An Orthogonal Locality and Globality Dimensionality Reduction Method Based on Twin Eigen Decomposition

SHUZHI SU\textsuperscript{1,2}, GANG ZHU\textsuperscript{1}, AND YANMIN ZHU\textsuperscript{3}

\textsuperscript{1}School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China
\textsuperscript{2}Institute of Energy, Hefei Comprehensive National Science Center, Hefei 230031, China
\textsuperscript{3}School of Mechanical Engineering, Anhui University of Science and Technology, Huainan 232001, China

Corresponding author: Yanmin Zhu (zyanmin1988@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61806006, in part by the China Postdoctoral Science Foundation under Grant 2019M660149, and in part by the Institute of Energy, Hefei Comprehensive National Science Center under Grant 19KZS203.

ABSTRACT Dimensionality reduction is a hot research topic in pattern recognition. Traditional dimensionality reduction methods can be separated into linear dimensionality reduction methods and nonlinear dimensionality reduction methods. Linear dimensionality reduction methods usually utilize Euclidean distances to explore global geometric structures, and nonlinear dimensionality reduction methods can preserve local manifold structures in learned low-dimensional subspaces. However, redundant information and noises of the raw high-dimensional data restrict the classification performance of these methods. To solve the problem, we propose a novel orthogonal dimensionality reduction method based on twin eigen decomposition called orthogonal locality and globality preserving projections (OLGPP). Orthogonality, as a commonly used criterion in pattern recognition, is insensitive to the redundant information and the noises. OLGPP not only combines the advantages of global Euclidean structures and local manifold structures, but also is insensitive to the redundant information and noises of the raw high-dimensional data by embedding the orthogonality criterion. Additionally, the objective function based on twin eigen decomposition can be solved sequentially to obtain analytical solutions. On four well-known datasets, we design extensive experiments to evaluate the performance of OLGPP. From the experimental results, OLGPP has the optimal experimental results, and its average recognition rates are about 10\% higher than the classic local dimensionality reduction method (i.e., locality preserving projection), which proves OLGPP is an effective dimensionality reduction method.

INDEX TERMS Dimensionality reduction, orthogonality criterion, global Euclidean structures, local manifold structures.

I. INTRODUCTION

With the rapid development of electronic information technology, the rapid increase of data generation and storage methods has led to the accumulation of massive data in many fields such as social media, knowledge graph, intelligent search engine and so on. Generally speaking, these data are usually accompanied by high-dimensional problems and complex structures [1]. If tasks such as classification and clustering are directly performed on these data, not only will tasks pay huge computational cost and storage cost, but also the curse of dimensionality [2] will appear. Especially in some fields of pattern recognition [3], machine learning [4] and neural networks [5], [6], how to deal with the massive data has become a problem with practical application significance. At present, dimensionality reduction (DR) [7], [8] is one of the effective methods to solve this problem. DR aims to reduce redundant information of raw high-dimensional data and retain effective information of the data as much as possible. And the complexity and dimensionality of raw high-dimensional data can be decreased while retaining intrinsic feature information by using DR methods.

In practical applications, principal component analysis (PCA) [9] and linear discriminant analysis (LDA) [10] are two representative linear DR methods. PCA attempts to find the optimal projection directions by maximizing the variance
matrix of samples. However, PCA does not consider class label information [11] of samples, which makes it different for PCA to perform well on classification tasks in most cases. Unlike PCA, LDA takes into accounts the class label information. LDA seeks to learn a set of projection directions by maximizing the inter-class scatter matrix and simultaneously minimizing the intra-class scatter matrix. The research results from subspace linear discriminant analysis (SLDA) [12] demonstrates that LDA exhibits favorable inter-class separability and intra-class aggregation in the learned projection subspace. When the complexity and dimensionality of data increase, the distribution rule of data is often nonlinear and more feature information are hidden in local nonlinear structures as pointed out in [13]. PCA and LDA usually play a significant role in some fields such as sentiment analysis [14], image segmentation [15] and memristive neural networks [16]. However, PCA and LDA only preserve the global linear structures in the learned projection subspace, which leads to poor classification performance.

To make up for the deficiency of linear DR methods, scholars had proposed many nonlinear DR methods, for instance, local linear embedding (LLE) [17], laplacian eigenmaps (LE) [18], multi-dimensional scaling (MDS) [19], isometric mapping (Isomap) [20], hessian LLE (HLLE) [21] and so on. Among them, LLE and Isomap are the most classical nonlinear DR methods. To achieve the purpose of retaining the sample manifold structures in the local neighborhood, LLE constructs a local reconstruction weight matrix of sample points and minimizes a reconstruction error in the learned projection subspace. Isomap is a method derived from MDS, which preserves the intrinsic geometric structures information of local manifold structures in the process of DR. Specifically speaking, the detailed ideas are as follows: (1) firstly, setting k-nearest neighbor parameter and constructing the nearest neighbor connection graph; (2) secondly, using the shortest path algorithm (Dijkstra algorithm [22] or Floyd algorithm [23]) to calculate the shortest path between two points on the graph; (3) finally, utilizing MDS method to obtain the projections in the learned projection subspace. In addition, since the shortest path algorithm will cause a huge computational cost, parallel technology [24], [25] can be used to improve the computational burden for DR. Unfortunately, the out-of-sample problem [26] exists in these methods. When the methods are applied to new data, these methods will have weak performance. To overcome this shortcoming, locality preserving projection (LPP) [27] and neighborhood preserving embedding (NPE) [28] are proposed based on the idea of traditional methods. There are some similarities between LPP and NPE. Both of them attempt to discover local nonlinear manifolds hidden in raw high-dimensional data. The differences are as follows: LPP constructs a Laplacian graph to keep the local structures information to find the projection directions in the learned projection subspace; NPE obtains the corresponding projection direction by calculating the edge weight between samples to retain the local structures information. Furthermore, LPP and NPE suffer from the small sample size (SSS) problem [29], which makes inferior performance. To avoid the interference of the SSS problem, kernel LPP (KLPP) [30] and kernel NPE (KNPE) [31] were proposed based on kernel technology. Many experimental results indicate that KLPP and KNPE have better classification performance compared with LPP and NPE. Besides, sample-dependent LPP (SLPP) [32] utilizes the similarities between sample pairs to settle the SSS problem, which is also a feasible method.

Combining the advantages of global Euclidean structures and local manifold structures, Zang et al. [33] proposed elastic preserving projections (EPP). EPP can preserve the global geometry structures and the local manifold structures into the learned projection subspace. However, due to the complexity of data distribution and the influence of noises, EPP will be difficult to fully explore the local manifold structures. In pattern recognition, orthogonality is a commonly used criterion. As [34], [35] pointed out, orthogonality criterion can be less sensitive to the complexity of data distribution and the influence of noises. And the orthogonality criterion can also enhance the extraction ability of local manifold structures. In this paper, to overcome the effect of noises and the complexity of data distribution and obtain more discriminative low-dimensional features, we propose a novel DR method named orthogonal locality and globality preserving projections (OLGPP). By constructing local undirected neighborhood graph and global geometric graph, OLGPP effectively utilizes the local manifold structures and the global Euclidean structures. Meanwhile, OLGPP optimizes the ratio between two structures so that the data mapped in the learned projection subspace will have more effective and comprehensive intrinsic feature information. By introducing the orthogonality constraints, OLGPP can have favorable discriminative power and can obtain more intrinsic local structures of the raw data. Experiments are designed to estimate the effectiveness of OLGPP on several real-world datasets, including Semeion handwritten digit dataset, COIL-20 object image dataset, ORL face image dataset and UMIST face dataset. The experimental results demonstrate that OLGPP always has better recognition ability under the same conditions.

The rest of this paper is as follows. Section II reviews some related methods. The next section introduces the construction and optimization of OLGPP in detail. Then we discuss the singularity problem in section IV. Section V presents and analyzes the experimental results. The paper is summarized in the final section.

**II. REVIEW OF RELATED WORK**

Supposing that \( X = \{x_1, x_2, x_3, \ldots, x_n\} \in \mathbb{R}^{m \times n} \) consists of \( n \) samples, where \( x_i (i = 1, 2, 3, \ldots, n) \) is the m-dimensional column vector and each column vector represents a sample. The main goal of DR methods is to find an optimal mapping: \( \mathbb{R}^{m} \rightarrow \mathbb{R}^{d (d < m)} \) for preserving the raw high-dimensional data information structures in the learned projection subspace as much as possible. Generally speaking, we can further describe this mapping as \( y_i = P^T x_i \), \( P \in \mathbb{R}^{m \times d} \).
generalized eigenvalue problem can be given:

\[ \sum_{i=1}^{n} x_i = 0. \]

PCA, as a mainstream DR method, has been widely used in data processing. Assuming that the sample set \( X \) is centralized, i.e., \( \sum_{i=1}^{n} x_i = 0 \). PCA intends to find a set of projection directions \( P = [P_1, P_2, P_3, \ldots, P_d] \) maximizing the variance between the samples. From the theory of PCA, the following formulation can be obtained:

\[
\max \quad tr(P^T XX^T P) \\
\text{s.t.} \quad P^T P = I
\]  

Using the Lagrange multiplier method [36], the following generalized eigenvalue problem can be given:

\[ XX^T P = \lambda P \]  

By solving (2), the projection directions \( P \) can be given by \( P = [P_1, P_2, P_3, \ldots, P_d] \).

### B. LOCALITY PRESERVING PROJECTION

LPP is a classical linear DR method based on the \( k \)-nearest neighbor method. The local graph is further constructed to retain the local manifold structures of the raw high-dimensional data. The corresponding nearest neighbor weight matrix is as follows:

\[ S_{ij} = \begin{cases} 
\exp \left( -\frac{\|x_i - x_j\|^2}{2\tau^2} \right) & \text{if } x_i \in N_k(x_j) \\
0 & \text{otherwise} \\
\end{cases} \]  

where \( N_k(x_j) \) (or \( N_k(x_i) \)) means \( x_j \) (or \( x_i \)) is a one of the \( k \)-nearest neighbors of \( x_i \) (or \( x_j \)) and \( \tau \in (0, +\infty) \) is a kernel parameter. The objective function of LPP can be given as follows:

\[ \min \sum_{i,j} \|y_i - y_j\|^2 S_{ij} \]  

By substituting \( y_j = P^T x_j \) in (4), we can obtain further optimization criterion for LPP:

\[
\min_p \quad \sum_{i,j} \|P^T x_i - P^T x_j\|^2 S_{ij} \\
= \min_p \quad tr(P^T XDX^T P) - P^T XSX^T P) \\
= \min_p \quad tr(P^T XLX^T P) \]  

subject to

\[ tr(P^T XDX^T P) = I \]

where \( S \) is the nearest neighbor weight matrix, \( D \) is the diagonal matrix, and \( L = D - S \) is a Laplacian matrix. Each element of \( D \) corresponds to the sum of each row vector of \( S \). The optimization criterion of LPP can be turned into the following solvable eigenvalue problem by using the Lagrange multiplier method

\[ XLX^T P = \lambda XDX^T P \]  

When the equivalent eigenvalue problem is solved, the eigenvectors corresponding to the first \( d \) largest eigenvalues can be obtained and corresponding low-dimensional features \( Y \) can be defined as \( Y = [y_1, y_2, y_3, \ldots, y_n] \), where \( y_i = P^T x_i \) (i = 1, 2, 3, \ldots, n).

### III. ORTHOGONAL LOCALITY AND GLOBALITY PRESERVING PROJECTIONS AND ITS OPTIMIZATION

In practical applications, the raw high-dimensional data is frequently complex and the distribution rule of the data is generally nonlinear so that intrinsic feature information often exists in the local manifold structures. Meanwhile, these data are often accompanied by large redundant information and noises. Most traditional DR methods do not pay attention to these manifold structures, and the existence of redundant information and noises also makes some traditional methods unable to absolutely achieve the feature information of the global and local structures. To enhance the ability of feature extraction and obtain more intrinsic local structures, we propose a novel DR method named OLGPP. Thanks to the orthogonality criterion, our method has better robustness and feature extract ability so that it can discover more efficient manifold structures and earn more comprehensive intrinsic feature information.

OLGPP combines the merits of traditional linear and nonlinear methods with the orthogonality criterion. Compared with the traditional methods, our methods can not only preserve the elasticity between samples, but also can reduce the sensitivity to the noise problems. Our method can be divided into the following parts. First, we construct a local undirected neighborhood graph and further formulate the local structure objective function for preserving the local manifold structures. Second, we describe the global Euclidean structures by constructing a global geometric graph and then give the global structure objective function for keeping the global Euclidean structures. Finally, combined with the orthogonality criterion, we present the complete objective function, which considers the global and local structures and give a twin eigen decomposition method [34] to solve this problem.

### A. LOCAL UNDIRECTED NEIGHBORHOOD GRAPH CONSTRUCTION AND ITS OBJECTIVE FUNCTION

In order to preserve the local manifold structures, we adopt the \( k \)-nearest neighbor method to construct the local undirected neighborhood graph \( G_{local} \). We can get this graph by using the following method: if \( x_j \) belongs to one of the \( k \)-nearest neighbors of \( x_i \), put an undirected edge between \( x_i \) and \( x_j \). The corresponding weights of \( G_{local} \) are described as

\[ S_{ij}^{local} = \begin{cases} 
\exp \left( -\frac{\|x_i - x_j\|^2}{2\tau^2} \right) & \text{if } x_i \in N_k(x_j) \\
& \text{or } x_j \in N_k(x_i) \\
0 & \text{otherwise} \\
\end{cases} \]
According to the corresponding linear transformation method, we give the local structure objective function as following:

\[
\min_{p} \sum_{i,j}^{n} \|y_i - y_j\|^2 S_{local}^{i,j} = \sum_{i,j}^{n} \|p^T x_i - p^T x_j\|^2 S_{local}^{i,j} = \sum_{i,j}^{n} p^T x_i D_{local}^{i,j} x_j^T p - \sum_{i,j}^{n} p^T x_i S_{local}^{i,j} x_j^T p = tr(p^T X L_{local}^{i,j} X^T p) \tag{9}
\]

where \( L_{local} = D_{local} - S_{local} \), \( D_{local} \) is the diagonal matrix, and the elements on the diagonal are the sum of the elements in each row of \( S_{local} \). That is, \( D = \text{diag}(\sum_{j=1}^{n} S_{ij}, \sum_{j=1}^{n} S_{1j}, \sum_{j=1}^{n} S_{2j}, \ldots, \sum_{j=1}^{n} S_{nj}) \).

Minimizing the objective function can acquire the more compact local features.

**B. GLOBAL GEOMETRIC GRAPH CONSTRUCTION AND ITS OBJECTIVE FUNCTION**

Inspired by the Undirected Complete Graph, we construct the global geometric graph \( G_{global} \) to describe the global Euclidean structures. In other words, there is an undirected edge between any two points in the graph. The corresponding weights of \( G_{global} \) are defined as

\[
S_{global}^{i,j} = \begin{cases} \|x_i - x_j\|^2 \exp\left(-\|x_i - x_j\|^2 / 2\sigma^2\right) & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{10}
\]

Similarly, the global structure objective function can be given as follows:

\[
\max_{p} \sum_{i,j}^{n} \|y_i - y_j\|^2 S_{global}^{i,j} = \sum_{i,j}^{n} \|p^T x_i - p^T x_j\|^2 S_{global}^{i,j} = \sum_{i,j}^{n} p^T x_i D_{global}^{i,j} x_j^T p - \sum_{i,j}^{n} p^T x_i S_{global}^{i,j} x_j^T p = tr(p^T X L_{global}^{i,j} X^T p) \tag{11}
\]

where \( S_{global} \) is the nearest neighbor weight matrix, \( D_{global} \) is the diagonal matrix, and \( L_{global} = D_{global} - S_{global} \) is the Laplacian matrix. Each element of \( D_{global} \) corresponds to the sum of each row vector of \( S_{global} \). The purpose of maximizing the objective function is to obtain more comprehensive global structure information.

**C. ORTHOGONAL LOCALITY AND GLOBALITY PRESERVING PROJECTIONS OBJECTIVE FUNCTION**

From Section III.A and Section III.B, it can be learned that we should minimize the objective function \( tr(p^T X L_{local}^{i,j} X^T p) \) so that the local manifold structures can be used to explore the intrinsic local features. Meanwhile, in order to keep the global Euclidean structures, we are necessary to maximize the objective function \( tr(p^T X L_{global}^{i,j} X^T p) \).

Inspired by [37], we use a method based on twin eigen decomposition to solve our objective function. According to the idea of this method, the first projection direction is required and then the other projection directions are solved sequentially. Therefore, the final objective function can be provided in two parts:

\[
\text{model } 1 \left\{ \begin{array}{l} \max_{p} p^T X \left((1 - \alpha) S_{local} + \alpha L_{global}\right) X^T p \\ \text{s.t. } p^T p = 1 \end{array} \right. \tag{12}
\]

\[
\text{model } 2 \left\{ \begin{array}{l} \max_{p} tr\left(p^T X \left((1 - \alpha) S_{local} + \alpha L_{global}\right) X^T p\right) \\ p^T p = 0 \quad (i = 1, 2, 3, \ldots, k - 1) \end{array} \right. \tag{13}
\]

where \( \alpha \in (0, 1) \) is a trade-off parameter used to balance the local manifold structures and the global Euclidean structures.

In addition, to ensure the uniqueness of the solutions, we further add constraints:

\[
\text{model } 1 \left\{ \begin{array}{l} \max_{p} p^T X \left((1 - \alpha) S_{local} + \alpha L_{global}\right) X^T p \\ \text{s.t. } p^T X(D_{local} - D_{global}) X^T p = 1 \end{array} \right. \tag{14}
\]

\[
\text{model } 2 \left\{ \begin{array}{l} \max_{p} tr\left(p^T X \left((1 - \alpha) S_{local} + \alpha L_{global}\right) X^T p\right) \\ \text{s.t. } tr\left(p^T X(D_{local} - D_{global}) X^T p\right) = 1 \\ p^T p = 0 \quad (i = 1, 2, 3, \ldots, k - 1) \end{array} \right. \tag{15}
\]

In conclusion, the first projection direction can be acquired by solving model 1 and then the \( k \) th orthogonal projection direction can be given from model 2.

**D. OPTIMIZATION**

To acquire the first orthogonal projection direction and the \( k \) th orthogonal projection direction, we respectively give two theorems to solve model 1 and model 2. From the Theorem 1, the first orthogonal projection direction can be obtained accurately and the \( k \) th orthogonal projection direction can be gotten sequentially according to Theorem 2.

**Theorem 1:** Let \( p \) is the solution of the following generalized eigenvalue function, then \( p \) is also the first orthogonal projection direction of OLGPP:

\[
(X H_{dd} X^T)^{-1} X H_{d} X^T p = \lambda p \tag{16}
\]

where \( H_{d} = (1 - \alpha) S_{local} + \alpha L_{global} \)

\[
H_{dd} = D_{local} - D_{global} \tag{17}
\]

**Proof:** By using the Lagrange multiplier method, (14) can be transformed into the solution problem as follows:

\[
L(p) = p^T X H_{dd} X^T p - \frac{\lambda}{2} (p^T X H_{dd} X^T p - 1) \tag{19}
\]

where \( \lambda \) is a Lagrange multiplier. Setting the partial derivative of \( L(p) \) with respect to \( p \) and letting it to zero, we can get

\[
\frac{\partial L(p)}{\partial p} = X H_{d} X^T p - \lambda X H_{dd} X^T p = 0 \tag{20}
\]

Obviously, we multiply both sides of (20) by \((X H_{dd} X^T)^{-1}\) to get (16), which means the Theorem 1 is established.
Theorem 2: Let \( P \) be a set of eigenvectors corresponding to the maximum eigenvalue of the following solvable eigenvalue problem, then \( P \) is also the \( k \)th orthogonal projection directions:
\[
(\text{XH}_{dd}X^T)^{-1} (I - G) \text{XH}_{dd}X^T p = \lambda p
\]
(21)

where \( p^{(k-1)} = (p_1, p_2, p_3, \ldots, p_{k-1}) \)
(22)

\[
G = p^{(k-1)} \left( p^{(k-1)} \right)^T \left( \text{XH}_{dd}X^T \right)^{-1} p^{(k-1)} \left( \text{XH}_{dd}X^T \right)^{-1} \left( p^{(k-1)} \right)^T \times \left( \text{XH}_{dd}X^T \right)^{-1}
\]
(23)

Proof: By means of the Lagrange multiplier method, \( (15) \) can be transformed into the following solution problem:
\[
L(p) = p^T \text{XH}_{dd}X^T p - \frac{\lambda}{2} (p^T \text{XH}_{dd}X^T p - 1) - \sum_{i=1}^{k-1} \sigma_ip_i^T p_i
\]
(24)

where \( \lambda \) and \( \sigma_i \) are Lagrange multipliers, set \( \sigma^{(k-1)} = \sigma_1, \sigma_2, \sigma_3, \ldots, \sigma_{k-1} \). Taking the partial derivative of \( L(p) \) with respect to \( p \) and setting it to zero, we can get
\[
\frac{\partial L(p)}{\partial p} = \text{XH}_{dd}X^T p - \lambda \text{XH}_{dd}X^T p - \sigma^{(k-1)} p^{(k-1)} = 0
\]
(25)

Multiplying the left side of (25) by \( p_i^T (\text{XH}_{dd}X^T)^{-1}, p_i^T (\text{XH}_{dd}X^T)^{-1}, \ldots, p_i^T (\text{XH}_{dd}X^T)^{-1} \) respectively, and the following formulation can be further obtained. For the convenience of calculation, the variables of each equation in the \( (26), \) as shown at the bottom of the page, can be combined in the form of a matrix, so that we can get the equation as follows
\[
\left[ p^{(k-1)} \right]^T (\text{XH}_{dd}X^T)^{-1} \text{XH}_{dd}X^T p - \left[ p^{(k-1)} \right]^T (\text{XH}_{dd}X^T)^{-1} \times \sigma^{(k-1)} p^{(k-1)} = 0
\]
(27)

From (27), we can further derive the expression of the Lagrange multipliers \( \sigma^{(k-1)} \) from the known variables
\[
\sigma^{(k-1)} = \left( \left[ p^{(k-1)} \right]^T (\text{XH}_{dd}X^T)^{-1} \right)^{-1} \left[ p^{(k-1)} \right]^T \times (\text{XH}_{dd}X^T)^{-1} \text{XH}_{dd}X^T p
\]
(28)

At present, the expression of \( \sigma^{(k-1)} \) is substituted into (25) to obtain
\[
\text{XH}_{dd}X^T p - \lambda \text{XH}_{dd}X^T p - p^{(k-1)} \times \left[ p^{(k-1)} \right]^T (\text{XH}_{dd}X^T)^{-1} \times (\text{XH}_{dd}X^T)^{-1} \text{XH}_{dd}X^T p = 0
\]
(29)

Algorithm 1 The OLGPP Method
Input: Training samples \( X = \{x_1, x_2, x_3, \ldots, x_n\} \in R^m \times n \)
Output: orthogonal projection matrix \( P = \{P_1, P_2, P_3, \ldots, P_d\} \)
1 Computer the weights \( S^{ij}_{\text{local}} \) and \( S^{ij}_{\text{global}} \) by (8) and (10) respectively
2 Obtain \( L_{\text{global}} \) and \( L_{\text{local}} \) by (9) and (11)
3 Obtain the first orthogonal projection direction \( p \) by using the (16)
4 for \( k = 2, 3, 4, \ldots, d \) do
5 computer the \( k \) projection direction from solving (21)
6 end for
7 Obtain orthogonal projection matrix \( P = \{P_1, P_2, P_3, \ldots, P_d\} \)

After optimizing (29) according to (22), we can get the final formulation
\[
(\text{XH}_{dd}X^T)^{-1} (I - G) \text{XH}_{dd}X^T p = \lambda p
\]
(30)

Above all, the Theorem 2. can be proved. Therefore, the \( k \) projection direction can be obtained by solving (29). In Algorithm 1, we show our OLGPP method in detail.

By solving model 1 and model 2, we can get the orthogonal projection matrix \( P = \{P_1, P_2, P_3, \ldots, P_d\} \) corresponding to the first \( d \) largest eigenvalues. According to \( y_i = P^T x_i, \ P \in R^m \times d \), the orthogonal features of the sample set \( X \) can be obtained as \( Y = \{P^T x_1, P^T x_2, P^T x_3, \ldots, P^T x_n\} \), \( Y \in R^d \times n \). Finally, we utilize the nearest classifier to obtain the final recognition results.

E. Time Complexity Analysis

The time complexity of our method is not very high in comparison with other related methods. In essence, the formulation of our method is a generalized eigenvalue problem. In other words, we can get the projection matrix by solving this problem. Specifically, the most computational step of OLGPP is step 1 and step 5. According to the Algorithm 1, time complexity of step 1 is \( O(n^2m + m^2n) \), and time complexity from step 5 is \( O((d - 1)m^2) \). So the total time complexity is \( O(n^2m^2 + m^2n + (d - 1)m^2) \).

IV. DISCUSSION ON SINGULARITY

In practical applications, the SSS problem is a common problem. When the SSS problem occurs, the number of samples \( n \) is lower than the dimensionality of data \( m \). In this section, we discuss the singularity of constraint matrix
\[
\begin{bmatrix}
p_1^T (\text{XH}_{dd}X^T)^{-1}XH_{dd}X^T p - p_1^T (\text{XH}_{dd}X^T)^{-1}p^{(k-1)}p^{(k-1)} = 0 \\
\vdots \\
\hat{p}_{k-1} (\text{XH}_{dd}X^T)^{-1}XH_{dd}X^T p - p_{k-1} (\text{XH}_{dd}X^T)^{-1}p^{(k-1)}p^{(k-1)} = 0
\end{bmatrix}
\]

55718
\( X(D_{\text{local}} - D_{\text{global}})X^T \) when the SSS problem exists in our method.

(a) if the diagonal elements of \( D_{\text{local}} - D_{\text{global}} \) do not contain zeros, we can give some conclusions as follows:

\[
\text{det}(D_{\text{local}} - D_{\text{global}}) \neq 0 \\
\iff D_{\text{local}} - D_{\text{global}} \text{ is invertible} \\
\iff D_{\text{local}} - D_{\text{global}} \text{ is nonsingular} \\
\iff \text{rank}(D_{\text{local}} - D_{\text{global}}) = n \tag{31}
\]

According to the property of the rank of matrix product, i.e., \( \text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B)) \), we have

\[
\text{rank}(X(D_{\text{local}} - D_{\text{global}})X^T) \
\leq \min(\text{rank}(X), \text{rank}((D_{\text{local}} - D_{\text{global}})X^T)) \leq n \tag{32}
\]

Since \( \text{rank}(D_{\text{local}} - D_{\text{global}}) = n \), we can further get

\[
\text{rank}((D_{\text{local}} - D_{\text{global}})X^T) \leq n.
\]

Above all, we can prove

\[
\text{rank}(X(D_{\text{local}} - D_{\text{global}})X^T) \
\leq \min(\text{rank}(X), \text{rank}((D_{\text{local}} - D_{\text{global}})X^T)) \leq n \tag{33}
\]

(b) if the diagonal elements of \( D_{\text{local}} - D_{\text{global}} \) contain zeros, some conclusions can be gotten as follows:

\[
\text{det} (D_{\text{local}} - D_{\text{global}}) = 0 \\
\iff D_{\text{local}} - D_{\text{global}} \text{ is irreversible} \\
\iff D_{\text{local}} - D_{\text{global}} \text{ is singular} \\
\iff \text{rank}(D_{\text{local}} - D_{\text{global}}) < n \tag{34}
\]

Similarly, according to the property of the rank of matrix product

\[
\text{rank}(X(D_{\text{local}} - D_{\text{global}})X^T) \
\leq \min(\text{rank}(X), \text{rank}((D_{\text{local}} - D_{\text{global}})X^T)) \leq n \tag{35}
\]

Now \( \text{rank}((D_{\text{local}} - D_{\text{global}})X^T) < n \), we can have

\[
\text{rank}(X(D_{\text{local}} - D_{\text{global}})X^T) \leq \min(\text{rank}(X), \text{rank}((D_{\text{local}} - D_{\text{global}})X^T)) < n \tag{36}
\]

In summary, \( \text{rank}(X(D_{\text{local}} - D_{\text{global}})X^T) \) is always less than \( n \) when the SSS problem exists, which means the constraint matrix is singular. In order to avoid the influence of the SSS problem, we can adopt two methods to process the data: (1) utilizing PCA to process the original data; (2) adopting regularization techniques to transform the original generalized eigenvalue decomposition problem is into a solvable problem.

### V. EXPERIMENTS AND ANALYSIS

To estimate the performance of OLGPP, we design experiments [38] on the Semeion handwritten digit dataset, COIL-20 object image dataset, UMIST face dataset and ORL face image dataset. To avoid the SSS problem from influencing the final experimental results, we use PCA to reduce the dimensionality of each sample to 100. As in [39], we select the kernel parameter from a series of \( \alpha = [0.2l, 0.4l, 0.6l, 0.8l, l, 2l, 4l, 6l, 8l, 10l] \), where \( l \) is average Euclidean distance of all samples. For the trade-off number \( \alpha \) and \( k \)—nearest neighbor parameters, they are set to 0.05 and 5 respectively according to [37], and all comparison methods adopt the same parameter settings if these parameters are involved. Moreover, in order to avoid the influence of the sample randomness on the final experimental results, we conduct random experiments ten times independently and take the average of ten random experiments to computer recognition rates. Additionally, all the recognition rates shown in the Section V are obtained by using the nearest classifier.

#### A. EXPERIMENTS ON SEMEION HANDWRITTEN DIGIT DATASET

There are 1593 handwritten digits on the Semeion handwritten digit dataset, which are composed of about 80 persons who have handwritten them twice. The first is to write numbers in a normal way (accurately write down all numbers from 0 to 9), and the second is to write numbers quickly (ignoring accuracy). In this section, we randomly elect \( \chi (\chi = 30, 40, 50, 60) \) images from per class of dataset as training samples and the other images are used as testing samples. Such random selections are executed ten times, and Table 1 lists the average recognition rates of ten random experiments.

LDA is a DR method that takes consideration of the global Euclidean structures, while LPP and NPE are DR methods to explore local manifold structures hidden in the raw data. Table 1 shows that the average recognition rates of LDA is better than these of LPP and NPE when the number of training samples is relatively small. However, the recognition ability of LPP and NPE is significantly better than LDA with the number of training samples increases. This also indicates that in the complex high-dimensional data, the distribution

### TABLE 1. The average recognition rates on the Semeion handwritten digit dataset.

| Methods | 30Train | 40Train | 50Train | 60Train |
|---------|---------|---------|---------|---------|
| LPP     | 81.80±1.16 | 85.25±0.89 | 86.96±1.24 | 88.65±0.96 |
| NPE     | 82.05±1.38 | 85.83±1.02 | 87.07±1.03 | 89.32±1.00 |
| LDA     | 83.20±0.94 | 85.57±1.13 | 86.78±1.00 | 87.75±1.05 |
| EPP     | 87.11±0.79 | 89.24±0.80 | 89.85±0.80 | 90.51±0.92 |
| SLPP    | 83.02±0.92 | 85.88±1.33 | 86.51±1.40 | 87.04±1.62 |
| OLGPP   | 87.50±0.74 | 89.55±0.69 | 90.19±0.88 | 90.63±0.78 |

\( A \pm B \): A means the average recognition rates (%), and B denotes the standard deviation.
rule of the data is nonlinear, and effective feature information generally exists in local structures. If we still adopt linear DR methods to process these data, intrinsic feature information will be lost in the process of DR, which is one of the reasons for low recognition rates. Compared with LPP and SLPP, EPP considers the local structures and the global structures. When reconstructing data in the learned projection subspace after dimensionality reduction, it retains more structure information in the learned subspace, which is also the reason why EPP in the Table 1 has always higher recognition rates than other comparison methods. By introducing the orthogonality criterion, our method can further capture the intrinsic local and global structures of the raw data and it is one of the reasons why our method has the best average recognition rates. In Table 1, our method outperforms other comparison methods in the recognition rates, which also proves the advantage of our method.

Besides, to more intuitively demonstrate the advantages of the method we proposed, Figure 1 reveals the variation of the average recognition rates in different dimensionalities.

![Graphs showing average recognition rates versus dimensionality](image)

**FIGURE 1.** The average recognition rates (%) versus dimensionality under different training samples of the Semeion dataset for per class: (a) 30 training samples; (b) 40 training samples; (c) 50 training samples; (d) 60 training samples.

of EPP are superior to those of LPP, NPE and SLPP from the beginning to the end and it also indicates that EPP can preserve the two structures effectively and reconstruct the feature information in the learned projection subspace. As the number of the dimensionality increases, the average recognition rates of all methods also begin to increase. And the recognition rates of our method exhibit the fastest increasing trend when the dimensionality increases gradually, which also indicates that our method can be effectively applied to the high-dimensional data. Different from other methods, the curve of SLPP grows slowly. However, the average recognition rates of all the methods tend to be stable when the dimensionality increases to a certain value, and our method appears a smoother change trend than the compared method. It also illustrates that our method has better robustness.

### B. EXPERIMENTS ON COIL-20 OBJECT IMAGE DATASET

The COIL20 object image dataset collects images of 20 kinds of objects, in which one image is taken every $5^\circ$ rotation of each category. Therefore, there are 72 images in each object, and 1440 object images are collected in the whole dataset. Figure 2 shows a positive image of 20 objects. $\chi(\chi = 6, 7, 8, 9)$ images are selected randomly from per class as training as samples and the other images are regarded as testing samples. We conduct ten such random experiments repeatedly and the average recognition rates are listed in Table 2. From Table 2, it can observe that the recognition...
FIGURE 2. The frontal image of 20 objects on the COIL20 image dataset.

TABLE 2. The average recognition rates on the COIL20 object image dataset.

| Methods | 6Train | 7Train | 8Train | 9Train |
|---------|--------|--------|--------|--------|
| LPP     | 73.27±1.86 | 79.21±1.57 | 81.87±1.18 | 84.86±0.62 |
| NPE     | 70.18±1.85 | 77.47±1.66 | 81.20±2.10 | 84.75±0.70 |
| LDA     | 73.64±4.00 | 81.89±2.14 | 85.48±1.48 | 89.16±1.02 |
| EPP     | 83.57±2.16 | 86.77±2.36 | 87.98±1.54 | 90.44±1.26 |
| SLPP    | 70.42±3.58 | 78.70±1.79 | 82.20±1.27 | 85.98±0.81 |
| OLGPP   | 86.61±1.60 | 88.82±1.17 | 89.63±1.34 | 91.73±1.02 |

A±B: A means the average recognition rates (%), and B denotes the standard deviation.

TABLE 3. The average recognition rates on the UMIST face image dataset.

| Methods | 6Train | 7Train | 8Train | 9Train |
|---------|--------|--------|--------|--------|
| LPP     | 74.81±3.19 | 83.89±1.92 | 88.58±2.25 | 92.28±2.66 |
| NPE     | 71.71±3.64 | 83.06±2.96 | 87.90±2.85 | 91.54±2.55 |
| LDA     | 81.93±3.72 | 90.90±2.14 | 94.12±2.45 | 96.25±2.13 |
| EPP     | 85.69±3.69 | 90.44±1.31 | 92.60±2.02 | 93.62±2.34 |
| SLPP    | 77.10±3.35 | 86.64±1.67 | 89.59±2.14 | 91.62±2.42 |
| OLGPP   | 87.98±3.48 | 91.36±1.44 | 92.60±1.86 | 94.51±1.65 |

A±B: A means the average recognition rates (%), and B denotes the standard deviation.

rates of all methods present an upward trend with the increase of training samples, and the average recognition rates of EPP consistently outperform those of LPP, NPE and SLPP, while our method always maintains the optimal recognition rates. Different from the Semeion dataset, NPE has the lowest average recognition rates in Table 2. Furthermore, the standard deviation of OLGPP is generally lower than that of other methods in Table 2. From the viewpoints of mathematical meaning, standard deviation reflects the degree of dispersion between objects in a group. In other words, it means that the recognition rates obtained in each random experiment are extremely close to the average recognition rates and there are few differences between the recognition rates, which just proves that our method has favorable stability for different samples.

C. EXPERIMENTS ON UMIST FACE DATASET

The UMIST face dataset consists of 20 persons with a total of 564 images, and each person is collected a series of posture images from the side view to the front view. In our experiments, we randomly choose \( \chi(\chi = 6, 7, 8, 9) \) images from each class of the dataset as training samples while the remaining images are regarded as testing samples. In Table 3, we tabulate the average recognition rates for ten random experiments. From Table 3, it can be learned that the overall change trend in the average recognition rates is the same as that in the COIL20 image dataset, and NPE always performs the worst classification performance in the all methods. Because of considering the global and local structures, the average recognition rates of the EPP method is optimal in the four compared methods, and our method still maintains the best classification performance.

D. EXPERIMENTS ON ORL FACE IMAGE DATASET

The ORL face image dataset takes face images of 40 persons in a dark and uniform background. Each person is collected respectively 10 images at different times, different illumination, different facial expressions, and different facial details.
TABLE 4. The average recognition rates on the ORL face image dataset.

| Methods | 3Train | 4Train | 5Train | 6Train |
|---------|--------|--------|--------|--------|
| LPP     | 71.64±3.22 | 88.50±1.57 | 93.10±2.02 | 96.06±1.32 |
| NPE     | 70.68±2.94 | 87.92±1.52 | 93.45±1.91 | 95.63±0.93 |
| LDA     | 72.71±4.05 | 89.88±1.66 | 94.90±1.78 | 96.63±1.80 |
| EPP     | 88.07±2.45 | 92.75±1.73 | 94.65±2.31 | 96.69±1.03 |
| SLPP    | 76.36±3.24 | 88.42±1.33 | 91.80±1.96 | 93.63±1.01 |
| OLGPP   | 89.25±2.88 | 92.67±2.00 | 94.85±2.10 | 96.69±0.84 |

A±B: A means the average recognition rates (%), and B denotes the standard deviation.

![FIGURE 3. The best recognition rates of different methods in ten random experiments on COIL20 object image dataset: (a) 6 training samples for per class; (b) 7 training samples for per class; (c) 8 training samples for per class; (d) 9 training samples for per class.](image)

We randomly select $\chi (\chi = 3, 4, 5, 6)$ images from per class as testing samples while the remaining images are treated as testing samples and repeat ten time, and we tabulate the average recognition rates in Table 4. Similar to the experiment results on COIL20 and UMSIT datasets, the recognition ability of NPE is the worst in Table 4. On the ORL face image dataset, our method always has the optimal classification performance, and the standard deviation is consistently lower than other methods.

E. EXPERIMENTAL RESULTS ANALYSIS

From all the experiment results in Section V, some conclusions can be given as follows: (1) the recognition ability of all the methods increases quickly as the number of training samples increases, which indicates abundant training samples play an important role in the improvement of recognition ability; (2) compared with linear DR methods, nonlinear methods can discover the intrinsic local structures hidden in the raw high-dimensional data effectively, which are beneficial to DR; (3) our method integrated with the orthogonality criterion always has low the standard deviation, which means that our method has favorable robustness and can be capable of handling the complex high-dimensional data and noises in the datasets; (4) in all the experiments, our method has the optimal experimental results, which demonstrates that the method proposed is a feasible DR method.
F. THE IMPACT OF THE SAMPLE RANDOMNESS

In the experiments, the recognition rates in the above tables are the average values computing ten random independent experiments. To more intuitively illustrate whether the sample randomness affects the recognition rates, in the form of histograms, we enumerate the recognition rates in each random experiment on the COIL20 and the ORL datasets respectively when images are selected as training samples in
per class. By analyzing Figure 3 and Figure 4, we can get some conclusions: (1) in each random experiment, the recognition rates of all the methods has changed, but the change trend of our method is not obvious. In terms of the overall change trend, compared with other methods, the change trend of our method is relatively gentle, and almost close to a fixed value, which is the reason why the standard deviation of our method is smaller than that of other methods in Table 2 and Table 4; (2) our method has the highest recognition rates in every random experiment, which further illustrates the method feasibility.

G. THE IMPACT OF KERNEL PARAMETER

In our objective function, we are necessary to use a kernel parameter $t$ to compute the weights of the global geometric graph and the local undirected neighborhood graph. The size of the kernel parameter directly affects the construction of the two graphs, which means that the kernel parameter may influence the global and local structures preserved in the learned subspace. In this section, we analyze whether the kernel parameters affect the recognition ability of our method. In Figure 5, we illustrate the best recognition rates of different parameters when $\chi$ images are selected for training in all experimental datasets. As can be seen from Figure 5, when the number of the training samples is fixed, the recognition rates corresponding to different parameters have little change, which indicates that the kernel parameter rarely affects the classification performance. Simultaneously, we can observe the common change trend on the four datasets: the recognition rates of our method will increase as the number of training samples increases. Associated with Table 1-4, we can conclude that the classification performance of OLGPP is hardly influenced by the kernel parameter, and our method has strong robustness to the kernel parameter. Besides, OLGPP is more sensitive to the number of training samples and increasing the number of training samples is a practicable way to improve the recognition rates.

VI. CONCLUSION

In this paper, a novel DR method named OLGPP is proposed. In our method, the local manifold structures and the global Euclidean structures are utilized to explore the intrinsic structure information hidden in the raw high-dimensional data and the discriminate power of low-dimensional orthogonal features will be further enhanced by the orthogonality criterion. Furthermore, our method is also insensitive to the noises, which is conducive to the recognition ability. By constructing the orthogonal objective function and solving this function sequentially, low-dimensional orthogonal features are obtained. Extensive experimental results on several datasets demonstrate that our method is an effective DR method.

In addition, some the raw high-dimensional data also contains class label information. As reported in [11], classification performance can be improved from using the class label information. Our method is an unsupervised method, and our future work will focus on the embedding of class label information for further improving the classification performance of our method.

REFERENCES

[1] S. Ayesha, M. K. Hanif, and R. Talib, “Overview and comparative study of dimensionality reduction techniques for high dimensional data,” Inf. Sci., vol. 59, pp. 164–182, 2020.
[2] C. C. Aggarwal, “On k-anonymity and the curse of dimensionality,” in Proc. VLDB, vol. 5, 2005, pp. 901–909.
[3] M. Paolanti and E. Frontoni, “Multidisciplinary pattern recognition applications: A review,” Comput. Sci. Res., vol. 37, Aug. 2020, Art. no. 100276.
[4] T. Meng, J. J. Zhang, Z. Yan, and W. Pedrycz, “A survey on machine learning for data fusion,” Inf. Fusion, vol. 57, pp. 115–129, May 2020.
[5] Y. Wang, Y. Cao, Z. Guo, T. Huang, and S. Wen, “Event-based sliding-mode synchronization of delayed memristive neural networks via continuous/periodic sampling algorithm,” Appl. Math. Comput., vol. 383, Oct. 2020, Art. no. 125379.
[6] S. Wang, Y. Cao, T. Huang, Y. Chen, and S. Wen, “Event-triggered distributed control for synchronization of multiple memristive neural networks under cyber-physical attacks,” Inf. Sci., vol. 518, pp. 361–375, May 2020.
[7] S. Vellaiangiri, S. Alagumuthukrishnan, and S. I. T. Joseph, “A review of dimensionality reduction techniques for efficient computation,” Procedia Comput. Sci., vol. 165, pp. 104–111, Jan. 2019.
[8] Z.-K. Feng, W.-J. Niu, and C.-T. Cheng, “China’s large-scale hydropower system: Operation characteristics, modeling challenge and dimensionality reduction possibilities,” Renew. Energy, vol. 136, pp. 805–818, Jun. 2019.
[9] H. Abdi and L. J. Williams, “Principal component analysis,” Wiley Interdiscip. Rev., Comput. Statist., vol. 2, no. 4, pp. 433–459, 2010.
[10] S. Balakrishnana and A. Ganapathiraju, “Linear discriminant analysis—A brief tutorial,” Inst. Signal Inf. Process., vol. 18, pp. 1–8, Mar. 1998.
[11] L. Yu, F. Cao, X. Zhao, X. Yang, and J. Liang, “Combining attribute content and label information for categorical data ensemble clustering,” Appl. Math. Comput., vol. 381, Sep. 2020, Art. no. 125280.
[12] W. Zhao, R. Chellappa, and P. Jonathon, “Philips. Subspace linear discriminant analysis for face recognition,” Comput. Vis. Lab., Center Automat. Res., Univ. Maryland, College Park, MD, USA, Tech. Rep., 1999.
[13] R. Hajizadeh, A. Aghagolzadeh, and M. Ezoji, “Local distances preserving based manifold learning,” Expert Syst. Appl., vol. 139, Jan. 2020, Art. no. 112860.
[14] S. Wen, H. Wei, Y. Yang, Z. Guo, Z. Zeng, T. Huang, and Y. Chen, “Memristive LSTM network for sentiment analysis.” IEEE Trans. Syst., Man, Cybern. Syst., vol. 51, no. 3, pp. 1794–1804, Mar. 2021.
[15] S. Wen, M. Dong, Y. Yang, P. Zhou, T. Huang, and Y. Chen, “End-to-end detection-segmentation system for face labeling,” IEEE Trans. Emerg. Topics Comput. Intell., early access, Nov. 6, 2019, doi: 10.1109/TETCI.2019.2947319.
[16] Y. Cao, Y. Cao, Z. Guo, T. Huang, and S. Wen, “Global exponential synchronization of delayed memristive neural networks with reaction–diffusion terms,” Neural Netw., vol. 123, pp. 70–81, Mar. 2020.
[17] S. T. Roweis, “Nonlinear dimensionality reduction by locally linear embedding.” Science, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
[18] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” Neural Comput., vol. 15, no. 6, pp. 1373–1396, Jun. 2003.
[19] M. A. A. Cox and T. F. Cox, “Multidimensional scaling,” in Handbook of Data Visualization, Berlin, Germany: Springer, 2008, pp. 315–347.
[20] J. B. Tenenbaum, V. De Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction.” Science, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.
[21] D. L. Donoho and C. Grimes, “Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data,” Proc. Nat. Acad. Sci. USA, vol. 100, no. 10, pp. 5591–5596, May 2003.
[22] M. Iqbal, K. Zhang, S. Iqbal, and I. Tariq, “A fast and reliable dijkstra algorithm for online shortest path,” Int. J. Comput. Sci. Eng., vol. 5, no. 12, pp. 24–27, Dec. 2018.
[23] L. Zhao and J. Zhao, “Comparison study of three shortest path algorithm,” in Proc. Int. Conf. Comput. Technol., Electron. Commun. (ICCTEC), Dec. 2017, pp. 748–751.
[24] W.-J. Niu, Z.-K. Feng, B.-F. Feng, Y.-S. Xu, and Y.-W. Min, “Parallel computing and swarm intelligence based artificial intelligence model for multi-step-ahead hydrological time series prediction,” Sustain. Cities Soc., vol. 66, Mar. 2021, Art. no. 102686.

[25] Z.-K. Feng, W.-J. Niu, and C.-T. Cheng, “Optimization of hydropower reservoirs operation balancing generation benefit and ecological requirement with parallel multi-objective genetic algorithm,” Energy, vol. 153, pp. 706–718, Jun. 2018.

[26] Y. Bengio, “Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and spectral clustering,” in Proc. Adv. Neural Inf. Process. Syst., 2004, pp. 1–8.

[27] X. He and P. Niyogi, “Locality preserving projections,” in Proc. Adv. Neural Inf. Process. Syst., 2004, pp. 153–160.

[28] X. He, D. Cai, S. Yan, and H.-J. Zhang, “Neighborhood preserving embedding,” in Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV), vol. 2, Oct. 2005, pp. 1208–1213.

[29] S.-J. Wang, B.-J. Li, Y.-J. Liu, W.-J. Yan, X. Ou, X. Huang, F. Xu, and X. Fu, “Micro-expression recognition with small sample size by transferring long-term convolutional neural network,” Neurocomputing, vol. 312, pp. 251–262, Oct. 2018.

[30] G. Feng, D. Hu, D. Zhang, and Z. Zhou, “An alternative formulation of kernel LPP with application to image recognition,” Neurocomputing, vol. 69, nos. 13–15, pp. 1733–1738, Aug. 2006.

[31] X. Tao, H. Ji, and J. Men, “Kernel neighborhood preserving embedding for classification,” J. Electron., vol. 26, no. 3, pp. 374–379, May 2009.

[32] B. Yang and S. Chen, “Sample-dependent graph construction with application to dimensionality reduction,” Neurocomputing, vol. 74, nos. 1–3, pp. 301–314, Dec. 2010.

[33] F. Zang, J. Zhang, and J. Pan, “Face recognition using elasticfaces,” Pattern Recognit., vol. 45, no. 11, pp. 3866–3876, Nov. 2012.

[34] E. Kokio-poulou and Y. Saad, “Orthogonal neighborhood preserving projections: A projection-based dimensionality reduction technique,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 12, pp. 2143–2156, Dec. 2007.

[35] D. Cai, X. He, J. Han, and H.-J. Zhang, “Orthogonal Laplacianfaces for face recognition,” IEEE Trans. Image Process., vol. 15, no. 11, pp. 3608–3614, Nov. 2006.

[36] D. P. Bertsekas, Constrained Optimization and Lagrange Multiplier Methods. New York, NY, USA: Academic, 2014.

[37] X.-B. Shen, Q.-S. Sun, and Y.-H. Yuan, “Orthogonal canonical correlation analysis and its application in feature fusion,” in Proc. 16th Int. Conf. Inf. Fusion, 2013, pp. 151–157.

[38] S.-J. Wang, B.-J. Li, Y.-J. Liu, W.-J. Yan, X. Ou, X. Huang, F. Xu, and X. Fu, “Micro-expression recognition with small sample size by transferring long-term convolutional neural network.” Neurocomputing, vol. 312, pp. 251–262, Oct. 2018.

[39] S. Su, H. Ge, and Y.-H. Yuan, “Kernel propagation strategy: A novel out-of-sample propagation projection for subspace learning,” J. Vis. Commun. Image Represent., vol. 36, pp. 69–79, Apr. 2016.

SHUZHI SU received the Ph.D. degree from the School of Internet of Things Engineering, Jiangnan University, Wuxi, China. He is currently an Associate Professor with the School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, China. His research interests include multimodal pattern recognition, information fusion, feature learning, and image processing.

GANG ZHU received the B.S. degree from the Department of Internet of Things Engineering, Anhui University of Science and Technology, Huainan, China, where he is currently pursuing the M.S. degree in computer science and engineering. His research interests include dimensionality reduction and image recognition.

YANMIN ZHU received the M.S. degree from the School of Mechanical and Electronic Engineering, Shandong Agricultural University, Tainan, China. She is currently pursuing the Ph.D. degree with the School of Mechanical Engineering, Anhui University of Science and Technology, Huainan, China. Her research interests include multi-modal information fusion, correlation projection theory, and image processing.