Motion-Aware Mosaicing for Confocal Laser Endomicroscopy

Jessie Mahé[1⋆], Nicolas Linard[1⋆], Marzieh Kohandani Tafreshi[1,2], Tom Vercauteren[3], Nicholas Ayache[2], Francois Lacombe[1], and Remi Cuingnet[1]

[1] Mauna Kea Technologies, Paris, France
[2] Inria Asclepios Project-Team, Sophia Antipolis, France
[3] Translational Imaging Group, University College London, London, UK

Abstract. Probe-based Confocal Laser Endomicroscopy (pCLE) provides physicians with real-time access to histological information during standard endoscopy procedures, through high-resolution cellular imaging of internal tissues. Earlier work on mosaicing has enhanced the potential of this imaging modality by meeting the need to get a complete representation of the imaged region. However, with approaches, the dynamic information, which may be of clinical interest, is lost. In this study, we propose a new mosaic construction algorithm for pCLE sequences based on a min-cut optimization and gradient-domain composition. Its main advantage is that the motion of some structures within the tissue such as blood cells in capillaries, is taken into account. This allows physicians to get both a sharper static representation and a dynamic representation of the imaged tissue. Results on 16 sequences acquired *in vivo* on six different organs demonstrate the clinical relevance of our approach.

Introduction

Probe-based Confocal Laser Endomicroscopy (pCLE) [10] is a recent modality for *in situ* and *in vivo* imaging in the context of endoscopy. Basically, the objective of a confocal microscope is replaced by a flexible optical probe of length and diameter compatible with the working channel ($\varnothing \sim$ a few mm) of an endoscope in order to be able to perform *in situ* and *in vivo* imaging. Thus, pCLE provides physicians with real-time access to histological information during standard endoscopy procedures, through high-resolution cellular imaging of internal tissues, which is of particular interest for early detection of cancer (e.g. [2]).

Since pCLE is a contact real-time imaging modality, there is an inevitable hardware trade-off between invasiveness, frame rate, resolution and field-of-view. A typical pCLE acquisition system images, with a micrometrical resolution at 12 frames per second, a 600×600 μm optical section parallel to the tissue surface. Since most pCLE videos interpretations are based on the morphological characteristics of micro-cellular architecture in deep layers of the epithelium with structures' sizes ranging from one to a few hundred microns (e.g. ~ 200 μm for

⋆ Authors have contributed equally to the paper.

a colonic crypt in a healthy tissue), an increase of the field of view helps the clinicians in analyzing pCLE videos. Therefore, earlier work on mosaicing [12,16,19] has enhanced the potential of this imaging modality by meeting the need to get a complete representation of the imaged region. However, with such approaches, the dynamic information, which may be of clinical interest, is lost. For some organs such as the pancreas, dynamic information is required to draw a proper diagnosis. For instance, it helps distinguishing vessels from other structures by visualizing cells circulating in the blood stream.

In this study, we propose a new mosaicing construction algorithm to enlarge the field of view of pCLE sequences while preserving their motion information. It is based on a min-cut optimization and gradient domain composition. Our work has been inspired by the works of Agarwala et al., Ravi-Acha et al. and Joshi et al. on *panoramic video texture* (PVT) [1], *dynamosaicing* [15] and *cliplets* [7] respectively for manual video editing. The main advantage of our approach over previous works [12,16,19] is that the motion of some structures within the tissue, such as blood cells in capillaries, is taken into account. This allows physicians to get a dynamic representation of the imaged tissue whose field of view is no longer defined by the imaging system but by the size of the imaged region.

1 Dynamic Mosaicing

1.1 Problem Statement and Related Work

The goal of our method is to enlarge the field of view of a pCLE sequence while preserving its motion information. To that end, we begin by assuming that all the images have been spatially registered into a single coordinate system [19] and have an isotropic resolution. Thus, the registered pCLE sequence can be represented by a real value function I defined on $\mathcal{D} = \bigcup_{t\in[0,T]} \Omega_t \times \{t\}$ where T is the video duration and Ω_t is the imaged region at time t. We would like to create from I a new function $\tilde{I}$ defined on $\Omega \times [0,T]$ where $\Omega = \bigcup_{t\in[0,T]} \Omega_t$ is the whole imaged region. Since we visualize a 2D optical section at 12 Hz, motion quantification cannot be directly obtained from the pCLE sequence. Indeed, estimating out-of-plane motions is an ill-posed problem. Furthermore, the Nyquist-Shannon sampling theorem shows that the motion of blood cells cannot always be properly estimated. Motion information is mainly used to distinguish vessels from other structures by vizualising moving blood cells.

One way to deal with this problem would have been to consider it as an inpainting or video completion problem and search $\tilde{I}$ among functions extending I. This is the solution chosen by Matsushita et al. [13]. However, their method based on motion inpainting cannot be applied to our problem since motion estimation in a 2D section is an ill-posed problem. Rav-Acha et al. [15] tackled this problem differently by considering $\tilde{I} = I \circ \phi$ where ϕ is a deformation field. ϕ is chosen as a trade-of between a user defined deformation field ϕ_0 and a image regularization term chosen to reduce stitching artifacts. While their method is well adapted for manual video editing, it has several limitations for dynamic

mosaicing of pCLE sequences. First, chosing a suitable ϕ_0 is not obvious. Moreover, it does not have the ability to have discrete jumps in time, which is more adapted to repetitive stochastic textures, neither has it the ability to generate infinite dynamics. Agarwala et al. [1] proposed a very similar approach that does not have all these limitations: the *panoramic video textures* (PVT).

1.2 Markov Random Field Formulation

Agarwala et al. [1] considered $\tilde{I} = I \circ \phi$ with $\phi : (\mathbf{x}, t) \mapsto (\mathbf{x}, t + \Delta(t) \bmod T_{\max})$ where Δ is a time-offset function and $T_{\max}$ is the ouput video duration. In the following, we drop the modulo notation for clarity sake. In a nutshell, ϕ is obtained by solving a 3D Markov random field (MRF) problem where $\Omega \times [0, T_{\max}]$ is the domain and Δ are the free variables taking values in $\{0, 1, \cdots, T\}$. The unitary potential function U^{PVT}, defined as $U_i^{\text{PVT}}(\Delta) := 0$ if $(\mathbf{x}_i, t_i + \Delta(\mathbf{x}_i)) \in \mathcal{D}$ and $U_i^{\text{PVT}}(\Delta) := +\infty$ otherwise, ensures that all the pixels of the output video are defined. Agarwala et al. [1] defined the pairwise potential function as

$$V_{(i,j)}^{\text{PVT}}(\Delta_i, \Delta_j) := \sum_{k=i,j} \|I(\mathbf{x}_k, t_k + \Delta_i) - I(\mathbf{x}_k, t_k + \Delta_j)\|^n \tag{1}$$

for every adjacent points (i, j) in the spatio-temporal volume (6-connectivity) where I is defined and $V_{(i,j)}^{\text{PVT}}(\Delta_i, \Delta_j) = +\infty$ otherwise.

When applied to pCLE sequences, solving PVT energy function may yield either rather static videos or oscillation and jitter motions. In [1], the rationale behind the temporal pairwise potential is to consider a transition to be correct provided a similar one exists at the same position in the original (registered) video. As a matter of fact, with such regularization term, only temporal transitions that are similar to existing ones are considered, even if it means only keeping a small proportion of existing transitions. In [1], removing motion information is less critical than adding temporal artifacts. Since we rather avoid removing motion information, our temporal regularization penalizes label changes without regard to the pixel intensity. We therefore modify the energy pairwise potential to enforce temporal consistency with the original video.

To do so, we do not make any assumptions on the intensity evolution. Instead, we consider that a temporal intensity change of a given pixel is plausible if and only if it exists in the original video for the same pixel. Hence, in our approach, the pairwise potential function V is defined, for two adjacent points $i = (\mathbf{x}_i, t_i)$ and $j = (\mathbf{x}_j, t_j)$ in the spatio-temporal volume (6-connectivity), as

$$V_{(i,j)}(\Delta_i, \Delta_j) := \begin{cases} C & \text{if } \Delta_i \neq \Delta_j \text{ and } \mathbf{x}_i = \mathbf{x}_j \quad \text{(temporal)} \\ V_{(i,j)}^{\text{PVT}}(\Delta_i, \Delta_j) & \text{otherwise} \qquad\qquad\quad \text{(spatial)} \end{cases} \tag{2}$$

where C is a positive constant. The unitary potential function is $U = U^{\text{PVT}}$.

Since the pairwise potential V is a metric, we solve this MRF energy's optimization problem by α-expansion [4,8]. The time-offset function Δ is initialized to 0, which is equivalent to consider the original mosaic video as initialization.

1.3 Hierachical Optimization

Basically, the α-expansion algorithm [4] is a sequence of min-cut / max-flow optimization problems. To solve a min cut optimization problem, algorithms such as the the *relabel-to-front* algorithm are running in $O(|v|^3)$ where $|v|$ is the number of nodes of the optimization graph. When applied to our problem, each node represent one pixel in the reconstructed video. Thus, the number of nodes is $|v| = |\Omega \times [0, T_{\max}]|$.

For typical pCLE sequences, Ω ranges from 500×500 to 1000×1000 pixels and T from 30 to 200 frames. Therefore, to lessen the computational burden, hierachichal min-cut optimizations were performed (e.g. [1, 11]). The heuristic we followed is based on the assumption that the computed seams at the finer resolution are roughly similar to the ones at a coarser resolution. Hence, the problem is first solved at a coarse resolution. At finer resolution, the optimization is then only performed within the neighborhood of the seams.

1.4 Gradient-Domain Composition

In the reconstructed video, the computed seams may remain visible after optimization. This is mainly due to intensity variation resulting from the photobleaching of optical probe's doping fluorophore as well as an evolution of the fluorophore concentration in the tissue. Residual errors in the registration process may also yield similar visual artifacts. To reduce them, a gradient-domain composition [14, 21] with mixed Neuman and Dirichlet boundary conditions was performed to create the final video. For sake of efficiency, we solve the Poisson equation frame by frame instead of considering the whole 3D volume. To avoid bleaching artifacts, we used a still mosaic reconstruction to define the same Dirichlet boundary condition for all the pixels lying on still regions along the boundary of the entire imaged region.

1.5 Looping Mosaic

The time required to analyze a dynamic mosaic might be longer than its duration $T_{\max}$. Therefore, generating infinite dynamics is of interest. A first approach would be to post-process the constructed video $\tilde{I}$ with the *video texture* algorithm [18]. Note that the frame transition obtained in [18] cannot be directly used in the MRF formulation since they are not symmetric. As for the PVT method, Agarwala et al. [1] proposed to generate infinite dynamics by merely playing the video in an infinite loop mode. An optimal duration $T_{\max}$ of the video is then chosen to reduce temporal artifacts.

In our algorithm, the output video duration ($T_{\max}$) is defined by the user. We create a looping mosaic by slighlty changing the MRF graph. Every point $(\mathbf{x}, T_{\max})$ of the last frame is then connected to the point $(\mathbf{x}, 0)$ of the first frame. Thus, the infinite dynamics directly results from the temporal consistency constraints. Our approach is rather simple and yet gives good results.

2 Combining Static and Dynamic Mosaics

Acquired pCLE videos are rarely solely composed of dynamic regions. Hence, solving the MRF optimization problem described in section 1.2 on the whole input image unnecessarily increases the computational burden. Besides, residual registration errors on static region may result in flickering artifacts when considered as dynamic. Therefore, we combine still and dynamic mosaic [1, 7].

2.1 Texture Preserving Static Mosaicing

Reconstruction used in standard mosaicing algorithm [12,16,19] is an average of registered frames. This can bring superresolution or increase the signal-to-noise ratio provided there is no residual registration error. However, this in not always the case when observing *in vivo* tissues. Moving structures such as blood cells in capillaries appear blurred. Besides, texture information, which may be used for diagnosis in some organs (e.g. eosophagus), is lost when averaging. Several methods for mosaicing non rigid dynamical scenes have been proposed in the literature (e.g. [6, 15, 20]). In this study, we simply notice that using the MRF formulation (section 1.2) with $T_{\max} = 0$ results in a static mosaic obtained by stitching. This optimization is similar to texture synthesis [9] with the *image quilting*'s pairwise potential function [5]. This constructs a static mosaic that preserves texture information. Besides, it does not blur moving structures neither regions with residual registration error. Note that we do not use the pairwise potential described in [9], for it is a semi-metric and not a metric. Hence, α-expansion does not apply [4]. The α-β-swap algorithm [4] could be used instead but there is no guarantee to converge to a solution close to the global minimum.

2.2 Static Background Detection

Combining still and dynamic mosaicing requires a partition of imaged tissue into static and dynamic regions. In [1], this partition required user inputs. To do it automatically we focused on background substraction and motion detection algorithms. There is an extended literature on this subject [3]. To the best of our knowledge, these methods based on motion estimations or density estimations are not adapted to our problem for the reasons described in section 1.2. Therefore, our approach consists in thresholding the temporal variance computed on locally normalized frames (zero mean and unit variance).

2.3 Combining Still Images and Video Segments

To combine still images and video segments, Joshi et al. [7] used feathering and Laplacian blending. Agarwala at al. [1] adapt the unitary potential function on the boundary to stitch the video segments to a static mosaic. Using either of these approaches on thin structures such as vessels results in rather static videos. To solve this problem, we drop the stitching constraint in the MRF and we use a Dirichet boundary condition on the frontier between static and dynamic region for the gradient domain composition.

Table 1. Consistency of visual summaries with the original video rated by four experts with a five-level Likert scale (SD: strongly disagree; D: disagree; NAND: neither agree nor disagree; A: agree; SA: strongly agree).

Expert	SD	D	NAND	A	SA
1	00%	00%	06%	31%	63%
2	00%	31%	00%	44%	25%
3	00%	00%	00%	13%	87%
4	00%	00%	13%	06%	81%

Stitching Mosaicing

Expert	SD	D	NAND	A	SA
1	13%	19%	31%	31%	06%
2	13%	69%	00%	19%	00%
3	06%	00%	44%	50%	00%
4	00%	13%	00%	56%	31%

Averaging Mosaicing

Expert	SD	D	NAND	A	SA
1	06%	00%	00%	06%	88%
2	00%	00%	00%	25%	75%
3	00%	00%	00%	13%	87%
4	00%	06%	00%	13%	81%

Dynamic Mosaicing

3 Experiments and Results

3.1 Materials

The validation of our method was performed by four experts on a dataset composed of 16 sequences coming from six organs (oesophagus, stomach, pancreas, bladder, biliary duct and colon) with various conditions. Sequences have been acquired and preprocessed following [10,17]. To register all the input video frames we used the deformation fields obtained with the algorithm proposed by Vercauteren et al. [19]. In our experiments, we chose $T_{\max} = T$ and $n = 2$ (note that we also tried $n = 1$ and $n = 8$ and obtained similar results). As for the trade-off parameter C, we set it relatively to the dynamic of the image intensity. Parameter C was set proportionally to the median of $\max_{\Delta_i, \Delta_j} V_{(i,j)}(\Delta_i, \Delta_j)$ over all couples (i, j) of spatially adjacent points. The proportion parameter was set to 10, which we found to be a good trade-off between the spatial and temporal coherency constraints. Examples of reconstructed videos are available as supplementary materials (`https://sites.google.com/site/motionawaremosaicingpcle/`).

3.2 Consistency of the Visual Summary

To validate the clinical relevance of our mosaicing method, we asked four experts to assess the clinical consistency of the constructed mosaics with the original pCLE videos. Each expert rated, with a five-level Likert scale, a static mosaic reconstructed by weighting average as well as both a static and a dynamic mosaic constructed using our method. Results are presented in Table 1.

We also asked each expert to rank static mosaics obtained by averaging and by stitching. Our method was considered to give significantly better results (p-value < 0.05 with a Wilcoxon signed rank test). Stitched mosaics were preferred 97% of the time. We carried out the same experiment between the dynamic mosaics and the best static mosaics. Dynamic mosaicing was considering to give

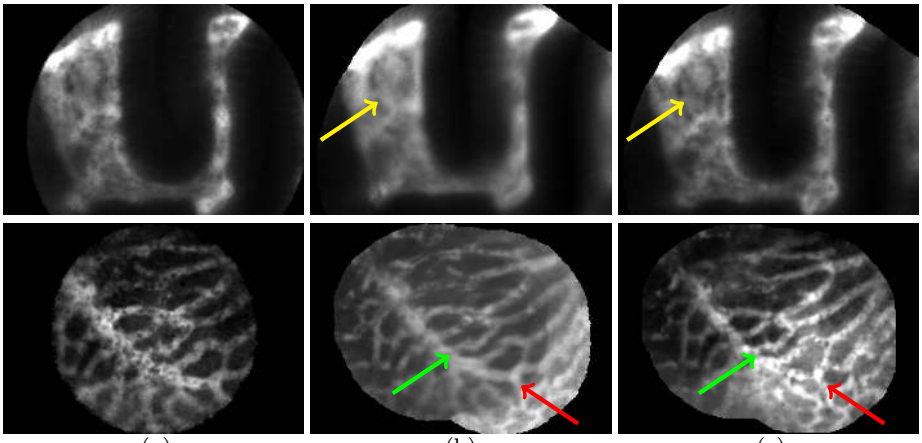

Fig. 1. (a) frame from a pCLE sequence; (b) static mosaic reconstructed by averaging; (c) static mosaic reconstructed with our method; it yields sharper images (yellow), preserves moving cells (green) in vessels and removes intensity change artifacts (red).

a significantly more consistent summary (p-value < 0.05). Dynamic mosaics were preferred 78% of the time over static mosaics.

We also show visually the benefits of our algorithm over the averaging reconstruction methods in Figure 1. Compared to the averaging approach, the stitching method better preserves texture information and does not blur moving structures such as blood cells. Besides, the gradient domain composition removes bleaching artifacts.

4 Discussion

By introducing time information in the mosaicing construction algorithm, we manage to have a dynamic mosaic reconstruction that preserves texture information and maintains motion appearance. Such mosaics help clinicians to distinguish specific structures such as capillaries by visualizing moving cells in the blood stream. We also derived a method to get a static and a dynamic mosaic. Both mosaics have sharper reconstructed images by reducing blur induced by residual registation error.

The clinical relevance of our methods was assessed by four expert users on 16 pCLE sequences aquired *in vivo* and *in situ* on six different organs. Results showed that our methods yield more consistent visual summaries of the original videos.

References

1. Agarwala, A., Zheng, K.C., Pal, C., Agrawala, M., Cohen, M., Curless, B., Salesin, D., Szeliski, R.: Panoramic video textures. In: TOG, SIGGRAPH. vol. 24, pp.

821–827. ACM (2005)
2. Becker, V., Vercauteren, T., von Weyhern, C.H., Prinz, C., Schmid, R.M., Meining, A.: High-resolution miniprobe-based confocal microscopy in combination with video mosaicing (with video). Gastrointest Endosc. 66(5), 1001–1007 (2007)
3. Benezeth, Y., Jodoin, P.M., Emile, B., Laurent, H., Rosenberger, C.: Comparative study of background subtraction algorithms. J Electron Imaging 19(3) (2010)
4. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. TPAMI 23(11), 1222–1239 (2001)
5. Efros, A.A., Freeman, W.T.: Image quilting for texture synthesis and transfer. In: Proc. SIGGRAPH. pp. 341–346. ACM (2001)
6. Fitzgibbon, A.W.: Stochastic rigidity: Image registration for nowhere-static scenes. In: Proc. ICCV. vol. 1, pp. 662–669. IEEE (2001)
7. Joshi, N., Mehta, S., Drucker, S., Stollnitz, E., Hoppe, H., Uyttendaele, M., Cohen, M.: Cliplets: juxtaposing still and dynamic imagery. In: Proc. 25th annual ACM symposium on User interface software and technology. pp. 251–260. ACM (2012)
8. Kolmogorov, V., Zabin, R.: What energy functions can be minimized via graph cuts? TPAMI 26(2), 147–159 (2004)
9. Kwatra, V., Schödl, A., Essa, I., Turk, G., Bobick, A.: Graphcut textures: image and video synthesis using graph cuts. In: ACM Transactions on Graphics (ToG). vol. 22, pp. 277–286. ACM (2003)
10. Le Goualher, G., Perchant, A., Genet, M., Cavé, C., Viellerobe, B., Berier, F., Abrat, B., Ayache, N.: Towards optical biopsies with an integrated fibered confocal fluorescence microscope. In: Proc. MICCAI, pp. 761–768. Springer (2004)
11. Lombaert, H., Sun, Y., Grady, L., Xu, C.: A multilevel banded graph cuts method for fast image segmentation. In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on. vol. 1, pp. 259–265. IEEE (2005)
12. Mahé, J., Vercauteren, T., Rosa, B., Dauguet, J.: A viterbi approach to topology inference for large scale endomicroscopy video mosaicing. In: Proc. MICCAI, pp. 404–411. Springer (2013)
13. Matsushita, Y., Ofek, E., Ge, W., Tang, X., Shum, H.Y.: Full-frame video stabilization with motion inpainting. TPAMI 28(7), 1150–1163 (2006)
14. Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. In: TOG. vol. 22, pp. 313–318. ACM (2003)
15. Rav-Acha, A., Pritch, Y., Lischinski, D., Peleg, S.: Dynamosaicing: Mosaicing of dynamic scenes. TPAMI 29(10), 1789–1801 (2007)
16. Rosa, B., Erden, M.S., Vercauteren, T., Herman, B., Szewczyk, J., Morel, G.: Building large mosaics of confocal edomicroscopic images using visual servoing. Biomedical Engineering, IEEE Transactions on 60(4), 1041–1049 (2013)
17. Savoire, N., André, B., Vercauteren, T.: Online blind calibration of non-uniform photodetectors: Application to endomicroscopy. In: Proc. MICCAI, pp. 639–646. Springer (2012)
18. Schödl, A., Szeliski, R., Salesin, D.H., Essa, I.: Video textures. In: Proc. SIGGRAPH. pp. 489–498. ACM (2000)
19. Vercauteren, T., Perchant, A., Malandain, G., Pennec, X., Ayache, N.: Robust mosaicing with correction of motion distortions and tissue deformations for in vivo fibered microscopy. Medical image analysis 10(5), 673–692 (2006)
20. Vidal, R., Ravichandran, A.: Optical flow estimation & segmentation of multiple moving dynamic textures. In: Proc. CVPR. vol. 2, pp. 516–521. IEEE (2005)
21. Wang, H., Raskar, R., Ahuja, N.: Seamless video editing. In: Pattern Recognition. Proc. ICPR. vol. 3, pp. 858–861. IEEE (2004)