Cytometry
PART A
Journal of the
International Society for
Advancement of Cytometry

A Color and Shape Based Algorithm for Segmentation of White Blood Cells in Peripheral Blood and Bone Marrow Images

Salim Arslan,[1] Emel Ozyurek,[2,3] Cigdem Gunduz-Demir[1]*

• **Abstract**

Computer-based imaging systems are becoming important tools for quantitative assessment of peripheral blood and bone marrow samples to help experts diagnose blood disorders such as acute leukemia. These systems generally initiate a segmentation stage where white blood cells are separated from the background and other nonsalient objects. As the success of such imaging systems mainly depends on the accuracy of this stage, studies attach great importance for developing accurate segmentation algorithms. Although previous studies give promising results for segmentation of sparsely distributed normal white blood cells, only a few of them focus on segmenting touching and overlapping cell clusters, which is usually the case when leukemic cells are present. In this article, we present a new algorithm for segmentation of both normal and leukemic cells in peripheral blood and bone marrow images. In this algorithm, we propose to model color and shape characteristics of white blood cells by defining two transformations and introduce an efficient use of these transformations in a marker-controlled watershed algorithm. Particularly, these domain specific characteristics are used to identify markers and define the marking function of the watershed algorithm as well as to eliminate false white blood cells in a postprocessing step. Working on 650 white blood cells in peripheral blood and bone marrow images, our experiments reveal that the proposed algorithm improves the segmentation performance compared with its counterparts, leading to high accuracies for both sparsely distributed normal white blood cells and dense leukemic cell clusters. © 2014 International Society for Advancement of Cytometry

• **Key terms**

cell segmentation; white blood cells; leukemia; blasts; peripheral blood images; bone marrow images; marker-controlled watersheds; microscopy

TRADITIONAL techniques for diagnosis of leukemia require a careful observation of peripheral blood and bone marrow samples under a microscope, which relies on the morphology of white blood cells. This process is time consuming and greatly depends on the skills and experience of an expert. Although there exist more sophisticated techniques (such as surface antigen analysis made by flow cytometry) that produce more precise results (1,2), these techniques drastically increase the cost of diagnosis and treatment monitoring. Computerized systems that provide an automated image-based framework (3,4) alleviate the effects of human factor, and thus, have potential to increase the reproducibility and throughput of the assessments and help the expert in decision-making. The first step of these systems is usually cell segmentation, in which individual cells are separated from each other and the background. After this step, it is possible to extract properties from the segmented cells and distinguish leukemic cells from normal ones based on their properties. Therefore, it is crucial for the segmentation algorithm to accurately work on both normal and leukemic cells for precise decision-making. The focus of this article is implementing such a robust cell segmentation algorithm for peripheral blood and bone marrow images.

ISAC
International Society for Advancement of Cytometry
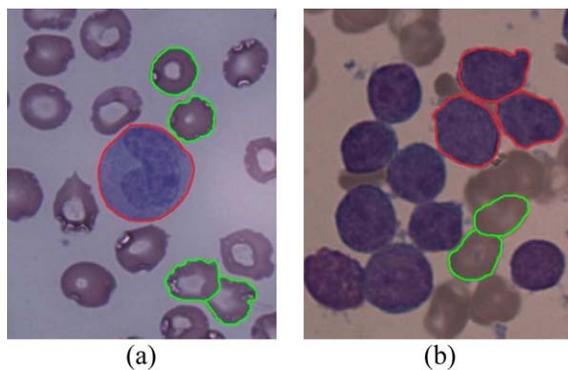
Figure 1. Example peripheral blood and bone marrow images containing (**a**) a normal white blood cell and (**b**) leukemic cells (both are indicated with red). Red blood cells also exist in both (a) and (b), some of which are indicated with green. These images were taken using a 100× objective lens and cropped from the original images. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Like in many segmentation problems, a successful solution of cell segmentation requires incorporating domain specific characteristics of images into the segmentation algorithm. Peripheral blood and bone marrow images of patients with acute lymphoblastic leukemia usually consist of normal white blood cells, blasts, red blood cells, platelets, and background pixels. Figure 1 demonstrates two sample images containing different types of cells, each of which has its own characteristics. Blue-like color of white blood cells usually makes them easy-to-differentiate from red blood cells and the background. However, red blood cells abutting these cells might produce some false positives since their boundary regions may show similar coloration (Fig. 1a). In this case, shape information is useful, since the shapes of white blood cells are more regular and compact compared with red blood cells, which look like thick-edged rings. Moreover, normal white blood cells (Fig. 1a) are sparsely distributed across the background whereas leukemic cells (Fig. 1b) tend to grow over each other forming cell clusters. The shape information is also useful to decompose these leukemic cell clusters into individual cells. The definition and use of different transformations that represent color and shape information constitute the main motivation of our proposed algorithm.

In the literature, there exist several algorithms for segmenting white blood cells in peripheral blood and bone marrow images. When white blood cells are sparsely distributed over the background and do not have adjacent red blood cells, straightforward methods such as thresholding (5–7), edge detection (8–10), and region growing (11) are usually sufficient for delineating cell boundaries. When red blood cells or other noisy fragments are also found in the results, it is possible to refine these results by applying morphological operations (12–14) and/or evolving active contours on the segmented cell boundaries (15–17). After eliminating such fragments, one can also reconstruct the boundaries of white blood cells by spiral interpolation or using color differences between overlapping cells (18,19).

Color-based segmentation is another widely used method for segmenting white blood cells. Unsupervised clustering algorithms (20–22) and supervised classification models (23–26) have been used for this purpose. These methods usually work on the RGB color space; however, there also exist some studies that work on other color spaces [such as the HSV (21,27) and La*b* (28) color spaces] for better quantifying visual differences between different image components. Additionally, it is possible to preprocess images with different techniques including filtering (29,30), histogram equalization (31), and contrast enhancement (32) for noise elimination.

These previous methods implemented their algorithms mainly focusing on segmentation of isolated white blood cells, which are typically normal. However, leukemic cells tend to grow in over layers forming cell clusters. These clusters should be decomposed into single cells for accurate segmentation. Marker-controlled watershed algorithm is a powerful image processing technique that is widely used to separate clumped cells in dense cellular images (33–35). Some variations of these algorithms are also adapted to solve the white blood cell segmentation problem. These methods usually employ distance transformations (36,37) for marker detection and use gradients/intensities for marking function definition (38). However, they usually yield oversegmented cells and/or irregular cell boundaries, which are refined afterwards (37,39). These methods address the problem of segmenting leukemic cells to an extent (36–38), but there still remain challenges to overcome when predefined markers do not represent cells accurately. For white blood cell segmentation, this problem arises when an image contains highly over-layered leukemic cells with fuzzy boundaries and aggravates if the image also contains confluent red blood cells adjacent to white blood ones.

In this article, we devise and implement a new algorithm for segmenting white blood cells in peripheral blood and bone marrow images. In this algorithm, we propose to model color and shape characteristics of white blood cells through transformations and use these transformations in a marker-controlled watershed algorithm for segmentation. The main contribution of this article is the definition of these transformations and their efficient use in a segmentation algorithm. Particularly, we define a simple but an efficient color transformation to better reveal chromatic characteristics of white blood cells. This transformation successfully suppresses background pixels while preserving cell pixel intensities. We then define a two-stage procedure on this new color space for calculating a distance transformation, which represents shape characteristics of white blood cells. We finally use these two transformations to define markers and a marking function of the watershed algorithm. Working on images that contain both isolated normal white blood cells and clustered leukemic cells, our experiments reveal that the proposed transformations and their use in a watershed algorithm yield more precise segmentation results compared to previous algorithms.

The remainder of this article is organized as follows. In Section "Methodology," we explain our methodology giving details of the segmentation stages. In Section "Experiments," we describe the experimental setup and evaluation algorithms for performance assessments, and concisely explain the methods we use in our comparisons. In Section "Results," we
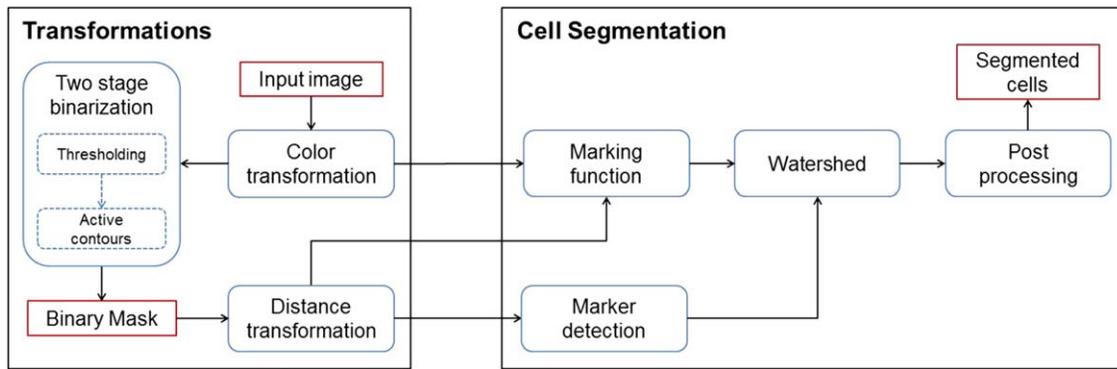
Figure 2. Flowchart of the proposed segmentation algorithm. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

present visual and quantitative results of our proposed algorithm as well as those of the comparison methods. Finally, in Section "Conclusions," we conclude the article and discuss its future work.

METHODOLOGY

Our algorithm relies on combining shape and color information in a marker-controlled watershed algorithm for segmenting white blood cells. To this end, it defines two transformations based on the shape and color characteristics of the white blood cells and uses these two transformations to define markers and the marking function of the watershed algorithm. The flowchart of our segmentation algorithm is given in Figure 2; the details of its steps are explained in the following subsections.

Transformations

Color transformation. Our observations on peripheral blood and bone marrow images show that simply reducing an RGB image into grayscale yields poor segmentation results since the contrast between foreground and background pixels in the grayscale is typically not sufficient to classify the pixels precisely. However, each color band of an RGB image comprises great amount of information about their specific characteristics. In a typical peripheral blood and bone marrow image, there exists high amount of contrast between fore-

ground and background pixels of the blue band, which facilitates finding a good threshold level for separating these pixels. More importantly, white blood cells in the green band have darker intensities compared with the other cellular objects. To incorporate these image characteristics into our segmentation algorithm, we propose to transform an RGB image into a new intensity map based on its green and blue bands.

Let I_B and I_G be the blue band and green band of the RGB image, respectively, defined in the 2D domain $\Omega \rightarrow \mathbb{Z}^2$, where $(x, y) \in \Omega$ represents the x and y coordinates of a pixel. The intensity map I_M is defined on Ω as follows:

$$g(x, y) = I_B(x, y) - I_G(x, y)$$

$$I_M(x, y) = \begin{cases} g(x, y), & g(x, y) > 0 \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

For an example subimage, Figure 3 demonstrates the blue band I_B, the green band I_G, and the intensity map I_M obtained via the transformation. As shown in this figure, the transformation in Eq. (1) makes the pixels of white blood cells more distinguishable than those of the other objects and the background. Note that this transformation also makes the pixel values of the other objects and the background close to each other, which facilitates to differentiate white blood cell pixels. We will use the intensity map obtained by this
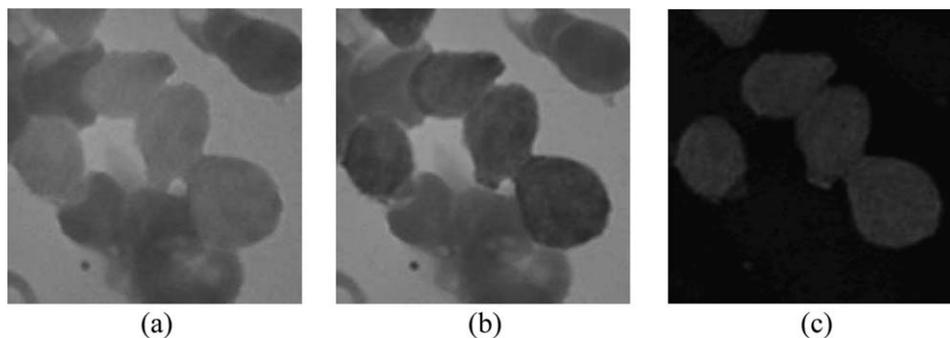
Figure 3. For an example subimage, (**a**) the blue band I_B, (**b**) the green band I_G, and (**c**) the intensity map I_M obtained via our color transformation.
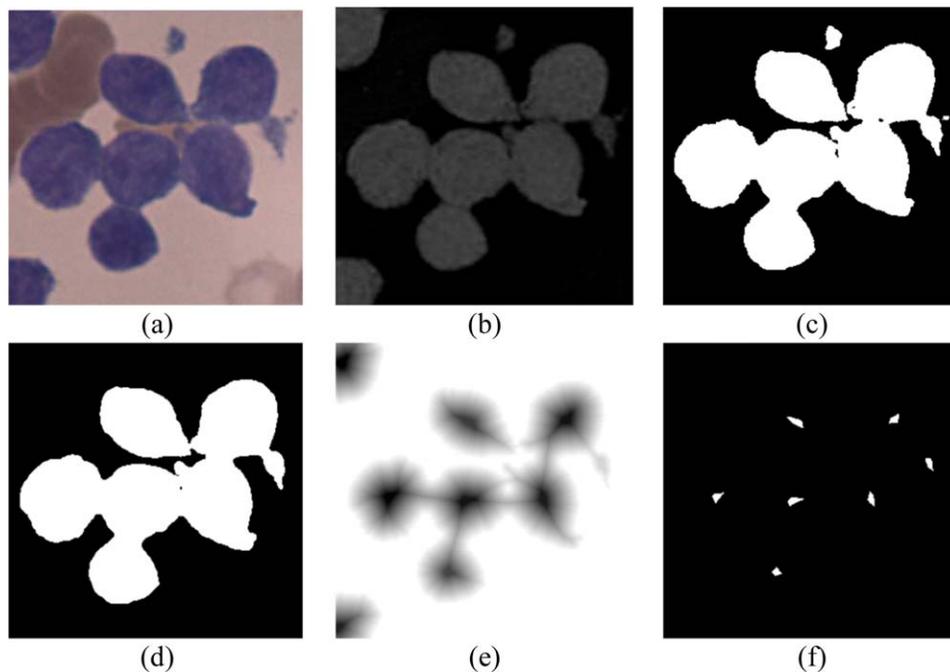
Figure 4. Marker identification step: (**a**) original subimage, (**b**) intensity map after color transformation, (**c**) coarse mask obtained by thresholding, (**d**) refined mask obtained after active contours, (**e**) inverse of the inner distance transform map after minima suppression, and (**f**) identified markers, each of which represents a single cell's location. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

transformation to define the distance transformation as well as the marking function of the watershed algorithm.

Distance transformation. We calculate the distance transformation on a binary mask that is obtained processing the intensity map I_M. To obtain this binary mask, we implement a two-stage segmentation algorithm. In the first stage, we threshold the intensity map by the Otsu's method (40). This stage provides us with a coarse mask where cellular regions and the background are clearly separated from each other. Nevertheless, the cellular regions may contain both white blood cells and other false positives. Therefore, we run a second stage to refine the boundaries of the cellular regions. In this second stage, we refine the boundaries by applying active contours without edges (41), in which we evolve an active contour by minimizing an energy function on the intensity map I_M. To find a threshold level for the initial spline of the active contour, we make use of the coarse mask. Here we obtain the initial spline by again applying the Otsu's method on the intensity map, but this time by considering only the pixels delimited by the coarse mask. We also evolve the active contour only within the boundaries of the coarse mask since the previous stage successfully separates the pixels of cellular regions and the background.

After obtaining the binary mask, we transform it into a distance map by inner distance transformation, which computes the minimum Euclidean distance from every foreground (cellular region) pixel to a background pixel. Formally, let F and B be two closed sets of foreground and background pixels

in the mask and $p = (x_p, y_p)$ and $q = (x_q, y_q)$ be arbitrary points selected on F and B, respectively. The inner distance transformation $\Delta(p, F)$ is defined as

$$\Delta(p, F) = \min_{q \in B}\{d(p, q), \quad p \in F\} \quad (2)$$

where $d(p, q)$ is the Euclidean distance between p and q. Note that we will use the distance map in defining both markers and the marking function of the watershed algorithm.

Cell Segmentation

The proposed cell segmentation method delineates cell boundaries by using a priori color and shape information obtained from the intensity and distance maps and combining them in a marker-controlled watershed algorithm. To this end, it first defines a set of markers on the distance map, where each marker corresponds to an estimated location of a single cell and then grows these markers using the marking function, which is defined as a combination of the intensity and distance maps, in the flooding process of the watershed. At the end, it postprocesses the grown markers to eliminate false positive cells and smooth the segmented cell boundaries. The details are explained in the following subsections.

Marker identification. Identifying markers constitutes the core step in a marker-controlled watershed algorithm since these markers correspond to cell locations, from which flooding starts. Defining markers more than actual cells

leads to oversegmentation whereas defining a single marker for a cell clump yields undersegmentation. In our algorithm, we first apply h-minima transform on the inverse of the inner distance map to suppress undesired minima and reduce the likelihood of identifying false markers. The transform eliminates all minima whose depth is smaller or equal to the threshold h. Here, h is a model parameter and should be selected based on the cell characteristics. After applying h-minima transform, we identify the remaining regional minima as the cell markers. It is worth to noting that in our algorithm, we use h-minima transform beforehand instead of directly finding the regional minima, since otherwise would result in oversegmentation due to high amount of spurious minima. For an example subimage, Figure 4 demonstrates marker identification. Note that this figure shows only the markers defined for a single connected component of the binary mask since the others are not seen in their entirety in this cropped subimage.

Marking function. The marking function in a watershed algorithm represents the topographic surface where the water rises and determines the watershed lines at the locations where two floods meet. These watershed lines will correspond to cell boundaries; therefore, the marking function should reflect the image characteristics for an accurate segmentation. With this motivation, we define a new marking function that combines the color and shape characteristics of white blood cells through two transformations we defined. Let R be the regional minima in the inverse of the distance map Δ and I_M be the intensity map, both defined in the 2D domain $\Omega \rightarrow \mathbb{R}^2$, where $(x,y) \in \Omega$. First, for every pixel, we define our marking function $\Phi(x,y)$ by multiplying their values in R and I_M, as also given below.

$$\Phi(x, y) = R(x, y) \cdot I_M(x, y) \qquad (3)$$

Then considering the markers as the starting points, we use this marking function in the flooding process of the watershed and obtain the watershed lines. Finally, we obtain the segmented cells by superimposing the binary mask onto these watershed lines. This process is illustrated in Figure 5.

Note that it is common to define marking functions on only the binary mask and its distance transform. In that case, floods always meet at the equidistant points from the markers due to the distance transform definition. This however might lead to inaccurate (and typically jagged) boundaries between two adjacent cells especially when their sizes and shapes are different. Moreover, watershed lines (and thus cell boundaries) are highly dependent on the positions and shapes of the markers. This causes problems when the markers are not precisely located, which is indeed mostly the case. However, we also incorporate the intensity map into our marking function definition. This makes the marking function less dependent on the markers' shape and positions as well as the assumption of having cells of the same size and shape. For an example subimage, Figure 6 compares cell boundaries obtained when only a distance map is used and when a combination of distance and intensity maps is used. This figure shows that combining these two maps yield more natural boundaries. Here note that this use improves segmentation for the boundary pixels. Since the number of these pixels is much lower compared to that of all cell pixels, it only slightly changes the quantitative error rates that we use in our experiments. However, finding more natural boundaries might be important especially in some applications (e.g., when morphological features are defined on the boundaries to quantify the segmented cells).

Postprocessing. Peripheral blood and bone marrow images do not only comprise white blood cells; they also contain red blood cells, segmentation of which is beyond the scope of this article. Since their outer regions show similar color characteristics with white blood cells, the watershed segmentation may partially find red blood cells (mostly the outer regions of these cells). Thus, the shape characteristics of the segmented white and red blood cells are different from each other. In our algorithm, we make use of this difference to eliminate the falsely segmented red blood cells. To this end, we identify narrow components and eliminate them from the results. For

(a) (b) (c) (d)
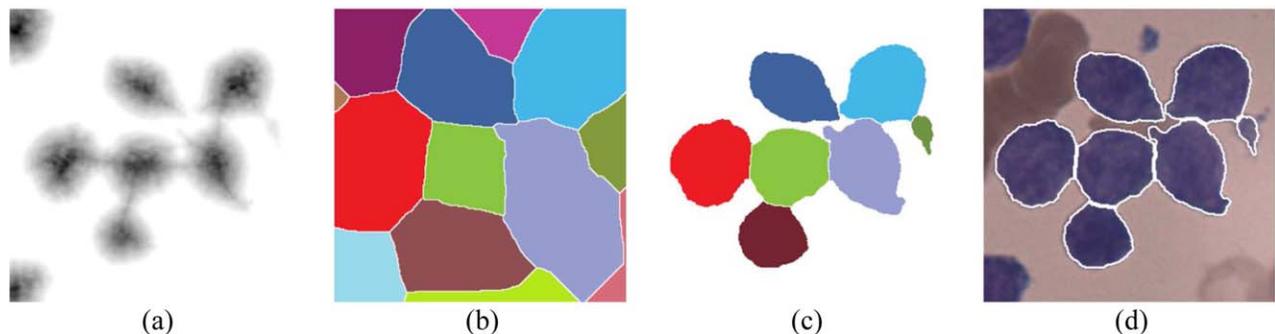
Figure 5. Marking function definition and watershed segmentation steps: (**a**) topographic surface used as the marking function, (**b**) watershed lines, (**c**) segmented cells after superimposing the binary mask on the watershed lines (cells found on other connected components are not shown here as they are not seen in their entirety), and (**d**) segmented cell boundaries superimposed on the original subimage. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

identification, we employ the circle-fit algorithm (42), in which circles are iteratively located provided that their radii are greater than a circle threshold r_{thr}. We then keep the segmented cells on which the circle-fit algorithm can locate at least one circle; we eliminate the rest from the results. After this elimination, we apply majority filtering on the remaining cells to smooth their boundaries. Here, we use a circle-shaped kernel with a radius of W. For an example subimage, Figure 7 shows the results obtained by the steps of postprocessing.

EXPERIMENTS

Dataset

The samples were obtained from bone marrow and peripheral blood smears of four children. Two of them were diagnosed as acute lymphoblastic leukemia and the other two had iron deficiency anemia. After smearing peripheral blood and bone marrows on lamella, the slides were dyed by Wright's stain with standard methods. The data were collected under the full consent of the patients' legal guardians. This consent explicitly includes the use of the samples for scientific study. To properly handle the data, the identifiers of the patients were completely removed and slides were numerically recoded corresponding to their diagnoses by the hematologist, prior to obtaining their digital images. Thus, two nonmedical investigators in this work had access to images, without retraceable personal identifiers. The images of these slides were taken under a Nikon 50i microscope with a digital camera using a 100× objective lens.

We conduct our experiments on a total of 650 cells in 31 peripheral and bone marrow images. The first set comprises 15 peripheral normal blood images containing 26 normal white blood cells. The second set comprises the remaining 16 images containing 624 leukemic cells. Images in the first set contain only a few white blood cells since normal ones are very sparsely distributed. However, images in the second set contain dense leukemic cell clusters due to the fact that leukemic cells tend to reproduce rapidly growing over each other. This dense and confluent structure of leukemic cells increases the difficulty of segmentation in the second set. Moreover, both sets also include red blood cells surrounding the white ones. This adjacency between red and white blood cells is another factor that increases the segmentation difficulty. Note that to evaluate the robustness of our proposed algorithm, we test it on the images containing both normal and leukemic cells with the same parameter configuration and observe the segmentation performance.

Evaluation

We evaluate the performance of our proposed algorithm both visually and quantitatively. For quantitative evaluation, we use two assessments: cell-based and boundary-based. The cell-based assessment evaluates the performance in terms of the number of correctly identified cells. To this end, we first find the correct segmentations, which correspond to one-to-one matches between segmented and annotated cells. For that, we first match a segmented cell with an annotated one if at least half of its pixels overlap with those of the annotated cell,
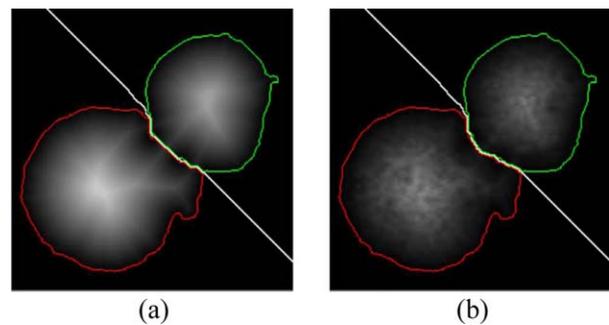
Figure 6. Cell boundaries obtained (**a**) when only a distance map is used and (**b**) when a combination of distance and intensity maps is used. Here, white lines correspond to the watershed ridge and colored lines indicate the segmented cells' boundaries. Note that the maps shown in these images are negated for better visualization. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and vice versa. Then, we consider an annotated cell A being correctly identified, if there is exactly one segmented cell C matching with A and if A also matches with this cell C. Using the number of one-to-one matches, we compute cell-based precision, recall, and F-score measures. To analyze the results better, we also calculate and report the number of oversegmentations, undersegmentations, false positives, and false negatives. An annotated cell is considered as oversegmented if it matches with more than one segmented cell. On the contrary, a segmented cell is considered as undersegmented if it matches with more than one annotated cell. A segmented cell is considered as a false positive if it does not match with any annotated cells and an annotated cell is considered as a false negative if it does not match with any segmented cells.

The boundary-based assessment evaluates the performance in terms of the precision of delineating the cell boundaries. It is important to note that it only evaluates this for the correctly identified cells, which correspond to one-to-one matches between annotated and segmented cells. For that, we compute the number of overlapping pixels of these annotated-segmented pairs, which correspond to true positives, and then calculate the boundary-based precision, recall, and F-score measures.

Parameter Selection

The proposed algorithm has three external model parameters: (1) the depth threshold h, which is used in h-minima transform to suppress spurious minima in marker identification step of the cell segmentation, (2) the circle radius threshold r_{thr}, which determines the radius of the smallest circle located on the segmented components in the postprocessing step, and (3) the radius W, which is the kernel size used by majority filtering in the postprocessing step. In our experiments, we selected them as $h = 2$, $r_{thr} = 20$, and $W = 4$. Note that we discuss the effects of the selection of these parameters to the segmentation performance in the following section.

Additionally, we have an internal parameter α to determine the smoothness of the active contour for obtaining a binary mask. Within a range of 0 to 1, its higher values yield
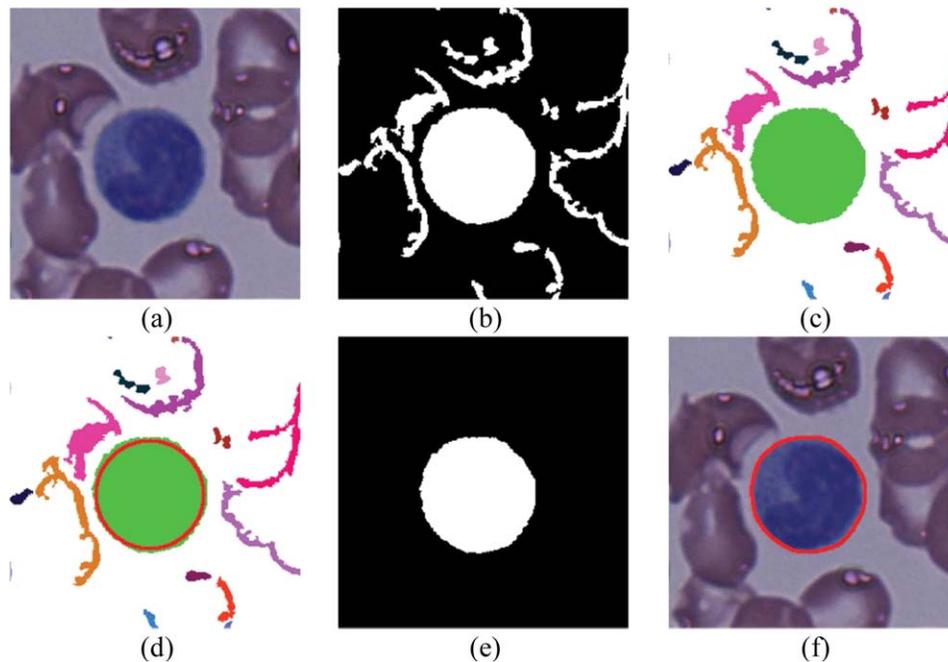
Figure 7. Postprocessing step: (**a**) original subimage, (**b**) binary mask, (**c**) result of the watershed algorithm (each component corresponding to a single segmented cell is shown with a different color), (**d**) circles located on the segmented components (indicated by red), (**e**) remaining cell after eliminating components on which no circles are located, and (**f**) smoothed cell boundary after majority filtering. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

smoother boundaries, causing some cellular regions to be excluded from the binary mask. However, its smaller values lead to more precise boundaries, resulting in a more accurate binary mask. In our experiments, we internally set the value of this parameter to 0.1; its smaller values do not improve the boundary quality further, but elevate the running time of the algorithm.

Comparisons

We compare our algorithm with two methods: color-based clustering [28] and conditional erosion [33]. Color-based clustering [28] is specifically implemented for cell segmentation in microscopic blood images. This method first enhances an RGB image with median filtering and unsharp masking, for removing noise and sharpening image details, respectively. Transforming the image into La*b* color space, it then clusters image pixels by running k-means on their a* and b* channels and identify those belonging to the clustering vector with the minimum b* value as white blood cell pixels. At the end, it fills the holes in the identified pixels and considers each connected component on these pixels as a white blood cell.

The conditional erosion method [33] is originally presented as a shape-based marker-controlled watershed algorithm for cell nucleus segmentation in fluorescence microscopy images. We adapt this method to our domain by providing it with the binary mask produced by our algorithm so that it is not negatively biased with the differences of fluorescence and peripheral blood images with respect to color and texture. This is a two-stage method to extract the markers of a watershed algorithm. In its first stage, it iteratively erodes each connected component of the mask with two coarse structuring elements until the size of each component drops under a predefined threshold. In its second stage, it repeats the same process but this time uses two finer structuring elements and a second threshold. Taking the eroded components obtained at the end of this second stage as the markers and using the distance transform of the binary mask as the marking function, it delineates cell boundaries by a marker-controlled watershed algorithm.

In our experiments, we use these two comparison methods to understand the importance of combining the color and shape information instead of using them alone. Note that the first comparison method makes use of only the color information without considering the shape information. The second one employs only the shape information without combining this information with color values. However, our proposed method combines the color and shape information for white blood cell segmentation.

Results

We present visual results obtained by our proposed algorithm as well as the comparison methods in Figure 8. Note that although the results are obtained on the original images, this figure shows the results on subimages cropped from some of these original images and rescaled for better visualization. This figure shows that our proposed algorithm identifies both normal and leukemic cells more accurately compared to the other methods.

These visual results are also consistent with the quantitative ones that we report in Tables 1 and 2. These tables
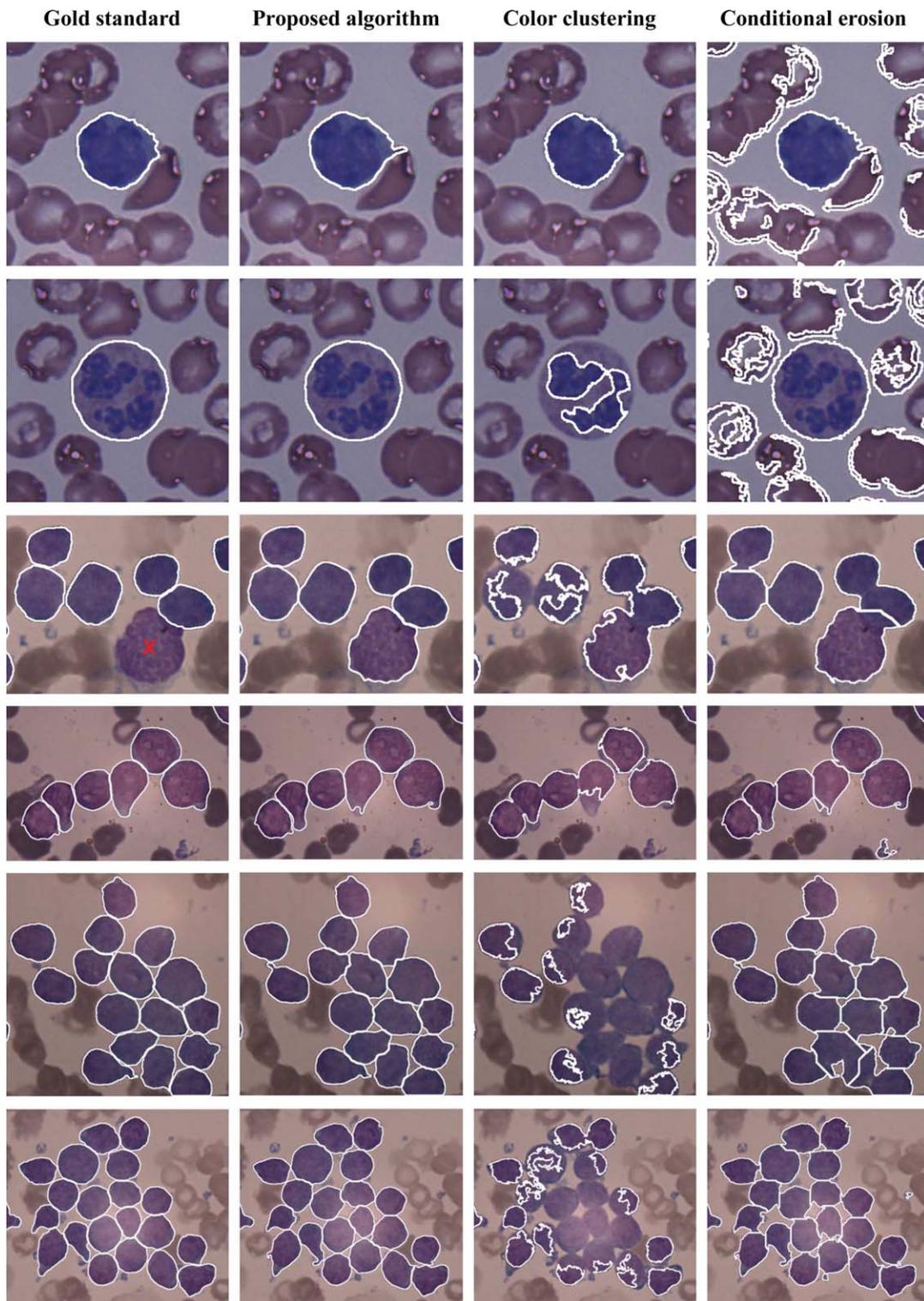
Figure 8. For example subimages, gold standards (first column) as well as the visual results obtained by our proposed algorithm (second column), the color clustering method (third column), and the conditional erosion method (fourth column). These images were taken using a 100× objective lens and cropped from the original images. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

reveal that the proposed algorithm identifies cells more correctly, producing more one-to-one matches and less oversegmentations, undersegmentations, false positives, and false negatives. As a result, it yields relatively high cell-based precision, recall, and F-score measures. This indicates the effectiveness of using color and shape characteristics in the

Table 1. Quantitative comparison of the algorithms in terms of the types of the annotated and segmented cells.

	ONE-TO-ONE	OVERSEGM.	UNDERSEGM.	FALSE POS.	FALSE NEG.
Proposed algorithm	637	6	3	31	3
Color clustering	352	6	58	231	147
Conditional erosion	610	32	9	2,638	6

marker identification and postprocessing steps. Besides, the boundary-based assessment results given in Table 2 shows that the proposed algorithm locates boundaries better for the correctly identified cells. This, however, shows the benefit of combining color and shape characteristics in defining the marking function.

Both the visual and quantitative comparisons indicate the usefulness of employing both color and shape characteristics of the cells in a segmentation algorithm. First, peripheral blood and bone marrow images contain both white and red blood cells and it is of great importance to differentiate these two types of cells for accurate segmentation. However, the use of only the color information for this differentiation usually leads to incorrect segmentations. For example, in the first two rows of Figure 8, the conditional erosion method falsely identifies lots of red blood cell pixels as belonging to white blood cells because it uses a binary mask obtained using only the color information. (Here it is worth to noting that this is the binary mask that is produced by our algorithm. However, the use of this binary mask yields much more accurate results compared to using the mask produced by the original conditional erosion method. When its original mask is used, false detections double its current value in the quantitative results.) On the contrary, when the algorithm is designed to eliminate such false positives using only the color information, this may also eliminate some true positives, as shown in the results of the color clustering method. As opposed to these methods, our proposed algorithm makes use of the color information in defining its binary mask and the shape information in postprocessing to eliminate the false positives. This leads to more accurate results.

Second, especially leukemic cell images contain touching and confluent cells (last four rows of Fig. 8) and they should be decomposed into individual cells. When we compare the results of our proposed algorithm and the conditional erosion method with those of the color clustering method, we observe that the use of shape information is very effective in this decomposition. However, when only the shape information is used, the boundaries between adjacent cells could be incorrectly located. This can be observed in the comparison of the

results of our algorithm, which uses both the color and shape information in defining its marking function, and the conditional erosion method, which uses just the shape information in the form of a distance transform for the same purpose. As results indicate, our algorithm gives more natural boundaries between the adjacent cells.

In our method, we use a simple color transformation $g(x,y)$, in which we subtract the green channel of a pixel from its blue channel. This color transformation indeed corresponds to defining a linear combination of the pixel's red, blue, and green channels, where the coefficient vector in this linear combination is $V = [0 \ -1 \ 1]^t$. In our experiments, we also investigate the consistency of this use with the image data. For that, we use manual segmentations to select white blood cell and background pixels in the images and apply the Fisher's discriminant analysis on the RGB values of these pixels. This analysis gives a vector $W = [0.1 \ -1 \ 0.6]^t$, which best separates the white blood cell and background pixels (note that we normalize the magnitudes of this vector such that its second entry becomes -1). Although the vectors W and V are not exactly equal to each other, they are still consistent in terms of their magnitudes. The further investigation of this analysis can be considered as a future work.

One limitation of our algorithm is its limited capability to eliminate some false positives. Although it gives promising results to eliminate most of the red blood cells or dead cell fragments in segmentation, it may fail for eliminating fragments that show similar coloration and shape with leukemic cells. In the third row of Figure 8, one example of such fragments is shown with a red cross on the gold standard image. As shown in the results, all of the algorithms incorrectly locate this fragment as a white blood cell. Such fragments might be eliminated by analyzing their textures; this could be considered as another future work of the proposed algorithm.

Parameter Analysis

We analyze the effects of the model parameters on the segmentation performance. For this purpose, for each parameter, we run our algorithm with its different values while fixing the values of the remaining ones and observe the changes

Table 2. Quantitative comparison of the algorithms in terms of the cell-based and boundary-based precision, recall, and F-score measures.

	CELL-BASED			BOUNDARY-BASED		
	PRECISION	RECALL	F-SCORE	PRECISION	RECALL	F-SCORE
Proposed algorithm	94.09	98.00	96.01	88.76	95.78	92.14
Color clustering	54.40	54.15	54.27	55.82	40.26	46.78
Conditional erosion	18.54	93.84	30.97	59.54	90.02	71.67
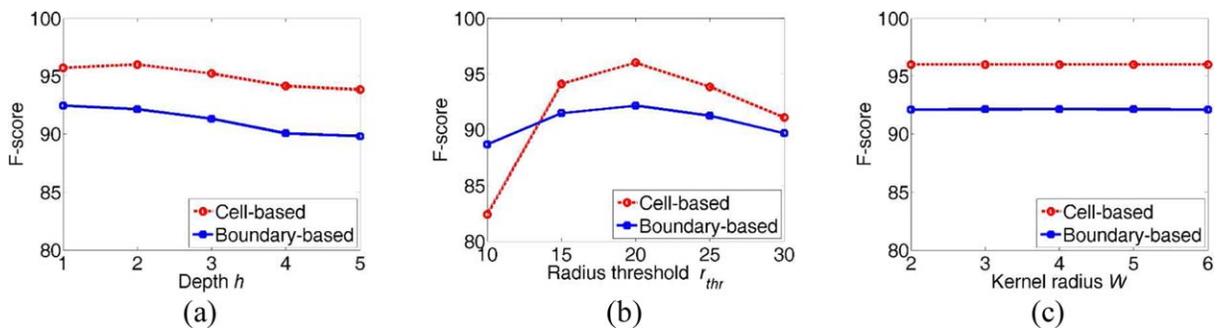
Figure 9. Cell-based and boundary-based F-score measures as a function of the (**a**) depth threshold h, (**b**) the circle radius threshold r_{thr}, and (**c**) the kernel radius W. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

in cell-based and boundary-based F-score measures. We present the results obtained for each parameter in Figure 9.

The first parameter is the depth threshold h, which is used by h-minima transform for suppressing undesired spurious minima in the marker identification step. Increasing the value of this parameter eliminates more minima than desired. This typically leads to undersegmentation, which decreases the F-score measures (Fig. 9a). On the other hand, using too small values may cause oversegmentations, which slightly decreases the F-scores in our experiments.

The second parameter is the circle radius threshold r_{thr}, which determines the size of the smallest circle located on the segmented components in the postprocessing step. This is the most important parameter that dramatically changes the segmentation performance. Since this parameter is directly related with white blood cells' size, it should be set according to the average size of these cells. Selecting too small values cannot successfully eliminate noisy components such as red blood cell fragments and thrombocytes (platelets), whose segmentation is out of the scope of this work. This increases the number of false positives and oversegmentations. However, selecting values larger than the size of a typical white blood cell results in also eliminating most of the white blood cells; this increases the number of false negatives. Both of these cases decrease the number of one-to-one matches, leading lower F-score measures as observed in Figure 9b.

The last parameter is the radius W, which determines the size of the kernel used by majority filtering to smooth the boundaries in the postprocessing step. As observed in Figure 9c, this parameter only slightly affects the segmentation performance.

Conclusions

In this article, we present a new algorithm for segmenting white blood cells in peripheral blood and bone marrow images. In our algorithm, we model color and shape characteristics of these cells by defining two transformations and efficiently employ these transformations in a marker-controlled watershed. The experiments show that this algorithm more accurately locates white blood cells and better delineates their boundaries compared to its counterparts. This improvement is attributed to the fact that the proposed algorithm makes use of the

domain specific color and shape information in its four core steps: First, it identifies white blood cell regions using an intensity map obtained by color transformation defined on the two bands of an image. Second, it locates initial cell locations using the shape information quantified in the form of distance transformation. Third, it delineates the cell boundaries combining the color and shape information in a marking function. Last, it uses the shape information imposed by the circle-fit algorithm in its postprocessing mechanism. The experiments show that these uses of color and shape information in its different steps make the algorithm more robust to segmenting both isolated normal white blood cells and confluent leukemic cell clusters.

As previously mentioned, one future work is to enhance the postprocessing step by also using texture information to eliminate false positive cells. For that, one could consider extracting texture features from the segmented cells and carrying out elimination based on these texture features. As another future work, one might consider classifying white blood cells into further subgroups based on the features extracted from the segmented cells.

Literature Cited

1. Weir EG, Borowitz MJ. Flow cytometry in the diagnosis of acute leukemia. Semin Hematol 2001;38:124-138.
2. Coustan-Smith E, Sancho J, Behm FG, Hancock ML, Razzouk BI, Ribeiro RC, Rivera GK, Rubnitz JE, Sandlund JT, Pui C, et al. Prognostic importance of measuring early clearance of leukemic cells by flow cytometry in childhood acute lymphoblastic leukemia. Blood 2002;100:52-58.
3. Swolin B, Simonsson P, Backman S, Löfqvist I, Bredin I, Johnsson M. Differential counting of blood leukocytes using automated microscopy and a decision support system based on artificial neural networks–evaluation of DiffMaster™ Octavia. Clin Lab Hematol 2003;25:139-147.
4. Münzenmayer C, Schlarb T, Steckhan D, Haßlmeyer E, Bergen T, Aschenbrenner S, Wittenberg T Weigand C, Zerfaß T. HemaCAM–A computer assisted microscopy system for hematology. In: Heuberger A, Elst G, Hanke R, editors. Microelectronic Systems. Heidelberg: Springer; 2011. p 233-242.
5. Rezatofighi SH, Soltanian-Zadeh H, Sharifian R, Zoroofi RA. A new approach to white blood cell nucleus segmentation based on Gram-Schmidt orthogonalization. In Proceeding of International Conference on Digital Image Processing, Bangkok, Thailand; 2009. p 107-111.
6. Wu J, Zeng P, Zhou Y, Olivier C. A novel color image segmentation method and its application to white blood cell image analysis. In Proceeding of 8th International Conference on Signal Processing, Beijing, China; 2006. Vol 2.
7. Liao Q, Deng Y. An accurate segmentation method for white blood cell images. In Proceeding of IEEE International Symposium on Biomedical Imaging, Washington, DC; 2002. p 245-248.
8. Scotti F. Automatic morphological analysis for acute leukemia identificaiton in peripheral blood microscope images. In Proceeding of IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, Giardini Naxos, Italy; 2005. p 96-101.
9. Sadeghian F, Seman Z, Ramli AR, Kahar BHA, 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.

10. Kumar BR, Joseph DK, Sreenivas TV. Teager energy based blood cell segmentation. In Proceeding of 14th International Conference on Digital Signal Processing, Santorini, Greece; 2002. p 619-622.

11. Chassery JM, Garbay C. An iterative segmentation method based on a contextual color and shape criterian. IEEE Trans Pattern Anal 1984;6:794-800.

12. Madhloom HT, Kareem SA, Ariffin H. An image processing application for the localization and segmentation of lymphoblast cell using peripheral blood images. J Med Syst 2012;36:2149-2158.

13. Sheeba F, Thamburaj R, Nagar AK, Mammen JJ. Segmentation of peripheral blood smear images using tissue-like P systems. In Proceeding of Sixth International Conference on Bio-Inspired Computing: Theories and Applications, Penang, Malaysia; 2011. p 257-261.

14. Piuri V, Scotti F. Morphological classification of blood leucocytes by microscope images. In Proceeding of IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, Boston, MA USA; 2004. p 103-108.

15. Ongun G, Halici U, Leblebicioglu K, Atalay V, Beksac M, Beksac S. An automated differential blood count system. In Proceeding of 23rd Annual EMBS International Conference, Istanbul, Turkey; 2001. p 2583-2586.

16. Zamani F, Safabakhsh R. An unsupervised GVF snake approach for white blood cell segmentation based on nucleus. In Proceeding of 8th International Conference on Signal Processing, Beijing, China; Vol 2.

17. Rezatofighi SH, Khaksari K, Soltanian-Zadeh H. Automatic recognition of five types of white blood cells in peripheral blood. Comput Med Imaging Graph 2011;35:333-343.

18. Gu G, Cui D. Polar angle detection and image combination based leukocyte segmentation for overlapping cell images. Comput Inform 2011;30:189-199.

19. Gu G, Cui D, Li X. Segmentation of overlapping leucocyte images with phase detection and spiral interpolation. Comput Method Biomed 2012;15:425-433.

20. Ko BC, 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.

21. Nasir ASA, Mashor MY, Rosline H. Unsupervised colour segmentation of white blood cell for acute leukemia images. In Proceeding of IEEE International Conference on Imaging Systems and Techniques, Penang, Malaysia; 2011. p 142-145.

22. Theera-Umpon N. White blood cell segmentation and classficiaiton in microscpic bone marrow images. Lect Notes Comput SC 2005;3614:787-796.

23. Gua N, Zeng L, Wu Q. A method based on multispectral imaging technique for white blood cell segmentation. Comput Biol Med 2006;37:70-76.

24. Yi F, Chongxun Z, Chen P, Li L. White blood cell image segmentation using online trained neural network. In Proceeding of 27th Annual EMBS International Conference, Shanghai, China; 2005. p 6476-6479.

25. Shitong W, Min W. A new detection algorithm (NDA) based fuzzy cellular neural networks for white blood cell detection. IEEE T Inf Technol B 2006;10:5-10.

26. Sabino DMU, Costa LF, Rizzatti EG, Zago MA. A texture approach to leukocyte recognition. Real-Time Imaging 2004;10:205-216.

27. Harun NH, Mashor NR, Mokhtar NR, Aimimi Salihah AN, Hassan R, Raof RAA, Osman MK. Comparison of acute leukemia image segmentation using HSI and RGB color space. In Proceeding of 10th International Conference on Information Sciences Signal Processing and their Applications, Kuala Lumpur, Malaysia; 2010. p 749-752.

28. Mohapatra S, Patra D. Automated cell nucleus segmentation and acute leukemia detection in blood microscopic images. In Proceeding of International Conference on Systems in Medicine and Biology, Kharagpur, India; 2010. p 49-54.

29. Zerfab T, Rehn T, Wittenberg T. Boundary-precise segmentation of nucleus and plasma of leukocytes. In Proceeding of SPIE Medical Imaging, San Diego, CA USA; 2008. p 1-6.

30. Eom S, Kim S, Shin V, Ahn B. Leukocyte segmentation in blood smear images using region-based active contours. Lect Notes Comput SC 2006;4179:867-876.

31. Madhloom HT, Kareem SA, Ariffin H, Zaidan AA, Alanazi HO, Zaidan BB. An automated white blood cell nucleus localization and segmentation using arithmetic and automatic threshold. J Appl Sci 2010;10:959-966.

32. Aimi-Salhah AN, Mashor MY, Harun NH, Abdullah AA, Rosline H. Improving colour image segmentation on acute myelogenous leukemia images using contrast enhancement techniques. In Proceeding of 32nd Annual EMBS International Conference, Buenos Aires, Argentina; 2010. p 246-251.

33. Yang X, Li H, Zhou X. Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy. IEEE T Circuits-I 2006;53:2405-2414.

34. Cheng J, Rajapakse JC. Segmentation of clustered nuclei with shape markers and marking function. IEEE Trans Biomed Eng 2009;56:741-748.

35. Malpica N, de Solorzano CO, Vaquero JJ, Santos A, Vallcorba I, García-Sagredo JM, del Pozo F. Applying watershed algorithms to the segmentation of clustered nuclei. Cytometry 1997;28:289-297.

36. Osowski S, Markiewicz T, Marianska B, Moszczynski L. Feature generation for the cell image recognition of myelogenous leukemia. In Proceeding of 12th European Signal Processing Conference, Vienna, Austria; 2004. p 753-756.

37. Nilsson B, Heyden A. Segmentation of complex cell clusters in microscpic images: application to bone marrow samples. Cytometry Part A 2005;66A:24-31.

38. Pan C, Fang Y, Yan X, Zheng C. Robust segmentation for low quality cell images from blood and bone marrow. Int J Control Autom 2006;4:637-644.

39. Kocalev VA, Grigoriev AY, Ahn H-S, Myshkin NK. Automatic localization and feature extraction of white blood cells. In Proceeding of SPIE Medical Imaging, San Diego, CA USA; 1995. p 177-182.

40. Otsu N. A threshold selection method from gray-level histograms. IEEE Trans Sys Man 1979;9:62-66.

41. Chan TF, Vese LA. Active contours without edges. IEEE Trans Image Process 2001; 10:266–277.

42. Tosun AB, Kandemir M, Sokmensuer C, Gunduz-Demir C. Object-oriented texture analysis for the unsupervised segmentation of biopsy images for cancer detection. Pattern Recognit 2009;42:1104–1112.

Color and Shape Based Cell Segmentation