Robert BURDUK[1], Bartosz KRAWCZYK[1]

AUTOMATIC DETECTION AND COUNTING OF PLATELETS IN MICROSCOPIC IMAGE

In this paper we present a machine learning-based approach for detecting platelet cells in microscopic smear images. Counting how many platelets appeared in each smear image is one of the basic tasks done in many laboratories. In many cases this is still done by a human – laboratory technician. Due to very small size and often great quantity of those cells, precise estimating of the number of platelets is not a trivial task. As in all man-dependent problems the whole process is very sensitive to errors, time-consuming and its accuracy is limited by human perception. We propose alternative, fully automatic solution that is free of those drawbacks. Our idea is based on the combination of techniques driven from two fields of modern computer science: the image analysis and pattern recognition / machine learning. It not only reduces the error rate, but, what is more important, also decreases the time needed for each smear image analysis. The obtained results are very satisfying and our solution is more precise than estimation based on human perception. This will improve the quality of laboratory work and allow to save time that can be spent on other important tasks.

1. INTRODUCTION

In modern days health and bioinformatics are one of the most vital fields for introducing new ideas and technologies. We also observe rapidly growing interest in using computational techniques for improving the quality of laboratory analysis and diagnosis. Modern advantages of computer science allow us to simplify the work in these regions by automating many procedures, simplifying experiments and supporting decisions. The main goals of implementing these techniques are reducing the amount of mistakes made by laboratory technicians due to routine or limitation of human perception and decreasing the time spent on each task. One of most commonly done analysis is the examination of microscopic smear images. Observing and counting leukocytes, erythrocytes or platelets gives researchers and laboratory technicians various valuable information. In our paper we deal with the problem of detecting and precisely counting the platelets. In many laboratories this task is still done manually. Normally this procedure is done as follows: each digital image is loaded into computer, a technician locates platelets and marks them, e.g. by clicking the mouse on it. Therefore, due to a small size and often numerous appearances of those cells this solution is not very efficient. Examining every microscopic smear image takes from a few to sometimes several or so minutes. Each day in laboratories there are analysed lots of these images. This job is very repetitive and time consuming for the technician. With routine comes higher probability of overlooking some platelets. The typical error rate oscillates about 10% of all thrombocytes. We propose an automatic approach to solving this problem, based on techniques driven from image analysis [1,8] and machine learning / pattern recognition [2,9], that reduces error rate and allows to save precious time.

---
[1] Department of Systems and Computer Networks, Wroclaw University of Technology, Poland.
 e-mail : robert.burduk@pwr.wroc.pl

2. BACKGROUND

2.1. BLOOD SMEARS

Blood smears, the main tool for blood analysis, are thin films of peripheral blood or bone marrow which are examined under a microscope to perceive the morphological features of the cells. A blood smear is made by placing a drop of blood or bone marrow sample on a very pure glass slide, which is corrosion-resistant, and then it is dispersed using a spreader slide. It is then fixed and stained to enhance the morphological cell characteristics. In a blood smear we can also spot other things such as parasites. Peripheral blood typically contains only mature leukocytes, erythrocytes and platelets. Platelets (thrombocytes) are the main object of interest in this paper, so they will be described more precisely.

2.2. PLATELETS

Platelets (or thrombocytes) are very sparse, small, disk shaped and non-nucleated corpuscles, whose size is between 1-15 microns. They come from fragmentation of precursor megakaryocytes. The average lifespan of a platelet is just 5 to 9 days. Thrombocytes have a fundamental role in hemostasis, leading to the formation of blood clots. Low numbers of those cells lead to excessive bleeding, high numbers lead to forming thrombosis, obstructing blood vessels which may result in a stroke or myocardial infarction.

3. SUPPORT VECTOR MACHINE (SVM)

Machine learning is one of the most vital sub-fields of AI research. Nowadays it is widely used for analyzing medical data [7] and it had proven itself as most useful many times, from typical decision support, to specialized data analysis and diagnosis. In our approach to detecting platelets we decided to use the Support Vector Machine technique.

The basic idea of the Support Vector Machine (SVM) is to construct a hyper-plane as the decision surface in such a way that the margin of separation between positive and negative examples is maximized. The idea of SVM is best explained in [2] and [9].

3.1. GENERAL IDEA

In order to properly classify and count the platelets that appear in analysed smear image we combine the methods from two fields of computer science. We need image analysis techniques to enhance the image, conduct the object segmentation and measure cells parameters. Machine learning, based on extracted information, gives us a powerful tool to quickly distinguish between objects of interest and unnecessary data. The proposed process is shown in Figure 1.
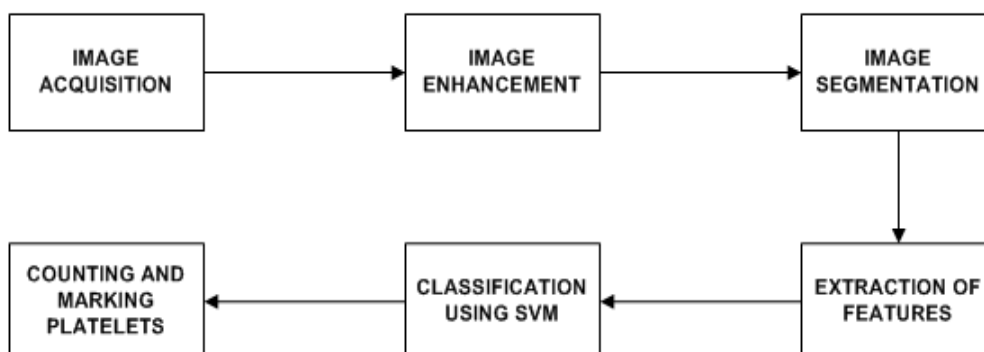

Fig. 1. Schema of proposed method.

3.2. IMAGE ENHANCEMENT

After the process of acquisition, the obtained image must be prepared for further segmentation. This is achieved by applying the image enhancement techniques, which makes the regions of interest more visible and easier to detect. Also the possible appearing noise must be reduced. To minimize the noise we use a traditional approach – the median filtering [1]. After that, to make platelets more visible, we consider only the value of green component of the analysed RGB image. Many authors had proven that the most efficient way to distinguish cells is operating on certain component of RGB : [3] shown that nuclei of white cells are most visible on the green component, and the difference between cytoplasm and erythrocytes is best seen at the blue component. [4] and [5] used this for the segmentation of leukocytes and hemoparasites. In our case platelet cells are best visible on the green component. After this enhancement the centre of platelet is clearly white, the edges are clearly black and the background is in many levels of grey. This simplifies further segmentation, making objects easy to detect. Smear images before and after enhancement are shown in Figure 2.
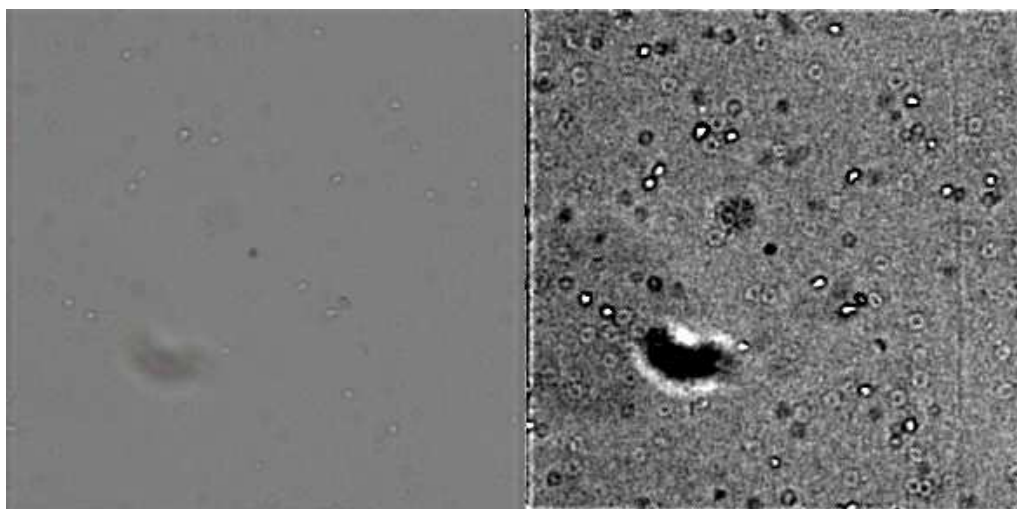

Fig. 2. Images: a) raw from the microscope b) after enhancement.

There is a possibility that objects other than platelets also may appear in each image and also will be enhanced, but by using the SVM we will discard them during the classification process.

3.3. IMAGE SEGMENTATION

Cell images are commonly segmented using algorithms such as watershed segmentation [6], mean shift or typical thresholding [1]. In our proposal we use heuristic method for finding regions of interest in each image, based on colour values of pixels. After enhancement of the analysed image we can assume that white regions represent cells, black region around them represent their edges and grey areas are cytoplasm or other artefacts. Because only pixels that may represent cells or edges carry useful information for classification process, we must ensure that background can be easily discarded. Therefore we need to receive our image segmented into three types of regions. Pixels with values 0 or 255 are left untouched and are considered as two types of areas. Pixels differing about 3% from 0 or 255 are replaced with extreme values. The background is segmented by taking all other pixels that are neither white or black and replacing their values with one level of grey (127). In Table 1 we show example pixel values before and after segmentation.

Table 1. Pixel values before and after segmentation.

Before segmentation	After segmentation
0	0
3	0
54	127
198	127
248	255
253	255
255	255

By this we ensure that extracting interesting features from objects will be easy. Results of segmentation are shown in Figure 3.
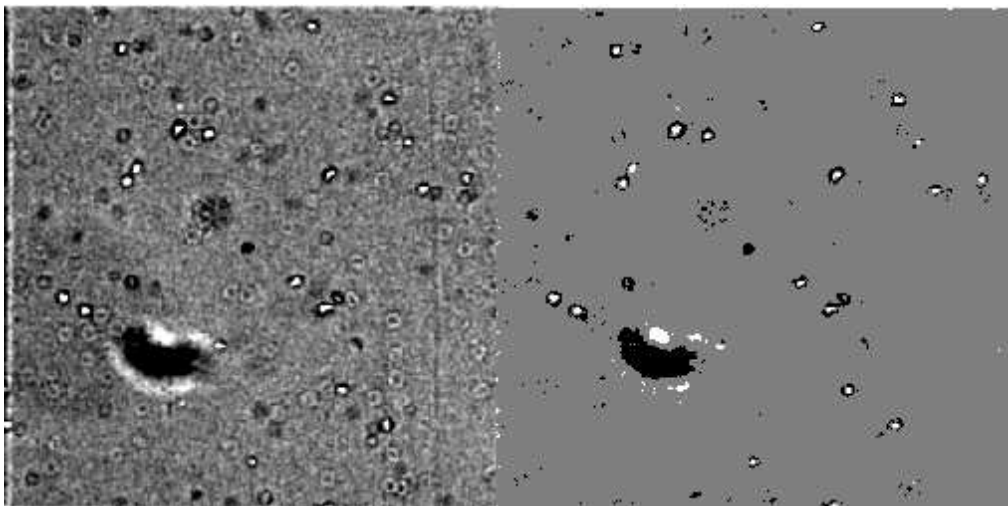

Fig. 3. Images: a) before segmentation b) after segmentation.

This heuristic approach is very fast, efficient and allows further proper classification. As said before any objects that will be segmented and are not platelets, will be discarded by using SVM.

3.4. FEATURE EXTRACTION

To clearly distinguish platelets from other objects that may appear in smear image, we need four features: the size of a cell (white region), the size of edges (black region surrounding the cell), does the object have anything inside e.g. nucleus (in practice – is there any other region inside the white one) and is the shape of the white region close to regular (circular/square). Objects that consisted of only white or black areas (e.g. no white region inside) were automatically discarded. This was achieved by examining the neighbourhood of each object of interest. Thanks to those four features we can easily discard any objects that are not platelets. Typically appearing patterns are shown in Figure 4, with additional explanation why those four features are sufficient for the classification. The location of each white region on the image is also noted, but not used in the classification.
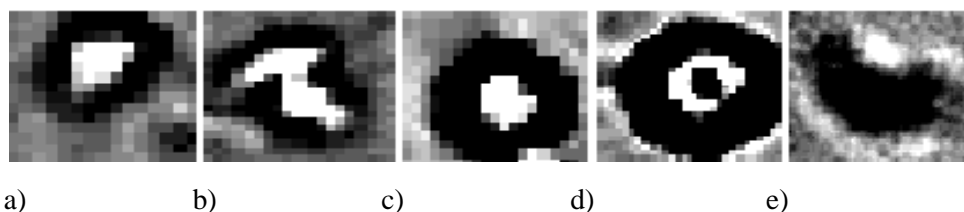
Fig. 4. Objects acquired after the segmentation. Object: a) is a platelet and others are not, b) – not a regular size, c) – too thick edge, d) – contains something inside the cell, e) – does not contain white region inside.

3.5. CLASSIFICATION USING SUPPORT VECTOR MACHINE

To train SVM we used data collected from forty smear images with platelets marked by an expert from Institute of Bioengineering from the Wroclaw University of Technology. These images were taken in normal laboratory conditions, by the same equipment that is used for everyday analysis. Therefore we are certain that only correctly classified patterns take part in the supervised learning process. Their features were extracted and turned into the training set. An expert marked both platelets (first category) and other objects (second category). Thrombocytes were put into class with label $y = 1$, and other objects, shown in previous point, were classified with label $y = -1$. In case of object sets that are not linearly separable we additionally implemented the soft margin method presented in [9]. By usage of slack variables, which measure the degree of misclassification, we ensure that hyperplane classifies data as cleanly as possible, while still maximizing the distance to the nearest cleanly split examples. The SVM was trained on four-dimensional patterns (the size of white area, the size of black area, does the white region contains something inside, is the white area of a regular shape). The first two feature values were natural numbers, the other two had binary ones (1 if condition is met, 0 otherwise). After the learning process the SVM model was classifying patterns as platelets if following conditions were achieved: the size of white and black regions did not diverge significantly from learned values, there was nothing inside the white region (value 1) and the size of the white region was regular (value 1). Tests showed that there is no need to use the kernel trick [9]. Each object that will be classified as a platelet is automatically denoted on the image by the green marker (this is the reason for noting the location of each white area). After the classification process is over the user receives an image with each detected thrombocyte marked. Additionally the information about the total number of classified platelets is displayed.

4. TESTING

Output of our presented algorithm was compared to decisions made by the laboratory technician. Testing set consisted of sixty images, which represented the typical collection of obtained images during a laboratory workday. The number of overlooked cells by a human expert oscillated around 10%. Our algorithm reduces this error rate to circa 4%. When the numbers of platelets were small the time computational methods and a qualified technician spent on finding them did not significantly differ. But in cases where the number of cells exceeded one hundred our automatic approach was superior. It should be emphasized that our method does not consume the time of staff members – they only upload the set of images and later receive the results.

5. CONCLUSIONS

We have presented an image analysis / machine learning approach for automatic detection and counting the numbers of platelets appearing in smear images. By using techniques driven from these two sub-fields of computer science we created a method that allows to save time and effort of laboratory technicians, while preserving low error rate. This will contribute to the improvement of quality of laboratory work and allow researchers to spend their time not on routine task of marking thrombocytes by hand, but on more creative work. The first user of our program will be the Institute of Bioengineering from the Wroclaw University of Technology.

BIBLIOGRAPHY

[1] BANKMAN I., Handbook of medical image processing and analysis, John Hopkins University, 2008.

[2] CRISTIANINI N., SHAWE-TAYLOR J., An introduction to support vector machines and other kernel-based learning methods, Cambridge University Press, 2000.

[3] CSEKE I., A fast segmentation scheme for white blood cell images, Proceedings of the IAPR, 1992.

[4] International Conference on Image, Speech and Signal Analysis.

[5] KATZ A. R. J., Image analysis and supervised learning in the automated differentiation of white blood cells from microscopic imag., RMIT University, 2000.

[6] LE M., BRETSCHNEIDER T. R., KUSS C., PREISER P. R., A novel semi-automatic image processing approach to determine Plasmodium falciparum parasitemia in Giemsa-stained thin blood smears, BMC Cell Biology, 2008.

[7] LEZORAY O., CARDOT H., Cooperation of color pixel classification schemes and color watershed: a study for microscopic images, IEEE Transactions on Image Processing, 2002, pp. 783–789.

[8] MICHELI-TZANAKOU E., SHEIKH H., ZHU B., Neural networks and blood cell identification, 1997.

[9] Journal of Medical Systems, Vol. 21, pp. 201–210.

[10] SURI S. S., WILSON D. L., LAXMINARAYAN S., Handbook of biomedical image analysis, Vol. 1-3, Kluwer Academic, 2005.

[11] VAPNIK V., Statistical learning theory, Wiley, 1998.