Automatic classification of breast tissue

A. Oliver[†], J. Freixenet[†], A. Bosch[†], D. Raba[†], and R. Zwiggelaar[‡]

[†] Institute of Informatics and Applications, University of Girona, Spain
[‡] Department of Computer Science, University of Wales, Wales, UK
{aoliver,jordif,aboschr,draba}@eia.udg.es;rrz@aber.ac.uk

Abstract. A recent trend in digital mammography are CAD systems, which are computerized tools designed to help 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 is increased. In addition, the suitability of abnormality segmentation approaches tends to depend on breast tissue density. In this paper we propose a new approach to the classification of mammographic images according to the breast parenchymal density. Our classification is based on gross segmentation and the underlying texture contained within the breast tissue. Robustness and classification performance are evaluated on a set of digitized mammograms, applying different classifiers and leave-one-out for training. Results demonstrate the feasibility of estimating breast density using computer vision 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 study developed in 1998 by the American Cancer Society estimates that in western cultures between one in eight and one in twelve women will develop breast cancer during their lifetime. Breast cancer remains the leading cause of death for women in their 40s in the United States [1]. However, although breast cancer incidence has increased over the past decade, breast cancer mortality has declined among women of all ages. This favorable trend in mortality reduction may relate to the widespread adoption of mammography screening, in addition to improvements made in therapy [1].

Mammography remains the key screening tool for breast abnormalities detection, because it allows identification of tumour before being palpable. In a recent study, Vacek et al. [2] show that the proportion of breast tumours that were detected in Vermont by screening mammography increased from 2% during $1974 - 1984$ to 36% during $1995 - 1999$. 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 with the difficulty of obtaining accurate diagnosis [3]. In this sense, computerized image analysis is going to play an important role in improving the issued diagnosis. Computer-Aided Diagnosis (CAD) systems are composed of a set of tools to help radiologists to detect and diagnose new cases. 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].

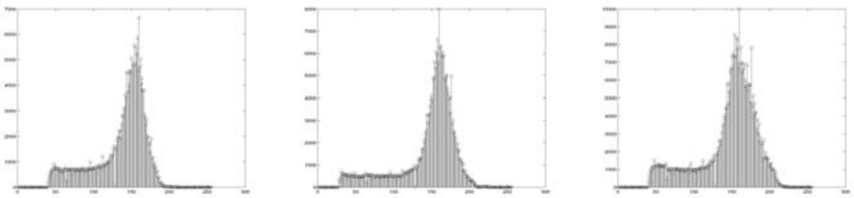
The origins of breast parenchymal classification are found in the work of Wolfe [5], who showed the relation between mammographic parenchymal patterns and the risk of developing breast cancer, classifying the parenchymal patterns in four categories. Since the discovery of this relationship automated parenchymal pattern classification has been investigated. Boyd et al. [6] proposed a semiautomatic computer measure based on interactive thresholding and the percentage of the segmented dense tissue over the segmented breast area. Karssemeijer [7] developed an automated method where features were computed from the grey-level values and the distance from the skin-line and are used in a k-Nearest Neighbour (kNN) classifier. Recently, Zhou et al. [8] proposed a rule-based scheme in order to classify the mammograms in four classes according to the characteristic features of the gray-level histograms.

A small number of previous papers have suggested texture representations of the breast. Miller and Astley [9] investigated texture-based discrimination between fatty and dense breast types applying granulometric techniques and Laws texture masks. Byng et al. [10] used measures based on fractal dimension. Bovis and Singh [11] estimated features from the construction of Spatial Gray Level Dependency matrices. Recently, Petroudi et al. [12] used textons to capture the mammographic appearance within the breast area. The approach developed by Blot and Zwiggelaar [13] is based on the statistical difference between local and median co-occurrence matrices computed over three regions of the breast. Related to this work, Zwiggelaar et al. [14] estimated the breast density by using co-occurrence matrices as features and segmentation based on the Expectation-Maximization algorithm. PCA was used to reduce the dimensionality of the feature space. This work was extended in [15] were a transportation algorithm was used for the feature selection process.

Our approach is also based on grouping those pixels with similar behaviour (gross segmentation), in our case consisting of similar tissue. Subsequently, texture features extracted from each region are used to classify the whole breast in one of the three categories that appears in the MIAS database [16]: fatty, glandular or dense breast. 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 appear in Section 3. Finally, conclusions are given in Section 4.

2 Methodology

As we have mentioned in the introduction, previous work has used histogram information to classify breast tissue. However, in our experience and with our database, histogram information is not sufficient to classify the mammogram as



(a) (b) (c)

Fig. 1. Three similar histograms, each belonging to a different class of tissue. Concretely, (a) corresponds to a fatty breast, (b) to a glandular breast and (c) to a dense breast.

fatty, glandular or dense tissue. Figure 1 shows histograms for three different mammograms, each belonging to a different class.

Our approach is based on gross segmentation and the extraction of texture features of those pixels with similar tissue appearance of the breast. Using this set of features we train different classifiers and test them. But first of all, our approach begins with the segmentation of the profile of the breast.

2.1 Breast Profile Segmentation

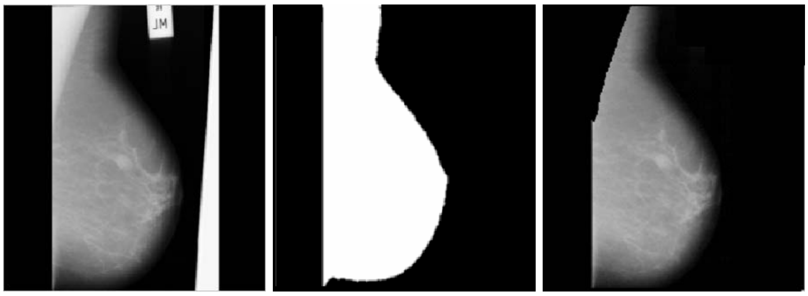
The segmentation of the foreground breast object from the background is a fundamental step in mammogram analysis. The aim of this process is to separate the breast from the rest of objects that could appear in a digital mammography: the black background; some annotations or labels; and the pectoral muscle.

In this work we used a previous developed algorithm [17] based on gray-level information. This algorithm begins by finding a threshold using histogram information, in order to separate the background from the rest of the objects of the mammogram, that is, the annotations and the union of the breast and pectoral muscle. The breast and pectoral muscle object is segmented looking for the largest object in the image. In order to separate the breast profile from the pectoral muscle we used an adaptive region growing approach, initializing the seed inside the pectoral muscle, and controlling the growing step using information about gray-level and the growth area.

Figure 2(a) shows a typical mammographic image. Applying the threshold and detecting the largest object, the union of the pectoral muscle and the breast area is found(Figure 2(b)). In the last image, the region of interest of the breast has been extracted from the pectoral muscle using the adaptive region growing algorithm described above.

2.2 Finding Regions with Similar Tissue

We consider that pixels from similar tissue have similar gray-level values, as can be seen in Figure 3. Hence, we use the k-Means algorithm [18] to group these



(a) (b) (c)

Fig. 2. Sequence of the breast profile segmentation. (a) original image, (b) result of thresholding the image and detecting the largest region, and (c) segmented image without background and pectoral muscle.

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.

The k-Means algorithm [18] is a popular clustering algorithm. It is defined as an error minimization algorithm where the function to minimize is the sum of errors squared:

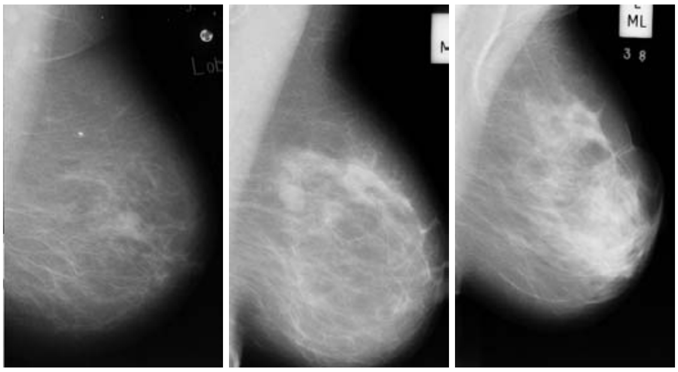
$$e^2(K) = \sum_{k=1}^{K} \sum_{i \in C_k} (x_i - c_k)^2 \tag{1}$$

where x_i are feature vectors, c_k is the centroid of cluster C_k, and K the number of clusters, which have to be known a priori. In our work, we selected $K = 2$ with the aim to obtain representative instances of two classes: fatty tissue and dense tissue.

When using the k-Means algorithm, the placement of the initial seed points plays a central role in obtaining the final segmentation results. Despite their importance, seeds are usually initialized randomly. In order to make a more informed decision, in our approach, the k-Means is initialized using histogram information. We initialized the two seeds with the gray level values that represent 15% and 85% of the accumulative histogram, with the objective to cluster fatty and dense tissue, respectively.

2.3 Extracted Features

Subsequent to k-Means, a set of features from the two classes that form the breast are extracted. In fact, a simple view of the feature space shows that using only the morphological features (like the centers of mass of both classes), is enough to distinguish between dense and fatty tissue. However, in order to distinguish glandular tissue, texture features needs to be considered. We used features derived from co-occurrence matrices [19].



(a) (b) (c)

Fig. 3. Examples of different types of breast tissue in the MIAS database [16]. (a) fatty, (b) glandular, and (c) dense.

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}, 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 [19]. Here we use 4 different directions: $0°$, $45°$, $90°$, and $135°$; and a distance equal to 1. For each co-occurrence matrix we determine the contrast, energy, entropy, correlation, sum average, sum entropy, difference average, difference entropy, and homogeneity features.

2.4 Classification

We evaluated two different kind of classifiers: the k-Nearest Neighbours algorithm and a Decision Tree classifier. The k-Nearest Neighbours classifier [20] (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 normalized 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 separates 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 [20] to determine thresholds values from the training data.

3 Experimental Results

The method was applied on a set of 270 mammograms taken from the MIAS database, 90 of each class (fatty, glandular and dense). The spatial resolution of the images is 50μm x 50μm and the optical density is linear in the range $0 - 3.2$ and quantized to 8 bits. To evaluate the method we performed three experiments.

The first experiment was performed over the set of fatty and dense mammograms, and using only morphological features extracted from the segmented clusters. We calculated the relative area, the center of masses and the medium intensity of both clusters. These features formed the input parameters for the classification stage. In order to evaluate the results, we used a leave-one-out method, in which each sample is analysed by a classifier which is trained using all other samples except for those from the same woman. The results showed that 87% and 82% of mammograms were correctly classified using the kNN classifier and the ID3 decision tree. However, when including the glandular class, both results were drastically decreased.

(a)

		Automatic Classification		
		Fatty	Glandular	Dense
Truth	Fatty	23	6	1
	Glandular	2	22	6
	Dense	1	14	15

(b)

		Automatic Classification		
		Fatty	Glandular	Dense
Truth	Fatty	18	10	2
	Glandular	4	21	5
	Dense	2	4	24

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

The second experiment was performed using 30 cases per class, and using the morphological features cited above as well the texture features. The efficiency of the classifiers were computed using again the same leave-one-out approach. Experimental results showed that classification results were improved when the cluster means were subtracted from the feature vectors. The reason for this can be found in the fact that increasing the dense area of the breast, results in a larger difference between the two tissue type clusters.

The confusion matrices for both classifiers are shown in Table 1. Confusion matrices should be read as follows: rows indicate the object to recognize (the true class) and columns indicate the label the classifiers associates at this object. As should be clear, the ID3 decision tree classifier has in general, 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. We can also note in Table 1 that mammograms belonging to fatty class are better classified than the rest of mammograms when using the kNN approach. On the other hand, dense mammograms are better classified by the ID3 approach.

The last experiment was performed by using a set of 90 mammograms per class, from which 30 were manually extracted from the set for training both classifiers, while the rest of mammograms constituted the testing set. The confusion matrices for this classification are shown in Table 2. Note that the ID3 classifier have again better results than the kNN. Moreover, it can be seen that the percentage of well-classified mammograms drastically decrease in comparative with the previous experiment. Concretely, in the leave-one-out method, the accuracy of the system was around 67% and 73% for the kNN and the ID3 classifiers respectively, while in the second experiment, the accuracy is about 56% and 61%. The main reason for this is likely to be the enlarged variance for the different mammographic tissue types.

		Automatic Classification		
		Fatty	Glandular	Dense
Truth	Fatty	38	19	3
	Glandular	9	33	18
	Dense	5	25	30

(a)

		Automatic Classification		
		Fatty	Glandular	Dense
Truth	Fatty	34	17	9
	Glandular	5	35	20
	Dense	2	18	40

(b)

Table 2. Confusion matrices of (a) the k-NN classifier and (b) ID3 decision tree.

4 Conclusions

This paper has presented an automatic classification method for the identification of breast tissue 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 k-Means algorithm is used to segment the different tissue types of the mammograms. Morphological and texture features are extracted in order to characterize the breast tissue for each cluster. Finally, k-NN and ID3 are used to classify the breast as dense, fatty or glandular. Experimental results demonstrate the effectiveness of the proposed algorithm. Compared to published work, we can say that the developed method has a similar performances.

Further work will be focused on the characterization of the mammographic tissue in four classes, as are described in the Breast Imaging Reporting and Data System (BI-RADS). Moreover, other databases will be tested in order to evaluate in depth the efficiency of our proposal.

References

1. Buseman, S., Mouchawar, J., Calonge, N., Byers, T.: Mammography screening matters for young women with breast carcinoma. Cancer **97** (2003) 352–358

2. Vacek, P., Geller, B., Weaver, D., Foster, R.: Increased mammography use and its impact on earlier breast cancer detection in vermont. Cancer **94** (2002) 2160–2168
3. Basset, L., Gold, R.: Breast Cancer Detection: Mammograms and Other Methods in Breast Imaging. Grune & Stratton, New York (1987)
4. Ho, W., Lam, P.: Clinical performance of computer-assisted detection (cad) system in detecting carcinoma in breasts of different densities. Clinical Radiology **58** (2003) 133–136
5. Wolfe, J.: Risk for breast cancer development determined by mammographic parenchymal pattern. Cancer **37** (1976) 2486–2492
6. Boyd, N., Byng, J., Jong, R., Fishell, E., Little, L., Miller, A., Lockwood, G., Tritchler, D., Yaffe, M.: Quantitative classification of mammographic densities and breast cancer risk: results from the canadian national breast screening study. J. Natl Cancer Inst. **87** (1995) 670–675
7. Karssemeijer, N.: Automated classification of parenchymal patterns in mammograms. Physics in Medicine and Biology **43** (1998) 365–378
8. Zhou, C., Chan, H., Petrick, N., Helvie, M., Goodsitt, M., Sahiner, B., Hadjiiski, L.: Computerized image analysis: Estimation of breast density on mammograms. Medical Physics **28** (2001) 1056–1069
9. Miller, P., Astley, S.: Classification of breast tissue by texture and analysis. Image and Vision Computing **10** (1992) 227–282
10. Byng, J., Boyd, N., Fishell, E., Jong, R., Yaffe, M.: Automated analysis of mammographic densities. Physics in Medicine and Biology **41** (1996) 909–923
11. Bovis, K., Singh, S.: Classification of mammographic breast density using a combined classifier paradigm. In: International Workshop on Digital Mammography. (2002) 177–180
12. Petroudi, S., Kadir, T., Brady, M.: Automatic classification of mammographic parenchymal patterns: A statistical approach. In: International Conference of the IEEE Engineering in Medicine and Biology Society. Volume 2. (2003) 416–423
13. Blot, L., Zwiggelaar, R.: Background texture extraction for the classification of mammographic parenchymal patterns. In: Medical Image Understanding and Analysis. (2001) 145–148
14. Zwiggelaar, R., Blot, L., Raba, D., Denton, E.: Set-permutation-occurrence matrix based texture segmentation. In: Iberian Conference on Pattern Recognition and Image Analysis. (2003) 1099–1107
15. Zwiggelaar, R., Denton, E.: Optimal segmentation of mammographic images. In: International Workshop on Digital Mammography. (2004)
16. Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., Taylor, P., Betal, D., Savage, J.: The mammographic image analysis society digital mammogram database. In: International Workshop on Digital Mammography. (1994) 211–221
17. Raba, D., Oliver, A., Martí, J., Peracaula, M., Espunya, J.: Breast segmentation with pectoral muscle suppression on digital mammograms. In: Iberian Conference on Pattern Recognition and Image Analysis. (2005) to appear
18. MacQueen, J.: Some methods of classification and analysis of multivariate observations. In: Berkeley Symposium on Mathematical Statistics and Probability. Volume 1. (1967) 281–297
19. Haralick, R., Shanmugan, K., Dunstein, I.: Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics **3** (1973) 610–621
20. Duda, R., Hart, P., Stork, D.: Pattern Classification. 2 edn. John Wiley & Sons, New York (2001)