

Automatic *in vivo* Microscopy Video Mining for Leukocytes

Chengcui Zhang, Wei-Bang Chen, Lin Yang, Xin Chen, and John K. Johnstone

Department of Computer and Information Sciences, University of Alabama at Birmingham
115A Campbell Hall, 1300 University Boulevard, Birmingham, Alabama 35294, USA
{zhang, wbc0522, galabing, chenxin, jj}@cis.uab.edu

ABSTRACT

Biological videos are very different from conventional videos. Automatic spatiotemporal mining of moving cells from *in vivo* microscopy videos is extremely difficult because of the severe noises, camera/subject movements, deformations, and strong dependencies on microscopy operators. In this paper, we present an automatic spatiotemporal mining system of rolling and adherent leukocytes for intravital videos. The magnitude of leukocyte adhesion and decrease in rolling velocity are common interests in inflammation response studies. Currently, there is no existing system which is perfect for such purposes. Several approaches have been proposed for tracking leukocytes. However, these approaches can either only track leukocytes that roll along the centerline of the blood vessel, or can only handle leukocytes with fixed morphologies. In addition, the camera/subject movement is a severe problem which occurs frequently while analyzing *in vivo* microscopy videos. In this paper, we proposed a new method for automatic recognition of non-adherent and adherent leukocytes. The proposed method includes three steps: (1) camera/subject movement alignment; (2) moving leukocytes detection; (3) adherent leukocytes detection. The experimental results demonstrate the effectiveness of the proposed method.

Keywords

Temporal and spatial mining of biological videos, leukocytes

1. BACKGROUND INTRODUCTION

The *in vivo* microscopy technology enables biologists to create a natural environment and to monitor phenomena in live animal for studying biological problems. However, in contrast to mining the conventional surveillance video, mining *in vivo* microscopy videos is a more challenging since the respiration and movement of a living creature may cause unexpected camera/subject movements which change the spatial information of the monitored subject. Among mining various types of *in vivo* microscopy videos, the automatic detection of leukocytes is probably one of the toughest tasks since leukocytes are capable of amoeboid movements which may cause changes in their morphology.

In this paper, we focus on mining the *in vivo* microscopy video for leukocyte detection since the process of leukocyte migration is a very important phenomenon in the inflammatory tissue. It is well known in biologists' communities that leukocytes roll along vascular beds, arrest, and transmigrate before they are recruited to inflammatory sites and secondary lymphoid tissues during an inflammation response [9]. The magnitude of leukocyte adhesion and the decrease in rolling velocity are the main predictors of the inflammatory response. In [2] and [14], the measurement of leukocyte rolling and adhesion is done manually with a frame-to-frame video analysis. This type of manual data collection is time and labor consuming and subject to bias from observers. Automatic spatiotemporal mining of rolling and adherent leukocytes from intravital videos can significantly increase the

accuracy of the data collection and liberate biologists from the unnecessarily tedious analyses.

Several approaches have been proposed for tracking moving leukocytes from *in vivo* microscopy video sequences. In [5], local features such as color and temporal features are combined to develop a tracking system. It is reported to be capable of automatically tracking moving leukocytes. In order to perform motion correspondence between frames, they assume that all leukocytes roll along the vessel centerline. However, this assumption is not appropriate for all leukocytes activated during an inflammation response. Our video clips show that a significant amount of activated leukocytes roll along the vessels' boundaries. In another method of tracking moving leukocytes [1], after background removal, morphological filters are used to remove noises. However, with this method, the shape/size changes of leukocytes can pose a big challenge for selecting a fixed structure element for morphological operations. Furthermore, neither of these two papers deals with detecting adherent leukocytes at all.

2. OVERVIEW OF OUR APPROACH

In this paper, the closed cranial window model on rats was used to study the brain microcirculation. The scalp and the tissue from a 1.5×1.5 cm area bilaterally over the parietal cortex of rats were removed and a glass plate was glued to the surrounding bone to create a window. After recovery, animals were given 20 Gy radiation locally delivered to the brain. Ionizing radiation has been known to induce inflammatory responses in normal tissues including the central nervous system [2]. Prior to microscopy video recording, rhodamine 6G was injected through the tail vein to fluorescently label leukocytes in order to visualize the blood vessels. Leukocytes rolling and adhesion are thus visible in grayscale videos (see Figure 1).

Figure 1 shows typical frames of rolling and adherent leukocytes from *in vivo* grayscale video sequences. It is apparent that both moving and adherent leukocytes have relatively higher intensities compared with surrounding pixels. This observation prompts us to consider using the intensity threshold method to extract leukocytes in each frame. However, a simple global threshold method cannot extract all moving leukocytes because some of them tend to appear blurred in some frames as shown in Figure 1(c)-(d). Furthermore, as the separation of adherent and moving leukocytes is desired in most inflammation response studies, the presence of adherent leukocytes frequently introduces false correspondence in the temporal tracking of moving leukocytes. From previous studies in [1] and [5], the spatiotemporal feature between frames seems to be a good target for detecting moving leukocytes. Therefore, in our approach, we detect moving leukocytes first from spatiotemporal features and then use this information to facilitate the detection of adherent leukocytes with the intensity threshold method. By the end, we have two separate data sets on hands: moving and adherent leukocytes.

Our proposed method consists of three steps: (1) camera/subject movement alignment; (2) moving leukocytes detection; (3) adherent leukocytes detection.

The camera/subject movement must be dealt with prior to leukocyte tracking. We adopt a 2D homography based method for frame alignment. The proposed method begins with detecting camera/subject movements and then matches the selected frames to their predecessors by extracting point correspondences, which are used to compute a homography matrix that captures the transformation between two consecutive frames. Then, a robust estimator is adopted to identify the matched points that are geometrically consistent. These points are used to calculate the final transformation matrix to align the two frames [7].

In the second step, we propose and compare two approaches for detecting moving leukocytes. The first approach is based on probabilistic learning. This approach assumes that the pixels at the same location of consecutive frames are considered to follow a Gaussian distribution by themselves. Therefore, the probabilistic density can be estimated and compared with a global threshold. The second approach adopts neural network which learns the pattern of a background pixel (non-moving). We implement a two-layer neural network with the ability to handle temporal relationships by adopting a sliding window technique.

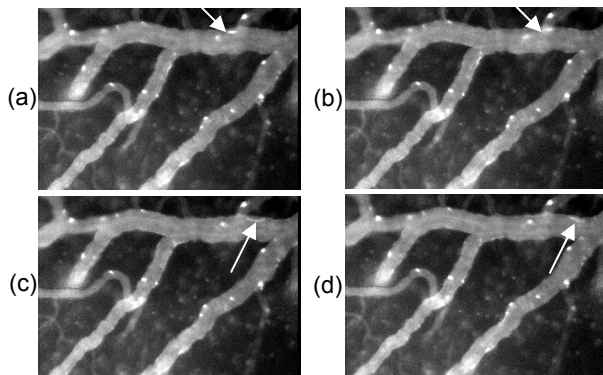


Figure 1. Four consecutive frames (a)-(d) from an *in vivo* microscopy grayscale video sequence. The bright white dots are mostly leukocytes and the tree-like structures are the blood vessels of study. The arrows mark one moving leukocyte over the four consecutive frames.

The third step is adherent leukocytes detection. We adopt an automatic thresholding method which approximates the optimal threshold value by using image histogram curve fitting. The real part of the second largest root of the curve derivative is close to the ideal threshold. We then implement a denoise method by inspecting the probability of a leukocyte pixel appearing at a specific location in consecutive frames. It is worth mentioning that the majority of both slow-rolling and firmly adherent leukocytes can be detected with this approach.

The contributions of this paper are in the following aspects:

- A homography based method is proposed to deal with camera/subject movement, which, to our best knowledge, has never been addressed before for microscopy video mining.
- A new algorithm of using probabilistic learning, filtering, and centroid trackers for automatic spatiotemporal mining of moving leukocytes from *in vivo* grayscale video sequences is proposed. Our algorithm has shown a reasonable recall rate and

a very low false positive rate (1%) and is therefore capable of correctly detecting the average velocity automatically.

- By detecting moving leukocytes, we are able to extract adherent leukocytes in a more robust way with an automatic thresholding method which approximates the optimal threshold value by using image histogram curve fitting. This adaptive method is able to find the optimal threshold for each frame respectively since the intensity level does change among frames in *in vivo* video microscopy. Our experiment results show a recall rate of 95% and a low false positive rate ($\approx 2\%$) as verified manually.

Section 3 describes the camera/subject movement alignment. Sections 4 and 5 present the detection methods for non-adherent and adherent leukocytes. Section 6 shows the experimental results. Section 7 concludes the paper.

3. VIDEO FRAME ALIGNMENT

The automatic mining of *in vivo* microscopy video is difficult. One of the reasons is the unexpected camera/subject movements, which would cause devastating noises in the spatial information of the moving cells. Thus, before any tracking tasks can be done, a preprocessing must be performed to make sure that all frames are aligned to the same coordinates. In this paper, we draw some ideas from the research of 2D projective transformations, namely, the 2D homography, as well as some recent developments in the area of image matching, and apply them to the problem of camera alignment. We embed it into a spatiotemporal mining system of rolling and adherent leukocytes for intravital videos.

The algorithm begins with a global scan of all N input frames to detect which part of the video contains camera movements. Once a frame is selected for alignment, it is matched with its predecessor to extract point correspondences, which are used to compute the homography, a 3×3 matrix that encapsulates the transformation between two images. Due to the possible errors caused by mismatches, the homography computed may not always reflect the real geometry. Thus, a robust estimator, RANSAC [7] is adopted to identify the matches that are geometrically consistent (called inliers). Only these inliers are used to calculate the final transformation matrix. Because both rotation and projective distortion between consecutive frames are small enough, our final transformation only consists of a translation matrix, which is calculated based on the average shift of all inlier matches.

3.1 Camera/Subject Movement Detection

Camera movement can be captured by measuring the difference between consecutive frames. Direct subtraction is probably the simplest way to do it. In our case, however, the foreground leukocytes shall be excluded since they are moving. Observing the different intensities of background tissues, vessels and leukocytes, we apply the Otsu's method [12] to turn all frames into binary images. After this, background tissues will be black while vessel and leukocyte areas are white (see Figure 2). Thus only the static part (tissues and vessels) will take effect in the subtraction.

Since each frame, if it needs alignment, will be aligned to their immediate predecessor, accumulating errors will be introduced if too many frames are selected for alignment. However, in a normal *in vivo* microscopy video, only a small portion of frames are observed to contain camera movements that are severe enough to affect the tracking performance. Thus, instead of using direct

subtraction, we adopt a simple similarity metric that allows a little tolerance in spatial variation:

$$S(f_1, f_2) = \sum_{(x,y)} \min_{|dx,dy| \leq r} |f_1(x+dx, y+dy) - f_2(x, y)| \quad (1)$$

where f_1 and f_2 are binary images of two consecutive frames and (x, y) are the x- and y-coordinates of a pixel. The search radius r is set to 1 in our case to account for any errors and tiny shifts.

After one scan over the N input frames, $N-1$ similarity values are computed for each consecutive pair. So is the standard deviation σ . If we assume that the movements of camera conform to the Gaussian distribution, frames with severe camera movements can be identified as the outliers of this distribution. That is, their similarity values S is larger than k times the variance (σ). The selection of k needs some training for different video contents. In our case, it achieves the optimal performance when k is set to 0.9.

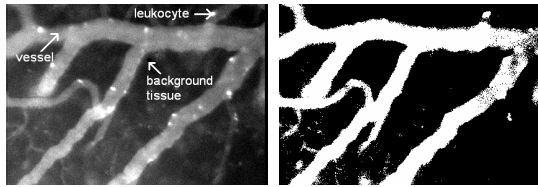


Figure 2. Left: the original video frame. Right: the binary image obtained by Otsu's method.

3.2 Frame Matching

Once a frame is selected, it is matched to its predecessor for point correspondences. Many such approaches exist in the area of image matching, and an evaluation of their performances can be found in [11]. In our work, we use SIFT [10] as our feature descriptor, since it is invariant to image translation, scaling, rotation, and partially invariant to illumination changes and affine projection.

SIFT [10] first locates feature points by detecting local extrema in a series of difference-of-Gaussian functions over all scales, and the sub-pixel accuracy is then achieved by fitting in a 2D quadratic and interpolating the location of the maximum. At each feature location, a dominant orientation is determined so that the features are invariant to image rotation. Finally, instead of recording the local pixel intensities, the SIFT feature vector is formed by accumulating local gradient values weighted by a Gaussian window, which provides certain invariance to affine transformations. For a typical frame, SIFT is able to generate around 2000 stable feature points.

After a database of feature vectors is generated, the match for each feature point is identified as its nearest neighbor in the feature space. Because of the large volume of feature points in the database and the high dimensionality (128) of SIFT feature vectors, brute force matching (based on exhaustive search) could be very inefficient. Regarding the problem of the nearest neighbor search in high dimensional space, Beis proposed a modified k-d tree structure, called "Best Bin First" or BBF [3]. It stores all the feature points in a k-d tree, and checks only a small portion of the leaf bins in the increasing order of Euclidean distance from the query point. It can efficiently locate the approximate nearest neighbor for each feature point. In the mean while, it may lose some correct matches.

Due to the relatively low intensity and the small frame size of our microscopy videos, much fewer feature points (around 200) are extracted from each frame. Therefore, the database of feature

points is very small and the speedup brought by BBF is not quite noticeable. In order to find as many matches as possible for the following stages, we decide to settle for the little time overhead of brute force matching, which is able to detect all correct matches.

3.3 2D Homography

After a set of matched points are obtained for two consecutive images, we can estimate the 2D projective transformation of them. This transformation is represented by a 3×3 matrix called the homography, which satisfies the following equation:

$$Hx_i = x_i' \quad (2)$$

where x_i and x_i' are matched points represented in the homogenous coordinates. Using the normalized direct linear transformation (normalized DLT) algorithm suggested by [11], H can be resolved based on 4 or more pairs of matched points. First the input coordinates are normalized with similarity transformations T and T' , each consisting of a translation and an isotropic scaling, such that the centroid of the points is the coordinate origin and their average distance from the origin is $\sqrt{2}$. Then the normalized coordinates are piled up into a $2n \times 9$ matrix (there are n matching pairs and each accounts for 2 rows), and the normalized homography \tilde{H} is given by solving for its right null space. This can be done efficiently using singular value decomposition. Finally, \tilde{H} is denormalized to get the homograph H for the original matched points by $H = T'^{-1}\tilde{H}T$.

3.4 Robust Estimation

Up to this point it has been assumed that the point matches we have do not contain any errors, but this is not always valid because points may be mismatched. These mismatched points (or outliers) may severely disturb the estimation of homography and thus must be identified before proceeding to the next stage.

We adopt the robust estimator, the RANdom SAMple Consensus (RANSAC) algorithm [7] to iteratively find the largest subset of geometrically consistent matches (inliers). It begins with randomly sampling 4 point matches out of the sample set (all matches), based on which a homography H' is estimated. Then H' is applied to the rest of the point matches to find the inliers based on the following equation:

$$|H'x_i - x_i'| \leq \varepsilon \quad (3)$$

The distance threshold ε is set to 0.5 in our case. This process is repeated a number of times to find the largest subset of inliers. In practice, however, it is not necessary to try every possible sample. In our work, 143 RANSAC iterations are performed for each pair of frames. If we assume the probability that any correspondence is valid to be 0.5, then the likelihood that all RANSAC trials will all fail is $(1-0.5^4)^{143} < 0.01\%$, which is good enough for our purposes.

After the largest subset of inliers is selected, we can now re-run the algorithm described in Section 3.3. However, this time we only use all the inliers to compute the homography. An example result of the matched points can be found in Figure 3.

3.5 Frame Alignment

Given the homography of two images, we can easily align one to the other based on Equation (2). However, there is a disadvantage doing it that way: what homography describes is the projective transformation between two views, which also includes the change of perspectives. In most microscopy videos, this change is negligible, especially between consecutive frames. On the other hand, if we do consider projective distortion, the accumulated

error will grow significantly as subsequent images are aligned and the total number of frames that can be aligned is thus very much restricted. Also note that the camera/subject movement between consecutive frames can be well described by a “shift” (without significant rotation or zooming effect). Taking these into account, our final transformation matrix only consists of a translation T .

Since we already have geometrically consistent point matches, we use their average shift in x- and y-axis to calculate T , and the bilinear interpolation is used in the pixel transformation.

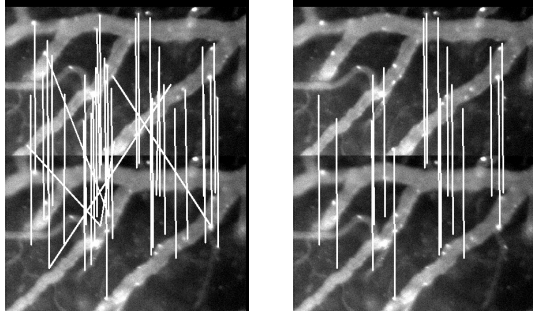


Figure 3. Left: matched points before RANSAC. Right: matched points after running RANSAC. Only those geometrically consistent matches are preserved.

4. MOVING LEUKOCYTES DETECTION

The optical flow techniques and background subtraction are popular approaches for tracking moving objects. The basis of optical flow is to estimate the optical flow at all points of a frame and significant points are grouped based on principles of motion coherence [15]. However, since there are also erythrocytes and other blood cells circulating inside and outside the targeted vessels, the optical flow technique can lead to a high false positive rate and make it inappropriate for tracking moving leukocytes from microcirculation video. Another drawback of the optical flow is its sensitivity to contrast changes.

Background subtraction comes to everyone’s mind naturally since ideally we can consider that all have steady backgrounds and the rolling leukocytes are the only moving parts. The simplest background subtraction is to calculate an average image of all frames and then subtract each frame from this average, and finally threshold the result. Yet, the performance of tracking cells from *in vivo* videos is affected by many factors due to poor video quality such as noise and clutter, cell deformation, and contrast changes. Hence, the simplest form of background subtraction is not good enough and needs further improvement.

In this paper, we propose and compare two approaches for background subtraction and detecting moving leukocytes, including a probabilistic learning approach and a neural network approach. Both methods are followed by a median and location-based filtering to reduce noise. After locating moving leukocytes for each frame, we perform motion correspondence with centroid trackers and compute the average rolling velocity.

4.1 Probabilistic Learning Approach

Given $x_{1j}, x_{2j}, \dots, x_{Nj}$ be the grayscale intensity values (0 to 255) of a pixel at the location j ($1 \leq j \leq \text{Total number of pixels in a frame}$) over N consecutive frames. The probability density function that this pixel will have intensity value x_{ij} in the frame t can be non-parametrically estimated with the kernel estimator K [6]. We choose the kernel estimator function K to be a normal distribution,

which means that the pixels at the same location of the N frames are considered to follow a Gaussian distribution by themselves. Therefore, the probabilistic density can be estimated with

$$P(x_{ij}) = 1/N * \sum_{i=1}^N (1/\sqrt{2\pi\sigma_j^2}) * e^{-0.5*(x_{ij}-x_{ij})^2/\sigma_j^2} \quad (4)$$

σ_j is the temporal invariance of intensity I for a pixel at the location j over N frames and is calculated as

$$\sigma_j = 1/(N-1) * \sum_{i=1}^N (I_i - \bar{I})^2 \quad (5)$$

Using this probability estimate, a pixel at the location j in the frame t is considered to be a foreground pixel if $P(x_{ij}) < th$ where th is a global threshold over all frames that can be adjusted to achieve a desired percentage of false positives. With this simple model, a foreground pixel is a part of moving leukocytes in each frame since adherent leukocytes and blood vessels will be identified as background by this model.

Simple as it seems, it is not easy to determine the threshold value th . This requires expert knowledge or many trial and error experimentations. Another problem of this method is that it is only effective on linearly separable datasets. In view of these problems of the thresholding method, we choose to use One-class SVM [13] to partition the pixels into two classes (i.e., foreground and background). The SVM approach is more desirable because it discards the notion that there exists a straight line separating foreground pixels from background pixels. It is safer to assume that the probability values are not linearly separable. During experiments, it was also found that a one-class SVM approach was better than the binary classification.

One-Class classification is a kind of unsupervised learning mechanism. It tries to assess whether a test point is likely to belong to the distribution underlying the training data. The objective of One-Class SVM is to create a binary-valued function that is positive in those regions where the data predominantly lies and negative elsewhere. More details of Schölkopf’s One-Class SVM can be found in [13].

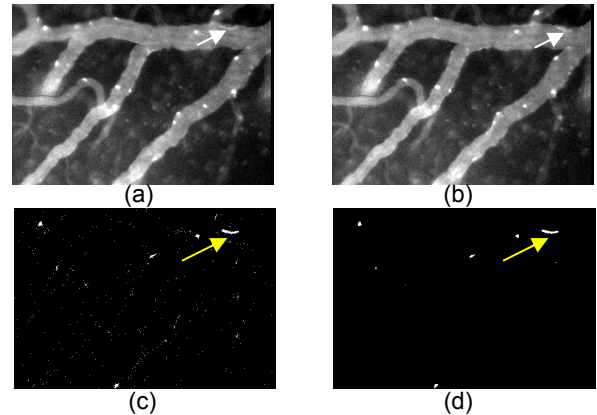


Figure 4. Extract moving leukocytes via probabilistic learning. (a) The current frame; (b) the next frame from the same video sequence; (c) the moving leukocytes extracted through probabilistic learning; (d) the result of (c) after noise removal. The arrows point to a moving leukocyte in (a) and (b) which is being tracked and becomes evident in (c) and (d).

The probabilities (based on Equation (4)) of a small set of pixels which are part of some moving leukocytes are used to train a one-class SVM, and the trained model is used to classify pixels into

foreground (moving leukocytes) and background. Based on our experiments, the probabilistic learning from temporal features, when coupled with one-class SVM, is able to identify all moving leukocytes from each frame. An example of extracted moving elements is shown in Figure 4. The arrows in Figures 4(a) and (b) point to the appearance of a moving leukocyte in two consecutive frames. In Figure 4(c), the extracted moving elements are depicted as white dots. As expected, the extracted moving elements include not only moving leukocytes, but also noises.

4.2 Neural Network Approach

Moving and adhering leukocytes co-exist in blood cells. As adhering leukocytes are static or rolling very slowly, they can be regarded as part of the background. In order to detect moving leukocytes, we only need to single out the foreground of a frame.

Each pixel can be represented by a sequence of intensity values extracted from each frame, which is referred to as the intensity sequence in this paper. Intensity sequence of a pixel is time series data. The foreground pixels' intensity sequences are different from those of background pixels in that the former vary dramatically along time while the latter only show small variations. That is to say, the background pixels' intensity sequences demonstrate stable and consistent patterns while the foreground pixels' intensity sequences are comparatively more unpredictable. Therefore, we can analyze the predictable background pixel patterns and hence detect the foreground pixels (i.e., moving leukocytes) by identifying the intensity sequences that do not follow the learned background pattern. Classical linear autoregressive models in modeling time series data are rather limited, since they assume linear relationship among consecutive data series. As illustrated in [4], Multi-Layer Perceptron (MLP) and Radial Basis Function (RBF) offer an extension to the linear model by using a non-linear function, which can be estimated by such learning and optimization techniques as back propagation and conjugate gradient. Therefore, we choose to use the temporal model of a neural network for time series data prediction. The neural network has two layers with sigmoid transfer in the first layer and linear transfer in the second layer.

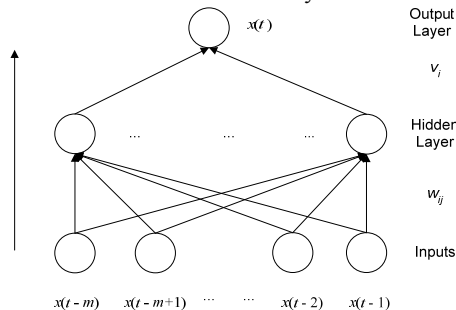


Figure 5. Temporal Model of a Neural Network

The neural network is supposed to learn the consistent pattern of a background pixel. As can be seen from Figure 5, the input nodes of the neural network have temporal relationships. These input nodes are sub-sequences extracted from the intensity sequences by a sliding window technique. For each pixel, its intensity sequence can be obtained by a window of size m one step a time. Each time, a sub-sequence of size m can be extracted. With the intensity sequence length being n , the total number of sub-sequences is $n-m+1$. The intensity of the pixel at time t can be predicted by the sub-sequence of size m preceding it. In this way, we can predict the intensity value of a pixel at each time point

(i.e., at each frame). There are certain parts of the blood cell that we know for sure are background. We choose the intensity sequence S of a background pixel as the training data, where $S = [s_1, \dots, s_t, \dots, s_{n-m+1}]$ and s_t is a subsequence $\{x(t-m), \dots, x(t-1), x(t)\}$. $x(t)$ is the output of the network, given $\{x(t-m), \dots, x(t-1)\}$ as the input. The neural network is trained to learn the pattern of the background pixel. In testing, the trained neural network assumes each input sub-sequence follows this pattern and predict the pixel in the next frame according to this pattern. We test the trained neural network on intensity sequences of all the pixels and generate their prediction values on each frame.

In each frame, the pixel's real intensity is compared to its predicted intensity. A big difference between these two signifies that the preceding sub-sequence does not match the background pattern, and the pixel in the current frame is foreground. Otherwise it is background. A threshold is set up to differentiate foreground and background.

4.3 Noise Removal

As shown in Figure 4, the data obtained from probability learning is not perfect and we can see a lot of noise signals in Figure 4(c). Further enhancement is necessary to differentiate noise from real signals. As mentioned earlier, one of the problems faced in vivo microcirculation videos is fluids flowing outside the vessels. Leukocytes moving outside the target vessels are certainly not of our interests. We can remove moving objects outside vessels with a location-based filtering. This can be achieved with a binary image of vessel segmentation. A simple way of vessel segmentation is to create a binary image of each frame using a global threshold method based on Otsu's thresholding algorithm [12]. Figure 2 shows such a binary image example. It is worth mentioning that each frame is associated with its own vessel binary image which might be slightly different from the others. By doing this, the slight changes in vessel shapes due to animal respiration or other reasons can be accommodated.

Based on our experiments, the vessel segmentation is effective in filtering out moving objects outside vessels. However, even after we remove the objects outside the vessels, there are still random noises which are relatively small and mostly isolated. Thus, the median filtering is a good choice to remove them without losing real signals. With the median filtering, the value of an output pixel is determined by the median of the neighborhood pixels. Figure 4(d) shows an example of cleaned result after noise removal.

4.4 Motion Correspondence and Velocity Computation

As aforementioned, the velocity of moving leukocytes is a qualitative measurement of inflammation responses. Current methods of tracking cells from *in vivo* videos include correlation trackers [1] [16] and centroid trackers [17]. Correlation trackers use a fixed template for target cells and correlate it with the images to trace the target cells. This type of correlation is inappropriate for tracing moving leukocytes during inflammation responses because activated leukocytes tend to change their shapes a lot. We indeed observe this from the video sequence used in this paper. In contrast, centroid trackers are able to trace deformable cells by following their intensity mass center positions over frames. Therefore, we are going to use centroid trackers to trace moving leukocytes in this paper.

The binary image obtained after noise removal (Figure 4(d) is an example) for each frame actually contains spatial information about moving leukocytes in each frame. Then a seeding and growing approach is used to extract each group of spatially connected leukocytes, and finally the centroid of each of such group can be located. After all centroid positions are located, we apply our matching algorithm to each frame and its previous frame. The matching algorithm is designed as below:

- Start with a centroid in the current frame, find the centroid in the previous frame with the smallest distance;
- If the smallest distance is within a pre-defined limit, we consider this pair to be a match;
- Remove the matched pairs from the previous and the current frames and repeat the first step until all the centroid positions in the current frame are tried.

In our experiments, the limit is set at 30 pixels, which is determined by the camera calibration parameters and the maximal velocity we can expect. Our results show a very low false positive rate and a reasonable recall rate when verified manually, which proves that the chosen limit is reasonable for our case.

After acquiring all matched centroid positions between frames, we can compute the mean velocity V_e by averaging the Euclidean distances between matched pairs. We get an average velocity of 6.218 pixels per frame in our experiments. It is easy to convert it to a real-world velocity measurement V_c by the following equation when a calibration of *in vivo* microscopy c (length per pixel) and the time elapse between frames t are available:

$$V_c = V_e * c / t \quad (6)$$

5. ADHERENT LEUKOCYTE DETECTION

From Figure 1, it can be seen that leukocytes tend to stand out from their surrounding areas in each frame. This suggests that local features (e.g., local intensity distribution) in each frame might be sufficient for extracting leukocytes. The local range feature, where each pixel is represented by its range value (maximum_value–minimum_value) of its neighborhood pixels, is tried first. The choice of the neighborhood turns out to be very tricky. A circular shape neighborhood is reported to be a good choice by Acton et al. for tracking leukocytes [1], but we find it very difficult for our case. Even though normal leukocytes have a disk shape, leukocytes in inflammation responses tend to change their shapes a lot and the shapes can be quite irregular. We finally find out that the traditional 3-by-3 rectangle neighborhood is actually the best choice. The local range filtering indeed captures all moving and adherent leukocytes in the frame. However, the resulting local range image also outlines the vessel boundaries.

The single global intensity threshold is another method which might be used to extract leukocytes of our interest. Yet, the intensity change between frames from *in vivo* microscopy video sequences could not justify a single intensity threshold for all frames. Therefore, we design an adaptive algorithm to select an intensity threshold for each frame, respectively.

To detect adherent leukocytes, we observe that each frame mainly consists of three types of regions: tissues, blood vessels, and leukocytes. These three regions are very different in their pixel intensity values. The intensity values of tissue pixels are often on the low side, and the pixel intensities of blood vessel regions are generally in the middle range, while the pixel intensities of leukocytes are typically on the high end. This is exemplified in

Figure 6, where the pixel intensity histograms of the above three types of regions overlap with each other in a frame.

While the detection of tissue regions is relatively easy, problems remain in finding a threshold that can best separate leukocytes regions from blood vessel regions. As shown in Figure 6, the ideal threshold is around the intersection of the two curves representing the regions of leukocytes and blood vessels, respectively. However, finding the ideal threshold is difficult since the above two curves are unknown.

To approximate the ideal threshold value, our first step is to find the best fitting curve for a given image histogram. We tried several curve fitting methods, and found that the 8th degree polynomial which is shown as the red curve in Figure 6 is the best fit. We then calculate the first derivative and the roots of the 8th degree polynomial equation. We found that the real part of the second largest root is about the ideal threshold value. Though the majority of leukocytes can be detected with this threshold, some low-intensity leukocyte pixels are excluded and some high-intensity blood vessel pixels are falsely included. In the latter, those pixels are present in the form of noise as shown in Figure 7.

To eliminate noise, we observe that the rolling velocity of adherent leukocytes is relatively slow compared with the exposure time (30fps). This implies an adherent leukocyte should stay at almost the same location in adjacent frames. On the contrary, noise pixels often appear randomly in a frame. Therefore, we may determine whether a pixel belongs to leukocytes or noise based on the frequency of a pixel being recognized as a candidate leukocyte pixel at the same location in consecutive frames.

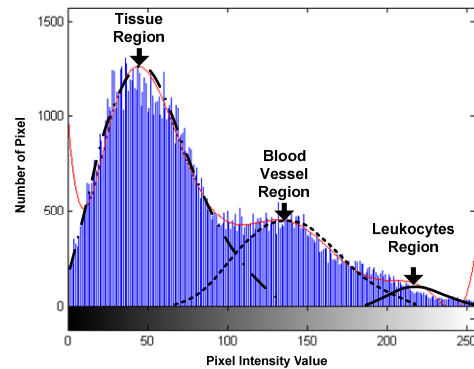


Figure 6. A sample intensity distribution of tissues, blood vessels, and leukocyte regions in an image histogram.

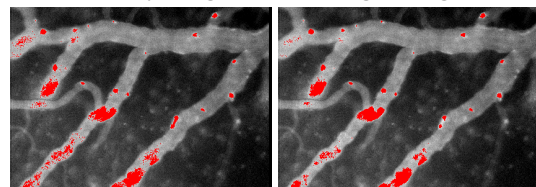


Figure 7. Left: The candidate adherent leukocytes (marked with red) before denoising. Right: The result after denoising.

In particular, to determine whether a pixel (x, y) in frame f_k is a leukocyte pixel or noise, we first obtain the adjacent frames around f_k -- $f_i \dots f_k \dots f_j$, where $j-i = m$ (m is an odd number), $i=k - \lfloor m/2 \rfloor$ and $j=k + \lfloor m/2 \rfloor$. Assume there are n pixels locate at (x, y) in adjacent frames $f_i \dots f_j$ being recognized as candidate leukocyte pixels by the thresholding method in the first step, a probability $p = n/m$ is calculated for this pixel. If the probability is greater than

0.5, this pixel is considered to be a leukocyte pixel. Otherwise, this pixel is considered to be a noise pixel. We tested on different values of m and found that $m \approx 5$ performs the best for the given testing video. Based on this, a median filter is further applied to eliminate those relatively small and isolated noise pixels. A sample result of this denoise process is demonstrated in Figure 7.

It is worth noting that this adaptive thresholding method is applied after the moving leukocytes are removed (see Section 4). This way, we observe much less false positives.

6. RESULTS AND DISCUSSIONS

Our testing video sequence comes from an in vivo video microscopy during a typical inflammation response where activated leukocytes roll and adhere to the vessels. Our strategy is to detect moving leukocytes first and then utilizes the information of moving leukocytes to detect adherent leukocytes. The testing video consists of 148 frames with several places of camera/subject movements. As the preprocessing step, the entire video sequence is first aligned by the 2D homography based method.

Our video mining of moving leukocytes include three steps – locating moving leukocytes through the spatiotemporal probabilistic learning, noise removal with median filtering and location-based filtering, and centroid trackers for tracing moving leukocytes over the frames and computing the average rolling velocity. As aforementioned, there are several approaches locating moving leukocytes and computing rolling velocity from in vivo microscopy videos based on background subtraction with temporal features [1] [5]. In [1], the background subtraction is achieved by subtracting the average of all frames from each frame. In [5], a temporal invariance image is computed for each frame and a single threshold is then selected to remove the background. We test both methods with our test video and both yield poor performances. This is largely due to severe noises, cell deformations and background movements. In order to overcome the problems, we introduce the use of spatiotemporal probabilistic learning to extract moving leukocytes as foreground pixels and the performance is satisfactory. This demonstrates that a probability threshold is more appropriate than a low level intensity threshold, especially for those video sequences full of severe noises and contrast changes. However, the background subtraction alone is likely to suffer from excessive noise. Further enhancements following background subtraction are necessary to overcome this problem. In [5], a location-based filtering is used to remove false positives outside a vessel region. Acton et al. [1] approach this problem by using morphological filters. However, none of them deal with the detection and separation of adherent leukocytes from moving leukocytes. In addition, the morphological filtering method performs poorly on our video sequence. This is probably due to the irregular shapes of leukocytes and their deformations. In any morphological operation, a structuring element has to be defined to specify the neighborhood of the pixel of interest. It is a matrix consisting of only 0's and 1's that can have any arbitrary shape and size. In [1], they use a structuring element with a disk shape because ideal leukocytes have disk shapes. However, this is not true in our video sequence and it is easy to spot that leukocytes come in all shapes and sizes. After location-based filtering, we also apply median filtering on our data to further remove those isolated random noises. After locating moving leukocytes in each frame, we continue our with centroid trackers to obtain motion

correspondence and compute the rolling velocity. The rolling velocity is the main interest of moving leukocytes for many inflammation response studies because it offers biologists a qualitative measurement of each inflammation response. In [5], it is assumed that leukocytes roll along the vessel centerline and therefore a vessel centerline extracted by thinning can help in predicting the direction of moving leukocytes. However, this assumption is against our observations that under inflammation response, leukocytes rarely roll along the centerline and move along the vessel boundaries instead. Acton et al. 2002 use correlation trackers to match a template of the target leukocyte to each frame in order to find the instances of the target. This method is insufficient and very vulnerable to leukocytes' deformations. We decide to apply the centroid trackers in our motion correspondence because it is less vulnerable to the frequent shape changes of leukocytes. We match leukocytes between frames by finding the pairs with the minimal distance which is below a selected limit.

Our experiments on detecting moving leukocytes show a false positive rate as low as 1% when double-checked manually, while the neural network based approach has a false positive rate of nearly 49%. The rolling velocity is 6.218 pixels per frame, which is verified by manually. It is easy to convert it to a conventional speed representation by Equation 6. However, we observe a relatively low recall rate ($\approx 50\%$) for tracking moving leukocytes since some leukocytes tend to disappear in some frames and reappear afterwards. Since the measurement of the average rolling velocity is the main interest, this is justified as long as enough leukocytes are tracked with a very low false positive rate. More discussion about this issue is presented at the end of this section.

There are no studies about automatic detection of adherent leukocytes that can be found in the literature. This is probably due to the emphasis on leukocyte rolling in most inflammation responses studies. However, the magnitude of adherent leukocytes is certainly a way to predict the strength of one inflammation response. In this paper, we are able to extract adherent leukocytes by combining an adaptive global intensity threshold method with the information about the moving leukocytes detected earlier. Our experimental results show a recall rate of 95% with a low false positive rate ($\approx 2\%$) when double-checked manually.

It is observed from our video sequence that moving leukocytes occasionally disappear in one frame and reappear again in the next frame. We are concerned that our motion correspondence algorithm may be vulnerable to the errors caused by this phenomenon and draw a potential criticism. In [11], Ghosh and Webb address the similar concern in their approach to automatically detect cell receptors. They propose including a storage matrix for the unmatched pixels in the previous frame against the current frame. Then when the matching for the next frame starts, a scan through the storage matrix of the previous frame is also performed to find whether there is any reappearance. We test this algorithm on our video sequence. However, our results show the false positive rate of this algorithm is as high as 61% when double-checked manually. By including those false leukocytes into our measurement, we are also introducing more significant errors into our velocity measurement. Therefore, we argue that in this case, precision is more important than recall, and that omitting those disappearing and reappearing leukocytes is okay for the measurement of the average rolling velocity as long as enough leukocytes are traced. Another potential criticism is the

inclusion of one empirical parameter in learning motion correspondence. We use a distance threshold for leukocytes matching. This concern will be addressed in our future work when we include relevance feedback in our algorithm to automatically adjust those two parameters.

7. CONCLUSIONS

In this paper, we present an automatic intravital video mining system of leukocytes rolling and adhesion. Video mining of in vivo microscopy video sequences is very difficult due to severe noises, background movements, leukocytes' deformations, and contrast changes. Currently, there are a few approaches attempting to automatically track rolling leukocytes in the literature but none of them suit our needs of separating moving leukocytes from adherent ones. In our approach, we first align video frames to eliminate noise caused by camera movement. We then locate the moving leukocytes by applying and comparing a spatiotemporal probabilistic learning method and a neural network framework for time series data. We further remove noises by applying median and location-based filtering. Another contribution from this paper is that we extract the information about moving leukocytes first and therefore are able to extract adherent leukocytes with an image histogram curve fitting method.

8. REFERENCES

- [1] Acton, S.T., Wethmar, K., and Ley, K. Automatic tracking of rolling leukocytes in vivo. *Microvasc Res.*, 63 (2002), 139-148.
- [2] Anderson, S.I., Shiner, R., Brown, M.D., and Hudlicka, O. Icam-1 expression and leukocyte behavior in the microcirculation of chronically ischemic rat skeletal muscles. *Microvasc Res.*, 71 (2006), 205-211.
- [3] Beis, J., and Lowe, D.G. Shape indexing using approximate nearest-neighbor search in high-dimensional spaces, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1997.
- [4] Dorffner, G. Neural network for time series processing. *Neural Network World*, 6, 4 (1996), 447-468.
- [5] Eden, E., Waisman, D., Rudzsky, M., Bitterman, H., Brod, V., and Rivlin, E. An automated method for analysis of flow characteristics of circulating particles from in vivo video microscopy. *IEEE Transactions on Med Imaging*, 24 (2005), 1011-1024.
- [6] Elgammal, A.M., Harwood, D., and Davis, L.S. Non-parametric model for background subtraction. In *Proceedings of the 6th European Conference on Computer Vision-Part II*, (2000). Springer-Verlag.
- [7] Fischler, M.A., and Bolles, R.C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24 (1981), 381-395.
- [8] Hartley, R., and Zisserman, A., *Multiple view geometry in computer vision*. Cambridge University Press, 2003.
- [9] Lawrence, M.B., and Springer, T.A. Leukocytes roll on a selection at physiologic flow rates: Distinction from and prerequisites for adhesion through integrins. *Cell*, 65 (1991), 859-873.
- [10] Lowe, D.G. Object recognition from local scale-invariant features, in *Proceedings of the International Conference on Computer Vision*, 1999.
- [11] Mikolajczyk, K. and Schmid, C. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27 (2005), 1615-1630.
- [12] Otsu, N. A threshold selection method from gray-level histogram. *IEEE Transactions Systems, Man, Cybernetics*, 9 (1979), 62-66.
- [13] Schölkopf, B. and Platt, J.C. Estimating the support of a high-dimensional distribution. Microsoft Research Corporation Technical Report MSR-TR-99-87 (1999).
- [14] Yuan, H., Goetz, D.J., Gaber, M.W., Issekutz, A.C., Merchant, T.E., and Kiani, M.F. Radiation-induced up-regulation of adhesion molecules in brain microvasculature and their modulation by dexamethasone. *Radiat Res.*, 163 (2005), 544-551.
- [15] Mclvor, A., Zang, Q., and Klette, R. The background subtraction problem for video surveillance systems. *Lecture Notes in Computer Science*, Springer Berlin / Heidelberg, International Workshop RobVis, (Auckland, New Zealand, 2001), 176-188.
- [16] Schutz, G.J., Schindler, H., and Schmidt, Th. Single molecule microscopy on model membranes reveals anomalous diffusion. *Biophysical Journal*, 73, (1997), 1073-1080.
- [17] Ghosh, R.N., and Webb, W.W. Automated detection and tracking of individual and clustered cell surface low density lipoprotein receptor molecules. *Biophysical Journal*, 66, (1994), 1301-1318.

About the authors:

Chengcui Zhang is an Assistant Professor of Computer and Information Sciences at University of Alabama at Birmingham (UAB) since August, 2004. She received her Ph.D. from the School of Computer Science at Florida International University, Miami, FL, USA in 2004. She also received her bachelor and master degrees in Computer Science from Zhejiang University in China. Her research interests include multimedia databases, multimedia data mining, image and video database retrieval, bioinformatics, and GIS data filtering. She is the recipient of several awards, including the IBM Unstructured Information Management Architecture (UIMA) Innovation Award, UAB ADVANCE Junior Faculty Research Award from the National Science Foundation, and UAB Faculty Development Award.

Wei-Bang Chen is a Ph.D. candidate in the Computer and Information Sciences Department at UAB. He received a Master's degree in Genetics from National Yang-Ming University in Taipei, Taiwan and a Master's degree in Computer Sciences from UAB. His main research area is bioinformatics. His current research involves microarray image and data analysis, biological sequence clustering, and biomedical video and image mining.

Lin Yang is a Ph.D. student in the Computer and Information Sciences Department at UAB. His research interests lie at the intersection of computer vision and graphics, which include multiple view geometry, visual surveillance system, and multimedia data mining. He received the BS degree in computer science from Fudan University, China in 2006.

Xin Chen received her Master's degree in Computer Science from University of Science and Technology Beijing, China, in 2002. From 2004 to present, she has been pursuing her Ph.D. degree in the Computer and Information Sciences Department at UAB. Her research interests include Content-based Image Retrieval, multimedia data mining, and spatiotemporal data mining.

John K. Johnstone is an Associate Professor of Computer and Information Sciences at UAB. His research is in geometric modeling, with recent interest in structure from motion. He received his Ph.D. from Cornell University and has also been on the faculty at Johns Hopkins University.