Pattern Recognition 133 (2023) 109034

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Low-rank 2D local discriminant graph embedding for robust image feature extraction

Minghua Wan[a,b,c], Xueyu Chen[a], Tianming Zhan[a], Guowei Yang[a,d], Hai Tan[a], Hao Zheng[b,*]

[a] School of Information Engineering, Nanjing Audit University Nanjing, Nanjing 211815, China
[b] Key Laboratory of Intelligent Information processing, Nanjing Xiaozhuang University, Nanjing 211171, China
[c] Jiangsu Key Laboratory of Image and Video Understanding for Social Safety, Nanjing University of Science and Technology, Nanjing 210094, China
[d] School of Electronic Information, Qingdao University, Qingdao 266071, China

ARTICLE INFO

Article history:
Received 8 February 2022
Revised 7 August 2022
Accepted 6 September 2022
Available online 11 September 2022

Keywords:
Feature extraction
Two-dimensional locality preserving projections (2DLPP)
Low-rank
Graph embedding (GE)
Discrimination information

ABSTRACT

As a popular feature extraction algorithm, the 2D local preserving projections (2DLPP) algorithm has been successfully applied in many fields. Using 2D image representation, the 2DLPP algorithm preserves the manifold attributes and retains the local information of high-dimensional space data. However, the 2DLPP algorithm may encounter some problems in real-world applications, such as a lack of discriminatory ability, singularity problems, and sensitivity to occlusion and noise in data. Therefore, this paper introduces low-rank into the 2DLPP algorithm and proposes a new feature extraction algorithm, which is the low-rank two-dimensional local discriminant graph embedding (LR-2DLDGE), to solve these problems. To improve the LR-2DLDGE algorithm robustness, we fuse the discriminant information in graph embedding and the low-rank properties of the data. The algorithm has three advantages: First, the algorithm uses a graph embedding (GE) framework to maintain the local neighbourhood discrimination information between data. Second, the LR-2DLDGE algorithm ensures that the data points are as independent as possible from different classes in the feature space. Finally, the algorithm uses the L_1-norm as a constraint and reduces the influence of noise and corruption through low-rank learning. The theoretical computational complexity and convergence of the algorithm are explicated and proved. Extensive experimental results on three occluded and noisy image datasets confirm the effectively and robustness of LR-2DLDGE, respectively.

1. Introduction

In recent years, with the rapid development of information society, multimedia image data has the characteristics of large scale, many classes, shorten production period, great value but low density. How to analyze these data, mine the key variables of the data, obtain the key knowledge hidden in the data, and extract more effective features for image storage and retrieval is an important development direction of current data processing. At the same time, large-scale multimedia image retrieval has broad market and academic significance. Therefore, the "curse of dimensionality" [1] problem has been a hot topic for many researchers. The most commonly used method to overcome this problem is dimension reduction. Proposed solutions include many linear feature extraction algorithms, including nonpeaked discriminant analysis (NDA) [2], local manifold-based sparse discriminant learning (LMSDL) [3], low-rank matrix regression (LMR) [4], and low-rank adaptive graph embedding (LAGE) [5], to solve the problems in feature extraction.

However, nonlinear data can be divided into non-Gaussian data and manifold data in the real-world applications. Linear feature extraction algorithms may not be able to discover the essential nonlinear data structures. The maintenance of the local geometric structure of non-Gauss and manifolds has been an important research field in the past ten years. We usually use local patches to process non-Gaussian data and local Euclidean observations to process curve manifold data [6].

Therefore, some representative nonlinear manifold learning techniques, such as Laplacian eigenmaps (LE) [7], ISOMAP [8], and locally linear embedding (LLE) [9], have been proposed to reveal hidden semantics while maintaining the geometric structure of manifolds. However, the above nonlinear algorithms all have the same problem: they are out of sample [10]. The graph embedding (GE) framework [11] unifies most existing graph-based subspace learning algorithms to ensure that the relationship between vertices in the projected low and high-dimensional space is as simi-

* Corresponding author.
E-mail address: zhh710@163.com (H. Zheng).

https://doi.org/10.1016/j.patcog.2022.109034
0031-3203/

lar as possible. This framework emphasizes the importance of constructing a similarity matrix and proposes a new GE formula.

Therefore, LE and LLE algorithms and their linearized versions, such as locality preserving projections (LPP) [12,13], neighbourhood preserving projections (NPP) [14,15], and neighbourhood preserving embedding (NPE) [16,17], represent advanced ways to solve the out-of-sample problem.

Due to the singularity of the matrix in the generalized eigen-equation, LPP and LDA encounter the small sample sizes (SSS) problem [18]. The proposed 2D local preserving projections (2DLPP) [19] is inspired by the direct action of 2DPCA and 2DLDA on a 2D image matrix and is used for linear feature extraction and dimension reduction. However, 2DLPP is an unsupervised algorithm, and some new supervised versions of 2DLPP have been also proposed [20,21].

However, the 2DLPP may suffer from some problems, such as (i) it is an unsupervised learning algorithm without considering the class information of training samples; (ii) it uses the L_2-norm criterion to measure the similarity of projection data pairs, so it is sensitive to outliers; and (iii) it has singularity and cannot solve the eigenvalue problem.

The 1D vector-based algorithms or the 2D matrix-based algorithms discussed above all use the L_2-norm as a metric, and these algorithms are very sensitive to noise or outliers. These problems may reduce the performance of the algorithms and have a negative impact on the key information.

Compared with algorithms based on the L_2-norm or L_1-norm, low-rank representation (LRR) [22] has good performance in recovering a clean matrix from noisy data. Reference [23] used the nuclear norm to represent reconstruction error and proposed a 2DPCA algorithm based on the nuclear norm (N-2DPCA) [23] to improve image representation. The supervised low-rank discriminant algorithm is proposed in reference [24], and robust PCA (RPCA) is proposed to recover the noise data matrix in reference [25]. However, in practical applications, most of the feature extraction algorithms discussed above are easily affected by illumination, corrosion, and noise.

In recent years, the robustness of many feature extraction algorithms based on LRR has attracted much attention [26–28] to noise contaminated data. The LRR in reference [22] introduces the single subspace clustering problem into multiple subspace clustering to maintain the lowest rank representation and the global structure of data. In reference [26], robust PCA (RPCA) is proposed by introducing the nuclear norm, which can recover the subspace structure from noisy or occluded data. In reference [27], Laplace regularization LRR is proposed, which uses a regularization term with manifold structure in clustering data. The nonnegative low-rank sparse graph (NNLRS) [28] combines with the NNLRS-graph and LRR is proposed by introducing the low roughness to maintain the global structure.

According to these analyses, our goal is to learn a good optimal projection matrix, which can simultaneously perform supervised feature selection and subspace clustering in this work. Motivated by [27] and [28], we introduce the low-rank and discriminant Laplace regularization constraint into NMF to utilize a seamless model. Therefore, we propose a new feature extraction and dimension reduction algorithm, named the low-rank two-dimensional local discriminant graph embedding (LR-2DLDGE) algorithm, to overcome the sensitivity of the 2DLPP algorithm in this paper. The algorithm is implemented in two steps. First, the intraclass weighted matrix graph and the interclass weighted matrix graph are constructed to maintain the discriminant information of the local neighbourhood. Second, the given data are divided into two parts: 1) the low-rank feature coding part, and 2) the sparse noise error part. According to the identification ability of graph embedding and the sparsity of low- rank learning in the objective function, the features are kept to retain the optimal features.

The main contributions of LR-2DLDGE method are summarized as follows:

- It learns a low-rank matrix based on the graph embedding that can simultaneously perform subspace learning, graph Laplacian regularization, and low-rank learning in a unified strategy is proposed and an iterative solution to the convex optimization problem is provided;
- It combines the graph embedding framework with the low-rank matrix, two intraclass and interclass weighted matrix graphs are proposed, which fully discover the manifold structural information of the neighbourhood and improve the recognition ability in 2D images;
- It proposes to ensure that the given data are divided into a low-rank feature coding part and a sparse noise error part to improve the recognition ability, which can weaken the influence of noise and occlusion when learning the optimal projection.

The remaining sections of this article are as follows: The section II mainly introduces the related works, such as 2DLPP, LRR, and LRMD-SI. In section III, the model, convergence, and complexity analysis of the LR-2DLDGE algorithm are introduced in detail. The results of the FERET, ORL, COIL 100, AR, Yale, and PolyU databases show the effectiveness of the algorithm in section IV. Finally, section V summarizes the whole work and discusses the future work.

2. Related works

To facilitate the understanding of our proposed method, we will cite some related work including GE learning, low-rank learning and structurally incoherent learning, i.e., the 2DLPP, LRR, and 2DRLPP algorithms.

First, the matrix $A=[a_{ij}] \in R^{m\times n}$ is defined, and then A_i or $A_{\cdot j}$ is used to represent the i^{th} or j^{th} row of A, respectively. These vector norms are interpreted by the Frobenius norm $\|A\|_F = \sqrt{\sum \|A_i\|_2^2}$, the L_1 norm $\|A\|_1 = \sum_{i,j} |a_{ij}|$, and the $L_{2,1}$ norm $\|A\|_{2,1} = \sum_j \|A_{\cdot j}\|_2$, and the nuclear norm $\|A\|_*$ is interpreted as the sum of the singular matrix.

We consider N high-dimensional space samples $X_i = \{X_1, X_2, \cdots, X_N\} \in R^{m\times n}$ into low-dimensional space features$Y_i = \{Y_1, Y_2, \cdots, Y_N\} \in R^{d\times n}$, $P \in R^{m\times d}$is the projection matrix.

Then, we can obtain the formula as follows:

$$Y_i = P^T X_i, i = (1, 2, \cdots, N) \tag{1}$$

where $n > d$, m, n and d represent the size of matrix.

In the real-world applications, there are a large number of nonlinear data, including non Gaussian data and manifold data. GE framework usually uses local patches to process non Gaussian data, and local Euclidean observations to process curve manifold data, so as to ensure that the relationship between vertices in projection low and high-dimensional space is as similar as possible, such as 2DLPP algorithm.

We supposed that the graph $G = \{X, W\}$ contains similarity matrix $W \in R^{N\times N}$ and vertex set X, in which an image X_i corresponds to a node. The similarity matrix W in the algorithm is obtained by the k-neighbourhood or ε-neighbourhood, which has a uniform weight of Euclidean distance or Gaussian weight. The equation can be obtained as follows:

$$W_{ij} = \begin{cases} 1, & \left\|X_i - X_j\right\|^2 < \varepsilon \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

Therefore, we can define the 2DLPP algorithm equation:

$$\min \sum_{i,j} \left\|Y_i - Y_j\right\|^2 W_{ij} \tag{3}$$

where the symbol $\| \bullet \|$ represents the L_2 norm. In this paper, the linear transformation $Y_i = P^T X_i$ is replaced by Eq. (3), and the minimization problem of this equation can be transformed into the following equation through the matrix transformation step:

$$\begin{aligned} &\arg\min_P P^T X(L \otimes I_n) X^T P, \\ &\text{s.t.} P^T X(D \otimes I_n) X^T P = 1 \end{aligned} \tag{4}$$

where $\otimes$ is defined as the Kronecker product matrix. I_n is defined as an identity matrix of n. The diagonal matrix is D, and its entries are the sum of rows or columns of W, and $L = D - W$.

The generalized eigenvalue problem can be used to obtain the d minimum optimal projection vectors of the objective function.

$$X(L \otimes I_n) X^T P = \lambda X(D \otimes I_n) X^T P \tag{5}$$

LRR can obtain clean data to enhance the robustness to noise as well discloses the relationship of samples by low-rank structure. Thus, exploring the low-rank subspace structures becomes a challenging problem. As the linear combination of other samples, it represents the data vector to solve the subspace segmentation problem. The linear combination matrix F can be represented by every column of X, for example:

$$X = FZ, \tag{6}$$

where $Z = [Z_1, Z_2, \ldots Z_n]$ is obtained by the LRR algorithm which is the coefficient matrix:

$$\min_Z \|Z\|_*, \text{s.t.} X = FZ, \tag{7}$$

where $\| \cdot \|_*$ represents the nuclear norm.

The damaged data can be represented by the noise term E, so we obtain Eq. (7).

$$\min_Z \|Z\|_* + \lambda \|E\|_{2,1}, \text{s.t.} X = FZ + E, \tag{8}$$

where λ is an adjusted parameter and $\|E\|_{2,1} = \sum_{j=1}^{n} \sqrt{\sum_{i=1}^{m} (E_{ij})^2}$.

Recently, many dimensionality reduction algorithms using the L_1-norm as a distance criterion have been proposed [29–39]. Algorithms based on the L_1-norm, such as L_1 principal component analysis (L_1-PCA) [29] and PCA-L_1 [30], are solved by an optimization problem that reduces noise and outlier sensitivity in the data. The rotation invariant L_1-norm PCA (R1-PCA) [31] is proposed based on L_1-PCA, and it also shares some properties of PCA that L_1-PCA does not have. A nongreedy algorithm is used to search the maximum projection principal component of the L_1-norm to solve the general L_1-norm maximization problem [32]. Tensor PCA of the L_1-norm (TPCA-L_1) [33] uses the L_1-norm as the distance measure of tensor analysis. In reference [34], an optimization algorithm is advanced to compute the L_1principal component by maximizing the L_1energy in the projection space. An iterative algorithm for L_1-norm difference problems using a greedy strategy is proposed in reference [35]. Reference [36] extended the PCA-L_1 algorithm to the 2DPCA (2DPCA-L_1) algorithm based on the L_1-norm. To solve the outliers and corrosion in reference [37], the 2DLPP algorithm based on the PCA-L_1 algorithm (2DLPP-L_1) [38] is proposed. In reference [39], 2DLPP based on the L_1-norm (2D-DLPP-L_1) is proposed to preserve the spatial topological structure more effectively.

Robust two-dimensional locality preserving projection with regularization (2DRLPP) [21] combined 2DLPP and L_1-norm to improve the recognition accuracy and discrimination ability of the learning basis matrix. The final function of the 2DRLPP algorithm is as follows:

$$\begin{aligned} &\min_{W'W=I} \sum_{i,j}^{m} \left\| Z_{ij} P \right\|_1 W_{ij} + \sigma \|P\|_1 \\ &s.t. \sum_{i,j}^{m} \|X_i P\|_1 D_i = 1, \end{aligned} \tag{9}$$

where matrix $\|.\|_1$ is the L_1-norm of the matrix and $\sigma > 0$ is the regularization parameter. In Eq. (9), the first term represents the local manifold structure of the image matrix space, while the second term represents an L_1-norm regularization term.

Many optimization algorithms can be used to solve the above methods. The Dwarf Mongoose Optimization (DMO) [40] algorithm mimics the foraging behavior of the Dwarf Mongoose to solve the optimization problems. By the Aquila's behaviors, in nature during the process of catching the prey Aquila Optimizer (AO) [41] algorithm proposes the meta-heuristic optimization algorithm. By the hunting behaviour of Crocodiles, Reptile Search Algorithm (RSA) [42] algorithm advances the nature-inspired meta-heuristic optimizer. To address the effectiveness of this disease transmission strategy, Ebola Optimization Search Algorithm (EOSA) [43] algorithm presents a novel metaheuristic algorithm. Arithmetic Optimization Algorithm (AOA) [44] algorithm utilizes the distribution behavior of arithmetic operators in mathematics, which is a new meta-heuristic method. Applications, Deployments, and Integration of Internet of Drones (IoD) [45] algorithm is a new comprehensive survey of IoD and its applications, deployments, and integration methods.

2DRLPP does not consider the noise of data, and it is also an unsupervised method. Therefore, how to remove the noise of data and obtain relatively clean data, while adding identification information is also the focus of our work.

3. Low-rank 2D local discriminant graph embedding

First, the objective function of the LR-2DLDGE algorithm is given, and its solution is optimized. Then, the convergence and complexity of the algorithm are analysed.

3.1. Motivation

Currently, many 2DLPP algorithm versions [19] have been proposed to overcome the limitations of the LPP algorithm. In this paper, we use a low-rank learning algorithm to improve the sensitivity of the 2DLPP algorithm and its versions to noise and outliers observed in data. The 2DLDGE [46] algorithm uses an image matrix to design the input and the MMC criterion function to minimize the difference between the intraclass scatter matrix S_w and the interclass scatter matrix S_b to construct the objective function:

$$J(P) = \min_P (S_w - S_b) \tag{10}$$

$$\begin{aligned} J(P) &= \min_P \left(\sum_{j=1}^{k_1} \left\| Y_i - Y_i^j \right\|^2 - \gamma \sum_{q=1}^{k_2} \left\| Y_i - Y_i^q \right\|^2 \right) \\ &= \min_P \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \left\| Y_i - Y_i^j \right\|^2 S_{ij}^w \right. \\ &\quad \left. - \gamma \sum_{i=1}^{N} \sum_{j=1}^{N} \left\| Y_i - Y_i^q \right\|^2 S_{iq}^b \right) \\ &= tr\left(P^T X(L^c \otimes I_n) X^T P - \gamma P^T X \left(L^b \otimes I_n \right) X^T P \right) \end{aligned} \tag{11}$$

where γ is used to balance $P^T X(L^c \otimes I_n) X^T P$ and $P^T X(L^b \otimes I_n) X^T P$. S_{ij}^w is the intraclas similarity matrix and S_{iq}^b is the interclass similarity matrix. They can be defined as follows::

$$S_{ij}^w = \begin{cases} 1, \ if \ X_i \in N_{K_c}^+(X_j) \ or \ X_j \in N_{K_c}^+(X_i) \\ 0, \ otherwise. \end{cases} \tag{12}$$

$$S_{iq}^b = \begin{cases} 1, \ X_i \in N_{K_b}^+(X_q) \ or \ X_q \in N_{K_b}^+(X_i) \\ 0, \ otherwise. \end{cases} \tag{13}$$

where $N_{K_c}^+(X_i)$ or $N_{K_c}^+(X_j)$ is the index set of K_c nearest neighbours of X_i or $X_j$in the same class, respectively. $N_{K_b}^+(X_i)$ or $N_{K_b}^+(X_q)$ is the

index set of K_b nearest neighbours of X_i or X_q in different classes, respectively.

Therefore, we propose a new algorithm to obtain low-rank matrix A. First, it is assumed that X can be divided into clean data A and sparse noise E, that is, $X = A + E$. Second, the nuclear norm is added to A, and the L_1-norm is used as E.

$$\begin{aligned} &\min_{A,E} rank(A) + \lambda \|E\|_1, \\ &\text{s.t.} X = A + E, \end{aligned} \tag{14}$$

By combining (11) with (14), we have:

$$\begin{aligned} &\min_{A,P,E} \sum_{i,j,q} \left\|Y_i - Y_j\right\|_2^2 S_{ij}^w - \gamma \|Y_i - Y_q\|_2^2 S_{iq}^b + \alpha rank(A) + \beta \|E\|_1 \\ &s.t. Y_i = P^T A_i, X = A + E \end{aligned} \tag{15}$$

where the balance parameters $\alpha > 0$ and $\beta > 0$.

To find the solution, we transform the NP hard problem of Eq. (15) into the following functions:

$$\begin{aligned} &\min_{A,P,E} \sum_{i,j,q} \left\|Y_i - Y_j\right\|_2^2 S_{ij}^w - \gamma \|Y_i - Y_q\|_2^2 S_{iq}^b + \alpha \|A\|_* + \beta \|E\|_1 \\ &s.t. Y_i = P^T A_i, X = A + E \end{aligned} \tag{16}$$

The first term in Eq. (16) retains the learned clean data neighbourhood and projects it into a new subspace. The second term guarantees the low-rank matrix, which can be recovered from noisy data. The third term ensures the noise is sparse in the data.

3.2. The objective function of LR- 2DLDGE

First, the objective function can be separated by introducing auxiliary variables. Then, the alternating direction method of multiplier algorithm (ADMM) [46] is often used to advance the optimization problem in Eq. (16). Finally, the following equation is obtained:

$$\begin{aligned} &\min_{A,E,P} \sum_{i,j,q} \left\|Y_i - Y_j\right\|_2^2 S_{ij}^w - \gamma \|Y_i - Y_q\|_2^2 S_{iq}^b + \alpha \|A\|_* + \beta \|E\|_1 \\ &s.t. A = B, Y_i = P^T A_i, X = A + E \end{aligned} \tag{17}$$

For the convenience of calculation, we introduce the auxiliary matrix B. Using the augmented Lagrange function, Eq. (18) is obtained:

$$\begin{aligned} &L(A, B, \mu, P, E, M_1, M_2) \\ &= \sum_{i,j,q} \left\|Y_i - Y_j\right\|_2^2 S_{ij}^w - \gamma \|Y_i - Y_q\|_2^2 S_{iq}^b + \alpha \|B\|_* + \beta \|E\|_1 \\ &\quad + tr\left(M_2^T (A - B)\right) + tr\left(M_1^T (X - A - E)\right) \\ &\quad + \frac{\mu}{2}\left(\|A - B\|_F^2 + \|X - A - E\|_F^2\right) \\ &= 2P^T A(L^w \otimes I_n) A^T P - 2\gamma P^T A\left(L^b \otimes I_n\right) A^T P + \alpha \|B\|_* + \beta \|E\|_1 \\ &\quad + \frac{\mu}{2}\left(\left\|A - B + \frac{M_2}{\mu}\right\|_F^2 + \left\|X - A - E + \frac{M_1}{\mu}\right\|_F^2\right) \\ &\quad - \frac{1}{2\mu}\left(\|M_1\|_F^2 + \|M_2\|_F^2\right) \end{aligned} \tag{18}$$

where M_1 and M_2 are Lagrangian multipliers and $\mu > 0$ is the penalty parameter.

3.3. Optimization

3.3.1. Fix E, A, and P, and update B

If the variable B is updated and the values of other variables are fixed, then the solution of variable B is as follows:

$$\min_B \alpha \|B\|_* + \frac{\mu}{2}\left\|B - A - \frac{M_2}{\mu}\right\|_F^2 \tag{19}$$

Given the matrix $Q \in R^{q \times q}$, we can obtain the solution of the above equation by the singular value decomposition (SVD) threshold operator [47,48] of matrix Q as follows:

$$Q = U_{a \times r} \Sigma V_{r \times q}, \ \Sigma = diag(\sigma_1, \cdots, \sigma_r) \tag{20}$$

where $\{\sigma_1, \cdots, \sigma_r\} > 0$ and the rank of matrix Q is r. $U_{a \times r}$ and $V_{r \times q}$are corresponding orthogonal column matrices. Therefore, the singular value contraction operator $D_\zeta(Q)$ is explicated as follows:

$$D_\zeta(Q) = U_{a \times r} diag\left(\left\{\max\left(0, \sigma_j - \zeta\right)_{1 \le j \le r}\right\}\right) V_{r \times q}, \tag{21}$$

where $\tau > 0$.

Theorem 1 [48]**.** For each matrix $Q \in R^{q \times q}$ and parameter $\zeta > 0$, the $D_\zeta(Q)$ in Eq. (21) is rewritten as follows:

$$D_\zeta(Q) = \arg\min_B \left(\zeta \|B\|_* + \frac{1}{2}\|B - Q\|_F^2\right) \tag{22}$$

The optimal solution of Eq. (8) is obtained by Theorem 1:

$$B = D_{\frac{1}{\mu}}\left(A + \frac{M_2}{\mu}\right) = U_{a \times r}^1 diag\left(\left\{\max\left(0, \sigma_j - \frac{1}{\mu}\right)_{1 \le j \le r}\right\}\right) V_{q \times r}^1 \tag{23}$$

where $A + [M_2/\mu] = U_{a \times r}^1 \Sigma V_{r \times q}^1$, $\Sigma = diag(\sigma_1, \cdots, \sigma_r)$.

3.3.2. Fix B, A, and P, and update E

The problem of Eq. (18) can be rewritten by updating variable E and fixing other variable values:

$$\Psi = \arg\min_E \frac{\mu}{2}\left\|X - A - E + \frac{M_1}{\mu}\right\|_F^2 + \beta \|E\|_1 \tag{24}$$

We use the shrinkage operator [49] to find the solution of Eq. (24). The soft threshold operator $S_\varepsilon[X] = \max(|X| - \varepsilon, 0).sign(X)$ is defined to promote shrinkage [50]. Finally, we obtain the solution of Eq. (24).

$$E = S_{\frac{\beta}{\mu}}\left(X - A + \frac{M_1}{\mu}\right) \tag{25}$$

3.3.3. Fix B, E, and A, and update P

We fix the value of variable P in Eq. (18) and change the value of other variables. We obtain the following equation:

$$\Psi = \arg\min_P 2P^T A(L^w \otimes I_n) A^T P - 2\gamma P^T A\left(L^b \otimes I_n\right) A^T P \tag{26}$$

We can rewrite Eq. (26) as:

$$\begin{aligned} &2P^T A\left(\left(L^w - L^b\right) \otimes I\right) A^T P \\ &= 2\gamma P^T (X - E)\left(\left(L^w - L^b\right) \otimes I_n\right)(X - E)^T P \end{aligned} \tag{27}$$

To solve the problem, we add a constraint, as shown in the following equation:

$$P^T (X - E)\left(D^w - D^b\right)(X - E)^T P = 1 \tag{28}$$

Finally, Eq. (27) and Eq. (28) can be rewritten as follows:

$$\arg\min_{P^T (X-E)\left(D^w - D^b\right)(X-E)^T P = 1} P^T (X - E)\left(\left(L^w - L^b\right) \otimes I_n\right)(X - E)^T P \tag{29}$$

To solve Eq. (29), we transform it into the following equation to solve the generalized eigenvalue problem of Eq. (30):

$$(X - E)\left(\left(L^w - L^b\right) \otimes I\right)(X - E)^T P = \Lambda (X - E)\left(D^w - D^b\right)(X - E)^T P \tag{30}$$

where Λ is the set of eigenvalues.

Algorithm 1 The proposed LR-2DLDGE method

Input: Parameter α, β, γ in Eq (15); and training set X;
Initialization: $M1 = 0$, $M2 = 0$, $B = 0$, $E = 0$, $\rho > 0$.
Repeat
Step 1. Fixing E, A and P, update B by Eq. (23);
Step 2. Fixing B, A and P, update E by Eq. (25);
Step 3. Fixing E, A and B, update P by Eq. (30);
Step 4. Fixing E, B and P, update A by Eq. (32);
Step 5. Update parameter $M1$ and $M2$ as follows:
$M1 \leftarrow \mu(X - A - E) + M1$;
$M2 \leftarrow \mu(A - B) + M2$.
Step 6. $\mu \leftarrow \min(\max\mu, \rho\mu)$.
Step 7. $t \leftarrow t + 1$;
Until Eq.(17) is converged.
Step 8. After many iterations, the optimal solution (B, E, A, P) is obtained.
Output: The projection matrix P.

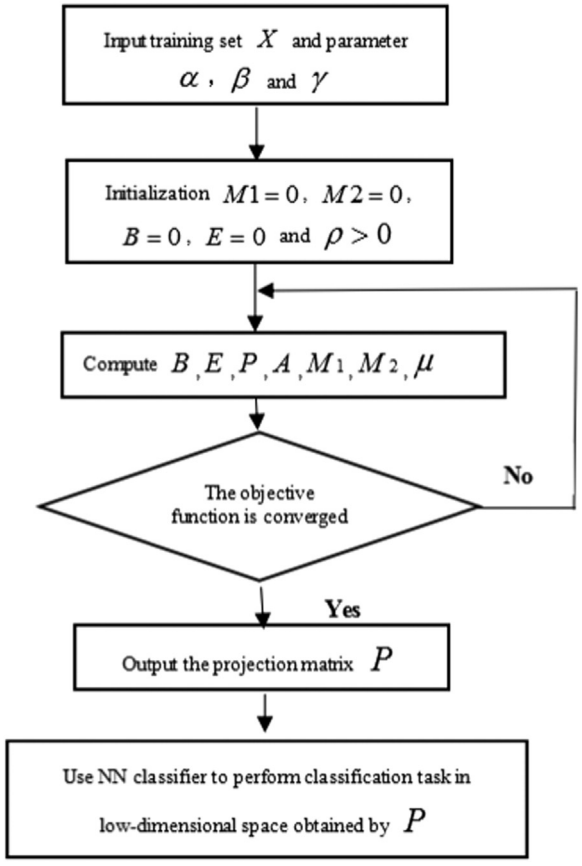


Fig. 1. The flow chart of the proposed method.

3.3.4. Fix B, E, and P, and update A

We fix the value of variable A in Eq. (18) and change the value of other variables. We obtain the following equation:

$$\Psi = \arg\min_{A} 2P^T A(L^w \otimes I_n)A^T P - 2P^T A\left(L^b \otimes I_n\right)A^T P + \frac{\mu}{2}\left(\left\|X - \left(A + E - \frac{M_1}{\mu}\right)\right\|_F^2 + \left\|A - B + \frac{M_2}{\mu}\right\|_F^2\right) \tag{31}$$

We obtain the following equation to find the solution by setting the derivative $\frac{\partial\psi}{\partial A} = 0$:

$$4PP^T A\left(\left(L^w - L^b\right) \otimes I_n\right) + 2\mu A - \mu(Q_1 + Q_2) = 0 \tag{32}$$

where $Q_1 = B - \frac{M_2}{\mu}$ and $Q_2 = X - E + \frac{M_1}{\mu}$. Update A by solving the Sylvester equation.

Algorithm 1 gives the concrete steps of **LR-2DLDGE**. The flow chart of the proposed method is shown in Fig 1.

Algorithm 2 ADMM

Set $\mu > 0$, v_0, d_0, $t = 0$;
Repeat
Step 1. $u^{t+1} \leftarrow \arg\min_u gf(v) + \frac{\mu}{2}\|Gv^{t+1} - u - d^t\|_2^2$;
Step 2. $v^{t+1} \leftarrow \arg\min_v f(v) + \frac{\mu}{2}\|Gv - u^t - d^t\|_2^2$;
Step 3. $d^{t+1} \leftarrow d^t - (u^{t+1} - Gv^{t+1})$;
Step 4. $t \leftarrow t + 1$;
Stopping conditions are met.
End

3.4. Convergence Analysis

The convergence of the LR-2DLDGE algorithm is analysed. Using the following structured optimization problems, we prove its convergence with linear constraints:

$$\min_{u,v} g(u) + f(v), \quad s.t. Gv = u \tag{33}$$

where $g(u) : R^n \to R$ and $f(v) : R^m \to R$ are convex functions. We can obtain the augmented Lagrangian function of the above equation as follows:

$$L(u, v, \alpha) = g(u) + f(v) + \alpha^T(Gv - u) + \frac{\mu}{2}\|Gv - u\|_2^2 = g(u) + f(v)\frac{\mu}{2}\|Gv - u - d\|_2^2 + constant \tag{34}$$

where the Lagrange multiplier is α, the penalty parameter is $\mu > 0$, and the parameter $d = -\frac{\alpha}{u}$.

Algorithm 2 introduces the algorithm steps of the ADMM, and Theorem 2 describes the convergence of the ADMM. Unlike the classical augmented Lagrangian algorithm, ADMM uses the Gauss-Seidel algorithm [38] to minimize $L(u, v, a)$ relative to u and v.

Theorem 2 [51]. *If f, g are proper, closed, convex functions and Ghas full rank, then the following conditions hold in Eq. (33). For* arbitrary u_0, v_0, d_0, and$\mu > 0$, *if Eq. (33) has no solution, Algorithm 2 generated the sequence* $\{u_t, v_t, d_t\}$, *which converges to the solution; otherwise, the sequences* $\{d_t\}$and $\{(u_t, v_t)\}$ *have at least one divergence.*

Eq. (16) is represented as an example of Eq. (34). By setting Theorem 2, the convergence of Algorithm 1 can be guaranteed:

$$f(u) = \left\|Y_i - Y_j\right\|_2^2 S_{ij}^w - \gamma\left\|Y_i - Y_j\right\|_2^2 S_{ij}^b + \alpha\|A\|_* + \beta\|E\|_1 \tag{35}$$

$$g(v) = \left\|Y_i - Y_j\right\|_2^2 S_{ij}^w - \gamma\left\|Y_i - Y_j\right\|_2^2 S_{ij}^b + \alpha\|A\|_* + \beta\|E\|_1 \tag{36}$$

where O_1, O_2, O_3 *and* $V = \begin{matrix}K\\[A]\\E\end{matrix}$, $U = \begin{matrix}K\\[A]\\E\end{matrix}$, *and* $K = \begin{bmatrix} VO_1 \\ O_2O_2 \end{bmatrix} \in R^{a\times q}$ *denote zero matrices.*

Therefore, Eq. (16) can be written as *(34)*. If $G = I$, then $Gu = v$. If the functions f and g are proper, closed, and convex, then G is a full column rank matrix. The convergence of Algorithm 1 is guaranteed when the condition of Theorem 2 is satisfied.

3.5. Computational Complexity

The complexity of the LR-2DLDGE algorithm is mainly determined by steps 1, 3, and 4 in the process of Algorithm 1. The complexity of step 3 is$O(t(2a^3))$ for calculating the SVD of the $q \times q$ matrix, and the complexity of step 1 and step 4 is the same, which is $O(t(2q^3))$ at most in Algorithm 1, and t is the number of iterations. Therefore, the total complexity of the LR-2DLDGE algorithm is $O(t(2(a^3 + q^3)))$.

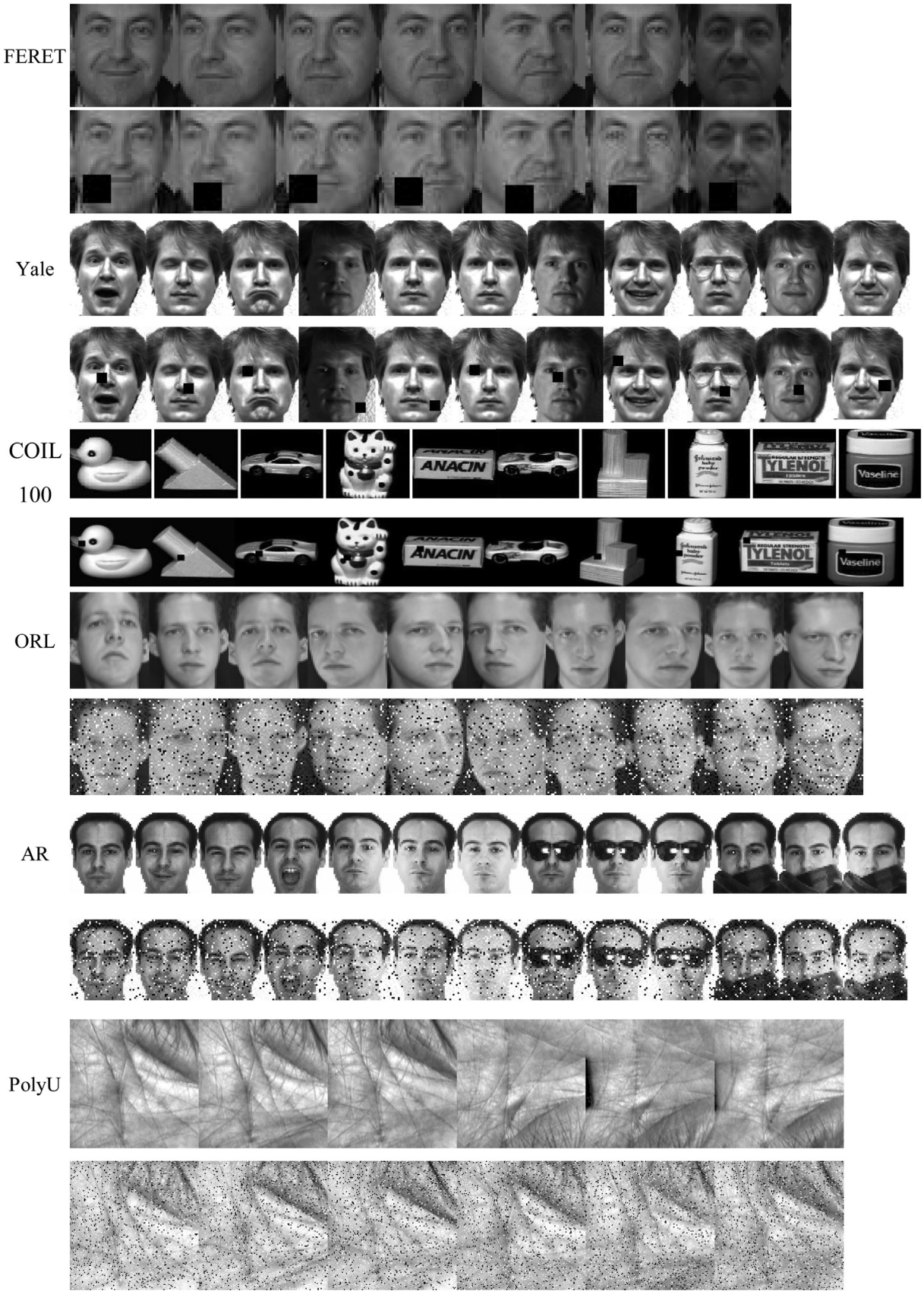
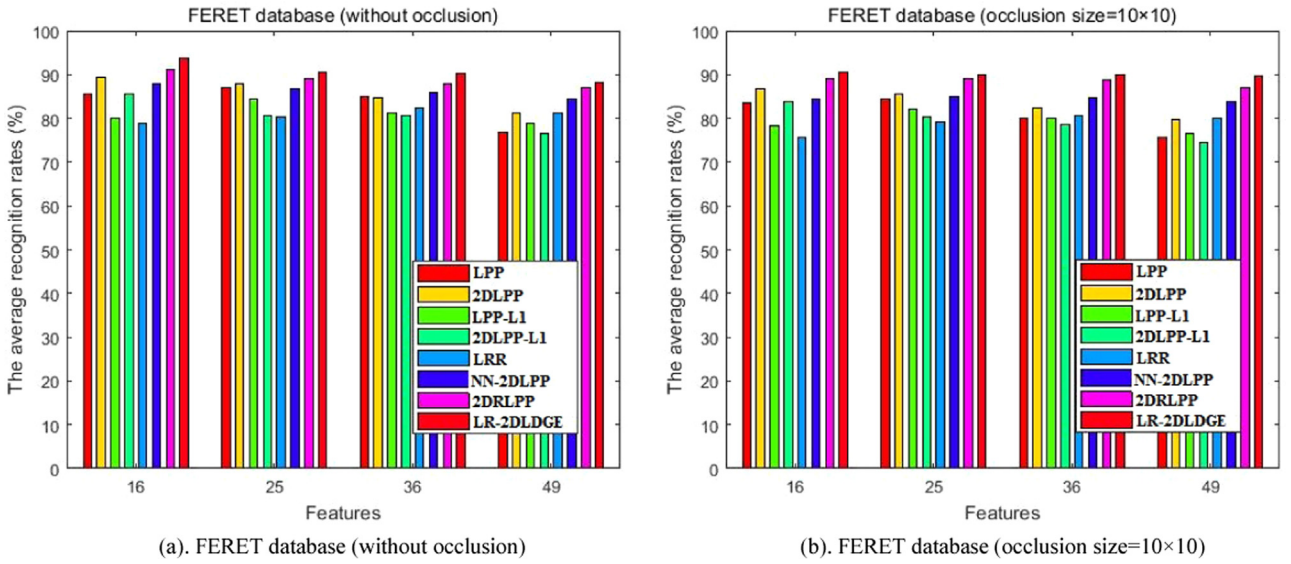
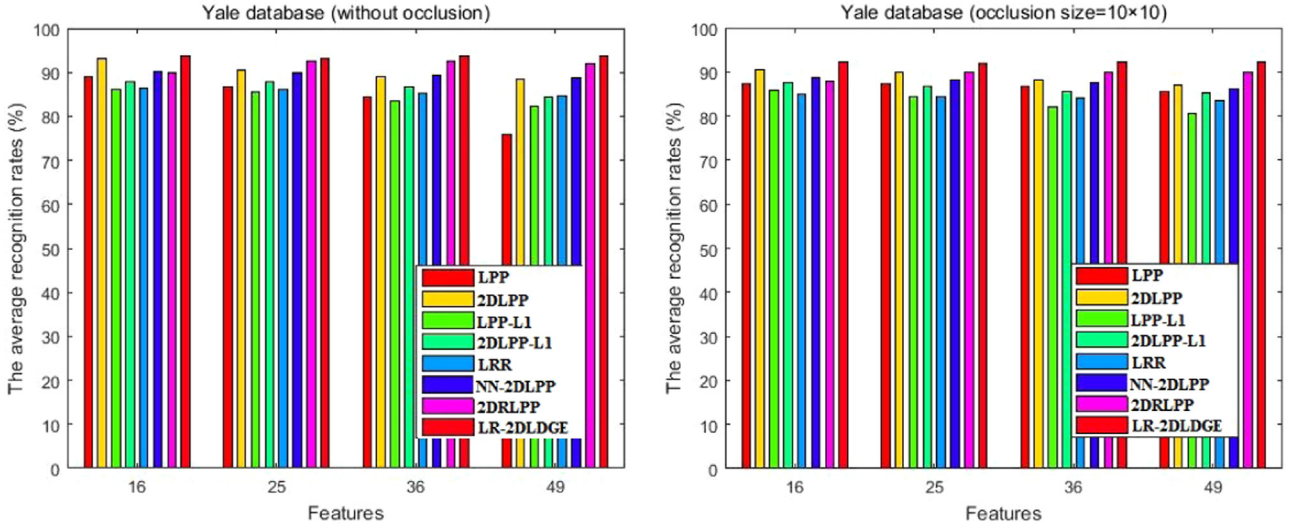


Fig. 2. shows some images in six different databases. (a) FERET, (b) Yale, (c) COIL 100, (d) ORL, (e) AR, and (f) PolyU.



(a). FERET database (without occlusion) (b). FERET database (occlusion size=10×10)

Fig. 3. The average recognition rates (%) and corresponding different dimension changes on the FERET database



(a). Yale database (without occlusion) (b). Yale database (occlusion size=10×10)

Fig. 4. The average recognition rates (%) and corresponding different dimension changes on the Yale database

3.6. The connections among the LPP, 2DLPP, and LR-2DLDGE algorithms

If $\alpha = \beta = 0$ is assumed, Eq. (17) can be developed into Eq. (37).

$$\min_{p} \sum_{i,j} \left\| Y_i - Y_j \right\|_2^2 S_{ij}^w - \gamma \left\| Y_i - Y_j \right\|_2^2 S_{ij}^b \tag{37}$$

If $\gamma = 0$ is assumed, Eq. (37) can be developed into Eq. (38).

$$\min_{p} \sum_{i,j} \left\| Y_i - Y_j \right\|_2^2 S_{ij}^w \tag{38}$$

If the input image is a vector, Eq. (38) is the LPP algorithm. If the input image is a two-dimensional matrix, Eq. (38) is the 2DLPP algorithm. Therefore, if $\alpha = \beta = 0$, the LPP algorithm and 2DLPP algorithm can be regarded as special cases of the LR-2DLDGE algorithm. Different from LPP and 2DLPP, the LR-2DLDGE algorithm eliminates the noise E and projects the clean data points A learned into the new subspace. Therefore, although the LR-2DLDGE algorithm is not disturbed by noise in the data, its measurement is based on Euclidean distance.

3.7. The connections between the LR-2DLDGE, LRR, and RPCA algorithms

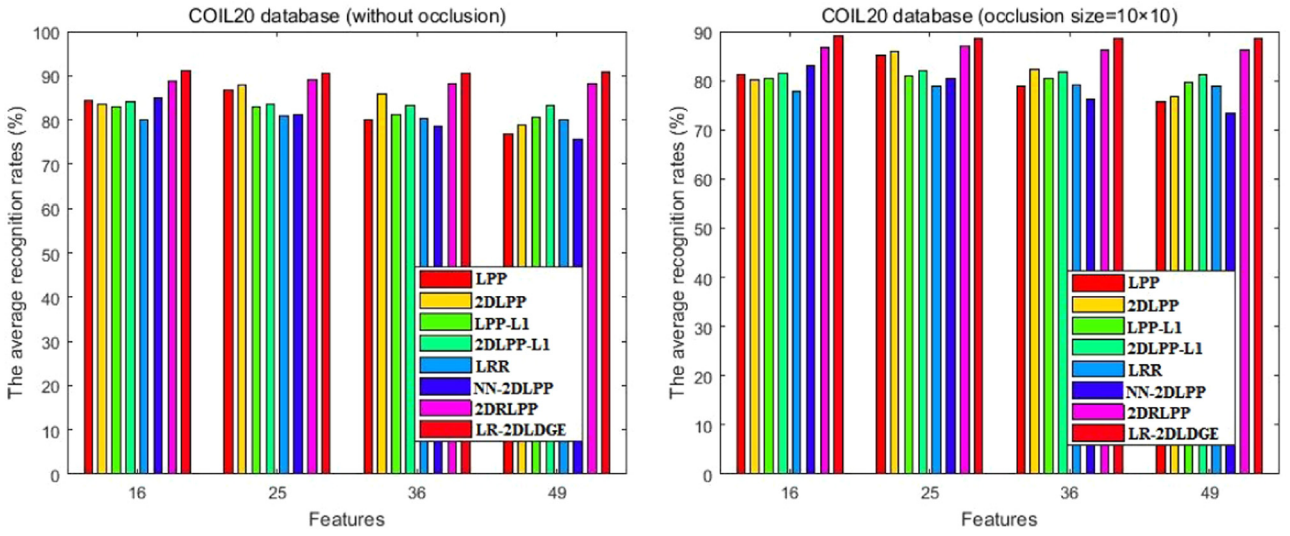
Next, we will analyse the relationship between the LR-2DLDGE, LRR [33] and RPCA algorithms. If the last two terms of Eq. (17) are retained, they are converted to the following:

$$\min_{A,E,X=A+E} \alpha \|A\|_* + \beta \|E\|_1, \tag{39}$$

The $A = XZ$ and Z norms are used instead of the A norm and L_1 is replaced to constrain the error matrix with $L_{2,1}$, we obtain the following equation:

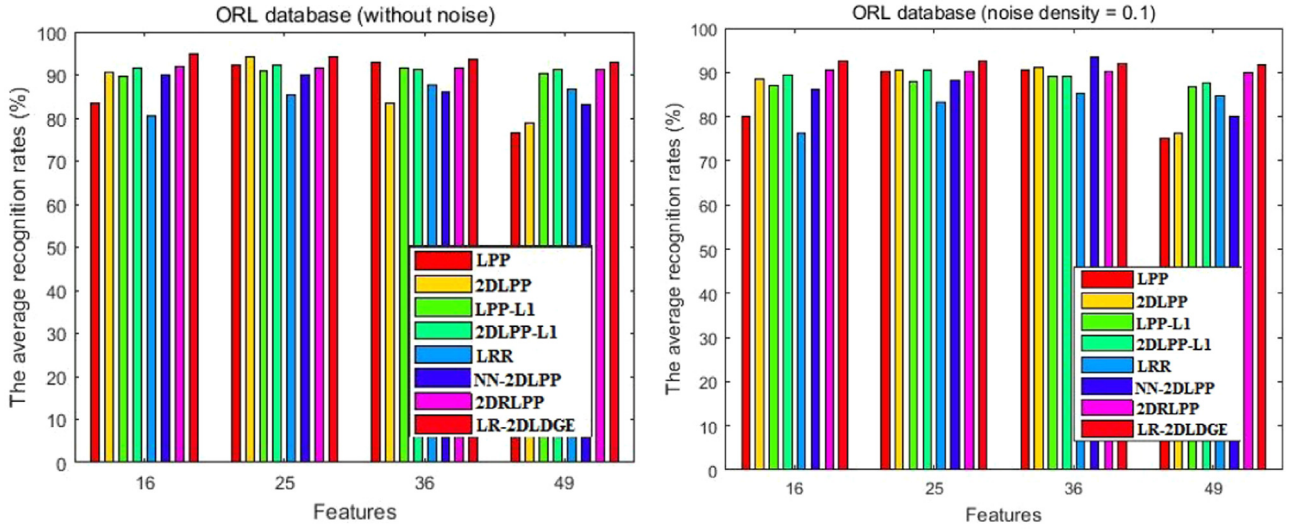
$$\min_{Z,E,X=XZ+E} \|Z\|_* + \beta \|E\|_{2,1}, \tag{40}$$

Therefore, Eq. (40) is the LRR algorithm. Finally, through the above analysis, it is found that the special cases of the LR-2DLDGE



(a). COIL20 database (without occlusion) (b). COIL20 database (occlusion size=10×10)

Fig. 5. The average recognition rates (%) and corresponding different dimension changes on the COIL20 database



(a). ORL database (without noise) (b). ORL database (noise density = 0.1)

Fig. 6. The average recognition rates (%) and corresponding different dimension changes on the ORL database

algorithm are the RPCA algorithm and LRR algorithm. If $\alpha = 1$, Eq. (39) becomes the RPCA algorithm.

Therefore, Eq. (40) is the LRR algorithm. Finally, through the above analysis, it is found that the special cases of the LR-2DLDGE algorithm are the RPCA algorithm and LRR algorithm.

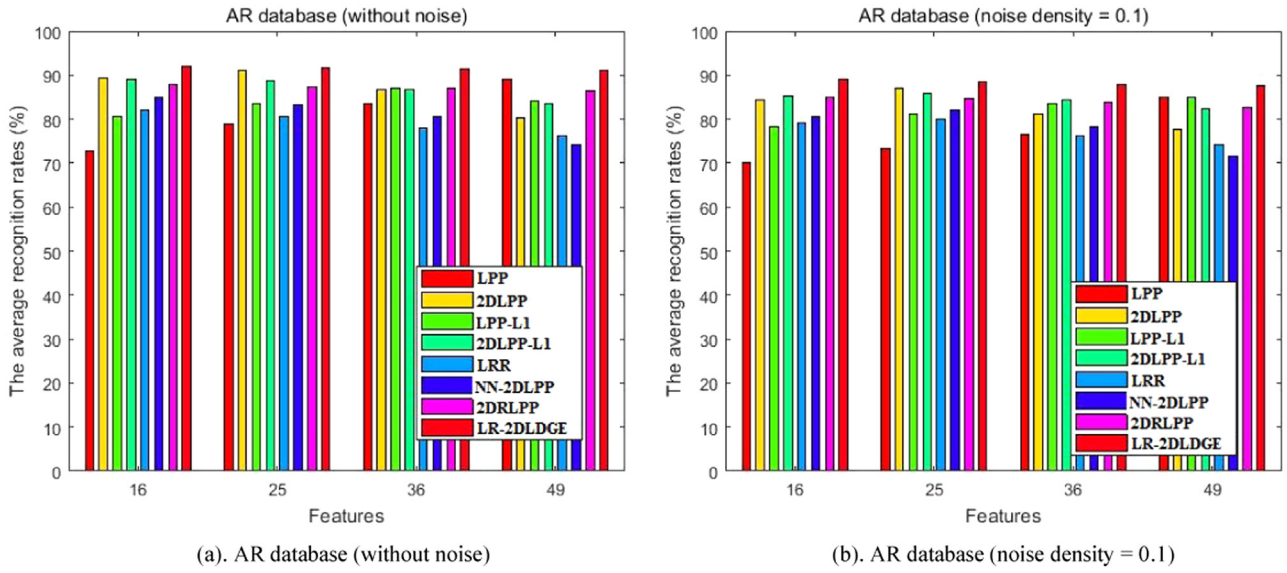
The RPCA algorithm, LRR algorithm, and LR-2DLDGE algorithm are robust image classification algorithms based on the nuclear norm. The RPCA algorithm and LR-2DLDGE algorithm do not consider the embedded manifold structure information and encode only the low-rank characteristics. The LR-2DLDGE algorithm preserves manifold structure information and lowers data rank. In addition, both the RPCA algorithm and LRR algorithm are based on a one-dimensional vector input algorithm, which may cause structural loss of the image information. The LR-2DLDGE algorithm can be directly inputted into a 2D image and retain the structural information of the image.

4. Experimation results

We will introduce several sets of experiments to verify the proposed LR-2DLDGE algorithm in this section. We compare the results of the algorithms with the LPP [12,13], 2DLPP [19], LPP-L_1 [38], 2DLPP-L_1 [37], LRR [22], 2DRLPP [21], and NN-2DLPP[20] algorithms on the FERET, ORL, coil100, Yale, AR, and PolyU databases. The experimental results are completed with an Intel i5 CPU with 16 GB of RAM.

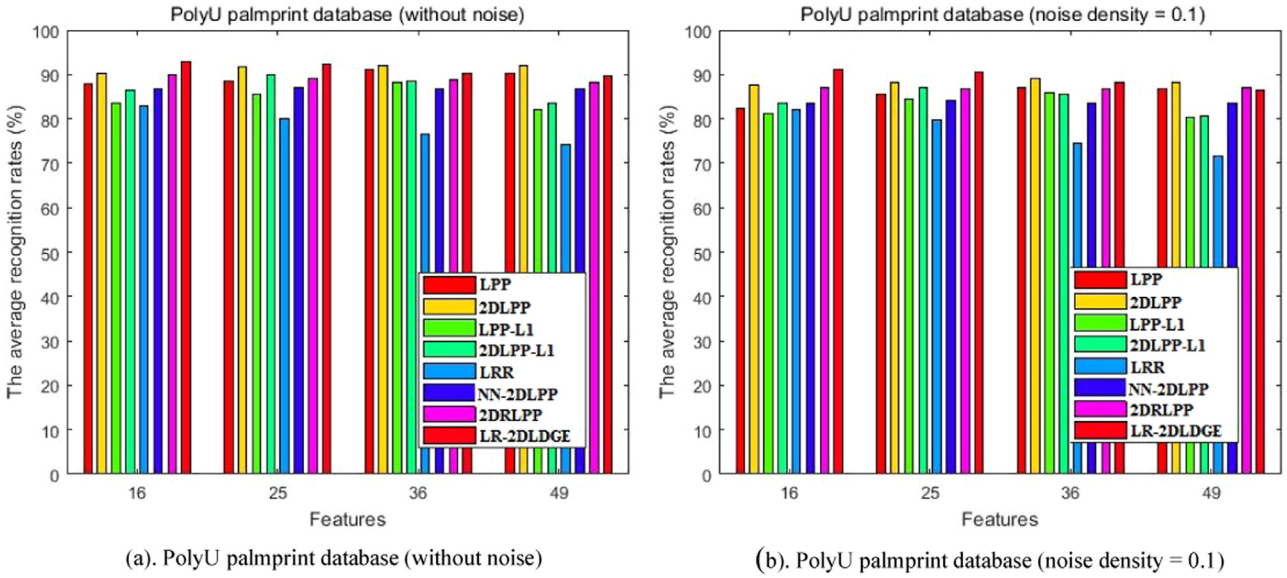
4.1. Database Descriptions

Fig. 2 shows some images in six different databases. The following describes six common image databases that will be used to evaluate the effectiveness and performance of LR-2DLDGE algorithm.



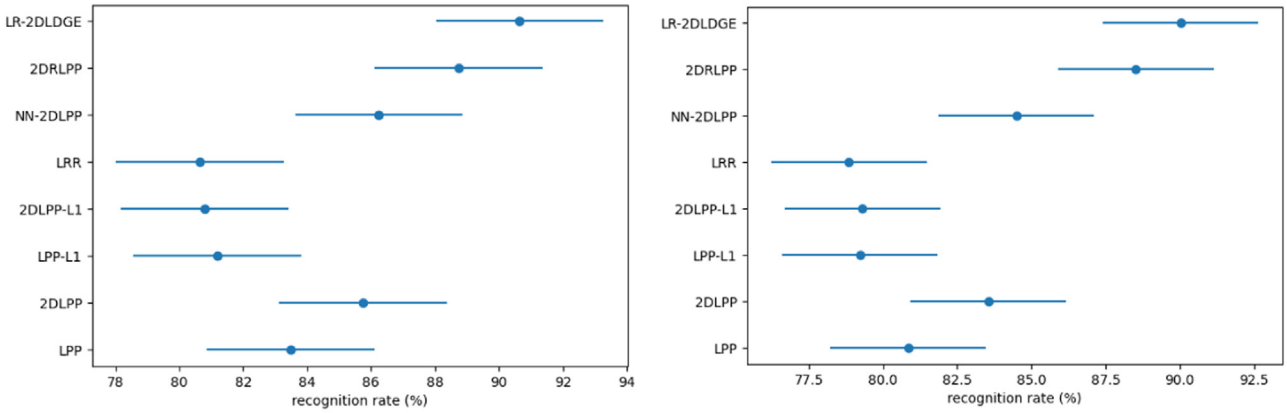
(a). AR database (without noise) (b). AR database (noise density = 0.1)

Fig. 7. The average recognition rates (%) and corresponding different dimension changes on the AR database



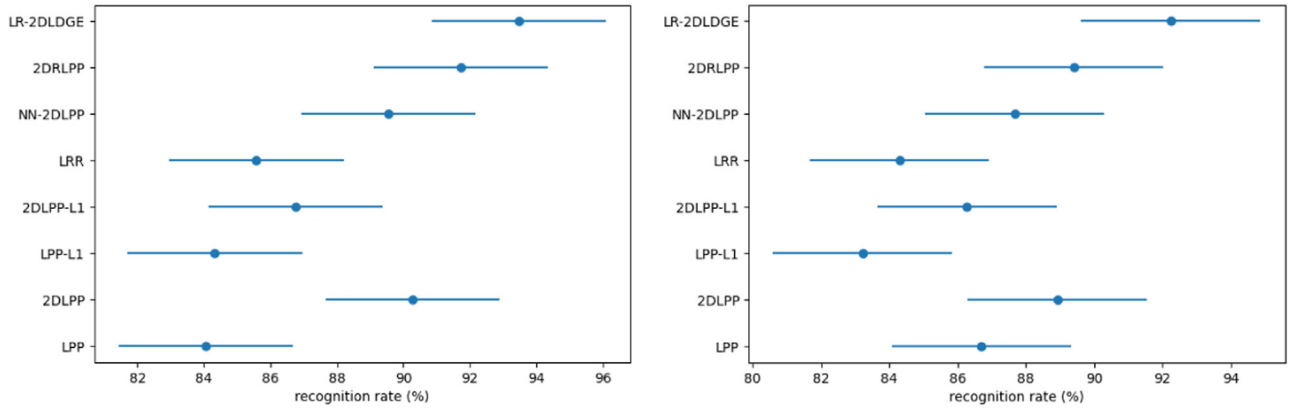
(a). PolyU palmprint database (without noise) (b). PolyU palmprint database (noise density = 0.1)

Fig. 8. The average recognition rates (%) and corresponding different dimension changes on the PolyU palmprint database



(a). FERET database (without occlusion) (b). FERET database (occlusion size=10×10)

Fig. 9. The Friedman test of the average recognition rates (%) on the FERET database



(a). Yale database (without occlusion)

(b). Yale database (occlusion size=10×10)

Fig. 10. The Friedman test of the average recognition rates (%) on the Yale database

4.1.1. FERET database

The FERET database is a standard face database, including different sexes, ethnicities, and ages. It is mainly used to verify the diversity of face recognition algorithms. In the experiment, 200 people were selected, with 7 images for each person, and a total of 1400 images were selected, including the changes in pose, illumination, and facial expression. All face images are adjusted to 40 × 40 pixels to reduce the computational burden.

4.1.2. ORL database

There are 40 classes, and each class includes 10 face images with different expressions, postures, and illumination on the ORL face database. All face images are adjusted to 56 × 46 pixels to reduce the computational burden.

4.1.3. COIL 100 database

The COIL 100 object database has 100 subjects, and each subject contains 72 images. To improve the computational efficiency, all original images are adjusted to 32 × 32 resolution pixels and converted into grey images

4.1.4. Yale database

The Yale face database includes 15 faces, which are used to show facial expression changes. In the Yale database, each class has 11 person, for a total of 165 greyscale images, which are adjusted to 50 × 40 pixels to reduce the computational burden.

4.1.5. AR database

The AR database contains 4000 colour images, including 70 men and 56 women, for a total of 126 people. These classes have different lighting conditions, facial expressions, and occlusion. All face images are adjusted to 50 × 40 pixels to reduce the computational burden.

4.1.6. PolyU palmprint database

PolyU contains 100 different palms, and each class has 6 images, for a total of 600 grey images. To improve the computational efficiency, each image is adjusted to 64 × 64 pixels in the PolyU database.

Experiments are carried out on different levels of random pixel corruptions and continuous occlusion data to test and verify the robustness of the algorithm, respectively. For continuous occlusion experiments, we randomly add blocks of different sizes to different positions of the image: 10 × 10 in the FERET, Yale, and COIL100 databases. The “salt & pepper” noise with a density of 0.1 was added to the images of the ORL, AR, and PolyU databases for corrosion experiments.

4.2. Baseline and Parameter Selection

In each experiment, we randomly choose T samples from each class, each experiment is repeated 10 times, and the NN classifier is used for classification. Our algorithm runs on six public datasets and we compare its results with other algorithms. The experimental operations of the following six databases are shown in Table 1.

In all experiments, we will set the maximum number of iterations equal to 500 in running the low-rank correlation algorithm. The local manifold graphs of LPP, 2DLPP, LPP-L_1, and 2DLPP-L_1 are constructed by the k-nearest neighbours ($k = T - 1$, where T is the training image from each class in the database), which are well gathered in the observation space [52]. Additionally, an iterative algorithm is used to solve the solutions of the L_1-norm algorithms. In the LRR 2DRLPP and NN-2DLPP models, we choose the parameters described in their references. The three parameters α, β, and γ in the LR-2DLDGE algorithm are selected from [0.001, 0.01, 0.1, 1, 10,100, 1000]. In this study, the classification accuracy is used to compare the performance of different algorithms. Classification accuracy is defined as $((N-T)\ \mathrm{cor}/(N-T)) \times 100\%$, where $(N-T)$cor is the number of test samples correctly classified by the nearest neighbor classifier, and $N-T$ is the total number of test samples.

4.3. Experiment results

In order to verify the robustness of the LR-2DLDGE method to occlusion (size=10 × 10) and noise (density = 0.1), Figs. 3 to 8 show the average recognition rates (%) of all methods with 4 different values (16, 25, 36 and 49) when the images shown in Table 1 are selected as training samples and the other images are selected as testing samples, respectively Figs. 4–8.

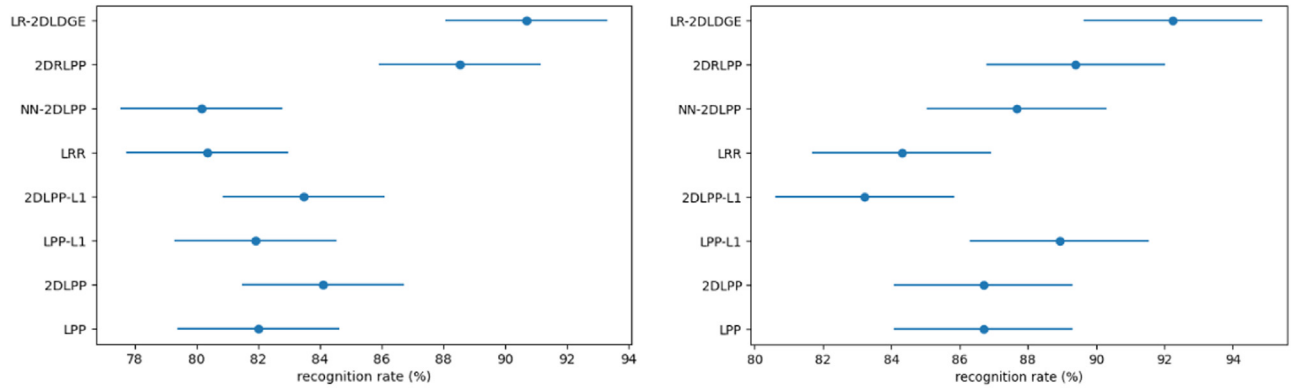
Figs. 3–8 the classification recognition rates (%) of the LPP, 2DLPP, LPP-L1, 2DLPP-L1, LRR, LRMD-S, and LR-2DLDGE algorithms and corresponding dimensions on the FERET, ORL, COIL 100, Yale, AR, and PolyU databases.

We can see from the curve changes in Fig. 3 to Fig. 8 that the recognition rate of the algorithm on six databases is higher than other algorithms and is not affected by the change of dimension.

In order to further verify the effectiveness of the LR-2DLDGE method proposed in this paper, we performed the Friedman test and Nemenyi post hoc test on six databases. The Friedman test is generally used to judge whether the performance of the algorithm is exactly the same. Obviously, the performance of each algorithm is different. Therefore, we use the Nemenyi test for the post hoc test to further demonstrate the advantages of the LR-2DLDGE

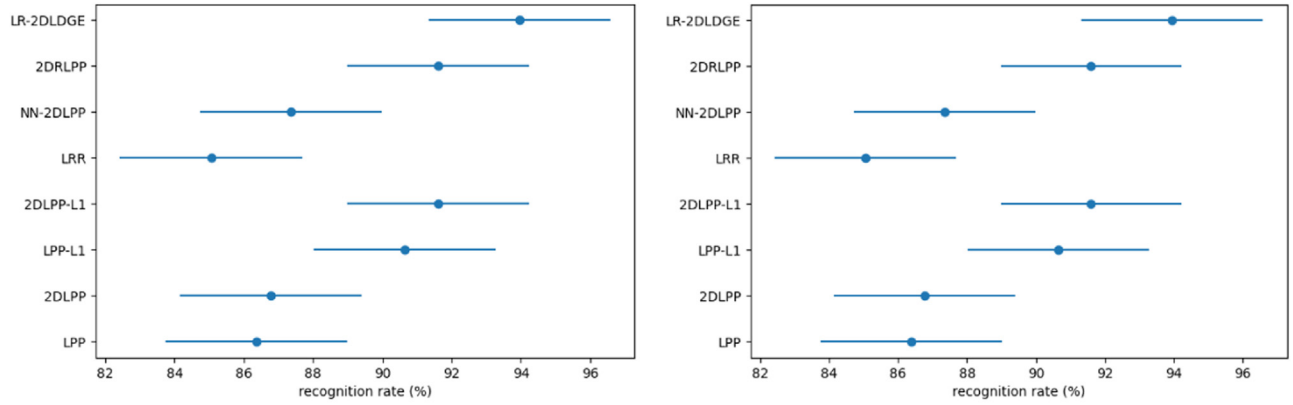
Table 1
The description of six different databases

Database	Sample size in each class	Dimension	Class number	Training size in each Class(T)
FERET	7	1600 (40 × 40)	200	5
Yale	11	2000 (50 × 40)	15	5
COIL 100	72	1024(32 × 32)	20	10
ORL	10	2576(56 × 46)	10	6
AR	14	2000 (50 × 40)	120	6
PolyU	6	4096(64 × 64)	100	3



(a). COIL20 database (without occlusion) (b). COIL20 database (occlusion size=10×10)

Fig. 11. The Friedman test of the average recognition rates (%) on the COIL20 database



(a). ORL database (without noise) (b). ORL database (noise density = 0.1)

Fig. 12. The Friedman test of the average recognition rates (%) on the ORL database

algorithm. $C = q_\alpha \sqrt{k(k+1)/6N}$ is the critical threshold, where $k = 8$ represents the number of algorithms, $N = 4$ represents the number of experimental dimensions, and the test level α is 0.05, so we can obtain $q_\alpha = 3.031$ and $C = 5.250$. The above test comparison is shown in Fig. 9 to Fig. 14, where for each algorithm, dots represent the mean value of the average recognition rate on different dimensions, and the horizontal line represents the value of the critical threshold C. Fig. 10, Fig. 11, Fig. 12, Fig. 13, Fig. 14

4.4. Maximal average recognition rates

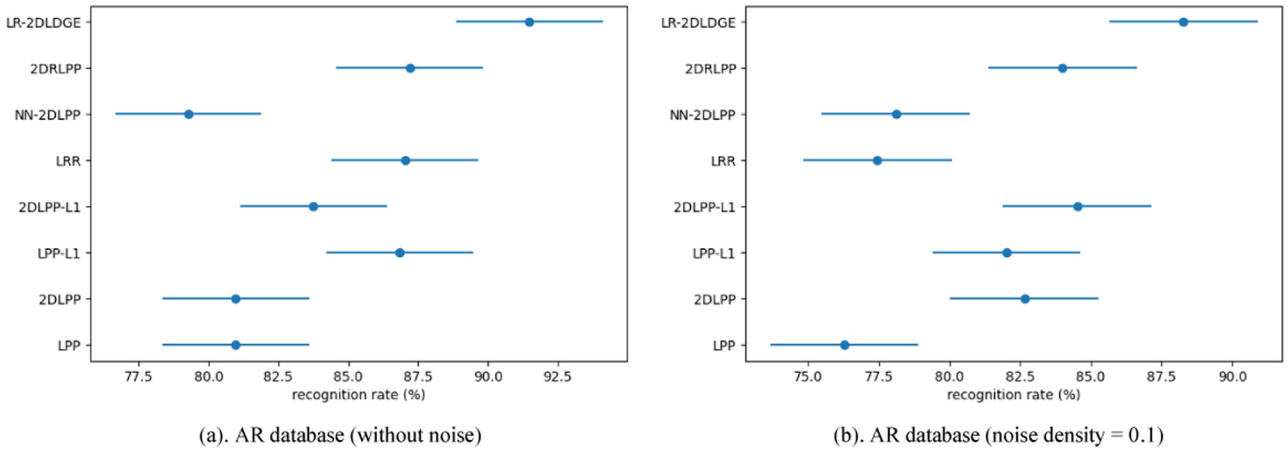
According to the settings in Table 1, when each class in the database randomly selects T sample points to form the training sample set, the highest average recognition rate (%) of each algorithm is shown in Table 2 and Table 3. The LR-2DLDGE algorithm has the highest maximum average recognition rate on different databases, indicating that our algorithm is robust in Table 2 and Table 3.

In this section, we run a set of experimental results and compare them with LPP, 2DLPP, LPP-L1, 2DLPP-L1, LRR, 2DRLPP and NN-2DLPP algorithms to evaluate the proposed LR-2DLDGE algorithm.

4.5. Convergence Study and computational time

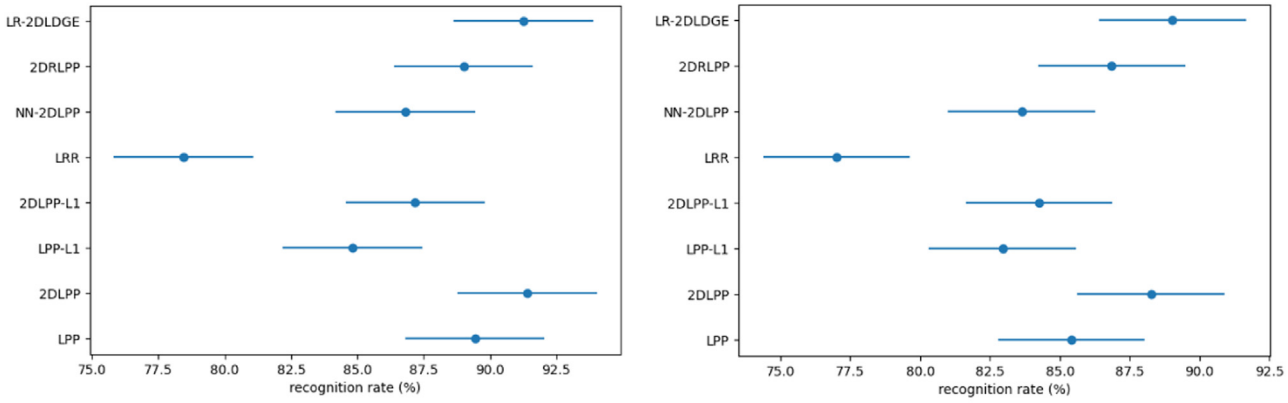
In the previous section, the theoretical analysis and objective function of LR-2DLDGE were proven and showed to converge to the local optimum. The convergence of the algorithm is further illustrated by experiments on the FERET and AR databases. As the number of iterations increases, the values of the objective function decrease monotonically, which shows that the proposed algorithms are convergent in Fig. 15 (a) and (b), respectively. It can be found that the proposed LR-2DLDGE can converge within 50 to 60 iterations.

In addition, when using training samples with T=600 from the PolyU palmprint database, Table 4 shows the highest average recognition rate (%), corresponding dimensions and average CPU time for each method.



(a). AR database (without noise) (b). AR database (noise density = 0.1)

Fig. 13. The Friedman test of the average recognition rates (%) on the AR database



(a). PolyU palmprint database (without noise) (b). PolyU palmprint database (noise density = 0.1)

Fig. 14. The Friedman test of the average recognition rates (%) on the PolyU palmprint database

Table 2
The highest average recognition rate (%) of different algorithms on the FERET, Yale and COIL100 databases and corresponding dimension (D).

Database							
algorithm		without occlusion			occlusion size=10 × 10		
		FERET	Yale	COIL20	FERET	Yale	COIL100
LPP	(%)	87.16	89.50	86.69	84.35	87.36	85.90
	(D)	25	18	20	26	20	24
2DLPP	(%)	89.28	93.11	88.35	86.76	90.55	86.86
	(D)	40 × 16	50 × 14	32 × 20	40 × 18	50 × 16	32 × 22
LPP-L_1	(%)	89.28	93.11	88.35	82.18	85.87	81.65
	(D)	28	22	26	30	24	28
2DLPP-L_1	(%)	85.42	88.36	84.35	83.82	87.58	82.36
	(D)	40 × 12	50 × 14	32 × 18	40 × 16	50 × 16	32 × 20
LRR	(%)	82.65	86.85	81.20	80.63	85.25	79.95
	(D)	32	16	28	36	18	32
NN-2DLPP	(%)	87.50	90.30	85.28	85.70	88.85	83.25
	(D)	18	16	14	20	20	18
2DRLPP	(%)	91.25	92.59	89.56	89.32	90.86	87.82
	(D)	40 × 18	50 × 26	32 × 20	40 × 22	50 × 30	32 × 20
LR-2DLDGE	(%)	93.76	95.26	91.45	90.16	93.68	89.28
	(D)	40 × 16	50 × 20	32 × 18	40 × 20	50 × 18	32 × 16

D: dimension

Table 3
The highest average recognition rate (%) of different algorithms on the ORL, AR, and PolyU databases and corresponding dimension (D).

Database							
algorithm		without noise			noise density = 0.1		
		ORL	AR	PolyU	ORL	AR	PolyU
LPP	(%)	90.31	88.93	89.08	89.65	85.43	87.75
	(D)	32	40	36	36	40	40
2DLPP	(%)	91.18	90.35	91.88	90.56	87.39	88.86
	(D)	56 × 26	50 × 24	64 × 30	56 × 30	50 × 28	64 × 36
LPP-L_1	(%)	91.87	89.36	90.56	89.08	86.45	88.82
	(D)	36	45	38	32	45	36
2DLPP-L_1	(%)	92.98	90.16	91.08	90.66	88.45	89.56
	(D)	56 × 20	50 × 18	64 × 24	56 × 26	50 × 22	64 × 28
LRR	(%)	87.65	82.96	83.25	85.36	80.40	82.82
	(D)	36	18	16	40	20	20
NN-2DLPP	(%)	90.18	85.40	87.32	88.90	82.28	84.88
	(D)	20	18	2	26	28	22
2DRLPP	(%)	94.15	91.96	92.84	91.65	88.20	90.45
	(D)	56 × 12	50 × 16	64 × 10	56 × 14	50 × 14	64 × 12
LR-2DLDGE	(%)	95.32	92.06	93.87	92.56	89.88	91.27
	(D)	56 × 10	50 × 12	64 × 14	56 × 16	50 × 16	64 × 12

D: dimension

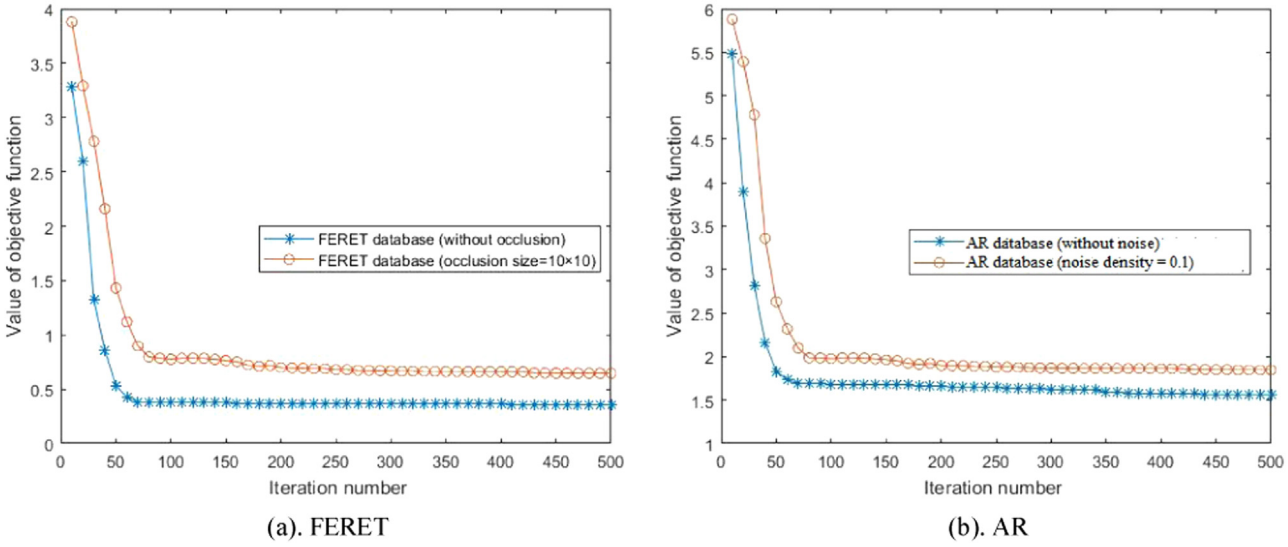


Fig. 15. Convergence of LR-2DLDGE algorithm on the FERET and AR image database. (a) FERET. (b) AR.

Table 4
The highest average recognition rate (%), corresponding dimensions and average CPU time consumed for each method, when T=600 training samples from the PolyU palm print database.

PolyU algorithm	LPP	2DLPP	LPP-L_1	2DLPP-L_1	LRR	NN-2DLPP	2DRLPP	LR-2DLDGE
nois(%)	87.75	88.86	88.82	89.56	82.82	84.88	90.45	91.27
den(D)	40	64 × 36	36	64 × 28	20	22	64 × 12	64 × 12
sity(S)0.1	0.981	0.818	1.236	0.852	0.765	0.838	0.689	0.5412

5. Discussion

By analyzing the above experimental results, we can draw the following conclusions:

[1] Considering the experimental results in Figs. 3 to 8, the average recognition rates (%) of LR-2DLDGE are significantly higher than other algorithms under different dimension changes. The results show that the LR-2DLDGE algorithm can obtain more distinctive and sparse features than other algorithms, because we fused the discriminant information in GE and the low-rank properties of the data.

[2] Considering the experimental results of Tables 2–3, with the increase of the number of training samples, the highest average recognition rates (%) of LR-2DLDGE are always higher than other algorithms. The results show that the proposed algorithm preserve the manifold structure which is more discriminative than other algorithms, which can obtain more discriminative information for effective feature selection and extraction.

[3] Considering the experimental results in Figs. 9 to 14, the performance of each algorithm is apparently different. Therefore, we use the Nemenyi test for the post-hoc test to further reflect the advantages of the LR-2DLDGE algorithm, which can obtain

low-rank property of the projection and simultaneously maintain the neighborhood structure of data.

[4] Considering the experimental results of the convergence of the LR-2DLDGE algorithm on the FERET and AR images in Fig. 15, the values of the objective function decrease monotonically with an increasing number of iterations. From the experimental results in Table 4, the average CPU time consumed by the LR-2DLDGE algorithm is always lower than that of the other seven algorithms on the PolyU palmprint database. This shows that our algorithm is efficient.

6. Conclusion

This paper mainly combines the 2DLPP algorithm with low-rank representation learning and proposes the low-rank two-dimensional local discriminant graph embedding (LR-2DLDGE) algorithm based on low-rank sparsity and graph embedding. Local structure information and low-rank representation exist simultaneously in the algorithm. In addition, the algorithm makes the data points as distinct as possible from different classes. In particular, the algorithm uses the L_1-norm as a constraint to reduce the influence of noise and corruption. The optimal solution of LR-2DLDGE algorithm can be obtained by using the ADMM. At the same time, the convergence and complexity are analyzed. The experiments on six public databases show that the LR-2DLDGE algorithm can obtain better recognition performance than other methods. However, in many applications such as contiguous occlusions, random pixel corruptions, noises and errors in data, the performance of the proposed algorithm faces challenges. In the future work, we will combine kernel and tensor forms to improve the discrimination ability and robustness evaluation of LR-2DLDGE algorithm. Besides, parameter selection is a trivial task. So, it is necessary to further study how to determine the optimal parameters, such as parameter α, β, γ, K_c and K_b, and perform more tests on other large databases.

Declaration of Competing Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

This work is partially supported by the National Science Foundation of China under Grant Nos. 61876213, 62172229, 6177227, 61976117, 71972102, 61861033, 61976118, the Key R&D Program Science Foundation in Colleges and Universities of Jiangshu Province Grant Nos.18KJA520005, 19KJA360001, 20KJA520002, the Natural Science Fund of Jiangsu Province under Grants Nos. BK20201397, BK20191409,BK20211295. Jiangsu Key Laboratory of Image and Video Understanding for Social Safety of Nanjing University of Science and Technology under Grants J2021-4. Funded by the Qing Lan Project of Jiangsu University, Funded by the Future Network Scientific Research Fund Project SRFP-2021-YB-25, and China's Jiangxi Province Natural Science Foundation (No. 20202ACBL202007). This work is funded in part by the "Qinglan Project" of Jiangsu Universities under Grant D202062032.

References

[1] C Zhao, X Wang, W Zuo, et al., Similarity learning with joint transfer constraints for person re-identification[J], Pattern Recognit. 97 (2020) 107014.
[2] Qiaolin Ye, Zechao Li, Liyong Fu, et al., Nonpeaked Discriminant Analysis, IEEE Trans. Neur. Netw. Learn. Syst. 30 (12) (2019) 3818–3832.
[3] Y Duan, H Huang, Z Li, et al., Local Manifold-Based Sparse Discriminant Learning for Feature Extraction of Hyperspectral Image[J], IEEE Trans Cybern (99) (2020) 1–14.
[4] A Hy, A Jl, A Lll, et al., Low-rank matrix regression for image feature extraction and feature selection - ScienceDirect[J], Information Sciences 522 (2020) 214–226.
[5] J Lu, H Wang, J Zhou, et al., Low-rank adaptive graph embedding for unsupervised feature extraction[J], Pattern Recognit. (2020) 107758.
[6] C Zhao, X Wang, D Miao, et al., Maximal granularity structure and generalized multi-view discriminant analysis for person re-identification[J], Pattern Recognit. 79 (2018) 79–96.
[7] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput 15 (6) (Jun. 2003) 1373–1396.
[8] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[9] S. Roweis, L. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[10] Y. Bengio, et al., Out-of-sample extensions for LLE, isomap, MDS, eigenmaps, and spectral clustering, in: Proc. Adv. Neural Inf. Process. Syst., Vancouver, BC, Canada, 2004, pp. 1–8.
[11] S. Yan, D. Xu, B. Zhang, H. 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.
[12] J Zhou, W Pedrycz, X Yue, et al., Projected fuzzy C-means clustering with locality preservation[J], Pattern Recognit. 113 (6) (2020) 107748.
[13] X. He, S. Yan, Y. Hu, P. Niyogi, H.J. Zhang, Face recognition using laplacianfaces, IEEE Trans. Pattern Analysis and Machine Intelligence 27 (3) (2005) 328–340.
[14] Y. Pang, N. Yu, H. Li, R. Zhong, Z. Liu, Face recognition using neighborhood preserving projections, in: Advances in Multimedia Information Processing (LNCS 3768), Springer, Berlin, Germany, 2005, pp. 854–864.
[15] G Wang, N Shi, Collaborative representation-based discriminant neighborhood projections for face recognition[J], Neural Comput Appl 32 (10) (2020) 5815–5832.
[16] Y Zhang, D Ye, Y. Liu, Robust locally linear embedding algorithm for machinery fault diagnosis[J], Neurocomputing 273 (2018) 323–332.
[17] X. He, D. Cai, S. Yan, H.J. Zhang, Neighborhood preserving embedding, in: Proc. Int. Conf. Comput. Vis. (ICCV), Beijing, China, 2005, pp. 1208–1213.
[18] X. He, D. Cai, P. Niyogi, Tensor Subspace Analysis, Adv Neural Inf Process Syst 18 (2005) Vancouver, Canada.
[19] Yang Ben, Pal Shiu, Two-dimensional Laplacianfaces algorithm for face recognition, Pattern Recognit. 41 (10) (2008) 3237–3243.
[20] Y Lu, C Yuan, Z Lai, et al., Nuclear Norm-Based 2DLPP for Image Classification[J], IEEE Trans. Multimedia 19 (11) (2017) 2391–2403.
[21] W J Chen, C N Li, Y H Shao, et al., 2DRLPP: Robust two-dimensional locality preserving projection with regularization[J], Knowledge Based Systems 169 (APR.1) (2019) 53–66.
[22] F. Zhang, J. Yang, J. Qian, Y. Xu, Nuclear norm-based 2-DPCA for extracting features from images, IEEE Trans Neural Netw Learn Syst 26 (10) (Oct. 2015) 2247–2260.
[23] P. Zhou, Z. Lin, C. Zhang, Integrated low-rank-based discriminative feature learning for recognition, IEEE Trans Neural Netw Learn Syst 27 (5) (May 2016) 1080–1093.
[24] E.J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis? J. ACM 58 (3) (May 2011) 1–37.
[25] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, Y. Ma, Robust recovery of subspace structures by low-rank representation, IEEE Trans Pattern Anal Mach Intell 35 (1) (Jan. 2013) 171–184.
[26] E.J. Candès, X. Li, Y. Ma, J. Wright, Robust principal component analysis, J. ACM 58 (3) (2009) 1–17.
[27] J. Liu, Y. Chen, J. Zhang, Z. Xu, Enhancing low-rank subspace clustering by manifold regularization, IEEE Trans Image Process 23 (9) (Sep. 2014) 4022–4030.
[28] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans Pattern Anal Mach Intell 31 (2) (Feb. 2009) 210–227.
[29] D. Meng, Q. Zhao, Z. Xu, Improve robustness of sparse PCA by L1-norm maximization, Pattern Recognit 45 (1) (2012) 487–497.
[30] N. Kwak, Principal component analysis based on L1-norm maximization, IEEE Trans Pattern Anal Mach Intell 30 (9) (Sep. 2008) 1672–1680.
[31] C. Ding, D. Zhou, X. He, H. Zha, R1-PCA: Rotational invariant L1-norm principal component analysis for robust subspace factorization, in: Proc. 23rd Int. Conf. Mach. Learn, 2006, pp. 281–288.
[32] F. Nie, H. Wang, C.H. Ding, D. Luo, H. Huang, Robust principal component analysis with non-Greedy l1-Norm maximization, in: Proc. Int. Joint Conf. Artif. Intell., 2011, pp. 1433–1438.
[33] Y. Pang, X. Li, Y. Yuan, Robust tensor analysis with L1-norm, IEEE Trans Circuits Syst Video Technol 20 (2) (Feb. 2010) 172–178.
[34] P.P. Markopoulos, G.N. Karystinos, D.A. Pados, Optimal algorithms for L1-subspace signal processing, IEEE Trans. Acoust., Speech, Signal Process. 62 (19) (Oct. 2014) 5046–5058.
[35] X. Chen, J. Yang, Z. Jin, An improved linear discriminant analysis with L1-norm for robust feature extraction, in: Proc. Int. Conf. Pattern Recognit., Stockholm, Sweden, 2014, pp. 1585–1590.
[36] X. Li, Y. Pang, Y. Yuan, L1-norm-based 2DPCA, IEEE Trans Syst Man Cybern B Cybern 40 (4) (Aug. 2009) 1170–1175.
[37] Y. Pang, Y. Yuan, Outlier-resisting graph embedding, Neurocomputing 73 (4–6) (2010) 968–974.

[38] H.X. Zhao, H.J. Xing, X.Z. Wang, J.F. Chen, L1-norm-based 2DLPP, in: Proc. Chinese Conf. Control Decis, 2011, pp. 1259–1264.
[39] Y. Tang, Z. Zhang, Y. Zhang, F. Li, Robust L1-norm matrixed locality preserving projection for discriminative subspace learning, in: Proc. Int. Joint Conf. Neural Netw, 2016, pp. 4199–4204.
[40] J O Agushaka, A E Ezugwu, L. Abualigah, Dwarf mongoose optimization algorithm[J], Comput. Meth. Appl. Mech. Eng. 391 (2022) 114570.
[41] L Abualigah, D Yousri, M Abd Elaziz, et al., Aquila optimizer: a novel meta-heuristic optimization algorithm[J], Comput. Ind. Eng. 157 (2021) 107250.
[42] L Abualigah, M Abd Elaziz, P Sumari, et al., Reptile Search Algorithm (RSA): A nature-inspired meta-heuristic optimizer[J], Expert Syst. Appl. 191 (2022) 116158.
[43] O N Oyelade, A E S Ezugwu, T I A Mohamed, et al., Ebola optimization search algorithm: A new nature-inspired metaheuristic optimization algorithm[J], IEEE Access 10 (2022) 16150–16177.
[44] L Abualigah, A Diabat, S Mirjalili, et al., The arithmetic optimization algorithm[J], Comput. Meth. Appl. Mech. Eng. 376 (2021) 113609.
[45] L Abualigah, A Diabat, P Sumari, et al., Applications, deployments, and integration of internet of drones (iod): a review[J], IEEE Sensors J. (2021).
[46] Minghua Wan, Ming Li, Guowei Yang, Shan Gai, Zhong Jin, Feature extraction using two-dimensional maximum embedding difference, Inf. Sci. 274 (2014) 55–69.
[47] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier algorithm for exact recovery of corrupted low-rank matrices," Univ. Illinois at Urbana–Champaign, Champaign, IL, USA, Rep. UILU-ENG-09-2215, 2009.
[48] J.-F. Cai, E.J. Candès, Z. Shen, A singular value thresholding algorithm for matrix completion, SIAM J Optim 20 (4) (2010) 1956–1982.
[49] P. Drineas, A. Frieze, R. Kannan, S. Vempala, V. Vinay, Clustering large graphs via the singular value decomposition, Int. J. Mach. Learn. Cybern. 56 (1) (Jul. 2004) 9–33.
[50] E. Hosseini-Asl, J.M. Zurada, L. Nasraoui, Deep learning of part-based representation of data using sparse autoencoders with nonnegativity constraints, IEEE Trans Neural Netw Learn Syst 27 (12) (Dec. 2016) 2486–2498.
[51] J. Eckstein, D.P. Bertsekas, On the Douglas–Rachford splitting algorithm and the proximal point algorithm for maximal monotone operators, Math. Program. 55 (3) (Jun. 1992) 293–318.
[52] J. Yang, D. Zhang, J. Yang, B. Niu, Globally maximizing, locally minimizing: unsupervised discriminant projection with applications to face and palm biometrics, IEEE Trans Pattern Anal Mach Intell 29 (4) (2007) 650–664.

Minghua Wan received the B.S. degree in automated institute from the Nanchang University of Aviation in 2003, the M.S. and PhD degrees in pattern recognition and intelligent systems from the Nanjing University of Science and Technology (NUST) in 2007 and 2011, respectively. He is the author of more than 50 scientific papers in pattern recognition and computer vision. He is a professor at the Nanjing Audit University. His current research interests include face recognition and detection, and image processing.

Xueyu Chen received the B.S. degree from the Nanjing Audit University, Nanjing, China, in 2020. She is currently M.S. candidate with the school of Information and Engineering, Nanjing Audit University. Her research interests include pattern recognition and image processing.

Tianming Zhan received the B.S. degree and M. S. degree from the School of Math and Statistics, Nanjing University of Information Science and Technology, Nanjing, China, in 2006 and 2009, respectively, and the Ph. D degree with the School of Computer Science and Engineering, Nanjing University of Science and Technology, in 2013. He is currently an associated professor with the School of Information and Engineering, Nanjing Audit University. His research interests include medical image processing, hyperspectral image processing, machine learning, and data

Guowei Yang received the B.S. and the M.S. degrees in mathematics from the Jiangxi Normal University, Nanchang, China, in 1985 and 1988, and the Ph.D. degree from the University of Science and Technology Beijing, China, in 2004. He was appointed as Professor in the Qingdao University, China, in 1999. Now, he is a professor in the School of Technology of Nanjing Audit University (NAU). His current research interests include artificial intelligence, artificial life, artificial neural network, pattern recognition, innovative and creative design, etc.

Hai Tan received his M.E. degree in the application of computer from Taiyuan University of Technology, Taiyuan, China, in 2003. His PhD degree was received from Beijing Institute of Technology. He is a professor in the School of Mechanical and Electronic Engineering at East China University of Technology, located in Nanchang, China. His current research interests include many-core architecture, big data computing, artificial intelligence, and Data processing. He is a member of CCF and ACM.

Hao Zheng received the B.S. degree from Southeast University, in 1998, the M.S. degree from Nanjing University Posts and Telecommunications, in 2005, and the Ph.D. degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology, in 2013. He visited the QCIS of University of Technology Sydney, Australia, from September 2003 to March 2004. He also visited the ITC of The Hong Kong Polytechnic University, Hong Kong, from July 2018 to January 2019. He is currently a Postdoctoral with Southeast University. He is also a Professor with the College of Information Engineering, Nanjing Xiaozhuang University (NJXZC). His research interests include pattern recognition, image processing, face recognition, facial expression recognition, and computer vision.