Optimal subset-division based discrimination and its kernelization for face and palmprint recognition

Xiaoyuan Jing\textsuperscript{a,b,c,*}, Sheng Li\textsuperscript{b}, David Zhang\textsuperscript{d}, Chao Lan\textsuperscript{b}, Jingyu Yang\textsuperscript{e}

\textsuperscript{a}State Key Laboratory of Software Engineering, Wuhan University, Wuhan, China
\textsuperscript{b}College of Automation, Nanjing University of Posts and Telecommunications, China
\textsuperscript{c}State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, China
\textsuperscript{d}Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong, China
\textsuperscript{e}College of Computer Science, Nanjing University of Science and Technology, China

A R T I C L E  I N F O

Article history:
Received 16 August 2010
Received in revised form 18 February 2012
Accepted 1 April 2012
Available online 14 April 2012

Keywords:
Optimal subset-division
Discriminant analysis
Improved stability criterion
Kernel method
Face recognition
Palmprint recognition

A B S T R A C T

Discriminant analysis is effective in extracting discriminative features and reducing dimensionality. In this paper, we propose an optimal subset-division based discrimination (OSDD) approach to enhance the classification performance of discriminant analysis technique. OSDD first divides the sample set into several subsets by using an improved stability criterion and K-means algorithm. We separately calculate the optimal discriminant vectors from each subset. Then we construct the projection transformation by combining the discriminant vectors derived from all subsets. Furthermore, we provide a nonlinear extension of OSDD, that is, the optimal subset-division based kernel discrimination (OSKD) approach. It employs the kernel K-means algorithm to divide the sample set in the kernel space and obtains the nonlinear projection transformation. The proposed approaches are applied to face and palmprint recognition, and are examined using the AR and FERET face databases and the PolyU palmprint database. The experimental results demonstrate that the proposed approaches outperform several related linear and nonlinear discriminant analysis methods.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Discriminant analysis is an important research topic in the field of pattern recognition and computer vision [1], especially in face recognition [2,3] and palmprint recognition [4]. Linear discriminant analysis (LDA) is a popular and widely used supervised discriminant analysis method [5]. LDA calculates the discriminant vectors by maximizing the between-class scatter and minimizing the within-class scatter simultaneously. It is effective in extracting discriminative features and reducing dimensionality. Many methods have been developed to improve the performance of LDA, such as direct LDA [6], enhanced Fisher linear discriminant model (EFM) [7], regularized discriminant analysis [8], uncorrelated optimal discriminant vectors (UODV) [9], 2D-LDA [10], LDA/QR [11], discriminant common vectors (DCV) [12], improved LDA [13], incremental LDA [14], semi-supervised discriminant analysis (SSDA) [15], local uncorrelated discriminant transform (LUDT) [16], tensor rank one discriminant analysis (TR1DA) [17], Perturbation LDA (P-LDA) [18], and subclass discriminant analysis (SDA) [19,20]. SDA approximates the underlying distribution with mixture of Gaussians by dividing a class into suitable subclasses. It reformulates the between-class scatter matrix and employs a stability criterion to determine the number of subclasses. Sugiyama put forward a local Fisher discriminant analysis (LFDA) method [21], which uses local neighborhood information to construct the weighted between-class and within-class scatter matrices, and then performs discriminant analysis.

Unlike LDA, LFDA and SDA that extract discriminative features from the whole dataset, two discriminant methods, i.e., Multiple LDA (MLDA) [22] and Random sampling LDA (RLDA) [23,24], have been recently proposed to perform discriminant analysis by splitting the sample set into several subsets. Multiple LDA (MLDA) randomly splits the original dataset into several subsets, constructs the LDA classifier for each subset and selects an appropriate classifier to do recognition [22]. Random sampling LDA (RLDA) randomly chooses training samples to construct some subsets and randomly chooses principal components (e.g., eigenfaces or eigen-palms) to construct other subsets, then produces multiple weak classifiers and integrates them by using a fusion rule [23,24]. Both RLDA and MLDA employ the random sampling strategies.

In recent years, many kernel discriminant methods have been presented to extract nonlinear discriminative features and enhance the classification performance of linear discrimination techniques, such as kernel discriminant analysis (KDA) [25,26],...
kernel direct discriminant analysis (KDDA) [27], improved kernel Fisher discriminant analysis [28], complete kernel Fisher discriminant (CKFD) [29], kernel discriminant common vectors (KDCV) [30], kernel subclass discriminant analysis (KSDA) [31], kernel local Fisher discriminant analysis (KLFDA) [32], kernel uncorrelated adjacent-class discriminant analysis (KUADA) [33], and mapped virtual samples (MVS) based kernel discriminant framework [34].

1.1. Motivation and contribution

In this paper, we propose an optimal subset-division based discrimination (OSDD) approach. According to the distribution of samples, OSDD divides the sample set into subsets by using the K-means algorithm and attains the discriminant vectors from each subset. We then construct the projection transformation by combining the discriminant vectors derived from each subset. How to find the appropriate number of subsets. Inspired by the stability criterion used in subclass discriminant analysis (SDA), which determines the number of subclasses, we put forward an improved stability criterion to determine the optimal number of subsets. This criterion first computes the between-class scatter matrices and the within-class scatter matrices of all subsets, and then measures the average conflict degree of eigenvectors of between-class and within-class scatter matrices. The less average degree is, the better separability within each subset and the lower correlation between all subsets are. OSDD determines the optimal number of subsets with the least average conflict degree.

The differences between OSDD and two related methods (i.e., RLDA and MLDA) is that (1) OSDD employs the improved stability criterion to get the optimal number of subsets, while RLDA and MLDA empirically select the number of subsets and (2) OSDD divides the sample set according to the distribution of samples by using the K-means algorithm, while RLDA and MLDA use the random sampling strategies.

OSDD is expected to handle non-linearly separable datasets better than many linear approaches, since it divides the whole dataset into several subsets and consequently simplifies datasets distributions. Nevertheless, it is still worthy to explore the non-linear discriminant capability of OSDD, in particular, when some obtained subsets still show complicated and non-linearly separable data distribution. Hence, in this paper, we further extend OSDD in the kernel space and present an optimal subset-division based kernel discrimination (OSKD) approach to extract nonlinear discriminative features. It maps the sample set from the input space to the kernel space, and employs the kernel K-means algorithm to divide the sample set. The improved stability criterion can be used to find the optimal number of subsets in the kernel space. The nonlinear discriminant vectors are calculated from each subset, and combined as a whole projection transformation.

Using the AR and FERET face databases and the PolyU palmprint database, the experimental results show that the proposed approaches outperform several related linear and nonlinear discriminant analysis methods. The discriminative features achieved by our approaches are demonstrated to have favorable local discriminabilities, and also have good generalization abilities for the whole sample set.

1.2. Organization

The rest of this paper is organized as follows: in Section 2, we outline some related work. In Section 3 we describe our OSDD approach and propose an improved stability criterion to determine the optimal number of subsets. We describe the nonlinear extension of OSDD in Section 4. Experimental results on the face and palmprint databases are presented in Section 5. In Section 6, we provide the further experimental analysis, before drawing conclusions in Section 7.

2. Related work

2.1. Linear discriminant analysis and kernel discriminant analysis

Assume that the original sample set \( X = \{x_1, x_2, \ldots, x_N\} \) is composed of \( c \) classes, and there are \( n_i \) training samples in the \( i \)-th class. Linear discriminant analysis (LDA) tries to find a set of optimal projections \( W = [w_1, w_2, \ldots, w_m] \), such that the ratio of the determinant of between-class scatter to the determinant of within-class scatter is maximized, i.e.,

\[
J(W) = \arg \max_W \frac{|W^T S_b W|}{|W^T S_w W|},
\]

(1)

where \( S_b \) and \( S_w \) are the between-class scatter matrix and the within-class scatter matrix, respectively [5]. Generally, \( W \) is constructed by the eigenvectors of \((S_p)^{-1} S_b\). Kernel discriminant analysis (KDA) is based on a conceptual transformation from original input space into a nonlinear high-dimensional kernel space [25,26]. For a given nonlinear mapping function \( \phi \), the input data space \( R^d \) can be mapped into the kernel space \( F \):

\[
\phi : R^d \to F, \quad x \to \phi(x).
\]

(2)

KDA reformulates the scatter matrices \( S_b^\phi \) and \( S_w^\phi \). In the kernel space, the Fisher criterion is defined as

\[
J(W^\phi) = \arg \max_W \frac{|W^T S_b^\phi W^\phi|}{|W^T S_w^\phi W^\phi|},
\]

(3)

\( W^\phi \) is constructed by the eigenvectors of \((S_w^\phi)^{-1} S_b^\phi\).

2.2. Subclass discriminant analysis and its kernelization

Subclass discriminant analysis (SDA) was presented to adapt to a large variety of data distributions [19,20]. It employs a simple criterion that could provide a good estimate for the number of subclasses needed to represent each class. SDA divides classes into the optimal number of subclasses. The extracted discriminant vectors can maximize the distance between the class means and the means of the subclasses that correspond to the same class.

SDA redefines the between-class scatter \( \Sigma_b \) as

\[
\Sigma_b = \sum_{l=1}^{C-1} \sum_{j=1}^{N_l} \sum_{k=1}^{C} \sum_{i=1}^{N_j} p_{ij} p_{kl} (\mu_{ij} - \mu_{ij})(\mu_{kl} - \mu_{ij})^T.
\]

(4)

where \( N_l \) is the number of subclass divisions in class \( l \), \( p_{ij} = n_{ij}/n \), and \( \mu_{ij} \) is the mean of the \( j \)-th subclass in class \( i \).

Kernel subclass discriminant analysis (KSDA) is the kernel version of SDA [31]. KSDA adopts nonlinear clustering technique to find the underlying distributions of datasets in the kernel space. It reformulates Eq. (4) in the kernel space and then performs SDA.

2.3. Random sampling LDA

Random sampling LDA (RLDA) [23,24] constructs two kinds of weak classifiers. First, it randomly selects the samples whose dimensions are reduced by principal component analysis (PCA) [1], and constructs a set of null-space LDA classifiers. Second, it randomly selects a number of principal components (e.g., eigenfaces or eigenpalms) derived by PCA transform to generate random subspaces and constructs multiple LDA classifiers. Finally,
these two kinds of complementary classifiers are integrated using the majority-voting fusion rule.

2.4. Multiple LDA subspaces

Multiple LDA (MLDA) [22] randomly splits the original dataset into multiple clusters and computes a single LDA subspace for each cluster. In MLDA, each cluster contains the same number of classes. For a new unknown test sample, it is projected onto all subspaces, and the class label for this sample can be obtained by searching for the closest class center over all projected spaces.

Table 1 shows the comparison of LDA, SDA, RLDA, MLDA and OSDD in the aspects of subset’s division strategy, the fusion method of subset’s discriminant information, and whether performing the kernel extension.

<table>
<thead>
<tr>
<th>Method</th>
<th>Subset division strategy</th>
<th>Fusion method</th>
<th>Kernel extension</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td>N/A</td>
<td>N/A</td>
<td>Yes</td>
</tr>
<tr>
<td>SDA</td>
<td>N/A*</td>
<td>N/A</td>
<td>Yes</td>
</tr>
<tr>
<td>RLDA</td>
<td>Random division</td>
<td>Classifiers fusion</td>
<td>No</td>
</tr>
<tr>
<td>MLDA</td>
<td>Random division</td>
<td>Classifiers fusion</td>
<td>No</td>
</tr>
<tr>
<td>OSDD</td>
<td>Improved stability criterion and K-means clustering based division</td>
<td>Discriminant features fusion</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* Subclass division using a stability criterion.

3. Optimal subset-division based discrimination

In this section, we first provide the description of OSDD approach. Then we introduce how to determine the optimal number of subsets and divide the sample set into several subsets.

3.1. OSDD approach

Assume that we have divided the original sample set $X = \{x_1, x_2, \ldots, x_N\}$ into $L$ subsets, $X = \{X_1, X_2, \ldots, X_L\}$. The $j$th subset $X_j$ contains $N_j$ samples that belong to $c_j$ classes, $\sum_{j=1}^{L} N_j = N$, and $\sum_{j=1}^{L} c_j = c$, where $c$ is the number of classes in the sample set $X$. Here, $L_o$ is the optimal number of subsets determined by the optimal subsets division strategy described in Section 3.2. We realize OSDD by the following three steps.

3.1.1. Perform PCA for each subset

For high-dimensional data such as images, discriminant analysis methods always suffer from the so-called singularity problems, that is, the scatter matrices are always singular. To overcome the singularity problems, PCA can be used to reduce the dimensionality of the data [5].

For the $j$th subset, we define the total scatter matrix as

$$S_j = \frac{1}{N_j} \sum_{l=1}^{N_j} (x_{jl} - m_j)(x_{jl} - m_j)^T,$$

(5)

where $c_j$, $N_j$, $N$ and $m_j$ are the number of classes, the number of samples of $j$th class, the total number of samples and the mean of all the samples in the $j$th subset, respectively.

The PCA projection transformation $w_{\text{pca}} = [w_{\text{pca}1}, w_{\text{pca}2}, \ldots, w_{\text{pcaN-cj}}]$ is chosen to maximize $J(w_{\text{pca}}) = w_{\text{pca}}^T S_j w_{\text{pca}}$. To ensure the non-singularity of $S_j$ in the following procedures, we select at most $N - c_j$ principal components for each subset. Then we project the samples of the $j$th subset onto the subspace spanned by $w_{\text{pca}}$ and obtain a low-dimensional sample set $z_{\text{pca}}^j = w_{\text{pca}}^T x$.

3.1.2. Calculate discriminant vectors in the PCA transformed spaces

For the $j$th subset, the between-class scatter matrix $S_b^j$ and the within-class scatter matrix $S_w^j$ are constructed in the PCA transformed space as follows:

$$S_b^j = \frac{1}{N_j} \sum_{l=1}^{N_j} (x_{jl} - m_j)(x_{jl} - m_j)^T,$$

(6)

$$S_w^j = \frac{1}{N_j} \sum_{l=1}^{N_j} \sum_{i=1}^{c_j} (x_{jl} - m_i)(x_{jl} - m_i)^T,$$

(7)

where $x_{jl}^j$ is a sample in $Z_{\text{pca}}$, $m_j$ is the mean of the $j$th class in $Z_{\text{pca}}$, and $m_i$ is the global mean of $Z_{\text{pca}}$.

We calculate the discriminant vectors derived from different subsets. For each subset, we obtain a set of discriminant vectors that maximize the following objective function:

$$J(w_{\text{lda}}) = \frac{w_{\text{lda}}^T S_b^j w_{\text{lda}}}{w_{\text{lda}}^T S_w^j w_{\text{lda}}}$$

(8)

It is easy to prove that the optimal solutions $w_{\text{lda}}^j = [w_{\text{lda}1}^j, w_{\text{lda}2}^j, \ldots, w_{\text{lda}(c_j-1)}^j]$ are the eigenvectors of $(S_w^j)^{-1} S_b^j$ corresponding to the $l$ largest eigenvalues. According to Eq. (6), the rank of $S_b^j$ is at most $c_j - 1$. Hence, there are at most $c_j - 1$ discriminant vectors for the $j$th subset. Generally, the discriminant vectors calculated from a subset have a better local discriminabilities than the discriminant vectors extracted from the whole sample set.

Finally, we obtain $c - L_o$ discriminant vectors from all subsets, where $L_o$ is the number of subsets.

3.1.3. Construct the linear projection transformation and do classification

The dimensions of the $c - L_o$ discriminant vectors are the same. We integrate the discriminant vectors derived from all subsets and construct the projection transformation $W$ as

$$W = [w_{\text{lda}1}^j, w_{\text{lda}2}^j, \ldots, w_{\text{lda}(c_j-1)}^j]$$

$$= [w_{\text{lda}}^1, w_{\text{lda}}^2, \ldots, w_{\text{lda}(c-1)}^j] w_{\text{pca}}^T,$$

(9)

After projecting original samples on $W$, we employ the nearest neighbor classifier with cosine distance to classify the projected samples. Fig. 1 shows the flowchart of our optimal subset-division based discrimination (OSDD) approach.

In Section 3.2, we will show that according to the optimal subsets division strategy, each subset does not overlap other subsets, and it holds a low conflict degree with other subsets. Therefore, the discriminant vectors calculated from each subset have good generalization abilities for other subsets. We can combine the discriminant vectors derived from all subsets as a whole projection transformation.

3.2. Optimal subsets division

3.2.1. Improved stability criterion

A stability criterion was proposed to judge whether the linear feature extraction methods are applicable [35]. In the case of LDA, when the smallest angle between the $i$th eigenvector given by the between-class scatter matrix ($S_b$) to be maximized and the first $i$ eigenvectors given by the within-class scatter matrix ($S_w$) to be minimized is close to zero, the results of LDA are not guaranteed.
K between the eigenvectors of conflict value L sample set is divided into k where b is a nonnegative constant, and b approach.

Fig. 1. Flowchart of optimal subset-division based discrimination (OSDD) approach.

to be correct. The stability criterion was formulated as [20]

$$K = \frac{1}{c-1} \sum_{i=1}^{c-1} \max\left(v_i^T u_i\right)^2.$$  \hspace{1cm} (10)

where v_i is the ith eigenvector of S_x and u_i is the jth eigenvector of S_w.

In Eq. (10), the value of K measures the average conflict degree between the eigenvectors of S_x and S_w. In practice, the value of K is desired to be as small as possible. A smaller value of K means a lower average conflict degree, which indicates a better linear separability of the sample set.

However, the stability criterion defined in Eq. (10) just reflects the average conflict degree between eigenvectors of a single dataset. In OSDD, the sample set is split into several subsets. We need to measure the conflict degree of eigenvectors both within the subsets and between the subsets. Therefore, the original stability criterion of SDA is not suitable for OSDD. We define an improved stability criterion described as follows.

Let v_ij denote the ith eigenvector of S_x, u jm and u jm denote the mth eigenvector of S_x and S_w, respectively. Assume that the sample set is divided into L subsets, the within-sets average conflict value K_1(L) is formulated as

$$\bar{K}_1(L) = \frac{1}{L} \sum_{l=1}^{L} \sum_{i=1}^{c-1} \sum_{j=1}^{c-1} K_{ij}^l,$$  \hspace{1cm} (11)

where

$$K_{ij}^l = \begin{cases} \max_{m \leq c_l-1} (u_{jm}^T v_{ij})^2 & \text{if } \max_{m \leq c_l-1} (u_{jm}^T v_{ij})^2 \geq \beta \\ 0 & \text{otherwise} \end{cases}.$$

\(\beta\) is a nonnegative constant, and l_i is the number of \(K_{ij}^l\) whose values are unequal to zero.

And the between-sets average conflict value \(\bar{K}_2(L)\) is formulated as

$$\bar{K}_2(L) = \frac{1}{\sum_{j=1}^{c-1} \sum_{i=1}^{c-1} K_{ij}^0},$$  \hspace{1cm} (12)

where

$$K_{ij}^0 = \begin{cases} \max_{m \leq L < r} (u_{jm}^T v_{ij})^2 & \text{if } \max_{m \leq L < r} (u_{jm}^T v_{ij})^2 \geq \beta \\ 0 & \text{otherwise} \end{cases}.$$

\(\beta\) is a nonnegative constant, and \(l_2\) is the number of \(K_{ij}^0\) whose values are unequal to zero.

Then, we define the improved stability criterion as

$$\tilde{K}(L) = \begin{cases} \tilde{K}_1(L) & \text{if } L = 1 \\ \frac{\tilde{K}_1(L) + \tilde{K}_2(L)}{2} & \text{otherwise} \end{cases}.$$  \hspace{1cm} (13)

where \(\tilde{K}(L)\) indicates the overall average conflict value.

In Eq. (13), \(\tilde{K}_1(L)\) represents the conflict degree within each subset, \(\tilde{K}_2(L)\) indicates the conflict degree between all subsets, and \(\tilde{K}(L)\) represents the overall average conflict degree both within the subsets and between the subsets. In OSDD, the discriminant vectors derived from each subset contribute to the whole sample set. Therefore, the generalization of the discriminant vectors should be considered, and \(\tilde{K}(L)\) may give a proper estimation. In Eqs. (11) and (12), the threshold value \(\beta\) is a small nonnegative constant (in the experiments, \(\beta = 0.05\)). When the conflict value of eigenvectors is smaller than \(\beta\), we consider that the relationship between these eigenvectors is an agreement rather than a conflict. Therefore, it is not appropriate to take this value into account when calculating \(\tilde{K}_1(L)\) or \(\tilde{K}_2(L)\).

3.2.2. Dividing sample set into subsets

To use the improved stability criterion and determine the optimal number of subsets, we need to calculate the values of \(\tilde{K}(L)\) for each of the possible values of L. Then we select the number corresponding to the minimum value.

Therefore, the optimal number of subsets \(L_o\) is given by

$$L_o = \arg\min_{L} \tilde{K}(L).$$  \hspace{1cm} (14)

Then, we use the K-means algorithm to divide the original sample set X into \(L_o\) disjoint subsets \(X = \{X_1, X_2, \ldots, X_{L_o}\}\). The K-means algorithm is widely used for clustering. It can divide the sample set into several disjoint clusters according to its spatial distribution. This is different from other procedures used to divide the dataset like the random sampling procedures employed by Random sampling LDA (RMLDA) and Multiple LDA (MLDA).

In this paper, we use the mean samples of all classes instead of all samples. If the mean vector \(m_j\) is assigned to the 4th subset, all the samples of the jth class are assigned to the 4th subset. That is to say, the subsets division does not change the class attributes of the samples. The procedures of optimal subsets division are summarized in Algorithm 1. Fig. 2 shows the distribution of subsets on three face and palmprint image databases, where the number of subsets in Fig. 2a, b and c are 3, 8 and 7, respectively. It shows that, by using the improved stability criterion and K-means algorithm, the datasets could be divided into several disjoint subsets according to their spatial distributions.

Algorithm 1. Optimal Subsets Division

for \(L=2\) to \(L^*\) do

Divide the sample set into \(L\) subsets using the K-means algorithm and mean samples.

Calculate \(S_{ij}, j = 1, \ldots, L.\)
Compute $S_j V_j = l_j V_j$, $V_j = \frac{v_j}{C_1}$, $v_j \frac{c_j}{C_0}$, $j = 1, \ldots, L$.

Calculate $S_j = 1, \ldots, L$.

Compute $S_j U_j = l_j U_j$, $U_j = \frac{u_j}{C_1}$, $u_j \frac{c_j}{C_0}$, $j = 1, \ldots, L$.

Calculate $K(l)$ using Eq. (13).

end for

$L_0 = \arg\min L k(l)$.

Divide the sample set into $L_0$ subsets and calculate the projection transformation according to Section 3.1. The value of $L$ can be specified for different datasets.

In addition, the proposed OSDD approach is well suited for the large-scale classification problems. By using the optimal subsets division strategy, experimental results show that the number of subsets is small, and each set contains many classes. It’s a rare case that one class is considered as one cluster. Moreover, if only one class is assigned into a subset, this class is well-separated, and it cannot contribute effective discriminant information according to the LDA technique. Since this class is well-separated, it could be easily separated by the discriminant vectors derived from other subsets.

3.2.3. Difference of stability criterions between OSDD and SDA

It’s important to note that the stability criterion used in OSDD is essentially different from that of subclass discriminant analysis (SDA). SDA uses a stability criterion to divide each class into several subclasses and constructs overall scatter matrices, while OSDD splits the sample set into several subsets based on the improved stability criterion, and constructs scatter matrices for each subset. That is, SDA treats the sample set as a single dataset, whereas OSDD treats the sample set as multiple datasets (subsets) and each subset contains lots of classes. Moreover, since the stability criterion of SDA could not measure the conflict degree of multiple datasets, it is not suitable for OSDD.

3.3. Computational complexities

Table 2 shows the computational complexities of LDA, SDA, MLDA, RLDA and the proposed OSDD approach. Ye and Li [11] pointed out that the computational complexity of LDA is $O(N^2d)$, where $N$ is the total number of training samples and $d$ is the dimension of sample. In SDA, $l$ denotes the times of attempting to divide the subclasses. The value of $l$ can be specified by the user or be set so as to guarantee that the minimum number of samples per subclass is sufficiently large. Therefore, SDA solves the eigen-decomposition problem $l$ times, and the computational complexity of SDA is $O(N^2d + lN^3)$. The complexity of LFDA is $O(cN^2d)$, RLDA first costs $O(N^2d + lN^3)$ to perform PCA on the dataset, and then it constructs $h$ LDA or N-LDA (Null space LDA) subspaces that costs $O(hN^2d)$, where $N_i$ is the dimension of the subspace. Hence, the complexity of RLDA is $O(N^2d + hN^2d)$.
original dataset into \( k \) clusters, and computes a single LDA subspace for each cluster that costs \( O(N^2d) \), where \( N \) is the number of samples in one cluster. Meanwhile, MLDA has to make \( s \) attempts to divide the datasets and get the proper number of clusters [22]. Therefore, the total computational complexity of MLDA is \( O(skN^2d) = O((s/k)N^2d) \).

In our OSDD approach, the computational complexity caused by the improved stability criterion is \( \sum_{L=2}^{L^2-2} N_L^2d \), where \( L \) denotes the number of subsets and \( N_L \approx N/L \) is the number of samples in one subset. In practice, the value of \( L \) is fairly small, and it's easy to see that \( \sum_{L=2}^{L^2-2} N_L^2d = \sum_{L=2}^{L^2-2} (N/L)^2d \). Since \( \sum_{L=2}^{L^2-2} 1/L^2 = \pi^2/6 \) [36], we have \( \sum_{L=2}^{L^2-2} (N/L)^2d < ((\pi^2/6)-1)N^2d \approx (3/5)N^2d \). Therefore, the complexity of OSDD is \( O(N^2d) \). It's obvious that OSDD, MLDA and LDA have lower computational complexities than other compared methods.

### 4. Optimal subset-division based kernel discrimination (OSKD)

In this section, the OSDD approach is performed in a high dimension space by using the kernel trick. We realize the OSKD in the following four steps:

#### 4.1. Nonlinear mapping and subsets division in the kernel space

Let \( \phi: \mathbb{R}^d \to \mathbb{F} \) denote a nonlinear mapping. The original sample set \( X \) is injected into \( F \) by \( \phi: x_i \to \phi(x_i) \). We obtain a set of mapped samples \( \mathcal{Y} = \{\phi(x_1), \phi(x_2), \ldots, \phi(x_N)\} \).

The improved stability criterion described in Section 3.2 can also be reformulated to determine the optimal number of subsets in the kernel space. We use the kernel K-means algorithm [37] to divide the mapped samples into \( L \) subsets, \( \mathcal{Y} = \{\psi_1, \psi_2, \ldots, \psi_L\} \).

#### 4.2. Perform KPCA for each subset

For the \( j \)th subset, we perform KPCA by maximizing the following equation:

\[
J(w_{\text{kpca}}^j) = w_{\text{kpca}}^j \mathbf{T} w_{\text{kpca}}^j,
\]

where \( S_j = \sum_{i=1}^{N_j} (\phi(x_i) - m_j^\phi)(\phi(x_i) - m_j^\phi)^T \), \( m_j^\phi \) is the global mean of the \( j \)th subset in the kernel space.

According to the kernel reproducing theory [37], the projection transformation \( w_{\text{kpca}}^j \) in \( F \) can be linearly expressed by using all the mapped samples:

\[
w_{\text{kpca}}^j = \sum_{i=1}^{N} z_i \phi(x_i) = \Psi z,
\]

where \( z = (z_1, z_2, \ldots, z_N)^T \) is a coefficient matrix.

Substituting Eq. (16) into Eq. (15), we have

\[
J(w^j) = z^T (\Psi^T \Psi) z = z^T K^j K^j^T z,
\]

where \( K^j = \Psi^T \Psi \), which indicates an \( N \times N \) non-symmetric kernel matrix whose element is \( K^j_{ij} = \left< \phi(x_i), \phi(x_j) \right> \).

The solution of Eq. (17) is equivalent to the eigenvalue problem:

\[
z^j K^j K^j^T z^j = K^j z^j.
\]

The optimal solutions \( z^j = (z^j_1, z^j_2, \ldots, z^j_{N-L})^T \) are the eigenvectors corresponding to \( N-L \) largest eigenvalues of \( K^j K^j^T \).

### Table 2

<table>
<thead>
<tr>
<th>Method</th>
<th>LDA</th>
<th>SDA</th>
<th>LFDA</th>
<th>RLDA</th>
<th>MLDA</th>
<th>OSDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complexity</td>
<td>( O(N^2d) )</td>
<td>( O(N^2d + (N^2) )</td>
<td>( O(dN^2) )</td>
<td>( O(N^2d + hN^2) )</td>
<td>( O((L/k)N^2d) )</td>
<td>( O(N^2d) )</td>
</tr>
</tbody>
</table>

### Fig. 3. Demo images of one subject from the AR face database.

### Table 3

<table>
<thead>
<tr>
<th>( L )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rates (%)</td>
<td>78.23</td>
<td>85.79</td>
<td>86.75</td>
<td>86.53</td>
<td>85.84</td>
<td>85.50</td>
<td>85.42</td>
<td>85.16</td>
<td>84.63</td>
<td>83.56</td>
</tr>
<tr>
<td>( \bar{k} )</td>
<td>0.1654</td>
<td>0.1527</td>
<td>0.1461</td>
<td>0.1493</td>
<td>0.1577</td>
<td>0.1551</td>
<td>0.1536</td>
<td>0.1562</td>
<td>0.1555</td>
<td>0.1634</td>
</tr>
</tbody>
</table>
project the mapped training sample set $\Psi^d$ on $W^d_{\text{kpca}}$ by
\[ Z^d_{\text{kpca}} = W^d_{\text{kpca}} \Psi^d = \Psi^d W^d_{\text{kpca}}. \]  
(19)

4.3. Calculate kernel discriminant vectors in the KPCA transformed space

By using the KPCA transformed sample set $Z^d_{\text{kpca}}$, we reformulate Eqs. (6) and (7) as
\[ S^d_b = \sum_{i=1}^{c} \sum_{n_i=1}^{n_i} (z^d_{ni} - m^d_i)(z^d_{ni} - m^d_i)^T, \]  
(20)
\[ S^d_w = \frac{1}{N'} \sum_{i=1}^{c} \sum_{n_i=1}^{n_i} (z^d_{ni} - m^d_i)(z^d_{ni} - m^d_i)^T, \]  
(21)
where $z^d_{ni}$, $m^d_i$ and $m^d$ are the sample, the mean of the $i$th class and the mean of all samples in $Z^d_{\text{kpca}}$, respectively.

\[
\text{Fig. 4. Recognition rates of all compared methods on the AR face database.}
\]

\[
\begin{array}{cccccccc}
\end{array}
\]

4.4. Construct the nonlinear projection transformation and do classification

We construct the nonlinear projection transformation $W^d$ as
\[ W^d = \begin{bmatrix} w^d_{11} & w^d_{21} & \cdots & w^d_{c1} \\ w^d_{12} & w^d_{22} & \cdots & w^d_{c2} \\ \vdots & \vdots & \ddots & \vdots \\ w^d_{1L} & w^d_{2L} & \cdots & w^d_{cL} \end{bmatrix}, \]
\[ \Psi = \begin{bmatrix} \Psi^d_{11} \Psi^d_{21} \cdots \Psi^d_{c1} \\ \Psi^d_{12} \Psi^d_{22} \cdots \Psi^d_{c2} \\ \vdots \vdots \ddots \vdots \\ \Psi^d_{1L} \Psi^d_{2L} \cdots \Psi^d_{cL} \end{bmatrix} \]
\[ Y = \begin{bmatrix} y^d_{11} y^d_{21} \cdots y^d_{c1} \\ y^d_{12} y^d_{22} \cdots y^d_{c2} \\ \vdots \vdots \ddots \vdots \\ y^d_{1L} y^d_{2L} \cdots y^d_{cL} \end{bmatrix}, \]
\[ K = Y^T \Psi. \]

Then, the OSKD features can be generated by
\[
Z^d = W^d \Psi = \begin{bmatrix} w^d_{11} \phi^T x_{11} \\ w^d_{12} \phi^T x_{12} \\ \vdots \vdots \\ w^d_{c1} \phi^T x_{c1} \\ w^d_{12} \phi^T x_{12} \\ \vdots \vdots \\ w^d_{c2} \phi^T x_{c2} \\ \vdots \vdots \ddots \vdots \\ w^d_{cL} \phi^T x_{cL} \end{bmatrix}, \]
\[ \Psi = \begin{bmatrix} \Psi^d_{11} \Psi^d_{21} \cdots \Psi^d_{c1} \\ \Psi^d_{12} \Psi^d_{22} \cdots \Psi^d_{c2} \\ \vdots \vdots \ddots \vdots \\ \Psi^d_{1L} \Psi^d_{2L} \cdots \Psi^d_{cL} \end{bmatrix}, \]
\[ Y = \begin{bmatrix} y^d_{11} y^d_{21} \cdots y^d_{c1} \\ y^d_{12} y^d_{22} \cdots y^d_{c2} \\ \vdots \vdots \ddots \vdots \\ y^d_{1L} y^d_{2L} \cdots y^d_{cL} \end{bmatrix}, \]
\[ K = Y^T \Psi. \]

where $K = \Psi^T \Psi$ indicates an $N \times N$ symmetric kernel matrix whose element is $K_{ij} = \langle \phi(x_i), \phi(x_j) \rangle$.

Finally, the nearest neighbor classifier with cosine distance is employed to perform classification.

In nonlinear methods, the construction of kernel matrix costs $O(N^2d)$, and the eigen-decomposition of kernel matrix costs $O(N^3)$. Thus, the computational complexity of KDA, KSDA and KLFDA are $O(N^2d + N^3)$ [38], $O(N^2d + N^3)$ and $O(N^2d + N^3)$, respectively. OSKD employs the improved stability criterion to obtain the optimal number of subsets in the kernel space, which costs $(3/5)N^2d$. And OSKD costs $L - N^2 = L(N/L)^2 = N^2/L^2$ to solve the L eigen-problems for L subsets, where $N^2 = N/L$ is the number of samples in one subset. The computational complexity of OSKD is $O(N^2d + N^3/L^2)$, which is lower than the complexities of KDA, KSDA and KLFDA.

5. Experiments

In this section, we compare the proposed approach with several representative discriminant methods on the public AR and FERET face databases and the PolyU palmprint database. For all compared methods, we use the nearest neighbor classifier with the cosine distance to do classification. The cosine distance
between two vectors \( x \) and \( y \) is defined as follows:

\[
d_{\text{cos}}(x, y) = \frac{x^T y}{\|x\| \cdot \|y\|},
\]

where \( \| \cdot \| \) is the notation of Euclidean norm.

5.1. Experiments using the AR face database

The AR face database \[39\] contains over 4000 color face images of 126 people (70 men and 56 women), including frontal views of faces with different facial expressions, under different lighting conditions and with various occlusions. Most of the pictures were taken in two sessions (separated by two weeks). Each session yielded 13 color images, with 119 individuals (65 men and 54 women) participating in each session. We selected images from 119 individuals for use in our experiment for a total number of 3094 (\( \frac{119 \times 119}{2} \)) samples. All color images are transformed into gray images and each image was scaled to 60 by 60 with 256 gray levels. Fig. 3 illustrates all of the samples of one subject, where (a)–(m) are from Session 1 and (n)–(z) are from Session 2. The details of the images are (a) and (n), neutral expression; (b) and (o), smile; (c) and (p), anger; (d) and (q), scream; (e) and (r), left light on; (f) and (s), right light on; (g) and (t), all sides light on; (h) and (u), wearing sun glasses; (i) and (v), wearing sun glasses and left light on; (j) and (w), wearing sun glasses and right light on; (k) and (x), wearing scarf; (l) and (y) wearing scarf and left light on; (m) and (z), wearing scarf and right light on.

In order to evaluate the impact of different variations to the recognition results, we run each compared methods 30 times. In each time, we randomly choose 6 images of every subject as training samples. The rest is for testing.

By using the improved stability criterion described in Eq. (13) and Eq. (14), we find that the optimal number of subsets (i.e., the value of \( L_0 \)) is 3. Table 3 shows the recognition rates and the overall conflict values (\( \bar{K} \)) varying with the different number of subsets \( L \). It shows that the recognition rates are almost on the opposite trend compared to the overall conflict values.

Fig. 4 shows the recognition rates of 30 random tests of our approaches and other compared methods: (a) OSDD, LDA, SDA, LFDA, MLDA (Multiple LDA) and RLDA (Random sampling LDA); (b) OSKD, KDA, KSDA and KLFDA. To get the best recognition results, RLDA constructs 10 LDA subspaces and 10 N-LDA (Null space LDA) subspaces, and MLDA divides the dataset into 2 clusters. The average recognition results of related methods are shown in Table 4. In this paper, we consider the Gaussian kernel \( k(x, y) = \exp(-\|x-y\|^2/2\delta_i^2) \) for the compared kernel methods, and set the parameter \( \delta_i = \delta \times \delta_i \), \( i \in \{1, \ldots, 20\} \), where \( \delta \) is the standard deviation of training dataset. For each compared kernel method, the parameter \( \delta \) was selected such that the best classification performance was obtained.

Table 4 shows that OSDD and OSKD perform better than other compared linear and nonlinear methods. Compared with LDA, SDA, LFDA, MLDA and RLDA, OSDD improves the average recognition rate at least by 2.78% (\( \frac{84.36\% - 81.58\%}{84.36\%} \)). And the average recognition rate of OSKD is at least 3.07% (\( \frac{85.40\% - 82.33\%}{85.40\%} \)) higher than KDA, KSDA and KLFDA.

5.2. Experiments using the FERET face database

The FERET database \[40\] contains 2200 facial images from 200 individuals with each person contributing 11 images. The images, named with two-character strings ranging from “ba” to “bk”, were numbered as shown in Fig. 3. The random testing results of all compared methods on the FERET face database are shown in Fig. 6. Table 5 shows the recognition rates and the values of \( \bar{K} \) varying with the different number of subsets \( L \) on the FERET face database.

![Fig. 5. Demo images of one subject from the FERET face database.](image)

![Fig. 6. Recognition rates of all compared methods on the FERET face database.](image)

<table>
<thead>
<tr>
<th>( L )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rates (%)</td>
<td>51.57</td>
<td>69.93</td>
<td>77.43</td>
<td>79.86</td>
<td>79.86</td>
<td>80.93</td>
<td>81.50</td>
<td><strong>81.87</strong></td>
<td>81.06</td>
<td>80.94</td>
</tr>
<tr>
<td>( \bar{K} )</td>
<td>0.2477</td>
<td>0.2145</td>
<td>0.2183</td>
<td>0.2231</td>
<td>0.2203</td>
<td>0.2127</td>
<td>0.2074</td>
<td><strong>0.2062</strong></td>
<td>0.2151</td>
<td>0.2193</td>
</tr>
</tbody>
</table>
captured under various illuminations and display a variety of facial expressions and poses. Each image is 384×256 with 256 gray levels. Since many images in this database include the background and the body chest region, we automatically cropped every image sample. That is, we crop the image and preserve the rows from the 40th to the 340th in the original image, producing an image size of 300×256. We scaled the intercepted images to 60×50 to preserve the aspect ratio. Fig. 5 shows all of the samples of one subject. The details of the images are as follows: (a) regular facial status; (b) ±15° pose angle; (c) −15° pose angle; (d) +40° pose angle; (e) −40° pose angle; (f) +25° pose angle; (g) −25° pose angle; (h) alternative expression; (i) different illumination; (j) ±60° pose angle; (k) −60° pose angle. We run each compared method 30 times. In each time, 4 images of each person are randomly selected as training samples. And the remainder 7 images are regarded as testing samples.

The number of subsets in this experiment is set as 8 according to the improved stability criterion. Table 5 shows the recognition rates and the values of $K$ varying with the different number of subsets $L$.

Fig. 6a and b show the recognition rates across 30 runs of OSDD, OSKD and related linear and nonlinear discriminant methods, respectively. To obtain the best recognition performance, RLDA constructs 10 random subspaces, and the optimal number of clusters in MLDA is 10. The average recognition results are listed in Table 6. OSDD boosts the average recognition rate at least by 3.6% (=83.32%−79.72%) than other linear methods, and OSKD boosts the average rate at least by 6.81% (=84.43%−77.62%) than other kernel methods.

5.3. Experiments using the PolyU palmprint database

We use a palmprint database provided by the Hong Kong Polytechnic University (HK PolyU). This database collected palmprint images from 189 individuals. The subjects mainly consisted of volunteers from the students and staff at HK PolyU. The subject was asked to provide about 10 images, each of the left palm and the right palm. Therefore, each person provided 20 images so that the database contains a total of 3780 images from 189 individuals. The size of all the original palmprint images is 384×284 pixels with 75 dpi resolution. Using the processing method in [41], the sub-images with a fixed size (128×128) are extracted from the original images. To reduce the computational cost, each sub-image is compressed to 64×64. We use these sub-images to represent the original palmprint images and to conduct our experiments. Fig. 7 shows all the image samples of one subject. The major changes are in illumination, position including shift and rotation, and texture details. Similar to the kinds of changes encountered in facial expressions, the image may also be slightly affected by the way the hand is posed, shrunk, or stretched.

Table 7
Recognition rates and the values of $K$ varying with the different number of subsets $L$ on the PolyU palmprint database.

<table>
<thead>
<tr>
<th>$L$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rates (%)</td>
<td>89.45</td>
<td>97.59</td>
<td>97.06</td>
<td>97.16</td>
<td>98.64</td>
<td>98.68</td>
<td><strong>99.04</strong></td>
<td>98.97</td>
<td>98.11</td>
<td>98.24</td>
</tr>
<tr>
<td>$\bar{K}$</td>
<td>0.1788</td>
<td>0.1774</td>
<td>0.1657</td>
<td>0.1746</td>
<td>0.1561</td>
<td>0.1579</td>
<td><strong>0.1507</strong></td>
<td>0.1531</td>
<td>0.1609</td>
<td>0.1685</td>
</tr>
</tbody>
</table>

Fig. 7. Demonstration images of one subject from the palmprint database.
We compare the proposed methods with other compared methods 30 times. In each time, we randomly choose 4 images of each person as training samples. The rest is for testing. In this experiment, the optimal number of subsets is 7. Table 7 shows the recognition rates and the values of $K$ varying with the different number of subsets $L$.

Fig. 8a and b separately show the recognition rates of 30 random tests of OSDD, OSKD and their compared methods. RLDA constructs 6 random subspaces, and MLDA divides the dataset into 4 clusters. Our proposed OSDD and OSKD approaches outperform other compared linear and nonlinear methods in all cases. Table 8 shows the average recognition results of OSDD, OSKD and related methods. OSDD improves the average recognition rate at least by 2.71% ($=97.56\% - 94.85\%$) than other linear methods, and OSKD improves the average rate at least by 2.40% ($=98.77\% - 96.37\%$) than other kernel methods.

### 6. Further experimental analysis

In this section, we analyze the performance of our OSDD approach and other compared methods on the aspects of computational complexities, robustness, and stability.

<table>
<thead>
<tr>
<th>Method</th>
<th>Complexity</th>
<th>Average computing time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDA</td>
<td>$O(N^2d)$</td>
<td>9.39</td>
</tr>
<tr>
<td>SDA</td>
<td>$O(N^2d + N^3)$</td>
<td>25.61</td>
</tr>
<tr>
<td>LFDA</td>
<td>$O(cN^2d)$</td>
<td>24.66</td>
</tr>
<tr>
<td>MLDA</td>
<td>$O(s/kN^2d)$</td>
<td>6.86</td>
</tr>
<tr>
<td>RLDA</td>
<td>$O(N^2d + hN^3)$</td>
<td>40.84</td>
</tr>
<tr>
<td>OSDD</td>
<td>$O(N^2d)$</td>
<td>7.31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Complexity</th>
<th>Average computing time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KDA</td>
<td>$O(N^2d + N^3)$</td>
<td>10.97</td>
</tr>
<tr>
<td>KSDA</td>
<td>$O(N^2d + N^3)$</td>
<td>27.05</td>
</tr>
<tr>
<td>KLFDA</td>
<td>$O(cN^2d + N^3)$</td>
<td>21.78</td>
</tr>
<tr>
<td>OSKD</td>
<td>$O(N^2d + (N^3/L^2))$</td>
<td>9.16</td>
</tr>
</tbody>
</table>

Fig. 8. Recognition rates of all compared methods on the PolyU palmprint database.

Fig. 9. Total average computing time (s) of all compared methods on three databases.
6.1. Comparison of computing time

Table 9 shows the computational complexities and average computing time of OSDD and other linear compared methods on three databases, and Table 10 shows those results of OSKD and related nonlinear methods. Fig. 9 shows the total average computing time of all methods on three databases. Table 9 and Fig. 9 show that OSDD and MLDA have lower computational complexities than other methods. And OSDD spends less computing time than other compared methods on the FERET face and PolyU palmprint databases. Table 10 shows that OSKD consumes less computing time than related nonlinear methods on three databases.

6.2. Robustness analysis of the OSDD approach

To give a sense of the robustness of the proposed OSDD approach, we compare the recognition results based on optimal subset-division (i.e., OSDD) and non-optimal subset-division (i.e., NOSDD). OSDD adopts the optimal subset-division by using the improved stability criterion. And NOSDD employs a non-optimal subset-division strategy using random sampling method. Fig. 10 shows the average recognition rates of OSDD and NOSDD (5 different random sampling realizations) across 30 random tests on the AR, FERET face and PolyU palmprint databases. It shows that the optimal subset-division is superior to the non-optimal subset-division, which demonstrates the robustness of our OSDD approach.

6.3. McNemar’s test

To statistically analyze the classification results given in Tables 4, 6 and 8, we conduct a statistical test, McNemar’s test, which gives statistical significance of OSDD and OSKD compared with related methods. In our experiments, the McNemar’s test uses a significance level of 0.05. If the \( p \)-value is below the significance level 0.05, the performance difference between two compared methods is considered to be statistically significant. Table 11 shows the \( p \)-values between OSDD and other compared linear methods, and Table 12 shows the \( p \)-values between OSKD and related nonlinear methods. According to Tables 11 and 12, the following conclusions are reached:

1. The proposed OSDD approach statistically significantly outperforms other compared linear methods, including LDA, SDA, LFDA, MLDA and RLDA, on all the three databases.
2. The proposed OSKD approach is significantly better than related nonlinear methods including KDA, KSDA and KLFDA on the AR and FERET face databases.

### Table 11

<table>
<thead>
<tr>
<th>Method A</th>
<th>Method B</th>
<th>AR face database</th>
<th>FERET face database</th>
<th>PolyU palmprint database</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSDD</td>
<td>LDA</td>
<td>(3.8671 \times 10^{-8a})</td>
<td>(3.5309 \times 10^{-8a})</td>
<td>(8.6825 \times 10^{-8a})</td>
</tr>
<tr>
<td>OSDD</td>
<td>SDA</td>
<td>(3.3145 \times 10^{-4a})</td>
<td>(2.5395 \times 10^{-4a})</td>
<td>(2.0966 \times 10^{-7a})</td>
</tr>
<tr>
<td>OSDD</td>
<td>LFDA</td>
<td>(1.8759 \times 10^{-8a})</td>
<td>(2.8052 \times 10^{-8a})</td>
<td>(0.0280a)</td>
</tr>
<tr>
<td>OSDD</td>
<td>MLDA</td>
<td>(0.0297a)</td>
<td>(2.1881 \times 10^{-8a})</td>
<td>(9.7824 \times 10^{-5a})</td>
</tr>
<tr>
<td>OSDD</td>
<td>RLDA</td>
<td>(0.0417a)</td>
<td>(7.4826 \times 10^{-7a})</td>
<td>(4.1930 \times 10^{-8a})</td>
</tr>
</tbody>
</table>

\(a\) Statistically significant difference between Method A and Method B at a significance level of 0.05.

### Table 12

<table>
<thead>
<tr>
<th>Method A</th>
<th>Method B</th>
<th>AR face database</th>
<th>FERET face database</th>
<th>PolyU palmprint database</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSKD</td>
<td>KDA</td>
<td>(3.7665 \times 10^{-5a})</td>
<td>(7.0114 \times 10^{-8a})</td>
<td>(0.0397a)</td>
</tr>
<tr>
<td>OSKD</td>
<td>KSDA</td>
<td>(0.0356a)</td>
<td>(3.1638 \times 10^{-5a})</td>
<td>(0.1506)</td>
</tr>
<tr>
<td>OSKD</td>
<td>KLFDA</td>
<td>(9.4147 \times 10^{-5a})</td>
<td>(1.1958 \times 10^{-8a})</td>
<td>(1.3940 \times 10^{-8a})</td>
</tr>
</tbody>
</table>

\(a\) Statistically significant difference between Method A and Method B at a significance level of 0.05.
(3) On the PolyU palmprint database, OSDK significantly outperforms KDA and KLFDA. Although the recognition rates of OSDK is better than KSDA (see Table 8), the performance difference between OSDK and KSDA on the PolyU database is not statistically significant.

7. Conclusions

In this paper, we propose an optimal subset-division based discrimination (OSDD) approach and its kernel extension (OSKD) for feature extraction and classification. OSDD first divides the sample set into several subsets by using the optimal subsets division strategy. It then calculates the discriminant vectors from all subsets, and combines them to construct the projection transformation. OSDK acquires nonlinear projection transformation from all subsets in the kernel space. Experimental results show that the AR and FERET face databases and the PolyU palmprint database show that the proposed approaches outperform several related discriminant analysis methods. OSDK improves the average recognition rate at least by 2.71% in contrast with linear discriminant methods including LDA, SDA, LFDA, MLDA and RLDA, and OSDK improves average rate at least by 2.40% in contrast with KDA, KSADA and KLFDA. Experimental results also demonstrate the higher computational efficiency, robustness and statistically significance of our approaches. Hence, the proposed approaches enhance the classification performance of discriminant analysis effectively.

Acknowledgment

The work described in this paper was fully supported by the NSFC under Project nos. 61073113 and 60772059, the New Century Excellent Talents of Education Ministry under Project no. NCET-09-0162, the Doctoral Foundation of Education Ministry under Project no. 20090223110001, the Qing-Lan Engineering Academic Leader of Jiangsu Province, and the Foundation of Jiangsu Province Universities under Project no. 09KJB510011.

References

Xiao-Yuan Jing received his M.Sc. and Ph.D. in Pattern Recognition from Nanjing University of Science and Technology (1995 and 1998, respectively), China. He was a Professor at the Bio-computing Research Center of Shenzhen Graduate School, Harbin Institute of Technology, China. He is now a Professor at the State Key Laboratory of Software Engineering, Wuhan University, and at the College of Automation, Nanjing University of Posts and Telecommunications, China. His research interest includes biometrics, pattern recognition, image processing, neural networks, machine learning and artificial intelligence.

Sheng Li received his B.Sc. degree in Computer Science from the Nanjing University of Posts and Telecommunications in 2010. He is now reading for the M.Sc. degree in Information Security at the Nanjing University of Posts and Telecommunications, China. His research interest includes pattern recognition, biometrics and machine learning.

David Zhang received his M.Sc. and Ph.D. in Computer Science from the Harbin Institute of Technology (1982 and 1985, respectively). In 1994, he received his second Ph.D. in Electrical and Computer Engineering from the University of Waterloo, Canada. Currently he is a professor at the Hong Kong Polytechnic University, where he is the founding director of the Biometrics Technology Center supported by Hong Kong SAR government. His research interest includes biometrics, image processing and pattern recognition.

Chao Lan received his M.Sc. degree in Pattern Recognition from the Nanjing University of Posts and Telecommunications in 2011. He is now reading for the Ph.D. in Computer Science at the University of Kansas, USA. His research interest includes pattern recognition and machine learning.

Jing-Yu Yang is now a Professor at the College of Computer Science, Nanjing University of Science and Technology, China, where he is the founder of Research Institute of Pattern Recognition and Intelligent Systems. His research interest includes pattern recognition, artificial intelligence, biometrics and robotics.