Kent Academic Repository
Full text document (pdf)

Citation for published version

Zhang, Lin and Ye, Menglong and Giataganas, Petros and Hughes, Michael and Bradu, Adrian and Podoleanu, Adrian G.H. and Yang, Guang-Zhong (2017) From Macro to Micro: Autonomous Multiscale Image Fusion for Robotic Surgery. IEEE Robotics & Automation Magazine, PP (99). p. 1. ISSN 1070-9932.

DOI

https://doi.org/10.1109/MRA.2017.2680543

Link to record in KAR

http://kar.kent.ac.uk/61760/

Document Version

Publisher pdf

From Macro to Micro

Autonomous Multiscale Image Fusion for Robotic Surgery

By Lin Zhang, Menglong Ye, Petros Giataganas, Michael Hughes, Adrian Bradu, Adrian Podoleanu, and Guang-Zhong Yang

Minimally invasive surgery (MIS), performed through a small number of keyhole incisions, has become the standard of care for many general surgical procedures, reducing trauma, blood loss, and other complications and offering patients the prospect of a faster recovery with less postoperative pain. These improvements for the patient, however, require higher dexterity and complex instrument control by the surgeons. Keyhole incisions constrain the motion of surgical instruments, while the loss of stereovision when using a laparoscope or endoscope means that depth perception is much poorer than in traditional open surgery. The desire to tackle these issues has been the main driver

behind the development of robotic MIS systems with stereovision. In particular, the da Vinci robot (Intuitive Surgical, Inc., Sunnyvale, California) is a successful surgical platform, used widely in the treatment of gynecological and urological cancers. While human guidance is essential for MIS, recent studies [1] have suggested that automation of some surgical subtasks, particularly those that are tedious and repetitive or require high precision, can be beneficial in improving accuracy and reducing the cognitive load of the surgeon. For example, several studies have investigated automation of surgical suturing subtasks, including using a suturing tool under fluorescence guidance [2], and other studies have explored areas such as autonomous tissue dissection [3].

A less-explored potential beneficiary of automation is optical biopsy. High-resolution pCLE and OCT offer

visualization of cellular-scale tissue details in situ, providing a real-time alternative to conventional biopsy and histopathology. There has been extensive use of pCLE in diagnostic and surgical procedures in the gastrointestinal tract and abdominal organs [4], [5], and there are a number of other potential applications, particularly in the identification of tumor resection margins. However, the small field of view (FoV) of the probes (typically under 1.0 mm) means that scanning is needed if significant areas of tissue are to be analyzed. For individual images to be stitched together and to form large FoV mosaics, several mosaicing algorithms have been developed, including both real-time [6] and more robust offline processing [7] approaches. However, the operator is required to maintain optimal probe orientation while performing smooth and controlled scanning motions with submillimeter accuracy. For pCLE in particular, which requires direct tissue contact, the probe–tissue force is critical [8], both to ensure good-quality images and to limit the amount of tissue deformation occurring during scanning. Mosaicing under manual control is therefore extremely challenging.

To address this issue, several studies have attempted to robotize pCLE probes with the aim of achieving consistent imaging over large areas of tissue. These developments have included dedicated robotic mechanisms for two-dimensional (2-D) scanning [9]–[11] and for maintaining the desired probe–tissue contact force [8]. Integration with existing robotic systems, such as the da Vinci [8], [12], has also been investigated. However, even with robotic assistance, limited depth perception and poor ergonomics mean that it can still be difficult for clinicians to perform a continuous and smooth scan during teleoperated surgery; maintaining the probe in a specific orientation and ensuring continuous contact with the surface remains difficult and tedious.

OCT differs from pCLE in that it provides cross-sectional images of the tissue's structure. The nominal depth range is typically several millimeters, although penetration into scattering tissue is limited to 1–2 mm in practice. OCT has achieved significant clinical success in ophthalmology, and endoscopic OCT has extended the range of applications to areas like the gastrointestinal tract [13]. It has also been used in laparoscopic and robotic prostatectomy [14] for characterizing neurovascular bundles along the prostate in real time. As for pCLE, it is necessary to mosaic OCT images if a large FoV is required, and this shares many of the difficulties with pCLE mosaicing. However, because OCT is a cross-sectional technique that does not require tissue contact, it can also be used to determine the distance between the probe tip and the tissue. Several studies have suggested the potential benefits of combining cross-sectional OCT images and fluorescence microscopy (e.g., [15]), although most of the designs for combined probes require sacrificing the resolution of the fluorescence channel. Here, by mounting the two probes side by side, we combine the advantages of both techniques to provide complementary diagnostic information— high-resolution surface information and lower-resolution subsurface information.

Autonomous scanning systems have the potential to support the use of optical biopsy in surgical operations by reducing the cognitive load on surgeons and improving the feasibility of scanning over larger surfaces. Zhang et al. [16] proposed such an autonomous framework using the da Vinci robot, where both the pCLE images and the laparoscope were used to guide the scanning. Estimation of the required pCLE probe position for good surface contact was through stereo surface reconstruction. However, due to inaccuracies of this reconstruction, manual adjustment may be required to maintain consistent contact forces and hence good pCLE images. To address this issue, this article visually servoes the scanning motion not only using pCLE images (in the direction transverse to the tissue surface) but also OCT images (in the axial direction). With this approach, the robot is able to ensure that optimal contact force is maintained between the pCLE probe and the tissue, assisting with smooth motion (avoiding stick-slip) and helping to form continuous mosaics.

Furthermore, to facilitate intraoperative tissue diagnosis and identification, a three-dimensional (3-D) visualization method is proposed to fuse the reconstruction of the 3-D tissue surface with 2-D endomicroscopy mosaics and 3-D OCT volumes on the fly, an improvement over the work of [16]. This 3-D fusion approach is designed to provide the surgeon with intuitive real-time visualization of multiscale image information, supporting surgical diagnosis and decision making. The framework has been validated with a series of phantom and ex vivo tissue experiments, with the results demonstrating the potential clinical value of the approach.

Methodology

Framework Overview

The system links custom pCLE and OCT systems with a stereo laparoscope and a patient-side manipulator (PSM) of a da Vinci surgical robot with dVRK controllers. The dVRK controllers (Medical Motion Corporation, Massachusetts) allow the conventional master console of the da Vinci to be bypassed and replaced by the autonomous control system. They are connected to a host personal computer (PC) via an IEEE 1394 firewire interface in a daisy-chain topology. The stereoscopic system provides standard-definition (720×576) video streaming (for both left and right channels) at 25 Hz, which is captured by the host PC using a Kona 4 Peripheral Component Interconnect Express frame grabber (AJA Video Systems, Grass Valley, California).

The pCLE and OCT systems are developed with in-house designs, both consisting of a flexible fiber-optic probe connected to an external base unit. The distal tips of the fiber probes are held by a pickup mount that can be grasped by da Vinci instruments. The pCLE system is a high-speed line-scanning endomicroscope [17] coupled to a Cellvizio ultrahigh-density-fiber probe (Mauna Kea Technologies, Paris, France). It excites tissue at a wavelength of 488 nm and collects fluorescence emission above 500 nm. It is used with typical fluorescent stains that can be excited at 488 nm,

such as acriflavine. The probe, which consists of a 30,000-core fiber imaging bundle and a microlens, provides a FoV of 240 μm and a fiber-sampling limited resolution of approximately 2.4 μm. Because each fiber core within the bundle transmits a pixel of the image, no scanning mechanism is needed at the distal tip of the probe; it is entirely passive. The external line-scanning system illuminates and images a single line of the tissue (via the bundle) at a time, a technique that ensures light is predominantly collected only from an in-focus plane approximately 9 μm in depth. The line-scan design (rather than point scanning, as in conventional confocal microscopes) sacrifices some axial resolution in exchange for a maximum frame rate of 120 frames per second (fps), which is beneficial for this application, as it allows for higher-speed scanning.

The OCT system uses a swept-source laser (Axsun Technologies, Billerica, Massachusetts) with a central wavelength of 1,300 nm. The laser sweeps over a wavelength of 110 nm at a frequency of 100 kHz, with each sweep allowing acquisition of a single axial line (an A scan) through the tissue by the technique of spectral domain low-coherence interferometry. The forward-viewing fiber probe is 13 mm long (fabricated at the Institute of Applied Physics, Nizhny Novgorod, Russia) and has a 2.7-mm-diameter rigid tip. This tip incorporates copper wires and an NdFeB magnet, providing a mechanism for scanning the fiber at 40 Hz and hence generating 2-D images. A gradient-index microlens focuses the beam onto the tissue, and the assembly is covered with Teflon tubing. The FoV of the probe is approximately 1 mm.

Both optical biopsy systems are controlled via a single software interface, developed in LabVIEW (National Instruments, Austin, Texas), running on a secondary PC. Images are acquired and processed in real time, resized to 300×300 (pCLE) and 90×300 pixels (OCT), corresponding to real dimensions of 240×240 μm and $900 \times 3{,}000$ μm, respectively, and are streamed to the host PC at 40 Hz via a Transmission Control Protocol/Internet Protocol (TCP/IP) connection. The software component of the framework is implemented using the robot operating system (ROS) across two computers, with the workload of robot motion control separated from imaging processing and 3-D visualization. The ROS uses TCP/IP to communicate with the pCLE/OCT PC and the servoing PC; this has a round-trip time of about 5 ms. The 2-D mosaic-processing runs at 10 Hz (100 ms). Therefore, for lateral motion compensation using pCLE images, the delay is approximately 110 ms. For vertical motion compensation using OCT distance estimation, the delay is approximately 40 ms.

As shown in Figure 1, the stereo images captured by the camera are used for 3-D tissue-surface reconstruction via a stereo-matching method. Images captured by the pCLE system are registered pairwise by normalized cross correlation and combined by the dead-leaf method. Concurrently, OCT images are mapped into a 3-D volume based on the current probe pose, providing a 3-D volumetric reconstruction over the scanned tissue. The pCLE mosaic, OCT volume, and reconstructed surface are fused together on the fly to provide both macro and microviews of the scanned region.

To generate a mosaic over a defined area, the probe is scanned over a predefined trajectory using the robot, with the trajectory planned in the coordinate system of the mosaic. A visual control component closes the loop by comparing the current and desired probe poses and driving the robot to minimize their difference. This results in the desired region of tissue being imaged regardless of kinematic errors or tissue motion and deformation. Separate from the visual serving loop, the end-effector pose of the robot (in Cartesian space) is then read and set via a dVRK-ROS component, which is connected to a low-level proportional–integral differential controller implemented by the Surgical Assistant Workstation package using the *cisst* library [18].

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

$$b = \frac{\Delta r}{2\pi}. \tag{1}$$

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

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

Finally, the kth trajectory point is defined as

$$\begin{cases} x_k = b \cdot \theta_k \cdot \cos \theta_k \\ y_k = b \cdot \theta_k \cdot \sin \theta_k, \end{cases} \tag{3}$$

where θ_k is obtained using (1) and (2):

$$\theta_k = \sqrt{\frac{2 l_k}{b}}, l_k \in [0, l_{sp}]. \tag{4}$$

Closed-Loop Scanning

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

$$\mathbf{T}_B^{E^*} = \mathbf{T}_B^E \cdot \mathbf{T}_E^M \cdot \mathbf{T}_M^P \cdot \mathbf{T}_P^{P^*} \cdot \mathbf{T}_{P^*}^{M^*} \cdot \mathbf{T}_{M^*}^{E^*}, \tag{5}$$

where $\mathbf{T}_1^2$ denotes a transformation from coordinate {1} to {2}. The definition of the different coordinate systems is illustrated in Figure 2. $\mathbf{T}_B^E$ is the end-effector pose in robot base coordinates, calculated using forward kinematics. The transformation $\mathbf{T}_P^{P^*}$ is the transformation between the current and desired probe positions. $\mathbf{T}_P^{P^*}$ is calculated in every iteration during closed-loop control using the displacement between the current and desired probe positions. $\mathbf{T}_P^{P^*}$ is defined as
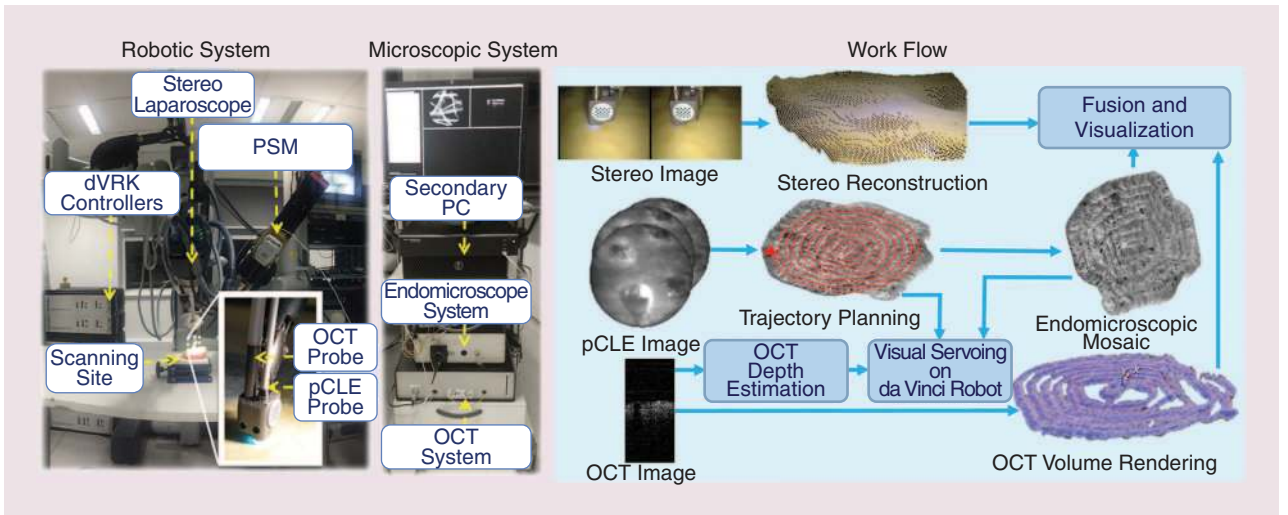
Figure 1. An overview of our experimental setup and the steps involved for autonomous optical biopsy probe scanning and multiscale fusion. The robotic system consists of a set of dVRK controllers, and both the pCLE and OCT probes are grasped by a da Vinci PSM. The microscopic system consists of an endomicroscope (pCLE) system, an OCT system, and a PC used to capture and process pCLE and OCT images. The data flow streaming from the different imaging modalities is processed for visualization and servoing purposes. From a pair of stereo images, a surface of the scene is reconstructed as a point cloud. By stitching pCLE images, a mosaic image can be created, and a 3-D volume can be built from OCT images. These results are fused into a unified window for multiscale visualization.

$$\mathbf{T}_P^{P^*} = \begin{bmatrix} \mathbf{I} & \Delta\mathbf{t} \\ 0 & 1 \end{bmatrix}, \tag{6}$$

where $\Delta\mathbf{t} = [\Delta t_x, \Delta t_y, \Delta t_z]^\top$ is a displacement vector and $\mathbf{I}$ is a 3×3 identity matrix. To obtain the displacement vector, we use the information from either or both of the pCLE and OCT images. The rest of the transformations in (5) are constants that can be measured or calibrated in advance. In particular, the end-effector-to-marker transformation $\mathbf{T}_E^M$ is calibrated using a standard hand–eye calibration method [19]. The marker-to-probe transformation $\mathbf{T}_M^P$ is determined from the computer-aided design model of the adapter. We also note that $\mathbf{T}_{M^*}^{E^*} = (\mathbf{T}_E^M)^{-1}$ and $\mathbf{T}_{P^*}^{M^*} = (\mathbf{T}_M^P)^{-1}$. Continuous detection of the marker is not required during local scanning with closed-loop servoing, as it is only used to determine the transformation between the robot end effector and probe, which does not change in time with a rigid setup. Note that the scanning surface is assumed to be a plane, as each individual scanning region is usually small (about 2×2 mm).

Servoing Using pCLE Images

To allow a continuous mosaic to be generated over the desired area of tissue, the visual servoing loop is closed on the mosaic image. The pCLE mosaicing is performed using an approach similar to the standard real-time technique described in [6] and [17], using normalized cross correlation to estimate the relative shift between each pair of consecutive frames. At the beginning of each scanning procedure, the probe position in the mosaic image $^m p(t = 0) = (0,0)$ is located at the center of the image. During scanning, an estimate of the current probe position $^m p(t)$ at time t is obtained from the mosaic (i.e., from integrated pairwise image shifts over time). Next, by comparing $^m p(t)$ with the kth desired probe position in

the trajectory $^m p^*(k)$, the displacement between the current and desired position is calculated as

$$\begin{bmatrix} \Delta t_x \\ \Delta t_y \end{bmatrix} = k_m \cdot \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} \cdot \begin{bmatrix} ^m p_x^* - {}^m p_x \\ ^m p_y^* - {}^m p_y \end{bmatrix}, \tag{7}$$

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

Servoing Using OCT Images

The distance from the OCT probe to the tissue can be estimated by detecting the top surface of the tissue in the OCT cross-sectional image, which can be seen in Figure 1. The OCT probe is mounted slightly higher than the pCLE probe, so that the top surface of the tissue appears at a nonzero depth in the OCT image. Because the top surface usually maintains the most intense signal, this can be found by simple peak detection, taking the first peak above a user-defined threshold. To mitigate the influence of noise, a Kalman filter is applied to assist accurate distance estimation for servoing. Here, we consider a constant velocity model for the Kalman filter, as the motion of the robot along the depth direction is smooth. At the beginning of each scan, we assume that a good initial contact has been made by the user (i.e., clear pCLE images can be seen), and the current OCT distance estimate d_{oct} is recorded. This distance is then set as the desired distance d_{oct}^*, and the robot is thus required to maintain this distance during scanning. The displacement along the z-axis is defined as
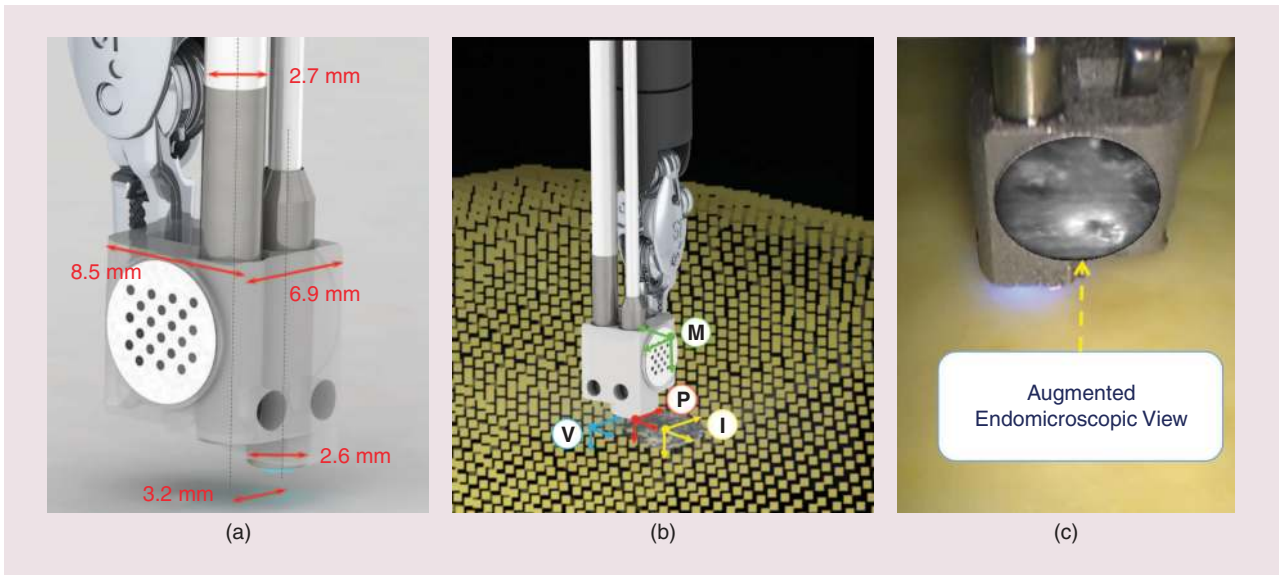
Figure 2. An illustration of the coordinate frames and dimension of the hardware. (a) A da Vinci tool grasping the pickup adapter holding the pCLE and OCT probes. The dimension of the adapter and probes is labeled. (b) The definition of the coordinate frames used in the visual servoing and fusion. The marker and pCLE probe coordinate frames are denoted as Ⓜ and Ⓟ, respectively. The volume and mosaic image coordinate frames are denoted as Ⓥ and Ⓘ, respectively. (c) An augmented laparoscopic view, where a pCLE image is overlaid on the adaptor.

$$\Delta t_z = k_o \cdot (d_{oct}^* - d_{oct}). \tag{8}$$

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

Multiscale Fusion and Visualization

In this work, three different imaging modalities are fused together in a unified visualization framework including both macro- and microviews. For the macroview, a pair of stereo images is used to reconstruct a 3-D surface of the scanning region using an efficient stereo-matching method [20]. The microview then consists of a mosaic image obtained from pCLE images overlaid onto a 3-D volume reconstructed from OCT images.

To place the mosaic image with the OCT volume, we take account of the known lateral offset between the OCT and pCLE probes due to the design of the pickup mount, as shown in Figure 2. The size of the volume is set according to the defined size of the scanning region such that it will be large enough to contain all OCT images during scanning. The size of each voxel in the volume is set equal to the size of the OCT image pixels, such that each voxel represents approximately 10 μm in the plane of the cross section and 10 μm along the out-of-plane direction. A 2-D–3-D mapping allows each pixel in each OCT image to be mapped to a voxel in the volume. For each scan, the volume is initialized such that all voxels are set to invisible. The pose of each OCT image in the volume coordinate frame is then obtained from the current position of the probe in the mosaic image coordinate frame, taking into account the known offset between the two probes. The value of the voxel is simply set equal to the intensity of the corresponding pixel in the OCT image. If the voxel has

already been set using a previous OCT image, its value is updated by averaging the new and current values.

The mosaic image is then scaled according to its real size and placed in the same visualization framework as the OCT volume. To register the OCT volume and pCLE mosaics with the macrosurface reconstruction, we make use of a circular-dot marker attached to the adaptor [Figure 2(a)]. The definition of different coordinate frames is defined in Figure 2(b). When a scan is started, the initial pose of the marker in camera coordinates $\mathbf{T}_C^M$ is recorded. Because there is a known transformation between the marker and the probe $\mathbf{T}_M^P$ and a known offset $\mathbf{T}_V^I$ between the two probes, the registration matrices $\mathbf{T}_C^I$ and $\mathbf{T}_C^V$ can be calculated as

$$\begin{cases} \mathbf{T}_C^I = \mathbf{T}_C^M \cdot \mathbf{T}_M^P \cdot \mathbf{T}_P^I \\ \mathbf{T}_C^V = \mathbf{T}_C^I \cdot \mathbf{T}_I^V. \end{cases} \tag{9}$$

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

Experiments and Results

Accuracy of pCLE Visual Servoing

To evaluate the accuracy and consistency of the pCLE visual servoing, we manufactured a test phantom containing a known grid pattern. The phantom was printed by a laser printer on a sheet of paper and coated by a fluorescent marker, making it visible in the pCLE channel. Every square in the grid pattern had a line thickness of 35 μm and width of 208 μm, as shown in Figure 3. To avoid the effect of
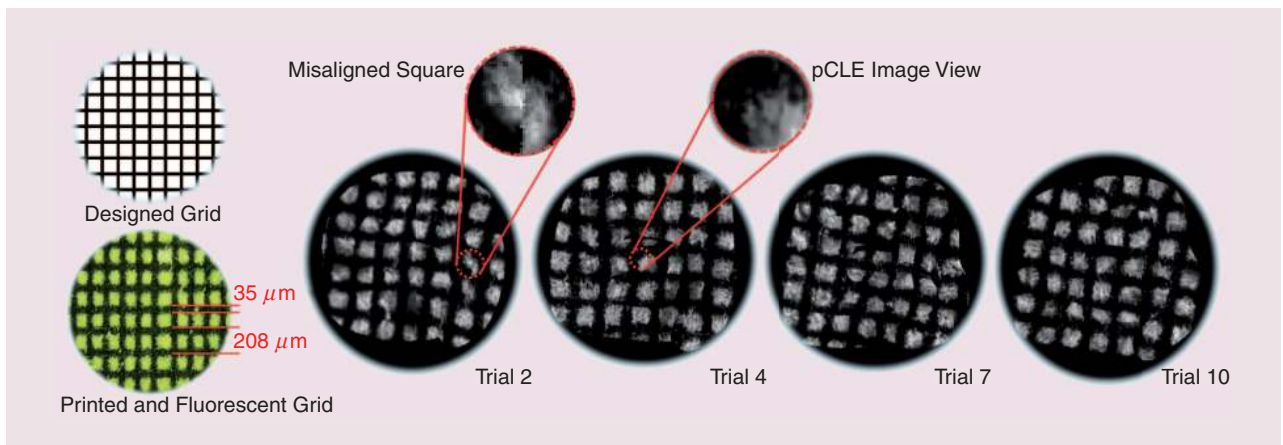
Figure 3. The resulting mosaics from pCLE servoing on a printed grid phantom. Here we show results from trials 2, 4, 7, and 10. In all trials, the mosaic image shows the grid pattern clearly. An example of a misaligned square due to the inaccuracy in the visual servoing is indicated in trial 2.

printing inaccuracies, which can be seen in Figure 3 where the printed grid pattern is not the same as the design, a benchtop microscope was used to take several measurements of the width of individual squares. The average of the measurements was then taken as the ground-truth width of a square. For the results given, the measured ground-truth width of a square was 213.24 μm (compared to an intended width of 208 μm).

To test the consistency of the visual servoing, ten scans were performed over the same phantom, generating ten mosaic images. To maintain the independence of the trials, the phantom was repositioned between each run. In each mosaic image, we measured the width of every square and compared it with the ground-truth value. We took 36 measurements along the horizontal and vertical direction for each image. Due to the intrinsic inaccuracy of the servoing, squares in the pattern can be misaligned, as shown in Figure 3. The fraction of misaligned squares is presented in Table 1. The root-mean-square error (rmse) and interquartile

mean error (IME) were then calculated for each trial, with the IME ignoring the effect of outliers. These results are shown in Table 1, where the total rmse is 14.8 μm and the total IME is 10.0 μm. If this is compared to the FoV of the probe (240 μm), it is clear that both qualitative and quantitative results show that the proposed pCLE servoing method can achieve a high accuracy and good consistency using the da Vinci robot, for the case of a nondeforming sample.

Tendon-driven robots, such as the da Vinci system, are often considered incapable of performing tasks that require such high precision due to backlash and kinematic inaccuracy. To demonstrate how the proposed method improves the robot's capability for pCLE scanning, we compare scanning results with and without the pCLE servoing. As shown in Figure 4(a) and (b), although the kinematic model appears to show that the robot should follow the desired trajectory correctly, the actual mosaic image does not show the corresponding motion, indicating that there is an error in the kinematic model. With pCLE servoing, a comparison of the trajectory and resulting mosaic image demonstrates that the controller corrects the kinematic error to follow the desired trajectory in the space of the mosaic image.

We then tested the framework on ex vivo porcine stomach tissue stained with acriflavine. Unlike the phantom experiments, deformation is now caused by probe interaction with the tissue during scanning. In Figure 4(g), the kinematic trajectory indicates that the robot attempts to compensate for this deformation because the deviation is 0.4 mm more than for the nondeforming phantom shown in Figure 4(c). Figure 4(c) and (g) also shows that the scanning trajectory of an ideal spiral is typically squared off. The most likely causes of this are backlash in the tendon-driven robot and the limited bandwidth of the visual servoing control of the robot position.

Robustness to Unexpected Motion

The system is also somewhat robust to unexpected motion and deformation of the scanning target. To show this qualitatively, we manually moved the phantom to interrupt the scanning and

Table 1. An evaluation of endomicroscopic servoing.

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

mosaicing process. The first row of Figure 4(e) and (f) shows that the servoing is able to recover following the unexpected phantom motion. To confirm the capability of the system under unexpected motion of the tissue, the tissue was placed on a motorized translation stage that was moved laterally with different velocities during the robotic scanning. The mosaic results are shown in Figure 5. The maximal velocity of the lateral motion that would not cause failure (discontinuities in the mosaic image) was 0.5 mm/s. When the velocity was increased further to 0.63 mm/s, there was an obvious discontinuity when the system attempted to correct the motion and continue the mosaicing. Improving the system to deal with faster motion would require an increase in the mosaic-processing frame rate, which is currently set to 10 fps. However, this is not limited in practice by the mosaic algorithm but by the use of a tendon-driven robot that is not capable of fast and accurate motion. If the mosaic frame rate were increased without an increase in velocity, the mosaic algorithm would fail, as the positional shift between frames would be too small to detect accurately.

Validation of OCT Distance Servoing

To obtain high-quality pCLE images, as shown in Figure 6, the probe should be placed in gentle contact with the tissue at all times. In the probe holder, the OCT probe is positioned with an axial offset, slightly higher than the pCLE probe. Optimal pCLE images should therefore be obtained when the distance to the tissue, given by the position of the top surface in the OCT cross-sectional image, is at a fixed value around this distance (the exact distance required is determined experimentally). To evaluate how effective the OCT distance servoing is at maintaining continuous tissue contact, and hence good image quality, we used a translation stage to move a phantom in a cyclic motion with a constant linear velocity along the axial direction of the OCT probe (both toward and away from the probe). Trials were performed
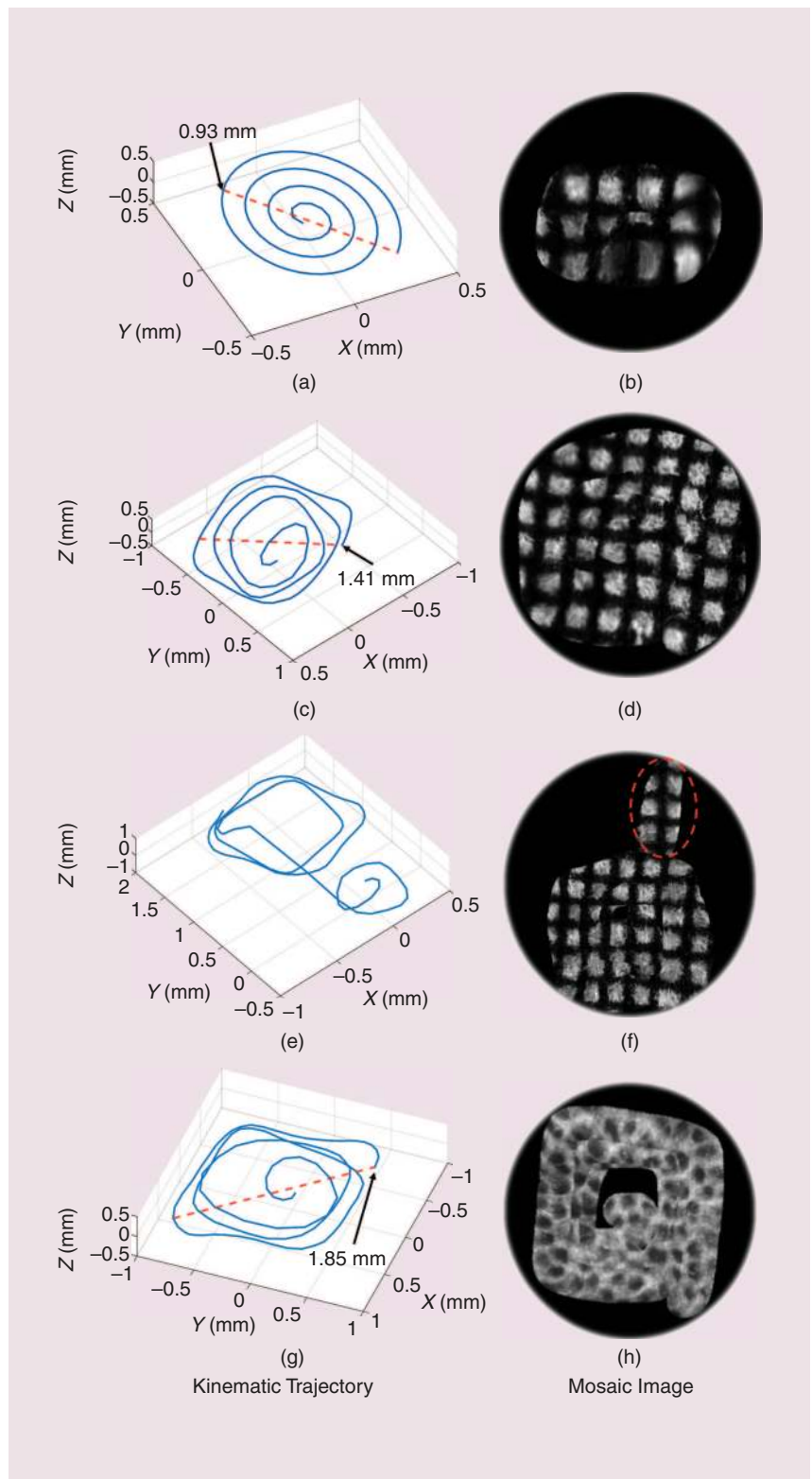

Figure 4. An illustration of how pCLE servoing improves mosaicing results: (a), (c), (e), and (g) show the scanning trajectory based on kinematic reading of the robot's end effector. (b), (d), (f), and (h) show the corresponding mosaic results on the same phantom. (a) and (b) show the result of using only the kinematic model, while (c) and (d) use closed-loop scanning. (e) and (f) show the effect of unexpected motion on the mosaic image. (f) shows the unexpected motion in the red-dashed circle before the closed-loop servoing recovers and returns the probe to the desired trajectory relative to the tissue surface. The diameter of the spiral trajectory from the kinematic model is about 0.93 mm and from the pCLE servoing on the printed pattern and porcine tissue is 1.41 mm and 1.85 mm, respectively.
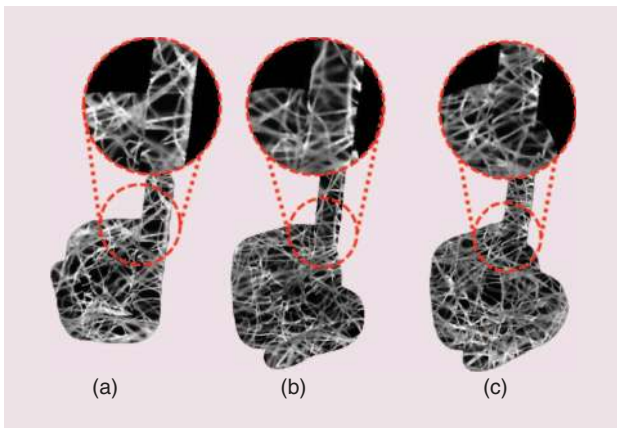
Figure 5. The mosaic results obtained using the framework under unexpected motion with various velocities: (a) 0.38 mm/s, (b) 0.50 mm/s, and (c) 0.63 mm/s. The arm connecting the mosaic over the planned region and the additionally scanned region (caused by the motion) are enlarged to show that it is continuous, suggesting that the probe correctly returned to its planned trajectory on the tissue surface.
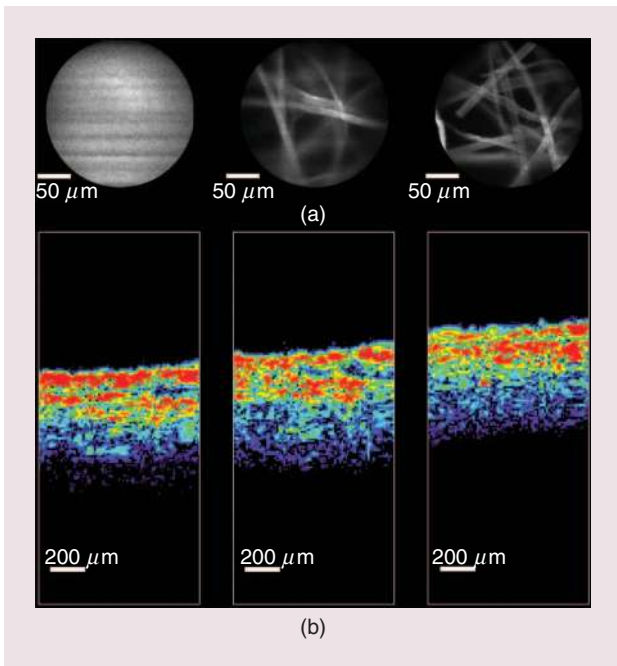

Figure 6. A demonstration of the correlation between (b) the distance to the tissue surface measured in the OCT image and (a) the quality of pCLE images. As the surface moves closer to the probe, from left to right in the figure, the probe-to-tissue distance measured by the OCT probe becomes smaller, while the pCLE image quality improves. Note that the pCLE images are autocontrasted.

with different linear velocities ranging from 16 to 80 μm/s. The results in Figure 7 show how well the robot maintains a desired distance relative to the moving phantom under different peak velocities (16, 32, and 80 μm/s). When the phantom moves, the deviation between the current and desired distances can be generally kept within four pixels (approximately 40 μm). When the phantom stops, the current distance successfully converges to the desired one.
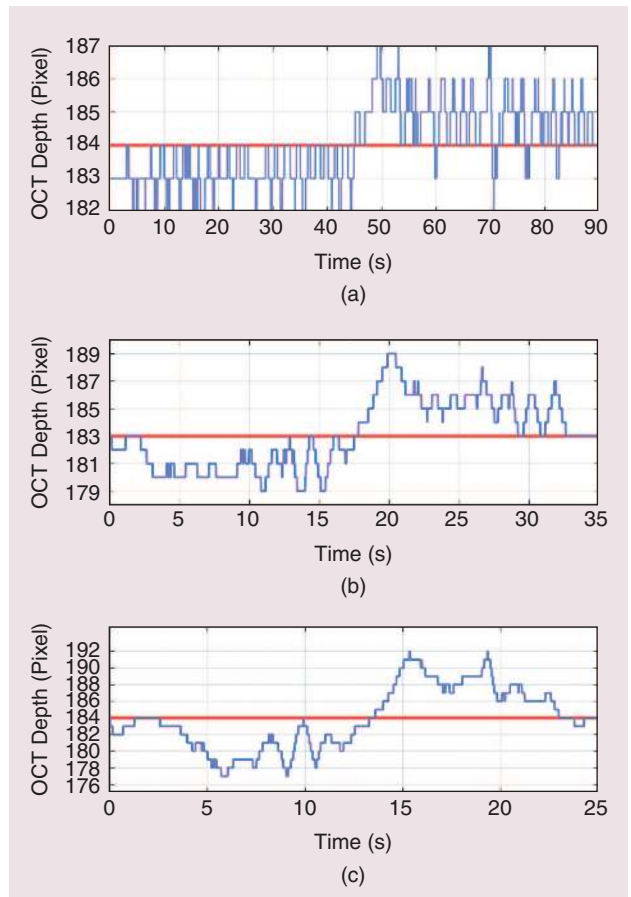

Figure 7. (a) Velocity 16 µm/s, (b) velocity 48 µm/s, and (c) velocity 80 µm/s. The desired OCT distance shown in red is assumed to be constant during servoing. As the phantom starts moving, the deviation between the desired position and the current position measured by the OCT channel initially increases and then decreases toward zero as the system corrects for the motion. Finally, the deviation converges to zero when the phantom stops. One pixel in the OCT image corresponds to approximately 10 μm in real space.

To determine whether the OCT distance servoing improves the robustness of the pCLE scanning, a spiral-scanning task on a curved surface phantom was tested with and without the distance servoing. Figure 8 shows an example where the mosaicing cannot continue without OCT distance servoing, as the pCLE probe loses contact with the phantom and the image becomes blurred. In contrast, as shown in the zoomed region of Figure 8, when contact begins to be lost, the robot is successfully able to recover.

OCT Volume Rendering and Fusion

To validate the OCT volume rendering, we scanned a surgical needle with a thickness of 0.3 mm. In this experiment, the OCT distance servoing was disabled so that only pCLE images were used for the control. As shown in Figure 9, a segment of the needle is scanned and reconstructed in an OCT volume. From the volume, we can clearly see that the needle is placed on a flat tissue surface. To visualize both macro and microinformation, a multiscale image fusion is shown in Figure 10. From the stereo reconstruction in
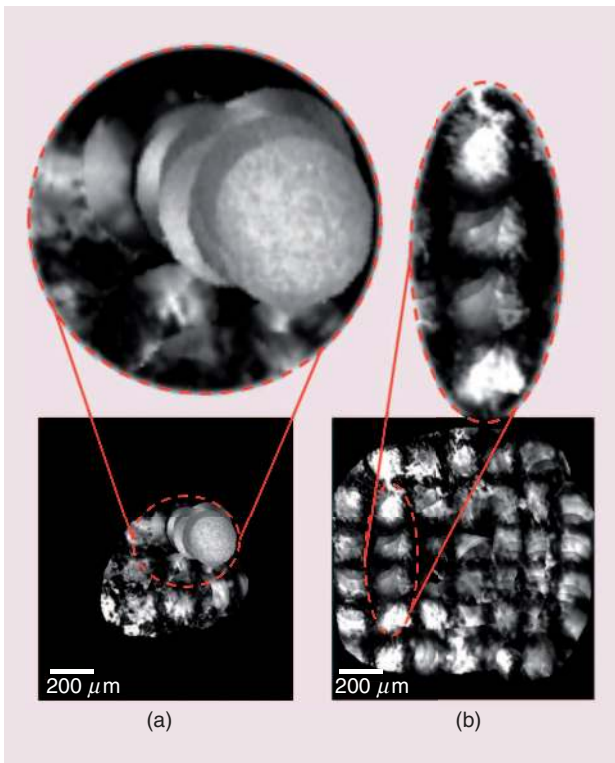
Figure 8. A demonstration of pCLE scanning (b) with and (a) without OCT distance compensation on a curved surface. Without the OCT distance servoing, the mosaic could only continue for a short distance before the probe lost contact, the pCLE image was degraded, and the lateral visual servoing failed. When OCT distance servoing was turned on, a mosaic was generated over the whole trajectory.

macroscale, we can zoom in to the scanned region in microscale, where a mosaic image and an OCT volume are presented.

Conclusions and Future Work

In this article, we have presented an automated scanning framework for pCLE and OCT optical biopsy using the da Vinci surgical robot. It is capable of generating large-area mosaics of both pCLE and OCT images. A crucial feature is that pCLE images are used to close the control loop, and mosaicing results from both static and deforming phantoms demonstrated that this effectively compensates kinematic errors. Furthermore, by using OCT images to maintain a constant distance to the tissue, and hence ensure consistent contact between the pCLE probe and the tissue, the system is able to compensate for target motion along the axial direction. This visual servoing allows for the correction of errors due to tissue deformation, robot positioning, and grasping of the pickup probe. The accuracy of this correction is better than the FoV of the pCLE probe, resulting in continuous 2-D mosaics without gaps or discontinuities, which represent a common problem for open-loop control. Finally, the resulting high-resolution tissue maps at microscale can be fused in real time with a stereo reconstruction at macroscale from the laparoscopic image feed, providing the surgeon with a
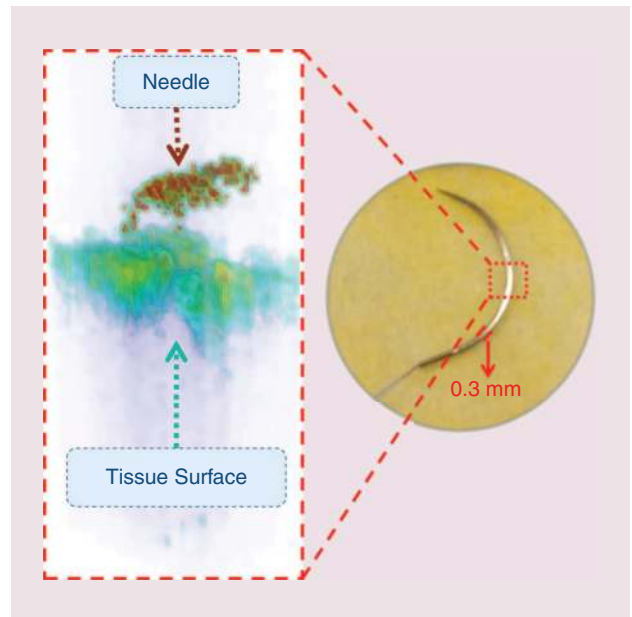

Figure 9. A volume reconstruction of a surgical needle with a thickness of 0.3 mm. The red rectangle indicates the segment that has been reconstructed in the microscopic volume shown on the left.
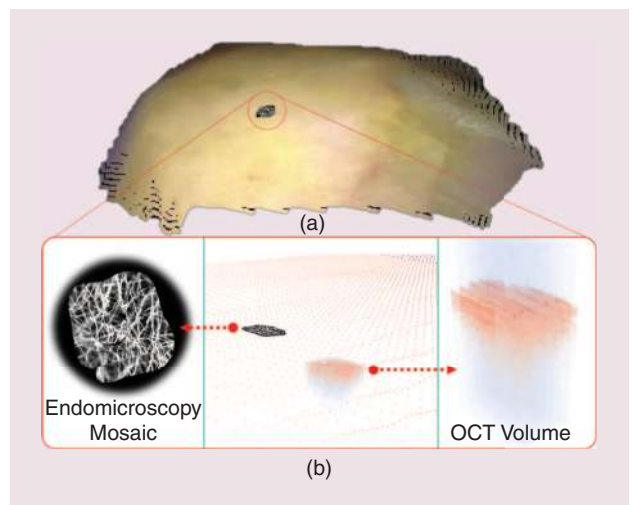

Figure 10. An example of multiscale fusion. (a) A macroscale stereo reconstruction. (b) To link with the microscale, the mosaic image is overlaid on the stereo reconstruction.

multiscale 3-D view of the operating site. This augmented visualization provides a range of potential benefits for intraoperative tissue characterization and surgical planning. Future work will investigate complementary servoing using the laparoscopic image, both for robot motion to the desired scan site and to improve robustness to loss of contact (and hence loss of the pCLE images for servoing), and an investigation of the potential of the system to handle patient motion. A study comparing the results to master–slave-controlled scanning will confirm the practical clinical benefits of the system in providing the surgeon with real-time intraoperative cellular-scale tissue analysis.

References

[1] G. Moustris, S. Hiridis, K. Deliparaschos, and K. Konstantinidis, "Evolution of autonomous and semi-autonomous robotic surgical systems: A review of the literature," *IJMRCAS*, vol. 7, no. 4, pp. 375–392, 2011.

[2] A. Shademan, R. S. Decker, J. D. Opfermann, S. Leonard, A. Krieger, and P. C. Kim, "Supervised autonomous robotic soft tissue surgery," *Sci. Trans. Med.*, vol. 8, no. 337, pp. 337ra64, 2016.

[3] P. Pratt, A. Hughes-Hallett, L. Zhang, N. Patel, E. Mayer, A. Darzi, and G.-Z. Yang, "Autonomous ultrasound-guided tissue dissection," in *Proc. Medical Image Computing and Computer-Assisted Intervention,* 2015, pp. 249–257.

[4] M. B. Wallace, P. Sharma, C. Lightdale, H. Wolfsen, E. Coron, A. Buchner, M. Bajbouj, A. Bansal, A. Rastogi, J. Abrams, J. E. Crook, and A. Meining, "Preliminary accuracy and interobserver agreement for the detection of intraepithelial neoplasia in Barrett's esophagus with probe-based confocal laser endomicroscopy," *Gastrointestinal Endoscopy*, vol. 72, no. 1, pp. 19–24, 2010.

[5] A. Meining, Y. K. Chen, D. Pleskow, P. Stevens, R. J. Shah, R. Chuttani, J. Michalek, and A. Slivka, "Direct visualization of indeterminate pancreaticobiliary strictures with probe-based confocal laser endomicroscopy: A multicenter experience," *Gastrointestinal Endoscopy*, vol. 74, no. 5, pp. 961–968, 2011.

[6] T. Vercauteren, A. Meining, F. Lacombe, and A. Perchant, "Real time autonomous video image registration for endomicroscopy: Fighting the compromises," *Proc. Society of Photo-Optical Instrumentation Engineers,* vol. 6861, p. 68610C, 2008.

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

[8] P. Giataganas, M. Hughes, and G.-Z. Yang, "Force adaptive robotically assisted endomicroscopy for intraoperative tumour identification," *Int. J. Comput. Assist. Radiol. Surg.*, pp. 1–8, 2015.

[9] B. Rosa, M. S. Erden, T. Vercauteren, B. Herman, J. Szewczyk, and G. Morel, "Building large mosaics of confocal endomicroscopic images using visual servoing," *IEEE Trans. Biomed. Eng*, vol. 60, no. 4, pp. 1041–1049, 2013.

[10] G. Dwyer, P. Giataganas, P. Pratt, M. Hughes, and G.-Z. Yang, "A miniaturised robotic probe for real-time intraoperative fusion of ultrasound and endomicroscopy," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2015, pp. 1196–1201.

[11] S. Zuo, M. Hughes, C. Seneci, T. Chang, and G.-Z. Yang, "Towards intraoperative breast endomicroscopy with a novel surface scanning device," *IEEE Trans. Biomed. Eng*, vol. PP, no. 99, p. 1, 2015.

[12] A. Patsias, L. Giraldez-Rodriguez, A. D. Polydorides, R. Richards-Kortum, S. Anandasabapathy, T. Quang, A. G. Sikora, and B. Miles, "Feasibility of transoral robotic-assisted high-resolution microendoscopic imaging of oropharyngeal squamous cell carcinoma," *Head Neck*, vol. 37, no. 8, pp. E99–E102, 2015.

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

[14] R. K. Goel and J. H. Kaouk, "Optical coherence tomography: The past, present and future," *J. Robot. Surg.*, vol. 1, no. 3, pp. 179–184, 2007.

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

[16] L. Zhang, M. Ye, P. Giataganas, M. Hughes, and G.-Z. Yang, "Autonomous scanning for endomicroscopic mosaicing and 3d fusion," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2017.

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

[18] P. Kazanzides, Z. Chen, A. Deguet, G. S. Fischer, R. H. Taylor, and S. P. DiMaio, "An open-source research kit for the da Vinci surgical system," in *Proc. IEEE Int. Conf. Robotics and Automation*, 2014, pp. 6434–6439.

[19] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration," *IEEE Trans. Robot. Autom.*, vol. 5, no. 3, pp. 345–358, 1989.

[20] A. Geiger, M. Roser, and R. Urtasun, "Efficient large-scale stereo matching," in *Proc. Asian Conf. Computer Vision (ACCV 2014)*, 2011, pp. 25–38.

Lin Zhang, Imperial College, London, United Kingdom. E-mail: lz1611@imperial.ac.uk.

Menglong Ye, Imperial College, London, United Kingdom. E-mail: menglong.ye11@imperial.ac.uk.

Petros Giataganas, Imperial College, London, United Kingdom. E-mail: petros.giataganas11@imperial.ac.uk.

Michael Hughes, Imperial College, London, United Kingdom. E-mail: m.hughes@imperial.ac.uk.

Adrian Bradu, University of Kent, United Kingdom. E-mail: a.bradu@kent.ac.uk.

Adrian Podoleanu, University of Kent, United Kingdom. E-mail: ap11@kent.ac.uk.

Guang-Zhong Yang, Imperial College, London, United Kingdom. E-mail: g.z.yang@imperial.ac.uk.