

Nuclear norm-based matrix regression preserving embedding for face recognition

Yang-Jun Deng^a, Heng-Chao Li^{a,*}, Qi Wang^b, Qian Du^c

^aSichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University, Chengdu 611756, China

^bSchool of Computer Science and Center for OPTICAL IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China

^cDepartment of Electrical and Computer Engineering, Mississippi State University, MS 39762, USA



ARTICLE INFO

Article history:

Received 3 January 2018

Revised 30 April 2018

Accepted 18 May 2018

Available online 26 May 2018

Communicated by Dr. Ran He

Keywords:

Nuclear norm

Matrix regression

Dimensionality reduction

Face recognition

ABSTRACT

Recently, using linear reconstruction technique to construct intrinsic graph for projection-based dimensionality reduction (DR) has aroused broad interest in face recognition. However, current methods either lack robustness to corruptions or require to perform vectorization which causes loss of local geometrical information of images. To this end, a novel nuclear norm-based matrix regression preserving embedding (NN-MRPE) method is proposed in this paper. First, NN-MRPE constructs an intrinsic graph by using the nuclear norm to evaluate the residual errors to resist data corruptions. Second, a matrix-based embedding cost function is formulated to seek two transformation matrices which can preserve the geometrical structure reflected by the intrinsic graph exactly. Finally, based on the linear regression theory, we summarize a general DR framework called linear regression preserving embedding that preserves the intrinsic structure of data by recovering the reconstruction relationship in the original space. Specifically, many existing approaches are the special cases of the linear regression preserving embedding. Experiments on five public face databases with different types of corruptions are conducted to demonstrate the efficiency of the proposed NN-MRPE method.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Face recognition has received great attention in pattern recognition and computer vision fields during the past decades [1]. In order to solve the problem known as “curse of dimensionality” [2,3], dimensionality reduction (DR) also has aroused broad interest. Up to now, numerous DR methods have been developed [4]. Among them projection-based DR methods have become a hot topic due to their simplicity and efficiency [5]. These projection-based DR methods can be generally divided into three categories: supervised, unsupervised, and semi-supervised scenarios. The most well-known supervised methods include linear discriminant analysis (LDA) [6] and its variants, such as maximum margin criterion (MMC) [7], adaptive maximum margin analysis (AMMA) [8], locality Fisher discrimination analysis (LFDA) [9], and so on. The unsupervised methods include principal component analysis (PCA) [10], locality preserving projections (LPP) [11], and collaborative representation based projections method (CRP) [12], etc. The semi-supervised methods include semi-supervised discriminant analysis

(SDA) [13], semi-supervised Laplacian eigenmaps [14], etc. Since the prior knowledge is often difficult and expensive to obtain in many practical applications, this paper will only focus on unsupervised scenario.

PCA is the most popular unsupervised method which aims at mapping high dimensional data into a lower dimensional space by seeking the direction of maximum variance [15]. However, PCA fails to discover the nonlinear structure of data. To overcome this disadvantage, the kernel-based techniques are developed to deal with nonlinear DR problems implicitly [16,17]. However, most of them are unable to explicitly treat data manifold structure. Furthermore, kernel selection and optimal kernel parameter assignment are generally difficult in many real applications.

To address the above problems, many manifold learning methods have been proposed for nonlinear DR. The representative methods include locally linear embedding (LLE) [18], isometric feature mapping (ISOMAP) [19], and Laplacian eigenmaps (LE) [20]. However, these nonlinear methods often suffer from the so-called out-of-sample problem [21]. Thus, linear approximate technique is introduced to improve manifold learning methods. For example, neighborhood preserving embedding (NPE) proposed in [22] is a linearization of LLE, which aims at preserving local neighborhood

* Corresponding author.

E-mail address: lihengchao_78@163.com (H.-C. Li).

structure by constructing an affinity graph. The LPP proposed by He is a linearization of LE [11]. Furthermore, since the orthogonality constraint on projection directions is demonstrated to be more effective in preserving intrinsic geometrical structure, orthogonal LPP (OLPP) [23] and orthogonal NPE (ONPE) [24] are proposed for face recognition.

The aforementioned methods are almost related to vector-based methods, which require to transform image matrices into vectors as input. However, the vectorization destroys the local geometrical structure in the image resulting in spatial information lost [25]. To capture spatial information effectively, a lot of matrix-based methods have been proposed in recent works [26–28]. For example, to avoid local information losing caused by vectoring, Lee and Choi proposed two-dimensional canonical correlation analysis (2DCCA) [29]. Similar to 2DCCA, two-dimensional PCA (2DPCA) was proposed for face recognition in [30]. Hu et al. [31] proposed two-dimensional LPP (2DLPP) which can extract features from image matrices directly and approximate the original images more accurately than LPP.

Actually, most of previously mentioned methods can be referred to the generalized framework termed graph embedding (GE) in [32]. GE attempts to map the high-dimensional data points into a low-dimensional space with the intrinsic geometrical structure of data preserved by an affinity graph. However, these methods are based on Euclidean distance which is sensitive to noise. To overcome this disadvantage, some robust metrics [33–35] are introduced to replace the L_2 norm and Frobenius norm (F-norm). Under the GE framework, the simplest and frequently used technique is to construct a more accurate affinity graph with robust metrics to reduce the impact of noise. Sparsity preserving projection (SPP) [36] and two-dimensional SPP (2DSPP) [37] were proposed with the same motivation of constructing the affinity graph by using the sparse reconstruction of data under an L_1 norm constraint. By considering the low rank property of data, the work [38] introduced a nuclear norm-based method into the affinity graph construction and proposed low rank preserving embedding (LRPE). Besides the ability of describing the low rank structure, nuclear norm has attracted much attention also because its robustness on noise and occlusions has been demonstrated by robust PCA (RPCA) [39] and nuclear norm-based matrix regression (NMR) [40].

In this paper, inspired by the motivation of constructing a more accurate affinity graph from the literature [41], to handle the DR of corrupted 2D images, we propose a novel projection-based DR method, termed nuclear norm-based matrix regression preserving embedding (NN-MRPE), which reduces the impact of corruptions by nuclear norm-based reconstruction. In NN-MRPE, the learned robust projections transform the corrupted image matrices into a low dimensional space with more accurate reconstruction relationships between images preserved. To demonstrate the performance of the proposed NN-MRPE algorithm, extensive experiments on five public face databases are conducted in the paper. Experimental results validate the efficiency of the proposed method. Several characteristics of the paper are enumerated as follows.

- A novel nuclear norm-based matrix regression graph is built to eliminate effect of corruptions for the matrix graph embedding framework. The novel affinity graph captures more accurate reconstruction relationships of the corrupted images which can enhance the robustness of learned projections.
- A matrix-based projection learning formulation is constructed to seek two transformation matrices that can ensure the geometrical structures of data effectively preserved in a low dimensional space. Moreover, an alternating iteration algorithm is proposed to solve the objective function and its convergence is mathematically proven.

- A generalized framework, called linear regression projection embedding, is developed for DR, which is a regularized extension of LLE.

2. Related work

This section briefly reviews some projection-based DR methods, e.g., collaborative representation based projection method (CRP) and sparsity preserving projection (SPP).

2.1. Collaborative representation based projection method (CRP)

Collaborative representation based projection method (CRP) is based on an L_2 graph [12]. Similar to NPE, CRP consists of two main parts: weighted graph construction and projection learning. Let X be a set of data points $x_1, x_2, \dots, x_n \in R^m$, the CRP algorithm can be formally stated as below:

Part 1. Weighted graph construction: The first step is to construct an L_2 graph based on collaborative representation. Let w_i be the reconstruction weight for each x_i , which can be obtained by minimizing the following objective function

$$w_i = \arg \min \|x_i - Xw_i\|^2 + \lambda \|w_i\|_2^2, \quad (1)$$

where $w_i = [w_{i,1}, \dots, w_{i,i-1}, 0, w_{i,i+1}, \dots, w_{i,n}]^T$, $w_{ij} (i \neq j)$ represents the contribution of x_j to represent x_i .

After solving w_i for x_i , the L_2 graph can be defined as $G(X, W)$ with the sample set X as the vertices and $W = [w_1, w_2, \dots, w_n]$ as the weight matrix.

Part 2. Projection learning: CRP aims at seeking projections that not only preserve the reconstruction relationship but also maximize the total separability information. Therefore, the optimal map U can be solved from the following objective function

$$J(U) = \arg \min \frac{U^T S_L U}{U^T S_T U}, \quad (2)$$

where $S_L = X(I - W - W^T + WW^T)X^T$ and $S_T = \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$. The optimal solution to problem (2) is actually equivalent to the eigenvectors associated with the d_k smallest nonzero eigenvalues of the generalized eigenvalue problem $S_L \mathbf{u} = \lambda S_T \mathbf{u}$.

2.2. Sparsity preserving projection (SPP)

Sparsity preserving projection (SPP) is a method belonging to the generalized graph embedding framework, whose adjacency graph construction is based on a modified sparse representation (MSR) [36]. Given a set of training samples $X = [x_1, x_2, \dots, x_n] \in R^{m \times n}$, where $x_i \in R^m$ denotes the i th training sample. Similar to sparse representation, the modified sparse representation also aims to reconstruct each sample x_i with samples as few as possible. The sparse reconstruction weight vector s_i for each x_i is the solution to the following optimization problem:

$$\min_{s_i} \|s_i\|_0 \quad s.t. \quad x_i = Xs_i, \mathbf{1} = \mathbf{1}^T s_i, \quad (3)$$

where $s_i = [s_{i,1}, s_{i,2}, \dots, s_{i,i-1}, 0, s_{i,i+1}, \dots, s_{i,n}] \in R^n$, s_{ij} represents the contribution of each x_j for reconstructing x_i , and $\mathbf{1} \in R^n$ denotes a vector of all ones.

The sparse reconstruction weight matrix $S = [s_1, s_2, \dots, s_n]$ is obtained by solving the optimal solution s_i of MSR problem. Subsequently, the objective function of SPP is defined to seek the projections that preserve the reconstruction weight matrix S , which can be formulated as

$$\min_U \sum_{i=1}^n \|U^T x_i - U^T X s_i\|^2. \quad (4)$$

To avoid degenerate solutions, constraint $U^T X X^T U = I$ is imposed on Eq. (4). Thus, with simple algebraic transformation, the objective function (4) can be reformulated as:

$$\min \frac{U^T X (I - S - S^T + S^T S) X^T U}{U^T X X^T U} \propto \max \frac{U^T X (S + S^T - S^T S) X^T U}{U^T X X^T U}. \quad (5)$$

The optimal projection matrix U consists of the eigenvectors associated with the d_k largest eigenvalues of the following generalized eigenvalue problem

$$X(S + S^T - S^T S) X^T U = \lambda X X^T U. \quad (6)$$

3. Methodology

In this section, we first introduce the nuclear norm-based matrix regression model and its optimization algorithm. Then, the nuclear norm-based matrix regression preserving embedding (NN-MRPE) method is formulated.

3.1. Nuclear norm-based matrix regression

Denote $\mathbf{X} = \{X_i \in R^{m_1 \times m_2}\}_{i=1}^n$ as a set of image matrices, where X_i is the i th image. For a given image $Y \in R^{m_1 \times m_2}$, a linear combination of matrices in \mathbf{X} can represent Y approximately. Then, the linear matrix regression model can be defined as

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n + E, \quad (7)$$

where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]$ is a vector of coding coefficients and E is the residual. For convenience, we denote $\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n$ as (α) , then Eq. (7) can be rewritten as

$$Y = (\alpha) + E. \quad (8)$$

In many applications, the optimal residual image $(\alpha) - Y$ is typically low rank (or approximately low rank). Motivated by this fact, we would like to estimate the regression coefficients via solving the following nuclear norm approximation problem

$$\min_{\alpha} \|(\alpha) - Y\|_{*}. \quad (9)$$

Moreover, adding an L_2 regularization term [42] to Eq. (9) obtains the regularized nuclear norm-based matrix regression (NMR) [40] model as

$$\min_{\alpha} \|(\alpha) - Y\|_{*} + \frac{1}{2} \lambda \|\alpha\|_2^2. \quad (10)$$

In order to solve the above model, we first have to rewrite it as

$$\min_{\alpha} \|E\|_{*} + \frac{1}{2} \lambda \|\alpha\|_2^2 \quad \text{s.t.} \quad (\alpha) - Y = E. \quad (11)$$

Thus, the corresponding augmented Lagrangian function is formulated as

$$\begin{aligned} L_{\mu}(E, \alpha, W) = & \|E\|_{*} + \frac{1}{2} \lambda \|\alpha\|_2^2 + \text{tr}(W^T ((\alpha) - Y - E)) \\ & + \frac{\mu}{2} \|(\alpha) - Y - E\|_F^2. \end{aligned} \quad (12)$$

Then, the regularized matrix regression problem can be solved by the alternating direction method of multipliers (ADMM) algorithm summarized in Algorithm 1.

3.2. Motivation

The methods related to GE explore the manifold structure of data by constructing an affinity graph. According to [41], it can be found that the performance of these methods are significantly related to the quality of affinity graph. However, the vector-based

Algorithm 1: ADMM algorithm for NMR.

Input : A set of image matrices X_1, X_2, \dots, X_n and an image matrix $Y \in R^{m_1 \times m_2}$, the parameters λ and μ ;

1 Initialize $E^k = -Y$, $W^k = 0$, $k = 0$;

2 **repeat**

3 Update α :

4 $\alpha^{k+1} = \arg \min_{\alpha} \left(\frac{\lambda}{2} \|\alpha\|_2^2 + \frac{\mu}{2} \|(\alpha) - Y - E^k + \frac{1}{\mu} W^k\|_F^2 \right)$;

5 Update E :

6 $E^{k+1} = \arg \min_E \left(\frac{1}{\mu} \|E\|_{*} + \frac{1}{2} \|E - ((\alpha^{k+1}) - Y + \frac{1}{\mu} W^k)\|_F^2 \right)$;

7 Update W :

8 $W^{k+1} = W^k + \mu((\alpha^{k+1}) - E^{k+1} - Y)$;

9 Update $k = k + 1$;

10 **until** convergence;

Output: Reconstruction coefficient vector α^{k+1} .

methods (such as LPP, CRP, and SPP) require to transform 2D images into 1D vectors, which causes the loss of local structure information in images. Although 2DLPP overcomes this shortage, the Euclidean distance based graph is sensitive to the noise and corruptions. However, corruptions have a profound influence on local structural information of the data. Recently, robust metric is proven to be able to deal with corruptions effectively. Thus, some robust metrics are introduced to affinity graph construction for GE framework. Under this motivation, 2DSPP is proposed with a L_1 norm graph. However, the projection learning function of 2DSPP only consider one direction (i.e., relationships between rows) of image matrices, which neglects the relationships between columns in images.

To address the problems mentioned before, we propose nuclear norm-based matrix regression preserving projection (NN-MRPE). First, in order to improve the robust to the noise and corruptions, NMR is used to construct an affinity graph. Second, to consider horizontal and vertical directions of image matrices simultaneously, a matrix-based cost function is constructed to learn two robust transformation matrices.

3.3. Nuclear norm-based matrix regression preserving embedding(NN-MRPE)

Since projection-based DR under the GE framework is mainly characterized by a specific affinity weight matrix of data, a matrix regression framework is considered to construct the affinity matrix. Under the assumption that each image matrix X_i can be represented as a linear combination of other image matrices [43], the nuclear norm-based matrix regression model is employed to reconstruct X_i , i.e.,

$$\min_{\alpha_i} \|_i(\alpha_i) - X_i\|_{*} + \frac{1}{2} \lambda \|\alpha_i\|_2^2, \quad (13)$$

where $_i(\cdot)$ is a linear combination of all matrices belonging to X except X_i , i.e., $_i(\alpha_i) = \alpha_{i,1} X_1 + \cdots + \alpha_{i,i-1} X_{i-1} + \alpha_{i,i+1} X_{i+1} + \cdots + \alpha_{i,n} X_n$, and λ is a balance parameter. The Algorithm 1 is utilized to solve the optimization problem (13). Then, the element w_{ij} of affinity weight matrix W can be constructed as

$$w_{ij} = \begin{cases} 0, & i = j; \\ \alpha_{i,j}, & i > j; \\ \alpha_{i-1,j}, & i < j. \end{cases} \quad (14)$$

Each w_{ij} in W reflects the relationship between X_i and X_j . The larger value of w_{ij} implies that there is a larger probability for X_i and X_j belonging to the same class.

After the affinity matrix W is formed, we thereby expect that the desirable characteristics in the original high-dimensional

space can be preserved in a low-dimensional embedding subspace. Therefore, the following objective function is defined to find the projections that can preserve the optimal weight matrix W effectively.

$$\min_{U,V} \sum_{i=1}^n \|U^T X_i V - \sum_{j=1}^n U^T X_j V w_{ij}\|^2 \quad (15)$$

Let $Y_i = U^T X_i V$ and $D = [d_{ij}]_{n \times n}$ be an identity matrix. Since $\|A\|^2 = \text{tr}(AA^T)$, we see that:

$$\begin{aligned} & \sum_{i=1}^n \|U^T X_i V - \sum_{j=1}^n U^T X_j V w_{ij}\|^2 = \sum_{i=1}^n \|Y_i - \sum_{j=1}^n Y_j w_{ij}\|^2 \\ &= \sum_{i=1}^n \text{tr} \left(\left(Y_i - \sum_{j=1}^n Y_j w_{ij} \right) \left(Y_i - \sum_{j=1}^n Y_j w_{ij} \right)^T \right) \\ &= \sum_{i=1}^n \text{tr} \left(\left(\sum_{j=1}^n Y_j (d_{ij} - w_{ij}) \right) \left(\sum_{j=1}^n Y_j (d_{ij} - w_{ij}) \right)^T \right) \\ &= \sum_{i=1}^n \text{tr} \left(\left(\sum_{j=1}^n U^T X_j V (d_{ij} - w_{ij}) \right) \left(\sum_{j=1}^n U^T X_j V (d_{ij} - w_{ij}) \right)^T \right) \\ &= \sum_{i=1}^n \text{tr} \left(U^T \left(\sum_{j=1}^n X_j (d_{ij} - w_{ij}) \right) V V^T \left(\sum_{j=1}^n X_j^T (d_{ij} - w_{ij}) \right) U \right) \\ &= \sum_{i=1}^n \text{tr} (U^T (\mathbf{d}_i - \mathbf{w}_i) V V^T (\mathbf{d}_i - \mathbf{w}_i)^T U) \\ &= \text{tr} \left(U^T \left(\sum_{i=1}^n (\mathbf{d}_i - \mathbf{w}_i) V V^T (\mathbf{d}_i - \mathbf{w}_i)^T \right) U \right) \\ &= \text{tr}(U^T S_V U), \end{aligned} \quad (16)$$

where $S_V = \sum_{i=1}^n (\mathbf{d}_i - \mathbf{w}_i) V V^T (\mathbf{d}_i - \mathbf{w}_i)^T$, $(\mathbf{d}_i - \mathbf{w}_i)$ represents a linear combination of X_j ($j = 1, 2, \dots, n$) with the coefficient vector $\mathbf{d}_i - \mathbf{w}_i$, and $\mathbf{d}_i - \mathbf{w}_i$ represents a linear combination of X_j^T ($j = 1, 2, \dots, n$) with the coefficient vector $\mathbf{d}_i - \mathbf{w}_i$. \mathbf{d}_i and \mathbf{w}_i denote the i th column of D and W , respectively.

Similarly, since $\|A\|^2 = \|A^T\|^2$ and $\|A^T\|^2 = \text{tr}(A^T A)$, we also have

$$\begin{aligned} & \sum_{i=1}^n \|U^T X_i V - \sum_{j=1}^n U^T X_j V w_{ij}\|^2 = \sum_{i=1}^n \|Y_i^T - \sum_{j=1}^n Y_j^T w_{ij}\|^2 \\ &= \sum_{i=1}^n \text{tr} \left(\left(Y_i^T - \sum_{j=1}^n Y_j^T w_{ij} \right) \left(Y_i^T - \sum_{j=1}^n Y_j^T w_{ij} \right)^T \right) \\ &= \sum_{i=1}^n \text{tr} \left(\left(\sum_{j=1}^n Y_j^T (d_{ij} - w_{ij}) \right) \left(\sum_{j=1}^n Y_j^T (d_{ij} - w_{ij}) \right)^T \right) \\ &= \sum_{i=1}^n \text{tr} \left(\left(\sum_{j=1}^n V^T X_j^T U (d_{ij} - w_{ij}) \right) \left(\sum_{j=1}^n V^T X_j^T U (d_{ij} - w_{ij}) \right)^T \right) \\ &= \sum_{i=1}^n \text{tr} \left(V^T \left(\sum_{j=1}^n X_j^T (d_{ij} - w_{ij}) \right) U U^T \left(\sum_{j=1}^n X_j (d_{ij} - w_{ij}) \right) V \right) \\ &= \sum_{i=1}^n \text{tr} (V^T (\mathbf{d}_i - \mathbf{w}_i)^T U U^T (\mathbf{d}_i - \mathbf{w}_i) V) \end{aligned}$$

$$\begin{aligned} &= \text{tr} \left(V^T \left(\sum_{i=1}^n {}^T(\mathbf{d}_i - \mathbf{w}_i) U U^T (\mathbf{d}_i - \mathbf{w}_i) \right) V \right) \\ &= \text{tr}(V^T S_U V), \end{aligned} \quad (17)$$

where $S_U = \sum_{i=1}^n {}^T(\mathbf{d}_i - \mathbf{w}_i) U U^T (\mathbf{d}_i - \mathbf{w}_i)$ and the other notations have the same meaning as before. Therefore, our task is to minimize $\text{tr}(U^T S_V U)$ and $\text{tr}(V^T S_U V)$ simultaneously.

In addition, in order to avoid degenerate solutions, the constraints [44]

$$\text{tr} \left(U^T \left(\sum_{i=1}^n X_i V V^T X_i^T \right) U \right) = 1 \text{ and } \text{tr} \left(V^T \left(\sum_{i=1}^n X_i^T U U^T X_i \right) V \right) = 1 \quad (18)$$

are imposed on U and V , respectively. Thus, the optimization problems are formulated as

$$\min_U \frac{\text{tr}(U^T S_V U)}{\text{tr}(U^T (\sum_{i=1}^n X_i V V^T X_i^T) U)}, \quad (19)$$

$$\min_V \frac{\text{tr}(V^T S_U V)}{\text{tr}(V^T (\sum_{i=1}^n X_i^T U U^T X_i) V)}. \quad (20)$$

It is easy to see that their optimal solutions are equivalent to the generalized eigenvectors of the following two generalized eigenvalue equations, respectively.

$$S_V \mathbf{u} = \beta D_V \mathbf{u}, \quad \text{where } D_V = \sum_{i=1}^n X_i V V^T X_i^T; \quad (21)$$

$$S_U \mathbf{v} = \beta D_U \mathbf{v}, \quad \text{where } D_U = \sum_{i=1}^n X_i^T U U^T X_i. \quad (22)$$

However, U and V can not be solved independently since they depend on each other. Thus, an iterative method is used to estimate the optimal U and V . Then, the complete NN-MRPE algorithm can be expressed as the Algorithm 2.

Algorithm 2: NN-MRPE algorithm.

Input : Training image matrices $X_1, X_2, \dots, X_n \in R^{m_1 \times m_2}$, the parameters λ and μ ;

- 1 Initialize transformation matrix V with an unit matrix;
- 2 **for** each $i = 1, 2, \dots, n$ **do**
- 3 Solve α_i from Eq. (13) by Algorithm 1;
- 4 **end**
- 5 Compute the affinity matrix W by (14);
- 6 **repeat**
- 7 Update S_V and D_V respectively by
- 8 $S_V = \sum_{i=1}^n (\mathbf{d}_i - \mathbf{w}_i) V V^T (\mathbf{d}_i - \mathbf{w}_i)^T$ and $D_V = \sum_{i=1}^n X_i V V^T X_i^T$;
- 9 Update S_U and D_U respectively by
- 10 $S_U = \sum_{i=1}^n {}^T(\mathbf{d}_i - \mathbf{w}_i) U U^T (\mathbf{d}_i - \mathbf{w}_i)$ and $D_U = \sum_{i=1}^n X_i^T U U^T X_i$;
- 11 Update V by (22)
- 12 $S_V \mathbf{u} = \beta D_V \mathbf{u}$;
- 13 Update S_U and D_U respectively by
- 14 $S_U = \sum_{i=1}^n {}^T(\mathbf{d}_i - \mathbf{w}_i) U U^T (\mathbf{d}_i - \mathbf{w}_i)$ and $D_U = \sum_{i=1}^n X_i^T U U^T X_i$;
- 15 Update U by (21)
- 16 $S_U \mathbf{v} = \beta D_U \mathbf{v}$;
- 17 Update $t = t + 1$;
- 18 **until** convergence;

Output: Transformation matrices U and V .

4. Algorithm analysis

First, the convergence and computational complexity of the proposed algorithm are analyzed. Then, a generalized framework for some existing methods is given in this section.

4.1. Convergence analysis

The convergence of ADMM in [Algorithm 1](#) has been mathematically proven in [\[40\]](#). Thus, this subsection mainly presents the convergence analysis of the alternating iterative method in [Algorithm 2](#).

Denote

$$J(U, V) = \sum_{i=1}^n \|U^T X_i V - \sum_{j=1}^n U^T X_j V W_{ij}\|^2, \quad (23)$$

we can obtain the following Theorem.

Theorem 1. The iterative scheme in [Algorithm 2](#) monotonically decreases the objective function value of $J(U, V)$ in each iteration.

Proof. Suppose in the t th iteration, we have the following result

$$J(U_t, V_t) = \sum_{i=1}^n \|U_t^T X_i V_t - \sum_{j=1}^n U_t^T X_j V_t W_{ij}\|^2, \quad (24)$$

First, for given V_t , U_{t+1} is obtained by solving an eigen-decomposition problem, which further reduces the objective function value. Thus, we get

$$J(U_t, V_t) \geq J(U_{t+1}, V_t). \quad (25)$$

Similarly, when fix U_{t+1} , the following inequality holds, i.e.,

$$J(U_{t+1}, V_t) \geq J(U_{t+1}, V_{t+1}). \quad (26)$$

Then, it can obtain

$$J(U_t, V_t) \geq J(U_{t+1}, V_t) \geq J(U_{t+1}, V_{t+1}). \quad (27)$$

Therefore, we know that the objective function value $J(U, V)$ is monotonically decreasing in each iteration.

Since the objective function (23) has positive lower bound, from [Theorem 1](#) we can know that [Algorithm 2](#) converges.

4.2. Computational complexity

Given n training samples belonging to $R^{m_1 \times m_2}$ and $m_1 \geq m_2$, the computational complexity of [Algorithm 1](#) is $O(k(m_1 m_2^2 + m_1 m_2 n))$ according to [\[40\]](#), where k is the number of iterations. For [Algorithm 2](#), the major computation is in steps 3, 8, and 10. Step 3 has the same computational complexity as in [Algorithm 1](#). Since Step 3 is repeated n times with $n-1$ samples in [Algorithm 2](#), the complexity of this part is $O(k(m_1 m_2^2 n + m_1 m_2 n^2))$. Steps 8 and 10 require solving two generalized eigenvalue problems with different sizes. The complexity of Step 8 is $O(m_1^3)$ and the complexity of Step 10 is $O(m_2^3)$. Taking the number of iterations t into consideration, the complexity of NN-MRPE algorithm is $O(k(m_1 m_2^2 n + m_1 m_2 n^2) + t(m_1^3 + m_2^3))$.

4.3. A generalized framework: linear regression preserving embedding

By analyzing the commons of NPE, CRP, SPP, 2DSPP, and NN-MRPE, we develop a generalized DR framework termed linear regression preserving embedding which preserves the local/global linear reconstruction relationships of data in a low dimensional space. Given a training set X_1, X_2, \dots, X_n , where each X_i is a vector/matrix sample, we expect to recast each X_i by other data points through the linear regression model, i.e.,

$$X_i = \alpha_{i,1} X_1 + \dots + \alpha_{i,i-1} X_{i-1} + 0 \cdot X_i + \alpha_{i,i+1} X_{i+1} + \dots + \alpha_{i,n} X_n + E,$$

Table 1

The special setting of several algorithms related to linear regression preserving embedding. Atoms: the samples in training set used for each X_i reconstruction, where All means all samples except X_i itself; Dimension: the dimension of sample X_i .

Methods	Atoms	Dimension	$dist(\cdot, \cdot)$	λ	p
LLE	k neighborhoods	1D	$\ \cdot\ _2^2$	$\lambda = 0$	null
NPE	k neighborhoods	1D	$\ \cdot\ _2^2$	$\lambda = 0$	null
CRP	All	1D	$\ \cdot\ _2^2$	$\lambda > 0$	2
SPP	All	1D	$\ \cdot\ _2^2$	$\lambda > 0$	1
2DSPP	All	2D	$\ \cdot\ _F^2$	$\lambda > 0$	1
NN-MRPE	All	2D	$\ \cdot\ $	$\lambda > 0$	2

$$(28)$$

where $\alpha_i = [\alpha_{i,1}, \dots, \alpha_{i,i-1}, 0, \alpha_{i,i+1}, \dots, \alpha_{i,n}]^T$ is the reconstruction coefficient vector and E is the reconstruction residual. For convenience, denote $(\alpha_i) = \alpha_{i,1} X_1 + \dots + \alpha_{i,i-1} X_{i-1} + 0 \cdot X_i + \alpha_{i,i+1} X_{i+1} + \dots + \alpha_{i,n} X_n$. Then model (28) can be rewritten as

$$X_i = (\alpha_i) + E. \quad (29)$$

To find the optimal reconstruction coefficient vector α_i for X_i , we formulate the following minimization problem

$$\min_{\alpha_i} dist(X_i, (\alpha_i)) + \lambda \|\alpha_i\|_p^p \text{ subject to } \alpha_{ii} = 0, \quad (30)$$

where $dist(\cdot, \cdot)$ denotes a distance metric which can be chosen from F -norm, L_2 norm, and nuclear norm, etc. λ is a balance parameter and p can be set as 1 or 2. Different metric and parameter assignments carry different meanings for the coefficient vectors.

After solving each weight vector α_i from model (30), we can define the linear reconstruction weight matrix $W^\alpha = [\alpha_{ij}]_{n \times n}$ as

$$W^\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]. \quad (31)$$

Suppose there exists a linear mapping that can keep the intrinsic geometric property of data reflected by W^α exactly invariant. Based on this assumption, each high dimensional data point X_i is projected to a low dimensional Y_i in the last step of linear regression preserving embedding. Each Y_i can be solved by minimizing the following embedding cost function

$$J(Y) = \sum_i \|Y_i - \sum_j \alpha_{ij} Y_j\|^2. \quad (32)$$

The above cost function can be solved in the similar way as LLE. Actually, the linear regression preserving embedding method is an extension of LLE.

In fact, it is easy to find that NPE, CRP, SPP, 2DSPP, and NN-MRPE belonging to linear regression preserving embedding. NPE is a linear version of LLE which preserves the local neighborhood structure by recovering the local neighborhood relationship in the original space. The weight matrix of NPE is calculated under the least squares reconstruction error by using the neighborhoods sought by the nearest neighbor method when $\lambda = 0$. Furthermore, CRP and SPP are regularized extensions of NPE. The difference is that CRP constructs the weighted graph by L_2 -based Ridge regression (RR), whereas SPP builds the weighted graph by L_1 -based sparse representation. 2DSPP is a naive extension of SPP which actually uses the same weight graph with SPP. The proposed NN-MRPE not only extends the vector-based LR graph to linear matrix regression graph but also alleviates the effect of corruptions by using nuclear norm. The detailed comparison between these methods is shown in [Table 1](#).

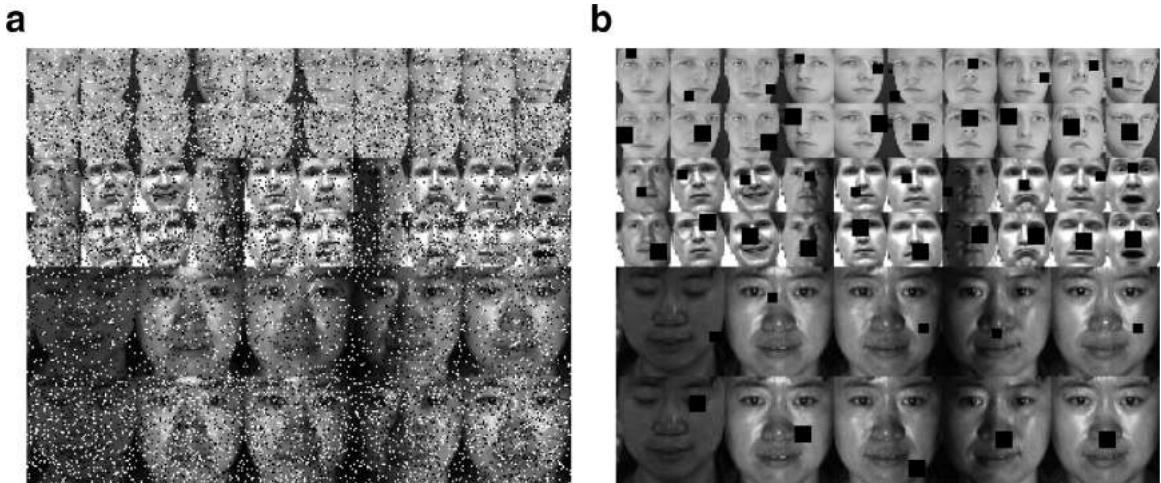


Fig. 1. Image examples from ORL (The first two rows), Yale (The middle two rows), and PIE (The last two rows) databases. (a) Example images with different densities of “salt & pepper” noise, (b) example images with different sizes of block occlusions.

5. Experiments

5.1. Database description

ORL [45], Yale [46], CMU PIE [47], AR [48], and LFW [49] are five face databases used in the experiments, which are simply described as follows.

The ORL database composes of 400 face images from 40 individuals, each object provides 10 different images. The face images were captured under varying facial expressions and facial details, such as, smile or no smile, glasses or no glasses, and so on. The size of each face image is normalized to 32×32 . Yale database contains 165 face images of 15 subjects. For each subject, 11 images are captured under various facial expression and lighting conditions. Each image is cropped and resized to 32×32 pixels. CMU PIE face database contains 68 individuals with 41,368 images. The face images were taken under various pose, illumination, and expression. In our experiments, the subset C29 which contains 1632 images of 68 individuals is selected to evaluate the proposed algorithm. All of these face images are cropped to have 64×64 pixels. The AR database consists of over 4000 face images of 126 people (70 men and 56 women). For each person, 26 images were taken in two sessions (separated by two weeks). Each section contains 13 images under different facial expression, lighting conditions, and occlusions. In our experiments, a subset of the AR database including 50 men and 50 women is used to test the performance of the proposed algorithm. Each face image is cropped and resized to 55×40 . LFW is a large-scale database of face photographs designed for unconstrained face recognition with extreme variations in pose, illumination, expression, occlusion and so on. In the experiments, a subset containing 1840 images from 120 persons is used. Each image in LFW is resized to 60×60 .

5.2. Experiment setup

To validate the robustness of the NN-MRPE algorithm, we test the algorithm by the images in ORL, Yale, and PIE databases with three different kinds of random corruptions: random pixel corruptions, contiguous occlusions of various levels, and complex corruptions. The pixel corruptions are “salt & pepper” noise with two different probabilities 0.1 and 0.2. Some image examples from the three databases with “salt & pepper” noise are shown in Fig. 1 (a). Block occlusions corrupted images are employed to test the robust-



Fig. 2. Example images simultaneously contain “salt & pepper” noise and blocks from ORL (The 1st row), Yale (The 2nd row), PIE (The 3rd row) databases.

ness of the algorithm with various levels of contiguous occlusions. The sizes of blocks added in images have two different sizes: 6×6 and 10×10 as shown Fig. 1 (b). The complex corruptions added in images compose of “salt & pepper” noise (with 0.1 probabilities) and blocks (with 6×6). Fig. 2 shows some example images that contain complex corruptions. The AR and LFW databases is used to test the robustness on realistic occlusions of the proposed method. Figs. 3 and 4 show some example images from AR and LFW, respectively.

For the ORL and PIE databases, 50% of the images of each individual are randomly selected as training samples and the rest for testing. For the Yale database, 6 images of each subject are chosen randomly to construct the training set and the remaining images as testing samples. For the AR database, fixed training samples and testing set are used in the experiments and the detail is shown in Fig. 3. For the LFW database, 8 images of each person are randomly selected as training samples and the rest as testing samples. We compare the proposed method with the algorithms of CRP, SPP, LRPE, 2DPCA, 2DLPP, and 2DSPP. The parameters λ and μ of NN-MRPE are selected from the ranges (0,1] and $[10^{-5}, 10^5]$, respectively. Fig. 5 shows the classification accuracy versus different values of λ and μ on the ORL, Yale, and PIE databases with the “salt & pepper” noise. The nearest neighbor classifier is employed to compute the classification accuracy of testing samples. The results reported are the average recognition accuracy of 20 times independent trials. In addition, in order to study the relationship between recognition accuracy and dimensions, the number of smallest nonzero eigenvectors is set as $\min(n_r, 15)$ in our experiments, where n_r denote the actual number of nonzero eigenvalues.



Fig. 3. Example images of training and test in AR database. Image examples in the first row are from training set and the second row are example images from testing samples.



Fig. 4. Example images from the LFW database.

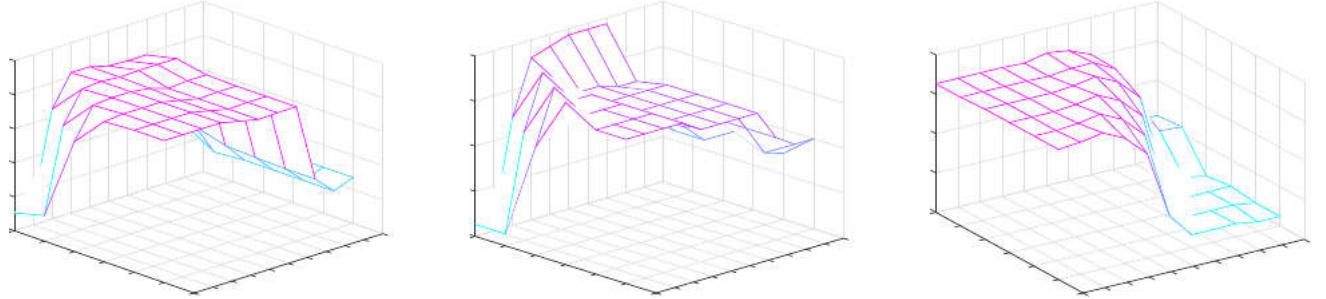


Fig. 5. The classification accuracy versus different values of λ and μ in NN-MRPE for ORL, Yale, and PIE databases with the “salt & pepper” noise. Experiments on (a) ORL database (b) Yale database (c) PIE database.

5.3. Experimental results and analysis

In order to test the robustness of NN-MRPE on against random pixel corruptions, we conduct several experiments on images as shown in Fig. 1 (a). Fig. 6 presents the recognition accuracies versus different dimensions of ORL, Yale, and PIE databases with two different densities of the “salt & pepper” noise. As can be seen from Fig. 6, NN-MRPE performs better than other compared methods in low dimensional space with different dimensions.

The images shown in Fig. 1 (b) are used to validate the robustness of NN-MRPE against various level contiguous occlusions. Fig. 7 shows the recognition accuracies of CRP, SPP, LRPE, 2DPCA, 2DLPP, 2DSPP, and NN-MRPE with different dimensions on various level contiguous occlusions. From Fig. 7, it can be seen that NN-MRPE is superior to other compared methods in recognition performance for each database with different sizes of blocks. These results illustrate that NN-MRPE has stronger robustness than other methods.

To test the robustness of the proposed method on complex corruptions, experiments on images with salt & pepper noise and blocks are conducted and experimental results are shown in Fig. 8. In Fig. 8, the recognition accuracy obtained by NN-MRPE outperforms that of other compared methods in different dimensions for ORL database. For Yale and PIE databases, it is obvious to find that NN-MRPE also achieves the highest recognition rate.

The AR and LFW databases are used to test the robustness of the NN-MRPE algorithm on realistic occlusions. Fig. 9 (a) shows the recognition accuracy versus different dimensions of AR database obtained by CRP, SPP, LRPE, 2DPCA, 2DLPP, 2DSPP, and NN-MRPE algorithms. Table 2 lists the highest correct recognition rate of AR database achieved by the aforementioned methods respectively. The superiority of NN-MRPE is obviously presented in Fig. 9 (a) and Table 2. The experimental results of LFW yielded by all the compared methods are presented in Fig. 9 (b). From Fig. 9 (b), it can be seen that the performance of classification on LFW database is not well since the images in the LFW exhibit extreme pose, illumination, and background variations. However, the proposed NN-MRPE still outperforms other compared methods.

Although the 2DSPP and 2DLPP algorithms present competitive robustness on different corruptions for each database by using image matrices as input directly, NN-MRPE still yields higher recognition accuracy than 2DSPP and 2DLPP, which indicates that the constructed matrix-based embedding cost formulation can provide more robust transformation matrices for DR. It can be found that most of methods related to linear regression embedding (e.g., SPP, 2DSPP, NN-MRPE) outperform the other compared methods (e.g., 2DPCA). This illustrates that the methods related to linear regression embedding can map the intrinsic geometric structure and discriminant information into the subspace more effectively. In particular, NN-MRPE is more robust than CRP, SPP and 2DSPP, demonstrating that nuclear norm can improve reconstruction accuracy for subspace embedding. The comparison

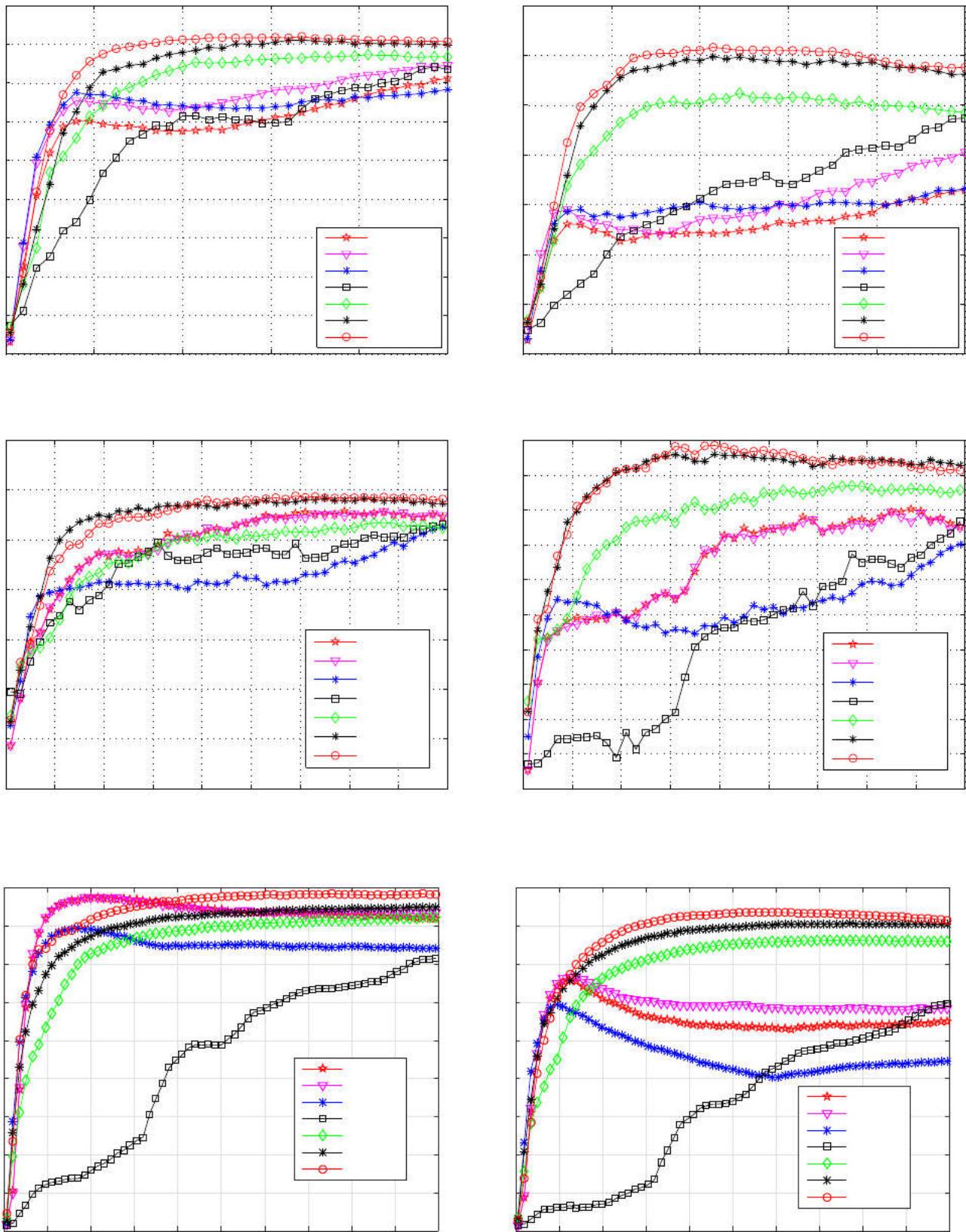


Fig. 6. The classification accuracy versus different dimensions of ORL, Yale, and PIE databases with the salt & pepper noise by using different DR methods. Experiments on (a) ORL database (noise density=0.1), (b) ORL database (noise density=0.2), (c) Yale database (noise density=0.1), (d) Yale database (noise density=0.2), (e) PIE database (noise density=0.1), (f) PIE database (noise density=0.2).

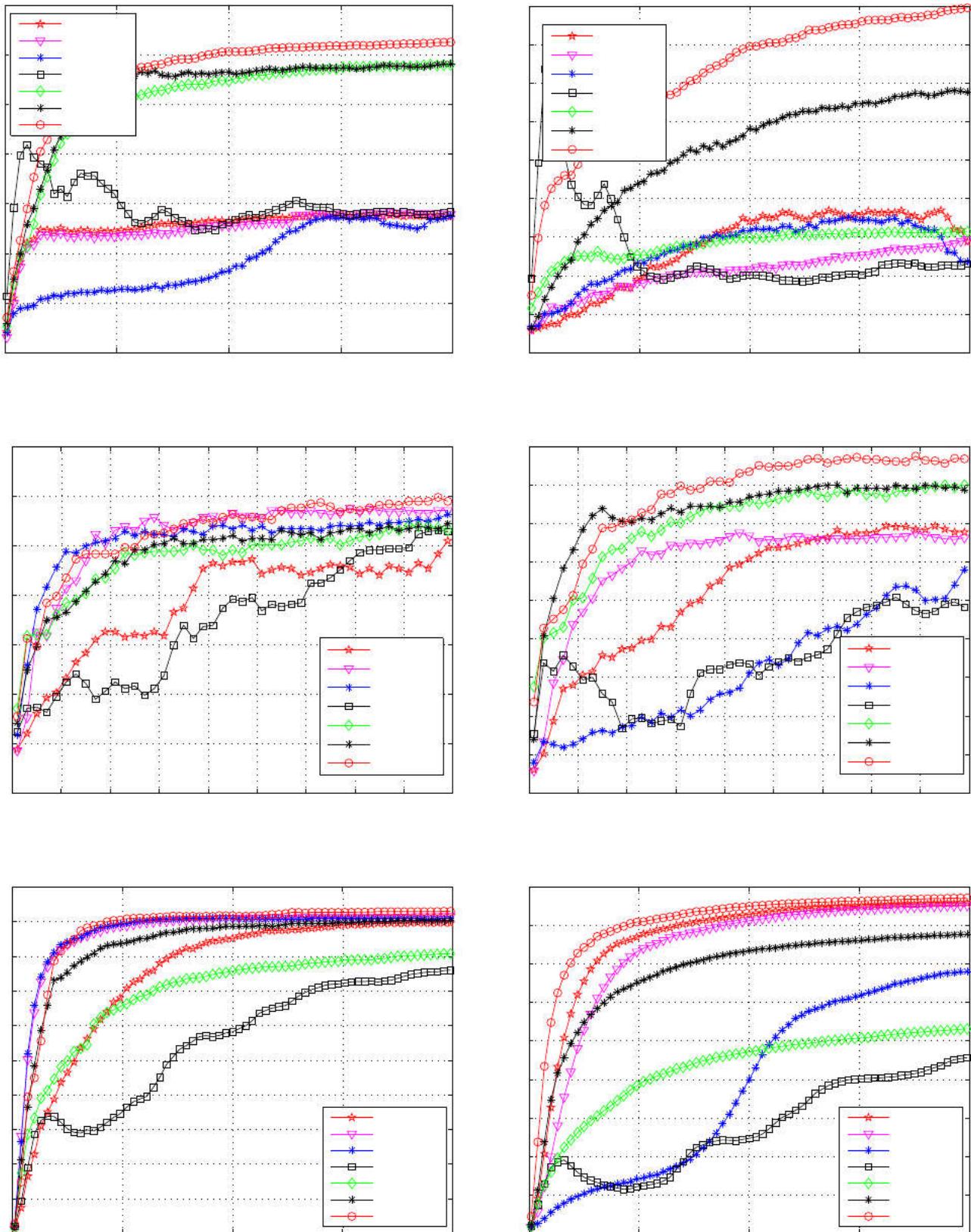


Fig. 7. The classification accuracy versus different dimensions of ORL, Yale, and CMU PIE databases with different block occlusions by using different DR methods. Experiments on (a) ORL database (occlusion size=6×6), (b) ORL database (occlusion size=10×10), (c) Yale database (occlusion size=6×6), (d) Yale database (occlusion size=10×10), (e) PIE database (occlusion size=6×6), (f) PIE database (occlusion size=10×10).

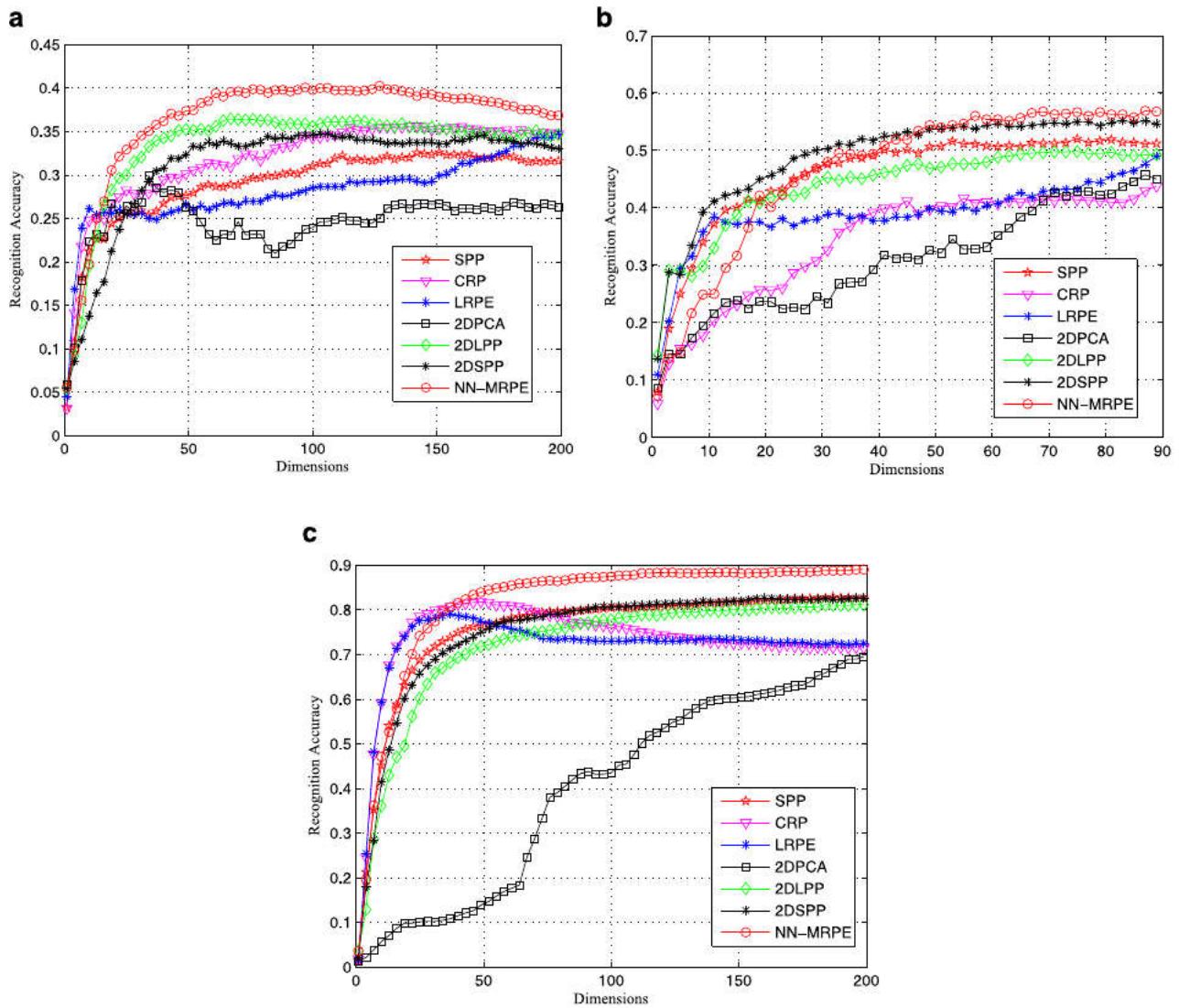


Fig. 8. The classification accuracy versus different dimensions of ORL, Yale, and PIE databases with the salt & pepper noise and blocks simultaneously by using different DR methods. Experiments on (a) ORL, (b) Yale, (c) PIE.

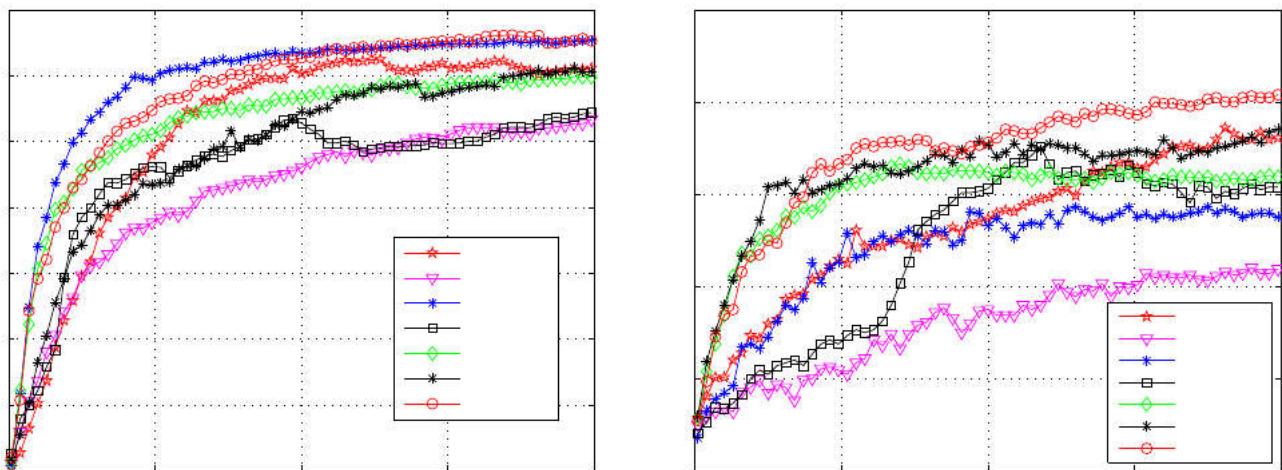


Fig. 9. The classification accuracy versus different dimensions of AR and LFW databases by using different DR methods. Experiments on (a) AR (b) LFW.

Table 2

The optimal recognition accuracy of AR database obtained by different algorithms.

Algorithms	CRP	SPP	LRPE	2DPCA	2DLPP	2DSPP	NN-MRPE
Accuracy	0.533	0.625	0.654	0.545	0.600	0.645	0.662

between 2DSPP and SPP showed that the 2D matrix-based methods generally provides better classification performance than the same type of methods based on vectors.

6. Conclusion

In this paper, to address the DR problem for images, we proposed a novel projection-based DR method (i.e., NN-MRPE), which learns two transformation matrices to map high dimensional data into a low dimensional space with the reconstruction relationship of the original data being preserved. In NN-MRPE, the robust metric known as nuclear norm was used to eliminate the effect caused by data corruptions for graph construction. In order to obtain robust projections to preserve the weight matrix exactly, a matrix-based embedding cost function was constructed for projection learning. We also mathematically proved the convergence and analyzed the main computational complexity of the proposed algorithm. By analysing the commons of some current works, a general linear regression preserving embedding framework was developed in the end. Experiments on five public face databases with different corruptions were conducted to test the robustness of the proposed algorithm and the experimental results demonstrated the superiority of NN-MRPE compared with other DR methods.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China [Grant 61371165] and by the Frontier Intersection Basic Research Project for the Central Universities [Grant A0920502051814-5].

References

- [1] E. Sariyanidi, H. Gunes, A. Cavallaro, Automatic analysis of facial affect: a survey of registration, representation, and recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 37 (6) (2015) 1113–1133, doi:[10.1109/TPAMI.2014.2366127](https://doi.org/10.1109/TPAMI.2014.2366127).
- [2] J. Zhang, J. Wang, X. Cai, Sparse locality preserving discriminative projections for face recognition, *Neurocomputing* 260 (2017) 321–330. [10.1016/j.neucom.2017.04.051](https://doi.org/10.1016/j.neucom.2017.04.051)
- [3] X. Ning, W. Li, B. Tang, H. He, BULDP: biomimetic uncorrelated locality discriminant projection for feature extraction in face recognition, *IEEE Trans. Image Process.* 27 (5) (2018) 2575–2586, doi:[10.1109/TIP.2018.2806229](https://doi.org/10.1109/TIP.2018.2806229).
- [4] X. Chen, H. Yin, F. Jiang, L. Wang, Multi-view dimensionality reduction based on universum learning, *Neurocomputing* 275 (2018) 2279–2286. [10.1016/j.neucom.2017.11.006](https://doi.org/10.1016/j.neucom.2017.11.006)
- [5] W.K. Wong, Z.H. Lai, J. Wen, X. Fang, Y. Lu, Low-rank embedding for robust image feature extraction, *IEEE Trans. Image Process.* 26 (6) (2017) 2905–2917, doi:[10.1109/TIP.2017.2691543](https://doi.org/10.1109/TIP.2017.2691543).
- [6] R.P.W. Duin, M. Loog, Linear dimensionality reduction via a heteroscedastic extension of LDA: the Chernoff criterion, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (6) (2004) 732–739, doi:[10.1109/TPAMI.2004.13](https://doi.org/10.1109/TPAMI.2004.13).
- [7] H.F. Li, T. Jiang, K.S. Zhang, Efficient and robust feature extraction by maximum margin criterion, *IEEE Trans. Neural Netw.* 17 (1) (2006) 157–165, doi:[10.1109/TNN.2005.860852](https://doi.org/10.1109/TNN.2005.860852).
- [8] Q. Wang and L. Ma and Q. Gao and Y. Li and Y. Huang and Y. Liu, Adaptive maximum margin analysis for image recognition, *Pattern Recognit.* 61 (2017) 339–347.
- [9] H. Huang and H. Feng and C. Peng, Complete local Fisher discriminant analysis with Laplacian score ranking for face recognition, *Neurocomputing* 89 (2012) 64–77. [10.1016/j.neucom.2012.02.020](https://doi.org/10.1016/j.neucom.2012.02.020)
- [10] A.M. Martinez, A.C. Kak, PCA versus LDA, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (2) (2001) 228–233, doi:[10.1109/34.908974](https://doi.org/10.1109/34.908974).
- [11] X.F. He, S.C. Yan, Y.X. Hu, H.-J. Zhang, Learning a locality preserving subspace for visual recognition, in: Proceedings of the Ninth International Conference on Computer Vision (ICCV), 2003, pp. 385–392 vol.1, doi:[10.1109/ICCV.2003.1238370](https://doi.org/10.1109/ICCV.2003.1238370).
- [12] W.K. Yang, Z.Y. Wang, C.Y. Sun, A collaborative representation based projections method for feature extraction, *Pattern Recognit.* 48 (1) (2015) 20–27.
- [13] D. Cai, X. He, J. Han, Semi-supervised discriminant analysis, in: Proceedings of the Eleventh International Conference on Computer Vision (ICCV), 2007, pp. 1–7, doi:[10.1109/ICCV.2007.4408856](https://doi.org/10.1109/ICCV.2007.4408856).
- [14] F. Zheng, N. Chen, L.Q. Li, Semi-supervised Laplacian eigenmaps for dimensionality reduction, in: Proceedings of the International Conference on Wavelet Analysis and Pattern Recognition, 2, 2008, pp. 843–849, doi:[10.1109/ICWAPR.2008.4635894](https://doi.org/10.1109/ICWAPR.2008.4635894).
- [15] J. Yang, D. Zhang, J.Y. Yang, Constructing PCA baseline algorithms to reevaluate ICA-based face-recognition performance, *IEEE Trans. Syst. Man Cybern. B Cybern.* 37 (4) (2007) 1015–1021, doi:[10.1109/TSMCB.2007.891541](https://doi.org/10.1109/TSMCB.2007.891541).
- [16] C. Zhang and F. P. Nie and S. Xiang, A general kernelization framework for learning algorithms based on kernel PCA, *Neurocomputing* 73 (4) (2010) 959–967. [10.1016/j.neucom.2009.08.014](https://doi.org/10.1016/j.neucom.2009.08.014)
- [17] C. Varon, C. Alzate, J.A.K. Suykens, Noise level estimation for model selection in kernel PCA denoising, *IEEE Trans. Neural Netw. Learn. Syst.* 26 (11) (2015) 2650–2663, doi:[10.1109/TNNLS.2015.2388696](https://doi.org/10.1109/TNNLS.2015.2388696).
- [18] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locality linear embedding, *Science* 290 (500) (2000) 2323–2326, doi:[10.1126/science.290.5500.2323](https://doi.org/10.1126/science.290.5500.2323).
- [19] A. Najafi, A. Joudaki, E. Fatemizadeh, Nonlinear dimensionality reduction via path-based isometric mapping, *IEEE Trans. Pattern Anal. Mach. Intell.* 38 (7) (2016) 1452–1464, doi:[10.1109/TPAMI.2015.2487981](https://doi.org/10.1109/TPAMI.2015.2487981).
- [20] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, *Neural Comput.* 15 (6) (2003) 1373–1396, doi:[10.1162/089976603321780317](https://doi.org/10.1162/089976603321780317).
- [21] B.Y. Paiement, J.F. Vincent, P. Delalleau, O. Roux, N. L., M. Ouimet, Out-of-sample extensions for LLE, Isomap, MDS, eigenmaps, and spectral clustering, in: *Proceedings of the Advances in Neural Information Processing Systems*, 16, 2004, pp. 177–184.
- [22] B.H. Wang, C. Lin, X.F. Zhao, Z.M. Lu, Neighbourhood sensitive preserving embedding for pattern classification, *IET Image Process.* 8 (8) (2014) 489–497, doi:[10.1049/iet-ipr.2013.0539](https://doi.org/10.1049/iet-ipr.2013.0539).
- [23] R. Wang and F. Nie and R. Hong and X. Chang and X. Yang and W. Yu, Fast and orthogonal locality preserving projections for dimensionality reduction, *IEEE Trans. Image Process.* 26 (10) (2017) 5019–5030, doi:[10.1109/TIP.2017.2726188](https://doi.org/10.1109/TIP.2017.2726188).
- [24] X. Liu, J. Yin, Z. Feng, J. Dong, L. Wang, Orthogonal neighborhood preserving embedding for face recognition, in: *Proceedings of the IEEE International Conference on Image Processing*, 1, 2007, doi:[10.1109/ICIP.2007.4378909](https://doi.org/10.1109/ICIP.2007.4378909). I – 133–I–136.
- [25] C. Lu and X. Liu and W. Liu, Face recognition based on two dimensional locality preserving projections in frequency domain, *Neurocomputing* 98 (2012) 135–142. [10.1016/j.neucom.2011.08.045](https://doi.org/10.1016/j.neucom.2011.08.045).
- [26] S.H. Lee, L.-L. Chen, J.W. Lin, T.C. Woo, Mass customization: 2-d projection of high dimensional data, *IEEE Trans. Syst. Man Cybern. A, Syst. Humans* 33 (1) (2003) 129–138, doi:[10.1109/TSMCA.2003.808250](https://doi.org/10.1109/TSMCA.2003.808250).
- [27] Q. L. Dong, Two-dimensional relaxed representation, *Neurocomputing* 121 (2013) 248–253. [10.1016/j.neucom.2013.04.044](https://doi.org/10.1016/j.neucom.2013.04.044).
- [28] M.A. Ochoa-Villegas, J.A. Nolazco-Flores, O. Barron-Cano, I.A. Kakadiaris, Addressing the illumination challenge in two-dimensional face recognition: a survey, *IET Comput. Vis.* 9 (6) (2015) 978–992, doi:[10.1049/iet-cvi.2014.0086](https://doi.org/10.1049/iet-cvi.2014.0086).
- [29] S.H. Lee, S. Choi, Two-dimensional canonical correlation analysis, *IEEE Signal Process. Lett.* 14 (10) (2007) 735–738, doi:[10.1109/LSP.2007.896438](https://doi.org/10.1109/LSP.2007.896438).
- [30] J. Yang, D. Zhang, A.F. Frangi, J.Y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (1) (2004) 131–137, doi:[10.1109/TPAMI.2004.1261097](https://doi.org/10.1109/TPAMI.2004.1261097).
- [31] D.W. Hu, G.Y. Feng, Z.T. Zhou, Two-dimensional locality preserving projections (2DLP) with its application to palmprint recognition, *Pattern Recognit.* 40 (1) (2007) 339–342. [10.1016/j.patcog.2006.06.022](https://doi.org/10.1016/j.patcog.2006.06.022)
- [32] S.C. Yan, D. Xu, B. Zhang, H. J. Zhang, Q. Yang, S. Lin, Graph embedding and extensions: a general framework for dimensionality reduction, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (1) (2007) 40–51, doi:[10.1109/TPAMI.2007.250598](https://doi.org/10.1109/TPAMI.2007.250598).
- [33] Q. Gao, L. Ma, Y. Liu, X. Gao, F. Nie, Angle 2DPCA: a new formulation for 2DPCA, *IEEE Trans. Cybern.* 48 (5) (2018) 1672–1678, doi:[10.1109/TCYB.2017.2712740](https://doi.org/10.1109/TCYB.2017.2712740).
- [34] Y.W. Lu, C. Yuan, Z.H. Lai, X.L. Li, W.K. Wong, D. Zhang, Nuclear norm-based 2DLP for image classification, *IEEE Trans. Multimed.* 19 (11) (2017) 2391–2403, doi:[10.1109/TMM.2017.2703130](https://doi.org/10.1109/TMM.2017.2703130).
- [35] Q. Gao, S. Xu, F. Chen, C. Ding, X. Gao, Y. Li, R1- 2DPCA and face recognition, *IEEE Trans. Cybern.* (2018) 1–12, doi:[10.1109/TCYB.2018.2796642](https://doi.org/10.1109/TCYB.2018.2796642).
- [36] L.S. Qiao, S.C. Chen, X.Y. Tan, Sparsity preserving projections with applications to face recognition, *Pattern Recognit.* 43 (1) (2010) 331–341. [10.1016/j.patcog.2009.05.005](https://doi.org/10.1016/j.patcog.2009.05.005)
- [37] X. Jing, Z. Sui, Y. Yao, J. Sun, Two-dimensional sparse preserving projection for

- face recognition, in: Proceedings of the International Conference on Automatic Control and Artificial Intelligence (ACAI), 2012, pp. 2223–2226, doi:[10.1049/cp.2012.1441](https://doi.org/10.1049/cp.2012.1441).
- [38] Y. Zhang, M. Xiang, B. Yang, Low-rank preserving embedding, *Pattern Recognit.* 70 (Supplement C) (2017) 112–125. [10.1016/j.patcog.2017.05.003](https://doi.org/10.1016/j.patcog.2017.05.003).
- [39] G.C. Liu, Z.C. Lin, S.C. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (1) (2013) 171–184, doi:[10.1109/TPAMI.2012.88](https://doi.org/10.1109/TPAMI.2012.88).
- [40] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, Y. Xu, Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (1) (2017) 156–171, doi:[10.1109/TPAMI.2016.2535218](https://doi.org/10.1109/TPAMI.2016.2535218).
- [41] Y.G. Zhai, L.F. Zhang, N. Wang, Y. Guo, Y. Cen, T.X. Wu, Q.X. Tong, A modified locality-preserving projection approach for hyperspectral image classification, *IEEE Geosci. Remote Sens. Lett.* 13 (8) (2016) 1059–1063, doi:[10.1109/LGRS.2016.2564993](https://doi.org/10.1109/LGRS.2016.2564993).
- [42] W.F. Liu, P.P. Pokharel, J.C. Principe, The kernel least-mean-square algorithm, *IEEE Trans. Signal Process.* 56 (2) (2008) 543–554, doi:[10.1109/TSP.2007.907881](https://doi.org/10.1109/TSP.2007.907881).
- [43] I. Naseem, R. Togneri, M. Bennamoun, Linear regression for face recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (11) (2010) 2106–2112, doi:[10.1109/TPAMI.2010.128](https://doi.org/10.1109/TPAMI.2010.128).
- [44] X.F. He, D. Cai, P. Niyogi, Tensor subspace analysis, in: *Proceedings of the NIPS, 2005*, pp. 5–8.
- [45] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: *Proceedings of the Second IEEE International Workshop Application Computer Vision*, 1994, pp. 138–142, doi:[10.1109/ACV.1994.341300](https://doi.org/10.1109/ACV.1994.341300).
- [46] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (7) (1997) 711–720, doi:[10.1109/34.598228](https://doi.org/10.1109/34.598228).
- [47] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression database, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (12) (2003) 1615–1618, doi:[10.1109/TPAMI.2003.1251154](https://doi.org/10.1109/TPAMI.2003.1251154).
- [48] A.M. Martinez, R. Benavente, The AR face database, Center de Visio per Computador, Universitat Autònoma de Barcelona, Barcelona, Spain, Technical Report, 24, 1998.
- [49] G. B. Huang M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: a database for studying face recognition in unconstrained environments, Technical Report 07-49, University of Massachusetts, Amherst, 2007.



Yang-Jun Deng received the B.S. degree in mathematics from Wuyi University in 2012 and the M.S. degree in mathematics from University of South China in 2015. He is currently pursuing the Ph.D. degree in signal and information processing in the School of Information Science and Technology, Southwest Jiaotong University. His current research interests include pattern recognition and remote sensing image processing.



Heng-Chao Li received the Ph.D. degree in information and communication engineering from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2008. He is currently a Professor with the Sichuan Provincial Key Laboratory of Information Coding and Transmission, Southwest Jiaotong University. His research interests include statistical analysis of synthetic aperture radar images, and remote sensing image processing. Prof. Li is currently serving as an Associate Editor of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing.



Qi Wang received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent system from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively. He is currently a Professor with the School of Computer Science and the Center for Optical Imagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.



Qian Du received the Ph.D. degree in electrical engineering from the University of Maryland-Baltimore County, Baltimore, MD, USA, in 2000. She is currently a Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, pattern classification, data compression, and neural networks. Dr. Du is a fellow of IEEE and SPIE. She was a recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS). She served as a Co-Chair for the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013. She was the

Chair with the Remote Sensing and Mapping Technical Committee of International Association for Pattern Recognition from 2010 to 2014. She was the General Chair for the fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held at Shanghai, China, in 2012. She served as an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS), the Journal of Applied Remote Sensing, and the IEEE SIGNAL PROCESSING LETTERS. Since 2016, she has been the Editor-in-Chief of the IEEE JSTARS.