*, 1979, pp. 446–451. 4.2