Online robust endomicroscopy video mosaicking using robot prior

B. Rosa[1], B. Dahroug[2], B. Tamadazte[2], K. Rabenorosoa[2], P. Rougeot[2], N. Andreff[2], P. Renaud[1]

***Abstract*—This paper discusses the development of a mosaicking algorithm for building large and high resolution confocal images. Due to the nature of optics and vision systems in general, there is still a dilemma between choosing a wide field-of-view (FOV) and high-resolution. The most accepted solution is to opt for a high-resolution optics and expand the FOV *algorithmically* thanks to mosaicking approaches. The study reported in this paper consists of online and real-time construction of large mosaics using individual confocal images with a micrometer resolution. These individual images are provided by a confocal laser endomicroscopy system which can grab *in vivo* real-time images through a minimally invasive access. The acquisition of the confocal images is achieved by moving the imaging probe on the studied sample surface with a constant contact between the probe and the sample.**

The mosaicking algorithm proposed in this paper deals with the combination of both the robot inputs and the image registrations. The proposed method has demonstrated very promising performances in terms of accuracy and robustness with regard to image noise (poor image quality or loss of contact between the probe and the sample) as well as misregistration issues. Experiments carried out with a highly accurate robotic system and a ground truth obtained by conventional optical microscopy demonstrate the robustness of the proposed approach.

***Index Terms*—Medical Robots and Systems; Computer Vision for Medical Robotics; Sensor Fusion**

I. Introduction

OPTICAL biopsy techniques, in opposition to the physical ones, are increasingly used in clinical investigations, thanks to the ability to directly visualize microscopic cellular structures without the need to take a physical tissue sample. In fact, optical biopsy images can be useful in several clinical scenarios for: i) reducing sampling errors and costs; ii) reducing the need for excision, transport, storage and examination of the sampled tissue, and iii) providing *in situ*, *in vivo*, and real-time feedback during (micro)surgical procedures. Among the imaging techniques used for optical biopsy, probe-based confocal laser endomicroscopy (pCLE) is a very promising modality [1]. A common problem, however, is that the field-of-view (FOV) of micrometer resolution optical biopsies is too narrow for a proper diagnosis. Consequently, several image mosaicking techniques have been reported in the literature. The principle is to move the imaging system (or the tissue sample with respect to the imaging system) to collect high resolution images, which will be used to compute the mosaic [2], [3]. While mosaicking of endomicroscopic images has been validated in clinical studies [3], [4], the microscale movements necessary to produce good quality large FOV mosaics were a limiting factor. As a result, many studies proposed robotic assistance. Different combinations of embedded microactuators [5], [6], sensors [7], stabilizers [5], [8], [9], and control architectures [6], [10] were proposed to reliably displace the probe with respect to the tissue in order to produce large FOV mosaics. Introducing robotic-assisted endomicroscopy has radically improved the quality of the produced mosaics. However, many of those studies were performed in *ex vivo* conditions. In real surgical settings, various phenomena – such as partial loss of contact between the probe and the tissue (e.g., due to non-planar tissue geometry [11] or inaccurate depth control [8]), debris on the tissue surface, or nonlinearities of the robotic actuators (local accelerations, mechanical

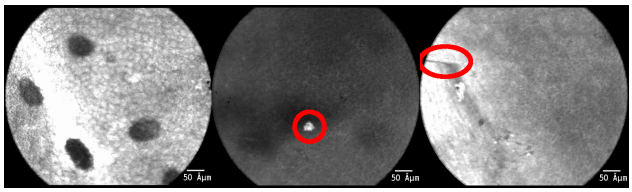


Fig. 1. Sample images acquired by a CellVizio endomicroscopy probe on a *ficus benjamina leaf*. *Left*: good contact conditions. *Middle*: Bad contact conditions. *Right*: motion artifact (circle) induced by local accelerations. Note how the middle image is dominated by noise, with artifacts from the fiber bundle showing up (circle).

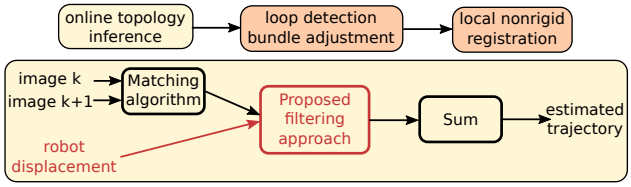


Fig. 2. General structure of the mosaicking algorithm. Top: mosaicking pipeline from [2], [3] (lighter colored boxes are computed in real-time, others offline). Bottom: detail of the online topology inference. The parts added in the proposed algorithm are in red.

Manuscript received: Feb, 25, 2018; Revised May, 27, 2018; Accepted July, 7, 2018.

This paper was recommended for publication by Editor Ken Masamune upon evaluation of the Associate Editor and Reviewers' comments. This work was supported by France Life Imaging (grant ANR-11-INBS-0006'), and by the French National Agency for Research (grant NEMRO ANR-14-CE17-0013), Labex CAMI (ANR-11-LABX-0004) and Labex ACTION (ANR-11-LABX-0001-01).

[1] ICube-AVR, UMR7357, CNRS, Université de Strasbourg, INSA, Strasbourg, France. firstname.lastname@icube.unistra.fr

[2] FEMTO-ST Institute, Univ. Bourgogne Franche-Comté, CNRS, Besançon, France. firstname.lastname@femto-st.fr

Preprint version prepared by the authors after review for IEEE Robotics and Automation Letters. The official version can be found on IEEEXplore, searching for the DOI 10.1109/LRA.2018.2863372

backlash, hysteresis and/or creep effects) – will create image artifacts. In some cases, this can result to complete image loss (Fig. 1). Those artifacts and image losses are detrimental for the mosaicking process. A typical mosaicking pipeline includes image matching and warping, loop closure and bundle adjustment, and mosaic construction [12], [13]. This is a well established process, which can run in real-time in some contexts (e.g. panorama reconstruction in modern smartphones). In the context of endomicroscopic image mosaicking [2], [3], the typical pipeline is a similar three-stage process (Fig. 2). First, an online topology inference is constructed by registering successive frames together and summing those registrations. In the second pass, loop closure detection and bundle adjustment are performed. Finally, a third pass may be performed to account for local tissue deformations. It should be highlighted that variations in contact conditions and other physical phenomena can cause large illumination and aspect differences between non-sucessive images, which make direct image matching for online loop closure and bundle adjustment extremely challenging (Fig. 3). As a result, the two main clinically-validated methods for endomicroscopic image mosaicking use the result of the online topology inference to perform loop closure in a second step, by detecting images which are close spatially but not temporally. Those images are then registered together, and the result is used as a constraint in the bundle adjustment step [2], [3].

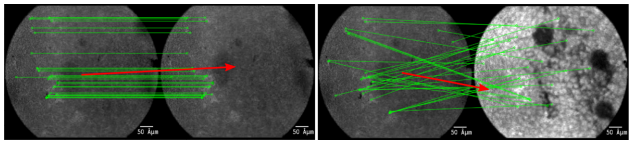


Fig. 3. Direct matching of sequential (left) and topologically close but not sequential (right) images using ORB features [14], in the event of image contrast loss. Green lines: found matches; Red arrow: image displacement obtained by careful manual alignment.

In the event that large image losses and artifacts are present locally, registering consecutive images together gets very difficult. Usual registration algorithms will likely fail at those points (see Fig. 3 for an example using ORB [14] feature matching), causing an accumulation of large errors in the online topology inference. Since this first topology inference is used as a first guess in the subsequent steps of the algorithm, large errors will cause the whole process to fail. Moreover, in the case of robot-assisted mosaicking, several studies have shown the importance of using visual feedback for controlling the robot movement and producing large FOV mosaics [6], [10]. If the motion estimation is erroneous, this could cause control issues and instabilities. Finally, clinical studies have also shown the importance of the online topology inference for the clinician [4]. In fact, displaying the online topology inference, even if inaccurate, on the screen, will help the surgeon assess whether the final offline-computed mosaic will be correct. For all those reasons, the online topology inference algorithm needs to be robust to large artifacts and local image losses.

A. Contributions

This paper proposes an algorithm for robust online topology inference in the context of robot-assisted endomicroscopic image mosaicking. We propose to use extra information coming either from the robot commanded trajectory (or from position sensors placed on the probe) to robustify the online estimation. Mahé *et. al.* [15] developed a topology inference algorithm using weak shape priors from the robot trajectories, specifically tailored at spiral ones. More recently, Vyas and her colleagues [7] proposed tracking the probe motion using an electromagnetic tracker in order to provide a first guess to the image matching algorithm. This approach is however limited by the fact that the tracker might not give an accurate representation of the probe/tissue displacement (for instance due to tissue deformations [11]), therefore providing the image matching algorithm with an incorrect first guess.

In this paper, we propose a novel method for online topology inference based on a Kalman filter.The combination of the robot data and the image information in the Kalman filter scheme allows avoiding the image quality and spurious matching issues. In order to efficiently fuse the image and robot information, we also propose a method to estimate the relative confidence in those two estimates. It is designed to run online, and to be robust to image contrast loss, misregistrations, and non-uniform robot motion (e.g., local accelerations, mechanical backlash, etc.). As displayed in Fig. 2, the proposed method is an alternative, robust online topology inference. Subsequent loop closure detection and bundle adjustment steps could be performed using methods from [2], [3].

The developed method is detailed in Section II. Section III presents the used materials and devices to perform the experimental validation. Finally, Section IV discusses the validation scenarios as well as the ground truth comparison of the obtained results.

II. Kalman Filtering For Online Topology Inference

This section deals with the introduction of the topology inference filtering method that occurs during the first pass of the mosaicking process. The proposed method runs online and in real-time (i.e. greater than the image acquisition rate of 10-12Hz). With such a framerate, the mosaic image reconstruction allows giving feedback to the clinician during the acquisition process, and allows him/her identifying whether the mosaic will be correctly formed in the subsequent off-line optimization process [4].

A. Filtering the Topology Inference

Let us consider an image $\mathcal{I}^i$ grabbed at time i. The position of $\mathcal{I}^i$ in the mosaic is noted $X_m(i)$, and its registration with respect to $\mathcal{I}^{i-1}$ is noted dX^i_{i-1}. As such, the simple image-based topology inference (i.e. first pass of the mosaic) can be expressed as follows

$$X_m(i) = X_m(i-1) + dX^i_{i-1} \quad (1)$$

$$= \sum_{k=1}^{i} dX^k_{k-1} \quad (2)$$

At the same time i, the robot is controlled with a velocity $V_r(i)$ in order to track accurately a predefined trajectory $\Gamma(i)$. Similarly, and by assuming that the robot control frequency $f_r = 1/T_s$ (where T_s is the sampling time) is synchronized with the image framerate, one can write

$$X_r(i) = \sum_{k=1}^{i} V_r(k) T_s, \tag{3}$$

It should be underlined that both X_m and X_r are relative to the first acquired image.

With the aim of filtering out the bad matches, we design a Kalman filter which will take both X_m and X_r into account to estimate a filtered position X_f. Following a Bayesian notation, one can write the belief in a given state as a Gaussian $\mathcal{N}(\mu, \sigma^2)$, where μ is the estimated position and σ^2 is the associated variance. At a time instant i, the belief in the current state can then be written as $\mathcal{N}(X_f(i), \sigma_f^2)$. By considering the robot control speed as a process model, then the prior $\mathcal{P}(i+1)$ can be expressed as follows

$$\begin{aligned} \mathcal{P}(i+1) &= \mathcal{N}(\mu_{\mathcal{P}}(i),\ \sigma^2_{\mathcal{P}}(i)) \\ &= \mathcal{N}\Big(X_f(i) + V_r(i+1)T_s,\ \sigma^2_{rob}(i)\Big), \end{aligned} \tag{4}$$

where $\sigma^2_{rob}(i)$ is the variance associated with the robot prediction.

The likelihood $\mathcal{L}(i)$, coming from the image estimate dX_i^{i+1}, is written as follows

$$\begin{aligned} \mathcal{L}(i+1) &= \mathcal{N}\Big(\mu_{\mathcal{L}}(i),\ \sigma^2_{\mathcal{L}}(i)\Big) \\ &= \mathcal{N}\Big(X_f(i) + dX_i^{i+1},\ \sigma^2_{img}(i)\Big) \end{aligned} \tag{5}$$

Similarly to $\sigma^2_{rob}(i)$, $\sigma^2_{img}(i)$ is the variance associated with the image-based estimation. Using those notations, the filtered state $\mathcal{F}(i+1)$ can be defined as the multiplication of the two Gaussians

$$\begin{aligned} \mathcal{F}(i+1) &= \mathcal{N}\Big(X_f(i), \sigma^2_{est}(i)\Big) \\ &= ||\mathcal{P}(i).\mathcal{L}(i)|| \end{aligned} \tag{6}$$

The filtered state $X_f(i)$ and the associated variance $\sigma^2_{est}(i)$ are then computed using the standard formula for the multiplication of two Gaussians:

$$X_f(i) = \frac{\sigma^2_{\mathcal{P}}(i)\mu_{\mathcal{L}}(i) + \sigma^2_{\mathcal{L}}(i)\mu_{\mathcal{P}}(i)}{\sigma^2_{\mathcal{P}}(i) + \sigma^2_{\mathcal{L}}(i)} \tag{7}$$

$$\sigma^2_{est}(i) = \frac{\sigma^2_{\mathcal{P}}(i)\sigma^2_{\mathcal{L}}(i)}{\sigma^2_{\mathcal{P}}(i) + \sigma^2_{\mathcal{L}}(i)} \tag{8}$$

Because endomicroscopic images are confocal, the mosaic is built in the $x - y$ plane. Therefore, the variables $X_{m,r,f}$ and V_r have two components, which means the Gaussians are theoretically multivariate in the previous equations. We do not, however, have specific information as to how the x and y elements of the image or robot speed vary in relation to one another. As such, we chose to develop two independent filters for the x and y components of the displacement. Each of those filters is governed by Eqs. 1–8.

B. Prior and Likelihood Variance

To function properly, a key element of a Kalman filter is to have a good estimate of the Prior and Likelihood variances (respectively $\sigma^2_{rob}(i)$ and $\sigma^2_{img}(i)$ in our framework). Indeed, those variances will govern how much information from the Prior and Likelihood is incorporated into the filtered output (Eqs 7–8). In our case, neither the robot trajectory or the image-based estimations are perfect along the whole scanning path, due to the various phenomena discussed earlier (tissue deformations, loss of probe/sample contact, poor image contrast, nonlinearities on the robot motion, etc.). To take this into account, we propose a method which modulates the covariances depending on the current *confidences* c_{img} and c_{rob} corresponding to the robot and the image estimations, respectively. This modulation is expressed as follows

$$\sigma^2_{rob}(i) = \frac{\sigma^2_{r0}(i)}{c_{rob}(i)} \tag{9}$$

$$\sigma^2_{img}(i) = \frac{\sigma^2_{i0}(i)}{c_{img}(i)} \tag{10}$$

with σ_{r0} and σ_{i0} initial values. Because we do not have any information about how the probe interacts with the tissue, we choose to rely more on the image data than on the robot data at the start, thus initializing these parameters as follows : $c_{rob}(0) = 10$ and $c_{img}(0) = 1$. These values will be updated as the scanning task progresses, as detailed in the two following subsections.

C. Image Matching Confidence Estimation

As defined above, $c_{img}(i)$ represents the confidence in a given image estimation at time i, i.e. the likelihood of the translation dX_{i-1}^i to be accurate. Using the output metric of the registration algorithm (i.e. normalized cross correlation (NCC) in [3], [4] or the sum of squared differences (SSD) in [2]) is not necessarily suitable, especially in the case of low contrast images or presence of image artifacts, which are generally characterized by a unfavorable signal-to-noise ratio (see Fig. 1).

Rather, we hypothesize that good image quality matches are consistent with the direct preceding matches. This assumes that the scanning path followed by the probe with respect to the tissue is smooth. Obviously, it is entirely possible, even if the robot input trajectory is smooth, to induce local accelerations in the image, either due to mechanical imperfections of the robot, or to stick/slip effects (during the probe/tissue contacts). However, as shown in [3], such accelerations induce image deformations due to the physical image acquisition process, which make the registration unreliable or at least inaccurate. Hughes et *al.* recently proposed a promising high frame rate endomicroscope which partially tackles this problem [16]. This prototype is however neither available to the public, nor marked for clinical uses.

Let us note x and y the two components of the image-estimated displacement at time i, i.e. $dX_{i-1}^i = (x; y)^\top$. We

define $\alpha(i)$ as the angle of the vector dX_{i-1}^{i} and $N(i)$ its norm :

$$\alpha(i) = \arctan 2\Big(y(i), x(i)\Big) \quad (11)$$

$$N(i) = \Big(x(i)^2 + y(i)^2\Big)^{\frac{1}{2}} \quad (12)$$

Using this formalism, we estimate the probability of having a good match at time i by looking at the smoothness of α and N. Therefore, to estimate this smoothness, we compare the direction and the norm of the current velocity vector with low-pass filtered values. This means that, if $h_\alpha(i,n) = [\alpha(i-n), ...\alpha(i)]$ is the history of α (respectively h_N for N) over the n time-steps directly preceding i, one can write

$$c_{angle}(i) = normalize\Big(\alpha(i) - f\big(h_\alpha(i,n)\big)\Big) \quad (13)$$

$$c_{speed}(i) = normalize\Big(N(i) - f\big(h_N(i,n)\big)\Big) \quad (14)$$

where $f(.)$ is a low-pass filtering operator which will eliminate the oscillations.

To do this, we use a median filter (to remove large peaks) followed by a 3^{rd} order polynomial fitting (to smoothen the curve). The $normalize$ function is defined as follows

$$normalize(x) = \frac{\Big(1 - erf\big(k * (x - x_r)\big)\Big)}{2}, \quad (15)$$

where erf is the Gauss error function. This function is chosen to normalize the confidence score between 0 and 1, with k and x_r setting the slope and the 0.5 confidence level, respectively.

One can also estimate the image matching confidence by looking at the similarity between aligned images. In this case, the score c_{match} is estimated by computing the structural similarity [17] of the overlapping part of two successive images $\mathcal{I}^i$ and $\mathcal{I}^{i-1}$, once aligned. As this score is by definition between 0 and 1, it is not necessary to normalize the obtained value.

Finally, the image matching confidence is defined as the geometric mean of c_{angle}, c_{speed}, and c_{match} :

$$c_{img}(i) = \Big(c_{angle}(i) * c_{speed}(i) * c_{match}(i)\Big)^{\frac{1}{3}} \quad (16)$$

D. Robot Trajectory Confidence Estimation

Due to probe/tissue interactions, the robot commanded trajectory is very likely to be different from the probe/tissue displacements, even with a high-accuracy robot [10]. For this reason, it is also necessary to infer a confidence score c_{rob} for the robot trajectory. We propose a score based on both the image and the robot estimations. As a reminder, it has been stated that at a given time i, the estimated displacement in the image $\mathcal{I}$ is noted dX_{i-1}^{i}, when the one performed by the robot trajectory is $dX_r(i) = V_r(i)T_s$. Let us introduce $n_I(i)$ and $n_r(i)$ the respective norms of those vectors. Hence, it is possible to define a speed confidence metric c_{sp} as a score between 0 and 1 reflecting the difference between two displacements. This score can be obtained by

$$c_{sp}(i) = normalize\left(\min\left(\frac{n_I(i)}{n_r(i)}, \frac{n_r(i)}{n_I(i)}\right)\right) \quad (17)$$

where c_{sp} is the difference between the robot velocity and the one estimated in the image (please note that c_{sp} is different from c_{speed} in the sense that the former represents the speed difference between the robot and image estimates, while the latter computes a difference between successive image estimates).

This score will tend to 0 if one displacement is significantly smaller than the other one, and to 1 if both displacements get close from one another. To estimate where discrepancies between robot and image-estimated speeds come from, we propose a combined score between c_{img} and c_{sp}, as follows

$$c_{rob} = c_{sp} * c_{img} - c_{img} + 1 \quad (18)$$

The obtained *mixed* score is built so that the confidence in the robot trajectory gets to 0 if c_{img} is high and c_{sp} is low. This means that there is an important difference in speed between the robot input and the image-based estimation. In this case, we consider that the image-based estimation is more reliable. Alternatively, c_{rob} tends to 1 when c_sp tends to 1, and when c_{img} tends to 0 (see Fig. 4).

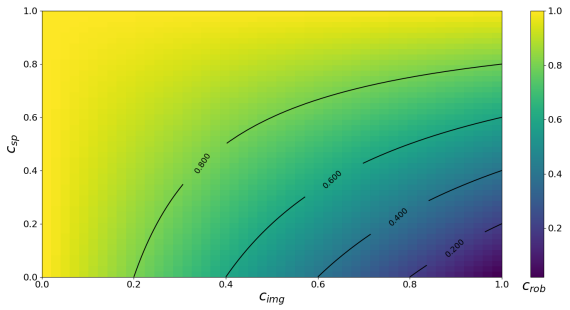


Fig. 4. Evolution of c_{rob} as a function of c_{img} and c_{sp}.

III. Materials and Devices

The endomicroscopy mosaicking problem described in this paper was experimentally validated using using several evaluation scenarios. The experimental tests were made on a benchtop robotic setup including of a highly accurate 6 degrees-of-freedom (DOF) parallel structure. Furthermore, the used endomicroscopy imaging system consists of the CellVizio laser confocal microscopy from Mauna Kea Technologies Inc.[1] The CellVizio was mounted in an *eye-to-hand* configuration which allows visualizing the sample carried by the robotic platform (Fig. 5). In others words, the endomicroscopy probe remains fixed when the viewed sample moves relatively to the probe.

A. Robotic Setup

The robotic setup i.e., the sample holder (Fig. 5), consists of a 3PPSR robot SpaceFAB SF-3000 BS from Micos[2]. The latter is characterized with the following features: translation ranges $(t_x, t_y, t_z)^\top_{max}$ = (50, 100, 12.7)$^\top$ $[mm]$ and rotation

[1]www.maunakeatech.com
[2]www.pimicos.com

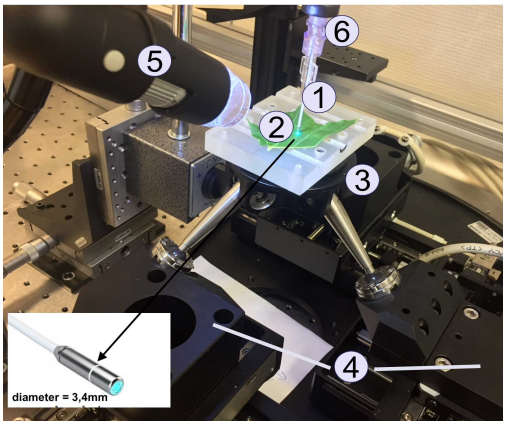


Fig. 5. Photography of the experimental setup. ① pCLE imaging system, ② scanned sample, ③ 6 DOF sample holder platform, ④ positioning stages, ⑤ external optical microscope for ground truth validation, ⑥, pCLE holder.

ranges $(r_x, r_y, r_z)^\top_{max}$ = (10, 10, 10)$^\top$ [°], a linear resolution of 0.2μm (repeatability of $\pm$0.5μm) and an angular resolution of 0.0005° (repeatability of $\pm$0.0011°).

Two computers equip the experimental platform: the first one (a 3.20-GHz i5 core Intel CPU with a MacOS X distribution) is dedicated to the endomicroscopy images acquisition when the second one (a 2.33-GHz Xeon Intel CPU with a Windows distribution) is used for the robot inner control (inner PID loop, static and differential kinematic models). The computers communicate asynchronously using a TCP/IP protocol.

B. Confocal Laser Endomicroscopy

The CellVizio endomicroscopy system is a standalone imaging system based on a fibered technology capable to achieve real-time (9 to 12 Hz) and *in situ* optical biopsies via minimally invasive access (Fig. 5). The Cellvizio incorporates a proximally-scanned fiber bundle to deliver 488 nm wavelength laser light to the sample and acquire a fluorescence signal in return. In our study, we used Z1800 probe, which incorporates a fiber bundle composed of 30,000 optical fibers, providing a lateral resolution of 3.5 μm with a FOV of 500 microns, at a framerate of 9-10 images/second. After exporting the images, the resolution is 512x448 pixels.

C. Robot Trajectory Generation

In order to ensure an accurate achievement of the scanning path during the mosaicking process, the robot is controlled using a path following scheme. This means that, in addition to the inner PID controller which controls each robotic stage, we implemented an external closed-loop controller. The developed path following approach has the advantage of decoupling the velocity profile from time, geometric shape, size, ... of the scanning curve $\Gamma(i)$. In other words, the path following accuracy is expected to be independent from the velocity amplitude which can be tuned by the operator independently. Actually, the controller needs only the cartesian coordinates $(x_i,\ y_i)$ of sampled 2D points (respectively, 3D points) to perform the path tracking. For more details, please refer to [18], [19].

The pCLE probe was placed with its imaging plane parallel to the $x-y$ plane of the robot, and axes were calibrated by doing a simple straight line scan. After this calibration, the curves $\Gamma(i)$ were programmed in the same $x-y$ plane. In order to simulate varying contact conditions, which occur in clinical practice, variations in the z axis of the robot were introduced during the scanning trajectory. Those take the form of a sinusoidal oscillation of amplitude 150 μm around the nominal contact point between the probe and the sample. This led to portions of the trajectory where the probe/tissue contact was almost lost, and other portions where the contact force was too important. In the first case, the image contrast is gradually lost, whereas in the second case, the large contact forces lead to adherence between the probe and the sample, and stick-slip effects (i.e. a very still image while the robot keeps moving, and a subsequent acceleration when the elasticity of the fiber bundle overcomes the frictional forces).

Finally, one should note that the robotic setup used in our experiments is a commercial micropositionner with excellent repeatability (see section III-A). In minimally invasive settings, however, mechanical performances of scanning devices are typically more modest due to the lower performance of micro-mechanisms and noise on sensors. As a result, robotic trajectories followed with miniature mechanisms are typically noisier [10], [15]. We chose, nevertheless, to have the robot follow a smooth trajectory $\Gamma(i)$. Indeed, the inner control loop of the robot imposes limits on its dynamics, making the simulation of noise and mechanical vibrations very difficult. To simulate a difference between the commanded trajectory and the input given to our algorithm, we add artificial noise to the trajectory estimate $V_r(i)$. This noisy estimate of the robot effectively followed trajectory (resp. speed) is noted $\hat{X}_r(i)$ (resp. $\hat{V}_r(i)$) in the following.

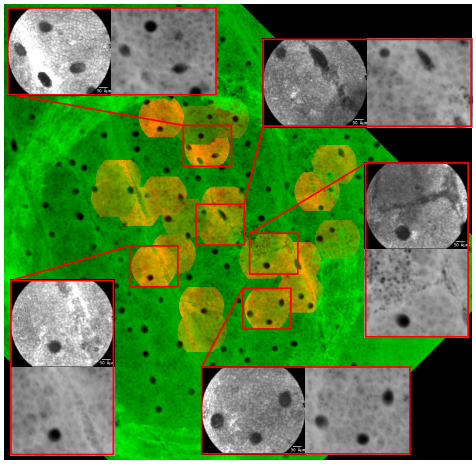
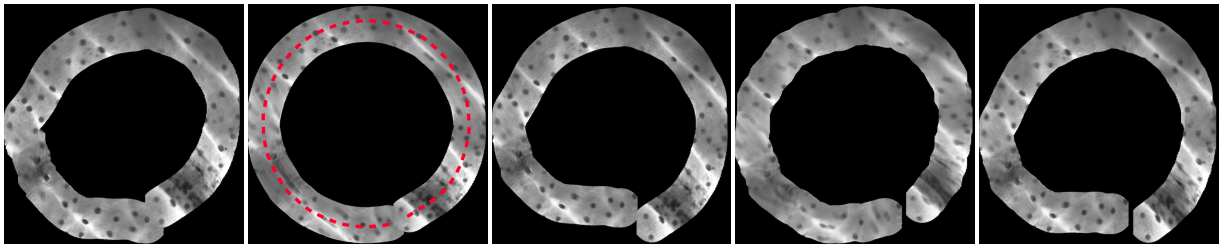


Fig. 6. Result of images registration for ground truth construction. Zoomed areas represent the endomicroscopic image, together with the corresponding area in the standard microscopy image.



(a) **Test #1** (b) **Test #2** (c) **Test #3** (d) **Test #4** (e) **Test #5**

Fig. 7. Reconstructed mosaics in the case of a circular trajectory, using the estimated trajectories from the five different test cases. The reference robot trajectory is overlaid over **Test #2**.

D. Implementation

The mosaicking algorithm was developed in python, using image processing routines from OpenCV and scikit-image. Normalized cross correlation was used for registering images and estimate dX^i_{i-1}, finding the maximum of the correlation between a template at the center of image $\mathcal{I}^{i-1}$ and image $\mathcal{I}^i$. After trying several template sizes, a template of 268x307 pixels (i.e. 3/5 of the original image size) was found to give the best results. In order to avoid side effects, the same process was repeated in the backward direction (i.e. matching a template from $\mathcal{I}^i$ into image $\mathcal{I}^{i-1}$), and the average displacement was taken. This displacement was subsequently corrected for motion artifacts. Finally, image blending was performed by placing all pixels from different images in a common reference frame and taking an average value. The interested reader can refer to [10] for further details on motion artifact compensation and image blending.

The implementation was running at a framerate higher than the pCLE video rate (12 Hz), therefore allowing the whole inference and online mosaic reconstruction process to run in real-time. Code optimization, as well as a C++ implementation, could allow much faster framerates.

IV. Experimental Evaluation

Different experimental scenarios using the test-bench presented in Section III were considered in order to judge the effectiveness of the proposed algorithm and methods.

A. Test cases

Two test cases were designed for testing our algorithm. In the first one, the robot was following a circular trajectory, while in the second one the reference trajectory was a spiral (with a straight line at the beginning and at the end). In both cases, we compared different outputs for the topology inference:

1) **Test #1**: using only the image measurements X_m
2) **Test #2**: using only the perferct robot trajectory X_r
3) **Test #3**: using the filtered output X_f with X_m and X_r as inputs.
4) **Test #4**: using only the noisy robot trajectory $\hat{X}_r$
5) **Test #5**: using the filtered output X_f with X_m and $\hat{X}_r$ as inputs.

One should appreciate that, even though all test cases are needed to evaluate the algorithm, Test #5 is the one that effectively simulates best the realistic conditions. In this case, the estimate of the robot trajectory is imperfect, and the contact conditions between the probe and the sample are varying.

B. Ground Truth and Evaluation Metrics

In order to assess the performance of the proposed algorithm, we built a ground truth scenario. We imaged the tissue sample (which, since it was taped on a rigid plate, is assumed to be rigid) using a standard optical microscope. Using a backlight to illuminate the sample, cell nuclei (which are typically seen in Cellvizio images) can be observed. We then used a coarse manual alignment, followed by a refinement using mutual information (the *imregister.m* function in MATLAB) in order to build a reference trajectory. Since this is a tedious process, n randomly selected images $\mathcal{I}_k, k \in [1, n]$ approximately evenly spaced along the trajectory, were selected for building this ground truth. n was 13 for the circular trajectory, and 25 for the spiral (which is longer). Figure 6 shows the result of the registration for obtaining the ground truth position of images in the spiral trajectory case.

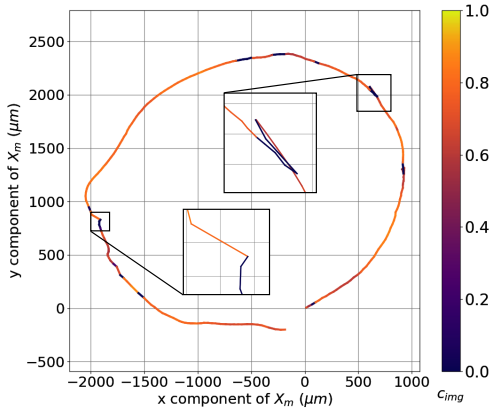
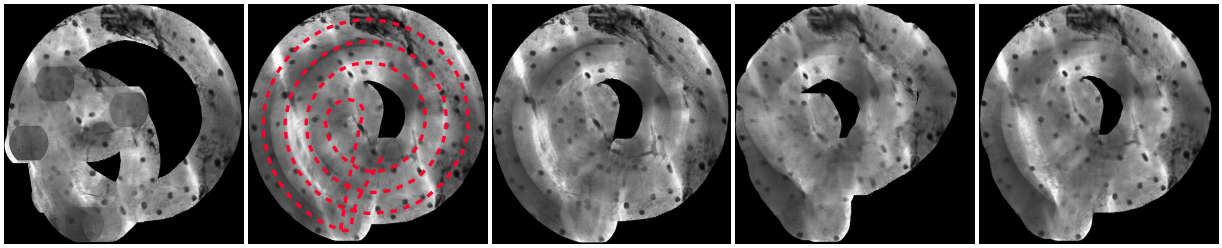


Fig. 8. Image-estimated trajectory X_m for the circular robot trajectory. The colormap corresponds to the values of c_{img} as they are estimated along the trajectory.

In order to estimate to goodness of fit between an estimated mosaicking trajectory and the reference optical microscope image, the distance between the postions of corresponding images in the ground truth and the mosaic was computed. Classical statistics such as average (mean), maximum error



(a) **Test #1** (b) **Test #2** (c) **Test #3** (d) **Test #4** (e) **Test #5**

Fig. 9. Reconstructed mosaics in the case of a spiral trajectory, using the different estimated trajectories from the five different test cases (Sec. IV-A). The reference robot trajectory is overlaid over **Test #2**

(max), standard deviation (std), and median error were then computed and reported in the following. Statistically significant differences are tested with the non-parametric Wilcoxon U-test, with a significance level set at $\alpha = 0.05$.

C. Results for a Circular Trajectory

Figure 7 depicts the reconstructed mosaic images using the five tests introduced above. The mosaic reconstructed using the robotic trajectory X_r looks visually accurate (Fig. 7(b)), which is confirmed by the error metrics on TABLE I. On such a simple trajectory, adding the information from the image registrations in order to filter the estimation adds little value (Fig. 7(c)). However, as soon as noise is added to the robot estimates, errors increase and the mosaic gets locally blurred (Fig. 7(d)). Filtering the estimation using the image information helps restoring a better mosaic shape (Fig. 7(e)), which is confirmed when looking at quantitative errors *w.r.t* the ground truth data (TABLE I). These results are further confirmed by statistical tests, which show a statistically significant difference between Test #4 and Test #5 ($p < 0.01$), but not between Test #2 and Test #3 ($p = 0.95$).

TABLE I
NUMERICAL ERROR VALUES FOR CIRCULAR MOSAIC RECONSTRUCTION. ALL THE VALUES ARE IN μM.

error *w.r.t.* ground truth	mean	max	std	median
X_m (Test #1)	274.3	543.1	190.2	249.2
X_r (Test #2)	116.5	251.6	73.3	108.4
X_f with X_r (Test #3)	142.3	320.8	86.8	108.6
$\hat{X}_r$ (Test #4)	489.1	752.8	201.8	551.8
X_f with $\hat{X}_r$ (Test #5)	226.8	372.1	86.4	224.0

Figure 8 represents the image-based estimated trajectory X_m with a colormap showing the estimated matching confidence c_{img}. One can see that the value of the confidence is very close to 1 over a large part of the trajectory except in rare positions which correspond to poor contacts between the probe and the tissue.

D. Results for a Spiral Trajectory

The same scenario as for the circle test is repeated with a more complex spiral trajectory. Figure 9 shows some examples of reconstructed mosaics using the five different tests mentioned in Sec. IV-A. One can see that the image-based estimation gets very far from a spiral (Fig. 9(a)). The trajectory being perfectly executed by the robot, the resulting mosaic using X_r is, again, close to perfect (Fig. 9(b)). As a result, filtering in this case adds little value (Fig. 9(c)), similarly to the circular case. However, as soon as the robot trajectory gets noisy (Fig. 9(d)) the errors get higher and the mosaic of worse visual quality. Our proposed fusion algorithm helps restoring a good topology inference (Fig. 9(e)), while being robust to very large local image losses and noisy robot trajectory inputs. Those results are confirmed by the error values reported in TABLE II. Again, statistical tests further confirm, with a statistically significant difference between Test #4 and Test #5 ($p < 0.001$), but not between Test #2 and Test #3 ($p = 0.07$).

TABLE II
NUMERICAL ERROR VALUES IN CASE FOR A SPIRAL MOSAIC RECONSTRUCTION. ALL THE VALUES ARE IN μM.

errors	mean	max	std	median
X_m (Test #1)	686.7	800.2	1080.8	375.3
X_r (Test #2)	139.4	220.9	51.3	141.8
X_f with X_r (Test #3)	135.7	263.1	70.7	133.4
$\hat{X}_r$ (Test #4)	291.0	517.2	134.9	312.9
X_f with $\hat{X}_r$ (Test #5)	100.8	260.8	63.9	105.9

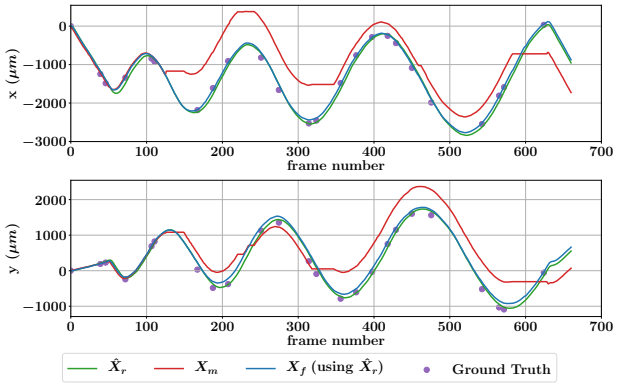


Fig. 10. Comparison of robot inputs, image estimations and filtered outputs (x and y coordinates as a function of the frame number).

Fig. 10 represents the x and y components of the estimated trajectories $\hat{X}_r$ (noisy robot inputs), X_m (image measurements), X_f (Kalman filtered outputs) and the ground truth (violet dots). Again, one can notice that inaccurate image

matches at some places in the trajectory create high accumulated errors in the end, which can also visually be seen on Fig. 9(a). Our approach allows filtering out those bad matches, using the robot inputs. In fact, as one can see on Fig. 11, the confidence score of the image matches is generally close to 1, and it's mostly at places where the trajectory is visually easily identifiable as wrong (zoomed areas where the image-estimated trajectory shows erratic local movements) that the confidence dramatically drops. This validates our trajectory smoothness assumption.

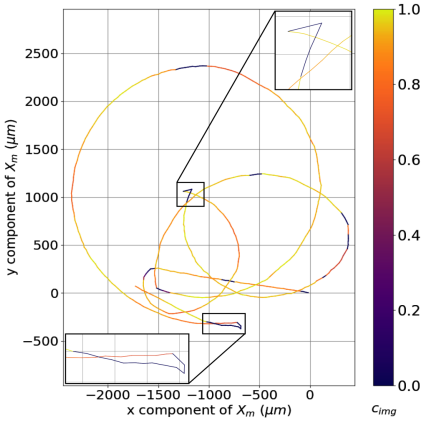


Fig. 11. Image-estimated trajectory X_m for the spiral trajectory case, with a colormap corresponding to values of c_{img}.

V. Conclusion and Perspectives

We proposed a novel online and real-time mosaicking method which uses both robot inputs and image measurements. While it is not a perfect topology inference, it is robust to image noise (i.e., artifacts, poor texture, etc.) and misregistrations. As such, it is useful to keep the structural coherence of the topology inference in the first pass of the mosaicking process, which helps subsequent passes to converge.

The method was experimentally validated under rigorous *ex vivo* scenarios using a high-accuracy robot, and using the established CellVizio confocal laser endomicroscopy system. Through the experimental results, it has been demonstrated that the developed method goes beyond current methods, especially in case of unfavorable conditions of use. Our approach is accurate and robust, allowing a reliable registration (i.e., large FOV mosaicking) even under non-smooth sample scanning process (contact loss between the probe and the tissue, robot nonlinearities, etc.). The ground truth validation was evaluated using different metric scores which showed promising performances in terms of accuracy, rapidity and robustness.

Future work will consist of validations using ex vivo and in vivo experimental setups. On deformable tissue samples, we plan to cope with tissue deformations by implementing the model from [11] in the process model of the filter. Further work will also include speeding up the subsequent optimization steps of the algorithm in [4] by using the confidence information. We will also investigate the integration of online bundle adjustment into the algorithm during the online mosaicking construction phase. Finally, we will integrate the pCLE imaging system in a concentric tube robot for *in vivo* tissue characterization.

References

[1] K. K. Wang, D. Carr-Locke, S. Singh, H. Neumann *et al.*, "Use of probe-based confocal laser endomicroscopy (pcle) in gastrointestinal applications. a consensus report based on clinical evidence," *United European gastroenterology journal*, vol. 3, no. 3, pp. 230–254, 2015.

[2] K. Loewke, D. Camarillo, W. Piyawattanametha, M. Mandella, C. Contag, S. Thrun, and J. Salisbury, "In vivo micro-image mosaicing," *IEEE Reviews in Biomedical Engineering*, vol. 58, no. 1, pp. 159–171, 2011.

[3] 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.

[4] T. Vercauteren, A. Meining, F. Lacombe, A. Perchant, J. Conchello, C. Cogswell, T. Wilson, and T. Brown, "Real time autonomous video image registration for endomicroscopy: fighting the compromises," in *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XV*, vol. 6861, 2008, p. 68610C.

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

[6] L. Zhang, M. Ye, P. Giataganas, M. Hughes, A. Bradu, A. Podoleanu, and G.-Z. Yang, "From macro to micro: Autonomous multiscale image fusion for robotic surgery," *IEEE Robotics & Automation Magazine*, vol. 24, no. 2, pp. 63–72, 2017.

[7] K. Vyas, M. Hughes, and G.-Z. Yang, "Electromagnetic tracking of handheld high-resolution endomicroscopy probes to assist with real-time video mosaicking," in *Endoscopic Microscopy X; and Optical Techniques in Pulmonary Medicine II*, vol. 9304. International Society for Optics and Photonics, 2015, p. 93040Y.

[8] R. Newton, D. Noonan, C. Payne, J. Andreyev *et al.*, "Probe tip contact force and bowel distension affect crypt morphology during confocal endomicroscopy," *Gut*, vol. 60, pp. A12–A13, 2011.

[9] P. Giataganas, M. Hughes, and G. Yang, "Force adaptive robotically assisted endomicroscopy for intraoperative tumour identification," *Int. J. of Computer Assisted Radiology and Surgery*, vol. 10, no. 6, pp. 825–832, 2015.

[10] B. Rosa, M. S. Erden, T. Vercauteren, B. Herman, J. Szewczyk, and G. Morel, "Building large mosaics of confocal edomicroscopic images using visual servoing," *IEEE transactions on biomedical engineering*, vol. 60, no. 4, pp. 1041–1049, 2013.

[11] M. Erden, B. Rosa, J. Szewczyk, and G. Morel, "Understanding soft-tissue behavior for application to microlaparoscopic surface scan," *IEEE Trans. on Biomedical Engineering*, vol. 60, no. 4, pp. 1059–1068, 2013.

[12] D. Capel, "Image mosaicing," in *Image Mosaicing and Super-resolution*. Springer, 2004, pp. 47–79.

[13] D. Ghosh and N. Kaabouch, "A survey on image mosaicing techniques," *Journal of Visual Communication and Image Representation*, vol. 34, pp. 1–11, 2016.

[14] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Computer Vision (ICCV), 2011 IEEE international conference on*. IEEE, 2011, pp. 2564–2571.

[15] J. Mahé, T. Vercauteren, B. Rosa, and J. Dauguet, "A viterbi approach to topology inference for large scale endomicroscopy video mosaicing," in *nt. Conf. on Medical Image Computing and Computer-Assisted Intervention*, 2013, pp. 404–411.

[16] 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.

[17] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[18] J. A. Seon, B. Tamadazte, and N. Andreff, "Decoupling path following and velocity profile in vision-guided laser steering," *IEEE Trans. on Robotics*, vol. 31, no. 2, pp. 280–289, 2015.

[19] B. Dahroug, B. Tamadazte, and N. Andreff, "Visual servoing controller for time-invariant 3d path following with remote centre of motion constraint," in *IEEE Int. Conf. on Robotics and Automation*, 2017, pp. 3612–3618.