INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Noise Removal from Digital Mammogram Images for Proper Diagnosis of Breast Cancer

Mrs. Sangita Bhattacharjee[1], Dr. Indra Kanta Maitra[2], Prof. Samir Kumar Bandyopadhyay[3]
[1]PhD Scholar, Lincoln University College, Malaysia,
[2]Assistant Professor, B.P. Poddar Institute of Management and Technology, Kolkata, India,
[3]Advisor to Chancellor, JIS University, India,

Abstract: The introduction of digital mammography represents a significant technological advance in breast imaging. However, these images may contain artefacts. Commonly encountered artefacts include patient-related artefacts, hardware-related artefacts, detector-associated artefacts, collimator misalignment and underexposure and grid lines. This paper proposed algorithm for elimination of artefacts and noise in medio-lateral oblique (MLO) view of mammograms.
Keywords: Digital mammography, Noise, Artefacts, SRGA, MLO

I.   INTRODUCTION

Digital mammography images may contain artefacts. Commonly encountered artefacts include patient-related artefacts, hardware-related artefacts, detector-associated artefacts, collimator misalignment and underexposure and grid lines. Software processing artefacts like vertical processing bars, loss of edge and high-density artefacts can also be present. Although some of these artefacts are similar to those seen with screen-film mammography, many are unique to digital mammography. Such artefacts and noise in mammogram images is the major obstacle to develop fully automated Computer Aided Diagnosis (CAD) systems. This affects the result and accuracy of algorithms. Hence it is essential to perform preparation steps to suppress these artefacts and noise from background and enhance the breast region.

This paper proposed for elimination of artefacts and noise in medio-lateral oblique (MLO) view of mammograms. The proposed Seeded Region Growing Algorithm (SRGA) will remove artefacts followed by image orientation to obtain uniform images for each pair of mammograms. Finally, Gaussian filter is used to eliminate noise from the mammogram image.

II. LITERATURE REVIEW

In the preparation process on digital mammogram it is essential to understand the different types of artefacts and noise which are present within the breast and non-breast region of digital mammogram. While the incidence of artefact on digital mammographic images are typically less than with film based mammography, artefacts can be produced on digital systems.   Some researchers reviewed all the artefacts in mammography encountered and classified the causes of these artefacts into four categories such as patient information, technologist-related, related to the mammographic unit and related to processing and the processor[1].

Other researchers described the process of improvement of image quality for digital mammography and proposed method optimises the visualisation of breast cancers [2]. Researchers found that it was difficult to standardise and compare the image quality of Full-Field Digital Mammography (FFDM) versus screen-film mammography in a screening population[3]. Researchers also recommended set of tests including additional and improved tests, which they believe meet the intent and spirit of the Mammography Quality Standards Act regulations to ensure that full field digital mammography systems are functioning correctly and consistently producing mammograms of excellent image quality [4].

Some other researchers classified the artefact especially on digital mammogram [5-7]. According to them "some of these artefacts are similar to those seen with screen-film mammography, many are unique to digital mammography—specifically, those due to software processing errors or digital detector deficiencies. In addition, digital mammographic artefacts depend on detector technology (direct vs. indirect) and therefore can be vendor specific. It is important that the technologist, radiologist and physicist become familiar with the spectrum of digital mammographic artefacts and pay careful attention to digital quality control procedures to ensure optimal image quality".

Researchers also proposed one of the well-known artefact suppression algorithm based on "area morphology" to remove radiopaque artefacts from the background region of mammograms [6]. Here, a comprehensive technique is proposed to suppress the unwanted artefact from the digital mammogram along with noise reduction to improve the quality of the image. At the same time images will be homogeneously oriented to meet the uniformity.

III. PROPOSED METHODOLOGY

Artefacts are common to digital mammograms. Recognition of these artefacts is critical for achieving optimal image quality. Digital mammography systems differ in the way they acquire, process, and display images, and artefacts can be a result of problems involving any one of these components. Many artefacts specific to screen-film mammography have been well documented and are recognisable on patient images. Some examples include dust artefacts, pickoff, processor roller artefacts, static artefacts, and fogging artefacts. Some artefacts may be seen at both screen-film mammography and digital mammography, such as patient-related artefacts and hardware-related artefacts.

External artefacts are represented by tapes and other identification marks that are used on the mammogram to identify the patient and related data that may be required for identification of the mammogram. These markings provide high intensity regions on the mammogram and are inconsequential to the investigation of abnormalities within the mammogram. Presence of such artefacts also changes the intensity levels of the mammogram image significantly that may affect statistical analysis on the image. In this chapter the algorithm proposed by me attempts to remove all such artefacts, markings on the non-breast region of the mammogram and replace them with the background colour. The resulting image is free from any other object except the breast region.

To achieve the desired goal a new algorithm has been introduced by the combination of modified seeded region growing with thresholding. It has been observed that in MLO mammogram images, the breast portion is placed in the middle of the object irrespective of left and right breast. As per the characteristic feature of mammogram, breast region is represented by high intensity pixels, where as background consist of chest wall and skin to air are represented by ideally zero intensity or by very low intensity i.e. not more than 10 in grey scale images. The external artefacts and other irrelevant object are present in background, more specifically in the skin to air part of mammogram image.

The proposed algorithm is to search a seed from breast region of mammogram. It has been observed that the pixel that is located at height/2, width*3/4 may always lie in the breast region of mammogram with higher intensity value. Initially this pixel is considered as the seed of region growing algorithm. Otherwise, it will continue to search a new location given by height/2, width*5/8 and if necessary height/2, width/2. It is observed that in all the cases the seed will be obtained from these locations. The Original Mammogram with the preferred pixel location is shown in figure 1.
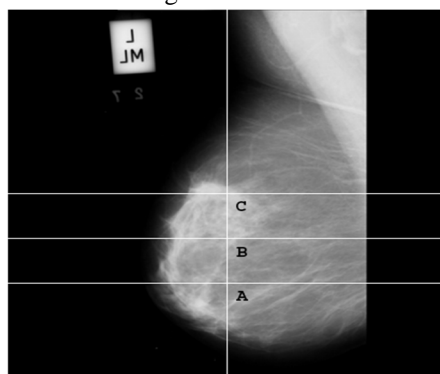

Figure 1 Original Mammogram with the preferred pixel location (A, B and C) to collect seed for region growing algorithm

After finding the seed for a region, next objective is to extract only the breast region from the background, leaving the background with artefacts behind and copying the breast region on another blank image. The seed pixel is copied to the new image at the same location of the pixel in the original image. This seed pixel is being coloured black in the original image. It is known that the breast regions of the mammogram have high intensity values whereas the background contains low or zero intensity. The algorithm starts searching for pixels that bounds the seed pixel. For each seed pixel the four boundary pixels located north, east, west and south of the pixel is also checked to find out whether they have high intensity value. Here algorithm uses the thresholding technique to divide the breast region pixel with background. If the pixels are with high intensity value, these will be used by the algorithm as seed , stored in a stack, for further searching.   The process continues by popping a seed from the stack and checking its intensity. If the seed is of high intensity value becomes the next seed. The value of intensity is copied to the new image at the same location of the pixel in the original image and the pixel is coloured black in the original image. This process continues till the stack is empty. The region grows from the single seed and stops when the entire breast region is blackened on the original image and the corresponding entire breast region is copied to the new image. The output image consists of only the breast region and remaining artefact are left behind in the original image. Now the external artefact free mammogram image will be the input for next processing.

1) Algorithm: Proposed Seeded Region Growing Artefact Removal Algorithm

SEEDED-REGION-GROWING (OrgImage, ImgWidth, ImgHeight)

$\Delta t \leftarrow 10$

If OrgImage[ImgHeight/2, ImgWidth*3/4]. Intensity $> \Delta t$
Then   GROW-REGION (OrgImage, ImgHeight/2, ImgWidth*3/4, Δt)
      Else If OrgImage[ImgHeight/2, ImgWidth*5/8]. Intensity $> \Delta t$
            Then   GROW-REGION (OrgImage, ImgHeight/2, ImgWidth*5/8, Δt)
                  Else If OrgImage[ImgHeight/2, ImgWidth/2]. Intensity $> \Delta t$
                       Then   GROW-REGION (OrgImage, ImgHeight/2, ImgWidth/2, Δt)
                       Else   Error

Return

GROW-REGION (OrgImage, h, w, Δt)

Stack $\leftarrow$ New Empty Stack

NewImage $\leftarrow$ New Blank Image

Stack.Push (h)

Stack.Push (w)

While Stack $\neq$ Empty
    Do   x $\leftarrow$ stack.Pop ()
              y $\leftarrow$ stack.Pop ()
              GreyValue $\leftarrow$ OrgImage[x,y].Intensity
              If   GreyValue $> \Delta t$
                Then   NewImage[x, y].Intensity $\leftarrow$ GreyValue
                       OrgImage[x, y].Intensity $\leftarrow$ 0
                       If   x-1 > 0 AND OrgImage[x-1, y].Intensity $> \Delta t$
                            Then   Stack.Push (y)
                                         Stack.Push (x-1)
                       If x+1 < OrgImage.height AND OrgImage[x+1, y].Intensity $> \Delta t$
                            Then   Stack.Push (y)
                                             Stack.Push (x+1)
                       If y-1 > 0 AND OrgImage[x, y-1].Intensity $> \Delta t$
                            Then   Stack.Push (y-1)
                                           Stack.Push (x)
                       If y+1 < OrgImage.width AND OrgImage[x, y+1].Intensity $> \Delta t$
                            Then   Stack.Push (y+1)

                                     Stack.Push (x)

Return (NewImage)

The output of algorithm is shown in figure 2.

A. Complexity Analysis of Algorithm

Here the image size is N*N. But there is no iteration in the algorithm to read the entire image. So, complexity will never exceed n^2. The pixels are traversed directly using x and y value. The numbers of pixels accessed on the image depend on the area covered by the breast region. By observation, it is found that half of the mammogram image is covered by the high intensity breast region. So, it may be said that half of the pixels out of N*N pixel are traversed by the algorithm. Hence the average time complexity of the algorithm will be approximately $n^2/2$.
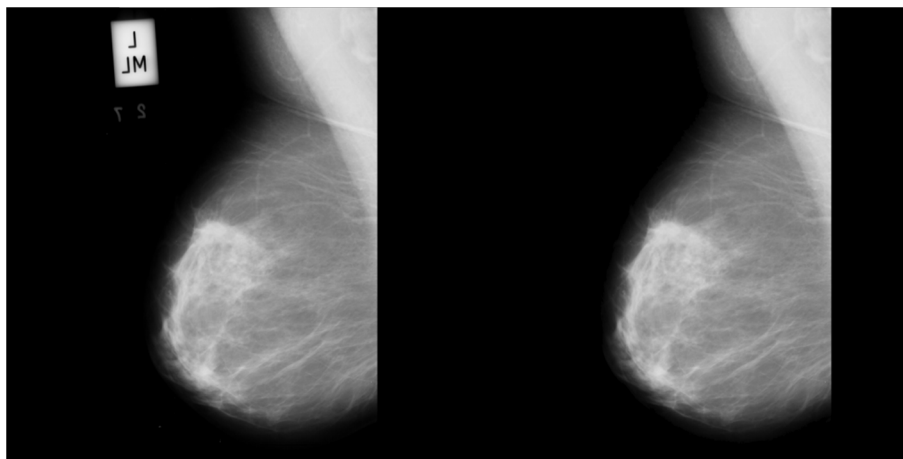
Figure 2 Original Mammogram and Mammogram without external artefact after SRGA

One of the major conditions of fully automated system is to standardise the input. The mammogram image needs to be transformed, so that, the chest wall location, i.e., the side of the image containing the pectoral muscle is on the upper left corner of the image in MLO view. Each mammogram pair represents a set containing the left and the right breast. So at this phase it is needed to identify the entire breast mammograms that have a left orientation, which is desirable from the point of view of execution of subsequent algorithms. These left breast mammograms are represented by their chest wall on the left side and the pectoral muscles on the top left corner of the mammogram while the breast boundary (skin-air surface) and nipple on the right side. The right breast mammogram needs to be flipped horizontally at 180°, so that, it is an exact mirror reflection of the image. So the image now obtained after flipping is of similar orientation to the left breast mammogram. This process will allow the breast regions to be compared and analysed in a similar way by applying same automated algorithms for both the mammogram images. This is especially significant when two mammograms of the same pair needs to registered for determining the symmetry between a pair of mammograms.

The process of image orientation starts with the identification of the mammogram, whether it represents the left or right breast region. In case of left, algorithm will ignore the orientation process as it is already in the desired orientation. If the mammogram is the right breast, it is needed to be flipped horizontally. To identify the orientation, algorithm scans the pixel intensities from left to right within a row of pixel intensities, at fixed intervals of rows (Δd). The background of the mammogram is represented by very low intensity but the breast regions have higher intensities. As soon as, it finds a high intensity value, it stops scanning. On performing such scans on a number of rows at fixed interval the algorithm tries to find out the column value (K) that represents the breast region. If it falls on a straight line it can be safely concluded that it is the chest wall and the orientation is left. So, it does not require further checking. If the subsequent column values ($\neq$K) do not fall in a straight line then algorithm can concludes that it represents the skin air interface of the breast region and it is right side up. Such mammograms will be required to be flipped.

The process of Image flipping can be done with any general purpose image processing software but in this research it has been done automatically . The process involves the scanning of the image and copying the mirror reflection of the image pixels on another image. The process starts by scanning the pixel intensities of the image and then copying the pixel intensity to the resulting image exactly at position that is obtained by subtracting the position from the width of the image.

$$P_I = \sum_{i=0}^{height} \sum_{j=0}^{width} (P_{i,j} = Q_{i,width-j}) \qquad (1)$$

This process continues for all the pixels of each row and the subsequent rows till the last pixel of the last row are copied. The resulting image is horizontally flipped at 180° and the right breast mammogram becomes similar to the left breast mammogram with their chest wall on the left side and the pectoral muscles on the top left corner of the mammogram, while the breast boundary (skin-air surface) and nipple on the right side.

1) Algorithm: Homogeneous Orientation using Image Orientation Algorithm

HOMO-ORIENTATION (OrgImage, Height, Width)

ChestWall ← 0

Δd ← Constant

Loop i ← 0 to Height

```
Do      Loop j ← 0 to Width
            Do      If   OrgImage[i, j] > 0
                    Then If i = 0
                            Then ChestWall ← j
                        If ChestWall ≠ j
                            Then HORZ-FLIP (OrgImage, Height, Width)
                                    Return
                        Break
        i ← i + Δd
Return
HORZ-FLIP (OrgImage, Width, Height)
NewImage ← New Blank Image
Loop i ← 0 to Height
    Do      Loop j ← 0 to Width
            Do      NewImage[i, j].Intensity ← OrgImage[i, Width - (j + 1)].Intensity
Return (NewImage)
```

B. Complexity Analysis of Algorithm

Here the image size is N*N. The proposed method is consisting of two parts; initial part is to detect orientation and latter part is only applicable for right breast. By observation, it may be said that the chest wall part approximately occupies ¼ part of the width of mammogram image. Initial process will start scanning from OrgImage[0, j] and terminated to Chest Wall i.e. OrgImage[N/4, j]. So, N/4 numbers of pixels are scanned. Further the algorithm scans the pixel intensities at fixed intervals of rows (Δd). So, it N/Δd number of row are scanned. Finally, it may be concluded that time complexity of the initial part of the method is (N/4)*(N/Δd) i.e. $n^2/(4*Δd)$. The later optional part, will require N*N processing and generate time complexity of n^2. If the image is consisting of left breast then it will be much faster. It is shown in figure 3.
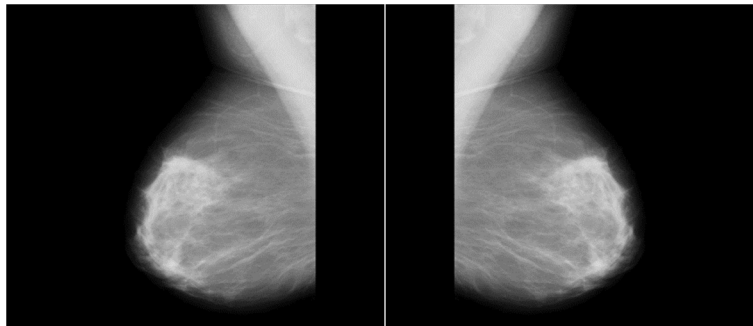

Figure 3 Right Mammogram and Mammogram after Flipping 180°

There are different types of noises, which appear in digital mammogram images. High intensity noise is characterised by high values of optical densities like shadows presenting themselves as horizontal running strips or ghost images of previously performed mammography. These noises are embedded to the breast region of the mammogram thus resulting in loss of information from the breast region. These noises also make detection process for an automated CAD process to yield false results or negative detection. Such noise must be removed from the image to provide accurate results in the detection processes. In this research the well-known Gaussian filter is used to remove such noise by blurring these noises before performing edge detection or other processing on the mammogram images. It is is used to `blur' images and remove details and noise. Gaussian form is used for the proposed method is as follows:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (2)$$

Since the image is stored as a collection of discrete pixels it is needed to produce a discrete approximation to the Gaussian function before it can perform the convolution.   Once a suitable kernel is obtained then the Gaussian smoothing can be performed using standard convolution methods. This is shown in figure 4.
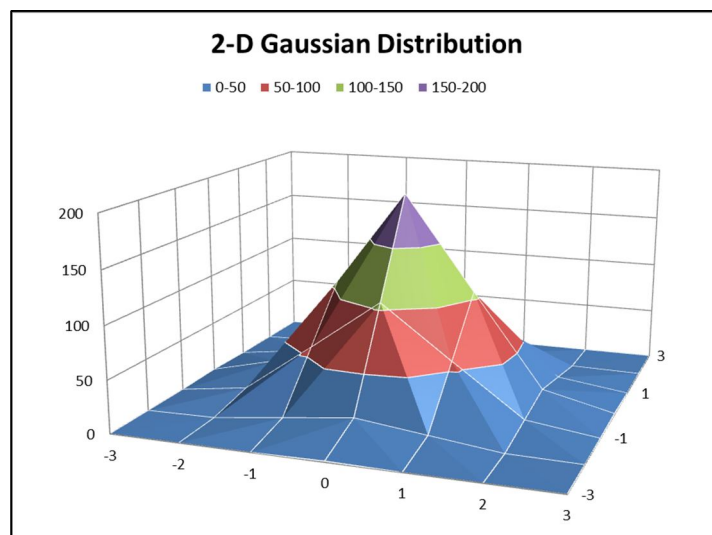
Figure 4 Proposed Method uses 7×7 kernel as a Convolution Filter

In this proposed method, 7×7 kernel has been taken as a convolution filter. The mammogram images are broadly categorized into 3 categories namely Fatty, Fatty-Glandular and Dense-Glandular depending on the density of fatty tissues and abundance of glands in the breast. Close observation of all the mammograms reveal that each category of mammogram displayed a varied intensity value which is distinct for each category. This property of the mammograms has helped the choice of value of deviation (Ω) for each category, thus able to adjust the level of smoothening for each category. Figure 5 shows Mammogram before and after Gaussian Smoothening
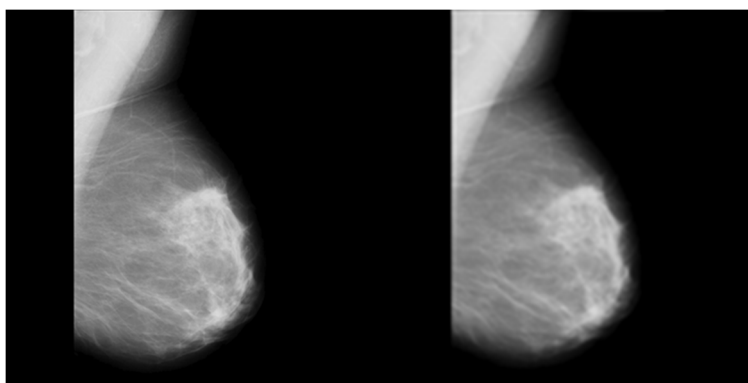

Figure 5 Mammogram before and after Gaussian Smoothening

The algorithms have been tested with several mammographic images including on all mammograms from MIAS mammogram database and other available databases containing normal and abnormal cases. Almost all cases output is as expectation. Some of the appropriate test results are depicted in figure 6 with the mammograms taken from the MIAS database to prove the accuracy of the algorithm.
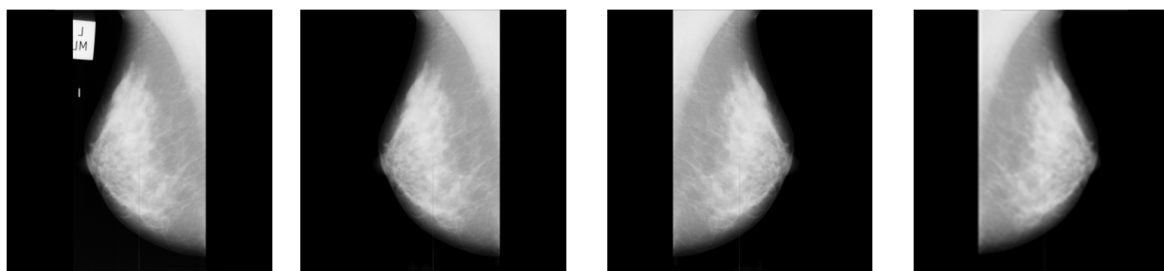

Figure 6 Images showing Original Mammogram followed by image after Artefacts Removable, Flipping and Noise Removable (Images from left to right)

The preparation includes three broad areas namely, artefact removal, flipping of right sided breast and noise elimination. All the three process is done automatically by the system without any user intervention which is a prerequisite for any real-time system. Very few authors included this vital step of preparation in their dissertations. Noise elimination has been done with standard Gaussian kernel. Flipping method has been successfully done by the proposed method except for 5 images in MIAS database where there are operator induced errors are present. The Accuracy of the proposed method is 100%. It can be noted that flipping may not be relevant to other methods proposed by different authors but for the newly proposed method, it is of vital importance as registration of mammogram pairs depends on the success of flipping.

Flipping failed in these few cases are due to the presence of some shadow or vertical high intensity line, as noise on one of the sides of the mammogram. This line mimics the chest wall so the algorithm fails to distinguish a right breast mammogram, falsely interprets it to be a left mammogram and it does not perform flipping. Artefact removal method is implemented on 322 mammograms but showed failure in 8 cases and falsely classified two mammograms of artefact where there was none present. A detailed statistical analysis has been performed using Receiver operating characteristic (ROC) analysis. ROC methodology is a popular method for comparing the performances of two or more imaging modalities. ROC is a binary concept where the object is present or not present and the result is in binary. The resulting 2 x 2 truth-response table defines positive decisions (true positives, true negatives) and negative decisions (false positives and false negatives). Using the obtained ROC quantities one can define True Positive Fraction (TPF), False Positive Fraction (FPF) and resultant ROC curve is shown in figure 7.

The findings of ROC curve in total number of cases 322, number of correct cases is 312 with Accuracy of 96.9%, Sensitivity of 99.1% and Specificity value of 91.9%. Total positive cases missed is 2 and negative cases missed is 8. The Empiric ROC curve area enclosed is 0.955.
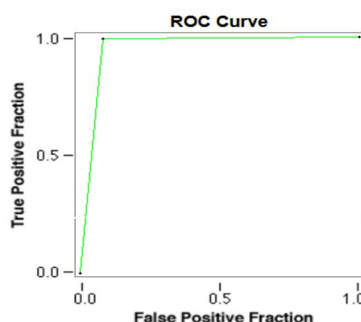

Figure 7 Empirical ROC Curve for Artefact Removal

IV. CONCLUSIONS

In this proposed method, three distinct algorithms are used and the combinations of these have provided elimination of different artefacts, making homogeneous orientation and excellent removal of noise from mammogram images. The fully automated and accurate elimination was achieved by the proposed Seeded Region Growing Algorithm (SRGA). This algorithm removes the external artefacts in most cases. This process is followed by changing the orientation of mammogram image to get uniform mammogram image for both left and right pairs of mammogram. Finally, Gaussian smoothening is used to remove noise that is internal to the breast region. The output of this preparation processing are the mammograms that are free from most of the artefacts and noise; can be used for other medical image processing applications and further studies on mammogram for detection of abnormalities.

REFERENCES

[1] Chaloeykitti et al, "Artefacts in mammography: ways to identify and overcome them", Singapore Medical Journal, Vol. 47, Issue-7, pp. 634–640, 2006

[2] Williams et al, "Image quality in digital mammography: image acquisition", Journal of the American College of Radiology, Vol. 3, Issue-8, pp. 589–608, 2006

[3] Yaffe et al, "Quality control for digital mammography. II. Recommendations from the ACRIN DMIST trial", Medical Physics, Vol. 33, Issue-3, pp. 737–752, 2006

[4] Rama et al, "Digital Mammographic Artefacts on Full-Field Systems: What Are They and How Do I Fix Them?", Radio Graphics, Vol. 28, Issue-7, pp. 1999-2009, 2008

[5] American Cancer Society, "Cancer Facts & Figures", Atlanta: American Cancer Society, 2012

[6] Bahlmann et al, "Automated Detection of Diagnostically Relevant Regions in H&E Stained Digital Pathology Slides", Proceedings SPIE 8315 Medical Imaging 2012, Computer-Aided Diagnosis, 23 Feb, 2012

[7] Fung et al, "Mammographic Determination of Breast Volume by Elliptical Cone Estimation", World Journal of Surgery, Vol. 34, Issue-7, pp.1442–1445, 2010

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY