AUTOMATIC CLASSIFICATION OF BREAST DENSITY

Arnau Oliver, Jordi Freixenet

Computer Vision and Robotics Group
University of Girona
Campus de Montilivi s/n. 17071 Girona, Spain
{*aoliver,jordif*}*@eia.udg.es*

Reyer Zwiggelaar

Department of Computer Science
University of Wales
Aberystwyth SY23 3DB, Wales, UK
rrz@aber.ac.uk

ABSTRACT

A recent trend in digital mammography are Computer-Aided Diagnosis systems, which are computerised tools designed to assist radiologists. Most of these systems are used for the automatic detection of abnormalities. However, recent studies have shown that their sensitivity is significantly decreased as the density of the breast increases. This dependence is method specific. In this paper we propose a new approach to the classification of mammographic images according to their breast parenchymal density. Our classification uses information extracted from segmentation results and is based on the underlying breast tissue texture. Classification performance was based on a large set of digitised mammograms. Evaluation involves different classifiers and uses a leave-one-out methodology. Results demonstrate the feasibility of estimating breast density using image processing and analysis techniques.

1. INTRODUCTION

Breast cancer is considered a major health problem in western countries, and indeed it constitutes the most common cancer among women. A recent study developed by the American Cancer Society estimates that, in this country, between one in eight and one in twelve women will develop breast cancer during their lifetime [1]. Mammography remains the key screening tool for abnormality detection, because it allows identification of tumour before being palpable. However, of all lesions previously diagnosed as suspicious and sent for biopsy, approximately 25% were confirmed malignant lesions, and approximately 75% were diagnosed benign. This high false-positive rate is related to the difficulty of obtaining accurate diagnosis from image information alone [2]. Computerised image analysis is starting to play an important role in improving the issued diagnosis. Computer-Aided Diagnosis (CAD) systems are composed of a set of tools to assist radiologists to detect and diagnose new cases [3]. However, recent studies have shown that the sensitivity of these systems is significantly decreased as the density of the breast increased while the specificity of the systems remained relatively constant [4].

The origins of breast density classification are the work of Wolfe [5], who showed the relation between mammographic parenchymal patterns and the risk of developing breast cancer, classifying the parenchymal patterns in four categories. Since the discovery of this relationship, automated parenchymal pattern classification has been investigated [6, 7]. However, only a small number of previous papers have suggested that texture representation of the breast might play a significant role. Miller and Astley [8] investigated texture-based discrimination between fatty and dense breast types applying granulometric techniques and Laws texture masks. Byng et al. [9] used measures based on fractal dimension. Bovis and Singh [10] estimated features from the construction of Spatial Gray Level Dependency matrices. Recently, Petroudi et al. [11] used textons to capture the mammographic appearance within the breast area. Zwiggelaar et al. [12] segment mammograms into density regions based on a set of co-occurrence matrices. Density classification uses the size of the density regions as the feature space. This work was extended in [13] were a transportation algorithm was used to select an *optimal* set of co-occurrence matrices for the segmentation process.

The American College of Radiology (ACR) Breast Imaging Reporting and Data System (BIRADS) [14] is becoming a standard on the assessment of mammographic images. Breast density is classified in four categories:

- BIRADS I: the breast is almost entirely fatty.

- BIRADS II: there is some fibrogandular tissue.

- BIRADS III: the breast is heterogeneously dense.

- BIRADS IV: the breast is extremely dense.

Our approach assumes that mammograms belonging to different BIRADS categories will be represented by tissue
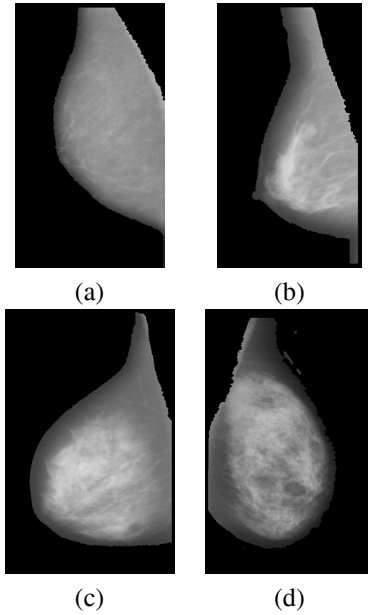

Fig. 1. Four mammograms, each of different BIRADS category: (a) BIRADS I, (b) BIRADS II, (c) BIRADS III and (d) BIRADS IV.

with different texture features. The novelty of our approach is that the grouping of pixels with similar tissue is used as a segmentation strategy. Subsequently, extracting and comparing texture features from each cluster, the system trains two different classifiers. The remainder of this paper is structured as follows: Section 2 describes the proposed segmentation and classification method. Experimental results proving the validity of our proposal are presented in Section 3. Conclusions are given in Section 4.

2. METHODOLOGY

In the literature different approaches for classifying breast tissue based only on the use of histogram information have been proposed [7]. In our experience, however, histogram information seems not to be sufficient to classify mammograms into breast density categories. Our approach is based on segmenting pixels with similar tissue appearance, followed by the extraction of texture features of each segmented group. Using this set of features, we can train different classifiers and use them to classify new images.

However, the initial step of our approach is the segmentation of the profile of the breast [15]. The aim of this segmentation is to extract the breast from other objects that could be present in a mammographic image (background, annotations, pectoral muscle in MLO images) with a minimum loss of breast area. Fig. 1 shows the result of our approach.

2.1. Finding Regions with Similar Tissue

We consider that those pixels from similar tissue have similar gray-level values, as can be seen in Fig. 1. We use the Fuzzy C-Means algorithm [16] to group pixels into separate categories. However, to avoid effects from microtexture that could appear in some regions, we first smooth the breast region with a median filter of size 5×5.

Fuzzy C-Means is an extension of the well known k-Means algorithm. The main difference is that Fuzzy C-Means lets each pattern of the image to be associated with every cluster using a fuzzy membership function (in k-Means, each pattern belongs to one and only one cluster). In our implementation, the function criterion that the algorithm minimises is defined by:

$$e^2(\Xi, U) = \sum_{i=1}^{N} \sum_{k=1}^{K} u_{ik} ||x_i - c_k||^2 \qquad (1)$$

where Ξ is the partition of the image, U is the membership matrix: u_{ik} represents the membership of pattern x_i to belong to cluster k, which have centre $c_k = \sum_{i=1}^{N} u_{ik} x_i$, N is the number of patterns of the image (number of pixels), and K the number of clusters, which has to be known a priori.

When using partitional clustering algorithms, like Fuzzy C-Means, the placement of the initial seed points is one of the central issues in the robustness of segmentation results. Despite their importance, seeds are usually initialised randomly. In our approach, the Fuzzy C-Means is initialised using histogram information, with the aim to obtain representative instances of two classes: normal and dense tissue. Hence, we initialised the two seeds with the gray level values that represent 15% and 85% of the accumulative histogram (of the breast pixels).

2.2. Extracted Features

The results of Fuzzy C-Means is, therefore, a breast divided into two clusters. A set of features for each class can be directly extracted. We used a set of morphological and texture features. As morphological features we simply calculated the relative area, the centre of masses and the medium intensity of both clusters, whilst as textures features, we calculated features derived from co-occurrence matrices [17].

Co-occurrence matrices are essentially two-dimensional histograms of the occurrence of pairs of grey-levels for a given displacement vector. Formally, the co-occurrence of grey levels can be specified as a matrix of relative frequencies $P_{ij}(d, \theta)$, in which two pixels separated by a distance d and angle θ have gray levels i and j. Co-occurrence matrices are not generally used as features, rather a large number of textural features derived from the matrix have been proposed [17]. Here we use 4 different directions: $0°, 45°, 90°,$ and $135°$; and a distance equal to 1. For each co-occurrence

		Automatic Classification			
		BIRADS I	BIRADS II	BIRADS III	BIRADS IV
Truth	BIRADS I	17	24	5	4
	BIRADS II	27	35	30	8
	BIRADS III	3	25	54	18
	BIRADS IV	3	9	23	15

(a)

		Automatic Classification			
		BIRADS I	BIRADS II	BIRADS III	BIRADS IV
Truth	BIRADS I	24	20	3	3
	BIRADS II	34	49	10	7
	BIRADS III	18	17	40	25
	BIRADS IV	9	5	19	17

(b)

Table 1. Confusion matrices of the k-NN classifier (a) and ID3 decision tree (b).

matrix we determine the contrast, energy, entropy, correlation, sum average, sum entropy, difference average, difference entropy, and homogeneity features.

2.3. Classification

We evaluated two different kind of classifiers: the k-Nearest Neighbours algorithm and a Decision Tree classifier. The k-Nearest Neighbours classifier [18] (kNN) consists of classifying a non-classified vector into the k most similar vectors presents in the training set. Because kNN is based on distances between sample points in feature space, features need to be re-scaled to avoid that some features are weighted much more strongly than others. Hence, all features have been normalised to unit variance and zero mean.

On the other hand, a decision tree recursively subdivides regions in feature space into different subspaces, using different thresholds in each dimension to separate the classes "as much as possible". For a given subspace the process stops when it only contains patterns of one class. In our implementation we used the ID3 information criterion [18] to determine thresholds values from the training data.

As was shown in Bovis and Singh [10], the combination of classifiers can reach better performances than a single one. Therefore, we use a fuzzy combination of the above two classifiers to construct a third one. The output of the above classifiers is a membership value for each class instead of a single class. The average combination of these membership values will be used to produce the final result.

For the kNN classifier, the membership value of a class is different from zero if there is at least one neighbour (of k possible neighbours) belonging to this class. The membership value for each class will be the sum of the inverse Euclidean distances among each neighbour of the class and the pattern. A final normalisation to one between all the membership values is required. On the other hand, on the ID3 decision tree, a new pattern is classified according the value of one of its features in each node. Instead of looking for one node alone (as is traditional), we calculated a fuzzy membership according to the distance between the corresponding value of the feature of the new pattern and the node. This value is carried forward to each child of the node. As a result, a value is associated with each tree leaf.

We consider this value as the distance between the leaf and the pattern. Therefore, the normalised sum of the inverses of the distances is used to give a membership value for each class.

3. EXPERIMENTAL RESULTS

The method was applied to a set of 300 Medio-Lateral Oblique right mammograms taken from the Digital Database of Screening Mammographies (DDSM) [19]. This database provides for each mammogram additional information, including the density of the breast determined by an expert according to BIRADS categories. In order to simulate the real world [20], our database is formed by 50 mammograms with BIRADS I and IV, and 100 with BIRADS II and III.

To evaluate the results we used the leave-one-image-out method, in which each sample is analysed by a classifier which is trained using all other samples. The confusion matrices for both classifiers are shown in Table 1. Confusion matrices should be read as follows: rows indicate the object to be recognised (the true class) and columns indicate the label the classifiers associates with this object. As should be clear, the ID3 decision tree classifier has a higher efficiency compared to the kNN classifier. This is due to the fact that the ID3 classifier contains a feature selection discrimination process. This ensures it avoids non-discriminant features to weight in the classification step. kNN classifiers do not have such feature selection and, therefore, the set of discriminant and non-discriminant features are weighted equally in the classification procedure.

Some previous works [10, 11] reduced the dimensionality of this four-class classification to a two-class problem. In this case, classes with low density (BIRADS I and II) are compared with classes with high density (BIRADS III and IV). Note that using this approach most of the incorrect classified mammograms in the four-class problems are now correctly classified. Moreover, we can see that for high density breasts the kNN has better performances than the ID3 tree, whilst for low dense breasts, the ID3 is better. This is the main reason to construct the fuzzy combination of them. With this approach, we reach the performances shown in Table 2. This shows an increase of the overall performance when compared to the individual classifiers. We

Automatic Classification			
BIRADS I	BIRADS II	BIRADS III	BIRADS IV

Confusion matrix with Truth rows:

		BIRADS I	BIRADS II	BIRADS III	BIRADS IV
Truth	BIRADS I	23	21	3	3
	BIRADS II	32	48	12	8
	BIRADS III	3	22	52	23
	BIRADS IV	5	7	20	18

Table 2. Confusion matrix of the fuzzy combination of the ID3 and kNN classifiers.

reach 47% of correct classification when combining classifiers, opposed to 43.3% for ID3, and 40.3% for kNN, respectively.

4. CONCLUSIONS AND FURTHER WORK

This paper has presented an automatic classification method for the identification of breast density in mammographic images. The method is based on the integration of texture and gray level information. An initial method based on gray-level information starts segmenting the profile of the breast. Subsequently, the Fuzzy C-Means algorithm is used to segment the different tissue types in the mammograms. Morphological and texture features are extracted to characterise the breast tissue for each cluster. Finally, kNN, ID3 and a fuzzy combination of them are used to classify the breast into BIRADS categories. Experimental results demonstrate the effectiveness of the proposed strategy. Further work will focus on the extraction of more texture features and the use of feature selection algorithms in order to provide a more reliable subset of features.

5. REFERENCES

[1] American Cancer Society, "Breast cancer: facts and figures. 2003-04," *ACS*, 2003.

[2] L. Basset and R. Gold, *Breast Cancer Detection: Mammograms and Other Methods in Breast Imaging*, Grune & Stratton, New York, 1987.

[3] R2 Technology Inc., "www.r2tech.com," 2004.

[4] W.T. Ho and P.W.T. Lam, "Clinical performance of computer-assisted detection (CAD) system in detecting carcinoma in breasts of different densities," *Clinical Radiology*, vol. 58, pp. 133–136, 2003.

[5] J.N. Wolfe, "Risk for breast cancer development determined by mammographic parenchymal pattern," *Cancer*, vol. 37, pp. 2486–2492, 1976.

[6] N. Karssemeijer and G. te Brake, "Combining single view features and asymmetry for detection of mass lesions," in *Int. Work. on Dig. Mammography*, June 1998, pp. 95–102.

[7] C. Zhou, H.P. Chan, N. Petrick, M.A. Helvie, M.M. Goodsitt, B. Sahiner, and L.M. Hadjiiski, "Computerized image analysis: Estimation of breast density on mammograms," *Med. Phys*, vol. 28, no. 6, pp. 1056–1069, 2001.

[8] P. Miller and S. Astley, "Classification of breast tissue by texture and analysis.," *Image and Vision Computing*, vol. 10, pp. 227–282, 1992.

[9] J.W. Byng, N.F. Boyd, E. Fishell, R.A. Jong, and M.J. Yaffe, "Automated analysis of mammographic densities," *Phys. Med. Biol.*, vol. 41, pp. 909–923, 1996.

[10] K. Bovis and S. Singh, "Classification of mammographic breast density using a combined classifier paradigm," in *Int. Work. on Dig. Mammography*, 2002, pp. 177–180.

[11] S. Petroudi, T. Kadir, and M. Brady, "Automatic classification of mammographic parenchymal patterns: A statistical approach," in *IEEE Conf. Eng. Med. Biol. Soc.*, 2003, vol. 2, pp. 416–423.

[12] R. Zwiggelaar, L. Blot, D. Raba, and E.R.E Denton, "Set-permutation-occurrence matrix based texture segmentation," in *Ib. Conf. on Patt. Rec. and Image Anal.*, June 2003, pp. 1099–1107.

[13] R. Zwiggelaar and E.R.E Denton, "Optimal segmentation of mammographic images," in *Int. Work. on Dig. Mammography*, June 2004.

[14] American College of Radiology, *Illustrated Breast Imaging Reporting and Data System BIRADS*, American College of Radiology, 3rd edition, 1998.

[15] D. Raba, A. Oliver, J. Martí, M. Peracaula, and J. Espunya, "Breast segmentation with pectoral muscle suppression on digital mammograms," in *Ib. Conf. on Patt. Rec. and Image Anal.*, Estoril, Portugal, June 2005, pp. 471–478.

[16] J.C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*, Plenum Press, New York, 1981.

[17] R.M. Haralick, K.S. Shanmugan, and I. Dunstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. 3, no. 6, pp. 610–621, 1973.

[18] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, John Wiley & Sons, New York, 2 edition, 2001.

[19] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P.J. Kegelmeyer, "The digital database for screening mammography," in *Int. Work. on Dig. Mammography*, June 2000.

[20] P.B. Lam, P.M. Vacek, B.M. Geller, and H.B. Muss, "The association of increased weight, body mass index, and tissue density with the risk of breast carcinoma in Vermont," *Cancer*, vol. 89, no. 2, pp. 369–375, 2000.