International Journal of Computational Vision and Biomechanics
Vol. 3 No. 1 (January-June, 2017)

MUK PUBLICATIONS
Open Access Publisher

Coarse Segmentation of Suspicious Tissues in Digital Mammogram Images using Bayesian-Based Threshold Estimation

Amr R. Abdel-Dayem[1] and Mahmoud R. El-Sakka[2]

[1]*Department of Mathematics and Computer Science, Laurentian University, Sudbury, Ontario, Canada*

[2]*Computer Science Department, University of Western Ontario, London, Ontario, Canada*

In this paper, we propose a coarse segmentation scheme for highlighting suspicious lesions in digital mammogram images. The proposed scheme is intended to be used in a multi-stage segmentation paradigm for accurate localization of suspicious masses. The major objective of the proposed scheme is to reduce the search space when further stages search for abnormalities. The proposed scheme uses the image histogram to estimate the Bayes threshold that can segment suspicious lesions from normal breast tissues with minimum probability of classification error. We also present a block-based measure that can objectively assess the computer-segmented images, compared with the clinician-segmented ones.

Experimental results over a set of sample images (consists of 50 normal and 50 abnormal cases) showed that the proposed scheme produces accurate highlighting results, compared with the manual results produced by clinicians. It achieves a true positive fraction, a precision and an overlap ratio of 1.0 for the entire fifty abnormal cases (when used in the screening mode). Meanwhile, the 95% and the 99% confidence intervals for the false positive fraction, calculated over the fifty normal cases, are [0.017, 0.183] and [0, 0.209], respectively (when used in the screening mode).

When the proposed scheme is used in diagnosis or follow up mode, we used our block-based measure with 32×32 block size to report the performance of the system. The results shows that the 95% and 99% confidence intervals (calculated over the fifty abnormal images) for the true positive fraction are [0.842, 0.938] and [0.827, 0.953], for the false positive fraction are [0.101, 0.203] and [0.084, 0.219], for the precision are [0.538, 0.691] and [0.514, 0.715], and for the overlap ratio are [0.483, 0.623] and [0.461, 0.645], respectively. Meanwhile, the 95% and 99% confidence intervals for the false positive fraction (calculated over the fifty normal images) are [0.002, 0.078] and [0, 0.09], respectively. However, if we consider all hundred images together, the 95% and 99% confidence intervals for the false positive fraction are [0.062, 0.130] and [0.052, 0.140], respectively.

It is worth mentioning that, the output produced by the proposed scheme represents preliminary estimates that will be fine-tuned using more advanced stages that employ both pattern classification and artificial intelligence techniques (future work).

Key Words—Digital mammogram image, image segmentation, Bayes decision theory, thresholding, block-based performance measure.

I. INTRODUCTION

According to the World Health Organization, more than 1.2 million women worldwide is diagnosed with breast cancer per year. Every year, nearly 21,000 women are diagnosed with breast cancer in Canada alone. Breast cancer is the most frequently diagnosed cancer in women. Early detection of breast abnormalities plays a vital role in increasing the survival rate.

Currently, mammography is the most reliable method for detecting breast abnormalities. Till now, it is the only exam approved by the U.S. Food and Drug Administration (FDA) to screen breast cancer in women who do not show any signs or symptoms of the disease, where the diagnostic accuracy depends on the clinician's experience. In order to increase the accuracy of the diagnostic process, an automated scheme for detecting breast abnormalities that can accurately segment suspicious regions for further investigations is needed.

Segmentation is the process of extracting an object of interest in an image from its background. Binary thresholding is the simplest segmentation technique that converts an image into a binary one. This segmentation method depends on the appropriate selection of the threshold value. Various criteria can be exploited during the threshold selection process [1].

Kobatake *et al*. [2] proposed a scheme for segmenting suspicious masses on digital mammograms. The proposed segmentation scheme was based on the use of an IRIS filter, which has the capability of capturing circular and rounded convex objects even if they have intensities very close to the background. However, the design of the IRIS filter is not a trivial task. The researchers empirically adjusted the filter parameters based on their own database. This makes the system cumbersome when used on different image sets.

Brake *et al*. [3] proposed a segmentation scheme based on a statistical analysis of the gradient-orientation

map of the image. It is assumed that masses appear in mammograms as nearly circular bright regions. Thus in the gradient orientation map, suspicious masses are expected to appear as circular regions with many gradients pointing toward the center. The analysis of the gradient orientation map was carried out over a local window around each pixel. However, the major drawback of this method is the need to select an appropriate window size for optimal detection. Moreover, this scheme is very sensitive to the accuracy of calculating the orientation map and the spatial scale used to compute the derivatives.

Liu *et al*. [4] proposed a segmentation scheme based on multiresolution analysis. The wavelet transform is used to decompose the image into a pyramid of images at different resolutions. Then, several features from the images in different resolutions are extracted using a certain window size. A binary tree classifier is used to generate a probability map of the image. The probability map is filtered using a median filter to eliminate isolated positive responses and then, a low pass filter is used to smooth the probability map. Finally, the probability map is thresholded to segment the image into suspicious and non-suspicious tissues. However, this scheme has several drawbacks. First, like all classification techniques, the success of this method depends on the appropriate selection of the feature vector, as well as the availability of representative training data. Second, the proposed scheme is sensitive to the used wavelet filter and the number of decomposition levels. Moreover, the proposed scheme depends on several parameters including the window size used to extract the features, the sizes of both the median and the low pass filters, and the threshold used to segment the probability map. The authors did not include any information on how to adjust those parameters. Hence, the scheme can not be reproduced.

Valverde *et al*. [5] proposed a deformable-based model for segmenting suspicious masses on digital mammogram images. The forces acting on the deformable model vertices are defined such that the points do not cluster on the contour. However, this scheme violates the main objectives of a *Computer Aided Diagnostic* (CAD) system which is to provide assistance in detecting suspicious lesions that may have been missed by the clinician. With this scheme, the clinician has to detect and draw the contours of all suspicious regions and hence, if he misses any region there is no way to compensate for this human error. Moreover this scheme is sensitive to the initial contour estimation. This makes the proposed scheme cumbersome for clinical purposes.

Catanzariti *et al*. [6] proposed a segmentation scheme based on Gabor filters, feature extraction and classification. First, the image is decomposed into different frequency bands using a set of Gabor filters. From these images, various features are extracted. Finally, a neural network is used to classify the image pixels into suspicious and non-suspicious regions. However, designing Gabor filters is not a trivial process. The filter choice and the features' selection will have a crucial effect on the results. Moreover, the neural network parameters and the choice of the training data will affect the overall system performance.

Abdel-Dayem *et al*. [7] proposed a scheme for detecting suspicious regions in digital mammogram images. The proposed scheme is based on image thresholding, where the optimal threshold is determined by minimizing the fuzzy entropy of the image. While this scheme produces excellent results for abnormal breast tissue cases, it produces fair results when dealing with normal cases.

Various promising schemes [8] [9] [10] [11] [12] [13] were introduced. They employ different strategies (e.g., neural networks, pattern classification, wavelets, fractal modeling, and stochastic modeling) to extract breast abnormalities. While each scheme has its own merit, it suffers from various shortcomings. We are not going into a detailed discussion regarding those schemes, as we do not intend to compete with any of them. Our goal is to propose a coarse segmentation scheme for preliminary extraction of a limited small set of candidate suspicious lesions in mammogram images for further comprehensive processing. Our scheme can be viewed as a preprocessing stage, which can be incorporated into other advanced detection schemes, aiming at reducing the search space for improved performance. The scheme uses the image histogram to estimate the Bayes threshold value that can identify suspicious areas. It is fully image dependant and automated, where all parameters are calculated based on the image under consideration. Hence, the proposed scheme can be used with any database of images without any customization or user interaction.

The rest of this paper is organized as follows. Section II describes the proposed scheme in more detail. Section III presents the results of the proposed scheme, as well as the block-based performance evaluation measure. Finally, Section IV offers the conclusions of the paper.

II. THE PROPOSED SOLUTION

The proposed scheme consists of five stages, namely: preprocessing to remove markers and other labels from the image, histogram analysis and curve fitting to approximate the image histogram as a summation of Gaussian functions, Bayesian threshold estimation to find a threshold that can discriminate between normal and abnormal tissues with minimal probability of classification error, image thresholding to produce a binary mask image highlighting the locations of suspicious tissues (if any) and finally, contour extraction

to find the contours of the highlighted areas. Fig. 1 and Fig. 2 show the block and the pseudo code of the proposed scheme, respectively. In the following subsections, a detailed description of each stage is presented.

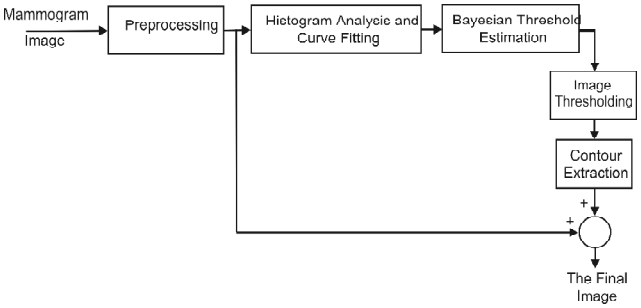


Figure 1: The block diagram of the proposed scheme

Module Name **Suspicious Tissues Detection**
Input Mammogram image *I*
Output Mammogram image with suspicious tissues are highlighted *O*
Variables The *Bayesian_threshold*
T /*The final threshold value*/
Description
Begin
Step 1: *(Histogram Analysis and Curve Fitting)*
Find the histogram of image *I*.
Approximate the image histogram as the summation of 3 Gaussian functions by applying the least square errors curve fitting scheme to the image histogram to get the values of three Gaussian function parameters, a_i , μ_i and σ_i, where i = 1, 2, and 3.
Step 2: *(Bayesian Threshold Estimation)*
Find the *Bayesian_threshold* as the intersection between the 2nd and 3rd Gaussian functions.
IF (*Bayesian_threshold* > μ_2) **and** (*Bayesian_threshold* < μ_3)
Then set *T* ? *Bayesian_threshold*
Else set *T* ? 1+ maximum intensity value of image *I*
Step 3: *(Image Thresholding)*
Threshold the input image *I* using the calculated threshold *T*.
Remove small objects from the segmented image.
Step 4: *(Contour Extraction)*
Find the contour of the segmented area and superimpose it on the original image *I* to generate the output image *O*.
Return (the output image *O*).
End

Figure 2: The pseudo code of the proposed algorithm

(a) Preprocessing

According to Mammography Quality Standards Act (MQSA), all mammogram images should have markers to identify both view and lateral information. These markers, as well as pectoral muscles and chest ribs which appear in some mammographic views, may severely affect any automated processing of mammogram images. Hence, it is crucial to remove these regions prior to any processing. Various schemes [14] [15] [16] [17] [18] [19] found in literature, were proposed to address this challenging task. However, this direction is outside the scope of our paper. For simplicity, we manually preprocessed the images in our database to void those regions from contributing to the segmentation process.

(b) Histogram Analysis and Curve Fitting

This stage consists of two parts. First, the probability density function of image pixels is estimated. There are various techniques that can be used [20] [21], including image histogram. Due to its simplicity, we use the image histogram as an estimator for the probability density function. The second part is a smoothing, or curve fitting, of the estimated probability density function.

The inspection of mammogram images reveals that they usually have two or three distinguished ranges of intensity levels. These ranges represent the background of the x-ray film (which appears as the darkest area in the images), the normal breast tissues in the middle of the image histogram and finally, suspicious regions (if they exist) appear as the brightest areas in the mammogram image. Hence, the histograms of mammogram images are expected to have bimodal (normal mammogram images) or trimodal (suspicious mammogram images) shapes. This fact motivates the use of curve fitting techniques to approximate the histogram data as a weighted sum of positive Gaussian-shaped functions, each with a different mean and variance. In the proposed scheme we assume that the image histogram can be approximated using three Gaussian functions as shown in (1):

$$\hat{h}(x) = \sum_{i=1}^{3} a_i \times N(x, \mu_i, \sigma_i) \tag{1}$$

where $\hat{h}(x)$ is the approximation of the image histogram, a_i is a scaling parameter, and $N(x, m_i, \sigma_i)$ is a Gaussian function with mean m_i and standard deviation s_i defined by:

$$N(x, \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} \times e^{\frac{(x-\mu_i)^2}{-2\sigma_i^2}} \tag{2}$$

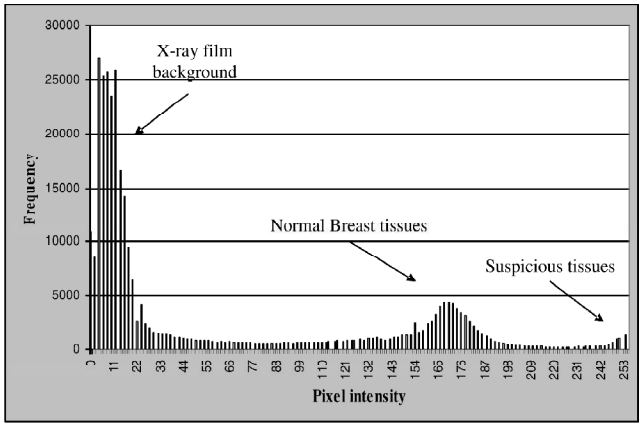
where

$i = 1$ in the darkest area of the image histogram,

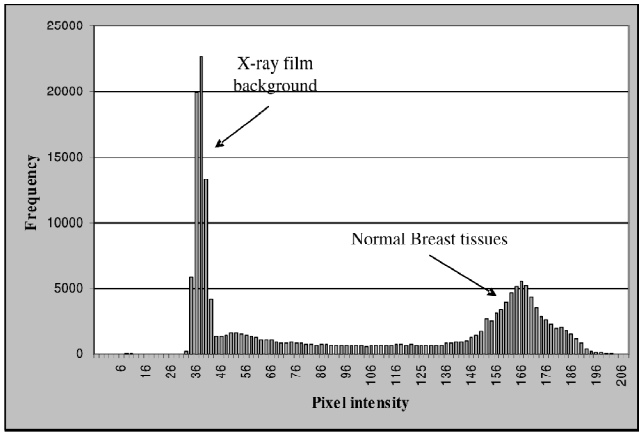
$i = 2$ in the mid tone area of the histogram, and

$i = 3$ in the brightest area of the histogram.

This assumption will be validated in Section III using the correlation coefficient. Note that, the use of three Gaussian functions is suitable for approximating both trimodal and bimodal shapes. In the bimodal case, our model will force the data to be clustered into three classes. As the first Gaussian represents homogeneous area (the background of the x-ray film), the algorithm tends to split the area representing normal breast tissues into two overlapping classes, where their means will be very close to each other and they will be located at the same side of their intersection point. The scheme checks this condition to distinguish between normal and abnormal images and hence, it reacts accordingly. Fig. 3(a) shows a typical histogram of a mammogram image which has a trimodal shape, whereas Fig. 3(b) shows an example of a bimodal shape. It is worth mentioning that, theoretically, it is



(a)



(b)

Figure 3: Typical histograms for mammogram images: (a) trimodal shape histogram; (b) bimodal shape histogram

possible to extract the breast area only and neglect the background of the x-ray film. In this case, the image histogram can be modeled as a summation of two Gaussian functions instead of three. However, this requires the use of an *efficient* automatic breast border extractor. To avoid, these complications, our scheme deals with the whole image, where the background of the x-ray film is modeled as a Gaussian function. After the curve fitting stage, the scheme focuses on the second and the third Gaussian functions and neglects the first one (which corresponds to the background of the x-ray film).

Various curve fitting techniques can be used to fit the histogram data [22] [23]. In the proposed scheme, we used the Least Squares Error (LSE) curve fitting technique [22] [23], however, any other curve fitting scheme can be used. In LSE, the parameters a_i, m_i and s_i are selected to minimize the following cost function:

$$C = \sum_{x=0}^{L-1} [h(x) - \hat{h}(x)]^2 \qquad (3)$$

where $h(x)$ is the actual histogram function, $\hat{h}(x)$ is the approximation function defined in (1), and L is the number of intensity levels in the mammogram image. Minimizing (3) requires that:

$$\frac{\partial C}{\partial a_i} = 0, \qquad (4)$$

$$\frac{\partial C}{\partial \mu_i} = 0, \qquad (5)$$

and

$$\frac{\partial C}{\partial \sigma_i} = 0. \qquad (6)$$

This produces a set of nonlinear equations that can be numerically solved. In the proposed scheme, we used the TABLE CURVE package which has several built-in curve fitting functions as defined in [24].

(c) Bayesian Threshold Estimation

Bayesian decision theory is an important approach in the field of statistical pattern classification [25] [26]. Consider a classification problem in which there are two classes, ω_1 and ω_2. The *a posteriori* probability $P(\omega_i/x)$ measures the probability that the state of nature being É$_i$ given that the value x has been measured. For a given decision boundary α, the probability of the classification error is calculated according to the following equation:

$$\begin{aligned} P(error) &= P(\text{ } x \text{ is assigned to the wrong class}) \\ &= P(\text{assigning } x \text{ to } \omega_1, \text{ while } x \text{ actually belongs to } \omega_2) \\ &+ P(\text{assigning } x \text{ to } \omega_2, \text{ while } x \text{ actually belongs to } \omega_1) \end{aligned} \qquad (7)$$

Assuming that the *a posteriori* probability functions for both classes are known, the classification error probability for a certain decision boundary ± can be calculated as shown in Fig. 4. According to Bayesian decision theory, the probability of error can be minimized by moving the decision boundary to the point of intersection of the two *a posteriori* probability functions (point x^* in Fig. 4). This point represents the Bayes decision boundary.

The objective of this stage is to estimate a threshold that can discriminate normal breast tissues, which are approximated by the second Gaussian function in (1), from suspicious regions, which are approximated by the third Gaussian function in (1), if any. The threshold is selected to be the point of intersection between these two Gaussians. This threshold represents the Bayes decision boundary that produces the minimum probability of classification error between normal and suspicious tissues.

In case of normal images, the means of the second and the third Gaussian functions should be located at the

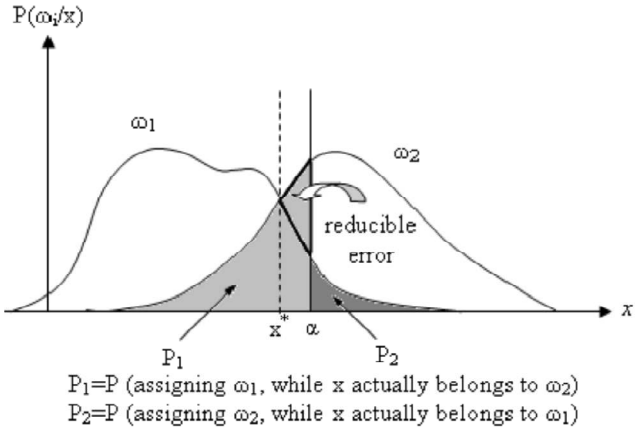


Figure 4: The probability of classification error for a non-optimal decision boundary ±. If the decision boundary is moved to x*, the area labeled as reducible error is eliminated. This point represents the Bayes decision point

same side of their intersection point (the hypothetical decision boundary) as shown in Fig. 5. The proposed scheme checks for this condition and if it is satisfied, the threshold is set to a value that is greater than the maximum intensity level of the mammogram image. Hence, the thresholding operation will not yield any suspicious regions.

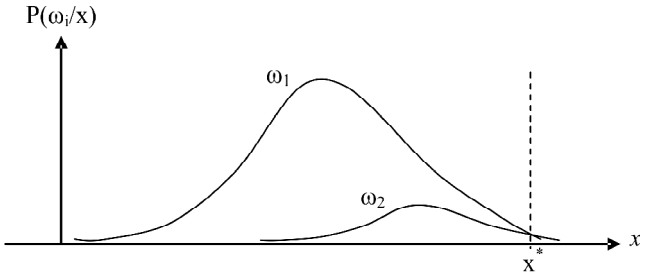


Figure 5: A sample of two Gaussian functions in which their means are located at the same side of their intersection point. This is what happens when dealing with normal mammogram image cases

(d) Image Thresholding

In this stage, the original mammogram image is thresholded using the threshold (which is determined in Section II.C) to produce a binary image that highlights the locations of suspicious tissues.

Analysis of the mammogram images revealed that very small isolated regions are usually irrelevant. Fig. 6 shows a mammogram image where different block sizes were highlighted. From this figure, we can conclude that regions smaller than sixteen pixels (4×4 blocks) have higher probability of being artifacts than being suspicious lesions. It is worth mentioning that these small regions are smaller than the clinician's resolution, which will be empirically determined in Section III.

In our scheme we group pixels in the segmented image into connected objects and produce a suspicious objects' table containing information about image objects. This information includes a pixel to represent each object and the number of pixels in each object. Table entries that correspond to objects smaller than a certain threshold value (16 in our experiments) are eliminated from the suspicious object's table.

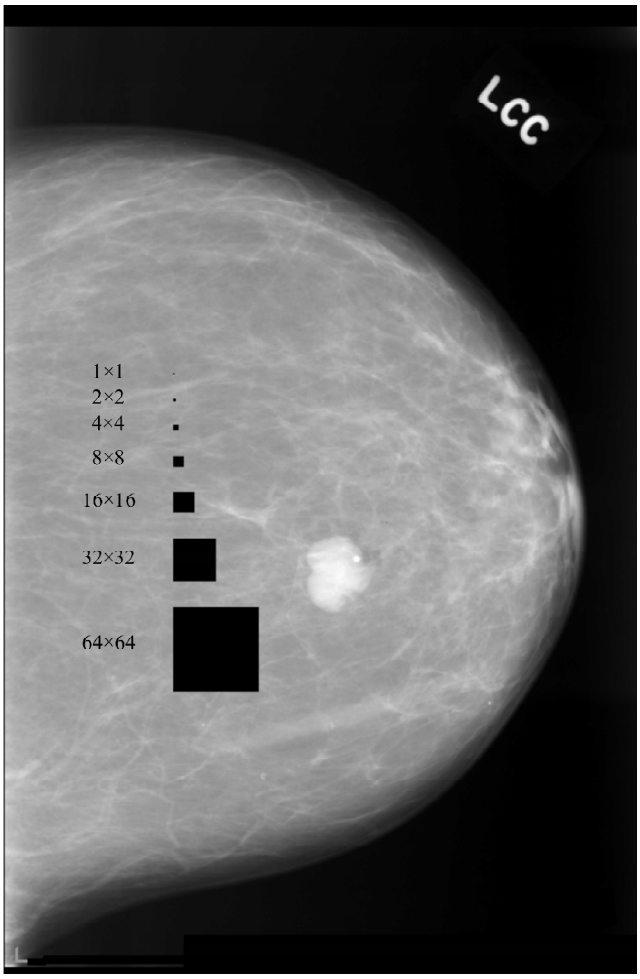


Figure 6: Blocks with various block sizes superimposed over a mammogram image

(e) Contour Extraction

The objective of this stage is to extract the boundary of the area that is segmented in the previous stage (Section II.D). We used a morphology edge detector [27] [28]. First, the segmented area is morphologically eroded using a 3×3 rounded structuring element. Then the eroded image is subtracted from the segmented image to produce the contour of the segmented area. This operation can be described by the following equation:

$$Boundary\ (A) = A - (A \ominus B), \quad (8)$$

where A is the segmented image, B is a 3×3 rounded structuring element, and $\ominus$ is the erosion operation. Finally the contours of the segmented areas are superimposed on the original image to highlight the suspicious masses in the mammogram image.

III. RESULTS

(a) Experimental Setup

The scheme is implemented using C language. The test images used in this research were obtained from the USF Digital Database for Screening Mammography (DDSM) [29]. These images were digitized and saved in a compressed raw format. The database contains different sets of normal and abnormal images collected from patients belonging to different race groups. These images come from two different hospitals; namely Massachusetts General Hospital and the Wake Forest University School of Medicine. Table 1 shows the percentage breakdown of patients by race. The database also offers gold standard images which were segmented by experienced clinicians and validated by biopsy tests. In our experimentations, we divided the images into two sets. The first set consists of fifty images that contain masses with different sizes. The second set consists of fifty images representing normal breast images. The proposed scheme can be used with both *Cranio-Caudal* (CC) and *Mediolateral-Oblique* (MLO) views. However, in case of MLO views, a breast boundary extractor is needed, as a pre-processing stage, to remove the areas representing the chest muscles which usually appear in these views. For simplicity, we used CC views only in our tests.

Table 1
The Percentage Breakdown of Patients by Race [29]

Group Name	*Massachusetts General Hospital*	*Wake Forest University School of Medicine*
Asian	2.06 %	0.2 %
Black	4.12 %	20.4 %
Spanish Surname	6.55 %	1.8 %
American Indian	0.00 %	0.1 %
Other	0.75 %	0.1 %
Unknown	30.34 %	0.3 %
White	56.18 %	77.0 %

We calculated the average correlation coefficient for the entire sets of mammogram images to measure how well the calculated curves fit the original data. The correlation coefficient is defined as:

$$R = \frac{\sum_i (x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2} \times \sqrt{\sum_i (y_i - \bar{y})^2}} \qquad (9)$$

where,

y_i is the actual data value,

x_i is the estimate of y_i,

$\bar{y}$ is the mean of actual data values, and

$\bar{x}$ is the mean of estimated values

The correlation coefficient value ranges from -1 to 1. The closer the absolute value of the correlation coefficient to 1, the better the fit is. We found that the average absolute correlation coefficient for the entire sets of mammogram images approached 0.93 (the minimum absolute correlation coefficient = 0.77, while the maximum = 0.99). These high values of the correlation coefficients support our claim that the histograms of mammogram images can be approximated as the summation of three Gaussian functions.

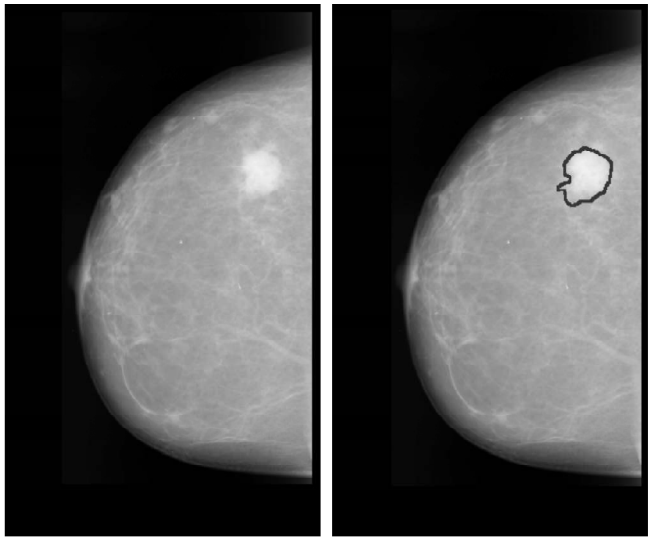
We also found that for the fifty normal images, the means of the second and the third Gaussian functions appear at the same side of their intersection point. Hence, this result validates our claim shown in Fig. 5.

The results produced by our scheme were compared to the clinician segmented images (the gold standard images). This comparison was based on both subjective and objective analysis. Section III.B presents the subjective comparison results for both abnormal (Section III. B.1) and normal cases (Section III. B.2). In Section III.C, we present our proposed block-based objective measure. Section III. C.1 and Section III. C.2 offer analysis for some cases in which the proposed segmentation scheme produced unexpected objective performance assessment in both the abnormal and normal cases, respectively. Section III. D presents general comments about the proposed scheme.

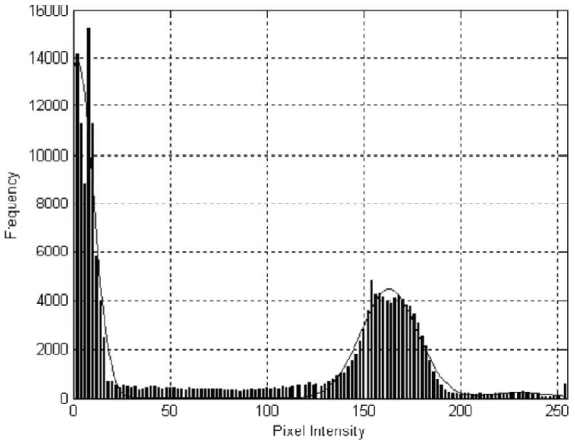
(b) Subjective Analysis

(i) Abnormal Case

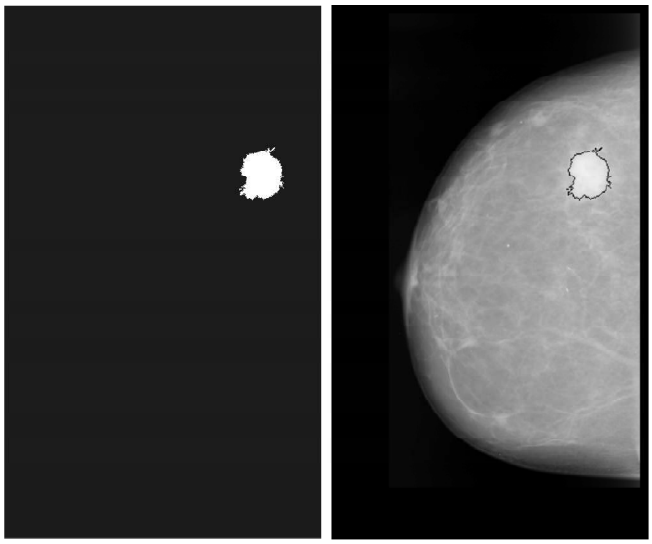
Fig. 7(a) shows one of the mammogram images that were used to test the proposed scheme. Fig. 7(b) shows the manually highlighted image (The gold standard) for the case shown in Fig. 7(a). Fig. 7(c) shows the histogram of the original image with the approximation function $\hat{h}(x)$ produced from the first stage of the proposed scheme (i.e. histogram analysis and curve fitting stage). The second and the third Gaussian functions intersect at intensity value equal to 205.90, which is greater than μ_2 (i.e., 163.22) and less than μ_3 (i.e., 230.29). This means that the decision boundary appears between the means of the two Gaussian functions. Hence this value is used as the Bayes threshold (Bayes decision boundary). Fig. 7(d) shows the segmented suspicious area produced by thresholding the original mammogram image shown in Fig. 7(a) using the calculated Bayes threshold. Fig. 7(e) shows the final output from the proposed scheme, where the input image with the suspicious mass is highlighted. Finally, Fig. 7(f) shows the parameters of the approximation function $\hat{h}(x)$, defined in (1). By comparing Fig. 7(b) and Fig. 7(e), we can subjectively



(a) (b)



(c)



(d) (e)

Parameter	1st Gaussian	2nd Gaussian	3rd Gaussian
a	14062.15	4463.52	243.25
μ	1.97	163.22	230.29
σ	8.17	14.73	15.31

(f)

Figure 7: Abnormal case subjective analysis: (a) the input mammogram image "suspicious case 11"; (b) the mammogram image with the suspicious mass manually highlighted (gold standard image); (c) the histogram of the input mammogram image with the approximation curve, where the second and third Gaussian functions intersect at intensity 205.90 (the Bayes threshold); (d) the segmented mass using the Bayes threshold (t* = 205.90); (e) the final output of the proposed scheme; and (f) the parameters of the function $\hat{h}(x)$ produced by the curve fitting stage

conclude that the proposed scheme was able to accurately highlight the suspicious mass from the mammogram image.

(ii) Normal Case

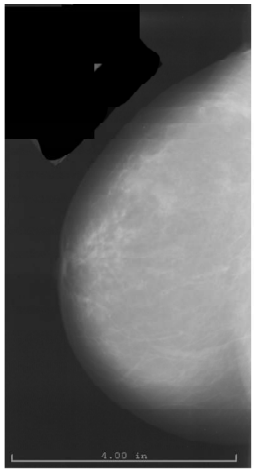
Fig. 8(a) shows a mammogram image for a normal breast. Fig. 8(b) shows the histogram of the image with the approximation function $\hat{h}(x)$. Fig. 8(c) shows the parameters a_i, μ_i, and σ_i, defined in (1). The second and the third Gaussian functions intersect at infinity. The means of both functions (i.e., 167.07 and 188.30) are located at the same side of the intersection point. Hence, the threshold is set to a large number, yielding no suspicious regions in this image.

(c) Objective Analysis

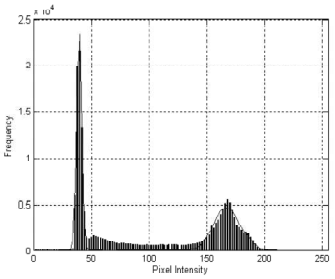
Typically, the terms *true positive (TP)*, *false positive (FP)*, *true negative (TN)* and *false negative (FN)* are used to measure the performance of detecting the presence or absence of an abnormality in an image. From these terms, the *sensitivity*, the *specificity*, the *precision* and the *overlap ratio* of the system are defined as:

$$\textit{Sensitivity (a.k.a., the true positive fraction)} = \frac{TP}{TP + FN} \quad (10)$$

$$\textit{Specificity (a.k.a., the true negative fraction)} = \frac{TN}{TN + FP} \quad (11)$$



(a)



(b)

Parameter	1st Gaussian	2nd Gaussian	3rd Gaussian
a	23793.61	4732.60	418.22
μ	39.56	167.07	188.30
σ	2.26	11.21	4.79

(c)

Figure 8: Normal case subjective analysis: (a) the input mammogram image "normal case 24"; (b) the histogram of the input mammogram image with the approximation curve produced by the first stage of the proposed scheme, where the second and third Gaussian functions intersect at infinity. The means of both functions are located at the same side of the intersection point. Hence, the threshold is set to a large number, yielding no suspicious regions in the image; and (c) The parameters of the function $\hat{h}(x)$ produced by the curve fitting stage

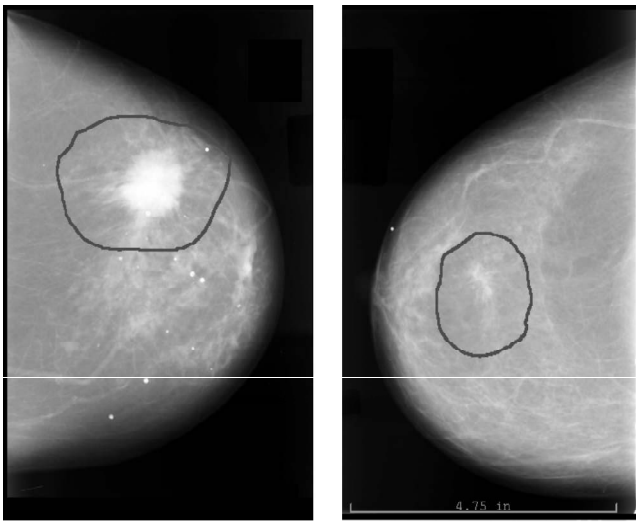
Sometimes the *false positive fraction* is reported instead of the specificity where,

$$False\ positive\ fraction = \frac{FP}{TN + FP} = 1 - Specificity \quad (12)$$

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

$$Overlap\ ratio = \frac{TP}{TP + FP + FN} \quad (14)$$

Unfortunately, there is no standard method for counting the TP, FP, TN and FN terms. They may be considered as image-based (very coarse) or pixel-based (very fine). Image-based measures may be suitable for screening mammograms. However, these methods do not consider the localization of detected masses. On the other hand, using pixel-based measures may provide misleading results. This can be attributed to the fact that most clinicians usually over segment suspicious tumor regions. Fig. 9 shows sample of mammogram images, where suspicious tumor regions were over segmented. This poor annotation negatively affects any pixel-based performance measure. Due to the lack of a standard evaluation method, individual researchers usually attempt



(a) (b)

Figure 9: Sample of mammogram images where the clinician over estimated the size of the suspicious tumors. This poor annotation negatively affect any pixel-based performance measure

to define their own criterion. For example, both Kobatake *et al.* [2] and Catanzariti *et al.* [6] did not mention how they counted the TP, FP, TN and FN terms. Brake *et al.* [3] assumed the tumor was detected if the pixel with the highest degree of suspiciousness was located inside the clinician's annotation. Fig. 10 shows a hypothetical example that illustrates the weakness of Brake's assumption. In this figure, even thought the highest suspicious pixel lies within the clinician's annotation, there are huge areas of both FNs and FPs that were completely ignored in their reported performance measure. Liu *et al.* [4] considered a computer finding as a true positive, if its area is overlapped by at least 50% with the clinician's annotation.

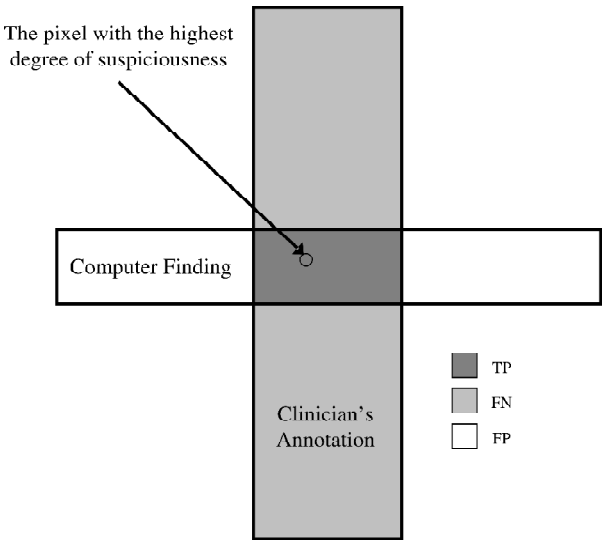
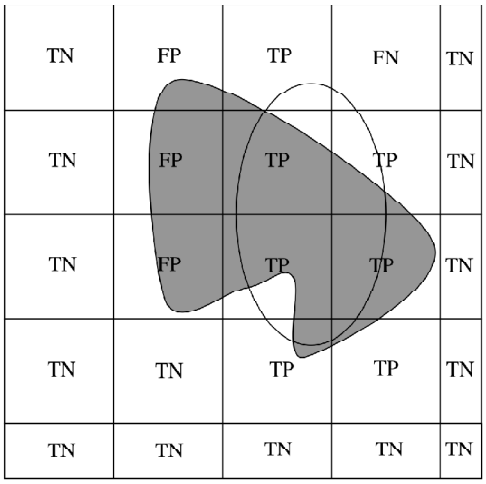


Figure 10: A hypothetical example in which the pixel having the highest degree of suspiciousness lies completely within the area annotated by the clinician. This example demonstrates the weakness in the assumption made by Brake *et al.* [3] to asses the performance of their scheme

The lack of a standard measures makes it difficult to compare between different schemes. Hence, we presented our *block-based* objective performance criterion in an attempt to unify the way of calculating these performance measures. An early version of this *block-based* objective performance criterion was introduced in our previous work [7].

In this measure, both the computer-segmented and clinician-segmented images are divided into *n´n* blocks. A block is labeled as a suspicious block if it contains at least one suspicious pixel. Then, the corresponding blocks in both images are compared. If both blocks are labeled as suspicious blocks, it is considered a *true positive* finding. If a block in the computer-segmented image is marked as suspicious and the corresponding block in the clinician-segmented image is not marked, the block is considered a *false positive* finding. However, if a block in the clinician-segmented image is marked as suspicious and the corresponding computer-segmented block is not marked, the block is considered a *false negative* finding. Other blocks are considered as *true negative* findings. Fig. 11(a) shows a hypothetical graphical example of the performance evaluation labeling process. Note that, at image borders, for simplicity, we consider a part of a block as a complete block (border condition). Fig. 11(b) shows the definition of the TP, FP, TN and FN terms. Note that, if *n* is set to equal one, our measure converges to a *pixel-based* criterion. On the other hand, if *n* is set to equal the entire image size, our measure converges to an *image-based* screening criterion.

We started our analysis by calculating the *true positive fraction* and the *false positive fraction* for our system when operated in the screening mode. In this mode, the main emphasis is on whether the image has suspicious regions or not, without taking into account the localization factor (i.e., the overlap with the gold standard annotations). Table 2 and Table 3 show the TP, FP, TN, FN block counts, the *true positive fraction*, the



(a)

	Computer finding	
Clinician finding	Marked as suspicious	Not marked as suspicious
Marked as suspicious	TP	FN
Not marked as suspicious	FP	TN

(b)

TP = 7 blocks

TN = 14 Blocks

FP = 3 Blocks

FN = 1 Blocks

$$The\ true\ positive\ fraction = \frac{7}{7+1} = 0.875$$

$$The\ true\ negative\ fraction = \frac{14}{14+3} = 0.824$$

$$The\ false\ positive\ fraction = \frac{3}{14+3} = 0.176$$

(c)

Figure 11: The block-based performance measure: (a) a hypothetical example to demonstrate the used performance criteria, where the image is divided into non blocks, the ellipse represents the clinician-segmented area, the shaded area represents the computer findings, and a block is labeled as a suspicious block if it contains at least one suspicious pixel.; (b) definition of the true positive (TP), false positive (FP), true negative (TN) and false negative (FN). Near the image borders, we consider a block part as a complete block for simplicity; and (c) the TP, TN, FP and FN counts, as well as the calculations of the true positive, true negative and false positive fractions for the hypothetical example in (a)

Table 2
The Performance Criterion Applied to Cancer Cases using the Entire Image (in Block Counts, where the Block Size Equals the Entire Image)

Case #	*Case Name*	*TP*	*FP*	*TN*	*FN*	*True Positive Fraction*	*False Positive Fraction*	*Precision*	*Overlap ratio*
1	c1_0020_LCC.pgm	1	0	0	0	1	N/A	1	1
2	c1_0011_RCC.pgm	1	0	0	0	1	N/A	1	1
3	c1_0014_RCC.pgm	1	0	0	0	1	N/A	1	1
4	c1_0016_RCC.pgm	1	0	0	0	1	N/A	1	1
5	c1_0019_RCC.pgm	1	0	0	0	1	N/A	1	1
6	c1_0020_RCC.pgm	1	0	0	0	1	N/A	1	1
7	c1_3022_RCC.pgm	1	0	0	0	1	N/A	1	1
8	c1_3049_LCC.pgm	1	0	0	0	1	N/A	1	1
9	c1_3065_RCC.pgm	1	0	0	0	1	N/A	1	1
10	c1_3066_LCC.pgm	1	0	0	0	1	N/A	1	1
11	c1_3084_RCC.pgm	1	0	0	0	1	N/A	1	1
12	c2_0022_RCC.pgm	1	0	0	0	1	N/A	1	1
13	c2_0023_LCC.pgm	1	0	0	0	1	N/A	1	1
14	c2_0028_LCC.pgm	1	0	0	0	1	N/A	1	1
15	c2_0050_LCC.pgm	1	0	0	0	1	N/A	1	1
16	c2_0051_RCC.pgm	1	0	0	0	1	N/A	1	1
17	c2_0061_RCC.pgm	1	0	0	0	1	N/A	1	1
18	c2_0071_LCC.pgm	1	0	0	0	1	N/A	1	1
19	c2_0073_RCC.pgm	1	0	0	0	1	N/A	1	1
20	c2_0074_LCC.pgm	1	0	0	0	1	N/A	1	1
21	c2_0075_LCC.pgm	1	0	0	0	1	N/A	1	1
22	c2_0079_LCC.pgm	1	0	0	0	1	N/A	1	1
23	c2_0081_LCC.pgm	1	0	0	0	1	N/A	1	1
24	c2_0082_LCC.pgm	1	0	0	0	1	N/A	1	1
25	c2_0086_RCC.pgm	1	0	0	0	1	N/A	1	1
26	c2_0088_LCC.pgm	1	0	0	0	1	N/A	1	1
27	c2_0093_LCC.pgm	1	0	0	0	1	N/A	1	1
28	c2_0094_RCC.pgm	1	0	0	0	1	N/A	1	1
29	c2_0100_RCC.pgm	1	0	0	0	1	N/A	1	1
30	c2_0108_LCC.pgm	1	0	0	0	1	N/A	1	1
31	c2_0111_LCC.pgm	1	0	0	0	1	N/A	1	1
32	c2_0112_LCC.pgm	1	0	0	0	1	N/A	1	1
33	c2_0118_LCC.pgm	1	0	0	0	1	N/A	1	1
34	c2_0121_RCC.pgm	1	0	0	0	1	N/A	1	1
35	c2_0131_RCC.pgm	1	0	0	0	1	N/A	1	1
36	c2_0133_LCC.pgm	1	0	0	0	1	N/A	1	1
37	c2_0137_LCC.pgm	1	0	0	0	1	N/A	1	1
38	c2_3088_LCC.pgm	1	0	0	0	1	N/A	1	1
39	c3_1017_LCC.pgm	1	0	0	0	1	N/A	1	1
40	c3_1027_LCC.pgm	1	0	0	0	1	N/A	1	1
41	c3_1035_RCC.pgm	1	0	0	0	1	N/A	1	1
42	c3_1036_LCC.pgm	1	0	0	0	1	N/A	1	1
43	c3_1050_RCC.pgm	1	0	0	0	1	N/A	1	1
44	c3_1055_LCC.pgm	1	0	0	0	1	N/A	1	1
45	c3_1060_LCC.pgm	1	0	0	0	1	N/A	1	1
46	c3_1084_RCC.pgm	1	0	0	0	1	N/A	1	1
47	c4_1076_RCC.pgm	1	0	0	0	1	N/A	1	1
48	c4_1082_RCC.pgm	1	0	0	0	1	N/A	1	1
49	c4_1089_LCC.pgm	1	0	0	0	1	N/A	1	1
50	c4_1107_LCC.pgm	1	0	0	0	1	N/A	1	1
	Total	50	0	0	0				
	Sample Average					1	N/A	1	1
	Sample Standard Deviation					0	N/A	0	0
	95% confidence interval					[1,1]	N/A	[1,1]	[1,1]
	99% confidence interval					[1,1]	N/A	[1,1]	[1,1]

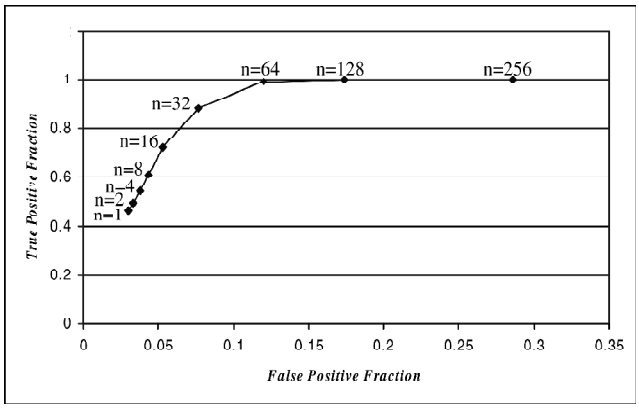
Table 3
The Performance Criterion Applied to Normal Cases using the Entire Image (in Block Counts, where the Block Size Equals the Entire Image)

Case #	*Case Name*	*TP*	*FP*	*TN*	*FN*	*True Positive Fraction*	*False Positive Fraction*	*Precision*	*Overlap ratio*
1	A_0061_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
2	A_0061_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
3	A_0062_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
4	A_0062_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
5	A_0074_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
6	A_0074_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
7	A_0076_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
8	A_0076_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
9	A_0096_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
10	A_0097_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
11	A_0097_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
12	A_0108_LCC.pgm	0	**1**	0	0	N/A	1	0	0
13	A_0108_RCC.pgm	0	**1**	0	0	N/A	1	0	0
14	A_0114_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
15	A_0114_RCC.pgm	0	**1**	0	0	N/A	1	0	0
16	A_0124_LCC.pgm	0	**1**	0	0	N/A	1	0	0
17	A_0124_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
18	A_0133_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
19	A_0133_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
20	A_0146_LCC.pgm	0	**1**	0	0	N/A	1	0	0
21	A_0146_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
22	A_0149_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
23	A_0149_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
24	A_0150_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
25	A_0150_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
26	A_0158_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
27	A_0203_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
28	A_0203_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
29	A_0215_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
30	A_0215_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
31	A_0216_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
32	A_0216_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
33	A_0223_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
34	A_0223_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
35	A_0238_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
36	A_0238_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
37	A_0239_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
38	A_0239_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
39	A_0243_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
40	A_0243_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
41	A_0246_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
42	A_0246_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
43	A_0250_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
44	A_0250_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
45	A_0254_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
46	A_0254_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
47	A_0255_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
48	A_0255_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
49	A_0260_LCC.pgm	0	0	1	0	N/A	0	N/A	N/A
50	A_0260_RCC.pgm	0	0	1	0	N/A	0	N/A	N/A
	Total	0	5	45	0				
	Sample Average					N/A	0.10	N/A	N/A
	Sample Standard Deviation					N/A	0.30	N/A	N/A
	95% confidence interval					N/A	[0.017, 0.183]	N/A	N/A
	99% confidence interval					**N/A**	**[0, 0.209]**	**N/A**	**N/A**

false positive fraction, the *precision* and the *overlap ratio* (calculated for every individual image) for the abnormal and normal images, respectively.

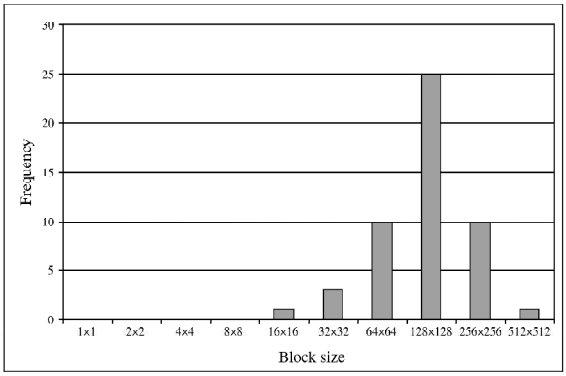
We also calculated the *true positive fraction* and the *false positive fraction* for our system using various block sizes. In this experiment, we dealt with the hundred test images as one large mosaic image, and then we calculated our measures based on the total number of blocks. Fig. 12 shows the *receiver operating characteristic* (ROC) curve for the *true positive fraction* and the *false positive fraction* of our system using various block sizes. The analysis of the ROC curve reveals that the system maintains the same level of *true positive fraction* of 1.0 with decreased *false positive fraction* when the block size is reduced from 256×256 to 128×128. By reducing the block size to 64×64, the *true positive fraction* approached 0.993 with 0.12 *false positive fraction*. With 32×32 block size, the proposed scheme has a *true positive fraction* of 0.885 with 0.077 *false positive fraction*. The *true positive fraction* drops faster when the block size is further reduced. Hence, there is a trade off between the *true positive fraction* and the *false positive fraction* when using a large block size (screening mode) or a small block size (maximizing the localization of the detected region).

To select a suitable block size, we analyzed the sizes of the areas annotated by clinicians. The analysis revealed that the average size of the gold standard areas (over the entire set of abnormal images) is 13190 pixels which can roughly be approximated by a square shape of 115×115 pixels. Fig. 6 shows various block sizes superimposed over a mammogram image to get a feeling of how big the blocks, relative to the total image size. Fig. 13 shows



Block Size (n)	*True Positive Fraction*	*False Positive Fraction*
1	0.465	0.030
2	0.498	0.033
4	0.543	0.038
8	0.612	0.043
16	0.724	0.053
32	0.885	0.077
64	0.993	0.120
128	1.000	0.171
256	1.000	0.286

Figure 12: ROC analysis for the entire sets of images using various block sizes



Block Size	1×1	2×2	4×4	8×8	16×16	32×32	64×64	128×128	256×256	512×512
Frequency	0	0	0	0	1	3	10	25	10	1

Figure 13: The histogram of the block sizes that approximate the areas annotated by the clinician

a histogram representing the distribution of the block sizes that can approximate the gold standard annotations. The histogram is calculated by finding the smallest block size that may completely cover each annotated object in the gold standard images. From this histogram, only one area (2% of all used abnormal images) needs a block smaller than 32×32 pixels and three areas (6% of all used abnormal images) need blocks smaller than 64×64 pixels. However, all other areas can be covered by blocks larger than 64×64 pixels (92% of all used abnormal images). This means that the clinician's precision is greater than 32×32 pixels in 98% of all cases (i.e., clinicians usually neglect regions smaller than 32×32 pixels). In other words, the gold standard data is reasonably accurate for regions greater than or equal to 32×32 pixels. This coincides with the results deducted from the ROC curve, Fig. 12, in which the *true positive fraction* does not change until we used a 32×32 blocks and then, it reduces dramatically. Hence, we find it quite reasonable to report our performance measure based on 32×32 block size.

Table 4 and Table 5 show the TP, FP, TN, FN block counts, the *true positive fraction*, the *false positive fraction*, the *precision* and the *overlap ratio* (calculated for every individual image), for the abnormal and normal images using 32×32 block size. The 95% and 99% confidence intervals (calculated over the fifty abnormal images) for the *true positive fraction* are [0.842, 0.938] and [0.827, 0.953], for the *false positive fraction* are [0.101, 0.203] and [0.084, 0.219], for the precision are [0.538, 0.691] and [0.514, 0.715], and for the overlap ratio are [0.483, 0.623] and [0.461, 0.645], respectively. Meanwhile, the 95% and 99% confidence intervals for the *false positive fraction* (calculated over the fifty normal images) are [0.002, 0.078] and [0, 0.09], respectively. However, if we consider all hundred images together, the 95% and 99% confidence intervals for the *false positive fraction* are [0.062, 0.130] and [0.052, 0.140], respectively.

Table 4
The Performance Criterion Applied to Cancer Cases (in Block Counts) using 32×32 Blocks

Case #	*Case Name*	*TP*	*FP*	*TN*	*FN*	*True Positive Fraction*	*False Positive Fraction*	*Precision*	*Overlap ratio*
1	c1_0020_LCC.pgm	39	14	50	0	1	0.22	0.74	0.74
2	c1_0011_RCC.pgm	15	2	103	0	1	0.02	0.88	0.88
3	c1_0014_RCC.pgm	3	3	40	1	0.75	0.07	0.50	0.43
4	c1_0016_RCC.pgm	6	3	52	0	1	0.05	0.67	0.67
5	c1_0019_RCC.pgm	48	9	97	0	1	0.08	0.84	0.84
6	c1_0020_RCC.pgm	17	20	72	0	1	0.22	0.46	0.46
7	c1_3022_RCC.pgm	3	1	86	0	1	0.01	0.75	0.75
8	c1_3049_LCC.pgm	4	**52**	55	0	1	0.49	0.07	0.07
9	c1_3065_RCC.pgm	16	11	98	0	1	0.1	0.59	0.59
10	c1_3066_LCC.pgm	6	4	70	0	1	0.05	0.60	0.60
11	c1_3084_RCC.pgm	6	0	76	0	1	0	1.00	1.00
12	c2_0022_RCC.pgm	18	2	23	1	0.95	0.08	0.90	0.86
13	c2_0023_LCC.pgm	4	0	88	1	0.8	0	1.00	0.80
14	c2_0028_LCC.pgm	16	15	69	**8**	0.67	0.18	0.52	0.41
15	c2_0050_LCC.pgm	12	17	27	0	1	0.39	0.41	0.41
16	c2_0051_RCC.pgm	14	6	62	0	1	0.09	0.70	0.70
17	c2_0061_RCC.pgm	24	2	0	0	1	1	0.92	0.92
18	c2_0071_LCC.pgm	97	8	31	**15**	0.87	0.21	0.92	0.81
19	c2_0073_RCC.pgm	2	0	86	**9**	0.18	0	1.00	0.18
20	c2_0074_LCC.pgm	11	7	38	1	0.92	0.16	0.61	0.58
21	c2_0075_LCC.pgm	12	10	42	0	1	0.19	0.55	0.55
22	c2_0079_LCC.pgm	9	7	49	4	0.69	0.12	0.56	0.45
23	c2_0081_LCC.pgm	27	0	52	**8**	0.77	0	1.00	0.77
24	c2_0082_LCC.pgm	5	20	165	5	0.5	0.11	0.20	0.17
25	c2_0086_RCC.pgm	4	6	63	2	0.67	0.09	0.40	0.33
26	c2_0088_LCC.pgm	17	0	69	3	0.85	0	1.00	0.85
27	c2_0093_LCC.pgm	5	10	60	0	1	0.14	0.33	0.33
28	c2_0094_RCC.pgm	18	**42**	50	0	1	0.46	0.30	0.30
29	c2_0100_RCC.pgm	10	20	19	0	1	0.51	0.33	0.33
30	c2_0108_LCC.pgm	9	0	34	5	0.64	0	1.00	0.64
31	c2_0111_LCC.pgm	8	0	28	3	0.73	0	1.00	0.73
32	c2_0112_LCC.pgm	45	**25**	45	**9**	0.83	0.36	0.64	0.57
33	c2_0118_LCC.pgm	2	3	53	2	0.50	0.05	0.40	0.29
34	c2_0121_RCC.pgm	9	8	59	3	0.75	0.12	0.53	0.45
35	c2_0131_RCC.pgm	6	0	87	0	1	0	1.00	1.00
36	c2_0133_LCC.pgm	28	22	25	1	0.97	0.47	0.56	0.55
37	c2_0137_LCC.pgm	13	11	62	3	0.81	0.15	0.54	0.48
38	c2_3088_LCC.pgm	5	2	30	0	1	0.06	0.71	0.71
39	c3_1017_LCC.pgm	34	13	90	4	0.89	0.13	0.72	0.67
40	c3_1027_LCC.pgm	13	2	78	1	0.93	0.02	0.87	0.81
41	c3_1035_RCC.pgm	6	**28**	50	0	1	0.36	0.18	0.18
42	c3_1036_LCC.pgm	11	12	95	0	1	0.11	0.48	0.48
43	c3_1050_RCC.pgm	4	9	69	0	1	0.12	0.31	0.31
44	c3_1055_LCC.pgm	4	11	105	0	1	0.09	0.27	0.27
45	c3_1060_LCC.pgm	5	4	230	0	1	0.02	0.56	0.56
46	c3_1084_RCC.pgm	2	8	130	0	1	0.06	0.20	0.20
47	c4_1076_RCC.pgm	2	0	116	0	1	0	1.00	1.00
48	c4_1082_RCC.pgm	9	7	62	0	1	0.1	0.56	0.56
49	c4_1089_LCC.pgm	5	9	69	1	0.83	0.12	0.36	0.33
50	c4_1107_LCC.pgm	2	**24**	85	0	1	0.22	0.08	0.08
	Total	690	489	3394	90				
	Sample Average					0.89	0.15	0.61	0.55
	Sample Standard Deviation					0.17	0.19	0.28	0.25
	95% confidence interval					[0.842, 0.938]	[0.101, 0.203]	[0.538, 0.691]	[0.483, 0.623]
	99% confidence interval					[0.827, 0.953]	[0.084, 0.219]	[0.514, 0.715]	[0.461, 0.645]

Table 5
The Performance Criterion Applied to Normal Cases (in Block Counts) using 32×32 Blocks

Case #	*Case Name*	*TP*	*FP*	*TN*	*FN*	*True Positive Fraction*	*False Positive Fraction*	*Precision*	*Overlap ratio*
1	A_0061_LCC.pgm	0	0	81	0	N/A	0	N/A	N/A
2	A_0061_RCC.pgm	0	0	94	0	N/A	0	N/A	N/A
3	A_0062_LCC.pgm	0	0	125	0	N/A	0	N/A	N/A
4	A_0062_RCC.pgm	0	0	110	0	N/A	0	N/A	N/A
5	A_0074_LCC.pgm	0	0	157	0	N/A	0	N/A	N/A
6	A_0074_RCC.pgm	0	0	162	0	N/A	0	N/A	N/A
7	A_0076_LCC.pgm	0	0	175	0	N/A	0	N/A	N/A
8	A_0076_RCC.pgm	0	0	173	0	N/A	0	N/A	N/A
9	A_0096_LCC.pgm	0	0	66	0	N/A	0	N/A	N/A
10	A_0097_LCC.pgm	0	0	83	0	N/A	0	N/A	N/A
11	A_0097_RCC.pgm	0	0	136	0	N/A	0	N/A	N/A
12	A_0108_LCC.pgm	0	**79**	51	0	N/A	0.61	0	0
13	A_0108_RCC.pgm	0	**54**	83	0	N/A	0.39	0	0
14	A_0114_LCC.pgm	0	0	112	0	N/A	0	N/A	N/A
15	A_0114_RCC.pgm	0	**73**	65	0	N/A	0.53	0	0
16	A_0124_LCC.pgm	0	**2**	98	0	N/A	0.02	0	0
17	A_0124_RCC.pgm	0	0	118	0	N/A	0	N/A	N/A
18	A_0133_LCC.pgm	0	0	122	0	N/A	0	N/A	N/A
19	A_0133_RCC.pgm	0	0	156	0	N/A	0	N/A	N/A
20	A_0146_LCC.pgm	0	**41**	49	0	N/A	0.46	0	0
21	A_0146_RCC.pgm	0	0	101	0	N/A	0	N/A	N/A
22	A_0149_LCC.pgm	0	0	142	0	N/A	0	N/A	N/A
23	A_0149_RCC.pgm	0	0	128	0	N/A	0	N/A	N/A
24	A_0150_LCC.pgm	0	0	89	0	N/A	0	N/A	N/A
25	A_0150_RCC.pgm	0	0	98	0	N/A	0	N/A	N/A
26	A_0158_RCC.pgm	0	0	112	0	N/A	0	N/A	N/A
27	A_0203_LCC.pgm	0	0	93	0	N/A	0	N/A	N/A
28	A_0203_RCC.pgm	0	0	169	0	N/A	0	N/A	N/A
29	A_0215_LCC.pgm	0	0	177	0	N/A	0	N/A	N/A
30	A_0215_RCC.pgm	0	0	168	0	N/A	0	N/A	N/A
31	A_0216_LCC.pgm	0	0	53	0	N/A	0	N/A	N/A
32	A_0216_RCC.pgm	0	0	66	0	N/A	0	N/A	N/A
33	A_0223_LCC.pgm	0	0	144	0	N/A	0	N/A	N/A
34	A_0223_RCC.pgm	0	0	68	0	N/A	0	N/A	N/A
35	A_0238_LCC.pgm	0	0	70	0	N/A	0	N/A	N/A
36	A_0238_RCC.pgm	0	0	130	0	N/A	0	N/A	N/A
37	A_0239_LCC.pgm	0	0	78	0	N/A	0	N/A	N/A
38	A_0239_RCC.pgm	0	0	80	0	N/A	0	N/A	N/A
39	A_0243_LCC.pgm	0	0	227	0	N/A	0	N/A	N/A
40	A_0243_RCC.pgm	0	0	148	0	N/A	0	N/A	N/A
41	A_0246_LCC.pgm	0	0	65	0	N/A	0	N/A	N/A
42	A_0246_RCC.pgm	0	0	61	0	N/A	0	N/A	N/A
43	A_0250_LCC.pgm	0	0	76	0	N/A	0	N/A	N/A
44	A_0250_RCC.pgm	0	0	63	0	N/A	0	N/A	N/A
45	A_0254_LCC.pgm	0	0	53	0	N/A	0	N/A	N/A
46	A_0254_RCC.pgm	0	0	51	0	N/A	0	N/A	N/A
47	A_0255_LCC.pgm	0	0	110	0	N/A	0	N/A	N/A
48	A_0255_RCC.pgm	0	0	96	0	N/A	0	N/A	N/A
49	A_0260_LCC.pgm	0	0	158	0	N/A	0	N/A	N/A
50	A_0260_RCC.pgm	0	0	168	0	N/A	0	N/A	N/A
	Total	0	249	5458	0				
	Sample Average					N/A	0.04	N/A	N/A
	Sample Standard Deviation					N/A	0.14	N/A	N/A
	95% confidence interval					N/A	[0.002, 0.078]	N/A	N/A
	99% confidence interval					N/A	[0, 0.09]	N/A	N/A

(i) Abnormal Cases Analysis

In some cases in Table 4, the produced objective results seemed unacceptable in terms of high false negative or high false positive block counts. Hence, we subjectively investigated these cases and compared the computer-highlighted with the clinician-highlighted images. In the following figures, we superimpose 32×32 blocks (appear in white color) over the computer highlighted images, to illustrate how the TP, FP, TN and FN terms are calculated.

Fig. 14 presents cases: "suspicious case 14", "suspicious case 18", "suspicious case 19", and "suspicious case 23", that produce high false negative blocks. The subjective comparison between the output produced by our scheme and the clinician images for these cases indicate that the high *false negative* blocks in these images are attributed to the fact that the clinician-highlighted regions are larger than the actual suspicious lesions. Since our scheme highlights the core of the same regions, it consequently fulfils the objective of CAD systems in attracting the clinician's attention to the specific location.

Fig. 15 shows sample cases, "suspicious case 8", "suspicious case 28", "suspicious case 41" and "suspicious case 50", which produce high *false positive* blocks. The subjective analysis of such cases indicated that the false detected regions usually represent either chest wall or milk ducts that appear brighter than the rest of the breast tissues, as in Fig. 15 (a) and (d). However, these areas are clearly unsuspicious, and hence they can be easily excluded by the clinician without confusion.

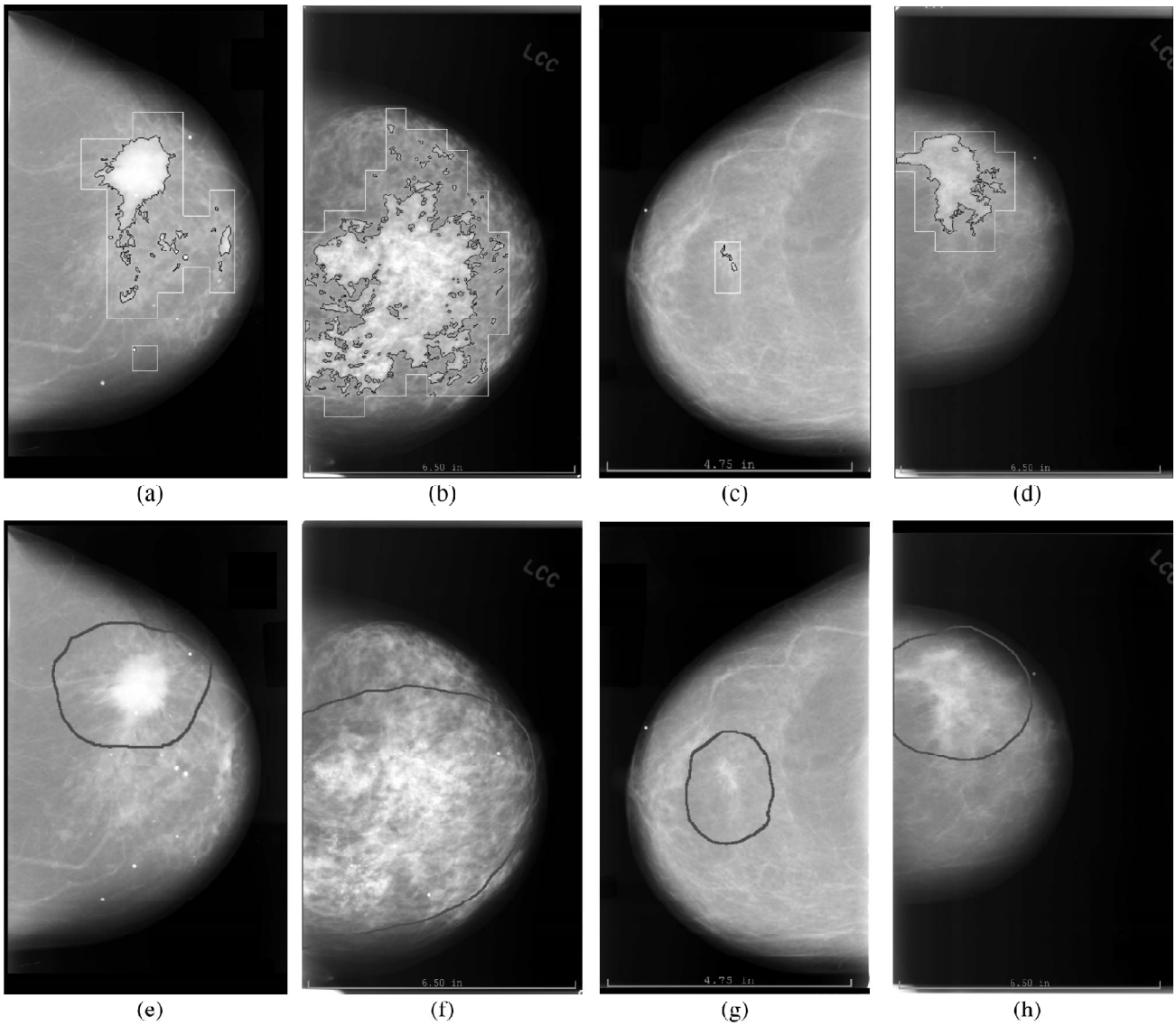


Figure 14: Samples of the abnormal cases in which the proposed scheme produces high false negative rates: (a)(b)(c)(d) are the images that produced by our scheme: suspicious cases 14, 18, 19 and 23, respectively; and (e)(f)(g)(h) are the clinician-highlighted images for the same case shown in (a)(b)(c)(d), respectively. The contour generated by the clinician is larger than actual suspicious region. This give rise to the false negative performance measure. Note that our scheme highlights the core of suspicious region and attracts the attention of the clinician to focus on this area. The white blocks that appear in the computer-highlighted images represent the 32×32 blocks that were used to count the TP, FP, TN and FN terms

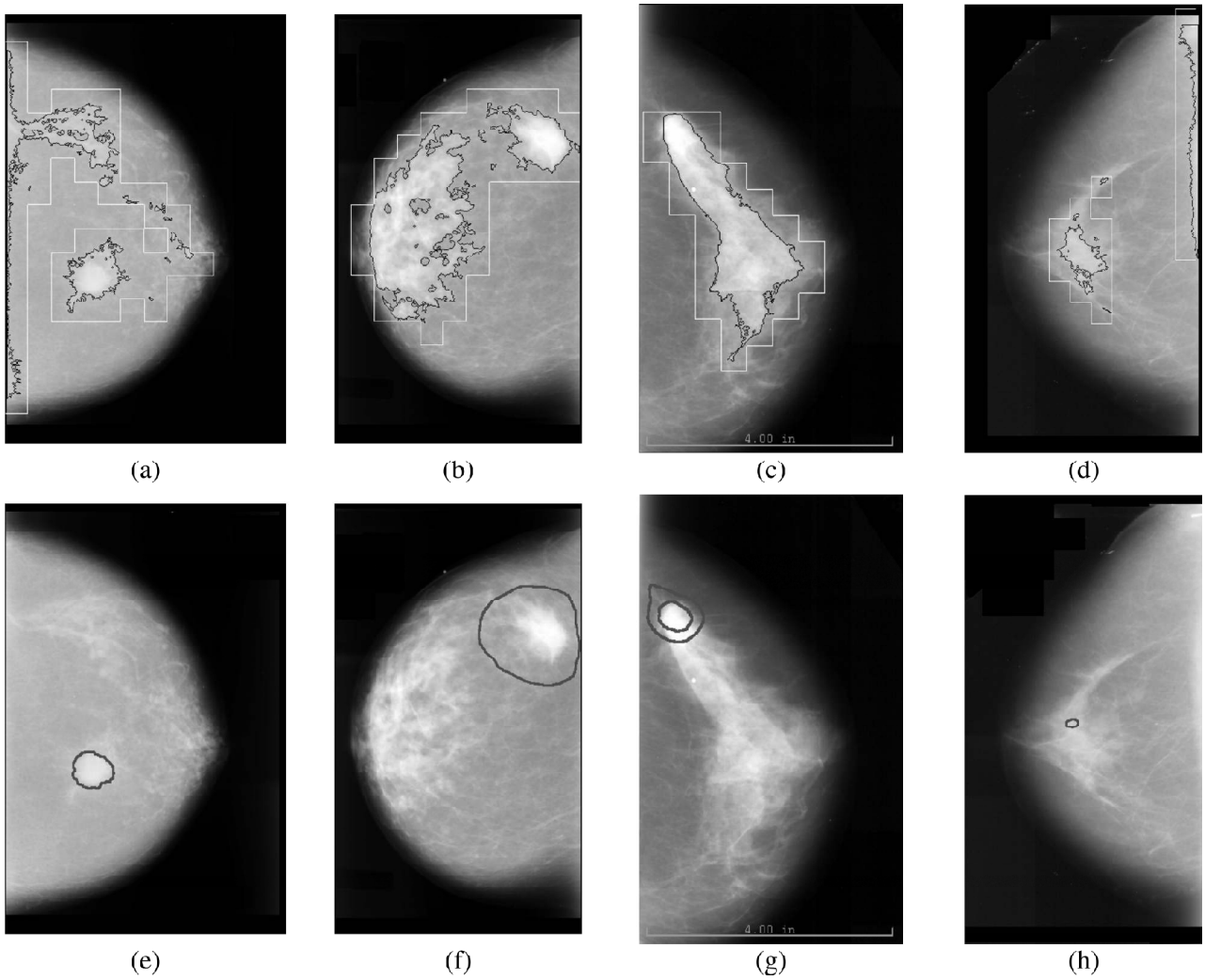


Figure 15: Sample of abnormal cases, in which the proposed scheme produces high false positive blocks: (a),(b),(c),(d) the images produced by our scheme for the suspicious cases 8, 28, 41 and 50, respectively; and (e),(f),(g),(h) the clinician-highlighted images for the same cases shown in (a),(b),(c),(d). The majority of the false detected areas are located near the chest walls (as in (a) and (d)). Other areas appear near the bright milk duct at the upper part of the breast (as in (a)). These areas can be easily excluded by the clinician. In (b) the proposed scheme successfully detected the cancerous regions in addition to other bright regions that need further clinician's investigation. In (c) the proposed scheme highlighted the whole cancerous regions, while the clinician just highlighted the mass center. Note that, the white blocks that appear in the computer-highlighted images represent the 32×32 blocks that were used to count the TP, FP, TN and FN terms

For "suspicious case 28", Fig. 15(b), the proposed scheme highlighted the cancerous regions in addition to other bright regions in the images that need further investigations. For "suspicious case 41", Fig. 15(c), the clinician just highlighted the mass center while the proposed scheme highlighted the entire suspicious region. Hence, our scheme takes extra precautions against missing a potential mass.

Fig. 16 shows "suspicious case 32" which produces both high *false positive* and *false negative* blocks. The subjective analysis of this case indicated that the clinician used a free-hand style plot to highlight the suspicious region. This gives rise to the *false negative* block counts. However, our schemes highlighted nearly the same region with better localization. Moreover, the clinician used his experience and intelligence to exclude the bright regions near the breast nipple, as they are more likely to represent network of milk ducts. Our scheme utilizes the pixel intensities only to highlight suspicious areas. That is why our scheme produced high *false positive* block count.

(ii) Normal Cases Analysis

Table 5 shows that the proposed scheme produces relatively high false positive results in few cases. Fig. 17 shows samples of such normal cases, "normal case 12", "normal case 13", "normal case 15" , "normal case 16", and "normal case 20", that produce high false positive blocks. As the figure shows, most of the false detected areas represent milk ducts and/or chest wall that are clearly unsuspicious and would not affect the clinician

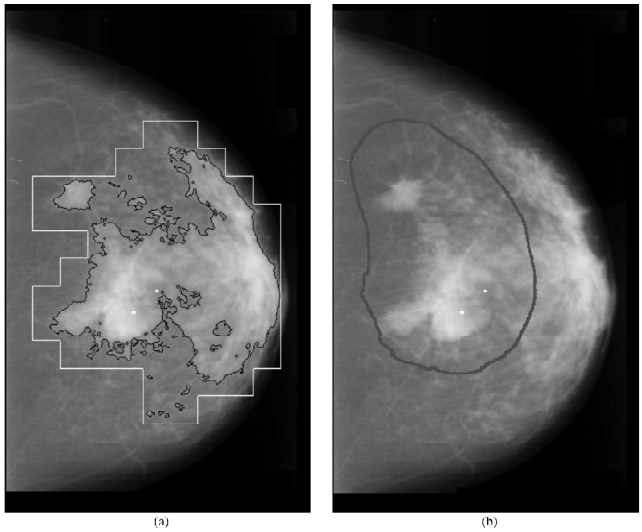


Figure 16: Sample of an abnormal case, in which the proposed scheme produces both high false positive and false negative blocks: (a) the images produced by our scheme for the suspicious cases 32; and (b) the clinician-highlighted image for the same cases shown in (a). The proposed scheme accurately highlighted the brightest suspicious regions in the image; while the clinician used a free-hand style plot to highlight nearly the same regions. Note that, the white blocks that appear in the computer-highlighted images represent the 32×32 blocks that were used to count the TP, FP, TN and FN terms

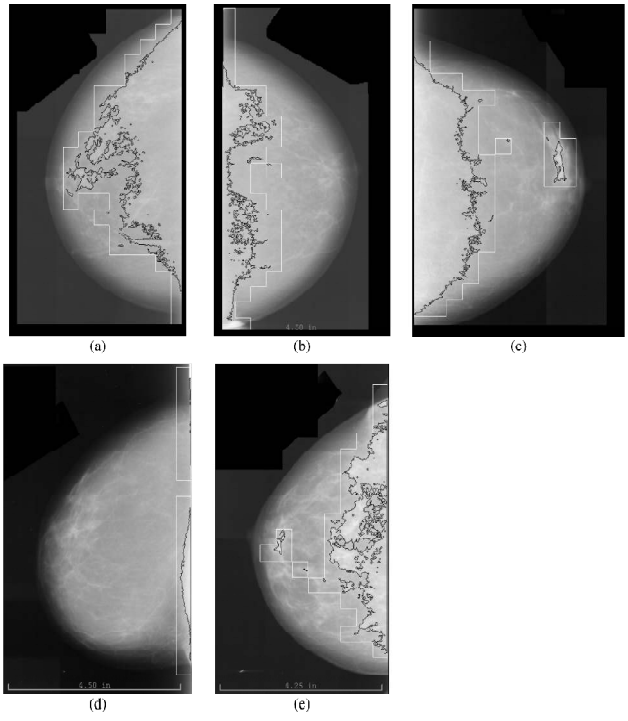


Figure 17: Samples of normal cases in which the proposed scheme produces high false positive rates: (a) normal case 12; (b) normal case 13; (c) normal case 15; (d) normal case 16; and (e) normal case 20. The false detected areas are located near the chest wall and milk ducts. These areas can be easily excluded by the clinician. Note that, the white blocks that appear in the computer-highlighted images represent the 32×32 blocks that were used to count the TP, FP, TN and FN terms

decision, i.e., similar to the high false positive in normal cases (Section III. C.1).

(d) General Comment

Unfortunately, we are unable to compare our scheme with others found in literature. We have neither the source code nor the set of images used by other researchers. However, the proposed scheme has several attractive features. First, it is fully automated (i.e. no user input is required). Second, all parameters are calculated based on the image under consideration. These features make the system portable (it operates with any image database without the need of any customization and produces the same level of performance). Finally, it is simple and easy to implement. Hence, the proposed scheme overcomes the shortcomings of the previous research works [2] [3] [4] [5] [6] [7] as discussed in Section I.

It is worth mentioning that the proposed scheme is designed to be used as a highlighting tool. In our future work, a module to automatically extract the chest wall, as well as further classification steps will be added to the proposed system to filter out many of the false positive areas.

IV. CONCLUSIONS

We proposed a scheme to provide an initial guess of suspicious masses in digital mammogram images, for further comprehensive processing. The scheme approximates the image histogram as a summation of three Gaussian functions with different means and standard deviations. These parameters are determined using the least square errors curve fitting technique. Bayesian decision theory is then applied to estimate the threshold that produces the minimum probability of classification error. Based on the value of this threshold, relative to the location of the second and the third Gaussian functions, the image is thresholded to highlight suspicious areas, if any. Finally, the contours of the highlighted regions, if any, are extracted and superimposed on the original mammogram image. The paper also presented a block-based performance criterion to assess the computer generated results.

The results produced by the proposed scheme (over a set of fifty abnormal and fifty normal images) were subjectively and objectively compared with the manually highlighted images that are provided by the DDSM database. The experimental results revealed that the proposed scheme achieved a *true positive fraction* of 1.0 for the entire fifty abnormal cases, when used in the screening mode. At the same time, the 95% and the 99% confidence intervals for the *false positive fraction*, calculated over the fifty normal cases, are [0.017, 0.183] and [0, 0.209], respectively. When used in diagnosis or follow up mode, more localization is required. In this

case, our proposed block-based measure is used to report the performance of our scheme. The analysis revealed that 32×32 block size is higher than the clinician's precision. Hence, we adopted this block size as our measuring unit to report the system's performance. We found that the 95% and 99% confidence intervals (calculated over the fifty abnormal images) for the *true positive fraction* are [0.842, 0.938] and [0.827, 0.953], for the *false positive fraction* are [0.101, 0.203] and [0.084, 0.219], for the precision are [0.538, 0.691] and [0.514, 0.715], and for the overlap ratio are [0.483, 0.623] and [0.461, 0.645], respectively. In the meantime, the 95% and 99% confidence intervals for the *false positive fraction* (calculated over the fifty normal images) are [0.002, 0.078] and [0, 0.09], respectively. On the other hand, if we consider all hundred images together, the 95% and 99% confidence intervals for the *false positive fraction* are [0.062, 0.130] and [0.052, 0.140], respectively. In conclusion, the proposed scheme can be utilized as a coarse segmentation stage to limit the search space for detecting/highlighting suspicious lesions on digital mammogram images.

ACKNOWLEDGMENT

This research is partially funded by the Natural Sciences and Engineering Research Council of Canada (NSERC). This support is greatly appreciated.

REFERENCES

[1] M. Sezgin and B. Sankur, "Survey Over Image Thresholding Techniques and Quantitative Performance Evaluation", *Journal of Electronic Imaging*, 13(1): 146–165, 2004.

[2] H. Kobatake, M. Murakami, H. Takeo, and S. Nawano, "Computerized Detection of Malignant Tumors on Digital Mammograms", *IEEE Transactions on Medical Imaging*, 18(5): 369–378, 1999.

[3] G. Brake and N. Karssemeijer, "Single and Multiscale Detection of Masses in Digital Mammograms", *IEEE Transactions on Medical Imaging*, 18(7): 628–639, 1999.

[4] S. Liu, C. Babbs, and E. Delp, "Multiresolution Detection of Spiculated Lesions in Digital Mammograms", *IEEE Transactions on Image Processing*, 10(6): 874–884, 2001.

[5] F. Valverde, N. Guil, J. Munoz, Q. Li, M. Aoyama, and K. Doi, "A Deformable Model for Image Segmentation in Noisy Medical Images", *International Conference on Image Processing*, 3: 82–85, 2001.

[6] E. Catanzariti, M. Ciminello and R. Prevete, "Computer Aided Detection of Clustered Microcalcifications in Digitized Mammograms Using Gabor Functions", *International Conference on Image Analysis and Processing*, pp. 266–270, 2003.

[7] A. Abdel-Dayem, and M. El-Sakka, "Fuzzy Entropy Based Detection of Suspicious Masses in Digital Mammogram Images", International Conference on Engineering in Medicine and Biology Society, pp. 4017–4022, 2005.

[8] R. Nakayama, Y. Uchiyama, K. Yamamoto, R. Watanabe, and K. Namba, "Computer-Aided Diagnosis Scheme Using a Filter Bank for Detection of Microcalcification Clusters in Mammograms", *IEEE Transactions on Biomedical Engineering*, 53(2): 273–283, 2006.

[9] S. Selvan, C. Xavier, N. Karssemeijer, J. Sequeira, R. Cherian, and B. Dhala, "Parameter Estimation in Stochastic Mammogram Model by Heuristic Optimization Techniques", *IEEE Transactions on Information Technology in Biomedicine*, 10(4): 685–695, 2006.

[10] D. Cascio, F. Fauci, R. Magro, G. Raso, R. Bellotti, F. De Carlo, S. Tangaro, G. De Nunzio, M. Quarta, G. Forni, A. Lauria, M. Fantacci, A. Retico, G. Masala, P. Oliva, S. Bagnasco, S. Cheran, and E. Torres, "Mammogram Segmentation by Contour Searching and Mass Lesions Classification with Neural Network", *IEEE Transactions on Nuclear Science*, 53(5): 2827 – 2833, 2006.

[11] A. Tripathy, S. Pathak, S. Chakrabarti, "Classifying Mammogram Images Using Fractal Features", *International Conference on Computational Intelligence and Multimedia Applications*, 2: 537–543, 2007.

[12] D. Sankar, T. Thomas, "Fractal Modeling of Mammograms Based on Mean and Variance for the Detection of Microcalcifications", International Conference on Computational Intelligence and Multimedia Applications, 2: 334–348, 2007.

[13] W. Ping, L. Junli, Z. Shanxu, L. Dongming, and C. Gang, "A Method of Detection Micro-Calcifications in Mammograms Using Wavelets and Adaptive Thresholds", 2nd International Conference on Bioinformatics and Biomedical Engineering, *ICBBE*, pp. 2361–2364, 2008.

[14] Y. Wu, C. Zhou, L. Hadjiiski, J. Shi, J. Wei, C. Paramagul, B. Sahiner, and H. Chan "A Dynamic Multiple Thresholding Method for Automated Breast Boundary Detection in Digitized Mammograms", *Proceedings of the SPIE Progress in Biomedical Optics and Imaging*, SPIE-6512, 2007.

[15] B. Shahedi, R. Amirfattahi, F. Azar, and S. Sadri, "Accurate Breast Region Detection in Digital Mammograms Using a Local Adaptive Thresholding Method", 8th International Workshop on Image Analysis for Multimedia Interactive Services, *WIAMIS'07*, pp. 26–26, 2007.

[16] K. Wei, W. Guangzhi, and D. Hui, "Segmentation of the Breast Region in Mammograms Using Watershed Transformation", 27th Annual International Conference of the Engineering in Medicine and Biology Society, *IEEE-EMBS*, pp. 6500–6503, 2006.

[17] D. Raba, A. Oliver, J. Martí, M. Peracaula, and J. Espunya, "Breast Segmentation with Pectoral Muscle Suppression on Digital Mammograms", Pattern Recognition and Image Analysis, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, 3523: 471–478, 2005.

[18] R. Ferrari, A. Frère, R. Rangayyan, J. Desautels, and R. Borges, "Identification of the Breast Boundary in Mammograms using Active Contour Models", *Medical and Biological Engineering and Computing Journal, Springer Berlin / Heidelberg*, 42(2): 201–208, 2004.

[19] R. Ferrari, R. Rangayyan, J. Desautels, R. Borges, A. Frere," Automatic Identification of the Pectoral Muscle in Mammograms", *IEEE Transactions on Medical Imaging*, 23(2): 232–245, 2004.

[20] L. Wasserman, "All of Statistics, A Concise Course in Statistical Inference", New York, Springer-Verlag, chapter 20, pp. 304–326, 2004.

[21] L. Hays and R. Winkler, "Statistics: Probability, Inference and Decision", Holt Rinhart and Winston Inc., chapter 12, pp. 778 – 850, 1971.

[22] P. Hoel, "Introduction to Mathematical Statistics", Fifth Edition, John Wiley & Sons, chapter 7, pp. 192–207, 1984.

[23] G. Snedecor and W. Cochran, "Statistical Methods", eighth edition, Iowa State University Press, Ames, chapter 9, pp. 149 – 174, 1989.

[24] *http://www.systat.com/products/TableCurve2D*

[25] R. Duda, P. Hart, D. Stork, "Pattern Classification", Second Edition, John Wiley & Sons, chapter 2, pp. 20–63, 2001.

[26] R. Schalkoff, "Pattern Recognition Statistical, Structural and Neural Approaches", John Wiley & Sons, chapter 2, pp. 34–54, 1992.

[27] G. Gonzalez and E. Woods, "Digital Image Processing", 2nd Edition, Prentice Hall, 2002.

[28] E. Dargherty and R. Lotufo, "Hands–on Morphological Image Processing", The Society of Photo-Optical Instrumentation Engineers, 2003.

[29] M. Heath, K Bowyer, D. Kopans *et al*., "Current status of the Digital Database for Screening Mammography", *Digital Mammography*, pp. 457–460, 1998.

[30] M. Sonka and J. Fitzpatrick, "Handbook of Medical Imaging", SPIE- The International Society for Optical Engineering, Vol. 2, chapter 10, pp. 569–604, 2000.