

Development of automated computer vision methods for cell counting and endometrial gland detection for medical images processing

Daniel Sergeev
ICST
SPBSTU
SPB, Russia
densvr1@gmail.com

Anna Drobintseva
SPSPMU
SPB, Russia
anna.drobintseva@gmail.com

Nikola Kukavica
ICST
SPBSTU
SPB, Russia
nikola.kukavica.94@gmail.com

Alexander Andreev
SPSPMU
SPB, Russia
alexander597@mail.ru

Slobodanka Cenevska
ICST
SPBSTU
SPB, Russia
slobodankacenevska@yahoo.com

Pavel Drobintsev
ICST
SPBSTU
SPB, Russia
drob@ics2.ecd.spbstu.ru

Abstract - current work is focused on the processing of medical images obtained by performing a pathomorphological analysis of preparation. The algorithms for processing images of nuclei of light and confocal microscopy and tissue of light microscopy were considered in particular. The application of the proposed algorithms and software for detecting pathologies was justified.

Keywords - computer vision, image processing, digital pathology, detection and classification, cell and tissue nuclei, light microscopy, confocal microscopy.

I. INTRODUCTION

The processing of medical images is an extremely important issue for biology and medicine. Pathomorphologists have to process hundreds of images of preparations per day. Their work can be automated due to computer analysis.

Medical image processing can be performed in semi-automatic and automatic modes. Semi-automatic mode is based on manual adjustment of simple, intuitive parameters for evaluating single microphotographs. Automatic mode otherwise does not require both direct operator intervention and initial settings in the processing of preparations.

Modern experts in the field of pathomorphology have an access to a wide range of technologies that make it possible to carry out various measurements depending to required tasks. A striking example of such technologies is the universal ImageJ software [11], which is capable for performing operations aimed at evaluating the geometry and color gamut of the resulting images. The main idea of this software is to write macroses that require a minimum understanding of computer technology from a specialist. This approach makes ImageJ flexible, but not user friendly. More intuitive tools are commercial software such as VideoTestMorphology [15] and ImgeProPlus [17]. There are various microcopy solutions. The most popular software at the moment are compared in Table 1.

Thus, there are a lot of modern software, designed for morphometry. However, there are a number of facts that severely limit the domain specialist in choosing his own tools. Firstly, most of the free automatic and semi-automatic programs are incomplete and require bioinformatics in the team. Secondly, the already collected, semi-automatic programs in most cases are universal, which means the absence in the process of taking into account the parameters

of the microscope, the type of tissue being examined and the lighting of a particular picture. These parameters must be driven manually, relying on the empirical experience of the specialist responsible for setting up the program. Thirdly, the most convenient and user-friendly programs have a high cost. Therefore they are not available in small laboratories.

The aim of this work is the development of software that partially eliminate the shortcomings of modern non-commercial software for processing digital images with automatic objects recognition by series of images and extracting from them the minimum set of basic features necessary for researchers to work with.

Typical operations for this task are initialization, localization, segmentation, shape analysis, modeling, analysis of cell parameters, etc. [2]. Although there are many methods of segmentation, precise segmentation is a complex task, and it plays a significant role in biological imaging studies

II. TYPICAL TASKS

A. Processing of cell and tissue preparations

Processing of preparations is carried out both at the cellular and tissue levels. In both cases, the cells are usually tinted with the help of their special reaction to the examined "marker". In this study, nuclear markers such as estrogen receptors (ER) and progesterone (PR) were detected. A quantitative analysis of nuclei with receptors for ER and PR is essential, since they are involved in the mechanisms of growth and metastasis of tumors. The research of the expression of ER and PR is included in the standard of examining patients with breast cancer, as it allows us to determine the sensitivity of the tumor to hormone therapy and to clarify the prognosis of the disease.

The research of ER and PR is also used in the diagnosis of infertility, endometrial hyperplasia [3].

Pathomorphologist needs to calculate the number of cell nuclei highlighted in color on the preparation, which correspond to the expression of the researched markers, as well as the total number of nuclei per unit area.

Another assignment of the pathomorphologist is to isolate the contours of glands and tissues on the preparation and determine the number of glands with high total marker expression.

TABLE I. SOFTWARE FEATURES

Software	Language	Automatic mode	Machine learning	Coding	Open source
Image J	Java	macros	plugin	no	no
Cell profiler	Python	macros	no	yes	no
Orbit Image Analysis	Java, python	yes	yes	no	yes
Axio vision 4.8	-	macros	yes	no	no
Video test morphology 5.2	-	macros	yes	no	no
CellSens	-	macros	no	no	no
IMAGE-PRO-Premium	.NET	yes	yes	no	no
BioVision	-	yes	yes	no	no

According to statistical data, such as the number of nuclei and glands with and without marker expression, conclusions about the structure of the tissue are made, the effectiveness of treatment is considered, a diagnosis and the prognosis of the disease are specified.

B. Light and confocal microscopy

Medical preparations can be obtained in various ways. Preparations obtained using light and confocal microscopes are studied in this work. Confocal microscopy has been used relatively recently and, compared to the light one, gives more contrasting color images, with staining of the nuclei with different fluorescent dyes with antibodies of the corresponding markers. However, confocal microscopy requires special equipment and the quality of the preparation, which makes it impossible for mass application. Therefore, for the diagnosis of diseases light microscopy is still used in most cases [4].

III. POPULAR ALGORITHMS

Medical imaging algorithms typically consist of the following steps:

A. Image pre-processing

It is used to create conditions that increase the efficiency and quality of the isolation and recognition of nuclei in medical preparations. It includes morphological operators and filters, border detectors, filters with brightness normalization [5].

B. Detection of objects of interest

At this stage, the X and Y coordinates of the proposed center of each object of interest (nucleus) are determined. As a result of the stage, a set of objects of interest, which are probably nuclei is obtained. The algorithm parameters at this step are set in the particular way in order to create the redundant number of objects of interest. In other words, the detection of false objects is acceptable, but the admission of real cores is not. Basically, the following algorithms are used: active contour algorithm, watershed algorithm, image segmentation by known classes, segmentation with preliminary detection of class boundaries [5].

C. Selection of characteristics of objects of interest, classification and arrival at a decision

The final stage allows to attribute each of the objects of interest to one of the target classes. For the task of classifying

nuclei using fluorescence and light microscopy preparations, the topological, texture, and color intensity-based characteristics of following classes are distinguished.

At the classification stage, objects of interest classes are determined by markers used for coloring the nuclei and include:

- A nucleus not highlighted with a marker;
- Background (stroma);
- A core highlighted with marker
- Several nuclei

The case with several nuclei should be considered separately (see paragraph D).

The method for determining the intensity and clustering of the color histogram allows you to automatically determine the number of markers and their colors. This saves the precious time of pathomorphologists.

The following classifiers are used for pathomorphological analysis: the support vector method, Bayes classifier, Haar cascade, convolutional neural network [6].

D. Methods for the separation of overlapping nuclei

A specific task is the separation of overlapping nuclei. This can be caused by cell division, the camera's viewing angle in the process of shooting the preparation, and also, the location of the nuclei on top of each other in the depth of the examined tissue. The following approaches can be used to separate the fused nuclei and accurately determine their number:

- The method of active contours with the preliminary use of erosion;
- Classifiers (convolutional neural network), previously trained in classes that determine the number of cells in the area of a given size;
- Watersheds algorithms;
- Segmentation algorithms focused on topological features of objects [7].

IV. IMAGE DATASET

Many images of cell structures and tissues preparations of light and confocal microscopy were collected and labeled (see Table 2).

Material and equipment for shooting images was provided by the Institute of Obstetrics, Gynecology and Reproductology Ott. The shooting of individual classes of images was carried out with a fixed scale of the microscope. Image preparations of various types of tissues with different lighting conditions and marker colors were collected and labelled.

To accelerate nuclei and glands labelling on images, a software was developed that allows a specialist to set a marker on an object of interest in the image. Subsequently, the coordinates of the centers of these markers (x, y), as well as the length (SizeY) and width (SizeX) were recorded in a csv file (see Table 3). Thus, a numerical data of the location and shape of the investigated structures were obtained. The markup was carried out by an employee from the laboratory by a cell biologist at the Ott Research Institute.

TABLE 2. COLLECTED IMAGES

Collected image class	Number of images
Confocal microscopy cell preparations	30
Confocal microscopy tissue preparations	92
Lightinh microscopy cell preparations	100

TABLE 3. A DATASET PART FOR THE CELLS

Image name	sizeX	sizeY	x	y
1(5).jpg	963	963	278	100
1(5).jpg	963	963	441	201
1(5).jpg	963	963	795	246
1(5).jpg	963	963	911	417
1(5).jpg	963	963	627	475

V. FORMULATION OF THE PROBLEM

The aim of the work is the research and development of the following algorithms:

- The cell nuclei number estimation with and without researched marker expression on light microscopy image preparations;
- The cell nuclei number estimation with and without researched marker expression on confocal microscopy image preparations;
- Highlighting the internal and external glands borders in the confocal microscopy image preparations.

For all algorithms, the following requirements are established:

- work without an operator;
- resistance to changes in the brightness of the preparation;
- resistance to various colors of markers;
- image scale is an input parameter of the algorithms.

VI. SUGGESTED ALGORITHMS

The algorithms were developed in the PyCharm environment in Python 3.7 using the OpenCV-python 4.0.0.21 library.

The source code of developed algorithms is available on github.com [10].

A. Counting the number of cell nuclei in confocal microscopy images.

- Enter the scaling parameter *expectedPixelsPer100Nm* - the number of pixels per 100 nanometers
- Read color image *I* in RGB format, depth 8 bits per channel;

- Bring the image *I* to a scale of 1.5 nanometers per pixel;
 - Convert *I* to HSV format, write the *V* component to the variable *V*;
 - Apply the contrast limited adaptive histogram equalization method with *clipLimit* = 2 and *tileGridSize* = 8 on image *V*;
 - Perform erosion on image *V* with an ellipse core of size 3;
 - Calculate mean as the average value of pixels *V*;
 - For each pixel V_{ij} : if $V_{ij} > mean + 20$, assign $V_{ij} = \min(V_{ij} + 100, 255)$ if $V_{ij} < mean - 20$, assign $V_{ij} = \max(V_{ij} - 100, 0)$
 - Apply a median filter with a core of size 5 to the image *V*;
 - Perform threshold binarization of image *V* with a threshold 127. Write the result to variable *B*;
 - Perform a contour search on image *B*, leaving contours that do not have nested paths;
- Calculate the centers of mass (C_x, C_y) for each contour using formulas (1):

$$C_x = \frac{m_{10}}{m_{00}}; C_y = m_{01}/m_{00}$$

$$m_{p,q} = \sum_{x,y=1..n} x^p y^q \quad (1)$$

Leave only those contours for which $|C_x - C_y| < 1$

- The number of contours received will be the total number of nuclei.
- ### B. Detection of the internal contours of the glands in confocal microscopy images
- Read color image *I* in RGB format, depth 8 bits per channel;
 - Bring the image to a scale of 1.5 nanometers per pixel;
 - Convert image *I* to HSV format, write the *V* component to the variable *V*;
 - Use the contrast limited adaptive histogram equalization method with parameters *clipLimit* = 2 and *tileGridSize* = 8 on image *V*;
 - Perform threshold binarization on *V* component of HSV with a threshold of 127;
 - Perform 27 erosion steps on the image *V* with the ellipse core of size 3;
 - Perform a contour detection on image *B*, leaving paths that do not have nested paths. Write the result to the contours variable;
 - For each contour: calculate the area, count the number of pixels V_{ij} falling into this contour, taking into account the V_{ij} exceeding 15, and write to the variable *nonZeroPixelsArea*. Suppose that the contour is the inner border of the gland if the nonzero pixels of the region exceed the product of the contour of the region 0.4.

- Recognized glands boundaries will be in the contours list on the output image

C. Counting the number of cell nuclei in light microscopy images

- Enter image I in RGB format and bring it to a scale of 1.5 nanometers per pixel (similar to algorithms A and B);
- Convert I to HSV format, write the V component to the variable V (similar to algorithms A and B);
- Equalize the histogram of the V component in intensity. To do this, obtain the hist distribution vector of colors (0..255); accumulate the histogram values in $chist$, where $chist[n] = \sum_{i=0..n} hist[i] (hist[i])$. Calculate the V_{max} as: $V_{max} = chist[len(chist) - 1] * \frac{(100 - thresh)}{100}$, where $thresh$ is a threshold value equal to 1. Replace V pixels with a value greater than V_{max} with a value of V_{max} ;
- Perform Laplace transforms on image V with sigma equal to 3, 6 and 9, write the results in L_1, L_2, L_3 , respectively. Calculate the sum of the images L_1, L_2, L_3 by pixels and write to $laplaceSum$. Normalize L_{sum} in the range from 0 to 255. Perform Otsu binarization on L_{sum} . Write the result to variable B ;
- Search for contours in image B , leaving contours that do not have nested paths (similar to algorithm A). The result is written to the contours variable.
- Calculate the moments of the contour C_x, C_y according to formulas (1). Leave only those contours for which

$$|C_x - C_y| < 1;$$

The task of scaling the image to 1.5 nanometers per pixel was solved by manual measuring the length of the scale bar on the image in pixels. However, we plan to develop automated recognition of the scale bar's length on images

The coefficients of the algorithms were selected by optimizing the criteria by the simplex method and the gradient descent method when testing on the analyzed preparations. The criteria are given in the section "Results"

VII. RESULTS

A. The algorithm for counting the number of cell nuclei in confocal microscopy images

The percent of successful recognition in the researched images was 84%. An example of image recognition is presented in Fig. 1s

For the task of counting nuclei, the percentage of successful recognition was presented as the average value for the analyzed images. For each image I , there were the number of detected nuclei N_d and the number of labeled nuclei N_l . The percent of recognition π for image I was calculated by the formula (2):

$$P = \frac{\min(N_d, N_l)}{\max(N_d, N_l)} * 100\% \quad (2)$$

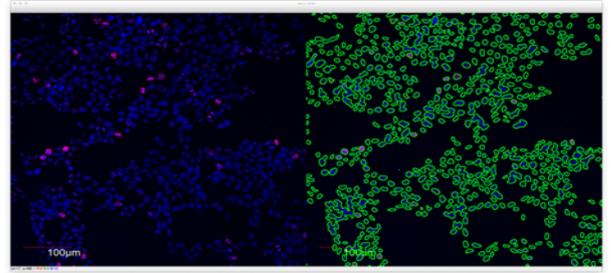


Fig. 1. Highlighting of nuclei on the preparations of confocal microscopy. Initial image (left), image with selected nuclei (right)

B. Highlighting of the internal contours of the glands on confocal microscopy images.

The percentage of successful recognition in the researched images was 70%. An example of image recognition is shown in Fig. 2.

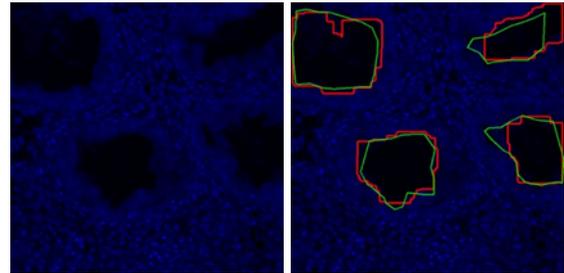


Fig. 2. Highlighting of the internal borders of the glands using confocal microscopy preparations. Initial image (left), image with highlighted glands (right)

For the problem of counting glands, the percentage of successful recognition was given as the average value for the analyzed images. For each image I , there were internal contours of the glands detected by the algorithm - $detectedContours$ and internal contours of the glands marked by the pathomorphologist - $labelledContours$. The intersection area of the marked and detected contours A_d and the total area of the marked out contours - A_l were calculated. The percent π of recognition for image I was calculated by the formula (3):

$$P = \left| \frac{A_d}{A_l} - 1 \right| * 100\% \quad (3)$$

C. The estimation of the number of cell nuclei in light microscopy images

The percentage of successful recognition was 90% in the researched images. An example of image recognition is presented in Fig. 3.

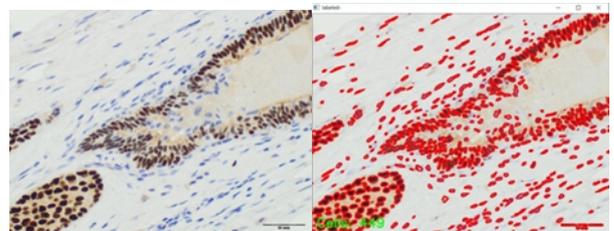


Fig. 3 Highlighting of nuclei on light microscopy preparations. Initial image (left), image with selected nuclei (right)

For the task of counting nuclei on light microscopy preparations, the percentage of successful recognition was analyzed by the formula (2) similarly to the percentage of recognition for confocal microscopy.

VIII. COMPARISON WITH EXISTED SOFTWARE

The purpose of last stage of this work was to compare the obtained data with the results of other applications used to evaluate microphotographs.

Since most of the software intended for cytological studies are either expensive or require a long study of the manual and programming languages, it was decided to compare the performance of the created software with Fiji - the ImageJ plugin for evaluating microscopy images, which is the most balanced among its plugins. The initial task was to estimate the number of cells in 30 microphotographs in the jpg format. The comparison results between proposed algorithm and Fiji for confocal imagery task is presented in Table 4.

TABLE 4. COMPARISON OF CREATED ALGORITHM AND FIJI FOR CONFOCAL IMAGERY

No	Labelled	Fiji	% Fiji	Created algorithm	% of created algorithm
1(5)	11	15	73,3	14	78,5
1(6)	11	17	64,7	18	61,1
1(7)	95	109	87,2	94	98,9
1(9)	12	25	48,0	21	57,1
1(13)	360	424	84,9	195	54,2
1(30)	145	261	55,6	165	87,8
1(31)	128	161	79,5	128	100
1(38)	45	59	76,3	57	78,9
1(41)	189	192	98,4	182	96,2
1(42)	123	121	98,4	117	95,1
...					
Result of Fiji				76% +/- 16.1%	
Result of created algorithm				83,8% +/- 11.8%	

For the task of counting nuclei on light microscopy preparations, the percentage of successful recognition was analyzed by the formula (2) similarly to the percentage of recognition for confocal microscopy preparations.

Thus, the developed algorithm exceeds the Fiji accuracy by 7.8%. It should be noted that in order to achieve maximum accuracy in Fiji, the threshold parameter is required and the estimated radius of the object of interest should be introduced as the lower limit, whereas in the written code the determination of the size of objects takes place automatically, which excludes the element of subjectivity from the study and the necessity for preliminary processing of photographs. For example, to evaluate the image, the created algorithm does not require preliminary removal of the scale bar from the image. Also, for the various color markers expression in the nuclei classification task using "Fiji", it is necessary to set the "Color Threshold" value for each type of marker in each photo separately, which not only significantly increases the time of analyze carried out by the pathomorphologist, but also greatly reduces the quality of this analysis, since nuclei having weak expression are most likely will not be included in the corresponding group. Unlike Fiji, the created algorithm

equally effectively copes with the task of counting the total number of cores as well as with the task of classifying them.

IX. FURTHER RESEARCH

In the future, it is planned to continue research in this area, improving the reliability of the algorithms, particularly:

- Use CNN and U-net networks for better segmentation of core images [8, 9].
- Detect the contours of the scale bar for automatic scaling of the drug.
- Automatically recognize marker colors (support more than two colors)

REFERENCES

- [1] Kuznets S.M., Panteleev V.G. *The application the Hardware - Software Complex "VideoTesT - Morphology" for the differentiation of tumor cells / Clinic. N.1. 2011, pp. 122-123.*
- [2] Berezsky O.N., Melnik G.N. *Information technology for the analysis and synthesis of histological images in automated microscopy systems / Control systems and machines. 2013. N.4. S. 26-32. - Bibliogr.: 27*
- [3] Carrarelli P., Rocha A. *Increased expression of antimullerian hormone and its receptor in endometriosis // Fertility and sterility. 2014; 101 (5) P. 1353-8.*
- [4] Drobintseva A.O., Polyakova V.O., Masing D.S., Matyushkin LB *Confocal microscopy. Role and importance in the study of the reproductive system. Tutorial. FGBNU, 2015.18 s.*
- [5] Chen, S., Zhao, M., Wu, G., Yao, C., Zhang, J. *Recent Advances in Morphological Cell Image Analysis Computational and Mathematical Methods in Medicine Volume 2012 (2012), Article ID 101536, 10 pages.*
- [6] Gurcan M. N., Boucheron L., Can A., Madabhushi A., Rajpoot N., & Yener B. (2009). *Histopathological Image Analysis: A Review. IEEE Reviews in Biomedical Engineering, 2, 147–171.* <http://doi.org/10.1109/RBME.2009.2034865>.
- [7] A. Kovrigin (2007), "Application of the principles of constructing machine vision systems in the task of analyzing images of cellular structures". *Scientific journal KubSAU, No29 (5), May 2007.*
- [8] Xue, Y., & Ray, N. (2017). *Cell Detection with Deep Convolutional Neural Network and Compressed Sensing. CoRR, abs/1708.03307.*
- [9] Falk, T., Mai, D., Bensch, R. et al. *U-Net: deep learning for cell counting, detection, and morphometry. Nat Methods 16, 67–70 (2019).* <https://doi.org/10.1038/s41592-018-0261-2>
- [10] Sergreev D.I., Drobinceva A.O., ets (2020). *Biomedicine diagnostic repository. Source code of this paper. Available at <https://github.com/densvr/biomedicine-diagnostic>*
- [11] *ImageJ software homepage (2020). Available at <https://imagej.net>*
- [12] *CellProfiler software homepage (2020). Available at <https://cellprofiler.org>*
- [13] *OrbitImageAnalysis software homepage (2020). Available at <https://www.orbit.bio>*
- [14] *AxioVision 4.8. software homepage (2020). Available at <https://www.micro-shop.zeiss.com/en/us/system/software-axiovision+software-products/1007/>*
- [15] *VideoTestMorphology software homepage (2020). Available at http://www.digitalimagingystems.co.uk/pdfs/morpho_en.pdf*
- [16] *CellSens software homepage (2020). Available at <https://www.olympus-lifescience.com/en/software/cellsens/>*
- [17] *IMAGE-PRO-Premium software homepage (2020). Available at <https://www.mediacy.com/imagepro>*
- [18] *BioVision software homepage (2020). Available at <https://www.biovision.com>*
- [19] *Fiji plugin home page (2020). Available at <https://imagej.net/Fiji>*