Pattern Recognition 52 (2016) 249–259

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Sparse discriminative multi-manifold embedding for one-sample face identification

Pengyue Zhang [a,c], Xinge You [a,*], Weihua Ou [b], C.L. Philip Chen [a,d], Yiu-ming Cheung [e,f]

[a] Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
[b] School of Mathematics and Computer Science, Guizhou Normal University, Guiyang 550001, China
[c] Department of Computer Science, Stony Brook University, Stony Brook, NY, 11790, USA
[d] Faculty of Science and Technology, University of Macau, Macau, China
[e] Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China
[f] BNU-HKBU United International College, Zhuhai, China

ARTICLE INFO

Article history:
Received 19 January 2014
Received in revised form
14 January 2015
Accepted 21 September 2015
Available online 29 October 2015

Keywords:
Face recognition
Structured sparse representation
Manifold embedding

ABSTRACT

In this paper, we study the problem of face identification from only one training sample per person (OSPP). For a face identification system, the most critical obstacles towards real-world applications are often caused by the disguised, corrupted and varying illuminated images in limited sample sets. Meanwhile, storing fewer training samples would essentially reduce the cost for collecting, storing and processing data. Unfortunately, most methods in the literature basically need large training sets for good representation and generation abilities and would fail if there is only one training sample per person. In this paper, we propose a two-step scheme for the OSPP problem by posing it as a representation and matching problem. For the representation step, we present a novel manifold embedding algorithm, namely sparse discriminative multi-manifold embedding (SDMME), to learn the intrinsic representation beneath the raw data. We construct two sparse graphs to measure the sample similarity, based on two structured dictionaries. Multiple feature spaces are learned to simultaneously minimize the bias from the subspace of the same class and maximize the distances to the subspaces of other classes. For the matching step, we use a distance metric based on the manifold structure to identify the person. Extensive experiments demonstrate that the proposed method outperforms other state-of-the-art methods for the problem of one-sample face identification, while the robustness with occlusion and illumination variances highlights the contribution of our work.

1. Introduction

The problem of automatic face identification has attracted considerable attentions in the field of computer vision and pattern recognition [1–4]. Though it is a simple task for human beings to recognize a familiar person, it nevertheless remains a challenging task for machine vision, especially under complex conditions such as partial occlusion, varying illumination and limited labeled training samples. In a real-world scenario, such as ID card identification and airport surveillance, because of the difficulty in grasping face photographs or limited ability of the system for data storage, there is usually only one image for each individual. On the other hand, storing only one sample for each person in the database has many advantages: they are easy to collect, save storage space and reduce computational cost. Therefore, developing an efficient method for the task of face identification with one-sample per person is of great significance.

In recent decades, many efforts have been devoted to human face identification. There are many methods targeting at the problem of face recognition with corruption, occlusion and illumination variances. They have achieved impressive face recognition performance on some databases with a large number of accessible training samples. Unfortunately, most of these approaches rely to a great degree on the size and representativeness of the training dataset [5]. While these methods focus on improving the recognition accuracy, they ignore the problem originated from the dataset without sufficient training samples [6,7]. For the problem of one-sample face identification with gross corruption and illumination variation, many popular methods perform terribly or even fail due to the ignorance of the sample deficiency. For instance, the sparse representation-based classification (SRC) [8] and its extensions [9,10] are effective methods for the problem of face recognition with varying expression and illumination, as well as occlusion and disguise. Nevertheless, the SRC-based methods

* Corresponding author. Tel.: +86 27 87544014 8328, fax: +86 27 87542831.
E-mail addresses: penryzhang@gmail.com (P. Zhang), youxg@mail.hust.edu.cn (X. You), ouweihuahust@gmail.com (W. Ou).

http://dx.doi.org/10.1016/j.patcog.2015.09.024
0031-3203/

basically need a sufficiently large set of training data to construct a well-defined dictionary, which is not applicable in the scenario of one-sample face identification.

In the literature, the problem of one-sample face identification can be addressed by roughly two ways: holistic methods and local methods. The holistic method includes linear subspace learning methods, such as PCA [11–13], $(PC)^2A$ [14] and LPP [15,16], and manifold learning methods like LLE [17], ISOMAP [18] and LE [19]. This kind of methods aims to learn a low-dimensional feature space by preserving the desired relationships modeled by the similarity graphs and suppressing the undesired ones. The PCA-based approaches can efficiently recover the subspace structure of the high-dimensional data contaminated by small Gaussian noise, but is extremely fragile to gross corruption [20]. Meanwhile, these learning mechanisms basically need a sufficiently large training dataset for a good generation performance [3]. As regards the problem of one-sample face identification, these approaches may easily fall into the trap of overfitting and lose the generalization capability. Moreover, the graph structure of these approaches highly depends on the choice of parameters, which may result in an unstable graph. For example, the structure of the adjacency graph is changing with parameters such as the ϵ in ϵ-ball graph and k in k-NN graph. However, it is difficult to obtain a theoretical suitable and adaptive parameter setting for the optimal graph [21].

The other kinds of holistic methods, such as SVD perturbation [22] and Enhanced $(PC)^2A$ ($E(PC)^2A$) [23], generate new samples or different representations from the original training samples. Adding these imitated images could effectively enlarge the sample number and transform the problem of face identification with one training sample per person to an ordinary face identification problem. However, one major drawback of this kind of methods is that the augmented samples are highly correlated to the original ones and thus cannot be considered as independent samples [24].

For the local methods, such as local probabilistic subspace [24], component-based SVM [25] and self-organizing maps (SOM [26,27]), the face images are represented by a set of local features. Heisele et al. [25] proposed a component-based framework for face detection and identification. The facial semantic parts, such as eyes, nose and mouth, are automatically detected and passed to a combined SVM for identification. Matinez et al. [24] proposed a probabilistic approach for face identification with partial occlusion and expression variation. Six eclipse-shaped areas are used to learn the eigenspace, where the distribution is modeled by Gaussian mixture model. Finally a probabilistic identification approach is adopted to compute the similarity. Tan et al. [27] extended the probabilistic work by representing the face subspace with self-organizing maps (SOM [26]). This representation is proved to be more compact, intuitively comprehensible and robust to noise. Lu et al. [28] proposed the DMMA approach to learn discriminative features from image patches. Each image is partitioned into local patches of the same size and each patch is treated as a data point. The DMMA method aims to learn N different feature spaces for N classes. By utilizing the class label information, the manifold margins are maximized to learn more discriminant features. A local manifold distance is used for manifold matching. Compared to the holistic methods, the local methods have several advantages. Firstly, by learning low-dimensional local features, we can effectively reduce the feature dimension. Secondly, representing the image by a set of local features can effectively overcome the sample deficiency. Thirdly, local methods are more robust to corruption and illumination, since local patches are less likely to be corrupted than the whole image. However, using only the local features may lead to a loss in holistic information. Therefore, one has to provide a suitable way combining both the local features and holistic information, which is a quite challenging task.

In this paper we present a two-step scheme for one-sample face identification with gross corruption and varying illumination. Motivated by the Graph Embedding framework [29], we propose a new manifold embedding approach for data representation, namely sparse discriminative multi-manifold embedding (SDMME). First we partition each image into several non-overlapping patches with the same size. Based on two structured dictionaries, we build two sparse graphs to model the interconnection among different patches. We learn multiple discriminative feature spaces for each individual by maximizing the inter-manifold separability and minimizing the intra-manifold variance. For recognition, we use a global manifold distance to measure the similarity between samples. We evaluate the proposed SDMME approach on two publicly available facial image databases. The contribution of our work includes the following:

1. We present a class-conditional graph learning algorithm. From the perspective of classification, instead of a global graph of all classes, we learn the intrinsic graph (intra-manifold graph) and the penalty graph (inter-manifold graph) within the training set. Given N samples (each partitioned into M patches of the same size), we have $2*M*N$ graphs overall. With these graphs, we can efficiently characterize the relationships between data points. It is worthy to be noted that with sparse regularization we will learn a more stable and robust graph with fewer manually tuned parameters.
2. Based on the two kinds of graphs, we propose a novel manifold embedding approach (SDMME) for feature learning. We learn the feature spaces to preserve the characteristics of the intra-manifold graph and suppress the characteristics of the inter-manifold graph at the same time. In the feature spaces, the bias from the manifold of the same class is expected to be small and the distances to the manifolds of other classes are expected to be high. This would greatly improve the discriminant power of the learned feature.
3. We develop an efficient approach for the challenging problem of one-sample face identification. With the features learned by SDMME, the original problem of face identification with one training sample is cast as a problem with two parts: data representation and manifold matching. We use a global manifold metric which is more robust to occlusion and illumination variations. The proposed method outperforms the state-of-the-art face identification approaches handling the one training sample problem.

The paper is organized in the following manner. In Section 1, we introduce the problem of face recognition with a single training sample per person and review some classical face recognition methods. In Section 2, we present the details of the proposed method. Experimental results and analysis are provided in Section 3. In Section 4, we conclude the paper and suggest possible future research directions.

2. Sparse discriminative multi-manifold embedding

2.1. Sparse representation

In the statistical signal processing community, the problem of sparse representation had received much attention [30–32]. Sparse representation was originally designed to deal with compress sensing problems, searching for the possibility of using lower sampling rates than the Shannon–Nyquist bound [33]. Recently the sparse representation technique had been proved to be an effective tool in computer vision problems, such as image restoration, object detection and face recognition [8,34,35]. In

signal theory, a signal observation y with noise is usually modeled as $y=\hat{y}+e$, where $\hat{y}$ denotes the original signal and e is the error term that compensates the noise or the reconstruction error. Classic coding theories such as the Fourier transform and the Wavelet transform assume that there is a set of basis functions ψ_n, such that the observation can be represented as a linear superposition of the weighted basis functions $y=\sum_{i=1}^{N}\alpha_i\psi_i+e$. In sparse coding, with an overcomplete dictionary and sparse regularization, most weights of the basis function are set to zero, i.e. the observation signal is sparsely coded with the basis functions. Wright et al. [8] define the basis functions by the set of all training samples $A=[x_1,x_2,\ldots,x_N]$, where $\{x_i, i=1,\ldots,N\}$ are column-wise concatenated vectors of training samples. Then the observation sample can be represented as

$$y=A\alpha+e=A\alpha+I\alpha_e=B\beta \tag{1}$$

where $\alpha=[\alpha_1,\alpha_2,\ldots,\alpha_N]^T\in\mathcal{R}^{N\times 1}$, $\alpha_e\in\mathcal{R}^{N\times 1}$, $B=[A,I]$ and $\beta=[\alpha;\alpha_e]$.

The sparsity of a vector α is measured by the value of its l_0 norm $\|\alpha\|_0$. Therefore, sparse representation is obtained through solving an optimization problem:

$$\hat{\alpha}=\arg\min\|\alpha\|_0,\quad \text{s.t. } y=B\beta \tag{2}$$

The l_0 regularized minimization problem is proven to be NP-hard [36]. However, it has been found that if the solution $\hat{\alpha}$ is sparse enough, the l_0 minimization problem is equal to the solution of the l_1 regularized version [30,31]. Thus $\hat{\alpha}$ is approximated through l_1 regularized minimization:

$$\hat{\alpha}=\arg\min\|\alpha\|_1,\ \text{s.t. } y=B\beta \tag{3}$$

or the relaxed version

$$\hat{\alpha}=\arg\min\|\alpha\|_1,\ \text{s.t. } \|y-B\beta\|_2<\epsilon \tag{4}$$

This l_1-minimization problem can be effectively solved through a linear programming process [37,38]. The sparse coefficient vector $\hat{\alpha}$ can be regarded as an image feature and be applied to classification in later steps.

2.2. Sparse neighborhood preserving embedding

Cheng et al. [21] proposed the sparse neighborhood preserving embedding (SNPE) method for multiple vision tasks of semi-supervised learning, subspace learning and data clustering. Similar to LLE and its linear extension NPE, sparse neighborhood preserving embedding (SNPE) aims to obtain an optimal embedding which preserves the neighborhood similarity relationship. Different from LLE, which models the neighborhood similarity relationship with a k-NN graph, the SNPE approach constructs a sparse graph and preserves the graph relation through the embedding. The target graph is $G=(X,W)$, where $X=[x_1,x_2,\ldots,x_i,\ldots,x_N]\in\mathcal{R}^{D\times N}$ is the set of training samples and W is the graph weight matrix. For x_i in the sample set, we denote the sample dictionary as $A_i=[x_1,x_2,\ldots,x_{i-1},x_{i+1},\ldots,x_N]\in\mathcal{R}^{D\times(N-1)}$ and the expanded dictionary as $B_i=[A_i\ I]\in\mathcal{R}^{D\times(N+D-1)}$. The sparse coding is achieved by solving the l_1-minimization problem:

$$\hat{\alpha}=\arg\min\|\alpha\|_1,\ \text{s.t. } x_i=B_i\beta \tag{5}$$

Note that $\alpha\in\mathcal{R}^{N-1}$ and $\beta=[\alpha;\alpha_e]$. Then the graph weight W is set as

$$w_{ij}=\begin{cases}\hat{\alpha}_j, & i>j\\ 1, & i=j\\ \hat{\alpha}_{j-1}, & i<j\end{cases} \tag{6}$$

where $\hat{\alpha}_j$ is the j-th element of $\hat{\alpha}$. To achieve the optimal embedding P, the SNPE approach aims to minimize the reconstruction error. The SPNE approach is formulated as the following optimization problem:

$$\min_P\sum_{i=1}^{N}\|P^Tx_i-P^TXw_i\|^2,\ \text{s.t. } P^TXX^TP=I \tag{7}$$

The constraint $P^TXX^TP=I$ is added to avoid degenerate solutions. By solving the objective function, we have a unified projection matrix P which preserves the graph relationship. The projection matrix P represents a transformation from the original high-dimensional datum into a low-dimensional one. The low-dimensional representation of the original manifold point x_i can be calculated as $y_i=P^Tx_i$. The reconstruction error is measured by the l_2 norm for simplicity and efficiency.

The optimization problem can be solved by transforming it into a generalized eigenvalue problem:

$$X(I-W)^T(I-W)X^Tp=\lambda XX^Tp \tag{8}$$

The projection P is computed as the eigenvectors corresponding to the largest d eigenvalues:

$$P=[\,p_1,p_2,\ldots,p_d\,]\in\mathcal{R}^{D\times d} \tag{9}$$

2.3. Class-conditional graph learning with sparse criterion

We denote the sample matrix as $X=[x_1,x_2,\ldots,x_N]$. First we partition each sample into M non-overlapping patches of the same size (shown as in Fig. 1) and concatenate the patches into vectors by column. For the i-th subject x_i, let $X_i=\{x_{i,j}|j=1,2,\ldots,M\}$ be the set of vectors of corresponding patches, where $x_{i,j}$ is the j-th vector of x_i. For local methods, with loss of global information in the partitioning process, it is essential to incorporate global information to the system. The most commonly used way of incorporating global information is to construct a graph that measures the similarity relationship between local patches, such as k-NN graph. However, the k-NN graph heavily depends on the graph parameters. For the problem of face identification, the sparse graph (or l_1-graph) has proved to be more robust to noise and more adaptive in neighborhood structures [21].

In this section, we design two kinds of graph based on class label information: the intrinsic graph (or intra-manifold graph) that characterizes the intra-class compactness and the penalty graph (or inter-manifold graph) that characterizes the inter-class separability. For graph construction, we must make three observations:

- All the graph points should be interrelated, including the pair of points with large Euclidean distance, though their relationship may be loose.
- In sparse coding, a sample is most likely to be sparsely represented by the dictionary atoms from the same class, which implies that the similarity between the same class sample should be larger than others.
- In the graph construction step, all of the patch vectors are unified with a unit l_2 norm to avoid the imbalance between dictionary atoms. In the later steps of recognition, however, there is no need for the normalization.

To measure the representation capability of the samples from the same class and all other classes, we design two different kinds of dictionaries (taking the i-th class as an example): inter-manifold dictionary:

$$A_i^{inter}=X/X_i=[x_{1,1},\ldots,x_{i-1,M},x_{i+1,1},\ldots,x_{N,M}]\in\mathcal{R}^{d\times(N-1)M} \tag{10}$$

and intra-manifold dictionary:

$$A_i^{intra}=X_i/x_{i,j}=[x_{i,1},\ldots,x_{i,j-1},x_{i,j+1},\ldots,x_{i,M}]\in\mathcal{R}^{d\times(M-1)} \tag{11}$$

To initialize the inter-manifold graph, we solve the l_1 regularized optimization problem

$$\alpha_{i,j}^{inter}=\arg\min\|\alpha\|_1,\ \text{s.t. } \|x_{i,j}-B_i^{inter}\beta\|_2<\epsilon \tag{12}$$

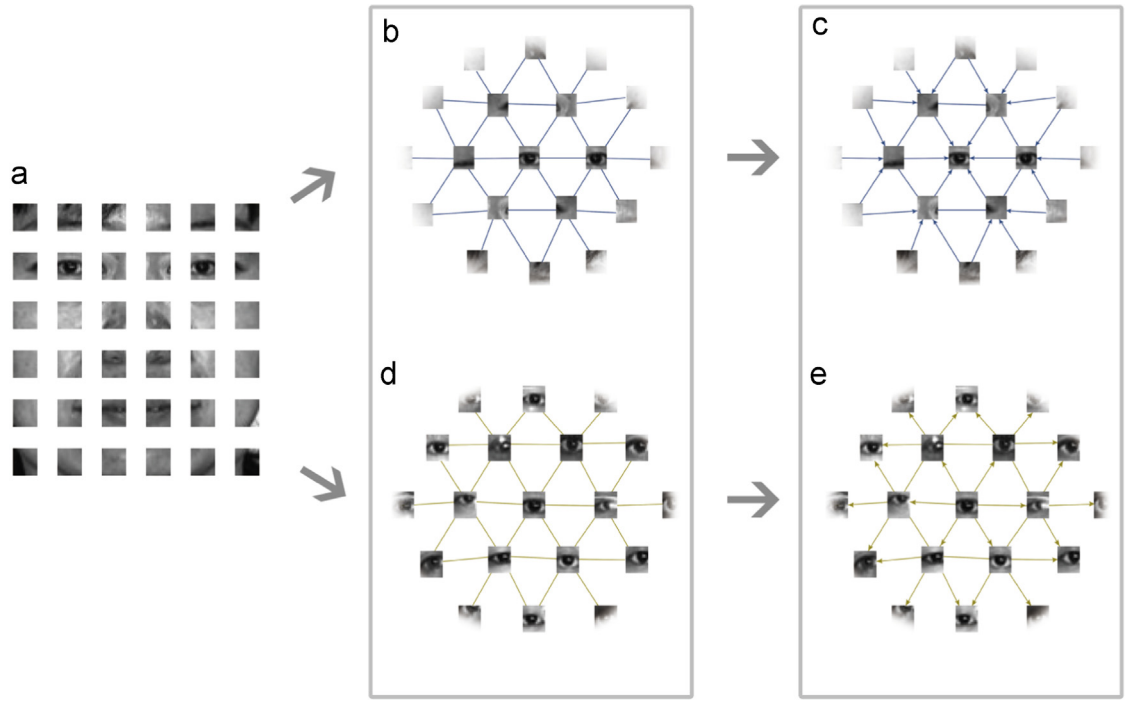

Fig. 1. Class-conditional Graph Learning: (a) Non-overlapping patches. There are $6 \times 6 = 36$ patches in this example. (b) In the intra-manifold graph, the vertices are patches from the same person. Each vertex is connected to the center vertex, indicating that the similarity between each pair of vertices is taken into consideration. Because of the sparse regularization, however, most graph weights are set to small values. (c) The intra-manifold graphs are expected to "shrink" through the embedding. (d) In the inter-manifold graph, the vertices are from different classes. Note that the closest vertices are usually the same semantic parts of different people. (e) The inter-manifold graphs are expected to "expand".

where $B_i^{inter} = [A_i^{inter}, I] \in \mathcal{R}^{d \times [(N-1)M+d]}$ and $\beta = [\alpha; \alpha_e]$. Then the j-th column inter-manifold graph of the i-th subject is

$$w_{i,j,k}^{inter} = \begin{cases} |\alpha_{i,j,k}^{inter}|, & 0 < k \le (i-1)M \\ 1, & (i-1)M < k \le iM \\ |\alpha_{i,j,k-M}^{inter}|, & iM < k \le NM \end{cases} \tag{13}$$

where $\alpha_{i,j,k}^{inter}$ and $w_{i,j,k}^{inter}$ denote the k-th element of $\alpha_{i,j}^{inter}$ and $w_{i,j}^{inter}$, respectively. The inter-manifold graph of the i-th sample is $W_i^{inter} = [w_{i,1}^{inter}, \ldots, w_{i,M}^{inter}] \in \mathcal{R}^{NM \times M}$. Similarly for intra-manifold graph,

$$\alpha_{i,j}^{intra} = \arg\min \|\alpha\|_1, \quad \text{s.t. } \|x_{i,j} - B_i^{intra}\beta\|_2 < \epsilon \tag{14}$$

where $B_i^{intra} = [A_i^{intra}, I] \in \mathcal{R}^{d \times [M-1+d]}$ the j-th column of the i-th intra-manifold graph is

$$w_{i,j,k}^{intra} = \begin{cases} |\alpha_{i,j,k}^{intra}|, & 0 < k < j \\ 1, & k = j \\ |\alpha_{i,j,k-1}^{inter}|, & j < k < M \end{cases} \tag{15}$$

where $\alpha_{i,j,k}^{intra}$ and $w_{i,j,k}^{intra}$ denote the k-th element of $\alpha_{i,j}^{intra}$ and $w_{i,j}^{intra}$, respectively. The intra-manifold graph of the i-th sample is $W_i^{intra} = [w_{i,1}^{intra}, \ldots, w_{i,M}^{intra}] \in \mathcal{R}^{M \times M}$.

The graph construction procedure is summarized in Algorithm 1.

Algorithm 1. Class-conditional graph construction based on l_1 optimization

Input: Training dataset $X = [x_1, x_2, \ldots, x_N]$;
Output: The sparse class conditional graph W_i^{inter} and W_i^{intra};
Step 1: Partition each training image into M patches. Concatenate the columns of each patch. The patches of x_i are denoted as $x_{i,j}, j = 1, \ldots, M$;
Step 2: Normalize each data vector $x_{i,j}$ to have a unit norm;
Step 3: Initialize the inter-manifold dictionary A_i^{inter} and the intra-manifold dictionary A_i^{intra} for each class i, $i = 1, 2, \ldots, N$;
Step 4: Solve the l_1 regularized optimization problem such as in (12) and (14);
Step 5: Construct the sparse graph such as in (13) and (15);

For the representation step, our goal is to learn feature spaces to preserve the characteristics of the intra-manifold graph and suppress the characteristics of the inter-manifold graph. Intuitively, we need to pull together the vertices of the same class and push away the vertices of different classes.

2.4. Sparse discriminative multi-manifold embedding

In this section, we present a sparse discriminative multi-manifold embedding (SDMME) approach based on the class-conditional sparse graph. Previous works, such as SNPE, usually assume that all data points intrinsically lie on a single low-dimensional manifold. This assumption, however, is not appropriate in the case of limited training samples for the purpose of discriminative classification for the following reasons:

- As regards face recognition, the person-specific shape and depth information in a facial image can be quite critical for recognition. Modeling all the samples as a single manifold may cause a loss in such information.
- From the perspective of subspace learning, the intrinsic feature dimensions of different faces will not be the same. Therefore, the face features with same dimensions from a unified manifold may not accurately represent the faces.
- For classification tasks, supervised methods are more likely to achieve better recognition accuracy. Therefore, considering the discriminative information, modeling the data as N different

manifolds is more appropriate for precise recognition. Therefore, we learn N SDMME feature spaces for N classes.

The class-conditional sparse graphs intrinsically model the similarity relationship between data points. We desire to preserve the graph structure through the manifold learning. For OSPP problem, to overcome the shortage of training samples, we partition each image into t non-overlapping patches. Throughout the manifold learning process, the image patch is considered as the process unit. In order to learn the low-dimensional representation, we design two kinds of dictionaries, i.e. intra-manifold dictionary and inter-manifold dictionary. We then construct the class-conditional sparse graph, with the technique introduced in Section 2.3. For samples from class i, we have an inter-manifold graph

$$W_i^{inter} = [w_{i1}^{inter}, w_{i2}^{inter}, \ldots, w_{ij}^{inter}, \ldots, w_{it}^{inter}] \tag{16}$$

and an intra-manifold graph

$$W_i^{intra} = [w_{i1}^{intra}, w_{i2}^{intra}, \ldots, w_{ij}^{intra}, \ldots, w_{it}^{intra}] \tag{17}$$

For simplicity, we omit the commas. For each class i, we aim to search for a best projection matrix $P_i \in \mathcal{R}^{M\times D_i}$. Different from SNPE, N different discriminative subspaces will be learned. We aim to learn a set of projection matrices $P = \{P_1, P_2, \ldots, P_N\}$. The overall objective function is

$$R(P) = \min_P \sum_{i=1}^{N} R(P_i) \tag{18}$$

where

$$R(P_i) = R_1(P_i) - R_2(P_i) \tag{19}$$

$$R(P_i) = \min_{P_i} \sum_{j=1}^{M} \| P_i^T x_{ij} - P_i^T X_i w_{ij}^{intra} \|^2 - \| P_i^T x_{ij} - P_i^T X w_{ij}^{inter} \|^2 \tag{20}$$

In the learning process, we minimize the objective function to learn an optimal embedding. The N projection matrices are independent of each other and thus the original objective function can be considered as the sum of N subfunctions $R(P_i)$ which are separately optimized. The optimization problem can be solved with the generalized eigenvalue decomposition approach [39] as

$$(M_i^{intra} - M_i^{inter})p_k = \lambda_k p_k \tag{21}$$

where

$$M_i^{intra} = \sum_{j=1}^{M} (x_{ij} - X_i W_{ij}^{intra})(x_{ij} - X_i W_{ij}^{intra})^T \tag{22}$$

and

$$M_i^{inter} = \sum_{j=1}^{M} (x_{ij} - X W_{ij}^{inter})(x_{ij} - X W_{ij}^{inter})^T \tag{23}$$

The deductive process is given in Appendix A. For simplicity, we omit the class label i and denote the k-th largest eigenvalue as λ_k and the corresponding eigenvector p_k. Then we rank the eigenvalues as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{D_i} \geq 0 > \lambda_{D_i+1} \geq \cdots \geq \lambda_D$. All the eigenvectors with a positive eigenvalue are selected to form the embedding matrix $P_i = [p_1, p_2, \ldots, p_{D_i}]$. For a data point x in the original data manifold, the corresponding low-dimensional feature in the i-th manifold is $y_i = P_i^T x$. In this manner, for an unlabeled sample, we are able to attain N low-dimensional feature, and calculate its bias from the manifolds. For the classification task, all the biases from N manifolds are calculated and the label corresponding to the smallest bias is chosen as the classification result.

Algorithm 2. Sparse discriminative multi-manifold embedding.

Input: Training dataset $X = [x_1, x_2, \ldots, x_N]$;
Class-conditional sparse graph W_i^{intra}, W_i^{inter}, $i = 1, 2, \ldots, N$
Output: Projection matrices P_i, $i = 1, 2, \ldots, N$
Step 1: Initialize the inter-manifold dictionary A_i^{inter} and the intra-manifold dictionary A_i^{intra} for each class i, $i = 1, 2, \ldots, N$;
Step 2: Transform the optimization problem (19) to a generalized eigenvalue problem (21); calculate intra-manifold reconstruction matrix M_i^{intra} and inter-manifold reconstruction matrix M_i^{inter};
Step 3: Solve problem (21) and sort the eigenvalue as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d \geq 0 > \lambda_{d+1} \geq \cdots \geq \lambda_D$;
Step 4: Select the eigenvectors that correspond to positive eigenvalues to form the projection matrix.

Fig. 2 visualizes the distribution in the original manifold and the manifold learned by SDMME. Two kinds of data points are exhibited. The red stars represent all of the data points (36 points in this case) of a person, whereas the blue circles represent randomly selected 100 points from other classes. (a) shows the original distribution in the data space. (b) shows the distribution of

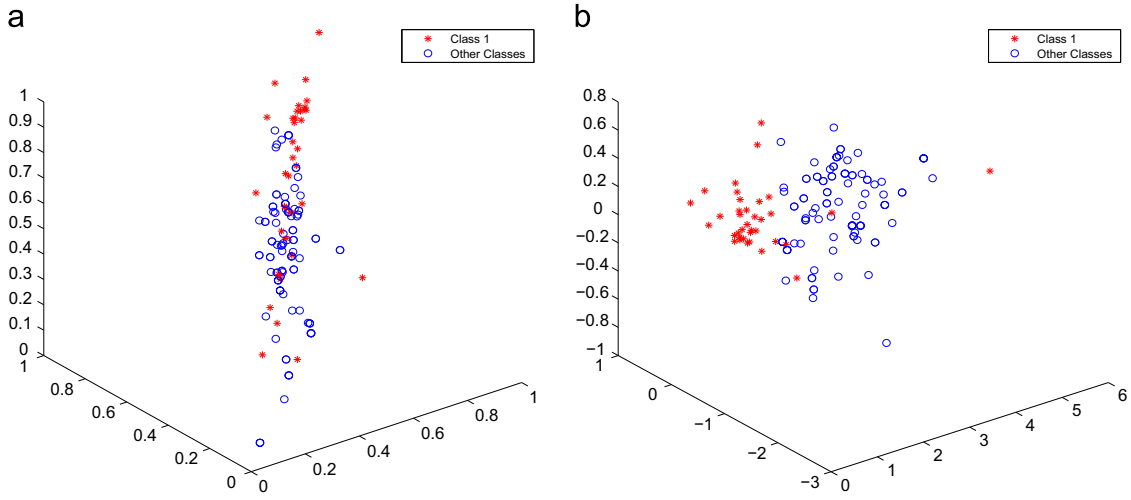

Fig. 2. Visualization of distributions of (a) original manifold and (b) low-dimensional manifold learned by SDMME. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

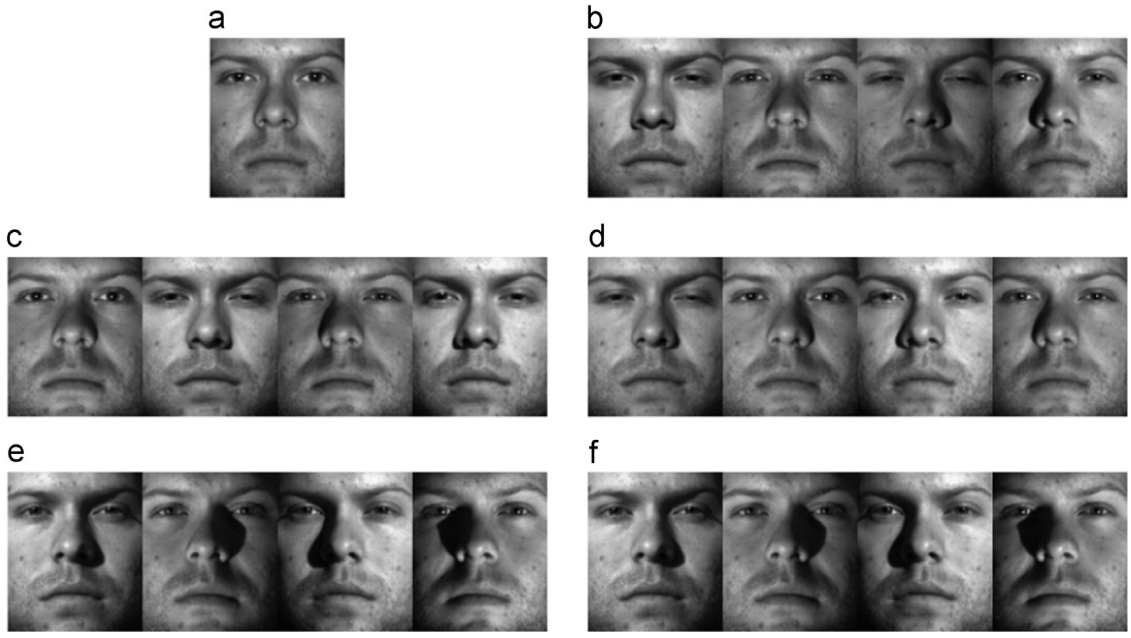

Fig. 3. Sample images from subset A–subset F of Yale B face database.

the data points in the low-dimensional manifold learned by SDMME. We observe that in (a) the points of different classes strongly overlap, and in (b) the two kinds of points are clearly separated. In the learning step, the i-th low-dimensional manifolds are learned by reducing the intra-class variation of the i-th class and expanding the separability from the other classes, i.e. by pulling the points of the i-th class towards each other and pushing the other points away. In this view, the subspaces learned by SDMME show the introduction of discriminative information, which can be rather helpful in terms of recognition.

2.5. Inherent structure-based manifold matching

In this section, we propose a global manifold matching technique based on the global manifold distance. After discriminative feature learning, the original recognition problem has been transformed into a point-manifold matching problem. From the previous steps, we have learned N SDMME feature spaces. For an unlabeled sample u, the image is first partitioned into t non-overlapping patches $\{u_1, u_2, \ldots, u_j, \ldots, u_t\}$ in the same way. Then the patches are projected to the spaces for a low-dimensional representation. To compute the distances from sample u to each of the N classes, we calculate the patch-manifold distances first and sum these distances as the sample-manifold distance:

$$d(u, M_i) = \sum_{j=1}^{t} d(u_j, M_i) \tag{24}$$

We model the patch-manifold distance by means of the reconstruction capability of the reference manifold. The distance is computed by solving the minimization problem below:

$$d(u_j, M_i) = \min_{\sum_l \pi_l = 1} \| P_i^T u_j - \sum_{l=1}^{M} \pi_l P_i^T m_{il} \|^2 \tag{25}$$

where m_{il} is the l-th vector of M_i and π_l is a weight factor corresponding to m_{il}.

With the global manifold metric, we attain the label L by

$$L = \arg\min_i d(u, M_i), \quad i = 1, 2, \ldots, N \tag{26}$$

For this sample-manifold distance, we note three facts:

- The distance $d(u_j, M_i)$ is based on the SDMME feature space of the i-th class. Thus we need to first project the patches onto the subspace and then compute the distance. Note that to measure the bias between u_j and another manifold, the distance will be computed in a different space.
- Rather than the recognition metric used in DMMA, which exploits only the k-nearest patches to calculate the distance, we propose here a global distance in which all of the t patches are involved in the metric. This is because in complex environments, such as extreme illumination, the data points in the manifold are distributed with large variance. Therefore, the k-nearest patches in the manifold cannot truly reflect the real distribution overall and the local distance is thus not accurate.
- In fact, we minimize the cost function with the sum-to-one constraint: $\sum_l \pi_l = 1$. This is because the sum-to-one constraint can effectively enforce the invariance of translation, as the effect of translation terms in data will be offset by each other.

3. Experiments

In this section, we test the effectiveness of the proposed SDMME method for the one-sample face identification problem. Our main concerns are the robustness of our method to partial occlusion and illumination variance. Two publicly available face datasets, namely extended Yale B and CMU PIE facial image database, are used for experiments. We compare the SDMME method with representative face recognition methods such as PCA, Block-PCA, 2DPCA, $(PC)^2$A, LPP, SOM, SNPE, and DMMA.

3.1. Datasets

The extended Yale B face dataset [40] consists of 2414 frontal-face images of 38 different individuals. All of the images are photographed with light sources from different directions. Consequently, there may be shadows on part of the face, which may pose difficulties in terms of recognition. In the experiment, all of the images are cropped and resized to the size of 48×42 pixels. We use a subset of the Yale B dataset to test the availability of our method. The images are divided into six subsets according to the

illumination angle. In the experiment, we use the image with normal illumination for training (subset A), and the other five subsets are used for testing (subset B to subset F) (Fig. 3).

The CMU Pose, Illumination and Expression (CMU PIE) face dataset [41] includes over 40,000 facial images of 68 people. The images are 640×486 color images. Each person is photographed in 13 different poses, under 43 different illumination conditions and with 4 different expressions. The database consists of two partitions, with pose and illumination variation for the first partition, with pose and expression variation for the second one. To obtain significant illumination variation, the people are photographed with 21 flashes from different directions. Therefore, images captured with and without the background lighting give $21 \times 2+1=43$ different illuminations overall. In the experiment, the first parts of frontal facial images (1496 images of 68 people, 22 images each person) are used, which are taken in 22 different illumination conditions. The experimental images are first cropped and resized to the size of 48×42 pixels. For each person, the image without flashes is used as the training image (subset A) and the left 21 images as the testing images. We partition the testing images into four different subsets (subset B to subset E) depending on the flash direction (Fig. 4).

3.2. Robustness to illumination variation

In this section, we focus on evaluating the robustness of face recognition algorithms under variant illumination conditions. Illumination variation is one of the most critical obstacles in robust face identification. In the literature, state-of-the-art methods can be grouped into three categories: illumination modeling, illumination invariant feature extraction, and face image normalization. Nevertheless, these approaches require prior knowledge of the illumination model, which is not satisfied in most cases. In the experiment, we use a subset of the Yale B database for the experiment. For each individual, the neutral face (subset A) is selected as the only training sample. The remaining subsets (subset B to subset F) are used for testing. Table 1 exhibits the face recognition accuracy of different methods in Yale B dataset. To ensure the fairness of the comparison, we have tuned the parameters of all the methods as appropriate. For block PCA and DMMA, each image is partitioned into 6×6 patches and the size of the patches is 8×7 pixels, which is the same as in SDMME. For $(PC)^2A$, the weighting parameter α is set as 0.25. For $E(PC)^2A$, the two weighting parameters are set as 0.25 and 0.5, respectively. For LPP, the size of neighborhood k is set as 4 and the heat kernel parameter t is set as 100. For DMMA, the parameters k_1, k_2, k and σ are set as 15, 5, 4 and 100, respectively. For SNPE and SDMME, the l1-ls toolbox [42] is used to solve the l_1 minimization problem. The balance factor λ is set as $0.001\lambda_{max}$.

Table 1 shows that the proposed SDMME method outperforms all the other methods on the Yale B database. In the light of the recognition result, we comment as follows:

- Many face recognition methods report excellent recognition performance with multiple training samples. It still remains a most challenging problem for one-sample face identification, however. Some of the most popular face recognition algorithms, such as PCA-based and LPP-based methods, do not perform well with such problems.
- For some methods such as LPP and SPNE, a PCA preprocessing step is added to avoid singularity.
- Compared with the DMMA approach, the recognition accuracy of the SDMME method outperforms it in all cases. Specifically, the SDMME achieved a gain of 17.76 and 8.55 percent in subset E and subset F, which consist of samples with more extreme illumination conditions. This result shows that the DMMA

Table 1
Recognition accuracy (percent) on the Yale B database.

Method	Subset B	Subset C	Subset D	Subset E	Subset F
PCA	**99.34**	95.39	76.32	32.24	19.08
2DPCA	98.03	92.11	76.97	33.55	25.66
Block PCA	**99.34**	96.05	76.32	32.89	21.05
$(PC)^2A$	96.05	85.53	85.53	36.18	40.79
$E(PC)^2A$	91.45	86.84	76.32	37.50	23.68
LPP	**99.34**	**98.68**	82.24	35.53	33.55
SPNE	97.37	96.05	94.74	47.37	38.16
DMMA	96.71	97.37	**98.03**	40.79	34.21
SDMME+5-NN	93.42	90.13	78.95	11.84	13.16
SDMME+Global Distance	**99.34**	**98.68**	**98.03**	**58.55**	**42.76**

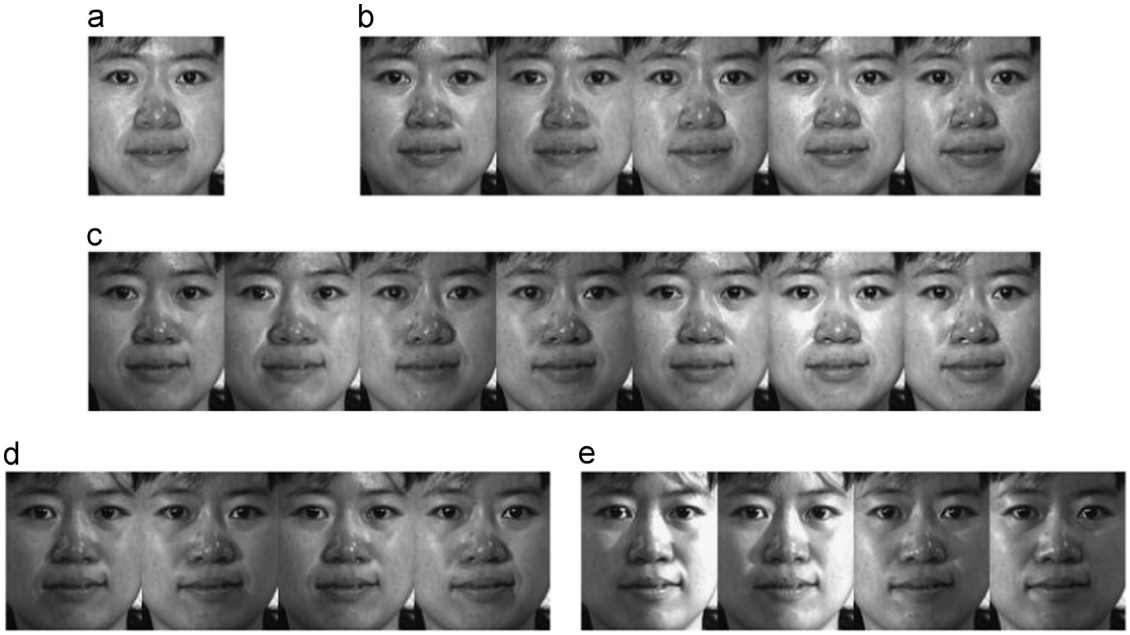

Fig. 4. Sample images from subset A to subset E of the PIE face database.

approach is not as nearly robust as the proposed SDMME method in the case of extreme illumination.

- The original DMMA method adopted a local manifold–manifold distance to fulfill the classification task. In the experiment, we test this metric for its robustness in extreme illumination and compare it with the global manifold distance metric. The recognition accuracy of these metrics shows that the local manifold distance is not nearly robust enough under extreme illumination, whereas the global manifold distance can still work perfectly in similar cases.

Table 2 shows the recognition accuracy obtained by different methods in the CMU PIE database. Though the recognition accuracy of these methods in the PIE database is generally lower than in the Yale B database, the proposed SDMME method still outperforms all the other methods with considerable accuracy gain. The main reasons for the lower recognition accuracy in the PIE database may be the harder illumination condition and larger database size. Especially in subset D and subset E, we observe a recognition rate below 50% for many recognition methods. The proposed SDMME method, however, can still achieve a 73.80% and 59.19% recognition rate in subset D and subset E, respectively.

Table 2
Recognition accuracy (percent) on the PIE database.

Method	Subset B	Subset C	Subset D	Subset E
PCA	95.88	39.79	2.95	7.35
2DPCA	99.41	82.74	34.32	26.07
Block PCA	98.24	48.63	4.06	7.72
$(PC)^2A$	99.41	92.63	72.69	42.28
$E(PC)^2A$	94.12	44.21	6.64	9.93
LPP	99.12	72.84	21.03	26.10
SPNE	99.12	94.53	73.43	58.09
DMMA	99.12	91.16	72.32	56.99
SDMME+5-NN	90.59	46.11	11.07	13.60
SDMME+Global distance	**100**	**97.26**	**73.80**	**59.19**

We have several observations from the experiment result:

- The recognition accuracy of the PCA-based methods drops dramatically in the subsets with extreme illumination. There are two reasons for this phenomenon. First, in PCA-based algorithms, the illumination and corruption are not concerned and appropriately modeled. Second, its disfunction in dealing with the small sample problem is one of the intrinsic drawbacks of the PCA-based algorithm. In the experiment, we train the projection matrix with only one training sample per person, whereas the feature dimension is large. The curse of dimension affects the performance here and results in poor recognition accuracy.
- Compared with DMMA, the proposed SDMME method still achieves a gain in recognition accuracy of 6.1, 1.46 and 2.2 percent, respectively, in subset B, subset C and subset D. This result demonstrates that the proposed SDMME method is more robust to illumination variation than DMMA.

3.3. Robustness to random block occlusion

In this section, we test various scenarios of block occlusion by replacing a randomly located block with a black block and an unrelated image in the test images. We test the effectiveness of our approach on subset B of the PIE database. Blocks with a side length of 5, 10, 15, 20, and 25 pixels occlude the image at random locations. Fig. 5 shows several examples of faces occluded by a black block. We run the experiment 10 times in each case to

Fig. 5. Images with different occlusion size (block): (a) 5×5 pixels, (b) 10×10 pixels, (c) 15×15 pixels, (d) 20×20 pixels, and (e) 25×25 pixels.

eliminate the randomness. Fig. 6 displays the recognition accuracies of five different methods.

Fig. 6 shows the comparison of the five methods with random block occlusion. The proposed SDMME method outperforms the others in all cases. Note that the recognition accuracies decrease with greater block occlusion. This result is intuitively predictable as the facial features are severely affected by the size of occlusion (Fig. 7). We also note that even with 20×20 block occlusion, the proposed SDMME approach could still achieve a recognition rate of more than 90%, while none of the other methods could achieve an accuracy of more than 70%. Moreover, the relatively small variance of the recognition rate shows that the proposed SDMME approach is less affected by the location of the block.

We also simulate the scenario of contiguous occlusion with an unrelated image of gorilla face. Again, the location of the occlusion is randomly chosen for each image and is unknown to the computer. Fig. 8 demonstrates the comparison of the five methods in this case. The result shows that the proposed method consistently outperforms all the other methods.

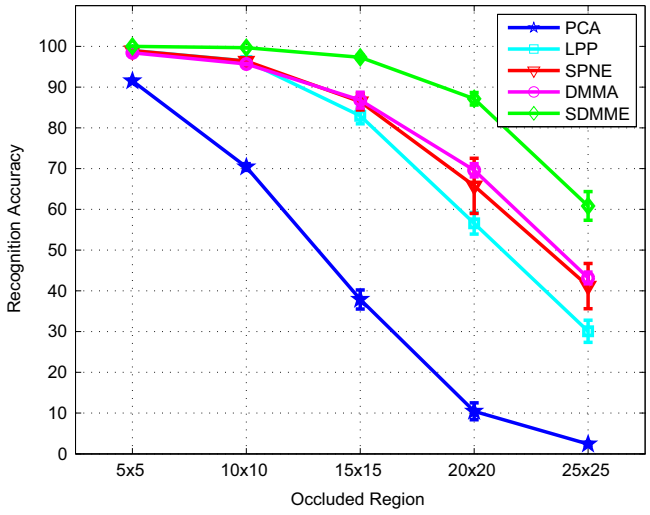

Fig. 6. Recognition accuracy with different occlusion sizes (block).

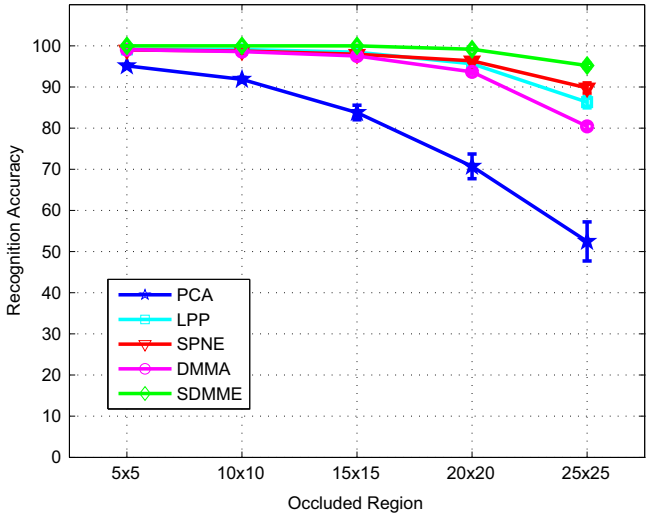

Fig. 8. Recognition accuracy with different occlusion size (gorilla).

Fig. 7. Images with different occlusion size (gorilla): (a) 5×5 pixels, (b) 10×10 pixels, (c) 15×15 pixels, (d) 20×20 pixels, and (e) 25×25 pixels.

4. Conclusion

We propose a novel approach for the problem of one-sample face identification. To overcome the limitation of inadequate training samples, the images are partitioned into non-overlapping patches of the same size and each point is considered as a data point in the manifold. We construct class-conditional sparse graphs based on two structure dictionaries. Rather than modeling all data points as one uniform manifold, we argue that the points of each class intrinsically lie in separate manifolds. We learn the subspaces by concentrating data points within classes and separating data points of one class from all the others. A global manifold distance is introduced for the classification task in the learned feature space. Experiments on several facial image databases demonstrate the effectiveness of the proposed approach.

Conflict of interest

None declared.

Acknowledgments

We want to thank the constructive comments and suggestions from the anonymous reviewers. This research was supported partially by the National Natural Science Foundation of China (Grant nos. 61272203, 61272366, 61402122, 61572540), the International Scientific and Technological Cooperation Project (Grant no. 2011DFA12180) and the Faculty Research Grants of Hong Kong Baptist University (project codes: FRG2/14-15/075 and FRG1/14-15/041).

Appendix A

$$
\begin{aligned}
R(P_i) &= \sum_{j=1}^{M} \| P_i^T x_{ij} - P_i^T A_i^{intra} w_{ij}^{intra} \|^2 - \| P_i^T x_{ij} - P_i^T A_i^{inter} w_{ij}^{inter} \|^2 \\
&= \sum_{j=1}^{M} \mathrm{tr}\left[P_i^T (x_{ij} - A_i^{intra} w_{ij}^{intra})(x_{ij} - A_i^{intra} w_{ij}^{intra})^T P_i \right] \\
&\quad - \mathrm{tr}\left[P_i^T (x_{ij} - A_i^{inter} w_{ij}^{inter})(x_{ij} - A_i^{inter} w_{ij}^{inter})^T P_i \right] \\
&= \mathrm{tr}\left[P_i^T \sum_{j=1}^{M} (x_{ij} - A_i^{intra} w_{ij}^{intra})(x_{ij} - A_i^{intra} w_{ij}^{intra})^T P_i \right] \\
&\quad - \mathrm{tr}\left[P_i^T \sum_{j=1}^{M} (x_{ij} - A_i^{inter} w_{ij}^{inter})(x_{ij} - A_i^{inter} w_{ij}^{inter})^T P_i \right] \\
&= \mathrm{tr}\left[P_i^T \left(\sum_{j=1}^{M} (x_{ij} - A_i^{intra} w_{ij}^{intra})(x_{ij} - A_i^{intra} w_{ij}^{intra})^T \right.\right. \\
&\quad \left.\left. - \sum_{j=1}^{M} (x_{ij} - A_i^{inter} w_{ij}^{inter})(x_{ij} - A_i^{inter} w_{ij}^{inter})^T \right) P_i \right] \\
&= \mathrm{tr}\left[P_i^T (M_i^{intra} - M_i^{inter}) P_i \right] \qquad (27)
\end{aligned}
$$

where

$$
M_i^{intra} = \sum_{j=1}^{M} (x_{ij} - A_i^{intra} W_{ij}^{intra})(x_{ij} - A_i^{intra} W_{ij}^{intra})^T \qquad (28)
$$

and

$$
M_i^{inter} = \sum_{j=1}^{M} (x_{ij} - A_i^{inter} W_{ij}^{inter})(x_{ij} - A_i^{inter} W_{ij}^{inter})^T \qquad (29)
$$

The optimization of the cost function

$$
\min_{P_i} R(P_i) = \min_{P_i} \mathrm{tr}\left[P_i^T (M_i^{intra} - M_i^{inter}) P_i \right] \qquad (30)
$$

can be transformed into the eigenvalue problem

$$
(M_i^{intra} - M_i^{inter}) p_k = \lambda_k p_k
$$

where λ_k denote the k-th largest eigenvalue and p_k denote the corresponding eigenvector.

References

[1] R. Chellappa, C.L. Wilson, S. Sirohey, Human and machine recognition of faces: a survey, Proc. IEEE 83 (5) (1995) 705–741.
[2] J. Daugman, Face and gesture recognition: overview, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 675–676.
[3] X. Tan, S. Chen, Z.H. Zhou, F. Zhang, Face recognition from a single image per person: a survey, Pattern Recognit. 39 (9) (2006) 1725–1745.
[4] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 35 (4) (2003) 399–458.
[5] A.K. Jain, B. Chandrasekaran, 39 dimensionality and sample size considerations in pattern recognition practice, in: Handbook of Statistics, vol. 2, 1982, pp. 835–855.
[6] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[7] A.M. Martínez, A.C. Kak, Pca versus lda, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2) (2001) 228–233.
[8] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[9] W. Deng, J. Hu, J. Guo, Extended src: undersampled face recognition via intraclass variant dictionary, IEEE Trans. Pattern Anal. Mach. Intell. 34 (9) (2012) 1864–1870.
[10] W. Ou, X. You, D. Tao, P. Zhang, Y. Tang, Z. Zhu, Robust face recognition via occlusion dictionary learning, Pattern Recognit. 47 (4) (2014) 1559–1572.
[11] I. Jolliffe, Principal Component Analysis, Springer, New York, NY, 2002.
[12] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: Computer Vision and Pattern Recognition, 1991. in: IEEE Computer Society Conference on Proceedings of CVPR'91, IEEE, Piscataway, NJ, 1991, pp. 586–591.
[13] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemom. Intell. Lab. Syst. 2 (1) (1987) 37–52.
[14] J. Wu, Z.-H. Zhou, Face recognition with one training image per person, Pattern Recognit. Lett. 23 (14) (2002) 1711–1719.
[15] X. He, S. Yan, Y. Hu, P. Niyogi, H.-J. Zhang, Face recognition using Laplacianfaces, IEEE Trans. Pattern Anal. Mach. Intell. 27 (3) (2005) 328–340.
[16] X. Niyogi, Locality preserving projections, in: Neural information processing systems, vol. 16, 2004, p. 153.
[17] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[18] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[19] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural computation 15 (6) (2003) 1373–1396.
[20] B. Bao, G. Liu, R. Hong, S. Yan, C. Xu, General Subspace Learning with Corrupted Training Data via Graph Embedding, IEEE Trans. Image Process 22 (11) (2013) 4380–4393.
[21] B. Cheng, J. Yang, S. Yan, Y. Fu, T.S. Huang, Learning with l_1 graph for image analysis, IEEE Trans. Image Process 19 (4) (2010) 858–866.
[22] D. Zhang, S. Chen, Z.-H. Zhou, A new face recognition method based on svd perturbation for single example image per person, Appl. Math. Comput. 163 (2) (2005) 895–907.
[23] S. Chen, D. Zhang, Z.-H. Zhou, Enhanced $(pc)^2$a for face recognition with one training image per person, Pattern Recognit. Lett. 25 (10) (2004) 1173–1181.
[24] A.M. Martínez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, IEEE Trans. Pattern Anal. Mach. Intell. 24 (6) (2002) 748–763.
[25] B. Heisele, T. Serre, T. Poggio, A component-based framework for face detection and identification, Int. J. Comput. Vis. 74 (2) (2007) 167–181.
[26] T. Kohonen, Self-Organizing Maps, 30, Springer, New York, NY, 2001.
[27] X. Tan, S. Chen, Z.-H. Zhou, F. Zhang, Recognizing partially occluded, expression variant faces from single training image per person with som and soft k-NN ensemble, IEEE Trans. Neural Netw. 16 (4) (2005) 875–886.
[28] J. Lu, Y.-P. Tan, G. Wang, Discriminative multi-manifold analysis for face recognition from a single training sample per person, in: 2011 IEEE

International Conference on Computer Vision (ICCV), IEEE, Piscataway, NJ, 2011, pp. 1943–1950.
[29] S. Yan, J. Liu, X. Tang, T.S. Huang, A parameter-free framework for general supervised subspace learning, IEEE Trans. Inf. Forens. Secur. 2 (1) (2007) 69–76.
[30] E.J. Candes, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Commun. Pure Appl. Math. 59 (8) (2006) 1207–1223.
[31] D.L. Donoho, For most large underdetermined systems of linear equations the minimal l_1-norm solution is also the sparsest solution, Commun. Pure Appl. Math. 59 (6) (2006) 797–829.
[32] P. Zhao, B. Yu, On model selection consistency of lasso, J. Mach. Learn. Res. 7 (2006) 2541–2563.
[33] E.J. Candès, M.B. Wakin, An introduction to compressive sampling, IEEE Signal Process. Mag. 25 (2) (2008) 21–30.
[34] S. Agarwal, D. Roth, Learning a sparse representation for object detection, in: Computer Vision—ECCV, vol. 2002, 2006, pp. 97–101.
[35] J. Mairal, M. Elad, G. Sapiro, Sparse representation for color image restoration, IEEE Trans. Image Process. 17 (1) (2008) 53–69.
[36] E. Amaldi, V. Kann, On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems, Theor. Comput. Sci. 209 (1) (1998) 237–260.
[37] D.L. Donoho, Y. Tsaig, Fast Solution of l_1-Norm Minimization Problems When the Solution May Be Sparse, Department of Statistics, Stanford University, 2006.
[38] K. Koh, S.-J. Kim, S.P. Boyd, An interior-point method for large-scale l_1-regularized logistic regression, J. Mach. Learn. Res. 8 (8) (2007) 1519–1555.
[39] F.R. Chung, Spectral Graph Theory, vol. 92, AMS Bookstore, Providence, RI, 1997.
[40] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.
[41] T. Sim, S. Baker, M. Bsat, The cmu pose, illumination, and expression database, IEEE Trans. Pattern Anal. Mach. Intell. 25 (12) (2003) 1615–1618.
[42] Koh, Kwangmoo, Seungjean Kim, Stephen Boyd, l1 ls: A Matlab Solver for Large-Scale l1-Regularized Least Squares Problems, Stanford University, Stanford, CA, 2007.

Pengyue Zhang is currently a Ph.D student at Stony Brook University. He received the B.S. degree and M.S. degree at Huazhong University of Science and Technology, Wuhan, China. His research interests include medical image analysis computer vision and machine learning.

Xinge You received the B.S. and M.S. degrees in mathematics from the Hubei University, Wuhan, China and the Ph.D. degree from the Department of computer Science from the Hong Kong Baptist University, Hong Kong, in 1990, 2000, and 2004, respectively. Currently, he is a Professor at the Department of Electronics and Information Engineering in Huazhong University of Science and Technology, China. His current research interests include wavelets and its application, signal and image processing, pattern recognition, machine earning, and computer vision. Prof. You is the co-chair of technical committee on Pattern recognition-IEEE Systems, Man, and Cybernetics Society. He is now serving as an associate editor of IEEE Transactions on Cybernetics and a guest editor of many journals.

Weihua Ou is an associate professor at the School of Mathematics and Computer Science at Guizhou Normal University, Guiyang, China. He received the M.S. degree in Mathematics from the Southeast University (SEU), Nanjing, China, in 2006 and received the Ph.D. degree in information and communication engineering at Huazhong University of Science and Technology (HUST), Wuhan, China, in 2014. His research interests include pattern recognition, machine learning and computer vision, such as the application of sparse or low rank representation in image processing and computer vision.

C.L. Philip Chen is currently a Dean and a Chair Professor of the Faculty of Science and Technology, University of Macau. He has been a Professor and the Chair of the Department of Electrical and Computer Engineering, an Associate Dean for Research and Graduate Studies of the College of Engineering, University of Texas at San Antonio, Texas. His current research interests include theoretic development in computational intelligence, intelligent systems, cyber-physical systems, robotics and manufacturing automation, networking, diagnosis and prognosis, and life prediction and life-extending control. He has been the President of the IEEE Systems, Man and Cybernetics Society (SMCS), 2012-2013 and currently is the Editor-in-Chief of the IEEE Transactions on Systems, Man, and Cybernetics: Systems. He is a member of Tau Beta Pi and Eta Kappa Nu honor societies and has been the Faculty Advisor for Tau Beta Pi Engineering honor society. In addition, he is an ABET (Accreditation Board of Engineering and Technology Education) Program Evaluator for Computer Engineering, Electrical Engineering, and Software Engineering programs.

Yiu-ming Cheung is a Full Professor at the Department of Computer Science in Hong Kong Baptist University. He received Ph.D. degree at Department of Computer Science and Engineering from the Chinese University of Hong Kong. His current research interests focus on artificial intelligence, pattern recognition, visual computing, and optimization. Prof. Cheung is the Founding Chairman of Computational Intelligence Chapter of IEEE Hong Kong Section. Also, he is now serving as an Associate Editor of IEEE Transactions on Neural Networks and Learning Systems, Knowledge and Information Systems, and International Journal of Pattern Recognition and Artificial Intelligence, among others. He is a senior member of IEEE and ACM. More details can be found at: "http://www.comp.hkbu.edu.hk/%7Eymc" http://www.comp.hkbu.edu.hk/~ymc.