Available online at www.sciencedirect.com
ScienceDirect
Computer Vision and Image Understanding 109 (2008) 86–96
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu

Note

Robust information clustering incorporating spatial information for breast mass detection in digitized mammograms

Aize Cao, Qing Song *, Xulei Yang

School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore

Received 25 February 2005; accepted 6 July 2007
Available online 1 September 2007

Abstract

In this paper, we investigate a robust information clustering (RIC) algorithm incorporating spatial information for breast mass detection in digitized mammograms. The detection system employs RIC algorithm based on the raw region of interest (ROI) extracted from global mammogram by two steps of adaptive thresholding. Pixels on the fuzzy margin of a mass and noisy data were identified by RIC through the minimax optimization of mutual information. The memberships of the identified pixels (outliers) were recalculated by incorporating spatial distance information that takes into account of the influence of a neighborhood of 3×3 window. The algorithm is robust in the sense that both peak and valley of image intensity histogram are estimated and the pixels corresponding to valley in the histogram are clustered adaptively to image content. The purpose of the detection system is to locate the suspicious regions of mass candidates in the mammograms which will be further examined by other diagnostic techniques or by radiologists. The proposed method has been verified with 60 mammograms in the MiniMIAS database. The experimental results show that the detection system has a sensitivity of 90.7% at 2.57 false positives (FPs) per image.

Keywords: Robust information clustering; Minimax optimization of mutual information; Spatial information

1. Introduction

The cancerous tumors are hard to be visually detected when they are embedded in or camouflaged by varying densities of parenchymal tissues, particularly, the dense parenchymal breast tissues. It is a challenging task to detect masses within such background tissues in screening mammography, which has been recommended as the most effective method for the early detection of breast cancer. In addition, masses with various sizes and shapes may fail to generate a template that presents the common geometric properties of tumors. A lot of research work based on various theories have been carried out to tackle the problems of computerized mass detection. Doi et al. [4,33] have developed several methods for automatic detection of masses in mammograms. One of the methods employs a nonlinear bilateral-subtraction technique that locates the initial mass candidates by using multiple subtracted images. An artificial neural network was used to reduce false-positive detection resulting from the bilateral subtraction based on various features. In classification, the main concern of a mass in question to be malignant was based on the degree of spiculation.

Other methods of computerized detection and analysis include the earlier work done by Li et al. [5]. They proposed a Markov random field approach that lies in the category of region based algorithm to do breast mass detection. The algorithm was reported to have achieved 90% sensitivity with 2 FPs/image. Perick et al. [16] proposed a two-stage adaptive density-weighted contrast enhancement (DWCE) filter in conjunction with a Laplacian-of-Gaussian edge detector for mass detection. They reported 96% detection accuracy at 4.5 FPs/image for 25 mammograms by using a set of morphological features. Texture features based on gray-level co-occurrence matrices

* Corresponding author.
E-mail address: eqsong@ntu.edu.sg (Q. Song).

1077-3142/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2007.07.004

were studied later for a dataset of 168 cases [17]. A detection accuracy of 80% was achieved at 2.3 FPs/image. Mudigonda and Rangayyan [18] proposed to employ Gaussian smoothing and sub-sampling operations as preprocessing steps in mass detection. The mass portions are segmented by establishing intensity links from the central portions of masses into the surrounding areas. A sensitivity of 81% was achieved at 2.2 FPs/image for mass versus normal tissue. Malignant mass versus benign case classification resulted in $Az = 0.9$ under ROC curve for 26 masses. Gradient and texture analysis was used in [19] to classify masses into benign and malignant cases. Some other methods were proposed for the detection of spiculated masses due to their high likelihood of malignancy [14,30].

Image segmentation is the fundamental task in these computer aided detection systems. It is indeed true that the image segmentation needs robust algorithms. The robustness means that the performance of an algorithm should be relatively stable for small deviations from the assumed model, or noise and outliers [29]. Recent years have seen increasing interests in developing robust clustering algorithms for the applications in computer vision [7,6], image segmentation [32,3,12] and pattern recognition [1]. Many robust algorithms have been proposed for fuzzy clustering and a comprehensive review can be found in [29]. Dave [28] proposed robust fuzzy clustering algorithm by introducing the concept of noise cluster in addition to c clusters. A noise prototype is defined such that it is equidistant to all the points in the data set, i.e. the noise cluster is determined based on distance measure. In the same year, Krishnapuram [27] published a possibilistic approach for c-means clustering. The membership of each data is a possibilistic membership. Many other robust clustering algorithms can be found in the literature, such as [26,13]. Wu et al. [9] proposed a new metric to replace the Euclidean norm in c-means clustering procedure which is robust to noisy data and outliers. The more recent work of robust clustering is presented in [8].

In this paper, a robust information clustering (RIC) algorithm incorporating spatial information is proposed for mammographic mass detection as the help of screening mammography. Pixels on the fuzzy margin of a mass and noisy data are identified by the RIC based on the minimax optimization of the mutual information. Similar to the general statistical theory in [31,22], the RIC algorithm is considered as a learning machine problem with the limited training data rather than based on the infinite number of training data. It is a two-phase minimax optimization procedure. Through a minimizing mutual information scheme and starting from one, the number of clusters is increased gradually, which minimizes the free energy function from high to low temperature. Starting from an initial single cluster of all input patterns, a new cluster is added each time as the temperature is lowered to a critical point. The robust estimation phase is carried out to maximize the mutual information and estimate the outliers. The algorithm is robust in the sense that both peak and valley of image intensity histogram are estimated simultaneously.

The detection system employs a bottom-up image interpretation strategy that includes three major steps: ROI extraction, RIC segmentation incorporating spatial information and false positive (FP) reduction. The raw regions of interest (ROIs) were first extracted from the global mammograms by two steps of adaptive thresholding. Then RIC algorithm was applied for further segmentation. In the step of FP reduction, both area and texture features were calculated for the ROIs in the suspicious cluster resulted from the segmentation step, and support vector machine (SVM) [31] was used as the classifier to further reduce the number of FPs.

Instead of constructing an artificial system as a high-tech radiologist, the objective of this study intends to serve as the first vision in the process. It aims to locate the suspicious regions of mass candidates in the mammograms that will be further examined by other diagnostic techniques or by radiologists. The RIC clustering algorithm was used here as a technique to partition the entire ROI into distinctive meaningful regions to isolate the mass candidates from their background tissues based on the statistical probability distribution of intensity feature. Spatial information was incorporated during the process of outliers partitions. Experimental results show that the proposed RIC algorithm incorporating spatial information is effective in mammographic mass detection.

2. Method

2.1. Raw ROI extraction

The RIC was applied within certain raw ROI. In this study, two steps of adaptive thresholding were used for ROI extraction. There are two main areas in a mammographic image: breast region and non-breast region. In general, the non-breast region of the digitized magmmogram has a very low intensity level and a maximum peak in the histogram. The intensities of the breast region, however, may vary with different patients and depend on the mammography conditions. It should be noted that the background objects with high intensity can affect the extraction result of ROI. Therefore, the breast region was first extracted from digitized mammograms by adaptive thresholding. The level of threshold for the breast region is determined as the mean of the intensities within the window around a pixel. Considering the processing speed, a window of 4×4 was chosen instead of each pixel. Some mammograms in MiniMIAS database contain background objects with intensity higher than the breast region, or even higher than mass region, as shown in Fig. 1. They may deteriorate the results of raw ROI extraction and final segmentation if they are segmented together with the breast region. Due to the breast region is the major region in the mammogram that occupies the largest area compared

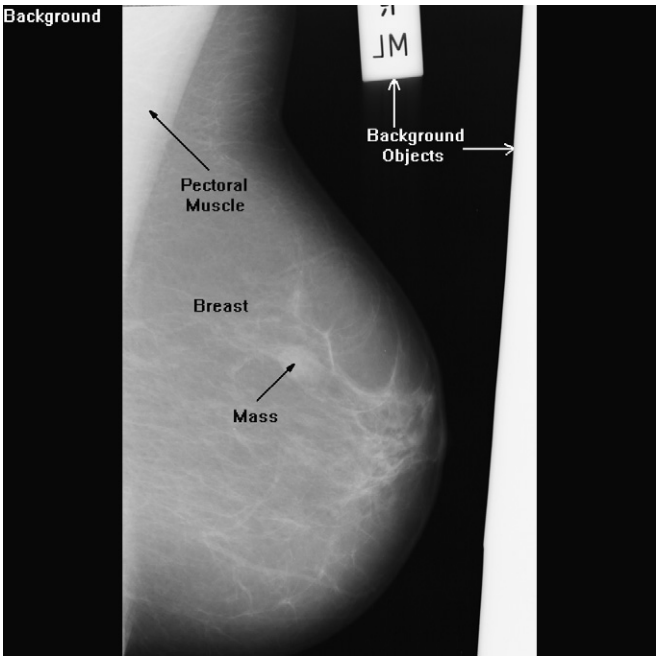

Fig. 1. Illustration of components in a mammogram. There are two major areas in a mammogram, breast region and non-breast region. The breast region contains pectoral muscle, breast tissues or mass. The non-breast region contains dark background and background objects, as indicated in the figure.

with the regions of other segmented objects, the segmented region with largest area was selected as breast region.

Based on breast region extraction, the second step of adaptive thresholding was carried out for the extraction of ROI. The threshold for each pixel in the breast region was computed as $\text{Threshold} = \text{mean} + \frac{\text{std}}{\text{factor}}$ [11], where 'std' represents the standard deviation. The 'mean' and standard deviation were computed for all the pixels in the breast region. A 'factor' of 3.0 is chosen empirically. There are a few sub-regions after the second step of adaptive thresholding, including some isolated points and very small regions, which may not be part of a mass. These subregions were dilated around their boundary for eight nearest neighborhoods and then went through an area filter. The sub-regions with areas larger than certain threshold were remained as part of the interested region for further processing. A raw ROI was formulated finally (see Fig. 2), based on which RIC was going to be used for further segmentation.

2.2. Robust information clustering (RIC) algorithm

The ROI normally contains subregions of fat, parenchymal tissues or masses. Each region corresponds to one group of intensity level. The mean intensity of each group varies according to their surrounding tissues. The proposed method to segment mass region is based on the assumption that the mean intensity of mass region is brighter than its surrounding tissues [5]. It should be noted that the fuzzy characters make it hard to find the boundary of a mass. Additionally, artifacts or noisy intensities may deviate the target patterns from the optimal positions, therefore, robust clustering algorithm is proposed here for the segmentation of mammograms.

In image segmentation, it requires that a decision should be made about the type of each pixel, i.e. which feature groups (or code vectors) the pixel belongs to. It is an observation at each pixel and describes the relation between source vector $\mathbf{x}_i \in \mathbf{X}$, $i = 1, \ldots, l$ and pattern class $\mathbf{w}_k \in \mathbf{W}$, $k = 1, \ldots, K$ (scalar if intensity only), which can also be stated as partitioning a given set of pixels into subgroups, each of which should be as homogeneous as possible within group, and should be as inhomogeneous as possible between groups. K is the maximum number of pattern partitions. It is an unsupervised clustering problem. The dissimilarity measure between the information source and target patterns is stated by the least mean square distance in Euclidean space

$$\min_{k,i} d(\mathbf{x}_i, \mathbf{w}_k) = \frac{1}{2}\|\mathbf{w}_k - \mathbf{x}_i\|^2 \tag{1}$$

Based on the least mean square measure, some of the popular clustering algorithms have been developed like the basic k-means, fuzzy c-means and improved robust versions,

Fig. 2. Raw ROI extraction. It is the case of mdb111 in MiniMIAS that contains an asymmetry malignant mass surrounded by dense glandular breast tissues. (a) Original image. The circled region contains mass. In addition to the breast region, there are two background objects. (b) Breast region, obtained by first step of thresholding. (c) Raw ROI, extracted by second step of adaptive thresholding.

like robust noise [29] and possibility clustering algorithm [15]. According to Huber [20], a robust procedure can be characterized as first, it should have a reasonably good accuracy at the assumed model; second, small deviations from the model assumptions should impair the performance only by a small amount; and third, larger deviations from the model assumptions should not cause a catastrophe. The first requirement is the consideration whether data are clean. It plays great importance in statistics and engineering since systems have to be designed such that their performance yields accurate estimation. The second property gains much attentions since the data is not clean in engineering practice, i.e. small deviations should not have a significant negative effect on the performance. The third requirement is related to the concept of breakdown point. According to these properties, the most important one is the robustness to initialization, namely, stable performance. Another common scheme of robustness is to reject or ignore a subset of the input patterns by evaluating the membership function, which can be viewed as a robust weight function to phase out outliers. The robust performance of the fuzzy algorithms in [28] is explained in a sense that it involves the memberships of all input patterns of a given cluster, but not the membership of a particular pattern in all clusters.

Suppose the probability distributions of $\mathbf{x}_i$ and $\mathbf{w}_k$ are $p(\mathbf{x}_i)$ and $p(\mathbf{w}_k)$, respectively, and conditional probability distributions are $p(\mathbf{x}_i|\mathbf{w}_k)$ and $p(\mathbf{w}_k|\mathbf{x}_i)$. For the two basic concepts in information theory, i.e. rate distortion and channel capacity, let us first consider the rate distortion function that is defined as [24]

$$R(D(p^*(\mathbf{X}))) = \min_{p(\mathbf{W}|\mathbf{X})} I(p^*(\mathbf{X}), p(\mathbf{W}|\mathbf{X})) \tag{2}$$

with the constraint

$$\sum_{i=1}^{l}\sum_{k=1}^{K} p^*(\mathbf{x}_i)p(\mathbf{w}_k|\mathbf{x}_i)d(\mathbf{w}_k,\mathbf{x}_i) \leqslant \sum_{i=1}^{l}\sum_{k=1}^{K} p^*(\mathbf{x}_i)\bar{p}(\mathbf{w}_k|\mathbf{x}_i)d(\mathbf{w}_k,\mathbf{x}_i) \tag{3}$$

where the mutual information $I(p^*(\mathbf{X}),p(\mathbf{W}|\mathbf{X}))$ is expressed as

$$I(p^*(\mathbf{X}),p(\mathbf{W}|\mathbf{X})) = p(\mathbf{W}|\mathbf{X})\left[\sum_{i=1}^{l}\sum_{k=1}^{K} p^*(\mathbf{x}_i)p(\mathbf{w}_k|\mathbf{x}_i)\ln\frac{p(\mathbf{w}_k|\mathbf{x}_i)}{\sum_{i=1}^{l}p(\mathbf{w}_k|\mathbf{x}_i)p^*(\mathbf{x}_i)}\right] \tag{4}$$

In these equations, $p^*(\mathbf{X})$ is the fixed unconditional prepobability, $p(\mathbf{W}|\mathbf{X})$ is the titled distribution, as

$$p(\mathbf{w}_k|\mathbf{x}_i) = \frac{p(\mathbf{w}_k)\exp(-d(\mathbf{w}_k,\mathbf{x}_i)/T)}{\sum_{k=1}^{K}p(\mathbf{w}_k)\exp(-d(\mathbf{w}_k,\mathbf{x}_i)/T)} \tag{5}$$

and $\bar{p}(\mathbf{w}_k|\mathbf{x}_i)$ is the titled distribution that achieves a minimum value in $R(D(p^*(\mathbf{X})))$. The rate distortion function is usually investigated in term of a parameter $s = -1/T$ with $T \in (0,\infty)$. It can be further formulated as a constraint optimization problem

$$R(D(p^*(\mathbf{X}))) = sD(p^*(\mathbf{X})) + \min_{p(\mathbf{W}|\mathbf{X})} F(p^*(\mathbf{X}), p(\mathbf{W}|\mathbf{X})) \tag{6}$$

where $F(p^*(\mathbf{X}),p(\mathbf{W}|\mathbf{X}))$ is the DA clustering objective function as (Rose [10])

$$F(p^*(\mathbf{X}), p(\mathbf{W}|\mathbf{X})) = I(p^*(\mathbf{X}), p(\mathbf{W}|\mathbf{X})) - s\left(\sum_{i=1}^{l}\sum_{k=1}^{K} p^*(\mathbf{x}_i)p(\mathbf{w}_k|\mathbf{x}_i)d(\mathbf{w}_k,\mathbf{x}_i) - D(p^*(\mathbf{X}))\right) \tag{7}$$

From these deduction we can see that minimizing the mutual information leads to deterministic annealing.

To gain robustness, we focus on the property of density estimation $p(\mathbf{x}_i) \in p(\mathbf{X})$. Instead of the fixed equal distribution $p^*(\mathbf{x}_i) = 1/l$ in the minimization of mutual information (DA) algorithm, maximization of the mutual information is considered to evaluate different distributions of all input patterns of a given cluster, in particular, seeking the outliers with $p(\mathbf{x}_i) = 0$ and eliminating them from the data set for the final data partitions. According to information theory, the mutual information can also be written in the following form

$$I(\mathbf{X},\mathbf{W}) = \max_{p(\mathbf{x}_i|\mathbf{w}_k)} \sum_{i=1}^{l}\sum_{k=1}^{K} p(\mathbf{x}_i)p^*(\mathbf{w}_k|\mathbf{x}_i)\ln\frac{p(\mathbf{x}_i|\mathbf{w}_k)}{p(\mathbf{x}_i)} \tag{8}$$

The capacity of a channel (or prototypes $\mathbf{w}_k$) expresses the maximum rate at which information can be reliably conveyed by the channel, defined as [23]

$$C = \max_{p(\mathbf{x}_i)} I(\mathbf{X},\mathbf{W}) \tag{9}$$

$$= \max_{p(\mathbf{x}_i)}\max_{p(\mathbf{x}_i|\mathbf{w}_k)} \sum_{i=1}^{l}\sum_{k=1}^{K} p(\mathbf{x}_i)p^*(\mathbf{w}_k|\mathbf{x}_i)\ln\frac{p(\mathbf{x}_i|\mathbf{w}_k)}{p(\mathbf{x}_i)} \tag{10}$$

According to Theorem 1 in [23], for a given $p^*(\mathbf{w}_k|\mathbf{x}_i)$, the capacity is maximized by

$$p(\mathbf{x}_i) = \frac{\exp\left(\sum_{k=1}^{K}p^*(\mathbf{w}_k|\mathbf{x}_i)\ln p(\mathbf{x}_i|\mathbf{w}_k)\right)}{\sum_{i=1}^{l}\exp\left(\sum_{k=1}^{K}p^*(\mathbf{w}_k|\mathbf{x}_i)\ln p(\mathbf{x}_i|\mathbf{w}_k)\right)}, \tag{11}$$

where $p^*(\mathbf{w}_k|\mathbf{x}_i)$ is the titled (or conditional probability) distribution obtained in the DA step, and the backward transition probability $p(\mathbf{x}_i|\mathbf{w}_k)$ is obtained through the Bayes formula

$$p(\mathbf{x}_i|\mathbf{w}_k) = \frac{p(\mathbf{x}_i)p^*(\mathbf{w}_k|\mathbf{x}_i)}{\sum_{i=1}^{l}p(\mathbf{x}_i)p^*(\mathbf{w}_k|\mathbf{x}_i)} \tag{12}$$

Note that the mutual information $I(\mathbf{X},W)$ is not negative. However, the individual item in the sum of the capacity can be negative. If the ith pattern $\mathbf{x}_i$ is taken into account and $p^*(\mathbf{w}_k|\mathbf{x}_i) < \sum_{i=1}^{l}p(\mathbf{x}_i)p^*(\mathbf{w}_k|\mathbf{x}_i)$, then the probability of the kth code vector (mass center) is decreased by the observed pattern and gives a negative information about pattern $\mathbf{x}_i$. Then the particular input pattern may be considered as an unreliable pattern (outlier) and its negative effect must be offset by other input patterns.

Therefore, the maximization of the mutual information provides a robust density estimation of the noisy pattern (outlier) in term that the average information is over all clusters and input patterns. The robust density estimation and optimization is now to maximize the mutual information against $p(\mathbf{x}_i)$ and $p(\mathbf{x}_i|\mathbf{w}_k)$, i.e., the maximization of mutual information in (9) is a double maximization for *a priori* $p(\mathbf{x}_i)$ via $p(\mathbf{x}_i|\mathbf{w}_k)$ in (10) [23]. For any value of i, if $p(\mathbf{x}_i|\mathbf{w}_k) = 0$, then $p(\mathbf{x}_i)$ should be set equal to zero in order to obtain the maximum, such that a corresponding training pattern $\mathbf{x}_i$ can be deleted and dropped from further consideration in the optimization procedure as an outlier.

The RIC is basically a two-phase iteration clustering algorithm, i.e. deterministic annealing clustering to minimize the mutual information through (2) and (3), and the robust estimation of the outliers to maximize mutual information through (9). For more details, comprehensive illustration and proof are in [21,23]. The implementation of RIC is summarized as follows.

1. Fixed $p^*(x_i) = 1/l$, do DA clustering [10], obtain the optimal titled distribution of $p(\mathbf{w}_k|\mathbf{x}_i)$.
2. Let $p^*(\mathbf{w}_k|\mathbf{x}_i) = p(\mathbf{w}_k|\mathbf{x}_i)$, initialize the prior probability of $p(\mathbf{x}_i) = 1/l$, where l is the number of the input patterns.
3. Set
$$c_i = \exp\left(\sum_{k=1}^{K} p^*(\mathbf{w}_k|\mathbf{x}_i) \ln \frac{p^*(\mathbf{w}_k|\mathbf{x}_i)}{\sum_{i=1}^{l} p(\mathbf{x}_i) p^*(\mathbf{w}_k|\mathbf{x}_i)}\right) \tag{13}$$
4. If
$$\ln \sum_{i=1}^{l} p(\mathbf{x}_i) c_i - \ln \max_{i=1,\ldots,l} c_i < \varepsilon \tag{14}$$
go to step 5, where $\varepsilon > 0$, otherwise, set
$$p(\mathbf{x}_i) = p(\mathbf{x}_i) \frac{c_i}{\sum_{i=1}^{l} p(\mathbf{x}_i) c_i} \tag{15}$$
go to step 3.
5. The input patterns with $p(\mathbf{x}_i) = 0$ are estimated as outliers.
6. Do DA clustering for input patterns of all $\mathbf{x}_i$ with $p(\mathbf{x}_i) > 0$, stop.

The $\varepsilon > 0$ is an accuracy control parameter. At the final point of estimation, pixels will be deemed as outliers with $p(\mathbf{x}_i) = 0$. Until now all the pixels in ROI belong to two clusters: outliers and non-outliers. Label zero was assigned to pixels that belong to outliers, and label one was assigned to pixels that belong to the non-outliers. Then the outliers were going to be partitioned by incorporating spatial information.

2.3. Outlier partitions by incorporating spatial information

Since image data has inherent spatial ordering, the spatial information can influence their clustering results, particularly for noisy data and ambiguous data around an object, which are called outliers that were identified in the RIC step. Therefore, we introduce the spatial context [2] into the outliers clustering based on the concept that the prototypes after RIC clustering are optimized, i.e. the cluster centers are optimal that are fixed for the further outliers clustering. The objective distances were calculated from the identified outliers to the fixed prototypes (centroids).

Consider a 3×3 window centered at each outlier. Suppose the image is $M \times N$, the image patterns $\mathbf{x}_i \in \mathbf{X}$ are converted into $M \times N$ matrix and the location of an outlier in the image is (m,n), where $m \in (1,M)$ and $n \in (1,N)$. The distance of the outlier to kth cluster center $\mathbf{w}_k$ is updated as
$$\widehat{d}_{k,m,n} = \frac{1}{8} \sum_{h_1=-1}^{1} \sum_{h_2=-1}^{1} \left(d^2_{k,m,n} \lambda^{m,n}_{h_1,h_2} + d^2_{k,m+h_1,n+h_2} \left(1 - \lambda^{m,n}_{h_1,h_2}\right)\right) \tag{16}$$
with $(h_1,h_2) \neq (0,0)$, and $d_{k,m,n}$ is the distance of the outlier to kth cluster center, $d_{k,m+h_1,n+h_2}$ is the distance of the neighbor located at $(m+h_1, n+h_2)$ to kth cluster center. The $\lambda^{m,n}_{h_1,h_2}$ is a weighted function as
$$\lambda^{m,n}_{h_1,h_2} = \frac{1}{1 + \exp(-(\delta_{(m,n),(m+h_1,n+h_2)} - \mu)/\sigma)} \tag{17}$$
where parameter δ is the distance between the outlier and its neighbors. For the outlier at location of (m,n), $\delta_{(m,n),(m+h_1,n+h_2)}$ gives the Euclidean distance between feature vectors $\mathbf{x}_{m,n}$ and its neighbor $\mathbf{x}_{m+h_1,n+h_2}$, i.e.
$$\delta_{(m,n),(m+h_1,n+h_2)} = (\mathbf{x}_{m,n} - \mathbf{x}_{m+h_1,n+h_2})^{\mathrm{T}} (\mathbf{x}_{m,n} - \mathbf{x}_{m+h_1,n+h_2}) \tag{18}$$
the feature vector $\mathbf{x}_{m,n}$ is updated by
$$\widehat{\mathbf{x}}_{m,n} = \frac{1}{8} \sum_{h_1=-1}^{1} \sum_{h_2=-1}^{1} (\lambda^{m,n}_{h_1,h_2} \mathbf{x}_{m,n} + (1 - \lambda^{m,n}_{h_1,h_2}) \mathbf{x}_{m+h_1,n+h_2}) \tag{19}$$
The parameter μ is defined as
$$\mu = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{1}{8} \left(\sum_{h_1=-1}^{1} \sum_{h_2=-1}^{1} \delta_{(m,n),(m+h_1,n+h_2)} \right) \tag{20}$$
with $(h_1,h_2) \neq (0,0)$.

From these definition we can see that if the feature vector of an outlier is close to its neighbors in feature space, then $\delta_{(m,n),(m+h_1,n+h_2)}$ is small. The distance of an outlier is greatly influenced by its neighbors. Otherwise, it is largely independent of its neighbors. The implementation of outliers clustering is given as follows:

1. Obtain outliers and the optimal cluster prototypes (or centers) from RIC.
2. Convert the feature vector into $M \times N$ matrix that corresponds to the image formulation. Set the maximum iteration number L and the value of parameter σ.
3. Calculate $\delta_{(m,n),(m+h_1,n+h_2)}$, μ, $\lambda^{m,n}_{h_1,h_2}$ according to (18), (20) and (17).

4. Update the weighted distance $\widehat{d}_{k,m,n}$ and feature vector $\widehat{\mathbf{x}}_{m,n}$ according to (16) and (19).
5. Go to step 3 if the iteration does not reach the maximum number L.
6. Assign labels to outliers with minimum distance $\widehat{d}_{k,m,n}$, stop.

2.4. Feature extraction

Statistical image texture analysis, which has been described by Raeth et al. [34], was used to improve the ability of distinguishing normal from abnormal cases. One area and nine texture features were extracted within ROIs. Among these texture features, two are the first order texture features, i.e. the mean of gray level and gray variance that describe the hight and shape of the gray level histogram. The other seven features are second order texture features. Two features were calculated from the first order gradient distribution: mean gradient and variance of gradient. Five features were calculated from the gray level co-occurrence matrix: namely energy, inertia, entropy, homogeneity and correlation. The co-occurrence matrices were taken in the North–South, and East–West directions at pixels spacing of 1.

Based on the feature extraction, support vector machine (SVM) [31] was used as classifier to further classify the ROI into normal and abnormal case as a help of FP reduction.

3. Experimental results

In this section, the experimental results for breast mass detection are presented. A testing data set of 60 mammograms were used to test and evaluate the performance of the proposed algorithms. Among these mammograms, six contain normal breast tissue only, and 54 have at least one mass surrounded by glandular and dense glandular breast tissues with 30 benign and 24 malignant masses.

3.1. The performance of RIC for mammographic image segmentation

An example of RIC for outliers identification and image segmentation for mammographic image is provided in the following paragraph and shown in Fig. 3. The maximum cluster number is chosen as two. There are three true clusters after RIC clustering, namely, normal breast tissues, mass body, and the outlier. Fig. 3(a) is the raw ROI that contains mass. Its corresponding histogram is illustrated in Fig. 3(d). During RIC clustering, the whole image data was first partitioned into two clusters: outliers and non-outliers. As shown in Fig. 3(b), the estimated outliers are represented by intensity value of zero, i.e. dark points in the image, where bright points are the normal data. It can be seen that the outliers represent pixels on the boundary around the mass region and some disconnected noisy data. The three clusters after RIC clustering, i.e. outliers, normal breast tissue and mass region, are illustrated in

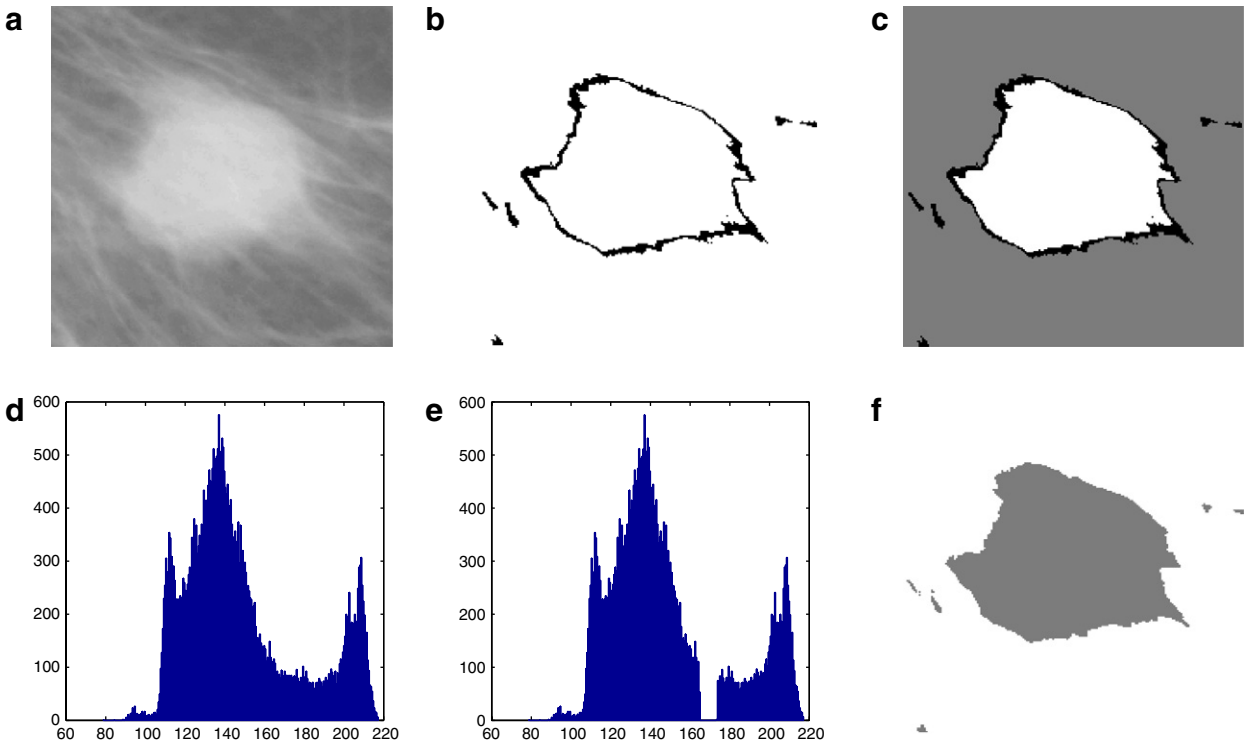

Fig. 3. The performance of RIC on outliers identification and image segmentation for mammographic image. The number of cluster is two. (a) Raw ROI that contains a case of mass. (b) Outliers (dark points) that are estimated by RIC. (c) Segmentation result by RIC, where dark points represent outliers. (d). Histogram of the original ROI. (e) Histogram of the original image with outliers removed. (f) Segmentation result by DA.

Fig. 3(c). The outlier-free image data were partitioned into two clusters, each of which has distinctive intensity. The histogram of the original image with outliers removed by RIC algorithm is illustrated in Fig. 3(e). Fig. 3(f) shows the segmentation result by DA.

Fig. 3(e) demonstrates that a portion of gray levels in the proximity of the valley between the two peeks in histogram are identified automatically by RIC, rather than the one by the standard clustering method. Because of the estimation of outliers, the suspicious cluster that contains mass has less number of subregions than that obtained by standard DA. As a result, it reduced the complexity of further analysis. Certain criteria is necessary to relocated the identified outliers into the two clusters optimally.

3.2. Segmentation result for outlier partitions by incorporating spatial information

We further partitioned the estimated outliers into different clusters incorporating the spatial information. For the purpose of comparison, the mammograms were segmented not only by the proposed spatial RIC algorithm, they were also segmented by fuzzy *c*-means (FCM), DA, and one of the most popular robust clustering algorithms, named possibilistic *c*-means (PCM) clustering algorithm [29] as illustrated in Fig. 4. The raw ROIs are shown in the row Fig. 4(a) that contain mass. The segmentation results are given in row Fig. 4(b)–(e), respectively. Pixels in the same cluster are represented by the same intensity value. The

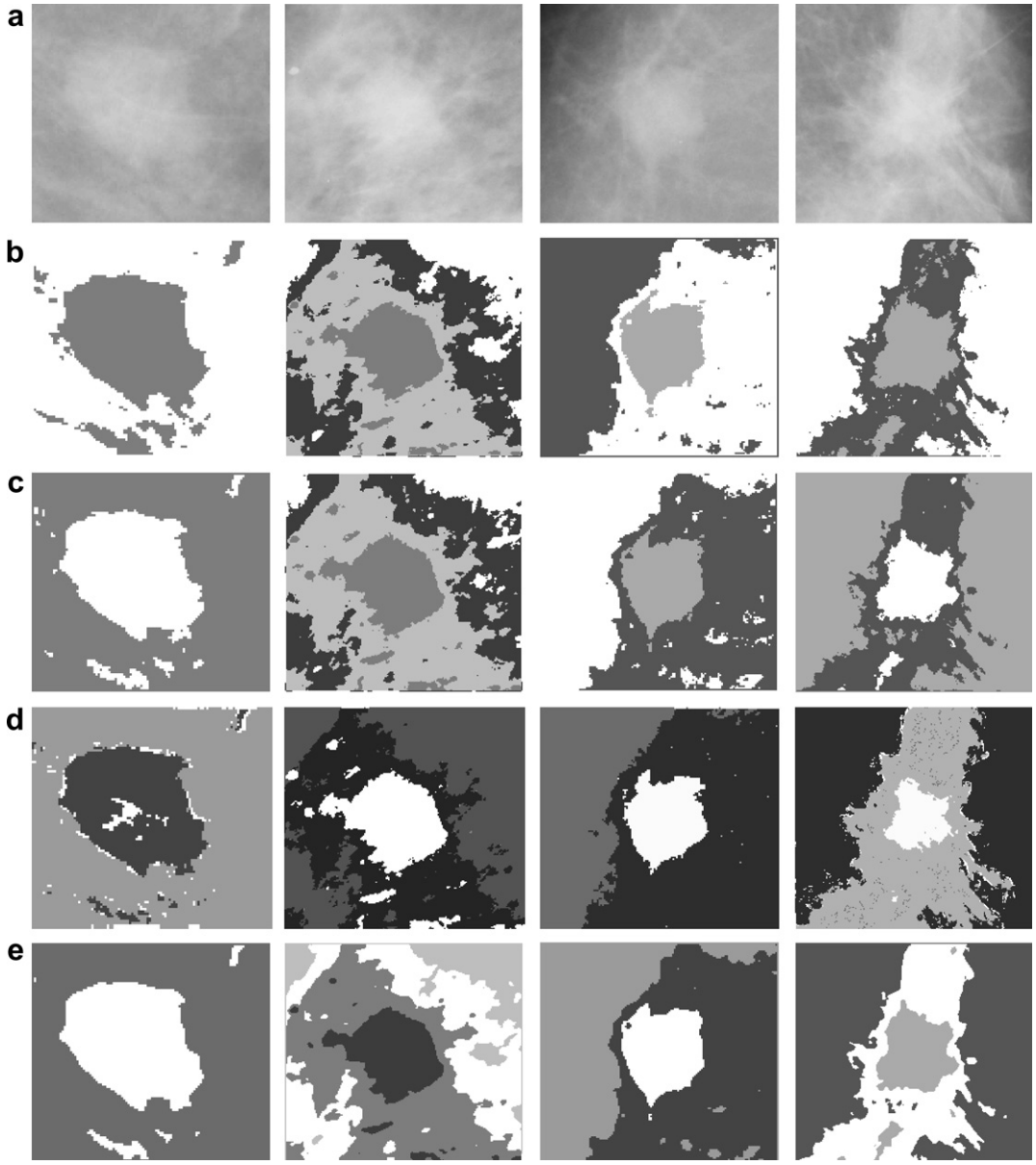

Fig. 4. Segmentation results comparison. Pixels in the same cluster represented by the same intensity. The number of cluster is two for the segmentation results in first column, four for the second column, three for the third and fourth columns. (a) The raw ROI in mammograms. (b) Segmentation results by FCM. (c) Segmentation results by DA. (d) Segmentation results by PCM. (e) Segmentation results by RIC incorporating spatial information.

same location in ROI may be represented by different intensities for different segmentation since the labels are assigned randomly at the end of clustering. It can be seen that the segmentation results by RIC incorporating with spatial information are able to find optimal boundary of mass with less noisy regions compared with FCM, DA and PCM methods.

3.3. FP reduction

FP reduction was performed during the two steps of segmentation, i.e. adaptive thresholding and RIC. Since some subregions were segmented together with ROIs and masses, which may lead further analysis more complex. Thus, following each step of segmentation, region growing was performed for each subregion that belongs to the suspicious group, and the area of each subregion was calculated. In the first step of adaptive thresholding, the object with the largest area was selected as breast region since it occupies the major structure in the mammogram. The raw ROI was extracted within the segmented breast structure, in which two main regions that may contain mass have high intensity: the pectoral muscle and the glandular breast tissues. The pectoral muscle represents a predominant density region in most mammograms with medio-lateral oblique (MLO) views. The segmentation results will be affected or biased if it is included in ROI for intensity-based image processing methods. Ferrari et al. [25] proposed a new method for the identification of the pectoral muscle based upon a multi-resolution technique using Gabor wavelets. In this study, the pectoral muscle was identified as the largest object among the first ten subregions with area larger than 200 pixels according to the searching method. Raw ROI was thus formulated based on the pectoral muscle and noisy points removal.

After the segmentation by RIC incorporating spatial information, all the pixels in ROI were grouped into K clusters. The mean intensity of each cluster is different, where the mean intensity of the cluster that contain mass has the highest value (according to assumption), which is called suspicious cluster. There were some subregions or isolated points in the suspicious cluster. The final abnormal cases were identified by SVM classifier based on both area and texture features extracted from the ROIs in the suspicious cluster, where radius based function(RBF) was used as the kernel function.

4. Detection results

The purpose of final FP reduction is to increase the detection rate. For the complicated diagnosis system, it is able to decrease the computational complexity and allows more computationally intensive classifiers to be applied to as few objects as possible. In this step, one area and nine texture features were extracted from each suspicious ROI in the selected cluster, and SVM was applied as the classifier to classify the these ROIs into normal and abnormal cases. Corresponding to the target of mammography screening, the purpose of this study is to locate the suspicious areas that may contain mass,

Table 1
The detection results of the number of suspicious regions per image for RIC, DA, FCM and PCM based on the same raw ROI

Mammogram	RIC	DA	FCM	PCM	Mammogram	RIC	DA	FCM	PCM
mdb001	1	5	5	1	mdb117	3	3	3	3
mdb002	1	1	1	1	mdb120	2	3	4	1
mdb013	7	7	7	8	mdb121	2	1	2	2
mdb015	8	12	12	6	mdb127	4	7	7	18
mdb017	3	4	4	4	mdb145	3	5	4	8
mdb019	7	6	11	7	mdb165	1	1	1	1
mdb021	7	8	7	3	mdb170	2	2	3	2
mdb023	4	8	7	5	mdb171	1	2	2	1
mdb030	10	15	14	10	mdb175	1	2	1	6
mdb032	10	17	10	7	mdb178	1	2	2	3
mdb058	4	7	7	5	mdb179	1	2	1	1
mdb063	2	3	2	6	mdb181	1	1	1	1
mdb072	2	2	2	3	mdb186	1	4	4	1
mdb081	1	1	1	1	mdb188	2	1	1	1
mdb083	9	6	10	10	mdb190	1	1	1	1
mdb090	5	5	6	0	mdb191	3	5	6	0
mdb099	14	9	10	8	mdb198	2	5	5	7
mdb102	3	3	3	8	mdb199	7	5	6	7
mdb104	5	5	4	3	mdb202	1	7	7	7
mdb105	4	3	4	1	mdb207	1	1	1	1
mdb107	6	7	7	5	mdb244	4	5	5	2
mdb110	3	3	3	3	mdb264	1	1	1	1
mdb111	2	3	3	2	mdb265	3	4	3	5
mdb115	1	1	2	1	mdb315	1	1	1	1
mdb163	2	5	5	2	mdb270	9	8	9	6
mdb193	2	8	7	5	mdb290	1	1	1	1
mdb124	2	3	2	3	mdb125	4	4	4	5

therefore, the detection results are evaluated by terms of sensitivity and the number of false positives per image.

For the comparison of the detection results, standard clustering methods of FCM and DA, as well as one of the most popular robust clustering algorithms PCM, were also implemented based on the same detection system as spatial RIC, i.e. the image segmentation by spatial RIC, DA, FCM and PCM were based on the same ROIs. By regulating the cluster number in image segmentation, the parameters of threshold and classifier in the step of raw ROI selection and abnormal case classification, different detection results were obtained. In the experiments, the numbers of cluster for RIC and PCM were regulated from two to three, and the numbers of cluster for DA and FCM were from two to six. The detection results for mammograms with mass are illustrated in Table 1.

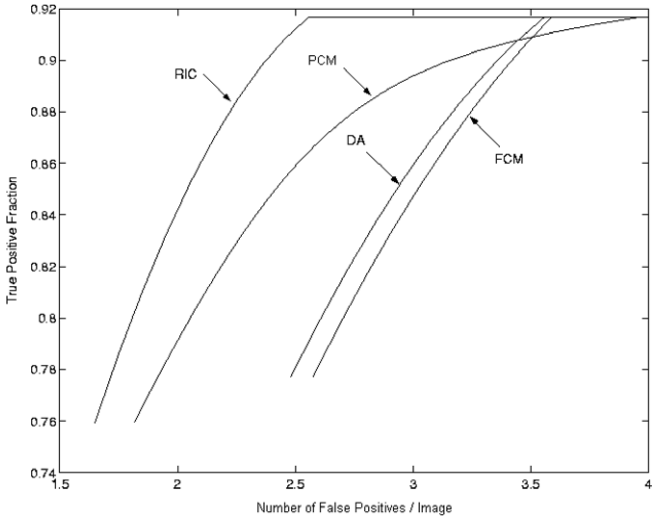

Fig. 5. FROC analysis for the detection systems based on RIC incorporating spatial information, DA, FCM and PCM.

Among the 54 mammograms, thirty masses are embedded in glandular breast tissues, and twenty-four masses are within dense glandular breast tissues. The proposed detection system incorporating RIC algorithm obtains a sensitivity of 90.7% at the false positive of 2.57 per image. The detection result of the detection system incorporating with DA is 3.56 FPs per image and 3.59 FPs per image for the incorporation of FCM. When incorporation of PCM, the detection result is 3.91.

Fig. 5 is the free-response ROC (FROC). From this figure, we can see that at the same TP fraction, the number of FPs by RIC is less than that of DA, FCM and PCM. When FP is under 3.5, the performance of RIC and PCM are better than DA and FCM, i.e. at the same number of FPs per image, the TP fraction by RIC and PCM are higher than that being produced by DA and FCM, where RIC outperformed PCM. When the number of false positives is larger than certain value, as 3.9 in this study, the TP fraction of these four methods are the same.

Because of the estimation of outliers, RIC can segment mass body from its background even when it is embedded in the dense glandular breast tissue with relatively less number of FPs compared with DA, FCM and PCM methods, i.e. RIC has robust capacity in data splitting and estimating the noisy points. In addition, due to the incorporation of spatial information, the detected results are of benefit for the further mass pattern analysis(i.e. benign or malignant), which is beyond the focus of this paper and can be discussed in further study. We also observed that the segmented mass region by DA and FCM includes large area of background breast tissues which leads to the number of FPs relatively less in some images. Based on the experimental analysis we can see that the detection system incorporating spatial RIC works more effective in false positive reduction than the systems based on DA, FCM and PCM.

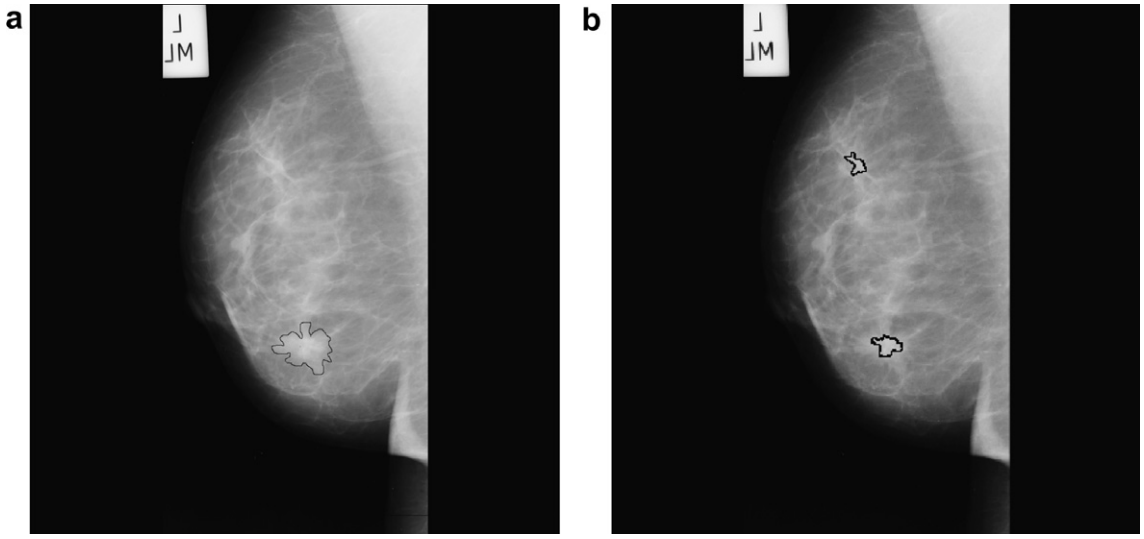

Fig. 6. Detection results for mdb181. (a) Original image. The circled region contains mass. (b) Detection result of RIC incorporating spatial information.

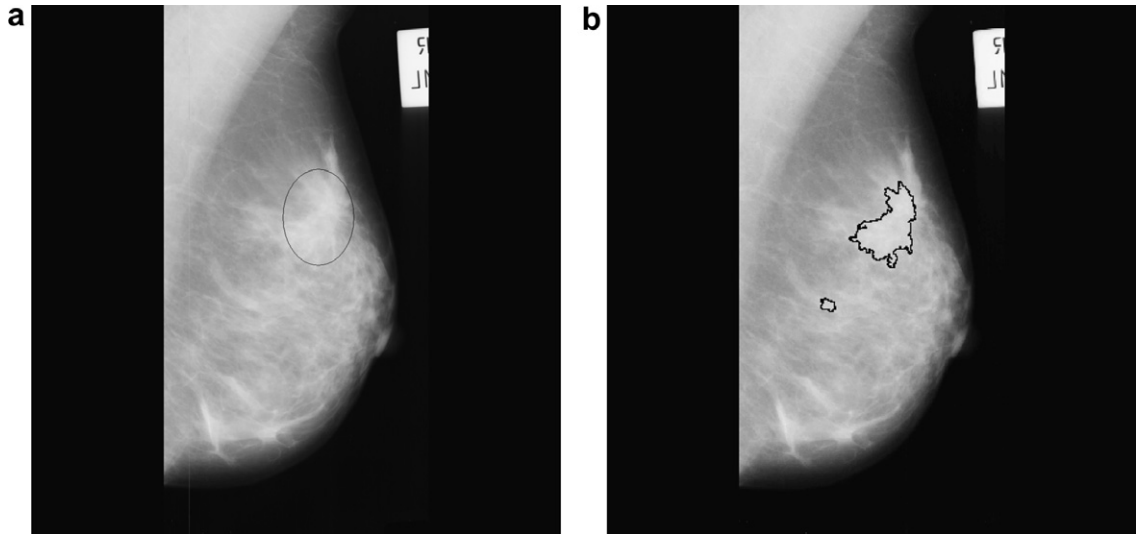

Fig. 7. Detection results for mdb198. (a) Original image. The circled region contains mass. (b) Detection result of RIC incorporating spatial information.

The detection results by RIC for four different kinds of masses embedded in glandular or dense glandular breast tissues are illustrated in Figs. 6 and 7. Fig. 6 shows the final FP reduction results of the case mdb181, which contains spiculated mass within glandular breast tissue. Fig. 7 is the detection results of the case mdb198 that has spiculated mass in dense breast tissue.

5. Conclusion

We have presented a RIC algorithm for mammographic mass detection. It can identify outliers and clusters the mammographic image into corresponding clusters by using feature space and spatial contextual information. Unlike the classical robust clustering algorithm, RIC has the capability to find outliers automatically by the minimax optimization of mutual information. In addition, the incorporation of spatial information with RIC algorithm reduces the number of noisy regions and results in optimal mass boundary. Experimental results show that the proposed algorithm works effective for mammographic image segmentation and detection.

References

[1] A.K. Jain, P.W. Duin, J. Mao, Statistical pattern recognition: a review, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (2000) 4–37.
[2] A.W.C. Liew, S.H. Leung, W.H. Lau, Fuzzy image clustering incorporating spatial continuity, in: IEE Proceedings-Vision, Image and Signal Processing 147 (2000) 185–192.
[3] D.L. Pham, J.L. Prince, An adaptive fuzzy c-means algorithm for image segmentation in the presence of intensive inhomogeneities, Pattern Recognition Letter 20 (1999) 57–68.
[4] F. Yin, M.L. Giger, K. Doi, C.E. Metz, C.J. Vyborny, R.A. Schmidt, Computerized detection of masses in digital mammograms: Analysis of bilateral subtraction images, Medical Physics 18 (1991) 955–963.
[5] H.D. Li, M. Kallergi, L.P. Clarke, V.K. Jain, R.A. Clark, Markov random field for tumor detection in digital mammography, IEEE Transactions on Medical Imaging 14 (3) (1995) 565–576.
[6] H. Frigui, R. Krishnapuram, A robust competitive clustering algorithm with applications in computer vision, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (1999) 450–465.
[7] J.M. Jolion, P. Meer, S. Bataouche, Robust clustering with applications in computer vision, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (8) (1991) 791–802.
[8] J.S. Zhang, Y.W. Leung, Robust clustering by pruning outliers, IEEE Transactions on Systems, Man, and Cybernetics, Part B 33 (6) (2003) 983–999.
[9] K.L. Wu, M.S. Yang, Alternative c-means clustering algorithms, Pattern Recognition 35 (2002) 2267–2278.
[10] K. Rose, Deterministic annealing for clustering, compression, classification, regression, and related optimization problems, Proceedings of the IEEE 86 (11) (1998) 2210–2239.
[11] K. Woods, Automated image analysis techniques for digital mammography. Ph.D. Dissertation, University of South Florida, December, 1994.
[12] L. Cinque, G. Foresti, L. Lombardi, A clustering fuzzy approach for image segmentation, Pattern Recognition 37 (9) (2004) 1797–1807.
[13] L.N.F. Ana, A.K. Jain, Robust data clustering, Proceedings on Computer Vision and Pattern Recognition 2 (2003) 128–133.
[14] M.A. Kupinski, M.L. Giger, Automated seeded lesion segmentation on digital mammograms, IEEE Transactions on Medical imaging 17 (1998) 510–517.
[15] M. Shen, K.L. Wu, A similarity-based robust clustering method, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (4) (2004) 434–448.
[16] N. Perick, H-P. Chan, B. Sahiner, D.T. Wei, An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection, IEEE Transactions on Medical Imaging 15 (1996) 59–67.
[17] N. Petrick, H.P. Chan, D. Wei, M.A. Helvie, B. Sahiner, D.D. Adler, Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification, Medical Physics 23 (1996) 1685–1696.
[18] N.R. Mudigonda, R.M. Rangayyan, Detection of breast masses in mammograms by density slicing and texture flow-field analysis, IEEE Transactions on Medical Imaging 20 (2001) 1215–1227.
[19] N.R. Mudigonda, R.M. Rangayyan, E.L. Desautels, Gradient and texture analysis for the classification of mammographic masses, IEEE Transactions on Medical Imaging 19 (10) (2000) 1032–1043.
[20] P.J. Huber, Robust Statistics, Wiley, New York, 1981.
[21] Q. Song, A robust information clustering algorithm, Neural Computation 17 (12) (2005) 2672–2698.

[22] Q. Song, W.J. Hu, W.F. Xie, Robust support vector machine for bullet hole image classification, IEEE Transactions on Systems, Man and Cybernetics, Part C 32 (2002) 440–448.
[23] R.E. Blahut, Computation of channel capacity and rate-distortion functions, IEEE Transactions on Information Theory IT-18 (4) (1972) 460–473.
[24] R.E. Blahut, Principle and Practice of Information Theory, Addison-Wesley, 1988.
[25] R.J. Ferrari, R.M. Rangayyan, J.E.L. Desautels, R.A. Borges, F. Frère, Automatic identification of the pectoral muscle in mammograms, IEEE Transactions on Medical Imaging 23 (2) (2004) 232–245.
[26] R.J. Hathaway, J.C. Bezdek, Fuzzy c-means clustering of incomplete data, IEEE Transactions on Systems, Man, and Cybernetics Part B 31 (2001) 735–744.
[27] R. Krishnapuram, J.M. Keller, A possibilistic approach to clustering, IEEE Transactions on Fuzzy Systems 1 (2) (1993) 98–110.
[28] R.N. Dave, Robust fuzzy clustering algorithms, in: Second IEEE International Conference on Fuzzy Systems, vol. 2, 1993, pp. 1281–1286.
[29] R.N. Dave, R. Krishnapuram, Robust clustering methods: a unitied view, IEEE Transactions on Fuzzy Systems 5 (1997) 270–293.
[30] R. Zwiggelaar, T.C. Parr, J.E. Schumm, I.W. Hutt, C.J. Taylor, S.M. Astley, C.R.M. Boggis, Model-based detection of spiculated lesions in mammograms, Medical Image Analysis 3 (1998) 1–23.
[31] V.N. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[32] X.Q. Shen, M. Spann, P. Nacken, Segmentation of 2d and 3d images through a hierarchical clustering based on region modelling, Pattern Recognition 31 (9) (1998) 1295–1309.
[33] Z. Huo, M.L. Giger, C.J. Vyborny, C.E. Metz, Effectiveness of cad in the diagnosis of breast cancer: an observer study on an independent database of mammograms, Radiology 224 (2002) 560–568.
[34] U. Raeth, D. Schlaps, B. Limberg, I. Zuna, A. Lorenz, G.V. Kaick, W.J. Lorenz, B. Kommerell, Diagnostic accuracy of computerized B-scan texture analysis and conventional ultrasonography in diffuse parenchymal and malignant liver disease, Journal Clinical Ultrasound 13 (1985) 87–99.