

# Breast Density Segmentation: A Comparison of Clustering and Region Based Techniques

A. Torrent<sup>1</sup>, A. Bardera<sup>2</sup>, A. Oliver<sup>1</sup>, J. Freixenet<sup>1</sup>, I. Boada<sup>2</sup>, M. Feixes<sup>2</sup>, R. Martí<sup>1</sup>, X. Lladó<sup>1</sup>, J. Pont<sup>3</sup>, E. Pérez<sup>3</sup>, S. Pedraza<sup>3</sup>, and J. Martí<sup>1</sup>

<sup>1</sup> Computer Vision and Robotics Group, University of Girona, Catalonia, Spain  
aoliver@eia.udg.edu

<sup>2</sup> Graphics & Imaging Laboratory, University of Girona, Catalonia, Spain

<sup>3</sup> Department of Radiology, Hospital Josep Trueta of Girona, Catalonia, Spain

**Abstract.** This paper presents a comparison of two clustering based algorithms and one region based algorithm for segmenting fatty and dense tissue in mammographic images. This is a crucial step in order to obtain a quantitative measure of the density of the breast. The first algorithm is a multiple thresholding algorithm based on the excess entropy, the second one is based on the Fuzzy C-Means clustering algorithm, and the third one is based on a statistical analysis of the breast. The performance of the algorithms is exhaustively evaluated using a database of full-field digital mammograms containing 150 CC and 150 MLO images and ROC analysis (ground-truth provided by an expert). Results demonstrate that the use of region information is useful to obtain homogeneous region segmentation, although clustering algorithms obtained better sensitivity.

## 1 Introduction

Recent studies have shown that the performance of mammographic Computer-Aided Detection systems decreases as the density of the breast increases, either decreasing the sensitivity [1,2] or increasing the specificity [3]. Thus, the segmentation of breast density is beneficial not only for estimating the quantity of breast dense tissue, but also for establishing specific abnormalities detection procedures according to the type of tissue (i.e. fatty or dense).

The different algorithms proposed so far for breast density segmentation can be classified into two main categories: clustering or region based. In the former category, each pixel is classified using only its characteristics, i.e. without taking the information of neighboring pixels into account. Examples of algorithms belonging to this category are the works of Boyd et al. [4] and Sivaramakrishna et al. [5] who used a gray-level thresholding technique to segment the breast into dense and fatty regions, or the proposals of Aylward et al. [6] and Ferrari et al [7], who used Gaussian mixtures for segmenting the breast in five and four regions respectively. On the other hand, region based algorithms relies on classifying the pixels of the mammogram taking neighborhood information into account, i.e. considering the class of its neighbors. Examples of this approach are the work of Saha et al. [8] who used a fuzzy affinity function, or the works of Zwiggelaar and

Denton [9] and Petroudi et al. [10] where textural features were used to take the spatial distribution of the pixel and its neighborhood into account.

In this paper we provide a quantitative comparison among two clustering based algorithms and one region based. The evaluation of our experiments is done using the well-known ROC analysis. Note that one of the goals of our work is to provide a novel benchmark for automatic mammographic density quantification from which newly developed techniques could be directly compared. Remember that nowadays this is usually done by manual inspection of the segmented images. The rest of this paper is organized as follows. The following Section describe the analyzed algorithms. Afterwards, experimental results are presented and the application of our work to breast density classification is described. The paper ends with the conclusions.

## 2 Breast Density Segmentation

### 2.1 Thresholding Based on Excess Entropy (*EE*)

This approach is based on information theory. Given a sequence of random variables  $X_0, X_1, \dots$  taking values  $x$  in  $\mathcal{X}$  with probability distribution  $p(x)$ , its structure can be expressed by the excess entropy [11], which is defined by:

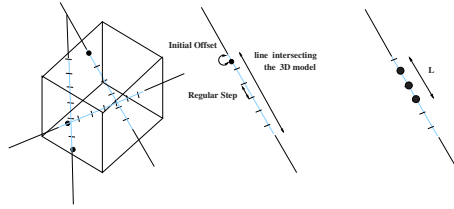
$$E = \sum_{L=1}^{\infty} (h^x(L) - h^x) = \lim_{L \rightarrow \infty} [H(X^L) - h^x L] \quad (1)$$

where  $h^x(L) = H(X_L | X_{L-1} \dots X_1)$  is the entropy of a symbol conditioned on a block of  $L - 1$  adjacent symbols,  $H(X^L)$  is the Shannon entropy of a sequence of  $L$  random variables, and  $h^x$  is the entropy rate of the sequence [11,12].

From the assumption that a medical image is structured in regions, the optimal thresholding should provide us with the maximum structure, i.e., the maximum excess entropy. In consequence, this proposal consists in quantizing the histogram by using the maximization of excess entropy. In the context of an image,  $\mathcal{X}$  represents the set of clusters or bins of the image histogram and  $x^L$  is given by a set of  $L$  neighbor intensity values. To compute the excess entropy, the concept of neighborhood has to be defined. This notion is unique and unambiguous in 1D, because only two neighbors per intensity value are possible. However, the extension to 2D or 3D introduces ambiguity, since the voxels can be selected in different ways [11].

On the other hand, excess entropy (1) has been defined from  $L$ -block entropies when  $L \rightarrow \infty$ . In practice,  $L$ -block entropies for high  $L$  can not be computed since the number of elements of the joint histogram is given by  $N^L$ , where  $N$  is the cardinality of the system (the number of clusters or bins of the segmented image histogram). Thus, a tradeoff between the accuracy of the measure and the number of clusters has to be assumed.

To overcome the neighborhood problem, the uniformly distributed random lines approach is used [13]. Points chosen on each line provide us with the intensities to calculate the  $L$ -block entropies, required to compute the excess entropy



**Fig. 1.** Uniformly distributed random lines are cast from the walls of the bounding box. Neighbor intensity values are taken in  $L$ -blocks.

(see Figure 1). In this manner, the 3D-neighborhood problem is reduced to a 1D problem in which neighborhood is well defined.

## 2.2 Fuzzy C-Means Approach (*FCM*)

The Fuzzy C-Means algorithm [14] is an extension of the well known k-Means algorithm. The main difference is that Fuzzy C-Means allows each pattern of the image to be associated with every cluster using a fuzzy membership function. The function criterion minimized by the algorithm is defined by:

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

where  $\Xi$  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 centered at  $c_t = (\sum_{n=1}^N u_{nt} p_n) / (\sum_{n=1}^N u_{nt})$ , where  $N$  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.

Once the algorithm has divided the image into clusters, the gray-level mean of each cluster is used to establish a ranking of cluster densities: a higher mean corresponds to a higher probability of being a cluster of dense tissue.

## 2.3 Fisherfaces Based Method (*FF*)

The third approach is based on a statistical analysis of the breast. We consider this approach as region based, because an analysis of the neighborhood of each pixel is performed to classify it as a fatty or dense pixel.

Patches of  $50 \times 50$  pixels of a set of mammograms are extracted and used as the ground-truth in order to segment the rest of mammograms. Some of the patches represent dense tissue while the other ones represent non-dense tissue. Using the fisherfaces approach [15], a set of models from each tissue type are created. Subsequently, a patch is extracted for each pixel of the mammogram and, using the created model, it is classified as belonging to fatty or dense.

The fisherfaces approach is a linear discriminant based algorithm. Firstly, the mean patches of dense and fatty classes are found, as well as the overall mean

patch. Subsequently, each patch is subtracted from the mean of its corresponding class, and this value is used to compute the between-class scatter (the dispersion among each class means) and the within-class scatter (the overall sum of the dispersion inside each class). The final aim is to find the subspace where both the between-class scatter is maximized and the within-class minimized. Subsequently, each patch of the training is projected in this subspace and a model is created. When classifying a new patch, it is also projected into that subspace and classified according to the most similar class using the k-Nearest Neighbor algorithm.

In contrast to both described approaches, in this work we do not perform a final classification of the pixels between both classes, but we give a membership ( $\Phi$ ) for each pixel. This is found as:

$$\Phi = \frac{1}{2} \frac{\frac{1}{d_D+1} - \frac{1}{d_F+1}}{\frac{1}{d_D+1} + \frac{1}{d_F+1}} + \frac{1}{2} \quad (3)$$

where  $d_D$  and  $d_F$  are the Euclidean distances between the patch centered at the pixel and the closest dense and fatty instances, respectively. Note that the first and final fractions are constant values that normalize this value between 0 (when it is sure the patch is not dense tissue) and 1 (it is sure that the patch is dense tissue).

### 3 Experimental Results

To evaluate the proposed methods we used a database of digital mammograms, containing 125 CC and 125 MLO views of the same breast of 75 women (there are 50 complete cases including left and right mammograms, and 25 where only one view is provided). The mammograms are acquired using a full-field digital mammograph (Siemens Mammomat Novation), and stored in DICOM format in a PACS server. The mammograms are 70 micron pixel edge.

Table 1 shows the used mammograms classified by the experts according to the BI-RADS lexicon [16]. Note that most of the images are classified into BI-RADS I and BI-RADS III categories, where mammograms with low BI-RADS correspond to non-dense breasts while mammograms with high BI-RADS correspond to dense breasts. Besides, the experts have annotated the dense regions of the mammographic images, surrounding accurately the region using a digital pen.

The comparison of the methods is provided using a ROC analysis. This is done in this way because, as explained before, the output of the segmentation

**Table 1.** Distribution of the mammograms used according to BI-RADS density categories [16]. Usually, the density of the breast increases as the BI-RADS class increases.

	BI-RADS I	BI-RADS II	BI-RADS III	BI-RADS IV	Total
MLO	50	19	36	20	125
CC	46	21	38	20	125
Total	96	40	74	40	250

**Table 2.** ROC analysis of each algorithm detailed for both mammographic views. Az mean is showed for the global analysis while Az mean and standard deviation are showed for the local analysis.

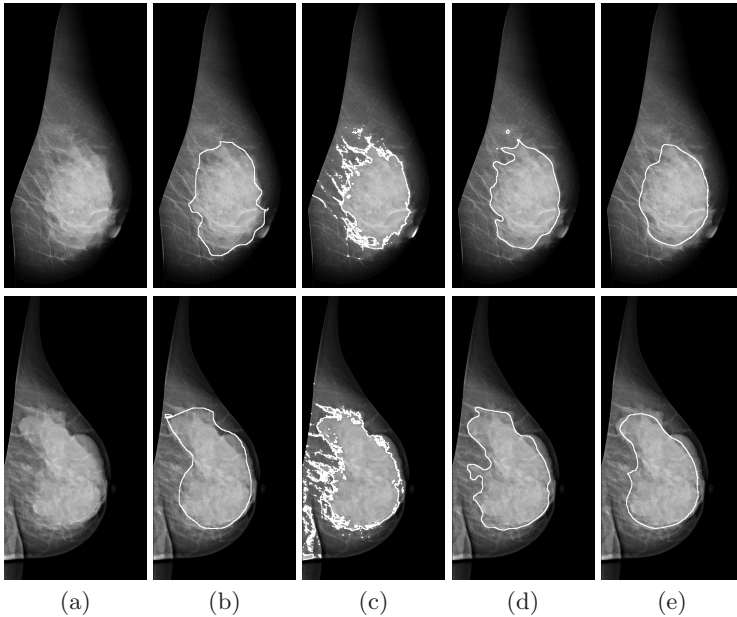
	Global		Local	
	MLO	CC	MLO	CC
<i>EE</i>	0.818	0.803	$0.8613 \pm 0.0939$	$0.8572 \pm 0.0927$
<i>FCM</i>	0.813	0.826	$0.9018 \pm 0.1067$	$0.9030 \pm 0.0935$
<i>FF</i>	0.852	0.824	$0.8798 \pm 0.1174$	$0.8753 \pm 0.1135$

algorithms is not a simple segmentation in two different classes, but a degree of membership for each class. Therefore, different results can be obtained when thresholding the images at different thresholds providing, hence, the ROC curve. The area under the curve ( $Az$ ) is used to compare the algorithms. The parameters of the algorithms have been empirically optimized, although the results do not depend strongly on this choice, surely because we are segmenting the breast in only two clusters.

Table 2 shows the  $Az$  values computed using a global ROC analysis. This consists in computing each point of the ROC curve using all correctly and incorrectly classified pixels of all the mammograms at the same threshold. This measure shows the overall performance of the algorithms. Note that for MLO images, the *FF* algorithm showed the best performance, while for CC views the *FCM* and the *FF* algorithms obtained similar results, being the former slightly better. Comparing both MLO and CC views, the performance of the algorithms was similar, except for algorithm *FF* which had an  $Az$  decrease of 0.028. This is probably due to the fact that we trained the algorithm using RoIs extracted from only MLO images.

Table 2 also shows the mean and the standard deviation of the  $Az$  value computed individually for each image. Note this value can only be used to show the performance of the algorithm per image. Observe that now the performance of the algorithms is substantially different. For instance, for MLO images, the best performance was obtained when using the *FCM* algorithm, while *FF* had intermediate values, and *EE* obtained the worst results. Note that, in contrast with the global  $Az$ , the same behavior was obtained for the CC views.

Finally, Figure 2 shows the segmentation results of two examples per algorithm. From left to right respectively, original mammogram, manual segmentation by the expert, *EE* segmentation, *FCM* segmentation, and *FF* segmentation. Note that the clustering based algorithms (*EE* and *FCM*), although obtaining disconnected regions, follow accurately the boundary between the fatty and dense classes in those regions where this boundary is clear. However, the *EE* algorithm loses accuracy in those regions where this boundary is not so clear. On the other hand, the region based approach clearly obtains a single region. However, the boundary of this region is less accurate than the obtained by the *FCM* algorithm. This is consistent with the quantitative results obtained in the ROC analysis: when comparing the performance individually, the clustering algorithms show more accuracy than the region based algorithm. In contrast,



**Fig. 2.** Segmentation examples of two different MLO mammograms. (a) original image, (b) expert segmentation, (c) *EE* segmentation, (d) *FCM* segmentation, and (e) *FF* segmentation.

when evaluating the overall mean, the number of disconnected regions decrease the performance of the clustering based algorithms.

## 4 Application to Breast Density Classification

We extend this work to classify the breast according to their internal density in terms of BI-RADS categories. Note that this classification is not necessary related with the density of the breast, because it is more referred to the type of internal tissue than to the density itself. Hence, a simple ratio of the dense tissue versus the total area of the breast could not be enough to correctly classify the breast.

To classify the breast we used a previous developed approach [17]. The first step is the aim of this paper, i.e. the segmentation of the breast according to the internal density. We used the Fuzzy C-Means segmentation explained here since it provided the best results. Following, we extract a set of features from each cluster in order to characterize the fatty and the dense regions. From this set, the first ones are related with the morphology of the clusters (mean gray-level, deviation, skewness, and kurtosis), while the rest are texture features extracted from co-occurrence matrices. Finally, a classifier is need to categorize each mammogram according to the BI-RADS categories. The classifier used here is a Bayesian combination of the C4.5 decision tree and the k-Nearest Neighbor algorithm.

**Table 3.** Confusion matrix for breast density estimation. (a) MLO views and (b) CC views.

		(a)			
		B-I	B-II	B-III	B-IV
Truth	B-I	46	4	0	0
	B-II	3	15	1	0
	B-III	0	2	28	6
	B-IV	0	1	5	14

		(b)			
		B-I	B-II	B-III	B-IV
Truth	B-I	37	6	3	0
	B-II	6	13	2	0
	B-III	0	3	30	5
	B-IV	0	0	6	14

To evaluate the results we used a leave-one-woman-out methodology, in which each sample is analyzed by a classifier which is trained using all other mammograms except for those of the same woman. The obtained confusion matrices for both MLO and CC views are shown in Table 3. Confusion matrices should be read as follows: rows indicate the object to be recognized (the true class) and columns indicate the label associated by the classifiers to this object.

Looking at Table 3, the shown algorithm clearly obtained better performance for MLO mammograms than for CC ones. The overall percentage of correct classification for the former is 82.4%, while for CC views is 75.2%. Looking at class level, note that mammograms belonging to BIRADS I are almost all classified correctly for MLO mammograms, while for CC views the performance is reduced.

## 5 Conclusions

A comparison of three different methods for segmenting the breast density has been presented in this paper: a multiple thresholding based on information theory, a Fuzzy C-Means based clustering, and a statistical region based approach supported by the fisherfaces algorithm. The performance of all the methods is analyzed using full-field digital mammograms and ROC analysis, with the ground-truth provided by a set of experts. The results show that the region based approach allows us to obtain single and homogeneous regions, although the clustering based obtain better sensitivity at individual level. Moreover, we have shown how to extent this work to qualitatively classify the breast according to the BI-RADS breast composition categories.

Current research is directed on merging the result of region and clustering based algorithms in order to obtain best performance not only at local but also at global level. Further research is directed to establish correspondences between the mammographic projections and the MRI images of the same breast by using the segmented density in both modalities.

## Acknowledgments

This work was supported by Ministerio de Educación y Ciencia of Spain under Grant TIN2007-60553 and by the UdG under Grant IdiBGi-UdG91060080.

## References

1. Ho, W.T., Lam, P.W.T.: Clinical performance of computer-assisted detection (CAD) system in detecting carcinoma in breasts of different densities. *Clinical Radiology* 58, 133–136 (2003)
2. Obenauer, S., Sohns, C., Werner, C., Grabbe, E.: Impact of breast density on computer-aided detection in full-field digital mammography. *J. Digit. Imaging* 19(3), 258–263 (2006)
3. Brem, R.F., Hoffmeister, J.W., Rapelyea, J.A., Zisman, G., Mohtashemi, K., Jindal, G., DiSimio, M.P., Rogers, S.K.: Impact of breast density on computer-aided detection for breast cancer. *Am. J. Roentgenol.* 184(2), 439–444 (2005)
4. Boyd, N.F., Byng, J.W., Jong, R.A., Fishell, E.K., Little, L.E., Miller, A.B., Lockwood, G.A., Tritchler, D.L., Yaffe, M.J.: Quantitative classification of mammographic densities and breast cancer risk: results from the Canadian national breast screening study. *J. Natl Cancer Inst.* 87, 670–675 (1995)
5. Sivaramakrishna, R., Obuchowski, N.A., Chilcote, W.A., Powell, K.A.: Automatic segmentation of mammographic density. *Acad. Radiol.* 8(3), 250–256 (2001)
6. Aylward, S.R., Hemminger, B.H., Pisano, E.D.: Mixture modelling for digital mammogram display and analysis. *Int. Work. Dig. Mammography*, 305–312 (1998)
7. Ferrari, R.J., Rangayyan, R.M., Borges, R.A., Frere, A.F.: Segmentation of the fibro-glandular disc in mammograms via Gaussian mixture modelling. *Med. Biol. Eng. Comput.* 42, 378–387 (2004)
8. Saha, P.K., Udupa, J.K., Conant, E.F., Chakraborty, P., Sullivan, D.: Breast tissue density quantification via digitized mammograms. *IEEE Trans. Med. Imag.* 20(8), 792–803 (2001)
9. Zwiggelaar, R., Denton, E.R.E.: Optimal segmentation of mammographic images. In: *Int. Work. Dig. Mammography*, pp. 751–757 (2004)
10. Petroudi, S., Brady, M.: Breast density segmentation using texture. In: Astley, S.M., Brady, M., Rose, C., Zwiggelaar, R. (eds.) *IWDM 2006*. LNCS, vol. 4046, pp. 609–615. Springer, Heidelberg (2006)
11. Feldman, D.P., Crutchfield, J.P.: Structural information in two-dimensional patterns: Entropy convergence and excess entropy (2002)
12. Cover, T.M., Thomas, J.A.: *Elements of Information Theory*. Wiley Series in Telecommunications (1991)
13. Bardera, A., Feixas, M., Boada, I., Sbert, M.: High-dimensional normalized mutual information for image registration using random lines. In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) *WBIR 2006*. LNCS, vol. 4057, pp. 264–271. Springer, Heidelberg (2006)
14. Bezdek, J.C.: *Pattern Recognition With Fuzzy Objective Function Algorithms*. Plenum Press, New York (1981)
15. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Machine Intell.* 19(7), 711–720 (1997)
16. American College of Radiology: *Illustrated Breast Imaging Reporting and Data System BIRADS*. 3rd edn. American College of Radiology (1998)
17. Oliver, A., Freixenet, J., Martí, R., Pont, J., Pérez, E., Denton, E., Zwiggelaar, R.: A novel breast tissue density classification methodology. *IEEE Trans. Inform. Technol. Biomed.* 12(1), 55–65 (2008)