Computerized Medical Imaging and Graphics 33 (2009) 415–422

Contents lists available at ScienceDirect

Computerized Medical Imaging and Graphics

journal homepage: www.elsevier.com/locate/compmedimag

A textural approach for mass false positive reduction in mammography

X. Lladó*, A. Oliver, J. Freixenet, R. Martí, J. Martí

Computer Vision and Robotics Group, IIiA-IdIBGi, University of Girona, Campus Montilivi s/n, 17071 Girona, Spain

ARTICLE INFO

Article history:
Received 31 October 2008
Received in revised form 25 March 2009
Accepted 26 March 2009

Keywords:
Breast cancer
Image analysis
Mammographic automatic mass detection
False positive reduction
Textural information
Local binary patterns

ABSTRACT

During the last decade several algorithms have been proposed for automatic mass detection in mammographic images. However, almost all these methods suffer from a high number of false positives. In this paper we propose a new approach for tackling this false positive reduction problem. The key point of our proposal is the use of Local Binary Patterns (LBP) for representing the textural properties of the masses. We extend the basic LBP histogram descriptor into a spatially enhanced histogram which encodes both the local region appearance and the spatial structure of the masses. Support Vector Machines (SVM) are then used for classifying the true masses from the ones being actually normal parenchyma. Our approach is evaluated using 1792 ROIs extracted from the DDSM database. The experiments show that LBP are effective and efficient descriptors for mammographic masses. Moreover, the comparison with current methods illustrates that our proposal obtains a better performance.

1. Introduction

Breast cancer is considered a major health problem in western countries, and indeed it constitutes the most common cancer among women in the European Union [1]. A study developed in 2003 by the American Cancer Society estimates that in the United States between one in eight and one in twelve women will develop breast cancer during their lifetime [2]. In the European Community, breast cancer represents 19% of cancer deaths and the 24% of all cancer cases [3]. Nearly 25% of all breast cancer deaths occur in women diagnosed between ages 40 and 49 years. In the United States, for instance, breast cancer remains the leading cause of death for women in their forties [4]. However, although breast cancer incidence has increased over the past decade, breast cancer mortality has declined among women of all ages [5]. This favourable trend in mortality reduction is considered to be related to the widespread adoption of mammography screening [5–8] which allows to detect the cancer at its early stages, and to the improvements made in breast cancer treatment [4].

Among the different imaging modalities used for breast cancer detection, mammography remains the key screening tool for the detection of breast abnormalities. In a recent study, Vacek et al. [9] show that the proportion of breast tumours that were detected in Vermont (US) by screening mammography increased from 2% during 1974–1984 to 36% during 1995–1999. However, it is also well known that expert radiologists can miss a significant portion of abnormalities [10,11]. In addition, a significant number of mammographic abnormalities turn out to be benign after biopsy [12,13].

The introduction of digital mammography gave the opportunity of increasing the number of commercial Computer-Aided Detection (CAD) systems to help radiologists to interpret and diagnose mammograms. The idea of computer systems aiding radiologists to detect breast cancer is not recent [14]. However, the nowadays rapid development of full digital mammographic systems has being accompanied by the natural increase of such systems. A CAD is a set of automatic or semiautomatic tools developed to assist radiologists in the detection and / or evaluation of mammographic images [11,15,16].

Back in 2001 Freer and Ulissey [15] using a database containing 12860 patients conclude that the use of CAD in the interpretation of screening mammograms can increase the detection of early-stage malignancies. However, controversial results published in 2005 using a database of 6111 women claimed to show that CAD, in its present form, is not effective in the sense that there was no significant difference observed in cancer detection rates when using or not the CAD [17]. The main drawback of CAD systems is the number of false positives[1] obtained, which makes the radiologist not really trust them [18]. This is a major issue for the low number of malignancies within the screening population, which is supposed to be around 6 out of 1000 screened cases [19]. The so called false positive reduction algorithms try to solve this drawback, i.e. given a Region

* Corresponding author. Tel.: +34 972418878; fax: +34 972418259.
E-mail addresses: llado@eia.udg.edu (X. Lladó), aoliver@eia.udg.edu (A. Oliver), jordif@eia.udg.edu (J. Freixenet), marly@eia.udg.edu (R. Martí), joanm@eia.udg.edu (J. Martí).

[1] A false positive is a region being normal tissue but interpreted by the automatic algorithm as a suspicious one.

0895-6111/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compmedimag.2009.03.007

of Interest (ROI) – a sub-image containing the suspicious region – the aim is to validate whether it contains a real lesion or it is only a region depicting normal parenchyma. This will be the focus of our work: reduction of the false positive detection of masses, which are a common lesion found in mammography.

In this paper we propose a novel approach to perform mass false positive reduction by using the textural properties of the masses. The idea of using textural information for solving this problem is not new and has been previously introduced in several works [20–22,38]. However, we study here the use of Local Binary Patterns (LBP) [23] to characterize micro-patterns (i.e. edges, lines, spots, flat areas) and preserve at the same time the spatial structure of the masses. To our knowledge, our recent seminal work [24] is the first attempt to use LBP in the field of mammographic mass detection. LBP and its extensions have performed very well in various comparative studies and have been applied successfully in different real-world texture analysis problems [25,26]. In this work we show that using a LBP characterization and Support Vector Machines (SVM) in order to classify the ROIs between real masses and normal parenchyma we are able to improve the results on mass false positive reduction. We perform experiments on a complete set of 1792 ROIs extracted from the Digital Database for Screening Mammography (DDSM) database, analyzing, evaluating and discussing 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 obtained results and the comparison with previous works demonstrate the validity of our approach for reducing false positives. The rest of the paper is organized as follows. Section 2 reviews the false positive reduction methods in the mass detection problem. In Section 3 we describe our methodology, introducing LBP descriptors and presenting our proposed extension. Section 4 presents the obtained results. Finally, the paper ends with discussion and conclusions.

2. Background

During the last decade several algorithms have been proposed for the automatic mass detection purpose [27–29]. However, as we have seen in the introduction the main drawback of these methods is the high number of false positives obtained [18]. Almost all works trying to automatically detect masses in mammography need this final step in order to reduce the high number of false positives. Fig. 1 shows different examples of suspicious masses (represented in black squares) detected by a mass detection algorithm. Note that only one detected ROI per image is a true mass while the remaining ones are 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.

A set of different techniques for false positive reduction have been developed during the last years. These algorithms typically work with one view, although recent approaches have tackled the case of more than one view [30,31]. However, these approaches based on multiple views have three main disadvantages: the images must be correctly registered to allow a correct region comparison, the gray-level values have also to be registered, and finally, there are some specific cases where the comparison is not feasible, for instance, those cases in which patients have suffered a previous breast operation. In contrast, is important to notice that algorithms based on only one view can always be used.

Algorithms for mammographic mass false positive reduction using a single image are typically based on the classification of the ROIs as normal tissue or as depicting the abnormality. Hence, these algorithms are based on a typical classifier scheme: given a database of known cases the system learns how to differentiate between both kinds of ROIs. Subsequently, once the system has been trained, a new ROI can be classified. Among all these false positive reduction algorithms we can distinguish between two main strategies. The first one includes the set of algorithms which firstly extracts features from the ROIs, usually related to their texture, and subsequently trains the classifier. On the other hand, the second strategy handles this problem as a template matching approach. Each new ROI image is compared to all the ROI images of the database and then it is classified as an image containing a mass or not. Note that the first strategy is based on feature vectors extracted from the ROIs while the second one is based on comparing directly the new ROI image with all the images of the database using a similarity measure. Table 1 summarizes some works belonging to both strategies.

Observing the table we can see that one of the main differences among all these works 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 wrongly classified ROIs is likely to increase as the number of normal ROIs increases. One should remember that the aim of this work is to reduce the number of false positives. However, all these methods have to choose the trade-off between reduction of the false positive fraction and increase of the false negative fraction. This step is usually done optimizing the parameters to obtain the best results. Observe also that the evaluation of the methods illustrated in Table 1 is presented using Receiver Operating Characteristics (ROC) analysis [32]. In such analysis, widely used in the medical field, a graphical curve which represents the true positive rate as a function of the false positive rate is computed. The percentage area under the curve (known as Az) is an indication for the overall performance of the method, and is typically used to analyze the performance of the algorithms. Sahiner et al. [20] extracted a huge set of features, and used genetic

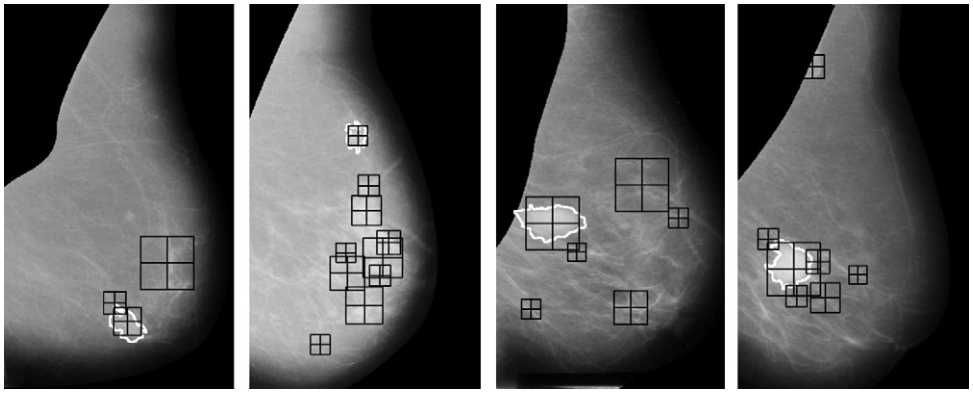

Fig. 1. Example of suspicious masses (black squares) obtained by a mass detection algorithm. Real masses are represented by a white contour. Note that almost all the suspicious masses are false positives.

Table 1
Summary of the reviewed works on false positive reduction. We detail the features used, the classifier/similarity used (where LDA means linear discriminant analysis; NNet, neural network analysis; NN, nearest neighbor classifier; ICA, independent component analysis; SVM, support vector machines; and PCA, principal 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 works of Christoyianni et al. [22] which just gives the correct classification percentage and Angelini et al. [33] where the results are given in terms of ROC curves.

Classifier-based					
Author	Year	Features	Classifier	ROIs	Results
Sahiner et al. [20]	1996	Texture, morphologic	LDA, NNet	168/504	$Az = 0.90$
Qian et al. [21]	2001	Texture, shape	NNet	200/600	$Az = 0.86$
Christoyianni et al. [22]	2002	Gray-level, texture, ICA	NNet	119/119	88.23%
Tourassi et al. [39]	2005	Gray-level	NNet	681/984	$Az = 0.84$
Angelini et al. [33]	2006	Wavelet decomposition	SVM	1000/5000	ROC
Oliver et al. [34]	2006	PCA	C4.5 + NN	196/392	$Az = 0.83$
Varela et al. [38]	2007	Gray-level, morphologic	NNet	60/60	$Az = 0.90$
Oliver et al. [36]	2007	2DPCA	NN	256/1536	$Az = 0.86$
Template-based					
Author	Year	Features	Similarity	ROIs	Results
Chang et al. [40]	2001	Gray-level, shape	Likelihood function	300/300	$Az = 0.83$
Tourassi et al. [41]	2003	Gray-level	Mutual information	809/656	$Az = 0.87$
Tourassi et al. [42]	2007	Gray-level, entropy	Mutual information	901/919	$Az = 0.81$

algorithms to select the most discriminative ones. With this subset of features, a neural net (NNet) and a linear classifier (LDA) were trained and used to classify a new ROI. A similar strategy was used by Christoyianni et al. [22], who extracted gray-level, texture, and features related to independent component analysis (ICA) to train a neural network. Note also, that they applied a principal component analysis (PCA) pre-processing step to reduce the complexity and dimensionality of the problem. On the other hand, Qian et al. [21] analyzed the implementation of an adaptive module to improve the performance of an automatic procedure which consists in training a Kalman-filter based neural net using features obtained from a wavelet decomposition. Angelini et al. [33] also proposed to use a wavelet decomposition to extract features. SVM were used in their work in order to classify the ROIs. It is important to notice that they only provide ROC curves to evaluate the results. Therefore, no *Az* value is available for comparison. A different strategy was recently proposed by Oliver et al. [34], who adapted the idea of the eigenfaces approach [35] to the mass detection problem. They introduced the concept of eigenrois, which span the ROI subspace of the original image space. The result of this transformation was a vector of weights describing the contribution of each eigenroi in representing the corresponding input image. They proposed to use these vectors to construct the models for the training step. In [36] the authors extended their previous method by using the 2DPCA method [37] instead of the standard PCA technique, improving the performance of the false positive reduction. Finally, Varela et al. [38] proposed an strategy based on extracting gray-level and morphologic features, and training a neural net (NNet) used to classify the new ROI.

As shown in Table 1, the works of Chang et al. [40] and Tourassi et al. [41,42] are based on comparing a new ROI with all the ROIs in the database (template-based approach). 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 gray-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 works of Tourassi et al. [41,42] consisted in 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 was labeled as belonging to the closest class. Note that with the methods based on the template-based strategy, the similarity measure used for classifying the ROIs 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.

In the following section we will introduce our false positive reduction methodology based on a classifier-based scheme and LBP texture features extracted from the ROIs.

3. Methods

Ojala et al. [43] introduced the original texture LBP operator with the idea to perform gray scale invariant two-dimensional texture analysis. Being theoretically simple, LBP has been demonstrated to be a rich descriptor in many applications. The LBP operator labels the pixels of an image by thresholding the neighborhood (i.e. 3×3) of each pixel with the center value and considering the result of this thresholding as a binary number. Fig. 2 (a) shows an example of how to compute a LBP code. Note that the gray-level of the center pixel is used as a threshold (1 is assigned to higher or equal gray levels, while 0 to the lower ones). 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. Initially, the limitation of this basic LBP operator was its small 3×3 neighborhood since it can not deal with dominant features with large scale structures. Due to this fact, the operator was later extended to use neighborhoods of different sizes [23]. The idea of this operator is the detection of LBP at circular neighborhoods 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 neighborhood of P members on a circle of radius R (Fig. 2(b) illustrates examples of circular neighborhoods). The gray values of neighbors which do not fall exactly in the center of pixels are estimated by interpolation. In addition to evaluating the performance of individual operators of a particular configuration (P, R), one could analyze and combine responses of multiple operators using different (P, R) parameters.

Another extension of LBP was the use of the so called uniform patterns [23]. 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. For example, 00011100 and 11100011 are uniform patterns. As stated by Ojala et al., the uniform patterns account for nearly 90% of all patterns in the (8,1) neighborhood and for about 70% in the (16,2) neighborhood in texture images. In this paper, we shall refer the uniform LBP operator as $\mathrm{LBP}_{P,R}^{u2}$, where the sub-

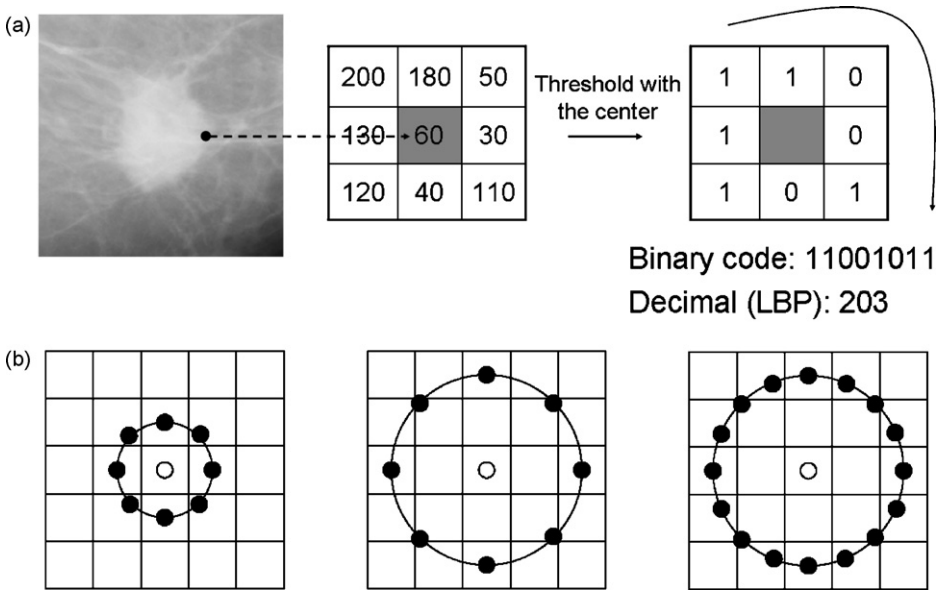

Fig. 2. LBP computation. (a) Example of the basic LBP operator. (b) Examples of the common circular LBP neighborhoods: (8,1), (8,2) and (16,2) respectively.

script represents using the operator in a (P, R) neighborhood and the superscript $u2$ indicates using uniform patterns.

As in the original work of Ojala et al. the histogram of the labeled image $f_l(x, y)$ is used as a descriptor. We can define this histogram as

$$H_i = \sum_{x,y} I(f_l(x, y) = i), \quad i = 0, \ldots, n-1 \tag{1}$$

where n is the number of different labels produced by the LBP operator, and $I(A) = 1$ when A is true, while $I(A) = 0$ when A is false. This discrete occurrence histogram of the uniform patterns computed over an image or an image region contains information about the distribution of the local micro-patterns such as edges, spots and flat areas, and has been demonstrated to be a very powerful texture descriptor. Note that all these patterns are typical descriptors of mammographic lesions such as spiculated lesions (edges), small masses (spots) and larger masses (flat areas).

3.1. Using LBP for mass false positive reduction

Texture features have been utilized for many medical image applications, and have proven to be useful in discriminating texture classes in X-ray mammography, as well as in other modalities for breast cancer detection. Specifically, in mammographic images, texture and its gray-level spatial information play also a key role in correctly detecting masses (see related works [20–22,38]). Due to this fact, we study here the use of LBP for combining characteristics of statistical and structural texture analysis, describing the texture with micro-primitives (often called textons) and their statistical placement rules. This idea has been successfully applied in other domains [26] where the analyzed image is divided into several regions from which the LBP feature distributions are extracted and concatenated into a new feature vector used as a final descriptor.

Following the same idea, 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 the one used to reduce the false positives, classifying the ROIs between true masses and normal tissue (see Fig. 3 for some examples of ROI images).

Initially, the ROI image – which contains the suspicious mass – is divided into several local regions. See Fig. 4 for an example of a mass image (left image) divided into 5×5 rectangular regions. Notice that in this step one could use different divisions of different size and shape. This will be analyzed in the results section where different divisions ($1 \times 1, 3 \times 3, 5 \times 5$, and 7×7) are tested. From these regions, LBP histograms are independently extracted and then concatenated to form a global description of the ROI. Observe that we could analyze and combine responses of multiple LBP operators using different (P, R) parameters. Moreover, we could also compute a set of different LBP operators for some specific regions of the ROI aiming to improve the final texture descriptor. This is also illustrated in Fig. 4 where the regions with higher probability to contain mass information – the central 3×3 regions showed in the right image – are used to compute different LBP operators. This point will be further discussed in the experimental section where several combinations of LBP operators are evaluated.

Following our methodology, the basic LBP histogram is extended into a spatially enhanced histogram which encodes both the local region appearance and the spatial relations of the mass. The ROI image is divided into m small regions $R_0, R_1, \ldots, R_m$ and the spatially enhanced histogram is defined as

$$H_{i,j} = \sum_{x,y} I(f_l(x, y) = i), \quad (x, y) \in R_j \tag{2}$$

where $i = 0, \ldots, n-1, j = 0, \ldots, m-1$. In this 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 of our proposal is the mass classification. For this purpose we use the well-known SVM technique [44] which performs an implicit mapping of data into a higher dimensional feature space, where linear algebra and geometry can be used to separate data. SVM based approaches have provided good results in many applications, including texture classification problems. For our specific mass false positive reduction problem, SVM with a polynomial kernel is used to provide a membership between ROIs depicting a true mass and ROIs depicting normal parenchyma. In the results section we will compare the results obtained by using SVM with those obtained using a Nearest Neighbor (NN) classifier.

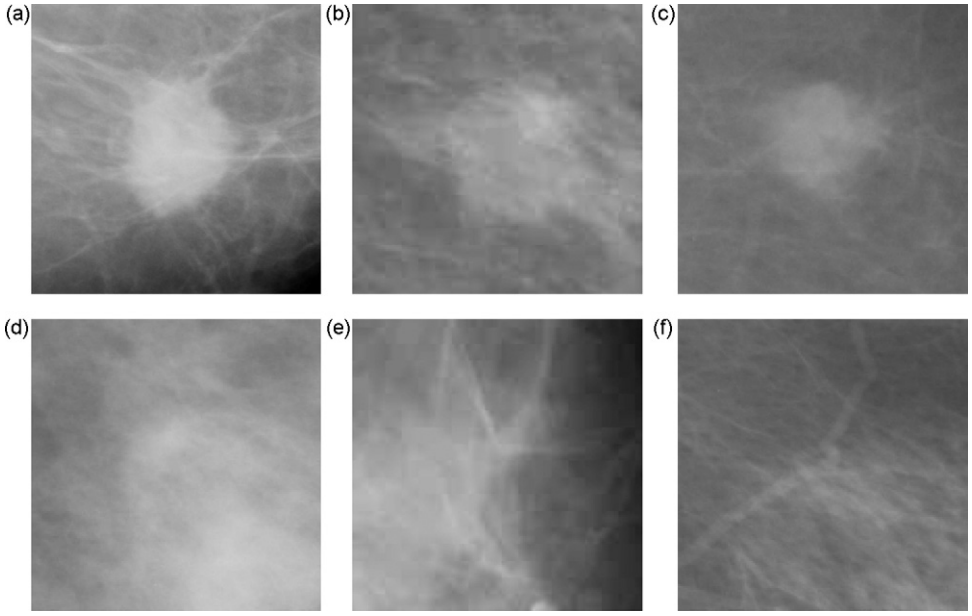

Fig. 3. Examples of ROIs: (a–c) ROIs with masses; (d–f) ROIs without masses.

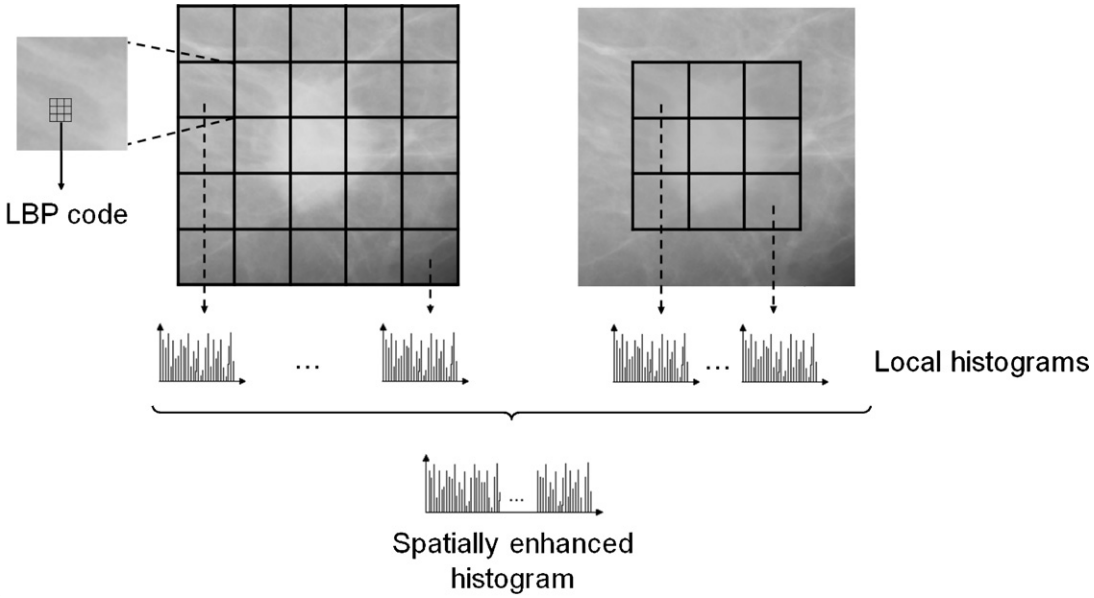

Fig. 4. Strategy for computing the LBP descriptor. The ROI image is divided into small regions, where LBP feature distributions are extracted and concatenated to form the spatially enhanced histogram used as a final descriptor. Note that one could compute different LBP operators for some specific regions (3×3 central regions in the example) and combine them into the final descriptor.

4. Results

This section is divided into the following subsections. First of all, the database and the methodology used to evaluate the results are presented. Afterwards, the parameter optimization done in order to compute the final LBP descriptors is analyzed. We then present the obtained results when using different ratios of number of ROIs depicting masses and ROIs depicting normal tissue. Finally, we include the obtained results when using different ROI image sizes.

4.1. Database and evaluation methodology

Our LBP approach has been evaluated using a database of 1792 ROIs extracted from the DDSM mammographic database [45]. From this set, 256 depicted a true mass, while the rest 1536 were normal, but suspicious tissue. Different scanners were used in the DDSM for digitizing the mammograms of the database: a DBA M2100 ImageClear ($42 \times 42\,\mu m$ pixel resolution), a Howtek 960 ($43.5 \times 43.5\,\mu m$ pixel resolution) a Lumisys 200 Laser ($50 \times 50\,\mu m$ pixel resolution), and a Howtek MultiRad850 ($43.5 \times 43.5\,\mu m$ pixel resolution). All the images were 12 bits per pixel. Finally, the images were re-scaled to have the same resolution: $50\,\mu m$. According to the size of the lesion, we use six different groups of ROI images, corresponding to the following mass sizes intervals: $<10\,mm^2$, $(10\text{-}60)\,mm^2$, $(60\text{–}120)\,mm^2$, $(120\text{–}190)\,mm^2$, $(190\text{–}270)\,mm^2$, $>270\,mm^2$. The number of masses in each interval were 28, 32, 37, 57, 69, and 33, respectively. Note we are dealing with different lesion sizes, an important aspect for correctly classifying the masses.

The evaluation of our experiments is done by using a leave-one-out strategy and ROC analysis. In the leave-one-out methodology, a specific input ROI is selected and classified according to the model obtained by training the system with the remaining ROIs in the database. This procedure is repeated until all the ROIs have been used as an input image. The SVM classifier provides a numerical value related to the membership of each class. Thus, varying the threshold of this membership it is possible to generate a ROC curve and obtain the corresponding *Az* value.

In order to perform a more global evaluation we compute the *Az* 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 compare our proposal with existing methods. Notice that previous works only provide results for specific ratios (see Table 1). Hence, analyzing all these ratios will enable the comparison with them.

4.2. Parameters optimization

As explained in Section 3.1, some parameters may be optimized in order to obtain the final LBP descriptor. For instance, the number of regions in which a ROI image is divided or the parameters (P, R) used to obtain the LBP responses. With the aim of choosing the number of divisions, we tested the performance of LBP on four different configurations ($1 \times 1, 3 \times 3, 5 \times 5$ and 7×7 regions) for all the ROI image sizes, all ratios, and using the basic $LBP_{8,1}{}^{u2}$, $LBP_{8,2}{}^{u2}$, $LBP_{16,1}{}^{u2}$, and $LBP_{16,2}{}^{u2}$ operators (see Fig. 5 (a)). The best result was obtained when dividing each ROI image into 5×5 squared regions and computing $LBP_{8,1}{}^{u2}$ operators for each one. Using this configuration we obtained an overall mean *Az* value of 0.882 ± 0.051. Note that the size of each subregion depended on the original ROI image size (i.e. larger ROIs will have larger regions). This was done to preserve the spatial information of the ROIs at different mass sizes. With the aim of improving these results and obtaining a more accurate description per ROI, we studied also the effect of adding a new set of LBP operators for the 3×3 central regions. We computed two new LBP responses varying the radius *R* as a function of the size of the ROI (i.e. *R*size $= 4$ and 6 was used for the smallest ROI image size and *R*size $= 14$ and 21 for the largest one). Using the *R*size parameter we obtained similar LBP operators independently of the subdivisions done per each ROI image size. The global descriptor was then obtained concatenating the 43 histograms of all the regions and LBP operators. For this new configuration, an overall mean *Az* value of 0.911 ± 0.043 was obtained. Observe that better results were obtained when including LBP operators with different radius *R* in the central regions. We then repeated the same experiment varying also the quantization of the angular space *P* in the basic 5×5 LBP computation and in the 3×3 central regions, using $LBP_{8,1}{}^{u2}$, $LBP_{8,2}{}^{u2}$, $LBP_{16,1}{}^{u2}$, and $LBP_{16,2}{}^{u2}$. Similar results were obtained, although, the basic $LBP_{8,1}{}^{u2}$ combined with the $LBP_{8,Rsize}{}^{u2}$ provided the best overall mean results for different ROI image sizes and ratios. This was the final descriptor we used for our experiments since it provided a good trade-off between performance and feature vector length.

4.3. Results varying the ROIs ratio

Fig. 5(b) shows the obtained mean *Az* values for each specific ratio when testing our proposal for all ROI image sizes. We include a quantitative comparison with the recent works of Oliver et al. [34,36], using our database of ROIs. While the first approach is based on using standard PCA, the second introduced a variation on their previous work by using the 2DPCA method [37] instead of the standard PCA technique (we shall refer to these methods as PCA and 2DPCA, respectively). Observe that the performance of our proposal 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 *Az* value of 0.686 ± 0.095 for different ROI image sizes and ratios, using 2DPCA the *Az* value was 0.868 ± 0.087, while using our LBP approach with SVM the *Az* value was 0.906 ± 0.043. Note that we also show the results of LBP when using a Nearest Neighbor (NN) classifier, obtaining an overall mean *Az* value of 0.871 ± 0.036. Observe that LBP with NN provided better performance than 2DPCA for smaller ratios, although the best results were provided by LBP with SVM.

4.4. Results varying the ROI image sizes

Fig. 5(c) shows the obtained mean *Az* values for each specific image ROI size when testing our proposal for all ratios of number of ROIs depicting masses and number of ROIs depicting normal tissue (from ratio 1/1 to ratio 1/6). As in the previous section we include our results obtained by using LBP, and those obtained by using PCA and 2DPCA approaches. Analyzing this figure we are able to observe the behavior of our approach when varying the ROI image sizes. Note that LBP provided better and more constant overall results than PCA and 2DPCA methods.

5. Discussion and conclusions

We have studied a novel method for mammographic mass false positive reduction based on textural features. Our approach divides a ROI image into small regions and computes local texture descriptions using local binary patterns. The combination of these local

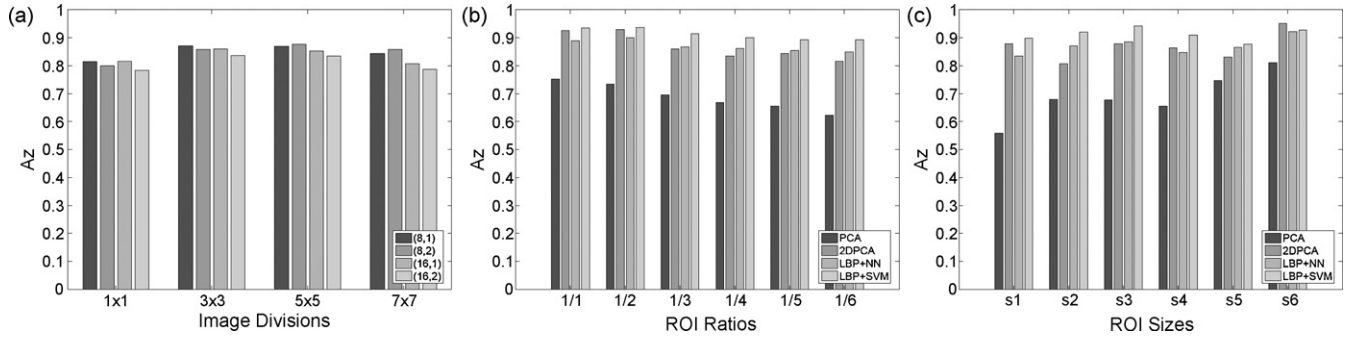

Fig. 5. Experimental results. (a) *Az* values obtained when using LBP and varying the subdivision sizes. Note the 5×5 division obtained the best results. (b) *Az* values obtained with the methods PCA, 2DPCA, LBP + NN and LBP + SVM. Each *Az* value is the mean computed from the results of different ROI image sizes. (c) *Az* values per each ROI image size obtained with the methods PCA, 2DPCA, LBP + NN and LBP + SVM. Each *Az* value is the mean computed from the results of all the ROI ratios (from ratio 1/1 to ratio 1/6).

Table 2
Obtained results per lesion size. *Az* values for ratio 1/1 and 1/3 respectively, detailed per size (in cm^2).

	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	Mean
PCA	0.67	0.70	0.76	0.77	0.79	0.81	0.75
2DPCA	0.90	0.93	0.90	0.90	0.86	0.98	0.91
LBP + SVM	0.88	0.95	0.99	0.92	0.93	0.95	0.94
PCA	0.53	0.70	0.70	0.68	0.72	0.83	0.69
2DPCA	0.81	0.83	0.87	0.84	0.89	0.93	0.86
LBP + SVM	0.91	0.91	0.95	0.92	0.89	0.91	0.91

descriptors in a spatially enhanced histogram which encodes both the local region appearance and the spatial structure of the masses provides our final feature descriptor. Afterwards, these descriptors are used to classify the ROIs between true masses and normal parenchyma.

The results presented in Section 4.4 have shown that LBP features are effective and efficient for false positive reduction at different ROI image sizes, an important aspect for correctly classifying the masses. Table 2 shows in more detail the *Az* values obtained for the specific ratios 1/1 and 1/3 respectively for different lesion sizes. Note that all the approaches (PCA, 2DPCA and LBP + SVM) 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 gray-level contrast with respect to their surrounding tissue than smaller masses, which are usually more subtle, even for an expert.

On the other hand, the results presented in Section 4.3 have also shown that even when using different ratios of number of ROIs with masses and number of ROIs with normal tissue it is possible to obtain reliable results using our LBP approach.

There are two aspects to mention regarding the computational cost of our experiments. Firstly, the inclusion of the leave-one-out strategy has increased the overall computational time for training and testing. In particular, most of the computational time was due to the training of all the sets using SVM. Secondly, we should emphasize that LBP is an efficient descriptor and its computational simplicity makes it possible to combine and apply different operators. For instance, computing our LBP descriptor for a ROI of 400×400 pixels takes 30 ms approximately using the Matlab code provided by Ojala et al. [23].

Finally, in Table 3 we present a qualitative comparison – in terms of *Az* value– with the most representative approaches of the state of the art described in Section 2. Note that our efforts are concentrated on obtaining the same ratio of masses used in their experiments. However, we want to clarify that not all the methods used the same databases and number of ROIs and therefore our aim is only to provide a general trend of the performance of our LBP approach with respect to different strategies. For instance, the works of Chang et al. [40], Tourassi et al. [41] and Varela et al. [38], which used a ratio 1/1, obtained *Az* values of 0.83 ± 0.02, 0.87 ± 0.01, and 0.90 ± 0.02 respectively. Note that for this ratio better performances are clearly obtained using our proposal, being statistically significant at *p*-value <0.0001. A similar behavior is observed for the works which used the ratios 1/2 and 1/3. Observe that the difference between the performance reported in the original PCA work of Oliver et al. (*Az* of 0.83) and the ones showed in our quantitative comparison (*Az* value of 0.69) is due to the use of different ROI image databases and their particular level of difficulty (the database used in [34] had less images and only 4 different ROI image sizes). In summary, the comparison of our proposal with current methods shows the feasibility of LBP descriptors for mass false positive reduction.

Table 3
Az comparison of different methods dealing with mass false positive reduction. We detail the number of ROIs and ratio used in the original works. We also include the results obtained with our LBP + SVM approach. Note that for some works the standard deviation is not available.

	ROIs	Ratio	*Az*
Sahiner et al. [20]	672	1/3	0.90
Qian et al. [21]	800	1/3	0.86
Chang et al. [40]	600	1/1	0.83 ± 0.02
Tourassi et al. [41]	1465	$\cong 1/1$	0.87 ± 0.01
Oliver et al. [34]	588	1/2	0.83
Oliver et al. [36]	1792	1/2	0.83
Tourassi et al. [42]	1820	$\cong 1/1$	0.81 ± 0.02
Varela et al. [38]	120	1/1	0.90 ± 0.02
LBP + SVM	512	1/1	0.94 ± 0.02
LBP + SVM	768	1/2	0.94 ± 0.02
LBP + SVM	1024	1/3	0.91 ± 0.04

Acknowledgements

This work was supported by the Ministerio de Educación y Ciencia of Spain under Grant TIN2007-60553, by the UdG under Grant IdIBGi-UdG, and by the CIRIT and CUR of DIUiE of Generalitat de Catalunya under grant 2008SALUT00029.

References

[1] Eurostat, Health statistics Atlas on mortality in the European Union, Office for Official Publications of the European Union; 2002.
[2] American Cancer Society Breast cancer: facts and figures. ACS; 2003–2004.
[3] Esteve J, Kricker A, Ferlay J, Parkin D. Facts and figures of cancer in the European Community. In: Tech. rep., International Agency for Research on Cancer. 1993.
[4] Buseman S, Mouchawar J, Calonge N, Byers T. Mammography screening matters for young women with breast carcinoma. Cancer 2003;97(2):352–8.
[5] Sickles EA. Breast cancer screening outcomes in women ages 40–49: clinical experience with service screening using modern mammography. J Natl Cancer Inst: Monogr 1997;22:99–104.
[6] De Koning HJ, Fracheboud J, Boer R, Verbeek AL, Collette HJ, Hendriks JHCL. Nation-wide breast cancer screening in the Netherlands: support for breast cancer mortality reduction. National evaluation team for breast cancer screening, Int J Cancer 1995;60(6):777–80.
[7] Hendee WR, Beam C, Hendrick E. Proposition: all mammograms should be double-read. Med Phys 1999;26:115–8.
[8] Thurfjell EL, Lernevall KA, Taube AAS. Benefit of independent double reading in a population-based mammography screening program. Radiology 1994;191:241–4.
[9] Vacek PM, Geller BM, Weaver DL, Foster RS. Increased mammography use and its impact on earlier breast cancer detection in Vermont. Cancer 2002;94(8):2160–8.
[10] Bird RE, Wallace TW, Yankaskas BC. Analysis of cancers missed at screening mammography. Radiology 1992;184:613–7.
[11] Birdwell RL, Ikeda DM, O'Shaughnessy KD, Sickles EA. Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection. Radiology 2001;219:192–202.
[12] Basset LW, Gold RH. Breast Cancer Detection: Mammograms and Other Methods in Breast Imaging. New York: Grune & Stratton; 1987.
[13] Hall FM, Storella JM, Siverstond DZ, Wyshak G. Nonpalpable breast lesions: recommendations for biopsy based on suspicion of carcinoma at mammography. Radiology 1988;167:353–8.
[14] Winsberg F, Elkin M, Macy J, Bordaz V, Weymouth W. Detection of radiographic abnormalities in mammograms by means of optical scanning and computer analysis. Radiology 1967;89(2):211–5.
[15] Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12860 patients in a community breast center. Radiology 2001;220:781–6.
[16] Zwiggelaar R, Astley SM, Boggis CRM, Taylor CJ. Linear structures in mammographic images: detection and classification. IEEE Trans Med Imag 2004;23(9):1077–86.
[17] Khoo LAL, Taylor P, Given-Wilson RM. Computer-aided detection in the United Kingdom national breast screening programme: prospective study. Radiology 2005;237(2):444–9.
[18] Nishikawa RM, Kallergi M. Computer-aided detection, in its present form, is not an effective aid for screening mammography. Med Phys 2006;33:811–4.
[19] Taylor P, Champness J, Given-Wilson R, Johnston K, Potts H. Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography. Health Techn Assess 2005;9(6):1–58.
[20] Sahiner B, Chan HP, Wei D, Petrick N, Helvie MA, Adler DD. Image feature selection by a genetic algorithm: Application to classification of mass and normal breast tissue. Med Phys 1996;23:1671–84.

[21] Qian W, Sun X, Song D, Clark RA. Digital mammography - wavelet transform and Kalman-filtering neural network in mass segmentation and detection. Acad Radiol 2001;8(11):1074–82.
[22] Christoyianni I, Koutras A, Dermatas E, Kokkinakis G. Computer aided of breast cancer in digitized mammograms. Comp Med Imag Grap 2002;26:309–19.
[23] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Machine Intell 2002;24(7):971–87.
[24] Oliver A, Lladó X, Freixenet J, Martí J. False positive reduction in mammographic mass detection using local binary patterns. Int Conf Med Image Comput Comput Assist Interv 2007;4478:286–93.
[25] Zhang J, Tan T. Brief review of invariant texture analysis methods. Pattern Recognit 2002;35(3):735–47.
[26] Ahonen T, Hadid A, Pietikainen M. Face detection with local binary patterns: application to face recognition. IEEE Trans Pattern Anal Machine Intell 2006;28(12):2037–41.
[27] Rangayyan RM, Ayres FJ, Desautels JEL. A review of computer-aided diagnosis of breast cancer: toward the detection of subtle signs. J Frankl Inst 2007;344(3–4):312–48.
[28] Suliga M, Deklerck R, Nyssen E. Markov random field-based clustering applied to the segmentation of masses in digital mammograms. Comp Med Imag Grap 2008;32(6):502–12.
[29] Rojas A, Nandi A. Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection. Comp Med Imag Grap 2008;32(4):304–15.
[30] Wu Y, Wei J, Hadjiiski LM, Sahiner B, Zhou C, Ge J. Bilateral analysis based false positive reduction for computer-aided mass detection. Med Phys 2007;34(8):3334–44.
[31] van Engeland S, Karssemeijer N. Combining two mammographic projections in a computer aided mass detection method. Med Phys 2007;34:898–905.
[32] Metz CE. Evaluation of digital mammography by ROC analysis. Int Work Dig Mammogr 1996:61–8.
[33] Angelini E, Campanini R, Iampieri E, Lanconelli N, Masotti M, Roffilli M. Testing the performances of different image representations for mass classification in digital mammograms. Int J Mod Phys C 2006;17(1):113–31.
[34] Oliver A, Martí J, Martí R, Bosch A, Freixenet J. A new approach to the classification of mammographic masses and normal breast tissue. IAPR Int Conf Pattern Recognit 2006;4:707–10.
[35] Turk MA, Pentland AP. Eigenfaces for recognition. J Cognitive Neurosci 1991;3(1):71–86.
[36] Oliver A, Lladó X, Martí J, Martí R, Freixenet J. False positive reduction in breast mass detection using two-dimensional PCA. Lect Not Comp Sci 2007;4478:154–61.
[37] Yang J, Zhang D, Frangi AF, Yang J. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans Pattern Anal Machine Intell 2004;26(1):131–7.
[38] Varela C, Tahoces PG, Méndez AJ, Souto M, Vidal JJ. Computerized detection of breast masses in digitized mammograms. Comput Biol Med 2007;37(2):214–26.
[39] Tourassi GD, Eltonsy NH, Graham JH, Floyd CE, Elmaghraby AS. Feature and knowledge based analysis for reduction of false positives in the computerized detection of masses in screening mammography. IEEE Conf Eng Med Biol Soc 2005:6524–7.
[40] Chang YH, Hardesty LA, Hakim CM, Chang TS, Zheng B, Good WF. Knowledge-based computer-aided detection of masses on digitized mammograms: a preliminary assessment. Med Phys 2001;28(4):455–61.
[41] Tourassi GD, Vargas-Vorecek R, Catarious DM, Floyd CE. Computer-assisted detection of mammographic masses: A template matching scheme based on mutual information. Med Phys 2003;30(8):2123–30.
[42] Tourassi GD, Harrawood B, Singh S, Lo JY, Floyd CE. Evaluation of information-theoretic similarity measures for content-based retrieval and detection of masses in mammograms. IEEE Trans Med Imag 2007;34(1):140–50.
[43] Ojala T, Pietikainen M, Harwood D. A comparative-study of texture measures with classification based on feature distributions. Pattern Recognit 1996;29(1):51–9.
[44] Vapnik V. Statistical Learning Theory. New York: John Wiley & Sons; 1998.
[45] Heath M, Bowyer K, Kopans D, Moore R, Kegelmeyer PJ. The digital database for screening mammography. Int Work Dig Mammogr 2000:212–8.

Xavier Lladó received the BSc degree in Computer Science from the University of Girona in 1999 before joining the Computer Vision and Robotics Group. He also received the PhD in Computer Engineering from the University of Girona. From 2004 until 2006, he was working as a Post-doctoral Research Assistant in the Department of Computer Science at the Queen Mary, University of London. Currently, he is a lecturer in the Computer Vision and Robotics Group at the University of Girona. His research interests are in the field of image processing and computer vision, focusing on medical image analysis, object recognition and non-rigid structure from motion.

Arnau Oliver received the BSc degree in physics from the Universitat Autònoma de Barcelona (UAB), Barcelona, Spain, in 1999, and the PhD degree in information technology from the University of Girona, Girona, Spain, in 2007. Since 2002, he has been with the Computer Vision and Robotics Group, Department of Electronics, Informatics, and Applications, University of Girona, where he is currently an Assistant Lecturer. His current research interests include pattern recognition and the development of automatic tools for breast cancer detection.

Jordi Freixenet received the MSc degree in computer science from the Polytechnical University of Catalonia, Barcelona, Spain, in 1994, and the PhD degree in computer engineering from the University of Girona, Girona, Spain, in 2000. He is with the Computer Vision and Robotics Group, University of Girona, Girona, Spain,where he is currently a Lecturer. His current research interests include the field of image processing and computer vision, focusing on medical image analysis, object recognition, image classification, and segmentation. He is also engaged in the improvement in detection and diagnosis of breast cancer.

Robert Martí received the MSc degree in computer science from the University of Girona, Girona, Spain, in 1999, and the PhD degree from the University of East Anglia, Norwich, U.K., in 2002, for his work on image registration applied to multimodal mammography. He is currently a Lecturer at the University of Girona. His current research interests include medical image analysis, image registration, pattern recognition, and feature extraction techniques specially focusing on mammographic and prostatic data.

Joan Martí received the BSc degree in Physics from the Universitat Autònoma de Barcelona (UAB) in 1986. He received the PhD degree in Computer Vision from the Polytechnical University of Catalonia (UPC) in 1998. He is currently a senior lecturer in Computer Vision at the University of Girona (UdG), in the Institute of Informatics and Applications. His research interests include knowledge-based systems, segmentation of biomedical images, and colour and texture- based segmentation for object recognition purposes.