Multi Features Based Approach for White Blood Cells Segmentation and Classification in Peripheral Blood and Bone Marrow Images

Mohammed Lamine Benomar, Amine M. Chikh, Xavier Descombes, Mourtada Benazzouz

Multi features-based approach for white blood cells segmentation and classification in peripheral blood and bone marrow images

Mohammed Lamine Benomar* and
Amine Chikh

Computer Science Department,
Genie Biomedical Laboratory,
Abou-Bekr Belkaid University,
Tlemcen, Algeria
Email: mohamedamine.benomar@univ-tlemcen.dz
Email: mea_chikh@univ-tlemcen.dz
*Corresponding author

Xavier Descombes

French Institute for Research in Computer
Science and Automation (INRIA),
Laboratoire d'Informatique,
Signaux et Systèmes de Sophia Antipolis (I3S),
Côte d'Azur University,
Nice, France
Email: Xavier.Descombes@inria.fr

Mourtada Benazzouz

Computer Science Department,
Genie Biomedical Laboratory,
Abou-Bekr Belkaid University,
Tlemcen, Algeria
Email: m_benazzouz@univ-tlemcen.dz

Abstract: In this paper, we propose a complete automated framework for white blood cells differential count in peripheral blood and bone marrow images, in order to reduce the analysis time and increase the accuracy of several blood disorders diagnosis. A new colour transformation is first proposed to highlight the white blood cells regions; then, a marker controlled watershed algorithm is used to segment the region of interest. The nucleus and cytoplasm are subsequently separated. In the identification step, a set of colour, texture and morphological features are extracted from both nucleus and cytoplasm regions. Next, the performances of a random forest classifier on a set of microscopic images are compared and evaluated. The obtained results reveal high recognition accuracies for both segmentation and classification stage.

Keywords: white blood cells; WBCs; cells segmentation; cells classification; colour transformation; texture features; morphological features, peripheral blood images; bone marrow images.

Reference to this paper should be made as follows: Benomar, M.L., Chikh, A., Descombes, X. and Benazzouz, M. (2021) 'Multi features-based approach for white blood cells segmentation and classification in peripheral blood and bone marrow images', *Int. J. Biomedical Engineering and Technology*, Vol. 35, No. 3, pp223–241.

Biographical notes: Mohammed Lamine Benomar received his Magister in Science and Technology of Information and Communication at the Tlemcen University, Algeria, in 2012. He is currently a PhD student at the Biomedical Engineering Laboratory at the Tlemcen University, Algeria. His research interests include artificial intelligence, medical image analysis and computer-aided diagnosis.

Amine Chikh is a Professor of Biomedical Engineering and the Head of CREADOM Team at the Tlemcen University. He received his PhD in Electrical Engineering from the University of Tlemcen, Algeria and the INSA of Rennes, France, in 2005. He published over 90 journal and conference papers to date and is involved in a variety of funded research projects related to biomedical engineering. His current research interests include artificial intelligence, big data classification and pattern recognition.

Xavier Descombes is the Research Director at the INRIA since 2008. He is currently the Head of a joint team between INRIA, CNRS and the University of Nice dedicated to biological imagery and a member of the scientific committee of the Labex SIGNALIFE. He is a member of the IEEE (Biomedical Image and Signal Processing) BISP committee, a member of the Strategical Committee of Optitech and an expert for the DRRT. He has been the Principal Investigator of several international projects including the ODESSA Associated Team and two ECONET projects. He is the Associated Editor of the *Digital Signal Processing*. His scientific interests include stochastic modelling (Markov random fields, marked point process), optimisation (RJMCMC, birth and death process). His current projects concern computational biology. He is a co-author of more than 100 publications in international journals or conferences.

Mourtada Benazzouz received his MCA in Computer Science at the Tlemcen University, Algeria. He received his PhD in Computer Vision from the University of Tlemcen, Algeria, in 2014. He is interested in medical image analysis and computer-aided diagnosis.

1 Introduction

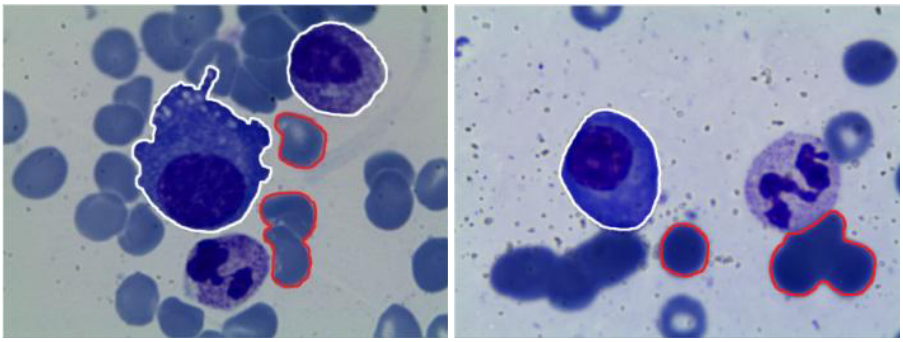
The differential count of white blood cells (WBCs) for medical diagnosis requires a careful observation in peripheral blood and bone marrow microscopic images in order to detect abnormal or suspicious cells. However, this process (screening) is time consuming even for an experiment expert.

The diagnosis relevance of several blood disorders such as leukaemia and myeloma, through the analysis of WBCs or leukocytes, depends on the correct recognition of cells. To achieve this goal, a computer analysis image system is required to automate the

process in order to help experts, to reduce the time of analysis and increase the accuracy. The main steps in such systems are segmentation and classification of WBCs. In this paper, we present a method to identify and classify a set of peripheral blood and bone marrow WBCs that includes basophil, neutrophil, eosophil, monocyte, lymphocyte and plasma cells.

In many researches, cell segmentation is the most challenging step and there are not standard techniques for each domain, but the processing must be adapted to the context. Usually, peripheral blood and bone marrow images consist of WBCs, red blood cells (RBCs), platelets and plasma. Figure 1 shows two images where the most represented cells are RBCs and WBCs. Although WBCs are easily identifiable with their colour, texture and shape characteristics from RBCs and background, the images tend to have complex contents due to the proliferation, maturation and the wide variations in cell shape, dimensions and edges. Therefore, the proposed method to identify WBCs exploits this information to segment the nucleus and cytoplasm regions that will be used to classify the cell type.

Figure 1 Example of peripheral blood and bone marrow images (see online version for colours)



Note: Some WBCs are outlined with white contour and RBCs with red.

The remaining of this paper is organised as follows. In Section 2, we present different methods to identify the WBCs. Next, in Section 3, we detail each step of the proposed method. In Section 4, we describe the implementation and we discuss results. Finally, in Section 5, we conclude the paper and present some possible future works.

2 Related works

In the literature, there exist various methods for segmenting WBCs in order to facilitate the classification in peripheral blood and bone marrow images. To this end, there are mainly two approaches. In the first approach, the WBCs nucleus are identified and then, adequate features are extracted to classify cells. Theera-Umpon and Dhompongsa (2007) propose a differential WBCs count framework in bone marrow images and show that nucleus alone can be used to classify cells, since its segmentation is much easier than the entire cell. They extracted morphological granulometries nucleus features, followed by WBCs classification using Bayes classifiers and artificial neural networks. Leukocyte cell nucleus enhancer using RGB and HSV property to segment nucleus region is

proposed in Huang and Hung (2012). In the recognition steps, they reduced 85 textural and shape features by PCA and classified cells using a genetic algorithm. Madhloom et al. (2010) work focus on five types of WBCs nucleus segmentation using a combination of contrast stretching and image arithmetic operation.

In the second approach, the idea is to segment the entire WBCs individually and then to separate nucleus from cytoplasm in the second step. Recently, Arslan et al. (2014) have implemented WBCs segmentation in peripheral blood and bone marrow images based on colour and shape transformation. They transformed the RGB image into a new intensity map based on its green and blue bands to make the pixels of WBCs more distinguishable. Then, they calculated a distance transform on a binary mask of this new intensity map. They implemented a two-stage segmentation algorithm. First, the cell regions are separated from background using Otsu threshold method. The second stage, based on active contour, is to refine the WBCs boundaries regions and remove false positives. Finally, they combined the new intensity and distance maps in a marker-controlled watershed algorithm to delineate cell boundaries. Indeed, it was found that the proposed algorithm improve the WBCs segmentation performance. Putzu and Di Ruberto (2013) use colour transformation (RGB to CMYK) as WBCs are more contrasted in the Y component, followed by a redistribution of image grey levels by contrast stretching or histogram equalisation in order to simplify the process. The segmentation is realised using the triangle threshold method and arithmetic operations to remove the background, followed by a modified watershed to separate grouped WBCs from which are extracted various features for the classification phase. A system to locate WBCs in microscopic blood smear is proposed in Prinyakupt and Pluempitiwiriyawej (2015). The concept of the segmentation is based on morphological properties of the real cells. Therefore, the nucleus and cytoplasm were segmented separately after colour transformation, thresholding and ellipse curve fitting to overcome the shapes and sizes variability of WBCs. They extracted 15 features from the segmented nucleus and cytoplasm regions to classify five types of leukocytes using linear and naïve Bayes classifiers. Chu et al. (2015) introduce a method inspired by cosegmentation to delineate the entire WBCs contour. Colour transformation and thresholding are employed to obtain a reference subimage. Then, to cosegment the reference image and the other subimages a similarity measurement is used. Madhloom et al. (2012) integrate colour features with morphological operations to localise WBCs in peripheral blood images and extract each individual cell separately in a subimage. In a continuation of their experiments, Madhloom et al. (2013) develop a computerised recognition system of normal and abnormal lymphocytes cells based on shape and texture features extraction, selection and cell classification. Rezatofighi and Soltanian-Zadeh (2011) propose a system to classify five major groups of WBCs (eosinophil, basophil, monocyte, lymphocyte and neutrophil). Nucleus and cytoplasm were segmented using Gram-Schmidt orthogonalisation and a snake algorithm after the pre-processing procedures. Three kinds of features (colour, morphological and textural) are elicited and selected. Finally, they compared the results of two classifiers (ANN and SVM). A texture approach to WBCs recognition was presented by Sabino et al. (2004). Ramoser et al. (2005) employ colour transformation and K-means clustering for WBCs segmentation. A set of colour and shape features is performed with a polynomial support vector classifier to discriminate between different cell types.

In previous work, we have identified plasma cell in bone marrow images algorithm in two phases. Firstly, nucleus extraction is performed by Otsu thresholding from green

channel, then a region growing with circularity criterion delimitates the cytoplasms. Features extraction and cells classification is presented in Benazzouz et al. (2015). Segmentation scheme using pixel classification based on the fusion of information and evidential algorithm to segment blood cell images is reported in Benazzouz et al. (2013, 2016) and Baghli et al. (2014).

The previous studies based on segmentation of nucleus regions alone are limited when considering the identification of cells types, since the cytoplasm is essential for the classification of several WBCs kinds. Moreover, the published methods show that the cells segmentation and features extraction are the most important steps. In this paper, we present the different steps of a differential WBCs counting system based on a new colour transformation, texture and shape properties leading to faster and more accurate results.

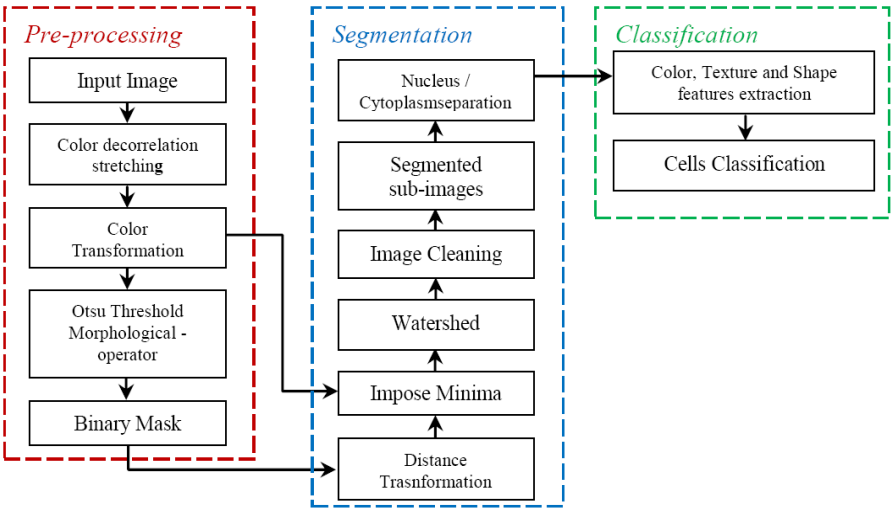
3 Proposed methods

In this paper, we propose to locate the entire WBCs in peripheral blood and bone marrow microscopic smear in three main steps: pre-processing, segmentation and classification as shown on Figure 2. The main properties used by the WBCs segmentation and classification algorithm are colour, texture and morphology. The first step reveals chromatic characteristics of the WBCs by applying decorrelation stretch to multichannel RGB image. Then, a simple colour transformation and Otsu thresholding suppresses background and most of the RBCs. In the segmentation step, two techniques have been used which are marker controlled watershed based on the colour transformation and distance maps, to separate grouped WBCs, followed by an image cleaning step to differentiate between WBCs, false positives and artefacts using shape, colour and texture features. Then the nucleus and cytoplasm separation is based on both green and a* bands of the RGB and L*a*b colour system. The result consists of a binary subimage showing the individual WBCs. Finally, WBCs are classified into categories; this phase is based on features extraction followed by a classifier.

3.1 Pre-processing

Since peripheral blood and bone marrow images captured at the microscope are all in RGB colour space (Figure 1), it becomes necessary to exploit these characteristics consistently to several works which conclude that reducing images into greyscale yields poor segmentation results (Arslan et al., 2014; Benazzouz et al., 2013; Putzu and Di Ruberto, 2013). However, the microscopic images suffer from uneven lighting and staining during acquisition process. Moreover, WBCs can present some complications due to the variations in cell contrast, texture and morphology. Therefore, a pre-processing step is necessary in order to derive a robust and consistent segmentation for a large image dataset. The pre-processing is twofold.

Figure 2 Block diagram of the proposed cells identification system (see online version for colours)



3.1.1 Colour decorrelation stretching

The nucleus regions are more contrasted than other components as shown in Figure 1. Moreover, the nucleus regions have lower value in the green channel compared with other regions and often the cytoplasm colour is indistinguishable from adjacent RBCs. Therefore, decorrelation stretching is necessary in order to enhance the colour differences in peripheral blood and bone marrow images by mapping the original colour values to a new set of colour with a wider range. In Figure 3(a), we can see that the colour points are highly correlated and concentrated. Decorrelation stretching method was introduced by Soha and Schwartz (1978) based on a principal component (PC) transformation of the acquired image (Gonzales and Wintz, 1977). This method is accomplished in four steps (Gillespie et al., 1986):

- extract the PCs of the image
- rotate and translate along the axes of the PCs
- apply contrast stretching separately.
- calculate the inverse transformation.

The results of applying decorrelation stretching on RGB image are shown on Figure 3(b). The pixels of WBCs are more distinguishable and the WBCs can be easily segmented from the images.

3.1.2 Colour transformation

After decorrelation stretching, as shown in Figure 3(b), the RBCs regions have a greater value in the green channel than nucleus and cytoplasm regions. Moreover, the red and blue bands show the WBCs regions as the brightest objects in the image (Figure 4).

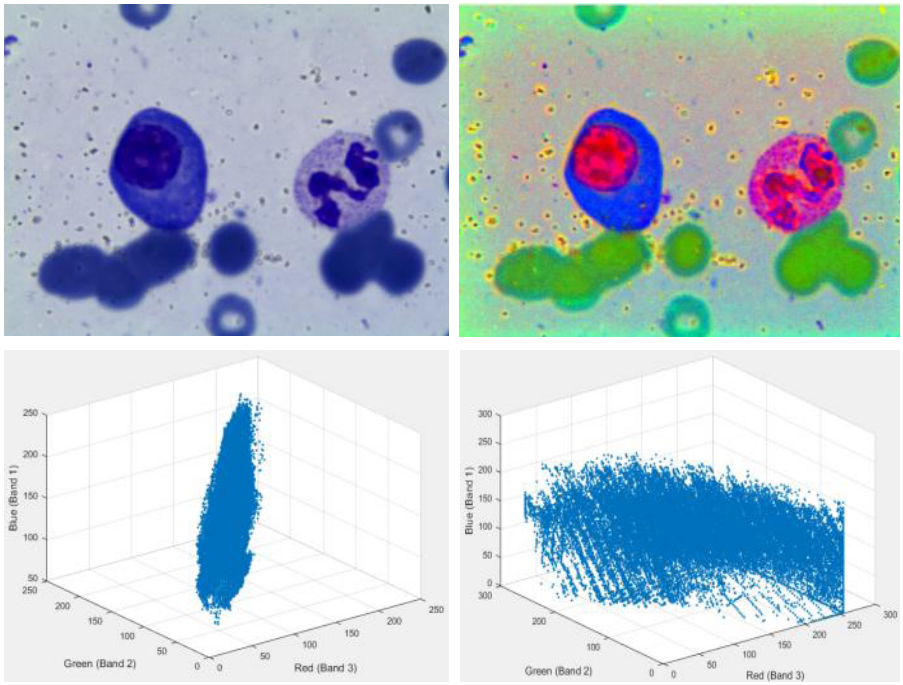
Therefore, we propose to enhance the WBC regions in the image by adding the pixel values in the red and blue bands and then subtract the green band value. Let I be the decorrelation stretch image. The I_R, I_G and I_B denote the red, green and blue bands, respectively, in RGB colour space of the latter image. The enhanced I_E can be denoted for every pixel (x, y) as:

$$I_E(x,y) = \begin{cases} T(x,y), T(x,y) > 0 \\ 0, \quad \text{otherwise} \end{cases} \tag{1}$$

where

$$T(x,y) = \left(I_R(x,y) + I_B(x,y)\right) - I_G(x,y)$$

Figure 3 Distribution of RGB pixel values (a) before and (b) after decorrelation stretch (see online version for colours)



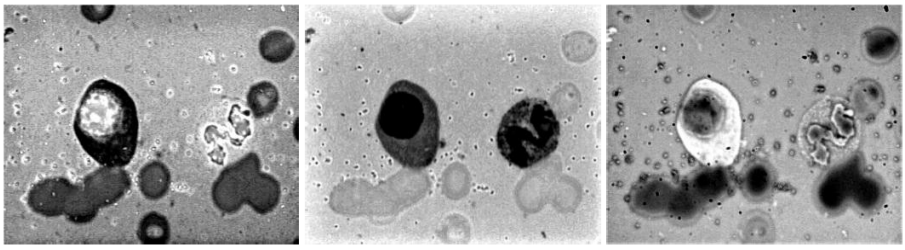
(a) (b)

Notes: a Original.
b Decorrelation stretching.

Figure 5(a) show the new intensity map via the transformation (I_E). Here, the entire WBCs are brighter than the RBCs and background, which simplifies the segmentation of WBCs from the image. We use Otsu (1979) threshold method to obtain the binary mask containing WBCs regions as shown in Figure 5(b). To refine the cell boundaries and remove the small artefacts in the background, we apply morphological operators (dilatation, erosion) as it can be seen in Figure 5(c). Nevertheless, the mask obtained may

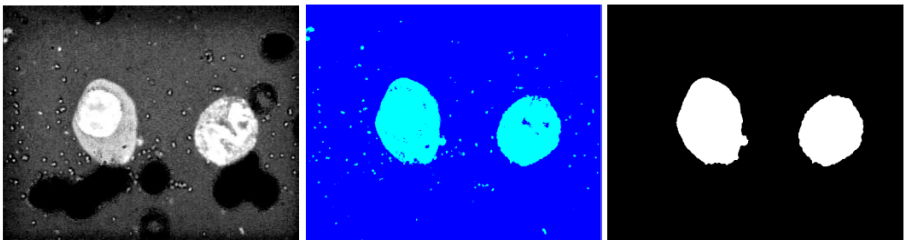
contain some false positives cells (that are not WBCs) or damaged RBCs as shown in Figure 6. These false positives are filtered using colour, shape and texture features in the next segmentation step.

Figure 4 Result of decorrelation stretching, (a) red band (b) green band (c) blue band



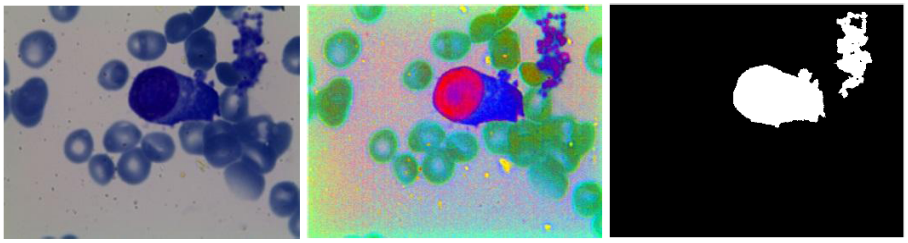
(a) (b) (c)

Figure 5 Colour transformation, (a) new intensity map (b) thresholding process (c) refined binary mask (see online version for colours)



(a) (b) (c)

Figure 6 Example of misfiled image, (a) input image (b) decorrelation stretching (c) refined binary mask (see online version for colours)



(a) (b) (c)

3.2 Segmentation

The input image in the segmentation step is the binary mask. It can contain single or connected WBCs. To separate adjacent cells the segmentation process is divided into two parts. In the first part, we consider the marker controlled watershed algorithm

(Lindblad, 2002) which uses the distance and the new intensity (I_E) maps to delineate cells boundaries (Arslan et al., 2014). Then, image cleaning is applies to remove all the false positives cells by using the colour, shape and texture features of the WBCs.

3.2.1 Marker controlled watershed

Watershed segmentation is a mathematical method based on the theory of topology for morphological segmentation (Beucher, 1982). The main drawback of this method is the over segmentation, to improve the performance of watershed segmentation, marker-controlled watershed transformations have been proposed by combining the shape and intensity maps (Arslan et al., 2014).

Firstly, we transform the binary mask into a distance map by inner distance transformation using the Euclidean metric from every region pixel to the border and then we identify the markers from which flooding starts by applying H-minima transform as shown in Figure 7(a). At this stage, applying marker-controlled watershed, provide us inaccurate separation between adjacent WBCs [Figure 7(b)]. For this reason, it is necessary to refine the contours extracted. Therefore, we define a new marking function that combines the colour and shape characteristics of WBCs (Arslan et al., 2014). Let D be the distance transform and I_E be the new intensity map. We define the marking function F for every pixel (x, y) as follows:

$$F(x,y) = D(x,y).I_E(x,y) \tag{2}$$

By exploiting this new marking function in watershed flooding process, we obtain more natural contour of WBCs, as we can see in Figure 7(c).

Figure 7 (a) Distance map, (b) watershed results with the original distance map and (c) the new marking function

(a) (b) (c)

3.2.2 Image cleaning

The extracted WBCs mask by marker-controlled watershed step contains all the WBCs and sometimes other abnormal components or RBCs that show similar colour characteristics with WBCs (see Figure 6). Therefore, image cleaning is an important stage to remove false positives cells and avoid errors in the classification process. To achieve this goal, we consider texture and shape properties of the segmented WBCs. Thus, we calculate descriptors for each connected component in the cells binary mask, which are:

- Area: the actual number of pixels in the connected region.

- Roundness:

$$Roundness = \frac{4\pi * Area}{Perimeter^2} \tag{3}$$

- Mean intensity: of the three bands (red, green, blue) in the decorrelated stretch image.

$$Mean = \frac{1}{N}\sum_{i=1}^{N} I_i \tag{4}$$

- Variance: here, we use the red and blue bands of the decorrelated stretch image to describe intensities similarity within the region. If we denote μ the mean of intensity I_i and N the pixel number of each connected component, we have:

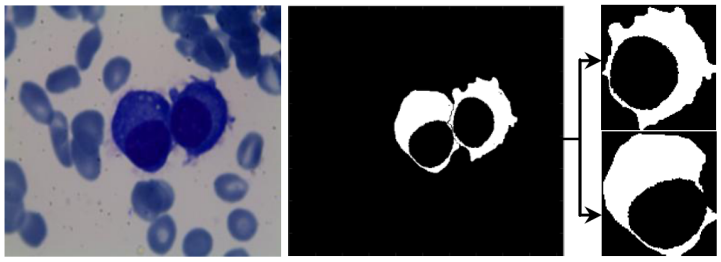
$$Variance = \frac{1}{N-1}\sum_{i=1}^{N} |I_i - \mu|^2 \tag{5}$$

The latter descriptors are computed for two objectives: as a factor to remove the noise and to take as features in the cell recognition (classification) step. To eliminate the abnormal components, we employ random forest classification algorithm (Breiman, 2001), in which several classification trees grow using a training set. Therefore, to classify a new object from an input vector, each tree gives a classification and the tree votes for that class. The forest chooses the classification having the greatest number of votes over all the trees in the forest (Berkley Statistics, https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm). Thus, the WBCs can be extracted and the noise objects are eliminated.

3.2.3 Nucleus and cytoplasm separation

The goal of this stage is to divide the WBC to its basic components which are nucleus and cytoplasm. Before that, we cut out subimages containing only single WBC from the cleaned image (Figure 8) in order to avoid problems due to signal heterogeneity among different cells. However, the WBCs can have different shape and size. Therefore, we use the bounding box size which is the smallest rectangle containing each connected component in the WBCs binary mask, the result is a binary subimage which delimitate the entire cell as shown in Figure 8.

Figure 8 Binary subimages extraction (see online version for colours)



In peripheral blood and bone marrow images, the nucleus regions are more contrasted in the green channel of the RGB colour space (Cseke, 1992; Sabino et al., 2004). However, a simple Otsu threshold in this colour band provide inaccurate nucleus regions, since there are granules in the cytoplasm region selected erroneously as part of the nucleus (Putzu and Di Ruberto, 2013). Moreover, the nucleus regions are more distinguishable in the a* channel of the L*a*b colour system. Thus, we make use of these properties by combining the binary image of both green and a* bands threshold. Combining these two colour bands, yield more accurate nucleus regions. Once the nucleus binary subimage have been created, to obtain the cytoplasm regions, we perform a subtraction between the entire cell and nucleus binary subimages.

3.3 Classification

In practice, the expert uses visual WBCs characteristics (nucleus and cytoplasm shape, texture and colour) to count the cells in the human blood or bone marrow smear and at the same time, identify the cells type. However, data extraction from WBCs can present some complications due to wide variations in cells morphologies, dimensions and boundaries. We automatically quantify these properties in order to provide a complete framework to support the medical activity, able to classify the major types of WBCs in bone marrow and peripheral blood images. These groups are basophil, neutrophil, eosinophil, monocyte, lymphocyte and plasma cell (derived from a type of lymphocyte) which include normal and dystrophic cells (LaFleur-Brooks, 2008; Sun, 2009) as shown in Table 1. To this end, morphological, colour and texture features are computed from the segmented nucleus and cytoplasm regions and used in a random forest classification (Breiman, 2001) to identify the cells types.

Table 1 WBCs types and diameters

WBC type	*Granulocytes*			*Agranulocytes*	
	Basophil	*Eosinophil*	*Neutrophil*	*Lymphocyte*	*Monocyte*
Diameter (μm)	around 10	10–12	10–12	Small lymphocyte 7–8 Large lymphocyte 12–18	12–20

3.3.1 Morphological features

To obtain a robust classification, we extract morphological features based on the biological aspects of WBC subtype. These features include nucleus and the whole cell area and perimeter, since the monocyte and plasma cell size is high compared with basophil, neutrophil and eosinophil which have intermediate size, whereas, the lymphocyte size is very low (see Table 1). We use the ratio between nucleus and cytoplasm areas to determinate the spread between the two regions, this ratio is very high for lymphocyte and allows to differentiate it from the other WBCs kinds since the nucleus occupies the major cell area. In addition, roundness [see equation (1)], solidity and extent of nucleus and cell body are given by:

$$Solidity = \frac{Area}{Convex\ hull} \tag{6}$$

$$Extent = \frac{Area}{Bonding\ box\ area} \tag{7}$$

The nucleus shape differentiates clearly the WBCs. The lymphocyte and plasma cell nucleus shape are closer to a circle (Putzu and Di Ruberto, 2013). Thus, they have higher values of nucleus roundness and extent, whereas, neutrophil have lower values. Moreover, solidity differentiates WBCs with irregular nucleus and body cell shape. Therefore, dystrophic plasma cells have lower cell solidity value than normal plasma cell. To these features are added two specific measures, the number of nucleus concavities and nucleus connected components. Hence, if there are multiple nucleus regions in the same cell the respective features are averaged. The number of concavities is found by subtracting the nucleus image from its convex hull; however, we consider only the concavities with a significant size and remove the small concavities. In lymphocyte, plasma cell and basophil the number of concavities is low when compared with monocyte and eosinophil, which have intermediate values, however, neutrophil have the highest values.

Table 2 Features extracted for cells classification

Region of interest	*Morphological features*	*Colour and texture features*
Nucleus	Area	Energy
	Nucleus to cytoplasm ratio	Contrast
	Roundness	
	Solidity	
	Extents	
	Number of connected components	
	Number of concavities	
Cytoplasm		Mean intensity in red channel
		Mean intensity in green channel
		Energy
		Homogeneity
		Entropy
		Correlation
		Contrast
Whole cell	Area	
	Perimeter	
	Roundness	
	Solidity	

3.3.2 Colour and texture features

In addition to the morphological features, we also take into account colour and texture information, since neutrophil, basophil and eosinophil contain granules, called granulocytes (see Table 1) and the other cells are smooth called agranulocytes (Putzu and Di Ruberto, 2013). Moreover, the difference in smear colour values is used to distinguish between normal and dystrophic plasma cells. In this paper, we used mean intensity in

each of the red and green band of the segmented regions. While textural features were computed from the co-occurrence matrix (energy, homogeneity, entropy, correlation and contrast) (Haralick et al., 1973). We computed the average of each features at four angles (0°, 45°, 90° and 135°) to make features rotation invariant. Table 2 illustrate the total employed features and the corresponding region of interest.

4 Results

The proposed method was tested on a set of 87 colour images containing 155 WBCs, obtained from marrow and peripheral blood smears dyed by May-Grunwald Giemsa (MGG) staining method. The images were taken on a Leica microscope with 100× magnification achromatic lens and recorded by a digital camera with a 1,024 × 768 pixels resolution. The WBCs have been classified by an expert to evaluate the segmentation and identification results, thus, each microscopic image has an associated ground truth image where nucleus, cytoplasm, background and RBCs regions are clearly separated and coloured with green, yellow, black and red respectively [see Figure 9(c)].

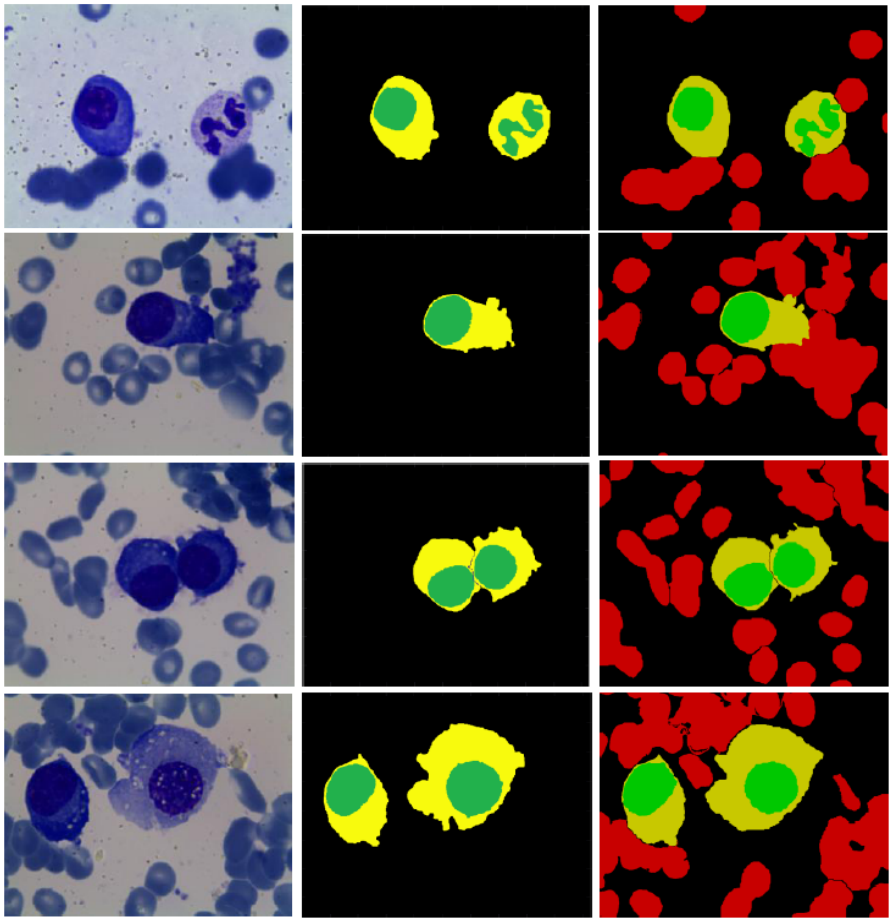
4.1 Segmentation results

To evaluate our proposed method, we use both visual and quantitative measurements. Figure 9 shows samples segmentation results. First column denotes the input images, the second are the segmented images results where nucleus and cytoplasm are coloured with green and yellow for each WBCs and the third are ground truth images. The performances of the proposed method are excellent in most cases, since the background and the RBCs surrounding the WBCs are completely removed from the image in the pre-processing stage, in fact, the adjacency between cells increases the difficulty in many previous researches. Moreover, the damaged cells and the false positive objects are cleaned from the binary mask before the identification stage. Note that errors in the cells segmentation may affect the efficiency of the classification and identification process. In the same way, the non-entire cells located on the edge of the images affect the segmentation accuracy. We eliminate ten non-entire cells, located on the image border. Table 3 shows the obtained matrix confusion. Indeed, our segmentation focuses on the WBCs which include nucleus and cytoplasm. Even the diagnostic of experts is based essentially on this type of cells.

Table 3 Segmentation confusion matrix

	Nucleus	*Cytoplasm*	*Other (red blood cells/background)*
Nucleus	0.9687	0.0313	0.0000
Cytoplasm	0.0431	0.9250	0.0319
Other (red blood cells/background)	0.0000	0.0025	0.9975

Figure 9 Segmentation samples, (a) first column demonstrate the original images (b) second are segmentation results and (c) third are ground truth images (see online version for colours)



(a) (b) (c)

Table 4 Quantitative comparison in terms of segmented regions

	Rate (%)		*Accuracy (%)*	
	Nucleus	*Cytoplasm*	*Nucleus*	*Cytoplasm*
Benazzouz et al. (2013)	94.63	90.25	95.02	84.53
Benazzouz et al. (2016)	93.53	90.04	96.42	50.77
Benazzouz et al. (2015)	98.73	94.04	96.67	91.44
Proposed method	98.81	95.24	96.87	92.50

Table 5 Confusion matrix, accuracy and overall accuracy

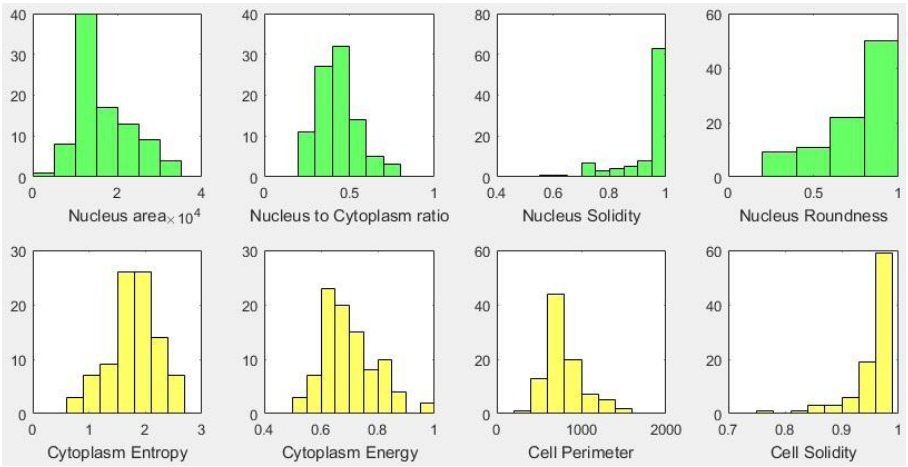
	Plasma cell		*Lymphocyte*	*Monocyte*	*Eosinophil*	*Neutrophil*	*Basophil*	*Accuracy*
	Normal	*Dystrophic*						
Plasma cell normal	46	3	0	0	0	0	0	93.87%
Plasma cell dystrophic	1	24	0	0	0	0	0	96.00%
Lymphocyte	0	1	30	0	0	0	0	96.77%
Monocyte	0	0	0	11	0	0	0	100%
Eosinophil	0	0	0	0	5	1	0	83.33%
Neutrophil	0	0	0	0	0	20	0	100%
Basophil	0	0	0	0	0	0	3	100%
Overall								95.86%

For a quantitative evaluation, we compare the performance of our experiments to the methods mentioned in the related work on the same images dataset (Benazzouz et al., 2013, 2015, 2016). As shown in Table 4, an average accuracy of 96.87% and 92.50% was obtained for nucleus and cytoplasm segmentation, respectively. Therefore, the proposed method achieved better results than Benazzouz et al. (2013, 2015, 2016) especially in cytoplasm regions. Indeed, our segmentation extracts the cytoplasm regions precisely even when the shape boundaries are irregular. We can note that the circularity criterion that prevents the deformation of the region growing in Benazzouz et al. (2015) and the misclassification between RBCs and some cytoplasm regions in Benazzouz et al. (2013, 2016) affect the segmentation accuracy.

4.2 Classification results

The proposed method classifies the WBCs into seven types: lymphocyte, monocyte, eosinophil, neutrophil, basophil, plasma cell including normal and dystrophic cells (LaFleur-Brooks, 2008; Sun, 2009) with a set of 20 extracted features to represent the WBCs (see Table 2), Figure 10 presents some features histograms that have been extracted on the WBCs subimages dataset. The classification is performed using the random forest algorithm that requires as inputs the number of trees forming the forest. After multiple tests, we opt for the typical value of 100 trees. Also, the segmentation stage extracts properly 145 subimages containing individual WBCs from the image dataset. The learning was done on 92 cells images and 53 were used for testing.

Figure 10 The histograms of eight extracted features from a set of twenty classification features (see online version for colours)



For the purpose of evaluation we calculate the class accuracy (8) and the overall accuracy (9) from the confusion matrix as:

$$AC_k = 100 \times \frac{T_{kk}}{\sum_{j=1}^{M} T_{kj}} \tag{8}$$

$$AC_{overall} = 100 \times \frac{\sum_{i=1}^{M} T_{ii}}{\sum_{i=1}^{M} \sum_{j=1}^{M} T_{ij}} \tag{9}$$

where M is the classes number, T_{ij} is the number of samples of class i that are classified as samples of class j.

As shown in Table 5, we obtain an overall accuracy of 95.86% and a robust recognition of the majority classes (lymphocyte, monocyte, neutrophil and basophil). However, we observe some misclassification of eosinophil into neutrophil since the eosinophil and neutrophil nucleus have similar shape.

A first improvement with respect to the method described earlier by Benazzouz et al. (2015), where the authors separate WBCs into two classes (normal and dystrophic plasma cells vs. other types) in order to diagnose myeloma pathology, lies in the number of WBC classes that the proposed method can distinguish. The obtained plasma cells accuracy in their work is 75.25% due to the misclassification of most dystrophic plasma cells into other cells type. The proposed classification method provides an important improvement with an accuracy of 93.87% and 96.00% for normal and dystrophic plasma cells respectively (see Table 5). The main reason is that our classification algorithm employs morphological cell features in addition to the colour and texture features, since the segmentation stage is able to find the cells boundaries precisely. Nevertheless, it should be noted that some misclassification between normal and dystrophic plasma cells can be attributed to similar cells shape when even the human expert hardly recognises the difference.

5 Conclusions

In this paper, we have proposed an automatic differential WBCs count system to assist expert in medical diagnosis. The proposed system segments the WBCs nucleus and cytoplasm and then identifies the cell types by using colour, texture and shape properties. The experiments show good results in both segmentation and classification stage, considering the cells difference and the complex scenes, with an overall accuracy of 95.86%. These results show that the WBCs identification depend on both the nucleus and cytoplasm segmentation and the choice of discriminative characteristics. This approach could be generalised to a greater number of cells types by introducing new discriminative features.

References

Arslan, S., Ozyurek, E. and Gunduz-Demir, C. (2014) 'A color and shape based algorithm for segmentation of white blood cells in peripheral blood and bone marrow images', *Journal of the International Society for Advancement of Cytometry*, Vol. 85, No. 6, pp.480–490.

Baghli, I., Nakib, A., Sellam, E., Benazzouz, M., Chikh, A. and Petit, E. (2014) 'Hybrid framework based on evidence theory for blood cell image segmentation', *Proceeding of SPIE Medical Imaging 2014 Biomedical applications in Molecular, Structural, and Functional Imaging*, Vol. 9038, DOI: 10.1117/12.2042142.

Benazzouz, M., Baghli, I. and Chikh, M.A. (2013) 'Microscopic image segmentation based on pixel classification and dimensionality reduction', *International Journal of Imaging Systems and Technology*, Vol. 23, No. 1, pp.22–28.

Benazzouz, M., Baghli, I., Benomar, A., Ammar, M., Benmouna, Y. and Chikh, M.A. (2016) 'Evidential segmentation scheme of bone marrow images', *Advanced in Image and Video Processing Journal*, Vol. 4 No. 1, p.37 [online] http://scholarpublishing.org/index.php/AIVP/article/view/1924.

Benazzouz, M., Baghli, I., Benomar, A., Benmouna, Y. and Ammar, M. (2015) *Automated Identification of Plasma Cell in Bone Marrow Images* [online] https://hal.archives-ouvertes.fr/hal-01466770 (accessed 2 January 2017).

Berkley Statistics, *Random Forests (Version 2004)* [online] https://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm (accessed 2 January 2017).

Beucher, S. (1982) 'Watersheds of functions and picture segmentation', *IEEE International Conference on Acoustics, Speech, and Signal Processing*, DOI: 10.1109/ICASSP.1982.1171424.

Breiman, L. (2001) 'Random forests', *Machine Learning*, Vol. 45, No. 1, pp.5–32.

Chu, R., Zeng, X., Han, L. and Wang, M. (2015) 'Subimage cosegmentation in a single white blood cell image', *in IEEE 2015: Computational Intelligence, Communication Systems and Networks International Conference*, DOI: 10.1109/CICSyN.2015.36.

Cseke, I. (1992) 'A fast segmentation scheme for white blood cell images', *Proceedings of the IAPR International Conference, IEEE, Pattern Recognition*, DOI: 10.1109/ICPR.1992.202041.

Gillespie, A.R., Kahle, A.B. and Walker, R.E. (1986) 'Color enhancement of highly correlated images. I. Decorrelation and HSI contrast stretches', *Remote Sensing of Environment*, Vol. 20, No. 3, pp.209–235.

Gonzales, R.C. and Wintz, P. (1977) *Digital Image Processing*, pp.103–112, 309–317, Addison-Wesley, Reading, MA.

Haralick, R.M., Shanmugam, K. and Dinstein, I. (1973) 'Textural features for image classification', *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-3, No. 6, pp.610–621.

Huang, D.C. and Hung, K.D. (2012) 'Leukocyte nucleus segmentation and recognition in color blood-smear images', *in I2MTC 2012: The Instrumentation and Measurement Technology Conference, IEEE international*.

LaFleur-Brooks, M. (2008) *Exploring Medical Language: A Student Directed Approach*, 7th ed., Mosby Elsevier, St. Louis, Missouri, USA.

Lindblad, J. (2002) *Development of Algorithms for Digital Image Cytometry, Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 789*, Uppsala University, Sweden, ISBN: 91-554-5497-6.

Madhloom, H.T., Kareem, S.A. and Ariffin, H. (2012) 'An image processing application for the localization and segmentation of lymphoblast cell using peripheral blood images', *Journal of Medical Systems*, Vol. 36, No. 4, pp.2149–2158.

Madhloom, H.T., Kareem, S.A. and Ariffin, H. (2013) 'A robust extraction and selection method for the recognition of lymphocytes versus acute lymphoblastic leukemia', *Proceeding IEEE 2013, International Conference on Advanced Computer Science Application and Technologies*, Vol. 36, No. 4, pp.2149–2158.

Madhloom, H.T., Kareem, S.A., Ariffin, H., Zaidan, A.A., Alanazi, H.O. and Zaidan, B.B. (2010) 'An automated white blood cell nucleus localization and segmentation using image arithmetic and automatic threshold', *Journal of Applied Sciences*, Vol. 10, No. 11, pp.959–966.

Otsu, N. (1979) 'A threshold selection method from gray-level histograms', *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 9, No. 1, pp.62–66.

Prinyakupt, J. and Pluempitiwiriyawej, C. (2015) 'Segmentation of white blood cells and comparison of cell morphology by linear and naïve Bayes classifiers', *BioMedical Engineering OnLine*, Vol. 14, p.63 [online] http://doi.org/10.1186/s12938-015-0037-1.

Putzu, L. and Di Ruberto, C. (2013) 'White blood cells identification and counting from microscopic blood image', *International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering*, Vol. 7, No. 1, p.363.

Ramoser, H., Laurain, V., Bischoft, H. and Ecker, R. (2005) 'Leukocyte segmentation and classification in blood-smear images', *Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Conference*.

Rezatofighi, S.H. and Soltanian-Zadeh, H. (2011) 'Automatic recognition of five types of white blood cells in peripheral blood', *Journal of Computerized Medical Imaging and Graphics*, Vol. 35, No. 4, pp.333–343.

Sabino, D.M.U., Costa, L.F., Rizzatti, E.G. and Zago, M.A. (2004) 'A texture approach to leukocyte recognition', *Real-Time Imaging Journal*, Vol. 10, No. 4, pp.205–216.

Soha, J.M. and Schwartz, A.A. (1978) 'Multispectral histogram normalization contrast enhancement', *Proceeding of the 5th Canadian Symposium on Remote Sens.*, Victoria, BC, Canada, pp.86–93.

Sun, T. (2009) *Atlas of Hematologic Neoplasms*, 1st ed., Springer, USA.

Theera-Umpon, N. and Dhompongsa, S. (2007) 'Morphological granulometric features of nucleus in automatic bone marrow white blood cell classification', *IEEE Transactions on Information Technology in Biomedicine*, Vol. 11, No. 3, pp.353–359.