IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 6, 2020, accepted February 18, 2020, date of publication February 21, 2020, date of current version March 3, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2975622

A Kernel-Based Probabilistic Collaborative Representation for Face Recognition

JENG-SHYANG PAN[1,2], XIAOPENG WANG[1,2], QINGXIANG FENG[3], AND SHU-CHUAN CHU[1]

[1]College of Computer Science and Engineering, Shandong University of Science and Technology, Qingdao 266590, China
[2]Fujian Provincial Key Laboratory of Big Data Mining and Applications, Fujian University of Technology, Fuzhou 350118, China
[3]Department of Biomedical Engineering, Tulane University, New Orleans, LA 70118, USA

Corresponding author: Shu-Chuan Chu (scchu0803@gmail.com)

This work was supported by the National Natural Science Foundation of China, under Grant 61872085.

ABSTRACT A novel classifier for face recognition using an improved probabilistic collaborative representation named IPCR is proposed in this paper. The purpose of this paper is to improve the accuracy of face recognition. The testing sample is assumed to be linearly combined by a part of training samples in feature space. There is two-phase framework in IPCR. In the first phase, an adjusted parameter of the nearest neighbors of the samples is chosen for classification. In the second phase, a linear combination of the features and the sparse coefficients are used for new patterns. In the process of two-phase framework, the weight matrix is obtained according to the distance between all the training samples and each testing sample, and then it is applied to weight probabilistic collaborative representation coefficients. The kernel trick is implemented for the high-dimensional nonlinear information instead of linear information of data to improve the class separability. The second classifier named KPCR uses a kernel probabilistic collaborative representation for face recognition. Several renowned face databases, e.g., AR, GT, PIE, FERET, and LFW-crop are used for evaluating the performances of the proposed classifiers. The experimental results demonstrate that the proposed classifiers outperform the collaborative representation-based classification (CRC), the probabilistic collaborative representation-based classifier (ProCRC), and the other state-of-the-art classifiers in recognition accuracy.

INDEX TERMS Computer vision, face recognition, sparse representation, probabilistic collaborative representation.

I. INTRODUCTION

Object recognition is a fundamental problem of computer vision and machine learning. Sparse coding is derived from the theory of image compressive sensing, which has been widely applied in many fields [1]–[4]. Sparse representation-based classification (SRC) [5], [6] was proposed by Wright for the first time, which has shown promising performance on face recognition in the past few years. The sparsity constraint on representation coefficients is a key point in robust classification. However, Zhang et al. found that the sparsity can be further improved when applying all training samples collaboratively to represent a testing sample. Moreover, they used l_2 norm regularization instead of l_1 regularization to solve the coefficients, and proposed a collaborative representation based classification (CRC) algorithm [7], [8]. Due to the simplicity and effectiveness of CRC, it has attracted extensive attention and has been successfully applied in many pattern recognitions. Cai et al. analyzed the classification mechanism of CRC from a probabilistic viewpoint and proposed Probabilistic collaborative representation based classification (ProCRC) [9]. Xu et al. proposed a two-phase testing sample sparse representation (TPTSR) [10], differing from the original SRC but internally borrowed the idea from sparse representation and made coarse to fine classification decisions for the testing sample. Akhtar et al. augmented a dense collaborative representation with an efficiently computed sparse representation and proposed a sparsity augmented collaborative representation based classification (SA-CRC) [30]. Motivated by observation in biological

The associate editor coordinating the review of this manuscript and approving it for publication was Sudipta Roy.

founds, Peng et al. proposed locality constrained collaborative representation (LCCR) [31] which can accurately and robustly identify various occlusion and corruption issues. Additionally, truncated collaborative representation based classification (TCRC) [32], squared and fused versions of SRC and CRC [33], kernel sparse representation to perform coarse-to-fine recognition (KCF) [34] proposed by Zeng et al. also show great competitiveness for classification. At present, these methods are highly represented in the fields of face recognition and sparse representation.

In this paper, an improved probabilistic collaborative representation (IPCR) for face recognition is proposed. We adopt two-phase framework and extend the idea of spatial weight to reconstruct the testing sample. The regularization of ProCRC is weighted by the spatial distance between each training sample and the testing sample, which makes the contribution of each training sample to represent the testing sample more precisely [11]–[13].

The first phase of IPCR linearly combines the overall training samples with sparse coefficients to reconstruct the testing sample and exploits the representation ability of each training sample in order to select the M nearest neighbors. The M nearest neighbors of the testing sample form a determined subspace from all training samples. The labels of M nearest neighbors are used as candidates for the testing sample label, so the classification problem becomes to determine the class of the testing sample from relatively smaller candidates. This is very effective for accurate classification in the second phase.

The second phase of IPCR uses M nearest neighbors to represent the testing sample to obtain the sparse coefficients. Then, linearly combine the training samples of each class with its sparse coefficients to reconstruct the testing sample separately. The errors between the reconstructed results of different classes and the test sample are calculated. Ultimately classify the testing sample into the minimum error class.

The second classifier for face recognition named KPCR is proposed. We adopt the kernel trick based on IPCR for the high-dimensional nonlinear information instead of linear information of data to improve the class separability. Then, the same weighting and two-phase methods as IPCR are used to classify the testing sample. Finally, the testing sample is definitely divided into the class with the minimum error.

Two new classifiers IPCR and KPCR are contributed. Additionally, a large number of face experiments are conducted, and the testing results show that our methods are very competitive in terms of the recognition accuracy in comparison to some existing classifiers in the literature.

The remainder of this paper is organized as follows: Section 2 describes the IPCR method. Section 3 introduces our KPCR. Section 4 analyses our proposed methods. In Section 5, several experiments are carried on the public face databases to evaluate the performance of the proposed classifiers. Finally, the conclusion is discussed in Section 6.

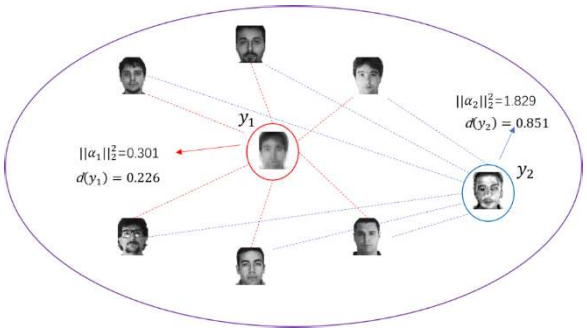


FIGURE 1. Motivation: ProCRC denoted that y_1 has a smaller sum of e2-norm-coefficients, and is more likely to be a face image than y_2. We also find that y_1 has a smaller sum of distances than y_2. It means that sums of e2-norm-coefficients and distances are both useful for classification.

II. IMPROVED PROBABILISTIC COLLABORATIVE REPRESENTATION (IPCR)

In this section, we will present the details of our proposed IPCR. Supposed that there is a date set with n training samples $Y = \{y_i\}_{i=1}^n$ in R^d (d is the dimension of the sample) and C classes. If a training sample y_i is from jth class ($j = 1, 2, \ldots, C$), we take j as the class label of the y_i.

A. MOTIVATION

To explain the motivation of our method, we give an example in Fig. 1. ProCRC [9] also gives a similar example, but ProCRC only denoted that the smaller sum of e2-norm-coefficients is helpful for classification. In this paper, we also find that the smaller sum of distances is useful, which means that sums of e2-norm-coefficients and distances are both helpful for classification. Motivated by this, we propose a new model

$$P\left(l\left(x\right) \in l_Y\right) \propto exp(-\lambda||W\alpha||_2^2), \tag{1}$$

where λ is a constant, $l\left(x\right)$ denotes the label of x, $P\left(l\left(x\right) \in l_Y\right)$ is higher when the sum of e2-norm-coefficients is smaller, and W is a matrix of distances.

$$W = \begin{bmatrix} \|x - y_1\|_2 & & 0 \\ & \ddots & \\ 0 & & \|x - y_n\|_2 \end{bmatrix} \tag{2}$$

For kth class, we get its probability as

$$\begin{aligned} &P\left(l\left(x\right) = k\right) \\ &= P\left(l\left(x\right) \in l_Y\right) \cdot P\left(l\left(x\right) = k | l\left(x\right) \in l_Y\right) \\ &\propto exp\left(-\left(||x - Y\alpha||_2^2 + \lambda||W\alpha||_2^2 + \gamma||Y\alpha - Y_k\alpha_k||_2^2\right)\right), \end{aligned} \tag{3}$$

where γ is a constant. Next, we maximize the joint probability $P\left(l\left(x\right) = k\right)\ k = 1, 2, \cdots, C$ *as*

$$\begin{aligned} max \prod\nolimits_k P\left(l\left(y\right) = k\right) \propto max\ exp\Big(-\Big(&||x - Y\alpha||_2^2 + \lambda||W\alpha||_2^2 \\ &+ \frac{\gamma}{C}\sum\nolimits_{c=1}^{C}||Y\alpha - Y_c\alpha_c||_2^2\Big)\Big) \end{aligned} \tag{4}$$

B. THE FIRST PHASE OF IPCR

The first phase uses the linear combination [14]–[18] of all training samples to represent the testing sample and determines the M nearest neighbors that are the most similar to the testing sample. The testing sample x can be approximately calculated as $Y\alpha$, where the size of Y is $d \times n$, and α is the sparse coefficient which has the size $n \times 1$. We use a spatial distance between each training sample and the testing sample x to weight regularization of ProCRC [9], and the sparse coefficient α for linear combination can be solved by regularization of l_2- norm.

$$\hat{\alpha} = \underset{\alpha}{arg\,min} \begin{pmatrix} \|x - Y\alpha\|_2^2 + \lambda \|W\alpha\|_2^2 \\ +\frac{\gamma}{C} \sum_{c=1}^{C} ||Y\alpha - Y_c\alpha_c||_2^2 \end{pmatrix}, \quad (5)$$

where $\hat{\alpha} = [\alpha_1, \ldots, \alpha_n]^T$, W is a weighted diagonal matrix, λ, γ are constants, C is the total number of classes, $Y_c\alpha_c$ is a matrix composed of the training samples from the *cth* class.

Then, the sparse coefficient $\hat{\alpha}$ can be recovered in a closed-form solution by applying the ordinary least squares technique.

$$\hat{\alpha} = (Y^T Y + \lambda W^T W + \frac{\gamma}{C} \sum_{c=1}^{C} \bar{Y}_c'^T \bar{Y}_c')^{-1} Y^T x, \quad (6)$$

where Y_c' be a matrix which has the same size as Y, while only the samples of the *cth* class will be assigned to Y_c' at their corresponding location in Y, $Y_c' = [0 \ldots, Y_c, \ldots, 0]$, $\bar{Y}_c' = Y - Y_c'$.

$$e_i = ||x - y_i\alpha_i||_2^2 \quad (7)$$

When the testing sample x is represented, the representation ability of the *ith* training sample is shown as $y_i\alpha_i$. We can exploit Eq. (7) to calculate the parameter e_i. It can be regarded as a measurement of the distance between y_i and x. We consider that the smaller e_i, the larger contribution of the *ith* training sample in representing the testing sample x, and vice versa. According to e_i, we determine the M training samples with the greatest contribution, named M nearest neighbors of the testing sample, denoted as $\tilde{Y}$, $\tilde{Y} = \{\tilde{y}_i\}_{i=1}^{M}$. If a nearest neighbor $\tilde{y}_i$ comes from the *jth* ($j = 1, 2, \ldots, C$) class, we will use the j as the label of this nearest neighbor $\tilde{y}_i$. Obviously, $\tilde{Y}$ is a subspace in Y. If $\tilde{Y}$ does not contain the nearest neighbor from the *rth* class, testing sample x will not be ultimately classified into *rth* class.

C. THE SECOND PHASE OF IPCR

The second phase uses the selected M nearest neighbors of the testing sample to get sparse coefficients. Then, the training samples of each class with its sparse coefficients are linearly combined to represented the testing sample separately. Finally, we classify the testing sample x based on the represented result. The testing sample x can be approximately calculated as $\tilde{Y}\beta$, $\tilde{Y} \in R^{d \times m}$, $\beta \in R^{m \times 1}$. Similar to Eq. (5), we get the parameter β as follows.

$$\hat{\beta} = \underset{\beta}{arg\,min} \Big(||x - \tilde{Y}\beta||_2^2 + \lambda ||\tilde{W}\beta||_2^2 + \frac{\gamma}{L} \sum_{l=1}^{L} ||\tilde{Y}\beta - \tilde{Y}_l\beta_l||_2^2 \Big), \quad (8)$$

where $\hat{\beta} = [\beta_1, \ldots, \beta_M]^T$, $\tilde{W}$ is a weighted diagonal matrix and λ, γ are positive constants. L is the number of all classes in $\tilde{Y}$, $\tilde{Y}_l\beta_l$ is a matrix composed of the training samples from the *lth* class.

$$\tilde{W} = \begin{bmatrix} \|x - \tilde{y}_1\|_2 & 0 \\ & \ddots \\ 0 & \|x - \tilde{y}_M\|_2 \end{bmatrix}. \quad (9)$$

where $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_M$ are the columns of $\tilde{Y}$ matrix. Then, the sparse vector $\hat{\beta}$ can be recovered in a closed-form solution.

$$\hat{\beta} = (\tilde{Y}^T \tilde{Y} + \lambda \tilde{W}^T \tilde{W} + \frac{\gamma}{L} \sum_{l=1}^{L} \bar{\tilde{Y}}_l'^T \bar{\tilde{Y}}_l')^{-1} \tilde{Y}^T x, \quad (10)$$

where $\tilde{Y}_l'$ be a matrix which has the same size as $\tilde{Y}$ matrix, while only the sample of $\tilde{Y}_l$ will be assigned to $\tilde{Y}_l'$ at their corresponding location in $\tilde{Y}$, $\tilde{Y}_l' = [0 \ldots, \tilde{Y}_l, \ldots, 0]$, $\bar{\tilde{Y}}_l' = \tilde{Y} - \tilde{Y}_l'$.

The M nearest neighbors of the testing sample come from different sample classes. We calculate the sum of the contribution values of the nearest neighbors in each class separately to classify x. For instance, all nearest neighbors from the *hth* class in $\tilde{Y}$ can be denoted as $\tilde{y}_r, \ldots, \tilde{y}_t$, and the sum of their contribution values, S_h can be shown as follows.

$$S_h = \tilde{y}_r\beta_r + \cdots + \tilde{y}_t\beta_t \quad (11)$$

We use the following Eq. (12) to calculate the residual between S_h and x. A small residual S_h means that the *hth* class has a great contribution in representing x, and then the testing sample is divided into the class with the smallest residual.

$$R_h = \|x - S_h\|_2^2 \quad (12)$$

In summary, the main steps of the proposed IPCR are as follows.

Input: Training data $Y = \{y_i\}_{i=1}^{n}$, the class labels $(1, 2, \ldots, C)$, testing sample x in R^d, the number of nearest neighbors M, and the parameters λ, γ.

Step 1: Calculate the weighted diagonal matrix according to Eq. (2), select M the nearest neighbors for testing sample x by using the first phase Eq. (6), (7).

Step 2: Obtain $\tilde{W}$ of the M nearest neighbors according to Eq. (9), and use the second phase Eq. (10) to get the parameter $\hat{\beta}$.

Step 3: Compute the residual of each class in the M nearest neighbors according to Eq. (12).

Step 4: Classify the testing sample into the class that has the smallest residual.

Output: The x class label.

III. KERNEL BASED PROBABILISTIC COLLABORATIVE REPRESENTATION (KPCR)

Based on the kernel concept, KPCR first maps the original feature to nonlinear high-dimensional feature and uses the high-dimension nonlinear information instead of the linear information to improve the class separability. Then, the first phase of KPCR determines M nearest neighbors of the testing sample according to the deviation between the testing sample and the represented result of each training sample. The second phase of KPCR represents the test sample x by exploiting the linear combination of the M nearest neighbors and then obtains the sparse coefficients. Moreover, calculate the results of the reconstructed testing sample for the different classes, and divide the testing sample into the class with the minimum error. In the process, the diagonal matrix based on the distance between each training sample and the testing sample in high-dimensional space is used to weight the regularization of the KPCR [19]– [21].

A. KERNEL TRICK

Appropriate selection of a kernel function can reflect similarities between samples, the kernel trick uses a linear method to get its nonlinear counterpart without calculating the mapping explicitly. In this paper, the Gaussian radial basis function (RBF) kernel is used, which can be written as

$$k\left(x, x'\right) = \phi\left(x\right)^T \phi\left(x'\right) = exp\left(-\delta||x - x'||_2^2\right), \quad (13)$$

where $\delta > 0$ is the parameter of RBF kernel, x and x' are two original data points. In the kernel method, we use function $k(*, *)$ to convert to the feature space, but $\phi(*)$ is unclear.

B. THE FIRST PHASE OF THE KPCR

The main purpose of the first phase is to determine the M nearest neighbors that are most similar to the testing sample. Assume that there is a nonlinear feature mapping function $\phi\left(.\right) : R^d \rightarrow R^q \ (d \ll q)$, which maps the testing sample x and training dataset Y into a high dimensional feature space as

$$\begin{aligned} &x \rightarrow \phi\left(x\right) \\ &Y \rightarrow \phi\left(Y\right) \quad \text{and } \phi\left(Y\right) = \left[\phi\left(y_1\right)\phi\left(y_2\right), \ldots, \phi\left(y_n\right)\right] \end{aligned}$$

In this part, based on the distance between $\phi\left(x\right)$ and $\phi\left(y_i\right)$ $(i = 1, 2 \ldots, n)$ data points in the high-dimensional feature, the regularization term is weighted to obtain the sparse parameter α as follows.

$$\begin{aligned} \hat{\alpha} = \underset{\alpha}{arg\,min} &\left(||\phi\left(x\right) - \phi\left(Y\right)\alpha||_2^2 + \lambda||W\alpha||_2^2\right. \\ &\left. + \frac{\gamma}{C}\sum\nolimits_{c=1}^{C}||\phi\left(Y\right)\alpha - \phi\left(Y\right)_c\alpha_c||_2^2\right) \end{aligned} \quad (14)$$

$$W = \begin{bmatrix} \|\phi(x) - \phi\left(y_1\right)\|_2, & & 0 \\ & \ddots & \\ 0 & & \|\phi(x) - \phi\left(y_n\right)\|_2 \end{bmatrix}, \quad (15)$$

where, W is the weighted parameter, $\|\phi\left(x\right) - \phi\left(y_i\right)\|_2 = \left[k\left(x, x\right) + k\left(y_i, y_i\right) - 2k\left(y_i, x\right)\right]^{1/2}$. $\phi\left(Y\right)_c\alpha_c$ is a matrix composed of the training samples from the *cth* class.

Then, the $\hat{\alpha}$ can be recovered in a closed-form solution.

$$\hat{\alpha} = \left(K + \lambda W^T W + \frac{\gamma}{C}\sum\nolimits_{c=1}^{C}\bar{K}_c'\right)^{-1}k, \quad (16)$$

where, $K = \phi\left(Y\right)^T\phi\left(Y\right) \in R^{n\times n}$ is the Gram matrix with $K_{i,j} = k\left(y_i, y_j\right)$. $k = \left(k\left(y_1, x\right), k\left(y_2, x\right), \ldots, k(y_n, x)\right)^T \in R^{n\times 1}$. Let K_c' be a matrix which has the same size as K. K_c is the Gram matrix of the samples in *cth* class, and only K_c will be assigned to K_c' at their corresponding location in K, $K_c' = [0 \ldots, K_c, \ldots, 0]$, $\bar{K}_c' = K - K_c'$. According to Eq. (16), sparse parameter $\hat{\alpha}$ is obtained. We can calculate the deviation e_i from the *ith* training sample representation.

$$\begin{aligned} e_i &= ||\phi\left(x\right) - \phi\left(y_i\right)\alpha_i||_2^2 \\ &= k\left(x, x\right) + \alpha_i^2 k\left(y_i, y_i\right) - 2\alpha_i k\left(y_i, x\right) \end{aligned} \quad (17)$$

The testing sample $\phi(x)$ is represented as $\phi\left(y_i\right)\alpha_i$ in the *ith* training sample, the contribution can be estimated by deviation e_i. The smaller the e_i, the larger the contribution of the *ith* training sample, when the testing sample $\phi(x)$ is represented, and vice versa. According to e_i, the M training samples with the greatest contribution are selected, called M nearest neighbors, expressed as $\phi'\left(Y\right), \phi'\left(Y\right) = \left\{\phi\left(y_i'\right)\right\}_{i=1}^{M}$. If the nearest neighbor $\phi\left(y_i'\right)$ comes from the *jth* $(j = 1, 2, \ldots, C)$ class, we will determine the j as the label of this nearest neighbor $\phi\left(y_i'\right)$. If $\phi'\left(Y\right)$ does not contain the nearest neighbor from the *rth* class, testing sample $\phi(x)$ will not be classified into the *rth* class.

C. THE SECOND PHASE OF THE KPCR

In the second phase of the KPCR, the testing sample $\phi(x)$ is classified by the linearly combining the M nearest neighbors with the sparse coefficients. $\phi(x)$ can be approximated as $\phi'\left(Y\right)\beta$ where $\phi'\left(Y\right) \in R^{q\times M}$, $\beta \in R^{M\times 1}$. Similar to Eq. (14), β is written as follows.

$$\begin{aligned} \hat{\beta} = \underset{\beta}{arg\,min} &\left(||\phi\left(x\right) - \phi'\left(Y\right)\beta||_2^2 + \lambda||\tilde{W}\beta||_2^2\right. \\ &\left. + \frac{\gamma}{L}\sum\nolimits_{l=1}^{L}||\phi'\left(Y\right)\beta - \phi'\left(Y\right)_l\beta_l||_2^2\right) \end{aligned} \quad (18)$$

$$\tilde{W} = \begin{bmatrix} \left\|\phi(x) - \phi\left(y_1'\right)\right\|_2 & & 0 \\ & \ddots & \\ 0 & & \left\|\phi(x) - \phi\left(y_M'\right)\right\|_2 \end{bmatrix}, \quad (19)$$

where $\hat{\beta} = \left[\beta_1, \ldots, \beta_M\right]^T$, $\tilde{W}$ is a weighted diagonal matrix, λ, γ are positive constants, L is the number of training samples class in $\phi'\left(Y\right), \phi'\left(Y\right)_l\beta_l$ is a data point consisting of training samples from the lth class of the $\phi'\left(Y\right)$.

Then, the sparse vector $\hat{\beta}$ can be recovered in a closed-form solution.

$$\hat{\beta} = (\tilde{K} + \lambda \tilde{W}^T \tilde{W} + \frac{\gamma}{L} \sum_{l=1}^{L} \bar{\tilde{K}}_l')^{-1} k', \tag{20}$$

where, $\tilde{K} = \phi'(Y)^T \phi'(Y) \in R^{M \times M}$ is the Gram matrix with $\tilde{K}_{i,j} = k\left(y_i', y_j'\right)(ij = 1, 2, \ldots, M)$. $k' = \phi'(Y)^T \phi(x) = \left(k\left(y_1', x\right), k\left(y_2', x\right), \ldots, k(y_M', x)\right)^T \in R^{M \times 1}$. Let $\tilde{K}_l'$ be a matrix which has the same size as $\tilde{K}$. $\tilde{K}_l$ is the Gram matrix of the samples in *lth* class, and only $\tilde{K}_l$ will be assigned to $\tilde{K}_l'$ at their corresponding location in $\tilde{K}$, $\tilde{K}_l' = [0 \ldots, \tilde{K}_l, \ldots, 0]$, $\bar{\tilde{K}}_l' = \tilde{K} - \tilde{K}_l'$.

M nearest neighbors come from different sample classes. The number of training samples in each class may be different. Sum the contribution values of the nearest neighbors in each class separately to classify the testing sample $\phi(x)$. For instance, the nearest neighbors from the *hth* class in $\phi'(Y)$ can be represented as $\phi\left(y_r'\right), \ldots, \phi\left(y_t'\right)$, and the sum of contribution values is as follows

$$S_h = \phi\left(y_r'\right)\beta_r + \cdots + \phi\left(y_t'\right)\beta_t \tag{21}$$

The residual between S_h and $\phi(x)$ is calculated according to Eq. (22). A small residual S_h means that the *hth* class has a great contribution in representing the $\phi(x)$, and ultimately classify x into the class with the smallest residual.

$$\begin{aligned} R_h &= \|\phi(x) - S_h\|_2^2 \\ &= (\phi(x) - S_h)^T (\phi(x) - S_h) \\ &= k(x, x) + \beta_h'^T \tilde{K}_h \beta_h' - 2\beta_h'^T k_h', \end{aligned} \tag{22}$$

where $\beta_h' = (\beta_r, \ldots, \beta_t)^T$, $\tilde{K}_h$ is the Gram matrix of the samples in *hth* class, $k_h' = \left[k\left(y_r', x\right), \ldots, k\left(y_t', x\right)\right]^T$.

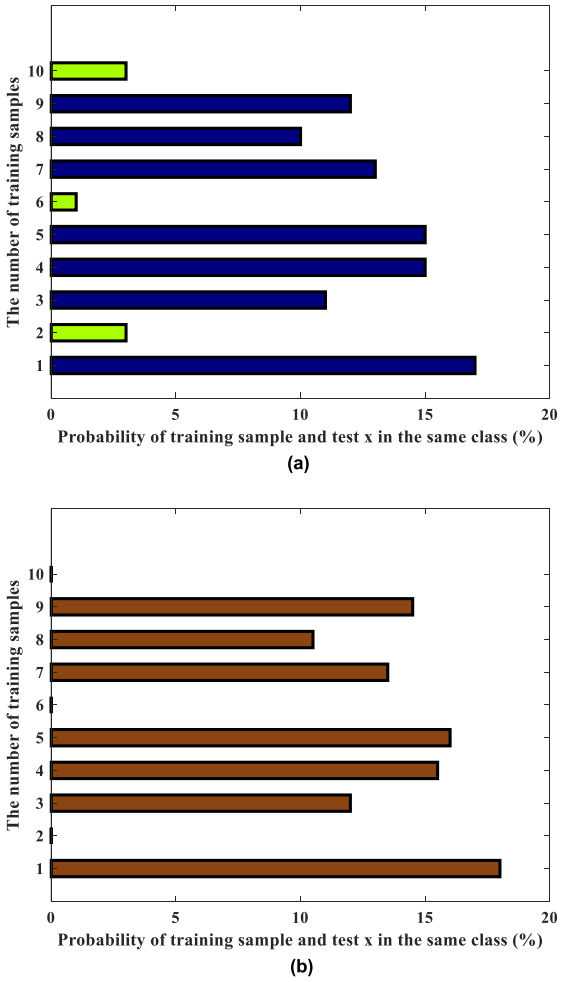


FIGURE 2. (a) Possible probability distribution of all training samples, and (b) Possible probability distribution of M nearest neighbors.

IV. ALYSIS OF THE PEOPOSED METHOD

In this section, we will analyze the ideas and principles of the IPCR and compare it with some methods.

A. METHOD COMPARISON

CRC and ProCRC were proposed in [8] and [9]. These methods perform well on face recognition. They use all the training samples to linearly represent the testing sample for classification.

Compared with ProCRC, when IPCR finds sparse parameter α, we add the weight matrix and convert $\lambda||\alpha||_2^2$ to $\lambda||W\alpha||_2^2$. In this way, we modify the parameter α according to the spatial distance of each training sample relative to the testing sample, so that it is more reasonable relative to $\lambda||\alpha||_2^2$, and the experimental result verification is more effective.

IPCR adopts the two-phase framework. The first phase selects the M nearest neighbors according to the deviation e_i, and the second phase uses the linear combination of M nearest neighbors to classify the testing sample x. It is a sparse form with supervision. Meaning that the sparse parameters of M nearest neighbors are usually not zero, and the remaining training sample sparse coefficients are set to zero according to the deviation e_i. Therefore, we know the sparse coefficients of the training samples when representing the testing sample. However, CRC, ProCRC and other methods are called an unsupervised sparse form. When the training samples linearly represent the testing sample x, we do not know which sparse coefficient is zero or close to zero [22], [23].

B. PROBABILITY DESCRIPTION

Our IPCR method can be described as follows: $c_h(h \in (1, 2, \ldots, C))$ represents the *hth* class in the samples. The first phase of IPCR uses e_i to evaluate the probability that the testing sample x and the *ith* training sample belong to the same class, denoted as $p(y_i|x)(i \in (1, 2, \ldots, n))$. The first phase of IPCR assumes that $p(y_i|x) \propto 1/e_i$, $e_i = ||x - y_i\alpha_i||_2^2$. The smaller e_i is, the greater the probability $p(y_i|x)$ that the x belongs to the same class as the *ith* training sample. As shown in Fig. 2(a), if the training sample has a small $p(y_i|x)$, the first phase will set the $p(y_i|x)(i = 2, 6, 10)$ of green bars to zero when M nearest neighbors are selected. However, the value $\sum_i p(y_i|x) = 1$ is constant, so the $p(y_i|x)$ (blue bars) of the other training samples will increase, and Fig. 2(a) will be converted to Fig. 2(b). The first phase removes the training

sample with small value $p(y_i|x)$ and retains the M training samples with high similarity to the testing sample x.

The second phase of IPCR uses $p(c_h|x)$ to represent the posterior probability that test x belongs to the hth class. The second phase assumes that $p(c_h|x) \propto 1 \Big/ \|x - S_h\|_2^2$, S_h is the reconstructed result of the training samples from the hth class. A minimum $\|x - S_h\|_2^2$, which means a maximum posterior probability $p(c_h|x)$. The second phase of IPCR will classify test x into the hth class with the maximum $p(c_h|x)$. If there is no training sample from the sth class in the M nearest neighbors, then the test x will not be classified into the sth class in the second phase, $p(c_h|x) = 0$.

In contrast to the IPCR, its global version uses all training samples to represent the test x [10]. Almost every $p(c_h|x)$ is a non-zero value, and the global version needs to find a maximum posterior probability $p(c_h|x)$ from all C classes for the test x. IPCR sets some posterior probabilities to zeros. The result may produce a more ideal probability distribution to classify the test x relative to the global version. We assume that test x comes from the second class in Fig. 3. As can be seen from Fig. 3(a), the global version will incorrectly classify x into the fifth class due to noise interference, but IPCR can accurately classify x to the second class with the highest posterior probability in Fig. 3(b).

V. EXPERIMENT RESULTS

In this section, we conducted a lot of testing experiments on AR, GT, PIE, FERET, and LFW-crop. The experimental results demonstrate the effectiveness of our methods in recognition accuracy.

A. FACE DATABASE INTRODUCTION

AR Face Database: AR Face Database was created by Aleix Martinez and Robert Benavente at the Ohio State University. It contains more than 4,000 color images, corresponding to 126 faces (70 males and 56 females). The images have different facial expressions, lighting conditions and occlusion (sunglasses and scarves). For fair comparison to show the results, we used 2600 images from 100 people, every individual has 26 images, and these images are resized to 30×40 for our experiments [24].

1) GT FACE DATABASE

Georgia Tech Face Database contains images of 50 people taken in two or three sessions at the Center for Signal and Image Processing at Georgia Institute of Technology. There are 15 color images, the images were taken into account the variations in illumination conditions, facial expression, and appearance. In addition to this, the faces were captured at different scales and orientations and we resized the size of images to 30×40 for experimental testing [25].

2) PIE FACE DATABASE

PIE Face Database was provided by the face research team at CMU, the database contains 41,368 images of 68 people, each person under 13 different poses, 43 different illumination conditions, and with 4 different expressions. For PIE, we only used a subset from 68 individuals with each person providing 39 images, and each image from the PIE database was resized to 30×40 in our experiments [26].

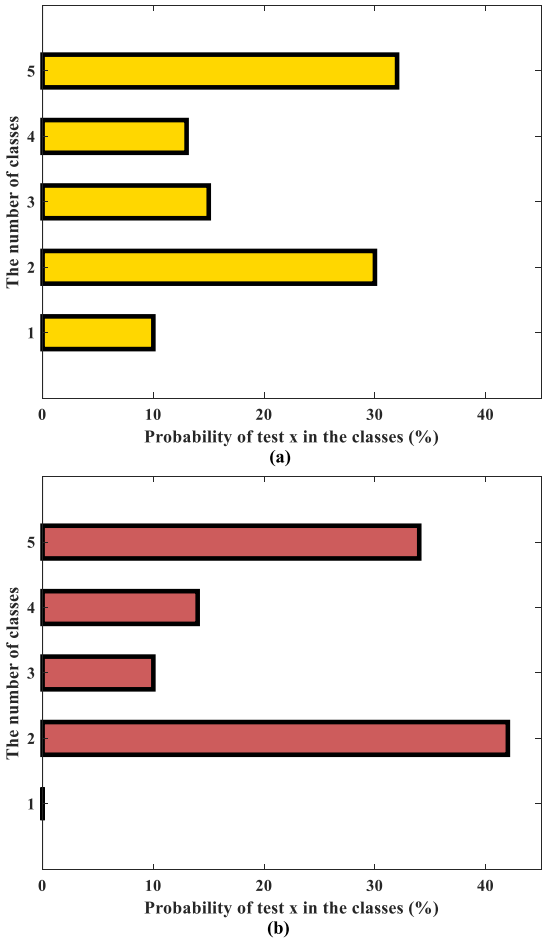


FIGURE 3. (a) Possible posterior probability distribution of the global IPCR, and (b) Possible posterior probability distribution of the IPCR.

3) FERET DATABASE

FERET is created by the Face Recognition Technology project, this collection of images contains a large number of face images, and each image has only one face. In this database, images of the same person have different expressions, changes in light, posture, and age. FERET with more than 10,000 multi-poses and illumination is one of the most widely used face databases. In the experiment, we selected a subset with 1400 images containing 200 people, 7 images per person, and each image was resized to 40×40 [27].

4) LFW-CROP DATABASE

LFW-crop is a cropped version of the Labeled Faces in the Wild (LFW) database, keeping only the center portion of each image (i.e. the face). In the majority of images, almost all of the background is omitted. 158 people, 10 pictures per person were selected for the experiment. Each image was resized to 30×30 [28].

FIGURE 4. Some face images from the AR Face Database. The images shown in first, second and third rows are from three different classes.

FIGURE 5. Some face images from the GT Face Dataset. The images shown in first, second and third rows are from three different classes.

B. SAMPLE SELECTIONS AND PARAMETER SETTINGS

1) SAMPLE SELECTIONS

For each face database AR, GT, PIE, FERET and LFE-crop, we randomly select q samples from the p samples in each class as the training sample set, and the remaining part as the test set, including C_P^q combination possibly. As a result, there are C_P^q training sets and test sets. For each database, we have done a random selection of the number of training samples in three cases. For the AR face database, in the first case, we randomly selected 6 out of 26 samples of each class as the training set, and the rest were test sets, which were tested with different classification methods. In the second, third cases, 10 and 16 samples in each class were randomly selected as the training set. Fig. 4 shows some face images from the AR face database. For the GT face database, we randomly selected 5, 8 and 11 images from the 15 images of each class as the training images and the rest as test images. Some images of the GT face database are shown in Fig. 5. We also performed the same sample selection on the PIE, FERET and LFW-crop face databases. Finally, different classification methods were applied to test the recognition rate on PIE, FERET, and LFW-crop.

2) PARAMETER SETTINGS

The proposed IPCR contains three parameters, λ, γ and M. On the AR face database, $\lambda = 1e^{-4}, \gamma = 1\ M = 0.5$; On the GT face database, $\lambda = 1, \gamma = 1, M = 0.5$; On the PIE face database, $\lambda = 1, \gamma = 1, M = 0.5$; On the FERET face database, $\lambda = 1e^{-1}, \gamma = 1, M = 0.5$; On the LFW-crop face database, $\lambda = 1e^{-4}, \gamma = 1, M = 0.5$.

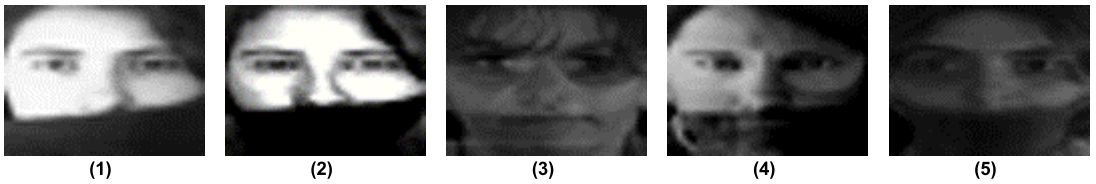
The proposed KPCR contains four parameters, δ, λ, γ and M. On the AR face database, $\delta = 1e^{-1}, \lambda = 1e^{-4}, \gamma = 1e^{-3}, M = 0.5$; On the GT face database, $\delta = 1, \lambda = 1e^{-1}, \gamma = 1e^{-3}, M = 0.5$; On the PIE face database, $\delta = 1e^{-1}, \lambda = 1e^{-3}, \gamma = 1e^{-3}, M = 0.5$; On the FERET and LFW-crop face databases, $\delta = 1, \lambda = 1e^{-1}, \gamma = 1e^{-3}, M = 0.5$. For the other competitive classification methods, we also achieved the best classification accuracy level in each experiment by tuning the parameters.

C. EXPERIMENTAL DATA DISPLAY AND ANALYSIS

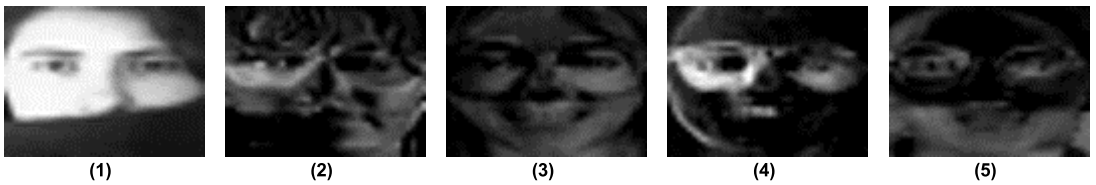
As shown earlier, S_h in Eq. (11) is the reconstructed result of the training samples of the *hth* class to the testing sample x. If the deviation between S_h and x is the smallest, x will eventually be assigned to the *hth* class. We reconstructed the test x using the proposed IPCR, then converted the obtained $S_h(h \in (1, 2, \ldots C))$ into a matrix of the same size as the test face image. The reconstructed result when the IPCR method was used as shown in Fig. 6(a) and M was selected as half of the training samples (1000 samples). Image (1) is the original data from the AR face database, image (2)-(5) are the reconstructed results of the four classes with the smallest residual, and the testing sample is correctly divided into the class to which it belongs.

At the same time, we also used the global version (do not choose M nearest neighbors) and the version without weight (called TPPCR) methods to reconstruct the same test image. Therefore, as shown in Fig. 6(b) and 6(c), image (1) is still the original data from the AR database, and image (2)-(5) are the reconstructed results of the four classes with the smallest residuals, respectively. Obviously, the final result is misclassified.

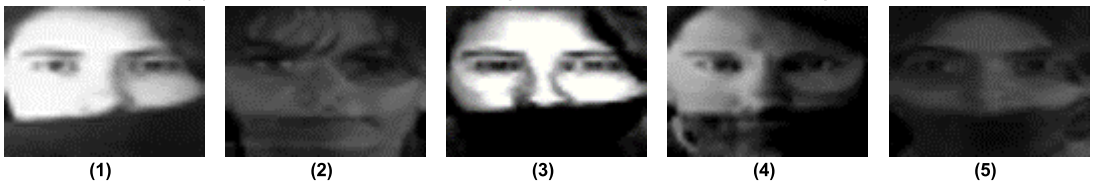
We conducted a lot of the testing experiments on the GT and FERET databases. On GT, 4 and 6 images randomly selected from 15 of each person were used as the training samples. On FERET, 2 and 4 from 7 images of each person were selected as the training sample set, and the remaining part was used as the test set for experimental comparison. Fig. 7 and Fig. 8 show the mean of the error rates of 10 experimental results by using different classification methods. As can be seen from Fig. 7 and Fig. 8, compared to TPPCR, the proposed IPCR has a much lower recognition error rate, which shows that the weighting technique is very effective. Moreover, our KPCR performed further improvement in recognition accuracy compared to the IPCR.



(a) The reconstructed results of our IPCR on the testing sample. Image (1) is the original testing sample, images (2)-(5) are reconstructed from the four classes with the first four smallest residuals, respectively.



(b) The reconstructed results of the global version of IPCR on the testing sample.



(c) The reconstructed results of the TPPCR on the testing sample.

FIGURE 6. The reconstructed results of the testing sample from AR face database.

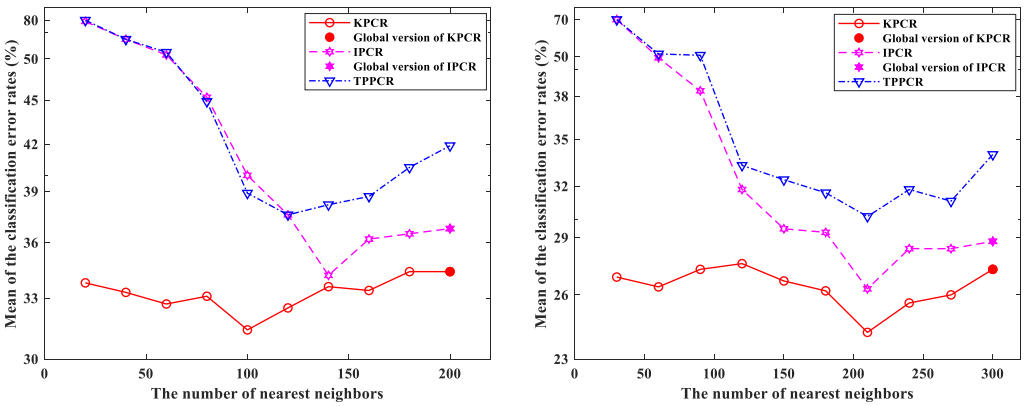


FIGURE 7. Mean of the classification error rates on the GT face database.

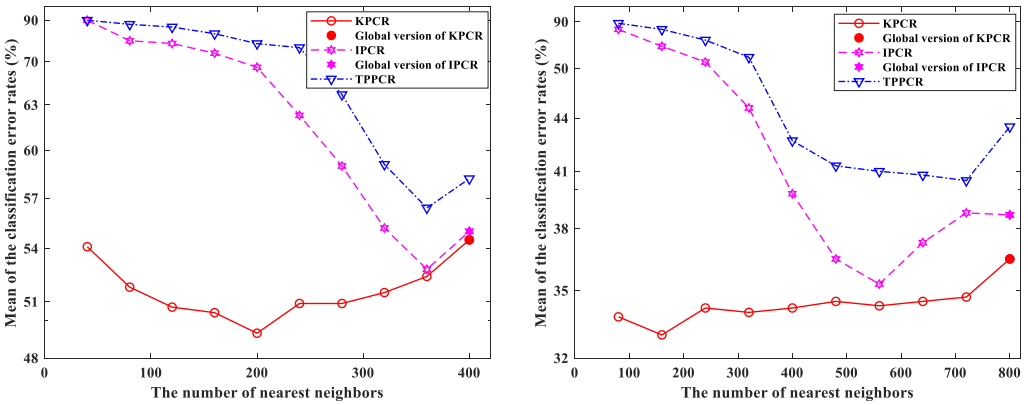


FIGURE 8. Mean of the classification error rates on the FERET face database.

TABLE 1. AR-100-26_30-40 database.

Methods / Samples	Recognition rate													
	KPCR		IPCR		CRC		ProCRC		KCRC		TPTSR		l_2RDSRA	
	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std
6	**0.923**	0.003	0.914	0.002	0.830	0.007	0.881	0.010	0.902	0.006	0.896	0.012	0.875	0.005
10	**0.978**	0.003	0.976	0.002	0.787	0.011	0.813	0.011	0.972	0.007	0.933	0.011	0.950	0.006
16	**0.985**	0.004	0.984	0.003	0.947	0.012	0.949	0.011	0.978	0.009	0.940	0.011	0.979	0.006

TABLE 2. GT-50-15_30-40 database.

Methods / Samples	Recognition rate													
	KPCR		IPCR		CRC		ProCRC		KCRC		TPTSR		l_2RDSRA	
	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std
5	**0.702**	0.014	0.692	0.018	0.501	0.015	0.501	0.016	0.686	0.016	0.601	0.015	0.674	0.021
8	**0.809**	0.022	0.786	0.029	0.517	0.022	0.517	0.021	0.760	0.023	0.664	0.019	0.745	0.026
11	0.855	0.009	**0.870**	0.007	0.525	0.017	0.520	0.018	0.840	0.021	0.720	0.013	0.805	0.022

TABLE 3. PIE-68-39_40-40 database.

Methods / Samples	Recognition rate													
	KPCR		IPCR		CRC		ProCRC		KCRC		TPTSR		l_2RDSRA	
	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std
10	0.507	0.013	0.506	0.009	0.445	0.016	0.345	0.017	**0.508**	0.017	0.493	0.010	0.478	0.013
15	0.662	0.009	**0.698**	0.013	0.554	0.018	0.556	0.021	0.678	0.010	0.632	0.014	0.639	0.016
25	0.881	0.013	**0.894**	0.016	0.664	0.011	0.652	0.015	0.873	0.018	0.797	0.016	0.875	0.015

TABLE 4. FERET-200-7_40-40 database.

Methods / Samples	Recognition rate													
	KPCR		IPCR		CRC		ProCRC		KCRC		TPTSR		l_2RDSRA	
	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std
3	**0.613**	0.023	0.510	0.021	0.343	0.019	0.369	0.030	0.573	0.021	0.418	0.020	0.600	0.017
5	**0.740**	0.016	0.702	0.018	0.320	0.025	0.314	0.026	0.678	0.030	0.456	0.019	0725	0.021
6	**0.810**	0.020	0.772	0.023	0.288	0.025	0.278	0.034	0.760	0.027	0.448	0.034	0.770	0.031

Especially when the number of selected the M nearest neighbors is different, the KPCR shows good robustness. By comparing the data results obtained with different M nearest neighbors, we can find that the appropriate selection of the M nearest neighbors can reduce the error rate of recognition. Our KPCR and IPCR can improve the recognition rate by about 5% compared to their global version. Therefore, it can be said that the adopted two-phase framework is also effective.

Table 1 to Table 5 show the classification results of the proposed KPCR, IPCR and the other state-of-the-art methods including CRC, KCRC [29], ProCRC, TPTSR, and l_2 Regularization Based Discriminative Sparse Representation Algorithm (l_2 RDSRA). Table 1 is the recognition results of the AR database, and Table 2 to Table 5 are the results on GT, PIE, FERET, and LFW-crop, respectively. Taking the AR database as an example, we randomly chose 6, 10 and 16 from the 25 samples of each class as the training set, and the remaining part was used as the testing set. Repeated the above test 10 times to get the experimental mean and standard deviation. In the 6-sample case, the recognition rate generated by KCPR is 92.3%, the recognition rate generated by IPCR is 91.4%, the recognition rate generated by CRC is 83.0%, the recognition rate generated by ProCRC is 88.1%, the recognition rate generated by KCRC is 90.2%, the recognition rate generated by TPTSR is 89.6%, the recognition rate generated by l_2 RDSRA is 87.5%. In the 10-sample case,

TABLE 5. LFW-crop-158-10_30-30 database.

Methods / Samples	Recognition rate													
	KPCR		IPCR		CRC		ProCRC		KCRC		TPTSR		l_2RDSRA	
	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std	Ave	Std
4	**0.471**	0.008	0.433	0.014	0.165	0.008	0.196	0.010	0.450	0.007	0.300	0.013	0.465	0.012
5	**0.506**	0.011	0.475	0.012	0.089	0.005	0.134	0.009	0.472	0.013	0.295	0.014	0.477	0.012
7	**0.623**	0.007	0.515	0.011	0.196	0.007	0.014	0.012	0.550	0.011	0.310	0.010	0.568	0.011

TABLE 6. KPCR vs. Other contrasted methods under Wilcoxon signed-rank test.

Database	Compared methods					
	IPCR	CRC	ProCRC	KCRC	TPTSR	l_2RDSRA
AR						
6	+	+	+	+	+	+
10	=	+	+	=	+	+
16	=	+	+	+	+	+
GT						
5	+	+	+	+	+	+
8	+	+	+	+	+	+
11	-	+	+	+	+	+
PIE						
10	=	+	+	=	+	+
15	-	+	+	-	+	+
25	=	+	+	+	+	=
FERET						
3	+	+	+	+	+	+
5	+	+	+	+	+	+
6	+	+	+	+	+	+
LFW						
4	+	+	+	+	+	+
5	+	+	+	+	+	+
7	+	+	+	+	+	+

the recognition rate generated by KCPR is 97.8%, the recognition rate generated by IPCR is 97.6%, the recognition rate generated by CRC is 78.7%, the recognition rate generated by ProCRC is 81.3%, the recognition rate generated by KCRC is 97.2%, the recognition rate generated by TPTSR is 93.3%, the recognition rate generated by l_2 RDSRA is 95.0%. In the 16-sample case, the recognition rate generated by KCPR is 98.5%, the recognition rate generated by IPCR is 98.4%, the recognition rate generated by CRC is 94.7%, the recognition rate generated by ProCRC is 94.9%, the recognition rate generated by KCRC is 97.8%, the recognition rate generated by TPTSR is 94.0%, the recognition rate generated by l_2 RDSRA is 97.9%. The standard deviation produced by each algorithm is also very small. For detailed experimental data of the remaining databases, refer to Table 2 to Table 5. All the highest accuracy recognition results obtained have been bolded.

From the data results in Table 1 to Table 5, it can be seen that the performance of KPCR is the most outstanding compared to the other methods in terms of recognition accuracy. For the case of selecting the 16 training samples on the AR database, the mean recognition rate reached 98.5%. IPCR also performed very well, especially on AR, GT, and PIE, the recognition accuracy is very close to KPCR. This proves that the proposed KPCR and IPCR can perform well in face recognition.

In order to comprehensively compare KPCR with the other methods, Wilcoxon signed-rank test with a level of significance $\alpha = 0.05$ is applied to evaluate the results of 10 times running [40]. Combined to obtain the mean value, the final test results are shown in Table 6. When KCPR is compared with other methods, the symbol "+" means "win", symbol "-" means "lose", and symbol "=" means "draw". In Table 6, compared to IPCR, KPCR produced 9 better

results, 4 similar results, and 2 worse results. Compared with KCRC, KPCR produced 12 better results, 2 similar results, and 1 worse result. Compared with l_2 RDSRA, KPCR produced 14 better results, 1 similar result. Compared to CRC, ProCRC, and TPTSR, KPCR all produced 15 better results. From the overall experimental results, our proposed KPCR is superior to the contrast methods in face recognition. The efficiency of the proposed schemes can be further improved by adopting some relevant methods [35]–[39].

VI. CONCLUSION

In this paper, two novel classifiers for face recognition, namely IPCR and KPCR are presented. The weighted based on space distance and two-phase framework are applied to proposed IPCR. We also carried out a clear probability interpretation of the two-phase framework and analyzed the feasibility of the method. Moreover, we adopted the kernel concept, utilizing the high-dimensional nonlinear information instead of linear information of data to the proposed KPCR. The popular AR, GT, PIE, FERET, and LFW face databases are used, a large number of experiments showed the high recognition accuracy of our classifiers. For instance, in the case of selecting the 4 training samples on the LFW-crop database, IPCR improved the recognition rate of 23.7%, and KPCR improved 27.5% respectively compared to ProCRC. Comparing with more existing classifiers, such as CRC, KCRC, TPTSR, and l_2 Regularization Based Discriminative Sparse Representation Algorithm (l_2 RDSRA), our methods also performed better in face recognition. The proposed KPCR and IPCR may be implemented for the various pattern classification problems.

REFERENCES

[1] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[2] B.-D. Liu, B. Shen, and Y.-X. Wang, "Class specific dictionary learning for face recognition," in *Proc. IEEE Int. Conf. Secur., Pattern Anal., Cybern. (SPAC)*, Hubei, China, Oct. 2014, pp. 229–234.

[3] B.-D. Liu, B. Shen, L. Gui, Y.-X. Wang, X. Li, F. Yan, and Y.-J. Wang, "Face recognition using class specific dictionary learning for sparse representation and collaborative representation," *Neurocomputing*, vol. 204, pp. 198–210, Sep. 2016.

[4] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.

[5] H. Zhang, Y. Zhang, and T. S. Huang, "Pose-robust face recognition via sparse representation," *Pattern Recognit.*, vol. 46, no. 5, pp. 1511–1521, May 2013.

[6] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T.-S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

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

[8] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang, "Collaborative representation based classification for face recognition," 2012, *arXiv:1204.2358*. [Online]. Available: http://arxiv.org/abs/1204.2358

[9] 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 Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2950–2959.

[10] 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.

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

[12] R. Timofte and L.-V. Gool, "Weighted collaborative representation and classification of images," in *Proc. Int. Conf. Pattern Recognit.*, Tsukuba, Japan, Apr. 2012, pp. 1606–1610.

[13] X. Dong, H. Zhang, L. Zhu, W. Wan, Z. Wang, Q. Wang, P. Guo, H. Ji, and J. Sun, "Weighted locality collaborative representation based on sparse subspace," *J. Vis. Commun. Image Represent.*, vol. 58, pp. 187–194, Jan. 2019.

[14] J. Wen, X. Fang, J. Cui, L. Fei, K. Yan, Y. Chen, and Y. Xu, "Robust sparse linear discriminant analysis," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 2, pp. 390–403, Feb. 2019.

[15] 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.

[16] Y. Lu, X. Fang, and B. Xie, "Kernel linear regression for face recognition," *Neural Comput. Appl.*, vol. 24, nos. 7–8, pp. 1843–1849, Jun. 2013.

[17] 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, Feb. 2016.

[18] J. Wen, Y. Xu, Z. Li, Z. Ma, and Y. Xu, "Inter-class sparsity based discriminative least square regression," *Neural Netw.*, vol. 102, pp. 36–47, Jun. 2018.

[19] Q. Zhu, Y. Xu, J. Wang, and Z. Fan, "Kernel based sparse representation for face recognition," in *Proc. Int. Conf. Pattern Recognit.*, Tsukuba, Japan, Nov. 2012, pp. 1703–1706.

[20] Q. Feng and Y. Zhou, "Kernel combined sparse representation for disease recognition," *IEEE Trans. Multimedia*, vol. 18, no. 10, pp. 1956–1968, Oct. 2016.

[21] 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.

[22] H. Wang, C. Yuan, W. Hu, and C. Sun, "Supervised class-specific dictionary learning for sparse modeling in action recognition," *Pattern Recognit.*, vol. 45, no. 11, pp. 3902–3911, Nov. 2012.

[23] J. Gou, L. Wang, Z. Yi, J. Lv, Q. Mao, and Y.-H. Yuan, "A new discriminative collaborative neighbor representation method for robust face recognition," *IEEE Access*, vol. 6, pp. 74713–74727, 2018.

[24] A. M. Martinez, "The ar face database," CVC, New Delhi, India, Tech. Rep. 24, 1998.

[25] A. V. Nefian, "Georgia Tech face database," Georgia Inst. Technol., Atlanta, GA, USA, Tech. Rep., 1999.

[26] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recognit.*, Washington DC, USA, May 2002, pp. 53–58.

[27] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, 2000.

[28] L. J. Karam and T. Zhu, "Quality labeled faces in the wild (QLFW): A database for studying face recognition in real-world environments," *Proc. SPIE Hum. Vis. Electron. Imag.*, vol. 9394, Mar. 2015, Art. no. 93940B.

[29] 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.*, Vancouver, BC, Canada, May 2013, pp. 2877–2881.

[30] N. Akhtar, F. Shafait, and A. Mian, "Efficient classification with sparsity augmented collaborative representation," *Pattern Recognit.*, vol. 65, pp. 136–145, May 2017.

[31] X. Peng, L. Zhang, Z. Yi, and K. K. Tan, "Learning locality-constrained collaborative representation for robust face recognition," *Pattern Recognit.*, vol. 47, no. 9, pp. 2794–2806, Sep. 2014.

[32] S. Zeng, B. Zhang, Y. Lan, and J. Gou, "Robust collaborative representation-based classification via regularization of truncated total least squares," *Neural Comput. Appl.*, vol. 31, no. 10, pp. 5689–5697, Feb. 2018.

[33] S. Zeng, J. Gou, and X. Yang, "Improving sparsity of coefficients for robust sparse and collaborative representation-based image classification," *Neural Comput. Appl.*, vol. 30, no. 10, pp. 2965–2978, Feb. 2017.

[34] S. Zeng, X. Yang, and J. Gou, "Using kernel sparse representation to perform coarse-to-fine recognition of face images," *Optik*, vol. 140, pp. 528–535, Jul. 2017.
[35] J. Wang, Y. Gao, W. Liu, W. Wu, and S.-J. Lim, "An asynchronous clustering and mobile data gathering schema based on timer mechanism in wireless sensor networks," *Comput., Mater. Continua*, vol. 58, no. 3, pp. 711–725, 2019.
[36] J.-S. Pan, L. Kong, T.-W. Sung, P.-W. Tsai, and V. Snášel, "α-Fraction first strategy for hierarchical model in wireless sensor networks," *J. Internet Technol.*, vol. 19, no. 6, pp. 1717–1726, 2018.
[37] J.-S. Pan, C.-Y. Lee, A. Sghaier, M. Zeghid, and J. Xie, "Novel systolization of subquadratic space complexity multipliers based on toeplitz Matrix–Vector product approach," *IEEE Trans. Very Large Scale Integr. (VLSI) Syst.*, vol. 27, no. 7, pp. 1614–1622, Jul. 2019.
[38] T.-T. Nguyen, J.-S. Pan, and T.-K. Dao, "A compact bat algorithm for unequal clustering in wireless sensor networks," *Appl. Sci.*, vol. 9, no. 10, p. 1973, May 2019.
[39] T.-T. Nguyen, J.-S. Pan, and T.-K. Dao, "An improved flower pollination algorithm for optimizing layouts of nodes in wireless sensor network," *IEEE Access*, vol. 7, pp. 75985–75998, 2019.
[40] R.-F. Woolson, "Wilcoxon signed-rank test," in *Wiley Encyclopedia of Clinical Trials*. Hoboken, NJ, USA: Wiley, 2007, pp. 1–23355.

JENG-SHYANG PAN received the B.S. degree in electronic engineering from the National Taiwan University of Science and Technology, in 1986, the M.S. degree in communication engineering from National Chiao Tung University, Taiwan, in 1988, and the Ph.D. degree in electrical engineering from the University of Edinburgh, U.K, in 1996. He is currently the Dean of the College of Information Science and Engineering, Fujian University of Technology. He joined the Editorial Board of *LNCS Transactions on Data Hiding and Multimedia Security*, the *Journal of Computers*, and the *Chinese Journal of Electronics*. His current research interests include soft computing, information security, and signal processing.

XIAOPENG WANG received the B.S. degree in electrical engineering from the Fujian University of Technology, China, in 2016, where he is currently pursuing the M.S. degree with the College of Information Science and Engineering. His current research interests include pattern recognition and swarm intelligence.

QINGXIANG FENG received the Ph.D. degree from the Department of Computer Science, University of Macau, Macau, China. He is currently a Postdoctoral Fellow with the Department of Biomedical Engineering, Tulane University, New Orleans, LA, USA. As the first author, he has published more than 20 scientific articles, including CVPR, AAAI, TMM, TCSVT, and TCYB. His research interests include artificial intelligence for computer vision and medical analysis.

SHU-CHUAN CHU received the Ph.D. degree from the School of Computer Science, Engineering and Mathematics, Flinders University of South Australia, in 2004. She joined Flinders University, in December 2009, after nine years at Cheng Shiu University, Taiwan. She has been a Research Fellow with the College of Science and Engineering, Flinders University, Australia, since December 2009. She has also been a Research Fellow with the Ph.D. Advisor with the College of Computer Science and Engineering, Shandong University of Science and Technology, since September 2019. Her research interests are mainly in swarm intelligence, intelligent computing, and data mining.