# Detection of Breast Cancer in the JPEG2000 Domain

Fayez M. Idris, and Nehal I. AlZubaidi

***Abstract*—**Breast cancer detection techniques have been reported to aid radiologists in analyzing mammograms. We note that most techniques are performed on uncompressed digital mammograms. Mammogram images are huge in size necessitating the use of compression to reduce storage/transmission requirements. In this paper, we present an algorithm for the detection of microcalcifications in the JPEG2000 domain. The algorithm is based on the statistical properties of the wavelet transform that the JPEG2000 coder employs. Simulation results were carried out at different compression ratios. The sensitivity of this algorithm ranges from 92% with a false positive rate of 4.7 down to 66% with a false positive rate of 2.1 using lossless compression and lossy compression at a compression ratio of 100:1, respectively.

***Keywords*—**Breast cancer, JPEG2000, mammography, microcalcifications.

## I. Introduction

BREAST cancer is one of the most common causes of death among women around the world [1]. Early detection of breast cancer is the best hope for reducing the death rate caused by the invasive disease [2]. X-ray mammography is considered by general consent to be the best for early detection of breast cancer [2]. Nevertheless, some problems stand up to this technology. First, it is a subjective task leading to multiple interpretations to each single mammogram. Second, mammograms have low contrast causing the analysis of these images to be difficult. Third, the number of mammograms to be analyzed by a radiologist per day is limited.

To overcome these problems, computerized assisted diagnostic (CAD) techniques of breast cancer have been raised. We note that most of these techniques are performed in the spatial domain. Digital mammograms are huge in size, a feature inherent with medical images, leading to annoying latency in image processing and transmission and to wasted storage. This necessitates the use of compression techniques on images to reduce storage/transmission requirements. Many image compression techniques have been reported in the literature [3], [4]. The Digital Imaging and Communications in Medicine (DICOM) has adopted the JPEG2000 standard to compress medical images [22].

Manuscript received September 15, 2005.

F. M. Idris is with the Department of Computer Engineering, Jordan University of Science and Technology, P.O. Box 3030, Irbid 22110, Jordan; (phone: +9622-7201000 Ex 22131; fax:+9622-70954046; e-mail: fidris@just.edu.jo).

N. I. AlZubaidi, was with the Department of Computer Engineering, Yarmouk University, Irbid; (e mail: znehal2000@yahoo.com).

In this paper we present an algorithm for the detection of microcalcifications in the JPEG2000 domain. The algorithm is based on the statistical properties of the wavelet transform that the JPEG2000 coder employs. Simulation results have demonstrated that the proposed algorithm yields an excellent sensitivity and specificity at a reasonable false positive detection rate.

This rest of this paper is organized as follows. In Section II, related work is reviewed. In Section IV, the proposed algorithm is detailed. Finally, experimental results are presented in Section IV.

## II. Related Work

The reported algorithms of breast cancer detection can be classified into two classes; spatial domain and wavelet based techniques.

### A. Spatial Domain Techniques

In these techniques, an image is preprocessed firstly in order to enhance its quality. Enhancement techniques involve morphological operations [5]-[6], Laplacian filtering [7], fuzzification [8], fractal analysis [5], higher order statistics [9], and histogram manipulation [10]. The use of image enhancement increases the contrast of a mammogram image. Secondly, all regions that include microcalcifications are identified by segmenting the enhanced image using thresholding or neural network techniques [11]. Thresholding techniques include histogram thresholding and entropy based thresholding [12]. In this step, all candidate pixels of being microcalcifications are identified. This might result in false positives. Lastly, false positive findings are eliminated using texture analysis.

The drawback of these methods is that digital mammograms are huge in size and it is expected that mammogram images will be compressed in order to reduce transmission/memory requirements. Therefore, it is necessary to decompress digital mammograms before applying the spatial domain techniques. Decompressing is a time consuming process and needs large memory size.

### B. Wavelet based Techniques

In these techniques, a digital mammogram is first decomposed by a wavelet filter. The wavelet decomposition of a mammogram embeds the fine details in the high resolution levels of the decomposed image. The common wavelets that are used to detect microcalcifications in digital mammograms are: Least Asymmetric Daubechies' wavelet transform, with a finite basis of length 8 (LAD8) [13]-[17], cubic spline [18], a'trous [19], and Mallat [20]. Secondly, the wavelet coefficients of the decomposed image are modified to enhance small details. Thirdly, a new enhanced image with clear microcalcifications and suppressed background is reconstructed. After that, thresholding is applied to the reconstructed image to detect microcalcifications as well as to eliminate false positives.

The drawbacks of these techniques are as follow. The time and space wasting is still persist because mammogram images are still need to be decompressed. Further complexity is added by the wavelet transformation step. Another disadvantage is in using wavelet filters that are not used widely in the medical imaging area.

We recall from Section I that many compression techniques have been reported in the literature [3], [4]. The JPEG2000 compression standard have been prominently useful in medical imaging as well as in image processing [21] and it is included in the Digital Imaging and Communications in Medicine (DICOM) [22]. This standard is introduced by the Joint Photographic Experts Group (JPEG). The reader can find a detailed explanation of the JPEG2000 standard in [23].

## III. Microcalcification Detection in the JPEG2000 Domain

In this section, we present the proposed algorithm for early detection of breast cancer in the domain of JPEG2000. The input to the algorithm is a compressed mammogram image. The output of the algorithm is a binary image where its white spots point to microcalcifications that has been detected. The steps of the algorithm are illustrated in the block diagram that is shown in Fig. 1. In this algorithm, the wavelet transform coefficients are extracted from the JPEG2000 mammogram image. The subimages of the second and third level are extracted from the wavelet transformed image. The subimages of the third level are downsampled and correlated with the second level subimages. The correlated subimages are scaled and then an image is resulted by applying interband standard deviation. Automatic thresholding is applied to yield candidates of microcalcifications. From those candidates, isolated pixels are eliminated and a final binary image is obtained. The following subsections explain in detail each step in the algorithm.

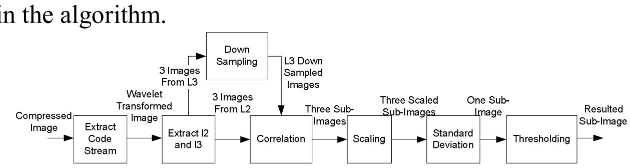


Fig. 1 Block diagram for the detection of micrcalcifications

### A. *Wavelet Transformed Image Extraction*

In this step, the wavelet coefficients of the compressed image are extracted from the code stream of the JPEG2000 file. Fig. 2 shows a digital mammogram in the spatial domain. The extracted coefficients of the wavelet transformed mammogram image that corresponds to the mammogram shown in Fig. 2 are shown in Fig. 3. Fig. 3 shows the wavelet resolution levels with the three oriented subimages at each resolution level. Each resolution level of the wavelet transformed image includes four blocks. The upper left block is the approximation subimage that is supported to further decomposition. The remaining three blocks are the high-frequencies decomposed bands: horizontal (HL), vertical (LH), and diagonal (HH) [25]. Our concern is the high frequency bands because microcalcifications are of high frequency. From our experiments, we found that the second and third levels of the wavelet transformed mammogram image are the best to apply our algorithms on. This is because although fine details are clear in the higher levels, those levels are very noisy and resulted with high rates of false positives. On the other hand, fine details are lost in the lower levels and resulted with very low sensitivities.

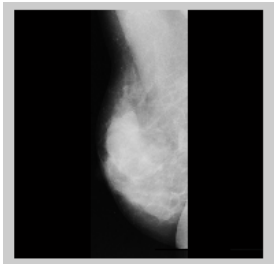

Fig. 2 Mammogram image in the spatial domain

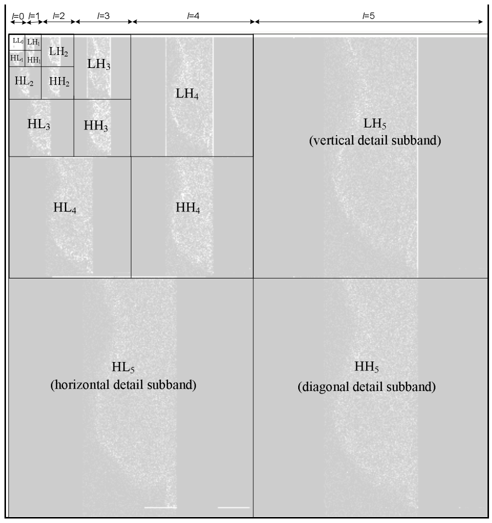


Fig. 3 Extracted wavelet transformed image

### B. *Correlation*

Correlation is calculated by multiplying two adjacent scales of the wavelet transformed image in order to magnify significant structures and suppress noise. This is based on the fact that edge structures present observably at each subband of the wavelet domain while noise decreases rapidly along the scales [20]. The aim of the correlation step here is not matching as in conventional correlation procedures. The matching between the two wavelet scales exists since those scales are obtained from the same original image by correlating this original image with the translated and scaled mother wavelet function. The correlation $C_{2,3}^i$ is performed simply by multiplying each wavelet coefficient in level 2, $W_2^i$, by its corresponding coefficient in the downsampled subimages of level 3, $[W_3^i]_d$,

$$C_{2,3}^i = W_2^i \times [W_3^i]_d \quad (1)$$

Where $i = LH$, $HL$, and $HH$.

*C. Scaling*

Scaling is required to return the total power in the correlated coefficients $C_{2,3}^i$ to be the same as the total power in the wavelet coefficient set of the 2nd level, $W_2^i$, namely,

$$SC_{3,2}^i = C_{3,2}^i \times r = C_{3,2}^i \times \sqrt{\frac{P_{W_2^i}}{P_{C_{3,2}^i}}} \quad (2)$$

Where

$$P_{W_2^i} = \sum W_2^i(x,y) \times W_2^i(x,y) \quad (3)$$

And

$$P_{C_{2,3}^i} = \sum C_{2,3}^i(x,y) \times C_{2,3}^i(x,y) \quad (4)$$

*D. Determination of Standard Deviation*

This step is required to identify pixels that are candidates to be microcalcifications. In this step, we calculate the standard deviation among the three corresponding wavelet coefficients in the three scaled subimages. The output of this step is a subimage that we called *STD* where each pixel in it is resulted from calculating the standard deviation as defined in,

$$STD(x,y) = \sum_{i=HL,LH,HH} SC_{3,2}^i(x,y) - m \quad (5)$$

Where $m$ is the mean of the three oriented pixels and is given by,

$$m = \frac{1}{3} \sum_{i=HL,LH,HH} SC_{3,2}^i(x,y) \quad (6)$$

The use of the standard deviation among bands that are resulted from levels 2 and 3 after correlating and scaling them is based on the fact that edges have high frequency components in the wavelet transformed image and have strong orientation in the wavelet domain bands where flat areas are suppressed because of its low frequency components and have no orientation. So microcalcifications that are considered as edges in the frequency domain will have the horizontal, vertical and diagonal orientations in the bands and so the standard deviation among those will be much greater than the standard deviation among flat areas.

*E. Thresholding*

Clusters of microcalcifications are isolated from all candidates that resulted from the previous step by obtaining a suitable threshold. The selection of the threshold is automatic and it is based on the non-Gaussian model of wavelets, i.e. the highly peaked around zero [26] and the heavy tail distribution of the transform [27]. Briefly, we calculate the wavelet coefficients threshold by finding the valley in the histogram that resides after the median of the wavelet coefficients of the *STD* subimage and assign it to our threshold. We choose left most and right most valleys. We start from the right half (i.e. the median) of the histogram, because the histogram of the microcalcifications affected subimages are skewed right distribution (i.e. the heavy tail is located on the right side of the histogram) and so the microcalcifications are located in the right tail of the skewed right histogram. The left side of the histogram of the wavelet transformed mammogram image often contains wavelet coefficients for the black background. Next we move to the decision rule thresholding, namely,

$$f(x,y) = \begin{cases} 1, & STD(x,y) \geq TH \\ 0, & STD(x,y) < TH \end{cases} \quad (7)$$

Any pixel in the *STD* subimage that has a wavelet coefficient value which is larger than or equal to the threshold, *TH*, is set to one; else it is cleared to zero. The result of this step is a binary image *f(x,y)* contains all candidates of pixels to be microcalcifications. False signals are found here. To decrease the false positives, we calculate white dots in an overlapped 5 by 5 window that we move on the entire binary subimage. If there is only one white point in the window we remove it else a cluster of microcalcifications is identified.

## IV. Experimental Results

The Digital Database for Screening Mammography (DDSM) [28] is used to test the performance of the developed algorithms. The proposed algorithm is evaluated using 100 mammograms where 70 are affected by microcalcifications and 30 are normal. The diagnosed images and the specification file of each image are used to compare our detection results with the actual results. The mammogram images of the database are compressed using JPEG2000 to obtain 4 image sets. The first set is based on the lossless JPEG2000 compression (compression ratios 2:1 to 3:1). The second, third, and fourth sets are based on the lossy JPEG2000 compression at compression ratios of 10:1, 50:1 and 100:1, respectively. Our simulations demonstrated that setting the threshold as the minimum value of the least count wavelet coefficients set gives better results than setting it to the maximum value. Sensitivity and specificity were used to evaluate the proposed algorithm. Sensitivity evaluates how often the algorithm correctly identifies microcalcifications, while specificity evaluates how often the algorithm correctly specifies mammograms without microcalcifications.

The results of applying the proposed algorithm on the 4 sets of mammogram images are shown in Table 1. For lossless compressed mammograms the proposed algorithm has yielded a sensitivity of 92% at a false positive detection rate of 4.7. For lossy compressed images at 10:1, 50:1, and 100:1, the obtained sensitivities are 92%, 77%, and 66%, respectively. The corresponding false positive rates are 2.4, 2.3. and 2.1, respectively. The specificity for the lossless compressed images is 87% at a false positive rate of 1.3. We note that the algorithm sensitivity and specificity are comparable to spatial domain and other wavelet based techniques. However, the proposed algorithm is based on standard image compression technique (JPEG200) and has excellent false positive rates. More importantly, the proposed technique saves time and memory space by eliminating the need to decompress mammograms for processing.

TABLE I
RESULTS OF APPLYING THE PROPOSED ALGORITHM ON DDSM MAMMOGRAMS

| Compression Ratio | False Positive Rate | Sensitivity |
|---|---|---|
| 100:1 | 2.1 | 66% |
| 50:1 | 2.3 | 77% |
| 10:1 | 2.4 | 92% |
| Lossless | 4.7 | 92% |

## ACKNOWLEDGMENT

The authors would like to thank Dr. Hassan Balas, from King Abdullah University Hospital, Irbid, Jordan, for his great help in understanding digital mammograms.

## REFERENCES

[1] M. Nimry, *Health experts plan major study on breast cancer patterns in jordan*, Jordan Times (1998).

[2] S. J. Nass, I. C. Henderson and J. C. Lashof, "Mammography and beyond: Developing technologies for the early detection of breast cancer," INSTITUTE OF MEDICINE and Division of Earth and Life Studies NATIONAL RESEARCH COUNCIL, United States of America, Washington, DC, 2003, pp. 1-267.

[3] C. Fabregas, O. Egger, F. Moscheni and V. Vaerman, "Compression of medical images," Swiss Federal Institute of Technology EPFL, Switzerland, 1997, pp. 1-13.

[4] D. A. Clunie, *Lossless compression of grayscale medical images - effectiveness of traditional and state of the art approaches*, Quintiles Intelligent Imaging.

[5] W. M. Diyana, J. Larcher and R. Besar, *A comparison of clustered microcalcifications automated detection methods in digital mammogram*, IEEE International Conference on Acoustics, Speech, and Signal Processing Proceedings, 2003, pp. 385-388.

[6] J. M. Mossi and A. Albiol, *Improving detection of clustered microcalcifications using morphological connected operators*, IEEE Seventh International Conference on Image Processing and Its Applications, 1999.

[7] T. Netsch and H. Peitgen, *Scale-space signatures for the detection of clustered microcalcifications in digital mammograms*, IEEE Transactions on Medical Imaging **18** (1999), no. 9.

[8] H. Cheng, Y. M. Lui and R. I. Freimanis, *A new approach to microcalcification detection in digital mammography*, IEEE Transactions on Medical Imaging (1997).

[9] M. Gurcan, Y. Yardimci, A. Cetin and R. Ansari, *Detection of microcalcifications in mammograms using higher order statistics*, IEEE Signal Processing Letters **4** (1997), no. 8.

[10] M. Wilson, R. Hargrave, S. Mitra, Y. Shieh and G. H. Roberson, *Automated detection of microcalcifications in mammograms through application of image pixel remapping and statistical filter*.

[11] A. Wroblewska, P. Boninski, A. Przelaskowski and M. Kazubek, *Segmentation and feature extraction for reliable classification of microcalcifications in digital mammograms*, Opto-Electronics Review **11** (2003), no. 3, pp. 227-235.

[12] M. Melloul and L. Joskowicz, *Segmentation of microcalcification in x-ray mammograms using entropy thresholding*, Computer Assisted Radiology and Surgery CARS: Proceedings of the 16th International Congress and Exhibition **6** (2002), no. 671.

[13] M. J. Lado, P. G. Tahoces, A. J. Mendez, M. Souto and J. J. Vidal, *Computer-assisted diagnosis: Application of wavelet transform techniques to the detection of clustered microcalcifications in digital mammograms*, 10th Portuguese Conference on Pattern Recognition, Marzo, 1998, pp. 26-27.

[14] H. Yoshida, K. Doi, R. M. Nishikawa, K. Muto and M. Tsuda, "Application of wavelet transform to automated detection of clustered microcalcifications in digital mammograms," Tokyo Institute of Polytechnics 17, 1583 liyama, Atsugi, Kanagawa, Japan, 1994, pp. 24-37.

[15] H. C. Choe and A. K. Chan, *Microcalcification cluster detection in digitized mammograms using multiscale techniques*, IEEE Southwest Symposium on Image Analysis and Interpretation, 1998, pp. 23-28.

[16] D. Nesbitt, F. Aghdasi, R. Ward and J. Morgan-Parkes, *Detection of microcalcifications in digitized mammograms film images using wavelet enhancement and local adaptive false positive suppression*, Proceedings. IEEE Pacific Rim Conference on Communications, Computers, and Signal Processing, 1995, pp. 594 - 597.

[17] T. C. Wang and N. B. Karayiannis, *Detection of microcalcifications in digital mammograms using wavelets*, IEEE Transactions on Medical Imaging **17** (1998), no. 4, pp. 498-509.

[18] A. Laine, M. Lewis and F. Taylor, *A wavelet based mammographic system*, IEEE International Conference on Acoustics, Speech, and Signal Processing, 1994,.

[19] T.J.Brown, *An adaptive strategy for wavelet based image enhancement*, Irish Machine Vision & Image Processing Conference IMVIP 2000 Programme, Image and Vision Systems Group, School of Computer Science, The Queen's University of Belfast., 2000.

[20] L. Zhang and P. Bao, *A wavelet-based edge detection method by scale multiplication*, Proceedings of IEEE International Conference on Pattern Recognition, 2002, pp. 501-504.

[21] J. Z. Wang, *Wavelets and imaging informatics: A review of the literature*.

[22] "Jpeg2000 for medical applications link," Aware, Inc. the source for broadband intellectual property.

[23] "Jpeg-2000 home page.," URL http://www.jpeg.org/jpeg2000/index.html.

[24] M. Unser and T. Blu, *Mathematical properties of the jpeg2000 wavelet filters*, IEEE transactions on image processing **12** (2003), no. 9, pp. 1080-1090.

[25] W. Kou, *Digital image compression algorithms and standards*, Kluwer Academic Publishers, Boston/Dordrecht/London, 1995.

[26] R. C. Gonzalez and R. E. Woods, *Digital image processing*, Prentice Hall, New Jersey, 2002.

[27] A. Azimifar, P. Fieguth and E. Jernigan, *Towards random field modeling of wavelet statistics*, ICIP, IEEE, 2002.

[28] "Ddsm: Digital database for screening mammography.," University of South Florida Digital Mammography Home Page. Available from URL http://marathon.csee.usf.edu/Mammography/Database.html.