IOP PUBLISHING PHYSICS IN MEDICINE AND BIOLOGY

Phys. Med. Biol. **57** (2012) 7029–7052 doi:10.1088/0031-9155/57/21/7029

Multiresolution local binary pattern texture analysis combined with variable selection for application to false-positive reduction in computer-aided detection of breast masses on mammograms

Jae Young Choi[1] **and Yong Man Ro**[2]

Image and Video Systems Lab, Department of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST) 335, Gwahak-ro, Yuseon-gu, Daejeon, 305-701, Korea

E-mail: Jae.YoungChoi@uphs.upenn.edu and ymro@ee.kaist.ac.kr

Received 18 April 2012, in final form 11 August 2012
Published 10 October 2012
Online at stacks.iop.org/PMB/57/7029

Abstract

In this paper, a new and novel approach is designed for extracting local binary pattern (LBP) texture features from the computer-identified mass regions, aiming to reduce false-positive (FP) detection in a computerized mass detection framework. The proposed texture feature, the so-called multiresolution LBP feature, is well able to characterize the regional texture patterns of core and margin regions of a mass, as well as to preserve the spatial structure information of the mass. In addition, to maximize a complementary effect on improving classification accuracy, multiresolution texture analysis has been incorporated into the extraction of LBP features. Further, SVM-RFE-based variable selection strategy is applied for selecting an optimal subset of variables of multiresolution LBP texture features to maximize the separation between breast masses and normal tissues. Extensive and comparative experiments have been conducted to evaluate the proposed method on two public benchmark mammogram databases (DBs). Experimental results show that the proposed multiresolution LBP features (extracted from automatically segmented mass boundaries) outperform other state-of-the-art texture features developed for FP reduction. Our results also indicate that combining our multiresolution LBP features with variable selection strategy is an effective solution for reducing FP signals in computer-aided detection (CAD) of mammographic masses.

(Some figures may appear in colour only in the online journal)

1. Introduction

Screening mammography is the most common cost-effective method to detect early breast cancer. Breast cancers are generally presented as microcalcifications, masses or asymmetry in

[1] Present Address: Department of Radiology, University of Pennsylvania School of Medicine, PA, USA.
[2] Author to whom any correspondence should be addressed.

mammography (Freer and Ulissey 2001). In general, breast masses are much more difficult to detect and classify than microcalcifications because masses can be simulated or obscured by normal glandular tissue in dense breast (Chan *et al* 1995). Computer-aided detection (CAD) of breast cancers in mammography is an active area of research (Suri and Rangayyan 2006). CAD is designed to provide visual prompts (mainly masses and microcalcification clusters) to the interpreting radiologists in specific areas on mammograms. In particular, masses are important indicators of malignancy on mammograms. However, as reported in a large body of studies (Tang *et al* 2009, Zheng *et al* 2001, Birdwell *et al* 2001, Suri and Rangayyan 2006, Eltonsy *et al* 2007, Lladó *et al* 2009), the sensitivity of current CAD systems for masses is high, but the specificity is not due to high false-positive (FP) detection rates. In CAD systems, a FP is usually a region being normal tissue, but interpreted by the computer algorithms as a suspicious one (Chan *et al* 1995, Wei *et al* 1997). High FP rates increase the radiologists' recall rate in clinical CAD environment, while it induces unnecessary biopsies (Freer and Ulissey 2001, Tang *et al* 2009, Zheng *et al* 2001). Hence, in order to reduce costs and patient discomfort, it is of paramount importance for currently developed CAD systems to reduce the number of FP signals without missing any malignant masses (Wei *et al* 1997, Sahiner *et al* 1996, 1998, 2001).

In recent years, many researchers have investigated the use of computer-extracted image features for the detection and classification of breast masses. In particular, considerable research efforts (Wei *et al* 1997, Sahiner *et al* 1996, 1998, 2001) have been devoted to the development of *texture* features for the classification of lesions. The results reported in these works indicate that texture features can play a crucial role in encoding the characteristics of masses (such as the margin and density of mammographic abnormalities (Huo *et al* 1998)), and they can be used to considerably improve classification accuracy for differentiating masses and breast normal tissues. Until so far, the most commonly used texture features for mammographic mass detection (classification) can be roughly divided into four classes (Cheng *et al* 2005): (1) spatial gray-level dependence (SGLD)-based texture features (Wei *et al* 1997), (2) gray-level difference statistics (GLDS)-based texture features (Sahiner *et al* 1996), (3) run length statistics (RLS)-based texture features (Sahiner *et al* 2001) and (4) rubber-band straightening transform (RBST)-based texture features (Sahiner *et al* 1998). In Wei *et al* (1997), texture features were extracted from a SGLD matrix, which characterizes the average spatial relationship of pairs of gray levels in the region of interest (ROI). Experimental results showed that using texture features derived from SGLD matrices is effective for the classification of masses and normal breast tissue on mammograms. In Sahiner *et al* (1996), the authors investigated the use of GLDS-based local texture features. They demonstrated the usefulness of using GLDS-based texture features for the classification of masses and normal tissue on mammograms. The authors in Sahiner *et al* (2001) proposed the RLS-based texture features for classifying a mass as malignant or benign. In Sahiner *et al* (2001), RLS texture features were extracted by computing a gray-level run (a set of consecutive, collinear pixels with the same gray-level value) and a run length (the number of pixels in a run). The results indicate that RLS-based features are effective for the characterization of breast masses on mammograms. In Sahiner *et al* (1998), the authors introduced an RBST-based texture features to analyze the margin characteristics of a mass. The RBST maps the pixel values in a band of pixels surrounding the mass onto the Cartesian plane (Sahiner *et al* 1998). This work reports that features extracted from RBST images are significant in the differentiation of malignant and benign masses.

While many texture feature extraction methods have been explored in the CAD of breast masses in past years, developing effective and robust texture features —which allow the highest classification accuracy for an FP reduction purpose—still remains challenging. In this paper,

we present a new texture feature extraction approach based on local binary patterns (LBP), which aims to considerably reduce the FP rate in a practical CAD system. To our knowledge, the methods explicitly and systematically devoted to the development of LBP-based texture features for the purpose of FP reduction in CAD systems are quite a few, although the idea of using LBP texture information for mammographic mass classification has been previously introduced by the authors in Lladó *et al* (2009) and Oliver *et al* (2007). It should be noted, however, that our proposed approach to extracting LBP texture features from mammographic masses has been devised in a novel and quite different way. In particular, differing from the previous approach proposed in Lladó *et al* (2009) and Oliver *et al* (2007), key technical contributions of this work can be summarized in the following aspects.

- In a previous approach, following the same original idea of the LBP operator proposed by Ojala *et al* (1996, 2002), the ROI image is divided into several *rectangular* local regions, and LBP texture patterns, each extracted from a corresponding local region, are concatenated to produce the final LBP-based texture descriptor. However, in the proposed approach, two individual LBP patterns are independently extracted from the so-called *core* and *margin* regions of pixels (a band of pixels surrounding the segmented mass object in the ROI) of a given mass region, respectively. The proposed approach allows for characterizing the inherent LBP-based texture features of a mass region in a more effective and robust way, as well as at the same time for well preserving the spatial structure information of the mass.
- The previous approach only makes use of a *single*-resolution-based LBP operation during the extraction of LBP texture features. On the other hand, *multiresolution* LBP texture analysis has been incorporated into the proposed LBP feature extraction solution—leading to the proposed multiresolution LBP features. As demonstrated in section 5.4, using multiresolution LBP texture analysis is beneficial for facilitating a *complementary* effect on improving classification accuracy.
- The previous approach simply applies the LBP features to a classifier model (for mass classification purpose) without any sophisticated feature (or variable) selection strategy. However, it is important to note that the LBP features generally result in the high dimensionality and the redundant information (Ojala *et al* 2002). This might cause the degradation in classification performance (Bishop 2006). To tackle the aforementioned limitation, we introduce an effective variable selection mechanism (designed for classification) to maximize classification accuracy obtained using the proposed multiresolution LBP features. Toward this goal, SVM-RFE-based *variable selection* strategy (Guyon and Elisseeff 2003, Guyon *et al* 2002) is adopted to select an optimal subset of variables of multiresolution LBP features to maximize the separation between masses and normal tissues. To the best of our knowledge, combining the proposed multiresolution LBP features with SVM-RFE-based variable selection is a new approach in the task of differentiating ROIs on mammograms as either mass or normal tissue.

Comparative and extensive experiments have been conducted to investigate the effectiveness of proposed multiresolution LBP features in conjunction with SVM-RFE-based variable selection. Two benchmark mammogram datasets Digital Database for Screening Mammography (DDSM) (Heath *et al* 2000) and Mammographic Image Analysis Society (MAIS) (Suckling *et al* 1994) databases (DBs) are used. Our experiments show that the proposed solution to extract multiresolution LBP features is able to achieve a high level of classification accuracy—in the presence of the errors caused by the inaccurate location of the

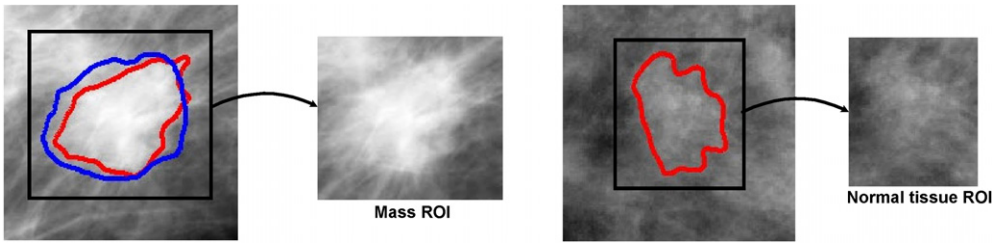


Figure 1. (a) Mass ROI. (b) Normal tissue ROI. See the text for explanation.

boundaries of masses automatically segmented using computer algorithms (see figure 1)—that meets the requirement of practical clinical applications. In addition, we clearly validate the feasibility of proposed multiresolution texture features by making comparisons with other state-of-the-art texture features developed for mammographic mass classification. In particular, it has been found that the proposed method considerably outperforms the best LBP feature extraction approach previously proposed in Lladó *et al* (2009) and Oliver *et al* (2007), in terms of classification performance for test data. Moreover, our results demonstrate that the performance (in terms of reducing FP detections) can be further improved by combining our multiresolution LBP features with variable selection strategy.

The remainder of this paper is organized as follows. In section 2, ROI detection and the segmentation method used in this study is briefly explained. Section 3 details the proposed mutiresolution LBP feature extraction approach. In section 4, we discuss the SVM-RFE algorithm for variable selection. In section 5, we present the experimental results. The conclusion is drawn in section 6.

2. ROI detection and segmentation

Note that a typical CAD system largely consists of three parts (Petrick *et al* 1999, Mudigonda *et al* 2001, Dominguez and Nandi 2008): (1) automated detection and segmentation of mammographic masses from a given mammogram, which yields ROIs; (2) automated feature extraction for generated ROIs; (3) automated classification of ROIs into true and false positives for an FP reduction purpose. Before explaining the extraction of multiresolution LBP features for classifying ROIs, we briefly describe the ROI detection and segmentation method employed in this study.

In our work, as recommend in Chan *et al* (1995), Wei *et al* (1997), Sahiner *et al* (1996) and Mudigonda *et al* (2001) to perform a more realistic assessment of the classification process, the ROI regions were automatically detected and segmented from each mammogram by using a fully automated segmentation method. For this purpose, one popular approach to use a multi-level thresholding algorithm (Mudigonda *et al* 2001, Dominguez and Nandi 2008, Hong and Sohn 2010) was adopted for segmenting masses. We chose this segmentation approach because it has been well documented in previous publications in that it can provide 'successful' segmentation results. Also note that the mass detection and segmentation method used in our research adopts an 'unsupervised' scheme (Tang *et al* 2009, Oliver *et al* 2010). Unsupervised approaches do not generally require the training stage in order to optimize their performances (Oliver *et al* 2010). Instead, the optimal performance of the used algorithm entirely depends on tuning parameters, by adjusting a threshold value (not using training images) to find the

proper balance between sensitivity and specificity (Mudigonda *et al* 2001, Dominguez and Nandi 2008, Hong and Sohn 2010). This fact enables us to evaluate FP reduction performance without any potential bias (especially due to the overlap of training and testing images) induced by the optimization (or tuning) process of ROI detection and segmentation used in this study.

The segmentation algorithm used consists of the following three sequential steps: (1) the construction of the isocontour map; (2) the formation of the inclusion tree and (3) the computation of the minimum nesting depth (MND). The aforementioned individual steps are briefly explained in the following. The isocontour map is obtained from a set of isocontours at multiple different threshold values over the intensity range of the mammogram. From the isocontour map representation of the image, suspicious regions, which form a dense quasi-concentric pattern, are detected by considering the nesting patterns of isocontours, as well as gradient between isocontours, for the purpose of characterizing mass objects. Next, the inclusion tree is generated from the isocontour map to represent the enclosure relationship between contours. The inclusion tree is used to determine ROIs from the suspicious regions by analyzing the inclusion tree. Note that the probability (confidence) that a segmented ROI is truly mass is measured by the MND, which is defined as the number of isocontours that are embedded within a contour in question. The MND is used as a threshold value to eliminate less suspicious regions. The implementation details of each step have been described in the literature (Mudigonda *et al* 2001, Dominguez and Nandi 2008, Hong and Sohn 2010).

Figure 1 shows an example of segmented ROIs generated by the segmentation algorithm used. Note that in figure 1, the red line is successfully segmented mass contour (or boundary) identified by the segmentation algorithm, and the blue line is the mass outline (as ground truth) marked by experienced radiologists. Note that the size of the ROI is chosen such that the computer-marked mass contour and a band of pixels surrounding a segmented object are included in the ROI (a detailed description of determining the width of band is given in section 3.2). The mass or normal tissue ROIs are then used as input for the extraction of multiresolution LBP features to be discussed in the following section.

3. Proposed texture feature extraction

The proposed multiresolution LBP features (extracted from given ROIs) are computed based on the LBP operator introduced by Ojala *et al* (2002). In this section, we will briefly review the LBP operator and describe in detail the proposed texture feature extraction method.

3.1. Review of the LBP operator

The basic LBP operator labels the pixels of an image by thresholding the neighborhood of each pixel with center pixel value and considering the result of thresholding with center pixel as a binary pattern number (Ojala *et al* 1996). The limitation of the basic LBP operator is that its small 3×3 neighborhood cannot capture dominant features with large-scale structures. Thus, the LBP operator was extended to use circular neighborhood suggested in Ojala *et al* (2002). In this study, the LBP operator with a circularly symmetric neighborhood of P members on a circle radius of R is employed. Also, note that in the computation of LBP for a given image, the uniform LBP operator (Ojala *et al* 2002) is adopted because it is reasonable to assume that a typical mass ROI image is likely to contain only a smaller number of LBP values (called uniform patterns (Ojala *et al* 2002), which contain at most two bitwise transitions from 0 to 1 or vice versa). The rationale behind this assumption is that the gray-level changes at the masses are usually constant and at least gradual (rather than abrupt), as reported in a large

number of literatures (Chan *et al* 1995, Wei *et al* 1997, Sahiner *et al* 1996, Mudigonda *et al* 2001, Dominguez and Nandi 2008, Hong and Sohn 2010).

Let us denote z_c to represent the center pixel position of an image and $z_n (n = 0, \ldots, P-1)$ are P equally spaced pixels (or sampling points) on a circle of radius R ($R > 0$) that form a circular neighborhood of the center pixel z_c. The uniform LBP operator for the center pixel position z_c of an image is then defined as follows (Ojala *et al* 2002):

$$\mathrm{LBP}_{P,R}(z_c) = \begin{cases} \sum_{n=0}^{P-1} \delta(r_n - r_c) 2^n & \text{if } H \leqslant 2 \\ P(P-1)+2 & \text{otherwise,} \end{cases} \tag{1}$$

where

$$H = |\delta(r_{P-1} - r_c) - \delta(r_0 - r_c)| + \sum_{n=1}^{P-1} |\delta(r_n - r_c) - \delta(r_{n-1} - r_c)|, \tag{2}$$

and $\delta(x) = \begin{cases} 1 & x \geqslant 0 \\ 0 & x < 0 \end{cases}$, $r_n (n = 0, \ldots, P-1)$ denote the pixel values of an image at z_n and r_c denotes the pixel value at z_c of a circular neighborhood.

3.2. Computing LBP histograms of core and margin regions

In order to characterize the texture pattern information of the masses *at a regional level*, as well as to capture the *spatial structure* (global geometry) of the masses, we construct a band of pixels along the boundary of the segmented object in the ROI. Note that the regions that are segmented by the method described in section 2 vary greatly in size and shape. To cope with this problem, we adaptively compute the width of the band based on the segmented region's size and shape. Specifically, the diameter of the bounding circle enclosing a given segmented region is first obtained by calculating the maximum distance between any two points on its boundary. Next, the area of the segmented region S_r and the bounding circle S_c enclosing the region are calculated. The width of the band is then computed as $W_b = R_c(S_r/S_c)$, where R_c is the radius of the bounding circle. Note that the ratio S_r/S_c is a simple indicator of representing the narrowness and complexity of the segmented region. Finally, the band of pixels can be formed by applying morphological dilation operation (Gonzalez and Woods 2008) on the boundary of the segmented object. Note that for the dilation operation, the diameter of the circular morphological operator is set to the width of the band (W_b). The bands of pixels extracted across the boundaries of the segmented regions are illustrated in figure 2(a).

Based on the band of pixels, a segmented object is divided into two individual subregions (as shown in figures 2(b) and (c)), as follows: the core and margin regions each denoted by $\mathbf{R}_{\mathrm{core}}$ and $\mathbf{R}_{\mathrm{margin}}$, respectively. The rationale behind such region partition is that within a given ROI, there may be several subregions showing different texture statistics, for instance, the region inside the mass (i.e. core region), the transition region between the mass and surrounding tissue (i.e. margin region) and surrounding tissue. Based on (1), let $\mathrm{LBP}^{\mathrm{core}}_{P,R}(z_c)$ be the LBP label at a given pixel location $z_c = (x, y)$ of the core region $\mathbf{R}_{\mathrm{core}}$. The LBP histogram $\mathbf{h}_{\mathrm{core}}$ corresponding to a core region is then extracted as follows:

$$\mathbf{h}_{\mathrm{core}} = [h_{\mathrm{core}}(1), \ldots, h_{\mathrm{core}}(P(P-1)+3)]^{\mathrm{T}}, \tag{3}$$

where

$$h_{\mathrm{core}}(k) = \sum_{z_c=(x,y)\ \mathrm{in}\ \mathbf{R}_{\mathrm{core}}} T(\mathrm{LBP}^{\mathrm{core}}_{P,R}(z_c) = l_k), \text{ for } 1 \leqslant k \leqslant P(P-1)+3, \tag{4}$$

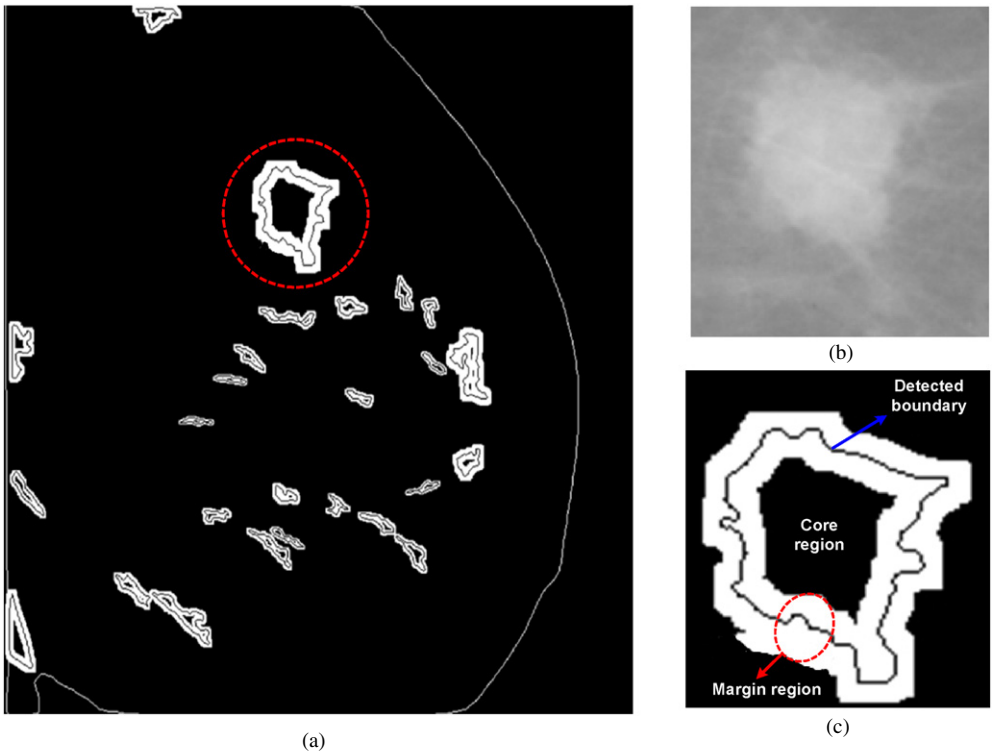


Figure 2. (a) Band of pixels (white) constructed adaptively across the boundaries (black) of the segmented regions detected in the mammogram from MIAS DB. Note that the region enclosed by a dot-line circle corresponds to true mass region, while the others are FP regions. (b) Corresponding mass ROI image of a true mass region shown in (a). (c) Core and margin regions extracted from the mass ROI shown in (b).

and $T(\cdot)$ is an indicator function that returns 1 when the Boolean logic of its input augments is true and returns 0 otherwise, T represents the transpose operator of the matrix and l_k denotes LBP label of the kth bin and, therefore, $h_{\text{core}}(k)$ denotes the number of pixels with the LBP value l_k within the core region $\mathbf{R}_{\text{core}}$. Note that in (3), the LBP histogram consists of $P(P-1)+3$ bins: $P(P-1)$ bins for the patterns with two transitions, two bins for the patterns with zero transitions and one bin for all non-uniform patterns. Using (3) and (4), we can also readily compute the LBP histogram (denoted by $\mathbf{h}_{\text{margin}}$) corresponding to the margin region $\mathbf{R}_{\text{margin}}$.

Figure 3 shows the two histograms of core and margin regions, respectively. Note that each histogram shown in figure 3 has been averaged over a total of 246 histograms —each computed using one of the 246 mass ROIs —identified from 303 mammograms collected from DDSM (Cheng *et al* 2005) DB. Further, each histogram has been normalized by dividing each of its components by the total number of LBP labels. From figure 3, we can see that the histogram of LBP labels for core regions can be concentrated to a small range of bins, while histograms of LBP labels for margin regions should be more spread out. This observation demonstrates that LBP histograms computed from the core and margin regions of masses well represent typical characteristics of masses. The change of pixel values of mass core region is nearly uniform, at least gradual change. Hence, the mass core region has a uniform circular structure that contains a small number of spatial transitions. This observation is consistent

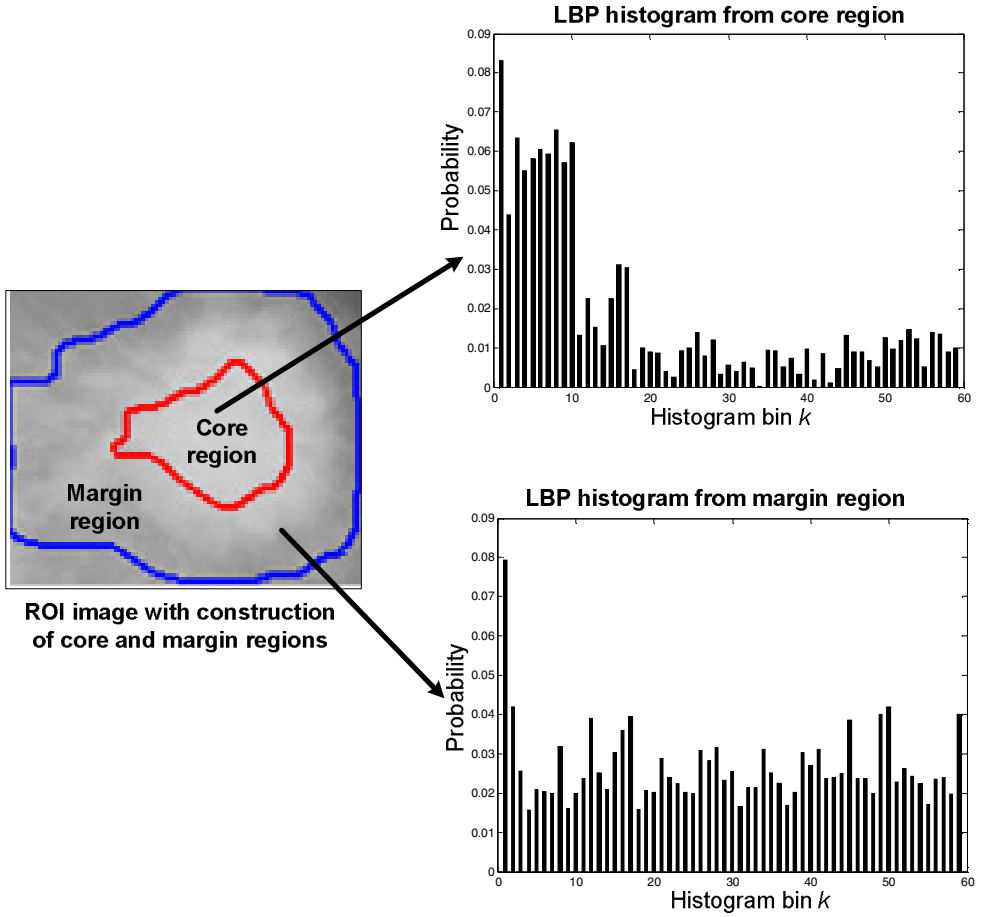


Figure 3. Core and margin regions and their corresponding histograms obtained using uniform LBP operator with parameters $P = 8$ and $R = 2$.

with a previous one that reported that the frequency components of masses are usually in the low-frequency region (Wei *et al* 1995). On the other hand, the change of pixel values of mass margin region may be more abrupt, especially in the margin between mass and normal tissues, so that the LBP labels should often be quite different.

Note that as demonstrated in figure 3, $\mathbf{h}_{\text{core}}$ and $\mathbf{h}_{\text{margin}}$ can provide regional LBP histogram information for core and margin regions, respectively. In order to capture the additional information about the *spatial structure* (or *configuration*) of core and margin regions of masses, the two corresponding LBP histograms $\mathbf{h}_{\text{core}}$ and $\mathbf{h}_{\text{margin}}$ are combined by concatenating them into a longer histogram vector as follows:

$$\mathbf{h}_{\text{combined}} = [(\mathbf{h}_{\text{core}})^{\text{T}}(\mathbf{h}_{\text{margin}})^{\text{T}}]^{\text{T}}. \tag{5}$$

3.3. Generation of multiresolution LBP features

Besides spatial structure information (of the mass) encoded in $\mathbf{h}_{\text{combined}}$, we exploit additional discriminant information through using multiresolution analysis. Note that by altering

parameters P and R, we can realize LBP operation for any quantization of the angular space and for any spatial resolution. Hence, to perform multiresolution texture analysis, we define a set $\{(P_k, R_k)\}_{k=1}^{M}$ consisting of M different pairs (P, R) in the LBP operator defined in (1). Multiresolution analysis can be accomplished by combing the responses of M multiple LBP operators realized with different configurations of (P_k, R_k). Specifically, by using $\{(P_k, R_k)\}_{k=1}^{M}$ and (5), we can compute M individual combined LBP histograms, denoted by $\mathbf{h}_{\text{combined}}^{(m)}(m = 1, \ldots, M)$, for the segmented region under consideration.

In order to extract the proposed *multiresolution LBP feature*, individual $\mathbf{h}_{\text{combined}}^{(m)}$ are fused by concatenating them into an augmented vector:

$$\mathbf{x}_{\text{multi\_LBP}} = \left[\left(\mathbf{h}_{\text{combined}}^{(1)}\right)^{\text{T}} \left(\mathbf{h}_{\text{combined}}^{(2)}\right)^{\text{T}} \cdots \left(\mathbf{h}_{\text{combined}}^{(M)}\right)^{\text{T}}\right]^{\text{T}}. \tag{6}$$

This expression is based on information fusion theory (Jain *et al* 2005, Mangai *et al* 2010) for pattern classification. Techniques for fusing multiple evidences (i.e. multiple classification results or multiple features) can be generally classified into two classes: fusion at the 'feature-level' and fusion at the 'decision-level'. In the proposed multiresolution LBP features, we make the use of feature-level fusion strategy for the following reason: 'feature-level fusion' can generally achieve better classification result compared to performance obtained using 'decision-level' fusion; this is mainly due to the fact that feature-level fusion can provide a complementary effect by eliminating noisy information arising from the components (to be fused) to a certain degree (Jain *et al* 2005, Mangai *et al* 2010). From this perspective, combining multiple $\mathbf{h}_{\text{combined}}^{(m)}$ based on feature-level fusion strategy is expected to produce more robust features against mass appearance variations. It should also be noted that in (6), the underlying reason for the adoption of the multirsolution framework is that we are able to extract additional and complementary discrimination information (by exploiting different spatial relationships) than what a single-resolution LBP operator does. In addition, it is important to note that our multiresolution LBP features are gray-scale and rotation invariant as the LBP operation defined in (1) is performed based on the use of signed difference and circular neighborhood, as discussed in section 3.1. This is a particularly useful property in extracting the texture features of masses, since the location and orientation of mass in the ROI can be arbitrary.

The segmented mass regions are encoded via our multiresolution LBP features on three different levels of discriminating information: (1) the histograms each computed from core and margin regions, respectively, encode the associated regional LBP distribution information, (2) these two individual histograms are combined to produce spatial structure (global geometry) information of masses and (3) the combined histograms are concatenated to exploit additional discriminating information provided by using multiresolution analysis. Indeed, the proposed mutiresolution LBP features condense all of the aforementioned discriminant information in the form of hierarchically encoded structure as shown in figure 4. The benefit of the hierarchical encoding approach for characterizing masses is that the proposed multiresolution LBP features are able to handle a wide variability in the appearance (e.g., shapes and sizes) of masses found in clinical mammograms for a classification purpose. Further, the hierarchical encoding structure of our multiresolution LBP features allows for realizing a robust mass classification solution against the imprecise location of the boundaries of the masses posed by computer segmentation algorithms. The rationale behind these advantages is that although a loss in discriminating information (e.g., regional LBP distribution of masses) available at a lower level of a hierarchical encoding structure may occur, the discriminating patterns (e.g., spatial structure of masses) at a higher level are able to be reserved to some extent. This is mainly because that, as demonstrated in our experiment, discriminating patterns available at all levels of hierarchical encoding structure potentially offer different complementary information about the complex patterns of masses and, therefore, they are likely to be mutually compensational

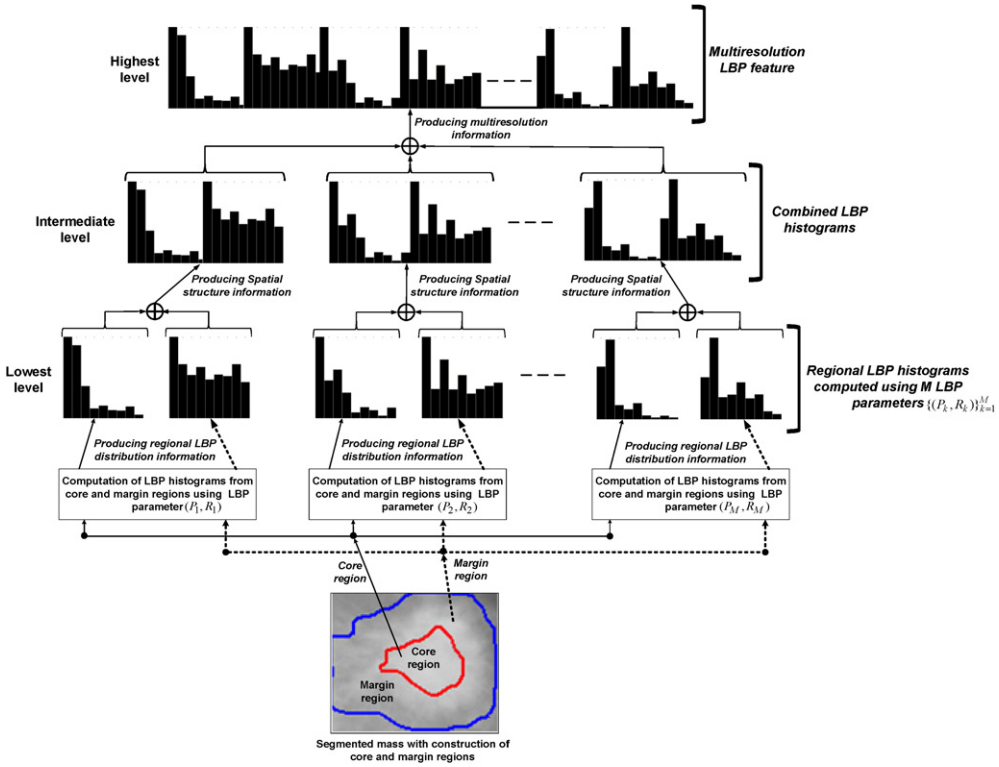


Figure 4. Illustration of the hierarchical encoding structure of masses via the proposed multiresolution LBP features. Note that the region enclosed by red line represents the core region, while the region located between red and blue lines represents margin region. Also note that in order to illustrate multiresolution analysis, LBP histograms from core and margin regions are computed using individual LBP operators each associated with a particular (P, R).

for mammographic mass classification, as well as for improving classification accuracy (for more details, please refer to the experimental assessment in section 5.4).

4. Variable selection of multiresolution LBP features

For classification, one straightforward way is to apply multiresolution LBP features defined in (6) to a classifier (such as support vector machines (SVMs)). However, it should be noted that the multiresolution LBP feature may result in the high dimensionality and the redundant information. This might cause the degradation in classification performance (Bishop 2006). In addition, the computational cost of inputting high-dimensional features into a classifier is generally prohibitive (Bishop 2006). Hence, it is necessary to identify the *proper set of variables* of the multiresolution LBP feature and filter the undesired ones which confuse the classifier and produce worse performance. For this purpose, the multiresolution LBP features are applied to *variable selection* (Guyon and Elisseeff 2003) analysis to reduce dimensionality, ultimately aiming to maximize classification accuracy. In this paper, an SVM-RFE (recursive feature elimination) (Guyon *et al* 2002) is adopted as a variable selection framework because of its lower computational complexity, compared to randomized or exponential algorithms,

and its optimality in the subset selection algorithm (Guyon *et al* 2002). Since the ranking criterion of the SVM-RFE is derived from the formulation of SVMs, we briefly review the SVMs classification method before describing the details on the SVM-RFE-based variable selection.

SVMs have been the most widely used approach for the classification problem. It constructs an optimal hyperplane decision function in the so-called feature space that is mapped via a mapping function Φ from the original input space. Let us use $\mathbf{x}_i$ to denote the ith example vector in the original input space and $\mathbf{z}_i$ to denote the corresponding vector in the feature space, $\mathbf{z}_i = \Phi(\mathbf{x}_i)$. For a typical classification problem with N training samples $\{\mathbf{x}_i, y_i\}_{i=1}^N$, where y_i denotes the class label of $\mathbf{x}_i$ and $y_i \in \{+1, -1\}$, finding the discriminant function $f(\mathbf{x}) = \mathbf{w} \cdot \Phi(\mathbf{x}) + b$ is formulated using the following optimization problem (Vapnik 1998):

$$\begin{aligned} &\min_{w,b,\xi_i} \frac{1}{2}||\mathbf{w}||^2 + C\sum\nolimits_{i=1}^{N} \xi_i \\ &\text{subject to } l_i(\mathbf{w} \cdot \mathbf{z}_i + b) \geqslant 1 - \xi_i, \quad \xi_i \geqslant 0, \end{aligned} \tag{7}$$

where $C > 0$ is predefined higher level parameter, besides the kernel function parameter. The solution (weight vector $\mathbf{w}$) of this optimization problem is of the form

$$\mathbf{w} = \sum\nolimits_{i=1}^{N} \alpha_i y_i \mathbf{z}_i, \tag{8}$$

where α_i is the solution of the following quadratic optimization problem (Vapnik 1998):

$$\begin{aligned} &\min_{\alpha_i} \frac{1}{2}\sum\nolimits_{i=1}^{N}\sum\nolimits_{j=1}^{N} \alpha_i\alpha_j y_i y_j(\mathbf{z}_i \cdot \mathbf{z}_j) - \sum\nolimits_{i=1}^{N} \alpha_i \\ &\text{subject to } 0 \leqslant \alpha_i \leqslant C, \quad \sum\nolimits_{i=1}^{N} \alpha_i y_i = 0. \end{aligned} \tag{9}$$

We now describe the SVM-RFE algorithm for the variable ranking and selection of our multiresolution LBP features. SVM-RFE was originally proposed to perform gene selection for cancer classification (Guyon *et al* 2002). In general, variable selection involves ranking the variables based on their importance (Guyon and Elisseeff 2003). In the SVM-RFE, nested subsets of feature variables are selected in a sequential backward elimination manner, which starts with all the variables and removes one variable. In this way, all the feature variables have been ranked. At each step, the coefficients of the weight vector $\mathbf{w}$ defined in (8) of an SVM are used to compute the variable ranking score. The variable (i.e. the ith variable) with the smallest ranking score $c_i = (w_i)^2$ is eliminated, where w_i represents the corresponding component in the weight vector $\mathbf{w}$. Using $(w_i)^2$ as ranking score corresponds to removing the variable whose removal changes the objective function least. Here, the objective function is defined as $J = (1/2)||\mathbf{w}||^2$ in the SVM-RFE.

To explain the usefulness of choosing this objective function, we approximate the change in the objective function caused by removing the ith feature by expanding the objective function in the Taylor series to second order:

$$\Delta J(i) = \frac{\partial J}{\partial w_i}\Delta w_i + \frac{\partial^2 J}{\partial w_i^2}(\Delta w_i)^2. \tag{10}$$

At the optimum of J, the first-order term can be neglected and with $J = (1/2)||\mathbf{w}||^2$, the equation shown in (10) becomes

$$\Delta J(i) = (\Delta w_i)^2. \tag{11}$$

Note that in (11), $\Delta w_i = w_i$ corresponds to removing the ith variable, which represents that the smallest weight value contributes the least to the resulting hyperplane and can be discarded.

It should also be noted that in order for RFE to work, variable scaling (Guyon and Elisseeff 2003) has to be performed.

To perform variable selection on multiresolution LBP features, the recursive elimination procedure of the SVM-RFE is as follows.

(1) Start: ranked variable list $L = []$; selected variable subset $S = \{1, \ldots, d\}$.
(2) Repeat until all variables of multiresolution LBP feature are ranked.
    (a) Train an SVM with multiresolution LBP features in set S as input variables.
    (b) Compute the weight vector using (8).
    (c) Compute the ranking scores for variables in S: $c_i = (w_i)^2$.
    (d) Find the variable with the smallest ranking score: $e = \arg\min_i c_i$.
    (e) Update: $L = [e, L]$, $S = S - \{e\}$.
(3) Output: ranked variable list L.

5. Experiments

In this section, we have conducted a series of experiments to study the effectiveness of the proposed approach for extracting multiresolution LBP features and to study the effect of applying variable selection on multiresolution LBP features to improve accuracy in FP reduction. The details on our experimental studies are given in the following subsections.

5.1. Experimental setup and condition

In our experiments, two public benchmark mammogram datasets MIAS (Suckling *et al* 1994) and DDSM (Heath *et al* 2000) DBs were used. MIAS DB consists of film mammography images having a pixel size of 200 μm with 8-bit gray-level resolution. A total of 89 single-view mammograms (each containing one mass) were collected from the MIAS DB in our study. These collected 89 mammograms consisted of 34 fatty, 31 fatty-glandular and 24 dense-glandular background tissue. Note that these mammograms were chosen to well characterize the diversity of masses, encountered in clinical practice. As for another different dataset, the 303 single-view mammograms were collected from the DDSM DB. The selection of these mammograms was based on the following criteria: (1) in DDSM DB, mammogram images are digitized by different scanners with different resolutions (Heath *et al* 2000), (Catarious *et al* 2006). Hence, for data consistency purposes, all 303 mammograms were randomly selected from the same type of scanner and resolution; (2) we chose 303 mammograms acquired using the scanner type Howtek 960 because a large number of cases are digitized by this type (Heath *et al* 2000). Collected images from the DDSM were digitized with a pixel size of 43.5 μm with 12-bit gray level resolution. In this study, all images collected from the DDSM were subsampled to 200 μm and quantized to 8 bits per pixel for computational efficiency. The 200 μm pixel size is consistent with that used by another group (Catarious *et al* 2004, 2006).

As described in section 2, using a computer segmentation algorithm, two sets of ROI images were generated from the MAIS and the DDSM DBs, respectively. Note that, during the generation of ROIs, we chose operating threshold (point) which led to an average number of 11 FPs per image at a detection sensitivity of about 80% with MAIS DB. Also, when using DDSM DB, operating threshold at 8 FPs per image with about 81% sensitivity was selected. With these fixed operating thresholds, FP reduction analysis was performed on all approaches used in the following experiments. Using 89 images collected from the MAIS DB, a total of 1 693 ROIs were generated: 72 mass ROIs and 1 621 normal tissue ROIs (i.e. FP regions).

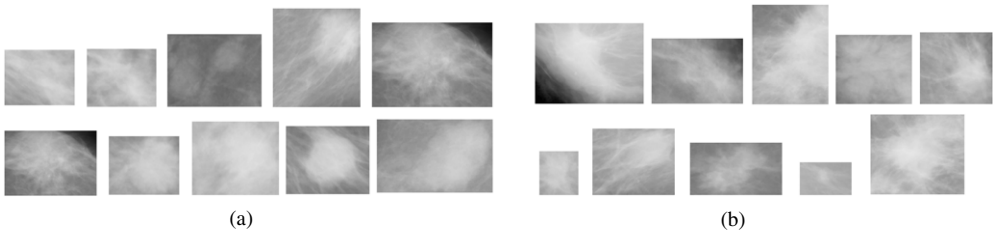


Figure 5. The example of ROIs containing masses with different image sizes that are automatically extracted from mammograms by using computer segmentation algorithm described in section 2. (a) MIAS DB. (b) DDSM DB.

72 mass ROIs included 20 circumscribed masses, 15 spiculated masses, 14 ill-defined masses, 12 architectural distortions and 11 asymmetries. Note that an ROI extracted from the MIAS DB is regarded as a true mass ROI when the ROI includes the center of abnormality provided by the MIAS DB (Mudigonda *et al* 2001, Dominguez and Nandi 2008). Moreover, a total of 2 743 ROIs were automatically generated by using 303 images collected from the DDSM DB: 246 mass ROIs and 2 497 normal tissue ROIs. DDSM DB provides annotations of the masses presented in each image. These annotations were considered as the ground truth in our experiments as already described in figure 1. Using ground truth information, a generated ROI was considered as a true mass ROI only if it met the following two criteria (Eltonsy *et al* 2007, Hsu *et al* 2010, Yu *et al* 2010): (1) the centroid of a segmented region is included in the DDSM annotated area and (2) a segmented region intersects with the true mass region more than 25%.

Figure 5 shows examples of ROIs of the mammograms collected from both MIAS and DDSM DBs. In addition, figure 6 shows information of mass dataset constructed using the DDSM DB. It can be seen from figure 6 that the masses with different shapes and density found in clinical practice were fairly well represented in the data set by containing a broad variety of mass shapes, margin characteristics and breast densities. Also note that as shown in figure 1(b), errors caused by mass segmentation (using computer algorithm) were not excluded from our experiments. Therefore, it should be emphasized that this allows testing the proposed method in realistic clinical conditions, as also encountered by real-world CAD applications.

The SVM classifier which utilizes a radial basis function (Guyon *et al* 2002, Vapnik 1998) (as kernel) was used for a classifier. A total of ROIs for two datasets were randomly divided into two sets of equal size: training and testing sets. For each random partition, we train an SVM classifier (including the process of optimizing both the associated parameter of the kernel function and the regularization parameter to achieve the smallest generalization error (Samulski and Karssemeijer 2011)) on the training set and then test it on the corresponding test set. Note that, to guarantee stable classification results, 40 independent runs of aforementioned random partitions were executed. Thus, all of the results reported in this section were averaged over 40 runs. The SVM classification outputs were used as the decision variable in receiver operating characteristics (ROC) analysis (Suri and Rangayyan 2006) to evaluate the classification performance. ROC analysis evaluates the relationship between the true-positive fraction (TPF) or sensitivity and the false-positive fraction (FPF) or specificity (1–FPF) at differing decision thresholds. The area under the ROC curve denoted by Az was used as an index of classification accuracy. Moreover, to test the statistical significance of the difference in Az (the area under the ROC curve) of different methods, a popular statistical test developed by Hanley and McNeil (1983) was used.

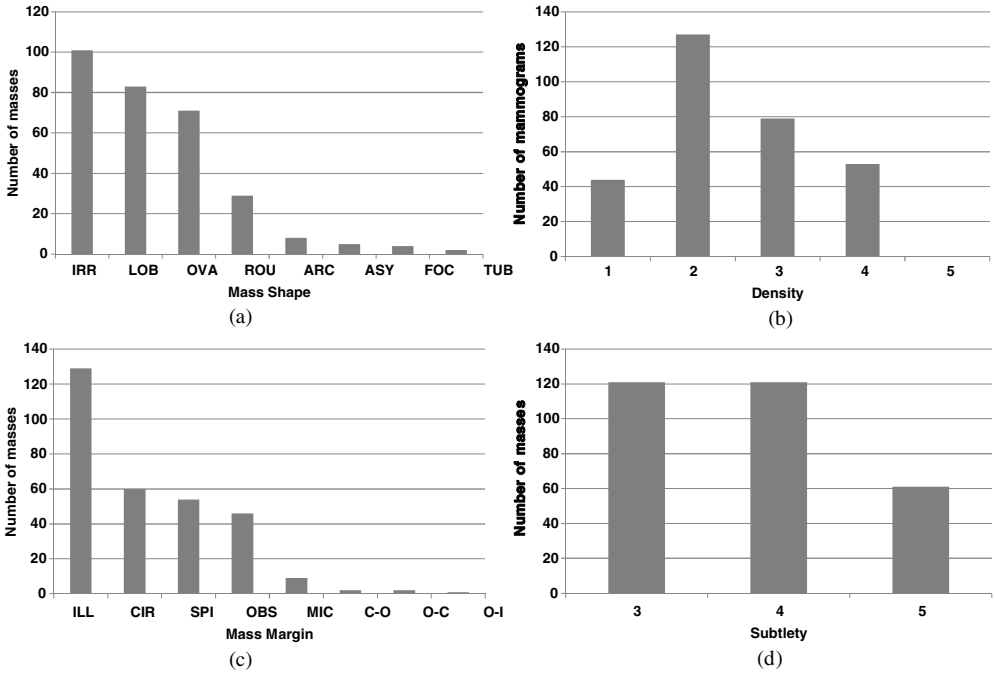


Figure 6. Information of our mass dataset in the case of the DDSM DB. (a) Distribution of mass shapes, IRR: irregular, LOB: lobulated, OVA: oval, ROU: round, ARC: architectural distortion, ASY: asymmetry, FOC: focal asymmetry, TUB: tubular, (b) distribution of the breast density in terms of BI-RADS category, (c) distribution of mass margins, ILL: ill defined, CIR: circumscribed, SPI: spiculated, OBS: obscured, MIC: microlobulated, C–O: circumscribed-obscured, O-C: obscured-circumscribed, O-I: obscured-ill defined and (d) distribution of mass subtlety ranging from 1 (subtle) to 5 (obvious).

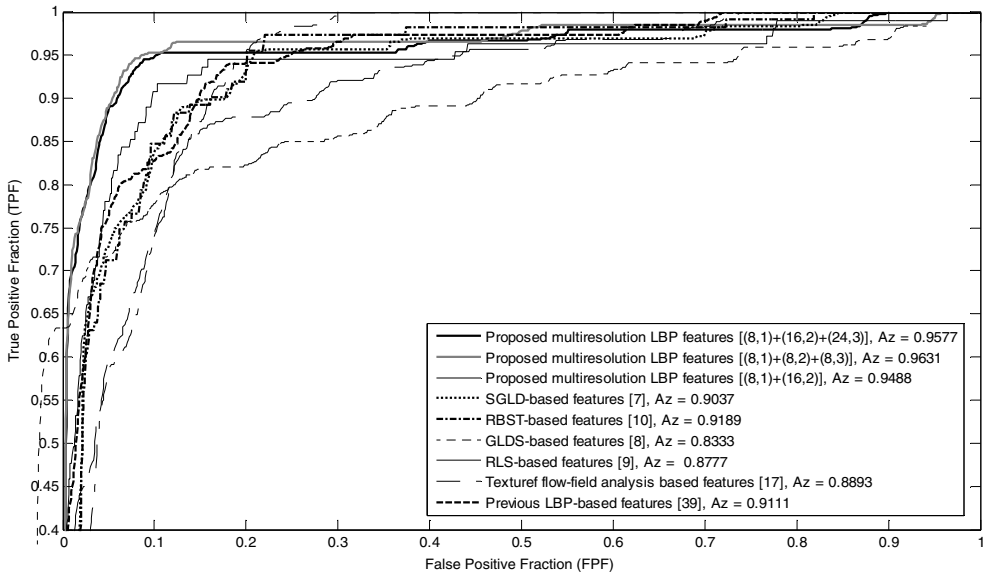
5.2. Classification with multiresolution LBP features

In this section, an extensive and comparative experimental study has been conducted to demonstrate the effectiveness of the proposed multiresolution LBP features. For this purpose, we compare the proposed texture feature against other six texture features (developed for the purpose of mammographic mass classification): (1) texture flow-field analysis-based features (Mudigonda *et al* 2001), (2) SGLD-based features (Wei *et al* 1997), (3) GLDS-based features (Sahiner *et al* 1996), (4) RLS-based features (Sahiner *et al* 2001), (5) RBST-based features (Sahiner *et al* 1998) and (6) previous LBP-based features (Lladó *et al* 2009). It should be noted that, when implementing other texture features, we have strictly followed the implementation guidelines (and parameter settings) listed in the respective published paper for the purpose of stable comparisons. For texture flow-field analysis-based features, the four features, namely 'contrast', 'coherence ratio', 'entropy of orientation', and 'variance of coherence-weighted angular estimates' were extracted based on textual flow-field analysis (Mudigonda *et al* 2001). For SGLD-based features, 13 features, namely 'correlation', 'energy', 'entropy', 'inertia', 'inverse difference moment', 'sum average', 'sum variance', 'sum entropy', 'difference energy', 'difference variance', 'difference entropy', 'information measure of correlation 1' and 'information measure of correlation 2' were extracted from each SGLD matrix at six different interpixel distances ($d = 1, 2, 4, 6, 8$ and 10) and in four directions ($\theta = 0^\circ, 45^\circ, 90^\circ$and 135°) were used to produce 24 SGLD matrices (Wei *et al*

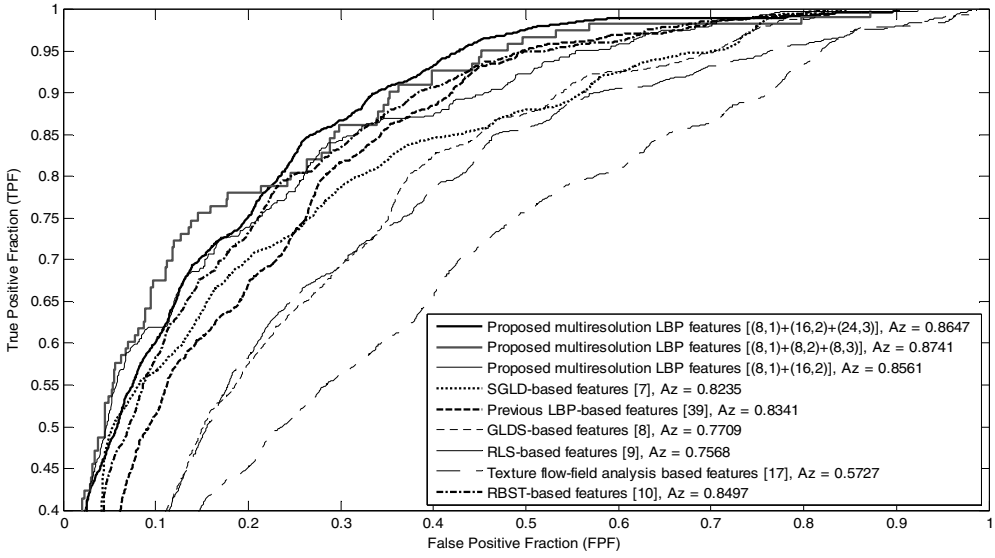
1997, Cheng *et al* 2005). Therefore, a total of 312 SGLD features were computed for each ROI. For GLDS-based features, the following four features 'contrast', 'angular second moment', 'entropy', and 'mean' were extracted from the gray-level difference statistics vector (Sahiner *et al* 1996, Cheng *et al* 2005). Analogous to SGLD matrix computation, six interpixel distances and four directions were used to calculate 24 GLDS vectors, yielding 96 GLDS features. For RLS-based texture features, five texture measures, namely 'short run emphasis', 'long runs emphasis', 'gray-level nonuniformity', 'run-length nonuniformity' and 'run percentage' were obtained from the gray-level run length matrices with four directions, $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ (Sahiner *et al* 2001, Cheng *et al* 2005). Hence, a total of 20 RLS-based features were calculated for each ROI image. For RBST-based texture features, using eight different pixel pair distances ($d = 1, 2, 3, 4, 6, 8, 12$ and 16) and in four directions ($\theta = 0^\circ, 45^\circ, 90^\circ$and 135°), the SGLD matrices were calculated from the RBST image representation (Sahiner *et al* 1998). Next, as recommended in Sahiner *et al* (1998), eight texture measures, namely 'correlation', 'energy', 'difference entropy', 'inverse difference moment', 'entropy', 'sum average', 'sum entropy' and 'inertia' were extracted from each SGLD matrix. As a result, a total of 256 SGLD features for each RBST image representation were used in our experimentation. Note that, following the same parameter values as used in Sahiner *et al* (1998), the 40-pixel-wide band was used to construct the RBST images. For implementing previously developed LBP features, we divided each ROI image into 5×5 squared local regions and using a uniform LBP operator with (P, R) values of (8,1), we extracted LBP histograms from these local regions and subsequently concatenated them (Lladó *et al* 2009). Next, as suggested by authors in Lladó *et al* (2009), this concatenated histogram was combined with the LBP histograms extracted from the 3×3 central regions to form the final texture descriptor.

For comparison purposes, the ROC curves obtained using seven different texture features are plotted in figure 7. Note that in figure 7, in order to guarantee the stability of classification performance with respect to different settings of LBP parameters (P, R), three different resolution combinations are used to compute the ROC curves obtained for the proposed multiresolution LBP features. Specifically, as suggested in Ojala *et al* (2002), we use the two 3-resolution combinations and the one 2-resolution combination for multiresolution analysis. As shown in figure 7, the proposed multiresolution LBP features achieve better classification performance, compared to the other six texture features, for both MAIS and DDSM DBs. In particular, the proposed multiresolution LBP features (for the case of using (P, R) values of (8, 1), (8, 2), and (8, 3)) attain the best A_z values up to 0.9631 and 0.8741 for MIAS and DDSM DBs, respectively, of all texture features under consideration. It should be also noted that our multiresolution LBP features considerably outperform the previous LBP-based features proposed in Lladó *et al* (2009) for both datasets. This result demonstrates that the proposed approach is more suitable for reliably encoding the texture patterns of the mass subregions (e.g., the region inside the mass) via conventional LBP operation, as well as for extracting complementary and discriminating information from the mass region by employing multiresolution analysis.

Clinically, the specificities at high sensitivity levels are important measures of classification accuracy because the cost of missing a cancer is greater than the cost of performing a biopsy to assess a lesion (Zheng *et al* 2001, Suri and Rangayyan 2006, Sahiner *et al* 1996). Thus, the performances, in terms of specificity at a given high sensitivity level, were also evaluated. Based on the results presented in figure 7, we then compared the FPF at TPFs of 95% and 90% obtained using the proposed multiresolution LBP features, and the second best RBST-based features. When using MIAS DB, at a TPF of 95%, the average FPF over the three different settings of LBP parameters (i.e. three different resolution combinations) is 0.1193



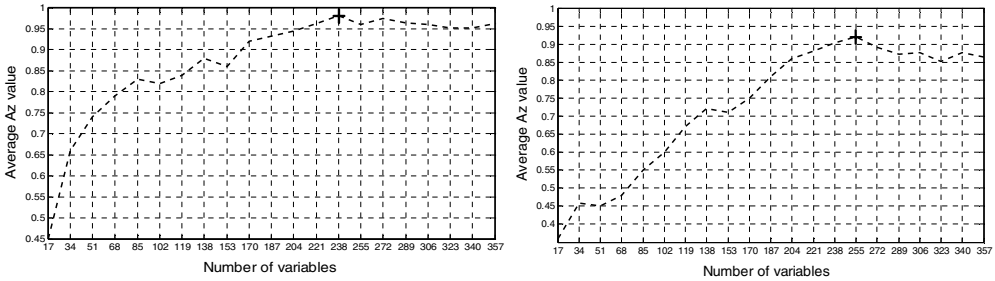
(a)



(b)

Figure 7. Comparisons of ROC curves, with respect to six different types of texture features. Note that ROC curves for the proposed multiresolution LBP features are obtained without variable selection process. Also note that in the proposed multiresolution LBP features, the values enclosed in square brackets represent the values of (P, R) for realizing LBP operation with different spatial resolutions and different angular resolutions. (a) MAIS DB. (b) DDSM DB.

for the multiresolution LBP feature, while the FPF is 0.2013 for the RBST-based feature. At a TPF of 90%, the average FPFs are 0.0710 for the multiresolution LBP feature and 0.1609 for RBST-based feature. Note that similar results are also obtained for the case of using DDSM DB. These encouraging results demonstrate the potential of applying our multiresolution LBP features to a mammographic mass classification in a practical CAD system.



(a) (b)

Figure 8. Average A_z values at different sizes of variable subsets selected by SVM-RFE algorithm. The performance of the best variable subset selected is denoted by '+'. (a) MIAS DB. (b) DDSM DB.

Table 1. Classification performances of the proposed multiresolution LBP features with and without variable selection by SVM-RFE or SFS methods. Note that the average number of variables selected by SVM-RFE and SFS are tabulated. (a) MIAS DB. (b) DDSM DB.

		Performance measurement		
Approach	Number of variables	A_z	TPF (%) at FPF of 10%	TPF (%) at FPF of 20%
	(a)			
Without variable selection	Full (357)	0.963 ± 0.021	95.30 ± 2.09	96.54 ± 1.34
With SFS	202	0.972 ± 0.014	95.47 ± 2.41	97.81 ± 1.77
With SVM-RFE	238	0.988 ± 0.010	96.99 ± 2.37	98.57 ± 0.45
	(b)			
Without variable selection	Full (357)	0.874 ± 0.019	67.47 ± 3.23	78.04 ± 1.37
With SFS	198	0.891 ± 0.023	73.95 ± 2.91	87.86 ± 0.51
With SVM-RFE	255	0.921 ± 0.026	78.68 ± 3.63	91.67 ± 1.01

5.3. Effect of incorporating variable selection for FP reduction

In this section, we present the results for evaluation of effectiveness of combining the SVM-RFE-based variable selection with the proposed multiresolution LBP features. The variable selection procedure (described in section 4) was performed on multiresolution LBP features to further improve classification performance. It should be noted that the goodness of a variable subset selected by the SVM-RFE method was evaluated by the performance of an SVM classifier built on the training set. More specifically, the goodness of a variable subset is examined by the performance of an SVM classifier trained with variables in the subset as input variables. Also note that in the computation of multiresolution LBP features, (P, R) values of (8,1), (8,2) and (8,3) were chosen in the following experiments because this resolution combination shows the best classification performance of all kinds of resolution combinations represented in figure 7.

We plot the average A_z value versus the size of variable subsets selected by the SVM-RFE in figure 8. Based on the results shown in figure 8, the variable subset with the largest value of A_z is chosen as the best variable subset. To validate the feasibility of the SVM-RFE (as a variable selection approach) for enhancing classification accuracy, the test performances of the SVM with full number of variables are reported in table 1, together with those of the best variable subset selected by the SVM-RFE. In addition, for the purpose of comparison,

we also report the classification accuracy of the same SVM classifier using a well-established feature selection method for mass classification, the so-called stepwise feature selection (SFS) (Norusis 1933). At each step of the SFS procedure, one feature (or variable) is entered into or removed from the selected feature pool by analyzing its effect on a selection criterion. In this experiment, we employed the Wilks' lambda as our selection criterion, which is defined as the ratio of features selected by this method is controlled by two parameters, called *F*-to-enter and *F*-to-remove (Norusis 1933). A detailed description of implementing the SVM classifier in combination with the SFS procedure can be found in the literature (Way *et al* 2010). Looking at the results in table 1(a) and (b), we can see that the classification performances with variable selection using either the SVM-RFE or SFS are better than those obtained using all variables (i.e. without variable selection) as input feature to an SVM classifier for both MAIS and DDSM datasets. In particular, for the case of using MAIS DB, a statistical comparison between the proposed the SVM-RFE and the method using all variables (i.e. without variable selection) yields a two-tailed *p*-value of 0.031 (deemed statistically significant at a level of 0.05 (Zheng *et al* 2001, Birdwell *et al* 2001)) for rejecting the null hypothesis that their corresponding ROC curves have the same area under them; moreover a 95% confidence interval on the difference in *Az* between the two methods is (0.008, 0.071). Also, when using DDSM DB, a statistical comparison between two methods yields *p*-value of 0.018 and also 95% confidence interval is (0.012, 0.096). It can also be observed that the classification results with the SVM-RFE algorithm are better than their counterparts with the SFS. These results indicate that selecting an appropriate subset of the multiresolution LBP feature using the SVM-RFE algorithm is feasible for providing further improvement in classification accuracy.

Figure 9 shows some mammography examples for validating the usefulness of the proposed solution (to combining our multiresolution LBP features with SVM-RFE variable selection) for the purpose of correctly identifying masses from FP regions. From figure 9, we can see that almost all FP regions have been removed after the FP reduction (without missing true mass regions) performed via the proposed approach. These examples demonstrate that the proposed approach has the potential for aiding radiologists in considerably reducing the number of unnecessary biopsies performed without misclassifying mammographic masses.

5.4. Investigating effectiveness of multiresolution LBP features

Recall from section 3 that the following three different types of information in characterizing masses are encoded into the multiresolution LBP features in a hierarchically manner: (1) regional LBP distribution information, (2) spatial structure information and (3) multiresolution information. To validate the potential advantage of fusing all of the aforementioned information based on our method, we have performed additional experiments. The first part of the analysis was focused on investigating the effectiveness of combining two individual LBP distribution patterns each computed from core and margin regions, respectively, and compared against a framework utilizing separately these LBP distribution patterns. The second part was focused to examine the effect of exploiting multiresolution information on the improvement in classification accuracy by comparing the cases of using only single resolution during the computation of LBP texture features.

The results are given in figure 10 and table 2. From figure 10 and table 2, the following observations can be made. First, the combination of LBP distribution patterns extracted from both core and margin regions achieve much better results, compared to the cases of separately using them. This result confirms that the LBP distribution patterns for core and margin region parts are able to provide different discriminant information and to be mutually compensational in terms of improving classification accuracy. Second, LBP distribution information extracted

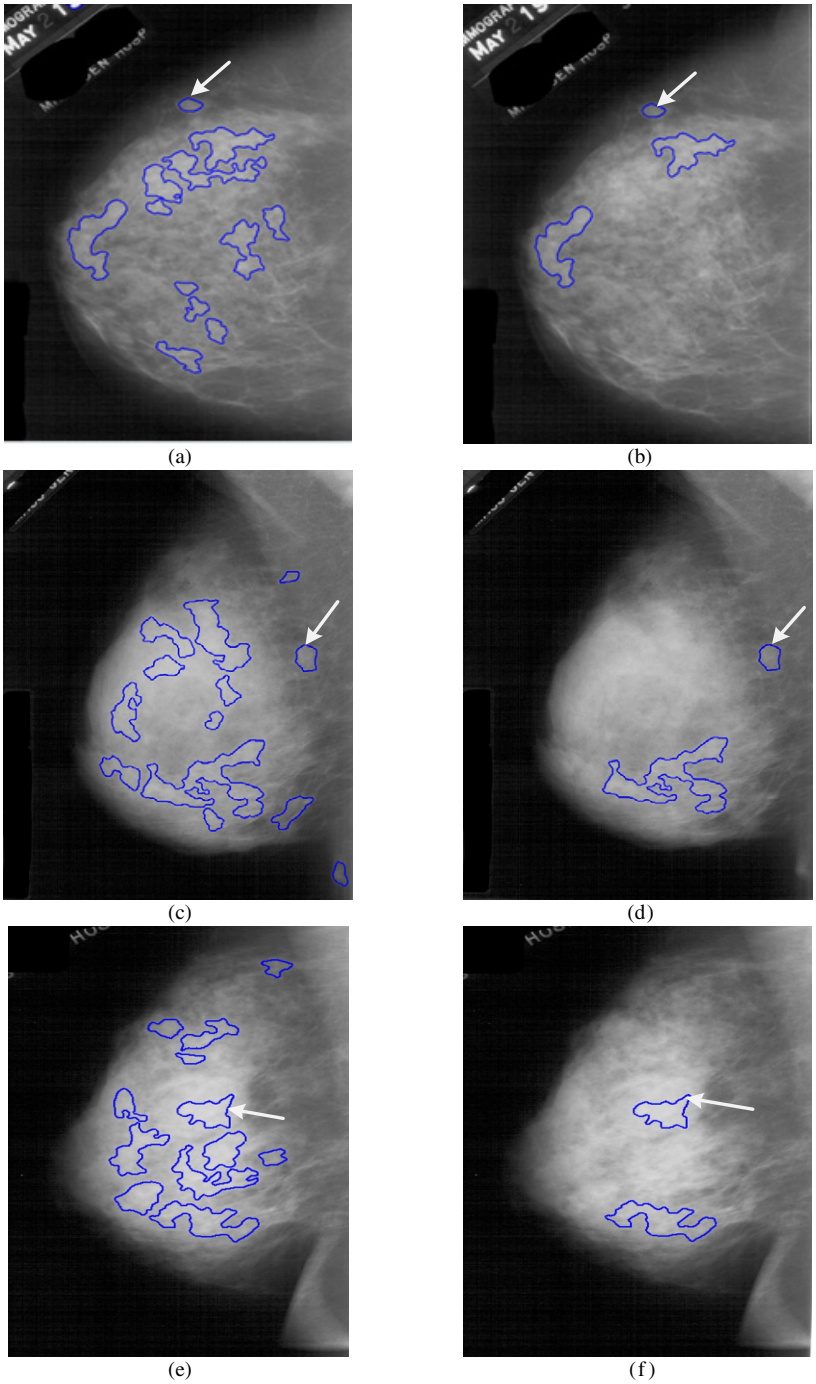


Figure 9. Examples of mammograms (chosen from the DDSM DB) for demonstrating the effectiveness of using our multiresolution LBP features together with SVM-RFE for reducing the number of FP objects. (a), (c) and (e) The mammograms superimposed with detected suspicious ROIs; true masses are indicated by white colored arrows in all mammograms. (b), (d) and (f) The corresponding mammograms resulting from FP reduction performed via the proposed approach.

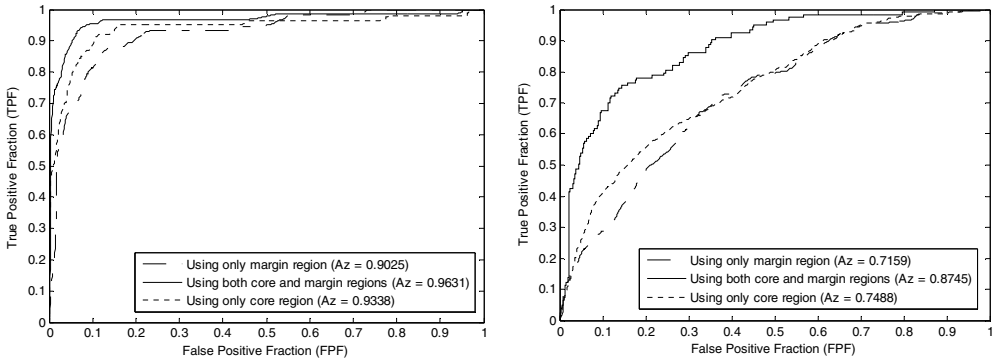


Figure 10. ROC curves to demonstrate the effectiveness of 'combined (or fused)' LBP distribution patterns extracted using both core and margin regions. Note that LBP operation with (P, R) values of (8,1), (8,2) and (8,3) was used to compute ROC curves. (a) MIAS DB. (b) DDSM DB.

Table 2. Comparisons of classification performances obtained for single resolution vs. multiresolution analysis to show the effectiveness of incorporating multiresolution information. (a) MIAS DB. (b) DDSM DB.

		Performance measurement		
Approach	Values of (P, R)	A_z	TPF (%) at FPF of 10%	TPF (%) at FPF of 20%
	(a)			
Single resolution analysis	(8,1)	0.903 ± 0.029	88.76 ± 2.67	92.02 ± 1.77
	(16,2)	0.908 ± 0.025	88.89 ± 2.02	93.11 ± 0.98
	(24,3)	0.896 ± 0.038	85.54 ± 3.88	90.59 ± 2.39
Multiresolution analysis	(8,1)+(8,2)+(8,3)	0.963 ± 0.021	95.30 ± 2.09	96.54 ± 1.34
	(8,1)+(16,2)+(24,3)	0.957 ± 0.017	94.95 ± 1.99	95.36 ± 0.83
	(8,1)+(16,2)	0.948 ± 0.011	90.84 ± 1.81	94.49 ± 0.65
	(b)			
Single resolution analysis	(8,1)	0.829 ± 0.017	55.93 ± 1.68	70.95 ± 1.88
	(16,2)	0.835 ± 0.022	57.17 ± 2.90	72.02 ± 1.61
	(24,3)	0.823 ± 0.011	54.91 ± 2.41	68.56 ± 1.03
Multiresolution analysis	(8,1) + (8,2) + (8,3)	0.874 ± 0.019	67.47 ± 3.23	78.04 ± 1.37
	(8,1) + (16,2) + (24,3)	0.864 ± 0.023	61.09 ± 2.11	75.28 ± 1.40
	(8,1) + (16,2)	0.856 ± 0.016	61.95 ± 1.46	73.98 ± 0.32

from the core region yields better classification accuracy compared to that extracted from the margin region. This result is an indication that texture statistics (or patterns) extracted from the 'inner regions' of the masses are important characteristics in the classification of ROIs as either containing mass or normal tissue. This observation can be supported by the following two commonly used assumptions (in the detection of mammographic masses) (Dominguez and Nandi 2008, Hong and Sohn 2010, Eltonsy *et al* 2007): (1) compared to normal tissues, the regions inside the masses are generally brighter and more dense than their surrounding regions; (2) the gray levels at the inner regions of the masses are relatively uniform; at least, they usually change in a gradual manner. Finally, a multiresolution analysis allows for achieving higher classification accuracy in comparison to the results obtained with LBP features at a single resolution. In support of this observation, a statistical comparison of the results obtained using a multiresolution analysis (for the case of using (P, R) values of (8, 1), (8, 2), and (8, 3)) with

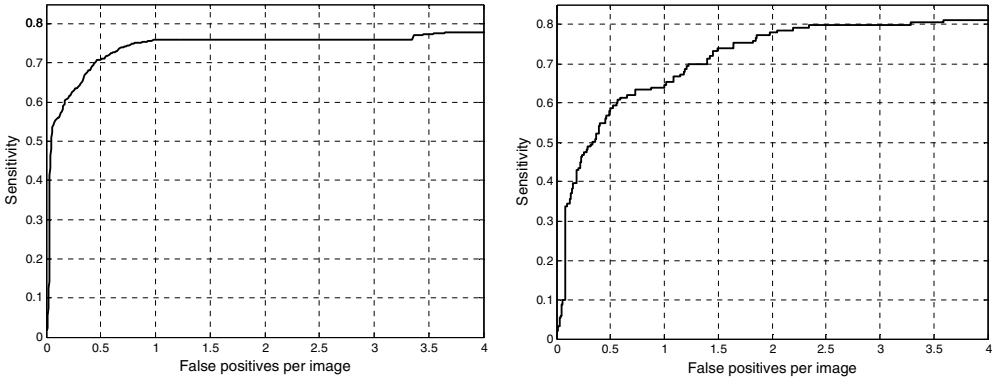


Figure 11. FROC results of a full CAD system based on the proposed multiresolution LBP features in conjunction with SVM-RFE variable selection. (a) MIAS DB. (b) DDSM DB.

those obtained using single resolution (with (P, R) values of (8, 1)) yields a two-tailed p-value of about 0.008 and 0.012 for MAIS and DDSM DBs, respectively. Also, when considering a 95% confidence interval on the difference in Az, we obtain (0.031, 0.088) and (0.027, 0.063) for MAIS and DDSM DBs, respectively. Underlying reason for above result can be explained as follows. The masses found in clinical mammograms have different sizes and shapes (Huo *et al* 1998). Thus, it is generally difficult to define *a priori* an optimal and universal resolution for the ROIs to be used for texture analysis. However, it can be expected that a multiresolution analysis is beneficial when dealing with the aforementioned problem because it could provide a scale-invariant interpretation of an image (Ojala *et al* 2002).

5.5. Evaluating mass detection and number of FPs

In this section, we present overall detection performance of a full CAD system based on the proposed multiresolution LBP features in conjunction with SVM-RFE variable selection. For this purpose, free-response receiver operating characteristic (FROC) curves (Suri and Rangayyan 2006, Oliver *et al* 2010) were used. A FROC curve is obtained by plotting sensitivity on the y axis and the number of FPs per image on the x axis. The FROC results obtained using the sets of 89 MAIS mammograms and 303 DDSM mammograms are shown in figures 11(a) and (b), respectively. For the images from the MAIS DB, we can obtain a sensitivity of 75.86% at 1 FP per image, while we can achieve a sensitivity of 77.88% at 2 FPs for the DDSM DB. Note that, compared to the number of FPs per image at initial detection sensitivity, a significant FP reduction can be made by using the proposed method; for the MAIS DB, the overall FP rate is reduced as much as 90.91% (from 11 to 1 per image) at only cost of 4.98% sensitivity loss (from 80.84% to 75.86%), while it can be greatly reduced for 75.00% (from 8 to 2 per image) with only loss of 3.30% sensitivity rate (from 81.18% to 77.88%) for the DDSM DB.

In a previous report on the detection performance on the MAIS DB, the detection accuracy of masses has been reported as 62% sensitivity at 1 FP per image (Sampat *et al* 2008). In Mudigonda *et al* (2001), the CAD system is reported to have 81% sensitivity at 2.2 FPs per image for mass versus normal tissue classification. For the detection of masses in the DDSM, the authors in Catarious *et al* (2004) reported a sensitivity of 80% at 4.5 FPs per image for the testing set. In another mass detection study, which used images from the DDSM, the authors in Catarious *et al* (2006) reported a sensitivity of 72% at 2 FPs per image for overall mass detection performance (including the malignant and benign masses). The authors in Brake

et al (2000) reported 81% sensitivity with FPs (per image) ranging from 1.35 to 3.66 FPs for malignant mass detection performance. Note that the results of any given CAD algorithms can vary significantly depending on the dataset used (Nishikawa *et al* 1994) and two CAD algorithms can only be directly compared when the same training and testing sets are used (Suri and Rangayyan 2006). Hence, it should be pointed out that the aforementioned discussion on the comparison of results of previous studies aims at providing context for our work rather than a direct comparison with our results, which is not meaningful because a different dataset was used in the reference study.

6. Conclusions

We have developed a new and novel solution to extracting texture feature, referred to as multiresolution LBP feature, for the characterization of mammographic masses. In our approach, features describing the texture statistics of the mass were extracted from the core and margin regions of the segmented mass, respectively. It was found that combining texture patterns extracted from both core and margin regions was considerably more effective than using texture patterns extracted from using only core or margin regions. In addition, in our study, we demonstrated that incorporating multiresolution analysis into the extraction of LBP texture features is helpful for enhancing the classification of masses and normal tissues. It was also validated that combining our multiresolution LBP features and SVM-RFE variable selection can further improve the classification accuracy. Based on our experimental results, the proposed LBP feature extraction solution has clearly proven to be significant for the purpose of FP reduction and is not sensitive to the precise location of boundaries of masses. Thus, it can be expected that the proposed method could considerably improve radiologists' accuracy in the characterization of masses in CAD systems, and thereby might reduce unnecessary biopsies.

For future work, in order to further guarantee the feasibility of our method in realistic clinical conditions, as well as the generalizability of our results, the proposed method should be tested on a large mammogram DB collected from a larger number of patients undergoing routine mammographic screening at some medical centers. Also, future work includes extending application of the proposed method to other different classification tasks of importance in a CAD system, such as the differentiation of malignant and benign tissue in mammography (Sahiner *et al* 1998).

Acknowledgments

This paper is partially supported by the development of MEXA image enhancement and processing (A080782) of the Ministry for Health, Welfare and Family Affairs in the Republic of Korea.

References

Birdwell R L, Ikeda D M, O'Shaughnessy K F and Stickles E A 2001 Mammographic characteristics of 115 missed cancers later detected with screening mammography and the potential utility of computer-aided detection *Radiology* **219** 192–202

Bishop C M 2006 *Pattern Recognition and Machine Learning* (Berlin: Springer)

Brake G M, Karssemeijer N and Hendriks J H 2000 An automatic method to discriminate malignant masses from normal tissue in digital mammograms *Phys. Med. Biol.* **45** 2843–57

Catarious D M, Baydush A H and Floyd C E 2004 Incorporation of an iterative, linear segmentation routine into a mammographic mass CAD system *Med. Phys.* **31** 1512–20

Catarious D M Jr, Baydush A H and Floyd C E Jr 2006 Characterization of difference of Gaussian filters in the detection of mammographic regions *Med. Phys.* **33** 4104–14

Chan H P, Wei D, Helvie M A, Sahiner B, Adler D D, Goodsitt M M and Petrick N 1995 Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space *Phys. Med. Biol.* **40** 857–76

Cheng H D, Shi X J, Min R, Hu L M, Chi X P and Du H N 2005 Approaches for automated detection and classification of masses in mammograms *Pattern Recognit.* **39** 646–68

Dominguez A R and Nandi A K 2008 Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection *Comput. Med. Imaging and Graph.* **32** 304–15

Eltonsy N H, Tourassi G D and Elmaghraby A S 2007 A concentric morphology model for the detection of masses in mammography *IEEE Trans. Med. Imaging* **26** 880–9

Freer T W and Ulissey M J 2001 Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center *Radiology* **220** 781–6

Gonzalez R C and Woods R E 2008 *Digital Image Processing* 3rd edn (New York: Prentice-Hall)

Guyon I and Elisseeff A 2003 An introduction to variable and feature selection *J. Mach. Learn. Res.* **3** 1157–82

Guyon I, Weston J, Barnhill S and Vapnik V 2002 Gene selection for cancer classification using support vector machines *Mach. Learn.* **46** 389–422

Hanley J A and McNeil B J 1983 A method of comparing the areas under receiver operating characteristic curves derived from the same cases *Radiology* **148** 839–43

Heath M, Bowyer K, Kopans D, Moore R and Kegelmeyer P J 2000 The digital database for screening mammography *Proc. 5th Int. Workshop on Digital Mammography* pp 212–8

Hong B W and Sohn B S 2010 Segmentation of regions of interest in mammograms in a topographic approach *IEEE Trans. Inform. Technol. Biomed.* **14** 129–39

Hsu C-W, Chang C-C and Lin C-J 2010 A practical guide to support vector classification *Bioinformatics* **1** 1–16

Huo Z, Giger M L, Vyborny C J, Wolverton D E, Schmidt R A and Doi K 1998 Automated computerized classification of malignant and benign masses on digitized mammograms *Acad. Radiol.* **5** 155–68

Jain A, Nandakumar K and Ross A 2005 Score normalization in multimodal biometric systems *Pattern Recognit.* **38** 2270–85

Lladó X, Oliver A, Freixenet J, Marti R and Marti J 2009 A textural approach for mass false positive reduction in mammography *Comput. Med. Imaging Graph.* **33** 415–22

Mangai U G, Samanta S, Das S and Chowdhury P R 2010 A survey of decision fusion and feature fusion strategies for pattern classification *IETE Tech. Rev.* **27** 293–307

Mudigonda N R, Rangayyan R M and Leo Desaultels J E 2001 Detection of breast masses in mammograms by density slicing and texture flow-field analysis *IEEE Trans. Med. Imaging* **20** 1215–27

Nishikawa R M, Giger M L, Doi K, Metz C E and Yin F F 1994 Effect of case selection on the performance of computer-aided detection scheme *Med. Phys.* **21** 265–9

Norusis M J 1933 *SPSS Professional Statistics 6.1* (Chicago: SPSS Inc)

Ojala T, Pietikainen M and Harwood D 1996 A comparative study of texture measures with classification based on feature distributions *Pattern Recognit.* **29** 51–9

Ojala T, Pietikainen M and Maenpaa T 2002 Multiresolution gray-scale and rotation invariant texture classification with local binary patterns *IEEE Trans. Pattern Anal. Mach. Intell.* **24** 971–87

Oliver A, Freixenet J, Marti J, Perez E, Pont J, Denton E R E and Zwiggelaar R 2010 A review of automatic mass detection and segmentation in mammographic masses *Med. Image Anal.* **14** 87–110

Oliver A, Lladó X, Freixenet J and Marti J 2007 False positive reduction in mammographic mass detection using local binary patterns *Proc. 10th Int. Conf. on Medical Image Computing and Computer Assisted Intervention* part 1 pp 286–93

Petrick N, Chan H P, Sahiner B and Helvie M A 1999 Combined adaptive enhancement and region-growing segmentation of breast masses on digitized mammograms *Med. Phys.* **26** 1642–54

Sahiner B, Chan H P, Petrick N, Helvie M A and Hadjiiski L M 1998 Computerized characterization of masses on mammograms: the rubber band straightening transform and texture analysis *Med. Phys.* **25** 516–26

Sahiner B, Chan H P, Petrick N, Helvie M A and Hadjiiski L M 2001 Improvement of mammographic mass characterization using speculation measures and morphological features *Med. Phys.* **28** 1455–65

Sahiner B, Chan H P, Petrick N, Wei D, Helvie M A, Adler D D and Goodsitt M M 1996 Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images *IEEE Trans. Med. Imaging* **15** 598–609

Sahiner B, Chan H P, Wei D, Petrick N, Helvie M A, Adler D D and Goodsitt M M 1996 Image feature selection by a genetic algorithm: application to classification of mass and normal breast tissue *Med. Phys.* **23** 1671–84

Sampat M P, Bovik A C, Whitman G J and Markey M K 2008 A model-based framework for the detection of spiculated masses on mammography *Med. Phys.* **35** 2110–23

Samulski M and Karssemeijer N 2011 Optimizing case-based detection performance in a multiview CAD system for mammography *IEEE Trans. Med. Imaging* **30** 1001–9

Suckling J *et al* 1994 The mammographic image analysis society digital mammogram database *Proc. 2nd Int. Workshop on Digital Mammography* pp 375–8

Suri J S and Rangayyan R M 2006 *Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer* (Washington: SPIE Press)

Tang J, Rangayyan R M, Xu J, Naqa I E and Yang Y 2009 Computer-aided detection and diagnosis of breast cancer with mammography *Recent Adv. IEEE Trans. Inf. Technol. Biomed.* **13** 236–51

Vapnik V N 1998 *Statistical Learning Theory* (New York: Wiley)

Way T W, Sahiner B, Hadjiiski L M and Chan H P 2010 Effect of finite sample size on feature selection and classification: a simulation study *Med. Phys.* **37** 907–20

Wei D, Chan H P, Helvie M A, Sahiner B, Petrick N, Adler D D and Goodsitt M M 1995 Classification of mass and normal breast tissue on digital mammograms: multiresolution texture analysis *Med. Phys.* **22** 1501–13

Wei D, Chan H P, Petrick N, Sahiner B, Helvie M A, Adler D D and Goodsitt M M 1997 False-positive reduction technique for detection of masses on digital mammograms: global and local multiresolution texture analysis *Med. Phys.* **24** 903–14

Yu Z, Noriko T, Jacob F and Daniela S R 2010 Image enhancement and edge-based mass segmentation in mammogram *Proc. SPIE* **7623** 76234P

Zheng B, Ganott M A, Britton C A, Hakim C M, Hardesty L A, Chang T S, Rockette H E and Gur D 2001 Soft-copy mammographic readings with different computer-assisted detecting cuing enviorments: preliminary findings *Radiology* **221** 633–40