Proceedings of the 1999 IEEE Canadian Conference on Electrical and Computer Engineering
Shaw Conference Center, Edmonton, Alberta, Canada May 9-12 1999

Concavity and Convexity Analysis of Mammographic Masses via an Iterative Boundary Segmentation Algorithm

NAGA R. MUDIGONDA[1], RANGARAJ M. RANGAYYAN[1,2]
J. E. LEO DESAUTELS[3,2]

[1]Department of Electrical and Computer Engineering , University of Calgary,
2500 University Drive N.W., Calgary, Alberta, CANADA T2N 1N4 ranga@enel.ucalgary.ca
[2]Department of Radiology, University of Calgary, Calgary, Alberta, CANADA T2N 1N4
[3]Alberta Cancer Board, 120, 1040 7th Ave. S.W., Calgary, Alberta, CANADA T2P 3G9

Abstract

The problem of computer-aided classification of benign and malignant breast masses as seen on mammograms using morphological features is addressed in this paper. We propose methods of shape analysis treating the object's boundary in terms of local details. We use an iterative boundary segmentation method to separate major portions of the boundary and label them as concave or convex segments. In order to analyze the shape information localized in each segment, we compute features through polygonal modeling of the tumor boundaries. The features developed in the present study when used in combination with the global shape feature of compactness resulted in a classification accuracy of 81% with a database of 28 benign masses and 25 malignant tumors.

1 Introduction

Many studies have shown that early detection through periodic mammographic screening of asymptomatic women can reduce breast cancer mortality. Computer-based methods to accurately distinguish between benign and malignant diseases may help in performing an initial screening or a second reading of mammograms, and may lend objective tools to help the radiologist in analyzing difficult cases and deciding on biopsy recommendations. Computer detection of breast tumors is difficult due to the nature of mammographic images and features; in the present study we focus on the equally important problem of their classification as benign or malignant based on their morphology.

Breast abnormalities present varying diagnostic information on a mammogram. The diagnostic features vary in terms of shape, density, and textural contents. Most benign masses are homogeneous and possess fairly well-defined edges; malignant tumors typically have fuzzy or ill-defined boundaries. Benign masses possess smooth, round, or oval shapes with possible macrolobulations, as opposed to malignant tumors which typically exhibit rough contours with microlobulations, spiculations, and concavities. Many studies have therefore focused on analyzing the shapes of mammographic masses.

Kilday et al. [1] computed measures representing tumor circularity and radial distances from the centroid to the points on the boundary, and used the measures in a linear discriminant function for the classification of breast tumors. Their measures work well with a generally round boundary. However, in the case of complex shapes the centroid may lie outside the tumor region and may not be a valid point to measure the distances to the boundary. Ackerman et al. [2] analyzed breast lesions on xeroradiograms and investigated the use of four measures of malignancy: calcification, spiculation, roughness, and area-to-perimeter ratio. Brzakovic et al. [3] proposed an automated detection and Bayesian classification scheme on a data set of 25 tumors using the tumor size, shape, and intensity changes of extracted regions. Rangayyan et al. [4] used moments of distances of contour points from the centroid, compactness of the boundary, Fourier descriptors, and chord-length statistics to characterize the roughness of tumor boundaries. Other studies on detection and analysis of breast tumors have quantified features based on tumor edge sharpness [4] and textural information; these methods will not be reviewed here as the scope of the present work is limited to analyzing the shape of the mass contour or boundary.

Most of the shape analysis methods that have been applied to breast mass discrimination have focused on computing global measures characterizing the boundary's shape. Such methods are relatively insensitive to important local changes manifested by spicules and microlobulations. While the majority of benign masses on mammograms are well circumscribed, some do possess stellate or spiculated distortions. Discrimination between the microlobulations in malignant tumors and the macrolobulations of benign masses requires detailed analysis of local characteristics of mass boundaries. Recently, Menut et al. [5] performed parabolic modeling of tumor boundaries and used the

0-7803-5579-2/99/$10.00©1999 IEEE

narrowness and width of individual parabolic segments for classification. We use the same data set in the present study and propose a method of shape analysis that treats a mass boundary as a union of piecewise continuous and locally salient concave and convex parts. A convex part is defined as a segment of the boundary that encloses a portion of the mass, while a concave part is one formed by the presence of a background region within the segment. In order to identify and analyze the concavities and convexities possessed by a mass boundary in a localized manner, we propose methods based on segmentation and polygonal modeling. The segments are used to compute the angles subtended by the individual concave and convex segments at their vertices. This information is used in characterizing the narrowness and depth of spiculations or indentations.

The paper is organized as follows: Section II describes the segmentation and polygonal modeling methods. Section III presents the details on extraction of features. Section IV contains details of the database used. Section V presents the results of application of the proposed techniques to mammograms for benign versus malignant mass discrimination.

2 Methods

2.1 Concavity-convexity analysis

The first step in the proposed method is to identify and label the prominent concave and convex parts of a mass boundary. For this purpose, we segment the mass boundary into a set of piece-wise continuous curves by locating points of inflexion on the boundary.

Let $\vec{U_o} = (x(n), y(n)), n = 0, 1, 2, \ldots, N-1$, be the vector of the (x, y) coordinates of N points on the boundary of a mass. The points of inflexion on the boundary are defined by

$$\begin{aligned} \vec{U_o}' \times \vec{U_o}'' &= 0 \,, \\ \vec{U_o}' \times \vec{U_o}''' &\neq 0 \,, \end{aligned} \tag{1}$$

where $\vec{U_o}'$, $\vec{U_o}''$, and $\vec{U_o}'''$ are the first, second, and third derivatives of $\vec{U_o}$. Solving (1) is equivalent to solving the system of equations given by

$$\begin{aligned} x''(n)y'(n) - y''(n)x'(n) &= 0 \,, \\ x'(n)y'''(n) - x'''(n)y'(n) &\neq 0 \,, \end{aligned} \tag{2}$$

where $x'(n)$, $y'(n)$, $x''(n)$, $y''(n)$, $x'''(n), and\ y'''(n)$ are the first, second, and third derivatives of $x(n)$ and $y(n)$, respectively.

The portions of the contour between successive inflexion points were modeled as parabolas as in the work of Menut et al. [5]. Difficulty lies in the fact that contours of masses are, in general, not smooth. Many false or irrelevant inflexion points appear on relatively straight parts of a tumor boundary when $x''(n)$ and $y''(n)$ are not far from zero. To solve this problem, derivatives at each contour point were computed by considering weighted and averaged differences of a certain number of pairs of points on either side of the point under consideration. Variable numbers of pairs of points were used to compute derivatives that resulted in varying numbers of inflexion points for each contour. A hysterisis procedure was then applied on the resulting data to determine the optimum number of pairs of differences that provide the most appropriate inflexion points; details of this procedure may be found in Menut et al. [5].

After performing segmentation of the boundary as explained above, we label the individual segments between successive inflexion points as concave or convex. A convex part is defined as a segment of the boundary that encloses a portion of the mass, while a concave part is one formed by the presence of a background region within the segment. Fig. 1 shows a 450×650 pixel section of a mammogram (pixel size = 50 μm) with a circumscribed benign mass overlaid with the contour, the black and white parts representing the concave and convex parts, respectively. Fig. 2 shows the result of concavity/convexity labeling of the boundary of a spiculated malignant tumor (a 600×770 pixel section of a mammogram with pixel size = 62 μm). Once the boundary of a mass is identified in terms of its concave and convex parts, we propose to compute fractional concavity and fractional convexity values to use as features in further pattern classification. The boundaries used in the present work were manually drawn, and include artifactual, minor modulations that could lead to inefficient representation for pattern classification. We propose a polygonal model in order to extract features that can effectively characterize the shape complexities represented by the individual concave and convex segments identified as above.

2.2 Polygonal modeling of a mass boundary

Ventura and Chen [6] presented an algorithm for segmenting two-dimensional curves in which the number of segments is to be prespecified for initiating the process, in relation to the complexity of the shape. This is not a desirable step when dealing with complex or spiculated shapes of breast tumors [4]. We use the segments obtained via the points of inflexion as the initial input to the polygonal modeling procedure. This step helps in automating the polygonalization algorithm: the proposed method does not require any interaction from the user in terms of the starting number of segments.

The first step in the polygonal modeling algorithm

0-7803-5579-2/99/$10.00 © 1999 IEEE

Figure 1: Concave and convex parts of the boundary of a benign mass. The concave parts are shown in black and the convex parts in white.

Figure 2: Concave and convex parts of the boundary of a spiculated malignant tumor. The concave parts are shown in black and the convex parts in white.

is to represent each segmented curve by a pair of linear segments based on its arc-to-chord deviation. The procedure is iterated subject to predefined boundary conditions so as to minimize the error between the length of the initial boundary and the cumulative length computed from the polygonal segments.

For the application to mammographic mass boundary analysis, we have developed the following set of conditions:

1. If the arc-to-chord deviation at a contour pixel exceeds 0.25 mm (5 pixels in our images with a pixel size of 50 μm), the curve is segmented at that point irrespective of the length of the resulting segments.

2. If the arc-to-chord length deviation at a contour pixel is less than 0.25 mm and greater than 0.1 mm, the curve is segmented at that point subject to the condition that the resulting segments satisfy a minimum-length criterion, which is chosen as 1 mm.

The criterion for choosing the threshold for arc-to-chord deviation is based on the assumption that any segment possessing a lesser deviation is insignificant in terms of its degree of spiculation for further analysis (as indicated by the radiologist involved in this work).

The result of application of the polygonal modeling algorithm to the circumscribed benign mass boundary in Fig. 1 is shown in Fig. 3. Fig. 4 shows the result for the spiculated malignant case of Fig. 2. Fig. 5 shows the convergence plot of the algorithm for the malignant case. The number of segments required for the approximation of a boundary increased with its shape complexity (in the range 20-400 for the database used). However, the number of iterations required for the convergence of the algorithm did not vary much for different mass boundary shapes (remaining within the range of 3-5). This is due to the fact that the relative complexity of the boundary to be segmented is taken into consideration during the initial pre-processing step of locating the points of inflexion; hence the subsequent polygonalization process is robust and computationally efficient. The algorithm performed well and delivered satisfactory results on various irregular shapes of spiculated cases of benign and malignant masses in our database. The techniques proposed can successfully divide a tumor's boundary into concave and convex parts and further approximate each part by a set of linear segments leading to a polygonal model of the whole boundary.

3 Feature extraction

The features computed from the individual segments of a mass boundary consist of measures representing the concave or convex fraction of the total boundary length and an index of spicularity.

0-7803-5579-2/99/$10.00 © 1999 IEEE

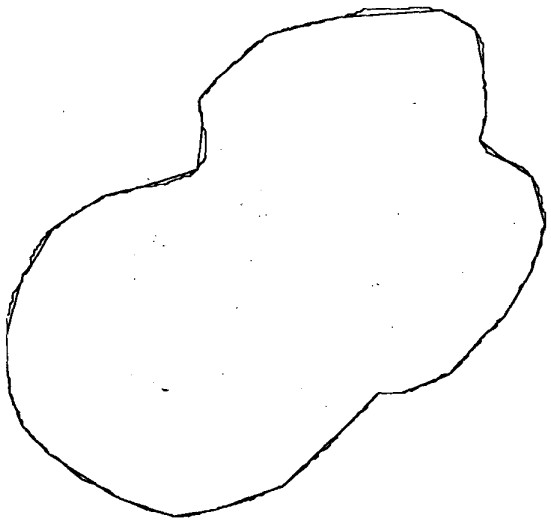

Figure 3: Polygonal model of the boundary of the benign mass shown in Fig. 1. Number of sides = 36.

Figure 4: Polygonal model of the boundary of the spiculated malignant tumor shown in Fig. 2. Number of sides = 146.

3.1 Concavity and convexity fractions

Most benign mass boundaries have major portions of convex macrolobulations. Some benign masses may have minor concavities and spicules. On the other hand, malignant tumors typically possess both concave and convex segments as well as microlobulations and prominent spicules. We compute the fractional concavity and fractional convexity values of each mass boundary in order to characterize and quantify these properties.

Let $S_i, i = 1, 2, 3, \ldots, M$, be the lengths of M segments obtained from the polygonal model of a mass boundary. The total length of the boundary T_l is computed as the cumulative length of the M segments:

$$T_l = \sum_{i=1}^{M} S_i. \tag{3}$$

Let $CC_i, i = 1, 2, 3, \ldots, P$, be the respective lengths of P concave segments, and $CV_i, i = 1, 2, 3, \ldots, Q$, be the lengths of Q convex segments. The cumulative concave length (CC_l) and convex length (CV_l) are computed as

$$CC_l = \sum_{i=1}^{P} CC_i\ , \tag{4}$$

$$CV_l = \sum_{i=1}^{Q} CV_i\ . \tag{5}$$

The fractional concavity (f_{cc}) and convexity (f_{cv}) values are computed by normalizing the respective cumulative lengths by the total boundary length so that the sum of the concavity and convexity fractions for a given boundary equals unity:

$$f_{cc} = \frac{CC_l}{T_l}\ , \tag{6}$$

$$f_{cv} = \frac{CV_l}{T_l}\ . \tag{7}$$

Since the lengths of the segments and the total length of the boundary are computed using a polygonal model of the boundary, the resulting parameters are not sensitive to noise and artifactual modulations present in the boundary; they are also independent of translation and rotation of the boundary. Moreover, the fractional parameters are normalized with respect to the total length of the boundary, thus making them independent of the size of the boundary.

3.2 Spiculation index

It is known that invasive carcinomas, by virtue of their infiltrating nature into the surrounding tissues, form narrow, stellate distortions at their boundaries. Based on this feature, we propose a Spiculation Index *(SI)* to represent the degree of spiculation of a mass boundary. In order to emphasize narrow spicules and microlobulations, we apply a function that enhances the

0-7803-5579-2/99/$10.00©1999 IEEE

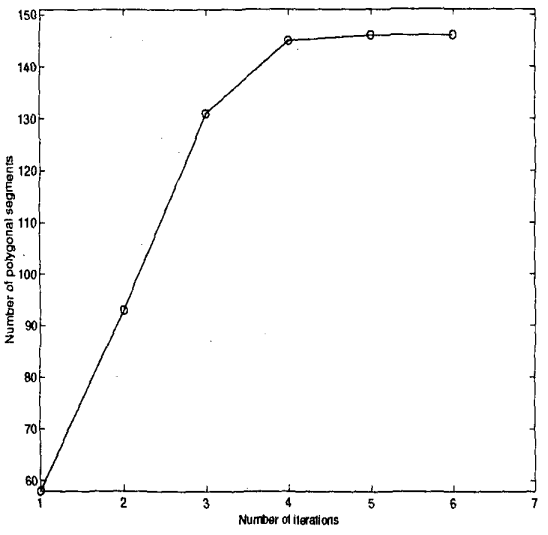

Figure 5: Convergence plot of the iterative polygonal modeling procedure for the spiculated malignant tumor in Fig. 2.

contribution of narrow spicules in the computation of SI. For each curved segment of a mass boundary or the corresponding polygonal model segment, the ratio of its length to the base width can represent its degree of narrowness or spiculation. We propose a nonlinear weighting function based upon the segment's length and angle of spiculation that delivers progressively increasing weighting with increase in the narrowness of spiculation of the segment.

Let SL_i and $\theta_i, i = 1, 2, 3, \ldots, N$, be the length and angle of N sets of polygonal model segments of a mass boundary. Then, SI is computed as

$$SI = \frac{\sum_{i=1}^{N}(1+\cos\theta_i)SL_i}{\sum_{i=1}^{N} SL_i}. \quad (8)$$

The factor $(1+\cos\theta_i)$ in Eqn. 8 modulates the length of each segment (possible spicule) according to its narrowness. Spicules with narrow angles between 0^o and 30^o get maximum weighting, as compared to macrolobulations which usually form obtuse angles and hence get lesser weighting. The majority of the spicules in the mass boundaries of our database were found to have angles ranging between 30^o and 150^o: the function $(1+\cos\theta)$ is progressively decreasing within this range, giving reduced weighting to segments with larger angles. Relatively flat segments having angles ranging between 150^o and 180^o receive the least weighting, and hence are treated as least significant.

The denominator in Eqn. 8 serves as a normalization factor to take into account the effect of the size of the contour. It ensures that the spiculation index represents only the severity of the spiculated nature of the mass boundary, which in turn may be linked to the invasive properties of the mass under consideration. Thus circumscribed mass shapes should have lower SI values and sharp, stellate shapes with acute spicules should have higher SI values.

3.3 Compactness

In order to compare the performance of the proposed features with a global shape measure, we computed the compactness of each tumor boundary. Compactness is a simple measure of the efficiency of a contour to contain a given area, and is commonly defined as $C = P^2/A$, where P and A are the contour perimeter and area, respectively. A malignant tumor with a number of concavities or spicules could be expected to possess a higher compactness value than a smooth and round benign mass. In order to restrict the range of compactness to $(0, 1)$ and to obtain increasing values of C with increase in complexity of the shape, we use a modified expression of compactness given by [7]

$$C = 1 - \frac{4\pi A}{P^2}. \quad (9)$$

With this expression, the compactness of a circle will be zero; C increases with the complexity of the contour to a maximum value of 1.

4 Image Database

Thirty-eight mammographic images including 16 circumscribed benign, four circumscribed malignant, 12 spiculated benign, and six spiculated malignant masses were selected from the Mammographic Image Analysis Society (MIAS, UK) database. The spatial resolution of the images is 50 $\mu m \times 50$ μm. Although most of the malignant tumors encountered in mammography are spiculated and a majority of the benign masses are well-circumscribed, the MIAS database has a relatively large number of spiculated benign cases. In order to augment the numbers of the two types of malignant tumors, 15 images from Screen Test: Alberta Program for the Early Detection of Breast Cancer, digitized at a spatial resolution of 62 $\mu m \times 62$ μm, were added to the study, increasing the number of cases examined to 53, with 28 benign and 25 malignant cases. Sections of interest of the mammographic images were displayed on a Sun SPARC Station 2 and the boundary of each mass was traced and input to the computer by an expert radiologist specialized in mammography (JELD) using $XPAINT$ for $X - Windows$ [4].

0-7803-5579-2/99/$10.00 © 1999 IEEE

5 Results of Pattern Classification

For each mass or tumor in the database we computed compactness, fractional concavity, fractional convexity, and the spiculation index. All of the features, individually and in different combinations, could effectively discriminate circumscribed benign masses from spiculated malignant tumors. Compactness failed to classify almost all spiculated benign and circumscribed malignant masses. Fractional concavity-convexity analysis could distinguish two out of four circumscribed malignant cases correctly; it however failed to distinguish between spiculated benign masses and spiculated malignant tumors.

Since the degree of spiculation of boundary segments is emphasized characterized by the spiculation index, a significant portion of the spiculated benign cases (5 out of 12) with large convexities and a few narrow spicules were correctly classified as benign. This is an encouraging result with the present methods as none of the other parameters developed by us in the past could be effective in separating the spiculated benign cases in our database. Spiculation index, however, failed in three of the four circumscribed malignant cases, although with narrow margins. These cases, although malignant, do not have prominent spicules in their boundaries.

To further validate our results, we used the computed features in the BMDP 7M step-wise discriminant analysis program [8] to perform pattern classification. The program realizes a jack-knife validation procedure using the leave-one-out algorithm. We obtained the best benign versus malignant classification accuracy of 81% by combining the global shape feature compactness with fractional concavity/convexity and SI, which are sensitive to local variations in the boundary. Compactness, fractional concavity or convexity, and SI individually resulted in classification accuracies of 72%, 74%, and 79%, respectively. Given the fact that the database we have used includes an unusual proportion of spiculated benign cases (12 out of 53), the performance of the proposed features could be regarded to be very good. Addition of other features based upon density variations [4] and textural information may result in improved benign versus malignant discrimination.

6 Conclusion

We have proposed new methods of shape analysis to analyze and classify breast tumors. The methods include boundary segmentation algorithms to identify major concave and convex portions of the boundary, and features that are sensitive to local variations in the boundary computed through a polygonal modeling algorithm. The shape measures achieved a benign versus malignant classification accuracy of 81%, an improvement of 6% over other shape features applied to the same database of 53 cases [4, 5].

Acknowledgments

This project is supported by grants from the Alberta Heritage Foundation for Medical Research (AHFMR), the Natural Sciences and Engineering Research Council (NSERC) of Canada, and the Whitaker Foundation.

References

[1] Kilday J, Palmieri F, and Fox MD. Classifying mammographic lesions using computerized image analysis. *IEEE Transactions on Medical Imaging*, 12(4):664–669, December 1993.

[2] Ackerman LV and Gose E. Breast lesion classification by computer and xeroradiograph. *Cancer*, 30:1025–1035, 1972.

[3] Brzakovic D, Luo XM, and Brzakovic P. An approach to automated detection of tumors in mammograms. *IEEE Transactions on Medical Imaging*, 9(3):233–241, 1990.

[4] Rangayyan RM, El-Faramawy NM, Desautels JEL, and Alim OA. Measures of acutance and shape for classification of breast tumors. *IEEE Transactions on Medical Imaging*, 16(6), 1997.

[5] Menut O, Rangayyan RM, and Desautels JEL. Parabolic modeling and classification of breast tumors. *International Journal of Shape Modeling*, 3(3 & 4):155–166, 1998.

[6] Ventura JA and Chen JM. Segmentation of two-dimensional curve contours. *Pattern Recognition*, 25(10):1129–1140, 1992.

[7] Shen L, Rangayyan RM, and Desautels JEL. Detection and classification of mammographic calcifications. *International Journal of Pattern Recognition and Artificial Intelligence*, 7(6):1403–1416, 1993.

[8] Brown MB and Engelman L. *BMDP Statistical Software Manual*. University of California, Berkeley, CA, 1988.