Image Stabilization and Registration for Tracking Cells in the Microvasculature

Adam P. Goobic, Jinshan Tang, *Senior Member, IEEE*, and Scott T. Acton*, *Senior Member, IEEE*

***Abstract*—We propose a registration system to be used for tracking cells in intravital video microscopy that 1) stabilizes *jitter*—the undesired translational displacement of frames due to respiratory movement, *etc*., and 2) registers frames in a moving field of view (FOV) to allow for cell tracking over an extended range. For the first time, tracking of rolling leukocytes *in vivo* over a moving FOV is demonstrated. In a fixed FOV, stable background regions are located using a morphological approach. Template subregions are then selected from the stable regions and matched to corresponding locations in a reference frame. We show the effectiveness of the stabilization algorithm by using an active contour to track 15 leukocytes previously untrackable due to jitter. For 30 fixed FOV sequences containing rolling leukocytes, the resulting root-mean-square error (RMSE) is less than 0.5 μm. To align frames in a moving FOV, we present a modified correlation approach to estimate the common region between two consecutive fixed FOVs. We correlate the overlapping regions of the initial frame of the current fixed FOV and the final frame of the previous fixed FOV to register the images in the adjoining moving FOV. The RMSE of our moving FOV registration technique was less than 0.6 μm. In 10 sequences from different venules, we were able to track 11 cells using an active contour approach over moving FOVs.**

***Index Terms*—Active contour, leukocyte, registration, stabilization.**

I. Introduction

THE velocity of rolling leukocytes (white blood cells) is the preeminent indicator predicting the intensity of inflammatory cell recruitment, the innate immune response to cellular injury [1]–[3]. Inflammation aides the human body by locating, restricting, and eliminating undesired irritants and damaged tissue [4]. However, unwanted inflammation can initiate diseases such as arthritis, heart disease, and multiple sclerosis [5]. To ultimately understand the inflammatory process researchers must investigate how leukocytes stop rolling and become adherent to inflamed venules [2]. Thus, medical research groups studying inflammatory disease desire the ability to accurately and efficiently track rolling leukocytes, which is necessary to analyze the effect of anti-inflammatory treatments in living animals [6]–[8].

Automated tracking of rolling leukocytes for intravital microscopy is necessary 1) to increase the speed of inflammation research by significantly reducing the number of hours required to analyze data and 2) to increase the accuracy of displacement measurements and velocity calculations by removing investigator bias. Currently, *in vivo* tracking must be performed by hand, a lengthy and tedious task. Researchers manually record cell positions using a frame-by-frame analysis method, which requires tens of hours of examination [1], [3], [9]. The analysis of hundreds of rolling leukocytes is required to develop a meaningful velocity distribution, the prevailing manner to describe leukocyte rolling velocities [3].

A major impediment to automated tracking is background movement or jitter—the rapid shifting of frames caused by abrupt movement from the subject due to the circulatory and respiratory systems [10]. In murine experiments, on average, displaced frames, called δ-frames (for convenience, we list the definitions of some symbols used in this paper in Table I), occur in near periodic cycles of twice per second, in groups of about five frames. δ-frames 1) can cause the tracker to lose the target cell, 2) lead to erroneous information such as displacement measurements and velocity calculations, and 3) introduce motion blur. Background movement can cause the tracker to jump from the target cell to a nearby cell [10]. Spurious tracking data result since the background is not stable [10]. Thus, background registration of the intravital video microscopy is requisite for 1) the realization of a robust tracking system and 2) to ensure accurate cell motion statistics.

Recently, automated trackers were developed to track rolling leukocytes in a fixed field of view (FOV) [6]. However, tracking over a moving FOV is requisite to investigate a single leukocyte as it travels through the microvasculature [2]. Specifically, we desire the ability to track a leukocyte from the point it enters the post-capillary venule and follow it through the stage of rolling until firm adhesion to the endothelium [11]. Displacement measurements and velocity calculations over this path give insight to the progressive leukocyte activation phenomenon from exposure to chemoattractants found on the surface of the endothelium [2]. Extended range tracking will provide information involving long endothelial contact times, which are an apparent requirement for a leukocyte to transition from rolling to firm adhesion through a gradual deceleration process requiring increased β_2 integrin adhesiveness [2]. Currently, researchers use a manual frame-by-frame tracking method to understand the molecular requirements of leukocyte arrest under these physiological conditions. To determine leukocyte displacements, each

Manuscript received May 20, 2003; revised July 18, 2004. This work was supported in part by the National Institutes of Health (NIH) under Grant HL68510. *Asterisk indicates corresponding author.*

A. P. Goobic is with the Air Force Research Laboratory Sensors Directorate at Hanscom AFB, Bedford, MA 01731 USA.

J. Tang is with the Virginia Image and Video Analysis (VIVA), Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904 USA.

*S. T. Acton is with the Virginia Image and Video Analysis (VIVA), Department of Electrical and Computer Engineering, University of Virginia, P.O. Box 400743, Charlottesville, VA 22904 USA (acton@virginia.edu).

Digital Object Identifier 10.1109/TBME.2004.840468

TABLE I
DEFINITIONS OF SYMBOLS USED

F1	Frame to initial frame correlation.
FA	Frame to average frame correlation.
FX	Frame to frame X correlation.
δ-frame	Frame displaced from majority of frames in sequence defined by having a correlation coefficient less than the expected input correlation value.
S-frame	Fixed field of view frame that does not experience jitter; defined by having a correlation coefficient greater than the expected input correlation value.
F-frame	Fixed field of view frame.
F_f	Final frame in $\mathbf{F}_g$.
$\mathbf{F}_g$	Fixed field of view frame set where g is the index for the number of fixed fields of view in the sequence.
F_i	Initial frame in $\mathbf{F}_{g+1}$.
Fixed FOVBR	Background Registration system for fixed field of view sequences
Moving FOVBR	Background Registration system for moving of view sequences
M-frame	Moving field of view frame.
$\mathbf{M_g}$	Moving field of frame set where g is the index for the number of moving fields of view sequence.
J	A set of contiguous δ-frames
I	Original input video sequence

frame must be manually registered. Registration over a moving FOV presents the additional challenges of 1) registration over large displacements and 2) tracking in the presence of motion blur.

In this paper we present two Virginia Image and Video Analysis (VIVA) background registration solutions that compensate for jitter: fixed FOV background registration (fixed FOVBR) and moving FOV background registration (moving FOVBR). The fixed FOVBR uses a morphological approach to determine stable background regions to register the δ-frames through correlation method. The moving FOVBR defines frames as either fixed FOV (F-frames) or moving FOV (M-frames) and then utilizes the fixed FOVBR to minimize the jitter from the δ-frames in the fixed FOVs in the sequence. M-frames are registered to a mosaic built from the initial and final F-frames of successive fixed FOVs. We then integrate an active contour (snake) tracker to delineate the moving leukocyte.

Section II provides the necessary background, discussing conventional registration and stabilization techniques. Section III presents the fixed FOVBR in a step-by-step explanation of the stabilization process. The section also gives an outline of the gradient vector flow (GVF) active contour tracker. Section IV describes the operation of the moving FOVBR. Section V details the fixed FOVBR and moving FOVBRs performance and tracking results. Section VI provides a conclusion.

II. BACKGROUND

Image registration is the spatial and intensity mapping between two images [12] and is defined by the process of geometrically transforming points in two images so corresponding features in the two images have the same coordinates after transformation [13]. The spatial transformation or mapping employed to correctly overlay two images is the quantitative objective of the image registration process [12].

A. Conventional Stabilization

An array of conventional stabilization techniques employ mechanical [14], [15] and electronic methods to compensate for jitter. Mechanical means to reduce sensor platform vibrations include accelerometers, gyros, and mechanical dampers [14]. The mechanical stabilization techniques minimize large motion jitter. Electronic stabilization systems range in methods but share a common task—the minimization of small motion jitter. Several algorithms use local motion vectors from subimages to estimate global frame motion between successive frames or a reference frame [15]–[24]. For images only experiencing translation (with no interior movement), a corrective vector compensates for global motion and properly aligns the image [19], [23]. Correcting images that experience affine transformations and three dimensional rotation [25] require complex solutions. Several algorithms have been proposed for these applications. These algorithms include the mosaic-based registration technique implemented in pyramidal hardware by Hansen *et al.* in [14], a two-dimensional feature-based multiresolution motion estimation system to compensate for global motion proposed by Yao *et al.* in [25], a Kalman filter smoothing approach presented in [26], an integral projection matching algorithm aimed to provide visually "jerk-free" video for consumer viewing proposed by Ratakonda in [19], a real-time tracking of multiple moving objects in a moving FOV using a least squares method in [16] and so on.

Intravital microscopy presents unique challenges for registration not common to imaging large-region scenes—those recorded using hand-held cameras, aerial photography, or satellite imagery. First, the source of jitter is the movement of a living subject, not instability of the camera. Second, intravital video microscopy images are micro-region scenes where the FOV is approximately 10 000 $\mu\mathrm{m}^2$ (120 $\mu\mathrm{m}\times$ 90 $\mu\mathrm{m}$ in our experiments). These images include innate registration complexities: nondistinct landmark features, rapid changes in camera focus, and multiple motion regions. Intravital images capture ever-present motion found in the main feature, a post-capillary venule, with a diameter between 20–40 $\mu\mathrm{m}$ [10]. Moreover, capillaries terminate at the venule and emit the targets of interests, leukocytes. Thus, roughly a third of the image is constantly changing, most of which is moving too fast for the camera to capture at a frame rate of 30 fps. Therefore,

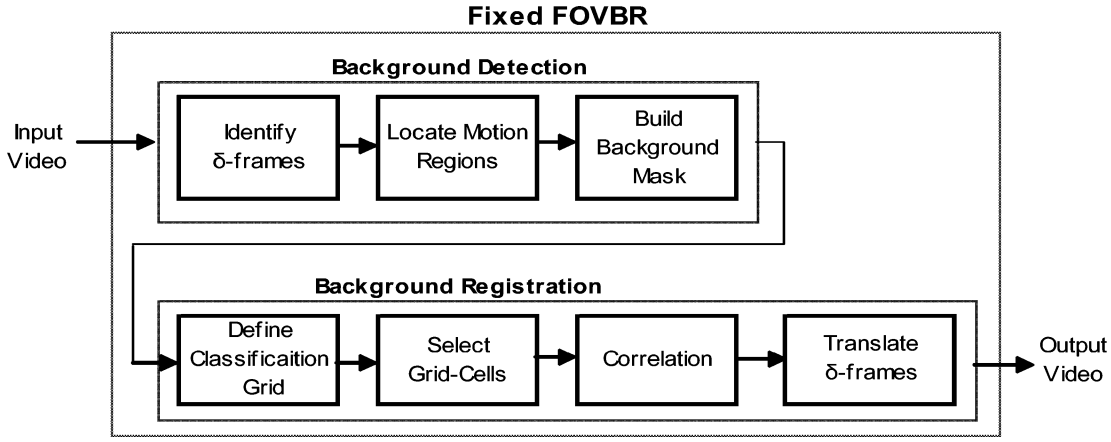

Fig. 1. Flow chart of the fixed FOVBR operation.

only static regions within the image can be considered to accurately estimate global motion. Furthermore, local regions across an intravital microscopy image are highly correlated, thus, projective matching cannot be used. And because the optics are adjusted to follow leukocytes located in the foreground, background features, ideal for matching, lack contrast, yet another challenge to intravital registration.

Although registration is not the focus of their paper, Sato *et al.* mention an intensity-based registration technique for microscopic images of a rat mesentery microvessel [27] in a fixed FOV. Translation is estimated by combining the quadratic interpolations of the cross correlation values from the current frame to a key frame. In contrast, we first eliminate the motion regions, which will give better results for our applications, and then use a group of subimage "patches" to determine the translation. Here, we also address registration and tracking in a moving FOV.

To track leukocytes over a moving FOV, researchers currently use an off-line digital image processing system to manually register each frame of the video sequence [28], [29]. In this paper, we introduce an automated method for registration and tracking in a moving FOV, which allows prolonged observation of cell motility.

III. Fixed FOV Background Registration

In this section, we describe the VIVA background registration system for fixed FOVBR sequences. The fixed FOVBR can be broken down into two parts: 1) detection of the background region and 2) background registration through local region correlation. Fig. 1 provides a flow chart of the fixed FOVBR process which is discussed in detail in the next subsections. Fig. 2 shows a graphical user interface that provides access to the software for minimizing jitter by performing background registration for intravital video microscopy in a fixed FOV.

From qualitative observation of more than eight hours of video we observe that sets of contiguous δ-frames, $\mathbf{J}$, repeat every r frames. Typically $\mathbf{J}$ possesses four to six δ-frames and $13 \leq r \leq 20$ frames when the camera frame rate is 30 fps. Because the circulatory and respiratory systems are unique to each living subject and change with time, we choose not to use the periodicity information in our registration system for the purpose of anticipating jitter. Throughout the duration of an experiment, the number of δ-frames per $\mathbf{J}$ and the repetition rate can change due to the semi-regular and semi-periodic respiratory system. Also note that for our intravital experiments, the frame-to-frame transformation is translational. By manually registering 30 sequences, each containing 100 frames, we determined the rigid transformation of our intravital images is limited to translation. The maximum, minimum and average rotations were 0.487°, 0.000°, and 0.033°, respectively. The maximum, minimum and average scale factors were 1.004, 0.994, and 1.000, respectively. In terms of maximum pixel movement, this means that error due to both rotation and scaling is less than one micron.

A. Background Detection

Background detection is a multistep process to identify regions of the video that do not contain moving objects. The process consists of three steps (see Fig. 1): 1) Identify the δ-frames in the video sequence; 2) locate motion regions, and 3) build a background mask.

Step 1) Identify the δ-Frames: First we define each frame in the input video sequence as an S-frame or δ-frame. This is accomplished using the normalized correlation among the frames in the video sequence. For doing this, we need to define the normalized correlation between the reference frame (which will be selected before computing the normalized correlation) and the frame in the video sequence, and the average normalized correlation of a video sequence.

The normalized correlation between the reference frame I_r and the tth frame I_t with displacement in the video sequence is defined as

$$C_{I_r,I_t}(x,y) = \frac{\sum_m \sum_n I_r(m,n) I_t(m-x,n-y)}{\sqrt{\left(\sum_m \sum_n {I_r}^2(m,n)\right)\left(\sum_m \sum_n {I_t}^2(m-x,n-y)\right)}} \quad (1)$$

where (x,y) is the translation and $I_r(i,j)$ is the pixel value of image I_r at location (i,j).

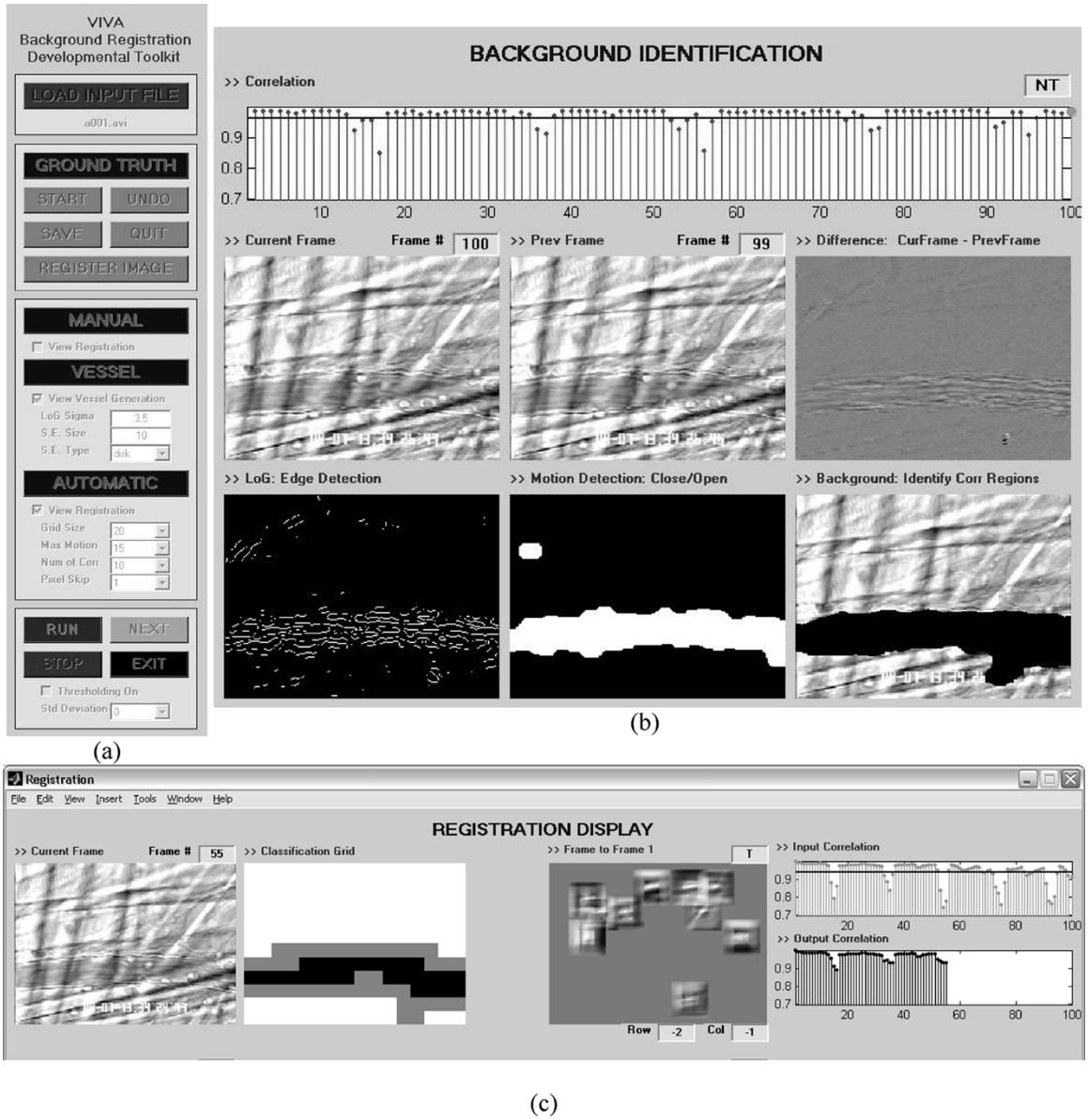

Fig. 2. Screen shots of the (a) fixed FOVBR controller, (b) background identifier, and (c) registration window showing the current frame, the classification grid, and the cross correlation of the random grid-cell locations for the one stabilization method.

For each sequence, we define the average normalized correlation without displacement as follows:

$$\Phi_{I_r} = \frac{1}{N}\sum_{t=1}^{N} C_{I_r,I_t}(0,0) \tag{2}$$

where N is the number of frames in the video sequence and the subscript of Φ_{I_r} means the average normalized correlation depends on the reference frame. The VIVA background registration system provides three alternatives for the reference frame I_r as described in detail in Section III-C.

Now let us define the S-frames and δ-frames in the sequence. The tth frame I_t in the sequence is defined as a δ-frame if $C_{I_r,I_t}(0,0) < \Phi_{I_r}$, and I_t is defined as an S-frame if $C_{I_r,I_t}(0,0) \geq \Phi_{I_r}$. Note that this correlation is a full frame cross correlation without translation, as opposed to the correlation of patches with limited translation used in the actual registration step. Also note that a more aggressive threshold could be used in determining the δ-frames, but we stipulate that a more conservative measure is appropriate, as registering stable frames (when misclassified as δ-frames) does not cause error in the registration process.

Step 2) Locate Motion Regions: We use stable frames to locate the regions of motion, avoiding any degradation caused by δ-frames. This is carried out by first determining a difference frame, which is found by pointwise subtracting the previous frame from the current frame. The difference frame provides intensity changes, or edges, where regions of motion exist [30]. For example, Fig. 2(b) contains an image of a difference frame in the upper right labeled as "Difference". The edges in the motion regions are created from venule contents, i.e., leukocytes, erythrocytes and platelets moving faster than can be recorded by the video camera.

Step 3) Build a Background Mask for Stabilization: We locate edges in the difference frame using the Laplacian-of-a-Gaussian (LoG) edge detector, a linear operator that uses Gaussian filtering to decrease the effect of noise and

detail and applies Laplacian filtering to determine the position of inflection points in intensity transitions [10], [31]. Edges are found by locating the zero crossings of the processed image W

$$W = \nabla^2 \left[G_\sigma(i,j) * I_d(i,j) \right] \tag{3}$$

where ∇^2 is the Laplacian operator $\partial^2/\partial x^2 + \partial^2/\partial y^2$, and G_σ is a Gaussian filter with standard deviation σ, $G_\sigma(x,y) = e^{-\left((x^2+y^2)/2\pi\sigma^2\right)}$, I_d is the difference frame. σ dictates the degree of smoothing performed on the original image, for the purpose of smoothing less significant edges found in the background and retaining prominent edges existing within the venule. To retain the prominent edges within the venule, we choose σ based on the radius, η, of the largest moving objects within the venule, that is $\eta = 2\sqrt{2}\sigma$. Here, we are observing leukocytes, where typically $\eta \approx 4\ \mu\text{m}$. Application of the LoG filter returns an image with the minimum distance between two edges, $\eta = 2\sqrt{2}\sigma$ [32]. Application of the LoG edge detector creates a binary image I_W. (The binary image is created by first computing the zero-crossings of the Laplacian of a Gaussian, and then defining edges at the zero crossing points for which the image intensity gradient magnitude exceeds a threshold determined by the average gradient magnitude in the image.) The bottom left image in Fig. 2(b) shows the result after applying LoG edge detector on the difference image. We combine edges to define the motion region by using the morphological close-open procedure: $Q = (I_W \bullet K) \circ K$ where Q is the mask of the motion regions, I_W is the input image, and K is a structuring element. Using dilation only to combine edges impedes shape consistency of the venule. A disk-shaped structuring element with a diameter $\lambda \geq \eta$ ensures the combination of edges having the minimum possible separation, since the minimal separation distance is defined by η.

The center image at the bottom of Fig. 2(b) shows the result of applying the close-open process to create an image mask, Q, which is composed of two pixel values, "0" and "1"; value "1" is associated with the background and value "0" the foreground. The image mask assigns each pixel in the current frame to the motion foreground region (areas inside the venules and capillaries) or to stable background regions (e.g., muscle striation and tissue). For each frame in the video sequence, we can obtain a motion mask. Hence, we can use these masks to create a background mask A for the entire sequence. The pixel value of mask A at location (i,j) is set to be 1 if the average of the pointwise sum of the masks at location (i,j) is larger than a prefixed threshold , otherwise, it is set to be 0. The threshold is typically set to ½. Simliar to the mask Q, mask A is also composed of two pixel values, "0" and "1"; "1" is associated with the background and "0" the foreground. A will be used for background registration in the second step of fixed FOVBR.

B. Background Registration

The second step of the jitter minimization procedure is background registration through local region correlation, which also employs a multistep process.

Step 1) Definition of the Classification Grid: We divide A obtained in the background detection stage into $u \times v$ nonoverlapped blocks which are called grid cells in this paper. Thus, we have total $[M/u] \times [N/v]$ cells, where M and N are the size of image A. Using the pixel values of the grid cell, each grid cell can be labeled as B, representing the background, V, the venule, and U, a combination of motion and stable regions [see Fig. 2(c)]. Let the (m,n)th grid cell be denoted by $H_{m,n}$, and $G_{m,n}$ be the corresponding label of $H_{m,n}$. $G_{m,n}$ can be obtained as follows: if all of the pixel values of $H_{m,n}$ equal to 1, then $G_{m,n}$ is set to label B; if all of the pixel values of $H_{m,n}$ are 0, then $G_{m,n}$ is set to label V; otherwise, $G_{m,n}$ is set to label U. We call $G = \{G(m,n)\}$ the classification grid of cells. Grid cells are labeled as B, V or U if all, some, or none of their pixels (i,j), respectively, correspond to pixel value 1 in A.

Step 2) Grid-Cell Selection: We build a set with K grid cells, $H = \{H_{m_1,n_1}, H_{m_2,n_2}, \ldots, H_{m_K,n_K}\}$, at random from locations where $G_{m_t,n_t} = B(t = 1,2,\ldots,K)$ for each δ-frame. The grid cells are the "patches" to be used in correlating two frames. We do not consider grid cells along the border of G in order to eliminate border correlation issues and any artifacts created during data collection and digitization. We size each grid cell to capture one distinct background feature. Muscle striation intersections are ideal landmark features because of their size, stability, and distinction from tissue. The width of a typical muscle striation is about two microns and can range in length, but a single striation commonly extends for more than 30 μm. We typically choose $(u,v) = (20,20)$ pixels $\sim (7,7)$ μm for $(M,N) = (280,200)$ pixels.

Step 3) Correlation: In this step, we determine the corrective translations from the δ-frames to the reference frame. In order to determine the corrective translation, we need to define the local cross correlation between a δ-frame and the reference frame. For convenience, we reorder the cells in $\mathbf{H}$ as $H_1, H_2, \ldots, H_K$ and let H_k be a grid cell selected from $\mathbf{H}$, I_t be a δ-frame, and I_r be the reference frame selected before the correlation is computed. I_{r,H_k} is the subimage in I_r covered by H_k, and $I_{t,\mathrm{H}_k,T_x,T_y}$ is the subimage in the δ-frame I_t covered by H_k translated by T_x, T_y. Then the local correlation between δ-frame, I_t, and the reference I_r is defined as the correlation between I_{r,H_k} and $I_{t,\mathrm{H}_k,T_x,T_y}$,denoted by $C^{H_k}_{I_r,I_t}(T_x,T_y)$.

The corrective translation for the tth frame using local cell H_k can be obtained by

$$(T_x(t,H_k), T_y(t,H_k)) = \underset{(T_x,T_y)\in\Theta}{\arg\max} \{C^{H_k}_{I_r,I_t}(T_x,T_y)\} \tag{4}$$

where Θ is the search space for corrective translation. In theory, Θ can be the entire two dimensional domain. However, because the size of the image is limited, thus, we can limit the values of T_x, T_y such that $I_{t,\mathrm{H}_k,T_x,T_y}$ is a subimage of I_t. Through close observation of hours of video we know δ-frames translate to a maximum of 5 μm ($\sim$ 15 pixels) in our experiments. Thus, we can define maxmotion, $m = [m_x, m_y]$, a parameter to limit the range of cross correlation, reducing correlations producing incorrect translations and computational cost. So, the search space is fixed by $\Theta = \{(T_x,T_y) : -m_x \leq T_x \leq m_x, -m_y \leq T_y \leq m_y\}$.

Step 4) Registration: For the tth frame (a δ-frame), we can obtain a corrective translation for each $H_k \in \mathbf{H}$, thus, we obtain a corrective translation vector related to $\mathbf{H}$, denoted by

$$T_{\mathrm{H}}(t) = [T_{\mathrm{H}_1}(t), T_{\mathrm{H}_2}(t), \ldots, T_{\mathrm{H}_K}(t)] \tag{5}$$

where

$$T_{\mathrm{H}_k}(t) = [T_x(t,H_k), T_y(t,H_k)]^T. \tag{6}$$

To align the tth frame, which is a δ-frame, we can apply corrective translation $\mathrm{mean}(T_{\mathrm{H}}(t))$, $\mathrm{median}(T_{\mathrm{H}}(t))$, or $\mathrm{mode}(T_{\mathrm{H}}(t))$ (where the mode is computed over the rounded integer $T_{\mathrm{H}}(t)$ values). Through 10 sequences (of 100 frames each) we compared the performance using $\mathrm{mean}(T_{\mathrm{H}}(t))$, $\mathrm{median}(T_{\mathrm{H}}(t))$, or $\mathrm{mode}(T_{\mathrm{H}}(t))$ and determined that $\mathrm{median}(T_{\mathrm{H}}(t))$ provides the best corrective translation in terms of average root-mean-square error (RMSE) (see the definition in Section V-B1), where the three statistics led to errors of 0.38, 0.28, and 0.29 μm, respectively.

C. Reference Frame Selection for Correlation Computation

From Sections III-A and III-B, we note that correlation computation requires a reference frame. We will consider three alternatives for the reference frame. The first alternative is to use the initial frame of the sequence as a reference frame and correlates every frame to the initial frame. One advantage of this approach is that every frame is correlated to the same reference frame, thus eliminating the problem of additive errors [21]. However, one drawback is the necessary condition for the initial frame to belong to S_1, the first set of stable frames. If the initial frame is a δ-frame, then all future frames will be corrected using a frame displaced from the majority of frames in the sequence, thus, registration cost increases.

The second alternative is to use the (pointwise) average frame of the sequence as the reference frame. The average frame of sequence $\{I_t, t = 1, \ldots, N\}$ is defined by $F_{\mathrm{ave}} = 1/N \sum_{t=1}^{N} I_t$, where $\sum$ is a pointwise sum of image intensities. The average frame is a smoothed image that eliminates moving features, i.e., the cells. In our experiments with murine venules, two-thirds of the frames in the original sequence are stable on average. Thus, the average frame retains the stable features such as the endothelium and muscle striations (although they are somewhat smoothed out by the jittered frames that enter the average).

The third alternative is to use 'frame X' as the reference frame. We define frame X as the frame I_t that maximizes the normalized correlation $C_{F_{\mathrm{ave}},I_t}(0,0)$, the correlation coefficient of I_t with F_{ave}. We assume at least two-thirds of the frames in the video sequence are stable; thus, Frame X is a stable frame. Frame X, like F_{ave}, retains the large stable features, but also maintains feature contrast, which reduces correlation error. For convenience, we refer to the correlation methods as F1, FA, or FX for reference frame chosen as the initial frame, the average frame, or frame X, respectively.

D. Intravital Cell Tracker

Although the focus of the paper is registration, we briefly discuss the method used to track cells observed *in vivo*, since the goal of the registration is the enabling of tracking.

Tracking of rolling leukocytes in a stable, fixed FOV may be accomplished *in vivo* using active contours. For tracking cells observed in a jittery video (and later in a moving FOV), we employ an active contour similar to that introduced by Kass *et al.* [33] but instead utilize gradient vector flow [34] as the external force that attracts the contour to the cell border. We also utilize the shape and size, and constraints introduced by Ray *et al.* [3].

Consider a closed active contour $\mathrm{r}(s) = (x(s), y(s))$ that is parameterized by $s (s \in [0,1])$. By minimizing an energy functional, we can compute the cell boundary for a single leukocyte. The functional is

$$E_{\mathrm{snake}} = \int_0^1 [E_{\mathrm{int}}(\mathrm{r}(s)) + E_{\mathrm{ext}}(\mathrm{r}(s)) + E_{\mathrm{shape}}(\mathrm{r}(s))]ds + E_{\mathrm{size}} \tag{7}$$

where E_{int} and E_{ext} represent the internal energy and the external energy, respectively, which is defined in [33]. E_{shape} represents the shape constraint energy which is defined as follows [3]:

$$E_{shape} = \frac{1}{2}\big(R_x(s,x(s)) - \bar{R}(\cos(2\pi s)\big)^2 + \frac{1}{2}\big(R_y(s,y(s)) - \bar{R}(\sin(2\pi s)\big)^2 \tag{8}$$

where $R_x(s,x(s)) = x(s) - \int_0^1 x(r)dr$, $R_y(s,y(s)) = y(s) - \int_0^1 y(r)dr$ and

$$\bar{R} = \int_0^1 \sqrt{R_x^2(s,x(s)) + R_y^2(s,y(s))}ds. \tag{9}$$

E_{shape} penalizes deviations from a circular contour. E_{size} represents the size constraint energy, which is defined as [3]

$$E_{\mathrm{size}} = \frac{1}{2}\big(\bar{R} - K\big)^2 \tag{10}$$

where K is an expected cell radius.

Using variational calculus techniques to minimize the energy functional (7), we can obtain the following Euler equations:

$$\alpha x_{ss} - \beta \mathrm{x}_{ssss} - \frac{\partial E_{\mathrm{ext}}}{\partial x} - \left\{x(s) - \int_0^1 x(r)dr - \bar{R}\cos(2\pi s)\right\} - \frac{(\bar{R} - K)(x(s) - \int_0^1 x(r)dr)}{\sqrt{R_x^2(s,x(s)) + R_y^2(s,y(s))}} = 0 \tag{11}$$

$$\alpha y_{ss} - \beta \mathrm{y}_{ssss} - \frac{\partial E_{\mathrm{ext}}}{\partial y} - \left\{y(s) - \int_0^1 y(r)dr - \bar{R}\sin(2\pi s)\right\} - \frac{(\bar{R} - K)(y(s) - \int_0^1 y(r)dr)}{\sqrt{R_x^2(s,x(s)) + R_y^2(s,y(s))}} = 0 \tag{12}$$

where $\mathrm{x}_{ss}(s)$, $\mathrm{x}_{ssss}(s)$ and $\mathrm{y}_{ss}(s)$, $\mathrm{y}_{ssss}(s)$ are the second and the fourth derivatives of $\mathrm{x}(s)$ and $\mathrm{y}(s)$, respectively. Because the external energy is application-based, thus, we can design different external energy from [33]. In our solution, we use GVF field defined in [34] as the external force. The gradient vector flow field is defined as the vector field $\mathrm{V}(x,y) = [u(x,y), v(x,y)]$ that minimizes the following energy functional [34]

$$\varepsilon = \int\int \mu(u_x + u_y + v_x + v_y) + |\nabla f|^2 |\mathrm{V} - \nabla f|^2 dxdy \tag{13}$$

where f is the edge map of the image (the gradient magnitude of the Gaussian-smoothed image), u_x, u_y, v_x, v_y are the partial derivatives of $u(x,y), v(x,y)$ and μ weights the importance of field smoothness in homogenous regions.

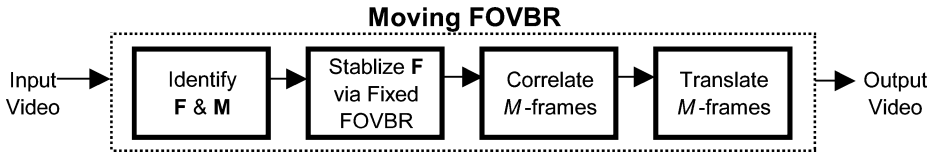

Fig. 3. Moving FOVBR flow chart. Input sequences are sectioned into fixed FOV frame sets, $\mathbf{F}$, and moving FOV frame sets M. The fixed FOVBR stabilizes all $\mathbf{F}$, then M-frames are registered and a registered output video is created.

Replacing the external forces $(-\partial E_{\text{ext}}/\partial x, -\partial E_{\text{ext}}/\partial y)$ in (11)–(12) with the gradient vector flow field $V(x(s), y(s)) = (u(x(s), y(s)), v(x(s), y(s)))$, and treating $\mathrm{r}(s)$ as function of time t (i.e., $\mathrm{r}(s,t)$) , we can obtain the active contour update as follows

$$\begin{aligned} x_t(s,t) =& \alpha x_{ss}(s,t) - \beta \mathrm{x}_{ssss}(s,t) + u(x(s,t), y(s,t)) \\ &- \left\{ x(s,t) - \int_0^1 x(r,t)dr - \bar{R}\cos(2\pi s) \right\} \\ &- \frac{(\bar{R}-K)(x(s,t) - \int_0^1 x(r,t)dr)}{\sqrt{R_x^2(s,x(s,t)) + R_y^2(s,y(s,t))}} \end{aligned} \tag{14}$$

$$\begin{aligned} y_t(s,t) =& \alpha y_{ss}(s,t) - \beta \mathrm{y}_{ssss}(s,t) + v(x(s,t), y(s,t)) \\ &- \left\{ y(s,t) - \int_0^1 y(r,t)dr - \bar{R}\sin(2\pi s) \right\} \\ &- \frac{(\bar{R}-K)(y(s,t) - \int_0^1 y(r,t)dr)}{\sqrt{R_x^2(s,x(s,t)) + R_y^2(s,y(s,t))}}. \end{aligned} \tag{15}$$

For tracking, the final contour in a given frame is used as the initial contour in the subsequent frame. As long as the distance between two consecutive cell center positions is less than the cell diameter, the active contour will be able to capture the leukocyte in the subsequent frame [35], which, at 30 frames per second, corresponds to a maximum cell velocity of roughly 200 μm/s [35].

IV. Moving FOV Background Registration

In this section, we describe moving FOVBR. A moving FOV is created when the stage of the microscope is translated in the x, y-plane by some amount. In our system, a joystick, operated by the biomedical researcher, controls the speed and direction of the stage movement. Thus, frames within a video sequence can be defined as either *fixed* – frames belonging to a fixed FOV set of contiguous frames, $\mathbf{F}$, or *motion* – frames belonging to a moving FOV set of contiguous frames, $\mathbf{M}$. Generally, a video sequence includes a series of fixed FOV sets and moving FOV sets. Assume that video sequence has L fixed FOV sets where the first and last fixed FOV set are fixed FOV sets. Therefore, we have $L-1$ moving FOV sets. Listing all of the successive fixed FOV sets in the sequence by $\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_L$, and all of the successive moving FOV sets by $\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_{L-1}$, we can define two supersets $\hat{\mathbf{F}}$ and $\hat{\mathbf{M}}$, where $\hat{\mathbf{F}} = \{\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_L\}$ and $\hat{\mathbf{M}} = \{\mathbf{M}_1, \mathbf{M}_2, \ldots, \mathbf{M}_{L-1}\}$. For convenience, we call $\hat{\mathbf{F}}$ and $\hat{\mathbf{M}}$ the fixed set and motion set within a sequence, respectively.

The moving FOVBR accepts audio-video interlaced (avi) files video sequences having a combination of fixed FOV sets and moving FOV sets. We assume 1) the first and last frame sets are fixed FOV sets, 2) the speed and direction of the stage during a motion set $\mathbf{M}$ is constant, and 3) successive frames in a motion set $\mathbf{M}$ share a common overlap region, W. The moving FOVBR registers the video in three steps, first identifying $\hat{\mathbf{F}}$ and $\hat{\mathbf{M}}$, next stabilizing F-frames, and finally registering M-frames. Fig. 3 provides a flow chart describing each step of the moving FOVBR.

A. Fixed and Motion Frame Identification

We classify every frame in $\mathbf{I}$ as fixed or motion using two main criteria, frame correlation between the frame I and the reference frame I_r, $C_{I_r,I}(0,0)$ and Laplacian sum, which is defined as

$$\Gamma = \sum_{i=1}^{m}\sum_{\mathrm{j}=1}^{n} \left|\nabla^2 I(i,j)\right| \tag{16}$$

where

$$\nabla^2 I(i,j) = \frac{\partial^2 I(i,j)}{\partial x^2} + \frac{\partial^2 I(i,j)}{\partial y^2}. \tag{17}$$

We call I an M-frame if $C_{I_r,I}(0,0) < 1/N \sum_{i=1}^{N} C_{I_r,I_i}(0,0)$ and $\Gamma_I < 1/N \sum_{i=1}^{N} \Gamma_{I_i}$ where $N = |I|$. Otherwise, I is classified as an F-frame when $C_{I_r,I}(0,0) \geq 1/N \sum_{i=1}^{N} C_{I_r,I_i}(0,0)$ or $\Gamma_I \geq 1/N \sum_{i=1}^{N} \Gamma_{I_i}$. M-frames have less contrast than F-frames due an insufficient camera frame rate. Fig. 4 displays stem plots of frame to frame correlations and Laplacian sums which define two moving FOV sets for sequence $\mathbf{I}$ where $|\mathbf{I}| = 199, |\hat{\mathbf{M}}| = 2, |\mathbf{M}_1| = 30$, and $|\mathbf{M}_2| = 27$.

To minimize jitter we apply fixed FOV registration to $\mathbf{F}$ using the fixed FOVBR. According to our analysis and experiments, the most effective choice for the reference frame in the fixed FOVBR is frame X – the frame with the highest correlation to the pointwise average of the frames in the set. For completeness, we tested each reference frame alternative (F1, FA, FX) and used the reference frame that yielded the highest frame-to-frame correlation. Three videos are created for each $\mathbf{F}$ using F1, FA, and FX methods. We select the registered video for which the Laplacian magnitude sum of the pointwise average frame, $\bar{\Gamma}$ [which can be computed using the average frame in place of the individual frame in (16)], of the average frame is maximal. Finally, to attain a uniform frame size, we automatically detect and crop regions in which $I(i,j) = 0$ (introduced along the borders of the stabilized video).

B. Fixed FOV Registration Within Motion FOVBR

The normalized cross correlation based registration includes two steps. The first step is the computation of the correlation between two images, I_1 and I_2 with displacement, $C_{I_1,I_2}(x,y)$.

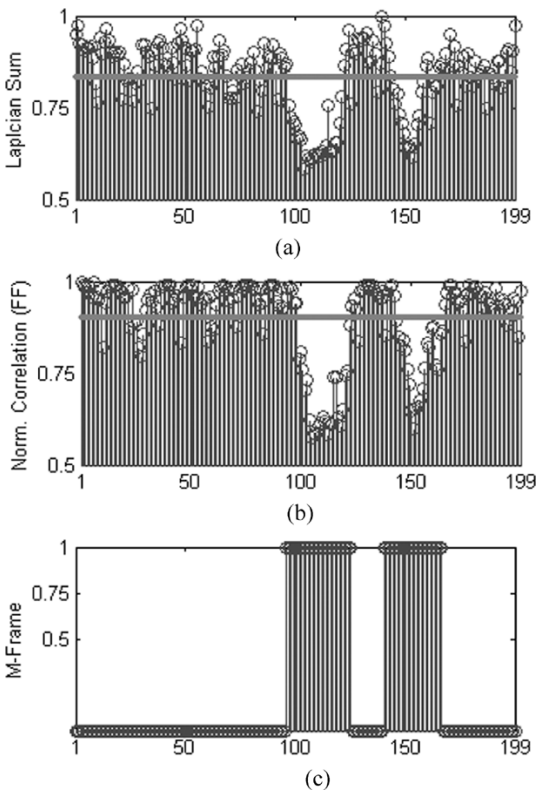

Fig. 4. Plots describing fixed and moving FOVs identification. (a) Laplacian sum of each frame in the sequence with expected Laplacian energy line. (b) Frame to frame correlation with expected correlation line. (c) Fixed FOV frames identified as zeros and motion frames shown as ones. Thus, a grouping of zeros is a fixed FOV set, $\mathbf{F}$, and a grouping of ones is a motion set, $\mathbf{M}$.

The second step is the computation of the translation which has the largest associated correlation, denoted by $T = [x_o, y_o]$

$$T = [x_o, y_o] = \arg\max_{x,y}[C_{I_1,I_2}(x,y)]. \tag{18}$$

We use T to register the image being matched, I_2 to image I_1.

The goal of successive fixed FOV registration is to register two pivot frames, the final frame, F_f, of a fixed FOV set $\mathbf{F}_g$ and the initial frame, F_i, of the following fixed FOV set $\mathbf{F}_{g+1}$, where g is the index numbering the fixed FOV sets.

Although we can use the translation obtained by (18) to register the final frame of a fixed FOV set and the initial frame of the subsequent fixed FOV set directly, the results may be unreliable due to similar features shared by these two "pivot" frames. Therefore, we use a multistep process to refine the corrective translation between the pivot frames. The first step is to estimate the translation between the pivot frames. This estimation can be accomplished by computing the translation between two frames in the moving FOV set $\mathbf{M_g}$ by (18). Here, $\mathbf{M_g}$ is the moving FOV set positioned in time between the two fixed FOV sets $\mathbf{F}_g$ and $\mathbf{F}_{g+1}$. The two frames used in $\mathbf{M_g}$ to estimate the translation between the pivot frames are the $(k/2-1)$th frame and the $(k/2+2)$th frame, where k is the number of frames in $\mathbf{M_g}$. Denoting the $(k/2-1)$th frame by I_1 and the $(k/2+2)$th frame by I_2 and applying (18), we can obtain the translation, denoted by T_{AB}, between I_1 and I_2. Because there are two frames between $(k/2-1)$th frame and the $(k/2+2)$th frame, the coarse estimation of the average translation between any two neighboring frames in $\mathbf{M_g}$ is $T_{AB}/3$. Thus, the translation between the two pivot frames F_f and F_i is estimated to be $(k+1)T_{AB}/3$.

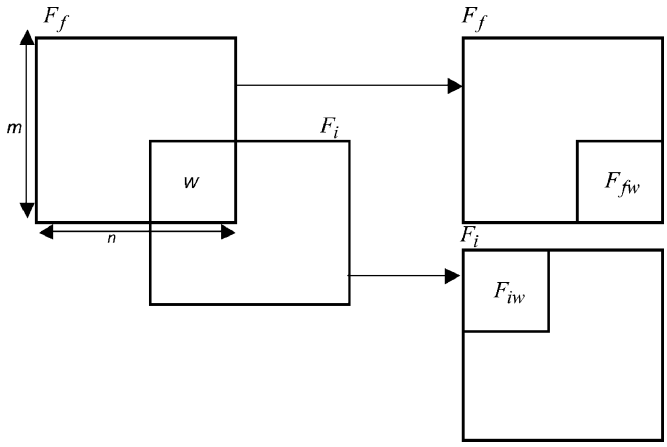

Fig. 5. Common overlap region of adjacent pivot frames. Diagram shows pivot frames F_f and F_i sharing common region W. F_{fw} and F_{iw} denote the common region in each pivot frame.

Using the coarse estimation of the translation between the two pivot frames, we can obtain a common region W, shared by the two pivot frames. We denote these two regions as F_{fw} in the final frame of the $\mathbf{F}_g$ and F_{iw} the initial frame of $\mathbf{F}_{g+1}$ as shown in Fig. 5. We then refine the corrective translation by applying the translation, T_W, found using (18), where I_1 and I_2 are replaced by F_{fw} and F_{iw} . Here, the corrected translation between the two pivot frames is $(k+1)T_{AB}/3+T_W$. Using the two pivot frames from $\mathbf{F}_g$ and $\mathbf{F}_{g+1}$, we can create a mosaic which is called the *framemap* (see Fig. 6 for an example) for the gth moving FOV set $\mathbf{M_g}$, denoted by β_g. This technique is applied to all frames in $\mathbf{F}$ and used to create a set of framemaps $\beta = \{\beta_g, g = 1, 2, \ldots, L-1\}$ for the entire sequence.

C. Motion Frame Registration Within Motion FOVBR

We use the obtained framemap β_g to register each frame in $\mathbf{M}_g$ In order to register the tth M-frame, I_t, we reduce the size $(m \times n)$, defining a placement frame, $R_t = I_t((m/r):(3m/r),(n/r):(3m/r))$, where typically $r = 4$. We determine the translation of the current M-frame using (18), where I_1 and I_2 are replaced by R_t *and* β_g. Correlating with R_t, instead of I_t minimizes incorrect translations due to correlation errors caused by framemap regions where $\beta_g(i,j) = 0$ (which draws I_t to the $\beta_g(i,j) = 0$ edges since the edges of the FOV are artificial features that do not correspond to physical features in the scene). Here, $\beta_g(i,j)$ is the pixel value of β_g at location (i,j). We apply this technique to each moving FOV set and create an *overlay map*, Ω, which mosaics every frame in the sequence. The overlay map simply provides a translation of every point in every frame with reference to the first frame of the first fixed FOV portion. Since every frame is registered, we are able to track leukocytes in a video sequence containing both fixed and moving FOVs. Fig. 7 shows an example overlay map Ω, where $|\mathbf{I}| = 160$, $|\mathbf{M}| = 2$, and $||\mathbf{M}|| = 37$, and the leukocyte motion is denoted by the dotted path. The rectangular outlines in Fig. 7 show the boundaries of the constituent frames that form the sequence.

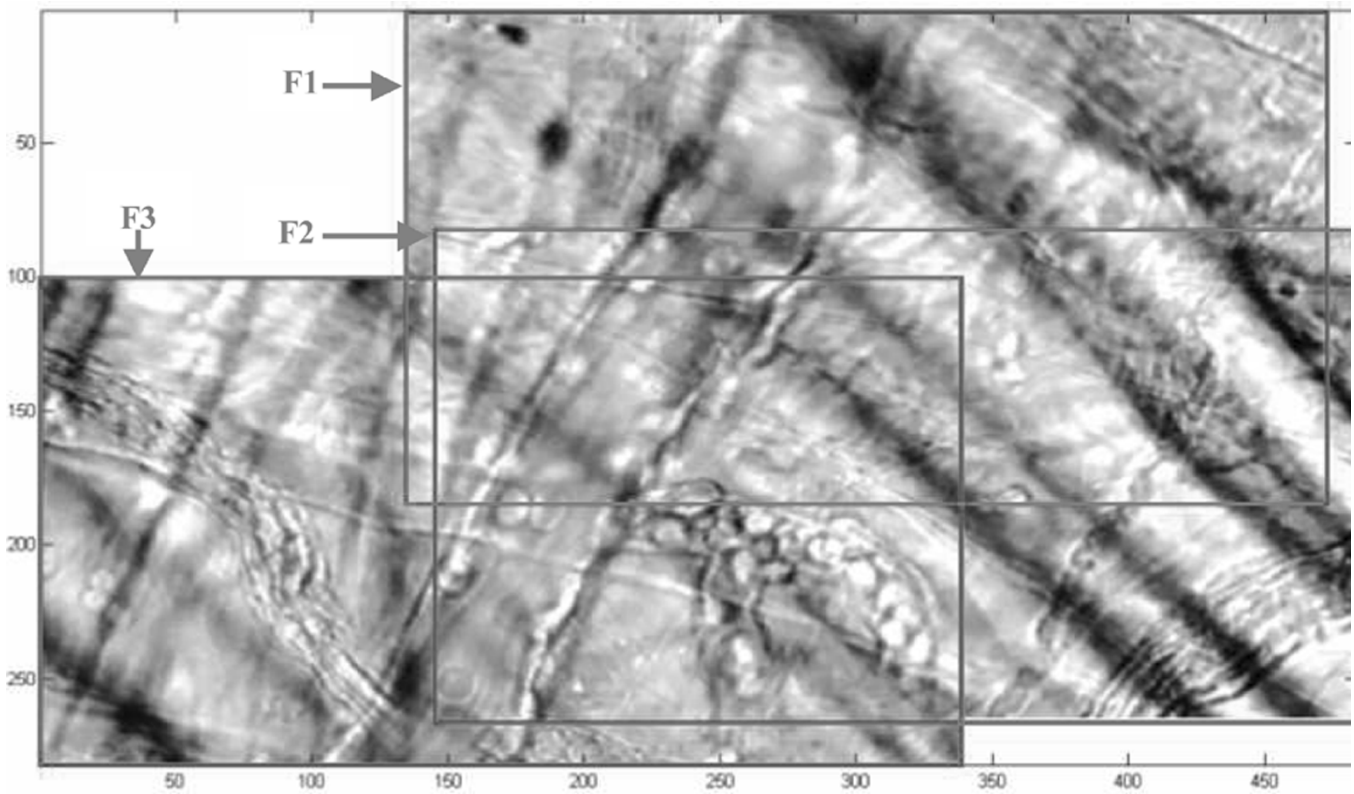

Fig. 6. Example of a framemap, β, built using 4 pivot frames (the final frame, of a fixed FOV set and the initial frame of the subsequent fixed FOV set are shown). The framemap of a sequence is an image mosaic that gives an extended FOV. β is instrumental for registering all frames within the motion sets. When viewing the sequence from which this framemap was obtained, the stage movement produces the impression of the camera moving down and to the right over the first FOV, M_1 and then down and to the left over the second FOV M_2. Superimposed lines represent constituent frame boundaries.

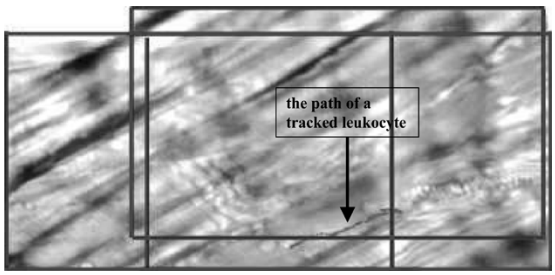

Fig. 7. Example overlay map (mosaic) showing the path of a tracked leukocyte.

V. Results

In this section, we first outline the data, equipment, and video specifications. Next we provide fixed FOVBR stabilization and tracking results. We conclude the FOVBR (with the three reference frame choices of F1, FA, or FX) is capable of minimizing jitter as necessary to accommodate accurate cell tracking. We then give the moving FOVBR registration and tracking results, demonstrating the realization of automatic tracking over moving FOVs.

A. Data, Equipment, and Video Specifications

All intravital video microscopy data were collected from the cremaster muscle of C57BL/6 wild-type mice. The cremaster muscle was prepared for intravital video microscopy as described in [36], where the cremaster microcirculation is exposed by pinning the epididymis and testes to the side [10]. The cremaster is superfused with thermocontrolled (35°C) bicarbonate-buffered saline [36]. Post-capillary venules between 20–40 μm in diameter were observed with a charge-coupled device camera system (model VE-1000CD; Dage-MTI, Michigan City, IN) and recorded using a Panasonic S-VHS recorder. A JVC S-VHS recorder was connected to a Macintosh G4 (dual 1.0-GHz processors, 1.0 GB RAM) with an IEEE 1394 cable. Video was digitized and stored as avi files (uncompressed, 320×240 pixels with a pixel-to-micron ratio of 3.16, 256 gray levels, 30 fps) using iMovie and Quicktime Pro. At these specifications, a 100 frame video sequence requires 7509 KB of storage. All registration experiments were performed on a Dell 4500 desktop computer (2.4-GHz processor, 1.0 GB DDR RAM). On average, one fixed FOVBR experiment on a 100-frame sequence, executing all four methods and saving 4 avi sequences, 13 jpeg images, 2 Microsoft Excel files, and 1 Matlab "mat" file (a file format used by Matlab for storing data) takes just over 2 min. One moving FOVBR trial on a 458 frame sequence, saving 16 avi sequences, 36 jpeg images, 1 Excel file, and 2 mat files, requires less than 4 min of processing.

B. Fixed FOVBR Results

1) Performance Measures for Stabilization: We determine the performance of the stabilization methods by way of the RMSE and the normalized Laplacian sum. We calculate the

RMSE of the corrective translations obtained by stabilization method described in Section III. Using the ground truth (manual) translations as the true translations

$$\text{RMSE} = \sqrt{\text{mean}[(x_{gi} - x_i)^2 + (y_{gi} - y_i)^2]} \quad (19)$$

where x_i, y_i are the corrective translations obtained by stabilization method and x_{gi}, y_{gi} are ground truth (manual) translations, here g denotes the ground truth position. If we assume the ground truth sequences to be jitter-free, then the position error provided by the RMSE serves to validate the fixed FOVBR acting as a stabilization measure. For the second measure, we use the maximum error of the corrective translations, which is defined as

$$\text{Maximum error} = \left\{ \sqrt{\max[(x_{gi} - x_i)^2 + (y_{gi} - y_i)^2]} \right\} \quad (20)$$

where x_i, y_i and x_{gi}, y_{gi} have the same meaning as (19).

For the third measure we compare the high frequency content by computing the sum of the Laplacian magnitude, $\bar{\Gamma}$, of the average frame for the original input video, the average frame of fixed FOVBR generated output videos, and the average frame of manually registered video. Besides computing the sum of the Laplacian magnitude of the average frame (obtained using all frames in the video sequence) of video sequences, we also compute the sum of the Laplacian magnitude of the average δ-frame obtained. This average is computed using only δ-frames defined by the ground truth sequence – frames where $T_{xy} \geq 0.5$ pixels. For convenience, we denote the latter sum of the Laplacian magnitude by $\bar{\Gamma}_{\text{j}}$. Because the fixed FOVBR and manual registration tool stabilize δ-frames, the average frame (either obtained using all of the frames or just δ-frames) from the output videos will be sharper, containing more contrast, thus having stronger high frequency components. The method with the greatest $\bar{\Gamma}$ signifies the output video with the least amount of jitter.

2) Results: We first provide the experimental results regarding performance of our methods for δ-frames identification. We viewed each frame of four sequences (100 frames each) to manually identify δ-frames and then compared the results of the manually detected δ-frames to the δ-frames detected automatically by the fixed FOVBR. For four sequences we computed the average number of true positives (frames labeled δ-frames both manually and by the fixed FOVBR), true negatives (frames considered S-frames both manually and by the fixed FOVBR), false positives (frames labeled S-frames manually, but deemed δ-frames by the fixed FOVBR), and false negatives (frames marked δ-frames manually, but labeled S-frames by the fixed FOVBR). The success of the fixed FOVBR varies according to the reference frame used in terms of true positives (20.8 for F1, 18.8 for FA, 20.0 for FX), true negatives (70.8 for F1, 73.2 for FA, 73.0 for FX), false positives (6.6 for F1, 3.8 for FA, 3.8 for FX), and false negatives (1.8 for F1, 4.2 for FA, 3.2 for FX). While the number of false results in summation (false positives plus false negatives) is approximately the same for F1, FA, and FX, the use of F1 led to a lower number of false negatives and a higher number of false positives. A false positive, in this application, may be more desirable as an S-frame can be registered with zero registration. A false negative means that a δ-frame goes unregistered, which could be deleterious to the cell tracking system.

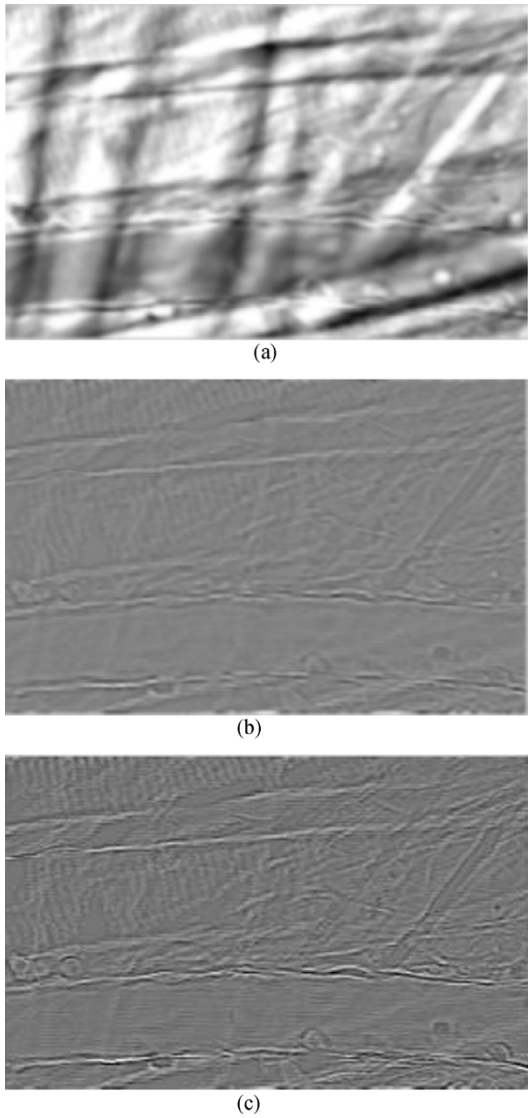

Fig. 8. (a) The pointwise average of the original sequence; (b) Laplacian image of (a); and (c) the Laplacian image of the average frame of the stabilized videos using FX method.

We also provide the experimental results on performance of our method for registration. Thirty video sequence were used in the experiments. The RMSE, maximum error and Laplacian sum defined in Section V-B1. were used as performance measures. Note that the Laplacian sum is computed by summing the Laplacian magnitude of the average frame, where the average frame refers to the pointwise average taken after registration. Fig. 8 shows the average frame of an input video and the Laplacian image of the average frame of the stabilized video using the FX reference frame. From Fig. 8, we can see qualitatively that the Laplacian magnitude of the average frame after registration is higher that that of the image before registration. Table II shows the results for 30 fixed FOV sequences of 100 frames where the maximum translation was $\mathbf{m} = (15, 15)$ pixels. In the experiments, we found that the correlation alternatives using a fixed reference frame (F1, FA, and FX) produced an average

TABLE II
STABILIZATION RESULTS FOR 30 FIXED FOV SEQUENCES

Average	I	F1	FA	FX	Manual
RMSE [μm]	1.921	0.434	0.448	0.462	n/a
Maximum Error	4.472	0.626	0.972	0.982	n/a
$\overline{\Gamma}$	0.843	0.928	0.973	0.978	0.949
$\overline{\Gamma}_{\delta}$	0.002	0.910	0.980	0.979	0.826

I – Original input sequence
F1 – frame to initial frame
FA – frame to average frame
FX – frame to frame X
RMSE – root mean-squared error
$\overline{\Gamma}$ – Laplacian sum
$\overline{\Gamma}_{\delta}$ – Laplacian sum computed only using δ- frames
Maximum Error- the maximum RMSE in the sequence

$\mathrm{RMSE} < 0.5\ \mu\mathrm{m}$ in 30 sequences. Fixed FOVBR with reference frame F1 produced the lowest average RMSE of 0.434 μm for all thirty sequences. Using FX as the reference frame produced the highest Laplacian sum, while FOVBR with the FA reference frame achieved the highest Laplacian sum when computed only on the δ-frames (0.980, see Table II). Although physically less meaningful than the RMSE in microns, the Laplacian sum gives an unbiased measure of stabilization. The RMSE has a slight bias toward the F1 frame selection, as this is the frame reference used in the manual ground truth measurements.

The parity in results between F1, FA, and FX suggests each method successfully stabilizes jitter. All three methods improve Γ by at least 10% for all frames and by nearly three orders of magnitude when only considering δ-frames. The FA and FX approaches performed in a similar range to method F1 with regard to average RMSE (0.434 μm for F1, 0.448 μm for FA, 0.462 μm for FX), maximum RMSE (0.626 μm for F1, 0.972 μm for FA, 0.982 μm for FX), and minimum RMSE (0.125 μm for F1, 0.133 μm for FA, 0.125 μm for FX). The slight improvement in RMSE of F1 can be explained by the similarity to the manual correlation, in which all frames are registered to the initial frame of the sequence. Thus, we regard the RMSE results of the three methods using a reference frame to be comparable. With a lack of "gold standard" data, the manual measurements, based on matching with the initial frame, are the best possible "ground truth" results attainable.

Third, we provided the experimental results on performance comparison of cell tracking before and after the stabilization. In the experiments, 15 sequences, each of which has 31 frames (1 s of video), were used. In 15 sequences, where the modified GVF active contour failed to track due to jitter, the fixed FOVBR stabilized the sequence to allow the tracker to succeed in tracking the leukocyte for the entire 31 frames. Table III shows the percentage of frames tracked for the original and stabilized sequences and root-mean-square difference (RMSD) between cell positions as computed with and without registration. The percentage of frames tracked uses manually collected ground truth – this measure shows the improvement obtained from fixed FOVBR. (We consider the leukocyte "tracked" if the computed cell center is within one cell radius ($\sim 3-4\ \mu$m) of the manually observed cell center [3].) The RMSD measure does not validate the fixed FOVBR registration, but serves as an indication of the amount of adjustment incurred by automated registration.

Given the original unregistered sequences, the automated tracker, on average, tracked cells for just over one-third of the frames. When jitter was minimized by the fixed FOVBR, the target cell was tracked in every frame of the sequence.

TABLE III
TRACKING RESULTS FOR 15 FOV SEQUENCES

	% Frames Tracked		RMSD [μm]
Sequence	Original	Stabilized	
1	29.0	100.0	6.35
2	48.4	100.0	4.65
3	64.5	100.0	1.90
4	77.4	100.0	1.36
5	6.5	100.0	3.31
6	38.7	100.0	3.03
7	6.5	100.0	4.20
8	58.1	100.0	1.69
9	19.4	100.0	1.82
10	25.8	100.0	3.43
11	25.8	100.0	1.99
12	19.4	100.0	6.23
13	25.8	100.0	2.58
14	54.8	100.0	1.70
15	51.6	100.0	2.30
Ave	36.8	100.0	3.10

RMSD measures position difference between active contour tracker in unregistered and registered sequence.

3) Fixed FOVBR Discussion: Tracking while minimizing misinformation on cell statistics is the ultimate goal of intravital registration; thus, we define success for the fixed FOVBR as $\mathrm{RMSE} < 0.5\ \mu$m, nearly one-sixteenth the diameter of a typical leukocyte. To achieve an average $\mathrm{RMSE} < d$, a desired error, for the entire sequence, the normalized correlation coefficient must be greater than a threshold for at least $\lfloor k/2 \rfloor + 1$ grid-cells on each δ-frame, where k is the number of grid cells in $\mathbf{H}$. Experimentally, we found only when the correlation coefficient is more than 0.93 can we attain an $\mathrm{RMSE} < 0.5\ \mu$m, which was the result for 24 of the 30 sequences.

C. Moving FOVBR Results

In 33 sequences the moving FOVBR identified all frames in 32 of 53 motion sets without recording a false negative (moving FOVBR grouped the frames in $\mathbf{F}$ while manual assessment examines the frames only in $\mathbf{M}$) but on average, each $\mathbf{M}$ included 1.6 false positives. Of the total number of false detections, (positive and negative), 19% were false negatives.

We registered 33 video sequences containing fixed and moving FOVs, 16 sequences with one motion set, 12 having two motion sets, and five containing three motion sets. We validated our registration by manually marking a common landmark feature and determining the RMSE for 10 representative sequences. The sequences contained an average of 42.5 M-frames. We recorded an average RMSE of 0.58 μm. The average error is on par with the average error reported for the tracker itself [3].

We deem the moving FOVBR successful for this data set, because the RMSE for each sequence was $< 1\ \mu$m, about 1/4 the radius of a leukocyte. For intravital cell tracking, the error for successful estimation of position has been bounded above by one cell radius [10].

We were able to track 11 leukocytes in 11 sequences containing moving FOVs. Table IV shows the number of frames

TABLE IV
CELL TRACKING STATISTICS FOR 11 LEUKOCYTES OVER CHANGING FOVs

Sequence	Frames Tracked	$\|\mathbf{M}\|$	v [μm/s]	Max Error [μm]	RMSE [μm]
1	140	35	0.22	3.98	0.92
2	280	67	0.61	2.87	1.05
3	190	15	0.72	5.52	0.95
4	100	37	2.03	4.27	1.88
5	80	67	1.76	3.53	1.39
6	80	21	1.71	4.24	1.53
7	170	15	0.86	3.60	0.85
8	160	37	2.00	2.96	0.85
9	100	37	0.38	2.08	1.00
10	30	17	1.01	4.94	2.43
11	140	40	1.10	2.44	1.13
Ave					1.27

$\|\mathbf{M}\|$ - total number of M-Frames in all motion sets, M
v - average leukocyte velocity calculated using the automated tracker
Max Error - max distance between the automated tracker and manual tracking position
RMSE - position error between the automated and manual tracking displacements

the leukocyte was tracked, the number of motion frames ($||\mathbf{M}||$), velocities (v) are calculated and tabulated (in Table IV) using position measurements for every tenth frame. To validate the moving FOVBR we manually tracked each cell in the registered sequences and determined the RMSE for tracker position. The RMSE calculation utilizes the leukocyte position in every frame.

D. Moving FOVBR Discussion

Accurately identifying fixed and motion frames is a crucial step in registering frames in changing FOVs. Identifying fixed frame sets is critical to stabilizing jitter while identifying desired stage motions is essential to proper alignment of successive $\mathbf{F}$. Cross correlation of local regions proves to be an accurate matching metric for registering intravital microscopy images in a fixed FOV, recording an average $\mathrm{RMSE} < 0.5\ \mu\mathrm{m}$, while cross correlation of R on β yields an $\mathrm{RMSE} < 0.6\ \mu\mathrm{m}$ for registration of moving FOV frames.

To reduce the amount of δ-frames within the set of frames $\hat{\mathbf{F}}$ that are identified as M-frames it is preferable that $|\mathbf{M}| \geq k$, where typically $k = 10$. For improved motion frame detection we increase the size of $\mathbf{M}$ by $2b$ frames, changing b F-frames prior to $\mathbf{M}$ and b F-frames following $\mathbf{M}$ to M-frames; typically $b = 3$. Although over-estimating $|\mathbf{M}|$ (labeling frames as motion frames that are not motion frames) is not desired, effects are beneficial (when $|\mathbf{M}| < a$, $a \approx 3$) by reducing the number harmful false negative identifications—when the moving FOVBR incorrectly identifies an M-frame as an F-frame. M-frames commonly have displacements greater than $\mathbf{m}$, the maximum translation the fixed FOVBR can correct, thus, misidentification of M-frames is detrimental to the stabilization of $\mathbf{F}$. Often the moving FOVBR fails to correctly label the moving FOV frames at the beginning (between 1 and 3 frames) and at the end of the moving FOV set of frames (also between 1 and 3 frames). So, by changing b frames before and after the moving FOV from fixed FOV frames to moving FOV frames, we protect these frames from undergoing fixed FOV registration. We do not want the moving FOV frames to be labeled as fixed FOV frames because fixed FOV registration will actually crop the entire fixed FOV set of frames by an amount equal to the greatest shift of the moving FOV frame. So, if a moving FOV frame is labeled as a fixed FOV, the entire fixed FOV set suffers in terms of limitation in image size.

One obstacle to tracking in a moving FOV is image blur. Fast stage movements produce blurry images that cause the tracker to lose the target cell. Fortunately, we can control the speed of the stage and experimentally determine an optimum stage speed. To improve registration, automated stage movement is necessary. Other tracking challenges include changing camera focus, microbursts, and clutter. Since we are tracking over an extended range, changes in the camera focus are often necessary to keep the cell in view. Microbursts are sudden jumps by the leukocyte, often greater than a leukocyte diameter, between consecutive frames [37] and always cause the tracker to lose the target cell. Clutter, such as muscle striations and other cells can generally cause the tracker to fail. These challenges set the current limits to tracking in a moving FOV.

VI. CONCLUSION

This paper introduced a system capable of registering intravital microscopy video for fixed and moving FOVs providing improved tracking of leukocytes *in vivo*. The novelty of the fixed FOV registration system lies in the approach to automatically eliminate motion regions from possible influence on frame registration. We have validated our method and demonstrated its effectiveness using position error. Our tracking data for 15 leukocytes in a fixed FOV verifies that without video stabilization to minimize jitter, accurate cell tracking is not possible. Even minimal frame displacements cause misinformation. Moreover, we have used a Laplacian-based stabilization metric, Γ, that allows for numerical comparison between three capable correlation methods.

We demonstrated the feasibility of registration in a moving FOV, provided the stage move at a constant speed, constant direction, and there exists a common region between successive fixed FOVs. Using the registered video, we were able to track 11 leukocytes in a moving FOV using a special-purpose active contour. Our new method will be useful for tracking rolling leukocytes *in vivo* over an extended range, providing more accurate information more efficiently for the validation of anti-inflammatory drugs.

REFERENCES

[1] U. Jung *et al.*, "Transit time of leukocytes rolling through venules controls cytokine-induced inflammatory cell recruitment *in vivo*," *J. Clin. Investigat.*, vol. 102, pp. 1526–1533, 1998.
[2] E. J. Kunkel, J. L. Dunne, and K. Ley, "Leukocyte arrest during cytokine-dependent inflammation *in vivo*," *J. Immunol.*, vol. 164, pp. 3301–3308, 2000.
[3] N. Ray, S. T. Acton, and K. Ley, "Tracking leukocytes *in vivo* with shape and size constrained active contours," *IEEE Trans. Med. Imag.*, vol. 21, no. 10, pp. 1222–1235, Oct. 2002.
[4] D. Stanimirovic and K. Satoh, "Inflammatory mediators of cerebral endothelium: A role in ischemic brain inflammation," *Brain Pathol.*, vol. 10, pp. 113–126, 2000.
[5] H. Weiler *et al.*, "Characterization of a mouse model for thrombomodulin deficiency," *Arterioscler. Thromb. Vasc. Biol.*, pp. 1531–1537, 2001.
[6] A. P. Goobic, M. E. Welser, S. T. Acton, and K. Ley, "Biomedical application of target tracking in clutter," presented at the 35th Asilomar Conf. Signals, Systems and Computers, Pacific Grove, CA, Nov. 4–7, 2001.
[7] K. Ley, "Leukocyte recruitment as seen by intravital microscopy," in *Physiology of Inflammation*, K. Ley, Ed. New York: Oxford Univ. Press, 2001, pp. 303–337.
[8] N. Manjunath, P. Shankar, B. Stockton, P. D. Dubey, J. Lieberman, and U. H. von Andrian, "A transgenic mouse model to analyze CD8+ effectro T cell differentiation *in vivo*," *Proc. Nat. Acad. Sci. USA*, vol. 96, pp. 13 932–13 937, 1999.

[9] S. B. Forlow *et al.*, "Severe inflammatory defect and reduced viability in CD18 and E-selectin double mutant mice," *J. Clin. Investigat.*, vol. 106, pp. 1457–1466, 2000.
[10] S. T. Acton, K. Wethmar, and K. Ley, "Automatic tracking of rolling leukocytes *in vivo*," *Microvasc. Res.*, vol. 63, pp. 139–148, 2002.
[11] J. Dunne, "Role of Beta 2 Integrins and ICAM-1 in Cytokine-Dependent Leukocyte Rolling and Adhesion," Ph.D. dissertation, Univ. Virginia, Sch. Eng. Appl. Sci., 2003.
[12] L. G. Brown, "A survey of image registration techniques," in *ACM Computing Surveys*, vol. 24, Dec. 1992, pp. 325–376.
[13] L. Shapiro and G. Stockman, *Computer Vision*. NJ: Prentice Hall, 2001.
[14] M. Hansen *et al.*, "Real-time scene stabilization and mosaic construction," in *Proc. 2nd IEEE Workshop Applications of Computer Vision*, Dec. 5–7, 1994, pp. 54–62.
[15] R. Kurazume and S. Hirose, "Development of image stabilization system for remote operation of walking robots," in *Proc. Int. Conf. Robotics Automation*, San Francisco, CA, Apr. 2000, pp. 1856–1861.
[16] S. Araki, T. Matsuoka, H. Takemura, and N. Yokoya, "Real-time tracking of multiple moving objects in moving camera image sequences using robust statistics," in *Proc. 14th Int. Conf. Pattern Recognition*, vol. 2, Aug. 16–20, 1998, pp. 1433–1435.
[17] B. Liu and A. Zaccarin, "New fast algorithms for the estimation of block motion vectors," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 3, no. 2, pp. 148–157, Apr. 1993.
[18] C. Morimoto and R. Chellappa, "Fast electronic image stabilization," in *IEEE Proc. 13th Int. Conf. Pattern Recognition*, vol. 3, 1996, pp. 284–288.
[19] K. Ratakonda, "Real-time digital video stabilization for multi-media applications," in *Proc. Int. Symp. Circuits and Systems*, vol. 4, May-Jun. 31–3, 1998, pp. 69–72.
[20] K. Uomori *et al.*, "Automatic image stabilizing system by full-digital signal processing," *IEEE Trans. Consumer Electron.*, vol. 36, no. 3, pp. 510–519, Aug. 1990.
[21] V. Zagorodnov and P. Ramadge, "Error stabilization in successive estimation of registration parameters," in *Proc. Int. Conf. Image Processing*, vol. 1, 2000, pp. 228–231.
[22] T. Vlachos, "Simple method for estimation of global motion parameters using sparse translational motion vector fields," *Electron. Lett.*, vol. 34, pp. 60–62, Jan. 1998.
[23] S. Erturk, "Image sequence stabilization: Motion vector integration (MVI) versus frame position smoothing (FPS)," in *Proc. 2nd Int. Symp. Image and Signal Processing and Analysis*, 2001, pp. 266–271.
[24] Q. Zheng and R. Chellappa, "Automatic feature point extraction and tracking in image sequences for unknown camera motion," in *Proc. 4th Int. Conf. Computer Vision*, May 11–14, 1993, pp. 335–339.
[25] Y. S. Yao, P. Burlina, R. Chellappa, and T. H. Wu, "Electronic image stabilization using multiple visual cues," in *Proc. Int. Conf. Image Processing*, vol. 1, Oct 23–26, 1995, pp. 191–194.
[26] S. Erturk, "Image sequence stabilization based on kalman filtering of frame positions," *Electron. Lett.*, vol. 37, Sep. 2001.
[27] Y. Sato, J. Chen, R. A. Zoroofi, N. Harada, S. Tamura, and T. Shiga, "Automatic extraction and measurement of leukocyte motion in microvessels using spatiotemporal image analysis," *IEEE Trans. Biomed. Eng.*, vol. 44, no. 4, pp. 225–236, Apr. 1997.
[28] K. E. Norman, "An effective and economical solution for digitizing and analyzing video recordings of the microcirculation," *Microcirculation*, vol. 8, pp. 243–249, 2001.
[29] A. R. Pries, "A versatile video image analysis system for microcirculatory research," *Int. J. Microcirc.*, vol. 7, pp. 327–435, 1988.
[30] S.-Y. Chien, S.-Y. Ma, and L.-G. Chen, "Efficient moving object segmentation algorithm using background registration technique," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 7, pp. 577–586, Jul. 2002.
[31] D. Marr and E. Hildreth, "Theory of edge detection," *Proc. Roy. Soc.*, vol. B207, pp. 187–217, 1980.
[32] D. Marr, *Vision*. San Francisco, CA: Feeman, 1982.
[33] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, pp. 321–331, 1987.
[34] C. Xu and J. L. Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. Image Process.*, vol. 7, pp. 359–369, 1998.
[35] N. Ray and S. T. Acton, "Tracking fast-rolling leukocytes *in vivo* with active contours," in *Proc. Int. Conf. Image Processing*, Rochester, NY, Sep. 2002.
[36] K. Ley, D. C. Bullard, M. L. Arbones, R. Bosse, D. Vestweber, T. F. Tedder, and A. L. Beaudet, "Sequential contribution of L- and P-selectin to leukocyte rolling *in vivo*," *J. Exp. Med.*, vol. 181, pp. 669–75, 1995.
[37] A. Hafezi-Moghadam *et al.*, "L-Selectin shedding regulates leukocyte recruitment," *J. Exp. Med.*, vol. 193, pp. 863–872, 2001.

Adam P. Goobic received the B.S. (with distinction) and M.S. degrees in electrical engineering from the University of Virginia, Charlottesville, in 2003.

He was commissioned a Second Lieutenant in the United States Air Force and is currently serving as an infrared applications developmental engineer at the Air Force Research Laboratory Sensors Directorate at Hanscom AFB, MA. His research interests include biomedical image analysis, image morphology, image segmentation, sensor fusion, and tracking.

Mr. Goobic is a member of Eta Kappa Nu.

Jinshan Tang (M'01–SM'03) received the Ph.D. degree from Beijing University of Posts and Telecommunications, P.R. China, in 1998.

From September 1998 to June 2000, he worked as an Invited Researcher in ATR Media Integration and Communications Research Laboratories (MIC), Kyoto, Japan. In June 2000, he went to Harvard Medical School to work on image enhancement technology. Since Dec. 2001, he has been working as a Research Scientist in Department of Electrical and Computer Engineering, University of Virginia. His research interests are medical image analysis, smart computers, content based image retrieval, active contour model, face detection and recognition, image enhancement for low-vision.

Scott T. Acton (S'89–M'93–SM'99) received the B.S. degree in Electrical Engineering from Virginia Tech, Blacksburg, in 1988 as a Virginia Scholar. He received the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Texas at Austin in 1990 and 1993, respectively, where he was a Microelectronics and Computer Development Fellow.

He has worked in industry for AT&T, Oakton, VA, the MITRE Corporation, McLean, VA, and Motorola, Inc., Phoenix, AZ, and in academia for Oklahoma State University, Stillwater. His research interests include anisotropic diffusion, active contours, biomedical segmentation problems, biomedical tracking problems and war.

For his research in video tracking, Dr. Acton was given an ARO Young Investigator Award. He received the Halliburton Outstanding Young Faculty Award in 1998. In 1997, he was named the Eta Kappa Nu Outstanding Young Electrical Engineer – a national award that has been given annually since 1936. At the University of Virginia, Charlottesville, he was named the Outstanding New Teacher in 2002, elected a Faculty Fellow in 2003, and holds the Walter N. Munster chair in Electrical and Computer Engineering and Biomedical Engineering. He is the recipient of a Whitaker Foundation Biomedical Engineering Research Grant for work in cell detection and tracking. He is an active participant in the IEEE, served as Associate Editor for the IEEE TRANSACTIONS ON IMAGE PROCESSING and as Associate Editor the IEEE SIGNAL PROCESSING LETTERS. He is the 2004 Technical Program Chair and the 2006 General Chair of the Asilomar Conference on Signals, Systems and Computers.