

Detection of Microcalcifications in Mammograms Using Local Maxima and Adaptive Wavelet Transform Analysis

A. Murat Bagci and A. Enis Cetin

Abstract: A method for computer-aided diagnosis of microcalcification clusters in mammogram images is presented. Microcalcification clusters which are an early sign of breast cancer appear as isolated bright spots in mammograms. Therefore they correspond to local maxima of the image. The local maxima of the image is first detected and they are ranked according to a higher order statistical test performed over the subband domain data.

Introduction: Microcalcification clusters are an early sign of breast cancer. The survival rate approaches 100 percent if cancer is detected early. Microcalcifications (MC) appear as isolated bright spots on mammograms images [1]-[4].

MC's correspond to local maxima of a mammogram as they are relatively bright and tiny regions in the image. The first step of our method is the detection of the local maxima of the mammogram image. Although a typical mammogram is much smoother than most natural images there are thousands of local maxima in a mammogram image. After detecting the maxima locations we rank them according to a higher order statistical test performed over the subband domain data obtained by the adaptive wavelet transform. The distribution of wavelet data corresponding to the regular breast tissue is almost Gaussian [3, 4]. On the other hand MC's are different in nature than regular breast tissue and they produce outliers in the subband domain. We take advantage of this fact and rank the local maxima according to a higher order statistical test estimated in the neighborhood of each local maximum. When the data is Gaussian the test statistics becomes zero. Higher the value of the test higher the rank of the maximum. Peaks due to MC's receive high ranks. The maxima due to small variations in the pixel values and smooth edges get low ranks.

Recently, we developed methods for detection of MC's based on higher order statistics, and wavelet analysis [3]-[4]. In these schemes the subband (or wavelet) domain image $|x_{lh}| + |x_{hl}| + |x_{hh}|$ of the mammogram image x is divided into overlapping small windows and a higher order statistic (HOS) [6] is estimated in each window. The windows with HOS values higher than a threshold value, T are marked as regions containing MC clusters. A weakness of the methods [3]-[4] is that the threshold T should be estimated from a set of training images. The threshold has to be adjusted from scanner to scanner and according to the data set. In addition, we compute the HOS test only around maxima locations instead of the entire image, thus achieve a computationally more efficient method than [3]-[4]. The HOS test is reviewed in Section III.

Another important feature of this paper is that adaptive wavelet (subband) transform[5] is used instead of a regular wavelet transform (WT). It is experimentally observed that adaptive WT provides better results than the ordinary Daubechies WT.

Adaptive Wavelet Transform: Classical adaptive prediction concepts are combined with the perfect reconstruction filter bank theory in [5] where the key idea is to decorrelate the polyphase components of the multichannel structure by using an adaptive predictor P as shown in Figure 1. Adaptation of the predictor coefficients are carried out by an Least Mean Square (LMS) type algorithm.

In Figure 1, $x_1(n)$ is the downsampled version of the original signal, $x(n)$, thus it consists of the even samples of $x(n)$. Similarly, the signal $x_2(n)$ consists of the odd samples. An LMS based FIR predictor of $x_2(n)$ from $x_1(n)$ can be expressed as $\hat{x}_2(n) = \mathbf{w}(n)\mathbf{x}_1^T(n)$ where $\mathbf{x}_1(n) = [x_1(n-L), \dots, x_1(n+L)]^T$ is the observation vector, and the $2L+1$ vector $\mathbf{w}(n)$ is the vector of predictor coefficients which is adapted by the equation

$$\mathbf{w}(n+1) = \mathbf{w}(n) + \mu \frac{\mathbf{x}_1(n)e(n)}{\|\mathbf{x}_1(n)\|^2}, \quad (1)$$

where the error signal $e(n) = x_2(n) - \hat{x}_2(n)$.

The filterbank structure shown in Figure 1 is the simplest adaptive wavelet transform (AWT) structure. In this structure, the “highband signal” is essentially the prediction error and as a result the subsignals are expected to be decorrelated. Other AWT structures with antialiasing filters for the upper branch signal can be found in [5].

This one dimensional filterbank is extended to two dimensions in a separable manner. The main advantage of the AWT over the standard WT is the use of adaptive prediction to compute the subimages x_{lh} , x_{hl} , and x_{hh} . Since MC’s are isolated bright spots, they cannot be predicted using the background pixels corresponding to the breast tissue. As a result they produce outliers and the prediction error image pixels deviate from Gaussianity around MC locations. On the other hand background pixels corresponding to the breast tissue are very effectively estimated by the adaptive filter.

The downsampling operation in the AWT does not cause any data loss even if a MC consists of a single pixel. The AWT structure either tries to estimate this pixel by its neighbors or it is used in prediction of other pixels. In both cases it produces outliers in subimages as in the first case it cannot be predicted from its neighbors and in the second case it destroys the prediction process.

HOS Test: The Higher Order Statistical (HOS) test used here is based on the sample estimates of the first four moments I_1, I_2, I_3 and I_4 of the wavelet data. Estimates of the moments are given by $I_k = \frac{1}{M \times N} \sum_{m=1}^M \sum_{n=1}^N e^k[m, n]$, $k = 1, 2, 3, 4$ where, $e = (|x_{lh}| + |x_{hl}| + |x_{hh}|)$. The subimages x_{lh} , x_{hl} , x_{hh} are obtained from the AWT of the mammogram image. In the limit $I_1 \rightarrow \mu$, $I_2 \rightarrow \sigma^2 + \mu^2$, $I_3 \rightarrow \mu^3 + 3\sigma^2\mu$ and $I_4 \rightarrow \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4$, where μ and σ^2 denote the mean and the variance of the error e , respectively. As a result, the statistics

$$h_3(I_1, I_2, I_3) = I_3 - 3I_1(I_2 - I_1^2) - I_1^3 \quad (2)$$

and

$$h_4(I_1, I_2, I_4) = I_4 + 2I_1^4 - 3I_2^2 \quad (3)$$

are equal to zero for Gaussian distributed data in the limit. When there are outliers in the data, h and H deviate from zero. The test statistic used in this paper is $H = h_3 + h_4$.

Detection Method and Experimental Results: The detection algorithm consists of three steps:

We first analyze the mammogram image and detect the local maxima. We simply do this in two steps. We determine pixels with the property that $x(i, j) \geq x(i + l, j + k)$, $k, l = -2, -1, 1, 2$. We also divide the mammogram image into overlapping windows of size $M = 15 \times N = 15$ and determine the pixels which are greater than 5% of the mean of each block even if they do not satisfy the above condition. In a 2K by 2K mammogram image we determine about 1000 candidate pixels for MC locations. In the second step we process the image using the analysis part of the AWT structure described in Section II and we obtain the quarter size image $|x_{lh}| + |x_{hl}| + |x_{hh}|$. In the third step, we estimate the higher order statistic, H , in an M by N window around each maximum location and the maxima are ranked according to their H value. A typical output image with the first 10 maxima is shown in Figure 2.

Detection results are summarized in Table 1 for thirteen mammograms obtained from the Nijmegen University and University of Florida databases which have been studied by radiologists. For example, the image c05o has 2 regions containing MC clusters. The adaptive wavelet based method marks the MC's in the first (second) cluster as 2 (5) whereas in the wavelet transform based ranking the same MC's are ranked as 14 and 121 respectively. The AWT based test produces better results in terms of ranking accuracy as shown in Table 1.

In another experiment the maxima are ranked according to the variance around the maxima locations instead of the higher order statistical (HOS) test. The average of ranks assigned to MC clusters by AWT+HOS method is 4.28 while it is 9.6 in AWT and variance based ranking.

Radiologists can use the proposed method as a computer-aided diagnosis tool to get a second opinion or to verify their findings.

References

- [1] H. Yoshida et.al., “An improved computer-assisted diagnostic scheme using wavelet transform for detecting of clustered microcalcifications in digital mammograms,” *Acad. Radiol.*, vol. 3, pp. 621-627, 1996.
- [2] R. N. Strickland and H. I. Hahn, “Wavelet transforms for detecting microcalcifications in mammograms,” *IEEE Trans. Med. Imag.*, vol. 15, pp. 218-229, Apr. 1996
- [3] M. Nafi Gurcan, Yasemin Yardimci, A. Enis Cetin, “Microcalcification Detection Using Adaptive Filtering and Gaussianity Tests,” *Proc. of the 4th Int. Workshop on Digital Mammography*, pp. 157-164, Nijmegen, The Netherlands, 1998.
- [4] M. Nafi Gürcan, Yasemin Yardımcı, A. Enis Çetin, Rashid Ansari, “Detection of Microcalcifications in Mammograms Using Higher Order Statistics,” *IEEE Signal Processing Letters*, vol. 4, no. 8, pp 213-216, 1997.
- [5] Omer N. Gerek, A. Enis Cetin, “Adaptive Polyphase Subband Decomposition Structures for Image Compression,” *IEEE Trans. on Image Processing*, pp. 1649-1660, Oct. 2000.
- [6] R. Ojeda, J. Cardoso, E. Moulines, “Asymptotically Invariant Gaussianity Test For Causal Invertible Time Series,” *Proc. of IEEE ICASSP*, vol. 5, pp. 3713-3716, 1997.

Authors' Affiliations:

A. M. Bagci, A. Enis Cetin, Department of Electrical and Electronics Engineering, Bilkent University, Ankara, 06533, Turkey. E-mail: cetin@ee.bilkent.edu.tr

Captions

Fig.1: Analysis stage of the 2-channel adaptive filter bank structure (P_1 represents an adaptive predictor).

Fig. 2: An 384 by 384 region of the image c06c: Peaks marked with 3,7,8; 4; and 5 are microcalcifications.

Table 1: Ranks given to MC's by algorithm based on adaptive wavelet transform followed by the HOS test (upper) and ordinary WT followed by the HOS test (lower).

.

.

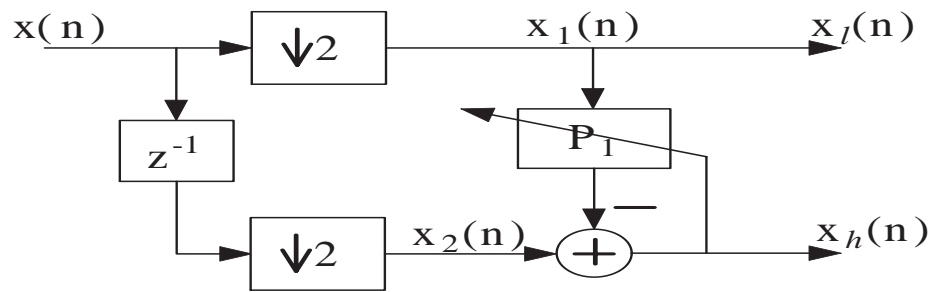


Figure 1: Analysis stage of the 2-channel adaptive filter bank structure (P_1 represents an adaptive predictor).

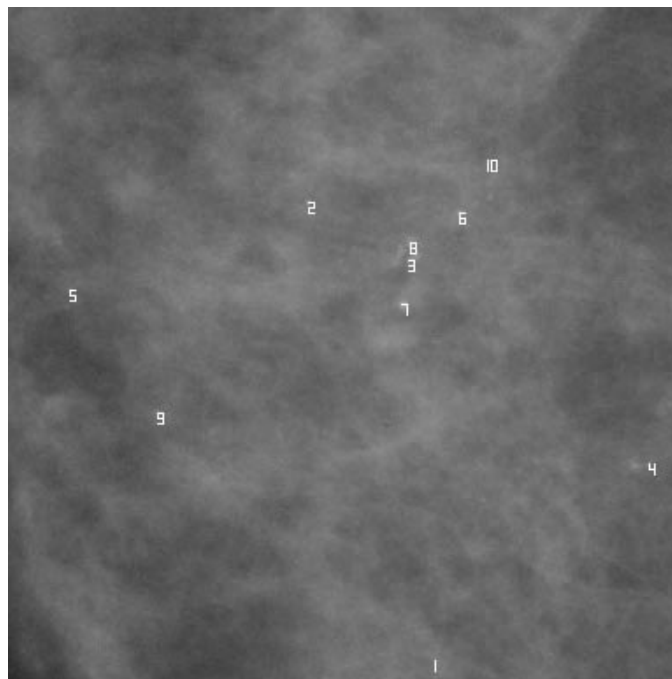


Figure 2: An 384 by 384 region of Image c06c: Peaks marked with 3, 7, 8; 4; and 5 are MC's.

Table 1: Ranks given to MC's by algorithm based on adaptive wavelet transform followed by the HOS test (upper) and ordinary WT followed by the HOS test (lower).

Image	Ranks given to microcalcifications		
	Region 1	Region 2	Region 3
c05o	2	5	
2 regions	14, 19	121	
c06c	5	3	4
3 regions	13	6	62
c07c	11		
1 region	11, 15		
c08c	2	11	4, 5
3 regions	1, 19	4, 5	13, 14
c09c	2		
1 region	3		
c10c	3, 4		
1 region	8, 11		
c13c	2		
1 region	1, 4, 6		
c14c	5	2, 3	
2 regions	13, 15	1	
c15c	3, 7		
1 region	24		
c16c	2, 3		
1 region	3, 4		
c18c	1, 3	5	
2 regions	3, 4	11, 17	
c19c	4	10	
2 regions	24	23	
c20c	4		
1 region	35		