Building robust confocal endomicroscopy mosaics despite image losses

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

[1] *ICube, UDS-CNRS-INSA, 300 bd Sébastien Brant - Illkirch, 67000 Strasbourg, France*
[2] *FEMTO-ST, University of Bourgogne Franche-Comté, CNRS, 25000 Besançon, France*
{*b.rosa*}*@unistra.fr* {*kanty.rabenorosoa*}*@femto-st.fr*

INTRODUCTION

Probe-based confocal laser endomicroscopy (pCLE) is a promising image modality for early cancer screening in various clinical applications. A typical limitation of probe-based systems, however, is the limited field-of-view (FOV) achievable with miniature optics. This is especially true with the high magnifications required for clinical assessment, namely for optical biopsy-based investigations. A widely accepted solution is to opt for high-resolution optics, and enlarge the FOV algorithmically by sweeping the probe along the tissue and reconstructing a mosaic.

While mosaicing effectively enhances the FOV, its accuracy is limited by the fact that the microscale movements required to sweep the probe are difficult to generate manually, especially in minimally invasive settings. For this reason, various approaches using a robotic micromanipulator and visual feedback control have been developed [1, 2, 3].

The above-mentioned methods have in common the fact that they rely on an accurate image-based motion estimation, which should be computed in real-time. Typically, pCLE-based visual servoing methods use a normalized cross correlation (NCC) computation between successive overlapping frames [4]. While this may be sufficient in *ex vivo* conditions, the constraints imposed by the *in vivo* environment make it more difficult. The image quality might be affected by a partial loss of contact with –or excessive force applied on– the tissue (e.g., due to non planar tissue geometry), or simply due to surgical debris present on the surface. Moreover, accelerations in the probe/tissue movement (due to various effects such as stick/slip effects or robot manufacturing inaccuracies) are detrimental for image quality and real-time matching. In summary, there may be parts of the trajectory where the image quality is insufficient for visual servo control, leading to the production of poor mosaics.

This abstract presents a Kalman filter-based approach, where both the image estimation and the (possibly inaccurate) robot trajectory are fused. We validate the proposed approach in controlled bench-top experiments, where the loss of contact with tissue is simulated. We show that it allows computing online mosaics with a coherent topology, despite an important loss of probe-tissue contact at several points along the trajectory. The method could be used for robustly estimating the probe-tissue movement online in a visual servo control loop.

MATERIALS AND METHODS

Let's consider the typical mosaicing situation in which a probe is moving with respect to a tissue. The robot moves with a speed $V_r(k)$, with k being a discrete time instance. $V_r(k)$ can be integrated in time to estimate a position $X_r(k)$. Due to various phenomena (encoder noise, mechanical inaccuracies, etc.), this position is estimated with a certain inaccuracy. We denote the noisy robot inputs $\hat{X}_r(k)$.

In parallel, thanks to online image-based methods such as the NCC method [4], an image trajectory $X_m(k)$ (i.e., the trajectory followed by successive images in the mosaic) can be estimated by integrating estimated displacements $dX_m(k)$.

We propose to estimate a filtered trajectory $X_f(k)$ by fusing together $\hat{X}_r(k)$ and $X_m(k)$ using a Kalman filter. The robot velocity at a given instant k is chosen as the process model, which governs the prediction we make about the next position. It is subsequently corrected using the measurement $dX_m(k+1)$, yielding a filtered position estimate $X_f(k+1)$.

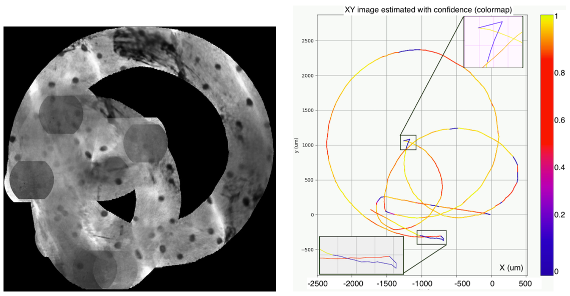
At places where the image quality is good, the image estimates are likely to be more precise than the robot ones. Conversely, when there are image losses, the robot input is likely to be more accurate. We encode this information in the process and estimation covariances. To do so, we estimate an image matching confidence c_{img}. We hypothesize that at places where the image matches are of good quality, they are locally consistent with one another in direction. Therefore, we estimate c_{img} by subtracting the angle of the $dX_m(k)$ vector with its median-filtered value over the last five time instances. This value is subsequently normalized to be between 0 and 1. It is then integrated in the Kalman Filter by dividing the measurement covariance values by c_{img}.

The method was validated using a Mauna Kea Cellvizio system for pCLE imaging, and a *Ficus Benjamina* leaf as a tissue sample. The latter was placed on a high-accuracy 6-degrees-of-freedom

robot, which was moved with respect to the probe to produce a relative probe-tissue movement. In order to simulate varying contact conditions (which, for instance, would be created by breath in *in vivo* conditions), the z-axis of the robot was controlled with an oscillating movement of amplitude 150μm around the ideal contact point during the scanning movement. Since the robotic setup is of high accuracy, the programmed spiral trajectory was perfectly executed. Therefore, noise was artificially added to the robot trajectory to produce the noisy robot trajectory $\hat{X}_r$. This trajectory is similar to the ones obtained in *in vivo* conditions with minimally invasive settings, such as reported in [2].
In order to obtain a ground truth validation, the *Ficus Benjamina* leaf was imaged under a conventional microscope with back-light. pCLE images were then registered to this ground truth using Mutual Information (*imregister* function of MATLAB).

RESULTS

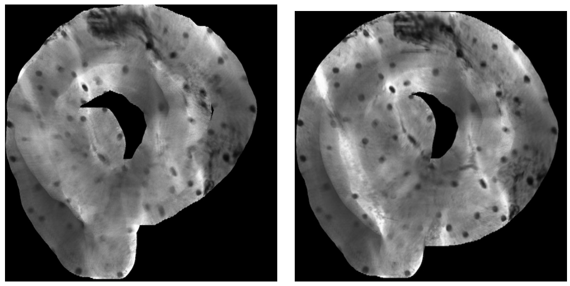
Figure 1 presents the mosaic obtained for a spiral trajectory when using only the image measurements. As expected, local image losses create estimation errors, which drive the overall mosaic topology very far from the input spiral trajectory. One can see on Fig. 1b that the estimated image confidence gets very low at places where the image trajectory is locally incoherent.



(a) Image-estimated mosaic (b) c_{img} values

Fig. 1: Reconstructed mosaic using only image measurements. Left: mosaic; Right: image-estimated trajectory X_m, with c_{img} shown as a colormap.

When using the noisy robot trajectory $\hat{X}_r$ (Fig. 2(a), the mosaic shape is partially recovered. However, branches of the spiral are superimposed, leading to a visually blurry mosaic very difficult to analyze. Finally, using the filtered output X_f for the mosaic yields the best result (Fig. 2(b)). Note that the mosaic could be further improved by offline bundle adjustment, but it is nonetheless important for the clinician to be able to assess the mosaic quality in real time [4].
Those results are confirmed by comparing the positions in the mosaic with the conventional microscopy ground truth. Using only the image estimates X_m, the average position error is very high (mean 686.7μm, std. dev. 1080.8μm). As expected, it decreases when using $\hat{X}_r$ (mean 291.0μm, std 134.9μm). The error is, however, high with respect to the pCLE images FOV (500μm). Finally, the filtered positions X_f exhibit the lowest errors (mean 100.8μm, std 63.9μm). Similar results were obtained using different trajectories (circle, line).



(a) Using $\hat{X}_r$ (b) Using X_f

Fig. 2: Reconstructed mosaics using robot measurements. Left: using the noisy robot inputs $\hat{X}_r$; Right: Using the filtered outputs X_f.

DISCUSSIONS

This abstract presents a method for robustly estimating the topology of online pCLE-based mosaics. Estimating the confidence of online image matches allows fusing them with noisy robot trajectory inputs, leading to improved mosaics, despite important local losses in image quality.
The method was validated on a bench-top setting using a high-accuracy robotic setup. Results show that mosaics obtained using the filtered outputs are of better quality, both in terms of visual appearance and with respect to the ground truth image positions. This is likely to help the surgeon perform real-time visual assessment, as well as to shorten computing time of subsequent bundle adjustment.
Further work will include validation on *ex vivo* and *in vivo* settings.

REFERENCES

[1] B. Rosa, et al. *Building large mosaics of confocal edomicroscopic images using visual servoing.* IEEE Trans. Biomed. Eng., 60(4), p. 1041-1049, 2013.

[2] M. S. Erden, et al. Conic-Spiraleur: A miniature distal scanner for confocal microlaparoscope. IEEE/ASME Trans. Mechatronics, 19(6), p. 1786-1798, 2014.

[3] Zhang, L., et al. *From Macro to Micro: Autonomous Multiscale Image Fusion for Robotic Surgery.* IEEE Robotics & Automation Mag., 24(2), p. 63-72, 2017.

[4] Vercauteren, T. et al. *Real time autonomous video image registration for endomicroscopy: fighting the compromises.* SPIE BIOS, vol. 6861, pages 68610C, 2008.