GENETIC ALGORITHM FOR PARAMETER OPTIMIZATION OF IMAGE SEGMENTATION ALGORITHM

Sándor Szénási
Óbuda University, John von Neumann Faculty of Informatics, Budapest, Hungary
szenasi.sandor@nik.uni-obuda.hu

***Abstract*— In the current practice of medicine, histopathological examinations are some of the most important tools for clinical diagnoses of a large group of diseases. To help pathologists and to reduce the subjectivity level, it has been proposed that computer-aided procedures be used to provide objective results. The first step of these procedures is the segmentation of the tissue image. In our research, we try to detect nuclei, glands and surface epithelium in Haematoxylin and Eosin (HE) stained colon tissue samples. This paper focuses on the identification of epithelial cell nuclei.**

Keywords—GPGPU,medical image segmentation, genetic algorithm, colon tissue, epithelium detection

I. EPITHELIUM DETECTION

The processing of microscopic tissue images and especially the detection of cell nuclei is nowadays done more and more using digital imagery and special immunodiagnostic software products [1][2] and hardware improvements [3][4]. These are fast and accurate products and can serve several additional functions, like remote access, archiving, searching, tagging [5] etc.

Epithelial tissues line the cavities and surfaces of structures throughout the body. They also form many glands. In HE stained colon tissue samples (Fig. 1), epithelial cells appear around the glands and at the edge of the whole sample (surface epithelium). A number of published papers can be found dealing with the nuclei searching or gland detection. However, only a few deal with epithelial cell identification.

For gland detection, one of the most promising procedures is the object-graph method [6], "which decomposes the tissue image into a set of primitive objects and segments glands making use of the organizational properties of these objects, which are quantified with the definition of object-graphs". This approach employs the object based information for the gland segmentation problem, instead of using the pixel-based information.

There are some other methods, like the technique developed by Wu and Gil. They introduce [7] "a biased median filtering image segmentation algorithm for intestinal cell glands consisting of goblet cells". Four biased median filters with long rectangular windows of identical dimensions, but different orientations, are designed based on the shapes and distributions of cells. These identify a part of gland segments in a particular direction. Finally, the complete gland regions are the combined responses of the already mentioned filters.

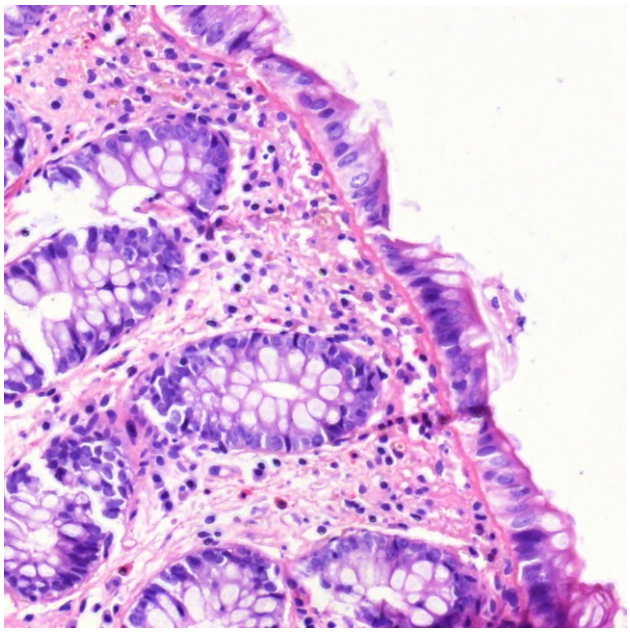

Fig. 1. HE stained colon tissue image

In most cases, we have to know the accurate position of cell nuclei. In several papers the authors usually use some threshold based method for this segmentation, but (based on our previous study) with region growing technique better accuracy was achieved. Although it is a fact that this method is an order of magnitude slower, but we can use GPGPUs to accelerate the process [8].

II. DETECTION ALGORITHM

After the cell nuclei detection, we have to determine that an appropriate nucleus belongs to an epithelial cell or not. The method developed by us is based on the idea that that the epithelial cells have some particular attributes compared to the other cells:

- Density of the epithelial cell nuclei groups differs from the average density of the other cells of the tissue sample:
    - High density in one direction.
    - Relatively small density orthogonally.
    - Average density in the other orthogonal direction.
- Epithelial cells usually form a chain.

Our algorithm tries to use both features in the detection process. Based on the first three features, it performs a preliminary test that tries to decide whether a given cell looks like epithelial. This is followed by a second examination (based on the fourth feature), in which the previously selected nuclei have been investigated on the basis that those are able to produce chains or not. If so, they are considered permanently epithelium, otherwise, they are rejected. For the best results, the initial test should be more lenient (to shift the results more to the false positive hits than to the false negatives), because the second step will further filter these results.

III. Parameters of the Density Test

The density scan searches through all of the nuclei individually. It examines the pixels of the image in all directions from the centre of the cell (like separate "beams" from the centre like in Fig. 2). If these pixels belong to a different nucleus, then the algorithm increases a counter that will finally be divided with the length of the beam. These lead to density values in all directions (and the algorithm stores the distance of the nearest pixels too). Based on the calculated data from all directions, the algorithm provides an estimation on whether the given cell is epithelial.

During the tests, we need the following parameters in order to get better accuracy:

- SCANN: Indicates the resolution of the scanning process. It actually represents the number of beams started from the centre of the cell. Therefore, the angular distance between these beams is 1 / SCANN * 2 * PI Rad. Choosing a larger number provides an increased resolution and better accuracy; however, it increases the processing time as well—something that we want to avoid. Our goal is to find the optimal value where the accuracy is maximal with minimal processing time.
- DENSITY_LENGTH: The length of the beam used in the density check. The algorithm only checks pixels that are not farther than this distance from the centre of the nucleus.
- FRONT_SCAN_WIDTH: It determines that how wide of an area the algorithm can examine in the forward direction. Practically, it represents the number of beams (symmetrically distributed between the two directions). It is critical because if too high of a value is chosen, the algorithm will calculate only the average density of the environment and it will not find the cells where there is a real densification in one direction. Too low of a value can cause other problems since nuclei forming a chain are not necessary form a straight line (even nuclei of glands clearly form an ellipse). If the algorithm uses only one beam it will not find the required density in the case of curves.
- SIDE_SCAN_WIDTH: Similar to the previous parameter; however, this is responsible not for the forward direction but for the lateral direction. It represents the number of beams (distributed symmetrically). Therefore the angular diameter is SIDE_SCAN_WIDTH / SCANN * 2 * PI Rad. Too high of a value may lead to checking too big of an area. In this case, the algorithm will not find the lower density places. Too small of a value may cause an increase in the rate of false positive results, because the usage of only one beam has a decent chance that it will not find any neighbouring cells in an area with average density.
- Weighting factors: For each nucleus, we perform a calculation based on the above parameters. The program scans the required beams and calculates the density and the distance of the nearest point into their direction. Based on these results it evaluates a fitness function that is a weighted sum of these values. This requires several weighting factors, these are all additional parameters.

IV. Parameters of Chain Search

In the chain search procedure, the program tries to build chains from the previously selected nuclei (qualified as epithelial cells). It performs a recursive depth search, which is designed to append as many epithelial cells into one chain as possible. The main parameters for this procedure are:

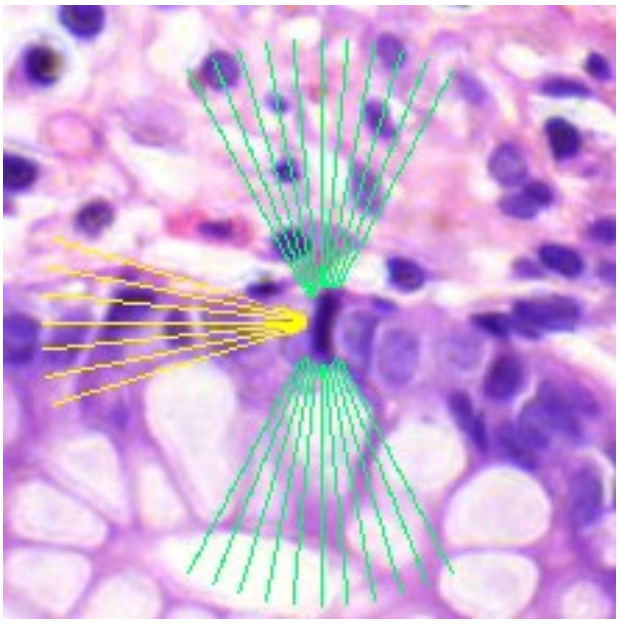

Fig. 2. Beams of density scan. Yellow lines: beams of front scan. Green lines: beams of lateral scan. Main scanning direction is left to right horizontally.

- CHAIN_CANDIDATE_LIMIT: The above mentioned density test determines only a score value (probability that the given cell is epithelial or not). This parameter defines the score limit above which the cells are likely to be epithelial (another approach can be a fuzzy model [9]). Only these cells will take part in the actual chain search procedure. Choosing a value that is too low can be problematic because it causes too many nuclei in the next phase, which greatly increases the running time and the number of false positive hits. On the contrary, if the value is too high, we reject many epithelial cells.
- DIRECTION_CHANGE_LIMIT: During the chain search, the algorithm tries to create long series of cells in line. The search is much more complicated than a simple linear search, because it is probable that the epithelial nuclei are not in a straight line, but that they form curves. The above parameter determines the maximal degree of direction change that we permit during the search process. Too low of a value is problematic because the program will not identify coherent items as a chain (for example, in the case of glands where the cells form ellipses). Too high of a value is also not good because it increases the proportion of false positive results.
- MIN_CHAIN_LENGTH: This determines the minimum length of the chain candidate that we accept. Not worth to choose too low value, because it is very easy to find randomly 2-3 cell length chain parts. Too high value is also not good, because it causes the rejection of several otherwise proper chain pieces.
- MAX_CHAIN_LENGTH: Basically, the limit for the maximum chain length is not necessary, because in case of big images (with high cell density) there may be really long valid chains. However, in practice, it is advisable to define a limit for the search that is meant not to do too many recursive steps, which can significantly increase the running time without any particular result.
- MIN_CHAIN_SCORE: After the chain searching, we can score all of the candidate chain parts. This is the final limit where we can decide that a nuclei chain candidate is acceptable or not. Obviously, our goal is to find a middle ground where the number of false positive and false negative hits is the most ideal.

V. GENETIC ALGORITHM

The fairly large number of parameters makes the manual optimization almost impossible. For this reason, we have developed a genetic algorithm and a distributed system [10] that helps in finding the optimal environment parameters.
The parameters of the genetic algorithm are as follows:

- All chromosomes contain 10 genes. These correspond to the above-described parameters. Each generation contains 200 pieces of individuals, except for the first generation, which was launched with 1000 instances.
- According to the elitism, the top 20 of all chromosomes are automatically transferred into the next generation.
- We have started the genetic algorithm in a distributed system and we have evaluated 70 generations.

The genetic algorithm runs the epithelium segmentation procedure for each chromosome (we did not have to run the cell nuclei detection each time because the region growing parameters did not change). Then the algorithm examined the accuracy of the results. In order to test the accuracy of a given parameter set, we create a manually annotated sample marking all cells in this tissue image on the basis that it is epithelial or

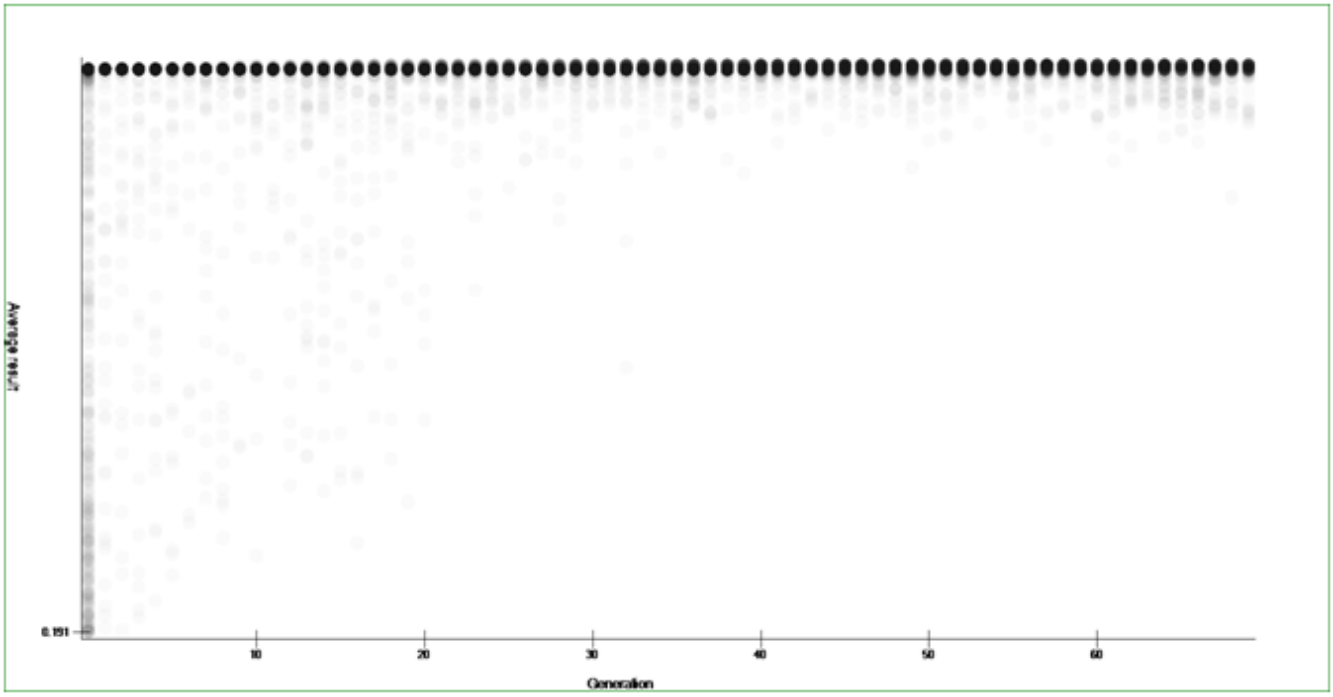

Fig. 3. Accuracy of all chromosomes in the given generations.

not. The accuracy of a given parameter set is simply calculated as [11]:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Where

- TP : Number of True-Positive hits
- TN : Number of True-Negative hits
- FP : Number of False-Positive hits
- FN : Number of False-Negative hits

The results of the first run are in the following diagram. (Fig. 3). The figure shows that there were some good results in the first random sample (first generation) and that after 15 generations it finds the final result. The implemented genetic algorithm has some fairly good parameter sets within a few generations. Because more than one chromosome represents equal accuracy, it is worth it to select the one that needs less computing capacity.

VI. RESULTS

With the obtained parameter set, we can determine more than 80% certainty that a given nucleus is epithelial or not. We consider this a success. When examining the results, it should be taken into account that the nuclei segmentation are not considered accomplished [12] (both the false positive and the false negative results of the nuclei detection can badly effect the results of the further processing) .

In the future, we might restart the genetic algorithm with other parameters and first generation; maybe we can achieve even better results, though we do not expect significantly better results. It would be better to speed up the epithelial search process (for example, with a GPGPU based implementation) and to use some advanced processing method [13]. Another promising way of improvements might be the investigation of some alternative methods [14][15][16][17].

ACKNOWLEDGEMENT

This work makes use of results produced by the Hungarian National Technology Programme, A1, Life sciences, the "Development of integrated virtual microscopy technologies and reagents for diagnosing, therapeutical prediction and preventive screening of colon cancer "Hungarian National Technology Programme, A1, Life sciences, (3dhist08) project and the ÓE-RH 1104/2-2011 project.

REFERENCES

[1] B. Kiss, J. Sápi, L. Kovács, "Imaging method for model-based control of tumor diseases," IEEE 11th International Symposium on Intelligent Systems and Informatics, 26-38 Sept. 2013, pp. 271-274

[2] S. Sergyán, "Content-based image retrieval using automatically determined color regions of images," 7th International Symposium on Applied Machine Intelligence and Informatics, Herlany, 30-31 Jan. 2009, pp. 41-45

[3] J. Kopják, J. Kovács, "EEE 10th Jubilee International Symposium on Applied Machine Intelligence and Informatics," Herlany, 26-28 Jan. 2012, pp. 377-382

[4] P. Kádár, M. Karacsi, „Stand alone island mode operation of microCHP device," 5th International Symposium on Exploitation of Renewable Energy Sources, Subotica, 21-23 March 2013, pp. 1-5

[5] Á. Bogárdi-Mészöly, A. Rövid, H. Ishikawa, S. Yokoyama, Z. Vámossy, "Tag and Topic Recommendation Systems", Acta Polytechnica Hungarica, vol. 10, issue 6, 2013, pp. 171-191

[6] C. Gunduz-Demir, M. Kandemir, A. Burak Tosun, C. Sokmensuer, "Automatic segmentation of colon glands using object-graphs," Medical Image Analysis, vol. 14, issue 1, Feb. 2010, pp. 1-12

[7] H.-S. Wu, J. Gil, "A biased median filtering algorithm for segmentation of intestinal cell gland images," The Scientific World Journal, vol. 6, pp. 200–220, Feb. 2006.

[8] Szénási, S.; Vámossy, Z.; Kozlovszky, M., "GPGPU-based data parallel region growing algorithm for cell nuclei detection," IEEE 12th International Symposium on Computational Intelligence and Informatics (CINTI), 2011, pp.493,499, 21-22 Nov. 2011

[9] E. Tóth-Laufer, M. Takács, "Comparative Analysis of Fuzzy Logic-based Risk Evaluation Models," International Engineering Symposium at Bánki (IESB 2011), Budapest, 2011, pp. 78-90

[10] S. Szénási, Z. Vámossy, "Implementation of a Distributed Genetic Algorithm for Parameter Optimization in a Cell Nuclei Detection Project", Acta Polytechnica Hungarica, 2013, vol. 10, issue 4, pp. 59-86, 2013

[11] R. Kohavi, F. Provost, "Glossary of Terms,", Machine Learning, vol.30 issue 2, Springer Netherlands, 1998, pp. 271-274

[12] S. Szénasi, Z. Vámossy, M. Kozlovszky, "Evaluation and comparison of cell nuclei detection algorithms", IEEE 16th International Conference on Intelligent Engineering Systems (INES), 13-15 June 2012, pp. 469-475.

[13] J. Tick, Cs. Imreh, Z. Kovács, "Business Process Modeling and the Robust PNS Problem", Acta Polytechnica Hungarica, 2013, vol. 10, issue 6, pp. 193-204.

[14] X. Zhao, M.-E. Lee, and S.-H. Kim, "Improved image thresholding using ant colony optimization algorithm," in Proc. International Conference on Advanced Language Processing and Web Information Technology (ALPIT '08), Dalian Liaoning, China, 2008, pp. 210-215.

[15] K. Venkatalakshmi and S. M. Shalinie, "A customized particle swarm optimization algorithm for image enhancement," in Proc. 2010 IEEE International Conference on Communication Control and Computing Technologies (ICCCCT 2010), Ramanathapuram, India, 2010, pp. 603-607.

[16] R.-E. Precup, R.-C. David, E. M. Petriu, M.-B. Radac, S. Preitl, and J. Fodor, "Evolutionary optimization-based tuning of low-cost fuzzy controllers for servo systems," Knowledge-Based Systems, vol. 38, pp. 74-84, Jan. 2013.

[17] P. Civicioglu and E. Besdok, "A conceptual comparison of the cuckoo-search, particle swarm optimization, differential evolution and artificial bee colony algorithms," Artificial Intelligence Review, vol. 39, pp. 315-346, Apr. 2013.