Discriminative Multi-Manifold Analysis for Face Recognition from a Single Training Sample per Person

Jiwen Lu[1], Yap-Peng Tan[1], Gang Wang[1,2]
[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]Advanced Digital Sciences Center, Singapore
lujiwen@pmail.ntu.edu.sg; eyptan@ntu.edu.sg; wanggang@ntu.edu.sg

Abstract

Conventional appearance-based face recognition methods usually assume there are multiple samples per person (MSPP) available during the training phase for discriminative feature extraction. In many practical face recognition applications such as law enhancement, e-passport and ID card identification, this assumption, however, may not hold as there is only a single sample per person (SSPP) enrolled or recorded in these systems. Many popular face recognition methods fail to work well in this scenario because there are not enough samples for discriminant learning. To address this problem, we propose in this paper a novel discriminative multi-manifold analysis (DMMA) method by learning discriminative features from image patches. First, we partition each enrolled image into several non-overlapping patches to form an image set for each sample per person. Then, we formulate the SSPP face recognition as a manifold-manifold matching problem and learn multiple DMMA feature spaces to maximize the manifold margins of different persons. Lastly, we propose a reconstruction-based manifold-manifold distance to identify the unlabeled subjects. Experimental results on three widely used face databases are presented to demonstrate the efficacy of the proposed approach.

1. Introduction

Appearance-based methods have been widely used in face recognition, and a large number of such algorithms have been proposed in recent years [1, 10, 28, 33]. The common goal of these methods is to learn a compact and low-dimensional feature subspace for face representation in a supervised, semi-supervised or unsupervised manner, such that the intrinsic characteristics of the original face samples are well preserved. Representative and popular algorithms include principal component analysis (PCA) [28], linear discriminant analysis (LDA) [1], locality preserving projections (LPP) [10], marginal fisher analysis (MFA) [33], and their weighted, kernelized and tensorized variants [16, 17, 35]. Despite the different assumptions of these methods, they can be unified into a general graph embedding (GE) framework [33] under different constraints.

The performance of the appearance-based methods in face recognition, however, is heavily affected by the number of training samples per person [11]. Specifically, if the number of training samples per person is much smaller than the feature dimension of face samples, it is generally inaccurate to estimate the within-class variance of LDA and exploit discriminant and geometrical information of most existing manifold learning algorithms [4, 18] for face recognition. In many practical face recognition applications, such as law enhancement, e-passport and ID card identification, there is usually only a single sample per person (SSPP) recorded in these systems because it is generally difficult to collect additional samples under these scenarios. Therefore, many existing appearance-based methods such as LDA and its variants [35] cannot be directly applied for feature extraction due to the lack of samples to compute the within-class scatter.

To address the SSPP problem in face recognition, there have been some attempts in the literature, which can be mainly classified into three categories [27]: generic learning, virtual sample generation, and image partitioning. For the first category, an additional generic training set with multiple samples per person (MSPP) is applied to extract discriminative features, which are then used to identify the persons each enrolled with a single sample. For example, Kim and Kitter [14] employed a generic training set to learn discriminative features to address the SSPP problem in pose-invariant face recognition. Wang *et al.* [31] proposed a general generic learning framework and presented several appearance-based discriminant feature extraction methods under this framework. While discriminative information can be exploited, there is one common assumption among these methods that the inter-class and intra-class variations of the MSPP generic training set and the SSPP gallery set

are similar, which may not hold in many real world applications because it is not easy to collect a MSPP generic training set which well represents all the gallery sets. To address this issue, Su *et al.* [25] proposed an adaptive generic learning method to infer the discriminative information of the SSPP gallery set by using a prediction model learned from the generic training set. Si *et al.* [24] proposed a transfer subspace learning approach to transfer the discriminative model learned from the MSPP training set to the SSPP gallery set for face recognition. Even if there are some improvements on addressing the SSPP face recognition, the performance of these methods is heavily affected by the selected generic training set, which is still very difficult to obtain in practical applications.

For the second category, some additional training samples for each person are virtually generated such that discriminant subspace learning can be used for feature extraction. For example, Zhang *et al.* [36] and Gao *et al.* [8] presented two singular value decomposition (SVD) based perturbation algorithms to obtain multiple images for each person and then applied the conventional LDA for feature extraction. While these methods can alleviate the SSPP problem to a certain extent, one common shortcoming among these methods is that there is high correlation among the virtually generalized samples as they cannot be considered as independent samples for feature extraction, and this may result in much redundancy in the learned discriminative feature subspace, as stated in [19].

For the third approach, each face image is first partitioned into several local patches and then some discriminant learning techniques can be applied for feature extraction. An early such attempt [20] divided each face image into six elliptical parts and learned a local probabilistic model for recognition. Tan *et al.* [26] extended this work by proposing an alternative way of representing each face subspace with self-organizing maps (SOM). Chen *et al.* [5] employed LDA and Kanan *et al.* [13] presented a weighted pseudo-Zernike moment to extract discriminative features of each local patch, respectively. However, these methods ignore the geometrical information of the local patches in the feature extraction procedure. Specially, when a face image is partitioned into several local patches, they represent different parts (semantics) of the original face image, such as nose, mouth and eyes, and may not be modeled accurately by a simple distribution. It is more likely that these patches reside in a manifold and each patch corresponds to a point in the manifold. Motivated by this observation, we consider the local patches of each person as a manifold, and the SSPP face recognition can be formulated as a manifold-manifold matching problem, as illustrated in Fig. 1.

Given a (frontal) face image, there are usually two eyes, one nose and one mouth in the image. Intuitively speaking, the similarity between two noses of different persons

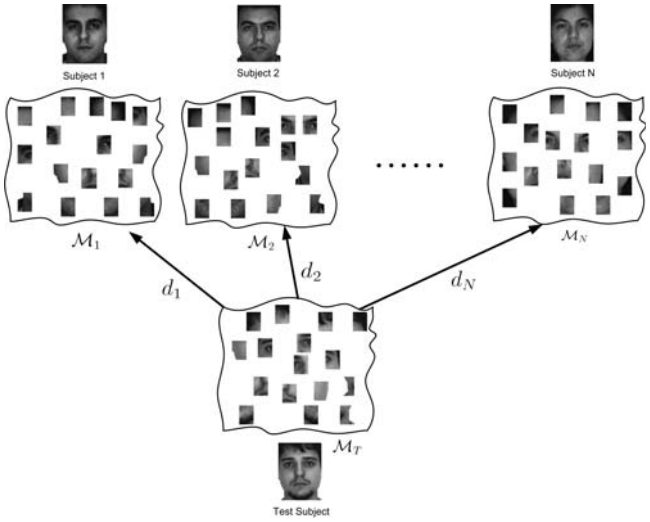

Fig. 1. Manifold-manifold matching for SSPP face recognition. Each image is partitioned into many local patches and represented by a manifold and the SSPP face recognition is converted into a manifold-manifold matching problem.

is usually higher than that between the nose and the mouth of the same person. Hence, there is high overlapping between these manifolds and the distances between patches at the same location of different images (persons) are usually smaller than those at different locations of the same image (person). Hence, a key challenge to address the SSPP problem is how to extract discriminative features such that the points in the same manifold become closer and those in different manifolds are far apart. To achieve this goal, we propose in this paper a novel discriminative multi-manifold analysis (DMMA) method to learn the local discriminant features to maximize the manifold margins of different persons, so that more discriminant information can be exploited for recognition. In the recognition phase, we propose a reconstruction-based manifold-manifold distance to identify the unlabeled subjects. Experimental results on three widely used face databases are presented to demonstrate the efficacy of the proposed approach.

The rest of the paper is organized as follows. Section 2 details the proposed approach. Section 3 provides the experimental results and Section 4 concludes the paper.

2. Proposed Approach

Let $X = [x_1, x_2, \cdots, x_N]$ be the training set, x_j is the training image of the jth person with a size of $m \times n$, $1 \leq j \leq N$, N is the number of persons in the training set. We first divide each image x_j into t non-overlapping local patches with an equal size of $a \times b$, where $t = \frac{m \times n}{a \times b}$. Let $M_i = [x_{i1}, x_{i2}, \cdots, x_{it}]$ be the image patch set of the ith person, which consists of a manifold $\mathcal{M}_i$.

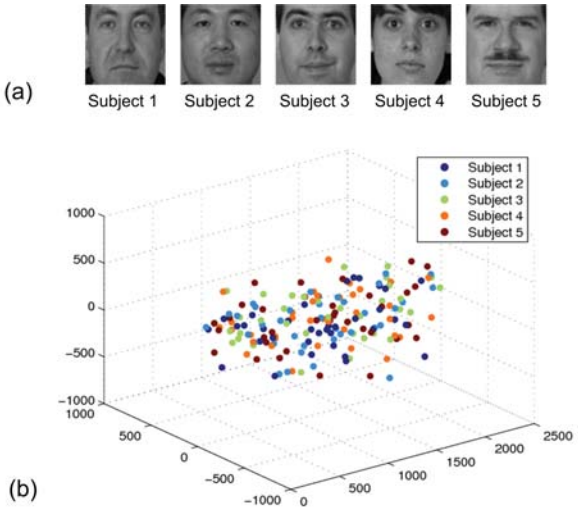

Fig. 2. Visualization of five manifolds corresponding to five persons from the FERET database.

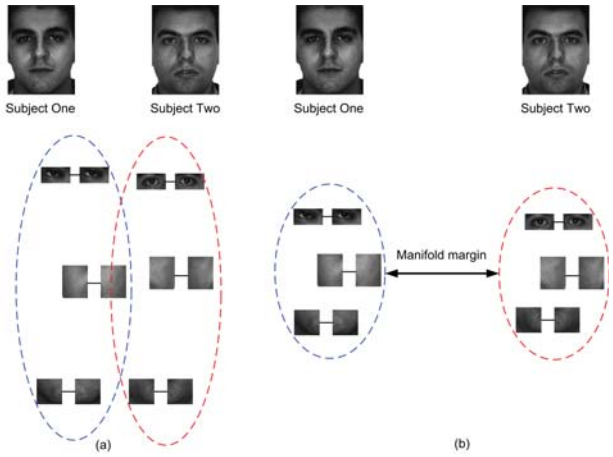

Fig. 3. Illustration of the proposed DMMA method. (a) Three pairs of patches of two subjects in the original high-dimensional space. (b) Desired distribution in the low-dimensional feature space.

2.1. Motivation

To explore the geometrical information of the partitioned local patches, we randomly selected five subjects from the FERET database [22] to visualize the manifold structure of these local patches. Fig. 2(a) shows five 60×60 well-aligned face images, each is partitioned into 36 10×10 non-overlapping patches. The distribution of all the image patches of the five subjects is shown in Fig. 2(b), in a three-dimensional space for ease of presentation. We can see that there is high overlapping among the manifolds corresponding to different persons. We can also see that there is large variation on the appearance of intra-subject patches.

As we mentioned above, the similarity between the same sematic patches of different subjects is usually higher than that of different sematic patches of the same subject. This is illustrated from Fig. 3(a), where the similarity of two eyes from two different subjects is usually higher than that of an eye and a cheek from the same person. Hence, in the original image patch space, different sematic patches of the same subject are well separated while the same sematic patches of different subjects are clustered, which is the main challenge in our task. We aim to learn some mappings to project the patches in the same manifold to be close and those in different manifolds to be far apart, respectively. Fig. 3(b) shows the desired result of the proposed method, where local patches of different subjects are well separated while those of the same subject become closer after the mappings. As a result, the manifold margin in the low-dimensional feature space is much larger and more discriminative information can be exploited for recognition. Yet most existing discriminative manifold learning methods assume that the samples from different classes define a single manifold in the feature space and seek a common feature space for all samples so that the ratio of between-class locality variance to within-class locality variance is maximized [4]. However, these methods may fail to uncover the geometrical structure of the data sampled from multiple manifolds which are of possibly different dimensionalities. This is because the manifolds corresponding to different classes may not have the same intrinsic dimensionality. In our proposed approach, we model the local patches of each subject as a manifold so that they can be better separated when the feature dimensions are selected to be different in the low-dimensional feature spaces. With this, we do away the assumption that each sample must have the same feature dimension and aim to seek multiple projections to project the data into different feature spaces to better separate them.

2.2. DMMA

Let $\mathcal{M} = [\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_N]$ be the training set of N persons and $\mathcal{M}_i = [x_{i1}, x_{i2}, \cdots, x_{it}]$ be the manifold of the ith person, $x_{ij} \in R^d$, $1 \leq i \leq N$, $1 \leq j \leq t$. The aim of DMMA is to seek N feature matrices $W_1, W_2, \cdots, W_N$, $W_i \in R^{d \times d_i}$, $i = 1, 2, \cdots, N$, to project the training set into N low-dimensional feature spaces, such that the manifold margins are maximized.

To the best of our knowledge, a formal definition of manifold margin has not been defined. In the literature, there have been a few attempts to compute the distance between two manifolds. For example, Wang *et al.* [30] proposed a maximal linear patch (MLP) clustering method to partition each manifold into several clusters, modeled each cluster with a subspace, and then computed the nearest subspace distance as the manifold distance. To address the unbalanced cluster issue in the MLP method, Wang and Chen [29] further proposed a hierarchical divisive clustering

method to better approximate the manifold by several subspaces, and then learned a discriminant function for feature extraction. While these methods can calculate the manifold distance approximately, they also assumed that all data are sampled from one single manifold and only one projection is derived for feature extraction. In our proposed approach, we aim to seek multiple projection matrices to uncover the geometrical information of these manifolds and better characterize the manifold margin in the low-dimensional feature spaces.

Given a sample x_{ij}, which is the jth patch of the ith manifold, there are usually two kinds of neighbors in these manifolds: intra-manifold neighbors N_{intra} and inter-manifold neighbors N_{inter}. From the viewpoint of classification, we aim to minimize the intra-manifold variance and maximize the inter-manifold separability in the low-dimensional feature spaces, simultaneously, so that the manifold margin can be maximized for feature extraction. To achieve this goal, we formulate the proposed DMMA as the following optimization problem:

$$\begin{aligned}
&\max_{W_1,W_2,\cdots,W_N} J(W_1,W_2,\cdots,W_N)\\
=\;& J_1(W_1,W_2,\cdots,W_N)-J_2(W_1,W_2,\cdots,W_N)\\
=\;& \sum_{i=1}^{N}\left(\sum_{r=1}^{t}\sum_{p=1}^{k_1}\|W_i^Tx_{ir}-W_i^Tx_{irp}\|^2A_{irp}\right)\\
&-\sum_{i=1}^{N}\left(\sum_{r=1}^{t}\sum_{q=1}^{k_2}\|W_i^Tx_{ir}-W_i^Tx_{irq}\|^2B_{irq}\right) \quad (1)
\end{aligned}$$

where x_{irp} represents the pth k_1-nearest inter-manifold neighbors and x_{irq} denotes the qth k_2-nearest intra-manifold neighbors of x_{ir}, respectively, A_{irp} and B_{irq} are two affinity matrices to characterize the similarity between x_{ir} and x_{irp} as well as x_{irq}, respectively. While many graph construction methods have been proposed in the machine learning and computer vision community recently [12], we apply the conventional k-nearest-neighbor method, for its high effectiveness and efficiency, to calculate the affinity matrices A and B, as follows:

$$A_{irp}=\begin{cases}\exp\left(-\|x_{ir}-x_{irp}\|^2/\sigma^2\right), & \text{if } x_{irp}\in N_{inter}^{k_1}(x_{ir})\\ 0, & \text{otherwise}\end{cases} \quad (2)$$

$$B_{irq}=\begin{cases}\exp\left(-\|x_{ir}-x_{irq}\|^2/\sigma^2\right), & \text{if } x_{irq}\in N_{intra}^{k_2}(x_{ir})\\ 0, & \text{otherwise}\end{cases} \quad (3)$$

where $N_{inter}^{k_1}(x_{ir})$ and $N_{intra}^{k_2}(x_{ir})$ denote the k_1-inter-manifold neighbors and k_2-intra-manifold neighbors of x_{ir}, respectively; k_1, k_2 and σ are three empirically pre-specified parameters.

The objective function of J_1 in Eq. (1) is to ensure that if x_{ir} and x_{irp} are close and from different subjects, then their low-dimensional representations are separated as far as possible. On the other hand, J_2 in Eq. (1) ensures that if x_{ir} and x_{irq} are close and from the same subject, then their low-dimensional representations are close as well.

To the best of our knowledge, there is no closed-form solution for the optimization problem defined in Eq. (1) as there are N projection matrices to be obtained simultaneously. We solve this problem in an iterative manner as inspired by recent advances in high order tensor decomposition [16, 33]. The basic idea is to first initialize $W_1,W_2,\cdots,W_{i-1},W_{i+1},\cdots,W_N$ with a valid initial solution, and then solve W_i sequentially.

Given $W_1,W_2,\cdots,W_{i-1},W_{i+1},\cdots,W_N$, Eq. (1) can be rewritten as

$$\begin{aligned}
&\max_{W_i} J(W_i)\\
=\;& (J_1(W_i)+F_1)-(J_2(W_i)+F_2)\\
=\;& \Big(\sum_{r=1}^{t}\sum_{p=1}^{k_1}\|W_i^Tx_{ir}-W_i^Tx_{irp}\|^2A_{irp}+F_1\Big)\\
-\;& \Big(\sum_{r=1}^{t}\sum_{q=1}^{k_2}\|W_i^Tx_{ir}-W_i^Tx_{irq}\|^2B_{irq}+F_2\Big) \quad (4)
\end{aligned}$$

where $F_1=\sum_{j=1,j\neq i}^{N}\Big(\sum_{r=1}^{t}\sum_{p=1}^{k_1}\|W_j^Tx_{jr}-W_j^Tx_{jrp}\|^2A_{jrp}\Big)$ and $F_2=\sum_{j=1,j\neq i}^{N}\Big(\sum_{r=1}^{t}\sum_{q=1}^{k_2}\|W_j^Tx_{ir}-W_j^Tx_{jrq}\|^2B_{jrq}\Big)$ are two constant matrices and which can be ignored as they don't affect the optimization of W_i.

After some algebraic manipulations, we can simplify $J_1(W_i)$ to the following form

$$\begin{aligned}
&J_1(W_i)\\
=&\sum_{r=1}^{t}\sum_{p=1}^{k_1}\|W_i^Tx_{ir}-W_i^Tx_{irp}\|^2A_{irp}\\
=&\sum_{r=1}^{t}\sum_{p=1}^{k_1}tr(W_i^Tx_{ir}-W_i^Tx_{irp})(W_i^Tx_{ir}-W_i^Tx_{irp})^TA_{irp}\\
=&\sum_{r=1}^{t}\sum_{p=1}^{k_1}tr\left(W_i^T[(x_{ir}-x_{irp})(x_{ir}-x_{irp})^TA_{irp}]W_i\right)\\
=&tr\left(W_i^T[\sum_{r=1}^{t}\sum_{p=1}^{k_1}(x_{ir}-x_{irp})(x_{ir}-x_{irp})^TA_{irp}]W_i\right)\\
=&tr\left(W_i^TH_1W_i\right) \quad (5)
\end{aligned}$$

where

$$H_1\triangleq\sum_{r=1}^{t}\sum_{p=1}^{k_1}(x_{ir}-x_{irp})(x_{ir}-x_{irp})^TA_{irp} \quad (6)$$

Input: Manifold $\mathcal{M}_1, \mathcal{M}_2, \cdots, \mathcal{M}_N$, $\mathcal{M}_i = [x_{i1}, x_{i2}, \cdots, x_{it}]$, $x_{ij} \in R^d$, $i = 1, \cdots, N$, $j = 1, 2, \cdots, t$, values of the parameters k_1, k_2 and σ, iteration number T, and convergence error ε.

Output: Projection matrices $W_i \in R^{d \times d_i}$, $i = 1, 2, \cdots, N$.

Algorithm:

Step 1 (Initialization):

Set $W_i^0 = I_{d \times d}$, $i = 1, 2, \cdots, N$.

Step 2 (Similarity matrices calculation):

For each sample x_{ij}, calculate two affinity matrices A and B as shown in Eqs. (2) and (3), respectively.

Step 3 (Local optimization):

For $r = 1, 2, \cdots, T$, repeat

3.1. Compute H_1 and H_2 as shown in Eqs. (6) and (8), respectively.

3.2. Solve the eigenvalue problem defined in Eq. (9).

3.3. Sort their eigenvectors $[w_1, w_2, \cdots, w_{d_i}]$ according to their associated eigenvalues: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{d_i}$.

3.4. Obtain $W_i^r = [w_1, w_2, \cdots, w_{d_i}]$. If $r > 2$ and $|W_i^r - W_i^{r-1}| < \varepsilon$, go to Step 4.

Step 4 (Output projection matrices):

Output projection matrices $W_i = W_i^r$, $i = 1, 2, \cdots, N$.

Fig. 4. Proposed DMMA algorithm.

Similarly, we can simplify $J_2(W_i)$ as follows

$$\begin{aligned} & J_2(W_i) \\ =& \sum_{r=1}^{t} \sum_{q=1}^{k_2} \|W_i^T x_{ir} - W_i^T x_{irq}\|^2 B_{irq} \\ =& tr\left(W_i^T [\sum_{r=1}^{t} \sum_{q=1}^{k_2} (x_{ir} - x_{irq})(x_{ir} - x_{irq})^T B_{irq}] W_i \right) \\ =& tr\left(W_i^T H_2 W_i \right) \end{aligned} \tag{7}$$

where

$$H_2 \triangleq \sum_{r=1}^{t} \sum_{q=1}^{k_2} (x_{ir} - x_{irq})(x_{ir} - x_{irq})^T B_{irq} \tag{8}$$

Having obtained H_1 and H_2, we can obtain the bases of W_i by solving the following eigenvalue equation:

$$(H_1 - H_2)w = \lambda w \tag{9}$$

Let $\{w_1, w_2, \cdots, w_{d_i}\}$ be the eigenvectors corresponding to the d_i largest eigenvalues $\{\lambda_j | j = 1, 2, \cdots, d_i\}$ ordered in such a way that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{d_i}$. Then $W = [w_1, w_2, \cdots, w_{d_i}]$ is the projection matrix of W_i. We can iteratively and sequentially solve Eq. (4) to determine the N projection matrices W_1, W_2, $\cdots$, W_N. The proposed DMMA algorithm summarized in Fig. 4.

Now, we discuss how to determine the feature dimension d_i for the ith projection matrix W_i. Most existing manifold learning algorithms [2, 3, 4, 10, 33] empirically select the optimal feature dimension because there is only one projection matrix to be sought for feature extraction. In our case, there are N feature projection matrices corresponding to N different manifolds. Hence, it is non trivial to determine the optimal feature dimensions because there are $\prod_{i=1}^{N} d_i$ candidates to be searched and it is time-communing, if not impossible, to find the best one. We propose a new feature dimension determination method by analyzing the eigenvalues of $(H_1 - H_2)$. Since $(H_1 - H_2)$ is not positive semi-definite, the eigenvalues of $(H_1 - H_2)$ may be positive, zero and negative. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{d_i} > 0 \geq \lambda_{d_{i+1}} \geq \cdots \geq \lambda_d$ and W_i be ordered according to their corresponding eigenvalues, We can select the first d_i eigenvectors to maximize Eq. (4). This is because when the samples are projected to one specific eigenvector w_j corresponding to an eigenvalue λ_j, Eq. (4) can be written as

$$\begin{aligned} J(w_j) &= J_1(w_j) - J_2(w_j) \\ &= w_j^T (H_1 - H_2) w_j \\ &= w_j^T \lambda_j w_j \\ &= \lambda_j \end{aligned} \tag{10}$$

If $\lambda_j > 0$, this means that the inter-manifold distance is larger than the intra-manifold distance along the direction of w_j, and samples can likely be correctly classified. According to this criterion, we can automatically determine the feature dimension d_i of each projection W_i.

Lastly, we briefly analyze the computational complexity of the DMMA method, which involves T iterations, and each iteration solves a generalized eigenvalue equation. Let N be the number of face samples in the training set, each image patch is $d_2 = a * b$, and t is the number of local patches of each image. Assume $k = \min\{N * t, d_2\}$, the computational complexity of our proposed DMMA method is $O(NTk^3)$.

2.3. Recognition

Given a test sample T, we first partition it into t non-overlapping local patches and model it as a manifold $\mathcal{M}_T = [x_{T1}, x_{T2}, \cdots, x_{Tt}]$. We then assign a label c to $\mathcal{M}_T$ as follows:

$$c = \arg\min_i d(\mathcal{M}_T, \mathcal{M}_i), \quad i = 1, 2, \cdots, N. \tag{11}$$

where $d(\mathcal{M}_T, \mathcal{M}_i)$ is the manifold distance between $\mathcal{M}_T$ and $\mathcal{M}_i$.

Different from existing manifold distance methods [29, 30], which calculate the nearest subspace-subspace distance to approximate the manifold-manifold distance, we propose here a novel reconstruction-based method to compute the

manifold distance by using all rather than selected samples in the two manifolds. The main reason is that there is high correlation among the samples used in existing manifold-manifold distance methods, and a subspace can well approximate their variations in each manifold. However, the samples used in our case are local patches of an image in the same manifold and we will lose some patch information if not all the patches are used to calculate the subspace distance.

Let $\mathcal{Y}_i = W_i^T \mathcal{M}_i = [y_{i1}, y_{i2}, \cdots, y_{it}]$ and $\mathcal{Y}_T = W_i^T \mathcal{M}_T = [y_{T1}, y_{T2}, \cdots, y_{Tt}]$ be the low-dimensional representations of manifolds $\mathcal{M}_i$ and $\mathcal{M}_T$. The manifold distance between them is defined as follows:

$$d(\mathcal{M}_T, \mathcal{M}_i) = \frac{1}{t} \sum_{j=1}^{t} d\left(y_{tij}, G_k(y_{tij})\right) \tag{12}$$

where $G_k(y_{tij})$ denotes the k-nearest neighbors of y_{tij} in $\mathcal{Y}_i$ and $d\left(y_{tij}, G_k(y_{tij})\right)$ can be easily obtained by solving the following constrained optimization problem, similar to the locally linear embedding method as discussed in [23], and described below as

$$d\left(y_{tij}, G_k(y_{tij})\right) = \min_{\sum_{s=1}^{k} c_s = 1} \|y_{tij} - \sum_{s=1}^{k} c_s G_s(y_{tij})\|^2 \tag{13}$$

where c_s is the reconstruction coefficient of the neighbor $G_s(y_{tij})$ to y_{tij}.

3. Experiments

We have evaluated the proposed DMMA method by conducting a number of SSPP face recognition experiments on three widely used face databases. The following describes the details of the experiments and results.

3.1. Datasets

Three publicly available face databases, namely AR [21], FERET [22] and FG-NET [15], are used for evaluation to demonstrate the efficacy of the proposed method. The AR face database contains over 4000 color face images of 126 people (70 men and 56 women), including frontal views of faces with different facial expressions, lighting conditions, and occlusions (sun glasses and scarves). There are 26 different images per person, taken in two sessions (separated by two weeks), and each session contains thirteen 768×576 color images. In our experiments, eight subsets of 800 images from 100 different subjects (50 men and 50 women) were used, which were taken from two different sessions and with different expressions. Table 1 provides the detailed information of each subset. Fig. 5(a) shows eight sample images of one subject from the Subset A to H of the AR database. In our experiments, we used Subset A for training and the remaining seven subsets for testing.

Table 1. Detailed information of all the eight subsets of the AR database used in our experiments.

Dataset	Collection condition
Subset A	Collected in the first session with neural expression
Subset B	Collected in the first session with smile expression
Subset C	Collected in the first session with anger expression
Subset D	Collected in the first session with scream expression
Subset E	Collected in the second session with neural expression
Subset F	Collected in the second session with smile expression
Subset G	Collected in the second session with anger expression
Subset H	Collected in the second session with scream expression

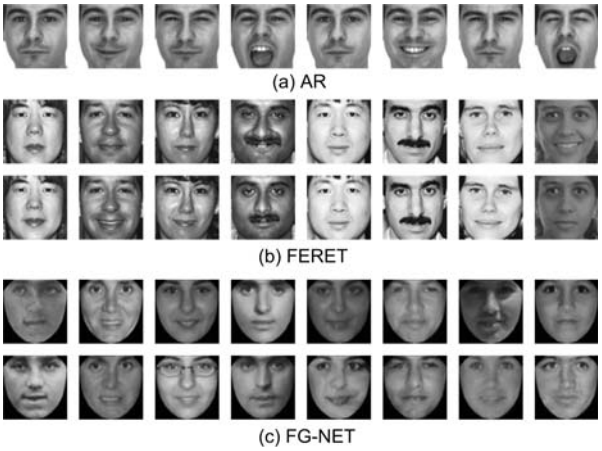

Fig. 5. Sample face images from the (a) AR, (b) FERET and (c) FG-NET database.

The FERET database consists of 13539 facial images corresponding to 1565 subjects, who are diverse across ethnicity, gender and age. Similar to [6, 26, 32, 36], we used a subset of the FERET database including 400 gray-level frontal face images comprising 200 different persons, with the size of 256×384. There are 71 females and 129 males, each of who has two two images (fa and fb), with different races, genders, ages, expressions, illuminations and scales, etc. For this database, we applied fa images for training and fb images for testing. Fig. 5(b) shows some sample images of one subject from the FERET database, where the first row are fa images and the second row are fb images.

The FG-NET database consists of 1002 face images with large variations in lighting, pose, and expression. There are 82 subjects (around 12 images/subject) in total with ages ranging from 0 to 69 years old. In our experiments, we constructed four subsets (Subsets A-D) from the FG-NET database. Each subset contains 82 subjects and there are two images (S_1 and S_2) for each subject. The age gap between S_1 and S_2 for Subsets A to D are $(0, 3]$, $(3, 6]$, $(6, 9]$ and $(9, 12]$, respectively. Fig. 5(c) shows some sample images from the FG-NET database extracted by using the active appearance model features, where the first row are S_1 images and the second row are S_2 images. In our experiments, we used S_1 for training and S_2 for testing.

3.2. Experimental Settings

For all images in the above three datasets, the facial part of each image was manually cropped, aligned and resized into 60×60 pixels according to the eyes' positions. We have compared our method with 11 algorithms which can be used to address the SSPP face recognition problem, including PCA [28], $(PC)^2A$ (projected-combined principal component analysis) [32], $E(PC)^2A$ (enhanced projected-combined principal component analysis) [6], 2DPCA (two-dimensional PCA) [34], $(2D)^2PCA$ (two-directional two-dimensional PCA) [37], SOM [26], LPP [10], SVD-LDA (singular value decomposition-based linear discriminant analysis) [36], Block PCA [9], Block LDA [5] and UP (uniform pursuit) [7]. The nearest neighbor classifier with the Euclidean distance was applied for recognition.

We implemented these methods ourselves and tuned the best parameters for each method for a fair comparison. For $(PC)^2A$, the weighting parameter α was tuned to be 0.25. For $E(PC)^2A$, two parameters α and β were used to weight the projected-combined principal component and they were empirically set to 0.25 and 0.5, respectively. For SVD-LDA, the first 3 singular values and the corresponding singular vectors were applied to construct a virtual sample to calculate the within-class scatter. For the LPP and UP methods, the number of nearest neighbors was selected to be 4 and t was set to be 100 to calculate the similarity matrix. For all block-based methods, such as Block PCA, Block LDA and our proposed DMMA method, the size of each block was empirically tuned to be 10×10. For our DMMA method, the values of the parameters k_1, k_2, k and σ were empirically tuned to be 15, 5, 4 and 100, respectively, and the feature dimensions of these manifolds vary from 15 to 25.

3.3. Results and Analysis

Table 2 tabulates the rank-one recognition rate of different methods on the AR and FERET databases. The best recognition accuracy of each method was recorded for a fair comparison. As can be seen from this table, the proposed DMMA consistently outperforms the 11 compared methods with the lowest gains in accuracy of 1%, 4%, 5%, 2%, 8%, 11% and 3% on Subsets B-H of the AR database, and 2% on the FERET database, respectively.

We have made three observations from the results listed in Table 2: 1) PCA and Block PCA obtain similar performance on both the AR and FERET databases, which indicates that there is no significant difference on the performance of either holistic-based or patch-based feature representation for unsupervised learning on the SSPP face recognition. 2) SVD-LDA obtains the worst performance on the AR database even though it is a supervised method. The reason is that the virtually generated samples by SVD-LDA are highly related to the original sample as the new sample is just an approximation by discarding some smaller singular values of the original image. Hence, the within-class variance cannot be accurately estimated in this scenario. Another reason is that LDA usually obtains better recognition when the number of samples per class is large. However, when the number of sample per class is small, the recognition performance of LDA is poor and even worse than PCA, as also shown in [20]. 3) DMMA obtains the best recognition performance on all the experiments, which implies that both discriminant and geometrical information of local patches are important for recognition.

Table 2. Rank-1 recognition accuracy (%) of different methods on different subsets of the AR and FERET databases.

Method	AR							FERET
	B	C	D	E	F	G	H	
PCA	97	87	60	77	76	67	38	84.0
$(PC)^2A$	97	87	62	77	74	67	40	84.5
$E(PC)^2A$	97	87	63	77	75	68	41	85.5
2DPCA	97	87	60	76	76	67	37	84.5
$(2D)^2PCA$	98	89	60	71	76	66	41	85.0
SOM	98	88	64	73	77	70	42	91.0
LPP	94	87	36	86	74	78	20	84.0
SVD-LDA	73	75	29	75	56	58	19	85.5
Block PCA	97	87	60	77	76	67	38	84.5
Blcok LDA	85	79	29	73	59	59	18	86.5
UP	98	88	59	77	74	66	41	90.0
DMMA	**99**	**93**	**69**	**88**	**85**	**79**	**45**	**93.0**

Table 3 shows the rank-one and rank-10 recognition accuracy obtained by different methods on the four subsets of the FG-NET database. We can observe that the recognition performance of all the methods are generally lower than those obtained in the AR and FERET databases, which implies that aging could affect the performance of the SSPP face recognition more than other factors such as lighting and expression variations. However, we can still see that that our proposed DMMA consistently outperforms other compared methods in terms of both rank-1 and rank-10 recognition accuracies. This is because compared with the holistic-based methods such as PCA, 2DPCA, LPP, and UP, the proposed DMMA extracts features from local patches and the appearance differences caused by aging are mitigated in the local patches and better age-invariant feature can be preserved. Compared with other block-based methods such as Block PCA and Block LDA, DMMA can exploit both discriminant and locality information to improve the recognition performance.

4. Conclusion and Future Work

We have proposed in this paper a novel discriminative multi-manifold analysis (DMMA) method to address the SSPP problem in face recognition. We partition each enrolled image into several non-overlapping patches, and construct an image set for each sample per person, and then learn multiple feature spaces to maximize the manifold margins of different persons. Experimental results on three widely used face databases are presented to demonstrate the

Table 3. Rank-1 and rank-10 recognition accuracies (%) of different methods on different subsets of the FG-NET database.

Method	Rank-1 accuracy				Rank-10 accuracy			
	A	B	C	D	A	B	C	D
PCA	23.2	15.9	11.0	8.5	39.0	32.9	35.4	25.6
$(PC)^2A$	24.4	17.1	11.0	8.5	40.2	34.2	36.6	26.8
$E(PC)^2A$	24.4	15.9	12.2	9.8	41.5	35.4	36.6	26.8
2DPCA	24.4	17.1	13.4	9.8	40.2	34.2	36.6	26.8
$(2D)^2PCA$	24.4	17.1	14.6	11.0	41.5	35.4	37.8	28.1
SOM	25.6	18.3	15.9	13.4	43.9	36.6	39.0	30.5
LPP	23.2	14.6	11.0	8.5	39.0	34.2	35.4	26.8
SVD-LDA	22.0	14.6	11.0	8.5	37.8	30.5	32.9	24.4
Block PCA	28.1	22.0	14.6	12.2	43.9	37.8	39.0	30.5
Blcok LDA	26.8	22.0	13.4	12.2	42.7	36.6	37.8	30.5
UP	24.4	19.5	12.2	11.0	41.5	35.4	36.6	28.1
DMMA	**32.9**	**26.8**	**18.3**	**14.6**	**48.8**	**42.7**	**43.9**	**36.6**

efficacy of the proposed approach. How to extend the proposed DMMA method to MSPP face recognition appears to be another interesting direction of future work.

References

[1] P. N. Belhumeur, J. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[2] D. Cai, X. He, and J. Han. Spectral regression for efficient regularized subspace learning. In *IEEE International Conference on Computer Vision*, pages 1–8, 2007.

[3] D. Cai, X. He, Y. Hu, J. Han, and T. Huang. Learning a spatially smooth subspace for face recognition. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 1–7, 2007.

[4] H. Chen, H. Chang, and T. Liu. Local discriminant embedding and its variants. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 846–853, 2005.

[5] S. Chen, J. Liu, and Z. Zhou. Making FLDA applicable to face recognition with one sample per person. *Pattern Recognition*, 37(7):1553–1555, 2004.

[6] S. Chen, D. Zhang, and Z. Zhou. Enhanced $(PC)^2A$ for face recognition with one training image per person. *Pattern Recognition Letters*, 25(10):1173–1181, 2004.

[7] W. Deng, J. Hu, J. Guo, W. Cai, and D. Feng. Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach. *Pattern Recognition*, 43(5):1748–1762, 2010.

[8] Q. Gao, L. Zhang, and D. Zhang. Face recognition using FLDA with single training image per person. *Applied Mathematics and Computation*, 205(2):726–734, 2008.

[9] R. Gottumukkal and V. Asari. An improved face recognition technique based on modular PCA approach. *Pattern Recognition Letters*, 25(4):429–436, 2004.

[10] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang. Face recognition using Laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):328–340, 2005.

[11] A. Jain and B. Chandrasekaran. Dimensionality and sample size considerations in pattern recognition practice. *Handbook of statistics*, 2:835–855, 1982.

[12] T. Jebara, J. Wang, and S. Chang. Graph construction and b-matching for semi-supervised learning. In *Proceedings of International Conference on Machine Learning*, pages 441–448, 2009.

[13] H. Kanan, K. Faez, and Y. Gao. Face recognition using adaptively weighted patch PZM array from a single exemplar image per person. *Pattern Recognition*, 41(12):3799–3812, 2008.

[14] T. Kim and J. Kittler. Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(3):318–327, 2005.

[15] A. Lanitis. Evaluating the performance of face-aging algorithms. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 1–6, 2008.

[16] H. Lu, K. Plataniotis, and A. Venetsanopoulos. MPCA: Multilinear principal component analysis of tensor objects. *IEEE Transactions on Neural Networks*, 19(1):18–39, 2008.

[17] J. Lu and Y. Tan. A doubly weighted approach for appearance-based subspace learning methods. *IEEE Transactions on Information Forensics and Security*, 5(1):71–81, 2010.

[18] J. Lu and Y.-P. Tan. Regularized locality preserving projections and its extensions for face recognition. *IEEE Transactions on Systems, Man, and Cybernetics: Part B, Cybernetics*, 40(2):958–963, 2010.

[19] A. Martínez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):748–763, 2002.

[20] A. Martínez and A. Kak. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):228–233, 2002.

[21] A. M. Martinez and R. Benavente. The AR face database. Technical report, CVC Technical Report, 1998.

[22] P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.

[23] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

[24] S. Si, D. Tao, and B. Geng. Bregman divergence based regularization for transfer subspace learning. *IEEE Transactions on Knowledge and Data Engineering*, (7):929–942.

[25] Y. Su, S. Shan, X. Chen, and W. Gao. Adaptive generic learning for face recognition from a single sample per person. In *IEEE International Conference on Computer Vision and Pattern Recognition*, pages 2699–2706, 2010.

[26] X. Tan, S. Chen, Z. 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*, 16(4):875–886, 2005.

[27] X. Tan, S. Chen, Z. Zhou, and F. Zhang. Face recognition from a single image per person: A survey. *Pattern Recognition*, 39(9):1725–1745, 2006.

[28] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[29] R. Wang and X. Chen. Manifold Discriminant Analysis. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2009.

[30] R. Wang, S. Shan, X. Chen, and W. Gao. Manifold-manifold distance with application to face recognition based on image set. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[31] X. Wang and X. Tang. Random sampling for subspace face recognition. *International Journal of Computer Vision*, 70(1):91–104, 2006.

[32] J. Wu and Z. Zhou. Face recognition with one training image per person. *Pattern Recognition Letters*, 23(14):1711–1719, 2002.

[33] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1):40–51, 2007.

[34] J. Yang, D. Zhang, A. Frangi, and J. Yang. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1):131–137, 2004.

[35] M. H. Yang. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In *IEEE International Conference on Face and Gesture Recognition*, pages 215–220, 2002.

[36] D. Zhang, S. Chen, and Z. Zhou. A new face recognition method based on SVD perturbation for single example image per person. *Applied Mathematics and Computation*, 163(2):895–907, 2005.

[37] D. Zhang and Z. Zhou. $(2D)^2PCA$: Two-directional two-dimensional PCA for efficient face representation and recognition. *Neurocomputing*, 69(1-3):224–231, 2005.