

Article

## Palmprint and Face Multi-Modal Biometric Recognition Based on SDA-GSVD and Its Kernelization

Xiao-Yuan Jing<sup>1,2,3,\*</sup>, Sheng Li<sup>2</sup>, Wen-Qian Li<sup>2</sup>, Yong-Fang Yao<sup>2</sup>, Chao Lan<sup>2</sup>, Jia-Sen Lu<sup>2</sup> and Jing-Yu Yang<sup>4</sup>

<sup>1</sup> State Key Laboratory of Software Engineering, Wuhan University, Wuhan 430072, China

<sup>2</sup> College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210046, China; E-Mails: lisheng1989@gmail.com (S.L.); 705353627@qq.com (W.-Q.L.); yaoyf@njupt.edu.cn (Y.-F.Y.); lanchao18@gmail.com (C.L.); echosenm@gmail.com (J.-S.L.)

<sup>3</sup> State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210046, China

<sup>4</sup> College of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, China; E-Mail: yangjy@mail.njust.edu.cn

\* Author to whom correspondence should be addressed; E-Mail: jingxy\_2000@126.com; Tel./Fax: +86-25-8579-2645.

Received: 1 March 2012; in revised form: 2 April 2012 / Accepted: 25 April 2012 /

Published: 30 April 2012

---

**Abstract:** When extracting discriminative features from multimodal data, current methods rarely concern themselves with the data distribution. In this paper, we present an assumption that is consistent with the viewpoint of discrimination, that is, a person's overall biometric data should be regarded as one class in the input space, and his different biometric data can form different Gaussians distributions, *i.e.*, different subclasses. Hence, we propose a novel multimodal feature extraction and recognition approach based on subclass discriminant analysis (SDA). Specifically, one person's different bio-data are treated as different subclasses of one class, and a transformed space is calculated, where the difference among subclasses belonging to different persons is maximized, and the difference within each subclass is minimized. Then, the obtained multimodal features are used for classification. Two solutions are presented to overcome the singularity problem encountered in calculation, which are using PCA preprocessing, and employing the generalized singular value decomposition (GSVD) technique, respectively. Further, we provide nonlinear extensions of SDA based multimodal feature extraction, that is, the feature fusion based on KPCA-SDA and KSDA-GSVD. In KPCA-SDA, we first apply Kernel PCA on each single modal before performing SDA. While in KSDA-GSVD, we

directly perform Kernel SDA to fuse multimodal data by applying GSVD to avoid the singular problem. For simplicity two typical types of biometric data are considered in this paper, *i.e.*, palmprint data and face data. Compared with several representative multimodal biometrics recognition methods, experimental results show that our approaches outperform related multimodal recognition methods and KSDA-GSVD achieves the best recognition performance.

**Keywords:** multimodal biometric feature extraction; palmprint and face; subclass discriminant analysis (SDA); generalized singular value decomposition (GSVD); kernel subclass discriminant analysis (KSDA)

---

## 1. Introduction

Multimodal biometric recognition techniques use multi-source features together in order to obtain integrated information to obtain more essential data about the same object. This is an active research direction in the biometric community, for it could overcome many problems that bother traditional single-modal biometric system, such as the instability in one's feature extraction, noisy sensor data, restricted degree of freedom, and unacceptable error rates. Information fusion is usually conducted on three levels, *i.e.*, pixel level [1,2], feature level [3–5] and decision level [6–9]. The former two levels mainly aim at learning descriptive features, while the last level aims at finding a more effective way to use learned features for decision making. Especially, at the pixel level and feature level, discriminant analysis technique always plays an important role to acquire more descriptive or more discriminative features.

Linear discriminant analysis (LDA) is a popular and widely used supervised discriminant analysis method [10]. 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 enhanced Fisher linear discriminant model (EFM) [11], improved LDA [12], uncorrelated optimal discriminant vectors (UODV) [13], discriminant common vectors (DCV) [14], incremental LDA [15], semi-supervised discriminant analysis (SSDA) [16], local Fisher discriminant analysis [17], Fisher discrimination dictionary learning [18], and discriminant subclass-center manifold preserving projection [19].

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) [20,21], kernel direct discriminant analysis (KDDA) [22], improved kernel Fisher discriminant analysis [23], complete kernel Fisher discriminant (CKFD) [24], kernel discriminant common vectors (KDCV) [25], kernel subclass discriminant analysis (KSDA) [26], kernel local Fisher discriminant analysis (KLFDA) [27], kernel uncorrelated adjacent-class discriminant analysis (KUADA) [28], and mapped virtual samples (MVS) based kernel discriminant framework [29].

In this paper, we have developed a novel multimodal feature extraction and recognition approach based on linear and nonlinear discriminant analysis technique. We adopt the feature fusion strategy, as

features play a critical role in multimodal biometric recognition. More specifically, we try to answer the question of how to effectively obtain discriminative features from multimodal biometric data. Some related works have appeared in the literature. In [1,2], multimodal data vectors are firstly stacked into a higher dimensional vector to form a new sample set, from which discriminative features are extracted for classification. Yang [3] discussed the feature fusion strategy, that is, parallel strategy and serial strategy. The former uses complex vectors to fuse multimodal features, *i.e.*, one modal feature is represented as the real part, and the other modal feature is represented as the imaginary part; while the latter stacks features of two modals into one feature, which is used for classification. Sun [4] proposed a method to learn features from data of two modalities based on CCA, but it has not been utilized in biometric recognition, and is not convenient to learn features from more than two modes of data.

While current methods generally extract discriminative features from multimodal data technically, they have rarely considered the data distribution. In this paper, we present an assumption that is consistent with the viewpoint of discrimination, that is, in the same feature space, one person's different biometric identifier data can form different Gaussians, and thus his overall biometric data can be described using mixture-Gaussian models. Although LDA has been widely used in biometrics to extract discriminative features, it has the limits that it can only handle the data of one person that forms a single Gaussian distribution. However, as we pointed out above, in multimodal analysis, different biometric identifier data of one person can form mixture-Gaussians. Fortunately, subclass discriminant analysis (SDA) [30] has been proposed to remove such a limit of LDA, and therefore could be used to describe multimodal data that lie in the same input space.

Based on the analysis above, in this paper we propose a novel multimodal biometric data feature extraction scheme based on subclass discriminant analysis (SDA) [20]. For simplicity, we consider two typical types of biometric data, that is, face data and palmprint data. For one person, his face data and palmprint data are regarded as two subclasses of one class, and discriminative features are extracted by seeking an embedded space, where the difference among subclasses belonging to different persons is maximized, and the difference within each subclass is minimized. Then, since the parallel fusion strategy is not suitable to fuse features from multiple modals, we fuse the obtained features by adopting the serial fusion strategy and use them for classification.

Two solutions are presented to solve the small sample size problem encountered in calculating the optimal transform. One is to initially do PCA preprocessing, and the other is to employ the generalized singular value decomposition (GSVD) [31,32] technique. Moreover, it is still worthy to explore the non-linear discriminant capability of SDA in multimodal feature fusion, in particular, when some single-modals still show complicated and non-linearly separable data distribution. Hence, in this paper, we further extend SDA feature fusion approach in the kernel space and present two solutions to solve the small sample size problem, which are KPCA-SDA and KSDA-GSVD. In KPCA-SDA, we first use KPCA to transform each single modal input space  $R^n$  into an  $m$ -dimensional space, where  $m = \text{rank}(K)$ ,  $K$  is the centralized Gram matrix. Then SDA is used to fuse the two transformed features and extract discriminative features. In KSDA-GSVD, we directly perform Kernel SDA to fuse multimodal data by applying GSVD to avoid the singular problem.

We evaluate the proposed approaches on two face databases (AR and FRGC), and the PolyU palmprint database, and compare the results with related methods that also tend to extract descriptive features from multimodal data. Experimental results show that our approaches achieve higher

recognition rates than compared methods, and also get better verification performance than compared methods. It is worthwhile to point out that, although the proposed approaches are validated on data of two modalities, it could be easily extended to multimodal biometric data recognition.

The rest of this paper is organized as follows: Section 2 describes the related work. Section 3 presents our approach. In Section 4, we present the kernelization of our approach. Experiments and results are given in Section 5 and conclusions are drawn in Section 6.

## 2. Related Work

In this section, we first briefly introduce some typical multimodal biometrics fusion techniques such as pixel level fusion [1,2], Yang's serial and parallel feature level fusion methods [3]. Further, three related methods, which are SDA, KSDA and KPCA, are also briefly reviewed.

### 2.1. Multimodal Fusion Scheme at the Pixel Level

The general idea of pixel level fusion [1,2] is to fuse the input data from multi-modalities in as early as the pixel level, which may lead to less information loss. The pixel level fusion scheme fuses the original input face data vector and palmprint data vector of one person, and then the discriminant features are extracted from the fused dataset. For simplicity and fair comparison, we testified the effectiveness of such scheme by extracting LDA features from the fused set in this paper.

### 2.2. Serial Fusion Strategy and Parallel Fusion Strategy

In [3], Yang *et al.* the authors discussed two strategies to fuse features of two data modes. One is called serial strategy and the other is called parallel strategy. Let  $x_i$ ,  $y_i$  denote the face feature vector and palmprint feature vector of the  $i^{th}$  person, respectively. The serial fusion strategy obtains the fused features by stacking two vectors into one higher dimensional vector  $\alpha_i$ , *i.e.*:

$$\alpha_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \quad (1)$$

On the other hand, the parallel fusion strategy combines the features into a complex vector  $\beta_i$ , *i.e.*,

$$\beta_i = x_i + i \cdot y_i \quad (2)$$

Yang *et al.* also pointed out that the fused feature set  $\{\alpha_i\}$  and  $\{\beta_i\}$  can either be used directly for classification, which is called feature combination, or can be input into a feature extractor to further extract more descriptive features with less redundant information, which is called feature fusion.

### 2.3. Subclass Discriminant Analysis (SDA) and Its Kernelization

Subclass discriminant analysis (SDA) [30] is an extension of LDA, which aims at processing data of one class that form mixture Gaussian distribution. It divides each class into a number of subclasses, and calculates a transform space where the distances between both class means and subclass means are maximized, and distances between samples of each subclass is minimized. SDA redefines the between-class scatter  $\Sigma_B$ , within-class scatter  $\Sigma_W$  as:

$$\Sigma_B = \sum_{i=1}^{C-1} \sum_{j=1}^{H_i} \sum_{k=i+1}^C \sum_{l=1}^{H_k} p_{ij} p_{kl} (\mu_{ij} - \mu_{kl})(\mu_{ij} - \mu_{kl})^T \quad (3)$$

$$\Sigma_W = \frac{1}{m} \sum_{i=1}^c \sum_{j=1}^H \sum_{k=1}^{n_{ij}} (x_{ij}^k - \mu_{ij})(x_{ij}^k - \mu_{ij})^T \quad (4)$$

where  $H_i$  is the number of subclasses of class  $i$ ,  $p_{ij} = n_{ij}/n$  is the prior of the  $j^{\text{th}}$  subclass of class  $i$ ,  $\mu_{ij}$  is the mean of the  $j^{\text{th}}$  subclass of class  $i$ . The advantage of this new definition of between class scatter is that it emphasizes the role of class separability over that of intra-subclass scatter. The optimal solution of SDA is the eigenvectors of matrix  $(\Sigma_W)^{-1}\Sigma_B$  associated with the largest eigenvalues.

Kernel subclass discriminant analysis (KSDA) is the nonlinear extension of SDA based on kernel functions [26]. The main idea of the kernel method is that without knowing the nonlinear feature mapping explicitly, we can work on the feature space through kernel functions. It first maps the input data  $x$  into a feature space  $F$  by using a nonlinear mapping  $\phi$ . KSDA adopts nonlinear clustering technique to find the underlying distributions of datasets in the kernel space. The between-class scatter matrix  $S_{KSDA}^{(b)}$  and within-class scatter matrix  $S_{KSDA}^{(w)}$  of KSDA are defined as:

$$S_{KSDA}^{(b)} = \sum_{i=1}^{C-1} \sum_{j=1}^{H_i} \sum_{k=i+1}^C \sum_{l=1}^{H_k} p_{ij} p_{kl} (\bar{\phi}_{ij} - \bar{\phi}_{kl})(\bar{\phi}_{ij} - \bar{\phi}_{kl})^T \quad (5)$$

$$S_{KSDA}^{(w)} = \frac{1}{m} \sum_{i=1}^c \sum_{j=1}^H \sum_{k=1}^{n_{ij}} (\phi_{ij}^k - \bar{\phi}_{ij})(\phi_{ij}^k - \bar{\phi}_{ij})^T \quad (6)$$

where  $\bar{\phi}_{ij}$  indicates the mean vector of  $j^{\text{th}}$  subclass of  $i^{\text{th}}$  class,  $\bar{\phi}$  is the global mean. Like SDA, KSDA tries to maximize the ratio  $|V^T S_{KSDA}^{(b)} V| / |V^T S_{KSDA}^{(w)} V|$  to find a transformation matrix  $V$ . The columns of  $V$  are the eigenvectors corresponding to the largest eigenvalues of  $(S_{KSDA}^{(w)})^{-1} S_{KSDA}^{(b)}$ .

#### 2.4. Kernel Principle Component Analysis

In kernel PCA [33], the input data  $x$  is mapped into a feature space  $F$  via a nonlinear mapping  $\phi$  and then perform a linear PCA in  $F$ . To be specific, we centralize the mapped data as  $\sum_{i=1}^M \phi(x_i) = 0$  firstly, where  $M$  is the number of input data. Then the covariance matrix of the mapped data  $\phi(x_i)$  is defined as follows:

$$C = 1/M \sum_{i=1}^M \phi(x_i) \cdot \phi(x_i)^T \quad (7)$$

Like PCA, the eigenvalue equation  $\lambda V = CV$  must be solved for eigenvalue  $\lambda \geq 0$  and eigenvector  $V \in F \setminus \{0\}$ . We can prove that all the solutions  $V$  lie in the space spanned by  $\phi(x_1), \dots, \phi(x_M)$ . Therefore, we may consider the equivalent system:

$$\lambda(\phi(x_k) \cdot V) = (\phi(x_k), CV) \text{ for all } k=1, \dots, M \quad (8)$$

and  $V$  can be represented as the linear combination of the mapped data  $\phi(x_i)$ : coefficients  $\alpha_1, \dots, \alpha_M$  such that:

$$V = \sum_{i=1}^M \alpha_i \phi(x_i) \quad (9)$$

where  $\alpha_1, \dots, \alpha_M$  denotes the coefficients. Substituting Equations (8) and (9) into Equations (7), and defining an  $M \times M$  matrix  $K$  by:

$$K_{ij} = (\phi(x_i)\phi(x_j)) \quad (10)$$

we arrive at:

$$\ell\lambda K\alpha = K^2\alpha \quad (11)$$

where  $\alpha$  denotes the column vector with entries  $\alpha_1, \dots, \alpha_M$ ,  $K$  is defined as the kernel matrix. To find solutions of Equation (11) we can solve the equivalent eigenvalue problem as follows:

$$\ell\lambda\alpha = K\alpha \quad (12)$$

for nonzero eigenvalues and obtain the optimal  $\alpha$ . Finally, we can project mapped  $\phi(x_i)$  onto  $V$  by using  $\alpha$  to get the KPCA-transformed features [33].

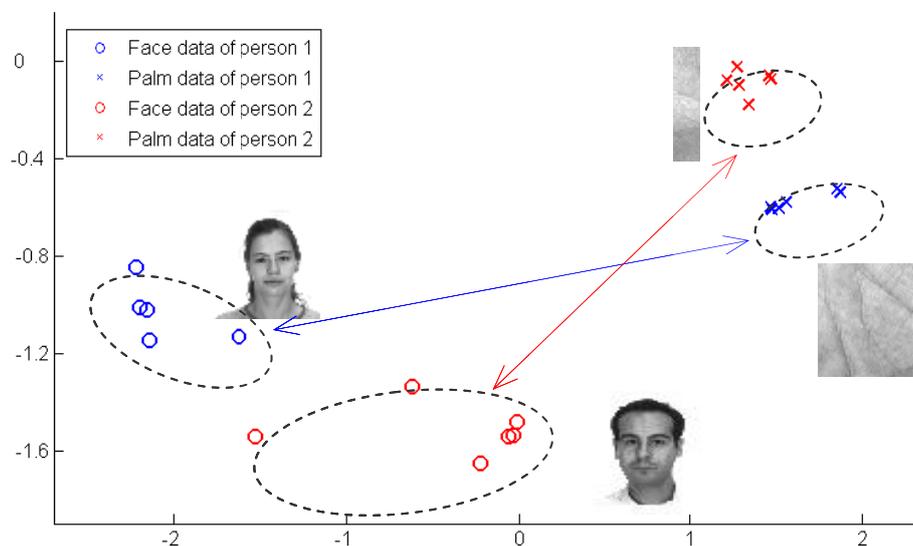
### 3. Subclass Discriminant Analysis (SDA) Based Multimodal Biometric Feature Extraction

In this section, we propose a novel multimodal biometric feature extraction scheme based on SDA. Two solutions are separately introduced to avoid the singular problem in SDA, which are PCA and GSVD. Then we present the algorithm procedures of the proposed SDA-PCA and SDA-GSVD approaches.

#### 3.1. Problem Formulation

For simplicity, we take two typical types of biometric data as examples in this paper. One is the face data, and the other is the palmprint data. From the viewpoint of discrimination, it is quite natural to assume that the overall biometric data one person may be regarded as one class. Moreover, his palmprint and face data can be regarded as two subclasses of this class in the same feature space. An example of two person's face and palmprint samples is shown in Figure 1.

**Figure 1.** Illustration of mix-Gaussian distribution of face data and the corresponding palmprint data. In this example, data of two persons are presented. Each contains 12 data, including six faces and s palmprints. We perform PCA on original data for demonstration, and the order of data magnitude is  $1e4$ .



As can be seen from Figure 1, identifier samples of one person show typical mix-Gaussian distribution, *i.e.*, the face data cluster together and form a Gaussian, while the palmprint data form another Gaussian. If we apply traditional LDA, which enforces both of face and palmprint data of one person to cluster together, then data of two persons would be very likely overlap in the embedded space. It is apparent that, in Figure 1, SDA is a better descriptor of such a data distribution.

Let  $x_{i1}^k$  and  $x_{i2}^k$  be the  $k^{\text{th}}$  face sample and palmprint sample of person  $i$ , respectively;  $n_c$  represent the sample number of each subclass. Then we construct the between-subclass scatter matrix  $S_B$  and within-subclass scatter matrix  $S_W$  as follows:

$$S_B = \sum_{i=1}^{c-1} \sum_{j=1}^2 \sum_{k=i+1}^c \sum_{l=1}^2 p_{ij} p_{kl} (\mu_{ij} - \mu_{kl})(\mu_{ij} - \mu_{kl})^T \quad (13a)$$

$$S_W = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^2 \sum_{k=1}^{n_c} (x_{ij}^k - \mu_{ij})(x_{ij}^k - \mu_{ij})^T \quad (13b)$$

where  $N = c \times n_c$ ,  $p_{ij} = p_{kl} = n_c/N$ ,  $\mu_{ij} = \sum_{k=1}^{n_c} x_{ij}^k/n_c$ .

Let  $\hat{w}$  be the optimal transform vector to be calculated, and then it can be obtained by:

$$\max_w \frac{w^T S_B w}{w^T S_W w} \quad (14)$$

The within-class matrix  $S_W$  is usually singular, and the solution cannot be calculated directly. We present two solutions below to solve this problem, *i.e.*, SDA-PCA and SDA-GSVD.

### 3.2. SDA-PCA

The first solution is to first apply PCA to project each image  $x_{ij}^k$  into a lower dimensional space, and then apply SDA to do feature extraction. By employing the Lagrange multipliers method to solve the optimization problem (15), we could obtain the optimal solution  $W_{SDA}$ , *i.e.*, the eigenvectors of matrix  $(S_W)^{-1} S_B$  associated with the largest eigenvalues.

Based on Formula (14), the rank of  $S_W$  is  $n - 2c$ , where  $n$  represents the total number of training samples (including face and palmprint images), and  $c$  represents the number of persons. Therefore, we can project original samples into a subspace whose dimension is no more than  $n - 2c$ , and then apply SDA to extract features.

Let  $W_{PCA}^1$ ,  $W_{PCA}^2$  separately denote the initial PCA transformations of the sample set of each modal, and  $W_{SDA}$  denote the later SDA transform. Then the final transformations for each modal are expressed as:

$$\hat{W}_1 = W_{PCA}^1 W_{SDA} \quad (15)$$

$$\hat{W}_2 = W_{PCA}^2 W_{SDA} \quad (16)$$

After the optimal transformations  $\hat{W}_1$  and  $\hat{W}_2$  are obtained, we project the face sample  $x_{i1}^k$  and palmprint sample  $x_{i2}^k$  on them:

$$y_{i1}^k = \hat{W}_1^T x_{i1}^k, y_{i2}^k = \hat{W}_2^T x_{i2}^k \quad (17)$$

Then, features derived from face and palmprint are fused used using serial fusion strategy and used for classification:

$$y_i^k = \begin{bmatrix} y_{i1}^k \\ y_{i2}^k \end{bmatrix} \quad (18)$$

### 3.3. SDA-GSVD

While PCA is a popular way to overcome the singular problem and accelerate computation, it may cause information loss. Therefore, we present a second way to overcome the singularity problem by employing GSVD. First, we rewrite the between-class scatter matrix and within-class scatter matrix as follows:

$$S_B = H_b H_b^T, S_w = H_w H_w^T \quad (19)$$

$H_b$  is obtained by transforming formula (13) as follows:

$$\begin{aligned} S_B &= \sum_{i=1}^{c-1} \sum_{j=1}^2 \sum_{k=i+1}^c \sum_{l=1}^2 p_{ij} p_{kl} (\mu_{ij} - \mu_{kl})(\mu_{ij} - \mu_{kl})^T \\ &= p_{ij} p_{kl} \sum_{i=1}^{c-1} \sum_{j=1}^2 [2(c-i)\mu_{ij} - \sum_{k=i+1}^c \sum_{l=1}^2 \mu_{kl}] \cdot \\ &\quad [2(c-i)\mu_{ij} - \sum_{k=i+1}^c \sum_{l=1}^2 \mu_{kl}]^T \end{aligned} \quad (20)$$

Compared with Equation (21),  $H_b$  is defined as:

$$H_b = [H_{(c-1)1}, H_{(c-1)2}, H_{(c-2)1}, \dots, H_{11}, H_{12}] \quad (21)$$

where  $H_{(c-m)n} = 2(c-N)\mu_{mn} - \sum_{k=m+1}^c \sum_{l=1}^2 \mu_{kl}$ .

According to Equation (14), we can easily achieve  $H_w$ :

$$H_w = [x_{ij}^1 - \mu_{ij}, x_{ij}^2 - \mu_{ij}, \dots, x_{ij}^c - \mu_{ij}]_{i=1, \dots, c, j=1, 2} \quad (22)$$

Then, we employ GSVD [31,32] to calculate the optimal transform, and the procedures are given in Algorithm 1.

#### Algorithm 1. Procedures of GSVD based LDA.

- 
- Step 1:** Define matrix  $K = [H_b, H_w]^T$ , and compute the complete orthogonal decomposition  $P^T K Q = \begin{Bmatrix} R & 0 \\ 0 & 0 \end{Bmatrix}$ .
- Step 2:** Compute  $G$  by performing SVD on matrix  $P(1:c, 1:t)$ , i.e.,  $U^T P(1:c, 1:t)G = \Sigma_A$ , where  $t$  is the rank of  $K$ .
- Step 3:** Compute matrix  $M = Q \begin{Bmatrix} R^{-1}G & 0 \\ 0 & I \end{Bmatrix}$ . Put the first  $c-1$  columns of  $M$  into matrix  $W$ . Then,  $W$  is the optimal transform matrix.
- 

Then, face data  $x_{i1}^k$  and palmprint data  $x_{i2}^k$  are separately projected on  $W$  and fused using serial fusion strategy:

$$y_i^k = \begin{bmatrix} y_{i1}^k \\ y_{i2}^k \end{bmatrix} = \begin{bmatrix} \hat{W}^T x_{i1}^k \\ \hat{W}^T x_{i2}^k \end{bmatrix} \quad (23)$$

$y_i^k$  is then used for classification.

### 3.4. Algorithmic Procedures

In this section, we summarize the complete algorithmic procedures of the proposed approach. In practice, if the dimension of two biometric data  $x_{i1}^k$  and  $x_{i2}^k$  are not equal, we could simply pad the lower-dimensional vector with zeros until its dimension is equal to the other one before fusing them using SDA. In case of SDA-PCA, after PCA projection, it is easy guarantee that  $x_{i1}^k$  and  $x_{i2}^k$  have the same dimension if we select the same number of principal components for them.

**Figure 2.** The complete procedures of SDA based multimodal feature extraction.

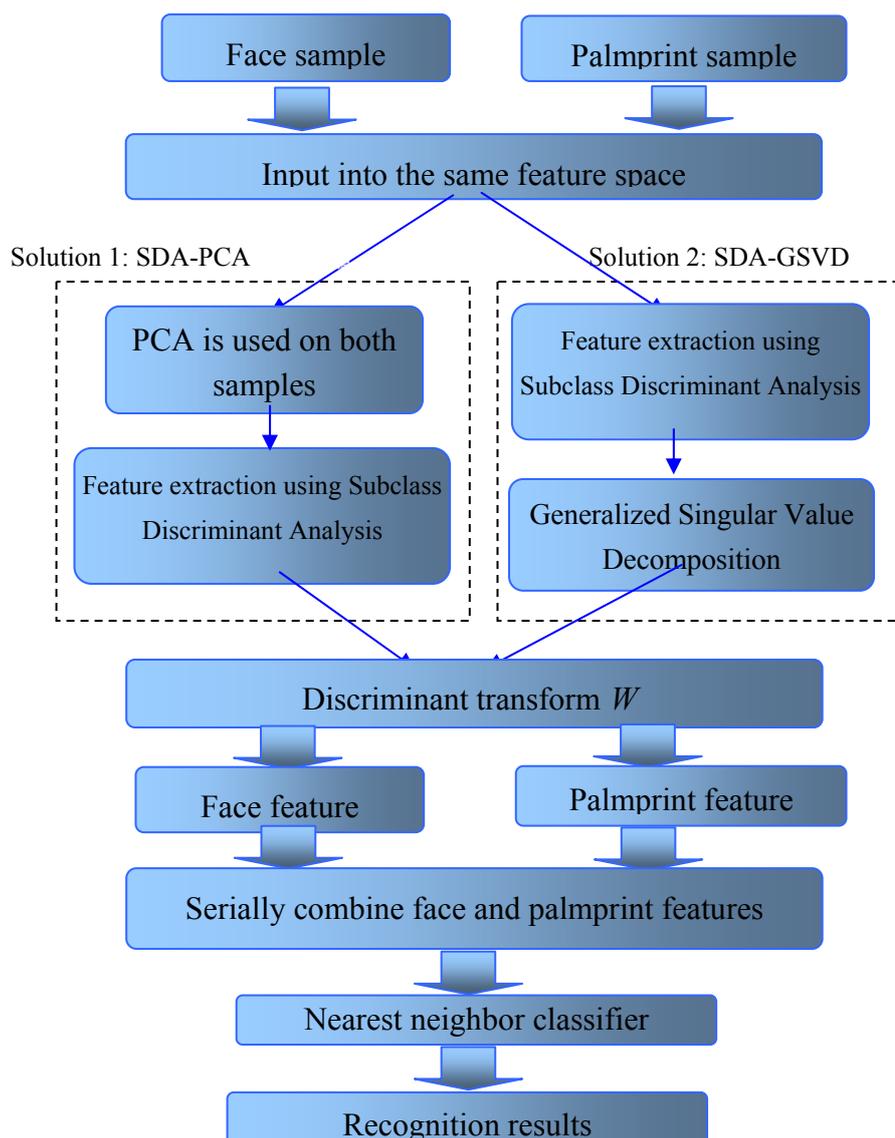


Figure 2 displays the complete procedure of the proposed approach for multimodal biometric recognition. It is worthwhile to note that, on one hand, our approach outputs features of each modal separately, which is convenient for later processing; on the other hand, discriminative information of different modals have been initially fused in the extraction process, since their features are extracted from the same input space and the transformed space also consider the distribution of data of other modals. Therefore, we think this approach can effectively obtain fused discriminative information from multimodal data.

#### 4. SDA Kernelization Based Multimodal Biometric Feature Extraction

In this section, we provide the nonlinear extensions of two SDA based multimodal feature extraction approaches, which are named KPCA-SDA and KSDA-GSVD. In KPCA-SDA, we first apply Kernel PCA on each single modal before performing SDA. While in KSDA-GSVD, we directly perform Kernel SDA to fuse multimodal data by applying GSVD to avoid the singular problem.

##### 4.1. KPCA-SDA

In this subsection, the SDA-PCA approach is performed in a high dimension space by using the kernel trick. We realized the KPCA-SDA in the following steps:

(1) Nonlinear mapping.

Let  $\phi: R^d \rightarrow F$  denote a nonlinear mapping. The original samples  $x_{i1}^k$  and  $x_{i2}^k$  of two modalities (face and palmprint) are injected into  $F$  by  $\phi: x_{i1}^k \rightarrow \phi(x_{i1}^k)$ ,  $x_{i2}^k \rightarrow \phi(x_{i2}^k)$ . We obtain two sets of mapped samples  $\Psi_1 = \{\phi(x_{11}^1), \phi(x_{11}^2), \dots, \phi(x_{c_1}^{n_c})\}$ ,  $\Psi_2 = \{\phi(x_{12}^1), \phi(x_{12}^2), \dots, \phi(x_{c_2}^{n_c})\}$ .

(2) Perform KPCA for each single modal database .

For the  $j^{th}$  modal, we perform KPCA by maximizing the following equation:

$$J(w_{kPCA}^{j\phi}) = w_{kPCA}^{j\phi T} S_i^{j\phi} w_{kPCA}^{j\phi} \quad (24)$$

where  $S_i^{j\phi} = \sum_{i=1}^c \sum_{k=1}^{n_c} (\phi(x_{ij}^k) - m_j^\phi)(\phi(x_{ji}^k) - m_j^\phi)^T$ , and  $m_j^\phi$  is the global mean of the  $j^{th}$  modal database in the kernel space.

According to the kernel reproducing theory [34], the projection transformation  $w_{kPCA}^{j\phi}$  in  $F$  can be linearly expressed by using all the mapped samples:

$$w_{kPCA}^{j\phi} = \sum_{i=1}^c \sum_{k=1}^{n_c} \alpha_{ij}^k \phi(x_{ij}^k) = \Psi_j \alpha_j \quad (25)$$

where  $\alpha_j = (\alpha_{1j}^1, \alpha_{1j}^2, \dots, \alpha_{c_j}^{n_c})^T$  is a coefficient matrix.

Substituting Equation (26) into Equation (25), we have:

$$J(w^{j\phi}) = \alpha_j^T \Psi_j^T \Psi_j \Psi_j^T \Psi_j \alpha_j = \alpha_j^T K^j K^{jT} \alpha_j \quad (26)$$

where  $K^j = \Psi_j^T \Psi_j$ , which indicates an  $N \times N$  non-symmetric kernel matrix whose element is  $K_{m,n}^j = \langle \phi(x_{jm}^k), \phi(x_{jn}^k) \rangle$ , where  $N$  denotes the total number of the samples,  $x_{jm}^k$  denotes the  $m^{th}$  sample of the  $j^{th}$  modal database.

The solution of Equation (27) is equivalent to the eigenvalue problem:

$$\lambda^j \alpha_j = K_j K_j^T \alpha_j \quad (27)$$

The optimal solutions  $\alpha_j = (\alpha_{j1}, \alpha_{j2}, \dots, \alpha_{j(N-c)})^T$  are the eigenvectors corresponding to  $N - c$  largest eigenvalues of  $K_j K_j^T$ . We project the mapped training sample set  $\Psi_j$  on  $w_{kPCA}^{j\phi}$  by:

$$Z_{KPCA}^{j\phi} = w_{kPCA}^{j\phi T} \Psi_j = \alpha_j^T \Psi_j^T \Psi_j = \alpha_j^T K_j \quad (28)$$

(3) Calculate kernel discriminant vectors in the KPCA transformed space.

By using the KPCA transformed sample set  $Z_{KPCA}^{j\phi}$ , we reformulate Equations (13) and (14) as:

$$S_B^\phi = \sum_{i=1}^{c-1} \sum_{j=1}^2 \sum_{k=i+1}^c \sum_{l=1}^2 p_{ij} p_{kl} (\mu_{ij}^\phi - \mu_{kl}^\phi)(\mu_{ij}^\phi - \mu_{kl}^\phi)^T \quad (29)$$

$$S_W^\phi = \frac{1}{N} \sum_{i=1}^c \sum_{j=1}^2 \sum_{k=1}^{n_c} (z_{ij}^{k\phi} - \mu_{ij}^\phi)(z_{ij}^{k\phi} - \mu_{ij}^\phi)^T \quad (30)$$

where  $Z_{ij}^{k\phi}$  is the sample in  $Z_{KPCA}^{j\phi}$ , and  $\mu_{ij}^\phi = \sum_{k=1}^{n_c} z_{ij}^{k\phi} / n_c$ .

We can obtain a set of nonlinear discriminant vectors  $W_{SDA}^{\phi}$ , *i.e.*, the eigenvector of matrix  $(S_W^\phi)^{-1} S_B^\phi$  associated with the largest eigenvalues.

(4) Construct the nonlinear projection transformation and do classification.

We then construct the nonlinear projection transformation  $W^{j\phi}$  as:

$$W^{j\phi} = w_{kPCA}^{j\phi} W_{SDA}^\phi \quad (31)$$

After the optimal transform  $W^{j\phi}$  is obtained, the fused features can be generated as:

$$y^\phi = \begin{bmatrix} W^{1\phi T} \Psi_1 \\ W^{2\phi T} \Psi_2 \end{bmatrix} \quad (32)$$

#### 4.2. KSDA-GSVD

In this subsection, the SDA-GSVD is performed in a high dimension space by using the kernel trick. Given two sets of mapped samples  $\Psi_1 = \{\phi(x_{11}^1), \phi(x_{11}^2), \dots, \phi(x_{c_1}^{n_c})\}$ ,  $\Psi_2 = \{\phi(x_{12}^1), \phi(x_{12}^2), \dots, \phi(x_{c_2}^{n_c})\}$ , that correspond to face and palmprint modalities, respectively. Afterwards,  $H_b$  and  $H_w$  are recalculated in the kernel space:

$$H_b^\phi = [H_{(c-1)1}^\phi, H_{(c-1)2}^\phi, H_{(c-2)1}^\phi, \dots, H_{11}^\phi, H_{12}^\phi] \quad (33)$$

$$H_w^\phi = [\phi(x_{ij}^1) - \mu_{ij}^\phi, \phi(x_{ij}^2) - \mu_{ij}^\phi, \dots, \phi(x_{ij}^{n_c}) - \mu_{ij}^\phi]_{i=1, \dots, c, j=1, 2} \quad (34)$$

$$\text{where } H_{(c-m)n}^\phi = 2(c-N)\mu_{mn}^\phi - \sum_{k=m+1}^c \sum_{l=1}^2 \mu_{kl}^\phi, \text{ and } \mu_{ij}^\phi = \sum_{k=1}^{n_c} \phi(x_{ij}^k) / n_c \quad (35)$$

Then, we apply GSVD to calculate the optimal transformation so that the singular problem is avoided. The procedures are precisely introduced in Algorithm 1. When the optimal  $W^\phi$  is obtained, the fused features can be generated as:

$$Y^\phi = \begin{bmatrix} y_{i1}^k \\ y_{i2}^k \end{bmatrix} = \begin{bmatrix} \hat{W}^{\phi T} \phi(x_{i1}^k) \\ \hat{W}^{\phi T} \phi(x_{i2}^k) \end{bmatrix} = \begin{bmatrix} \hat{W}^{\phi T} \Psi_1 \\ \hat{W}^{\phi T} \Psi_2 \end{bmatrix} \quad (36)$$

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

## 5. Experiments

In this section, we compare the proposed multimodal feature extraction approaches with single modal method and several representative multimodal biometric fusion methods. The identification and verification performance of our approaches and other compared methods is evaluated on two face databases and one palmprint database.

### 5.1. Introduction of Databases

Two public face databases (AR and FRGC) and one public palmprint database (PolyU palmprint database) are employed to testify our proposed approaches. The AR face database [35] contains over 4,000 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 3,094 ( $=119 \times 26$ ) samples. All color images are transformed into gray images and each image was scaled to  $60 \times 60$  with 256 gray levels. Figure 3 illustrates all of the samples of one subject.

**Figure 3.** Demo images of one subject from the AR face database.



The FRGC database [36] contains 12,776 training images that consist of both controlled images and uncontrolled images, including 222 individuals, each 36–64 images for the FRGC Experiment 4. The controlled images have good image quality, while the uncontrolled images display poor image quality, such as large illumination variations, low resolution of the face region, and possible blurring. It is these uncontrolled factors that pose the grand challenge to face recognition performance. We use the training images of the FRGC Experiment 4 as our database. We choose 36 images of each individual and then crop every image to the size of  $60 \times 60$ . All images of one subject are shown in Figure 4.

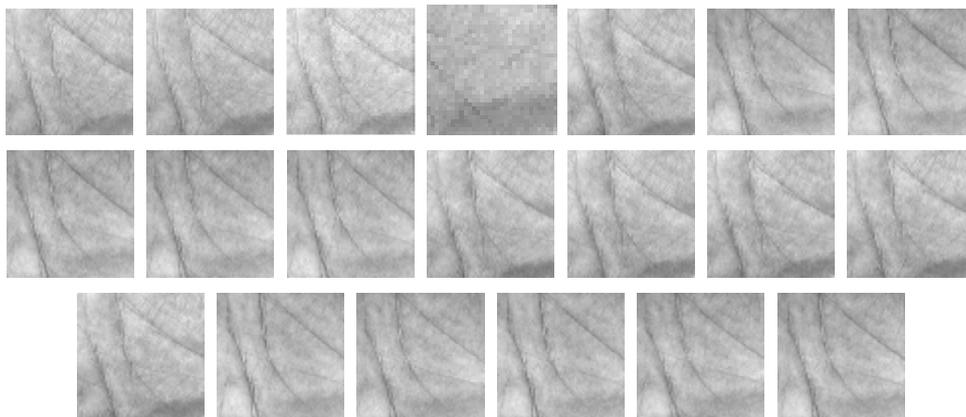
The palmprint database [37,38], which is provided by the Hong Kong Polytechnic University (HK PolyU), collected palmprint images from 189 individuals. Around 20 palmprint images from each individual were collected in two sessions, where around 10 samples were captured in the first session and the second session, respectively. Therefore, the database contains a total of 3,780 images from 189 palms. In order to reduce the computational cost, each subimage was compressed to  $60 \times 60$ . We

took these subimages as palmprint image samples for our experiments. All cropped images of one subject in Figure 5.

**Figure 4.** Demo images of one subject from the FRGC face database.



**Figure 5.** Demo images of one subject from the PolyU palmprint database.

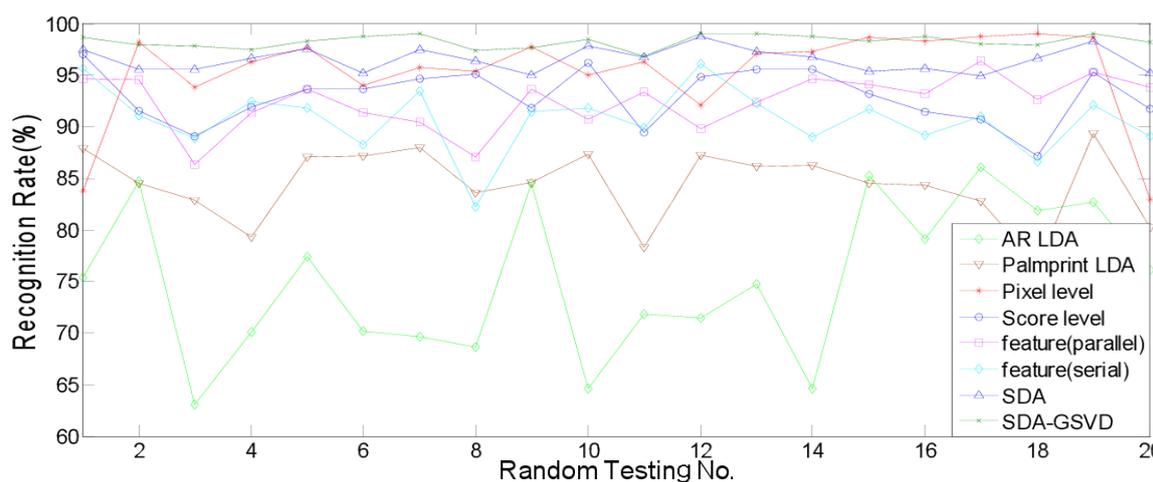


In order to testify the proposed fusion techniques, in the experiment which we fuse AR database and PolyU palmprint database, we choose 119 subjects from both face and palmprint database, and each class contains 20 samples. Similarly, in the experiment which we fuse FRGC database and PolyU palmprint database, we choose 189 subjects from both face and palmprint database, and each class contains 20 samples. We assume that samples of one subject in the palmprint database correspond to the samples of one subject in the face database. For the AR face database and PolyU palmprint database, we randomly select eight samples from each person (four face samples from AR database and four palmprint samples from PolyU database) for training, while use the rest for testing. For the FRGC face database and PolyU palmprint database, we randomly select six samples from each person (three face samples from FRGC database and three palmprint samples from PolyU database) for training, while use the rest for testing. We run all compared methods 20 times. In our experiments, 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  $\delta_i = i \times \delta$ ,  $i \in 1, \dots, 20$ , where  $\delta$  is the standard deviation of training data set. For each compared kernel method, the parameter  $i$  was selected such that the best classification performance was obtained.

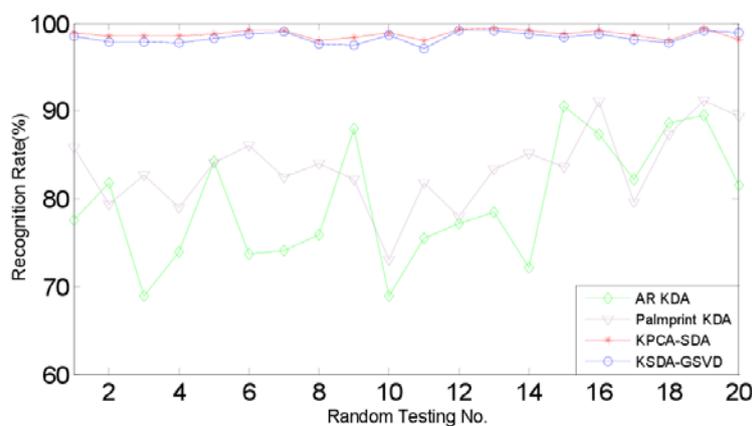
5.2. Experimental Identification Results

Firstly, the identification experiments are conducted. Identification is a one-to-many comparison which aims to answer the question of “who is this person?” We compare the identification performance of two proposed approaches, *i.e.*, SDA-PCA (which is abbreviated to SDA here), SDA-GSVD, with single modal recognition method using traditional LDA, a representative pixel level fusion method [1], parallel and serial feature level fusion [3], and score level fusion method using the sum rule [7], respectively. Further, we compare the proposed kernelizaion methods (KPCA-SDA and KSDA-GSVD), with single modal recognition method using KDA. Figures 6 and 7 show the recognition rates of 20 random tests of our approaches and other compared methods: (a) SDA, SDA-GSVD, LDA (single modal), Pixel level fusion, parallel feature fusion, Serial feature fusion and Score level fusion; (b) KPCA-SDA, KSDA-GSVD and KDA (single modal). The average recognition rates are given in Tables 1 and 2, which correspond to the figures above.

**Figure 6.** Recognition rates of compared methods on AR face and PolyU palmprint databases: (a) Linear methods; (b) Nonlinear methods.

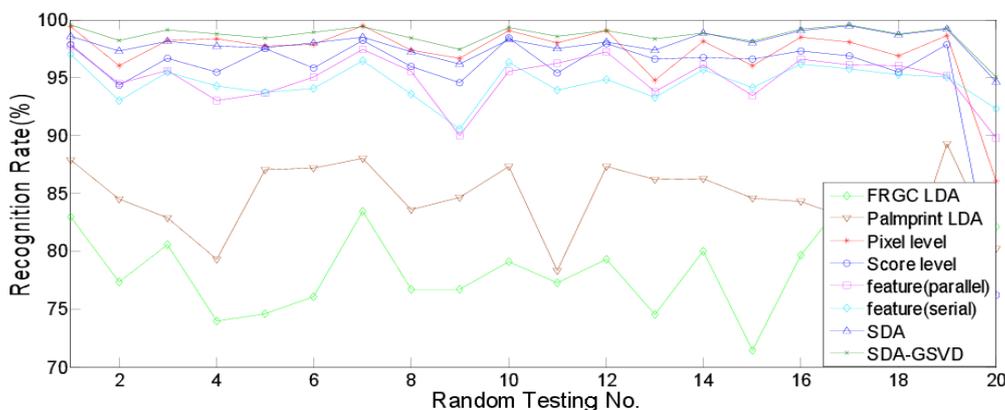


(a)

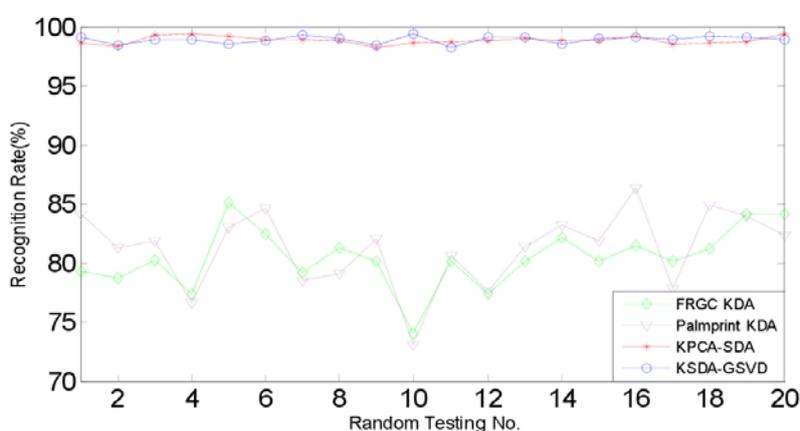


(b)

**Figure 7.** Recognition rates of compared methods on FRGC face and PolyU palmprint databases: **(a)** Linear methods; **(b)** Nonlinear methods.



**(a)**



**(b)**

**Table 1.** Average recognition rates of compared methods on AR face and PolyU palmprint databases

AR and palmprint		Average recognition rates (%)
Single modal recognition	AR LDA	75.09 ± 7.39
	Palmprint LDA	82.26 ± 3.50
	Pixel level fusion [1]	95.35 ± 4.50
	Parallel feature fusion [3]	92.48 ± 2.61
Multimodal recognition	Serial feature fusion [3]	90.71 ± 3.06
	Score level fusion [7]	92.99 ± 2.63
	SDA based feature extraction	96.52 ± 1.16
	SDA-GSVD based feature extraction	<b>98.23 ± 0.68</b>

**(a)** Linear methods

AR and palmprint		Average recognition rates (%)
Single modal recognition	AR KDA	79.50 ± 6.83
	Palmprint KDA	83.45 ± 4.47
Multimodal recognition	KPCA-SDA	98.74 ± 0.45
	KSDA-GSVD	<b>99.15 ± 0.63</b>

**(b)** Nonlinear methods

**Table 2.** Average recognition rates of compared methods on FRGC face and PolyU palmprint databases.

FRGC and palmprint		Average recognition rates (%)
Single modal recognition	FRGC LDA	78.26 ± 4.53
	Palmprint LDA	80.22 ± 3.26
	Pixel level fusion [1]	97.21 ± 2.89
	Parallel feature fusion [3]	94.92 ± 2.17
Multimodal recognition	Serial feature fusion [3]	94.54 ± 1.57
	Score level fusion [7]	95.59 ± 4.70
	SDA based feature extraction	98.06 ± 1.09
	SDA-GSVD based feature extraction	<b>98.61 ± 0.99</b>

(a) Linear methods

FRGC and palmprint		Average recognition rates (%)
Single modal recognition	AR KDA	80.44 ± 2.57
	Palmprint KDA	81.23 ± 3.26
Multimodal recognition	KPCA-SDA	98.82 ± 0.32
	KSDA-GSVD	<b>99.02 ± 0.31</b>

(b) Nonlinear methods

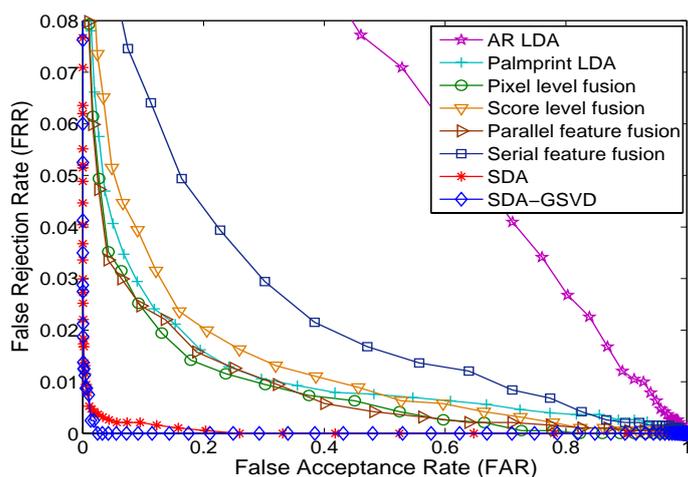
Table 1 shows that on the AR and PolyU palmprint databases, SDA and SDA-GSVD perform better than other compared linear methods. It also shows that KPCA-SDA and KSDA-GSVD achieve better recognition results than KDA (single modal). Compared with the single modal LDA, pixel level fusion, parallel feature fusion, parallel feature fusion, serial feature fusion and score level fusion, SDA improves the average recognition rate at least by 3.53% (=98.23%–92.99%), SDA-GSVD improves the average recognition rate at least by 5.24% (=98.23%–92.99%). And the average recognition rate of KPCA-SDA is at least 15.29% (=98.74%–83.45%) higher than that of KDA (single modal), and the average recognition rate of KSDA-GSVD is at least 15.7% (=99.15%–83.45%) higher than that of KDA (single modal). Table 2 shows a similar phenomenon on the FRGC and PolyU palmprint databases. SDA boosts the average recognition rate at least by 0.85% (=98.06%–97.21%), and SDA-GSVD boosts the average recognition rate at least by 1.40% (=98.61%–97.21%) than other linear methods. The average recognition rate of KPCA-SDA is at least 17.59% (=98.82–81.23) higher than that of KDA (single modal), and the average recognition rate of KSDA-GSVD is at least 17.79% (=99.02%–81.23%) higher than that of KDA (single modal).

### 5.3. Experimental Results of Verification

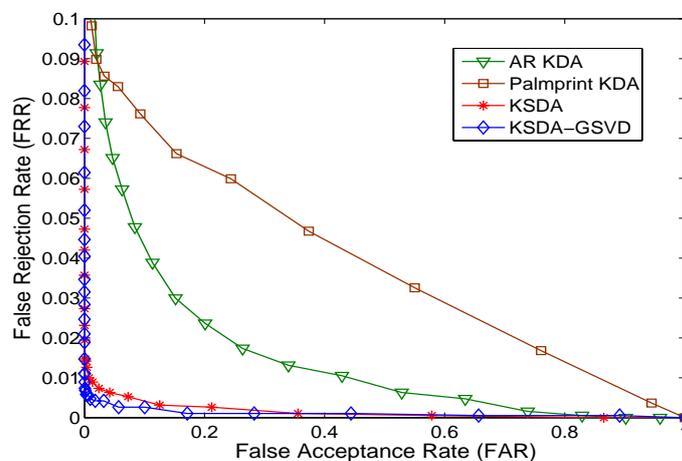
Verification is a one-to-one comparison which aims to answer the question of “whether the person is one he/she claims to be”. In the verification experiments, we show the receiver operating characteristic (ROC) curves, which plot the false rejection rate (FRR) *versus* the false accept rate (FAR), to report the verification performance. There is a tradeoff between the FRR and the FAR. It is possible to reduce one of them with the risk of increasing the other one. Thus the curve which is called receiver operating characteristic (ROC) reflects the tradeoff between the FAR and FRR, and FRR is plotted as a function of FAR.

Figures 8 and 9 show the Receiver Operating Characteristic (ROC) curves of our approaches and other compared methods on different databases. Table 3 shows the equal error rate (EER) of all compared methods. From the ROC curves shown in Figures 8–9 and the results listed in Table 3, we can see that our SDA based feature extraction approaches attains a significantly low EER (a point on the ROC curve where FAR is equal to FRR) than other representative multimodal fusion methods, including pixel level fusion method, score level fusion method and feature level fusion methods. On the AR face and PolyU palmprint databases, the lowest EER of related methods is 3.71%, while the EER of our approaches are all below 1%. And our KSDA-GSVD approach obtains the lowest EER 0.56% among all compared methods. On the FRGC face and PolyU palmprint databases, the lowest EER of other methods is 2.62%, while the EER of ours are all below 2%. Especially, the proposed SDA-GSVD approach gets the lowest EER that is 0.28%. The above experimental results demonstrate the superiority of our approaches.

**Figure 8.** ROC curves of all compared methods on AR face and PolyU palmprint databases: (a) Linear methods; (b) Nonlinear methods.

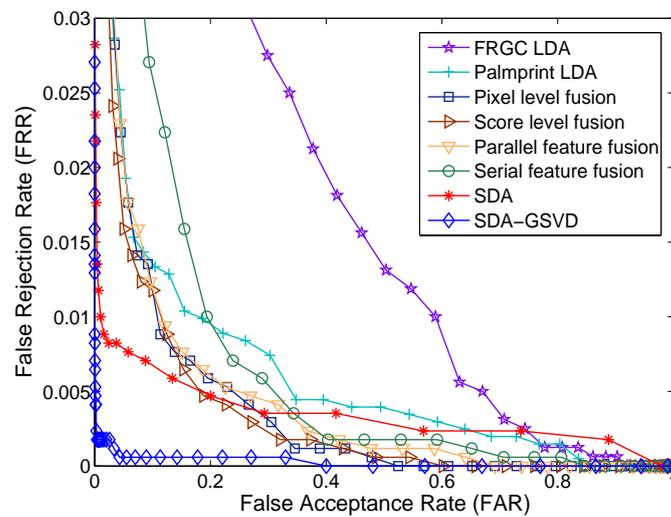


(a)

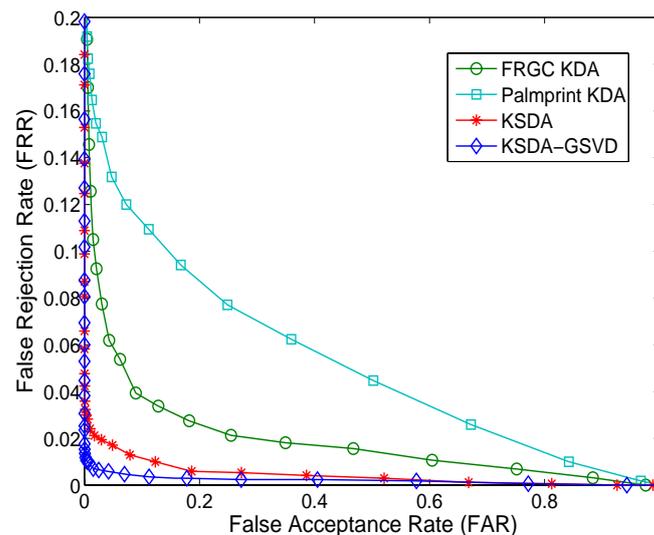


(b)

**Figure 9.** ROC curves of all compared methods on FRGC face and PolyU palmprint databases: **(a)** Linear methods; **(b)** Nonlinear methods.



(a)



(b)

**Table 3.** The equal error rate (EER) of all compared methods on different databases.

Method		AR and Palmprint EER (%)	FRGC and Palmprint EER (%)
Single modal recognition	Face LDA	15.45	8.13
	Palmprint LDA	4.32	3.14
Multimodal recognition	Face KDA	6.13	5.72
	Palmprint KDA	8.36	10.85
	Pixel level fusion [1]	3.95	3.25
	Parallel feature fusion [3]	3.71	3.27
	Serial feature fusion [3]	7.84	4.41
	Score level fusion [7]	5.12	2.62
	SDA based feature extraction	0.83	1.05
	SDA-GSVD based feature extraction	0.72	<b>0.28</b>
	KSDA based feature extraction	0.87	1.90
	KSDA-GSVD based feature extraction	<b>0.56</b>	0.84

## 6. Conclusions

In this paper, we present novel multimodal biometric feature extraction approaches using subclass discriminant analysis (SDA). Considering the nonsingularity requirements, we present two ways to overcome this problem. The first is to initially do principle component analysis before SDA, and the second is to employ generalized singular value decomposition (GSVD) to directly obtain the solution. Further, we present the kernel extensions (KPCA-SDA and KSDA-GSVD) for multimodal biometric feature extraction. We perform the experiments on two public face databases (*i.e.*, AR face database and FRGC database) and the PolyU palmprint database. In designing the experiments, we firstly do extraction on the AR and palmprint database, secondly on the FRGC and palmprint database. Compared with several representative linear and nonlinear multimodal biometrics recognition methods, the proposed approaches acquire better identification and verification performance. In particular, the proposed KSDA-GSVD approach performs best on all the databases.

## Acknowledgement

The work described in this paper was fully supported by the NSFC under Project No. 61073113, the New Century Excellent Talents of Education Ministry under Project No. NCET-09-0162, the Doctoral Foundation of Education Ministry under Project No. 20093223110001, the Qing-Lan Engineering Academic Leader of Jiangsu Province and 333 Engineering of Jiangsu Province.

## References

1. Jing, X.Y.; Yao, Y.F.; Zhang, D.; Yang, J.Y.; Li, M. Face and palmprint pixel level fusion and Kernel DCV-RBF classifier for small sample biometric recognition. *Pattern Recognit.* **2007**, *40*, 3209–3224.
2. Petrovi, V.; Xydeas, C. Computationally efficient pixel-level image fusion. *Proc. Euroficsion* **1999**, 177–184.
3. Yang, J.; Yang, J.Y.; Zhang, D.; Lu, J.F. Feature fusion: Parallel strategy vs. serial strategy. *Pattern Recognit.* **2003**, *36*, 1369–1381.
4. Sun, Q.S.; Zeng, S.G.; Heng, P.A.; Xia, D.S. Feature fusion method based on canonical correlation analysis and handwritten character recognition. In *Proceedings of the Control Automation Robotics and Vision Conference*, Kunming, China, 6–9 December 2004; pp. 1547–1552.
5. Yao, Y.F.; Jing, X.Y.; Wong, H.S. Face and palmprint feature level fusion for single sample biometrics recognition. *Neurocomputing* **2007**, *70*, 1582–1586.
6. Raghavendra, R.; Dorizzi, B.; Rao, A.; Kumar, G.H. Designing efficient fusion schemes for multimodal biometric systems using face and palmprint. *Pattern Recognit.* **2011**, *44*, 1076–1088.
7. Kumar, A.; Zhang, D. Personal authentication using multiple palmprint representation. *Pattern Recognit.* **2005**, *38*, 1695–1704.
8. Jain, A.K.; Ross, A. Learning user-specific parameters in a multibiometric system. In *Proceedings of the International Conference on Image Processing (ICIP)*, New York, NY, USA, 22–25 September 2002; pp. 57–70.

9. He, M.; Horng, S.; Fan, P.; Run, R.; Chen, R.; Lai, J.L.; Khan, M.K.; Sentosa, K.O. Performance evaluation of score level fusion in multimodal biometric systems. *Pattern Recognit.* **2009**, *43*, 1789–1800.
10. Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D.J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 711–720.
11. Liu, C.J.; Wechsler, H. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Trans. Image Process.* **2002**, *11*, 467–476.
12. Jing, X.Y.; Zhang, D.; Tang, Y.Y. An improved LDA approach. *IEEE Trans. Syst. Man Cybern. Part B* **2004**, *34*, 1942–1951.
13. Jing, X.Y.; Zhang, D.; Jin, Z. UODV: Improved algorithm and generalized theory. *Pattern Recognit.* **2003**, *36*, 2593–2602.
14. Cevikalp, H.; Neamtu, M.; Wilkes, M.; Barkana, A. Discriminative common vectors for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 4–13.
15. Kim, T.K.; Wong, S.F.; Stenger, B.; Kittler, J.; Cipolla, R. Incremental linear discriminant analysis using sufficient spanning set approximations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Minneapolis, MN, USA, 17–22 June 2007; pp. 1–8.
16. Zhang, Y.; Yeung, D.Y. Semi-supervised discriminant analysis using robust path-based similarity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
17. Sugiyama, M. Local fisher discriminant analysis for supervised dimensionality reduction. In *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, New York, NY, USA, 25–29 June 2006; pp. 905–912.
18. Yang, M.; Zhang, L.; Feng, X.; Zhang, D. Fisher discrimination dictionary learning for sparse representation. In *Proceedings of the International Conference on Computer Vision*, Barcelona, Spain, 6–13 November 2011; pp. 1–8.
19. Lan, C.; Jing, X.Y.; Zhang, D.; Gao, S.; Yang, J.Y. Discriminant subclass-center manifold preserving projection for face feature extraction. In *Proceedings of the International Conference on Image Processing (ICIP)*, Brussels, Belgium, 11–14 September 2011; pp. 3070–3073.
20. Mika, S.; Rätsch, G.; Weston, J.; Schölkopf, B.; Müller, K.R. Fisher discriminant analysis with kernels. In *Proceedings of the IEEE Neural Networks for Signal Processing Workshop*, Madison, WI, USA, August 1999; pp. 41–48.
21. Baudat, G.; Anouar, F. Generalized discriminant analysis using a kernel approach. *Neural Comput.* **2000**, *12*, 2385–2404.
22. Lu, J.; Plataniotis, K.N.; Venetsanopoulos, A.N. Face recognition using kernel direct discriminant analysis algorithms. *IEEE Trans. Neural Netw.* **2003**, *14*, 117–126.
23. Liu, Q.S.; Lu, H.Q.; Ma, S.D. Improving kernel fisher discriminant analysis for face recognition. *IEEE Trans. Circuits Syst. Video Technol.* **2004**, *14*, 42–49.
24. Yang, J.; Frangi, A.F.; Zhang, D.; Yang, J.Y.; Jin, Z. KPCA plus LDA: A complete kernel fisher discriminant framework for feature extraction and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 230–244.

25. Jing, X.Y.; Yao, Y.F.; Zhang, D.; Yang, J.Y.; Li, M. Face and palmprint pixel level fusion and KDCV-RBF classifier for small sample biometric recognition. *Pattern Recognit.* **2007**, *40*, 3209–3224.
26. Chen, B.; Yuan, L.; Liu, H.; Bao, Z. Kernel subclass discriminant analysis. *Neurocomputing* **2007**, *72*, 455–458.
27. Sugiyama, M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *J. Mach. Learn. Res.* **2007**, *8*, 1027–1061.
28. Jing, X.Y.; Li, S.; Yao, Y.F.; Bian, L.S.; Yang, J.Y. Kernel uncorrelated adjacent-class discriminant analysis. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, Istanbul, Turkey, 23–26 August 2010; pp. 706–709.
29. Li, S.; Jing, X.Y.; Zhang, D.; Yao, Y.F.; Bian, L.S. A novel kernel discriminant feature extraction framework based on mapped virtual samples for face recognition. In *Proceedings of the IEEE International Conference on Image Processing*, Brussels, Belgium, 11–14 September 2011; pp. 3066–3069.
30. Zhu, M.; Martinez, A.M. Subclass discriminant analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1274–1286.
31. Howland, P.; Park, H. Generalizing discriminant analysis using the generalized singular value decomposition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 995–1006.
32. Liu, F.; Sun, Q.; Zhang, J.; Xia, D. Generalized canonical correlation analysis using GSVD. In *Proceedings of International Symposium Computer Science Computational Technology*, Shanghai, China, 20–22 December 2008; pp. 136–141.
33. Kwang, I.K.; Keechul, J.; Hang, J.K. Face recognition using kernel principle component analysis. *IEEE Signal Process. Lett.* **2002**, *9*, 40–42.
34. Shawe-Taylor, J.; Cristianini, N. *Kernel Methods for Pattern Analysis*; Cambridge University Press: London, UK, 2004.
35. Martinez, A.M.; Benavente, R. *The AR Face Database*; CVC Technical Report; The Ohio State University: Columbus, OH, USA, 1998; Volume 24.
36. Phillips, P.J.; Flynn, P.J.; Scruggs, T.; Bowyer, K.; Chang, J.; Hoffman, K.; Marques, J.; Min, J.; Worek, W. Overview of the face recognition grand challenge. In *Proceedings of the IEEE Conference Computer Vision and Pattern Recognition*, San Diego, CA, USA, 20–25 June 2005; pp. 947–954.
37. Zhang, D.; Kong, W.K.; You, J.; Wong, M. Online palmprint identification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2003**, *25*, 1041–1150.
38. Zhang, L.; Zhang, D. Characterization of palmprints by wavelet signatures via directional context modeling. *IEEE Trans. Syst. Man Cybern. Part B* **2004**, *34*, 1335–1347.