

Registration of Microscopic Iris Image Sequences Using Probabilistic Mesh

Xubo B. Song¹, Andriy Myronenko¹, Stephen R. Plank²,
and James T. Rosenbaum²

¹ Department of Computer Science and Electrical Engineering
OGI School of Science and Engineering, Oregon Health and Science University, USA

² Department of Ophthalmology, Department of Cell and Developmental Biology,
and Department of Medicine

Casey Eye Institute, Oregon Health and Science University, USA
{xubosong, myron}@csee.ogi.edu, {rosenbaj, plancks}@ohsu.edu

Abstract. This paper explores the use of deformable mesh for registration of microscopic iris image sequences. The registration, as an effort for stabilizing and rectifying images corrupted by motion artifacts, is a crucial step toward leukocyte tracking and motion characterization for the study of immune systems. The image sequences are characterized by locally nonlinear deformations, where an accurate analytical expression can not be derived through modeling of image formation. We generalize the existing deformable mesh and formulate it in a probabilistic framework, which allows us to conveniently introduce local image similarity measures, to model image dynamics and to maintain a well-defined mesh structure and smooth deformation through appropriate regularization. Experimental results demonstrate the effectiveness and accuracy of the algorithm.

1 Introduction

Recent development of videomicroscopy technology for imaging the immune response is revolutionizing the way to study and understand the immune mechanism [1,2]. The motion patterns of leukocytes, specifically T cells, are directly related to the cellular and chemical environment in lymph node, thymus and a site of eye inflammation and can reveal underlying disease mechanisms [3,4,5]. Our specific interest has focused on ocular inflammatory disease, a leading cause of blindness. The eye is especially attractive for imaging studies because cellular migration can be recorded without introducing any surgical trauma. Microscopy videos can reveal patterns of T cell, neutrophil and antigen-presenting cell migration in the ocular uveal tract, indicating a complexity in immune responses that has not been closely examined before [1,3,4]. By studying these microscopy videos, we can characterize the migration of leukocytes within the iris stroma in disease models.

However, the characterization of leukocyte motility is made difficult by motion artifacts in videomicroscopy. The videos were taken of sedated murine eyes

affected by uveitis (inflammation of the uveal tract), with a static microscopic camera looking at a portion of the iris. The motion artifacts are caused by wandering of the eye, dilation and contraction of the pupil, head motion, and sometimes refocusing of the camera during imaging. The net result is jitter and distortion (both spatial and intensity) in the image plane, which subsequently obscures the leukocyte motion. Frame-by-frame image registration is needed to stabilize and rectify the image sequences, which will pave the way for subsequent cell tracking. The deformation across frames is locally non-linear, and it is not feasible to obtain an accurate closed-form deformation model by examining the image formation process. In addition, the motion artifacts can cause the region on the iris being imaged to go in and out of depth of field of the camera, resulting in local blurring and local intensity instability. In this paper, we focus on frame-by-frame registration of the image sequence, using a probabilistic mesh model to account for the nonlinear nonrigid nature of the image deformation and to accommodate the local intensity variations.

2 Method

Mesh-based deformable model has been successfully used for motion estimation, compensation, video compression and other applications [6,7,8,9]. A mesh consists of a set of control nodes, which define polygon elements (patches) in the image. The mesh nodes move freely. The displacement of an interior point in an image element can be interpolated from the corresponding nodal displacements. The motion field over the entire frame is described by the displacements of the nodes only. Very complex motion field can be reproduced by a mesh model, given that sufficient number of nodes are used. As long as the nodes form a feasible mesh, mesh-based representation is guaranteed to be continuous and thus free from the blocking artifacts. Another key advantage of mesh model is that it enables continuous tracking of the same set of nodes over consecutive frames, which is important for registration of image sequences. Triangular and quadrangular are the most common mesh elements. In this paper, we focus on quadrangular elements. Our approach is closely related to that in [6]. We generalize the original mesh by formulating mesh deformation in a Bayesian framework, which allows us to naturally introduce priors to model video dynamics, to constrain the mesh structure, and to account for the local intensity variations.

Consider two images $I_1(\mathbf{x}) = I(\mathbf{x}, t_1)$ and $I_2(\mathbf{x}) = I(\mathbf{x}, t_2)$, which are the reference image and the target image respectively. Let $\mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}$ represent the coordinates of the K control nodes placed in image $I_1(\mathbf{x})$, with known positions. These nodes move to locations $\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K}$ in image $I_2(\mathbf{x})$. The motion estimation problem is to find the location of all nodes $\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K}$, so that all image elements in the reference frame matches well with the corresponding deformed elements in the target frame. Given reference image I_1 , target image I_2 , and the location of control nodes $\mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}$ in image I_1 , the posterior probability of nodes locations $\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K}$ in image I_2 can be written as

$$\begin{aligned}
 & P(\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K} | I_1, I_2; \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}) \\
 & \propto P(I_2 | \mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K}; \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}; I_1) P(\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K} | \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}; I_1) \\
 & = P(I_2 | \mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K}; \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}; I_1) P(\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K} | \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}) \quad (1)
 \end{aligned}$$

The first term is the likelihood of observing I_2 given I_1 and the nodes location in both images. The second term is the prior of nodes location in I_2 given where they are in I_1 . In the following section, we will introduce approaches for modeling the likelihood function that reflect the local intensity properties and for modeling the priors that incorporate image dynamics and enforces well-defined mesh structures.

2.1 The Likelihood Term

Let D be the number of nodes in each image element. For quadrangular elements, $D = 4$. Let m be the element index. Denote B^m as the m th element the images, and $\mathbf{x}_i^{C_1^m}, \dots, \mathbf{x}_i^{C_D^m}$ as the D nodes that are responsible for defining element B^m in image $I_i, i = 1, 2$. Under the assumption that all elements in an image are conditionally independent given their corresponding defining nodes, and that an element in image I_2 only depends on its own nodes and the the same element in I_1 , the likelihood term in (1) becomes

$$\begin{aligned}
 & P(I_2 | \mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K}; \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}; I_1) \\
 & = \prod_m P(I_2(B^m) | \mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K}; \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}; I_1) \\
 & = \prod_m P(I_2(B^m) | \mathbf{x}_2^{C_1^m}, \dots, \mathbf{x}_2^{C_D^m}; \mathbf{x}_1^{C_1^m}, \dots, \mathbf{x}_1^{C_D^m}; I_1(B^m)). \quad (2)
 \end{aligned}$$

The likelihood term breaks down to measure the element-by-element image similarity between the reference and the target images, given the locations of the element nodes in both images.

With noisy images, it is desirable to have the element similarity measures dependent on the “distinctiveness” of the elements. For instance, two similar elements that have “distinct features” such as edges, corners, line crossings and rich textures should be given more confidence than two similar homogenous elements. This can be captured by modeling the element-by-element similarity with a Gaussian distribution given by

$$P(I_2(B^m) | \mathbf{x}_2^{C_1^m}, \dots, \mathbf{x}_2^{C_D^m}; \mathbf{x}_1^{C_1^m}, \dots, \mathbf{x}_1^{C_D^m}; I_1(B^m)) \sim N(I_2(B^m) - I_1(B^m), \Sigma_{B^m})$$

where the choice of Σ_{B^m} reflects the distinctiveness of patch B^m . For instance, we can use isotropic diagonal matrix $\Sigma_{B^m} = \sigma_m^2 \mathbf{I}$, where σ_m^2 is the local intensity variance in element m , and \mathbf{I} is the identity matrix. This is equivalent to assigning weighting to different element pairs in an error function according to the distinctiveness of the elements. Element-by-element similarity measure with such choice of Σ_{B^m} also has the property of being invariant with respect to local intensity scaling.

2.2 The Prior Term for Image Dynamics

One of the key advantages of mesh model is its ability for continuous tracking of the same set of nodes over consecutive frames. The image sequence often has its dynamics. Taking advantage of the dynamics can lead to improved tracking robustness and reduced search space for optimization. The image dynamics can be captured by properly defining the prior term $P(\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K} | \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K})$ in (1). The prior term depends only on nodes location and not on intensity. In case where the nodes locations across frames are first-order Markovian and go through a random walk, the prior can be modeled as a Gaussian distribution with $P(\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K} | \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}) \sim N(\mathbf{x}_2^{C_1} - \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_2^{C_K} - \mathbf{x}_1^{C_K}; \Sigma_C)$, where $\Sigma_C = \sigma_d^2 \mathbf{I}$ and σ_d^2 reflects the random walk step size.

2.3 The Prior Term for Maintaining a Well-Defined Mesh

When adapting the mesh nodes locations, it is necessary to ensure that the mesh structure is well-defined. In other words, the mesh doesn't change topology and there are no nodes flips overs or obtuse elements. Typically this is done by limiting the search range of the nodes location when they are updated during an iterative procedure [6]. Here we adapt a less ad hoc and more principled approach by introducing, as a prior term in Bayesian formulation, a node order preserving term that explicitly enforces the ordering of the nodes. Since we always start with well-defined mesh $\mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}$ in I_1 , this order preserving term will only be applied to constrain the nodes locations $\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K}$ in image I_2 . We can define the prior term as: $P(\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K} | \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}) = P(\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K}) \propto \exp\{-\frac{\beta}{2} \sum_{j,n} \|\mathbf{x}_2^{C_j} - \mathbf{x}_2^{C_n}\|^2\}$, where $\mathbf{x}_2^{C_j}$ and $\mathbf{x}_2^{C_n}$ are neighboring nodes. This term is similar to the one introduced in [10] for Elastic Nets. Since the spatial difference of neighboring nodes is an approximation of the first-order derivative of the deformation field, this order preserving term also serves as a Tikhonov regularization term that enforces the smoothness of the deformation field.

2.4 The Complete Posterior

Putting together the likelihood term and the two prior terms into (1), we have the posterior probability

$$\begin{aligned}
 & P(\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K} | I_1, I_2; \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}) \\
 & \propto \prod_m P(I_2(B^m) | \mathbf{x}_2^{C_1^m}, \dots, \mathbf{x}_2^{C_D^m}; \mathbf{x}_1^{C_1^m}, \dots, \mathbf{x}_1^{C_D^m}; I_1(B^m)). \\
 & \quad P(\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_K} | \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_K}) \\
 & \propto \prod_m \exp\{-\frac{1}{2}(I_2(B^m) - I_1(B^m))^T \Sigma_{B^m}^{-1} (I_2(B^m) - I_1(B^m))\}. \\
 & \exp\{-\frac{\alpha}{2}(\mathbf{x}_2^C - \mathbf{x}_1^C)^T \Sigma_C^{-1} (\mathbf{x}_2^C - \mathbf{x}_1^C)\} \exp\{-\frac{\beta}{2} \sum_{j,n} \|\mathbf{x}_2^{C_j} - \mathbf{x}_2^{C_n}\|^2\} \quad (3)
 \end{aligned}$$

where \mathbf{x}_1^C and \mathbf{x}_2^C are vectors formed by concatenating the control nodes coordinates $\mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_\kappa}$ and $\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_\kappa}$ in images I_1 and I_2 respectively, and α and β are the hyper-parameters controlling the strength of the two prior terms. The practically useful values for the hyper-parameters can be obtained manually for a given type of images.

An energy function can be defined as the negative log likelihood, given by

$$\begin{aligned}
 E(\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_\kappa}) &= -\log P(\mathbf{x}_2^{C_1}, \dots, \mathbf{x}_2^{C_\kappa} | I_1, I_2; \mathbf{x}_1^{C_1}, \dots, \mathbf{x}_1^{C_\kappa}) \\
 &= \sum_m (I_2(B^m) - I_1(B^m))^T \Sigma_{B^m}^{-1} (I_2(B^m) - I_1(B^m)) + \\
 &\quad \frac{\alpha}{2} (\mathbf{x}_2^C - \mathbf{x}_1^C)^T \Sigma_C^{-1} (\mathbf{x}_2^C - \mathbf{x}_1^C) + \frac{\beta}{2} \sum_{j,n} \|\mathbf{x}_2^{C_j} - \mathbf{x}_2^{C_n}\|^2, \tag{4}
 \end{aligned}$$

which can be minimized by gradient-based optimization.

3 Implementation

The image sequences were acquired of the iris and ciliary/limbal region of anesthetized animals with endotoxin-induced uveitis, observed by intravitral epifluorescence videomicroscopy with a modified DM-LFS microscope (Leica) and a CF 84/NIR Black-and-White camera from Kappa, Gleichen, Germany [3]. Time-lapse videos were recorded for 30 to 90 minutes at 3 frames per minute. The images are monotone of size 720x480.

3.1 Preprocessing

The images are first normalized to reduce the effect of global intensity variation across frames, followed by an edge-preserving smoothing process to reduce noise while preserving structural features in the images (e.g., vessel branches). A global affine registration is used to initiate the mesh deformation algorithm. The affine registration, even if not extremely accurate, is a good initial guess of the final registration. It also is crucial for speeding up the mesh-bases registration and for avoiding poor local minimum of the energy function.

3.2 Hierarchical Mesh-Based Registration

We adopt a hierarchical procedure, which successively approximate the control nodes locations. We start with an image down-sampled at the lowest resolution level L . At this image resolution, we start with a regular mesh such that each element covers a 16x16 image patch. The mesh nodes locations are updated according to the gradient descent of the energy function in (4), until the energy function reaches a preset threshold. Then these nodes locations are translated to the image with resolution level $L - 1$, and additional nodes are inserted so that in this image the elements maintain roughly the size of 16x16. The location of these newly inserted nodes are determined by linearly interpolating the existing nodes locations. The nodes locations in this newly formed mesh are once again updated

according to gradient descent of the energy function. This process is repeated until the images reach the highest resolution level and the energy function is reduced to a predetermined level.

Such a hierarchical process is important because it speeds up the optimization process and is crucial for avoiding getting stuck at a local minimum of the energy function by providing, at each iteration, a more reasonable starting point for optimization. By the construction of such hierarchy, we ensure an adequately complex, instead of an overtly complex, deformation field. This is consistent with the Occam's Razor principle, which prefers the simplest model among all models that are consistent with data.

4 Results

We tested the algorithm on Pentium4 3.5GHz machines with 4GB Ram. The code was implemented in Matlab with some subroutines written in C. It takes approximately 30 seconds to register 2 images frames.

We illustrate the effectiveness of the proposed algorithm on two microscopic iris video sequences, comparing the root of mean squared pixel-by-pixel intensity differences (RMSE) between two frames, for before and after registration. The RMSE are computed on images after intensity normalization. The first sequence has 51 frames. The average RMSE for this sequence is 0.1365 ± 0.0383 before registration, which reduces to 0.0074 ± 0.0011 after registration. The second

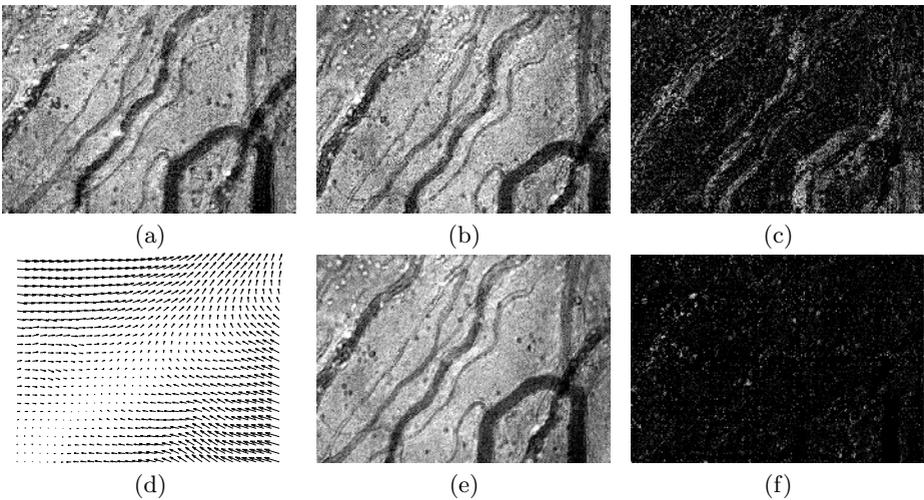


Fig. 1. Two image frames from the first video sequence: (a) frame 1; (b) frame 30; (c) the absolute intensity difference between the two frames before registration; (d) the estimated deformation field found by the algorithm; (e) the registration result when image (b) is aligned with image (a); (f) the absolute intensity difference between the two frames after image (b) is aligned with image (a)

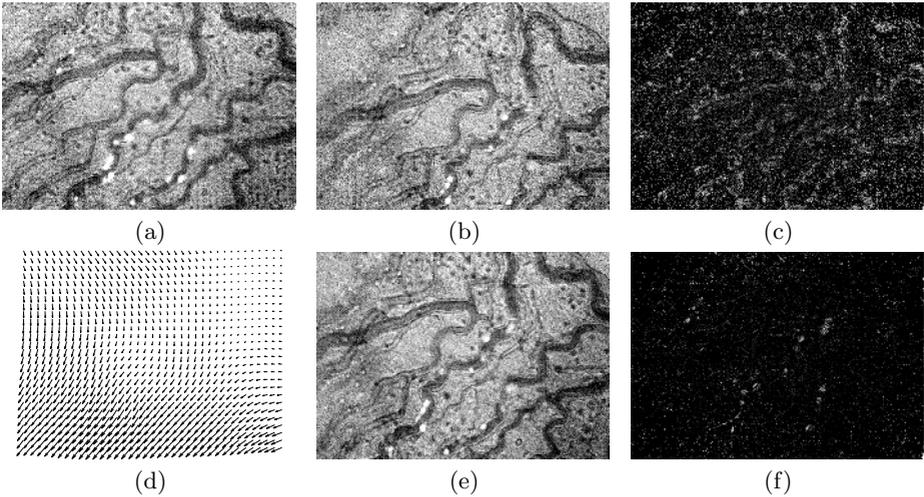


Fig. 2. Two image frames from the second video sequence: (a) frame 12; (b) frame 25; (c) the absolute intensity difference between the two frames before registration; (d) the estimated deformation field found by the algorithm; (e) the registration result when image (b) is aligned with image (a); (f) the absolute intensity difference between the two frames after image (b) is aligned with image (a)

sequence has 25 frames. The average RMSE for this sequence is 0.0912 ± 0.0193 before registration, which reduces to 0.0165 ± 0.0053 after registration. Visualization of the actual video sequences demonstrates significantly improved image stability that were originally severely jittery and deformed.

Figures 1 and 2 illustrate the registration result on a pair of image frames from these two sequence. In both figures, panels (a) and (b) are the two frames from the sequences to be registered. Panel (c) is the absolute frame difference before registration. Panel (d) is the deformation field between the two frames, and (e) shows the second image registers onto the first. Panel (f) is the absolute frame difference after registration. In both examples, we can see significant improvement in the difference images. The complex nonlinear deformations are effectively captured by the mesh-based model.

5 Conclusions and Discussions

Registration of microscopic iris images is an important step toward leukocyte tracking and characterization, but it is difficult due to the high local nonlinearity of the deformation. We generalize the mesh-based image registration method by formulating it in a probabilistic framework. Such formulation allows us the flexibility to measure local similarity between images and to encode priors. We define similarity measures that reflect the local image intensity properties, introduce priors that can capture the dynamics of image deformation in video sequence as well as the prior that enforces proper mesh structure and smooth

deformation. The algorithm was implemented using gradient descent optimization in a hierarchical fashion. The algorithm proves to be effective and accurate for registration of microscopic image sequences, which provide sufficient information for subsequent cell tracking.

Acknowledgement

This work is partially supported by NIH grants EY013093 and EY06484 and by Research to Prevent Blindness awards to SRP, JTR, and the Casey Eye Institute.

References

1. Rosenbaum, J.T. and Planck, S., Martin, T., Crane, I., Xu, H., Forrester, J.: Imaging ocular immune responses by intravital microscopy. *Int Rev Immunol* **21** (2002) 255–272
2. Halin, C., Rodrigo Mora, J., Sumen, C., von Andrian, U.: In vivo imaging of lymphocyte trafficking. *Annu Rev Cell Dev Biol.* **21** (2005) 581–603
3. Becker, M., Nobile, R., Planck, S., Rosenbaum, J.: Digital video-imaging of leukocyte migration in the iris: intravital microscopy in a physiological model during the onset of endotoxin-induced uveitis. *J Immunol Meth* **240** (2000) 23–27
4. Becker, M., Adamus, G., Martin, T., Crespo, S., Planck, S., Offner, H., Rosenbaum, J.: Serial imaging of the immune response in vivo in a t cell mediated autoimmune disease model. *The FASEB Journal* **14** (2000) A1118
5. Kawakami, N., Nagerl, U., Odoardi, F., Bonhoeffer, T., Wekerle, H., Flugel, A.: Live imaging of effector cell trafficking and autoantigen recognition within the unfolding autoimmune encephalomyelitis lesion. *J Exp Med* **201** (2005) 1805–1814
6. Wang, Y., Lee, O.: Active mesh: A feature seeking and tracking image sequence representation scheme. *IEEE Trans. Image Processing* **3** (1994) 610–624
7. Nosratinia, A.: New kernels for fast mesh-based motion estimation. *IEEE Trans. Circuits Syst. Video Technol.* **11** (2001) 40–51
8. Nakaya, Y., Harashima, H.: Motion compensation based on spatial transformations. **4** (1994) 339–356, 366–7
9. Toklu, C., Tekalp, A., Erdem, A., Sezan, M.: 2d mesh based tracking of deformable objects with occlusion. In: *International Conference on Image Processing.* (1996) 17A2
10. Durbin, R., Szeliski, R., Yuille, A.: An analysis of the elastic net approach to the traveling salesman problem. *Neural Computation* **1** (1989) 348–358
11. Bajcsy, R., Kovacic, S.: Multiresolution elastic matching. **46** (1989) 1–21
12. Ferrant, M., Warfield, S., Guttmann, C., Mulken, R., Jolesz, F., R., K.: 3d image matching using finite element based elastic deformation model. In: *MICCAI.* (1999) 202–209
13. Brown, L.: A survey of image registration techniques. **24** (1992) 325–376