

# Microcalcification Features Extracted from Principal Component Analysis in the Wavelet Domain

Nikolaos Arikidis, Spyros Skiadopoulos, Filippos Sakellaropoulos, George Panayiotakis, and Lena Costaridou  
Department of Medical Physics, School of Medicine, University of Patras,  
Patras, Greece, 265 00 Patras, Greece  
[costarid@upatras.gr](mailto:costarid@upatras.gr)

**Abstract.** In presence of dense mammographic parenchyma, microcalcifications (MCs) are obscured by anatomical structures, resulting in missed or/and false detections. Image analysis methods applied to improve visualization, detection and/or characterization of MCs, are targeted to MC SNR improvement and are unavoidably accompanied by MC background over-enhancement or false positive (FP) detections. A set of new features is proposed, extracted statistically with Principal Component Analysis from the wavelet coefficients of real subtle MCs in dense parenchyma. Candidate MCs are segmented and classified with the proposed features, using Linear Discriminant Analysis. Our method achieved 69% true positive fraction of MC clusters with 0.2 FPs per image in a dataset with 54 subtle MC clusters in extremely dense parenchyma.

## 1 Introduction

Mammography is currently the technique with the highest sensitivity available for early detection of breast cancer on asymptomatic women [1]. Detecting the disease in its early stages increases the rate of survival and improves quality of patient's life [2]. Detection of early signs of disease, such as microcalcifications (MCs), with screening mammography, is a particularly demanding task for radiologists. This is attributed to the high-volume of images reviewed, as well as the MC low contrast resolution, limited by their size, especially in case of dense breast, accounting for about 25% of the younger female population [3]. Although many analysis methods are reported [4], capable of enhancing or identifying specific image details as MCs, the most promising ones based on the wavelet transform, they typically produce disturbing background over-enhancement or false positive (FP) detections.

---

Please use the following format when citing this chapter:

Arikidis, Nikolaos, Skiadopoulos, Spyros, Sakellaropoulos, Filippos, Panayiotakis, George, Costaridou, Lena, 2006, in IFIP International Federation for Information Processing, Volume 204, Artificial Intelligence Applications and Innovations, eds. Maglogiannis, I., Karpouzis, K., Bramer, M., (Boston: Springer), pp. 730–736

In the framework of the wavelet transform, MCs contain relatively large amounts of high spatial frequency information. However, a large component of the power in a mammogram at high spatial frequencies is also noise [5,6] (dense tissue structure and film artifacts). Netch *et al.* [7], based on the circularly symmetric Gaussian model, used a Laplacian kernel to detect MCs as local maxima at different frequency bands. Strickland *et al.* [8] have shown that the average 2D gray level profile of MCs is well described by a circularly symmetric Gaussian function. Since the optimum detector of Gaussian functions is the Laplacian of Gaussian, they used a wavelet filter close to the Laplacian of Gaussian to detect significant peak responses to objects of similar shape and of the same size as the Gaussian filter. Soft or hard thresholding was used to set to zero the low amplitude wavelet coefficients, mostly dominated by noise. Other researchers [9-11] used globally or locally adapted linear enhancement functions to enhance high amplitude coefficients, corresponding to MCs, at various frequency bands. These methods assume statistical properties for the anatomical structure, which acts as structure noise in visualization (detection and/or characterization) of abnormalities.

Structure noise, especially in dense parenchyma, is highly correlated with abnormalities, such as MCs, producing wavelet coefficients comparable with those corresponding to MCs [12,13]. A method to describe the correlated noise is Principal Component Analysis (PCA), which replaces unknown image patterns with the linear combination of known image patterns and it is used for compression, classification or noise reduction tasks [14-17].

An MC specific method is proposed that detects image regions with very low contrast and uses a local method, trained by real MCs, to separate MC regions from structure noise and film artifacts.

## 2 Materials and Methods

One of the most successful paradigms of medical image analysis in mammography is Computer-Aided Detection (CADetection) systems for MC clusters [18]. The typical architecture of such a system [19] consists of a preprocessing step to increase MC SNR and segmentation of candidate MCs. Following, features of candidate MC regions are extracted and a classifier is trained for differentiating MCs from other image components. In a last step, a criterion is used to find only MCs that form clusters. In our approach, new features are suggested based on PCA of the wavelet coefficients of real MCs, capable of identifying individual MCs.

### 2.1 MC Eigen-Image Features in the Wavelet Domain

PCA is a mathematical tool that can find principal components from a set of real MC regions. Those principal components can be thought of as a set of images [15], named MC eigen-images, which together characterize the variation of MC regions. Then, each MC region is represented by the linear combination of the MC eigen-

images weights. For an unknown image, the segmented regions are replaced by the MC eigen-image weights and are classified.

Let a MC region  $I(x,y)$  be a two-dimensional  $N$  by  $N$  array, considered as a one-dimensional vector with length  $N^2$ . Considering a horizontal vector  $D=\{d_1, \dots, d_L\}$ , of  $L$  images of dimension  $N^2$  and denote  $M=\{\mu_1, \dots, \mu_L\}$  the mean vector of the population  $D$  ( $\mu_\lambda$  is the mean of the  $\lambda$ th image  $d_\lambda$ , where  $\lambda=1 \dots L$ ).

The covariance matrix  $C$  of  $D$  is defined by:

$$C = (D - M)(D - M)^t \quad (1)$$

where  $(D-M)^t$  is the transposed matrix of  $(D-M)$  and its size is of order  $N^2 \times \lambda N^2$ . To all vectors  $x_v = \{d_1(v), \dots, d_L(v)\}$  (where  $v=1, \dots, N^2$ ) the following transform is applied:

$$y_{v,k} = (x_v - M) \cdot A_k \quad (2)$$

where  $A$  is a matrix whose columns  $k=1 \dots L$  are formed from the eigenvectors of  $C$  [20], named MC eigen-images, ordered following the monotonic decreasing order of eigenvalues.

The wavelet transform can be considered as a mathematical microscope that emphasizes on image details, where the scale defines the detail size. Wavelet analysis is performed with Mallat's dyadic wavelet transform [21]. When the wavelet filter  $W_j f(x)$  is selected as the second derivative of the signal smoothed at scale  $j$ , high amplitude wavelet coefficients correlate with high curvatures. Gaussian functions, like MCs, are high curvature components at both vertical and horizontal direction and they can be differentiated from line-like structures. MCs have been highly correlated with the wavelet coefficients at scales 2 and 3 [10,22]. Thus, each MC region is replaced by four representations, which are the horizontal and vertical wavelet coefficients at the 2<sup>nd</sup> and 3<sup>rd</sup> scale (figure 1).

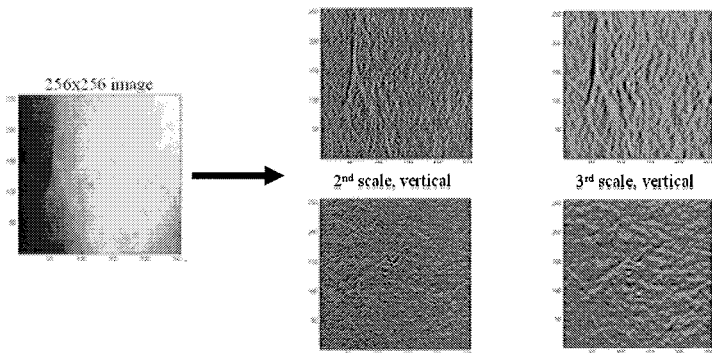


Fig. 1. The original image and its four (4) representations in the wavelet domain

When the wavelet transform is combined with PCA, the wavelet coefficients are used instead of using the pixel values to calculate the correlation matrix  $C$  [23]. The

input data set  $d$ , at the  $v$ th pixel is replaced by the wavelet coefficients  $W_j^{(v)}$  for each scale  $j$  of the wavelet transform.

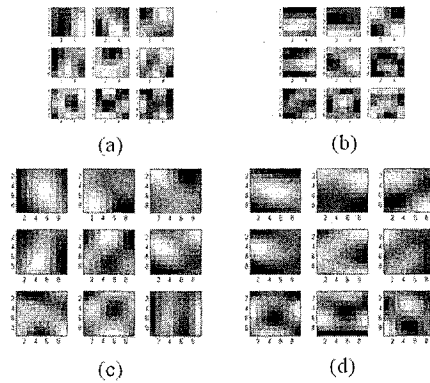
The correlation matrix  $C^{(j)}$  is calculated relative to the horizontal vector:

$$x_j = \{W_j^{(1)}, W_j^{(2)}, \dots, W_j^{(N^2)}\}. \tag{3}$$

For each band  $j$ , the transform matrix  $A^j$  contains the eigenvectors of  $C^j$ , which are the wavelet-based MC eigen-images (figure 2). Applying the matrix  $A^j$  to the  $\lambda$ th image, representing by the horizontal vector  $x_\lambda^j$ , the data are transformed to  $y_k^j$ :

$$y_k^j = (x_\lambda^j - M^j) \cdot A_k^j \tag{4}$$

where  $y_k^j$  is the projection of the wavelet coefficients  $x_\lambda^j$  to the  $k$ th principal component.



**Fig. 2.** Principal components (MC eigen-images) of the 2nd scale in the (a) horizontal and (b) vertical directions; and of the 3rd scale in the (c) horizontal and (d) vertical directions, respectively

Each MC eigen-image explains an amount of the variability of the MC regions, by the variance. The variability  $V_k$  is the variance of the  $k$ th MC eigen-image [24]:

$$V_k = \text{var}(y_k^j) \tag{5}$$

and can be measured as a percentage of the total variability. Nine (9) MC eigen-images are selected at each scale and direction, which accounts about 98% of MC regions variability.

The processing steps for MC eigen-images extraction are the following:

- 41 MC regions with sizes below 0.5mm, random shapes and contrasts were selected from an experienced radiologist specialized in mammography, as the training dataset.

- The MC regions are represented at the wavelet domain by the vertical and horizontal wavelet coefficients. At the 2nd scale, the MC region has 5x5 pixels size and at the 3rd scale, the MC region has 9x9 pixels size.
- From each scale and direction, 9 eigen-images are selected to describe 98% of MC regions variation. A total of 36 eigen-images are used as features to recognize MC regions from other image components.

## **2.2 CADetection for MCs**

### **2.2.1 Segmentation of Candidate MC Regions**

MCs are very small structures, visible as small bright spots in the mammogram, because their mass attenuation coefficient is higher than any other structure in the breast. However, due to the growth of the MCs, there is no absolute lower bound to their contrast. Very small MCs have low contrast relative to the background, which is sometimes close to the noise caused by either the film granularity or the inhomogeneous tissue background. Morrow [25] used the Weber ratio (2%) to segment MC regions. In our approach, an even lower contrast threshold criterion (0.5%) is proposed pointing at very low contrast MCs. The size criterion excludes signals below 3 pixels, which are likely to be noise, and signals above 100 pixels, which are likely to be macro-calcifications or line structures.

### **2.2.2 Feature Extraction**

Candidate MC regions are analyzed at the 2nd and 3rd scale, at the vertical and horizontal direction. Those regions are centered at the local maxima positions of the wavelet coefficients. Then, they are projected at the MC eigen-images and the resulting weights form a vector with 36 features.

### **2.2.3 Classification**

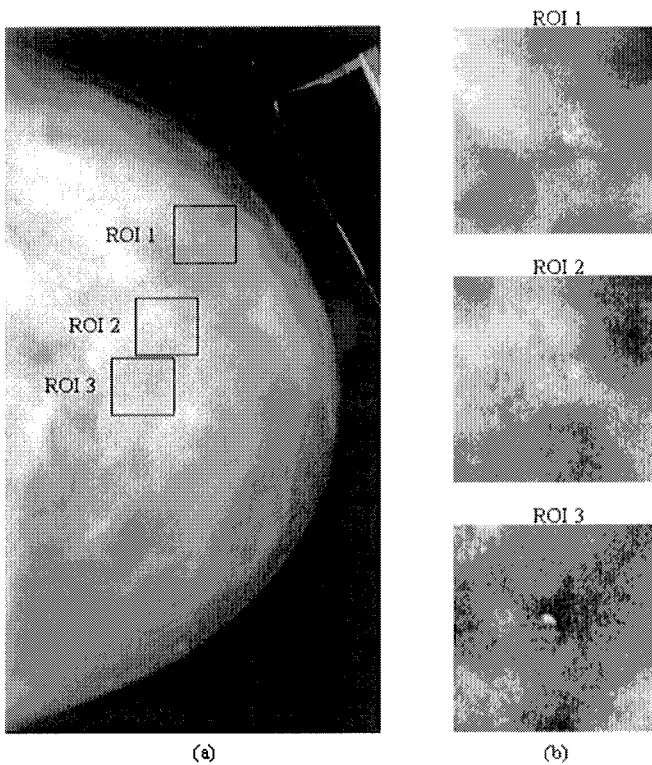
Linear Discriminant Analysis was used to classify candidate MC feature vectors in three classes – individual MCs, film artifacts and structure noise. The training set consisted of the dataset used in MC eigen-image extraction, as well as 20 verified film artifact regions and a large number of noisy structures.

### **2.2.4 Clustering**

Because isolated MCs are not clinically significant, the detection of clustered MCs is of paramount importance. Typically, at least 5 MCs per square centimeter are required to be considered a cluster, but three suspicious MCs could be enough to prompt a biopsy [26]. In our method, a cluster is defined and considered as a true positive (TP) when at least three candidate MCs have Euclidean distance less than 5mm.

### 3 Results

The method was tested on a dataset of 53 images. Specifically, 16 images were normal and 37 images contained 54 subtle MC clusters (46 malignant, 8 benign) in extremely dense parenchyma (density 4 of ACR BIRADS) with 12-bit pixel depth originating from Digital Database for Screen Mammography (DDSM) [27]. TP and FP clusters were counted for each mammogram and the number of FP clusters per image and the corresponding number of TP clusters are determined from an expert radiologist specialized in mammography. Our method achieved 69% TP fraction from 54 MC clusters with 0.2 FPs per image.



**Fig. 3.** (a) Mammographic image (A-1220\_1.RCC) with three TP detected ROIs. (b) Magnified ROIs (176x178 pixels size) with an intensity windowing function applied.

A representative example with the detected MC clusters is presented in figure 3 (a). Figure 3 (b) provides the magnified image regions (176x178 pixels size), where a window intensity function has been applied to better visualize the detected MC clusters.

## 4 Discussion and Conclusion

New features for the detection of individual MCs, based on PCA in the wavelet domain, are proposed. PCA analyses the region statistics of real MCs to produce a new feature set, named MC eigen-images. Candidate MC regions are represented as weights of the MC eigen-images. The wavelet transform focuses on the analysis at scales 2 and 3, where the SNR of the MCs is increased.

The sensitivity of the detection algorithm is controlled by the dataset of real MC regions, used for principal component eigen images extraction. Compared to other studies [7,8], the achieved FP rate is extremely low, lending itself for further processing for detection or classification. A feature step is to add more subtle MCs in the training set to improve sensitivity as well as FROC analysis to estimate the performance optimization.

## 5 Acknowledgements

We would like to thank the European Social Fund (ESF), Operational Program for Educational and Vocational Training II (EPEAEK II), and particularly the Program PYTHAGORAS I (B.365.011), for funding the above work. We also thank the staff of the Department of Radiology at the University Hospital of Patras for their contribution in this work.

## References

1. K. Smigel, Breast cancer death rates decline for white women, *J. Nat. Cancer Inst.* 87, 173 (1995).
2. L.A. Gaudette, R-N. Gao, M. Wysocki, F. Nault, Update on breast cancer mortality, *Health Reports* 9, 31-34 (1997).
3. V.P. Jackson, R.E. Hendrick, S.A. Feig, D.B. Kopans, Imaging of the radiographically dense breast, *Radiology* 188, 297-301 (1993).
4. R.N. Strickland, in: *Image-Processing Techniques for Tumor Detection* (Marcel Dekker Inc, New York, 2002)
5. R.M. Nishikawa, M.J. Yaffe, Signal-to-noise properties of mammographic film-screen systems, *Med. Phys.* 12, 32-39 (1985).
6. G.T. Barnes, D.P. Chakraborty, Radiographic mottle and patient exposure in mammography, *Radiology* 145, 815-821 (1982).
7. T. Netsch, H.O. Pitgen, Scale-space signatures for the detection of clustered microcalcifications in digital mammograms, *IEEE Trans. Med. Imag.* 18, 774-785 (1999).
8. R.N. Strickland, H.I. Hahn, Wavelet transforms for detecting microcalcifications in mammograms, *IEEE Trans. Med. Imag.* 15, 218-228 (1996).
9. P. Heinlein, J. Drexler, W. Schneider, Integrated wavelets for enhancement of microcalcifications in digital mammography, *IEEE Trans. Med. Imag.* 22, 402-413 (2003).
10. A.F. Laine, S. Schuler, J. Fan, W. Huda, Mammographic feature enhancement by multiscale analysis, *IEEE Trans. Med. Imag.* 13, 725-740 (1994).
11. P. Sakellariopoulos, L. Costaridou, G. Panayiotakis, A wavelet-based spatially adaptive method for mammographic contrast enhancement, *Phys. Med. Biol.* 48, 787-803 (2003).