Pattern Analysis & Applications (2000)3:303–313

Using Renyi's Information and Wavelets for Target Detection: An Application to Mammograms

G. Boccignone[1], A. Chianese[2] and A. Picariello[2]

[1]*DIIIE and INFM, Università di Salerno, Fisciano, Italy;* [2]*DIS, Università di Napoli, Napoli, Italy*

Abstract: In this paper we present a multi-scale method for the detection of small targets embedded in noisy background. The multi-scale representation is built using a weighted undecimated discrete wavelet transform. The method, in essence, is based on the maximisation of information available at each resolution level of the representation. We show that such objective can be achieved by maximising Renyi's information. This approach allows us to determine an adaptive threshold useful for discriminating, at each scale, between wavelet coefficients representing targets and those representing background noise. Eventually, avoiding inverse transformation, scale-dependent estimates are combined according to a majority vote strategy. The proposed technique is experimented on a standard data set of mammographic images.

Keywords: Mammograms; Medical imaging; Multi-resolution; Renyi's information; Scale-space; Target detection; Wavelet transform

1. INTRODUCTION

The problem of detecting targets through their automatic spatial localisation in a noisy background is of interest to several realms such as medical imaging, multispectral sensing, pattern recognition and information theory [1]. Generic targets exhibit a great variability of shape and appearance, scale and orientation, lighting and imaging conditions and natural background clutter. To address generic target detection, any method should in principle be powerful enough to cope with all such controversial features. Yet, it would be necessary to circumscribe the generality of these features by exploiting knowledge of the underlying nature of the world in which targets are generated and observed [2]. In practice, current research in this area is mainly dealing with algorithms relying on various restrictions on the applications.

In this work, we study the detection of small targets embedded within an inhomogeneous, textured background, and we assume that no other contextual knowledge is either taken into account or available. To be more precise, by detection we mean the spatial localisation of the targets, not being concerned with their exact shape reconstruction. To make progress, if we are not interested in gauging the structure of the background, the latter can be handled as an obscuring signal or 'noise'. Under this assumption, the target detection can be reformulated as a problem of signal/noise discrimination. This is generally known as de-noising problem: given an image, a finite energy function, $I \in \mathbf{L}^2(\mathbf{R}^2)$, the detection process can be expressed as an estimation problem of the 'true' but unknown signal I_τ (the target), hidden by a background 'noise' I_β, from the observed data I:

$$I(x,y) = I_\tau(x,y) + I_\beta(x,y) \tag{1}$$

where (x,y) is a point of the image domain.

Advanced statistical methods have been developed for this problem, from basically two different perspectives [1]. A first class of methods assumes prior knowledge to be available, and bayesian estimates are computed for the unknown signal. Such estimates are optimal under the 'true' model, but unfortunately, such a model is seldom if not at all available in practice. Then, non-parametric estimation methods often provide an appealing alternative.

Clearly, if we do not take into account any prior knowledge, the amount of information at hand is merely a function of the signal-to-noise ratio. On the other hand, in

Received: 29 June 1999
Received in revised form: 28 January 2000
Accepted: 28 January 2000

many applications one should cope with images where signal-to-noise ratio is rather poor. A good example is the detection of microcalcifications in X-ray mammography.

Mammograms are among the most difficult of radiological images to interpret, and microcalcification visual assessment, which is very important for the early diagnosis of breast cancer, turns 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 a non-stationary background. To shed light on the dimensions of the problem, the calcifications may vary in size from smaller than 0.1-mm to 5-mm in diameter. Figure 1 shows an example of mammogram with a selected Region Of Interest (ROI), including hardly visible microcalcifications, while Fig. 2 presents a magnified view of the same ROI. For dealing with such a challenging example, it is necessary to assume that objects of interest reveal some evident features at some specific scale, while being invisible at other scales.

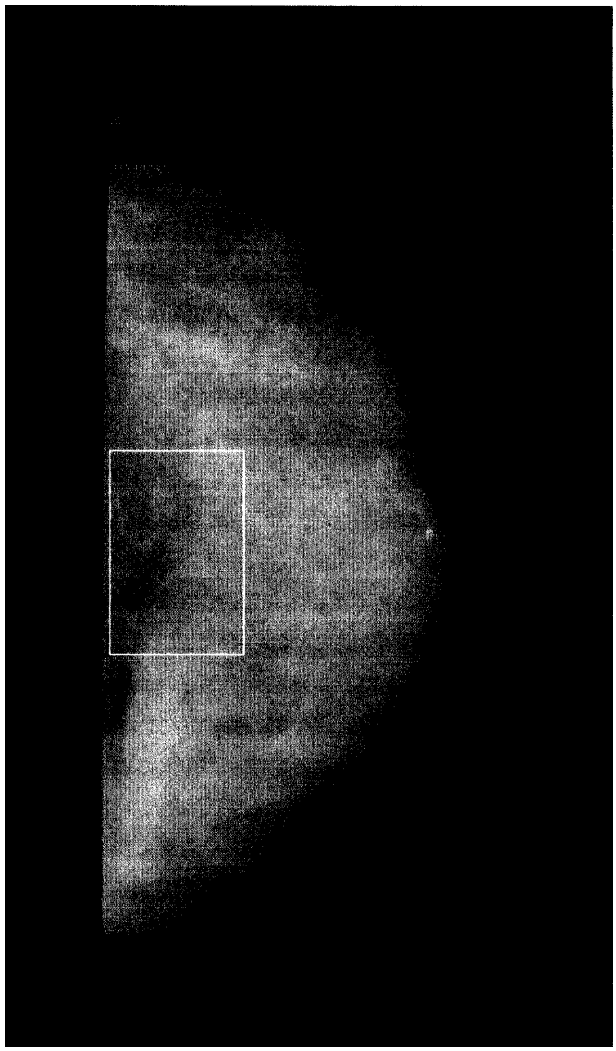

Fig. 1. An example of mammographic image where a Region Of Interest (ROI) has been selected that contains hardly visible microcalcifications.

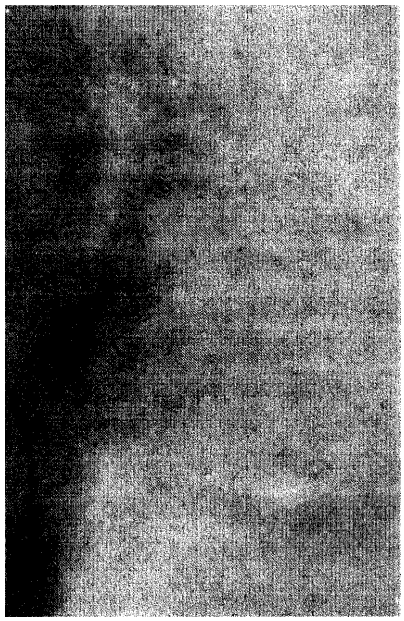

Fig. 2. A magnified view of the ROI selected in the mammogram of Fig. 1.

Wavelet theory is one important approach to scale analysis, due to the fact that the basis functions are well suited to the analysis of local scale phenomena; further, wavelets can also be used to model non-stationary processes. Fairly good results have been achieved in de-noising problems [3,4]. De-noising methods work as follows: first, they perform the wavelet transform of the observed data; secondly, they apply simple thresholding non-linearities to empirical wavelet coefficients; finally, signal estimates are computed by taking the inverse wavelet transform of the thresholded coefficients. However, many of the procedures derived to-date have been based upon the assumption of normal noise, and are therefore sensitive to noise distributions whose tails are heavier than the Gaussian distribution, as it is the case for images like mammograms. In addition, a reconstruction step is required, constraining the choice of the wavelet basis.

The rationale behind our research is that an ideal detector would gauge the image at multiple resolutions in order to pick out the objects of interest, and then thoroughly combine such resolution-dependent estimates according to some fusion procedure. Here, the multi-scale representation is obtained using a variation of the Undecimated Discrete Wavelet Transform (UDWT), which we call a weighted UDWT. In this set-up, we propose a method based on the maximisation of the information available at each resolution level of the weighted UDWT. Most importantly, we show that information maximisation can be achieved by maximising Renyi's information [5], and that this process eventually reduces to a simple thresholding procedure. The proposed method constitutes of the following steps: first, it builds a weighted UDWT of the digital image; then, wavelet coefficients are thresholded on the basis of scale dependent Renyi's information; finally, avoiding reconstruction, the

method combines results of scale-based estimation according to a majority vote strategy.

In the field of mammographic analysis, several target detection methods have been presented [6–8]. To overcome the limitations of these earlier methods, a number of publications have recently tried to exploit prior knowledge within the processing phase [9–11]. Highnam's model [9,10] uses image formation knowledge, like a scatter component, for mammographic enhancement and interesting tissue representation, like curvilinear structures of non-fat tissue. More significant is Karssemeijer's work [11], which incorporates prior knowledge via the random field model in a Bayesian framework; though effective, the method is very complex and computationally expensive, as noted by the author himself.

Multi-scale methods have also been adopted, either based on Laplacian pyramids [12] or, lately, on the wavelet transform [13–16]. Among these, Strickland's method [15] is appealing as regards both the approach and quality of experimental results achieved. It is worth noting that our work shares some common practical aspects with Strickland and Han [15], for instance, target detection is directly accomplished within the transformed domain, relying on the thresholding of wavelet coefficients, and also no prior knowledge is explicitly assumed. Therefore, results reported by Strickland and Han [15] are a suitable reference to compare with. In contrast, the overall rationale is quite different; most importantly, no general estimation problem is addressed by them: simply, the threshold is empirically chosen as a fixed percentile of the coefficient histogram, thus limiting the approach.

This paper is arranged as follows. Section 2 discusses the small target detection problem in terms of non-linear signal estimation in a multi-scale representation, and motivates the use of the wavelet transform obtained via a weighted undecimated scheme. Section 3 shows how, in such a set-up, the use of an information-theoretic approach turns into a simple but effective thresholding algorithm; in addition, the combination of information gathered at the different decomposition levels is introduced. Section 4 presents results achieved on mammographic images. To this aim, a public data set [11] has been used. The latter is specifically tailored for microcalcification analysis, and it allows performance characterisation using a Free Receiver Operating Characteristic (FROC) curve [11,15] as a figure of merit. Discussion is provided in Section 5, together with some concluding remarks.

2. THEORETICAL BACKGROUND

The reformulation of the small target detection problem as a non-linear signal estimation provides a framework to capture as much of the 'signal' as possible (the targets), while leaving out as much of the 'noise' as possible (background tissue plus actual noise). To this end, refer again to Eq. (1). If we recast the problem of estimating the true signal I_τ from the observed data I, a preliminary step is to find a suitable representation. The idea is to model the underlying signal and the background signal by expanding them using a suitable basis $\{\psi_s\}$: namely, $I_\tau(x,y) = \Sigma_s W_{s,\tau}\psi_s(x,y)$ and $I_\beta(x,y) = \Sigma_s W_{s,\beta}\psi_s(x,y)$. Then, using Eq. (1) and by linearity, the observed data can be expanded in the same fashion:

$$I(x,y) = \Sigma_s W_s \psi_s(x,y), \tag{2}$$

the coefficients given by

$$W_s = W_{s,\tau} + W_{s,\beta}. \tag{3}$$

In our case, the choice of basis functions is constrained by the fact we must deal with small targets of varying size and shape characterised by poor signal-to-noise ratio. A natural strategy is to think that the objects of interest convey information at some specific scale, while being invisible at other scales.

The analysis of images at multiple scales has proven insightful for image encoding, compression and feature extraction [17]. Gaussian and Laplacian pyramids have been used since the early 1970s to perform such multiresolution analyses of images [18], and have lately been reformulated in terms of scale-space methods employing diffusion processes and equations to evolve images at different scales [17,19,20]. On the other hand, the work on pyramids has led in recent times to the idea of wavelets and wavelet expansions (for a comprehensive review, see Mallat [21]).

Wavelets are basis functions generated from one single, zero mean function ψ, named 'mother wavelet', by dilations and translations. In two dimensions, a wavelet transform is computed with several wavelets $\{\psi_k\}_{1\le k\le K}$ that often have different spatial orientations, indexed by k. We denote $\psi_{k,s}(x,y) = \frac{1}{s}\psi_k\left(\frac{x}{s},\frac{y}{s}\right)$ and $\overline{\psi}_{k,s}(x,y) = \psi_{k,s}(-x,-y)$. The continuous wavelet transform of an image I at a scale s, in the direction k, is defined by the inner product

$$\begin{aligned} W_{k,s}I(u,v) &= \langle I(x,y), \psi_{k,s}(x-u, y-v)\rangle \\ &= \frac{1}{s}\iint I(x,y)\overline{\psi}\left(\frac{x-u}{s},\frac{y-u}{s}\right)dxdy, \end{aligned} \tag{4}$$

where the integral spans the domain of the image. This is equivalent to saying that $W_{k,s}I(u,v)$ is obtained by performing the convolution $I*\overline{\psi}_{k,s}(x,y)$. Since ψ has by definition a zero average, the convolution measures the variation of I in a neighbourhood of (u,v), whose size is proportional to scale s. Intuitively, this transformation can be seen as a mathematical microscope whose position and magnification are (u,v) and $1/s$, respectively, and whose optics is given by the choice of the specific wavelet ψ. It is worth noting that, by a suitable choice of ψ, the basis functions $\{\psi_{k,s}\}$ can provide an orthogonal basis. By means of such representation, it has been shown [21] that the image can be expanded as

$$I(x,y) = \sum_s \frac{1}{s}\left[\sum_{k=1}^{K} W_{k,s}I * \psi_{k,s}(x,y)\right]. \tag{5}$$

Clearly, Eq. (5) specifies the formula of Eq. (2) by using two-dimensional wavelet bases.

When dealing with digital images, we need to address an appropriate discrete wavelet transform. The continuous

transform of Eq. (4) is usually discretised as follows. A discrete scaling $s = s_0^l$ based on a dilation step $s_0 > 1$ is introduced; meanwhile, the translational parameters are chosen to depend both upon s and on translational steps u_0 and v_0, that is, $u = iu_0 s_0^l$, $v = jv_0 s_0^l$, where $(i,j,l) \in \mathbb{Z}^3$, and where $l \in [0,L]$ indexes the decomposition levels. The family of basis functions can thus be written as,

$$\frac{1}{s_0^l}\psi\left(\frac{x - iu_0 s_0^l}{s_0^l}, \frac{y - jv_0 s_0^l}{s_0^l}\right),$$

which substituted in Eq. (4) gives rise to the discrete wavelet transform $W_{k,s_0^l}I(iu_0,s_0^l, jv_0 s_0^l)$. In the following, for notational simplicity we denote $w_k^l(i,j) = W_{k,s_0^l}I(iu_0 s_0^l, jv_0 s_0^l)$ as the coefficients of the discrete wavelet transform. Depending upon the type of sampling, that is on the s_0, u_0 and v_0 values, the transform can be more or less redundant. In the case of critical sampling (occurring for $s_0 = 2$ and u_0, $v_0 = 1$), the wavelet family can be chosen as an orthonormal basis. A fast algorithm to compute an orthonormal transform, which is widely used for image processing applications, has been given by Mallat [22].

The problem this algorithm encounters is that it might not be convenient for pattern recognition purposes: first, it is not shift-invariant; secondly, image sub-bands result uncorrelated at the different scales. To overcome such drawbacks an UDWT can be used [23]. Such a scheme is characterised by octaves obtained by alternating a low-pass filter tapped by a band-pass filter (as in Mallat's algorithm), but inserting zeros between the elements of the filters in place of decimation. This approach seems to be much more effective for detection aims, counter-balancing a reasonable loss of computational efficiency. We extend it as follows.

By applying a UDWT to the discrete form of image I, namely $\{I(i,j)|i \in \{0,1,\ldots,M-1\}, j \in \{0,1,\ldots,N-1\}\}$, four coefficient planes $\{w_k^l(i,j)\}$ of size $M \times N$ are obtained at each scale or decomposition level. The plane indexed by $k = 0$ (low frequency sub-band) represents the smoothed image, and it does not convey useful directional information. For $k = 1,2,3$, the detail sub-bands are given, containing diagonal, horizontal and vertical directional information, respectively. Since we are not interested in image reconstruction via inverse transform, and provided that useful information mostly resides in detail sub-bands, a representation plane $\{w^l(i,j)|i \in \{0,1,\ldots,M-1\}, j \in \{0,1,\ldots,N-1\}\}$ is built at each level l, where each wavelet coefficient is the weighted linear combination

$$w^l(i,j) = \sum_{k=1}^{3} \alpha_k |w_k^l(i,j)|, \tag{6}$$

and $0 \leq \alpha_k \leq 1$. We define the collection of the L representation planes as the weighted UDWT.

Notice that, in this way, a more compact representation is obtained, as opposed to that achieved by the usual undecimated decomposition; namely, at each decomposition level there is only one coefficient plane whose domain is the same of the original image. In addition, we gain the possibility, by tuning the α_k weights, of adapting the representation to different kinds of images/applications. Since the transformation is applied on the discrete form of I, where I is modelled according to Eq. (1), each coefficient of the weighted representation plane can be written as

$$w^l(i,j) = w^l(i,j)_\tau + w^l(i,j)_\beta. \tag{7}$$

Summing up, we have precisely defined Eq. (3) in the framework of a computational procedure.

A key assumption we make here, and which we will use in the next section, is that for certain values of l, $w^l(i,j)_\beta = 0$. In other words, at certain scales the corresponding observation coefficients $w^l(i,j)$ represent the target signal rather than background and noise. This is reasonable in view of the spectral and structural differences between the signals I_τ and I_β across scales. It is worth noting that such an assumption shares some common aspects with multi-scale detection of image structures, as proposed in the scale-space literature. Actually, scale-space and wavelets are not unrelated, but as clearly stated by Lindeberg and ter Haar Romeny [24], the scale-space representation can be considered as a special case of continuous wavelet representation, where the scale-space axioms imply that the wavelet function must be selected as a derivative of the Gaussian kernel. In this respect, a first motivation to choose the wavelet representation as opposed to scale-space representation is to avoid constraining the method to an *a priori* chosen mathematical microscope, e.g. derivatives of Gaussians. A second motivation lies in providing a sufficiently simple, but effective, detection scheme. In the scale-space literature it has been recognised that linear processing with Gaussian kernels is not sufficient to handle the detection problem, and some nonlinear steps must be introduced in terms of differential geometric descriptors [20]. Interestingly, properties of wavelet basis endow the wavelet representation with a remarkable aptitude for estimating the signal with a smoothing that is locally adapted to the signal regularity. In fact, given an image modelled as in Eq. (1), it is well known that signal estimation can be achieved by applying a simple nonlinear thresholding filter to the wavelet transformed image [25]. It must be noted, however, that wavelet thresholding has been actually employed for true de-noising [3,4] rather than for pattern localisation purposes in the guise we are proposing. Further, although wavelets do provide an unconditional basis for a large class of signals, and do offer a simple framework for non-linear filtering, many of the procedures derived to date have been based upon limiting assumptions on noise and distributions. In the following section, we describe an information-theoretic approach to derive, in the context of the problem here examined, a thresholding filter under less restrictive assumptions.

3. COEFFICIENT THRESHOLDING VIA RENYI'S INFORMATION

The basic idea is the following. Consider the wavelet coefficients at the level l for simplicity, as a sequence of independent and identically distributed random variables characterised by a probability distribution function (pdf)

$p^l(x) = \Pr\{w^l(i,j) = x\}$. Clearly, taking into account Eq. (7), $p^l(x)$ can be considered as a mixture of two different distributions $\tau^l(x)$ and $\beta^l(x)$, representing at level l the distributions of wavelet coefficients relative to objects and background, respectively. The idea of discriminating targets from background can be restated in this framework as the process of maximising the distance between distributions τ^l and β^l in terms of some suitable parameter.

To this end, it is necessary to transform the coefficients $\{w^l(i,j)\}$ into distributions. First, each coefficient value is quantised according to the following linear quantisation:

$$\tilde{w}^l(i,j) = trunc\left[K\frac{w^l(i,j) - w^l_{\min}}{w^l_{\max} - w^l_{\min}}\right], \tag{8}$$

where $K = 2^k$ is the number of quantisation levels, k being the number of bits chosen for quantization. In the sequel, we will fix $k = 8$. The quantised coefficients $\tilde{w}^l(i,j)$ are thus in a suitable form to estimate probability distributions via histograms.

We denote by $n^l(x)$ the discrete histogram of the N coefficients $\tilde{w}^l(i,j)$. The histogram is normalised as $\hat{n}^l(x) = n^l(x)/N$, so that $\Sigma_x \hat{n}^l(x) = 1$. In general, for N sufficiently large and the histogram sufficiently regular, $\hat{n}^l(x)$ can be taken as an estimate of the probability $p^l(x) = \Pr\{\tilde{w}^l(i,j) = x\}$, namely an estimate of the pdf of the quantised wavelet coefficients $\tilde{w}^l(i,j)$. As discussed by different authors in the wavelet literature [21,26], this hypothesis holds for the histograms of wavelet coefficients of most 'natural' images. Notwithstanding, to assess how likely is such assumption for the case of mammographic images, we have experimentally determined on our data set (cf. Section 4), that the sample pdf $\hat{n}^l(x)$ is closely approximated by generalised Gaussian distribution, namely $a^l e^{-|b^l x|^{r^l}}$, where a^l, b^l, r^l characterise the pdf at level l. When the parameter r^l is computed with the chi-squared test, we obtain $r^l = 0.7$ [27], in agreement with results reported in Antonini et al [26]. For clarity's sake, in the remainder of this section we will simply refer to the distribution $p^l(x)$, but bearing in mind that we are indeed handling an empirical distribution.

At this point, we state the pdfs $\tau^l(x)$ and $\beta^l(x)$ to be defined on supports Ω^l_τ and Ω^l_β, respectively. Both supports depend upon the parameter t^l, namely $\Omega^l_\tau(t^l) = \{x : x^l_{\min} \le x \le t^l\}$ and $\Omega^l_\beta(t^l) = \{x : t^l < x \le x^l_{\max}\}$, where $x^l_{\min} = \min\{\tilde{w}(i,j)\}$ and $x^l_{\max} = \max\{\tilde{w}(i,j)\}$. Thus, we assume that $\tau^l(x) \to 0$ on $\Omega^l_\beta(t^l)$ and that $\beta^l(x) \to 0$ on $\Omega^l_\tau(t^l)$.

As stated from the start, to detect the objects of interest, one should determine t^l such that the distance between the two distributions is maximised. We say that such a distance represents the level l information, and we define it as the $\mathbf{L}^2$ distance $\int_{\Omega^l} (\tau^l(x) - \beta^l(x))^2 dx$ over the support $\Omega^l = \Omega^l_\tau(t^l) \cup \Omega^l_\beta(t^l)$. Clearly, since

$$\int_{\Omega^l} (\tau^l(x) - \beta^l(x))^2\, dx \tag{9}$$

$$= \int_{\Omega^l_\tau} \tau^l(x)^2\, dx + \int_{\Omega^l_\beta} \beta^l(x)^2\, dx - 2\int_{\Omega^l} \tau^l(x)\, \beta^l(x) dx,$$

the functional at the left-hand side of Eq. (9) is maximised iff $\int_{\Omega^l} \tau^l(x)\beta^l(x)dx$ is minimum. In fact, for $\int_{\Omega^l} \tau^l(x)\beta^l(x)dx = 0$ the two distributions can be considered 'orthogonal', in the sense that we exactly distinguish the spots from the background.

The left-hand side term of Eq. (9) denotes 'information'. In contrast, the term $\int_{\Omega^l} \tau^l(x)\beta^l(x)dx$ on the right-hand side of the same equation can be conceived as a sort of 'entropy', which destroys information while increasing. Define

$$H = \ln \int_{\Omega^l} \tau^l(x)\beta^l(x)dx \tag{10}$$

the entropy we want to minimise for maximising information. From Definition (10), by making use of the Cauchy–Schwartz inequality

$$\left(\int_{\Omega^l} \tau^l(x)\beta^l(x)dx\right)^2 \le \int_{\Omega^l} \tau^l(x)^2 dx \int_{\Omega^l} \beta^l(x)^2\, dx$$

and taking logarithms, entropy H can be upper bounded as

$$2H \le \ln \int_{\Omega^l_\tau} \tau^l(x)^2\, dx + \ln \int_{\Omega^l_\beta} \beta^l(x)^2\, dx. \tag{11}$$

Let $I_{\Omega_\alpha}(r) = \frac{1}{1-r} \ln\left(\int_{\Omega_\alpha} \alpha(x)^r dx\right)$ be the rth order Renyi's information of the distribution $\alpha(x)$ over the support Ω_α [5]. Then, from inequality (11), it follows that,

$$H < \frac{1}{2}\ln \int_{\Omega^l_\tau} \tau^l(x)^2 dx + \frac{1}{2}\ln \int_{\Omega^l_\beta} \beta^l(x)^2 dx, \textit{ i.e.}$$

$$H < -(I_{\Omega^l_\tau}(2) + I_{\Omega^l_\beta}(2)). \tag{12}$$

Thus, we have shown that the minimisation of entropy H can be achieved by maximising the second order Renyi's information of $\tau^l(x)$ and $\beta^l(x)$ with respect to t^l. Formally:

$$t^l = \text{Arg max} \lim_{r \to 2} \left\{\frac{1}{1-r} \ln\left(\int_{\Omega^l_\tau} \tau^l(x)^r\, dx\right) + \frac{1}{1-r}\ln\left(\int_{\Omega^l_\beta} \beta^l(x)^r\, dx\right)\right\}. \tag{13}$$

To summarise, this result shows that a difficult problem

such as the minimisation of the functional $\int_{\Omega^l} \tau^l(x)\beta^l(x)dx$, which would in principle require the modelling of a mixture of distributions with unknown parameters, can be turned into an effective procedure based on the maximisation of Renyi's information at the resolution level l. The parameter t^l that is obtained can then be used as a threshold for estimating the target at scale l.

A brief comment is due on how the performance of the threshold computed through Eq. (13) compares with thresholding proposed by classical de-noising methods. It is possible to show [27] that, by using Eq. (13), we obtain an estimator whose performance is comparable to that proposed by Donoho and Johnstone [28]. Most importantly, in our case the estimation error depends only upon the 'characteristic length' of the target distribution at a given scale l, whilst in classical thresholding it depends upon the number of samples/pixels in the image. In other words, the error is related to a lower bound on detection capabilities that depends upon the nature of targets to be detected and beyond the choice of a specific wavelet basis. In fact, in the case of images displaying small targets embedded within inhomogeneous background, most detection errors could be the result of the problem that in specific regions, such targets could be sparse and smaller than background structures. This is indeed the case of mammograms, where microcalcifications may be shadowed by the high frequency texture of parenchymal tissue structures.

In this set-up, the detection of microcalcifications is achieved in two steps. First, at each resolution l, a labelling is performed by constructing a saliency map in the form of a binary map M^l, where each connected set of non-zero locations constitutes an hypothesis target detected in the image. This is obtained by employing t^l as a threshold, whereby each site of the weighted representation plane is labelled $M^l(i,j) = 1$ if $\tilde{w}^l(i,j) > t^l$ and $M^l(i,j) = 0$, otherwise.

Secondly, the different saliency maps are combined according to a majority vote rule. The aim of this step is to provide the final detection map $M(i,j)$. Majority voting choice is motivated as follows.

From a general point of view, it is possible to describe in a bayesian framework the problem of classifying the information gathered at the different scales of a multiresolution representation. Assume that at each point (i,j) of the image domain, a vector of measurements is given, say $\mathbf{x} = x_1(i,j),\ldots,x_L(i,j)$. According to bayesian theory, classification is in our case equivalent to label each point as either τ (target) or β (background), provided that the *a posteriori* probability of that interpretation is maximum, i.e. $P(M(i,j) = \lambda|\mathbf{x}) = \max_{\lambda=\tau,\beta} P(M(i,j) = \lambda|\mathbf{x})$. The computation of the *a posteriori* probability functions would depend upon the knowledge of high order statistics described in terms of joint probability density functions $P(\mathbf{x}|M(i,j) = \lambda)$, which would be difficult to infer. However, it is possible to simplify the above rule and express it in terms of decision support computations performed by L classifiers, one for each resolution level. Each classifier only exploits the information conveyed by vector $\mathbf{x}$. Kittler et al [29] have shown that it may be true for some applications that these measurements are conditionally statistically independent. In our context, we can assume each labelled representation plane $M^l(i,j)$ as a 'hard' estimate of the *a posteriori* probability function for the resolution level l. When the available observational discriminatory information is highly ambiguous, for example due to a high level of noise, it may be appropriate to assume that the *a posteriori* probabilities will not deviate dramatically from the prior probabilities [29]. In this case, the computation of maximum *a posteriori* probability simplifies to the well-known majority vote rule.

4. APPLICATION TO MAMMOGRAPHIC IMAGES

4.1. Data Set

For evaluation purposes, experiments have been performed on the Nijmegen data set. The latter was chosen because it represents, to-date, a widely used public data set for specifically evaluating microcalcification detection performance without the need to account for other kinds of breast abnormalities. It is the same one used by Karssemeijer [11] and other recent papers (for a comprehensive review see [30,31]). The 40 images of this data set have been made available by the Department of Radiology of the University Hospital, Nijmegen, and can be obtained via anonymous ftp[1]. All images are in raw format, of size 2048×2048 pixels, 12 bits per pixel of grey level information. The images were digitised from film using an Eikonix 1412 12-bit CCD camera. A sampling aperture of 0.05 mm in diameter was used, with a 0.1 mm sampling distance. The images were corrected for inhomogeneity of the light source (Gordon plannar 1417). A fixed calibration of the CCD camera was adopted. The optical density of 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 ground truths and stored in separate files. Performance can then be evaluated in terms of FROC curves based on true/false cluster detection, as explained in the sequel (Section 4.3).

4.2. Experimental Set-up

First, the setting of the method was considered. At this stage we expressed the effectiveness of different solutions in terms of the number of individual calcification particles truly detected for each image against the number of false detections; 10 images of the data set were used. An individual microcalcification is considered as truly detected in the processed image, if it is also present in the accompanying truth image. Two major aspects have been taken into

[1] figment.csee.usf.edu in directory pub/mammograms/nijmegen-images

account: (1) the number of decomposition levels of the weighted UDWT; and (2) the selection of the wavelet basis to employ for decomposition. The choice of the number of decomposition levels is clearly a trade-off between the size of the targets to detect and the presence of noise. This parameter is related to the kind of images handled, rather than to the basis adopted. In the case of mammographic images, it is quite simple to establish, by visually inspecting the information present within the directional sub-bands at different scales, that an acceptable number of decomposition levels is $L = 6$. Running the detection module, the conservative value $L = 4$ was chosen; in fact, for higher values of L, we registered a fall of performance, which is likely to occur due to biasing of the majority voting process. Further, as expected, this behaviour was common to all bases.

As regards the second point – selection of the decomposition basis – standard wavelet bases have been employed: Burt-Adelson, Battle-Le Marie, B-spline I, B-spline II, Daubechies 4, Daubechies 6, Daubechies 8, Daubechies 10 [26]. In this phase we used Eq. (6), adopting uniform weights (i.e., $\alpha_1 = \alpha_2 = \alpha_3$) to build the representation planes $\{w^l(i,j)\}$ of the weighted UDWT. This can be justified by experimentally noting that when the different bases are compared with one another on the same image, their behaviour is reasonably independent with respect to weighting variations. The best performance has been achieved on an biorthogonal B-spline basis [26]. Summarising the results, the method, by using such a basis along four decomposition levels, truly detects individual particles, and most importantly, limits the detection of individual false calcifications to an average maximum of about 10 per image.

4.3. Performance Evaluation

The detection of individual particles of microcalcification is to be considered clinically significant if they appear in *clusters*, and there is a wide agreement that one of the most salient properties for discriminating benign from malignant clusters is the number of calcification particles in close proximity. Here, we adopted the standard cluster definition [11]: a cluster is observed if more than three microcalcifications are localised inside a circular region of radius 0.5 cm, marked around each detected microcalcification. In general, detection performance can be evaluated by counting the true positive clusters and false positive clusters per mammogram. The cluster is true positive if marked as such in the accompanying truth image; false positive, otherwise. To be precise, according to the literature [11,15], the following figures of merit have been adopted: the true positive fraction TPF = (#true positive clusters detected)/(#true positive clusters to be detected), and the false rating FPC = (#false positive clusters detected)/(#images).

Using the experimental set-up described in the previous section, the performance of the method was assessed by computing an FROC curve. This curve can be obtained by plotting the TPF and FPC ratios along the variation of a suitable parameter, which should be chosen as a critical control parameter of the method.

Here, the parameters of interest are the α_k values. In fact, when using the weighted UDWT, a variation of the α_k implies a different balance of the information conveyed by the three directional sub-bands (refer to Eq. (6)). A first reasonable assumption is that the strongest singularities (like background texture ones) are most likely to appear within the diagonal

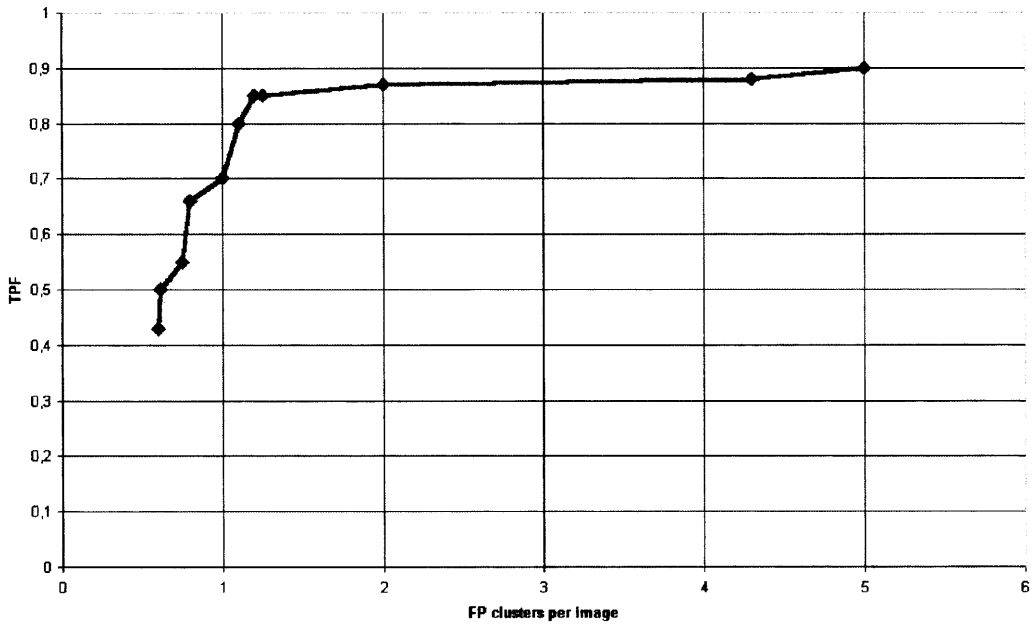

Fig. 3. FROC curve: cluster detection performance by varying the parameter α_1.

detail band ($k = 1$), rather than within horizontal and vertical detail sub-bands ($k = 2,3$). A second assumption is that horizontal and vertical detail sub-bands should be equally weighted (i.e. $\alpha_2 = \alpha_3$), reflecting the horizontal/vertical symmetry of each calcification. Summing up, by imposing that $\alpha_1 + \alpha_2 + \alpha_3 = 1$, under the assumption $\alpha_2 = \alpha_3$, it is possible to observe the TPF/FPC behaviour versus the α_1 variation in the [0, 1] range. The FROC curve obtained over the 40 images of the dataset, by using optimal settings, is plotted in Fig. 3. The parameter α_1 is not apparent from the graph; yet the curve provides the information on the various trade-offs between TPF and FPC that may be obtained by selecting the parameter. We achieved, as a best result, 66% of TPF at the FPC rate of 0.7 with $\alpha_1 = 0.56$.

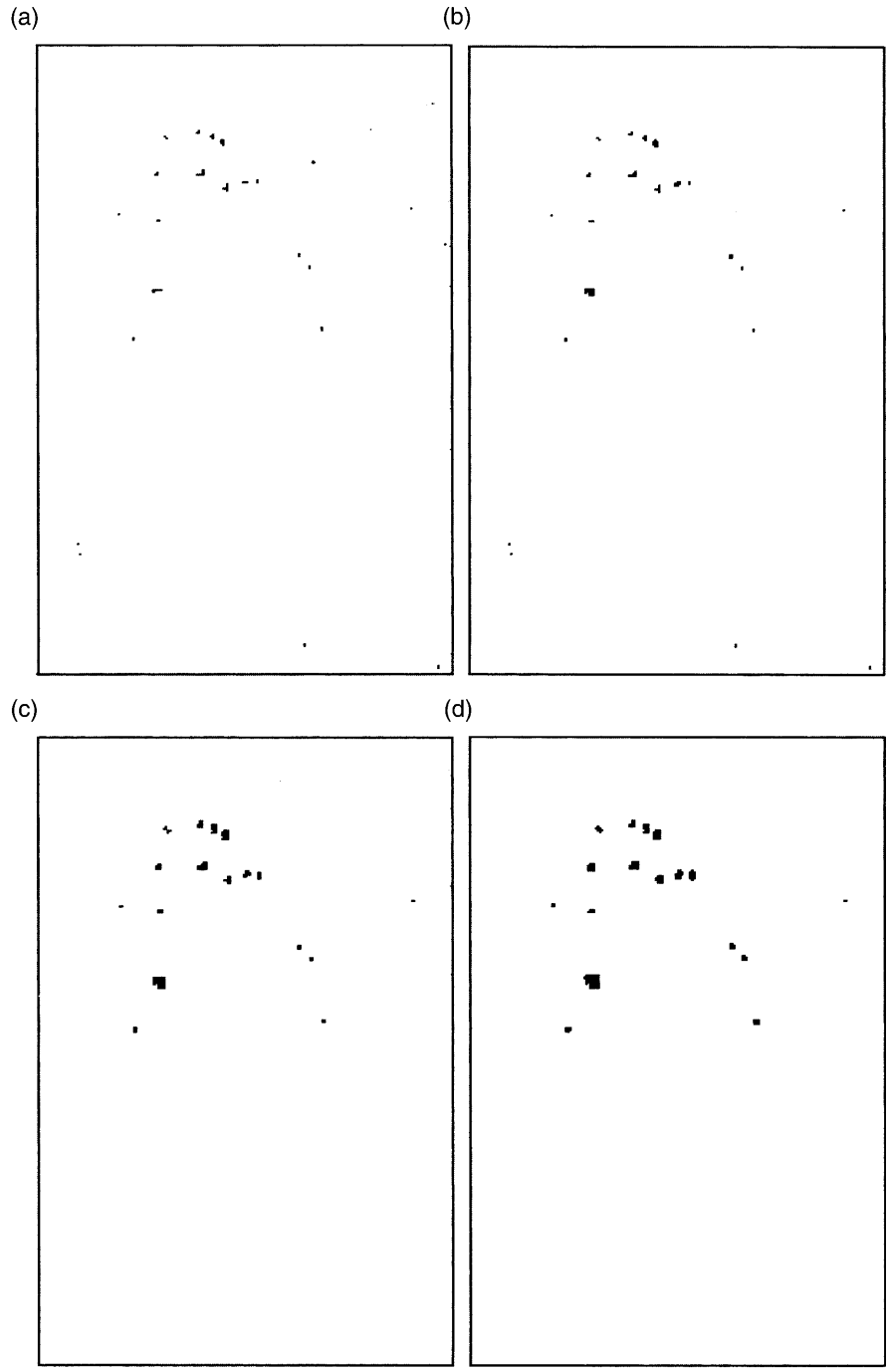

Fig. 4. Level dependent microcalcification detection on the ROI of Fig. 2. (a)–(d) show the saliency maps at decomposition levels 1, 2, 3 and 4, respectively. Each map includes the candidate individual targets proposed at that level. Detection is performed by B-spline II wavelet basis and fixing $\alpha_1 = 0.56$.

Comparing with Strickland and Han [15], 55% of TPF at the FPC rate of 0.7 is reported on the same data set. Thus, the proposed method gives an improvement of about 10% over Strickland's. This result is partly due to the optimal weighting among the wavelet sub-bands, although a major role is played by the use of the adaptive threshold t^l. As noted (cf. Section 1), in Strickland and Han [15] the threshold is empirically chosen as a fixed percentile of the coefficient histogram. Such a simple method, however, may cause drawbacks depending on the kind of distribution addressed. In general, when non-adaptive thresholding is employed, one usually relies on the bimodality of the empirical distribution. However, as previously discussed the wavelet coefficient pdf is shaped as a generalised Gaussian. In consequence, the bimodal assumption must be abandoned in this specific case, and a more sophisticated thresholding procedure is to be considered. For the sake of completeness, we recall that Karssemeijer [11] detects about 82% at a 0.7 FPC rate, which is still better than the result achieved here. However, this was to be expected, because in Karssemeijer [11], differently from our approach and that of Strickland and Han [15], prior knowledge is introduced to explicitly exploit specific features of mammogram structures.

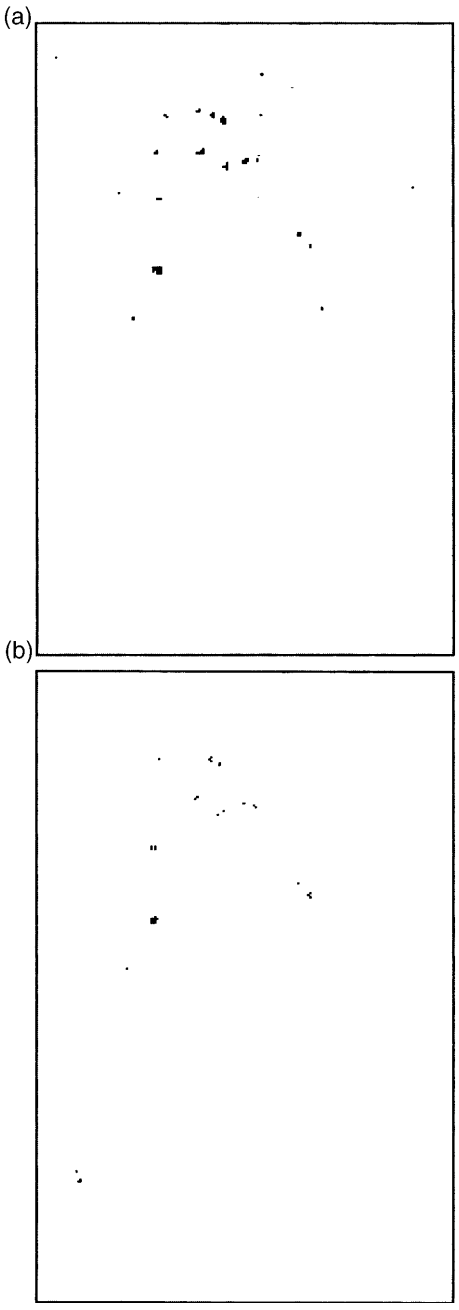

Fig. 5. (a) Final detection map obtained by applying majority voting on saliency maps shown in Fig. 4, (b) ground truth of the same ROI.

Eventually, to summarise the method at a glance, we apply the detection process to the ROI image of Fig. 2. Figures 4(a)–(d), show the saliency maps at decomposition levels 1, 2, 3 and 4, respectively; each map includes the candidate individual targets proposed at that level. Figure 5(a) presents the individual microcalcifications detected after outcomes obtained at levels 1, 2, 3 and 4 have been evaluated via majority voting; this final result may be compared with Fig. 5(b), which shows the ground truth of the ROI. To conclude, Fig. 6 displays the same individual calcifications, prompted on the original mammogram of Fig. 1, in the form of boxes, namely rectangles of minimum area surrounding each target.

5. DISCUSSION AND CONCLUSION

We have proposed a multi-scale method for the detection of small targets in noisy background. The multi-scale representation is designed using a weighted UDWT. In essence, the method is based on the maximisation of information available at each resolution level of such a representation, and we have shown that this goal can be achieved by maximising Renyi's information. Using Renyi's information, an adaptive threshold is determined, which is employed for discriminating between wavelet coefficients representing targets and those representing background noise. Eventually, resolution dependent estimates are combined according to a majority vote rule, while avoiding reconstruction.

Some remarks on major points introduced in this work may be useful here. The weighted UDWT we propose provides a more compact representation as opposed to that achieved by ordinary undecimated decomposition; meanwhile, we gain the possibility, by tuning the α_k weights, of adapting the representation to different kinds of images and applications. The introduction of such weighted representation is admissible, since we are not concerned with the reconstruction problem, the latter being avoided by the method.

We discuss the problem of target detection within an information-theoretic framework, which endows a general solution; on the other hand, what gives this approach a practical interest is that it allows us to set up a procedure to determine an adaptive threshold suitable for the non-linear filtering of targets hidden by background noise.

Information collected at the different scales is combined

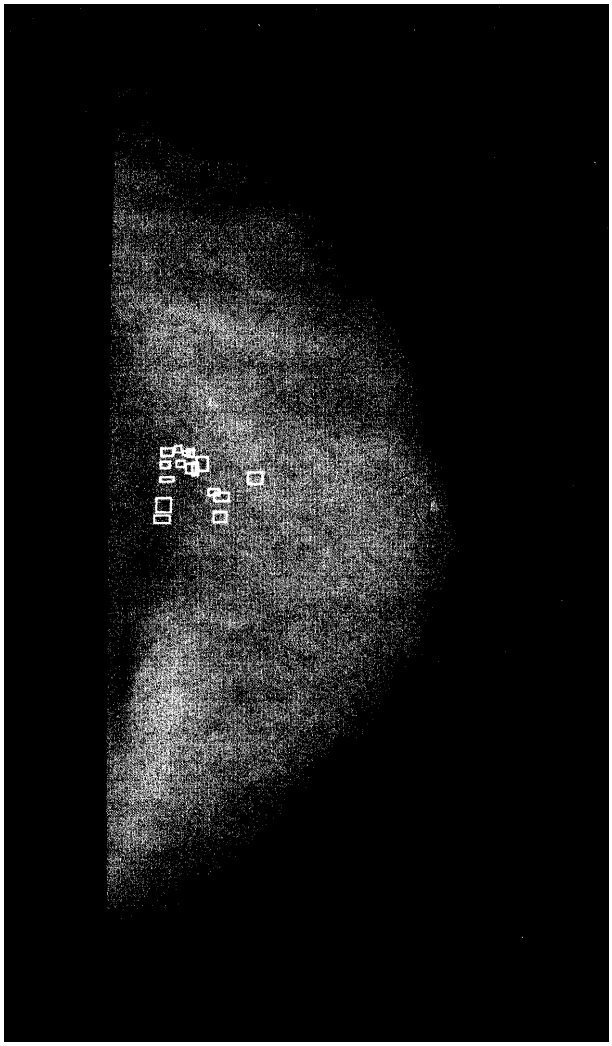

Fig. 6. The ROI calcifications, as detected in Fig. 5, displayed on the original mammogram of Fig. 1. Each target is marked by a surrounding box.

to obtain the final detection. To this end, a simple but effective combination scheme has been exploited. The scheme proposed shares some common aspects, in terms of nonlinear filtering coupled with a linear multi-scale representation, with other approaches like blob detection in scale-space [20]. Those methods achieve structure detection through nonlinear combination of Gaussian derivatives obtained at multiple scales, with the objective of evaluating features such as spatial extent, contrast and structure lifetime in scale-space. In our approach, thresholding naturally incorporates spatial extent and contrast, while lifetime is accounted for by the majority vote rule. Meanwhile, we do not have to restrict our representation space to Gaussian kernels.

The method has been applied to the difficult problem of microcalcification detection in mammographic images, obtaining interesting results. Experiments have been performed on a standard, Web-available, mammogram data set, namely the Nijmegen data set, which is considered a standard data set in this research realm.

To conclude, we have not limited our work to research experiments, but turned to a real clinical setting. Actually, much of the work described here has been carried out in the framework of a joint research project together with the Medical Physics Department at Sant'Orsola University Hospital, Bologna, Italy, and with Integris Biomedical Research Labs, Italy. In this project, a system for computer aided diagnosis has been designed with the goal of providing radiologists a toolbox of methods for the enhancement of mammograms and the early diagnosis of breast cancer [27]. A first prototype of the system has been written in the C++ language under the Windows NT Operating System, and developed in the framework of an innovative object-oriented radiological information system [32]. The detection method presented here has been conceived as a module of such a system, to serve either the visualisation of the clusters of microcalcifications to the radiologist in the course of a diagnostic session, or clinical follow-up of the patient.

Clinical testing of the system is in progress at Sant'Orsola Hospital. Images used in the testing are digitising mammograms at 300 dpi, at a resolution of 12 bit/pixel, successively windowed to 8 bit/pixel, obtained using a VIDAR VXR-12 scanner. Due to the lower resolution, such images are more challenging than those included in the Nijmegen data set. However, preliminary results collected adopting the same experimental settings previously described show a decrease in detection performance of less than 10% with respect to results reported here, giving evidence of a promising clinical relevance of the detection module.

Acknowledgement

The authors are grateful to the referees for their enlightening and valuable comments which have greatly improved the quality and clarity of an earlier version of this paper.

References

1. Kailath T, Poor HV. Detection of stochastic processes. IEEE Trans Information Theory 1998; 44:2230–2259
2. Mumford D. Pattern theory: a unifying perspective. In: Knill D, Richards W (eds). Perception as Bayesian Inference. Cambridge University Press, 1996: 25–62
3. Johnstone IM, Silverman BW. Wavelet threshold estimators for data with correlated noise. Journal of the Royal Statistical Society-B 1997; 59:100–115
4. Donoho DL. De-noising by soft-thresholding. IEEE Trans on Information Theory 1995; 41:613–627
5. Cover T, Thomas J. Elements of Information Theory. Wiley, New York, 1991
6. Davies DH, Dance DR. Automatic computer detection of clustered microcalcifications in digital mammograms. Physics in Medicine and Biology 1990; 35(8):1111–1118
7. Dengler J, Behrens S, Desaga JF. Segmentation of microcalcifications in mammograms. IEEE Trans Medical Imaging 1993; 12(4):634–642
8. Shen L, Rangayyan RM, Desautels JEL. Application of shape

analysis to mammographic calcifications. IEEE Trans Medical Imaging 1993; 13(2):263–274

9. Highnam RP, Brady JM, Shepstone BJ. Computing the scatter component of mammographic images. IEEE Trans Medical Imaging 1994; 13(2):301–313
10. Highnam RP, Brady JM, Shepstone BJ. A representation for mammographic image processing. Medical Image Analysis 1997; 1(1):1–18
11. Karssemeijer N. Adaptive noise equalization and recognition of microcalcifications clusters in mammograms. International Journal of Pattern Recognition and Artificial Intelligence 1993; 7(6):1357–1377
12. Brzakovic D, Neskovic M. Mammogram screening using multiresolution-based image segmentation. International Journal of Pattern Recognition and Artificial Intelligence 1993; 7(6):1437–1460
13. Dinten J, Darbaux M, Nicolas E. A global approach for localization and characterization of microcalcifications in mammograms. In: Doi K, Giger M, Nishikawa R, Schmidt R (eds). Digital Mammography. Elsevier, 1996: 235–238
14. Chitre Y, Dhawan AP. Adaptive wavelet analysis and classification of mammographic microcalcification. In: Doi K, Giger M, Nishikawa R, Schmidt R (eds). Digital Mammography. Elsevier, 1996: 323–326
15. Strickland RN, Han HI. Wavelet transform for detecting microcalcifications in mammograms. IEEE Trans Medical Imaging 1996; 15(2):218–229
16. Wang TC, Karayiannis NB, Detection of microcalcifications in digital mammograms using wavelets. IEEE Trans Medical Imaging 1998; 17(4):498–509
17. Witkin A. Scale space filtering. Proceedings of the International Joint Conference on Artificial Intelligence, Karlsruhe, Germany 1983:1019–1023
18. Burt P, Adelson E. The laplacian pyramid as a compact image code. IEEE Trans Communications 1983; 9:532–540
19. Koenderink J. The structure of images. Biological Cybernetics 1984; 50:363–370
20. Lindeberg T. Scale-space theory in computer vision. His Publisher, Erewhon, NC, 1999
21. Mallat S. A wavelet tour of signal processing. Academic Press, San Diego CA, 1998
22. Mallat S. Multifrequency channel decomposition of images and wavelet models. IEEE Trans Acoustics, Speech and Signal Processing 1989; 37(12):2091–2110
23. Shensa M. The discrete wavelet transform: Wedding the à trous and Mallat algorithm. IEEE Trans Signal Processing 1992; 40(10):2464–2482
24. Lindeberg T, ter Haar Romeny B. Linear scale-space 1: Basic theory. In: ter Haar Romeny B. (ed). Geometry-Driven Diffusion in Computer Vision. Kluwer Academic 1994: 1–38
25. Donoho DL. Unconditional bases are optimal bases for data compression and for statistical estimation. Applied and Computational Harmonic Analysis 1993; 1:101–115
26. Antonini M, Barlaud M, Mathieu P, Daubechies I. Image coding using wavelet transform. IEEE Trans Image Processing 1992; 1: 205–220
27. Picariello A. A system for Computer Aided Diagnosis of digital mammography: Model, design and implementation. PhD thesis, DIS-Università di Napoli, Naples, Italy, February 1998
28. Donoho D, Johnstone I. Ideal spatial adaptation by wavelet shrinkage. Biometrika 1994; 81(3):425–455
29. Kittler J, Hatef M, Duin RBW, Matas J. On combining classifiers. IEEE Trans Pattern Analysis and Machine Intelligence 1998; 20(3):226–239
30. Doi K, Giger M, Nishikawa R, Schmidt R (eds). Digital Mammography, Elsevier, 1996
31. Karssemeijer N, Thijssen M, .Hendreks J, Hardbound LVE (eds). Digital Mammography Kluwer Academic 1998
32. Boccignone G, Chianese A, De Santo M, Picariello A, Pierotti L, Prigione F. A system for computer aided mammography. In: Lemke H, Vannier M, Inamura K (eds). Computer Assisted Radiology and Surgery, Elsevier Science 1997: 197–202

Giuseppe Boccignone is an Assistant Professor and 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 chief researcher of the Computer Vision Lab at CRIAI, Naples, Italy. From 1992 to 1994, he has 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, ACM, and IAPR.

Angelo Chianese is Associate Professor of Computer Science at the Faculty of Engineering of the University of Napoli, Italy. He received the Laurea degree in electronic engineering from the University of Napoli, in 1980. From 1984 to 1992, he was assistant professor and researcher in the Computer Vision Lab of the same university. Since 1984 he has been with the Dipartimento di Informatica e Sistemistica of the University of Napoli. His current research interests include computer networks, character recognition, medical image processing, image databases, distance education and distance learning. Angelo Chianese is a member of AICA.

Antonio Picariello is Assistant Professor and Researcher at the Faculty of Engineering of the University of Napoli, Italy. He received the Laurea degree in electronic engineering in 1991 and a PhD in Computer and Electronic Engineering in 1998, both from the University of Napoli. 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, cursive script recognition, image databases. Antonio Picariello is a member of IAPR.

Correspondence and offprint requests to: G. Boccignone, DIIIE and INFM, Università di Salerno, via Ponte Melillo, 1, Fisciano, Italy. E-mail: boccig@diiie.unisa.it