Digital mammography: A weak continuity texture representation for detection of microcalcifications

Digital mammography: a weak continuity texture representation for detection of microcalcifications

Caputo, Barbara, Gigante, Giovanni

Barbara Caputo, Giovanni E. Gigante, "Digital mammography: a weak continuity texture representation for detection of microcalcifications," Proc. SPIE 4322, Medical Imaging 2001: Image Processing, (3 July 2001); doi: 10.1117/12.431057

SPIE.

Event: Medical Imaging 2001, 2001, San Diego, CA, United States

Digital mammography: a weak continuity texture representation for detection of microcalcifications

Barbara Caputo[a] and Giovanni E. Gigante[b]

[a]University of Erlangen, Chair of Pattern Recognition,
Erlangen, Germany
[b] University of Rome "La Sapienza", Physics Department - CISB,
Rome, Italy

ABSTRACT

This paper proposes a Weak Continuity Texture Representation (WCTR) method for detecting clustered microcalcifications in digitized mammograms. This technique is compared with other texture-analysis methods (Co-occurrence Matrices, Gabor Energy Mask, and Wavelet Filter). The WCTR is a new method for texture representation, based on the characterization of textures using statistics of their coarseness. From edge maps, obtained by a weak membrane at different noise levels, density values are computed which are representative of the texture coarseness. We chose six different noise levels; each texture class is then represented by six edge-density values. Textural features extracted using the four methods are used to discriminate between positive ROI's containing clustered microcalcifications and negative ROI's containing normal tissue; a three-layer backpropagation neural network is employed as a classifier. A ROC analysis is used to evaluate the classification performance. From an original database of 151 ROIs two different combinations of training and testing sets are used: 50/70 training cases and 101/81 testing cases. The best performance is obtained with the WCTR method in both cases (92% and 93% respectively). These results show the effectiveness of WCTR for the detection of microcalcifications in mammographic images.

Keywords: mammographic images, microcalcifications, computer-aided diagnosis, neural networks, texture analysis , weak continuity.

1. INTRODUCTION

Mammography has been used in clinical practice since 1927 in the diagnosis of breast abnormalities, reaching in the sixties a high degree of accuracy in the discrimination between benignant and malignant diseases.[2] Screening projects in the sixties and seventies showed as a strong decrease of breast cancer risk for women over 50 can be obtained with a systematical usage of mammography.[2] Actually, screen-film mammography associated with clinical breast examination and breast self-examination is widely recognized as the only effective imaging modality for early detection of breast cancer in women,[1] .[2] Among the early indicator of breast cancer, microcalcifications are one of the primary signs.[2] Microcalcifications are tiny, granule like deposits of calcium; several studies demonstrate that the presence of clustered microcalcifications in X-ray mammograms is a basical marker for the early detection of breast cancer, especially for individual microcalcifications with diameters up to about 0.7 mm and with an average diameter of 0.3 mm.[2] However, the interpretation of X-ray mammograms is very difficult because of the small differences in the image densities of various breast tissues, particularly for dense breast. The interpretation of mammograms by radiologists is performed by a visual examination of films for the presence of abnormalities that indicate cancerous changes. Unfortunately, for difficult-to-diagnose cases, the detection (or not detection) of supposed microcalcifications can lead to a false diagnosis, with the risk of unnecessary surgical intervention. Therefore, computerized analysis to help decision making, for biopsy recommendation and diagnosis of breast cancer, might be of significant value to improve the true-positive rate of breast cancer detection. In the actual interpretation of mammographic microcalcifications, the grey-level values defining local structures in the microcalcification clusters play a significant role.[2] It has been

[a]caputo@informatik.uni-erlangen.de; phone +49 9131 8527824; fac +49 9131 303811; http://www5.informatik.uni-erlangen.de; University of Erlangen, Chair of Pattern Recognition, Martensstrasse 3, 91058 Erlangen, Germany. [b]gigante@caspur.it, university of Rome "La Sapienza", Physics Department -CISB, Piazza A. Moro 3, 00161 Rome, Italy. This research work was developed while B. Caputo was at the University of Rome "La Sapienza", as partial fulfilment of her M.Sc.; the * indicates the corresponding author.

Medical Imaging 2001: Image Processing, Milan Sonka, Kenneth M. Hanson, Editors,
Proceedings of SPIE Vol. 4322 (2001) © 2001 SPIE · 1605-7422/01/$15.00

demonstrated in clinical studies described in,[2] that the grouping of microcalcification regions, in order to define the shape of the cluster, is highly dependent on the gray-level based-structure and texture of the image. Indeed, texture information plays an important role in image analysis and understanding, with potential applications in remote sensing, quality control, and medical diagnosis. Texture is one of the important characteristics used in identifying an object or a region of interest (ROI) in an image.[3] Moreover, the introduction of digital mammographic systems is pushing on the development of reliable methods for semiautomatic screening of mammograms for early detection of breast cancer.

Computerized mammographic image analysis can be divided into three steps: enhancement of mammographic features; detection and localization of suspicious areas and classification. Most of the techniques used in the computerized analysis of mammographic microcalcifications first segment the digitized gray-level image into binary regions representing the microcalcifications,[4] [5]; feature analysis is then performed on such binary images. Since mammographic images may often be poor in contrast, lacking in the definition of microcalcification region, the segmentation process for extraction of such regions is not reliable and accurate. Therefore, the shape features of the binary segmented individual microcalcifications may not be reliable, particularly for difficult-to-diagnose cases. It has been demostrated in clinical studies described in[2] that the grouping of microcalcification regions, in order to define the shape of the cluster, is highly dependent on the gray level based structure and texture of the image. For those reasons in the last years several researchers has investigated the effectiveness of texture analysis technique for microcalcification detection and analysis,[6] .[7]

In this paper we propose a *Weak Continuity Texture Representation* (WCTR) for texture classification. The WCTR is a new method for texture representation, proposed by one of the authors that has shown to be effective in representation of real textures for classification purposes.[8] The WCTR method is based on the characterization of textures using statistics of their coarseness.[3] From edges maps, obtained by a weak membrane at different noise level, density values are computed which are representative of the texture coarseness. Here we chose six different noise levels; each texture class is then represented by six edge-density values. In order to evaluate the WCTR method performance we made a comparative study with three very well known texture analysis methods, which were successfully employed for microcalcification detection,[9] ,[6] [7]: Co-occurrence Matrices (CM),[10] Gabor Energy Mask (GEM)[11] and Wavelet Filter (WF).[12] Textural features extracted using these four methods are used to classify Region Of Interests (ROIs) into positive ROIs containing clustered microcalcifications and negative ROIs containing normal tissue. To evaluate the classification efficacies of these texture-analysis methods, a three-layer backpropagation neural network was employed as a classifier. A Receiver Operating Characteristic (ROC) analysis was used to evaluate the classification performances.

The paper is organized as follows: Section 2 describes the employed texture analysis techniques, with a particular emphasis on WCTR. The experimental results are presented in Section 3; the three-layer backpropagation neural network used as classifier is also described in Section 3. Finally, conclusions are given in Section 4.

2. TEXTURE ANALYSIS

2.1. Weak Continuity Texture Representation

The WCTR is a model based on the characterization of textures using statistic of their coarseness,[3] .[13] It is based on the properties of the *Weak Continuity* (WC) approach,[14] which is a very effective tool for edge detection: the WC approach allows a multi-scale edge detection, and furthermore its non linearity leads to a selective smoothing and thus to a high robustness to the noise.

The basic idea is to represent the input signal by a *weak string*, that is an elastic string satisfying the weak continuity constraints. For two dimensional signals, such as images, the string will be substituted by an elastic membrane: the *weak membrane*. The general model consists on the minimization of a given functional, that for the weak membrane will be:

$$E = D + T + F, \tag{1}$$

with

$$D = \int_{\Omega} (u(x,y) - f(x,y))^2 dxdy, \tag{2}$$

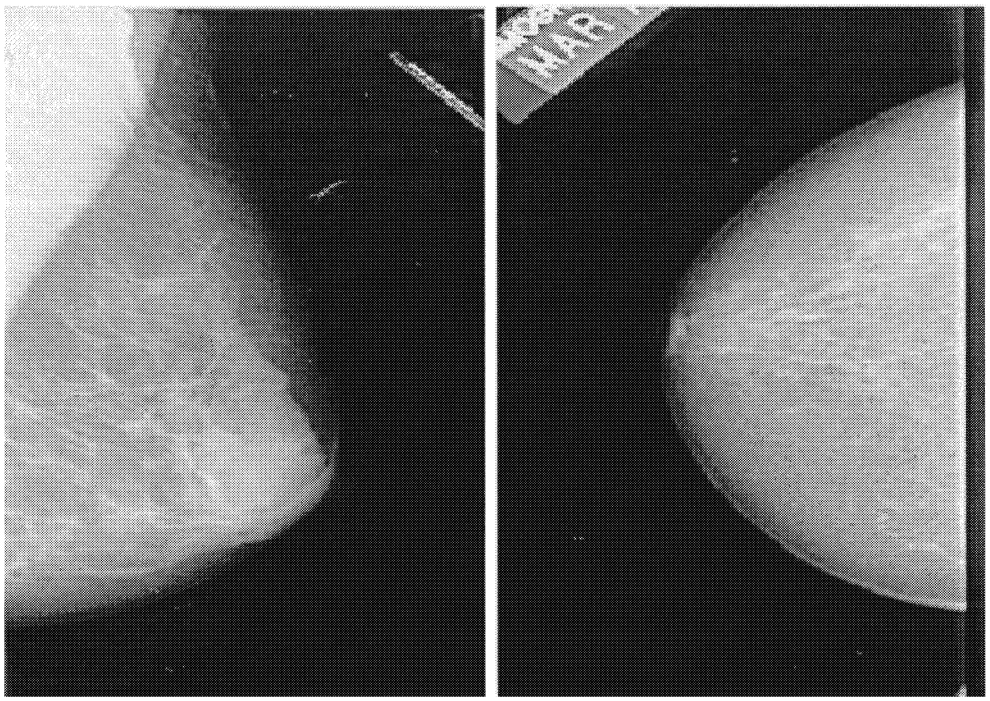

Figure 1. Two examples of mammographic images

$$T = l \int_{\Omega/K} \left(\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial x} \right)^2 \right) dxdy, \tag{3}$$

$$F = n \int_K dxdy = n|Z|, \tag{4}$$

and where $f(x,y)$ are the input data over a domain $\Omega \in \Re^2$, $u(x,y)$ is the function that will be restored, K represents the contour set, $|Z|$ is a measure of the contours contained in the image and, finally, l and n are two input parameters. The last two terms are weights determining a compromise among the three terms above. The first term in (1) assures the similarity of the computed u to the input data. We can say that, by means of this term, we lead the solution to be faithful to the data. The second term regulates the elasticity of the membrane: high values of l produce a smoothing of u while low values of the parameter itself assure that u is close to the input data f. The main property of this technique is that the continuity of the solution is weak: the solution tries to be continuous, but some discontinuities are allowed. Finally, the third term takes into account the "penalties" of the contours, that is for each discontinuity point there's a price to pay. It has to be stressed that the introduction of the last term, and the constraint on the domain of the second integral, causes the non-linearity of the functional, since the first two terms alone can be solved by the Euler-Lagrange equation:

$$u(x,y) - f(x,y) - l \left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) = 0. \tag{5}$$

This is equivalent to convolve the data with a linear filter whose impulse response is[15]

$$h_\lambda(x,y) = \frac{1}{2\lambda} \exp \left\{ -\frac{|x| + |y|}{\lambda} \right\}; \tag{6}$$

the introduction of the third term allows the functional to perfom a selective smoothing, preserving the image discontinuities and eliminating the noise overlapping the "good" signal. The minimization of (1) can be made by means of the line process $\{l_i, m_i\}$ indicating the continuity inside the interval $[u_i, u_{i-1}]$.

2.1.1. The Proposed Model

The WC representation introduced in the last section can be very useful for edge detection at different scale and noise levels. Here we propose to use these properties in order to classify texture which have different coarseness at different scales, in presence of (possibly high) noise.

For this purpose, we will use $|Z|$, which is the set of the contours of an image at a given scale level. Thus, given f on $\Omega \in \Re^2$, we are looking for the set $K_{l,n}$ such that:

$$E(u,K) = \int_{\Omega} (u(x,y) - f(x,y))^2 dxdy + l \int_{\Omega/K} \left(\left(\frac{\partial u}{\partial x} \right)^2 + \left(\frac{\partial u}{\partial x} \right)^2 \right) dxdy + n \int_K dxdy \tag{7}$$

will be minimized. Starting from $K_{l,n}$, we can define the characteristic function

$$\mathcal{H} : \Omega \rightarrow \{0,1\},$$

which is defined as:

$$\mathcal{H} = \begin{cases} 1 & \text{if } x,y \in K_{l,n} \\ 0 & \text{otherwise} \end{cases}$$

Thus, for given values of the parameters (l,n), $\mathcal{H}$ will represent a binary map of the detected edges in the original image $f(x,y)$. From $\mathcal{H}$ it is thus possible to compute the *density edges* for a given scale (l parameter) and for a given noise level (n parameter):

$$\Delta(H_{l,n}) = \frac{\int_k dxdy}{|R|} \tag{8}$$

where $|R|$ is the measure of the region $R \subseteq \Omega$. Different values of (8), computed for different values of (l,n), can be used for texture representation and classification.

2.2. Co-Occurrence Matrices Methods

The Co-occurrence Matrices Method (CMM) was proposed by Haralick in 1973,[10] and since then has been widely used for texture analysis. Consider an image $I[N_x \times N_y]$, with N_g gray-tone; if we assume that the textural information in I is given by the spatial relationship between the gray tone, this texture-context information will be adequately specified by the matrix of relative frequencies P_{ij} with which two neighboring resolution cells, separated by distance d, occur on the image, one with gray tone i and the other with gray tone j. Such matrices of gray-tone spatial-dependence frequencies are a function of the angular relationship between the neighboring resolution cells, as well as a function of the distance between them:

$$CM\Phi(d,\psi) = [P_{ij}(d,\psi)]. \tag{9}$$

In texture classification, individual elements of CMs are seldom used; instead, features are computed from the matrices. A large number of textural features have been proposed, starting with the original fourteen features described in.[10] Here we used the set of features listed above:

$$f_1 = \sum_{i,j} P_{i,j}^2, \tag{10}$$

$$f_2 = -\sum_{i,j} P_{i,j} log(P_{i,j}), \tag{11}$$

$$f_3 = \sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)P_{i,j}}{\sigma_i \sigma_j}, \tag{12}$$

$$f_4 = \sum_{i,j} (i-j)^2 P_{i,j}, \tag{13}$$

$$f_5 = \frac{\sum_{i,j}(ij)P_{i,j} - \mu_x \mu_y}{\sigma_x \sigma_y}, \tag{14}$$

where μ_x, μ_y, σ_x and σ_y are the means and standard deviations of row and column sums, respectively, and

$$\mu_i = \sum_i i \sum_j P_{i,j}, \qquad \mu_j = \sum_j j \sum_i P_{i,j}$$

$$\sigma_i = \sum_i (i-\mu_i)^2 \sum_j P_{i,j}, \qquad \sigma_j = \sum_j (j-\mu_j)^2 \sum_i P_{i,j}$$

2.3. Gabor Energy Filters

The Fourier Transform (FT) of a function $f(x)$ gives a measure of its irregularities (high frequencies), but this information is not spatially localized. For localizing the information obtained by the FT, Gabor[16] defined a new decomposition using a Gaussian window in the Fourier integral. These functions have been later extended to 2-D by Daugman.[17] A Gabor function is given by

$$G(x,y) = g(x',y') \exp[2\pi j(Ux+Vy)], \tag{15}$$

with

$$(x',y') = (x\cos\theta + y\sin\theta, -x\sin\theta + y\cos\theta).$$

They are rotated spatial-domain coordinates; (u,v) denote frequency-domain coordinates, and (U,V) represent a particular 2-D frequency.[17] The complex exponential is a 2-D complex sinusoid at frequency

$$\omega = \sqrt{U^2+V^2}$$

and

$$\phi = \arctan(V/U);$$

it specifies the orientation of the sinusoid. The function $g(x,y)$ is the 2-D Gaussian

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \cdot \exp\left\{-\frac{1}{2}\left[\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2\right]\right\}, \tag{16}$$

where σ_x and σ_y are related with the spatial extent and bandwidth of the filter. The Gabor function can thus be viewed as a Gaussian modulated by a complex sinusoid. In this paper we will assume that $\sigma_x = \sigma_y = \sigma$. This means that the parameter θ is not needed and the Gabor function becomes:

$$G(x,y) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{(x^2+y^2)}{2\sigma^2}\right\} \cdot \exp[2\pi j(Ux+Vy)]. \tag{17}$$

We can define now the Gabor Filter G_g:

$$G_g(I(x,y)) = |I(x,y) * G(x,y)|, \tag{18}$$

where $I(x,y)$ is an image. Gabor Filters applied to texture analysis measure the similarity between neighbourhoods in an image and Gabor functions. A family of Gabor functions can be generated for varying frequencies (ω) and Gaussian window standard deviations (σ); expressing (U,V) by means of the orientation θ,[18] we can write the Gabor function as

$$G_g(x,y|\lambda,\theta,\phi,x_0,y_0) = \exp\left\{-\frac{[(x-x_0)^2+(y-y_0)^2]}{2\sigma^2}\right\} \sin\left(\frac{2\pi}{\lambda}(x\cos\theta - y\sin\theta) + \phi\right), \tag{19}$$

where (x_0, y_0) specify the center of the Gaussian.

For texture analysis purpose, we will compute the Gabor Energy Filter (GEF) at each pixel for each combination of wavelength and orientation, where the energy is defined as the sum over the phases of the squared filter values. That is

$$S^2(x_0,y_0|\lambda,\theta) = \left[\sum_{x,y} G(x,y|\lambda,\theta,0,x_0,y_0)I(x,y)\right]^2 + \left[\sum_{x,y} G(x,y|\lambda,\theta,\pi/2,x_0,y_0)I(x,y)\right]^2. \tag{20}$$

Energy calculated using equation (20) for each combination of λ and θ may be used as textural features.[18]

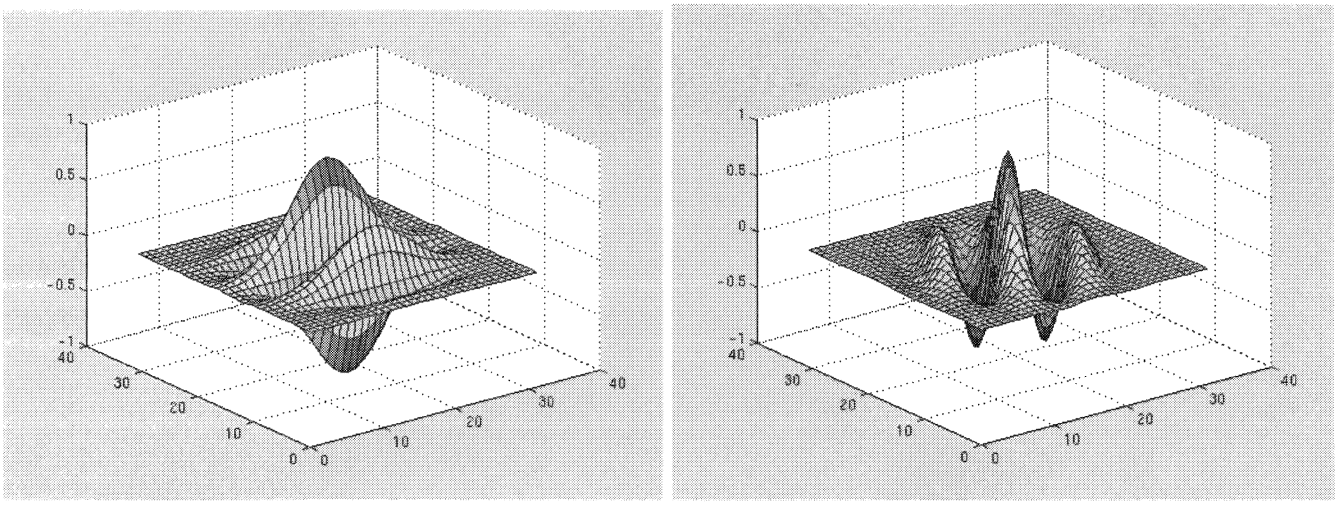

Figure 2. Two examples of 2D Gabor functions.

2.4. Wavelet Filters

Wavelets,[19 20] have been shown to be useful for texture classification in literature, due to their finite duration which provides both the frequency and spatial locality.[21] The hierarchical wavelet transform uses a family of wavelet functions and its associated scaling functions to decompose the original signal (image) into different subbands. The decomposition process is recursively applied to the low-frequency sub-band to generate the next level of the hierarchy. If an orthonormal wavelet basis has been chosen, the computed coefficients are independent and possess distinct features of the original signal. Wavelet Packets (WP) can be described by the following collection of basis function,[19 20]:

$$W_{2n}(2^{p-l}x - l) = \sqrt{2^{l-p}} \sum_m h(m - 2l)\sqrt{2^p} W_n(2^p x - m), \tag{21}$$

$$W_{2n+1}(2^{p-l}x - l) = \sqrt{2^{l-p}} \sum_m g(m - 2l)\sqrt{2^p} W_n(2^p x - m) \tag{22}$$

where p is a scale index, l is a translation index, h is a lower-pass filter, g is a hig-pass filter with

$$g(k) = (-1)^k h(l - k).$$

The function $W_o(x)$ can be identified with the scaling function Φ and W_1 with the mother wavelet Ψ. The inverse relationship between WP of different scales can be specified as follows:

$$\sqrt{2^p} W_n(2^p x - k) = \sum_l h(k - 2l)\sqrt{2^{p-l}} W_{2n}(2^{p-l}x - l) + \sum_l g(k - 2l)\sqrt{2^{p-l}} W_{2n+1}(2^{p-l}x - l). \tag{23}$$

Due to the orthonormal property, the WP coefficients at different scale and position of a signal $f(x)$ can be easily computed via

$$C^p_{n,k} = \sqrt{2^p} \int_{-\infty}^{\infty} f(x) W_n(2^p x - k) dx \tag{24}$$

for discrete signals. WP coefficients may be computed efficiently as follows: from (23), we have

$$C^p_{n,k} = \sum_l h(k - 2l) C^{p-1}_{2n,k} + \sum_l g(k - 2l) C^{p-1}_{2n+1,k}. \tag{25}$$

Using equations (21-22), we have

$$C^{p-1}_{2n,l} = \sum_m h(m - 2l) C^p_{n,m}, \tag{26}$$

$$C^{p-1}_{2n+1,l} = \sum_m g(m - 2l) C^p_{n,m}. \tag{27}$$

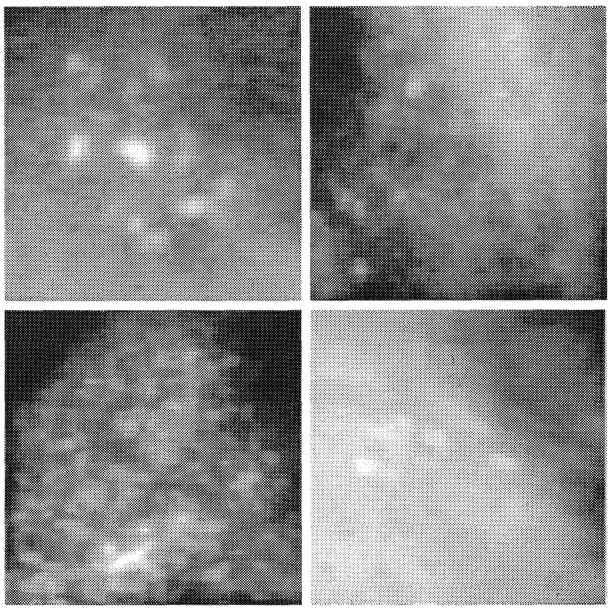

Figure 3. Four examples of ROIs containing microcalcifications.

Note that $C^0_{0,k}$ is given by

$$C^0_{0,k} = \int_{-\infty}^{\infty} f(x)\Phi(x-k)dx. \tag{28}$$

The 2-D WP basis functions can be expressed by the product of two 1-D WP basis functions along the horizontal and vertical directions respectively,[19] .[20] The corresponding 2-D filter coefficients have four groups:

$$hh(k,l) = h(k)h(l),$$

$$hg(k,l) = h(k)g(l),$$

$$gh(k,l) = g(k)h(l),$$

$$gg(k,l) = g(k)g(l).$$

3. EXPERIMENTS

3.1. Data Selection

We tested the performance of WCTR, CMM, GEF and WP for microcalcifications detection on a database of 81 images produced by the "Centro per la Cura e la Prevenzione dei Tumori" of the University of Rome "La Sapienza"; each image was digitized from film using a CCD camera operating at a spatial resolution of 604×575 pixels for image; the pixel rate was of $11,5MHz$, and the pixel size of $10\mu m \times 15\mu m$. From the 81 images, 151 Region of Interest (ROI) were selected by expert radiologists, each of 128×128 pixels. Among the selected 151 ROIs, 75 were positive and 76 were negative; examples of four different ROIs are shown in Figure 3. In a preprocessing step, each extracted ROI was stretched to the normalized gray-level range of 0-255.[3]

3.2. Feature Extraction

3.2.1. Weak Continuity Texture Representation

As shown in Section 2.1.1, each ROI was represented by edge density vales, computed for different values of the parameter l (scale parameter) and n (noise parameter). It has to be stressed that the choice of the parameters, and particularly the choice of l, it's a crucial point for the success of the representation. Actually, the scale level characterizes the binary matrix containing the edges of the analysed ROI; a too large value of l should lead to a characterization with few features, while a too small value of the same parameter should provide a too accurate

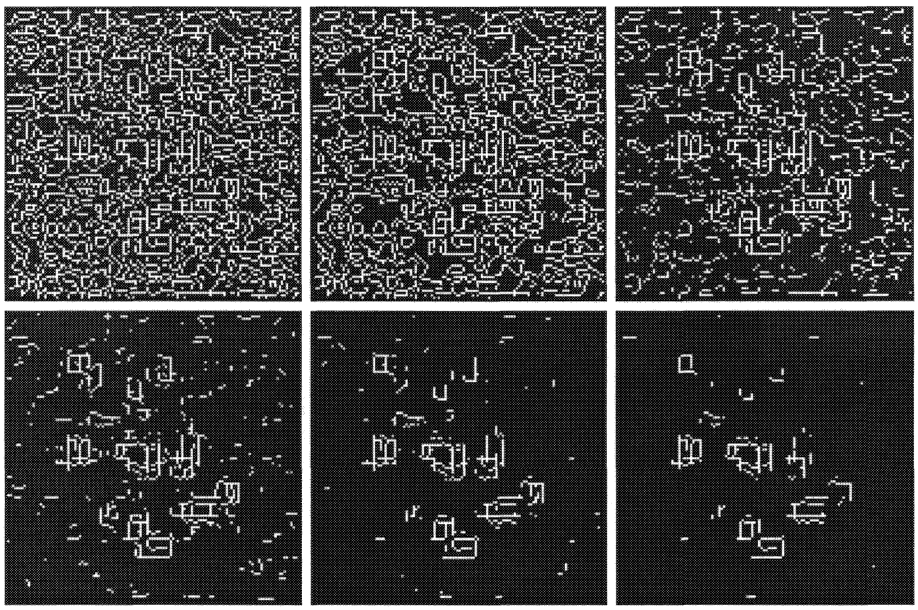

Figure 4. An example of six edge maps computed using the WCTR. The ROI analysed is the one showed in Figure 3, on top left.

description. On the other side, the noise level acts like a threshold parameter, which tells when the weak membrane can breaks. In this study we chose $l = 10$ and six different noise levels: $n = 5, 10, 15, 20, 25, 30$, thus obtaining six edge maps. From each map we computed the density value; so, each ROI was represented by a feature vector of six components (WCTR6). Figure 4 shows an example of computed edge maps.

3.2.2. Co-occurrence Matrices Method

As shown in Section 2.2, CMs are functions of the angular relationship between the neighboring pixel as well as functions of the distance between them. Typically choosed values for the angle θ are $0°, 45°, 90°, 135°$. For an N_g gray levels image, the CM will be of size $N_g \times N_g$. If N_g is too large, the number of pixel pairs contributing to each element of the CM will be low, and the statistical significance poor. On the other hand, if the number of gray levels is too low, much of the texture information may be lost in the image quantization. It must then be stressed that the value of d is critical too for the analysis and it must be compared with the typical size of the texture pattern element; if d is too large, we are averaging over several texture elements, and we get random correlations; if d is too small, we are looking at details of the pattern. In this study we chose $N_g = 16$, $d = 1$ and $\theta = 0°, 45°, 90°, 135°$; the five features described in Section 2.2 have been computed for these values, obtaining for every ROI a feature vector of 20 coefficients (CM20).

3.2.3. Gabor Energy Filters

As shown in Section 2.3, a GEF set is specified by the values of the parameters $\lambda, \theta, x_0, y_0$. According to the results obtained in,[9] here we used the following set: 4 frequencies (wavelengths of 32, 16, 8 and 4 pixels), 16 centers of the Gaussian ($x_0 = y_0 =$ 16, 48, 80, 112 pixels) and $\theta = 0°, 45°, 90°, 135°$. So, each ROI was represented by a feature vector of 256 components (GEF256).

3.2.4. Wavelet Packets

Following the procedure shown in Section 2.4, WP were applied to each ROI. We computed WP corresponding to a Level 1 and a Level 2 of decomposition. Following the example of many studies in literature,[21] ,[22] we chose to work with Daubechies wavelets. These are orthonormal and of compact support; thus they are suitable for analysis of images. Indeed, the orthonormality conditions ensures that the representation of a signal at different levels is uncorrelated; the regularity condition gives a sufficient decay of the mother wavelet in the frequency domain. We chose the Daubechies wavelets D_6 and D_{20}, which proved to give a good combination of regular prototype wavelets with varying sizes to extract texture information at different spatial frequencies. We computed the energy and

Layers Number	3
Output Neurons	1
Transfer Function	Log-sigmoid
Learning Rule	backpropagation
error goal	0.1

Table 1. Network architecture and learning parameters

entropy features from the Level 0 decomposition (as to say the original image) and from the level packet of the Level 1 and Level 2 decomposition according to the following equations:

$$Energy = \frac{\sum_{i,j} x_{ij}^2}{lenght * breadth}, \tag{29}$$

where x_{ij} is the computed wavelet packet value at the i-th row and j-th column of the wavelet packet and length and breadth are the dimensions of the wavelet packet,

$$Entropy = -\sum_{i,j} \left[\frac{xij^2}{norm}\right] log_{10} \left[\frac{x_{ij}^2}{norm}\right], \tag{30}$$

with

$$norm = \sum_{ij} x_{ij}^2. \tag{31}$$

Each ROI was then represented by a feature vector of 18 components (WP18).

3.3. Classifier

An artificial neural network is a computer architecture consisting of a single interconnected processing elements called neurons,[23] ,[24] .[25] A weight w_{ij} (coupling strength) characterizes the interconnections between any two neurons i and j. The input to each neuron is a weighted sum of the outputs incoming from the connected neurons. Each neuron operates on the input signal using his activation function b and produces the output response. The typical activation functions are linear, threshold and sigmoid,[23] ,[24] .[25] Normally the neurons are organized in an architecture with input nodes, interfacing the neural network and the external world, output nodes, producing the network's responses, and hidden nodes, having the task of correlating and building up an "internal representation" of the analyzed problem. Network's capacity and performance depends on the number of neurons, on the activation functions used, and on the neurons' interconnections. Another important attribute of artificial neural networks is that they can efficiently learn nonlinear mappings through examples contained in a training set, and use the learned mapping for complex decision making,[23] ,[24] .[25]

A three-layer, backpropagation neural network was employed as classifier in this research. In Table 1 are summarized the network architecture and the learning parameters; the initial weights are randomly selected from $[0.0, 1.0]$. Textural features extracted as described in Section 3.2, are used as the input signals of the input layer. There is a single output node for classification into positive or negative ROI. A non-linear sigmoid function with zero and one saturation values is used as the activation function for each neuron, and is defined as[11]

$$o_j = \frac{1}{1 + \exp\left(\sum_i w_{ij} o_i + v_j\right)} \tag{32}$$

where o_j is the output of the j-th neuron and v_j is the threshold value of the j-th neuron. The network is trained to provide a 1.0 output value for a positive ROI and a 0.0 output value for a negative ROI. In the training process, the weights between the neurons are adjusted iteratively so that the differences between the output values and the target values are minimized. In this study, the training process is stopped when the error per training case becomes smaller than 0.1.

	WCTR6	CM20	GEM256	WP18
$A_z(set1)$	0.92	0.81	0.84	0.86
$A_z(set2)$	0.93	0.87	0.89	0.90

Table 2. Classification results.

3.4. Results and Discussion

Textural features obtained using WCTR6, CM20, GEF256 and WP18 as described in Section 3.2 were used as input for the network described in Section 3.3. We used two different combinations of training and test sets: 50 training cases and 101 test cases for the $set1$, and 70 training cases and 81 test cases for the $set2$. For every set, we randomly chose 10 different partitions of the data; this procedure was applied in order to prevent a dependency of the results on a particular partitioning of the data. The results of the network for all the different partitions were analysed by using ROC analysis.[26] ROC analysis is based on statistical decision theory and has been applied extensively to the evaluation of clinical diagnosis. The ROC curve represents the relationship between the True-Positive Fraction (TPF) and the False-Positive Fraction (FPF) for variation of the decision threshold. The TPF and the FPF denote the fraction of patient actually having the disease in question that are diagnosed as positive and the fraction of patients actually without the disease in question that are diagnosed as positive, respectively. The area under the ROC curve A_z is used as a measure of the classification performance. A higher A_z indicates better classification performance because a larger value of TPF is achieved at each value of FTF. An ideal performance produces an area of 1.0.

ROC analysis was applied on the classification results obtained for $set1$ and $set2$, for each of the 10 different partitions of the data and for each different texture representation. In this way 10 different values of A_z, for each set, were obtained: for the $set1$ we obtained an average A_z of 0.92 for WCTR6, of 0.81 for CM20, of 0.84 for GEM256 and of 0.86 for WP18. For the $set2$ we obtained an average A_z of 0.93 for WCTR6, of 0.87 for CM20, of 0.89 for GEM256 and of 0.90 for WP18. These results are shown in Figure 5 and summarized in Table 2. The better performance achieved by all feature representations with $set2$ indicates that the network has generalized better with the bigger training set; for every data partition, the best results were obtained with the WCTR6.

4. CONCLUSIONS

In this paper we proposed WCTR for detection of microcalcifications in mammographic images. The extracted features constituted the input of a neural network trained to classify between ROIs containing microcalcifications and ROIs containing normal tissue. The performance of the network was evaluated by means of a ROC analysis, and compared with other texture analysis methods commonly employed in literature for microcalcification detection: Co-occurrence Matrices, Gabor Energy Masks and Wavelet Filters. The obtained results show the effectiveness and potential usefulness of this approach.

ACKNOWLEDGMENTS

We would like to thank Prof. V.Virno and the staff of the Radiology Department of the "Centro per la Cura e la Prevenzione dei Tumori" of the University of Rome "La Sapienza" for their cooperation.

REFERENCES

1. E. A. Sickles and D. B. Kopans, "Mammographic screening for women aged 40 to 49 years: the primary practitioner's dilemma", *Anna. Intern. Med.*, **122**, no. 7, pp. 534-538, 1995.
2. M. Lany, "*Diagnosis and Differential Diagnosis of Breast Calcifications*", New York: Springer-Verlag, 1986.
3. A. K. Jain, "*Fundamental of digital image processing*", Prentice Hall, Englewood Cliffs, 1989.
4. S. Morio and S. Kawahara et al., "Expert system for early detection of cancer of the breast", *Comp. Biol. Med.*, **19**, no. 5, pp. 295-305, 1989.
5. L.Shen, R. M. Rangayyan and J. E. L. Desautels, "Application of shape analysis to mammographic calcifications", *IEEE Trans. Med Imag.*, **13**, no. 2, pp. 263-274, 1994.

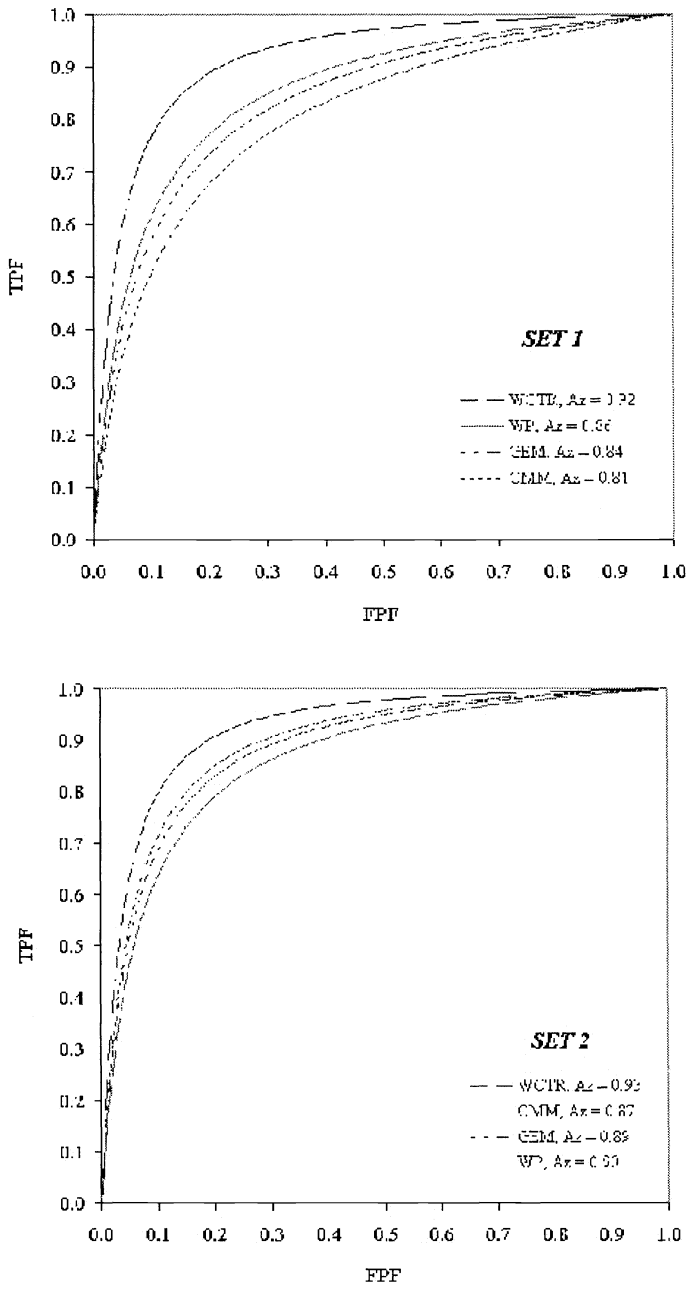

Figure 5. ROC analysis for *set1* (left) and *set2* (right)

6. J.K. Kim and H.W. Park, " Statistical textural features for detection of microcalcifications in digitized mammograms", *IEEE Trans Med Imag*, **18**, n 3, March 1999.
7. R.N. Strickland and H.I. Hahn, " Wavelet transform for detecting microcalcifications in mammograms", *IEEE Trans Med Imag*, **15**, no 2, pp 218-229, 1996.
8. B. Caputo, A. Troncone, D. Vitulano, "A Hierarchical Representation for Texture Classification", *Proceedings of VMV99*, pp 173-178, November 1999, Erlangen, Germany.
9. B. Caputo, G. E. Gigante, "Digital Mammography: Gabor Filters for Detection of Microcalcifications", *Proceedings of VMV00*, pp 375-381, November 2000, Saarbrucken, Germany.
10. R.M. Haralick, K. Shammugam, I. Dinstein " Textural Features for Image Classification", *IEEE Trans on Sys., Man., and Cyb.*, **3**, no 6, 1973.
11. A. C. Bovik, M. Clark and W. S. Geisler, "Mulichannel texture analysis using localized spatial filters", *IEEE Trans. Pattern Analysis and Machine Intelligence*, **12**, pp. 55-73, Jan 1990.
12. M. Vetterli and J. Kovacevic, *Wavelets and subband coding*, Englewood Cliffs, NJ: Prentice-Hall, 1995.
13. A. Rosenfeld, "Visual Texture Analysis: an Overview", *Tech. Rep. TR-406*, Computer Science Center, University of Maryland, August 1975.
14. A. Blake and A. Zissermann, *Visual Reconstruction*, MIT Press, 1987.
15. M. Gokmen and A. K. Jain, "$\lambda\tau$-Space Representation of Images and Generalized Edge Detector", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, **19**, No. 6, June 1997, pp. 545-563.
16. D. Gabor, "Theory of communication", *J. Inst. Elect. Eng.*, London, **93**, pp. 429-457, 1946.
17. J. D. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles", *Vision Res.*, **20**, pp. 847-856, 1980.
18. A. K. Jain and F. Farrokhnia "Unsupervised texutre segmentation using Gabor Filters", *Pattern Recognition*, **24**, no. 12, pp. 1167-1186, 1991.
19. I.Daubechies *Ten lectures on wavelets* Siam (1992).
20. M. Vetterli and J. Kovacevic, *Wavelets and subband coding*, Englewood Cliffs, NJ: Prentice-Hall, 1995.
21. S.G.Mallat "Multifrequency channel decompositions of images and wavelet models" *IEEE Transactions on Acoustics, Speech and signal processing*; **37**, n 12, Dec 1989.
22. A. P. Dhawan, Y. Chitre, C. Bonasso and M. Moskowitz, "Analysis of Mammographic microcalcifications using gray level image structure features", *IEEE Trans. on Med. Im.*, **15**, n.2, pp. 218-229, 1996.
23. R. P. Lippmann, "An introduction to computing with neural nets", *IEEE ASSP Magazine*, pp. 4-22, April 1987.
24. D. E. Rumelhart and C.R. Rosemberg, *Parallel Distributed Processing*, the MIT Press, Cambridge MA, 1986.
25. C. M. Bishop, *Neural Networks for Pattern Recognition*, Claredon Press - Oxford, 1995.
26. C. E. Metz, "ROC methodology in radiologic imaging", *Investigate Radiology*, **21**, no. 9, pp. 720-733, 1986.