An approach based on RDM for analysis in breast cancer detection

Atef BOUJELBEN, Ali Cherif CHAABANI, Hedi TMAR, Mohamed ABID
CES-Computer, Electronic and Smart engineering systems design Laboratory,
National School of Engineers of Sfax: Tunisia

Atef.Boujelben@fss.rnu.tn, hedi.tmar@ceslab.org, mohamed.abid@enis.rnu.tn

Abstract

The cancer treatment is currently effective only if it is detected at an early stage. In this state, Mammography is the most efficient method for early detection. Due to the complexity of Mammography, the distinction of microcalcifications or opacities is very difficult. This paper deals with the problem of shape feature extraction in digital mammograms, particularly the boundary information. In fact, we evaluate the efficiency on boundary information possessed by mass region. We propose to modify the boundary features extracted with Radial Distance Measure "RDM" which means that the important boundary features in Region Of Interested "ROI". The objective of this modification is to ameliorate the computation complexity and the diagnosis quality. We use the Digital Database for Screening Mammography "DDSM" for experiments. Some classifiers as Multilayer Perception "MLP" and k-Nearest Neighbors "kNN" are used to distinguish the pathological records from the healthy ones. Using kNN classifiers we obtained 90.28% as sensitivity (percentage of pathological ROIs correctly classified). The results in term of specificity (percentage of non-pathological ROIs correctly classified) grows around 92, 82% using MLP classifier.

***Keywords**: Shape, Features, RDM, Mammogram, Analysis*

1. Introduction

Breast cancer is considered as a major health problem and constitutes the most common mortality that causes cancer among women in the word. However, although breast cancer incidence has increased over the last decade, breast cancer mortality has declined among women of all ages [1]. This favorable trend in mortality reduction may relate to improvements made in the breast cancer treatment and the widespread adoption of mammography screening. In the past decade many research efforts attempts to a generalization of approaches used in general imaging processing to cope with specific one; namely, medical image processing or analysis. In the past several years there has been tremendous evolution in mammography process. In processing techniques every method based on segmentation and classification is adopted [2][3][4]. However, breast tumors and masses appear in mammograms with different shape characteristics: malignant tumors usually have rough, microlobulated, or spiculated contours; whereas benign masses commonly have smooth, round, macrolobulated, or oval contours. Measures that can quantitatively represent shape roughness and complexity can assist in the classification of malignant tumors and benign masses.

Thus, the technique of analysis composed of two approaches: shape analysis which includes features based on morphological lesion [3][5] and analysis of texture [6][7][8][9][10][11][12][13]. Analysis of shape is decomposed in analysis of boundary and region.

In this context, Alvarenga et Al [14] had evaluated the performance and relevance of seven shape features; namely perimeter "P", normalized radial length "NRL", standard deviation "DNRL", area ratio "AR", contour roughness "R", the circularity "C" and the morphological-closing ratio "Mshape". The performance of these features in distinguishing malignant from benign tumors reveals "NRL" as the best feature in terms of "Az" (0.91), TPR (88.2 %) and FPR (92.3 %). AR has presented the highest TPR (93%); its FPR is the worst one (34.6 %).

These characteristics were enriched by adding some other various ones in [15]. The most important added characteristics are zero crossing "ZC" (i.e. a count of the number of times the radial distance plot crosses the average radial distance) and convexity "CONV", who allow representing the studied shape better than the

characteristics cited above. Via these shape features, the authors attempts discrimination between malignant and benign masses by using classification techniques. This work was resumed in [16] to carry out a mammograms segmentation and classification of identified ROI.
On the basis of this state of the art, we include the approach of shape analysis in our analysis process of mammograms. Generally speaking, shape feature extraction methods can be classified in two major categories; namely region and boundary. We are interested in boundary analysis, particularly, in analysis with RDM. In opposition to related work, we present some extensions of the classic RDM approach to cope with the amelioration of the analysis performance. In fact, to use Radial Angle Measure, the calculation in every point of region can expand the time in implementation of this feature. To minimize this temporal complexity, the amelioration is obtained by computing the features only in the local concave and convex points. Hence, the performance is illustrated in two points of view; diagnostic relevance and computation time of optimization. In fact, two techniques of classification are used to classify ROI.
The rest of this paper is organized as follows: section 2 describes the proposed scheme. Section 3 illustrates the classic and extended proposed RDM in boundary analysis. Section 4 presents the results obtained by the proposed scheme. Finally, we draw conclusions and some future issues in section 5.

2. The flow proposed in detecting breast cancer

The proposed scheme consists of three stages: identification of ROI, features extraction, and vector classification. Figure 1 shows the bloc diagram of the proposed scheme. ROI is selected from the image by fixing a rectangular box around the suspicious lesion area. A classical method of segmentation based in Sobel filter and seuillage is adopted. After the isolation of the ROI, extraction of features is adopted in ROI: this is the stage of flow in which we are interest in this paper. After that a classification part is started. The features vector is entered to the classifier to make decision.

In the next section, we will illustrate the extraction of features adopted in this paper.

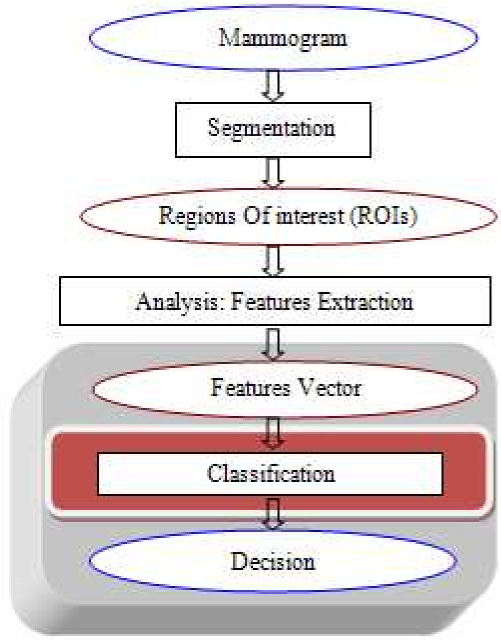

Figure 1: Proposed Flow

3. The Radial Distance Measure in analyzing boundaries

Boundary analysis which is often referred to as districting helps to define regions according to any criteria. The RDM is one of the most analysis methods.

3.1 Radial Distance Measure:

Radial Distance Measure "RDM" is a famous method used in shape analysis. However, an euclidian distance d(i) are calculated between the gravity center in region and all the points in boundary region (Figure2).

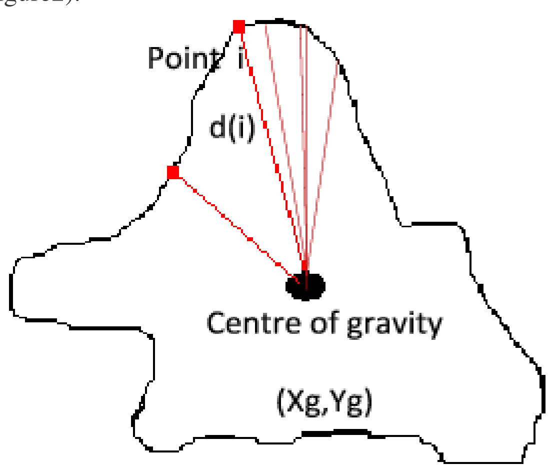

Figure 2: Illustrative Figure of RDM

$$d(i)=\sqrt{(Xi-Xg)^2+(Yi-Yg)^2} \quad (1)$$

To eliminate large calculations from characteristics, all radial distances were normalized by using the maximum value of the radial distances.

$$d_n(i)=\frac{d(i)}{\max[d(i)]} \quad (2)$$

Where n is the number of points (pixels) of the region boundary (the perimeter of region).

$$Xg=\frac{\sum_N X}{N} \quad (3)$$

$$Yg=\frac{\sum_N Y}{N} \quad (4)$$

Where N is the number of points (pixels) in the area region.

$$dmoy=\frac{\sum_N dn(i)}{N} \quad (5)$$

The features extracted in the RDM are cited below. We will only give the RDM features expressions related to this work.

- The Standard Deviation of the Normalized Radial Distance Measure "SDEV" is defined as the variance of the distances around the ray (the average distance dmoy previously definite) of a circle. This characteristic gives good quality information on the irregularity of contour. Indeed, when it is about a malignant tumor the value of SDEV tends towards a 0.5. On the other hand, in the case of benign tumor, the SDEV tends towards 0.

$$SDEV=\sqrt{\frac{\sum_N (dn(i)-dmoy)^2}{N}} \quad (6)$$

- The rugosity treats angular contours (contours which contain concave segments). It is given by the following equation:

$$R=\frac{1}{N}\sum_N \|dn(i)-dn(i+1)\| \quad (7)$$

- This characteristic discriminates between stellar contours and smooth contours. It is illustrated in the following equation:

$$Ar=\frac{1}{N*dmoy}*\sum_N (dn(i)-dmoy) \quad (8)$$

Where Ar=0, if $(dn(i) < dmoy)$

In practice, computation of these features increase the complexity of calculation. To make well the problem of complexity, we propose to calculate the features only in the concave and convex points. This will be the idea of the next subsection.

3.2 The extended Radial Distance Measure

To cure the problem of complexity, we propose to calculate the features cited in 3.1 only in the local concave and convex points (Figure 1). These points are defined as follows:

- The concave point (Pconcave (i)) of the contour is a point which has a radial distance d(i) lower than the radial distance d (i-1) and lower than the radial distance d(i+1).
- The convex point (Pconvexe (i)) of the contour is a point which has a radial distance d(i) higher than the radial distance d(i-1) and higher than the radial distance d(i+1).

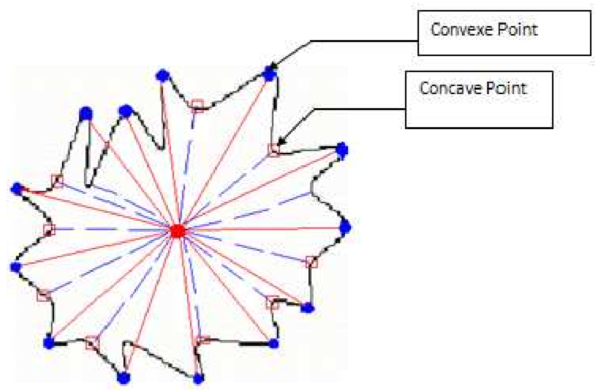

Figure 3: Illustrative figure of extended RDM

More formally:

Pconcave(i)=(i; $d(i) < d(i-1)$ and $d(i) < d(i+1)$)

Pconvexe(i)=(i; $d(i) \geq d(i-1)$ and $d(i) \geq d(i+1)$)

In this section, we showed how to optimize computation features time involve RDM so called extended RDM. In the next section, we will show the performance of this amelioration in analysis of the region in term of diagnosis relevance.

4. Classification and Test

The terminology used to determine the performance of a CADx System is defined as follows:

- Sensitivity: percentage of pathological ROIs correctly classified.
- Specificity: percentage of non-pathological ROIs correctly classified.

Because of the variation in the types of breast cancer, a larger number of cases can reduce the dependency "analysis techniques versus image sets". The performance of an algorithm is affected by the characteristics of a database like digitization techniques; namely pixel size, subtlety of cases, choice of training/testing subsets, etc.

4.1 The DDSM dataset:

The establishment of the DDSM allows the possibility of common training and testing data sets for the first time. The DDSM is the largest publicly available database of mammographic data. It contains approximately 2620 screening mammography cases.

From the total number of ROI included in the DDSM database, we used 200 ROI malignant and 100 ROI benign. For the evaluation we used the remaining 200 ROIs that contain 100 ground troughed abnormal regions together with 100 entirely benign.

4.2 Experimental results:

The basic classification is based on the two methods of classification KNN and MLP as shown in Table 1 and Table 2: where Table 1 and Table 2 show the results from discriminate analysis respectively RDM and extended RDM.

classifier	KNN	MLP
sensitivity	89.74 %	86.88 %
specificity	85.22 %	85.43 %

Table 1: Results from discriminant analysis: RDM

The table 1 illustrates the importance of RDM in shape analyzing. The result in term of sensitivity tends towards 89, 74 % in kNN classifier. The result in term of specificity tends towards more than 85 % in case of two classifiers kNN and MLP.

classifier	KNN	MLP
sensitivity	90,28 %	88,88 %
specificity	89,64 %	92,82 %

Table 2: Results from discriminant analysis: extended RDM

The table 2 illustrates the importance of extended RDM in shape analyzing. The result in term of sensitivity tends towards 90, 28 % in KNN classifier. The result in term of specificity tends towards 92, 82 % in MLP classifier. However, we can assume what the RDM is a good feature in differentiating the benign and malignant mass. But, we should not conclude if it is the better or the worst because of the experimental condition. In fact, the digitization can reflect the final result.

5. Conclusion

In this work, we attempted to improve the classification performance of RDM approach in analyzing the mammographic images. We presented some extensions and amelioration for this method which gives better performance in terms of diagnosis relevance and computation time optimisation. The results have been validated by two algorithms of classification: kNN and MLP. The results in term of sensitivity and specificity are variable tends towards 92.82 %. These results seem satisfactory, and the future work is focused on detection phase. The methods of detection like Level Set or wavelet will be considered. In the future work we will also illustrate the effectiveness of combination of the statistical texture features and shape ones in diagnostic process.

6. References

[1] M. Lanyi, “Breast calcifications”, *Springer-Verlag*, Berlin, Heidelberg, 1988.

[2] R. Amr Abdel-Dayem and R. Mahmoud El-Sakka, “Fuzzy Entropy Based Detection of Suspicious Masses in Digital Mammogram”, In *Images Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai*, China, September 2005, pp 1-4.

[3] C. Vachier, S. Muller and S. Bothorel, “Analyse morphologique des clichés mammographiques en vue de la détection des opacités du sein ”, In *Mathematics and Image Analysis*, Paris, septembre, 2000.

[4] M. Sameti “Detection of Soft Tissue Abnormalities in Mammographic Images for Early Diagnostis of Breast Cancer” University of British Columbia, November 1998.

[5]] U. Cascio, D. Fauci, B. Golosio, R. Magro, L.Masala, P. Oliva, G. Raso and S. Stumbo ,” Massive Lesions Classification using Features based on Morphological Lesion Differences”, In *Transactions on engineering computing and technology*, IMTC, March 2006.

[6] A. Boujelben, H. TMAR, J. M NIF and M. ABID, “Influence des caractéristiques issues de la matrice de co-occurrence sur le diagnostic du cancer du sein » In *deuxième Workshop AMINA, Application Médicale en Informatique: Nouvelle Approche*, Monastir, Tunisie, 17-19 Novembre 2006, pp 9-14.

[7] T. Lambrou , Alf D. Linney, R. D. Speller and A. Todd-Pokropek ,” Statistical classification of Digital

Mammograms Using Features from the Spatial and Wavelet Domains", In *Medical Image Understanding and Analysis* , volume 3, London UK, july, 2002.
[8] I. Christoyianni, E. Dermatas and G. Kokkinakis "Neural Classification Of abnormal tissue In Digital Mammography Using Statistical Features Of The texture", In *Proceedings of the 6th IEEE International. Conference on Electronics Circuits and Systems*, 1999, pp 117-120.
[9]R.Zwiggelaar and E. R. Denton, "Optimal segmentation of mammographic images", In *International Workshop on Digital Mammography*, Norwich, UK 2004.
[10] A.Oliver, X.Llado, R.Marti, J.Freixenet and R.Zwiggelaar, "Classifying mammograms using texture information", In *Medical Image Understanding and Analysis* Aberystwyth and Wales and UK, July 2007, pp 223-227.
[11] HS. Sheshadri and A. Kandaswamy,"Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms", In *Computerized Medical Imaging and Graphics*, 11 September 2006.
[12] HS. Sheshadri and A. Kandaswamy, "Breast Tissue Classification Using Statistical Feature Extraction of Mammograms*", In Indian J Med,* volume 23 No.3, 2 June 2006.
[13] R.N Mudigonda , R. M. Rangayyan and J. E. Leo Desautels, "Gradient and Texture Analysis for the Classification of Mammographic Masses". In *IEEE Transactions On Medical imaging*, volume 19, October 2000, pp 459-464.
[14] A.V. Alvarenga, A. Fernando and C. I. Wagner Coelho de Albuquerque, "Assessing the performance of the normalised radial length and convex polygons in distinguishing breast tumors on ultrasound images", Pereira Revista Brasileira de Engenharia Biomédica, volume 22, 2006, pp 181-189.
[15] A. Retico, P. Delogu and M.E. Fantacci and P. Kasae, "An Automatic System to Discriminate Malignant from Benign Massive Lesions on Mammograms", In *Medical Physics*, volume 14, Jan 2007, pp 596-600.
[16] P. Delogu, M.E Fantacci, Kasae P.and A. Retico," Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier", *In Comput Biol Med*, 10 Mar 2008.