Multimedia Tools and Applications (2022) 81:20781–20796
https://doi.org/10.1007/s11042-022-12708-3

Robust sparse manifold discriminant analysis

Jingjing Wang[1] · Zhonghua Liu[1] · Kaibing Zhang[2] · Qingtao Wu[1] · Mingchuan Zhang[1]

Received: 1 December 2020 / Revised: 4 March 2021 / Accepted: 21 February 2022 /
Published online: 12 March 2022

Abstract
Classical linear discriminant analysis (LDA) has been applied to machine learning and pattern recognition successfully, and many variants based on LDA are proposed. However, the traditional LDA has several disadvantages as follows: Firstly, since the features selected by feature selection have good interpretability, LDA has poor performance in feature selection. Secondly, there are many redundant features or noisy data in the original data, but LDA has poor robustness to noisy data and outliers. Lastly, LDA only utilizes the global discriminant information, without consideration for the local discriminant structure. In order to overcome the above problems, we present a robust sparse manifold discriminant analysis (RSMDA) method. In RSMDA, by introducing the $L_{2,1}$ norm, the most discriminant features can be selected for discriminant analysis. Meanwhile, the local manifold structure is used to capture the local discriminant information of the original data. Due to the introduction of $L_{2,1}$ constraints and local discriminant information, the proposed method has excellent robustness to noisy data and has the potential to perform better than other methods. A large number of experiments on different data sets have proved the good effectiveness of RSMDA.

Keywords Linear discriminant analysis · Robust sparse manifold discriminant analysis · Manifold learning · Feature selection

1 Introduction

Feature selection and extraction, which have received extensive attention in recent years, play an important role in pattern classification [10, 14, 41]. However an image raw data contains a large number of redundant features and noise, which make difficult for image recognition and image analysis. [18, 19, 27]. In this case, for classification tasks, how to select and extract the different categories of the most significant features in the whole exercise is among the most

✉ Zhonghua Liu
ZhonghuaLiu@haust.edu.cn

[1] Information Engineering College, Henan University of Science and Technology, Luoyang, China

[2] College of Electronics and Information, Xi'an Polytechnic University, Xi'an, China

difficult. It has proven that feature selection and extraction are effective tools in the field of machine learning and pattern classification. They can reduce the complexity, increase the efficiency and enhance the classification performance [20, 25, 26].

Whether it is feature selection or feature extraction, to a certain extent, they both are subspace learning methods [2, 11, 13]. They have common goal in finding a low-dimensional representation of the original high-dimensional data in a new learning space [28, 30, 33].

Various methods of feature extraction have been proposed in the past few decades. In this branch, principal component analysis (PCA) [38] is one of the most famous method, its main idea is to try to learn a projection that can preserve the main energy of the original data. Local preserving projection (LPP) [7], sparse preserving projection (SPP) [24] and neighborhood preserving embedding (NPE) [8] which they learn their projections are also feature extraction methods, these methods consider the local manifold geometry of the original data and try to preserve the local information in the projection space [4, 16, 29]. Subsequently, classifiers, such as K Nearest Neighbor (KNN) [42] and Support Vector Machine (SVM) [3], are usually used for classification. Although the above methods don't use the label information of the data, have different advantages, the classification performance of these algorithms are not good enough to some extent.

In fact, we hope that dimensionality reduction can be achieved for data with category labels (supervised), and each category can be better distinguished after dimensionality reduction. At this time, another classic feature extraction algorithm-linear discriminant analysis (LDA) appeared [43]. LDA is a classic linear learning method. The idea of linear discrimination is simple but very effective. LDA can greatly expand the distance between classes and reduce the distance within classes. And through the use of label information to learn discriminant projection, which improves the classification accuracy [9, 17, 40].To enhance performance and efficiency, many variants based on LDA are proposed [5, 12, 15] such as orthogonal LDA (OLDA) [31], unrelevant LDA (ULDA) [32]. OLDA and ULDA transform the image into a vector for learning projection, which aim to solve the problem of small LDA sample size. However, these methods, which based on LDA in the calculation of the L_2 norm scatter matrix, may cause the error seriously, and these methods are sensitive to outliers [1, 34, 44]. Later, sparse linear discriminant analysis (SLDA) [23] and sparse uncorrelated linear discriminant analysis (SULDA) [39] are proposed to learn sparse discriminant subspace for feature extraction. Robust Adaptive Linear Discriminant Analysis (RALDA) [6] method achieved an appropriate latent subspace for data representation where $L_{2,1}$ norm is adopted in the formulations of loss function,which the regularization term can reduce the impact of outliers and noise and predict the select discriminative features. Later, Ning et.al. proposed a method named BULDP: Biomimetic Uncorrelated Locality Discriminant Projection for Feature Extraction in Face Recognition, it is based on unsupervised discriminating projection and two human bionic where in: homologous and isomeric principle of continuity principle of similarity [21]. Later, Ning et.al. proposed Real-time 3D Face Alignment Using an Encoder-Decoder Network with an Efficient Deconvolution Layer [22]. Zhang et.al. proposed a new iterative reweight-based log-sum constraint channel estimation scheme. it used the structure sparsity of the mmWave channels by formulating the channel estimation problem as an objective optimization problem. [35].Later, Zhang et.al. proposed Block-Sparsity Log-Sum-Induced Adaptive Filter for Cluster Sparse System Identification The main idea of the proposed scheme is to add a new block-sparsity induced term into the cost function of the LMS algorithm [36]. In addition, he proposed Block-Sparsity Log-Sum-Induced Adaptive

Filter for Cluster Sparse System Identification [37] . The main idea is to add a sparsity lp norm penalty cost function of the LMS algorithm.

However, LDA has some obvious disadvantages. Firstly, each new feature combined with others, which the most of projection coefficients are not zero. The learned projection matrix cannot explain the features well. This also shows that LDA cannot select the most useful function from redundant data. Secondly, LDA selects k feature vectors as the projection of feature extraction, and the selected k feature vectors will have a one-to-one correspondence with the first k smallest feature values, and the number of k is related to the data. For dimensionality reduction, this makes the choice of the classification accuracy of linear discriminant analysis sensitive. Thirdly, LDA, and many methods based on linear discriminant analysis are sensitive to noise.

In order to overcome this problem, a new robust sparse manifold discriminant analysis (RSMDA) algorithm is proposed for dimension reduction. Especially, the proposed RSMDA method uses $L_{2,1}$ norm based sparse constraint to choose the important feature. At the same time, the manifold based local discrimination information is added on the basis of the original linear discrimination analysis. The innovations of this article mainly include the following aspects.

(1) The most useful and discriminative features can be selected by introducing the $L_{2,1}$ norm.
(2) Compared with other methods based on linear discriminant analysis, the proposed method is more robust to noise.
(3) By introducing the local discriminative information, the sparse projection can use the local identification information to enhance the discrimination of projection.

The remaining chapters of this article are specifically arranged as follows. Section 2 mainly reviews some related work. The third part describes the robust discriminant analysis of sparse manifolds in detail and gives the optimal iterative scheme of RSMDA. In Section 4, There are a large number experiments on different libraries to prove the good performance. Section 5 gives the corresponding conclusion.

2 Related work

In this section, we will briefly review the research work, which mainly includes LDA and manifold based local discriminant information learning. For convenience, Table 1 introduce some notations used through the paper.

2.1 Linear discriminant analysis (LDA)

Suppose there are c pattern classes, n_i represents the number of samples of the i^{th} class, $n = \sum_{i=1}^{c}$ is the total number of all samples, column vector $x_j^i \in R_m$ denotes the j^{th} sample of the i^{th} class. LDA tries to find a projection matrix, which makes the samples in the same category as close as possible, and makes the samples in different categories are as far away as possible. LDA uses the following fisher to obtain this projection vector.

$$a = \arg \max_a \frac{a^T S_b a}{a^T S_w a} \tag{1}$$

Table 1 Significant Notations Annotated in This Paper

Notations	Description
c	classes
n	Input data samples' number
m	Input data samples' dimension
x_i	Represents the i-th data samples
S_b	inter-class scatter matrix
S_w	intra-class scatter matrix
u_i	i-th row of $\boldsymbol{U}$
u	small positive constant
L	Laplacian matrix
D	Diagonal matrix
W	Weight matrix
w_i	i-th row of $\boldsymbol{W}$
w_{ij}	j-th element of w_i
d_i	i-th row of $\boldsymbol{D}$
G_w	within-class graph
G_b	between-class graph
P	orthogonal reconstruction matrix of $m \times d$
Q	discriminative projection matrix of $m \times d$
E	denotes random noise

where S_b is the inter-class scatter matrix, and S_w is the intra-class scatter matrix. S_b and S_w are calculated as following.

$$S_b = \frac{1}{n}\sum_{i=1}^{c} n_i (u_i - u)(u_i - u)^T \tag{2}$$

$$S_w = \frac{1}{n}\sum_{i=1}^{c} n_i \sum_{j=1}^{n_i} \left(x_j^i - u_i\right)\left(x_j^i - u_i\right)^T \tag{3}$$

where $u_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_j^i$ is the mean feature of samples of the ith class, $u_i = \frac{1}{n_i}\sum_{i=1}^{c}\sum_{j=1}^{n_i} x_j^i$ is the mean feature of all samples. Generally, problem (1) is equivalent to the following optimization problem .

$$\begin{aligned} a = \operatorname{argmin} & a^T (S_w - \mu S_b) a \\ s.t.\ & a^T a = I \end{aligned} \tag{4}$$

where μ is a small positive constant.

By solving Eq.(4), we can observe that the optimal projection vector a, it is the eigenvector corresponding to the minimum eigenvalue of $S_w a = \mu S_b a$. Generally, a single projection vector is not enough to distinguish multiple classes. In real world applications, we usually select a set of projection vectors which satisfy the optimal Fisher criterion $A = \arg \min_{A^T A = 1} \mathrm{Tr}\left(A^T (S_w - \mu S_b)\, A\right)$ for multiclass classification. The projection matrix X is selected as a set of eigenvectors corresponding to the first k smallest eigenvalues of $S_w A = \lambda S_b A$. Let $\mathrm{A} = [\mathrm{a}_1, \mathrm{a}_2, .., \mathrm{a}_k] \in \mathrm{R}^{m \times k}$ be the set of the selected k eigenvectors, we can obtain discriminative feature vector $y_j^i \in R^k$ of each sample by $y_j^i = A^T x_j^i$.

2.2 Local manifold learning

The nearest neighbor graph G is first constructed and its weight matrix is defined as follows:

$$W_{ij} = \begin{cases} 1, & if\ x_i \in N(x_j)\ or\ x_j \in N(x_i) \\ 0, & otherwise \end{cases} \tag{5}$$

where $N(\cdot)$ is a set of k nearest neighbors, the nearest neighbor graph G is divided into two graphs, i.e., within-class graph G_w and between-class graph G_b for extracting local discriminant information of samples. For each sample x_i, its k nearest neighbors are split into two subsets, $N_w(x_i)$ and $N_b(x_i)$. $N_w(x_i)$ contains the neighbors sharing the same label with x_i, whereas $N_b(x_i)$ contains the neighbors that have different labels. Let W^w and W^b be the weight matrices of G_w and G_b, respectively. Let W^w and W^b be the weight matrices of G_w and G_b, respectively. In each of the graphs defined above, if two samples have a nonzero weight, then they are referred to as connected samples. It is clear to see that $W = W^w + W^b$. In this way, the conversion matrix A can map samples from the original sample space to the label space.The result is that the connected samples of G_w stay as close together as possible, while the connected samples of G_b stay as far as possible. Given a map as $f_i = x_iA$ ($i = 1, \ldots, n$), it should satisfy the following objective functions:

$$\min \sum_{ij} \left\| f_i - f_j \right\|^2 W_{ij}^w \tag{6}$$

$$\max \sum_{ij} \left\| f_i - f_j \right\|^2 W_{ij}^b \tag{7}$$

In Eq. (6), if x_i and x_j are close and have the same label, then f_i and f_j should be close as well. In Eq. (7), if x_i and x_j are close but have different labels, then f_i and f_j should be far from each other. (6) and (7) can be converted into

$$\min \sum_{ij} \left\| f_i - f_j \right\|^2 W_{ij}^w = \min Tr\left(A^T X^T L_w X A\right) \tag{8}$$

$$\max \sum_{ij} \left\| f_i - f_j \right\|^2 W_{ij}^b = \max Tr\left(A^T X^T L_b X A\right) \tag{9}$$

where L_w and L_b are the Laplacian matrix of G_w and G_b. They are defined as $L_w = D^w - W^w$, $L_b = D^b - W^b$. D^w and D^b are diagonal matrices, which their diagonal entries are $D_{ii}^w = \sum_j W_{ij}^w$ and $D_{ii}^b = \sum_j W_{ij}^b$.

Finally, the Eqs. (8) and (9) can be rewritten as

$$\min_A \left(Tr\left(A^T X^T L_w X A\right) - Tr\left(A^T X^T L_b X A\right) \right) = \min_A Tr\left(A^T X^T (Lw - L_b) X A\right) \tag{10}$$

3 Robust sparse manifold discriminant analysis (RSMDA)

In this section, the motivation of our RSMDA is firstly introduced. Then the optimization solution of RSMDA is given.

3.1 The motivation of RSMDA

The main object of LDA is projection. The advantage of this algorithm is that the projection, it shortens the projection distance of the same type of sample, and it can also increase the distance of different types of samples. But the LDA also has its design defects. Most of the projection coefficients are non-zero, and the new feature samples are linear combinations of all sample features, which leads to the lack of good explanatoryness of the projective matrix to the characteristics, which is reflected in the following aspects. First of all, LDA does not have a good explanation for matrix features, but it cannot select the most suitable function in a large amount of redundant data. Second, when LDA selects k eigenvector corresponding to the first k minimum eigenvalue as a feature extraction projection, it is greatly affected by the data, and the LDA data classification varies greatly. This leads to great differences in LDA classification accuracy of different sizes. Third, most LDA-based implementation methods cannot be ignored by external environmental impacts, such as many methods that are vulnerable to noise. In this paper, we propose a robust sparse manifold discriminant analysis (RSMDA) to solve the above problems.

At the same time, in practical applications, the acquired data is generally high-dimensional. Data contains a large amount of redundant information, and some noise can corrupt data or images. Therefore, the choice of those essential features from the original complex data discriminant analysis in the whole learning process is vital, because it can effectively reduce the negative impact caused by redundant features. Applying a sparse norm constraint projection allows the model to perform feature selection, such as the $L_{2,1}$ norm. The $L_{2,1}$ norm has good line sparsity compared with the L_1 norm, precisely because of its property, which makes it easier for the projection to interpret elements. Sparse projection discrimination information can grasp the local identification of the projection to extend. At the same time, we are concerned about the situation of random noise. We use this term to compensate for the noise sparse so that can reduce the negative impact to some extent. Inspired by this motive, we recommend learning more powerful by using the following constraints discrimination subspace

$$\min_{P,Q,E} Tr\left(Q^T(S_w - \mu S_b)Q\right) + \lambda_1\|Q\|_{2,1} + \lambda_2\|E\|_1 \; s.t. X = PQ^TX + E, P^TP = I \quad (11)$$

where $Q \in R^{m \times d}(d < m)$ is divided into scattering matrices between different classes and within the same class. λ_1 and λ_2 are trade-off parameters, μ is a small positive constant used to balance the importance of S_b and S_w. LDA realizes maximizing the interclass scattering matrix S_w and minimizing the internal class scattering matrix S_b according to the balanced optimization projection matrix. E represents errors and it is used to simulate random noise. $\|\cdot\|_1$ is the L_1 norm. In some cases, $X = PQ^TX$ and $P^TP = I$ can be regarded as variants of PCA, which has the advantage that the original data can be recovered well. $P \in R^{m \times d}$ is an orthogonal reconstruction matrix. By reconfiguring the relationship between the sample and the original sample that is considering conversion, on a reduced dimension, the transformed data can be retained as much as possible the main energy of the original data. In this way, RSMDA not only learn discriminant subspace, but also learn optimization framework with the minimum loss of information, it is possible to perform better.

The research shows that the discriminant analysis method based on global structure information ignores the local information of the image. On the basis of manifold technology, the original image is reduced and dimensionally processed, the local data manifold structure in the nonlinear sub-manifold can be better maintained. By introducing the manifold based local discriminative information, the objective function of the proposed RSMDA can be rewritten as follows.

$$\begin{aligned} \min_{P,Q,E} & TQ^T(S_w - \mu S_b)Q + \mathrm{Tr}Q^T X^T(L_w - L_b)XQ \\ & + \lambda_1\|Q\|_{2,1} + \lambda_2\|E\|_1 \\ s.t. & X = PQ^T X + E,\ P^T P = I \end{aligned} \tag{12}$$

3.2 Optimization of RSMDA

In this section, we propose an iterative algorithm to update the rules to solve the optimal solution in eq. (12). Since the objective function is non-convex and contains four different variables. It's difficult for us to get the global optimal solution. Therefore, we can obtain the local optimal solution through continuous iteration. First of all, we use the Lagrange function to convert the problem (12) to the following form.

$$\begin{aligned} L(P,Q,E,Y) = & Tr\ (Q^T(Sw - \mu S_b)Q\ ) + Tr\ (Q^T X^T(L_w - L_b)QX\ ) \\ & + \lambda_1\|Q\|_{2,1} + \lambda_2\|E\|_1 + \langle Y, X - Q^T X - E\rangle \\ & + \frac{\beta}{2}\left\|X - PQ^T X - E\right\|_F^2 \\ = & Tr\ (Q^T\ (S_w - \mu S_b\ )Q\ ) + Tr\big(Q^T X^T\ (L_w - L_b)QX\ \big) \\ & + \lambda_1\|Q\|_{2,1} + \lambda_2\|E\|_1 - \frac{1}{2\beta}\|Y\|_F^2 \\ & + \frac{\beta}{2}\left\|X - PQ^T X - E + \frac{Y}{\beta}\right\|_F^2 \end{aligned} \tag{13}$$

where β is a penalty parameter, Y is the Lagrangian multiplier.Then P, Q, E can be alternately solved by minimizing the Lagrangian function L with other variables fixed. The solution scheme is as follows.

- Fix other variables to update Q

we fix P, E and update Q by solving the following problem

$$\begin{aligned} L(Q) = & Tr\big(Q^T(S_w - \mu S_b)Q\big) + TrQ^T X^T(L_w - L_b)QX \\ & + \lambda_1\|Q\|_{2,1} \\ & + \frac{\beta}{2}\left\|X - Q^T X - E + \frac{Y}{\beta}\right\|_F^2 \end{aligned} \tag{14}$$

Define $X - E + \frac{Y}{\beta} = M$, Q can be calculated by the derivative of L(Q) with respect to Q

$$\frac{\partial L(Q)}{\partial Q} = 2(Sw - \mu Sb) + \lambda_1 DQ + \beta\big(XX^T Q - XM^T\big) \tag{15}$$

where D is defined as q_i the ith row of Q and $Q = \begin{bmatrix} q_1 \\ \vdots \\ q_m \end{bmatrix}$. Let $\partial L(Q)/\partial Q = 0$, then we obtain

$$Q = \left(2(S_w - \mu S_b) + \lambda_1 D + \beta X X^T\right)^{-1}\left(\beta X M^T\right) \tag{16}$$

- Fix other variables to update P

We fix Q E and update P by minimizing the following problem

$$\min_{P^T P = I} \left\| X - P Q^T X - E + \frac{Y}{\beta} \right\|_F^2 \tag{17}$$

Let $X - E + \frac{Y}{\beta} = M$. The optimization (17) is converted to

$$\begin{aligned} &\min_{P^T P = I} \left\| M - P Q^T X \right\|_F^2 = \min_{P^T P = I} Tr\left(M^T M - 2 M^T P Q^T X\right) = \min_{P^T P = I} Tr\left(M^T P Q^T X\right) \\ &= \min_{P^T P = I} Tr\left(P^T M X^T Q\right) \end{aligned} \tag{18}$$

Problem (18) is an Orthogonal Procrustes problem and it can be simply solved. Suppose $SVD(MX^TQ) = USV^T$, then P is obtained as $P = UV^T$, where SVD represents the singular value decomposition operation.

- Fix other variables to update E

We fix P, Q and update E by solving the following problem.

$$\min_E \lambda_2 \|E\|_1 + \frac{\beta}{2} \left\| X - P Q^T X - E + \frac{Y}{\beta} \right\|_F^2 \tag{19}$$

If we define $\varepsilon = \frac{\lambda_2}{\beta}$ and $E_0 = X - Q^T X + \frac{Y}{\beta}$, in according to shrinkage calculator, problem (19) has the closed solution as follows.

$$E = shrink(E_0, e) \tag{20}$$

where *shrink* means the shrinkage calculator.

- Fix other variables to update Lagrange multiplier

Y and β are respectively updated by using the following formulas

$$Y = Y + \beta\left(X - P Q^T X - E\right) \tag{21}$$

$$\beta = \min(\rho\beta, \beta_{\max}) \tag{22}$$

where ρ and $\beta_{\max}$ are the constant.

3.3 Computational complexity and convergence analysis of RSMDA

We analyze the computational complexity of RSMDA. the major computational cost is the matrix inverse operation. For a m × m matrix, the computational complexity of inverse operation is $O(m^3)$. The whole computational complexity of the proposed method is $O(\tau(m^2n + m^3 + 2\max(m^2; mn)d + d^3))$, where τ is the iteration number. For simplicity, we suppose that $m \gg n$, thus the computational complexity of the proposed method is $O(\tau(m^2n + m^3 + 2m^2d + d^3))$.

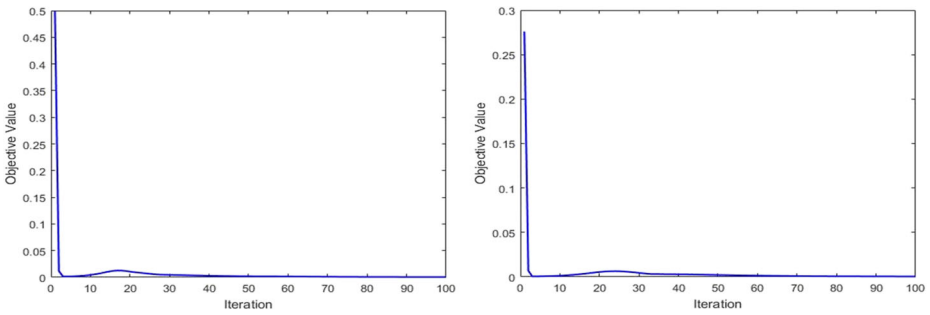
In this section, two databases are selected to analyze the convergence of the proposed RSMDA. The optimization solution process of RSMDA is realized by iteratively updating three variables. It can be seen from the experimental results in Fig. 1 that the RSMDA algorithm is convergent and converges to the local optimal solution.

4 Experiments

Some experiments have been carried out in this section to prove the good performance of the proposed RSMDA algorithm. This includes classification accuracy, convergence to global optimality, and the effectiveness of high-dimensional image maintenance. In this section, five public image databases are selected to evaluate the effectiveness of the proposed method, KNN and some supervised learning methods, including SVM, LDA, SLDA, OLDA, ULDA, SULDAare chose to compare with the proposed method. At the same time, the accuracy of all test results (AC) is used as a unified evaluation standard. The standard is based on the percentage and actual results of the correct classification results.

4.1 Experiments on the Yale B face database

There are 2432 face images from 38 subjects in the Expanded Yale B Face database. 64 images of faces under different lighting conditions were provided for each subject, and we manually converted each image to a 32 × 32 Gy scale image. One experiment, as show in Fig. 2. which 10, 15, 20 and 25 samples were randomly selected as training samples and the rest as test samples was repeated 10 times, and the classification accuracy of each algorithm was show in Table 2.



(a) Yale (b) ORL

Fig. 1 The convergence curve of the proposed NMF_ASGR on the two data sets. (a) Yale, (b) ORL

Fig. 2 Some typical images of the Extended Yale B face database

Table 2 shows the experimental results obtained by different methods on the extended Yale B database. It can be seen that the proposed RSMDA method has better performance than other supervised learning methods, especially KNN, SVM and LDA.

4.2 Experiments on the CMU PIE face database

There are 41,368 face images from 38 subjects in the CMU PIE face database. These facial images were obtained from 68 subjects in different poses and different lighting conditions. In this experiment, we used a subset of the CMU PIE face database, which contains 11,554 images from 68 subjects. We manually convert each image into a 32 × 32 Gy scale image. Figure 3 shows the sample image of one of them. In this experiment, the training set was the first 10, 15, 20, and 25 images of each person, and the test set was the remaining images. Repeat the algorithm 10 times. The classification accuracy of each algorithm is shown in Table 3.

It can be seen from Table 3 that as the number of training samples increases, the recognition rates of KNN, SVM, LDA, SLDA, OLDA, ULDA, SULDA and the proposed RSMDA are steadily increasing. Still, no matter how many samples are, the proposed methods perform better than other contrast algorithms. This proves that RSMDA can capture as much discriminative information as possible, which is better than these comparison methods to some extent.

4.3 Experiments on the AR face database

The AR face database contains color face images of 120 people, and the total number of face images exceeds 4000. Among them, 120 subjects were photo graphed twice with different

Table 2 Classification accuracy (%) of different methods on the Yale B database

Methods	The number of the training samples per class			
	10	15	20	25
KNN	34.28	42.83	49.17	55.40
SVM	63.90	75.97	82.55	86.30
LDA	81.09	87.53	90.57	92.41
SLDA	85.74	90.41	92.60	93.92
OLDA	87.38	91.32	93.20	94.33
ULDA	85.74	88.15	90.79	92.41
SULDA	82.97	88.11	90.72	92.50
RSMDA	**87.60**	**91.64;**	**93.59**	**94.57**

Bold indicates the highest recognition accuracy

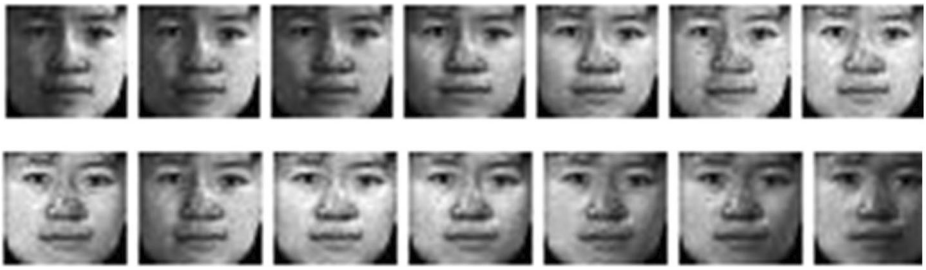


Fig. 3 Some faces images one object from CMU PIE database

facial expressions, light conditions and shade, with a 14-day interval, and each person produced 26 images. In our experiment, out of 26 facial images of 120 people, seven were selected from each stage, or 14 face images per person. We manually converted each image to 50 by 40 pixels. Figure 4. shows the sample image of one of them. In this experiment, the training set is the first 4, 6, 8, and 12 images of each person, and the test set is the remaining images. Repeat the algorithm 10 times. The classification accuracy of each algorithm is shown in Table 4.

Table 4 shows the experimental results of different methods on the AR face database. It is clear from the Table 4 that when the number of training samples increases from 6 to 8, the classification accuracy of ULDA and SULDA is decreasing, while the classification accuracy of RSMDA is steadily improving, and no matter how the number of training samples changes, this method achieves the best performance in many methods.

4.4 Experiments on ORL database

The ORL database includes a total of 400 facial images, collected from 40 people, and each person provides 10 facial images. And the images were taken at different times and under a different light. The image content includes different facial expressions and facial details. Figure 5. below shows a sample image of one of them. In this experiment, the training set is the first 3, 4, 5, and 6 images of each person, and the test set is the remaining images. Repeat the algorithm 10 times. The classification accuracy of each algorithm is shown in Table 5.

We can clearly see from Table 5 that the proposed RSMDA is superior to KNN, SVM, LDA, SLDA, OLDA,ULDA and SULDA.

Table 3 Classification accuracy (%) of different methods on the CMU PIE database

Methods	The number of the training samples per class			
	10	15	20	25
KNN	43.29	50.77	56.08	59.86
SVM	72.34	81.40	86.38	89.24
LDA	82.01	87.57	90.24	91.94
SLDA	83.77	88.97	91.74	93.31
OLDA	86.18	90.38	92.56	93.78
ULDA	82.49	88.20	91.02	92.63
SULDA	84.61	88.72	91.66	92.14
RSMDA	**87.46**	**88.72**	**93.26;**	**94.53**

Bold indicates the highest recognition accuracy

Fig. 4 Sample faces images from AR database

Table 4 Classification accuracy (%) of different methods on the AR database

Methods	The number of the training samples per class			
	4	6	8	10
KNN	53.88	62.92	69.15,	77.43
SVM	69.51	82.75	89.22	95.51
LDA	87.33	93.60	95.56	97.47
SLDA	89.83	94.00	95.83	97.38
OLDA	90.11	94.35	96.08	97.37
ULDA	86.16	92.56	91.02	97.02
SULDA	87.34	93.75	96.14	95.95
RSMDA	**90.48**	**94.57**	**96.94**	**98.17**

Bold indicates the highest recognition accuracy

4.5 Experiments on Georgia Tech database

The Georgia tech face database contains photos of 50 people taken during two or three sessions. The faces in these pictures may be front and tilted, or they may be front or tilted.

Fig. 5 Sample faces images from ORL face database

Table 5 Classification accuracy (%) of different methods on the ORL database

Methods	The number of the training samples per class			
	3	4	5	6
KNN	56.46	61.77	69.23	76.52
SVM	67.88	82.43	90.30	93.21
LDA	88.66	94.20	95.87	97.92
SLDA	89.96	95.10	95.83	98.33
OLDA	88.23	92.87	94.49	97.31
ULDA	88.71	94.07	95.79	97.41
SULDA	90.11	94.27	96.14,	97.59
RSMDA	**91.56**	**95.85**	**96.88**	**98.41**

Bold indicates the highest recognition accuracy

These images include different expression, illumination and proportion. Each image is manually cropped to 60 by 50 pixels. Be converted to grayscale images. Figure 6 shows the sample image of one of them. In this experiment, the training set was the first 5, to 8 images of each person, and the test set is the remaining images. Repeat the algorithm 10 times. The classification accuracy of each algorithm is shown in Table 6.

Table 6 shows the RSMDA method has better performance than KNN, LDA, OLDA and ULDA, and the recognition rate is higher. Compared with SULDA, the classification accuracy

Fig. 6 Sample faces images from Georgia Tech face database

Table 6 Classification accuracy (%) of different methods on the Georgia Tech database

Methods	The number of the training samples per class			
	5	6	7	8
KNN	81.91	86.58	89.31	92.72
SVM	83.95	90.16	93.05	95.45
LDA	77.57	79.13	87.56	93.33
SLDA	84.96	89.80	92.47	95.71
OLDA	77.10	84.61	89.16	93.28
ULDA	53.94	70.85	80.84	88.13
SULDA	74.12	75.08	67.50	82.08
RSMDA	**85.63**	**91.11**	**93.34**	**95.92**

Bold indicates the highest recognition accuracy

of RSMDA steadily increases with the increase in the number of training samples.This also shows that RSMDA is more stable than SULDA to a certain extent.

4.6 Parameters selection

There are two regularization parameters, which can affect the performance of RSMDA. In the following, we examine the effects of these two parameters λ_1 and λ_2 on the proposed algorithm by examining the changes in the recognition performance of RSMDA under different parameter values. We choose two databases as the test set, they are CMU PIE and Yale B databases. The experimental results of the algorithm are shown in Fig. 7. We take (10^{-5},10^{-4}, 10^{-3},10^{-2},10^{-1},1, 10^{1}, 10^{2}, 10^{3}, 10^{4}, 10^{5}) as the value range of the two regularization parameters λ_1 and λ_2, then execute the proposed method RSMDA. As can be seen from the Fig. 7, the proposed method performs best when the value λ_1 and λ_2 are close to 0.0001. Differences in the classification results show that the two parameters for the identification and classification of learning and performance projection algorithm has an important influence.

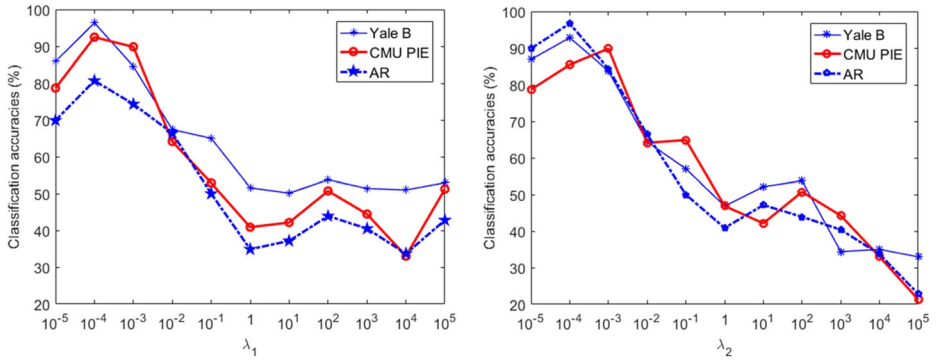


Fig. 7 The Classification accuracy (%) of the proposed method versus parameters λ_1 and λ_2. on the Yale B database, CMU PIE database and AR database

5 Conclusion

In this paper, we propose a novel supervised feature extraction method termed robust sparse manifold discriminant analysis (RSMDA), in which the global discriminative information, local manifold discriminative information and sparse representation are integrated into a framework. By using the $L_{2,1}$ sparse norm to limit the discriminative projection matrix, the proposed method can perform feature selection and feature extraction at the same time, and these features are more suitable for classification tasks because they have the most discriminative information. This reconstruction constraint minimizes the loss of difference information, thereby improving the accuracy of classification. In addition, the local identification information is introduced, which further enhances the discrimination of the extracted projection matrix. The experimental results on ~~six~~ five databases all show that this method performs better than other competing methods.

Acknowledgements This work was supported by Natural Science Foundation of China (U1504610, 61971339, 61471161), the Key Project of the Natural Science Foundation of Shanxi Province (2018JZ6002), Scientific and Technological Innovation Team of Colleges and Universities in Henan Province (20IRTSTHN018), the Doctoral Startup Foundation of Xi'an Polytechnic University (BS1616). the Natural Science Foundations of Henan Province (202300410148).

Declarations

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

References

1. Andekah Z, Naderan M, Akbarizadeh G (2017) Semi-supervised Hyperspectral image classification using spatial-spectral features and superpixel-based sparse codes, in: 25th Iranian Conference on Electrical Engineering (ICEE), pp 2229–2234
2. Belkin M, Niyogi P (2019) Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput 15:1373–1396
3. Chang C, Lin C(2011) Libsvm: a library for support vector machines. ACM transactions on intelligent systems and technology(TIST). Pp 1-27.
4. Zhu F, Gao J, Yang J, Ye N (2022) Neighborhood linear discriminant analysis. Pattern Recognition 123:108422
5. Gao G, Yu Y, Yang M, et al. (2020). Cross-resolution face recognition with pose variations via multilayer locality-constrained structural orthogonal procrustes regression. Information science, pp 19-36.
6. Guo J, Gao J, Hu Y (2020) Robust adaptive linear discriminant analysis with bidirectional reconstruction constraint. ACM, Transactions on Knowledge Discovery from Data 14(6):5
7. He X, Niyogi P (2004) Locality preserving projections. Neural Inf Process Syst:153–160
8. He X, Cai D, Yan S, et al(2005) Neighborhood preserving embedding. IEEE international conference on computer vision, pp 1208-1213.
9. Lai Z, Mo D, Wong W (2018) Robust discriminant regression for feature extraction. IEEE Trans Cybernet 48:2472–2484
10. Li Y, Tian X, Liu T (2018) On better exploring and exploiting task relationships in multitask learning: joint model and feature learning. IEEE TransNeural Netw Learn 29:1975–1985
11. Liu T, Huang G (2018) An adaptive graph learning method based on dual data representations for clustering. Parttern Recognition. pp 126–139 .
12. Liu Z, Liu G, Pu J(2017) Noisy label based discriminative least squares regression and its kernel extension for object identification, KSII trans Internet Inf Syst. pp 2523-2538.
13. Liu Z, Liu G, Pu J, et al.(2018) Orthogonal sparse linear discriminant analysis. International journal of systems science. Pp 848-858.
14. Liu Z, Shi K, Zhang K, Ou W, Wang L (2020) Discriminative sparse embedding based on adaptive graph for dimension reduction. Eng Appl Artif Intell 94:103758
15. Liu J, Song C, Zhao J,(2020) Manifold-preserving sparse graph-based ensemble FDA for industrial label-noise fault classification. IEEE Trans Instrum Meas. pp 2621-2634

16. Lu X, Wang Y (2017) Graph-regularized low-rank representation for destriping of hyperspectral images. IEEE Trans Geosci Remote Sens 51(7):4009–4018
17. Lu Y, Lai Z, Xu Y, Li X, Zhang D, Yuan C (2016) Low-rank preserving projections. IEEE Trans Cybernet 46:1900–1912
18. Luo M, Nie F , Chang X(2018) Adaptive unsupervised feature selection with structure regularization, IEEE trans. Neural Netw. Learn. Pp 944-956.
19. Mishra G (2020) Constrained L-1-optimal sparse representation technique for face recognition, Optics Laser Technol. pp 1975–1985 .
20. Modava M, Akbarizadeh G (2019) Coastline extraction from SAR images using spatial fuzzy clustering and the active contour method. Int J Remote Sens 38:355–370
21. Ning X, Li W,Tang B, He H (2018) BULDP: biomimetic uncorrelated locality discriminant projection for feature extraction in face recognition , IEEE Trans Image Process
22. Ning X, Li W, Tang B, He H (2020) Real-time 3D face alignment using an encoder-decoder network with an efficient deconvolution layer. Letters, IEEE Signal Processing
23. Qiao Z, Zhou L, Huang J(2009) Sparse linear discriminant analysiswith applications to high dimensional low sample size data. Int J Appl Math. pp 48-60.
24. Qiao L, Chen S, Tan X (2010) Sparsity preserving projections with applications to face recognition. Pattern Recogn 43:331–341
25. A. Sellami, M. Farah, I. Farah, et al.(2017) Hyperspectral imagery classification based on semi-supervised 3-D deep neural network and adaptive band selection, expert Syst. Pp 246–259 .
26. Sharifzadeh F, Akbarizadeh G, Kavian Y (2019) Ship classification in SAR images using a new hybrid CNN-MLP classifier. J Indian Soc Remote Sens 47:551–562
27. Taibi F, Akbarizadeh G, Farshidi E (2019) Robust reservoir rock fracture recognition based on a new sparse feature learning and data training method, Multidimension. Syst. Signal Process. pp 2113–2146 .
28. Wen J, Fang X, Cui J (2019) Robust sparse linear discriminant analysis. IEEE Trans Circuits Syst Video Technol 29:390–403
29. Liu Z, Lu Y, Lai Z, Ou W, Zhang K (2021) Robust sparse low-rank embedding for image reduction. Applied Soft Computing 113:20211129
30. Yang W, Wang Z, Sun C (2015) A collaborative representation based projections method for feature extraction, Pattern Recognit. pp 20–27.
31. Ye J, Xiong T(2006) Null space versus orthogonal linear discriminantanalysis. International conference on machine leaning, pp 1073-1080.
32. Ye J, Janardan R, Li Q (2006) Feature reduction via generalized uncorrelated linear discriminant analysis. IEEE Trans Knowl Data Eng. pp 1312–1322
33. Yu W, Zhang M, Shen Y(2019) Learning a local manifold representation based on improved neighborhood rough set and LLE for hyperspectral dimensionality reduction, signal process. Pp 20-29.
34. Zeng Z, Wang X, Yan F(2019) Local adaptive learning for semi-supervised feature selection with group sparsity, Knowl-Based Syst. pp 181.
35. Zhang A (2020) Channel estimation for MmWave massive MIMO with hybrid precoding based on log-sum sparse constraints. Transactions on Image Processing, IEEE
36. Zhang A (2020) Block-sparsity log-sum-induced adaptive filter for cluster sparse system identification. Access, IEEE
37. Zhang A (2020) Reweighted l p constraint LMS-based adaptive sparse channel estimation for cooperative communication system. IET Communications, IEEE Access
38. Zhang Z, Wang J, Zha H (2012) Adaptive manifold learning. IEEE Trans Pattern AnalMach Intell 34:253–265
39. Zhang X, Chu D, Tan R(2016) Sparse uncorrelated linear discriminant analysis for undersampled problems, IEEE Trans Neural Netw Learn Syst, pp 1469-1485.
40. Zhang J, Luo Z, Li C(2019) Manifold regularized discriminative feature selection for multi-label learning, pattern Recognit. Pp 136-150.
41. Zhang L, Liu Z, Pu J (2020) Adaptive graph regularized nonnegative matrix factorization for data representation. Appl Intell 50(2):438–447
42. Zhou T, Peng Y (2020) Kernel principal component analysis-based Gaussian process regression modelling for high-dimensional reliability analysis. Computers and Structures. pp. 241
43. Zhou Y, Sun S (2017) Manifold partition discriminant analysis. IEEE Trans Cybern 47:830–840
44. Zhu X, Ghahramani Z, Lafferty J (2013) Semisupervised learning using gaussian fields and harmonic functions. In: Machine learning, proceedings of the twentieth international conference (ICML). Washington, DC, USA, pp 21–24