USE OF 2D LINEAR PREDICTION ERROR TO DETECT MICROCALCIFICATIONS IN MAMMOGRAMS

Carmen Serrano*, Javier Díaz-Trujillo*, Begoña Acha* and Rangaraj M. Rangayyan**

*Área de Teoría de la Señal y Comunicaciones
Escuela Superior de Ingenieros, Universidad de Sevilla
Camino de los Descubrimientos, s/n. 41092-Sevilla, Spain

**Department of Electrical and Computer Engineering
University of Calgary, Calgary, Alberta, T2N 1N4, Canada

e-mail: cserrano@viento.us.es

ABSTRACT
We propose a new method to detect microcalcifications in mammograms. The method is based on region growing with pre-filtering and a seed selection procedure based on two-dimensional linear prediction error. The procedures are designed to reduce false positives, to improve detection capability, and to reduce computational time. When compared to a previous method proposed by Shen et al. [10], with the proposed method, computational time was reduced by up to 16% and detection accuracy was increased from 63% to 98%. The detection capability was 86% over all of the existing microcalcifications in three test mammograms.

Key words: mammogram, calcifications, breast cancer, region growing, 2D linear prediction error.

1. INTRODUCTION
The availability and proliferation of digital radiographic images have encouraged research in Computer-aided Diagnosis (CAD). A significant part of such research has concentrated on the detection of breast cancer, in view of the fact that women in Western countries have a higher than 1 in 10 chance of developing breast cancer during their life [3,2]. In particular, many researchers have focused on the detection of microcalcifications [10, 5, 11], which are early signs of breast cancer [4]. In this work, we propose an improvement to an algorithm proposed by Shen et al. [10] to detect microcalcifications based on region growing from seed pixels. The method has been proven to be effective in general, but fails in detecting microcalcifications in high-density mammograms; furthermore, its computational time is high. In order to reduce the high computational cost and to improve the capability to detect microcalcifications, especially those immersed in high-density regions appearing as bright or white background in mammograms, we propose a pre-filtering method and a seed selection algorithm based on two-dimensional (2D) linear prediction.

2. ALGORITHM DESCRIPTION
Region Growing Method: The region growing algorithm of Shen et al. [10] starts by selecting all pixels above the median gray level of the mammogram as seed pixels. For each seed, the 4-connected neighbors are checked for inclusion with the following condition:

\[(1 + \tau)(F_{\text{max}} + F_{\text{min}})/2 \geq p(i, j) \geq (1 - \tau)(F_{\text{max}} + F_{\text{min}})/2\]

where \(p(i, j)\) is the pixel being checked, \(F_{\text{max}}\) and \(F_{\text{min}}\) are current maximum and minimum pixels values of the region being grown, and \(\tau\) is the region growing tolerance parameter. The tolerance is varied between 0.01 and 0.4, with a step equal to the inverse of the seed pixel value. Features are extracted for regions grown for each value of the tolerance. The value that introduces the least change in the features from one step to the following one is chosen as the optimal tolerance, and the corresponding region and features are retained for further analysis. Finally, a region grown is treated as a microcalcification only if its size \(T\) in pixels meets the condition \(1 < T < 35\). This requirement is in view of the physical size of actual microcalcifications of interest being in the range 0.01 to 1mm², and due to the resolution of the images in our database being 35 pixels/mm².

Image Filtering: Considering the small size of microcalcifications, it is evident that a lowpass filter with a wide kernel can remove them from the image while conserving a background of high density in the image. Conversely, a highpass filter may be used to detect microcalcifications. Based upon a method proposed in our earlier work [1], we employ a highpass filter \(f(n) = 1 - g(n)\), where \(g(n)\) is a lowpass Gaussian function. In order to ensure that the filter does not eliminate any calcification, a variance of 2.75 and a filter support of 21x21 pixels were chosen.

Seed Selection: We propose a new method to select potential microcalcification points or seed pixels for subsequent region growing. The method consists of a 2D linear prediction error filter [7], followed by thresholding. A pixel is selected as seed for the region-growing algorithm if its prediction error is greater than an experimentally determined threshold. This is based
upon the observation that a microcalcification can be seen as a point of nonstationarity in an approximately homogeneous region or neighborhood in a mammogram; such a pixel cannot be predicted well by the linear predictor, and hence leads to a high error.

The adaptive linear predictor represents an $M \times N$ image as a 2D sequence $x(m,n)$, with $\{m,n\} = \{0,1,\ldots,M-1; 0,1,\ldots,N-1\}$. The basic assumption in linear predictive coding is that any pixel in an image may be predicted as a linear combination of a few neighboring pixels. The predicted value of a pixel $x(m,n)$ at $(m,n)$ is given by

$$\hat{x}(m,n) = \sum_{(i,j) \in \text{ROS}} a(i,j) x(m-i,n-j)$$

(1)

where $a(i,j)$ are known as the prediction coefficients and $\text{ROS}$ represents the region of support of the predictor, as shown in Fig. 1. The error of prediction is

$$e(m,n) = x(m,n) - \hat{x}(m,n)$$

(2)

and the mean-squared prediction error is

$$\varepsilon^2 = E\left[ e^2(m,n) \right]$$

(3)

where $E[\cdot]$ is the expectation operator. The prediction error values tend to have a smaller dynamic range and a more nonuniform distribution than the original pixel values [8].

![Fig. 1. Region of support of size $p_1 \times p_2$, for the predicted pixel P.](image)

The optimal prediction coefficients that minimize the mean-squared error are obtained by solving the 2D Yule-Walker equations:

$$\sum_{(i,j) \in \text{ROS}} a(i,j) r_{xx}(k-i,l-j) = \left\{ \begin{array}{ll} 0 & (k,l) \in \text{ROS} \\ \varepsilon^2 & (k,l) = (0,0) \end{array} \right.$$  

(4)

where $r_{xx}(k,l)$ is the 2D autocorrelation function of the image.

The computation of the prediction coefficients from the relationship above requires an estimate of the autocorrelation function of the image [9, 6]. However, the coefficients may be obtained directly from the data with the multichannel Burg algorithm [7]. In addition to computing the 2D linear prediction coefficients, the Burg algorithm may be used to compute the prediction errors directly from the data. This is advantageous in the present application as the error values are what are needed to select seeds for the region-growing algorithm.

The multichannel version of the Burg algorithm calculates the optimal prediction coefficients, for a $p_1 \times p_2$ 2D predictor, by computing prediction errors of order $p_i$. The prediction errors are represented as vectors of size $p_2 \times 1$ and are computed recursively, beginning with 0-order prediction errors. The error values are initialized to the original image as shown in Fig. 2. The prediction errors of higher orders are calculated by repeating recursively the following three steps, with $i = 0,1,\ldots,p_i-2$ [7]:

1. Compute the covariance matrices for the forward and backward prediction errors as:

$$E^f_i = \sum_m e^f_i[m] e^f_i[m]^T$$

$$E^b_i = \sum_m e^b_i[m] e^b_i[m]^T$$

$$E^{bf}_i = \sum_m e^f_i[m] e^b_i[m]^T$$

(5)

where $e^f_i(m)$ and $e^b_i(m)$ are, respectively, the forward and backward prediction errors of order $i$. As both of the error vectors are of size $p_2 \times 1$, $E^f_i$, $E^b_i$ and $E^{bf}_i$ will be $p_2 \times p_2$ matrices. The summation, depending on $m$, is performed for all the prediction error vectors of size $p_2 \times 1$ that the image can be partitioned into.

2. Calculate the prediction coefficient matrix $A_{i+1}[i+1]$ by solving the following equation:

$$E^f_i A_{i+1}[i+1] + A_{i+1}[i+1] E^f_i = -2E^{bf}_i$$

(6)

3. Compute the backward and forward prediction error vectors of the higher prediction order as:

$$e^f_{i+1}[m] = e^f_i[m] + A_{i+1}[i+1] e^b_i[m-1]$$

$$e^b_{i+1}[m] = e^b_i[m-1] + A_{i+1}[i+1] e^f_i[m]$$

(7)

Once we know the prediction error pixels, we use them as a measure of image stationarity at the corresponding points. Pixels where the prediction error is higher than a threshold are considered as points of nonstationarity, and therefore, seeds for the region growing algorithm.
3. EXPERIMENTAL RESULTS

The detection algorithm was tested with three mammograms containing a total of 428 microcalcifications of different nature and diagnosis. The mammograms were digitized with a Lumiscan75 scanner with resolution of 35 pixels/mm² or, equivalently, 6 pixels/mm, and 12 bits per pixel. Fig. 3a shows a 234×137 pixel portion of a mammogram with microcalcifications. It was experimentally determined that pixels with a prediction error higher than 125 were potential seeds for the region growing step to detect microcalcifications. The low threshold value (with 12 bits per pixel) is due to the small error value range, because of the homogeneity of mammograms and the good performance of the 2D prediction algorithm. Fig. 3b shows the result of this first step of the algorithm. Fig. 3c shows the calcification boundaries obtained by the region-growing step.

The results obtained with the algorithm were examined by a radiologist who determined the accuracy of the detection. Results, in comparison with the algorithm of Shen et al. [10], are summarized in Table I and Table II. From the tables, we can observe that the detection accuracy is increased significantly (lower number of false detections), although the detection capability is diminished, i.e., more calcifications were missed by the proposed method. In order to analyze the cause of this decreased detection capability, we examined which of the two parts of the algorithm (seed selection part and region growing part) missed microcalcifications. Table III shows the number of microcalcifications not detected.
by the seed selection part and those selected but not detected by the subsequent region-growing step. Results show that more than the 55% of the non-detected microcalcifications were marked by the seed-selection part, but were then rejected by the region-growing procedure. This indicates that, in addition to the improvement of the seed-selection algorithm that is proposed in the present work, an improvement in the pre-filtering step is required.

Table IV presents an estimation of the computational cost of both algorithms, when run on a Sparc Compliant Axil 320 workstation with 304 MB RAM.

**Table I**
Detection accuracy of the two algorithms over a total of 428 diagnosed microcalcifications

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Detected</th>
<th>Correct</th>
<th>False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shen et al.</td>
<td>646</td>
<td>409</td>
<td>237</td>
</tr>
<tr>
<td></td>
<td>63.3%</td>
<td>36.7%</td>
<td></td>
</tr>
<tr>
<td>Seed selection</td>
<td>374</td>
<td>366</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>97.9%</td>
<td>2.1%</td>
<td></td>
</tr>
</tbody>
</table>

**Table II**
Detection capability of the algorithms

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Existing</th>
<th>Correct</th>
<th>Missed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shen et al.</td>
<td>428</td>
<td>409</td>
<td>19</td>
</tr>
<tr>
<td>Seed select.</td>
<td>428</td>
<td>366</td>
<td>62</td>
</tr>
</tbody>
</table>

**Table III**
Analysis of missed microcalcifications: classification as those (a) not detected by the seed-selection part, and (b) detected by the seed-selection part but rejected by the region-growing part

<table>
<thead>
<tr>
<th>Image</th>
<th>(a) Not detected</th>
<th>(b) Rejected</th>
</tr>
</thead>
<tbody>
<tr>
<td>mammogram 1</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>mammogram 2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>mammogram 3</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>28</strong></td>
<td><strong>34</strong></td>
</tr>
<tr>
<td><strong>Percentage</strong></td>
<td><strong>45%</strong></td>
<td><strong>55%</strong></td>
</tr>
</tbody>
</table>

**Table IV**
Computational time required by the algorithms

<table>
<thead>
<tr>
<th>Image</th>
<th>Shen et al.</th>
<th>Seed selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>mammogram 1</td>
<td>93m 5s</td>
<td>78m 10s</td>
</tr>
<tr>
<td>mammogram 2</td>
<td>92m 50s</td>
<td>77m 46s</td>
</tr>
<tr>
<td>mammogram 3</td>
<td>92m 10s</td>
<td>78m 1s</td>
</tr>
<tr>
<td><strong>Average Time</strong></td>
<td><strong>92m 42s</strong></td>
<td><strong>77m 59s</strong></td>
</tr>
</tbody>
</table>

4. CONCLUSIONS AND FURTHER WORK

A new method for detecting microcalcifications is proposed in this work, based on the region-growing algorithm proposed by Shen et al. [10], augmented with pre-filtering and seed selection steps. The seed selection step uses 2D linear prediction error to detect potential microcalcifications. The proposed method shows a significant reduction in the computational time and an improvement in the detection accuracy; however, the number of correct detections was decreased. An analysis of the non-detected microcalcifications shows that it would be desirable to design a better filter that is tailored to the characteristics of the mammographic image to improve the detection performance. We are also investigating the design of adaptive filters for seed selection.

ACKNOWLEDGEMENTS

This work was supported by grants from the Natural Sciences and Engineering Council (NSERC) of Canada. We thank Liang Shen for providing the code for the region-growing algorithm and Dr. Pilar Serrano for the interpretation of the mammograms and the results.

REFERENCES

USO DEL ERROR DE PREDICCIÓN LINEAL EN 2D PARA DETECTAR MICROCALCIFICACIONES EN MAMOGRAMAS.

RESUMEN

Se propone un nuevo método para detectar microcalcificaciones en mamogramas. El método está basado sobre la región de crecimiento con prefiltrado y un proceso de selección basado en el error de predicción lineal en dos dimensiones. Los procedimientos están diseñados para reducir falsos positivos, incrementando la capacidad de detección y reduciendo el tiempo de cálculo computacional. Cuando se compara el método previo propuesto por Shen y colaboradores [10], con el método propuesto en este trabajo, el tiempo de ejecución computacional se reduce hasta un 16% y la exactitud en la detección es incrementada desde un 63% hasta un 98%. La capacidad de detección fue de 86% para todas las microcalcificaciones existentes en los tres mamogramas de prueba.