# Robust SIFT-Based Hierarchical Video Mosaicing for Endomicroscopy

Danica Chang, Nicolas Linard, Mahé Jessie, Tom Vercauteren, Julien Dauguet

# Robust SIFT-based Hierarchical Video Mosaicing for Endomicroscopy

Danica Chang, Nicolas Linard, Jessie Mahé, Tom Vercauteren and Julien Dauguet

***Abstract* — We present a method to perform video mosaicing for endomicroscopy with two major improvements compared to the state of the art. First, instead of using individual images directly, we start by creating sub-mosaics from short video sub-sequences using iconic registration. The sub-mosaics are then considered for feature-based registration. Second, groupwise estimation is performed between all sub-mosaics based on SIFT matching to infer globally consistent spatial transformations. Both improvements increase robustness of the reconstruction compared to a baseline mosaicing method.**

## I. Objectives

The goal of this work is to perform video mosaic reconstruction for long endomicroscopy sequences. We address several challenges with our datasets in this work: 1) the topology (knowledge of neighboring frames) is unknown between non successive frames, 2) some frames convey very little contrast and information making the registration very difficult, 3) the overlap between spatial neighbors can be very small or void.

## II. Material and Method

We acquired videos of about *n=1000* frames on a printed grid on paper, *ex vivo* chicken breast sample and *in vivo* pig liver during a pre-clinical trial [1]. The acquisition probe is held by a robot [2] that follows a noisy spiral trajectory (Fig. 1a).

First we create *sub-mosaics* which are small mosaics made of groups of *p=20* successive frames each and with each sub-mosaic overlapping by 5 frames.

We calculate SIFT descriptors [3] on a dense grid using a bin size of 20 pixels with *vlfeat*. We pair each SIFT descriptor from one sub-mosaic with its closest match (Euclidean distance in feature space) in each of the other sub-mosaics using *FLANN*.

We then look for a translation to match 2 sub-mosaics. Each SIFT matched pair provides one noisy measurement of the best translation. More precisely, if we consider a pair of images $F_i$ and $F_j$, we have:

$$T_i - T_j \approx T^k_{ij} \qquad (1)$$

where $T_i$ and $T_j$ are the best translations for frames $F_i$ and image $F_j$ respectively with respect to a common reference frame for the final mosaic image and $T^k_{ij}$ is one noisy measurement $k$ of the best translation registering frame $F_i$ to frame $F_i$. We then use a vote map to discard outlier translations. Since most pairs of frames do not actually overlap, we first run the SIFT matching and SIFT filtering on a coarse grid, we determine which pairs of frames overlap and we run the SIFT matching again on a fine grid.

We next find the best translation $T_i$ of each sub-mosaic with respect to a common reference frame by minimizing the global squared error compared to the observed translations $T^k_{ij}$ as expressed in equation (1). The final mosaic image is obtained using *Enblend* to merge the images together.

We compare our results to a baseline mosaicing method with topology inference using translations [4]. For the grid, we acquired a reference image acquired using a benchtop confocal microscope.

## III. Results and Discussion

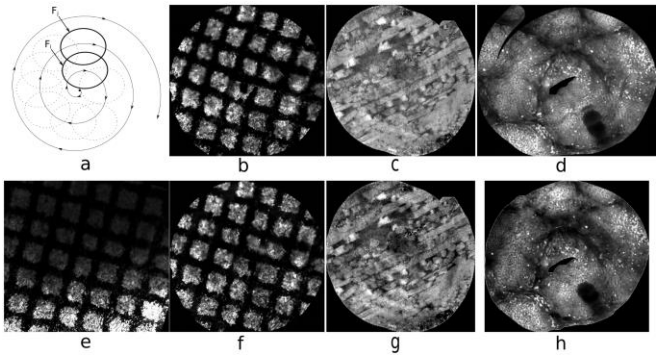


Figure 1. Spiral trajectory (a). Baseline reconstructions (b-d). Reference grid image (e). Proposed reconstructions (f-h).

For the grid, the proposed reconstruction (Fig.1f) is very similar to the reference image (Fig.1e) and shows straighter and more perpendicular lines compared to the baseline (Fig. 1b). We compute a per-block normalized correlation of the baseline and proposed reconstruction compared to the reference image. A *sign-test* on these two series shows that the proposed reconstruction is significantly better (*p-value* ≈ *2e-21*) than the baseline. For the chicken and the liver (Fig.1c-d), the reconstruction looks more compact and more contrasted compared to baseline (Fig.1g-h).

We proposed a method where we consider groups of frames instead of individual frames for registration in order to accumulate more landmarks and information. Also, we optimized all the transformations between groups of frames at once. These two improvements make the mosaicing reconstruction visually and statistically more robust compared to a state of the art method on our datasets.

## References

[1] J. Mahé et al. MICCAI '13 (2013).
[2] M. Erden et al. Proc. ICRA'13 (2013).
[3] D. Lowe, 2004. Int J Comput Vision, 60(2):91–110.
[4] T. Vercauteren et al., Med. Image Anal. 10(5), 673–692 (2006).

D. Chang is with the MIT Boston USA and was with Mauna Kea Technologies for a summer internship.

N. Linard, T. Vercauteren, J. Mahé and J. Dauguet are with Mauna Kea Technologies, 9 rue d'Enghien, 75010 Paris, France. (corresponding e-mail: nil@maunakeatech.com).