Joint Sparse Locality Preserving Projections

Haibiao Liu, Zhihui Lai(✉), and Yudong Chen

College of Computer Science and Software Engineering, Shenzhen University, Shen Zhen, Guang Dong, China
laizhihui@szu.edu.cn

Abstract. Manifold learning and feature selection have been widely studied in face recognition in the past two decades. This paper focuses on making use of the manifold structure of datasets for feature extraction and selection. We propose a novel method called Joint Sparse Locality Preserving Projections (JSLPP). In order to preserve the manifold structure of datasets, we first propose a manifold-based regression model by using a nearest-neighbor graph, then the $L_{2,1}$-norm regularization term is imposed on the model to perform feature selection. At last, an efficient iterative algorithm is designed to solve the sparse regression model. The convergence analysis and computational complexity analysis of the algorithm are presented. Experimental results on two face datasets indicate that JSLPP outperforms six classical and state-of-the-art dimensionality reduction algorithms.

Keywords: Manifold learning · Face recognition · Dimensionality reduction
Feature selection · Sparse feature extraction

1 Introduction

Dimensionality reduction is one of the most important topics in pattern recognition, machine learning and data mining [1–5]. Due to the curse of dimensionality, it's time-consuming to calculate the Euclidean distance between samples. In order to eliminate the redundant features and preserve meaningful features, many dimensionality reduction methods were proposed. Among them, feature extraction and feature selection are the two most important techniques. The purpose of feature extraction methods is to transform the original high-dimensional data into low-dimensional features by using a linear transformation matrix [1]. Therefore, feature extraction is also known as subspace learning. The classical subspace learning methods including Multiple Dimensional Scaling (MDS) [2], Principle Component Analysis (PCA) [3] and Linear Discriminant Analysis (LDA) [4].

MDS, PCA and LDA only consider the global information and fail to discover the underlying manifold structure of the datasets. Compared with the global Euclidean structure of the datasets, the intrinsic manifold structure embedded in the original high-dimensional space is more effective for pattern recognition [8].

Different from the KPCA and KLDA, many nonlinear manifold learning methods such as Isomap [5], Locally Linear Embedding (LLE) [6, 7], and Laplacian Eigenmap [8] can preserve the manifold structure in low-dimensional subspace with lower

M. Qiu (Ed.): SmartCom 2017, LNCS 10699, pp. 125–133, 2018.
https://doi.org/10.1007/978-3-319-73830-7_13

computational cost. However, these non-linear manifold learning methods lack of robustness and they fail to evaluate the map on testing data. Therefore, these nonlinear manifold learning techniques might not be suitable for some pattern recognition tasks including face recognition. To overcome the drawbacks, Locality Preserving Projections (LPP) [9, 10] and Neighborhood Preserving Embedding (NPE) [11] were proposed. LPP and NPE are the linear extensions of the traditional manifold learning methods and they were widely used in many applications because of their effectiveness and efficiency.

The above methods only focus on feature extraction and thus they all lack of the ability of feature selection. It's known that feature selection is also an important way to improve the performance on pattern recognition. An effective approach to obtain feature selection is to impose a regularization term on the model. For example, Bradley and Mangasarian proposed L_1-SVM for binary classification task [12]. Wang *et al.* proposed A Hybrid Huberized SVM (HHSVM) [13] combining both L_1-norm and L_2-norm regularization term for sparse feature selection. Unlike the L_1-norm regularization, $L_{2,1}$-norm regularization can generate jointly sparse projection matrix which has better explanation for the selected features. In order to perform subspace learning and feature selection simultaneously, Gu *et al.* proposed feature selection and subspace learning (FSSL) by imposing the $L_{2,1}$-norm on the graph embedding framework [14].

Motivated by previous researches [9, 14], in this paper, we propose a novel method called Joint Sparse Locality Preserving Projections. We construct a graph based regression model and then impose $L_{2,1}$-norm regularization term for feature selection. The main contributions of this paper are as follows:

(1) We propose a novel method called Joint Sparse Locality Preserve Projections (JSLPP) which combines manifold learning and feature selection techniques. We construct a regression model and impose $L_{2,1}$-norm regularization term on the modified regression model for feature selection. In the meantime, we design an iterative algorithm to solve the problem and obtain the optimal solution.
(2) We present a comprehensive theoretical analysis for the iterative algorithm, including the convergence analysis and computational complexity analysis.
(3) Experiments show that JSLPP performs better than the existing subspace learning and feature selection methods.

The rest of this paper is organized as follows. We propose the model and its theoretical analysis in Sect. 2. Experimental results are shown in Sect. 3, and the conclusion is given in Sect. 4.

2 Joint Sparse Locality Preserving Projections

In this section, we first give the motivation of this paper and then propose the model. At last, an iterative algorithm is designed to solve the optimization problem.

2.1 The Motivations

As mentioned in the introduction section, the $L_{2,1}$-norm based on jointly sparse feature selection can greatly improve the recognition performance. Moreover, a sparse projection can also give clearer explanation for the selected features [14]. On the other hands, the manifold learning methods can preserve the local structure of the datasets which are more useful than the global structure for feature extraction in some classification tasks [9]. Therefore, it is desirable to combine the advantages of sparse feature extraction and manifold learning for improving the recognition performance. Thus, we propose a novel manifold learning model called Joint sparse locality preserving projections (JSLPP) for feature extraction and selection by imposing $L_{2,1}$-norm regularization term on the projection matrix to guarantee the joint sparsity.

2.2 Objective Function and Its Solution

In order to integrate manifold learning and sparse regression together to improve the recognition performance, we present the objective function of JSLPP as follows:

$$\min_{\mathrm{A,B}} \sum_{i=1}^{\mathrm{n}} \sum_{j=1}^{\mathrm{n}} \left\|\mathbf{x}_i - \mathbf{A}\mathbf{B}^T\mathbf{x}_j\right\|_2^2 \mathbf{W}_{ij} + \lambda\|\mathbf{B}\|_{2,1} \; s.t. \; \mathbf{A}^T\mathbf{A} = \mathbf{I} \tag{1}$$

where $\mathbf{x}$ is a d-dimensional column vector, n is the training number of the samples, $\mathbf{A} \in R^{d\times k}$ is a basic matrix and $\mathbf{B} \in R^{d\times k}(k << d)$ is a projection matrix, $\mathbf{W} \in R^{d\times k}$ is a weight graph and λ is a regularization parameter. From (1), we have

$$\begin{aligned}
&\textstyle\sum_{ij} \left\|\mathbf{x}_i - \mathbf{A}\mathbf{B}^{\mathrm{T}}\mathrm{x}_j\right\|_2^2 W_{ij} + \lambda\|\mathbf{B}\|_{2,1} \\
&= \textstyle\sum_{ij} \mathrm{tr}(\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_i - 2\mathbf{x}_i^{\mathrm{T}}\mathbf{A}\mathbf{B}^{\mathrm{T}}\mathbf{x}_j + (\mathbf{A}\mathbf{B}^{\mathrm{T}}\mathbf{x}_j)^{\mathrm{T}}\mathbf{A}\mathbf{B}^{\mathrm{T}}\mathbf{x}_j)W_{ij} + \lambda\,\mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{B}) \\
&= tr(\textstyle\sum_{ij} \mathbf{x}_i^{\mathrm{T}}W_{ij}\mathbf{x}_i - 2\sum_{ij} \mathbf{A}\mathbf{x}_i^{\mathrm{T}}W_{ij}\mathbf{x}_j\mathbf{B}^{\mathrm{T}} + \sum_{ij} (\mathbf{A}\mathbf{B}^{\mathrm{T}}\mathbf{x}_j)^{\mathrm{T}}\mathbf{A}\mathbf{B}^{\mathrm{T}}\mathbf{x}_j W_{ij}) + \lambda\,\mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{B}) \\
&= \mathrm{tr}(\mathbf{X}\mathbf{D}\mathbf{X}^{\mathrm{T}} - 2\mathbf{A}\mathbf{X}\mathbf{W}\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}} + \mathbf{B}\mathbf{X}\mathbf{D}\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}) + \lambda\,\mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{B})
\end{aligned}$$

where $\mathbf{D}$ is a diagonal matrix, that is $\mathbf{D}_{ii} = \sum_j W_{ji}$. $\mathbf{\Lambda}$ is a diagonal matrix with the i-th diagonal element defined as $\mathbf{\Lambda}_{ii} = \frac{1}{2\left\|\mathbf{B}^i\right\|_2}$, where $\mathbf{B}^i$ is the i-th row of $\mathbf{B}$. Finally, we obtain the optimization problem as follows:

$$\min_{\mathbf{A},\mathbf{B}} \; \mathrm{tr}(\mathbf{X}\mathbf{D}\mathbf{X}^{\mathrm{T}} - 2\mathbf{A}\mathbf{X}\mathbf{W}\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}} + \mathbf{B}\mathbf{X}\mathbf{D}\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}) + \lambda\,\mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{B}) \quad s.t\; \mathbf{A}^T\mathbf{A} = \mathbf{I} \tag{2}$$

To obtain the optimal solutions of the two variables in (2), we design an alternately iterative algorithm. Suppose $\mathbf{A}$ is fixed, we have:

$$l(\mathbf{A},\mathbf{B}) = \mathrm{tr}(\mathbf{X}\mathbf{D}\mathbf{X}^{\mathrm{T}} - 2\mathbf{A}\mathbf{X}\mathbf{W}\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}} + \mathbf{B}\mathbf{X}\mathbf{D}\mathbf{X}^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}) + \lambda\,\mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{\Lambda}\mathbf{B}) \tag{3}$$

By taking the derivation of $l(\mathbf{A}, \mathbf{B})$ w.r.t $\mathbf{B}$ to be equal to zero, we have:

$$\mathbf{B} = (\mathbf{X}\mathbf{D}\mathbf{X}^{\mathrm{T}} + \lambda\mathbf{\Lambda})^{-1}\mathbf{X}\mathbf{W}\mathbf{X}^{T}\mathbf{A} \tag{4}$$

For given B, discarding the constant in (2), we can obtain the following optimization problem

$$\min_{\mathbf{A}} tr(-2\mathbf{B}^{T}\mathbf{X}\mathbf{W}\mathbf{X}^{T}\mathbf{A}) \ \ s.t. \ \ \mathbf{A}^{T}\mathbf{A} = \mathbf{I}. \tag{5}$$

Then, (5) is equal to the following maximization problem

$$\max_{\mathbf{A}} tr(\mathbf{A}^{T}\mathbf{X}\mathbf{W}\mathbf{X}^{T}\mathbf{B}) \quad s.t. \ \ \mathbf{A}^{T}\mathbf{A} = \mathbf{I}. \tag{6}$$

Let SVD of $\mathbf{X}\mathbf{W}\mathbf{X}^{T}\mathbf{B} = \ \mathbf{U}\mathbf{D}\mathbf{V}^{T}$ and from Theorem 4 in [17], we have

$$\mathbf{A} = \mathbf{U}\mathbf{V}^{T}. \tag{7}$$

By alternatively updating $\mathbf{A}$ and $\mathbf{B}$ with (4) and (7) respectively, we eventually obtain the optimal projection matrix $\mathbf{B}$ and the basic matrix $\mathbf{A}$.

2.3 The Convergence

In order to prove the convergence of the proposed algorithm, we need the following Lemmas.

Lemma 1. [15] For any two nonzero-constants a and b, we have the following inequality:

$$\sqrt{\mathrm{a}} - \frac{a}{2\sqrt{b}} \leq \sqrt{b} - \frac{b}{2\sqrt{b}} \tag{8}$$

Lemma 2. [15] For any nonzero vectors $\mathbf{p}$, $\mathbf{p}_t \in R^c$, the following inequality holds:

$$\|\mathbf{p}\|_2 - \frac{\|\mathbf{p}\|_2^2}{2\|\mathbf{p}_t\|_2} \leq \|\mathbf{p}_t\|_2 - \frac{\|\mathbf{p}_t\|_2^2}{2\|\mathbf{p}_t\|_2} \tag{9}$$

With Lemmas 1 and 2, we have the following theorem.

Theorem 1. The iteration approach presented in Sect. 2.2 will monotonically decrease the objective function value in each iteration and converge to the local optimum.

Proof: For ease of representation, we denote the objective function (1) as $J(\mathbf{B}, \mathbf{A}, \mathbf{W}, \mathbf{D}, \mathbf{\Lambda}) = J(\mathbf{B}, \mathbf{A}, \mathbf{\Lambda})$. Suppose in the $(t-1)$-th iteration, we have $\mathbf{B}_{t-1}$, $\mathbf{A}_{t-1}$ and $\mathbf{\Lambda}_{t-1}$. From (4), we can find that

$$J(\mathbf{B}_{(t)}, \mathbf{A}_{(t-1)}, \mathbf{\Lambda}_{(t-1)}) \leq J(\mathbf{B}_{(t-1)}, \mathbf{A}_{(t-1)}, \mathbf{\Lambda}_{(t-1)}) \tag{10}$$

For $\mathbf{A}_t$, as its optimal value comes from the SVD decomposition value of $\mathbf{XWX}^T\mathbf{B}$ which further decreases the objective function, we have

$$J(\mathbf{B}_{(t)}, \mathbf{A}_{(t)}, \mathbf{\Lambda}_{(t-1)}) \leq J(\mathbf{B}_{(t-1)}, \mathbf{A}_{(t-1)}, \mathbf{\Lambda}_{(t-1)}) \tag{11}$$

Once the optimal $\mathbf{B}_{(t)}$ and $\mathbf{A}_{(t)}$ are obtained, we have

$$\begin{aligned}&\mathrm{tr}(-2\mathbf{A}_{(t)}\mathbf{XWX}^{\mathrm{T}}\mathbf{B}_{(t)}^{\mathrm{T}} + \mathbf{B}_{(t)}\mathbf{XDX}^{\mathrm{T}}\mathbf{B}_{(t)}^{\mathrm{T}}) + \lambda\,\mathrm{tr}(\mathbf{B}_{(t)}^{\mathrm{T}}\mathbf{\Lambda}_{(t-1)}\mathbf{B}_{(t)})\\ &\leq tr(-2\mathbf{A}_{(t-1)}\mathbf{XWX}^{\mathrm{T}}\mathbf{B}_{(t-1)}^{\mathrm{T}} + \mathbf{B}_{(t-1)}\mathbf{XDX}^{\mathrm{T}}\mathbf{B}_{(t-1)}^{\mathrm{T}}) + \lambda\,\mathrm{tr}(\mathbf{B}_{(t-1)}^{\mathrm{T}}\mathbf{\Lambda}_{(t-1)}\mathbf{B}_{(t-1)})\end{aligned}$$

That is

$$\begin{aligned}&\mathrm{tr}(-2\mathbf{A}_{(t)}\mathbf{XWX}^{\mathrm{T}}\mathbf{B}_{(t)}^{\mathrm{T}} + \mathbf{B}_{(t)}\mathbf{XDX}^{\mathrm{T}}\mathbf{B}_{(t)}^{\mathrm{T}}) + \lambda\sum_i \frac{\left\|\mathbf{B}_{(t)}^i\right\|_2^2}{\left\|\mathbf{B}_{(t-1)}^i\right\|_2}\\ &\leq tr(-2\mathbf{A}_{(t-1)}\mathbf{XWX}^{\mathrm{T}}\mathbf{B}_{(t-1)}^{\mathrm{T}} + \mathbf{B}_{(t)}\mathbf{XDX}^{\mathrm{T}}\mathbf{B}_{(t)}^{\mathrm{T}}) + \lambda\sum_i \frac{\left\|\mathbf{B}_{(t-1)}^i\right\|_2^2}{\left\|\mathbf{B}_{(t-1)}^i\right\|_2}\end{aligned} \tag{12}$$

Then, we have

$$\begin{aligned}&\mathrm{tr}(-2\mathbf{A}_{(t)}\mathbf{XWX}^{\mathrm{T}}\mathbf{B}_{(t)}^{\mathrm{T}} + \mathbf{B}_{(t)}\mathbf{XDX}^{\mathrm{T}}\mathbf{B}_{(t)}^{\mathrm{T}}) + \lambda\sum_i \left\|B_{(t)}^i\right\|_2 - (\lambda\sum_i \left\|B_{(t)}^i\right\|_2 - \lambda\sum_i \frac{\left\|\mathbf{B}_{(t)}^i\right\|_2^2}{\left\|\mathbf{B}_{(t-1)}^i\right\|_2})\\ &\leq tr(-2\mathbf{A}_{(t-1)}\mathbf{XWX}^{\mathrm{T}}\mathbf{B}_{(t-1)}^{\mathrm{T}} + \mathbf{B}_t\mathbf{XDX}^{\mathrm{T}}\mathbf{B}_{(t)}^{\mathrm{T}}) + \lambda\sum_i \left\|B_{(t-1)}^i\right\|_2 - (\lambda\sum_i \left\|B_{(t-1)}^i\right\|_2 - \lambda\sum_i \frac{\left\|\mathbf{B}_{(t-1)}^i\right\|_2^2}{\left\|\mathbf{B}_{(t-1)}^i\right\|_2})\end{aligned} \tag{13}$$

Then combining (12) and (13) and Lemma 2, we further obtain

$$\begin{aligned}&tr(-2\mathbf{A}_{(t)}\mathbf{XWX}^{\mathrm{T}}\mathbf{B}_{(t)}^{\mathrm{T}} + \mathbf{B}_{(t)}\mathbf{XDX}^{\mathrm{T}}\mathbf{B}_{(t)}^{\mathrm{T}}) + \lambda\left\|\mathbf{B}_{(t)}^i\right\|_2\\ &\leq tr(-2\mathbf{A}_{(t)}\mathbf{XWX}^{\mathrm{T}}\mathbf{B}_{(t-1)}^{\mathrm{T}} + \mathbf{B}_{(t-1)}\mathbf{XDX}^{\mathrm{T}}\mathbf{B}_{(t-1)}^{\mathrm{T}}) + \lambda\left\|\mathbf{B}_{(t-1)}^i\right\|_2\end{aligned}$$

That is

$$J(\mathbf{B}_{(t)}, \mathbf{A}_{(t)}, \mathbf{\Lambda}_{(t)}) \leq J(\mathbf{B}_{(t-1)}, \mathbf{A}_{(t-1)}, \mathbf{\Lambda}_{(t-1)}) \tag{14}$$

Therefore, the algorithm will converge to the local optimum.

2.4 Computational Complexity Analysis

The algorithm first obtain the weight matrix $\mathbf{W}$, then get the optimal projection matrix $\mathbf{B}$ and the basic matrix $\mathbf{A}$ as well as the diagonal matrix $\mathbf{\Lambda}$. The main computational cost of the iterative algorithm is to compute the projection matrix $\mathbf{B}$, the basic matrix $\mathbf{A}$ and the diagonal matrix $\mathbf{\Lambda}$. Computing the projection matrix $\mathbf{B}$ needs $O(d^3)$, the basic matrix $\mathbf{A}$ needs $O(d^3)$ and the diagonal matrix $\mathbf{\Lambda}$ needs $O(d^2)$. If the algorithm needs T iteration steps, then the total computational complexity is $O(n^2 + Tnd^3 + Tnd^3 + Td^2)$.

2.5 JSLPP Algorithm

The code of JSLPP algorithm is as follows:

BEGIN

Input: the training data $\mathbf{X}$, the weight matrix $\mathbf{W}$, the diagonal matrix $\mathbf{D}$, the dimensionality d of the sample, the desired dimensionality k of matrix $\mathbf{A}$ and $\mathbf{B}$, and the regularization parameter λ.

Program:

1: $\mathbf{W}$ = constructW($\mathbf{X}$, options)

2: $\mathbf{A}$ = rand(d, k)

3: $\mathbf{B}$ = rand(d, k)

4: $\mathbf{D}_{ii} = \sum_j W_{ji}$

5: for iter←1 to maxIter

6: $\mathbf{B} = (\mathbf{X}\mathbf{D}\mathbf{X}^{\mathrm{T}} + \lambda\mathbf{\Lambda})^{-1}\mathbf{X}\mathbf{W}\mathbf{X}^{T}\mathbf{A}$

7: ($\mathbf{U}$, $\mathbf{V}$) = SVD($\mathbf{X}\mathbf{W}\mathbf{X}^{T}\mathbf{B}$)

8: $\mathbf{A} = \mathbf{U}\mathbf{V}^{T}$

9: Until Converge

10: End

11: $\mathbf{Y}$ = $\mathbf{B}$ * $\mathbf{X}$

Output: Low-dimensional features $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_n]$

END

3 Experiments

In this section, a set of experiments are presented to evaluate the proposed JSLPP algorithm for feature extraction and selection. We compared it with PCA, LPP, L_1-norm regularized sparse subspace learning methods SPCA, the most related $L_{2,1}$-norm based Feature Selection and Subspace learning (FSSL), RFS [15] and $L_{2,1}$-norm regularized discriminative feature selection for unsupervised learning UDFS, SAIR [16]. Six methods mentioned above were compared with the JSLPP in the same experimental condition. The datesets are all divided into training sets and test sets. The number of training samples are set as 4, 6, and the rest data are used as testing sets,

respectively. In all experiments, we first performed feature extraction and selection, then used the nearest neighborhood classifier to perform classification.

3.1 Experiments on the AR Face Database

There are over 4000 color face images of 126 people in AR face database, we selected 120 images of 120 people (65 men and 55 women) from this dataset. All images are the frontal views of faces with different facial expressions, lighting conditions, and occlusions, and they are normalized to 50 × 40 pixels.

In the experiment, the number of class is 120 and each class has 20 samples. $l(l = 4, 6)$ images of each class were randomly selected and used for training and the remaining images were used for test. The optimal value of parameter γ was selected form the set $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$, Table 1 lists the average performance of different methods on the AR face database based on 10 times running, and the average recognition rates versus the dimensions of the projection are shown in the Fig. 1.

Table 1. The performance (recognition rate, standard deviation and dimension) of different methods on the AR face database

Training samples	PCA	LPP	SPCA	FSSL	UDFS	SAIR	JSLPP
4	76.20	79.81	76.20	89.84	85.44	83.55	77.76
	±4.58	±8.86	±4.58	±9.45	±8.44	±8.20	±8.20
	110	95	110	110	105	100	75
6	79.96	85.57	79.96	95.39	89.73	91.05	87.87
	4.86	±7.31	±4.85	±7.38	±5.75	±6.17	±10.38
	110	105	110	115	100	100	60

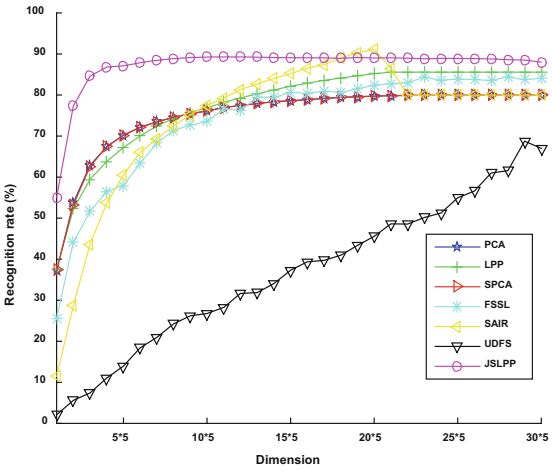

Fig. 1. The recognition rates (%) versus the dimensions of different methods on the AR face database.

3.2 Experiments on the ORL Face Database

The ORL face database have 40 people, and each person have 10 images. The images were taken at different times, varying lighting, facial expression (open or closed eyes, smiling or not smiling) and facial details (glasses or no glasses). In Table 2, we lists the average performance of different methods on the ORL face database, and the average recognition rates versus the dimensions of the projection are shown in Fig. 2.

Table 2. The performance (recognition rate, standard deviation and dimension) of different methods on the ORL face database

Training samples	PCA	LPP	SPCA	FSSL	L21R21	UDFS	SAIR	JSLPP
4	93.58	78.21	93.54	93.42	89.67	93.54	95.08	94.54
	±1.95	±2.96	±1.93	±1.70	±2.09	±2.10	±2.44	±1.72
	135	85	130	35	40	150	40	120
6	95.94	88.75	96.00	96.25	92.63	95.81	97.06	97.94
	±1.59	±2.76	±1.66	±1.85	±2.93	±1.59	±1.35	±1.41
	105	65	100	35	40	150	40	40

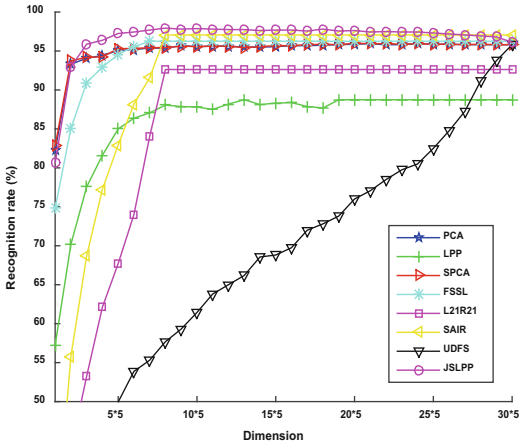

Fig. 2. The recognition rates (%) versus the dimensions of different methods on the ORL face databases.

4 Conclusion

In this paper, a novel method called Joint Sparse Locality Preserving Projection (JSLPP) is proposed for sparse subspace learning by considering manifold learning and feature selection techniques. The $L_{2,1}$-norm is introduced in the JSLPP model, an iterative algorithm is designed to solve the optimization problem. We prove the convergence of the proposed algorithm, and the computational complexity is also presented. Experiments on two well-known face datasets show that JSLPP performs better than the traditional feature extraction and linear manifold learning methods.

Acknowledgement. This work was supported in part by the Natural Science Foundation of China (Grant 61573248, Grant 61773328, Grant 61375012 and Grant 61703283), China Postdoctoral Science Foundation (Project 2016M590812 and Project 2017T100645), the Guangdong Natural Science Foundation (Project 2017A030313367 and Project 2017A030310067), and Shenzhen Municipal Science and Technology Innovation Council (No. JCYJ2017030215 3434048).

References

1. Batur, A.U., Hayes, M.H.: Linear subspace for illumination robust face recognition. In: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, December 2001
2. Cox, T.F., Cox, M.A.A.: Multidimensional scaling on a sphere. Commun. Stat. Theory Methods **20**(9), 2943–2953 (1991)
3. Turk, M., Pentland, A.P.: Face recognition using eigenfaces. In: IEEE Conference on Computer Vision and Pattern Recognition (1991)
4. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)
5. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**, 2319–2323 (2000)
6. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**, 2323–2326 (2000)
7. Saul, L.K., Roweis, S.T.: Think globally.: fit locally: unsupervised learning of low dimensional manifolds. J. Mach. Learn. Res. **4**, 119–155 (2003)
8. Belkin, M., Niyogi, P.: Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Proceedings of Conference on Advances in Neural Information Processing System, vol. 15 (2001)
9. He, X., Yan, S., Hu, Y., Niyogi, P., Zhang, H.J.: Face recognition using laplacianfaces. IEEE Trans. Pattern Anal. Mach. Intell. **27**(3), 328–340 (2005)
10. He, X., Niyogi, P.: Locality preserving projections. In: Neural Information Processing Systems, vol. 16, p. 153 (2004)
11. He, X., Cai, D., Yan, S., Zhang, H.J.: Neighborhood preserving embedding. In: ICCV, pp. 1208–1213 (2005)
12. Bradley, P., Mangasarian, O.: Feature selection via concave minimization and support vector machines
13. Wang, L., Zhu, J., Zou, H.: Hybrid huberized support vector machines for microarray classification. In: ICML (2007)
14. Gu, Q., Li, Z., Han, J.: Joint feature selection and subspace learning. In: International Joint Conference on Artificial Intelligence, pp. 1294–1299. AAAI Press (2011)
15. Nie, F., Huang, H., Cai, X., Ding, C.: Efficient and robust feature selection via joint $L_{2,1}$ norms minimization. In: Advances in Neural Information Processing Systems, vol. 23, pp. 1813–1821 (2010)
16. Ma, Z., Yang, Y., Sebe, N., Member, S., Zheng, K., Hauptmann, A.G.: Classifier-specific intermediate representation, **15**(7), 1628–1637 (2013)
17. Zou, H., Hastie, T., Tibshirani, R.: Sparse principal component analysis. J. Comput. Graph. Stat. **15**(2), 265–286 (2006)