Automatic BI-RADS Description of Mammographic Masses

Fabián Narváez, Gloria Díaz, and Eduardo Romero

Bioingenium Research Group, Departament of Medicine,
National University of Colombia, Bogotá, Colombia
{frnarvaeze,gmdiazc,edromero}@unal.edu.co
http://www.bioingenium.unal.edu.co

Abstract. This paper presents a CBIR (Content Based Information Retrieval) framework for automatic description of mammographic masses according to the well known BI-RADS lexicon. Unlike other approaches, we do not attempt to segment masses but instead, we describe the regions an expert selects, after the series of rules defined in the BI-RADS lexicon. The content based retrieval strategy searches similar regions by automatically computing the Mahalanobis distance of feature vectors that describe main shape and texture characteristics of the selected regions. A description of a test region is based on the BI-RADS description associated to the retrieved regions. The strategy was assessed in a set of 444 masses with different shapes and margins. Suggested descriptions were compared with a ground truth already provided by the data base, showing a precision rate of 82.6% for the retrieval task and a sensitivity rate of 80% for the annotation task.

Keywords: Automatic Annotation, BI-RADS, Computer Aided Diagnosis, Content-based Image Retrieval.

1 Introduction

Breast cancer is the most frequent type of cancer in women and is considered as the largest public health problem in women population [1,2]. This disease is fully curable if diagnosis is achieved early and mammography is the more efficient method for visualizing these first abnormalities [3,4]. However, mammographic interpretation is really hard, studies have shown that between 10% and 25% of breast cancers are not detected [5]. An agreement to reduce variability between radiologists resulted in the *Breast and Imaging Report and Database System* (BI-RADS), designed by the the American College of Radiology , as a standard description to report breast lesions for allowing to categorize different pathologies as well as their severity level [6]. This standard established that basic descriptors for masses are shape, margin and density. Thus, automatic mammography categorization based on BI-RADS descriptors is becoming important. Most of related works have classified the tissue density of mammograms based on the four BI-RADS categories [7,8,9]. But, shape and margin descriptors there are

J. Martí et al. (Eds.): IWDM 2010, LNCS 6136, pp. 673–681, 2010.

not considered. Recently, an approach to classify the shape and margin of mammographic mass into different categories was presented [10]. In this case masses were previously segmented in order to extract the boundaries that were then characterized and classified, results reported a precision lower than 70%.

On the other hand, development of Computer Assisted Diagnosis Systems (CAD) for mammography has decreased the variability effects, becoming a well accepted clinical practice to assist radiologists interpreting mammograms when they search and identify micro-calcification clusters [11]. However, the relatively low performance of CAD schemes in mass detection [12], make them less accepted as mass diagnosis tools. As an alternative, the interactive CAD systems, based on content-based information retrieval schemes [13,14,15], identify visually similar mass lesions that eventually are clinically relevant to the actual lesion [16]. Actually, CBIR-based CAD schemes have a potential to provide radiologist with visual aid and increase their confidence in accepting CAD-cued results in the decision making process. However, their main drawback is that they are also based on the segmentation of mass regions, which is a very difficult task especially for masses with blurred boundaries.

This paper proposes a new approach to support and assist the task of describing mass lesions from a set of Regions of Interests (RoIs). Given a particular query or region under examination, the method finds the most similar regions from a database, according to the BI-RADS lexicon, using the two most important diagnostic features for describing masses: morphology and texture. Morphology is described using the Zernike moments [17], and texture is captured via the Neighborhood Gray Tone Difference Matrix (NGTDM) [18]. Once these basic features are computed, a further reduction of dimensionality is achieved using a standard Principal Component Analysis (PCA), assembling a descriptor of 15 dimensions. Finally, a multiclass retrieval algorithm based on a k-NN scheme is constructed for the shape, margin descriptions and pathology classification.

The rest of this article is organized as follows: after this introduction, next section presents the methodology, then results are shown and last section discusses future work and conclusions.

2 Methods

An overview of the proposed framework is shown in Fig. 1. Firstly, a radiologist manually selects a region of interest as the queried RoI, which is preprocessed to improve the mass visual details. Afterward, morphology and texture features are extracted, compared with the information stored in the database (reference database) so that the most similar regions are retrieved. Finally, the BI-RADS is used to assign the most probable description to these regions.

2.1 RoI Pre-processing

Mammography analysis generally must deal with regions difficult to interpret [19] since they are associated to hard acquisition conditions. Every image was enhanced and two resulting images were prepared for analysis, the former aimed to

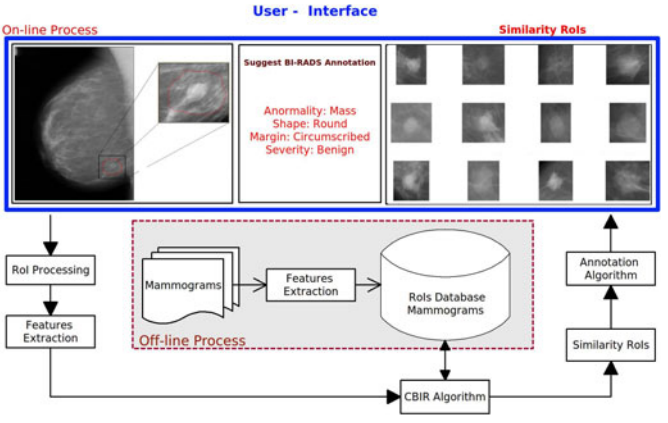

Fig. 1. Method overview

highlight mass edges by stretching the maximum and minimum gray level values to the interval $[0, 255]$, followed by a bin reduction from 256 to 12 bins. The latter captured differences between patterns associated with mass and parenchyma tissue by adaptively equalizing the histogram so that structural details were preserved. In both cases, resultant images were smoothed out by a median filter in order to remove remaining noise [20]. Mass descriptors were drawn from the former image while texture features were extracted from the latter.

2.2 Feature Extraction

According to the BI-RADS lexicon, morphology and texture are the most important criteria for mass diagnosis [5,21]. Tradionally, Zernike moments, a class of statistical moments [22], have shown to be very effective for morphological representation i.e. rotation-invariant and robust to the noise [23]. Furthermore, representation with Zernike moments allows reconstruction with minimal losses and constitute a classical multiresolution representation for shapes [24,23]. The Zernike polynomials are a set of complex polynomials that form an orthogonal complete set $V_{pq}(x, y)$ within the unitary circle, which are defined by the equation 1.

$$V_{pq}(x, y) = R_{pq}(r)e^{jq\theta}, r \in [-1, 1] \tag{1}$$

where $r = \sqrt{x^2 + y^2}$ is the vector magnitude and $\theta = tan^{-1}\left(\frac{y}{x}\right)$ its angle.

The complex Zernike moments are derived from the real-valued radial polynomials, given by [2].

$$R_{pq}(r) = \sum_{s=0}^{(p-|q|)/2} (-1)^s \frac{(p-s)!}{s!(\frac{p+|q|}{2}-s)!(\frac{p-|q|}{2}-s)!} r^{p-2s} \tag{2}$$

where p and q are subject to $p - |q|$ is even, $0 \leqslant |q| \leqslant p$, and $p \geqslant 0$. Then, the complex Zernike moments of order p, with q repetitions for an image intensity function f(x,y) are given by [3].

$$Z_{pq} = \frac{p+1}{\pi} \sum_{x} \sum_{y} V_{pq}^{*}(x,y) f(x,y) \tag{3}$$

where $*$ stands for the conjugated complex of $V_{pq}(x,y)$.

In this work the Hosny implementation [25] of Zernike moments was used for describing shapes. For achieving so, the RoI was mapped onto the unitary circle so that the image center coincides with the unitary circle center, subjected to the condition that every pixel is within the RoI. The number of order moments used for generating the shape descriptor was selected by minimization of the reconstruction error ε given by the equation 4. The first 60 order moments were heuristically selected for generating a descriptor of 961 features.

$$\varepsilon_n = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \{ \frac{[f(i,j) - F(i,j)_n]^2}{[f(i,j)]^2} \} \tag{4}$$

where $f(i,j)$ is the original image and $F(i,j)_n$ is the image reconstructed using n first order moments, as presented by Chong et al. [17].

On the other hand, the essential texture features were captured via the Neighborhood Difference Gray Tone Matrix [18] as follows: a neighborhood is firstly set and the absolute difference of the central pixel with its neighborhood average is computed. This average difference constitutes an occurrence that is stored in a histogram whose bins 1, 2, 3, 4 and 5 corresponded to the neighborhood sizes. So, this descriptor allows to capture pattern differences around to mass boundary in an area of 25 pixels, larger neighborhood sizes were evaluated but poor performance was obtained. So, five histograms of 256 positions were generated and five features were calculated from each, as described in [18]:

1. Contrast $= \left(\frac{1}{N_g(Ng-1)} \sum_{i=0}^{G_h} \sum_{j=0}^{G_h} p_i p_j (i-j)^2 \right) \left(\frac{1}{n^2} \sum_{i=0}^{G_h} s(i) \right)$
2. Busyness $= \left(\sum_{i=0}^{G_h} p_i s(i) \right) / \left(\sum_{i=0}^{G_h} \sum_{j=0}^{G_h} i p_i - j p_j \right)$
3. Complexity $= \sum_{i=0}^{G_h} \sum_{j=0}^{G_h} \left[|i-j| / (n^2(p_i + p_j)) \right] \left[p_i s(i) + p_j s(j) \right]$
4. Texture strength $= \left[\sum_{i=0}^{G_h} \sum_{j=0}^{G_h} (p_i + p_j)(i-j)^2 \right] / \left[\epsilon \sum_{i=0}^{G_h} s(i) \right]$
5. Coarseness $= \epsilon + \sum_{i=0}^{G_h} p_i s(i)$

where N_g is the total number of different gray levels with $n = N - 2d$, for an $N \times N$ image, G_h is the highest gray tone in the image, p_i is the probability of occurrence of the i^{th} gray tone, $s(i)$ is the histogram value at i^{th} gray tone and ϵ is a small number that prevents these values become infinite.

2.3 Automatic Mass Description

A content based information retrieval strategy that uses the K-NN rule (K-Nearest Neighbor [26]), was implemented. Such algorithm uses the information associated a certain number of images retrieved from the reference data base to assign a label to a particular RoI. Given the knowledge of N prototype RoIs (each marked by an expert radiologist) and their correct classification into several classes, it assigns an unclassified RoI to the class that is most heavily represented among its k nearest neighbors. The algorithm used a weighted Mahalanobis distance (w_d) to measure the similarity among the feature vectors describing both the database and queried RoIs. Once a set of K RoIs are retrieved, each shape, margin and pathology BI-RADS description is set using the decision rule in equation 5, where $Label(S)$ assigns the value description with the largest weight.

$$Label(S) = \arg\max_{S_i} |S_1, ..., S_n|, \quad S_i = \sum_{i=1}^{K} w_d^{s_i} \tag{5}$$

where $w_d = 1/d(\boldsymbol{x}, \boldsymbol{y})$ is the relative weight of each possible value description and S_n corresponds to number of possible labels for each shape, margin and pathology BI-RADS description.

3 Experimental Results

A total of 444 regions extracted from the *Digital Database for Screening Mammography (DDSM)* [27], which were previously BI-RADS annotated by a group of breast radiologists, were used to evaluate the performance of the proposed approach. These regions were split into training (344 RoIs) and testing (100 RoIs) sets. Both the content-based image retrieval scheme and the automatic annotation algorithm were independently assessed.

Before applying the retrieval process, a PCA analysis was used to reduce the feature vector dimensionality. From 986 extracted features (961 from shape descriptor and 25 from texture descriptor), 15 were selected, the ones which reported the larger eigen-value variability in the training subset.

3.1 Content-Based Image Retrieval Evaluation

Retrieval performance was assessed by computing the relevance of the recovered images, according to the ground truth of DDSM mammogram databases, identified by an experienced radiologists. Furthermore, shape, margin and pathology characteristics are taken into consideration in this evaluation. Therefore, We used four levels to describe the degree of relevance , namely $Score = 1$ for three correct targets, $Score = 0,66$ for two, $Score = 0,33$ for one and $Score = 0$ for zero. The Precision-Recall (P-R) graph was used for evaluating the performance of CBIR scheme. The precision (P) was defined as the relevance of images that the system was able to find among all images retrieved by the system, while

recall (R) was the relevance of images that the system was able to find among all the relevant images stored in the database.

$$P = \frac{\sum_{i=1}^{k} S_i}{K}, \qquad R = \frac{\sum_{i=1}^{k} S_i}{\sum_{i=1}^{n} S_i} \tag{6}$$

where S_i is the score assigned to the i^{th} RoI, K is the number of retrieved images and n is the total number of images in the reference database.

Retrieval performance for a set of 100 image queries was evaluated, using the 15 most similar images for annotation. An average precision and recall of respectively 82% and 42%, were obtained.

Figure 2 shows the Precision-Recall curve obtained when the average precision and recall measures were computed for a number of retrieved regions which varied from 1 to 15, with incremental steps of 1.The first point, the leftest curve point represents the average precision and recall rates for the first returned image. The second point corresponds to the precision and recall for the first two retrieved images, and so forth. It is observed from the graph that the first retrieved image report a relevance up to 80%. As expected, as long as the number of retrieved images increases, this precision decreases. However and interestingly, precision was higher than 65% in general. On the other hand, high recall values were found basically because the number of relevant images in the database was lower than the number of queried images.

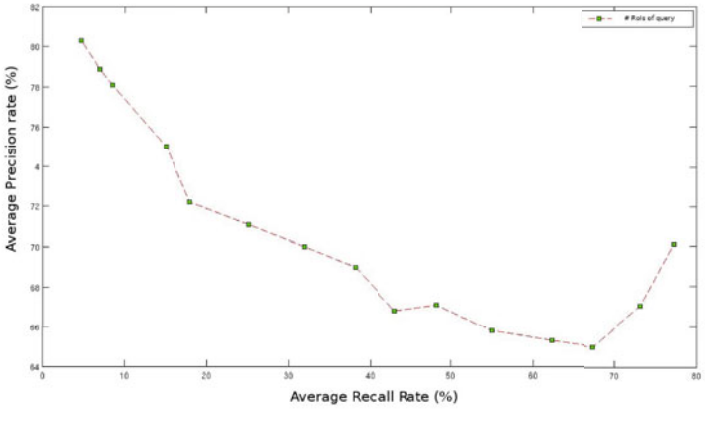

Fig. 2. Precision-Recall Average Curve

3.2 Automatic Annotation Assessment

Each shape, margin and severity BI-RADS description was assigned to each ROI query, based on equation 5. The optimal number of images used to assign each BI-RADS mass description was estimated by a 10-fold cross validation assessment. Results showed that a minimal of 7 images are needed for annotating shape

Table 1. Performance of Automatic BI-RADS description

	Accuracy	PPV	TPR
Shape	80.3	0.791	0.80
Margin	75.1	0.752	0.76
Pathology	85.3	0.861	0.87
Average	80.23	0.801	0.81

and margin descriptions while 9 are required for establishing a severity level. Table 1 shows the accuracy (ACC), positive predictive value (PPV) and true positive rate (TPR) computed from the automatic annotation for each BI-RAD description. The most difficult annotation was the margin, probably because of the blurred boundaries presented in many mass classes. However, average performance annotation is higher than 80%, a figure very acceptable for a system that tries to assist a diagnosis task.

4 Conclusions

In this paper a new strategy for assisting the diagnosis of mammography, based on a content based image retrieval scheme was proposed, implemented and evaluated. This strategy provided a BI-RADS mass description of a region of interest, which was supported by a set of diagnosed images that were shown to the expert. Instead of attempting to segment masses, we proposed a mass feature description, based on its internal structure with no explicit mass boundary detection.

The proposed approach was evaluated on a public image database (DDSM). Retrieval results have shown that this approach is successfully able to find the most similar images of a RoI query in a reference database. Likewise, the annotation results have also shown the capability of the method for generating the shape, margin and severity descriptions associated to the RoI, according to the BI-RADS lexicon, especially for discriminating the severity label i.e to decide whether a mass is benign or not.

These preliminary results have opened up new strategies for the computer assisting tools based on CBIR schemes in mammographic diagnosis, although a validation of the impact on diagnostic improvement of inexperienced radiologists is required.

References

1. Scottish Intercollegiate Guidelines Network : Management of breast cancer in women. A national clinical guideline 84, 1–3 (2005)
2. American Cancer Society : American cancer statistics (2007); Revisado el 2 Septiembre (2008)
3. Verma, K., Zakos, J.: A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques. IEEE Transactions on Information Technology in Biomedicine 16, 219–223 (2002)

4. Buseman, S., Mouchawar, J., Calonge, N., Byers, T.: Mammography screening matters for young women with breast carcinoma. Cancer 97, 352–358 (2003)
5. Bird, R., Wallace, T., Yankaskas, B.: Analysis of cancers missed at screening mammography. Radiology 178, 234–247 (1992)
6. American College of Radiology (ACR): Illustrated Breast Imaging Reporting and Data System (BI-RADS). ACR (1998)
7. Marias, K., Linguraru, M.G., Ballester, M.G., Petroudi, S., Tsiknakis, M., Sir, M.: Automatic labelling and bi-rads characterisation of mammogram densities. In: Proceedings of the IEEE Engineering in Medicine and Biology (2005)
8. Petroudi, S., Kadir, T., Brady, M.: Automatic classification of mammographic parenchymal patterns: A statistical approach. In: IEEE International Conference on Engineering in Medicine and Biology Society (2003)
9. Bovis, K., Singh, S.: Classification of mammographic breast density using a combined classifier paradigm. In: Medical Image Understanding and Analysis (2002)
10. Tao, Y., Lo Shih-Chung, B., Freedman Matthew, T., Erini, M., Xuan, J.: Automatic categorization of mammographic masses using bi-rads as a guidance. In: Proceedings of SPIE, the International Society for Optical Engineering (2008)
11. Nishikawa, R.M.: Current status and future directions of computer-aided diagnosis in mammography. Computerized Medical Imaging and Graphics 31, 224–235 (2007)
12. Gur, D., Stalder, J.S., Hardesty, L.A., Zheng, B., Sumkin, J.H., Chough, D.M., Shindel, B.E., Rockette, H.E.: Computer-aided detection performance in mammographic examination of masses: assessment. Radiology 233, 418–423 (2004)
13. Tao, Y., Lo, S.B., Freedman, M.T., Xuan, J.: A preliminary study of content-based mammographic masses retrieval. In: Proc. SPIE, vol. 6514, p. 65141Z (2007)
14. Zheng, B., Mello-Thoms, C., Wang, X.H., Abrams, G.S., Sumkin, J.H., Chough, D.M., Marie, G.A., Lu, A., Gur, D.: Interactive computer aided diagnosis of breast masses: Computerized selection of visually similar image sets from a reference library. Academical Radiology 14, 917–927 (2007)
15. Rosa, N.A., Felipe, J.C., Traina, A.J., Rangayyan, R.M., Azevedo-Marques, P.M.: Using relevance feedback to reduce the semantic gap in content-based image retrieval of mammographic masses. In: Conf. Proc. IEEE Med. Biol. Soc., pp. 406–409 (2008)
16. Rangayyana, R.M., Ayres, F.J., Desautelsa, J.E.L.: A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. Journal of the Franklin Institute 344, 312–348 (2007)
17. Wee, C.Y., Paramesran, R.: On the computational aspects of zernike moments. Image and Vision Computing 25, 967–980 (2007)
18. Amadasun, M., King, R.: Textural features corresponding to textural properties. IEEE Transactions on systems, man and Cybernetics 19, 1264–1274 (1989)
19. Cheng, H.D., Shi, X.J., Min, R., Hu, L.M., Cai, X.P., Du, H.N.: Approaches for automated detection and classification of masses in mammograms. Pattern Recognition 39, 646–668 (2006)
20. Petrick, N., Chan, H.P., Sahiner, B., Wei, D.: An adaptive density weighted contrast enhancement filter for mammographic breast mass detection. IEEE Trans. Med. Imaging 15(1), 59–67 (1996)
21. Maggio, C.D.: State of the art of current modalities for the diagnosis of breast lesions. Eur. J. Nucl. Med. Mol. Imaging 31(suppl.1), S59–S69 (2004)
22. Teague, M.R.: Image analysis via the general theory of moments. J. Optical Soc. Am. 70, 920–930 (1980)
23. Belkasim, S., Hassan, E., Obeidi, T.: Radial zernike moment invariants. In: The Fourth Int. Conf. on Computer and Information Tech., vol. 1, pp. 790–795 (2004)

24. Kim, H., Kim, J.: Region-based shape descriptor invariant to rotation, scale and translation. Signal Proc.: Image Communication 16, 87–93 (2000)
25. Hosny, K.M.: Fast computation of accurate zernike moments. J. Real-Time Image Proc. 3, 97–107 (2008)
26. Cover, T.M., Hart, P.E.: Nearest neighbor pattern classification. IEEE Trans. on Inf. Theo. IT-13, 21–27 (1967)
27. Heath, M., Bowyer, K., Kopans, D., Moore, R., Kegelmeyer, W.P.: The digital database for screening mammography. In: Yaffe, M.J. (ed.) Proceedings of the Fifth International Workshop on Digital Mammography, pp. 212–218. Medical Physics Publishing (2001)