Automatic microcalcification and cluster detection for digital and digitised mammograms

Arnau Oliver [a,*], Albert Torrent [a], Xavier Lladó [a], Meritxell Tortajada [a], Lidia Tortajada [b], Melcior Sentís [b], Jordi Freixenet [a], Reyer Zwiggelaar [c]

[a] Department of Computer Architecture and Technology, University of Girona, 17071 Girona, Spain
[b] UDIAT-Centre Diagnòstic, Corporació Parc Taulí, 08208 Sabadell, Spain
[c] Department of Computer Science, University of Wales, Aberystwyth SY23 3DB, UK

ARTICLE INFO

ABSTRACT

In this paper we present a knowledge-based approach for the automatic detection of microcalcifications and clusters in mammographic images. Our proposal is based on using local features extracted from a bank of filters to obtain a local description of the microcalcifications morphology. The developed approach performs an initial training step in order to automatically learn and select the most salient features, which are subsequently used in a boosted classifier to perform the detection of individual microcalcifications. Subsequently, the microcalcification detection method is extended in order to detect clusters. The validity of our approach is extensively demonstrated using two digitised databases and one full-field digital database. The experimental evaluation is performed in terms of ROC analysis for the microcalcification detection and FROC analysis for the cluster detection, resulting in better than 80% sensitivity at 1 false positive cluster per image.

1. Introduction

Breast calcifications are deposits of calcium inside breast tissue. They appear widespread in the breast and most women will have a few on their mammograms at some point in time, more commonly after menopause [1]. Most calcifications will not be detected during clinical exams or breast self-examination. However, mammography allows to find them long prior to they could move forward into an actual lump. This fact explains why developed countries are adopting the so-called screening programs, which mainly consist in promoting regular women examinations using mammography, usually starting at 40 years and performing them every 2 years.

It is usual to distinguish between two major types of calcifications according to the size: macrocalcifications and microcalcifications. While macrocalcifications are nearly always non-cancerous and need neither additional follow-up nor biopsy, microcalcifications should be diagnosed in more detail. Although about 80% of microcalcifications are typically non-cancerous, when the microcalcifications are new, clustered firmly together, and distributed in specific configurations, they are suspicious signs of breast cancer, most frequently a non-invasive ductal carcinoma in situ. Due

to its high spatial resolution, mammography allows to detect microcalcifications at an early stage, a fundamental step for improving prognosis [2,3]. In a mammogram, microcalcifications appear as small bright spots within an inhomogeneous background. Fig. 1 shows two mammograms from the MIAS database [4] containing a cluster of microcalcifications.

The automatic detection of microcalcifications and clusters is a well-known topic in mammography, as can be seen in the different surveys covering this topic [5,6]. More recent approaches are by Papadopoulos et al. [7], Pal et al. [8], Rizzi et al. [9] and Yu et al. [10]. Papadopoulos et al. [7] improve previous work [11], which was based on detecting microcalcifications using a neural network, by adding a pre-processing image enhancement step. In their work, different algorithms were tested, obtaining the best results when using the local range modification and the redundant discrete wavelet linear stretching and shrinkage enhancement algorithms. Pal et al. [8] also proposed to use neural networks for microcalcification detection. The first step of their approach consisted in using a multi-layered perceptron network for selecting 29 features that best account for the microcalcification detection from the 87 initially tested. These features are subsequently used to segment the mammograms using another perceptron network. A final step for false positive reduction was necessary for removing thin elongated regions. In this approach, clusters were detected by using a weighted density function which takes the position of the microcalcifications into account. Rizzi et al. [9] proposed a two-stage

* Corresponding author. Address: Ed. P-IV, Campus Montilivi, University of Girona, 17071 Girona, Spain. Tel.: +34 972418878; fax: +34 972 418259.
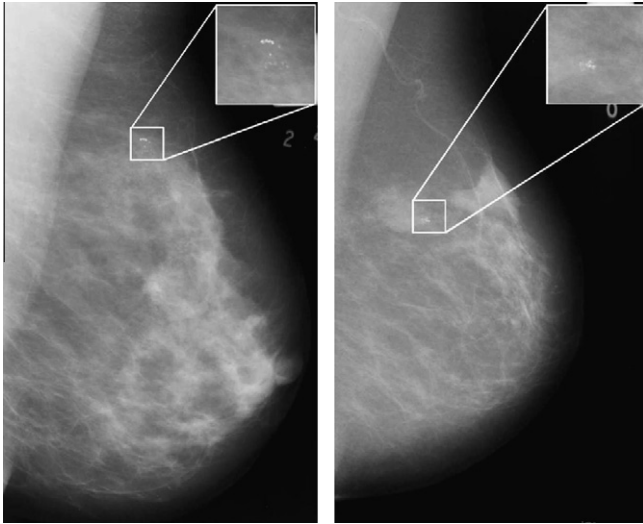E-mail address: aoliver@eia.udg.edu (A. Oliver).

Fig. 1. Two mammograms containing microcalcifications (extracted from the MIAS database). Both examples were selected for a good visualisation of the problem, although in general microcalcifications are more subtle and difficult to appreciate, even for experts radiologists.
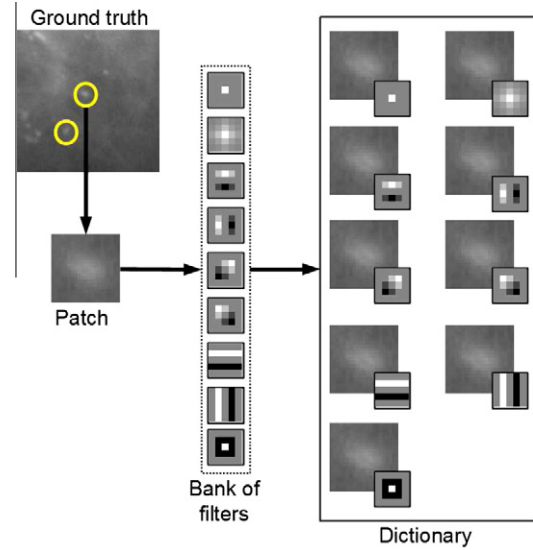

Fig. 3. Creation of the dictionary. From each manually marked microcalcification a patch is extracted and used together with the filter bank to create the words of the dictionary.

decomposition wavelet filtering for detecting microcalcifications. The first stage is used to reduce background noise preserving all suspect microcalcifications by thresholding mammograms according to image statistics (mean grey level pixel value and standard deviation), while the second one acts as a hard threshold technique, identifying the microcalcifications from the background. A cluster was considered if more than 3 microcalcifications were detected in a 1 cm^2 square area. Yu et al. [10] combined model-based and statistical textural features for clustered microcalcifications detection. Firstly, suspicious regions containing microcalcifications were detected using a wavelet filter and two thresholds. Secondly, textural features based on Markov random fields and fractal models together with statistical textural features were extracted from each suspicious region and were classified by a back propagated neural network.

In this paper we present a new approach for the detection of microcalcifications and clusters. The roots of this work can be

found in our previous work [12,13], which was only centred in the detection of individual microcalcifications. In this paper we increase the experimental evaluation of this part, and further, we extend the approach for the detection of clusters, which is more relevant from a clinical point of view. Briefly, the individual microcalcification detection is based on learning the variation in morphology of the microcalcifications using local image features. Afterwards, this set of features is used to train a pixel-based boosting classifier which at each round automatically selects the most salient microcalcification feature. Therefore, when a new mammogram is tested, only the salient features are computed and used to classify each pixel of the mammogram as being part of a microcalcification or actually being normal tissue. Afterwards, the microcalcification clusters are found by inspecting the local neighbourhood of each microcalcification. Note that with our boosting framework we are able to perform both microcalcification and cluster detection without requiring a further classification step as done in
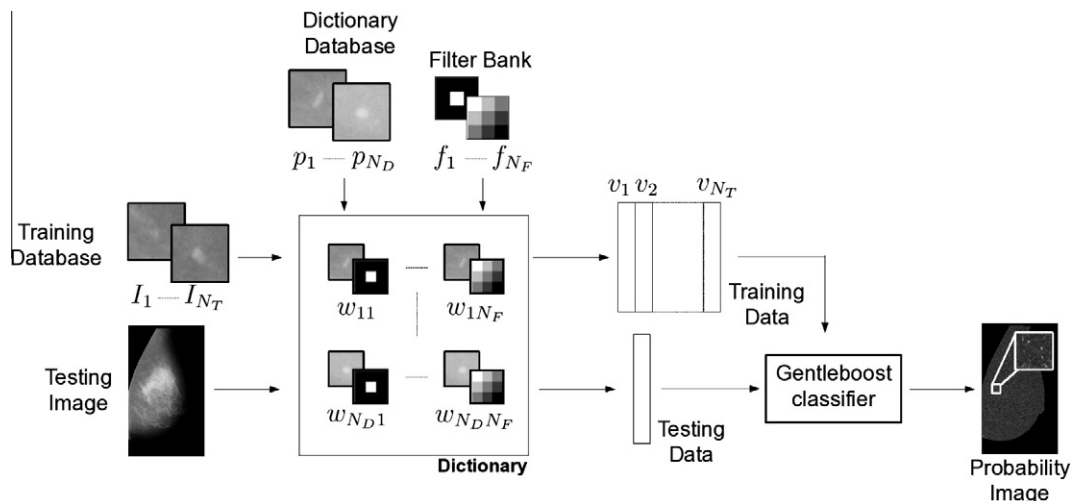

Fig. 2. Schematic representation of our approach. Firstly, a dictionary is created by convolving all the N_D patches of the dictionary database with all the N_F filters of the filter bank (obtaining an amount of $N_D \times N_F$ words). Afterwards, the training data is obtained by cross-correlating the N_T patches of the training data with all the dictionary words, and used as the learning of the classifier. When a new image is analysed, each pixel of the breast is used as the centre of a patch, which is subsequently analysed by the classifier. Therefore, the final result of the algorithm is a probability image where brighter pixels represent more confidence to be a microcalcification.

previous approaches [7–10]. Moreover, it is important to remark that we are not dealing with diagnosis in this paper, which is usually performed by means of knowledge-based systems [14–16].

It is well known that digital mammography allows to improve the detection of microcalcifications thanks to its superior sensitivity [17], and new approaches only dealing with full-field digital microcalcification detection are appearing. Unfortunately, this technology is not yet available in many countries and clinical centres due to its cost. Therefore, reliable automatic approaches able to detect microcalcifications clusters in digitised film plates are still necessary. In the experimental section of this paper, we validate our approach using both technologies. In particular, we used the whole set of 322 mammograms of the MIAS database [4] and a set of 280 mammograms extracted from a non-public full-field digital database. The results show the validity of our approach to deal with mammograms of both natures.

The remainder of this paper is structured as follows. The following section describes the proposed approach for detecting individual microcalcifications in a mammogram. In Section 3 we extend our approach to the detection of microcalcifications clusters. Section 4 presents the experimental set-up designed for testing our approach, while the results are presented in Section 5. The experimental evaluation is done in terms of ROC and FROC analysis. Finally, the paper ends with discussion and conclusions.

2. Microcalcification detection

The presented approach for microcalcification detection is based on the work of Murphy et al. [18] for object detection using local features and a boosting classifier. Their approach relies on detecting an object by learning their salient parts and the relative position of these parts to the object centre. The filtering of each of these patches with a bank of filters allows to create a dictionary of visual words, which represent the object morphology at a given position respect to the object centre. These words of the dictionary are subsequently used to extract a set of features from the training data, which will are used to learn the classifier. Afterwards, using the same dictionary, the features are extracted from the testing images and through the classifier, used for detecting the object. Therefore, its centre is found by combining all the relative positions of the analysed patches. Instead of following the same strategy, which is of general purpose, in this paper we propose to directly characterise the microcalcifications with one patch, since the centre and the boundaries of the microcalcification are close enough to be represented by a single patch. This represents a different challenging problem, since only one patch is used to characterise the object, instead of a set of patches.

The proposed approach, depicted in Fig. 2, is divided in three parts. Firstly, we create the word dictionary, which is obtained by convolving patches containing a microcalcification with a bank of filters. This dictionary allows to characterise examples of known microcalcifications and will be subsequently used to characterise unknown images. First, the training data is found by convolving positive samples (patches containing a microcalcification) and negative samples (patches of other tissues) with the words of the dictionary, and it is used as input to the Gentleboost classifier [19]. And second, new mammograms are classified pixel-by-pixel by this trained classifier. Hence, the detection problem is translated to a pixel-based classification approach. In the following subsections we describe in more detail the three parts of our approach.

2.1. Building the dictionary

The first task of the system consists in building the dictionary. This dictionary is similar to an atlas, since it contains samples (patches) of microcalcifications and is used for the system to learn their morphologies. Moreover, the dictionary contains also the convolution of the patches with a bank of filters. Fig. 3 shows an example of this process. A patch including in its centre a microcalcification is extracted from the mammogram, and together with the 9 filters is included in the dictionary. From top to bottom, the filters used in this work are: the delta function (which returns the original patch as a result), a Gaussian filter (which produces a smoothed version of the patch), the four Gaussian derivatives, the two Sobel filters, and the Laplacian one (which all of them return an image related with the gradient of the patch). In the figure, the mean grey value of the filters represents a 0, while brighter pixels are positive values and darker pixels are negative values. Therefore, the dictionary contains grey-level and gradient information of the microcalcifications and their neighbourhood.

Mathematically, each dictionary word w_{ij} can be understood as the ordered pair (p_i, f_j), where p_i represents the patch and f_j the filter. Assuming there are N_D patches in the dictionary database, there are $9 \times N_D$ words in the dictionary. In a recent study, Varma and Zisserman [20] show that for the general goal of characterising and classifying texture, like is done with textons, one could use the image patches themselves instead of convolving them using a bank of filters. However, in our experimental results we found that the use of filters improves the final detection results. This point will be discussed with more detail in Section 6.

2.2. Training step

Our approach is a supervised strategy, therefore requiring an initial training step before starting the detection process (testing step). The features used for detecting the microcalcifications are obtained using the words of the dictionary. Therefore, for the training step, we need a database of patches containing instances of both patches with microcalcifications and "normal" patches from the rest of tissues.

For each training image patch the feature extraction consists of two operations. Firstly, the patch is convolved with our filter bank, and secondly, the normalised cross-correlation with all the words is computed. Fig. 4 visually shows this operation, which can be mathematically summarised as:

$$x_{tk} = (I_t * f_j) \otimes w_{ij} = (I_t * f_j) \otimes (p_i * f_j) \tag{1}$$

where I_t is the training image patch convolved ($*$) with the filter f_j and cross-correlated ($\otimes$) with the word w_{ij} (kth word). The resulting value x_{tk} represents the similarity of the training patch and the kth
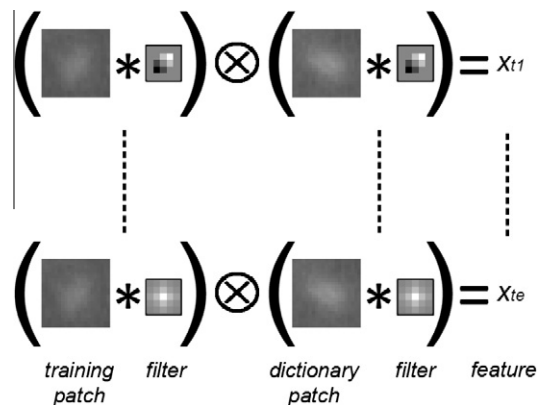

Fig. 4. Creation of the classifier training data. Each patch of the database is characterised using the words of the dictionary (actually, the length of the vector x_t is equal to $9 \times N_D$, i.e. the number of filters used multiplied by the number of patches of the dictionary database).

dictionary word. Therefore, for each training image patch, a vector of features x_t is constructed by cross-correlating all the dictionary words w_{ij} with the convolution of the patch itself with filter f_j. Notice here the necessity of keeping the filter as well as the patch in the dictionary word. Assuming that there are N_T patches in the training database, the training data x can be seen as a matrix of N_T column vectors of length $9 \times N_D$.

The image patches were manually selected from a set of (training) images. In particular, we select the centre of the microcalcifications as positive training examples and some random locations of the background containing examples of different tissues as negative training examples. Note that we could also use all the pixels not being microcalcifications instead of selecting only a subset, but this strategy will drastically increase the computational cost of the approach due to the large size of the mammograms. However, the total number of negative examples is still larger than the set of positive examples, since they should represent the different tissues that appear in a mammogram.

At this point, the positive and negative training examples have been characterised. The subsequent step is to use this data for training a classifier. For this task we used the Gentleboost algorithm [19,21]. Boosting algorithms are based on the idea that the sum of weak classifiers can produce a strong classifier [22,23]. In the Gentleboost algorithm, the weak classifiers (h_r) used at each round are simple regression stumps with one of the features:

$$h_r(x_t, k) = a\theta(x_{tk} > th) + b \qquad (2)$$

where a and b are the parameters of the regression stump, being $a = 2$, $b = -1$ the values obtained in the perfect case and $a = 0$ and $b = 0$ the values obtained in a random example, $\theta(\phi)$ refers to the Heaviside step function (0 if $\phi < 0$; 1 otherwise), th is the threshold that determines if pattern x_{tk} belongs to the object class, and x_{tk} refers to the kth dimension of column vector x_t (the feature actually selected). At each round, the values of the parameters a, b, and th are selected to minimise the error of the classifier:

$$e = \underset{k}{\arg\min} \sum_{t=1}^{N_T} \left(z_t^r.(y_t - h_r(x_t, k))^2 \right) \qquad (3)$$

where y_t is the data label (being 1 or -1) and z_t^r the training data weights of the t-patch at round r. The minimisation is done by exhaustively looking for all the values of th, since assuming a fixed th the values of a and b are automatically assigned. Hence, the value of b corresponds to the mean weighted label of the instances lower than the threshold, while a is the mean weighted label of the instances greater than the threshold (minus b to satisfy Eq. (2)):

$$b = \frac{\sum_{x_{tk} < th} y_t.z_t^r}{\sum_{x_{tk} < th} z_t^r} \qquad (4)$$

$$a = \frac{\sum_{x_{tk} \geq th} y_t.z_t^r}{\sum_{x_{tk} \geq th} z_t^r} - b \qquad (5)$$

At each round, only the feature k_r that obtains the minimum error e is selected and used in the testing step to classify the new data. Moreover, at each round of the boosting the weights z are updated, increasing in the following round the possibility of classifying correctly the previous incorrectly classified instances. In the Gentleboost algorithm the data weights are updated using:

$$z^{r+1} = z^r.e^{y.h_r(x, k_r)} \qquad (6)$$

Hence, when testing new data, the final (strong) classifier is computed using the weak classifier created at each round of the boosting. Therefore, the testing data is classified according to the sign of the sum of weak classifiers:

$$H(x) = \sum_{r=1}^{N_R} h_r(x, k_r) \qquad (7)$$

Therefore, pixels being part of a microcalcification should obtain positive values while the rest of pixels should obtain negative values. Furthermore, the absolute value of $H(x)$ shows the confidence of the classified data. Moreover, the classifier also returns the k_r features selected, i.e. the N_R filters and words that must be used at each round for extracting the features of the new testing mammograms, where N_R is the number of rounds used in the boosting (100 in this paper). Note that a filter and a word can be used more than once.

2.3. Testing step

Once the classifier is trained, the system is ready for the testing step. Therefore, the strong classifier $H(x)$ is applied to new mammograms in order to detect each individual microcalcification. Note that the classifier is pixel-based, i.e. it is applied one-by-one to all the image pixels. Therefore, the result of our approach after evaluating a mammogram is a probability image, where high values represents more confidence to be a microcalcification. Fig. 5 shows the result of applying the proposed approach to mammograms shown in Fig. 1.

Since the classifier is pixel-based, a pre-processing step is necessary in order to avoid the algorithm detecting microcalcifications in the background (in digitised images) and in the pectoral muscle. We used a previous developed algorithm to detect the skin-line border [24] and the approach of Kwok et al. [25] to remove the pectoral muscle. Recently, Paradkar and Pande [26] suggested to speed up the overall computational detection time by analysing only the image pixels of the breast region brighter than a certain threshold. Although this pre-processing allows to reduce the computational time, it introduces a new parameter that has to be properly tuned. Since the goal of our paper is to evaluate the boosting detection approach and its parameters, we have not included this extra step in our approach.
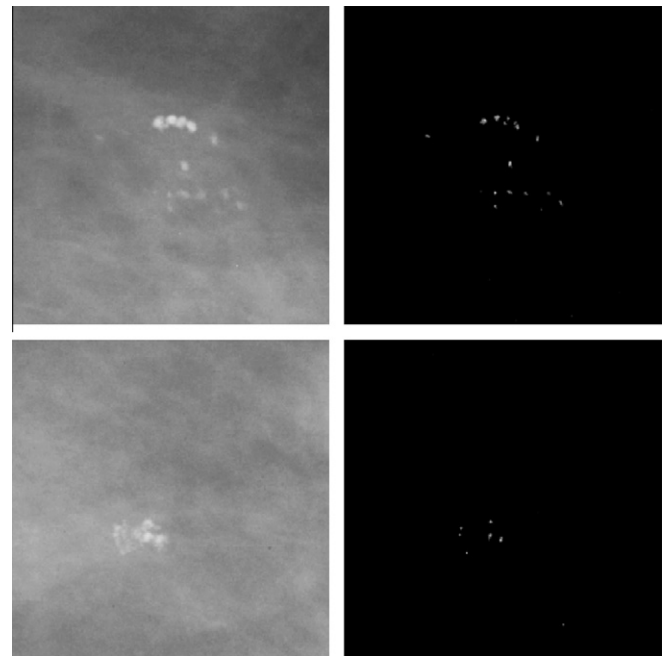

Fig. 5. Result of applying the proposed microcalcification detection approach to the mammograms shown in Fig. 1. Only the cropped regions are shown for easier visualisation.

3. Cluster detection

As stated in the introduction, most women will develop breast microcalcifications during their lifetime. If the microcalcifications are scattered in all the breast it is usually a sign of benign abnormality. However, when clustered together they may be a suspicious sign of breast cancer. Hence, the automatic detection of clusters is another important issue.

In order to deal with cluster detection, we use the probability image resulting of the microcalcification detection approach described in the previous section. When a set of microcalcifications is present in a region, the probability image should contain also high probabilities inside this region (see Figs. 1 and 5). Hence, the natural extension for cluster detection is to locally integrate this probability image. Notice that pixels in this region that are not microcalcification should have negative values and, therefore, decrease the output of such integration. In order to avoid this issue, we firstly threshold the negative values to zero:

$$I_Q(x,y) = \begin{cases} I_P(x,y), & \text{if } I_P(x,y) > 0 \\ 0, & \text{otherwise} \end{cases} \qquad (8)$$

where I_P is the probability image resulting from the microcalcification step. Therefore, the cluster probability is defined by:

$$I_C(x,y) = \int_\Omega I_Q(x',y')dx'dy' \qquad (9)$$

where x' and y' are the individual microcalcification probability included in the local neighbourhood Ω. Note that this extension for cluster detection is straightforward and only needs one additional parameter (the size of the neighbourhood). In the experimental section we will provide details of how to properly adjust this parameter. Fig. 6 shows the probability images obtained for cluster detection in the mammograms shown in Fig. 1.
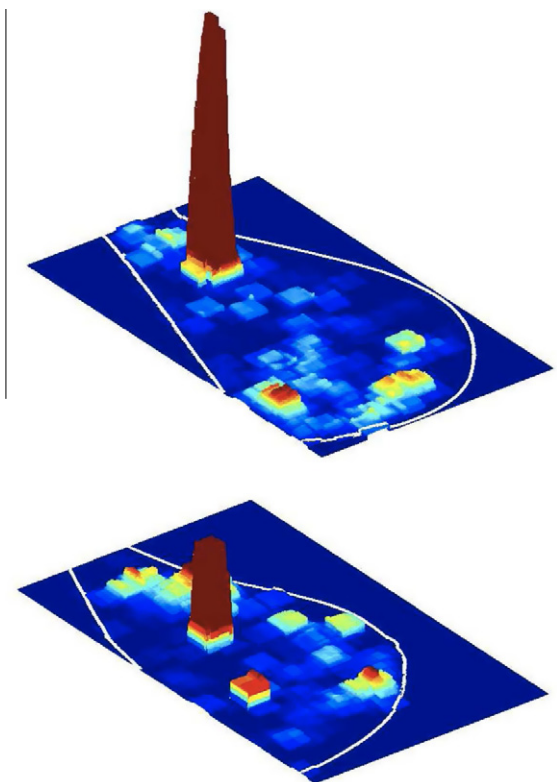

Fig. 6. Result of applying the proposed cluster detection approach to the mammograms shown in Fig. 1. Note that some false positive regions can appear depending on the final threshold.

The final step is to threshold the probability image to estimate if a mammogram contains microcalcifications or not. Note that if this threshold is high only few suspicious regions will be detected but with a great probability of being real clusters. In contrast, if the threshold is low, more suspicious regions will be detected but probably with some regions not being real clusters (i.e. false positive regions). This final threshold decision can be used according to the preference of the physicians.

4. Experimental set-up

The experimental results were performed using two different sets of mammograms. The first one was the full (digitised) MIAS database [4], which contained 207 normal mammograms, 25 mammograms with microcalcifications (with a total of 28 clusters), and 90 mammograms containing other types of abnormalities (masses, spiculations, architectural distortions, and asymmetries). The spatial resolution of the images was 50 μm × 50 μm and the optical density was linear in the range 0 − 3.2 and quantised to 8 bits. The second set of mammograms was a set of 280 full-field digital mammograms extracted from a non-public database, 90 of them containing microcalcifications and 190 mammograms without abnormalities. The mammograms were acquired using a Hologic Selenia mammograph, with resolution 70 micron-pixel, size 4096 × 3328, and 12-bit depth.

For the training step, our approach needs the exact location of some individual microcalcifications (positive examples). Therefore, an expert accurately marked among 5 and 15 microcalcifications in those mammograms containing microcalcifications in the MIAS and the digital databases. Hence, these two databases were used to train the system. The negative examples were obtained from the rest of the tissues of these mammograms and also from the normal ones, using around 20 marks in each mammogram. For testing the system, only an ellipse (or just a circle) circumscribing the clusters was necessary. Therefore, the public annotations of the MIAS database were used to test our approach. In addition, two experts annotated the corresponding ellipses for the testing images of the digital database (each radiologist annotated a different subset of images).

Since we are using the same database for training and testing the algorithm, we applied a 10-fold cross-validation methodology. Therefore, we divided each dataset in 10 different groups. One of the groups was used to create the dictionary, eight of them were merged for training the system, while the remaining one was used for testing it. This procedure was repeated until all groups were used for testing (in each fold the dictionary was also created using a different group). Hence, each mammogram appears in the test set only once.

To perform the quantitative evaluation of the microcalcification detection algorithm we used Receiver Operating Characteristic (ROC) analysis. In this analysis, a graphical curve represents the true positive rate (number of detected mammograms with microcalcifications divided by the total number of mammograms with microcalcifications) as a function of the false positives rate (number of normal mammograms incorrectly detected as containing microcalcifications divided by the total number of normal mammograms). Moreover, the percentage value under the curve (A_z) is an indication for the overall performance of the observer, and is typically used to analyse the performance of the algorithms. Note that the different points on the curve are obtained by thresholding equation (7) at different levels.

On the other hand, to evaluate the ability of the algorithm for the cluster detection we used Free-response Receiver Operating Characteristic (FROC) analysis. In FROC analysis the Lesion Localisation Fraction (LLF) is obtained as the number of correctly detected lesions relative to the total number of lesions and the

Non-lesion Localisation Fraction (NLF) as the number of incorrectly detected lesions relative to the total number of images. The FROC curve is the graphical summary of both measures [27]. Note that the definition of what is a detected cluster is needed. In this paper we assume that a cluster is detected if the centre of the automatically found cluster is inside the manually marked region (this is similar to the approach of Linguraru et al. [28]).

5. Results

5.1. Evaluation of the microcalcification detection

The first experimental evaluation is related to the ability of the algorithm to detect those mammograms containing microcalcifications. In order to empirically find the best number of dictionary words, we repeated the 10-fold cross-validation methodology above explained using 100, 250, 500, 750, and 1000 words for the MIAS database. As is graphically shown in Fig. 7, the best results were achieved when using 500 visual words for describing the different microcalcifications morphology. Note that similar results were obtained when increasing the number of words, while lower results were obtained using fewer words. However, the computational time of the whole process dramatically increased when increasing the number of patches and words used for building the dictionary, and the empirical values used here provided a good trade-off between performance and feature vector length.

On the other hand, when testing the digital database using again a 10-fold cross-validation methodology, we achieved an area under the ROC of $A_z = 0.918$, using again 500 words for describing the microcalcifications. This results shows that the algorithm can correctly detect almost all the mammograms containing microcalcifications without a large number of false positive mammograms. Table 1 summarises the result of the approach when testing the two different databases. We can observe that the results based are very similar, obtaining sensitivities higher than 90% at high specificities.

5.2. Evaluation of the cluster detection

To evaluate the cluster detection we used FROC curves, which are shown in Fig. 8. As in the previous experiment, similar results were obtained for the MIAS and the digital databases. In order to extract significant conclusions, in what follows we used the approach of Bornefalk [29,30] to compute the 95% confidence level
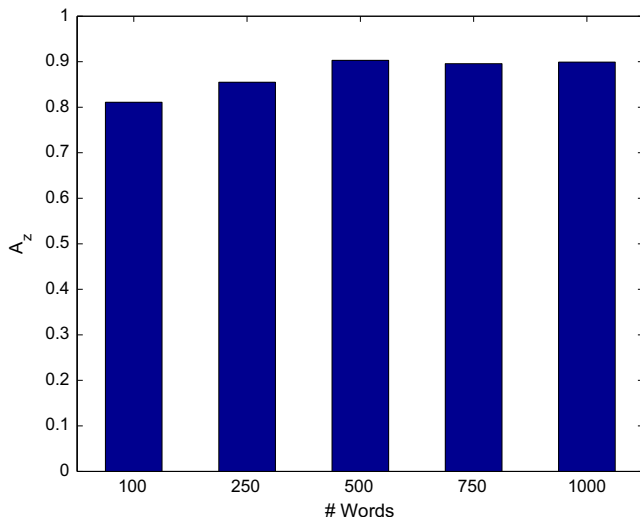

Fig. 7. 10-fold cross-validation results when using different number of words.

Table 1
Evaluation of the approach for detecting mammograms containing microcalcifications, detailing the database, the number of mammograms containing microcalcifications, the number of mammograms without microcalcifications, and the obtained area under the curve.

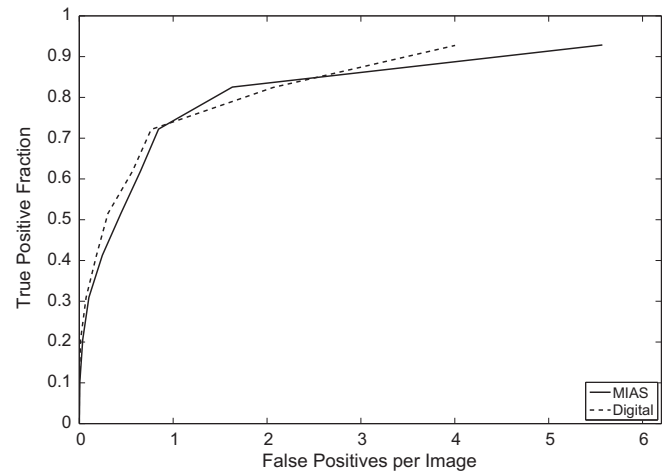| Database	# Micros	# Normals	Results (A_z)
MIAS	25	297	0.903
Digital	90	190	0.918

Fig. 8. Obtained FROC when testing the MIAS and the digital database. Note that we obtained better results when testing the latter.

of the number of false positives per image at a given sensitivity. Note that this approach allows us to obtain statistical meaning of the results avoiding executing multiple trials of our algorithm.

For the MIAS database, at a sensitivity of 80% we obtained a confidence interval of $(0.96, 1.73)$ false positives per image, while at 90% the false positive number per image ranges between $(3.23, 5.52)$. Looking at the results, we noticed that two of the 28 annotated clusters[1] were detected with very low probabilities. Inspecting them we noted that they were located in highly dense regions. These clusters were not detected with high probabilities due to the small number of similar cases in the database, and we assume that training with a larger database will be beneficial to avoid this issue. On the other hand, Fig. 9 shows the two mammograms where the algorithm obtained a false positive cluster with the highest probability. Note that in the first case, a cumulus of calcifications is present, while in the second case, the calcified vessel confused the algorithm.

In the cluster detection approach, the only parameter is the radius of the local neighbourhood (kernel). We used different kernel sizes in order to empirically find the best one. In particular, we used 10 different square kernel sizes: from 50 pixels to 500 pixels in steps of 50 pixels (other shape kernels can be used although the final result should be similar). The best results were obtained using the kernels of size 150×150 and 200×200 pixels (corresponding to 7.5×7.5 and 10×10 mm^2, respectively). Note that this is consistent with the annotations of the database, since the median of the diameter of the annotated clusters is 205 pixels.

Finally, when using the digital database we obtained the following confidence intervals: at 80% sensitivity, the false positives per image ranged between $(1.28, 3.02)$, while at 90%, they moved

---

[1] Note that this number (28) is different from the number of mammograms containing microcalcifications (25) since there are images with more than one annotated clusters. Moreover, 3 of the mammograms contained microcalcifications spread to the entire breast and were not considered in this evaluation.
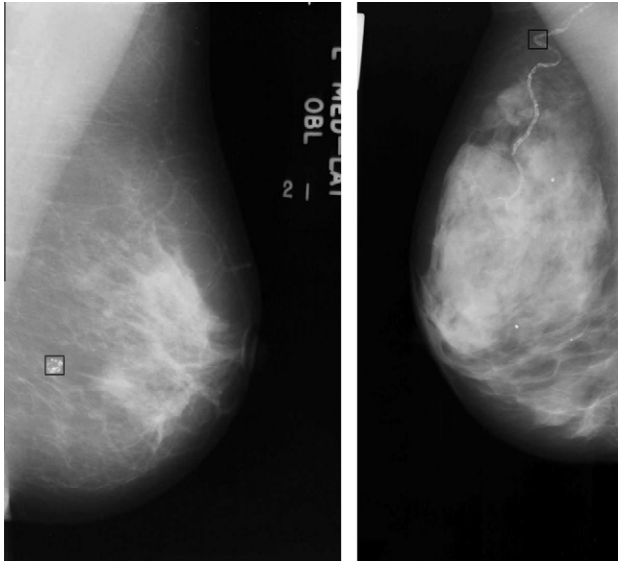
Fig. 9. Mammograms of the MIAS database where the algorithm obtained false positives. The location of the false positive is marked by the black square.

between (3.54, 4.09). Note that these results are similar to the ones obtained when testing the MIAS database. In this case, the best results were obtained using the kernels of size 50×50 and 100×100 (3.5×3.5 and 7.0×7.0 mm^2, respectively). This reduction of the kernel size may indicate that the clusters detected in this database are more subtle and hence difficult to detect than the ones detected in the digitised database. This is an expected behaviour, since digital mammography improves the contrast between the different internal structures.

6. Discussion and conclusion

We have presented a new fully automatic computer aided detection system for microcalcification detection. The core of the system is based on extracting local features for characterising the morphology of the microcalcifications. Afterwards, the proposed approach follows a boosting scheme, allowing the selection of the most salient features at each round. At the testing stage, only these features are computed and used to detect the individual microcalcifications. Subsequently, the cluster detection is performed by locally integrating the individual microcalcifications probability images.

The results of the individual microcalcifications detection are analysed at image level using ROC analysis, obtaining A_z values higher that 0.90 when using the same database for training and testing, in a cross-validation fashion. However, it is well known that using a cross-validation scheme may produce optimistic results and it is worth demonstrating the performance of the approach using less training data. In this sense, we carried out again the microcalcification detection experiment using the MIAS database but using half of the data for training and the other half for testing. In order to obtain significant results, we repeated the experiment 10 times. We obtained mean $A_z = 0.850 \pm 0.032$, which is slightly lower than the results obtained when using the 10-folder cross-validation scheme. This result can be seen as satisfactory considering the fact that the database used in this case was relatively small (just 12 mammograms contained microcalcifications). We can compare our results with current state-of-the-art approaches. Note that each approach used a different set of images coming also from different databases and hence the comparison is only done in a qualitative way. For instance, Chang et al. [31] obtained $A_z = 0.90$ with a database of

194 mammograms, Nunes et al. [32] obtained $A_z = 0.93$ with a database of 121 mammograms, Papadopoulos et al. [7] obtained $A_z = 0.92$ with a database of 60. Note that we obtained similar results but with two larger databases.

On the other hand, the results of the cluster detection are analysed at region level using FROC analysis. Summarising the results, we obtained 1 false positive per image at a sensitivity of 80%, and 4 false positives per image at a sensitivity of 90%. Comparing again with other approaches, Linguraru et al. [28] reached 95% sensitivity with 0.4 false positives per image, while Ge et al. [33] reached 90% sensitivity with 0.96 or 2.52 false positive per image when testing digitised or digital mammograms, respectively. However, the approach of Linguraru et al. [28] was based on using a small database of 82 mammograms from which 58 images contained microcalcifications and 24 were normal ones, while the approach of Ge et al. [33] used 96 mammograms with microcalcifications and 108 normal for the digital dataset and 96 with microcalcifications and 71 normal for the digitised subset. In contrast, we have used larger and more realistic databases, where the number of normal mammograms is large compared to the number of mammograms containing microcalcifications, which is the actual case in screening programs.

A different way to show the robustness of an algorithm is to use one database for training and another one for testing. Hence, with the same set of parameters optimised for the MIAS database and the same training dataset, we also tested a large subset of mammograms extracted from the DDSM database [34]. In order to obtain the results, we subdivided this dataset into three groups, depending on the hospital location and the used scanner machine. According to the DDSM database nomenclature, we used 420 A mammograms (112 of them including clusters), 441 B mammograms (205 including clusters), and 376 D mammograms (65 including clusters). The obtained mean area under the ROC curve in each case was 0.71, 0.75, and 0.84, respectively. Note that worse results were obtained compared to the MIAS dataset, mainly due to fact microcalcifications were detected with lower probabilities than in the MIAS database. This also affected the performance of the algorithm when looking for clusters, obtaining 80% sensitivity at (3.51, 5.13) false positives per image. Analysing the results, we noticed that the main problem was in the high number of false positives, since almost all clusters were correctly detected. These results show that the specificity of the algorithm highly depends on training the system using the own testing database.

The overall computational cost of the approach is relatively high. For instance, the time necessary to perform the training in one fold of the MIAS cross-validation was approximately 82 min, while the mean time for testing one case was 556.14 ± 112.33 s. This large deviation of the time is due to the fact that mammograms of the MIAS have four different sizes. In order to speed up the process we resized the images by a factor of four, needing approximately 65 min for training and only 24.75 ± 5.82 s for testing one case. However, the A_z value for the ROC analysis decreases from 0.903 to 0.856, indicating that it is not a good idea to downsample the mammograms when looking for microcalcifications. Note also that the programming language used to implement our approach has been Matlab and hence, this time can be largely reduced using other programming language like C++, or using the CUDA platform to exploit the benefits of using the GPU hardware acceleration and parallelisation. Another way to improve the velocity of the whole process would be to use directly the image patches to perform the characterisation instead of using the full bank of filters (similar to the approach of Varma and Zisserman [20]). When we performed this test on the MIAS database, although improving the computational time for the feature extraction process, we obtained an $A_z = 0.86$ instead of the $A_z = 0.90$ achieved when using the filters. However, we want to clarify that our image patch

description was not based on the whole framework described in Varma and Zisserman [20] since we only used the extracted image patches required to compute the correlation. Notice that when comparing both approaches using exactly the same image database, better results were obtained using the bank of filters. Note that another way to speed up the time would be to use only suspicious pixels, which can be found by using a pre-processing step [10,26].

In summary, a new approach for the detection of individual microcalcifications and clusters has been presented. The performed experiments have shown the validity of our approach when using either digitised or digital mammograms, obtaining slightly better results when testing the digital database. However, studies with larger databases will be needed in order to show the feasibility of the approach in the clinical routine of screening programs.

Acknowledgement

References

[1] D. Kopans, Breast Imaging, Lippincott-Raven, Philadelphia, 1998.
[2] R. Sivaramakrishna, R. Gordon, Detection of breast cancer at a smaller size can reduce the likelihood of metastatic spread: a quantitative analysis, Acad. Radiol. 4 (1) (1997) 8–12.
[3] S.F. Sener, D.J. Winchester, D.P. Winchester, E. Barrera, M. Bilimoria, E. Brinkmann, E. Alwawi, S. Rabbitt, M. Schermerhorna, H. Du, Survival rates for breast cancers detected in a community service screening mammogram program, Am. J. Surg. 191 (3) (2006) 406–409.
[4] J. Suckling, J. Parker, D.R. Dance, S.M. Astley, I. Hutt, C.R.M. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S.L. Kok, P. Taylor, D. Betal, J. Savage, The mammographic image analysis society digital mammogram database, in: Int. Work. Dig. Mammography, 1994, pp. 211–221.
[5] H.D. Cheng, X. Cai, X. Chen, L. Hu, X. Lou, Computer-aided detection and classification of microcalcifications in mammograms: a survey, Pattern Recogn. 36 (12) (2003) 2967–2991.
[6] R.M. Rangayyan, F.J. Ayres, J.E.L. Desautels, A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs, J. Frankl. Inst. 344 (3–4) (2007) 312–348.
[7] A. Papadopoulos, D.I. Fotiadis, L. Costaridou, Improvement of microcalcification cluster detection in mammography utilizing image enhancement techniques, Comput. Biol. Med. 10 (38) (2008) 1045–1055.
[8] N.R. Pal, B. Bhowmick, S.K. Patel, S. Pal, J. Das, A multi-stage neural network aided system for detection of microcalcifications in digitized mammograms, Neurocomputing 71 (13–15) (2008) 2625–2634.
[9] M. Rizzi, M. D'Aloia, B. Castagnolo, Computer aided detection of microcalcifications in digital mammograms adopting a wavelet decomposition, Integr. Comput.-Aided Eng. 16 (2) (2009) 91–103.
[10] S.N. Yu, Y.K. Huang, Detection of microcalcifications in digital mammograms using combined model-based and statistical textural features, Expert Syst. Appl. 37 (7) (2010) 5461–5469.
[11] A. Papadopoulos, D.I. Fotiadis, A. Likas, An automatic microcalcification detection system based on a hybrid neural network classifier, Artif. Intell. Med. 2 (25) (2002) 149–167.
[12] A. Torrent, A. Oliver, X. Lladó, R. Martí, J. Freixenet, A supervised microcalcification detection approach in digitised mammograms, IEEE Int. Conf. Image Proc. (2010) 4345–4348.
[13] A. Oliver, A. Torrent, M. Tortajada, X. Lladó, M. Peracaula, L. Tortajada, M. Sentís, J. Freixenet, A boosting based approach for automatic microcalcification detection, in: Lecture Notes in Computer Science, vol. 6136, 2010, pp. 251–258.
[14] E. Golobardes, X. Llorà, M. Salamó, J. Martí, Computer aided diagnosis with case-based reasoning and genetic algorithms, Knowledge-Based Syst. 1–2 (15) (2002) 45–52.
[15] J. Ren, ANN vs. SVM: Which one performs better in classification of MCCs in mammogram imaging, Knowledge-Based Syst. 26 (2012) 144–153.
[16] M. Salamó, M. López-Sánchez, Adaptive case-based reasoning using retention and forgetting strategies, Knowledge-Based Syst. 24 (2) (2011) 230–247.
[17] U. Fischer, F. Baum, S. Obenauer, S. Luftner-Nagel, D. von Heyden, R. Vosshenrich, E. Grabbe, Comparative study in patients with microcalcifications: full-field digital mammography vs screen-film mammography, Epidemiol. Rev. 11 (12) (2002) 2679–2683.
[18] K. Murphy, A. Torralba, D. Eaton, W.T. Freeman, Object detection and localization using local and global features, in: Lecture Notes in Computer Science, vol. 4170, 2006, pp. 382–400.
[19] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting, Ann. Stat. 28 (2) (2000) 337–374.
[20] M. Varma, A. Zisserman, A statistical approach to material classification using image patch examples, IEEE Trans. Pattern Anal. Mach. Intell. 31 (11) (2009) 2032–2047.
[21] A. Torralba, K.P. Murphy, W.T. Freeman, Sharing visual features for multiclass and multiview object detection, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) 854–869.
[22] J.R. Quinlan, Bagging, boosting, and C4.5, in: Nat. Conf. on Art. Intell., 1996, pp. 725–730.
[23] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, J. Comput. Syst. Sci. 55 (1) (1997) 119–139.
[24] R. Martí, A. Oliver, D. Raba, J. Freixenet, Breast skin-line segmentation using contour growing, in: Lecture Notes in Computer Science, vol. 4478, 2007, pp. 564–571.
[25] S.M. Kwok, R. Chandrasekhar, Y. Attikiouzel, M.T. Rickard, Automatic pectoral muscle segmentation on mediolateral oblique view mammograms, IEEE Trans. Med. Imag. 23 (9) (2004) 1129–1140.
[26] S. Paradkar, S.S. Pande, Intelligent detection of microcalcifications from digitized mammograms, Sādhāna 1 (36) (2011) 125–139.
[27] H.J. Yoon, B. Zheng, B. Sahiner, D.P. Chakraborty, Evaluating computer-aided detection algorithms, Med. Phys. 34 (6) (2007) 2024–2038.
[28] M.G. Linguraru, K. Marias, R. English, M. Brady, A biologically inspired algorithm for microcalcification cluster detection, Med. Image Anal. 10 (6) (2006) 850–862.
[29] H. Bornefalk, Estimation and comparison of CAD system performance in clinical settings, Acad. Radiol. 12 (2005) 687–694.
[30] H. Bornefalk, A. Bornefalk-Hermansson, On the comparison of FROC curves in mammography imaging, Med. Phys. 32 (2) (2005) 412–417.
[31] T.T. Chang, J. Feng, H.W. Liu, H.H.S. Ip, Clustered microcalcification detection based on a multiple kernel support vector machine with grouped features, in: IAPR Int. Conf. Pattern Recogn., 2008, pp. 1–4.
[32] F.L.S. Nunes, H. Schiabel, C.E. Goes, Contrast enhancement in dense breast images to aid clustered microcalcifications detection, J. Digit. Imag. 1 (20) (2007) 53–66.
[33] J. Ge, M. Hadjiiski, B. Sahiner, J. Wei, M.A. Helvie, C. Zhou, H.P. Chan, Computer-aided detection system for clustered microcalcifications: comparison of performance on full-field digital mammograms and digitized screen-film mammograms, Phys. Med. Biol. 52 (4) (2006) 981–1000.
[34] M. Heath, K. Bowyer, D. Kopans, R. Moore, P.J. Kegelmeyer, The digital database for screening mammography, in: Int. Work. Dig. Mammography, 2000, pp. 212–218.