The Cell Structures Segmentation

Robert Koprowski[1] and Zygmunt Wrobel[1]

University of Silesia, Institute of Computer Science, 41-200 Sosnowiec, Bedzinska 39, Poland koprow@us.edu.pl, wrobel@us.edu.pl

Summary. In this article we have presented an attempt to segmentation of cell structures images acquired while histological slides microscopic observation. The described algorithm of segmentation is also applicable in other matters, where the image segmentation is an important part.

1 Introduction

Reproducibility of cell structures microscopic-based images measurements is an important issue of biological slides analysis. For the effective tissue slides estimation and documentation the computed images analysis techniques are applied. The first step should thus be the automatic or semiautomatic picture segmentation necessary for morphometric analysis of particular cells. Microscope settings like light intensity or magnitude influence the picture quality. Hence, it is impossible to set constant thresholds, even for the same operator and the same slide, e.g. for the digitalization process. Provided biological experiment aimed to delimitation the proliferation in the rat's liver regenerating cells. The slides made of examined tissue after the proper staining (H/E), can be viewed at immersion magnitude of NIKON E 600 microscope (the lens: Plan Fluor 100x/130 Oil, the eyepiece: 10x/22) and resulted image was transferred into the PC (the processor: 3.06 GHz) using CCD Panasonic Colour camera.

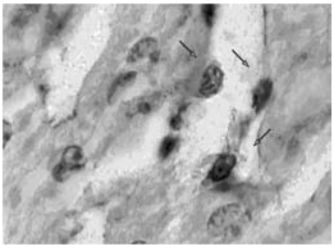

Fig. 1. An example of a microscopic- based cells image

2 The automatic image segmentation

The colour input image L_{RGB} (m,n,o) where (m - row, n - column, o - R, G or B component) at $M \times N$ resolution is inserted into the Matlab package workspace with Image Processing tool [7]. The proceeding process - described in the articles Bibliography [2], [4], [7]- towards cell segmentation is questionable at many points referring both to selection threshold and the results. Thus the modification of the presented and its expanding to image recognition parts based on decision trees seems to be advisable

3 The employment of decision trees in the segmentation process

An image segmentation process is a vitally important point of the discussed issue. The value of measured quantities (such as the object saturation level or the object area) entirely depends on the segmentation process results. On the other hand, in case of a larger number of images, the capability to automation of the segmentation process would be a deciding factor in computation acceleration. Among many segmentation methods such as: the area extending method, the watershed method or the patterns analysis method, the decision tree based method has been selected. The decision tree was used in the form of classification knowledge description for segmenting the objects (the classes space) in the L_{HSV} image. The choice of the HSV colour space has been determined here by the image contents that includes both objects of different colours and object of the same colour but different saturation levels. The proceeding process, which is presented below, can be also applied to other colour spaces (e.g. like RGB or L*a*b). The decision nodes in the mentioned case are described by features, which are the saturation levels of HSV components (generally) x, y, z. The edges of a tree qualify the possible values for the feature (the saturation level). The tree leafs are the values of the classification feature in which in that case belongs a cluster (object index). Thus the segmentation (classification) is performed by the reviewing the tree from the root up to leafs by the edges featured with features values. Decision trees in this particular application (i.e. image segmentation) have important limitations, such as:

- a risk of excessive complexity of the tree (an excessive fit problem),
- there is no easy way to update in case of a different images set.

Excessive fit for input data causes small classification error but too large real error. Such a tree, due to its complex structure, usually reflects random relations in the learning data set (e.g. image noise). Due to their characteristic structure which makes them suitable to represent any hypothesis, decision trees are particularly exposed to dangers of that kind

The solution to this problem is trees truncation (i.e. truncation of their excessive fit), which could be simply explained as replacement of the original tree with its sub-tree. The advantages of the decision trees application for the discussed purpose include:

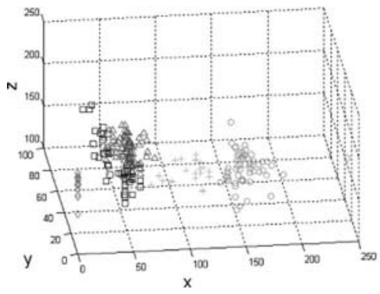

Fig. 2. The arranging of pixels to 5 clusters (marked with colours respectively)

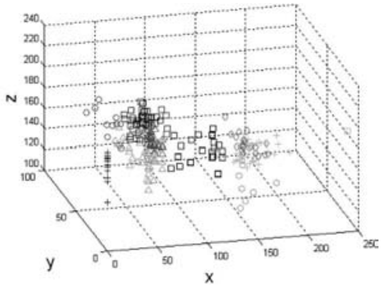

Fig. 3. The arranging of pixels to 10 clusters (marked with colours respectively)

- a capability for representing any hypothesis - a capability for performing image segmentation with any required precision,
- efficiency of the classification process - a capability for logic employment,
- efficiency in case of large learning sets - a capability for the classification trees employment for a segmentation process of any computation and logical complexity,
- a clear representation - assuming that the tree is not too complex and does not combine too many features.

The construction of the classification tree starts by L_{HSV} pixels clustering featured in the features space (L_{HSV}(m,n,1) the x coordinate, L_{HSV}(m,n,2) the y coordinate and L_{HSV}(m,n,3) the z coordinate) into 5 and 10 clusters (indexes) - see Fig. 2 and 3. The tests were performed for the following metrics: Minkowski, road, city and Chebyshev. In respect of strong influence of orders difference of features vector individual components, the normalization process was used. Finally the inner averages algorithm with Euclidean metric was used. The mentioned charts in the features space are shown below.

The object division created that way enables proceeding with decision tree creation. The decision tree generated that way is shown in the Fig. 4 and 5.

Based on this, the range of x, y, z values variability can be expanded against the entire features space containing the values range $x \in (0, 240), x \in (0, 100), x \in (100, 220)$, what was shown in Fig. 6, 7, 8 and 9.

According to mentioned disadvantages of excessively complex decision trees the MDL (Minimum Description Length) rule was used to the truncation purpose. The minimal cost rule allows to delimitation of optimal tree nodes quantity, what was shown on the chart in Fig. 11.

Based on this, in discussed case the optimal decision nodes quantity has been set to 9. The structure of truncated modified decision tree is shown in Fig. 12.

The result images, obtained on basis above, (each cluster reflects one index) are shown below (Fig. 13 - 17).

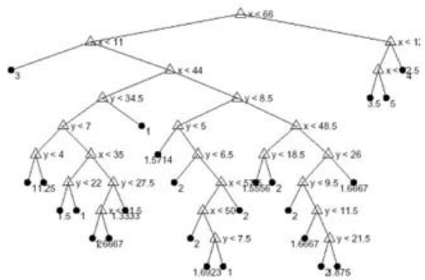

Fig. 4. The decision trees obtained for 5 clusters in the two features space (the H and S component)

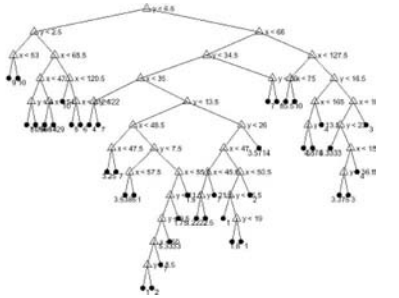

Fig. 5. The decision trees obtained for 10 clusters in the two features space (the H and S component)

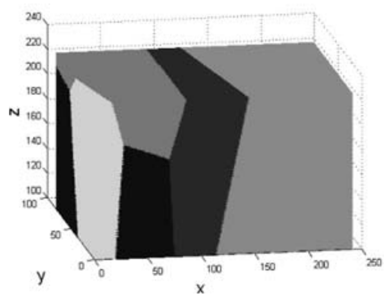

Fig. 6. The example of area boundaries in the features space for 5 clusters

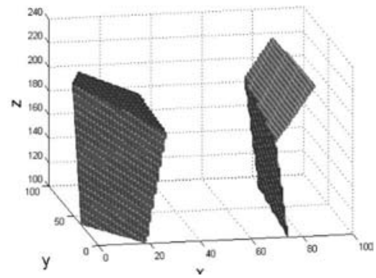

Fig. 7. The arranging of pixels to 5 clusters in the an example features space range (marked with colours respectively)

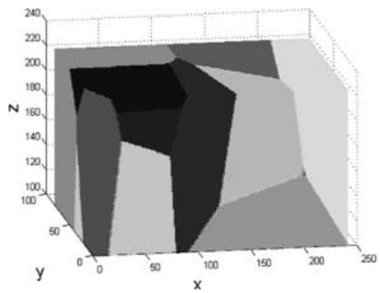

Fig. 8. The arranging of pixels to 10 clusters in the an example features space range (marked with colours respectively)

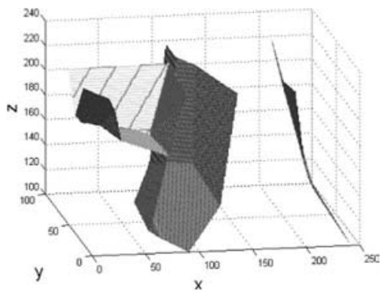

Fig. 9. The example of area boundaries in the features space for 10 clusters

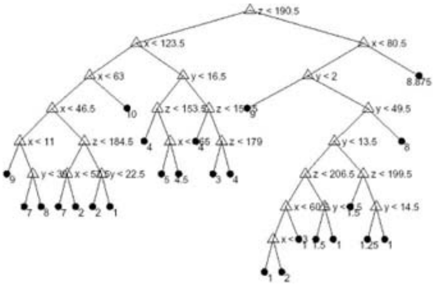

Fig. 10. The decision trees obtained for 10 clusters in respect of x, y and z variables

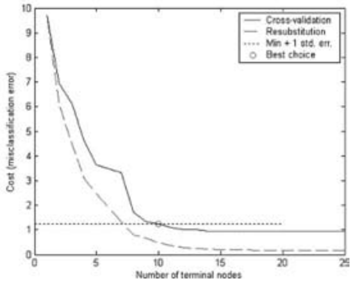

Fig. 11. The cost change chart as function of decision nodes quantity

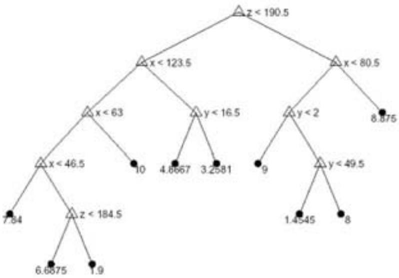

Fig. 12. The decision tree from Fig. 10 - truncated

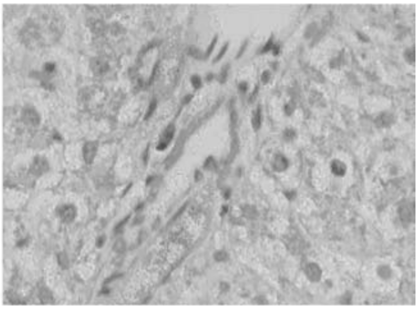

Fig. 13. The input image

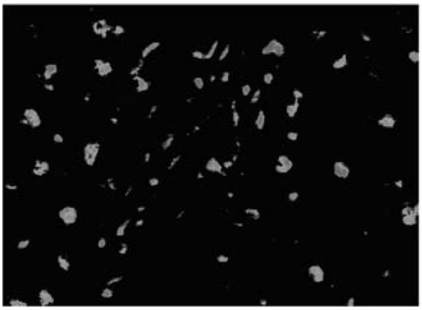

Fig. 14. The image obtained from the second cluster

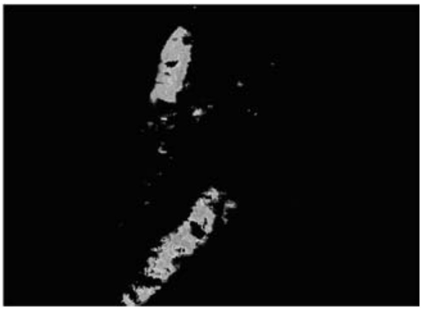

Fig. 15. The image obtained from the eight cluster

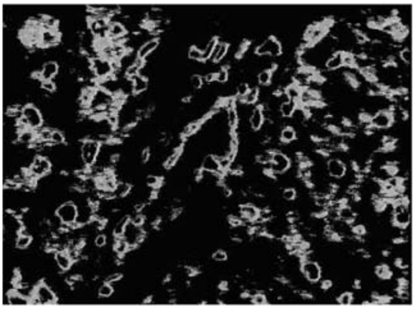

Fig. 16. The image obtained from the first cluster

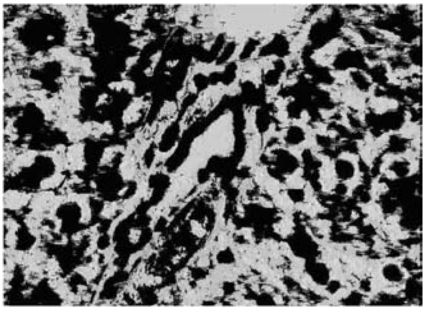

Fig. 17. The image obtained from the fifth cluster

In order to automate next algorithm phases the thresholding has been modified using the Nobuyuki Otsu formula. Presuming that grey scales or generally saturation levels i have the value i=0...k...g and searched threshold level is marked as k_c, we can write down:

$$\sigma_B^2(k) = \frac{[\mu_T - \omega(k)]^2}{\omega(k)[1-\omega(k)]} \tag{1}$$

where:

$$\omega(k) = \sum_{i=0}^{k}\sum_{m=1}^{M}\sum_{n=1}^{N}\frac{p(i/(m,n))}{M\cdot N} \tag{2}$$

$$p(i/(m,n)) = \begin{cases} 1 & dla \quad L_{GRAY}(m,n) = i \\ 0 & dla \quad L_{GRAY}(m,n) \neq i \end{cases} \tag{3}$$

- the zero order moment

$$\mu(k) = \sum_{i=0}^{k}\sum_{m=1}^{M}\sum_{n=1}^{N}\frac{i\cdot p(i/(m,n))}{M\cdot N} \tag{4}$$

- the first order moment

$$\mu_T = \sum_{i=0}^{g}\sum_{m=1}^{M}\sum_{n=1}^{N}\frac{i\cdot p(i/(m,n))}{M\cdot N} \tag{5}$$

- the maximal existing value for chosen k_v

$$\sigma_B^2(k_v) = \max_{0<k<g}\sigma_B^2(k) \tag{6}$$

If k_v is a vector of F elements, in case of finding more then one maximal value, then:

$$k_c = \frac{1}{F}\sum_{w=1}^{F} k_v(w) \tag{7}$$

The results are the objects, which are the cells with the background. So it seems reasonable to perform the binary-splitting process properly to provided relation for delimited areas, which contain also the cell area, according to relation (1) for obtained level k'_c [5], [7]. The binary image L_{BIN} can be written as:

$$L_{BIN} = L_{GRAY}(m,n) > k'_c \tag{8}$$

The closing operation using the structure element $H(j,k)$ of $J \times K$ size is applied to the result image i.e.:

$$L'_{BIN}(m,n) = \begin{cases} L_{BIN}(m,n) & for \;\; (k_w+1)\cdot k'_c > s_r \\ \max_{m_i,n_i \in H}(L_{BIN}(m_i,n_i)) & for \;\; (k_w+1)\cdot k'_c \leq s_r \end{cases} \tag{9}$$

where:

$$s_r = \sum_{m_i=1}^{J}\sum_{n_i=1}^{K} \frac{L_{GRAY}(m_i,n_i)}{J\cdot K} \tag{10}$$

and:

$$k_w = [0, 0.1, 0.2, ..., 0.5] \tag{11}$$

The conditional erosion has been defined analogous way, i.e.:

$$L'_{BIN}(m,n) = \begin{cases} L_{BIN}(m,n) & for \;\; (1-k_w)\cdot k'_c > s_r \\ \min_{m_i,n_i \in H}(L_{BIN}(m_i,n_i)) & for \;\; (1-k_w)\cdot k'_c \leq s_r \end{cases} \tag{12}$$

The (9) and (12) relations were used for the closing operation using the H structural element with the k_w coefficient arbitrarily set at 0.2 level [1], [6], [2], [4]. The selectivity of the algorithm was increased by decreasing the H mask size for following iterations up to the 5×5 level ($J \times K$). The results are visible at zoom (see Fig. 18 and Fig. 19).

The indexation process (L''_{BIN}) has been performed on the L''_{BIN} image, where pixels were included into an object if they are in the eight neighbour 'distance' from the others. Next the following morph-metric parameters were calculated for each of delimited objects (indexes) [1], [3]: area, the edge length using Crofton formula (according to Cauchy rule), shape parameter (Blair-Bliss). In selected instances of analysed cells the problem of connected objects occurs, which has been eliminated by the additional performance of the opening operation on objects which the square area stays in standard level while the diameters ratio is over threshold.

Fig. 18. The L_{BIN} binary image superimposed on a part of the L_{RGB} image

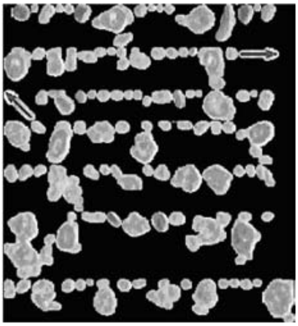

Fig. 19. The L''_{BIN} binary image superimposed on a part of the L''_{RGB} image

4 Conclusion

Fully automatic cell structures analysis gives new information like shape of particular objects. With employment of the presented decision trees algorithm, biological diagnostic support goes here fully automatically. Additionally created tools, which are a program option, allow on-line data import into the Excel spreadsheet. In addition, the described proceeding process of image processing - especially the image segmentation - can be extended on images of different contents. Independently from the fact that the way to a full expert system supporting diagnostic is still long, it seems that presented method is the beginning of new automatic morphometric factors measurements research direction.

References

1. Bajcsy R. (1973), Computer Description of Textured Surfaces, Proc, Int. Conf. Artificial Intell., Stanford, Calif, pp. 572-579.
2. Davis L.S. (1992), Hierarchical Generalized Hough Transforms and Line-Segment Based Generalized Hough Transforms, Pattern Recog., vol. 15, no. 4, pp. 277-285.
3. Forczek M. (2000), Colour Space Models in Colour Images Segmentation Algorithms, ZNPS, no. 38, pp. 171-195, Silesian Technical University Publisher, Gliwice (in Polish).
4. Fu, K.S., Mui J.K. (1981), A Survey of Image Segmentation, Pattern Recog., vol. 13, no. 1, pp. 3-16.
5. Gonzalez R.C. (1992), Computer Vision, Yearbook of Science and Technology, McGraw-Hill, New York, pp. 128-132.
6. Pavlidis T. (1972) Segmentation of Pictures and Maps Through Functional Approximation, Comp. Graph. Image Proc., vol. 1, pp. 360-372.
7. Wrobel Z., Koprowski R. (2004) The Practice of Image Processing in Matlab Application, Academic Publisher Exit, Warsaw (in Polish)