

Dealing with false positive reduction in mammographic mass detection

Xavier Lladó*, Arnau Oliver, Joan Martí, and Jordi Freixenet

Institute of Informatics and Applications, University of Girona, 17071, Girona, Spain

Abstract. In this paper we analyze a set of false positive reduction methods in the field of mammographic mass detection. The main goal of this false positive reduction process is the discrimination between the true recognized masses and the ones which actually are normal parenchyma. We describe three different approaches to extract breast mass image features. The first approach is based on modeling the tissue variation of both kinds of regions of interest (RoIs) by extracting the principal components (PCA) of a set of already classified RoIs. The second approach is based on an extension of the PCA approach by using the recently proposed 2DPCA algorithm. Finally, the third approach is based on Local Binary Patterns (LBP) for representing texture information and preserving at the same time the spatial structure of the masses. Once those image descriptors are extracted, the system is trained and used for classifying the unknown RoIs. We evaluate our false positive reduction approaches using a set of 1792 suspicious RoIs extracted from the DDSM database, providing a comparison when using different ratios of number of RoIs depicting masses and number of RoIs depicting normal tissue, and also when using different mass sizes, a critical aspect in mass detection systems.

1 Introduction

Breast cancer is one of the most devastating and deadly diseases for women in their 40s in the European Union as well in the United States. It is estimated that between one in eight and one in twelve women will develop breast cancer during their lifetime [1]. The most used method to detect breast cancer is mammography, because it allows the detection of the cancer at its early stages, a crucial issue for a high survival rate. During the last decade several algorithms have been proposed for the automatic detection of masses in mammograms [2, 3]. However, the main drawback of these methods is the high number of obtained false positives [4]. A false positive is a Region of Interest (RoI) – a sub-image containing the suspicious region – being normal tissue but interpreted by the automatic algorithm as a real mass. Therefore, almost all works trying to detect masses in mammography need a final step in order to reduce the number of false positives. This is due to the complexity of the internal breast tissue, which induces the detection of regions which are not masses, but normal variations in tissue characteristics.

In this paper we present three different approaches to perform the false positive reduction in mass detection. The first one is based on modeling the tissue variation by extracting the principal components of a set of already classified RoIs [5]. The second approach is based on an extension of the PCA approach by using the recently proposed 2DPCA algorithm [6]. Finally, our third approach is based on Local Binary Patterns (LBP) for representing texture information and preserving at the same time the spatial structure of the masses. The idea of this proposal is inspired on the recent work of Ahonen et al. [7] in which Local Binary Patterns (LBP) are successfully applied to the face recognition problem. To our knowledge this is the first attempt to use LBP in the field of mammographic mass detection. In order to analyze these approaches, we perform experiments on a complete set of 1792 RoIs extracted from the DDSM database, evaluating the results when using different RoI image sizes, and when using different ratios of number of RoIs depicting masses and RoIs depicting normal tissue in the database.

The rest of the paper is organized as follows. Section 2 presents a briefly state of the art on mass false positive reduction. Section 3 describes our three approaches for mass false positive reduction. Experimental results are presented in Section 4. Finally, the paper ends with conclusions.

2 Related work

Different algorithms have been proposed so far for the mass false positive reduction task. Analyzing these algorithms, we can distinguish between two main strategies. The first one includes the set of algorithms which firstly extracts image features from the RoIs, usually related to their texture, and subsequently trains the classifier. On the other hand, a second strategy handles this problem as a template matching algorithm. Each new image is compared to all the RoIs of the database and then it is classified as an image containing a mass or not. Table 1 summarizes some works belonging to both strategies.

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

Classifier-Based					
Author	Year	Features	Classifier	RoIs	Results
Sahiner [8]	1996	Texture, Morphologic	LDA, NN	168/504	$Az = 0.90$
Christoyianni [9]	2002	Grey-level, Texture, ICA	NN	119/119	88.23%
Qian [10]	2001	Texture, Shape	NN	200/600	$Az = 0.86$
Tourassi [11]	2005	Grey-level	NN	681/984	$Az = 0.84$

Template-Based					
Author	Year	Features	Similarity	RoIs	Results
Chang [12]	2001	Grey-level, shape	Likelihood function	300/300	$Az = 0.83$
Tourassi [13]	2003	Grey-level	Mutual Information	809/656	$Az = 0.87$

Table 1. Summary of the reviewed works on false positive reduction, with the features used, the classifier/similarity used (where LDA means linear discriminant analysis, NN neural network analysis, and ICA independent component analysis), the number of RoIs depicting masses vs the number of normal RoIs, and the results obtained. Note that for all works accuracy is given in terms of Az (the area under the ROC curve) except for the work of Christoyianni et al. [9] which just gives the correct classification percentage.

For instance, Sahiner et al. [8] extracted a huge set of features, and subsequently used genetic algorithms to select the most discriminative ones. With this subset of features, a neural net (NN) and a linear classifier (LDA) are trained and used to classify a new RoI. A similar strategy is used by Christoyianni et al. [9], who extracted grey-level, texture, and features related to independent component analysis (ICA), and use them to train a neural net. Note also, that they apply a principal component analysis (PCA) pre-processing step to reduce the complexity of the problem. On the other hand, Qian et al. [10] analyzed the implementation of an adaptive module to improve the performance of an automatic procedure which consists of training a Kalman-filter based neural net using features obtained from a wavelet decomposition.

The works of Chang et al. [12] and Tourassi et al. [13] are based on comparing a new RoI with all the RoIs in the database (template-based approaches). The two most clear differences between them arise from the similarity measure and the database used. More specifically, the former developed a likelihood measure which depends on the grey-level and the shape of the RoIs. Both parameters were compared with the new RoI and the set of RoIs present in the database, which was only composed by RoIs depicting masses. From this comparison a likelihood measure was computed. On the other hand, the work of Tourassi et al. [13] consists of comparing all the RoIs of the database (including RoIs with and without masses) with the new one using a mutual information based similarity measure. Thus, the new RoI will be labeled as belonging to the closest class.

Note that with the last strategy, the similarity used for classifying has to be re-computed for each new element, as it measures the difference between the new RoI and all the RoIs in the database. Observe also that among all those works showed in Table 1, one of the main differences in terms of producing their results is the ratio between the number of RoIs depicting masses and the total number of RoIs. This is an important issue because the number of wrong classified RoIs will increase as the number of normal RoIs increases. One should remember that the aim of this step is to reduce the number of false positives, which is usually higher than the number of true positives. We will analyze this issue in the experimental results.

3 Approaches to mass false positive reduction

In the following sections, we will present three different approaches to perform the false positive reduction: the first one based on the use of PCA like in the eigenfaces algorithm; the second one based on the recent developed 2DPCA decomposition instead of using the typical PCA; and finally, the third one based on the textural features extracted from LBP. Note that these methods should be grouped into the classifier-based strategy.

3.1 PCA-based false positive reduction

Our first approach to deal with the mass false positive reduction is based on the idea of the eigenfaces approach [14]. We assume that a database of already classified RoIs (RoIs containing masses and RoIs of normal tissue¹) is available. Different instances for each class are included in the database. Their intra class variability is mainly due to grey-level and texture differences and to the shape and size of the mass or other structures present in the RoI.

¹Note that although in this work the database contains only two types of RoIs (masses or normal tissue) this can be extended to include other RoIs containing microcalcifications or architectural distortions.

According to the Karhunen-Loeve transform, the usefulness of the different eigenvectors to characterize the variation among the images is ranked by the value of the corresponding eigenvalue. Note that in this case, instead of eigenfaces we should refer them as eigenrois. Thus, the eigenrois span the RoI subspace of the original image space, and each RoI can be transformed into this space by using them. The result of this transformation is a vector of weights describing the contribution of each eigenroi in representing the corresponding input image. Therefore, we propose to use this vectors to construct the models for the training step (see [5] for more details). In our experiments we retained 95% of the information when computing the eigenrois.

When a new RoI has to be tested, it will be classified as belonging to the most similar class. Although in this paper we use the k-Nearest Neighbour classifier to perform the classification, in our previous work [5] we also used a Bayesian combination of this algorithm with the C4.5 decision tree to compare the results.

3.2 2DPCA-based false positive reduction

The 2DPCA approach [15] is a recent improvement of the typical eigenfaces approach. As the authors argue 2DPCA has important advantages over PCA in two main aspects: firstly, it is simpler and more straightforward to use for image feature extraction since 2DPCA is directly based on the image matrix, and secondly, it is easier to accurately evaluate the covariance matrix. In the original eigenfaces approach, each input image of size $m \times n$ is transformed into a vector of size $m \cdot n$, in contrast to the natural way to deal with two dimensional data, which would be treating it as a matrix.

In [6] we presented an extension of the PCA approach to mass false positive reduction by using the 2DPCA algorithm. One of the main differences with respect the standard PCA approach is that while for PCA each principal component is a scalar, for 2DPCA each principal component is a vector. Therefore, it is this set of vectors for image that is used to construct the feature image. In a similar way to the eigenfaces approach, comparing images means to compare the constructed features. As the dimension of the feature space has increased in one dimension respect to the PCA approach, now the comparison of images is done by comparing matrices (see [6] for the detailed mathematical background).

3.3 LBP-based false positive reduction(LBP)

The original LBP operator was introduced by Ojala et al. [16] with the idea to perform grey scale invariant two-dimensional texture analysis. The LBP operator labels the pixels of an image by thresholding the neighbourhood (i.e. 3×3) of each pixel with the centre value and considering the result of this thresholding as a binary number. When all the pixels have been labeled with the corresponding LBP codes, the histogram of the labels is computed and used as a texture descriptor. The authors proposed to derive different LBP operators at circular neighbourhoods of any quantization of the angular space and at any spatial resolution. Therefore, it is possible to derive the operator for a general case based on a circularly symmetric neighbourhood of P members on a circle of radius R . Another extension of LBP was the use of the so called uniform patterns [17]. A LBP is called uniform if it contains at most two bitwise transitions from 0 to 1 or vice versa when the binary string is considered circular. In this paper, we shall refer the uniform LBP operator as $LBP_{P,R}^{u2}$, where (P, R) indicates the neighbourhood and $u2$ uniform patterns.

Inherently to the mass false positive reduction problem, and in particular with the images we are dealing with, texture and the image spatial information play a key role in correctly detecting the masses. Due to this fact, we base this approach on the use of LBP for representing salient micro-patterns, and an adaptation of these descriptors for preserving also the spatial structure of the masses. The idea of our proposal has been inspired on the recent work of Ahonen et al. [7] in which LBP is used to perform face recognition. Our general procedure consists in using the LBP texture descriptor to build several local descriptions of the RoI and combining them into a global description. Afterwards, this global LBP descriptor is used to reduce the false positives, classifying the RoIs between true masses and normal tissue.

Initially, the RoI image – which contains the suspicious mass – is divided into several local regions (for instance, 5×5 rectangular regions). From these regions, texture descriptors are independently extracted using LBP and then concatenated to form a global description of the RoI based on the textural information of each region and its spatial distribution. Note that following this methodology, the basic LBP histogram is extended into a spatially enhanced histogram which encodes both the local region appearance and the spatial relations (global geometry) of the mass. In the final histogram, the RoI is described on three different levels of locality: the labels for the histogram contain the pixel-level texture patterns, the labels are summed over a small region to produce information on a regional level and finally the regional histograms are concatenated to build a global description of the mass. The final step is the mass classification. For this purpose we use the well-known Support Vector Machines (SVM) [18] with polynomial kernel to perform a binary decision between RoIs depicting a true mass and RoIs depicting normal parenchyma.

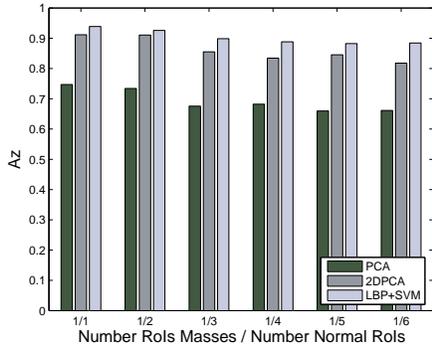


Figure 1. A_z values obtained with the methods PCA, 2DPCA and LBP+SVM. Each A_z value is the mean computed from the results of different RoI image sizes.

4 Experimental results

The three described approaches have been evaluated using a database of 1792 RoIs extracted from the DDSM mammographic database [19]. From this set, 256 depicted a true mass, while the rest 1536 were normal, but suspicious tissue. According to the size of the lesion, we use six different groups of RoI images, corresponding to the following mass sizes intervals: $< 0.10 \text{ cm}^2$, $(0.10 - 0.60) \text{ cm}^2$, $(0.60 - 1.20) \text{ cm}^2$, $(1.20 - 1.90) \text{ cm}^2$, $(1.90 - 2.70) \text{ cm}^2$, $> 2.70 \text{ cm}^2$. The number of masses in each interval were 28, 32, 37, 57, 69, and 33, respectively. The evaluation of our experiments is done by using a leave-one-out scheme and Receiver Operating Characteristics (ROC) analysis [20]. In such analysis, widely used in the medical field, a graphical curve represents the true positive rate as a function of the false positives rate. The percentage value under the curve (known as A_z) is an indication for the overall performance of the observer, and is typically used to analyze the performance of the algorithms.

In order to perform a more global evaluation we compute the A_z value for different ratios of number of RoIs depicting masses and number of RoIs depicting normal tissue (from ratio 1/1 to ratio 1/6). The idea of analyzing these different ratios is twofold: firstly, to evaluate the performance of our method on different levels of difficulty, and secondly, to allow the comparison of our proposal with existing methods (see the methods in Table 1). Notice that previous works only provide results for specific ratios. Hence, analyzing all these ratios will enable the comparison with them.

Figure 1 shows the obtained mean A_z values for each specific ratio when testing our proposal at different RoI image sizes. We include a quantitative comparison with the three described approaches: PCA, 2DPCA and LBP. For the LBP approach we empirically decided to use the basic $LBP_{8,1}^{u2}$ operator with the RoI image divided into 5×5 regions, obtaining an overall mean A_z value of 0.88 ± 0.05 . With the aim of improving these results and obtaining a more accurate description of the central area, we added a new set of LBP operators for the 3×3 central regions. In particular, we computed two new LBP responses varying the radius R according to each RoI image size (i.e. $R = 1, 4$ and 6 were used for the first RoI image size). The global descriptor was then obtained concatenating the 43 ($25 + 9 + 9$) histograms of all the regions and LBP operators. This was the final descriptor we used for the experiments presented in Figure 1 since provided a good trade-off between performance and feature vector length. Observe that the performance of LBP was clearly better than the PCA method. The results were also better than the 2DPCA approach, specially for the cases in which we had smaller ratios 1/4, 1/5 and 1/6. Note also that the use of 2DPCA itself allowed to obtain better results than the original PCA approach. Using PCA we obtained an overall mean A_z value of 0.67 ± 0.09 for different RoI image sizes and ratios, using 2DPCA the A_z value was 0.87 ± 0.08 , while using the LBP approach the A_z value was 0.91 ± 0.04 . Note that these results could be used to perform a qualitative comparison – in terms of A_z value – with the rest of the approaches described in Section 2. However, we want to clarify that not all the methods used the same databases and therefore our aim is only to provide a general view of the performance of our approaches. Note that using the LBP approach we obtained better results than those reported by the works which used ratios of 1/1 and 1/2, and similar results than those obtained by Sahiner et al. for the ratio 1/3.

Regarding the behaviour when varying the RoI image sizes, we observed that LBP provided better and more constant overall results than PCA and 2DPCA methods (see the standard deviations). Finally, Table 2 illustrates the A_z values for the specific ratio 1/3. Note that the three approaches are more suitable for false positive reduction of larger masses than smaller ones. This is due to the fact that larger masses have a larger variation in grey-level contrast with respect

		Lesion Size (in cm^2)					
		<0.10	0.10-0.60	0.60-1.20	1.20-1.90	1.90-2.70	>2.70
Alg.	PCA	0.53	0.70	0.70	0.68	0.72	0.83
	2DPCA	0.81	0.83	0.87	0.84	0.89	0.93
	LBP	0.86	0.90	0.96	0.91	0.86	0.92

Table 2. A_z results (ratio 1/3) for the classification of masses using RoIs extracted from the DDSM database, detailed per size (in cm^2).

to their surrounding tissue than smaller masses, which are usually more subtle, even for an expert.

5 Conclusions

In this paper we have presented three different approaches for the mass false positive reduction problem: one based on PCA, another based on 2DPCA, and the last one based on LBP. Our experiments have shown that LBP and 2DPCA features are effective and efficient for false positive reduction at different RoI image sizes, outperforming the traditional PCA approach. Moreover, LBP provides the best results when using different ratios of number of RoIs with masses and number of RoIs with normal tissue. Regarding the behaviour when varying the RoI image sizes, LBP also provides better and more constant results than PCA and 2DPCA methods.

Acknowledgments

This work was partially supported by MEC grant TIN2006-08035 and grant IdIBGi-UdG 91060080.

References

1. Eurostat. "Health statistics atlas on mortality in the European Union." *Official Journal of the European Union* 2002.
2. R. Zwiggelaar, T. C. Parr, J. E. Schumm et al. "Model-based detection of spiculated lesions in mammograms." *Med. Image Anal.* **3(1)**, pp. 39–62, 1999.
3. B. Sahiner, N. Petrick, H. P. Chan et al. "Computer-aided characterization of mammographic masses: Accuracy of mass segmentation and its effects on characterization." *IEEE Trans. Med. Imag.* **20(12)**, pp. 1275–1284, 2001.
4. R. M. Nishikawa & M. Kallergi. "Computer-aided detection, in its present form, is not an effective aid for screening mammography." *Med. Phys.* **33**, pp. 811–814, 2006.
5. A. Oliver, J. Martí, R. Martí et al. "A new approach to the classification of mammographic masses and normal breast tissue." In *IAPR Int. Conf. Pattern Recogn.*, volume 4, pp. 707–710, 2006.
6. A. Oliver, X. Lladó, J. Martí et al. "False positive reduction in breast mass detection using two-dimensional PCA." In *Lect. Not. Comp. Sc.*, volume 4478, pp. 154–161, 2007.
7. T. Ahonen, A. Hadid & M. Pietikäinen. "Face detection with local binary patterns: Application to face recognition." *IEEE Trans. Pattern Anal. Machine Intell.* **28(12)**, pp. 2037–2041, 2006.
8. B. Sahiner, H. P. Chan, D. Wei et al. "Image feature selection by a genetic algorithm: Application to classification of mass and normal breast tissue." *Med. Phys.* **23**, pp. 1671–1684, 1996.
9. I. Christoyianni, A. Koutras, E. Dermatas et al. "Computer aided of breast cancer in digitized mammograms." *Comp. Med. Imag. Grap.* **26**, pp. 309–319, 2002.
10. W. Qian, X. Sun, D. Song et al. "Digital mammography - wavelet transform and kalman-filtering neural network in mass segmentation and detection." *Acad. Radiol.* **8(11)**, pp. 1074–1082, 2001.
11. G. D. Tourassi, N. H. Eltonsy, J. H. Graham et al. "Feature and knowledge based analysis for reduction of false positives in the computerized detection of masses in screening mammography." In *IEEE Conf. Eng. Med. Biol. Soc.*, pp. 6524–6527, 2005.
12. Y. H. Chang, L. A. Hardesty, C. M. Hakim et al. "Knowledge-based computer-aided detection of masses on digitized mammograms: A preliminary assessment." *Med. Phys.* **28(4)**, pp. 455–461, 2001.
13. G. D. Tourassi, R. Vargas-Vorecek, D. M. Catarious et al. "Computer-assisted detection of mammographic masses: A template matching scheme based on mutual information." *Med. Phys.* **30(8)**, pp. 2123–2130, 2003.
14. M. A. Turk & A. P. Pentland. "Eigenfaces for recognition." *J. Cognitive Neurosci.* **3(1)**, pp. 71–86, 1991.
15. J. Yang, D. Zhang, A. F. Frangi et al. "Two-dimensional PCA: a new approach to appearance-based face representation and recognition." *IEEE Trans. Pattern Anal. Machine Intell.* **26(1)**, pp. 131–137, 2004.
16. T. Ojala, M. Pietikäinen & D. Harwood. "A comparative-study of texture measures with classification based on feature distributions." *Pattern Recogn.* **29(1)**, pp. 51–59, 1996.
17. T. Ojala, M. Pietikäinen & T. Mäenpää. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." *IEEE Trans. Pattern Anal. Machine Intell.* **24(7)**, pp. 971–987, 2002.
18. V. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, New York, 1998.
19. M. Heath, K. Bowyer, D. Kopans et al. "The Digital Database for Screening Mammography." In *Int. Work. Dig. Mammography*, pp. 212–218, 2000.
20. C. E. Metz. "Evaluation of digital mammography by ROC analysis." In *Int. Work. Dig. Mammography*, pp. 61–68, 1996.