

Autonomous Scanning for Endoscopic Mosaicing and 3D Fusion

Lin Zhang, Menglong Ye, Petros Giataganas, Michael Hughes and Guang-Zhong Yang

Abstract—Robotic-assisted Minimally Invasive Surgery (RMIS) can benefit from the automation of common, repetitive or well-defined but ergonomically difficult tasks. One such task is the scanning of a pick-up endomicroscopy probe over a complex, undulating tissue surface in order to enhance the effective field-of-view through video mosaicing. In this paper, the da Vinci[®] surgical robot, through the dVRK framework, is used for autonomous scanning and 2D mosaicing over a user-defined region of interest. To achieve the level of precision required for high quality mosaic generation, which relies on sufficient overlap between consecutive image frames, visual servoing is performed using a combination of a tracking marker attached to the probe and the endomicroscopy images themselves. The resulting sub-millimetre accuracy of the probe motion allows for the generation of large endomicroscopy mosaics with minimal intervention from the surgeon. It also allows the probe to be maintained in an orientation perpendicular to the local tissue surface, providing optimal imaging results. Images are streamed from the endomicroscope and overlaid live onto the surgeons view, while 2D mosaics are generated in real-time, and fused into a 3D stereo reconstruction of the surgical scene, thus providing intuitive visualisation and fusion of the multi-scale images. The system therefore offers significant potential to enhance surgical procedures, by providing the operator with cellular-scale information over a larger area than could typically be achieved by manual scanning.

I. INTRODUCTION

Minimally Invasive Surgery (MIS), in which surgical procedures are performed through small incisions, is a widely used and effective approach for surgical oncology. It has significant advantages over the traditional open techniques, in terms of reduced blood loss and postoperative pain, lower infection rates, and shorter hospital stays. However, MIS also has some drawbacks, which include constrained motion due to ‘keyhole’ access, poor depth perception from only a monocular laparoscope, and lack of direct tissue interaction. To provide better manipulation and depth perception, robotic systems with stereo vision have been developed for MIS. In particular, the da Vinci[®] robot (Intuitive Surgical Inc., CA) is a successful surgical platform which has been widely used in the treatment of gynecological and urological cancer. However, the da Vinci[®] robot is purely a master-slave system that only provides tele-operation abilities to the surgeons, with no automation or artificial intelligence features.

Whilst human guidance is essential for MIS, a recent review [1] has concluded that surgical automation can be used for some important steps of the operation, which can then reduce the cognitive load of the surgeon during certain surgical tasks that require repetitive and precise motion. To

this end, studies have been conducted, for instance, to maintain consistent motion for ultrasound elastography [2], or to automatically compensate heart motion in cardiovascular surgery [3]. In [4], a human-machine collaborative framework was proposed to improve surgeons’ performance by semi-automating surgical subtasks. Further studies have also investigated how to perform automatic surgical debridement [5], tissue dissection [6], and brain ablation [7].

Another obvious target for automation is the process of generating *in situ* microscopic images of tissue using probe-based confocal laser endomicroscopy (pCLE), such that the microscopic images can be used for *in vivo* and *in situ* pathology analysis. Endomicroscopy has been used extensively for diagnostic and surgical procedures in the gastrointestinal tract and abdominal organs [8], [9], mainly to discriminate between normal and cancerous tissue regions. However, the miniaturisation requirement for the probe results in a limited field-of-view (FoV) of typically 0.25-1 mm, and therefore, it is difficult to characterise larger areas of tissue or to return to previously sampled points [10]. Furthermore, retrieving high quality and stable microscopic images with a manually manipulated probe is a challenging task, particularly when the operator attempts to ‘mosaic’ images to synthesise a larger FoV comparable to histological images. Mosaicing requires the operator to maintain optimal probe pose and tissue-contact while performing slow and controlled scanning motions with sub-millimetre accuracy [11].

Recent work has proposed using robotised instruments to enhance control of the endomicroscopy probe, allowing large mosaics to be formed to assist real-time diagnosis. These developments have included novel robotic mechanisms for 2D scanning [12]–[14] or for maintaining a desired probe-tissue contact force [11]. Integration with existing robotic systems, such as the da Vinci[®] [15], [16], has also been investigated. However, the combination of weak-depth perception and poor ergonomics, meant that it is still difficult for clinicians to perform a continuous and smooth microscopic scan during tele-operated surgery. Maintaining the probe in a specific orientation (normal to the tissue surface) and ensuring continuous contact with the tissue remains a challenging and tedious task. Therefore, the development of an autonomous scanning system for endomicroscopy would benefit its use in surgical operations by reducing the cognitive load on users and improving scanning accuracies over larger 3D surfaces.

In this work, we present a 6 degree-of-freedom (DoF) visual servoing method based on the da Vinci[®] robot with the da Vinci Research Kit (dVRK). With our proposed approach, the robot is able to achieve smooth motion with

All authors are with The Hamlyn Centre for Robotic Surgery, Institute of Global Health Innovation, Imperial College London, SW7 2AZ, London, UK

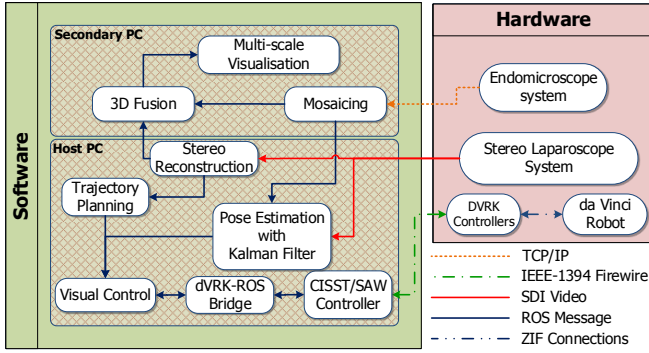


Fig. 1. An overview of our proposed system framework for autonomous endoscopic scanning and 3D mosaicing.

sub-millimetre accuracy, and we have demonstrated that an autonomous large-area endomicroscopy scan can be performed by the robot over a user-defined area. In addition, to facilitate intraoperative tissue diagnosis and identification, a 3D visualisation method is proposed to fuse the 3D tissue surface with multiple microscopic mosaics on-the-fly, which is an improvement over the work of [17]. This 3D fusion approach is designed to provide the surgeon with an intuitive and effective real-time visualisation of multi-scale imaging information, which supports surgical diagnosis and planning. Our framework has been tested on phantom and *ex vivo* tissue experiments and the results demonstrate its potential clinical value.

II. METHODOLOGY

A. System Overview

The hardware components of the proposed framework, highlighted in red in Fig. 1, consist of a patient side manipulator (PSM) of a da Vinci[®] robot with dVRK controllers, a stereo laparoscope, and a custom endomicroscopy system. The dVRK controllers are connected to a host PC via a IEEE 1394 firewire interface in a daisy chain topology. The stereoscopic system provides SD (720x576) video streaming for both left and right channels at 25 Hz. In the host PC, the stereo stream is captured by a Kona 4 PCIe frame grabber (AJA Video System).

The endomicroscope system is an in-house laser line-scanning fluorescence system, described fully in [18], coupled to a Cellvizio UHD Probe (Mauna Kea Technologies). 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 . The line-scanning system provides optical sectioning, meaning that the endomicroscope collects light only from a in-focus plane approximately 20 μm in depth. Processed microscopic images are scaled to 300x300 pixels and served to the controller at 80 Hz via a TCP/IP connection. In addition to the host PC, a second PC with modest configuration is dedicated to 2D mosaicing and 3D visualisation.

The software component of the proposed framework runs on the Robot Operating System (ROS), highlighted in green

in Fig. 1. The stereo images captured from the camera are used for 3D tissue surface reconstruction via a stereo matching method. The reconstructed 3D surface is used for two purposes: (1) to plan a scanning trajectory based on its positional and normal information; (2) to provide 3D fusion with the 2D image mosaics. The video stream from the laparoscope system is also used for estimating the pose of the endomicroscopy probe.

As shown in Fig. 2, the end-effector of the robot grasps an adapter that holds the endomicroscopy probe. A marker (KeyDot[®], Key Surgical, Minnesota) is attached to the adapter for pose estimation of the probe. The stereo system is calibrated using the method presented in [19] to obtain the intrinsic and extrinsic camera parameters such that the 3D depths of the tissue can be recovered via stereo matching. The trajectory planning consists of two parts: a global plan to guide the probe to the desired location of the scan (identified by the user), and a local plan to generate a mosaic.

A visual control component closes the loop by comparing the current and desired probe poses and commanding the robot to minimise their difference, causing the probe to follow the planned trajectory. The robot's end-effector Cartesian pose is 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 [20]. Microscopic images captured from the endomicroscopy system are stitched together in real-time, using pairwise registration of images frames by template matching via normalised cross-correlation. A position estimate obtained from the mosaicking registration process is fused with the pose estimation from the camera using a Kalman filter, to provide probe position information used for closed loop control during the mosaic scanning. The image mosaic is also fused with the reconstructed surface on-the-fly to provide both macro- and micro-views of the scanned region. By providing accurate and robust pose estimation, this visual servoing approach permits smooth and accurate probe scanning, allowing the formation of contiguous image mosaics.

B. Trajectory Planning Based on 3D Reconstruction

In this work, the scanning trajectory is planned to adapt to the 3D profile of the tissue surface. The Efficient LArge-scale Stereo (ELAS) [21] approach has been adopted for 3D surface reconstruction. To determine the normal to a point on the surface, the problem is approximated by the problem of estimating the normal of a plane tangent to the surface, similar to [22].

The scanning trajectory, which is a sequence of probe tip poses in the camera frame (T_C^D), is planned in two stages: the global trajectory planning, which moves the probe to the starting pose for scanning, and the local trajectory planning, which performs high resolution scanning from the starting pose. To generate the probe tip poses in the global trajectory, we interpolate between the current probe tip pose at a standby position, and the desired starting pose for scanning. More specifically, linear interpolation and spherical linear interpolation are used for translation and

estimate of the probe pose. Consider a discrete-time linear stationary signal model:

$$\mathbf{x}(k) = \mathbf{A}\mathbf{x}(k-1) + \mathbf{w}(k) \quad (6)$$

where $\mathbf{x}(k) \in \mathbb{R}^n$ is the state vector. $\mathbf{w}(k) \in \mathbb{R}^n$ is a sequence of process noise which is assumed to be drawn from a zero mean multivariate normal distribution whose covariance matrix is denoted by $\mathbf{Q}(k)$. In this work, the state vector $\mathbf{x}(k) = [t_x, t_y, t_z, \dot{t}_x, \dot{t}_y, \dot{t}_z, \alpha, \beta, \gamma, \dot{\alpha}, \dot{\beta}, \dot{\gamma}]^T$ represents the position and orientation (in Euler angle form) of the probe together with their first derivatives (translational and angular velocity). The state transition matrix is defined as: $\mathbf{A} = \begin{pmatrix} \mathbf{A}_T & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_T \end{pmatrix}$ in which $\mathbf{A}_T = \begin{pmatrix} \mathbf{I}_3 & \Delta t \cdot \mathbf{I}_3 \\ \mathbf{0} & \mathbf{I}_3 \end{pmatrix}$. The time interval between samples is denoted as Δt .

The relationship between the measurements from sensors and the states is defined as:

$$\mathbf{y}(k) = \mathbf{C}\mathbf{x}(k) + \mathbf{v}(k) \quad (7)$$

where $\mathbf{y}(k) \in \mathbb{R}^l$ is the measurement vector, and $\mathbf{v}(k) \in \mathbb{R}^l$ is a sequence of observation noise which is assumed to be zero mean Gaussian white noise with covariance $\mathbf{R}(k)$. Since we have two sensors, the measurement matrix C is defined separately for camera and pCLE:

$$\begin{cases} \mathbf{C}_c = \begin{pmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \end{pmatrix}, \\ \mathbf{C}_p = \begin{pmatrix} \mathbf{I}_2 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}. \end{cases} \quad (8)$$

The error covariance of states is updated in the prediction stage as:

$$\mathbf{P}(k|k-1) = \mathbf{A}\mathbf{P}(k-1|k-1)\mathbf{A}^T + \mathbf{Q}(k-1). \quad (9)$$

The Kalman gain for the camera and endomicroscope is given by:

$$\begin{cases} \mathbf{K}_c(k) = \mathbf{P}(k|k-1)\mathbf{C}_c^T(\mathbf{C}_c\mathbf{P}(k|k-1)\mathbf{C}_c^T + \mathbf{R}_c(k))^{-1}, \\ \mathbf{K}_p(k) = \mathbf{P}(k|k-1)\mathbf{C}_p^T(\mathbf{C}_p\mathbf{P}(k|k-1)\mathbf{C}_p^T + \mathbf{R}_p(k))^{-1}, \end{cases} \quad (10)$$

The measurement residual is given by: $\mathbf{r}(k) = \mathbf{y}(k) - \mathbf{C}\mathbf{x}(k-1)$. The correction is then performed using:

$$\begin{cases} \mathbf{x}(k) = \mathbf{x}(k|k-1) + \frac{1}{2}(\mathbf{K}_c(k)\mathbf{r}_c(k) + \mathbf{K}_p(k)\mathbf{r}_p(k)), \\ \mathbf{P}(k) = (\mathbf{I} - \frac{1}{2}(\mathbf{K}_c(k)\mathbf{C}_c + \mathbf{K}_p(k)\mathbf{C}_p))\mathbf{P}(k|k-1). \end{cases} \quad (11)$$

During the execution of global trajectory to the scanning region, there is no endomicroscopy image information (the probe is not in contact with the tissue) and thus only the camera data is fed into the Kalman filter for that case.

D. Closed Loop Scanning

In our framework, we compute the hand-tool transformation \mathbf{T}_E^M between the end-effector and marker on the adaptor via a standard hand-eye calibration approach [27]. The advantage of using the hand-tool transformation is that it will not be affected when the robot and camera are repositioned, making it more practical during surgical procedures. Although \mathbf{T}_E^M would change slightly if the adapter is released and re-grasped, we found that the resulting small

error (a few millimetres) in the hand-tool transformation is acceptable for the visual servoing control.

Given a current estimated probe pose \mathbf{T}_C^D and a desired pose $\mathbf{T}_C^{D^*}$ in the camera frame (in the planned trajectory), a new command pose can be calculated for the robot to minimise the difference. The relative probe pose, denoted as $\mathbf{T}_D^{D^*}$, is given by:

$$\mathbf{T}_D^{D^*} = (\mathbf{T}_C^D)^{-1} \cdot \mathbf{T}_C^{D^*} \quad (12)$$

where \mathbf{T}_C^D and $\mathbf{T}_C^{D^*}$ are obtained from the Kalman filter and the planned trajectory, respectively. Since the local trajectory is planned on the mosaic image coordinate, we need transformation $\mathbf{T}_D^I = \begin{pmatrix} \mathbf{R}_D^I & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \end{pmatrix}$ to convert the trajectory from 2D to 3D. \mathbf{R}_D^I is essentially a 2D rotation between the probe and the mosaic image, which can be calibrated by driving the robot with a horizontal line scan.

With this, we can calculate the robot command \mathbf{T}_B^E that minimises the relative probe pose:

$$\mathbf{T}_B^{E^*} = \mathbf{T}_B^E \cdot \mathbf{T}_E^M \cdot \mathbf{T}_M^D \cdot \mathbf{T}_D^{D^*} \cdot (\mathbf{T}_{M^*}^{D^*})^{-1} \cdot (\mathbf{T}_{E^*}^{M^*})^{-1} \quad (13)$$

The vision-based correction is performed on every pose along the planned trajectory. To produce a smooth scanning motion, the trajectory between two adjacent poses is interpolated and executed using the robot kinematic model.

E. Endoscopic 3D Fusion and Visualisation

Using the 3D reconstruction of the scanning region and the probe pose in camera coordinates, we are able to register the mosaic images with the tissue surface. This enables us to create a multi-scale map that links macro- and micro-views, which are obtained from the stereo camera and the endomicroscope, respectively. For endomicroscopy image mosaicing, we have used an approach similar to standard real-time techniques discussed in [28], using normalised cross-correlation to estimate the relative shift between each pair of consecutive frames. To avoid overwriting information from previous frames, and minimise boundary effects where images overlap, blending is applied. The mosaic image is then converted to a 3D mosaic point cloud in which each point represents a pixel in the mosaic image.

In order to register the mosaic point cloud with the surface, we need to find \mathbf{T}_I^C . As illustrated in 2 it can be calculated as:

$$\mathbf{T}_C^I = \mathbf{T}_C^{M_0} \cdot \mathbf{T}_{M_0}^D \cdot \mathbf{T}_D^I, \quad (14)$$

where M_0 represents the coordinate frame of the marker at the beginning of the local scan. With \mathbf{T}_I^C for each scanning region, we are able to register mosaics from multiple region scans onto the same 3D surface. The 3D mosaic point cloud is updated on-the-fly when the microscopic images are mosaicked. An illustration of this multi-scale 3D fusion can be seen in Fig. 7.

In this paper, we provide three ways to visualise the scanning and mosaicing on-the-fly: (1) An augmented view from the camera that allows showing both the robot and target tissue by overlaying the endomicroscope images onto

TABLE I
STEREO RECONSTRUCTION ACCURACY

		ELAS [21]	COCV [29]
Cardiac1	MAE (mm)	0.89 ± 0.70	1.24 ± 0.89
	RMS (mm)	1.31 ± 0.98	1.85 ± 0.82
Cardiac2	MAE (mm)	1.21 ± 1.56	1.47 ± 1.23
	RMS (mm)	1.77 ± 2.16	2.66 ± 1.47

TABLE II
CAMERA VISUAL SERVOING ERROR (MEAN ± STD)

Trial	Translation Error (mm)		Rotation Error (degrees)	
	VS	KM	VS	KM
Trial 1	0.238 ± 0.06	0.863 ± 0.26	0.89 ± 0.55	4.51 ± 1.45
Trial 2	0.222 ± 0.08	0.360 ± 0.21	1.21 ± 0.53	3.57 ± 0.79
Trial 3	0.189 ± 0.08	0.507 ± 0.16	1.19 ± 0.55	1.96 ± 1.00
Trial 4	0.185 ± 0.07	0.588 ± 0.37	1.54 ± 0.46	5.54 ± 1.27
Trial 5	0.172 ± 0.06	0.462 ± 0.27	1.44 ± 0.55	5.62 ± 1.11
Trial 6	0.226 ± 0.07	0.915 ± 0.40	1.57 ± 0.49	5.22 ± 1.08
Trial 7	0.247 ± 0.06	0.909 ± 0.37	1.11 ± 0.48	4.05 ± 0.87
Trial 8	0.303 ± 0.09	0.542 ± 0.17	1.25 ± 0.50	0.79 ± 0.52
Trial 9	0.165 ± 0.10	0.484 ± 0.13	0.84 ± 0.45	3.39 ± 0.80
Trial 10	0.164 ± 0.07	0.666 ± 0.13	1.21 ± 0.57	1.44 ± 1.04
Total	0.211 ± 0.07	0.630 ± 0.25	1.226 ± 0.51	3.61 ± 0.99

the tracking marker on the adaptor, as shown in Fig. 5(b) and (d); (2) A real-time, continuously updated 2D mosaic map; (3) An online 3D fusion of mosaics onto the reconstructed surface. These methods provide a clear and intuitive visualisation for the surgeon, aimed at assisting surgical planning.

III. EXPERIMENTS AND RESULTS

A. Stereo Validation

For assessing the accuracy of ELAS stereo reconstruction, we have provided a comparison of the method with the state-of-the-art [29] on the Hamlyn cardiac datasets (<http://hamlyn.doc.ic.ac.uk/vision/>). The results are provided in Table I, and show that the ELAS method provides competitive accuracies.

B. Hand-eye calibration accuracy

Due to the kinematic error of a tendon driven arm such as the da Vinci robot, the hand-eye calibration would only be accurate within a space where the calibration data has been taken. Given the calibrated hand-tool transformation \mathbf{T}_E^M and an initial marker pose $\mathbf{T}_C^{M_0}$, we can predict a marker pose using robot forward kinematics:

$$\hat{\mathbf{T}}_C^M = \mathbf{T}_C^{M_0} \cdot \mathbf{T}_{M_0}^{E_0} \cdot \mathbf{T}_{E_0}^B \cdot \mathbf{T}_B^E \cdot \mathbf{T}_E^M. \quad (15)$$

The ground truth of the marker pose \mathbf{T}_C^M was obtained from manual annotation. We calculate the error between the ground truth and the pose estimated by the hand-eye model as described by Eq. 15. The error pose is computed as: $\Delta\mathbf{T}_M^M = (\hat{\mathbf{T}}_C^M)^{-1} \cdot \mathbf{T}_C^M$. Starting from the initial pose $\mathbf{T}_C^{M_0}$, as the marker moves within the workspace for the scanning task, we calculate two values: translational error and rotational error. The translational error is the Euclidean norm of the translational part of $\Delta\mathbf{T}_M^M$ while the rotational error is the mean of the Euler angle of the rotational part.

TABLE III
ERROR OF HAND-EYE CALIBRATION

	Translation (mm)	Rotation (deg)
MAE	1.33 ± 0.78	0.64 ± 0.37
RMSE	1.54	0.74

As presented in III, the translational and rotational errors are approximately 1.5 mm and 0.7 degrees respectively, which is sufficient to support sub-millimetre-scale accuracy for the visual servoing.

C. Accuracy of Visual Servoing

To quantitatively validate the camera visual servoing approach, we used the Optotrak Certus system (Northern Digital Inc, Canada) which achieves 0.1 mm accuracy. Optical sensors were attached to a custom probe adapter, so that a rigid transformation could be found between the optical sensors and the marker frame. In this study, we ran ten trials for validation. Each trial consisted of two runs that included trajectory-following with Visual Servoing (VS) and Kinematic Motion (KM) only. The planned trajectories, which are defined in the camera coordinates, were used as the ground truth, and both the optical tracked results of two runs were transformed into the camera coordinate for comparison. To this end, we assume the starting pose of two runs should align with the first pose in the ground truth trajectory. For each pair of measured and ground truth poses at the same time instance, we calculate the pose difference as the error. For translational errors, we compute the Euclidean distance, and for rotational errors, we use the average of the Euler angles from the x, y and z axes. The measured translational errors and rotational errors are summarised in Table II. Both translational and rotational accuracies using the camera visual servoing are improved over kinematic-only control. The motion using the camera visual servoing was able to achieve on average approximately 0.21 mm for the translational error and 1.23 degrees for the rotational error.

Since the camera-based visual servoing is used to move the probe to an initial position for scanning, we can qualitatively evaluate the repeatability by comparing the differences among multiple mosaics acquired at the same nominal position. We ran three trials, in which the probe was moved to the same position before scanning and mosaicing over a printed grid phantom. As shown in Fig. 4, the mosaic images from different trials (identified by the colour) are well-aligned at their centres, both horizontally and vertically, and the camera visual servoing provides good repeatability with only about 50 to 150 μm deviation. This shows good correspondence with the quantitative accuracy estimated above (0.211 mm).

The mosaic images in the Fig. 4, we also provide us with the accuracy of camera and endomicroscope combined servoing. This is performed by comparing the original layout of the grid pattern with the obtained mosaicked grid, and in our experiment, the error is around 60 μm including the uncertainty introduced by the laser printer.

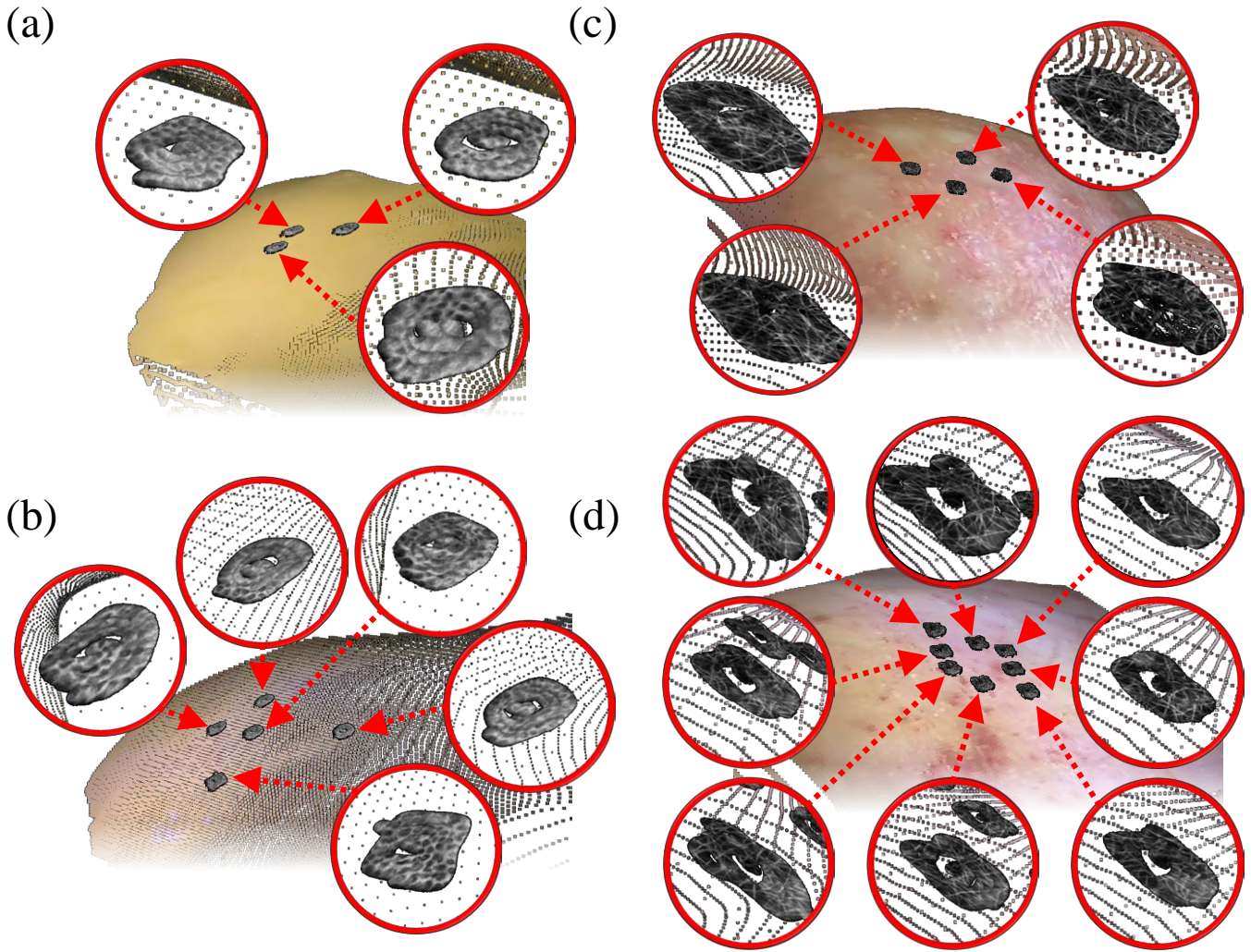


Fig. 7. Exemplar 3D fusion results. (a) and (b) Online multi-region fusion results from the *ex vivo* gastric mucosa experiment; (c) and (d) Online multi-region fusion results from phantom experiment.

D. Phantom and Ex Vivo Experiment

The experiment steps can be summarised as follow:

- 1) A stereo reconstruction of current surgical scene is generated.
- 2) A user places the probe at a standby location.
- 3) A user selects a target scanning region with the global and local trajectories being generated accordingly.
- 4) The robot moves the probe along the global trajectory with servoing from the laparoscopic camera images.
- 5) Probe arrives at the starting position for the mosaic scan over the desired region; robot drives probe to follow local trajectory using both microscopic and laparoscopic images for servoing.
- 6) After finishing the scan, the probe returns to the standby location, ready to scan the next region.

To validate the approach, we have performed autonomous scans on a custom-made PVA cryogel kidney phantom. In order to produce high quality microscopic images, lens tissue paper stained with the topical fluorescent contrast agent acriflavine, was placed on top of the surface. A snapshot

of the experiment is provided in Fig. 5(a-b). Scanning tasks were then performed on the phantom surface using both spiral and raster trajectories. An example of the 2D mosaics obtained is shown in Fig. 6. The actual trajectory line drawn on the mosaic image shows that the robot suffers from the raster motion due to backlash, as the raster pattern consists of abrupt changes of motion direction. The spiral pattern is considered to be superior as the motion direction changes gradually.

The 3D fusion results are presented in Fig. 7 which demonstrates that our approach is able to perform multi-region scanning. These results have shown our proposed framework is able to provide continuous image mosaicing, benefiting from the smooth and accurate motion generated by the visual servoing approach. Experiments were also performed using *ex vivo* animal tissue (see Fig. 5(c-d)). The gastric mucosa of a porcine stomach tissue sample was again stained with the topical acriflavine contrast agent. Examples of the 3D fusion results are provided in Fig. 7. Both the phantom and *ex vivo* experiments have demonstrated

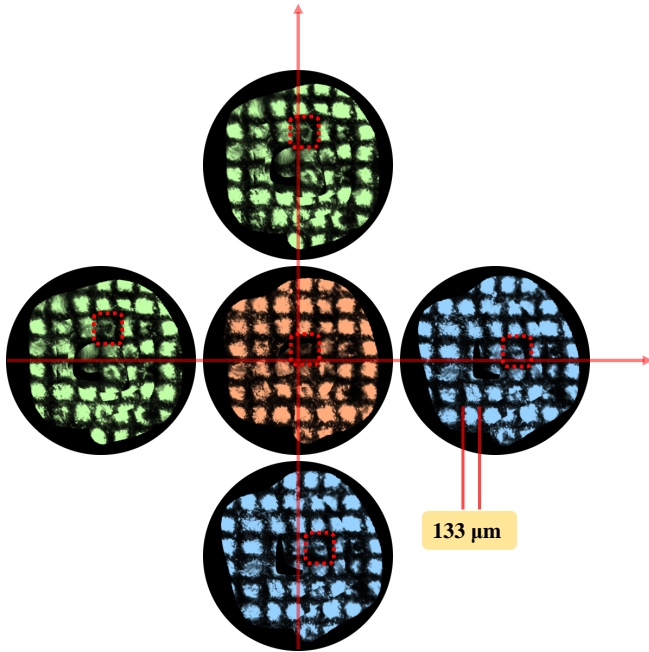


Fig. 4. Demonstration of mosaic results on a printed grid phantom with black holes (outlined by the dashed rectangle). Three trials are identified by three different colours and a red guide axis is placed for assisting comparison.

that our framework is capable of performing autonomous endomicroscopy scanning and generating high quality image mosaics. Again, due to the advantages of global surface reconstruction and visual servo control, we are able to scan multiple regions on the same surface. This also provides a global-local mapping where user can not only see each scanning location from macro-scale but also see a zoom-in mosaic image.

E. Discussion

Whilst one might expect that the accuracy of stereo reconstruction may be insufficient for an endomicroscopy scanning task, we found in the experiments that most of the scanning tasks can be completed successfully. However, it was observed during the stereo validation that the recovered surface tends to have a constant offset despite good shape matching to the ground truth. To address this erroneous offset, we adjusted the probe tip to marker transformation \mathbf{T}_D^M to ensure continuous contact with the surface. This solution is not ideal and would cause problems when the reconstruction error becomes too large ($> 2.0mm$). This is because the large error would cause the probe to either lose contact with the surface or apply too much pressure, leading to excessive deformation. In addition, we also observed that the image quality tends to decrease as the contact force increases. A comprehensive solution to this problem, would be to perform additional visual servoing based on the quality of microscopic images. With this, the image quality could then be used to detect loss of contact or excessive pressure, as well as to adjust the robot position accordingly.

The proposed framework presents good repeatability and

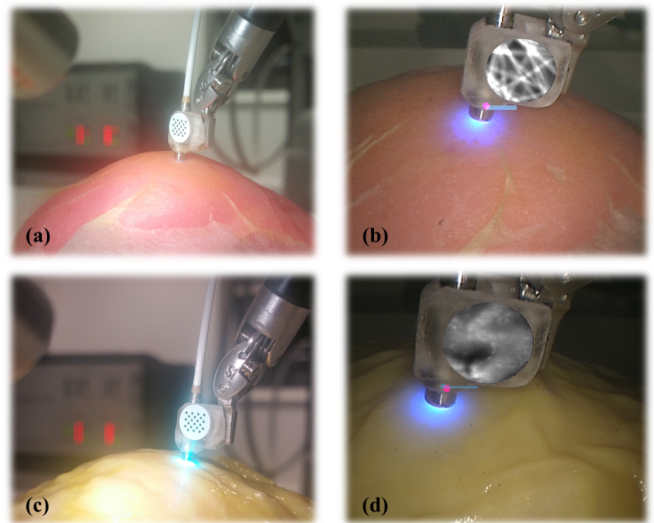


Fig. 5. Snapshots of the phantom and *ex vivo* experiments. (a) and (c) Experimental setups; (b) and (d) A snapshot of the surgical view captured by the laparoscope. The KeyDot[®] marker was overlaid with the streaming microscopic images for augmented reality.

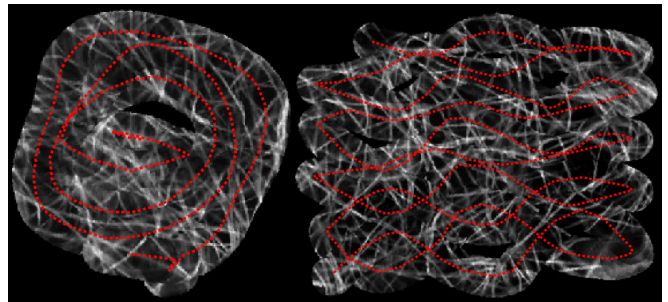


Fig. 6. Comparison of spiral and raster trajectories. The actual trajectories measured on the mosaic images are shown in red dotted line.

accuracy, where all trials present consistently low errors, sufficient to allow endomicroscopy scanning tasks to be completed. The mosaicing results from both phantom and *ex vivo* experiments, shown in Fig. 7, have demonstrated that the system is capable of scanning multiple regions while obtaining good quality mosaicing results. The mosaics generated in this work are a significant improvement over the basic linear mosaics produced by manually operated endomicroscopes, suggesting improved clinical value for intraoperative pathological analysis and surgical planning.

IV. CONCLUSIONS

In this paper, an autonomous framework using the da Vinci[®] robot is proposed for endomicroscopic mosaicing and online fusion of 3D reconstructed macroscopic images with microscopic images. We have used both laparoscopic camera and endomicroscopy images to close the loop to achieve continuous and smooth scanning. In addition, the proposed framework provides a comprehensive visualisation scheme for users, which includes: (1) an overlay of microscopic images into the surgical view for augmented reality; (2) an online updating 2D mosaic image; and (3) 3D fusion of

the mosaic images with the tissue surface on-the-fly. These visualisation options have a range of potential benefits for intraoperative tissue analysis and surgical planning, which will be explored in future work. The proposed framework has been tested on phantoms and *ex vivo* tissue, and quantitative results obtained from an external tracking system have shown that our framework can achieve accuracy better than the field-of-view of the endomicroscope, thus permitting smooth and accurate microscopic scanning.

ACKNOWLEDGMENT

The authors would like to thank Simon DiMaio from Intuitive Surgical Inc for providing the CAD model of the large needle driver, Konrad Leibrandt for his assistance with the software, and Stamatia Giannarou for valuable discussions.

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," *The International Journal of Medical Robotics and Computer Assisted Surgery*, vol. 7, no. 4, pp. 375–392, 2011.
- [2] S. Billings, N. Deshmukh, H. J. Kang, R. Taylor, and E. M. Boctor, "System for robot-assisted real-time laparoscopic ultrasound elastography," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2012, pp. 83 161W–83 161W.
- [3] A. Ruzskowski, O. Mohareri, S. Lichtenstein, R. Cook, and S. Salcudean, "On the feasibility of heart motion compensation on the davinci® surgical robot for coronary artery bypass surgery: Implementation and user studies," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015, pp. 4432–4439.
- [4] N. Padoy and G. D. Hager, "Human-machine collaborative surgery using learned models," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 5285–5292.
- [5] A. Murali, S. Sen, B. Kehoe, A. Garg, S. McFarland, S. Patil, W. D. Boyd, S. Lim, P. Abbeel, and K. Goldberg, "Learning by observation for surgical subtasks: Multilateral cutting of 3d viscoelastic and 2d orthotropic tissue phantoms," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, May 2015, pp. 1202–1209.
- [6] P. Pratt, A. Hughes-Hallett, L. Zhang, N. Patel, E. Mayer, A. Darzi, and G.-Z. Yang, "Autonomous ultrasound-guided tissue dissection," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*. Springer, 2015, pp. 249–257.
- [7] D. Hu, Y. Gong, B. Hannaford, and E. J. Seibel, "Semi-autonomous simulated brain tumor ablation with ravenii surgical robot using behavior tree," in *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015, pp. 3868–3875.
- [8] M. B. Wallace, P. Sharma, C. Lightdale, H. Wolfson, 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.
- [9] 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.
- [10] M. Ye, S. Giannarou, N. Patel, J. Teare, and G.-Z. Yang, "Pathological site retargeting under tissue deformation using geometrical association and tracking," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*. Springer, 2013, pp. 67–74.
- [11] W. T. Latt, R. Newton, M. Visentini-Scarzanella, C. Payne, D. Noonan, J. Shang, and G.-Z. Yang, "A hand-held instrument to maintain steady tissue contact during probe-based confocal laser endomicroscopy," *Biomedical Engineering, IEEE Transactions on*, vol. 58, no. 9, pp. 2694–2703, Sept 2011.
- [12] B. Rosa, M. S. Erden, T. Vercauteren, B. Herman, J. Szewczyk, and G. Morel, "Building large mosaics of confocal edomicroscopic images using visual servoing," *Biomedical Engineering, IEEE Transactions on*, vol. 60, no. 4, pp. 1041–1049, 2013.
- [13] 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 *Robotics and Automation (ICRA), 2015 IEEE International Conference on*. IEEE, 2015, pp. 1196–1201.
- [14] S. Zuo, M. Hughes, C. Seneci, T. Chang, and G.-Z. Yang, "Towards intraoperative breast endomicroscopy with a novel surface scanning device," *Biomedical Engineering, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015.
- [15] A. Patsias, L. Giraldez-Rodriguez, A. Polydorides, R. Richards-Kortum, S. Anandasabapathy, T. Quang, A. Sikora, and B. Miles, "Feasibility of transoral robotic-assisted high resolution microendoscopic imaging of oropharyngeal squamous cell carcinoma," *Head & neck*, 2014.
- [16] P. Giataganas, M. Hughes, and G.-Z. Yang, "Force adaptive robotically assisted endomicroscopy for intraoperative tumour identification," *International journal of computer assisted radiology and surgery*, pp. 1–8, 2015.
- [17] P. Giataganas, C. Bergeles, P. Pratt, M. Hughes, A. Darzi, and G.-Z. Yang, "Intraoperative 3d fusion of microscopic and endoscopic images in transanal endoscopic microsurgery," in *The Hamlyn Symposium on Medical Robotics*, 2014, p. 35.
- [18] 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.
- [19] Z. Zhang, "A flexible new technique for camera calibration," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 11, pp. 1330–1334, Nov 2000.
- [20] 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 *Robotics and Automation (ICRA), 2014 IEEE International Conference on*. IEEE, 2014, pp. 6434–6439.
- [21] A. Geiger, M. Roser, and R. Urtasun, "Efficient large-scale stereo matching," in *Computer Vision–ACCV 2010*. Springer, 2011, pp. 25–38.
- [22] R. B. Rusu, "Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments," Ph.D. dissertation, Technische Universität München, 2009.
- [23] M. S. Erden, B. Rosa, J. Szewczyk, and G. Morel, "Understanding soft-tissue behavior for application to microlaparoscopic surface scan," *IEEE Transactions on Biomedical Engineering*, vol. 60, no. 4, pp. 1059–1068, April 2013.
- [24] P. Pratt, A. Di Marco, C. Payne, A. Darzi, and G.-Z. Yang, "Intraoperative ultrasound guidance for transanal endoscopic microsurgery," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*. Springer, 2012, pp. 463–470.
- [25] J.-Y. Bouguet, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," *Intel Corporation*, vol. 5, pp. 1–10, 2001.
- [26] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnnp: An accurate o (n) solution to the pnp problem," *International journal of computer vision*, vol. 81, no. 2, pp. 155–166, 2009.
- [27] R. Y. Tsai and R. K. Lenz, "A new technique for fully autonomous and efficient 3d robotics hand/eye calibration," *Robotics and Automation, IEEE Transactions on*, vol. 5, no. 3, pp. 345–358, 1989.
- [28] M. Hughes and G.-Z. Yang, "High speed, line-scanning, fiber bundle fluorescence confocal endomicroscopy for improved mosaicking," *Biomedical optics express*, vol. 6, no. 4, pp. 1241–1252, 2015.
- [29] P.-L. Chang, D. Stoyanov, A. J. Davison *et al.*, "Real-time dense stereo reconstruction using convex optimisation with a cost-volume for image-guided robotic surgery," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*. Springer, 2013, pp. 42–49.