

Development algorithm- computer program of digital mammograms Segmentation for detection of masses breast using Marker-Controlled Watershed in MATLAB environment

تطوير خوارزمية برنامج حاسوبي لمعالجة صور الأشعة السينية الرقمية من أجل الكشف وتشخيص الكتل غير طبيعية لسرطان الثدي باستخدام برنامج الماتلاب

Dhirgaam A. Kadhim

Collage of Education for Pure Science/ Kerbala University

dhirgaam_lamy@uokerbala.edu.iq

Abstract

Breast cancer detection images are the standard clinical practice for the diagnosis and prognosis of breast cancer. Digital Mammogram has emerged as the most popular screening technique for early detection of Breast Cancer and other abnormalities. Raw digital mammograms are medical images that are difficult to interpret so we need to develop Computer Aided Diagnosis (CAD) systems that will improve detection of abnormalities in mammogram images. In this paper we present a fully algorithm for detection of abnormal masses by anatomical segmentation of Breast Region of Interest (ROI). We are using medio-lateral oblique (MLO) view of mammograms. We have proposed method Marker-Controlled Watershed Segmentation Algorithm (MCWSA). The algorithms proposed are fully autonomous, and are able to isolate and abnormal regions in the breast tissue, If any abnormalities are present it gets accurately highlighted by this algorithm thus helping the radiologists to further investigate these regions, a task very few existing mammogram segmentation algorithms can claim.

Keywords: Digital Mammogram, Breast cancer detection, Marker-Controlled Watershed Segmentation Algorithm (MCWSA), Computer program algorithm

الخلاصة

تعتبر صور كشف سرطان الثدي من التطبيقات السريرية القياسية لتشخيص سرطان الثدي. ان صور الاشعة السينية الرقمية تستخدم كتقنية للفحص الأكثر شعبيةً وللكشف المبكر عن سرطان الثدي والحالات الغير طبيعية الأخرى للمرض. ان صور الاشعة السينية الرقمية الطبية تعتبر صعبة الترجمة للكشف عن المرض لذا نحتاج برنامج حاسوبي مساعد لتحسين تشخيص الحالات غير الطبيعية لسرطان الثدي (كاد) وذلك من خلال معالجة صور الاشعة السينية. في هذا البحث قدمنا خوارزمية كاملة للكشف عن النقاط غير الطبيعية وذلك بتقطيع اجزاء الثدي (روي). استعملنا مقاطع مائل جانبي لصورة الاشعة السينية. إقترحنا طريقة خوارزمية إنقسام حد فاصل مدارة من قبل علامة (إم سي دبليو إس أي). إن الخوارزميات التي إقترحت مستقلة ذاتياً بالكامل، وقادرة على عزل ومناطق غير طبيعية في نسيج الثدي، اي ان حالات تصبح أبرزت بدقة بهذه الخوارزمية بحيث تُساعد أخصائيي الأشعة السينية لتحري هذه المناطق بصورة افضل.

1. INTRODUCTION

Globally, breast cancer is the most common cancer among women, comprising 23% of the 1.1 million female cancers that are newly diagnosed each year [1,2]. It is also the leading cause of cancer-related deaths worldwide, case fatality rates being highest in low resource countries [3]. Approximately 4.4 million women diagnosed with breast cancer in the last 5 years are still alive, making breast cancer the most prevalent cancer worldwide [1].

In Iraq, breast cancer is the commonest type of female malignancy, accounting for approximately one-third of the registered female cancers according to the latest Iraqi Cancer Registry [4]. This shows that the breast is the leading cancer site among the Iraqi population in general, surpassing even bronchogenic cancer.

As proposed by the World Health Organization, early detection and screening, especially when combined with adequate therapy, offer the most immediate hope for a reduction in breast cancer mortality [5]. This was the basis of the Iraqi national program for early detection of breast cancer, which was initiated in 2001 in an attempt to downstage this disease at the time of presentation. Since then specialized centers and clinics for early detection of breast tumors have been established in the major hospitals in all Iraqi provinces.

Digital mammography refers to the application of digital system techniques on digital mammograms. Digital systems have the capacity to bring revolutionary advantages to breast cancer detection [6]. Radiologists turn to digital mammography for an alternative diagnostic method due to the problems created by conventional screening programs. An automated system can overcome these problems by reducing the number of false positive and false negative readings from radiologists and increase the chance of detecting abnormalities early. Up to this moment mass detection from digitized mammograms is still very challenging. The main reason is that the abnormal masses usually mix with the homogeneous tissues in the breast. The images provided by different patients have different dynamics of intensity and present a weak contrast. Moreover the size of the significant details can be very small. Several research works have tried to develop computer aided diagnosis (CAD) tools that could help the radiologists in the interpretation of the mammograms and could be useful for an accurate diagnosis [7] [8].

2. MATERIALS AND METHODS

This paper presents the visual and statistical results of applying Computer Aided Diagnosis (CAD) System Based on Segmentation by using Marker Controlled Watershed Transform Algorithm for Detection of Masses in breast over the simulated and real image of mammogram breast image. In this study the original image works with gray level values, the original image is shown in Figure 1:

Surface image

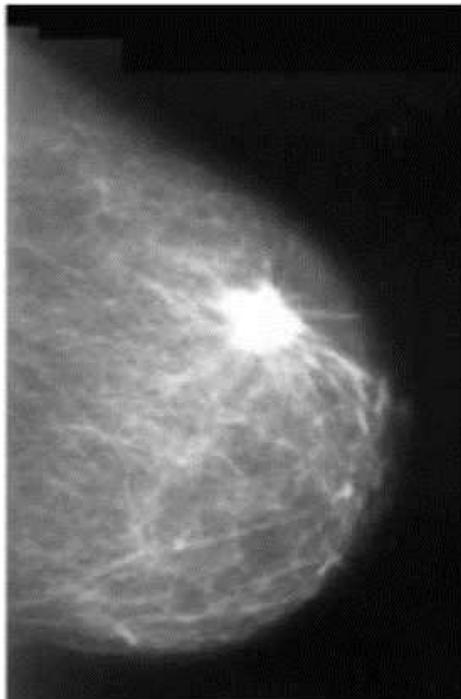


Figure 1: The original image

We used Matlab software to implement the algorithm because Matlab is a high-performance language for education and research as it integrates computation, visualization and programming in an easy-to-use environment where problems and solutions are expressed in familiar mathematical

notation and also it has toolboxes for signal processing, neural network, image processing, database ... etc., Matlab Image Processing Toolbox is a collection of functions that extend the capability of the Matlab numeric computing environment. The toolbox supports a wide range of image processing operations, such as Image analysis and enhancement. Region of interest operations, linear filtering and filter design all the functions and equations used in this study are from Matlab image processing toolbox (Matlab software).

2.1 Medical Imaging

Digital mammograms are medical images requires a preparation phase in order to improve the image The objective of medical image analysis is to acquire useful information about the physiological processes or organs of the body by using external and internal sources of energy. Breast image analysis can be performed using X-rays, magnetic resonance, nuclear medicine or ultrasound- So far the most effective and economical breast imaging modality has been X-ray mammography due to its simplicity, portability and cost effectiveness, an important source of radiological information for breast imaging is the presence and distribution of micro calcifications in the breast, this anatomical information can be obtained with high resolution technology using X-rays [9]. As yet there is no comprehensive imaging modality for all radiological applications and needs, although the ability to computerize and analyze medical images provides a powerful means to assist physicians; thus computer programs, processing methods that get the data and information from medical imaging scanners must be carefully developed to preserve and enhance the most important clinical information rather than introducing additional artifacts. The ability to improve diagnostic information from medical images can be further enhanced by designing computer processing algorithms, application and software intelligently, that is why we proposed an application based on matlab software to detect cancer in mammogram breast cancer images.

2.2 Marker Based Watershed Segmentation Method

According to the definition of geography, a watershed is the ridge that divides areas drained by different river systems and catchments basins is the geographical area draining into the river. The basic principle of watershed technique is to transform the gradient of a grey level image in a topographic surface, where the values of $f(x, y)$ are interpreted as heights and each local minima embedded in an image is referred as a catchments basins. If we imagine rain falling on the defined topographical surface, then water would be collected equally in all the catchments basins. The watershed transformation can be built up by flooding process on a gray tone image and may be illustrated by Figure 2.

The basic watershed algorithm is well recognized as an efficient morphological segmentation tool which has been used in a variety of gray scale image processes & video processing applications. However, a major problem with the watershed transformation is that it produces a large number of segmented regions in the image around each local minima embedded in the image.

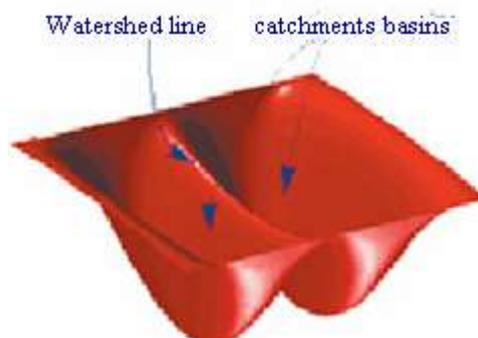


Figure 2. Watershed segmentation simplified to 2 Dimensions.

Over segmentation problem in the morphological watershed segmentation for irregular-shaped objects is usually caused by spurious minima in the inverse distance transform. Figure 3 shows the over segmentation result produced by the basic watershed segmentation method.

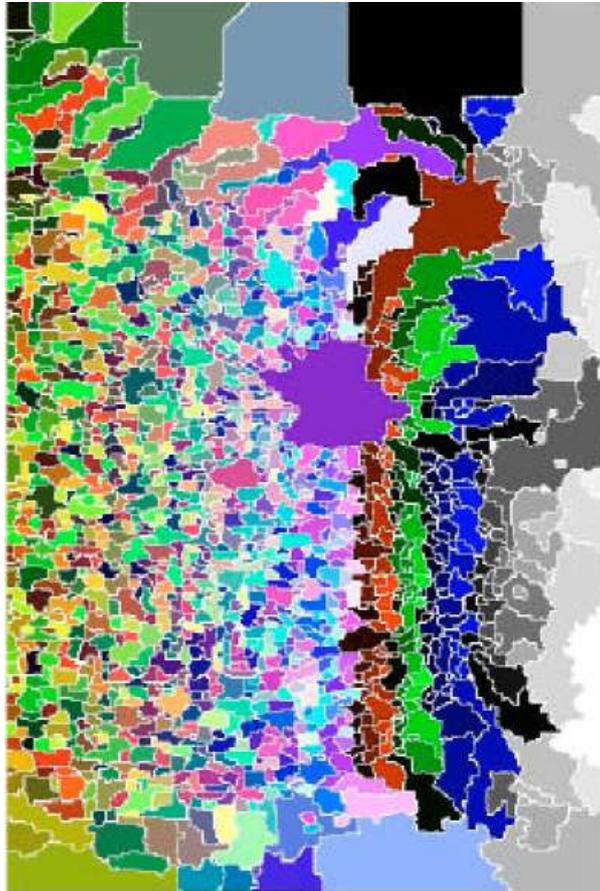
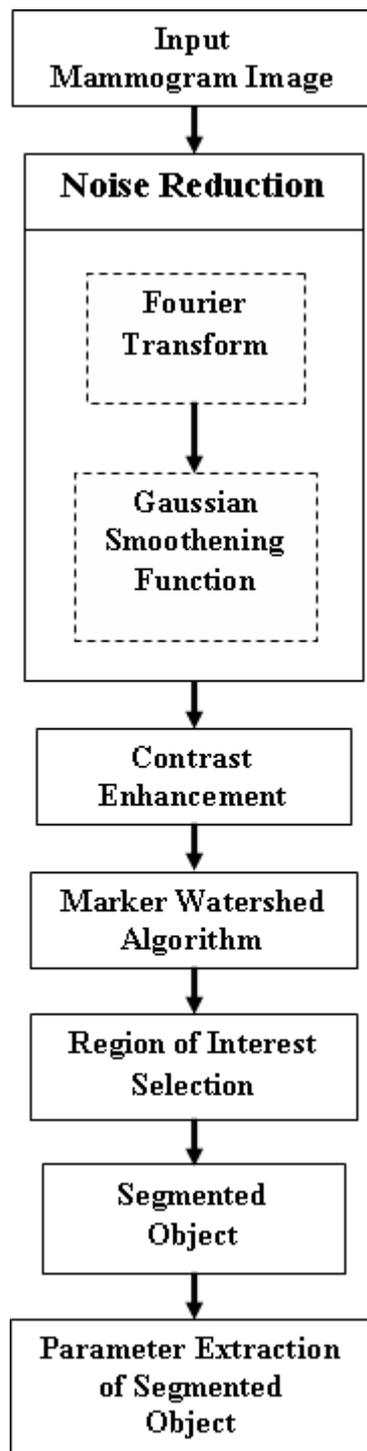


Figure 3. The over segmentation result produced by the basic watershed segmentation method.

In this, different catchment basin is represented by different colors and watershed ridgelines are shown by white lines. A solution to sort out this problem is to introduce markers and flood the gradient image starting from these markers instead of regional minima. The Marker based Watershed Segmentation method possesses several important properties that makes it highly usable for various kinds of image segmentation problems. Implementation of this method involves various processing steps which can be arranged in a meaningful manner that is shown in block diagram 1.



Block diagram 1. Overview of algorithm marker based watershed segmentation process

2.3 Image Pre-Processing

In this process, the image is converted into the accessible form. Image smoothing act as the pre-processing step for image segmentation, as almost all of the images suffers from the problem of noise effect, uneven illumination and local irregularities. In this present work, the noise and uneven illumination are filtered out from the mammogram image using Gaussian filter which is a type of a linear filter. The contrast of the smoothed image is enhanced using the image processing toolbox

functions available in Matlab 7.9. This improves the visualization effect of the original image. The `imadjust` function is used for this purpose. It has the following syntax

$$\mathbf{J} = \text{imadjust}(\mathbf{I}, [\text{low_in}; \text{high_in}], [\text{low_out}; \text{high_out}]);$$

This maps the values in intensity image I to new values in J such that values between `low_in` and `high_in` maps to values between `low_out` and `high_out`. After the intensity adjustment the image format is changed from 'indexed' to 'intensity' image into the 'binary' image. `Im2bw` function is used to do this process. It converts the input image to gray scale format, and then uses `threshold` to convert this gray scale image to binary. After this the complement of the binary image is taken to obtain the distant transform. Then the `watershed` function is applied on the complement of distance transform image. But `watershed` of negative distant transform leads to the serious over segmentation. Thus, this method is not preferred for the segmentation of breast mammogram image. So, after filtering and contrast enhancement, the image is first converted into double format to make it suitable for computing the gradient magnitude. The double format image is used as an input for the segmentation process.

2.3 Marker-Controlled Watershed Segmentation

Matlab programming is used to develop algorithm for marker controlled Watershed segmentation. The gradient magnitude is used often to preprocess a gray-scale image prior to using the watershed transform for segmentation which is computed using the linear filtering method [10]. For any gray scale image (x,y) , at co-ordinates (x,y) , the gradient vector magnitude and angle at which maximum rate of change of intensity level occurs at the specified co-ordinates (x,y) can be computed using the equation (1) and equation (2).

$$g(x, y) = \sqrt{(g_1^2(x, y) + g_2^2(x, y))} \quad (1)$$

$$\alpha(x, y) = \tan^{-1}(g_1(x, y)/g_2(x, y)) \quad (2)$$

Where $g_1(x, y)$ and $g_2(x, y)$ are the gradients in the x and y directions. Magnitude of these gradients is computed using the sobel mask H_1 and H_2 , which are defined by equation (3) given below:

$$H_1 = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad H_2 = \begin{bmatrix} -1 & -2 & 1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (3)$$

Watershed of simple gradient image does not produce good segmentation result [11]. There are too many watershed ridge lines that do not correspond to the object in which we are interested. To overcome such a difficulty, a new approach based on the concept of markers is introduced in this project and this approach is called Marker-Controlled Watershed Segmentation.

In the present method, the internal markers are produced from the gray scale image and then external markers are found by finding pixels that are exactly midway between the internal markers. This is done by computing the watershed transform of the distance transformed image of the internal markers. The gradient image is then modified by imposing regional minima at the location of both the internal and external markers. The next step involves the computation of the watershed transformation of the Marker modified gradient image to produce watershed ridge lines. Finally resulting watershed ridge lines are superimposed on the original image and produce the final segmentation shown in Figure 1 and then the output image produced as a result of Region of Interest (ROI) selection is converted into label matrix. All the parameters of an object are then extracted using the image processing toolbox functions.

3. EXPERIMENTAL RESULT EVALUATION

Results obtained by applying the proposed algorithms mammogram breast image, that's Take it from Iraqi national program for early detection of breast cancer This image represents the control image for our experiment indicating that mass cancer is detected.

Gaussian filter is used to remove the noise and detail form the original mammogram image, the Gaussian filter is very effective in the reduction of impulse and Gaussian noise as shown in figure 5. Figure 6. show the segmentation process that performed on the edge map differentiates various regions on the breast depending on their intensity values. Each region has a different intensity value. The fatty tissues, glands, lobules and the ducts display different intensity values and thus can be segregated into different regions. An abnormality such as a mass, tumors or calcifications may be present within the breast has distinctly higher intensity values than the normal tissues of the breast.

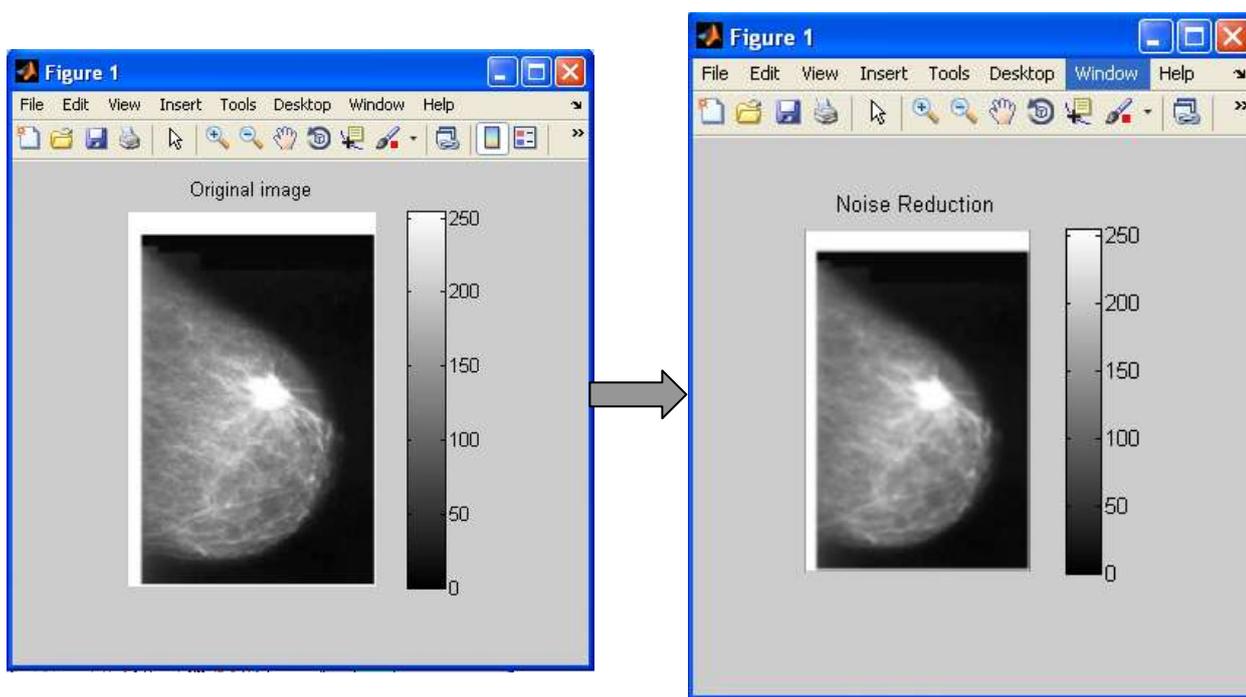


Figure 5. Original mammogram along with the image on which the noise is removed by the Gaussian filter.

So in this is figure we need to categorize all the obtained closed structures to their intensity values. The distribution of pixels intensities also vary within each segmented region but the majority of the pixels have similar intensity values. So for each region we calculate the arithmetic Mode value for the intensities from the original mammogram and replace those pixels in the region with the computed mode values.

The final segmentation result obtained by proposed Marker based Watershed algorithm is shown in Figure 7(a) and Figure 7(b) shows the ROI of a complete tumor.

Results obtained by applying the proposed algorithms on Digital mammogram image comprised predominantly Fatty tissues where abnormalities are present.

The accuracy of this technique “Marker–Controlled Watershed Segmentation.” was evaluated through quantitative measures derived through the comparison of each segmented mammogram representing an abnormal mass, as “mask” with its corresponding “gold standard”. The gold standard is generated by manually segmenting the breast region representing an abnormal mass from each mammogram. The boundary of the regions is then manually traced to extract the abnormal breast region to generate a ground truth (GT) image.

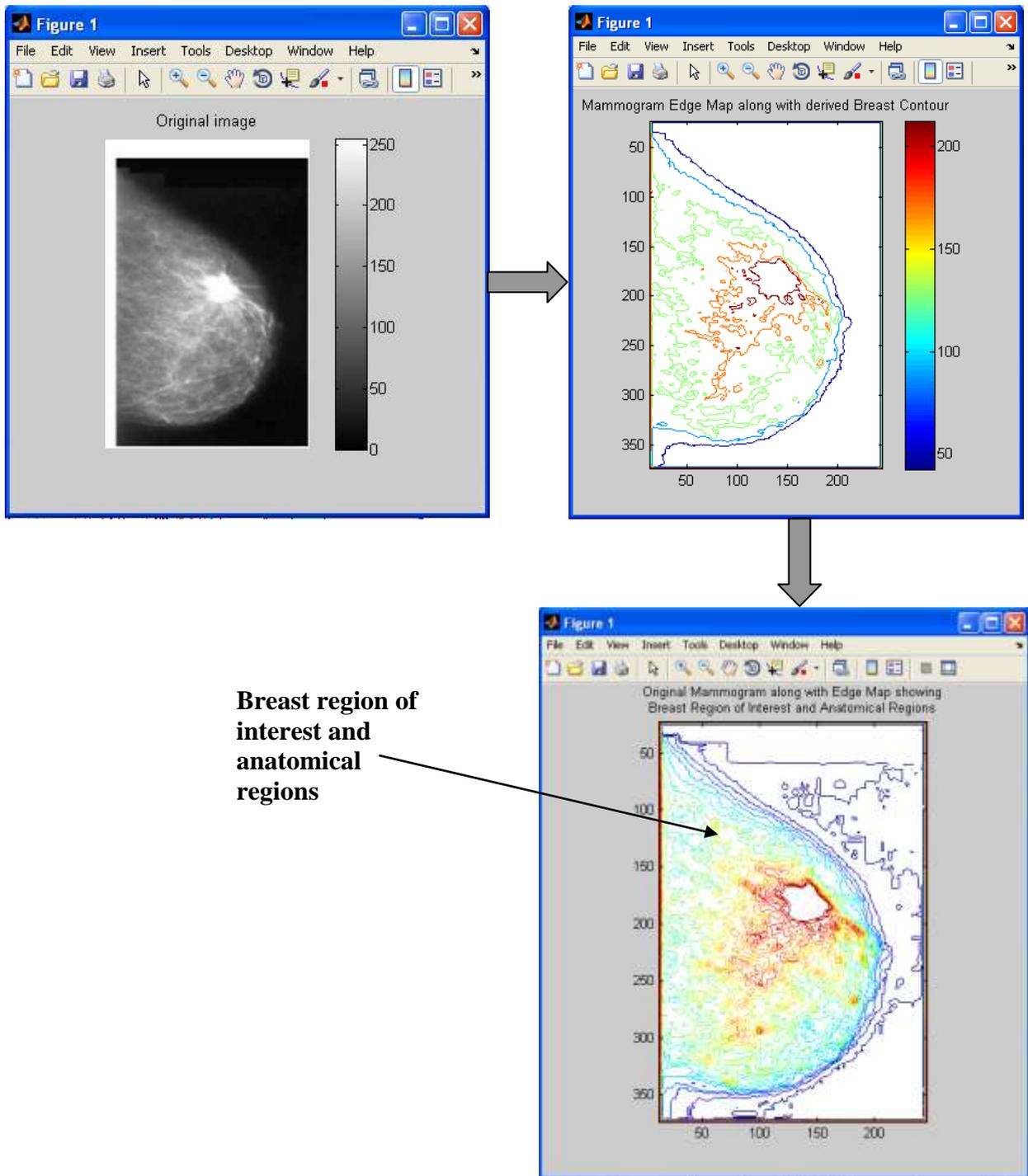


Figure 6. Original mammogram along with edge map derived breast contour and showing breast region of interest and anatomical regions.

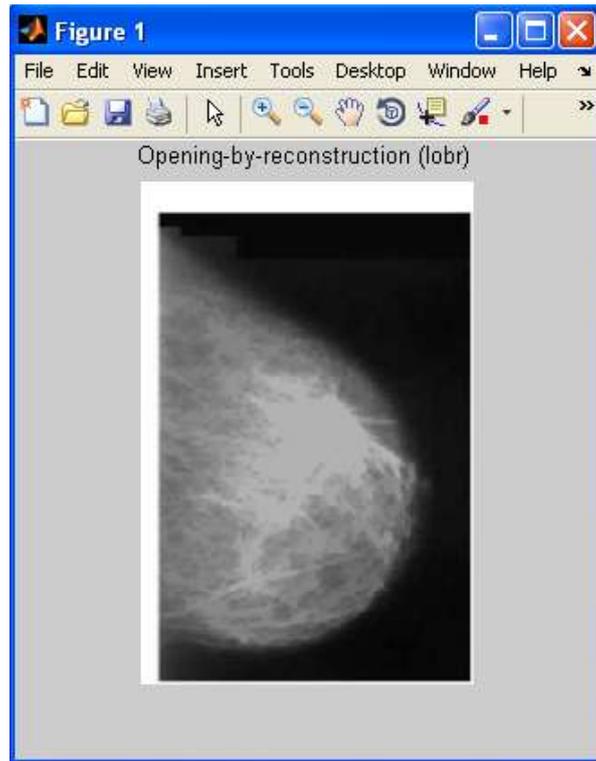


Figure 7(a) GT of Abnormal Region excluding insignificantly mark the Foreground abnormal regions.

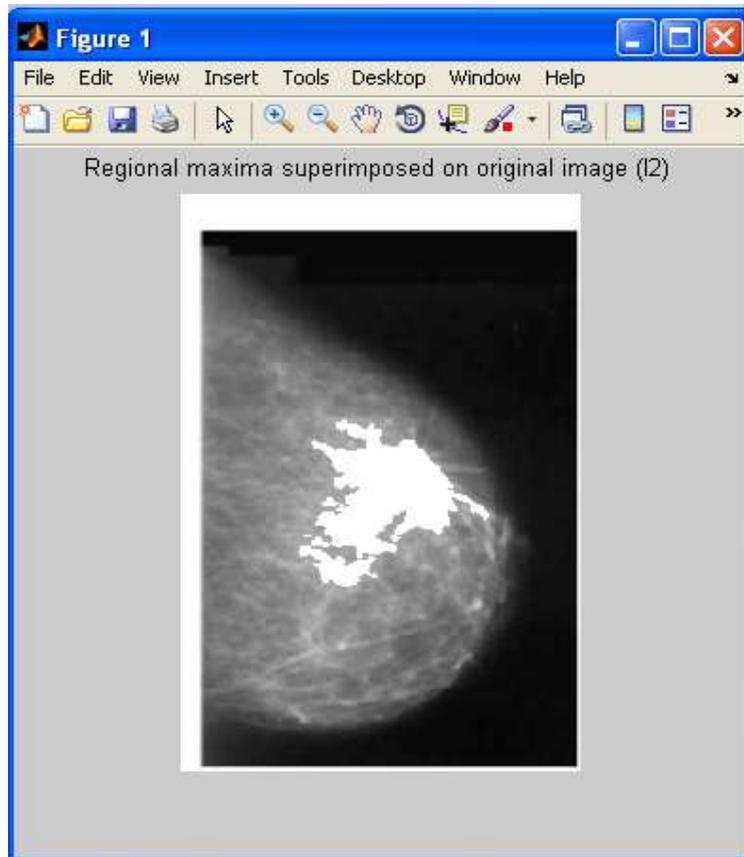


Figure 7(a) GT of Abnormal Region excluding insignificantly large abnormal regions.

4. CONCLUSIONS

The developed algorithm is used to know about the location and size of the of mass in mammograms having abnormal regions.. This method uses the user defined input parameters such as threshold for the analytical calculations. The optimal value of the threshold is highly dependent on shape and location of the tumor as well as on the different views of the images. The accurate mass detection was achieved by the proposed algorithm Marker–Controlled Watershed Segmentation.

The result obtained shows that the proposed algorithm has a higher computational efficiency along with the good performance for image segmentation problems as compared to the conventional tool. The algorithm used in this work is marker dependent which in turn depends upon the selected value of threshold. In this work, the optimal value of threshold is selected interactively. But this algorithm can be made more efficient and fast by providing the automatic value of threshold. This makes it faster and robust. In future, this work can be extended to classify the tumors using any of the techniques such as Fuzzy Expert Systems, Neural Network System, ANFIS and Hidden Markov Model.

REFERENCES

- [1] Parkin DM et al. Global cancer statistics 2002, CA: A Cancer Journal for Clinicians, 2005, 55:74–108.
- [2] Parkin DM, Fernandez LM. Use of statistics to assess the global burden of breast cancer. *Breast*, 2006, 12(1 Suppl.):S70–S80.
- [3] Anderson BO et al. Guideline implementation for breast healthcare in low-income and middle-income countries. Overview of the Breast Health Global Initiative Global Summit, 2007. *Cancer*, 2008, 113(8 Suppl.):2221–2243.
- [4] Iraqi Cancer Board. Results of the Iraqi Cancer Registry 2004. Baghdad, Iraqi Cancer Registry Center, Ministry of Health, 2007.
- [5] National Cancer Control Programs. Policies and managerial guidelines, 2nd. ed. Geneva, World Health Organization, 2002.
- [6] Pisano and F. Shtern, “Image processing and computer-aided diagnosis in digital mammography,” in *State of the Art of Digital Mammographic Image Analysis*. Singapore: World Scientific, 1994, vol. 7, pp. 280–291.
- [7] Highnam R and Brady M, “Mammographic Image Analysis”, Kluwer Academic Publishers, 1999. ISBN: 0-7923- 5620-9.
- [8] Kekre HB, Sarode Tanuja K and Gharge Saylee M, "Tumor Detection in Mammography Images using Vector Quantization Technique", *International Journal of Intelligent Information Technology Application*, 2009, 2(5):237-242
- [9] Dhawan P, A. *Medical Imaging Analysis*. Hoboken, NJ: Wiley-Interscience Publication, (2003).
- [10] Fu, K.S. & Mui, J.K. (1981). A Survey of Image Segmentation. *Pattern Recog.*, 13(1), 3-16. [http://dx.doi.org/10.1016/0031-3203\(81\)90028-5](http://dx.doi.org/10.1016/0031-3203(81)90028-5).
- [11] L. Vincent & P. Soille. (1991). Watersheds in Digital Spaces: An Efficient Algorithm based on Immersion Simulations. *IEEE Trans. Pattern and Machine Intell.*, 13, 583-598. <http://dx.doi.org/10.1109/34.87344>.