White Blood Cells Identification and Counting from Microscopic Blood Image

Lorenzo Putzu, and Cecilia Di Ruberto

Abstract—The counting and analysis of blood cells allows the evaluation and diagnosis of a vast number of diseases. In particular, the analysis of white blood cells (WBCs) is a topic of great interest to hematologists. Nowadays the morphological analysis of blood cells is performed manually by skilled operators. This involves numerous drawbacks, such as slowness of the analysis and a nonstandard accuracy, dependent on the operator skills. In literature there are only few examples of automated systems in order to analyze the white blood cells, most of which only partial. This paper presents a complete and fully automatic method for white blood cells identification from microscopic images. The proposed method firstly individuates white blood cells from which, subsequently, nucleus and cytoplasm are extracted. The whole work has been developed using MATLAB environment, in particular the Image Processing Toolbox.

Keywords—Automatic detection, Biomedical image processing, Segmentation, White blood cell analysis.

I. INTRODUCTION

THE observation of blood samples by expert operators is one of the diagnostic procedures available for the recognition of different diseases. The human visual inspection is tedious, lengthy and repetitive, and suffers from the presence of a non-standard precision: that is because it depends on the operator skills. These reasons have limited its statistical reliability. On the other side, the automated analysis by computer requires only one image and not a blood sample; for this, it turns out to be less expensive but at the same time more scrupulous in providing precise standards. The main goal of this work is the analysis and processing of a microscopic image, in order to provide an automated procedure to support the medical activity. On the market there are various systems for the automatic quantification of blood cells that allow to count the number of different types of cells within the blood smear. These counters make use of techniques of flow cytometry to measure some physical characteristics and/or chemical properties of the blood cells, passing through a light detector which, through the fluorescence or electrical impedance, allows to identify the type of cell. Although the results of quantification are very precise, morphological abnormalities of the cells are not detected by the machine and, therefore, it is seen necessary a subsequent analysis of blood under the microscope. The use of image processing techniques has grown rapidly in the recent years. These techniques help to count the cells in the human blood and, at the same time, provide information on the cells morphology. Unfortunately,

L. Putzu and C. Di Ruberto are with the Department of Mathematics and Computer Science, University of Cagliari, via Ospedale 72, 09124 Cagliari, Italy, (e-mail: lorenzo.putzu@gmail.com, dirubert@unica.it).

there are not standard techniques for the analysis and processing of the images valid for each type of them, but the processing must be adapted to the context. Regarding the microscopic images, processing techniques vary depending on the type of blood cells to be analyzed. Our work focuses on the analysis of white blood cells or WBCs. A typical blood image usually shows four components: plasma, red blood cells or erythrocytes, white blood cells or leukocytes, and platelets. The most represented cells in the blood are red blood cells and white blood cells. Leukocytes are easily identifiable, as their nucleus appears darker than the background (Fig. 1 shows an example).
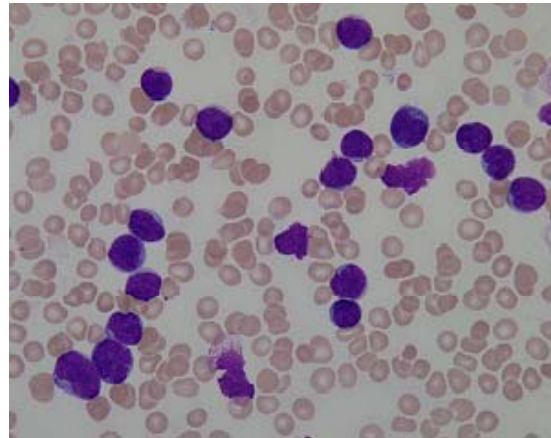

Fig. 1 Original blood sample image

However, the analysis and the processing of the data related to the white blood cells can present some complications due to wide variations in cell shape, dimensions and edges. The generic term leukocytes refers to a set of cells quite different each other as it can be seen in Fig. 2. The leukocyte cells containing granules are called granulocytes and include neutrophils, basophils and eosinophils. The cells without granules are called agranulocytes and include the lymphocytes and monocytes. Thus we can distinguish between them, not only according to the shape or size, but also thanks to the presence of granules in the cytoplasm and also by the number of lobes in the nucleus. The lobes are the most substantial part of the nucleus and are connected to each other by thin filaments.

Neutrophils are mainly present in human blood with a percentage ranging between 50 and 70%, have sizes around 10-12 microns and are distinguishable due to the number of lobes present in the nucleus, which can be up to a maximum of

5. Basophils instead represent only 0-1% of lymphocytes in human blood, have a diameter of about 10 microns and,
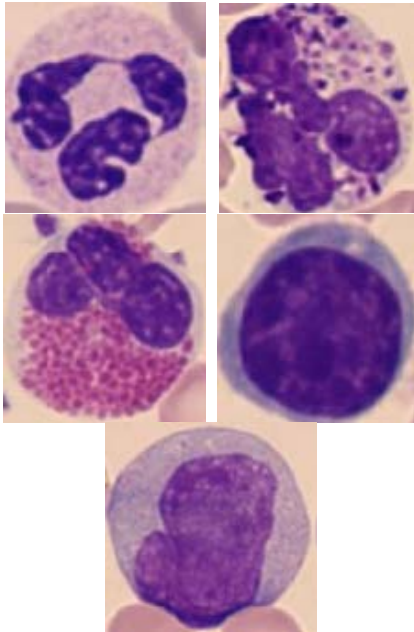

Fig. 2 Comparison between different types of white blood cells: neutrophil, basophil, eosinophil, lymphocyte, monocyte

generally, a nucleus with two lobes. Eosinophils are present for the 1-5% in human blood, have predominantly rounded shape with dimensions around 10-12 microns, and have a nucleus with more lobes, but not greater than 2. They differ from other white blood cells for the presence of granules, which include paracrystalline structures in the form of "coffee bean". In human blood is very common the presence of lymphocytes, with a percentage of 20-45% and a size of 7-15 microns, characterized by a rounded nucleus and a cytoplasm poor. Monocytes are the most voluminous white blood cells, with a diameter of 12-18 microns and representing 3-9% of circulating leukocytes. Therefore, in this paper, we present a method to identify all types of white blood cells present in the microscopic images, which need to various steps to reach the goal. The rest of this paper is organized as follows. Section II explains the proposed method: each phase of the method, applied on a sample image, is described in detail and compared with other approaches present in literature. Section III describes results and Section IV presents conclusions and some possible future works.

II. PROPOSED METHOD

The proposed method, differently from other methods in the literature, does not present separate steps of pre-processing and segmentation, but uses methods of pre-processing inserted between the various stages of the segmentation, in order to make the latter more simple and more robust. The identification of the leukocytes is carried out in the first stage,

this requires also to identify and remove the image background, in order to make the identification of leukocytes more efficient. The result consists of a binary image showing only leukocytes. The second step is the identification of
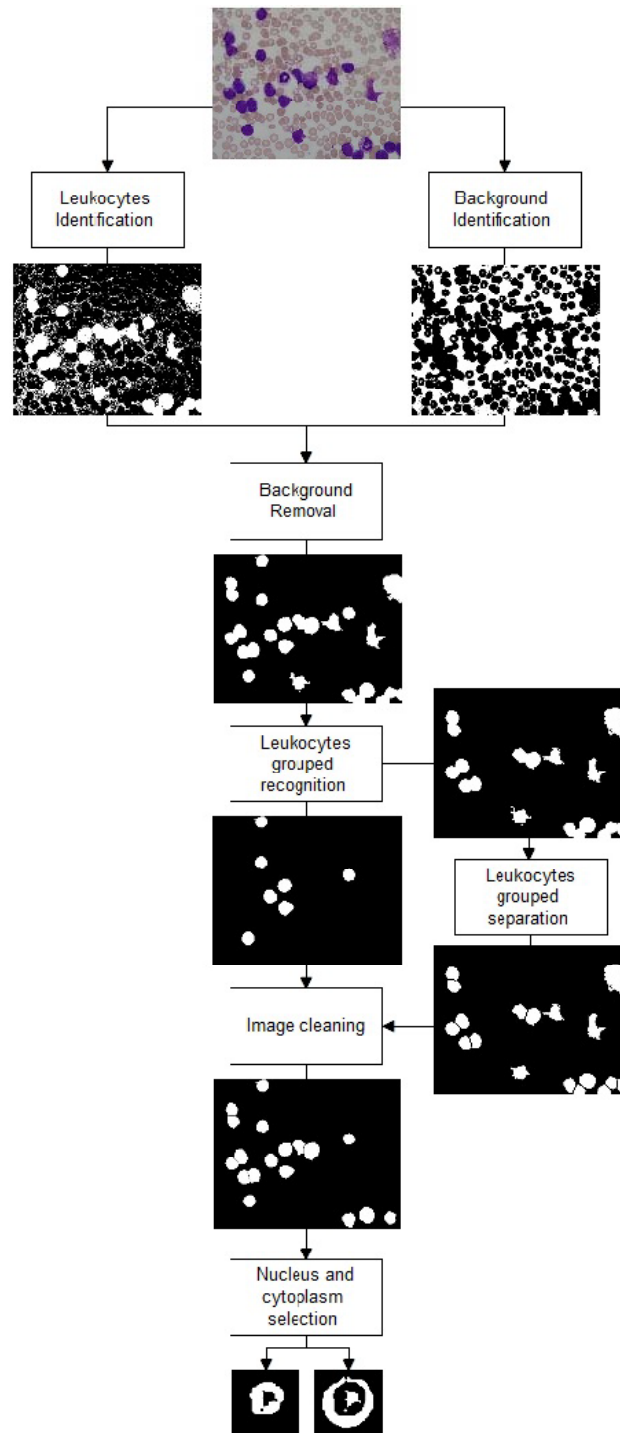

Fig. 3 Proposed method schema

leukocytes groups and produces a binary image showing the individual leukocytes and a binary image showing the adjacent leukocytes. The third phase takes care of separating the adjacent leukocytes and removing the elements localized on the border of the image. The fourth phase proceeds with a shape control, through which all the abnormal components are removed from the image. The fifth and last phase deal with the selection of the nucleus and the cytoplasm of each leukocyte. The whole process can be schematized as showed in Fig. 3.
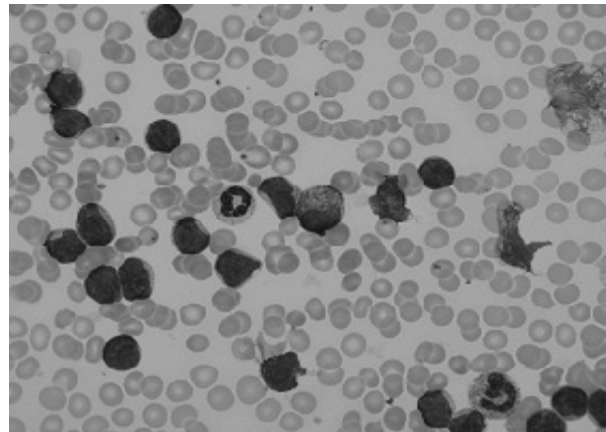
A. Background Identification

Since that the images captured at the microscope suffer from uneven lighting, it becomes necessary to remove the background because the segmentation methods based on threshold may suffer heavily for this problem. Some methods for background extraction are present in literature, but they use a collection of images captured with the same camera and the same microscope for estimation of the pixels belonging to the background [12], while others have a very high computational cost, not necessary for this application. The proposed approach involves the use of an automatic threshold to the original image in gray level (or along the green component, in this case the image chosen is the one presenting the greatest difference between minimum and maximum intensity level). There are many threshold techniques available in literature [4], [5]. Here, we use the threshold value based on triangle method or Zack algorithm [14]. The triangle method is applied to the image histogram, constructing a straight line that connects the highest histogram value $h[b_{max}]$ and the lowest histogram value $h[b_{min}]$, where b_{max} and b_{min} indicate the values of the gray levels where the histogram $h[x]$ reaches its maximum and minimum, respectively. The distance d between the marked line and the histogram values between b_{min} and b_{max} is then calculated. The intensity value, where the distance d reaches its maximum, defines the threshold value. This algorithm is particularly effective when the histograms show some clear valleys between an high peak and a weak peak, which are both present in the histograms, as it can be seen in Fig. 4(b), generated respectively from the background and from the leukocytes nucleus. Fig. 4(c) shows how the result may not be accurate in all its parts, for example, it has been detected as background even the center of the red blood cells. This does not preclude the achievement of an effective background removal, since in the later stage, during leukocytes identification also the red blood cells will be removed. Background removal can be performed later (as described in the proposed method) or directly in this phase, starting from the original image of the blood sample in the RGB color space or from the gray level image. In both cases the background removal can be performed by simple arithmetical operations.

B. Leukocytes Identification

In many methods present in the literature the idea is to identify firstly the nucleus which are more prominent than other components [10] and then the entire membrane, for example by region growing [1], [6], [7] and [8]. In the

proposed method instead, the membrane is detected firstly, in order to deal the subsequent separation of the adjacent cells more accurately. The white blood cells identification was made possible thanks to the conversion into the CMYK color model. In fact, we have observed that leukocytes are more contrasted in the Y component of CMYK color model, this is
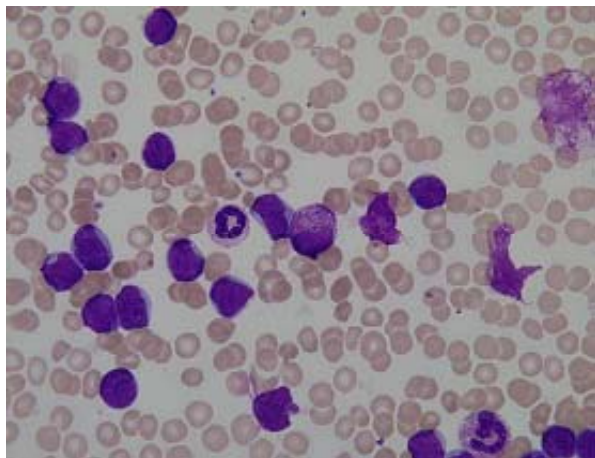

(a) Gray level image
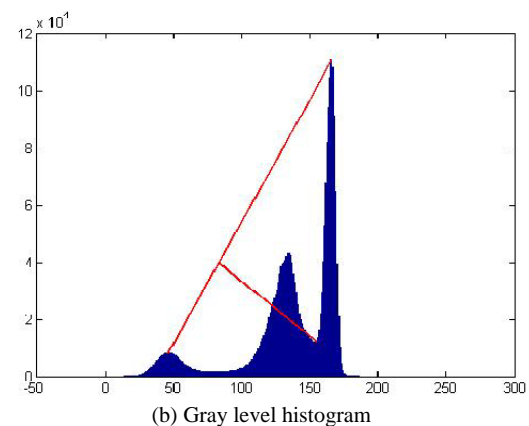
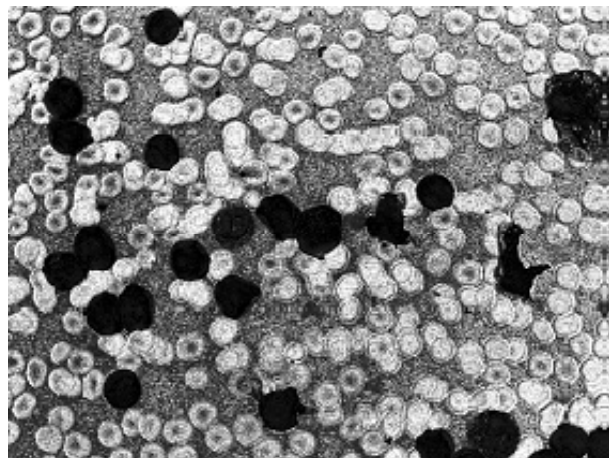

(b) Gray level histogram
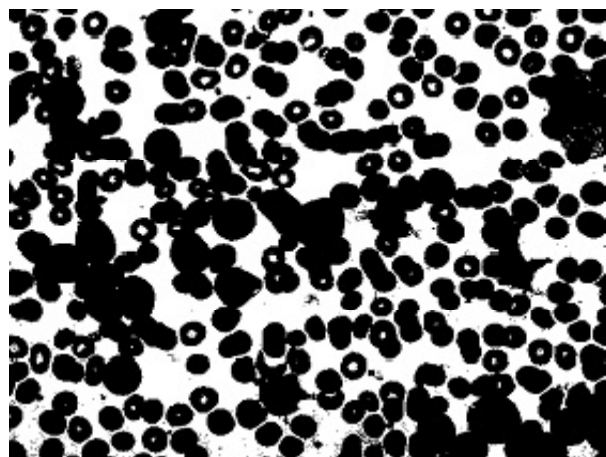
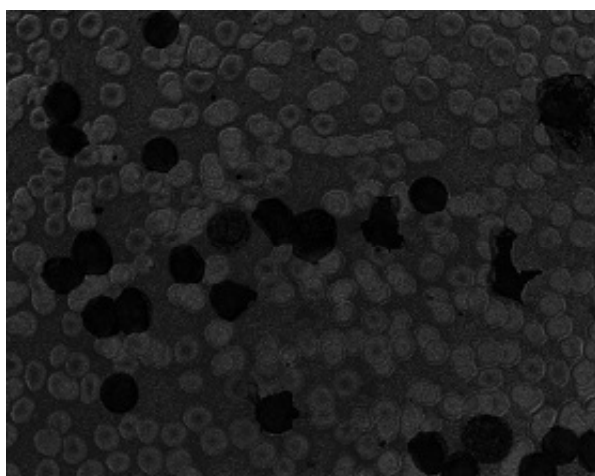(c) Threshold result

Fig. 4 The proposed approach for background identification by threshold
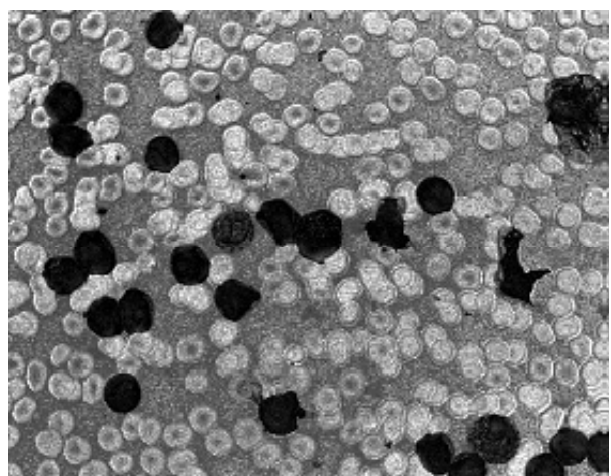
(a) RGB image

(a) Histogram equalization

(b) Y component image

(b) Contrast stretching

Fig. 5 An original blood sample image and its Y component image in the CMYK color space

Fig. 6 Histogram equalization and contrast stretching obtained from Fig. 5(a)

because the yellow color is present in all the elements of the image except in leukocytes, where it is practically absent (Fig. 5 shows an example). At this point a redistribution of image gray levels is necessary in order to make easier the subsequent segmentation process. Then, an histogram equalization or a contrast stretching can be used at this stage (see Fig. 6). The segmentation is realized again using a threshold automatically calculated by the triangle method. The weak peaks in the histograms represent again leukocytes, and the highest peak is no longer related to the background (darker in the Y component), but to the red blood cells. The complement image is then calculated in order to obtain white blood cells on a dark background (see Fig. 7).

C. Background Removal

The operation performed to identify more accurately the white blood cells is the same previously seen for the background removal (Fig. 8(a)). Obviously the background removal process does not produce a clean result in the whole

image. In order to clean up the image, the operation used, called area opening, allows to delete all the objects with a size smaller than the structuring element. The structuring element used has a circular shape and its size is calculated on the basis of the objects average size in the image (see Fig. 8(b)).

D. Identification of Grouped Leukocytes

Once obtained the image containing only the white blood cells it is possible to verify if there are adjacent cells (or agglomerates of leukocytes) and, therefore, provide for their separation. Several methods can be used to verify the presence of adjacent leukocytes [4]. In this work we have decided to use the roundness value, calculated for each connected component of the image. The connected components having a roundness value greater than a certain predetermined threshold are classified as individual leukocytes and so they go directly to the next step of the analysis process, while the connected components having a roundness value smaller than the threshold are classified as grouped leukocytes and so they
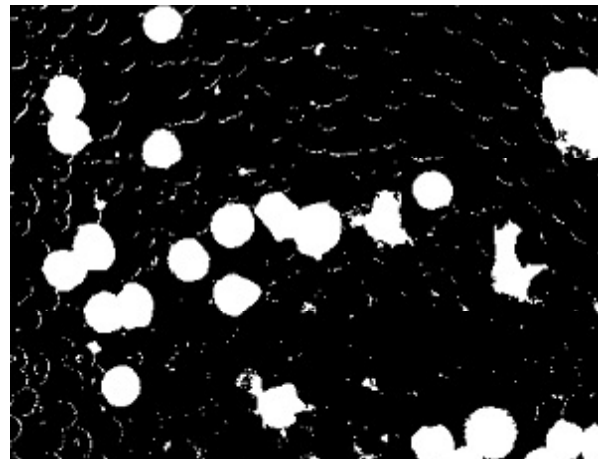
(a) Threshold result

(a) Background removal
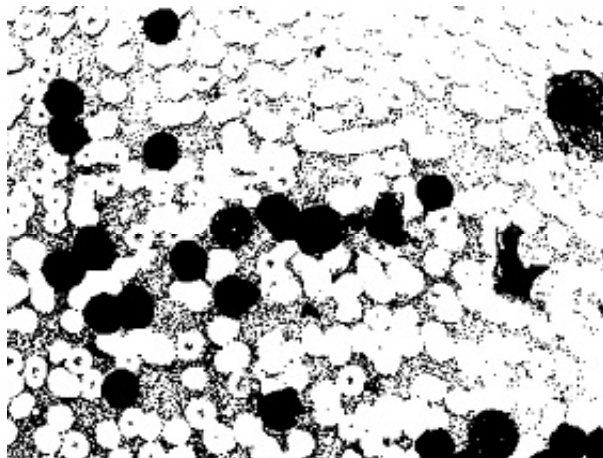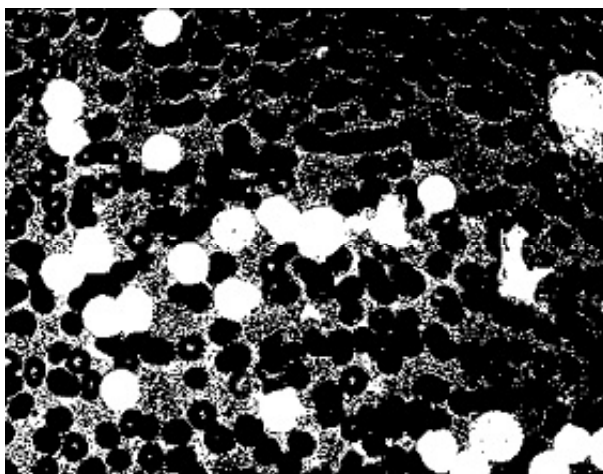
(b) Complementary image

(b) Area opening

Fig. 7 The image result of segmentation by threshold and its complementary image

Fig. 8 The image result of background removal and the enhancement by area opening

must deal with the separation process. This process creates two different images, as we can see in Fig. 9. Note that one of the two images may in some cases be empty, if it is the second, it means that the phase of separation of leukocytes will not take place.

E. Separation of Grouped Leukocytes

Many approaches have been proposed to separate the adjacent cells, some of which are included in the process of segmentation and other specifically dedicated to separate the overlapping cells. For example some approaches used by Kovalev [6], Sinha and Ramakrishnan [13], work on subimages extracted from the original image by cutting a square around the nucleus previously segmented. So, assuming that each sub-image has a single white blood cell, a clustering around the nucleus is performed, via the restrictions on the form and the color information. The proposed approach is divided into two parts. In the first part it is used the method proposed by Lindblad [9] which uses the distance transform.

The latter, applied to the binary image, associates to each pixel its distance from the border. Applying the distance transform of the watershed segmentation, it is possible to make a first roughly separation between adjacent leukocytes. The separation in this way tends to be inaccurate, as it uses the distance transform as a form delimiter, but performs well only in the presence of leukocytes adjacent with a nearly rounded shape, but it does not perform equally well in the presence of multiple complex forms, as we can see in the last image of Fig. 10. For this reason it is necessary a second step to refine the contours extracted through watershed transform. Then, all the pixels of the component under examination which are located at a distance not greater than a predetermined value from the watershed line concerned, are taken in consideration. These pixels are then used to derive the deepest concavity for which the line of exact separation will have to pass. Therefore, by exploiting the information of the points of concavity and the information related to the points of maximum image in gray tones, it is possible to obtain a cutting line that best fits the contour of the leukocytes, as we can see in Fig. 11.
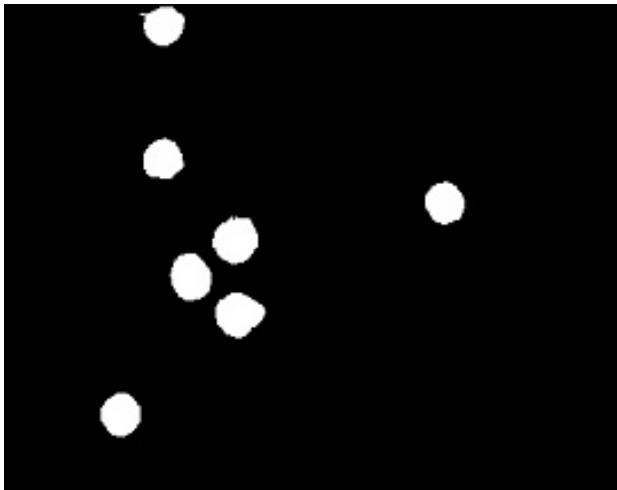
(a) Single leukocytes
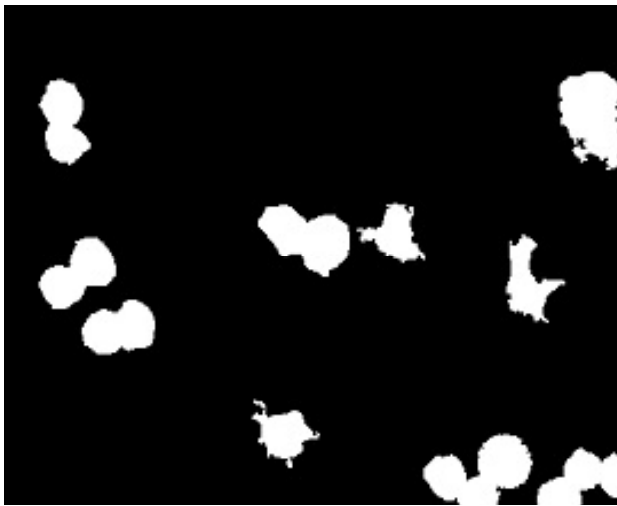
(b) Grouped leukocytes

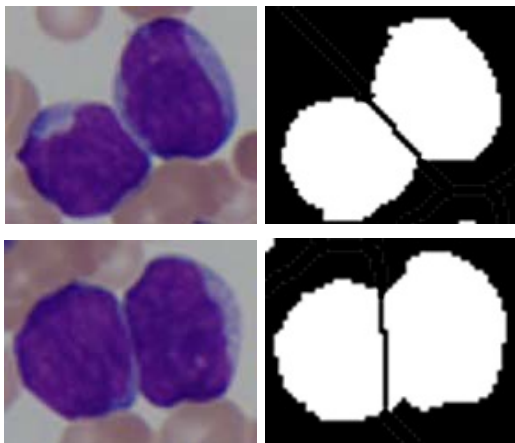Fig. 9 The images result from the identification of grouped leukocytes

Fig. 10 Two original blood sample sub-images and their respective watershed results


Fig. 11 Local maxima image and final separation results

F. Image Cleaning

The image cleaning requires the removal of all the leukocytes located along the border of the image and of all abnormal components (that are not leukocytes), in order to avoid errors in the later stages of the analysis process. The cleaning of the image border is a simple operation, while the removal of abnormal components is a more complex process. To do this it is necessary to first determine the number of leukocytes present in the image. For each of them it is then calculated the size of the area and the size of the convex area. The size of the area is used to calculate the mean area, necessary to determine and eliminate the components with irregular dimensions. For example, a very small area might indicate the presence of artifacts not removed adequately, on the other side, a very large area may indicate the presence of adjacent leukocytes not separated adequately. Area and convex area are then used in combination for the calculation of the solidity value. All objects with a solidity value less than a predetermined threshold are discarded. In fact, even in this case, a value of solidity less than the threshold value indicates the presence of artifacts not removed adequately. Fig. 12 shows the final results of the border cleaning and of the abnormal components removal.

G. Nucleus and Cytoplasm Selection

Once the leukocytes have been identified, it is possible to move to the second segmentation level that provides the selection of nucleus and cytoplasm. This step can be simplified performing an image crop using the bounding box size that is the smallest rectangle that completely contains the connected component, with the aim to have a single leukocyte for sub-image, as it is shown in Fig. 13. A border cleaning operation is again necessary in order to preserve only the white blood cell in question. Since by definition, leukocytes nucleus is internal to the membrane, it is possible to perform further simplification, through the crop of the entire portion of the image outside the leukocyte in question (see Fig. 13). This procedure allows a more robust nucleus selection, because it excludes completely artifacts from the selection. Nucleus selection approach takes advantage from Cseke's observations [2], who found that white blood cells nuclei are more in contrast in the green component of the RGB color space. Threshold operation using Otsu [11] in this color space,
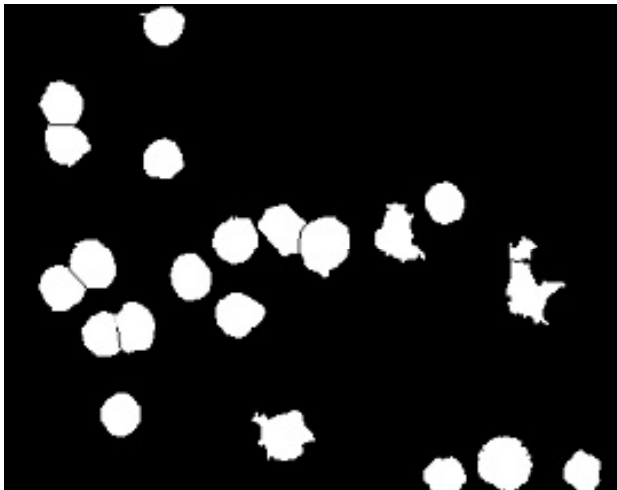
(a) Border cleaning

(b) Abnormal components removal
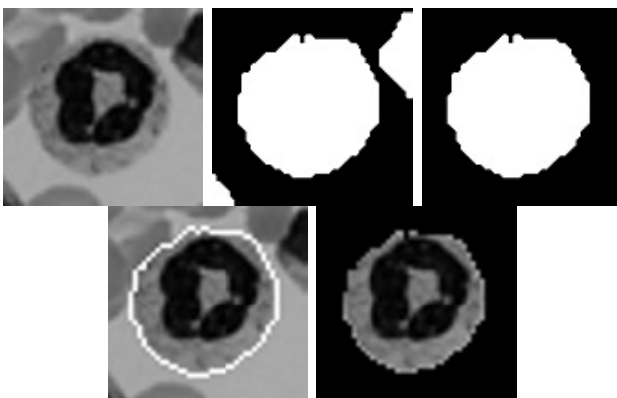
Fig. 12 The result from the image cleaning phase

Fig. 13 Gray level and binary sub-image of individual leukocytes. Border cleaning from binary image. Gray level sub-image with superimposed border. Gray level sub-image with external leukocyte image cropped

however, does not produce clean results, especially with the presence of granulocytes, whose granules are selected erroneously as part of the nucleus. For this the binary image obtained from the green component, is combined with the binary image, obtained from the a* component of the CIELab color space, again through a threshold operation. The mask obtained makes it possible to extract clearly the leukocytes nucleus. At the end, to obtain the cytoplasm you just have to perform a subtraction operation between the binary image containing the whole leukocyte and the image containing only the nucleus (see Fig. 14).
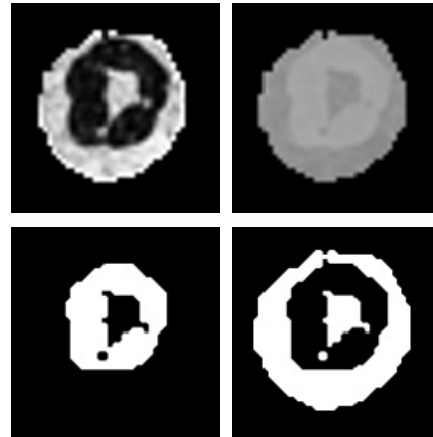

Fig. 14 Images of the G component in the RGB space and a* component in the CIELab space. Binary image after nucleus selection. Binary image after cytoplasm selection

III. Experimental Results

The proposed method was finally tested on the ALL-IDB1 database [3] which consists of 108 original blood sample images. The test was carried out with a sample of 33 images acquired from the same camera and under the same lighting conditions. These images were taken with an Olympus C2500L camera and have a resolution of 1712x1368. The proposed method has made possible to identify from these sample images 245 WBCs out of 267, with an average accuracy of 92%. The performances of the proposed method (shown in Table I) are excellent in most cases. The worst results achieved are from images that show significant overlapping between leukocytes, even difficult for human experts. The proposed method has also been tested on additional images taken with different camera and in different lighting conditions, showing strong results in all cases, as it can be seen in Fig. 15.

IV. Conclusion

In this work it has been proposed an innovative method for a completely automatic identification of leukocytes from microscopic images, in order to provide an automated procedure as support for medical activity. The results obtained show that the proposed method is able to identify in a robust way the white blood cells present in the image, separating the

TABLE I
PERFORMANCE OF THE PROPOSED METHOD

Image	Manual Count	Auto count	Accuracy
Image1	9	5	55
Image2	10	10	100
Image3	12	11	91
Image4	7	4	57
Image5	24	19	79
Image6	18	18	100
Image7	7	7	100
Image8	17	16	94
Image9	7	7	100
Image10	12	12	100
Image11	15	12	80
Image12	12	12	100
Image13	10	7	70
Image14	5	3	60
Image15	17	17	100
Image16	16	16	100
Image17	3	3	100
Image18	8	8	100
Image19	12	12	100
Image20	2	2	100
Image21	3	3	100
Image22	5	5	100
Image23	6	6	100
Image24	4	4	100
Image25	3	3	100
Image26	5	5	100
Image27	3	3	100
Image28	2	2	100
Image29	4	4	100
Image30	3	3	100
Image31	2	2	100
Image32	2	2	100
Image33	2	2	100
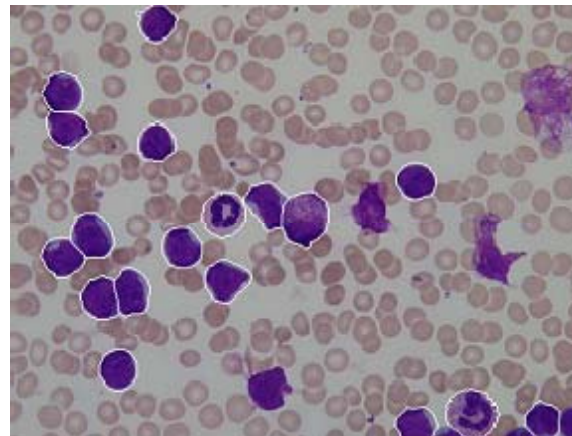
(a) Original blood image with WBCs border superimposed
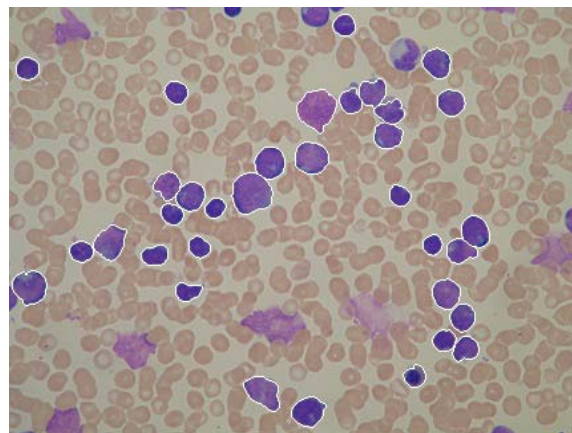
(b) Original blood image with WBCs border superimposed

Fig. 15 Leukocytes identification of two sample images

agglomerates of cells and selecting the nucleus and the cytoplasm. Further developments of the proposed method could affect the phase of separation of adjacent leukocytes, which, despite producing robust results, require a better tool than the current one, for both the identification of groups of leukocytes (even artifacts and strips of dye present in the image are identified as adjacent white blood cells), and for the separation itself, which generates incorrect results with the presence of holes to the whole leukocyte.

REFERENCES

[1] J. Cheewatanon, T. Leauhatong, S. Airpaiboon, and M. Sangwarasilp, A New White Blood Cell Segmentation Using Mean Shift Filter and Region Growing Algorithm, 2011.
[2] I. Cseke, A Fast Segmentation Scheme for White Blood Cell Images, 1992.
[3] R. Donida Labati, V. Piuri, F. Scotti, ALL-IDB: the Acute Lymphoblastic Leukemia Image DataBase for image processing, 2011.
[4] R. C. Gonzalez, R. E. Woods, Digital Image Processing, Prentice Hall Pearson Education, Inc.. New Jersey, USA, 2002.
[5] R. C. Gonzalez, R. E. Woods, S. L. Eddins, Digital Image Processing Using MATLAB, Pearson Prentice Hall Pearson Education, Inc., New Jersey, USA, 2004.
[6] V. A. Kovalev, A. Y. Grigoriev, H. Ahn, Robust Recognition of White Blood Cell Images, 1996.
[7] O. Lezoray, H. Cardot, Cooperation of Color Pixel Classification Schemes and Color Watershed: a Study for Microscopic Images, 2002.
[8] O. Lezoray, A. Elmoataz, H. Cardot, M. Revenu, Segmentation of Cytological Images Using Color and Mathematical Morphology, 1999.
[9] J. Lindblad, Development of Algorithms for Digital Image Cytometry, 2002.
[10] H. T. Madhloom, S. A. Kareem , H. Ariffin, A. A. Zaidan, H. O. Alanazi, B. B. Zaidan An Automated White Blood Cell Nucleus Localization and Segmentation using Image Arithmetic and Automated Threshold, 2010.
[11] N. Otsu, A Threshold Selection Method from Gray-Level Histograms, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 1, pp. 62-66, 1979.
[12] F. Scotti, Robust Segmentation and Measurements Techniques of White Cells in Blood Microscope Images, 2006
[13] N. Sinha, A. G. Ramakrishnan, Automation of Differential Blood Count, 2003.
[14] G. Zack, W. Rogers, S. Latt, Automatic measurement of sister chromatid exchange frequency, 1977.