IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 29 August 2022, accepted 14 September 2022, date of publication 22 September 2022,
date of current version 30 September 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3208901

RESEARCH ARTICLE

Marginal Fisher Analysis With Polynomial Matrix Function

RUISHENG RAN[1], ZHENG LI[1], TING WANG[1], JI FENG[1], AND BIN FANG[2], (Senior Member, IEEE)

[1]College of Computer and Information Science, College of Intelligent Science, Chongqing Engineering Research Center of Educational Big Data Intelligent Perception and Application, Chongqing Normal University, Chongqing 401331, China
[2]College of Computer Science, Chongqing University, Chongqing 400044, China

Corresponding author: Bin Fang (fb@cqu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61876026, in part by the Humanities and Social Sciences Project of the Ministry of Education of China under Grant 20YJAZH084 and Grant 18XJC880002, in part by the Chongqing Technology Innovation and Application Development Project under Grant cstc2020jscx-msxmX0190 and Grant cstc2019jscxmbdxX0061, and in part by the Science and Technology Research Program of Chongqing Municipal Education Commission under Grant KJZD-K202100505 and Grant KJQN201800539.

ABSTRACT Marginal fisher analysis (MFA) is a dimensionality reduction method based on a graph embedding framework. In contrast to traditional linear discriminant analysis (LDA), which requires the data to follow a Gaussian distribution, MFA is suitable for non-Gaussian data, and it has better pattern classification ability. However, MFA has the small-sample-size (SSS) problem. This paper aims to solve the small-sample-size problem while increasing the classification performance of MFA. Based on a matrix function dimensionality reduction framework, the criterion of the MFA method is reconstructed by using the polynomials matrix function transformation, and then a new MFA method is proposed, named PMFA (polynomial marginal fisher analysis). The major contributions of the proposed PMFA method are that it solves the small-sample-size problem of MFA, and it can enlarge the distance between marginal sample points of inter-class, so that it can get better pattern classification performance. Experiments on the public face datasets show that PMFA can get a better classification ability than MFA and its improved methods.

INDEX TERMS Dimensionality reduction, manifold learning, marginal fisher analysis, matrix function, the small-sample-size (SSS) problem.

I. INTRODUCTION

In some fields, such as data visualization and face recognition, features are usually represented as high-dimensional data. The original high-dimensional data contains redundant and noisy information, which can easily cause errors in practical applications, and the direct processing of high-dimensional data may cause considerable computational consumption. To better analyze and process data, it is necessary to find low-dimensional features that effectively represent high-dimensional data.

Many dimensionality reduction methods have been proposed. Principal component analysis (PCA) [1] and LDA [2] are widely used linear subspace algorithms. As an unsupervised learning algorithm, the principle of PCA is to maximize the covariance of reduced dimension samples. As a supervised learning algorithm, the principle of LDA is to make different classes as far as possible and the same classes as close as possible after dimensionality reduction.

There are many dimensionality reduction methods based on manifold, such as ISOMAP [3], LLE [4], LE [5], MVU [6], t-SNE [7], LPP [8], and NPE [9]. What they have in common is to find a neighborhood in each sample point, and retain the local structure information of the sample points while mapping the high-dimensional data into low-dimensional data. With the emergence of these classical manifold learning algorithms one after another, some researchers have hoped to unify manifold learning algorithms using a framework.

The associate editor coordinating the review of this manuscript and approving it for publication was Mehul S. Raval.

Bengio *et al.* unified ISOMAP, LLE, and LE into kernel functions [10]. Yan *et al.* proposed a unified graph embedding framework [11]. This framework can explain the existing dimensionality reduction algorithms with a linear, kernel, tensor extension of a certain intrinsic graph, or direct graph embedding.

In the graph embedding framework, the marginal fisher analysis method (MFA) is proposed. The main idea behind MFA is that it describes intra-class compactness by constructing an intrinsic graph and inter-class separability by constructing a penalty graph. Based on the design of the two graphs, it is also suitable for data of non-Gaussian distribution; therefore, there is no need for a priori knowledge of data distribution. However, the classical LDA method requires the data to follow a Gaussian distribution. Compared with LDA, MFA can find a more discriminative direction. However, the dimensions of the samples are usually much larger than the number of samples, resulting in the intra-class compactness matrix of the MFA being singular, and the generalized eigenvalue problem of MFA is difficult to solve. This may be called the small-sample-size (SSS) problem of MFA.

In recent years, much research has been conducted on MFA. Based on the separability criterion of MFA, Yu *et al.* proposed the orthogonal MFA (OMFA) and uncorrelated MFA (UMFA) method, in which orthogonal and irrelevant constraints are added to the base vector [12]. Siena proposed a coupled MFA (CMFA) method, which considers the nearest neighbor of each data sample and uses different rules to construct the penalty graph of the MFA [13]. Huang *et al.* proposed the multiple MFA (MMFA) method that uses edge data points and manifold structures and considers multiple marginal data pairs when constructing a penalty graph [14]. For the SSS problem of MFA, MFA uses PCA to reduce the dimensionality of the samples and then avoids the singularity of the intra-class compactness matrix, but this will lose some useful information about the original samples [11]. In addition, some methods have been proposed to solve the SSS problem of MFA. For example, Wang *et al.* proposed a general matrix exponential framework, the SSS problem of MFA [15] can be solved by an exponential MFA (EMFA) method which is based on this framework. Huang proposed a regularized MFA (RMFA) [16], which multiplies a unit matrix by a small number to construct a regular term into the inner-class compactness matrix of the MFA so that the resulting matrix is invertible. The graph embedding framework is reformulated in [17], as a special case of the GEU framework formula, the MFA technique is extended as GEU-MFA. Recently, there are many algorithms based on manifold learning, such as GDE [18], CR-DLPP [19] and LMGE-DDR [20].

In [21], a general framework for dimensionality reduction based on the manifold was proposed. The main idea of the framework is to use two scalar functions to map the scatter matrices into the corresponding matrix functions, and then to make dimensionality reduction and feature extraction. However, in real life, it is difficult to choose such two functions. In mathematics, any function can be approximated by a polynomial, so we can use the polynomial to uniformly represent the above optional and various forms of functions. Then, this framework is easier to use.

This paper aims to solve the SSS problem and improve the classification ability of MFA. So, based on the above idea, i.e., combining the polynomial function and the framework in [21], we proposed a new MFA method, named PMFA (polynomial marginal fisher analysis). Specially, we use two appropriate polynomials to map the scattering matrices of MFA to the new space, which can avoid the SSS problem and get better pattern classification performance. We also discuss the design of the two polynomial functions, and provide a theoretical analysis of the proposed method. Experiments are conducted on synthetic data set and some public face datasets, which show that the proposed PMFA is an effective method. As an effective feature extraction method, like MFA and its variants, PMFA can be applied in many fields, such as face recognition [13], [16], facial expression recognition [22], autism trait classification [23], image representation [24], etc.

The remainder of this paper is organized as follows. The Section II summarizes MFA and the matrix function dimensionality reduction framework. Section III presents polynomial marginal fisher analysis (PMFA). Section IV verifies PMFA with experiments. Finally, Section V summarizes the study and future directions.

TABLE 1. Notations and descriptions.

Symbol	Explanation
$\boldsymbol{X}=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\mathrm{L}\ ,\boldsymbol{x}_n\}$	the m-dimensional data in R^n space
$\boldsymbol{x}_i$	the original sample
m	the dimensionality of training samples
n	the number of training samples
c_i	the category label of the ith category and $c_i\in\{1,2,\mathrm{L}\ ,m_c\}$
m_c	the number of categories of the samples
$\boldsymbol{y}_i$	the projected sample
$\boldsymbol{L}$, $\boldsymbol{L}^p$	the Laplace matrix
$rank(\mathrm{g})$	the rank of the matrix

II. RELATED WORK

A. MARGINAL FISHER ANALISIS

MFA first constructs an intrinsic graph $\boldsymbol{G}^c=\{\boldsymbol{X},\boldsymbol{W}\}$ and penalty graph $\boldsymbol{G}^p=\{\boldsymbol{X},\boldsymbol{W}^p\}$, which are used to describe inter-class separability and intra-class compactness, respectively. Then, MFA tries to find an optimal projection matrix $\boldsymbol{U}$ and makes a projection $\boldsymbol{y}_i=\boldsymbol{U}^T\boldsymbol{x}_i$, so that the dimension of $\boldsymbol{y}_i$ is smaller than that of $\boldsymbol{x}_i$. To clarify the method, Table 1 summarizes the frequently used notations.

The intra-class compactness of the projected sample is defined as:

$$S_c = \sum_i \sum_{i \in N_{k_1}(j) or j \in N_{k_1}(j)} \left\| y_i - y_j \right\|^2 = 2tr\left\{ U^T X L X^T U \right\} \tag{1}$$

where $L = D - W$; D is a diagonal matrix composed of the sum of the rows of W, i.e., $D_{ii} = \sum_j W_{ij}$. $W = \left(W_{ij}\right)$, W_{ij} is defined as:

$$W_{ij} = \begin{cases} 1, & \text{if } i \in N_{k1}(j) \text{ or } j \in N_{k1}(i) \\ 0, & \text{else.} \end{cases} \tag{2}$$

where $N_{k_1}(i)$ represents an index set of the k_1 nearest neighbors of the sample x_i in the intrinsic graph G^c.

The projected inter-class data separation is defined as:

$$\begin{aligned} S_p &= \sum_i \sum_{(i,j) \in N_{k_2}(c_i) or (i,j) \in N_{k_2}(c_j)} \left\| y_i - y_j \right\|^2 \\ &= 2tr\left\{ U^T X L^p X^T U \right\} \end{aligned} \tag{3}$$

where $L^p = D^p - W^p$; D^p is a diagonal matrix composed of the row sum of the matrix W^p, i.e., $D^p_{ii} = \sum_j W^p_{ij}$. $W^p = \left(W^p_{ij}\right)$, W^p_{ij} is defined as:

$$W^p_{ij} = \begin{cases} 1, & \text{if } i \in N_{k_2}\left(c_j\right) \text{ or } j \in N_{k_2}(c_i) \\ 0, & \text{else.} \end{cases} \tag{4}$$

where $N_{k_2}(c_i)$ represents an index set of thek_2 nearest pairs for the class c_i in the penalty graph G^p.

The marginal fisher criterion is defined as:

$$J(U) = \max_U \frac{tr\left(U^T X L^p X^T U\right)}{tr\left(U^T X L X^T U\right)}. \tag{5}$$

Let the matrix $Z_p = X L^p X^T$, matrix $Z = X L X^T$, and upper Eq. (5) can be solved using the generalized eigenvectors problem:

$$Z_p u = \lambda Z u. \tag{6}$$

The basic MFA algorithm is given in Algorithm 1.

For the above Eq. (6), the rank of matrix Z has an inequality:

$$rank(Z) = rank\left(X L X^T\right) \leq rank(X) \leq \min(m, n).$$

Usually, the dimension of training samples is much larger than the number of training samples, that is $m \ll n$, so $rank(Z) < n$. And $Z = X L X^T$ is an n-order matrix, i.e., Z is a singular matrix, so Eq. (6) is difficult to solve directly, which is the SSS problem of MFA.

Algorithm 1 MFA

1) **Input** a dataset X

2) **PCA projection:** project the data set into the PCA subspace, let X_{PCA} denote the transformation matrix of PCA

3) **Constructing the intra-class compactness and inter-class separability matrixes:**

Adjacency matrix: $W_{ij} = W_{ji} = 1$ if x_j is among the k_1 nearest neighbors of x_i in the same class

Similarity matrix: $W^p_{ij} = W^p_{ji} = 1$ if the pair (i, j) is among the k_2 nearest pairs

4) **Marginal fisher criterion**

$$J(U) = \max_U \frac{tr\left(U^T X (D^p - W^p) X^T U\right)}{tr\left(U^T X (D - W) X^T U\right)}$$

5) **Compute the eigenvalues and eigenvectors** as:

$$X(D^p - W^p)X^T u = \lambda X(D - W)X^T u$$

6) **Output**the final linear projection direction as:

$$Y = U^T X_{PCA}$$

B. MATRIX FUNCTION AND ITS EIGEN-SYSTEM

The definition of the matrix function and the corresponding properties are presented, which is used in the paper.

Definition 1 [25]: Let A be a square matrix and $f(x)$ be a scalar function. If one replaces the variable x in $f(x)$ with the square matrix A, the resulting matrix $f(A)$ is called the matrix function of the matrix A.

Theorem 1 [26]: Let A be a real symmetric square matrix of n-order, $f(x)$ be a scalar function, λ_i be the eigenvalue of the matrix A and v_i be the eigenvector belonging to the eigenvalue λ_i, i.e., $Av_i = \lambda_i v_i (i = 1, 2, \cdots, n)$. For the matrix function $f(A)$, one has:

$$f(A)v_i = f(\lambda_i)v_i (i = 1, 2, \cdots, n), \tag{7}$$

where $f(\lambda_i)$ are the eigenvalues of $f(A)$, and v_i are the eigenvectors corresponding to $f(A)$.

Theorem 2 (Weierstrass Approximation Theorem [27]): Continuous functions on closed intervals can be approximated uniformly using a polynomial series.

C. MATRIX FUNCTION DIMENSIONALITY REDUCTION FRAMEWORK

In [21], a general matrix function dimensionality reduction framework is proposed for the dimensionality reduction method in manifold learning.

The criterion function of the manifold-based dimensionality reduction method is expressed as:

$$J(U) = \arg\max_U \frac{\left|U^T S_1 U\right|}{\left|U^T S_2 U\right|}, \tag{8}$$

where the matrices S_1 and S_2 have different forms for the different methods and $U$is the desired optimal projection matrix.

The main concept of the matrix function framework in [21] is to map the matrices $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ in Eq. (8) into the matrix functions $f(\boldsymbol{S}_1)$ and $g(\boldsymbol{S}_2)$with the scalar functions $f(x)$ and $g(x)$, so that the criterion Eq. (8) becomes:

$$J(\boldsymbol{U}) = \arg\max_{\boldsymbol{U}} \frac{\left|\boldsymbol{U}^T f(\boldsymbol{S}_1)\boldsymbol{U}\right|}{\left|\boldsymbol{U}^T g(\boldsymbol{S}_2)\boldsymbol{U}\right|}. \tag{9}$$

Eq. (9) can be reduced the generalized eigenvectors problem:

$$f(\boldsymbol{S}_1)\boldsymbol{u} = \lambda g(\boldsymbol{S}_2)\boldsymbol{u}. \tag{10}$$

Notice that the matrices $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ are composed of the original sample, and the above mapping $f(\cdot)$ and $g(\cdot)$ transform the matrices $\boldsymbol{S}_1$ and $\boldsymbol{S}_2$ from the original sample space to a new sample space.

The functions $f(x)$ and $g(x)$ in Eq. (9) are configurable. If selects different functions, the matrix function dimensionality reduction framework can construct a series of methods for dimensionality reduction. How to select the two functions is discussed in [21]. In conclusion, the two functions should satisfy the two conditions: (1) for any x, the function $g(x)$ should be $g(x) \neq 0$ and then the matrix function $g(\boldsymbol{S}_2)$ is non-singular so that the SSS problem can be avoided by the new dimensionality reduction method; (2) for better classification after mapping, the function $f(x)$ should be a monotonically increasing function and $f(x) > x$, at the same time, $g(x) \approx x$ (or$g(x) < x$).

III. THE PROPOSED NEW METHOD

A. MARGINAL FISHER ANALYSIS WITH POLYNOMIAL MATRIX FUNCTION

In the criterion of MFA, i.e., Eq. (5), denote the matrix $\boldsymbol{Z}_p = \boldsymbol{X}\boldsymbol{L}^p\boldsymbol{X}^T$ and the matrix $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^T$, then Eq. (5) can be expressed as:

$$J(\boldsymbol{U}) = \max_{\boldsymbol{U}} \frac{tr\left(\boldsymbol{U}^T\boldsymbol{Z}_p\boldsymbol{U}\right)}{tr\left(\boldsymbol{U}^T\boldsymbol{Z}\boldsymbol{U}\right)}. \tag{11}$$

According to section II.C, we can solve the small-sample-size problem and improve the classification performance of MFA by selecting two suitable functions to map the matrices $\boldsymbol{Z}_p$ and $\boldsymbol{Z}$ to the corresponding matrix functions, that is:

$$\boldsymbol{Z}_p \to f\left(\boldsymbol{Z}_p\right), \quad \boldsymbol{Z} \to g\left(\boldsymbol{Z}\right). \tag{12}$$

After mapping, we get a new criterion function:

$$J_p(\boldsymbol{U}) = \max_{\boldsymbol{U}} \frac{tr\left(\boldsymbol{U}^T f\left(\boldsymbol{Z}_p\right)\boldsymbol{U}\right)}{tr\left(\boldsymbol{U}^T g\left(\boldsymbol{Z}\right)\boldsymbol{U}\right)}. \tag{13}$$

According to section II.C, if $f(x)$ and $g(x)$ have satisfied the two conditions, the new method of Eq. (13) can avoid the small-sample-size problem of MFA and improve the classification ability of MFA. Although there are several functions to choose from, there are two problems. One is that although many functions meet the conditions, it is difficult to give a specific function form. Another problem is that among all the functions that meet the conditions, some are good, some are poor, so is there a standard form of the function? According to Theorem 2, the polynomial can approximate any function, that is, any function can be formulated by a polynomial. So, inspired by Theorem 2, we use polynomial functions to implement the above objectives.

We choose an n-order polynomial $f(x) = \sum_{k=0}^{n} a_k x^k$ to map the matrix $\boldsymbol{Z}_p$, thus the gotten matrix function is $f\left(\boldsymbol{Z}_p\right) = a_0\boldsymbol{I} + a_1\boldsymbol{Z}_p + \cdots + a_n\boldsymbol{Z}_p^n$. Simultaneously, we use a simple linear function $g(x) = b + x\ (b > 0)$ to map the matrix $\boldsymbol{Z}$, the gotten matrix function is $g(\boldsymbol{Z}) = b\boldsymbol{I} + \boldsymbol{Z}$. Thus, the Eq. (13) becomes:

$$J_p(\boldsymbol{U}) = \max_{U} \frac{tr\left(\boldsymbol{U}^T\left(\sum_{k=0}^{n} a_k\boldsymbol{Z}_p^k\right)\boldsymbol{U}\right)}{tr\left(\boldsymbol{U}^T\left(b\boldsymbol{I} + \boldsymbol{Z}\right)\boldsymbol{U}\right)}. \tag{14}$$

It can be reduced the generalized eigenvectors problem:

$$\left(\sum_{k=0}^{n} a_k\boldsymbol{Z}_p^k\right)\boldsymbol{u} = \lambda\left(b\boldsymbol{I} + \boldsymbol{Z}\right)\boldsymbol{u}. \tag{15}$$

Thus, a new MFA method has been presented. Since the polynomial is used to reconstruct the criterion of MFA, this new method is named polynomial marginal fisher analysis (PMFA). The PMFA algorithm is given below:

Algorithm 2 PMFA

1) **Input** a dataset $\boldsymbol{X}$

2) **Constructing the intra-class compactness and inter-class separability matrixes:**

Adjacency matrix: $\boldsymbol{W}_{ij} = \boldsymbol{W}_{ji} = 1$ if x_j is among the k_1 nearest neighbors of x_i in the same class

Similarity matrix: $\boldsymbol{W}_{ij}^p = \boldsymbol{W}_{ji}^p = 1$ if the pair (i, j) is among the k_2 shortest pairs

3) **New criterion:** select two suitable functions $f(x)$ and $g(x)$ to map the matrices $\boldsymbol{X}(\boldsymbol{D}^p - \boldsymbol{W}^p)\boldsymbol{X}^T$ and $(\boldsymbol{X}(\boldsymbol{D} - \boldsymbol{W}))\boldsymbol{X}^T$ to the corresponding matrix functions, get a new criterion function:

$$J(\boldsymbol{U}) = \max_{\boldsymbol{U}} \frac{tr\left(\boldsymbol{U}^T f(\boldsymbol{X}(\boldsymbol{D}^p - \boldsymbol{W}^p)\boldsymbol{X}^T)\boldsymbol{U}\right)}{tr\left(\boldsymbol{U}^T g(\boldsymbol{X}(\boldsymbol{D} - \boldsymbol{W})\boldsymbol{X}^T)\boldsymbol{U}\right)}$$

4) **Compute the eigenvalues and eigenvectors** as:

$$f(\boldsymbol{X}(\boldsymbol{D}^p - \boldsymbol{W}^p)\boldsymbol{X}^T)\boldsymbol{u} = \lambda g(\boldsymbol{X}(\boldsymbol{D} - \boldsymbol{W})\boldsymbol{X}^T)\boldsymbol{u}$$

5) **Output** the final linear projection direction as:

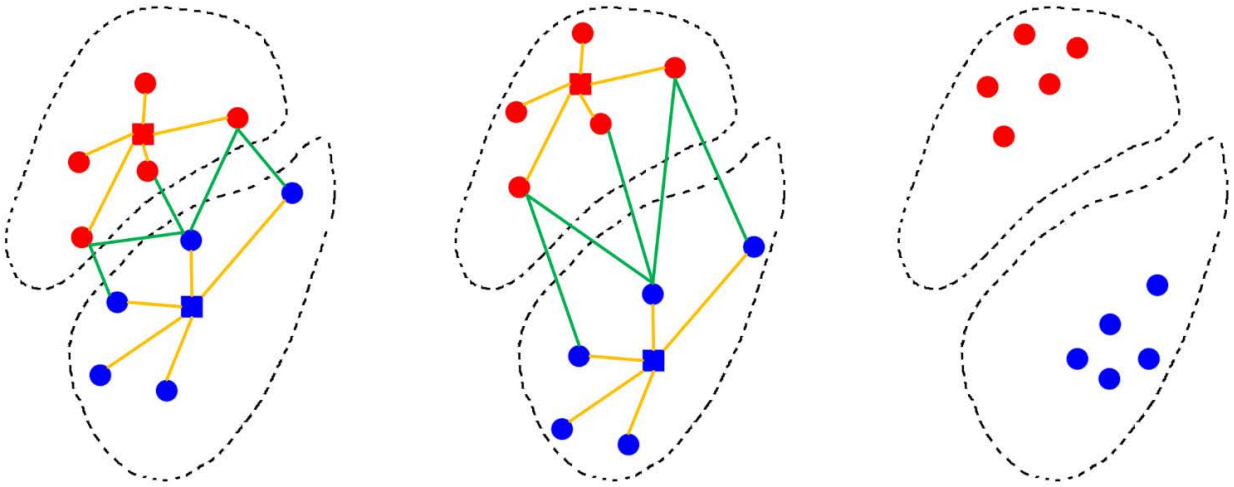
$$\boldsymbol{Y} = \boldsymbol{U}^T\boldsymbol{X}$$

B. THEORETICALLY ANALYSIS

In this section, we theoretically discuss why we chose $f(x) = a_0 + a_1x + \cdots + a_nx^n\ (a_k > 0(k = 0, 1, \cdots, n))$ and $g(x) = b + x\ (b > 0)$ to map matrices $\boldsymbol{Z}_p$ and $\boldsymbol{Z}$.

1) AVOID THE SMALL-SAMPLE-SIZE PROBLEM

After mapping the matrix $\boldsymbol{Z}$ with the linear function $g(x) = b + x\ (b > 0)$, the gotten matrix function is $g(\boldsymbol{Z}) = b\boldsymbol{I} + \boldsymbol{Z}$. Let λ_{wi} be the eigenvalues of the matrix $\boldsymbol{Z}$. We know that the matrix $\boldsymbol{Z}$ is semidefinite according to Section II. A, and so



(a) The initial samples space (b) The sample space mapped by the polynomial matrix function (c) The projected space

FIGURE 1. Geometric interpretation of PMFA.

$\lambda_{wi} \geq 0$. According to Theorem 1, $b+\lambda_{wi}$ are the eigenvalues of the matrix function $g(\boldsymbol{Z}) = b\boldsymbol{I} + \boldsymbol{Z}$. Since $b > 0$, we have $b + \lambda_{wi} > 0$ then the matrix function $g(\boldsymbol{Z}) = b\boldsymbol{I} + \boldsymbol{Z}$ will not have zero eigenvalues and will be nonsingular. Thus, Eq. (15) is solvable. Therefore, the small-sample-size problem of PMFA is avoided.

2) IMPROVE CLASSIFICATION PERFORMANCE

In the MFA method, $\boldsymbol{W}_{ij}^{p}$ is the correlation coefficient between the sample pairs $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ from the different classes, $\boldsymbol{W}_{ij}$ is the correlation coefficient between the samples $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ of the same class. The inter-class distance d_b and the intra-class distance d_w in the sample space can be expressed as:

$$d_b = \frac{1}{2}\sum_{i,j}\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2 \boldsymbol{W}_{ij}^{p} = tr\left(\boldsymbol{X}\boldsymbol{L}^{p}\boldsymbol{X}^{T}\right) = tr\left(\boldsymbol{Z}_p\right), \quad (16)$$

$$d_w = \frac{1}{2}\sum_{i,j}\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2 \boldsymbol{W}_{ij} = tr\left(\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{T}\right) = tr\left(\boldsymbol{Z}\right). \quad (17)$$

Let λ_{bi} and λ_{wi} be the eigenvalues of $\boldsymbol{Z}_p$ and $\boldsymbol{Z}$ respectively, the two distances can be written as:

$$d_b = tr\left(\boldsymbol{Z}_p\right) = \lambda_{b1} + \lambda_{b2} + \cdots \lambda_{bn}, \quad (18)$$

$$d_w = tr\left(\boldsymbol{Z}\right) = \lambda_{w1} + \lambda_{w2} + \cdots \lambda_{wn}. \quad (19)$$

For the PMFA method, the matrix $\boldsymbol{Z}_p$ is mapped into the matrix function $f\left(\boldsymbol{Z}_p\right) = \sum_{k=0}^{n} a_k \boldsymbol{Z}_p^k$ by the polynomial $f(x) = \sum_{k=0}^{n} a_k x^k$, and the matrix $\boldsymbol{Z}$ is mapped into the matrix function $g(\boldsymbol{Z}) = b\boldsymbol{I} + \boldsymbol{Z}$ by $g(x) = b + x\ (b > 0)$. According to Theorem 1, $f(\lambda_{bi}) = \sum_{k=0}^{n} a_k \lambda_{bi}^k$ are the eigenvalues of matrix function $f\left(\boldsymbol{Z}_p\right) = \sum_{k=0}^{n} a_k \boldsymbol{Z}_p^k$, and $g(\lambda_{wi}) = b + \lambda_{wi}$ are the eigenvalues of the matrix function $g(\boldsymbol{Z}) = b\boldsymbol{I} + \boldsymbol{Z}$. After mapping, the new inter-class distance d_b' and the intra-class distance d_w' are:

$$d_b' = tr\left(f(\boldsymbol{Z}_p)\right) = \left(\sum_{k=0}^{n} a_k \lambda_{b1}^k\right) + \left(\sum_{k=0}^{n} a_k \lambda_{b2}^k\right) + \cdots + \left(\sum_{k=0}^{n} a_k \lambda_{bn}^k\right), \quad (20)$$

$$d_w' = tr\left(g(\boldsymbol{Z})\right) = (b + \lambda_{w1}) + (b + \lambda_{w2}) + \cdots + (b + \lambda_{wn}). \quad (21)$$

Usually, the eigenvalues of the matrix $\boldsymbol{Z}_p$, λ_{bi}, take a larger value, so we have $f(\lambda_{bi}) = \sum_{k=0}^{n} a_k \lambda_{bi}^k \gg \lambda_{bi}$ and $g(\lambda_{wi}) = b + \lambda_{wi} \approx \lambda_{wi}$. Then, we have $d_b' \gg d_b$ and $d_w' \approx d_w$.

In this way, with the mapping of polynomial, the PMFA method keeps almost the intra-class distance while greatly enlarging the marginal space between the inter-class samples, which is beneficial to pattern classification.

To illustrate the main idea of the PMFA method visually, we present a geometric interpretation of PMFA in Fig. 1. For convenience, two class examples are used for illustration. The red circle and the blue circle represent two different classes respectively, the red square and the blue square represent the centers of the two different classes. In Fig. 1, orange lines represent the intra-class distance, and green lines represent the inter-class margin distance. Fig. 1(a) shows the initial samples space. In Fig. 1(b), PMFA uses the polynomial function to map the initial samples to the new space where the intra-class distance is almost unchanged, and the inter-class edge distance is enlarged. Fig. 1(c) shows a new space after the projection of the samples.

3) COMPUTATIONAL COMPLEXITY ANALYSIS

The computational complexity of the MFA and PMFA is evaluated in this section. For a given dataset $\boldsymbol{X} \in \boldsymbol{R}^{m\times n}$, m is the dimensionality of training samples, n is the number of the training samples. Both methods need to perform the following two steps:

(1) Construct an intrinsic graph and a penalty graph in $O\left((k_1 + k_2)n^2\right)$.
(2) Solve the eigenvalue problem in $O(m^3)$ [14].

For MFA, the covariance matrix needs to be calculated in $O(m^2n)$ and its eigenvalue decomposition is $O(m^3)$, thus, the complexity is $O\left((k_1+k_2)n^2+m^3+m^2n+m^3\right)$.

For PMFA, it requires the use of the polynomial matrix function transformation in $O(4m^3)$, thus computation complexity is $O\left((k_1+k_2)n^2+m^3+4m^3\right)$.

Usually, one has $n \gg m$, so there is little difference in computational complexity between MFA and PMFA.

C. THE DESIGN OF THE POLYNOMIAL

In this Section, we discuss how to select the polynomial $f(x)$ and $g(x)$ in Section III A. For $g(x) = b + x\ (b > 0)$, the constant b takes usually a little value, empirically, let $b = 0.01$, i.e., $g(x) = 0.01 + x$. The design of the polynomial $f(x)$ is discussed as follows.

It is well-known that, for any function $p(x)$, it has Taylor's expansion

$$p(x) = \frac{p(x_0)}{0!} + \frac{p'(x_0)}{1!}(x-x_0) + \cdots \frac{p^{(n)}(x_0)}{n!}(x-x_0)^n + o\left[(x-x_0)^n\right]$$

When $x_0 = 0$, it is reduced to McLaughlin's expansion

$$p(x) = \frac{p(0)}{0!} + \frac{p'(0)}{1!}x + \cdots + \frac{p^{(n)}(0)}{n!}x^n + o\left[x^n\right]. \tag{22}$$

In real application, the coefficients in Eq. (22) can be simplified to a certain constant c, so Eq. (22) can be rewritten as $p(x) = c\left(1 + x + \frac{x^2}{2!} \cdots + \frac{x^n}{n!}\right)$. According to the previous Section, the smaller the denominator of the polynomial, the larger the eigenvalue obtained by the polynomial $p(x)$ mapping, so the polynomial $p(x)$ can be simplified to $p(x) = 1 + x + \frac{x^2}{2} + \cdots + \frac{x^n}{n}$.

Let's review EMFA [15], which uses exponential function $y = e^x$ to map eigenvalues, and then to enlarge the distance of samples. In this paper, the designed polynomial should be stronger than the exponential function when the distance is extended.

For the polynomial $p(x)$, the larger the n value, the larger the eigenvalue after mapping. The smaller the n, the smaller the influence on the eigenvalues. Thus, we try to let $n = 3, 5, 7$, then the polynomial $p(x)$ becomes

$$s(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3},$$
$$h(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \frac{x^5}{5}, \text{ and}$$
$$f(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \frac{x^5}{5} + \frac{x^6}{6} + \frac{x^7}{7}.$$

In order to study the mapping ability of the functions $y = e^x$, $s(x)$, $h(x)$, and $f(x)$ to eigenvalues, we give their geometric illustration in Fig. 2. As can be seen that when $n = 7$ is, the value $f(\lambda)$ is the largest, so the ability to enlarge the distance is the strongest. However, the greater the n value, the greater the computational complexity of computing the n-order polynomial matrix function. We set $n = 7$, i.e., the seventh-order polynomial $f(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \frac{x^5}{5} + \frac{x^6}{6} + \frac{x^7}{7}$ is chosen.

Now, we have $f(x) = \sum_{k=0}^{7} \frac{x^k}{k}$ and $g(x) = 0.01 + x$. when mapping eigenvalues, we have $f(\lambda_{bi}) = \sum_{k=0}^{7} \frac{\lambda_{bi}^k}{k} \gg e^{\lambda_{bi}}$, and $g(\lambda_{wi}) = 0.01 + \lambda_{wi} \approx \lambda_{wi}$. Then, we have

$$\frac{f(\lambda_{bi})}{g(\lambda_{wi})} = \frac{\sum_{k=0}^{7} \frac{\lambda_{bi}^k}{k}}{0.01 + \lambda_{wi}} > \frac{e^{\lambda_{bi}}}{e^{\lambda_{wi}}} > \frac{\lambda_{bi}}{\lambda_{wi}}. \tag{23}$$

According to this analysis, the classification ability of PMFA should be better than that of EMFA, and much better than that of MFA.

TABLE 2. The comparisons of distance of MFA and PMFA.

	AR	ORL	FERET	PIE
d_b	367.59	8.27	159.91	143.68
d'_b	858.38	20.81	489.50	392.60
d_w	70.91	8.15	37.07	83.53
d'_w	71.72	8.29	37.22	83.93

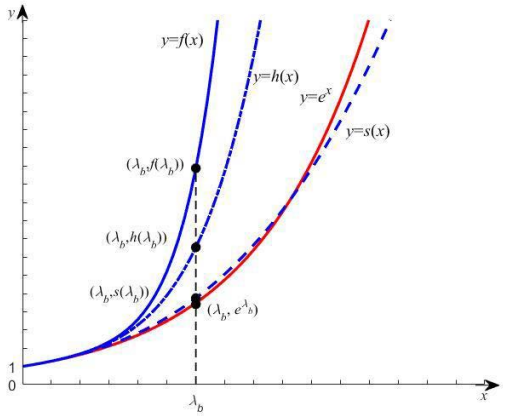
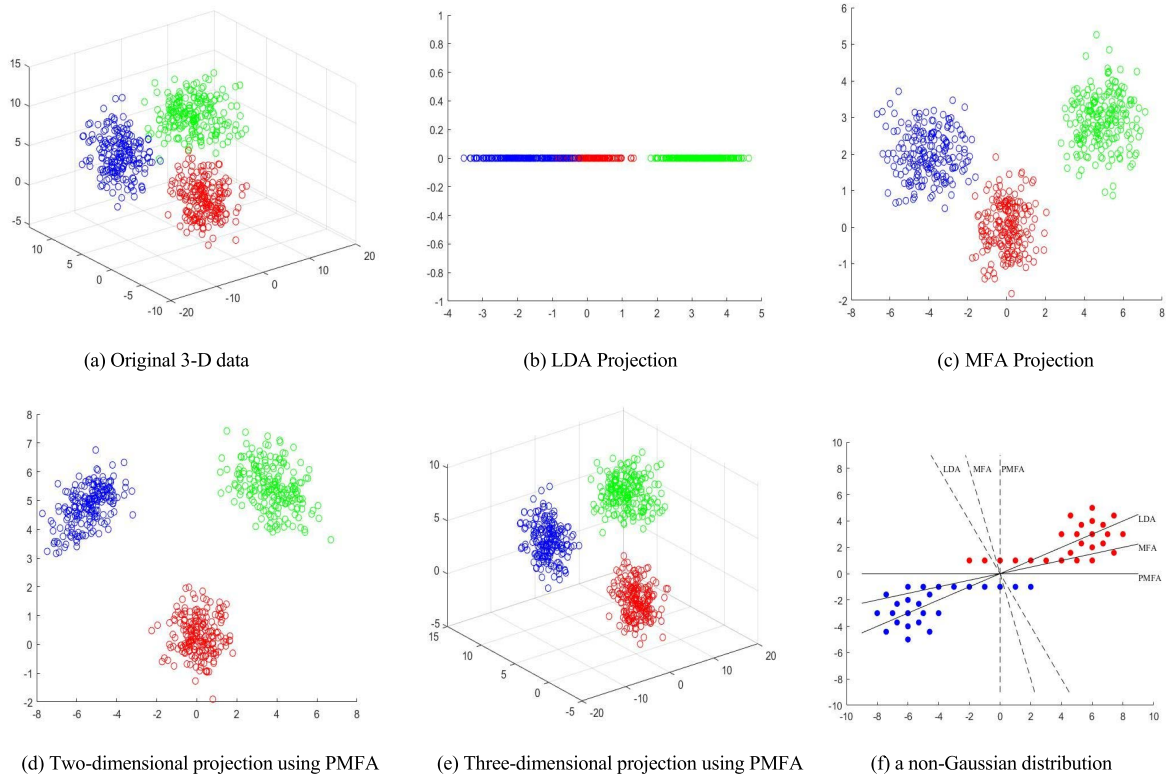


FIGURE 2. Geometric illustration of the functions.

D. AN EXPERIMENT OF THE DISTANCE DIFFUSION

After selecting the polynomial $f(x)$ and $g(x)$, we provide an experiment to illustrate the distance diffusion effect of PMFA in four face datasets: AR [28], ORL [29], FERET [30], and PIE [31]. In the experiment, for each data set, the original samples are normalized to avoid the large value, and $k_1 = 3$, $k_2 = 8$. The inter-class distance d_b and the intra-class distance d_w of MFA are calculated by Eq. (18) and (19), and the inter-class distance d'_b and the intra-class distance d'_w of PMFA are calculated by Eq. (20) and (21), where the first ten largest eigenvalues are used. Table 2 shows the comparisons of the results of the MFA and PMFA methods on four face datasets. As can be seen, compared with MFA, PMFA increases the inter-class distance and maintains the intra-class distance.



(a) Original 3-D data (b) LDA Projection (c) MFA Projection

(d) Two-dimensional projection using PMFA (e) Three-dimensional projection using PMFA (f) a non-Gaussian distribution

FIGURE 3. **The embedding of synthesized data.**

IV. EXPERIMENTAL RESULTS

A. EXPERIMENTAL SETUP

In the experiments, to verify the validity of the proposed PMFA method, the classical LDA [2], the original MFA [11], and the latest improvement of MFA: MMFA (Multiple MFA) [14], and two methods that solved the SSS problem of MFA: EMFA (Exponential MFA) [15] and RMFA (Regularized MFA) [16] are compared with the proposed PMFA method.

For the proposed PMFA method, according to section III C, two polynomial functions are selected to map the matrices $\boldsymbol{Z}_p$ and $\boldsymbol{Z}$ to the corresponding matrix functions:

$$f(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \frac{x^5}{5} + \frac{x^6}{6} + \frac{x^7}{7} \text{ and}$$
$$g(x) = 0.01 + x.$$

In the further experiment, the PMFA method is also compared with the latest proposed manifold-based learning algorithms, including GDE [18], CR-DLPP [19], LMGE-DDR [20], GEU-MFA-U, and GEU-MFA-S [17].

B. DATA VISUALIZATION

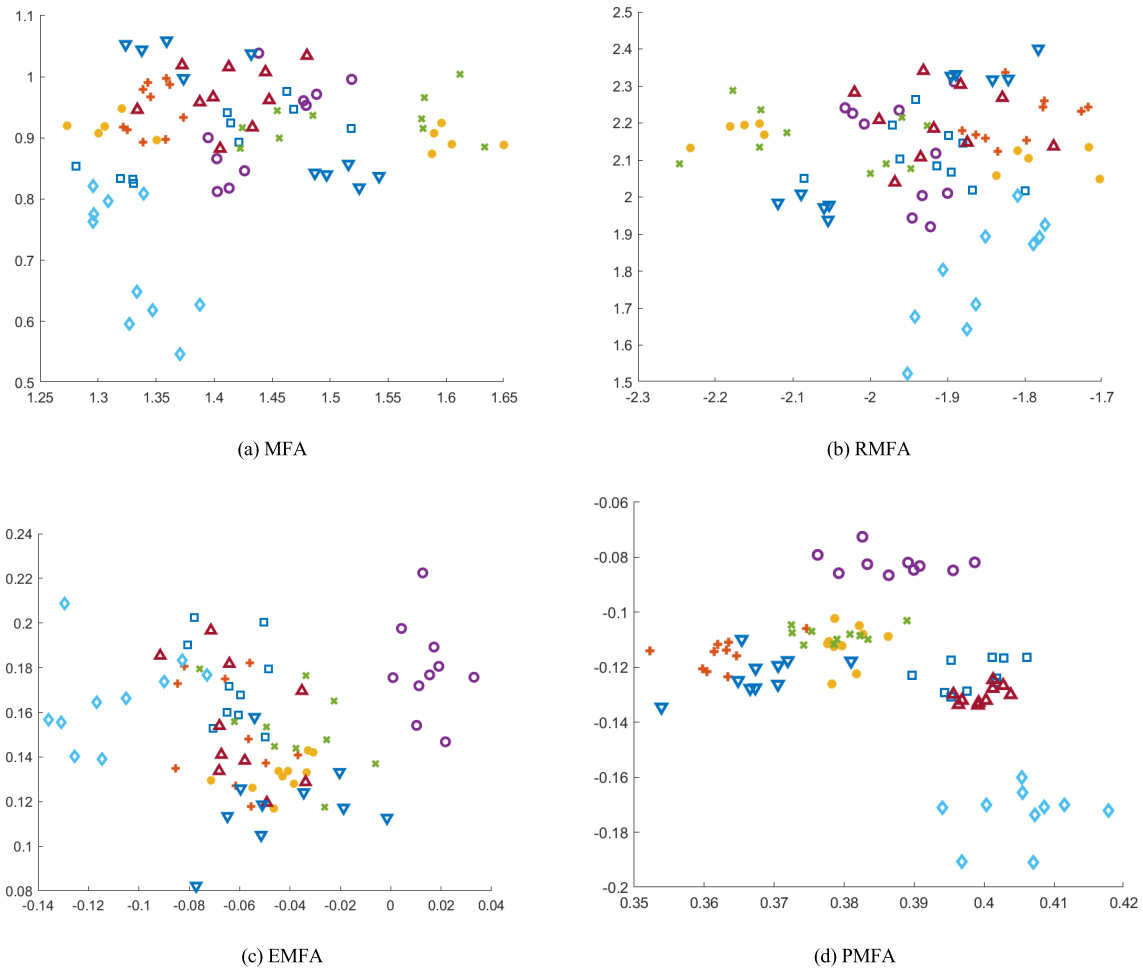
To illustrate the classification performance of PMFA, we conduct data visualization experiments on the synthetic data set (Gaussian and non-Gaussian distribution) and the real face data set.

The synthetic data of Gaussian distribution is a 3-dimensional dataset in Fig. 3(a). This three-class set contains 600 points. Each class is generated using a single Gaussian.

Fig. 3(b) shows the projection in a 1-D subspace using LDA, Fig. 3(c), (d) is the 2-D projection of MFA and PMFA, respectively. Fig. 3(e) shows the projection using PMFA in a 3-D subspace. As we can see, that both MFA and PMFA provide prediction data with good distinction capabilities. We can also see that, compared with MFA, PMFA makes the intra-class samples more compact, and the marginal distance of the inter-class samples larger.

In Fig. 3(f), a binary classification problem shows the classification ability of LDA, MFA, and PMFA in the case of non-Gaussian distribution. Where, the red solid circles and blue solid circles are two different classes of synthetic data, which do not follow the Gaussian distribution. The solid lines represent optimal classification lines and the dotted lines represent the optimal projection directions learned from LDA, MFA, and PMFA, respectively. The results show that: (1) in the case of non-Gaussian distribution, LDA does not work well, but MFA and PMFA can still find the best projection directions; (2) the best projection direction learned by PMFA is better than that of MFA, because PMFA not only considers the edge points but also enlarges the distance of the inter-class samples.

Fig. 4 shows the 2-D projection on the Georgia Tech face database using the MFA, RMFA, EMFA, and PMFA methods. It can be observed that PMFA is more effective than MFA, RMFA and EMFA.



(a) MFA (b) RMFA (c) EMFA (d) PMFA

FIGURE 4. The two-dimension projection of Georgia Tech face database. (a) MFA. (b) RMFA (c) EMFA. (d) PMFA.

TABLE 3. Details of the datasets.

Dataset	Number of classes	Sample number of each class Number of samples per class	Image size	Data dimension
ORL	40	10	32×32	1024
Yale	15	11	24×24	576
Georgia Tech	126	14	32×32	1024
AR	50	15	40×50	2000

C. FACE RECOGNITION

In this part, we conduct experiments on four face databases: ORL, Yale, AR, and Georgia Tech [32]. Fig. 5 shows some of the sample pictures from the datasets. It showed the details of the datasets used in the experiment in Table 3.

In the experiment, the matrix $\boldsymbol{Z}_p$ and $\boldsymbol{Z}$ are normalized to their norm, then mapping $\boldsymbol{Z}_p$ with $f(x)$, and adds 0.01 to the diagonal element of $\boldsymbol{Z}$. For each class in all the datasets, we select p samples as the training samples and the rest as the test samples. For the p training samples, to get stable results, the experiment is repeated three times and the p samples are randomly selected for each time. In each experiment, for the p training sample, the dimension of subspace range between 10 and 100 with the step size being 5. The recognition rate corresponding to the optimal subspace dimension is the optimal recognition rate. Therefore, for the three experiments, there are three optimal recognition rates. Finally, the average

TABLE 4. The recognition accuracy, standard deviation and optimal dimensions of the ORL dataset.

Method	4 trains	5 trains	6 trains
LDA	91.67±1.21(30)	93.33±1.64(35)	95.23±1.98(39)
MFA	92.78±1.97(55)	95.83±2.02(65)	97.92±1.44(60)
RMFA	95.39±1.05(55)	96.83±1.61(55)	97.86±1.30(85)
EMFA	95.97±1.88(55)	97.00±1.73(40)	98.13±0.98(65)
MMFA	94.41±1.61(55)	97.24±1.46(95)	98.09±2.01(60)
PMFA	**96.25±0.42(100)**	**98.00±0.50(70)**	**98.54±1.57(60)**

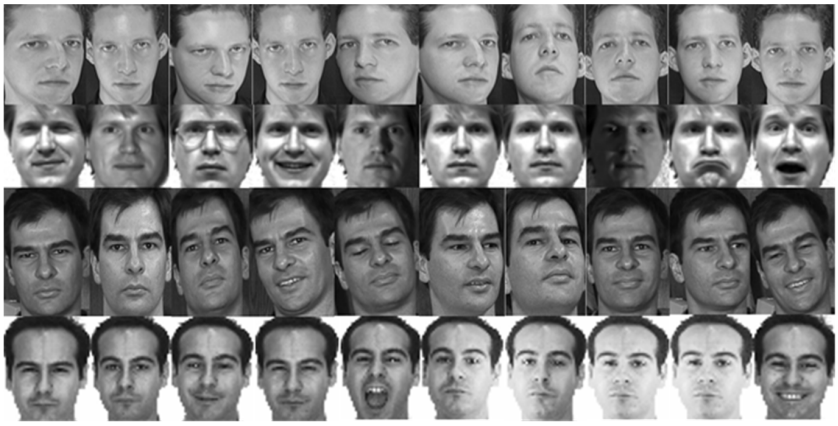


FIGURE 5. The sample pictures taken from the datasets in the experiment. The first line is from the ORL face dataset, the second line is from the Yale face dataset, the third line is from the Georgia Tech face dataset, and the last line is from the AR face dataset.

TABLE 5. The recognition accuracy, standard deviation and optimal dimensions of the Yale dataset.

Method	4 trains	5 trains	6 trains
LDA	67.30±6.56(100)	75.30±7.81(35)	81.00±7.19(35)
MFA	79.36±2.40(75)	81.48±3.80(50)	86.66±5.81(25)
RMFA	80.90±1.96(35)	81.56±1.44(20)	88.67±3.34(65)
EMFA	79.56±0.55(20)	80.74±1.24(50)	83.11±6.01(25)
MMFA	79.92±0.84(75)	81.76±2.97(65)	87.96±5.36(40)
PMFA	**82.54±2.20(75)**	**82.96±3.71(50)**	**90.22±6.84(25)**

TABLE 6. The recognition accuracy, standard deviation and optimal dimensions of the Georgia Tech dataset.

Method	6 trains	7 trains	8 trains
LDA	70.13±0.97(30)	72.50±1.50(49)	73.87±2.04(60)
MFA	77.78±2.94(90)	79.33±0.95(100)	82.38±1.19(100)
RMFA	78.67±1.39(60)	79.58±2.13(100)	81.76±1.58(90)
EMFA	78.29±1.12(50)	79.83±0.80(80)	83.34±1.35(65)
MMFA	78.37±1.86(40)	79.90±0.49(60)	83.06±1.71(30)
PMFA	**80.15±2.93(40)**	**82.25±1.73(35)**	**85.33±2.22(65)**

TABLE 7. The recognition accuracy, standard deviation and optimal dimensions of the AR dataset.

Method	3 trains	4 trains	5 trains
LDA	90.15±2.25(80)	95.03±0.69(100)	96.79±0.49(100)
MFA	95.47±2.32(85)	98.39±0.21(85)	98.70±0.43(85)
RMFA	95.88±0.42(45)	96.75±(0.54)90	97.01±0.52(95)
EMFA	95.99±1.00(85)	97.28±0.76(85)	98.03±0.54(95)
MMFA	96.06±1.55(75)	98.13±1.14(75)	98.56±0.45(95)
PMFA	**98.03±0.35(85)**	**98.69±0.42(85)**	**99.07±0.34(80)**

value is used as the recognition rate when the training sample is p. Tables 4-7 show the recognition results (recognition accuracy (%) $\pm$ standard deviation and optimal dimension) of these methods.

We also evaluate the performance of these methods when the subspace dimension takes different values. In each experiment, for a training sample p and a subspace dimension, there is a recognition rate. When the subspace dimension is between 10 and 100, the recognition rate of each method in each dimension can be got, Figs. 6-9 show how the recognition rate varies with the dimension.

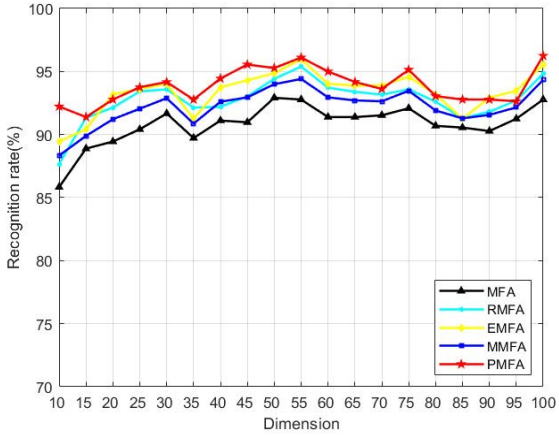


FIGURE 6. Comparison of performance and dimension on ORL dataset (training sample $p = 4$).

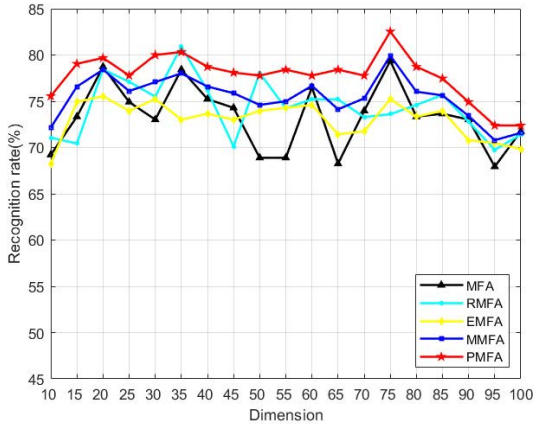


FIGURE 7. Comparison of performance and dimension of the Yale dataset (training sample $p = 4$).

TABLE 8. The comparison of recognition rates (%) on PIE, AR, Yale, FERET and Yale B datasets.

Methods	PIE			AR		Yale			FERET		Yale B
	$p=6$	$p=7$	$p=8$	$p^1=7$	$p^2=7$	$p=4$	$p=5$	$p=6$	$p=5$	$p=6$	$p=10$
MFA	93.16	93.71	93.14	95.44	96.39	73.65	77.04	82.22	88.58	88.50	88.77
EMFA	93.71	93.37	93.38	96.54	96.87	73.96	75.19	80.00	83.38	79.33	86.68
GDE [18]	59.80	69.46	82.72	71.90	71.55	-	-	-	-	-	-
LMGE-DDR [20]	91.35	-	93.31	-	-	-	-	-	89.71	90.73	-
CR-DLPP [19]	93.96	-	-	-	-	76.57	80.81	83.38	87.68	88.52	-
GEU-MFA-U [17]	-	-	-	-	-	-	-	-	-	-	92.20
GEU-MFA-S [17]	-	-	-	-	-	-	-	-	-	-	92.10
PMFA	94.26	94.29	94.02	97.90	96.51	77.46	81.48	84.89	89.92	91.33	92.43

[1] the training selects the first seven images of each class, and the remaining images are used for testing
[2] the training selects the last seven images of each class, and the rest are used for testing

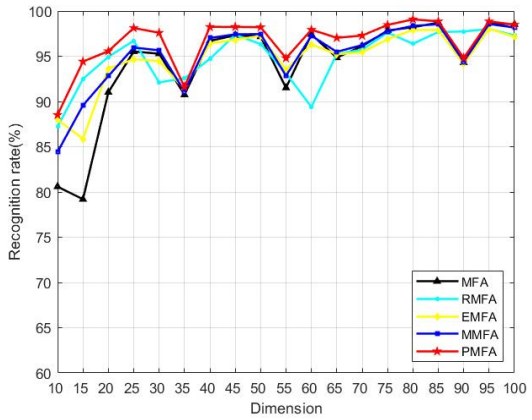


FIGURE 8. Comparison of performance and dimension of the AR dataset (training sample $p = 5$).

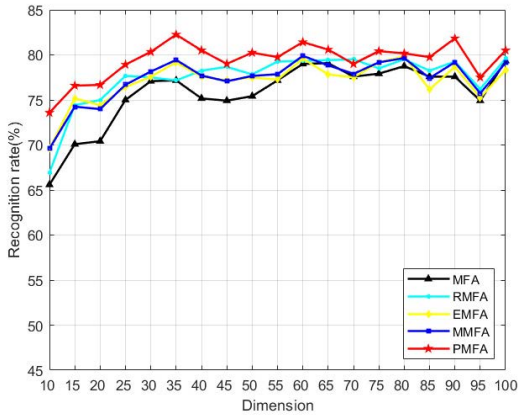


FIGURE 9. Comparison of performance and dimension of the Georgia Tech dataset (training sample $p = 7$).

D. THE FURTHER EXPERIMENTS

In this section, we also compare the PMFA method and other latest proposed manifold-based learning algorithms on PIE, AR, Yale, FERET and Yale B [33] datasets in Table 8. These including GDE [18], CR-DLPP [19], LMGE-DDR [20], GEU-MFA-U, and GEU-MFA-S [17]. Where GEU-MFA-U and GEU-MFA-S are the latest methods of MFA. We use the recognition rates reported in the literature for comparison, because the source code of these methods is not available. The results show that the recognition rate of the PMFA method is better than the latest methods.

V. CONCLUSION AND FUTURE DIRECTIONS

In this paper, based on the framework of matrix function dimensionality reduction, when the polynomial matrix function transformation is used in the MFA method, PMFA (Polynomial Marginal Fisher Analysis) is proposed. The proposed PMFA method has two advantages: one is that it solves the small-sample-size problem of MFA, and the other is that it can enlarge the distance of the inter-class sample, then have a good ability for pattern classification. However, the computational complexity and classification ability are closely related to the selected polynomial function, so enough research is needed when choosing an appropriate polynomial. In the future, the design of the polynomial functions can be improved, and the idea of the polynomial matrix function transformation can also be used to some other dimensionality reduction methods.

REFERENCES

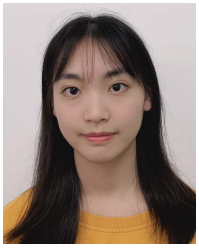
[1] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, Jan. 1991.
[2] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
[3] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, Dec. 2000.
[4] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
[5] M. Belkin and P. Niyogi, "Laplacian genotypes and spectral techniques for embedding and clustering," in *Proc. Conf. Adv. NIPS*, vol. 14, no. 1, 2002, pp. 585–591.
[6] K. Q. Weinberger, F. Sha, and L. K. Saul, "Learning a kernel matrix for nonlinear dimensionality reduction," in *Proc. 21st Int. Conf. Mach. Learn. (ICML)*, 2004, p. 106.
[7] J. Yu, M. Tan, H. Zhang, Y. Rui, and D. Tao, "Hierarchical deep click feature prediction for fine-grained image recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 2, pp. 563–578, Jul. 2019.
[8] X. F. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. NIPS*, Vancouver, BC, Canada, vol. 16, 2004, pp. 153–160.
[9] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proc. 10th IEEE Int. Conf. Comput. Vis. (ICCV)*, Beijing, China, Oct. 2005, pp. 1208–1213.
[10] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. Roux, and M. Ouimet, "Out-of-sample extensions for LLE, Isomap, MDS, Eigenmaps, and spectral clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 16, 2003, pp. 177–184.

[11] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2006.
[12] Y. L. Yu and L. M. Zhang, "Orthogonal MFA and Uncorrelated MFA," *Pattern Recognit. Artif. Intell.*, vol. 21, no. 5, pp. 603–608, 2008.
[13] S. Siena, V. N. Boddeti, and B. V. K. V. Kumar, "Coupled marginal Fisher analysis for low-resolution face recognition," in *Proc. Eur. Conf. Comput. Vis.* Berlin, Germany: Springer, 2012, pp. 240–249.
[14] Z. Huang, H. Zhu, J. T. Zhou, and X. Peng, "Multiple marginal Fisher analysis," *IEEE Trans. Ind. Informat.*, vol. 66, no. 12, pp. 9798–9807, Dec. 2019.
[15] S.-J. Wang, S. Yan, J. Yang, C.-G. Zhou, and X. Fu, "A general exponential framework for dimensionality reduction," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 920–930, Feb. 2014.
[16] K. Huang, "Regularized marginal Fisher analysis and sparse representation for face recognition," *J. Comput. Appl.*, vol. 33, no. 6, pp. 1723–1726, Oct. 2013.
[17] F. Laakom, J. Raitoharju, N. Passalis, A. Iosifidis, and M. Gabbouj, "Graph embedding with data uncertainty," *IEEE Access*, vol. 10, pp. 24232–24239, 2022.
[18] P. Huang, T. Li, G. Gao, and G. Yang, "Feature extraction based on graph discriminant embedding and its applications to face recognition," *Soft Comput.*, vol. 23, no. 16, pp. 7015–7028, Aug. 2019.
[19] T. Su, D. Feng, and H. Hu, "Collaborative representation based discriminant local preserving projection," *Neural Process. Lett.*, vol. 2022, pp. 1–28, Apr. 2022.
[20] Y. Tian and X. Feng, "Large margin graph embedding-based discriminant dimensionality reduction," *Sci. Program.*, vol. 2021, pp. 1–12, Dec. 2021.
[21] R. S. Ran, J. Feng, S. Zhang, and B. Fang, "A general matrix function dimensionality reduction framework and extension for manifold learning," *IEEE Trans. Cybern.*, vol. 99, pp. 1–12, 2020.
[22] Z. Wang, L. Zhang, and B. Wang, "Sparse modified marginal Fisher analysis for facial expression recognition," *Int. J. Speech Technol.*, vol. 49, no. 7, pp. 2659–2671, Jul. 2019.
[23] N. Zhang, M. Ruan, S. Wang, L. Paul, and X. Li, "Discriminative few shot learning of facial dynamics in interview videos for autism trait classification," *IEEE Trans. Affect. Comput.*, early access, May 30, 2022, doi: 10.1109/TAFFC.2022.3178946.
[24] X. Cai, J. Chai, Y. Gao, S. Li, and B. Zhu, "Deep marginal Fisher analysis based CNN for image representation and classification," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 181–189.
[25] G. H. Golub and C. F. V. Loan, *Matrix Computation*, 4th ed. Baltimore, MD, USA: Johns Hopkins Univ. Press, 2013, pp. 513–515.
[26] R. Ran, Y. Ren, S. Zhang, and B. Fang, "A novel discriminant locality preserving projections method," *J. Math. Imag. Vis.*, vol. 2021, pp. 541–554, Feb. 2021.
[27] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York, NY, USA: McGraw-Hill, 1976, pp. 159–161.
[28] A. Martinez and R. Benavente, "The AR face database," CVC, Tech. Rep. 24, Jun. 1998.
[29] (1994). Olivetti & Oracle Research Laboratory. *The Olivetti & Oracle Research Laboratory Face Database of Faces*. [Online]. Available: https://www.cam-orl.co.uk/facedatabase.html
[30] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
[31] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database," in *Proc. 5th IEEE Int. Conf. Autom. Face Gesture Recognit.*, May 2002, pp. 53–58.
[32] A. V. Nefian. (1999). *Georgia Tech Face Database*. [Online]. Available: http://www.anefian.com/research/face_reco.htm
[33] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.

RUISHENG RAN received the B.S. degree in mathematics from Chongqing Normal University, Chongqing, China, and the M.S. degree in computational mathematics and the Ph.D. degree in computer application technology from the University of Electronic Science and Technology of China, Chengdu, China.

He is currently a Professor with the College of Computer and Information Science, Chongqing Normal University. His research interests include pattern recognition and computer vision.

ZHENG LI is currently pursuing the master's degree with the College of Computer and Information Science, Chongqing Normal University, Chongqing, China. His research interests include machine learning and computer vision.



TING WANG is currently pursuing the master's degree with the College of Computer and Information Science, Chongqing Normal University, Chongqing, China. Her research interests include machine learning and computer vision.

JI FENG received the Ph.D. degree in computer science from Chongqing University, China, in 2016. He is currently a Lecturer at the Computer and Information Science College, Chongqing Normal University, China. His research interests include artificial intelligence and pattern recognition and applications.

BIN FANG (Senior Member, IEEE) received the B.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, the M.S. degree in electrical engineering from Sichuan University, Chengdu, China, and the Ph.D. degree in electrical engineering from The University of Hong Kong, Hong Kong.

He is currently a Professor with the Department of Computer Science, Chongqing University, Chongqing, China. He has published more than 100 technical papers. His research interests include computer vision, pattern recognition, medical image processing, biometrics applications, and document analysis. He is an Associate Editor of the *International Journal of Pattern Recognition and Artificial Intelligence*.