Classifying mammograms using texture information

Arnau Oliver[a]*, Xavier Lladó[a], Robert Martí[a], Jordi Freixenet[a], and Reyer Zwiggelaar[b]

[a]Institute of Informatics and Applications, University of Girona, 17071, Girona, Spain
[b]Department of Computer Science, University of Wales, Aberystwyth SY23 3DB, UK

Abstract. In an ongoing effort to assist radiologists in detecting breast cancer early, this paper focuses on breast characterisation according to internal tissue characteristics. This is an important feature because it has been demonstrated that women with dense breasts are more likely to suffer breast cancer, and also, the performance of automatic mass detection methods decreases in dense breasts. The strategy of our proposal firstly identifies regions with similar grey-level by using a clustering strategy. Subsequently, texture descriptors are extracted from each cluster by using Local Binary Patterns and Co-occurrence Matrices, and finally used to train a classifier. Results obtained from the complete MIAS database and using a leave-one-out strategy show a correct classification of 78% when compared to expert assessment.

1 Introduction

Mammographic Computer Aided Diagnosis (CAD) systems are being developed to assist radiologists in the evaluation of mammographic images [1]. However, recent studies have shown that CAD performance decreases as the density of the breast increases, either decreasing the sensitivity [2] (the accuracy in which the lesions are detected) or decreasing the specificity [3] (increasing the number of regions being normal tissue but marked by the automatic systems as suspicious lesions). In addition, it is well-known that there is a strong positive correlation between breast parenchymal density in mammograms and the risk to develop breast cancer [4]. Therefore, as Taylor [5] suggested, the development of automatic methods for classification of breast tissue are justified by, at least, two factors:

- To permit a better use of the time and skills of expert radiologists by allowing difficult mammograms to be examined by the most experienced readers.
- To increase the scope for CAD of abnormalities by concentrating on the easier (fatty) mammograms.

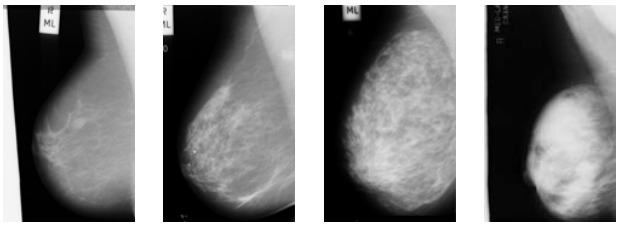
Classification in mammographic risk assessment can be based on a number of categories which might not describe the same mammographic features [6]. However, the American College of Radiology Breast Imaging Reporting and Data System (BIRADS) [7] is becoming a standard on the assessment of mammographic images and uses four categories for density evaluation:

- BIRADS I: the breast is almost entirely fatty,
- BIRADS II: there is some fibro-glandular tissue,
- BIRADS III: the breast is heterogeneously dense,
- BIRADS IV: the breast is extremely dense.

Figure 1 shows example mammograms of each class (the mammograms are extracted from the MIAS database [8]). Note how the internal density of the breasts increases from BIRADS I (left) to BIRADS IV (right). It should be noted that besides density these BIRADS classes also included patterns that can be described as various textures. As such, it seems appropriate to include both aspects in an automatic classification approach.

Although BIRADS is becoming the radiologic standard, it has not been commonly used in the evaluation of automatic breast density classification approaches. Exceptions to this are the works of Bovis and Singh [9], Petroudi et al. [10], and our previous work [11–13]. Bovis and Singh [9] extracted features from the global breast and used a combination of classifiers for training and testing the system. Petroudi et al. [10] used textons to obtain a visual dictionary for breast classification. This work was extended in [12] adding also SIFT features, although we found that textons outperformed them. The strategy we used here is similar to the work presented in [11,13], which in turn follows ideas from Bovis and Singh [9] and more straightforward compared to Petroudi et al. [10] and Bosch et al. [12]. We used the Fuzzy C-Means [14] for clustering the pixels with similar grey-level values into two classes and, subsequently, a set of features derived from co-occurrence matrices was extracted from each cluster. In [11] the used classifier was a Bayesian combination of kNN and a decision tree. Here, we use two more stable clustering algorithms for grouping the tissue with similar grey-level. In addition to using texture features extracted from the co-occurrence matrices, we also propose to use the Local Binary Patterns [15] for extracting the texture information. To our knowledge this is the first attempt to use LBP in the mammographic field analysis.

*Corresponding author, e-mail: aoliver@eia.udg.es



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

Figure 1. Example mammograms, where the breast density increases from (a) BIRADS I to (d) BIRADS IV.

2 Clustering the tissue with similar appearance

As a first step, all mammograms are pre-processed to identify the breast region. Hence, the background, labels, and pectoral muscle are removed from the image. We used a previous developed approach [16] which results in a minor loss of skin-line pixels in the breast area, but those pixels are deemed not to be relevant for tissue estimation and the relative number of potentially affected pixels is small.

Subsequently, using a clustering algorithm, the pixels with similar grey-level values are grouped together. We use a clustering algorithm in order to obtain only two clusters of pixels, representing fatty tissue (the cluster with lower mean intensity) and dense tissue (the cluster with higher mean intensity). In previous works we showed that the Fuzzy C-Means algorithm outperforms the results obtained using other segmentation strategies [13]. In this work we want to compare it with two alternative clustering algorithms, which are the Normalised Cuts [17] and the Mean Shift [18] algorithms. All these approaches are summarised below:

- Fuzzy C-Means clustering allows each pattern of the image to be associated with every cluster using a fuzzy membership function. In our implementation, the function minimised by the algorithm is defined by:

$$e^2(\Xi, U) = \sum_{n=1}^{N} \sum_{t=1}^{T} u_{nt} ||p_n - c_t||^2 \tag{1}$$

  where Ξ is the partition of the image, U is the membership matrix: u_{nt} represents the membership of pattern p_n to belong to cluster t, which is centred at $c_t = \frac{\sum_{n=1}^{N} u_{nt} p_n}{\sum_{n=1}^{N} u_{nt}}$, N is the number of patterns in the whole image (i.e. the number of pixels), and T the number of clusters, which has to be known a priori. Instead of randomly initialising the algorithm, the two cluster centres were initialised with the grey-level values that represent 15% and 85% of the accumulative histogram of the mammogram pixels, respectively representing fatty and dense tissue.

- The Normalised Cuts algorithm [17] deals with segmentation as a graph partitioning problem, where each node represents a pixel and the edges the pixel neighbourhood. Given a similarity measure between each pair of points in the set, it tries to group together points having large affinity between them. Namely, the algorithm consists in four steps: 1. Given an image build a weighted graph and set the weight (w) on the edge connecting two nodes to be a measure of their similarity. 2. Solve $(D - W)x = \lambda Dx$ as an eigen decomposition problem, where D is a $N \times N$ diagonal matrix with $d(i) = \sum_j w(i, j)$. 3. Use the eigenvector with the second smallest eigenvalue to bipartition the graph. 4. Decide if the current partition should be subdivided and recursively divide the segmented parts if necessary. Note that as we only need two clusters, this last step is not necessary.

- The Mean Shift algorithm [18] is a robust, non-parametric density estimation based clustering approach for image segmentation. It assumes that the feature space can be regarded as an empirical probability density function where dense regions correspond to the modes of the unknown density. A two-step kernel-based iterative procedure is used for the localisation of the modes. First, the mean shift vector m is computed as:

$$m(x) = \frac{\sum_{i=1}^{n} x_i g(||\frac{x - x_i}{h}||^2)}{\sum_{i=1}^{n} g(||\frac{x - x_i}{h}||^2)} - x \tag{2}$$

  where x is the actual mean, x_i all the points in the kernel, h the bandwidth (the size of the kernel), and g is the profile of the used kernel (we used a flat kernel). And secondly, the centre of the kernel is translated by $m(x)$. Both steps are repeated until convergence.

3 Characterising the breast tissue

Once the pixels with the same grey-level have been grouped in the fatty and dense clusters, we extract a set of features from both regions, and a feature vector is constructed by concatenating the fatty and dense features. In particular, we test in this paper the use of Local Binary Patterns (LBP) [15] and the use of features extracted from co-occurrence matrices [19].

The LBP operator labels the pixels of an image by thresholding the neighbourhood of each pixel with the centre value and considering the result of this thresholding as a binary number. When all the image pixels have been labeled with the corresponding LBP codes, the histogram of the labels is computed and used as a texture descriptor. However, this representation is not directly applicable to our work, because we need to characterise the clusters instead of the full image. Thus, we slightly modify the LBP in two different ways. Firstly, those pixels with a neighbourhood belonging to both clusters are not considered when extracting the texture. And secondly, instead of extracting one histogram from each mammogram, we extract two histograms per mammogram: one for the fatty cluster and the other one for the dense cluster. Hence, the concatenation of both histograms is taken as the full description of a mammogram.

Co-occurrence matrices are essentially two-dimensional histograms of the occurrence of pairs of gray-levels for a given displacement vector. Formally, the co-occurrence of gray-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. The full co-occurrence matrices are not generally used as features (mainly due to their high dimensionality and potential sparseness), but instead a large number of features derived from such matrices have been proposed [19].

4 Classifying the features

Once the features are extracted, a classifier is trained and used in a posterior step to classify new mammograms. In this paper we test the use of four different classifiers [20]: kNN in combination with SFS feature selection, Fisher discriminant analysis using a linear discriminant feature selection algorithm, C4.5 decision tree classifier, and support vector machine.

5 Results

We test our approach using all 322 mammograms of the MIAS database [8], classified in BIRADS categories using the majority opinion of three experts (in the 12 cases where there was complete disagreement, the intermediate class was used). The results are obtained using a leave-one-woman-out strategy, where a query mammogram is tested by a classifier trained on the rest of the mammograms belonging to different women, and this procedure is repeated until all the mammograms have been used as a query image.

In these experiments we tested different parameters for LBP (we refer to the work of Ojala et al. [15] for details), including the mapping (uniform, rotation-invariant, and uniform rotation-invariant), neighbourhood (4, 8, and 16), and distance (1,2,4). In what follows we used the best results experimentally obtained, which were the ones using the uniform rotation-invariant mapping and 8-neighbourhood with distances 1 and 2, and 16-neighbourhood with distances 1 and 4. The complete size of the used descriptor was 112. We also tested different parameters for the co-occurrence matrices. In this paper we use four different directions: $0°$, $45°$, $90°$, and $135°$, and three distances equal to 1, 5, and 9 pixels. Note that these values were empirically determined in our experiments and are related to the scale of the texture features found in mammographic images. For each co-occurrence matrix the following features were used: contrast, energy, entropy, correlation, sum average, sum entropy, difference average, difference entropy, and homogeneity features. As each of these features is extracted from each class, we deal with 216 co-occurrence features in total.

5.1 Segmentation comparison

In this section we used the same parameters for LBP and the same classifier (kNN+SFS) while varying the segmentation algorithm used for segmenting the fatty and dense regions. This allows us to compare how the segmentation strategy modifies the final results.

Table 1 shows the results obtained in each case, obtaining 65%, 78%, 73%, and 77% correct classification when no segmentation, Fuzzy C-Means, Normalised Cuts, and Mean Shift segmentation algorithms were used respectively.

Note that the best performances are obtained when using the Fuzzy C-Means algorithm and the Mean Shift, while Normalised Cuts obtained intermediate results. The no segmentation strategy provided the worst results. This shows the benefits of the proposed segmentation step for clustering the tissue with similar grey-level. On the other hand, the low performance of the Normalised Cuts algorithm is due to the fact that the affinity function of this algorithm takes the distance between pixels into account, which results in filled-in regions in the images. However, in this not the aim of our work, because usually there are some small clusters of dense tissue inside large regions of fatty tissue, and viceversa. These regions are effectively detected using the Fuzzy C-Means and the Mean Shift algorithms.

		Automatic			
		B-I	B-II	B-III	B-IV
Manual	B-I	71	13	1	2
	B-II	14	70	17	2
	B-III	4	24	57	10
	B-IV	1	7	17	12

(a)

		Automatic			
		B-I	B-II	B-III	B-IV
Manual	B-I	73	13	0	1
	B-II	7	77	17	2
	B-III	0	14	75	6
	B-IV	1	1	9	26

(b)

		Automatic			
		B-I	B-II	B-III	B-IV
Manual	B-I	70	13	4	0
	B-II	9	70	21	3
	B-III	2	16	69	8
	B-IV	1	4	6	26

(c)

		Automatic			
		B-I	B-II	B-III	B-IV
Manual	B-I	75	6	4	2
	B-II	8	83	10	2
	B-III	0	18	70	7
	B-IV	1	2	14	20

(d)

Table 1. Confusion matrices obtained when (a) no segmentation, (b) Fuzzy C-Means, (c) Normalised Cuts, and (d) Mean Shift were used in the segmentation of fatty and dense tissue. B-I stands for BIRADS I, etc.

5.2 Features comparison

The aim of this section is to demonstrate the usefulness of computing texture information. Thus, the Fuzzy C-Means is used as a segmentation strategy, kNN+SFS as the classifier, and we compare the results obtained using LBP with those obtained using co-occurrence matrices. Table 2 shows the confusion matrices for both cases, obtaining 79% correct classification when using features derived from co-occurrence matrices and 78% for LBP. Note that co-occurrence matrices discriminate better low dense breasts, while Local Binary Patterns performs better the discrimination in dense breasts. The difference between these two results is not significant and, in fact, these results can well be influenced by the different dimensionality of the two set of features and to evaluate this is part of our future research.

		Automatic			
		B-I	B-II	B-III	B-IV
Manual	B-I	75	8	1	3
	B-II	7	77	15	4
	B-III	1	10	79	5
	B-IV	3	3	7	24

(a)

		Automatic			
		B-I	B-II	B-III	B-IV
Manual	B-I	73	13	0	1
	B-II	7	77	17	2
	B-III	0	14	75	6
	B-IV	1	1	9	26

(b)

Table 2. Confusion matrices obtained when (a) using co-occurrence matrices and (b) using Local Binary Patterns for texture extraction.

5.3 Classifiers comparison

Finally, we tested here the use of different classifiers. Hence, the same segmentation algorithm is used (the Fuzzy C-Means) and the same set of LBP features as it allows obtaining similar results with lower computational cost.

Table 3 shows the confusion matrices for kNN with SFS, linear discriminant analysis, C4.5 decision tree, and support vector machine with a polynomial kernel, obtaining respectively, 78%, 65%, 68%, and 65% correct classification. The high performance of the kNN+SFS algorithm compared to the other algorithms shows the difficulty of the problem, being hard to find a real boundary separating the four classes (for the SVM we used a 10-fold cross-validation instead of the leave-one-out to avoid overfitting). The unexpected poor performance of SVM is an aspect which will need further investigation.

		Automatic			
		B-I	B-II	B-III	B-IV
Manual	B-I	73	13	0	1
	B-II	7	77	17	2
	B-III	0	14	75	6
	B-IV	1	1	9	26

(a)

		Automatic			
		B-I	B-II	B-III	B-IV
Manual	B-I	78	9	0	0
	B-II	15	59	28	1
	B-III	0	31	54	10
	B-IV	2	2	15	18

(b)

		Automatic			
		B-I	B-II	B-III	B-IV
Manual	B-I	74	10	1	2
	B-II	14	65	22	2
	B-III	2	29	56	8
	B-IV	2	2	8	25

(c)

		Automatic			
		B-I	B-II	B-III	B-IV
Manual	B-I	77	9	0	1
	B-II	16	55	31	1
	B-III	0	35	50	10
	B-IV	1	4	14	18

(d)

Table 3. Confusion matrices obtained when using (a) kNN classifier (k=3), (b) Fisher Discriminant Analysis, (c) C4.5 Decision Tree, and (d) Support Vector Machine.

6 Conclusions

In this paper we have quantitatively compared the use of Fuzzy C-Means, Normalised Cuts, and Mean Shift for grouping the tissue with similar grey-level, as a first step of a full strategy for classifying the breasts according their internal density. We have found that, due to the internal nature of the algorithm (the use of the distance between pixels into the affinity function), the performance of the Normalised Cuts is lower compared to the use of both other algorithms. On the other hand, Mean Shift and Fuzzy C-Means obtained similar performance, around 78% correct classification. We consider these results in-line with the expected results, because when comparing the agreement between each individual expert annotations and the consensus opinion we obtained 78%, 89%, and 72% agreement.

In addition, we effectively tested the use of Local Binary Patterns and co-occurrence matrices for describing the breast tissue textural information. Results obtained from the complete MIAS database and using a leave-one-woman-out strategy show that LBP and co-occurrence matrices features provide similar overall results, although LBP performs better in dense breasts while co-occurrence matrices in fatty breasts. Future work will concentrate on further evaluation of variation in these features.

Acknowledgements

This work was partially supported by MEC grant nb. TIN2006-08035 and grant IdIBGi-UdG 91060080. We also thank Dr. Josep Pont and Dr. Elsa Pérez from the Hospital Josep Trueta of Girona (Spain) and Dr. Erika R.E. Denton from the Norfolk and Norwich University Hospital for providing the ground-truth data used in this paper.

References

1. T. W. Freer & M. J. Ulissey. "Screening mammography with computer-aided detection: Prospective study of 12860 patients in a community breast center." *Radiology* **220**, pp. 781–786, 2001.
2. S. Obenauer, C. Sohns, C. Werner et al. "Impact of breast density on computer-aided detection in full-field digital mammography." *J. Digit. Imaging* **19(3)**, pp. 258–263, 2006.
3. R. F. Brem, J. W. Hoffmeister, J. A. Rapelyea et al. "Impact of breast density on computer-aided detection for breast cancer." *Am. J. Roentgenol.* **184(2)**, pp. 439–444, 2005.
4. J. N. Wolfe. "Risk for breast cancer development determined by mammographic parenchymal pattern." *Cancer* **37**, pp. 2486–2492, 1976.
5. P. Taylor, S. Hajnal, M. H. Dilhuydy et al. "Measuring image texture to separate "difficult" from "easy" mammograms." *Brit. J. Radiol.* **67**, pp. 456–463, 1994.
6. I. Muhimmah, A. Oliver, E. R. E. Denton et al. "Comparison between Wolfe, Boyd, BI-RADS and Tabár based mammographic risk assessment." In *Int. Work. Dig. Mammography*, pp. 407–415. 2006.
7. American College of Radiology. *Illustrated Breast Imaging Reporting and Data System BIRADS*. American College of Radiology, third edition, 1998.
8. J. Suckling, J. Parker, D. R. Dance et al. "The Mammographic Image Analysis Society digital mammogram database." In *Int. Work. Dig. Mammography*, pp. 211–221. 1994.
9. K. Bovis & S. Singh. "Classification of mammographic breast density using a combined classifier paradigm." In *Med. Image Underst. Anal.*, pp. 177–180. 2002.
10. S. Petroudi, T. Kadir & M. Brady. "Automatic classification of mammographic parenchymal patterns: A statistical approach." In *IEEE Conf. Eng. Med. Biol. Soc.*, volume 1, pp. 798–801. 2003.
11. A. Oliver, J. Freixenet & R. Zwiggelaar. "Automatic classification of breast density." In *IEEE Int. Conf. Image Proc.*, volume 2, pp. 1258–1261. 2005.
12. A. Bosch, X. Muñoz, A. Oliver et al. "Modeling and classifying breast tissue density in mammograms." In *IEEE Conf. Comput. Vision Patt. Rec.*, volume 2, pp. 1552–1558. 2006.
13. A. Oliver, J. Freixenet, R. Martí et al. "A comparison of breast tissue classification techniques." In *Int. Conf. Med. Image Comput. Comput. Assist. Interv.*, volume 2, pp. 872–879. 2006.
14. J. C. Bezdek. *Pattern Recognition With Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981.
15. T. Ojala, M. Pietikäinen & T. Mäenpää. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." *IEEE Trans. Pattern Anal. Machine Intell.* **24(7)**, pp. 971–987, 2002.
16. D. Raba, A. Oliver, J. Martí et al. "Breast segmentation with pectoral muscle suppression on digital mammograms." In *Lect. Not. Comp. Sc.*, volume 3523, pp. 471–478. 2005.
17. J. Shi & J. Malik. "Normalized cuts and image segmentation." *IEEE Trans. Pattern Anal. Machine Intell.* **22(8)**, pp. 888–905, 2000.
18. D. Comaniciu & P. Meer. "Mean sift: a robust approach toward feature space analysis." *IEEE Trans. Pattern Anal. Machine Intell.* **24(5)**, pp. 603–619, 2002.
19. R. M. Haralick, K. S. Shanmugan & I. Dunstein. "Textural features for image classification." *IEEE Trans. Syst., Man, Cybern.* **3(6)**, pp. 610–621, 1973.
20. R. O. Duda, P. E. Hart & D. G. Stork. *Pattern Classification*. John Wiley & Sons, New York, second edition, 2001.