

Detecting Architectural Distortion in Mammograms Using a Gabor Filtered Probability Map Algorithm

O'tega Ejofodomi, Michael Olawuyi, Don Uche Onyishi, and Godswill Ofualagba

Department of Electrical Engineering, Federal University of Petroleum Resources, P.M.B.
1221, Effurun, Nigeria
tegae@yahoo.com

Abstract. Breast Cancer is a disease that is prevalent in many countries. Computer-Aided detection (CAD) systems have been developed to assist radiologists in detecting breast cancer. This paper discusses an algorithm for architectural distortion (AD) detection with a better sensitivity than the current CAD systems.

19 images containing ADs were preprocessed with a median filter and Gabor filters to extract texture information. AD probability maps were generated using a maximum amplitude map and histogram analysis on the orientation map of the Gabor filter response. AD maps were analyzed to select ROIs as potential AD sites.

AD map analysis yielded a sensitivity of 79% (15 out of 19 cases of AD were detected) with a false positive per image (FPI) of 18. Future work involves the development of a second stage in the algorithm to reduce the FPI value and application of the algorithm to a different set of database images.

Keywords: Breast Cancer, Architectural Distortion, Gabor Filters.

1 Introduction

Breast cancer is a disease that is prevalent in many countries. In the United States of America, breast cancer is the most common type of cancer, side from skin cancer. The American Cancer Society estimates that there will be 234, 580 new cases of breast cancer (232,340 female, 2,240 male) and that 40,030 people (39,620 female, 410 male) will die from breast cancer in 2013 [1]. In the United Kingdom, it is estimated that there were 49,961 new cases (49,564 female, 397 male) of breast cancer, and 11, 633 deaths (11,556 female, 77 male) related to breast cancer in 2010 [2]. In some countries, the statistics for breast cancer is unreliable.

The current gold standard for detecting breast cancer is screening mammography, which is X-ray imaging of the breast. Mammography can often detects breast cancer at an early stage, when treatment is more effective and a cure is more likely, especially for women over 50 years. This reduces the mortality rate of the disease. Numerous studies have shown that early detection with mammography saves lives and increases treatment options [1, 3-4]. In mammography, two types of X-ray images are usually obtained: the Mediolateral Oblique (MLO) view, which is a sideways view of the breast, and the Craniocaudal view (CC), which is a top to bottom view of the breast.

Radiologists search each image for signs of lesions and abnormalities that may represent breast cancer. Typically, they search for masses, calcifications and architectural distortions. Breast lesions are described and reported according to the Breast Imaging Reporting and Data System (BIRADS™), which is a terminology developed by the American College of Radiology (ACR), for the description of mammographic lesions [5]. A mass is defined as a space-occupying lesion seen in at least two different projections. Calcifications are calcium deposits in the breasts that show up as white, bright spots in mammograms. Architectural distortions (AD) are abnormal breast lesions in which the breast parenchyma is distorted with no visible mass. The distortion includes spiculations radiating from a point, and focal retraction or distortion of the edge of the parenchyma [5].

Breast cancer detection using mammograms is far from perfect. It is estimated that radiologists fail to detect about 10-30% of breast cancers [6-8]. Computer-Aided Detection (CAD) systems has been developed to enable radiologists detect lesions that may be indicative of cancer. These systems serve as support to the radiologist and are not used by themselves as a breast cancer detection tool. Commercial CAD systems accurately identified 90% cases of masses and microcalcifications [9, 10]. However, the sensitivity for AD is low. The R2 Image Checker system successfully identified 49% cases containing ADs with 0.7 false positives per image. The CADx Second Look system [9] successfully detected 33% cases of ADs with 1.27 false positives per image [11].

There is need to improve the detection of ADs in CAD systems. The goal of this work is to develop an algorithm that can detect ADs with higher sensitivity than the currently existing CAD systems. Details of the algorithm are presented in the methods section. The results of the algorithm are shown in the results section. Analysis of the performance of the algorithm is provided in the discussion section. Future improvements to the algorithm are highlighted in the conclusion section.

2 Materials and Methods

2.1 Materials

All of the images used in this study were obtained from the Mammographic Image Analysis Society (MIAS) Digital Mammogram Database [12]. All the images were digitized at 50 micron pixel edge. The sizes of all images were 1024 by 1024 pixels. There were a total of 19 images containing ADs. The pixel coordinates of the AD sites, the pathology of the AD, and the nature of the background breast tissue are also provided. Image views were both MLO and CC views of the breast. These 19 images were used in this pilot study. Figure 1 shows a MIAS image with a circle marking the site of the AD. The ground truth provided by the database was used in determining the effectiveness of the algorithm.

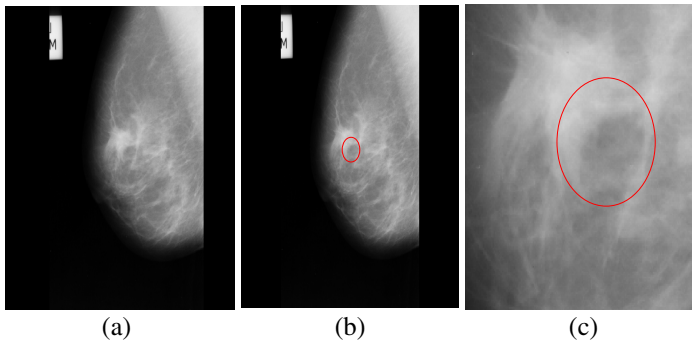


Fig. 1. (a) Original MIAS image (b) Ground truth: Original image with a boundary containing AD overlaid. (b) Enlarged ground truth image.

2.2 Methods

Most of the reported automated detection of AD in literature consists of several steps. There may be some preprocessing to remove noise from the mammograms and to select the breast region within the mammogram. The next stage involves the extraction of the texture orientation (line structures oriented at different angles within the image). The texture orientation is then analyzed to produce regions of interest (ROI) which are potential AD sites. Estimation of the characteristic features of ADs are then used to narrow down the number of potential AD sites, thereby reducing the number of false positives in automated detection. The methodology employed in this work consists of four major steps: preprocessing, Gabor filtering to extract texture information, generation of AD maps, and finally, analysis of AD maps to select ROIs. The flow chart for the algorithm is given in Figure 2.

2.2.1 Image Preprocessing

The images were filtered with a median filter to reduce the noise in the image. No region growing methods were employed to segment the breast region from the background.

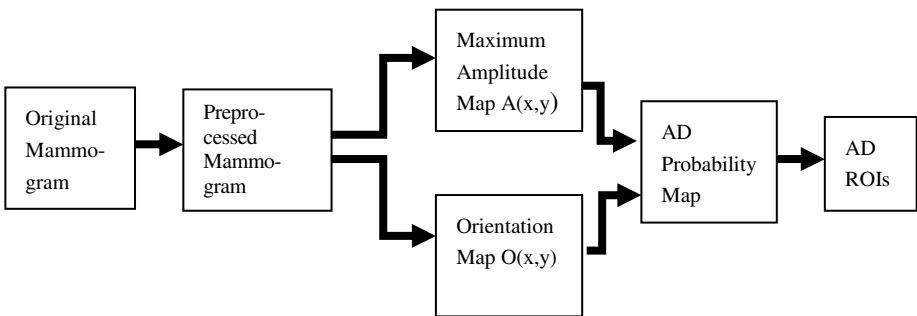


Fig. 2. Flow Chart for AD probability map algorithm

2.2.2 Gabor Filtering

There are many different techniques employed to extract the texture orientation from an image. The method used in this study is the Gabor Filters. The Gabor filter is a sinusoidally modulated Gaussian function. Mathematically, a 2D Gabor function g , is the product of a 2D Gaussian and a complex exponential function. The general expression is given by:

$$g_{\theta,\lambda,\sigma_1,\sigma_2}(x, y) = \exp -1/2\{x \ y\}M \{x \ y\}^T \exp\{j\frac{\pi}{\lambda}(x \cos\theta + y \sin\theta)\} \quad (1)$$

where $M = \text{diag}(\sigma_1^{-2}, \sigma_2^{-2})$. The parameter θ represents the orientation, λ is the wavelength, and σ_1 and σ_2 represent scale at orthogonal directions. When the Gaussian part is symmetric, we obtain the isotropic Gabor function:

$$g_{\theta,\lambda,\sigma}(x, y) = \exp\{-x^2 + y^2/2\sigma^2\} \exp \{j\frac{\pi}{\lambda}(x \cos\theta + y \sin\theta)\} \quad (2)$$

A bank of Gabor filters was applied to the images. The angle of orientation in the bank of filters was varied from $\theta \in [-\pi/2;\pi/2]$, the frequency of sinusoidal function was set to $f = 0.022$, and the variance S was set to 0.05. These parameters were chosen experimentally. A total of 19 Gabor filters were applied to the images.

2.2.3 AD Map Generation

Two maps were created from the output of the Gabor filtration: a maximum amplitude map $A(x, y)$ and a dominant orientation map $O(x,y)$. The maximum amplitude for each point (x,y) in the responses to the Gabor filters was extracted to form the amplitude map $A(x,y)$. The angle of orientation of the maximum amplitude was considered to be the dominant angle and was used in creating the orientation map $O(x,y)$. Both maps are used to create the final Architectural Distortion map $AD(x,y)$.

To create the AD map, histogram analysis was performed on $O(x,y)$. Both the amplitude and the orientation map were divided into blocks of 16 by 16 pixels. The maximum amplitude for each block was obtained from $A(x,y)$. Histogram analysis was performed on each block of $O(x,y)$. It is expected that blocks containing ADs should show several dominant peaks (corresponding to dominant orientation angles), while blocks of normal tissue should only have one dominant peak. Therefore, in the histogram analysis an experimentally determined threshold T_1 was set to determine if a peak is dominant or not. If a peak was within a certain percentage of the maximum orientation angle, it was classified as a dominant orientation angle.

The AD map was created by multiplying the maximum grayscale amplitude in block with the number of dominant angles and then dividing by a characteristic number N . The characteristic number was established experimentally. Mathematically,

$$AD_i = \frac{Dn_i * A_{max}}{N} \quad (3)$$

Where i is the index number of the block, Dn is the number of dominant angles in block i , A_{max} is the maximum grayscale amplitude in block i , and N is the characteristic number. AD is a value that quantifies the probability of the presence of AD in block i [13].

2.2.4 AD Map Analysis

In the final stage, the AD maps were analyzed. The blocks with the highest AD values were selected as regions of interest (ROI). The image containing ROIs was then compared with the ground truth mammogram. A true positive (TP) was recorded when a ROI was within the region identified as AD by the ground truth. A false positive (FP) was recorded when a ROI was not within the region identified as AD by the ground truth.

3 Result

Our data consisted of the 19 images identified with AD in the MIAS database. Table 1 shows the number of true positives (TP) and false positives (FP) identified in each image. Images in which ADs were unidentified have 0 TPs. In one case (Image 18) the AD has been marked twice (2 TPs). We achieved a sensitivity of 78.9% (15 cases of AD were detected) with a false positive per image (FPI) of 18.4 (see Table 2). 10 out of 19 ADs were classified as malignant, while 9 was classified as benign AD in the ground truth data. Of the 10 malignant ADs, the AD Map analysis correctly identified 80% (8 out of 10) malignant ADs, and 78% (7 out of 9) of benign ADs. The total number of ROIs obtained was 291, of which 16 ROIs positively identified AD. The results of the AD Map analysis algorithm are compared with two commercial CAD systems in Table 2. The sensitivity for the AD Map analysis algorithm is much higher than that of the commercial CAD systems. However, the FPI is also relatively high. Most AD detection algorithms employ two stages: Identification of ROIs in the first stage and reduction of the number of FPs in the second stage [13, 14]. The AD map analysis algorithm is currently a one stage detection algorithm. A second stage would need to be included to reduce the FPI to a more reasonable number.

Table 1. True Positives (TP) and False Positives (FP) identified in Images

Image No.	No. of True Positives (TP)	No. of False positives (FP)
1	1	41
2	1	5
3	1	5
4	1	10
5	1	17
6	0	>> 41
7	1	10
8	1	17
9	1	40
10	1	40
11	1	33
12	1	13
13	1	19
14	0	>> 41
15	0	>> 41
16	1	11
17	0	>>41
18	2	7
19	1	7

Table 2. Comparison of AD Map Analysis Algorithm to Commercial CAD systems

	AD Map Analysis	Image Checker R_2	CADx Look	Second
Sensitivity (%)	79	38	21	
No of Images	19	45	45	
FPI	18	0.7	1.27	
% of malignant ADs detected	80%	-	-	
% of benign ADs detected	70%	-	-	
Total No of ROIs	291	-	-	
Total No of ROIs with AD	16	-	-	

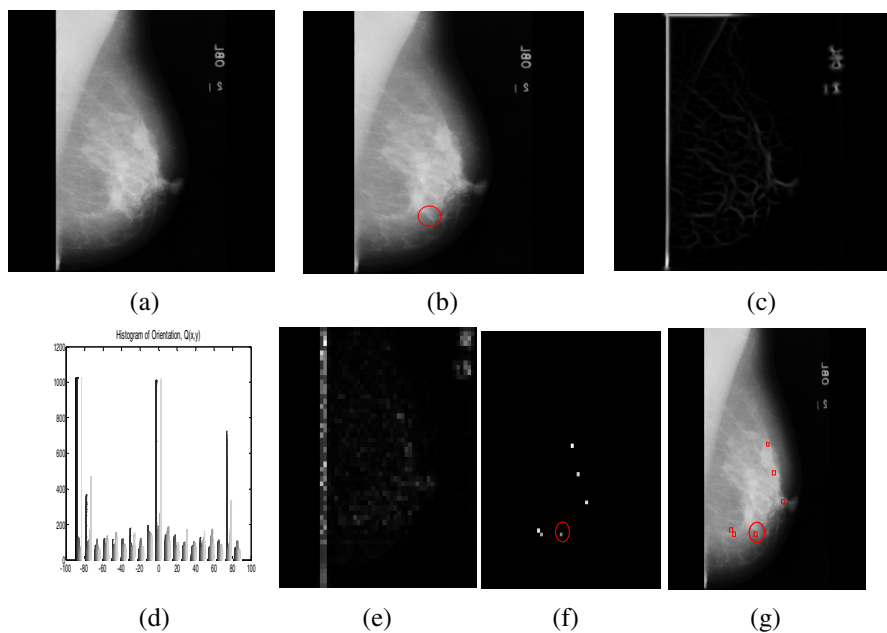


Fig. 3. AD Detection Algorithm. (a) Original MIAS image (b) Preprocessed Image (c) Maximum Amplitude Map $A(x,y)$ (d) Orientation Map $O(x,y)$ (e) AD map (f) ROIs with highest AD values (g) AD detection results overlaid over original image.

Figure 3 shows the outputs of the algorithm for one of the 19 cases. The original image from the MIAS database is shown in Figure 3a. The median filtered image is shown in Figure 3b. The maximum amplitude map $A(x,y)$ obtained after Gabor filtering is shown in Figure 3c. The histogram of the orientation map $O(x,y)$ is shown in

Figure 3d. The generated AD map is shown in Figure 3e. The selected ROI from the AD map analysis is shown in Figure 3f. Finally, the selected ROIs overlaid over the original image is shown in Figure 3f. The ground truth is also labeled in Figure 3f. In this particular case, one of the ROIs is located within the region identified as AD by the ground truth.

4 Discussion

The sensitivity of the Gabor filtered probability map algorithm is far greater than that of the two commercial CAD systems shown in Table 2. The high number of FPI can be attributed to several factors. First as was duly noted, no segmentation to separate the breast region was performed on the images in the preprocessing step. Non-identification of the breast boundary (skin-air) caused the algorithm to flag ROIs outside of the breast region. This influenced the number of FPI in each image. The boundary between breast and film showed up consistently as a bright vertical line in the AD maps (see Figure 3e). This had to be excluded in the AD map analysis. Secondly, the pectoral muscle in the MLO view shows up as a bright white region. Since the algorithm relies on the fact that AD sites possess maximum grayscale values, the pectoral muscle region is often flagged as a ROI even when AD is not present. Thirdly, images in the MIAS database contain labels identifying the image as either an MLO view or CC view (see Figure 1a). These labels were imprinted on the films at different locations and so the algorithm did not remove them in the preprocessing step. These also contributed to the high FPI.

There are several things that can be done to improve the AD probability map algorithm. Most importantly, a second stage needs to be incorporated into the algorithm to reduce the number of FPI. This can be accomplished by filtering and thresholding in the Radon Domain, feature extraction and classification of ROIs [12, 13]. The number of filters in the Gabor banks can be increased to include more orientation angles. A breast segmentation algorithm can also be implemented in the preprocessing stage to extract the breast region from the surrounding to reduce the FPI due to the labels and the breast-film boundary. Finally, it should be noted that the algorithm has been applied to a single dataset from the MIAS database. The results could vary if applied to a different set of images.

5 Conclusion

We have developed an algorithm to detect ADs in mammograms. The Gabor filtered probability map algorithm has a sensitivity of 79% and an FPI of 18, more than two times that of commercial CAD systems. Future work involves the development of a second stage in the algorithm to reduce the FPI value and application of the algorithm to a different set of database images.

References

1. American Cancer Society: Cancer Facts and Figures (2013). American Cancer Society, Inc., <http://www.cancer.org> (accessed February 28, 2013)
2. Cancer Research, UK., <http://www.cancerresearchuk.org> (accessed February 28, 2013)
3. Magnus, M.C., Ping, M., Shen, M.M., Bourgeois, J., Magnus, J.H.: Effectiveness of mammography screening in reducing breast cancer mortality in women aged 39-49 years: a meta-analysis. *Journal of Women's Health* 20(6), 845-852 (2011)
4. Mandelblatt, J.S., Cronin, K.A., Bailey, S., et al.: Effects of mammography screening under different screening schedules: model estimates of potential benefits and harms. *Annals of Internal Medicine* 151(10), 738-747 (2009)
5. American College of Radiology ACR BI-RADS - Mammography, Ultrasound & Magnetic Resonance Imaging, 4th edn. Reston, VA (2003)
6. Kerlikowske, K., Carney, P.A., Geller, B., Mandelson, M.T., Taplin, S.H., Malvin, K., Ernster, V., Urban, N., Cutter, G., Rosenberg, R., Ballard-Barbash, R.: Performance of screening mammography among women with and without a first-degree relative with breast cancer. *Annals of Internal Medicine* 133, 855-863 (2000)
7. Kolb, T.M., Lichy, J., Newhouse, J.H.: Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. *Radiology* 225, 165-175 (2002)
8. Bird, R.E., Wallace, T.W., Yankaskas, B.C.: Analysis of cancers missed at screening mammography. *Radiology* 184, 613-617 (1992)
9. The CADx Second Look system, <http://www.icadmed.com> (accessed February 20, 2013)
10. The R2 Technology's Image Checker, <http://www.r2tech.com> (accessed February 20, 2013)
11. Sampat, M.P., Markey, M.K., Bovik, A.C.: Computer-aided detection and diagnosis in mammography. IEEE Publication (2004)
12. Suckling, J., et al.: The Mammographic Image Analysis Society Digital Mammogram Database. In: *Excerpta Medica. International Congress Series*, vol. 1069, pp. 375-378 (1994)
13. Jasionowska, M., Przelaskowski, A., Rutczyńska, A., Wroblewska, A.: A two-step method for detection of architectural distortions in mammograms. In: *Piejka, E., Kawa, J. (eds.) Information Technologies in Biomedicine. AISC*, vol. 69, pp. 73-84. Springer, Heidelberg (2010)