Portable Dual Sensor Large Area Visualization System for Robotic Laser Surgery

Ravi Prakash*[1,3], Vincent Wang[1,3], Zacharias Chen[1], Leila Bridgeman[1,3], Patrick J. Codd[1,2,3]

***Abstract*— Laser scalpels are precise, dexterous, and efficient tools for soft tissue surgeries. However, surgical lasers are hard to control manually, require experience, and are often incompatible with conventional intraoperative imaging sensors. Integrating compatible sensing technology with laser scalpels for precise soft tissue surgery opens vital avenues for widespread adoption and previously unrealized automation. This paper proposes a dual-sensor strategy to generate high-resolution surgical scene visualization based on surgeon feedback for robotic laser surgery. The proposed method uses a coarse depth sensor to localize the tissue of interest in the surgical scene, and a fine optical coherence tomography (OCT) sensor to create a detailed ($< 30\mu m$ lateral resolution) tissue representation. The method achieves RMSE error in the range of 0.0878mm to 0.102mm in large-area tissue reconstruction and 0.050mm to 0.427mm in pattern-based laser ablation using user feedback over various fiducial samples. The findings demonstrate the proposed system's capability in large-area tissue imaging for precise laser-based surgery.**

I. Introduction

The need for precise, safe, and accurate minimally invasive surgery has naturally led to the development of precision scalpels and detailed surgical scene visualization tools. With over 40 million major surgeries taking place annually in the United States [1], and the ratio of patient per clinician increasing [2], there is a growing need to improve post-operative patient outcomes by decreasing recovery time and utilizing novel technologies to automate repetitive lower-level surgical tasks to lessen clinical burden.

The increasing adoption of robot-assisted minimally invasive surgery (RAMIS) has resulted in significant safety improvements, better patient care, and reduced recovery time through highly dexterous surgical scalpels with complementary imaging modalities. Energy-based scalpels, such as lasers, are well-suited for RAMIS owing to their versatility (fiber-based or free-form), precision (sub-millimeter), and tunability (resection vs. coagulation) [3, 4]. Currently, RAMIS is performed using conventional mechanical tools repackaged in novel form factors with complementary imaging sensors and teleoperation capability [5].

Despite rapid advancements in robotic surgery, existing laser surgery techniques remain limited in their broader deployment and application in RAMIS systems. Laser-based devices are often incompatible with conventional sensing modes and prone to hand tremors and inconsistencies during manual operation. Additionally, though laser scalpels can provide sub-millimeter resection cavities, they do not provide tactile feedback, and the quality of resection is difficult to accurately predict owing to the complex bio-physical laser-tissue interaction phenomenon that dictates the result. The efficacy of the laser scalpel can be determined by the fluence delivered onto the tissue surface. As laser fluence is dependent on laser beam distance and orientation, a high-resolution representation of the tissue surface is necessary to calculate the optimal laser pose and position [6–8]. We hypothesize that using high-resolution sensing technologies compatible with laser-tissue interaction, such as optical coherence tomography (OCT), will improve accuracy and precision when using intelligent laser-based robotic platforms.

In this work, a novel dual-sensor strategy is proposed for accurate 3-D surgical scene visualization for robotic laser surgery. Owing to imaging sensors' varied spatial resolutions and fields of view, our proposed method utilizes a coarse depth camera to visualize the surgical scene, followed by a fine 3-D visualization of the tissue of interest using an OCT system. This enables faithful representation of tissues with an uneven surface contour larger than OCT field-of-view (FOV) and at sub-micron ($< 30\mu m$) resolution, a key component of near-real-time feedback for closed-loop control in laser-based robotic tissue surgery. We demonstrate our proposed method in Figure 1, through:

Surgeon-based tissue visualization: The initial surgical scene is captured by the depth camera, allowing the surgeon to choose a region of interest (ROI) for finer visualization.

Large-area robotically assisted OCT scan for heterogeneous tissue surface: A robotically-mounted OCT device scans the selected tissue region while maintaining a consistent distance from the tissue surface, keeping the tissue within the focal distance across the scanning region.

Evaluation of OCT representation through laser ablation of user-defined patterns: Reproducibility of surgical resection patterns defined in the OCT visualization is tested using a fiber-coupled laser scalpel to evaluate the accuracy of distance estimates between large-area OCT volumes. The proposed method can be easily adapted to perform surgeon-in-loop tissue resection with feedback.

A. Large-area OCT

The invention and subsequent adoption of OCT has significantly impacted the field of ophthalmology and broader

[1]Thomas Lord Department of Mechanical Engineering and Materials Science, Duke University, NC, United States

[2]Department of Neurosurgery, School of Medicine, Duke University, NC, United States

[3]Supported by the National Science Foundation under Grant Number 2303158.

*Corresponding author: Ravi Prakash (rp247@duke.edu)

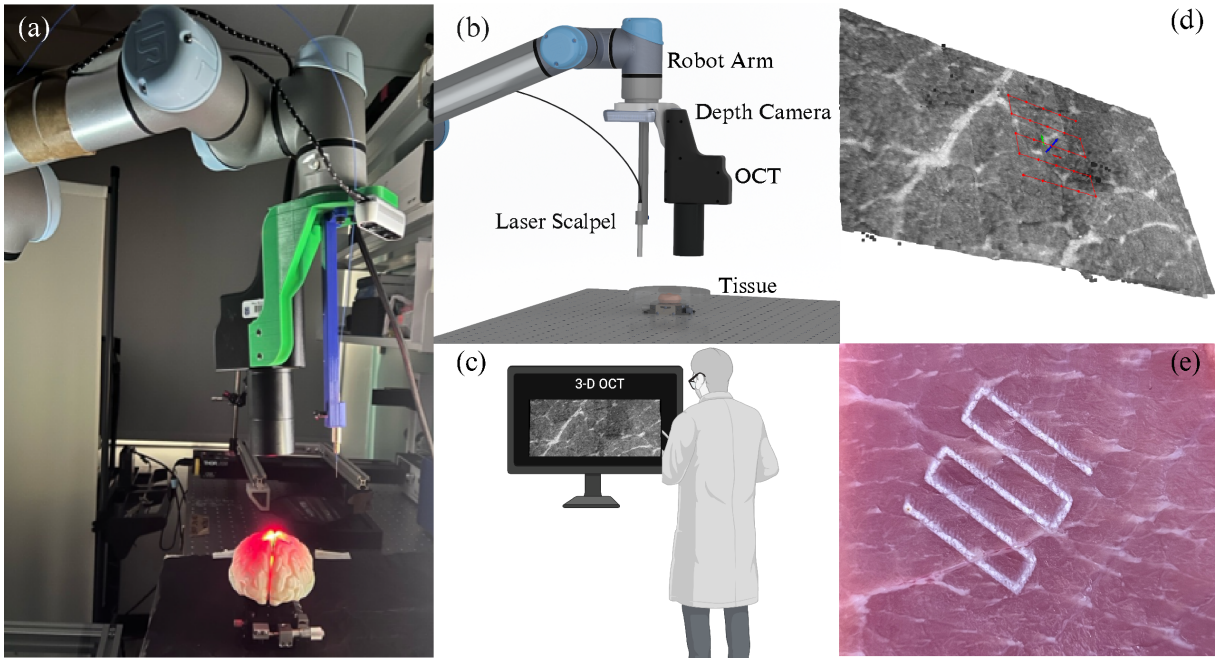

Fig. 1: **System overview.** (a) Overall system setup, (b) dual-sensor imaging system and laser scalpel, (c) interface for visualization and surgeon feedback [9], (d) surgeon-feedback based laser ablation pattern on porcine muscle, (e) laser ablation pattern on porcine muscle.

medical science [10–12]. OCT is a high-resolution laser-based imaging sensor working on the principle of interferometry. It constructs cross-sectional images (*B-scan*) based on reflected light from the tissue medium or echo time delay of backscattered light. However, OCT systems have a limited range ($<$14 x 14 mm^2) and suffer from image quality degradation outside of working distance. As most tissue surfaces are larger than the OCT FOV and have uneven contours, standalone OCT systems do not suffice for surgical needs. However, a combination of OCT with a dexterous robotic system is a possible solution to overcome the limitations of traditional OCT.

Several works in recent literature have been proposed to perform large-area robotic OCT scanning, with the central component being an OCT system mounted on a robotic arm [13–18]. The goal is to augment the OCT capabilities by accounting for tissue contour change by adjusting attitude and stitching multiple individual 3-D tissue representations (*C-scan*) in real-time or post-processing. However, these works do not involve surgical planning or tissue resection. In one work [13], authors used a depth camera on an optic table to guide a robot arm-mounted OCT system, with an additional RGB camera for fine alignment to the target. The tissue surface was divided into several voxels within a single OCT scan, and the final volume was manually registered and fused using recorded robot probe pose information. Another common method involves recording of continuous *B-scans* during robot motion with either a structured light camera for initial scene visualization and pose estimation [15], or OCT image feedback for real-time pose adjustment [18]. The resulting volumes are then registered using robot pose information and cross-correlation of imaging features [14], manually finding common points, or commercial software packages. Despite the ability to obtain continuous OCT scan data, the OCT systems used can be cost-prohibitive to buy or build in-house. In this work, we utilize a low-cost, off-the-shelf OCT system to drive our robotic laser surgery platform.

B. OCT and Laser-based Surgery

While different sensing modalities have been proposed to guide lasers for tissue resection [3, 4], OCT provides high spatial resolution, real-time imaging, intraoperative application, and subsurface visualization. Previous works have combined a 1940 nm Thulium Fiber (Tm) and a 2940 nm Erbium-doped yttrium-aluminum-garnet laser (Er: YAG) with a 1310 nm OCT system to scan and resect tissue [19, 20]. While their integration of various wavelength lasers in common optic paths is commendable, the components are fixed on an optic table and do not afford the dexterity needed to perform complex surgical maneuvers. Fan et al. [21], combined an OCT system with an off-axis laser fiber mounted on a 3-DoF linear stage for tissue resection. Li et al. [22] proposed a dexterous robotic system with OCT and fiber laser scalpel mounted on the end-effector of the robot arm. The system uses a preoperative MRI scan of the tissue surface with fiducial markers to perform an eye-to-hand depth camera calibration to align the preoperative scan with the current tissue position. Their work achieved scanning of tissue volume with height adjustment based on preoperative MRI data and performed tissue resection with a learning-based segmentation algorithm.

To develop a functional robotic laser surgery platform, real-time (or near real-time), intraoperative sensors are needed to visualize the surgical scene. This would admit accurate and precise tissue surface representations, enabling intraoperative adjustments using real-time control. Our proposed dual-sensor strategy is low-cost, near real-time, provides intraoperative sensing, and packages all the imaging

sensors and laser scalpel onto a robot arm, a critical step towards a multi-purpose robotic laser surgery platform.

II. Methods

A. System Overview

Our proposed hardware system consists of a robotic arm, a depth camera, a portable OCT system, and a fiber-coupled laser source for resection, as shown in Figure 1(a). We use a 6 DoF UR5e (Univeral Robots, Denmark) robot arm with an 850mm reach and payload capacity of 5 kg. An Intel Realsense 435i (Intel Corporation, United States) camera is used to capture depth information about the surgical scene. A low-cost commercial spectral domain OCT system with a center wavelength of 1310 nm and a bandwidth of 60 nm (OQ Stratascope 1.0, Lumedica Inc, United States) is used to generate a high-resolution visualization of the ROI. A Thulium (Tm) fiber laser scalpel, 1940 nm (IPG Photonics, United States), is used with a fiber diameter $400\mu m$ close to the tissue surface without any additional optics. All sensors are mounted and calibrated in the eye-in-hand configuration.

The robot arm initially obtains an RGB-D scan of the surgical scene from a predefined home position. Within the resulting point cloud, the user is prompted to select a ROI for higher-resolution visualization by selecting a center point and a scan radius. Scanning waypoints are generated by dividing the selected region into a grid. Using the RGB-D point cloud, the waypoints are height adjusted to keep the tissue surface within working distance of the OCT by calculating the average tissue surface elevation of a small voxel region around each waypoint. The voxel size is defined based on OCT FOV [22]. The robot sequentially reaches each waypoint and scans the ROI with the OCT scanner. Once the OCT volume is stitched, the user can define regions in the stitched volume for further laser resection.

B. Intraoperative Image Guidance

We aim to use real-time image guidance for our robotic surgery systems during surgery. We obtain both depth and color images of the surgical scene from the Realsense 435i camera. The depth information is incorporated with the 2-D color image of the scene and intrinsic calibration parameters to generate a 3-D color pointcloud, allowing the surgeon to easily choose a ROI. Points farther than 2 m are clipped, removing extraneous scan regions irrelevant to the scene. Compared to structured vision cameras and preoperative imaging (MRI), the RGB-D camera provides lower-depth resolution but is a fraction of the cost, size, and time.

C. OCT Data Acquisition

The OCT system comes with a software application (Lumedica OQ LabScope, Lumedica Inc, United States) to acquire *A-scans* and generate cross-sectional *B-scans*. During operation, we pass serial commands to the OCT to perform *C-Scans* and save the *B-scans* as individual 2-D images. Each *C-scan* has a voxel resolution of 512 x 512 x 128. Using the *C-scans*, the tissue surface boundary can be extracted. As

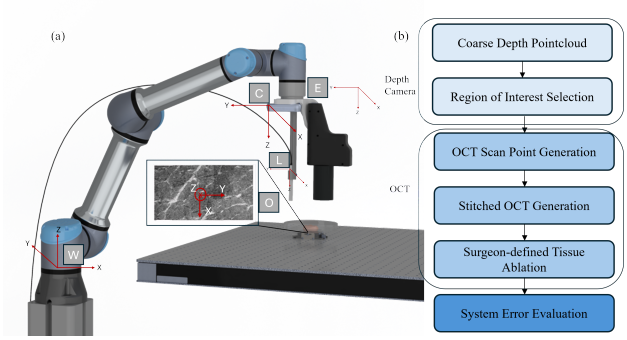

Fig. 2: **Experiment Flow.** (a) Coordinate frame of robot base, W, robot end-effector E, depth camera coordinate system C, OCT 3-D image coordinate system O, and fiber laser scalpel L (b) flow chart of the experiment process

OCT images are generated based on interferometry, the air-tissue boundary has a higher intensity than the neighboring pixels, allowing for inference of the boundary location. A bilateral filter is applied to smooth out the noise and preserve edges, followed by the extraction of the brightest pixel value from each image column. Furthermore, we define the relationship of pixels from an OCT volume to physical distances. Based on the measurement of fiducial objects of known lengths, we obtain an axial resolution of $14.59\mu m$, a lateral resolution of $28.00\mu m$, and a distance of $111.7\mu m$ between two adjacent B-scans. The scaled tissue volume is then represented by a point cloud in the OCT frame of reference. Each OCT volume is 14.336 x 14.298 mm^2 in X & Y and takes about 13 seconds to complete.

D. System Calibration

As shown in Figure 1(b), the system comprises two sensing elements and one actuating element: an Intel Realsense RGB-D camera, a Lumedica OCT system, and a Tm fiber laser scalpel. To obtain accurate point cloud positioning and laser targeting, each of these must be calibrated to the world frame by calculating the homogeneous transformation matrices T_C^E, T_O^E, and T_L^E (Figure 2 (a)). Once the transformation to the end effector (E) is known, transformation to the world frame (W) is found using the robot forward kinematics to obtain T_E^W. The camera calibration matrix T_C^E was obtained using an iterative eye-in-hand calibration method similar to the one used in [23]. The two transformation matrices T_O^W and T_L^W were both obtained using the hardware design specifications based on 3D models of the OCT system and laser fiber [22]. Additionally, specifications provided by Lumedica (e.g., focal distance) were used to position the focal frame of the OCT system, allowing for an accurate transformation to the robot world frame.

E. Large-area OCT Volume Stitching

To perform surgery, the OCT scanner must scan a large area of the tissue surface. Based on height-adjusted scan points generated using a depth camera, the robot moves the OCT scanner to each point to perform a volumetric scan. The distance between two scan points is kept at 7.5mm. However, this value can be optimized if time is a constraint.

The robot pose corresponding to the scan point location is saved for post-processing. The stitched large-area OCT volume is generated by transforming each OCT volume to the robot base frame using T_O^W. We account for non-telecentric distortions in the XY plane by truncating 1.5mm from each side of the OCT voxel in the X direction. As OCT *B-scans* have the least distortion in the center of the image, and the distortion increases towards the edges, truncating the edges helps to reduce lateral wrapping.

III. Experiments & Results

In this section, we perform experiments to validate the efficacy of using a dual-sensor imaging strategy for generating an accurate representation of tissue for precision surgery. First, we perform a qualitative experiment to generate large-area OCT stitched volume with minimal distortion. Second, we use known ground truth lengths to quantify errors in stitched OCT representations. Third, we demonstrate that our method can perform tissue resection tasks based on user-defined patterns in high-resolution OCT volumes.

A. Automatic Scanning of Large-area Object

We test our coarse (depth camera) to fine scanning (OCT) pipeline by placing the target object on a 3-axis stage. The depth camera takes the initial image of the scene and allows the user to select a point on the object of interest. A square of side length 50mm is defined around the chosen point, and height-adjusted scan points are generated. We scan a 2-D checkerboard to observe any distortion in the XY direction. Similarly, 3-D checkerboards of varied depths (1mm, 2mm) are also scanned validate depth reconstruction in Figure 3. When needed, we remove floating outliers for ease of visualization.

Finally, we also scan a sample of porcine muscle tissue to mimic a surgical scene (Figure 1). A 30 mm square scanning region was chosen based on the size of commonly occurring muscle tumor masses (soft tissue sarcoma).

B. Accuracy of Stitched OCT Representation

A quantitative evaluation of fiducial length representation in stitched OCT volume allows us to estimate the total system error. Obtaining correctly stitched high-resolution volumes involves several steps where errors can occur, thus distorting the final representation. Common possible error sources include: OCT pixel-to-physical coordinate scaling, lateral wrapping, robot forward kinematic estimation, and sensor extrinsic calibration. We use known length values from one 2-D and two 3-D checkerboards to measure errors in representing known distance measurements over multiple stitched OCT voxels. Each checkerboard was 3-D printed using a 5 x 5 mm^2 grid, with depths of 2mm and 1mm for the 3-D boards. As seen in Figure 4, distances are calculated over a stitched OCT region by using Segment Anything Model [24] to segment out square regions. By finding the centroid of each segmented region, a center-to-center distance can be found, which is also known from both caliper measurement and the 3-D model to be 5.0mm for the 3-D boards and 4.7mm for the 2-D board. Overall, we obtain an average error of 0.0878mm, 0.0844mm, and 0.102mm for the 3-D 2mm, 3-D 1mm, and 2-D checkerboard targets respectively, which are below the 0.5mm accuracy target commonly used in surgical applications.

C. Evaluation of OCT Representation Through Laser Ablation of User-defined Patterns

With the high-resolution representation of the object completed and system error quantified, we demonstrate our proposed method for downstream surgical tasks using a fiber-coupled Tm laser mounted to the robot end-effector.

To define a cutting region, we scan and stitch an OCT volume consisting of a dotted pattern with known spacing. We employ user feedback to select points on the stitched OCT volume and generate a trajectory connecting each point in the sequence. If the OCT volume representation is accurate, the distance between objects in the OCT representation should match the ground truth length between them. Various lengths were scanned to test representation accuracy across a varied number of voxel stitches, shown in Table II. We observe straight ablation paths for lengths longer than 6.85mm as seen in Figure 5(b), with one side of the square being of smaller size than programmed Figure 5(a). We select a raster pattern for ablating tissue, though other ablation patterns are possible. We demonstrate ablating a raster scan pattern on scanned porcine tissue based on user choice, as seen in Figure 5(d) in OCT representation and Figure 5(e) in 2-D color camera. We observe a positional error of 1.875mm for 1x1 case, 1.825mm for 1x3 case, and 2.15 mm for 3x5 case on average when measured from the center of the target circle to the nearest corner (Figure 5 (a) and (b)).

IV. Discussion

Our proposed work addresses the sensing challenges intimately associated with barriers to minimally invasive precision surgeries. We do so by proposing a dual-sensor strategy of obtaining a coarse representation of a noisy surgical scene and a fine representation of the surgical ROI. Our work employs real-time, depth camera, and near real-time OCT sensors to inform the surgeon and adapt to changes during surgery. Our choice of sensors complements the FOV and resolution limitations of each other and fit on the robot arm, to provide a low-cost, robotic platform with ease of

TABLE I: 3D Checkerboard Distance Metric

	3D-2mm Board	3D-1mm Board	2D Board
Ground Truth (mm)	**5.0**	**5.0**	**4.7**
Mean (mm)	4.959	4.963	4.705
Std. Dev. (mm)	0.0773	0.0758	0.102
RMSE (mm)	0.0878	0.0844	0.102

TABLE II: Tm Laser Cutting Distance using OCT Point Selection

Actual Distance (mm)	Avg. Cut Distance (mm)	RMSE (mm)	n
6.85	6.9	0.180	4
20.55	20.7	0.427	2
34.25	34.3	0.050	1

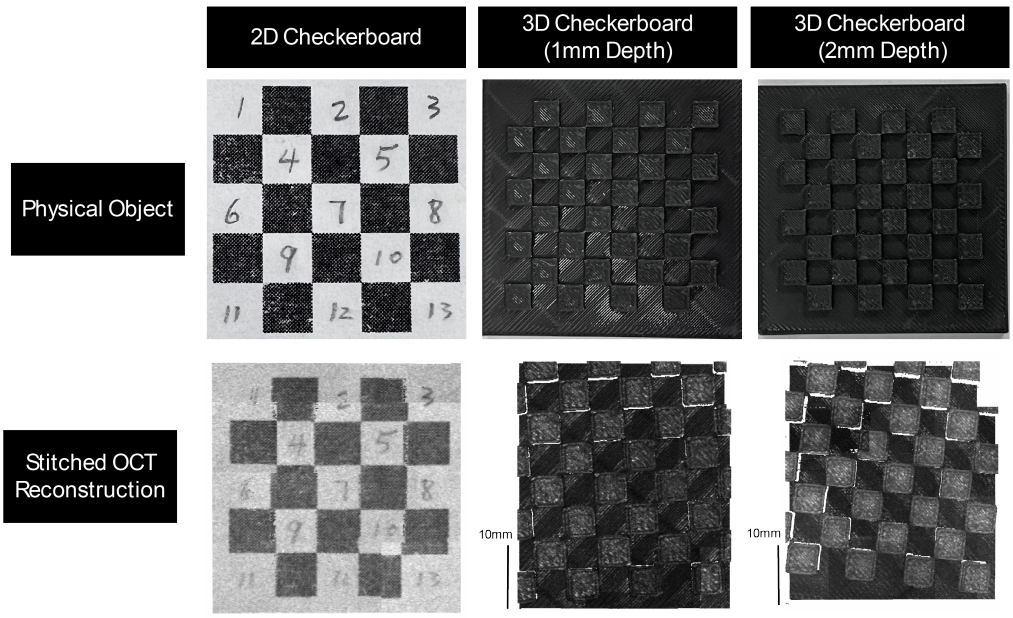

Fig. 3: **OCT Large-area reconstruction.** Examples of reconstructed physical objects using forward-kinematic-stitched OCT volumes.

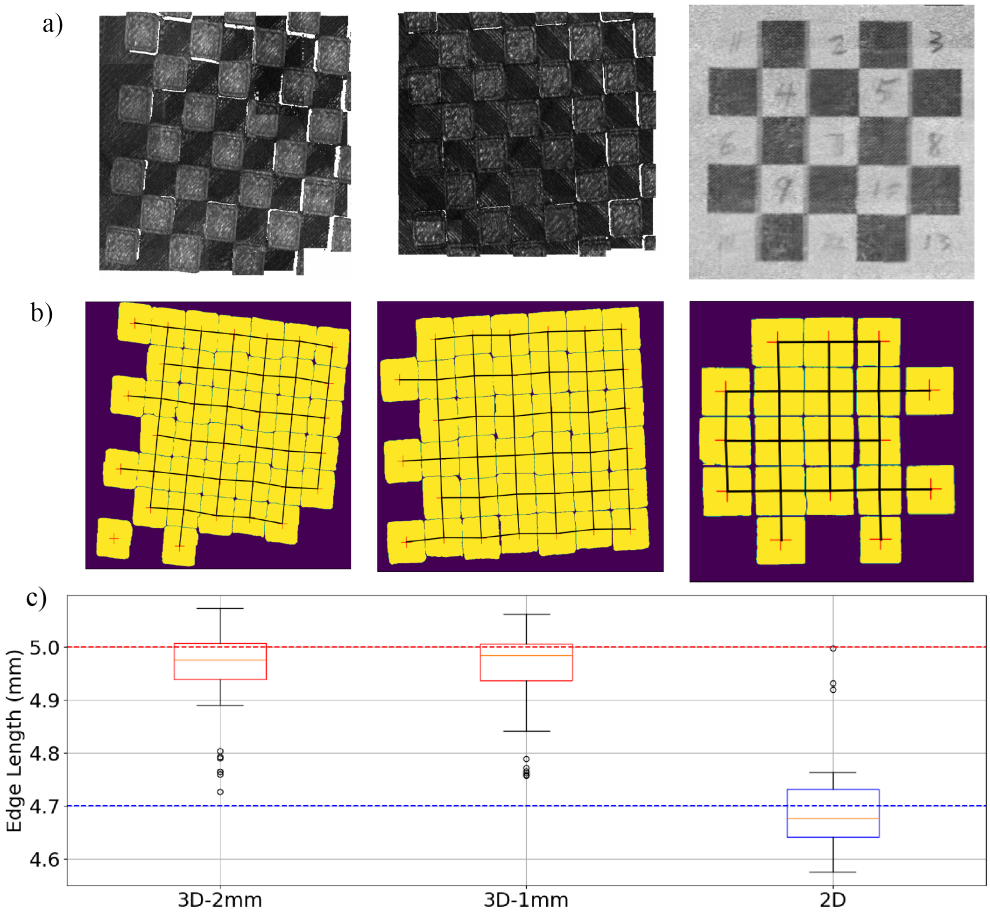

Fig. 4: **Checkerboard OCT Distance Measurement.** (a) OCT scan of wide-area, (b) Segmented squares and center distance measurement, (c) Box plot of measured lengths with ground truth values.

incorporating surgeon feedback. We anticipate our system to be especially useful for surgical scalpels requiring microscale precision, such as laser scalpels, where accurate knowledge of tissue surface information is needed.

Compared to previous work in large-area OCT scanning, our OCT scanner is less cost-prohibitive. It allows quick

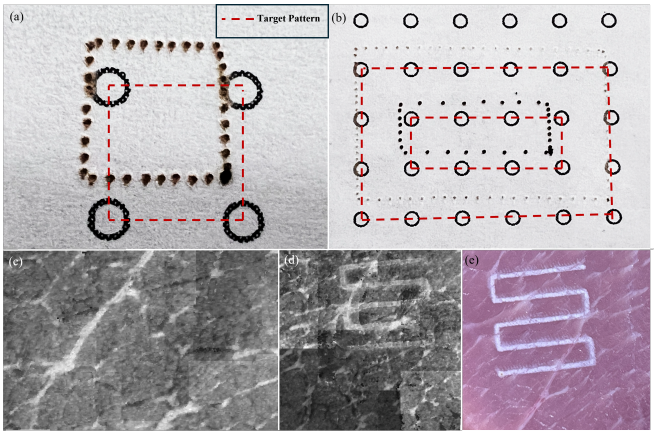

Fig. 5: **Laser ablation based on user-defined patterns.** (a) 1x1 square (b) 1x3 and 3x5 over multiple OCT voxels,(c) pre-ablation, (d) post-ablation OCT volume of porcine muscle, (e) color camera view.

prototyping with an easy interface, albeit at the cost of limited flexibility in accessing raw OCT data [15–18]. Using a "stop-and-stare" scanning strategy maintains high lateral resolution with a tradeoff in imaging time. Unlike other large-area OCT visualization literature [13, 14, 18] where individual OCT voxels are stitched together through manual or commercial software-based registration, our method uses robot pose information and careful OCT volume reconstruction to preserve the spatial relationship between points in stitched OCT volume and perform downstream tasks of resection. We observe that coarse depth sensor visualization has substantial information about tissue surface contour to determine the initial position and pose of the OCT scanner and does not warrant a pre-scan by the OCT.

There are multiple avenues of error propagation when building multimodal multisensor robotic systems for automated laser tissue surgery. We quantify system error to calculate net error that may have propagated during the experimental process. Our target was to keep the net error below $500\mu m$ for delicate surgical tasks involving laser energy as might be seen during pathological tissue resection at a normal-tumor tissue boundary. We were able to represent large-area objects with a maximum error of $102\mu m$ and for experiments involving laser scalpel to ablate patterns defined by users, we had a maximum positional offset of 2.15 mm with ablated region lengths within $427\mu m$ or less of user-defined length. Our proposed work is the first step in moving towards a low-cost robust robotic laser system for automated and semi-automated surgical tasks.

As seen in Figure 3, a slight misalignment between stitched voxels can be observed in the 2-D checkerboard around the first row (1,2,3 fiducial marker) and number 4 is smaller than the rest. There are several areas where our approach can be improved. The current system is limited in speed and ease of data acquisition and transmission compared to other OCT systems and structured light cameras. Though a structured light camera will struggle to scan a point resection, its efficacy needs to be compared in the case of surface resection, as that is of higher practical value for volumetric tissue resection. In the current system, the OCT *B-scan* artifacts can be corrected by implementing a non-telecentric scanning distortion correction [25]. OCT scanner can be calibrated using the Tsai-Lenz method [26] to lower scanner positioning error. Though surgical scene segmentation is beyond this study's scope, methods such as 3DSAM-adapter [27] and UNET-based automatic segmentation [28] can provide additional semantic information to the surgeon. Our current OCT FOV spans a 14.336 x 14.298 mm^2 region which may still have gradual but substantial variation in tissue surface contour. One possible solution could be to use the mean and standard deviation of tissue elevation in a voxel to determine the need for sub-dividing the voxel into smaller sub-voxels or modifying the OCT pose to account for large deviations in surface contour.

In our experiment for ablating patterns based on user feedback, we noticed a consistent positional offset from the center of the target. This could be due to a combination of T_O^E, and T_L^E transform errors. Future design can integrate surgical laser beam and OCT beam through the use of a dichroic mirror and calibrating the laser beam offset with the OCT. Having both the laser beams in a common optic path will also simplify the calibration process though care must be taken to consider the damage threshold of the optics used. To add a close-loop control, it is important to track the laser spot on the tissue. The modified method can be tested on tissue phantoms compatible with laser scalpels and OCT [29]. We can adapt the "Track Anything Model" from [30] for real-time laser tracking and [8] for path planning. This will allow us to maximize the utility of large-area OCT volume and perform volumetric tissue resection while maintaining consistent positioning of the laser beam.

V. Conclusion

We introduce a dual-sensor coarse-to-fine imaging strategy for robotic laser surgery. Our system incorporates real-time surgical scene visualization in varied resolution levels, allowing surgeons to provide feedback for tissue visualization and controlling laser ablation parameters through low-cost yet effective hardware integration and algorithm design. Various system components, depth camera, OCT, and laser-fiber are calibrated to overcome limited FOV and resolution challenges of each sensor. Detailed assessment was performed to quantify system error in high-resolution OCT representation and its impact on laser ablation. Notably, our proposed method demonstrates the first step in developing an autonomous robotic laser system. We anticipate that our system will find use in tumor resections such as meningioma, soft tissue sarcoma, and osteosarcoma. Future research will focus on reducing the OCT distortion, laser-fiber calibration error, and adding a closed-loop controller for volumetric tissue resection utilizing feedback from the OCT.

VI. Acknowledgements

The authors wish to thank Arpit Mishra, Aislinn Hurley, Ryan McNabb, Guangshen Ma and members of Brain Tool Lab for their valuable feedback and support.

References

[1] G. P. Dobson, "Trauma of major surgery: a global problem that is not going away," pp. 47–54, 2020.

[2] X. Zhang, D. Lin, H. Pforsich, and V. W. Lin, "Physician workforce in the united states of america: forecasting nationwide shortages," *Human resources for health*, vol. 18, pp. 1–9, 2020.

[3] H. C. Lee, N. E. Pacheco, L. Fichera, and S. Russo, "When the end effector is a laser: A review of robotics in laser surgery," *Advanced Intelligent Systems*, vol. 4, no. 10, p. 2200130, 2022.

[4] Y. Fan, L. Xu, S. Liu, J. Li, J. Xia, X. Qin, Y. Li, T. Gao, and X. Tang, "The state-of-the-art and perspectives of laser ablation for tumor treatment," *Cyborg and Bionic Systems*, vol. 5, p. 0062, 2024.

[5] N. Feizi, M. Tavakoli, R. V. Patel, and S. F. Atashzar, "Robotics and AI for teleoperation, tele-assessment, and tele-training for surgery in the era of Covid-19: Existing challenges, and future vision," *Frontiers in Robotics and AI*, vol. 8, p. 610677, 2021.

[6] G. Ma, W. Ross, M. Tucker, and P. Codd, "Characterization of photoablation versus incidence angle in soft tissue laser surgery: an experimental phantom study," in *Optical Interactions with Tissue and Cells XXXI*, vol. 11238. SPIE, 2020, pp. 69–80.

[7] G. Ma, W. Ross, and P. J. Codd, "Robotic laser orientation planning with a 3D data-driven method," *arXiv preprint arXiv:2201.01401*, 2022.

[8] V. Y. Wang, R. Prakash, S. R. Oca, E. J. LoCicero, P. J. Codd, and L. J. Bridgeman, "Sampling-based model predictive control for volumetric ablation in robotic laser surgery," *arXiv preprint arXiv:2410.03152*, 2024.

[9] "Scientific Image and Illustration Software — BioRender — biorender.com," https://www.biorender.com/, [Accessed 15-11-2024].

[10] D. Huang, E. A. Swanson, C. P. Lin, J. S. Schuman, W. G. Stinson, W. Chang, M. R. Hee, T. Flotte, K. Gregory, C. A. Puliafito, *et al.*, "Optical coherence tomography," *science*, vol. 254, no. 5035, pp. 1178–1181, 1991.

[11] J. Fujimoto and E. Swanson, "The development, commercialization, and impact of optical coherence tomography," *Investigative ophthalmology & visual science*, vol. 57, no. 9, pp. OCT1–OCT13, 2016.

[12] J. Wang, Y. Xu, and S. A. Boppart, "Review of optical coherence tomography in oncology," *Journal of biomedical optics*, vol. 22, no. 12, pp. 121 711–121 711, 2017.

[13] Y. Huang, X. Li, J. Liu, Z. Qiao, J. Chen, and Q. Hao, "Robotic-arm-assisted flexible large field-of-view optical coherence tomography," *Biomedical Optics Express*, vol. 12, no. 7, pp. 4596–4609, 2021.

[14] P. Ortiz, M. Draelos, A. Narawane, R. P. McNabb, A. N. Kuo, and J. A. Izatt, "Robotically-aligned optical coherence tomography with gaze tracking for live image montaging of the retina," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 3783–3789.

[15] X. Li, Y. Huang, and Q. Hao, "Automated robot-assisted wide-field optical coherence tomography using structured light camera," *Biomedical Optics Express*, vol. 14, no. 8, pp. 4310–4325, 2023.

[16] Y. Li, Y. Fan, and H. Liao, "Robot-assisted optical coherence tomography for automatic wide-field scanning," in *Asian-Pacific Conference on Medical and Biological Engineering*. Springer, 2023, pp. 65–72.

[17] S. Lotz, M. Göb, S. Böttger, L. Ha-Wissel, J. Hundt, F. Ernst, and R. Huber, "Large area robotically assisted optical coherence tomography (LARA-OCT)," *Biomedical Optics Express*, vol. 15, no. 6, pp. 3993–4009, 2024.

[18] X. Ma, M. Moradi, X. Ma, Q. Tang, M. Levi, Y. Chen, and H. K. Zhang, "Large area kidney imaging for pre-transplant evaluation using real-time robotic optical coherence tomography," *Communications Engineering*, vol. 3, no. 1, p. 122, 2024.

[19] N. Katta, A. D. Estrada, A. B. McElroy, A. Gruslova, M. Oglesby, A. G. Cabe, M. D. Feldman, R. D. Fleming, A. J. Brenner, and T. E. Milner, "Laser brain cancer surgery in a xenograft model guided by optical coherence tomography," *Theranostics*, vol. 9, no. 12, p. 3555, 2019.

[20] N. Katta, A. D. Estrada, A. B. McErloy, and T. E. Milner, "Fiber-laser platform for precision brain surgery," *Biomedical optics express*, vol. 13, no. 4, pp. 1985–1994, 2022.

[21] Y. Fan, B. Zhang, W. Chang, X. Zhang, and H. Liao, "A novel integration of spectral-domain optical-coherence-tomography and laser-ablation system for precision treatment," *International journal of computer assisted radiology and surgery*, vol. 13, pp. 411–423, 2018.

[22] Y. Li, Y. Fan, C. Hu, F. Mao, X. Zhang, and H. Liao, "Intelligent optical diagnosis and treatment system for automated image-guided laser ablation of tumors," *International Journal of Computer Assisted Radiology and Surgery*, vol. 16, no. 12, pp. 2147–2157, 2021.

[23] G. Ma, S. R. Oca, Y. Zhu, P. J. Codd, and D. M. Buckland, "A novel robotic system for ultrasound-guided peripheral vascular localization," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 12 321–12 327.

[24] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, *et al.*, "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4015–4026.

[25] Z. S. Navabi, R. Peters, B. Gulner, A. Cherkkil, E. Ko, F. Dadashi, J. O'Brien, M. Feldkamp, and S. B. Kodandaramaiah, "Computer vision guided rapid and precise automated cranial microsurgeries in rodents," *bioRxiv*, pp. 2024–09, 2024.

[26] R. Y. Tsai and R. K. Lenz, "Real time versatile robotics hand/eye calibration using 3D machine vision," in

Proceedings. 1988 IEEE International Conference on Robotics and Automation. IEEE, 1988, pp. 554–561.

[27] S. Gong, Y. Zhong, W. Ma, J. Li, Z. Wang, J. Zhang, P.-A. Heng, and Q. Dou, "3DSAM-adapter: Holistic adaptation of SAM from 2D to 3D for promptable tumor segmentation," *Medical Image Analysis*, vol. 98, p. 103324, 2024.

[28] M. Moradi, X. Du, T. Huan, and Y. Chen, "Feasibility of the soft attention-based models for automatic segmentation of oct kidney images," *Biomedical Optics Express*, vol. 13, no. 5, pp. 2728–2738, 2022.

[29] R. Prakash, K. K. Yamamoto, S. R. Oca, W. Ross, and P. J. Codd, "Brain-mimicking phantom for photoablation and visualization," in *2023 International Symposium on Medical Robotics (ISMR)*. IEEE, 2023, pp. 1–7.

[30] S. Li, A. Gunalan, M. A. Azam, V. Penza, D. G. Caldwell, and L. S. Mattos, "Auto-calm: Autonomous computer-assisted laser microsurgery," *IEEE Transactions on Medical Robotics and Bionics*, 2024.