Kernel Combined Sparse Representation for Disease Recognition

Qingxiang Feng and Yicong Zhou, *Senior Member, IEEE*

Abstract—**Motivated by the idea that the correlation structure of the entire training set can disclose the relationship between the test sample and the training samples, we propose the combined sparse representation (CSR) classifier for disease recognition. The CSR classifier minimizes the correlation structure of the entire training set multiplied by its transposition and the sparse coefficient together for classification. Including the kernel concept, we propose the kernel combined sparse representation classifier utilizing the high-dimensional nonlinear information instead of the linear information in the CSR classifier. Furthermore, considering the information of the training samples and the class center, we then propose the center-based kernel combined sparse representation (CKCSR) classifier. CKCSR uses the center-based kernel matrix to increase the center-based information that is helpful for classification. The proposed classifiers have been evaluated by extensive experiments on several well-known databases including the EXACT09 database, Emphysema-CT database, mini-MIAS database, Wisconsin breast cancer database, and HD-PECTF database. The experimental results demonstrate that the proposed classifiers achieve better recognition rates than the sparse representation-based classification, collaborative representation based classification, and several state-of-the-art methods.**

Index Terms—**Collaborative representation, disease recognition, sparse representation.**

I. INTRODUCTION

DUE to the fast development of medical imaging technologies, a large number of medical images have been generated every single minute in clinic applications all over the world. Disease recognition based on medical images is an effective tool in the computer-aid system that helps doctors to quickly locate and diagnose the disease in the patient's body. Disease recognition is also a challenging problem in the medical field because of complex and diversity of individual patient health conditions. This attracts more and more attentions of researchers. To perform disease recognition [1], [2], accurate and efficient image classification, recognition and retrieving methods are demanded. To meet this need, this paper focuses on medical image classification for disease recognition.

The authors are with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: fengqx1988@gmail.com; yicongzhou@umac.mo).

In general, image classification includes two stages: feature extraction and classification. Existing research on feature extraction manly works on incorporating dictionary learning techniques [3]–[5], generating new feature descriptors [6]–[8] and performing optimized feature selection [9], [10]. For classification, the classifiers play an important role for the overall classification performance. Therefore, we aim to develop the better classifiers that can help doctors to efficiently determine the disease's categories. The supervised classifier can be roughly divided into two categories: The class-subspace- based classifiers such as the nearest neighbor (NN) [11] and linear regression [12]–[15]; and the entire-space-based classifiers such as sparse representation classification (SRC)[16], [17]. The latter one has attracted growing attention in recent years.

In the above description, the hospitals and medical institutions need the better methods to meet the disease recognition task. As the well-known classification method, sparse representation based classification (SRC) [16], [17] attracts a lot of researchers' attention and obtains the good performance for image classification task [18]–[29]. However, the above sparse-based methods fail to consider the correlation structure of the prototype set, which is also useful for classification [30], [31]. Motivated by this, this paper tries to minimize the correlation structure of the entire training set multiplied by its transposition and the sparse coefficient together. Based on this idea, the combined sparse representation (CSR) classifier is proposed for disease recognition. CSR can find the relationship between the test sample and the training samples. Moreover, we propose the kernel combined sparse representation (KCSR) classifier to map the original feature to a nonlinear high-dimensional feature such that it can obtain the nonlinear information for classification. In order to increase the information of the training samples and the corresponding class mean sample, we further propose the center-based kernel combined sparse representation (CKCSR) classifier using the center-based kernel matrix. Due to the increase of the center-based nonlinear information, the CKCSR classifier can obtain better performance than KCSR for the most situations. Experiments on Emphysema-CT database, EXACT09 database, mini-MIAS database, WBC database and HD-PECTF database are used to evaluate the proposed algorithms. The experimental results show that the proposed methods achieve better recognition rates than sparse representation based classification (SRC), collaborative representation based classification (CRC), and several state-of-the-arts approaches. To sum up, the main contributions of this paper can be summarized as follows.

1) We propose the combined sparse representation (CSR) classifier, which considers the correlation structure and

sparsity together. In the CSR classifier, the correlation structure of the entire training set multiplied by its transposition is used such that the correlation can be encoded for classification.

2) Based on the kernel concept, the kernel combined sparse representation (KCSR) classifier is also proposed. It utilizes the high-dimension nonlinear information instead of the linear information in the CSR classifier.

3) Considering the information of the training samples and the class center, we further propose center-based kernel combined sparse representation (CKCSR) classifier. It uses the center-based kernel matrix to increase the center-based information, which is helpful for classification.

4) Extensive experiments are carried out on five medical disease databases including the EXACT09 database, Emphysema-CT database, mini-MIAS database, WBC database and HD-PECTF database. The experiment results show that the proposed classifiers obtain better performance than several state-of-the-art methods.

II. RELATED WORK

In the class-subspace-based classifier [32], the number of prototype samples is usually quite small, which makes the classification much difficult. The nearest feature line (NFL) proposed by Li *et al.* [33], [34] attempts to enhance the representation capacity of the limited sample set by using the line passing through each pair of the samples in the same class. After NFL, Pan *et al.* proposed the neighborhood feature line segment (NFLS) [35] selecting the useful lines rather than using all the lines for classification; Chien *et al.* proposed the nearest feature plane (NFP) [36], which uses the feature plane instead of the feature line in NFL; Feng *et al.* proposed the center-based nearest feature plane (CNFP) and line-based nearest feature plane (LNFP) [37] constituting the new feature plane.

Different from the class-subspace-based classifier, sparse representation based classification (SRC) [16], [17] uses all class-subspaces to solve the L_1-norm minimum problem and to classify the test sample. Based on the SRC classifier, some researchers use the idea of SRC to perform the feature selection [18], [19], [20], while other researchers pay attention to improve the classifier. Yang *et al.* proposed the regularized robust coding [38] using a new data distribution method for classification. Zhang *et al.* proposed the collaborative representation classification (CRC) [39], which solves the L_2-norm minimum problem instead of L_1-norm minimum problem in SRC, and uses the collaborative representation instead of the sparse representation in SRC. Based on CRC, Yang *et al.* proposed the relaxed collaborative representation [40] and Cai *et al.* proposed probabilistic collaborative representation based classification [41] for pattern recognition. Xu *et al.* proposed two-phase sparse representation (TPSR) [42], which chooses some useful samples while removing the remaining samples in the first phase and solves the L_2-norm minimum problem in the second phase. In [22], Fang *et al.* proposed the non-negative sparse graph, which solves the non-negative sparse coefficient for classification. In [24] and [23], Lai *et al.* used the sparse idea to perform the dimension re-

duction. In [25], Feng *et al.* proposed the superposed method for the sparse representation and obtained the good performance. In [43], Xu *et al.* proposed a novel sparse representation method using data uncertainty idea for face recognition. In [27], Gu *et al.* proposed dictionary pairing learning (DPL) achieving from dictionary learning to dictionary pair learning. In [28], Yang *et al.* proposed Fisher Discrimination Dictionary Learning (FDDL) to make the dimension reduction. In [44], Deng *et al.* proposed the superposition sparse representation classification (SSRC) using the prototype plus variations to optimize the prototype set.

All above methods are based on the linear information and fail to consider the nonlinear information. Motivated by this, researchers utilize the kernel idea to solve it. For example, Gao *et al.* proposed kernel sparse representation (KSR) [29]. It maps the original feature to a nonlinear high-dimension feature and uses the feature searching method to solve the optimization problem. Based on KSR, Zhang et al. proposed kernel sparse representation based classification (KSRC) [45]. It uses the dimension reduction method to reduce the dimension of the nonlinear feature so that the computation cost is much less than that of KSR. Wang *et al.* proposed kernel collaborative representation classification (KCRC) [46] utilizing the collaborative representation instead of the sparse representation in KSRC. Lu *et al.* proposed kernel linear regression classification (KLRC) [47], which maps the original feature to a nonlinear high-dimension feature and uses the least square errors for classification. Feng *et al.* proposed the center-based weighted kernel linear regression classification (CWKLRC) [48], which brings the "weighted idea" to KLRC. It can encode the important information of the training samples. Wang *et al.* proposed a kernel classification framework (KCF) [49] for metric learning, which gains a fast speed for the metric leaning. Li *et al.* proposed kernel collaborative representation classification with Tikhonov regularization (KCRC-TR) [50] using the Tikhonov Matrix to encode the important information. Chou *et al.* proposed class-specific kernel linear regression classification (CKLRC) [51] increasing the class-specific for KLRC. Zhang *et al.* proposed kernel sparse representation- based classifier ensemble (KSRCE) [52] considering the ensemble [53], [54],[55], [56] information for KSRC. Moreover, Shrivastava *et al.* used the multiple kernels learning for sparse representation [57], which can benefit from both multiple kernels and sparsity. Lan *et al.* proposed the joint sparse representation [58] and Robust joint discriminative feature learning [59], which both obtained good performance for visual tracking. Xu *et al.* proposed learning a structured dictionary for video-based face recognition [60] and obtained good performance.

Recently, some researchers have applied SRC to medical images such that the doctors can efficiently determine the disease's categories. For example, Song *et al.* proposed large margin local estimate (LMLE) for medical image classification [1]. Kong *et al.* proposed the jointly sparse learning for breast cancer [2]. Tong *et al.* utilized the sparse coding for MR images segmentation [61]. Zhang *et al.* used the sparse representation for deformable segmentation [4]. Wang *et al.* integrated the sparse multi-modality representation and anatomical constraint for isointense infant brain MR image segmentation [62].

Song *et al.* combined the sparse and boost for ILD classification [63]. Liu *et al.* proposed localized sparse code gradient in Alzheimers disease staging [5]. Weiss *et al.* utilized the sparse coding for multiple sclerosis lesion segmentation [64]. Zhang *et al.* [65] applied the sparse representation for higher-order functional interaction patterns in task-based fMRI data.

III. Proposed CSR Classifier

In this section, we first introduce the objective function of the proposed CSR classifier. Section III-B describes the solution method of the objective function. Next, the detailed classification procedures of the CSR classifier are described in Section III-C. Finally, Section III-D describes the CSR vs methods in [30], [31].

Definition: Let $X = \{x_i^c\}, c = 1, 2, \ldots, M, i = 1, 2, \ldots, N_c$ denote the prototype set, where x_i^c is the ith prototype belonging to the cth-class, M is the number of classes, and N_c is the number of prototype samples in the cth-class.

A. Objective Function of CSR

In [30], [31], authors showed that the sparsity is effective in sample selection for representation while the correlation structure helps to find the relationship between the query sample and the training samples. Therefore, they tried to minimize the structure of entire training set X and the sparse coefficient together. In order to emphasize the importance of the correlation structure of the entire training set, this paper tries to increase the entire training sets transposition for optimization. That is, this paper tries to minimize the correlation structure of the entire training set multiplied by its transposition, $X^T X$, and the sparse coefficient together. Motivated by the methods in [30], [31], we try to utilize the L_p-minimization ($1 < p < 2$) to solve the sparse problem, which is different from the L_1-minimization optimization problem in SRC and L_2-minimization optimization problem in CRC. The objective function of CSR can be described as follows:

$$\min ||X^T X \, Diag(\beta)||_*$$
$$s.t. \quad x = X\beta \tag{1}$$

where $|| \bullet ||_*$ is the trace or nuclear norm. In order to explain the relationship between the proposed objective function and the objective functions in SRC and CRC, two examples are given as follows. Suppose that the subjects of X are different from each other, the columns in X are orthogonal, that is, $X^T X = I$, where I denotes the identity matrix. Then, the decomposition is

$$||X^T X \, Diag(\beta)||_* = ||Diag(\beta)||_*$$
$$= Tr[(Diag(\beta))^T (Diag(\beta))]^{1/2}$$
$$= ||\beta||_1. \tag{2}$$

Therefore, the objective function in (1) is equivalent to that of SRC. On the other hand, suppose that the subjects of X are similar to x_1, the first column of X, that is $X^T X = 11^T$ (1 is a vector of size $n = \sum_{c=1}^{M} N_c$, each element of which is one).

According to [31], the decomposition is described as

$$||X^T X Diag(\beta)||_* = ||11^T Diag(\beta)||_*$$
$$= ||11^T Diag(\beta)||_F$$
$$= \sqrt{n}||\beta||_2 \tag{3}$$

where $|| \bullet ||_F$ is the Frobenius norm. The derivation procedure of (3) is described in Appendix A. Notice: For a specific classification task, the training set is fixed. That is, the number of training samples n is a fixed value. Thus, minimization of $\sqrt{n}||\beta||_2$ is equivalent to minimization of $||\beta||_2$. Generally, the samples of the prototype set X are neither distinct from each other nor similar to each other. Thus, the objective function in (1) can be treated as a combination of the L_1-minimization in SRC and L_2-minimization in CRC. That is, the objective function in (1) can benefit from both L_1-norm minimization and L_2-norm minimization according to the correction of the samples in the prototype set X. Because $x = X\beta$ implies $X^T x = X^T X\beta$, the (1) can be rewritten as

$$\min ||X^T X Diag(\beta)||_*$$
$$s.t. \quad X^T x = X^T X\beta. \tag{4}$$

It is noted that the objective function in (4) is designed for the situation that the samples have no noise. However, noise often exists in the real-world applications. Considering noise, the objective function in (4) becomes

$$\min ||X^T X Diag(\beta)||_*$$
$$s.t. \quad ||X^T x - X^T X\beta||_1 < \varepsilon. \tag{5}$$

Equation (5) is the final objective function of the CSR classifier.

B. Solution of the Optimization Problem

According to the optimization approaches in [66] and [67], the method of inexact augmented Lagrange multipliers (IALM) [68] is used to solve the minimization problem in (5), which can be converted to the following minimization problem:

$$\min_{E,e,\beta} ||E||_* + ||e||_1$$
$$s.t. \quad e = X^T x - X^T X\beta$$
$$E = X^T X Diag(\beta). \tag{6}$$

Instead of solving the minimization problem in (6), we solve the following augmented Lagrange multiplier problem in (7) as

$$L(E, e, \beta) = \lambda ||E||_* + ||e||_1 + y_1^T (a - A\beta - e)$$
$$+ Tr[Y_2^T (E - ADiag(\beta))] + \frac{u}{2}(||a - A\beta - e||_2^2$$
$$+ ||E - ADiag(\beta)||_F^2) \tag{7}$$

where $u > 0$ is a parameter, y_1 and Y_2 are Lagrange multipliers, $A = X^T X$ and $a = X^T x$.

Variables E, e and β in (7) can be optimized alternatively with the other two fixed. The detailed optimization procedures of E, e and β are described as follows.

Update E when e and β are fixed, which is equivalent to solve the minimization problem in (8)

$$
\begin{aligned}
E* &= \arg\min_{E} L(E, e, \beta) \\
&= \arg\min_{E} \lambda\|E\|_* + Tr(Y_2^T E) + \frac{u}{2}\|E - ADiag(\beta)\|_F^2 \\
&= \arg\min_{E} \frac{\lambda}{u}\|E\|_* + \frac{1}{2}\|E - \left(ADiag(\beta) - \frac{1}{u}Y_2\right)\|_F^2.
\end{aligned}
\tag{8}
$$

The minimization problem in (8) can be solved approximately by the singular value thresholding (SVT) operator [69].

Afterwards, update β when e and E are fixed, which is equivalent to solve the minimization problem in (9)

$$
\begin{aligned}
\beta* &= \arg\min_{\beta} L(E, e, \beta) \\
&= \arg\min_{\beta} -y_1^T A\beta - Tr(Y_2^T ADiag(\beta)) \\
&\quad + \frac{u}{2}(\beta^T A^T A\beta - 2(a - e)^T A\beta - 2Tr(E^T ADiag(\beta)) \\
&\quad + Tr((ADiag(\beta))^T ADiag(\beta))) \\
&= \arg\min_{\beta} \frac{u}{2}\beta^T (A^T A + Diag(diag(A^T A)))\beta \\
&\quad - (A^T y_1 + uA^T A(a - e) + diag(Y_2^T A + uE^T A))^T \beta.
\end{aligned}
\tag{9}
$$

It is easy to solve the minimization problem in (9) by

$$
\begin{aligned}
\beta* &= (A^T A + Diag(diag(A^T A)))^{-1} A^T \left(\frac{1}{u}y_1 + A(a - e)\right) \\
&\quad + (A^T A + Diag(diag(A^T A)))^{-1} diag\left(A^T \left(\frac{1}{u}Y_2 + E\right)\right).
\end{aligned}
\tag{10}
$$

Next, update e when E and β are fixed, which is equivalent to solve the following minimization problem:

$$
\begin{aligned}
e* &= \arg\min_{e} L(E, e, \beta) \\
&= \arg\min_{e} \|e\|_1 - y_1^T e + \frac{u}{2}\|a - A\beta - e\|_F^2 \\
&= \arg\min_{e} \frac{1}{u}\|e\|_1 + \frac{1}{2}\|e - \left(a - A\beta + \frac{1}{u}y_1\right)\|_2^2.
\end{aligned}
\tag{11}
$$

The solution of the minimization problem in (11) can be solved by the soft thresholding (shrinkage) operator [70].

After updating E, e and β, the multipliers y_1 and Y_2 can be updated by

$$
\begin{aligned}
y_1 &= y_1 + u(a - A\beta - e) \\
Y_2 &= Y_2 + u(E - ADiag(\beta)).
\end{aligned}
\tag{12}
$$

The parameter u can be updated by

$$
u = \min(\rho u, u_{\max})
\tag{13}
$$

where $\rho, u_{\max}$ are constants and given in advance.

C. Classification

After obtaining the results of the optimization problem of the objective function, we calculate the sum of coefficients of each class by

$$
s_c = \sum_{i=1}^{N_c} \beta_i^c, c = 1, 2, ..., M.
\tag{14}
$$

Finally, the test sample will be classified into the class with the maximum sum of coefficients by

$$
c* = \arg\max(s_c).
\tag{15}
$$

The detailed classification procedures of CSR are summarized in Algorithm 1.

D. CSR Versus Methods in [30] and [31]

The objective function of methods in [30], [31] considers the correlation structure of the prototype set and sparsity together for classification. Their objective function can be described as $\min \|XDiag(\beta)\|_* \ s.t. \ x = X\beta$ and obtain the good performance for some classification tasks. Motivated this, CSR tries to emphasize the importance of the correlation structure of the entire training set. CSR increases the correlation structure of the entire training set's transposition for the optimization problem. That is, CSR tries to minimize the correlation structure of the entire training set multiplied by its transposition, $X^T X$, and the sparse coefficient together for classification. The objective function of CSR can be described as $\min \|X^T XDiag(\beta)\|_* \ s.t. \ x = X\beta$. Therefore, the main difference between CSR and methods in [30], [31] is that CSR uses the entire training set multiplied by its transposition, $X^T X$, instead of the entire training set, X, as in [30], [31].

IV. PROPOSED KCSR AND CKCSR CLASSIFIERS

In this section, we first introduce the kernel trick in Section IV-A. Afterwards, the proposed KCSR classifier is described in Section IV-B. Next, Section IV-C presents the proposed CKCSR classifier. In Section IV-D, KCSR and CKCSR vs methods in [30], [31] is described. Finally, KCSR and CKCSR vs KSRC is described in Section IV-E.

A. Kernel Trick

In machine learning, the kernel trick as a well-known technique utilizes a linear algorithm to its nonlinear counterpart without ever having to compute the mapping explicitly. Two popular kernels are polynomial kernel and Gaussian radial basis function (RBF) kernel. For simplicity, this paper uses the Gaussian radial basis function (RBF) kernel, which can be represented as

$$
k(x, y) = \phi(x)^T \phi(y) = \exp\left(-\frac{\|x - y\|^2}{\sigma}\right)
\tag{16}
$$

where x and y are two original samples, and σ is the parameter. In kernel methods, $\phi(*)$ is unknown. The only way to access the feature space is using $k(*, *)$.

Algorithm 1: Combined Sparse Representation (CSR) Classifier.

Require: The entire prototype set X with M class models $X_c \in R^{q \times N_c}$ for $c = 1, 2, \ldots, M$ and a test sample vector $x \in R^{q \times 1}$. Initialize the value of $E, e, \beta, y_1, Y_2, u, \rho, \varepsilon$ and $u_{\max}$

Ensure: The class index of x.

1: Use the entire prototype set X and test sample x to constitute the objective function as

$$\beta^* = \arg\min ||X^T x - X^T X \beta||_1 + \lambda ||X^T X Diag(\beta)||_*$$

2: Transform the problem in Step 1 to the Lagrange multiplier problem as follows and solve it with the while circle.

$$L(E, e, \beta) = \lambda ||E||_* + ||e||_1 + y_1^T (a - A\beta - e) + Tr[Y_2^T (E - ADiag(\beta))]$$
$$+ \frac{u}{2}(||a - A\beta - e||_2^2 + ||E - ADiag(\beta)||_F^2)$$

3: **while** until convergence **do**

4:     Update the E when the others are fixed by $E* = \arg\min_E \frac{\lambda}{u} ||E||_* + \frac{1}{2} ||E - (ADiag(\beta) - \frac{1}{u}Y_2)||_F^2$

5:     Update the β when the others are fixed by

$$\beta* = (A^T A + Diag(diag(A^T A)))^{-1} A^T (\frac{1}{u} y_1 + x - e)$$
$$+ (A^T A + Diag(diag(A^T A)))^{-1} diag(A^T (\frac{1}{u} Y_2 + E))$$

6:     Update the e when the others are fixed by $e* = \arg\min_e \frac{1}{u} ||e||_1 + \frac{1}{2} ||e - (x - A\beta + \frac{1}{u}y_1)||_2$

7:     Update the multipliers y_1 and Y_2 by $y_1 = y_1 + u(a - A\beta - e)$ and $Y_2 = Y_2 + u(E - ADiag(\beta))$

8:     Update the parameter u by $u = \min(\rho u, u_{\max})$

9:     Check the convergence conditions $||a - A\beta - e||_\infty \le \varepsilon$ and $||E - ADiag(\beta)||_\infty \le \varepsilon$

10: **end while**

11: Compute the sum of coefficients of each class by $s_c = \sum_{i=1}^{N_c} \beta_i^c, c = 1, 2, \ldots, M$.

12: Classify the test sample x into the class with the maximum sum of coefficients by $c* = \arg\max(s_c)$

B. Proposed KCSR Classifier

Suppose that there exists a nonlinear feature mapping function $\Phi(.) : R^q \to R^Q (q << Q)$, which maps the test sample x and the prototype set X to a high dimensional feature space as

$$x \to \Phi(x)$$
$$X \to \Phi(X) = [\Phi(x_1^1) \quad \ldots \quad \Phi(x_i^c) \quad \ldots \quad \Phi(x_{N_c}^M)]. \quad (17)$$

Similar to the proposed CSR classifier in Section III, the objective function of KCSR can be described as follows:

$$\min ||\Phi(X)^T \Phi(X) Diag(\gamma)||_*$$
$$s.t. \quad \Phi(X)^T \Phi(x) = \Phi(X)^T \Phi(X)\gamma. \quad (18)$$

The objective function in (18) can be also treated as the combination of the L_1-minimization and L_2- minimization. That is, the objective function in (18) can also benefit from both L_1-norm minimization and L_1-norm minimization according to the correction of the samples in the prototype set X.

The objective function in (18) is also designed for the situation that the samples have no noise. Considering noise, the objective function in (18) can be rewritten as

$$\min ||\Phi(X)^T \Phi(X) Diag(\gamma)||_*$$
$$s.t. \quad ||\Phi(X)^T \Phi(x) - \Phi(X)^T \Phi(X)\gamma||_1 \le \varepsilon. \quad (19)$$

Using the optimization approaches in [66] and [67], the method of inexact augmented Lagrange multipliers (IALM) [68] is used

to solve the minimization problem in (19), which can be converted to the following minimization problem:

$$\min_{H, h, \gamma} ||H||_* + ||h||_1$$
$$s.t. \quad h = \Phi(X)^T \Phi(x) - \Phi(X)^T \Phi(X)\gamma$$
$$H = \Phi(X)^T \Phi(X) Diag(\gamma). \quad (20)$$

Suppose that $K = \phi(X)^T \phi(X)$ and $k = \phi(X)^T \phi(x)$, which can be computed as follows:

$$K = \phi(X)^T \phi(X)$$
$$= \begin{bmatrix} k(x_1^1, x_1^1) & k(x_1^1, x_2^1) & \ldots & k(x_1^1, x_{N_c}^M) \\ k(x_2^1, x_1^1) & k(x_2^1, x_2^1) & \ldots & k(x_2^1, x_{N_c}^M) \\ \ldots & \ldots & \ldots & \ldots \\ k(x_{N_c}^M, x_1^1) & k(x_{N_c}^M, x_2^1) & \ldots & k(x_{N_c}^M, x_{N_c}^M) \end{bmatrix} \quad (21)$$

and

$$k = \phi(\hat{X})^T \phi(x) = [k(x_1^1, x) \quad k(x_2^1, x) \quad \ldots \quad k(x_{N_c}^M, x)]. \quad (22)$$

Algorithm 2: Kernel Combined Sparse Representation (KCSR) Classifier.

Require: The entire prototype set X with M class models $X_c \in R^{q \times N_c}$ for $c = 1, 2, \ldots, M$ and a test sample vector $x \in R^{q \times 1}$. Initialize the value of

Ensure: The class index of x.

1: Use the entire prototype set X and test sample x to constitute the kernel matrix K and test vector k by (21) and (22).

2: Define the objective function as

$$l\gamma^* = \arg\min ||\Phi(X^T)\Phi(x) - \Phi(X^T)\Phi(X)\gamma||_1$$
$$+ \lambda||\Phi(X^T)\Phi(X)Diag(\gamma)||_*$$

3: Transform the problem in Step 2 to the Lagrange multiplier problem. The solution is similar to Algorithm 1.

4: Compute the sum of coefficients of each class by

$$s_c = \sum_{i=1}^{N_c} \gamma_i^c \quad c = 1, 2, \ldots, M.$$

5: Classify the test sample x into the class with the maximum sum of coefficients by $c* = \arg\max(s_c)$.

Using the (21) and (22), the (20) can be rewritten as

$$\min_{H,h,\gamma} ||H||_* + ||h||_1$$
$$s.t. \quad h = k - K\gamma$$
$$H = KDiag(\gamma). \tag{23}$$

Afterwards, we can solve the following augmented Lagrange multiplier problem in (24) instead of solving the minimization problem in (23):

$$L(H,h,\gamma) = \lambda||H||_* + ||h||_1 + y_1^T(k - K\gamma - h)$$
$$+ Tr[Y_2^T(H - KDiag(\gamma))]$$
$$+ \frac{u}{2}(||k - K\gamma - h||_2^2 + ||H - KDiag(\gamma)||_F^2) \tag{24}$$

where $u > 0$ is a parameter, y_1 and Y_2 are Lagrange multipliers. Similar to CSR classifier, variables H, h and γ in (24) can be optimized alternatively with the other two fixed. The detailed classification procedures of KCSR classifier are summarized in Algorithm 2.

C. Proposed KCSR Classifier

Let m_c be the mean sample of the cth class. A novel prototype-set can be described as

$$\hat{X}_c = [x_1^c \; x_2^c \; \ldots \; x_{N_c}^c \; m_c] \in R^{q \times (N_c+1)}. \tag{25}$$

Form all-class-subspaces model as

$$\hat{X} = [\hat{X}_1 \; \hat{X}_2 \; \ldots \; \hat{X}_M] \in R^{q \times M(N_c+1)}. \tag{26}$$

Similar to KCSR, CKCSR Maps the test sample x and the prototype set to a high dimensional feature space as

$$x \rightarrow \Phi(x)$$
$$\hat{X} \rightarrow \Phi(\hat{X}) = [\Phi(x_1^1) \quad \ldots \quad \Phi(x_i^c) \quad \ldots \quad \Phi(x_{N_c+1}^M)]. \tag{27}$$

Similar to the proposed KCSR classifier, the objective function of CKCSR can be described as follows:

$$\min ||\Phi(\hat{X})^T \Phi(\hat{X})Diag(\xi)||_*$$
$$s.t. \quad \Phi(\hat{X})^T \Phi(x) = \Phi(\hat{X})^T \Phi(\hat{X})\xi. \tag{28}$$

The objective function in (28) can be also treated as the combination of the L_1-minimization and L_2-minimization. Similarly, the objective function in (28) is also designed for the situation that the samples have no noise. Considering noise, the objective function in (28) can be rewritten as

$$\min ||\Phi(X)^T \Phi(X)Diag(\beta)||_*$$
$$s.t. \quad ||\Phi(X)^T \Phi(x) - \Phi(X)^T \Phi(X)\beta||_1 \leq \varepsilon. \tag{29}$$

Using the optimization approaches in [66] and [67], the method of inexact augmented Lagrange multipliers (IALM) [68] is used to solve the minimization problem in (29), which can be converted to the following minimization problem:

$$\min_{P,p,\xi} ||P||_* + ||p||_1$$
$$s.t. \quad p = \Phi(\hat{X})^T \Phi(x) - \Phi(\hat{X})^T \Phi(\hat{X})\xi$$
$$P = \Phi(\hat{X})^T \Phi(\hat{X})Diag(\xi). \tag{30}$$

Suppose that $\hat{K} = \phi(\hat{X})^T \phi(\hat{X})$ and $\hat{k} = \phi(\hat{X})^T \phi(x)$ can be computed as follows:

$$\hat{K} = \phi(\hat{X})^T \phi(\hat{X})$$
$$= \begin{bmatrix} k(x_1^1, x_1^1) & k(x_1^1, x_2^1) & \ldots & k(x_1^1, x_{N_c+1}^M) \\ k(x_2^1, x_1^1) & k(x_2^1, x_2^1) & \ldots & k(x_2^1, x_{N_c+1}^M) \\ \ldots & \ldots & \ldots & \ldots \\ k(x_{N_c+1}^M, x_1^1) & k(x_{N_c+1}^M, x_2^1) & \ldots & k(x_{N_c+1}^M, x_{N_c+1}^M) \end{bmatrix} \tag{31}$$

and

$$\hat{k} = \phi(\hat{X})^T \phi(x)$$
$$= [k(x_1^1, x) \; k(x_2^1, x) \; \ldots \; k(x_{N_c+1}^M, x)]. \tag{32}$$

Using the (31) and (32), the (30) can be rewritten as

$$\min_{P,p,\xi} ||P||_* + ||p||_1$$
$$s.t. \quad p = \hat{k} - \hat{K}\xi$$
$$P = \hat{K}Diag(\xi). \tag{33}$$

Next, instead of solving the minimization problem in (33), we can solve the following augmented Lagrange multiplier problem

---

Algorithm 3: Center-based Kernel Combined Sparse Representation (CKCSR) Classifier.

Require: The entire prototype set X with M class models $X_c \in R^{q \times N_c}$ for $c = 1, 2, \ldots, M$ and a test sample vector $x \in R^{q \times 1}$. Initialize the value of

Ensure: The class index of x.

1: Use the entire prototype set X to constitute the center-based entire prototype set $\hat{X}$ by (26)

2: Utilize the center-based entire prototype set $\hat{X}$ to constitute center-based kernel matrix $\hat{K}$ and test vector $\hat{k}$ by (31) and (32), respectively.

3: Define the objective function as

$$\xi^* = \arg \min ||\Phi(\hat{X}^T)\Phi(x) - \Phi(\hat{X}^T)\Phi(\hat{X})\xi||_1 \\ + \lambda ||\Phi(\hat{X}^T)\Phi(\hat{X})Diag(\xi)||_*$$

4: Transform the problem in Step 3 to the Lagrange multiplier problem. The solution is similar to Algorithm 1.

5: Compute the sum of coefficients of each class by

$$s_c = \sum_{i=1}^{N_c+1} \xi_i^c, c = 1, 2, \ldots, M.$$

6: Classify the test sample x into the class with the maximum sum of coefficients by $c* = \arg \max(s_c)$.

---

in (34) as

$$L(P, p, \xi) = \lambda ||P||_* + ||p||_1 + y_1^T (\hat{k} - \hat{K}\xi - p) \\ + Tr[Y_2^T (P - \hat{K}Diag(\xi))] \\ + \frac{u}{2}(||\hat{k} - \hat{K}\beta - p||_2^2 + ||P - \hat{K}Diag(\xi)||_F^2) \tag{34}$$

where $u > 0$ is a parameter, y_1 and Y_2 are Lagrange multipliers. Variables P, p and ξ in (34) can be optimized alternatively with the other two fixed. The detailed classification procedures of CKCSR classifier are summarized in Algorithm 3.

D. KCSR and CKCSR Versus Methods in [30] and [31]

Similar to the CSR classifier, the proposed KCSR and CKCSR try to minimize the correlation structure of the entire training set multiplied by its transposition, and the sparse coefficient together for classification. Therefore, the first difference among KCSR, CKCSR and methods in [30], [31] is that KCSR and CKCSR use the entire training set multiplied by its transposition while the methods in [30], [31] use the entire training set. Moreover, KCSR and CKCSR utilize the nonlinear information for classification instead of using the linear information in [30], [31]. Besides, CKCSR increases the nonlinear information of the training samples and the class center while methods in [30], [31] and KCSR fail to do so.

E. KCSR and CKCSR Versus KSRC

They all use the nonlinear information for classification. The main difference among KCSR, CKCSR and KSRC is that KCSR and CKCSR try to minimize the correlation structure of the entire training set multiplied by its transposition and the sparse coefficient together for classification while the KSRC considers only the sparsity for classification. Besides, CKCSR increases the nonlinear information of the training samples and the class center while KSRC and KCSR fail to consider this.

V. EXPERIMENTAL RESULTS

To assess the effectiveness of the proposed methods, the proposed CSR, KCSR and CKCSR classifiers are compared with the existed methods: LRC [12], SRC [17], CRC [39], DPL [27], FDDL [28], SSRC [44], ASRC [31], NFLS-I [35], NFLS-II [35], LMLE [1], KSRC [45], KCRC [46], KLRC [47], CWKLRC [48] and KCRC-TR [10] with five medical disease databases. In the experiments, the parameter for the minimization optimization problem is set 0.001. The parameter $\rho, u, u_{\max}$ for ADM solver are set as 1.4, 1e-8 and 1e-6. ϵ is set as 1e-9. The max number of iterations is set as 50. The parameter for the RBF kernel is set as

$$\sigma = \frac{\sum_{c=1}^M \sum_{i=1}^{N_c} ||x_i^c - m||^2}{\sum_{c=1}^M \sum_{i=1}^{N_c} 10}$$

where m denotes the mean sample of all the training samples. In the following experiments, the training and testing samples are randomly selected. We repeat the experiments 20 times and report their averages as the final results.

A. Experiment on EXACT09 Database

Extraction of Airways from CT 2009 (EXACT09) database is a dataset of chest CT scans [71]. This database uses the DICOM format to store the CT scan images. Following [72], 675 CT images of CASE23 of testing set of EXACT09 are used in this experiment. It includes 19 categories including 36, 23, 30, 30, 50, 42, 20, 45, 50, 24, 28, 24, 35, 40, 50, 35, 30, 28 and 55 CT images, respectively. Each CT image of this database is 512×512. Fig. 1(a) shows several images of this database.

In this experiment, all images of EXACT09 database are manually resized into 128×128. Two cases are used. Case 1: Each class selects four images for training, and the rest images are used as the test set. Case 2: The training set chooses five images from each class, the rest images are used as the test set. Afterwards, the PCA feature [73] is used in the experiment. The chosen Dims of the PCA feature are 50, 100, 150, 200, 250, 300, 350, 400, 450, and 500, respectively. The proposed CSR, KCSR and CKCSR classifiers are compared with the LRC [12], SRC [17], CRC [39], DPL [27], FDDL [28], SSRC [44], ASRC [31], NFLS-I [35], NFLS-II [35], LMLE [1], KSRC [45], KCRC [46], KLRC [47], KCRC-TR [10] and CWKLRC [48] classifiers. The experimental results are exhibited in Fig. 2. The max and mean recognition rates over the various Dims in Fig. 2 are shown in Table I. From the results, the outcomes on EXACT09 database can be summarized as follows.
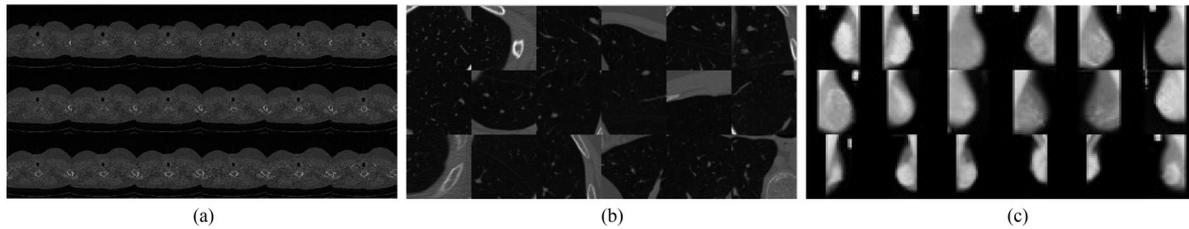
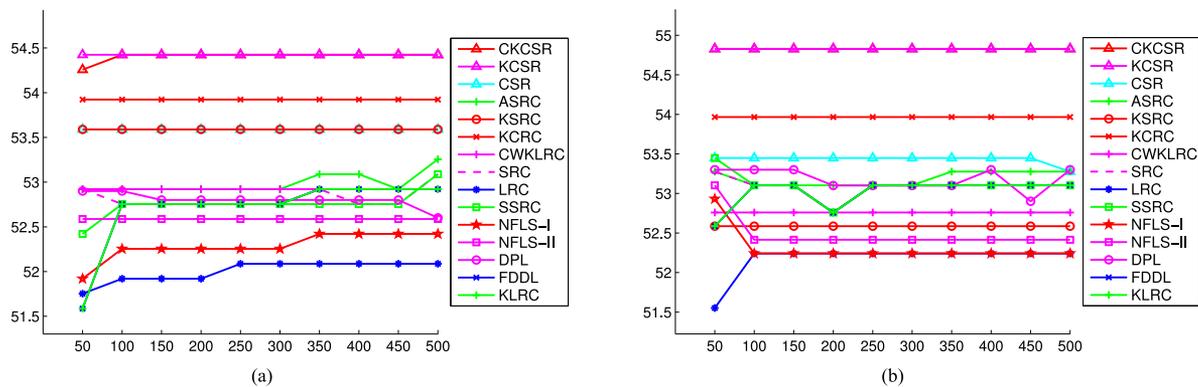Fig. 1. Some images from the (a) EXAC09 database, (b) EMPHYSEMA-CT database, and (c) mini-MIAS database.

Fig. 2. Performance of several classifiers with various Dims of PCA feature on EXACT09 database. The horizontal axis denotes the chosen Dim of PCA feature.

TABLE I
RECOGNITION RATES (RRS) AND AVERAGED RRS OF
SEVERAL CLASSIFIERS ON EXACT09 DATABASE

Type	Classifier	Case 1		Case 2	
		Max RR	Mean RR	Max RR	Mean RR
Linear-based	LRC	52.09	52.00	52.24	52.17
	SRC	53.09	52.83	53.28	53.09
	CRC	53.59	48.16	53.45	48.45
	DPL	52.90	52.80	53.30	53.18
	FDDL	52.92	52.70	53.10	53.02
	ASRC	53.26	52.99	53.28	53.19
	SSRC	53.09	52.76	53.45	53.10
	NFLS-I	52.42	52.29	52.93	52.31
	NFLS-II	52.59	52.59	53.10	52.48
	LMLE	53.09	52.80	53.30	53.10
Kernel-based	KSRC	53.59	53.59	52.59	52.59
	KCRC	53.93	53.93	53.97	53.87
	KLRC	52.92	52.92	52.76	52.76
	CWKLRC	52.92	52.92	52.76	52.76
	KCRC-TR	54.09	54.09	54.10	54.88
Proposed	CSR	53.59	53.59	53.45	53.43
	KCSR	54.24	54.24	54.83	54.83
	CKCSR	54.24	54.41	54.83	54.83

1) The proposed KCSR and CKCSR classifiers obtain the best recognition rates among all the classifiers.
2) The proposed CSR classifiers obtain the best recognition rates among linear-based classifiers.
3) The linear-based classifiers and the kernel-based classifiers obtain the competitive performance.
4) The kernel-based classifiers with the various dimensions have similar performance. However, the performance of the linear-based classifiers with the various dimensions is different.

B. Experiment on Emphysema-CT Database

Emphysema is characterized by loss of lung tissue and the recognition of healthy and emphysematous lung tissue is quite useful for analyzing the disease [74], [75]. Three categories are included in the Emphysema-CT database. They are Normal Tissue (NT), Centrilobular Emphysema (CLE), and Paraseptal Emphysema (PSE). These three categories collected from 39 subjects have 59, 50 and 59 images, respectively. Fig. 1(b) shows several selected images of this database.

In this experiment, all images in Emphysema-CT database are manually cropped into 128×128. Case 1: Each class selects four images for training, and the rest images are used as the test set. Case 2: The training set chooses five images from each class, the rest images are used as the test set. All cropped images are transformed to the PCA feature [73] for the experiment. The chosen Dims of the PCA feature are 50, 100, 150, 200, 250, 300, 350, 400, 450, and 500, respectively. The proposed CSR, KCSR and CKCSR classifiers are compared with the LRC [12], SRC [17], CRC [39], DPL [27], FDDL [28], SSRC [44], ASRC [31], NFLS-I [35], NFLS-II [35], LMLE [1], KSRC [45], KCRC [46], KLRC [47], KCRC-TR [61] and CWKLRC [48] classifiers. The experimental results are exhibited in Fig. 3. The maximum and mean recognition rates over the various Dims in Fig. 3 are shown in Table II. The conclusions for the experiment results on the Emphysema-CT database can be drawn as follows.

1) The proposed CKCSR classifier obtains the best performance among all classifiers. It has the significantly improvement contrasted to the most compared methods.
2) The proposed CSR classifier only obtains the intermediate performance among these linear-based classifiers.
3) The kernel-based classifiers obtain better performance compared to the linear-based classifiers.
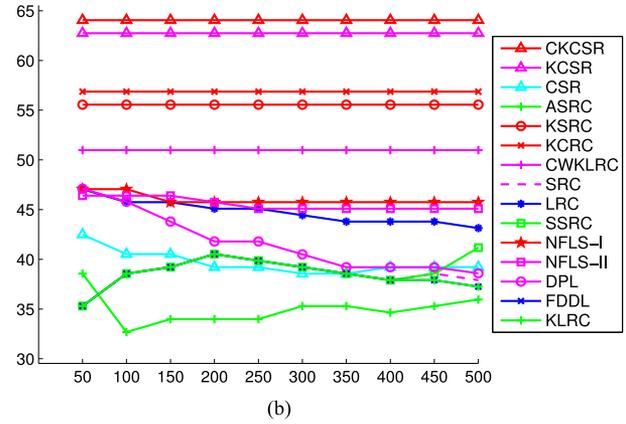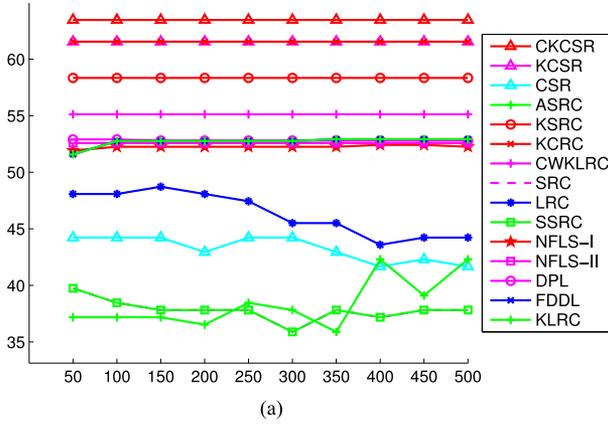
Fig. 3.   Performance of several classifiers with various Dims of PCA feature on EMPHYSEMA-CT database. The horizontal axis denotes the chosen Dim of PCA feature.

TABLE II
RECOGNITION RATES (RRS) AND AVERAGED RRS OF SEVERAL CLASSIFIERS ON EMPHYSEMA-CT DATABASE

Type	Classifier	Case 1		Case 2	
		Max RR	Mean RR	Max RR	Mean RR
Linear-based	LRC	48.72	46.35	47.06	44.77
	SRC	39.74	37.82	40.52	38.56
	CRC	43.59	41.99	41.18	40.52
	DPL	52.90	52.82	47.10	41.70
	FDDL	52.92	52.70	40.52	38.43
	ASRC	42.31	38.40	38.56	34.97
	SSRC	39.74	37.82	41.18	40.52
	NFLS-I	52.42	52.25	47.06	46.01
	NFLS-II	52.59	52.59	46.41	45.56
	LMLE	54.92	54.25	53.52	52.56
Kernel-based	KSRC	58.33	58.33	55.56	55.56
	KCRC	61.54	61.54	56.86	56.86
	KLRC	52.92	52.92	50.98	50.98
	CWKLRC	55.13	55.13	50.98	50.98
	KCRC-TR	60.50	60.50	58.98	58.98
Proposed	CSR	44.23	43.27	42.48	39.67
	KCSR	61.54	61.54	62.75	62.75
	CKCSR	63.46	63.46	64.05	64.05

4) The performance of the kernel-based classifiers with the various dimensions is similar while the performance of the linear-based classifiers with the various dimensions is different.

C. Experiment on Mini-MIAS Database

The mini-MIAS database [76] is developed by the Mamography Images Analysis Society. The X-ray films of this database have been digitized with a Joyce-Lobel scanning microdensitometer to a resolution of 50×50 um, 8-bit word. Each image of the mini-MIAS database is 1024×1024 pixels. Fig. 1(c) shows several selected images of this database.

In this experiment, all images of mini-MIAS database are manually resized into 60×60. Each class selects four images for training, the rest images are used as the test set. All resized images are transformed to the PCA feature [73],[21] for the experiment. The chosen Dims of the PCA feature are 50, 100, 150, 200, 250, 300, 350, 400, 450, and 500, respectively. The
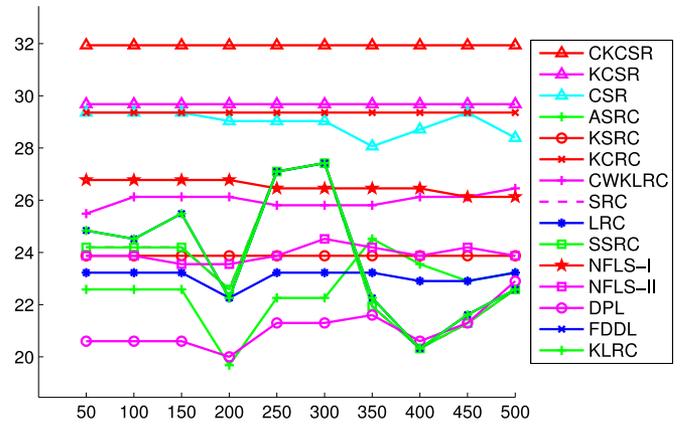

Fig. 4.   Performance of several classifiers with various Dims of PCA feature on mini-MIAS database. The horizontal axis denotes the chosen Dim of PCA feature.

proposed CSR, KCSR and CKCSR classifiers are compared with the LRC [12], SRC [17], CRC [39], DPL [27], FDDL [28], SSRC [44], ASRC [31], NFLS-I [35], NFLS-II [35], LMLE [1], KSRC [45], KCRC [46], KLRC [47], KCRC-TR [10] and CWKLRC [48] classifiers. The experimental results are exhibited in Fig. 4 and Table III. The outcomes can be drawn as follows.

1) The proposed CKCSR classifier obtains the best recognition rates among all the classifiers and has the significantly improvement over the most methods.
2) The proposed CSR classifier obtains the best recognition rates among linear-based classifiers.
3) The linear-based and the kernel-based classifiers obtain the competitive performance.
4) The performance of the kernel-based classifiers with the various dimensions is similar while the performance of the linear-based classifiers with the various dimensions is different.

D. Experiment on Two UCI Databases

The cardiac Single Proton Emission Computed Tomography (SPECT) database [77] contains two categories: Normal and Abnormal, which describes diagnosing of cardiac Single Proton

TABLE III
RECOGNITION RATES (RRS) AND AVERAGED RRS OF
SEVERAL CLASSIFIERS ON MINI-MIAS DATABASE

Type	Classifier	Max RR	Mean RR
Linear-based	LRC	23.23	23.06
	SRC	27.42	23.58
	CRC	25.48	23.65
	DPL	22.90	21.08
	FDDL	27.42	23.84
	ASRC	24.52	22.61
	SSRC	27.42	23.65
	NFLS-I	26.77	26.52
	NFLS-II	24.52	23.94
	LMLE	28.48	26.61
Kernel-based	KSRC	23.87	23.87
	KCRC	29.36	29.36
	KLRC	21.94	21.94
	CWKLRC	26.45	26.00
	KCRC-TR	30.00	30.00
Proposed	CSR	29.36	28.97
	KCSR	29.68	29.68
	CKCSR	31.94	31.94

TABLE IV
RECOGNITION RATES (RRS) OF THE CLASSIFIERS ON
WBC DATABASE AND HD-SPECTF DATABASE

Type	Classifier	WBC		HD-SPECTF	
		Case 1	Case 2	Case 1	Case 2
Linear-based	LRC	75.33	62.26	64.71	66.80
	SRC	63.49	68.22	41.96	35.22
	CRC	71.86	70.68	41.18	35.63
	DPL	67.80	70.83	53.33	56.30
	ASRC	60.03	69.81	43.92	44.13
	SSRC	63.49	64.54	41.96	35.22
	NFLS-I	90.33	72.42	61.60	66.40
	NFLS-II	74.46	70.83	61.37	62.87
Kernel-based	KSRC	88.89	86.36	58.43	68.42
	KCRC	89.61	89.99	63.67	69.47
	Triplet-SVM	89.47	86.94	56.86	61.54
	Doublet-SVM	92.67	86.65	54.11	55.87
Proposed	CSR	64.65	71.11	49.02	53.85
	KCSR	92.50	91.15	65.88	70.04
	CKCSR	93.36	92.31	67.45	70.45

TABLE V
AVERAGED RECOGNITION RATES OF THE CLASSIFIERS
WITH VARIOUS KERNEL ON EXACT09 DATABASE

Classifier	Polynomial kernel	Laplacian kernel	Cauchy kernel	Log Kernel	Power kernel
KSRC	51.72	53.27	53.44	53.28	53.28
KCRC	51.24	52.45	53.79	53.10	53.10
KLRC	51.24	52.93	52.93	53.10	53.10
CWKLRC	51.24	52.76	52.93	53.10	53.10
KCRC-TR	51.90	52.97	54.00	53.28	53.28
KCSR	52.07	53.62	54.14	55.17	55.17
CKCSR	52.07	53.79	54.31	55.35	55.35

Emission Computed Tomography (SPECT) images. The 267 SPECT image set (patients) of this database was processed to extract features instead of the original SPECT images. As a result, 44 continuous feature patterns were created for each patient. Two cases are used in this experiments. Case 1: Each class selects six samples for training, the rest samples are used as the test set. Case 2: The training set chooses ten samples from each class, the rest samples are used as test set. The Wisconsin Breast Cancer (WBC) database [78] was developed by the University of Wisconsin Hospitals. This databases samples are collected periodically in the clinical cases. There are 699 samples in the database, which belong to two classes: Benign and Malignant. In the experiment, we uses two cases. Case 1: Three samples are chosen from each class for training, the rest samples are used as the test set. Case 2: The training set selects five samples from each class, the rest samples are used as test set.

In the experiment, the proposed CSR, KCSR and CKCSR classifiers are compared with the LRC [12], SRC [17], CRC [39], DPL [27], SSRC [44], ASRC [31], NFLS-I [35], NFLS-II [35], KSRC [45], KCRC [47], Triplet-SVM [49] and Doublet-SVM [49] classifiers. The experimental results are exhibited in Table IV. The outcomes on the basis of the experiment results on these two database can be drawn as follows.

1) The proposed CKCSR classifier obtains the best recognition rates among all the classifiers.
2) The proposed CSR classifier obtains the competitive performance among linear-based classifiers.
3) The linear-based classifiers gain the worse performance than the kernel-based classifiers.

E. Evaluation on Various Kernel Functions

In above experiments, KCSR and CKCSR classifiers use the Gaussian radial basis function (RBF) to project the original space into a high dimensional space. In this Section, we provide experiment results to compare the performance of several classifiers using other kernels [79], such as the polynomial kernel,

Laplacian Kernel, Cauchy Kernel [80], Log Kernel and Power Kernel. The experiment results are shown in Table V. These kernel are described as follows. Given two original samples x and y.

1) *Polynomial Kernel:* The Polynomial kernel is a non-stationary kernel. Its kernel function is defined as

$$k(x, y) = (\alpha x^T y + c)^d$$

where, c is a constant term and d is the polynomial degree. In our experiment, c is set to zero and d is set to 3.

2) *Laplacian Kernel:* This kernel is closely relative to the Gaussian kernel (RBF). Its kernel function is defined as

$$k(x, y) = \exp\left(-\frac{||x - y||}{\sigma}\right)$$

where, σ is a parameter. In this experiment σ is set to the square root of the parameter of the RBF kernel.

3) *Cauchy Kernel:* The Cauchy kernel is based on the Cauchy distribution. Its kernel function is defined as

$$k(x, y) = \frac{1}{1 + \sigma ||x - y||^2}$$

where, σ is a parameter. In this experiment σ is set to the parameter of the RBF kernel.

4) *Log Kernel:* The kernel function of the Log kernel is defined as

$$k(x, y) = -\log(||x - y||^d + 1)$$

where, d is a parameter. In this experiment d is set to 3.

5) *Power Kernel:* The kernel function of the Power kernel is defined as

$$k(x, y) = -||x - y||^d$$

where, d is a parameter. In this experiment d is set to 3.

In our experiment, the proposed KCSR and CKCSR classifiers are compared with the KSRC [45], KCRC [46], KCRC-TR [10], KLRC [47] and CWKLRC [48] classifiers when various kernels are used. The experimental set follow Section V-A. The experimental results are exhibited in Table V. The outcomes of the experiment results over the EXACT09 database can be concluded as follows.

1) The proposed KCSR and CKCSR classifiers obtain the better recognition rates than several kernel-based methods on various kernel functions.

2) The performance of all the classifiers with various kernel functions is different.

VI. CONCLUSION

In this paper, a novel combined sparse representation (CSR) classifier and its kernel version, kernel CSR (KCSR) and center-based kernel CSR (CKCSR), have been proposed for disease recognition. The proposed three approaches consider the correlation structure of the entire prototype set multiplied by its transposition, for classification. The proposed methods can be treated as a combination of L_1-minimization and L_2- minimization such that it can benefit from the L_1-minimization and L_2- minimization. Experimental results have shown that the proposed CSR, KCSR and CKCSR classifiers achieve better recognition rates than the well-known SRC, CRC, and several state-of-the-art classifiers. The analyses and the experimental results on several famous medical databases have confirmed the effectiveness of the proposed classifiers for disease recognition.

APPENDIX A
THE DERIVATION PROCEDURE OF (3)

Suppose that the subjects of X are similar to x_1, the first column of X, that is $X^T X = 11^T$ (1 is a vector of size $n = \sum_{c=1}^{M} N_c$, where each element is one). The objective function can be re-written as

$$||X^T X Diag(\beta)||_* = ||11^T Diag(\beta)||_*. \tag{35}$$

For matrix $B \in R^{m \times t}$ of rank r, we know the following inequalities [81], [82] as [1]

$$||B||_F \leq ||B||_* \leq \sqrt{r}||B||_F. \tag{36}$$

It is easy to know that the rank of matrix $11^T Diag(\beta)$ is one. That is $r = 1$. With (35)-(36), the objective function will become

$$||X^T X Diag(\beta)||_* = ||11^T Diag(\beta)||_F. \tag{37}$$

[1][Online]. Available: https://en.wikipedia.org/wiki/Matrix_norm

The definition of Frobenius norm is as

$$||B||_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{t} |b_{ij}|^2}. \tag{38}$$

With (37)-(38), the objective function will be

$$||X^T X Diag(\beta)||_* = \sqrt{n}||\beta||_2. \tag{39}$$

Notice: For a specific classification task, the training set is fixed. That is, the number of training samples n is a fixed value. Thus, minimization of $\sqrt{n}||\beta||_2$ is equivalent to minimization of $||\beta||_2$.

REFERENCES

[1] Y. Song *et al.*, "Large margin local estimate with applications to medical image classification," *IEEE Trans. Med. Imag.*, vol. 34, no. 6, pp. 1362–1377, Jun. 2015.

[2] H. Kong, Z. Lai, X. Wang, and F. Liu, "Breast cancer discriminant feature analysis for diagnosis via jointly sparse learning," *Neurocomputing*, vol. 177, pp. 198–205, 2016.

[3] Y. Song, W. Cai, H. Huang, Y. Wang, and D. D. Feng, "Object localization in medical images based on graphical model with contrast and interest-region terms," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog. Workshops*, Jun. 2012, pp. 1–7.

[4] S. Zhang, Y. Zhan, and D. N. Metaxas, "Deformable segmentation via sparse representation and dictionary learning," *Med. Image Anal.*, vol. 16, no. 7, pp. 1385–1396, 2012.

[5] S. Liu *et al.*, "Localized sparse code gradient in alzheimer's disease staging," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2013, pp. 5398–5401.

[6] A. Depeursinge *et al.*, "Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 4, pp. 665–675, Jul.2012.

[7] Y. Song, W. Cai, Y. Zhou, and D. D. Feng, "Feature-based image patch approximation for lung tissue classification," *IEEE Trans. Med. Imaging*, vol. 32, no. 4, pp. 797–808, Apr. 2013.

[8] F. Zhang *et al.*, "Lung nodule classification with multilevel patch-based context analysis," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 4, pp. 1155–1166, Apr. 2014.

[9] E. Parrado-Hernández "Discovering brain regions relevant to obsessive–compulsive disorder identification through bagging and transduction," *Med. Image Anal.*, vol. 18, no. 3, pp. 435–448, 2014.

[10] M. Yaqub, M. K. Javaid, C. Cooper, and J. A. Noble, "Investigation of the role of feature selection and weighted voting in random forests for 3-d volumetric segmentation," *IEEE Trans. Med. Imaging*, vol. 33, no. 2, pp. 258–271, Feb. 2014.

[11] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 1, pp. 21–27, Jan. 1967.

[12] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2106–2112, Nov. 2010.

[13] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally linear regression for pose-invariant face recognition," *IEEE Trans. Image Process.*, vol. 16, no. 7, pp. 1716–1725, Jul. 2007.

[14] Q. Feng and Y. Zhou, "Iterative linear regression classification for image recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2016, pp. 1566–1570.

[15] Q. Feng, Q. Zhu, L. Tang and J. Pan, "Double linear regression classification for face recognition," *J. Mod. Opt.*, vol. 62, no. 4, pp. 288–295, 2015.

[16] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[17] J. Wright *et al.*, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[18] J. Yang, D. Chu, L. Zhang, Y. Xu, and J. Yang, "Sparse representation classifier steered discriminative projection with applications to face recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 7, pp. 1023–1035, Jul. 2013.

[19] L. Ma, C. Wang, B. Xiao, and W. Zhou, "Sparse representation for face recognition based on discriminative low-rank dictionary learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 2586–2593.

[20] A. V. Nefian, "Embedded bayesian networks for face recognition," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Aug. 2002, vol. 2, pp. 133–136.

[21] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image Vis. Comput.*, vol. 28, no. 6, pp. 902–913, 2010.

[22] X. Fang, Y. Xu, X. Li, Z. Lai, and W. K. Wong, "Learning a nonnegative sparse graph for linear regression," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2760–2771, Sep. 2015.

[23] Z. Lai, W. K. Wong, Y. Xu, C. Zhao, and M. Sun, "Sparse alignment for robust tensor learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 10, pp. 1779–1792, Oct. 2014.

[24] Z. Lai, Y. Xu, Q. Chen, J. Yang, and D. Zhang, "Multilinear sparse principal component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 10, pp. 1942–1950, Oct. 2014.

[25] Q. Feng *et al.*, "Superimposed sparse parameter classifiers for face recognition," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2016.2516239.

[26] Y. Xu *et al.*, "Using the idea of the sparse representation to perform coarse-to-fine face recognition," *Inf. Sci.*, vol. 238, pp. 138–148, 2013.

[27] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Projective dictionary pair learning for pattern classification," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 793–801.

[28] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based fisher discrimination dictionary learning for image classification," *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 209–232, 2014.

[29] S. Gao, I. W.-H. Tsang, and L.-T. Chia, " Sparse representation with kernels," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 423–434, Feb. 2013.

[30] E. Grave, G. R. Obozinski, and F. R. Bach, "Trace lasso: A trace norm regularization for correlated designs," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 2187–2195.

[31] J. Wang *et al.*, "Robust face recognition via adaptive sparse representation," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2368–2378, Dec. 2014.

[32] Z. Yu *et al.*, "Hybrid k -nearest neighbor classifier," *IEEE Trans. Cybern.*, vol. 46, no. 6, pp. 1263–1275, Jun. 2016.

[33] S. Z. Li and J. Lu, "Face recognition using the nearest feature line method," *IEEE Trans. Neural Netw.*, vol. 10, no. 2, pp. 439–443, Mar. 1999.

[34] S. Z. Li, K. L. Chan, and C. Wang, "Performance evaluation of the nearest feature line method in image classification and retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1335–1349, Nov. 2000.

[35] J.-S. Pan, Q. Feng, L. Yan, and J.-F. Yang, "Neighborhood feature line segment for image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 3, pp. 387–398, Mar. 2015.

[36] J.-T. Chien and C.-C. Wu, "Discriminant waveletfaces and nearest feature classifiers for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 12, pp. 1644–1649, Dec. 2002.

[37] Q. Feng, J.-S. Pan, and L. Yan, "Two classifiers based on nearest feature plane for recognition," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 3216–3219.

[38] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1753–1766, May 2013.

[39] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 471–478.

[40] M. Yang, L. Zhang, D. Zhang, and S. Wang, "Relaxed collaborative representation for pattern classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 2224–2231.

[41] S. Cai, L. Zhang, W. Zuo, and X. Feng, "A probabilistic collaborative representation based approach for pattern classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 2950–2959.

[42] Y. Xu, D. Zhang, J. Yang, and J.-Y. Yang, "A two-phase test sample sparse representation method for use with face recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 9, pp. 1255–1262, Sep. 2011.

[43] Y. Xu *et al.*, "Data uncertainty in face recognition," *IEEE Trans. Cybern.*, vol. 44, no. 10, pp. 1950–1961, Oct. 2014.

[44] W. Deng, J. Hu, and J. Guo, "In defense of sparsity based face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2013, pp. 399–406.

[45] L. Zhang *et al.*, "Kernel sparse representation-based classifier," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1684–1695, Apr. 2012.

[46] B. Wang, W. Li, N. Poh, and Q. Liao, "Kernel collaborative representation-based classifier for face recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 2877–2881.

[47] Y. Lu, X. Fang, and B. Xie, "Kernel linear regression for face recognition," *Neural Comput. Appl.*, vol. 24, no. 7/8, pp. 1843–1849, 2014.

[48] Q. Feng, C. Yuan, J. Huang, and W. Li, "Center-based weighted kernel linear regression for image classification," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2015, pp. 3630–3634.

[49] F. Wang, W. Zuo, L. Zhang, D. Meng, and D. Zhang, "A kernel classification framework for metric learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 1950–1962, Sep. 2015.

[50] W. Li, Q. Du, and M. Xiong, "Kernel collaborative representation with tikhonovregularization for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 48–52, Jan. 2015.

[51] Y.-T. Chou, S.-M. Huang, and J.-F. Yang, "Class-specific kernel linear regression classification for face recognition under low-resolution and illumination variation conditions," *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, pp. 1–9, 2016.

[52] L. Zhang, W.-D. Zhou, and F.-Z. Li, "Kernel sparse representation-based classifier ensemble for face recognition," *Multimedia Tools Appl.*, vol. 74, no. 1, pp. 123–137, 2015.

[53] Z. Yu, L. Li, J. Liu, and G. Han, "Hybrid adaptive classifier ensemble," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 177–190, Feb. 2015.

[54] Z. Yu *et al.*, "Double selection based semi-supervised clustering ensemble for tumor clustering from gene expression profiles," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 11, no. 4, pp. 727–740, Jul./Aug. 2014.

[55] Z. Yu, L. Li, J. Liu, J. Zhang, and G. Han, "Adaptive noise immune cluster ensemble using affinity propagation," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 12, pp. 3176–3189, Dec. 2015.

[56] Z. Yu *et al.*, "Incremental semi-supervised clustering ensemble for high dimensional data clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 701–714, Mar. 2016.

[57] A. Shrivastava, V. M. Patel, and R. Chellappa, "Multiple kernel learning for sparse representation-based classification," *IEEE Trans. Image Process.*, vol. 23, no. 7, pp. 3013–3024, Jul. 2014.

[58] X. Lan, A. J. Ma, P. C. Yuen, and R. Chellappa, "Joint sparse representation and robust feature-level fusion for multi-cue visual tracking," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5826–5841, Dec. 2015.

[59] X. Lan, S. Zhang, and P. C. Yuen, "Robust joint discriminative feature learning for visual tracking," in *Proc. Int. Joint Conf. Artif. Intell.*, 2016, pp. 3403–3410.

[60] H. Xu, J. Zheng, A. Alavi, and R. Chellappa, "Learning a structured dictionary for video-based face recognition," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2016, pp. 1–9.

[61] T. Tong *et al.*, "Segmentation of mr images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling," *NeuroImage*, vol. 76, pp. 11–23, 2013.

[62] L. Wang *et al.*, "Integration of sparse multi-modality representation and anatomical constraint for isointense infant brain mr image segmentation," *NeuroImage*, vol. 89, pp. 152–164, 2014.

[63] Y. Song *et al.*, "Boosted multifold sparse representation with application to ILD classification," in *Proc. IEEE 11th Int. Symp. Biomed. Imaging*, 2014, pp. 1023–1026.

[64] N. Weiss, D. Rueckert, and A. Rao, "Multiple sclerosis lesion segmentation using dictionary learning and sparse coding," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2013, pp. 735–742.

[65] S. Zhang *et al.*, "Sparse representation of higher-order functional interaction patterns in task-based fmri data," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervention*, 2013, pp. 626–634.

[66] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, 2011, Art. no. 11.

[67] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 663–670.

[68] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *CoRR*, 2010. [Online]. Available: http://arxiv.org/abs/1009.5055

[69] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[70] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for |1 minimization: Methodology and convergence," *SIAM J. Optim.*, vol. 19, no. 3, pp. 1107–1130, 2008.

[71] P. Lo *et al.*, "Extraction of airways from CT (EXACT'09)," *IEEE Trans. Med. Imag.*, vol. 31, no. 11, pp. 2093–2107, Nov. 2012.

[72] S. R. Dubey, S. K. Singh, and R. K. Singh, "Local wavelet pattern: A new feature descriptor for image retrieval in medical CT databases," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5892–5903, Dec. 2015.

[73] A. M. Martínez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 228–233, Feb. 2001.

[74] L. Sorensen, S. B. Shaker, and M. De Bruijne, "Quantitative analysis of pulmonary emphysema using local binary patterns," *IEEE Trans. Med. Imaging*, vol. 29, no. 2, pp. 559–569, Feb. 2010.

[75] S. R. Dubey, S. K. Singh, and R. K. Singh, "Local diagonal extrema pattern: A new and efficient feature descriptor for CT image retrieval," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1215–1219, Sep. 2015.

[76] J. Suckling *et al.*, "The mammographic image analysis society digital mammogram database," in *Proc. Exerpta Medica. Int. Congr. Series*, 1994, vol. 1069, pp. 375–378.

[77] L. A. Kurgan, K. J. Cios, R. Tadeusiewicz, M. Ogiela, and L. S. Goodenday, "Knowledge discovery approach to automated cardiac spect diagnosis," *Artif. Intell. Med.*, vol. 23, no. 2, pp. 149–169, 2001.

[78] J. Zhang, "Selecting typical instances in instance-based learning," in *Proc. 9th Int. Conf. Mach. Learn.*, 1992, pp. 470–479.

[79] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[80] N. L. Johnson, S. Kotz, and N. Balakrishnan, "Continuous univariate distributions," New York, NY, USA: Wiley, 1994, p. 163.

[81] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD, USA: JHU Press, 2012, vol. 3.

[82] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

Qingxiang Feng received the B.S. degree in computer science and technology from Henan University, Kaifeng, China, the M.S. degree in computer science and technology from the Harbin Institute of Technology, Harbin, China, and is currently working toward the Ph.D. degree at the University of Macau, Macau, China.

He has authored or coauthored more than 20 scientific papers, which includes 10 papers indexed by Science Citation Index. His research interests include pattern recognition and machine learning.

Yicong Zhou (S'08-M'10-SM'14) received the B.S. degree from Hunan University, Changsha, China, and the M.S. and Ph.D. degrees from Tufts University, Medford, MA, USA, all in electrical engineering.

He is currently an Associate Professor and Director of the Vision and Image Processing Laboratory with the Department of Computer and Information Science, University of Macau, Macau, China. His research interests include chaotic systems, multimedia security, image processing and understanding, and machine learning.

Dr. Zhou was the recipient of the third place prize of the Macau Natural Science Award in 2014.