A Fuzzy Logic C-Means Clustering Algorithm to Enhance Microcalcifications Clusters in Digital Mammograms

L. Vivona, D. Cascio, R. Magro, F. Fauci, G. Raso

Abstract–The detection of microcalcifications is a hard task, since they are quite small and often poorly contrasted against the background of images. The Computer Aided Detection (CAD) systems could be very useful for breast cancer control. In this paper, we report a method to enhance microcalcifications cluster in digital mammograms. A Fuzzy Logic clustering algorithm with a set of features is used for clustering microcalcifications. The method described was tested on simulated clusters of microcalcifications, so that the location of the cluster within the breast and the exact number of microcalcifications is known.

I. INTRODUCTION

THE main difficulties encountered by radiologists in the detection of microcalcifications are due to the small size of microcalcifications (0.1-1 mm, mean diameter ~ 0.3 mm) and the fact that they are often hidden by the breast tissue. In this context, a CAD system may aid the physician to improve his performance [1]-[4]. Microcalcifications must bind together because clusters of microcalcifications may be an important warning for a breast cancer, present in 30%-50% of screening mammography.

A cluster is a set of localized microcalcifications, therefore a potential microcalcification cannot be associated with a cluster spatially "distant". However, the spatial association of microcalcifications is a necessary but not sufficient condition for proper clustering, therefore in the following analysis we will take into account other significant characteristics (features).

The proposed technique uses a Fuzzy C-means algorithm [5], [6] suitably modified by checking the distance of microcalcifications before proceeding with the association to the cluster: the matrix of data related to spatial coordinates is separated from that related to other features, and the distance is calculate only in geometrical space. If this distance is larger than the characteristic distance between microcalcifications, the degree of membership of the microcalcification to the cluster is minimized. In this way, the clustering is performed in the space of features without losing the information on the location of the microcalcifications.

II. METHOD AND PROCEDURE

The first phase of testing focused on verifying the proper operation of the algorithm on simulated microcalcifications. The choice to simulate clusters of microcalcifications can make possible to assess the efficiency of the clustering algorithm, since it is known not only the location of the cluster, but the exact number of microcalcifications present.

A. Microcalcifications segmentation

In the first step it was implemented a database of microcalcifications clusters extracted by mammographic images.

For the contours detection we used a filter which combine the properties of the Gaussian filter with the search properties of the Laplacian edge (1):

$$-\nabla^2 H_\sigma(x,y) = -\frac{1}{\pi\sigma^2}\left(\frac{x^2+y^2+\sigma^2}{2\sigma^4}\right)e^{-\frac{x^2+y^2}{2\sigma^2}} \qquad (1)$$

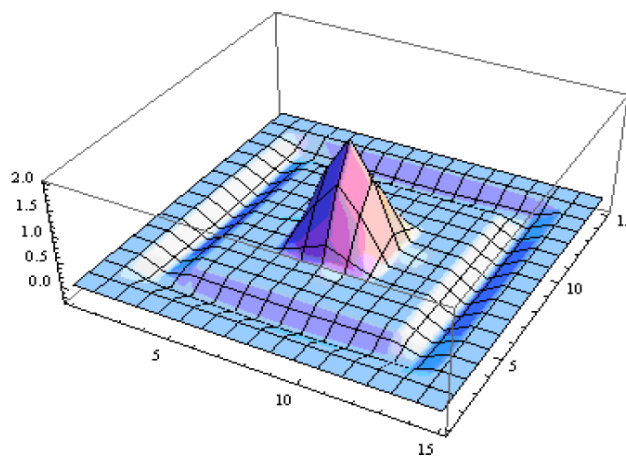The three-dimensional representation of the filter is shown in Fig. 1:

Fig. 1. 3 D representation of developed filter.

B. Clusters database creation

From real pathological segmented microcalcifications, correctly reported by radiologist, we created a database in which we have:

- for each cluster: the number of microcalcifications
- for each microcalcification: the cluster membership microcalcification ID, the mean intensity, the number of pixels, the coordinates and intensity of each pixel.

An example of database is shown in Table I.

TABLE I. EXAMPLE OF DATABASE

Cluster 1 N.Micro = 15

Micro ID	Mean Intensity	N.Pixels	Pixels Coordinates and Intensity
1	117	25	(1277, 894); (118) (1278, 894); (117) (1279, 894); (119) ……
2	120	31	(1269, 957); (119) ……..
…..	….	….	……
15	118	26	(1222, 959); (120)

C. Clusters simulation on healthy images

After database preparation, a random amount of microcalcifications, belonging to the same pathological cluster, is extracted from database and is injected on healthy images, taking care that each microcalcification maintain geometrical and morphological properties. In this injection we rescale the average intensity of each cluster taking into account the intensity of the parenchyma.

D. Algorithm testing

The proposed method uses a Fuzzy clustering algorithm for checking the distance of microcalcifications before proceeding with the association to the cluster (Fig. 3). The clustering is performed in the features space. The features with which the algorithm was tested are the follows: area, perimeter, eccentricity, mean intensity, standard deviation of intensity, skewness of intensity, defined as (2)

$$y = \frac{E(x-\mu)^3}{\sigma^3} \qquad (2)$$

kurtosis of intensity, defined as (3)

$$k = \frac{E(x-\mu)^4}{\sigma^4} \qquad (3)$$

In addition to these features, the coordinates of the centers of each microcalcification have been obtained.

Since the variables are on completely different scales of measurements, normalization was necessary to balance the contributions. The features were normalized in their range of variability, thus the value of each feature is always between 0 and 1:

$$X_{norm} = \frac{X - \min(X)}{\max(X) - \min(X)} \qquad (4)$$
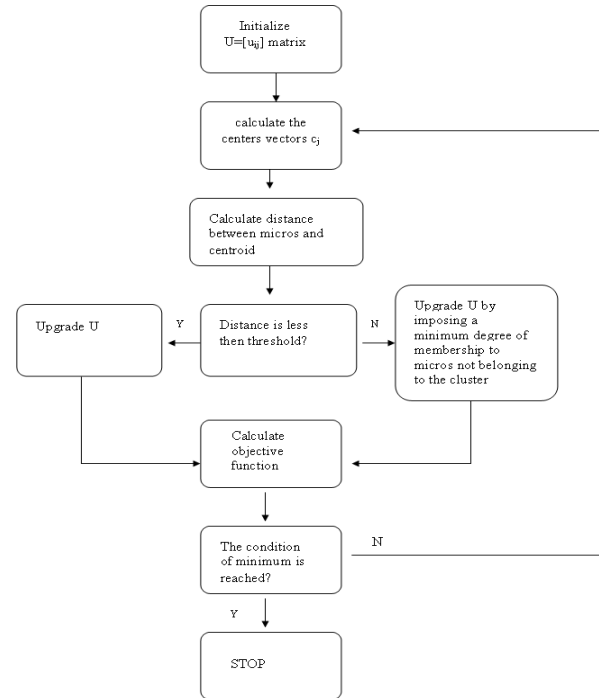

Fig. 3. Scheme of proposed algorithm.

We don't impose *a priori* a number of clusters but we let to automatically derive the optimal number of clusters (Fig. 4).
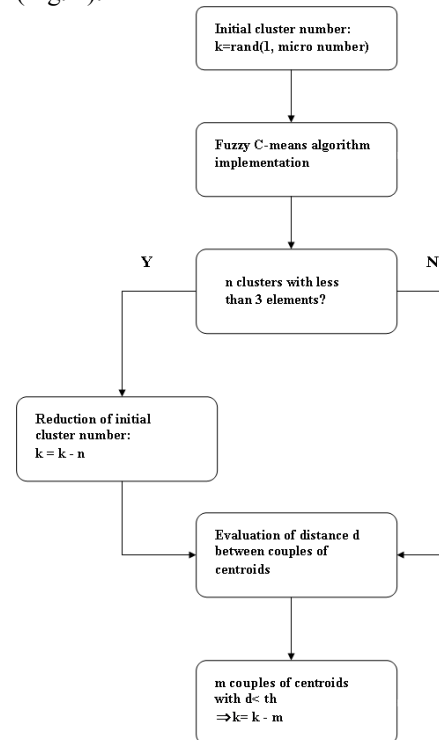

Fig. 4. Cluster number determination.

III. RESULTS AND DISCUSSION

We have applied the proposed algorithm to a set of simulated images and we have evaluated the performances of the algorithm compared to the performance of Fuzzy C-means algorithm applied to the same set.

Figure 5 shows the result of algorithms application: part a) shows the result of the clustering algorithm run with the C-means, part b) shows the result of clustering performed with proposed algorithm. Clustering is performed in multi-dimensions features space and the result is displayed in bidimensional features space of normalized geometrical centers coordinates X and Y .
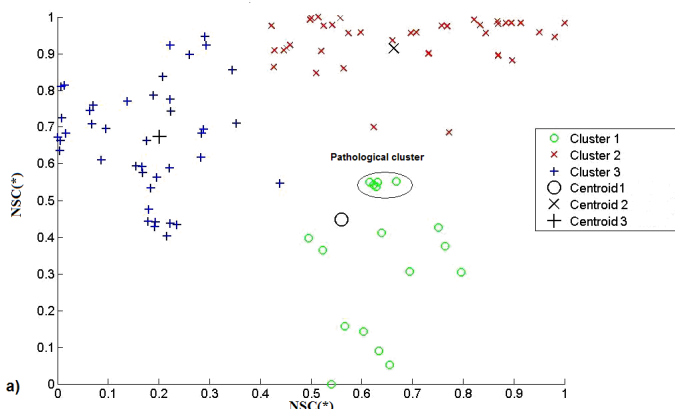

Fig. 5. a) Results of Fuzzy C- Means algorithms application
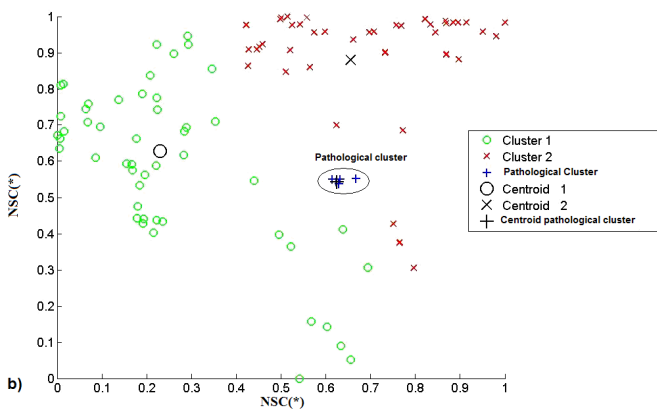

Fig. 5. b) Results of proposed algorithms application

(*) NSC = Normalized Spatial Coordinates

By comparing figures 5 a) and b) it is clear that:
- in fig. 5 a), pathological microcalcifications are erroneously associated by the Fuzzy C-means algorithm to other not pathological microcalcifications;

- in fig. 5 b), the proposed algorithm correctly perform the association of pathological microcalcifications in only one cluster.

IV. CONCLUSIONS

The proposed algorithm allows a microcalcifications clustering in the features space that takes into account correctly the Euclidean distance between microcalcifications. Despite these results are satisfactory, they are considered preliminary because at this early stage, the aim of this study was to test the ability of clustering algorithm on simulated microcalcifications.

We scheduled to test the functionality of the algorithm on a huge database of real data currently being acquired.

REFERENCES

[1] S. Ciatto, D. Cascio, F. Fauci, R. Magro, G. Raso, R. Ienzi, F. Martinelli, M. Vasile Simone, "Computer-Assisted Diagnosis (CAD) in Mammography: Comparison of Diagnostic Accuracy of a New Algorithm (Cyclopus®, Medicad) with Two Commercial Systems", Radiol. Med. DOI 10.1007/s11547-009-0396-4, 114, 626-635 (2009).
[2] N. Karssemeijer, J.H.C.L. Hendriks, "Computer-Assisted Reading of Mammograms", Eur. Radiol. 7, 743-748 (1997).
[3] W.J.H. Vedkamp, N. Karssemeijer, "Automated Classification of Clustered Microcalcifications into Malignant and Benign Types", Med. Phys. 27 (11), 2600-2608 (2000).
[4] R.M. Nishikawa, M.L. Giger, K. Doi, C.J. Vyborny, R.A. Schmidt, "Computer-Aided Detection of Clustered Microcalcifications: An Improved Method for Grouping Detected Signals", Med. Phys. 20 (6), 1661-1666 (1993).
[5] A. K. Jain, R. C. Dubes, Algorithms for clustering data, Prentice Hall, 1988, pp. 55-143.
[6] M. Berthold, D. J. Hand, Intelligent data analysis, Springer, 1999, pp. 321-351.