Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

A NOVEL WHITE BLOOD CELL DETECTION METHOD BASED ON BOUNDARY SUPPORT VECTORS

Min Wang[1], *Rong Chu*[2]

1. College of Computer & Information Engineering, Hohai University, Nanjing, 210098, China
2. College of Electrical Engineering, Hohai University, Nanjing, 210098, China

ABSTRACT

White blood cell (WBC) detection is one of the most basic and key steps in the automatic WBC recognition system. Its accuracy and stability greatly affect the recognition accuracy of the whole system. This paper presents a novel method for WBC detection based on boundary support vectors (BSVs). Firstly, ν-Support Vector Regression (ν-SVR) is introduced. Then sparse BSVs are obtained while fitting the 1D histogram by ν-SVR. Next so-needed threshold value is directly sifted from these limited support vectors. Finally the entire connective WBC regions are segmented from the original cell image. The proposed method successfully works for WBC detection, and effectively reduces the influence brought by illumination and staining. It also has the advantages, such as high computing efficiency and easy parameter setting. Experimental results demonstrate its good performances.

Index Terms— Object detection, support vector machine, image segmentation

1. INTRODUCTION

White blood cell (WBC) differential count as a percentage is one of the most frequently performed blood tests and provides a lot of valuable information for the quantitative analysis, life study and early diagnoses of various diseases. In hospital, manual differential is usually performed by hematology experts using microscope, which is unfortunately a highly complex, tedious and time-consuming task. The development of computer and image processing technology make "brain-eye system" [1] possible, i.e., an attempt to automate the task performed manually by experts in routine. With the system, we can alleviate workload, shorten the hemanalysis time, eliminate the influence of subjective factors, and improve the recognition accuracy at the same time.

The process of an automatic differential blood counter system may require four steps: acquisition, detection, feature extraction and classification. In the first step, the blood smear is magnified to a suitable scale under the microscope, then we transform it into digital image with modern charge-coupled devices (CCD) camera; In the second step, cell segmentation yields a number of single-cell images and each single-cell image is segmented into three regions: cell nucleus, cytoplasm and background; In the third step, the feature extraction engine analyzes each segmented cell and its nucleus to form a feature vector from color, shape and texture features; In the last step, each WBC is labeled by the classifier according to its feature vectors. It is easy to see that WBC detection, i.e. detecting WBC regions from microscopic images is one of the most basic and key steps in the whole system. Its accuracy and stability have direct and great influence on the recognition accuracy of the whole system.

Up to now, numerous WBC detection methods for microscopic cell images have been proposed [2-10], most of which are edge-based or region-based scheme. Edge-based scheme performs poorly on cell images because not all cell boundaries are sharp and hence it is difficult to get all the edge information and locate the cells accurately. The region growing methods, composing one branch of region-based scheme, rely greatly on the proper markers extraction by empirical threshold using prior knowledge. They also require the color homogeneity in each component of cell images, which is often unsatisfied for the cytoplasm. As another typical class of region-based scheme, although threshold segmentation was proposed a long time ago, it attracts the attentions of the researchers worldwide till now, because it is a simple, fast yet effective tool to separate objects (WBCs) from the background. It is just helpful for the improvement of the recognition accuracy and running speed of the whole automatic recognition system. Thus, in this paper, we will follow this line.

As everyone knows, threshold selection plays the key role in threshold segmentation. Otsu method is one of the most referenced threshold selection methods. Recently, an iterative Otsu method (IterOtsu for short) based on circular histogram by taking full use of HSI space information is presented in [10]. It is a pity that the traditional methods just based on histogram fitting, such as IterOtsu, is generally difficult in threshold selection, because the resulting fitting function has many local optimal values to be determined by iterations. As a result, such a process is usually time-consuming.

In contrast, our method is essentially different. We view the histogram of a given image as a frequency distribution

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

of its gray levels. Then the 1D histogram is fitted by ν-support vector regression (ν-SVR) [11] in order to obtain all boundary support vectors (BSVs). At last the so-needed threshold value is sifted or selected automatically and directly from the BSVs rather than the extremes of the fitted histogram. Obviously, it is not necessary for us to search for the optimal threshold in the whole range of gray levels like the conventional threshold selection methods. Thanks to the sparseness of BSVs, the computation efficiency can be greatly improved. Besides, such a method can avoid the oversegmentation of the traditional segmentation methods based on histogram fitting.

It should be mentioned that, in recent years, support vector machines have been introduced into the field of medical image segmentation [12-14]. The common characteristic of these methods is that the threshold determination problem is usually treated as a case of classification. In other words, in these methods, support vector machines are just regarded as a better substitution for the traditional classifiers. The proposed method in this paper is obviously different from the above methods based on support vector machines.

The rest of the paper is organized as follows. After the presentation of a brief review of ν-SVR, the details of our novel detection method are given in Section II. In Section III, experimental results for the proposed method and comparison with IterOtsu in detection ratio and completeness degree are presented. Section IV concludes this paper.

2. A NOVEL DETECTION METHOD BASED ON BOUNDARY SUPPORT VECTORS

Before the detailed description of our method, it is necessary to give a brief review of ν-SVR first.

2.1. ν-Support Vector Regression (ν-SVR)

Let the training set D be $\{(\mathbf{x}_i, y_i)\}_{i=1}^{l}$, with input $\mathbf{x}_i \in \mathbb{R}^d$ and output $y_i \in \mathbb{R}$. To estimate a linear function $f(\mathbf{x}) = (\mathbf{w} \cdot \mathbf{x}) + b$ from the training data, at each point $\mathbf{x}_i$, ν-SVR allows an error of ε. Everything above ε is captured in slack variables $\xi_i^{(*)}$ ($^{(*)}$ being a shorthand implying both the variables with and without asterisks), which are penalized in the objective function via a regularization constant C, chosen a priori [15]. The tube size ε is traded off against model complexity and slack variables via a constant $\nu \geqslant 0$:
minimize

$$\tau(\mathbf{w}, \xi^{(*)}, \varepsilon) = \|\mathbf{w}\|^2 / 2 + C \cdot (\nu\varepsilon + \frac{1}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*)) \quad (1)$$

subject to

$$((\mathbf{w} \cdot \mathbf{x}_i) + b) - y_i \leq \varepsilon + \xi_i \quad (2)$$

$$y_i - ((\mathbf{w} \cdot \mathbf{x}_i) + b) \leq \varepsilon + \xi_i^* \quad (3)$$

$$\xi_i^{(*)} \geq 0, \varepsilon \geq 0 \quad (4)$$

Introducing a Lagrangian with multipliers $\alpha_i^{(*)}, \eta_i^{(*)}, \beta \geq 0$, the Wolfe dual problem is obtained. Moreover, a kernel k is substituted for the dot product, corresponding to a dot product in some feature space related to input space via a nonlinear map Φ,

$$k(\mathbf{x}, \mathbf{y}) = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})) \quad (5)$$

This leads to the ν-SVR optimization problem: for $\nu \geqslant 0$, $C>0$, maximize

$$W(\alpha^{(*)}) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)y_i - \frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)k(\mathbf{x}_i, \mathbf{x}_j) \quad (6)$$

subject to

$$\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \quad (7)$$

$$0 \leq \alpha_i^{(*)} \leq \frac{C}{l} \quad (8)$$

$$\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) \leq C \cdot \nu \quad (9)$$

The regression estimate can be shown to take the form

$$f(\mathbf{x}) = \sum_{i=1}^{l}(\alpha_i^* - \alpha_i)k(\mathbf{x}_i, \mathbf{x}) + b \quad (10)$$

All $\mathbf{x}_i$ corresponding to nonzero $\alpha_i^{(*)}$ constitute the support vector set (SV set for short). And all $\mathbf{x}_i$ on bound, i.e., satisfying α_i or $\alpha_i^* = C$, are members of the boundary support vector set (BSV set for short).

The theoretical and experimental analysis have suggested that ν provides a way to control an upper bound on the number of training errors which is tighter than the parameter ε used in standard SVR. Asymptotically, it directly controls the number of SVs [16]. This makes our method more convenient in parameter setting. Besides, the desirable properties of ν-SVR, including the formulation as a definite quadratic program, and the sparse SV representation of the solution, are retained. In a word, ν-SVR is more robust than ε-SVR in real applications. That is the reason why we choose ν-SVR as the fitting tool rather than ε-SVR in this paper.

2.2. The Detail of the Novel Detection Method

As shown by Chassery and Garbay [17], nucleus pixels can be easily separated from others in the image histogram of green layer. But an automatic threshold value may not be easily obtained because WBCs cannot form an obvious cluster in histogram due to a relatively small area in whole cell image. To avoid this problem, a method based on BSVs is proposed as a solution to automatically find the threshold value.

Let each pixel of green layer image have gray level in [0, 1, 2, …, L-1] , and commonly L = 256. The number of pixels with gray level i is denoted by n_i, i=0, 1, 2, …, L-1, and the total number of pixels is denoted $N = n_0+n_1+\ldots+n_{L-1}$. Thus, the gray level histogram is defined as a probability distribution: $p(i) = n_i/N$, $p(i)\geq 0$, and $\sum_{i=0}^{L-1} p(i) = 1$. Forming a histogram $p(x)$ of the image results in an ordered set of discrete values, $p(0)$, $p(1)$, …, $p(L\text{-}1)$. Suppose those zero $p(j)$s' are deleted and we pack the remaining nonzero $p(k)$'s, say m nonzero $p(k)$'s, into an array with size m. We thus have a compact image histogram with the probability distribution $\langle p(i_0), p(i_1), \ldots, p(i_{m-1})\rangle$ for $0\leq i_j\leq L$-1, j=0, 1, 2, …, m-1, and $p(i_j)\neq 0$. The total number of pixels is denoted by $N = n_{i_0} + n_{i_1} + \cdots + n_{i_{m-1}} = n_0' + n_1' + \cdots + n_{m-1}'$, where $n_{i_j} = n_j' = N \times p(i_j)$.

Our aim is to approximate the set of discrete values by some fitting method such that the p_j, j=0, 1, 2, …, m-1 can be replaced by q_j, j=0, 1, 2, …, m-1. In this paper, we choose the SVR as the fitting tool rather than the traditional histogram smoothing technique. An advantage of doing so is that the generated support vectors on boundary can provide us good candidate thresholds, consequently, it is not necessary for us to employ some complicated optimization technique to find so-needed threshold from the extremes of the fitted or regression histogram. Besides, the SVR has much better noise-tolerance than the traditional smoothing technique.

Next we will seek out the optimum threshold value directly from the BSV set. We calculate first order derivative of the regression histogram with respect to each member in the BSV set. From this derivative we can obtain the threshold value by identifying the gray levels or BSVs within the areas where negative to positive transition is occurring.

According to the above description, the proposed algorithm is summarized below:

1. Form a compact histogram from the green layer of a color microscopic image;
2. Use the ν-SVR to approximate the above compact histogram and let BSV be the set consisting of all support vectors on boundary;
3. Seek the threshold from the BSV set to ensure it to lie in the areas where negative to positive transition of the first order derivative is occurring;
4. Use the so-obtained threshold to segment the given image;
5. Perform morphological operations to the above image in order to obtain the entire connective WBC region.

3. EXPERIMENTAL RESULTS AND COMPARISONS

In this section, we compare the performance of our method with the existing method IterOtsu [10], and apply the two methods to fifty microscopic images under different circumstances of staining and illumination. In our method, RBF kernel is selected as the kernel function. And the parameter ν is fixed at 0.2. Experimental results are shown in Fig.1.

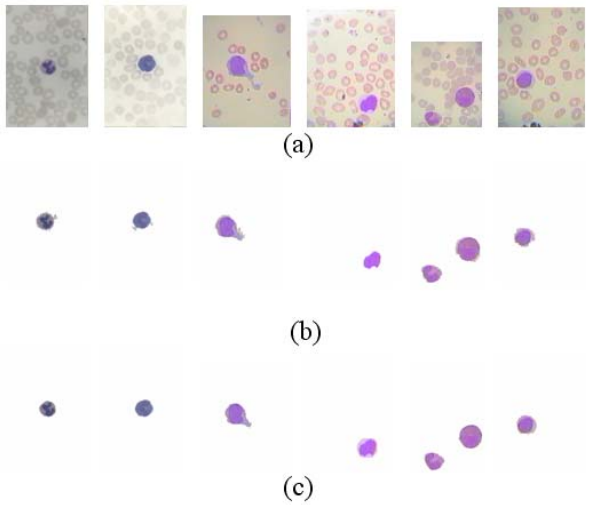

Fig. 1 Experimental comparisions between our method and IterOtsu. (a) The original color microscopic images. (b) The results obtained by IterOtsu. (c) The results obtained by our method.

From the six groups of experiments, we can see that most WBCs can be detected by IterOtsu. In other words, the ratio of the number of the detected WBCs by IterOtsu to the number of the existing WBCs in all the microscopic images is acceptable. However, in general, IterOtsu cannot make sure that the contour of a detected WBC is complete. For example, the cytoplasm in the last three microscopic images is partly or even entirely omitted (see Fig.1(b)). As we know, the completeness degree of the detected WBC, especially the contour, has close relation with the classification accuracy of the whole system because the features extracted from the detected WBC affect the classification performance very much. In the case of incomplete cells detected by IterOtsu, the extra assistance of other modification algorithms or man-machine interface methods is needed. Moreover, this method greatly suffers from the factors such as illumination and staining (see Fig.1(b)). Besides, IterOtsu entails an iteration procedure, which is wasteful of computation. All these disadvantages limit its real application.

Our method based on BSVs retains the merits of IterOtsu in detection ratio and avoids its disadvantages. With the proposed method, we can detect almost all WBCs in the microscopic images and each cell is nearly complete. Some of the experimental results are demonstrated in Fig.1(c). Obviously, our method has stronger adaptability to staining and illumination than IterOtsu. Owing to the sparseness of BSV set, we can expect that the running speed of our method is comparatively much higher against the other threshold selection methods including IterOtsu, because the latter usually adopt the traditional combination optimization technique. From the above analysis, we can easily see that our method is superior to IterOtsu.

4. CONCLUSIONS

The WBC detection step in the automatic recognition system of microscopic images has been addressed in this paper. Different from the other existing threshold selection methods, a new WBC detection method based on boundary support vectors is proposed. By use of support vector regression, we can obtain the limited and sparse BSV set and then sift the so-needed threshold value from it directly and efficiently. With this novel method, we detect almost all WBCs and each detected cell is nearly complete. Its adaptability is strong and the running speed is comparatively high. Experimental results show its good performances.

ACKNOWLEGEMENTS

This paper is supported by the National Science Foundation of China under Grant No. 30700183, the Doctor Point Foundation of Education Ministry of China under Grant No. 20070294001, and the Innovative Talents Project of Hohai University.

REFERENCES

[1] I.T. Young, "The classification of white blood cells," *IEEE Trans. BME*, vol.19, pp. 291-298, 1972.

[2] S.S.S. Poon, R.K.Ward, and B. Palcic, "Automated image detection and segmentation in blood smears," *Cytometry*, vol. 13, pp. 766-774, 1992.

[3] I. Cseke, "A fast segmentation scheme for white blood cell images," in Proc. International Conference on IAPR, vol. 3, pp. 530-533, 1992.

[4] Yang Xiao-min, Luo Li-min and Wei Yu, "Automatic classification system for leukocytes in human blood," *Chinese Journal of Computers*, vol. 17, no. 2, pp. 130-136, 1994 (in Chinese).

[5] V.A.Kovalev, A.Y.Grigoriev, and H-S. Ahn, "A reliable segmentation technique for white blood cell image recognition," in Proc. of the 3rd International Conf. on Pattem Recognition and Inf. Analysis, Sept. 1995, Minsk, pp. 177-182, 1995.

[6] O. Lezoray, A. Elmoataz, H. Cardot, et al., "Segmentation of cytological images using color and mathematical morphology," *Eur Congr Stereol*, vol. 18, no. 1, pp. 1-14, 1999.

[7] Q. Liao, and Y. Deng, "An accurate segmentation method for white blood cell images," in Proc. IEEE International Symposium on Biomedical Imaging, 7-10, pp. 245-248, 2002.

[8] B. Kumar, and T. Sreenivas, "Teager energy based blood cell segmentation," in Proc. International Conference of Digital Signal Processing, vol. 2, pp. 619-622, 2002.

[9] E. Montseny, P. Sobrevilla, and S. Romani, "A fuzzy approach to white blood cells segmentation in color bone marrow images," in Proc. IEEE International Conference on Fuzzy Systems, vol. 1, pp. 173-178, 2004.

[10] J. Wu, P. Zeng, Y. Zhou, and C. Olivier, "A novel color image segmentation method and its application to white blood cell image analysis," in Proc. IEEE International Conference on Signal Processing, vol. 2, 2006.

[11] B. Schölkopf, A.J. Smola, R.C. Williamson, and P.L. Bartlett, "New support vector algorithms," *Neural Computation*, vol. 12, pp. 1207-1245, 2000.

[12] E.N. Issam, Y. Yang, et al., " A support vector machine approach for detection of microcalcifications," *IEEE Trans. MI*, vol. 21, no. 12, pp. 1552-1563, 2002.

[13] H.X. Xu, L. Yu, G.X. Zhu, et al., "Segmentation of magnetic resonance brain tissues image based on support vector machines," *Journal of Image and Graphics*, vol. 10, no. 10, pp. 1275-1280, 2005 (in Chinese).

[14] Z.P. Hu, Y. Wu, Y. Zhang, "Micro-calcifications detection based on iterative rank-order filter subspace and SVM with rejection feature," *Acta Electronica Sinica*, vol. 34, no. 2, pp. 312-316, 2006 (in Chinese).

[15] V. Vapnik, *The nature of statistical learning theory*, Springer Verlag, New York, 1995.

[16] B. Schölkopf, P. Bartlett, A. Smola, and R. Williamson, "Shrinking the tube: a new support vector regression algorithm," in M.S. Kearns, S.A. Solla, D.A. Cohn (Eds.), Advances in neural information processing systems, 11, Cambridge, MA: MIT Press, 1999.

[17] J.M. Chassery, C. Garbay, "An iterative segmentation method based on a contextual color and shape criterion," IEEE Trans. PAMI, vol. 6, no. 6, pp. 794-800, 1984.