World Scientific
www.worldscientific.com

A robust automatic leukocyte recognition method based on island-clustering texture

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

A leukocyte recognition method for human peripheral blood smear based on island-clustering texture (ICT) is proposed. By analyzing the features of the five typical classes of leukocyte images, a new ICT model is established. Firstly, some feature points are extracted in a gray leukocyte image by mean-shift clustering to be the centers of islands. Secondly, the growing region is employed to create regions of the islands in which the seeds are just these feature points. These islands distribution can describe a new texture. Finally, a distinguished parameter vector of these islands is created as the ICT features by combining the ICT features with the geometric features of the leukocyte. Then the five typical classes of leukocytes can be recognized successfully at the correct recognition rate of more than 92.3% with a total sample of 1310 leukocytes. Experimental results show the feasibility of the proposed method. Further analysis reveals that the method is robust and results can provide important information for disease diagnosis.

Keywords: Image processing; leukocyte recognition; texture analysis; island-clustering texture.

1. Introduction

Recognition and counting of human peripheral blood leukocyte (white blood cell, WBC) image play an important role in diseases diagnosis, but the manual process is time consuming and prone to error. For a trained medical technician, it may take more than 15 min to evaluate and count 100 WBCs for each blood smear.[1] Therefore, an automatic leukocyte image recognition method is needed. At present, most automatic leukocyte counting methods are based on laser-light scattering,[2,3] and flow-cytochemical principles which can only classify leukocytes from different sizes, but cannot recognize the five types of WBC for more detailed information. However, microscopic imaging-based automatic leukocyte recognition and counting

*Corresponding author.

methods can classify the five types of WBC and also give doctors more visual information of color, texture, shape, size, etc.

WBC image processing covers segmentation, features extraction and classification.[4] For segmentation, methods such as HSV segmentation,[5,6] image enhancement segmentation,[7] gradient vector flow snake segmentation[8] and supported vector machine (SVM) segmentation[9] are standard in literature. In general, well extracted features can compensate the effectiveness of segmentation for better recognition rate. Commonly used features of WBC include geometric features,[10–12] histogram-based features[6,13,14] and texture features.[1,11,15] The geometric features are effective in normal conditions, but may fail in the case of some distorted cells and hard to be differentiated WBCs such as basophil and lymphocyte. In this case, texture features could be more useful. However, general texture processing methods, like gray level co-occurrence matrix (GLCM)[1,16] and local binary pattern (LBP)[17] may not be intuitive. It may be hard to decide predetermined parameters used in these methods. Therefore, developing a more visual and robust texture is necessary for our applications. On classification, SVM,[6,12,18,19] artificial neural network (ANN)[12,15,19] and decision tree are popular. In this work, HSI segmentation is firstly applied, and geometric features are extracted, and then a new texture named island-clustering texture (ICT) is proposed to extract texture features of WBC, and finally ANN is used as classifier.

2. Related Works

At present, GLCM and LBP are the most frequently used texture analysis tools, so they are evaluated in this work.

2.1. *Texture analysis based on GLCM*

Textural features can be statistically evaluated by using GLCM. For an analyzed image I_c as shown in Fig. 1(a), its gray processing is dependent of the gray level (g). Figure 1(b) is the gray image of I_c at $g = 256$, Fig. 1(c) is the gray image of I_c at $g = 8$. In GLCM, for a given g, p_θ is the joint probability distribution of two pixels away from a given distance (d) along a direction (θ). Assume that GLCM is an isotropic matrix, i.e., GLCM is insensitive to

(a) Analyzed image I_c (b) Gray image at $g = 256$
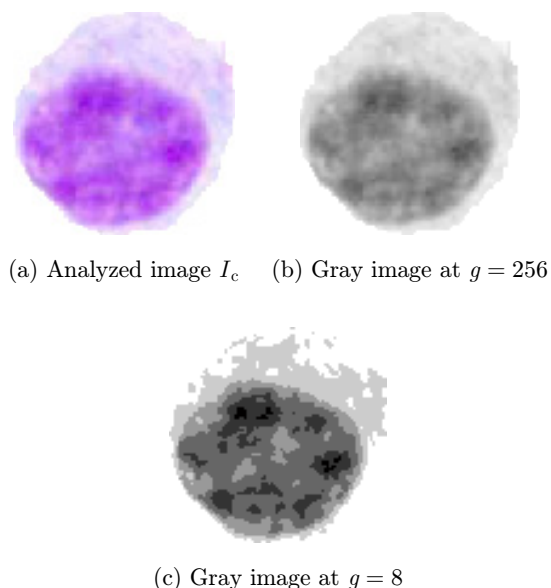

(c) Gray image at $g = 8$

Fig. 1. Gray processing at different gray levels of a leukocyte image.

180° rotation and invariant under different θ values: 0°, 45°, 90° and 135°. So the normalization of GLCM P is described by averaging p_θ for the above angles with its transposed matrix p_θ^T as Eq. (1):

$$P = \sum_{\theta=0,45,90,135} \frac{(p_\theta + p_\theta^T)}{8}. \qquad (1)$$

It should be noted that g may affect the dimensions seriously. The relationship of g and the size (S) of GLCM is shown in Eq. (2):

$$S = g^2. \qquad (2)$$

Gray level of $g = 256$ will make $S = 65,536$. Decreasing g to 8 will drop S to 64 for more practical but with accompanying cell detail information lost as shown in Fig. 1. The clear texture of cytoplasm in Fig. 1(b) disappeared in Fig. 1(c).

GLCM is just a matrix, and it needs 22 second-order statistics of GLCM for describing a texture, and the frequently used parameters are listed in Eqs. (3)–(5), where $p_{i,j}$ is the joint-probability distribution function, i and j keep track of cells by their horizontal and vertical coordinates. Angular second moment (ASM) described in Eq. (3) reflects the image gray distribution uniformity, ASM uses each $p_{i,j}$ as a weighting, and high values of it occur when the window is very orderly. Equation (4) can describe the Contrast of an image, but it may be influenced by noises. The Homogeneity in Eq. (5) represents the local changes of texture, and the

more orderly the texture is, the higher value it will possess.

$$ASM = \sum_{i,j=0}^{n-1} (p_{i,j})^2, \tag{3}$$

$$Contrast = \sum_{i,j=0}^{n-1} pi, j(i-j)^2, \tag{4}$$

$$Homogeneity = \sum_{i,j=0}^{n-1} \frac{pi, j}{1 + (i-j)^2}. \tag{5}$$

Although GLCM could extract some texture features of WBC, the gray levels, the angle and the distance affect the dimension of GLCM seriously, and another 22 second-order statistics of GLCM would be varied with the above three parameters. Obviously, changing any of them will produce more data to deteriorate the processing speed. In addition, choosing a set of proper parameter is not easy. On the other hand, deducing gray level may lose useful information. Furthermore, it will not be robust enough when processing WBC texture.

2.2. Texture analysis based on LBP

LBP proposed by Ojala,[17] is another useful texture tool that analyzes texture in different radii. It has multi-resolution [see Fig. 2] and has been used in leukocyte recognition.[19] LBP has only two necessary parameters, which makes it less complex than GLCM. One represents the size of texture structure, and the second one quantifies its intensity variance. As seen in Fig. 2, there are three sample circles with radii R and C points on the periphery. If the gray of central pixel is g_x and each peripheral pixel g_i, $i \in [0, C-1]$, then a vector T is obtained as follows:

$$T = t(g_x, g_0 - g_x, g_1 - g_x, \ldots, g_{C-1} - g_x). \tag{6}$$

Choose a proper binary function $B(x)$, then get the binary vector $T1$:

$$T1 = t(g_x, B(g_0 - g_x), B(g_1 - g_x), \ldots, B(g_{C-1} - g_x)). \tag{7}$$

LBP texture features can be expressed as $LBP_{C,R}$:

$$LBP_{C,R} = \sum_{c=0}^{C-1} B(g_c - g_x) \cdot 2^C. \tag{8}$$

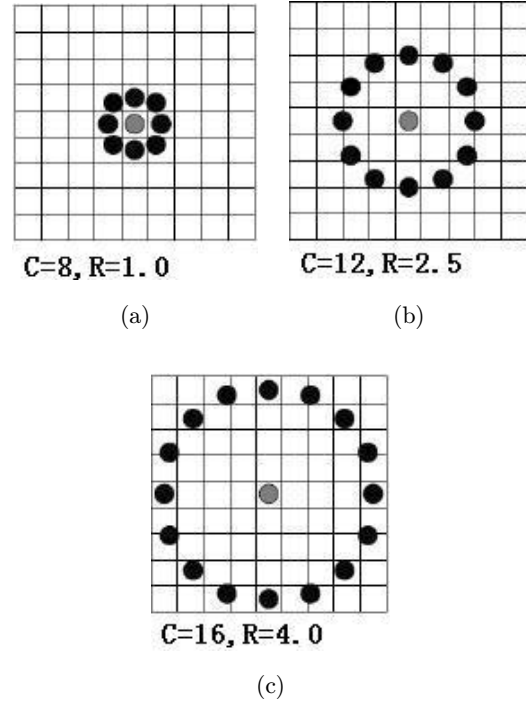C=8, R=1.0 (a)   C=12, R=2.5 (b)

C=16, R=4.0 (c)

Fig. 2.   Different radii of LBP.

LBP has too many patterns and is not rotation invariant. To solve this problem, $T1$ is circularly shifted with its decimal values, and the smallest one is chosen as the present pattern. Ojala found that most LBP patterns just have less than two times 0-1 or 1-0 fluctuations, so all the patterns can be divided into so-called "Uniform Pattern" and "Mix Pattern". Thus, the modes are decreased immensely. The rotation invariant LBP is defined as $LBP_{C,R}^{riu2}$ in Eq. (9), where the function U is employed to count the number of fluctuations between 0 and 1, and $U(LBP_{C,R}) <= 2$ represents "Uniform Pattern".

$$LBP_{C,R}^{riu2} = \begin{cases} \sum_{c=0}^{C-1} B(g_c - g_x), & U(LBP_{C,R}) <= 2 \\ c + 1, & \text{otherwise} \end{cases} \tag{9}$$

$LBP_{C,R}^{riu2}$ can only get the structure of texture, so feature $VAR_{C,R}$ quantifies the variation intensities of those circular pixels as:

$$VAR_{C,R} = \sum_{c=0}^{C-1} (g_c - \bar{g})^2, \quad \bar{g} = \sum_{c=0}^{C-1} g_c. \tag{10}$$

Now LBP could work well to deal with most textures, but may not fit it to some natural ones.

Because a WBC may have many sizes of LBP, but only a few of them are effective, so it is necessary to choose multi-sizes of LBP. On the other hand, LBP is a set of binary values which is not fit for our vision, so it cannot bring benefit to medical experts.

3. The ICT Model

One may not definitely say what the texture is, but some words like rough and smooth, uniform and nonuniform, can describe it to some degree. The question arises: how to calculate the degree of rough or uniform texture? GLCM and LBP are useful tools to extract features of the texture, but GLCM has so many parameters is that it hard to choose the best ones, and LBP also has to predetermine radius, and both of them should handle large data as shown in Sec. 2, so some degree of simplifying and approximating must be done to lower the data dimension with less information lost as far as possible. Furthermore, it is important to establish a more visual mode of texture which can simulate and benefit from human eyes.

3.1. *Island centers auto-extraction based on mean-shift clustering*

Mean-shift[20] is a nonparametric algorithm which can locate the maxima of a density function that may express some modes or features generally.

Given n data points x_i, $i = 1, \ldots, n$, on a d dimensional space R^d, and the basic mean-shift vector is defined as:

$$M_h(x) = \frac{1}{k_n} \sum_{x_i \in S_h} (x_i - x), \qquad (11)$$

where S_h is a higher dimensional ball area whose radius is h, and y is a set of points which satisfy the following relations:

$$S_h(x) = \{y : (y - x)^T(y - x) \leq h^2\}. \qquad (12)$$

There are k_n sample points in S_h area, and $M_h(x)$ always points toward the probability density gradient. Regardless of whether they are far or near, these points are contributed to $M_h(x)$ equally. In order to get various weights of sample points, kernel function $K(x)$ is added, and a probability density function $f(x)$ can be described as:

$$f(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right). \qquad (13)$$

If using radially symmetric kernels, a profile of the kernel $k(x)$ is defined to satisfy:

$$K(x) = c_{k,d} k(\|x\|^2), \qquad (14)$$

where $c_{k,d}$ is a normalization constant which assures $\int k(x)dx = 1$. The modes of the density function $f(x)$ can be located at the zeros of the gradient function $\nabla f(x) = 0$. The gradient of the density estimator is shown in Eq. (15).

$$\nabla f(x) = \frac{2c_{k,d}}{nh^{d+2}} \left[\sum_{i=1}^{n} g\left(\left\|\frac{x - x_i}{h}\right\|^2\right) \right]$$
$$\times \left[\frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \right], \quad (15)$$

where $g(x) = -k'(x)$, the first term is proportional to the density estimate at x computed with kernel $G(x) = c_{g,d} g(\|x\|^2)$, and the second term [see Eq. (16)] is just the mean-shift vector which points to the direction of maximum increasing probability density gradient.

$$m_h(x) = \frac{\sum_{i=1}^{n} x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x. \qquad (16)$$

Given the iteration times t, the mean-shift modes finding procedure is shown as:

(1) compute $m_h(x^t)$.
(2) shift the window $x^{t+1} = x^t + m_h(x^t)$.
(3) repeat 1 and 2 to locate the stationary points of the density function.
(4) remove these points and record them as one type of mode.
(5) repeat 1 to 4 until all modes are founded.

As shown in Fig. 3, the island centers of a leukocyte image is auto-extracted successfully by mean-shift modes finding procedure, especially the cytoplasm texture of the lymphocyte which disappeared in Fig. 1(c), and now is clearly labeled in Fig. 3(b).

3.2. *Definition of ICT*

The ICT is a type of texture model which regards a texture as a collection of islands, each of them has its own properties, and the texture is expressed by the spatial relation of these islands and statistical data of their properties. The ICT can be used in extracting texture feature, identifying or creating texture.
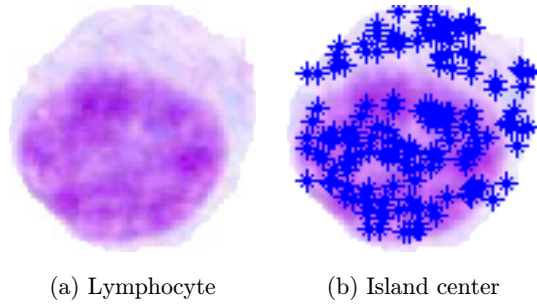
(a) Lymphocyte   (b) Island center

Fig. 3. Lymphocyte image processed by mean-shift modes finding procedure.

In order to give the ICT more visual attributes, a texture unit which simulates a real island is defined as a property vector I:

$$I = t(h_{\max}, h, s, v, l, lbp, \ldots), \qquad (17)$$

where $h_{\max}$ is max height of an island, h is its average height, s is area, v is volume, l is length-width ratio of its top view, lbp is the LBP value of its clustering center point. Other custom features or parameters such as center of gravity, torque, etc, can be added, not all of them are listed. Once most key islands have been detected, pure statistic methods or classic texture methods come in handy. s and v can describe the rough degree, $h_{\max}$ and h express the contrast. However, l and lbp or all kinds of invariant moments could reveal the details or micro shape of a texture, etc. Obviously, if all of them are chosen, it would get more effective, but it would consume more processing time. In short, this vector contains many features, most of them are visual, thus it is easier to choose the feature parameters than other methods.

3.3. *The procedure of ICT extraction*

There are five steps in ICT extraction:

Step 1. Main parameter choosing: There are many parameters which can be chosen for creating islands, such as gray level, G channel of RGB color space, or S of HSV color space, etc. In leukocytes recognition, the gray level is the best one as the main parameter, because it is less likely affected by lighting and dyeing conditions than colors.

Step 2. Island centers auto-extraction by the mean-shift as described in Sec. 3.1.

Step 3. Islands creation: In Step 2, a series of island centers is acquired. But the regions of the islands are hard to compute precisely, because some islands are overlapped partly. A simplified method is suggested to solve this puzzle. The growing region is applied to create the region of an island, whose seed point is just its center, and the threshold is lower at three gray level of the center. Although this method is somewhat not precise, it is fairly effective enough to satisfy the later recognition. The ICT procedure for five typical classes of leukocytes are shown as Figs. 4–8.

As shown in Fig. 4(a), one of the main visual features of a neutrophil is that it has light cytoplasm with many small and uniform granules, each of them is lifted as a peak after clustering. Let each peak be the center of an island, then the size and the distance of these granules can be measured as features.

As shown in Fig. 5(c), after clustering, the uniform and rough texture features of a eosinophil are protruded, and the height of these granules is taller than that of the neutrophil's, therefore the eosinophil can be identified by the height and the area of the islands.

As shown in Fig. 6(a), a basophil has nonuniform and density texture features, which are more visible when shown in Fig. 6(c). Obviously, the average height of the granules is bigger than the neutrophil, and the variance of height must be the biggest than other leukocytes for its nonuniform texture, thus the basophil can be recognized.

As shown in Fig. 7(a), the lymphocyte has high density texture like the basophil, but has less islands
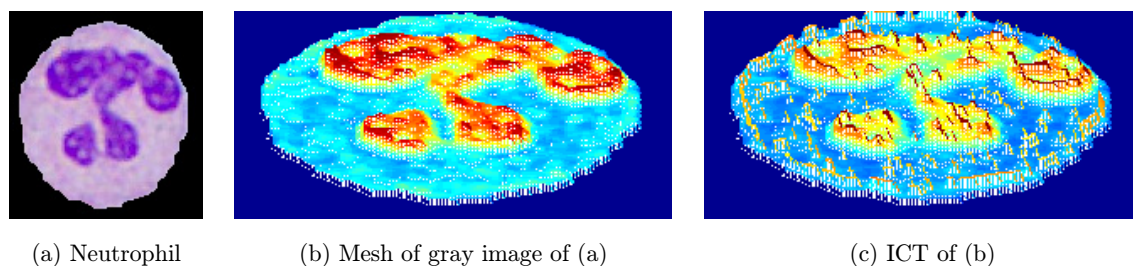

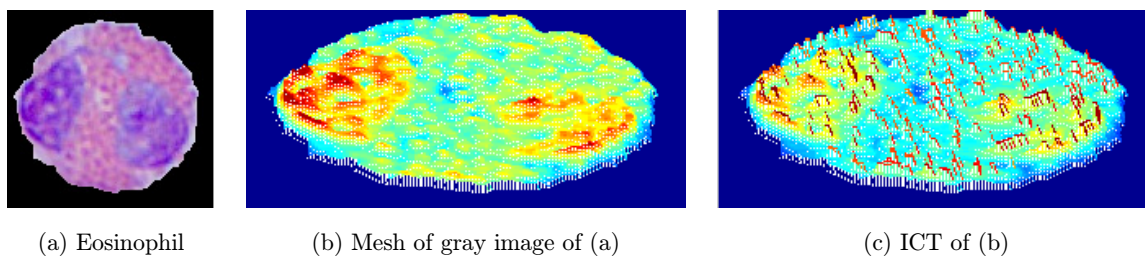(a) Neutrophil   (b) Mesh of gray image of (a)   (c) ICT of (b)

Fig. 4. Clustering of neutrophil.

(a) Eosinophil   (b) Mesh of gray image of (a)   (c) ICT of (b)

Fig. 5.   Clustering of eosinophil.

(a) Basophil   (b) Mesh of gray image of (a)   (c) ICT of (b)

Fig. 6.   Clustering of basophile.
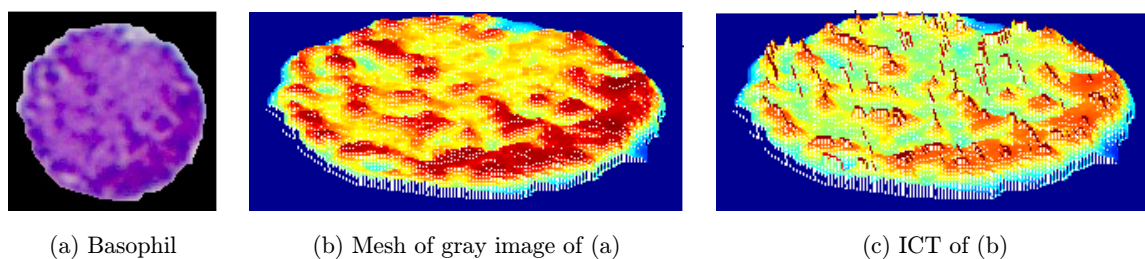

(a) Lymphocyte   (b) Mesh of gray image of (a)   (c) ICT of (b)
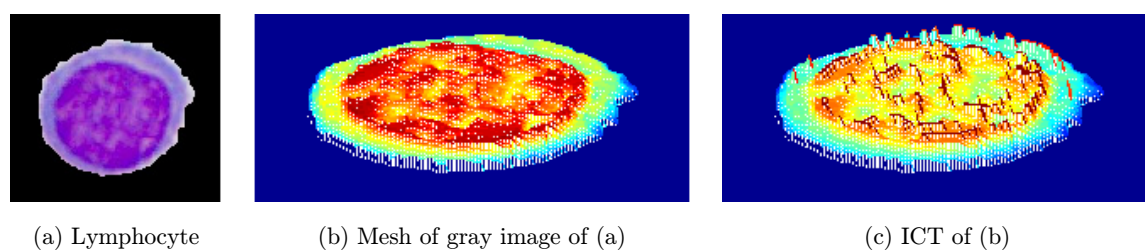
Fig. 7.   Clustering of lymphocyte.

[see Fig. 7(c)] than other leukocytes, so it can also be distinguished by the count of islands.

As shown in Fig. 8(a), the monocyte has light cytoplasm, low density texture, which is shown more clearly after clustering [see Fig. 8(c)]. Therefore, the height and counts of the islands are still effective to mark it.

In short, after clustering, all granules or protruding blocks of the leukocytes are lifted and shrunk to corresponding center points, which possess various gray level info and spatial structures of the textures at the same time. Therefore, by choosing proper ICT feature parameters, it is possible to distinguish the five types of leukocytes and to recognize them.

Step 4. Choose favorite properties and establish a property vector: measure or calculate every property of each island, optionally add any statistic
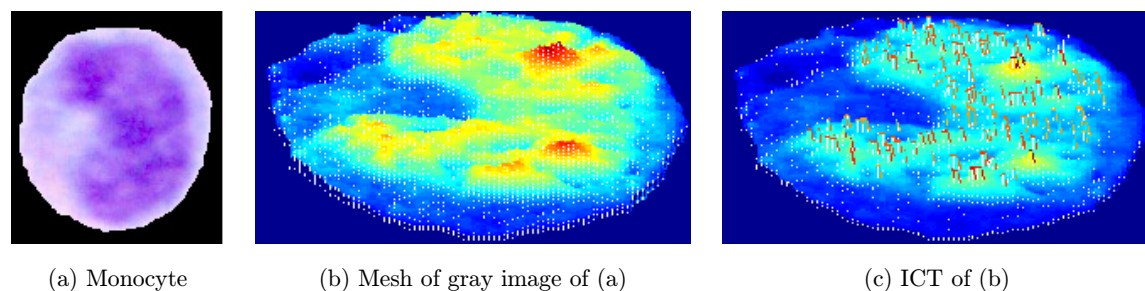

(a) Monocyte   (b) Mesh of gray image of (a)   (c) ICT of (b)

Fig. 8.   Clustering of monocyte.

Table 1.   ICT features of five types of WBC.

Properties/types	Neutrophil	Eosinophil	Basophil	Lymphocyte	Monocyte
$\overline{h}_{\max}$	101.15	110.47	152.15	142.57	69.61
$\overline{s}$	7.71	3.21	5.37	7.09	12.41
$v1$	25.12	22.39	19.39	14.95	35.94
$v2$	65.32	45.16	70.55	51.93	33.21
n	153	182	126	70	208

parameters. As seen in Figs. 4–8, neutrophil, eosinophil and basophil belong to granulocytes for their cytoplasms have many granules, instead, lymphocyte and monocyte have fewer granules, and the texture of these granules are different for their average height ($\overline{h}_{\max}$), average area ($\overline{s}$), and variance of distance ($v1$), variance of $h_{\max}(v2)$, and the count of islands (n), therefore, the feature vector of ICT can be constructed of the three properties, as shown in Eqs. (18)–(20):

$$I = t(\overline{h}_{\max}, \overline{s}, v1, v2, n), \qquad (18)$$

$$v1 = \sqrt{\frac{1}{n-1} \sum_{i=0}^{n-2} (d_i - d_{i+1})^2}, \qquad (19)$$

$$v2 = \sqrt{\frac{1}{n} \sum_{i=0}^{n-1} (h_{\max i} - \overline{h}_{\max})^2}. \qquad (20)$$

In Eq. (19), d_i is the distance of two adjacent peaks.

Step 5. Analysis ICT vector values.

Statistic data of Figs. 4–8 computed from Eqs. (18)–(20) are listed in Table 1.

As seen in Table 1, the texture of WBC can be clearly described by the feature of ICT, and values of these properties are fit for observation by human eyes. The lymphocyte and the basophil have more dark paint and contrast than others for their average of $h_{\max} > 140$, but $h_{\max}$ of the monocyte is low at 69.61, and the rest of WBCs have middle values of about 100. The average of S reflects the texture block size and roughness, and the small value expresses a uniform and tiny texture, such as eosinophil. As a contrast, monocytes have the biggest value for they have rough blocks. The $v1$ describes the uniformity of position, and it can be used to identify lymphocyte. The $v2$ is used to distinguish lymphocyte and monocyte. At last, n is easier to identify lymphocyte for it has lowest count of granules.

In the general condition, ICT may be effective in extracting the texture features of leukocytes, and it will be evaluated in special conditions to test its robustness.

3.4.   *Robustness testing of ICT*

A lot of leukocytes images have been analyzed using ICT. Figure 9 shows a typical case when handling weak repeatability of a texture [see Fig. 9 (a)]. It is hard to classify Fig. 9(a) and Fig. 9(c) as the same type of leukocyte, and Fig. 9(a) is more likely classified as lymphocyte by general texture tools for its texture is more smooth like lymphocyte, but their ICT features listed in Table 2 show they belong to basophil. Although the average of $h_{\max}$ of basophil and lymphocyte are very similar, the count of granules of lymphocyte is much smaller than basophil, at the same time, the average of S of lymphocyte is higher than basophil. So some basophils with weak repeatability texture can be distinguished from lymphocyte effectively by ICT.
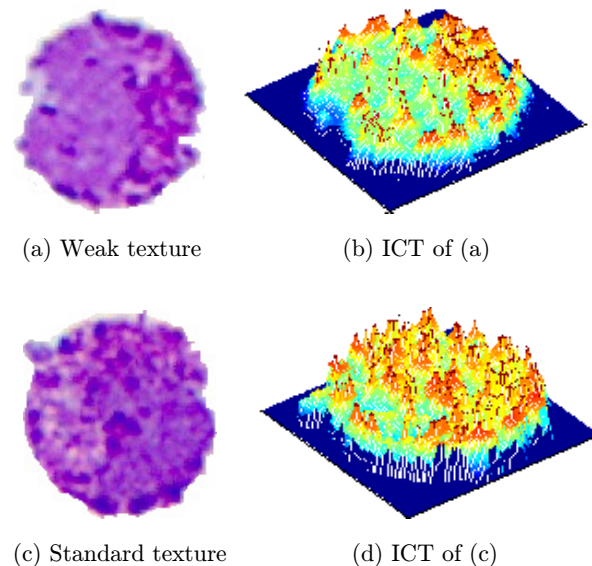

(a) Weak texture

(b) ICT of (a)

(c) Standard texture

(d) ICT of (c)

Fig. 9.   Different repeatability in the same type of basophil.

Table 2.   ICT features of basophil vs lymphocyte.

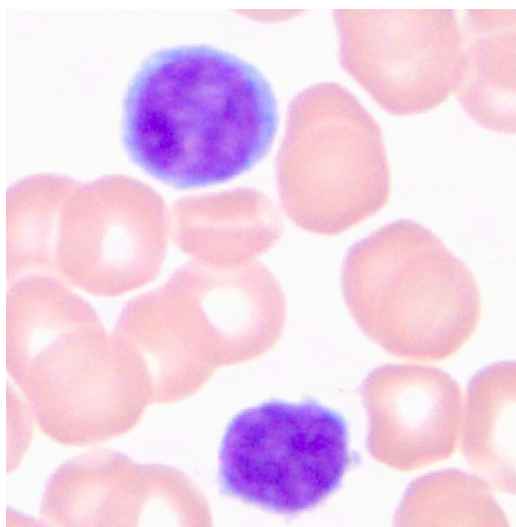| Properties/types	Image1	Image2	Image3	Image4
$\overline{h}_{\max}$	131.30	145.93	152.15	142.57
$\overline{s}$	4.32	3.56	5.37	7.09
$v1$	20.34	23.38	19.39	14.95
$v2$	59.98	46.37	70.55	51.93
n	124	134	126	70

Fig. 11.   Distorted lymphocyte.

Fig. 10.   Special lymphocyte without cytoplasm (the bottom one).

Note in Table 2: image1 is Fig. 9(a), image2 is Fig. 9(c), image3 is Fig. 6(a), all are images of basophil, but image4 is lymphocyte in Fig. 7(a).

The other typical condition is that a few special lymphocytes have so little cytoplasm [see Fig. 10, the bottom one] that they will be identified as basophils, because the area of cytoplasm is a key geometric feature to identify WBC for most of basophils have little cytoplasm and most of lymphocytes have notable ones. Fortunately now though the other ICT features of them are very similar, the n value of the top one is 87, and the bottom one is 57, while the basophil is more than 120. Therefore, lymphocytes can be rightly recognized even if they have little cytoplasm and look much like basophils by ICT.

Another case is the sample blood is not too fresh to make some lymphocyte distorted as Fig. 11. In general condition, the circularity of lymphocyte is about 1, and can be used as an important geometric feature to discriminate lymphocyte, but it will fail in this case. However, the n value of ICT in Fig. 11 is low at 32, so it is easy to be classified as lymphocyte by n.

4.   Experimental Methods

There are five steps to do experiment. Firstly, a group of micrograph of human peripheral blood cells is captured and every WBC is located. Then a WBC is segmented by HSI color system and shape filter. Next, the geometric and ICT features of the WBC are extracted. Finally, all the features are flushed into ANN classifier and the recognition result is figured out.

In order to evaluate the proposed method deeply as far as possible, first a less number of 150 sample images and a bigger number of 1310 sample images with 11 nonWBC images in it were collected and classified manually.

4.1.   *WBC segmentation*

A typical ROI usually includes a big WBC and parts of red blood cell (RBC) or other small cells [see Fig. 12(a)]. Obviously, there are two features to distinguish them: hue channel of hue-saturation-intensity (HSI) and shape. First, we extract hue channel and paint its histogram, and then use auto-threshold method[21] to cut background [see Fig. 12(b)]. Second, although morphology algorithms can be used to get their shape, for reason of savings in time, the shape filter is used instead. Then the small blocks or particles or noises are wiped out, and only a WBC or some big blocks of dyestuff are left. Finally, some WBCs may have hole, where the filling operation should do for features extraction [see Fig. 12(c)] later. As seen in Fig. 13, some leukocytes are segmented not very precisely [see Fig. 13(d) and
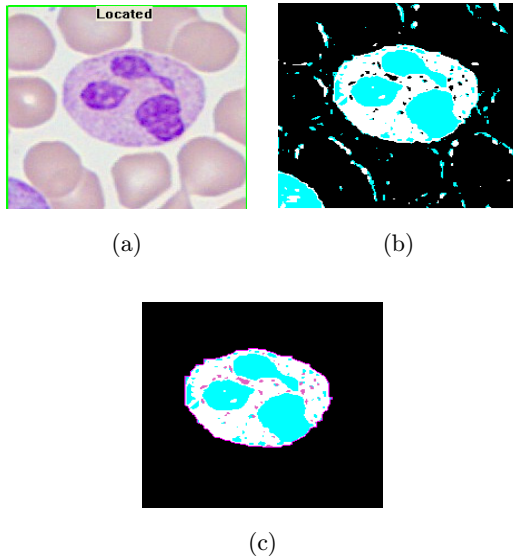
Fig. 12. (a) Located ROI, (b) segmented image by auto-threshold of Hue and saturation of HSI and (c) final result of shape filter and filled mask.

has high saturation above 0.45. Now, a single and complete mask image of WBC is obtained, and then a segmented WBC can be obtained by logic "and" operation.

Geometric features are frequently used in the leukocytes recognition,[4] such as number of segmented nucleus (NSN), area, perimeter, circularity, ratio of areas of nucleus and cytoplasm (RNC), etc. [see Table 3]. However not all of these are listed in Table 3, because texture features is our aim.

$$c = \frac{4\pi S}{L^2}, \tag{21}$$

where c, S and L are circularity, area and perimeter of an individual leukocyte.

These geometric features are useful for classifying WBC, but some cells still cannot be identified precisely because some of the features overlap with each other, for example, basophil and lymphocyte only have one NSN and overlapped circularity, monocyte and eosinophil have overlapped RNC and area [see Table 3], etc. Sometimes, they have very similar geometric values. Therefore, texture features are employed in the next step.

13(e)], even so, the result can be used in the succeeding procedures. Furthermore, dividing a WBC region into nucleus and cytoplasm just needs saturation threshold for the nucleus of WBCs always
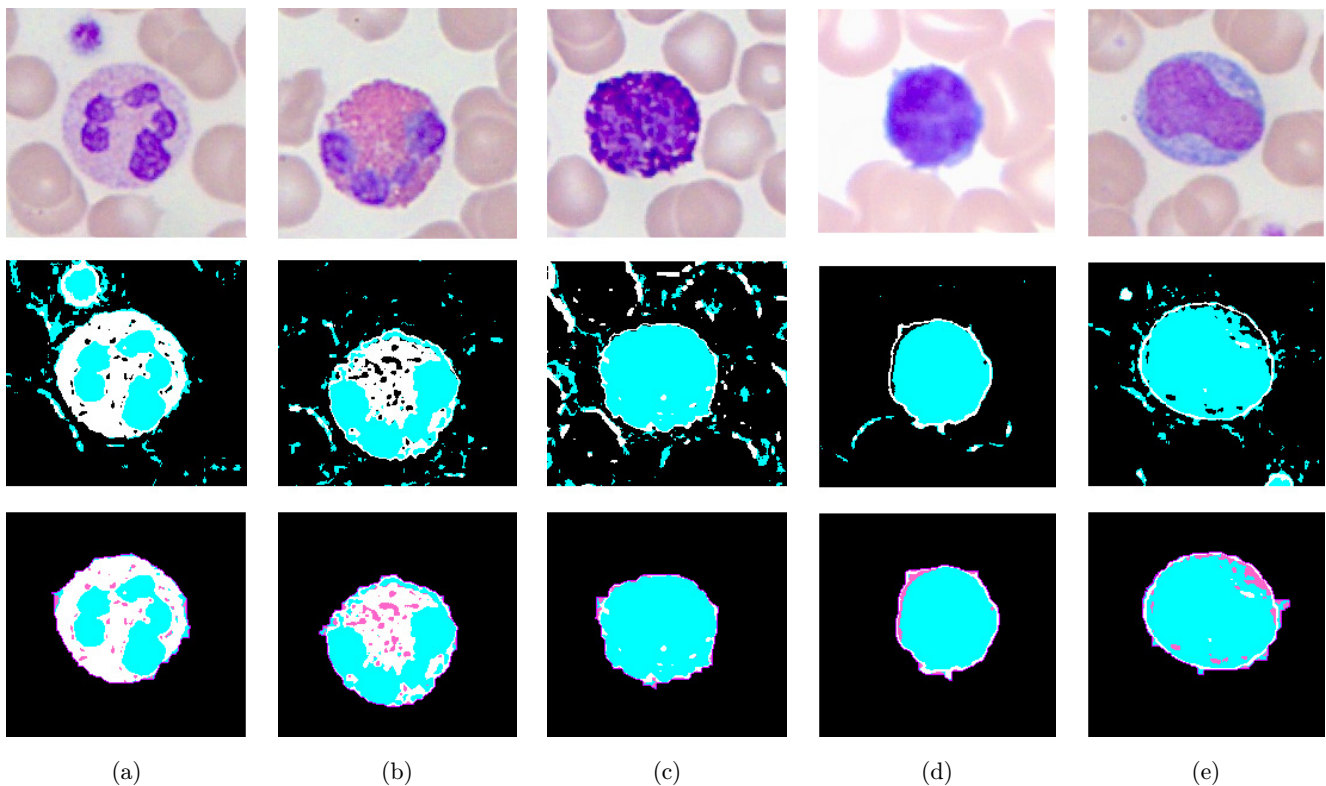

Fig. 13. Five types of leukocyte are segmented by HSI color system and shape filter. (a) Neutrophil, (b) eosinophil, (c) basophil, (d) lymphocyte and (e) Monocyte.

Table 3. Common geometric features of WBC.

Types	Neutrophil	Eosinophil	Basophil	Lymphocyte	Monocyte
NSN	2–5	2–3	1	1	1–2
Area (pixels)	8503–11856	7725–12622	7636–10789	4055–7768	7439–15170
Circularity	0.25–0.52	0.28–0.55	0.30–0.50	0.30–0.72	0.31–0.52
RNC	0.2–0.5	0.3–2.0	2.0–8.6	0.5–6.0	1.0–4.0

Table 4. Texture features of WBC extracted by ICT.

Properties/types	Neutrophil	Eosinophil	Basophil	Lymphocyte	Monocyte
$\bar{h}_{\max}$	92.30–110.12	98.74–115.33	131.52–162.15	135.14–165.25	68.20–112.91
$\bar{s}$	6.82–8.01	2.01–0.4.21	3.32–0.6.55	6.89–13.84	7.03–16.77
$v1$	23.23–30.43	21.45–24.19	19.12–23.81	9.2–15.46	22.56–37.94
$v2$	56.01–73.90	30.21–47.27	45.00–73.45	32.34–66.73	28.40–36.38
n	125–163	154–187	126–132	14–87	126–215

4.2. *Extraction for texture features of WBC by ICT*

As stated above, geometric features are not effective enough to identify WBCs precisely, so texture features are necessary to be added to increase the correct recognition rate and the robustness of machine. 1310 sample images of WBCs have been used to extract texture features by ICT, and the result is shown in Table 4. Although some parameter values partially overlap, others are effective to distinguish the WBCs.

4.3. *Leukocytes recognition*

This is the last step of leukocytes recognition. SVM and ANN are frequently used as classifier in the recognition of WBC, however, ANN is less complex than SVM, and it works in a manner that can be trained to recognize something like human brain, so ANN is applied.

5. Result and Analysis

5.1. *Hard-ware and computer platform*

The digital images of blood smears are captured by using a microscope imaging system at $100 \times$ magnification and a USB CMOS camera with 1024×768 pixels. In term to ensure that less time wasted in user interface and more time was spent in optimizing codes, our program code was written by BCB C++, avoiding popular but slow Java or C#

language with gigantic runtime packages. All the programs were run in a computer typed Lenovo T61 equipped with Intel Core 2.4G CPU and 2G RAM.

5.2. *Software flow chart and architecture*

The soft-ware flow chart and architecture is shown in Fig. 14, from WBC location and segmentation to features extraction, and recognition.

5.3. *Recognition rate*

As shown in Fig. 15, the features (properties) are listed in top right, and the details of procedure in bottom right, and the segmented parts of a WBC are shown in the middle.

The experiment results without using ICT and just with geometric and color features are shown in Table 5. The correct recognition rate of the neutrophil is fairly high above 94.6%, while the lymphocyte and the basophil are under 83%. Because the neutrophils have remarkable geometric features, low RNC and high NSN [see Table 3], it is not very hard to identify them, but the geometric features and color features of lymphocytes and basophils are similar, thus some lymphocytes with fewer cytoplasms were recognized as basophils and some big lymphocytes were identified as monocytes. At the same time, some small monocytes were also regarded as lymphocytes for the same reason. The monocyte is a special type of WBC for its geometric
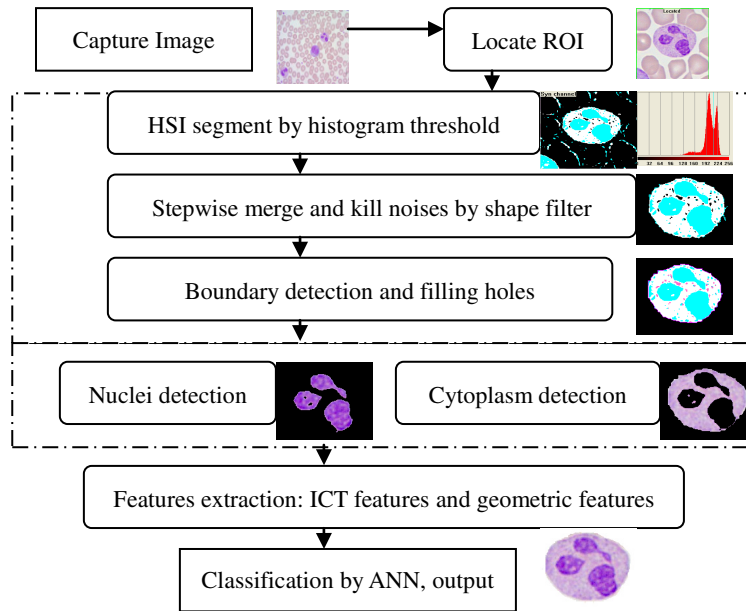
Fig. 14.   Software flow chart and architecture.

features [see Table 3] overlapping with other four types of WBC, and would be wrongly recognized as any WBC.

As seen in Table 6, the correct recognition rate is improved markedly when ICT features are added, especially basophil. Fifteen samples were completely recognized, because basophil always has some deep color granules which can be clustered as high islands with small area by ICT. The correct recognition rate of lymphocyte is lifted from 83.0% to 93.1%, because the count of islands of lymphocyte can help to identify them. The increasing correct recognition
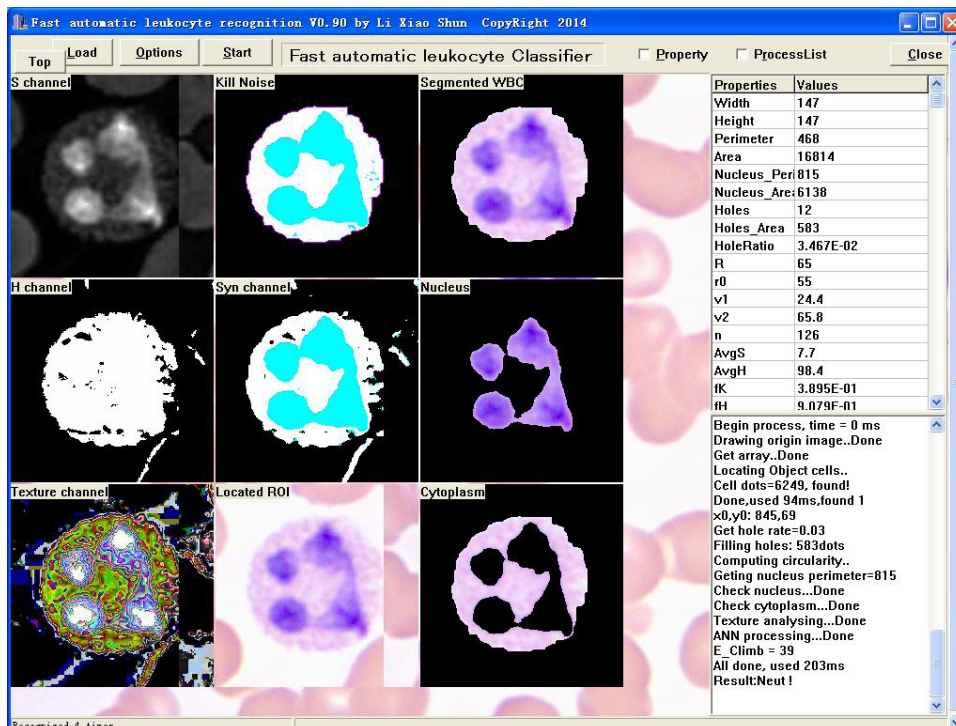

Fig. 15.   Software interface: leukocyte classifier.

Table 5.   Recognition rate with geometric and color features.

Types	Neutrophil	Eosinophil	Basophil	Lymphocyte	Monocyte	Others
Total mount	452	128	15	335	380	11
Right	428	117	12	278	345	6
Wrong	24	11	3	57	35	5
Correct recognition rate	94.6%	91.4%	80.0%	83.0%	90.8%	54.5%

Table 6.   Recognition rate with ICT.

Types	Neutrophil	Eosinophil	Basophil	Lymphocyte	Monocyte	others
Total mount	452	128	15	335	380	11
Right	431	120	15	312	351	10
Wrong	21	8	0	22	29	1
Correct recognition rate	95.4%	93.8%	100.0%	93.1%	92.3%	91.0%

rate of monocyte is not more than lymphocyte, because the texture of monocyte is very like lymphocyte and confuses us in some cases and needs to be identified by specialists. Sometimes, monocytes have some little granules like neutrophils and thus are prone to error. At last, the nonWBC images [see Fig. 16] which are very susceptible to error now can be recognized rightly, only one nonWBC is recognized as basophil for their textures are similar [see Fig. 16(a)].

Table 7 shows the comparison results of different methods, where the results of former three methods are just originally cited in corresponding literatures,[1,11] and,[19] the parameters of ICT are shown in Table 4. As seen in Table 7, the accuracy rate of basophil is similar to GLCM, LBP and ICT, but the ICT is the best one in other four types of WBC recognition. When the sample count is enlarged to 1310, the accuracy rates of ICT reduce slightly, because in large random samples, there are some

Table 7.   The counts of sample images and accuracy rate comparison.

Methods	GLCM[1]	GLCM[11]	LBP[19]	ICT	
Samples	54	140	150	150	1310
Neutrophil	82.5%	81.3%	96.43%	96.6%	95.4%
Eosinophil	76.79%	98%	100.00%	100.0%	93.8%
Basophil	97.73%	98%	98.64%	100.0%	100%
Lymphocyte	83.06%	84.3%	93.10%	95.0%	93.1%
Monocyte	76.79%	93.3%	95.83	97.7%	92.3%

fuzzy images which are very hard to be recognized indeed and do not appear in small samples generally. Nevertheless, it shows that the ICT is robust enough. Indeed, due to the rareness of basophil, the result of basophils may not well reflect the ground truth. We have been collecting more basophil samples to do further testing.

In short, ICT is not a powerful tool which can solve all the puzzles because there always are some cases where the texture features of lymphocyte, basophil and monocyte are very similar. However, it is true that ICT can solve some cases which are difficult with general methods, and ICT can increase the robustness of system and the correct recognition rate of WBC.

(a)　　　(b)

Fig. 16.   NonWBC images (a) recognized as basophil and (b) recognized rightly.

6.   Conclusion

This paper presents an image processing system based on texture analysis which can recognize five types of leukocytes in human peripheral blood automatically. Compared with previous works, HSI
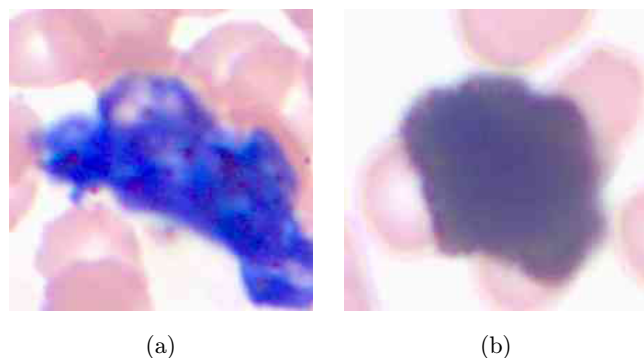
color system-based segmentation and geometric features of leukocyte extracting are the same, while the texture features of WBC are extracted not by commonly used GLCM or LBP, but by a new texture model ICT which is tested and proved to be more simple, effective and robust.

Finally, a complete software frame and hardware configuration has been realized and presented. A total of 1310 sample images including some fuzzy ones which can only be identified by specialists were used to evaluate this proposed method. Relatively high correct recognition rate above 92.3% has been achieved.

Acknowledgment

References

1. D. M. U. Sabino, "A texture approach to leukocyte recognition," *Real-Time Imaging* **10**, 205–216 (2004).

2. U. Neugebauer, J. H. Clement, "Identification and differentiation of single cells from peripheral blood by Raman spectroscopic imaging," *J. Biophotonics* **3**, 8–9 (2010).

3. M. Mihailescu, J. Costescu, "Diffraction pattern study for cell type identification," *Opt. Express* **20** (2), 1465 (2012).

4. E. A. Mohammed, M. M. Mohamed, B. H. Far, C. Naugler, "Peripheral blood smear image analysis: A comprehensive review," *J. Pathol. Inform.* **28**, 5–9 (2014).

5. P. Em Pavlova, "Application of HSV colour system in identification by colour of biological objects on the basis of microscopic images," *Comput. Med. Imaging Graph.* **5**, 357–364 (1996).

6. P. Chen, S. P. Dong, Y. Sook, C. Y. Ju, "Leukocyte image segmentation using simulated visual attention," *Expert Syst. Appli.* **39**, 7479–7494 (2012).

7. F. Scotti, "Robust segmentation and measurements techniques of white cells in blood microscope images," *Conf. Rec. IEEE Instrum. Meas. Technol. Conf.* pp. 43–48 (2006).

8. B. C. Ko, J.-W. Gim, J.-Y. Nam, "Automatic white blood cell segmentation using stepwise merging rules and gradient vector flow snake," *Micron* **42**, 695–705 (2011).

9. E. A. Mohammed, M. M. Mohamed, C. Naugler, B. H. Far, "Application of support vector machine and k-means clustering algorithms for robust chronic lymphocytic leukemia color cell segmentation," *Proc. 15th IEEE Int. Conf. e-Health Networking, Application and Services* pp. 622–626 (2013).

10. V. Piuri, "Morphological classification of blood leucocytes by microscope images," *IEEE Int. Conf. Comput. Intell. Meas. Syst. Appl.* pp. 103–109 (2004).

11. D.-C. Huanga, K.-D. Hunga, Y.-K. Chanb, "A computer assisted method for leukocyte nucleus segmentation and recognition in blood smear images," *J. Syst. Softw.* **85**, 2104–2118 (2012).

12. P. S. Hiremath, B. Parashuram, G. Sai, "Automated identification and classification of white blood cells (Leukocytes) in digital microscopic images," IJCA Special Issue on Recent Trends in Image Processing and Pattern Recognition, RTIPPR, pp. 59–63 (2010).

13. M. Habibzadeh, A. Krzyżak, T. Fevens, "White blood cell differential counts using convolutional neural networks for low resolution images," *Artificial Intelligence and Soft Computing*, Lecture Notes in Computer Science, Vol. 7895, pp. 263–274 (2013).

14. L. A. Chris, B. Mulyawan, "Focused color intersection for leukocyte detection and recognition System", *Int. J. Inf. Electron. Eng.* **3**, 498–501 (2013).

15. F. Chastine, L. T. Martin, R. W. Muhammad, "Parameter optimization of local fuzzy patterns based on fuzzy contrast measure for white blood cell texture feature extraction," *J. Adv. Comput. Intell. Intell. Inform.* **16**, 412–419 (2012).

16. R. M. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE* **67**(5), 786–804 (1979).

17. T. Ojala, M. Pietikainen, T. Maenpaa, "Multiresolution gray scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 971–987 (2002).

18. H. Mehdi, K. Adam, "Analysis of white blood cell differential counts using dual-tree complex wavelet transform and support vector machine classifier," *Computer Vision and Graphics*, Lecture Notes in Computer Science, Vol. 7594, pp. 414–422 (2012).

19. S. H. Rezatofighi, K. Khaksari, H. Soltanian-Zadeh, "Automatic recognition of five types of white blood cells in peripheral blood," *Comput. Med. Imaging Graph.* **35**, 333–343 (2011).

20. D. Comaniciu, P. Meer, "Mean-shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Recogn. Mach. Intel.* **24**, 603–619 (2002).

21. N. A. Otsu, "Threshhold selection method from gray-level histograms," *IEEE Trans. Syst. Man Cybern.* **9**, 62–66 (1979).