An Improved method of Two Stage Linear Discriminant Analysis

Yarui Chen, Xin Tao, Congcong Xiong, Jucheng Yang*
Computer Science and Information Engineering Institute,
Tianjin University of Science and Technology, Tianjin 300222, China
[e-mail: yrchen@tust.edu.cn, xtao@mail.tust.edu.cn, xiongcc@tust.edu.cn, jcyang@tust.edu.cn]
*Corresponding author: Jucheng Yang

Received June 22, 2017; revised October 19, 2017; accepted November 19, 2017; published March 31, 2018

Abstract

The two-stage linear discrimination analysis (TSLDA) is a feature extraction technique to solve the small size sample problem in the field of image recognition. The TSLDA has retained all subspace information of the between-class scatter and within-class scatter. However, the feature information in the four subspaces may not be entirely beneficial for classification, and the regularization procedure for eliminating singular metrics in TSLDA has higher time complexity. In order to address these drawbacks, this paper proposes an improved two-stage linear discriminant analysis (Improved TSLDA). The Improved TSLDA proposes a selection and compression method to extract superior feature information from the four subspaces to constitute optimal projection space, where it defines a single Fisher criterion to measure the importance of single feature vector. Meanwhile, Improved TSLDA also applies an approximation matrix method to eliminate the singular matrices and reduce its time complexity. This paper presents comparative experiments on five face databases and one handwritten digit database to validate the effectiveness of the Improved TSLDA.

Keywords: Two-stage linear discrimination analysis, Small size sample problem, Selection and compression method, Approximation matrix method, Regularization

http://doi.org/10.3837/tiis.2018.03.015 **ISSN : 1976-7277**

1. Introduction

Linear discriminant analysis (LDA) is a classical method for dimensionality reduction and feature extraction, and it has been widely used in the face recognition, fingerprint recognition, gait recognition and other fields. LDA projects the training samples from higher dimensional space to lower dimentisonal space through maximizing the between-class scatter and minimizing the within-class scatter in the lower dimensional space. Classical LDA aims to find an optimal projection matrix $\mathbf{W} \in R^{d \times h} (h < d)$ by maximizing the loss function $J(\mathbf{W})$, which is the radio of the between-class scatter matrix $\mathbf{S}_b \in R^{d \times d}$ to the within-class scatter matrix $\mathbf{S}_w \in R^{d \times d}$, such that $J(\mathbf{W})=|\mathbf{W}^{\mathrm{T}}\mathbf{S}_b\mathbf{W}|/|\mathbf{W}^{\mathrm{T}}\mathbf{S}_w\mathbf{W}|$ [1,2]. For the situation, the projection matrix $\mathbf{W}$ is determined by the eigenvectors of $\mathbf{S}_w^{-1}\mathbf{S}_b$ using eigenvalue decomposition (EVD) . However, in many pattern classification applications, the matrix $\mathbf{S}_w$ is singular and the projection $\mathbf{W}$ can not be directly computed using EVD of $\mathbf{S}_w^{-1}\mathbf{S}_b$. This problem is commonly called small size sample (SSS) problem, where the number of sample dimension is larger than the number of samples. [3]. Classical LDA cann't be applied to the SSS problem directly in the applications of face recognition, image recognition and so on.

For the SSS problem, current researches based on LDA mainly can be divided into two categories: one is to eliminate the singularity of $\mathbf{S}_w$ directly, and the other is based on linear subspace analysis. The former class technique contains Fisherface[4], Regularized LDA(RLDA) [5], Direct Regularized LDA (DRLDA) [6] and Approximate LDA (ALDA) [7]. Fisherface firstly reduces dimension of samples using principal component analysis (PCA), which makes the total scatter matrix $\mathbf{S}_t$ to be full rank, and then applies LDA to obtain optimal projection space. However, some important feature information may be discarded in the dimensionality reduction of PCA, and it also can not be guaranteed that $\mathbf{S}_w$ is a non-singular matrix. RLDA makes matrix $\mathbf{S}_w$ non-singular through adding a regularization term $\alpha\mathbf{I}$ to the diagonal elements of $\mathbf{S}_w$, where $\alpha > 0$ is regularzation parameter and $\mathbf{I}$ is identity matrix, and then obtains an optimal projection space by EVD of $(\mathbf{S}_w + \alpha\mathbf{I})^{-1}\mathbf{S}_b$ [8,9]. RLDA eliminates the singularity of matrix and retains both the null space and range space of $\mathbf{S}_w$. But it has high time complexity since the regularization parameter α is evaluated by cross-validation method, and the granularity selection for cross-validation is important [10]. And a poor value of α can degrade the generalization performance of RLDA[11]. DRLDA calculates regularization parameters directly and avoids the cross-validation process of RLDA to improve training efficiency [6]. Instead of heuristic methods for estimating the regularization parameters, ALDA introduces a reversible approximation matrix to eliminate matrix singularity by replacing original eigenvalue matrix of $\mathbf{S}_w$. Particularly, ALDA presents better recognition accuracy than RLDA and DRLDA and less time complexity than DRLDA [7]. These algorithms mentioned above are all based on eliminating the singularity of $\mathbf{S}_w$ to solve SSS problem directly.

The other category technique, which is based on linear subspace analysis, incorporates null space and range space information of $\mathbf{S}_w$ and $\mathbf{S}_b$ respectively. This class technique mainly contains the Null LDA(NLDA) [12], Direct LDA(DLDA) [13] and Two Stage LDA(TSLDA) [14]. For $\mathbf{S}_w$ and $\mathbf{S}_b$, there are four feature subspaces namely null space of $\mathbf{S}_w$ ($\mathbf{S}_w^{null}$), range space of $\mathbf{S}_w$ ($\mathbf{S}_w^{range}$), null space of $\mathbf{S}_b$ ($\mathbf{S}_b^{null}$), range space of $\mathbf{S}_b$ ($\mathbf{S}_b^{range}$). NLDA firstly projectes the training samples on the null space of $\mathbf{S}_w$, and then finds $\mathbf{W}$ in the range space of $\mathbf{S}'_b$ which satisfies $\mathbf{S}'_b\mathbf{W} \neq 0$ and maximizes $\left|\mathbf{W}^{\mathrm{T}}\mathbf{S}'_b\mathbf{W}\right|$. DLDA firstly transforms the training samples to the range space of $\mathbf{S}_b$, and then find $\mathbf{W}$ in the range space of $\mathbf{S}'_w$ by minimizing $\left|\mathbf{W}^{\mathrm{T}}\mathbf{S}'_w\mathbf{W}\right|$. Based on the above analysis, NLDA retains $\mathbf{S}_w^{null}$ and $\mathbf{S}_b^{range}$, while DLDA contains $\mathbf{S}_w^{range}$ and $\mathbf{S}_b^{range}$ [15]. However, all these four individual subspaces may have some significant feature information for classification [3].

TSLDA exploits all the four subspaces $\mathbf{S}_w^{null}$, $\mathbf{S}_b^{range}$, $\mathbf{S}_w^{range}$ and $\mathbf{S}_b^{null}$ to constitute the optimal projection space $\mathbf{W}$ [16-18], and it has been confirmed to own better results in feature extraction than NLDA and DLDA[19]. But TSLDA has high time complexity in eliminating singular matrix since it determines regularization parameters using cross-validation method. Meanwhile, the feature information in these subspaces may not be entirely beneficial for classification. Therefore, it is necessary to extract superior feature vectors in the projection space of TSLDA to improve the classification performance.

Face image classification is an important type of SSS problem. Recently, there are many researches for face recognition methods, including singular value decomposition frameworks for low resolution image and face-hallucination [20,21], the hierarchical scheme for facial-feature detection and localization [22], and the potential-field representation method for face-image retrieval [23]. Face feature learning and represetation approaches also have develeped rapidly, including the discriminative feature learning approach for deep face recognition [24], the structured subspace learning approach [25], and the clustering-guided sparse structural learning approach [26].

This paper focus on the linear discriminant analysis method to solve the SSS problem. The TSLDA method has used all the four subspaces $\mathbf{S}_w^{null}$, $\mathbf{S}_b^{range}$, $\mathbf{S}_w^{range}$ and $\mathbf{S}_b^{null}$ to constitute the optimal projection space, but it has high time complexity and the full feature information in the four subspaces may not be entirely beneficial for classification. In order to address the drawbacks of TSLDA, this paper proposes an improved method of TSLDA(Improved TSLDA). On the one hand, the Improved TSLDA eliminates the singular matrix $\mathbf{S}_w$ and $\mathbf{S}_b$ using an approximate matrix method to reduce the time complexity, where it approximately computes the inverse of original eigenvalue matrix with a reverse eigenvalue matrix. On the other hand, the Improved TSLDA explores a selection and compression method to extract superior feature vectors in the four subspaces, where we defines a single Fisher criterion to measure the importance of single feature vector. This paper also presents comparative experiments on five face recognition databases as ORL, YALE, AR, FERET and CMU-PIE and a handwirtten digit database as MNIST to validate the effectiveness of the Improved TSLDA.

2. Related Work

Let $\{\mathbf{x}_i, t_i\}_{i=1}^n, \mathbf{x}_i \in R^d$ denote n training samples in a d-dimensional space having class labels $t_i \in \{1,2,...,c\}$,where c is the number of classes. The dataset $\{\mathbf{x}_i, t_i\}_{i=1}^n$ can be divided into c subsets $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_c], \mathbf{X}_j \in R^{d\times n_j}$ where $\mathbf{X}_j$ belongs to class j and consists of n_j number of training samples such that $n = \sum_{j=1}^{c} n_j$. The within-class scatter matrix $\mathbf{S}_w \in R^{d\times d}$, the between-class scatter matrix $\mathbf{S}_b \in R^{d\times d}$ and total scatter matrix $\mathbf{S}_t \in R^{d\times d}$ are respectively denoted as:

$$\mathbf{S}_w = \sum_{j=1}^{c} \sum_{\mathbf{x}_i \in \mathbf{X}_j} (\mathbf{x}_i - \mathbf{m}_j)(\mathbf{x}_i - \mathbf{m}_j)^{\mathrm{T}} \tag{1}$$

$$\mathbf{S}_b = \sum_{j=1}^{c} n_j (\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^{\mathrm{T}} \tag{2}$$

$$\mathbf{S}_t = \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^{\mathrm{T}} \tag{3}$$

where $\mathbf{m} = (1/n)\sum_{i=1}^{n} \mathbf{x}_i$ is the centriod of $\mathbf{X}$, and $\mathbf{m}_j = (1/n_j)\sum_{j=1}^{c} \sum_{\mathbf{x}_i \in \mathbf{X}_j} \mathbf{x}_i$ is the centriod of $\mathbf{X}_j$.

Classical LDA aims to find a projection space $\mathbf{W}$ based on the Fisher criterion such that $\mathbf{W} = \arg\max_{\mathbf{W}}\{tr((\mathbf{W}^{\mathrm{T}}\mathbf{S}_w\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\mathbf{S}_b\mathbf{W})\}$. To maximize the Fisher criterion, the optimal projection space $\mathbf{W}$ can be computed by EVD of $\mathbf{S}_w^{-1}\mathbf{S}_b$ with the top c-1 eigenvalues corresponding to eigenvectors[27,28]. However, it is impossible to obtain the eigenvectors of $\mathbf{S}_w^{-1}\mathbf{S}_b$ directly since the $\mathbf{S}_w$ is singular in image recognition.

TSLDA utilizes four subspaces and the optimal projection space can be computed from the input samples by carrying out two discriminant analysis in parallel [29]. In the first analysis, the projection space $\mathbf{W}_1$ is computed by retaining nonzero eigenvalues of $\mathbf{S}_w'^{-1}\mathbf{S}_b$ corresponding to eigenvectors, where non-singular matrix $\mathbf{S}'$ is the regularization form of $\mathbf{S}$. The $\mathbf{W}_1$ includes null space of $\mathbf{S}_w$ and range space of $\mathbf{S}_b$. In the second analysis, the projection space $\mathbf{W}_2$ that is computed by retaining top eigenvalues of $\mathbf{S}_b'^{-1}\mathbf{S}_w$ corresponding to eigenvectors, and it includes null space of $\mathbf{S}_b$ and range space of $\mathbf{S}_w$. The projection spaces obtained by the two-stage analysis are concatenated to get total projection space $\mathbf{W}$,i.e., $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2]$. The details of TSLDA are as follows:

Firstly, singluar matrices $\mathbf{S}_w$ and $\mathbf{S}_b$ are regularized to be non-singular matrices $\mathbf{S}_w'$ and $\mathbf{S}_b'$ using cross-validation method in determining regularization parameter:

$$\mathbf{S}_w' = \mathbf{S}_w + \alpha_1 \mathbf{I} \tag{4}$$

$$\mathbf{S}_b' = \mathbf{S}_b + \alpha_2 \mathbf{I} \tag{5}$$

In order to extract the range space of $\mathbf{S}_w'^{-1}\mathbf{S}_b$, TSLDA carries out the EVD of $\mathbf{S}_w'^{-1}\mathbf{S}_b$:

$$\mathbf{S}_w'^{-1}\mathbf{S}_b\mathbf{E}_1 = \sigma\mathbf{E}_1 \tag{6}$$

where $\sigma = diag(\sigma_1,\sigma_2,...,\sigma_{r_b},...,\sigma_d)$ is the diagonal matrix of eigenvalue that satisfies $\sigma_1 > \sigma_2 > ... > \sigma_{r_b} > \sigma_{r_b+1} = ... = \sigma_d = 0$, and $r_b = rank(\mathbf{S}_b) = c-1$. The eigenvector matrix $\mathbf{E}_1$ is given by:

$$\mathbf{E}_1 = [\mathbf{E}_{R_1}, \mathbf{E}_{N_1}] \tag{7}$$

where $\mathbf{E}_{R_1} \in R^{d\times r_b}$ and $\mathbf{E}_{N_1} \in R^{d\times(d-r_b)}$ respectively represent range space and null space of $\mathbf{S}_w'^{-1}\mathbf{S}_b$. Particularly, $\mathbf{E}_{R_1} \in R^{d\times r_b}$ is the only effective projection space to be reserved, i.e., $\mathbf{W}_1 = \mathbf{E}_{R_1}$.

Secondly, TSLDA introduces range space of $\mathbf{S}_b'^{-1}\mathbf{S}_w$ to approximate the null space of $\mathbf{S}_w'^{-1}\mathbf{S}_b$ [14], and it also carries out EVD of $\mathbf{S}_b'^{-1}\mathbf{S}_w$:

$$\mathbf{S}_b'^{-1}\mathbf{S}_w\mathbf{E}_2 = \delta\mathbf{E}_2 \tag{8}$$

where $\delta = diag(\delta_1,\delta_2,...,\delta_{r_b},...,\delta_{r_w},...,\delta_d)$ is the diagonal matrix of eigenvalue that satisfies $\delta_1 > \delta_2 > ... > \delta_{r_b} > \delta_{r_b+1} > ... > \delta_{r_w} = ... = 0$ and $r_w = rank(\mathbf{S}_w) = n-c$. The eigenvector matrix $\mathbf{E}_2$ is given by:

$$\mathbf{E}_2 = [\mathbf{E}_{R_2}, \mathbf{E}_{N_2}] \tag{9}$$

where $\mathbf{E}_{R_2} \in R^{d\times r_w}$ and $\mathbf{E}_{N_2} \in R^{d\times(d-r_w)}$ respectively denote range space and null space of $\mathbf{S}_b'^{-1}\mathbf{S}_w$. Due to $r_b < r_w$, the important discriminant information in $\mathbf{E}_{R_2} \in R^{d\times r_w}$ is calculated by the top r_b eigenvalues corresponding to eigenvectors to constitute the effective projection space $\mathbf{W}_2 = \mathbf{E}_{RL} \in R^{d\times r_b}$. Finally, these two projection spaces are concatenated to get the total projection space **W**:

$$\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2] \tag{10}$$

and $\mathbf{W} \in R^{d\times 2r_b}$.

3. Improved TSLDA

3.1 Motivation Analysis

For the within-class scatter matrix $\mathbf{S}_w$ and the between-class scatter matrix $\mathbf{S}_b$, there are four feature subspaces namely null space of $\mathbf{S}_w$ ($\mathbf{S}_w^{null}$), range space of $\mathbf{S}_w$ ($\mathbf{S}_w^{range}$), null space of $\mathbf{S}_b$ ($\mathbf{S}_b^{null}$), and range space of $\mathbf{S}_b$ ($\mathbf{S}_b^{range}$). Traditional researches show that spaces $\mathbf{S}_w^{null}$ and $\mathbf{S}_b^{range}$ obtain the main information [12,13]. Moreover, some resent works show that the spaces $\mathbf{S}_w^{range}$ and $\mathbf{S}_b^{null}$ also can improve the classification accuracy [14,15]. Hence, all these four individual subspaces may have some significant feature information for classification.

TSLDA computes the optimal projection space through the four subspaces by carrying out two discriminant analysis in parallel [29]. The first projection space $\mathbf{W}_1$ is computed by retaining nonzero eigenvalues of $\mathbf{S}_w'^{-1}\mathbf{S}_b$ corresponding to eigenvectors, where non-singular matrix $\mathbf{S}'$ is the regularization form of S. The $\mathbf{W}_1$ includes null space of $\mathbf{S}_w$ and range space of $\mathbf{S}_b$. The second projection space $\mathbf{W}_2$ that is computed by retaining top eigenvalues of $\mathbf{S}_b'^{-1}\mathbf{S}_w$ corresponding to eigenvectors, and it includes null space of $\mathbf{S}_b$ and range space of $\mathbf{S}_w$. TSLDA combines the two projection spaces, that is $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2]$.

Since TSLDA algorithm determines regularization parameters using cross-validation method, and it has high time complexity in eliminating singular matrix. Meanwhile, the full feature information in the four subspaces may not be entirely beneficial for classification, and it is necessary to extract superior feature vectors in the four projection spaces to improve the classification performance. This paper proposes an improved method of Two Stage Linear Discriminant Analysis (Improved TSLDA). Improved TSLDA eliminates the singular matrices $\mathbf{S}_w$ and $\mathbf{S}_b$ and reduces its time complexity using an approximate matrix method. Meanwhile, the Improved TSLDA proposes a selection and compression method to extract superior feature vectors from the four projection spaces and compresses original projection space.

3.2 Improved TSLDA Algorithm

The Improved TSLDA firstly reduces dimensionality of all samples to simplify calculation using PCA method. Then an approximate matrix method is introduced to estimate the singular matrices $\hat{\mathbf{S}}_w$ and $\hat{\mathbf{S}}_b$ and eliminate matrix singularity through approximately computing the inverse of original eigenvalue matrix with a reverse eigenvalue matrix. Next, Improved TSLDA integrates null space and range space of $\hat{\mathbf{S}}_w$ and $\hat{\mathbf{S}}_b$ to constitute projection space $\mathbf{W}$. Finally, we apply selection and compression method to extract superior feature vectors in $\mathbf{W}$ to obtain optimal projection space $\mathbf{W}_{opt}$. The **Fig. 1** shows the a schematic diagram of Improved TSLDA, and the detailed description of this algorithm is listed as follows:

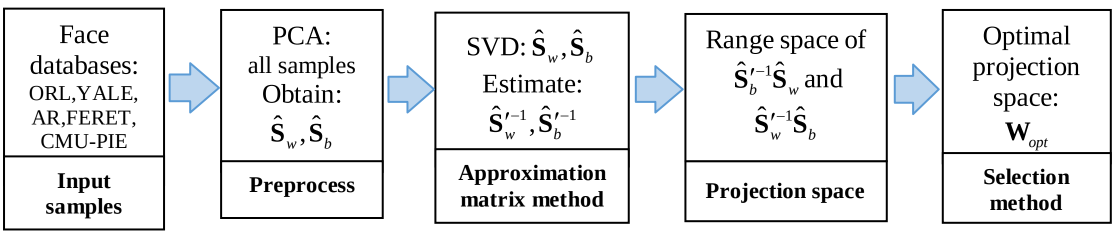


Fig. 1. Schematic diagram of Improved TSLDA

(1) Pre-processing stage: apply PCA to reduce dimension of all samples. The SVD of total scatter matrix $\mathbf{S}_t$ is given by:

$$\mathbf{S}_t = \mathbf{U}_t \Sigma \mathbf{U}_t^{\mathrm{T}} \tag{11}$$

where $\Sigma = \begin{bmatrix} \Sigma_t & 0 \\ 0 & 0 \end{bmatrix}, \Sigma_t \in R^{r_t \times r_t}$ is eigenvalue matrix with the rank $r_t = n-1$, $\mathbf{U}_t = [\mathbf{U}_{TR}, \mathbf{U}_{TN}]$ is eigenvector matrix of $\mathbf{S}_t$, and $\mathbf{U}_{TR} \in R^{d \times r_t}$, $\mathbf{U}_{TN} \in R^{d \times (d-r_t)}$ respectively denote range space and null space of $\mathbf{S}_t$. All samples are projected on $\mathbf{U}_{TR}$ and dimensionality is reduced from d to r_t (d>r_t), and the transformed between-class scatter matrix is $\hat{\mathbf{S}}_w \in R^{r_t \times r_t}$ and the transformed within-class scatter matrix is $\hat{\mathbf{S}}_b \in R^{r_t \times r_t}$.

(2) Eliminate singluar matrices: Eliminate the singularity of $\hat{\mathbf{S}}_w, \hat{\mathbf{S}}_b$ using approximation matrix method. Due to $rank(\hat{\mathbf{S}}_w) = n - c$ and $rank(\hat{\mathbf{S}}_b) = c - 1$, the rank relationship of each scatter matrix becomes $rank(\hat{\mathbf{S}}_t) > rank(\hat{\mathbf{S}}_w) > rank(\hat{\mathbf{S}}_b)$. It is obvious that $\hat{\mathbf{S}}_w$ and $\hat{\mathbf{S}}_b$ are still singular matrices in a r_t–dimensional space, hence performs SVD with them as follow:

$$\hat{\mathbf{S}}_w = \hat{\mathbf{U}}_w \hat{\mathbf{D}}_w \hat{\mathbf{U}}_w^{\mathrm{T}} \tag{12}$$

$$\hat{\mathbf{S}}_b = \hat{\mathbf{U}}_b \hat{\mathbf{D}}_b \hat{\mathbf{U}}_b^{\mathrm{T}} \tag{13}$$

where $\hat{\mathbf{U}}_w \in R^{r_t \times r_t}$ and $\hat{\mathbf{U}}_b \in R^{r_t \times r_t}$ are eigenvectors, $\hat{\mathbf{D}}_w = \begin{bmatrix} \mathbf{\Lambda}_w & 0 \\ 0 & 0 \end{bmatrix}$ and $\hat{\mathbf{D}}_b = \begin{bmatrix} \mathbf{\Lambda}_b & 0 \\ 0 & 0 \end{bmatrix}$ are eigenvalues. The inverse computation of $\hat{\mathbf{S}}_w$ and $\hat{\mathbf{S}}_b$ are shown as:

$$\hat{\mathbf{S}}_w^{-1} = \hat{\mathbf{U}}_w \hat{\mathbf{D}}_w^{-1} \hat{\mathbf{U}}_w^{\mathrm{T}} \tag{14}$$

$$\hat{\mathbf{S}}_b^{-1} = \hat{\mathbf{U}}_b \hat{\mathbf{D}}_b^{-1} \hat{\mathbf{U}}_b^{\mathrm{T}} \tag{15}$$

Since the eigenvalue matrices $\hat{\mathbf{D}}_w$ and $\hat{\mathbf{D}}_b$ are singluar and irreversible, let us denote:

$$\hat{\mathbf{D}}_{\alpha w} = \alpha \mathbf{I} - \hat{\mathbf{D}}_w \tag{16}$$

where $\alpha = \max(diag(\hat{\mathbf{D}}_w))$, $\mathbf{I}$ is identity matrix, and we can substitute $\hat{\mathbf{D}}_w^{-1}$ with the nonsingular eigenvalue matrix $\hat{\mathbf{D}}_{\alpha w}$. Thus, $\hat{\mathbf{S}}_w'^{-1}, \hat{\mathbf{S}}_b'^{-1}$ are denoted to approximately estimate $\hat{\mathbf{S}}_w^{-1}, \hat{\mathbf{S}}_b^{-1}$ and the inverse of $\hat{\mathbf{S}}_w$ can be given by:

$$\hat{\mathbf{S}}_w'^{-1} = \hat{\mathbf{U}}_w \hat{\mathbf{D}}_{\alpha w} \hat{\mathbf{U}}_w^{\mathrm{T}} \tag{17}$$

The inverse computation of $\hat{\mathbf{S}}_b$ can be treated in the same manner,

$$\hat{\mathbf{S}}_b'^{-1} = \hat{\mathbf{U}}_b \hat{\mathbf{D}}_{\alpha b} \hat{\mathbf{U}}_b^{\mathrm{T}} \tag{18}$$

(3) Analyze projection space: compute and concatenate two projection space to get $\mathbf{W}$. Improved TSLDA obtains the feature space $\mathbf{E}_1 = [\mathbf{E}_{R_1}, \mathbf{E}_{N_1}]$ of $\hat{\mathbf{S}}_w'^{-1}\hat{\mathbf{S}}_b$ using EVD method and selects the range space $\mathbf{E}_{R_1} \in R^{r_t \times r_b}$ of $\hat{\mathbf{S}}_w'^{-1}\hat{\mathbf{S}}_b$ as projection space $\mathbf{W}_1 = \mathbf{E}_{R_1}$. Similarly, the

projection space $\mathbf{W}_2 = \mathbf{E}_{RL} \in R^{r_t \times r_b}$ is constituted by the top r_b eigenvectors in the range space $\mathbf{E}_{R_2}$ in $\mathbf{E}_2 = [\mathbf{E}_{R_2}, \mathbf{E}_{N_2}]$ where $\mathbf{E}_2$ denotes the eigenvector space of $\hat{\mathbf{S}}_b'^{-1}\hat{\mathbf{S}}_w$. The projection spaces $\mathbf{W}_1$ and $\mathbf{W}_2$ are concatenated to get the total projection space $\mathbf{W}$:

$$\mathbf{W} = \left[\mathbf{W}_1, \mathbf{W}_2\right] \tag{19}$$

(4) Selection and compression method: define a single Fisher criterion to measure the importance of single feature vector. Under the condition of maximizing single Fisher criterion, Improved TSLDA removes the noise information, and extracts superior feature vectors in the projection space of TSLDA to get optimal projection space. The single Fisher criterion d_i is defined as:

$$d_i = \frac{\mathbf{W}_{\cdot i}^{\mathrm{T}} \hat{S}_b \mathbf{W}_{\cdot i}}{\mathbf{W}_{\cdot i}^{\mathrm{T}} \hat{S}_w \mathbf{W}_{\cdot i}}, i = 1, 2, ..., 2r_b \tag{20}$$

Each feature column vector $\mathbf{W}_{\cdot i}$ in $\mathbf{W}$ is substituted into (20) that will obtain the set of single Fisher criterion $D = \{d_1, d_2, ..., d_{2r_b}\}$. In order to measure the important feature vectors in $\mathbf{W}$, the element value in set D is limited to $d_i \geq \varphi > 0$, and the favorable elements are selected to constitute a new set $D' = \{d_1, d_2, ..., d_g\}, 1 < g \leq 2r_b$. The selected single Fisher criterion $\{d_1, d_2, ..., d_g\}$ that corresponds the g-th column vectors in $\mathbf{W}$ are retained to constitute optimal projection space $\mathbf{W}_{opt}$:

$$\mathbf{W}_{opt} = \left[\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, ..., \mathbf{W}_{\cdot g}\right], 0 < g \leq 2r_b \tag{21}$$

where g is defined as optimal projection parameter, and all samples owning maximum class separability and best classification performance in this projection space. However, the parameter g is various for each database, and the computational cost is much expensive to find it by traversing all values. Thus, Improved TSLDA integrates random sampling and key point selection methods to find parameter g. The random sampling means the value of d_i is under such constraints $\boldsymbol{\varphi}$=0.2**a**, $\boldsymbol{\varphi}$=0.1**a** and $\boldsymbol{\varphi}$=0.05**a** where **a**=max(d_i), and key point selection aims to select two key value of $\boldsymbol{\varphi}$ for $\boldsymbol{\varphi}$=1 and $\boldsymbol{\varphi}$=0, which respectively represent the boundary of projection space $\mathbf{W}_1$ and projection space $\mathbf{W}$. Only the projection space with optimal training accuracy can be retained as optimal projection space $\mathbf{W}_{opt}$, and the pseudo-code description of Improved TSLDA is illustrated as **Algorithm 1**.

3.3 Computational Complexity Analysis

For Improved TSLDA, the time complexity of each step in **Algorithm 1** are respectively represented by O(dc), O(dn^2), O(n^3), O(n^3), O(c), O(dn^2), and the final time complexity can be estimated as O(dn^2). The time complexity of TSLDA can be estimated as O(d^2n). ALDA, NLDA and Fisherface has the same time complexity for O(dn^2). Since $d >> n, n > c$, it is obvious that $O(dn^2) << O(d^2n)$ and time complexity of Improved LDA is significantly lower than that of TSLDA. Therefore, according to the theoretical derivation above, the proposed algorithm has a positive effect in improving classification performance, and it will be confirmed in the next experiments.

Algorithm 1. An Improved method of Two Stage Linear Discriminant Analysis

Data: A training dataset $\{\mathbf{x}_i, t_i\}_{i=1}^{n}, \mathbf{x}_i \in R^d$, number of class c,
Result: Optimal projection space $\mathbf{W}_{opt}$, class label $\boldsymbol{t_i}$

Perform iteratively until no test sample.

1. **Calculate scatter matrices :**

   within-class scatter matrix: $\mathbf{S}_w = \sum_{j=1}^{c} \sum_{\mathbf{x}_i \in \mathbf{X}_j} (\mathbf{x}_i - \mathbf{m}_j)(\mathbf{x}_i - \mathbf{m}_j)^{\mathrm{T}}$;

   between-class scatter matrix: $\mathbf{S}_b = \sum_{j=1}^{c} n_j (\mathbf{m}_j - \mathbf{m})(\mathbf{m}_j - \mathbf{m})^{\mathrm{T}}$;

   total scatter matrix: $\mathbf{S}_t = \sum_{i=1}^{n} (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^{\mathrm{T}}$.

2. **Pre-processing stage:** reduce dimension of all samples using PCA.

   The transformed between-class scatter matrix: $\hat{\mathbf{S}}_w = \mathbf{U}_{TR}^{\mathrm{T}} \mathbf{S}_w \mathbf{U}_{TR} \in R^{r_t \times r_t}$.

   The transformed within-class scatter matrix: $\hat{\mathbf{S}}_b = \mathbf{U}_{TR}^{\mathrm{T}} \mathbf{S}_b \mathbf{U}_{TR} \in R^{r_t \times r_t}$.

3. **Eliminate singluar matrices $\hat{\mathbf{S}}_w, \hat{\mathbf{S}}_b$ using approximation matrix method.**

   SVD with $\hat{\mathbf{S}}_w : \hat{\mathbf{S}}_w = \hat{\mathbf{U}}_w \hat{\mathbf{D}}_w \hat{\mathbf{U}}_w^{\mathrm{T}}$,

   SVD with $\hat{\mathbf{S}}_b : \hat{\mathbf{S}}_b = \hat{\mathbf{U}}_b \hat{\mathbf{D}}_b \hat{\mathbf{U}}_b^{\mathrm{T}}$.

   Approximation matrix method: substitute $\hat{\mathbf{D}}_w$ with $\hat{\mathbf{D}}_{\alpha w} = \alpha \mathbf{I} - \hat{\mathbf{D}}_w$ in $\hat{\mathbf{S}}_w$.

   Inverse with $\hat{\mathbf{S}}_w : \hat{\mathbf{S}}_w'^{-1} = \hat{\mathbf{U}}_w \hat{\mathbf{D}}_{\alpha w} \hat{\mathbf{U}}_w^{\mathrm{T}}$,

   Inverse with $\hat{\mathbf{S}}_b : \hat{\mathbf{S}}_b'^{-1} = \hat{\mathbf{U}}_b \hat{\mathbf{D}}_{\alpha b} \hat{\mathbf{U}}_b^{\mathrm{T}}$.

4. **Construct projection space W:**

   The range space of $\hat{\mathbf{S}}_w'^{-1} \hat{\mathbf{S}}_b : \mathbf{W}_1 = \mathbf{E}_{R_1}$

   The top r_b eigenvectors in range space of $\hat{\mathbf{S}}_b'^{-1} \hat{\mathbf{S}}_w : \mathbf{W}_2 = \mathbf{E}_{RL} \in R^{r_t \times r_b}$

   The projection space: $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2]$.

5. **Selection and compression method:**

   Maximize single Fisher criterion: $d_i = \dfrac{\mathbf{W}_{\cdot i}^{\mathrm{T}} \hat{S}_b \mathbf{W}_{\cdot i}}{\mathbf{W}_{\cdot i}^{\mathrm{T}} \hat{S}_w \mathbf{W}_{\cdot i}}, i = 1, 2, \ldots, 2r_b$.

   Limit condition: $\{d_1, d_2, \ldots, d_g\}$ satisfying $d_i \geq \varphi \geq 0$.

   Optimal projection space: $\mathbf{W}_{opt} = [\mathbf{W}_{\cdot 1}, \mathbf{W}_{\cdot 2}, \ldots, \mathbf{W}_{\cdot g}], 0 < g \leq 2r_b$.

6. **Classification (KNN classifer):**

   Projected test samples $\mathbf{x}_i^L = \mathbf{W}_{opt} \mathbf{x}_i$ and outputs class label $\boldsymbol{t_i}$.

4. Experimental Results and Analysis

4.1 Experimental databases

In this section, we compare the performances of Improved TSLDA, TSLDA, ALDA, NLDA and Fisherface algorithms on five face databases and a handwritten digit database. These face recognition databases contain ORL [30], YALE [31], AR [32], FERET [33] and CMU-PIE [34], and the handwritten digit database is MNIST [35]. The ORL contains 400 images of 40 persons having 10 images per class, which are captured under various postures and expressions with the front face. The dimension of the original space is 10304. The YALE contains 165 images of 15 volunteers having 11 images per class, including changes in illuminations, expressions and postures with the front face. The dimension of the original space is 6400. The AR contains 2600 images from 126 people with 14 images per individual, including different postures and expression in the front face. We randomly choose 15 people with 210 images for experiment. The sample dimension is 4980. The FERET contains 200 classes with 1400 face images having 7 images per class, including different postures and expressions in the front face. We randomly choose 30 people with 210 images for experiment. The sample dimension is 6400. The CMU-PIE contains 68 classes 41368 face images under various postures, illuminations and expressions with multi-angle face. We choose pose09_64x64 dataset in PIE for experiment, which is composed by 24 images per individual and 1632 images in total. The sample dimension is 4096. The MNIST contains 10 classes 6000 handwritten digits with 784 dimension for each digit. We randomly choose 100 samples for each class. The detailed information of the databases is shown in **Table 1**, and some samples in six experimental databases are shown in **Fig.2**. The information of experimental platform includes: CPU： Inter(R) Core(TM) i7-3520M CPU@2.90GHz, RAM: 8GB, operating system: MAC OS X 10.11, software: MATLAB 2014a.

The pre-processing stage is applied to all these images firstly where image size is scaled down to 65*51. Then Improved TSLDA, TSLDA, ALDA, NLDA and Fisherface are respectively conducted to extract sample feature to obtain optimal projection space. Finally, the projected samples are classified using k-nearest neighbor classifier.

Table 1. Information of each database

Database	Image size	No. samples per class	No. Class
ORL	92*112	10	40
YALE	80*80	11	15
AR	40*50	14	15
FERET	80*80	7	30
CMU-PIE	64*64	24	68
MINIST	28*28	100	10

Fig. 2. Some samples in each experimental database

4.2 Parameter Selection

This section analysizes the selection of optimal projection parameter in Improved TSLDA. We give the accuracy of Improved TSLDA with different parameter φ and corresponding optimal projection parameter g in the specific number of training samples on six databases.

First, all samples are cropped and normalized to 65×51 gray images. Then, Improved TSLDA is applied to extract sample feature and calculate optimal projection space $\mathbf{W}_{opt}$. Finally, the projected samples are classified using k-nearest neighbor classifier. The accuracy and optimal projection parameter g of Improved TSLDA on each database are respectively shown in **Table 2** and **Table 3**. In **Table 2,** the highest classification accuracies are depicted in bold fonts, and Num denotes the number of training samples per class. According to **Table 2**, the corresponding optimal projection parameters g on each database are depicted in bold fonts in **Table 3**.

Table 2. Classification accuracy of Improved TSLDA with different φ on databases (%)

Database	φ=0.2a	φ=0.1a	φ=0.05a	φ=1	φ=0
ORL(Num=6)	95.00	98.75	98.12	98.12	93.75
YALE(Num=8)	80.00	95.56	95.56	95.56	93.33
AR(Num=7)	75.24	79.05	80.00	80.00	85.71
FERET(Num=4)	84.90	90.50	94.00	89.50	86.90
CMU-PIE(Num=16)	99.67	99.51	99.51	98.86	96.57
MNIST(Num=70)	73.92	85.98	94.17	96.87	97.50

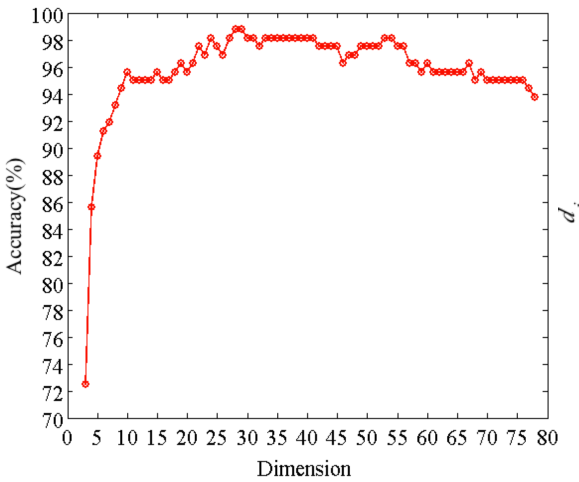
Table 3. Optimal projection parameter g of Improved TSLDA on databases

Database	φ=0.2a	φ=0.1a	φ=0.05a	φ=1	φ=0
ORL(Num=6)	11	25	38	39	78
YALE(Num=8)	8	14	14	14	28
AR(Num=7)	8	11	14	14	28
FERET(Num=4)	9	55	63	193	386
CMU-PIE(Num=16)	108	128	132	67	134
MNIST(Num=70)	3	4	8	9	18

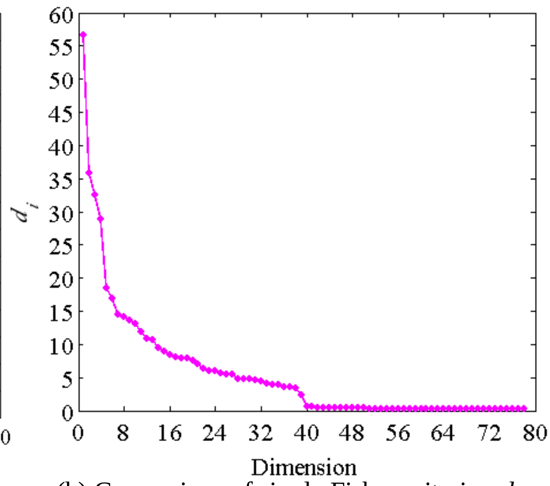
The **Table 2** shows that the highest classification accuracies of Improved TSLDA on ORL, YALE, AR, FERET, CMU-PIE and MNIST are 98.75% (φ=0.1a,Num=6), 95.56% (φ=1,Num=8), 85.71% (φ=0,Num=7), 94.00% (φ=0.05a,Num=4), 99.67% (φ=0.2a,Num=16), and 97.50% (φ=0,Num=70) respectively. For ORL, YALE, FERET databases, the accuracy increases at first and then decreases as parameter $\boldsymbol{\varphi}$ constantly closes to zero. For AR and MVIST databases, the accuracy shows a trend of constantly increasing as parameter $\boldsymbol{\varphi}$ decreasing. For CMU-PIE database, the accuracy shows a decreasing trend as parameter $\boldsymbol{\varphi}$ decreasing. Since d_i in set D' is limited to $d_i \geq \varphi \geq 0$, it is obvious that the optimal projection space $\mathbf{W}_{opt}$ may be a subset of $\mathbf{W}$ and is determined by optimal projection parameter g. The value of g is further explored and shown in **Table 3**.

The **Table 3** shows that the optimal projection parameter g of $\mathbf{W}_{opt}$ is 25, 14, 28, 63, 108, 18 for each database. For ORL and FERET, the corresponding optimal projection parameters are g=25<39<78 and g=63<193<386 respectively. It demonstrates that optimal projection $\mathbf{W}_{opt}$ only retains partial feature space (discriminant information) in $\mathbf{W}_1$ and the remaining feature space is discarded as the noise information. The effective discriminant information exists in subspace $\mathbf{W}_1$. For YALE, optimal projection parameter is g=14=14<28, which means optimal projection $\mathbf{W}_{opt}$ contains all discriminant information in W_1 and the W_2 is discarded as noise information. The effective discriminant information only exists in W_1. For CMU-PIE and MNIST, the corresponding optimal projection parameter are 67<g=108<134 and 10<g=18<20 respectively, that indicates optimal projection $\mathbf{W}_{opt}$ is constituted by all feature space in W_1 and some feature space in W_2. For AR, optimal projection parameter of $\mathbf{W}_{opt}$ is g=28>14=28 that means the discriminant information in both W_1 and W_2 are significant for classification.

The **Fig. 3, 4, 5, 6, 7, 8** further explore optimal projection parameter g, variation trend of d_i and accuracy with different feature dimension for Improved TSLDA on ORL, YALE, AR, FERET, CMU-PIE and MNIST. In **Fig. 3(a), 4(a), 5(a), 6(a), 7(a)** and **8(a)**, the horizontal axis means the i selected feature $W._i$ in W, and the vertical axis means accuracy for different dimension. In **Fig. 3(b), 4(b), 5(b), 6(b), 7(b)** and **8(b)**, the horizontal axis means i-th selected feature vector $W._i$ in W, and the vertical axis means the corresponding single Fisher criterion d_i. According to above figures, we find that: For ORL, YALE, FERET, CMU-PIE and MNIST, the classification accuracy obtained by constantly adding single feature does not show a increasing trend until reaching maximum and then slightly decrease on the whole. For AR, the classification accuracy has increased to the optimal value by adding single feature vector $\mathbf{W}._i$ constantly. Besides, the corresponding single Fisher criterion d_i for all database behaves a rapid decline trend to a stable level, and declines rapidly again to zero.

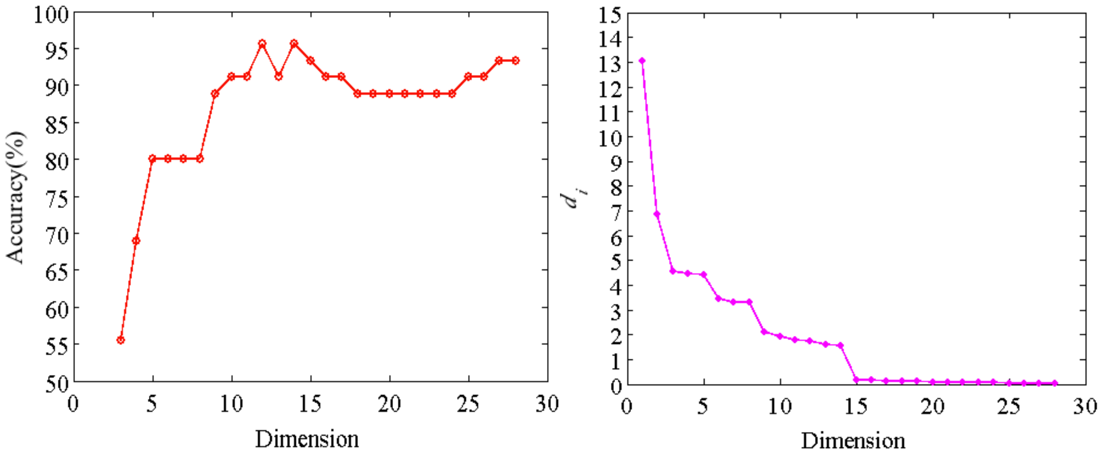


(a) Accuracy comparion



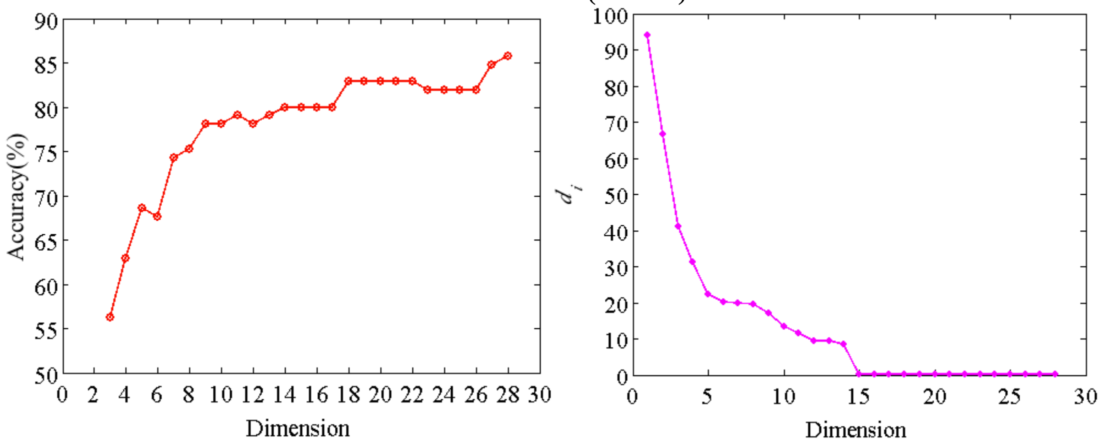
(b) Comparison of single Fisher criterion d_i

Fig. 3. Comparisons of accuracy and d_i with different feature dimension for Improved TSLDA on ORL database (Num=6).



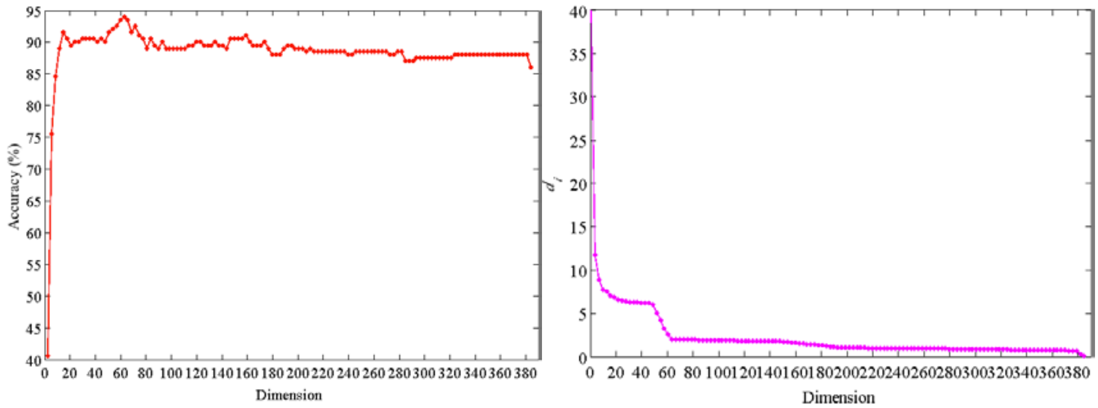
(a) Accuracy comparion (b) Comparison of single Fisher criterion d_i

Fig. 4. Comparisons of accuracy and d_i with different feature dimension for Improved TSLDA on YALE database (Num=8).



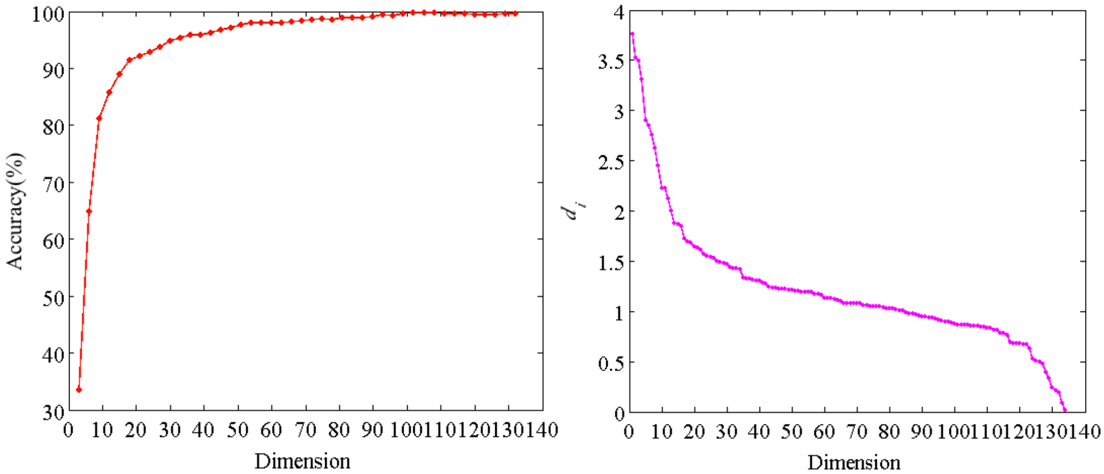
(a) Accuracy comparion (b) Comparison of single Fisher criterion d_i

Fig. 5. Comparisons of accuracy and d_i with different feature dimension for Improved TSLDA on AR database (Num=7).



(a) Accuracy comparion (b) Comparison of single Fisher criterion d_i

Fig. 6. Comparisons of accuracy and d_i with different feature dimension for Improved TSLDA on FERET database (Num=4).



(a) Accuracy comparion (b) Comparison of single Fisher criterion d_i

Fig. 7. Comparisons of accuracy and d_i with different feature dimension for Improved TSLDA on CMU-PIE database (Num=16).

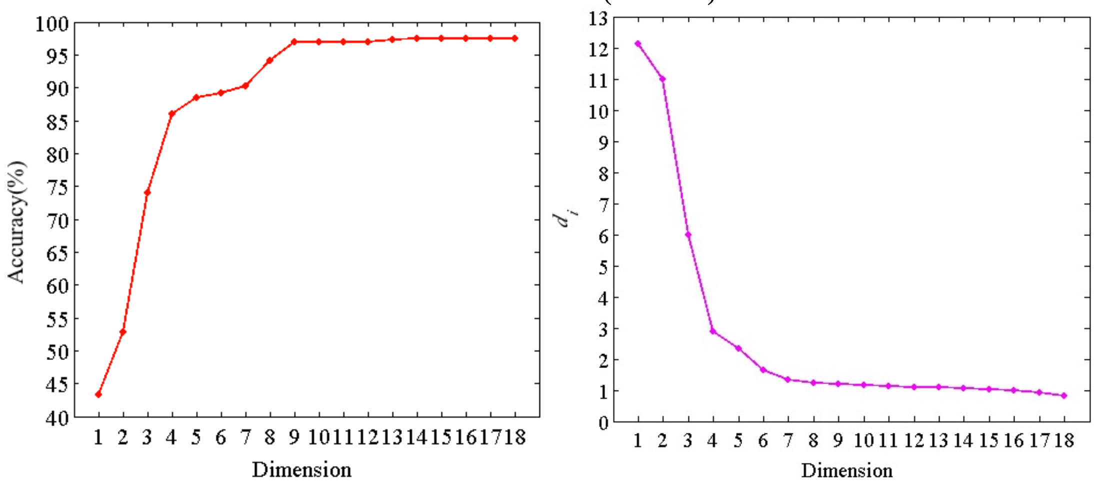


Fig. 8. Comparisons of accuracy and d_i with different feature dimension for Improved TSLDA on MNIST database (Num=70).

According to these trends, we can draw two conclusions that: ①The discriminant information in projection space of TSLDA may not be entirely effectively, and there may be some noise information in it. ② The optimal projection space $\mathbf{W}_{opt}$ is determined by optimal projection parameter g. Improved TSLDA can extract superior feature vectors and eliminate noise feature information in $\mathbf{W}$.

4.3 Algorithm Comparisons

This section compares the Improved TSLDA algorithm with TSLDA, ALDA, NLDA and Fisherface on the six databases, including ORL, YALE, AR, FERET, CMU-PIE and MNIST. We analyze the accuracy and time of each compatitive algorithm.

4.3.1 Accuracy Comparisons

We compare the accuracy of Improved TSLDA with other four algorithm on each database. For each class, we utilize **Num** samples for training and **(n*-Num)** samples for test on the six databases, where **Num** denotes number of training sample per class and **n*** is total sample number per class. The experimental results are shown in **Table 4-9**. In **Table 4-9**, the best results are highlighted in bold face font. Besides, the corresponding changing trend of accuracy are repectively plotted in **Fig. 9 (a)**, **Fig. 9 (b)**, **Fig. 9 (c)**, **Fig. 9(d)**, **Fig. 9(e)** and **Fig. 9(f)**. In the **Fig. 9**, the horizontal axis means the number of trainning samples, and the vertical axis means classification accuracy for each database.

Table 4. Classification accuracy on ORL with different number of training samples (%)

Algorithm	Num=2	Num=3	Num=4	Num=5	Num=6	Num=7	Num=8	AVG
Improved TSLDA	88.44	93.21	94.58	97.00	98.75	99.17	100	95.88
TSLDA	82.81	92.14	93.75	92.50	95.00	95.83	96.25	92.61
ALDA	88.44	91.43	94.58	97.00	98.12	99.17	100	95.53
NLDA	87.81	91.07	92.17	95.50	96.88	99.17	98.75	94.48
Fisherface	83.44	81.43	80.42	69.50	85.62	82.50	93.75	82.38

Table 5. Classification accuracy on YALE with different number of training samples (%)

Algorithm	Num=3	Num=4	Num=5	Num=6	Num=7	Num=8	Num=9	AVG
Improved TSLDA	78.33	83.81	88.89	85.33	93.33	95.56	90.00	87.89
TSLDA	75.00	82.86	83.33	80.00	90.00	88.89	93.33	84.77
ALDA	78.33	83.81	88.89	85.33	93.33	95.56	90.00	87.89
NLDA	80.00	82.86	82.22	80.00	90.00	84.44	93.33	84.69
Fisherface	72.50	80.95	81.11	50.67	76.67	66.67	76.67	72.17

Table 6. Classification accuracy on AR with different number of training samples (%)

Algorithm	Num=4	Num=5	Num=6	Num=7	Num=8	Num=9	AVG
Improved TSLDA	77.33	84.44	86.67	85.71	100	100	89.03
TSLDA	76.67	82.96	88.33	84.76	100.00	100.00	88.79
ALDA	75.33	83.70	82.50	80.00	100.00	100.00	86.92
NLDA	78.00	82.22	87.50	83.81	100.00	100.00	88.59
Fisherface	79.33	80.74	86.67	84.76	98.89	99.17	88.26

Table 7. Classification accuracy on FERET with different number of training samples (%)

Algorithm	Num=2	Num=3	Num=4	Num=5	AVG
Improved TSLDA	76.25	84.44	94.00	100.00	88.67
TSLDA	75.00	73.33	90.00	96.67	83.75
ALDA	76.25	80.00	93.33	100.00	87.40
NLDA	68.88	83.33	93.33	93.33	84.72
Fisherface	61.50	74.44	88.33	100.00	81.07

Table 8. Classification accuracy on CMU-PIE with different number of training samples (%)

Algorithm	Num=4	Num=8	Num=12	Num=16	Num=20	AVG
Improved TSLDA	27.25	67.69	87.50	99.67	100	76.24
TSLDA	50.66	67.56	84.80	96.88	100	79.98
ALDA	24.93	63.97	76.84	95.62	100	72.27
NLDA	33.75	64.34	79.41	97.24	100	74.94
Fisherface	48.68	67.00	82.23	99.38	100	79.45

Table 9. Classification accuracy on MNIST with different number of training samples (%)

Algorithm	Num=30	Num=40	Num=50	Num=60	Num=70	Num=80	AVG
Improved TSLDA	87.15	89.42	91.00	95.33	97.50	97.84	93.04
TSLDA	83.17	88.75	92.17	94.36	95.83	96.25	91.76
ALDA	82.10	84.18	91.00	93.75	96.17	96.33	90.59
NLDA	89.24	90.33	90.50	91.33	93.17	93.75	91.39
Fisherface	69.75	75.33	81.50	85.33	87.50	88.97	81.40

According to **Table 4-9,** we can find that: On ORL, Improved TSLDA has better accuracy than other comparative algorithms with various **Num** and performs the best in seven experiments while ALDA performs the best in five experiments. Improved TSLDA owns the best classification performance while Fisherface is worst. On YALE, both Improved TSLDA and ALDA perform 5 times of the highest accuracy out of all the seven comparative experiments. For classification performance, the accuracy relationship can be described as Improved TSLDA=ALDA>TSLDA≈NLDA>Fisherface. On AR, the Improved TSLDA owns the best accuracy for three experiments while ALDA, NLDA and Fisherface only have one time. The classification performance of Improved TSLDA is still the best. On FERET, the Improved TSLDA still owns the best accuracy for all comparative experiments. For classification performance, we have Improved TSLDA>ALDA>NLDA>TSLDA>Fisherface. On CMU-PIE, since the number of training samples are significantly small, the accuracy of all comparative algorithms are abnormal and useless. As the **Num** increasing, the accuracy of Improved TSLDA has improved rapidly and is still higher than TSLDA, ALDA, NLDA and Fisherface from **Num**≥8. For classification performance, we have Improved TSLDA>TSLDA≈Fisherface>NLDA>ALDA. On MNIST, the Improve TSLDA provides the best accuracy for four experiments while ALDA, NLDA, and Fisherface only have one time respectively. Improved TSLDA also proposes the best classification performance. The average accuracy indicates that Improved TSLDA provides higher accuracy than other comparative algorithms on ORL, YALE, AR, FERET, CMU-PIE and MNIST.

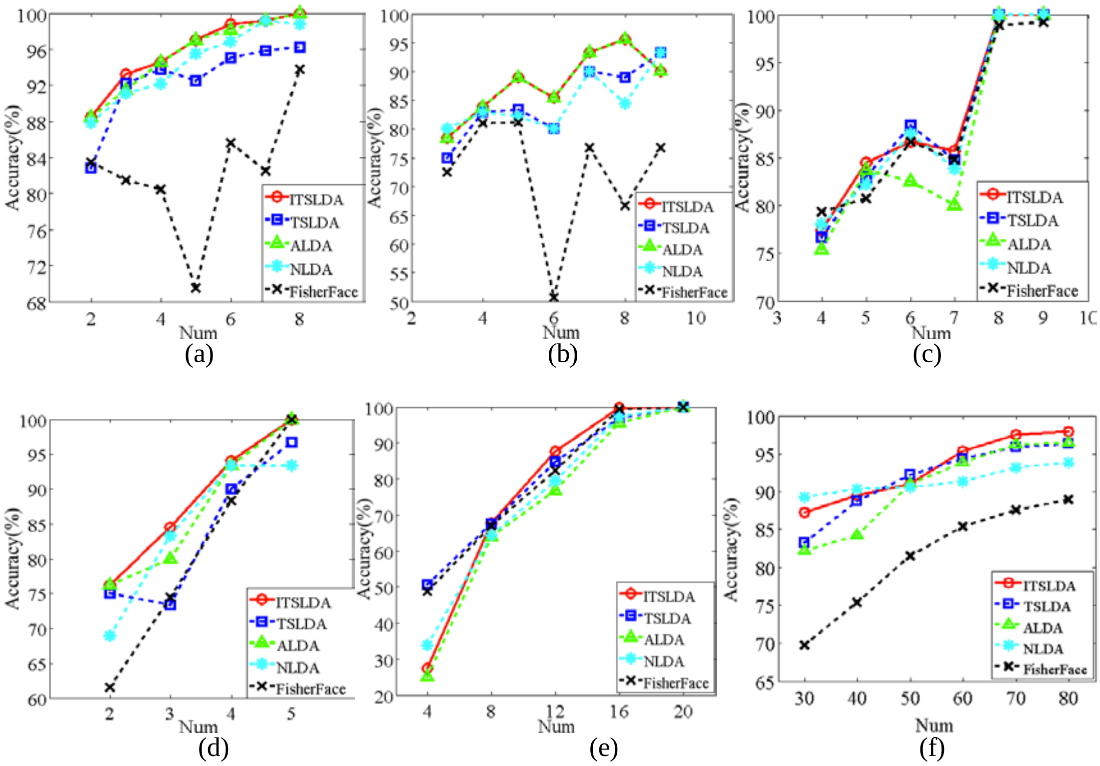


Fig. 9. Classification accuracy with various number of training samples on different databases. (a) Accuracy on ORL. (b) Accuracy on YALE. (c) Accuracy on AR. (d) Accuracy on FERET. (e) Accuracy on CMU-PIE. (f) Accuracy on MNIST.

We can draw three conclusions that: ① For ORL, AR, FERET, and MNIST, Improved TSLDA owns better accuracy than TSLDA, NLDA, ALDA and Fisherface. It is obvious that Improved TSLDA is excellent on classification and its projection space contains more effective discrimination information than other comparative algorithms. ②For YALE, since the significant information only exit in projection $\mathbf{W}_1$, the Improved TSLDA and ALDA have the same projection space and accuracy. This demonstrates that the selection and compression method of Improved TSLDA not only can retain useful feature and discard noise information automatically, but also own the advantage of ALDA. ③ For CMU-PIE, due to small number of training samples, the accuracy of all comparative algorithm are abnormal and useless. As the **Num** increasing, the classification performance Improved TSLDA has improved rapidly and is still superior than TSLDA, ALDA, NLDA and Fisherface.

Based on above analysis, we consider that Improved TSLDA has better classification performance than TSLDA, ALDA, NLDA and Fisherface than TSLDA on the above six databases. Besides, the number of the best performing experiments and total times are decipted in **Fig. 10,** where the overall performance of Improved TSLDA is best in five comparative algorithms.

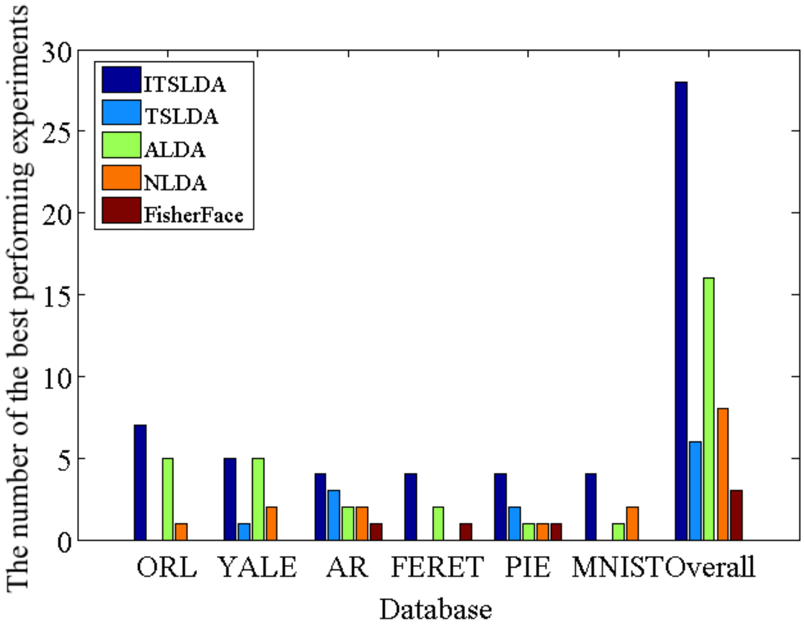


Fig. 10. The number of the best performing experiments on each database

4.3.2 Time Comparisons

The paper analyzes the average time of five comparative algorithms in **Table 10**. It is obvious that the Improved TSLDA, ALDA, NLDA and Fisherface have almost the same training time, and less training time than TSLDA on each database. These results indicate that the theoretical derivation of time complexity in section 3 is accurate. Thus, we consider that the Improved TSLDA using an approximate matrix method has eliminated the singular matrix and reduced the time complexity of the TSLDA effectively.

Table 10. Training time of each comparative algorithm (s)

Database	Improved TSLDA	TSLDA	ALDA	NLDA	Fisherface
ORL(Num=6)	1.28	2.89	0.94	0.93	0.98
YALE(Num=8)	0.60	2.19	0.50	0.55	0.51
AR(Num=7)	0.63	2.05	0.50	0.52	0.47
FERET(Num=4)	4.63	10.33	4.60	5.02	3.90
CMU-PIE(Num=16)	7.63	19.60	6.30	6.89	6.02
MNIST(Num=7)	1.93	3.11	1.87	1.67	1.72
Average	2.78	6.70	2.45	2.60	2.27

4.4 Summary

The Improved TSLDA applies selection and compression method to extract superior feature information from the four subspaces to constitute optimal projection, and uses an approximation matrix method to eliminate the singular matrices. For the experiments with selection of parameter g, we verify the discriminant information in projection space of TSLDA may not be entirely effectively. The selection and compression in Improved TSLDA can extract superior feature in the four subspaces to improve classification performance. For experiments with algorithm comparisons, it concludes that Improved TSLDA not only retains useful feature information and discard noise information in **W** with less time, but also can select optimal feature extraction method automatically. The experimental results show that Improved TSLDA has more excellent classification performance and less time complexity with various number of training samples.

5. Conclusion

In this paper, we have designed an improved method of TSLDA to solve the SSS problem and select effective feature information automatically. First, the Improved TSLDA has introduced a selection and compression method to improve classification performance, where it extracts superior feature vectors and discards noise information automatically from four subspaces. Then, the Improved TSLDA applies an approximate matrix method to eliminate the singular matrix $\mathbf{S}_w$ and $\mathbf{S}_b$, and reduce its time complexity. Theoretical analysis and experimental results indicate that the Improved TSLDA provides better and more stable classification performance and less time complexity. In future work, the paper will focus on: (1) We will explore a more efficient method to select optimal projection parameter in minimal computational complexity. (2) We will have a further study on geometric meaning of the single Fisher criterion and try to introduce the selection and compression method into other feature extraction algorithms.

Acknowledgement

This work has been partly supported by the National Natural Science Foundation of China (61402332, 61502338); Tianjin Municipal Science and Technology Commission (17JCQNJC00400,15JCQNJC00700); The Foundation of Tianjin University of Science and Technology (2017LG10); Supported by the Key Laboratory of food safety intelligent monitoring technology, China Light Industry.

References

[1] Christopher Bishop, "Pattern Recognition and Machine Learning," Springer, Germany, 2008. Article(CrossRef Link)

[2] M Imani and H Ghassemian, "Two Dimensional Linear Discriminant Analyses for Hyperspectral Data," *Photogrammetric Engineering & Remote Sensing*, vol.81, no.10, pp.777-786, 2015. Article (CrossRef Link)

[3] A Sharma and KK Paliwal, "Linear discriminant analysis for the small sample size problem: an overview," *International Journal of Machine Learning & Cybernetics*, vol.6, no.3, pp. 443-454, 2015. Article (CrossRef Link)

[4] P.N. Belhumeur, J.P. Hespanha and D.J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol.19, no.7, pp. 711-720, 1997. Article (CrossRef Link)

[5] JW Lu, KN Plataniotis and AN Venetsanopoulos, "Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition," *Pattern Recognition Letter*, vol.26, no.2, pp. 181-191, 2005. Article (CrossRef Link)

[6] A Sharma and KK Paliwal, "A deterministic approach to regularized linear discriminant analysis," *Neurocomputing*, vol.151, no.11, pp. 207-214, 2015. Article (CrossRef Link)

[7] A Sharma and KK Paliwal, "Approximate LDA Technique for Dimensionality Reduction in the Small Sample Size Case," *Journal of Pattern Recognition Research*, pp.298-306, 2011. Article(CrossRefLink)

[8] S Wang, "Linear Discriminant Analysis Based on Clustering Regularization," *Master's thesis of Tianjin University*, 2013. Article(CrossRef Link)
[9] A Sharma, KK Paliwal, Seiya Imoto and Satoru Miyano, "A feature selection method using improved regularized linear discriminant analysis," *Machine Vision and Applications*, vol.25, no.25, pp. 775-786, 2014. Article (CrossRef Link)
[10] Y Liu and SZ liao, "Granularity selection for cross-validation of SVM," Information Sciences, vol. 378, pp. 475-483,2017 Article(CrossRef Link)
[11] WY Yang and HY Wu, "Regularized complete linear discriminant analysis," *Neurocomputing*, vol. 137, no.11, pp. 185-191, 2014. Article (CrossRef Link)
[12] LF Chen, HYM Liao, MT Ko, JC Lin and GJ Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, vol.33, no.10, pp. 1713-1726, 2000. Article (CrossRef Link)
[13] H Yu and H Yang, "A direct LDA algorithm for high-dimensional data-with application to face recognition," *Pattern Recognition*, vol.34, no.10, pp. 2067-2070, 2001. Article (CrossRef Link)
[14] A Sharma and Kuldip K Paliwal, "A two-stage linear discriminant analysis for face-recognition," *Pattern Recognition Letters*, vol.33, no.9, pp. 1157-1162, 2012. Article (CrossRef Link)
[15] KK Paliwal and A Sharma, "A Improved direct LDA and its application to DNA microarray gene expression data," Pattern Recognition Letters, vol.31, no.16, pp. 2489-2492, 2010. Article(CrossRef Link)
[16] XS Zhuang, DQ Dai, "Inverse Fisher discriminate criteria for small sample size problem and its application to face recognition," *Pattern Recognition*, vol.38, no.11, pp.2192-2194, 2005. Article (CrossRef Link)
[17] Pi-Fuei Hsieh, "Impact and realization of increased class separability on the small sample size problem in hyperspectral classification," *Canadian Journal of Remote Sensing*, vol.25, no.3, pp: 248-261, 2009. Article(CrossRef Link)
[18] Fadi Dornaika and ALireza Bosagzadeh, "On Solving the Small Sample Size Problem for Marginal Fisher Analysis," *Image Analysis and Recognition*, pp.116-123, 2013. Article (CrossRef Link)
[19] JH Zhao, L Shi and J Zhu, "Two-Stage Regularized Linear Discriminant Analysis for 2-D Data," *IEEE Transactions on Neural Networks And Learning Systems*, vol.26, no.8, pp.1669-1681, 2015. Article(CrossRef Link)
[20] MW Jian and KM Lam, "Simultaneous Hallucination and Recognition of Low-Resolution Faces Based on Singular Value Decomposition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol.25, no.11 pp. 1761-1772, 2015. Article (CrossRef Link)
[21] MW Jian, KM Lam and JY Dong. "A Novel Face-Hallucination Scheme Based on Singular Value Decomposition," *Pattern Recognition*, vol. 46, no. 11, pp. 3091-3102, November 2013. Article (CrossRef Link)
[22] MW Jian, KM Lam and JY Dong. "Facial-Feature Detection and Localization Based on a Hierarchical Scheme," Information Sciences, vol. 262, pp. 1-14, 2014. Article (CrossRef Link)
[23] MW Jian and KM Lam, "Face-Image Retrieval Based on Singular Values and Potential-Field Representation," Signal Processing, vol. 100, pp. 9-15, 2014. Article (CrossRef Link)
[24] YD Wen, KP Zhang, ZF Li and Y Qiao, "A Discriminative Feature Learning Approach for Deep Face Recognition," in *Proc. of the 14th European Conference of Computer Vision* (ECCV 2016), Amsterdam, Netherlands, 499-515, October 2016. Article (CrossRef Link)
[25] ZC Li, J Liu, JH Tang and HQ Lu, "Robust Structured Subspace Learning for Data Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.37, no.10, pp.2085-2098. Article (CrossRef Link)
[26] ZC Li, J Liu, Y Yang, XF Zhou and HQ Lu, "Clustering-Guided Sparse Structural Learning for Unsupervised Feature Selection," *IEEE Transactions On Knowledge And Data Engineering*, vol.26, no.9, pp. 2138-2150, 2014. Article(CrossRef Link)
[27] J Gui, ZN Sun, J Cheng and SW Ji, "How to Estimate the Regularization Parameter for Spectral Regression Discriminant Analysis and its Kernel Version?" *IEEE Transactions On Circuits And Systems For Video Technology*, vol.24, no.2, pp.211-223, 2014. Article(CrossRef Link)

[28] Andrew R. Webb, *Statistical Pattern Recognition, 3nd Edition*, Wiley, America, 2011. Article(CrossRef Link)
[29] JP Ye and Q Li, "A two-stage linear discriminant analysis via QR-decomposition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol.27, no.6, pp.929-941, 2005. Article(CrossRef Link)
[30] ORL Database of Faces. Article(CrossRef Link)
[31] Yale Face Database. Article(CrossRef Link)
[32] AR Face Database. Article(CrossRef Link)
[33] FERET Database. Article(CrossRef Link)
[34] CMU-PIE. Article(CrossRef Link)
[35] MNIST Database. Article(CrossRef Link)

Yarui Chen is an associate professor in College of Computer Science and Information Engineering, Tianjin University of Science and Technology, Tianjin, P.R.China. She received her B.S. Degree from Hebei University of Technology, China, M.S. and Ph.D. degrees from Tianjin University, China. She has published 14 papers in international journals and conference, such as Expert Systems With Application, Neural Computing and application, CIKM, and so on. Her research interests include machine learning, neural networks, probabilistic inference, and approximate inference.

Xin Tao received his B.E. and M.S. degrees in Computer Science and Information Engineering Institute from Tianjin University of Science and Technology, Tianjin, P.R. China, in June 2014 and 2017. His current research interests include machine learning algorithms, neural networks and deep learning.

Congcong Xiong is a professor at Tianjin University of Science and Technology, Tianjin, P. R. China. She has published 43 papers in journals and international conference, such as Journal of Tianjin University of Science & Technology, Computer Engineering and Design, International Conference on Fuzzy Systems & Knowledge Discovery and so on. Her current research interests include Computer network, embedded system and parallel computing.

Jucheng Yang is a full professor in College of Computer Science and Information Engineering, Tianjin University of Science and Technology, Tian jin, P.R.China. He is member of IEEE, ACM and CCF. He received his B.S. degree from South-Central University for Nationalities, China, M.S. and Ph.D. degrees from Chonbuk National University, Republic of Korea. He did his postdoctoral work at the Advanced Graduate Education Center of Jeonbuk for Electronics and Information Technology-BK21 (AGECJEIT-BK21) in Chonbuk National University, too. He has published over 80 papers in related international journals and conferences, such as IEEE Trans. on HMS, Expert Systems with Applications, IEEE Systems Journal and so on. He owns seven patents in biometrics. His research interests include image processing, biometrics, pattern recognition, and neural networks.