Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 1527965, 17 pages
https://doi.org/10.1155/2020/1527965

Research Article

Jointly Learning the Discriminative Dictionary and Projection for Face Recognition

Chao Bi,[1,2] **Yugen Yi**,[3] **Lei Zhang**,[4] **Caixia Zheng**,[5] **Yanjiao Shi**,[6] **Xiaochun Xie**,[1,2] **Jianzhong Wang**,[5] **and Yan Wu**[1,2]

[1]*School of Psychology, Northeast Normal University, Changchun 130024, China*
[2]*Jilin Provincial Experimental Teaching Demonstration Center of Psychology, Northeast Normal University, Changchun 130024, China*
[3]*School of Software, Jiangxi Normal University, Nanchang 330022, China*
[4]*Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130024, China*
[5]*School of Information Sciences and Technology, Northeast Normal University, Changchun 130117, China*
[6]*School of Computer Science and Information Engineering, Shanghai Institute of Technology, Shanghai 200235, China*

Correspondence should be addressed to Jianzhong Wang; wangjz019@nenu.edu.cn and Yan Wu; 1901984590@qq.com

Received 21 February 2020; Revised 22 June 2020; Accepted 28 July 2020; Published 24 August 2020

Academic Editor: Giuseppe D'Aniello

Recently, dictionary learning has become an active topic. However, the majority of dictionary learning methods directly employs original or predefined handcrafted features to describe the data, which ignores the intrinsic relationship between the dictionary and features. In this study, we present a method called jointly learning the discriminative dictionary and projection (JLDDP) that can simultaneously learn the discriminative dictionary and projection for both image-based and video-based face recognition. The dictionary can realize a tight correspondence between atoms and class labels. Simultaneously, the projection matrix can extract discriminative information from the original samples. Through adopting the Fisher discrimination criterion, the proposed framework enables a better fit between the learned dictionary and projection. With the representation error and coding coefficients, the classification scheme further improves the discriminative ability of our method. An iterative optimization algorithm is proposed, and the convergence is proved mathematically. Extensive experimental results on seven image-based and video-based face databases demonstrate the validity of JLDDP.

1. Introduction

Face recognition (FR) is an imperative issue in the field of image processing and computer vision. Recently, plenty of face recognition methods have been proposed [1–5]. However, the problems of occlusion, illumination, pose, and small sample size are still huge challenges for face recognition [6–8]. Currently, sparse representation-based classification (SRC) [9] has been successfully employed, in which the overcomplete dictionary can represent the query face image well. Significantly, the dictionary designed for SRC utilizes all training images. SRC has shown favorable properties in FR, particularly when images are partly occluded. Nevertheless, the unsure and noisy components may lead to the ineffectiveness of the dictionary in representing query samples. Moreover, the dictionary's size is consistent with the number of training images. Thus, the computational cost of solving sparse representation coefficients will increase if the training samples' number is large. At last, the dictionary does not take the structure of the training set or class label into account, which will make the dictionary lack discriminant information. To address these issues, predefined dictionaries that use bases such as Haar or Gabor wavelet instead of training samples are presented [10, 11], but none of these bases is proposed for SRC [12].

Dictionary learning (DL) is significant for SRC because it can suppress the useless information to promote the representation and discrimination [13]. To learn a

discriminative and small-sized dictionary, a substantial amount of methods have been presented [14–16], which can be roughly divided into two categories: unsupervised and supervised. Unsupervised DL methods have achieved satisfactory results by minimizing the representation error. The method of optimal directions (MOD) [17] was proposed for unsupervised DL. MOD updated the dictionary by minimizing the representation error and achieved the convergence by an iteration-based strategy. However, the computation of the inverse matrix in the MOD was very complicated. The K-singular value decomposition (K-SVD) [18] method was proposed based on the MOD, which performed SVD decomposition on the representation error term and selected the decomposition terms as the updated dictionary atoms and the corresponding coding coefficients. The most substantive difference between MOD and K-SVD is the dictionary updating strategy, in which K-SVD updates one atom and its corresponding coding coefficients each time until all atoms are updated. Therefore, the MOD can be considered as a simplified version of K-SVD. Although the performance of the K-SVD method has been improved, the computational complexity of updating atoms is also high. To enhance the efficiency of DL, an effective reconstructed DL method was presented in [19], which was based on alternating optimization over two subsets of variables. Skretting and Engan [20] introduced a forgetting factor λ into the DL algorithm to make the algorithm less dependent on the initial dictionary. In [21], metafaces were learned from the training samples, which can promote the representation ability of the dictionary. Although unsupervised DL methods have achieved impressive recognition results, there still exists a limitation in their practical applications. Due to the absence of label information, the dictionaries obtained by unsupervised DL methods were always lacking the discriminative ability. To overcome this problem, many supervised DL methods that utilize the label information have been proposed. In [22], a discriminative K-SVD algorithm was proposed to ensure the representative and discriminative abilities of the learned dictionary. To better utilize the correspondence between the dictionary and labels, the label consistent K-SVD [23] algorithm, which associated the label information with each atom to promote the discriminative ability of the dictionary, was put forward. Recently, the Fisher discrimination dictionary learning (FDDL) [24] algorithm was proposed to learn a class-specific dictionary for FR. Based on the Fisher discrimination criterion [25], the representation error associated with each class was employed for classification. Ding and Ji [26] applied a kernel-based robust disturbance dictionary to significantly enhance the recognition accuracy of occluded faces. Since the supervised DL methods explored the label information of training samples to promote the discriminative ability of the learned dictionary, they have achieved well performance for FR. Recent progresses in SRC have made video-based face recognition become a growing research topic. The video can be treated as a set of images obtained from different poses, illuminations, and expressions. The main difficulty is how to effectively use the multiframe information. In [27], a video dictionary was adopted to encode different video information, i.e., pose, temporal, and illumination. In [28], a multivariate sparse representation method was suggested for video-based face recognition, which was robust to noise and occlusion. These two methods learned the dictionary for FR, but they did not consider the impact of other constraints on algorithm performance. Xu et al. [29] proposed a method to learn a structured dictionary for video-based face recognition, which adopted the nuclear norm to make the coding coefficient matrix be low-rank. However, this method did not enhance the discriminative ability of the representation coefficients. In addition, it utilized the samples in the original space to learn the dictionary and the coding coefficient matrix, which ignores the influence of noise and other irrelevant information.

Dimensionality reduction (DR) is an essential step to decrease the cost of data computation and storage. It also eliminates the irrelevant information to enhance the discriminative ability of features [30–33]. Zhang et al. [34] proposed a novel unsupervised algorithm to obtain the orthogonal projection, which can ensure that the samples were well reconstructed in the projected subspace. Clemmensen et al. [35] utilized the sparseness criterion to realize linear discriminant analysis so that the classification and feature selection can be achieved concurrently. In [36], a linear discriminative projection was learned by maximizing the ratio of the between-class representation error to the within-class representation error in the projected space. In [37], the sparsity criterion and the maximum margin criterion [38] were combined to obtain the discriminant projection. Although these SRC-based DR methods yielded notable results, they only acquired the low-dimensional features of the samples and failed to supply an explicit discriminative dictionary.

To overcome this limitation, a series of methods have been suggested to combine DR and DL into a unified framework. By combining the sparseness criterion with PCA, Nguyen et al. [39] presented a sparse embedding method for simultaneously solving the DR and DL problems. The projection matrix was learned for retaining the sparse structure of samples, and the dictionary was learned in the reduced space simultaneously. However, it ignored the distinguish ability of different class samples in the subspace. In [40], the sigmoid function and the ratio of intraclass representation error to interclass representation error were utilized to learn the discriminative dictionary and projection simultaneously, but it ignored both the intraclass and interclass scatter matrix of the coefficients and low-dimensional samples. To address this problem, Feng et al. [41] introduced an orthogonal projection matrix, which can be obtained through maximizing the total scatter and between-class scatter of the training set, in the projection and dictionary simultaneously learning framework. Liu et al. [42] utilized the discriminative graph constraints to achieve nonnegative feature projection and dictionary learning simultaneously. Lu et al. [43] also presented a framework, which can simultaneously learn low-dimensional features and dictionaries, to deal with the video-based face recognition problem. Although these jointly learning methods have achieved success, they did not exploit the discriminative

relationship between low-dimensional features and dictionary. To address this issue, a novel method called jointly learning the discriminative dictionary and projection (JLDDP), which simultaneously learns the dictionary and projection in a unified framework, is proposed for FR in this paper. Compared with the existing methods, JLDDP has four characteristics. First, the discriminative ability of the dictionary can be enhanced via imposing the Fisher discrimination criterion on the coding coefficients. Second, the projection learned by our approach enables the closeness of samples from the same class, while keeping the samples from different classes far away in the low-dimensional subspace. Third, JLDDP combines the processes of projection learning and DL into a uniform framework, so the dictionary and projection can be automatically optimized. Last, we design an iterative optimization algorithm to solve our model and provide a theoretical proof for its convergence.

The remaining part is organized as follows. Some of the related work is briefly reviewed in Section 2. The details of JLDDP are provided in Section 3. Experiments and comparisons are carried out in Section 4, and conclusions are provided in Section 5.

2. Related Work

2.1. SRC. SRC was proposed by Wright et al. [9] for face recognition. Assume there are n classes of samples, and the training set can be expressed as $X = [X_1, \ldots, X_i, \ldots, X_n] \in R^{m\times n}$, where $X_i = [x_{i,1}, \ldots, x_{i,j}, \ldots, x_{i,n_i}] \in R^{m\times n_i}$ denotes the subset of the training samples that contains n_i samples of class i. Let $x_{i,j}$ $(j = 1, 2, \ldots, n_i)$ represent the m-dimensional vector stretched by the j-th sample of class i. SRC assumes that a testing sample can be well estimated by the linear combination of the training samples from the same class, so let $y \in R^m$ denote a testing sample of class i; it can be expressed as $y = a_{i,1}x_{i,1} + a_{i,2}x_{i,2} + \cdots + a_{i,n_i}x_{i,n_i}$, where $a_{i,j}$ is the corresponding coding coefficient. Suppose we utilize the training set to represent y, the corresponding coefficient vector entries except those related to the i-th class should be zero. In SRC, the l_1-minimization is applied to handle the coefficient vector, i.e., $\widehat{a} = \arg\min_a \|y - Xa\|_2^2 + \lambda\|a\|_1$, where λ is a tradeoff parameter. $e_i = \|y - X\delta_i(\widehat{a})\|_2$ denotes the representation error of class i, where $\delta_i(\cdot)$: $R^n \longrightarrow R^{n_i}$ can choose the coefficients of class i. The classification criterion is $\text{identity}(y) = \arg\min_i\{e_i\}$.

2.2. Dictionary Learning. In this section, the DL methods, including unsupervised K-SVD [18] and supervised FDDL [24], will be reviewed.

2.2.1. K-SVD. In the K-SVD algorithm [18], an overcomplete dictionary is learned from the training set for image compression and denoising. The objective function of K-SVD is formulated as

$$\begin{aligned} &\min_{D,\alpha} \|X - D\alpha\|_2 \\ &\text{s.t. } \|\alpha\|_0 \le T, \end{aligned} \tag{1}$$

where X is the training set, D is the dictionary, α is the sparse coding coefficient matrix of X over D, and T is the parameter to adjust the sparsity. To optimize equation (1), the sparse coding coefficient α and the dictionary D are updated iteratively. However, there is no corresponding relation between the class label and the dictionary atoms. Thus, K-SVD is unsuitable for solving classification problems.

2.2.2. FDDL. Different from K-SVD, FDDL [24] combines the class label information and the Fisher discrimination criterion to learn a structured discriminative dictionary, which performs classification by the representation error for each class. The FDDL model is formulated as

$$\begin{aligned} &\min_{D,\alpha} \left\{ \sum_{i=1}^{c} \left(\left\|X_i - D\alpha_i\right\|_F^2 + \left\|X_i - D_i\alpha_i^i\right\|_F^2 + \sum_{j=1, j\neq i}^{c} \left\|D_j\alpha_i^j\right\|_F^2 \right) + \lambda_1\|\alpha\|_1 + \lambda_2\left(tr\left(S_W(\alpha) - S_B(\alpha)\right) + \eta\|\alpha\|_F^2\right) \right. \\ &\text{s.t. } \left\|d_k\right\|_2^2 = 1, \quad \forall k, \end{aligned} \tag{2}$$

where X is the training set, λ_1 and λ_2 are tradeoff parameters, and each column of D is normalized to a unit vector. $\sum_{i=1}^{c}(\|X_i - D\alpha_i\|_F^2 + \|X_i - D_i\alpha_i^i\|_F^2) + \sum_{j=1, j\neq i}^{c}\|D_j\alpha_i^j\|_F^2$ is the discriminative term, $\|\alpha\|_1$ is the sparse regularization term, and $tr(S_W(\alpha) - S_B(\alpha)) + \eta\|\alpha\|_F^2$ is the discriminative coefficient term to enforce the discriminative ability of the sparse representation coefficients. The objective function of FDDL can be optimized by updating the dictionary and sparse representation coefficients iteratively. Although FDDL has achieved a good performance for FR, the process is time-consuming. Therefore, PCA is applied to extract features from all samples firstly in FDDL.

3. Methodology

In this section, we firstly describe the proposed JLDDP, which incorporates DL and projection learning into a unified framework. Secondly, the novel iterative update algorithm of JLDDP is deduced. Thirdly, the convergence analysis is given. Fourthly, we provide the classification schemes which characterize the class-specific representation error for FR. Finally, we analyze the guideline for parameter setting.

3.1. Modeling. Let $Y = [Y_1, Y_2, \ldots, Y_c]$ denote the set of d-dimensional training samples with c classes, where Y_i is the i-th class subset of Y. Let P be the projection that reduces the feature dimension of samples. The structured (class-specific) dictionary is denoted by $D = [D_1, D_2, \ldots, D_c]$, where D_i is the i-th class subdictionary. The coding coefficient matrix of P^TY over D is denoted by X, which can be refined to $X = [X_1, X_2, \ldots, X_c]$, where X_i is the i-th class submatrix of coding coefficient X. Actually, X_i can also be expressed as $X_i = [X_i^1, \ldots, X_i^j, \ldots, X_i^c]$, where X_i^j is the coding coefficient of P^TY_i over the subdictionary D_j. In JLDDP, the projection, dictionary, and coding coefficients are jointly learned with the following model:

$$\begin{aligned} &\min_{P,D,X} R(P,D,X) + \omega_1 \|X\|_1 + \omega_2 C(X) + \omega_3 S(P), \\ &\text{s.t. } \|d_k\|_2^2 = 1, \quad \forall k, \end{aligned} \tag{3}$$

where $R(P,D,X)$ denotes the representation error term, $\|X\|_1$ is the l_1-regularization on X, $C(X)$ is the coding coefficient term imposing discriminative label information on DL, and $S(P)$ is the projection learning term projecting the samples into a more discriminative space. ω_1, ω_2, and ω_3 are the tradeoff parameters. Each atom d_k in the dictionary has a unit norm. Next, more detailed descriptions of the terms in equation (3) will be given.

3.1.1. Representation Error Term. When the training samples are represented by a dictionary, we expect the dictionary to have both strong reconstructive ability and strong discriminative ability. In addition, the samples can be reconstructed not only by the whole dictionary but also by the subdictionary from the same class. Therefore, the representation error term is expressed as

$$\begin{aligned} R(P,D,X) = \sum_{i=1}^{c} \Bigg(&\left\|P^TY_i - DX_i\right\|_F^2 + \left\|P^TY_i - D_iX_i^i\right\|_F^2 \\ &+ \sum_{j=1, j\neq i}^{c} \left\|D_jX_i^j\right\|_F^2 \Bigg). \end{aligned} \tag{4}$$

The representation error term is designed to obtain a small representation error that is calculated by the low-dimensional training samples P^TY_i and the structured dictionary D. First, each class of low-dimensional training samples P^TY_i should be well represented by the structured dictionary D, i.e., $P^TY_i \approx DX_i = D_1X_i^1 + \cdots + D_iX_i^i + \cdots + D_cX_i^c$. Second, each class of low-dimensional training samples should be well represented by the dictionary from the same class, rather than other classes, which indicates that P^TY_i should be well represented by D_i as much as possible, but not by $D_j\,(j \neq i)$. Hence, X_i^i should have some significant coefficients, and $X_i^j\,(j \neq i)$ should have nearly zero coefficients.

3.1.2. Coding Coefficient Term. We can make the dictionary discriminative by constraining the coding coefficients [24]. According to the Fisher discrimination criterion, the within-class scatter should be minimized, and the between-class scatter should be maximized, which can make the coding coefficients have discriminative ability. Hence, the coding coefficient term is formulated as

$$C(X) = tr\left(S_w(X) - S_b(X)\right) + \|X\|_F^2, \tag{5}$$

where $S_w(X) = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - m_i)(x_k - m_i)^T$ is the within-class scatter of X, m_i is the mean vector of X_i, $S_b(X) = \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^T$ is the between-class scatter of X, n_i is the number of samples in class i, and m is the mean vectors of X. We impose the Fisher discrimination criterion on X to improve the discriminative ability, which indicates the within-class scatter $S_w(X)$ should be minimized, and the between-class scatter $S_b(X)$ should be maximized. $\|X\|_F^2$ is an elastic term, and the convexity of equation (5) is proved in [24].

3.1.3. Projection Learning Term. The projection matrix P should preserve the energy of samples as much as possible and make the samples from different classes separable in the low-dimensional space. Therefore, the projection learning term is expressed as

$$S(P) = tr\left(S_w\left(P^TY\right)\right) - tr\left(S_b\left(P^TY\right)\right) - \left\|P^TY\right\|_F^2, \tag{6}$$

where $S_w(P^TY) = \sum_{i=1}^{c} \sum_{y_k' \in P^TY_i} (y_k' - m_i')(y_k' - m_i')^T$ and $S_b(P^TY) = \sum_{i=1}^{c} n_i (m_i' - m')(m_i' - m')^T$ are the within-class scatter and the between-class scatter of P^TY, respectively. y_k' denotes the k-th sample from class i in the low-dimensional space. m_i' and m' denote the mean vectors of P^TY_i and P^TY, respectively. We adopt the Fisher discrimination criterion on low-dimensional samples, i.e., $tr(S_w(P^TY)) - tr(S_b(P^TY))$, to enhance the discriminative ability of features. Moreover, we minimize the term $-\|P^TY\|_F^2$ to guarantee that the energy of Y can be well preserved.

By incorporating equations (4)–(6), we obtain the JLDDP model as shown in equation (3). The iterative update scheme is adopted to optimize the objective function, and the detailed optimization process of JLDDP is presented in the following section.

3.2. Optimization. The objective function of JLDDP is not convex for P, D, and X jointly, but it is convex with regard to each of them when the others are fixed. Thus, equation (3) can be divided into three subproblems and optimized by an iterative update scheme.

3.2.1. Updating X with Fixed P and D. Suppose that P and D are fixed, we can update $X = [X_1, X_2, \ldots, X_c]$ class-by-class, i.e., we fix all $X_j\,(j \neq i)$ to update X_i. Therefore, the simplified form of equation (3) can be obtained as follows:

$$\min_{X_i}\left\{\sum_{i=1}^{c}\left(\left\|P^TY_i-DX_i\right\|_F^2+\left\|P^TY_i-D_iX_i^i\right\|_F^2+\sum_{j=1,j\neq i}^{c}\left\|D_jX_i^j\right\|_F^2\right)+\omega_1\|X_i\|+\omega_2\left(\left\|X_i-M_i\right\|_F^2-\sum_{k=1}^{c}\left\|M_k-M\right\|_F^2+\left\|X_i\right\|_F^2\right),\right. \tag{7}$$

where M_k and M_i are the mean vector matrices of class k and class i, respectively. M is the mean vector matrix of all classes. Except for $\|X_i\|_1$, the other terms in equation (7) are differentiable. Since equation (7) is strictly convex, we can employ iterative projection methods (IPM) [44] to solve it.

3.2.2. Updating D with Fixed P and X. To obtain the optimal structured dictionary D, we need to update the subdictionary D_i class-by-class, while P, X, and all other D_j $(j\neq i)$ are fixed. Then, equation (3) can be simplified as

$$\begin{aligned}&\min_{D_i}\left\|P^TY-D_iX^i-\sum_{j=1,j\neq i}^{c}D_jX^j\right\|_F^2+\left\|P^TY_i-D_iX_i^i\right\|_F^2+\sum_{j=1,j\neq i}^{c}\left\|D_iX_j^i\right\|_F^2\\&\text{s.t. } d_k^Td_k=1,\quad\forall k,\end{aligned} \tag{8}$$

where X^i represents the coding coefficients of P^TY over the subdictionary D_i. We can employ the algorithm in [19] to solve equation (8), i.e., update D_i atom-by-atom.

3.2.3. Updating P with Fixed D and X. When the dictionary D and the coding coefficient matrix X are fixed, equation (3) can be simplified to

$$\begin{aligned}\min_{P}\sum_{i=1}^{c}&\left(\left(\left\|P^TY_i-DX_i\right\|_F^2+\left\|P^TY_i-D_iX_i^i\right\|_F^2\right)\right.\\&\left.+\omega_3\left(tr\left(S_w\left(P^TY\right)\right)-S_b\left(P^TY\right)\right)-\left\|P^TY\right\|_F^2\right).\end{aligned} \tag{9}$$

We can obtain equation (10) by the mathematical derivation of equation (9):

$$\begin{aligned}\min_{P}\left\{\sum_{i=1}^{c}\right.&\left(P^TY_iY_i^TP-2P^TY_iX_i^TD^T+DX_iX_i^TD+P^TY_iY_i^TP-2P^TY_iX_i^{iT}D_i^T+D_iX_i^iX_i^{iT}D_i^T\right)\\&\left.+\omega_3\left(tr\left(P^T\left(S_w(Y)-S_b(Y)\right)P\right)-tr\left(P^TYY^TP\right)\right)\right.\end{aligned} \tag{10}$$

If we set the derivative of P as zero in equation (10), we acquire

$$\begin{aligned}&\sum_{i=1}^{c}\left(2Y_iY_i^TP-Y_iX_i^TD^T-Y_iX_i^{iT}D_i^T\right)\\&\quad+\omega_3\left(\left(S_w(Y)-S_b(Y)\right)P-YY^TP\right)=0.\end{aligned} \tag{11}$$

For convenience, we define $t_1=\sum_{i=1}^{c}Y_iY_i^T$, $t_2=\sum_{i=1}^{c}Y_iX_i^TD^T$, $t_3=\sum_{i=1}^{c}Y_iX_i^{iT}D_i^T$, $t_4=S_w(Y)-S_b(Y)$, and $t_5=YY^T$ to replace the corresponding parts of equation (11). Then, we gain the explicit solution of the projection matrix P as shown in the following:

$$P=\left(2t_1-\omega_3t_5+\omega_3t_4\right)^{-1}\left(t_2+t_3\right). \tag{12}$$

The above iterative optimization process of JLDDP will stop when the algorithm is convergent or the maximum number of iterations is attained. Algorithm 1 is the summary of the whole optimization process.

3.3. Convergence. The optimization process of JLDDP can be simplified into three subproblems that can be solved iteratively, as formulated in equations (7), (8), and (12). It has been proved that the subproblem in equation (7) is convex in [24]. Obviously, equation (8) is quadratic programming, so it is convex. In each iteration, the value will decline after solving X and D via equations (7) and (8), respectively, as proved in [21, 44]. Moreover, the subproblem in equation (12) can obtain an explicit solution. Thus, to justify the convergence of JLDDP, we need to demonstrate that the value of equation (3) is nonincreasing after optimization. For convenience, let $\phi(P,D,X)$ denote the objective function of JLDDP. Before proving the convergence of Algorithm 1, we should establish Theorem 1 first.

Theorem 1. *If Algorithm 1 is used to solve $\phi(P,D,X)$, the objective function value is nonincremental.*

Proof. Let $\phi(P^t,D^t,X^t)$ indicate the value in the t-th iteration.

When solving the subproblem $\min_X\phi(P^t,D^t,X)$, we utilize the method in [44] to obtain the optimal value of X^{t+1} with fixed P^t and D^t. This subproblem is convex, so we can obtain

(1) **Input**: the training set $Y = [Y_1, Y_2, \ldots, Y_c]$, iteration number T, parameters ω_1, ω_2, and ω_3.
(2) **Initialize**: projection matrix $P = P^0$, structured dictionary $D = D^0$, $t = 1$.
(3) **Repeat** steps 3–6 until convergence or $t < T$ conditions.
(4) Update X^t with fixed P^{t-1} and D^{t-1} by equation (7).
(5) Update D^t with fixed P^{t-1} and X^t by equation (8).
(6) Update P^t with fixed D^t and X^t by equation (12).
(7) **Output**: projection matrix P, structured dictionary D, coding coefficient matrix X.

Algorithm 1: The algorithm of JLDDP.

$$\phi\left(P^t, D^t, X^{t+1}\right) \le \phi\left(P^t, D^t, X^t\right). \tag{13}$$

When solving the subproblem $\min_D \phi(P^t, D, X^t)$, we employ the method in [21] to obtain the optimal value of D^{t+1} with fixed P^t and X^t. It is still a convex problem, so we have

$$\phi\left(P^t, D^{t+1}, X^t\right) \le \phi\left(P^t, D^t, X^t\right). \tag{14}$$

When solving the subproblem $\min_P \phi(P, D^t, X^t)$, we can obtain the explicit solution with fixed D^t and X^t based on equation (12). Therefore,

$$\phi\left(P^{t+1}, D^t, X^t\right) \le \phi\left(P^t, D^t, X^t\right). \tag{15}$$

Combining equations (13)–(15), we have

$$\phi\left(P^{t+1}, D^{t+1}, X^{t+1}\right) \le \phi\left(P^t, D^t, X^t\right). \tag{16}$$

Now, the theorem has been proved.

Since each term in equation (3) is nonnegative, the objective function value has a low bound. According to Theorem 1 and the Cauchy convergence criterion [45], the optimization algorithm presented for JLDDP is convergent.

3.4. Classification. The learned projection P can reduce the dimension of the testing sample y_t, and the low-dimensional feature $P^T y_t$ can be coded over the learned dictionary D. Therefore, we can obtain the coding coefficient $x\prime$ by

$$x\prime = \arg\min_x \left\|P^T y_t - Dx\right\|_2^2 + \alpha\|x\|_1, \tag{17}$$

where $x\prime = [x_1', \ldots, x_i', \ldots, x_c']$ is the coding coefficient and x_i' is the coding coefficient vector associated with class i. α is a tradeoff parameter.

The structured dictionary D is learned to ensure the coding coefficients of the identical class are similar, and the coding coefficients of various classes are different. In addition, the coding coefficients have a stronger discriminative ability through the constraints of the Fisher discrimination criterion. Therefore, not only the representation error but also the distance information of the coding coefficients obtained by equation (17) is useful for classification. We classify the testing sample y_t by

$$\text{label}\left(y_t\right) = \arg\min_i \left\|P^T y_t - D_i x_i'\right\|_2^2 + \gamma\left\|x\prime - \overline{x_i'}\right\|_2^2, \tag{18}$$

where $\overline{x_i'}$ is the mean vector of $x\prime$ related to class i and γ is a tradeoff parameter.

3.5. Parameter Analysis. There are three parameters in the proposed JLDDP, i.e., ω_1, ω_2, and ω_3. Therefore, how to properly set their values is important. Fortunately, each parameter has a clear physical meaning, which can supply a guideline for setting the value. The parameter ω_1 is used to control the sparsity of the coding coefficient matrix, whose value needs to be set as a moderate value. The parameter ω_2 can adjust the coding coefficient term based on the Fisher discrimination criterion, whose value should not be set either too small or too large. Since an extremely small ω_2 value will lead to the loss of latent discrimination information, a too large ω_2 value will make other terms be neglected. The parameter ω_3 is used to constrain the projection learning term based on the Fisher discrimination criterion. Analogous to the parameter ω_2, a relatively small ω_3 value can decrease the projection learning term effect. However, a relatively large ω_3 value will make the objective function dominated by the projection learning term, and the role of other terms will be neglected.

3.6. Comparison with the Existing Work. In order to highlight the novelty of our work, we compare the proposed JLDDP method with some related studies. First, although some terms in the objective function of FDDL [24] are similar to those in our JLDDP, they are different from each other. Specifically, FDDL utilizes PCA to project original features into a low-dimensional subspace, which is separated from the process of dictionary learning. Thus, FDDL does not exploit the relationship between the low-dimensional features and the learned dictionary, which cannot effectively learn the appropriate features for the discriminative dictionary learning task. To solve this problem, our proposed JLDDP simultaneously learns the feature projection matrix and dictionary in a unified framework, which can ensure that the learned projection matrix is most beneficial for discriminative dictionary learning. That is, the learned projection matrix and dictionary in our JLDDP are relevant and mutually beneficial. Hence, jointly optimizing them can achieve better performance for face recognition. Second, the proposed JLDDP also seems like the dictionary learning methods in [46–48]. However, there exist some significant differences between them. To be specific, (1) the methods in [46–48], respectively, learn multiple class-specific

subdictionaries and a common subdictionary shared by all classes. Then, they combine the learned class-specific subdictionaries and common subdictionary to achieve the recognition task. In our JLDDP, we only need to learn a subdictionary for each class and combine all subdictionaries as a whole dictionary. Therefore, there is no need to learn and update the common dictionary during the model optimization, which can make sure that our model has a fast convergence speed and high computational efficiency. (2) Similar to FDDL, the methods in [46–48] do not consider feature projection matrix learning in the process of dictionary learning. Thus, the feature projection is separated from the process of dictionary learning in them, which cannot learn the best combination of the low-dimensional feature and dictionary for face recognition. (3) The regularization criteria in the objective functions adopted in [46–48] were different from our proposed JLDDP, e.g., [46, 48] used l_1-norm, and [47] used $l_{2,1}$-norm to enforce the learned coefficients of the dictionary to be sparse, while our proposed JLDDP utilizes the intraclass and interclass scatter of coefficients as constraints, which can improve the discrimination of the model. Third, Lin et al. [49] proposed a RDCDL method which utilizes the low rank and sparse constraint to extract the disturbance components (e.g., noise, outliers, and occlusion) in the training samples. In RDCDL, a set of training samples and a set of alternative training samples with simulated facial variation are employed to build a dictionary learning model with a complex and comprehensive dictionary. The comprehensive dictionary includes a class-shared dictionary, a class-specific dictionary, a simulated disturbance dictionary, and a real disturbance dictionary. The main difference between our JLDDP and RDCDL lies in that we only adopt class-specific dictionary to construct the whole dictionary, which is simpler than Lin's model and can deeply decrease the computational complexity. Besides, RDCDL utilizes PCA to reduce the feature dimension of samples, which is separated from the process of dictionary learning. However, our JLDDP combines the processes of feature projection and dictionary learning into a unified framework to obtain a more suitable low-dimensional feature, which is quite different from RDCDL. Moreover, it is worth noting that RDCDL only adopts the intraclass scatter of coefficients as the discrimination constraint but neglects the interclass scatter of coefficients, while our JLDDP utilizes both the intraclass scatter and the interclass scatter to improve the discriminative ability of the learned dictionary. Fourth, Zhang et al. [40] proposed a SS-DSPP model which can simultaneously learn the dictionary and the projection matrix, but it is still very different from our JLDDP in the following aspects. SS-DSPP takes advantage of the relationship between the reconstruction error of training samples by the same class dictionary and the reconstruction error of training samples by different classes. Nevertheless, the discrimination constraint on coefficients is not considered in it. In addition, SS-DSPP also ignores the class information of low-dimensional features obtained after projection but only imposes an orthogonal constraint on the projection matrix, which leads to reducing the discrimination capability of the model to some extent. To solve these problems, our JLDDP utilizes the Fisher discrimination criterion to constrain the intraclass and interclass scatters of coefficients and low-dimensional samples, which can ensure the discrimination ability of the JLDDP model. In summary, although the proposed method shares several similarities with the aforementioned approaches [24, 40] and [46–49], our JLDDP is different from them in the dictionary learning process, projection learning process, or coefficient constraint. Specifically, JLDDP simultaneously learns the dictionary and projection matrix in a unified framework by adopting the intraclass and interclass scatter as the constraint of coefficients and the samples. Thus, JLDDP can explore the intrinsic relationship between the dictionary and the feature learning, which can improve the classification performance of both the image-based and the video-based face recognition.

4. Experimental Results

We conduct extensive experiments on image-based and video-based face databases to confirm the validity of JLDDP.

4.1. Image-Based Face Recognition Results and Analysis

4.1.1. Image Database Description. ORL [50], CMU PIE [51], FERET [52], and LFW [53] databases are used to prove the validity of JLDDP for image-based face recognition. Some examples from the ORL, CMU PIE, FERET, and LFW databases are shown in Figure 1.

The ORL face database includes 400 images of 40 subjects. The images reflect the changes of illumination, pose, expression, and whether glasses are worn. The CMU PIE face database includes 41,368 images of 68 subjects. In 43 distinct illumination conditions, images are taken across 13 various poses and with 4 diverse expressions. We adopt a subset of 24 images for each person in this experiment. The FERET database is recorded in a real environment with a lot of images. It includes 14,051 face images of more than 1,000 subjects. The face images have the characteristics of different expressions, postures, and illuminations. In addition, the time span of image acquisition in the FERET database is very large. We adopt a subset which contains 1,400 images of 200 subjects in this experiment. The LFW database is collected in unconstrained environments, which is very challenging. This database contains 13,233 face images of 5,749 subjects. However, most of the people have only one image in the database. Therefore, we select 158 subjects from LFW, which has at least 10 distinct images, to verify the effectiveness of algorithms. In [54], a new sparse representation-based alignment method is proposed for real-world images, which can eliminate the variety of orientations, expressions, and other factors as much as possible. We use this method to deal with the original LFW database for all the recognition methods. Table 1 provides the detailed database information. All images are clipped by selecting eye coordinates manually and normalized to 32×32 pixels.

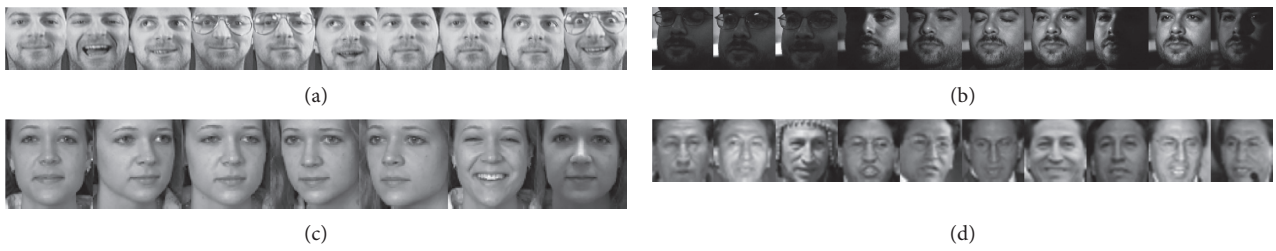

Figure 1: Examples from different databases: (a) ORL, (b) CMU PIE, (c) FERET, and (d) LFW.

Table 1: Details of the four image-based databases.

Databases	Images	Classes	Number
ORL	400	40	10
CMU PIE	1,632	68	24
FERET	1,400	200	7
LFW	1,580	158	10

4.1.2. Experiment Setting. In the image-based face recognition task, we compare our method with some representative methods, including SRC [9] with PCA and LDA, LCK-SVD [23], FDDL [24], DRSRC [34], LSD [29], DSRC [40], JDDRDL [41], and JNPDL [42]. The l_1-l_s toolbox [55] is adopted to handle the l_1-minimization problem in the SRC-related algorithms. The source code of the l_1-l_s toolbox can be found at http://web.stanford.edu/~boyd/l1_ls/. The source code of FDDL can be found at http://www4.comp.polyu.edu.hk/cslzhang/code/FDDL.zip. The source code of LC-KSVD can be found at http://users.umiacs.umd.edu/~zhuolin/projectlcksvd.html. The other methods are based on our implementations, and the parameters are tuned based on the settings reported in their papers. We set the number of atoms for each class of the dictionary in JLDDP as half of the training samples. Through randomly chosen training and testing samples, experiments are conducted 10 times totally, and the average recognition accuracies and standard deviations are reported. All the methods are developed in MATLAB and implemented on a computer with an Intel Core i3-2100 CPU at 3.2 GHz and 8 GB physical memory.

We first compare the recognition performance under various feature dimensions, and next, we compare the recognition performance under various number of training samples. For convenience, the number of training and testing samples is represented by l and h, respectively. Tables 2 and 3 show the data descriptions.

We compare the recognition performance under different parameter values. We adjust the parameter values by searching the grid {0, 0.0001, 0.001, 0.01, 0.1, 1} in an alternate manner to obtain the optimal parameter combination. Finally, we provide the convergence evaluation. We set the number of atoms for each class of the dictionary in JLDDP as half of the training samples. Through randomly chosen training and testing samples, experiments are conducted 10 times totally, and the average recognition accuracies and standard deviations are reported.

Table 2: Data description for different feature dimensions.

Databases	Train (l)	Test (h)	The reduced feature dimension
ORL	5	5	50, 60, 70, 80, 90
CMU PIE	7	17	50, 100, 150, 200, 250
FERET	4	3	50, 100, 150, 200, 250
LFW	5	5	450, 500, 550, 600, 650

Table 3: Data description for different number of training samples.

Databases	First round		Second round		Third round	
	Train (l)	Test (h)	Train (l)	Test (h)	Train (l)	Test (h)
ORL	2	8	5	5	7	3
CMU PIE	2	22	7	17	12	12
FERET	2	5	4	3	6	1
LFW	2	8	5	5	7	3

4.1.3. Recognition Results and Analysis. (1) Recognition Performance under Different Feature Dimensions. In the first experiment, we employ different feature dimensions to verify the performance of various methods. Table 2 shows the number of training samples and the reduced feature dimensions. The reduced feature dimension of LDA can be one less than the number of classes at most, and we cannot vary the feature dimensions as other methods. Thus, the results of LDA + SRC are not shown in the first experiment. In LC-KSVD and FDDL, PCA is adopted to reduce the sample dimension. Tables 4–7 demonstrate the recognition accuracies on the four databases by various number of dimensions. In most instances, the performance of JLDDP is better than the other methods. Moreover, several points can be seen from the tables. First, DRSRC is an unsupervised DR method that is designed based on SRC, so the accuracy is higher than PCA + SRC in most cases. This illustrates that the well-designed projection is more suitable for the classification. Second, compared with PCA + SRC and DRSRC, the average recognition accuracies of LCK-SVD, FDDL, and LSD are higher. The reason is that, after reducing the dimension of the samples with PCA and LCK-SVD, FDDL and LSD can learn a representative and discriminative dictionary, which is a key role in SRC. Third, LCK-SVD, FDDL, and LSD enhance the discrimination ability of the dictionary, but they do not jointly learn the projection that can preserve much discriminative information. Therefore, their performance is not as good as JDDRDL, DSRC, JNPDL, and

Table 4: The recognition accuracies (the top average recognition accuracy ± standard deviation %) of methods on the ORL database under various feature dimensions.

Method	dim = 50	dim = 60	dim = 70	dim = 80	dim = 90
PCA + SRC	84.50 ± 2.04	87.70 ± 1.57	89.25 ± 1.92	90.55 ± 1.82	91.50 ± 1.75
DRSRC	92.90 ± 1.41	93.25 ± 1.69	93.35 ± 1.49	93.40 ± 1.33	93.30 ± 0.79
LCK-SVD	84.00 ± 2.98	85.40 ± 1.90	87.60 ± 1.98	90.65 ± 1.60	90.90 ± 2.13
FDDL	86.35 ± 2.93	88.05 ± 2.07	88.85 ± 2.42	89.20 ± 2.37	89.05 ± 2.70
LSD	86.77 ± 1.72	88.78 ± 1.91	89.44 ± 1.60	90.67 ± 1.39	91.72 ± 1.58
JDDRDL	90.90 ± 1.96	92.30 ± 1.96	92.80 ± 1.40	94.35 ± 1.90	94.50 ± 1.55
DSRC	91.87 ± 1.12	92.03 ± 1.24	92.62 ± 1.34	93.13 ± 1.60	93.77 ± 1.71
JNPDL	92.44 ± 1.71	93.17 ± 1.63	93.86 ± 1.49	94.52 ± 1.53	94.98 ± 1.24
JLDDP	**97.25 ± 0.78**	**97.00 ± 1.03**	**97.30 ± 0.79**	**96.70 ± 1.21**	**97.05 ± 0.90**

Table 5: The recognition accuracies (the top average recognition accuracy ± standard deviation %) of methods on the CMU PIE database under various feature dimensions.

Method	dim = 50	dim = 100	dim = 150	dim = 200	dim = 250
PCA + SRC	89.55 ± 0.72	91.44 ± 0.70	92.08 ± 0.60	92.26 ± 0.62	92.23 ± 0.59
DRSRC	89.04 ± 0.84	91.38 ± 0.57	91.89 ± 0.62	92.08 ± 0.59	92.11 ± 0.70
LCK-SVD	74.80 ± 1.84	88.12 ± 1.10	88.91 ± 0.99	89.31 ± 0.89	89.16 ± 1.03
FDDL	78.52 ± 1.42	89.47 ± 0.86	91.72 ± 0.58	92.54 ± 0.44	92.90 ± 0.61
LSD	76.98 ± 0.51	89.43 ± 0.69	90.97 ± 0.55	92.37 ± 0.71	93.12 ± 0.93
JDDRDL	83.75 ± 2.27	91.59 ± 0.65	92.94 ± 0.65	93.34 ± 0.70	**93.30 ± 0.71**
DSRC	84.42 ± 1.54	88.37 ± 1.28	91.24 ± 0.97	91.65 ± 0.63	92.44 ± 0.53
JNPDL	86.63 ± 1.19	91.23 ± 1.35	93.18 ± 0.69	92.64 ± 0.77	92.15 ± 0.64
JLDDP	**90.09 ± 1.36**	**92.87 ± 0.99**	**94.07 ± 0.82**	**93.89 ± 0.87**	93.23 ± 0.79

Table 6: The recognition accuracies (the top average recognition accuracy ± standard deviation %) of methods on the FERET database under various feature dimensions.

Method	dim = 50	dim = 100	dim = 150	dim = 200	dim = 250
PCA + SRC	27.92 ± 1.08	43.12 ± 2.02	50.35 ± 1.87	53.48 ± 2.15	55.00 ± 1.69
DRSRC	28.67 ± 1.23	47.00 ± 1.78	51.17 ± 1.51	54.33 ± 1.67	55.33 ± 1.24
LCK-SVD	19.40 ± 1.46	22.83 ± 1.26	23.43 ± 1.61	23.37 ± 1.99	23.58 ± 1.28
FDDL	47.05 ± 1.90	63.12 ± 1.85	68.12 ± 2.36	70.37 ± 2.01	71.20 ± 1.81
LSD	39.79 ± 1.51	50.64 ± 1.93	63.77 ± 1.77	69.71 ± 1.59	71.88 ± 1.45
JDDRDL	56.65 ± 1.43	67.33 ± 1.74	69.10 ± 2.04	69.03 ± 1.85	67.95 ± 1.72
DSRC	60.79 ± 1.89	68.73 ± 2.04	70.46 ± 2.23	72.33 ± 2.16	73.72 ± 2.04
JNPDL	63.24 ± 2.47	72.68 ± 2.59	75.24 ± 2.47	77.18 ± 2.64	78.35 ± 2.71
JLDDP	**80.17 ± 2.08**	**79.79 ± 1.78**	**80.13 ± 2.57**	**80.25 ± 2.95**	**79.79 ± 2.19**

Table 7: The recognition accuracies (the top average recognition accuracy ± standard deviation %) of methods on the LFW database under various feature dimensions.

Method	dim = 450	dim = 500	dim = 550	dim = 600	dim = 650
PCA + SRC	32.10 ± 1.31	32.84 ± 2.29	32.91 ± 0.99	33.09 ± 1.39	33.09 ± 1.37
DRSRC	33.21 ± 1.54	34.07 ± 1.83	34.93 ± 1.31	35.77 ± 1.63	36.06 ± 1.42
LCK-SVD	37.78 ± 2.45	38.04 ± 2.57	38.67 ± 2.42	39.73 ± 2.61	41.73 ± 2.78
FDDL	44.53 ± 1.72	45.61 ± 2.03	47.46 ± 1.72	48.44 ± 1.93	49.20 ± 1.82
LSD	41.72 ± 2.69	43.37 ± 2.44	43.63 ± 2.38	44.78 ± 2.53	46.93 ± 2.31
JDDRDL	51.90 ± 2.47	53.69 ± 2.73	59.78 ± 2.60	57.42 ± 2.92	55.68 ± 2.84
DSRC	64.38 ± 2.86	67.24 ± 3.19	68.52 ± 3.25	66.49 ± 2.50	63.74 ± 2.79
JNPDL	72.02 ± 2.46	71.77 ± 2.57	70.82 ± 2.92	71.39 ± 2.58	70.87 ± 2.68
JLDDP	**73.28 ± 2.61**	**76.77 ± 2.35**	**74.63 ± 2.57**	**75.06 ± 2.44**	**73.29 ± 2.73**

JLDDP under different feature dimensions. Fourth, JLDDP outperforms JDDRDL, DSRC, and JNPDL significantly under different feature dimensions on the four databases, except when the feature dimension is 250 on the CMU PIE database, in which the best average recognition result of JDDRDL is only 0.07% higher than that of JLDDP.

Nevertheless, the experimental results still indicate that JLDDP can achieve relatively stable and high recognition accuracy in general under different feature dimensions. The superiority of our approach is due to that JLDDP can discover the latent discriminative ability of samples in the low-dimensional space and learn the class-specific dictionary simultaneously.

(2)·Recognition Performance under Various Number of Training Samples. The effectiveness of JLDDP under various number of training samples is compared with other methods on the ORL, CMU PIE, FERET, and LFW databases. The number of training samples and test samples used is listed in Table 3. Tables 8–11 show the recognition accuracies and the corresponding feature dimensions. The corresponding feature dimensions are annotated in parentheses. When there are only 2 training samples per subject, JDDRDL, DSRC, JNPDL, and JLDDP that learned the dictionary and projection jointly obtain better performance than other methods. When the number of training samples is increased, the performance of all the methods is improved in general, except for the LDA + SRC and LCK-SVD methods in the FERET database. Compared with other methods, JLDDP can achieve the best average recognition accuracies and a relatively small feature dimension, which demonstrate its capability to address practical applications.

(3)·Recognition Performance under Different Parameter Values. We test the impacts of various parameter values on four image-based face recognition databases. Since there are three parameters in the proposed JLDDP, we fix two of them and then analyze the influence of the remaining parameter. The physical meaning of the parameters is described in Section 3. For the ORL, CMU PIE, FERET, and LFW databases, the number of training samples is set as 5, 7, 4, and 5, respectively. The top average recognition results obtained by JLDDP under various parameter values are shown in Figure 2. When the parameter values of ω_1, ω_2, and ω_3 equal to zero, the recognition accuracy of JLDDP is relatively low, which indicates that each term in the objective function of JLDDP is significant for classification. With the increasing of each parameter value, the performance of JLDDP improves gradually. When $\omega_1 = 0.0001$, $\omega_2 = 0.0001$ or 0.001, and $\omega_3 = 0.001$ or 0.01, the proposed JLDDP performs best on the four databases. However, after achieving its best performance, the recognition accuracy dramatically decreases with the increase of each parameter value. Hence, ω_1, ω_2, and ω_3 should be set as moderate values to obtain a good performance, which is conform to our analysis in Section 3. That is, if the parameter value is too large, the corresponding term in equation (11) will play a leading role, which makes other terms be neglected. In contrast, if the parameter value is too small, the corresponding term will lose its constraint ability.

To further evaluate the role of each term in our model, we, respectively, set the parameter values of ω_1, ω_2, and ω_3 as zero to test the performance of JLDDP. Here, the number of training samples is set as 5, 7, 4, and 5 for ORL, CMU PIE, FERET, and LFW databases, respectively. The top average recognition results obtained by JLDDP under various situations are shown in Table 12. In this table, the baselines are results obtained by the optimal parameter combination in Tables 9–11. From the experimental results, we can see that the proposed method cannot achieve its best recognition accuracies when one of the parameters ω_1, ω_2, and ω_3 is equal to zero, which indicates that the sparse constraint term, the coding coefficient term, and the projection learning term are all essential to improve the recognition performance of our JLDDP method. Besides, the recognition accuracies are dramatically decreased when ω_1 is set as zero, that is, the sparse constraint term is omitted, which indicates the sparse constraint in the dictionary representation is very important to improve the discriminative ability of our model. Furthermore, the recognition accuracies are very close when ω_2 or ω_3 is set as zero, but much lower than the baselines. This means the coding coefficient term and the projection learning term are also indispensable in our JLDDP since they can bring the intraclass and interclass information into our model to ensure the discrimination of coefficients and low-dimensional features.

(4)·Convergence Evaluation. Figure 3 demonstrates the convergence curves of JLDDP on the ORL, CMU PIE, FERET, and LFW databases. In each figure, the x-axis represents the iteration number, and the y-axis represents the value of the objective function. From this figure, we can find that the proposed iterative updating algorithm of JLDDP is convergent, which is conformable to our convergence analysis in Section 3.

4.2. Video-Based Face Recognition Results and Analysis

4.2.1. Classification Scheme. To further evaluate the performance of JLDDP, we perform face recognition experiments on video. Here, we suppose $V^t = \{v_1^t, \ldots, v_j^t, \ldots, v_{n_t}^t\}$ is a testing face video, where v_j^t is the j-th $(1 \le j \le n_t)$ frame and n_t is the total number of frames. According to Lu et al. [43], we project each frame into a low-dimensional feature space by the learned projection P and then obtain the corresponding coding coefficients by equation (17). Finally, the class label of the frame can be obtained by the following equation as [42]

$$\text{label}\left(v_j^t\right) = \arg\min_i \left\| P^T v_j^t - D_i D_i^\dagger v_j^t \right\|_2^2, \tag{19}$$

where $D_i^\dagger = (D_i^T D_i)^{-1} D_i^T$ is the pseudo-inverse of D_i and $D_i D_i^\dagger v_j^t$ is the projection of v_j^t onto the span of atoms in D_i [26]. Finally, we apply the majority voting to determine the testing video's label after obtaining the entire frames' label:

$$i^* = \arg\max_i Z_i, \tag{20}$$

where Z_i denotes the total votes from the i-th class.

4.2.2. Video Database Description. The Honda [56], MoBo [57], and YTC [58] databases are employed to verify the performance of JLDDP. All the videos in the Honda database are recorded indoors with normal lighting conditions and

Table 8: The recognition accuracies (the top average recognition accuracy ± standard deviation %) of methods on the ORL database under various number of training samples.

Method	$l = 2$	$l = 5$	$l = 7$
PCA + SRC	75.47 ± 2.38 (78)	95.10 ± 0.99 (188)	92.50 ± 2.52 (77)
LDA + SRC	77.12 ± 2.01 (39)	93.20 ± 1.51 (39)	95.17 ± 1.61 (39)
DRSRC	75.91 ± 2.68 (77)	95.20 ± 0.89 (197)	96.00 ± 2.28 (200)
LCK-SVD	77.34 ± 1.93 (1024)	94.25 ± 1.72 (1024)	95.50 ± 1.72 (1024)
FDDL	79.78 ± 2.04 (1024)	96.20 ± 0.75 (1024)	97.00 ± 1.31 (1024)
LSD	80.63 ± 2.43(1024)	97.15 ± 0.44 (1024)	97.91 ± 1.73 (1024)
JDDRDL	80.19 ± 1.86 (50)	95.55 ± 1.83 (100)	93.92 ± 1.25 (70)
DSRC	76.31 ± 1.22 (100)	93.79 ± 1.93 (110)	95.69 ± 1.46 (130)
JNPDL	80.25 ± 1.13 (90)	95.67 ± 1.52 (100)	96.36 ± 1.27 (110)
JLDDP	**82.94 ± 1.78 (60)**	**97.30 ± 0.79 (70)**	**97.33 ± 1.35 (70)**

Table 9: The recognition accuracies (the top average recognition accuracy ± standard deviation %) of methods on the CMU PIE database under various number of training samples.

Method	$l = 2$	$l = 7$	$l = 12$
PCA + SRC	77.11 ± 1.44 (150)	92.34 ± 0.67 (450)	93.77 ± 0.97 (450)
LDA + SRC	75.86 ± 1.09 (67)	89.46 ± 0.88 (67)	89.62 ± 0.89 (67)
DRSRC	77.85 ± 1.32 (150)	92.34 ± 0.65 (450)	93.73 ± 0.87 (450)
LCK-SVD	73.19 ± 1.83 (1024)	89.19 ± 0.68 (1024)	90.45 ± 1.14 (1024)
FDDL	49.49 ± 1.45 (1024)	93.32 ± 0.56 (1024)	93.93 ± 0.79 (1024)
LSD	46.33 ± 1.46 (1024)	91.93 ± 0.78 (1024)	94.13 ± 1.02 (1024)
JDDRDL	78.50 ± 1.51 (500)	93.34 ± 0.70 (200)	94.55 ± 1.05 (250)
DSRC	78.04 ± 1.66 (300)	92.96 ± 0.61 (250)	94.07 ± 1.45 (200)
JNPDL	78.72 ± 1.67 (250)	93.22 ± 0.63 (200)	94.18 ± 1.73 (200)
JLDDP	**79.61 ± 1.30 (100)**	**94.07 ± 0.82 (150)**	**95.27 ± 0.80 (250)**

Table 10: The recognition accuracies (the top average recognition accuracy ± standard deviation %) of methods on the FERET database under various number of training samples.

Method	$l = 2$	$l = 4$	$l = 6$
PCA + SRC	39.22 ± 1.41 (400)	57.60 ± 1.94 (600)	68.50 ± 2.11 (350)
LDA + SRC	26.50 ± 1.12 (199)	19.87 ± 1.66 (199)	20.30 ± 2.65 (199)
DRSRC	38.73 ± 1.50 (400)	56.33 ± 1.93 (300)	68.70 ± 1.95 (350)
LCK-SVD	37.73 ± 1.85 (1024)	37.30 ± 1.43 (1024)	38.30 ± 2.41 (1024)
FDDL	49.49 ± 1.45 (1024)	76.25 ± 1.65 (1024)	74.50 ± 2.05 (1024)
LSD	39.97 ± 1.68 (1024)	63.28 ± 1.86 (1024)	75.69 ± 1.91 (1024)
JDDRDL	47.37 ± 1.02 (100)	69.10 ± 2.04 (150)	77.50 ± 1.99 (200)
DSRC	50.17 ± 1.52 (150)	74.66 ± 2.13 (200)	83.93 ± 1.87 (200)
JNPDL	55.49 ± 1.37 (200)	75.31 ± 2.05 (250)	86.29 ± 1.90 (250)
JLDDP	**58.08 ± 1.62 (50)**	**80.13 ± 2.57 (150)**	**88.50 ± 1.25 (200)**

Table 11: The recognition accuracies (the top average recognition accuracy ± standard deviation %) of methods on the LFW database under various number of training samples.

Method	$l = 2$	$l = 5$	$l = 7$
PCA + SRC	26.37 ± 2.51 (250)	43.14 ± 2.64 (200)	46.91 ± 3.02 (300)
LDA + SRC	20.45 ± 3.48 (157)	27.69 ± 3.72 (157)	31.92 ± 3.54 (157)
DRSRC	32.73 ± 3.14 (300)	46.48 ± 3.57 (400)	49.38 ± 3.17 (400)
LCK-SVD	31.62 ± 2.08 (1024)	42.05 ± 2.11 (1024)	44.73 ± 2.32 (1024)
FDDL	42.68 ± 2.82 (1024)	53.41 ± 3.05 (1024)	58.66 ± 3.24 (1024)
LSD	38.76 ± 2.66 (1024)	47.60 ± 2.86 (1024)	57.38 ± 2.90 (1024)
JDDRDL	44.91 ± 2.39 (500)	59.78 ± 2.60 (550)	66.62 ± 2.14 (450)
DSRC	47.23 ± 2.53 (450)	68.52 ± 3.25 ((450)	70.31 ± 2.78 (500)
JNPDL	50.37 ± 3.11 (550)	73.02 ± 2.61 (450)	75.31 ± 2.93 (450)
JLDDP	**53.46 ± 2.42 (450)**	**76.77 ± 2.35 (500)**	**80.06 ± 2.29 (450)**

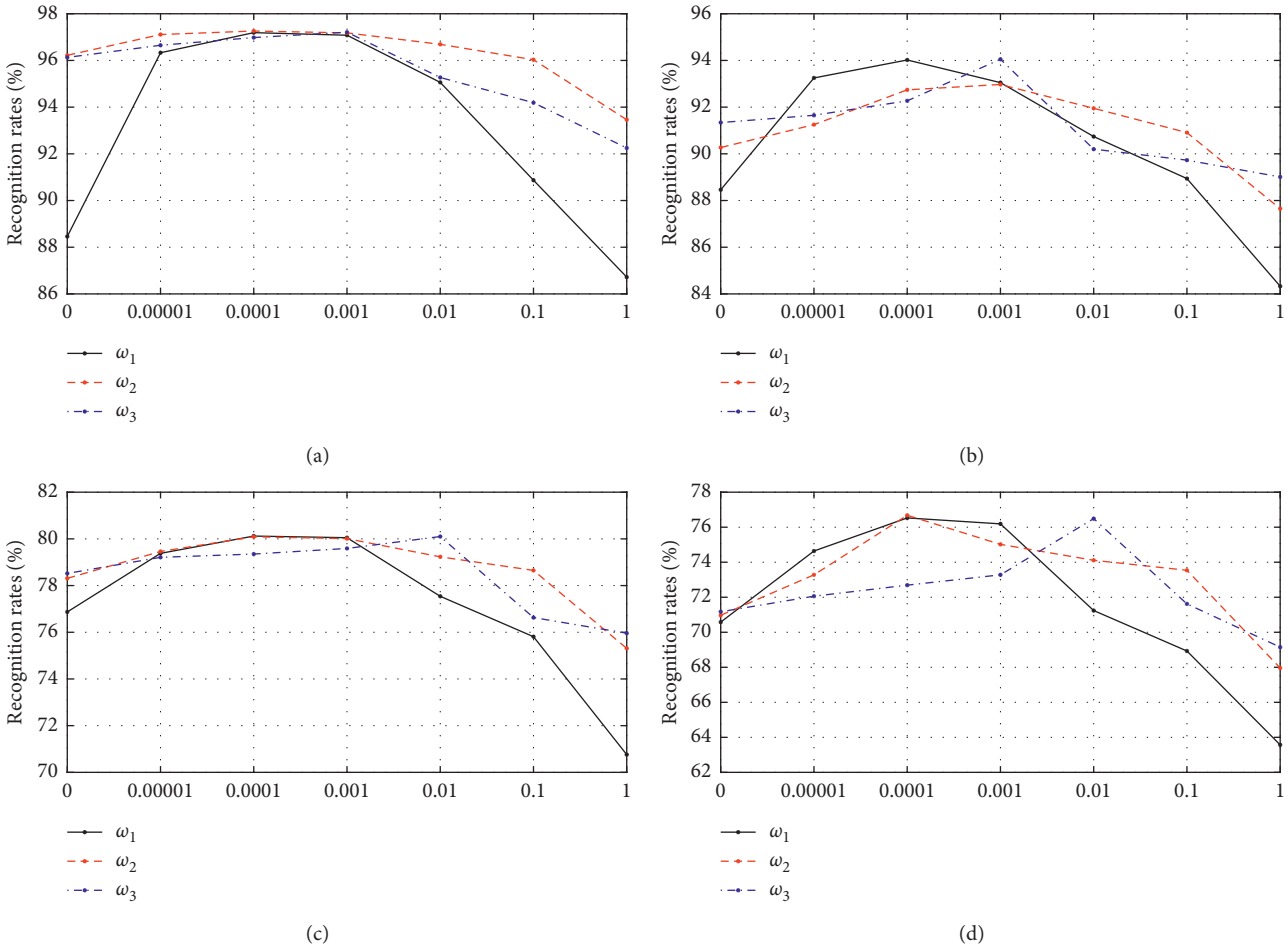

Figure 2: The performance of JLDDP under various parameter values on the (a) ORL, (b) CMU PIE, (c) FERET, and (d) LFW databases.

Table 12: The recognition accuracies (the top average recognition accuracy ± standard deviation %) of methods on the ORL, CMU PIE, FERET, and LFW databases under various situations.

Databases	$\omega_1 = 0$	$\omega_2 = 0$	$\omega_3 = 0$	Baseline
ORL	88.46 ± 1.31	96.22 ± 1.07	96.13 ± 1.22	97.26 ± 1.19
CMU PIE	88.71 ± 1.64	90.27 ± 1.20	91.34 ± 1.31	94.02 ± 1.07
FERET	76.87 ± 1.73	78.31 ± 1.98	78.52 ± 2.19	80.12 ± 2.32
LFW	70.58 ± 1.88	70.97 ± 2.31	71.17 ± 2.67	76.68 ± 2.54

include different facial expressions and a large range of head movement. The Honda database contains 59 videos of 20 subjects. Each video clip comprises 12 to 645 frames. The MoBo database is designed for the identification of long-distance people, which is captured with fixed-position cameras. The MoBo database comprises 96 videos of 24 subjects, which include large head-pose variations. Each subject comprises 4 videos, about 300 frames per video. The YTC database is collected from YouTube, which has 1,910 videos of 47 subjects. These subjects are politicians, actors, or actresses. It is a large low-resolution video database for face recognition, which is highly compressed. Each video contains 8 to 400 frames. In the experiment, the cascaded face detector [59] is used to detect the face, and then all the faces are resized to grayscale images with 30×30 pixels.

4.3. Experiment Setting. We compare the proposed JLDDP with several existing classical video-based face recognition methods, including MSM [60], DCC [61], MMD [62], MDA [63], AHISD [64], CHISD [64], SANP [65], DFRV [27], LSD [29], and SFDL [43]. The source code of DCC can be found at http://mi.eng.cam.ac.uk/~tkk22. The source code of AHISD and CHISD can be found at http://mlcv.ogu.edu.tr/softwareimageset.html. Since the source codes of other methods are not provided by their authors, we implement them by ourselves and follow the same parameter settings in their corresponding papers. In the video-based experiments, the parameters ω_1, ω_2, and ω_3 of JLDDP are empirically set as 0.0001, 0.0005, and 0.005, respectively. The number of atoms per class for the Honda, MoBo, and YTC databases is set as 20, 25, and 40, respectively. We select the best accuracy that

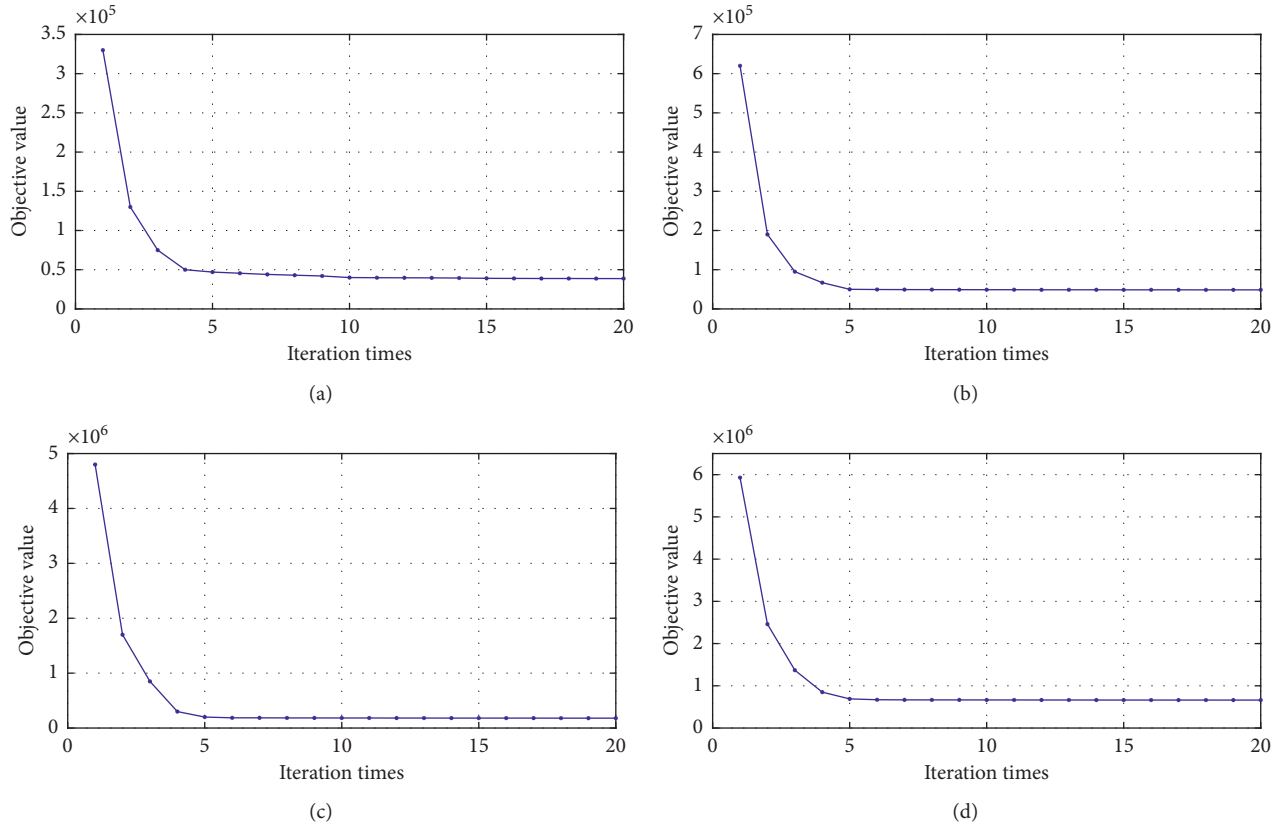

Figure 3: The convergence curves of JLDDP on four databases: (a) ORL, (b) CMU PIE, (c) FERET, and (d) LFW.

Table 13: The classification accuracies (the top average recognition accuracy ± standard deviation %) of methods on the Honda, MoBo, and YTC databases.

Method	Honda	MoBo	YTC
MSM	90.26 ± 2.15	85.57 ± 3.74	60.78 ± 4.38
DCC	94.87 ± 1.46	91.53 ± 1.66	63.91 ± 4.71
MMD	94.87 ± 2.05	89.72 ± 3.48	66.47 ± 3.77
MDA	97.44 ± 1.22	95.97 ± 1.90	67.89 ± 4.62
AHISD	89.74 ± 2.69	94.58 ± 2.57	66.71 ± 5.36
CHISD	92.31 ± 2.17	96.52 ± 1.18	67.34 ± 6.12
SANP	93.69 ± 1.79	97.08 ± 1.03	67.96 ± 5.91
DFRV	97.44 ± 2.73	94.47 ± 2.10	73.49 ± 5.20
LSD	**100.00 ± 0.00**	95.69 ± 2.04	72.93 ± 6.07
SFDL	**100.00 ± 0.00**	96.71 ± 1.77	76.24 ± 5.48
JLDDP	**100.00 ± 0.00**	**97.72 ± 1.45**	**78.31 ± 5.04**

JLDDP achieves with projected dimensions from 50, 100, 150, 200, and 300. All results are the average value of 10 times' independent experiments with different training set selection.

In the first experiment, the proposed JLDDP is compared with the state-of-the-art methods. The training set of the Honda and MoBo databases contains one video of each subject, and the testing set contains the remaining videos. If the subject has only one video, we separate the video into two clips and select one video for training and another video for testing randomly. The training set of the YTC database contains 3 videos of each subject, and the testing set contains 6 videos of each subject. In the second experiment, the influence of different training and testing frames on the performance of various methods is tested. We randomly choose 50, 100, and 200 frames from each video as the training set and another 50, 100, and 200 frames as the testing set.

4.4. Recognition Results and Analysis

4.4.1. Comparison with the Contrast Methods. In the first experiment, our JLDDP is compared with several existing methods. Table 13 tabulates the recognition accuracies of the methods on the Honda, MoBo, and YTC databases. The

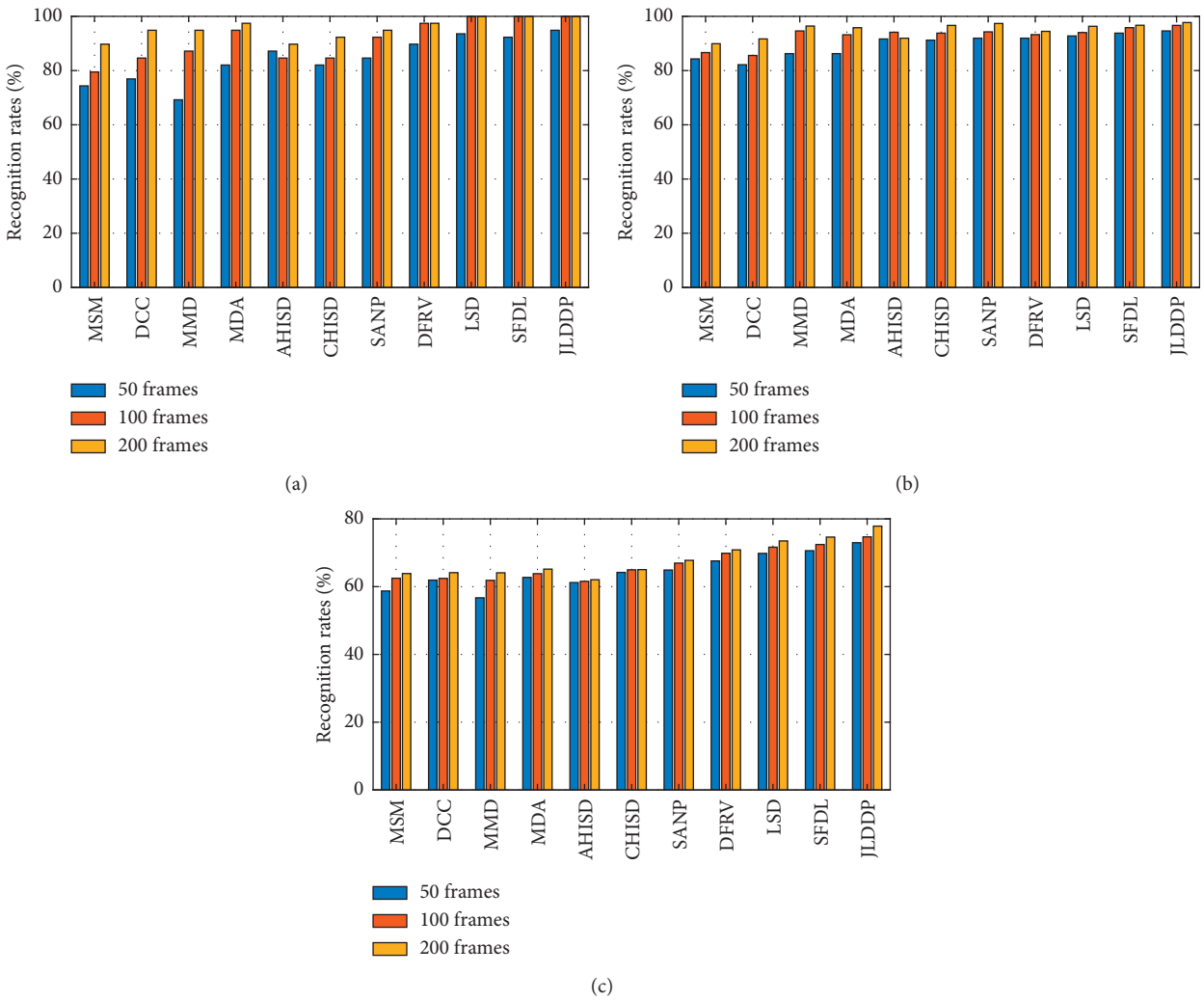

Figure 4: Comparison of the top average recognition rates (%) on (a) Honda, (b) MoBo, and (c) YTC databases with various number of frames.

recognition accuracies of MDA, LSD, SFDL, and JLDDP are higher than those of MSM, DCC, MMD, AHISD, CHISD, SANP, and DFRV in most cases. Therefore, we can infer that the supervised methods can exploit more discriminative information than the unsupervised methods. Moreover, our JLDDP surpasses the compared methods. The main reason is JLDDP can project the frames into a discriminative low-dimensional subspace, which is beneficial to obtain the discriminative coding coefficients with the class-specific dictionary.

4.4.2. Comparison under Various Number of Frames. In the second experiment, various number of frames are selected as the training set to compare the robustness of JLDDP with other methods. Figure 4 shows the top average recognition accuracies of different methods on the Honda, MoBo, and YTC databases with various number of frames. The recognition accuracies are improved with increasing of the number of frames. JLDDP can achieve the best recognition accuracy with different numbers of frames. This is because joint learning of the projection and dictionary can enable JLDDP to obtain more discriminative information.

5. Conclusions

This paper presents a JLDDP method for sparse representation-based face recognition. By combining DL and DR into a unified framework, our JLDDP obtains the adaptive projection and dictionary. The proposed JLDDP achieves commendable performance and robustness on seven benchmark image-based and video-based databases. Moreover, an effective iterative algorithm is proposed to solve the optimization problem, and the convergence is strictly proven.

Data Availability

The data are derived from public domain resources.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant nos. 61602221, 61672150, and 61806126, in part by the Fund of the Jilin Provincial Science and Technology Department under Grant nos. 20200201199JC, 20180201089GX, 20190201305JC, 20200401081GX, and 20200401086GX, in part by the Fund of Education Department of Jilin Province under Grant nos. JJKH20190294KJ and JJKH20190291KJ, in part by the Natural Science Foundation of Jiangxi Province under Grant no. 20171BAB212009, in part by the Science and Technology Research Project of Jiangxi Provincial Department of Education under Grant no. GJJ160333, and in part by the Funds for the Central Universities under Grant nos. 2412018QD029, 2412019FZ049, and 2412020FZ031.

References

[1] C. Ding and D. Tao, "Trunk-branch ensemble convolutional neural networks for video-based face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 4, pp. 1002–1014, 2018.

[2] W. Liu, Y. Wen, Z. Yu et al., "Sphereface: deep hypersphere embedding for face recognition," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 212–220, July 2017.

[3] C. Ding, J. Choi, D. Tao, and L. S. Davis, "Multi-directional multi-level dual-cross patterns for robust face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 3, pp. 518–531, 2016.

[4] J. Gou, L. Wang, B. Hou, J. Lv, Y. Yuan, and Q. Mao, "Two-phase probabilistic collaborative representation-based classification," *Expert Systems with Applications*, vol. 133, pp. 9–20, 2019.

[5] X.-Y. Jing and D. Zhang, "A face and palmprint recognition approach based on discriminant DCT feature extraction," *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 6, pp. 2405–2415, 2004.

[6] C. Bi, L. Zhang, M. Qi et al., "Supervised filter learning for representation based face recognition," *PLoS One*, vol. 11, no. 7, Article ID e0159084, 2016.

[7] Y. Yi, C. Bi, X. Li, J. Wang, and J. Kong, "Semi-supervised local ridge regression for local matching based face recognition," *Neurocomputing*, vol. 167, pp. 132–146, 2015.

[8] J. Wang, Y. Yi, W. Zhou et al., "Locality constrained joint dynamic sparse representation for local matching based face recognition," *PLoS One*, vol. 9, no. 11, Article ID e113198, 2014.

[9] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[10] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary," in *Computer Vision-ECCV 2010*, pp. 448–461, Springer, Berlin, Germany, 2010.

[11] K. Huang and S. Aviyente, "Sparse representation for signal classification," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 609–616, Vancouver, Canada, December 2006.

[12] 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.

[13] Y. Xu, Z. Li, J. Yang, and D. Zhang, "A survey of dictionary learning algorithms for face recognition," *IEEE Access*, vol. 5, pp. 8502–8514, 2017.

[14] X.-Y. Jing, X. Zhu, F. Wu et al., "Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning," in *Proceedings of the IEEE 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 695–704, Boston, MA, USA, June 2015.

[15] M. Zhou, H. Chen, J. Paisley et al., "Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 130–144, 2012.

[16] C. Zheng, Y. Yi, M. Qi et al., "Multicriteria-based active discriminative dictionary learning for scene recognition," *IEEE Access*, vol. 6, pp. 4416–4426, 2017.

[17] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *Proceedings of the 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 2443–2446, Kobe, Japan, December 1999.

[18] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[19] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 801–808, Vancouver, Canada, December 2007.

[20] K. Skretting and K. Engan, "Recursive least squares dictionary learning algorithm," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2121–2130, 2010.

[21] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Metaface learning for sparse representation based face recognition," in *Proceedings of the 2010 IEEE International Conference on Image Processing*, pp. 1601–1604, Hong Kong, September 2010.

[22] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2691–2698, San Francisco, CA, USA, June 2010.

[23] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651–2664, 2013.

[24] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based Fisher discrimination dictionary learning for image classification," *International Journal of Computer Vision*, vol. 109, no. 3, pp. 209–232, 2014.

[25] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley, Hoboken, NY, USA, 2012.

[26] B. Ding and H. Ji, "Learning kernel-based robust disturbance dictionary for face recognition," *Applied Sciences*, vol. 9, no. 6, pp. 2178–2191, 2019.

[27] Y.-C. Chen, V. M. Patel, P. J. Phillips, and R. Chellappa, "Dictionary-based face recognition from video," *Computer Vision-ECCV 2012*, Springer, Berlin, Germany, pp. 766–779, 2012.

[28] Y. C. Chen, V. M. Patel, S. Shekhar, R. Chellappa, and P. J. Phillips, "Video-based face recognition via joint sparse representation," in *Proceedings of the 2013 10th IEEE*

International Conference and Workshops on Automatic Face and Gesture Recognition (FG), pp. 1–8, Shanghai, China, April 2013.

[29] H. Xu, J. Zheng, A. Alavi, and R. Chellappa, "Learning a structured dictionary for video-based face recognition," in *Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–9, Lake Placid, NY, USA, March 2016.

[30] C. Zheng, R. Zhao, F. Liu et al., "Dimensionality reduction via multiple locality-constrained graph optimization," *IEEE Access*, vol. 6, pp. 54479–54494, 2018.

[31] X.-Y. Jing, X. Zhang, X. Zhu et al., "Multiset feature learning for highly imbalanced data classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, In press.

[32] Y. Yi, J. Wang, W. Zhou, C. Zheng, J. Kong, and S. Qiao, "Non-Negative matrix factorization with locality constrained adaptive graph," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 2, pp. 427–441, 2020.

[33] Y. Yi, J. Wang, W. Zhou, Y. Fang, J. Kong, and Y. Lu, "Joint graph optimization and projection learning for dimensionality reduction," *Pattern Recognition*, vol. 92, pp. 258–273, 2019.

[34] L. Zhang, M. Yang, Z. Feng, and D. Zhang, "On the dimensionality reduction for sparse representation based face recognition," in *Proceedings of the 20th International Conference on Pattern Recognition*, pp. 1237–1240, Istanbul, Turkey, August 2010.

[35] L. Clemmensen, T. Hastie, D. Witten, and B. Ersbøll, "Sparse discriminant analysis," *Technometrics*, vol. 53, no. 4, pp. 406–413, 2011.

[36] J. Yang, D. Chu, L. Zhang, Y. Xu, and J. Yang, "Sparse representation classifier steered discriminative projection with applications to face recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 7, pp. 1023–1035, 2013.

[37] J. Gui, Z. Sun, W. Jia, R. Hu, Y. Lei, and S. Ji, "Discriminant sparse neighborhood preserving embedding for face recognition," *Pattern Recognition*, vol. 45, no. 8, pp. 2884–2893, 2012.

[38] G.-F. Lu, Z. Lin, and Z. Jin, "Face recognition using discriminant locality preserving projections based on maximum margin criterion," *Pattern Recognition*, vol. 43, no. 10, pp. 3572–3579, 2010.

[39] H. V. Nguyen, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Sparse embedding: a framework for sparsity promoting dimensionality reduction," *Computer Vision-ECCV 2012*, Springer, Berlin, Germany, pp. 414–427, 2012.

[40] H. Zhang, Y. Zhang, and T. S. Huang, "Simultaneous discriminative projection and dictionary learning for sparse representation based classification," *Pattern Recognition*, vol. 46, no. 1, pp. 346–354, 2013.

[41] Z. Feng, M. Yang, L. Zhang, Y. Liu, and D. Zhang, "Joint discriminative dimensionality reduction and dictionary learning for face recognition," *Pattern Recognition*, vol. 46, no. 8, pp. 2134–2143, 2013.

[42] W. Liu, Z. Yu, Y. Wen, R. Lin, and M. Yang, "Jointly learning non-negative projection and dictionary with discriminative graph constraints for classification," in *Proceedings of the British Machine Vision Conference 2016*, pp. 1–12, York, UK, September 2016.

[43] J. Lu, G. Wang, and J. Zhou, "Simultaneous feature and dictionary learning for image set based face recognition," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 4042–4054, 2017.

[44] L. Rosasco, S. Mosci, M. Santoro, A. Verri, and S. Villa, "Iterative projection methods for structured sparsity regularization," MIT Technical reports, MIT-CSAIL-TR-2009-050, CBCL-282, Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.

[45] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, New York, NY, USA, 1964.

[46] D. Wang and S. Kong, "A classification-oriented dictionary learning model: explicitly learning the particularity and commonality across categories," *Pattern Recognition*, vol. 47, no. 2, pp. 885–898, 2014.

[47] Y. Sun, Q. Liu, J. Tang, and D. Tao, "Learning discriminative dictionary for group sparse representation," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3816–3828, 2014.

[48] S. Gao, I. Tsang, and Y. Ma, "Learning category-specific dictionary and shared dictionary for fine-grained image categorization," *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, vol. 23, no. 2, pp. 623–634, 2014.

[49] G. Lin, M. Yang, J. Yang, L. Shen, and W. Xie, "Robust, discriminative and comprehensive dictionary learning for face recognition," *Pattern Recognition*, vol. 81, pp. 341–356, 2018.

[50] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the 1994 IEEE Workshop on Applications of Computer Vision*, pp. 138–142, Saratosa, FL, USA, December 1994.

[51] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proceedings of the Fifth IEEE International Conference on Automatic Face Gesture Recognition*, pp. 53–58, Washington, DC, USA, May 2002.

[52] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[53] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Techical report 07-49, University of Massachusetts Amherst, Amherst, MA, USA, 2007.

[54] X. Fontaine, R. Achanta, and S. Süsstrunk, "Face recognition in realworld images," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1–5, New Orleans, LA, USA, March 2017.

[55] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale -regularized least squares," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 606–617, 2007.

[56] K. C. Lee, J. Ho, M. H. Yang, and D. Kriegman, "Video-based face recognition using probabilistic appearance manifolds," in *Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 313–320, Madison, WI, USA, June 2003.

[57] R. Gross and J. Shi, "The CMU motion of body (MoBo) database," CMU Technical reports, CMU-RI-TR-01-18, Carnegie Mellon University, Pittsburgh, PA, USA, 2001.

[58] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, "Face tracking and recognition with visual constraints in real-world videos," in *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1787–1794, Anchorage, AK, USA, June 2008.

[59] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[60] O. Yamaguchi, K. Fukui, and K. Maeda, "Face recognition using temporal image sequence," in *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 318–323, Nara, Japan, April 1998.

[61] T.-K. Kim, J. Kittler, and R. Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1005–1018, 2007.

[62] R. Wang, S. Shan, X. Chen, and W. Gao, "Manifold-manifold distance with application to face recognition based on image set," in *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, Anchorage, AK, USA, June 2008.

[63] R. Wang and X. Chen, "Manifold discriminant analysis," in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 429–436, Miami, FL, USA, June 2009.

[64] H. Cevikalp and B. Triggs, "Face recognition based on image sets," in *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2567–2573, San Francisco, CA, USA, June 2010.

[65] Y. Hu, A. S. Mian, and R. Owens, "Sparse approximated nearest points for image set classification," in *Proceedings of the IEEE Conference Computer Vision Pattern Recognition*, pp. 121–128, Colorado Springs, CO, USA, June 2011.