Mosaicing of Confocal Microscopic In Vivo Soft Tissue Video Sequences

Tom Vercauteren, Aymeric Perchant, Xavier Pennec, Nicholas Ayache

Mosaicing of Confocal Microscopic *In Vivo* Soft Tissue Video Sequences

Tom Vercauteren[1,2], Aymeric Perchant[2], Xavier Pennec[1], and Nicholas Ayache[1]

[1] Projet Epidaure, INRIA Sophia-Antipolis, France
[2] Mauna Kea Technologies, 9 rue d'Enghien Paris, France

Abstract. Fibered confocal microscopy allows *in vivo* and *in situ* imaging with cellular resolution. The potentiality of this imaging modality is extended in this work by using video mosaicing techniques. Two novelties are introduced. A robust estimator based on statistics for Riemannian manifolds is developed to find a globally consistent mapping of the input frames to a common coordinate system. A mosaicing framework using an efficient scattered data fitting method is proposed in order to take into account the non-rigid deformations and the irregular sampling implied by *in vivo* fibered confocal microscopy. Results on 50 images of a live mouse colon demonstrate the effectiveness of the proposed method.

1 Introduction

Fibered confocal microscopy (FCM) is a potential tool for *in vivo* and *in situ* optical biopsy [1]. FCM is based on the principle of confocal microscopy which is the ability to reject light from out-of-focus planes and provide a clear in-focus image of a thin section within the sample. This optical sectioning property is what makes the confocal microscope ideal for imaging thick biological samples. The adaptation of a confocal microscope for *in vivo* and *in situ* imaging can be viewed as replacing a microscope objective by a probe of adequate length and diameter in order to be able to perform *in situ* imaging. For such purpose, a fiber bundle is used as the link between the scanning device and the microscope objective. After image processing of the FCM raw output, the available information is composed of a video sequence irregularly sampled in the space domain, each sampling point corresponding to a fiber center [1].

This imaging modality unveils the cellular structure of the observed tissue. The goal of this work is to enhance the possibilities offered by FCM by using image sequence mosaicing techniques in order to widen the field of view (FOV). Several possible applications are targeted. First of all, the rendering of wide-field micro-architectural information on a single image will help experts to interpret the acquired data. This representation will also make quantitative and statistical analysis possible on a wide field of view. Moreover, mosaicing for microscopic images is a mean of filling the gap between scales and allows multiscale information fusion for probe positioning and multi-modality fusion.

Each frame of the input sequence is modeled as a deformed partial view of a ground truth 2D scene. The displacement of the fiber bundle probe across the tissue is described by a rigid motion. Due to the interaction of the contact probe with the soft tissue, a small non-rigid deformation appears on each input frame. Because of those non-linear deformations, and the irregular sampling of the input frames, classical video mosaicing techniques need to be adapted.

In Section 2, our mosaicing framework is described. The first main contribution presented in Section 3 is the use of Riemannian statistics to get a robust estimate of a set of mean rigid transformations from pairwise registrations results. The second main contribution is proposed in Section 4 where we develop a mosaicing framework using an efficient scattered data fitting method well-suited for non-linear deformations and irregularly sampled inputs. Finally real experiments described in Section 5 demonstrate the effectiveness of our approach and show the significant image improvements obtained by using non-linear deformations.

2   Problem Formulation and Mosaicing Method

The goal of many existing mosaicing algorithms is to estimate the reference-to-frame mappings and use these estimates to construct the mosaic [2]. Small residual misregistrations are then of little importance because the mosaic is reconstructed by segmenting the field into disjoint regions that use a single source image for the reconstruction [3,4]. Since our input frames are rather noisy, we would like to use all the available information to recover an approximation of the true underlying scene. We will therefore estimate the frame-to-reference transformations (instead of the usual reference-to-frame) and consider all the input sampling points as shifted sampling points of the mosaic. This has several advantages for our problem. First of all, this is really adapted to irregularly sampled input frames because we will always use the original sampling points and never interpolate the input data. This approach is also more consistent with a model of noise appearing on the observed frames rather than on the underlying truth. Finally in this framework, it will be possible to get a mosaic at a higher resolution than the input frames. The drawback is that we need an accurate estimate of the unknown transformations.

Let I be the unknown underlying truth and I_n be the observed frames. Our algorithm makes use of the following observation model,

$$I_n(p) = I(f_n(p)) + \epsilon_n(p), \quad \forall p \in \Omega_n, \tag{1}$$

where $\epsilon_n(p)$ is a noise term, Ω_n is the coordinate system associated with the nth input frame and $f_n : \Omega_n \to \Omega$ is the unknown frame-to-reference mapping composed of a large rigid mapping r_n and a small non-rigid deformation b_n,

$$f_n(p) = b_n \circ r_n(p). \tag{2}$$

By making the reasonable assumption that consecutive frames are overlapping, an initial estimate of the global rigid transformations can be obtained by
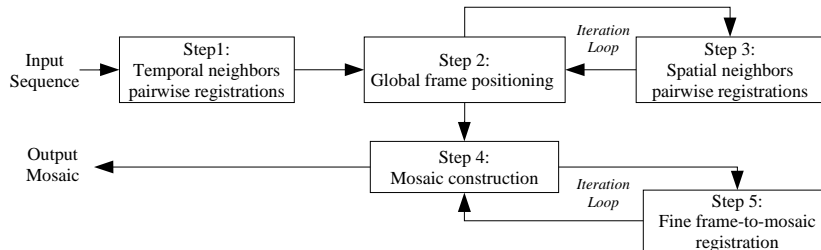
Fig. 1. Block diagram of the mosaicing algorithm

using an image registration technique to estimate the motion between the consecutive frames. Global alignment is then obtained by composing the local motions. This initial estimate suffers from a well-known accumulation of error problem. Our algorithm depicted in Fig. 1 iteratively refines the global positioning by adding new pairwise rigid registration results to estimate the global parameters. Once a consistent set of rigid transformations is found, the algorithm constructs an initial mosaic by mapping all observed sampling points into the common reference coordinate system and using an efficient scattered data fitting technique on this point cloud. The residual non-rigid deformations are finally taken into account by iteratively registering an input frame to the mosaic and updating the mosaic based on the new estimate of the frame-to-mosaic mapping.

3   Global Frame Positioning from Pairwise Registrations

The first task of our algorithm is to register pairs of overlapping input images under a rigid transformation assumption. For that purpose, we use a classical registration framework based on a similarity criterion optimization but any other technique (e.g. block matching framework [5], Mellin transform [3], feature-based registration [6] etc.) can be used. Let $r_{j,i}^{(\mathrm{obs})} : \Omega_i \to \Omega_j$ be the pairwise rigid registration result between input frames i and j. This result is considered as a noisy observation of $r_j^{-1} \circ r_i$. Based on the set of all available observations, our algorithm looks for a globally consistent estimate of the global parameters $[r_1, \ldots, r_N]$. This problem is addressed in [3] where a least-square solution is given when linear transformations are considered. This technique cannot readily be adapted to rigid transformation. In [7], the authors propose a more general approach. Some chosen corner points are transformed through r_i and $r_j \circ r_{j,i}^{(\mathrm{obs})}$. The squared distance between the transformed points added to a regularization term is then minimized. These techniques are sensitive to outliers, and are either tailored to a specific type of transformation or need a somewhat ad hoc choice of points. In this paper statistics for Riemannian manifolds are used to provide a globally consistent and robust estimate of the global rigid transformation.

The computational cost of registering all input frames pairs is prohibitive and not all pairs of input frames are overlapping. It is therefore necessary to choose which pairs could provide informative registration results. For that purpose, we chose the topology refinement approach proposed in [7]. An initial guess of the

global parameters $[r_1, \ldots, r_N]$ is obtained by registering the consecutive frames, the algorithm then iteratively chooses a next pair of input frames to register (thus providing a new observation $r_{j,i}^{(\text{obs})}$) and updates the global parameters estimation. As we only consider the pairwise registration results as noisy observations, we need many of them. In order to minimize the computational cost of those numerous registrations, we use a multiresolution registration technique using a Gaussian image pyramid that stops at a coarse level of the optimization.

Given the set of all available pairwise registration results Θ, we need to estimate the true transformations. A sound choice is to consider a least-square approach. However the space of rigid transformations is not a vector space but rather a Lie group that can be considered as a Riemannian manifold. Classical notions using distances are therefore not trivial to generalize. In what follows, we provide an extension of the Mahalanobis distance for Riemannian manifolds [8] and propose an optimization algorithm to find the least-square estimate. Let r be an element of the 2D rigid transformations Riemannian manifold. From the theory of Riemannian manifolds, the vector $\mathcal{V}(r)$ representing r by its angle $\alpha(r) \in [-\pi, \pi]$ and translation components $t_x(r)$, $t_y(r)$ can be considered as a mapping of the transformation r onto the vector space defined by the tangent space at the identity point Id of the manifold. This tangent space will be denoted as Id-tangent space. By using the canonical left-invariant Riemannian metric for rigid transformations, the distance of r to the identity is defined as the norm of the representation $\mathcal{V}(r)$ in the Id-tangent space, $\text{dist}(r, \text{Id}) = ||\mathcal{V}(r)||$. The distance between two transformations is given by

$$\text{dist}(r_a, r_b) = \text{dist}(r_b^{-1} \circ r_a, \text{Id}) = ||\mathcal{V}(r_b^{-1} \circ r_a)||. \tag{3}$$

Using this distance, it is possible to define a generalized mean for a random transformation r, the Fréchet mean $\text{E}_f[r] = \arg\min_{r_f} \text{E}[\text{dist}(r_f, r)^2]$. If e is a random error whose Fréchet mean is the identity, its covariance matrix is simply defined as $\Sigma_{ee} = \text{E}[\mathcal{V}(e)\mathcal{V}(e)^T]$. The squared Mahalanobis distance between e and the identity is given by $\mu^2(e, \text{Id}) = \mathcal{V}(e)^T \Sigma_{ee}^{-1} \mathcal{V}(e)$. We now have the tools to derive the global parameters estimator. The observation model is given by $r_{j,i}^{(\text{obs})} = r_j^{-1} \circ r_i \circ e_{j,i}^{(\text{obs})}$, where $e_{j,i}^{(\text{obs})}$ is a random error whose Fréchet mean is assumed to be the identity and whose covariance matrix is Σ_{ee}. An estimate of $[r_1, \ldots, r_N]$ is given by the set of transformations that minimizes the total Mahalanobis distance:

$$[\hat{r}_1, \ldots, \hat{r}_N] = \arg \min_{[r_1, \ldots, r_N]} \sum_{(i,j) \in \Theta} \mu^2(e_{j,i}^{(\text{obs})}, \text{Id}) \tag{4}$$

This equation does not admit a closed form solution but an efficient optimization method can be designed by a simple modification of a usual non-linear least square optimizer such as the Gauss-Newton descent. The Riemannian structure of the transformation space is taken into account by adapting the intrinsic geodesic gradient descent in [8]. The idea is to walk towards an optimum by a series of steps taken along a geodesic of the manifold rather than walking in the tangent vector space. Let $r^{(t)}$ be an estimate of a rigid transformation r at step

t, the intrinsic geodesic walking is achieved by finding a direction $\delta r^{(t)}$ and a step length λ for the following update equation in the Id-tangent space:

$$\mathcal{V}(r^{(t+1)}) = \mathcal{V}(r^{(t)} \circ \lambda \delta r^{(t)}). \tag{5}$$

By contrast, if a usual optimization routine on the Id-tangent space is used, we get a walking direction $\Delta \mathcal{V}^{(t)}$ such that the update would be

$$\mathcal{V}(r^{(t+1)}) = \mathcal{V}(r^{(t)}) + \lambda \Delta \mathcal{V}^{(t)}. \tag{6}$$

Using (6) directly can be problematic because we are not assured to remain on the manifold. It is however possible to combine the power of intrinsic geodesic walking and the ease of use of classical optimization routine by mapping a walking direction found in the Id-tangent space onto the manifold. For that purpose, a first order Taylor expansion of (5) around the identity is used:

$$\mathcal{V}(r^{(t)} \circ \lambda \delta r^{(t)}) = \mathcal{V}(r^{(t)}) + \lambda JL(r^{(t)}) \cdot \mathcal{V}(\delta r^{(t)}) + O(\lambda^2), \tag{7}$$

where $JL(r) = \partial \mathcal{V}(r \circ s)/\partial \mathcal{V}(s)|_{s=\mathrm{Id}}$. By identifying (6) and (7), we see that $\Delta \mathcal{V}^{(t)} = JL(r^{(t)}) \cdot \mathcal{V}(\delta r^{(t)})$. The walking direction in the manifold is thus

$$\mathcal{V}(\delta r^{(t)}) = JL(r^{(t)})^{-1} \cdot \Delta \mathcal{V}^{(t)}. \tag{8}$$

Within this general framework, several improvements can easily be added. The terms of the cost function can be weighted by some confidence measure, robust estimation techniques such as M-estimators can be used to discard outliers, the noise variance can be re-estimated based on the measurements etc. We are now able to get robust and globally consistent estimates of the rigid transformations which we use as initial estimates of the complete global transformations $[f_1, \ldots, f_N]$. The resulting mosaics in Fig. 2(a) show an accurate global positioning of the input frames that is robust to erroneous pairwise registrations.

4   Frame to Mosaic Fine Registration

Once a globally consistent mosaic has been constructed, it is possible to make a fine multi-image registration by iteratively registering each input frame to the mosaic and updating the mosaic. The mosaicing problem can be written as an optimization problem over the unknown underlying image I and the unknown transformations $[f_1, \ldots, f_N]$ of the following multi-image criterion, $\boldsymbol{S}(f_1, \ldots, f_N, I) = \sum_{n=1}^{N} S(I_n, I \circ f_n)$, where $S(I_a, I_b)$ is a usual pairwise similarity criterion between the two images I_a and I_b. With this framework our mosaic refinement procedure can be seen as an alternate optimization scheme.

We divide the fine frame-to-mosaic registrations into two loops of increasing model complexity. First we refine the global rigid mappings. Then, in order to account for the small non-rigid deformations, the residual deformation fields $[b_1, \ldots, b_N]$ are modeled by using B-splines tensor products on a predefined grid.

This framework can easily be extended to use any other non-rigid registration methods using landmarks-based schemes or more accurate deformation models.

This iterative mosaic refinement scheme requires a new mosaic construction at each iteration. It is therefore necessary to use a very efficient reconstruction algorithm. Furthermore, since we want to register input frames with the mosaic, the reconstruction needs to be smooth enough for the registration not to be trapped in a local minimum but detailed enough for the registration to be accurate. Once an estimate $\hat{f}_n$ of f_n is available, we get a point cloud composed of all transformed sampling points from all the input frames

$$\{(p_k, i_k)\} = \{(\hat{f}_n(p), I_n(p)) | p \in \Omega_n^s, n \in [0, \dots, N]\}, \tag{9}$$

where Ω_n^s is the set of sampling points in the input frame n. The usual algorithms for scattered data approximation do not simultaneously meet the requirements of efficiency, and control over the smoothness of the approximation. In the sequel we develop our second main contribution which is an efficient scattered data fitting algorithm that allow a control over the smoothness of the reconstruction. The main idea is to get an approximation of the underlying function by using a method close to Shepard interpolations. The value associated with a point in Ω is a weighted average of the nearby sampled values,

$$\hat{I}(p) = \sum_k w_k(p) i_k = \sum_k \frac{h_k(p)}{\sum_l h_l(p)} i_k. \tag{10}$$

The usual choice is to take weights that are the inverse of the distance, $h_k(p) = \text{dist}(p, p_k)^{-1}$. In such a case we get a true interpolation [9]. An approximation is obtained if a bounded weighting function $h_k(p)$ is chosen. We choose a Gaussian weight $h_k(p) = G(p - p_k) \propto \exp(-||p - p_k||^2 / 2\sigma_a^2)$ and thus (10) can be rewritten as
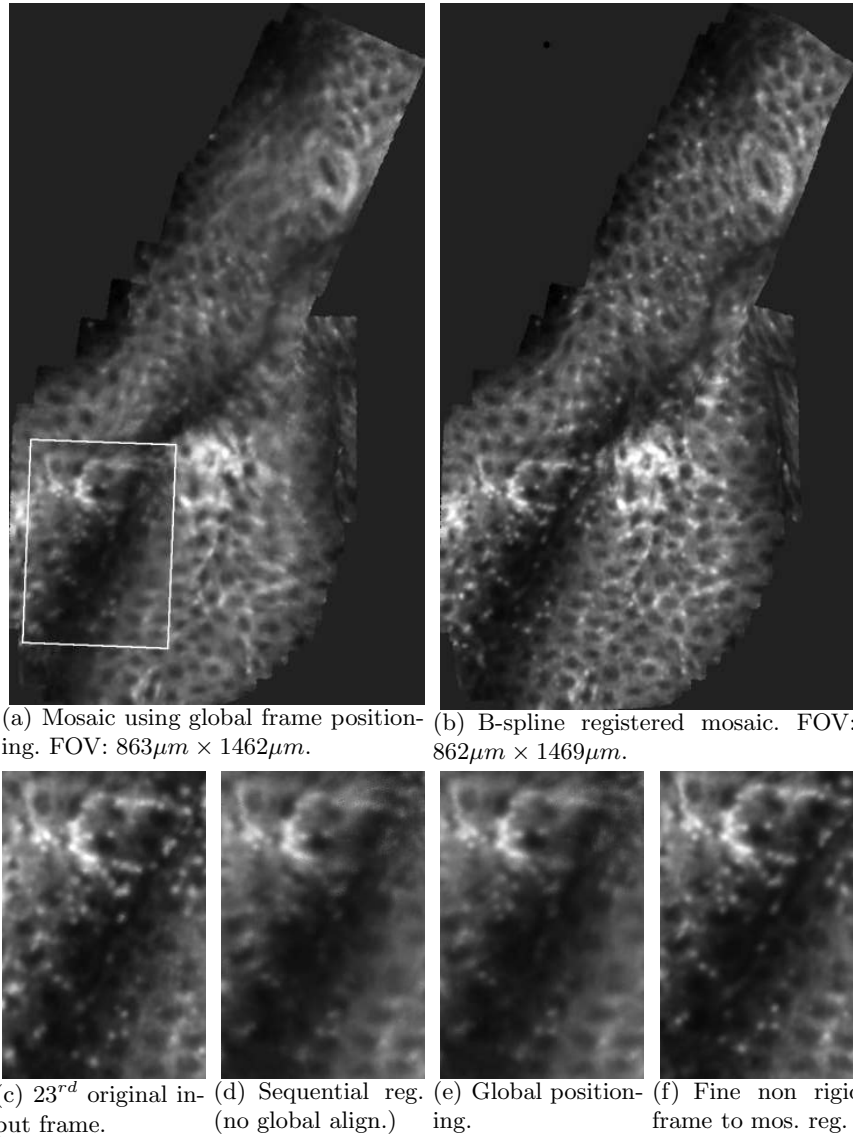
$$\hat{I}(p) = \frac{\sum_k i_k G(p - p_k)}{\sum_k G(p - p_k)} = \frac{[G \star \sum_k i_k \delta_{p_k}](p)}{[G \star \sum_k \delta_{p_k}](p)}, \tag{11}$$

where δ_{p_k} is a Dirac distribution centered at p_k. This scattered data approximation technique requires only two Gaussian filtering and one division and is thus very efficient. The smoothness is controlled by the variance σ_a^2 of the Gaussian kernel. Thanks to this approximation method, we obtain a mosaic that is sharper and smoother than the original input frames as demonstrated in Section 5.

5   Results

In the field of colon cancer research, the development of methods to enable reliable and early detection of tumors is a major goal. In the colon, changes in crypt morphology are known to be early indicators of cancer development. The crypts that undergo these morphological changes are referred to as Aberrant Crypt Foci (ACF) and they can develop into cancerous lesions. Compared to the standard methods of ACF screening, fluorescence FCM enables the operator to see the lesions in real-time and to make an almost immediate evaluation.

However, in many cases the limited field of view restricts the confidence that the operator has in the ACF counting. By offering an extended field of view, mosaicing techniques can be an answer to this restriction.

(a) Mosaic using global frame positioning. FOV: $863\mu m \times 1462\mu m$.

(b) B-spline registered mosaic. FOV: $862\mu m \times 1469\mu m$.

(c) 23rd original input frame.

(d) Sequential reg. (no global align.)

(e) Global positioning.

(f) Fine non rigid frame to mos. reg.

Fig. 2. Mosaic of 50 live mouse colon images (Fluorescence FCM). The zone corresponding to the 23rd input frame is detailed for different steps of the algorithm. FOV one input: $303\mu m \times 425\mu m$. Images are courtesy of Danijela Vignjevic, Sylvie Robine, Daniel Louvard, Institut Curie, Paris, France

The effectiveness of the proposed algorithm is shown on a sequence that has been acquired *in-vivo* on a mouse colon stained by acriflavine at 0,001%. The mouse was treated with azoxymethane (AOM) to induce a colon cancer. As shown in fig.2(b), our algorithm allows for a simultaneous visualization of normal crypts and ACFs.

The global frame positioning mosaicing took approximately 1 min on a 2GHz P4 and 12 min if the non-rigid deformations are compensated. The imaged tissue is really soft and non-linear deformations occur. Figure 2(b) illustrates the gain we obtain by taking into account those non-rigid deformations. Some details are lost if we only use rigid deformations and appear again on our final mosaic. The results shown here prove the feasibility of mosaicing for *in-vivo* soft-tissue microscopy. Current work is dedicated to the validation of the proposed approach.

6   Conclusion

The problem of video mosaicing for *in-vivo* soft tissue confocal microscopy has been explored in this paper. A fully automatic robust approach based on Riemannian statistics and efficient scattered data fitting techniques was proposed. The results shown for two different imaging modalities are promising and encourage the application of the proposed method for qualitative and quantitative studies on the mosaics.

References

1. G. Goualher, A. Perchant, M. Genet, C. Cave, B. Viellerobe, F. Berier, B. Abrat, and N. Ayache, "Towards optical biopsies with an integrated fibered confocal fluorescence microscope," in *Proc. of MICCAI'04*, 2004, pp. 761–768.
2. M. Irani, P. Anandan, and S. Hsu, "Mosaic based representations of video sequences and their applications," in *Proc. ICCV'95,*, June 1995, pp. 605–611.
3. J. Davis, "Mosaics of scenes with moving objects," in *Proc. CVPR'98*, 1998, pp. 354–360.
4. S. Peleg, B. Rousso, A. Rav-Acha, and A. Zomet, "Mosaicing on adaptive manifolds," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 10, pp. 1144–1154, Oct. 2000.
5. S. Ourselin, A. Roche, S. Prima, and N. Ayache, "Block matching: A general framework to improve robustness of rigid registration of medical images," in *Proc. MICCAI'00*, 2000, pp. 557–566.
6. A. Can, C. V. Stewart, B. Roysam, and H. L. Tanenbaum, "A feature-based technique for joint linear estimation of high-order image-to-mosaic transformations: Mosaicing the curved human retina," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, no. 3, pp. 412–419, Mar. 2004.
7. H. S. Sawhney, S. Hsu, and R. Kumar, "Robust video mosaicing through topology inference and local to global alignment," in *Proc. ECCV'98*, vol. 2, 1998, pp. 103–119.
8. X. Pennec, "Probabilities and statistics on Riemannian manifolds: Basic tools for geometric measurements," in *Proc. NSIP'99*, vol. 1, 1999, pp. 194–198.
9. I. Amidror, "Scattered data interpolation methods for electronic imaging systems: a survey," *J. Electron. Imaging*, vol. 11, no. 2, pp. 157–176, Apr. 2002.