

A Leukocyte image fast scanning based on max–min distance clustering

Yapin Wang and Yiping Cao*
Department of Optical Electronics
Sichuan University, Chengdu
Sichuan 610064 P. R. China
**ypcao@scu.edu.cn*

Received 20 August 2015
Accepted 24 November 2015
Published 21 January 2016

A leukocyte image fast scanning method based on max-min distance clustering is proposed. Because of the lower proportion and uneven distribution of leukocytes in human peripheral blood, there will not be any leukocyte in larger quantity of the captured images if we directly scan the blood smear along an ordinary zigzag scanning routine with high power ($100\times$) objective. Due to the larger field of view of low power ($10\times$) objective, the captured low power blood smear images can be used to locate leukocytes. All of the located positions make up a specific routine, if we scan the blood smear along this routine with high power objective, there will be definitely leukocytes in almost all of the captured images. Considering the number of captured images is still large and some leukocytes may be redundantly captured twice or more, a leukocyte clustering method based on max–min distance clustering is developed to reduce the total number of captured images as well as the number of redundantly captured leukocytes. This method can improve the scanning efficiency obviously. The experimental results show that the proposed method can shorten scanning time from 8.0–14.0 min to 2.5–4.0 min while extracting 110 nonredundant individual high power leukocyte images.

Keywords: Leukocyte image; fast scanning; scanning routine; max–min distance clustering; window clustering; microscopic imaging; image segmentation.

1. Introduction

Quantities of different types of leukocytes¹ in human peripheral blood is the most commonly performed medical data to play a vital role in the diagnosis of various diseases, therefore, researching

leukocyte image fast scanning is meaningful. There are five kinds of leukocytes with different functions in human peripheral blood. Blindly counting the total number of leukocytes is meaningless. It is more reasonable to use differential counting for disease

*Corresponding author.

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.

diagnosis. The morphology structures of leukocyte can help doctors to analyze the pathogenesis. Only the high power microscopic images can deliver the detailed morphology structures of leukocyte. If the current leukocyte scanning mode is directly using high power (100^x) objective to scan along an ordinary zigzag routine, lots of images without any leukocyte are captured due to the lower proportion and uneven distribution of leukocytes in human peripheral blood. There is no point in capturing the regions without any leukocyte, besides it will increase follow-up workload. Thus, the scanning efficiency is low, and many computer resources are wasted in saving and processing the useless images without any leukocyte. Due to the low power (10^x) objective's larger field of view, there are definitely many leukocytes in the captured low power images though the morphology structures of leukocyte are not detailed. Because the leukocyte has nucleus while the erythrocyte does not have, leukocyte nuclei can be segmented by image segmentation²⁻⁴ and all the leukocyte nuclei that have been segmented can be positioned by image processing.⁵⁻⁷ This paper uses image binarization⁸⁻¹⁰ to achieve

the segmentation of leukocyte nuclei. If these leukocyte nucleus' centroids can compose a special routine, all the leukocyte nuclei that have been segmented will be efficiently captured on purpose when scanning with high power objective. But scanning along each leukocyte nucleus' centroid one by one is still ineffective. The number of totally captured images is still large and this method will lead to redundantly capturing of many leukocytes. Applying clustering algorithm^{11,12} on leukocyte nucleus' centroids in leukocyte low power microscopic image based on max-min distance clustering means¹³⁻¹⁵ can greatly reduce the total number of captured images and the proportion of redundantly capturing. Experimental results indicate that leukocyte image fast scanning method based on max-min distance clustering has a higher efficiency.

2. Leukocyte Localization

If we directly scan a blood smear along a zigzag routine with high power objective, it is certainly that the expected images containing one or more intact leukocytes can be captured as shown in Fig. 1(a).

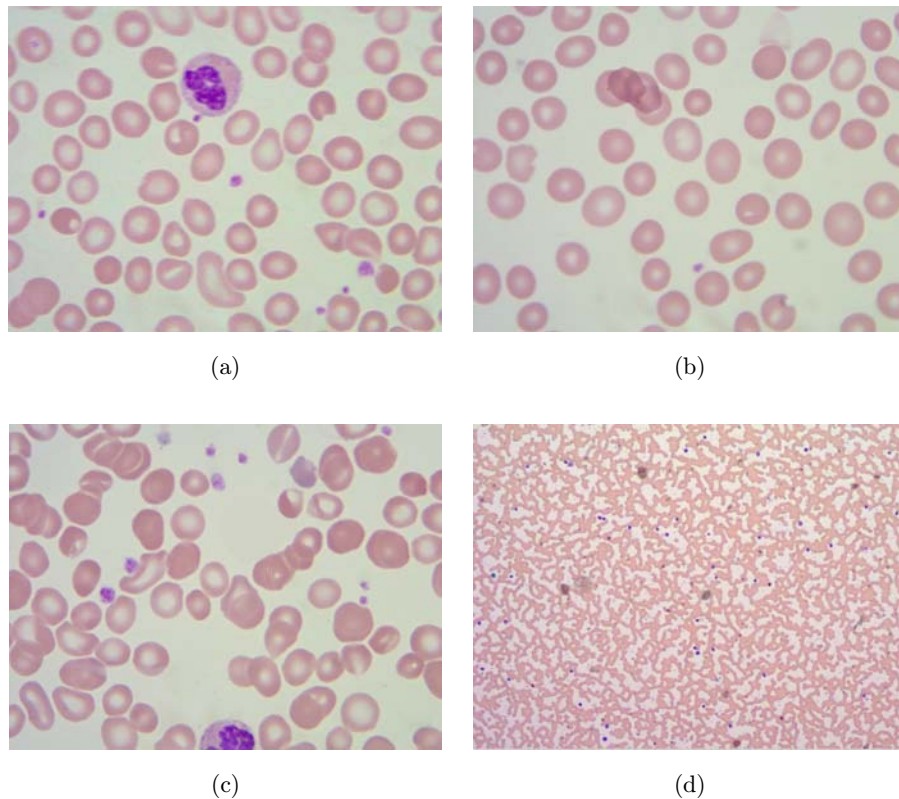


Fig. 1. Captured images while scanning at different power objectives. (a) Expected leukocyte image (b) Image without any leukocyte (c) Image with an incomplete leukocyte (d) Captured low power image.

But because of the lower proportion and the uneven distribution of leukocytes in blood smear, many captured high power images contain no leukocyte completely or contain only an incomplete leukocyte as shown in Fig. 1(b) and 1(c) while scanning. So many unexpected images will not only waste saving space, but also increase follow-up workload. Due to the low power objective's larger field of view, the captured low power blood smear image as shown in Fig. 1(d) may contain many unevenly distributing leukocytes (with nucleus) though there are not enough details of leukocytes' morphology structure in it. If we pre-scan the blood smear with low power objective along the zigzag routine as shown in Fig. 2 (a), several images containing different amounts of leukocytes as shown in Fig. 2(b) may be captured and the total amounts of leukocytes in these images are more than 110, which is sufficient to the medical needs. If we manage to locate these leukocyte nuclei by image segmentation and all of the located positions make up a specific routine, while we scan the blood smear along this specific routine with high power objective, there will be definitely leukocyte with details of leukocytes' morphology structure in almost all of the captured images. Locating leukocytes in low power microscopic images firstly is a way to improve scanning efficiency.

As shown in Fig. 3(a), leukocyte low power microscopic image is a color image. Commonly used color spaces are RGB color space and HSI color space. R represents red component, G represents green component, B represents blue component.

HSI color space includes three components: hue (H), saturation (S) and intensity (I). Conversion formulas between RGB color space and HSI color space are as follows:

$$H = \begin{cases} \arccos\left\{\frac{[(R-G)+(R-B)]/2}{[(R-G)^2+(R-G)(G-B)]^{1/2}}\right\} & B \leq G \\ 2\pi - \arccos\left\{\frac{[(R-G)+(R-B)]/2}{[(R-G)^2+(R-G)(G-B)]^{1/2}}\right\} & B > G, \end{cases} \quad (1)$$

$$S = 1 - \frac{3}{(R+G+B)}[\min(R, G, B)], \quad (2)$$

$$I = \frac{1}{3}(R+G+B). \quad (3)$$

Image segmentation is a process which differentiates leukocyte regions and some impurity regions from background. Lots of experimental results show that S component can segment the leukocytes and some impurities while B component can be used to filter the impurities of blood smear. Figure 3(b) is the S component grayscale of Figure 3(a). It shows that leukocyte regions and some impurity regions have the higher grayscale. Fig. 3(c) is the B component grayscale of Fig. 3(a). It shows that some impurity regions have the lowest grayscale. Thus, we use S component to locate leukocyte and the impurities, and use B component to filter impurities.

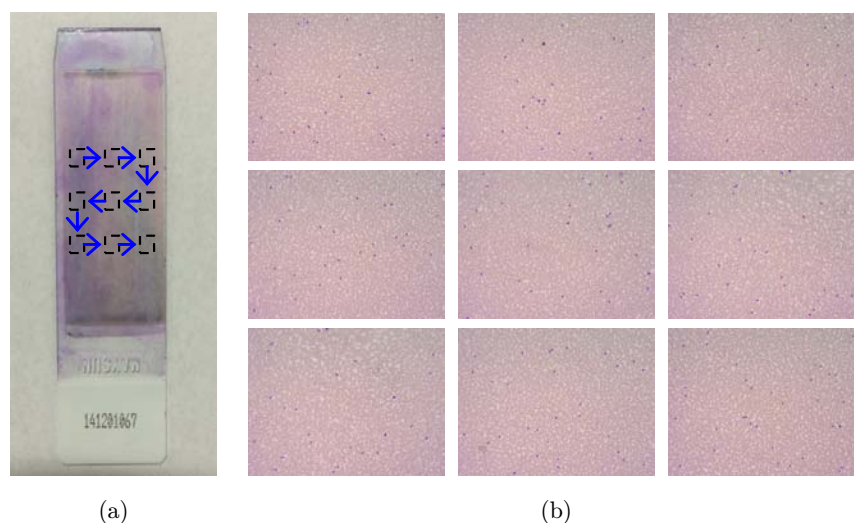


Fig. 2. Pre-scanning procedure with low power objective. (a) Blood smear and pre-scanning routine and (b) captured low power images while pre-scanning.

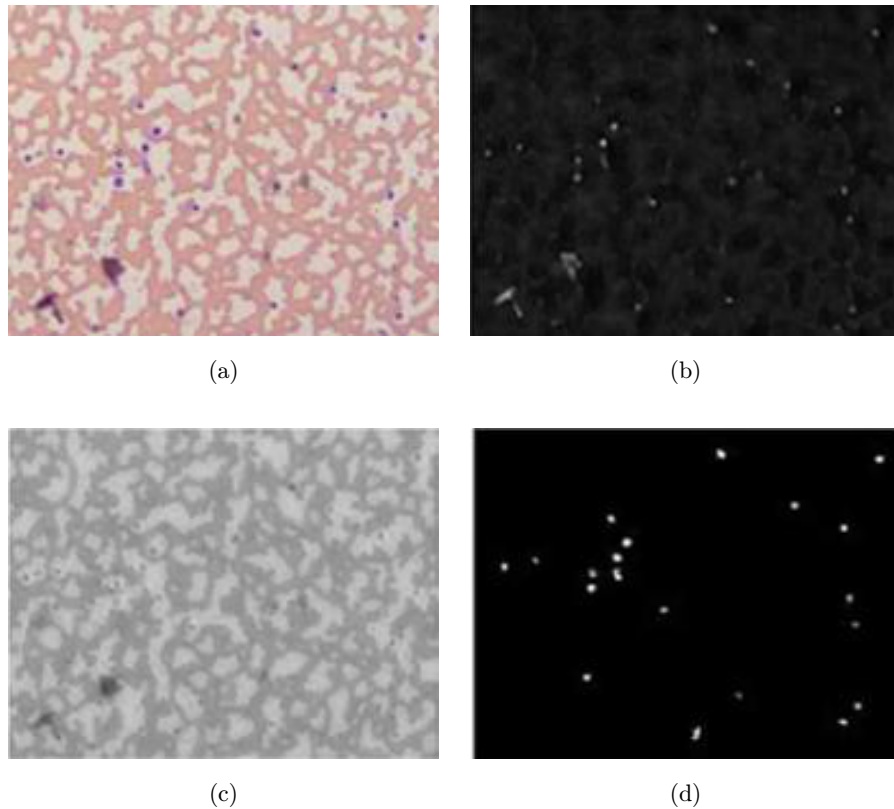


Fig. 3. Leukocyte low power microscopic image and its processed images. (a) Low power microscopic image, (b) S component image of (a) (c) B component image of (a), (d) binarization image of (a) without any impurity.

Image threshold segmentation also called image binarization is an important image segmentation method. Threshold selection^{16,17} is the key step in image binarization process. Otsu^{18,19} algorithm is a favorable threshold selection method, but for the consideration of shorter time consumption, we use the mean value threshold method. It goes like this:

$$T = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} f(i, j). \quad (4)$$

In formula (4) where T is the threshold, N is total number of pixels per row and M is total number of rows of the selected image $f(i, j)$.

After solving the problem of threshold selection, the specific procedures of leukocyte location is as follows:

Step 1. Separating R , G and B component of the captured leukocyte low power microscopic image by image processing; calculating S component of the image by formula (2).

Step 2. Calculating thresholds of S component and B component using formula (4); getting

binarization images of S component image and B component image.

Step 3. Obtaining coordinate range of impurities in binarization images of B component; removing the coordinate range of impurities obtaining above in binarization images of S component.

Step 4. Connecting the areas whose centroid are very close, thereby singling out paging lymphocytes; removing the areas too small or too large, so as to filter out impurities.

After the above procedure, we can get the binarization image of leukocytes without any impurity, as shown in Fig. 3(d).

3. Fast scanning method based on max-min distance clustering

By morphological detection, the leukocytes' centroids in the binarization image with only leukocyte nuclei as shown in Fig. 3(d) can be figured out. If we directly use each leukocyte nucleus' centroid as the corresponding scanning center of high power objective, all the located leukocytes will be definitely

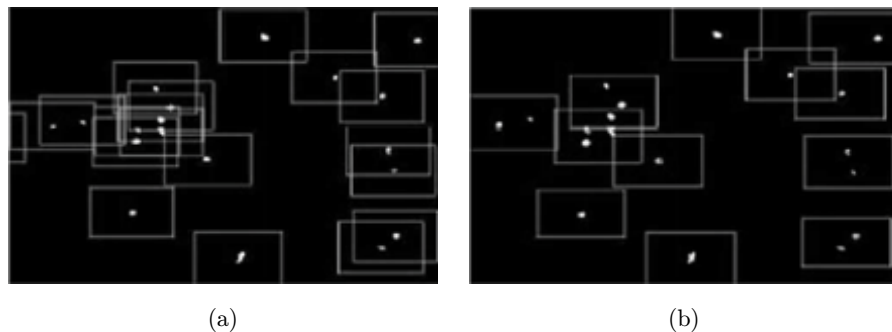


Fig. 4. Comparison of scanning routine before and after clustering. (a) Scanning routine composed of leukocyte nucleus' centroids and (b) Scanning routine composed of clustering centers.

captured. But because the high power objective's field of view is much bigger than the size of any leukocyte, some leukocytes may be captured several times when they are closer to one another. As shown in Fig. 4(a), the rectangular boxes represent the high power objectives' field of view, those leukocytes on the left-center and right-bottom corner may be captured several times. Clustering can reduce the total number of captured images and the number of redundantly captured leukocytes.

Compared with K -means clustering,^{20–22} max-min distance clustering method does not need to determine the number of initial clustering centers, which is more suitable for leukocytes clustering. The basic principles of max-min distance clustering are as follows:

Step 1. Suppose there are N leukocytes ($N = 19$ in Fig. 4(a)), their centroids are the set $S_n = \{Z_1, Z_2 \dots Z_n\}$.

Step 2. Randomly selecting an leukocyte, for example Z_1 , take Z_1 as the first clustering center; selecting the leukocyte which has the maximum distance to Z_1 from the set S_n as the next clustering center.

Step 3. For the remaining leukocytes Z_i in the set S_n , respectively calculating the distance to $Z_1, Z_2 \dots Z_j$ ($Z_1, Z_2 \dots Z_j$ are the established clustering centers), the minimum is signed as DZ_i .

Step 4. Calculating the $\max S_n \{DZ_i\}$, if $\max S_n \{DZ_i\} > K$ (K is the set threshold), then take the Z_i as the new clustering center.

Step 5. Repeating **Step 3** and **Step 4**, until we can not find a new clustering center which can satisfy the above constraints.

The superiority of max-min clustering in leukocyte image scanning is mainly embodied in capturing the regions that contain dense leukocytes.

Comparing Figs. 4(a) with Fig. 4(b), after clustering, the total number of captured images on the left-center with denser leukocytes distribution is reduced obviously, from 9 to 4 frames. Clustering can reduce both the total number of captured images and the proportion of redundantly captured leukocytes.

After max-min distance clustering on leukocytes in low power microscopic image, the number of clustering centers is much less than the number of leukocytes. As shown in Figs. 4(a) and 4(b), after clustering on leukocytes, the number of captured images has been reduced from 19 to 11, and there is only one redundantly captured leukocyte dash circled in Fig. 4(b). Thus, scanning along the routine composed of clustering centers can greatly reduce the number of captured images. Actually, the redundantly captured leukocytes need to be manually singled out and will lead to additional time consumption.

4. Experimental Results and Analysis

In order to verify the feasibility of the proposed method, lots of experiments using an improved OLYMPUS BX53 microscope which can realize leukocyte image auto-scanning have been carried out. Ten medical professionals have been invited helping us to complete the experiments. The analysis of experimental result is as follow:

4.1. Scanning routine design

Figure 5(a) is one of the captured low power microscopic images while scanning the blood smear along the zigzag routine. The total number of

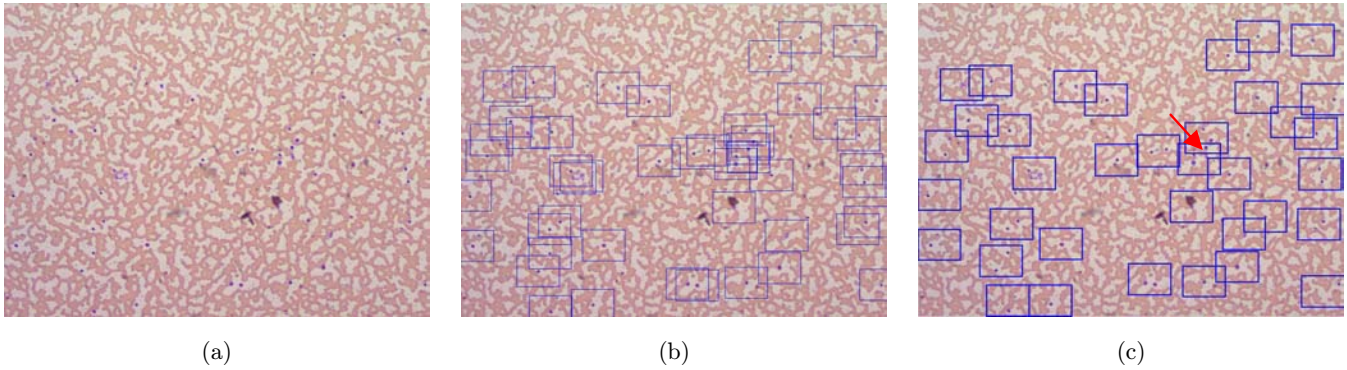


Fig. 5. Scanning routine design. (a) One of the captured low power microscopic image while scanning, (b) scanning routine composed of leukocyte nucleus' centroids and (c) Scanning routine composed of max-min distance clustering centers.

available leukocytes is 49 confirmed by the invited medical professionals. By using the above leukocyte segmentation and the nucleus' centroid location method, the scanning routine composed of nucleus' centroids is shown in Fig. 5(b). The 49 nucleus centroids are just figured out which shows the above leukocyte segmentation and the nucleus' centroid location method are validity. If we scan the blood smear along this routine with high power objective, because 19 leukocytes will be captured twice or even four times, it will take much more time to single out the 49 nonredundant leukocytes from 68 or even 106 captured leukocytes. After the max-min distance clustering on leukocyte nucleus centroids, the clustering result is shown in Fig. 5(c). There are 36 clustering centers in it. If we scan the blood smear along this routine, only 36 frames of leukocyte images need to be captured. It saves lots of scanning time. And as shown in Fig. 5(c), there is only one leukocyte captured twice in the region marked by the arrow.

4.2. Redundant capturing analysis

With the help of the 10 medical professionals, we obtain 110 individual leukocyte high power microscopic images. The result of directly using leukocyte nucleus' centroids (no clustering) as guidance of high power objective is shown in Fig. 6(a). As can be seen, there are massive leukocytes redundantly captured, among them, 16 leukocytes are captured twice, 4 leukocytes are captured three times and 2 leukocytes are captured four times. It contains only 80 nonredundant leukocytes in Fig. 6(a). Because of the massive redundant capture, to single out 100 nonredundant leukocytes for medical diagnosis we need about 140-170 individual leukocyte high power microscopic images using this method. The result of using max-min distance clustering is shown in Fig. 6(b). Only 2 leukocytes are redundantly captured. By this method, 110 individual leukocyte high power microscopic images can guarantee 100 nonredundant leukocytes which is the minimum amount of leukocytes for diagnosis.

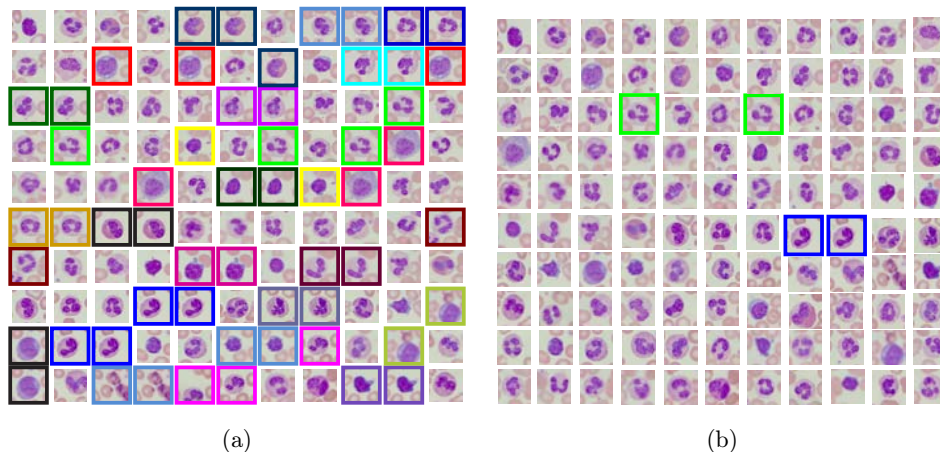


Fig. 6. Results of different methods. (a) Result of no clustering (b) Result of using max-min clustering

Table 1. One of the 10 medical professionals' time consumption recorded (unit: minute).

Number of blood smears	1	2	3	4	5	6	7	8	9	10
No clustering	13.7	13.9	8.9	8.0	9.5	10.1	9.4	12.5	9.4	10.7
Max-min clustering	2.9	3.7	3.2	3.3	4.0	3.5	4.0	2.6	3.8	2.6

4.3. Comparison of time consumption

There are five kinds of leukocytes with different functions in human peripheral blood. Blindly counting the total number of leukocytes is meaningless. It is more reasonable to use differential counting to make diagnosis of disease. The redundantly captured leukocytes will influence the differential counting and ultimately lead to misdiagnosis for patients. Thus the redundantly captured leukocytes need to be manually singled out. This work is time consuming.

To compare time consumption of different scanning routines, we record the time consumption for extracting 110 nonredundant individual leukocytes. Table 1 is the time consumption recorded for scanning 10 blood smears by one of the 10 invited medical professionals. The mean time consumption by the scanning routine composed of leukocyte nucleus centroids (No clustering) is 10.61 min. The mean time consumption by scanning routine composed of max-min distance clustering centers is 3.36 min, much shorter than that of the former.

Figure 7 shows the time consumption recorded for all the 100 blood smears tested by the 10 medical professionals. The time consumption by scanning routine composed of leukocyte nucleus' centroids

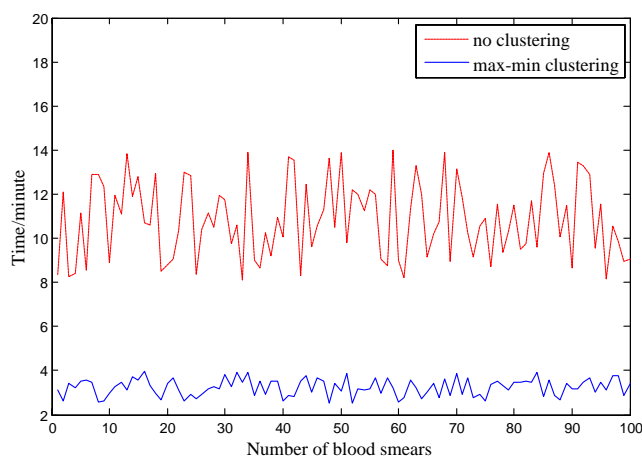


Fig. 7. Comparison of time consumption.

ranges from 8.0–14.0 min and presents an apparent volatility. This apparent volatility is mainly caused by different uneven leukocyte distributions in different blood smears. The further experimental analysis reveals that the denser the leukocyte distribution is, the more redundantly captured leukocytes there will be. The redundantly captured leukocytes need to be manually singled out and lead to additional time consumption. Time consumption of scanning the blood smear along the routine composed of max-min distance clustering centers is greatly shortened to 2.5–4.0 min and presents a little bit volatility, it is because the total number of captured images and the proportion of the redundantly captured leukocytes are both reduced.

The above experimental results show that the max-min clustering method can observably shorten the scanning time. And the total scanning time is less than 4.0 min for each blood smear to extracted 110 nonredundant leukocytes.

5. Conclusion

A leukocyte image fast scanning method based on max-min distance clustering is proposed. Due to the low power objective's larger field of view, the captured low power blood smear images during scanning can be used to locate leukocytes and guide the leukocyte image scanning on purpose with high power objective. By establishing the max-min distance clustering model, the total number of captured images can be significantly reduced. The experimental results show the feasibility of the proposed method. It takes only less than 4.0 min to extract 110 nonredundantly leukocytes in a blood smear to be tested. That means more than 15 pieces of blood smear can be tested in an hour. This will increase efficiency of blood smear test apparently.

Acknowledgments

This work was supported by the 863 National Plan Foundation of China under Grant No.

2007AA01Z333 and Special Grand National Project of China under Grant No. 2009ZX02204-008.

References

1. R. W. Cain, A. C. Anderson, "Parametric and non-parametric analysis for leukocyte classification," *Proc. Southeastcon'77 on Imaginative Engineering thru Education and Experience*, Proc. SOUTH-EASTCON Reg. 3 Conf. Williamsburg, VA, USA, pp. 111–15 (1977).
2. J. R. Cheng, J. C. Rajapakse, "Segmentation of clustered nuclei with shape markers and marking function," *IEEE Trans. Biomed. Eng.* **56**(3), 741–748 (2009).
3. Y. Wang, Z. Zhang, "Segmentation of the clustered cells with optimized boundary detection in negative phase contrast images," *PLOS One* **10**(6), e0130178 (2015).
4. Y. Al-Kofahi, W. Lassoued, B. Roysam, "Improved automatic detection and segmentation of cell nuclei in historoutineology images," *IEEE Trans. Biomed. Eng.* **57**(4), 841–52 (2010).
5. Y. Zheng, B. Jeon, "Image segmentation by generalized hierarchical fuzzy C-means algorithm," *J. Intell. Fuzzy Syst.* **28**(2), 961–973 (2015).
6. S. Srivastava, N. Sharma, "A combined approach for the enhancement and segmentation of mammograms using modified fuzzy C-means method in wavelet domain," *J. Med. Phys.* **39**(3), 169–183 (2014).
7. C. Xiaohao, "Variational image segmentation model coupled with image restoration achievements," *Pattern Recog.* **48**(6), 2029–2042 (2015).
8. N. Mitianoudis, N. Papamarkos, "Document image binarization using local features and Gaussian mixture modeling," *Image Vision Comput.* **38**, 33–51 (2015).
9. D. M. Hancock, "Improved dynamic thresholding algorithm for image binarization," *IBM Tech. Discl. Bull.* **21**(6), 2295–2297 (1958).
10. F. Saitoh, "Integrated local image binarization using genetic algorithm," *Trans. Inst. Electr. Eng. Jpn. C* **119**-C(8–9), 942–947 (1999).
11. Y. Yoshii, K. Sasaki, "Cluster analysis for the probability of DSB site induced by electron tracks," *Nucl. Instrum. Methods Phys. Res. B, BEAM Interact. Mater. At.* **350**, 55–59 (2015).
12. S. Seaman, M. Pavlou, "Review of methods for handling confounding by cluster and informative cluster size in clustered data," *Stat. Med.* **33**(30), 5371–5387 (2014).
13. A. A. Kumar, S. Sriparna, "A min–max distance based external cluster validity index: MMI," *12th Int. Conf. Hybrid Intelligent Systems (HIS)*, Pune, India, pp. 354–359 (2012).
14. Mohd. Wan Maseri Binti Wan, A. H. Beg, "An improved parameter less data clustering technique based on maximum distance of data and lioyd K-means algorithm," *1st World Conf. Innovation and Software Development (INSODE)*, Bahcesehir Univ, Istanbul, Turkey **1**, 367–371 (2012).
15. Q. Y. Xie, Y. Cheng, "K-centers min-max clustering algorithm over heterogeneous wireless sensor networks," *2013 Wireless Telecommunications Symp. (WTS 2013)*, Phoenix, AZ, USA, pp. 142–7 (2013).
16. M. Zeng, T. M. Han, Q. H. Meng, Image thresholding based on edge information analysis, *2012 5th Int. Congress on Image and Signal Processing (CISP)*, Chongqing, China, pp. 1371–1375 (2012).
17. H. Chia-Shan, R. Shanq-Jang, "Adaptive thresholding algorithm: Efficient computation technique based on 2-D intelligent block detection for degraded images," *IEICE Trans. Fund. Electron. Commun. Comput. Sci.* **E97**-A(2), 717–18 (2014).
18. N. Otsu, "A threshold selection method from gray level histograms," *IEEE Trans. Syst. Man Cybern.* **9**(1), 62–66 (1979).
19. X. Yuan, L. Wu, "An improved Otsu method using the weighted object variance for defect detection," *Appl. Surf. Sci.* **349**, 472–484 (2015).
20. X. Huang, W. Su, "An improved k-means clustering algorithm," *J. Netw.* **9**(1), 161–167 (2014).
21. N. K. Visalakshi, J. Suguna, "K-means clustering using max-min distance measure," *2009 Annual Meeting of the North American Fuzzy Information Processing Society*, Cincinnati, OH, United States, 74–79 (2009).
22. M. E. Celebi, H. A. Kingravi, "A comparative study of efficient initialization methods for the K-means clustering algorithm," *Expert Syst. Appl.* **40**(1), 200–210 (2013).