Feature Extraction from contours shape for tumor analyzing in Mammographic images

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, aly_soft@yahoo.fr, hedi.tmar@ceslab.org, mohamed.abid@enis.rnu.tn

Abstract

The cancer treatment is effective only if it is detected at an early stage. In this context, Mammography is the most efficient method for early detection. Due to the complexity of this last, 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 evaluated the efficiency on boundary information possessed by mass region. We propose a feature vector based on boundary analysis to ameliorating three caracteristics like RDM, convexity and angular ones. We use the Digital Database for Screening Mammography DDSM for experiments. Some classifiers like Multilayer Perceptron MLP and k-Nearest Neighbors kNN are used to distinguish the pathological records from the healthy ones. Using MLP classifiers we obtained 94,2% as sensitivity(percentage of pathological ROIs correctly classified). The results in term of specificity (percentage of non-pathological ROIs correctly classified) grows around 97,9% using MLP classifier.

***Index Terms*—Medical Applications; Boundary; Shape analysis.**

I. Introduction

Breast cancer is considered as major health problem and constitutes the most common mortality causes 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 breast cancer treatment and the widespread adoption of mammography screening. In the past decade many research efforts attempts 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]. 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 techniques of analysis take two directions: analysis of shape which includes features based on morphological lesion [3] [4] and analysis of texture [5] [6] [7] [8]. Analysis of shape is decomposed into analysis of boundary and region. In the boundary analysis, many work applied methods based on Radial Distance Measure *RDM*, the convexity and angular measures. In this context, methods used the angular properties include two approaches: Radial Angle [9] [10] and Turning angle [11] [12]. In fact, breast tumors and masses appear in mammograms with different shape characteristics: malignant tumors usually have rough, spiculated, or microlobulated contours, whereas benign masses commonly have smooth, round, oval, or macrolobulated contours [13].

In this context, Sheng-Chih et al [10] used the Radial Angle measure. The Radial Angle is the smaller included angle between the direction of the gradient and the radial direction of the edge. It differentiates the malignant mass from benignant one by calculing the angle value: the mass is benignant if angle value approaches 180 and the average radial is large; otherwise, the mass is malignant.

Rangayyan et al [12] had demonstrated the relevance of Tuning angle technique in mammography context. The tuning angles (or tangent function) consists of computing, at each point of the contour, the angle value formed by the intersection between the tangent function and the horizontal one. Because of the larger number of points, the temporal complexity of the calculus increase. To optimize computation, the authors reduce the number of points by considering only the convex ones. Guliato et al [11] extended this work by normalizing the computation. In the others hand, 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*, 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), *sensitivity* (88.2 %) and *specificity* (92.3%). *AR* has presented the highest *sensitivity* (93%); its *specificity* 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 to represent the studied shape better than the characteristics cited above. Via these shape features, the authors attempts a 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, particularely, a vector based in *RDM*, convexity and an angular feature. In opposition to related work, we present some extensions of the classic *RDM* approach to cope with the amelioration of the analysis performance, we also present a characteristic based an angular calcul named Index Angle *IA* and we changed in calculing convexity features. 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*. To evaluate the performance of the efficiency in the classification of breast mass, we calculate the result of specificity and sensibility which using the DDSM base.
The rest of this paper is organized as follows: section II describes the proposed scheme. Section III illustrates the features extraction in boundary analysis. Section IV presents the results obtained in the proposed scheme. Finally, we draw conclusions and some future issues in section V.

II. The proposed flow in detecting breast cancer

The proposed scheme consists of three stages: identification of *ROI*, the features extraction, and the 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 on Sobel filter and seuillage is adopted. After the isolation of the *ROI*, the extraction of features is adopted in *ROI*: this is the flow stage in which we are interest in this paper. After this, a classification part is started. The features vector is entered to the classifier to make decision.
In the next paper, extraction of feature is adapted in ROI isolate.

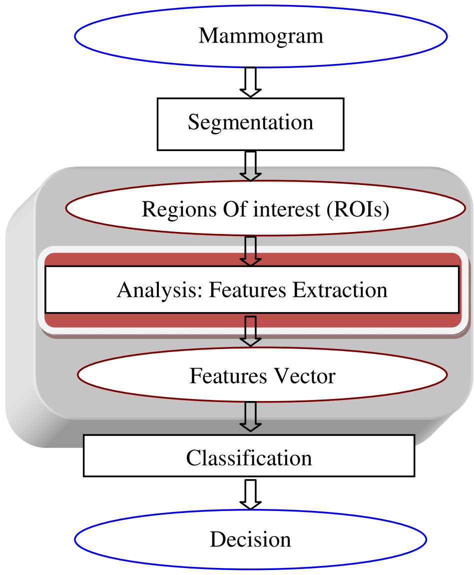

Fig. 1. Proposed Flow

III. Extraction of features

Boundary analysis which is often referred to as districting helps to define regions according to any criteria. We illustrate a method based on *RDM*, convexity and *IA*.

A. 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.

1) Radial Distance Measure: Radial Distance Measure *RDM* is a famous method used in shape analysis. However, an euclidian distances d(i) are calculated between the gravity center in region and all the points in boundary region (Figure 2).

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

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

$$dn(i) = \frac{d(i)}{max[d(i)]} \tag{2}$$

where n: the number of points (pixels) of the region boundary (the perimeter of region).

$$Xg = \frac{\sum_N X}{N} \tag{3}$$

$$Yg = \frac{\sum_N Y}{N} \tag{4}$$

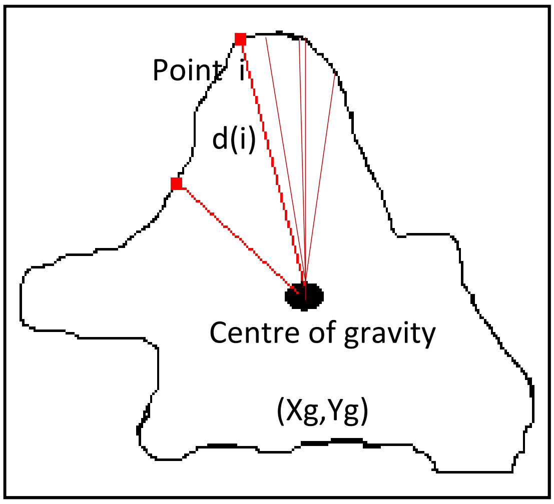

Fig. 2. Illustrative Figure of RDM

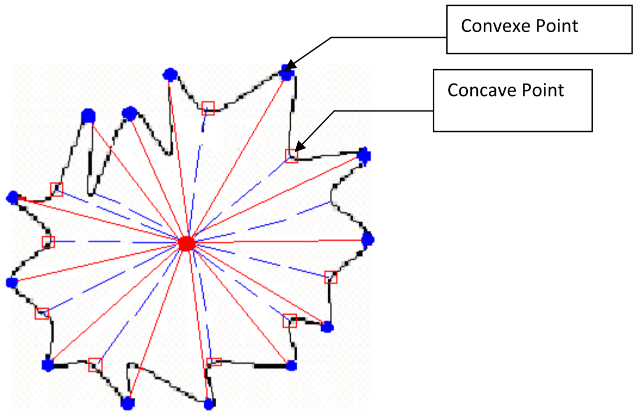

Fig. 3. Illustrative figure of eXtended RDM

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

$$dmoy = \frac{1}{n} \sum_{N} dn(i) \tag{5}$$

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.

$$SDEV = \sqrt{\frac{1}{N} \sum_{N} (dn(i) - dmoy)^2} \tag{6}$$

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

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

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

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

where Ar=0, if (dn(i)$\leq dmoy$)

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.

2) The extended Radial Distance Measure: To cure the problem of complexity, we propose to calculate the features cited in subsection III-A1 only in the concave and convex points (Figure 3). 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).

More formally:
Pconcave (i) = (i; $d(i) \leq d(i-1)$ and $d(i) \leq d(i+1)$)
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 in vovel *RDM* so called eXtended *RDM*. In the next section, we will show a feature based an angular analysis.

B. Index Angle

The Index Angle *IA* is the ratio of all the external angles by internal ones: the external is the angle between a central convex point (Convex point pi) and their "next least" convexes points (Convex point pi-1, p+1), however, the internal is the angle between a central convex point and their "next least" concaves points (Fig 4). The *IA* is applied to make a distinction between the shapes of edge of the mass as speculated or as round. When the mass tends to be more rounds, its *IA* tends to be near the 1. Conversely, a mass with speculated edge will have a *IA* smaller than 0.5. Then, the *IA* can be calculated as follows:

$$E = \frac{\sum_i \phi(i)}{\sum_i \theta(i)} \tag{9}$$

The concave and convex points are defined as follows subsection III-A2.

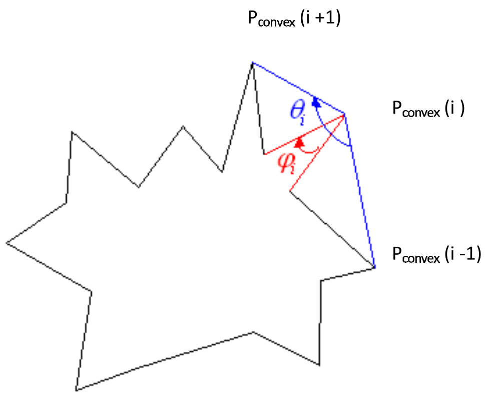

Fig. 4. An example of Index Angle

So, the *IA* is used only in the concave and convex point and not to any point to minimize the temporal complexity differently with the radial angle used in [9]. On the other hand, the advantage of this characteristic is that it is standardized and is invariant with any affine transformation.
In the next section, we will show another feature based in boundary analysis.

C. Convexity

In literature we used the convexity *CVX* which feature depend of region (ratio of Air of region detected by perimeter/or Air for his Convex envelop). In this study, we attempt to improve the importance of analysis with boundary shape. However, *CVX* is the ratio of perimeter of the region detected by perimeter for his Convex envelop. When the mass tend to be more round, its *CVX* tend to be near the 1. Conversely, a mass with speculated edge will have a *CVX* smaller than 0.5. Then the *CVX* can be calculated as follows:

$$E = \frac{Perimeter(Region)}{Perimeter(ConvexEnvelop)} \quad (10)$$

The advantage of this feature, is that it is standardized and it is invariant with any affine transformation.
Their five features represent the input of classifier. In the next section, we will show the performance of this shape vector in analysis of region in term of diagnosis relevance.

IV. Result and Discussion

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

- *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 numerisation techniques; namely pixel size, subtlety of cases, choice of training/testing subsets, etc.

A. The DDSM dataset:

The establishment of the *DDSM* allows the possibility of the common training and testing Dataset 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 evaluation we used the remaining 200 *ROIs* that contain 100 ground truthed abnormal region together with 100 entirely benign.

B. Experimental results:

The basic classification is based on the two methods of classification *KNN* and *MLP* as shown in Table 1. It represent the results from discriminant analysis in shape ”‘or boundary”’ vector.

classifier	KNN	MLP
sensitivity	93,67%	94,2 %
specificity	95,1%%	97,9%
Exactitude	94,37%	95,98%

TABLE I. Results from discriminant analysis with boundary vector

The table I illustrates the importance of boundary features in shape analyzing. The result in term of *sensitivity* tends towards 94,2% in *KNN* classifier. The result in term of *specificity* tends towards more than 97,9% in case of *MLP*.
By comparing thes results with other related works, we notice important ameliorations. In fact, in [14] Alvarenga et al obtained 88% in sensitivity and 90% in specificity. In heir experimental he classified a local base by LDA method. In the other hand, Rangayyan et Al [17] used the LDA classifier and his result attempt 95%. Differently, the result of Retico et al[15] using a *MLP* classifier can attempt 78,1 and 79,1 respectively in sensitivity and specificity. However, using a SVM classifier [18] Chang et al obtained 88,89% in sensitivity and 92,5% in specificity. But, both in [14] [17] [15] [18] we could not compare the feature vector because we did not use the same Database. In fact, the digitization can reflect the final result. However, we can assume that this feature vector 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.

V. Conclusion

In this work, we attempted to improve the classification performance of boundary shape method in analyzing the mammographic images. We presented some amelioration of features like the *RDM* method, the convexity and a new feature based an angular detail. To this method which gives better performance in term 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 97.9%. The results have been validated via two algorithms of classification: *kNN* and *MLP*. The results in term of *sensitivity* and *specificity* tends towards 91.9%. These results seem to be sufficient, and the future work is focused on the 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.

References

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

[2] R. A. Abdel-Dayem and R. M. El-Sakka, "Fuzzy entropy based detection of suspicious masses in digital mammogram," 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 clichs mammographiques en vue de la dtection des opacits du sein," in *Mathematics and Image Analysis*, Paris, septembre 2000.

[4] U. Cascio, D. Fauci, B. Golosio, R. Magro, L.Masala, P. Oliva, G. Raso, and S. Stumbo, "Transactions on engineering computing and technology," in *Massive Lesions Classification using Features based on Morphological Lesion Differences*, IMTC, March 2006, p. v12.

[5] 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.

[6] A. Boujelben, H. TMAR, J. M. NIF, and M. ABID, "Influence des caractristiques issues de la matrice de co-occurrence sur le diagnostic du cancer du sein," in *second Workshop AMINA, Application Medicale en Informatique: Nouvelle Approche*, Monastir, Tunisie, November:17-19: 2006, pp. 9–14.

[7] H. Sheshadri and A. Kandaswamy, "Breast tissue classification using statistical feature extraction of mammograms," *Indian J Med*, vol. 23 No.3, 2 June 2006.

[8] R. Zwiggelaar and E. .R.Denton, "Optimal segmentation of mammographic images," in *International Workshop on Digital Mammography*, Norwich, UK, 2004.

[9] R. Naga, Mudigonda, M. Rangaraj, Rangayyan, and J. E. L. Desaults, "Concavity and convexity analysis of mammographic masses via an iterative boundary segmentation algorithm," *In Proceedings of the 1999 IEEE Canadian Conference on Electrid and Computer Engineering*, pp. 9–12, May 1999.

[10] S.-C. Yang, C.-M. Wang, Y.-N. Chung, G.-C. Hsu, S.-K. Lee, P.-C. Chung, and C.-I. Chang, "A computer-aided system for mass detection and classification in digitized mammograms," *In Biomed Eng Appl Basis Comm*, pp. 215–228, October 2005.

[11] D. Guliato, M. Rangaraj, Rangayyan, D. J. Carvalho, and A. S. Santiago, "Polygonal modeling of contours of breast tumors with the preservation of spicules," *IEEE Transations In Biomedical Engineering*, vol. 55, January 2008.

[12] Rangayyan, M. Rangaraj, D. Guliato, J. D. Carvalho, and S. A. Santiago, "Feature extraction from the turning angle function for the classification of contours of breast tumors," *The International Special Topics Conference on Information Technology in Biomedicine IEEE-ITAB2006*, pp. 1–6, 2006.

[13] A. C. of Radiology and R. VA, "Illustrated breast imaging reporting and data system (bi-radstm)," p. third edition 1998.

[14] A. Alvarenga, A. Fernando, and C. I. W. C. de Albuquerque, "Assessing the performance of the normalised radial length and convex polygons in distinguishing breast tumours on ultrasound images," *Pereira Revista Brasileira de Engenharia Biomdica*, vol. 22, pp. 181–189, 2006.

[15] A. Retico, P. Delogu, M. Fantacci, and P. Kasae, "An automatic system to discriminate malignant from benign massive lesions on mammograms," *Medical Physics*, vol. 14, pp. 596–600, January 2007.

[16] P. Delogu, M. Fantacci, P. Kasae, and A. Retico, "Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier," *In Comput Biol Med*, 10 Mar 2008.

[17] R. M. Rangayyan, N. M. El-Faramawy, J. E. L. Desautels, and O. A. Alim, "Measures of acutance and shape for classification of breast tumors," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 799–810, 1997.

[18] R.F.Chang, W.Wu, W. Moon, and D.R.Chen, "Automatic ultrasound segmentation and morphology based diagnosis of solid breast tumors," *Springer*, vol. 89, p. 179185, 2005.