Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Spiculation-preserving Polygonal Modeling of Contours of Breast Tumors

Denise Guliato*, Rangaraj M. Rangayyan, Juliano Daloia de Carvalho, and Sérgio Anchieta Santiago

Abstract— **Malignant breast tumors typically appear in mammograms with rough, spiculated, or microlobulated contours, whereas most benign masses have smooth, round, oval, or macrolobulated contours. Several studies have shown that shape factors that incorporate differences as above can provide high accuracies in distinguishing between malignant tumors and benign masses based upon their contours only. However, global measures of roughness, such as compactness, are less effective than specially designed features based upon spicularity and concavity. We propose a method to derive polygonal models of contours that preserve spicules and details of diagnostic importance. We show that an index of spiculation derived from the turning functions of the polygonal models obtained by the proposed method yields better classification accuracy than a similar measure derived using a previously published method. The methods were tested with a set of 111 contours of 65 benign masses and 46 malignant tumors. A high classification accuracy of 0.93 in terms of the area under the receiver operating characteristics curve was obtained.**

I. CHARACTERISTICS OF BREAST TUMORS

Breast tumors and masses usually appear in the form of dense regions in mammograms. A typical benign mass has a round, smooth, and well-circumscribed boundary; on the other hand, a malignant tumor usually has a spiculated, rough, and blurry boundary [1], [2]. There do exist atypical cases of macrolobulated or spiculated benign masses, as well as microlobulated or well-circumscribed malignant tumors [3], [4]. On the basis of the shape differences between benign masses and malignant tumors, objective features of shape complexity such as compactness (C), fractional concavity (F_{cc}), spiculation index (SI), a Fourier-descriptor-based factor (FF), moments, chord-length statistics, and wavelet transform modulus-maxima have been developed for pattern classification [3], [4], [5], [6], [7], [8].

Although textural differences have been noted between benign masses and malignant tumors, they do not appear to be significant or consistent [3], [6], [9], [10]. Sahiner et al. [6] and Alto et al. [10] evaluated many combinations of shape, edge-sharpness, and texture measures to classify breast masses; shape factors such as F_{cc}, FF, and SI provided much higher classification accuracies than measures related to texture and density characteristics.

In spite of the established importance of shape factors in the analysis of breast tumors and masses, difficulties exist in obtaining accurate and artifact-free boundaries of the related regions from mammograms. Automatic detection and segmentation of breast masses is difficult [6], [9], [11]. The studies of Rangayyan et al. [3], [4] and Alto et al. [10] were based on contours of masses drawn manually on mammograms by a radiologist. Sahiner et al. [6] evaluated the performance of several features with automatically detected boundaries of masses, and found FF to be the most discriminant feature. Whereas manually drawn contours could contain artifacts related to hand tremor and are subject to intra-observer and inter-observer variations, automatically detected contours could contain noise and inaccuracies due to limitations or errors in the procedures for the detection and segmentation of the related regions.

A. Global versus local analysis of the shape of breast masses

Most of the shape analysis methods that have been applied for the discrimination of breast masses have focused on computing global measures to characterize the shape complexity of their boundaries. Two commonly used global measures of shape roughness are compactness and the ratio of the major axis to the minor axis of the bounding ellipse. Such methods are relatively insensitive to diagnostic relevant local changes manifested in terms of spicules or microlobulations. Few works presented in the literature explicitly take into account these characteristics [4], [12].

Menut et al. [12] proposed a method to fit each piecewise-continuous part of a given boundary of a tumor with a parabolic model. Contours of benign masses were typically segmented into a few wide parabolas and several small and flat sections due to smooth boundaries and large lobulations. On the other hand, contours of malignant tumors were typically modeled with a large number of narrow parabolas and few flat sections. The parameters of the parabolic segments were used for the classification of masses.

Rangayyan et al. [3], [4] used a modified version of the method of Ventura and Chen [13] to fit polygonal models to hand-drawn contours of breast masses and derive shape factors. Rangayyan et al. [4] proposed methods to obtain indices of spiculation (SI) and concavity (F_{cc}) based on polygonal models of contours of breast masses. The polygonal modeling procedure was found to aid in reducing noise and artifacts in the hand-drawn contours.

D. Guliato, J. D. de Carvalho, and S. A. Santiago are with the Faculty of Computation, Federal University of Uberlândia, Uberlândia, MG, Brazil.

* To whom all correspondence should be addressed. guliato@ufu.br

R. M. Rangayyan is with the Department of Electrical and Computer Engineering, Schulich School of Engineering, and Department of Radiology, University of Calgary, Calgary, AB, Canada T2N 1N4. ranga@ucalgary.ca

This work was supported by the Conselho Nacional Desenvolvimento Científico e Tecnológico of Brazil, and a Catalyst grant from Research Services, University of Calgary. We thank Fábio José Ayres, University of Calgary, for assistance with the ROC procedures.

The objective of this paper is to present a polygonal modeling method that reduces the influence of noise and artifacts while preserving the diagnostically relevant features, in particular the spicules and lobulations, in the original contour. The generation of the model is controlled by conditions related to the lengths of the sides of the polygon as well as its angles. To evaluate the performance of the modeling procedure in terms of the efficiency in the classification of breast masses, we derive an index of spiculation from the model and compare the results with those provided by the method of Rangayyan et al. [4] to obtain SI.

II. DATASETS OF CONTOURS OF MASSES

The dataset of contours of breast masses used in this study includes contours obtained in two preceding studies. One set of contours was derived from mammograms of 20 cases obtained from Screen Test: the Alberta Program for the Early Detection of Breast Cancer [14], [10], [15]. The mammograms were digitized using the Lumiscan 85 scanner at a resolution of 50 μm with 12 b/ pixel. The set includes 57 regions of interest (ROIs), of which 37 are related to benign masses and 20 are related to malignant tumors [10]. Most of the benign masses in this dataset are smooth or macrolobulated, whereas most of the malignant tumors are spiculated or microlobulated.

Another set of images was obtained from the Mammographic Image Analysis Society (MIAS, UK) database [16], [17] and the teaching library of the Foothills Hospital (Calgary) [4], [3]. The MIAS images were digitized at a resolution of 50 μm; the Foothills Hospital images were digitized at a resolution of 62 μm. This set includes smooth, lobulated, and spiculated contours in both the benign (28) and malignant (26) categories.

The contour of each mass was manually drawn by an expert radiologist specialized in mammography. The combined dataset has 111 contours, including both typical and atypical shapes of benign masses (65) and malignant tumors (46). The diagnostic classification was based upon biopsy.

III. METHODS: POLYGONAL MODELING

Polygonal modeling is used to approximate a closed contour with linear segments; segmentation and modeling of planar curves has been studied extensively [18], [19], [20], [4], [13], [12], [21]. Pavlidis and Ali [19] and Pavlidis and Horowitz [20] proposed methods for segmentation and polygonal modeling for the representation of electrocardiograms (ECGs), contours of cells, and handwritten numerals. In general, modeling methods use, as approximation criteria, the minimization of the error between the given contour and the model, the minimal polygon perimeter, the maximal internal polygon area, or the minimal area external to the polygon but contained by the given contour.

The polygonal modeling procedure proposed by Rangayyan et al. [4] starts by dividing the given boundary (a closed contour) into a set of piecewise-continuous curved parts by locating the points of inflection on the boundary based on its first, second, and third derivatives. Each segmented curved part is represented by a pair of linear segments with a vertex at the point of maximal arc-to-chord distance. The procedure is iterated subject to predefined boundary conditions so as to minimize the error between the perimeter of the initial contour provided and the perimeter of the polygon. The criterion for choosing the threshold on the arc-to-chord distance is based on the assumption that any segment that possesses a smaller distance is insignificant in terms of the accuracy of the contour for further analysis. Specifically, the maximal arc-to-chord distance permitted was 0.25 mm or 5 pixels (at a pixel resolution of 50 μm), and the smallest side of the polygon permitted was 1.0 mm. The procedure as above eliminates certain disadvantages of the method of Ventura and Chen [13], such as the requirement to provide the number of sides of the desired polygonal model. However, it retains the points of inflection, thereby constraining the fit of the model to the contour provided; furthermore, the criteria used do not specifically relate to the notion of preserving spicules.

A. Spiculation-preserving polygonal modeling

The proposed method starts by identifying all of the linear segments of the given contour. (Some of the segments could be as short as 2 pixels.) Let M, M_i, and N be the number of the points of the given contour, the number of points in the ith linear segment, and the number of the linear segments of the contour, respectively. Then, the original contour is given by $S = \{(x_j, y_j)\}|\ j = 1, 2, 3, \ldots, M$ that is partitioned into N linear segments $S_i = \{(x_{ij}, y_{ij})\}|j = 1, 2, 3, \ldots, M_i$, $i = 1, 2, 3, \ldots, N$, where $M = M_1 + M_2 + \ldots + M_N$ and $S_k \cap S_l = \emptyset\ \forall(k, l),\ k \neq l$.

The next step is to reduce the influence of noise while maintaining the diagnostically relevant characteristics of the given contour, and attempting to reduce in each iteration of the algorithm the value of N as well as to increase the value of M_i, $i = 1, 2, 3, \ldots, N$. The algorithm to obtain the polygonal model executes two rules for every linear segment in each iteration:

Rule 1: if two adjacent segments S_i and S_{i+1} are shorter than a threshold $S_{\min}$, then join S_i and S_{i+1}.

Rule 2: if S_i or S_{i+1} is greater than the threshold $S_{\min}$, then analyze the angle between S_i and S_{i+1}:
if the angle is greater than a given threshold $\theta_{\max}$, then join S_i and S_{i+1}, else retain S_i and S_{i+1}.

The threshold $S_{\min}$ represents the relevance of a segment and $\theta_{\max}$ indicates the relevance of the angle between two adjacent linear segments being analyzed. The angle used is always the smaller of the two angles between the adjacent sides being analyzed, which could be the related internal or the external angle of the polygon. The algorithm stops when no two linear segments are joined in an iteration. Fig. 1 and Fig. 2 illustrate the results obtained by the proposed method superimposed on the original contours of a benign mass and a malignant tumor, respectively. Note that, in the case of the malignant tumor, all of the spicules and lobulations have been preserved. On the other hand, in the case of the benign mass, the convex nature of the contour has been preserved

(and made more obvious). The captions of the figures also provide the number of vertices in the polygonal models N_p and the number of points in the original contours N_c, which indicate the data compression provided by the models.

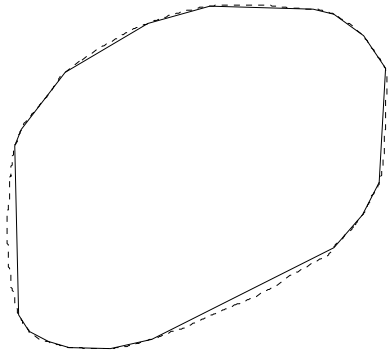

Fig. 1. Polygonal modeling of a benign mass with a relatively smooth contour. $S_{\min} = 10$ pixels. $\theta_{\max} = 170^o$. $N_p = 18$ and $N_c = 936$.

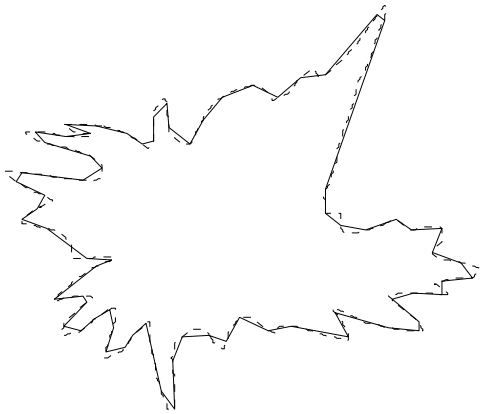

Fig. 2. Polygonal modeling of a malignant tumor with a spiculated contour. $S_{\min} = 10$ pixels. $\theta_{\max} = 170^o$. $N_p = 72$ and $N_c = 1,318$.

B. Derivation of an index of spiculation

In order to derive an index of spiculation or lobulation from the polygonal model obtained as above, we use the turning function of the polygon. The turning function $T_S(S_i)$ at the segment S_i of a contour (polygon) S is the angle, measured in the counterclockwise direction, of the segment S_i with reference to the x axis, expressed as a function of the length of the segment. Fig. 3 and Fig. 4 show the turning functions for the polygonal models shown in Fig. 1 and Fig. 2. For a convex contour, as the case in Fig. 1, the turning function is a monotonically increasing function; see Fig. 3. For a contour that presents concave and convex regions, the turning function begins to decrease at the beginning of a concave portion, and keeps on decreasing until the direction of a segment of the contour changes at the beginning of the next convex portion. In the turning function in Fig. 4, a spicule in the contour is related to a portion bounded by a pair of successive significant drops in the angle.

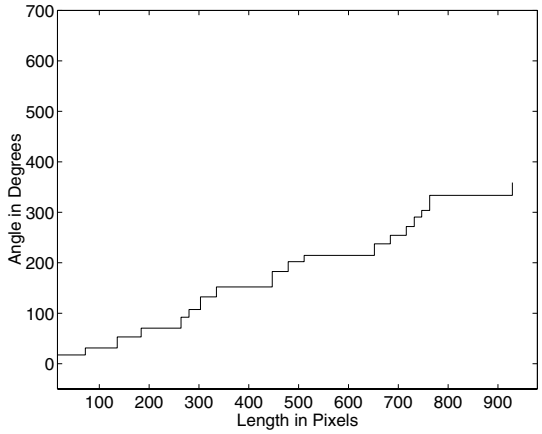

Fig. 3. Turning function of the polygonal model for the benign mass in Fig. 1.

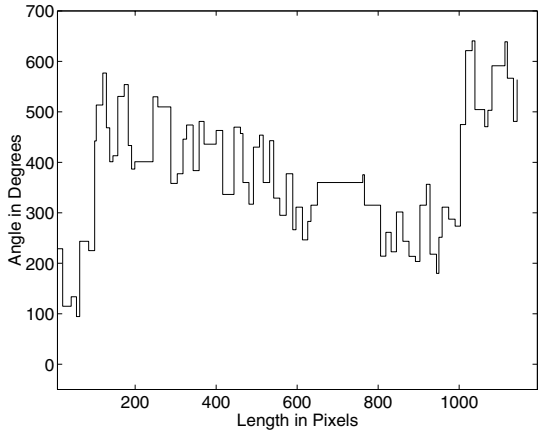

Fig. 4. Turning function of the polygonal model for the malignant tumor in Fig. 2.

The turning function may be used as a signature to represent the shape of the original contour (or its polygonal model) [21], [18]. In the following, we describe a procedure to derive an index of spiculation from the turning function.

To obtain a spiculation index from the turning function (SI_{TF}), the length of each possible spicule is multiplied by $(1 + \cos\psi)$, where ψ is the angle of the spicule (as obtained from the turning function). The weighted lengths of the spicules are summed, and normalized by twice the sum of their unweighted lengths.

IV. RESULTS AND DISCUSSION

The proposed methods were applied to a set of 111 contours of breast masses (see Section II). A sliding threshold was applied to classify the features, and receiver operating characteristics (ROC) curves [22] were generated. The area A_z under each ROC curve was computed to serve as a measure of the classification performance of the corresponding feature. Table I lists the values of A_z obtained by the

TABLE I

COMPARISON OF CLASSIFICATION PERFORMANCE OF INDICES OF SPICULATION. A_z: AREA UNDER THE ROC CURVE. N_p AND N_c ARE THE NUMBERS OF POINTS IN THE POLYGONAL MODEL AND THE ORIGINAL CONTOUR, RESPECTIVELY.

Method	A_z	Average N_p/N_c
SI of Rangayyan et al. [4]	0.91	0.0386
SI_{TF} of the present paper using:		
(a) $S_{\min} = 5$ pixels, $\theta_{\max} = 150^o$	0.85	0.0226
(b) $S_{\min} = 5$ pixels, $\theta_{\max} = 160^o$	0.88	0.0339
(c) $S_{\min} = 5$ pixels, $\theta_{\max} = 170^o$	0.93	0.0806
(d) $S_{\min} = 10$ pixels, $\theta_{\max} = 150^o$	0.86	0.0179
(e) $S_{\min} = 10$ pixels, $\theta_{\max} = 160^o$	0.91	0.0237
(f) $S_{\min} = 10$ pixels, $\theta_{\max} = 170^o$	0.93	0.0365
(g) $S_{\min} = 20$ pixels, $\theta_{\max} = 150^o$	0.89	0.0130
(h) $S_{\min} = 20$ pixels, $\theta_{\max} = 160^o$	0.90	0.0162
(i) $S_{\min} = 20$ pixels, $\theta_{\max} = 170^o$	0.91	0.0219

proposed method for various parameters. Also listed in the table is the ratio of the number of vertices in the polygonal models N_p to the number of points in the original contours N_c averaged over all of the 111 contours; this measure indicates the data compression provided by the modeling methods as a side benefit. It is seen that the proposed method provides polygonal models that have, on the average, values of the ratio N_p/N_c lower than that obtained by the method of Rangayyan et al. [4] (except for the case with $S_{\min} = 5$ pixels, $\theta_{\max} = 170^o$). The index of spiculation SI_{TF} derived from the turning function obtained using the proposed methods has provided a higher classification accuracy than the similar feature of spiculation index SI as defined by Rangayyan et al. [4], with the parameters ($S_{\min} = 5$ pixels, $\theta_{\max} = 170^o$) or ($S_{\min} = 10$ pixels, $\theta_{\max} = 170^o$). (For an image digitalized at a resolution of 50 μm, the value of $S_{\min}$ set at 10 represents a segment length around 0.5 mm).

V. CONCLUSION

We have proposed an efficient method to obtain polygonal models of contours. The procedure is useful in modeling contours of a vast range of applications. The values of $S_{\min}$ and $\theta_{\max}$ depend on the features that have to be preserved in the polygonal modeling, for further analysis. In particular, when applied to model contours of breast masses and tumors, the proposed polygonal modeling method preserved all diagnostically important features related to spicules and lobulations. In addition, we have proposed an index of spiculation derived from the turning function of a polygonal model; the index has provided a high classification accuracy in discriminating between benign breast masses and malignant tumors. The methods should be useful in computer-aided diagnosis of breast cancer.

REFERENCES

[1] Homer MJ. *Mammographic Interpretation: A Practical Approach*. McGraw-Hill, Boston, MA, 2nd edition, 1997.

[2] American College of Radiology, Reston, VA. *Illustrated Breast Imaging Reporting and Data System (BI-RADS™)*, third edition, 1998.

[3] 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):799–810, 1997.

[4] Rangayyan RM, Mudigonda NR, and Desautels JEL. Boundary modelling and shape analysis methods for classification of mammographic masses. *Medical and Biological Engineering and Computing*, 38:487–496, 2000.

[5] Bruce LM and Adhami RR. Classifying mammographic mass shapes using the wavelet transform modulus-maxima method. *IEEE Transactions on Medical Imaging*, 18(12):1170–1177, 1999.

[6] Sahiner BS, Chan HP, Petrick N, Helvie MA, and Hadjiiski LM. Improvement of mammographic mass characterization using spiculation measures and morphological features. *Medical Physics*, 28(7):1455–1465, 2001.

[7] Huo Z, Giger ML, Vyborny CJ, Wolverton DE, and Metz CE. Computerized classification of benign and malignant masses on digitized mammograms: A study of robustness. *Academic Radiology*, 7(12):1077–1084, 2000.

[8] Huo Z, Giger ML, and Vyborny CJ. Computerized analysis of multiple-mammographic views: Potential usefulness of special view mammograms in computer-aided diagnosis. *IEEE Transactions on Medical Imaging*, 20(12):1285–1292, 2001.

[9] Mudigonda NR, Rangayyan RM, and Desautels JEL. Detection of breast masses in mammograms by density slicing and texture flow-field analysis. *IEEE Transactions on Medical Imaging*, 20(12):1215–1227, 2001.

[10] Alto H, Rangayyan RM, and Desautels JEL. Content-based retrieval and analysis of mammographic masses. *Journal of Electronic Imaging*, 14(2):023016:1–17, 2005.

[11] Wei D, Chan HP, Helvie MA, Sahiner B, Petrick N, Adler DD, and Goodsitt MM. Classification of mass and normal breast tissue on digital mammograms: multiresolution texture ananlysis. *Medical Physics*, 22(9):1501–1513, 1995.

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

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

[14] Alberta Cancer Board, Alberta, Canada, www. cancerboard. ab.ca/ screentest. *Screen Test: Alberta Program for the Early Detection of Breast Cancer – 2001/03 Biennial Report*, 2004.

[15] Alto H, Rangayyan RM, Paranjape RB, Desautels JEL, and Bryant H. An indexed atlas of digital mammograms for computer-aided diagnosis of breast cancer. *Annales des Télécommunications*, 58(5-6):820–835, 2003.

[16] The Mammographic Image Analysis Society digital mammogram database. http://www.wiau.man.ac.uk/ services/MIAS/ MIASweb.html, accessed June, 2004.

[17] Suckling J, Parker J, Dance DR, Astley S, Hutt I, Boggis CRM, Ricketts I, Stamatakis E, Cerneaz N, Kok SL, Taylor P, Betal D, and Savage J. The Mammographic Image Analysis Society digital mammogram database. In Gale AG, Astley SM, Dance DR, and Cairns AY, editors, *Proceedings of the 2nd International Workshop on Digital Mammography*, volume 1069 of *Excerpta Medica International Congress Series*, pages 375–378, York, England, July 1994.

[18] Latecki LJ and Lakämper R. Application of planar shape comparisons to object retrieval in image databases. *Pattern Recognition*, 35(1):15–29, 2002.

[19] Pavlidis T and Ali F. Computer recognition of handwritten numerals by polygonal approximations. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-5:610–614, November 1975.

[20] Pavlidis T and Horowitz SL. Segmentation of plane curves. *IEEE Transactions on Computers*, C-23:860–870, August 1974.

[21] Arkin EM, Chew LP, Huttenlocher DP, Kedem K, and Mitchell JSB. An efficiently computable metric for comparing polygonal shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:209–216, March 1991.

[22] Metz CE. Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, VIII(4):283–298, 1978.