Abstract—In recent years, minimally invasive robotic surgery has shown great promise for enhancing surgical precision and improving patient outcomes. Despite these advances, intraoperative tissue characterisation (such as the identification of cancerous tissue) still relies on traditional biopsy and histology, a process that is time-consuming and often disrupts the normal surgical workflow. In order to provide effective intra-operative decision-making, emerging optical biopsy techniques, such as probe-based confocal laser endomicroscopy (pCLE) and optical coherence tomography (OCT), have been developed to provide real-time assessment of tissue micro-structures. Clinical deployment of these techniques, however, requires large area surveillance, from macro to micro coverage in order to differentiate underlying tissue structures. This article provides a real-time multi-scale fusion scheme for robotic surgery. It demonstrates how the da Vinci® surgical robot, used together with the da Vinci® Research Kit (dVRK), can be used for automated 2D scanning of pCLE/OCT probes, providing large area tissue surveillance by image stitching. Open-loop control of the robot provides insufficient precision for probe scanning, and therefore the motion is visually servoed using the live pCLE images (for lateral position) and OCT images (for axial position). The resulting microscopic tissue maps can then be fused in real-time with a stereo reconstruction from the laparoscopic video, providing the surgeon with a multi-scale 3D view of the operating site.

I. INTRODUCTION

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 stereo vision 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 stereo vision. In particular, the da Vinci® robot (Intuitive Surgical Inc., CA) is a successful surgical platform, used widely in the treatment of gynaecological 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 probe-based confocal laser endomicroscopy (pCLE) and optical coherence tomography (OCT), offer visualisation of cellular scale tissue details in situ, providing a real-time alternative to conventional biopsy and histopathology. pCLE has been applied extensively to diagnostic and surgical procedures in the gastrointestinal tract and abdominal organs [4], [5], and there are a number of potential applications in surgery, particularly for the identification of tumour resection margins. However, the small Field-of-View (FoV) of the probes (typically less than 1.0 mm) means that scanning is needed if significant areas of tissue are to be analysed. 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 sub-millimetre 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 robotise pCLE probes with the aim of achieving consistent imaging over large areas of tissue. These developments have included dedicated robotic mechanisms for 2D scanning [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 tele-operated surgery; maintaining the probe in a specific orientation and ensuring continuous contact with the surface remains difficult and tedious.

OCT differs from pCLE in that it provides cross-sectional images of the tissue’s structure. The nominal depth range is typically several millimetres, although penetration into scattering tissue is limited to 1-2 mm in practice. OCT has achieved significant clinical success in ophthalmology, and endoscopic OCT has extended the range of applications to areas such as the gastrointestinal tract [13]. It has also been used in laparoscopic and robotic prostatectomy [14] for characterising neurovascu-
lar bundles along the prostate in real-time. As for pCLE, it is necessary to mosaic OCT images if a large FoV is required, and this shares many of the difficulties with pCLE mosaicing. However, since OCT is a cross-sectional technique that does not require tissue contact, it can also be used to determine the distance between the probe tip and the tissue. Several studies have suggested the potential benefits of combining cross-sectional OCT images and fluorescence microscopy (e.g., [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 sub-surface 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 in order to maintain consistent contact forces and hence good pCLE images. To address this issue, in this paper the scanning motion is visually servoed not only using pCLE images (in the direction transverse to the tissue surface) but also by OCT images (in the axial direction). With this approach, the robot is able to ensure that optimal contact force is maintained between the pCLE probe and the tissue, assisting with smooth motion (avoiding stick-slip) and helping to form continuous mosaics.

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

II. METHODOLOGY

A. 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 (da Vinci Research Kit, Medical Motion) allow the conventional master console of the da Vinci® to be bypassed and replaced by the autonomous control system. They are connected to a host PC via an IEEE 1394 firewire interface in a daisy chain topology. The stereoscopic system provides standard-definition (720x576) video streaming (for both left and right channels) at 25 Hz, which is captured by the host PC using a Kona 4 PCIe frame grabber (AJA Video System).

The pCLE and OCT systems are developed with in-house designs, both consisting of a flexible fibre optic probe connected to an external base unit. The distal tips of the fibre probes are held by a pick-up mount which can be grasped by da Vinci® instruments. The pCLE system is a high-speed line-scanning endomicroscope [17] coupled to a Cellvizio UHD
fibre probe (Mauna Kea Technologies). It excites tissue at a wavelength of 488 nm and collects fluorescence emission above 500 nm. It is used with typical fluorescent stains which can be excited at 488 nm, such as acriflavine. The probe, which consists of a 30,000-core fibre imaging bundle and a micro-lens, provides a FoV of 240 µm and a fibre-sampling limited resolution of approximately 2.4 µm. Since each fibre core within the bundle transmits a ‘pixel’ of the image, no scanning mechanism is needed at the distal tip of the probe; it is entirely passive. The external line-scanning system illuminates and images a single line of the tissue (via the bundle) at a time, a technique which ensures light is predominantly collected only from an in-focus plane approximately 9 µm in depth. The line-scan design (rather than point-scanning as in conventional confocal microscopes) sacrifices some resolution in exchange for a maximum frame rate of 120 fps, which is beneficial for this application as it allows for higher speed scanning.

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

Both optical biopsy systems are controlled via a single software interface, developed in LabVIEW (National Instruments), running on a secondary PC. Images are acquired and processed in real time, resized to 300x300 (pCLE) and 90x300 pixels (OCT), corresponding to real dimensions of 240 x 240 µm and 90 x 3000 µm respectively, and are streamed to the host PC at 40 Hz via a TCP/IP connection. The software component of the framework is implemented using the Robot Operating System (ROS) across two computers, with the workload of robot motion control separated from imaging processing and 3D visualisation. The ROS uses TCP/IP to communicate with the pCLE/OCT PC and the servoing PC; this has a round trip time of about 5 ms. The 2D mosaic 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 Fig. 1, the stereo images captured by the camera are used for 3D tissue surface reconstruction via a stereo matching method. Images captured by the pCLE system are registered pairwise by normalised cross-correlation and combined by the dead-leaf method. Concurrently, OCT images are mapped into a 3D volume based on the current probe pose, providing a 3D volumetric reconstruction over the scanned tissue. The pCLE mosaic, OCT volume and reconstructed surface are fused together on-the-fly to provide both macro- and micro-views of the scanned region.

To generate a mosaic over a defined area, the probe is scanned over a pre-defined trajectory using the robot, with the trajectory planned in the coordinate system of the mosaic. A visual control component closes the loop by comparing the current and desired probe poses and driving the robot to minimise their difference. This results in the desired region of tissue being imaged regardless of kinematic errors or tissue motion and deformation.

Separate from the visual servoing loop, the end-effector pose of the robot (in Cartesian space) is then read and set via a dVRK-ROS component which is connected to a low-level PID controller implemented by the SAW package using the cisst library [18].

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

\[ b = \frac{\Delta r}{2\pi}. \]  

In order to reduce the likelihood of producing gaps in the mosaic image, whilst preventing excessive over-sampling, ∆r 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:

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

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

\[ \begin{align*}
    x_k &= b \cdot \theta_k \cdot \cos \theta_k \\
    y_k &= b \cdot \theta_k \cdot \sin \theta_k,
\end{align*} \]  

where \( \theta_k \) is obtained using equation 1 and 2:

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

### B. Closed Loop Scanning

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

\[ T_{E^*}^B = T_{B}^E \cdot T_{M}^M \cdot T_{p}^{p^*} \cdot T_{M^*}^{E^*}, \]  

where \( T_{E^*}^B \) denotes a transformation from coordinate \{1\} to \{2\}. The definition of the different coordinate systems is illustrated in Fig. 2. \( T_{E^*}^B \) is the end-effector pose in robot base coordinates, calculated using forward kinematics. The transformation \( T_{p}^{p^*} \) is the transformation between the current and desired probe positions. \( T_{p}^{p^*} \) is calculated in every iteration during closed loop control using the displacement between the current and desired probe positions. \( T_{p}^{p^*} \) is defined as:

\[ T_{p}^{p^*} = \begin{bmatrix} I & \Delta t \\ 0 & 1 \end{bmatrix} , \]
marker and pCLE probe coordinate frames are denoted as \{OCT\} probes. The dimension of the adapter and probes is labelled; (b) shows the definition of the coordinate frames used in the visual servoing and fusion. The volume and mosaic image coordinate frames are denoted as \{V\} and \{I\} respectively; (c) shows an augmented laparoscopic view where a pCLE image is overlaid on the adaptor.

where \( \Delta t = [\Delta t_x, \Delta t_y, \Delta t_z]^\top \) is a displacement vector and \( I \) is a 3×3 identity matrix. In order 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 equation 5 are constants that can be measured or calibrated in advance. In particular, the end-effector to marker transformation \( T_{E}^{M} \) is calibrated using a standard hand-eye calibration method [19]. The marker to probe transformation \( T_{P}^{M} \) is determined from the CAD model of the adapter. We also note that \( T_{E}^{M} = (T_{E}^{M})^{-1} \) and \( T_{P}^{M} = (T_{P}^{M})^{-1} \). Continuous detection of the marker is not required during local scanning with closed loop servoing, as it is only used to determine the transformation between the robot end-effector and probe, which does not change in time with a rigid setup. Note that the scanning surface is assumed to be a plane as each individual scanning region is usually small (about 2x2mm).

1) Servoing using pCLE Images: To allow a continuous mosaic to be generated over the desired area of tissue, the visual servoing loop is closed on the mosaic image. The pCLE mosaicing is performed using an approach similar to the standard real-time technique described in [6], [17], using normalised cross-correlation to estimate the relative shift between each pair of consecutive frames. At the beginning of each scanning procedure, the probe position in the mosaic image \( m_p(t=0) = (0,0) \) is located at the centre of the image. During scanning, an estimate of the current probe position \( m_p(t) \) at time \( t \) is obtained from the mosaic (i.e. from integrated pair-wise image shifts over time). Next, by comparing \( m_p(t) \) with the \( k \)-th desired probe position in the trajectory \( m_p^*(k) \), the displacement between the current and desired position is calculated as:

\[
\begin{bmatrix}
\Delta t_x \\
\Delta t_y \\
\Delta t_z 
\end{bmatrix} = k_m \cdot \begin{bmatrix}
\cos \alpha & -\sin \alpha \\
\sin \alpha & \cos \alpha 
\end{bmatrix} \begin{bmatrix}
m_{p_x}^* - m_{p_x} \\
m_{p_y}^* - m_{p_y}
\end{bmatrix},
\]

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

2) Servoing using OCT Images: The distance from the OCT probe to the tissue can be estimated by detecting the top surface of the tissue in the OCT cross-sectional image, which can be seen in Fig. 1. The OCT probe is mounted slightly higher than the pCLE probe, so that the top surface of the tissue appears at a non-zero depth in the OCT image. Since the top surface usually maintains the most intense signal, this can be found by simple peak detection, taking the first peak above a user-defined threshold. To mitigate the influence of noise, a Kalman filter is applied to assist accurate distance estimation for servoing. Here, we consider a constant velocity model for the Kalman filter as the motion of robot along the depth direction is smooth. At the beginning of each scan, 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:

\[
\Delta t_z = k_o \cdot (d_{oct}^* - d_{oct}).
\]
Here $k_o$ is a constant that converts the distance from pixels to real distance.

C. Multiscale Fusion and Visualisation

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

To place the mosaic image with the OCT volume, we take account of the known lateral offset between the OCT and pCLE probes due to the design of the pick-up mount, as shown in Fig. 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 $\mu$m in the plane of the cross-section, and 10 $\mu$m along the out-of-plane direction. A 2D-3D mapping allows each pixel in each OCT image to be mapped to a voxel in the volume. For each scan, the volume is initialised such that all voxels are set to invisible. The pose of each OCT image in the volume coordinate frame is then obtained from the current position of the probe in the mosaic image coordinate frame, taking into account the known offset between the two probes. The value of the voxel is simply set equal to the intensity of the corresponding pixel in the OCT image. If the voxel has already been set using a previous OCT image, its value is updated by averaging the new and current values.

The mosaic image is then scaled according to its real size and placed in the same visualisation framework as the OCT volume. To register the OCT volume and pCLE mosaics with the macro surface reconstruction, we make use of a circular-dot marker attached to the adaptor Fig. 2(a). The micro-view then consists of a mosaic image obtained from pCLE images overlaid onto a 3D volume reconstructed from OCT images.

To test the consistency of the visual servoing, 10 scans were performed over the same phantom, generating 10 mosaic images. To maintain the independence of the trials, the phantom was repositioned between each run. In each mosaic image, 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 Fig. 3. The fraction of misaligned squares is presented in Table I. 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 I where the total RMSE is 14.8 $\mu$m and the total IME is 10.0 $\mu$m. If this is compared to the FoV of the probe (240 $\mu$m), it is clear that both qualitative and quantitative results show that the proposed pCLE servoing method can achieve a high accuracy and good consistency using the da Vinci® robot, for the case of a non-deforming sample.

Tendon-driven robots such as the da Vinci® system, are often considered incapable of performing tasks that require such high precision due to backlash and kinematic inaccuracy. To demonstrate how the proposed method improves the robot’s capability for pCLE scanning, we compare scanning results with and without the pCLE servoing. As shown in Figs. 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, comparison of the trajectory and resulting mosaic image demonstrates that the controller corrects the kinematic error in order to follow the desired trajectory in the space of the mosaic image.

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 Fig. 4(g), the kinematic trajectory indicates that the robot attempts to compensate for this deformation, since the deviation is 0.4 mm more than for the non-deforming phantom shown in Fig. 4(c). It can also be seen that the scanning trajectory of an ideal spiral is typically ‘squared-off’ as shown in Figs. 4(c) and (g). The most likely causes of this are backlash in the tendon-driven robot, and the limited bandwidth of the visual servoing control of the robot position.

III. Experiments and Results

A. 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 $\mu$m and width of 208 $\mu$m, as shown in Fig. 3. To avoid the effect of printing inaccuracies, which can be seen in Fig. 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 shown below, the measured ground truth width of a square was 213.24 $\mu$m (compared to an intended width of 208 $\mu$m).

To test the consistency of the visual servoing, 10 scans were performed over the same phantom, generating 10 mosaic images. To maintain the independence of the trials, the phantom was repositioned between each run. In each mosaic image, 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 Fig. 3. The fraction of misaligned squares is presented in Table I. 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 I where the total RMSE is 14.8 $\mu$m and the total IME is 10.0 $\mu$m. If this is compared to the FoV of the probe (240 $\mu$m), it is clear that both qualitative and quantitative results show that the proposed pCLE servoing method can achieve a high accuracy and good consistency using the da Vinci® robot, for the case of a non-deforming sample.

Tendon-driven robots such as the da Vinci® system, are often considered incapable of performing tasks that require such high precision due to backlash and kinematic inaccuracy. To demonstrate how the proposed method improves the robot’s capability for pCLE scanning, we compare scanning results with and without the pCLE servoing. As shown in Figs. 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, comparison of the trajectory and resulting mosaic image demonstrates that the controller corrects the kinematic error in order to follow the desired trajectory in the space of the mosaic image.

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 Fig. 4(g), the kinematic trajectory indicates that the robot attempts to compensate for this deformation, since the deviation is 0.4 mm more than for the non-deforming phantom shown in Fig. 4(c). It can also be seen that the scanning trajectory of an ideal spiral is typically ‘squared-off’ as shown in Figs. 4(c) and (g). The most likely causes of this are backlash in the tendon-driven robot, and the limited bandwidth of the visual servoing control of the robot position.
Fig. 3. Resulting mosaics from pCLE servoing on a printed grid phantom. Here we show results from trial 2, 4, 7 and 10. In all trials, the mosaic image shows the grid pattern clearly. An example of misaligned square due to the inaccuracy in the visual servoing is indicated in trial 2.

| Trial | 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 |

### 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     |

### B. 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 mosaicing process. The first row of Figs. 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 motorised translation stage which was moved laterally with different velocities during the robotic scanning. The mosaic results are shown in Fig. 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 frames per second. However, this is not limited in practice by the mosaic algorithm but by the use of a tendon-driven robot that is not capable of fast and accurate motion. If the mosaic frame rate was increased without an increase in velocity, the mosaic algorithm would fail as the positional shift between frames would be too small to detect accurately.

### C. Validation of OCT Distance Servoing

In order to obtain high quality pCLE images, as shown in Fig. 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 towards and away from the probe). Trials were performed with different linear velocities ranging from 16 to 80 µm/s. The results in Fig. 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 4 pixels (approximately 40 µm). When the phantom stops, the current distance successfully converges to the desired one.

To determine whether the OCT distance servoing improves the robustness of the pCLE scanning, a spiral scanning task on a curved surface phantom was tested with and without the distance servoing. Fig. 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 Fig. 8, when contact begins to be lost, the robot is successfully able to recover.

### D. 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 Fig. 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 visualise both macro and micro information, a multiscale image fusion is shown in Fig. 10. From the stereo reconstruction in macro scale, we can zoom in to the scanned region in micro scale where a mosaic image and an OCT volume are presented.

IV. CONCLUSION 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
Fig. 7. The desired OCT distance shown in red is assumed to be constant during servoing. As the phantom starts moving, the deviation between the desired potion and the current position measured by the OCT channel initially increases and then decreases towards zero as the system corrects for the motion. Finally, the deviation converges to zero when the phantom stops. 1 pixel in the OCT image corresponds to approximately $10 \mu m$ in real space.

Fig. 8. Demonstration of pCLE scanning with and without OCT distance compensation on a curved surface. Without the OCT distance servoing, the mosaic could only continue for a short distance before the probe lost contact, the pCLE image was degraded and the lateral visual servoing failed. When OCT distance servoing was turned on, a mosaic was generated over the whole trajectory.

Fig. 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.

ACKNOWLEDGMENT

Lin Zhang is financially supported by China Scholarship Council (CSC) and the Hamlyn Centre during his PhD study. This work was partly funded by EPSRC Grant EP/N019318/1: REBOT: Robotic Endobronchial Optical Tomography; and EP/N022521/1: Translational Alliance: SMART-Endomicroscopy. Data underpinning this study was demonstrated that this effectively compensates kinematic errors. Furthermore, by using OCT images to maintain a constant distance to the tissue, and hence ensure consistent contact between the pCLE probe and the tissue, the system is able to compensate for target motion along the axial direction. This visual servoing allows for the correction of errors due to tissue deformation, robot positioning and grasping of the pick-up probe. The accuracy of this correction is better than the FoV of the pCLE probe, resulting in continuous 2D mosaics without gaps or discontinuities, which represent a common problem for open loop control. Finally, the resulting high-resolution tissue maps at micro-scale can be fused in real-time with a stereo reconstruction at macro-scale from the laparoscopic image-feed, providing the surgeon with a multi-scale 3D view of the operating site. This augmented visualisation provides a range of potential benefits for intraoperative tissue characterisation and surgical planning. Future work will investigate complementary servoing using the laparoscopic image, both for robot motion to the desired scan site, and to improve robustness to loss of contact (and hence loss of pCLE images for servoing), and an investigation of the potential of the system to handle patient motion. A study comparing the results to master-slave controlled scanning will confirm the practical clinical benefits of the system in providing the surgeon with real-time intraoperative cellular-scale tissue analysis.

Lin Zhang is financially supported by China Scholarship Council (CSC) and the Hamlyn Centre during his PhD study. This work was partly funded by EPSRC Grant EP/N019318/1: REBOT: Robotic Endobronchial Optical Tomography; and EP/N022521/1: Translational Alliance: SMART-Endomicroscopy. Data underpinning this study was demonstrated that this effectively compensates kinematic errors. Furthermore, by using OCT images to maintain a constant distance to the tissue, and hence ensure consistent contact between the pCLE probe and the tissue, the system is able to compensate for target motion along the axial direction. This visual servoing allows for the correction of errors due to tissue deformation, robot positioning and grasping of the pick-up probe. The accuracy of this correction is better than the FoV of the pCLE probe, resulting in continuous 2D mosaics without gaps or discontinuities, which represent a common problem for open loop control. Finally, the resulting high-resolution tissue maps at micro-scale can be fused in real-time with a stereo reconstruction at macro-scale from the laparoscopic image-feed, providing the surgeon with a multi-scale 3D view of the operating site. This augmented visualisation provides a range of potential benefits for intraoperative tissue characterisation and surgical planning. Future work will investigate complementary servoing using the laparoscopic image, both for robot motion to the desired scan site, and to improve robustness to loss of contact (and hence loss of pCLE images for servoing), and an investigation of the potential of the system to handle patient motion. A study comparing the results to master-slave controlled scanning will confirm the practical clinical benefits of the system in providing the surgeon with real-time intraoperative cellular-scale tissue analysis.

ACKNOWLEDGMENT

Lin Zhang is financially supported by China Scholarship Council (CSC) and the Hamlyn Centre during his PhD study. This work was partly funded by EPSRC Grant EP/N019318/1: REBOT: Robotic Endobronchial Optical Tomography; and EP/N022521/1: Translational Alliance: SMART-Endomicroscopy. Data underpinning this study was
generated by code which is openly accessible via https://github.com/hamlyn-centre/auto_scan.

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,” International Journal of Medical Robotics Computer Assisted Surgery, 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,” Science translational medicine, vol. 8, no. 337, pp. 337ra64–337ra64, 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 Medical Imaging Computing and Computer-Assisted Intervention–MICCAI 2015. Springer, 2015.

[4] M. B. Wallace, P. Sharma, C. Lightdale, H. Wolfsen, E. Coron, A. Buchner, M. Bajbouj, A. Bansal, A. Rastogi, J. Abrams et al., “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 pancreatobiliary 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. SPIE, vol. 6861, pp. 68 610C–68 610C–8, 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 Analysis, 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,” International Journal Computer Assisted Radiology and Surgery, 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. Robot. Automat., May 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, pp. 1–1, 2015.

[12] A. Patsias, L. Giraldez-Rodriguez, A. D. Polydorides, R. Richards-Kortum, S. Anandababapathy, 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,” Optics letters, vol. 24, no. 19, pp. 1358–1360, 1999.

[14] R. K. Goel and J. H. Kaouk, “Optical coherence tomography: the past, present and future,” Journal of Robotic Surgery, 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,” Optics Letters, 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. Robot. Automat., 2017, In press.

[17] M. Hughes and G.-Z. Yang, “Line-scanning fiber bundle endomicroscopy with a virtual detector slit,” Biomedical Optics 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. Robot. Automat., May 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 Computer Vision – ACCV 2014. Springer, 2011, pp. 25–38.