J Med Syst
DOI 10.1007/s10916-011-9691-4

ORIGINAL PAPER

Mammographic Image Based Breast Tissue Classification with Kernel Self-optimized Fisher Discriminant for Breast Cancer Diagnosis

Jun-Bao Li

Received: 1 January 2011 / Accepted: 25 March 2011

Abstract Breast tissue classification is an important and effective way for computer aided diagnosis of breast cancer with digital mammogram. Current methods endure two problems, firstly pectoral muscle influences the classification performance owing to its texture similar to parenchyma, and secondly classification algorithms fail to deal with the nonlinear problem from the digital mammogram. For these problems, we propose a novel framework of breast tissue classification based on kernel self-optimized discriminant analysis combined with the artifacts and pectoral muscle removal with multi-level segmentation based Connected Component Labeling analysis. Experiments on mini-MIAS database are implemented to testify and evaluate the performance of proposed algorithm.

Keywords Breast tissue classification · Breast cancer · Kernel self-optimized fisher discriminant · Connected component labeling

Introduction

One of approximately 11 women at some stage of their life in the Western world endures breast cancer as a major health problem in western countries. With the developing of all medical and technological advances, breast cancer cases are keeping still increasing in the recent 50 years and it is alarming but it is true that there is more women affected by breast cancer. For any form of cancer including breast cancer, early detection is one of the most important factors affecting the possibility of recovery from the disease. Diagnosis using mammography screening programs assisted by computers is an effective early detection of breast cancer. In the past one and a half decades, several researchers have studied computer-aided detection and classification of abnormalities about breast cancer in digital mammograms. Breast density, an important indicator of breast cancer risk in digital mammogram, an X-ray image of the breast tissue, is widely used for computer aided diagnosis of breast cancer. Mammographic density has relations to the breast cancer tumor characteristics. In the previous works, many statistical feature extraction methods have been proposed, but all the them extract the statistical features of the entire breast image including pectoral muscle and artifacts. Artifacts are from the processor, the technologist, the mammography and patient, and most of these artifacts are removed in the previous work. But for the patient identification and the type of X-ray view are marked and remained in the digital mammograms. The texture of pectoral muscle has high influence on classification owing to its similarity to the parenchyma. In the previous work, many methods are proposed for breast tissue density classification. Wang [1] presented histogram features based neural classification for breast tissue, and Muhimmah [2] applied multi-resolution histogram features based support vector machine into mammographic density classification, and Oliver [3] classified the breast tissue density through extracting morphological and textural features and k nearest neighbor classifier. In the past research work, researcher proposes informative genes selection based cancer classification with different ways [19–21], and researchers presented medical image tilt correction method

J.-B. Li (✉)
Department of Automatic Test and Control,
Harbin Institute of Technology,
P.O. Box 3033, No. 2 Yi-Kuang Street,
Nan-Gang District, Harbin 150080, China
e-mail: junbaolihit@hotmail.com

Published online: 08 April 2011

[22], and other relative work including DNA splicing system [23] and image recognition [24, 25]. Recently, Fisher classifier is a traditional dimensionality reduction technique for feature extraction, which has been widely used and proven successful in a lot of real-world applications. Fisher classifier works well in some cases, but it fails to capture a nonlinear relationship with a linear mapping. In order to overcome this weakness of linear Fisher classifier, the kernel trick is used to represent the complicated nonlinear relationships of input data to develop kernel Fisher classifier. Kernel-based nonlinear classifier techniques have attracted much attention in the areas of pattern recognition and machine learning [6–8]. Some algorithms using the kernel trick are developed in recent years, such as kernel Fisher classifier [9] and support vector machine (SVM) [11]. KPCA was originally developed by Scholkopf et al. [10], while kernel Fisher classifier was firstly proposed by Mika et al. [9]. Kernel Fisher classifier has been applied in many real-world applications owing to its excellent performance on feature extraction. Researchers have developed a series of kernel Fisher classifier algorithms (Lu [12] Yang [13], Yeung [14]). Because the geometrical structure of the data in the kernel mapping space, which is totally determined by the kernel function, has significant impact on the performance of these kernel Fisher classifier. The separability of the data in the feature space could be even worse if an inappropriate kernel is used. In order to improve the performance of KFD, many methods of optimizing the kernel parameters of the kernel function are developed in recentyears (Huang [15], Chen [16]). However, choosing the parameters for kernel just from a set of discrete values of the parameters does not change the geometrical structures of the data in the kernel mapping space. In order to overcome the limitation of the conventional KFD, in this paper, we present a novel Kernel Self-optimized Fisher Discriminant (KSFD) classifier, and KSFD adaptive changes the geometrical structure of data in the feature space with the different parameters of the quasiconformal kernel. Consequently Kernel Self-optimized Fisher Analysis is more adaptive to the input data for classification than KFD. All these methods extract the entire image including pectoral muscle, while pectoral muscle influences the classification performance owing to its texture similar to parenchyma. Moreover, the current classification methods are linear methods for feature classification, and they fails to deal with the nonlinear problem from the digital mammogram. So it is feasible to enhance the classification performance using the nonlinear kernel based classification methods.

As above discussion on medical image classification and its relative research topic, the current research has its research emphasis on classification but less attention on image preprocessing and feature extraction. Researches on the whole framework on the preprocessing, feature extraction and classification of medical images have not been paid attentions. In this paper, we propose the framework for breast tissue classification through eliminating the influence of background, artifacts and pectoral muscle in the digital mammograms and improving the nonlinear classification performance with kernel self-optimized discriminant analysis. Compared with the current research work, the proposed framework performs well on breast tissue classification under considering the background, artifacts and other influences on classification. We implement a set of experiments on the real database, MIAS database, from two parts, one is to evaluate the image preprocessing methods, second is to evaluate the performance on breast tissue classification with comparing with other popular breast tissue classification methods.

Classification method: Kernel self-optimization fisher discriminant

In this section, we propose a novel classification method, Kernel Self-optimization Fisher Discriminant (KSFD), for breast tissue classification. Firstly, we review the traditional kernel Fisher discriminant (KFD) algorithm, and secondly we present the theoretical derivation in detail of KSFD and describe its feasibility and affectivity with the 2D Gaussian distribution data.

Kernel Fisher discriminant

Firstly, we describe the detailed KFD algorithm, and the main idea of KFD is to map the original training samples to the feature space F with the nonlinear mapping Φ, and the linear discriminant analysis is implemented in the feature space F. Supposed that N dimensional M training sample $\{x_1, x_2, \ldots, x_M\}$ from L classes, then $\Phi : \mathbb{R}^N \to F, x \mapsto \Phi(x)$. The dimension of feature space F is very high, in order to avoid to deal with the mapped samples, we introduce the kernel function, which can be calculated by $k(x, y) = <\Phi(x), \Phi(y)>$. Through the nonlinear mapping, the scatter matrix between classes and total scatter matrix in the feature space F are defined as follows. $S_B^\Phi = \sum_{i=1}^{L} \frac{n_i}{M}\left(m_i^\phi - m^\phi\right)\left(m_i^\phi - m^\phi\right)^T$ and $S_T^\Phi = \frac{1}{M}\sum_{i=1}^{M}(\Phi(x_i) - m^\phi)(\Phi(x_i) - m^\phi)^T$, where $m^\phi = \frac{1}{M}\sum_{i=1}^{M}\Phi(x_i)$ and $m_j^\phi = \frac{1}{M}\sum_{i=1}^{n_j}\Phi(x_i)$. Then

$$J(V) = \frac{V^T S_B^\Phi V}{V^T S_T^\Phi V} \tag{1}$$

where V is the discriminative vector. According to the Mercer function theory, any solution vector V must be lies in the feature space consisted of $\{\Phi(x_1), \Phi(x_2), \ldots, \Phi(x_M)\}$, that is, there is one coefficient $c_p(p = 1, 2, ..., M)$ cause the following equation:

$$V = \sum_{p=1}^{M} c_p \{\Phi(x_p) = \Psi \alpha \tag{2}$$

where $\Psi = [\Phi(x_1), \Phi(x_2), ..., \Phi(x_M)]$ and $\alpha = [c_1, c_2, ..., c_M]^T$. The equation (1) is transformed to

$$J(\alpha) = \frac{\alpha^T KGK\alpha}{\alpha^T KK\alpha} \tag{3}$$

where $G = diag(G_1, G_2, ..., G_L)$,$G_i$ is a $n_i \times n_i$ matrix consisted of $\frac{1}{n_i}$, K is the kernel matrix calculated by $k(x, y)$. In fact, let $A_{opt} = [\alpha_1, \alpha_2, ..., \alpha_d]$ consisted of d discriminant vector $\alpha_1, \alpha_2, ..., \alpha_d$. The matrix A_{opt} satisfies

$$A_{opt} = \arg\max_A \frac{|A^T KGKA|}{|A^T KKA|} \tag{4}$$

Then the feature vector y of sampe x is $y = A_{opt}^T [k(x, x_1), k(x, x_2), ..., k(x, x_n)]^T$.

KFD algorithm:

Step 1. Select kernel function $k(x, y)$ and its parameters, calculate the kernel matrix K;
Step 2. Calculate the projection matrix A_{opt};
Step 3. The feature of the sample x is $y = A_{opt}^T [k(x, x_1), k(x, x_2), ..., k(x, x_n)]^T$.

As the above discussion, the projection matrix is the function of kernel matrix. If the kernel function and its parameter are not appropriately chosen, the projection matrix is not optimal for the mapping from the input space to the feature space. So the traditional KFD is not automatically to adjust the data structure in the feature space for feature extraction. We evaluate the performance on 2D Gaussian data, and the examples are shown in Fig. 1 (a). As shown in Fig. 1(b) and Fig. 1(c), KFD performs worse for Gaussian kernel and Polynomial kernel.

Kernel self-optimized Fisher discriminant

As the previous discussion, KFD finds an optimal linear projection from the kernel feature space to the projection subspace. Supposed that the nonlinear mapping Φ is inappropriately chosen, KFD can not find the optimal linear projection. In our algorithm, we optimize the nonlinear map Φ to maximize the class separability in feature space by optimizing the kernel, and then find the optimal transformation to maximize the class separability in projection subspace. Based on the above idea, we propose two stages of KFD algorithm, the first one is to optimize the kernel and the second is to find the optimal projection with the traditional method same as KFD. The geometry structure of sample data in the nonlinear projection space is different with the different kernel function. Accordingly,

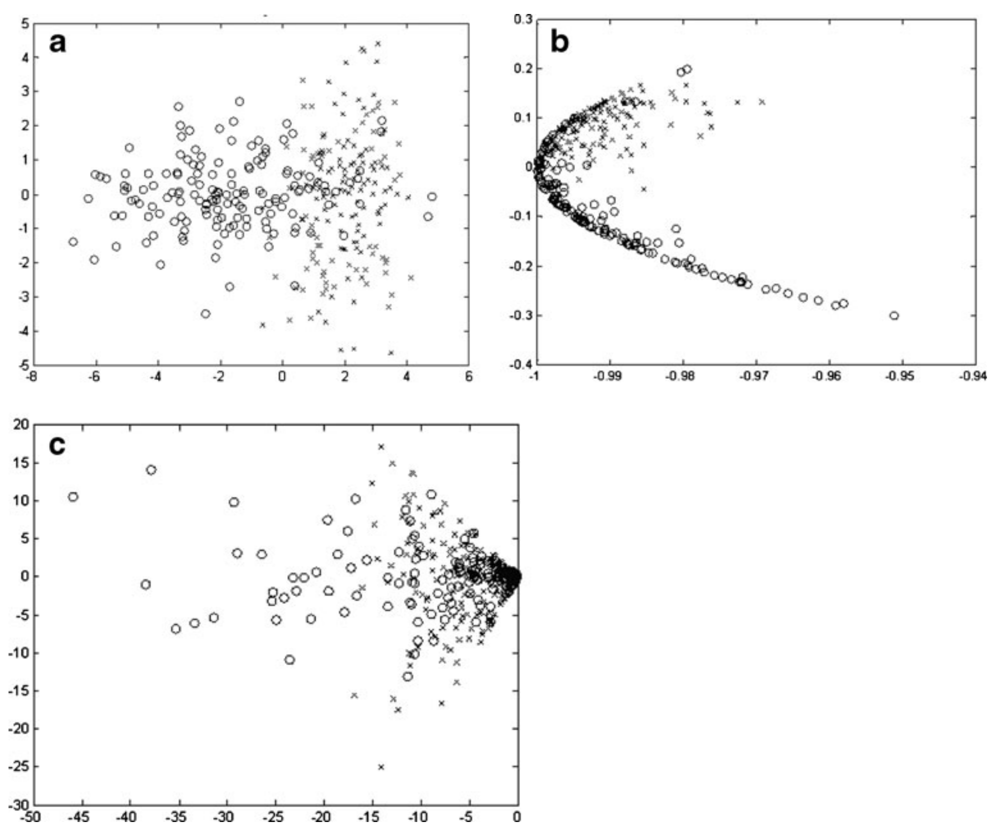


Fig. 1 Example with 2D data. (**a**) 2D Gaussian Data, (**b**) Projection data with Gaussian kernel, (**c**) Projection data with Polynomial kernel

data in the nonlinear projection space has the different class discriminative ability. So the kernel function should be dependent to the input data, which is the main idea of data-dependent kernel which was proposed in [17]. The parameter of the data-dependent kernel is changed according to the input data so that the optimal geometry structure of data in the feature space is achieved for the classification. In this paper, we extend the definition of the data-dependent kernel $k(x,y)=f(x)f(y)k_0(x,y)$ as the objective function for creating the constrained optimization equation to solve the solution, where $k_0(x,y)$ is the basic kernel function, such as polynomial kernel and Gaussian kernel. The function $f(x)$ is defined as $f(x)=\sum_{i\in SV} a_i e^{-\delta\|x-\widetilde{x}_i\|^2}$, where $\widetilde{x}_i$ is the support vector, SV is the set of support vector, a_i denotes the positive value which represent the distribution of $\widetilde{x}_i$, δ is the free parameter. We extend the definition of data-dependent kernel through defining the function $f(x)$ with the different ways as $f(x)=b_0+\sum_{n=1}^{N_{XV}} b_n e(x,\widetilde{x}_n)$, where δ is the free parameters, $\widetilde{x}_i$ is the expansion vectors (xvs) and N_{XV} is the number of expansion vectors, b_n (n=0,1,2,…,N_{XV}) is the according expansion coefficients. In our previous work [18], we present four methods of defining $e(x,\widetilde{x}_n)$.

Fisher criterion is to measure the class discriminative ability of the samples in the empirical feature space. The discriminative ability of samples in the empirical feature space is defined as

$$J_{Fisher}=\frac{tr\left(S_B^{\Phi}\right)}{tr\left(S_W^{\Phi}\right)} \tag{5}$$

where J_{Fisher} measure the linear discriminative ability, S_B^{Φ} is the between class scatter matrix, S_W^{Φ} is inter class scatter matrix, and tr denotes the trace. Let K is the kernel matrix with its element $k_{ij}(i,j=1,2,\dots,n)$ is calculated with x_i and x_j. The matrix $K_{pq},p,q=1,2,\dots,L$ is the $n_p\times n_q$ matrix with p and q class. Then in the empirical feature space, we can obtain $tr(S_B^{\Phi})=1_n^TB1_n$ and $tr(S_W^{\Phi})=1_n^TW1_n$, where $B=diag(\frac{1}{n_1}K_{11},\frac{1}{n_2}K_{22},\dots,\frac{1}{n_L}K_{LL})-\frac{1}{n}K$. The class discriminative ability is defined as

$$J_{Fisher}=\frac{1_n^TB1_n}{1_n^TW1_n} \tag{6}$$

According to the definition of the data-dependent kernel, let $D=diag(f(x_1),f(x_2),\dots,f(x_n))$, the relation between the data-dependent kernel matrix K and the basic kernel matrix K_0 calculated with basic kernel function $k_0(x,y)$ is defined as $K=DK_0D$. Accordingly, $B=DB_0D$ and $W=DW_0D$. Then

$$J_{Fisher}=\frac{1_n^TDB_0D1_n}{1_n^TDW_0D1_n} \tag{7}$$

where 1_n is n dimensional unit vector, according to the definition of data-dependent kernel, then

$$D1_n=E\alpha \tag{8}$$

where $\alpha=[a_0,a_1,a_2,\dots,a_{N_{XVs}}]^T$, the matrix E is consisted of $e(x,\widetilde{x}_n)$. Then

$$J_{Fisher}=\frac{\alpha^TE^TB_0E\alpha}{\alpha^TE^TW_0E\alpha} \tag{9}$$

where E^TB_0E and E^TW_0E are constant matrix, J_{Fisher} is a function with its variable α. Under the different expansion coefficient vector α, the geometry structure of data in the empirical space causes the discriminative ability of samples. Our goal is to find the optimal α to maximize J_{Fisher}. Supposed that α is an unit vector, i.e., $\alpha^T\alpha=1$., the constrained equation is created to solve the optimal α as follows.

$$\max \quad J_{Fisher}(\alpha)$$

$$subject \quad to \quad \alpha^T\alpha-1=0 \tag{10}$$

There are many methods of solving the above optimization equation. The following method is a classic method. Let $J_1(\alpha)=\alpha^TE^TB_0E\alpha$ and $J_2(\alpha)=\alpha^TE^TW_0E\alpha$, then

$$\begin{cases}\frac{\partial J_1(\alpha)}{\alpha}=2E^TB_0E\alpha\\ \frac{\partial J_2(\alpha)}{\alpha}=2E^TW_0E\alpha\end{cases} \tag{11}$$

Then

$$\frac{\partial J_{Fisher}(\alpha)}{\partial\alpha}=\frac{2}{J_2^2}\left(J_2E^TB_0E-J_1E^TW_0E\right)\alpha \tag{12}$$

In order to maximize J_{Fisher}, let $\frac{\partial J_{Fisher}(\alpha)}{\partial\alpha}=0$, then

$$J_1E^TW_0E\alpha=J_2E^TB_0E\alpha \tag{13}$$

If $(E^TW_0E)^{-1}$ exists, then

$$J_{Fisher}\alpha=\left(E^TW_0E\right)^{-1}\left(E^TB_0E\right)\alpha \tag{14}$$

J_{Fisher} is equal to the eigenvalue of $(E^TW_0E)^{-1}(E^TB_0E)$, and the corresponding eigenvector is equal to expansion coefficients vector α. In many applications, the matrix $(E^TW_0E)^{-1}(E^TB_0E)$ is not symmetrical or E^TWE is singular. So the iteration method is to solve α as follows.

$$\alpha^{(n+1)}=\alpha^{(n)}+\varepsilon\left(\frac{1}{J_2}E^TB_0E-\frac{J_{Fisher}}{J_2}E^TW_0E\right)\alpha^{(n)} \tag{15}$$

ε is the learning rate as follows. The definition of learning rate is

$$\varepsilon(n)=\varepsilon_0\left(1-\frac{n}{N}\right) \tag{16}$$

where ε_0 is the initialized learning rate, n and N is the current iteration number and the total iteration number in advance respectively.

The initialized learning rate ε_0 and the total iteration number N is set in advance for the solution of the expansion coefficient. The initial learning rate ε_0 influences the convergence speed of the algorithm, and the total iteration number N determines the time of solution. Only when the parameter ε_0 and N are chosen appropriately we choose the optimal expansion coefficient vector. So the solution of expansion coefficient is not unique, which is determined by the selection of learning parameter. The iteration algorithm costs much time. So we select the maximum margin criterion as the objective function to solve the optimal expansion coefficients. As shown in Fig. 2, KSFD performs better than KFD.

Breast tissue classification with Kernel self-optimized Fisher discriminant

In this section, we present a unified framework of breast tissue classification with kernel self-optimized Fisher discriminant for computer-aided diagnosis. The proposed framework of breast tissue classification contains four stages as shown in Fig. 3. The first stage is artifact removal, in this stage we implement the radio opaque artifacts removal algorithm using histogram threshold based Connected Component Labeling (CCL) segmentation to obtain the preprocessed image I without artifacts. The second is pectoral muscle removal, in which stage we remove the pectoral muscle from the breast using the morphological filter and CCL algorithm. The third is feature extraction, in this stage we compute the statistical features of breast tissue without background, artifacts and pectoral muscle including the mean pixel of image, the standard deviation, the smoothness, the asymmetry of pixels around the image mean, and average histogram, the uniformity, the Kurtosis. The fourth is classification, and in this stage we use the enhanced Kernel Fisher Discriminant, Kernel Self-optimized Fisher Discriminant (KSFD) to classify the statistical features of the input image.

Step 1. Artifacts removal

The intensity histogram of the original image reflects the pixel statistical distribution of background, breast, artifacts and pectoral muscle. So we apply histogram based segmentation and Connected Component Labeling (CCL) to extract the artifacts from the breast tissue. The threshold value used in multi-level histogram based segmentation is chosen in advance. The detail procedure of artifact removal is shown as follows.

Input: Original mammogram.

Output: Preprocessed mammogram without artifacts.

Step 1.1 Create the intensities histogram based on the original image.

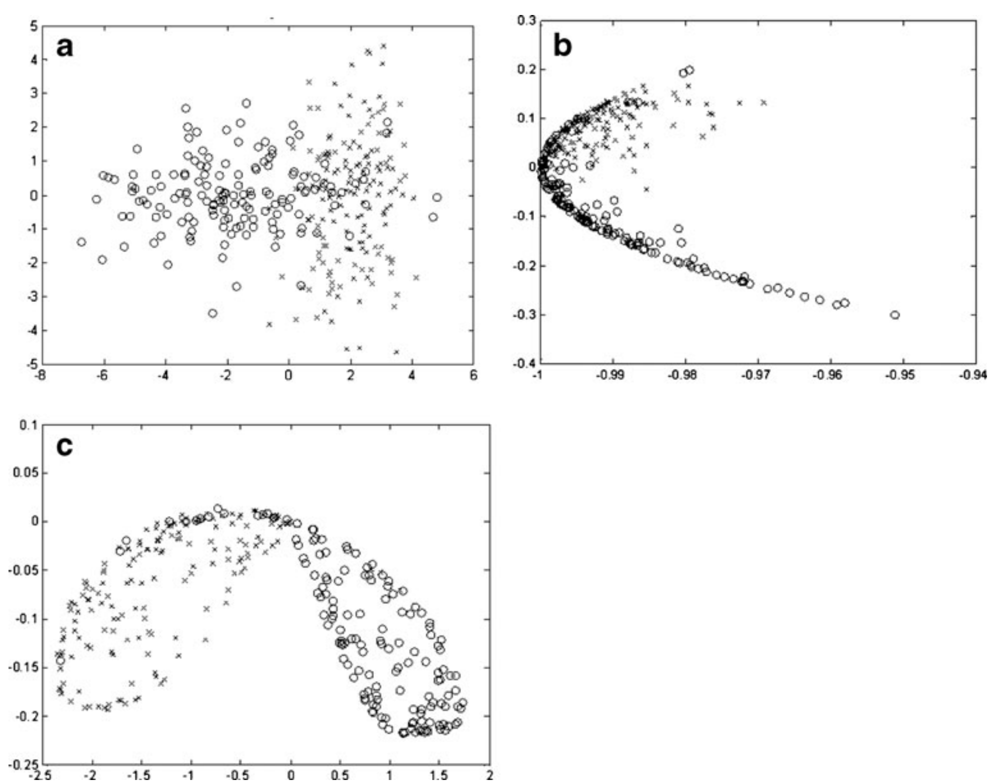


Fig. 2 Example with 2D data with KSFD. (**a**) 2D Gaussian Data, (**b**) Projection data with Gaussian kernel under KFD, (**c**) Projection data with KSFD

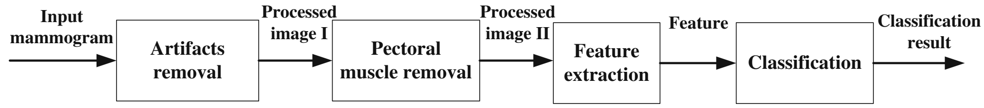


Fig. 3 Framework of breast tissue classification

Step 1.2 Binarize image based the threshold value for image segment

Step 1.3 Implement Connected Component Labeling (CCL).

Step 1.4 Recovery the image without artifacts.

Step 2. Pectoral muscle removal

This step aims to extract the pectoral muscle from the breast with histogram based threshold, morphological dilation, and CCL algorithm. The detail algorithm procedure is shown as follows.

Input: Preprocessed mammogram without artifacts.
Output: Preprocessed mammogram without artifacts and pectoral muscle.

Step 2.1 Removing the noise with median filtering.

Step 2.2 Binarized image with histogram-based segmentation

Step 2.3 Binarized image with morphological dilation

Step 2.4 Implement 8 pixels based CCL algorithm on the binarized image.

Step 2.5 Remove the pectoral muscle part from the image without artifacts

Step 3. Statistical features extraction

On preprocessed Mammograms, we implement the statistical feature extraction algorithm for classification. 7-dimensional feature vector is calculated for each mammogram. These seven features include mean pixel of image, the standard deviation, the smoothness, the asymmetry of pixels around the image mean, and average histogram, the uniformity, the Kurtosis. The statistical feature vector is constructed with the above 7 elements for classification.

Input: Preprocessed mammogram without artifacts and pectoral muscle.
Output: Statistical features extraction vector

Step 3.1 Calculate the statistical feature elements.

Step 3.2 Construct the statistical feature vector using the feature elements.

Step 4. Classification

We conduct the classification task with Kernel Self-optimized Fisher Discriminant (KSFD) proposed in our previous work [4] is used for classification. KFD finds the optimal projection from the feature space to the projection subspace. KSFD implement classification with two stages, the first is to find the optimal nonlinear mapping from input space to the feature space through optimizing the data-dependent kernel function, and the second is to seek the optimal linear projection from the feature space to the low dimensional projection subspace.

Input: A N-dimensional statistical feature vector x of the training preprocessed digital mammogram.
Output: A low-dimensional representation y of x with enhanced discriminating power.

Step 4.1. Calculate the basic kernel matrix with chosen kernel function in advance.

Step 4.2. Seek the optimal parameters of the data-dependent kernel with the basic kernel matrix with the kernel optimization criterion.

Step 4.3. Implement KFD algorithm with the optimized data-dependent kernel adaptive to the training feature vectors of preprocessed digital mammogram.

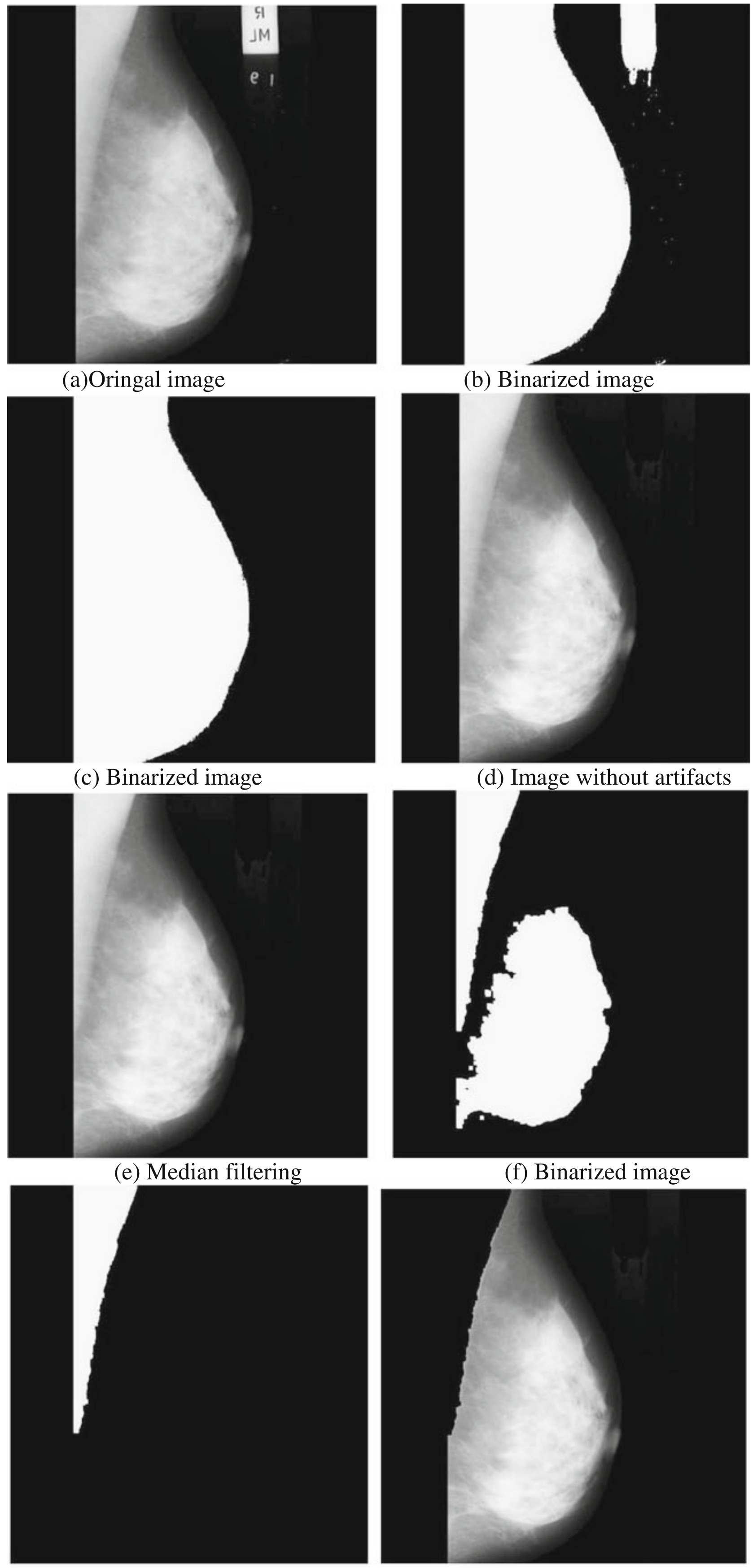
Step 4.4. Classify the new test sample with KSFD.

Experimental results

In this section, we testify and evaluate the performance and feasibility of the proposed algorithm with simulation conditions. We implements the algorithms on Matlab platform and use the public databases to test algorithms. The set of experiments are implemented in a digital mammography database, mini-MIAS database [5], which is developed by the Mammography Image Analysis Society. In the database, the X-ray films are have been digitized with a Joyce-Lobel scanning microdensitometer to a resolution of 50 μm$\times$50 μm, 8-bit word, and the original image has the size of 1,024$\times$1,024 pixels. In the experiments, firstly, we evaluate the feasibility of artifacts and pectoral muscle removal algorithm, statistical features extraction. Secondly, we evaluate the performance of framework compared with other algorithms. We choose the three types, Fatty, Glandular and Dense, from the norm samples to test the performance of the proposed framework. 12 mammograms have fatty tissue, and 14 mammograms have the glandular tissue and the rest 16 have the dense tissue. We evaluate the recognition accuracy with cross-validation method.

In this set of experiments, we evaluate the performance of algorithms with the following four steps.

Step 1). Implement artifact removal algorithm. The radio opaque artifacts removal algorithm using histo-



(a)Oringal image (b) Binarized image (c) Binarized image (d) Image without artifacts (e) Median filtering (f) Binarized image (g) 8 pixels based CCL algorithm (h) Preprocessed mammogram

Fig. 4 Experimental results of mammograms preprocessing

Table 1 Performance comparison on mini-MIAS database

Algorithms	Features	Classification accuracy
Proposed framework	Statistical features	94.46%
KFD	Statistical features	93.33%
kNN	Statistical features	92.17%
Neural classifier [1]	Histogram features	75.31%
SVM [2]	Multi-resolution histogram features	79.76%
kNN [3]	Morphological and textural features	91.58%

gram threshold based Connected Component Labeling (CCL) segmentation is implemented to preprocess image.

Step 2). Implement pectoral muscle removal algorithm. the morphological filter and CCL algorithm is implemented to remove the pectoral muscle from the breast.

Step 3). Implement feature extraction algorithm. We compute the statistical features of breast tissue without background, artifacts and pectoral muscle including the mean pixel of image, the standard deviation, the smoothness, the asymmetry of pixels around the image mean, and average histogram, the uniformity, the Kurtosis.

Step 4). Implement classification algorithm. Kernel Self-optimized Fisher Discriminant (KSFD) is used to classify the statistical features of the input image.

The results of Step 1) and 2) are shown with figures, while step 3) and 4) is evaluated with parametric level in tables. The detailed descriptions are shown as follows.

Firstly, we implement performance evaluation on digital mammogram preprocessing including artifacts and muscle removal algorithm. The results on one example (mdb004.pgm) are shown in Fig. 4. Figure 4(a) is the original image with ID mdb004, and Fig. 4(b) shows the segmented image with the threshold value determined by the histogram, and Fig. 4(c) shows the binarized image without artifacts after CCL algorithm, and Fig. 4(d) shows the image without artifacts. Secondly, we implement pectoral muscle removal algorithm on the image without artifacts, and the result is shown in Fig. 4. Figure 4(e) is the result after median filtering, and Fig. 4(f) shows the binarized image with histogram and morphological dilation based segmentation, and Fig. 4(g) is the muscle region in the whole mammogram with 8 pixels based CCL algorithm, and Fig. 4(h) is the final preprocessed mammogram used for the next statistical feature extraction. We also implement the same preprocessing algorithm on other examples from the database, the similar results are achieved. The proposed algorithm performs well on the mammogram preprocessing including removing of noise, background, artifacts, pectoral muscle.

Secondly, we compare the proposed framework with other popular classification algorithms on breast tissue classification. On the one hand, we implement KSFD classification together with other popular classification methods including k Nearest Neighbor (kNN) classifier, Bayesian classifier, Fisher classifier and kernel Fisher discriminant classification methods. On the other hand, we also implement other algorithms proposed in the previous works including histogram feature based neural network[1], multi-resolution histogram features based SVM [2], Morphological and textural features based kNN classification [3]. The comparison results are shown Table 1. The proposed framework outperforms other methods under the same training samples and test samples set. So, it is feasible to improve the classification performance using the proposed preprocessed digital mammogram and KSFD classification methods. In the experimental results, we achieve 94.46% recognition accuracy higher than other popular algorithms. Moreover, the proposed statistical feature extraction on preprocessed mammogram without artifacts and pectoral muscle performs better than morphological and textural feature extraction on the entire mammogram under the same classifier.

From the experimental results, breast tissue classification algorithm with kernel self-optimized Fisher discriminant

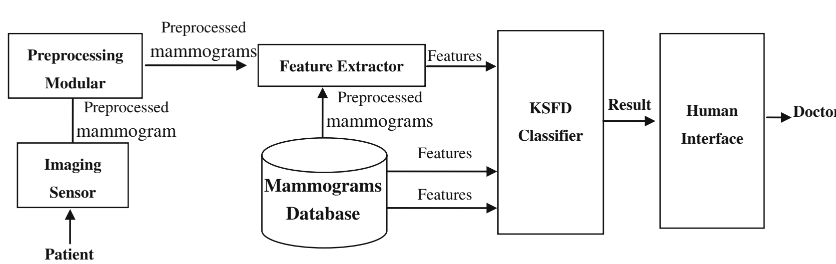


Fig. 5 The framework of a real system example

analysis is feasible and effective to for mammographic image based beast cancer diagnosis. In the applications, the algorithm is realized to the software of computer-aided diagnosis. The framework of the real application system is shown in Fig. 5. Firstly, the system collects the mammograms of the patient with the imaging sensor, and then the mammograms are preprocessed with the image preprocessing modular. Then the features of the preprocessed mammograms are extracted with the feature extractor which is realized with the software in the system. On the other hand, there are many mammograms are saved in the mammograms database. When we implement the system, the enquired patient's mammogram is classified into the classification with matching with the samples in the database. The KSFD classifier is applied into classification. The classification results are display with human interface to doctor as the diagnosis reference. In this system, all algorithms are implemented with the software, and imaging modular is realized with hardware.

Conclusion

In this paper, we propose a novel framework of breast tissue density classification in digital mammogram. The framework includes the novel mammogram preprocessing method using artifacts and pectoral muscle removal based on multi-level segmentation based CCL analysis, the statistical feature extraction method, and the novel classification method Kernel Self-optimized Fisher Discriminant (KSFD). The contributions are summarized as follows. Firstly, muscle removal algorithm is applied preprocess the original image together with artifacts, which deals with the problem that the pectoral muscle influences the classification performance owing to its texture similar to parenchyma. Secondly, we applied the enhanced Kernel Discriminant Analysis with self-optimized data-dependent kernel into feature classification, which deals with the problem that the breast tissue classification algorithms fail to deal with the nonlinear problem from the digital mammogram.Experimental results on the real database show that the proposed algorithm is feasible and effective for breast tissue classification.

Acknowledgments This work is supported by National Science Foundation of China under Grant No. 61001165 and Natural Science Foundation of Heilongjiang Province under Grant No. QC2010066.

References

1. Wang, X. H., Good, W. F., Chapman, B. E., Chang, Y. H., Poller, W. R., Chang, T. S., and Hardesty, L. A., Automated assessment of the composition of breast tissue revealed on tissue-thickness-corrected mammography. *Am. J. Roentgenol.* 180:227–262, 2003.
2. Muhimmah, I., and Zwiggelaar, R., "Mammographic density classification using multiresolution histogram information", Proceedings of the International Special Topic Conference on Information Technology in Biomedicine, Ioannina, Greece, 2006
3. Oliver, A., Freixenet, J., Marti, R., Pont, J., Perez, E., Denton, E. R. E., and Zwiggelaar, R., A novel breast tissue density classification methodology. *IEEE Trans. Inf. Technol. Biomed.* 12(1):55–65, 2008.
4. Li, J. B., Pan, J. S., and Lu, Z. M., Kernel optimization-based discriminant analysis for face recognition. *Neural Comput. Appl.* 18(6):603–612, 2009.
5. Suckling, J., Parker, J., Dance, D., Astley, S., Astley, I., Hutt, I., and Boggis, C., The mammographic images analysis society digital mammogram database. *Exerpta Med.* 1069:375–378, 1994.
6. Li, J.-B., Chu, S.-C., Pan, J.-S., and Ho, J.-H., Adaptive data-dependent matrix norm based Gaussian kernel for facial feature extraction. *Int. J. Innovative Comput. Inf. Control.* 3(5):1–10, 2007.
7. Sahbi, H., Kernel PCA for similarity invariant shape recognition. *Neurocomput* 70:3034–3045, 2007.
8. Wu, M., Schölkopf, B., and Bakır, G., A direct method for building sparse kernel learning algorithms. *J. Mach. Learn Res.* 7:603–624, 2006.
9. Mika, S., Ratsch, G., Weston, J., Schölkopf, B., and Muller, K.-R., Fisher discriminant analysis with kernels, in: Proceedings of IEEE International Workshop on Neural Networks for Signal Processing IX, pp. 41–48, 1999
10. Schölkopf, B., Smola, A., and Muller, K. R., Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* 10(5):1299–1319, 1998.
11. Tao, D., Tang, X., Li, X., and Wu, X., Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* 28(7):1088–1099, 2006.
12. Lu, J., Plataniotis, K. N., and Venetsanopoulos, A. N., Face recognition using kernel direct discriminant analysis algorithms. *IEEE Trans. Neural Netw.* 14(1):117–226, 2003.
13. Yang, J., Frangi, A. F., Yang, J.-Y., Zhang, D., and Jin, Z., KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* 27(2):230–244, 2005.
14. Yeung, D.-Y., Chang, H., and Dai, G., Learning the kernel matrix by maximizing a KFD-based class separability criterion. *Pattern Recogn.* 40(7):2021–2028, 2007.
15. Huang, J., Yuen, P. C., Chen, W.-S., and Lai, J. H., Kernel subspace LDA with optimized kernel parameters on face recognition, in: Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition, 2004.
16. Chen, W.-S., Yuen, P. C., Huang, J., and Dai, D.-Q., Kernel machine-based one-parameter regularized fisher discriminant method for face recognition. *IEEE Trans. Syst. Man. Cybern. B Cybern.* 35(4):658–669, 2005.
17. Amari, S., and Wu, S., Improving support vector machine classifiers by modifying kernel functions. *Neural Netw.* 12 (6):783–789, 1999.
18. Pan, J. S., Li, J. B., and Lu, Z. M., Adaptive quasiconformal Kernel discriminant analysis. *Neurocomputing* 71:2754–2760, 2008.
19. Mohamad, M. S., Omatu, S., Yoshioka, M., and Deris, S., A three-stage method to select informative genes for cancer classification. *Int. J. Innovative Comput. Inf. Control* 6(1):117–126, 2010.
20. Mohamad, M. S., Omatu, S., Yoshioka, M., and Deris, S., A cyclic hybrid method to select a smaller subset of informative genes for cancer classification. *Int. J. Innovative Comput. Inf. Control* 5(8):2189–2202, 2009.
21. Mohamad, M. S., Omatu, S., Yoshioka, M., and Deris, S., A two-stage method to select a smaller subset of informative genes for

cancer classification. *Int. J. Innovative Comput. Inf. Control* 5 (10):2959–2968, 2009.

22. Pan, M., Tang, J., and Yang, X., An algorithm for medical image tilt correction using B-Spline and moment invariants. *ICIC Express Lett.* 4(1):57–64, 2010.
23. Zhou, C., Wei, X., Zhang, Q., and Wang, B., Image encryption based on DNA splicing system and henon model. *ICIC Express Lett.* 4 (3B):1011–1018, 2010.
24. Arora, S., Bhattacharjee, D., Nasipuri, M., Basu, D. K., and Kundu, M., Complementary features combined in a MLP-based system to recognize handwritten devnagari character. *J. Inf. Hiding Multimedia Signal Process.* 2(1):71–77, 2011.
25. Parviz, M., and Shahram Moin, M., Boosting approach for score level fusion in multimodal biometrics based on AUC maximization. *J. Inf. Hiding Multimedia Signal Process.* 2 (1):51–59, 2011.