Dariusz FREJLICHOWSKI

West Pomeranian University of Technology, Szczecin

Detection of Erythrocyte Cells in Microscopy Images

Abstract. *In the paper an algorithm for automatic detection of red blood cells in microscopic images is described and investigated. For this purpose digital microscopic images stained by means of the MGG (May-Grunwald-Giemsa) method are applied. This method consists in three main stages. The first one is composed of the conversion to binary image using the modified histogram thresholding. The second one comprises the localisation and extraction of each separate object. The last stage covers the selection of erythrocytes and the rejection of other cells. The method described in the paper method gave the average efficiency above 83%, which is a promising result considering the specific characteristics of the investigated data as well as the large number of cells within an image.*

Streszczenie. *W artykule przedstawiono i omówiono algorytm przeznaczony do automatycznej detekcji czerwonych krwinek na obrazach mikroskopowych. Do tego celu użyto cyfrowych obrazów mikroskopowych, barwionych metodą MGG (May-Grunwald-Giemsa). Metoda opiera się na trzech głównych etapach. Pierwszy to konwersja do obrazu binarnego przy użyciu zmodyfikowanego progowania histogramowego. Drugi to lokalizacja i ekstrakcja każdego obiektu oddzielnie. Ostatni to selekcja erytrocytów i odrzucenie pozostałych komórek krwi. Opisywana w artykule metoda dała średnią skuteczność przekraczającą 83 %, co jest obiecującym wynikiem, biorąc pod uwagę specyficzne cechy badanych danych oraz dużą liczbę komórek na obrazie. (**Detekcja komórek erytrocytów na obrazach mikroskopowych na potrzeby automatycznej lub pół-automatycznej diagnozy**)*

Keywords: Computer-Assisted Diagnosis, Detection of Blood Cells, Erythrocyte, MGG images.
Słowa kluczowe: Komputerowo wspomagana diagnoza, Detekcja komórek krwi, erytrocyt, obrazy MGG.

Introduction

The abnormalities in erythrocyte shapes can give rise to diseases, e.g. various kinds of anaemia and malaria. These deformations influence the circulation of oxygen due to the fact that it cannot be properly delivered to lungs by the affected red blood cells. Hence, the diagnosis of the above-mentioned, as well as a number of other diseases, can be based on the investigation of abnormal erythrocytes. Not only the modified, distorted shape is crucial. In some cases, the unusual number of particular types of cells is also a clue. For example, a larger amount of ovalocytes – exceeding 10 percent of all – is associated with the occurrence of a disease. This means that not only the deformation in the shape, but also the number of particular types of red blood cells has to be investigated.

In the analysis of erythrocyte shapes one can make use of digital microscopic images of human blood stained by means of the MGG (May-Grunwald-Giemsa) method. The whole process can be divided into several main stages. The first stage is the pre-processing one. It starts with the conversion to greyscale. It comes from the basic characteristic of the MGG staining method, which produces images that are mainly pink and purple. If necessary, some other initial operations on the image can also be performed in order to enhance its quality, e.g. noise reduction, histogram equalisation or high-pass filtering. Moreover, if the planned automatic or semi-automatic diagnosis will be based on a shape as a cell's feature, the thresholding has to be performed in order to obtain the binary image with white background and black blood particles.

The second stage consists of the localisation and extraction of single particular cells in the image. While this stage constitutes the focus of this paper, it will be described in greater detail in the subsequent sections. The first important problem encountered at this step is the presence of occluded shapes, which occurs very often when working with microscopic images. The second one is the selection of particular objects of interest and rejection of other ones. In the described problem it is equivalent to the detection of particular blood cells, based on their size. The leukocytes are significantly larger than RBCs, and – on the other hand – the thrombocytes are usually smaller. Both types of blood cells are outside our area of interest and therefore are rejected.

The last main stage is the recognition of an erythrocyte type. This process is performed separately for each detected and extracted particle and it is aimed at identifying the amount of each particular red blood types in an image. This stage enables to determine the possibility of a disease.

The remaining part of the paper is organised as follows. The second section describes briefly the existing approaches to automatic diagnosis based on microscopic images of human blood. The third section presents in details the proposed and applied algorithm for the detection and extraction of erythrocytes in digital MGG images. The fourth section presents the results of the experiment, and finally, the last section concludes the paper.

Brief Review of Algorithms for the Automatic Diagnosis Based on Blood Microscopy Images

The most desirable method would be to perform the automatic diagnosis based on all types of particles present in a microscopic image. More diseases could be diagnosed and hence, the system would be more advanced. However, taking into consideration the variety of cells and problems involved in the automatic analysis of digital microscopic images, this goal is very difficult to reach. Nevertheless, one can find apply several methods for the diagnosis, which are based on various kinds of particles. An example is the approach proposed in [1], where twelve categories of particles in human urine were classified. The second one is the usage of the neural network-based fuzzy classifier for the recognition of various particles in human blood ([2]).

Usually, one selected type of particle is investigated. With relation to human blood, the most popular amongst researchers is the analysis of leukocytes. In [3] the fusion of contextual information was used for this purpose. In [4] the textural information was utilised. The approach was based on the Haralick's technique. In [5] the shape features of leukocytes were used. The authors obtained an accuracy of 80 percent. In [6] the analysis was focused on one particular type of white blood cells – lymphocytes. For this purpose, the cell geometry, its area, perimeter, and several other characteristics were analysed.

The recognition of erythrocytes was also performed by several approaches. In [7] properties of a derived histogram were utilised for this purpose. In [8] the deformable templates were applied. In [9] the simple morphological operators were used. Several shape descriptors were

tested in [10] for the recognition of red blood cells types. In [11] a similar approach was investigated, but for the description of particle's shape the combination of logarithmic-polar transform and mathematical morphology was used. An unusual method for description was proposed and investigated in [12]. Detected erythrocytes were represented by means of the Polar-Fourier Greyscale Descrpitor. The method works on the extracted subpart of the image in greyscale.

The detection of particular cells in an image, preceding the above-mentioned analysis, applies various approaches, e.g. graphs ([13]) mathematical morphology ([14]), deformable templates ([15]), Voronoi diagrams ([16]), active contours ([17]), the watersheds ([18]).

The Proposed Algorithm for the Detection of Erythrocyte Cells in Microscopy Images

The method described in this paper was successfully applied at the first stages of works described in [10] and [12]. However, in the mentioned papers the emphasis was placed on the last stage, i.e. the identification of red blood types.

The described approach which is designed for the detection of erythrocytes is based on three main stages that are described in details in the following subsections.

Conversion to Binary Image

As has been already mentioned, firstly the processed MGG image is converted to greyscale. Next, its quality is enhanced through histogram equalisation. The result of this operation is important not only t because of the image quality, but it will be useful later, at the stage of cell identification.

Later, the binary image has to be obtained. For this purpose the modified histogram thresholding is applied. It starts with the derivation of the histogram. The result of this process can be represented as a function $h(l_k)$:

$$h(l_k) = \sum_{k=1}^{m} b(k, l_k), \tag{1}$$

where:

$$b(k, l_k) = \begin{cases} 1, & if\ k = l_k, \\ 0, & if\ k \neq l_k. \end{cases} \tag{2}$$

The obtained representation is smoothed using bins averaging. For this purpose the window with experimentally established size is applied. The formula for the new histogram value is:

$$c(j) = \frac{\sum_{i=j-m}^{j+m} h(i)}{2m+1}, \tag{3}$$

where:
j – number of a bin,
$c(j)$ – averaged histogram value for a bin,
$h(i)$ – histogram value before averaging,
m – number of bins taken in left and right neighbourhood of j-th bin.

The next stage is the calculation of the threshold for the derivation of the binary image. For this purpose the common property of the microscopic blood images in greyscale is utilized, i.e., the presence of two distinct peaks. The first one contains the number of pixels belonging to the cytoplasm, and the second one – background pixels. A threshold is established between them. The proper value was established experimentally and, as it turned out, this parameter differs and is correlated with the number of distinct grey-levels present in an image. If it is higher than 150, the threshold t is calculated as the number of a bin with the minimal value in the histogram and belonging to the interval:

$$t \in (c_{mac} - \left\lfloor \frac{v}{4} \right\rfloor, c_{mac}), \tag{4}$$

where:
c_{max} – number of the highest bin,
v – number of grey-level found in particular image (number of non-zero bins in histogram).

If the number of non-zero values in the histogram is lower than 150, the interval is different:

$$t \in (c_{\max} - \left\lfloor \frac{v}{4} \right\rfloor, c_{\max} - 20). \tag{5}$$

Localisation of Particular Objects

The second stage was performed on the obtained binary image. The goal was the localisation and extraction of each particular cell for further analysis. Roughly speaking, it was carried out through tracing the boundary of each separate object present in an image. However, before this operation, the objects placed on boundaries of the image were rejected, since they did not preserve the desired information about the whole shape. Hence, only the objects entirely placed in the image were later taken into account.

Next, the extraction of each separate cell was performed. Firstly, the search for a pixel not belonging to the background was conducted. This entailed scanning the image and the selection of the first white element. This pixel was the first one for the processed object (a cell). Later, the pixels belonging to the boundary were traced until the initial pixel was reached again. For the obtained object position the maximal co-ordinates in each direction were stored.

Finally, the cell was extracted from the binary image, stored, and the pixels belonging to it were marked as the background to prevent them from further processing. The described method for localisation was repeated for each object and was ended if the scanning for the white pixel did not give the result – the down-right corner of the image was reached, what indicated the lack of other objects and ended this stage.

Selection of the detected objects for further work – rejection of thrombocytes, leukocytes and occluded shapes

The microscopic image of the human blood contains various particles. Since erythrocytes are the only objects of our interest, other cells have to be removed. The first property that can be applied for this purpose is the size of the extracted shapes. Normally, leukocytes are significantly larger than red blood cells. On the other hand, thrombocytes are smaller. It means that we can simply calculate the set of appropriate area values in order to perform further analysis with erythrocytes. That process could provide us with an additional benefit – the rejection of some occluded shapes, which are impossible to be recognised. Unfortunately, the discussed assumption is insufficient, since some leukocytes and thrombocytes can have a similar area to erythrocytes. That is why another property of those three particles has to be used. The rejection can be based on a histogram, because thrombocytes and leukocytes have some parts that are very

dark, almost black. In result, the derived histogram for erythrocytes varies from the one obtained for the other two blood particles.

For the calculation of the histogram we have to return for a moment to the greyscale representation. That is why the positions of pixels for the processed detected object are used for the extraction of its version in greyscale. Then the histogram is calculated and the maximal value in it is selected – I_{max}. If it lies in the left part of the histogram the object is rejected. If this peak is situated in the right part, one can conclude that we are dealing with the red blood cell. The decision about the appropriate part of the histogram is made basing on the maximal non-zero value in the histogram. The borderline lies in the middle – the half of this value.

A very common problem is the occlusion of cells. Usually it results from the projection of three-dimensional objects into a planar image, which is a microscopic image. This problem can also occur if the preparation process is performed in a fully correct way. Moreover, the tendency of overlapping of erythrocytes in a form similar to the 'roll' is a common property of this type of blood cells. Despite of these reasons, the influence of the discussed problem on the shape object is obvious. It clearly differs from the expected templates and therefore the occluded objects have to be rejected before the later stages of the approach for automatic (or semi-automatic) diagnosis.

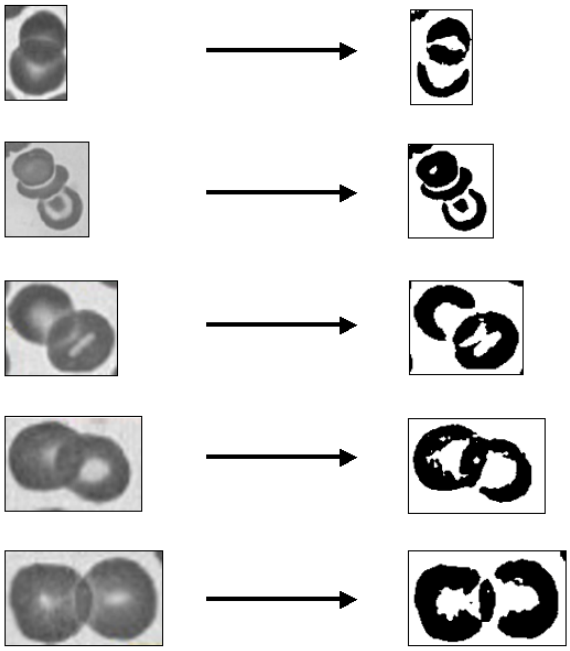

Fig. 1. The result of the processing of the occluded objects.

The decision about the additional work for particular, probably resulted from the influence of occlusion, object can be made basing on two clues. If it is larger than expected and the dissimilarity measure for it calculated for all template erythrocytes is high then we can assume that the processed object is composed of two or more occluded particles. In that case an additional step is performed. For the pixels belonging to the object (in greyscale) the histogram is built and later equalised and thresholded using the same method as described in the first subsection. Later, again the resultant objects are counted. If the obtained number is higher than one, all of them are rejected, due to the fact that the occlusion makes the proper identification impossible. The new objects, obtained after the described separation of occluded cells, should not be taken into consideration, since their shape is completely unpredictable and unusual. The result of the discussed method is provided in figure 1.

Experimental Results

The efficiency of the above-described approach for the automatic localisation and extraction of erythrocytes from digital microscopic images was experimentally investigated using ten images of human blood stained by means of the May-Grunwald-Giemsa method.

The evaluation of the results was not obvious. It could not be performed automatically. The only possible solution was counting the properly localised and extracted shapes as opposed to those that have been wrongly processed. The precise results are provided in table 1. Some exemplary images with obtained locations of cells are presented in figure 2.

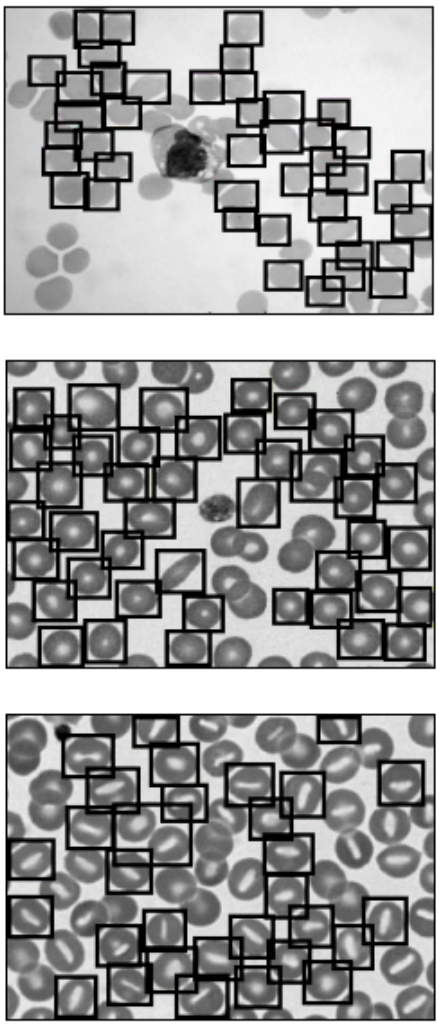

Fig. 2. The results of the approach described in the paper.

As can be seen in table 1 and figure 2, the obtained results, especially when taking into account the difficulties connected with processed microscopic images, are not ideal, however they are very promising. The above-mentioned influence of the image quality on the results was dominant, since it distorted the outline shape of the cells, and hence gave in some cases the improper identification. The second problem was the similarity between some of the classes. For example, the acanthocytes were could be mistakenly identified as echinocytes. The difference between ovalocytes and normocytes was in some cases to

small, if the first ones were visible from an unusual angle. Finally, the only difference between normocytes, stomatocytes or spherocytes is the darker area inside the cell. Therefore, if the image quality did not allowed for proper pre-processing, sometimes the identification was wrong. Nevertheless, the average efficiency of the proposed approach was above 83%. While at first sight this may seem not enough, this result suggests that the method is rigorous and rejects objects that in further stages could be wrongly recognised. Moreover, as it was stated in [12], owing to the large number of objects in the microscopic image of human blood, the automatic diagnosis will be appropriate even if some of the objects will not be extracted for further recognition.

Table 1. The results of the experiment on localisation and detection of erythrocyte shapes.

Test image	Rejected cells placed on image boundary or occluded	Properly located cells	Cells not or wrongly located	Efficiency
1	16	20	2	90,91 %
2	14	17	14	54,84 %
3	44	30	8	78,95 %
4	16	43	10	81,13 %
5	35	20	4	83,33 %
6	30	42	4	91,30 %
7	70	55	7	88,71 %
8	24	32	11	74,42 %
9	19	28	4	87,50 %
10	45	52	5	91,23 %
SUM/ AVG	**313**	**339**	**69**	**83,09 %**

Conclusions

In the paper an approach for the localisation and extraction of the erythrocyte shapes from digital microscopic images of human blood was described and experimentally investigated. This method constitutes one of the stages in the developed approach for automatic (or semi-automatic) diagnosis of some selected diseases (e.g. anaemia or malaria) based on the deformations in the shape of red blood cells. It consists of three steps. Firstly the input greyscale image is converted into the binary one using modified histogram thresholding. Then each particular cell is separately localised and extracted. Finally, some of them are rejected from further work. Since the whole approach is assumed to make use of solely red blood cells, the thrombocytes, leukocytes and occluded shapes are rejected based on the knowledge about their important properties, e.g. size and lightness level.

In the paper the experimental results performed using several digital microscopic images, stained by means of the May-Grunwald-Giemsa method, were described. In total, more than seven hundred erythrocytes were present in test images. Amongst them 313 were properly rejected, as they were placed on the image boundary and hence they could not be used in further recognition and diagnosis. Only 69 objects were not located or located in a wrong way (e.g. they were occluded and hence they should be rejected). Finally, what is most important, 339 red blood cells were properly localised and extracted. That gave the average efficiency above 83 percent.

Future analysis will be mainly concentrated on the last step of the developed approach – the automatic diagnosis based on a classification using shape description algorithms. For this purpose it is important to formulate decision rules for the diagnosis.

Wydanie publikacji zrealizowano przy udziale środków finansowych otrzymanych z budżetu Województwa Zachodniopomorskiego.

REFERENCES

[1] Ranzato M., Taylor P.E., House J.M., Flagan R.C., LeCun Y., Perona P., Automatic recognition of biological particles in microscopic images, *Pattern Recognition Letters* 28 (2007), iss.1, 31–39

[2] Uebele V., Abe S., Lan M.S., A neural-network-based fuzzy classifier, *IEEE Transactions on System, Man and Cybernetics*, 25 (1995), no. 2, 353-361

[3] Song X.B., Abu-Mostafa Y., Sill J., Kasdan H., Pavel M., Robust image recognition by fusion of contextual information, *Information Fusion*, 3 (2002), iss. 4, 277–287

[4] Sabino D.M.U., da Fontoura Costa L., Rizzatti E.G., Zago A.M., A texture approach to leukocyte recognition, *Real-Time Imaging*, 10 (2004), iss. 4, 205–216

[5] Yang X., Luo L., Study of Leukocyte Features in Peripheral Blood for Automatic Classification, *Journal Of Applied Sciences*, 12 (1994), iss. 2, 120–126

[6] Nin X., Tang J., Cell morphology Extraction and Identification Technology, *Laser Journal*, 26 (2005), iss. 4, 95–97

[7] Dıaz G., Gonzalez F.A., Romero E., A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic imageses, *Journal of Biomedical Informatics*, 42 (2009), iss. 2, 296–307

[8] Bronkorsta P.J.H., Reinders M.J.T., Hendriks E.A., Grimbergen J., Heethaar R.M., Brakenhoff G.J., On-line detection of red blood cell shape using deformable templates, *Pattern Recognition Letters*, 21 (2000), iss. 5, 413–424

[9] Di Ruberto C., Dempster A., Khan S., Jarra B., Analysis of infected blood cell images using morphological operators, *Image and Vision Computing*, 20 (2002), iss. 2, 133–146

[10] Frejlichowski D., Pre-processing, Extraction and Recognition of Binary Erythrocyte Shapes for Computer-Assisted Diagnosis Based on MGG Images, In: L.Bolc et al. (Eds.): ICCVG 2010, Part I, *Lecture Notes in Computer Science*, 6374 (2010), 368-375

[11] Luengo-Oroz M.A., Angulo J., Flandrin G., Klossa J., Mathematical Morphology in Polar-Logarithmic Coordinates. In: Marques, J.S., Perez de la Blanca, N., Pina, P. (Eds.) IbPRIA 2005, *Lecture Notes in Computer Science*, 3523 (2005), 199–205

[12] Frejlichowski D., Identification of Erythrocyte Types in Greyscale MGG Images for Computer-Assisted Diagnosis, In: J. Vitrıá, J.M. Sanches, and M. Hernández (Eds.): IbPRIA 2011, *Lecture Notes in Computer Science*, 6669 (2011), 636-643

[13] Ta Vinh-Thong, Lézoray O., Elmoataz A., Schüpp S., Graph-based tools for microscopic cellular image segmentation, *Pattern Recognition*, 42 (2009), iss. 6, 2009

[14] Angulo J., Flandrin G., Automated detection of working area of peripheral blood smears using mathematical morphology, *Anal. Cell. Pathol.*, 25 (2003), iss. 1, 37–49

[15] Garrido A., de la Blanca N.P., Applying deformable templates for cell image segmentation, *Pattern Recognition*, 33 (2000), iss. 5, 821–832

[16] Jones T., Carpenter A., Golland P., Voronoi-based segmentation of cells on image manifolds, In: ICPR (2002), vol. 2, 286–289

[17] Zimmer C., Labruyere E., Meas-Yedid V., Guillen N., Olivo-Marin J.-C., Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours: a tool for cellbased drug testing, *IEEE Trans. Med. Imaging,* 21 (2002), no. 10, 1212–1221

[18] Malpica N., de Solórzano C.O., Vaquero J.J., Santos A., Vallcorba I., Garcıa-Sagredo J.M., del Pozo, F., Applying watershed algorithms to the segmentation of clustered nuclei, *Cytometry*, 28 (1997), iss. 4, 289–297

Author:
Dariusz Frejlichowski, Ph.D. Eng., West Pomeranian University of Technology, Szczecin, Faculty of Computer Science nad Information Technology, Zolnierska 52, 71-210, Szczecin, Poland, E-mail: dfrejlichowski@wi.zut.edu.pl.