

AUTOMATIC CELL RECOGNITION IN IMMUNOHISTOCHEMICAL GASTRITIS STAINS USING SEQUENTIAL THRESHOLDING AND SVM NETWORK

Tomasz Markiewicz^{1,2}, Cezary Jochymski¹, Robert Koktysz¹, Wojciech Kozlowski¹

¹Department of Pathology Military Institute of the Health Services, Poland

²Warsaw University of Technology, Poland

ABSTRACT

The paper presents program for automatic cell recognition and counting in selected immunohistochemical stains in the gastritis diseases. It is applied to cytoplasm reactivity markers, such as chromogranin A, serotonin and somatostatin antibodies. The program uses the sequential thresholding algorithm in combination with artificial neural network of Support Vector Machine (SVM) type, to recognize the nuclei of the separated cells. The constructed algorithm imitates the human view of the image. The Support Vector Machine is used for recognition of the immunoreactivity of the separated cell. The results corresponding to the exemplary images, confirm good accuracy, comparable to the human expert.

Index Terms— immunohistochemistry, cell recognition, cell extraction, Support Vector Machine, neuroendocrine cells

1. INTRODUCTION

The chronic gastritis belongs to the frequently gastric diseases. An experienced researcher assesses the relevance of the distribution and density of the neuroendocrine cells. The reported results confirm the dependence of the density of the neuroendocrine cells such as D, EC or ECL on the type of stomach disease [1,2]. However there are significant differences in the used methods. Some researchers count the neuroendocrine cells in few fields of view, with magnification 200x and results are calculated as the mean of cell quantity per one field [3]. Other calculate a number of cells per mm² corresponding to the area of the mucosa glandular epithelium or per one glandular [4,5,6]. Generally these methods require the recognition of immunoreactivity of the cells in the view field and counting them. In the last years many computer programs for automation of this process, based on the image processing, have been developed [7,8,9,10]. They used mainly the mathematical morphology operations, such as erosion, dilation, opening, closing, watershed method etc. [11,12]. However the algorithms used in these programs must be specialized for the specific stains and features of the recognized cells.

In this paper the authors propose a new concept to solve the problem of recognition of the cells in the image. It is similar to the human analysis of the viewed structures. In the considered images this is the easier applicable form, thanks to the stable area of the recognized cells. The described algorithm imitates the extraction of the separate cells (or their nuclei) from the background, based on the lighter border. It uses the sequential thresholding operation combined with checking the area of the fully separated objects in the image. We used this idea with satisfied results for other types of tissues and stains, such as chromatine stains in the brain tissues with neuroblastoma tumor or in the breast tumor tissues [13]. However we still use (as a input to this algorithm) the gray scale images representing the difference between immunonegative blue colored cells and the other cells or background, created by the arithmetic operation on the RGB color components. However this approach is applicable only for these objects, for which one dominating color should be extracted.

Better results may be obtained if we rely the recognition on the real color of the cells and not on the difference of their RGB components. Such opportunity offers the artificial neural network of SVM type [14,15]. The output of this network working in the classification mode can be understood as an indicator of reliability of the recognized type (color) of the cells. After the extraction of the whole cells with blue colored nuclei, the program recognizes them as the member of two groups: immunopositive and immunonegative, based on the color of the cytoplasm. This recognition is done by the SVM and morphological operations.

The presented method was applied to three types of the immunohistochemistry cytoplasm stains in the gastric disease: chromogranin-A, serotonin and somatostatin antibody. For this stains the following markers have been applied:

- Monoclonal Mouse Anti-Human Chromogranin A, clone DAK-A3, code M 0869 DAKO
- Monoclonal Mouse Anti-Serotonin, clone 5HT-H209, code M 0758 DAKO
- Rabbit Polyclonal Anti-Human Somatostatin, code A0566 DAKO

We used statistical density of the counted cells in the field of view with magnification 400x as an evaluation method. The results of the numerical experiments performed for ganglion diseases show good agreement with the human experts results.

2. ALGORITHM OF CELL EXTRACTION

The main task in the appropriate evaluation of the distribution density of the selected stained endocrine cells is a correct recognition of the separated nuclei. This problem can be solved in two steps: extraction of the nuclei of the cells and classification of them into two groups – immunopositive and immunonegative cells. The used stains (chromogranin A, serotonin and somatostatin) are types of the cytoplasm indicators. All nuclei are blue and only cytoplasm in the immunopositive cells is brown. However usually in the field of view some cells are fully brown colored. This is the effect of the fact that cutting plane of the tissue specimen may go through any possible cutting levels of the cell. Moreover some cells in the slide are only viewed as a part of the cytoplasm. Additionally the nuclei are sometimes covered by the cytoplasm. According to this fact the segmentation algorithm should possess two extraction lines: one for the cells with recognized blue nucleus and second for recognition of the immunopositive cells without the blue nucleus.

The input image for segmentation and counting program is in the form of RGB standard file, acquired from the microscope with the help of camera. The first step in our segmentation algorithm is the image standardization. It means the calculation of the average RGB values of the lighter area in the field of view. Then the RGB values of all pixels in the image are linearly transformed in such way that the lighter area will be exactly white. This step eliminates the influence of the differences in the glass transparency, lighting and other unstable values.

The next step is extraction of the blue nuclei as the separated cells. To this task we used the threshold operation, which is defined as

$$T_{[0,t_2]}[f(x)] = \begin{cases} 1 & \text{if } 0 \leq f(x) \leq t_2 \\ 0 & \text{else} \end{cases} \quad (1)$$

where $f(x)$ is the value of pixel of the image f in the position x and t_2 is the threshold value. Generally this is a very difficult task because the nuclei are stained at different intensity and selecting one threshold value for this operation is problematic, and for some images even not possible.

We propose a new concept to solve this problem. We use the thresholding operation sequentially and apply the artificial neural network of SVM type [14,15]. The idea of this network is to create a hyperplane dividing the feature space of the input data into two separated parts with the maximum margin between them. In our case the input data are the pixels selected from the image for learning phase in

such a way, that they are representative for three classes: blue nuclei, brown cytoplasm and light background. The input vector is composed of three color components in RGB standard. Because one network recognizes only two classes, we must build three SVMs for recognizing pixels between all pairs of classes and then use the one-against-one strategy to find the winner [15]. The learning data are manually selected from the sampled images and their quantity was 150 pixels per class. The output of the network is determined using the formula

$$D(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (2)$$

where \mathbf{w} is the weight vector, \mathbf{x} is the input vector and b is the bias. For the learning process $D(\mathbf{x})$ is 1 for the first class and -1 for the alternative class in any pair of the classes. The learning of the network is understood as a task to find the optimal \mathbf{w} and b values. It is transformed to the quadratic programming with the help to Lagrange multipliers α_k [14,15]. As a result of training we get

$$\mathbf{w} = \sum_k \alpha_k d_k \mathbf{x}_k \quad (3)$$

and

$$b = d_k - \mathbf{w}^T \mathbf{x}_k \quad (4)$$

where d_k is 1 or -1 and α_k is the nonzero Lagrange multiplier corresponding to the k th training data. Usually most multipliers are zero and these data points don't influence for the solution.

After learning of the network the weights and bias are fixed. In the testing phase the input vectors represent the color values of all pixels in the image. They are put to the SVM. It is evident that the output value $D(\mathbf{x})$ will be different, depending on its relation to the respected class. For example, if the pixel is light blue, the $D(\mathbf{x})$ signal of the SVM recognizing the nuclei will take values $0 < D(\mathbf{x}) < 1$. For the dark blue pixels this value will be higher than 1. Based on this relevance we use $D(\mathbf{x})$ value as an indicator of the recognized class. In the constructed algorithm we use sequential thresholding operation starting from the minimum $D(\mathbf{x})$ value as a threshold. Then this threshold is increased step by step until the maximum. In any step $D(\mathbf{x})$ for the whole image is thresholded and the separated objects, whose area are in the selected range, are added to the mask of the recognized nuclei. The result of this process will be the mask of all blue nuclei for which the area is in the preselected range. This range is selected on the basis of some knowledge and the image resolution with some margin. This process is supported by the watershed operation for the bigger cells to divide them into two cells if there is a narrow space between the two or more parts of them. This will help in the case of the overlapping nuclei of the cells.

3. RECOGNITION OF THE IMMUNOREACTIVITY OF CELLS

The next step is the recognition of the immunopositive cells. This is performed with the use of the SVM and mathematical morphology operations such as closing and reconstruction. First the brown cytoplasm is extracted by using the one-against-one strategy with three SVMs. Next all brown areas are closed by the structural element with the disk shape of the diameter of 7 pixels long. This operation outputs the mask of the brown cytoplasm with filled internal area of them. This filled area should be only on the internal side of the cytoplasm that is on the nuclei of the cell. If the immunopositive cell is touched with the other immunonegative cell, the second nucleus shouldn't be selected. In practice we can tolerance some errors in the case of overlapping.

The last process of the segmentation is the extraction of still unrecognized immunopositive cells without the distinct blue nuclei. They are created from the brown mask, received in the previous step of the algorithm. Any separated brown objects that didn't possess the blue nucleus and lie in the selected range of the cell area are added to the set of the immunopositive cells after a watershed operation dividing touched items.

The final task is to count the cells according to their immunoreactivity classes. This is done in the form of simple counting any separated objects in the masks of immunopositive and immunonegative cells, independently.

4. RESULTS

According to the statistical method for the ratio of immunopositive cell, the chromogranin A, serotonin and somatostatin antibodies stain specimens should be evaluated. We have selected some patients suffering for stomach diseases from the Military Hospital in Warsaw in Poland from the years 1999 – 2007. The requirement experiments have been done using the acquired images of the magnification 400x, at resolution of 768x576 pixels. We used the Olympus AX70 microscope with the Olympus Camedia C-3030 camera.

The slides used in further quantitative analysis should fulfill the following histological criteria: proper sections (automatic fixed paraffin sections, slide thickness from 3 to 5 micrometers), the right technical quality of tissue samples and the correct histo- and immunohistochemical staining, lack of any mechanical injury or any artifact (fragmentation, hemorrhages, etc).

The main problem is the correct staining of the slides in the somatostatin antibody, because in some items of our database there were brown colored nuclei or/and mucous membrane. Such patients have to be eliminated. Generally for somatostatin evaluating we used samples of higher level of brown color used in learning of the SVM classifier for the

immunopositive cell recognition. The pixels of the light brown color were treated as a background. In chromogranin A and Serotonin the most specimens were stained correct.

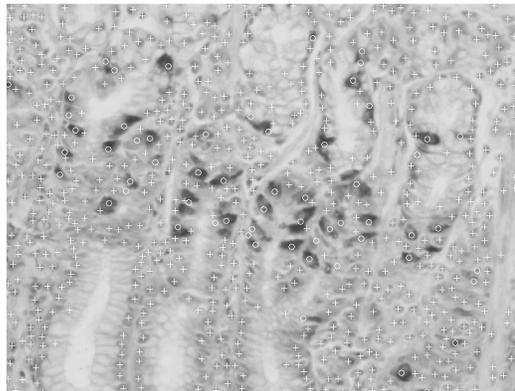


Figure 1. The sample result of the recognition of the cells in chromogranin-A stain



Figure 2. The sample result of the recognition of the cells in serotonin stain

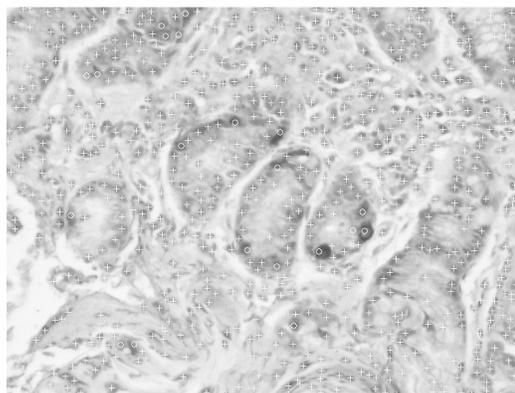


Figure 3. The sample result of the recognition of the cells in somatostatin stain

The evaluated specimens were taken from patients of the following diagnosis:

- gastritis chronica
- gastritis chronica superficialis
- dysplasia regenerativa
- metaplasia intestinalis
- hyperplasia foveolaris
- limphadenoplasia
- gastritis refluxiva

Additionally the presence of *Helicobacter pylori*, the age of patient and the sex have been taken into account.

The specimens were divided relative to the oligobiopsy anatomical region – prepyloric or oxyntic part of stomach. The images of them have been acquired in the manually selected regions, with the higher number of immunopositive cells in the center, mainly three per slide. Because significantly more these cells are in the chromogranin A antibody stain, the regions were selected on the basis of this stain. In the serotonin and somatostatin stains we were get images from the too same regions. Figure 1 presents a sample result of recognition in the chromogranin stain. The immunopositive cells are marked “o”, the immunonegative “+”, both in white color. The results for the serotonin and somatostatin stains of the same patient specimens are presented in figure 2 and 3 respectively. We have evaluated more than 30 patients and the achieved accuracy was on the acceptable level. Less than 5% of the cells were unrecognized or misclassified.

The developed program was written in Matlab language and tested of PC Certino Duo 1.86 MHz, 2GB RAM. The result for one image is received in less than 1 minute. It is possible to correct manually the classification results of the cell by the intervention of a human expert.

5. CONCLUSION

The presented algorithm combines the mathematical morphology operations with artificial neural network of SVM type. It can be helpful for researches in checking the endocrine cells in gastric disease and can be easily adapted to the other cytoplasm immunohistochemistry stains. It imitates the human view strategy in recognition of the separated nuclei of the cells. The received accuracy is on good level and fully repeatable. Because some researchers use the counting of cells by the one ganglion we plan to continue our work for the applied ganglion recognition with the cells created them.

6. ACKNOWLEDGMENT

This work is supported by the grant of Polish Ministry of Science and Higher Education for the years 2006-2009.

7. REFERENCES

- [1] C. Bordi, T. D'Adda, C. Azzoni, and G. Ferraro, “Classification of gastric endocrine cells at light and electron microscopical levels”, *Microsc Res Tech*, pp. 48, 5:2000.
- [2] S. Falkmer, and E. Wilander, “The endocrine cell population”, in *Gastrointestinal and oesophageal pathology*, Longman Group Ltd., pp. 63-71, 1995.
- [3] D.M. Green, A.E. Bishop, G. Rindi, F.I. Lee, M.J. Daly, J. Domin, S.R. Bloom, and J.M. Polak, “Enterochromaffin-like cell populations in human fundic mucosa: quantitative studies of their variations with age, sex, and plasma gastrin levels”, *J Pathol*, pp. 235-241, 157:1989.
- [4] L.G. Aruin, V.A. Vinogradov, J.V. Zverkov, and P.A. Nikiforov, “The endocrine stomach cells containing endocrine gastrin and somatostatin in patients with duodenal ulcer”, *Arkhiv. Patologii*, pp. 15-20, XLVI(6):1984.
- [5] M. Tzaneva, and A. Julianov, “Chromogranin A-, somatostatin- and serotonin-containing endocrine cells in the corporal gastric mucosa of patients with *Helicobacter pylori* associated gastritis”, *Endocr Regul*, pp. 79-82, 33(2): 1999.
- [6] C. Azzoni, C. Doglioni, and G. Viale, et al, “Involvement of BCL-2 Oncoprotein in the Development of Enterochromaffin-like Cell Gastric Carcinoids”, *Am J Surg Pathol*, pp. 433-41, 20:1996.
- [7] P.H. Bartels, D. Thompson, R. Montironi, G. Mariuzzi, and P.W. Hamilton, “Automated reasoning system in histopathologic diagnosis and prognosis of prostate cancer and its precursors”, *European Urology*, pp. 222-233, 30:1996.
- [8] O. Lezoray, A. Elmoataz, H. Cardot, G. Gougeon, M. Lecluse, H. Elie, and M. Revenu, “Segmentation of color images from serous cytology for automated cell classification”, *Analytical and Quantitative Cytology and Histology*, pp. 311-322, 22:2000.
- [9] T. Markiewicz, S. Osowski, J. Patera, and W. Kozłowski, “Image processing for accurate recognition and counting of cells of the histological slides”, *Analytical and Quantitative Cytology and Histology*, pp. 281-292, 28:2006.
- [10] G. Kayser, D. Radziszowski, P. Bzdyl, R. Sommer, and K. Kayser, “Theory and implementation of an electronic, automated measurement system for images obtained from immunohistochemically stained slides”, *Analytical and Quantitative Cytology and Histology*, pp. 27-38, 28:2006.
- [11] P. Soille, *Morphological image analysis, principles and application*, Springer, Berlin, 2003.
- [12] *Matlab Image Processing Toolbox, user's guide*, MathWorks, Natick, 2002.
- [13] T. Markiewicz, P. Wisniewski, S. Osowski, J. Patera, W. Kozłowski, and R. Koktysz, “Comparative analysis of the methods for accurate recognition of cells in the nuclei staining of the Ki-67 in neuroblastoma and ER/PR status staining in breast cancer”, *Analytical and Quantitative Cytology and Histology*, 2008 (in press).
- [14] V.Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.
- [15] B. Schölkopf, and A. Smola, *Learning with Kernels*, MIT Press, Cambridge, 2002.