

# Modeling and Classifying Breast Tissue Density in Mammograms

Anna Bosch, Xavier Muñoz, Arnau Oliver and Joan Martí  
Computer Vision and Robotics Group  
University of Girona  
Campus Montilivi, Girona  
{aboschr,xmunoz,aoliver,joanm}@eia.udg.es

## Abstract

*We present a new approach to model and classify breast parenchymal tissue. Given a mammogram, first, we will discover the distribution of the different tissue densities in an unsupervised manner, and second, we will use this tissue distribution to perform the classification. We achieve this using a classifier based on local descriptors and probabilistic Latent Semantic Analysis (pLSA), a generative model from the statistical text literature.*

*We studied the influence of different descriptors like texture and SIFT features at the classification stage showing that textons outperform SIFT in all cases. Moreover we demonstrate that pLSA automatically extracts meaningful latent aspects generating a compact tissue representation based on their densities, useful for discriminating on mammogram classification. We show the results of tissue classification over the MIAS and DDSM datasets. We compare our method with approaches that classified these same datasets showing a better performance of our proposal.*

## 1 Introduction

Breast cancer is considered a major health problem in western countries. A recent study from the National Cancer Institute (NCI) estimates that, in the United States, about 1 in 10 women will develop breast cancer during their lifetime [1]. Moreover, in such country, breast cancer remains the leading cause of death for women in their 40s [6].

Although manual screening of mammographies remains the key screening tool for the detection of breast abnormalities, it is widely accepted that automated Computer Aided Diagnosis (CAD) systems are starting to play an important role in modern medical practices. Most of the commercially available CAD systems and research efforts in breast mammography focuses only on the automatic detection of abnormalities. However, from a medical point of view, it is well-known that there is a strong positive correlation be-

tween high breast parenchymal density and high breast cancer risk [25]. For instance, the relative risk is estimated to be about 4 to 6 times higher for women whose mammograms have parenchymal densities over 60% of the breast area, as compared to women with less than 5% of parenchymal densities [26]. Thus, the development of automatic methods for classification of breast tissue is justified for an automatic risk assessment framework in prospective CAD systems. However, developments in this area have been limited.

Several techniques have been proposed for breast density classification [12, 26], but only a small number of previous works have suggested that texture representation of the breast might play a significant role. Miller and Astley [16] investigated texture-based discrimination between fatty and dense breast types applying granulometric techniques and Laws texture masks. Byng et al. [7] used measures based on fractal dimension. The work of Bovis and Singh [5] first estimated features from the construction of Spatial Gray Level Dependency matrices and second, it trains multiple Neural Nets (ANN) to classify the parenchymal density. Zwiggelaar et al. [27] segmented mammograms into density regions based on a set of co-occurrence matrices, and density classification used the size of the density regions as the feature space. Similarly, Oliver et al. [17, 19] proposed to extract texture features after the segmentation of the breast in two clusters which represent dense and fatty tissue.

Reviewing the literature of texture and materials classification we can also learn [13, 24]. These texture models first identify a dictionary of textons and then for each category of texture, a model is learnt to capture the signature distribution of these textons. Recently, Petroudi et al. [20] defined mammographic appearance making use of a technique proposed for texture classification. They defined parenchymal tissue as statistical distributions (histograms) over texton dictionary developed from a training set. Textons are defined as clustered filter responses which are rotational invariant.

Some works that classified object and scene categories can also provide us good ideas. For example, Sivic et

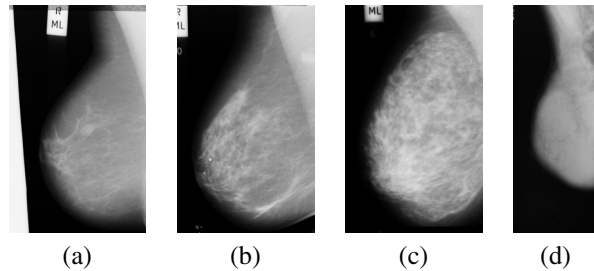
al. [22] used the probabilistic Latent Semantic Analysis (pLSA) for object recognition and in [4, 21] it is used for scene classification. pLSA is a generative model from the statistical text literature [11]. In text analysis this is used to discover topics in a document using the bag-of-words document representation. In these cases, there are *images* as *documents* and they discover *topics as object categories* (e.g. grass, houses, bikes, planes), so that an image containing instances of several objects is modelled as a mixture of topics. The models are applied to images by using a *visual* analogue of a *word*, formed by vector quantizing visual features (colour, texture, etc.) like region descriptors.

Motivated by the excellent results in object and scene classification, we propose to classify the mammograms by using pLSA. To carry out the adaption of this method to the medical image domain, we established the following analogies: in tissue classification, the *images* will be the *mammogram*, the *topics* will be the different *densities of the tissue* and we also will talk about *visual words* as the analogue of a *word*. pLSA is appropriate here because it provides the correct statistical model for clustering in the case of multiple tissues densities per image. We will have to study which are the best descriptors when classifying parenchymal densities as well as which is the best representation for this kind of images. Our main contribution in mammogram tissue classification is that this algorithm is able to learn relevant intermediate representation of tissue density automatically and without supervision. The previous approach of Petroudi et al. [20] which uses histogram models of textons does not provide a strong statistical model as our and can not differentiate the different densities in a mammogram automatically.

Nowadays, the American College of Radiology (ACR) Breast Imaging Reporting and Data System (BI-RADS) [2] is becoming a standard on the assessment of mammographic images. This standard provides four categories according to breast parenchymal density (see also Figure 1):

- BI-RADS I: the breast is almost entirely fatty.
- BI-RADS II: there is some fibroglandular tissue.
- BI-RADS III: the breast is heterogeneously dense.
- BI-RADS IV: the breast is extremely dense.

The rest of the paper is described below. Section 2 presents a detailed overview of the proposed system. Sections 2.1 and 2.2 describe a previous segmentation step and how we are going to represent the image using local descriptors. Section 2.3 describes the latent aspect and the pLSA process used to classify the mammograms according their parenchymal density. Section 3 describes the dataset and the followed methodology to test the approach. Section 4 show the results obtained and a brief comparison. The paper ends with conclusions and outlines possible future work.



**Figure 1. Four images belonging to one of each BI-RADS category extracted from MIAS dataset: from (a) BI-RADS I to (d) BI-RADS IV.**

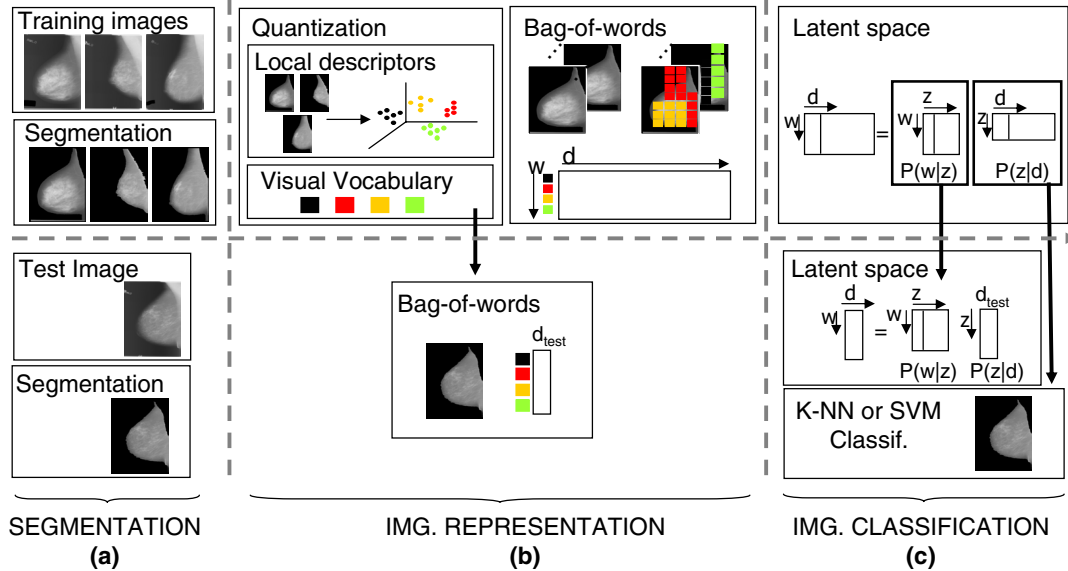
## 2 System Overview

In this Section we will explain the three processes involved on the mammographic tissue classification: (i) segmentation of the breast profile, (ii) breast tissue representation using a bag-of-words, and (iii) the use of probabilistic Latent Semantic Analysis (pLSA) to obtain the tissue classification according to the BI-RADS standard. Figure 2 shows the schema of the system.

### 2.1 Pre-processing steps

The initial step of our approach is the segmentation of the profile of the breast. Previous works on breast tissue classification and abnormalities detection noticed that the feature extraction process is affected if the region processed is not well focused. Thereby, it is important to segment the mammogram in order to extract the breast from other objects that could be present in a mammographic image (background, annotations, pectoral muscle in MLO images) and to achieve optimal breast parenchyma measurements. We used a *two-phase* based method:

- **Breast Segmentation.** The algorithm computes a global gray histogram for the image. The gray values are represented by a histogram with 8 bins. We compute an automatic threshold which is the minimum value over the 8-histogram. This one is used to threshold the image obtaining a collection of different regions. The largest region (the union of the breast and the pectoral muscle) is extracted using a Connected Component Labeling algorithm. As a result we delete the labels and the information which is not necessary and we obtain an image with the segmented breast.
- **Pectoral Muscle Extraction.** This operation is important in mediolateral oblique view (MLO), where the pectoral muscle, slightly brighter compared to the rest of the breast tissue, can appear in the mammogram. We used the approach of Ferrari et al. [8] who propose a polynomial modeling of the pectoral muscle.



**Figure 2. Schema of the learning and classification process. The first row shows the learning process while the second one shows the classification process. The figure is divided in three subparts corresponding to the three main process of our approach: (a) segmentation; (b) image representation; and (c) use of the latent space for learning and image classification.**

This segmentation results in a minor loss of skin-line pixels in the breast area, but those pixels are not relevant for tissue estimation.

## 2.2 Image Representation

We will represent the images as a co-occurrence table (bag-of-words) built from automatically extracted and quantised descriptors. Given the set of training images, local descriptors are computed around the pixels of the tissue (we do not take into account points close to the border) and a vocabulary of visual words (visual vocabulary) is obtained. In order to obtain the visual vocabulary, we used two different measures: the first one based on the appearance (textons) and the second one based on the edge orientation (Scale Invariant Feature Transform - SIFT).

- **Textons:** As in [24], a  $N \times N$  square neighbourhood is opened around each pixel. The pixels are row re-ordered to form a vector in an  $N^2$  dimensional feature space. The patch size tested are  $N = 3, 5, 7, 11, 15$  and  $21$ . The patches are spaced by  $M$  pixels on a regular grid over the area of the tissue. The patches do not overlap when  $M = N$ , and do overlap when  $M = 2$  (for  $N = 3, 5, 7$ ) and  $M = 7$  (for  $N = 11, 15$  and  $21$ ).
- **SIFT:** SIFT descriptors [14] are computed at: (i) points on a regular grid with spacing  $M$  pixels, here  $M = 5$  and  $10$ . At each grid point SIFT descriptors are computed over circular support patches with radii  $r = 8$  and  $16$  pixels; (ii) Affine co-variant regions are

computed for each grey scale image, constructed by elliptical shape adaptation about an interest point [15]. Consequently each point is represented by a 128-dim SIFT descriptors. Note, the descriptors are rotation invariant.

The number of descriptors is around 35000 and depends on how big is the area of the tissue and the parameters  $N$  and  $M$ . The visual vocabulary ( $V$ ) is obtained by vector quantising the descriptors computed from the training images using k-means.

Once we obtain the vocabulary, we represent the mammogram. Suppose we have a collection of images (mammograms)  $D = d_1, \dots, d_N$  with words from a visual vocabulary  $W = w_1, \dots, w_V$ . One may summarize the data in a  $V \times N$  co-occurrence table of counts  $N_{ij} = n(w_i, d_j)$ , where  $n(w_i, d_j)$  denotes how often the term  $w_i$  occurred in an image  $d_j$ .

## 2.3 Image Classification

Once we have built the bag-of-words we will use pLSA [11] to automatically find the topic (tissue) distribution for each mammogram. These distributions will be further used by the K-Nearest Neighbour (K-NN) or Support Vector Machines (SVM) to perform the mammogram classification as it is shown in Figure 2c. In pLSA, there is a latent variable (latent aspect or topic) model for co-occurrence data which associates an unobserved class variable  $z \in Z = z_1, \dots, z_Z$  with each observation (the occurrence of a visual word in an image). A joint probability model

$P(w, d) = P(d)P(w|d)$  over the co-occurrence table ( $V \times N$ ) is defined by the mixture:

$$P(w|d) = \sum_{z \in Z} P(w|z)P(z|d) \quad (1)$$

where  $P(w|z)$  are the topic specific distribution, and each image is modeled as a mixture of topics  $P(z|d)$ .

In training stage, the topic specific distributions  $P(w|z)$  are learnt from the set of training images. Each training image is then represented by a  $Z$ -vector  $P(z|d_{train})$ , where  $Z$  is the number of topics learnt. Determining both  $P(w|z)$  and  $P(z|d_{train})$  simply involves fitting the pLSA model to the entire set of training images. In particular it is not necessary to supply the identity of the images (i.e. which category they are in).

Classification of an unseen test image proceeds in two stages. First the document specific mixing coefficients  $P(z|d_{test})$  are computed, and following these are used to classify the test images. In more detail, document specific mixing coefficients  $P(z|d_{test})$  are computed using the fold-in heuristic described in [10]. The result is that the test image is represented by a  $Z$ -vector. The test image is then classified using a K-NN or SVM on the  $Z$ -vectors of the training images.

### 3 Datasets & Methodology

In order to test our method two public and widely known databases have been used: MIAS -Mammographic Image Analysis Society- database [23] and DDSM -Digital Database of Screening Mammographies- database [9]. Both are explained following:

- **MIAS.** This database is composed by the Medio-Lateral Oblique views of both breasts of 161 women (322 mammographies). The MIAS database provides annotations for each mammogram, and one of them is referred to the breast density. The images are labelled as: (i) *fatty* (106 images) if the breast is almost entirely fatty, (ii) *glandular* (104 images) if the breast contains some fibroglandular tissue, or (iii) *dense* (112 images) if the breast is extremely dense. Moreover, two experts mammographic readers, from the *Hospital Universitari Josep Trueta of Girona*, classified the MIAS database according to BI-RADS categories: BI-RADS I (128 images), BI-RADS II (80 images), BI-RADS III (70 images), and BI-RADS IV (44 images). Note that although a strong correlation exists between fatty class and BI-RADS I, glandular and dense tissue are distributed among the rest of BI-RADS categories.
- **DDSM.** We use a set which consists of 500 Medio-Lateral Oblique mammograms from the right breast: BI-RADS I (125 images), BI-RADS II (125 images), BI-RADS III (125 images), and BI-RADS IV (125 images). 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 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. However when working with the MIAS dataset, we leave the two images (left and right breast) from the same woman. This has to be done in order not to bias the results, because both breasts of the same woman have very similar tissue features. Therefore for the MIAS database we use 320 training images and 2 for testing 161 times, changing the test and train images every time. For the DDSM database we use 499 training images and 1 for testing 500 times.

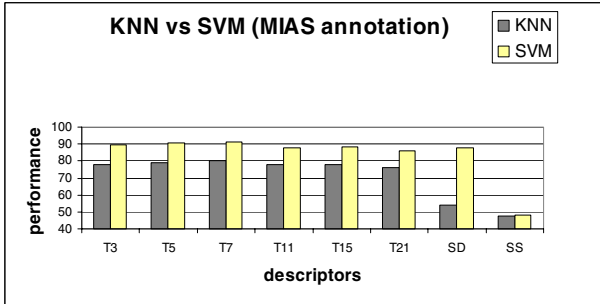
The classification task is to assign each test image to one category. In more detail, when using the K-NN, it selects the K nearest neighbours of the new image within the training database. Then, it assigns to the new mammogram the label of the category which is most represented within the K nearest neighbours. An Euclidean distance function is used. When using the SVM a gaussian kernel is used, and the multi-class classification is done using the one-versus-all rule: a classifier is learned to separate each class from the rest, and a test image is assigned the label of the classifier with the highest response. Overall performance rates are measured by the average value of the diagonal entries of the confusion table.

## 4 Experimental Results

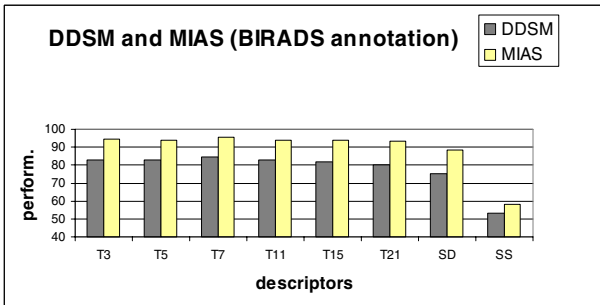
We divided this Section in three Subsections. The first one shows the results obtained when classifying the MIAS dataset using its own annotation: fatty, glandular and dense. The second one shows the results when BI-RADS annotation is used over both the MIAS and DDSM databases. Last subsection shows a comparison with other works. We investigated the classification performance when using K-NN and SVM classifiers over  $P(z|d)$  and when changing the value of different parameters: N (size of the patch when using textons), r (radii of the patch when using SIFT descriptors), M (space between patches), V (number of visual words of the vocabulary obtained using k-means), K (number of neighbours when using K-NN) and the two descriptors explained in Section 2.2.

### 4.1 MIAS annotation

The best results here have been obtained when  $V = 1600$ ,  $Z = 20$  and  $K = 6$ . Note that  $K$  has only sense if K-NN classifier is used. Results increase around 2% when using overlap between patches ( $M < N$ ). Figure 3 shows the results when classifying using the MIAS annotation and the two tested classifiers K-NN and SVM. Results using different descriptors (textons, dense and sparse SIFT) are



**Figure 3. Performance according to MIAS annotation when changing the values of parameters  $N$  and  $r$  and fixing  $V = 1600$ ,  $Z = 20$  and  $M = 2$  for K-NN and SVM, and  $K = 6$  for K-NN.  $T3 =$  Textons with  $N = 3$  and so on;  $SD =$  SIFT Dense;  $SS =$  SIFT Sparse.**



**Figure 4. Performance according to BI-RADS annotation over MIAS and DDSM datasets when changing the values of parameters  $N$ ,  $r$ ,  $M$  and fixing  $V = 1600$ ,  $Z = 20$ . SVM are used.  $T3 =$  Textons when  $N = 3$  and so on;  $SD =$  SIFT Dense;  $SS =$  SIFT Sparse.**

shown. The best rate classification is obtained when texton vocabulary is used with  $N = 7$  and  $M = 2$ . When using the K-NN the performance is 80.00% and increases up to 91.39% when using SVM. SVM always outperforms the K-NN classifier. The percentages drastically decreases to 54.2% (for K-NN) and to 87.98% (for SVM) when using the vocabulary obtained from the dense SIFT descriptors with  $r = 8$  and  $M = 5$ . This could be due to the nature of this kind of features: they are local histograms of edge directions computed over different parts of the local patch. In all kind of tissues provided from the mammograms there are a lot of edges and changes in the gradient orientation, so in this case, SIFT features are not a good discriminant to classify the tissue density. Better performances have been obtained with dense descriptors and high degree of overlap.

## 4.2 BI-RADS annotation

Best results are obtained when  $V = 1600$ ,  $Z = 20$  and  $K = 7$ . Figure 4 shows the results when classifying

**Table 1. Confusion table when using BI-RADS annotation. Texton vocabulary and  $N = 7$ ,  $M = 2$ ,  $V = 1600$ ,  $Z = 20$  are used.**

| -     | B-I    | B-II   | B-III  | B-IV   |
|-------|--------|--------|--------|--------|
| B-I   | 96.06% | 3.93%  | 0%     | 0%     |
| B-II  | 5.12%  | 93.58% | 1.28%  | 0%     |
| B-III | 0%     | 2.85%  | 94.28% | 2.85%  |
| B-IV  | 0%     | 0%     | 2.27%  | 97.72% |

the MIAS and DDSM datasets using BI-RADS annotation, SVM and different descriptors. This annotation is the one that specialists use when classifying the tissue density. For MIAS dataset, best result (95.42%) is obtained when using textons with  $N = 7$  and with overlap ( $M = 2$ ). More accurate results are obtained when using dense SIFT descriptors (88.19%) than when using the sparse ones (58.34%). Best results with DDSM dataset is 84.75% also with  $N = 7$  and  $M = 2$ . Results when using K-NN are around 18% worse.

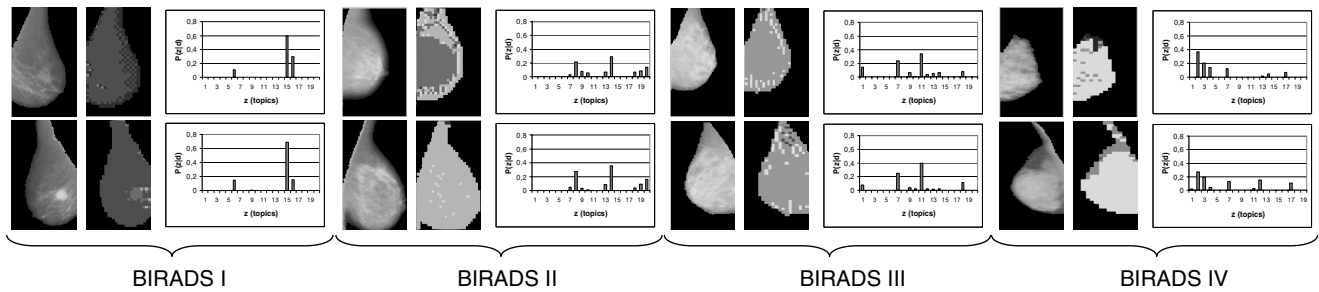
As can be seen from the confusion matrix of Table 1, the best classified tissue belongs to BI-RADS IV and the most difficult to classify and the ones which present most confusion are BI-RADS II and III. However, following previous works on breast tissue classification according to BI-RADS categories [5, 17, 20], we can reduce this four-class classification problem to the following two-class problem: (BI-RADS I and II) vs (BI-RADS III and IV). In other words, breasts with low density against breast with high density. With this supposition, a classification accuracy of 99.51% and 98.24% respectively is achieved.

Figure 5 shows examples of the spatial distribution of a number of topics (tissue densities) and their histogram of topic distributions ( $P(z|d)$ ). Patches are painted according to the maximum posterior  $P(z|w, d)$ :

$$P(z|w, d) = \frac{P(w|z)P(z|d)}{\sum_{z_i \in Z} P(w|z_i)P(z_i|d)} \quad (2)$$

For each visual word in the image we choose the topic with maximum posterior  $P(z|w, d)$  and paint the patch with its associated colour, so each colour represents a different topic (the topic colour is chosen randomly).

The images of Figure 5 are the segmentation of the parenchymal densities in mammograms. They illustrate that topics are representing consistent density tissues across images, there is a similar topic distribution (similar colour) for images from the same BI-RADS category. See for example that images belonging to BI-RADS I are very dark, while images from BI-RADS IV are lighter, showing that there is a different tissue density. If we observe the histograms we can see that those from the images of the same BI-RADS category, have a similar behaviour and topic distribution is consistent across the four BI-RADS categories.



**Figure 5. Segmentation results of two mammograms of each class: BI-RADS I, BI-RADS II, BI-RADS III and BI-RADS IV with the histogram of their topic distribution ( $P(z|d)$ ). The parameters used are: visual textons to compute the vocabulary,  $N = 7$ ,  $M = 2$ ,  $V = 1600$  and  $Z = 20$ .**

### 4.3 A Brief Comparison

We can compare the obtained results when using the own annotations of the MIAS dataset with those obtained by Blot and Zwiggelaar [3] and Oliver et al. [17]. The first one used a subset of the MIAS database (about 100 images per class) and obtained a 50% of correct classified mammograms. The second one increased this result to 73.00% and used a subset of 60 images per category. Our proposal outperforms both methods obtaining an score of 91.39% of correct classification.

When classifying the MIAS dataset using BI-RADS annotation we can compare the result with [19]. They obtained a 50% of correct classification when classifying the four categories while our performance is of 95.42%. Moreover we improve their results when classifying with only two categories (low and high density). In [19] they obtained 80.00% and we obtain 98.88%.

Bovis and Singh [5] and Oliver et al. [18] worked with the DDSM dataset obtaining a 71% and 50% of correct classified images, while working with DDSM we obtain 84.75%. Note that in [18], they only used a subset of 300 images whereas in [5] and our approach used a subset of 500. Other authors classified the tissue density using other datasets. For example Petroudi et al. [20] obtained a 76% of correct classified tissues and also works with BI-RADS annotation. However we can not compare this last result directly because their approach was developed by using a different database. Table 2 summarises these results.

### 4.4 Summary & Discussion

We have demonstrated the performance of our approach to classify tissue in mammograms. We investigated performances when working with K-NN and SVM and showed that SVM always outperform the K-NN classifier. We also investigated two kinds of descriptors: textons and SIFT features and our results showed that textons work better over this kind of images. Even though SIFT features have been stated as very useful for object and scene classification, they present a worst performance in our work. This is because

**Table 2. Comparison summary of the proposed method with other works that classify parenchymal density. Note that the approach of Petroudi et al. [20] work with a different dataset and we could not give a direct comparison. MIAS annotation is with 3 classes (fatty, glandular and dense) while BI-RADS annotation is 4 classes (from I to IV).**

| #Ref | Database | Annot.  | Author (%) | Our (%) |
|------|----------|---------|------------|---------|
| [3]  | MIAS     | MIAS    | 50%        | 91.39%  |
| [17] | MIAS     | MIAS    | 73%        | 91.39%  |
| [19] | MIAS     | BI-RADS | 50%        | 95.42%  |
| [18] | DDSM     | BI-RADS | 47%        | 84.75%  |
| [5]  | DDSM     | BI-RADS | 71%        | 84.75%  |
| [20] | OXFORD   | BI-RADS | 76%        | —       |

SIFT features work with histograms of edge directions and all the tissues in mammograms have a lot of lines. Thus, we can not disambiguate tissue density with this feature (edges). We also have demonstrated that the classification process works better with a high degree of overlap between patches.

Best results are obtained with SVM classifier when working with textons vocabulary and  $V = 1600$ ,  $Z = 20$ ,  $N = 7$ ,  $M = 2$ . Specifically, when classifying with MIAS annotation (3 categories) we obtained a 91.39% of correct classified images. When classifying with the same database with BI-RADS annotation (4 categories) the score obtained is 95.39% and for DDSM dataset the accuracy is 84.75%. We also compared our proposal with several previous approaches that worked with the same databases, and our results outperformed all of them. The main drawback of these techniques is they rely on an initial segmentation of the breast. We think this may be a reason of the superiority of our results. As it is well known, the segmentation is always a very hard task, and specially on medical image. Hence, a wrong segmentation can imply errors on the characterisation and later classification.

## 5 Conclusions

We have demonstrated the successful application of pLSA and SVM techniques to medical image domain when classifying breast tissue in mammograms. We have represented the images according to their tissue densities and we have shown that the distribution for the same category are similar. Besides, we have studied the influence of various descriptor parameters and have shown that using texture descriptors with overlap works better than SIFT features when working with mammograms.

As further work, we want to study the influence of abnormalities over our results. Masses, microcalcifications, and spicular lesions present a very different texture from the remaining tissue, which can affect negatively our results. Hence, we want to work on the previous detection of these abnormalities to avoid its inclusion on the model. Moreover, we are working together with hospital *Dr Josep Trueta* from *Girona* to increase the mammogram dataset.

## Acknowledgements

Thanks to Dr. Josep Pont and Dra. Elsa Pérez for labeling the MIAS dataset according to the standard BIRADS, and to Andrew Zisserman for his discussions. This work was partially founded by research grant BR03/01 from the University of Girona and by the MEC grant nb. TIN2005-08792-C03-01.

## References

- [1] American Cancer Society. Breast cancer: facts and figures. 2003-04. ACS, 2003.
- [2] American College of Radiology. *Illustrated Breast Imaging Reporting and Data System BIRADS*. American College of Radiology, 3rd edition, 1998.
- [3] L. Blot and R. Zwiggelaar. Background texture extraction for the classification of mammographic parenchymal patterns. In *MIUA*, pages 145–148, 2001.
- [4] A. Bosch, A. Zisserman, and X. Muñoz. Scene classification via plsa. In *ECCV*, May 2006.
- [5] K. Bovis and S. Singh. Classification of mammographic breast density using a combined classifier paradigm. In *IWDM*, pages 177–180, 2002.
- [6] S. Buseman, J. Mouchawar, N. Calonge, and T. Byers. Mammography screening matters for young women with breast carcinoma. *Cancer*, 97(2):352–358, 2003.
- [7] J. Byng, N. Boyd, E. Fishell, R. Jong, and M. Yaffe. Automated analysis of mammographic densities. *Physics in Medicine and Biology*, 41:909–923, 1996.
- [8] R. Ferrari and R. Rangayyan. Automatic identification of the pectoral muscle in mammograms. *IEEE T-MI*, 23:232–245, 2004.
- [9] M. Heath, K. Bowyer, D. Kopans, R. Moore, and P. J. Kegelmeyer. The digital database for screening mammography. In *International Conference in Medicine and Biology Society*, pages 211–221, June 2000.
- [10] T. Hofmann. Probabilistic latent semantic indexing. In *SIGIR Conference on Research and Development in Information Retrieval*, 1998.
- [11] T. Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 41(2):177–196, 2001.
- [12] N. Karssemeijer. Automated classification of mammographic parenchymal pattern. *Physics in Medicine and Biology*, 28(6):365–378, 1998.
- [13] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textures. *IJCV*, 43(1):29–44, June 2001.
- [14] D. Lowe. Distinctive image features from scale invariant keypoints. *IJCV*, 60(2):91–110, 2004.
- [15] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *IJCV*, 60(1):63–86, 2004.
- [16] P. Miller and S. Astley. Classification of breast tissue by texture and analysis. *IVC*, 10:227–282, 1992.
- [17] A. Oliver, J. Freixenet, A. Bosch, D. Raba, and R. Zwiggelaar. Automatic classification of breast tissue. In *IbPRIA*, pages 431–438, Estoril, Portugal, June 2005.
- [18] A. Oliver, J. Freixenet, and R. Zwiggelaar. Automatic classification of breast density. In *ICIP*, volume 2, pages 1258–1261, 2005.
- [19] A. Oliver, J. Martí, J. Freixenet, J. Pont, and R. Zwiggelaar. Automatic classification of breast density according birads categories using a clustering approach. In *Computed Aided Radiology and Surgery*, Berlin, Germany, June 2005.
- [20] S. Petroudi, T. Kadir, and M. Brady. Automatic classification of mammographic parenchymal patterns: A statistical approach. In *IEEE EMBS*, Cancun, Mexico, volume 2, pages 798–802, 2003.
- [21] P. Quelhas, F. Monay, J. Odobez, D. Gatica-Perez, T. Tuytelaars, and L. Van Gool. Modeling scenes with local descriptors and latent aspects. In *ICCV*, Beijing, China, October 2005.
- [22] J. Sivic, B. C. Russell, A. Efros, A. Zisserman, and W. T. Freeman. Discovering objects and their locations in images. In: *ICCV*, Beijing, China, 2005.
- [23] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggs, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage. The Mammographic Image Analysis Society digital mammogram database. In *IWDM*, pages 211–221, 1994.
- [24] M. Varma and A. Zisserman. Texture classification: Are filter banks necessary? In *CVPR*, volume 2, pages 691–698, Madison, Wisconsin, June 2003.
- [25] J. Wolfe. Risk for breast cancer development determined by mammographic parenchymal pattern. *Cancer*, 37:2486–2492, 1976.
- [26] C. Zhou, H. Chan, N. Petrick, M. Helvie, M. Goodsitt, B. Sahiner, and L. Hadjiiski. Computerized image analysis: Estimation of breast density on mammograms. *Medical Physics*, 28(6):1056–1069, 2001.
- [27] R. Zwiggelaar and E. Denton. Optimal segmentation of mammographic images. In *IWDM*, June 2004.