ENTROPY-BASED DETECTION OF MICROCALCIFICATIONS IN WAVELET SPACE

Giuseppe Boccignone, Member, IEEE, Angelo Chianese[1], Antonio Picariello[1]
DIIIE - Dipartimento di Ingegneria dell'Informazione e Ingegneria Elettrica
Universita' di Salerno - 84084 Fisciano (Salerno), Italy.
bebo@diiie.unisa.it
[1]DIS - Dipartimento di Informatica e Sistemistica
Universita' di Napoli - v. Claudio, 21 - 80125 Napoli, Italy.
picus@unina.it

ABSTRACT

In this paper we present a method for the detection of microcalcifications in digital mammographic images. Our approach is based on the wavelet transform, but differently from other techniques proposed in the literature, the detection is directly accomplished into the wavelet domain and no inverse transform is required. After a preliminary de-noising pass, microcalcifications are separated from background tissue. This is performed by exploiting information gained through evaluation of Renyi's entropy at the different decomposition levels of the wavelet space. Experimental results achieved on the standard Nijmegen data set are shown and discussed.

1. INTRODUCTION

Microcalcification analysis is an appealing tool for early breast cancer detection, but mammograms are among the most difficult of radiological images to interpret. A mammographic image I can be modeled at each point (x,y) as

$$I(x,y) = I_\mu(x,y) + I_\beta(x,y) + n(x,y) \quad (1)$$

where I_μ represents the brightness contribute due to microcalcifications, I_β the background breast tissue, and n the noise. Note that each microcalcification is very small (about 0.7 mm in size, 0.3 mm of average diameter) and the breast tissue inhomogeneous. Consequently, the clusters of microcalcifications result embedded in such varying background, while exhibiting low contrast and poor signal-to-noise characteristics. Furthermore, the distribution of noise in mammograms is actually unknown.

In the literature, several methods have been proposed for detection and segmentation aims. A complete review of the early ones has been given in [1].

More recently, methods based either on the Wavelet Transform (WT) [2,3] or on the hybrid techniques combining the WT with textural information [4] and neural networks [5], have also been proposed. The general approach of employing the WT for feature enhancement and detection is the following: compute the forward wavelet transform (FWT) of the image; modify the wavelet coefficients by a non-linear function; compute the inverse wavelet transform (IWT); eventually, if required, perform detection.

In our research we follow a different strategy. The microcalcification detection is directly performed in the transformed domain (briefly, wavelet space) and the IWT computation is avoided. After the FWT has decomposed the image at multiple scales (levels), a preliminary de-noising step is performed, in order to reduce noise contribute. Then, microcalcifications are separated from background tissue, by exploiting information gained through evaluation of Renyi's entropy at the different scales constituting the wavelet space.

In Section 2, we discuss the exploitation of the WT for object detection purposes and motivate the WT algorithm used in this work. In Section 3 we present de-noising and detection steps. In Section *4*, we provide results obtained on the set of *4*0 mammograms of the Nijmegen database.

2. WAVELET DECOMPOSITION

Be an image a finite energy function $I \in L^2(R^2)$, defined on a support $\Omega \subset R^2$. The continuous WT of the image I is the functional

$$W_a I(\vec{r}) = \frac{1}{\sqrt{a}} \int_\Omega \overline{\psi}\left(\frac{\vec{b}-\vec{r}}{a}\right) I(\vec{b})\, d\vec{b} \quad (2),$$

where a is the scaling term and $\vec{r}, \vec{b} \in \Omega$. ψ is the 'mother' wavelet satisfying regularity constraints [6,7]. $\overline{\psi}$ represents the conjugate complex of ψ.

We will deal with images canonically discretized on a square lattice (in the following, $I(i,j)$ represents the value of I at node (i,j) of the lattice).

Discretization of eq. 2 can be performed either using critical sampling, thus giving rise to orthogonal representations, or adopting non-critical sampling. A widely used orthogonal scheme for image processing applications is Mallat's algorithm [8]. However, the same algorithm might not be convenient for pattern recognition purposes: on the one hand, it is not shift-invariant, on the other hand image sub-bands result uncorrelated at the different scales. On the contrary, in the case of non-critical sampling, the discretization computes the WT on denser grids than their orthogonal counterparts; so, non critical sampling is known to produce a redundant representation. In signal analysis,

unlike compression applications, a redundant expansion of the signal is often desired. Therefore, several redundant decompositions have been proposed in the literature. One such scheme, the *a trous* (with holes) algorithm [6], is well known for its computational efficiency with respect to image analysis applications. The WT performed by this algorithm produces at each decomposition level l a so called wavelet plane $W^l=\{w^l(i,j)\}$ whose dimensions are equal to those of the original image I^0. The coefficients of the plane are computed as $w^l(i,j)=I^{l-1}(i,j)-I^l(i,j)$. $I^l(i,j)$ is obtained by applying a low-pass filter f to $I^{l-1}(i,j)$. However, it is worth noticing that by using the single wavelet plane W^l, no kind of directional information, like that provided by sub-bands in Mallat's decomposition, is available.

In a more recent work [9], Shensa has shown that the *a trous* algorithm bears an intimate relationship to Mallat's decomposition. Both can be considered as instances of a single filter bank structure, the discrete wavelet transform (DWT). The latter instance is simply the decimated output of the former. The decimated DWT is characterized by octaves obtained alternating the low-pass filter f with decimation and tapped by a band-pass filter g to produce the output (Mallat's algorithm). The undecimated DWT inserts i zeros between the elements of the filters in place of decimation (*a trous* algorithm). We exploit such property in two dimensions, thus obtaining an *a trous* algorithm which preserves sub-band information, as in Mallat's scheme. Our solution is represented in Fig. 1. In the latter, D^if and D^ig represents f and g with 2^i-1 zeros between each pair of filter coefficients and $w_{SBk}{}^l(i,j)$ are the *detail* wavelet coefficients in the sub-bands SB_k, k=1,2,3, at level l.

3. DETECTION

The model specified in eq. 1 can be represented in wavelet space by applying eq. 2 and exploiting linearity property:

$$WI(x,y) = W I_\mu(x,y) + WI_\beta(x,y) + Wn(x,y) \qquad (3).$$

The same remark holds for the discrete case. Clearly, the optimal detection strategy is to reinforce the WI contribute due to the microcalcifications, while reducing the one due to both background and noise.

As a first step, let us consider noise reduction. In general, the WT compresses the energy of the image into a small number of big coefficients, while noise preserves the same structure and spreads over all coefficients []. Thus, $Wn(x,y)$ can be partially removed from the data without affecting the $W I_\mu$ and WI_β components

The de-noising technique we use to select significant wavelet coefficients, is similar to Donoho and Johnstone's [10] shrinkage. The latter consists in the cancellation of those coefficients which are close to the noise levels. At a given scale, it is sufficient to compare the coefficients to the noise level in the sub-band SB_k, namely σ_{SBk}, choosing the significative ones as follows. For each detail sub-band SB_k, we empirically estimate the noise information as:

$$\sigma^l{}_{SB_k} = \sqrt{\frac{\sum_{i,j}(w^l{}_{SB_k}(i,j) - \overline{w}^l{}_{SB_k})}{\sum_{i.j} w^l{}_{SB_k}(i,j)}} \qquad (4),$$

$\overline{w}^l{}_{SB_k}$ being the average of $w^l{}_{SB_k}(i,j)$. The coefficients $w^l{}_{SB_k}$ are then transformed as $t^l(w^l{}_{SB_k}(i,j),\lambda)$ according to the following:

$$t^l(w^l{}_{SB_k}(i,j),\lambda) = \begin{cases} w^l{}_{SB_k}(i,j) & \text{if } w^l{}_{SB_k}(i,j) \geq \lambda \\ 0 & \text{otherwise} \end{cases} \qquad (5).$$

To calculate the threshold λ, we use the estimate:

$$\lambda = \sigma^l_{SB_k} \rho N^{-1/2} \qquad (6),$$

where ρ is a constant introduced for normalization aims, N is the number of wavelet coefficients, $\sigma^l{}_{SBk}$ is computed through eq. 4.

After the denoising step, detection is performed.

Let $d^l_{SB_k}(i,j) = t^l(w^l{}_{SB_k}(i,j),\lambda)$ be the shrinked detail coefficients. For each scale l, we use a linear combination of the weighted absolute values of such coefficients:

$$d^l(i,j) = \sum_{k=1}^{3} \alpha_k \, | d^l_{SB_k}(i,j) | \qquad (7).$$

It is worth noticing that the α weights are image dependent weights. For the purposes of the present work, these are determined so as to enhance information provided by sub-band 3 ("diagonal" details), with respect to the other sub-bands. Notice that the coefficient set $\{d^l(i,j)\}$ is a kind of *a trous* wavelet plane, but where spatial direction information has been taken into account.

Then, the $d^l(i,j)$ are used to extract microcalcifications from tissue background. We assume that such operation be a kind of object/background segmentation performed in wavelet space.

It is well known that for grey-level images object/background separation can be achieved by a thresholding process [11]. Notice that, in ideal bilevel thresholding, it is assumed that the probability distribution function (p.d.f.) of the grey levels is bimodal: the optimal threshold is chosen so as to coincide with the minimum of the p.d.f. The multilevel case is usually handled as an extension of the bilevel case. Obviously, these methods do not produce effective threshold values if the assumption of bimodality or multimodality is not met.

In our work we extend to the wavelet space, such techniques usually applied within the image domain. To this end, it is necessary to transform the coefficients $d^l(i,j)$ into distributions, and this can be done as follows. First, coefficient quantization is performed. Second, we consider the quantized coefficients $\tilde{d}^l(i,j)$ as a sequence of independent and identically distributed random variables whose p.d.f. is. $F(x) = \Pr\{\tilde{d}^l(i,j) = x\}$. Let

$$p^l{}_x = F^l(x) \cong n^l{}_x / N \qquad (8)$$

be an estimate of the probability distribution of the $\tilde{d}^l(i,j)$, where n^l_x represents the number of coefficients $\tilde{d}^l(i,j)$ assuming value x at level l, N is the total number of coefficients. We experimentally found that the p.d.f. of the coefficients may be interpolated by a generalized gaussian, namely $p^l{}_x = a^l e^{-|b^l x|^{r^l}}$, where a^l, b^l and r^l characterize the gaussian parameters at level l. In order to match the real p.d.f., the parameter r^l has been computed according to the χ^2 test (r^l=0.7). Notice that this result extends that of

Daubechies et al,.[7] referring to the p.d.f. of a single wavelet subband.

In consequence of the above result, the bimodal assumption must be abandoned in our specific case. Recently, some authors have shown how thresholding selection on a unimodal distribution may be efficiently performed by taking into account Renyi's entropy [12]. The a priori Renyi's entropy of order ν, with respect to the probability distribution Γ, can be defined as

$$H^l{}_\Gamma = \frac{1}{1-\nu}\ln\sum_{k=0}^{L}(p^l{}_k)^\nu \qquad (9),$$

where ν ($\nu \neq 1$) is a positive real parameter. We use Renyi's entropy as follows. From distribution $p^l{}_x$, estimated according to eq. 8, two probability distributions for the object class (microcalcifications) and for the background class (breast tissue), $\Gamma^l{}_1$ and $\Gamma^l{}_2$ respectively, are derived:

$$\Gamma^l{}_1 : \frac{p^l{}_0}{p^l(\Gamma^l{}_1)}, \frac{p^l{}_1}{p^l(\Gamma^l{}_1)}, ..., \frac{p^l{}_t}{p^l(\Gamma^l{}_1)}$$
$$\Gamma^l{}_2 : \frac{p^l{}_{t+1}}{p^l(\Gamma^l{}_2)}, \frac{p^l{}_{t+2}}{p^l(\Gamma^l{}_2)}, ..., \frac{p^l{}_L}{p^l(\Gamma^l{}_2)} \qquad (10),$$

where $p^l(\Gamma^l{}_1) = \sum_{i=0}^{t} p^l{}_i$, $p^l(\Gamma^l{}_2) = \sum_{i=t+1}^{L} p^l{}_i$, $p^l(\Gamma^l{}_1) + p^l(\Gamma^l{}_2) = 1$

The optimal threshold is that which maximizes $H^l{}_{\Gamma 1}(t)+H^l{}_{\Gamma 2}(t)$, and it is a function of ν. Sahoo et al. [] found by numerical simulation that

$$t^l(\nu) = \begin{cases} t^l_1 & 0<\nu<1 \\ t^l_2 & \nu \to 1 \\ t^l_3 & 1<\nu<\infty \end{cases}$$

The optimal threshold is thus chosen as

$$t_{opt}{}^l = \omega_1 t^l_1 + \omega_2 t^l_2 + \omega_3 t^l_3 \qquad (11),$$

where $\omega_1, \omega_2, \omega_3$ are weights determined by using local information [12].

Finally, the outcomes of the detection process, gathered at each level, are combined in a binary map $M(i,j)$, namely the *multiresolution support*. This is done according to the following rule:

$$M(i,j) = \bigcup_l M^l(i,j) \quad \text{with } M^l(i,j) = 1 \text{ if } d^l(i,j) > t^l_{opt} \qquad (12),$$

where $t^l{}_{opt}$ is calculated as in eq. 11.

Each connected set of non-zero locations of $M(i,j)$ is a microcalcification detected in the mammographic image.

4. EXPERIMENTAL RESULTS

The method has been tested on the Web available Nijmegen database [13]. This test set is formed by 40 mammograms, including both benign and malignant cases. Each mammogram is accompanied by a "truth" image produced by two expert radiologists.

In a preliminary stage, the detection of the single microcalcifications was considered. The aim was to tune method's parameters for maximizing the true calcification detection (true positive, TP) while minimizing the false detections (false positive, FP). Clearly, the choice of the decomposition levels to use is a trade-off between the size of the object to detect and the presence of noise. Our experiments have shown that a reasonable number of decomposition levels is l=4. The results of these preliminary experiments can be summarized as follows: a maximum average of about 10 FP are detected per image; the method detects the presence of microcalcifications in the same regions as in the accompanying truth images. Fig. 2 shows a typical example of achieved results.

A second set of experiments aimed at comparing our method performance with respect to others in the literature, in particular to the most recent one of Strickland [2]. To this end, microcalcification cluster detection is considered, counting true positive clusters (TPC) and false positive clusters (FPC). We adopt the standard cluster definition proposed by Karssemeijer [14]. A cluster is observed if more then two microcalcifications are localized inside a circular region of radius 0.5 cm, marked around each detected microcalcification. The cluster is then classified as a TPC if marked in the accompanying truth image, FPC otherwise.

By varying the wavelet basis, a typical receiver operating characteristic (ROC) curve is plotted (Fig 3). Evidence is given to the higher effectiveness of Daubechies' biorthogonal B-spline basis [7], with respect to TPC/FPC ratio. According to our method, we have as best result about 0.7 FPC and 66 % of TPC. On the same database, Strickland achieves 55% of TPC at the FPC rate of 0.7.

5. REFERENCES

[1] Dengler J., Behrens S. and Desaga J.F., "Segmentation of Microcalcifications in Mammograms", IEEE Transactions on Medical Imaging, vol. 12, n.4, December 1993, pp. 634-642

[2] Strickland R. N., Han H.I., "Wavelet Transform for Detecting Microcalcifications in Mammograms", IEEE Transactions on Medical Imaging, vol. 15, n.2, April 1996, pp. 218-229

[3] Dinten J.M. , Darboux M. and Nicolas E., "A Global Approach for Localization and Characterization of Microcalcifications in Mammograms", Proceedings 3rd International Conference on Digital Mammograms, Chicago, 1996

[4] Dahwan A.P., Citre Y, Bonasso C.K. and Moskowitz M., "Analysis of Mammographic Microcalcifications usign Gray Level Image Structure Feature", IEEE Transactions on Medical Imaging, vol. 15, n.3, June 1996, pp. 246-258

[5] Kocur C.M., *et. Al*, "Using Neural Networks to Select Wavelet Features for Breast Cancer Diagnosis", IEEE Engineering in Medicine and Biology, vol. 15, n.3, May/June 1996, pp. 95-101

[6] Vetterli M., Kovacevic J., *Wavelets and Subband Coding*. 1995, Prentice Hall, Englewood Cliffs, NewJersey

[7] Antonini M., Barlaud M., Mathieu P. and Daubechies I., "Image Coding Using Wavelet transform", IEEE Transactions on Image Processing, vol. 1, n.2, April 1992, pp. 205-220

[8] Mallat S., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", IEEE Transactions PAMI, vol. 11, n.7, July 1989, pp. 674-693
[9] Shensa M. J., "The Discrete Wavelet Transform: Wedding the à trous and Mallat Algorithm", IEEE Transactions on Signal Processing, vol. 40, n. 10, October 1992, pp. 2464-2482
[10] Donoho D. and Johnstone I., "Ideal spatial adaptation by wavelet shrinkage", Biometrika, vol. 81, n. 3, 1994, pp. 425-455
[11] Sahoo P.K., Soltani S.,Wong A.K.C. and Chen Y.C., "A Survey of the Thresholding Techniques", Computing Vision, Graphics Image Processing, vol. 41, 1988, pp. 233-260
[12] Sahoo P.K., Wilkins C. and Yeager J., "Threshold Selection using Renyi's Entropy", Pattern Recognition, vol. 30, n.1, 1997, pp. 71-84
[13] http://marathon.csee.usf.edu/Mammography/Nijmegen
[14] Karssemeijer N., "Adaptive Noise equalization and Recognition of Microcalcifications Clusters in Mammograms", IJPRAI, vol. 7, n. 6, December 1993, pp. 1357-1377

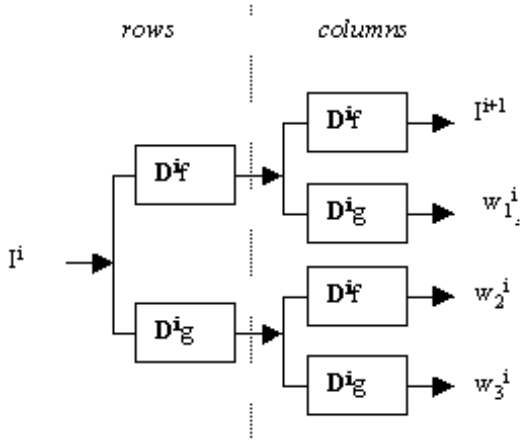

Figure 1. The proposed decomposition scheme at level l.

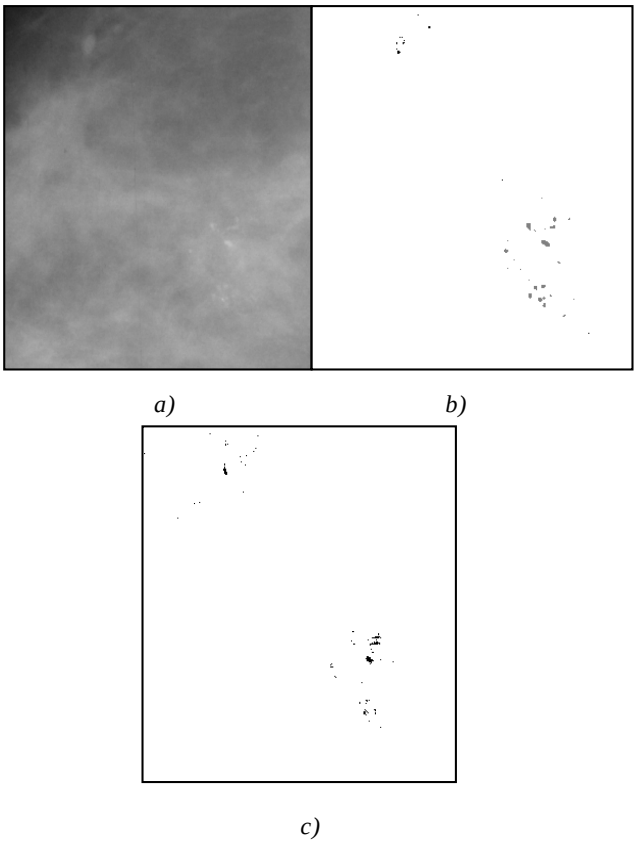

Figure 2. An example of detection. a) Original image. b) Truth image. c) Detected microcalcifications

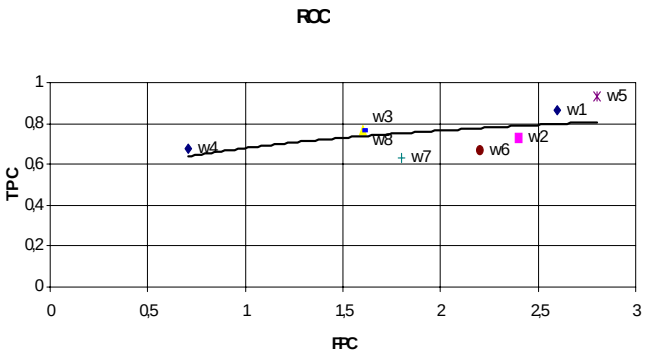

Figure 3. Cluster detection performance measured on 40 images of the Nijmegen database. The following bases have been used[7]: Burt-Adelson (W1), Battle-Le Marie (W2), B-spline I (W3), B-spline II (W4), Daubechies 4 (W5), Daubechies 6 (W6), Daubechies 8 (W7), Daubechies 10 (W8)