Detection and Shape Feature Extraction of Breast Tumor in Mammograms

Zaheeruddin , Z. A. Jaffery and Laxman Singh

Abstract − **An accurate and standard techniques for breast tumor segmentation plays a pivotal role in detecting and quantifying breast cancers. Segmentation of breast tumor in mammograms presents many challenges related to selection of optimal threshold in various segmentation techniques. In this paper, a mean based region growing segmentation (MRGS) is presented that automatically finds the seed pixel and optimal threshold value and thus makes the segmentation process very fast and accurate. Furthermore, experimental results are compared with the findings of expert radiologist and marker controlled watershed segmentation approach. A set of 3 mammogram images is used to demonstrate the effectiveness of the segmentation methods. Numerical validation of the results is also provided.**

Keyword: **breast tumor, mean based region growing segmentation, marker controlled watershed segmentation, mammograms, threshold.**

1. Introduction

Breast cancer continues to be the most common diagnosed cancer among women in US [1]. In the United States, every year, approximately, 182,000 new cases of breast cancer are diagnosed and more than 46,000 women die of it [2-3]. Since the causes of breast cancer still remain unknown, early detection is the key to control the breast cancer [4]. Now days, mammography is considered to be the most reliable imaging modality for an early detection of breast carcinomas [5].

If the tumor is detected at an early stage, the chances of successful treatment as well as patient survival rate increases significantly. These days, in most of the hospitals, radiologists performs the diagnosis of breast tumor manually on mammographic images, which is time consuming and error prone process due to (i) small size and various shapes, (ii) low contrast and unclear boundary from surrounding normal tissues. It has been observed that, 10-30 percent of breast cancers are missed even by experienced radiologists during routine screening in manual diagnosis [4]. Therefore, a number of computer based techniques have been proposed for identifying and segmenting breast tumors from mammograms.

The authors are with Dept. of Electrical Engineering, Jamia Milia Islamia, New Dehli, India
Email: zaheer_2k@hotmail.com, zajaffery@yahoo.com, laxman_mehlawat@yahoo.co.in

In last twenty years, several techniques [5] have been applied for tumor detection from various modalities, namely, X-ray mammography, CT-Scan and magnetic resonance imaging. Yu Sung et al. [6] investigated the performance of two stage method for the detection of clustered micro-calcifications (MCs) in digital mammograms using combined model-based and statistical textural features. Celia et al. [7] studied the behavior of an iris filter at different scales to detect masses on mammograms. Strickland et al. [8] applied the wavelet transform for detecting micro calcifications in mammograms. Shen et al. [9] developed a set of shape factors for the analysis of calcifications in mammograms.

Among the available techniques, region growing has been considered as an effective and most frequently used technique in image processing applications. Region growing method, initially introduced by Zucker, is based on a similarity measure between neighboring pixels and object of interest [10-11]. But the limitation of this method is that it produces a very poor segmentation results for the images representing variable intensity and requires lot of computation time. In region growing method, an important step is the selection of appropriate value of threshold. But, finding an optimal value of threshold by trial and error method or on the basis of histogram analysis is very error prone and time consuming. So, in this paper, a mean based region growing segmentation (MRGS) is presented that automatically finds the seed pixel and optimal threshold value. The proposed approach consists of three steps: (i) in the first step, an image quality is improved using appropriate noise removal and contrast enhancement techniques, (ii) in the second step, a "mean" criterion is used as a seed pixel and optimal threshold value after a proper formulation in the ordinary region growing (RG) method, (iii) Finally, in third step, morphological filtering operations are applied on the segmented images to improve the results by removing the unnecessary spots and noise. Then, MRGS method is compared with marker controlled watershed segmentation (MCWS) [12] approach.

2. Proposed Methodology

Image segmentation refers to the techniques of dividing an image into different regions. Basic region growing method is known to be the most effective tool for performing the quantitative analysis of anatomical structures in medical images. But, it does not lead to the accurate detection of an object, when directly applied on raw input images containing the noise and poor contrast. So, the proposed algorithm could be applied on the mammographic images effectively, when

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

the noise and other local irregularities are removed from the input images.

2.1. Image Smoothing and Contrast Enhancement

The major limitations of mammograms are low contrast, uneven illumination and presence of noise, because of which feature extraction and object detection in mammograms have become a challenging task. Image smoothing and contrast enhancement are two major steps of pre-processing. These steps are very essential in image segmentation to enlarge the intensity difference between object and background and to reduce the noise without destroying the important features of an image. In the present work, the noise and uneven illumination are removed using a weiner filter, a type of a linear filter. Then, we applied the histogram equalization method to enhance the image contrast that scales the gray level of each voxel by pre-defined factor. This improves the visualization effect of the original image and thus makes the object of interest more clearly visible as shown in Fig. 1(a), and Fig. 2 (a) respectively.

2.2. Mean based region growing segmentation (MRGS)

This algorithm starts with an approach of choosing a pixel called the seed pixel, as the first region pixel. Then, the neighboring pixel is compared with the seed pixel and is grouped into larger regions based on homogeneity criterion [11]. Candidate pixel is merged with the seed pixel if a criterion of homogeneity is satisfied. In other words, we can say, that the intensity of the candidate (neighboring) pixel should be similar to that of seed pixel to be grown into single region. In this work, 8-connectedness was used to segment the mass structures successfully. In the proposed method, appropriate values of seed pixel and threshold are computed automatically on the basis of mean and entropy.

If the gray level value of the neighboring pixel is within the deviation range of given threshold, then, it is labeled as foreground otherwise as a background pixel. The steps which illustrate the proposed approach can be explained here as follows:

- Firstly, input image is denoised using Weiner filter.
- Secondly, histogram equalization method is applied to enhance the denoised image.
- In next step, the appropriate value of seed pixel (S_f) and threshold (T) is selected using the Equation (1) and (2) respectively.

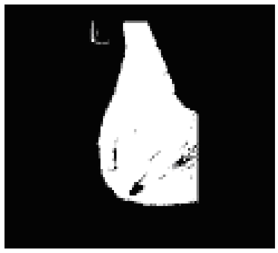
$$S_f = \frac{1}{M \times N} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x,y) \quad (1)$$

Where $f(x, y)$ is a grayscale image with co-odinates (x, y) of size M by N and

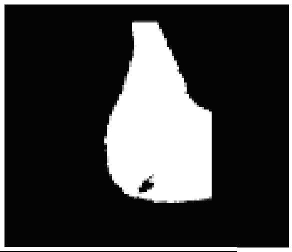
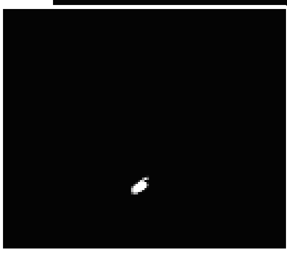
$$T = \frac{S - \beta * Entropy\{f(x,y)\}}{4} \quad (2)$$

The value of β depend upon the variability in the intensities of foreground and background. Here, its value is assumed between 0 and 1.

- Then, use the above found values of seed points and threshold in the basic region growing method to segment the image more precisely.
- Then, apply the opening and closing morphological operations to improve the result.

(a) **(b)**

(c) **(d)**

Fig. 1. (a) Original image obtained after filtering and contrast enhancement, (b) Tumor segmented by MRGS method, (c) spots removal with opening-closing morphological filtering operations, (d) tumor extracted after ROI

2.3. Filtering by Morphological Operations

. It plays a pivotal role in image analyses, machine vision and pattern recognition [13]. In this approach, set theory is basically used to identify the objects.

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

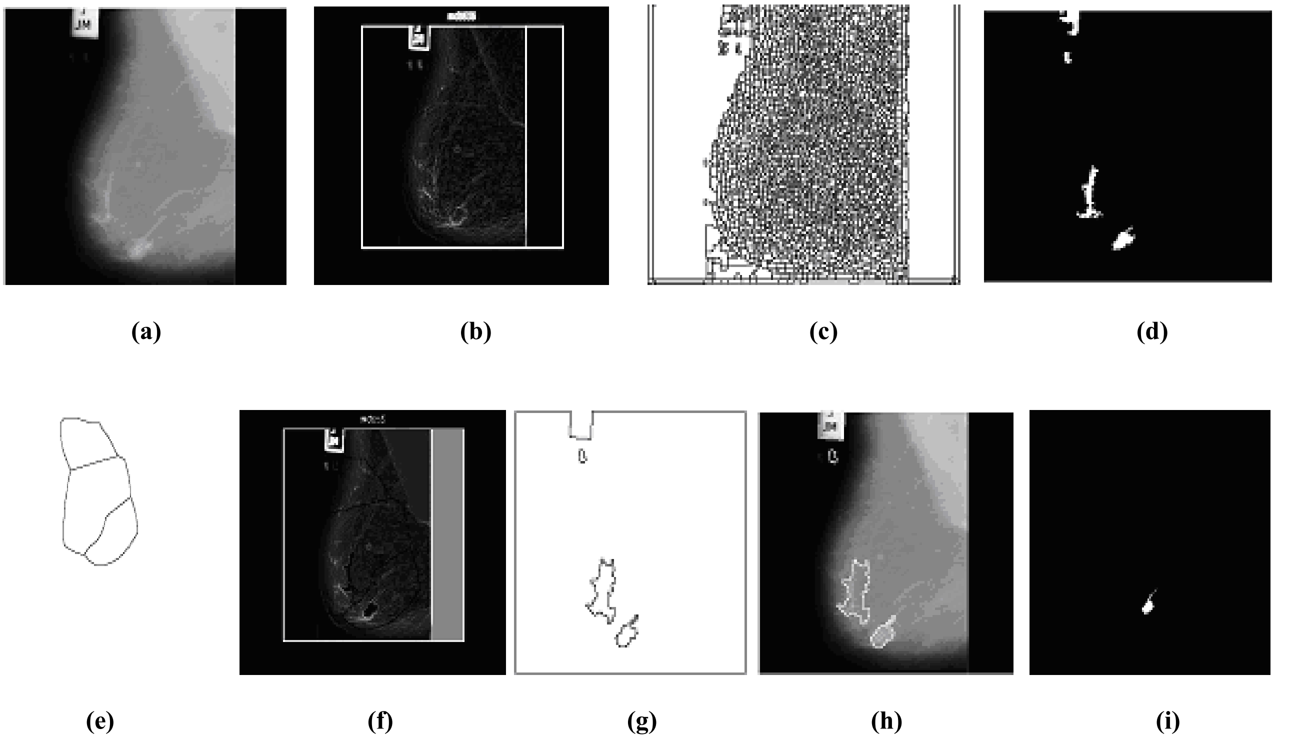


Fig. 2. (a) Original image obtained after filtering and contrast enhancement, (b) gradient image, (c) watershed of gradient image, (d) internal markers, (e) external markers, (f) markers modified gradient image, (g) watershed ridge lines, (h) final segmentation with watershed ridge lines, (i) extracted tumor image after ROI.

Dilation and erosion are the two fundamental operators of mathematical morphology. These two operators constitute two additional operations of opening and closing which are sensitive to the specific shapes embedded in the input image. Dilation adds up the pixels to the edges of the objects and thereby making the object boundaries thick, while erosion operator removes the disconnected pixels in the input image.

The dilation is written as $M \oplus N$ and for binary images [14], it is defined as

$$\boldsymbol{M} \oplus \boldsymbol{P} = \{ \boldsymbol{x} \in \boldsymbol{R}^N \mid \exists\, \boldsymbol{m} \in \boldsymbol{M} \;\&\; \boldsymbol{n} \in \boldsymbol{P} \text{ such that } \boldsymbol{x} = \boldsymbol{m} + \boldsymbol{n} \} \tag{3}$$

The erosion of M by P is defined as

$$\boldsymbol{M} \,\Theta\, \boldsymbol{P} = \{ \boldsymbol{x} \in \boldsymbol{R}^N \mid \exists\, \boldsymbol{m} \in \boldsymbol{M} \;\&\; \boldsymbol{n} \in \boldsymbol{P} \text{ such that } \boldsymbol{x} = \boldsymbol{m} - \boldsymbol{n} \} \tag{4}$$

The opening of image M by structuring element P is defined as

$$\boldsymbol{M} \circ \boldsymbol{P} = (\boldsymbol{M} \,\Theta\, \boldsymbol{P}) \oplus \boldsymbol{P} \tag{5}$$

The closing of image M by structuring element P is defined as

$$\boldsymbol{M} \bullet \boldsymbol{P} = (\boldsymbol{M} \oplus \boldsymbol{P}) \,\Theta\, \boldsymbol{P} \tag{6}$$

Where, M and P represents input image and the structuring elements respectively. The members of sets M and P define structuring element in an N - dimensional space, where value of N is 2 for binary images.

2.4. Marker controlled watershed segmentation (MCWS)

According to the definition of geography, a watershed line is defined as the line separating two catchment's basins, as shown in Fig. 3. The basic principle of watershed technique is to view the gradient of a gray level image as a topographic surface [12]. Each local minima embedded in an image is referred to as a catchments basins. The rain that falls on the either side of the watershed lines would be collected equally in catchments basins [15]. Morphological based watershed is known as the efficient tool for image segmention and video processing applications. But due to the presence of noise and several kinds of artifacts, it produces the over segmentation [5]. Noise and irregularities could be eliminated by using several denoising techniques, viz., weiner filter, median filter, Gaussian filter and morphological top hat filtering. Vincent and soille [16] proposed the novel approach for finding the watershed lines by using the immersion simulation algorithm.

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

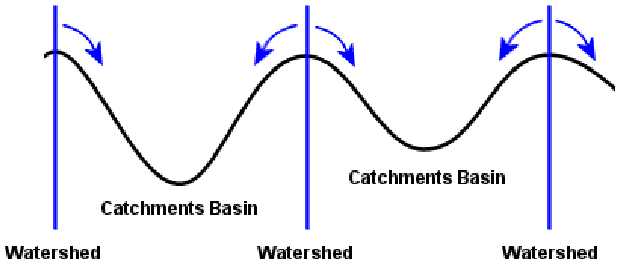


Fig. 3. Watershed lines with catchments basins

In watershed transformation, over segmentation problems occurs mainly because of large number of minima's embedded in input image. This problem can be minimized greatly if we introduce markers and flood the gradient image starting from these markers [12]. In the proposed work, the over segmentation problem of the simple watershed method is reduced by developing the marker's around only concerned deep minima's The threshold which controls the marker size can be selected either interactively or automatically.

The Algorithm

The gradient magnitude is basically used to highlight the object edges on gray-scale images. So in this method, before applying the watershed algorithm, the gradient magnitude of the gray-scale image is computed using the linear filtering method [5]. The gradient magnitude of the original image is shown in Fig. 2 (b). For any gray scale image f(x, y) with co-ordinates (x, y), the gradient vector magnitude and angle is computed using the equation (7) and (8). Here, the angle represents maximum rate of change of intensity level that occurs at the specified co-ordinates (x, y).

$$g(x, y) = \sqrt{(g1^2(x,y) + g2^2(x,y))} \quad (7)$$

$$\alpha(x, y) = \tan^{-1}(g2(x, y)/ g1(x, y)) \quad (8)$$

Where, g1(x, y) and g2(x, y) shows the values of gradients in the x and y direction. Magnitude of these gradients is computed using the sobel mask H1 and H2, which are illustrated by equation (9).

$$H_1 = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, H_2 = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (9)$$

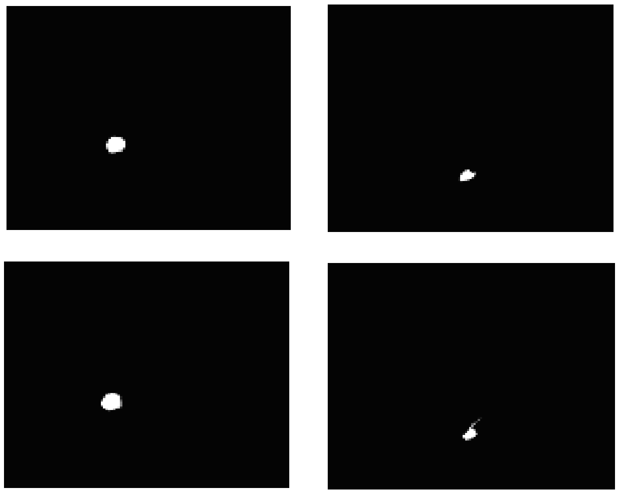


Fig. 4. Results of ROI (tumor) selection for sample no. 2, and 3 after applying (a) mean based region growing segmentation (top row), (b) marker controlled watershed segmentation (bottom row).

Watershed of simple gradient image produces too many watershed ridge lines that do not correspond to the region of interest. To sort out this problem, a novel approach based on the concept of markers is discussed in this paper and this approach is called Marker–Controlled Watershed Segmentation (MCWS). The steps which illustrate the MCWS approach can be explained here as follows:

- Firstly, pre-process the input image using appropriate filtering and image enhancement techniques.
- Compute the magnitude gradient of the pre-processed image.
- Compute the internal and external marker in order to identify all those regional minima's which have higher values than a specified threshold.
- Suppress all other minima's except the minima's we specify from the gradient image and obtain the modified gradient image.
- Compute the watershed of the modified gradient image in order to produce the watershed ridge lines.
- Superimpose the resulting watershed ridge lines on the original image to show the final segmentation result.
- Separate out the region of interest (ROI) from the segmented image.
- Compute the shape parameters of the ROI (i.e. basically tumor parameters).

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

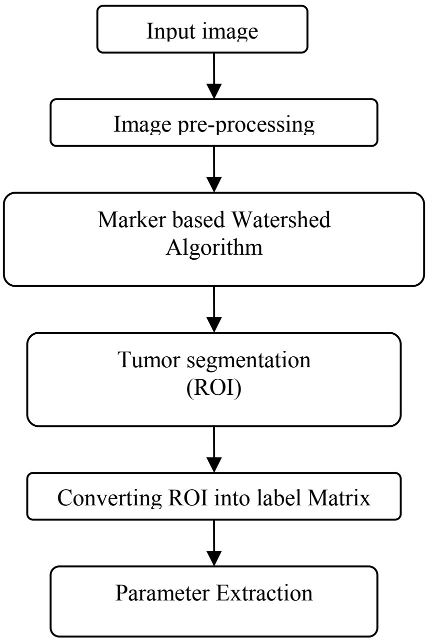


Fig. 5. Flow Chart of segmentation process by MCWS method

3. Results

In mammograms, masses are assumed to be distinctive regions that are relatively brighter than their surrounding tissues as shown in Fig. 1(a). In this paper, we investigated the performance of two novel approaches, viz., marker controlled watershed approach and mean based region growing approach. Result of extracted tumor by MRGS method is shown in Fig. 1(b).

Numerical result of MRGS method is displayed in Table 1. The morphological operation removes the small spots from the segmented image and thereby produces the better result as illustrated by Fig. 1 (c). Table 3 display the result of the proposed methods validated against visual inspection of expert radiologists.

The relative error (RE) of tumor area is computed as

$$RE\ (\%) = \left(\left| \frac{D - D^{'}}{D^{'}} \right| \right) \times 100 \qquad (10)$$

Where D is the tumor area measured by MRGS method and marker controlled watershed method and $D^{'}$ represents the tumor area as furnished by an experts.

Fig. 2 illustrates the various steps involved in the marker controlled watershed segmentation approach. The shape characteristics of extracted tumor obtained after applying marker controlled watershed approach are tabulated in table 2. Final segmentation results obtained after applying MRGS method and MCWS method for the sample images 2 and 3 are shown in Fig. 4.

4. Conclusion

In this paper, mean based region growing segmentation (MRGS) method is presented which has the improvement over ordinary region growing (RG) method with regard to the selection of threshold. The proposed approach reduces computational complexity of ordinary RG method and can be helpful in assisting the radiologist in performing the in-depth diagnosis of breast tumor at considerably reduced time (see table 4). Table 3 makes a comparison of numerical result of two methods namely; MRGS method and MCWS method with an expert radiologist data and shows that MRGS method performs better than the MCWS method. In the present study, experimental results shows that we have successfully achieved our goal of obtaining high segmentation accuracy and reducing computation cost by using the proposed method.

Table 1: Tumor characteristics obtained after applying MRGS method.

Sample no.	Area (mm^2)	Major Axis (mm)	Minor Axis (mm)	Eccentricity ($0<e<1$)	Solidity ($0<s<1$)	Perimeter (mm)
01	27.8	7.7	4.9	0.76	0.89	22.1
02	75.6	10.4	9.4	0.43	0.97	33.1
03	58	11.8	6.4	0.84	0.93	8.6

Table 2: Tumor characteristics obtained after MCWS method.

Sample no.	Area (mm^2)	Major Axis (mm)	Minor Axis (mm)	Eccentricity ($0<e<1$)	Solidity ($0<s<1$)	Perimeter (mm)
01	24.2	8.4	4.4	0.85	0.81	25.9
02	83.3	10.9	10	0.4	0.92	36.6
03	57.8	11.97	6.5	0.84	0.93	31.7

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Table 3: Comparison of different tumor area (extracted by MRGS and MCWS method) with an expert radiologist results.

Sample no.	Area (MRGS) (mm^2)	Area (MCWS) (mm^2)	Area (Expert radiologist) (mm^2)	Relative error (%) (MRGS)	Relative error (%) (MCWS)
01	27.8	24.2	26.5	4.9	8.7
02	75.6	83.3	77	1.8%	8.2%
03	58	57.8	61	4.9%	5.2%

Table 4: Computation cost of RG*, MRGS**, and MCWS* methods.

Sample no.	RG method Time (sec)	MRGS method Time (sec)	MCWS method Time (sec)
01	240	10	60
02	300	11	70
03	250	10	65

* In case of RG and MCWS methods, computation cost includes the manual threshold selection time and execution time of the input image.
** In case of MRGS method, computation cost includes automatic threshold selection time and execution time of the input image.

References

[1] American Cancer Society, "American Cancer Society: Breast cancer facts and figures 2005-2006", pp. 1-28, 2006.

[2] X. Zhou, R. Gordon, "Detection of early breast cancer: an overview and future prospects", *Crit. Rev. Biomed. Eng*., vol. 17, pp. 203-255, 1989.

[3] H.P. Chan, K. Doi, C.J. Vyborny, K.L. Lam, R.A. Schmidt, "Computer-aided detection of microcalcifications in mammograms-methodology and preliminary clinical study", *Invest. Radiol,* vol. 23, no. 9, pp. 664-671, 1988.

[4] H. D. Cheng, Xiaopeng Cai, Xiaowei Chen, Liming Hu, Xueling Lou,"Computer-aided detection and classification of microcalcification in mammograms: a survey", *Pattern Recognition,* vol. 36, pp. 2967-2991, 2003.

[5] Laxman Singh, R. B. Dubey, Z. A. Jaffery, Z. Zaheeruddin, "Segmentation and Characterization of Brain Tumor from MR Images", *2009 IEEE Int. Conf., Computer Society*, 2009, pp. 815-819.

[6] Yu Sung-Nien, Huang Yu-Kun, "Detection of micro-calcification in digital mammograms using combined model-based and statistical textural features", *Expert Systems with Applications,* vol. 37, pp. 5461-5469, 2010.

[7] Varela Celia, G.Tahoces Pablo, J. Mendez Arturo, Souto Miguel, J. Vidal Juan, "Computerized detection of breast masses in digitized mammograms", *Computers in Biology and Medicine,* vol. 37, pp. 214-226, 2007.

[8] Robin N Strickland, "Wavelet Transform for Detecting Microcalcifications in Mammograms", *IEEE Transactions on Medical Imaging,* vol-15, no.2, pp. 218-229, April 1996.

[9] L. Shen, M. Rangayyan Rangaraj, and Desautls Leo J.E, "Aplication of Shape Analysis to mammographic Calcifications", *IEEE Transactions of Medical Imag.,* vol. 13, no. 2, june 1994, pp. 263-274.

[10] S. W. Zucker, "Region growing: childhood and adolescence", *Computer Graphics Image Processing,* vol. 5, pp. 382-399, 1976.

[11] R. Adams and L. Bischof, "Seeded Region Growing", *IEEE Transactions on Pattern Analysis and Machine Intellegence,* vo. 16, no. 6, June 1994.

[12] Po-Whei Huang, Yan-Hao Lai. "Effective segmentation and classification for HCC biopsy images", *Pattern Recognition*, vol. 43, no. 4, pp. 1550-1563, 2010.

[13] Pawael Tadejko, Waldemar Rakowski, "Mathematical morphology based ECG feature extraction for the purpose of heartbeat classification", *in proc. IEEE Int. Conf., Computer Society, 2007,* pp. 322-327.

[14] Stephen S. Wilson, "Theory of Matrix Morphology", *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 14, no.6, pp. 636-652, June 1992.

[15] M. Veta, A. Huisman, M.A. Viergever, P.J.Diest, J.P.W. Pluim, "Marker Controlled Watershed Segmentation of Nuclei in H&E Stained Breast Cancer Biopsy Images", *IEEE Int. Symposium on Biomed. Imag.,* 2011, pp. 618-621.

[16] L.Vincent and P.Soille, "Watersheds in Digital Spaces: An Efficient Algorithm based on Immersion Simulations," *IEEE Trans. Pattern and Machine Intell.,* vol.13, pp.583-598, June 1991.

ISBN: 978-988-19252-1-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)