Robust Image Segmentation Method for Cytological Slides

Ramona Galatus[1], Sorina Persa[2], Tiberiu Marita[2], Viorel Trifa[3],
[1]*Basis of Electronics Department,* [2]*Computer Science Department,* [3]*Electrical Drives and Robots*
Technical University of Cluj-Napoca, Romania
Galatus.Ramona@bel.utcluj.ro, Tiberiu.Marita@cs.utcluj.ro

Abstract

The traditional process for detecting the cervical cancer is called Pap smear testing and it is the most widely used screening technique. The pathologists diagnose the smear according to its normality or abnormality. The huge number of slides to be analyzed requires an automated computer-aided system which can help in diagnosis process. This paper presents a robust image segmentation method able to detect from each slide only the views containing abnormal nuclei (suspicious to precancerous and cancerous changes). A hybrid, multi-spatial method is proposed which combines fundamental techniques used in image segmentation: thresholding, labeling, color clustering, morphological operations etc. in an original way for the nuclei detection. After the segmentation phase a set of relevant features categorizing the cervical cancer evolution are extracted. The method can be used in conjunction with an automated slide scanning system and a classifier method in order to help the pathologist in the final diagnosis.

1. Introduction

Mass screening programs of the population to identify seemingly healthy individuals has been a growing trend for over 30 years, to prevent some diseases which represent a substantial public health burden. Cervical Intraepithelial Neoplasia (cervical cancer, called CIN) is one of the most common form of cancer among women. As all the tumors, cervical intraepithelial neoplasia changes continuously in time. Over longer periods of time, its physical and functional characteristics change significantly. CIN has three grades of evolution: 1 (mild), 2 (moderate) and 3 (severe) and is characterized by some distinct features [1]: disproportionate nuclear enlargement, which leads to high nuclear-to-cytoplasm ratio, hyperchromasia, irregularity in form and outline of the nuclei, irregular chromatin distribution, presence of keratinization, abnormalities in the number, size and form of the nucleoli, multi-nucleation. If the CIN is detected in the first stage of disease it can be easy treated.

The traditional detection process, called Pap smear testing, consists in the examination of cells collected from the uterine cervix in a cytology laboratory under the microscope. Even the best laboratories can miss from 10 up to 30% cancerous cases (false negatives), due to the following reasons: huge number of normal slides being analyzed, each containing a huge number of cells, the big number of slide's zones (views) that must be explored under microscope, and the short time allocated to each slide (10-15 minutes).

The impossibility for the human eye to detect all cases of cancer, lead to the need of creating some automated methods of cancer detection. The most significant researches done so far are: :the Papnet (a system based on neural networks) [2], a fractal analysis-based approach [3], an optics-based approach [4], visual inspection detection [5], direct visual inspection or self-adaptive methods based on the existence of localized group of discriminatory elements [6] or image-processing based approaches applied on single cell images [7], [8]. Unfortunately, most of these approaches are not completely satisfactory and none detects perfectly the tumor. Besides, no such system is used in our country.

Existing cell segmentation techniques are boundary-based, region-based or model based approaches. Variations of these techniques were applied also in pap smears segmentation. For example local methods like region growing and edge detection in [6], or morphological and multi-resolution segmentation which work either globally or locally on

image [9] or neural network based approaches as the CHAMP tool [8] have improved the performances of traditional segmentation methods. But all these methods were tested on limited sets of images in databases which cover only particular casees: images with a single nucleus, acquired in the best conditions, even pathologist-pre-classified interest cases in conformity with some MAC (Malignancy Associated Changes) features [9].

Some high cost systems as Bethesda [10] and LifeLinger [11] are able to work on large image database and can evaluate samples based on a standard set of criteria using technology available now for many physicians. These systems are in progress permanently to cover MAC features but are removing specimens that can't be processed by some reasons: wrong biochemical specimen preparation or failure to distinguish between cancer cells and other kinds of objects in the specimen (cellular clusters, debris, degenerate cells etc.).

An additional low-cost solution appeared in 2002 and patented as NMP179 is based on a nuclear matrix protein marker for detection of pre-neoplastic lesions of the cervix. The large-scale pre-clinical testing began in USA in 2004 by Matritech Inc. sustained by American Cancer Society. It was tested for high-grade squamous intraepithelial lesions [12] and was 92% sensitive, but it was not yet large scale adopted as a standard process (less than the 97% reported by Papnet [2]).

Since the traditional detection method is subjective valuate and not very reliable, in this paper a solution intended to back up the pathologists by the means of an automated tool is proposed. The detection of CIN calls for an automatic inspection consisting in nuclei detection. In the case of a positive authentication (abnormal nuclei having enlarged diameters), a set of relevant features are extracted. The pathologist will analyze only the slide regions containing abnormalities and will take the final decision in classifying the evolution grade of the CIN using the detected features.

The segmentation of the nuclei is perhaps the most difficult part of the whole process because upon its accuracy depend the next steps. A major factor in the segmentation is the quality of sampled smear slides. Although there are some techniques to improve the quality of slide preparation, such as ThinPrep Test, AutoCyte Prep and Sure-Path System [13] the available slides used for our experiments were exhibiting some unexpected features like, non-uniform cellular cluster distribution, cell clumps, mucus etc. A large number of algorithms were tested to find an optimal and robust solution. A hybrid, multi-spatial method was proposed which combines fundamental techniques used in image segmentation: thresholding, object labeling, color clustering and morphological operations in an original way for the current problem

2. System overview

The automated scanning of the smear slides is perfumed using a manipulating robot (Figure 1) able to supply the microscope with slides and position the slides under the microscope with planar movements [14]. For each slide a set of several views (images) covering its whole surface (the view of the microscope is much smaller then the size of the slide) are acquired. Further each view is analyzed in order to detect the presence of neoplasia.

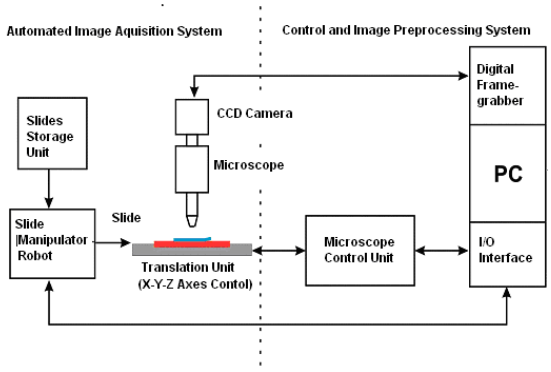


Figure 1. Schematic View of the system architecture

Neoplasia covers a wide range of abnormalities which occur in the skin of the uterine cervix. Detecting the presence of cervical cancer in a smear is a process that can be decomposed into the following steps:

1. Nuclei segmentation,
2. Abnormal nuclei detection,
3. Features extraction (at both the nucleus and cytoplasm level)
4. CIN grade of evolution classification.

Since first and second step are applied on the whole images needed to be analyzed, the third and fourth steps are applied only if abnormal nuclei are detected. The features extracted are used in the last phase when the pathologist has to take the final decision regarding CIN's grade of evolution for the suspicious images. In the following figure are summarized the steps of CIN detection process.

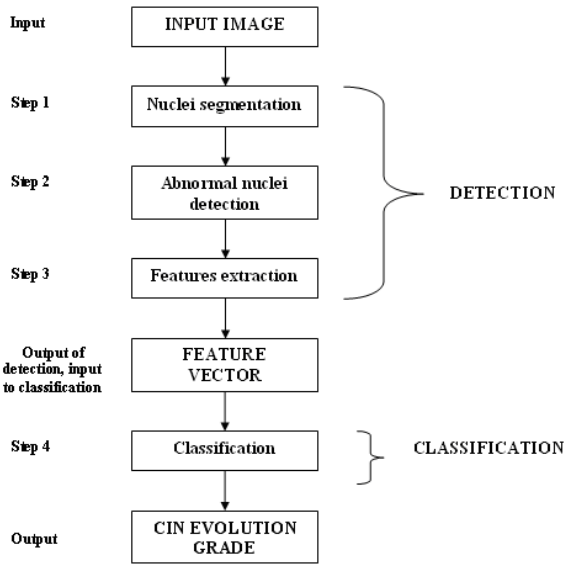


Figure 2. CIN detection phases

3. Nuclei segmentation

The purpose of the nuclei segmentation is to outline each nucleus for further analyzes regarding its size and structure. The segmentation method based on edge detection and reconstruction (region filling) implied in [15] worked well with perfectly prepared smear slides, but for slides prepared with traditional techniques (as is in the case of many of the laboratories) we had to deal with non-uniform cellular cluster distribution, cell-clumps, mucus etc.

Therefore we propose an original method for robust segmentation of nuclei that overcomes the bad quality of the prepared smear slides. The method is based on the combination of some dedicated segmentation techniques. The algorithm was named *Multi-spatial hybrid segmentation algorithm* of nuclei and it consists in the following steps:

1. Thresholding using Kapur's method [16] on each RGB color channel, followed by objects labeling and noise reduction (small objects are rejected - Figure 3.b).
2. Thresholding with an arbitrary threshold on the grayscale image followed by labeling and noise (small objects) elimination. For every labeled object the following processing steps are applied:
   - Thresholding using Kapur's method on its intensity/grayscale component;
   - Thresholding using minimum error method [17] on its grayscale image;
   - Combining the results of the two thresholding methods.
3. For every label obtained at step 1 a spatially correspondent label is searched in the result of step 2. If its correspondent exists, the label obtained in step 2 is stored. Otherwise:
   - For each of the labels resulted from the gray combination (step 2), its cluster in on the hue component of the HSI space [18] and its maximum radius (we assume that for each label we have a single cluster) are computed;
   - For each label remained from the color segmentation (step 1), a checking is performed: in order to see if a certain percentage of its pixels belongs to one of the clusters previously computed (assumption is useful for removing the mucus if it hasn't been detected by the grayscale combination and it is based on the supposition that in an image the abnormal nuclei have similar characteristics);
   - If it does, the label remains, otherwise is deleted.
4. On the labels remained from the color segmentation and on those from the grayscale combination whose area exceeds the maximum abnormal nuclei threshold area (since the abnormal nuclei are the objects with the biggest size in the image, obtaining labels with an area bigger than this maximum means that the segmentation did not succeed):
   - The *multi-space segmentation algorithm* is applied
   - Noise (small objects) is removed

The *Multi-space segmentation algorithm* is proposed in order to encompass the maximum information from the image by using as many as possible color/intensity components: the color (Hue) and the intensity (I) of the HIS color space and the Red, Green and Blue components of the RGB color space. The steps of the algorithm are as follows

1. An initial segmentation by thresholding with Kapur's method on the color (RGB) image is perfomed.
2. For each of the following components (hue, intensity, red, green, blue):
   - The histogram is computed;
   - The threshold by the means of the global adaptive threshold method [18] is computed (in this case the hue space will be considered

as if it were a linear space and not a circular one);
- The shape of the histogram is checked.

3. For those components that proved to have a bimodal histogram, thresholding is applied and the results are composed:
   - Thresholding is applied with the purpose of obtaining a binary image, so the results won't be made separately on the components, but on the final image;
   - The composed result is obtained by performing an AND between the thresholding results;
   - If none of the components histograms proves to be bimodal, thresholding in the intensity space is applied

4. **Abnormal nuclei detection**

From the nuclei and the mucus segmented so far (Figure 3.c), we have to identify the atypical nuclei and eliminate the objects without interest: small and healthy nuclei or mucus cells (small or irregular shaped cells). An improved version of the "hit&miss" algorithm [18] was adopted. The morphology-based classical "hit&miss" algorithm described in [15] has some major disadvantages:

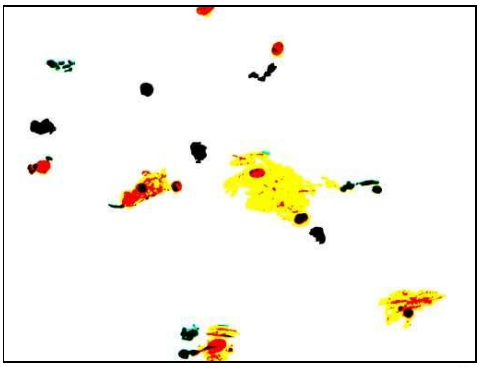
- Increased processing time while applied on the whole image, depending on the image size and nuclei size
- The detection fails when there are noise elements or other nuclei around the nucleus we are trying to identify
- If the nucleus of interest has a single background (noise) pixel inside, the identification fails

The first measure that must be taken is to reduce the search space. Therefore, square-shaped regions of interest were defined around the segmented nuclei having their origins in the centers of mass of the nuclei, and sizes equal with:

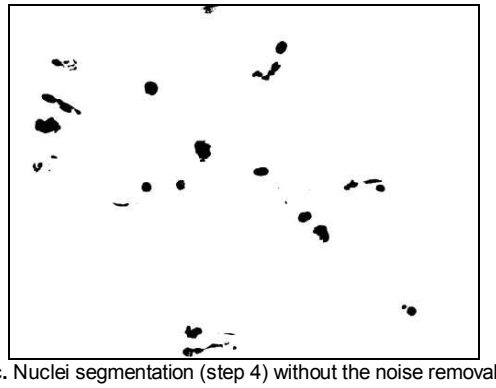
- half of the minimum diameter of an abnormal nucleus in the case of the hit operation;
- half of the maximum diameter of an abnormal nucleus In the case of the miss operation.

In each search area, we perform first of all what is it meant by *hit*, which is actually a pattern matching operation with a disk kernel (Figure 4.a). If the number of points matched is bigger than a certain percentage of the disk kernel's area (90% in our case), the hit part is considered accomplished. Otherwise, the hit process is considered failed for that label.

a. Original image (color image). Small objects (dark colored) are mucus

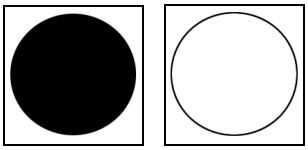


b. Nuclei segmentation (steps 1-3). Each label has a different grayscale value in the image



c. Nuclei segmentation (step 4) without the noise removal

Figure 3. Nuclei segmentation results using the proposed *multi-spatial hybrid algorithm*



a. **b.**

Figure 4. The kernels used for the *hit* operation (a) and for the *miss* operation (b)

The miss part of the transform is performed only on the labels whose hit operation was a success:

- The local background kernel is a circle kernel (Figure 4.b), also represented as a matrix, where the circle's points value is zero, and background points value is one (255);
- While matching the kernel's origin with each point from the search space, we check if the kernel's circle points in the label space differ from the label's value. If at least one point has the label's value, the miss is considered failed;
- If all the points are fulfilling the condition previously mentioned, the hit is considered successful and the current label/object is considered an abnormal nucleus.

It can be noticed that while in [15] the hit and miss transform was implemented as a morphological operation, in this case it was not anymore. The conditions corresponding to the both operations were loosened:

- In the hit case, by introducing the condition that only a percentage of the area must be covered, which implies that we can give up the reconstruction process needed for filling the nuclei [15];
- In the miss case, with the condition that the circle kernel's points should not equal the background, but it is enough to differ from the current label.

The results of the abnormal nuclei detection applied on the labels/objects obtained in Figure 5.b can be seen in the Figure 5.c.

5. Feature extraction of atypical nuclei

In order to quantify the CIN evolution grade, along with the disproportionate nuclear enlargement which was the selection criteria of the abnormal nuclei detection, a set of features characterizing hyperchromasia, irregular chromatin distribution, presence of keratinization [1] were extracted. The features were grouped into two categories: abnormal nuclei features and abnormal cell features.

In the category of the abnormal nuclei features color and texture based features characterizing hyperchromasia and irregular chromatin distribution were considered. The features were extracted on the area covered by each abnormal nucleus segmented in the previous steps.

- Color-related features: number of colors (*nColors*) and number of clusters (*nClusters*) computed on the Hue (color) component in the transformed HSI space;
- Texture related features: *entropy*, *homogeneity*, *coarseness*, and *business* computed using the Gray Levels Co-occurrence Matrix. And Autocorrelation Matrix [19].

Another feature of the nucleus that we have considered was the *fractal dimension*. As the results of [3] proved, the fractal dimension of abnormal nuclei is distinct from that of the typical nuclei. Besides, the fractal dimension of different evolution grades differs, but it cannot be used as the only differentiation feature but in conjunction with other color and texture features could be a valuable prognostic marker. To compute the fractal dimension of a nucleus, a region of interest is selected as the biggest square that can be contained in it. The fractal dimension is estimated applying a box-counting algorithm on the grayscale image of the selected region of interest[20],

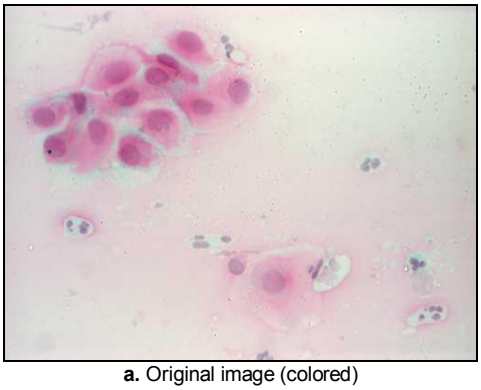
Regarding the abnormal cell features, the presence of the keratinization, which is represented as the presence of an orangeophilic tone in the cytoplasm, is searched. To determine if the cytoplasm has an orangeophilic tone, a rectangular search space around the abnormal nucleus is defined and in that space, we are looking for pixels whose hue value and saturation value are comprised in the intervals mentioned below:

$$15<H(x, y)<35$$
$$30<S(x,y)<100$$

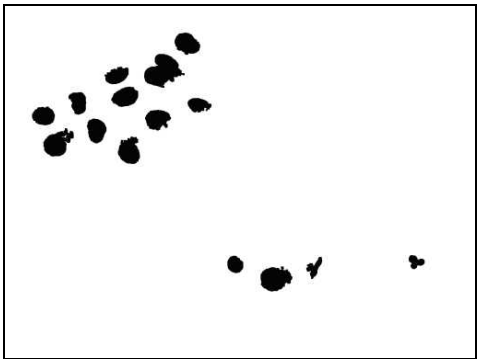
We count the pixels having the previously mentioned attributes and if their sum is bigger than a predefined threshold, we can conclude that keratinization is present.

6. Results

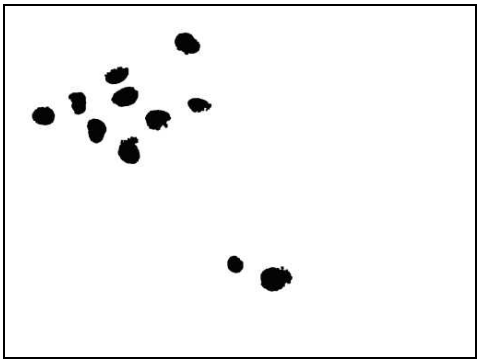
The image segmentation and abnormal nuclei detcetion algorithms were tested on the available smear images database. The slides were prepared with traditional techniques, therefore have presented unexpected features like cell clumps or mucus as it can be seen in figures 3.a and 5.a. In Figure 5 the results of the complete processing used to detect the abnormal nuclei is presented. It can be seen that the nuclei from the upper left cell group (merely clumped and hardly distinguishable from each other and from their cytoplasm) were properly segmented. Also the mucus (dark colored features) representing noise in our case was correctly eliminated.



a. Original image (colored)



b. Nuclei segmentation



c. Abnormal nuclei detection

Figure 5. The complete process of image segmentation and abnormal nuclei detection

For the detected nuclei from Figure 5, their features are computed (Table 1). From the computed features, the relevant ones (with discrimination power in terms of sensitivity) are plotted in Figure 6. As it can be seen the *nColors*, *nClusters*, *business* and *coarsness* features have similar shapes (the *coarseness* has only inverted values) and can be used for further classification purposes, while the fractal dimension can be used as an additional discriminating feature in ambiguous cases.

Table 1. Extracted features of the abnormal nuclei (CIN2) from Figure 5

nCo.	nCl.	Ent.	Hom.	Busy.	Coars.	FD
2441	20	2.53	0.45	0.661	0.013	1.96
1027	12	2.57	0.40	0.611	0.022	4.71
2436	19	2.78	0.43	0.660	0.013	2.14
1230	16	2.47	0.45	0.639	0.016	4.65
1548	16	2.51	0.43	0.644	0.016	2.27
1310	12	2.48	0.44	0.633	0.018	4.34
1122	14	2.54	0.43	0.628	0.019	1.97
1840	13	2.80	0.38	0.631	0.018	1.88
1792	15	2.67	0.45	0.646	0.015	2.02
1308	15	2.53	0.47	0.636	0.017	2.55
739	7	2.30	0.47	0.618	0.021	2.36

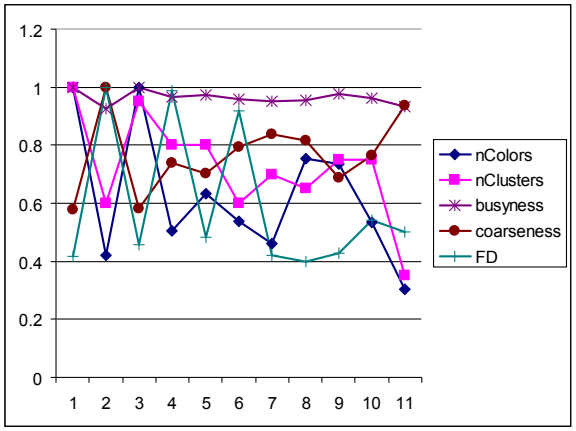


Figure 6. Normalized features' values vs. nuclus number.

7. Conclusions

In this paper we have purposed a method for automated examination of the cytological slides. The image processing algorithms that we have used were intended to automatically identify the image regions that contain tissue with uncommon characteristics that suggest abnormality. We have tried to find a solution as robust as possible which has to work even for not perfectly prepared slides (non-uniform cellular cluster distribution, cell clumps, mucus etc).

Therefore a hybrid, multi-spatial method was proposed which combines fundamental techniques used in image segmentation: thresholding, labeling, color clustering and morphological operations in an original way for the abnormal nuclei detection. After the segmentation phase a set of relevant features categorizing the cervical cancer evolution degree were extracted.

Further work will be focused on an automated classification tool of the CIN using a neural network approach. The neural network will have as inputs a combination of the already extracted features for each nucleus and will have as output the evolution degree of the CIN, quantified in three classes: (I) - mild, (II) – moderate, (III) – severe. For that purpose database with a large training set will be built up. The system will be tested in a cytological laboratory using a robot for automated slide manipulation and image acquisition.

10. References

[1] W. Gray, Diagnostic Cytopathology, *Churchill Livingston*, 1995

[2] D.L. Rosenthal., D Acosta., R.K Peters, Computer-assisted re-screening of clinically important false negative cervical smears using the PAPNET testing system. *Acta Cytol* 1996, no. 40, pp.120-126.

[3] R Sedivy & all, Fractal Analysis an Objective Method for Identifying Atypical Nuclei in Dysplastic Lesions of the Cervix Uteri, *Gynecologic Oncology,* No. 75/1999, pp.78-83.

[4] S. Bains, Optical correlators may reduce false negatives in cervical cancer screening, *OE Reports*, May, 2000.

[5] R. Sankaranarayanan, Early detection of cervical cancer with visual inspection methods: A summary of completed and on-going studies in India, *Salud Publica Mex,* 2003.

[6] R. F. Walker, Adaptive Multi-Scale Texture Analysis With Application to Automated Cytology, *PhD Thesis*, *Queensland University of Technology, Australia*, 1997.

[7] J. Bryel, Neuro-Fuzy Classification of Cells in Cervical Smears, *MSc Thesis, Tehnical University of Denmark*, 1999.

[8] CHAMP – Cytology and Hystology Analysis Modular Package, *DIMAC Company*, Denmark.

[9] J. S. Hallinan, Detection of Malignancy Associated Changes in Cervical Cells using Statistical and Evolutionary Computation Techniques, *PhD Thesis, Queensland University of Technology,* Australia, 1999.

[10] The Bethesda system of reporting cervical/vaginal cytologic diagnoses. *Acta Cytol* 1993; 37: 115-124.

[11] Lifelinger – www.lifelinger.net/menu/home.htm

[12] NMP179 – www.matritech.com , www.sysmex.com

[13] CIGNA Health Corporation Report, Cervical Cancer Screening Technologies, Coverage Position 2005.

[14] R. Galatus, T. Marita, S. Persa, D. Moga, M. Dumitrean, V. Trifa,, Proposal for a Computerized Cytological Screening System, *ICCCP2005*, 24-26 May, Lilafured-Miskolc, Hungary, pp.443-448,

[15] R. Galatus, S. Persa, D. Moga, V. Trifa, L. Neaga, T. Marita, R. Munteanu jr., Toward Abnormal Nuclei Detection Using an Integrated Automatic System.. *Acta Electrotehnica*, vol. 45, nr.5/2004, pp. 553-560.

[16] J. N Kapur, P. K. Sahoo, A.K.C. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, *Graphical Models and Image Processing*, 1985, No. 29, 273-285.

[17] J Kittler, J. Illingworth, Minimum error thresholding, *Pattern Recognition*, Vol. 19, 1986, pp. 41-47.

[18] C. Zhang, P Wang, A New Method of Color Image Segmentation Based on Intensity and Hue Clustering, *Proceedings of the 15th International Conference on Pattern Recognition,* 2000, pp. 3617-3620,

[18] R. C. Gonzales, R. E. Woods, *Digital Image Processing*, 2001

[19] M Sharma, M. Markou, S. Singh, Evaluation of texture methods for image analysis , *7-th Australian and New Zealand Intelligient Information Systems Conference*, Perth, 2001, pp. 117-121.

[20] R. Galatus, Metode de extragere automată a caracteristicilor structurale şi funcţionale ale zonelor de interes din imaginile medicale, *PhD Report, from „Tecnical University of Cluj-Napoca",* 2002.