EyeSLAM: Real-Time Simultaneous Localization and Mapping (SLAM) of Retinal Vessels during Intraocular Microsurgery

Daniel Braun, **Sungwook Yang**, **Joseph N. Martel, M.D.**, **Cameron N. Riviere**, and **Brian C. Becker**

Carnegie Mellon University, The Robotics Institute

Abstract

Background—Fast and accurate mapping and localization of the retinal vasculature is critical to increasing the effectiveness and clinical utility of robotic-assisted intraocular microsurgery such as laser photocoagulation and retinal vessel cannulation.

Methods—The proposed EyeSLAM algorithm delivers 30 Hz real-time simultaneous localization and mapping of the human retina and vasculature during intraocular surgery, combining fast vessel detection with 2D scan-matching techniques to build and localize a probabilistic map of the vasculature.

Results—In the harsh imaging environment of retinal surgery with high magnification, quick shaky motions, textureless retina background, variable lighting, and tool occlusion, EyeSLAM can map 75% of the vessels within two seconds of initialization and localize the retina in realtime with a Root Mean Squared (RMS) error of under 5.0 pixels (translation) and 1.0 degree (rotation).

Conclusions—EyeSLAM robustly provides retinal maps and registration that enable intelligent surgical micromanipulators to aid surgeons in simulated retinal vessel tracing and photocoagulation tasks.

Introduction

A challenging area of surgery is intraocular microsurgery because high precision is required to work with the tiny and fragile retinal tissues, which is made more difficult by physiological tremor decreasing surgical performance at such small scales (1, 2). For example, surgeons must be able to accurately place laser burns within fractions of a millimeter but not accidentally burn retinal vessels (3) and peel retinal membranes under 10 μm thick (4). Advances in retinal vessel cannulation require the microsurgeon to inject anti-coagulants into vessels less than 100 μm in diameter (5).

Corresponding Author, Cameron N. Riviere (camr@ri.cmu.edu), 5000 Forbes Avenue, Pittsburgh PA 15213-3890, US, 1-412-268-3818.

Micromanipulation aids often depend on robotics and computer assisted surgery and various approaches have been built (6–8). Our lab has developed Micron, a fully handheld micromanipulator paired with vision-based control to help surgeons performing retinal surgery (9). Robotic control for microsurgery has advanced from scaling motion, limiting velocity, and regulating force to more sophisticated control methods made possible by understanding the tool position relative to the retinal structures in real-time. Merging camera imagery of the surgical scene with semi-automatic micromanipulation, vision-based control assists procedures and prevents mistakes (10). For instance, if laser burn placement during retinal laser photocoagulation is imprecise or errant, as in inadvertent retinal vasculature or macular photocoagulation, then excessive retinal ischemia or direct laser-induced retinal damage may impair visual function (11). During retinal vessel cannulation, robot aid can help guide the cannula to enable a surgeon to more easily inject drugs into a vessel only 100–200 μm across (12).

Although many algorithms increase the effectiveness of robotic aid for retinal surgery, this paper focuses on localization and mapping of retinal vessels. We roughly categorize the related work into vessel detection, retinal registration, and the more general robotic approach of simultaneous localization and mapping (SLAM). Highlights of each are provided in Table 1. Most methods are slow and high performance, which limits their usefulness for real-time microsurgery applications. Becker and Riviere introduced a formulation of simultaneous localization and mapping for retinal vasculature that ran in real-time using fast vessel detection and smoothing noisy observations over time while building and localizing to a map (13). We extend this work and introduce a new algorithm, called EyeSLAM. It exhibits robustness to variable illumination conditions, high magnification, quick shaky motion, textureless retina, and transient occlusions. Specifically, this paper introduces two new components of the EyeSLAM algorithm: (a) a scan-based localization algorithm with a dynamic map for higher convergence accuracy and increased tolerance of quick motion; and (b) more robust vessel detection with better rejection of spurious detections. The paper includes more quantitative results showing the effectiveness of EyeSLAM on simulated and recorded surgical video imagery, along with qualitative demonstrations of its usefulness in simulated surgical scenarios.

Vessel Detection

Given a single image of the retina, vessel detection extracts information such as location, width, and orientation of the visible vasculature. Per-pixel approaches classify each location in the image as vessel or non-vessel (14), identifying vasculature using popular approaches such as matched filters (15, 16) or Gabor filters (17, 18). Focusing on high performance using static, high-resolution fundus images at low magnification, many of these approaches' runtimes exceed 1 second, which is too slow for incorporation into a high-speed robotic feedback loop. Some approaches decrease runtime requirements via faster algorithms (16, 18) or hardware optimizations (19). Can et al. (16) is interesting in that it balances speed against performance by finding sparse sets of points on vessels and then tracing each vessel with a set of matched filters, dynamically estimating the direction and size of the vessel at each step. The entire vasculature is then obtained without having to touch pixels in large

expanses of the retina where there are no vessels. While very fast, the results of (16) are usually less complete and accurate than competing methods mentioned.

Retinal Image Registration

Retinal image registration takes in a set of images of the same retina and registers them by calculating the relative motion. Sparse keypoint descriptors like SIFT (20) matched between images are a popular approach in panorama stitching and have been extended to creating retinal mosaics (21, 22). Some method use custom descriptors or match uniquely identifiable vessel locations like bifurcations (23–25). Other methods match the shape of the entire vasculature tree between images (26, 27). More recently, hybrid methods combine sparse keypoints with tracking on the pixel level to create a mosaic from a video in real-time (28). Adapting many of these algorithms for the challenges of operating during intraocular surgeries in real-time would be difficult. Local keypoint methods (21, 22) struggle to find uniquely identifiable points at high magnification because of the low texture of the retina and contrast in surgical imaging. Likewise (23, 25) run on high-resolution fundus images and find few, if any, uniquely identifiable vessel locations at the high magnifications required for intraocular procedures. Furthermore, most approaches are not designed to cope with occlusion from instruments, harsh illumination, or variable lensing. Most importantly, these algorithms only perform localization and do not build a map of the vasculature, which is critical for robotic aid that depends on using the vasculature structure to improve micromanipulation.

Comparing to up-to-date technology, Richa et al. (28) performs registration of high-definition imagery to provide a full mosaic view of the retina along with on-the-fly registration of a pre-operatively selected location from a fundus image. It is very fast (15 ms per frame) and robust to full frame occlusion. While similar to the proposed EyeSLAM algorithm, there are important differences. First, the output of Richa et al. is a registered image built from microscope frames stitched together over time with emphasis on photorealism designed to aid tutoring eye exams; our output is a full, probabilistic map of all the vessels seen from each frame designed to enhance robotic aid during intra-ocular procedures. Second, there is the difference of evaluation being limited on the single application of slit-lamp videos from eye exams (28), EyeSLAM works across a diverse set of videos at various magnifications collected in our lab and gathered online, spanning several different procedures: photocoagulation, membrane peeling, and cannulation. Finally, Richa et al. do not address issues critical to providing robotic aid during intraocular procedures: they do not model occlusion by the surgeon's instrument or provide any vasculature information. Our approach is robust to transient tool occlusion and EyeSLAM paired with a robotic micromanipulator can actively aid the operator avoid or target the retinal vessels in experiments of several different simulated retinal surgeries.

Our approach incorporates an adaptation of the real-time correlative scan-matching method proposed by Olson et al. (29), using vasculature trees to build maps and register motion. We do not include (29) in Table 1 because it is not originally designed for retinal application, but instead more traditional SLAM applications. While EyeSLAM may not have as high accuracy as some algorithms listed in Table 1, it is unique because it operates in real time

and provides both vasculature maps and retinal registration, making it suitable for tight control loops in robotic surgical assistance.

Simultaneous Localization and Mapping (SLAM)

A related problem in robotics is that of simultaneous localization and mapping (SLAM) where a robot with noisy sensors traverses the world and wants to both incrementally build a global map of everything it has seen and determine its own location in that map (30). SLAM attempts to take the joint probability over all the observations and optimize both the global map and robot location at the same time. Original attempts performed poorly at scale and could not disambiguate between landmarks with similar features. Modern particle filter methods improved both speed and robustness, yielding methods like FastSLAM, which can represent maps as 2D images of dense probabilities of each point in space being occupied (31).

The problem of creating a global map of all the retinal vessels and localizing the current vessels seen in the image to the map is similar to that of SLAM. The most significant differences are that most SLAM methods assume space-carving sensors like laser range-finders instead of overhead cameras and depend on a reasonably good model of the robot motion, which is lacking in retinal localization. As a result, most SLAM algorithms are not immediately suitable for solving the problem of mapping and localizing retinal vessels. However, EyeSLAM uses and extends the core ideas of both SLAM and other retinal algorithms to achieve mapping of vasculature and real-time localization, all operating in a surgical environment to provide manipulation aid via a robotic platform.

Materials and Methods

Our goal was to develop an approach that maps and localizes vasculature in the eye by fusing temporal observations from retinal vessel detection approaches in a probabilistic framework similar to existing SLAM algorithms. Currently, there is not a good solution to this problem that works with the constraints of intraocular surgery. Out of the box SLAM algorithms do not work with the harsh intraocular environment. Previous work in registration fuses temporal information well to build a mosaic of the entire retina but do not focus on extracting the vasculature (26, 27) while fast vessel detection algorithms are low quality and do not cope well with occlusions or temporal viewpoint changes (16). An important constraint is the algorithm must run real-time to be suitable for integration into a robotic aid system. We introduce a new approach named EyeSLAM, an extension of a previous approach (13) that merges concepts introduced by (16, 29–31) to rapidly detect vessels, build a probabilistic map over time, and localize using scan matching in an intraocular environment. It is robust to harsh lighting conditions and transient occlusions.

2D vs. 3D Models of the Retina

When considering a model of a retina, a 3D sphere seems the most suitable representation for building a map of the inside of the eye. However, a full 3D representation is problematic because 3D estimation in the eye is challenging. Microscope calibration can be difficult (32) and modeling the lens of the eye to achieve intraocular localization is an area of active

research (33), especially in conjunction with the nonlinear vitrectomy lenses often used during intraocular surgery. Further complicating matters is the deformation caused by the tool inserted through the sclera (or white of the eye). In practice, most approaches to retinal mapping choose a simpler 2D representation, assuming a roughly planar structure for the retina with an X translation, Y translation, and an in-plane rotation. Scaling can be added, but often has local minima that cause poor tracking (28), so we use the 3DOF representation without scaling. We have found the planar assumption has sufficient power to compactly represent the retina, which is especially true at high magnification where only a small subset of the retina is seen and can be treated as a plane. However, to compensate for small shifting vessel locations caused by 3D rotation, we do add a dynamic aspect to the map.

Problem Definition

The algorithm takes in a series of images of the retina from a temporally consistent video with no underlying assumptions about camera or eye motion except that it is roughly planar and there is no change in magnification. At each time t, EyeSLAM takes an input image I_t and produces two outputs: (1) a dynamically expanding 2D grayscale image representing the global occupancy map O_t of all the vasculature seen so far, with each pixel encoding the probability of a vessel at that location and (2) camera location viewpoint L_t into that map which is the result of the registration of the current image to the global map that changes as either the camera or eyeball moves. A 3-DOF (x translation, y translation, planar theta rotation) represents the possible camera registration to the map and our experiments show it approximates eyeball motion well, even for low magnification.

Feature Extraction via Vessel Detection

Sparse keypoints or vessel landmarks (crossovers or bifurcations) are difficult to track during intraocular surgery because of a lack of texture at high magnification and harsh illumination. We instead focus on semi-sparse points extracted along vessels with the idea that instead of trying to match points based on local texture, we can match the overall visible structure of vessels frame to frame. EyeSLAM uses the low quality but very fast algorithm of Can et al. (16) to extract approximate vessel locations that enable frame to frame matching (see Fig. 1b). To reduce false positives and improve quality, points detected as vessels are first filtered. Locations that are too dark, too bright, insufficiently red, or in the microscope fringing region are rejected, which improves performance in the presence of glare, low contrast, or distortion. These detections form the set of 2D points we want to match, denoted as the current observation Z_t. In addition to better filtering, we have improved (13) by allowing the orientation of the vessel to change more quickly and then smoothing the vessels after tracing to provide higher quality vessel detections. False positive vessels are occasionally produced but do not usually gain enough evidence in the map to appear as high probability vessels.

Mapping via Occupancy Grids

EyeSLAM incorporates all observations over time by calculating the probabilistic occupancy map O_t as a 2D grayscale image where each location is represented by the log probability of a vessel occupying the pixel (see Fig. 1(c)). With each new image and associated detected vessel observation points Z_t, the map is updated using the estimated

camera location viewpoint L_t (described in the next section). Each newly detected vessel point is transformed into the map with L_t^{-1} and the log probability at that map location is increased; all other visible map locations that did not have any detected vessels have their associated log probabilities decreased. This allows vessels that have not been seen for a while to be removed from the map, which allows the algorithm to more gracefully handle false positives, out of plane rotation, imaging distortions, or tissue deformations. In practice, the viewport into the map is calculated (depicted by the yellow oriented rectangle in Figure 1(c)), and all the log probabilities in that area is first decayed by a fixed amount U_{decay}. Vessel points detected in the current viewpoint are added to the map as a 3×3 Gaussian filter with an initial value of $U_{initial}$ to represent the uncertainty of each detected vessel point. Log probabilities in the occupancy map are capped at U_{max} to keep a cell from becoming too certain and being unable to respond to changes. See Fig. 3 for an example of the probabilistic map of vessels at two resolutions. Tuning these parameters allow us to better model uncertainty of the vessel detection quality on real-world observations; our implementation uses $U_{decay} = 0.01$, $U_{initial} = 0.25$, and $U_{max} = 5.0$ as those values tended to work well across different types of intraocular environments.

Unlike (13), the EyeSLAM algorithm uses a dynamically sized map that can automatically expand to accommodate newly detected vessels that were located outside the previous map boundaries as described on Fig. 2 on the "Dynamic Map Expansion" and "Map Updating" steps. The map update calculates the camera viewpoint on the map to not decay the probability of the grid cells located outside the currently observable field of view. By not decaying unseen areas of the map, we assume the vasculature is nominally stationary, and the map of the vessels does not change when out of sight. The formulation of the occupancy map reasons about uncertainty over time, smoothing noise while handling transient occlusions and deformation. Our previous work required finding the centerlines of the vessels as a necessary part of the registration step, which required an expensive image skeletonization and subsequent Iterative Closest Point (ICP) process. The new registration method avoids this time-consuming procedure and operates directly on the occupancy grid map. The final map is now generated using all the vessel points with a high probability value in the occupancy grid map. It is possible to calculate the centerlines if necessary for robotic control, but it is no longer integral to the internal workings of the algorithm, which is faster than (13).

Localization via Scan Matching

To localize eyeball motion (which is mathematically equivalent to localizing camera motion), a 3-DOF planar motion model is chosen. The problem of localization is then to estimate the 2D translation and rotation of the camera C_t between the current observations Z_t and the occupancy grid map O_t. The original formulation (13) used Iterative Closest Point (ICP) algorithm for registration between a skeletonized version of the occupancy map and the current vessel observations, similar to (34). While this worked well for smooth motions, it was slow and had a tendency to fail with large, jerky motions, which caused divergence and would reset the tracker.

The biggest improvement compared to (13) is the replacing of ICP with the fast correlative scan-matching method proposed by Olson et al. (29). It is used to scan the whole 3D search window of solutions W_l, parametrized by x_l, y_l, and θ_l, and to find the best match B_t between the map M_t and the current observations P_t transformed to the map with W_l^{-1}. The solution associated with the best match is considered as the best solution for camera registration L_t. The challenge is to minimize the processing time while maximizing the quality and the robustness of the solution.

A brute-force method that scans every solution is too slow. As detailed in (29), we adopt a multi-resolution approach consisting in scanning the 3-DOF search window (two translations and a rotation) with two different map resolutions (Fig. 3). A first scan quickly identifies the approximate best solution in the low-resolution map. Afterwards, a second scan on the high-resolution map initialized around the low-resolution transformation more precisely finds the best approximation of L_t. At $1/4^{th}$ map size, we see about a 16X speed up.

Incomplete vessel detections at each frame can be noisy, so the final scan-matching estimation of the localization is smoothed using a constant-velocity Kalman filter, yielding the localization of the camera L_t. At most 500 vessel points are selected for scan matching (at random) to improve runtime. If too few observations are found, they are discarded and the current localization is kept. Once scan matching completes, the occupancy map is then updated with the newly registered vessel points Z_t to close the feedback loop on the algorithm.

Video Sequences

For ease of robotic testing in our lab, color video recorded with a surgical microscope is captured at 30 Hz with a resolution of 800x608 at a variety of high magnifications (10–25X). We also tested with human retinas *in vivo* from videos taken of real human eye surgeries (available publicly via YouTube). Those videos have variable resolution at different magnification, which are specified if relevant in the results. Fig. 4 shows the proposed algorithm output on a human retina during surgery *in vivo*.

Results

We have evaluated EyeSLAM both quantitatively and qualitatively on a variety of videos of paper slides, porcine retina *ex vivo*, and human procedures *in vivo*.

Retina Localization Results

Fig. 4 shows translation and rotation transformation with error between EyeSLAM estimates and human-labeled ground truth for a video sequence of human retina during an epiretinal peeling procedure *in vivo*. Quantitatively in Table 2, EyeSLAM with scan matching outperforms the previous ICP method for experiments on three video sequences. Error is measured relative to a transformation calculated from tracked fiducials for the paper slide or human labeled fiducials as ground truth. As seen in Fig. 4, significant translations in excess of 100 pixels are evident on these video sequences. The global nature of the low-resolution

scan-matching step helps prevent jumps due to poor vessel registration. In the planar phantom experiment, where the vessels are well defined, the lighting uniform, and the motion smooth, the two results are close, although EyeSLAM localizes 50% better. On the porcine and human retina, earlier versions of the algorithm have very large Root Mean Squared (RMS) errors because of lost tracking and jumps that were not recoverable with the older ICP registration. The new EyeSLAM algorithm provides superior localization, which is important during critical microsurgical operations.

Vasculature Mapping Results

Figure 5 shows a visual, qualitative evaluation of the quality of maps EyeSLAM builds in a variety of simulated and real retinal surgical applications. While there are some incomplete detections and false positives in the map, EyeSLAM is able to largely map and localize in a diverse and challenging set of lighting conditions and environments.

To measure map quality quantitatively, Fig. 6 evaluates vessel coverage and false positives on a typical human video sequence *in vivo*. Calculations are made using hand-labeled vasculature (including very thin vessels) in nine frames in the first five seconds. Initialization is fast, requiring less than 15 frames (0.5 sec) to start building the map and only a few seconds to fully build the map, depending on the image quality and the ability of the vessel detection algorithm. On this sequence, EyeSLAM maps 50% of the hand-labeled vessels within 30 frames (one second) and achieves 75% coverage within 60 frames (two seconds). It does not achieve 100% coverage because it misses some very thin, faint vessels, and the ones located very close together. The false positive rate is under 10% after five seconds. Overall, EyeSLAM converges quickly with good coverage of the well-defined vasculature structures of the retina.

Timing Performance

For speed, images are resized in half, yielding resolutions in the range of 400×304 to 380×360, depending on video sequence source. On an Intel i5-3570K computer, EyeSLAM implemented in C++ runs at 50–100 Hz with a mean runtime of 15 ms on the three videos listed in Table 2. This time includes all vessel feature detection, correlative scan-matching localization, and dynamic occupancy-grid mapping running in a single thread. This is about 2X faster than the previous work (13), and is sufficient to run simultaneous EyeSLAM algorithms on stereo microscope views in real-time (>30 Hz).

Surgically-Applied Results with Micron

To test the usefulness of the EyeSLAM algorithm combined with robotic aid in intraocular surgical environments, we performed surgically-applied tests with the robot Micron developed in our lab (9). Micron is a 6-DOF handheld micromanipulator with motors in the handle to control the tip of the instrument semi-independently of hand motion within a small volume of area (several mm^3). The Micron setup has custom-built optical trackers on the instrument to provide 1 kHz 3D positioning information with μm-level precision and 30 Hz stereo cameras applying computer vision and stereo algorithms to the left and right microscope view. Calibration between the microscope stereo view and custom 3D optical trackers that measure the position of Micron is more fully described in (35). Micron's

control system uses EyeSLAM running on the stereo camera stream to identify where the vessels are and enforce virtual fixtures that constrain and scale motion of the tip about those vessels. These control system sets the goal position of the instrument tip based on a combination of the user hand movement and the virtual fixtures in effect (for instance, allowing the user to slide the tip along the vessel direction but not orthogonal to it). In the case of Micron outfitted with laser, EyeSLAM's tracking of the vasculature helps guide the laser beam finder (also tracked in the microscope stereo cameras) and controls the firing of the full-power laser so as to not accidentally burn the vessels.

In the first surgically-applied experiment using the Micron and EyeSLAM robotic aid system, we simulated the eye and retina in a rubber eyeball phantom filled with saline and fitted with a vitrectomy lens. A surgeon performed a vessel-tracing task under a board-approved protocol. Fig. 7(a) shows unaided human performance and Fig. 7(b) shows performance aided with the Micron robot enforcing virtual fixtures (35) derived from EyeSLAM. In (a) the tracing is imprecise because of normal physiological hand tremor at sub-millimeter scales, whereas in (b) the tracing is smoother because the Micron robot knows the vasculature map from EyeSLAM and can enforce virtual fixtures in the control loop to help keep the tip of the instrument on the vessel.

Finally, we tested efficacy of pairing EyeSLAM localization and mapping with robotic aid in a simulated retinal surgery. Research performed in our lab has used the EyeSLAM algorithm with the Micron manipulator during simulated photocoagulation surgeries performed with vitreoretinal surgeons on synthetic paper slides of retina (36). In laser photocoagulation, the goal is to accurately place burns on the retina while avoiding the vasculature. EyeSLAM provides the map of the vessels from which the system automatically plans the pattern of laser burns as well as the localization necessary to register that pattern to the retina as it moves during the procedure. The robotic aid aims the laser at the proper location on the retina, compensating for motion measured by EyeSLAM, and activates the laser. The surgeons using the Micron micromanipulator reduced burn placement error by over 50% while providing regular burn sizes with the robotic aid. Fig. 8 shows the automated avoidance of vessels using EyeSLAM during automated laser photocoagulation.

Discussion

We have presented a new algorithm, EyeSLAM, for retinal mapping and localization that operates in real time at >30 Hz. Designed to handle the dynamic environment of high magnification, variable illumination, and rapid motion inherent in retinal surgery, our approach converges quickly and is robust to occlusion. Fusing ideas from vessel detection, retinal registration, and SLAM, it has proven to be an effective method to temporally smooth vessel detections and build a comprehensive map of the vasculature. EyeSLAM localization is accurate within five pixels in translation and one degree in rotation on representative video sequences and initializes quickly, covering 80% of the vasculature within two seconds. Compared to (13), the new EyeSLAM approach addresses a lot of the earlier shortcomings, especially lag and loss of tracking. EyeSLAM has greatly improved localization, especially being robust to rapid motions common in retinal surgeries, along with additions to make the algorithm faster with more comprehensive vessel detections and an enhanced dynamic map

that expands as needed. EyeSLAM is least 50% better than our earlier formulation with significantly more consistent tracking in more difficult video sequences. EyeSLAM operates in real time in challenging intraocular environments, providing both mapping and localization of the retinal vasculature to the Micron robot, improving operator performance in synthetic vessel tracing and photocoagulation experiments.

Future improvements should include more robust vessel detection. Another focus is reducing false positives and better modeling occlusion with more sophisticated tool tracking such as (37). More advanced 3D models could be beneficial and accomplished with stereo vision. More sophisticated closing the loop SLAM algorithms could be studied and applied to reduce the map from drifting as the FOV changes. Finally, optimization to run in real-time on high-definition video may increase localization accuracy and map quality.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Holtz, F., Spaide, RF. Medical Retina. Krieglstein, GK., Weinreb, RN., editors. Berlin: Springer; 2007.

2. Singh, SPN., Riviere, CN. Physiological tremor amplitude during retinal microsurgery; Proc. IEEE Northeast Bioeng. Conf; 2002. p. 171-172.

3. Frank RN. Retinal laser photocoagulation: Benefits and risks. Vision Res. 1980; 20(12):1073–1081. [PubMed: 7196642]

4. Brooks HL. Macular hole surgery with and without internal limiting membrane peeling. Ophthalmology. 2000; 107(10):1939–1948. [PubMed: 11013203]

5. Weiss JN, Bynoe LA. Injection of tissue plasminogen activator into a branch retinal vein in eyes with central retinal vein occlusion. Ophthalmology. 2001; 108(12):2249–2257. [PubMed: 11733266]

6. Ueta T, Yamaguchi Y, Shirakawa Y, et al. Robot-assisted vitreoretinal surgery: Development of a prototype and feasibility studies in an animal model. Ophthalmology. 2009; 116(8):1538–1543. [PubMed: 19545902]

7. Uneri A, Balicki Ma, Handa J, Gehlbach P, Taylor RH, Iordachita I. New Steady-Hand Eye Robot with Micro-Force Sensing for Vitreoretinal Surgery. Proc IEEE RAS EMBS Int Conf Biomed Robot Biomechatron. 2010; 2010(26–29):814–819. [PubMed: 21461178]

8. Hubschman JP, Bourges JL, Choi W, et al. The Microhand: A new concept of micro-forceps for ocular robotic surgery. Eye. 2009; 24(2):364–367. [PubMed: 19300461]

9. Yang S, MacLachlan RA, Riviere CN. Manipulator design and operation for a six-degree-of-freedom handheld tremor-cancelling microsurgical instrument. IEEE/ASME Trans Mechatron. 2015; 20(2):761–772.

10. Bettini A, Marayong P, Lang S, Okamura AM, Hager GD. Vision-assisted control for manipulation using virtual fixtures. IEEE Trans Robot. 2004; 20(6):953–966.

11. Infeld DA, O'Shea JG. Diabetic retinopathy. Postgrad Med J. 1998; 74(869):129–133. [PubMed: 9640436]

12. Lin, HC., Mills, K., Kazanzides, P., et al. Portability and applicability of virtual fixtures across medical and manufacturing tasks. Proc. IEEE Int. Conf. Robot. Autom; Citeseer; 2006. p. 225-231.

13. Becker, BC., Riviere, CN. Real-time retinal vessel mapping and localization for intraocular surgery; IEEE International Conference on Robotics and Automation (ICRA2013); 2013. p. 5360-5365.

14. Chanwimaluang, T. An efficient blood vessel detection algorithm for retinal images using local entropy thresholding. International Symposium on Circuits and Systems; IEEE; 2003. p. 21-24.

15. Chaudhuri S, Chatterjee S, Katz N, Nelson M, Goldbaum M. Detection of blood vessels in retinal images using two-dimensional matched filters. IEEE Trans Med Imag. 1989; 8(3):263–269.

16. Can A, Shen H, Turner JN, Tanenbaum HL, Roysam B. Rapid automated tracing and feature extraction from retinal fundus images using direct exploratory algorithms. IEEE Trans Inform Technol Biomed. 1999; 3(2):125–138.

17. Soares JVB, Leandro JJG, Cesar RM, Jelinek HF, Cree MJ. Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification. IEEE Trans Med Imag. 2006; 25(9):1214–1222.

18. Bankhead, P., Scholfield, CN., McGeown, JG., Curtis, TM. Fast retinal vessel detection and measurement using wavelets and edge location refinement. In: Serrano-Gotarredona, T., editor. PLoS One. Vol. 7. 2012.

19. Koukounis D, Ttofis C, Papadopoulos A, Theocharides T. A high performance hardware architecture for portable, low-power retinal vessel segmentation. Integr VLSI J. 2014; 47(3):377–386.

20. Lowe DG. Distinctive image features from scale-invariant keypoints. Int J Comput Vis. 2004; 60(2):91–110.

21. Cattin, P., Bay, H., Van Gool, L., Székely, G. Retina mosaicing using local features. In: Larsen, R.Nielsen, M., Sporring, J., editors. Medical Image Computing and Computer-Assisted Intervention. Vol. 4191. Lecture Notes in Computer Science; Springer Berlin / Heidelberg: 2006. p. 185-192.

22. Wang Y, Shen J, Liao W, Zhou L. Automatic fundus images mosaic based on SIFT feature. J Image Graph. 2011; 6(4):533–537.

23. Becker DE, Can A, Turner JN, Tanenbaum HL, Roysam B. Image processing algorithms for retinal montage synthesis, mapping, and real-time location determination. IEEE Trans Biomed Eng. 1998; 45(1):105–118. [PubMed: 9444845]

24. Can A, Stewart CV, Roysam B, Tanenbaum HL. A feature-based, robust, hierarchical algorithm for registering pairs of images of the curved human retina. IEEE Trans Pattern Anal Mach Intell. 2002; 24(3):347–364.

25. Broehan AM, Rudolph T, Amstutz CA, Kowal JH. Real-time multimodal retinal image registration for a computer-assisted laser photocoagulation system. IEEE Trans Biomed Eng. 2011; 58(10): 2816–2824. [PubMed: 21689999]

26. Stewart CV, Tsai C-L, Roysam B. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. IEEE Trans Med Imag. 2003; 22(11):1379–1394.

27. Chanwimaluang T, Fan G, Fransen SR. Hybrid retinal image registration. IEEE Trans Inform Technol Biomed. 2006; 10(1):129–142.

28. Richa R, Linhares R, Comunello E, et al. Fundus Image Mosaicking for Information Augmentation in Computer-Assisted Slit-Lamp Imaging. Med Imaging, IEEE Trans. 2014; 33(6):1304–1312.

29. Olson, EB. Real-Time Correlative Scan Matching; Proc. IEEE Int. Conf. Robot. Autom; 2009. p. 4387-4393.

30. Thrun, S. Simultaneous Localization and Mapping. In: Jefferies, M., Yeap, W-K., editors. Robotics and Cognitive Approaches to Spatial Mapping. Vol. 38. Springer Tracts in Advanced Robotics; Springer Berlin / Heidelberg: 2008. p. 13-41.

31. Hahnel D, Burgard W, Fox D, Thrun S. An efficient FastSLAM algorithm for generating maps of large-scale cyclic environments from raw laser range measurements. Proc. IEEE Intl. Conf. Intell. Robot. Syst. 2003; 1:206–211.

32. Zhou Y, Nelson BJ. Calibration of a parametric model of an optical microscope. Opt Eng. 1999; 38(12):1989–1995.

33. Bergeles C, Shamaei K, Abbott JJ, Nelson BJ. Single-camera focus-based localization of intraocular devices. IEEE Trans Biomed Eng. 2010; 57(8):2064–2074. [PubMed: 20442042]

34. Stewart CV, Tsai C-L, Roysam B. The dual-bootstrap iterative closest point algorithm with application to retinal image registration. IEEE Trans Med Imaging. 2003; 22(11):1379–1394. [PubMed: 14606672]

35. Becker BC, MacLachlan RA, Lobes LA Jr, Hager G, Riviere C. Vision-based control of a handheld surgical micromanipulator with virtual fixtures. IEEE Trans Robot. 2013; 29(3):674–683. [PubMed: 24639624]

36. Yang S, Lobes LA Jr, Martel JN, Riviere CN. Handheld Automated Microsurgical Instrumentation for Intraocular Laser Surgery. Lasers Surg Med. 2015; 47(8):658–668. [PubMed: 26287813]

37. Sznitman R, Richa R, Taylor RH, Jedynak B, Hager GD. Unified detection and tracking of instruments during retinal microsurgery. Pattern Anal Mach Intell IEEE Trans. 2013; 35(5):1263–1273.

38. Chen L, Huang X, Tian J. Retinal image registration using topological vascular tree segmentation and bifurcation structures. Biomed Signal Process Control. 2015; 16:22–31.
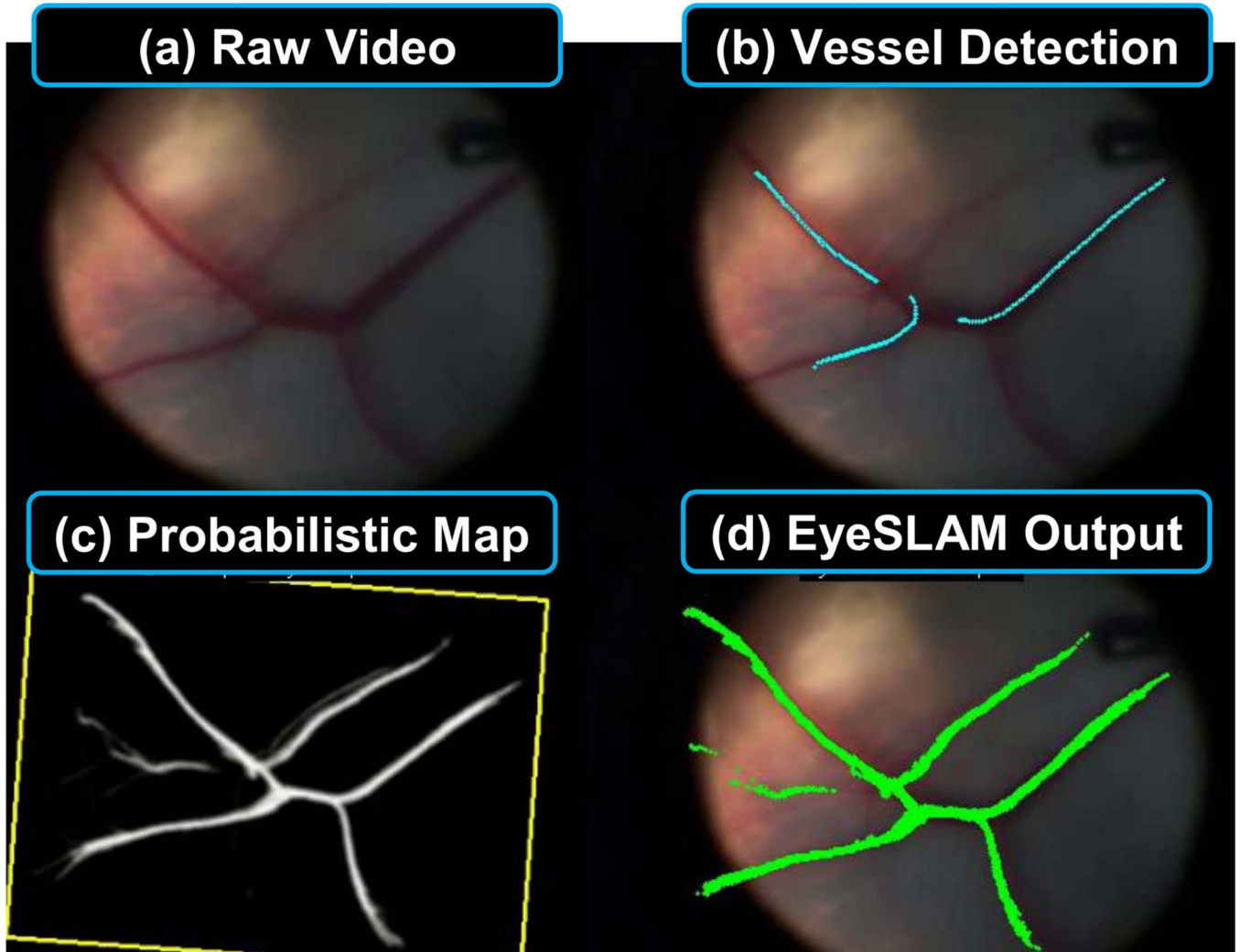
Fig. 1.
Summary of the EyeSLAM algorithm running on *ex vivo* porcine retina under surgical microscope. (a) Raw input video of retina during surgery via high-magnification microscope is run through (b) fast vessel detection, forming noisy observations which is used to build and localize (c) a 2D occupancy map representing probabilities of vasculature at each point (with yellow box showing the localized camera view), yielding (d) EyeSLAM output of the full vessel map localized to the current frame.
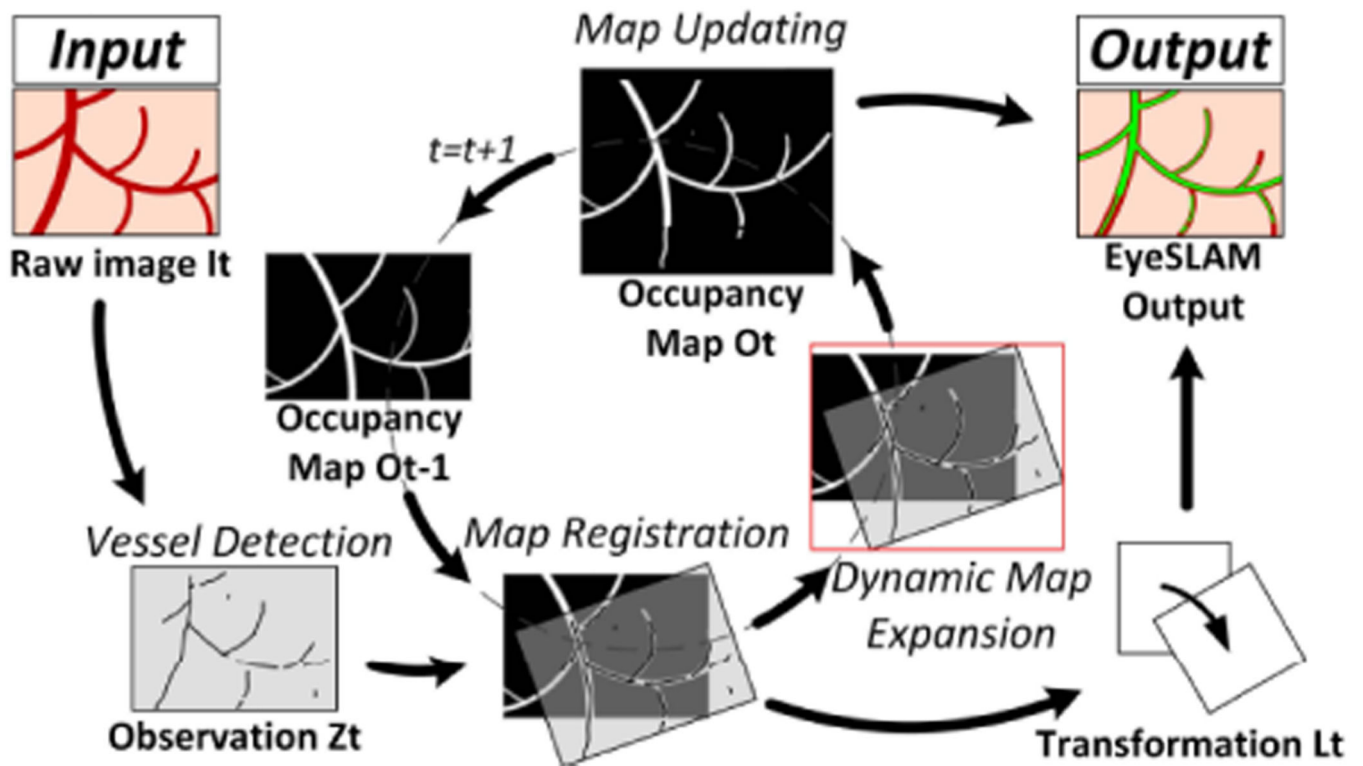
Fig. 2.
Block diagram showing the steps of EyeSLAM algorithm that maps and localizes retinal vessels during intraocular surgery. Vessels are detected and registered to the map, building an occupancy grid. The map is initialized with the raw vessel detections on the first frame and resized dynamically as needed. The registration allows the map to be transformed back into the current frame, providing full vasculature map along with localization.
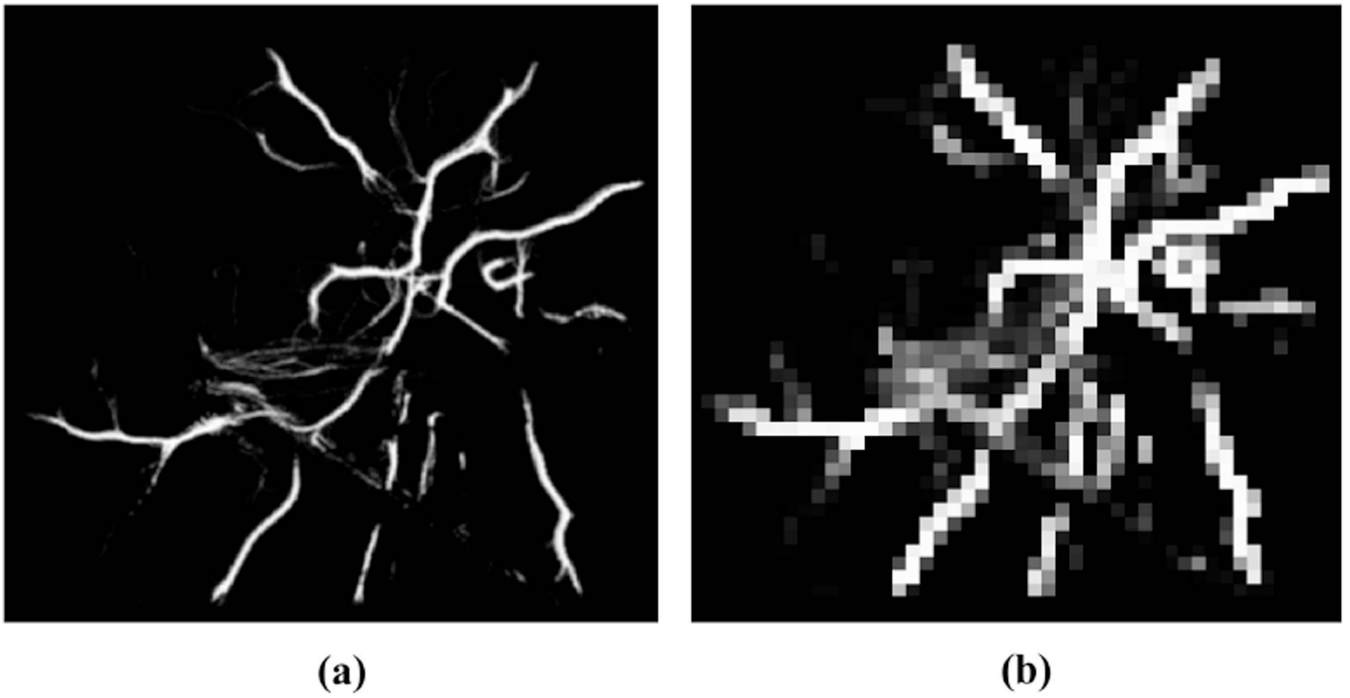
(a) **(b)**

Fig. 3.
Multi-resolution maps used by the scan-matching registration algorithm. (a) High resolution map. (b) Low resolution map.
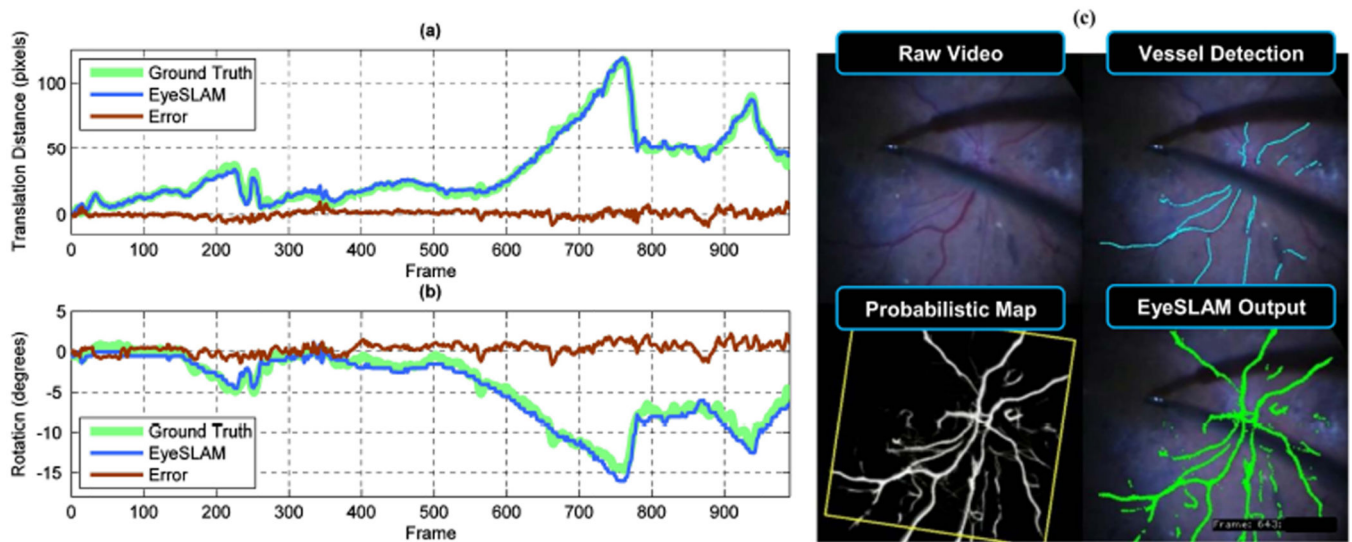
Fig. 4.

Localization accuracy and error compared to labeled video of human retina *in vivo* of a patient with retinopathy. (a) Translation component representing the L2 norm of X and Y (b) Rotation component (c) Sample output of raw video, vessel detection, map, and localized EyeSLAM output from frame 643. This system was able to perform vessel detection and localization in diseased eyes with retinal hemorrhages. Notice the accumulated evidence of seeing a vessel over many frames adds persistence of the mapped vessel structure through transient false negatives (no observed vessels because of tool occlusion). One failure mode is persistent occlusions will yield a repeated lack of vessel observations and remove vessels from the map once the probability decays too low. Loop closure global optimization would help with the small amount of angular drift seen in the latter part of the sequence.
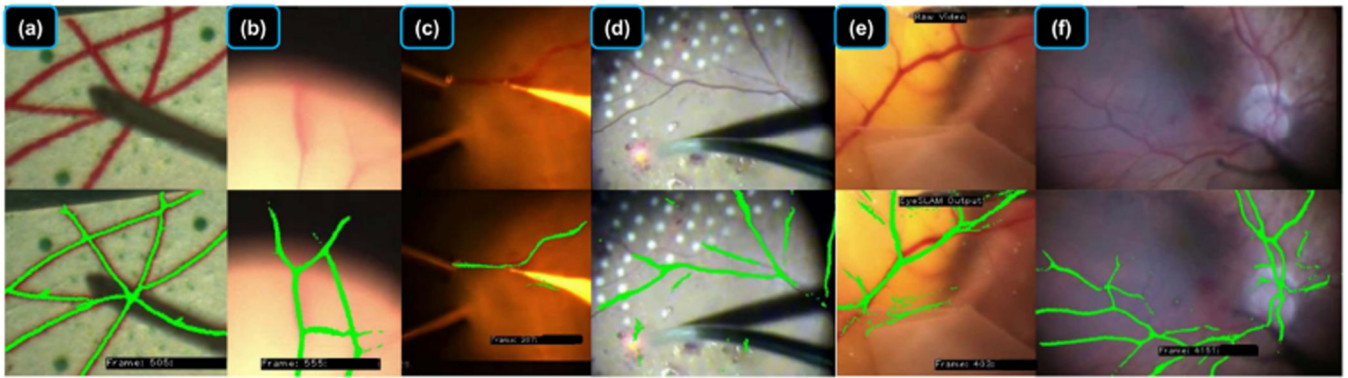
Fig. 5.
Visual evaluation of EyeSLAM operating in diverse environments. (a) Printed retinal image
on paper: note that EyeSLAM can maintain vasculature structure even during tool occlusion.
(b) Porcine retina *ex vivo* at high magnification during rapid movement where some of the
vessels have moved outside of the view of the microscope; note that EyeSLAM is able to
remember where the vessels are, even outside of the microscope FOV. (c) Porcine retina *ex
vivo* during retinal vessel cannulation experiment; notice very good performance in a
challenging light environment despite a few false positives on the edge of the tool where the
color filter is failing. (d) Human retina *in vivo* during panretinal photocoagulation surgery;
there are a few false positives on the red laser dot where the color filter is failing. (e) Chick
chorioallantoic membrane (CAM) *in vivo*, a model for retinal vessels that is considerably
more elastic than the retina; EyeSLAM does not easily correct for deformations induced by
the tool, and has some false positive responses on the edge of the membrane. (f) Human
retina *in vivo* during epiretinal membrane peeling; note the good mapping and localization.
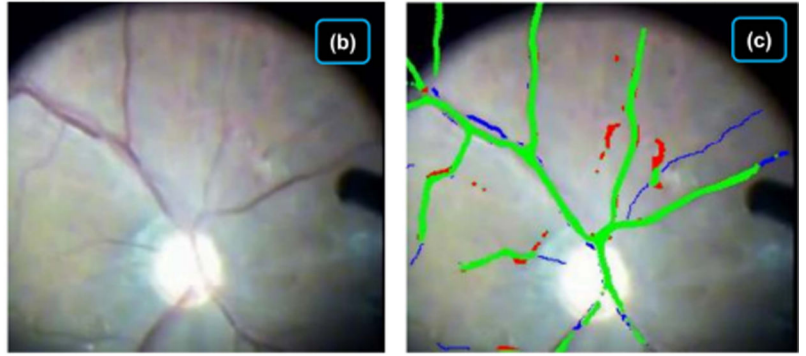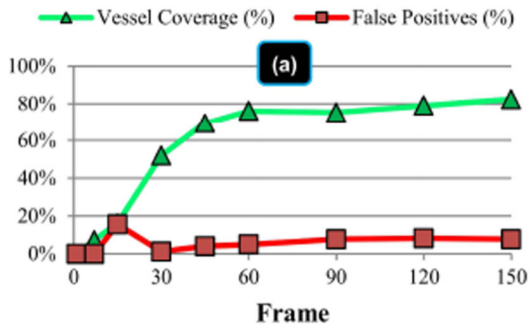
Fig. 6.
Vasculature map initialization statistics including coverage and false positives for the first five seconds of a video sequence of human retina *in vivo* being prepared to receive photocoagulation treatment. (a) Vessel coverage and false positives of the thresholded probabilistic map evaluated at frame 0, 7, 15, 30, 60, 90, 120, and 150 compared to hand-labeled vasculature at each frame. Coverage is the percentage of hand-labeled vasculature that EyeSLAM correctly built. False positives are the places in the EyeSLAM maps that do not correspond to true vessels. The spike in false positives at frame 15 is due to slight mis-registration during initialization. (b) Raw video of frame 150 after five seconds. (c) Color-coded map at frame 150 with green representing true positives (correctly matched vessels), red representing false positives (spurious vessels), and blue representing false negatives (missed vessels). Best viewed in color.
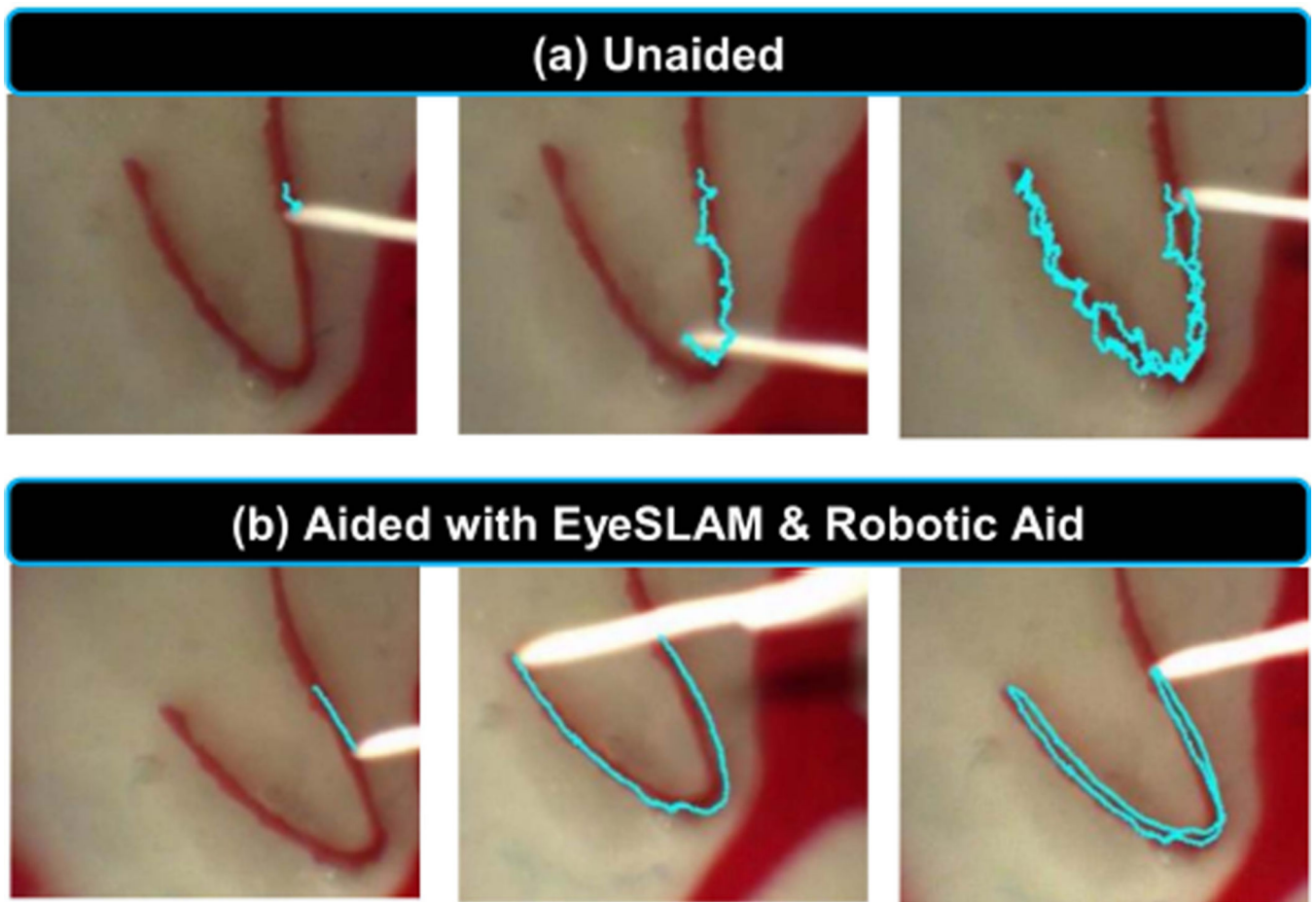
Fig. 7.
Tracing a retinal vessel in an eyeball phantom. (a) Unaided attempt to trace the vessel and (b) Aided attempt with a robotic micromanipulator enforcing virtual fixtures based off EyeSLAM mapping and localization. The blue line indicates the path of the tip of the instrument registered to each frame of the video using EyeSLAM. Note in both cases, the entire phantom eyeball is moving due to movement of the tool through the sclera port.
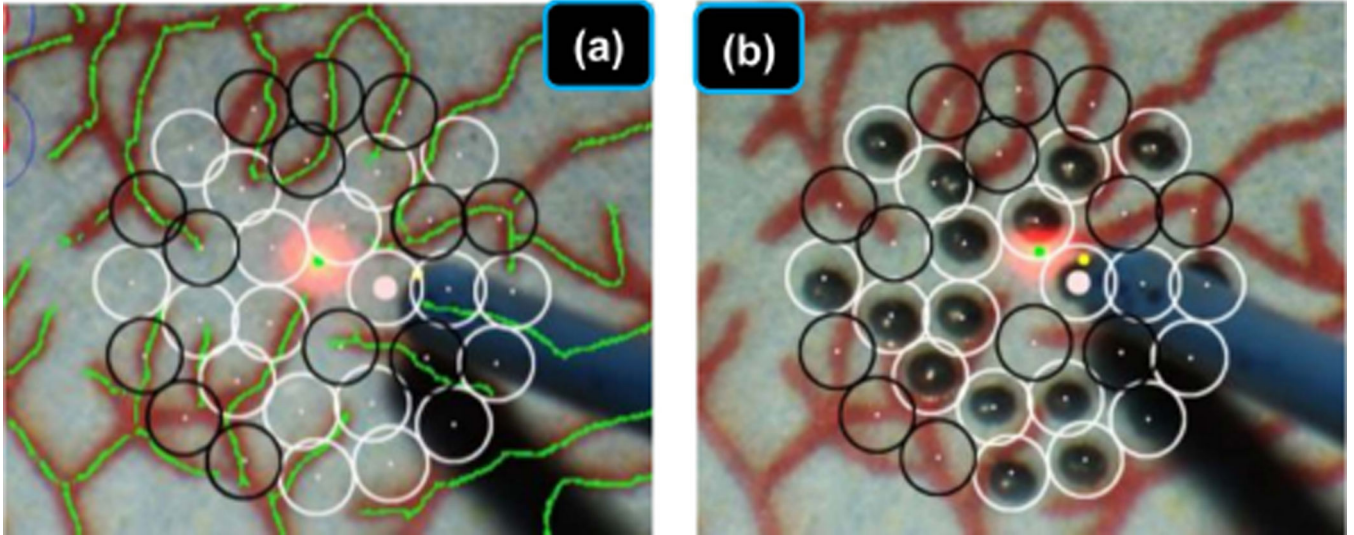
Fig. 8.
EyeSLAM can be used in the robotic micromanipulator Micron control system to provide accurate targeting information on a paper phantom during a simulated retinal photocoagulation experiment. EyeSLAM provides the necessary localization information to compensate for motion of the retina in real time and register the burn pattern to the pre-operatively specified placement. The EyeSLAM map is used to automatically plan burn patterns that do not overlap or touch the vasculature structure, thus protecting the vessels while applying the needed treatment to the retina. (a) Target placement before laser photocoagulation. Empty black circles represent targets to be avoided within 100 μm of a vessel while white circles show valid targets (b) After the completion of the automated laser photocoagulation.

Table 1

Comparison of Various Vessel Detection, Registration, and Mapping Algorithms for the Retina

A sample of representative algorithms for detection, registration, and mapping of the retina and retinal vessels sorted by publication date. Feature abbreviations are VD: Vessel Detection, RR: Retina Registration, RM: Retinal Mosaicking, and VM: Vasculature Mapping. Core capabilities and underlying methods used are noted. Vessel detection is the ability to return where retinal vessels are from a single image of the retina. Registration is the ability to determine the transformation between two images of the retina. Mosaicking is the ability to build up a map of the observed retina, while vasculature mapping is the ability to build up a map of the seen retinal vessels. Times for detection algorithms are for a single image, whereas times for registration algorithms are often for a pair of images. No attempt has been made to normalize for computing advances, so estimates are upper bounds. EyeSLAM fills a niche for real-time performance while achieving retinal vessel detection, registration, and vasculature mapping simultaneously.

Year	Algorithm	Features VD	RR	RM	VM	Method	Image Size	Reported Time
1989	Chaudhuri et al. (15)	X				Matched Filters	*	50 s
1998	D. E. Becker et al. (23)		X	X		Vasculature Landmarks	640×480	0.9 s
1999	Can et al. (16)	X				Tracing Templates	1024×1024	30 ms
2002	Can et al. (24)		X	X		Vasculature Landmarks	1024×1024	*
2002	Stewart et al. (26)	X	X			Vasculature Landmarks & Tree + ICP	1024×1024	5 s
2003	Chanwimaluang et al. (14)	X				Matched Filters + Local Entropy	605×700	3 min
2006	Cattin et al. (21)		X	X		Keypoint Features (SURF)	1128×1016	*
2006	Chanwimaluang et al. (27)	X	X			Vasculature Tree + ICP & Correlation	600×900	20 s
2006	Soares et al. (17)	X				Gabor Wavelets	768×584	3 min
2010	Wang et al. (22)		X	X		Keypoint Features (SIFT)	640×480	1 s
2011	Broehan et al. (25)	X	X			Vasculature Tree	720×576	40 ms
2012	Bankhead et al. (18)	X				Wavelets + Spline Fitting	564×584	0.6 sec
2013	B. C. Becker et al. (13)	X	X		X	Vasculature Tree + ICP	402×300	25 ms
2014	Koukounis et al. (19)	X				Matched Filters on FPGA	640×480	35 ms
2014	Richa et al. (28)		X	X		Keypoint Features (SIFT) + SSD	720×1280	15 ms
2015	Chen et al. (38)	X	X			Vasculature Landmarks	*	*
2017	EyeSLAM (our approach)	X	X		X	Vasculature Tree + Scan Matching	400×304	15 ms

Abbreviations: Iterative Closest Point (ICP), Speeded-Up Robust Features (SURF), Field Programmable Gate Array (FPGA), and Sum of Squared Differences (SSD).

* No data reported.

Table 2

RMS Error of EyeSLAM Localization

Root Mean Squared (RMS) error of 2D localization (translation + rotation) for three video sequences. The existing approach and improved EyeSLAM algorithm are compared with the percentage error reduced listed in parenthesis (these quantities represent improvement). The approach of Becker et al. (13) works well for gentle transitions present in the synthetic example, but for more challenging video sequences with jerky motion, it jumps and cannot recover, causing very high overall error. On these videos, the proposed EyeSLAM algorithm is much more robust, with an average RMS error of under 5 pixels in X and Y and under 1° in rotation. Furthermore, this represents a 43 − 98% improvement over the earlier algorithm of Becker et al. (13) in ideal conditions and over 20X more accurate localization in sequences with high motion, where EyeSLAM is more able to maintain consistent tracking.

Video Sequence	Frames	Becker et al. (13)		EyeSLAM (ours)	
		Distance (px)	Theta (deg)	Distance (px)	Theta (deg)
Paper Slide (Synthetic)	598	2.8	0.7	1.6 (43%)	0.4 (43%)
Porcine Retina (*ex vivo*)	795	204.2	43.8	7.6 (96%)	0.9 (98%)
Human Retina (*in vivo*)	997	115.1	31.6	4.8 (96%)	0.8 (97%)