Detection of Leukocytes from In vivo Videos

Chengcui Zhang, Wei-Bang Chen, Lin Yang, and Xin Chen
Department of Computer and Information Sciences, University of Alabama at Birmingham, USA

Abstract

Leukocyte migration is an important phenomenon in the inflammatory tissue. The migration process includes the rolling velocity decreasing and the leukocytes adhesion. However, the analysis of in vivo microscopy video is a labor-intensive and time consuming task. Several approaches have been proposed for tracking leukocyte movements. 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 experimental results demonstrate the effectiveness of the proposed method.

1. 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 videos, mining in vivo microscopy videos is a more challenging task 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. The process of migration includes the rolling velocity decreasing and the leukocyte adhesion [1]. In the past, leukocytes detection in vivo microscopy video is a labor intensive task since leukocytes must be manually selected from each frame [2, 3]. Several approaches have been proposed for tracking leukocyte movements and can be categorized as two groups: 1) those that can only track those leukocytes rolling along the centerline of the blood vessel [4], or 2) those that can only handle leukocytes with fixed morphologies [5]. These approaches are not sufficient because leukocytes may roll along the endothelial lining of blood vessels and they may change their morphology due to amoeboid movements.

In this paper, we proposed an effective method for detecting non-adherent and adherent leukocytes from in vivo microscopy videos. This 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 any tracking of leukocytes in blood vessels. For that, we adopt a 2D homography based method to align the video frames in the entire video sequence.

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. The second approach adopts neural network which learns the pattern of a background pixel. We implement a two-layer neural network capable of handling temporal relationships.

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. 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.

Section 2 describes the details of the camera/subject movement alignment. Sections 3 and 4 present the detection methods for non-adherent and adherent leukocytes, respectively. Section 5 shows the experimental results. Section 6 concludes this paper.

2. Video frame alignment

2.1. Camera/subject movement detection

Camera/subject movements can be captured by measuring the difference between consecutive frames, and maybe the simplest way to do it is the direct subtraction. In our case, however, foreground leukocytes should be excluded from this operation since they are known to be moving through frames. By observing the different intensities of background tissues, vessels and leukocytes, we apply the Otsu's method [7] to turn all frames into binary images. After this operation, background tissues will appear in black

0-7695-3006-0/07 $25.00 © 2007 IEEE
DOI 10.1109/IPC.2007.19

while all vessel and leukocyte areas are white as shown in Figure 1, and the subsequent subtractions are applied to the binary images only.

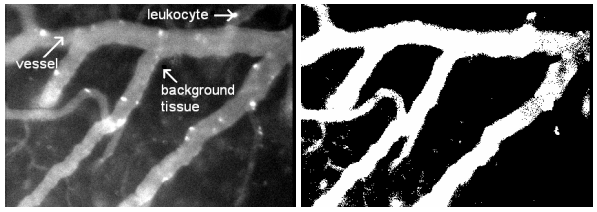

Figure 1. Left: the original video frame. Right: the binary image obtained by Otsu's method.

Since frames (if aligned) are aligned to their immediate predecessor, the accumulated errors will be significant if too many frames are selected for alignment. However, in a typical vivo microscopy video, only a small portion of frames are observed to contain camera/subject movements that are significant 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 variations:

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

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

After one scan over the N input frames, N-1 similarity values are computed for each consecutive pair, and so is the standard deviation σ. If we assume that the movements of camera/subject conform to the Gaussian distribution, those frames with significant 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 is optimal when k is set to 0.9.

2.2. Frame matching

Once a frame is selected for alignment, it is matched to its predecessor for point correspondences. In this study, we use SIFT as our feature descriptor, since it is invariant to image translation, scaling, rotation, and partially invariant to illumination changes and affine projection.

SIFT [8] first locates feature points by detecting local extrema in a series of difference-of-Gaussian functions over all scales, and sub-pixel accuracy is then achieved by fitting in a 2D quadratic and interpolating the location of the maximum. 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. Due to 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 a high dimensional space, Beis proposed a modified k-d tree structure, called "Best Bin First" or BBF [9]. It stores all the feature points in a k-d tree, and checks only a small portion of leaf bins in the increasing order of Euclidean distance from the query point. In our case, due to the low pixel intensity and the small frame size of microscopy videos, much fewer feature points (around 200) are extracted from each frame, so 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 subsequent processing, we decide to settle for the little time overhead of the brute force matching, which on the other hand, is able to detect all correct matches.

2.3. 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 [10], 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 $H = T'^{-1}\tilde{H}T$.

2.4. Robust estimation and frame alignment

Up to this point it has been assumed that the matched points do not contain any errors, but this is not always true 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 [6], to iteratively find the largest subset of geometrically consistent matches (inliers). After the largest subset of inliers is selected, we can now re-run the algorithm described in Section 2.3, but this time only using all the inliers to compute the homography.

Given the homography of two images, we can easily align one to the other based on Equation (2). However, in most microscopy videos, the change of perspectives is negligible, especially between consecutive frames. On the other hand, if we do consider projective distortion, then 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/zooming). Taking these into account, our final transformation matrix only consists of a translation T.

3. Moving leukocytes detection

In this paper, we propose and compare two approaches for detecting moving leukocytes, including the probabilistic learning and the neural network. Both are followed by a median filtering to reduce noise.

3.1. Probabilistic learning based approach

Given x_{1j}, x_{2j}, … x_{Nj} the grayscale intensity values (0 to 255) of a pixel at the location j ($1 \le j \le$ Total number of pixels in a frame) over N consecutive frames. The probability density function that this pixel will have intensity value x_{tj} in Frame t can be non-parametrically estimated with the kernel estimator K [11]. We choose the kernel estimator function K to be a normal distribution, meaning 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

$$P(x_{tj}) = 1/N \times \sum_{i=1}^{N} (1/\sqrt{2\pi\sigma_j^{\,2}}) \times e^{-0.5\times(x_{tj}-x_{ij})^2/\sigma_j^{\,2}} \quad (3)$$

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

$$\sigma_j = 1/(N-1) \times \sum_{i=1}^{N} (I_i - \bar{I})^2 \quad (4)$$

Using this probability estimate, a pixel at the location j in Frame t is considered to be a foreground pixel if $P(x_{tj}) < th$ where th is a global threshold and is automatically determined by a simple learning process. In particular, the probability values of a small set of pixels which are part of some moving leukocytes are used to train a one-class SVM with Gaussian radial basis kernel, and the trained model is then used to classify pixels into foreground (moving leukocytes) and background according to their probability values.

3.2. Neural network based 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. 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. The 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. 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.

In this study, a neural network is built to learn the pattern of a background pixel. The input nodes of the neural network have temporal relationships. These input nodes are sub-sequences extracted from the intensity sequences by the sliding window technique. From the perspective of one pixel, its intensity sequence can be scanned by a window of size m one step at a time. Each time, a sub-sequence of size m can be extracted. 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 unit (i.e., at each frame).

In each frame, a pixel's real intensity is compared with 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 pixels.

4. Adherent leukocytes detection

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 tissues 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.

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. Typically, the ideal threshold is around the

intersection of the two histogram curves representing the intensity distributions 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 a best fitting curve for a given image histogram. We tried several curve fitting methods, and found that the 8^{th} degree polynomial is the best fit. We then calculate the first derivative and the roots of the 8^{th} 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 case, those pixels are present in the form of noise.

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 the 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. In particular, a sliding window based approach is implemented in this study to achieve such a goal. Based on this, a median filter is further applied to eliminate those relatively small and isolated noise pixels.

5. Results and discussions

The in vivo microscopy video used in our experiments consists of 148 frames with several places of camera/subject movements. As the preprocessing step, the entire frame sequence is aligned by the 2D homography based method.

Our experiments on detecting moving leukocytes show a false positive rate of 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 manually. However, we observe a relatively low recall rate (≈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.

For adherent leukocytes detection, we randomly pick 11 adherent leukocytes from the video sequence. The total appearance count of the selected leukocytes is 1505 for the testing video. The total number of recognized leukocytes is 1479, making the accuracy rate of detecting adherent leukocytes about 98.37%.

6. 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.

7. References

[1] M. B. Lawrence and T. A. Springer, "Leukocytes roll on a selection at physiologic flow rates: distinction from and prerequisite for adhesion through integrins," *Cell*, vol. 65, pp. 859-73, 1991.

[2] S. I. Anderson, R. Shiner, M. D. Brown, and O. Hudlicka, "ICAM-1 expression and leukocyte behavior in the microcirculation of chronically ischemic rat skeletal muscles," *Microvasc Res*, vol. 71, pp. 205-11, 2006.

[3] H. Yuan, D. J. Goetz, M. W. Gaber, A. C. Issekutz, T. E. Merchant, and M. F. Kiani, "Radiation-induced up-regulation of adhesion molecules in brain microvasculature and their modulation by dexamethasone," *Radiat Res*, vol. 163, pp. 544-51, 2005.

[4] E. Eden, D. Waisman, M. Rudzsky, H. Bitterman, V. Brod, and E. Rivlin, "An automated method for analysis of flow characteristics of circulating particles from in vivo video microscopy," *IEEE Trans. on Med Imaging*, vol. 24, pp. 1011-24, 2005.

[5] S. T. Acton, K. Wethmar, and K. Ley, "Automatic tracking of rolling leukocytes in vivo," *Microvasc Res*, vol. 63, pp. 139-48, 2002.

[6] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, pp. 381-395, 1981.

[7] N. Otsu, "A threshold selection method from gray-level histogram," *IEEE Trans. on Systems, Man, Cybernetics*, vol. 9, pp. 62-66, 1979.

[8] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the 1999 International Conference on Computer Vision*, pp. 1150-57, 1999.

[9] J. S. Beis and D. G. Lowe, "Shape Indexing Using Approximate Nearest-Neighbour Search in High-Dimensional Spaces," in *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR'97)*: IEEE Computer Society, 1997.

[10] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*: Cambridge University Press, 2003.

[11] A. M. Elgammal, D. Harwood, and L. S. Davis, "Non-parametric Model for Background Subtraction," in *Proceedings of the 6th European Conference on Computer Vision-Part II*: Springer-Verlag, 2000.