16th IMEKO TC4 Symposium
Exploring New Frontiers of Instrumentation and Methods for Electrical and Electronic Measurements
Sept. 22–24, 2008, Florence, Italy

Features Extraction to Classify Microcalcification Clusters in Mammography

Arianna Mencattini[1], Giulia Rabottino[1], Marcello Salmeri[1],
Federica Caselli[2], Roberto Lojacono[1], Giulia Frondizi[1]

[1] Department of Electronic Engineering, University of Rome "Tor Vergata",
[2] Department of Civil Engineering, University of Rome "Tor Vergata",
Viale del Politecnico, 1 – 00133 – Roma, Italy,
Phone: +39–06 72597368, Email: {rabottino, mencattini, salmeri, caselli, lojacono}@ing.uniroma2.it

Abstract – Breast cancer is the first leading cause of death by cancer for women. To increase the survival rate it is necessary to detect lesions as soon as possible. Most early breast cancer can be diagnosed by detecting microcalcification clusters in mammographic images. The clusters appear as groups of small, bright particles with arbitrary shapes and distribution. Because of human factors such as subjective or varying decision criteria, distraction by other image features, large number of images to be inspected, or simple oversight, some diagnosis are missed. In this paper, we propose a method to classify clusters of microcalcifications characterizing the lesion by the extraction of geometrical (2D) and textural (3D) features. Then, through a statistical analysis of these features, we can choose the most discriminating between benign and malignant lesions and so design the classifier.

I. Introduction

Mammography is the most effective method for early detection of breast diseases [1]. However, because of human factors several breast lesions are missed during routine screening. The most important signs of breast cancer are masses and microcalcifications. Masses are difficult to be detected because of the minor difference in x–ray attenuation between normal and malignant glandular tissue, while microcalcifications, although they have high inherent attenuation properties, appear with low contrast due to their small size. In order to increase radiologists diagnostic performance, a lot of computerized image analysis techniques have been developed in last years with the aim of enhancing mammographic features, without emphasizing noise in order to make more effective the following detection phase [11]. Among these techniques, enhancement methods based on wavelet transform are proved to be very useful because of their multi–resolution properties [2]. In fact, in mammograms different features, such as microcalcifications, masses, background, and noise appear at different scales, and so they can be selectively enhanced, detected or reduced within different resolution levels. In this paper we consider mainly image regions containing microcalcifications, that represent the most important sign for an early diagnosis of breast cancer.

II. The whole CADe system

The whole Computer Aided Detection (CADe) system for microcalcifications segmentation and classification is described in Fig. 1. The main steps are summarized as follows:

(a) Extraction of a Region of Interest (ROI), actually manually performed.
(b) Image enhancement step by denoising and contrast enhancement.
(c) Segmentation of microcalcifications [3].
(d) Features extraction of segmented microcalcifications.
(e) Classification of microcalcifications in order to assign a malignancy index.

III. Performance evaluation of the CADe

A major limitation in developing a CADe system is that the performance evaluation is difficult and often subjective. The purpose of algorithm testing is twofold: firstly, a quantitative measure of the results should be provided, and then the comparison with existing algorithms should be performed.

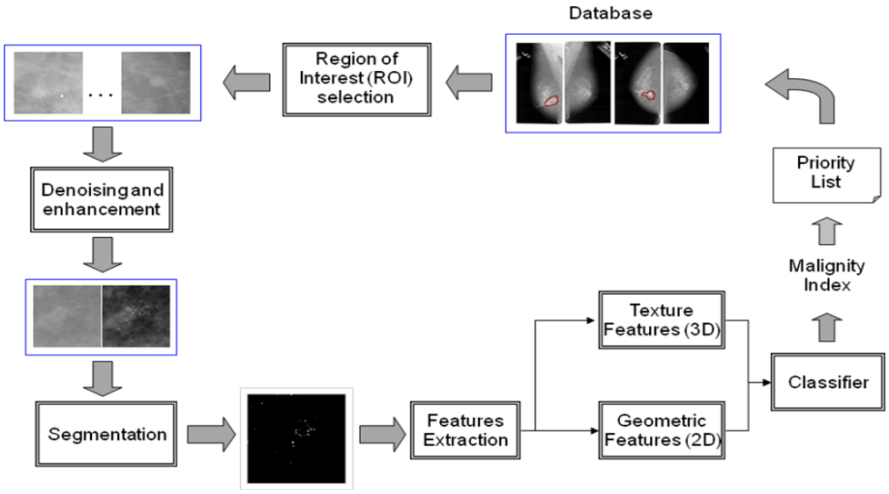


Figure 1. The whole CADe system for microcalcifications segmentation and classification.

Anyway, the critical point is to evaluate the accuracy of the results, since the ground–truth is often unavailable. Typically [6], two kind of evaluations can be performed: (1) testing each individual component of the CADe evaluating performance by suitable metrics, or (2) testing the whole CADe system. Obviously, the latter approach yields to the evaluation of the classifier performance, since a system designed to detect microcalcifications should provide a ratio of true microcalcifications detected versus false positives. Anyway, in order to metrologically investigate the performance of the whole CADe, we need to inspect separately each component.

(a) The ROI extraction is performed manually, by visual inspection of the mammographic image. In a future work we will study possible solutions for an automatic microcalcifications detection. So, at this point, we cannot evaluate the performance of this manual step.

(b) This step implements the denoising and the enhancement of the ROI, performed by using wavelet decomposition and applying wavelet thresholding. Denoising is critical in this context, since microcalcifications are bright spots, not much similar to their surroundings, but with similar characteristics in size and in luminance as noise. By using an undecimated discrete wavelet transform [2], we isolate microcalcifications separating them from noise and from background tissue and large object. Moreover, by performing a pseudo parallel processing by DDWT we extract separately background and foreground. To evaluate the performance of this step we compute the Contrast Improvement Index (CII) [7], the Peak to Signal to Noise Ratio (PSNR) [8,9], and the Average Signal to Noise Ratio (ASNR) [8,9], with respect to the original ROI. Table 1 shows the values of CII, PSNR, and ASNR obtained for the example in Fig. 2. All the images we consider are taken from Digital Database for Screening Mammography (DDSM) [4–5] that contains more than 3000 12–16 bpp grayscale images acquired by Screen Film Mammography and then digitalized by three kind of scanner devices, with a spatial resolution in the range [43–50] μm.

Figure 2. Original ROI (left); enhanced ROI (middle); segmented microcalcifications (right).

	C_{Orig}	C_{proc}	CII	$PSNR_{Orig}$	$PSNR_{Proc}$	$ASNR_{Orig}$	$ASNR_{Proc}$
ROI	0.0617	0.1546	2.5082	2.8390	9.7597	1.4644	4.1221

Table 1. Denoising and Contrast Enhancement performance metrics.

(c) The segmentation of microcalcifications is performed by applying a suitable threshold to the enhanced image, as shown in Fig. 2 (right). The difficulty of an objective evaluation of segmentation results, due to the small size and low contrast of microcalcifications in the original image, has led to leave to the radiologist the refinement of the threshold.
Some results obtained by applying denoising, contrast enhancement in a wavelet based framework, and segmentation are shown in Fig. 3 for five ROIs extracted from DDSM.

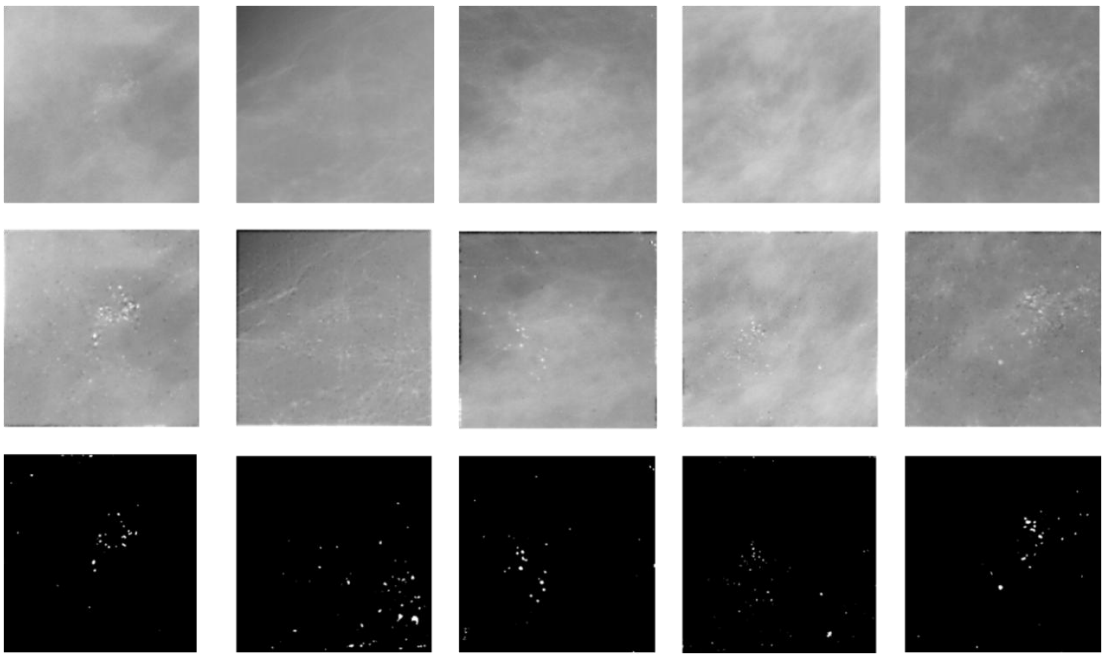


Figure 3. Microcalcifications enhancement and segmentation

(d) Features extraction step is the most critical in the context of microcalcifications segmentation and classification [10,12]. Owing to the great variability in distribution, size, luminance, cluster shape, several features are needed in order to classify microcalcification clusters into malignant and benign. In this work we propose the use of 42 different parameters in the preliminary step of the features extraction. Then we reduce the number of features after the investigation of their mutual correlation. The parameters that we consider are subdivided into: *geometrical features* [13,14] and *textural features* [13,14]. The first ones are related to the shape of microcalcifications, to the distances among them, and to many statistical properties of these geometric features, and also to the shape of the cluster, represented by the convex polygon containing the microcalcifications (*convex hull)*. Instead, the second set of features deals with the interior of microcalcifications, taking into account the intensity variation within each microcalcifications and within the convex hull. Finally, we also use more sophisticated texture features called *Haralick parameters.* The features are described below.

Geometrical Features

- Single microcalcification features
  - *Standard deviation of Area* $\sigma = \sqrt{A_k - A^2 / N-1}$, where A_k is the area (pixels) of the k^{th} microcalcification, A is the mean area, N is the number of microcalcifications in the cluster.
  - *Mean perimeter* of microcalcifications, *p.*
  - *Total area*, evaluated by the sum of the pixels representing the microcalcifications.
  - *Mean eccentricity*. Values close to 1 mean high circularity of the microcalcifications, whereas values close to 0 denote lengthened shape of microcalcifications.
  - *Mean circularity* evaluated as $C=4\pi A/p^2$ where A and p are defined above.

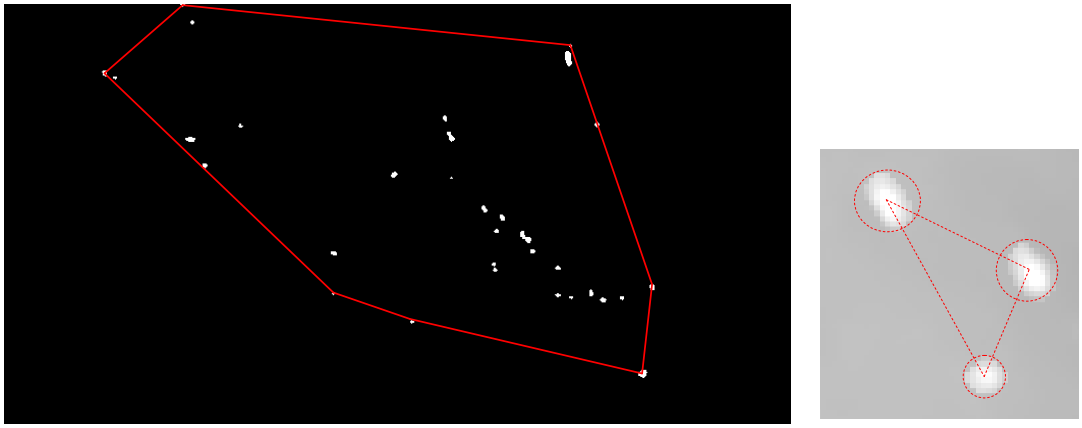
- Microcalcifications cluster features
  - Standard deviation of mutual distances.
  - Cluster perimeter, where the cluster is represented by the *convex hull* as shown in Fig. 3.
  - Approximative cluster area evaluated by the area of the convex hull.

Textural Features

- Haralick parameters [16]
  - Angular second moment, contrast, correlation, variance, inverse different moment, sum average, sum variances, sum entropy, difference variance, difference entropy, information measure correlation, maximum correlation coefficient.

- Histogram–based features
  - Mean intensity value of microcalcifications $\bar{I}$.
  - Skew function that measures the asymmetry of the histogram of microcalcifications.
  - Mean intensity value of the boundary of the microcalcifications $\overline{I_b}$.
  - Difference: $\bar{I} - \overline{I_b}$.
  - Contrast: $\bar{I} - \overline{I_b} \;/\; \bar{I} + \overline{I_b}$.

In order to better understand the meaning of the parameters, consider Fig. 3(a), where, as an example, the binary mask extracted for the ROI in Fig. 2 is used for a visual description of some features. In particular, the red polygon identifies the convex hull of the cluster. Moreover, in Fig. 3 (b) a portion of the same ROI containing only three microcalcifications is enlarged so that their *circularity* can be represented by dotted red circles and the mutual distances are shown by red dotted lines.



(a) (b)

Figure 3. (a) Convex hull of the cluster of microcalcifications, superimposed on the binary mask. (b) A portion containing three microcalcifications is enlarged so that circularity of each microcalcification and their mutual distances are represented by dotted circles and lines.

For the same portion Fig. 4 (a) shows the mean intensity value of the three microcalcifications, whose luminance is here visualized as a 3D surface. Finally, Fig. 4 (b) shows the histogram of the luminance of only microcalcifications (the so called foreground). We recall that the skewess of such histogram is often symptom of a malignancy.
Features selection is followed by a *features reduction* implemented by *statistical analysis* or by more sophisticated methods such as Principal Component Analysis (PCA) [17]. Here, we propose the representation of all features two by two on different plots, evaluated on a set of images containing both benign and malignant cases.
A possible result is shown in Fig. 5 (a)–(b), where two couples of features are high discriminating and the dotted curves describe the *decision functions* for the two choices. Figure 5 (c)–(d) represents two

couples of features that cannot be used to separate benign and malignant clusters.

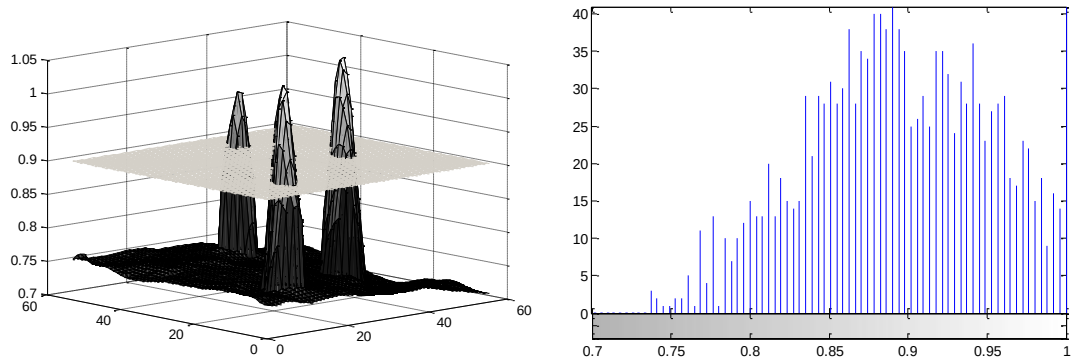


Figure 4. (a) Microcalcifications luminance visualized as a 3D surface and the mean intensity represented here by a horizontal plane. (b) The histogram of microcalcifications luminance.

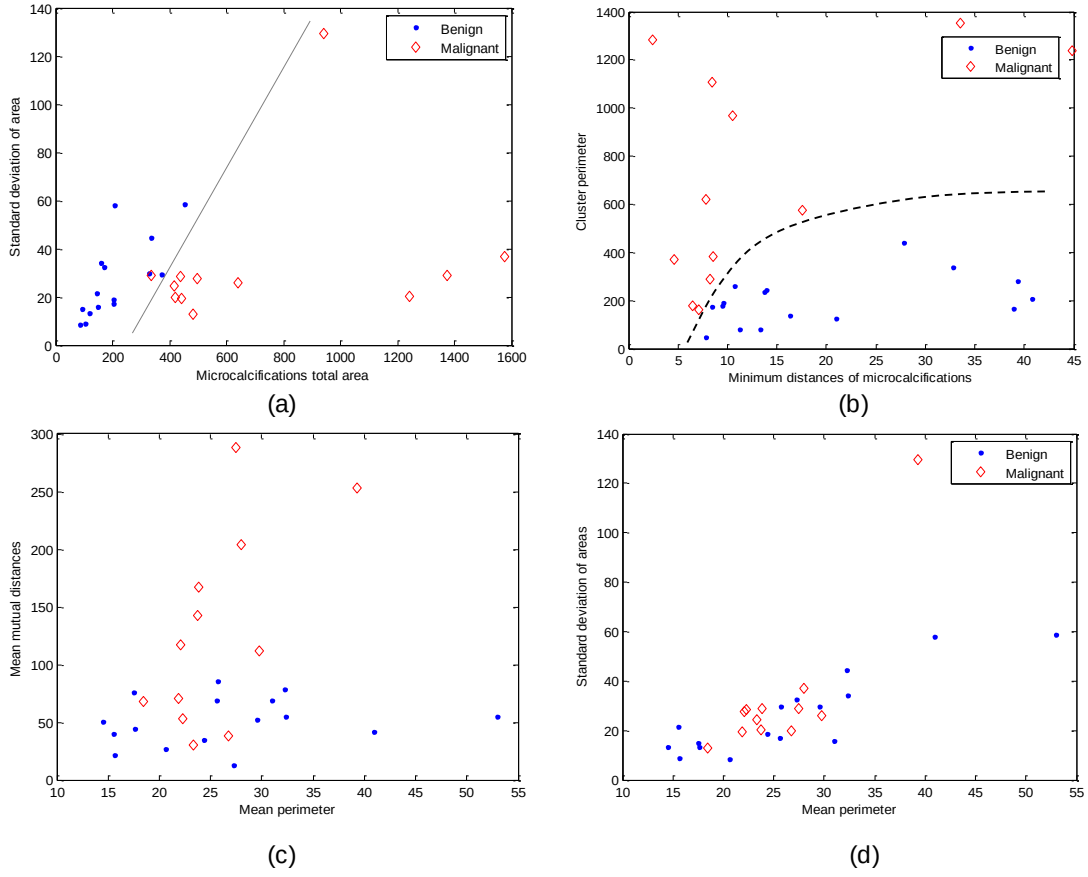


Figure 5. (a)–(b) High discriminating features. (c)–(d) Low discriminating features.

When a reliable subset of features is selected then a classifier can be trained to assign to the clusters a malignancy degree. The use of a fuzzy classifier will be investigated in a future work and compared with a neural network classifier [15] in terms of easiness of use and understandability for radiologists. Also the accuracy and the capability of improving sensitivity and specificity will be investigated.

IV. Conclusions

As a preliminary investigation toward the implementation of a complete CADe system for the early detection of breast cancer, in this paper we present an algorithm for the segmentation and features extraction of microcalcifications. Specific aspects of contrast enhancement and features selection have been described, and preliminary results, obtained on mammographic images, are provided. In a future work we will investigate and compare different classifiers suitable for this kind of application.

References

[1] World Health Organization (February 2006). Fact sheet No. 297: Cancer. Retrieved on 2007–04–26.
[2] A. Mencattini, M. Salmeri, R. Lojacono, M. Frigerio, F. Caselli, "Mammographic images enhancement and denoising for detection of tumoral signs using dyadic wavelet processing", *IEEE Trans. on Instrumentation and Measurement*, vol. 57, n. 5, pp. 1422–1430, Jul 2008.
[3] I. N. Bankman, T. Nizialek, I. Simon, O. B. Gatewood, I. N. Weinberg, W. R. Brody, "Recent advances in breast imaging, mammography, and computer–aided diagnosis of breast cancer", chapter *Algorithms for segmenting small low–contrast objects in images*, pp. 723 – 738. SPIE Press, Bellingham, USA, 2006.
[4] M. Heath, K.W. Bowyer, D. Kopans, R. Moore, P. Kegelmeyer, "Digital Mammography", chapter *Current status of the Digital Database for Screening Mammography*, pp. 457 – 460, Kluwer Academic Publishers, 1998.
[5] University of South Florida, University of South Florida digital mammography home page, 2000.
[6] M. A. Wirth, "Performance Evaluation of CADe Algorithms in mammography", chapter *Recent advances in breast imaging, mammography, and computer aided diagnosis of breast cancer*, pp. 639 – 699, SPIE Press, Bellingham, USA, 2006.
[7] P. Sakellaropoulos, L. Costaridou, G. Panayiotakis, "A wavelet–based spatially adaptive method for mammographic contrast enhancement", *Phys. Med. Biol.*, vol. 48, pp. 787–803, 2003.
[8] H. Li, K. Liu, S. Lo, "Fractal modeling of mammogram and enhancement of microcalcifications", *Nuclear Science Symposium & Medical Imaging Conference*, vol. 3, 1996.
[9] H. Li, K. Liu, S. Lo, "Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms", *IEEE Trans. on Medical Imaging*, vol. 16, vol. 6, pp. 785–798, 1997.
[10] H. D. Cheng, X. Cai, X. Chen, L. Hu, X. Lou, "Computer aided detection and classification of microcalcifications in mammograms: a survey", *Pattern Recogn.*, vol. 36, pp. 2967–2991, 2003.
[11] M. Kallergy, "Computer aided diagnosis of mammographic microcalcification clusters", *Med. Phis.*, vol. 31, n. 2, pp. 314–326, 2004.
[12] C. Y. Enderwick, E. Micheli–Tzanakou, "Classification of mammographic tissue using shape and texture features", *Int. Conf. of IEEE/EMBS*, pp. 810–813, 1997.
[13] Y. Jiang, R. M. Nishikawa, D.E. Wolverton, et al., "Mammographic features analysis of clustered microcalcifications for classification of breast cancer and benign breast diseases", *IEEE Int. Conf. Engeneering in medicine and biology*, pp. 594–595, 1994.
[14] H. Soltanian–Zadeh, S. Puorabdollah–Nezhad, F. Rafiee–Rad, "Shape–based and texture based feature extraction for classification of microcalcifications in mammograms", *Proceedings of SPIE*, vol. 4322, pp. 301–310, 2001.
[15] L. Wei, Y.Yang, R.M. Nishikawa, Y. Jiang, "A study on several machine learning methods for classification of malignant and benign clustered microcalcifications", *IEEE Trans. on Medical Imaging*, vol. 24, n. 3, pp. 371–380, 2005.
[16] R. M. Haralick, K. Shanmugam, I. Dinstein, "Textural features for image classification", *IEEE Trans. on Systems, Man and Cybernetics*, vol. 3, n. 6, pp. 610–621, Nov. 1973.
[17] N. Arikidis, S. Skiadopoulos, F. Sakellaropoulos, G. Panayiotakis, L. Costaridou, "Artificial Intelligence Applications and Innovations", chapter *Microcalcification Features Extracted from Principal Component Analysis in the Wavelet Domain*, vol. 204, pp. 730–736, Springer, 2006.