Microscopic Image Segmentation Method based on SVM

Li Jianyi[1], Wang Huijuan[2*] and Li Shufeng[3]

[1,2] *North China Institute of Aerospace Engineering, Langfang, Hebei, 065000, P.R. China*
[3]*Institute of Technology, Langfang, Hebei, 065000, P.R. China*
[1]*lijianyi001@126.com,* [2] *wanghj323@126.com,* [3] *395012418@qq.com*

Abstract

In this paper, the existing microscopic image segmentation method is studied, by comparing the segmentation result, the support vector machine (SVM) of microscopic image segmentation method has good precision, deserves further research.

Keywords: SVM, white blood cell, image segmentation

1. Introduction

Image segmentation is the first step to realize the automatic blood cell identification analysis, which belongs to the lowest layer analysis technology of system. The result of image segmentation directly affects the following a series of analysis and recognition results. The current segmentation method is not able to produce satisfactory segmentation results to all real or synthetic images. If the target is clear and the background is uniform, segmentation problem becomes quite simple. Because of the complexity of the object in the medical image, image segmentation is one of the most difficult problems to solve. This paper mainly studies the microscopic image segmentation of white blood cells.

2. The Existing Cell Nucleus Extraction Method

Considering that determining the kinds of leukemia types is on the basis of WBC in cell classification, and white blood cell classification is based primarily on the nucleus, the segmentation process is to divide white blood cells of the nucleus with other background (cytoplasm, mature red blood cells and white blood cells).

The main existing cell nucleus extraction methods include three categories:

2.1. The Method Based on Threshold

At first, according to the distribution of image color (gray histogram), this kind of method makes pixel values, multi-value classification, and then in connected domain makes tags, analysis, processing (split and merge) and filter, in order to gain the nucleus area. When the nuclei are uneven color, this kind of method often generates hole in the cell nucleus by mistake, also generates the nucleus area. Although this can be achieved by mathematical morphology method for reprocessing, it is difficult to automatically select the appropriate structure elements, and more manual work is needed.

2.2. The Method based on Gradient (including Amplitude and Direction)

This kind of method gets the first image gradient information through differential operator or morphological operation, and then extracts nuclei according to the distinctive characteristic of gradient (local extreme). The method mainly includes the method based

---

on edge, deformable model method and watershed method, *etc*. Color change of the nucleus area will produce a lot of noise.

2.3. The Method based on Region

This kind of method mainly includes regional division merger and region growing two ways. Regional division combined methods don't recursively split the entire image according to the first four, until every uniform meets color consistency. Then form a recursive merge area which is spatial adjacency and consistent color. This method determines the color consistency of area only through average color. The nucleus with uneven color does not meet this requirement, will be the over-segmentation or not.

In region growing method, at first, a pixel or area is set as a seed domain, then the seed domain and the adjacent pixel domain is compared in color. If the difference is less than the setting threshold, the seed domain grows. Due to non-uniformity of the nucleus color, the growth based on a single pixel domain is easy to be over-segmented or owe devised.

After dyeing, the color of the nuclei is varied, and the shade is also uneven. The existing methods of extracting the nucleus from the microscopic image of blood are just based on the local information of a single pixel, which often cause the nucleus of over-segmentation and owe segmentation, and then segmentation result is difficult to meet the requirements of subsequent aided analysis.

3. Support Vector Machine (SVM) Method

3.1. The Thoughts of Support Vector Machine (SVM) Applied to Split

We divided all the pixels in blood image into two classes: class is pixels belong to the white blood cell nucleus area; Negative class is pixels beyond white blood cell nucleus.

Observe whole the blood image, choose some representative pixel as the training sample. Train them by support vector machine (SVM), find classification decision function belongs to white blood cell nuclei and other pixel, and finally save it. For any new blood image for segmentation, bring all the pixel values of image into the decision function; determine which category the pixel belongs to. In image class point is marked as 1, and negative class point marked as 0. We get the binary image of image segmentation.

Process the binary image after splitting, filter using morphological operator. In the binary image after filtering, find the boundary of the segmentation. The segmentation is over.

The specific work process is shown in Figure 1:

3.2. Support Vector Machine Segmentation Experiments

3.2.1. Data collection: Prepare for 200 images with 2 to 4 cell in each image.

3.2.2. The selection of training samples: Observe all the collected images, choose some representative white pixel points in the cell nucleus as samples, the pixel points outside the nucleus as negative samples. In order to investigate the influence of different sample size for segmentation results, we select the several groups of different Numbers of samples.
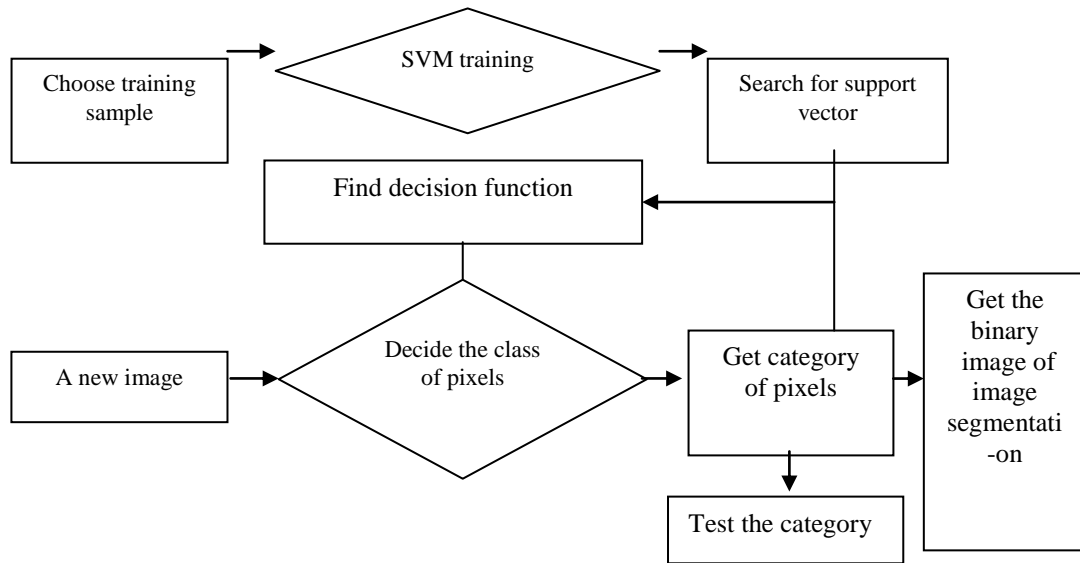
Figure 1. Process of Support Vector Machine (SVM) used to Split

3.2.3. The training process: Training selected sample with support vector machine (SVM). In the process of using support vector machine (SVM), problem of parameter selection is involved, which is selection of linear classification method or kernel function method, and when using soft interval optimization, how to determine the penalty factor. After repeated trials, I found that soft interval linear classifier can solve the segmentation problem, and the effect of soft interval linear classifier is better than kernel function to sample space projection to the high-dimensional space. In addition, when the default value of the interval linear soft penalty factor is 1 (C = 1), the effect is best.

In 2722 training sample, we found 94 support vectors, which is 3.453% support vector in the training sample. In the resulting classification function, there are 2719 sample points are strictly on both sides of the classification hyper plane according to their real category, and only three sample points are wrong points on the other side of the hyper plane, which is 99.896% accuracy for the training sample classification. Figure 2 is the diagram of this. Blue point in figure is a kind of sample points; green dot is negative sample points. The purple dot is all the support vectors by training. Upper yellow areas represent the class areas; the white areas represent the negative class areas. The straight line between the two regions is the optimal separating hyper plane.

3.2.4. Classification Accuracy (i.e. the segmentation accuracy): After the test, we get the test accuracy of 99.130%. Figure 3 is a classification diagram of the test data using classification decision function. The classification hyper plane in the graph is still to the one in the Figure 2, and support vector is still the one in Figure 2. We can see from the diagram decision function can correctly classify most points.
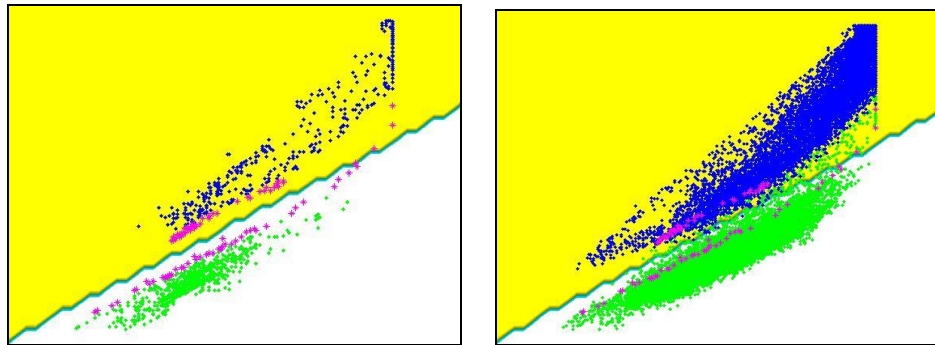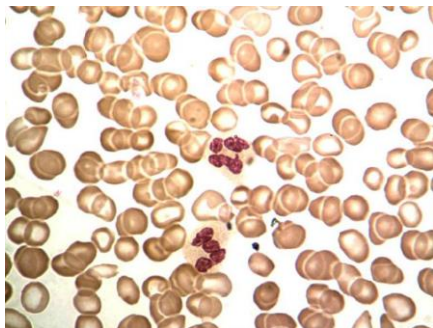
Figure 2. The Optimal Separating Hyperplane (left)

Figure 3. The Classification in the Face of the Test Data (right)

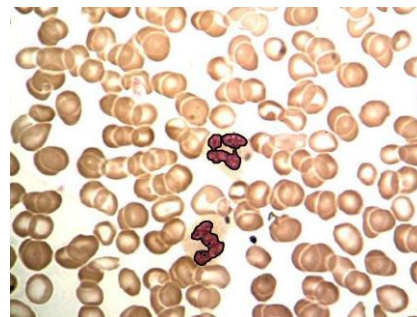### 3.2.5. Segmentation of Rendering:

(a) Original image

(b) The binary image after segmentation

(c) The binary image after filtering

(d) The final result of segmentation

Figure 4. Segmentation Effect

In order to filter out the noise signal, we process the image using morphological operator. We presented here the segmentation effect image. Figure 4 is a segmentation effect of the first image, in which Figure (a) is the original image, Figure (b) is the binary image after segmentation, Figure (c) is the binary image after filtering, Figure (d) is the final result of segmentation.

3.3.Effects of Different Samples Number for the Accuracy of the Segmentation

As we know, we use a priori knowledge into training sample selection. Selected sample, decides what support vector is, and the decision-making function. Selected number of samples is a great impact on the classification. How many samples should be selected to make the most accurate classification and to make the fastest computing speed? This is also an important task. We obtained six different sets of random samples, the number of samples which 6 groups of samples is incremented. Table 1 shows the accuracy of these six groups of samples were used to train the support vector obtained, the classification decision function segmentation.

Table 1. Effects of Different Samples Number for the Accuracy of the Segmentation

The number of positive class samples	The number of negative class sample	The number of support vector	The proportion of the total sample number of support vectors	Correct classification rate of training samples	Test classification accuracy
29883	21375	660	1.288%	98.855%	99.563%
18907	10520	386	1.312%	99.745%	99.301%
9739	7500	249	1.444%	99.785%	99.257%
4097	2910	161	2.298%	99.715%	99.155%
1393	1329	94	3.453%	99.896%	99.130%
728	432	47	4.052%	99.914%	99.103%

As shown in Table 1, With the increase of the number of samples, SVM and reduce the proportion of the number of samples, while testing classification accuracy increases. We can get the conclusion that the larger the sample size, the more we can find a good sample as a support vector classification results, the better the result is, since we selected samples randomly.

4. The analysis of Result

In order to illustrate the effect of SVM, we compare the result of SVM with other classification method. Our original image is shown in Figure 5.
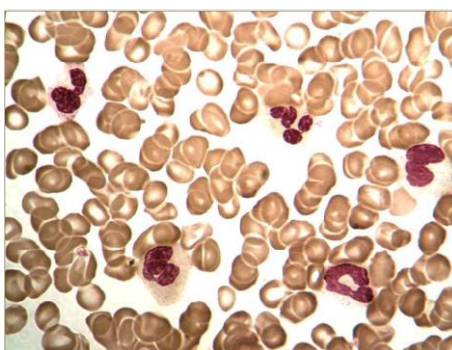
The result of SVM is shown in Figure 6.
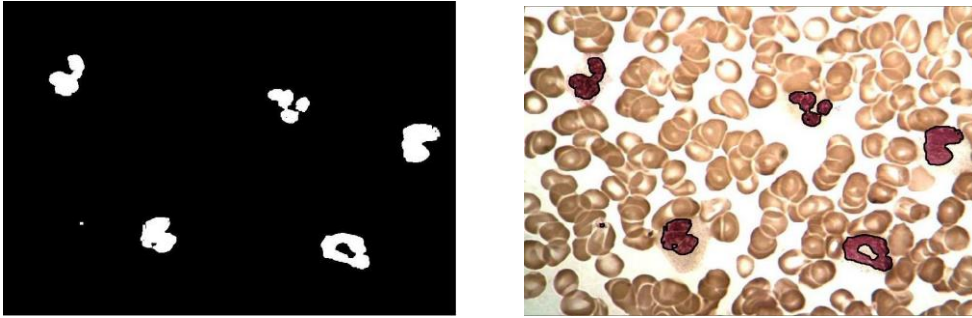

Fig.5 Original Image

Figure 6. Support Vector Machine Segmentation Binary Image (left) and Result Image (right)

4.1. Threshold Method

We consider threshold segmentation of G components in RGB color space and the H component in the HSV color space. These two components produce the best segmentation results.

We get the segmentation binary image and result image as follow in Figure 7. We can clearly find that effect on the threshold of G component is not as good as the support vector machine.
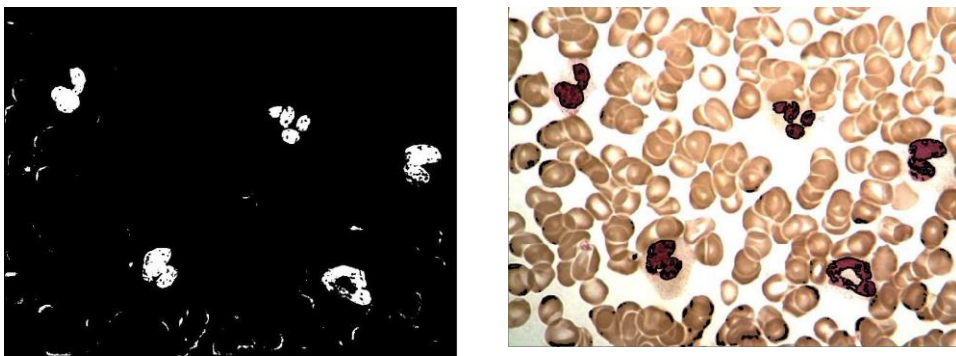

Figure 7. G Components Threshold Segmentation Binary Image (left) and Effect (right)

After many times of trial and error, the H component select threshold value of 0.95. The segmentation effect is shown in Figure 8.
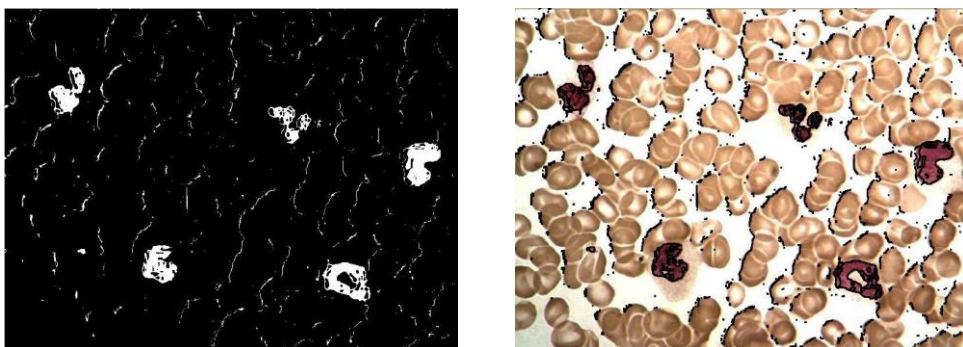

Figure 8. H Components Threshold Segmentation Binary Image (left) and Result Image (right)

If using threshold method, blood cell image must be converted into a gray scale image, and we can not automatically obtain the threshold function to obtain the desired threshold. Threshold has greater impact on the results, it need more work to select by repeated testing.

4.2.   The Gradient Operator Method and K-Means Clustering Algorithm in the Lab Color Space

We get the best segmentation results using canny operator as shown in Figure 9. We can also see, it is difficult to distinguish the the gradient of color change around the nucleus and other parts of, so the weak gradient operator to extract the nucleus.
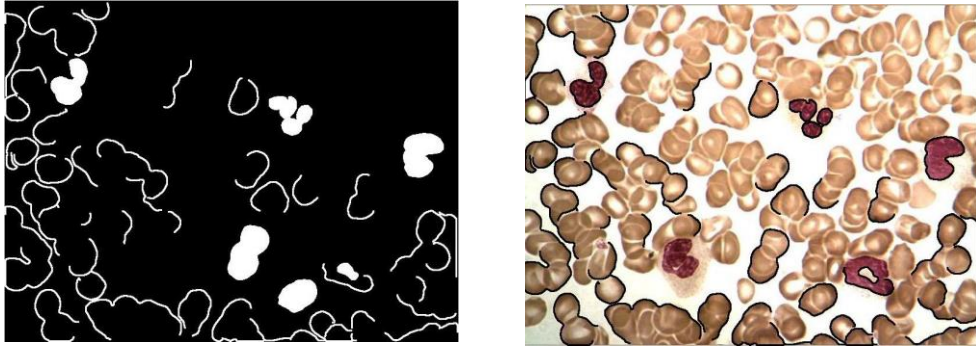

Figure 9. Gradient Operator Segmentation Binary Image (left) and Effect (right)

When using k-means clustering algorithm in the Lab color space, k= 4. The results of clustering is as shown in Figure 10. The algorithm needs long time and the stability of it is bad, while it needs to manually choose k value, which makes poor generalization. Figure 10 and Figure 11 is the segmentation effect of k-means clustering method. Figure 10 is a diagram of all the pixels clustered into four groups clustering results, and a kind of a nucleus containing the positive class, remaining as a negative type, is converted to the binary image, Figure 11 is the effect image after filterring.
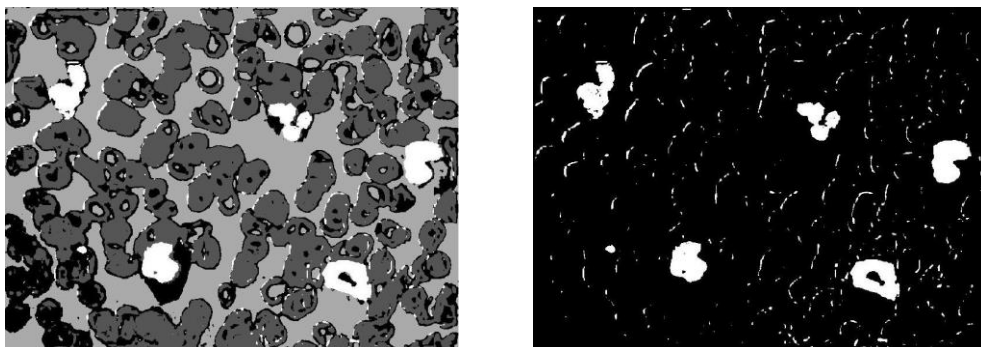

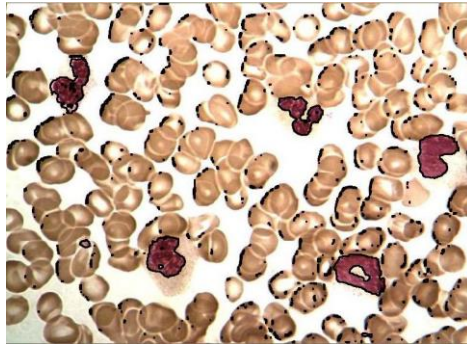**Figure 10. Clustering Results (left) and Binary Image (right)**

Figure 11. Segmentation Results of K-Means Clustering Method

5. Conclusion

This paper studied the segmentation method based on the support vector machine (SVM) and other method in contrast. We can find that SVM introduced prior knowledge to the algorithm, improved information selectively and judgment of the algorithm, and can make good use of the RGB tricolor information. This algorithm The algorithm requires less artificial participation, so it can be further promoted.

In order to get more types of white blood cell classified, we need to find more effective feature, such as texture inside the nucleus, cytoplasm of the particle size and density, and so on. To get these features, mathematical methods such as wavelet transform might be uses.

Nucleolus nucleus also has a lot of information, which is also very useful to identification of leukocytes. Because the nucleus nucleolus feature relative to the background is relatively weak, so it is easy to recognize nucleoli exist or not by artificial visual, and automatic detection method to accurately extract the feature is less easy. In the classification process, there are a lot of decision-making methods in a variety of categories, such as k- nearest neighbor, Fisher linear discriminant analysis method and so on. Which Used in this paper is one way to support vector machines. For the classification of blood cells which method is more suitable needs to be studied.

Acknowledgments

References

[1] R C. Gonzalez and R E. Woods, "Digital Image Processing [J]", Addison Wesley Publishing Co Reading Mass London Amsterdam, **(2008)**.

[2] J. M. S. Prewitt and M. L. Mendelsohn, "THE ANALYSIS OF CELL IMAGES [J]", Annals of the New York Academy of Sciences, vol. 128, **(1966)**, pp. 1035–1053.

[3] I. T. Young, "The Classification of White Blood Cells [J]", Biomedical Engineering IEEE Transactions on, vol. 4, **(1972)**, pp. 291 - 298.

[4] E. S. G, H. F. B and A.W. S, "Application of the method of multiple thresholding to white blood cell classification[J]", Computers in Biology & Medicine, vol. 18, no. 2, **(1988)**, pp. 65–74.

[5] J. F. Brenner, J. M. Lester and W. D. Selles, "Scene segmentation in automated histopathology: techniques evolved from cytology automation [J]", Pattern Recognition, vol. 13, no. 1, **(1981)**, pp. 65–77.

[6] K P, C. M. Blood analysis using black and white digital images.[J]. J Biomed Eng, 1992, 14(4):287-292.

[7] Haussmann G, Liedtke C -. A region extraction approach to blood smear segmentation [J]. Computer Vision Graphics & Image Processing, 1984, 25(2):133–150.

[8] JM C, C. G. An Iterative Segmentation Method Based on a Contextual Color and Shape Criterion[J]. IEEE Trans Pattern Anal Mach Intell, 1984, 6(6):794-800.

[9] C G. Image Structure Representation and Processing: A Discussion of Some Segmentation Methods in Cytology[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1986, 8(2):140-146.

[10] Poon S S S, Ward R K, Palcic B. Automated image detection and segmentation in blood smears.[J]. Cytometry, 1992, 13(7):766–774.

[11] Hong-jun L, Ji-wan H. Digital Image Processing and Its Application[J]. Computer Automated Measurement & Control, 2002, 10(9):620-622.

[12] Bamford P C. Segmentation of Cell Images with Application to Cervical Cancer Screening[J]. Cervical Cancer, 1999.

[13] F Y, T. J. Cell Image Segmentation with KernelBased Dynamic Clustering and an Ellipsoidal Cell Shape Model[C]. //Journal of Biomedical Informatics. 2001:67-73.

[14] Park J, Keller J M. Snakes on the Watershed[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2001, 23(10):1201-1205.

[15] Horowitz S L. Picture segmentation by a tree traversal algorithm. Journal of Acm. 1976:368--388.

[16] Mehnert A, Jackway P. An improved seeded region growing algorithm. Pattern Recognition Letters. 1997:1065–1071.

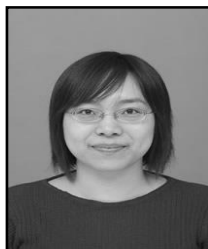# Authors

Li Jianyi

Born on 20 March 1974.

Scientific interest: Parallel and distributed algorithms and Embedded System., study in Beijing Information Science & Technology University and get Bachelor degree from 1993.9 to 1997.6, and study in BeiJing University of Posts and Telecommunications and get Master degree of Engineering from 2000.9 to 2004.6, be a teacher in North China Institute of Aerospace Engineering from 1997.7 up to now.

Wang Huijuan

Born on 23 March 1982.

Scientific interest: Parallel and distributed algorithms and Embedded System., study in Nankai University and get Bachelor degree and Master degree of Engineering from 2001.9 to 2008.6, be a teacher in North China Institute of Aerospace Engineering from 2008.6 up to now.