

# Automatic counting method for complex overlapping erythrocytes based on seed prediction in microscopic imaging

Xudong Wei and Yiping Cao\*

*Department of Optical Electronics, Sichuan University  
Chengdu, Sichuan 610064, P. R. China  
\*ypcao@scu.edu.cn*

Received 31 July 2015

Accepted 6 November 2015

Published 24 December 2015

Blood cell counting is an important medical test to help medical staffs diagnose various symptoms and diseases. An automatic segmentation of complex overlapping erythrocytes based on seed prediction in microscopic imaging is proposed. The four main innovations of this research are as follows: (1) Regions of erythrocytes extracted rapidly and accurately based on the G component. (2) K-means algorithm is applied on edge detection of overlapping erythrocytes. (3) Traces of erythrocytes' biconcave shape are utilized to predict erythrocyte's position in overlapping clusters. (4) A new automatic counting method which aims at complex overlapping erythrocytes is presented. The experimental results show that the proposed method is efficient and accurate with very little running time. The average accuracy of the proposed method reaches 97.0%.

*Keywords:* Image segmentation; erythrocyte; cell counting; k-means; seed prediction.

## 1. Introduction

Erythrocytes (also referred to Red Blood Cells, RBCs) are a crucial kind of blood cells in human blood due to their main devotions to blood gas transport, especially oxygen. RBCs are much more common than the other blood particles: White Blood Cells (WBCs) and thrombocytes (platelets, PLTs). For a healthy adult, about  $2.5 \times 10^{11}$  new RBCs are released from the bone marrow into the blood circulation per day. Such a large amount leads to a high RBC component

proportion of 97% in the blood and its mass density is equal to about  $4.5-6.2 \times 10^6/\text{mm}^3$  in men.<sup>1</sup> Mature RBCs are nonnucleated, biconcave disc-shaped cells of  $7-8 \mu\text{m}$  in diameter,  $2 \mu\text{m}$ -thick with an average volume of  $80-100 \mu\text{m}^3$ .<sup>2</sup> RBC counting is one of the most commonly performed medical tests as it plays a vital role in the diagnosis of various diseases.<sup>3</sup> Pathological alterations in RBC have been associated with various diseases<sup>4</sup> such as malaria,<sup>5</sup> sickle cell anemia, diabetes,<sup>6</sup> hereditary disorders<sup>7</sup> and so on.<sup>8</sup>

\*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.

Traditional counting process, when performed manually, has been proved to be tedious, time consuming and subjective. Some automatic and accurate mechanisms are invented to cope with the aforementioned problems such as automatic flow cytometry.<sup>9</sup> Various automatic, computer-aided blood cell counting techniques have been proposed in recent decades.<sup>10–16</sup> A large number of experiments show that most of the existed methods are untenable with a low precision in the case of cells overlapping together (also called clustering, some cells clustering together and forming into a big area) which often appears in actual blood smear images. To optimize this shortcoming, an automatic segmentation of complex overlapping RBCs based on seed prediction is presented. In this method, after regions of erythrocytes extracted rapidly and accurately based on the G component of original images, traces of erythrocytes' biconcave shape are utilized as seeds to predict erythrocyte's position in overlapping clusters. Besides, K-means algorithm is applied on internal edge detection of overlapping erythrocytes.

## 2. Related Works

Segmentation of the RBCs in a microscopic blood smear image is a potential research problem due to their medical significances. Researchers have been putting serious efforts for this purpose. Grishagin<sup>10</sup> presented a complete solution for automatic cell counting in which a conventional light microscope is equipped with a web camera to obtain images of a suspension of mammalian cells in a hemocytometer assembly. Based on the ImageJ toolbox, we devised two algorithms to automatically count these cells. On this basis of Thomasset's<sup>11</sup> level set, Lu<sup>12</sup> presented an improved algorithm for the segmentation of cytoplasm and nuclei from clumps of overlapping cervical cells by utilizing a joint optimization of multiple level set functions. Saima *et al.*<sup>13</sup> presented a parametrized segmentation algorithm called capture largest included circles (CLIC) that captures largest possible circles in an object boundary. Ge *et al.*<sup>14</sup> counted RBCs by extracting regional maxima of overlapping erythrocytes and defined a distance-dependent rule to eliminate those spurious maxima. An adaptive marker-controlled watershed approach, aimed at improving the automatic extraction of markers was proposed by Tonti *et al.*<sup>15</sup> It

takes advantage of domain-specific knowledge about the textural and geometrical characteristics of cells to reduce the sensitivity to uneven illumination and over-segmentation errors. Genctav *et al.*<sup>16</sup> segmented overlapping cells by ranking the cells based on their feature characteristics computed from the nuclei and cytoplasm regions. The ranking was generated via linearization of the leaves of a binary tree that was constructed using hierarchical clustering.

## 3. Proposed Complete RBCs Counting Method

### 3.1. Pre-processing

RGB color space is commonly used to describe color space of images.<sup>17</sup> Figure 1(a) is a typical cell color image and Fig. 1(b) shows the G component in RGB color space of this image. It can be easily found from Fig. 1(b) that there lies distinct difference between background and cells. The gray value of background is higher than that of the cells as shown in Fig. 1(c). While applying Otsu's method<sup>18</sup> to the G component, a gray value threshold  $T_0$  is acquired self-adaptively. By doing image binarization through  $T_0$  on Fig. 1(b), every cell can be figured out from background rapidly and accurately as shown in Fig. 1(d). After analyzing further, gray value in nucleus regions in G component are obviously lower than that of other regions as shown in Fig. 1(e). So a binary image which only contains nucleated parts through an empirical value  $T_1$  will reveal itself as shown in Fig. 1(f).

By image binarization on Fig. 1(b). With the guide of Fig. 1(f), regions of the nucleated cells can be wiped off from Fig. 1(d) as shown in Fig. 1(g) in which WBCs and PLTs are no longer existed but some noises, whose area is much smaller than the cell size, may invade due to the blemish on blood slides. By small area filtering, the binary image only contains RBCs is extracted successfully as shown in Fig. 1(h). Figure 2 shows the flowchart of the pre-processing.

Sequential steps of pre-processing are as follows:

**Step1.** Splitting the G component of the input image and extracting a self-adaptive  $T_0$  by Otsu's method on G component; Image binarization of the G component with  $T_0$ ; Hole-filling the binary image by morphological operations (see Fig. 1(d)).

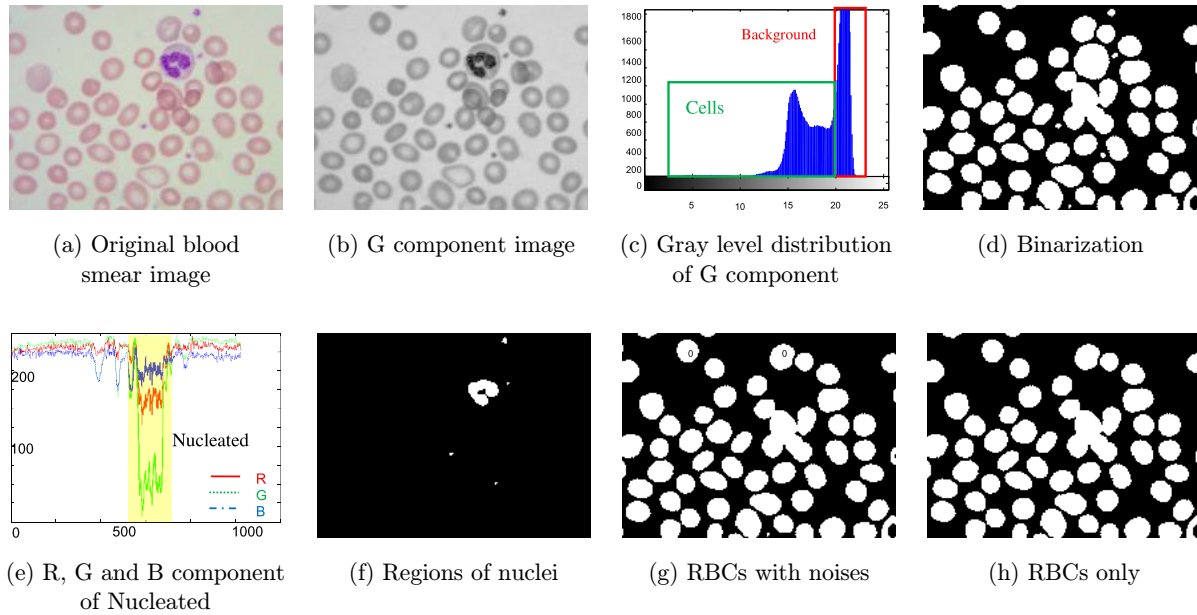


Fig. 1. Image pre-processing.

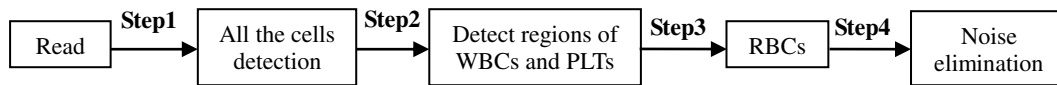


Fig. 2. The flowchart of pre-processing.

**Step2.** Image binarization of the G component with  $T_1$  (see Fig. 1(f)).

**Step3.** Reading the first pixel of each individual nucleus region in Fig. 1(f) and storing their coordinates; Wiping off closed regions who conations these coordinates in the Fig. 1(d).

**Step4.** Deleting regions whose area is too much small.

### 3.2. RBCs classification

Due to RBCs' large amount and high density in blood, distribution of RBCs is complex and varied

in blood smear images. In principle, RBCs can be classified into two classes: individual RBCs and overlapped RBCs as shown in Figs. 3(a) and 3(b)). Much more attention should be paid to the different shape characteristic between these two classes. It has been corroborated by repeated experiments that while taking count of RBCs with a well-worked method for individual RBCs, the accuracy of overlapped ones is too low to meet the application demand; on the other hand, while employing a complex algorithm which fares well in overlapping clusters, over-segmentation always occurs on individual RBCs and also these algorithms consume

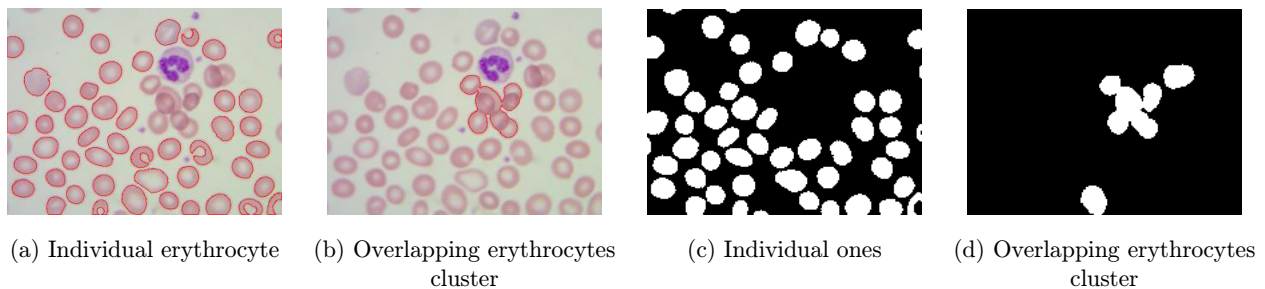


Fig. 3. Classification of erythrocytes.

much more time. In purpose of enhancing the precision, it is necessary to classify individual ones and overlapping clusters into different classes, respectively before taking count.

8249 erythrocytes in 100 blood images with  $1027 \times 768$  sizes at  $100\times$  were used to analyze the morphological characters of individuals and overlapping ones. Notice that the area sizes of the two classes are obviously different: area sizes of overlapping clusters are at least 1.48 times of individual RBCs. Individual RBCs contain 5275 pixels on average while 7858 in overlapping clusters. Another conspicuous difference between individual RBCs and clusters is the circularity index calculated using the following formula:

$$CI = 4\pi \frac{A}{P^2}, \quad (1)$$

where  $CI$ ,  $P$ , and  $A$  represent the circularity index, perimeter, and area, respectively. This results in a range of  $CI$  values between 0 and 1, with a  $CI$  of 1 indicating a perfect circle.<sup>19</sup> RBCs can be classified into two classes: individual RBCs and overlapping clusters, via both the two parameters mentioned above. Regions who meet both the following two measures would be judged as suspected overlapping clusters: (1) area is larger than the average size of individual RBCs  $T_A$ ; (2) circularity index is smaller than the average  $CI$  of individual RBCs  $T_{CI}$ . It is inescapable that a minority of individual RBCs whose area and circularity are both close to the thresholds and be classified into overlapping class, they were called suspected ones. Actually, only 43 individual ones in the 100 images were misclassified. Results of classification are shown in Figs. 3(c) and 3(d). Figure 4 shows the flowchart of the RBCs classification algorithm.

### 3.3. RBC segmentation

Different classes of RBCs will be counted, respectively. The number of individual RBCs  $N_i$ , can be calculated by setting a counter in the circulation in Fig. 4. In order to overcome the limitations due to the complex overlapping groups' incomplete shape and ambiguous edges, an algorithm based on K-means, seed location and seed prediction is proposed.

#### 3.3.1. Pixel clustering

By using the binary image Fig. 3(d) as a rectangular window function, the overlapping groups in the G

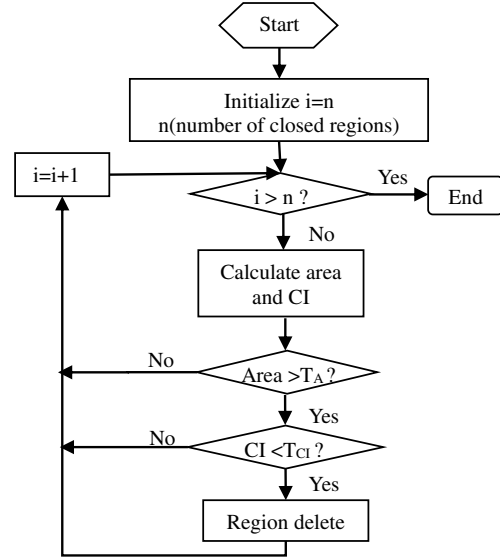


Fig. 4. The flowchart of the classify algorithm.

component can be extracted from Fig. 1(b) as shown in Fig. 5(a). When applying K-means cluster algorithm<sup>20–22</sup> on Fig. 5(a), pixels would be ranged into K clusters. Pixels belong to the same cluster have the same gray value, thus the edge will be much more distinct to be detected. Notice that background's gray value is 0 and a healthy human being's RBCs are biconcave shape structure, so the RBCs' sunken parts have higher gray value than other regions as shown in Fig. 5(b). It is reasonable to cluster pixels into three clusters by setting the parameter K to 3 in the K-means algorithm. Figures 5(c) and 5(e) shows the details.

#### 3.3.2. Edge detection and restoring

Canny edge detection algorithm is a classical and robust method for edge detection in gray-scale images.<sup>23</sup> It looks like a doughnut due to the biconcave shape of RBC, so the peripheral of the cell is called external edge while the internal edge corresponding to the trace of biconcave region (see Fig. 5(e)). After edge detection with canny edge detection method,<sup>24</sup> details of the internal edges of every RBC in the overlapping groups emerged to be legible (see Figs. 5(d) and 5(h)). However, some internal edges are not closed and these edges cannot help to locate a seed of RBCs accurately, so restoring of these edges is indispensable before the next manipulation. When the distance between two edges is closer than 5 length units (set one pixel as 1 length unit), these two lines will be deemed as different parts of one RBC's

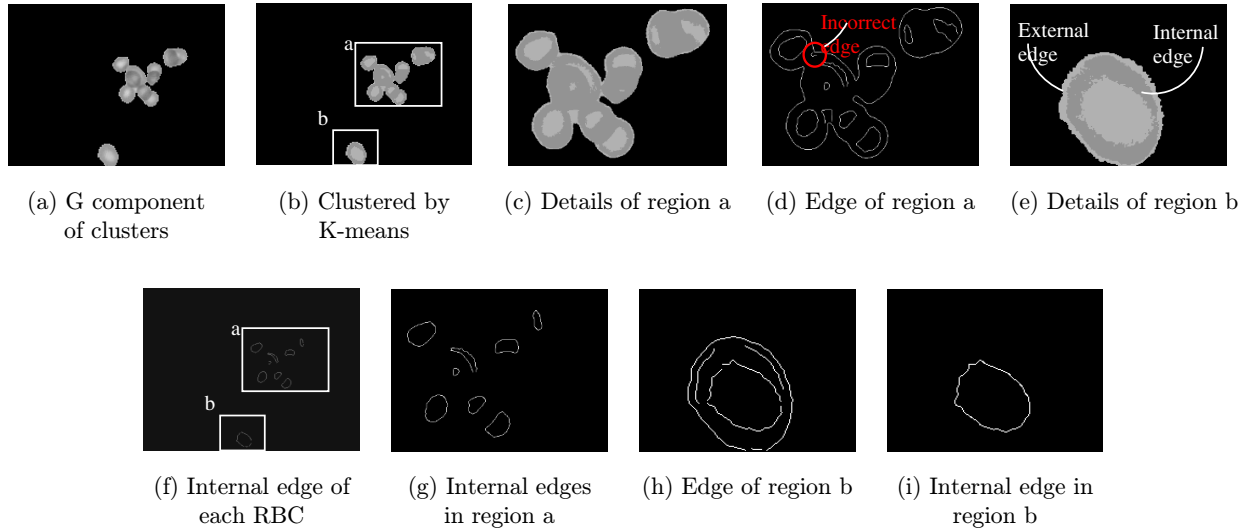


Fig. 5. Processing of overlapping cluster (1).

internal edge and then be linked. Another preponderance of restoring internal edges is that some exiguous incorrect edges as shown in Fig. 5(d) will be isolated and wiped off due to these incorrect edges cannot be linked as a part of internal edges. A perfect edge will come out after edge restoring as shown in Fig. 5(f). Restoring effect is obvious via the contrast between Figs. 5(h) and 5(i).

### 3.3.3. Prediction and segmentation

Internal edges and RBCs are one-for-one corresponding. Each centroid of internal edges can be used as a seed to predict the location of a corresponding RBC. A fitting circle on the basis of each internal edge can be used to simulate a corresponding RBC. With the centroids and circles, the original overlapping situation in the clusters will be predicted as close as possible (see Fig. 6(a)). As a matter of fact, these simulated circles cannot reflect overlapping clusters objectively and exactly because

the radiuses are modulated by internal edges, which are smaller than RBCs' actual sizes. So a further step should be taken to improve seed prediction. Considering that the shape of an individual RBC is very close to a circle and the average radius of RBCs  $R_p$  can be calculated from individual RBCs, so we can draw fictitious circles with each seed and  $R_p$ , respectively as seed prediction (see Fig. 6(b)). By taking an intersection for Figs. 3(d) and 6(b), a more realistic prediction of clusters will exhibit as shown in Fig. 6(c). It is worth mentioning that some few RBCs would be missed in the edge detection due to their inconspicuous internal edge. Detection of these dropped ones will accomplish by subtracting Figs. 3(d) and 6(c) as shown in Fig. 6(d).

### 3.4. RBC counting

According to the above discussion, the number of overlapping RBCs contains the number of seeds  $N_s$  and the number of close regions in Fig. 6(d)  $N_c$ .

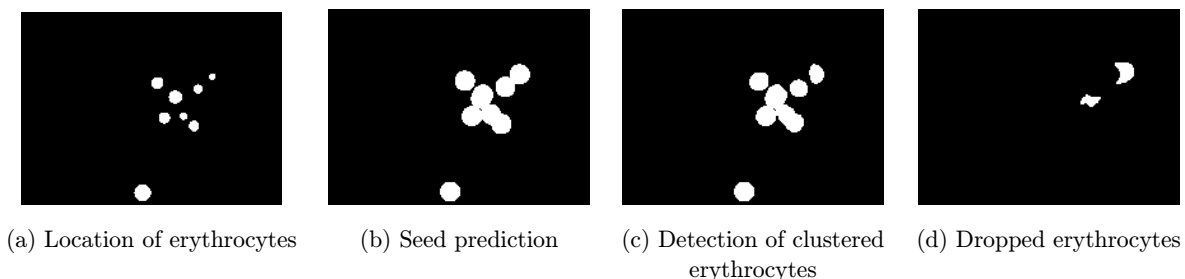


Fig. 6. Processing of overlapping clusters (2).



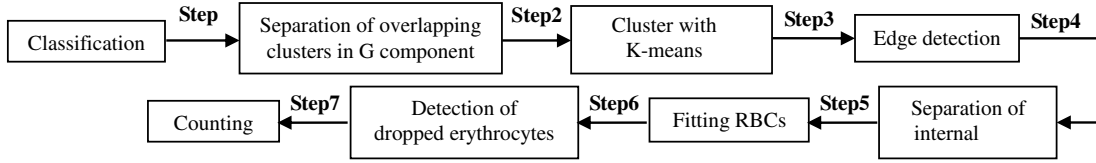


Fig. 7. The flowchart of counting.

Hence, the total number of RBCs  $N_R$  in the smear microscope image can be expressed as follows:

$$N_R = N_i + N_s + N_c, \quad (2)$$

Figure 7 shows the flowchart of the counting process.

This part of algorithm to complete RBC counting can be described step by step as follows:

- Step1.** Applying binary image Fig. 3(d) on the G component of original image Fig. 1(b) as a rectangular window function (see Fig. 5(a)).
- Step2.** Clustering with K-mean where K is set as 3 (see Fig. 5(b)).
- Step3.** Edge detection with canny edge detector (see Figs. 5(d) and 5(h)).
- Step4.** Internal edges restoring (see Figs. 5(g) and 5(i)).
- Step5.** Calculating centroids on the basis of internal edges; Calculating average radius  $R_p$  on the basis of individual RBCs; Drawing simulated circles (see Fig. 6(b)).
- Step6.** ‘IMAGE INTERSECTION’ between Figs. 3(d) and 6(b). ‘IMAGE COMPLEMENT’ between Figs. 3(d) and 6(c).
- Step7.** Summation of  $N_i$ ,  $N_s$  and  $N_c$ .

#### 4. Experiments and Result Analysis

Lots of microscopic images with 1024\*768 sizes were used to test and verify the universality and robustness of this proposed method. All the test images with 1027\*768 sizes were captured by a CCD camera at 100× and they were on-the-spot sampler collected from Sichuan Provincial People’s Hospital. A dataset consist of 100 test images was divided into 10 sets and the experimental results are shown in Table 1 to prove the performance of this proposed method.  $M_T$  and  $M_O$  represent manual counting result of total RBCs and manual counting result of overlapping RBCs by a hematologist, respectively.  $N_T$  is this proposed method’s counting result of total RBCs in the image while the data  $N_O$

Table 1. Experiment results of the dataset.

| Set   | Total |       | Overlapping |       | Acc of Total (%) | Acc of Clusters (%) |
|-------|-------|-------|-------------|-------|------------------|---------------------|
|       | $M_T$ | $N_T$ | $M_O$       | $N_O$ |                  |                     |
| 1     | 674   | 674   | 197         | 197   | 100              | 100                 |
| 2     | 785   | 784   | 217         | 216   | 99.9             | 99.5                |
| 3     | 738   | 721   | 251         | 234   | 97.7             | 93.2                |
| 4     | 876   | 847   | 288         | 259   | 96.7             | 89.9                |
| 5     | 901   | 871   | 326         | 296   | 96.7             | 90.8                |
| 6     | 955   | 916   | 352         | 313   | 95.9             | 88.9                |
| 7     | 693   | 693   | 225         | 225   | 100              | 100                 |
| 8     | 774   | 769   | 259         | 254   | 99.4             | 98.1                |
| 9     | 837   | 792   | 316         | 271   | 94.6             | 85.8                |
| 10    | 1016  | 938   | 374         | 296   | 92.3             | 79.1                |
| Total | 8249  | 8005  | 2805        | 2561  | 97.0             | 91.3                |

represents the counting result of overlapping clusters.

Three representative samples (i, ii and iii) were chosen to analyze the experimental results as shown in Fig. 8. Experiments in different cases obtained good results and the experiment results are shown in Table 2.

*Precision and accuracy:* Although the average precision reaches 97%, the weakness of this method should not be neglected: to a certain degree, the precision declines with the increase of the amount of RBC. In some excessive overlapping situation, few RBCs’ internal edges were impossible to be detected and they would also be covered during the seed prediction. This leads to a certain imperfection.

*Parameter selection:* In pre-processing,  $T_1$  was not adaptive. It was set to 100 via a large amount of experiments. In classification,  $T_A$  and  $T_{CI}$  are calculated on the basis of individual RBCs. Also, RBCs’ average radius  $R_p$  in seed prediction can be calculated.

*Restrictions:* As long as RBCs are biconcave shape, their internal edges can be obtained. Even irregularly shaped RBCs can be captured by the proposed method. Unfortunately, this proposed method cannot do well in taking count of hypotonic

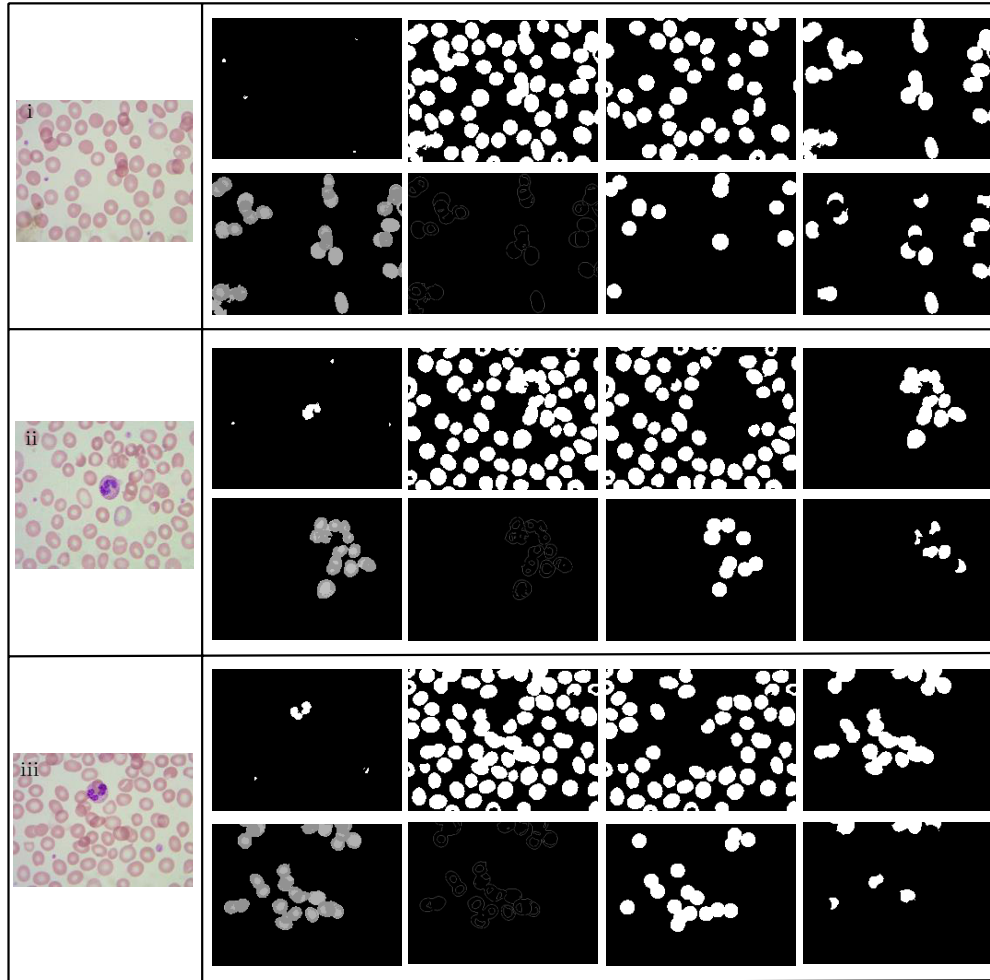


Fig. 8. Three representative experimental results.

Table 2. Cells counting using proposed algorithm.

| Image   | Total |       | Overlapping |       | Acc of Total (%) | Acc of Clusters (%) |
|---------|-------|-------|-------------|-------|------------------|---------------------|
|         | $M_T$ | $N_T$ | $M_O$       | $N_O$ |                  |                     |
| i       | 68    | 68    | 25          | 25    | 100              | 100                 |
| ii      | 73    | 73    | 14          | 14    | 100              | 100                 |
| iii     | 85    | 78    | 32          | 25    | 91.8             | 78.1                |
| Average |       |       |             |       | 96.9             | 90.1                |

or hypertonic cases RBCs, due to their smooth surface.

*Runtime:* Using nonoptimized computer program code on standard 2. GHz CPU, the average running time of this method was 31.04 s/image.

A comparison with other methods dealing with the above three representative images and the dataset is shown in Table 3.

Table 3. Comparison with other algorithms.

| Image/Method       | i         |           | ii        |           | iii       |           | Dataset   |           | Ave Acc of total (%) | Ave Acc of clusters (%) |
|--------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|----------------------|-------------------------|
|                    | $M_T/N_T$ | $M_O/N_O$ | $M_T/N_T$ | $M_O/N_O$ | $M_T/N_T$ | $M_O/N_O$ | $M_T/N_T$ | $M_O/N_O$ |                      |                         |
| Grishagin's method | 68/56     | 25/14     | 73/66     | 14/7      | 85/64     | 32/11     | 8249/6788 | 2805/1266 | 82.3                 | 45.1                    |
| Ge's approach      | 68/52     | 25/19     | 73/69     | 14/10     | 85/74     | 32/21     | 8249/7111 | 2805/1975 | 86.2                 | 70.4                    |
| CLIC               | 68/58     | 25/15     | 73/66     | 14/7      | 85/63     | 32/10     | 8249/6823 | 2805/1265 | 82.7                 | 45.1                    |
| Proposed           | 68/68     | 25/25     | 73/73     | 14/14     | 85/78     | 32/25     | 8249/8005 | 2805/2561 | 97.0                 | 90.1                    |

Comparing with the existing algorithms, the accuracy of total counting results increases 10% to 15% and the accuracy of clusters counting results rises almost 20% to 35%. Results reveal that the proposed scheme method outperforms traditional image segmentation algorithms by less calculation, less time-consumption and higher validity. It is evident that one of the major advantages of this new measure of RBCs counting is that it can detect overlapping clusters efficiently, which is reliable in improving the counting precision.

## 5. Conclusions

There is an increasing need for expanding the application of complete RBC counting method based on microscope smear blood image in computer background in medical domain. Significant progress has been made in blood cells counting in a remarkably short period of time. This study reveals the following three main findings: Firstly, regions of erythrocytes can be extracted rapidly and accurately from the smear image based on the G component. Secondly, an automatic segmentation of complex overlapping RBCs is proposed. Thirdly, a complete RBC counting method is presented. The experimental results have proven the superiority of this proposed scheme especially in case of complex overlapping RBC clusters. But there lies a reduction of the accuracy while RBCs in the smear image have no conspicuous internal edges. In future, adaptive threshold may be introduced to increase the accuracy and effectiveness of this approach.

## 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. M. Thiriet, Tissue Functioning and Remodeling in the Circulatory and Ventilatory Systems, *Biomathematical and biomechanical modeling of the circulatory and ventilatory systems* **5**, pp. 53–175 (2013).
2. J. Picot, P. A. Ndour, “A biomimetic microfluidic chip to study the circulation and mechanical retention of red blood cells in the spleen,” *Am. J. Hematol.* **90**(4), 339–345 (2015).
3. <http://www.nlm.nih.gov/medlineplus/ency/article/003644.htm> (site last visited on July 30, 2015).
4. S. Shattil, B. Furie, H. Cohen, L. Silverstein, P. Glave, M. Strauss, *Hematology: Basic Principles and Practice*, Churchill Livingstone, Philadelphia, (2000).
5. T. Wu, J. J. Feng, “Simulation of malaria-infected red blood cells in microfluidic channels: Passage and blockage,” *Biomicrofluidics* **7**(4), 044115 (2013).
6. A. V. Buys, E. Pretorius, “Changes in red blood cell membrane structure in type 2 diabetes: A scanning electron and atomic force microscopy study,” *Cardiovasc. Diabetol.* **12**, 25 (2013).
7. L. Da Costa, J. Galimand, O. Fenneteau, N. Mohandas, “Hereditary spherocytosis, elliptocytosis, and other red cell membrane disorders,” *Blood Rev.* **27**(4), 167–178 (2013).
8. G. Tomaiuolo, “Biomechanical properties of red blood cells in health and disease towards microfluidics,” *Biomicrofluidics* **8**, 051501 (2014).
9. Y. H. Lin, G. B. Lee, “Optically induced flow cytometry for continuous microparticle counting and sorting,” *Biosens. Bioelectron.* **24**, 572–578 (2008).
10. I. V. Grishagin, “Automatic cell counting with image,” *J. Anal. Biochem.* **473**, 63–65 (2015).
11. A. Dervieux, F. Thomasset, “A finite element method for the simulation of Rayleigh-Taylor instability,” *Lec. Notes Math.* **771**, 145–158 (1980).
12. Z. Lu, “An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells,” *IEEE Trans. Image Process.* **24**(4) 1261–1272 (2015).
13. S. Rathore, A. Iftikhar, A. Ali, M. Hussain, A. Jalil, “Capture largest included Circles: An approach for counting red blood cells,” *IMTIC 2012, CCIS* **281**, 373–384 (2012).
14. J. Ge, Z. Gong, Y. Sun, “A system for counting fetal and maternal red blood cells,” *IEEE Trans. Biomed. Eng.* **61**(12) 2823–2829 (2014).
15. S. Tonti, S. Cataldo, A. Bottino, E. Ficarra, An automated approach to the segmentation of HEP-2 cells for the indirect immunofluorescence ANA test, *Comput. Med. Imag. Graph.* **40**, 62–69 (2015).
16. A. Gencava, S. Aksoy, S. Onderb, “Unsupervised segmentation and classification of cervical cell images,” *Pattern Recognit.* **45**, 4151–4168 (2012).
17. Y. Yang, Y. Cao, W. Shi, “A method of leukocyte segmentation based on S component and B component images,” *J. Innov. Opt. Health Sci.* **7**(1) 1450007: 1–8 (2014).
18. N. Otsu, “A threshold selection method from gray level histograms,” *IEEE Trans. Syst. Man Cyber.* **9**(1) 62–66 (1979).
19. S. Nakano, T. Nakano, “Change in circularity index of cell lumen in a cross-section of wood induced by aqueous NaOH,” *J. Wood Sci.* **60**, 99–104 (2014).



20. L. Galluccio, "Graph based k-means clustering," *Signal Process.* **92**, 1970–1984 (2012).
21. D. MacKay, Chapter 20. An Example Inference Task: Clustering, *Information Theory, Inference and Learning Algorithms*, pp. 284–292. (Cambridge University Press, UK, 2012).
22. M. Mahajan, P. Nimbhorkar "The planar k-means problem is NP-hard," *Lect. Notes Comput. Sci.* **5431**, 274–285 (2009).
23. R. Biswas, J. Sil, "An improved canny edge detection algorithm based on type-2 fuzzy sets," *Procedia Technol.* **4**, 820–824 (2012).
24. H. Zhang, J. Jackman, "Feasibility of automatic detection of surface cracks in wind turbine blades," *Wind Eng.* **38**(6) 575–586 (2014).