Instantly Enhanced Image Selection for Mammogram Images Segmentation

Sura Jasim Mohammed¹, Dr. Thekra HayderAli Abbas²

Computer Science Department, College of Sciences, Mustansiriyah University, Baghdad, Iraq¹²

surags.dcc@gmail.com¹, thekra.abbas@yahoo.com²

Abstract

The proposed system is to design a CAD (Computer-aided diagnosis) system that would achieve higher precision of breast cancer masses diagnosis through image preprocessing step. Hybrid approach is proposed to filter and enhance mammogram images. Tow strong enhancement techniques are used, they are “contract limit adaptive histogram equalization” (CLAHE) and “Dynamic histogram equalization” DHE. When there are more than one technique nearly similar in effectiveness are applied on an image and their efficiency is varied at each enhancing case, the instant selection between those resulted images is the best-proposed way to passing the better image to the next step in the classification system; that is based on mathematical calculations. in order to achieve best results in terms of mean square error (MSE) and peak signal to noise ratio (PSNR), these calculations require an instantaneous collaboration between the implementing program (Matlab) and the memory that used for storing and comparing the calculated values. Instantly image selection provides an effective way to adopt the best enhanced image between the outputs of the applied techniques to be passing to segmentation stage. Seeded region growing (SRG) segmentation method and morphological operations are proposed for the purpose of eliminate labels and pectoral muscle part from the image. Extracting the suspected regions from the breast area is done using graph cut segmentation method.

Keywords: contrast limited adaptive histogram equalization, Filtering, Graph cut technique, enhancement and segmentation

اختيار الصورة المحسنة آنياً لتجزئة صور الماموكرام

سرى جاسم محمد¹، د. ذكرى حيدر علي عباس²

الملخص

النظام المقترح هو تصميم نظام (التشخيص بمساعدة الكمبيوتر) من شأنه تحقيق دقة عالية لتشخيص كتل سرطان الثدي خلال مرحلة المعالجة المسبيقة للصورة. اقترح أسلوب هجين لتحسين صور الماموكرام، تم استخدام اثنين من تقنيات التحسين القوية، هما "معادلة المدرج التكراري" و "معادلة المدرج التكراري الديناميكية". عند تطبيق أكثر من تقنية متشابهة تقييماً بالفعالية على صوره وكفاءة هذه التقنيات تختلف في كل حالة تحسين صوره. فأن الانتقاء اللحظي (الأني) بين تلك الصور
The outcome is the best approach suggested for transmitting the best image to the next step in the classification system, based on mathematical computations.

In order to achieve better results, the average mean square error and peak signal-to-noise ratio were calculated. These calculations require immediate cooperation between the running program (MATLAB) and the used memory for storing and comparing calculated values. Choosing the image immediately provides an effective method for choosing the best improved image among the outputs of the applied techniques to be transmitted to the segmentation stage. A hybrid approach was suggested to classify the tissue areas and morphological processes for removing the marks and the soft part of the image. The suspected areas were extracted from the breast area using the segmentation method, which was applied to the actual images. The experimental results indicate that the breast area may have a specific value for correct classification.

1. Introduction
Breast cancer compared to other types of cancer that affect women is considered to have the highest occurrences and mortality rate in several regions of the world [1]. Recent statistical analysis shows that breast cancer impacts one of every ten women in Europe and one of every eight in the US [2]. Education of the severity of cancer and early diagnosis of cancer at early stages combined with appropriate treatment have shown to be the most efficient approaches minimize its mortality rate. Currently, the first technique to detect abnormalities in breast tissue is mammography. With the current technology the variation in tissue’s density can provide some statistical errors in the diagnosis of cancer, false negatives results can arise with especially in higher density breast tissue. Incorporating the knowledge of breast density may also be helpful for setting up a suitable threshold in automatic detection algorithms. Three kinds of tumor classification can be identified with mammography: A typical tumor is a tissue mass happen at the healthy tissue. A benign tumor is a collection of harmless cells, with identified borders and can’t spread to other parts, could be removed completely by surgery with low chances to regrow again. A malignant tumor does not have distinguished border, with the ability to grow rapidly increasing the pressure through the cells of breast and can be propagate. In addition, due to their rapid increase in numbers they could pose a serious health problem if ignored [1]. Image processing method is implemented for extraction and analyzing the region of interest (ROI). Correct classification of tumors into normal, malignant, or benign isn’t a critical task and should be done after an application of an appropriate segmentation method. Because of the sensitivity of the segmentation stage in the general classification process, In this paper, a good approach presented which is hybrid algorithm techniques to enhance the image intensity to make detecting the breast tumors clear. The proposed algorithm includes finding MSE and PSNR values after filtering and enhancing an image using two different filtering and two enhancement methods, then a comparison on the two resulted enhanced images is accomplished and the one with the least MSE and highest PSNR is selected instantly in order to be segmented. We implemented a novel method based on mathematical approach for automatic segmentation of mammogram images. Theoretical background of filtering,
enhancement and segmentation methods implemented in this paper are discussed in section 2; Section 3 represents the experimental activities. Then, a conclusion and future work are presented in Section 4.

2. Theoretical Background

There are number different methods have been developed for denoising, enhancing and segmenting a mammogram image for the purpose of having a clear ROIs that could be used to extract important features for classifying it into their classes (normal, benign, and malignant).

2.1 Image Filtering

For removing digitization noise many of filtering methods can be use in this process of preprocessing the medical images. Mean filter working on the adjoining pixels’ brightness (intensity) and the mean calculated for those pixels; which then will be considered as the intensity of the new pixel [3]. Another common filtering technique is median filter; it is simple, hypothesized and easy to utilize approach of smoothing images, i.e. minimizing the value of intensity variation between neighboring pixels[4].

2.2 Image Enhancement

After removing noise; the second important step for increasing image contrast is utilizing the better enhancement techniques such that histogram equalization methods.

**Histogram Equalization (HE)** is a scheme operates by flattening the input image histogram, while the dynamic range of gray level will be extended by utilizing cumulative density function (CDF) of the image. A Histogram appears to the gray levels their relative frequency to maintain mean brightness of the input image [5]. The technique equalization of the histogram operation is as follows:

For a provided image \( X = X(jj, ii) \) composed of \( l \) discrete gray levels expressed as \( X_1, X_2, ..., X_{l-1} \). For image \( X \) probability density function \( P(X_k) \) is defined as in equation (1):

\[
P(X_k) = \frac{M^k}{M} \quad (1)
\]

Where \( M^k \) expresses number of times \( X_k \) presents in input image \( X \) and \( M \) is the entire number of samples in input image. Here, \( P(X_k) \) is the input image histogram which acts the number of pixels having a particular intensity \( X_k \). A graph of \( M^k \) vs. \( X_k \) denotes the X histogram. On the bases of PDF, the cumulative density function (CDF) is expressed as in equation (2):

\[
c(x) = \sum_{i=0}^{k} P(X_i) \quad (2)
\]

\( X_k = x, \) for \( k = 1, 2, ..., l - 1 \). Note, \( c(X_l, 1) = 1 \) by definition. HE is a scheme which plots input image into the whole dynamic range by utilizing (CDF) as a transformer function [5]. The technique of HE re-maps the input image gray levels by re-assigning intensity amount of pixels to create a unified intensity distribution.

Because the Histogram Equalization helps to increase the contrast globally, it is useful for applications of medical images especially when the background and the
interested area are presented by the closest contrast value. The histogram equalization is performed by using equation (3).

\[ H(V) = \text{round} \left( \frac{CDF(V) - CDF_{\text{min}}}{(M \times N) - CDF_{\text{min}}} \right) \times (L - 1) \quad (3) \]

Where the \( CDF_{\text{min}} \) is minimum value of cumulative frequency distribution and \( M \times N \) is image size and \( L \) is gray level [4].

**Adaptive Histogram Equalization (AHE)** is an expansion to the conventional Histogram Equalization technique where the image’s contrast is being enhanced by the value’s transformation in the concentrated image. Mainly, AHE regulates the local contrast and bring pure details. It fabricates over-magnifying also amplifying noise in a homogenous section of an image. For coping this issue, AHE’s advanced version is found which is named as **contrast limited adaptive histogram equalization (CLAHE)**. It splits the main image into several nonoverlapping sub-images. Histogram of the sub-images is sheared to restrict the value of enhancement of each pixel then to be equalized [6]. This technique was mainly founded for medical imaging [7]. There is other strong HE enhancement technique called **Dynamic Histogram Equalization (DHE)** it is performing an enhancement for the picture without losing its details. the input histogram divides into sub-histograms until newly constructed sub-histograms have no dominating part [8]. By partitioning histogram, whole technique is divided into three parts: partitioning the histogram, gray level GL ranges are allocated to each of sub histogram and HE is applied on each of them. Preprocessing also include eliminating lables and other extra objects. The algorithm of this function comes from Dr. Tong Zhang’s PhD Thesis (RPI, 2004), who utilized a fast marching level set technique and a quick change in the magnitude gradient for the segmentation [9].

### 2.3 Pectoral Muscle Elimination and ROIs Segmentation

After preprocessing stage; and in designing a classification system, an important stage is leading, it’s the segmentation. In which the image is divided into different regions to extract the (ROIs) through an automatic or semi-automatic process to classify all its lineaments according to obtained results. Mammogram segmentation generally classifies images into many region based on the main existing features, this includes breast border, nipple and pectoral muscle [10]. The pectoral muscle edge is beneficial in specifying mammogram sufficiency [11], comparison, mammogram-pair registration, and for limiting the screening space for lesion detection and calcification [12]. The outcomes of image processing is affected a region of a prevalent density in most mediolateral oblique vision of mammogram by the pectoral muscle. So, it is advised that the pectoral muscle eliminated through segmentation [13]. Mean filter and Sobel edge detector approach is used by Semmlow et al. to position the breast boundary on zero-mammograms [14]. This method might be considered as one of the premature methods for a segmentation of the breast contour. The most apparent approach would seem to be the utilization of
thresholding [15]. Normally, the interfering amount between the background breast
and region in such divisions will impossible results a miss-classification of some
the pixels of the background as breast region and vice versa [16]. A mammogram
breast anatomy is presented in Figure.1

Fig1. Mammographic breast anatomy (Laine and Huda, 1996).

3. Experimental work
Mammogram image analysis society is the selected dataset to be uses in this
paperwork. This dataset has 206 normal, 51 malignant, and 43 benign events. As we
mentioned above, the substantial goal of this contribution is to select the appropriate
enhanced image instantly which resulted from the two used enhancement techniques,
that to have the appropriate segmented ROIs, and for improving the system accuracy.
The stages that come after the segmentation will not be covered in this paper. The
proposed approach’s stages of this work are illustrated in figure 2.
3.1 Image Preprocessing

Preprocessing stage represent preparing the image for classification including removing noise, enhancing contrast and labels removal as explained below:

3.1.1 Filtering

In this step and in parallel, we implemented two main types of filtering techniques in image processing and measuring which image have better MSE and PSNR, and then the algorithm’s next step will be applied on the better filtered image.

a) Median Filtering.

Here a hybrid median filter instead of traditional filtering was implemented; a problem may arise like erase lines tighter than half neighborhood width. The hybrid median filter is a process of three steps, which implements two subsets of a $5 \times 5$ neighborhoods. These subsets are drawn from pixels parallel to the frame edges of the image, and at 45 degrees to the edges, centered on the reference pixel. The equation (4) of median filter is:

$$f(x, y) = \text{“median (g(s, t))”}$$

(4)
(s,t)∈ Sxy, where Sxy is the coordinates of sub-image window of size \( N \times M \).

b) Mean filtering
This filter is works on exchanging the value of each pixel in an image with the average neighbors of that pixel, including the pixel itself. It contains the impact of removing the value of pixel which is not true representative of their surroundings that as stated in equation (5):
\[
f( x, y ) = \frac{1}{mn} \sum_{s,t} S_{xy} g(s,t)
\]
After implementing these two aforementioned filtering techniques, the MSE and PSNR measurements are computed for every image. Below, the MSE and PSNR equations (6), (7).
\[
MSE = \frac{1}{IJ} \sum_{n=1}^{I} \sum_{m=1}^{J} [X(m, n) - Y(m, n)]^2
\]
\[
PSNR = 20 \log_{10} \left( \frac{255}{\sqrt{MSE}} \right)
\]
Where I, J are the images dimensions, \( X(m, n) \) is the original image and \( Y(m, n) \) is the approximated version. Lower value of MSE means minimum error value, and as shown by the inverse relation between the PSNR and MSE, this interprets to a high value of PSNR.

the average of (MSE) and the average of (SNR) are calculated to 25 images are selected randomly from each class of the mammogram images, after running statistical comparisons between the resulted images from the tow filters, that was find; the mean filter is giving better results than the median as illustrated in the table (1) below. For that; the mean filter is considered in the system’s algorithm while the median filter has been ignored.

**Table1: Mean values of MSE and PSNR for mean and median filtering for benign, malignant and normal images.**

<table>
<thead>
<tr>
<th>Class</th>
<th>Av-MSE mean filter</th>
<th>Av-MSE median filter</th>
<th>Av-PSNR mean filter</th>
<th>Av-PSNR median filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign</td>
<td>0.1088</td>
<td>0.8448</td>
<td>58.3653</td>
<td>40.9456</td>
</tr>
<tr>
<td>Malignant</td>
<td>0.1255</td>
<td>0.7926</td>
<td>57.5613</td>
<td>42.2300</td>
</tr>
<tr>
<td>Normal</td>
<td>0.0982</td>
<td>0.6683</td>
<td>58.6448</td>
<td>42.5964</td>
</tr>
</tbody>
</table>

Figure 3 shows the mammogram image after applying mean filter.
3.1.2 Image Enhancement

The prime goal of applying enhancement techniques is to improve contrast of mammography images so that the outcomes were more useable than the original image. Mammogram images are enhanced by utilizing histogram equalization methods. In this work and after removing noise step, two different and the more effective histogram equalization methods are implemented:

a. **Contrast limited adaptive histogram equalization**

The CLAHE technique demands on minimizing the noise and edge shadowing influence created irregular regions advantage of using CLAHE are overcoming over-enhancement and noise reduction by clipping spikes and also improved computation speed.

b. **Dynamic Histogram Equalization (DHE)**

This technique is performing an enhancement for the picture without losing its details. The input histogram divided into sub-histograms until newly constructed sub-histograms have no dominating part. Dynamic grey level for each of sub-histogram which is then mapped by HE. Total available dynamic grey level is partitioned between sub-histograms on basis of their dynamic range in input picture and CDF values of histograms.

After applying the two mentioned techniques at the same time on a selected mammogram image, standard deviation for every enhanced image is measured. The enhanced image that has better standard deviation value will be selected for the next algorithm’s step, that way is done because of the varying efficiency to the both used techniques in some cases, also for having maximum benefit from them. In other words; which method giving a better result than the other method, its output image will be taken to the coming system's step. Table (2) states the efficiency varying for
the used techniques on different images. Figure (3) showing sample of enhanced image.

Table 2: standard deviation values of enhancement techniques on some mammogram images

<table>
<thead>
<tr>
<th>Img_no.</th>
<th>STD_CLAHE</th>
<th>STD_DHE</th>
</tr>
</thead>
<tbody>
<tr>
<td>mdb019</td>
<td>23.3730</td>
<td>21.9407</td>
</tr>
<tr>
<td>mdb038</td>
<td>31.6386</td>
<td>25.8624</td>
</tr>
<tr>
<td>mdb006</td>
<td>24.6360</td>
<td>28.3202</td>
</tr>
<tr>
<td>mdb152</td>
<td>18.4963</td>
<td>23.3797</td>
</tr>
<tr>
<td>mdb110</td>
<td>25.2823</td>
<td>21.3742</td>
</tr>
<tr>
<td>mdb191</td>
<td>26.7720</td>
<td>25.7339</td>
</tr>
</tbody>
</table>

Fig4. mammogram image after applying enhancement CLAHE technique.

3.1.3 Labels Removal

The only interested area in mammogram image for classification is the breast area, for that, all labels, wedge and medical marks should be removing.

Firstly, create a binary map of all objects that are within the image, by using the lower threshold value for the convolution image; this threshold is used to binarize the image. Where every value less than or equal to this threshold is the background (false or 0) and what are more is foreground (true or 1), secondly, returns the connected components in the binary image, which is depends on a predefined connectivity of the connected components (objects), by using it for N-D input dimensions, a 4 connectivity had been used for 2D image. Thirdly; finding the areas of image objects; then the object with maximum area is presented as the breast profile and all other objects are being deleted from the image. Figure (5) shown the result obtained from this process.
3.2 Pectoral Muscle Elimination

The aim of concealment the pectoral muscle part from the breast area is to obtain only the ROIs and to minimize classification search area. The outcomes of image processing is affected a region of a prevalent density in most mediolateral oblique vision of mammogram by the pectoral muscle. So, it is advised that the pectoral muscle to be eliminated.

The pectoral muscle has been concealed by implementing the Seeded Region Growing (SRG) method. For the purpose of utilizing automated SRG, seed requires being placed inside the pectoral muscle area of the grayscale mammogram image. If the breast profile is right oriented a seed will be occurred inside the first 5th column and 5th row of the image. Figure (6) shows the result of this step.

SRG is segmenting the mentioned part according to a pre-specified point, known as “seed”. The neighbors pixels of seed which has same connectivity properties to that seed will appends forming the growth region Starting with a seed point. [17]. Let P is the set of unallocated pixels which it’s a neighbor for at least one of the regions which can be formulated as in equation (8):

$$P = \{x \in \bigcup_{i=0}^{n} A_i \mid N(x) \cap \bigcup_{i=0}^{n} A_i \neq \emptyset\}$$  \hspace{1cm} (8)

where $Ai$ is the set of iterative seed points.

$N(x)$ is the immediate neighbors set of the pixel $x$.

For $x \in P$, $N(x)$ meets just one of the $Ai$ using 4-connected neighbors to the pixel $x$. Then, $i(x) \in \{1, 2, 3, ..., n\}$ is defined to be that index such that $N(x) \cap Ai(x) \neq \emptyset$.
3.3 ROIs Segmentation
To obtain the ROIs from the image, graph cut segmentation method is implemented. The interested region in this work is the tumour region. The prime objective of implementing this algorithm for segmentation is splits the mammogram into several mammographic intensities. It is helpful for risk evaluation quantitative evaluation of intensity variations. Apart from the aforementioned features, it creates the contour (closed region) or a convex hull which is utilized for analyzing the morphological and original lineaments of the segmented regions. The earlier method performs in an effective shape classify of the extracted region into benign, malignant or normal. In recent years, Graph Cut has been implemented for image segmentation. With dynamic optimization method, the essence ideology of Graph Cut is to put an image onto a network graph and construct an energy function on the labeling [18]. Figure (7) illustrates the ideology of this method.

After extracting the ROIs from the image, the proposed system have been completed through feature extraction and tumor classification using wavelet discrete transform and multi-class support vector machine sequentially, the system reached 0.942 classification accuracy rate.

Fig6. Mammogram 019mdb image after eliminating the pectoral muscle process.
Fig7. (a) A straightforward 2D graph for a $3 \times 3$ image cut. (b) it’s minimal.

4. Conclusion and Future Work

This work here is ultimately to applying different preprocessing techniques in one time on image, then instantly, selecting the one with the better obtained results. This work involve applying the next steps of classification processing on the image by relying on the mathematical quantities returned from assessment functions of evaluating each step like MSE, PSNR and STD in the preprocessing stage. This will provides a more acceptable accuracy result in classification system.

therefore, we can conclude that, using a hybrid approaches in each one of the rest steps in classification system: enables the segmentation step to provide a more proper masses (ROI), in features extraction step also, for providing a good features among the others in the image, and last in the classification step, using a hybrid approach will produce result with high accuracy, and low false positive diagnoses. If the aforementioned approach is implemented, future works will involve the development of a CAD system that relies heavily on the outcomes presented in this contribution. The mammogram processing algorithm could be applied also for an automated detection of other types of abnormalities such as calcification, circumscribed masses, disguised masses and the work can be extended to be applied on other suspicious ill-defined masses, circumscribed lesions, asymmetry analysis etc.

References


