International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.7, No.1 (2014), pp.13-26
http://dx.doi.org/10.14257/ijsip.2014.7.1.02

Dimensionality Reduction Based on Supervised Slow Feature Analysis for Face Recognition

Xingjian Gu[1], Chuancai Liu[1] and Zhangjing Yang[1]

[1]*School of computer science and Engineering, Nanjing University of Science and Technology ,Nanjing, Jiangsu 210094, People's Republic of China*
guxingjian163@163.com

Abstract

Slow feature analysis (SFA) is motivated by biological model to extracts slowly varying feature from a quickly varying input signal. However, traditional slow feature analysis is an unsupervised method to extract slow or invariant feature and cannot be directly applied on the data set without an obvious temporal structure, i.e. face databases. In this paper, we propose a supervised slow feature analysis to do dimensionality reduction for face recognition. First, a new criterion is developed to construct a Pseudo-time series for data sets without an obvious temporal structure. Then, the first-order derivative at each point in the Pseudo-time series is computed in form of vectors. At last we construct the objective function of SSFA that ensures the secondary moment of first-order derivative as small as possible in the embedding space. SSFA is able to extract the invariant feature for each class and preserve the local structure in embedding space simultaneously. Experimental results on the Yale, ORL, AR, and FERET face databases show the effectiveness of the proposed algorithm.

Keywords: *slow feature analysis, dimensionality reduction, face recognition*

1. Introduction

Dimensional reduction is one of the most popular techniques in pattern recognition and computer vision. The goal of dimensional reduction is to find the compactness representations of high-dimensional data and simultaneously discover the intrinsic structure of the original data. Over the past decades, a number of dimensional reduction methods have been developed [1-11]which can be categorized into two classes: Linear subspace methods such as principle component analysis (PCA)[1]and linear discriminant analysis (LDA)[2], and nonlinear approaches such as kernel-based techniques, geometry-based techniques and subclass analysis.

Linear subspace method tries to find a linear subspace as feature space in order to preserve certain kind of characteristics of original data. Specifically, PCA aims to find projects by maximizing the total variance of features. While LDA intends to search the discriminant vectors by maximizing the between-class scatter and minimizing the within-class scatter simultaneously. Generally speaking, linear subspace method could obtain good performance when the data has a linear structure, but it may get suboptimal when the data structure is nonlinear and very complicated.

Recently, a number of nonlinear dimensionality reduction approaches have been proposed to deal with nonlinear structural data. Those nonlinear methods are mainly based on two kinds of attractive ideas: kernel-based techniques and geometry-based techniques. Kernel method is one common approach which implicitly expands the input data into high dimensional space in order to convert data from nonlinear structure to linear structure. For example, the kernel version of PCA and LDA are KPCA [3] and KLDA [4, 5] respectively. Compared with kernel-based methods, geometry-based methods are motivated by adopting geometrical perspective to explore the immanent

ISSN: 2005-4254 IJSIP

structure of data. There are many popular manifold learning algorithms including locally linear embedding (LLE)[6], ISOMAP[7], Laplacian eigenmaps (LE)[8], and local tangent space alignment (LTSA) [9]. These kinds of nonlinear methods aim to preserve local structures in small neighborhoods and successfully derive the intrinsic feature of nonlinear manifolds with the help of differential geometry. However, manifold learning methods obtain low-dimensional embeddings without an explicit mapping on new test data points for recognition problems. Many endeavors have been done to use the out-of-sample techniques [10] to deal with test data in practical application. To deal with the out-of-sample problem in LE, He *et al.*, [11] proposed locality preserving projection (LPP), a linearization of LE, to learn a linear mapping for dimensionality reduction. Neighborhood preserving embedding (NPE) [12] tries to find a linear subspace that preserves local structure under the same principle of LLE.

Over the past decades, there have been a number of research focuses on applying the biologically model to complex information tasks. Temporal coherence principle is one of the most attractive learning rules inspired by biological model. The first description of temporal coherence principle was given by Hinton [13] as a learning rule in back-propagation neural networks, and it was implemented by Mitchison [14] and Földiák [15]. Slow feature analysis (SFA), proposed by Wiskott and Sejnowski [16], is an attractive biologically inspired rule to extract invariant from vectorial temporal signals. It extracts slowly varying features from fast varying time series according to time scale. Recently, some researchers introduced the slowness principle to the applications of pattern recognition and machine learning. In[17] Huang has proposed a nonlinear dimensionality reduction framework using a temporal coherence principle which takes into considering both time series and data sets without an obvious temporal structure. Zhang[18] introduces the SFA framework to the problem of human action recognition which has an obvious temporal structure.

Inspired by slow feature analysis, we propose a novel supervised manifold learning based on supervised slow feature analysis (SSFA) which is able to extract the invariance feature of each class that can be favorable for classification. In order to deal with datasets without clear temporal structure, we need to construct a Pseudo-time series for each class. A new criterion is proposed in this paper to resort samples, which is equal to the famous traveling salesman problem. To solve the NP-hard problem, we apply the well-known Nearest Neighbor Procedure to find the shortest path containing all the points. Then we extract the invariant feature for each class and preserve the local structure in embedding space simultaneously.

The organization of the rest of this paper is as follows. In Section 2, we review briefly tradition slow feature analysis. In Section 3, we propose the supervised slow feature analysis (SSFA) for dimensional reduction and describe the proposed method in detail. We discuss the connections between SSFA and other DR methods in Section 4. In section 5, we perform experiment on four famous face databases. Conclusions are made in Section 6.

2. Traditional Slow Feature Analysis

Slow feature analysis (SFA) was firstly proposed by Wiskott and Sejnowski[16]. The basic idea of SFA is to extract slow feature from fast varying time series by learning proper functions. For a multidimensional time series $x(t) = (x_1(t), x_2(t), \cdots, x_n(t))^T$, $t \in [t_0, t_1]$, SFA finds a set of functions $y_k = g_k(x(t))$, $k = 1, 2, 3, \cdots, m$ following that

$$\begin{aligned} & \min \Delta(y_k) = \langle \dot{y}_k \rangle_t \\ \text{s.t.} \quad & \langle y_k \rangle_t = 0 \\ & \langle y_k^2 \rangle_t = 1 \\ & \forall i < j, \langle y_i y_j \rangle_t = 0 \end{aligned} \tag{1}$$

where $\langle \cdot \rangle_t$ and $\dot{y}_{k,t}$ indicates the time average and the time derivative of y_k receptively, the $\dot{y}_{k,t}$ can be approximated by a finite difference $\dot{y}_{k,t} = y_{k,t} - y_{k,t+1}$. In linear case $y_k = \left(w_k^T x \right)_t$, the equation can be written as:

$$\Delta(y_k) = \langle \dot{y}_k^2 \rangle_t = w_k^T \langle \dot{x}\dot{x}^T \rangle w_k = w_k^T \mathbf{A} w_k \tag{2}$$

$$\Delta(y_k y_i) = w_k^T \langle x x^T \rangle w_k = w_k^T \mathbf{B} w_k \tag{3}$$

The objection function can be reduced to the following:

$$\Delta(y_k) = \frac{w_k^T \mathbf{A} w_k}{w_k^T \mathbf{B} w_k} \tag{4}$$

It is a version of Rayleigh quotient and the problem to be solved is the generalized eigenproblem:

$$\mathbf{A W} = \mathbf{B W \Lambda} \tag{5}$$

where w is the matrix of the generalized eigenvectors and Λ is the diagonal matrix of the generalized eigenvalues respectively. The slowest feature corresponds to the smallest eigenvalues, the second slow feature corresponds to the second small eigenvalues, and so on.

Traditional slow feature analysis is an unsupervised method to extract slow varying feature from a quick varying signal with an obvious temporal structure. However, many real-world applications do not always have a clear temporal structure, *i.e.*, face recognition. We except that extracted the feature could be more favorable for classification by using the label informing. In the next section, we will propose a new supervised dimensional reduction based on supervised slow feature analysis.

3. Supervised Slow Feature Analysis (SSFA)

In this section, we introduce the SSFA algorithm that can be directly applied on datasets without obvious temporal structure. First, SSFA considers constructing a Pseudo-time series for those samples belonging to the same class for each class. Then, we construct the con-variance matrix of first derivative. Lastly, SSFA seeks the optimal projections that minimize the trace of con-variance matrix of first derivative in order to minimize the "distance" of the embedding coordinates of two frames if they are sequential frames in the original high dimensional space.

3.1. Pseudo-time Series Construction

Given a high D-dimensional dataset $\{ \tilde{x}_i^j \in R^D \mid i = 1,2,\cdots,c \ \ j = 1,2,\cdots,n_i \}$ that contains c classes and ith class contains n_i samples, to obtain the temporal structure, we should resort the samples for each class, and use a sequence to represent samples of one class. In temporal series case, the nearby frame is very "close" and correlative. We obey the following objective function to construct a Pseudo-time sequence $\{ x_i^1, x_i^2, \cdots, x_i^{n_i} \}$ to represent the samples belonging to ith class:

$$\min_{\{x_i^1, x_i^2, \cdots, x_i^{n_i}\}} \sum_{j=1}^{n_i - 1} \left\| x_i^j - x_i^{j+1} \right\|^2 \tag{6}$$

This problem is equivalence to the famous salesperson problem, which is an NP hard problem. In next subsection, we first use nearest neighbor procedure to approximate the path with a random point as the beginning "city". Then, the shortest one is chosen as the required pseudo-time sequence.

3.2. The Derivation of Algorithm

Assume a D-dimensional dataset $\mathbf{X} = \{x_i^j \in R^D \mid i = 1, \cdots, c, j = 1, 2, \cdots, n_i\}$ has been preprocessed by the algorithm detailed in pre-subsection, so as the dataset has temporal structure. The objection function of SSFA is to find low d-dimensional representation $\{\mathbf{y}_i^j \in R^d \mid i = 1, 2, \cdots, n, j = 1, 2, \cdots, n_i\}$, where $d << D$, so as the variable y_i^j and y_i^{j+1} change as little as possible if x_i^j and x_i^{j+1} are adjacent sequential frames. Thus we can obtain a "good" mapping by minimizing the following objective function:

$$\Delta y = \sum_{i=1}^{c} \sum_{j=1}^{n_i - 1} \left\| y_i^j - y_i^{j+1} \right\|^2 \tag{7}$$

Consider the linear case $y = \mathbf{W}^T x$, where $\mathbf{W} \in R^{D \times d}$, and give the matrix $\Delta \mathbf{X}_i = \left[\Delta x_i^1, \Delta x_i^2, \cdots, \Delta x_i^{n_i}\right]$, where $\Delta x_i^j = x_i^j - x_i^{j+1}$, the objective function can be reformulated as

$$\begin{aligned}
\Delta y &= \sum_{i=1}^{c} \sum_{j=1}^{n_i - 1} \left\| y_i^j - y_i^{j+1} \right\|^2 = \sum_{i=1}^{c} \sum_{j=1}^{n_i - 1} \left\| \mathbf{W}^T x_i^j - \mathbf{W}^T x_i^{j+1} \right\|^2 \\
&= tr\left(\sum_{i=1}^{c} \sum_{j=1}^{n_i - 1} \left(\mathbf{W}^T x_i^j - \mathbf{W}^T x_i^{j+1} \right) \left(\mathbf{W}^T x_i^j - \mathbf{W}^T x_i^{j+1} \right)^T \right) \\
&= tr\left(\mathbf{W}^T \left(\sum_{i=1}^{c} \sum_{j=1}^{n_i - 1} \left(x_i^j - x_i^{j+1} \right) \left(x_i^j - x_i^{j+1} \right)^T \right) \mathbf{W} \right) \\
&= tr\left(\mathbf{W}^T \left(\sum_{i=1}^{c} \Delta \mathbf{X}_i \Delta \mathbf{X}_i^T \right) \mathbf{W} \right)
\end{aligned} \tag{8}$$

We can solve this problem under the constraints of zero mean and unit variance, which can keep significant information as much as possible

$$\begin{aligned}
E(\mathbf{Y}) &= \mathbf{0} \\
E\left(\mathbf{Y}\mathbf{Y}^T\right) &= \mathbf{1}
\end{aligned} \tag{9}$$

The covariance matrix of $\mathbf{Y}$ can be calculated by

$$\mathbf{Y}\mathbf{Y}^T = \mathbf{W}^T \mathbf{X}\mathbf{X}^T \mathbf{W} \tag{10}$$

Finally, the minimization problem is reduced to find

$$\arg\min_{\mathbf{W}} \frac{\mathbf{W}^T \mathbf{A} \mathbf{W}}{\mathbf{W}^T \mathbf{B} \mathbf{W}} \tag{11}$$

Where $\mathbf{A} = \sum_{i=1}^{c} \Delta \mathbf{X}_i \Delta \mathbf{X}_i^T$, $\mathbf{B} = \mathbf{X}\mathbf{X}^T$

It is a version of Rayleigh quotient and the problem to be solved is the generalized eigenproblem

$$\mathbf{A}\mathbf{W} = \mathbf{\Lambda}\mathbf{B}\mathbf{W} \tag{12}$$

3.3. Algorithm Summarization

The SSFA algorithm is summarized as follows

Input: $\tilde{x} = \{\tilde{x}_i^j \in R^D \mid i = 1, \cdots, c,\ j = 1, 2, \cdots, n_i\}$

Step1: construct Pseudo-time series for each class

Step2: calculate $\Delta \mathbf{X}_i$ for each class

Step3: calculate $\mathbf{A}$ and $\mathbf{B}$

Step4: estimate $\mathbf{W}$ whose columns are the eigenvectors of $\mathbf{A}\mathbf{W} = \mathbf{\Lambda}\mathbf{B}\mathbf{W}$ and order the columns according to the eigenvalues

Step5: choose d small non-zeros eigenvalues and associated eigenvectors $\mathbf{U} = (w_1, w_2, \cdots, w_d)$, and obtain the d-dimensional coordinates by $\mathbf{Y} = \mathbf{U}^T \mathbf{X}$

---

4. Connections with other DR Methods

In this section, we discuss the relation between SSFA and LDA [2], LPP [11], NPE [12]. Meanwhile, we specify the advantages of SSFA and its differences with the existent popular DR methods.

4.1. Connections with LPP

LPP also seeks to preserve the intrinsic geometry of the data and local structure. The objective function is as follows:

$$\min \sum_{ij} \left(w^T x_i - w^T x_j \right)^2 S_{ij} \tag{14}$$

Where y_i is the one dimensional representation of x_i and the matrix S is a similarity. A possible way of defining S is as follows:

$$S_{ij} = \begin{cases} \exp\left(\left\| x_i - x_j \right\|^2 / t \right), & if\ x_i \text{ is among the k-nearest of } x_j \\ 0 & otherwise \end{cases} \tag{15}$$

After some simple algebraic steps, the minimize problem can reduce to the following:

$$\arg\min_{w, w^T XDX^T w = 1} w^T XLX^T w \tag{16}$$

Where $L = D - S$ and D is diagonal matrix; its entries are column sums of S , $D_{ii} = \sum_j S_{ij}$.

Although LPP has made a salience success, there is still an open problem, i.e. how to choose an optimal parameter k . In supervised scenarios, SLPP can easily project the samples belonging to same class into one point along the projection axes. In comparison with SLPP, the SSFA has the following advantage: a) SSFA compute first derivative by the nearby points that avoiding determining the parameter k; b) In SSFA, each point is just be used once so that it could avoid over-fitting phenomenon to some extent; c) SSFA can not only preserve the local geometric structure of original, but also the output series belonging to the same class varies slowly or keep invariant result in more classification ability.

4.2. Connections with NPE

To our knowledge, NPE is a linearization version of LLE which assumes that each data point and its neighbor lie on or close to a locally linear patch of some underlying manifold. NPE evaluates an affinity weight matrix by using a local least squares approximation, *i.e.*,

$$\begin{aligned} &\mathbf{W} = \arg\min_{\mathbf{W}} \sum_i \left\| x_i - \sum_j w_{ij} x_j \right\|^2 \\ &s.t. \ \sum_j w_{ij} = 1 \end{aligned} \tag{17}$$

A reasonable criterion for choosing a "good" projection is to minimizing the following cost function:

$$\Phi(p) = \sum_i \left\| p^T x_i - \sum_j w_{ij} p^T x_j \right\|^2 \tag{18}$$

By some simple algebraic derivation, the minimizing problem can be reduced to the following model:

$$\min_{\mathbf{p}} \frac{\mathbf{p}^T \mathbf{X} \mathbf{M} \mathbf{X}^T \mathbf{p}}{\mathbf{p}^T \mathbf{X} \mathbf{X}^T \mathbf{p}} \tag{19}$$

where $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W})$ is a symmetric and semi-positive definite matrix, $\mathbf{I}$ is an identity matrix. However NPE can obtain a high performance, similar to LPP, it still has the same open problem how to choose an optimal parameter. And more, it just considers the reconstruction error ignoring preserving the geometric structure of original data. When the density of data distribution is low in the original space, NPE may lose its advantage. In comparison to NPE, SSFA has the following advantage: a) it has no parameter to be determined and obtains a global solution as it converts to an eigen problem; b) it not only preserves the local structure but also extracts slow feature for each class that has high classification ability.

5. Experiments and Discussion

In this section, we evaluate the effectiveness of our proposed method for dimensional reduction and compare it with other DR methods including PCA[1], LDA[2], SLPP[19] and SNPE[12]on four famous facial database such as Yale, ORL, AR, and FERET. Before the implement of LDA, LPP and NPE, PCA method is utilized as preprocessing step by retaining 100% energy. The nearest neighbor classifier with Euclidean distance is employed to do classification in the low dimensional space. For LPP, the Gaussian Kernel $\exp\left(-\|x-y\|^2/\sigma^2\right)$ is used and parameter σ is set as the standard derivation of the training data set. LPP is an unsupervised DR method, which does not take the class information into account. In order to perform fair comparison with our SSFA, we propose Supervised LPP in which the similarity matrix including class information. Similar to SLPP, we introduce SNPE which also includes class information. The parameter k which determines the local neighborhood both in SLPP and SNPE is set as $n_{\text{train}} - 1$, where n_{train} is the number of training samples.

5.1. Experiments on Yale Database

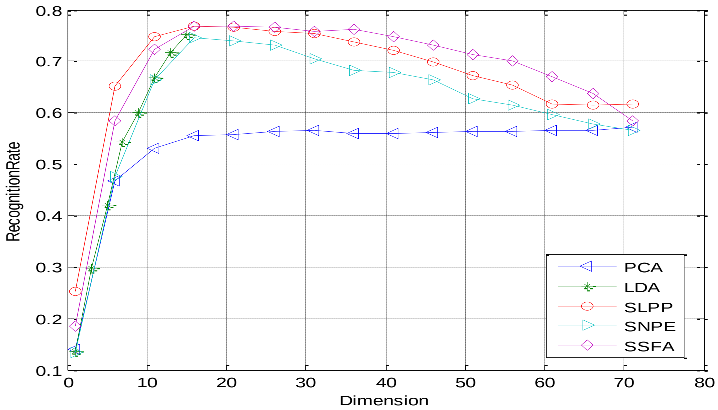
The Yale face database contains 165 images of 15 individuals and each person has 11 different images. The images demonstrate variations in lighting condition (left- light, center-light and right-light), facial expression (normal, happy, sad, sleepy, surprised, and wink), and with or without glasses. In our experiment, all images in Yale database are manually cropped and resized to the resolution of 32×32 pixels. Figure 1 shows 11 different images of one person from Yale database.

For each person, we randomly select p (p=5, 6) images for training, and the rest are used for testing. The image in the training set are used to learn a face subspace, and recognition was obtained in the subspace using NN classifier. For each giving p, we perform the experiment 20 times to randomly choose the training samples and calculate the average recognition rates as well as the standards deviations. We compare our method SSFA with PCA, LDA, SLPP and SNPE. Table 1 shows the maximal average recognition rate and standard deviations of each method. Figure 2 illustrates the plot of recognition rate vs. the dimension of reduced space for different methods.

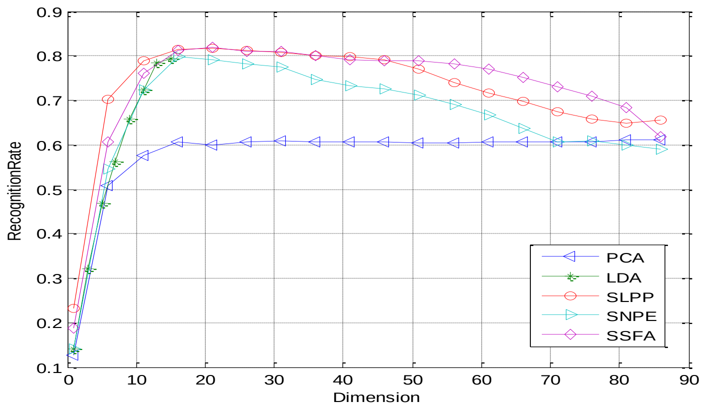
Figure 1. Images of One Person in Yale

Table 1. The Maximal Average Recognition (%) on Yale (mean±std)

Sample size	PCA	LDA	SLPP	SNPE	SSFA
5	57.7±3.9	75.2±3.6	77.0±3.3	75.0±3.3	**77.1±3.7**
6	61.3±4.5	79.3±4.0	81.7±4.1	80.1±3.7	**82.0±3.5**



(a)Recognition curve of all methods in Yale database using 5 trains



(b)Recognition curve of all methods in Yale database using 6 trains

Figure 2. Recognition Rate of PCA, LDA, SLPP, SNPE, and SSFA vs. Dimension of Reduced Space on the Yale Database, (a) 5 Training Samples, (b) 6 Training Samples

5.2. Experiments on ORL Database

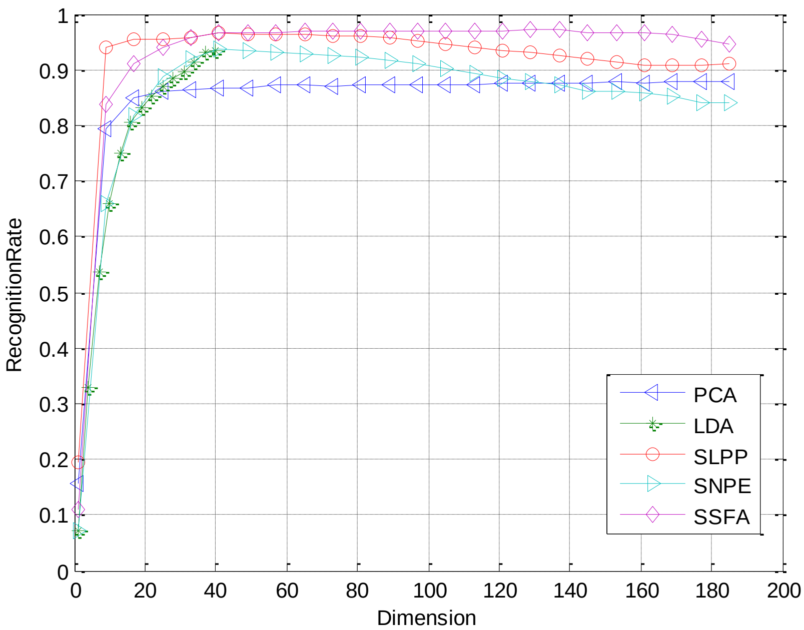
There are 400 images of 40 distinct individuals and each person contains 10 different images in ORL facial database. These images were taken at different times and demonstrates variations in lighting condition, facial expression (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). For computational convenience, we manually cropped the face portion of the image into the resolution of 32×32. Some example images of one person are shown in Figure 3.

We randomly select p (=5, 6) samples of each individual for training, and the rest of the ORL database are used for testing. For each giving p, we perform 20 times the experiments by random training set and show the result in the form of average recognition with standard deviations. Table 2 presents the maximal average recognition and standard deviations for each method. In addition, we draw the recognition rate curves of the representative algorithms in Figure 4.

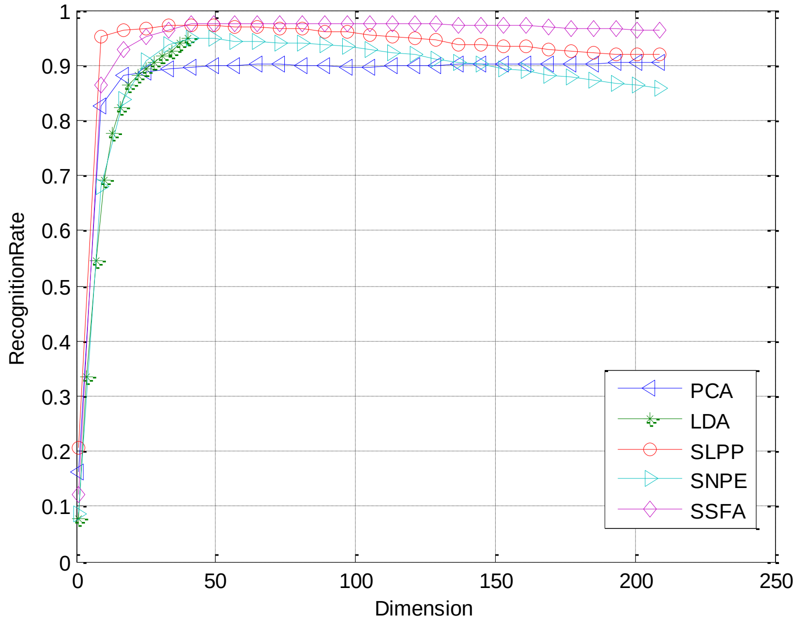
Figure 3. Images of One Person in ORL

Table 2. The Maximal Average Recognition (%) on ORL(mean±std)

Sample size	PCA	LDA	SLPP	SNPE	SSFA
5	87.9±2.5	93.4±1.9	96.6±1.5	94.0±1.8	**97.2±1.1**
6	90.5±2.3	95.0±2.2	97.5±1.5	95.1±2.2	**97.8±1.3**



(a) Recognition curve of all methods in ORL database using 5 trains



(b) Recognition curve of all methods in ORL database using 6 trains

Figure 4. Recognition Rate of PCA, LDA, SLPP, SNPE, SSFA vs. Dimension of Reduced Space on the ORL Database, (a) 5 Training Samples, (b) 6 Training Samples

5.3. Experiments on AR Database

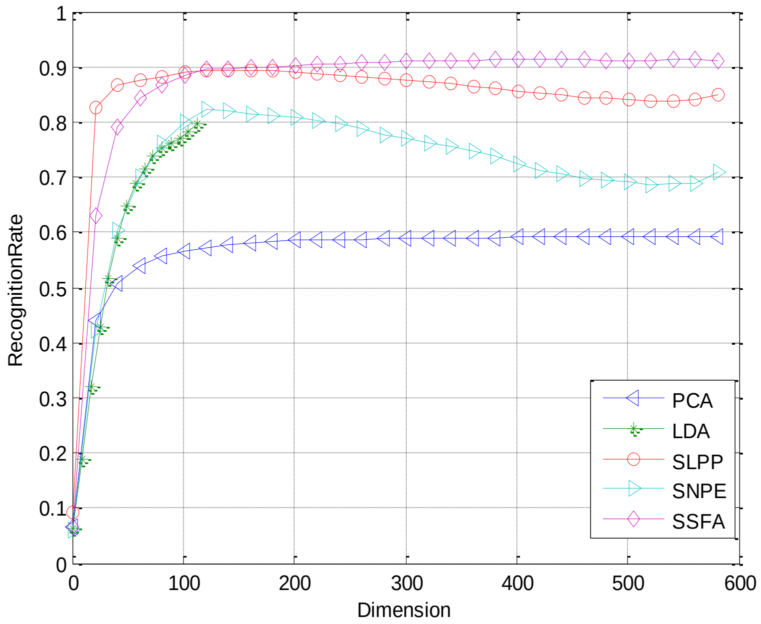
The AR face database consists of 126 subjects with 4000 color face images as a whole. These face images were taken under varying illumination, expression and occlusions. In our experiments, we used about 3120 face images corresponding to 120 persons (60 male and 60 female), each person has 26 face images. Before implement of dimensional reduction methods, we cropped the face portion of the image into the resolution of 32×32. Figure 5 shows some of the faces of one person in the AR database.

We randomly select p (=5, 6) samples of each individual for training, and the rest of the AR database are used for testing. For each giving p, we perform 20 times the experiments by random training set and show the result in the form of average recognition with standard deviations. Table 3 presents the maximal average recognition and standard deviations for each method. In addition, we draw the recognition rate curves of the representative algorithms in Figure 6.

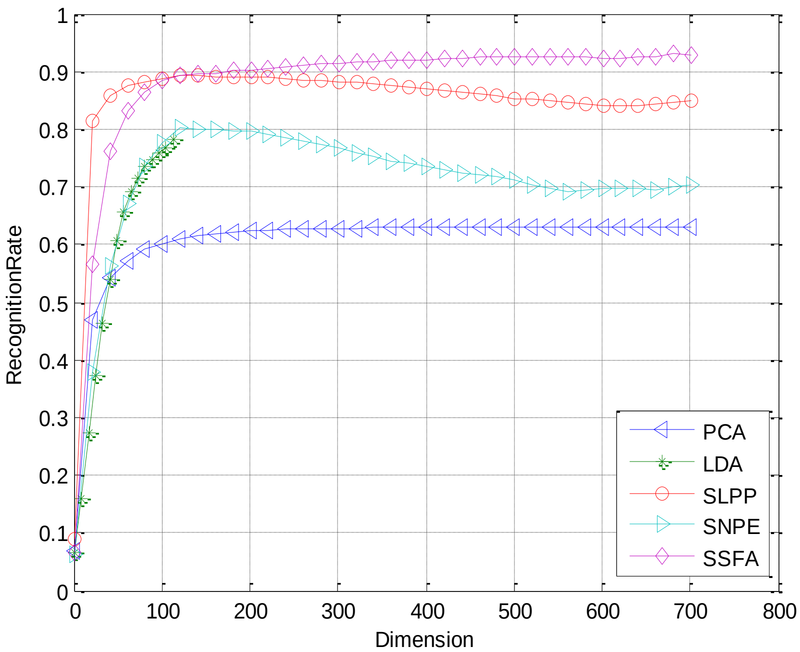
Figure 5. Images of One Person in AR

Table 3. The Maximal Average Recognition (%) on AR(mean±std)

Sample size	PCA	LDA	SLPP	SNPE	SSFA
5	59.2±1.3	81.8±0.8	89.4±0.7	82.4±0.8	**91.5±0.9**
6	63.1±0.8	80.1±1.0	89.3±0.8	80.4±1.0	**93.2±0.8**



(a) Recognition curve of all methods in AR database using 5 trains



(b) Recognition curve of all methods in AR database using 6 trains

Figure 6. Recognition Rate of PCA, LDA, SLPP, SNPE, SSFA vs. Dimension of Reduced Space on the AR Database, (a) 5 Training Samples, (b) 6 Training Samples

5.4. Experiments on Extended FERET Database

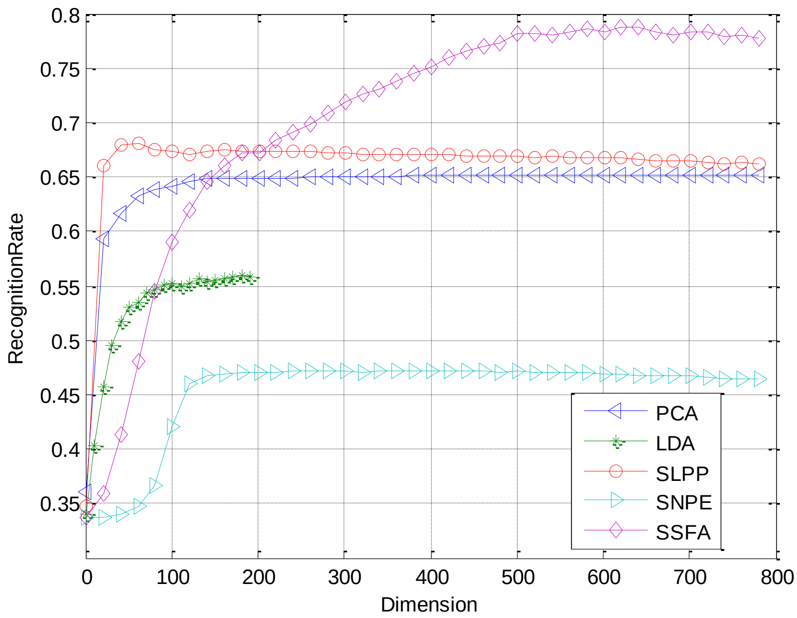
There are 14,126 facial images from 1199 different individuals in the FERET database. These images were taken under different laboratory-controlled lighting conditions. In our experiments, we select a subset which contains 1400 images of 200 persons. We preprocessed the images into the pixel resolution of 32×32. Figure 7 shows seven cropped sample images of one person in the FERET database.

For each individual, p images (p =5, 6) were randomly selected for training and the rest were used for testing. The experimental design is the same as in Section4.1. The maximal average recognition rate and the standard deviations calculated by 20 times random experiments of each method are shown in Table4. In addition, Figure 8 displays the recognition rate curves of the representative algorithms.

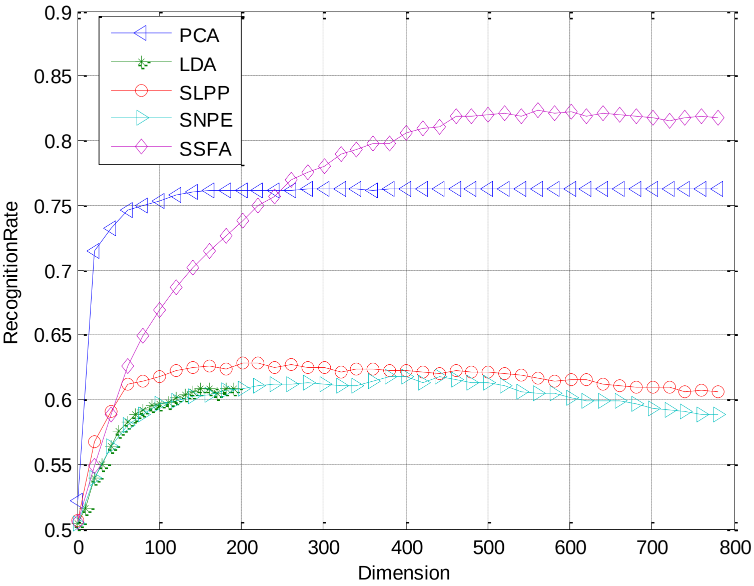
Figure 7. Images of One Person in FERET

Table 4. The Maximal Average Recognition (%) on FERET(mean±std)

Sample size	PCA	LDA	SLPP	SNPE	SSFA
5	65.2±0.9	56.0±1.4	68.2±2.1	47.2±1.4	**83.5±1.1**
6	76.2±1.7	61.0±1.4	63.0±1.5	61.8±1.6	**89.8±1.1**



(a)Recognition curve of all methods in FERET database using 5 trains



(b) Recognition curve of all methods in FERET database using 6 trains

Figure 8. Recognition Rate of PCA, LDA, SLPP, SNPE, and SSFA vs. Dimension of Reduced Space on the FERET Database, (a) 5 Training Samples, (b) 6 Training Samples

5.5. Discussion

From the experiments above, we notice that:

(1) From Table 1-4, it is obviously that our method SSFA outperforms PCA, LDA, SLPP, and SNPE in all four face databases, namely Yale, ORL, AR and FERET. The reason may be that SSFA is a supervised dimensional method which not only preserves neighboring relationship in the low dimensional space but also extracts slow feature for each class that is favorable for classification.

(2) Differing from PCA and LDA which attempts to preserve the global Euclidean structure, SLPP, SNPE and SSFA attempts to preserve the local geometric structure. It can be seen that Local structure based manifold learning algorithm are superior to the methods based on global structure from Figure 2. Especially SSFA obtain higher recognition rate the SLPP and SNPE since it is more robust to the distribution of the data.

(3) Figure 4 shows that SSFA outperforms other four methods markedly when feature dimension exceeds 50 and keep its superiority after that. LDA could get at most $c-1(c=40)$ dimensions, which restricts the ability of classification. Compared with volatility of recognition rate in SLPP and SNPE, SSFA exhibits a robust recognition performance versus the feature dimensions, since SSFA could preserve the underlying geometric structure but also extract invariance for each class which is obvious favorable to classification.

(4) From Table 3, we can see that SSFA enhances the maximal recognition rate for at least 2% compared with other four methods. This result demonstrates that SSFA outperforms PCA, LDA, SLPP, and SNPE both in 5 training samples and 6 training samples. Figure 6 shows the recognition curve versus the variation of dimensions, indicating that SSFA consistently performs better than other methods when feature dimension exceeds 200.

(5) Figure 8 shows that the recognition rate of SSFA is over 77% while other methods obtain their maximal recognition rate less than 77%. These results demonstrate that SSFA is capable of extracting effective features for face recognition task. Besides, we should notice that the five methods achieve high recognition performances with pretty

high-dimensional feature. The reason may be that a larger number of individuals in this experiment require relative larger feature elements for identifying each individual.

6. Conclusions and Future Work

In this paper we have proposed a novel dimensional reduction method based on supervised slow feature analysis (SSFA) to address the problem. We design SSFA for the purpose of preserving the relations of intra-class points and extracting invariant representation for each class though a linear projection. Experimental results on four benchmark face databases indicate the superiority of our method in terms of recognition accuracy.

Acknowledgments

This work is supported by the National Natural Science Fund of China (Grant No. 60632050, 9082004, 61202318), the National 863 Project (Grant No. 2006AA04Z238), the basic key technology project of the Ministry of Industry and Information Technology of China (Grant No. E0310/1112/JC01).

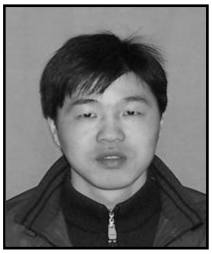
References

[1] M. Turk and A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, vol. 3, no. 1, **(1991)**, pp. 71-86.
[2] P. N. Belhumeur, J. P. Hespanha and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, **(1997)**,pp. 711-720.
[3] B. Schölkopf, A. Smola and K.-R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem", Neural Computation, vol. 10, no. 5, **(1998)** July, pp. 1299-1319.
[4] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, A. Smola, and K.-R. Müller, "Invariant feature extraction and classification in kernel spaces", Advances in Neural Information Processing Systems, vol. 12, **(2000)**,pp. 526–532.
[5] J. Yang, A. F. Frangi, J.-Y. Yang, D. Zhang and Z. Jin, "KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition", IEEE transactions on pattern analysis and machine intelligence, vol. 27, no. 2, **(2005)**,pp. 230–244.
[6] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding", Science (New York, N.Y.), vol. 290, no. 5500, **(2000)**,pp. 2323-2326.
[7] J. B. Tenenbaum, V. de Silva and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction", Science (New York, N.Y.), vol. 290, no. 5500, **(2000)**, pp. 2319-2323.
[8] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation", Neural Computation, vol. 15, no. 6, **(2003)**,pp. 1373-1396.
[9] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment", Journal of Shanghai University (English Edition), vol. 8, no. 4, **(2004)**, pp. 406-424.
[10] Y. Bengio, J. Paiement, P. Vincent, O. Delalleau, N. Le Roux and M. Ouimet, "Out-of-Sample Extensions for LLE, Isomap, MDS, Eigenmaps, and Spectral Clustering", vol. 16, **(2004)**, pp. 177–184.
[11] T. Zhang, J. Yang, D. Zhao and X. Ge, "Linear local tangent space alignment and application to face recognition", Neurocomputing, vol. 70, no. 7–9, **(2007)**, pp. 1547–1553.
[12] X. He, D. Cai, S. Yan and H.-J. Zhang, "Neighborhood preserving embedding", Tenth IEEE International Conference on Computer Vision,, **(2005)**, pp. 1208-1213.
[13] G. E. Hinton, "Connectionist learning procedures", Artificial Intelligence, vol. 40, no. 1-3, **(1989)**, pp. 185–234.
[14] G. Mitchison, "Removing Time Variation with the Anti-Hebbian Differential Synapse", Neural Computation, vol. 3, no. 3, **(1991)**,pp. 312–320.
[15] P. Földiák, "Learning Invariance from Transformation Sequences", Neural computation, vol. 3, no. 2, **(1991)**, pp. 194-200.
[16] L. Wiskott and T. J. Sejnowski, "Slow feature analysis: unsupervised learning of invariances", Neural computation, vol. 14, no. 4, **(2002)**, pp. 715-770.
[17] H. YaPing, Z. JiaLi, L. YunHui, L. SiWei, Q. Zou and M. Tian, "Nonlinear dimensionality reduction using a temporal coherence principle", Information Sciences, vol. 181, no. 16, **(2011)**,pp. 3284–3307.
[18] Z. Zhang and D. Tao, "Slow feature analysis for human action recognition", IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 3, **(2012)**,pp. 436–50.
[19] X. He, S. Yan, Y. Hu, P. Niyogi and H.-J. Zhang, "Face recognition using laplacianfaces", IEEE transactions on pattern analysis and machine intelligence, vol. 27, no. 3, **(2005)**, pp. 328-40.

Authors

XingJian Gu is now working for PhD degree at the department of computer science in Nanjing University of Science and Technology. He received his BS degree in the college of math and physics at Nanjing University of Information Science and Technology in 2009. His research interests mainly focus on computer vision and pattern recognition.

Chuancai Liu is a full professor in the school of computer science and engineering of Nanjing University of Science and Technology, China. He obtained his PhD degree from the China Ship Research and Development Academy in 1997. His research interests include AI, pattern recognition and computer vision. He has published about 50 papers in international/national journals.



Zhangjing Yang received his BS degree in Computer science and education from Nanjing normal university in 2004, his MS degree in computer applications from Nanjing University of Science and Technology in 2009, respectively. He is currently pursuing a PhD in computer applications at Nanjing University of Science and Technology, China. His current research interests include pattern recognition and computer vision.