A Framework for Sensorless and Autonomous Probe-Tissue Contact Management in Robotic Endomicroscopic Scanning

Rejin John Varghese, Pierre Berthet-Rayne, Petros Giataganas,
Valentina Vitiello, and Guang-Zhong Yang, *Fellow, IEEE*

***Abstract*— Advances in optical imaging, and probe-based Confocal Laser Endomicroscopy (pCLE) in particular, offer real-time cellular level information for *in-vivo* tissue characterization. However for large area coverage, the limited field-of-view necessitates the use of a technique known as mosaicking to generate usable information from the incoming image stream. Mosaicking also needs a continuous stream of good quality images, but this is challenging as the probe needs to be maintained within an optimal working range and the contact force controlled to minimize tissue deformation. Robotic manipulation presents a potential solution to these challenges, but the lack of haptic feedback in current surgical robot systems hinders the technology's clinical adoption. This paper proposes a sensorless alternative based on processing the incoming image stream and deriving a quantitative measure representative of the image quality. This measure is then used by a controller, designed using model-free reinforcement learning techniques, to maintain optimal contact autonomously. The developed controller has shown near real-time performance in overcoming typical loss-of-contact and excess-deformation scenarios experienced during endomicroscopy scanning procedures.**

I. Introduction

Early disease detection and timely treatment are key to tackling cancer, with early detection being the prerequisite to a good prognosis and timely treatment [1]. Histopathology analysis of biopsied tissue is still the gold-standard in tissue characterization, but it is time-consuming, operator-dependent, requires *ex-vivo* analysis of excised tissue, and cannot achieve results in real-time [2]. These limitations highlight the demand for alternative diagnostic modalities with real-time, *in-situ* tissue characterization capabilities. Endomicroscopy, an "optical-biopsy" based approach for visualization and diagnosis, has shown to be capable of continuous acquisition of cellular-level high-resolution histology-like images for *in-vivo* and *in-situ* tissue characterization [3].

But the technology has certain limitations that hinder translation into mainstream clinical practice. The most significant drawback is the limited Field-of-View (FOV), commonly only $0.24mm$ as a trade-off of the high-resolution cellular-level images generated by the system. Mosaicking algorithms have been used to compensate for the limited FOV [4]. These algorithms use a stream of good-quality microscopic images obtained by scanning over an area in a specific pattern, and stitching them together to obtain histology-like mosaics of larger areas. However, an important caveat to achieving the above is continuous good quality image acquisition. Acquiring a continuous stream of good quality images is challenging requiring the probe to be maintained within an optimal working distance, the range for which lies between $0 - 80\mu m$, based on the lens assembly at the tip of the fiber bundle [4]. Additionally, the tissue-contact force applied also needs to be regulated (in the order of $\sim 100mN$) to minimize tissue deformation which impacts image quality negatively.

Achieving these levels of accuracy consistently is challenging with manual scanning. Robotic platforms like the Da Vinci® system (Intuitive Surgical Inc., CA) with features including motion-scaling and tremor-compensation provide the necessary precision, accuracy and stability required for robot-assisted endomicroscopy scanning, but the lack of haptic feedback results in poor probe-tissue contact management [5]. Attempts at automation have been attempted with both active and passive force-control approaches. Probe-tissue contact force management has been attempted using active force-sensing instruments using load cells [6, 7], passive force-adaptive instruments with spring-based or pneumatic mechanisms [8, 9], distance-sensing [10] and stabilizing mechanisms [11]. But these solutions have certain drawbacks associated with them. The sensing modalities estimate image quality based on correlation with the sensor information and not by analyzing the image. This results in false-positives, *i.e.* a scenario where the sensing modality classifies the image to be acceptable, but the case is not, establishing the need for additional *in-vivo* calibration and baseline measurements leading to an increase in procedural time. Secondly, no information is available as soon as contact is lost with the tissue. Finally, additional sensing modalities to the tip increases probe footprint limiting applicability in small workspaces, negatively impacts probe dexterity/stiffness, and also leads to increased cost and reliability issues.

As stated earlier, the probe needs to be maintained within the optimal working range to ensure image acquisition of the highest quality. A drop in image sharpness is observed as the probe strays from the optimal working range. It has also been observed that this drop in sharpness occurs in both directions – *i.e.* if contact is lost or if excess deformation is induced. The drop in sharpness in the case of lost contact is more pronounced as compared to the drop when excess deformation is applied. This paper proposes a framework for probe-tissue contact control that does not use any external sensing modalities but leverages this observed phenomenon for contact-force management. A quantitative measure of image quality is extracted and is shown to be a good representation of the inherent image quality. This quantitative measure is then used to design a controller using model-free reinforcement learning (RL) techniques. Section II first discusses the methodology for extracting the quantitative measure from the

*All authors are with the Hamlyn Centre for Robotic Surgery, Institute of Global Health Innovation, Imperial College London, SW7 2AZ, London, UK (email: r.varghese15@imperial.ac.uk)

incoming image stream. The RL-based controller design methodology for the application is then presented. This is followed by a description of the controller implementation algorithm for maintaining optimal probe contact. Section III presents an overview of the experimental setup used and the experiments performed to derive, test and improve the controller. Section IV presents the outcome of the experiments and discusses the results. Section V concludes the paper highlighting the features and advantages of the proposed work, limitations of the system and also presents avenues for future work.

II. Methodology

A. Quantifying Image Quality

In order to manipulate the probe to maintain optimal contact, it is imperative to have a measurement signal. The loss of sharpness observed on straying from the optimal working range is quantified by processing the images using multiple blur-based imaging criteria which quantify the inherent sharpness/blur of an image. Well-known blur-based imaging criteria were analyzed on the basis of computational expense, the range of observed values, monotonicity, sensitivity to noise and ability to handle outliers. The analyzed metrics included Cumulative Probability for Blur Detection (CPBD) [12], Visual Blur Quality Evaluation Software (VBQUEST) - an implementation of the Marziliano Blurring Metric [13], an implementation of a blur metric by Crété-Roffet *et al.* in [14] by Do Quoc Bao, LPC-SI - an implementation of a Local Phase Coherence based Sharpness Index algorithm presented in [15], and image/information entropy. These metrics were studied based on a video of endomicroscopy images with 4133 frames, and the results revealing the best candidates for our application are summarized in Table 1. A graphical representation of the performance of each metric for 10 endomicroscopy images is also summarized in Fig. 5.

TABLE I
PERFORMANCE SUMMARY OF DIFFERENT IMAGING CRITERIA

	CPBD	VBQUEST	CRÉTÉ-ROFFET BM	LPC-SI	ENTROPY
RANGE OF VALUES	5	4	3	2	1
COMPUTATIONAL EXPENSE	1	3	4	2	5
MONOTONICITY	YES	YES	YES	NO	NO
NOISE SENSITIVITY	GOOD	GOOD	GOOD	GOOD	BAD
HANDLING OUTLIERS	3	5	4	1	1

Properties ranked between 1-5 with 1 being least acceptable and 5 most acceptable.

While CPBD, VBQUEST and Crété-Roffet blurring metric showed acceptable characteristics, no single metric satisfied all requirements. These three metrics were therefore blended together into a single classifier to compensate for the shortcomings of the individual metrics. This was done by using binary classification with supervised learning [16]. The images from a video dataset of previously performed scanning procedures were classified as acceptable and undesired by a human operator. A classifier with the normalized values of the chosen metrics was extracted using linear regression. The means associated with the good, bad and all data for the three metrics are: $\mu_g: (\mu_{1g}, \mu_{2g}, \mu_{3g})$, $\mu_b: (\mu_{1b}, \mu_{2b}, \mu_{3b})$ and $\mu_a: (\mu_{1a}, \mu_{2a}, \mu_{3a})$ respectively, where μ_{ig} and μ_{ib} are the means of metric i for the good and bad quality images respectively. The slope of the normal (m_N) to the classifier with the three metrics as axes can then be written as:

$$m_N = \left((\mu_{1g} - \mu_{1b}), (\mu_{2g} - \mu_{2b}), (\mu_{3g} - \mu_{3b})\right) \quad (1)$$

The slope of the classifier can then be written as:

$$m_C = -1/m_N \quad (2)$$

The slope of the classifier derived from (1) & (2) is equivalent to the weights associated with each metric in the classifier. A 2-D visualization of the above is shown in Fig. 1. The intercept of the classifier was further improved to maximize accuracy (using (3)) of the classification.

$$Accuracy = (TN + FP)/(TN + TP + FN + FP) \quad (3)$$

where TN, TP, FN & FP are true-negative, true-positive, false-negative and false-positive occurrences.

Positive classifier values are associated with images of desirable quality and vice-versa. This obtained classifier is the required quantitative measure representative of the inherent quality of the images. The workflow for the entire method is summarized as follows:

1) Extract frames from training video of scanning.
2) Compute imaging criteria metrics for extracted frames.
3) Acquire user feedback on image quality of frames.
4) Extract classifier of imaging criteria based on feedback.
5) Test and improve classifier with feedback from test set.
6) Optimize/threshold classifier intercept for imaging setup.

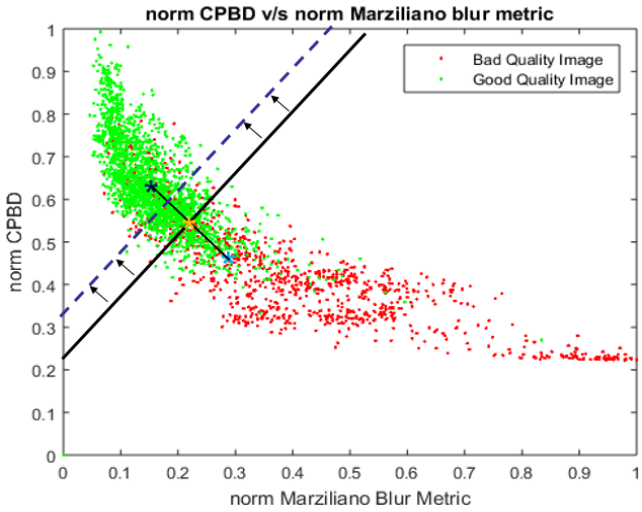

Fig. 1. Classifier derivation by image quality quantification. The black line is the classifier obtained from binary classification. The dashed blue line is obtained after optimizing the intercept for accuracy.

B. Controller Design

The classifier derived in Section II.A provides the measurement signal based on which the controller for maintaining optimal probe contact can be designed. Model-based controller design for autonomous control in medical robotics is extremely challenging. The non-linearities introduced by conventional surgical robotic systems in the form of tendon elasticity, backlash, friction, *etc.* along with the interaction of the robot with different tissues encountered *in-vivo* makes the extraction of accurate system dynamics very tedious, prone to error and a solution that is not generalizable. In this section, controller design is therefore attempted using model-free controller design methods. Reinforcement learning

(RL), which has been used extensively in literature to derive optimal control solutions to adaptive control problems has been used for developing the controller [17]. The system and environment dynamics are unknowns, and the best actions are learned by a reward-based feedback process.

A framework that is commonly used for formalizing sequential decision processes is the Markov Decision Process (MDP). The MDP is defined as: $MDP(X, U, P, R)$, where X is a set of states and U is a set of actions/commands. P is the set of transition probabilities $P: X \times U \times X \rightarrow [0,1]$, which describes for each state $x \in X$ and action $u \in U$, the conditional probability $P^{u}_{x,x'} = \Pr\{x'|x,u\}$, to transition to state $x' \in X$ given the agent is in state x and takes action u. The reward function $R: X \times U \times X \rightarrow R$, is the expected immediate reward $R^{u}_{x,x'}$, received after transition to state $x' \in X$, given the agent starts in state $x \in X$ and takes action $u \in U$ [17, 18]. The main goal with MDPs is to find the policy $\pi: X \times U \rightarrow [0,1]$ that gives for each state x and action u the conditional probability $\pi(x,u) = \Pr\{u|x\}$ of taking action u given the current state x, such that a particular objective is achieved [18].

The action-space in this application is the 1-D space defined by the axis normal to the scanning plane. A limit on the maximum action value is pre-set to address safety concerns during exploration. To reduce computational expense incurred during learning, the action-space is initialized with a vector of action values based on user experience. The definition of the state-space is critical for determining computational expense and global applicability of the learned controller. The three choices for the state-space were the 3-D Cartesian space, the 3-D imaging criteria space and 1-D classifier space. The 3-D imaging criteria space is made up of the three imaging criteria that were used in deriving the classifier and has the advantage of making the whole learning invariant in Cartesian space (a 2-D representation of this imaging criteria space is shown in Fig. 2(a)). The 1-D classifier space is a reduction of the 3-D imaging criteria space which exploits the strong correlation observed between the different metrics (leaving a significant part of the state-space unused as can be seen in Fig. 2(a)). The advantages of choosing the 1-D classifier space over the other two definitions are two-fold: 1). the state-space is Cartesian space invariant, allowing for the learned controller to be globally applicable irrespective of where the learning was done in Cartesian space, and 2). the state-space is extremely small as compared to both the Cartesian space and imaging criteria space.

The definition of the 1-D classifier space as the state-space is inspired by the approximate Q-learning method which uses feature-based representations [19]. Training can now be performed on a subset of the actual Cartesian state-space and then generalized to the entire state-space. The 1-D classifier state-space is discretized into a finite number of states as the Bellman equation in RL is defined for discrete state-spaces [17, 18]. The number of states the state-space is discretized into is based on balancing a trade-off between the robot's capabilities and the computational time needed for learning the controller. The transition probabilities are a representation of the system dynamics and are unknown. The immediate reward $R^{u}_{x,x'}$ is defined as follows:

$$R^{u}_{x,x'} = k.d^{c}_{x'} + P \tag{4}$$

The first component $k \times d^{c}_{x'}$, is based on the new state reached and the reward is proportional to the mean classifier value of this new state (refer Fig. 2(b)). The second component P, is based on how the state was reached. This is a fixed reward/penalty applied to downgrade actions that cause excess deformation and is identified by monitoring the classifier /state values attained while the robot is executing the action.

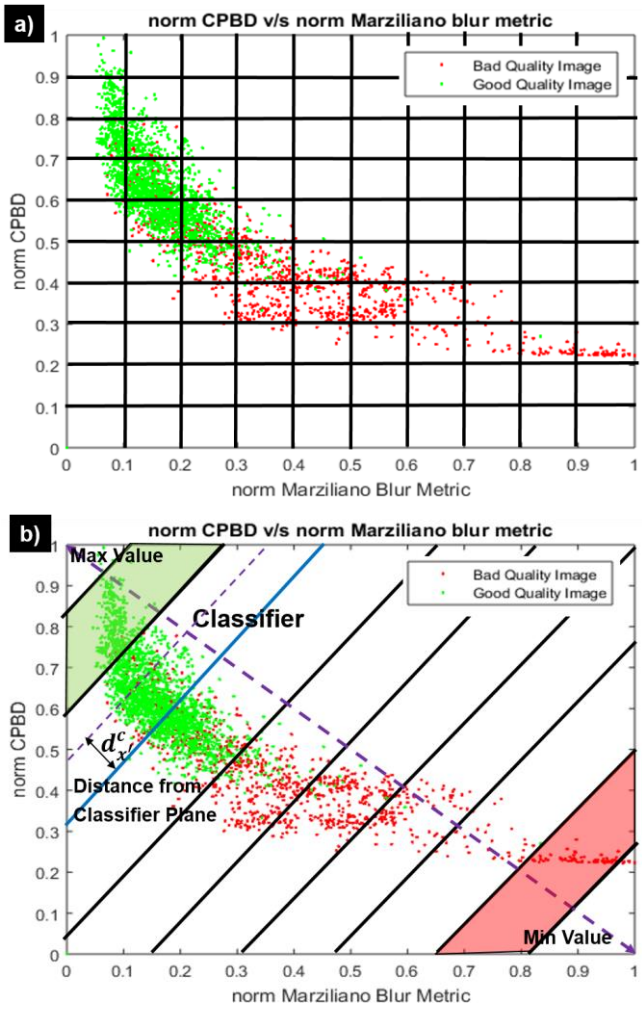

Fig. 2. State-space configuration choices: a) 2-D imaging criteria state-space, and b) 1-D classifier state-space

The Q-learning RL algorithm by Watkins [20], was used to learn the best actions associated with the MDP defined above. The update of Q-values can be written as:

$$Q_{k+1}(x,u) \leftarrow \sum_{x'} P^{u}_{xx'} \left[R^{u}_{xx'} + \gamma \max_{u'} Q_k(x',u') \right] \tag{5}$$

As the transition probabilities $P^{u}_{xx'}$ are unknown, an alternative approach of using running averages was implemented. The likelihood of occurrence of the state x' is a reflection of the unknown transition probabilities and gets accounted for over the course of the learning trials.

The Q-learning algorithm that was implemented in this work is as follows:

1. Perform a sample transition (x, u, x', r) and compute sample Q-value:

$$sample = r + \gamma \max_{u'} Q(x',u') \tag{6}$$

2. Update the corresponding Q-value $Q(x,u)$ associated with source state x and action u:

$$Q(x,u) \leftarrow (1-\alpha)Q(x,u) + \alpha(sample) \quad (7)$$

The optimal values and policy in terms of the Q-values can be written using the Bellman optimality equation [21]:

$$V_k^*(x) = \max_u Q_k^*(x,u) \quad (8)$$

$$u_k^* = \operatorname{argmax}_u Q_k^*(x,u) \quad (9)$$

γ is the discount rate for discounting future values. $\alpha \in [0,1]$ is the *learning* rate which blends in the latest sample value into the existing value of $Q(x,u)$ and gives the recently obtained samples more importance over the past values. α is reduced over time for convergence.

RL requires exploration of new actions so as to be able to learn the optimal policy and improve convergence. This was done by implementing the ε-greedy method [18]. Additionally, *random restarts* were implemented through random reset actions which allowed further exploration to be introduced to improve convergence rates. The RL implementation generates a look-up table associating each state (based on the degree of image sharpness/blur) with the optimal action to take from that state to reach the state with optimal image quality.

It is to be noted that the learning was done to learn actions only for the loss-of-contact scenario, as the range of classifier values encountered in the case of excess-deformation were very small and did not warrant learning with RL. Excess deformations result due to small movements made past optimal contact. To remove this excess deformation, a fixed step $(0.1mm)$ away from the tissue is taken repeatedly until the induced deformation is removed. Deformations experienced in typical scanning procedures were resolved in at most two steps.

C. Controller Implementation

Fig. 3 is a schematic of the controller implementation. The controller implementation uses two threads, one running the state/classifier acquisition and monitoring of these values, and the second thread for generating the control action (the two sections are partitioned by the dotted line in Fig. 3). To compute the states/classifier values, the incoming stream of image data is cropped to obtain a smaller region of interest (ROI). This was done to reduce the computational expense of the image processing without discarding useful information. The different imaging criteria perform computations on the incoming image stream by exploiting the parallel computing capabilities in MATLAB. Once the image processing is done, the classifier value and the MDP framework parameters presented earlier are computed. The aforementioned set of computations were successfully performed up to rates of 20 frames per second (fps) without frames being dropped. The framework is currently running at 10 fps. The above mentioned set of operations is represented by the left hand side off the dotted line in Fig. 3.

The previously computed state value is the control input for the controller and is transmitted via TCP/IP. When the computed state of the latest image falls below the minimum accepted state (T1) for a good quality image but is above a second threshold (T2), a check is performed (see Fig. 3). This check is to discern whether the drop in image quality was due to excess-deformation or loss-of-contact. This decides whether the probe needs to move towards or away from the tissue. The check is performed by moving towards the tissue by a small magnitude $(0.1mm)$. Post-movement, the image quality worsens in the case of excess-deformation and improves in the case of loss-of-contact. If the cause is loss-of-contact, the best action is taken. If the state value is less than T2, it definitely is a lost-contact scenario and in this case the best action is taken. In addition, while the robot is moving the interim state/classifier values during the movement is recorded in a vector. This is done to ascertain if a better quality image (optimal contact) was surpassed while the robot was moving, and if corrective action is needed. In the schematic in Fig. 3 the best action module is the look-up table obtained as a result of the RL performed in Section II.B.

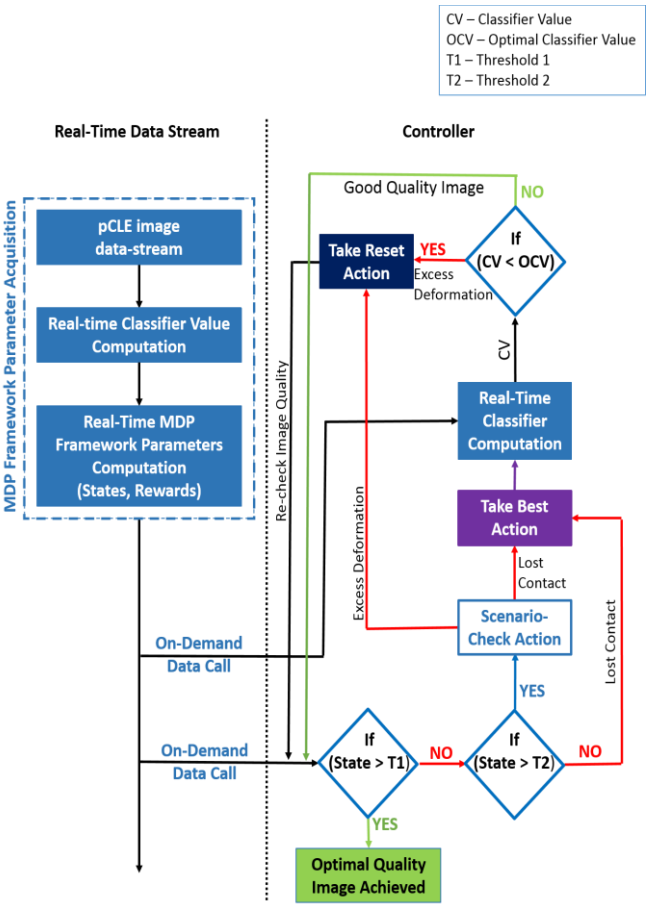

Fig. 3. Schematic of the implemented controller algorithm. Data is pulled on demand from the real-time data stream (left hand side) whenever the state value falls below T1. T1 is the minimum accepted state and T2 is the minimum state (T1>T2) at which a check is performed to see if the drop in image quality is due to loss-of-contact or excess-deformation.

III. Experimental Setup

The presented framework currently consists of the following hardware (shown in Fig. 4(a) and (b)):

- An endomicroscopy imaging system
- A surgical robot system
- A mechanical stage for manually regulating tissue-probe contact distance for learning, testing and validation.
- Load-cell with microcontroller (for force-measurements during validation experiments)

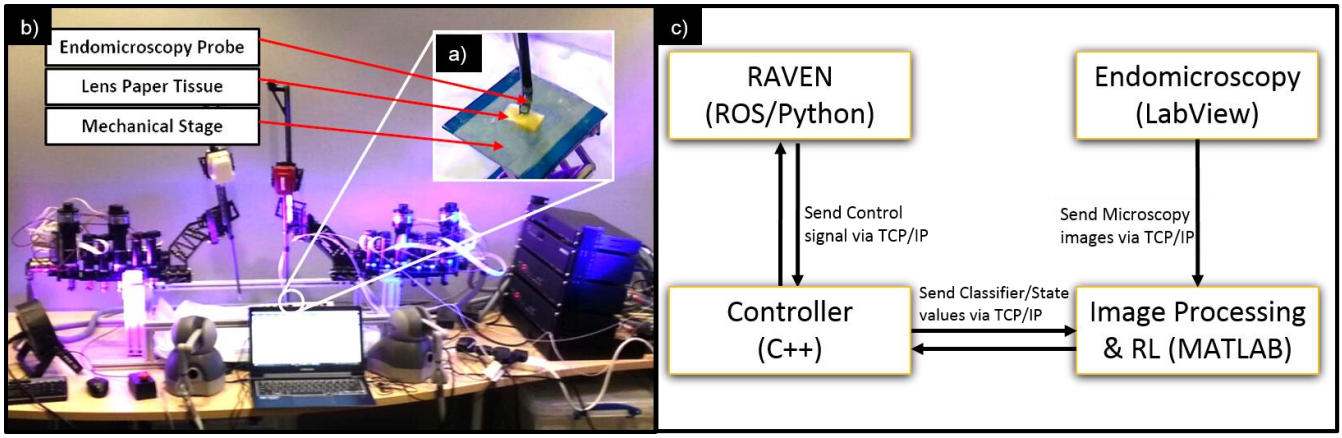

Fig. 4. Hardware & software setup: a) endomicroscopy probe scanning lens paper tissue (on a mechanical stage), b) hardware setup with Raven-II surgical robot, and c) information flow in software implementation.

A. *Endomicroscopy Imaging Systems*

In this work, two in-house probe-based endomicroscopy systems were used for image acquisition and validation of the framework. Both systems are fluorescence-based, as topical fluorescence contrast agent acriflavine is used to provide high-resolution microscopic images. The first system is a high-speed laser-scanning confocal endomicroscopy system developed by Hughes *et al.* [22] that is used for the training of the framework and provides optically-sectioned microscopic images at 120fps. The second in-house system, which is used for the validation of the framework, is a widefield optical microscopy system, similar to an established design by Pierce *et al.* [23], that provides non-optically-sectioned images at 30fps. With both systems, a commercial fibre bundle (Cellvizio™ UHD probe, Mauna Kea Technologies®, Paris, France) is used with approximately $30{,}000$ fibre cores coupled to a distal micro-objective and outer maximum diameter of $2.6mm$ at the distal tip. The resulting field-of-view is approximately $240\mu m$ and the resolution $2.4\mu m$.

B. *Surgical Robot System*

The Raven-II surgical robot system (Applied Dexterity, Seattle, WA) was used for manipulating the probe. The robotic platform has two 3-DOF (degrees-of-freedom) spherical positioning mechanisms capable of attaching interchangeable 4-DOF instruments [24]. The platform has been augmented with a custom controller previously developed in [25] and runs in parallel with the controller presented in this paper.

C. *Force-Sensing for Validation Experiments*

The load-cell and accompanying micro-controller (Cypress Semiconductor, CA) were used solely for validation of the proposed controller. It was used to ascertain: 1) no excess deformation was induced after the control action was taken, and 2) confirm that in case excess deformation was induced during scanning, the controller was able to recover. The load cell used had a range of $0-500g$. A linear fitting was derived to convert generated values to force in millinewtons (mN). A 4^{th} order low-pass digital Butterworth filter was applied to the signal to remove noise while maintaining adequate real-time performance.

All experiments were performed using lens paper tissue topically stained with fluorescent contrast agent acriflavine. The principle for the robot axis to move towards/away from the scanning plane is based on either loss-of-contact or excess-deformation. This cannot be achieved by lining a curved surface with lens paper, as deformation is not developed in this scenario. As a workaround, experiments for learning, improving and validating the controller were carried out by using a mechanical stage. Very small increments of positive or negative movements could be executed and this proved extremely useful for optimizing the controller.

The software components of the framework include:

- A LabView program which processes the images from the endomicroscopy systems and sends them via TCP/IP.
- A MATLAB script that runs image processing, and classifier/state value computation and transmission to the controller via TCP/IP during controller implementation. During learning, the image processing and RL are both implemented in MATLAB and control actions are directly transmitted to the Raven II robot.
- The developed controller programmed in C++ runs on a separate thread along with the custom controller. The custom controller receives commands and transmits it to the Raven-II robot.

The information flow between the different software components of the framework is shown in Fig. 4(c). For the learning experiments, the stage is mechanically moved a specific amount to achieve a certain state value. The RL framework then takes an action based on the ε-greedy method. The learning algorithm then updates the Q-values based on RL algorithm presented earlier. As mentioned earlier, providing initial values based on user-experience reduces the time for learning the best actions. The action-space for individual states was set as: $[bestAction-0.1mm, bestAction+0.1mm]$, in steps of $0.01mm$, during learning. This learning is done autonomously, assuming the initial set of best actions provided are validated as acceptable. When no user-experience was available, the learning of best actions was done hierarchically in two steps (broad-step and fine-step learning) to reduce learning times. For the validation experiments, a dashboard was created to monitor the performance of the controller.

IV. Results & Discussion

The two main objectives of this work were:

- To establish that the inherent information in the image stream from pCLE can be exploited to make it a self-sufficient sensing modality for closed-loop control.
- To use the generated measurement signal to design a controller that could successfully identify and overcome the two probe-tissue interaction scenarios: loss-of-contact and excess-deformation.

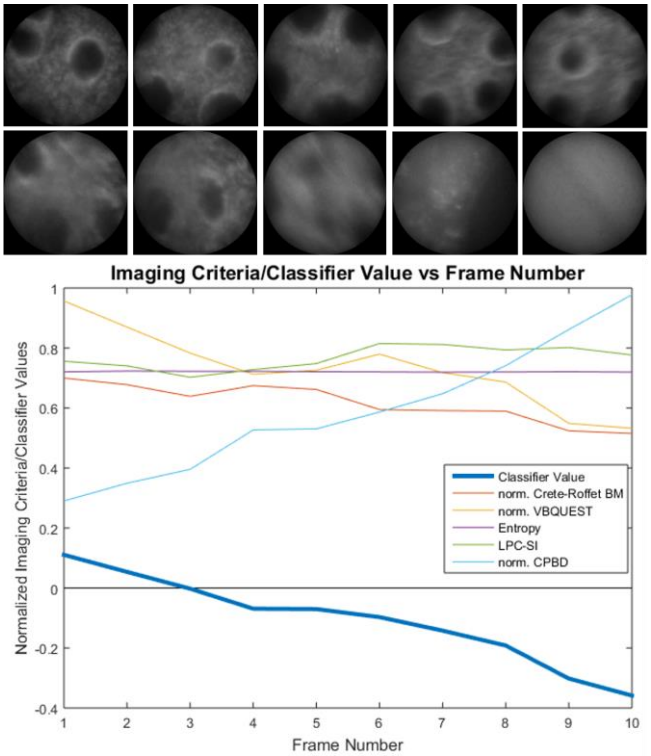

Fig. 5. Classifier performance for image classification with endomicroscopy images obtained from colon tissue samples. The classifier (bold blue line) accurately classifies the first 3 images as being good and the rest of undesirable quality. The other metrics (names in legend) are also plotted for comparing performance.

Fig. 5 presents the performance of the derived classifier in classifying the 10 reference images obtained from endomicroscopy of colon tissue. The classifier correctly classifies the first three images as acceptable and the rest seven of undesirable quality. A demonstration of the classifier's performance during real-time scanning operations can be found in the accompanying videos. The framework was able to process at $20fps$ which ensures high overlap (at recommended scanning velocities of $0.5mm/sec$ [26]) between consecutive images for good quality mosaicking. Fig. 5 also gives a comparison of the performance of the derived classifier against different blur metrics, and exhibits better performance than each metric. In the following sections, the controller performance is evaluated assuming the classifier to be an accurate representation of the image quality. The results of the controller performance in the loss-of-contact case and excess-deformation case are presented below. Fig. 6 and Fig. 7 are snapshots of the rolling window for load cell and classifier measurements from the framework dashboard. The first graph monitors load cell measurements measured in millinewtons (mN), and the second graph the classifier value. The experiments were performed using widefield microscopy instead of the pCLE system, while using the classifier derived from images obtained from pCLE to demonstrate the framework's ability to handle different imaging systems with minimal intervention. The threshold for the classifier was lowered from 0.0 to -0.13 as the inherent quality of images from the widefield microscopy system is lower compared to the pCLE system. It should be noted that the widefield microscopy system is less forgiving to the controller performance when compared to the pCLE system as there is no optical sectioning.

A. Controller Performance in Loss-of-Contact Case

Fig. 6 was obtained while testing the performance of the controller in the loss-of-contact scenario. The performance of the controller was analyzed in two scenarios: small and large loss-of-contact cases. The small loss-of-contact case would be the most predominantly encountered scenario during scanning operations. This scenario would be incurred when a negative gradient between the probe and tissue scanning plane would be experienced. This is an appropriate assumption as scanning velocities are recommended to not exceed $0.5mm/sec$ for good quality mosaicking based on the system's frame rate [26]. Also, the slopes encountered *in-vivo* while scanning tissue are not extremely high so as to cause a sudden large loss-of-contact. In Fig. 6, the small loss-of-contact scenario is tested at instance (C) and (G), and the controller successfully recovers from the lost contact and regains an acceptable image quality. These loss-of-contact scenarios were not induced intentionally by mechanical manipulation of the stage but are encountered occasionally during scanning operations. This is mostly due to the probe encountering a micro-fold in the tissue paper or a sudden reduction in contrast agent concentration on the lens paper. The image classifier value recovers above the preset threshold at instances (D) and (H) respectively. The actions taken by the robot in this case is extremely small (in the range of $\sim 0.1mm$) and is usually not discernible to the human eye. From the load cell measurements, it can be seen that there is no significant increase in measured contact force from the baseline measurements after contact has recovered. This indicates that no excess deformation was induced after the controller had taken the action.

From Fig. 6, it can be seen that in the case of the large loss-of-contact, the response time and frequency of the controller is not fast enough to compensate for physiological motion. This would be an area for improvement in the future. The highlight of the controller performance, however, is the one-step immediate recovery in the classifier value from the sub-optimal value to go above the threshold, once the action is taken. This is attributed to the actions learned during RL. It should be noted that this immediate recovery is limited to cases where the action-value required is less than the maximum limit imposed on the action-space in the MDP framework. The classifier value that the loss-of-contact was reduced to is different in both instances (A) and (E), *i.e.* different action values were applied. In both small and large

loss-of-contact scenarios, there is no significant increase in the measured contact force (and by extension no induced deformation) from the load cell measurements even though the action values are significantly larger. The contact force management is a result of the definition of the reward function during RL which rewards/penalizes based not only on what state is obtained but also on how that state was obtained. The gradual local rise and drop observed in classifier value between instances of lost-contact and excess-deformation in Fig. 6 and Fig. 7 respectively is due to the acquisition of relatively better/worse quality images while scanning. This can be attributed to the locally varying contrast agent concentrations on the tissue paper, folds in the tissue paper *etc.* causing variations in classifier values. These observed fluctuations are not due to probe manipulations. Scanning is currently stopped intermittently when the state value falls below the minimum accepted state.

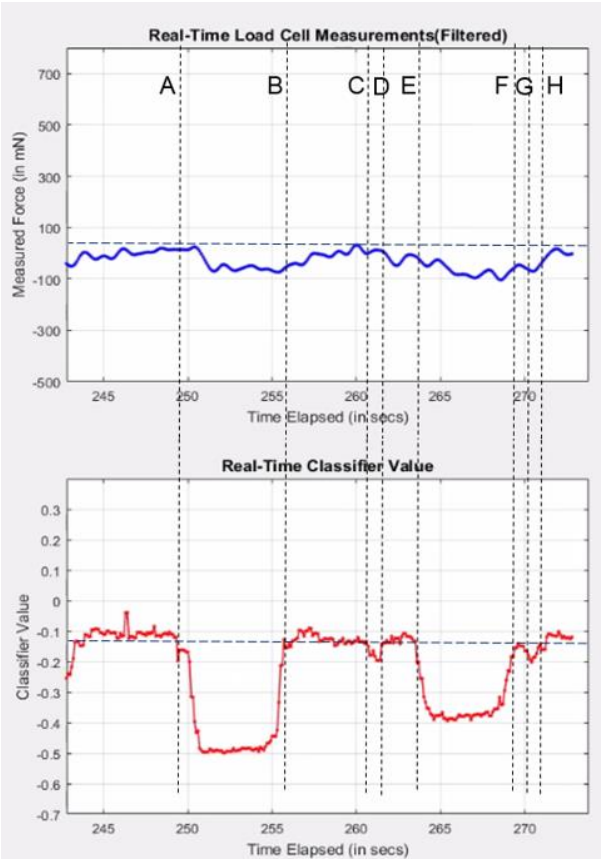

Fig. 6. Real-time controller performance in addressing image quality drop due to loss of probe-tissue contact. It can be seen that when image quality is regained the applied force on the tissue does not cross acceptable levels.

B. Controller Performance in Excess-Deformation Case

Fig. 7 was obtained while testing the performance of the controller in the excess-deformation scenario. It should be noted from Fig. 7 that the drop in classifier value when excess force is applied is extremely small as compared to the sub-optimal classifier values attained in the loss-of-contact scenario (see Fig. 6). As mentioned earlier, this is the reason why RL was not used to learn the control actions in the case of excess deformation. As a result, it can be seen that recovery after the compensatory action is much more gradual and not immediate as was demonstrated in the loss-of-contact case.

In Fig. 7, excess deformations were induced at instances (A) & (D) and the controller showed adequate performance in recovering from the excess deformation. The time taken for the recovery is proportional to the extent of deformation. As there are no best action values, it can be seen the action value had to be performed twice to remove the larger deformation induced in A and is the reason for the significant change in slope in the recovery from the first deformation. The improvement in classifier values at (C) in Fig. 7 was not due to a second corrective action, but due to a better image obtained after scanning was switched on post image quality recovery.

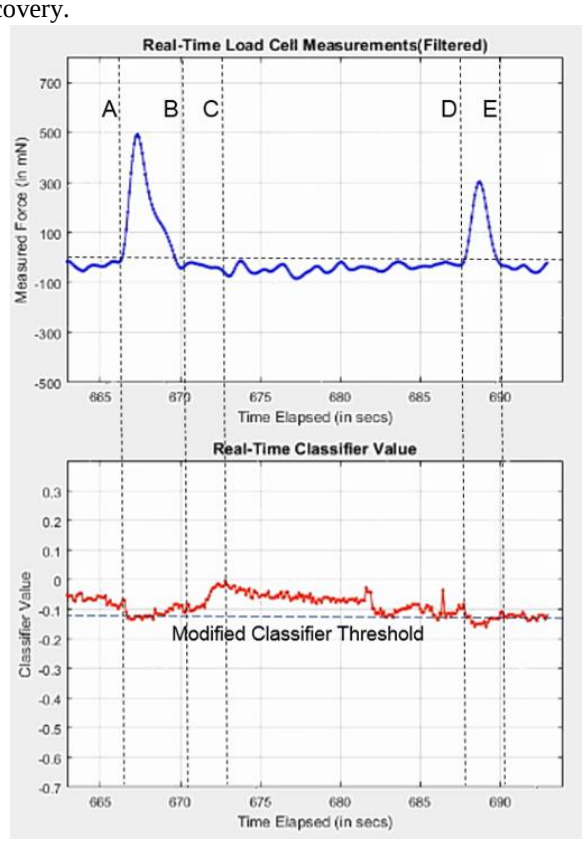

Fig. 7. Real-time controller performance in resolving excess-deformation induced on the tissue by the probe. When excessive force is applied (peak in force measurement graph), an image quality drop ensues (in the second graph), and the controller then acts to recover image quality and subsequently remove excess force applied on tissue

V. Conclusion

In this paper, a novel framework for autonomous and sensorless probe-tissue contact control has been proposed. A new method for obtaining a measurement signal based on image quality information has been presented, which could replace external sensing modalities for probe-tissue contact control. The requisite computation is extremely light and achieves real-time performance on conventional laptop computers. The derived classifier has also demonstrated versatility in being able to classify microscopy images from different systems with minimal tuning requirements.

The obtained measurement signal (classifier value) was

used to develop a controller for probe-tissue contact control using model-free learning techniques. The absence of any *apriori* knowledge of system dynamics during controller design facilitates cross-platform compatibility. The controller has demonstrated near real-time ability to resolve small loss-of-contact and excess-deformation scenarios. The controller is currently slow for large deformations to effectively compensate physiological motion. As the controller software runs as a layer on top of the robot controller software, it becomes independent of the inherent kinematics/dynamics of the system being controlled. Additionally, the design aspects of the framework (choice of state-space and reward functions) make the controller learning extremely fast and efficient.

It is worth noting that the biggest limitation of the current framework is the absence of global positioning. The current framework only allows relative positioning along the scanning plane normal, and the user needs to maintain the orientation. An extension of the framework could be to combine the work done by Zhang *et al*. in [10] incorporating stereo reconstruction, surface normals identification and subsequently achieving global positioning. Future work on the classifier could also include making the classifier more robust and adaptive to external disturbances.

Though the presented application is autonomous manipulation of a CLE probe for optimal quality data acquisition, this does not limit the developed framework's capabilities. Other potential applications could be systems requiring the manipulation of actuators to optimize imaging quality like autofocusing in imaging/visualization systems (*e.g.* microscopes), autonomous ultrasound scanning *etc*.

ACKNOWLEDGMENT

The authors would like to thank Dr. Mike Hughes, Lin Zhang and Khushi Vyas for all their help with the pCLE and widefield microscopy systems.

REFERENCES

[1] M. C. Pierce, D. J. Javier, and R. Richards-Kortum, "Optical contrast agents and imaging systems for detection and diagnosis of cancer," *International journal of cancer*, vol. 123, pp. 1979-1990, 2008.

[2] R. C. Newton, S. V. Kemp, P. L. Shah, D. Elson, A. Darzi, K. Shibuya, *et al.*, "Progress toward optical biopsy: bringing the microscope to the patient," *Lung*, vol. 189, pp. 111-119, 2011.

[3] J. M. Jabbour, M. A. Saldua, J. N. Bixler, and K. C. Maitland, "Confocal endomicroscopy: instrumentation and medical applications," *Annals of biomedical engineering*, vol. 40, pp. 378-397, 2012.

[4] M. Hughes and G.-Z. Yang, "Robotics and smart instruments for translating endomicroscopy to in situ, in vivo applications," *Computerized Medical Imaging and Graphics*, vol. 36, pp. 589-590, 2012.

[5] R. Newton, D. Noonan, C. Payne, J. Andreyev, A. Di Marco, M. V. Scarzanella, *et al.*, "Probe tip contact force and bowel distension affect crypt morphology during confocal endomicroscopy," *Gut*, vol. 60, pp. A12-A13, 2011.

[6] W. T. Latt, R. C. Newton, M. Visentini-Scarzanella, C. J. Payne, D. P. Noonan, J. Shang, *et al.*, "A hand-held instrument to maintain steady tissue contact during probe-based confocal laser endomicroscopy," *IEEE transactions on biomedical engineering*, vol. 58, pp. 2694-2703, 2011.

[7] W. T. Latt, T. P. Chang, A. Di Marco, P. Pratt, K.-W. Kwok, J. Clark, *et al.*, "A hand-held instrument for in vivo probe-based confocal laser endomicroscopy during minimally invasive surgery," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 1982-1987.

[8] S. Zuo, M. Hughes, C. Seneci, T. P. Chang, and G.-Z. Yang, "Toward Intraoperative Breast Endomicroscopy With a Novel Surface-Scanning Device," *IEEE Transactions on Biomedical Engineering*, vol. 62, pp. 2941-2952, 2015.

[9] 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*, vol. 10, pp. 825-832, 2015.

[10] L. Zhang, M. Ye, P. Giataganas, M. Hughes, and G.-Z. Yang, "Autonomous Scanning for Endomicroscopic Mosaicing and 3D Fusion," in *2017 IEEE International Conference on Robotics and Automation (ICRA)*, In press

[11] B. Rosa, B. Herman, J. Szewczyk, B. Gayet, and G. Morel, "Laparoscopic optical biopsies: in vivo robotized mosaicing with probe-based confocal endomicroscopy," in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011, pp. 1339-1345.

[12] N. D. Narvekar and L. J. Karam, "A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection," in *Quality of Multimedia Experience, 2009. QoMEx 2009. International Workshop on*, 2009, pp. 87-91.

[13] A. V. Murthy and L. J. Karam, "A MATLAB-based framework for image and video quality evaluation," in *Quality of Multimedia Experience (QoMEX), 2010 Second International Workshop on*, 2010, pp. 242-247.

[14] F. Crété-Roffet, T. Dolmiere, P. Ladret, and M. Nicolas, "The blur effect: Perception and estimation with a new no-reference perceptual blur metric," in *SPIE Electronic Imaging Symposium Conf Human Vision and Electronic Imaging*, 2007, pp. EI 6492-16.

[15] R. Hassen, Z. Wang, and M. M. Salama, "Image sharpness assessment based on local phase coherence," *IEEE Transactions on Image Processing*, vol. 22, pp. 2798-2810, 2013.

[16] S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, "Supervised machine learning: A review of classification techniques.", 2007.

[17] F. L. Lewis, D. Vrabie, and K. G. Vamvoudakis, "Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers," *IEEE Control Systems*, vol. 32, pp. 76-105, 2012.

[18] A. G. Barto and R. S. Sutton, "Reinforcement learning," *The handbook of brain theory and neural networks*, pp. 804-809, 1998.

[19] J. N. Tsitsiklis and B. Van Roy, "Feature-based methods for large scale dynamic programming," *Machine Learning*, vol. 22, pp. 59-94, 1996.

[20] C. J. C. H. Watkins, "Learning from delayed rewards," University of Cambridge England, 1989.

[21] R. E. Bellman and S. E. Dreyfus, *Applied dynamic programming*: Princeton university press, 1962.

[22] M. Hughes and G.-Z. Yang, "Line-scanning fiber bundle endomicroscopy with a virtual detector slit," *Biomedical Optics Express*, vol. 7, pp. 2257-2268, 2016.

[23] M. Pierce, D. Yu, and R. Richards-Kortum, "High-resolution fiber-optic microendoscopy for in situ cellular imaging," *JoVE (Journal of Visualized Experiments)*, pp. e2306-e2306, 2011.

[24] B. Hannaford, J. Rosen, D. W. Friedman, H. King, P. Roan, L. Cheng, *et al.*, "Raven-II: an open platform for surgical robotics research," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 954-959, 2013.

[25] P. Berthet-Rayne, M. Power, H. King, and G.-Z. Yang, "Hubot: A three state Human-Robot collaborative framework for bimanual surgical tasks based on learned models," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 715-722.

[26] M. S. Erden, B. Rosa, J. Szewczyk, and G. Morel, "Mechanical design of a distal scanner for confocal microlaparoscope: a conic solution," in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, 2013, pp. 1205-1212.