35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Measures of Radial Correlation and Trend for Classification of Breast Masses in Mammograms

Paola Casti[1*], Arianna Mencattini[1], Marcello Salmeri[1],
Antonietta Ancona[2], Fabio Mangieri[2], and Rangaraj M. Rangayyan[3]

Abstract— **In this paper, a novel approach for classification of breast masses is presented that quantifies the texture of masses without relying on accurate extraction of their contours. Two novel feature descriptors based on 2D extensions of the reverse arrangement (RA) and Mantel's tests were designed for this purpose. Measures of radial correlation and radial trend were extracted from the original gray-scale values as well as from the Gabor magnitude response of 146 regions of interest, including 120 benign masses and 26 malignant tumors. Four classifiers, Fisher-linear discriminant analysis, Bayesian, support vector machine, and an artificial neural network based on radial basis functions (ANN-RBF), were employed to predict the diagnosis, using stepwise logistic regression for feature selection and the leave-one-patient-out method for cross-validation. The ANN-RBF resulted in an area under the receiver operating characteristic curve of 0.93. The experimental results show the effectiveness of the proposed approach.**

I. INTRODUCTION

Methods for computer-aided diagnosis (CADx) of breast cancer have been investigated by researchers to classify benign masses and malignant tumors in mammograms. Most of the reported works have been focused on the analysis of masses via texture and morphological features based on reliable extraction of the contours of the masses. Lesions are analyzed in terms of the shape of their boundaries (shape features) [1]–[6], gray-scale variations within the margins or in a ribbon surrounding the margins (texture features) [5], [7]–[9], or changes in density across their boundaries (edge sharpness features) [5], [8], [10], [11].

Sahiner et al. [1] combined texture features with morphological features extracted from automatically segmented masses and reported $A_z = 0.91 \pm 0.02$. Most of the other methods described in the literature rely on boundaries traced by expert radiologists. Rangayyan and Nguyen [6] obtained $A_z = 0.93$ by using only the manually traced contours of the masses from which fractal dimension, fractional concavity, and other shape factors were extracted.

Accurate estimation of the contour of a mass, either drawn by a radiologist or extracted by segmentation procedures, is made difficult by indistinct and ill-defined boundaries of obscured masses and spiculated tumors. The aim of this study is to develop and test new measures for classification of breast masses that do not strictly depend on outlined contours and whose effectiveness is not influenced by the accuracy with which these contours depict the mass.

Statistical techniques are conventionally used to determine whether certain characteristics in terms of space and time are present, with an associated confidence, in a given signal. Khademi et al. [12] applied a two-dimensional ($2D$) extension of the reverse arrangement (RA) test [13] and Mantel's test for clustering [14] to examine, respectively, nonstationarity and spatial dependence of pixel values in a set of phantom images. They also explored the possibility of applying such tests to generate feature descriptors.

In this work, we propose two novel feature descriptors to quantify radial correlation and radial trend in the gray-level values of a given region of interest (ROI) containing a mass. A circular domain centered at the centroid of the mass is explored by means of concentric circles, without requiring any accurate boundary of the mass.

II. MATERIALS AND METHODS

A. Dataset of images

The mammograms used in this study were collected at the San Paolo Hospital of Bari, Italy, using the Senograph 2000D ADS 17.3 from GE Medical Systems. Our set of full-field digital mammograms (FFDMs) consists of 194 images from 88 cases with the spatial resolution of 94 μm and gray-scale resolution of 12 bits/pixel. A total of 146 ROIs, including 120 benign masses and 26 malignant tumors, were obtained from the mammograms using contours manually drawn by an expert radiologist specialized in mammography. The centroid of each mass was computed by using the first-order moments of the gray-scale values within its contour weighted by their distances from the center of the axes. Then, each ROI was automatically sized in order to include the manually drawn contour and centered at the centroid of the mass. Results of biopsy provided the diagnostic classification of each mass used to validate the proposed approach. Figures 1 (a) and (e) show two examples of the ROIs used in this work, including, respectively, a benign mass and a malignant tumor.

B. Preprocessing stage

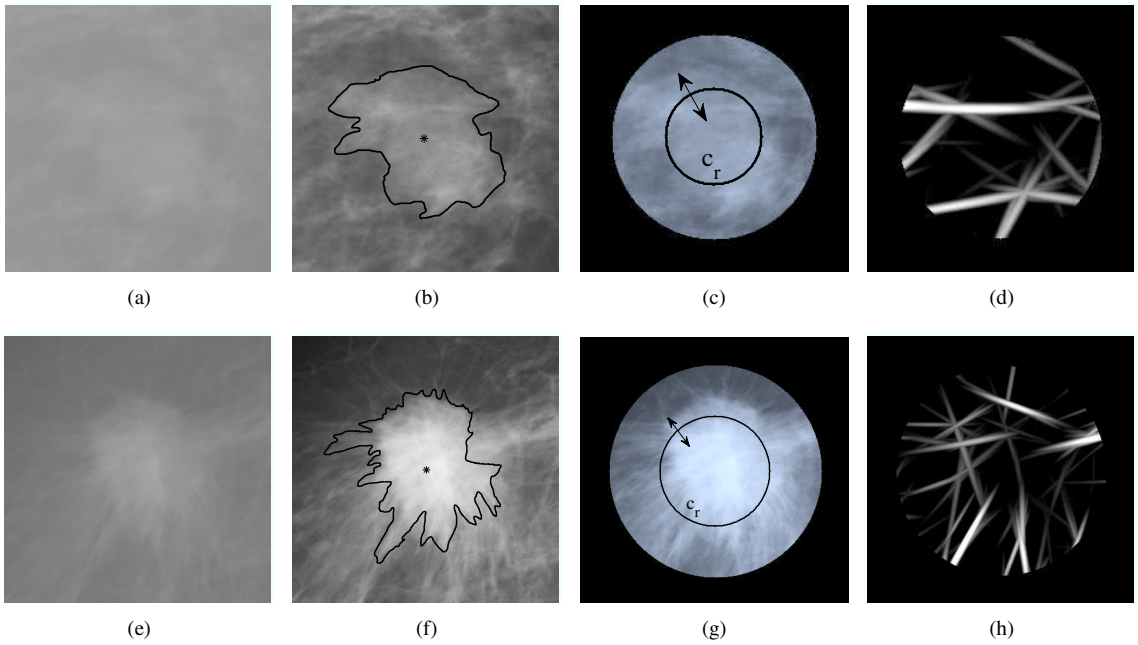
The original ROIs were downsampled to the spatial resolution of 200 μm for reducing the computational cost of the proposed methodology. Then, in order to enhance properly the contrast of the masses, the look-up table (LUT) information of the softer linear transformation encoded in the

[1]P. Casti, A. Mencattini, and M. Salmeri are with the Department of Electronic Engineering, University of Rome Tor Vergata, Italy *casti,mencattini,salmeri@ing.uniroma2.it

[2]A. Ancona and F. Mangieri are with the Diagnostic Radiology Unit, San Paolo Hospital of Bari, Bari, Italy

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

978-1-4577-0216-7/13/$26.00 ©2013 IEEE



(a) (b) (c) (d)

(e) (f) (g) (h)

Fig. 1. (a,e) Original ROI including a benign mass (a) and a malignant tumor (e). (b,f) ROI after the LUT transformation. The contour manually drawn by the expert radiologist and the computed centroid are superimposed on the image. (c,g) Circular domain centered at the centroid of the mass. A ring c_r of a given radius r, $r = 40$ pixels (c) and $r = 80$ pixels (g), is superimposed on the image. (d,h) Gabor magnitude response. Note that the two ROIs were resized to the same dimension for display purposes.

DICOM tags [15] of the FFDMs was applied to the gray-scale values of the ROIs. In Figs. 1 (b) and (f), two examples of ROIs after the LUT transformation are shown, including, respectively, a benign mass and a malignant tumor. The contours drawn by the expert radiologist and the centroids of the masses are superimposed on the ROIs.

A circular domain X of radius R centered at the centroid of the lesion was extracted from each ROI (see Figs. 1 (c) and (g)). Note that the two ROIs in Fig. 1 have been resized to the same dimension for display purposes. As a first attempt, for each ROI, the maximum radial length (R_m), computed as the maximum of the Euclidean distances from the centroid of the mass to each of the boundary coordinates, is used as the radius of the domain X.

C. Extraction of oriented patterns

In order to quantify trend and correlation of the directional components present within a mass, a set of 18 real Gabor filters [16] equally spaced in the angular range $(-\pi/2, \pi/2]$ was applied. The $2D$ Gabor kernel $g(x, y)$ oriented at $-\pi/2$ is defined as

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right)\right] \cos\left(2\pi\frac{x}{\tau}\right), \quad (1)$$

where σ_x and σ_y are the standard deviation values of the Gaussian envelope along the x and the y directions, and τ is the period of the cosine modulation. The filter bandwidth can be determined by two parameters: $\tau = 2\sigma_x \cdot \sqrt{2\log(2)}$ and $l = \sigma_y/\sigma_x$, ($\tau = 24$ pixels and $l = 4$ in this work). The magnitude response [shown in Figs. 1 (d) and (h)] was obtained using, for each pixel, the maximum value among the 18 filter responses [17].

D. Radial correlation measure

The first proposed feature was designed to measure radial correlation within a given domain X. Let us virtually divide the domain X into $R-1$ concentric rings of radius r_i, $i = 1, 2, \ldots, R-1$ pixels, and consider the average of the pixel values μ_i, $i = 0, 1, \ldots, R-1$, belonging to these rings. Note that μ_0 is the pixel value at the centroid of the mass. A generic ring of radius r is superimposed on the image in each case in Figs. 1 (c) and (g).

If radial correlation or radial dependence exists among the pixels in the domain X, rings close to one another will be coupled with average intensity values which are also similar. In order to quantify the extent of correlation in the radial direction, we define a radial correlation measure, $RaCo$, based on Mantel's test for clustering, as

$$RaCo = \frac{1}{R}\sum_{i=0}^{R-1}\sum_{j=0}^{R-1}|r_i - r_j||\mu_i - \mu_j|. \quad (2)$$

Note that a homogeneous distribution of pixel values in the domain X would result in $RaCo = 0$. Higher values of $RaCo$ are indicative of lower spatial dependency or heterogeneity among pixels in the radial direction, which is typical of malignant tumors, whereas lower values of $RaCo$ indicate radial clustering, which is typical of benign masses.

E. Radial trend measure

Let us consider $R-1$ circular regions C_i of radius $i = 1, 2, \ldots, R-1$, and the averages of the pixel values M_i,

$i = 0, 1, \ldots, R-1$, within these regions. We propose a radial trend measure based on a novel $2D$ extension of the RA test, which involves computing the number of times, starting with the first average intensity value M_0, that pixel values within the complement of C_0 in the domain X are less than M_0. The sum of all the counts defined as above is normalized by dividing by R, as

$$RaTr = \frac{1}{R} \sum_{i=0}^{R-1} \#X(\overline{C_i}) < M_i, \tag{3}$$

where $\overline{C_i}$ is the complement of C_i in the domain X. Higher values of $RaTr$ are indicative of radial trend (non-stationarity), which can be correlated to the characteristics of a malignant tumor infiltrating the surrounding tissue in the radial direction (e.g., radiating spicules).

III. RESULTS AND DISCUSSION

The two proposed measures of radial trend and radial correlation described above were computed for all of the ROIs obtained from the FFDM database of mammograms. For each ROI, four features were extracted by computing the two measures from the original ROI after the LUT transformation ($RaTr_1$, $RaCo_1$) and the Gabor magnitude response ($RaTr_2$, $RaCo_2$).

The performance of each feature was, at first, analyzed independently without training any classifier by means of the ROCKIT package [18]. The obtained individual area under the receiver operating characteristic curve (A_z) values are listed in Table I. The A_z values obtained indicate satisfactory to good performance in the classification of benign masses versus malignant tumors.

In order to evaluate the dependence of the performance of the features on the dimension of the domain X, the A_z value was computed, for each feature, by varying the radius of the domain in the range $[0.5R_m, R_m]$, where R_m is the maximum radial length defined in Section II-B. Figure 2 illustrates the effect of varying the radius on the A_z values. Experimental results show that the proposed features are robust to variations in the dimension of the domain X.

To avoid bias, feature selection using stepwise logistic regression [19] and pattern classification were performed using the leave-one-patient-out cross-validation method. All of the four features were selected more than 50% of the time. Four classifiers, Fisher Linear discriminant analysis (FLDA), the Bayesian classifier (quadratic discriminant analysis with Bayesian assumption, BAYES), support vector machine (SVM) [20], and an artificial neural network classifier based on radial basis functions (ANN-RBF) [21] were used. Further details of the classifiers used in this work can be found in Duda et al. [20] and Haykin [21]. A_z (and standard error, SE) values of 0.82 (0.05), 0.76 (0.06), 0.85 (0.05), and 0.93 (0.03) were obtained, respectively, with the four classifiers listed above. Table II lists the obtained results along with the asymmetric 95% confidence intervals estimated using ROCKIT. The high values of A_z obtained, especially 0.93 with the ANN-RBF, indicate that the four features complement one another and can be used together to achieve increased classification efficiency. The performance achieved is comparable to the best results reported in the literature [1], [3], [6], [7]. In particular, in the work by Sahiner et al. [1], $A_z = 0.91$ is reported when the leave-one-patient-out method was applied to partition the data set into the training and test sets.

The experiments were carried out using Matlab® of ver-

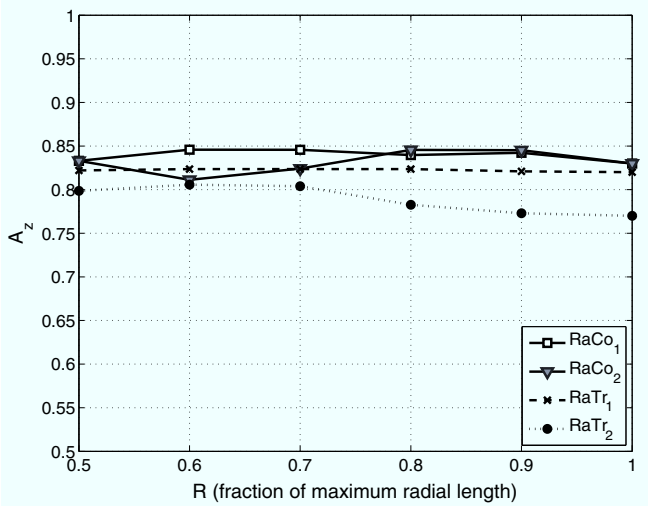


Fig. 2. Effect of varying the radius of the domain X on the A_z values of the proposed features. $RaTr_1$, $RaTr_2$: Radial trend measures; $RaCo_1$, $RaCo_2$: Radial correlation measures.

TABLE I

LIST OF FEATURES FOR CHARACTERIZATION OF RADIAL CORRELATION AND RADIAL TREND WITH A_z VALUES.

Symbol	Feature	A_z
$RaCo_1$	Original ROI after LUT transformation	0.83
$RaCo_2$	Gabor magnitude response	0.83
$RaTr_1$	Original ROI after LUT transformation	0.82
$RaTr_2$	Gabor magnitude response	0.77

$RaCo_1$, $RaCo_2$: radial correlation measures; $RaTr_1$, $RaTr_2$: radial trend measures. The A_z values were estimated using ROCKIT.

TABLE II

RESULTS OF ROC ANALYSIS USING DIFFERENT TYPES OF CLASSIFIERS.

Classifier	ROC analysis (A_z)	SE	$I_{95\%}$
FLDA	0.82	0.05	[0.72, 0.90]
BAYES	0.76	0.06	[0.63, 0.87]
SVM	0.85	0.05	[0.74, 0.92]
ANN-RBF	0.93	0.03	[0.85, 0.98]

FLDA: Fisher Linear discriminant analysis; BAYES: Bayesian classifier; SVM: support vector machine; ANN-RBF: artificial neural network classifier based on radial basis functions. The A_z values were estimated using ROCKIT along with the related standard error (SE) and the asymmetric 95% confidence interval ($I_{95\%}$).

sion R2012a on an Intel® Core™ i7 processor at 2.10 GHz and 4 GB of RAM. The average processing time of each ROI for the whole methodology described is 18.2 s using the ANN-RBF classifier.

The main advantage of the proposed methodology is that its effectiveness is independent of the accuracy with which the contours used depict the masses. In fact, the quantification of radial correlation and radial trend among pixels only requires a circular domain X properly centered on the lesion. This is confirmed by the low variability in the classification performance of the features in relation to the size of the spatial domain X used in their computation. Although the ROIs in the present work were initially obtained with reference to the contours of the masses drawn by a radiologist, this step is not required in the proposed procedure since only the spatial domain X is used for computation of the features and it will be removed in subsequent work. The centroid and a range of the expected size of the mass are required to define the domain X used to compute the features. However, the centroid of the mass could be detected by means of automated procedures [22], already developed by the authors, or manually pointed by a radiologist. The range of the size of masses expected in a certain dataset or population could be estimated from previous data or experience.

The obtained results show that the features proposed in the present work could be used to assess the likelihood of malignancy of breast masses in mammograms.

IV. CONCLUSIONS AND FUTURE WORK

New methods were proposed in the present paper for quantification of radial correlation and radial trend in breast masses as seen in mammograms. The developed features, inspired by Mantel's [14] and reverse arrangement [13] tests, can be used for classification of breast masses in mammograms without requiring the extraction of the precise contours of the masses.

Further work is in progress to optimize the proposed features, design contour-free features, and quantify, for instance, correlation and trend between pixels in the angular direction, so that the classification performance can be improved. Appropriate grid search will be also carried out in order to chose the optimal set of SVM and ANN-RBF parameters. The robustness of the method will be tested with larger and publicly available databases of screen-film mammograms (e.g., miniMIAS [23] and DDSM [24] databases). The proposed techniques should assist in the development of improved methods for screening and diagnosis of breast cancer.

ACKNOWLEDGMENT

We thank the Diagnostic Radiology Unit, San Paolo Hospital of Bari, Italy, for providing the digital mammograms used in this work.

REFERENCES

[1] Sahiner B, Chan HP, Petrick N, Helvie MA, and Hadjiiski LM, "Improvement of mammographic mass characterization using spiculation measures and morphological features," *Med Phys*, vol. 28, no. 7, pp. 1455 – 1465, Jul. 2011.

[2] Bruce LM and Adhami RR, "Classifying mammographic mass shapes using the wavelet transform modulus-maxima method," *IEEE Trans Med Imag*, vol. 18, no. 12, pp. 1170 – 1177, Dec. 1999.

[3] Rangayyan RM, El-Faramawy NM, Desautels JEL, and Alim OA, "Measures of acutance and shape for classification of breast tumors," *IEEE Trans Med Imag*, vol. 16, no. 6, pp. 799 – 810, Dec. 1997.

[4] Rangayyan RM, Mudigonda NR, and Desautels JEL, "Boundary modelling and shape analysis methods for classification of mammographic masses," *Med Biol Eng Comput*, vol. 38, no. 5, pp. 487 – 496, Sep. 2000.

[5] André TCSS and Rangayyan RM, "Classification of breast masses in mammograms using neural networks with shape, edge-sharpness, and texture features," *J Electron Imaging*, vol. 15, no. 1, Jan. 2006, Article 013019.

[6] Rangayyan RM and Nguyen TM, "Fractal analysis of contours of breast masses in mammograms," *J Digit Imaging*, vol. 4, no. 3, pp. 223 – 237, Sep. 2007.

[7] Sahiner B, Chan HP, Petrick N, Helvie MA, and Goodsitt MM, "Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis," *Med Phys*, vol. 25, no. 4, pp. 516 – 526, Apr. 1998.

[8] Mudigonda NR, Rangayyan RM, and Desautels JEL, "Gradient and texture analysis for the classification of mammographic masses," *IEEE Trans Med Imag*, vol. 19, no. 10, pp. 1032 – 1043, Oct. 2000.

[9] Rangayyan RM, Nguyen TM, Ayres FJ, and Nandi AK, "Effect of pixel resolution on texture features of breast masses in mammograms," *J Digit Imaging*, vol. 23, no. 5, pp. 547 – 553, Oct. 2010.

[10] Alto H, Rangayyan RM, and Desautels JEL, "Content-based retrieval and analysis of mammographic masses," *J Electron Imaging*, vol. 14, no. 2, May 2005, article 023016.

[11] Rojas-Domínguez A and Nandi AK, "Development of tolerant features for characterization of masses in mammograms," *Comput Biol Med*, vol. 39, no. 8, pp. 678 – 688, Aug. 2009.

[12] Khademi A, Hosseinzadeh D, Venetsanopoulos A, and Moody A, "Nonparametric statistical tests for exploration of correlation and nonstationarity in images," in *Proc. 16th International Conference on Digital Signal Processing*, Santorini-Hellas, Greece, Jul. 2009, pp. 1 – 6.

[13] Beck T, Housh T, Weir J, Cramer J, Vardaxis V, Johnson G, Coburn J, Malek M, and Mielke M, "An examination of the runs test, reverse arrangements test, and modified reverse arrangements test for assessing surface EMG signal stationarity," *J Neurosci Methods*, vol. 156, no. 1 – 2, pp. 242 – 248, Sep. 2006.

[14] Mantel N, "The detection of disease clustering and a generalized regression approach," *Cancer Res*, vol. 27, no. 2, pp. 209 – 220, Feb. 1967.

[15] National Electrical Manufacturers Association (NEMA), "Digital Imaging and Communication in Medicine (DICOM)," 2011, Rossyln, VA.

[16] Gabor D, "Theory of communications," *J Inst Electr Eng*, vol. 3, no. 93, pp. 429 – 457, Feb. 1967.

[17] Ayres FJ and Rangayyan RM, "Design and performance analysis of oriented feature detectors," *J Electron Imaging*, vol. 16, no. 2, Apr. 2007, article 023007.

[18] "ROCKIT software," http://www.radiology.uchicago.edu/ page/metz-roc-software.

[19] Ramsey FL and Schafer DW, *The Statistical Sleuth: A Course in Methods of Data Analysis*, Duxbury Press, Belmont, CA, 1997.

[20] Duda RO, Hart PE, and Stork DG, *Pattern Classification*, Wiley-Interscience, New York, 2nd edition, 2001.

[21] Haykin S, *Neural Networks: A Comprehensive Foundation*, Prentice Hall, Englewood Cliffs, NJ, 2nd edn. edition, 1999.

[22] Mencattini A and Salmeri M, "Breast masses detection using phase portrait analysis and fuzzy inference systems," *Int J Comput Assist Radiol Surg*, vol. 7, no. 4, pp. 573 – 583, Oct. 2011.

[23] Suckling J, Parker J, Dance DR, Astley, Hutt I, Boggis CRM, Ricketts I, Stamakis E, Cerneaz N, Kok SL, Taylor P, Betal D, and Savage J, "The Mammographic Image Analysis Society digital mammogram database," Exerpta Medica, International Congress Series 1069, 1994, pp. 242 – 248.

[24] Heath M, Bowyer K, Kopans D, Moore R, and Kegelmeyer WP, "The Digital Database for Screening Mammography," in *Proc. 5th International Workshop on Digital Mammography*, Ed. Medical Physics Publishing, 2001, pp. 212 – 218.