

Article

White Blood Cell Segmentation by Color-Space-Based K-Means Clustering

Congcong Zhang ¹, Xiaoyan Xiao ², Xiaomei Li ³, Ying-Jie Chen ³, Wu Zhen ³, Jun Chang ¹, Chengyun Zheng ^{4,*} and Zhi Liu ^{1,*}

¹ School of Information Science and Engineering, Shandong University, Jinan 250100, China; E-Mails: zhangcongcong@mail.sdu.edu.cn (C.Zha.); changjun@sdu.edu.cn (J.C.)

² Department of nephrology, Qilu Hospital of Shandong University, Jinan 250012, China; E-Mail: xiaoyanxiao2007@163.com

³ Department of Oncology, the Second Hospital of Shandong University, Jinan 250100, China; E-Mails: sdulixiaomei@163.com (X.L.); lz505@163.com (Y.-J.C.); qinghuanjn@gmail.com (W.Z.)

⁴ Department of Hematology, the Second Hospital of Shandong University, Jinan 250100, China

* Authors to whom correspondence should be addressed;

E-Mails: zhengzhengwei@hotmail.com (C.Zhe.); liuzhi@sdu.edu.cn (L.Z.).

Received: 12 May 2014; in revised form: 10 July 2014 / Accepted: 24 July 2014 /

Published: 1 September 2014

Abstract: White blood cell (WBC) segmentation, which is important for cytometry, is a challenging issue because of the morphological diversity of WBCs and the complex and uncertain background of blood smear images. This paper proposes a novel method for the nucleus and cytoplasm segmentation of WBCs for cytometry. A color adjustment step was also introduced before segmentation. Color space decomposition and k-means clustering were combined for segmentation. A database including 300 microscopic blood smear images were used to evaluate the performance of our method. The proposed segmentation method achieves 95.7% and 91.3% overall accuracy for nucleus segmentation and cytoplasm segmentation, respectively. Experimental results demonstrate that the proposed method can segment WBCs effectively with high accuracy.

Keywords: white blood cell; segmentation; color space decomposition; k-means clusters

1. Introduction

The immune system, which is the third line of defense of the human body, protects the body from viruses, bacteria, and pathogens. This natural defense identifies and eliminates abnormal cells, such as tumor cells. The immune system consists of immune organs, immune cells, and immune molecules. White blood cells (WBCs) are the principal components of immune cells and play an important role in our body's immunity.

In fact, WBCs normally have a constant concentration in the human blood. If the amount of WBC exceeds the normal range, then health problems may occur. The morphological analysis of WBCs is one of the basic steps of blood pathological analysis. Morphological analysis is traditionally performed manually, thus making this a tedious and time consuming process, even for an expert. Furthermore, morphological analysis is limited to the professional knowledge and eyesight of a pathologist. Automatic methods have been developed in recent years to overcome the limitations of manual methods, which can cause calculation inaccuracies. Automatic devices use the principle of the light scattering method to calculate red blood cells (RBCs) and WBCs [1]. The counting procedure of the light scattering method is invisible and can only achieve routine classification, merely counting the number of RBCs and WBCs. Actually, WBCs can be divided into five types, namely, basophils, eosinophils, neutrophils, lymphocytes, and monocytes [2,3]. The first three types are granular types, and the last two types are non-granular. Distinguishing these types by using the light scattering method proved unsuccessful in this study. Thus, we used image analysis, which can provide human-like assessments, to overcome this problem. Image analysis identifies abnormal cells and adopts the diagnosis experience summarized by pathologists. This method uses the high resolution and flexible extracted features of computer vision to improve work efficiency and accuracy. Researchers are currently focused on developing a system that can automatically identify WBCs by using blood images. However, accurate WBC segmentation before classification still poses challenges.

WBC segmentation is an important part of the WBC classification system, and the segmentation results directly affect the accuracy of cell recognition. Researchers have proposed several useful methods for achieving accurate segmentation results. Huang *et al.* [4–7] divided WBC segmentation into two parts: one for nucleus segmentation and the other for cytoplasm segmentation. The nuclei of WBCs are the areas of high contrast and are relatively easy to segment in blood smear image. Thus, most of the traditional methods [4–7] first segment the nucleus part of the image before the cytoplasm part.

Considerable research has been done in this field in recent years. All these methods can be divided into five categories: threshold-based methods, learning-based methods, active-contour-based methods, metaheuristic-based methods, and saliency-based methods [8]. Threshold-based methods include the region growing method, watershed method, Otsu's method, and their combination. Duan *et al.* [6] segmented the nucleus and then applied the region growth of color information according to certain rules to extract the cytoplasm. Putzu *et al.* [9] proposed a method based on the thresholding of the cyan, magenta, yellow, key plate (CMYK) color space. Threshold-based methods have high running speed and reliable performance for unified images. However, the performance of threshold-based methods is inconsistent for non-unified images. Learning-based methods include supervised methods, including support vector machine (SVM), artificial neural network, and unsupervised methods such as

k-means clustering and Fuzzy c-means. Zheng *et al.* combined the expectation-maximization clustering and SVM to segment the WBCs [10]. The performance of learning-based methods mainly depends on the selection of color feature vectors. Active-contour-based methods include level sets and snakes. Zamani *et al.* [7] segmented the nucleus first, applied a gradient vector flow (GVF) snake to segment WBCs, and used the boundary of the nucleus as the initial contour of the snake. Ko *et al.* [11] proposed a WBC segmentation method by using stepwise merging rules based on mean-shift clustering and boundary removal rules with GVF snake. Sadeghian *et al.* [12] presented a WBC segmentation framework that segments the cytoplasm by using GVF snake. This framework identifies the cytoplasm by applying Zack thresholding [13] to a gray image with the nucleus region removed. Active-contour-based methods are limited to the selection of initial contour, and overlapping objects will lead to poor performance of the methods. Metaheuristic-based methods include genetic algorithm (GA) and differential evolution, among others. Osowski *et al.* [14] presented a method that combines SVM and GA to recognize blood cells and used GA to select the features used by SVM. Metaheuristic-based methods regard segmentation problem as optimization problem, but the application of metaheuristic-based methods in WBC segmentation is very limited [8]. Saliency-based methods are based on visual saliency attention. Zheng *et al.* localized and segmented WBCs based on visual saliency attention [15]. Pan *et al.* realized WBC segmentation based on simulated visual attention, and gained satisfactory results [16]. Saliency-based methods are effective in the segmentation of WBCs. For some particular and practical applications, however, there is still some work to do. Some methods combine more than one of the categories listed above and achieve better results, such as combination of learning-based method and visual-attention-based method [17] and combination of learning-based method and thresholding-based method [18]. Other segmentation methods for WBC aside from the above five categories are available [19,20].

Saraswat *et al.* surveyed the current situation of WBC segmentation and indicated that the results of WBC segmentation are still not acceptable and need to be improved. They pointed out that the presence of artifacts, shape variations in the WBCs, and overlapped cells are the major problems and must be focused on [8]. Thus, aiming at the first two major problems, we present a novel, effective technique with good performance and robustness for WBC segmentation.

2. Technology Background

2.1. Color Transfer

Blood smear images are sampled with microscopic imaging systems. Given the different lighting conditions, varied dyeing time, unstable smear thickness, and different physical qualities of subjects, the color of blood smear image is changed in practice. Most of the published methods only work in databases constructed under strictly controlled conditions. By contrast, in practical application, the blood images to be detected do not always have good color consistency. Therefore, color adjustment for blood smear images is needed. This paper proposed a color adjustment method based on color transfer and k-means clustering, and this method will be discussed in the following sections. The color transfer method will also be introduced.

In 2001, Shirley *et al.* proposed the color transfer method between images [21]. This method is a general color correction method. The main idea of the method is to bring the color characters of one image to another image according to their mean value and standard deviation in Lab color model. In this paper, a standard template image that can be segmented accurately is chosen. The remaining images were then transformed according to the color characters of the template image. The algorithm can be elaborated into the following steps:

1. Transform the RGB color space of the input image and the template image to Lab color space.
2. Compute the mean value and standard deviation of the input image and the template image in Lab color space.
3. Subtract the mean value from all the pixels according to the following equation:

$$\begin{cases} l' = l - \bar{l} \\ a' = a - \bar{a} \\ b' = b - \bar{b} \end{cases} \quad (1)$$

where l, a, b denote the L, a, b components of each pixel; $\bar{l}, \bar{a}, \bar{b}$ denote the mean value of the L, a, b components; l', a', b' denote the resultant value of each pixel after subtraction.

4. Scale the pixel points of each color component of the synthetic image by factors determined by the ratio of respective standard deviations:

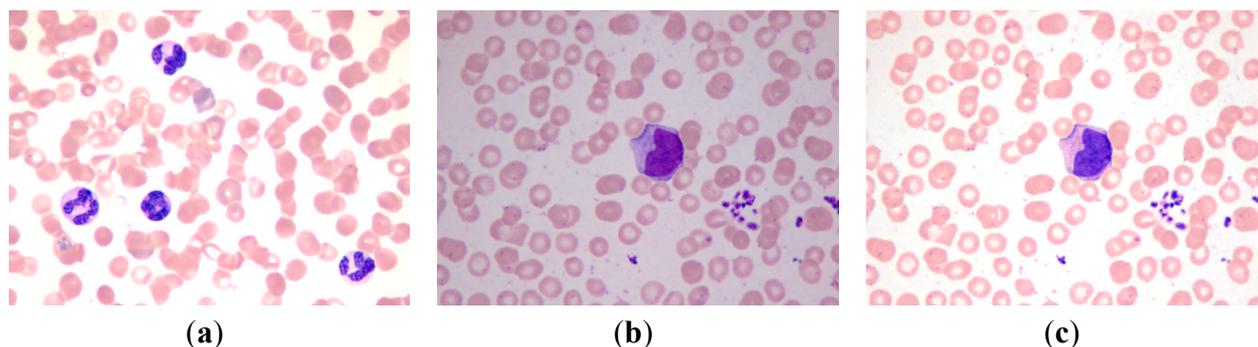
$$\begin{cases} l'' = \frac{\sigma_t^l}{\sigma_s^l} l' \\ a'' = \frac{\sigma_t^a}{\sigma_s^a} a' \\ b'' = \frac{\sigma_t^b}{\sigma_s^b} b' \end{cases} \quad (2)$$

where $\sigma_t^l, \sigma_t^a, \sigma_t^b$ denote the standard deviation of the L, a, b components of the template image; $\sigma_s^l, \sigma_s^a, \sigma_s^b$ denote the standard deviation of the L, a, b components of the input image; l'', a'', b'' denote the resultant value of each pixel.

5. Transform the Lab color space of the resultant image to RGB color space.

Through the steps above, the color characters of the template image are transferred to the input image, and the synthetic image possesses the color features that are beneficial to WBC segmentation. The template image, the input image and the synthetic image based on the color transfer method are shown in Figure 1.

Figure 1. (a) Template image; (b) Input image; (c) Synthetic image.

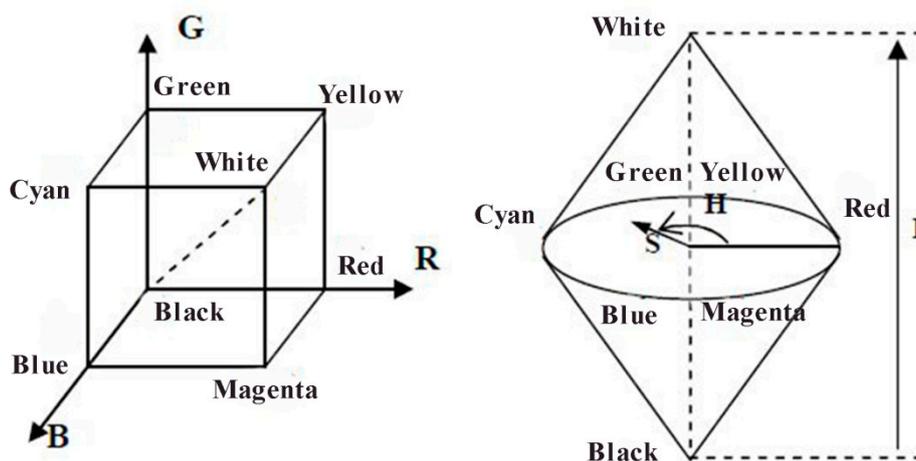


2.2. Different Color Spaces

The purpose of a color space (also called color model or color system) is to facilitate the specification of colors in some standard [22]. A color space is represented by matrix in mathematics, and typically, the matrix consists of three or four dimensions (e.g., RGB, HSI, and CMYK). The RGB color space is the most common color space used in electronic devices, such as cell phones and monitors. In the RGB color space, each color can be acquired by the addition of three primary colors: red, green, and blue. Therefore, this color space is an additive color space. The original stained blood smear image is represented by the RGB color space in the RGB model.

The HSI color space reflects the human vision system, where H denotes hue, S denotes saturation, and I denotes intensity. The I component is independent of color information, and the H and S components are closely related to the way humans perceive color. These characteristics make the HSI color space suitable for color detection and analysis. The schematic of the RGB and HSI color models is shown in Figure 2.

Figure 2. Schematic of the RGB color model (Left) and HSI color model (Right).



The CMYK color model is a subtractive color model and it is used in color printing. CMYK refers to the four inks used in color printing: cyan, magenta, yellow, and black. In the CMYK color space, C denotes cyan, M denotes magenta, Y denotes yellow, and K refers to the key plate (black).

The Lab color model is based on the way human perceive color, and it can describe all the color that human can see. This model minimizes the correlation between different color components. Color management system utilizes Lab as color patch to color transform one color space to another color space. Every color model has its own specific features, and the original stained blood smear image in the RGB color space shows useful features that benefit WBC segmentation when translated to HSI and CMYK color spaces. The details are discussed in the following section.

2.3. K-Means Clustering

K-means algorithm is a typical clustering algorithm based on distance. This algorithm adopts distance as the evaluation parameter of similarity, which means the shorter the distance is the more

similar the two objects are. K-means clustering aims to partition n observations into k clusters. Each observation in k-means clustering belongs to the cluster with the nearest mean, which serves as the prototype of the cluster. We take advantage of this feature to partition $M \times N$ pixels in the stained blood smear image into several clusters on the basis of the color information of the background, RBCs, cytoplasm, and white cell nuclei. M denotes the number of rows in the image, and N denotes the number of columns in the image.

In the clustering problem presented in this paper, the training set consisted of each pixel in the image and was denoted as $\{x^{(1)}, x^{(2)}, \dots, x^{(L)}\}$, where $x^{(i)} \in \mathbb{R}^n$, $i \in [1 L]$, $L = M \times N$. Each sample $x^{(i)}$ of the training set is a vector decided by the combination of color components of different color spaces. The adopted algorithm is described by the following steps:

- (1) Select an initial set of k -cluster centroids $\mu_1, \mu_2, \dots, \mu_k \in \mathbb{R}^n$ randomly.
- (2) After assigning the initial k -cluster centroids, the algorithm proceeds by alternating between two steps until the steps converge:

Assignment step: For each sample $x^{(i)}$, decide the cluster of a sample by calculating the squared Euclidean distance between sample $x^{(i)}$ and each cluster centroid μ_j ($j \in [1 k]$) according to Equation (3):

$$c^{(i)} := \arg \min \|x^{(i)} - \mu_j\|^2, \quad (3)$$

where $c^{(i)}$ denotes the cluster number of $x^{(i)}$.

Update step: For each cluster, recalculate the new means to be the centroids of the observations in the new clusters according to Equation (4):

$$\mu_j := \frac{\sum_{i=1}^L 1_{\{c^{(i)}=j\}} x^{(i)}}{\sum_{i=1}^L 1_{\{c^{(i)}=j\}}}. \quad (4)$$

3. Proposed Method

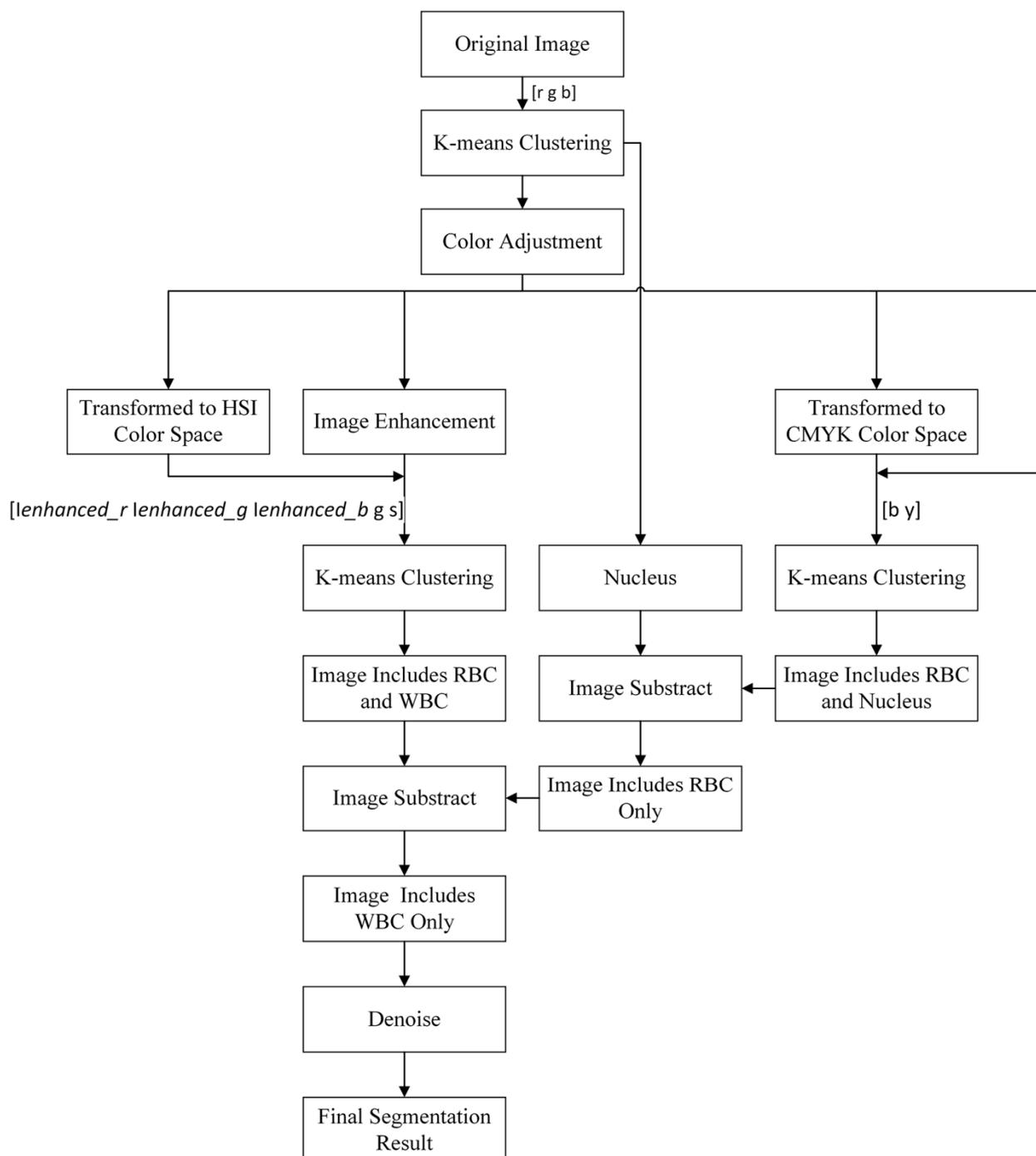
We adopted k-means clustering three times on the basis of the color information in different color spaces to segment the nucleus and cytoplasm of WBCs to achieve good performance. The entire flow chart of the proposed method is presented in Figure 3.

3.1. Color Adjustment

In Section 2.1 a general color correction method was introduced. However, complicated backgrounds degrade the performance of the method. To achieve satisfactory results, a color adjustment method based on color transfer and k-means clustering is introduced. The main idea is to divide the blood smear image into several parts by using k-means clustering. The color transfer algorithm will be applied for each part of the image, and the synthetic image can be obtained by reunion of the parts.

The stained blood smear images can be divided into three parts based on the following information: the background, the RBCs, and the WBC nuclei. Each pixel in each part has specific features. K-means clustering is used to segment the background, RBC, and the nucleus.

Figure 3. Flow chart of the proposed method.



The original image in the RGB color space was a $M \times N \times 3$ matrix. The value of each pixel was a 1×3 vector such as $[r \ g \ b]$, where r denotes the R component value of this pixel, g denotes the G component value, and b denotes the B component value. The $[r \ g \ b]$ of each pixel was treated as a feature vector; the original $M \times N \times 3$ matrix was reshaped to an $L \times 3$ matrix ($L = M \times N$), then k-means clustering was used to cluster every pixel into three partitions. Figure 4 shows the partition results.

After clustering, the template image and the input image are divided into three parts respectively. As shown in Figure 5, the color transfer algorithm is applied to each part of the template image and the

input image. The synthetic image can be obtained by the reunion of the three parts, and the subsequent steps are based on the synthetic image.

Figure 4. (a) Original image; (b–d) Partition results after clustering.

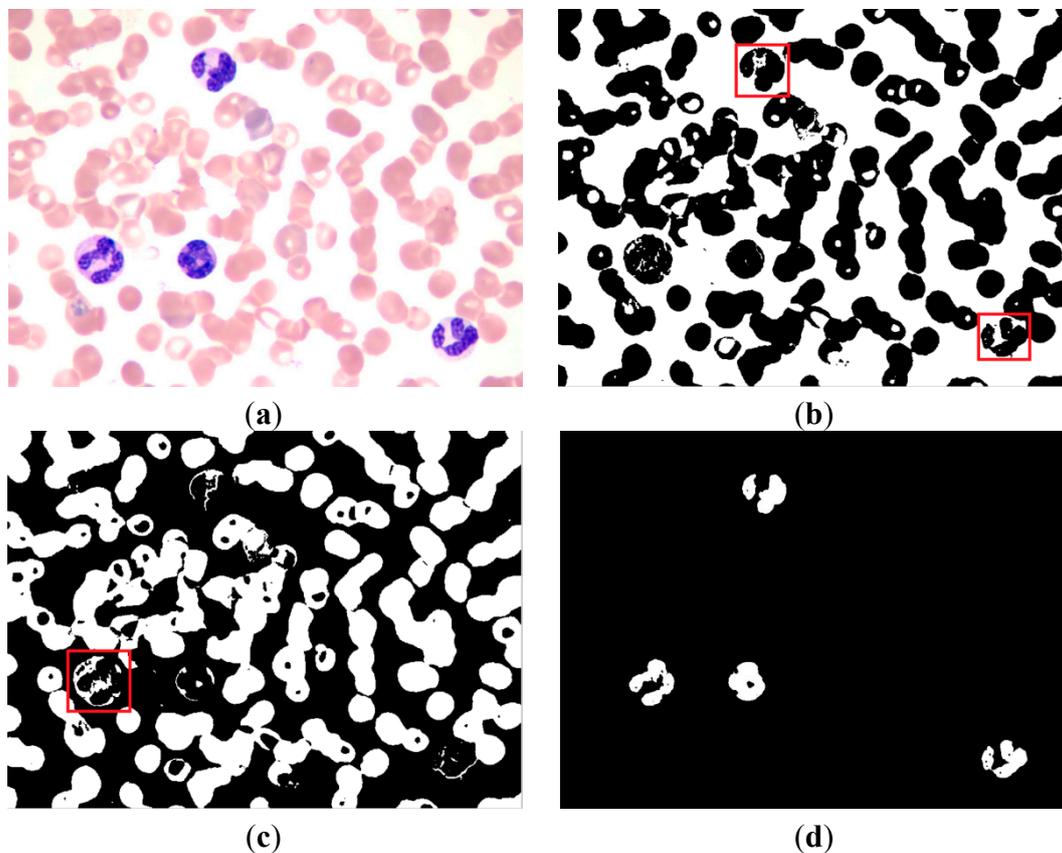
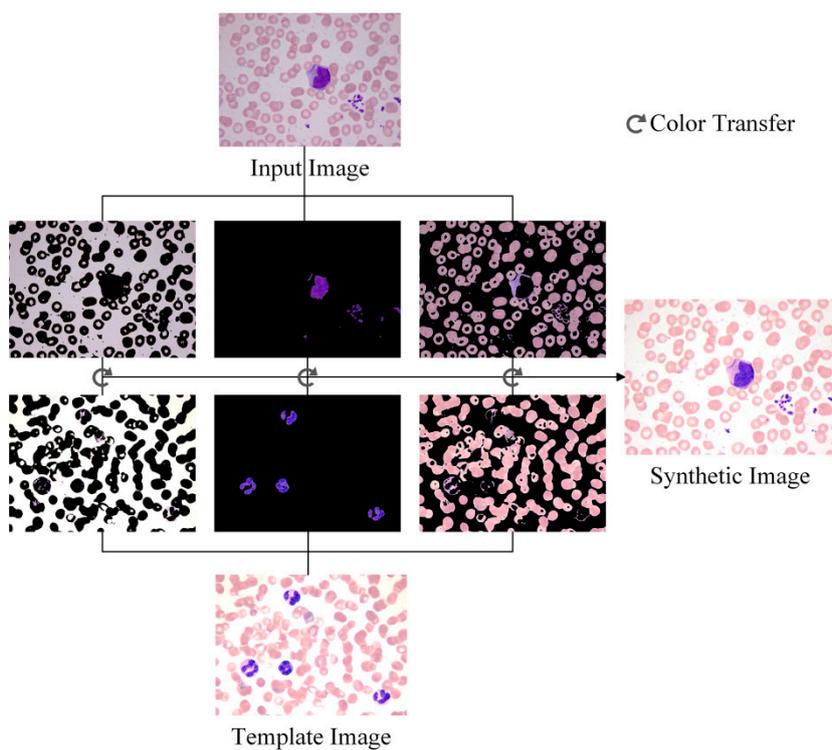


Figure 5. Schematic plot of the proposed color adjustment method.



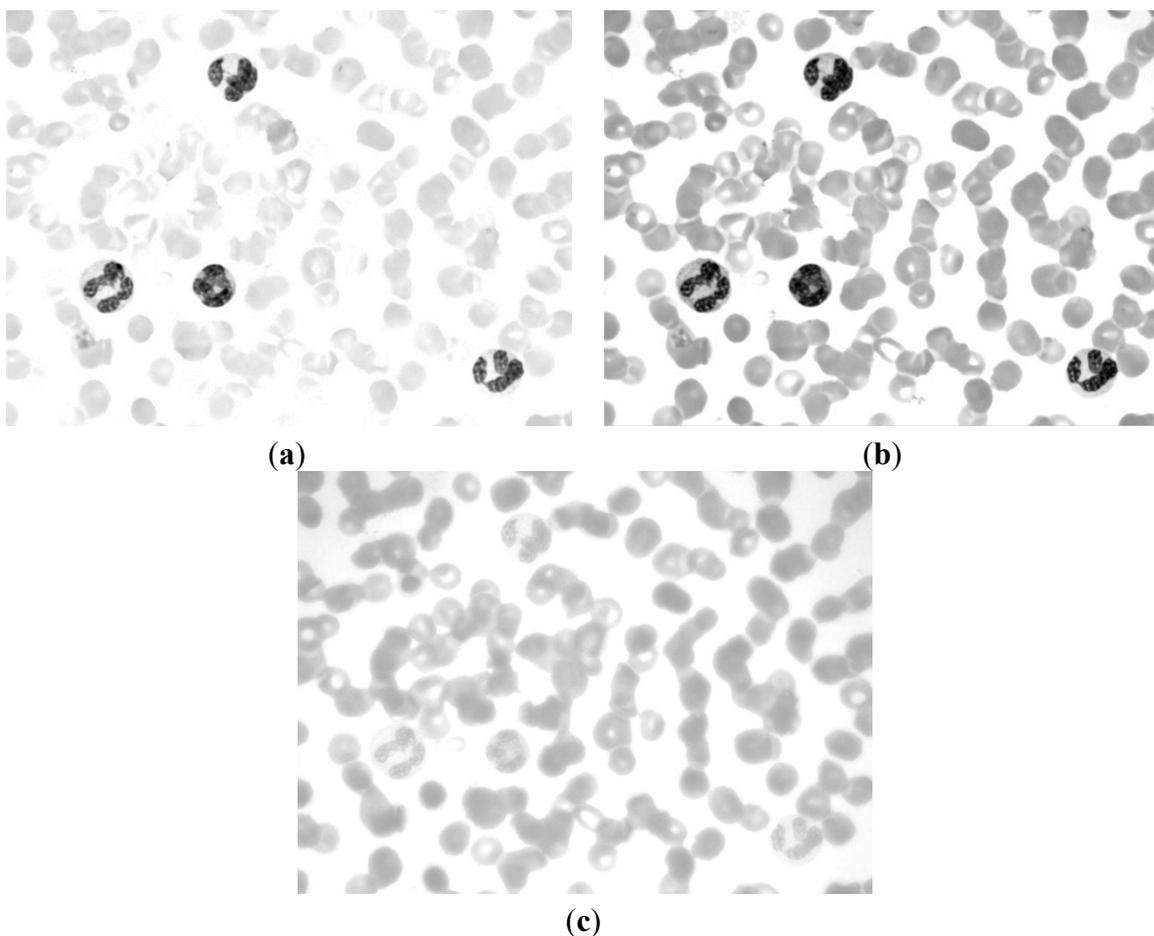
3.2. Red Blood Cell and Nucleus Segmentation

In Figure 4b, the white area represents the background of the image and contains part of the WBC cytoplasm because of the similar colors, as shown in the red block of the image. In Figure 4c, the white area represents the RBC region and part of the WBC cytoplasm, as shown in the red block of the image. The white area of Figure 4d represents the WBC nucleus region, and the segmentation result is relatively high because of high contrast.

We could not segment the cytoplasm of WBCs or RBCs directly from these results. Figure 4b shows that the white regions not only consists mostly of background but also contains the cytoplasm region; thus, removing the background of the image is unfeasible. Figure 4c shows that most of the white parts are RBCs but also contains the cytoplasm; thus, the cytoplasm region was difficult to remove and was deemed unsuitable to segment RBCs.

If we can determine a way to segment the RBC and nucleus together, then we can obtain the RBC region by subtracting the nucleus region obtained above. The R, G, and B components of the original image are shown in Figure 6. The CMYK color space image is translated from the RGB color space. The C, M, Y, and K components of the CMYK color space image are shown in Figure 7.

Figure 6. (a), (b) and (c) R, G, and B components of the original image respectively.



Figures 6c and 7c demonstrate that the RBC and nucleus regions have more contrasts than the cytoplasm in B and Y components. We can use this result to segment the RBCs and nuclei together. By combining B and Y components into a $M \times N \times 2$ matrix, the value of each pixel will be a 1×2 vector denoted as $[b \ y]$, where b denotes the B component's value of this pixel and y denotes the Y component's value of this pixel. The $[b \ y]$ of each pixel was considered a feature vector; the original $M \times N \times 2$ matrix was reshaped to an $L \times 2$ matrix ($L = M \times N$); and k-means clustering was adopted to cluster every pixel into two partitions. The results after partition are shown in Figure 8.

Figure 7. (a), (b), (c) and (d) C, M, Y, and K components of CMYK color space image respectively.

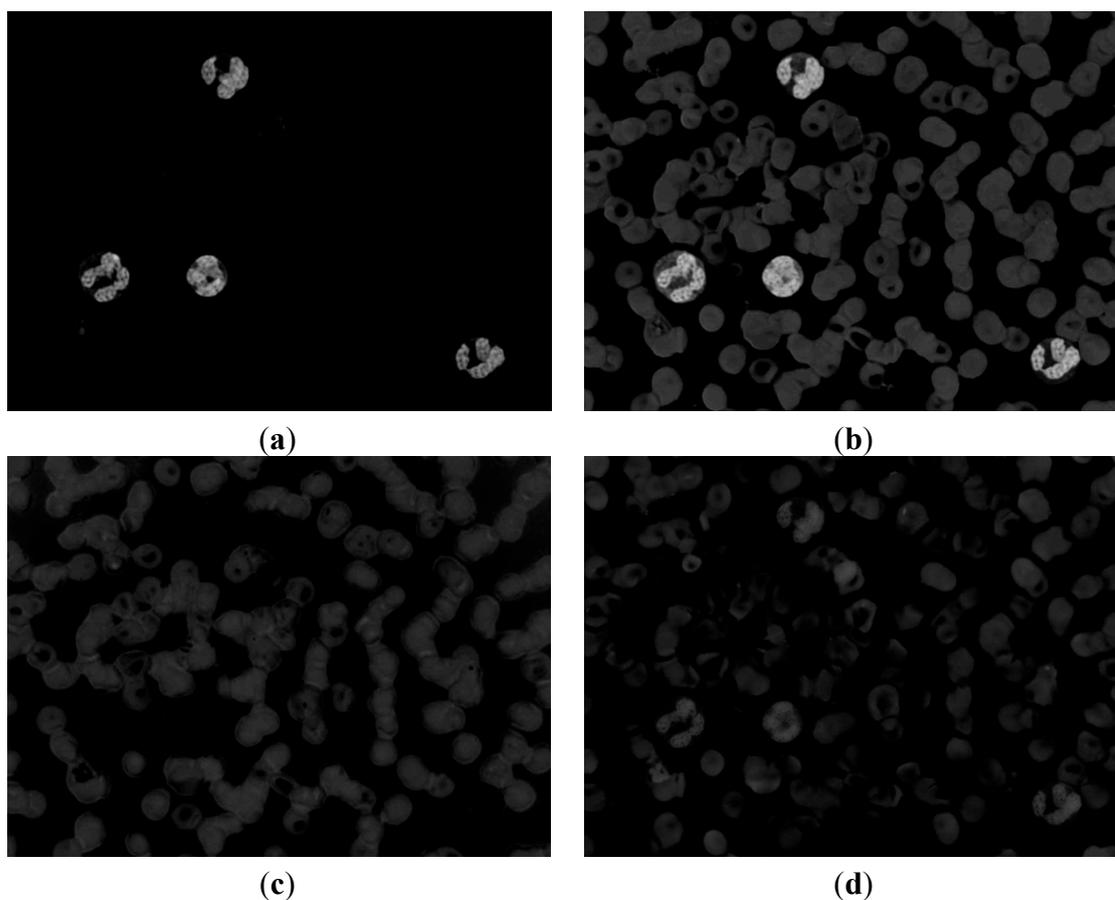
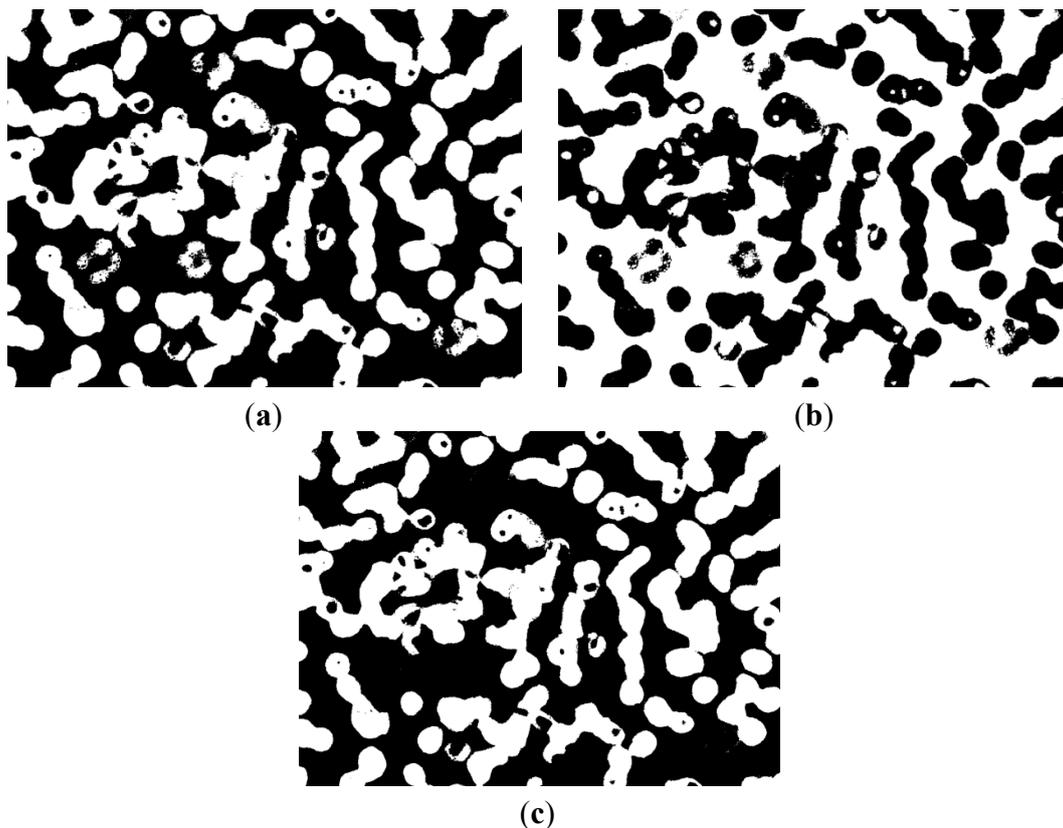


Figure 8a shows that this image contains RBC and nucleus regions only. The RBC region could be obtained by using Equation (5), and the result is shown in Figure 8c. The segmented RBCs may appear smaller than the actual cells; thus, morphological dilation was applied to expand the size of the RBC region to approximate the real one.

$$I_{RBC} = I_{RBC_Nucleus} - I_{Nucleus}, \quad (5)$$

where I_{RBC} denotes the binary image that contains RBCs only; $I_{RBC_Nucleus}$ denotes the binary image obtained above that contains RBC and nucleus (Figure 8a); $I_{Nucleus}$ denotes the binary image that contains the nucleus only (Figure 4d).

Figure 8. (a), (b) Clustering results; (c) RBC segmentation results.

3.3. Cytoplasm Segmentation

Cytoplasm segmentation is the most difficult part of WBC segmentation because the color of a cytoplasm is usually similar to the color of the RBCs or the background. The original image first needs to be enhanced to segment the cytoplasm. To enlarge the contrast of each content, the following equation was applied to every component of the original image:

$$I_{enhanced} = I + I_{bothat} - I_{tophat}, \quad (6)$$

where $I_{enhanced}$ denotes the enhanced RGB color space image, I denotes the original image, I_{bothat} denotes the image after bottom hat transformation, and I_{tophat} denotes the image after top hat transformation. Figure 9 shows that the cytoplasm region has high contrast in the enhanced image.

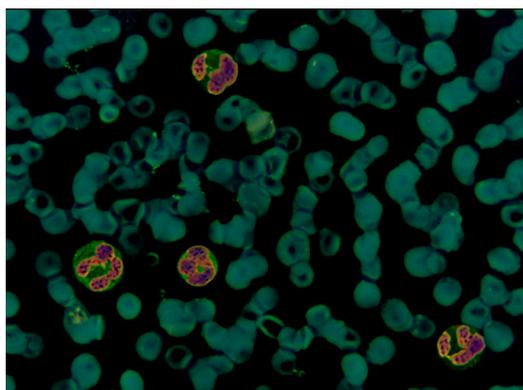
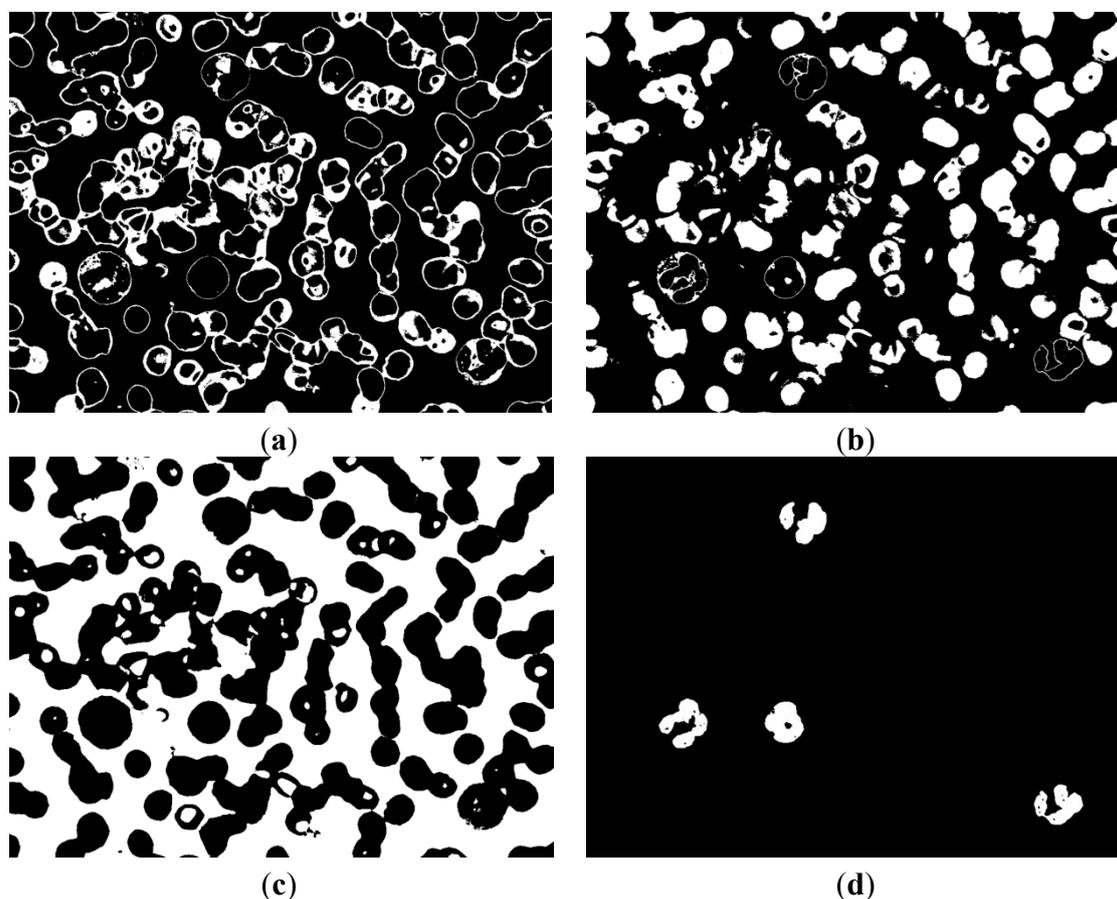
Figure 9. Enhanced image.

Figure 10c shows that the segmentation results of the background are better than the results in Figure 4b. Furthermore, the segmentation results of the background do not contain the cytoplasm region. Most of the cytoplasm parts are contained in Figure 10a, and some are contained in Figure 10b.

Figure 10. Clustering results of $I_{enhanced}$.



The entire WBC and RBC regions were obtained (Figure 11a) by combining the results of Figure 10a,b. The RBC region was also obtained (Figure 8c), and the WBC region could be acquired by subtracting Figure 8c from Figure 11a, with the result shown in Figure 11b. Figure 11b shows that the influence of noise on the WBC segmentation results. Thus, denoising is a necessary step. Smaller noises are prominent in the preliminary result of the entire WBC region segmentation; this result can be used to remove noises. First, an open operation was conducted to remove small dots and to separate WBC from small noises that connect to it. Second, all connected components were labeled and their sizes were computed before removing connected components with sizes smaller than the threshold value. After removing the noises outside the WBC, subsequent steps are needed to ensure the completeness of WBC. Third, the holes in WBC were filled and narrow gaps were connected by close operation. The WBC segmentation result after denoising is shown in Figure 12a. The cytoplasm part could be obtained by subtracting the nucleus region from the entire WBC region. The WBC nucleus and cytoplasm segmentation results are shown in Figure 12b.

Figure 11. (a) Result of the whole segmentation of the WBC and RBC regions; (b) Preliminary result of the entire WBC region segmentation.

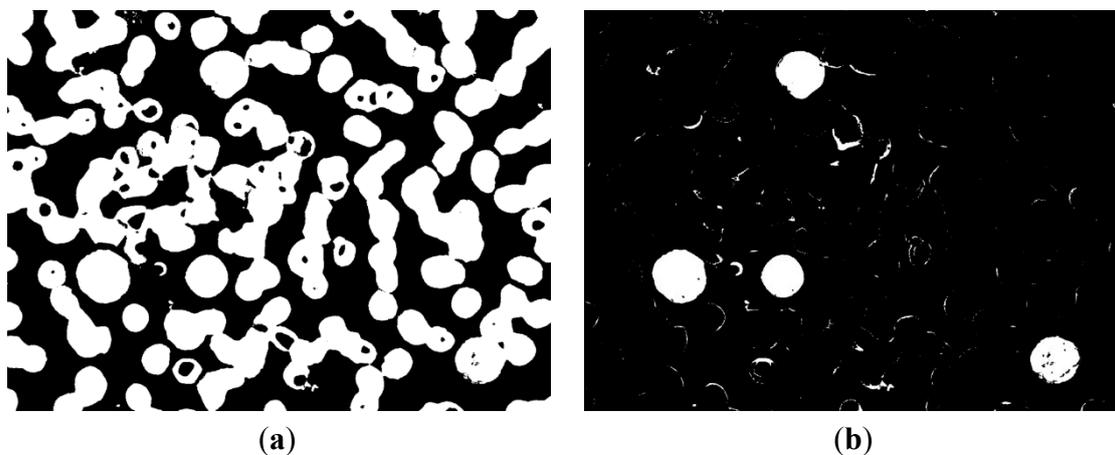
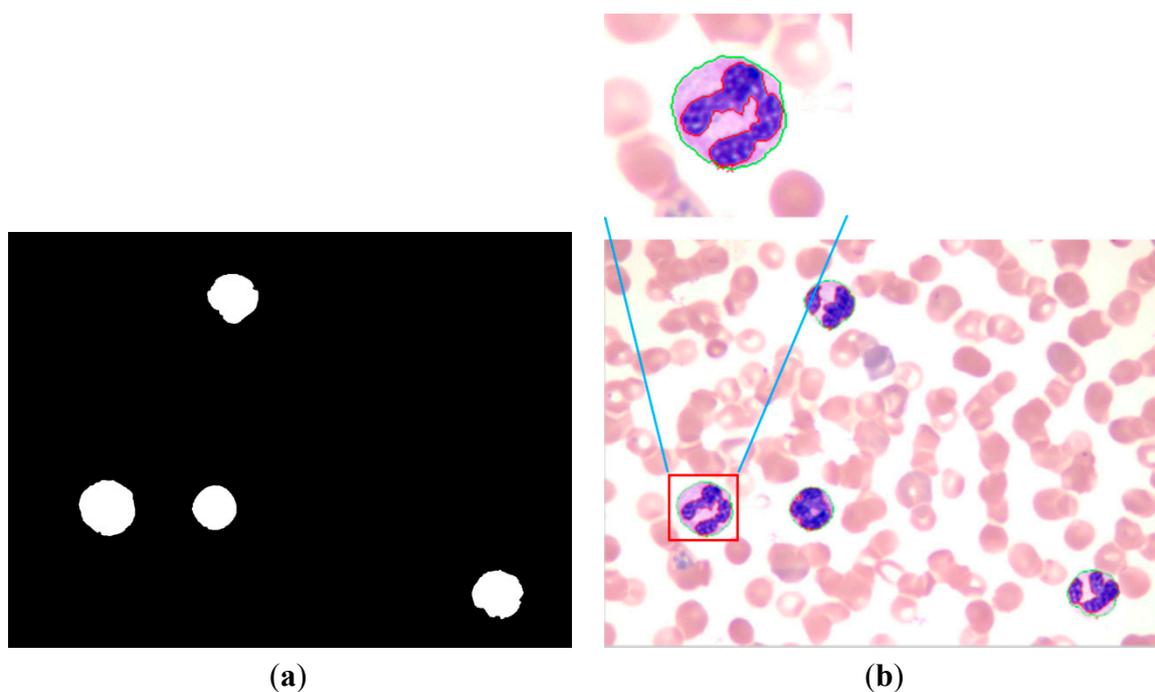


Figure 12. (a) WBC segmentation result after denoising; (b) Result of WBC nucleus and cytoplasm segmentation.

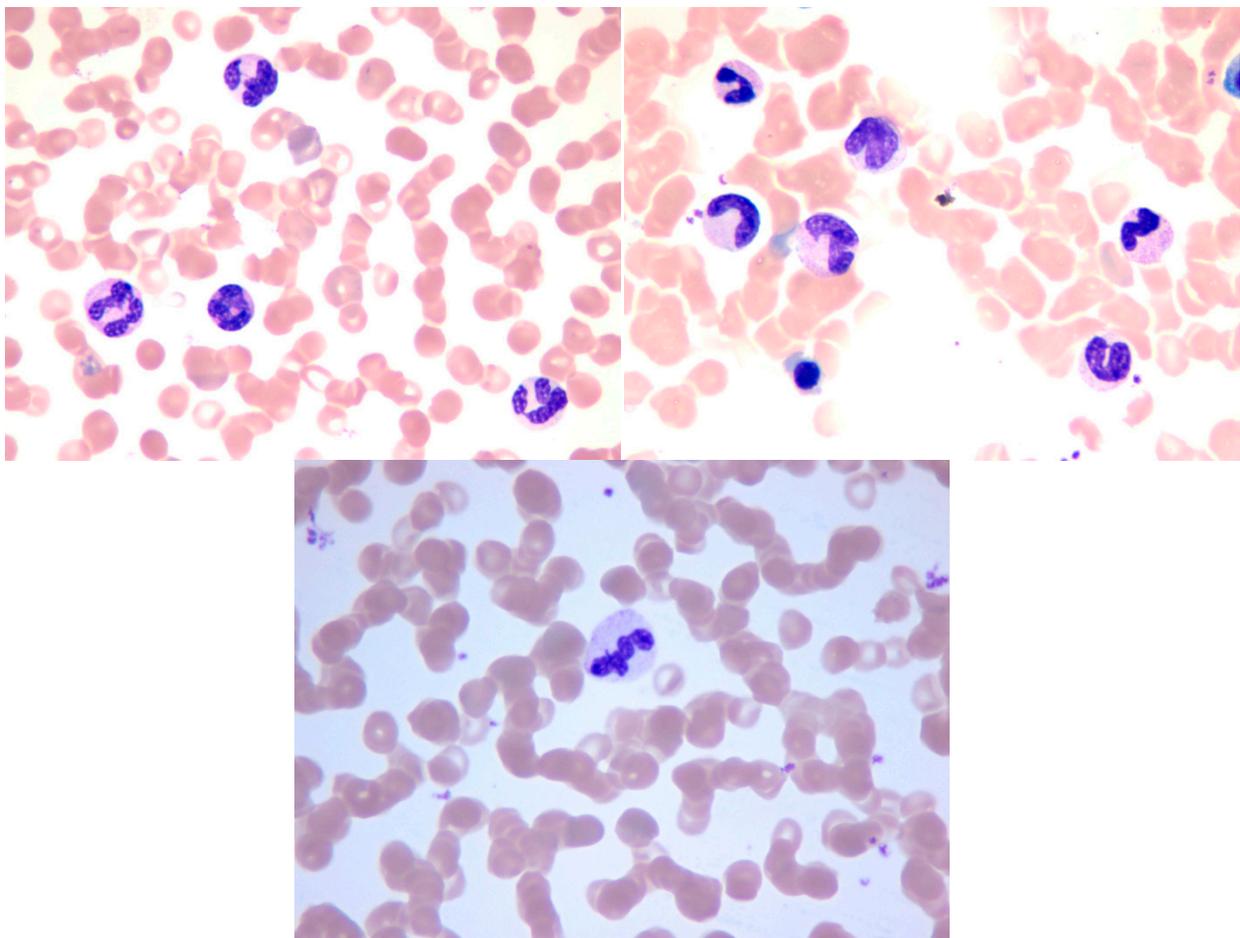


4. Experimental Results

Several methods for WBC segmentation have been proposed. However, the majority of these methods work only in specific databases wherein images are captured under certain circumstances to control light and other influencing factors. Our database was constructed by using an Olympus BX51 microscope in the cell lab at the Second Hospital of Shandong University under different circumstances. Our database consisted of 300 stained blood smear images. The images of the database were captured in three sections, and relatively long time gaps exist between each section. Approximately 100 images were obtained in each section of WBC image acquisition. In each acquisition section, the WBC images were captured under the same light condition from different

blood smears. Between each acquisition section the light condition differs from each other to some extent. The blood smears used for acquisition were stained with the same dyeing technique, and the dyeing time and the smear thick of each smear was slightly different according to the experience of the hematologist. Every WBC image has at least one WBC, and a number of them may have several WBCs. As previously mentioned, five types of WBCs can be found in the database. Every WBC image contains at least one type of WBC, and a number of them may contain several types of WBCs. The database has a total of 690 WBCs, which consists of 87 basophils, 94 eosinophils, 124 neutrophils, 203 lymphocytes, and 182 monocytes. Some sample images in our database are shown in Figure 13. The samples show that the color of each image differs from one another.

Figure 13. Sample images in the database.



Different databases differ in background color and dyeing conditions. Thus, not all of the previously proposed segmentation methods mentioned in the Introduction perform well all the time. The different conditions of the blood smear image hinder WBC segmentation. By conducting various experiments, we developed a method to segment WBC. Our method consists of two parts: (1) the segmentation of the nucleus and RBC; and (2) the segmentation of the WBC cytoplasm. By acquiring the whole WBC region and the nucleus region, the cytoplasm region could be obtained by a simple subtraction operation. During processing, we assumed that every input image had at least one WBC; no inferences were sizable compared with the WBC contained in the image. We also ignored the overlapping WBCs issues.

We present the performance test of the proposed WBC segmentation method on our database in this section, as discussed in Section 3. A total of 300 blood smear images were used to test the performance. We applied our method on each of the 300 images and then compared the results with the results of manual segmentation. The percentages of accuracy were calculated accordingly. Our database obtained an average overall accuracy of 95.7% for nucleus segmentation and 91.3% for cytoplasm segmentation. In this evaluation, the segmentation result is considered accurate when the auto-detected boundary closely matches the manually traced boundary.

We compared our method with the region growing method [6] and the snake method [8] to further evaluate the performance of our method. Table 1 shows the experimental results, wherein A1 denotes the ratio of the number of correctly detected WBCs to the total number of detected WBCs (Equation (7)). When $A1 = 1$, the WBCs detected in the microscopic images are all real WBCs, and no false detection occurs. The evaluation parameter A2 denotes the ratio of the number of WBCs detected to the total number of WBCs, as shown by Equation (8). When $A2 = 1$, no leak detection occurs. Equation (9) demonstrates the computation of false and leak detection ratios. A1 and A2 denote the detection accuracy of the entire system. In evaluating A1 and A2, we disregarded the accuracy of the traced boundaries. Therefore, some errors are allowed in a certain margin between the auto detected and manually traced boundaries; that is, the detection focuses only on the rough relative position between the traced boundaries and real WBCs. Therefore, the results are better than the results in the overall evaluation. D denotes the average Hausdorff distance between the auto detected and manually traced boundaries. Several experts are requested to manually trace the boundaries of WBC. We treated the average result of the experts as the ground truth and then applied the Hausdorff distance between the auto detected and manually traced boundaries to describe their contact ratios. A smaller D corresponds to better performance. D has a close relationship with the recognition accuracy of the entire system because the features extracted from the detected WBCs strongly rely on the segmentation results, whereas D describes the segmentation performance:

$$A1 = \frac{\text{The number of correctly detected WBCs}}{\text{Total number of detected WBCs}} \times 100\%, \quad (7)$$

$$A2 = \frac{\text{Total number of detected WBCs}}{\text{Total number of WBCs that exist in all images}} \times 100\%, \quad (8)$$

$$\begin{cases} E1 = 1 - A1 \\ E2 = 1 - A2 \end{cases}, \quad (9)$$

where E1 denotes the false detection ratio and E2 denotes the leak detection ratio.

The proposed method achieves a high overall accuracy for cytoplasm segmentation. Table 1 shows that the proposed method obtains low false and leak detection ratios and that the traced boundaries of the proposed method achieves high accuracy. The segmentation results for single WBC of the proposed method are compared with the original images and the manual segmentation results in Figure 14. The proposed method also performs well when multiple WBCs exists in one image, as shown in Figure 15.

Figure 14. Comparison of the proposed method and the manual method on single WBC segmentation.

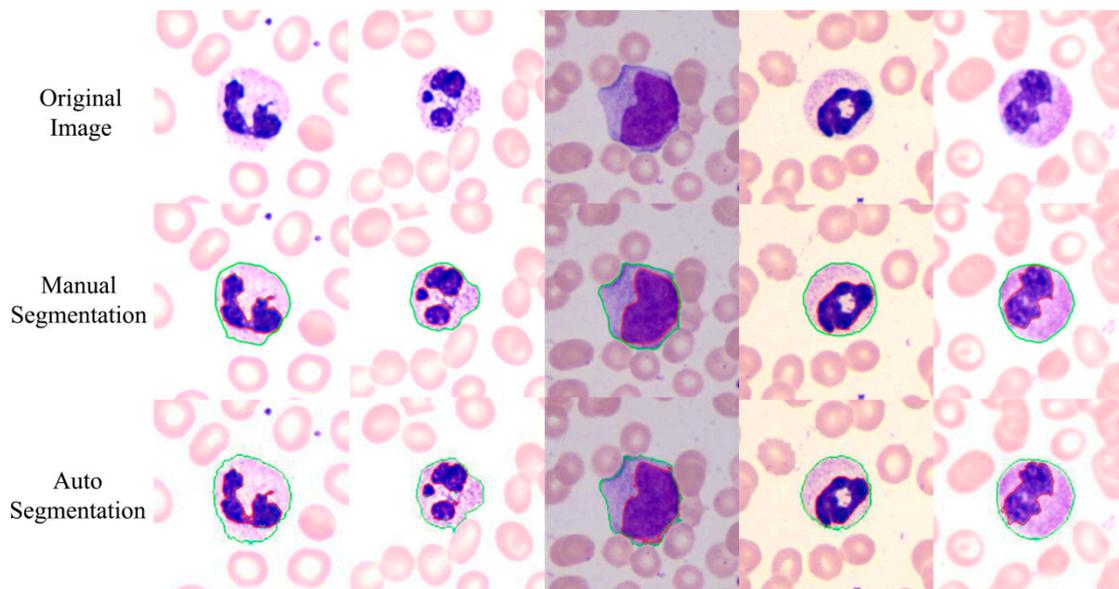


Figure 15. Comparison of the proposed method and the manual method on multiple WBC segmentation.

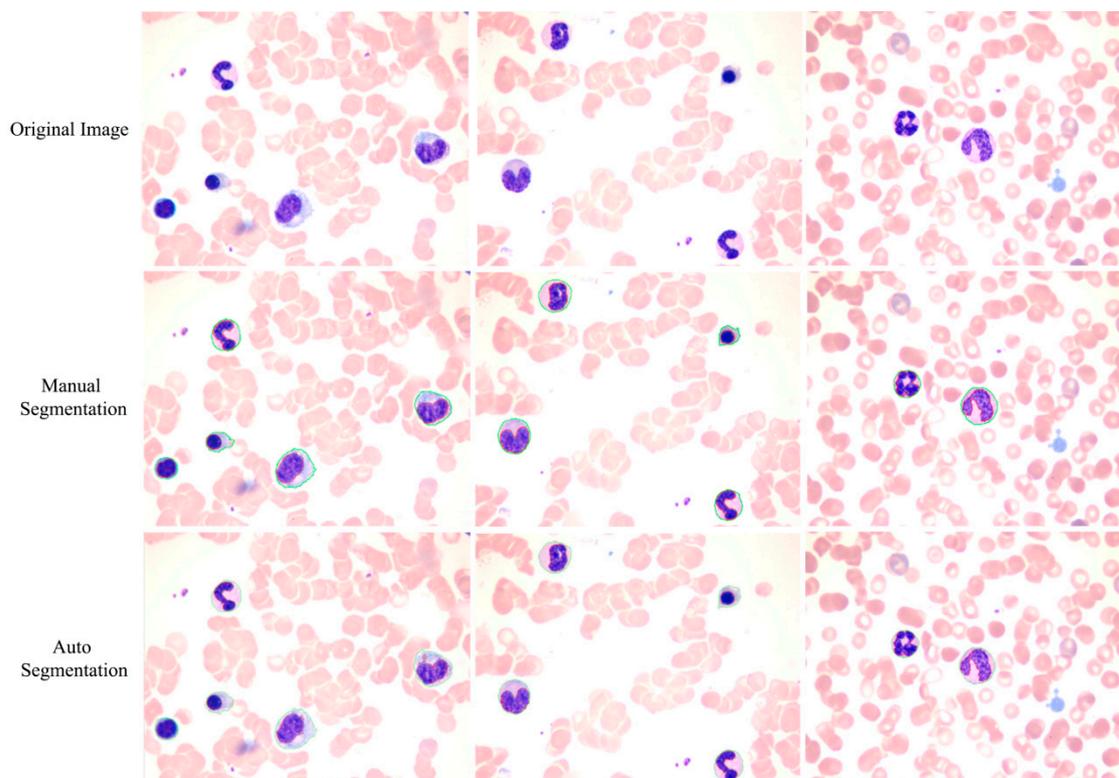


Table 1. Performance evaluation.

Parameter	Region Growing	Snake	Proposed Method
A1	88.4%	90.3%	94.6%
A2	89.2%	87.4%	95.1%
D	60.62	30.62	19.13

The performance of the proposed method, region growing method, watershed method and snakes' method was also evaluated via the parameter: OR, UR, ER and RDE [16,23,24]. OR denotes the over-segmentation rate, UR denotes the under-segmentation rate, and ER denotes the overall error rate. They are often used to evaluate the performance of segmentation methods. Pan *et al.* [16] and Liu *et al.* [23] have given a clear definition. RDE denotes the relative distance error, which is firstly proposed by Yang-Mao *et al.* [24] to evaluate the segmentation results. Assume that $e_1, e_2, e_3, \dots, e_{n_e}$ are the pixels of E, and $t_1, t_2, t_3, \dots, t_{n_t}$ are the pixels of T, in which E denotes the segmentation results, T denotes the ground truth, and n_e, n_t denote the number of pixels of E and T, respectively. The definition of RDE is illustrated as follows:

$$RDE = \frac{1}{2} \left(\sqrt{\frac{1}{n_e} \sum_{i=1}^{n_e} d_{e_i}^2} + \sqrt{\frac{1}{n_t} \sum_{j=1}^{n_t} d_{t_j}^2} \right), \quad (10)$$

where $d_{e_i} = \min\{\text{distance}(e_i, t_j) | j = 1, 2, \dots, n_t\}$, $d_{t_j} = \min\{\text{distance}(e_i, t_j) | i = 1, 2, \dots, n_e\}$, and $\text{distance}(e_i, t_j)$ denotes the Euclidean distance between e_i and t_j [16].

Table 2 shows the evaluation of the four methods via the average error measure of OR, UR, ER and RDE. The proposed method is only slightly worse than the region growing method in UR evaluation. It shows similar attributes to the watershed method in OR evaluation, however it is better in other indexes. Furthermore, the proposed method achieves a higher overall error rate of 14.3% and a lower relative distance error of 1.52 than the other compared methods. The evaluation demonstrates that the proposed method generally outperforms the traditional methods.

Table 2. Performance evaluation via average error measure of OR, UR, ER and RDE.

Parameter	Region Growing	Snake	Watershed	Proposed Method
OR	0.093	0.084	0.066	0.063
UR	0.059	0.081	0.068	0.067
ER	0.232	0.195	0.186	0.143
RDE	9.34	6.46	4.54	1.52

The performance of the proposed method on each type of WBC was evaluated. Table 2 shows the segmentation accuracy for each type of WBC. The performance of five kinds of segmentation methods was also compared. For each kind of method, a typical algorithm mentioned in the Introduction was evaluated as a representation. Duan's work [6] was adopted for the threshold-based method; Zheng's work [9] was adopted for the learning-based method; Zamani's work was adopted [7] for the active-contour-based method; Stanislaw's work [13] was adopted for the metaheuristic-based method; Zheng's work [14] was adopted for the saliency-based method. All of these methods were compared with the proposed method in Table 3, and the results show the proposed method achieved higher segmentation accuracy than those other methods.

In addition, owing to the color variety of blood smear images some segmentation methods that are sensitive to color space may fail to correctly segment WBCs. Most of the WBC segmentation methods focus mainly on the segmentation part, but few of them pay attention to color adjustment. However, the blood smear images captured in the clinic do not always have good color consistency. Thus, color adjustment is a necessary step in practical application. A color adjustment method based on k-means

clustering and color transfer was adopted in this paper. The color adjustment was treated equally as the segmentation part. Experiments show that the color adjustment can improve the accuracy of WBC segmentation. Table 4 shows the difference between the proposed method with color adjustment and without color adjustment.

Table 3. Performance evaluation of each types of WBC separately.

Methods	Neutrophil	Lymphocyte	Monocyte	Eosinophil	Basophil	Average
Threshold-based method [6]	88.7%	90.1%	92.3%	87.2%	83.9%	89.3%
Learning-based method [9]	94.4%	95.6%	97.3%	87.2%	85.0%	93.3%
Active-contour-based method [7]	91.9%	94.6%	94.5%	88.3%	86.2%	92.2%
Metaheuristic-based method [13]	94.4%	93.6%	96.2	93.8%	88.5%	93.5%
Saliency-based method [14]	95.2%	98.1%	95.6%	94.5%	85.1%	94.1%
Proposed method	97.6%	97.0%	97.8%	89.4%	89.7%	95.7%

Table 4. Performance evaluation with and without color adjustment.

Parameter	With Color Adjustment	Without Color Adjustment
A1	94.6%	90.2%
A2	95.1%	89.4%
D	19.13	29.43

5. Conclusions

This paper proposed a new WBC segmentation method by using color-space-based k-means clustering. A novel color adjustment method was applied before segmentation, thus improving the segmentation accuracy. The color components of RGB, HSI, and CMYK color spaces were applied to form the feature vectors of the k-means cluster. The k-means clustering results provide information that can be used for WBC segmentation. The experimental results show that the proposed method achieves precision rates of 94.6% and 95.1% for evaluation parameters A1 and A2, respectively, and higher segmentation accuracy for each type of WBCs. The experimental results indicate that the proposed method achieves high overall accuracy in cytoplasm segmentation and small leak and false detection ratios. The traced boundaries of the proposed method are more accurate than those of other methods. Moreover, the proposed method is immune to light conditions to some extent and is robust. All of the abovementioned characteristics demonstrate that the proposed method is accurate and effective for WBC segmentation, and outperforms traditional methods.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (No. 81100520), the Science and Technology Development Plans of Shandong Province (No. 2012GGE27073) and Independent Innovation Foundation of Shandong University, IIFSDU, (No. 2012JC015 and 2012DX001).

Author Contributions

Chengyun Zheng, Xiaoyan Xiao, Xiaomei Li and Zhi Liu conceived of the study, participated in its design and coordination, and helped to draft the manuscript. Congcong Zhang worked on the design of the study, carried out the laboratory experiments, analyzed the data, interpreted the results, and wrote the manuscript. Xiaomei Li, Ying-Jie Chen, Wu Zhen and Jun Chang carried out the laboratory experiments. All authors approved the final version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Yang, Y.; Zhang, Z.X.; Yang, X.H.; Jiang, D.Z. The blood cell counting and classification from stationary suspensions by laser light scattering method. In Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Hong Kong, China, 29 November 1998; Volume 4, pp. 1885–1888.
2. Wick, S. Human Physiology and Anatomy: Blood Cell Histology, 1997. Available online: <http://www.unomaha.edu/hpa/blood.html> (accessed on 23 December 2013).
3. Timby, B.K.; Smith, N.E. *Introductory Medical-Surgical Nursing plus Live Advice Online Student Tutoring Service*; Lippincott Williams and Wilkins: Philadelphia, PA, USA, 2006.
4. Dorini, L.B.; Minetto, R.; Leite, N.J. White blood cell segmentation using morphological operators and scale-space analysis. In Proceedings of SIBGRAPI 2007—XX Brazilian Symposium on Computer Graphics and Image Processing, Belo Horizonte, Brazil, 7 October 2007; pp. 294–304.
5. Huang, D.C.; Hung, K.D. Leukocyte nucleus segmentation and recognition in color blood-smear images. In Proceedings of IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Graz, Austria, 13 May 2012; pp. 171–176.
6. Duan, J.; Yu, L. A WBC segmentation method based on HSI color space. In Proceedings of the 4th IEEE International Conference on Broadband Network and Multimedia Technology (IC-BNMT), Shenzhen, China, 28 October 2011; pp. 629–632.
7. Zamani, F.; Safabakhsh, R. An unsupervised GVF snake approach for white blood cell segmentation based on nucleus. In Proceedings of the 8th International Conference on Signal Processing, Guilin, China, 16 November 2006; Volume 2.
8. Saraswat, M.; Arya, K.V. Automated microscopic image analysis for leukocytes identification: A survey. *Micron* **2014**, *65*, 20–33.
9. Putzu, L.; Di Ruberto, C. White blood cells identification and counting from microscopic blood images. In Proceedings of the WASET International Conference on Bioinformatics, Computational Biology and Biomedical Engineering, Guangzhou, China, 1 November 2013; Volume 73, pp. 268–275.

10. Zheng, X.; Wang, Y.; Wang, G. White blood cell segmentation using expectation-maximization and automatic support vector machine learning in Chinese. *J. Data Acquis. Process.* **2013**, *28*, 614–619.
11. Ko, B.C.; Gim, J.W.; Nam, J.Y. Automatic white blood cell segmentation using stepwise merging rules and gradient vector flow snake. *Micron* **2011**, *42*, 695–705.
12. Sadeghian, F.; Seman, Z.; Ramli, A.R.; Kahar, B.A.; Saripan, M.I. A framework for white blood cell segmentation in microscopic blood images using digital image processing. *Biol. Proced. Online* **2009**, *11*, 196–206.
13. Zack, G.; Rogers, W.; Latt, S. Automatic measurement of sister chromatid exchange frequency. *J. Histochem. Cytochem.* **1977**, *25*, 741–753.
14. Osowski, S.; Siroic, R.; Markiewicz, T.; Siwek, K. Application of support vector machine and genetic algorithm for improved blood cell recognition. *IEEE Trans. Instrum. Meas.* **2009**, *58*, 2159–2168.
15. Zheng, X.; Wang, Y.; Wang, G.; Chen, Z. A novel algorithm based on visual saliency attention for localization and segmentation in rapidly-stained leukocyte images. *Micron* **2014**, *56*, 17–28.
16. Pan, C.; Park, D.S.; Yoon, S.; Yang, J.C. Leukocyte image segmentation using simulated visual attention. *Expert Syst. Appl.* **2012**, *39*, 7479–7494.
17. Pan, C.; Park, D.S.; Yang, Y.; Yoo, H.M. Leukocyte image segmentation by visual attention and extreme learning machine. *Neural Comput. Appl.* **2012**, *21*, 1217–1227.
18. Pan, C.; Fang, Y.; Yan, X.G.; Zheng, C.X. Robust segmentation for low quality cell images from blood and bone marrow. *Int. J. Control Autom. Syst.* **2006**, *4*, 637–644.
19. Gudla, P.R.; Nandy, K.; Collins, J.; Meaburn, K.; Misteli, T.; Lockett, S. A highthroughput system for segmenting nuclei using multiscale techniques. *Cytom. Part A* **2008**, *73*, 451–466.
20. Reta, C.; Gonzalez, J.; Diaz, R.; Guichard, J. Leukocytes segmentation using Markov random fields. In *Software Tools and Algorithms for Biological Systems*; Springer: New York, NY, USA, 2011; pp. 345–353.
21. Shirley, P. Color Transfer between Images. *IEEE Corn.* **2001**, *21*, 34–41.
22. Gonzalez, R.C.; Woods, R.E. Color Image Processing. In *Digital Image Processing*; Prentice Hall: London, UK, 2002; pp. 416–478.
23. Liu, J.; Leong, T.Y.; Chee, K.B.; Tan, B.P.; Wang, S.C. Set-based cascading approaches for magnetic resonance (MR) image segmentation (SCAMIS). In Proceedings of the AMIA Annual Symposium, McLean, VA, USA, 27 April 2006; pp. 504–508.
24. Yang-Mao, S.F.; Chan, Y.K.; Chu, Y.P. Edge enhancement nucleus and cytoplasm contour detector of cervical smear Images. *IEEE Trans. Syst. Man Cybern. Part B* **2008**, *38*, 353–366.