A Viterbi Approach to Topology Inference for Large Scale Endomicroscopy Video Mosaicing

Jessie Mahé[1], Tom Vercauteren[1], Benoît Rosa[2], and Julien Dauguet[1]

[1] Mauna Kea Technologies, Paris, France
julien.dauguet@maunakeatech.com
[2] ISIR, UPMC-CNRS UMR7222, Paris, France

Abstract. Endomicroscopy allows in vivo and in situ imaging with cellular resolution. One limitation of endomicroscopy is the small field of view which can however be extended using mosaicing techniques. In this paper, we describe a methodological framework aiming to reconstruct a mosaic of endomicroscopic images acquired following a noisy robotized spiral trajectory. First, we infer the topology of the frames, that is the map of neighbors for every frame in the spiral. For this, we use a Viterbi algorithm considering every new acquired frame in the current branch of the spiral as an observation and the index of the best neighboring frame from the previous branch as the underlying state. Second, the estimated transformation between each spatial pair previously found is assessed. Mosaicing is performed based only on the pairs of frames for which the registration is considered successful. We tested our method on 3 spiral endomicroscopy videos each including more than 200 frames: a printed grid, an ex vivo tissue sample and an in vivo animal trial. Results were statistically significantly improved compared to reconstruction where only registration between successive frames was used.

1 Endomicroscopy during Surgical Intervention

Probe-based Confocal Laser Endomicroscopy (pCLE) is an imaging technique that provides in vivo video sequences of soft tissues at cellular level [8]. The work presented in this paper is part of a gastrointestinal surgery project where we aim to perform an *optical biopsy* during the procedure using pCLE. Like most microscopy imaging techniques, pCLE offers high resolution images at the expense of the field of view. Mosaicing techniques can be used to extend the field of view by stitching series of overlapping images and create a large field of view image. Mosaicing algorithms can be separated in two classes. In the first category are methods with no a priori on the acquisition trajectory (e.g. handheld microscopes) based on topology inference [6] where the configuration of the frames has to be entirely recovered from registration [1,4,7]. These methods are powerful but by definition do not take advantage of any topology information which tends to make them less robust and more computationally demanding on long videos. In the second category are methods adapted for known - usually robotized - acquisition trajectories (e.g. microscope with motorized platform) where the a

K. Mori et al. (Eds.): MICCAI 2013, Part I, LNCS 8149, pp. 404–411, 2013.

priori knowledge directly provides the topology [3]. For our project, a dedicated device was designed to allow video mosaic acquisitions wherein a robot holds the probe and follows a pre-defined spiral trajectory [2]. The theoretical trajectory given as a command to the robot can be actually very disturbed due to tissue friction and mechanical distortions (see Fig. 5, right). In a similar context, the mechanical perturbations were considered so dramatic that the estimation of the transformation between successive frames actually served as a velocity sensor to control the robot [5]. For mosaic reconstruction of videos acquired with our setup, theoretically we fall into the category of known trajectory. However, surgical conditions imply that we can not entirely rely on the trajectory information to infer the acquisition topology. Moreover, the actual neighbors might offer very little overlap preventing from performing reliable registration. The contribution of this paper is to propose a strategy to tackle the problem of mosaic reconstruction with weak a priori on the trajectory. We propose a three step method: first we infer the frames topology using Viterbi algorithm, second we estimate for each neighboring frames previously found the quality of the registration, and finally we rely on the best associations only to perform the final mosaic reconstruction.

2 Material and Method

2.1 Material

We used three film sequences acquired following the same spiral trajectory. The first sequence is a test sequence; it was acquired on a sheet of white paper with a printed image of a regular black grid using an industrial robot. The second sequence was acquired ex vivo on chicken breast using the same industrial robot. The last sequence was acquired in vivo on pig liver using the surgical actuator. Trajectories were three loop Archimedean spirals of polar equation $r = a\theta$ with $a = d_0/2\pi$ where $d_0 = 150\mu m$ is the theoretical inter-branch distance. The field of view of the mini-probe is approximately circular with a size of $200\mu m$, which implies a maximum theoretical overlap of $50\mu m$ between two frames from successive branches belonging to the same radius (see Fig. 1).

2.2 Method

The baseline information most video mosaicing algorithms rely on is the temporal registration between successive frames. Although fallible, this registration is considered sufficiently reliable since successive frames usually offer good overlapping provided that the speed of the probe is low enough. The goal of the topology inference is to add some extra associations between frames that are not temporal neighbors to constrain the global reconstruction and to prevent error propagation. We will refer to these extra associations as spatial neighbors. In this work, we will consider that the transformation between frames can be modeled by a translation estimated by maximization of the absolute value of correlation coefficient: we will refer to the estimated transformation as the *best* translation in the following.

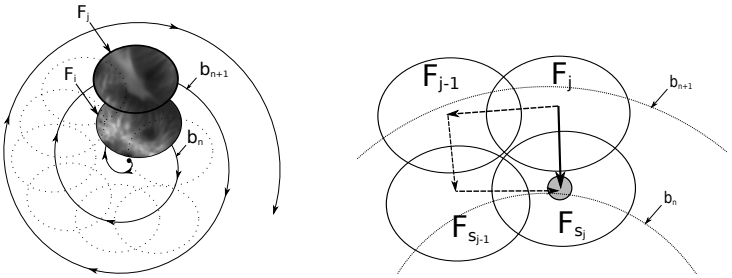

Fig. 1. The configuration of the spiral trajectory acquisition (left). The principle of checking the current spatial transformation by checking the registration consistency with previous frames (right).

Viterbi Algorithm. The first part of the method consists of estimating the spatial neighbor frames. Let us consider a hidden Markov model where at a given time index j, the observation is the current frame F_j and the hidden state s_j is the index of the frame in the previous branch that has the largest overlap with F_j (see Fig. 1). The frame F_{s_j} corresponding to the hidden state s_j will also be referred to as the *best match* for F_j. Following the Viterbi algorithm, we aim at recovering the most probable sequence of hidden states, $s^{0\to j} = \{s_0, s_1, \cdots, s_j\}$, given the sequence of observed frames, $F^{0\to j} = \{F_0, F_1, \cdots, F_j\}$:

$$\hat{s}^{0\to j} = \arg\max_{s^{0\to j}} P(s^{0\to j}|F^{0\to j}). \tag{1}$$

Let $\delta_j(i) = \max_{s^{0\to j-1}} P(s^{0\to j-1}, s_j = i, F^{0\to j})$ be the probability of the best sequence of states ending at state $s_j = i$. Similarly to the standard Viterbi algorithm but keeping the complete set of observations in the emission probability we find that:

$$\delta_j(i) = P(F_j|s_j = i, F^{0\to j-1}) \cdot \max_{i'} \left[P(s_j = i|s_{j-1} = i') \cdot \delta_{j-1}(i')\right]. \tag{2}$$

Thanks to (2), the dynamic programing approach of the Viterbi algorithm allows us to solve for (1) by keeping back-pointers to the antecedent of each possible last hidden state. In this work, the transition probability P_j^T is taken such that, if F_j has F_i as *best match*, i.e. $s_j = i$, the most likely a priori *best match* for F_{j+1} will be F_{i+1}, i.e. $s_{j+1} = i + 1$:

$$P_j^T(i, i') = P(s_{j+1} = i'|s_j = i) = \mathcal{N}(i'; i + 1, \sigma_T), \tag{3}$$

where $\mathcal{N}(i'; i + 1, \sigma_T)$ is the Gaussian probability function of mean $i + 1$ and standard deviation σ_T, with σ_T controlling the strength of the a priori on the temporal smoothness of the sequence of states. The emission probability P_j^E is chosen such that a frame F_j is likely to have F_i as *best match* if the registration between F_j and F_i leads to a good similarity criterion and to a transformation that is close to the expected one:

$$P_j^E(i, F_j) = P(F_j|s_j = i, F^{0\to j-1}) = \mathcal{N}(d_{ij}; d_0, \sigma_D) \cdot Corr(F_i, F_j), \tag{4}$$

where d_{ij} is the euclidean distance between frames F_i and F_j, d_0 is the expected distance between successive branches, σ_D reflects the expected precision of the robot and $Corr(F_i, F_j)$ is the best absolute value of the correlation coefficient when registering frames F_i and F_j with translation transformations.

Filtering of Associations. We deduct from the Viterbi path the global topology providing spatial associations between frames of the video sequence. However, some of these associations might be erroneous and even correct associations might not offer sufficient overlap and features for successful registration. To only keep pairs of frames that are both actual neighbors and successfully registered, we operate a filtering of the pairs. For a given pair of frames F_i of branch b_n and F_j of branch b_{n+1} supposedly spatial neighbors, we compute the following transformation consistency criterion TC: $TC_{ij} = \|T_{ji} - T_{(i-1)i} \circ T_{(j-1)(i-1)} \circ T_{j(j-1)}\|$ where T_{ji} is the best estimated translation between F_i and F_j (see Fig. 1). We then evaluate the reliability of the association between frames F_i and F_j based on three criteria: the transformation consistency TC_{ij}, the absolute value of the correlation coefficient CC_{ij} and the distance consistency $DC_{ij} =\mid d_0 - d_{ij} \mid$. We only keep spatial pairs respecting one of the following conditions:

- $TC_{ij} < TC_{strict}$ and $CC_{ij} > CC_{loose}$ and $DC_{ij} < DC_{loose}$
- $TC_{ij} < TC_{loose}$ and $CC_{ij} > CC_{strict}$ and $DC_{ij} < DC_{loose}$
- $TC_{ij} < TC_{loose}$ and $CC_{ij} > CC_{loose}$ and $DC_{ij} < DC_{strict}$
- $TC_{ij} < TC_{mid}$ and $CC_{ij} > CC_{mid}$ and $DC_{ij} < DC_{mid}$

where subscript tags $\{strict,\ mid,\ loose\}$ refer to $10^{th}, 20^{th}, 30^{th}$ percentiles of the set of values TC, DC and $90^{th}, 80^{th}, 70^{th}$ percentiles of the set of values CC estimated for all the pairs of frames derived from the Viterbi path and ranked in increasing order.

Mosaic Reconstruction. For the final mosaic reconstruction, we modified the mosaicing algorithm described in [7] so as to be able to inject the spatial pairs previously obtained to impose the topology. The rest of the algorithm remains unchanged: the best transformation is estimated between successive frames (temporal neighbors) as well as between given spatial pairs. A robust Fréchet mean of all the transformations - temporal and spatial - is then computed leading to the estimation of one unique transformation to the common reference per frame. Each point of each frame is then transformed to populate the final common reference. A smooth scattered data approximation is finally performed to construct the final mosaic image from the irregularly sampled point distribution.

Validation. We compared our results to results obtained using the mosaicing algorithm as described in [7]. Standard topology inference did not work in a satisfactory way on our large spiral datasets[1]. We therefore used translation mode with no topology inference as the baseline method for comparison.

[1] cf. supplemental material at `http://hal.archives-ouvertes.fr/hal-00830447`

As a ground truth for the paper grid, we acquired an image using a regular benchtop confocal microscope with a motorized platform performing raster scanning and we performed subsequence mosaic reconstruction using built in software from the microscope (see Fig. 3, left). We performed affine registration between the benchtop confocal image and the results of the mosaic reconstruction obtained with the baseline algorithm and our proposed method. We then estimated the correlation coefficient between each frame of the reconstructed mosaic and the confocal image for both methods.

For validation purpose of the chicken breast and pig liver reconstruction, since no benchtop confocal image was available as a reference, we relied on an *oracle* approach by manually infering the topology through visual assessment and we then injected the spatial associations obtained into the mosaicing algorithm. The reconstruction obtained was taken as reference and will be referred to as the *oracle* reconstruction. For both the baseline algorithm and our proposed method, we estimated the displacement between successive frames and compared it to the *oracle* results. More precisely, let X_i^O, X_i^B, X_i^V be the positions of frame F_i in the final mosaic image obtained using the *oracle*, the baseline algorithm and our proposed Viterbi framework. we computed for each frame $\Delta_i^B = (X_{i+1}^B - X_i^B) - (X_{i+1}^O - X_i^O)$ and $\Delta_i^V = (X_{i+1}^V - X_i^V) - (X_{i+1}^O - X_i^O)$.

3 Results

In Fig. 2, we present for the three video sequences the result of the emission probability estimation for all frames versus all frames of the video. We overlaid the Viterbi path in white (for all images) and the *oracle* path in red obtained manually by visual assessment of overlapping frames (for the chicken and liver acquisition only where no ground truth image was available). All estimations were computed using $\sigma_T = 2$ frames for the transition probability. We set $\sigma_d = 15\mu m$ for the industrial robot trajectories (grid and chicken breast) and $\sigma_d = 400\mu m$ for liver acquisition using the in vivo manipulator (where branches can

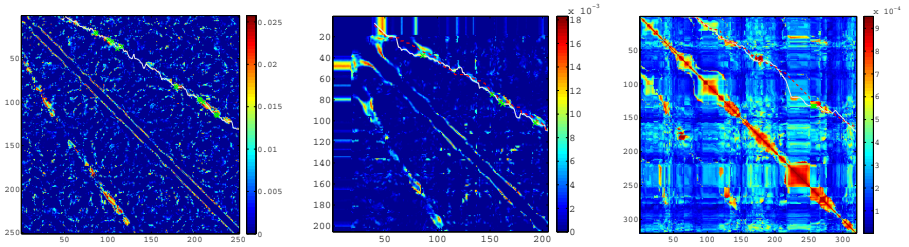

Fig. 2. Emission probability matrices for the grid (left), the chicken breast (middle) and the pig liver (right) spiral video sequences. The white line is the Viterbi path found. The red line is the oracle path obtained by visual identification of corresponding frames between successive branches (middle and right images only): the green crosses are the selected associations.

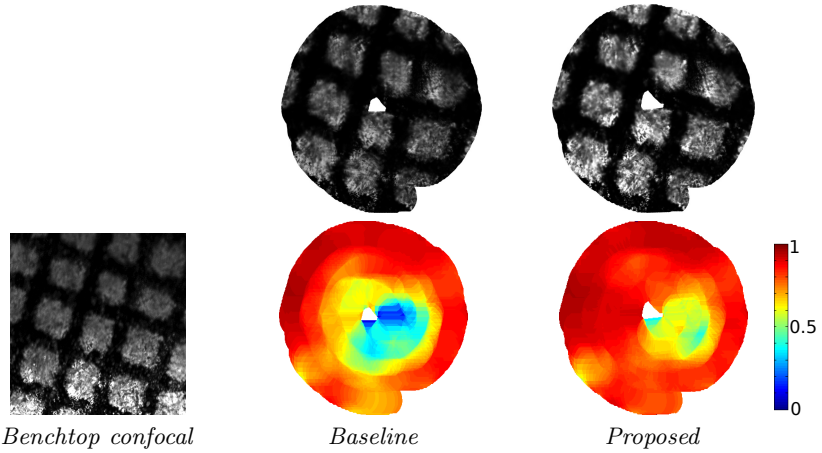

Fig. 3. Image representing the value of the correlation coefficient of each frame with the reference benchtop confocal image of the grid for the mosaic reconstruction using the baseline algorithm and using our proposed method

intersect). The selected pairs obtained after the filtering step are indicated with green crosses.

We also compared the mosaic reconstruction of the grid image obtained with the baseline algorithm and our method to the reference image of the grid acquired on a benchtop confocal microscope. For each method, we estimated the correlation coefficient of every frame with the registered reference confocal image. We then compared the sets of correlation coefficients for each frame of the baseline reconstruction and of our method using a sign test. Correlation coefficients proved significantly higher for our method (p-value=4.7037e-20). In Fig. 3,

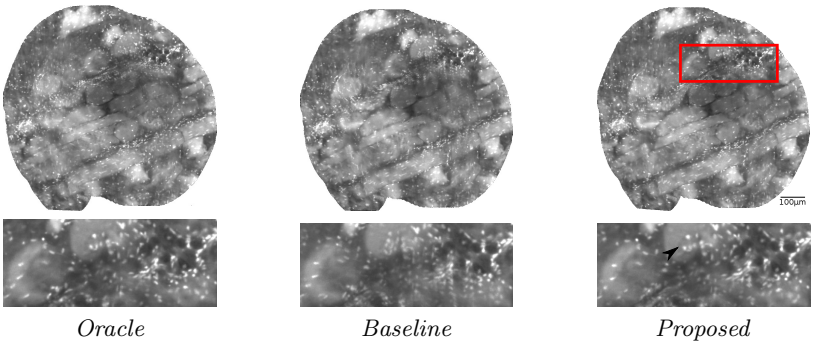

Fig. 4. Final result of mosaic reconstruction using the *oracle* reconstruction based on visual frame pairing, the baseline algorithm and our proposed method (top row). Magnification of the red frame region: the arrow indicates improvement in the registration with our proposed method.

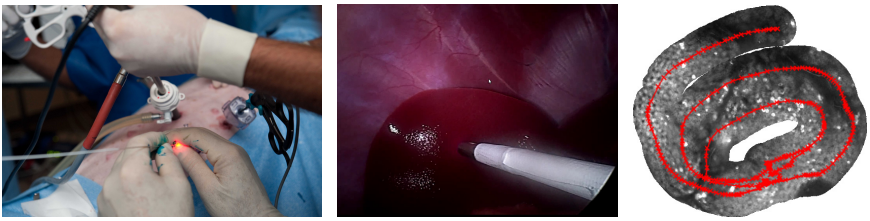

Fig. 5. In vivo clinical trial on pig liver: global setup when the probe is inserted (left), laparoscopic view of the robot-probe in contact with the liver (middle), mosaic reconstruction using our framework (right)

next to the confocal image (left), we present images of the mosaic reconstruction (top) and images of the correlation coefficients for each frame (bottom middle and right) for both methods.

The results of mosaic reconstruction obtained for the chicken breast and the pig liver using the baseline algorithm and our method were compared quantitatively to the *oracle* reconstruction: the sign test between populations Δ_i^B and Δ_i^V with hypothesis $\Delta_i^B > \Delta_i^V$ showed significant differences (p-value=5.5362e-13 for the chicken and p-value=2.1568e-08 for the pig liver) proving that the relative position of frames in the final mosaic reconstruction were closer to the *oracle* reconstruction using our method compared to the baseline. In Fig. 4, we also display both the mosaic reconstruction and a zoomed region for the chicken breast using the *oracle*, baseline and our method. For the in vivo pig liver dataset, the final mosaic reconstruction can be seen in Fig. 5 (right) using our method.

4 Discussion

The intuitive idea of using the equation of the trajectory directly to constrain the pairing and registration of frames could not be effectively used in our problem. In fact, the geometrical precision of the robot was not high enough to be relied on. Moreover, the angular speed of the robot holding the probe is instable due to tissue friction: the consequence of the instability on the position of the frames is amplified at each rotation making the pairing between frames from successive branches non trivially predictable. We thus only used as a priori basic geometrical properties of the spiral we programmed. These properties were 1) there was theoretical overlap between successive branches which meant that from a certain frame number corresponding to the beginning of the second branch, one could always find at least one frame from the previous branch overlapping it and 2) if frame F_i is overlapping frame F_j from next branch, then frame F_{i+1} has very high chances of overlapping a close neighbor of frame F_{j+1}. Formulating the problem as a hidden Markov model was a natural choice to implement these properties.

Correlation coefficient proved more robust than the natural sum of squared differences as similarity criterion due to tissue evolution during acquisition. The

Viterbi path providing spatial associations could have been sufficient to reconstruct the mosaic providing we could rely on the transformation estimated between every pair we found. However, in many occasions on real tissue inside the body, texture and features are very homogeneous and similar to one another, making the registration extremely challenging in some regions. Consequently, when the registration between spatial neighbors could not be performed with high confidence, we decided to discard it and to rely only on temporal transformations between frames in these regions. The filtered spatial pairs finally injected to the algorithm play the role of local anchors aiming at stabilizing the mosaic reconstruction. They ought to be regularly reparted along the trajectory: although we do not explicitly enforce the regularity of the pairs we inject, we keep a sufficiently high number of pairs that we can expect a frequency of anchor frames high enough to favorably help the reconstruction.

Conclusion. We presented a methodological framework to perform video mosaicing with a weak a priori on the trajectory. Our results showed statistically significant improvements compared to the baseline mosaicing method. Future work includes acquiring longer spiral videos in surgical conditions in vivo, using more flexible transformations to account for tissue deformations and making the mosaic reconstruction fast enough so that it may be used during surgery. We also plan on adapting the proposed framework to groupwise registration problems where images follow a pseudo-periodic pattern such as cardiac images.

References

1. Brown, M., Lowe, D.: Automatic panoramic image stitching using invariant features. Int. J. Comput. Vis. 74(1), 59–73 (2007)
2. Erden, M., Rosa, B., Szewczyk, J., Morel, G.: Mechanical design of a distal scanner for confocal microlaparoscope: a conic solution. In: Proc. ICRA 2013, pp. 1197–1204 (2013)
3. Gareau, D.S., Li, Y., Huang, B., Eastman, Z., Nehal, K.S., Rajadhyaksha, M.: Confocal mosaicing microscopy in mohs skin excisions: feasibility of rapid surgical pathology. J. Biomed. Opt. 13(5) (2008)
4. Loewke, K.E., Camarillo, D.B., Piyawattanametha, W., Mandella, M.J., Contag, C.H., Thrun, S., Salisbury, J.K.: In vivo micro-image mosaicing. IEEE Trans. Biomed. Eng. 58(1), 159–171 (2011)
5. Rosa, B., Erden, M., Vercauteren, T., Herman, B., Szewczyk, J., Morel, G.: Building large mosaics of confocal endomicroscopic images using visual servoing. IEEE Trans. Biomed. Eng. 60(4), 1041–1049 (2013)
6. Sawhney, H.S., Hsu, S., Kumar, R.: Robust video mosaicing through topology inference and local to global alignment. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 103–119. Springer, Heidelberg (1998)
7. Vercauteren, T., Perchant, A., Malandain, G., Pennec, X., Ayache, N.: Robust mosaicing with correction of motion distortions and tissue deformation for *in vivo* fibered microscopy. Med. Image Anal. 10(5), 673–692 (2006)
8. Wallace, M., Fockens, P.: Probe-based confocal laser endomicroscopy. Gastroenterology 136(5), 1509–1513 (2009)