UDK 004.89

Chetverykov G.[1], Tereshchenko G.[2], Konarieva I.[3]

[1] Doctor of Technical Sciences, Professor of Software Engineering Department, Kharkiv National University of Radio Electronics, grirorij.chetverykov@nure.ua, ORCID iD: 0000-0001-5293-5842

[2] Graduate students of the Department of Software Engineering, Kharkiv National University of Radio Electronics, hlib.tereshchenko@nure.ua, ORCID iD: 0000-0001-8731-2135

[3] Graduate students of the University Complutense, Madrid (UCM), Spain, iulikona@ucm.es, ORCID iD: 0000-0001-9266-9877

DETECTION OF BLOOD CELLS

The structure of the medical image analysis system is considered. The algorithm of the blood cell recognition system is given. Formulated the main tasks to be solved during the morphological analysis of blood. The requirements for the algorithm in determining the leukocyte formula and the detection of blood corpuscles on a smear were determined. A model of color-brightness characteristics is proposed for describing typical images of a blood smear. The threshold values of the sizes of objects are determined when searching for cells. A histogram of the brightness of a typical field of view was investigated. A two-step algorithm for detecting blood cells is described, as well as an algorithm for constructing a dividing line on the plane of relative colors. The results of experiments on real preparations are given. The causes of detection errors are considered.

CELL COUNTING, DIGITAL MICROSCOPY, IMAGE SEGMENTATION.

Introduction

In the tasks of analyzing images obtained with a microscope, in the framework of cytological studies, it is often necessary to count the number of cells of a certain type. When examining blood products, an important task is to count the number of red blood cells, based on the indices of which it is possible to diagnose disorders in blood formation or damage to red blood cells due to various factors.

Among the blood cells distinguish erythrocytes, leukocytes, platelets. The erythrocyte is a nuclear-free cell of pink color, having the form of a somewhat flattened ellipsoid with a depression in the center with an average size of 8 microns. Leukocytes differ from erythrocytes in their larger size, amounting to 9–20 μm, in the presence of the nucleus and in the nature of their color, which can be violet, pinkish, or bright red. Platelets are nuclear-free formations of a round or oval shape with a size of 1-3 microns, with a red-violet center and a pinkish-blue periphery.

There are various methods for counting red blood cells in the blood, some of them use the already modeled base of images of blood cells and their characteristics [1–4], some use threshold decomposition [5] or segmentation using the method of a controlled watershed [6]. There are approaches in which the color characteristics of the image [7, 8] or texture characteristics [8] are used for segmentation. In [9], it was proposed to use the algorithm of the active contour model for the selection of cell contours.

The main problem with counting cells is that they can overlap with each other, as well as change their shape in a certain range. The presence of extraneous noise, foreign objects in the field of view of the microscope further complicates image analysis.

In this work, in order to reduce the effect of noise, it is proposed to use median filtering of images [10] with the subsequent extraction of cell contours by the Canny boundary detector [11]. To improve the recognition of borders, the image is additionally contrasted.

1. General scheme of the algorithm

When leukocytes are detected, two methods of segmentation of "primary objects" are used. At the beginning of the screening, information on leukocyte and erythrocyte colors in this preparation is assumed to be unknown. Therefore, a one-dimensional iterative method is used, based on the study of the peaks of the brightness histogram. After it has found several nuclei of leukocytes and accumulated information about the colors of erythrocytes, a straight line separating the colors of the nuclei and the color of erythrocytes is built on the plane of relative colors fR, fB. If the border between the colors of erythrocytes and nuclei is drawn "with a margin" (errors of the first and second kind are small), it begins to be used for segmentation. So, when selecting primary objects (possibly, nuclei), the transition from the one-dimensional method to the two-dimensional method takes place: instead of segmentation of the brightness histogram, the segmentation of the color plane is performed. The latter option is more stable: there is no danger of missing light nuclei, it is easier to work with poorly focused frames, it is not necessary to accurately determine the position of the peak of the background [4].

The algorithm for extracting fragments consists of four steps.

1. The selection of primary objects - possibly nuclei. In this case, two different segmentation methods are used — according to a histogram of brightness or on a color plane. The second method is preferable, but at the initial stage of sample accumulation the first one is used.

2. Verification of primary objects for compliance with an already accumulated sample of nuclei, which is possible if the number of accumulated objects more than 5. Primary objects are classified into nuclear fragments and artifacts.

3. Fragments of nuclei are combined with each other, and an attempt is made to build cytoplasm around them. As a result, we obtain a rectangle, inside which there is one leukocyte.

4. If a brightness histogram was used for the segmentation of the primary objects, and there were several peaks on it, possibly corresponding to the nuclei, then the number of found nuclei pixels is compared with the estimated (number of pixels in the peak). As a result of testing the hypothesis of which peak should be considered the peak of the nuclei, it can change and the algorithm can be started from the first step anew.

As a rule, there is no need for iterations; the algorithm consists of three steps: primary objects → nuclear fragments → leukocytes.

2. A segmentation algorithm based on a brightness histogram

The proposed leukocyte detection algorithm consists of two stages, which can be repeated several times for the same frame. At the first stage (based on the study of the histogram of the brightness of the frame and the history of the search) are selected threshold values for brightness G and share blue fB. At the second stage, the sets of pixels that meet these conditions (primary objects) are examined to determine whether they can be considered as leukocyte nuclei. If the total number of pixels in these fragments is significantly less than the previously estimated number, then the selection of threshold values is considered unsatisfactory and the algorithm is launched again, etc.

The first stage is the study of the histogram, the choice of threshold values. The first step is to localize the peak associated with the image background, which will later be the reference, both when calculating the optical density and when determining the relative colors for the remaining pixels. Absolute values have to be used only if the background peak is not localized. An extreme right peak with a rather small dispersion is chosen as the peak of the background pixels: the standard deviation is less than 10 digits [5].

Next, a list of maxima (peaks) is compiled that could correspond to leukocyte nuclei. To do this, their optical density must be sufficiently large (the empirically found boundary > 0.6) and the average blue fraction $fB = B/(B + G + R)$ for pixels at this maximum should exceed the same value for the background by 0.03 (empirically found border). If there are several such suspicious maxima, they are selected sequentially one after the other (in this case the second stage of the algorithm is called), starting with the brightest one. Held threshold segmentation by brightness and relative proportion of blue $G < G\max$; $fB = B/(B + G + R) < fB\max$.

The obtained primary objects are compared with already existing nuclei. If they are not qualified as nuclei, and this is possible in the presence of optically dense and bluish red blood cells, stains of paint, large platelets, then the next maximum will be selected. If the leukocyte peak is not distinguished at all, then threshold values are used, which are no longer based on the current histogram, but on the prehistory of the search, and if there is none, then on a priori values.

The second stage is the study of the obtained fragments. This part is independent of the method by which the primary objects were obtained. The algorithm for checking selected objects consists of three cycles. In the first cycle, too large ($A > 2000$ μm^2) and too small ($A < 11$ μm^2) objects are discarded.

Further, the optical density and color characteristics are measured. If there is a prehistory of the search, then by the criterion of "three sigma" excessively light objects with a low optical density are discarded and the procedure for checking colors is called. The remaining objects are placed in the class of conditional nuclei of leukocytes.

In the second cycle, the completion of the cytoplasm around the nucleus takes place on the remaining objects. To the cytoplasm include a coherent set of nearby pixels, which with a high probability (more than 0.95) are not pixels of red blood cells or background. The constructed set is rejected if it is too large (more than 2000 μm^2) or the form factor of its external border (square of perimeter / area) exceeds a sufficiently large value (more than 50). It is often enough, with a close diligence of the white blood cells of similarly colored erythrocytes, the cytoplasm cannot be completed in such a simple way. Then the core or its fragment is placed inside a rectangle with added frames of 15 μm.

After completing the cytoplasm, agglutination of nuclear fragments is performed. This is necessary since the nuclei of neutrophils are detected in the form of several fragments. In this case, the separation of cells that are close to each other. This is possible if the cells lie in islands that are not connected with each other, surrounded by background.

After the second cycle, the objects obtained are considered as separate leukocytes. In the third cycle, the sizes of these objects are checked again, and too large objects are discarded.

If the segmentation used a brightness histogram, then at the end a check is made for the consistency of the assumptions and the results obtained. For this, the found number of pixels of leukocyte nuclei is compared with the number of pixels at the peak of the histogram, which was assumed to be the corresponding nuclei. If the differences are significant (exceed> 50%) and to the left of the peak of the supposed leukocytes there was another one, then the above algorithm runs again. In fig. 5 shows the scheme of the detection of secondary objects.

Algorithm for checking the primary object for belonging to the leukocyte nucleus group. Leukocyte nuclei do not constitute a homogeneous group. Therefore, it is not necessary to relate strictly to checking for the belonging of a new object to a two-dimensional, normal distribution, even for average values of relative colors. In addition, in the smear screening process, it is desirable to use the verification algorithm as early as possible when the number of objects accumulated is small. Therefore, the algorithm proposed below is heuristic. It is based on the following provisions.

1. Each new object is compared with two groups: a group of nuclei and a group of red blood cells.

2. When determining the probability of a new point belonging to an existing group, the probability is calculated twice. In the first case, the probability P_1 is calculated before the point is added to the existing statistics, and in the second case — P_2, after such an addition. Obviously, $P_2 > P_1$. Such a calculation of two probabilities at once is necessary if decisions are made on the basis of small samples, and the number of objects accumulated without checking is usually just a little — about 5. If the number of objects in a group exceeds several dozen, the difference between P_1 and P_2 almost disappears.

3. If the object under study is more "blue-red" than already accumulated nuclei, then it is accepted in any case. In other words, the possibility of emissions in this "blue-red" side is not taken into account. Conversely, if the object is more "green" than red blood cells, then it is rejected in any case.

4. If the probability P_2 calculated relative to the group of erythrocytes is greater than the corresponding probability calculated relative to the nuclei, then the object is rejected.

5. If a decision is not taken on the basis of the preceding paragraphs, then it is made taking into account three probabilities. The probabilities P_1 and P_2 estimate the deviation of the average value, and the probability P_3 — the deviation of the area of the ellipse of scattering from the characteristic for a group of nuclei. The object is rejected if $P_2 < 0.01$ or $P_3 < 0.01$.

For a given primary object, the corresponding probabilities $P_1 = 0.53$, $P_2 = 0.65$, $P_3 = 0.45$ are high, so it will be correctly qualified as a core and added to the statistics of nuclei.

3. Erythrocyte isolation and counting method

The one-dimensional median filter is a "sliding window" with a length of N samples, in which the central element is replaced by the median (ie, the middle element of the sequence, ordered in ascending order of the signal values in the "window"). Thus, the operation of the median filtering of a K-dimensional sequence of signal values $s(k) = s(x_k)$, $k = 1, \ldots K$ characterized by the ratio

$$\underset{1\le k\le N}{\text{med}}\{S_k\} = \begin{cases} 0{,}5(s_n + s_{n+1}), N = 2n \\ s_n, N = 2n-1, \end{cases}$$

where the fixed value $n = 1, 2, \ldots$ determines the filter aperture.

The next stage consists in the selection of boundaries, after which the method of connected components with a connectivity criterion along eight neighbors is separated into individual contours. For each connected region, the area of a convex polygon describing the contour is calculated. For given thresholds, sections that are too large or too small are cut off. Thresholds are selected based on the estimated real cell area.

Then, points are selected from each individual contour at equal intervals along the contour length. The points are connected in pairs with each other, and a perpendicular is drawn through the middle of the obtained segment. It can be described by the equation

$$y = -\frac{x_2 - x_1}{y_2 - y_1}x + \frac{y_2^2 - y_1^2 + x_2^2 - x_1^2}{2(y_2 - y_1)},$$

where (x_1, y_1) – coordinates of the first point;
(x_2, y_2) – coordinates of the second point.

The location of the intersection point of adjacent perpendiculars is preserved. The point of intersection of two perpendiculars is calculated as

$$x = \frac{b_2 - b_1}{a_1 - a_2},$$
$$y = a_1 x + b_1 = a_2 x + b_2,$$

where $a = \frac{x_2 - x_1}{y_2 - y_1}$ – coefficient of inclination of the perpendicular; $b = \frac{y_2^2 - y_1^2 + x_2^2 - x_1^2}{2(y_2 - y_1)}$ – the coefficient of perpendicular displacement.

The operation is performed for all pairs of perpendiculars for the various steps of taking points. As a result, a cloud of points is formed, which are located more densely in areas that are the centers of the radius of the contour curves.

Cells stuck together with one another or superimposed one upon the other are rather difficult to segmentation by methods based on analyzing the size or shape of the areas inside the contours. Gaps in the contours complicate the contour segmentation of cells. The proposed method is a single image for all areas of

the contours of their centers, thus allowing to solve the problems indicated above.

The picture, composed of the obtained points of intersection of perpendiculars, undergoes morpho logical processing, as a result of which only dense and rather large clusters of points remain. These clusters correspond to the putative cell centers. Using the method of connected components, clusters are counted, which should correspond to the number of cells in the image.

4. Experimental results

In an experimental study of the proposed method, an image of a blood sample was taken using a microscope (Fig. 1).

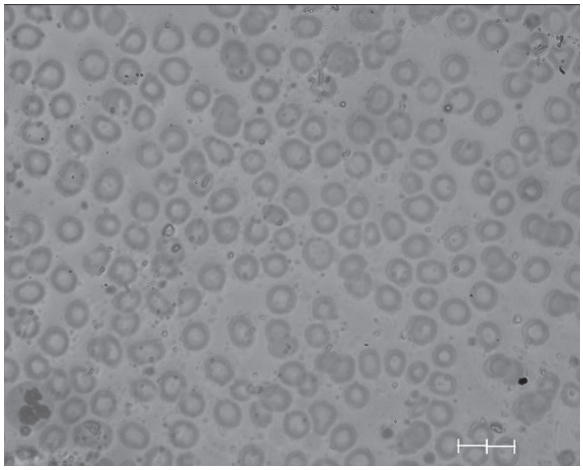

Fig. 1. Test picture

First of all, the image was converted from color to black and white. The window size in the median filtering was selected based on the average cell size and was 16x16 pixels, which corresponds to 20% of the cell diameter. The result of the median filtering of the test image is shown in Fig. 2.

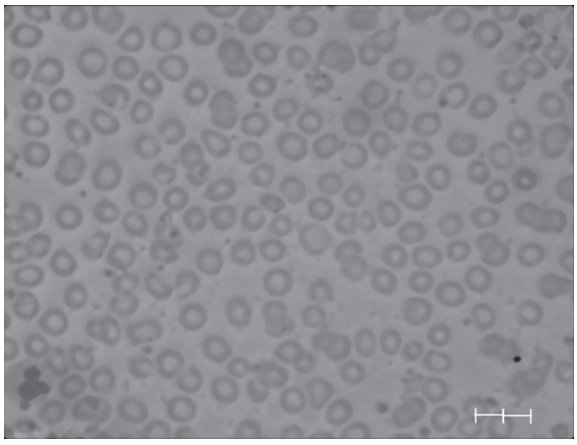

Fig. 2. The result of median filtering

After linear image contrasting, a Canny boundary detector was applied (Fig. 3). The Canny algorithm first smoothes the image to remove noise. Borders are then selected where the gradient of the image acquires the maximum value, with only local maxima marked as borders. The next step is to determine the potential boundaries of the double threshold filtering. Total boundaries are determined by suppressing all edges that are not associated with specific boundaries.

As thresholds for removing too large or too small contours in the image, two values were chosen: 0.05S as the lower threshold and 4S as the upper threshold, where S is the approximate area of the cell image calculated on the basis of its diameter.

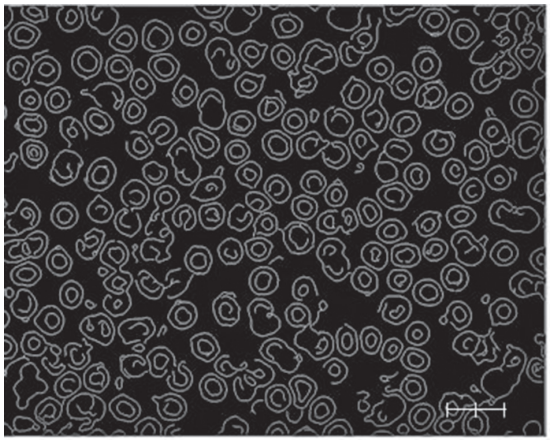

Fig. 3. Image of edges with large and small contours removed

As a result of the construction of perpendiculars (Fig. 4), for the segments between points taken with 3–40 pixel intervals, a cloud of intersection points was obtained for each contour (Fig. 5).

Fig. 4. Perpendiculars; spacing between points = 20

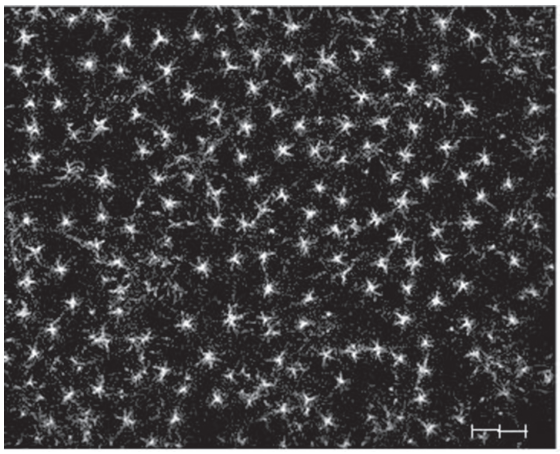

Fig. 5. Point clouds

The number of connected components was 220. The real number of cells in the image was obtained equal to 209. The number of false positives was 23, the number of unrecognized cells was 12. The method was tested on four different images containing blood cells. The average probability of correctly counting the number of erythrocyte cells was 86%. Compared with the methods using threshold decomposition [5] or segmentation by the method of controlled watershed [6], the proposed method gave the best results. However, in comparison with the methods proposed in [1–4, 9], the probability of correctly counting the number of cells turned out to be less, since the border detector incorrectly selected edges on test images due to strong noise and the presence of fuzzy boundaries. For such images, additional preprocessing methods are needed to improve the efficiency of edge extraction.

Plots with an area of less than 5 pixels were removed from the image of a cloud of points, and then the operation of closing a binary image with a mask of 8x8 pixels was performed. The result of the morphological processing of the image of intersection points superimposed on the original image is shown in Fig. 6.

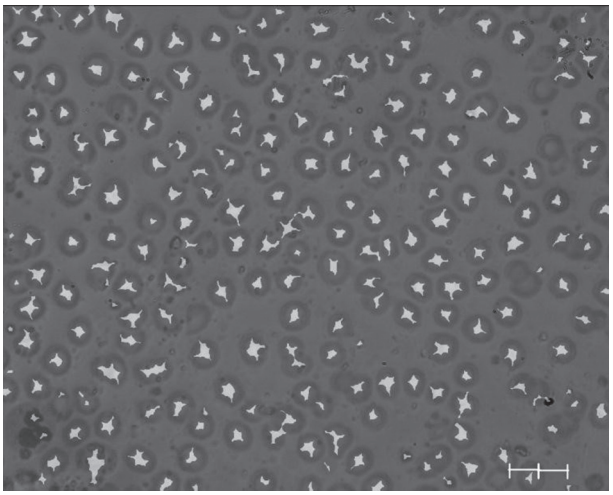

Fig. 6. Morphologically processed image of intersection points superimposed on the original image

Conclusions

Due to the fact that a border detector is used to identify cells, the segmentation results do not depend on the color of the cells, their texture and internal structure. The method with rather high accuracy segments the cells stuck together with each other or superimposed on each other. In conditions of noisy source image method showed good results.

The proposed algorithm allowed segmentation and counting of blood cells with an accuracy of 86%. The number of false cell detections is on average higher than that of other methods, which can be explained by the presence of a large amount of noise on the test images, as well as by fuzzy cell boundaries. A higher probability of correctly counting the number of cells can be achieved if, in parallel with the proposed method, you use others, specifying the result of the segmentation of one method by the results of another, as well as using other algorithms for preliminary processing of the original image. In the future we plan to develop an algorithm for image preprocessing to increase the efficiency of the proposed method, as well as a combination of the proposed method with others.

Reference*s*

[1] *Dahmen J., Hektor J., Perrey R., Ney H.* Automatic Classification of Red Blood Cells Using Gaussian Mix-ture Densities // Proc. Bildverarbeitung für die Medizin. – 2000. – P. 331–335.

[2] *Costrarido L.* Medical Image Analysis Methods: Evaluation Strategies for Medical-image Analysis. – Taylor & Francis, United Stated of America, 2005. – P. 433–471.

[3] *Kumar B.R., Joseph D.K., Teager T.V.S.* Energy Based Blood Cell Segmentation // 14th International Conference on Digital Signal Processing. – DSP, 2002. – 1–3 July. – Santorini, Greece. – V. 2. – P. 619–622.

[4] *Bamford P.* Empirical Comparison of Cell Segmentation Algorithms Using an Annotated Dataset // Proc. IEEE International Conference on Image Processing. – 2003. – V. 2. – P. 1073–1077.

[5] *Mukherjee D.P., Ray N., Acton S.T.* Level Set Analysis for Leukocyte Detection and Tracking // IP. – 2004. – V. 13. – № 4. – P. 562–572.

[6] *Park J., Keller J.M.* Snakes on the Watershed // IEEE Transactions on Pattern Analysis and Machine Intelligence. – PAMI, 2001. – V. 23. – № 10. – P. 1201–1205.

[7] *Sinha N., Ramakrishnan A.G.* Blood Cell Segmentation Using EM Algorithm // Proc. Third Indian Confer-ence on Computer Vision. Graphics Image Processing (ICVGIP), 2002. – Ahmadabad, India, 2002, Decem-ber 16–18. – P. 376–382.

[8] *Kumar R.S., Verma A., Singh J.* Color Image Segmentation and Multi-Level Thresholding by Maximization of Conditional Entrophy // International Journal of Signal Processing. – 2006. – V. 3. – № 1. – P. 121–125.

[9] *Mcinerney T., Terzopoulos D.* Deformable models in medical image analysis: A survey // Med Image Anal. – 1996. – P. 91–108.

[10] *Cormen T.H., Leisorson C.E., Rivest R.L.* Introduction to Algorithms. – MIT Press, 1990. – P. 185–191.

[11] *Canny J.F.* A computational approach to edge detection // IEEE Trans. Pattern Analysis and Machine Intelli-gence. – 1986. – P. 679–698.

The article was delivered to your editory stuff on the 24.05.2019