Generic Training Set based Multimanifold Discriminant Learning for Single Sample Face Recognition

Xiwei Dong^{1,3}, Fei Wu^{1} and Xiao-Yuan Jing^{1,2}
[1] College of Automation, Nanjing University of Posts and Telecommunications
Nanjing 210003, China
[e-mail: dxwdxw2005@126.com, wufei_8888@126.com, jingxy_2000@126.com]
[2] State Key Laboratory of Software Engineering, School of Computer, Wuhan University
Wuhan 430072, China
[3] School of Information Science and Technology, Jiujiang University
Jiujiang 332005, China
*Corresponding author: Xiao-Yuan Jing

Received April 11, 2016; revised July 2, 2017; accepted September 5, 2017; published January 31, 2017

Abstract

Face recognition (FR) with a single sample per person (SSPP) is common in real-world face recognition applications. In this scenario, it is hard to predict intra-class variations of query samples by gallery samples due to the lack of sufficient training samples. Inspired by the fact that similar faces have similar intra-class variations, we propose a virtual sample generating algorithm called k nearest neighbors based virtual sample generating (kNNVSG) to enrich intra-class variation information for training samples. Furthermore, in order to use the intra-class variation information of the virtual samples generated by kNNVSG algorithm, we propose image set based multimanifold discriminant learning (ISMMDL) algorithm. For ISMMDL algorithm, it learns a projection matrix for each manifold modeled by the local patches of the images of each class, which aims to minimize the margins of intra-manifold and maximize the margins of inter-manifold simultaneously in low-dimensional feature space. Finally, by comprehensively using kNNVSG and ISMMDL algorithms, we propose k nearest neighbor virtual image set based multimanifold discriminant learning (kNNMMDL) approach for single sample face recognition (SSFR) tasks. Experimental results on AR, Multi-PIE and LFW face datasets demonstrate that our approach has promising abilities for SSFR with expression, illumination and disguise variations.

Keywords: Single sample face recognition, virtual samples, illumination normalization, multimanifold discriminant learning, computer vision

This work is partially funded by the National Science Foundation of China (NSFC) with Grant numbers of 61272273 and 61702280, in part by the Natural Science Foundation of Jiangsu Province (No. BK20170900), in part by the Natural Science Fund for Colleges and Universities in Jiangsu Province (No. 17KJB520025). It is also supported by Nanjing University of Posts and Telecommunications under Grant number of XJKY14016 and the Scientific Research Staring Foundation for Introduced Talents in NJUPT (NUPTSF, No. NY217009). In addition, it is also funded by Education Department of Jiangxi Province under the Grant No. GJJ151076.

http://doi.org/10.3837/tiis.2018.01.018 **ISSN : 1976-7277**

1. Introduction

Face recognition (FR) has been attracting significant attention in many computer vision applications [1], and many researchers have proposed a number of methods to address various problems emerging in practical face recognition scenarios, such as face identification/verification in unconstrained or less constrained environment [2-5]. As a result, machine vision can surpass human vision when sufficient and representative training samples are supplied to those methods [6]. In fact, collecting plenty of samples means heavy workload, huge storage and processing expense. And unfortunately, in many real-world FR applications (e.g., law enforcement, e-passport, driver license, etc.), we can obtain only a single sample per person (SSPP). We call face recognition in this scenario as single sample face recognition (SSFR). If only a single sample per person is available, most current face recognition techniques would suffer degraded performance or fail to work due to lack of sufficient samples. For example, traditional discriminative subspace learning based face recognition methods may fail to work, because the intra-class variations cannot be well estimated in SSPP scenario [7]. Aiming to efficiently perform FR tasks with only a single sample per person, a number of SSFR methods were proposed [8-14]. When the single sample image of each person is divided into several local patches, face recognition tasks can be conducted by employing probabilistic model, discriminative multimanifold analysis, collaborative representation based classifier and linear discriminant analysis (LDA) [15-19]. However, the single training sample of each person usually just contains partial intra-class facial variations. If the intra-class variations (e.g., illumination, expression and disguise) of query sample images do not exist in the single training sample of each person, the performance of SSFR would be affected significantly.

In this paper, we propose an algorithm to tackle the problem of intra-class variations lackness in SSPP scenario. And then we propose a SSFR approach to efficiently perform SSFR tasks. The contributions of this paper are summarized as following three points:

(1) We propose a virtual sample generating algorithm called k nearest neighbors based virtual sample generating (kNNVSG) to enrich intra-class variation information, which may do not exist in the only single training sample of each person, by using generic training set.

(2) In order to use the intra-class variation information of virtual samples generated by kNNVSG to better learn low-dimensional feature space, we propose image set based multimanifold discriminant learning (ISMMDL) algorithm. It is different from single gallery sample based multimanifold discriminant learning algorithms and can use both the original single gallery sample image and generated virtual sample images of each class simultaneously.

(3) By comprehensively using our proposed kNNVSG and ISMMDL algorithms, we propose k nearest neighbor virtual image set based multimanifold discriminant learning (kNNMMDL) approach for performing SSFR tasks.

The rest of the paper is organized as follows. In Section 2, we briefly review related work. In Section 3, we give detailed description of our proposed approach. Experimental results are reported in Section 4. And conclusions are drawn in Section 5.

2. Related Work

In recent years, some methods have been proposed to address the SSFR problem. Those methods can be classified into three main categories: virtual sample generating methods, generic learning methods and image partitioning based methods.

(1) Virtual sample generating methods. For this category, some additional training samples for each person are virtually generated such that discriminative subspace learning can be used for feature extraction. For example, Zhang et al. [20] and Gao et al. [21] proposed methods to address the SSFR problem based on singular value decomposition (SVD), respectively. Vetter [22] proposed a 3D virtual sample generating method. Although these methods can alleviate SSFR problem to a certain extent, one common shortcoming of these methods is that there is high correlation among the virtual samples as they cannot be considered as independent samples for feature extraction [15].

(2) Generic learning methods. Methods of this type first extract discriminative features from an additional generic training set which contains multiple samples per person (MSPP), then those features are subsequently used for SSFR. For example, Su et al. [23] proposed to adapt the within-class and between-class scatter matrices computed from a generic training set. Si et al. [24] proposed a transfer subspace learning method which uses the discriminative model learned from generic training set to perform SSFR. Based on the simple observation that similar subjects have similar intra-personal variations, Wang et al. [25] designed an adaptive linear regression classifier (ALRC) for SSFR. Assuming that the intra-class variations of one subject can be approximated by a sparse linear combination of those of other subjects, Deng et al. [26] proposed extend SRC (ESRC) for SSFR. Considering the fact that the intra-class facial variations can be shared across different subjects, Zhu et al. [14] proposed a local generic representation (LGR) based framework for face recognition with SSPP. Based on the assumption that generic training set is full enough with sufficient variations, a collaborative probabilistic labels (CPL) method was developed [13]. By transferring the intra-class variations of the generic training set to that of the gallery set, a discriminative transfer learning (DTL) method was designed for SSFR [11, 27]. Yang et al. [28] proposed a sparse variation dictionary learning (SVDL) method.

(3) Image partitioning based methods. These methods first partition images of each person into local patches then extract discriminative features of those local patches and subsequently conduct classification based on the extracted discriminative features. For example, aiming to address the occlusion problem in SSFR, Martinez [15] proposed a method which divides a face image into local patches and then a probabilistic approach is used to find the best match. Chen et al. [19] proposed to divide each face image into several sub-images with the same size, therefore obtaining multiple training samples for each class, and then apply FLDA to the newly produced samples. Zhu et al. [18] proposed a patch based CRC (PCRC) method which applies the collaborative representation based classification (CRC) [29] to each patch. Gao et al. [30] proposed a regularized patch-based representation for SSFR, which represents each image by a collection of patches and seeks their sparse representations under the gallery image patches and intra-class variance dictionaries at the same time. The methods mentioned above ignore the geometrical information of local patches in feature extracting process. Aiming to seek multiple projection matrices to uncover the geometrical information of manifolds modeled by local patches, Lu et al. [17] formulated SSFR as a manifold-manifold matching problem in low-dimensional feature space. In [14], Zhang et al. also formulated SSFR task as manifold matching problem, and proposed a two-step scheme for SSFR task. Unlike the conventional image partitioning methods, Pei et al. [9] proposed a nonparametric method termed decision pyramid classifier (DPC), which does not require a training process.

In this paper, by comprehensively using our proposed kNNVSG and ISMMDL algorithms, we propose an approach called kNNMMDL for SSFR tasks. The main characteristics of kNNMMDL approach are the following three aspects. (1) For the given gallery set with only a single sample per person, we can obtain an extended gallery set with multiple samples per person by using our proposed kNNVSG algorithm and generic training set, which can enrich intra-class variation information for training samples. (2) By employing Weber-face algorithm [31] to conduct illumination normalization on sample images, we can obtain the illumination-insensitive representations of those sample images, which can alleviate the adverse effect caused by illumination variations on face recognition performance. (3) Both the only single gallery sample and generated virtual samples of each class can be used simultaneously in our proposed ISMMDL algorithm, which can take full advantage of the intra-class variation information of those samples to learn better low-dimensional feature space.

3. Our Approach

In this section, we first introduce our proposed virtual sample generating algorithm kNNVSG which is used to enrich the intra-class variation information for training samples. Then, the Weber-face algorithm [31] which is used to alleviate the illumination problem in face recognition is described. After that, we represent our proposed ISMMDL algorithm. Sequentially, we depict the classification scheme for query sample. Finally, we describe the procedures of our proposed SSFR approach kNNMMDL in detail, which comprehensively uses our proposed kNNVSG and ISMMDL algorithms.

Assume that $X=[x_1,x_2,...,x_N]$ denotes the given gallery set, where $x_i \in \Re^d$, $i=1,2,...,N$, N indicates the sample number of the gallery set, $d=m\times n$, m and n are width and height of original gallery images of the gallery set, respectively. The class number of gallery samples of the gallery set X is N, i.e., there is only a single sample per class in the gallery set X.

3.1 k Nearest Neighbors based Virtual Sample Generating (kNNVSG)

Inspired by generic learning [25, 26, 28], we propose kNNVSG algorithm. Assume that $G=[G_1,G_2,...,G_J]$ denotes the introduced generic training set, where J indicates the class number of the objects of the generic training set, $G_j=[r_j,Q_j]$ represents the face image sample subset corresponding to the jth class in the generic training set. In the subset G_j, r_j denotes a reference sample of the jth class, and $Q_j=[q_{j1},q_{j2},...,q_{jV}]$ represents a variation set of the jth class where V is the category number of intra-class variations of Q_j. Furthermore, we assume that the dimensionality of the samples in the generic training set is identical to that of the samples in gallery set, i.e., $r_j \in \Re^d$, $q_{jv} \in \Re^d$, $j=1,2,...,J$, $v=1,2,...,V$. **Fig. 1** illustrates a number of sample images of three persons from a generic training set, where each person includes a reference sample (shown in solid box) and a variation set (shown in dashed box) which includes illumination, expression and disguise variations.

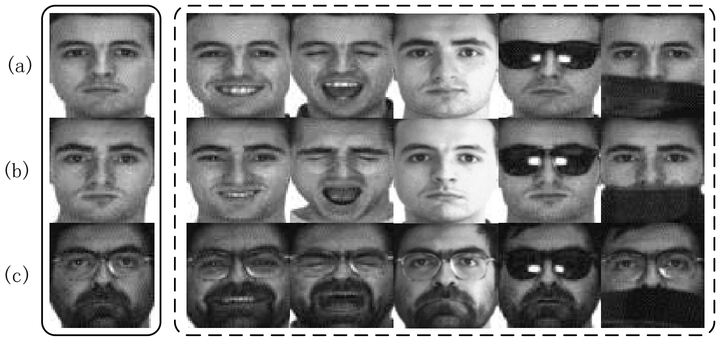


Fig. 1. Illustration of a number of sample images of three objects from a generic training set.

It is generally known that the shape of faces is highly similar. And, intra-class variations of the same category are significantly similar as well. For example, screaming is usually accompanied with mouth open. For human face images, similar objects typically have similar intra-class variations [25]. An intuitive illustration of this observation is displayed in **Fig. 1(a)** and **Fig. 1(b)**. Therefore, the intra-class variation of any gallery face can be approximated by those of the similar faces pulled from generic training set which contains sufficient samples. Hence, we can generate virtual samples by employing sample images from generic training set and the single gallery sample of each class to enrich the intra-class variation information of the class.

Given the only single gallery sample x_i of the ith class in the gallery set X, assume that $N(x_i)$ denotes the nearest neighbor sample set of x_i where the elements of $N(x_i)$ come from the reference sample set $R=[r_1,r_2,...,r_J]$, $C(N(x_i))$ represents the class label set which contains the class label of each sample in $N(x_i)$, and k_0 indicates the sample number of $N(x_i)$. Following the practice described in [19] for representing intra-class variation of face image, we give the procedures of our proposed virtual sample generating algorithm kNNVSG as follows.

First, we select variation sets which satisfy $j \in C(N(x_i))$ from variation sets $\{Q_j\}_{j=1}^{J}$, i.e., the selected variation sets are $\{Q_j \mid j \in C(N(x_i))\}$. Furthermore, by using the reference sample r_j and the vth sample q_{jv} in variation set Q_j, we extract the intra-class variation feature of the vth type of the jth class, which is denoted as $var_{jv} = q_{jv} - r_j$, where $j \in C(N(x_i))$, $v=1,2,...,V$. Finally, we fuse intra-class variation features of the vth type of k_0 objects with x_i to obtain the vth virtual sample $\tilde{x}_{iv}$ for the ith class, where $\tilde{x}_{iv} = x_i + VAR_v$, $VAR_v = \frac{1}{k_0}\sum_{j \in C(N(x_i))} var_{jv}$, $v=1,2,...,V$.

By combining the V virtual samples $\tilde{x}_{i1},\tilde{x}_{i2},...,\tilde{x}_{iV}$ and the gallery sample x_i together, we can obtain the extended training samples $\tilde{X}_i = [x_i,\tilde{x}_{i1},\tilde{x}_{i2},...,\tilde{x}_{iV}]$ of the ith class. For the sake of convenience, we denote $\tilde{X}_i$ as $\tilde{X}_i = [\tilde{x}_{i0},\tilde{x}_{i1},\tilde{x}_{i2},...,\tilde{x}_{iV}]$, where $\tilde{x}_{i0} = x_i$. Similarly, we can generate virtual samples for each class, and then the extended gallery set denoted by $\tilde{X} = [\tilde{X}_1,\tilde{X}_2,...,\tilde{X}_N]$ can be obtained, where $\tilde{X}_i = [\tilde{x}_{i0},\tilde{x}_{i1},\tilde{x}_{i2},...,\tilde{x}_{iV}]$, $i=1,2,...,N$. **Fig. 2** demonstrates the last step of the virtual sample generating algorithm kNNVSG. **Fig. 2(a)** illustrates a gallery sample image of original gallery set X . **Fig. 2(b)** shows the visual presentations of the intra-class variation features of five different types (illumination, simile, screaming, glasses and scarf), where the intra-class variation features are extracted from the samples of the variation sets of k_0 objects. And **Fig. 2(c)** illustrates five virtual sample images generated by fusing the images in **Fig. 2(a)** and **Fig. 2(b)**.

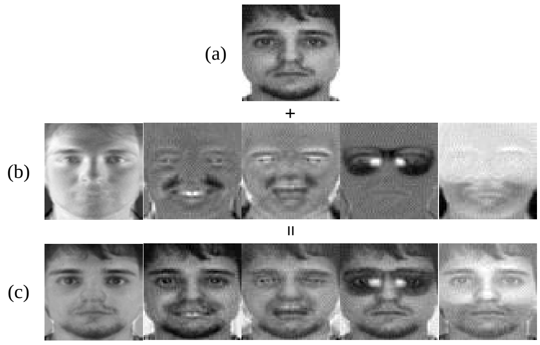


Fig. 2. Illustration of the last step of the kNNVSG method.

3.2 Illumination Process

Since facial appearance variations caused by illumination variations are more severe than those caused by the difference of identity, most existing face recognition methods are sensitive to illumination variations [32]. Weber's law suggests that for a stimulus, the ratio between the smallest perceptual change and the background is a constant, which implies stimuli are perceived not in absolute terms but in relative terms [33]. Given a face image, for each pixel we can compute a ratio between two terms: one is the relative intensity difference of the current pixel against its neighbors; the other is the intensity of the current pixel. Furthermore, we can obtain ratio image based on the ratio, which can extract the local salient patterns very well from input image and is an illumination-insensitive representation of input image. Wang et al. call the ratio image as Weber-face [31].

Algorithm 1. Weber-face algorithm.

Input: 2D face image $F(x,y) \in \Re^{m \times n}$.

Algorithm procedures:

(1) Smoothen $F(x,y)$ with a Gaussian filter to obtain $\hat{F}(x,y)$, i.e., $\hat{F}(x,y) = F(x,y) * G(x,y,\sigma)$, where $*$ is the convolution operator, and $G(x,y,\sigma) = \frac{1}{2\pi\sigma^2}\exp\left(-(x^2+y^2)/2\sigma^2\right)$ is the Gaussian kernel function with standard deviation σ .

(2) Process $\hat{F}(x,y)$ with Weber local descriptor to obtain $WF(x,y)$, i.e., $WF(x,y) = WLD(\hat{F}(x,y))$, where $WLD(\cdot)$ is the Weber local descriptor: $WLD(\hat{F}(x,y)) = \arctan\left(\alpha \sum_{i \in A} \sum_{j \in A} \frac{\hat{F}(x,y) - \hat{F}(x - i\Delta x, y - j\Delta y)}{\hat{F}(x,y)}\right)$, in which $A = \{-1, 0, 1\}$, Δx and Δy are positive integers used to indicate the size of local neighborhood of a pixel point.

Output: Weber-face $WF(x,y)$.

In order to alleviate the adverse effect caused by illumination variations on face recognition performance, we employ the Weber-face algorithm proposed by Wang et al. [31] to conduct illumination normalization for the sample images of the extended gallery set $\tilde{X}$. Assume that $F(x, y) \in \Re^{m\times n}$ denotes a 2D face image which needs to conduct illumination normalization. The procedures of Weber-face algorithm are shown in Algorithm 1.

For the extended gallery set $\tilde{X}$, when all sample images are normalized with Weber-face algorithm, we denote the normalized extended gallery set as $\hat{X} = [\hat{X}_1, \hat{X}_2, ..., \hat{X}_N]$, where $\hat{X}_i = [\hat{x}_{i0}, \hat{x}_{i1}, \hat{x}_{i2}, ..., \hat{x}_{iV}]$, $i = 1, 2, ..., N$, $\hat{x}_{iv} \in \Re^d$, $v = 0, 1, ..., V$, $d = m \times n$. **Fig. 3(a)** shows five images with different illumination variations. **Fig. 3(b)** illustrates five normalized images, which are normalized by using Weber-face algorithm, corresponding to the five images shown in **Fig. 3(a)**, respectively.

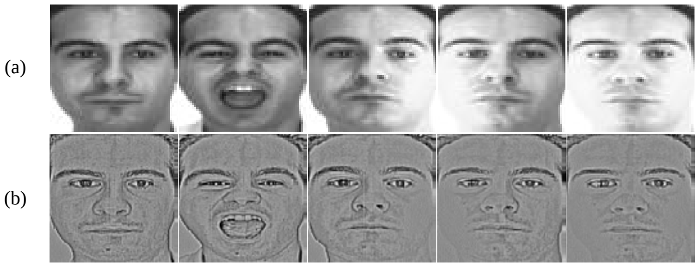


Fig. 3. Five images with different illumination variations and five images normalized by using Weber-face algorithm.

3.3 Image Set based Multimanifold Discriminant Learning (ISMMDL)

3.3.1 Manifold Modeling Procedures of ISMMDL

When a face image is divided into several local patches, they semantically represent different parts (e.g., nose, mouth and eyes) of the original face image, and may not be modeled accurately by a simple distribution [17]. It is more likely that these patches reside in a manifold and each patch corresponds to a point in the manifold [34, 35]. In this paper, we divide all sample images of each class of the normalized extended gallery set into a few non-overlapping local patches. Furthermore, we model all local patches of each class as a manifold and make each local patch corresponding to a point in the manifold. There are two main reasons for us to model all local patches of a class as a manifold. On the one hand, class-specific feature representation and extraction methods usually can achieve better recognition accuracy, which is because class-specific methods employ the information that is especially effective for a given class [36-38]. On the other hand, the optimal feature dimension for each specific class may be different due to the intrinsic differences of different classes. Therefore, we model all local patches of a class as a manifold such that we can optimize an optimal feature dimension for specific class.

The procedures of manifold construction based on image local patches are described in detail as follows.

(1) Assume that the size of one local patch is $a \times b$. For the vth training sample $\hat{x}_{iv} \in \Re^d$ ($d = m \times n$) of the ith class, we can obtain $S = (m \times n)/(a \times b)$ non-overlapping local patches by dividing the 2D image with the size of $m \times n$, which is reconstructed from $\hat{x}_{iv}$.

(2) Similarly, we divide each training sample image of the ith class into S non-overlapping local patches with the size of $a \times b$.

(3) Furthermore, we construct a manifold $M_i = \{\hat{x}_{ivs} \mid v = 0,...,V, s = 1,...,S\}$ by employing the column vectors vectorized from all local patches of the ith class, where $\hat{x}_{ivs} \in \Re^{d_p}$, $d_p = a \times b$. Meanwhile, we make each vector as a point in the manifold.

Similarly, we can construct N manifolds corresponding to N classes of the normalized extended gallery set, and then a manifold set $M = [M_1, M_2, ..., M_N]$ can be obtained, where $M_i = \{\hat{x}_{ivs} \mid v = 0,...,V, s = 1,...,S\}$, $i = 1,...,N$.

3.3.2 Objective Function of ISMMDL

In manifold set M, each manifold is constructed by a number of local patches. The same semantic local patches, such as the eye local patches or the nose local patches, always have a certain extent similarity in gray scale value, texture or directional amplitude features. Therefore, for the multiple manifolds in the original feature space, the distances between local patches of same semantic type in different manifolds are closer than those between local patches of different semantic types in the same manifold. That is, different manifolds are highly overlapped in original feature space [17].

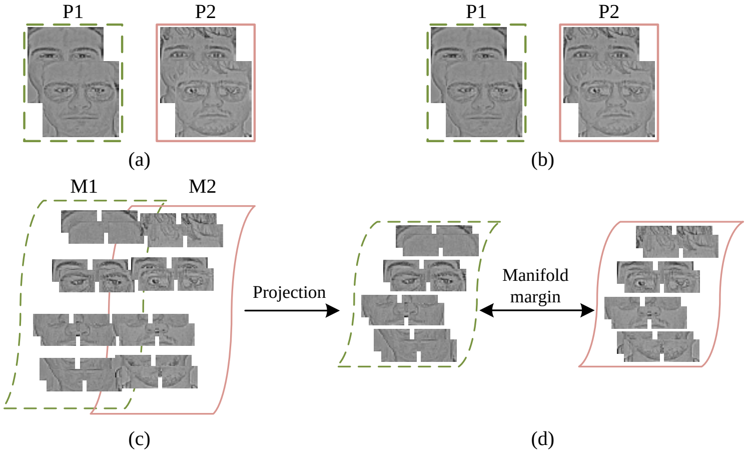


Fig. 4. Illustration of the traits of manifolds in original feature space and the aim of image set based multimanifold discriminant learning (ISMMDL) algorithm.

We illustrate the above phenomenon and the aim of our proposed algorithm ISMMDL using **Fig. 4**. In **Fig. 4(a)** and **Fig. 4(b)**, the images in the green dashed box and the red solid box are face images of two different classes (labeled as P1 and P2 respectively). We divide the images of the two classes into several local patches and model them as manifold M1 and manifold M2 respectively, which are shown in **Fig. 4(c)**. In original feature space, the distance between the eye local patches of manifolds M1 and M2 is small (the same phenomena also exists in nose local patches and mouth local patches of different manifolds). However, the distance between eye local patches and nose local patches (nose local patches and mouth local patches, etc.) in the same manifold (manifold M1 or M2) is large [17]. Hence, the distribution of the local patches of each manifold is not suitable for the classification of manifolds. These phenomena are also illustrated in **Fig. 4(c)**. Therefore, for ISMMDL, we aim to learn a discriminant projection matrix, which is used to construct a low-dimensional feature space, for each manifold (i.e., each class or each person) to better conduct classification. That is, the aim of ISMMDL is that the distances between local patches of different semantic types in the same manifold are minimized and the distances between local patches of the same semantic type in different manifolds are maximized simultaneously in low-dimensional feature space. In other words, the margins of intra-manifold need to be minimized and the margins of inter-manifold need to be maximized simultaneously in the low-dimensional feature space learned by ISMMDL. **Fig. 4(d)** is the illustration of two manifolds obtained by projecting the two original manifolds into the low-dimensional feature space which is constructed by two learned projection matrices corresponding to manifolds M1 and M2 respectively.

Given a local patch sample $\hat{x}_{ivs}$ of the ith manifold M_i, there are usually two types of neighbor in manifolds of manifold set M : (1) intra-manifold neighbors Nei_{intra} , i.e., the local patch samples of Nei_{intra} and the local patch sample $\hat{x}_{ivs}$ belong to the same manifold; (2) inter-manifold neighbors Nei_{inter} , i.e., the local patch samples of Nei_{inter} and the local patch sample $\hat{x}_{ivs}$ belong to different manifolds. From the viewpoint of classification, we need the distances between the nearest neighbors from different manifolds as far as possible and the distances between the nearest neighbors from the same manifold as close as possible in low-dimensional feature space.

In the learning process, we cannot priori determine the significance of the variation feature of each type and that of the variation feature of gallery samples. As a result, we need to learn weights for those variation features in our proposed multimanifold discriminant learning algorithm ISMMDL.

Based on the above analysis, we formulate the objective function of ISMMDL as follows:

$$\max_{W,\theta} \sum_{i=1}^{N}\sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_1}\left(\sum_{p=1}^{k_1}\left\|W_i^T\hat{x}_{ivs}-W_i^T\hat{x}_{ivsp}\right\|^2 A_{ivsp}\right)-\sum_{i=1}^{N}\sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_2}\left(\sum_{q=1}^{k_2}\left\|W_i^T\hat{x}_{ivs}-W_i^T\hat{x}_{ivsq}\right\|^2 B_{ivsq}\right), \quad (1)$$

$$s.t.\sum_{v=0}^{V}\theta_v=1, \theta_v\geq 0$$

where $W=\{W_1,W_2,...,W_N\}$, W contains N matrices learned for each manifold, $\theta=\{\theta_0,\theta_1,...,\theta_V\}$, θ_0 denotes the weight of the variation feature of gallery samples, $\theta_1,\theta_2,...,\theta_V$ denote the weights of variation features of V different types respectively, $\hat{x}_{ivsp}$ denotes the pth inter-manifold nearest neighbor of $\hat{x}_{ivs}$, $\hat{x}_{ivsq}$ denotes the qth intra-manifold nearest neighbor of $\hat{x}_{ivs}$, A_{ivsp} and B_{ivsq} are the elements of two affinity matrices. A_{ivsp} represents the similarity between $\hat{x}_{ivs}$ and $\hat{x}_{ivsp}$, and is used to ensure that if $\hat{x}_{ivs}$ and $\hat{x}_{ivsp}$ are close and from different manifolds, then their low-dimensional representations are separated as far as possible. B_{ivsq} represents the similarity between $\hat{x}_{ivs}$ and $\hat{x}_{ivsq}$, and is utilized to ensure that if $\hat{x}_{ivs}$ and $\hat{x}_{ivsq}$ are close and from the same manifold, then their low-dimensional representations are close as well. The definition of A_{ivsp} and B_{ivsq} are listed as follows:

$$A_{ivsp} = \begin{cases} \exp\left(-\left\|\hat{x}_{ivs} - \hat{x}_{ivsp}\right\|^2\right) \Big/ \sigma_1^{\ 2}, \text{if } \hat{x}_{ivsp} \in Nei_{inter}\left(\hat{x}_{ivs}\right) \\ 0, \qquad\qquad\qquad\qquad \text{otherwise} \end{cases}, \tag{2}$$

$$B_{ivsq} = \begin{cases} \exp\left(-\left\|\hat{x}_{ivs} - \hat{x}_{ivsq}\right\|^2\right) \Big/ \sigma_1^{\ 2}, \text{if } \hat{x}_{ivsq} \in Nei_{intra}\left(\hat{x}_{ivs}\right) \\ 0, \qquad\qquad\qquad\qquad \text{otherwise} \end{cases}. \tag{3}$$

where $Nei_{inter}\left(\hat{x}_{ivs}\right)$ and $Nei_{intra}\left(\hat{x}_{ivs}\right)$ denote the k_1 inter-manifold neighbors and k_2 intra-manifold neighbors of $\hat{x}_{ivs}$, respectively. σ_1 is a scaling parameter that controls how fast the affinity decreases as the distance between two local patch samples increases, and we set σ_1 as the standard deviation of all local patch samples, empirically.

3.3.3 Optimization of the Objective Function of ISMMDL

The unknown variables to be solved in the objective function (1) include two parts, i.e., manifold discriminant projection matrices $W = \{W_1, W_2, ..., W_N\}$ and variation feature weights $\theta = \{\theta_0, \theta_1, ..., \theta_V\}$. To our best knowledge, there is no closed-form solution for the optimization problem defined in (1) as there are N projection matrices and $V+1$ weights to be solved simultaneously. Therefore, in this paper, we apply an alternating optimization approach [39] to solve the problem defined in (1). Specifically, we first initialize θ with valid solution to solve W, and then update θ by fixing W.

In order to facilitate the representation of optimization process, we transform the objective function (1) into

$$\max_{W,\theta} J_1\left(W,\theta\right) - J_2\left(W,\theta\right) \quad s.t. \sum_{v=0}^{V} \theta_v = 1, \theta_v \geq 0, \tag{4}$$

where

$$J_1\left(W,\theta\right) = \sum_{i=1}^{N} \sum_{v=0}^{V} \sum_{s=1}^{S} \frac{\theta_v}{k_1} \left(\sum_{p=1}^{k_1} \left\| W_i^T \hat{x}_{ivs} - W_i^T \hat{x}_{ivsp} \right\|^2 A_{ivsp} \right), \tag{5}$$

$$J_2\left(W,\theta\right) = \sum_{i=1}^{N} \sum_{v=0}^{V} \sum_{s=1}^{S} \frac{\theta_v}{k_2} \left(\sum_{q=1}^{k_2} \left\| W_i^T \hat{x}_{ivs} - W_i^T \hat{x}_{ivsq} \right\|^2 B_{ivsq} \right). \tag{6}$$

(1) Solving W by fixing θ

When we initialize $\theta = \{\theta_0, \theta_1, ..., \theta_V\}$ with valid solution and fix them to solve W, we still cannot solve the N projection matrices of W simultaneously. Hence, we also solve this problem by employing alternating optimization technique. The basic idea is to first initialize $W_1, W_2, ..., W_{i-1}, W_{i+1}, ..., W_N$ with a valid initial solution, and then solve W_i sequentially.

Given θ and $N-1$ projection matrices $W_1, W_2, ..., W_{i-1}, W_{i+1}, ..., W_N$, $J_1\left(W,\theta\right)$ and $J_2\left(W,\theta\right)$ in objective function (4) can be written as

$$\begin{aligned} J_1\left(W_i\right) &= \sum_{v=0}^{V} \sum_{s=1}^{S} \frac{\theta_v}{k_1} \left(\sum_{p=1}^{k_1} \left\| W_i^T \hat{x}_{ivs} - W_i^T \hat{x}_{ivsp} \right\|^2 A_{ivsp} \right) + \sum_{j=1, j\neq i}^{N} \sum_{v=0}^{V} \sum_{s=1}^{S} \frac{\theta_v}{k_1} \left(\sum_{p=1}^{k_1} \left\| W_j^T \hat{x}_{jvs} - W_j^T \hat{x}_{jvsp} \right\|^2 A_{jvsp} \right) \\ &= \sum_{v=0}^{V} \sum_{s=1}^{S} \frac{\theta_v}{k_1} \left(\sum_{p=1}^{k_1} \left\| W_i^T \hat{x}_{ivs} - W_i^T \hat{x}_{ivsp} \right\|^2 A_{ivsp} \right) + C_1 \end{aligned}, \tag{7}$$

$$\begin{aligned} J_2\left(W_i\right) &= \sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_2}\left(\sum_{q=1}^{k_2}\left\|W_i^T\hat{x}_{ivs}-W_i^T\hat{x}_{ivsq}\right\|^2 B_{ivsq}\right)+\sum_{j=1,\,j\neq i}^{N}\sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_2}\left(\sum_{q=1}^{k_2}\left\|W_j^T\hat{x}_{jvs}-W_j^T\hat{x}_{jvsq}\right\|^2 B_{jvsq}\right) \\ &= \sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_2}\left(\sum_{q=1}^{k_2}\left\|W_i^T\hat{x}_{ivs}-W_i^T\hat{x}_{ivsq}\right\|^2 B_{ivsq}\right)+C_2 \end{aligned} , \quad (8)$$

where

$$C_1 = \sum_{j=1,\,j\neq i}^{N}\sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_1}\left(\sum_{p=1}^{k_1}\left\|W_j^T\hat{x}_{jvs}-W_j^T\hat{x}_{jvsp}\right\|^2 A_{jvsp}\right), \quad (9)$$

$$C_2 = \sum_{j=1,\,j\neq i}^{N}\sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_2}\left(\sum_{q=1}^{k_2}\left\|W_j^T\hat{x}_{jvs}-W_j^T\hat{x}_{jvsq}\right\|^2 B_{jvsq}\right). \quad (10)$$

In (9) and (10), C_1 and C_2 are two constant matrices which can be ignored as they do not affect the optimization of W_i. Hence, $J_1\left(W_i\right)$ can be written in the following form:

$$\begin{aligned} J_1\left(W_i\right) &= \sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_1}\left(\sum_{p=1}^{k_1}\left\|W_i^T\hat{x}_{ivs}-W_i^T\hat{x}_{ivsp}\right\|^2 A_{ivsp}\right) \\ &= \sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_1}\sum_{p=1}^{k_1} tr\left(\left(W_i^T\hat{x}_{ivs}-W_i^T\hat{x}_{ivsp}\right)\left(W_i^T\hat{x}_{ivs}-W_i^T\hat{x}_{ivsp}\right)^T A_{ivsp}\right) \\ &= \sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_1}\sum_{p=1}^{k_1} tr\left(W_i^T\left(\hat{x}_{ivs}-\hat{x}_{ivsp}\right)\left(\hat{x}_{ivs}-\hat{x}_{ivsp}\right)^T A_{ivsp}W_i\right) \\ &= tr\left(W_i^T\left(\sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_1}\sum_{p=1}^{k_1}\left(\hat{x}_{ivs}-\hat{x}_{ivsp}\right)\left(\hat{x}_{ivs}-\hat{x}_{ivsp}\right)^T A_{ivsp}\right)W_i\right) = tr\left(W_i^T D_1 W_i\right) \end{aligned} , \quad (11)$$

where $tr(\cdot)$ denotes the trace,

$$D_1 = \sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_1}\sum_{p=1}^{k_1}\left(\hat{x}_{ivs}-\hat{x}_{ivsp}\right)\left(\hat{x}_{ivs}-\hat{x}_{ivsp}\right)^T A_{ivsp}\,. \quad (12)$$

Similarly, we can simplify $J_2\left(W_i\right)$ as follows:

$$\begin{aligned} J_2\left(W_i\right) &= \sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_2}\left(\sum_{q=1}^{k_2}\left\|W_i^T\hat{x}_{ivs}-W_i^T\hat{x}_{ivsq}\right\|^2 B_{ivsq}\right) \\ &= tr\left(W_i^T\left(\sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_2}\sum_{q=1}^{k_2}\left(\hat{x}_{ivs}-\hat{x}_{ivsq}\right)\left(\hat{x}_{ivs}-\hat{x}_{ivsq}\right)^T B_{ivsq}\right)W_i\right) = tr\left(W_i^T D_2 W_i\right) \end{aligned} , \quad (13)$$

where

$$D_2 = \sum_{v=0}^{V}\sum_{s=1}^{S}\frac{\theta_v}{k_2}\sum_{q=1}^{k_2}\left(\hat{x}_{ivs}-\hat{x}_{ivsq}\right)\left(\hat{x}_{ivs}-\hat{x}_{ivsq}\right)^T B_{ivsq}\,. \quad (14)$$

In summary, the objective function of W_i can be simplified as follows:

$$\max_{W_i} J\left(W_i\right) = tr\left(W_i^T D_1 W_i\right) - tr\left(W_i^T D_2 W_i\right) = tr\left(W_i^T\left(D_1 - D_2\right)W_i\right). \quad (15)$$

The optimal solution of (15) is equivalent to the solution of the following eigenvalue equation [17]:

$$\left(D_1 - D_2\right)w = \lambda w\,. \quad (16)$$

Let $\{w_1, w_2, ..., w_{d_i}\}$ be the eigenvectors corresponding to the d_i largest eigenvalues $\{\lambda_1, \lambda_2, ..., \lambda_{d_i}\}$ ordered in such a way that $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_{d_i}$. Then, $W_i = [w_1, w_2, ..., w_{d_i}]$ is the optimal solution of projection matrix W_i.

Here, we discuss how to determine the feature dimension d_i for the ith projection matrix W_i. Because there are N feature dimension numbers corresponding to N projection matrices of N different manifolds to be determined, that is, there are $\Pi_{i=1}^{N} d_i$ candidates to be searched to determine the N optimal feature dimension numbers. Hence, it is time-communing if we empirically select the optimal feature dimensions for each projection matrix. According to the practice of literature [17], we adopt the automatic dimension determination method to determine the optimal feature dimensions by analyzing the eigenvalues of $(D_1 - D_2)$. That is, we select the eigenvectors whose corresponding eigenvalues are greater than zero to construct projection matrix W_i.

(2) Solving θ by fixing W

When we fix the projection matrices $W_1, W_2, ..., W_N$, the original objective function (1) is written as:

$$\max_{\theta} J(\theta) = \sum_{v=0}^{V} \theta_v tr\left(F_{1v} - F_{2v}\right) \quad s.t. \sum_{v=0}^{V} \theta_v = 1, \theta_v \geq 0\,, \tag{17}$$

where

$$F_{1v} = \sum_{i=1}^{N} \sum_{s=1}^{S} \frac{1}{k_1} \sum_{p=1}^{k_1} \left(W_i^T \left(\hat{x}_{ivs} - \hat{x}_{ivsp}\right)\left(\hat{x}_{ivs} - \hat{x}_{ivsp}\right)^T A_{ivsp} W_i \right), \tag{18}$$

$$F_{2v} = \sum_{i=1}^{N} \sum_{s=1}^{S} \frac{1}{k_2} \sum_{q=1}^{k_2} \left(W_i^T \left(\hat{x}_{ivs} - \hat{x}_{ivsq}\right)\left(\hat{x}_{ivs} - \hat{x}_{ivsq}\right)^T B_{ivsq} W_i \right). \tag{19}$$

Having solved θ of (17), we find that the solution to θ in (17) is $\theta_v = 1$ corresponding to the maximum $tr(F_{1v} - F_{2v})$ over different variation features, and $\theta_v = 0$ otherwise. This solution indicates that we do not take full advantage of the multiple facial variation features of generated virtual samples. Such an optimal solution violates the motivation of the ISMMDL algorithm. To address the problem, we modify θ_v to be $(\theta_v)^a$ in objective function (17), where $a > 1$, and the new objective function of (17) is defined as

$$\max_{\theta} J(\theta) = \sum_{v=0}^{V} (\theta_v)^a tr(F_{1v} - F_{2v}) \quad s.t. \sum_{v=0}^{V} \theta_v = 1, \theta_v \geq 0\,. \tag{20}$$

We solve (20) by employing Lagrangian multiplier method. First, we construct a Lagrange function as follows:

$$L(\theta, \lambda) = \sum_{v=0}^{V} (\theta_v)^a tr(F_{1v} - F_{2v}) - \lambda \left(\sum_{v=0}^{V} \theta_v - 1 \right). \tag{21}$$

Then, let $\dfrac{\partial L(\theta, \lambda)}{\partial \theta_v} = 0$ and $\dfrac{\partial L(\theta, \lambda)}{\partial \lambda} = 0$, we have

$$a(\theta_v)^{a-1} tr(F_{1v} - F_{2v}) - \lambda = 0\,, \tag{22}$$

$$\sum_{v=0}^{V} \theta_v - 1 = 0\,. \tag{23}$$

Combining (22) and (23), we can obtain

$$\theta_v = \frac{\left(1/tr(F_{1v} - F_{2v})\right)^{1/(a-1)}}{\sum_{v=0}^{V}\left(1/tr(F_{1v} - F_{2v})\right)^{1/(a-1)}} . \tag{24}$$

3.4 Classification Scheme for Query sample

In this section, we describe the classification scheme for query sample in detail, which mainly contains three procedures.

(1) Illumination normalization of query sample and manifold construction

Given a query sample $z \in \Re^d$ ($d = m \times n$), we deal it with the following procedures. The first step is transforming z into a 2D image with the size of $m \times n$. The second step is conducting illumination normalization on the 2D image to obtain a normalized 2D image $\hat{z} \in \Re^{m \times n}$ by employing Weber-face algorithm [31]. The third step is dividing $\hat{z}$ into $S = (m \times n)/(a \times b)$ non-overlapping local patches with the size of $a \times b$. The fourth step is representing each local patch as a vector, and modeling all vectors as a manifold $M_z = \{\hat{z}_s \in \Re^{d_p} \mid s = 1,...,S\}$ where $d_p = a \times b$.

(2) Computing the distances between the manifold of query sample and the manifolds in manifold set M

In order to label the query sample z, the most significant procedure is computing the distance between the manifold M_z and the manifold M_i in low-dimensional feature space. In this paper, we compute the distance between two manifolds based on the distances between point and nearest neighbor points. The distance between M_z and M_i is defined as follows:

$$d(M_z, M_i) = \frac{1}{S}\sum_{s=1}^{S}\frac{\theta_v}{k_3}\sum_{\hat{x}_{ivs} \in Nei(\hat{z}_s)} \left\| W_i^T \hat{z}_s - W_i^T \hat{x}_{ivs} \right\|^2 , \tag{25}$$

where $Nei(\hat{z}_s)$ denotes the nearest local patches in manifold M_i of the local patch $\hat{z}_s$, and $Nei(\hat{z}_s)$ includes k_3 nearest local patches. As can be seen from (25), the distance $d(M_z, M_i)$ between two manifolds is the mean distance of the distances between each local patch $\hat{z}_s$ in manifold M_z and the nearest local patches in manifold M_i.

(3) Obtaining the class label of query sample

We can obtain the distances $\{d(M_z, M_i)\}_{i=1}^{N}$ between the manifold M_z and N manifolds $M_1, M_2, ..., M_N$ by employing the (25) defined in procedure (2) above. Then, the class label of query sample z can be achieved by the equation as follows:

$$label(z) = \arg\min_i \{d(M_z, M_i)\} . \tag{26}$$

That is, the class label of query sample z is the class label of the manifold which has the minimal distance to the manifold M_z.

3.5 Our SSFR Approach kNNMMDL

Algorithm 2. kNNMMDL algorithm.

Input: gallery set $X=[x_1,x_2,...,x_N]$, generic training set $G=[G_1,G_2,...,G_J]$, query sample z.

Algorithm Procedures:

(1) Generating virtual samples to obtain extended gallery set $\tilde{X}=[\tilde{X}_1,\tilde{X}_2,...,\tilde{X}_N]$ by using kNNVSG algorithm based on gallery set X and generic training set G.

(2) Conducting illumination normalization for the sample images in the extended gallery set $\tilde{X}$ to obtain the normalized extended gallery set $\hat{X}=[\hat{X}_1,\hat{X}_2,...,\hat{X}_N]$, which is a illumination-insensitive representation of $\tilde{X}$ and can be used to alleviate the illumination problem in face recognition , by utilizing the Weber-face algorithm [31].

(3) Solving manifold discriminant projection matrices $W=\{W_1,W_2,...,W_N\}$ and variation feature weights $\theta=\{\theta_0,\theta_1,...,\theta_V\}$ by employing ISMMDL algorithm and the normalized extended gallery set $\hat{X}$.

(4) Computing $\{d(M_z,M_i)\}_{i=1}^{N}$ by using (25).

Output: $label(z)=\arg\min_i\{d(M_z,M_i)\}$.

In this section, we introduce how to comprehensively use our proposed two algorithms, i.e., kNNVSG and ISMMDL algorithms, and the Weber-face algorithm proposed by Wang et al. [31] to formulate a SSFR approach. We call our SSFR approach as k nearest neighbor virtual image set based multimanifold discriminant learning (kNNMMDL) which is detailedly described in Algorithm 2.

4. Experiments

We conduct experimental evaluation on three public datasets, which are AR [40], Multi-PIE [41] and LFW [42] datasets. Some example face images on the three datasets are shown in **Fig. 5**. We compare our kNNMMDL approach with the following methods of three different categories. (1) Virtual sample generating methods: extension of singular-value-perturbed version of PCA (SPCA+) [20] and singular value decomposition-based Fisher linear discriminant analysis (SVD-FLDA) [21]. (2) Generic learning methods: local generic representation (LGR) [14], adaptive generic learning (AGL) [23], extended SRC (ESRC) [26], collaborative probabilistic labels (CPL) [13] and sparse variation dictionary learning (SVDL) [28]. (3) Image partitioning based methods: discriminative multimanifold analysis (DMMA) [17], patch based CRC (PCRC) [18] and block linear discriminative analysis (Block LDA) [19]. We report the best recognition accuracy [21] of each method for different experimental scenarios. The recognition accuracy is defined as the ratio between the number of query samples which are correctly classified to the true classes and the number of all query samples. The higher the recognition accuracy is, the better performance the method owns.

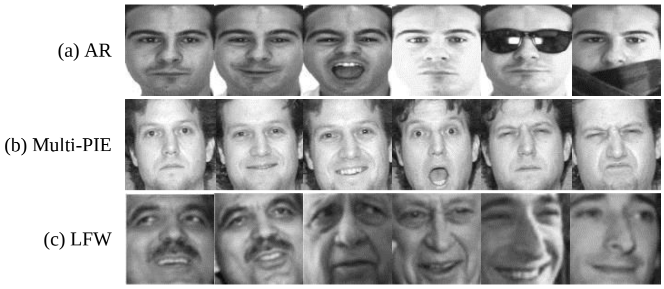


Fig. 5. Example face images of three datasets.

4.1 Parameter Settings

For all contrast methods, we carefully adjust their parameters so that they can achieve the optimal results. For SPCA+, parameters θ, α and n are set to 0.95, 0.25 and 1.25, respectively. For SVD-FLDA, the first three singular values and corresponding singular vectors are used to construct the virtual samples which are used to calculate within-class scatter matrix. For LGR, the regularization parameter λ is fixed as 0.001, and the kernel function parameter σ is set adaptively as stated in the literature [14]. For CPL, the parameter λ in collaborative representation is set to 0.01, and the parameters in group sparse coding are set following the suggestions in [13]. For SVDL, following the parameter settings of literature [28], parameters λ_1, λ_2 and λ_3 are set to 0.001, 0.01 and 0.0001 respectively; and we initialize the number of dictionary atoms as 400. For all image partitioning based methods, such as DMMA, PCRC, Block LDA and our proposed approach kNNMMDL, the corresponding patch size is set as 20×20 empirically. For DMMA, parameters k_1, k_2, k and σ are set to 15, 5, 4 and 100, respectively. For PCRC, the optimal regularization parameter is chosen from {0.0005,0.001,0.005,0.01}. Since AGL and Block LDA are sensitive to feature dimensions, we report the optimal results corresponding to different feature dimensions. For our proposed approach kNNMMDL, parameters k_0, k_1, k_2, k_3 and σ_1 are set to 3, 20, 5, 4 and 100, respectively.

4.2 Experiments on AR Dataset

The AR face dataset [40] contains about 4000 color face images of 126 people (70 males and 56 females), which consists of the frontal faces with different facial expressions, illuminations and disguises (glasses and scarf). There are two sessions and each session has 13 face images per object. Following the SSFR experiment setting in literature [26], a subset with face images of 50 males and 50 females is selected for experiments. The face images in the subset are cropped to the size of 80×80. We select the face images with non-illumination and natural expression of 80 objects in session 1 to construct the gallery set. All images (with different expressions, illuminations and disguises) in session 2 are used to construct the query set. Furthermore, in order to evaluate the robustness of our approach to illumination, expression and disguise, we split the query set into four subsets (i.e., Illumination subset, Expression subset, Disguise subset and Illumination+Disguise subset) to observe the face recognition performance of each subset respectively. We use the images of the remaining 20 objects in session 1 to build the generic training set, where the images of the 20 objects with non-illumination and natural expression are used to construct reference image set and the remainder images of the 20 objects are used to build variation image set.

Table 1 tabulates the recognition accuracies of all methods on the four different query subsets mentioned above. As shown in **Table 1**, our approach kNNMMDL achieves the best performance on all the four different query subsets. Specifically, our kNNMMDL approach outperforms the LGR method by 0.8%(=98.3%-97.5%), 7.5%(=92.5%-85.0%), 1.3%(=95.1%-93.8%) and 2.8%(=91.6%-88.8%) on illumination query subset, expression query subset, disguise query subset, and illumination+disguise query subset respectively. When the intra-class variations of a query sample image do not exist in the original single gallery sample image of the class of the query sample image, the intra-class variations of the query sample image cannot be predicted, which leads to the face recognition performance decrease. In DMMA method, since the DMMA method does not specifically handle the illumination, expression and occlusion variations, it achieves poor performance especially on the Disguise query subset and Illumination+Disguise query subset. The SPCA+ and SVD-FLDA methods achieve poor performance on all the four query subsets. This is because the gallery sample images which are used to train prediction models are images with non-illumination and natural expression and do not have the variation features of illumination, expression and disguise. Besides, there is no specific process to supplement the variation features used to predict the intra-class variations of query samples. Note that because the expressions and disguises (i.e., glasses and scarf) can be well handled by patch based methods, the face recognition performance of patch based methods such as LGR and our kNNMMDL are relatively competitive in this experiment. The other reasons that our kNNMMDL approach can achieve optimal face recognition performance are the following two aspects. On the one hand, using the Weber-face algorithm to normalize the illumination of sample images can alleviate the adverse effect caused by illumination variations on face recognition performance. On the other hand, employing generic training set to enrich the intra-class variation information for training samples can enable our proposed approach handling the intra-class variations which do not exist in the original single gallery sample of each class.

Table 1. Recognition accuracies (%) on four query subsets of AR dataset (bold numbers indicate the best results).

Method	**Illumination**	**Expression**	**Disguise**	**Illumination +Disguise**
SPCA+[20]	37.5	55.2	26.9	22.5
SVD-FLDA[21]	32.5	57.1	24.4	16.3
LGR[14]	97.5	85.0	93.8	88.8
AGL[23]	70.8	55.8	40.6	30.7
ESRC[26]	87.9	70.4	59.4	45.0
CPL[13]	95.7	88.3	71.6	69.3
SVDL[28]	87.1	74.2	61.3	54.1
PCRC[18]	88.8	71.7	81.8	63.1
Block LDA[19]	54.7	61.2	31.9	21.0
DMMA[17]	77.9	61.7	28.1	21.9
kNNMMDL	**98.3**	**92.5**	**95.1**	**91.6**

4.3 Experiments on Multi-PIE Dataset

The Multi-PIE dataset [41] contains more than 750000 images from 337 individuals, which are captured under 15 viewpoints and 19 illumination conditions in up to four recording sessions. Among the 337 people, 129 people have image acquisition in four sessions. In our experiments, images are cropped to the size of 80×80. We select frontal images in session 2 to construct the gallery set and generic training set. Concretely, we use images with non-illumination and natural expression of 100 objects to build the gallery set. And we use images of the remaining 103 objects to construct the generic training set, where images with non-illumination and natural expression are used as reference images, and images with natural, surprising and squint expressions under f06, f07 and f08 illumination conditions are used to construct variation set. Frontal images in session 3 and session 4 under f06, f07 and f08 illumination conditions are selected to construct four query sets. Specifically, we choose images under those illumination conditions separately with natural, smile and disgust expressions in session 3 and images under those illumination conditions with screaming expression in session 4 to construct the four query sets.

Table 2 tabulates the recognition accuracies of each method on the four query sets of session 3 and session 4. As can be seen from **Table 2**, our approach achieves distinctly better performance than the other methods on the four query sets. From the experimental results, we find that our approach kNNMMDL still achieves good face recognition performance for the images under f06, f07 and f08 illumination conditions in the four query sets even though the images in gallery set are images with non-illumination. This is because: (1) the generated virtual samples can enrich the illumination variations for training samples; (2) the illumination normalization for the sample images with illumination variations can alleviate the negative influence for face recognition performance. Meanwhile, we find that although our approach does not generate virtual samples to enrich disgust and screaming expression variations for training samples, it still achieves good face recognition performance on S3-Disgust and S4-Scream query sets. The reasons are that there are imges with surprising and squinting expressions in generic training set and those imges can be used to generate virtual samples to enrich the expression variations of surprising and squinting for training samples. Significantly, the expressions of squinting and surprising are separately similar to the expressions of disgust and screaming. In contrast, the face recognition performance is relatively poor in S3-Smile query set, which is because there do not exist imges with smile expression to be used for generating virtual samples to enrich the expression variation of smile for training samples. For CPL, it achieves a comparable recognition accuracies except in S3-Smile, S3-Disgust and S4-Scream query sets. This is because CPL is based on the assumption that generic training set is full enough with sufficient variations. However, there do not exist the intra-class variations of Smile, Disgust and Scream in the generic training set which is used in the experiments.

Table 2. Recognition accuracies (%) on four query sets of Multi-PIE dataset (bold numbers indicate the best results).

Method	**S3-Neutral**	**S3-Smile**	**S3-Disgust**	**S4-Scream**
SPCA+[20]	48.4	36.7	40.6	38.5
SVD-FLDA[21]	54.3	43.2	48.1	45.9
LGR[14]	91.8	81.6	86.3	84.1
AGL[23]	82.4	52.8	57.7	56.1
ESRC[26]	86.2	68.4	72.5	71.2
CPL[13]	92.6	70.5	77.2	76.3
SVDL[28]	88.5	72.4	76.8	74.9
PCRC[18]	74.2	67.5	71.5	70.9
Block LDA[19]	58.8	47.7	51.1	50.9
DMMA[17]	62.5	55.3	60.3	59.2
kNNMMDL	**93.6**	**84.1**	**88.4**	**87.7**

4.4 Experiments on LFW Dataset

The LFW dataset [42] contains more than 13000 images from 5749 different individuals in unconstrained environments, where 1680 people have two or more than two images per person. We use LFW-a [43] which is a version of LFW after alignment using commercial face alignment software in the experiment. One can see that although face alignment has been conducted, the intra-class variations in this dataset are still very large compared with the face datasets in the controlled environments. Following the experiment setting in literatures [18] and [28], a subset with face images of 158 objects is selected for experiment, and each object has more than 10 images. We resize the size of the images to 80×80. We select face images of the first 50 objects to construct the gallery set and the query set, and images of the remaining objects are used to build the generic training set. The mean face image of each object is used to construct the reference sample set in the generic training set because of the absence of frontal images with natural expression in this dataset. Since face images in the LFW dataset are captured in the unconstrained environments, the type of intra-class variations in each face image is uncertain. However, there always exist face images with smile expression or the expression similar to smile in face images of each object. Therefore, we only select face images with smile expression or the expression similar to smile for constructing the variation set.

Table 3 lists the recognition accuracies of all competeing methods. It can be seen from **Table 3** that all methods do not achieve very high face recognition performance. That is because images in the LFW dataset are collected in uncontrolled environments, which makes face images containing rich intra-class variations and increases the difficulty for face recognition. As a result, the performance of face recognition deteriorate. Nevertheless, our proposed approach is still superior to the other comparing methods. The reasons are the following two aspects. On the one hand, we handle illumination by employing Weber-face algorithm in our approach, which can alleviate the adverse influence caused by illumination variations for face recognition performance. On the other hand, the generated virtual samples can enrich the intra-class variation information for training samples to a certain extent. LGR and SVDL achieve good face recognition performance as well because the learned intra-class variation information from other objects of the generic training set can help improving the robustness of SSFR.

Table 3. Recognition accuracies (%) on the query set of LFW dataset (bold numbers indicate the best results).

Method	Accuracy
SPCA+[20]	14.9
SVD-FLDA[21]	15.5
LGR[14]	30.4
AGL[23]	19.2
ESRC[26]	27.3
CPL[13]	25.2
SVDL[28]	28.6
PCRC[18]	24.2
Block LDA[19]	16.4
DMMA[17]	17.8
kNNMMDL	**32.3**

4.5 The Effect of Virtual Sample Generating for kNNMMDL

In our proposed approach kNNMMDL, the aim of generating virtual samples by using the kNNVSG algorithm is to enrich the intra-class variation information of training samples (i.e., gallery samples) in the gallery set. When the kNNVSG algorithm is not used to generate virtual samples, the intra-class variation information of training samples in the gallery set is relatively simple. In the following, we verify whether the intra-class variation information contained in the virtual sample images can predict the intra-class variations of query samples and enhance the face recognition performance of the kNNMMDL approach. We conduct experiments in two cases, i.e., with and without using the kNNVSG algorithm in kNNMMDL approach, to observe the differences of the face recognition performance. Specifically, we separately conduct experiments on four query sets (Illumination, Expression, Disguise and Illumination+Disguise) of the AR dataset to observe their face recognition performance in the two cases mentioned above.

Table 4. Recognition accuracies (%) of kNNMMDL with and without using kNNVSG algorithm on four query sets of AR dataset (bold numbers indicate the best results).

Method	Illumination	Expression	Disguise	Illumination +Disguise
kNNMMDL (without kNNVSG)	88.6	70.7	37.9	34.8
kNNMMDL (with kNNVSG)	**98.3**	**92.5**	**95.1**	**91.6**

Table 4 lists the recognition accuracies of the kNNMMDL approach with and without using the kNNVSG algorithm in kNNMMDL approach on the four query sets of the AR dataset. As shown in **Table 4**, generating virtual samples by using the kNNVSG algorithm can improve the recognition accuracies on Illumination, Expression, Disguise and Illumination Disguise query sets by 9.7% (= 98.3% - 88.6%), 21.8% (= 92.5% - 70.7%), 57.2% (= 95.1% - 37.9%) and 56.8% (= 91.6% - 34.8%) as compared with the recognition results of doing not use the kNNVSG algorithm in kNNMMDL approach, respectively. It indicates that generating virtual samples by using the kNNVSG algorithm to enrich the intra-class variation information for the training samples is helpful to SSFR. Similar phenomena exist on Multi-PIE and LFW datasets as well.

4.6 The Effect of Illumination Normalization for kNNMMDL

In our proposed approach kNNMMDL, the aim of conducting illumination normalization on sample images by using Weber-face algorithm is to alleviate the adverse influence caused by illumination variations for face recognition performance. Illumination normalization procedure is a relatively independent procedure in kNNMMDL approach. Hence, it can be removed from kNNMMDL approach when we observe whether the illumination normalization procedure is beneficial to alleviate the adverse influence caused by illumination variations for the face recognition performance. Concretely, for comparison purposes, we conduct experiments on four query sets (Illumination, Expression, Disguise and Illumination+Disguise) of the AR dataset with and without performing illumination normalization respectively.

Table 5. Recognition accuracies (%) of kNNMMDL with and without performing illumination normalization on four query sets of AR dataset (bold numbers indicate the best results).

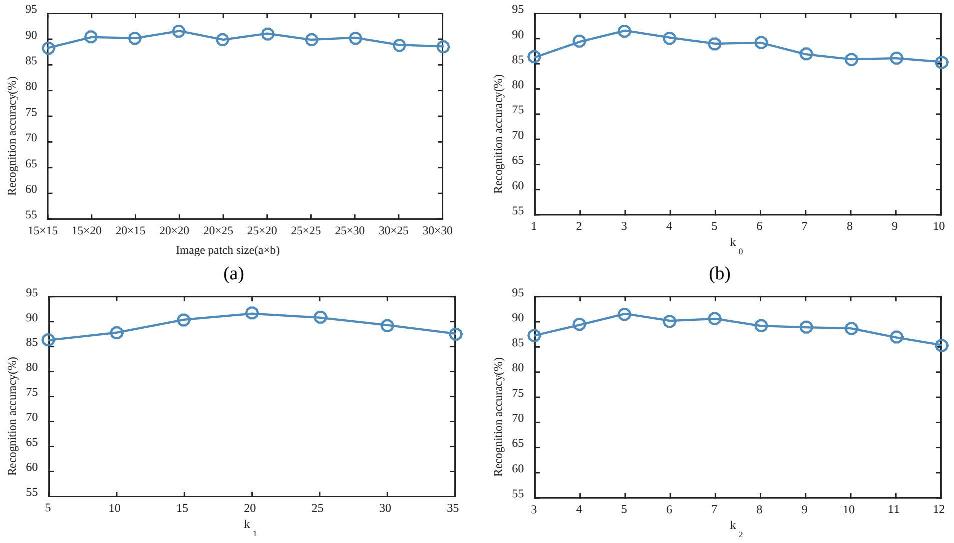
Method	Illumination	Expression	Disguise	Illumination +Disguise
kNNMMDL (without normalization)	89.1	92.1	94.9	80.8
kNNMMDL (with normalization)	**98.3**	**92.5**	**95.1**	**91.6**

Table 5 tabulates the recognition accuracies of the kNNMMDL approach on the four query sets of the AR dataset with and without performing illumination normalization respectively. From **Table 5**, it can be seen that the face recognition performance of kNNMMDL with performing illumination normalization are significantly better than those without performing illumination normalization, and the performance improvement of 9.2% (=98.3%-89.1%) and 10.8% (=91.6%-80.8%) on Illumination and Illumination+Disguise query sets can be obtained respectively. It indicates that we can enrich the illumination variations for training samples by generating virtual samples to better predict illumination variations of query samples. Besides, conducting illumination normalization to obtain the illumination-insensitive representations of sample images is helpful to alleviate the adverse effect caused by illumination variations on face recognition performance as well. Furthermore, the experimental results on Expression and Disguise query sets illustrate that the illumination normalization procedure basically does not loss useful discriminant information used for face recognition. Similar phenomena also exist on Multi-PIE and LFW datasets.

4.7 Parameter Analysis

The influence of key parameters of kNNMMDL for face recognition performance is studied. These key parameters include the size of the local patch $a \times b$, the size of nearest neighbor set of single gallery sample k_0, the size of the inter-manifold nearest neighbor set k_1 and the size of the intra-manifold nearest neighbor set k_2. Since each parameter could affect the face recognition accuracy, we should first fix the other three parameters when we test the effect of one parameter on the face recognition accuracy in the experiments.

Fig. 6 illustrates how these four parameters affect the face recognition accuracy of our approach. **Fig. 6(a)**, **Fig. 6(b)**, **Fig. 6(c)** and **Fig. 6(d)** show the face recognition accuracy variations separately versus different local patch size, different size of nearest neighbor set of single gallery sample, different size of inter-manifold nearest neighbor set and different size of intra-manifold nearest neighbor set on AR dataset when using Illumination+Disguise as the query set. As shown in **Fig. 6**, our proposed approach has relatively stable performance for these four parameters $a \times b$, k_0, k_1 and k_2. Therefore, in order to achieve good face recognition performance, it is relatively easy to select appropriate value for these four parameters.



(a) (b)

(c) (d)

Fig. 6. Parameter analysis of kNNMMDL. Face recognition accuracy versus (a) varying image patch size $a \times b$, (b) varying size of nearest neighbor of single gallery sample k_0, (c) varying size of inter-manifold nearest neighbor set k_1, and (d) varying size of intra-manifold nearest neighbor set k_2 on AR dataset.

5. Conclusion

In this paper, we propose a face recognition approach called k nearest neighbor virtual image set based multimanifold discriminant learning (kNNMMDL) to address the SSFR problem. In kNNMMDL approach, based on the idea that similar faces have similar intra-class variations, we propose k nearest neighbors based virtual sample generating (kNNVSG) algorithm to enrich the intra-class variation information of training samples. Aiming to use the intra-class variation information of virtual samples for better learning low-dimensional feature space, we propose image set based multimanifold discriminant learning (ISMMDL) algorithm. Besides, we introduce Weber-face algorithm to alleviate adverse influence caused by illumination variations to the face recognition performance. Experimental results on three widely used face datasets (i.e., AR, Multi-PIE and LFW datasets) illustrate that our proposed face recognition approach kNNMMDL is effective for SSFR tasks.

References

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: A literature survey,” *ACM Computing Surveys*, vol.35, no.4, pp.399-458, 2003. Article (CrossRef Link).

[2] G. B. Huang, V. Jain, and E. Learned-Miller, “Unsupervised joint alignment of complex images,” in *Proc. of the IEEE International Conference on Computer Vision*, pp.1-8, October 14-20, 2007. Article (CrossRef Link).

[3] L. Wolf, T. Hassner, and Y. Taigman, “Effective unconstrained face recognition by combining multiple descriptors and learned background statistics,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.33, no.10, pp. 1978-1990, 2011. Article (CrossRef Link).

[4] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, “Attribute and simile classifiers for face verification,” in *Proc. of the IEEE International Conference on Computer Vision*, pp.365-372, September 27 - October 4, 2009. Article (CrossRef Link).

[5] Q. Yin, X. Tang, and J. Sun, “An associate-predict model for face recognition,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.497-504, June 20-25, 2011. Article (CrossRef Link).

[6] A. J. O'Toole, P. J. Phillips, F. Jiang, J. Ayyad, N. Penard, and H. Abdi, “Face recognition algorithms surpass humans matching faces over changes in illumination, ” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.29, no.9, pp. 1642-1646, 2007. Article (CrossRef Link).

[7] P. Zhu, M. Yang, L. Zhang, and I. Y. Lee, “Local generic representation for face recognition with single sample per person,” in *Proc. of the Asian Conference on Computer Vision*, pp.34-50, November 1-5, 2014. Article (CrossRef Link).

[8] Y. Lei, Y. Guo, M. Hayat, M. Bennamoun, and X. Zhou, “A two-phase weighted collaborative representation for 3D partial face recognition with single sample,” *Pattern Recognition*, vol.52, pp. 218-237, 2016. Article (CrossRef Link).

[9] T. Pei, L. Zhang, B. Wang, F. Li, and Z. Zhang, “Decision pyramid classifier for face recognition under complex variations using single sample per person,” *Pattern Recognition*, vol.64, pp. 305-313, 2017. Article (CrossRef Link).

[10] Y. F. Yu, D. Q. Dai, C. X. Ren, and K. K. Huang, “Discriminative multi-scale sparse coding for single-sample face recognition with occlusion,” *Pattern Recognition*, vol.66, pp. 302-312, 2017. Article (CrossRef Link).

[11] J. Hu, “Discriminative transfer learning with sparsity regularization for single-sample face recognition, ” *Image and Vision Computing*, vol.60, pp.48-57, 2017. Article (CrossRef Link).

[12] M. Yang, X. Wang, G. Zeng, and L. Shen, “Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person,” *Pattern Recognition*, vol.66, pp.117-128, 2017. Article (CrossRef Link).

[13] H. K. Ji, Q. S. Sun, Z. X. Ji, Y. H. Yuan, and G. Q. Zhang, “Collaborative probabilistic labels for face recognition from single sample per person,” *Pattern Recognition*, vol. 62, pp. 125-134, 2017. Article (CrossRef Link).

[14] P. Zhang, X. You, W. Ou, C. P. Chen, and Y. M. Cheung, “Sparse discriminative multi-manifold embedding for one-sample face identification,” *Pattern Recognition*, vol. 52, pp. 249-259, 2016. Article (CrossRef Link).

[15] A. M. Martínez, “Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no .6, pp. 748-763, 2002. Article (CrossRef Link).

[16] X. Tan, S. Chen, Z. H. Zhou, and F. Zhang, “Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble,” *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp.875-886, 2005. Article (CrossRef Link).

[17] J. Lu, Y. P. Tan, and G. Wang, “Discriminative multimanifold analysis for face recognition from a single training sample per person,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 39-51, 2013. Article (CrossRef Link).

[18] P. Zhu, L. Zhang, Q. Hu, and S. C. Shiu, “Multi-scale patch based collaborative representation for face recognition with margin distribution optimization,” in *Proc. of the European Conference on Computer Vision*, pp. 822-835, October 7-13, 2012. Article (CrossRef Link).

[19] S. Chen, J. Liu, and Z. H. Zhou, “Making FLDA applicable to face recognition with one sample per person,” *Pattern Recognition*, vol. 37, no. 7, pp. 1553-1555, 2004. Article (CrossRef Link).

[20] D. Zhang, S. Chen, and Z. H. Zhou, “A new face recognition method based on SVD perturbation for single example image per person,” *Applied Mathematics and Computation*, vol. 163, no. 2, pp. 895-907, 2005. Article (CrossRef Link).

[21] Q. X. Gao, L. Zhang, and D. Zhang, “Face recognition using FLDA with single training image per person,” *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 726-734, 2008. Article (CrossRef Link).

[22] T. Vetter, “Synthesis of novel views from a single face image, ” *International Journal of Computer Vision*, vol. 28, no. 2, pp. 103-116, 1998. Article (CrossRef Link).

[23] Y. Su, S. Shan, X. Chen, and W. Gao, “Adaptive generic learning for face recognition from a single sample per person,” in *Proc. of the Conference on Computer Vision and Pattern Recognition*, pp. 2699-2706, June 13-18, 2010. Article (CrossRef Link).

[24] S. Si, D. Tao, and B. Geng, “Bregman divergence-based regularization for transfer subspace learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 7, pp. 929-942, 2010. Article (CrossRef Link).

[25] B. Wang, W. Li, Z. Li, and Q. Liao, “Adaptive linear regression for single-sample face recognition, ” *Neurocomputing*, vol. 115, no. 4, pp. 186-191, 2013. Article (CrossRef Link).

[26] W. Deng, J. Hu, and J. Guo, “Extended SRC: Undersampled face recognition via intraclass variant dictionary,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 9, pp. 1864-1870, 2012. Article (CrossRef Link).

[27] J. Hu, J. Lu, X. Zhou, and Y. P. Tan, “Discriminative transfer learning for single-sample face recognition,” in *Proc. of the International Conference on Biometrics*, pp.272-277, September 8-11, 2015. Article (CrossRef Link).

[28] M. Yang, L. Van Gool, and L. Zhang, “Sparse variation dictionary learning for face recognition with a single training sample per person,” in *Proc. of the IEEE International Conference on Computer Vision*, pp.689-696, December 1-8, 2013. Article (CrossRef Link).

[29] L. Zhang, M. Yang, and X. Feng, “Sparse representation or collaborative representation: Which helps face recognition?” in *Proc. of the IEEE International Conference on Computer Vision*, pp.471-478, November 6-13, 2011. Article (CrossRef Link).

[30] S. Gao, K. Jia, L. Zhuang, and Y. Ma, “Neither global nor local: regularized patch-based representation for single sample per person face recognition, ” *International Journal of Computer Vision*, vol.111, no.3, pp.365-383, 2015. Article (CrossRef Link).

[31] B. Wang, W. Li, W. Yang, and Q. Liao, “Illumination normalization based on Weber's law with application to face recognition,” *IEEE Signal Processing Letters*, vol. 18, no. 8, pp. 462-465, 2011. Article (CrossRef Link).

[32] Y. Adini, Y. Moses, and S. Ullman, “Face recognition: The problem of compensating for changes in illumination direction,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.19, no.7, pp.721-732, 1997. Article (CrossRef Link).

[33] A. K. Jain, “Fundamentals of digital signal processing,” *Fundamentals of Digital Signal Processing*, 1989.

[34] S. T. Roweis, and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” *Science*, vol. 290, no. 5500, pp. 2323-2326, 2000. Article (CrossRef Link).

[35] J. B. Tenenbaum, V. D. Silva, and J. C. Langford, “A global geometric framework for nonlinear dimensionality reduction,” *Science*, vol.290, no.5500, pp.2319-2323, 2000. Article (CrossRef Link).

[36] R. Gross, I. Matthews, and S. Baker, “Generic vs. person specific active appearance models,” *Image and Vision Computing*, vol. 23, no. 12, pp. 1080-1093, 2005. Article (CrossRef Link).

[37] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, and B. L. Lu, “Person-specific SIFT features for face recognition,” in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.593-596, April 15-20, 2007. Article (CrossRef Link).

[38] B. Yao, A. I. Haizhou, and S. Lao, “Person-specific face recognition in unconstrained environments: a combination of offline and online learning,” in *Proc. of the IEEE International Conference on Automatic Face and Gesture Recognition*, pp.1-8, September 17-19, 2008. Article (CrossRef Link).

[39] S. Yan, J. Liu, X. Tang, and T. S. Huang, “A parameter-free framework for general supervised subspace learning,” *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 1, pp. 69-76, 2007. Article (CrossRef Link).

[40] A. M. Martinez and R. Benavente, “The AR face database, ” *CVC Technical Report 24*, Barcelona, Spain, June 1998.

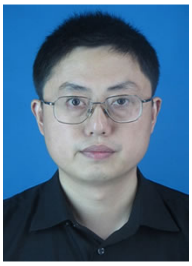
[41] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, “Multi-PIE,” *Image and Vision Computing*, vol. 28, no. 5, pp. 807-813, 2010. Article (CrossRef Link).

[42] G. B. Huang, Ramesh, T. Berg, and E. Learned-Miller, “Labeled faces in the wild: A database for studying face recognition in unconstrained environments,” *Technical Report 07-49*, Amherst, USA, October 2007.

[43] L. Wolf, T. Hassner, and Y. Taigman, “Similarity scores based on background samples, ” in *Proc. of the Asian Conference on Computer Vision*, pp. 88-97, September 23-27, 2009. Article (CrossRef Link).

Xiwei Dong is now a student with the College of Automation, Nanjing University of Posts and Telecommunications, and a visiting PhD student with the State Key Laboratory of Software Engineering, School of Computer, Wuhan University, China. His research interest includes pattern recognition, computer vision and machine learning.

Fei Wu is a lecturer in the College of Automation, Nanjing University of Posts and Telecommunications, China. He received his PhD from Nanjing University of Posts and Telecommunications in 2016. His research interest includes pattern recognition, machine learning and software engineering.



Xiao-Yuan Jing is now a Professor with the State Key Laboratory of Software Engineering, School of Computer, Wuhan University, and the College of Automation, Nanjing University of Posts and Telecommunications, China. He has published over 70 papers in the international conferences and journals like international conference on Computer Vision and Pattern Recognition (CVPR), International Joint Conference Artificial Intelligence (IJCAI), AAAI conference on Artificial Intelligence (AAAI) and IEEE Transactions on Image Processing.