Computer aided detection of microcalcifications in digital mammograms

Giuseppe Boccignone\textsuperscript{a,}\textsuperscript{*}, Angelo Chianese\textsuperscript{b}, Antonio Picariello\textsuperscript{b}

\textsuperscript{a}DIIIE and INFM, Università di Salerno, via Ponte Melillo 1, Fisciano, 84084, Italy
\textsuperscript{b}DIS, Università di Napoli, via Claudio 21, Napoli, 80100, Italy

Received 4 November 1999; accepted 13 April 2000

Abstract

Microcalcification detection is widely used for early diagnosis of breast cancer. Nevertheless, mammogram visual analysis is a complex task for expert radiologists. In this paper, we present a new method for computer aided detection of microcalcifications in digital mammograms. The detection is performed on the wavelet transformed image. The calcifications are separated from the background by exploiting the evaluation of Renyi's information at the different decomposition levels of the wavelet transform. Experiments are performed on a standard and publicly available dataset and the results are evaluated with respect to recent achievements reported in the literature. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Computer aided diagnosis; Image processing; Mammogram analysis; Microcalcification detection; Wavelet transform; Renyi’s entropy

1. Introduction

Breast cancer is by far the most common cancer among women and X-ray mammography is a very important tool for its early detection. Radiologists go through visual analysis of mammograms looking for certain abnormalities as signs of breast cancer. These signs are the presence of clusters of calcifications, spiculated lesions and circumscribed masses.

Unfortunately, mammograms are among the most difficult of radiological images to
interpret; in particular, microcalcification visual assessment turns out to be an actual challenging task. A microcalcification is a tiny calcium deposit that has accumulated in the breast tissue, and it appears in the mammogram as a small bright spot embedded within an inhomogeneous background. The calcifications vary in size from smaller than 0.1 mm to 5 mm in diameter, and a radiologist must carefully examine the mammogram with a magnifier to locate calcifications which may be embedded in dense tissue. Size, shape and radiographic density are the most important factors on analysing individual calcifications, whereas their number and distribution are taken into account when clusters are considered, a cluster being defined as a group of at least three to five microcalcifications within 1 cm² region.

Indeed, computer-aided analysis of microcalcifications in digital mammograms is widely reputed as a remarkable goal, and several methods have been proposed in the literature for their detection and segmentation. The effectiveness of such techniques can be assessed taking into account True Positive (TP) and False Positive (FP) detections [1]. FP is the probability of incorrectly classifying a non target object as a target object — the microcalcification — while TP detection rate is the probability of correctly classifying a target object.

One of the early methods achieving significant clinical results is that of Davies and Dance [2], with experiments performed over 50 test images, half of which contained no clusters. The authors report a 96% TP with an average of 0.18 FP clusters per image. However, the authors detect suspicious areas by using a local threshold and, to such end, the selected regions are limited by size criteria; further, those with an irregular shape are discarded.

The method due to Dengler et al. [3] exploits a two-stage algorithm for spot detection and shape extraction, based on Gaussian filter detection followed by morphological reconstruction. They report 70% of TP and 0.3 FP. Such a result is tested using a number of mammograms evaluated through the judgments of expert radiologists, although the mammographies are not publicly available and the outcome cannot be compared with other methods.

Shen et al. [4] propose a multi-tolerance region growing method. The resulting regions are then given in input to a neural network for classification. The results are the following: in mammograms containing benignant tumors, they achieve 81% of TP and zero FP, while 85% TP and 29 FP of malignant cases are reported. It has to be noted that the experiments, albeit biopsy proved, have been performed on a very low number of images, four real mammograms, that are not publicly available.

Other methods aim at incorporating a priori knowledge within the analysis phase. A fundamental contribution is that of Karssemeijer [5]. The method exploits structural geometric knowledge and it relies upon bayesian statistic techniques and upon the application of random field models. The pixels are assigned to a certain class, maximising the probability of belonging to one of the four considered classes: background, micro-calcifications, lines/edges and errors due to film erosion. The experimental work is extensive and based on a Web available database. As reported by the author himself, the method is very complex and computationally expensive.

Multi-scale detection has also been used for clustered microcalcifications. Motivations for adopting an image multi-scale representation ground in the wealth of physiological and psychophysical data demonstrating that the visual system analyzes images at different resolutions: visual information is processed in parallel by a number of spatial-frequency-tuned channels and in the visual pathway, filters of different size operate at the same location [6].
Brzakovic et al. [7] propose a detection method based on the Laplacian fuzzy pyramid, with experiments performed over both synthetic and real images. On synthetic images, the results vary from 60% to 80%, with an average of 2 FP per image due to the complexity of the background. The experiments on real images have been performed on the database distributed from the University of South Florida formed by 67 images (17 containing abnormalities and 50 without abnormalities). However, ROC analysis is not reported.

Recently, multi-scale methods based on the wavelet transform (WT) have been introduced [8–11].

In particular, Strickland’s approach [10] is appealing as regards both the results achieved and the methodology. Microcalcification detection is directly accomplished within the transform domain, relying on a thresholding of the wavelet coefficients to produce a detect/no detect result. However, the threshold is experimentally chosen as a fixed percentile of the histogram of each channel, thus limiting the approach in the capability of dealing with varying conditions due to the image formation and digitization process, noise level, etc.

In this paper, we propose a multi-scale detection of microcalcifications based on a wavelet representation of the digital mammogram. As in [10], the detection is directly accomplished within the transform domain, but, in contrast, our approach automatically determines the threshold function relying on an information theoretic tool, namely Renyi’s entropy. Successively, the estimates of objects locations that we collect at the different scales of the wavelet transform are combined to obtain the final detection.

In this way, images with varying characteristics can be successfully segmented using one procedure and without the need of a priori fixed thresholds.

In the following we will describe the theoretical foundation of the proposed method and its algorithmic realization. Experiments performed are presented and discussed.

2. Theory

In this section, we discuss and motivate the algorithm proposed to automatically detect the microcalcifications. The core of the algorithm relies on a representation of the image derived from the WT.

2.1. Wavelet transform computation

Since the method we present exploits the properties given by the WT representation of an image, we review some of the basic properties, with special reference to the two-dimensional theory. An in-depth presentation is given in [12]. The WT has been introduced to get some spatial information in the Fourier space. As long as we are satisfied with linear space invariant operators the Fourier transform is suitable for a wide range of applications such as image transmission or stationary image processing. However if transient phenomena — for instance, an object located in the right corner of an image — are to be investigated, the Fourier transform becomes a cumbersome tool. The uncertainty principle states that the energy spread of a function and its Fourier transform cannot be simultaneously arbitrarily small. If a specific
A frequency is detected in a Fourier spectrum, it is difficult to determine its spatial origin in the signal.

Following the pioneering work of Wigner in the context of quantum mechanics [13] and Gabor [14], Morlet and Grossmann formalized the continuous WT [15]. The basic idea is to use a family of functions localized both in space and in the frequency domain. These functions are related to each other by translations and by changes of scale. The frequency variable used with the Fourier transform is then replaced by the scale parameter.

Formally, be an image a finite energy function $I \in L^2(\mathbb{R}^2)$, defined on a support $\Omega \subset \mathbb{R}^2$, $\mathbb{R}$ being the set of reals. The representation of the image according to a family of wavelet functions is obtained as follows. First, a function $\psi \in L^2(\mathbb{R}^2)$, the 'mother wavelet', is chosen, which satisfies the admissibility condition

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\hat{\psi}(\omega_x, \omega_y)|^2 d\omega_x d\omega_y < \infty,$$

where $\hat{\psi}(\omega_x, \omega_y)$ is the Fourier transform of $\psi$, and $\omega_x, \omega_y$ are the spatial frequency components along the $x, y$ axes. Such condition is a regularity constraint so that $\psi$ have zero mean value and local oscillation quickly decaying to zero. Secondly, a family of functions is derived from scaling and translation of the mother wavelet:

$$\psi_{u_x, u_y}(x, y) = \frac{1}{|s|^p} \psi\left(\frac{x - u_x}{s}, \frac{y - u_y}{s}\right),$$

where $s$ is the scaling term and $u_x, u_y$ are the translational parameters. For the parameter $p \geq 0$, different choices are proposed in the literature ($p = 1, p = 1/2$), but its actual value is completely irrelevant for the basic theory; thus, in the following we will assume $p = 1$.

The continuous WT of $I$ at a scale $s$ and at position $(x, y)$ is computed by decomposing $I$ on such a family:

$$W_s I(u_x, u_y) = \langle I, \psi_{u_x, u_y, s} \rangle = \frac{1}{s} \int_{\Omega} I(x, y) \overline{\psi}\left(\frac{x - u_x}{s}, \frac{y - u_y}{s}\right) dx dy,$$

where $\overline{\psi}$ being the conjugate complex of $\psi$, and $\langle \rangle$ denoting the inner product. The square integrability of $\psi$ and the admissibility condition ensure that the transformation is reversible.

Intuitively, this transformation can be seen as a mathematical microscope whose position and magnification are $(u_x, u_y)$ and $1/s$, respectively, and whose optics is given by the choice of the specific wavelet $\psi$. Such a 'gauging' property is illustrated in Fig. 1.

In practice one deals with digital images, so $I(x, y)$ is canonically discretized on a square lattice (we will denote $I(i, j)$ the value of $I$ at node $(i, j)$ of the lattice). Taking into account a discrete scaling $s = s_0^j$ based on a dilatation step $s_0 > 1$, the translational parameters depend on this scale and on a translational step $u_0$:

$$u_x = nu_0s_0^j,$$

$$u_y = mu_0s_0^j$$
where \((n, m, l) \in \mathbb{Z}^3\). Then, the discrete wavelet transform is defined as the decomposition on the family of functions

\[
\psi_{n, m, l} = \left[ \frac{1}{s_0^l} \psi \left( \frac{x - n u_0 s_0^l}{s_0^l}, \frac{y - m u_0 s_0^l}{s_0^l} \right) \right]_{n, m, l}.
\]

Depending on the type of sampling, that is on the \(s_0\) and \(u_0\) values, the transform can be more or less redundant. In the case of critical sampling \(s_0 = 2\) and \(u_0 = 1\), the wavelet family can be chosen as an orthonormal (ON) basis. On the other hand, if one is tied to specific constraints, e.g. a symmetric shape of the wavelet function, then the orthonormality requirements must be relaxed by employing biorthogonal bases [12].

A fast algorithm to compute an ON transform, which is widely used for image processing applications, has been given by Mallat [16]. According to this algorithm, ON wavelet basis can be obtained by using two quadratic mirror filters \(H\) and \(G\) which are, respectively, a low pass and a band-pass filter.

The computing process is made recursive by filtering the image along the \(x, y\) axes. The resulting filterings by \(H\) along both axes give an approximation (smoothing) of the image on a scale of \(2^{-l}\). Similarly the filterings by \(H\) along the \(x\) axis and by \(G\) along the \(y\) axis, by \(G\) along the \(x\) axis and by \(H\) along the \(y\) axis, and by \(G\) along both the \(x\) and \(y\) axes provide the wavelets coefficients on a scale of \(2^{l-1}\).

The sub-images representing the smoothed image and the wavelet coefficients are named sub-

![Level 0](image1)

![Level 1](image2)

![Level 2](image3)

Fig. 1. An example of WT using B-Spline mother wavelet [17] over a scan line extracted from a mammography containing microcalcifications. At level 2 of the decomposition, a signal peak is shown corresponding to the target, whose magnitude is higher than the magnitude shown at level 1.
bands, which we will indicate as $SB_k$, $k = 0, 1, 2, 3$. Fig. 2 represents one stage of the decomposition. The filters $H$ and $G$ are one dimensional filters. This transform provides a decomposition of the original image into a set of independent, spatially oriented frequency channels (the sub-bands) at different resolution levels. Note that the algorithm can be adopted by biorthogonal bases too [12].

The representation properties of the WT can be grasped at a glance by considering Fig. 3. It shows a portion of mammogram decomposed according to Mallat’s scheme at $l = 1, 2, 3$ resolution levels.

In the figure, the channels are represented as sub-images. It can be noted that the $SB_0$ sub-image contains low spatial frequencies in both horizontal and vertical directions; in $SB_1$ the horizontal direction has low-frequencies, and the vertical one has high frequencies, thus representing vertical details of the original image, while the opposite holds for $SB_2$; eventually, the $SB_3$ sub-image shows high spatial frequencies along both directions, thus focusing on diagonal details.

2.2. Wavelet representation of mammographic images

The ON wavelet transform has been successfully applied for image processing and image compression. However, the same algorithm might not be convenient for image analysis purposes: on the one hand it is not shift-invariant, on the other hand image sub-bands result uncorrelated at the different scales.

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 à trous (with holes) algorithm [18], is well known for its computational efficiency with respect to image analysis applications. The WT performed by this algorithm produces a so-called wavelet plane $W^l = \{ w^l(i, j) \}$ at each decomposition level $l$, whose

![Diagram of Mallat's decomposition](image)

Fig. 2. One stage of Mallat’s decomposition.
dimensions are equal to those of the original image $I_0$. The coefficients of the plane are computed as $w'_l(i, j) = I_{l-1}^i(i, j) - I_l^i(i, j)$. $I_l^i(i, j)$ is obtained by applying a low-pass filter $H$ to $I_{l-1}^i(i, j)$. However, it is worth noticing that no kind of directional information, like that provided by sub-bands in Mallat’s decomposition, is available by using the single wavelet plane $W_l$.

In a more recent work, Shensa has shown that the à trous algorithm bears an intimate relationship to Mallat’s decomposition [19]. Both schemes can be considered as instances of a single filter bank structure, the discrete wavelet transform (DWT). The Mallat’s scheme is simply the decimated output of the DWT. As seen above, the decimated DWT is characterized by octaves obtained by alternating the low-pass filter $H$ with decimation and tapped by a band-pass filter $G$ in order to produce the output (Mallat’s algorithm). The undecimated DWT inserts $2^l-1$ zeros between the elements of the filters in place of decimation (à trous algorithm, Fig. 4).

It is well known that wavelets provide a suitable framework for non-linear estimation. So a possible solution to our initial problem is to recast the microcalcification detection problem as a non-linear signal estimation problem. To such purpose the observed image can be modeled as:

$$I = I_\mu + I_\beta,$$

where $I_\mu$ and $I_\beta$, respectively, represent the brightness contributes due to microcalcifications, and to inhomogeneous background plus noise.

By applying a wavelet transform $W$ to Eq. (2) and by exploiting the linearity property, it is possible to give a local description of the transformed domain as $WI = WI_\mu + WI_\beta$ where $WI$ indicates the collection $\bigcup_{k=1}^{3} w_{SB_k}^l(i, j)$, $w_{SB_k}^l(i, j)$ being the detail wavelet coefficient at position

Fig. 3. Wavelet decomposition of a mammogram following Mallat’s scheme and considering 3 levels of resolution (the original image is assumed, by default, at $l = 0$).
(i, j) in the sub-band \(SB_k\), at level \(l \in [0, L - 1]\). Notice that we are not considering the low frequency band \(SB_0\) (smoothed image), since useful information mostly resides in the detail sub-bands.

At each level \(l\), a representation plane \(\{w^l(i, j): 0 \leq i < N, 0 \leq j < M\}\) is built, where each coefficient is the weighted linear combination:

\[
w^l(i, j) = \sum_{k=1}^{3} a_k | w^l_{SB_k}(i, j) |
\]

Note that, differently from the classical \(à\) trous algorithm, we obtain a wavelet plane in which the directional information of the original sub-bands \(SB_k\) is implicitly retained through the \(a_k\) weights. Upon each plane, candidate microcalcifications are to be estimated by thresholding.

### 2.3. Wavelet thresholding and microcalcification estimation

In order to perform a non-linear estimation, a thresholding of the wavelet coefficients can be accomplished. Note that wavelet thresholding, as suggested by Donoho [20], has been actually employed for de-noising rather than true detection. Nevertheless the idea of using coefficient thresholding specifically for object localization is not new, as discussed in section 1 [10]. The original contribution of our approach consists into an automatic determination of the threshold function, relying on an information theoretic tool, namely Renyi’s information. Motivations are the following.

Consider at each level \(l\) the distribution \(p'(x)\) of the wavelet coefficients \(x = w^l(i, j)\). In principle, the distribution \(p'(x)\) can be conceived as the mixture of two distributions \(\mu'(x)\) and \(\beta'(x)\), namely the distributions of ‘microcalcification coefficients’ \(w^l_{\mu}\) and ‘background
coefficients’ $w_l^i$. More precisely, $\mu^l(x)$ and $\beta^l(x)$ are defined on the supports $\Omega_{\mu}$ and $\Omega_{\beta}$ both depending upon the parameter $l^t$, that is $\Omega_{\mu}^l(l^t) = \{x: x^l_{\min} \leq x \leq l^t\}$ and $\Omega_{\beta}^l(l^t) = \{x: l^t < x \leq x^l_{\max}\}$, where $x^l_{\min} = \min\{w^l(i, j)\}$ and $x^l_{\max} = \max\{w^l(i, j)\}$. Thus, we suppose that $\mu^l(x) \rightarrow 0$ on $\Omega_{\mu}^l(l^t)$ and that $\beta^l(x) \rightarrow 0$ on $\Omega_{\beta}^l(l^t)$.

In order to detect the objects of interest, at a fixed level $l$ of the transform, one should determine $l^t$ such that the functional $\int_{\Omega^l} (\mu^l(x) - \beta^l(x))^2 \, dx$, representing the distance between the two distributions, be maximized over $\Omega^l = \Omega_{\mu}^l(l^t) \cup \Omega_{\beta}^l(l^t)$. Denote

$$V_r(x) = \frac{1}{1-r} \ln \left( \int_{\Omega_x} x(x)^r \, dx \right)$$

(4)

the $r$-th order Renyi’s information of the distribution $x(x)$ over the support $\Omega_x$ [21]. It is easy to show (Appendix A) that the optimum $l^t$ can be suitably approximated by maximising the second order Renyi’s information of $\mu^l(x)$ and $\beta^l(x)$ with respect to $l^t$. Namely,

$$l^t = \text{Arg} \max \{V_r(\mu^l) + V_r(\beta^l)\},$$

(5)

where $r = 2$.

3. Implementation

3.1. Algorithm realization

The algorithmic implementation of the proposed method was approached through a series of phases: wavelet transformation of the digital mammographic image, average sub-band formation, coefficient quantization, computation of the thresholding values, thresholding, final microcalcification detection.

The first phase, WT computation, consists in exploiting an undecimated decomposition in two dimensions, so as to obtain an a trous algorithm that still preserves sub-band information, as in Mallat’s scheme. The scheme adopted is outlined in Fig. 4. $Df$ and $Dg$ represent $f$ and $g$ with $2^{i-1}$ zeros between each pair of filter coefficients.

The next phase takes into account the weighted composition, at each level, of the three sub-bands $SB_1$, $SB_2$, $SB_3$ according to Eq. (3). For the purposes of the present work, these are determined so as to enhance information provided by sub-band $SB_3$ with respect to the other sub-bands. The output of this phase is represented by the $l$ weighted sub-bands.

In order to determine the thresholds, it is necessary to transform the coefficients $\{w^l(i, j)\}$ into distributions, and this is done in the following phases. First, a quantization phase is performed. For each weighted sub-band, each coefficient value is quantized according to the following linear quantization:

$$\tilde{w}^l(i, j) = \text{trunc} \left[ \frac{K \cdot (w^l(i, j) - w^l_{\min})}{w^l_{\max} - w^l_{\min}} \right],$$

(6)

where $K = 2^k$ is the number of quantization levels, $k$ being the number of bits used for
The quantized coefficients \( \tilde{w}(i, j) \) are thus in a suitable form to determine probability distributions through histograms. To this end, we consider the coefficients as a sequence of independent and identically distributed random variables whose probability density function is \( p'(x) = \text{Pr}[\tilde{w}(i, j) = x] \).

Let
\[
p'(x) = \frac{n_x}{N}
\] (7)
be an estimate of the probability distribution of the \( \tilde{w}(i, j) \), where \( n_x \) represents the number of coefficients \( \tilde{w}(i, j) \) assuming value \( x \) at level \( l \), \( N \) is the total number of coefficients.

The continuous Renyi’s entropy of Eq. (4) is written in discrete form as:
\[
V_r(p') = \frac{1}{1 - r} \ln \sum_{k=0}^{K} p'(k)^r,
\] (8)
where \( r (r \neq 1) \) is a positive real parameter. We exploit Renyi’s entropy as follows. From distribution \( p'(x) \), estimated according to Eq. (7), two probability distributions for the object class (microcalcifications) and for the background class (breast tissue), \( \mu^l \) and \( \beta^l \), respectively, are derived:
\[
\mu^l = \frac{p'(0)}{p'_{\mu^l}}, \frac{p'(1)}{p'_{\mu^l}}, \ldots, \frac{p'(l)}{p'_{\mu^l}}
\]
\[
\beta^l = \frac{p'(l + 1)}{p'_{\beta^l}}, \frac{p'(l + 2)}{p'_{\beta^l}}, \ldots, \frac{p'(K)}{p'_{\beta^l}},
\] (9)
where \( p'(\mu^l) = \sum_{i=0}^{l} p'(i), p'(\beta^l) = \sum_{i=l+1}^{K} p'(i) \), with \( p'(\mu^l) + p'(\beta^l) = 1 \).

The optimal threshold \( t^l \) is that which maximizes the total Renyi’s entropy \( V_r(\mu^l) + V_r(\beta^l) \), and it is calculated according to Eq. (5). Practically, for each quantization level \( 0 \leq i < K \), \( V_r(\mu^l) \) and \( V_r(\beta^l) \) are computed using Eqs. (8) and (9), then summed to obtain the total Renyi’s entropy; the maximum value of the total entropies is the one for which \( i \) corresponds to the optimal threshold \( t^l \).

By using \( t^l \), at each level the detection is performed by constructing a binary map \( M^l(i, j) \) as a truth table such that \( M^l(i, j) = 1 \), if \( \tilde{w}(i, j) > t^l \), and \( M^l(i, j) = 0 \) otherwise.

The resulting \( M^l(i, j) \) are combined to produce a detection map \( M(i, j) \). To this end, we have adopted a majority vote criteria [22]: each \( M^l(i, j) \) is considered as an expert who votes for one class (i.e., the microcalcification class and the non-microcalcification class) and the estimated class is the one voted by the majority. The vote is also weighted by a number representing the reliability of the expert assigning a sample to the class. In this way, we selectively reward each decomposition level according to its information content as respect to the other ones (e.g., levels 2 and 3 are preferred as respect to level 1, which contains either details and noise). Eventually, the detection map \( M(i, j) \) is formed, where each connected set of non-zero locations of \( M(i, j) \) represents a target detected in the image.

The algorithmic realization of the proposed method can be summarized as a series of steps. These steps can be separated and listed as the following.
1. Computation of the wavelet transform of the digital mammogram.

2. For each decomposition level the following sub-steps are repeated.
   2.1. A weighted average of all the details images is formed
   2.2. The average detail image is quantized
   2.3. The threshold selection is accomplished by using Renyi’s entropy.
   2.4. The quantized image is thresholded and the binary map formed as an estimation of the microcalcification position.

3. The multiresolution supports that are produced at the various levels of the decomposition are combined in the detection map according to a majority vote policy.

When all the steps are completed, the connected regions of the latter map represent the detected microcalcifications.

3.2. Image data set

For evaluation purposes, experiments have been performed on the standard Nijmegen data set. The 40 images of this data set have been made available by the Department of Radiology, University Hospital Nijmegen, The Netherlands, and can be obtained via anonymous ftp from figment.csee.usf.edu in directory pub/mammograms/nijmegen-images.

All images are in raw format, of size 2048 by 2048 pixels, 12 bits per pixel of grey level information. The images were digitized from film using an Eikonix 1412 12 bits CCD camera. A sampling aperture of 0.05 mm in diameter was used, and a 0.1 mm sampling distance too. The images were corrected for inhomogeneity of the light source (Gordon plannar 1417). A fixed calibration of the CCD camera was adopted. Optical density 0.18 corresponds to the maximum output level (4095).

The position and size of the microcalcification clusters were marked by two expert radiologists, based on all patient data available (different views, magnifications). These annotations were put into the computer as circular areas and are stored in separate files.

This is the same data set used by Karssemeijer in [5] and other recent papers (for a comprehensive review see [23] and [24]).

3.3. Experimental framework

A first prototype of the system has been written in C++ language running on a Pentium 300 under Windows 95/Windows NT Operating Systems, and developed upon the FLOWeR® platform, an innovative object-oriented radiological information system produced by Integris-GROUPE BULL [25].

4. Procedures and results

In a preliminary stage, the aim was to tune method’s parameters for maximising the true detections while minimising false detections. As a figure of merit, the detection of
microcalcification clusters was considered [5], by counting the ratio

\[ \text{TPC} = \frac{\text{#true positive clusters detected}}{\text{#true positive clusters to be detected}} \]

and the false rating

\[ \text{FPC} = \frac{\text{#false positive clusters detected}}{\text{#images}} \]

A cluster is observed if more than two microcalcifications are localized inside a circular region of radius 0.5 cm, marked around each detected microcalcification. The cluster is true positive if marked as such in the accompanying truth image; false positive, otherwise.

Three aspects have been taken into account: (1) the choice of the wavelet bases; (2) the number of decomposition levels; (3) the tuning of the \( \alpha_k \) weights in Eq. (3).

Standard wavelet bases have been used: Burt–Adelson, Battle–Le Marie, B-spline I, B-spline II, Daubechies 4, Daubechies 6, Daubechies 8, Daubechies 10. The exact filter coefficients of such bases, and detailed descriptions can be found in [17] and [12]. The best TP/FP rate, at fixed parameters, has been obtained by biorthogonal B-spline II basis [17]. Fig. 5 displays in a graph the performance of the different bases.

For this and subsequent experiments, an acceptable number of decomposition levels resulted \( L = 4 \); clearly, this choice is a trade-off between the size of the object to detect and the presence of noise. As regards point (3), we initially set equal weights (Eq. (3)), i.e. \( \alpha_1 = \alpha_2 = \alpha_3 \).

The results of these preliminary stages, as far as individual microcalcifications are concerned, show that the method detects TP calcifications in the same regions as in the accompanying truth images and a maximum of about 10 FP cases are detected per image. Fig. 6 illustrates a typical detection example.

![Fig. 5. Cluster detection performance measured on 40 images of the Nijmegen database (\( \alpha_1 = \alpha_2 = \alpha_3, \ L = 4 \)). The following bases have been used: 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). W4 clearly shows the best TP/FP rate.](image)
A second set of experiments derived from a more subtle analysis of point (3). The \( \alpha_k \) value choice, \( 0 \leq \alpha_k \leq 1 \), implies a different weighting of the information in the three subbands \( SB_k \). A first reasonable assumption is that the strongest singularities (like background texture ones) are most likely to appear within the \( SB_1 \) band (diagonal details) rather than within \( SB_2 \) and \( SB_3 \) (vertical and horizontal details). A second assumption is that \( SB_2 \) and \( SB_3 \) should be equally weighted. By imposing that \( \alpha_1 + \alpha_2 + \alpha_3 = 1 \), with \( \alpha_2 = \alpha_3 \), it is possible to observe the TPC/FPC ratio versus the \( \alpha_1 \) variation in the \([0,1]\) range (FROC curve).

Our method was then compared with the most recent ones, in particular Strickland’s [10]. The obtained FROC is shown in Fig. 7. We obtained as best result 66% of TPC at the FPC rate of 0.7 (with \( \alpha_1 = 0.56 \)). On the same database, Strickland achieves 55% of TPC at the FPC rate of 0.7.

Fig. 6. An example of detection: (a) original image; (b) truth image; and (c) detected microcalcifications.
The average time needed to process a mammogram from the selected data base was about 1 minute on the described system.

5. Discussion

In its essence, the approach does not assume any contextual information about the presence or absence of targets in the image, neither an exact model of the background tissue and noise is needed. On the other hand, for effective use, the method requires some attention to be paid to the following points: the type of wavelet basis, the number of scales of the decomposition, and the \( \alpha \)-weighting of the information carried in the sub-bands. All these parameters are application dependent. However, the first two are generally common to any method relying upon the WT, whereas the \( \alpha \)-weighting is clearly specific of the method described here.

The choice of wavelets and the range of scales are two well known and crucial problems in the wavelet literature [12]. In particular, let us consider the choice of a suitable basis. If one wants to avoid the self-construction of a wavelet basis, then attention should at least be paid to the issues of orthonormality, compact support and regularity. Compact support improves the spatial resolution of wavelets. Wavelets are compactly supported if they have finite support with maximum number of vanishing moments for they support width; we recall that a wavelet \( \psi(x) \) has \( n \) vanishing moments if \( \int_{-\infty}^{\infty} x^k \psi(x) \, dx = 0 \), for \( 0 \leq k < n \). In general, a strict local support results in a lack of symmetry. Regularity relates to the order of differentiability and has a straightforward interpretation in terms of detectable structures. Since differentiation in Fourier domain amounts to a multiplication by \( j\omega \), existence of derivatives is related to sufficient decay of the Fourier spectrum. In other terms, regularity provides a sufficient decay of the mother wavelet in the frequency domain, which makes it continuous in the spatial domain.

![Fig. 7. Cluster detection performance measured by plotting TPC/FPC while varying \( z_0 \) (FROC curve).](image-url)
As reported in the previous Section, the following bases have been used: Burt–Adelson, Battle–Le Marie, B-spline I, B-spline II, Daubechies 4, Daubechies 6, Daubechies 8, Daubechies 10. Among these, B-spline II has achieved the best TP/FP rate. The reasons why one basis is better than the other is worth discussing at this point.

The different wavelet bases we employ for our method can be divided into orthonormal ones (Battle–Le Marie, Daubechies 4, Daubechies 6, Daubechies 8, Daubechies 10) and biorthogonal ones (Burt–Adelson, B-spline I, B-spline II). ON wavelets allow a compact coding and a perfect reconstruction of the image on one side, and for such reasons are and have been widely used, also in microcalcification analysis (namely, in the form of Daubechies bases [9,11]. However, on the other side, a trade-off has to be found between smoothness of the function, its shape and a strict finite support of the wavelet. Biorthogonal wavelet filters coefficients were chosen for this study because, due to their symmetry properties, they cause less image distortions than asymmetric filters such as the orthonormal Daubechies filters.

The Burt–Adelson basis (labelled as W1 in Fig. 5) is a quasi-orthogonal basis. It corresponds to the Laplacian pyramid filter proposed by Burt and Adelson. We employ it as a reference, general purpose basis since the Burt and Adelson decomposition has been widely used in the fields of image analysis and computational vision [6]. Specifically, we adopt it in the form proposed by [17], which is close to ON, nonsymmetric filters called ‘coiflets’. Due to the relationship between the number of vanishing moments and the symmetry of the function, these wavelet functions appear to be more symmetric than the Daubechies ones. However, in order to gain symmetry, some price has to be paid to the smoothness of the wavelet function. This results in an aptitude for capturing ‘spiky’ signals other than true microcalcifications. Such an effect can be noted in the graph of Fig. 5, where a relative higher rate of TP cases is achieved with respect to other bases, but at the cost of a high number of FP detections.

In contrast, the Battle–Lemarié wavelets (W2) represent a smoother example, than the Burt–Adelson basis. These wavelets are ON, polynomial spline wavelets that are computed from spline multiresolution approximations. Since the wavelet is a polynomial spline, it is continuously differentiable (which guarantees smoothness). As can be seen from the graph, Battle–Lemarié basis performs comparably to Burt–Adelson in terms of TP/FP ratio. Clearly, because of their improved smoothness, the false positive rate is reduced, but, meanwhile, some true positives are missed.

More interesting is to confront this basis against its non-orthonormal equivalent, namely B-spline wavelets (W3 and W4). These are the natural counterparts of the classic B-spline functions and are biorthogonal, regular wavelets of compact support. In this respect, B-spline I and B-spline II are very similar (same length and number of vanishing moments, precisely 4), but B-spline II is slightly more regular, which explains its smaller rate of FP detections. An essential property of these functions is their compact support, which makes them much better localized than the orthogonal Battle–Lemarié wavelets, thus achieving a better TP/FP rate on mammographic images.

It is worth noting that the best performance of the B-spline bases we have experimentally reported with respect to other bases, was not an unexpected result. They are a good approximation of the Laplacian of Gaussian function, which has been shown by Strickland to be a prewhitening matched (optimum) filter for objects of Gaussian shape (ideal microcalcification), embedded in background noise of Markov type. The suitability of B-spline
wavelets for processing images of the kind we are dealing with, also stems from a more general property, namely the asymptotic convergence of B-spline wavelets to Gabor functions (modulated Gaussian) [26]. It is well known that Gabor functions are optimally concentrated in both spatial and spatial-frequency domain, which endows them with a remarkable aptitude for the analysis of non-stationary signals and textures [6].

Eventually, we considered the Daubechies family of wavelets (W5, W6, W7 and W8) [12]. The bases of this family are usually named ‘Daubechies $k$', where $k$ denotes the length of the filter (and $k/2$ the number of vanishing moments). Daubechies wavelets are orthonormal, compactly supported and regular. These bases are in principle suitable for analysis of signals with finite support; further, they should provide a good combination of regular prototype wavelets with varying size to extract texture information with varying spatial frequency. The regularity of the Daubechies wavelets is proportional to the number of vanishing moments. For instance, the mother wavelet of the Daubechies 4 (2 vanishing moments) is more ‘spike-like’ compared with that of the smoother Daubechies 10 (5 vanishing moments): thus the former is eligible to detect more pixels of high spatial frequency, as opposed to the latter. These pixels may belong to microcalcifications, breast boundary, or background noise. In other words, the behavior of this family mirrors the fact that shorter wavelet filters are more sensible to existing microcalcifications but they tend to produce more false positives. Such behavior can be neatly seen in the graph displayed in Fig. 5, where the number of FP cases decrease as $k$ increases.

Once the wavelet basis has been selected, the range of scales can be derived by taking into account the size of the structure to be detected against the size of background structures.

In the case studied here, it is easy to forecast that most of the fine background structures are matched at the finest scale whereas they are discarded at subsequent, coarser scales ($l = 2, 3, 4$). Nevertheless, it has to be remarked that the voting procedure has given evidence of appreciable flexibility with respect to this specific problem (in the sense that, when the optimum coarsest scale $L = 4$ is relaxed to the $L = 6$ value, the system shows a very slight reduction of performance).

As regards the presented method, some particular attention should be devoted to the $\alpha$ - weighting of sub-band information. For this specific issue, ROC analysis, which is based on statistical decision theory, provides a reliable tool. We have employed it in the particular form of FROC curves, which constitutes a standard procedure in mammographic analysis. It is worth noting, however, that, in general, ROC curves provide objective estimates of probabilities of decision outcomes (in terms of true-positive and false-positive decisions), thus can be exploited for any and all of the criteria the system may have to tune.

6. Conclusion

In this study we have presented a new method for computer aided detection of microcalcifications in digital mammograms. The theoretical background of the method has been discussed and the theory substantiated with experiments on real mammographic images. These are part of a standard and publicly available database.

The method relies on a multi-scale representation of the mammographic image. Motivation
for adopting a multi-scale representation stems from the fact that microcalcifications reveal some evident features at some specific scale, while being invisible at other scales; on the other hand, directional information can be obtained at each scale in the form of sub-bands or detail images, in order to account for different kinds of image singularities — for instance, microcalcifications and background texture. In our case we choose a redundant representation based on an undecimated wavelet transform in order to avoid the lack of shift invariance and of spatial correlation across scales, which are typical of the most popular Mallat’s algorithm. Meanwhile, directional information has been taken into account by a weighted merge of the various sub-bands at a scale. Such a decomposition scheme has demonstrated to be a good trade-off between the effectiveness of completely redundant decomposition, like a Gabor representation, and the computational efficiency of an ON scheme.

Differently from other methods (see for instance [3,8]), microcalcification detection is thoroughly performed on the transformed image. The detection procedure relies on wavelet coefficient thresholding. In contrast to another approach [10], the threshold is automatically determined, at each scale, by maximising Renyi’s entropy. The Renyi’s entropy has deep roots in information theory [21] and to the best of our knowledge this is the first time it has been addressed for threshold selection in wavelet space.

So far candidate detections are proposed at each scale. Thus, a final combination stage, based on majority voting, selects the actual detected microcalcification.

It has to be noted that the only tuning specifically requested by the algorithm, apart from the selection of a suitable wavelet basis and the number of the decomposition scales, is the \( \alpha \) -weighting of the detail sub-bands. These parameters can be experimentally determined via ROC curves and no a priori knowledge is required. Therefore, while retaining effective detection, the algorithm avoids the complexity of other approaches [5].

Improvements to the method can be obtained in several ways. First of all, no correlation is considered among the several images of a single breast (cranio-caudal and lateral view). Taking into account both images may result in an enhancement of detection results.

Secondly, we argue that the TP/FP ratio may be improved if the detection results were given in input to a combination of classifiers or multiple experts. For instance, it is possible to conceive an architecture formed by wavelet detection modules coupled with neural networks trained at the several decomposition levels used as inputs of a validation net. It is likely that the use of a hybrid architecture would result in better performance than could be achieved by the method alone. However, this strategy should be followed only if the higher complexity of this approach were to be clearly counterbalanced by a noticeable improvement in detection performance.

7. Summary

This paper has been dealing with the application of computer image processing and analysis techniques to digital mammographic images. The purpose was the detection of clustered microcalcifications. Our algorithm was based on the wavelet transform. Unlike other techniques proposed in the literature, the detection is directly accomplished into the wavelet domain and no inverse transform is required. Further, microcalcifications are separated from
background tissue using a thresholding procedure. A novelty of the method is that the computation of the threshold is automatic and exploits Renyi’s entropy. The latter is evaluated at the different decomposition levels of the wavelet space and level dependent candidate detections are produced. Successively, these are combined via majority voting to achieve the final detection. Experiments are reported on the standard Nijmegen mammogram data set. Preliminary results are demonstrating a high clinical relevance of the system.

Appendix A

In order to detect the objects of interest, one should determine \( l' \) such that the distance between the two distributions \( \mu^l(x) \) and \( \beta^l(x) \) be maximized. We say that such distance represents the level \( l \) information, and we define it as the \( L^2 \) distance \( \int_{\Omega^l} (\mu^l(x) - \beta^l(x))^2 \, dx \) over the support \( \Omega^l = \Omega^l_{\mu}(l') \cup \Omega^l_{\beta}(l') \). Clearly, since

\[
\int_{\Omega^l} (\mu^l(x) - \beta^l(x))^2 \, dx = \int_{\Omega^l_{\mu}} \mu^l(x)^2 \, dx + \int_{\Omega^l_{\beta}} \beta^l(x)^2 \, dx - 2\int_{\Omega^l} \mu^l(x)\beta^l(x) \, dx
\]

the functional at the left-hand side of Eq. (10) is maximised if \( \int_{\Omega^l} \mu^l(x)\beta^l(x) \, dx \) is minimum. In fact, for \( \int_{\Omega^l} \mu^l(x)\beta^l(x) \, dx = 0 \) the two distributions can be considered ‘orthogonal’ (we exactly distinguish the targets from the background). If the left-hand side term of Eq. (10) denotes an ‘information’, the last term on the right-hand side can be conceived as a sort of ‘entropy’, which destroys information while increasing. Namely, denote \( H = \ln \int_{\Omega^l} \mu^l(x)\beta^l(x) \, dx \) the entropy we want to minimize for maximising information. By making use of the Cauchy–Schwartz inequality \( (\int_{\Omega^l} \mu^l(x)\beta^l(x) \, dx)^2 \leq \int_{\Omega^l} \mu^l(x)^2 \, dx\int_{\Omega^l} \beta^l(x)^2 \, dx \) and by taking into account Eq. (10), \( H \) can be upper bounded as

\[
2H \leq \ln \int_{\Omega^l_{\mu}} \mu^l(x)^2 \, dx + \ln \int_{\Omega^l_{\beta}} \beta^l(x)^2 \, dx
\]

Be \( I_{\Omega^l}(r) = 1/1 - r \ln(\int_{\Omega^l} z^r(x) \, dx) \) the \( r \)-th order Renyi’s information of the distribution \( z(x) \) over the support \( \Omega_x \) [21]. Then, from Eq. (11) follows

\[
H < -(I_{\Omega^l_{\mu}}(2) + I_{\Omega^l_{\beta}}(2))
\]

Consequently, the minimization of entropy \( H \) can be achieved by maximising the second order Renyi’s information of \( \mu^l(x) \) and \( \beta^l(x) \) with respect to \( l' \):

\[
l' = \text{Arg max } \lim_{r \to 2} \left\{ \frac{1}{1-r} \ln \left( \int_{\Omega^l_{\mu}} \mu^l(x)^r \, dx \right) - \frac{1}{2} \ln \left( \int_{\Omega^l_{\beta}} \beta^l(x)^r \, dx \right) \right\}
\]

QED.
References


Giuseppe Boccignone is an Assistant Professor and a Senior Researcher at the Faculty of Engineering of the University of Salerno, Italy. He received the Laurea degree in Physics from the University of Torino, Italy, in 1985. In 1986 he joined Olivetti Corporate
Research, Ivrea, Italy. From 1990 to 1992, he was a chief researcher of the Computer Vision Lab at CRIAI, Napoli, Italy. From
1992 to 1994, he held a Research Consultant position at Research Labs of Bull HN, Milano, Italy, leading projects on biomedical
imaging. Since 1994 he has been with the Dipartimento di Ingegneria dell'Informazione e Ingegneria Elettrica of the University of
Salerno and in 1997 he was appointed Research Associate at INFM (National Institute for the Physics of Matter). His current
research interests include physics based vision, medical image processing and video parsing. Giuseppe Boccignone is a member of
IEEE and IAPR.

Angelo Chianese is an Associate Professor of Computer Science at the Faculty of Engineering at the University of Naples ‘Federico
II’, Italy. He received the Laurea degree in Electronic Engineering from the University of Napoli, in 1980. From 1984 to 1992, he
was an Assistant Professor and a Researcher in the Computer Vision Lab of the Dipartimento di Informatica e Sistemistica of the
same university. His current research interests include image databases, distance education and distance learning, computer
networks, medical image processing and character recognition. Angelo Chianese is a member of AICA.

Antonio Picariello is an Assistant Professor and a Researcher at the Faculty of Engineering at the University of Napoli ‘Federico
II’, Italy. He received the Laurea degree in Electronic Engineering from the University of Napoli in 1991 and a Ph.D. in Computer
and Electronic Engineering in 1998 from the same university. Since 1995 he has been with the Dipartimento di Informatica e
Sistemistica of the University of Napoli. His current research interests include medical image processing, image databases, and
cursive script recognition.