Configurable Cell Segmentation Solution Using Hough Circles Transform and Watershed Algorithm

Oleh Hramm
Department of software engineering
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
oleg.gramm@gmail.com

Nataliya Bilous
Department of software engineering
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
bilous.n.v@gmail.com

Iryna Ahekian
Information and computer center
Kharkiv National University of Radio Electronics
Kharkiv, Ukraine
ostpc2013@gmail.com

***Abstract* — There are many methods for obtaining microscopic images, they can be obtained for different types of cells in different environments, which makes it impossible to create universal recognition algorithms. Therefore, for each specific task, it is necessary to implement a separate algorithm that will be effective for the existing set of images, and take into account the peculiarities of these images. The task of this work is the development of a flexible and customizable algorithm that can be configured to segment cells on different types of images. As a result, a solution was developed that has many customizable parameters to optimize the result for a specific data set. Also, this it is resistant to a lot of noise and artifacts that can occur on images, such as uneven background, small debris, loss of focus when shooting.**

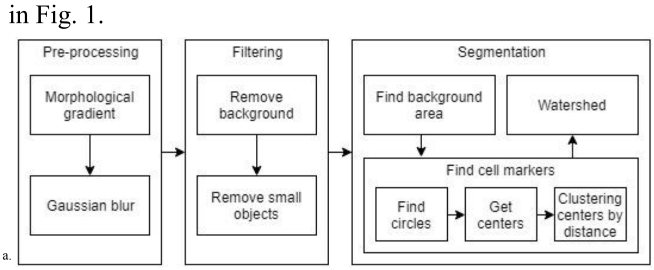
Keywords* — *cells recognition; segmentation; watershed; Hough transform.

I. INTRODUCTION

The principles of constructing the computer vision system imply the presence of few levels (pre-processing and filtration, segmentation, classification). The pre-processing phase changes the properties of the image to optimize the result in further processing steps. The filtering phase removes areas that do not contain important information and complicate the segmentation process. The segmentation phase separates the elements and their limits, which ultimately improves the quality and accuracy of the diagnosis [1]. Each of this phases have its own set of algorithms. However, if the existing set of algorithms at the lower level (pre-processing and filtering) allows qualitatively meeting all the requirements for this level when solving the task, then the solution existing at the middle level (segmentation) is insufficient. So traditional image segmentation algorithms based on the principle of proximity of object points by color (brightness) and don't take into account the intersections of objects, exclude the possibility of further classification with an acceptable [2]. Also cell structures can look completely different, have a transparent, translucent or opaque structure, have pronounced boundaries or have the following. Also, images can have a lot of noise and other objects that should be excluded from processing. It is also a big problem that all these properties can manifest themselves differently on different data sets. So it is actual task to develop algorithms for specific classes of images to increase effectiveness of used algorithm [3]. The purpose of the current work is to develop a solution that will allow the algorithm to be implemented in such a way that it is applicable to different data sets and has a high result. The best algorithm was found in the paper [4]. This algorithm uses a watershed algorithm for cell segmentation and Hugh transform for clustering intersecting cells. The purpose is achieved by giving ability to configure the algorithm for the properties of a specific data set. Accordingly, the results of this work can be used for future works with newer tasks of cell segmentation. So the novelty of this work lies in the versatility and the possibility of optimization for a different range of segmentation tasks. This paper describes the implementation of the modified method [4] so that we can parameterize this algorithm for the most efficient use of different data sets with the ability to optimize the result by configuring this algorithm for a specific data set. The implementation was performed in software application with open source code [5] using the OpenCV framework for Windows OS.

II. DESCRIPTION OF THE DEVELOPED SOLUTION

The algorithm was implemented for processing fluorescence and optical microscopy images of cell populations. It has a lot of configurable parameters, so this algorithm can be optimized for concrete properties of processed image. The structure of the algorithm is described in Fig. 1.



a.

Fig. 1. Algorithm structure

The process of the algorithm is shown by the example of processing a fragment of an image with cells. The image is presented in Fig. 2. The image has uneven backgrounds, as well as noise in the form of small objects, which should not be segmented as individual cells.

b.

Fig. 2. Original image

Then in this image cells will be segmented using the developed algorithm.

A. Pre-processing

On pre-processing stage, the input image is converted to another image with a different property. The properties of the image is changed so that it helps to more effectively carry out the process of filtering and segmentation.

1) Correcting background irregularities: The problem with many images is that they have different brightness of background areas, which complicates the task of removing it. To smooth the background, morphology transformation was applied, which allows to remove the dependence of the image on the heterogeneity of the background and leave only the boundaries of the contours [6]. The classical morphological gradient operator for greyscale images is the difference between a dilation and an erosion. This is an image in which each pixel value indicates the intensity of the contrast in close proximity to this pixel [7]. As a result, an image was obtained with a uniform level of background brightness for further processing.

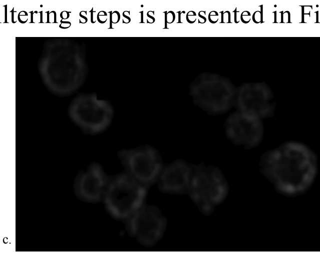
2) Reducing excess clarity: Reducing the clarity is necessary to remove unnecessary detail. The image will be processed using Gaussian blur. Thus, the image was spared from slight unevenness and noise and only the rounded contour of cells remains without image detail. This will make it possible to more effectively conduct the process of locating circles at the segmentation stage, because the search will only produce on the contours and the result will not be affected by noise.

B. Filtering

At the filtering stage, all unnecessary data from the image is deleted to get an image containing only the cell area. This is necessary to prepare the image for the next stage - segmentation, because the results of segmentation will be obtained only on the basis of cell data, without the influence of other factors.

1) Background removal: Removing the background is necessary to obtain object outlines, separate from the background of the image. It is implemented using threshold filtering.

2) Filtering small objects: The image of the objects is devoid of background, but in the picture there is noise in the form of small objects that are not cells. To filter small objects, all continuous groups of pixels, which have a smaller set of pixels than declared in parameters are deleted. The result of filtering steps is presented in Fig. 3.



c.

Fig. 3. Filtering result

This is the original image which the morphological gradient was processed, the background removed and small objects are deleted.

C. Segmentation

Segmentation of images is one of the most important image processing technologies for biological images. Its goal is to distribute the input image to the regions [6]. The increase in the reliability of the results of the segmentation of histological objects is achieved due to the determination of the specific properties of individual objects, it is of an individual nature and cannot be used for the analysis of histological images as a whole [8]. For segmentation, the watershed algorithm is used. For this, background area and cell markers are a set of points, each of which will correspond to one cell must be found.

1) Finding background area: To create a mask from background region, the thresholding binarization algorithm is applied, the result of which is shown in Fig. 4 [9].

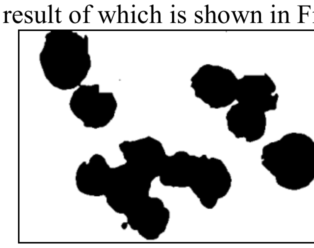


Fig. 4. Segments on original image

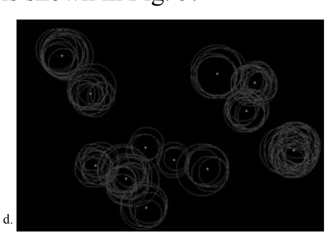
White marks an area that is certainly the background of the image and does not contain objects.

2) Finding cell markers: The algorithm for finding cell markers contains several steps.

a) Finding circles: Determination of circles is performed by Hough gradient method [10].

b) Detection of circle centers: In order to isolate cell markers, only the centers of the found circles should be left.

c) Clustering of centers by Euclidean distance: The centers of circles is clustered by Euclidean distance. As a result, an image was obtained with cell markers, which can later be used in the watershed algorithm. The result of the marker search is shown in Fig. 5.

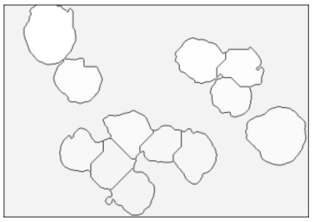


d.

Fig. 5. Result of finding circles and clustering

The figure shows the circles found and the result of the clustering of their centers. As a result, there is one cluster per cell, so the center of this cluster can be used as a marker to define the segment.

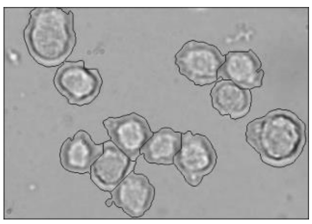
3) Accepting watershed algorithm: Applying the previous stages of image processing, an image with cell markers was obtained, and a mask with a region filled with cells. Having this data the watershed algorithm can be applied. As a result, an image was obtained *(Fig. 6).*



e.

Fig. 6. Watershed algorithm result

This is a set of segments, where each segment corresponds to one cell. The resulting segments is used on the source image (Fig. 7)



f.

Fig. 7. Segments on original image

As a result, an image with a set of segments was obtained, where each segment corresponds to one cell. This will allow further implementation of algorithms for analyzing the parameters of cells and their states for each of the segments.

III. CONFIGURABLE PARAMETERS

The algorithm has several configurable parameters. This will allow you to customize the algorithm for a particular data set, which will allow you to raise the efficiency when processing it. The available parameters are listed below.

1) Blur level: This parameter changes the intensity of the reduction in the sharpness of the image. The higher the level, the more effective the noise will be eliminated. At excessively high levels, the contours of the cells can be too blurred. If the level of intensity is insufficient, the remaining noise will negatively affect the finding of circles during the search for markers. Effective is the value at which the overall image noise is reduced, but the contours of the cells remain. If the image doesn't have background noise, the parameter value must be 0.

2) Background brightness threshold: This parameter determines the brightness value for which all areas with a value below the parameter are considered as the background image. All pixels with a brightness value below the specified parameter will be deleted. At excessively high level the background can be classified as areas that are cells. If the level is insufficient, parts of the background may not be removed and later they will enter the cell segment. Effective is the value at which the largest amount of background and the smallest number of areas with cells will be deleted.

3) External minimum size: This parameter determines the minimum size of objects that will be processed. All objects smaller than the specified size will be deleted from the image. At excessively high level, the areas that are actually engaged in the cell can be removed from the image. If the level is insufficient, the image may contain objects that are not cells and will be defined as separate segments. Effective is the value at which all unnecessary objects will be deleted, but all the areas with the cells will remain unchanged. If the image contains only cells and does not contain extraneous objects to be deleted, the value of the parameter must be 0.

4) Internal minimum size: This parameter is necessary in case the inner part of the area with cells has been deleted at the background filtration stage. All areas that are highlighted as a background and have a size smaller than the specified parameter will be assigned to the area with the cells. At an excessively high value, small areas of the background can be attached to the cell area, which are surrounded by cells. At insufficient level, the middle of some areas with cells can be interpreted as a background and the contour of a segment can be determined incorrectly. Effective is the value at which the contour of the regions with the cells will be as reliable as possible. If all the internal cells areas have not been deleted during the background removal process and the background area is determined correctly, the parameter value must be 0.

5) Sensitivity of circles detection: This is a parameter for adjusting the sensitivity of the definition of circles, during the search for markers. It determines how strong the contour of the circle should be. With an excessively high value of the parameter, the actual contour of the cell may not be defined as a circle. With insufficient level also increasing your probability of false circles which will make it difficult to further the process of clustering. The value at which the contours found will coincide with the contours of the cells will be effective.

6) Completeness of circles: This parameter determines how many points of the contour you need to find to define the contour as a circle. The effect is the same as when changing the previous parameter.

7) Minimum circle radius: Parameter defines the minimum size of the circle that can be defined. Corresponds to the minimum size of the cell in the image.

8) Maximum circle radius: Parameter defines the maximum size of the circle that can be defined. Corresponds to the maximum size of the cell in the image.

9) Minimum distance between clusters: Parameter defining the distance between the centers of circles, at which they will be defined in one cluster. With excessively high value of the parameter, the points belonging to the regions of different cells can be brought into one cluster. With an insufficient parameter, the points belonging to the region of one cell can be divided into different clusters. Effective is the parameter at which each cluster will correspond to one cell.

IV. ESTIMATION OF EFFICIENCY

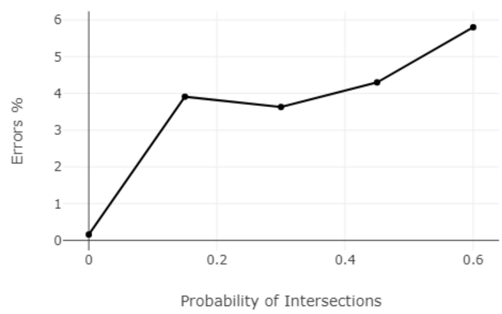
The evaluation of the efficiency was verified on the results of the work for the task of determining the number of cells in the image. To evaluate the efficiency, a set of images was processed [11, 12] with optimized parameters to this set.

A. Overall results

The general results were obtained by processing a set of images with a different number of objects and a different probability of intersection. The sample is 600 images. As a result of testing, an average error is 2.69%.

B. Dependence on the probability of intersection

The results of processing, depending on the probability of intersection, were obtained with the help of several sets of images, all of them have 300 objects, and have different probability of crossing cells. The results described in graph in Fig. 8.



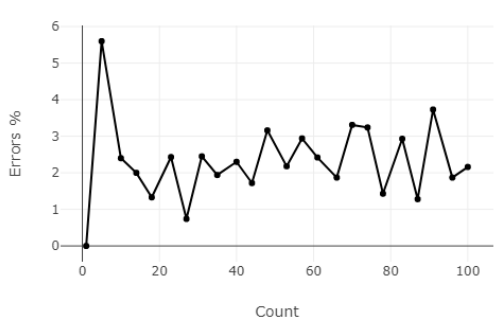
g.

Fig. 8. Dependence of probability of intersection to the error

On the results, the algorithm without the probability of crossing the cells has good results. The error in this case is 0.16%. But it has a clear correlation between the increase in the probability of crossing cells and the increase in error. The result of the intersection probability values from 0.15 to 0.4 is linear enough and don't change. But since the value of probability of intersection 0.4 there is an increase of error.

C. Dependence on the number of cells

The results of processing, depending on the number of cells, were obtained by processing several sets of images grouped by the number of cells in the picture. The results is described in graph in Fig. 9.



h.

Fig. 9. Dependence of number of cells to the error

On the results, there no correlation between the growth of the number of cells and the increase in the amount of error. The ideal work of the algorithm can be seen only on images with one cell. With an increase in the number of cells, the error increases. At most values of the number of cells, an error is between 1% and 3.5%.

IV Conclusion

General result of working algorithm has average error of 2.69% on a sample parameters. Also this algorithm allows many parameters to be configured, which will allow to obtain an acceptable level of error for a lots of data sets of images with different properties. So the data obtained with this algorithm can be used to interpret and analyze cells in the image and the algorithm itself can be improved for a specific data set by configuring parameters.

References

[1] D. Boyko, A. Podoroznyuk, A. Filatova. "The main stages of image processing in the design of biotechnical systems in medical radiology", The National Technical University "Kharkiv Polytechnic Institute", pp. 85 – 86, 2012. (in Russian)

[2] A. Kovrigin, "Application of the principles of building computer vision systems in the problem of image analysis of cellular structures", Scientific Journal of KubSAU, pp. 1 – 3, 2007. (in Russian)

[3] M. Tarkov, "Estimation of the number of cells on the images of cytological plant preparations", Rzhanov Institute of Semiconductor Physics, pp. 187 –190, 2013. (in Russian)

[4] M. Romero-Rondón, L. Sanabria-Rosas, Lola Bautista-Rozo, Alfonso Mendoza-Castellanos "Algorithm for detection of overlapped red blood cells in microscopic images of blood smears", DYNA, no. 84, pp. 187-194, 2016.

[5] Repository OlegGramm/CellsRecognition on the GitHub. Available from: https://github.com/OlegGramm/CellsRecognition

[6] Uchida S., "Image processing and recognition for biological images," Dev. Growth. Differ. 55, pp. 523–549, 2013.

[7] J. Rivest, P. Soille, S. Baucher, "Morphological gradients", Journal of Electronic Imaging, no. 2(4), October 1993.

[8] S. Ablameyko, A. Nedzved "Processing of optical images of cellular structures in medicine", OIIP NAS Belarus, pp. 35 – 55, 2005. (in Russian).

[9] R. Firdousi, S. Parveen, "Local Thresholding Techniques in Image Binarization", International Journal Of Engineering And Computer Science ISSN:2319-7242 vol. 3 issue 3 March, 2014 Page No. 4062-4065.

[10] G. Bradski, A. Kaehler, "Learning OpenCV: Computer vision with the OpenCV library", O'Reilly Media, Inc., 2008

[11] Image set BBBC004v1 from the Broad Bioimage Benchmark Collection. Available from: https://data.broadinstitute.org/bbbc/BBBC004/

[12] Image set BBBC005v1 from the Broad Bioimage Benchmark Collection. Available from: https://data.broadinstitute.org/bbbc/BBBC005/