Segmentation Method for ROI Detection in Mammographic Images using Wiener Filter and Kittler's Method

Pragathi. J

ME Student, Department of Instrumentation & Control,
Cummins College of Engineering,
Pune, India.

H. T. Patil

Assistant Professor, Department of Instrumentation & Control,
Cummins College of Engineering,
Pune, India.

ABSTRACT

Breast cancer is the second leading cause of death for women all over the world. Screening mammography is currently the best available radiological technique for early detection of breast cancer. However the presence of artifacts can disturb the detection of breast cancer and reduce the rate of accuracy in the computer aided detection (CAD) systems. For this reason, the pre-processing of mammogram images is very important in the process of breast cancer analysis because it reduces the number of false positive. This paper proposes various filtering techniques to solve the noise removal problems and separate the background region from the breast profile region using an automatic thresholding technique and Connected Component Labelling. We evaluated our pre-processing method on a set of images obtained from a private hospital. Thus this preparation phase improves the image quality and accentuates the CAD results more accurate.

General Terms

Medical image processing, Algorithms.

Keywords

Mammogram, Thresholding, Connected Component Labelling, Breast region extraction, Computer aided analysis, Otsu method, Kittler method, Wiener filter .

1. INTRODUCTION

Breast cancer or Malignant Breast Neoplasm is the most common type of cancer in women [1]. It originates most commonly from the inner lining of ducts or the lobules that supply the ducts with milk. Worldwide, breast cancer comprises 22.9% of cancers in women and causes approximately 500,000 deaths every year. It can be treated depending on the size, stage, growth rate and other characteristics of the tumour; but a late diagnosis can prove to be fatal too. For this reason, it is best if the cancer is diagnosed in its early stages. There are several ways of examining breasts for early detection of breast cancer, they are Mammography, Ultrasound, MRI (Magnetic Resonance Imaging), Clinical Breast Exam, Breast cancer screening etc. Mammography is currently the best available radiological technique for early detection of breast cancer [2]-[3]. Mammography is the use of low-energy X-rays to examine the human breast. However, the large number of mammograms generated by population must be interpreted by the relatively small number of expert radiologists and in order to reduce the workload on radiologists and reduce the number of unnecessary biopsies in patients with benign disease, a variety of CAD systems (computer-aided diagnosis (CADi) and computer-aided detection (CADe)) have been proposed [4]. These systems are used to identify a suspicious lesion (masses, calcification, etc) and to decrease the number of false positives cases. To accomplish those goals, most of the CAD systems composed of successive steps: pre-processing, segmentation, feature extraction, feature selection and classification. Before CAD algorithms can be applied for the task of classification and Identification, mammograms need to be pre-processed.

The main objective of this pre-processing is to improve the quality of the image to make it ready to further processing by removing the unrelated and surplus parts in the back ground of the mammogram. Breast border extraction is also a part of pre-processing. The types of noise observed in mammogram are high intensity rectangular label, low intensity label, tape artifacts etc [6]. The types of noises present in mammogram are represented in Figure 1. Therefore pre-processing is used in mammogram orientation, label and artifact removal, mammogram enhancement and mammogram segmentation.

In this paper, we propose an automatic pre-processing technique for mammogram images. In the algorithm various spatial filtering techniques are compared and best is used for noise equalisation. By employing these techniques, the noise levels in radiography images and mammograms are reduced, post which these images can be processed for contrast and sharpness enhancement without the risk of enhancing noise. After which the radiopaque artifacts are suppressed and Breast region is separated from background using thresholding techniques and the separated breast profile region can be used for further processing in CAD algorithms.
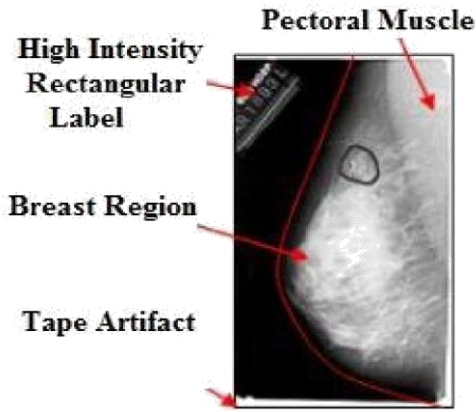
Fig 1: Types of noise observed in mammogram

2. RELATED WORKS

There have been various approaches proposed to the task of segmenting the breast profile region in mammograms. Some of these have focused on using thresholding [8], gradients [9], modelling of the non-breast region of a mammogram using a polynomial [6], or active contours [7], or by cropping. Mammogram pre-processing usually involves classifying mammograms into several distinct regions, including the breast border [6], the nipple and the pectoral muscle. In most of the approaches pectoral muscle is represented as a predominant density region in mediolateral oblique views of mammograms and is removed as a part of pre-processing step. But, an interesting article cited in [10]-[11], as shown that Mammography confirmed a 2×2 cm lump, suspicious of malignancy, in the upper outer quadrant of the right breast adjacent to the pectoralis major muscle, as shown in the Figure 2. Histological examination revealed this to be a benign granular cell tumour. If the pectoral muscle is removed in such cases, it would result in a wrong interpretation of the mammogram. So removal of pectoral muscle is not required and should be avoided to increase the accuracy in CAD algorithms.
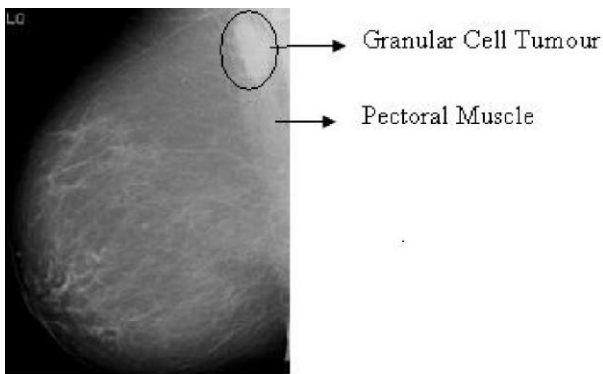

Fig 2: Mammogram showing a granular cell tumour in the pectoral muscle

3. MATERIALS

3.1 Mammography Technique

Mammography is performed using a low-dose X-ray system designed specifically for imaging the breasts. Each breast is lightly compressed between two flat plates with an X-ray film placed underneath. Compressing the breast helps by spreading the breast tissue out to obtain a clearer picture whilst using the

lowest possible dose. X-rays pass through the breast and onto the film beneath. The goal of mammography is the early detection of breast cancer, typically through detection of characteristic masses and Microcalcification. Like all x-rays, mammograms use doses of ionizing radiation to create images. It is normal and safe to use lower energy X-rays than those used for radiography of bones. Among all medical imaging techniques that used in breast cancer screening, digital mammogram is a convenient and easy tool in classifying tumours and also because of low cost and common in use, it is notable by all experts and physicians. Many applications in the literature prove mammogram effectiveness in breast cancer diagnosis. Currently, there are two types of mammography: one is film mammography and the other is digital mammography. Compared to digital mammography, screen-film mammography has some limitations, which include: 1) limited range of X-ray exposure; 2) image contrast cannot be altered after the image is obtained; 3) the film acts as the detector, display, and archival medium; and 4) film processing is slow and introduces artifacts. Spurgeon [12] showed that digital mammography depicts more tumours than screen-film mammography, especially lesions seen as micro calcifications (MCs). Pisano *et al.* [13] showed that digital mammography is more accurate in women under the age of 50, women with radio graphically dense breasts, and premenopausal women. There are two types of examinations performed using mammography: screening mammography and diagnostic mammography. Diagnostic mammography is often performed as a follow-up examination of an abnormal screening mammography in order to determine whether the area of concern on the screening examination needs additional breast imaging or a biopsy to determine whether the woman has breast cancer [1]. Screening mammography generally consists of four views, with two views of each breast: the craniocaudal (CC) view and the mediolateral oblique (MLO) view.

Radiologists then analyze the image for any abnormal findings. Tumours can be recognized as locally low density areas on mammograms. Their absolute values however, are not constant and they vary in size, background variations, imaging conditions and so on. Radiologists conclude that a patient has breast cancer if certain types of masses or calcifications (small white specks) in particular, spatial distributions are identified in the mammogram.

Occasionally, a mass commonly is described as a "spot" that may appear on the mammogram and has certain features which suggest a breast cancer. Similarly, calcifications which are deposits of calcium into various tissues within the breast may display a pattern, which the radiologist will identify.

3.2 Database

For the experimental work, we have 2 different databases. First of all, we have a set of 62 abnormal images and 19 normal images collected from a private Hospital & Research Centre, Pune, India. Then, we have access to the two main well known databases in the world of mammogram images, mini-MIAS and DDSM, for evaluation purpose.

4. METHODOLOGY

The framework proposed for automated mammogram pre-processing is shown in the Figure 3, which is implemented in MATLAB version 7.0. Since the images collected from the hospital are as jpeg version and is of size 1024×1024 pixels, before pre-processing these images have to be converted to gray scale and is resized to 256×256 pixels.

4.1 Noise Removal

Noise means an unwanted sound. In the context of images, image noise also known as image mottle, is an unwanted variation of brightness or colour information in a displayed image even when no image detail is present.
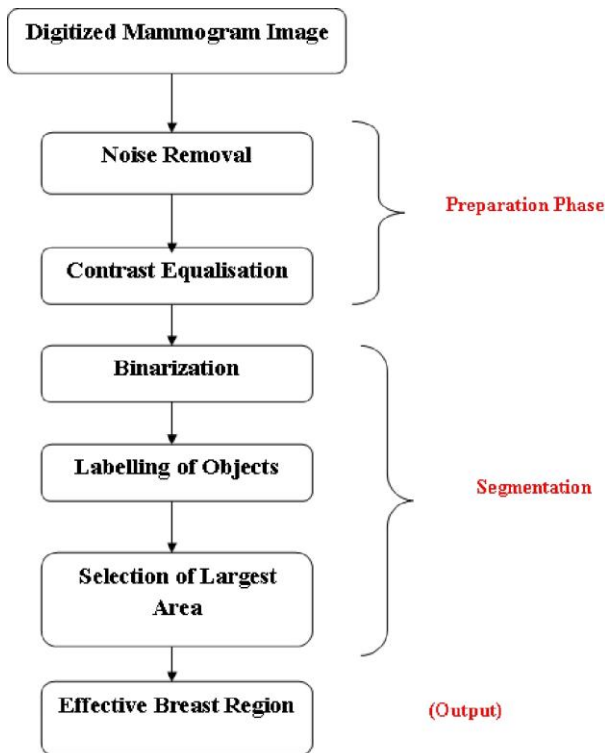

Fig 3: Proposed framework for mammogram preprocessing

This variation is usually random and has no particular pattern. This is especially significant when the images have a low contrast. All medical images tend to contain some visual noise. For the case of screen-film mammography, the structure of the film, intensifying screens, intensifier tube screens or digital receptors can introduce noise into images. There are several techniques employed for the removal of noise in images. Any technique applied to the images for noise reduction broadly falls under either of the two domains:
1) Spatial Domain
2) Frequency Domain

In this context, we used various spatial filters, like mean, median, wiener filter and frequency domain filters, like gaussian filter and Laplacian filter, for applying on image in order to choose the best; the output of various filters when applied on the mammogram gray scale image is shown in the Figure 4, as seen smoothing filter and gaussian filter not only removes noise but also blurs the edges. The Laplacian of an image highlights regions of rapid intensity change and is therefore often used for edge detection. Median filtering presents a significant reduction in noise while the edges are preserved better than by averaging, gaussian and Laplacian. A median filter sets each pixel to the median of itself and its neighbouring pixels. This filter modifies the noisy intensity values to a lower intensity value close to its neighbours and thus aims at lowering noise. This kind of filter is good at removing salt and pepper noise. This is a non-linear technique since the modification doesn't obey a linear equation or

manipulation. Wiener filter, known as "minimum mean square error filter", is an adaptive linear filter, applied to an image locally, by taking into account the local image variance. When the variance in an image is large the Wiener Filter results in light local smoothing, while when the variance is small, it gives an improved local smoothing (Math Works, 2004). Wiener filtered image when compared with median filtered image as shown in Figure 4, from which we can derive that wiener filter is choosed for image noise removal. Comparison of filter output implemented in MATLAB is shown in the Figure 4. From the results we see that average and gaussian filtered image shows blur in the image. But in median and wiener filtered image after noise removal also the edge information is saved.
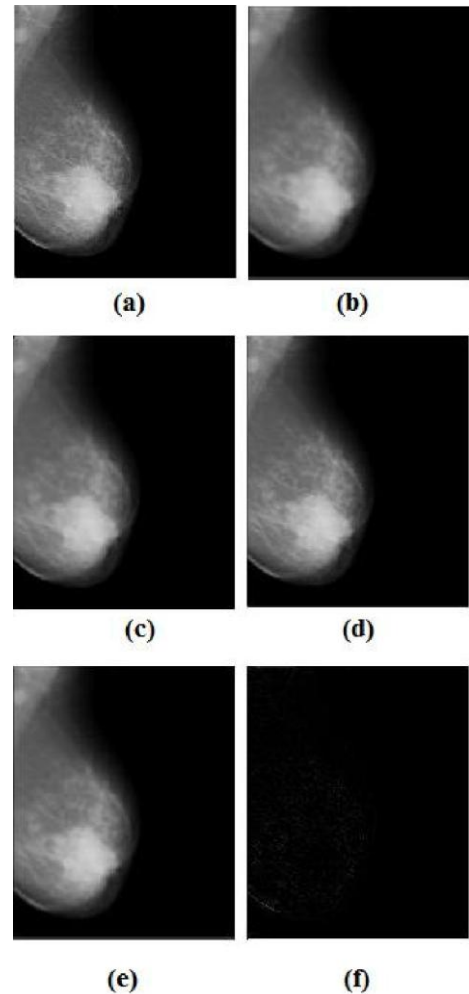

Fig 4: Results of mammogram digitization noise removal using different filters. (a) Original image (b) Average filtered image (c) Median filtered image (d) Wiener filtered image (e) Gaussian filtered image (f) Laplacian filtered image

4.2 Contrast Equalisation

In the majority of images, better visibility of the image for the reader has been achieved through histogram equalisation of the original image. This is a process where pixel brightness values are re-assigned in an attempt to have the same number of pixels correspond to each brightness value, resulting in a linear cumulative histogram. The basis of histogram equalisation is that peaks in a histogram identify high

probability pixels, often corresponding to some region. This is especially the case in the mammogram since tissue type corresponds directly to X-ray attenuation and as a result the gray scale intensity of that tissue type image. Valleys in the histogram are assumed to correspond to less important features or regions. The effect of re-distributing intensities through histogram equalisation is that the peaks are spread out, increasing the contrast of the regions, and the valleys are compressed as adjacent low probability histogram bins are aggregated to produce a bin with the same probability as the others in the histogram. To the output of wiener filtered image, histogram equalisation is applied to obtain a uniform histogram as shown in the Figure 5, using the command 'histeq' in MATLAB.
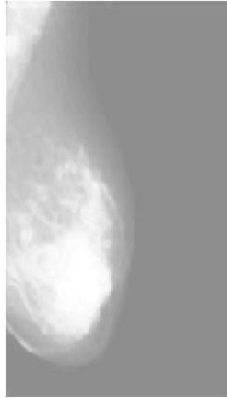
Fig 5: Histogram equalized image

5. SEGMENTATION

5.1 Binarization

A binary image is a digital image that has two possible values for each pixel. This means each pixel in the image is stored as a single bit (0 or 1). Binary images are produced from gray scale image or colour image by a process called segmentation. Segmentation is the process of assigning each pixel to two or more classes. Robust binarization gives the possibility of a correct extraction of the breast region from its background. For the binarization of images two algorithms have been implemented to choose the best.

5.1.1   Otsu Method

The simplest form of segmentation is Otsu thresholding in this step; we use an automated thresholding method to obtain a binarization of the enhanced image. In this context, we use Otsu's method to automatically perform histogram shape-based image thresholding, or, the reduction of a gray level image to a binary image. Here binarization is done on the histogram equalized image. Implementation is done according to the algorithm described below and the output obtained is shown in the Figure 6. In Otsu's method we exhaustively search for the threshold that minimizes the intra-class variance (the variance within the class), defined as a weighted sum of variances of the two classes:

$$\sigma^2_{within}(t) = W_1(t)\,\sigma^2_1(t) + W_2(t)\,\sigma^2_2(t)$$

Weights W_i are the probabilities of the two classes separated by a threshold t and σ^2_i variances of these classes. Otsu shows that minimizing the intra-class variance is the same as maximizing inter-class variance:

$$\sigma^2_{between}(t) = W_1(t).W_2(t)\,[\mu_1(t) - \mu_2(t)]^2$$

Which is expressed in terms of class probabilities W_i and class means μ_i.

The class probability W1 is computed from the histogram as t:

$$W_1(t) = \Sigma^t_0\ p(i)$$

While the class mean $\mu_1(t)$ is:

$$\mu_1(t) = \Sigma^t_0\ p(i)\,x(i)$$

Where x (i) is the value at the center of the *ith* histogram bin. The class probabilities and class means can be computed iteratively.

5.1.2 Kittler's method

Kittler and Illingworth have proposed a thresholding algorithm whose cost function to be optimized is based on the Bayesian classification rule. In this method, it is assumed that components in the bi-modal histogram in a gray-level image h (g) are normally distributed. Normal distributions are defined by their means μ_i, standard deviations σ_i, and a priori probabilities P_i. For a case of two different classes (i = 1; 2), the background and foreground, and given a threshold T, the parameters can be estimated from the following:

$$P_i(T) = \sum_{g=a}^{b} h(g)$$

$$\mu_i(T) = \frac{1}{P_{I(T)}} \sum_{g=a}^{b} h(g)\,g$$

$$\sigma^2_{i(T)} = \frac{1}{P_{I(T)}} \sum_{g=a}^{b} h(g)\,(g - \mu_i(T))^2$$

Now, the criterion function can be calculated as,

$$J(T) = 1 + 2[P_1(T)\log\sigma_1(T) + P_2(T)\log\sigma_2(T)] - 2[P_1(T)\log P_1(T) + P_2(T)\log P_2(T)]$$

And the minimum error threshold can be computed by minimizing the criterion J (T). Implementation of kittler method and Otsu method is shown in the Figure 6.

5.2 Artifact Suppression and Background Separation

In most images, the breast region is connected with tape artifact. For this reason, this step is used to separate the breast region from the background (tape artifact and labels). Connected component labelling scans an image and groups its pixels into components based on pixel connectivity. All pixels in a connected component share similar pixel intensity values and are in some way connected with each other. Once all groups have been determined, each pixel is labelled with a gray level or a colour according to the component it was assigned to. Connectivity is determined by the medium; image Graphs, for example, can be 4-connected or 8-connected. Connectivity checks are carried out by checking the labels of pixels that are North-East, North, North-West and West of the current pixel (assuming 8-connectivity). 4-connectivity uses only North and West neighbours of the current pixel. The following conditions are checked to determine the value of the label to be assigned to the current pixel (4-connectivity is assumed). Finally, we use the connected component labelling to divide the binary image into different labels. Implementation in MATLAB is as follows, the different objects in the binarized image is labled, using the command,

'bwlabel(BW, n)' returns a matrix L, of the same size as BW, containing labels for the connected objects in BW. 'n' can have a value of either 4 or 8; here we have choosed the n value to be 4. The label matrix returned by bwlabel or bwlabeln is of class double; it is not a binary image. One way to view it is to display it as a pseudocolor indexed image, using command, 'label2rgb', as shown in the Figure 7.
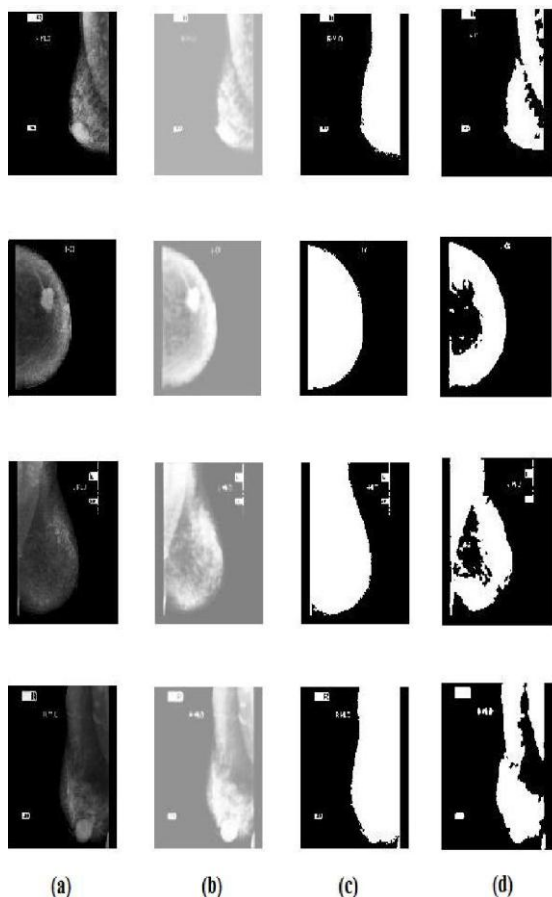

Fig 6: Binarization results of various methods. The (a) column of images is 4 different original images. The (b) column of images, are results of wiener filter and contrast enhanced image. (c), shows result of Kittler's binarized images and (d), shows the Otsu binarized images.

5.3 Region selection and Breast Region Extraction

Looking at the image generated by the last step, we note that breast region has the largest area. For this reason, we use the area criteria to select the label that represents the breast region and to eliminate the unlikely labels (tape artifacts and high intensity labels). Using the command, *'STATS= region props(L,properties)'* measures a set of properties for each labelled region in the label matrix L. Here we want to measure the area of largest object, the actual number of pixels in the region. The function *'ismember'* is useful in conjunction with *region props* for selecting regions based on certain criteria. Thus, the object with largest area is displayed. Finally, to obtain the effective breast region, the result of the previous step was multiplied with the original mammogram.
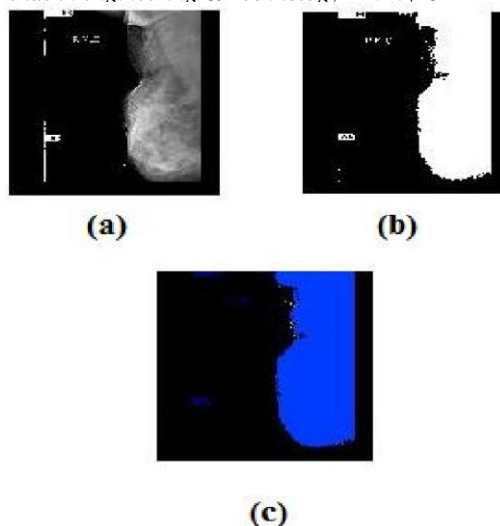

Fig 7: Component connected labelling to divide the binary image into different labels. (a) Original image (b) Binarized image (c) Labelled image
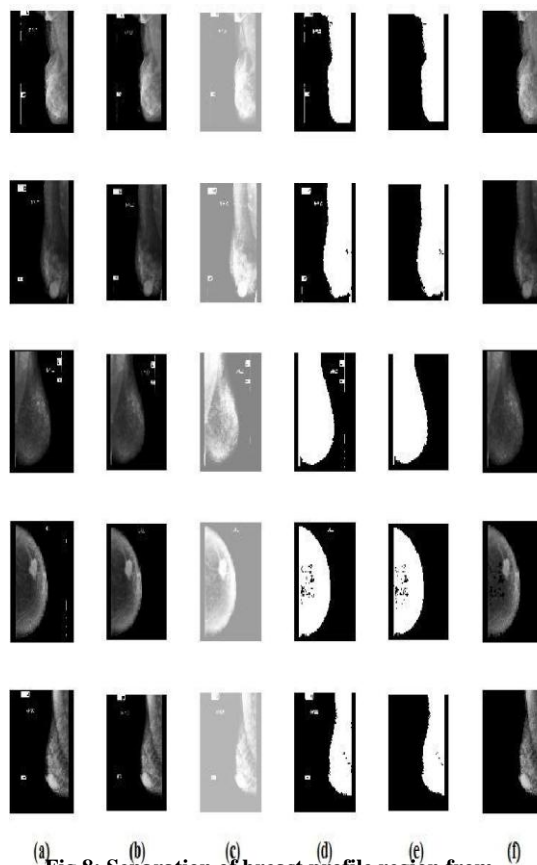

Fig 8: Separation of breast profile region from background by kittler method. The (a) column images are 5 different original images. The (b) column of images, are results of wiener filter, (c) shows contrast enhanced image.

(d) Shows result of Kittler's binarized images (e), shows largest area selected. (f), shows breast profile separated from the background.

6. EXAMPLES

In the following, we would like to give few examples to show the application of the proposed method. Here we used 5 different mammogram images, which were all digitized at a resolution of 1024×1024 pixels. Since these images were stored in jpeg version they were converted to grayscale images and resized to 256×256 pixels. The proposed algorithm uses wiener filter to denoise the image and kittler algorithm to binarize the denoised and contrast enhanced image. Then connected component technique was used to label the different objects in the binary image, and object with largest area is selected and all other objects are removed. The

resulting binary image is multiplied with the original image to form the final gray scale image. The following above steps were implemented in MATLAB 7.0 and results obtained are shown in the Figure 8. For comparison purpose, to the same images, wiener and Otsu algorithm were implemented as shown in the Figure 9. The results shown in Figure 8 (d) and Figure 9 (d) show that proposed algorithm which uses kittler method for binarization of images performs effectively without losing any information of the ROI. Comparing the final results shown in Figure 8 (f) and 9 (f), we can clearly say that the proposed method is able to accurately segment mammogram images for Region of Interest (ROI) detection.

7. CONCLUSION

In this paper an automated method for pre-processing the mammograms has been introduced. Since, the results of the CAD system depends strongly on the preprocessing step which includes noise equalization and extracting the breast area correctly in order to save processing time and avoid false detection made by markers and sharp edges near chest side. Here a comparative study was made by implementing different filters and different thresholding methods to segment the ROI in the image. The preparation of the image before filtering and binarization consists of converting the RGB image to gray scale image and resize the image. The best denoising filter is derived after comparison of various types of filters and more specifically the: Mean, Median and Wiener filters in spatial domain and Gaussian and laplacian filters in the frequency domain, in which wiener filter is proposed to remove the digitization noise present in mammogram images more effectively. Compare to Otsu thresholding, Kittler method promises to provide an efficient binarization output. Connected component labelling is used to label the different objects in the image and using the largest area criteria the breast region is extracted. Thus further processing is restricted to the breast area. The importance of pectoral muscle is also discussed in the paper. With these preprocessing results, the next objective of our future research work will involve the development of a computer aided detection system.
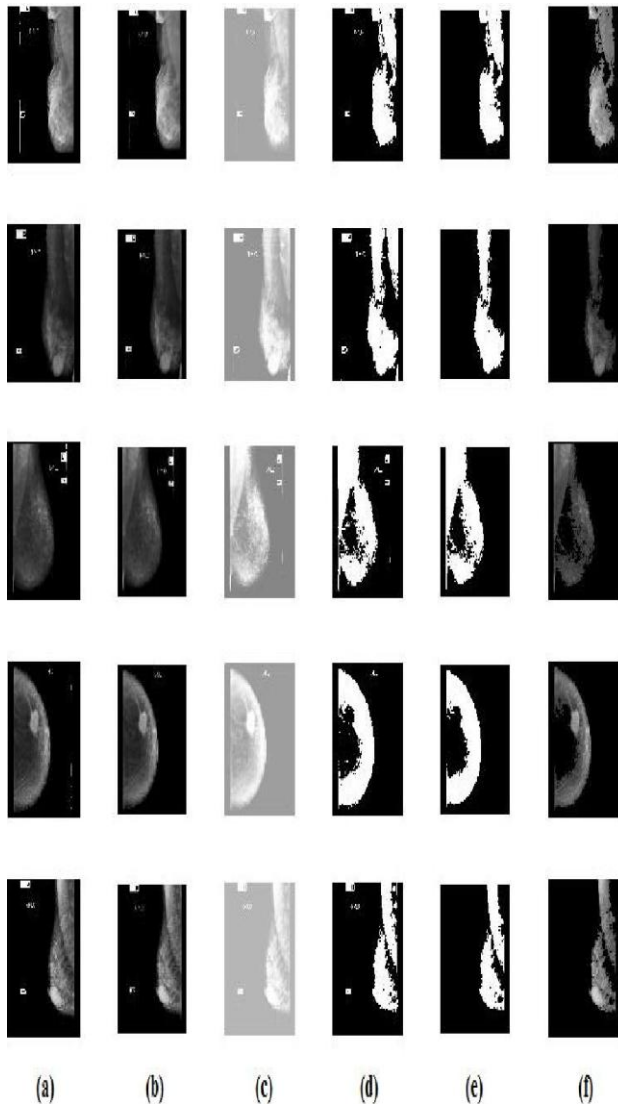

Fig 9: Separation of breast profile region from background by Otsu method. The (a) column images are 5 different original images. The (b) column of images, are results of wiener filter, (c) shows contrast enhanced image.

(d) Shows result of Otsu's binarized images (e), shows largest area selected. (f), shows breast profile separated from the background.

8. REFERENCES

[1] Breast Cancer Facts and Figures 2009-2010, ACS.

[2] Gøtzsche PC, Nielsen M (2011). "Screening for breast cancer with mammography". *Cochrane Database Syst Rev* (Online): Available. doi:10.1002/14651858.CD001877.pub4.

[3] K. H. Ng and M. Muttarak, "Advances in mammography have improved early detection of breast cancer," J. Hong Kong College Radiol., vol. 6, no. 3, pp. 126131, 2003.

[4] Jinshan Tang, Senior Member, IEEE, Rangaraj M. Rangayyan, Fellow, IEEE, Jun Xu, Issam El Naqa, Member, IEEE, and Yongyi Yang, Senior Member, IEEE "Computer-Aided Detection and Diagnosis of Breast Cancer with Mammography: Recent Advances", IEEE Transaction on Information Technology on Biomedicine, VOL. 13, NO. 2 March 2009.

[5] Samir Kumar Bandyopadhyay, pre-processing of Mammogram Images, International Journal of Engineering Science and Technology, Vol. 2(11), 2010, 6753-6758.

[6] R. Chandrasekhar, and Y. Attikiouzel, "Automatic Breast Border Segmentation by Background Modelling and Subtraction", in 5th International Workshop on Digital

Mammography (IWDM), (Yaffe M. ed.), Medical Physics Publishing, Madison, USA, pp. 560–565.

[7] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models", International Journal of

Computer Vision, vol. 1, pp. 321–331, 1998.

[8] U. Bick, M.L. Giger, R.A. Schmidt, R.M. Nishikawa, D.E. Wolverton, and K. Doi, "Automated Segmentation of Digitized Mammograms", Academic Radiology, vol. 2, no. 2, pp. 1–9, 1995.

[9] A.J. Mendez, P.J. Tahoces, M.J. Lado, M. Souto, J.L. Correa, and J.J. Vidal, "Automatic Detection of Breast

Border and Nipple in Digital Mammograms", Computer Methods and Programs in Biomedicine, vol. 49, pp. 253–262, 1996.

[10] Gordon AB, Fisher C, Palmer B, Greening WP. Granular cell tumour of the breast. Eur J Surg Oncol. 1985; 11:269–73. [PubMed].

[11] Anish Patel, Valentina Lefemine, Syed Mansoor Yousuf, and Walid Abou-Samra, Granular cell tumour of the pectoral muscle mimicking breast cancer, Eur J Surg Oncol. 1985; 11:269–73. [PubMed].

[12] D. Spurgeon, "Digital mammography is more accurate only for certain groups of women," Br. Med. J., vol. 331, no. 7518, pp. 653–653, 2005.

[13] E. D. Pisano, C. Gatsonis, E. Hendrick, M. Yaffe, J. Baum, S. Acharyya, E. Conant, L. Fajardo, L. Bassett, C.

D'Orsi, R. Jong, and M. Rebner, "Diagnostic performance of digital versus film mammography for breast cancer screening," New England J. Med., vol. 353, no. 17, pp. 1773–1783, 2005.

[14] N. Otsu, "A threshold selection method from grey level histogram", *IEEE Trans. Syst. Man Cybern.,* vol. 9 no. 1, 1979, pp. 62-66.

[15] He J, Q. D. M. Do, A. C. Downtown and J. H. Kim, A Comparison of Binarization Methods for Historical Archive Documents, Proc of the 8th International Conf on Document Analysis and Recognition (ICDAR'05) 2005 IEEE.

[16] J. Kittler and J. Illingworth. Minimum error thresholding. *Pattern Recognition*, 19:41.47, 1986.

[17] J. Bozek, M. Mustra, K. Delac, M. Grgic, "A Survey of Image Processing Algorithms in Digital Mammography", Springer-Verlag, pp. 631–657, Berlin Heidelberg, 2009.

[18] R. J. Ferrari, A.F. Frere, R.M. Rangayyan, J.E.L. Desautels, R. A. Borges, "Identification of the breast boundary in mammograms using active contour models", In Med. Biol. Eng. Comput., vol. 42, pp. 201-208, 2004.

[19] H. Mirzaalian, M.R Ahmadzadeh, S. Sadri, M. Jafari, "Pre-processing Algorithms on Digital Mammograms", In MVA2007 IAPR Conference on Machine Vision Applications, pp. 118-121, 2007.