Tomasz MARKIEWICZ[1,2], Żaneta ŚWIDERSKA[1], Bartłomiej GRALA[2], Wojciech KOZŁOWSKI[2]

Warsaw University of Technology (1), Military Institute of Medicine (2)

Automatic ROI selection in virtual slide images for assessment of pathomorphological diagnostic

Abstract. This paper describes the automatic region of interest selection method in virtual slide images for assessment of pathomorphological diagnostic. The proposed method for identification of high concentration of immunopositive cancer cells is based on specimen area recognition, segmentation of the immunopositive cells, local maxima recognition on cell density map and function of penalty to avoid the too high concentration of the selected ROIs. The obtained results confirm, that average 8.6±1.4 of 10 reaction fields chosen manually were compliant regionally with regions selected automatically, which is a high compliance of specimen histological examination.

Streszczenie. W pracy zaproponowano metodę automatycznego wyboru obszarów zainteresowań w obrazach wirtualnych preparatów jako narzędzie wspierające diagnostykę patomorfologiczną. Zaproponowane metoda identyfikacji obszarów o wysokiej koncentracji immunododatnich komórek nowotworowych bazuje na wydzielaniu obszaru tkanki z obrazu, segmentacji komórek immunododatnich, wykrywaniu lokalnych maksimów na mapie gęstości rozkładu komórek oraz zaproponowanej funkcji kary w celu uniknięcia nadmiernej koncentracji zwracanych obszarów zainteresowań. Wyniki liczbowe wskazują, iż średnio 8.6±1.4 na 10 wybranych pól manualnie oraz automatycznie jest tożsamych obszarowo, co skutkuje wysoką zgodność oceny histologicznej przypadków. (Automatyczna selekcja obszarów zainteresowania w wirtualnych preparatach jako narzędzie do wspomagania diagnostyki patomorfologicznej).

Keywords: image processing, mathematical morphology, pathomorphology, artificial intelligence.
Słowa kluczowe: przetwarzanie obrazów, morfologia matematyczna, patomorfologia, sztuczna inteligencja.

Background

Histopathological examination of tissues subject to immunostain tests is a basic method of recognizing a cancer and it provides valuable indicators concerning the choice of optimal therapy or defining prognostic indicators. It consists in microscopic assessment of taken material. Using appropriate immunostain tests allows to differentiate cells selectively and it is a basis of their quantitative assessment. One of the most important markers are expressions of oestrogen and progesterone receptor in breast cancers and mitotic receptor Ki-67 among others in central nervous system tumours, breast cancer, lymphomas and other cancers. The above mentioned tests are of nuclear staining type which means that immuno-positive cell nuclei are marked with a brown marker, whereas other cells nuclei are marked with purple markers. During the examination 10 regions of interest (ROI) are selected that correspond with the field of vision where the rate of cell nuclei is defined that show a positive receptors' reaction. The result of acquisition of a whole histological specimen on a scanner is RGB image called a virtual slide with very big size, for example 100 000 x 80 000 pixels. Despite development of computer vision and image processing techniques, there are only a few programme tools that are able to define ROI in an automated way. This paper contains presentation of an automatic method of ROI selection that contain a high concentration of immuno-positive cancer cells.

Materials and methods

10 cases of central nervous system cancers subject to Ki-67/MIB-1 tests come from the archives of Department of Pathomorphology, Military Institute of Medicine in Warsaw. Acquisition of images of some specimens of central nervous system tumours treated with Ki-67/MIB-1 reactions was performed on a microscope Olympus BX-61with DP-72 camera with a motorized table or on the Aperio scanner. Images were recorded in the following formats, respectively: vsi (Olympus) and svs (Aperio). Different types of generated files (svs, vsi) forced to adjust the functions of reading in of data and scaling images to a lower resolution, used in the ROI areas detection process.

Due to a very big size of images, it was necessary to reduce their resolution to such that is usable for the analysis and visualisation. In examinations we chose eight-folds reduction considered maximum to assess the image manually and by a computer. Operating the established method is connected with the following stages: defining a map of a specimen, recognising and selecting stained cells (segmentation) and selecting regions containing the biggest number of stained cells. Due to the large size of the input image and the small size of the individual cells, it is very important to properly select the parameters related to image processing, so as to obtain the most accurate information possible, while maintaining a short calculation time.

Available elaborations associated with the separation and counting of nucleus on histopathological images are based on mathematical morphology, or the use of neural networks. However, these methods are associated with the processing of a specific field of view, and not the entire virtual slide of specimen. This paper presents a method of determining of non-overlapping visual fields, which contain the largest number of cells stained to brown (immunopositive), on the basis of the analysis of the entire virtual slide.

The first stage of processing an image is defining a map of a specimen and at the same time possible fields of vision localisations showing a cancer completely. The process begins from evening brightness of an image produced with morphological operation of opening using a structural object shaped as a disc with a big radius (100 pixels), as below

$$(1) \qquad f = f / \Theta_{SE}^{f}$$

independently for every component. The differential image of components B and G is subject to processing. In order to eliminate the areas containing blood component, value of G component has been multiplied by a 0.85 factor. Through this operation, the red blood cells change the representation in the differential image to the value close to zero.

Figure 1 shows the image of the scanned specimen (virtual slide, Fig. 1a) and the created binary mask (Fig. 1b) of tissue for him. On this mask white color indicate areas of tissue, and black color indicate a background. The use of designated map of specimen helps to reduce the area to be processed only to the areas of tissue. The result is the reduction of the number of performed operations, allowing the shortening of the calculation time.
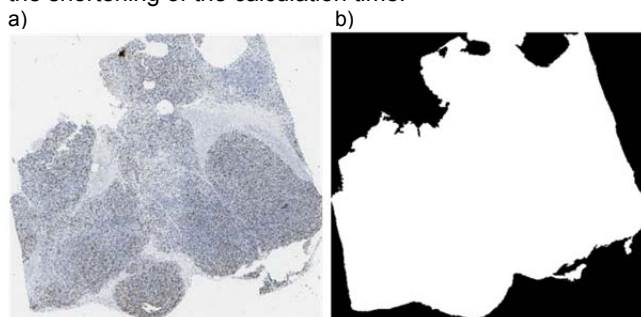
a)   b)

Fig. 1. Image of the scanned specimen (a) and the obtained binary mask of tissue area (b)

The next step is the segmentation of immunopositive cells implemented by multiple thresholding. First, only those pixels are qualified which brightness value is greater than the average for the entire image. The final threshold value is calculated using Otsu method, where global threshold is determined on the basis of the histogram and the inter-class variance, assuming the background and the object as a class. This corresponds to following criterion:

$$(2) \qquad \max \omega_{12}\{x \,|\, x > \overline{x}\}$$

The selection of ROI is associated with the designation of local maxima during the windowing of image obtained by thresholding. In order to avoid the selection of adjacent ROIs, function of penalty was proposed in the paper, associated with the distance between the designated areas and the location of another candidate for ROI. The position of ROIs depends on the value of local maxima and their arrangement. Operation of this algorithm is related to multiplication of the received mask of specimen by the mask of distance designated in accordance with the formula (3).

$$(3) \qquad
\begin{array}{l}
for \quad candidate \quad located \quad in \quad [x, y] \\[6pt]
penalty = 1 - \mu * \sum_{i} \dfrac{1}{\sqrt[4]{(x - x_i)^2 + (y - y_i)^2}}
\end{array}$$

The value of μ coefficient was indicated experimentally and amounts to 0.2. The increase of μ coefficient results in enforcing greater scattering of designated ROIs.

In case when one area has a value of the number of immunopositive cells exceeding other areas, the use of penalty allows to determine ROIs within the different locations of the tumour, and not just within a single dominating one, what would occur in case of the absence of penalty. As a result, the algorithm determines the ROIs of high immunopositive reaction and diverse location.
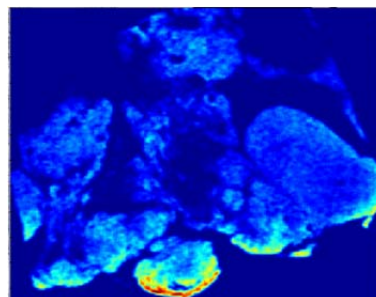

Fig. 2. Colorful map of the concentration of immunopositive cells for the specimen of Figure 1

Fig. 2 shows a colorful map of the concentration of immunopositive cells (local maxima) in the sample specimen. The blue colour represents low concentrations, while the red colour - high.

An interesting aspect of the conducted analysis may be a task of determining the contours of areas surrounding the local maxima. This allows the analysis of the distribution of ROIs selected by the expert, and regions designated by the developed algorithm with respect to local maxima. Determination of contours is related to the setting of the local maximum, and then the connection of the surrounding areas with a slightly lower concentration of immunopositive cells to it, according to the following criterion:

$$(4) \qquad \eta \geq \eta_{max} * \alpha,$$

where α coefficient was determined experimentally and amounts to 0.7. Reducing its value results in the increase of the amount of connected areas, and as a result obtaining large areas of the neighbourhood of local maxima.

Because of the possibility of occurrence of numerous artefacts, such as haemorrhages, dye stains or creases of specimen, it was necessary to conduct and expert's verification of the selected fields of view. In the event of statement of the presence of an artefact, the field of view was excluded from the analysis, and their number was supplemented by a further 10.

In order to compare the diagnostic value of ROIs selected by the expert, and ROIs selected by the program, a quantitative analysis was conducted consisting in the determination of the number of immunopositive cells and immunonegative cells, and then determining the percentage of immunopositive cells' share for each of the ROIs (Ki-67 index). In order to simultaneously test the possible variations of the obtained results and their reproducibility, an expert made an independent, double selection of ROIs for each of the specimens (Series I, Series II) with an interval of 6 months. Values of Ki-67 index calculated at the ROIs designated by the program are compared with the average value of two readings of ROIs taken by a doctor. Recognition and cell counting was carried out for the resolution of selected ROIs, compatible with the one used in the algorithm presented in the paper [3]. The figures obtained in this way allowed for assessing the correctness of ROI selection made by the developed algorithm and the differences in the numerical results of the diagnostic evaluation of the specimen between the expert and the system.

Results

Application of the developed algorithm allows to designate the map of specimen and to separate a certain number of ROIs, characterized by a high concentration of immunopositive cells. An exemplary result of the designation of the five ROIs (local maxima) with a high mitotic index are shown in Fig. 3. The automatically selected areas cover almost the entire area of increased proliferative activity of the tumour.
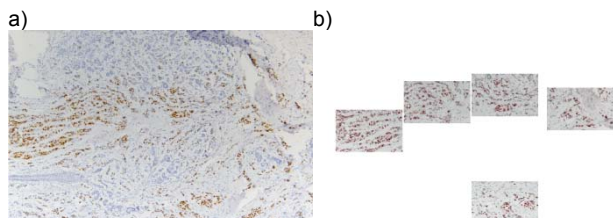

Fig.3. Processed image (a), and result of selecting five regions of interest with a high mitotic index (b)
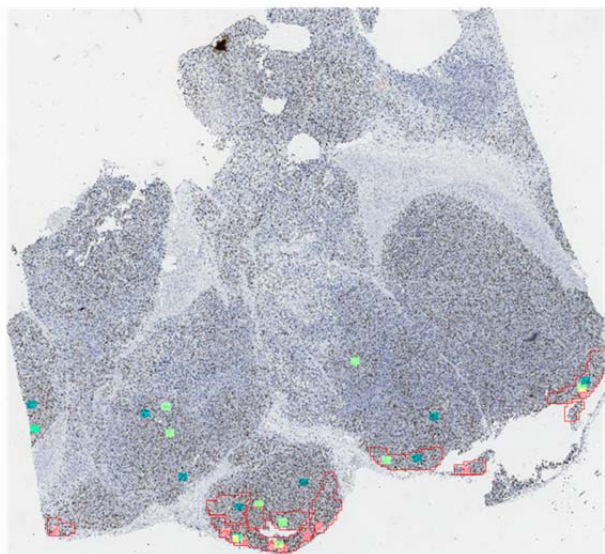

Fig. 4. Virtual slide with ROIs selected by an expert (Series I blue, Series II green), ROIs designated by the program (red) and areas of the neighbourhood of local maxima with a stable value of Ki-67 index (red contour)

Figure 4 shows the virtual slide with 10 ROIs marked by an expert selected in Series I (blue colour), 10 ROIs in Series II (green colour) and 10 ROIs designated by the program taking into account the formula (3) (red colour). Red outlines indicate areas of the neighbourhood of 10 local maxima set by the program with a stable value of Ki-67 index. On this basis it is possible to analyse the distribution of ROIs selected by the expert and the ROIs selected by the program.

In reference with results of the expert, very good compliance of localisations of searched fields in specimens was acquired. On average, 8.6±1.4 of 10 fields chosen manually were compliant regionally with regions selected automatically. Performing a quantitative analysis of immunopositive cells in the designated ROIs showed comparable or greater number of immunopositive cells in ROIs designated by the developed algorithm. An exemplary result of cell recognition in the selected ROI performed by the algorithm described in the paper [3] is shown on Fig. 5.

Comparison of the number of cells and Ki-67 index ("-" indicates immunonegative cells, "+" indicates immunopositive cells and "%" is a Ki-67 index) occurring in ROIs designated by the expert (Series I and II) and by the program for the same image in accordance with the order of their selection are shown in Table 1. The obtained results

for the investigated specimen have a high compatibility with a difference in expert's readings of about 0.6-9.6 % and between their mean value and the automatic assessment below 7.3%. It should be noted that in this specimen there are large areas of the tumour cells with high share of immunopositive cells. As a result, it is difficult to point out specific places with the highest value of Ki-67 index, and despite the considerable distance between the fields of view selected twice by the expert and by the algorithm, no significant differences in the quantitative assessment are stated.
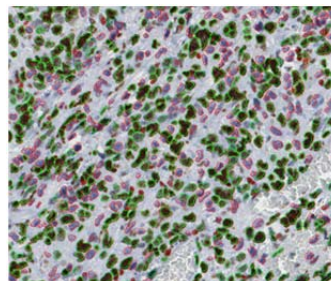

Fig. 5. The result of immunopositive (green contours) and immunonegative (red contours) cell recognition in the selected ROI

Table 1. The detailed results of the cell counting in 10 ROIs designated by the expert (Series I and II) and by the program

|	Expert						Program		
	Series I			Series II					
id	-	+	%	-	+	%	-	+	%
1	342	365	51,63	342	354	50,86	243	442	64,53
2	298	354	54,29	377	330	46,68	278	441	61,34
3	376	363	49,12	374	345	47,98	236	425	64,30
4	267	390	59,36	230	440	65,67	325	406	55,54
5	273	435	61,44	293	382	56,59	273	402	59,56
6	234	423	64,38	302	373	55,26	397	443	52,74
7	287	385	57,29	233	413	63,93	234	389	62,44
8	330	331	50,08	285	289	50,35	330	414	55,65
9	271	301	52,62	329	302	47,86	310	373	54,61
10	213	331	60,85	227	343	60,18	345	330	48,89
	mean		56,11	mean		54,54	mean		57,96
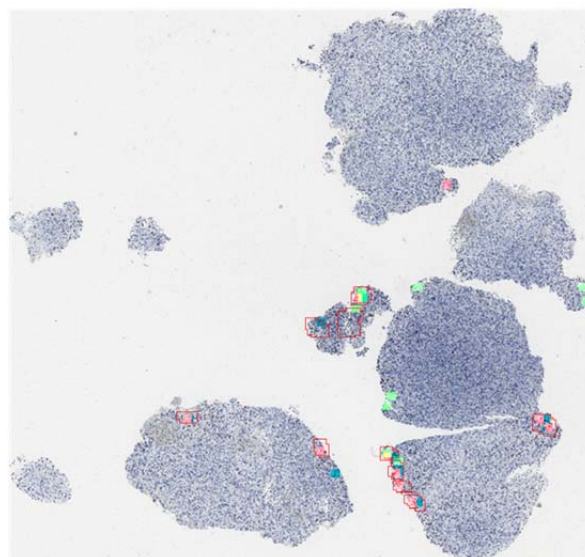
Fig. 6. The result of the analysis for the specimen with lower level of Ki-67 reaction

As a second graphical presentation of the results a specimen with lower level of Ki-67 reaction was selected, which, together with the result of the analysis, is presented on Fig. 6.

In the specimen shown in Figure 6 there are local proliferative areas of the tumour. Region with an increased Ki-67 index is located on the edge of the lower part of the specimen, which coincides with the ROIs selected by the physician (Series I and II) and the program. Comparison of quantitative assessment of the areas selected by the physician (double choice) and by the program did not show significant differences.

The analysis of 10 specimens for which ROIs were determined by an expert (physician) in two series and by the developed algorithm, showed that:

- In 8 cases the value of the percentage index Ki-67 for the fields determined automatically was comparable (the maximum difference of less than 3%) to the average value for Series I and II,
- In 1 case the automatic result was 5.4% higher than the result for the fields selected manually,
- In 1 case the automatic result was 7.3% lower than the result for the fields selected manually.

These relations can be seen on Fig. 7. The statistical examination of the difference between the average expert results and automatic results, based on Wilcoxon matched pairs test, confirmed no exist significance difference between these values ($p=0.575066$).
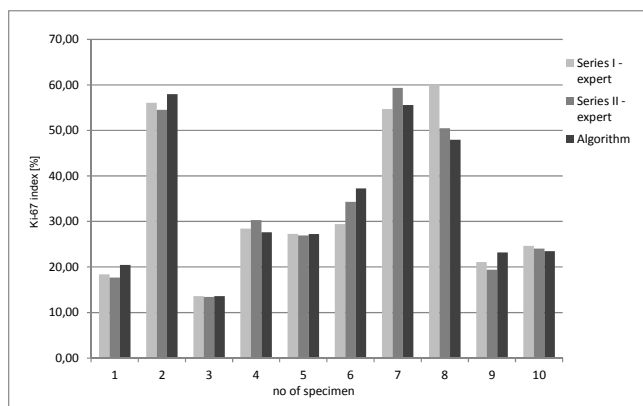

Fig. 7. The result of the analysis of 10 specimens for which ROIs were determined by an expert (physician) in two series and by the developed algorithm

Conclusions

The conducted study showed the effectiveness of the presented method of automatic designation of ROIs. Differences in the automatically obtained values of Ki-67 index do not exceed those occurring between the manual readings. The presented method is characterized by a short time of calculation and the reproducibility of results, which is a significant advantage over the traditional way of determining ROI by an expert, but also the opportunity of review of the result, selection of suggested fields, as well as increasing their number used to determine a diagnostic indicator.

REFERENCES
[1] Otsu, N.: "A threshold selection method from gray-level histograms", *IEEE Trans Sys Man Cyber*, 9 (1997), 62-66
[2] Soille, P.: „Morphological Image Analysis, Principles and Applications". Berlin, (2003)
[3] Markiewicz, T., Wiśniewski, P., Osowski, S., Patera, J., Kozłowski, W., Koktysz, R.: „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". *Anal Quant Cytol Histol*, 31 (2009), 49-62

Authors: dr hab. inż. Tomasz Markiewicz, Politechnika Warszawska, Instytut Elektrotechniki Teoretycznej i Systemów Informacyjno-Pomiarowych, ul. Koszykowa 75, 00-662 Warszawa, oraz Wojskowy Instytut Medyczny, Zakład Patomorfologii, ul. Szaserów 128, 04-141 Warszawa, E-mail: markiewt@iem.pw.edu.pl; mgr inż. Żaneta Świderska, Politechnika Warszawska, Instytut Elektrotechniki Teoretycznej i Systemów Informacyjno-Pomiarowych; mgr Bartłomiej Grala, prof. dr hab. n. med. Wojciech Kozłowski, Wojskowy Instytut Medyczny, Zakład Patomorfologii, ul. Szaserów 128, 04-141 Warszawa