Building large mosaics of confocal endomicroscopic images using visual servoing

Benoît Rosa, Mustapha Suphi Erden, Tom Vercauteren, Benoît Herman,
Jérome Szewczyk, Guillaume Morel

HAL Id: hal-01765030
https://hal.science/hal-01765030

Submitted on 12 Apr 2018

Building large mosaics of confocal endomicroscopic images using visual servoing

Benoît Rosa[1], Mustapha Suphi Erden[1], Tom Vercauteren[2], Benoît Herman[1], Jérôme Szewczyk[1], and
Guillaume Morel[1]

Abstract—Probe-based Confocal Laser Endomicroscopy (pCLE) provides real-time microscopic images of tissues contacted by a small probe that can be inserted *in vivo* through a minimally invasive access. Mosaicking consists in sweeping the probe in contact with a tissue to be imaged while collecting the video stream and process the images to assemble them in a large mosaic. While most of the literature in this field has focused on image processing, little attention has been paid so far to the way the probe motion can be controlled. This is a crucial issue since the precision of the probe trajectory control drastically influences the quality of the final mosaic.

Robotically controlled motion has the potential of providing enough precision to perform mosaicking. In this paper, we emphasize the difficulties of implementing such an approach. Firstly, probe-tissue contacts generate deformations that prevent from properly controlling the image trajectory. Secondly, in the context of minimally invasive procedures targeted by our research, robotic devices are likely to exhibit limited quality of the distal probe motion control at the microscopic scale. To cope with these problems visual servoing from real-time endomicroscopy images is proposed in this paper. It is implemented on two different devices (a precise industrial robot and a prototype minimally invasive device). Experiments on different kinds of environments (printed paper and *ex vivo* tissues) show that quality of the visually servoed probe motion is sufficient to build mosaics with minimal distortion in spite of disturbances.

Index Terms—Medical robotics, mosaicking, probe-based confocal laser endomicroscopy, visual servoing.

I. Introduction

Probe-based Confocal Laser Endomicroscopy (pCLE), [1], [2], is an imaging modality that requires placing a probe in contact with a tissue and provides real-time images of micrometer resolution. This resolution is sufficient to perform *in vivo* pathology in replacement to tissue biopsy, which would prevent from physically sampling the tissue and sending it to a pathologist. This is of particular interest for extemporaneous biopsies, *i.e.* when a patient is receiving surgery in an operating room while the surgeon is waiting for the results of a pathologist tissue sample analysis in order to orient the procedure. Current extemporaneous tissue biopsy practice requires a minimum of 20 minutes for the tissue sample to be sent to the pathologist laboratory, prepared for microscope examination, and be analyzed whereas real-time pCLE could save time and limit tissue damage. Our research is aimed at developing means for deploying a pCLE probe in the abdomen through a 5 mm trocar in order to image tissues throughout the abdomen in replacement of extemporaneous biopsies. It includes designing a minimally invasive device. We recently presented a first prototype in [3]. This device is able not only to position and stabilize, at a microscopic level, the probe in contact with the tissues but also to generate a probe motion with respect to the tissue.

Generating a probe sweeping motion is a key feature that is expected to allow overcoming a main drawback of current pCLE for replacing extemporaneous biopsies: the limitation of the images field of view. Indeed, the size of an image with micrometer resolution is typically 200×240 μm. This makes it possible to perform analysis of a cell, or a few cells. However, pathologists need not only to analyze the cells by themselves but also their relative organization from a larger picture. State of the art requires a total surface of typically 3 mm^2, which is the surface of a square with 1.7 mm side.

The mosaicking approach has been proposed in the literature in order to generate wide field of view images with super-resolution from a video sequence containing overlapping images. The approach has been successfully applied on different imaging modalities: video cameras [4], [5], endoscopic images [6], [7], retinal micro-images [8] and pCLE [9]–[11]. It requires estimating the displacement between spatially close images that are not necessarily temporally close. Typically, a first estimation of the image positions in the mosaics is made by computing the displacement between consecutive images [6], [12], then an optimization is performed to account for cumulative errors, [7], [8], [10], [11].

Figure 1 shows an example mosaic constructed *in vivo* with the device presented in [3]. It comprises 60 images and its size is rather small as compared to our objective of building 3 mm^2 mosaics. In this paper, we target building square mosaics of

[1]UPMC Univ Paris 06, UMR 7222, ISIR, F-75005, Paris, France and CNRS, UMR 7222, ISIR, F-75005, Paris, France (rosa,erden,szewczyk,morel) at isir.upmc.fr

[2]Mauna Kea Technologies, F-75010, Paris, France tom at maunakeatech.com

B. Herman is now F.R.S.-FNRS Postdoctoral Researcher with UCL, Center for Research in Mechatronics, B-1348 Louvain-la-Neuve, Belgium benoit.herman at uclouvain.be.
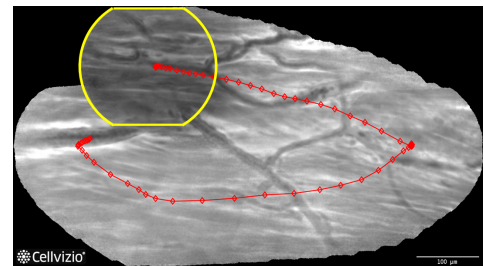
Figure 1.   *In vivo* mosaic of peritoneal tissue, [3]. Red line represents the position of the centers of the individual images registered thanks to a mosaicking algorithm proposed in [11]. Yellow line represents the limit of the first image. The continuity of the micro-vessels throughout the mosaic visually assesses for the quality of the mosaic.

1700 μm×1700 μm from 200 μm×240 μm individual images. For this, given the 100 μm overlap between two successive layers necessary to robustly compute the mosaics, and given the maximal velocity for proper imaging, we propose to follow a raster scan trajectory composed of sixteen 1500 μm parallel straight lines with a 0.2 mm/s velocity. This corresponds to a total duration of approximately 2 minutes for the acquisition of 1600 images. Clearly, it is not possible to pilot the probe motion manually, as it was done in [3] and robotic motion control is to be implemented.

Using robots to help generating mosaics has been proposed in the literature. For example a robot can be used as a sensor for the camera position and orientation, allowing real-time video mosaicking, see *e.g.* [5]. In the context of pCLE, robotic control has been investigated recently for image stabilization, [13], [14], but not to generate probe sweeping motion in the aim of mosaicking. Moreover, for mosaics involving micro displacements of a non contact camera, a high precision robot has been exploited in [6], [7] to generate the movements, using the Steady-Hand Robot as a camera holder. However, as it will be emphasized in Sec. III, this open-loop approach is inappropriate in the context of pCLE, where tissue deformations induced by the contact of the probe with the tissue result in distortions between the robot motion and the image motion, see [15]. To compensate for this disturbance, we propose in section IV a visual servoing method using the confocal images. This approach is successfully applied to both a precise industrial robot, presented in Sec. II and a minimally invasive instrument presented in Sec. V.

II. EXPERIMENTAL SETUP WITH A HIGH PRECISION ROBOT

The experimental platform is presented in Fig. 2. This platform is used for the development of our approach presented in sections III-IV. It consists of a pCLE system with its probe, a high precision robot for generating the probe movements, and a tissue to be imaged that is placed under the probe on a rigid surface.

The pCLE system is a Cellvizio device (Mauna Kea Technologies, Paris, France). It is composed of a central unit for laser scanning and data acquisition and treatment, and of an endomicroscopy probe. The probe used here is a Gastroflex UHD probe, which is made of a flexible fiber bundle (outer diameter 1.4 mm) and an optical head at its tip (outer diameter 2.6 mm). The excitation wavelength is 488 nm, and the contrast agent for fluorescence marking of the tissues is, for our *ex vivo* setup, Acriflavin. The images are acquired at 12 frames/s. They present a 1 μm lateral resolution and a field of view of 200 μm×240 μm. The depth of the focal plane is 50 μm under the tissue surface. This device is a contact imaging device, *i.e.* images are only acquired when the probe contacts the tissue surface.

The probe is fixed at the end effector of a TX40 robot (Stäubli, Faverges, France), through a prototype probe holder. The robot has 6 Degrees of Freedom (DoFs) and exhibits a 20 μm repeatability at its end effector. The tissue samples are put on a rough surface that is fixed with respect to the robot base and parallel to the horizontal plane. Considering
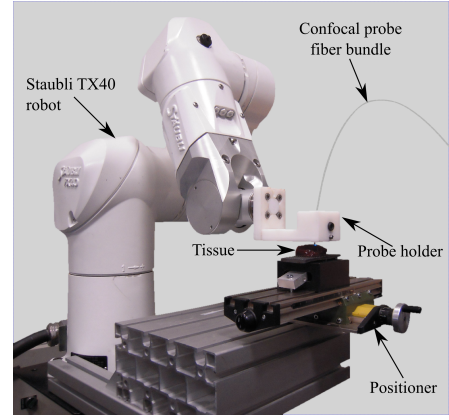

Figure 2. Test bench using a Stäubli TX40 robot and the Cellvizio probe.

the small size of the scanned areas (a few square millimeters), the scanned tissues are considered as planar in this study. Therefore, the robot movements are generated at a constant height z_0. Meanwhile, the probe orientation is kept constant, which finally results in a 2 DoFs planar positioning problem. The cartesian position of the optical head is controlled by the on-board controller of the TX40 robot which is interfaced for allowing real-time update.

The procedure for tissue experiments is the following: first, a piece of tissue is cut, and Acriflavin is delivered on the tissue surface for marking the first cellular layers. The robot is positioned with successive 100 μm steps down until an image appears on the Cellvizio screen. At this stage, we know that the probe has just contacted the tissue. A final 300 μm step down is made, which slightly compresses the tissue. With this procedure, it was experimentally verified that the normal elastic deformations of the tissue allows to guaranteeing that the contact is maintained during the scan despite any possible lack of planarity. Additionally, a drop of normal saline (pure water with 0.90% mass concentration of NaCl) is regularly delivered on the tissue surface in order to keep it hydrated. This is a standard procedure aimed at reproducing the contact conditions encountered during *in vivo* experiments despite *ex vivo* tissue drying.

III. OPEN-LOOP MOSAICKING

In order to build a large mosaic from the video stream provided by the Cellvizio system, it is required to scan the surface of the tissue to be imaged, while maintaining contact. The mosaicking algorithm is aimed at building one large image out of n small images recorded during the surface scan. It runs in two main phases. During a first phase, the displacement ζ_k in the image plane, between each acquired image $\mathcal{I}_k$ and its preceding image $\mathcal{I}_{k-1}$ is computed. Next, the displacements from the first image of the mosaic to the kth image are summed up in order to estimate the position of $\mathcal{I}_k$ in the mosaic with respect to the first mosaic image. This computation is light enough to be computable in real-time, at the Cellvizio image rate (12 images per second). The second phase is run off-line. It consists of a global optimization aimed at refining the estimation of the image positions in the mosaic

and uses as a first estimate the positions provided by the real-time mosaicking algorithm. Describing how this optimization works is out of the scope of the present paper; details can be found in [12].

A. Real-Time Mosaicking

To register two successive images we make the assumption that the motion of the probe can be modeled by a constant translation velocity on any time frame corresponding to the acquisition of a pair of successive images. This assumption allows us to perform image registration in real-time as presented in [12] with the drawback that only probe-tissue translations (*i.e.* no rotations) can be recovered. This limitation is not restrictive with the current setup, since the robot control imposes a translation of the probe. Following the approach of [12], fast normalized correlation matching [16] is used to estimate the translation between two consecutive images $\mathcal{I}_k$, $\mathcal{I}_{k+1}$. The main idea is to evaluate, in one pass with a Fourier transform, the correlation coefficient for every possible translation T of an image vector ζ having integer components:

$$\text{Sim}\left(\mathcal{I}_k, \mathcal{I}_{k+1} \circ T(\zeta)\right) =$$
$$\frac{\sum_j (I_k(j) - \bar{I}_k)(I_{k+1}(j+\zeta) - \bar{I}_{k+1})}{\sqrt{\sum_j (I_k(j) - \bar{I}_k)^2 + \sum_j (I_{k+1}(j+\zeta) - \bar{I}_{k+1})^2}} \quad (1)$$

where $\bar{I}_x$ is the mean of the pixel values in image $\mathcal{I}_x$, $I_x(.)$ is the value of pixel $(.)$ in image $\mathcal{I}_x$ and j stands for a pixel index. Given the full correlation coefficient map, we extract its maximum to get the optimal translation:

$$\hat{\zeta}(\mathcal{I}_k, \mathcal{I}_{k+1}) = \underset{\zeta}{\text{argmax}}\left(\text{Sim}\left(\mathcal{I}_k, \mathcal{I}_{k+1} \circ T(\zeta)\right)\right) \quad (2)$$

Fast normalized correlation matching provides globally optimal results, can be computed efficiently, and requires a fixed computational budget. This makes it an ideal candidate for real-time applications. Nonetheless, it has been designed for template matching rather than image matching and is theoretically correct only if the support of $\mathcal{I}_{k+1} \circ T$ is included in the support of $\mathcal{I}_k$, which is almost never fulfilled when working with images having the same size. In practice, this non-optimality leads to border effects and a lack of symmetry in the results, *i.e.* $\hat{\zeta}(\mathcal{I}_k, \mathcal{I}_{k+1}) \neq -\hat{\zeta}(\mathcal{I}_{k+1}, \mathcal{I}_k)$. To alleviate the lack of symmetry of the algorithm, we average the forward and backward estimation:

$$\hat{\zeta}^s(\mathcal{I}_k, \mathcal{I}_{k+1}) = \frac{\hat{\zeta}(\mathcal{I}_k, \mathcal{I}_{k+1}) - \hat{\zeta}(\mathcal{I}_{k+1}, \mathcal{I}_k)}{2} \quad . \quad (3)$$

As demonstrated in [12] this approach provides reliable registration results.

Given the estimation $\hat{\zeta}^s$ of the translation in pixels between two successive images, we now need to estimate the translation in μm on the tissue. Factory calibration of the probe provides the scaling factor s from which a first estimate of the velocity of the probe (subscript $_p$) with respect to the anatomical tissue (subscript $_a$) writes:

$$V_{p/a}^u = \frac{s}{T_{\text{acq}}}\hat{\zeta}^s \quad , \quad (4)$$

where T_{acq} is the acquisition time of an image. However, as explained in [11], the imager is a scanning device and compensating for the distortions of an image due to the motion is required. Following the derivations in [11] the estimated distortion-compensated velocity $\hat{V}_{p/a}$ is:

$$\hat{V}_{p/a} = \frac{1}{1 - \frac{(V_{p/a}^u)_y}{V_y^{\text{scan}}}}V_{p/a}^u = \frac{1}{1 - \frac{(V_{p/a}^u)_y}{V_y^{\text{scan}}}}\frac{s}{T_{\text{acq}}}\hat{\zeta}^s \quad (5)$$

where V_y^{scan} is the known vertical speed of the scanning and $(V_{p/a}^u)_y$ is the projection of $V_{p/a}^u$ on the y axis of the image frame.

The Cellvizio controller performs this computation in real-time and sends the result to the robot controller via an ethernet link. The integration of this data over time gives the current image position relative to the beginning of the trajectory:

$$\hat{X}_{p/a}(k) = \hat{X}_{p/a}(0) + \sum_1^k \hat{V}_{p/a}(k)T_{\text{acq}} \quad (6)$$

B. Mosaic Image Reconstruction

Given the estimated position $\hat{X}_{p/a}(k)$ of the probe for each image $\mathcal{I}_k$, we can reconstruct a large field of view mosaic image $\mathcal{M}$ as proposed in [4]. By mapping the pixels of all the single images into a common reference coordinate, we get the following point cloud:

$$\{(j_l^m, J_l^m)\} = \{(j + \hat{X}_{p/a}(k), I_k(j)) | j \in \mathcal{I}_k, k \in [1; n]\}. \quad (7)$$

The mosaic image pixel values $M(j)$ are reconstructed using a weighted average of all sampling points:

$$M(j) = \frac{\sum_l J_l^m . G(j_l^m - j)}{\sum_l G(j_l^m - j)} \quad , \quad (8)$$

where $G(.)$ is an isotropic Gaussian function centered on 0 whose parameter σ controls the smoothness of the reconstruction. In our experiments, we set it to 1 pixel. Note than only $\hat{X}_{p/a}(k)$ is reconstructed in real-time, the construction of $\mathcal{M}$ from Eqs. (7-8) is for display purpose and can be run asynchronously.

C. Experimental evaluation

Equation (6) provides a real-time measurement of the position of the probe with respect to the anatomical tissue. This measurement is subject to drift due to the possible accumulation of estimation errors through time integration.

In order to evaluate the drift, the following experiment is performed. On a sheet of plain paper, a black square grid with a 0.3 mm pitch is laser printed, see Fig. 3. A fluorescent highlighter is then applied to the paper. The paper fibers are then visible with the Cellvizio system, except when the black ink stops the light. Next, the robot is programmed to put the probe in contact with the paper and then to follow a 1 mm diameter circular trajectory with a 0.3 mm/s constant velocity. Meanwhile a real-time mosaic is computed. Figure 4 shows the obtained real-time mosaic manually superimposed on the grid picture. The effect of the drift is not visible in that sense that the grid pattern visible on the mosaic perfectly matches –
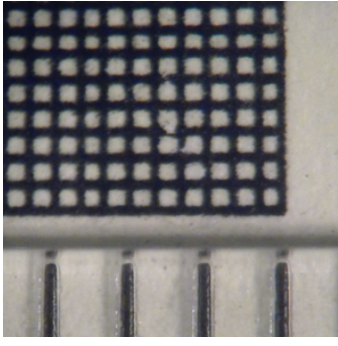
Figure 3. Picture of the grid used for real-time open-loop mosaicking experiments. Dashes on the bottom are those of a ruler with a 1 mm pitch.

at the scale of the picture – with the grid pattern visible on the macroscopic image.

In order to better quantify the drift at the microscopic scale, we take advantage of the overlap between the first image $\mathcal{I}_1$ and the last image $\mathcal{I}_n$. Because this operation is made off-line, time budget is not limited and we can compute, between these two images, both the translational and rotational displacement (respectively $\hat{X}^{\text{direct}}$ and $\hat{\alpha}^{\text{direct}}$) by maximizing the similarity criterion:

$$\left(\hat{X}^{\text{direct}}, \hat{\alpha}^{\text{direct}}\right) = \underset{\zeta,\alpha}{\text{argmax}} \left(\text{Sim}\left(\mathcal{I}_1, \mathcal{I}_n \circ \Delta(\zeta, \alpha)\right)\right) , \quad (9)$$

where $\Delta(\zeta, \alpha)$ is a displacement composed by a translation of a vector ζ and a rotation of an angle α. We then evaluate:

$$\begin{cases} \delta &= \left\| \hat{X}^{\text{direct}} - \hat{X}_{p/a}(n) \right\| \\ \beta &= \left| \hat{\alpha}^{\text{direct}} - \hat{\alpha}_{p/a}(n) \right| = \left| \hat{\alpha}^{\text{direct}} \right| \end{cases} , \quad (10)$$

since, by assuming that the motion is a pure translation, in order to reduce the real-time computational cost, we have set the on-line estimation of the angle $\hat{\alpha}_{p/a}(k) = 0, \forall k \in \{1 \cdots n\}$. The distance δ is an estimation of the translational drift due to integration in Eq. (6) while the angle β is used here to *a posteriori* verify the validity of the pure translation assumption.
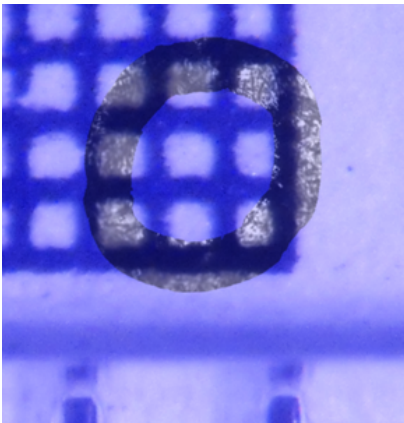

Figure 4. Mosaic of a known grid pattern printed on paper, obtained with a robot circular trajectory. The real-time mosaic (grey) is manually superimposed to the picture of the printed grid (blue).

Out of 5 trials, the mean estimation of the drift is $\bar{\bar{\delta}} = 9.1$ µm, with a maximum of $\delta_{\max} = 16$ µm, a minimum of $\delta_{\min} = 1.3$ µm and a standard deviation of $\sigma_\delta = 5.2$ µm. Such a drift is considered to be acceptable since it remains small as compared to the image size (200×240 µm) and the desired image overlap for mosaicking purposes (100 µm). The mean estimation of the final angle is $\bar{\beta} = 0.6$ deg, with a maximum of $\beta_{\max} = 1.0$ deg and a standard deviation of $\sigma_\beta = 0.4$ deg. This residual rotation is small enough to allow working in open-loop under the pure translation assumption as it will lead to subpixel displacements in the individual images.

D. Soft tissue deformation

A last experiment is now run in order to emphasize the effect of tissue deformation. A 1 mm diameter circular trajectory is commanded to the robot, while the probe is contacting a sample of chicken breast. The obtained real-time mosaic is presented in Fig. 5. Images $\mathcal{I}_1$ and $\mathcal{I}_n$ do not overlap anymore. This distortion can not be simply explained by a drift, as illustrated by the fact that the last and the first images do not show similar cellular structure. Rather, as experimentally evidenced and theoretically modeled in [15], this distortion is due to the deformations of the tissue generated by the tangential frictional forces between the probe and the tissue sample. The effect of these deformations can be estimated by computing the distance between the final image and the first image through integration: out of 12 trials, the mean final distance was estimated to be 111 µm, with a standard deviation of 95 µm, minimal value value of 13 µm and a maximal value of 301 µm, which is larger than one image. In average, during the motion, the distance between the position commanded to the robot and the estimated position of the image is 162 µm.

In summary, these open-loop experiments show that the real-time mosaicking is a mean of measuring in real-time the position of the probe with respect to the imaged sample. The observed measurement drift is small as compared to the geometric parameters of the mosaicking trajectory. However, for the tissue experiment, the image trajectory significantly differs from the robot commanded trajectory, due to tissue deformations. In the aim of building large mosaics, as required by the clinical application, these distortions may lead to unacceptable mosaics, exhibiting holes between the layers of images, and insufficient overlap between layers of images for

Figure 5. An open-loop real-time mosaic of a sample of chicken breast obtained with a robot circular trajectory.

image registration purposes. In order to control the image trajectory despite tissue deformations, we propose to exploit, in the next Section, the natural robustness properties of visual servoing by using the image-based measurement of the position described in Eqs. (1-6).

IV. VISUAL SERVO CONTROL OF THE PROBE POSITION

A. Extrinsic parameter calibration

To control the robot motion from the image-based measurement of the probe position, it is required to register the robot end-effector coordinates system with respect to the image coordinates system. In our setup, the $z-$axes of the robot end-effector frame and the probe frame are mechanically aligned. Therefore, for controlling the translational displacements of the probe in the $xy-$plane, it is sufficient to identify the angle θ between the robot $x-$axis (u_x) and the image $x-$axis (u_{xi}). This is straightforwardly achieved by running real-time mosaicking while the robot is programmed to follow a straight line along u_x. The resulting mosaic is a straight line in the image plane whose angle with respect u_{xi} is ($-\theta$).

Eight experiments are run. For each experiment a least square identification of the straight line is made out of the mosaic image coordinates and an angle θ is computed. The mean value is $\bar{\theta} = 162$ deg. with a standard deviation of 1.6 deg. and a maximum deviation of 4 deg. This precision is largely sufficient in the framework of visual servoing: in a similar hand-sensor configuration, robustness to a calibration error as large as 90 degrees for θ has been reported in the literature [17].

B. Control scheme

The proposed controller is aimed at computing the robot control input, which is the velocity $V_{r/0}$ of the robot end effector with respect to the fixed frame, in such a way that the estimated image position $\widehat{X}_{p/a}$ follows a given scanning trajectory $X_d(t)$. Usually, in visual servoing, when bandwidth is not an issue (which is the case here since the displacements are very slow), a simple proportional controller ensures an exponential convergence of the servoed error towards zero [18]. Additionnaly, a feedforward term is used when the desired image velocity is not null [19]. This approach is appropriate in conventional situations without disturbance. Here, tissue deformations under probe contact play the role of an external disturbance:

$$V_{p/a} = V_{p/0} - V_{a/0} \tag{11}$$
$$V_{p/0} = V_{p/r} + V_{r/0} \tag{12}$$

where $V_{p/0}$ and $V_{a/0}$ denote the velocities of the the probe (subscript $_p$) and the anatomical tissue (subscript $_a$) relative to a fixed base (subscript $_0$), respectively; $V_{p/r}$ denote the velocity of the probe relative to the the the robot.

Equation (11) highlights the fact that the probe motion relative to the tissue, $V_{p/a}$, depends not only on the probe motion relative to a fixed base $V_{p/0}$, but also on the tissue deformations $V_{a/0}$, as illustrated by the distortion observed in Fig. 5. Equation (12) highlights the fact that, relative to the
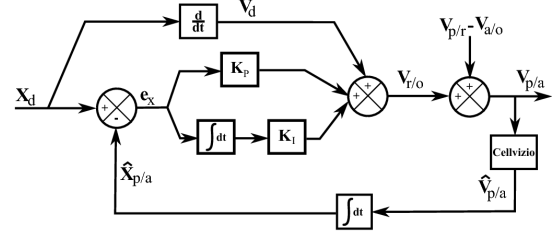

Figure 6. The proposed control scheme.

fixed frame, the robot velocity $V_{r/0}$ may differ from the probe velocity $V_{p/0}$ due to possible deformations of the probe holder $V_{p/r}$, as illustrated by the distortion observed in Fig. 4.

The disturbance is quasi-static since it changes slowly and it is even constant during constant velocity movements [15]. Therefore, a simple integral term can be used to reject it. The resulting control law becomes:

$$V_{r/0} = \frac{\mathrm{d}X_d}{\mathrm{d}t} + k_P \left(X_d(t) - \widehat{X}_{p/a}(t) \right) + \\ k_I \int_0^t \left(X_d(\tau) - \widehat{X}_{p/a}(\tau) \right) d\tau \tag{13}$$

where k_P and k_I are the proportional and the integral gains, respectively. They are tuned in a standard way, resulting in a proportional gain of $3\ s^{-1}$ and an integral gain of $0.15\ s^{-2}$. This provides a response time of less than 1 s with proper damping. The control scheme is presented in Fig. 6. It is comparable – except for the integral compensation – to the visual-servoing approach proposed in [20], where the current image position is also estimated using the integration of the measured image velocity.

C. Experimental results

For comparison with the results presented in Fig. 5 the 1 mm diameter circular experiment is performed on a piece of chicken breast in the closed loop mode. Figure 7 shows the $X_d(t)$ and $\widehat{X}_{p/a}(t)$ signals along x and y directions. Out of 3 trials, the mean tracking error during the whole trajectory is 39 μm, with a maximum value of 87 μm.

The mosaics obtained during this experiment are depicted in Fig. 8. Thanks to visual servoing, the first and the last images now overlap, contrarily to the open-loop result shown in Fig. 5. Thus, the drift of the measurement can be estimated. Out of
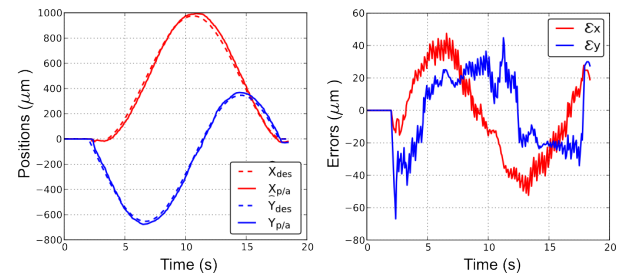

Figure 7. Data recorded during the realization of a 1 mm diameter circular mosaic on chicken breast under visual servoing. Left: time evolution of the desired and actual values of the image coordinates. Right: time evolution of the servoed error.
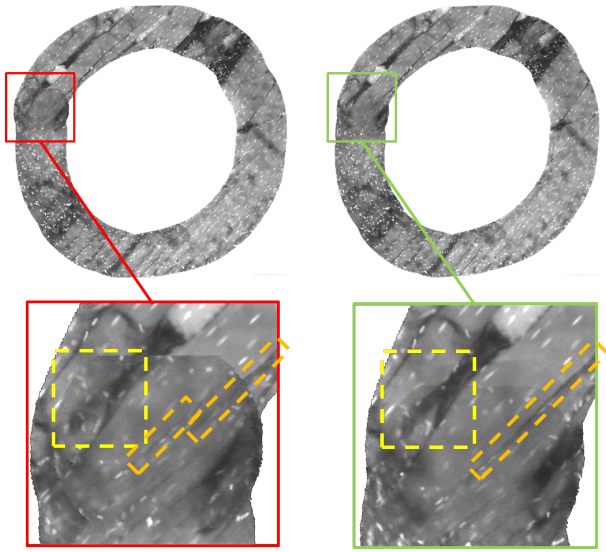
Figure 8. Mosaics of a sample of chicken breast obtained under visual servoing with a desired circular image trajectory. Left: real-time mosaic; right: optimized mosaic. Top: whole mosaic; bottom: zoom of the overlapping region. Detail comments in the text.

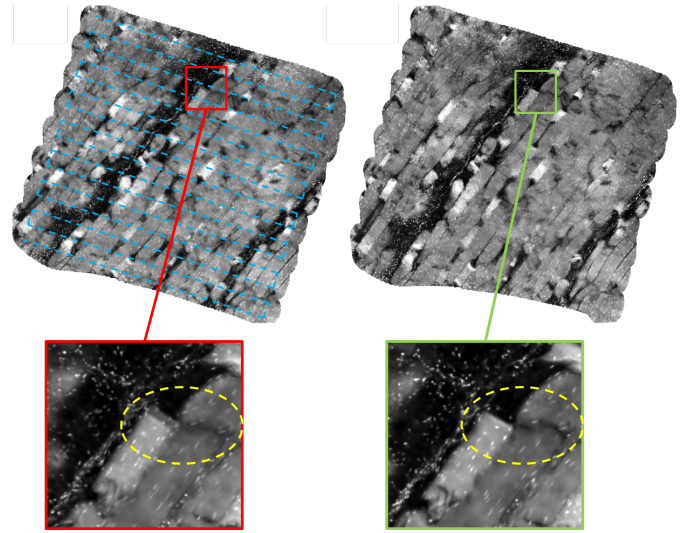Figure 9. Square 1.7×1.7 mm raster scan mosaics constituted with 1600 images obtained under visual servoing (chicken breast). Left: real-time mosaic with the raster scan desired trajectory in dashed blue; right: optimized mosaic. Zooming on details show that, for the optimized mosaic, the precision of the off-line registration allowed by proper layer overlapping leads to see microscopic details with reduced blur, for example in the yellow dashed ellipse.

3 experiments, the quantified translational drift was similar to the statistics of the paper experiments: $\bar{\delta} = 11$ μm, $\delta_{\max} = 20$ μm and $\delta_{\min} = 6.3$ μm. The mean angular deformation was $\bar{\beta} = 2.9$ deg, with $\beta_{\max} = 4.0$ deg and $\beta_{\min} = 2.1$ deg. This is slightly larger than the results obtained for the paper experiments, which is interpreted as a possible small rotational deformation of the tissue.

Moreover, quantifying the drift by a direct registration of $\mathcal{I}_1$ and $\mathcal{I}_n$ allows for computing an optimized mosaic accordingly, see [12] for details. In Fig. 8, it can be visually observed that, when zooming the overlapping region of the real-time mosaic (Fig. 8 – left) some details are blurred (yellow square) and some structures are misaligned (orange rectangles). Blur is due to the drift in the position estimation combined to the action of the filter used to build the mosaic image with Eqs. (7-8). Rather, in the optimized mosaic, which is observable in the right part of Fig. 8, detail structures in the yellow square appear more clearly while misalignments in the orange rectangle have been corrected.

The next experiment is a mosaic obtained by servoing the image to follow the 1.7 mm×1.7 mm raster scan desired trajectory described in Sec. I. The two mosaics (real-time and optimized) are presented in Fig. 9. Since, in the desired trajectory, $\mathcal{I}_1$ and $\mathcal{I}_n$ do not overlap, the estimation of the drift δ made for circular trajectories can not be performed here. However, the mosaic optimization algorithm can be run, taking advantage of multiple overlaps between images of two successive layers. This results in an optimized mosaic that is not subject to drift. The quantified mean distance between each individual image localization in the real-time mosaic and in the optimized mosaic is 50.9 μm, with a maximum of 160 μm and a standard deviation of 35.3 μm. These numbers are to be compared with the large size of the mosaic (1700 μm×1700 μm). Between the real-time mosaic and the optimized mosaic, the mean angular error was $\bar{\beta} = 3.1$ deg,

$\beta_{\max} = 8.5$ deg, with a standard deviation of 2.1 deg. Note than between successive images, the maximal angular displacement is only 0.4 deg. with a standard deviation of 0.4 deg. This justifies *a posteriori* the assumption that $\widehat{V}_{p/a}$ can be computed in real-time without taking rotations into account. From a clinical point of view, the interest of optimizing mosaics raises from blur reduction, which is visible in the zoomed regions at the bottom of Fig. 9.

V. APPLICATION TO A MINIMALLY INVASIVE SETUP

In the previous sections the visual servoing has proven to be efficient in rejecting the disturbances due to tissue deformations. However, the precise robot device that was used for this demonstration is not compatible with the targeted minimally invasive application. A minimally invasive device such as the one we have presented in [3] is likely to provide less precise transmission of the motion to the probe. In fact, this can also be viewed as an extra motion disturbance that can be rejected by the visual servoing approach.

The device we have presented in [3] features hydraulic actuation of the probe, along with a stabilization mechanism that solves the problem of physiological organ motion. The efficacy of the probe actuation was proven in an *in vivo* experiment, leading to the mosaic shown in Fig. 1. This mosaic was obtained by manually controlling, in open-loop, the probe motion through the hydraulic actuation. It was not possible to obtain larger mosaics with this mode of operation because of the difficulty of controlling the probe motion in open-loop.

A. Experimental setup

The tip of the micropositionner prototype represented in Fig. 10, integrated in a 5.45 mm outer diameter tube, is attached to a rapidly prototyped holder that is tightened to a
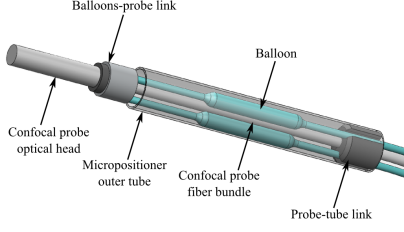
Figure 10. CAD view of the balloon actuation prototype. Only the distal part of the instrument is presented here, see [3] for more details.

manual micropositioning stage. It is used for depth calibration procedure similarly to section II. The probe movements are generated using differential volumes in the three balloons. The volumes of the three balloons are controlled using three syringes filled with water, and three linear stages that have 5 μm precision. For further details, please refer to [3].

In the following, we consider the three frames presented in Fig. 11-left. (u_x, u_y) is a frame attached to the micropositioner base body, (u_{xi}, u_{yi}) is the image frame, and u_1 (resp. u_2, u_3) is the direction along which the probe moves when balloon 1 (resp. 2, 3) is actuated alone from the central position.

B. Inverse kinematic model

To properly control the position of the probe, a model linking the velocities $(\dot{x}_1, \dot{x}_2, \dot{x}_3)$ of the three linear stages to the probe tip velocity $V_{r/0} = (V_x, V_y)$ is described.

First, for the small displacements involved, a linear dependency between balloon volumes and probe position is assumed. Second, we use hydraulic actuation. Since water can be considered as incompressible, the velocity of the linear stages is proportional to the time derivative of the balloon volume. Then the model writes :

$$\begin{pmatrix} V_x \\ V_y \end{pmatrix} = a \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} B \begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{pmatrix}, \qquad (14)$$

where a is a constant scaling the velocities of the linear stages into the velocity of the probe; θ is the angle from u_x
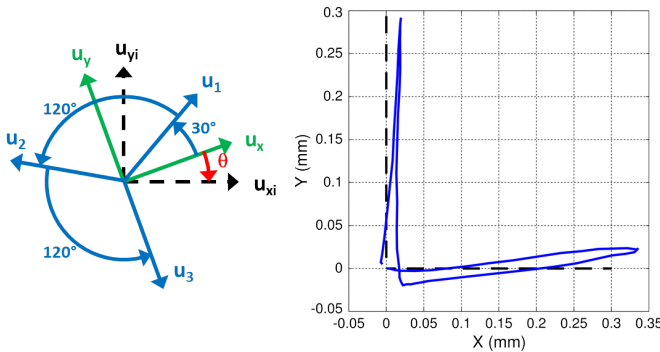

Figure 11. Left: frames attached to the balloon device and the image; right: image position recorded when scanning a sheet of paper with open-loop hydraulic actuation (blue solid: image trajectory ($\hat{X}_{p/a}$) – black dashed: reference trajectory).

to u_{xi} and B is a matrix whose columns are the projections of u_1, u_2 and u_3 in the (u_x, u_y) frame:

$$B = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{-\sqrt{3}}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 \end{pmatrix} . \qquad (15)$$

Equation (14) is a system of two equations and three unknowns. Therefore, another equation is needed to solve it. In [3], we have shown that imposing a constant sum of the three balloon volumes allows to maintain enough pressure from the balloons to the probe. Since balloon volumes are proportional to the positions of the linear stages, we thus impose $\dot{x}_1 + \dot{x}_2 + \dot{x}_3 = 0$. Under this constraint, Eq. (14) can be inverted as:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{pmatrix} = \frac{1}{3a} \begin{pmatrix} \sqrt{3} & 1 \\ -\sqrt{3} & 1 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} V_x \\ V_y \end{pmatrix} . \qquad (16)$$

Calibrating this model requires to identify the parameter a and the angle θ. This is done through an experimental procedure that is similar to the one described in Sec. IV-A: a constant velocity $V_x = V_0$ is programmed and converted into linear stages velocities thanks to Eq (16), where θ is arbitrarily set to 0. Meanwhile, the probe is contacting a rigid fluorescent environment and the real-time mosaicking is used to reconstruct the image trajectory. This trajectory is fitted with a straight line whose angle with respect to u_{xi} equals $-\theta$. Moreover, the average magnitude of the image velocity along the straight line allows to identify the parameter a.

Figure 11 illustrates the validity of the identified model through an open-loop experiment. It plots in the image frame the coordinates of the probe with respect to the environment, $\hat{X}_{p/a}$ reconstructed from real-time mosaicking. The input motion, from $X_d = (0, 0)$, consists of a constant positive velocity along u_{xi}, followed by a constant negative velocity along u_{xi}, and then the same sequence along u_{yi}. The obtained trajectories are not perfect straight lines, reflecting the modeling approximations. Here again, the robustness properties of visual servoing can be exploited to better control the probe motion in spite of the model uncertainties.

The control law combines Eq. (13) to compute $V_{r/0}$ from X_d and $\hat{X}_{p/a}$ and Eq. (16) to compute $(\dot{x}_1, \dot{x}_2, \dot{x}_3)$ from $V_{r/0} = (V_x, V_y)$. The control gains are the same as the one used for the Stäubli robot experiments.

C. Experimental Results

The last experiment is a raster scan mosaic executed with the minimally invasive device on deformable tissue (beef liver). Because of the limited range of the device, the scanning path is smaller than the previous experiments. The surface of the corresponding mosaic is 800×800 μm, see Fig. 12. The trajectory starts from the center of the device workspace and is first directed toward $X_d = (-300, -300)$ before the raster scan is performed. The mean tracking error is significantly larger than with the industrial robot due to modeling approximations in the device kinematic model. The distance from the individual images center to the desired path (geometric error) is in average 31 μm, with a maximal value of 90 μm and a
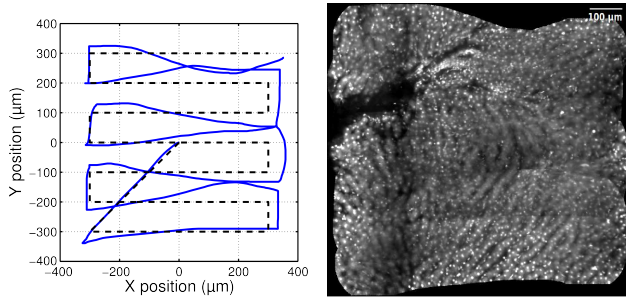
Figure 12. Raster scan mosaic made with the minimally invasive setup under visual servoing. Left: X_d (dashed dark) and $\hat{X}_{p/a}$ (blue) in the $xy-$plane. Right: real-time mosaic.

standard deviation of 19 μm. Tracking performance ensures enough overlapping between layers of images over the entire area, see Fig. 12: the shape of the mosaic is close to the expected square shape, and it does not exhibit any hole. This result shows that the proposed visual servo control algorithm is robust enough to compensate for both the modeling errors of the actuator and the tissue deformations at the same time.

VI. CONCLUSION

Building large mosaics from pCLE images is not only a problem of image processing. The control of sweeping probe movements is to be considered carefully. The required sub-millimeter precision excludes manual operation.

In this paper, robot control of the probe displacement has been investigated. We have shown that controlling these movements in open-loop may be inappropriate: the tissue deformation (Fig. 5) and/or the lack of precision in the distal probe motion control exhibited by a minimally invasive device (Fig. 11) result in large discrepancies between the input motion and the image motion.

To solve these problems, we have proposed the visual servoing approach which is aimed at directly controlling the displacements in the image plane. This approach was implemented on two devices: a high precision industrial robot and a prototype minimally invasive device. Two properties have been experimentally demonstrated. First, the drift inherent to the real-time image localization algorithm has been quantified and compared to the precision required for mosaicking. From these experiments, we conclude that the drift is compatible with the precision needs. It can thus be used for visual servoing purposes. Second, the disturbance rejection properties of the visual servoing has been proven to allow for proper control of the probe despite tissue deformations and distortions due the device actuation mode. This results in the ability of building mosaics that are significantly larger (1600 images with the precise robot; 1200 images with the minimally invasive device) than the largest pCLE mosaic of tissue found in the literature (300 images in [10] with manual operation).

Future works will be directed towards the exploration of the abdominal cavity during surgery. It is expected that the *in vivo* operation will challenge the robustness of the approach. Robustness will be increased firstly at the control level, by the inclusion of a disturbance model established from the deformation model (see [15]). Work will also be done on the design of an optimized reference trajectory, that will provide more smoothness.

REFERENCES

[1] M. Wallace, P. Fockens *et al.*, "Probe-based confocal laser endomicroscopy." *Gastroenterology*, vol. 136, no. 5, p. 1509, 2009.

[2] Y. Sabharwal, A. Rouse, L. Donaldson, M. Hopkins, and A. Gmitro, "Slit-scanning confocal microendoscope for high-resolution in vivo imaging," *Applied optics*, vol. 38, no. 34, pp. 7133–7144, 1999.

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

[4] D. Capel, "Image Mosaicing and Super-resolution," Ph.D. dissertation, University of Oxford, 2001.

[5] K. Loewke, D. Camarillo, C. Jobst, and J. Salisbury, "Real-time image mosaicing for medical applications." *Studies In Health Technology And Informatics*, vol. 125, pp. 304–309, 2007.

[6] S. Seshamani, W. Lau, and G. Hager, "Real-time endoscopic mosaicking," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pp. 355–363, 2006.

[7] S. Seshamani, M. Smith, J. Corso, M. Filipovich, A. Natarajan, and G. Hager, "Direct global adjustment methods for endoscopic mosaicking," in *Proceedings of SPIE*, vol. 7261, 2009, p. 72611D.

[8] A. Can, C. Stewart, B. Roysam, and H. Tanenbaum, "A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: Mosaicing the curved human retina," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 3, pp. 412–419, 2002.

[9] V. Becker, T. Vercauteren, C. H. von Weyhern, C. Prinz, R. M. Schmid, and A. Meining, "High-resolution miniprobe-based confocal microscopy in combination with video mosaicing," *Gastrointestinal Endoscopy*, vol. 66, no. 5, pp. 1001–1007, Nov. 2007.

[10] K. E. Loewke, D. B. Camarillo, W. Piyawattanametha, M. J. Mandella, C. H. Contag, S. Thrun, and J. K. Salisbury, "In vivo micro-image mosaicing," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 159–171, Jan. 2011.

[11] 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, Oct. 2006.

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

[13] W. Latt, R. Newton, M. Visentini-Scarzanella, C. Payne, D. Noonan, J. Shang, and G. Yang, "A hand-held instrument to maintain steady tissue contact during probe-based confocal laser endomicroscopy," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 9, pp. 2694–2703, 2011.

[14] D. Noonan, C. Payne, J. Shang, V. Sauvage, R. Newton, D. Elson, A. Darzi, and G. Yang, "Force adaptive multi-spectral imaging with an articulated robotic endoscope," in *Proc. IEEE International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI'10)*, 2010, pp. 245–252.

[15] M. S. Erden, B. Rosa, J. Szewczyk, and G. Morel, "Understanding soft tissue behavior for microlaparoscopic surface scan," in *Intelligent Robots and Systems (IROS), 2012 Int. Conf. on*.

[16] J. P. Lewis, "Fast template matching," in *Proceedings of the International Conference on Vision Interface (VI'95)*, 1995, pp. 120–123.

[17] C. Taylor, J. P. Ostrowski, and S. hack Jung, "Robust visual servoing based on relative orientation," in *In International Conf. on Robotics and Automation*, 1999, pp. 574–580.

[18] F. Chaumette and S. Hutchinson, "Visual servo control, part I: Basic approaches," *IEEE Robot. Autom. Mag.*, vol. 13, no. 4, pp. 82–90, 2006.

[19] ——, "Visual servo control, part II: Advanced approaches," *IEEE Robot. Autom. Mag.*, vol. 14, no. 1, pp. 109–118, 2007.

[20] A. Crétual and F. Chaumette, "Visual servoing based on image motion," *The International Journal of Robotics Research*, vol. 20, no. 11, pp. 857–877, 2001.