Pattern Recognition 89 (2019) 91–107

Contents lists available at ScienceDirect

Pattern Recognition

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

Robust heterogeneous discriminative analysis for face recognition with single sample per person

Meng Pang[a], Yiu-ming Cheung[a,*], Binghui Wang[b], Risheng Liu[c]

[a] Department of Computer Science, Hong Kong Baptist University, Hong Kong SAR, China
[b] Department of Electrical and Computer Engineering, Iowa State University, USA
[c] School of Software, Dalian University of Technology, Dalian, China

ARTICLE INFO

Article history:
Received 13 May 2018
Revised 16 October 2018
Accepted 4 January 2019
Available online 5 January 2019

Keywords:
Face recognition
Single sample per person
Heterogeneous representation
Fisher-like criterion
Joint majority voting

ABSTRACT

Single sample per person face recognition is one of the most challenging problems in face recognition (FR), where only single sample per person (SSPP) is enrolled in the gallery set for training. Although the existing patch-based methods have achieved great success in FR with SSPP, they still have limitations in feature extraction and identification stages when handling complex facial variations. In this work, we propose a new patch-based method called Robust Heterogeneous Discriminative Analysis (RHDA), for FR with SSPP. To enhance the robustness against complex facial variations, we first present a new graph-based Fisher-like criterion, which incorporates two manifold embeddings, to learn heterogeneous discriminative representations of image patches. Specifically, for each patch, the Fisher-like criterion is able to preserve the reconstruction relationship of neighboring patches from the same person, while suppressing the similarities between neighboring patches from the different persons. Then, we introduce two distance metrics, i.e., patch-to-patch distance and patch-to-manifold distance, and develop a fusion strategy to combine the recognition outputs of above two distance metrics via a joint majority voting for identification. Experimental results on various benchmark datasets demonstrate the effectiveness of the proposed method.

1. Introduction

Face recognition (FR) has been receiving considerable attentions in both computer vision and pattern recognition because of its potential applications in video surveillance [1–3], access control [4,5], person re-identification [6,7], visual tracking [8,9], just to name a few. In spite of great achievement in the past decades, FR is still becoming a challenging task due to many types of facial variations in a query face, e.g., illuminations, shadows, poses, expressions, disguises, occlusions, and misalignments [10].

In many practical FR systems, e.g., law enforcement, ID card identification, and airport surveillance, there is only one single sample per person (SSPP) when considering their limited storage and privacy policy [11]. As a result, it becomes particularly intractable for FR with SSPP when within-class information is not available to predict the unknown facial variations in query samples. Therefore, a variety of existing discriminative subspace learning methods such as linear discriminant analysis (LDA) [12] and other Fisher-based methods [13–15] would fail to work in such a scenario. Moreover, the recent emerging representation-based classifiers, e.g., sparse representation-based classifier (SRC) [16] and collaborative representation-based classifier (CRC) [17], also suffer from heavy performance degeneration, since these classifiers still require multiple within-class training samples to reasonably represent query samples.

To address the SSPP problem in FR, various methods have been developed recently, which can be roughly classified into two categories [18]: holistic methods and local methods. Holistic methods [19–22] identify a query sample using the whole face image as input. For holistic methods, the main idea is to enlarge training samples to acquire within-class information. As described in [23], there are two main directions: virtual sample generation and generic learning. Virtual sample generation synthesizes virtual samples by virtue of the real training samples. For example, SPCA [24] and SVD-LDA [25] generate virtual samples based on singular value decomposition (SVD) perturbation. Nevertheless, one major shortcoming of these methods is that the virtual samples are always highly correlated to the gallery samples and thus can hardly be considered as independent samples for feature extraction [23]. In contrast with virtue sample-based methods, generic learning methods usually introduce an auxiliary generic set with

* Corresponding author.
E-mail addresses: mengpang@comp.hkbu.edu.hk (M. Pang), ymc@comp.hkbu.edu.hk (Y.-m. Cheung), binghuiw@iastate.edu (B. Wang), rsliu@dlut.edu.cn (R. Liu).

https://doi.org/10.1016/j.patcog.2019.01.005
0031-3203/

persons not of interest to supplement the raw SSPP gallery set. For example, Wang et al. [26] proposed a generic learning framework to estimate approximated within-class scatter from generic set provided that different sets of persons share similar within-class variations. Representative methods under this framework include extended SRC (ESRC) [19], superposed SRC (SSRC) [27], sparse variation dictionary learning (SVDL) [20], collaborative probabilistic labels (CPL) [28], etc. Although this kind of holistic methods can alleviate the SSPP problem to some extent, their performance depends heavily on the elaborately selected generic set. For example, the desired generic set is always required to (1) share similar shooting situations with the gallery set, and (2) contain adequate facial variations to help predict the unknown variations in query samples. However, in practical applications, it would be a tough task to collect sufficient generic samples satisfying such requirements.

For local methods, they recognize a query sample by leveraging local facial features. Usually, a common way to generate local features is to partition a face sample into several overlapping/non-overlapping image patches. Thus, this type of local methods [29–32] is also called *patch-based* methods, in which each partitioned patch of a face sample is assumed to be an independent sample of this person (i.e., class). Based on this assumption, researchers extended conventional subspace learning methods and representation-based classifiers, e.g., PCA, LDA, SRC, and CRC, to the corresponding patch-based counterparts, i.e., modular PCA [33], modular LDA [34], patch-based SRC (PSRC) [16] and patch-based CRC (PCRC) [29]. Subsequently, they performed SSPP FR via integrating the recognition outputs of all partitioned patches. Furthermore, Lu et al. [23] developed a discriminative multi-manifold analysis (DMMA) method provided that the partitioned patches of each person lie in an individual manifold, hence converting FR to a manifold-manifold matching problem. Based on this work, Yan et al. [35] proposed a multi-feature multi-manifold learning method by combining multiple local features to promote recognition performance. Zhang et al. [31] modified DMMA and proposed a sparse discriminative multi-manifold embedding (SDMME) method by leveraging another sparse graph-based Fisher criterion to learn a discriminative subspace for partitioned patches.

Recently, a few attempts [36,37] have been made to incorporate generic learning into patch-based methods for SSPP FR. For example, Zhu et al. [36] extracted the patch variation dictionary from the generic set, then concatenated them with the gallery patch dictionary to measure the representation residual of each query patch. These methods have been reported to achieve much better performance compared to the existing patch-based methods for SSPP FR. However, the desired generic set of them is still difficult to be collected in practice, like generic learning methods. Therefore, we only focus on the patch-based methods without generic learning in this work.

Despite inspiring performance achieved by the existing patch-based methods for SSPP FR, these methods still suffer from two major drawbacks:

(i) *For feature extraction*, the graph-based Fisher criteria applied in the state-of-the-art patch-based methods, i.e., DMMA and SDMME, cannot generate representations (i.e., features) that are discriminative enough. Note that LE-graph [38] and l_1-graph [39] are two prevalent graphs that characterize the similarity relationships and reconstruction relationships for image data, respectively. Then, for the above two patch-based methods,
- DMMA preserves LE-graph for the within-class patches, meanwhile destroying LE-graph for the between-class patches (see Criterion I in Fig. 1). In doing so, the neighboring within-class patches will be pulled close to each other while the neighboring between-class patches will be kept far apart in the learned subspace. Nevertheless, in this criterion, the crucial reconstruction structure of the within-class patches is neglected, which may cause the recovered subspace structure to be skewed from the intrinsic subspace structure.
- SDMME preserves l_1-graph for the within-class patches, meanwhile destroying l_1-graph for the between-class patches (see Criterion II in Fig. 1). In this criterion, although the reconstruction structure and the similarity relationships of the within-class patches both can be reserved, the between-class patches may still have chance to stay nearby because destroying the between-class reconstruction relationship is a weaker penalty compared to directly suppressing the similarities of the between-class patches. Moreover, it would be time-consuming to compute the reconstruction coefficients of the between-class patches in SDMME, as the number of between-class patches is much larger than the number of within-class patches.

(ii) *For identification*, it is believed that, given a patch from a query sample, it should be (1) similar to the patch in the same position, or/and (2) well reconstructed by its neighboring patches, of the same person in the gallery set. However, existing patch-based methods only consider one of the two observations (i.e., distance metrics), which is inadequate when handling complex facial variations. For example, PSRC and PCRC simply leverage the similarities of patches in the same position to identify a query patch. However, when there exist pose variations or misalignments in the query sample, the patches of the same position between query and gallery sample do not match to each other probably, thus leading the query patch to be easily misclassified. Besides, DMMA and SDMME simply compute the reconstruction residual between the query patch and its neighboring patches in the gallery sample for identification. However, this reconstruction-based distance metric is quite sensitive to the facial variations such as severe illuminations and shadows in query samples.

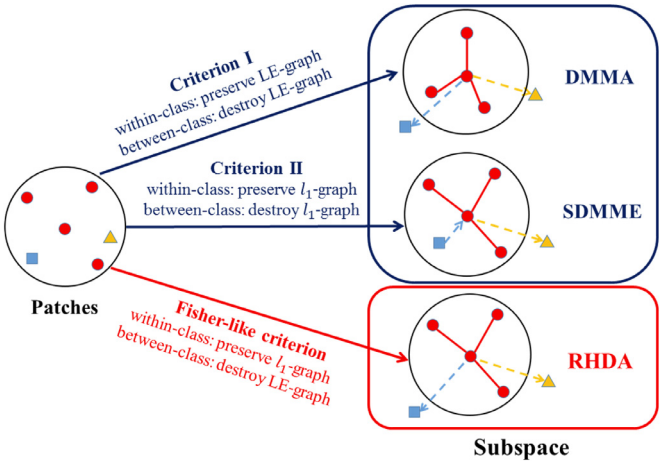

Fig. 1. Comparison of the proposed Fisher-like criterion in RHDA with the graph-based Fisher criteria in DMMA and SDMME. The points with the same color (or shape) indicate patches from the same person.

To address the above two issues, we propose a new patch-based method called Robust Heterogeneous Discriminative Analysis (RHDA), for FR with SSPP.

For the first issue, based on the purposes of the reconstruction-based l_1-graph and the similarity-based LE-graph, we propose a

new graph-based Fisher-like criterion in RHDA model to conduct discriminant analysis across both l_1-graph and LE-graph. The former preserves the sparse reconstruction relationship of the neighboring patches from the same person, and the latter suppresses the similarities of the neighboring patches from the different persons (see Fisher-like criterion in Fig. 1), so as to improve the discriminant ability of the patch distributions in the learned subspace compared with that of DMMA and SDMME. To the best of our knowledge, the Fisher-like criterion can be the first attempts to consider the cooperation of heterogeneous graphs to characterize the discriminant structure of image data. We will demonstrate the superiority of the proposed Fisher-like criterion in Section 2.1.3.

For the second issue, we present two different discriminative manifold embeddings, namely discriminative single-manifold embedding (DSME) and discriminative multi-manifold embedding (DMME). The two embeddings model the whole partitioned patches over all persons as a single manifold and multiple manifolds, respectively, and are then incorporated into the Fisher-like criterion to generate heterogeneous discriminative representations for image patches. Subsequently, we introduce two distance metrics, i.e., patch-to-patch distance and patch-to-manifold distance, associated with the single manifold and multiple manifolds, respectively; and develop a fusion strategy by assigning the heterogeneous representations to the two distance metrics and combining their recognition outputs via joint majority voting to identify the unlabeled query sample. In doing so, the proposed RHDA method can greatly enhance the robustness against complex facial variations and achieve promising recognition performance. Experimental results on five benchmark datasets, i.e., AR, CAS-PEAL, FERET, E-YaleB and Multi-PIE, verify the effectiveness of RHDA for FR with SSPP.

Moreover, it is worth noting that the deep learning based methods [40–44], e.g., DeepID [40], VGG-Face [41] and stacked denoising auto-encoders [42], have achieved great success in face verification and identification. Benefiting from these works, some attempts have been tried recently to employ the deep neural networks to address the SSPP FR problem. For instance, Gao et al. [45] proposed a stacked supervised auto-encoders (SSAE) method, in which the faces with different variations were treated as the contaminated samples, then a stacked denoising auto-encoders based deep neural network was leveraged to recover the clean part of the contaminated samples as well as to extract their common features for image representation. However, in SSAE, the training set that trained the network was directly partitioned from the whole evaluated dataset, which is not applicable from a practical view point and may also cause over-fitting problem. In contrast, Parchami et al. [46] and Yang et al. [47] utilized the convolutional neural networks (CNNs) to extract the deep features of input images (or image patches), and collected external face datasets in the web to train the networks, which could benefit the generalization ability of the deep model. Motivated by these, in this work, we also consider to apply the pre-trained CNNs to generate high-semantic features for the gallery and query samples, and explore the feasibility of combining our RHDA with the deep features to address the practical SSPP FR problem.

We highlight the contributions of our work as follows:

- We propose a new patch-based method called RHDA for single sampler per person face recognition.
- We develop a new graph-based Fisher-like criterion to conduct discriminant analysis across both l_1-graph and LE-graph, so as to improve the discriminative ability of patch distribution in the learned subspace.
- We present a joint majority voting strategy to both consider the patch-to-patch and patch-to-manifold distances for face identification.

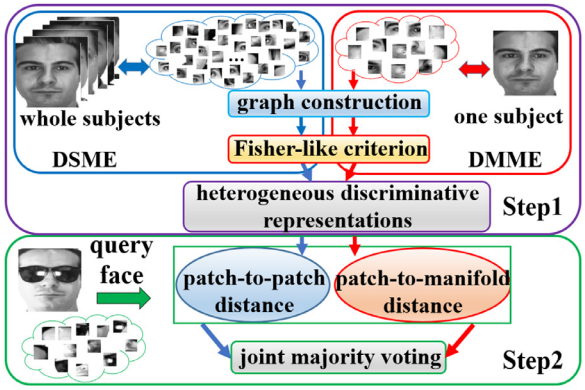

Fig. 2. The flowchart of the proposed RHDA method. The left blue arrows and the right red arrows represent the flows of the DSME associated with patch-to-patch distance metric and DMME associated with patch-to-manifold distance metric in RHDA, respectively.

Compared with our preliminary work in [48], this paper has made four major extensions: (1) We have improved the patch-to-patch distance metric by additionally leveraging the contributions of the neighboring patches at current position, to enhance the robustness against the mismatch between gallery and query samples. (2) We have further evaluated the performance of DSME with patch-to-patch distance and DMME with patch-to-manifold distance, respectively, and compared them with that of RHDA, to verify the effectiveness of the fusion strategy, i.e, joint majority voting, in face identification stage. (3) We have analyzed the computational complexity and studied the parameter sensitivity of RHDA. (4) We have conducted extensive experiments to evaluate the performance of RHDA, and compared it with other state-of-the-art methods, including the popular deep learning based methods on SSPP FR.

The reminder of the paper is organized as follows. Section 2 introduces the proposed RHDA method in detail. Section 3 evaluates the performance of RHDA, and provides the experimental results. Section 4 discusses other possible pattern recognition applications of RHDA. Finally, Section 5 concludes the paper.

2. The proposed method

This section presents the proposed RHDA in two steps: *heterogeneous feature extraction* and *face identification*. For heterogeneous feature extraction, we first construct an intrinsic graph and a penalty graph, then propose two discriminative manifold embeddings, i.e., DSME and DMME, and finally leverage a Fisher-like criterion to generate heterogeneous discriminative subspace representations for image patches. For face identification, we introduce two distance metrics, i.e., patch-to-patch distance and patch-to-manifold distance, and develop a fusion strategy to exploit the heterogeneous subspace representations and identify each unlabeled query sample via a joint majority voting. As described above, the flowchart of RHDA is illustrated in Fig. 2.

2.1. Heterogeneous feature extraction

2.1.1. Graph construction

Suppose $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N] \in \Re^{D\times N}$ is a gallery set with N persons, where $\mathbf{x}_i$ is the image of the ith person. We first partition each $\mathbf{x}_i$ into M non-overlapping local patches with an equal size d, then concatenate the patches column by column. For example, we define the patch set of the ith person as $\mathbf{X}_i = [\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \cdots, \mathbf{x}_{i,M}] \in \Re^{d\times M}$. Subsequently, we leverage l_1-graph and LE-graph to construct an intrinsic graph G and penalty graph G', respectively. $\mathbf{S}^w$

and $\mathbf{S}^b$ denote the corresponding reconstruction weight and affinity weight matrices in G and G', respectively.

In the intrinsic graph G, we aim to measure the representation ability of patches from the same person (i.e., class). Hence, we first design a within-class dictionary for each patch (e.g., $\mathbf{x}_{i,j}$) as follows:

$$\mathbf{A}_{i,j} = \mathbf{X}_i / \mathbf{x}_{i,j} = [\mathbf{x}_{i,1}, \cdots, \mathbf{x}_{i,j-1}, \mathbf{x}_{i,j+1}, \cdots, \mathbf{x}_{i,M}]. \tag{1}$$

Then, the representation coefficients of the remaining within-class patches for $\mathbf{x}_{i,j}$ can be computed as follows:

$$\boldsymbol{\alpha}_{i,j} = \arg\min_{\boldsymbol{\alpha}_{i,j}} ||\mathbf{x}_{i,j} - A_{ij}\boldsymbol{\alpha}_{i,j}||_F^2 + ||\boldsymbol{\alpha}_{i,j}||_1. \tag{2}$$

Therefore, the within-class reconstruction weight matrix $\mathbf{W}_i$ for the ith person is defined as follows:

$$\mathbf{W}_i = [\mathbf{W}_{i,1}, \cdots, \mathbf{W}_{i,j}, \cdots, \mathbf{W}_{i,M}] \in \Re^{M \times M}, \tag{3}$$

$$\mathbf{W}_{i,j}^p = \begin{cases} \boldsymbol{\alpha}_{i,j}^p & 0 < p < j \\ 0 & p = j \\ \boldsymbol{\alpha}_{i,j}^{p-1} & j < p < M, \end{cases} \tag{4}$$

where $\mathbf{W}_{i,j}^p$ denotes the pth element of $\mathbf{W}_{i,j}$. Hence, $\mathbf{S}^w$ for whole patches over all persons can be defined as $\mathbf{S}^w = \text{diag}(\mathbf{W}_1, \cdots, \mathbf{W}_i, \cdots, \mathbf{W}_N) \in \Re^{MN \times MN}$.

In the penalty graph G', we aim to measure the similarity of patches from different persons. For each $\mathbf{x}_{i,j}$, we let $\mathbf{x}_{i,j}^p$ represent its pth neighboring patch, and calculate the affinity weight between $\mathbf{x}_{i,j}$ and other patches as

$$\widehat{\mathbf{W}}_{i,j}^p = \begin{cases} exp(-\frac{||\mathbf{x}_{i,j} - \mathbf{x}_{i,j}^p||^2}{\sigma^2}) & \text{if } \mathbf{x}_{i,j}^p \in N_{k_1}(\mathbf{x}_{i,j}) \\ 0 & \text{otherwise}, \end{cases} \tag{5}$$

where $N_{k_1}(\mathbf{x}_{i,j})$ denote the k_1-nearest between-class patches of $\mathbf{x}_{i,j}$. Then, we let $\widehat{\mathbf{W}} \in \Re^{MN \times MN}$ represent the affinity weight matrix for the total MN patches by assigning the value of each $\widehat{\mathbf{W}}_{i,j}^p$ into the corresponding position. Hence, $\mathbf{S}^b$ in graph G' can be directly set as $\mathbf{S}^b = \widehat{\mathbf{W}} \in \Re^{MN \times MN}$.

2.1.2. Discriminative manifold embeddings

We propose discriminative single-manifold embedding (DSME) and discriminative multi-manifold embedding (DMME), respectively, as follows.

DSME: It models the whole patch set over all persons as a single manifold. For simplicity, we define the whole patch set as: $\widehat{\mathbf{X}} = [\widehat{\mathbf{x}}_1, \cdots, \widehat{\mathbf{x}}_q, \cdots, \widehat{\mathbf{x}}_{MN}] \in \Re^{d \times MN}$, where $\widehat{\mathbf{x}}_q = \mathbf{x}_{i,j}, i = \lceil \frac{q}{M} \rceil, j = q - Mi + M$. Then, on one hand, we aim to preserve the reconstruction structure of neighboring within-class patches. On the other hand, we also expect to suppress the similarities of neighboring patches from different classes. Formally, we can achieve the target by learning a shared projection basis $\mathbf{U} \in \Re^{d \times r}$ for all patches and optimizing the following two objective functions generated in graph G and G', respectively:

$$\min \Phi^w(\mathbf{U}) = \sum_i \|\mathbf{U}^T\widehat{\mathbf{x}}_i - \Sigma_j \mathbf{S}_{ij}^w \mathbf{U}^T\widehat{\mathbf{x}}_j\|^2, \tag{6}$$

$$\max \Phi^b(\mathbf{U}) = \sum_{i,j} \|\mathbf{U}^T\widehat{\mathbf{x}}_i - \mathbf{U}^T\widehat{\mathbf{x}}_j\|^2 \mathbf{S}_{ij}^b. \tag{7}$$

DMME: It models the whole patch set as a collection of multiple manifolds, and supposes that patches of each subject lie in an individual manifold. As a result, a set of N projection bases $\mathbf{V} = \{\mathbf{V}_1, \mathbf{V}_2, \cdots, \mathbf{V}_N\}$ will be learned for N persons. Formally, we need to optimize the following two objective functions:

$$\max J_1(\mathbf{V}_i) = \sum_{j=1}^{M} \sum_{p=1}^{k_1} \|\mathbf{V}_i^T\mathbf{x}_{i,j} - \mathbf{V}_i^T\mathbf{x}_{i,j}^p\|_F^2 \widehat{\mathbf{W}}_{i,j}^p, \tag{8}$$

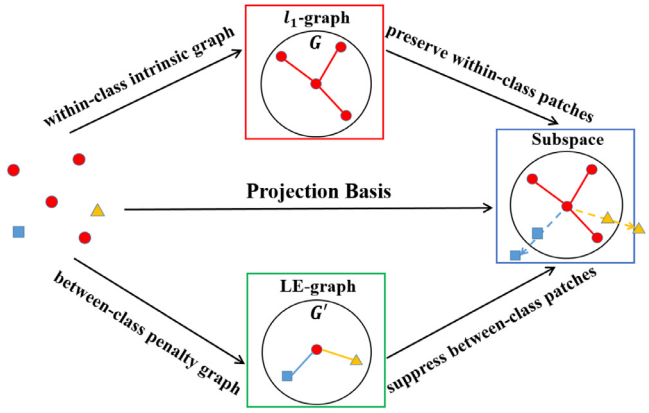

Fig. 3. Illustration of the Fisher-like criterion. The points with the same color (or shape) indicate the patches from the same person, while the points with different colors (or shapes) indicate the patches of different persons.

$$\min J_2(\mathbf{V}_i) = \sum_{j=1}^{M} \|\mathbf{V}_i^T\mathbf{x}_{i,j} - \mathbf{V}_i^T\mathbf{X}_i\mathbf{W}_{i,j}\|_F^2, \tag{9}$$

where J_1 generated from G' ensures that, if $\mathbf{x}_{i,j}$ and $\mathbf{x}_{i,j}^p$ are close but from different subjects, they should be separated as far as possible after projection. Moreover, J_2 from G is to preserve the reconstruction relationship of the within-class neighboring patches after projection.

2.1.3. Feature extraction via a fisher-like criterion

We develop a new Fisher-like criterion to extract discriminative features across the two heterogeneous adjacency graphs, i.e., l_1-graph and LE-graph. Specifically, it aims to simultaneously preserve the reconstruction relationship of neighboring within-class patches in l_1-graph, while suppressing neighboring patches of different classes in LE-graph. We provide the illustration of the Fisher-like criterion in Fig. 3.

Then, for DSME, Eqs. (6) and (7) are first rewritten in the following forms:

$$\begin{aligned} \min \Phi^w(\mathbf{U}) &= \|(\mathbf{I} - \mathbf{S}^w)\mathbf{U}^T\widehat{\mathbf{X}}\|_F^2 \\ &= tr\{\mathbf{U}^T\widehat{\mathbf{X}}[\mathbf{I} - (\mathbf{S}^w)^T](\mathbf{I} - \mathbf{S}^w)\widehat{\mathbf{X}}^T\mathbf{U}\} \\ &= tr(\mathbf{U}^T\widehat{\mathbf{X}}\mathbf{M}^w\widehat{\mathbf{X}}^T\mathbf{U}), \\ \max \Phi^b(\mathbf{U}) &= \sum_{i,j} tr\{[\mathbf{U}^T(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)(\widehat{\mathbf{x}}_i - \widehat{\mathbf{x}}_j)^T\mathbf{U}]\mathbf{S}_{ij}^b\} \\ &= 2tr[\mathbf{U}^T\widehat{\mathbf{X}}(\mathbf{D}^b - \mathbf{S}^b)\widehat{\mathbf{X}}^T\mathbf{U}] \\ &= 2tr(\mathbf{U}^T\widehat{\mathbf{X}}\mathbf{L}^b\widehat{\mathbf{X}}^T\mathbf{U}), \end{aligned}$$

where $\mathbf{M}^w = (\mathbf{I} - \mathbf{S}^w)^T(\mathbf{I} - \mathbf{S}^w)$, $\mathbf{D}^b$ is a diagonal matrix with $\mathbf{D}_{ii}^b = \Sigma_j \mathbf{S}_{ij}^b$, $\mathbf{L}^b = \mathbf{D}^b - \mathbf{S}^b$ is the Laplacian matrix.

By incorporating the proposed Fisher-like criterion, the final objective function for DSME becomes:

$$\max_{\mathbf{U}} \frac{\Phi^b(\mathbf{U})}{\Phi^w(\mathbf{U})} = \frac{tr(\mathbf{U}^T\widehat{\mathbf{X}}\mathbf{L}^b\widehat{\mathbf{X}}^T\mathbf{U})}{tr(\mathbf{U}^T\widehat{\mathbf{X}}\mathbf{M}^w\widehat{\mathbf{X}}^T\mathbf{U})}, \tag{10}$$

Then, the maximization problem in Eq. (10) can be transformed to the following generalized eigen-decomposition problem:

$$\widehat{\mathbf{X}}\mathbf{L}^b\widehat{\mathbf{X}}^T\mathbf{U} = \lambda\widehat{\mathbf{X}}\mathbf{M}^w\widehat{\mathbf{X}}^T\mathbf{U}, \tag{11}$$

where $\{\lambda_i\}_{i=1}^r$ denote the r largest positive eigenvalues with $\lambda_1 \geq \cdots \lambda_r > 0$, and $\mathbf{U}$ are the corresponding eigenvectors. Since the dimension of patches (i.e, d) is always smaller than the number

of total patches (i.e., MN), the matrices $\widehat{\mathbf{X}}\mathbf{L}^b\widehat{\mathbf{X}}^T$ and $\widehat{\mathbf{X}}\mathbf{M}^w\widehat{\mathbf{X}}^T$ can be nonsingular and the eigen-problem in Eq. (11) will be solved stably. For the extreme case that $d > MN$, we will adopt the SVD+LGE (Linear Graph Embedding) approach [49] to stably solve the above eigen-problem. After obtaining the shared projection basis $\mathbf{U}$, the subspace representation for each patch $\mathbf{x}_{i,j}$ is defined as $\mathbf{U}^T\mathbf{x}_{i,j}$.

For DMME, we define the Fisher-like criterion as

$$\max_{\mathbf{V}} J(\mathbf{V}) = \sum_{i=1}^{N}(J_1(\mathbf{V}_i) - J_2(\mathbf{V}_i)). \tag{12}$$

$J_1(\mathbf{V}_i)$ and $J_2(\mathbf{V}_i)$ can be simplified as follows:

$$\begin{aligned} J_1(\mathbf{V}_i) &= tr\{\mathbf{V}_i^T[\sum_{j=1}^{M}\sum_{p=1}^{k_1}(\mathbf{x}_{i,j} - \mathbf{x}_{i,j}^p)(\mathbf{x}_{i,j} - \mathbf{x}_{i,j}^p)^T\widehat{\mathbf{W}}_{i,j}^p]\mathbf{V}_i\} \\ &= tr(\mathbf{V}_i^T\mathbf{H}_1\mathbf{V}_i), \\ J_2(\mathbf{V}_i) &= tr[\mathbf{V}_i^T\sum_{j=1}^{M}(\mathbf{x}_{i,j} - \mathbf{X}_i\mathbf{W}_{i,j})(\mathbf{x}_{i,j} - \mathbf{X}_i\mathbf{W}_{i,j})^T\mathbf{V}_i] \\ &= tr(\mathbf{V}_i^T\mathbf{H}_2\mathbf{V}_i), \end{aligned}$$

where

$$\mathbf{H}_1 = \sum_{j=1}^{M}\sum_{p=1}^{k_1}(\mathbf{x}_{i,j} - \mathbf{x}_{i,j}^p)(\mathbf{x}_{i,j} - \mathbf{x}_{i,j}^p)^T\widehat{\mathbf{W}}_{i,j}^p, \tag{13}$$

$$\mathbf{H}_2 = \sum_{j=1}^{M}(\mathbf{x}_{i,j} - \mathbf{X}_i\mathbf{W}_{i,j})(\mathbf{x}_{i,j} - \mathbf{X}_i\mathbf{W}_{i,j})^T. \tag{14}$$

Note that the N projection bases are independent and thus $J(\mathbf{V})$ can be simply computed as the sum of N subfunctions $J_1(\mathbf{V}_i) - J_2(\mathbf{V}_i)$ of each $\mathbf{V}_i$, which can be separately solved via the following eigen-decomposition problem:

$$(\mathbf{H}_1 - \mathbf{H}_2)\mathbf{v} = \lambda\mathbf{v}. \tag{15}$$

Let $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{d_i}$ be the eigenvectors corresponding to the d_i largest positive eigenvalues $\{\lambda_j\}_{j=1}^{d_i}$ with $\lambda_1 \geq \cdots \geq \lambda_{d_i} > 0$. Then, the projection basis for the ith class is indicated as $\mathbf{V}_i = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{d_i}]$, and the subspace representation for each patch $\mathbf{x}_{i,j}$ is represented as $\mathbf{V}_i^T\mathbf{x}_{i,j}$.

Fig. 4 illustrates the partitioned image patches of six face subjects (each image is partitioned into 64 patches) from FERET dataset in the original manifold, and the subspaces learnt by DMMA [23], SDMME [31] and our DMME. For the sake of observation, we utilize the powerful visualization tool, i.e., t-SNE [50], to show the resulting map as a three-dimensional plot. From Fig. 4, it can be observed that:

- First, the patches in the original manifold are rather scattered, and there is a high overlapping among the manifolds corresponding to different subjects.
- Second, the scattering of patches in the subspaces from DMMA and SDMME are better than those in the original manifold, but there are still a small amount of overlapping patches among different subjects.
- Third, there is a clear separation for the patches of different subjects in the subspace learnt by our DMME, and the separability between different class clusters are better than those in DMMA and SDMME.

The inspiring results empirically verify the rationality of the Fisher-like criterion, and also demonstrate its superior discriminant ability over the graph-based Fisher criteria adopted in DMMA and SDMME methods.

2.2. Face identification

In this section, we introduce the patch-to-patch and patch-to-manifold distances, and present a fusion strategy based on the two distance metrics for identification.

2.2.1. Patch-to-patch distance

Given a query face sample $\mathbf{y}$, we partition it into M non-overlapping local patches $\{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_M\}$. Subsequently, inspired by PCRC [29], we introduce the patch-to-patch distance as the *first distance metric*, and utilize regularized least square to identify each query patch $\mathbf{y}_j$. Specifically, we apply the shared projection basis $\mathbf{U} \in \Re^{d\times r}$ generated by DSME to project $\mathbf{y}_j$ and each patch $\mathbf{x}_{i,j}$ of the same position in the gallery set into a common subspace, and construct a local patch dictionary $\mathbf{D}_j$ as:

$$\mathbf{D}_j = [\mathbf{U}^T\mathbf{x}_{1,j}, \cdots, \mathbf{U}^T\mathbf{x}_{i,j} \cdots, \mathbf{U}^T\mathbf{x}_{N,j}] \in \Re^{r\times N}, \tag{16}$$

where $\mathbf{U}^T\mathbf{x}_{i,j}$ denotes the subspace representation for the jth patch of the ith person in the gallery set.

To further enhance the representation ability of the local dictionary $\mathbf{D}_j$ and to better handle the mismatch (e.g., misalignment or pose variation) between gallery and query samples, we extract the neighboring patches of $\mathbf{x}_{i,j}$ at current position to supplement the local dictionary $\mathbf{D}_j$. Considering the facts that 1) large value of neighborhood size will introduce lots of irrelevant patches into the gallery patch dictionary to degrade the recognition performance, and 2) increase the recognition time cost, we thus set the value of the neighborhood size to be 1 in the following experiments. Specifically, the extraction strategy is presented as follows (refer to Fig. 5):

- First, as shown in Fig. 5(a), when $\mathbf{x}_{i,j}$ lies in the corner of the image, the neighboring 3 patches are extracted.
- Second, as shown in Fig. 5(b), when $\mathbf{x}_{i,j}$ lies in the edge (not the corner) of the image, the neighboring 5 patches are extracted.
- Third, as shown in Fig. 5 (c), when $\mathbf{x}_{i,j}$ does not lie in the edge of the image, the neighboring 8 patches are extracted.

Consequently, with such a dictionary expansion, the stability and robustness of the patch-to-patch distance metric can be improved.

Next, for $\mathbf{U}^T\mathbf{y}_j$, its representation coefficients over $\mathbf{D}_j$ are computed by

$$\widehat{\boldsymbol{\rho}}_j = \arg\min_{\boldsymbol{\rho}_j}\{\|\mathbf{U}^T\mathbf{y}_j - \mathbf{D}_j\boldsymbol{\rho}_j\|^2 + \lambda\|\boldsymbol{\rho}_j\|^2\}, \tag{17}$$

where $\widehat{\boldsymbol{\rho}}_j = [\widehat{\boldsymbol{\rho}}_{j,1}; \widehat{\boldsymbol{\rho}}_{j,2}; \cdots; \widehat{\boldsymbol{\rho}}_{j,N}]$. Hence, the identification output of the query patch $\mathbf{y}_j$ is defined as:

$$L^f(\mathbf{y}_j) = \arg\min_{k}\{\|\mathbf{U}^T\mathbf{y}_j - \mathbf{D}_{j,k}\widehat{\boldsymbol{\rho}}_{j,k}\|^2/\|\widehat{\boldsymbol{\rho}}_{j,k}\|^2\}. \tag{18}$$

2.2.2. Patch-to-manifold distance

Furthermore, we introduce the patch-to-manifold distance as the *second distance metric*, which targets to measure the reconstruction capability of the reference manifold. As depicted in DMME, the patch set $\mathbf{X}_i$ for the ith person is treated as an individual manifold. Then, the distance between the query patch $\mathbf{y}_j$ and $\mathbf{X}_i$ can be computed by

$$d(\mathbf{y}_j, \mathbf{X}_i) = \min\|\mathbf{V}_i^T\mathbf{y}_j - \Sigma_{p=1}^{k_2}\mathbf{c}_p G_{k_2}^p(\mathbf{V}_i^T\mathbf{y}_j)\|^2, \tag{19}$$

where $\mathbf{V}_i \in \Re^{d\times d_i}$ is the projection basis generated by DMME, $G_{k_2}^p(\mathbf{V}_i^T\mathbf{y}_j)$ denotes the pth member of k_2-nearest neighbors of $\mathbf{V}_i^T\mathbf{y}_j$ in $\mathbf{V}_i^T\mathbf{X}_i$, and $\mathbf{c}_p$ represents the reconstruction coefficient corresponding to $G_{k_2}^p(\mathbf{V}_i^T\mathbf{y}_j)$. For this distance metric, the identification output of the query patch $\mathbf{y}_j$ is computed by

$$L^s(\mathbf{y}_j) = \arg\min_{k}\ d(\mathbf{y}_j, \mathbf{X}_k). \tag{20}$$

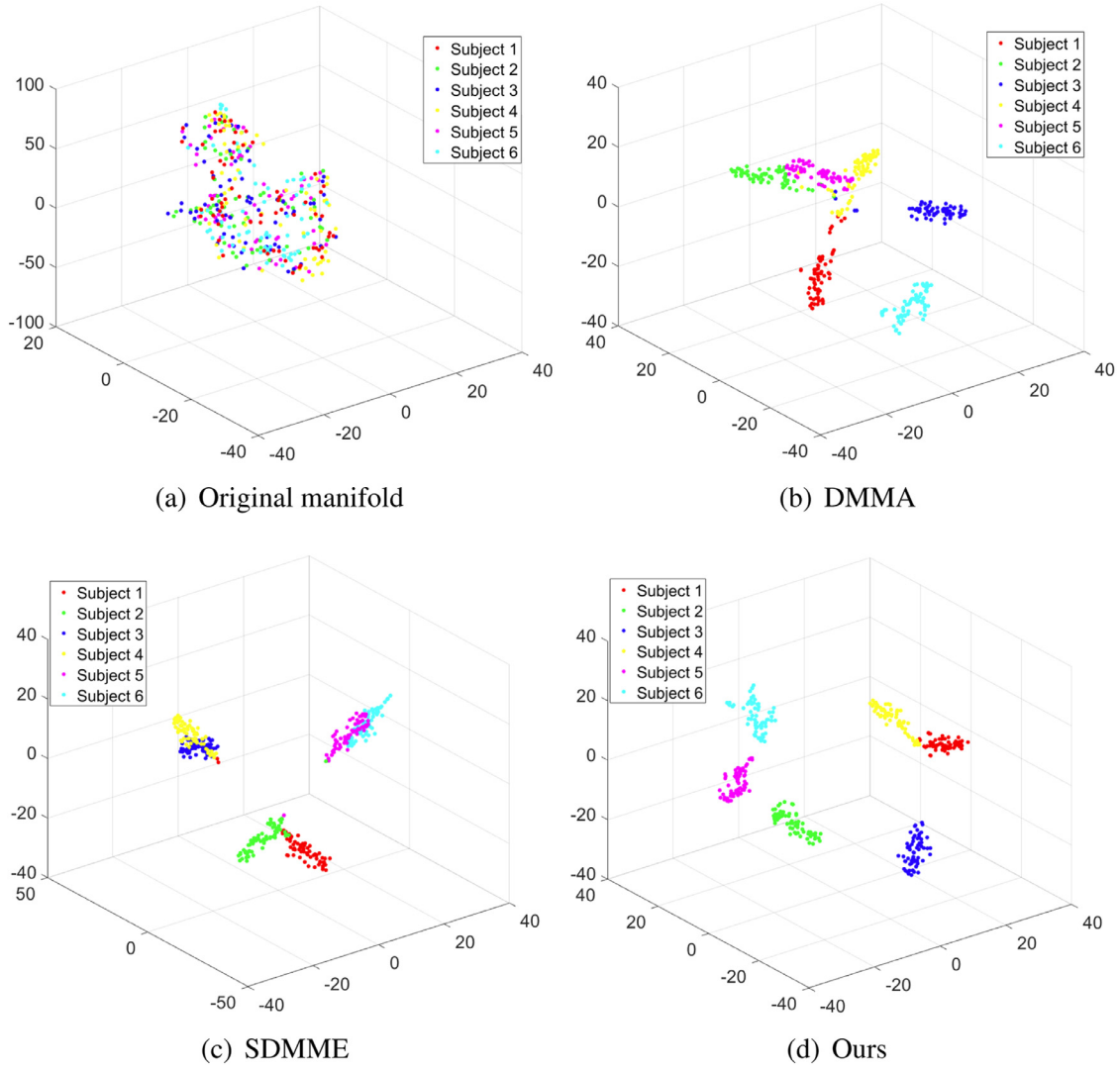

Fig. 4. Visualization of patch distributions in the (a) original manifold, and subspaces learnt by (b) DMMA, (c) SDMME and (d) our DMME.

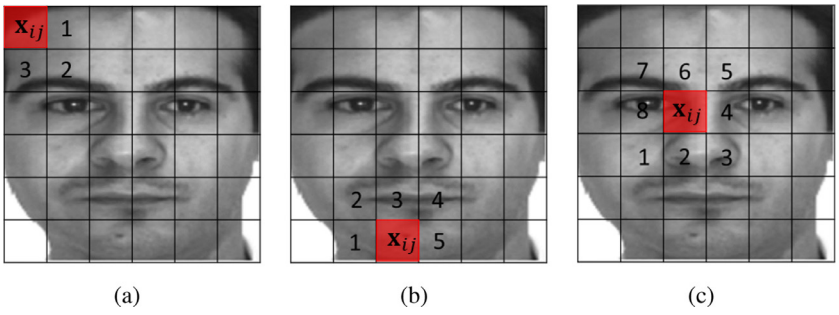

Fig. 5. Extraction strategy of neighboring patches for expanding the local dictionary **D**. The highlighted box indicates the position of the patch $\mathbf{x}_{ij}$.

2.2.3. Joint majority voting

In the final stage, we aim to identify the unlabeled query sample $\mathbf{y}$ by exploiting the identification outputs of all the query patches $\{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_M\}$. One should note that $L^f(\mathbf{y}_j)$ and $L^s(\mathbf{y}_j)$ obtained by two distance metrics may be different. Therefore, it would be difficult to decide which output should be adopted. To this end, we thus present a fusion strategy by leveraging both the outputs of two distance metrics and determine the final label of the query sample via a joint majority voting. Moreover, in the joint majority voting, we attempt to expand the voting rights from the top-1 predicted label to the top-k predicted label set for the two distance metrics, as it is believed that the correct label of the query patch is more likely to lie within the top-k predicted label candidates.

Specifically, we let $\{L_i^f(\mathbf{y}_j)\}_{i=1}^{T_1}$ and $\{L_i^s(\mathbf{y}_j)\}_{i=1}^{T_2}$ be the top-T_1 and top-T_2 predicted label set for the query patch $\mathbf{y}_j$ from the patch-to-patch and patch-to-manifold distance metrics, respectively, and define $\mathbf{vote}^f, \mathbf{vote}^s \in \Re^N$ as the initial zero vectors. Then, the label

of the query sample $\mathbf{y}$, i.e., $L(\mathbf{y}) \in \Re^N$, can be determined using the joint majority voting presented in Algorithm 1.

Algorithm 1 Joint majority voting.

Require: $\{L_i^f(\mathbf{y}_j)\}_{i=1}^{T_1}, j = 1, \cdots, M$; $\{L_i^s(\mathbf{y}_j)\}_{i=1}^{T_2}, j = 1, \cdots, M$; zero vectors $\mathbf{vote}^f, \mathbf{vote}^s \in \Re^N$

Ensure: $L(\mathbf{y}) \in \Re^N$

```
for j = 1 : M do
    for i = 1 : T_1 do
        vote^f(L_i^f(y_j)) = vote^f(L_i^f(y_j)) + 1
    end for
    for i = 1 : T_2 do
        vote^s(L_i^s(y_j)) = vote^s(L_i^s(y_j)) + 1
    end for
end for
L(y) = arg max_k(vote^f(k) + vote^s(k)), k = 1, ..., N
```

In fact, the proposed joint majority voting can be interpreted as a special case of ensemble learning involving two weak classifiers, i.e., DSME with patch-to-patch distance and DMME with patch-to-manifold distance. When facing with simplex facial variations such as expression, slight illumination and disguises, the votes for the correct label can be further reinforced in this fusion since the two distance metrics are both robust against these variations. Moreover, when facing with other challenging facial variations such as misalignment, pose and severe illumination, or the combinations of multiple variations, maybe not both of the voting vectors for patch-to-patch and patch-to-manifold distance metrics would be discriminative. However, the fusion of the two distance metrics can still (1) generate complementary information, and (2) increase the error tolerance, for identification, which is believed to achieve more stable and better performance than that using any of single distance metric.

3. Experimental results

In this section, five experiments in total in the subsequent subsections are performed to show the effectiveness of the proposed RHDA method. Specifically, in Section 3.1, we evaluate the performance of RHDA for FR with SSPP, on AR, FERET, CAS-PEAL, E-YaleB and Multi-PIE datasets. In Section 3.2, we evaluate the performance of the two discriminative manifold embeddings, i.e., DSME and DMME, respectively, and verify the effectiveness of the joint majority voting for face identification. In Section 3.3, we study the parameter sensitivity of RHDA. In Section 3.4, we analyze the computational complexity of RHDA. Lastly, in Section 3.5, we evaluate the performance of RHDA by combining it with the deep learning-based features on the unconstrained Labeled Faces in the Wild (LFW) dataset. All experiments are carried out on a host (CPU: Dual 4-core Intel Xeon X5570 2.93GHz 8MB L3 Cache, RAM: 32GB).

3.1. Performance of RHDA for SSPP FR

In this experiment, our purpose is to evaluate the performance of our RHDA for FR with SSPP. Subsequently, we perform FR experiments on five popular benchmark face datasets, including AR, FERET, CAS-PEAL, E-YaleB and Multi-PIE datasets.

Comparing algorithms: We compare our RHDA with 13 representative holistic and local methods that are used to address the SSPP FR problem, including PCA [51], $(PC)^2A$ [52], 2DPCA [53], laplacianfaces [54], representation-based classifiers, i.e., SRC and CRC, virtual sample-based method, i.e., SVD_LDA, generic learning methods, i.e., ESRC, and the state-of-the-art SVDL and CPL, and patch-based methods, i.e., DMMA and the state-of-the-art PCRC and SDMME. Among the 13 comparing methods, i.e., PCA, $(PC)^2A$, 2DPCA, Laplacianfaces, SVD_LDA, SRC, CRC, ESRC, PCRC, DMMA, SDMME, SVDL and CPL, we have implemented $(PC)^2A$, SVD_LDA, DMMA, SDMME and CPL by ourselves, and the codes of other 8 methods are obtained from the original authors.

Parameter setting: In SSPP FR experiments, the face images were resized to 48×48 pixels on AR, FERET, CAS-PEAL and E-YaleB and Multi-PIE datasets. For $(PC)^2A$, the weighting parameter α was set as 0.25. For SVD_LDA, the first three singular values and the corresponding singular vectors were applied to synthesize virtual samples. For laplacianfaces, the number of nearest neighbors k was selected as 3 to construct the adjacency graph. For all the patch-based methods such as PCRC, DMMA, SDMME and our RHDA, the *non-overlapping* patch size was set as 8×8 pixels for a fair comparison. In addition, the other values of parameters k_1, k_2, k, and σ in DMMA were empirically tuned to be 30, 2, 2, and 100, respectively. For SDMME, the l_1-ls toolbox was used to solve its l_1-minimization problem as suggested in [31], and the balance factor λ was tuned to be 0.001. For SRC, CRC, PCRC, ESRC, the values of the regularization parameter λ were searched from {0.001, 0.005, 0.01, 0.05, 0.1} to achieve the best results over five evaluated datasets. For SVDL and CPL, the parameters were set according to the suggestions in [20] and [28], respectively. Specifically, the parameters λ_1, λ_2 and λ_3 of SVDL were set to be 0.001, 0.01 and 0.0001, respectively, and the parameters λ, δ_1, δ_2, τ_1 and τ_2 of CPL were set to be 0.01, 0.3, 3, 1.618, 1.618, respectively. As to our RHDA, the value of the parameter k_1 in Eq. (8) was empirically set as N (the number of gallery subjects), k_2 in Eq. (19), λ in Eq. (17), and σ in Eq. (5) were fixed as 2, 0.001, and 1, respectively. The values of the combinations of T_1&T_2 in joint majority voting were fixed as $T_1 = T_2 = 1$ over five evaluated datsets except two cases on FERET and E-YaleB datasets, where we will describe their settings in the following experiments.

3.1.1. Evaluation on AR dataset

The AR dataset [55] consists of over 4000 frontal face images of 126 people from two sessions, and each session has 13 face images per subject, which involve different variations of facial expressions, illuminations and disguises (i.e., sunglasses and scarf). Following the setting in [18], the first 80 subjects from Session-1 were used for evaluation, while another 20 subjects were randomly selected from the remaining set in the same Session as the generic set for generic learning methods. The standard face images taken with neutral expression and under uniform illumination were selected to form the gallery set, while the rest 12 images of each subject were arranged to form 5 probe sets b-f (i.e., expression, illumination, sunglasses+illumination, scarf+illumination and disguises). Furthermore, to make the experiment more challenging, we designed a new probe set g by adding random block occlusion (i.e., 30% occlusion) into the expression probe set. The gallery sample and the 6 probe sets of one subject on AR dataset are shown in Fig. 6.

Table 1 lists the recognition results of all the methods on AR dataset. From Table 1, we have the following observations. First, RHDA achieves the best recognition performance in all cases we have tried (prob sets b-g). Second, RHDA outperforms the state-of-the-art generic learning CPL and SVDL methods. For example, for the variations in expression (probe set b) and disguises&occlusion (probe sets d-g), RHDA improves the recognition accuracies of CPL by 11.66%, 14.58%, 29.17%, 17.75% and 22.91%, respectively. In addition, for the variation with simplex illumination (probe set c), although SVDL and CPL have already achieved quite high performance (i.e., 94.58%), RHDA still outperforms these methods and obtains 97.00% recognition accuracy. Third, RHDA consistently outperforms the state-of-the-art patch-based PCRC method. Specifically, compared with PCRC, RHDA delivers 10.83%, 4.00%, 16.67%, 16.25%, 6.88% and 15.83% of improvements in probe sets b-g, re-

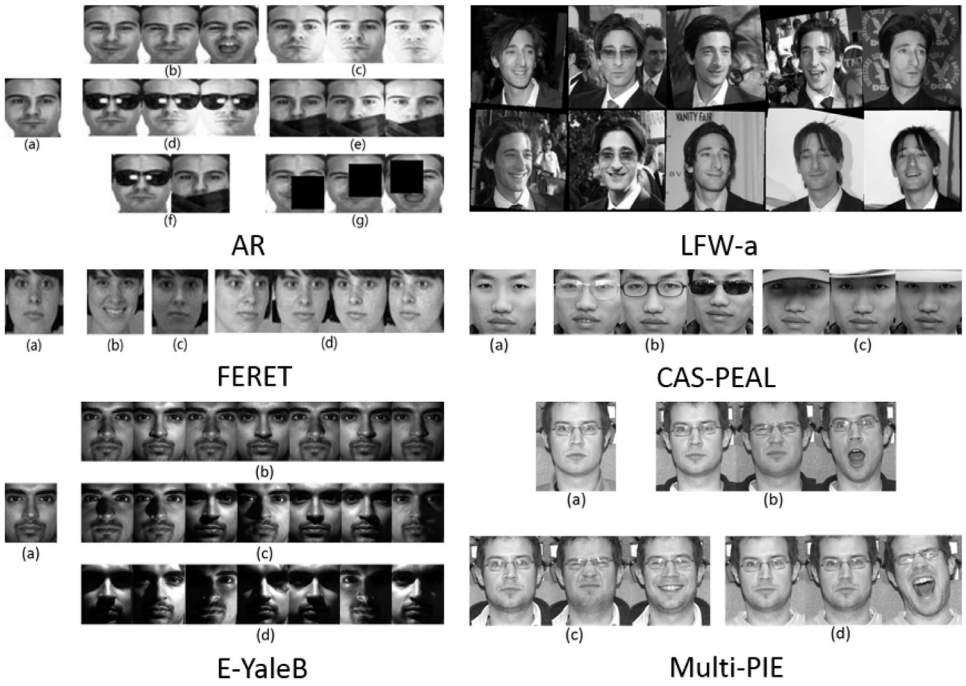

Fig. 6. The gallery and probe samples of one subject on the AR, FERET, CAS-PEAL, E-YaleB, Multi-PIE and LFW-a datasets, respectively.

Table 1
Recognition accuracy (%) on AR dataset (**Best**; *Second Best*).

Methods	Probe set b	Probe set c	Probe set d	Probe set e	Probe set f	Probe set g
PCA [51]	78.75	62.92	33.33	10.42	38.75	32.08
$(PC)^2A$ [52]	79.17	60.00	37.08	10.00	35.00	20.58
2DPCA [53]	83.33	68.33	38.33	12.50	40.00	26.25
Laplacianfaces [54]	77.08	73.75	45.83	15.83	48.13	24.17
SVD_LDA [25]	75.52	55.00	38.33	14.58	40.63	33.75
SRC [16]	85.42	77.08	46.67	24.58	53.75	39.58
CRC [17]	82.92	76.67	45.00	27.08	54.37	33.33
ESRC [19]	83.33	93.75	76.25	60.00	76.25	44.58
PCRC [29]	*86.25*	93.00	79.58	*79.17*	*92.50*	51.25
DMMA [23]	84.17	56.42	48.75	44.17	74.38	*55.83*
SDMME [31]	85.42	58.83	50.42	45.42	76.25	55.42
SVDL [20]	84.58	*94.58*	80.83	62.83	81.37	43.75
CPL [28]	85.42	*94.58*	*81.67*	66.25	81.63	44.17
RHDA	**97.08**	**97.00**	**96.25**	**95.42**	**99.38**	**67.08**

spectively. Forth, the recent patch-based methods such as DMMA and SDMME perform poor under the variations containing illumination (probe sets c-e). However, compared with the holistic representation-based classifiers such as ESRC, CRC and SRC, SDMME and DMMA have shown to be more robust against the random block occlusion (prob set g). Fifth, as to the other holistic methods like laplacianfaces, SVD_LDA, $(PC)^2A$ and 2DPCA, they acquire similar recognition results with PCA in most cases.

3.1.2. Evaluation on FERET dataset

The FERET dataset [56] is sponsored by the US Department of Defense through the DARPA Program, and consists of 14,126 images from 1199 subjects. In this experiment, we aim to evaluate the robustness of RHDA against the facial variations of poses, illuminations and expressions on FERET dataset. To this end, we selected 700 face images of 100 subjects from seven subsets (ba, bj, bk, bd, be, bf and bg) on FERET dataset. Following the strategy on AR dataset, we also utilized the first 80 subjects for evaluation, while the remaining 20 subjects were chosen as the generic set. The neutral images of all subjects were used to construct the gallery set, and the rest 6 images were arranged to form 3 probe sets b-d (i.e., expression, illumination and poses). Please note that, the values of T_1&T_2 of RHDA for probe set d were empirically set as $T_1 = 1, T_2 = 10$. The gallery sample and the 3 prob sets of one subject on FERET dataset are shown in Fig. 6.

Table 2 presents the recognition results of all the methods on FERET dataset. It is clear that RHDA again performs the best in the three cases. For example, compared with the second best method in each case, RHDA improves the recognition accuracies by 2.87%, 5.62% and 6.56%, respectively. Furthermore, we are interested to find that, the performance of PCRC degrades seriously for the variance of pose (prob set d). It may be because that, the pose variations always result in mismatch of the corresponding patches, and simply considering the patch-to-patch distance may lead PCRC to make misjudgment when identifying a query patch. By contrast, our RHDA exhibits greater robustness against pose variation as well as other facial variations compared with PCRC and other comparing methods owing to two important factors. On the one hand, the Fisher-like criterion in RHDA can extract highly discriminant information hidden in partitioned patches and meanwhile improves the discriminative ability of patch distribution in underlying subspaces; on the other hand, RHDA considers both the patch-to-patch

Table 2
Recognition accuracy (%) on FERET dataset (**Best**; *Second Best*).

Methods	Probe set b	Probe set c	Probe set d
PCA [51]	72.50	73.75	48.44
$(PC)^2A$ [52]	75.62	68.13	44.69
2DPCA [53]	78.13	75.00	51.56
Laplacianfaces [54]	72.50	71.25	22.81
SVD_LDA [25]	70.00	63.75	30.00
SRC [16]	75.00	72.50	43.13
CRC [17]	74.38	73.75	41.56
ESRC [19]	81.25	80.00	48.13
PCRC [29]	77.50	76.25	21.25
DMMA [23]	81.25	52.50	49.69
SDMME [31]	82.50	52.50	*54.69*
SVDL [20]	*84.38*	*80.63*	54.37
CPL [28]	83.75	79.37	50.06
RHDA	**87.25**	**86.25**	**61.25**

Table 3
Recognition accuracy (%) on CAS-PEAL dataset (**Best**; *Second Best*).

Methods	Probe set b	Probe set c
PCA [51]	70.33	27.11
$(PC)^2A$ [52]	70.44	24.22
2DPCA [53]	73.11	27.56
Laplacianfaces [54]	66.44	44.00
SVD_LDA [25]	69.78	39.33
SRC [16]	77.11	37.78
CRC [17]	78.22	59.78
ESRC [19]	81.78	65.56
PCRC [29]	64.89	66.89
DMMA [23]	71.00	38.44
SDMME [31]	70.78	36.00
SVDL [20]	*84.00*	63.11
CPL [28]	82.89	*69.33*
RHDA	**85.33**	**85.56**

and patch-to-manifold distances via joint majority voting, which can generate complementary information and increase the error tolerance for identification.

3.1.3. Evaluation on CAS-PEAL dataset

The CAS-PEAL dataset [57] contains 99,594 images of 1040 subjects (595 males and 445 females) with variations including expression, facing direction, accessory, lighting, age, etc. In this experiment, we aim to further evaluate the performance of RHDA under facial occlusions, as CAS-PEAL is considered to be the largest public dataset with occluded face images available. We utilized 200 subjects from the Normal and the Accessory categories of CAS-PEAL, thus each subject has 1 neutral image, 3 images with different glasses and 3 images with different hats. The first 150 subjects were used for evaluation, and another 50 subjects were selected as the generic set. The neutral images of all subjects were used to construct the gallery set, and the rest 6 images were arranged to form 2 probe sets b-c (i.e., glasses and hats). The gallery sample and the 2 probe sets of one subject on CAS-PEAL dataset are shown in Fig. 6.

Table 3 lists the recognition results of all the methods on CAS-PEAL dataset, where we can observe that RHDA still performs the best in both cases. Please note that, in CAS-PEAL dataset, except the glasses and hats disguises, there could also exist some slight misalignments because of the manually cropping. Moreover, the hats in some face images would bring about illumination variations or even shadows that hinder recognition (see Fig. 6). Consequently, the patch-based methods such as SDMME and DMMA perform poor in both cases. In contrast, the holistic generic learning methods such as CPL, SVDL and ESRC are less sensitive to illuminations and shadows, and obtain good results on CAS-PEAL by introducing the supplementary information (e.g., facial variations of wearing similar glasses or hats) from the generic set to help predict the query variations. Nevertheless, benefiting from the Fisher-like criterion and the joint majority voting strategy, RHDA still achieves promising recognition performance and outperforms the state-of-the-art generic learning CPL and SVDL methods even without the help of auxiliary generic set. The inspiring results demonstrate the effectiveness of RHDA when handling combinations of multiple facial variations.

Table 4
Recognition accuracy (%) on E-YaleB dataset (**Best**; *Second Best*).

Methods	Probe set b	Probe set c	Probe set d
PCA [51]	93.17	62.92	25.36
$(PC)^2A$ [52]	53.75	68.13	23.57
2DPCA [53]	61.25	75.00	25.36
Laplacianfaces [54]	64.17	71.25	22.50
SVD_LDA [25]	85.00	27.08	14.64
SRC [16]	96.67	56.67	15.00
CRC [17]	95.83	52.50	15.00
ESRC [19]	*99.83*	95.33	61.43
PCRC [29]	**100.00**	93.33	66.43
DMMA [23]	99.17	41.67	17.14
SDMME [31]	99.17	39.83	15.43
SVDL [20]	**100.00**	*98.58*	*71.07*
CPL [28]	**100.00**	98.33	70.71
RHDA	**100.00**	**99.17**	**74.29**

3.1.4. Evaluation on E-YaleB dataset

The Extended YaleB (E-YaleB) dataset [58] consists of 2414 images of 38 subjects under various illumination conditions, which can be divided into five subsets (i.e., 7, 12, 12, 14 and 19 images per subject). Subset 1 is under normal illumination condition (lighting angle: 0°-12°), Subsets 2–3 describe slight-to-moderate luminance variations (lighting angle: 13°–25° and 26°–50°), and Subsets 4–5 characterize severe illumination variations (lighting angle: 51°–77° and $>77°$). In this experiment, we aim to test the performance of RHDA under different illumination angles. Therefore, we selected the first 20 subjects for evaluation, and the rest 18 subjects were used as the generic set. The first image of each subject in Subset 1 was chosen as the gallery sample, while the samples in Subsets 2–4 were formed as 3 probe sets b-d. It is worth mentioning that, in probe set d, the values of T_1&T_2 of RHDA were set to be $T_1 = 5, T_2 = 1$. The gallery sample and the 3 probe sets of one subject on E-YaleB dataset are shown in Fig. 6.

Table 4 shows the recognition results of different methods on E-YaleB dataset. We can observe that, as the illumination angle increases, the performance of all the methods will degrade in different degrees. However, our RHDA still achieves the best performance in the three cases, and performs slightly better than the generic learning methods such as CPL and SVDL. Besides, we are interested to find the patch-based DMMA and SDMME suffer heavy performance decline under large illumination angle (i.e., prob set d), while the patch-based PCRC could maintain relative good results in this case. The plausible reason is that, both of DMME and SDMME are based on the patch-to-manifold distance, which may perform instable under large global variations such as severe illuminations; while PCRC leverages the patch-to-patch distance for identification, and this distance metric can be robust against such challenging variations. Moreover, RHDA improves 4.57%, 39.68% and 38.49%, w.r.t. the average recognition rates, over PCRC, SDMME and DMMA, respectively, which also explains that the fusion of the patch-to-patch and patch-to-manifold distance metrics can enhance the performance compared with that of using any single distance metric.

Table 5
Recognition accuracy (%) on Multi-PIE dataset (**Best**; *Second Best*).

Methods	Probe set b	Probe set c	Probe set d
PCA [51]	50.42	46.67	51.67
$(PC)^2A$ [52]	45.42	68.13	49.17
2DPCA [53]	50.42	46.67	51.67
Laplacianfaces [54]	47.50	71.25	44.58
SVD_LDA [25]	46.67	43.75	45.83
SRC [16]	55.42	53.75	56.67
CRC [17]	60.42	48.33	52.92
ESRC [19]	67.50	61.67	61.67
PCRC [29]	72.92	*80.42*	*67.50*
DMMA [23]	66.25	65.83	63.75
SDMME [31]	60.17	61.00	60.17
SVDL [20]	*73.75*	65.83	66.67
CPL [28]	72.50	65.42	*67.50*
RHDA	**85.42**	**92.92**	**75.42**

3.1.5. Evaluation on Multi-PIE dataset

The Multi-PIE dataset [59] is a comprehensive face dataset of 337 subjects with each containing faces images across 6 expressions (i.e., neutral, smile, surprise, squint, disgust and scream) in 4 different sessions, 15 poses, and 20 illuminations. In this subsection, we aim to evaluate the robustness of RHDA against different expressions in different shooting scenarios. Hence, in the experiment, we selected 120 subjects in expression subset of 4 different sessions, where the first 80 subjects were used for evaluation and the rest 40 subjects were used as generic set. The neutral image of each subject in session 1 was chosen as gallery sample, while the rest 9 images were formed as three probe sets b-d. The gallery sample and the 3 probe sets of one subject on Multi-PIE dataset are also illustrated in Fig. 6.

Table 5 presents the results of all the methods on Multi-PIE dataset. On this dataset, all the comparing patch-based methods including PCRC, SDMME and DMMA are observed to achieve goodish performance for probe sets b-d, which are comparable with or even better than that of the state-of-the-art generic learning methods such as CPL and SVDL. This is because the patch-to-patch distance metric in PCRC, and the patch-to-manifold distance metric in SDMME&DMMA are both robust against the local variations (e.g, expressions) in the three cases, even though the shooting sessions have been changed. In a similar fashion, it is believed that the fusion of the two distance metrics in our RHDA can further reinforce the discriminative ability of the joint voting results, thus making the predicted label in recognition stage more correct and reliable. Such analysis can be empirically verified by the superior performance of our RHDA in Table 5, where RHDA has delivered obvious improvements of 11.67%, 12.50% and 7.92% over the second best method in the three cases, respectively.

3.1.6. Summary

- RHDA consistently achieves promising recognition performance in all cases on AR, FERET, CAS-PEAL, E-YaleB and Multi-PIE datasets, which confirms the effectiveness of RHDA when dealing with complex facial variations in query samples.
- Compared with the state-of-the-art patch-based methods such as PCRC and SDMME, RHDA greatly enhances the recognition performance in all cases over five evaluated datasets, especially for occlusions and pose variation.
- Even without the help of auxiliary generic set, the patch-based RHDA still outperforms the state-of-the-art generic learning CPL and SVDL methods in almost all cases over five evaluated datasets.

3.2. Performance of DSME and DMME for SSPP FR

In this subsection, we evaluate the recognition performance of DSME and DMME on AR, FERET, CAS-PEAL, E-YaleB and Multi-PIE datasets, respectively. Note that, for DSME, the improved patch-to-patch distance is used for identification; while for DMME, the patch-to-manifold distance is employed. In the experiment, we first verify the effectiveness of the improved patch-to-patch distance, by comparing the performance of our DSME in this paper with that of DSME in [48] (using original patch-to-patch distance metric) over five evaluated datasets. As shown in Fig. 7, through extracting the neighboring patches in gallery samples to expand the local dictionary, the improved patch-to-patch distance metric enables DSME to perform better than DSME in [48]. Specifically, DSME improves the average recognition accuracies over DSME in [48] by 9.42%, 7.04%, 5.99%, 5.43% and 17.00% on AR, FERET, CAS-PEAL, E-YaleB and Multi-PIE datasets, respectively.

Moreover, we further test whether the joint majority voting combining the patch-to-patch and patch-to-manifold distance metrics can further enhance the robustness against complex query variations. To this end, we compare the performance of DSME, DMME and their fusion, i.e, RHDA, over five datasets in Table 6. It can be observed that, each of DSME and DMME has its own advantages and could handle different variations. For example, DSME shows greater robustness against illumination compared with DMME, but is more sensitive to pose variation and misalignment. Nevertheless, by reasonably combining the patch-to-patch distance metric in DSME and the patch-to-manifold distance metric in DMME via the joint majority voting, RHDA can take advantages of the two distance metrics to further enhance the robustness and achieve the best recognition performance in all probe cases over five evaluated datasets.

3.3. Parameter sensitivity study

This subsection studies the sensitivity of the parameters of our RHDA model. Note that the parameter σ in Eq. (5) has little effects on the performance of RHDA over five datasets, we thus show the results of the remaining parameters such as k_1 in Eq. (8), k_2 in Eq. (19), λ in Eq. (17), the combinations of $T_1\&T_2$ (T_1, $T_2 > 0$), and the patch size. Figs. 8–12 show the effects of k_1, k_2 and λ on the recognition accuracies of RHDA over AR, FERET, CAS-PEAL, E-YaleB and Multi-PIE datasets, respectively. It can be seen that the recognition performance of RHDA is insensitive to the selection of the above three parameters, when the values of k_1, k_2 and λ are set within the ranges from $N/4$ to $4N$, 1 to 5, and 0.001 to 0.01, respectively.

Fig. 13(a)–(o) illustrate the effects of the combination of $T_1\&T_2$ on 15 cases over five evaluated datasets, where we can observe that the performance of RHDA *changes little* when tuning the combinations of $T_1\&T_2$ in most cases on the five datasets. However, in few special cases, slightly expanding the value of T_1 or T_2 helps enhance the performance of RHDA. For example, for probe set d (i.e., pose variation) on FERET, increasing the value of T_2 could boost the performance of RHDA. It is because that patch-to-manifold distance metric is robust against the misalignments and pose variations, while patch-to-patch distance metric is not stable in this case due to the mismatch of the corresponding patches. In contrast, for probe set d (illuminations angles: 51°–77°) on E-YaleB, increasing the value of T_1 could benefit the final performance, as patch-to-patch distance metric is more robust against severe illuminations compared with patch-to-manifold distance metric.

Moreover, we empirically probe the effect of patch size on the accuracy of RHDA. The performance of RHDA using patches with different sizes on AR, FERET, CAS-PEAL, E-YaleB and Multi-PIE datasets are shown in Fig. 14, where we can see that RHDA is, to some extent, sensitive to the patch size. However, we also notice that RHDA always performs the best when the patch size is chosen as 8×8. A plausible explanation is that, the 8×8 patches usually cover the semantically meaningful parts of the face (taking 48×48

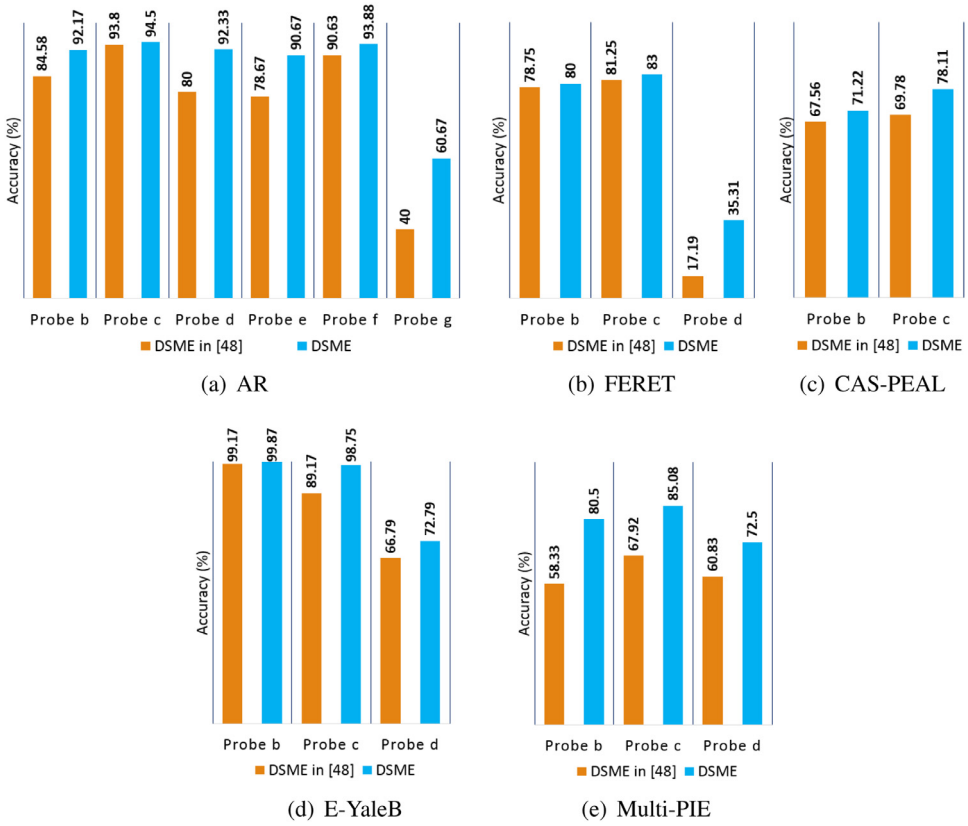

Fig. 7. The comparisons of our DSME in this work and DSME in [48] on (a) AR, (b) FERET, (c) CAS-PEAL, (d) E-YaleB and (e) Multi-PIE datasets.

Table 6
The comparisons of DSME, DMME and RHDA under various types of variations. For each tested case, the method with the highest performance is marked with "**1st**" to indicate the most robust method, and the latter two are marked with "***2nd***" and "*3rd*", respectively, according to their performance.

Variation factors		DSME	DMME	RHDA
Expression	probe set b on AR	92.17% (*3rd*)	92.50% (***2nd***)	97.08% (**1st**)
	probe set b on FERET	80.00% (*3rd*)	81.25% (***2nd***)	87.25% (**1st**)
Illumination	probe set c on AR	94.50% (***2nd***)	63.33% (*3rd*)	97.00% (**1st**)
	probe set c on FERET	83.00% (***2nd***)	57.50% (*3rd*)	86.25% (**1st**)
	probe set b on E-YaleB	99.87% (***2nd***)	99.58% (*3rd*)	100.00% (**1st**)
	probe set c on E-YaleB	98.75% (***2nd***)	43.67% (*3rd*)	99.17% (**1st**)
	probe set d on E-YaleB	72.79% (***2nd***)	18.83% (*3rd*)	74.29% (**1st**)
Disguise	probe set f on AR	93.88% (***2nd***)	90.62% (*3rd*)	99.38% (**1st**)
Pose	probe set d on FERET	35.31% (*3rd*)	53.44% (***2nd***)	61.25% (**1st**)
Expression + Illumination	probe set b on Multi-PIE	80.50% (***2nd***)	72.08% (*3rd*)	85.42% (**1st**)
	probe set c on Multi-PIE	85.08% (***2nd***)	77.08% (*3rd*)	92.92% (**1st**)
Expression + Block occlusion	probe set g on AR	60.67% (***2nd***)	55.42% (*3rd*)	67.08% (**1st**)
Sunglasses + Illumination	probe set d on AR	92.33% (***2nd***)	73.33% (*3rd*)	96.25% (**1st**)
Scarf + Illumination	probe set e on AR	90.67% (***2nd***)	68.33% (*3rd*)	95.42% (**1st**)
Glasses + Misalignment	probe set b on CAS-PEAL	71.22% (*3rd*)	81.56% (***2nd***)	85.33% (**1st**)
Hats + Shadow	probe set c on CAS-PEAL	78.11% (***2nd***)	44.89% (*3rd*)	85.56% (**1st**)
Expression + Illumination + Pose	probe set d on Multi-PIE	72.50% (***2nd***)	67.50% (*3rd*)	75.42% (**1st**)

face image for example), like the eyes, the lips, the nose, and these partitioned patches could possess the most informative and discriminative information for identification. It is worth noting that, in practical applications, the optimal patch size is also affected by the aligning and cropping way of face images as well as the final cropped size.

3.4. Computational complexity analysis

In this subsection, we briefly analyze the computational complexity of the training phase in our RHDA model, which involves two heterogeneous discriminative embeddings called DSME and DMME, respectively. Let N be the number of samples in the training set, and each sample image is partitioned into M non-overlapping local patches with an equal size d. The sparse optimization problem in Eq. (2) of the graph construction step is solved via the basis pursuit de-nosing (BPDN)-homotopy algorithm [60], and the number of iterations is denoted as t. For DSME, the complexities of computing the within-class reconstruction weights and the between-class affinity weights in graph construction step can be $O(td^2N + tdMN)$ [61] and $O(d(MN)^2)$, respectively. Besides, the solving of the eigen-problem in Eq. (11) requires $O((MN)^3)$ [49]. Let $Q = MN$, then the time complexity of DSME can be $O(td^2N + tdQ + dQ^2 + Q^3)$. For DMME, its major difference with DSME lies in the representation generation step, which involves N independent eigen-problems in Eq. (15) and costs time of $O(NM^3)$

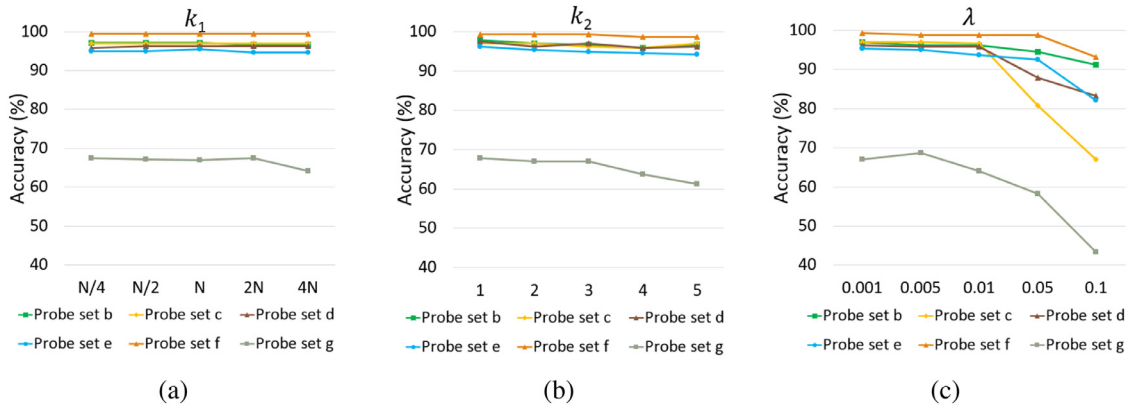

Fig. 8. (a) to (c) are the recognition accuracies of RHDA versus the parameters k_1, k_2 and λ on AR dataset. k_1 varies from $N/4$ to $4N$ (N indicates the number of gallery subjects), k_2 varies from 1 to 5 and λ varies from 0.001 to 0.1.

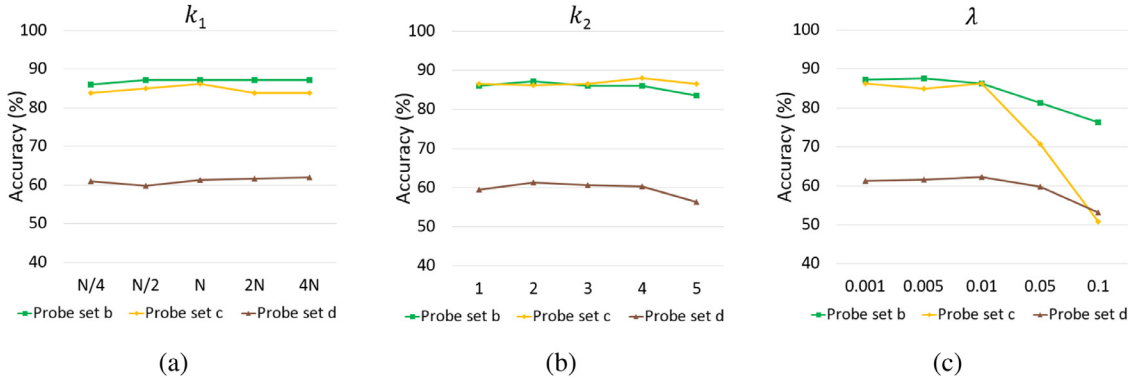

Fig. 9. (a) to (c) are the recognition accuracies of RHDA versus the parameters k_1, k_2 and λ on FERET dataset. k_1 varies from $N/4$ to $4N$ (N indicates the number of gallery subjects), k_2 varies from 1 to 5 and λ varies from 0.001 to 0.1.

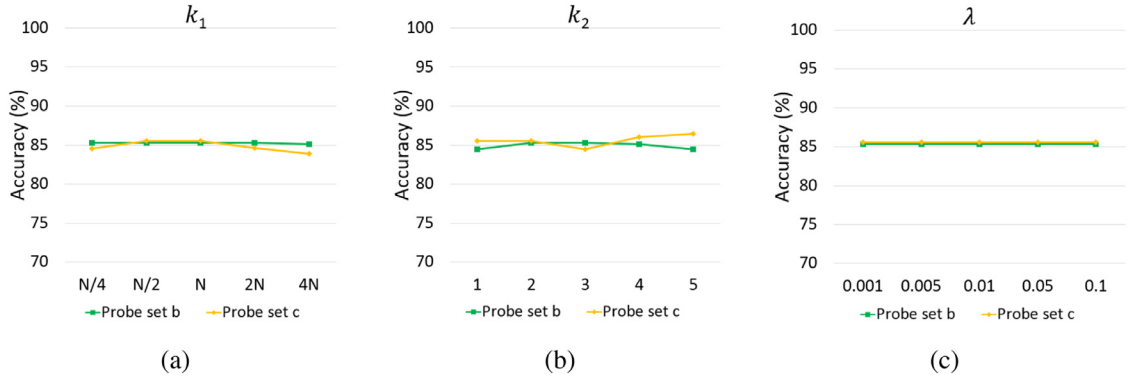

Fig. 10. (a) to (c) are the recognition accuracies of RHDA versus the parameters k_1, k_2 and λ on CAS-PEAL dataset. k_1 varies from $N/4$ to $4N$ (N indicates the number of gallery subjects), k_2 varies from 1 to 5 and λ varies from 0.001 to 0.1.

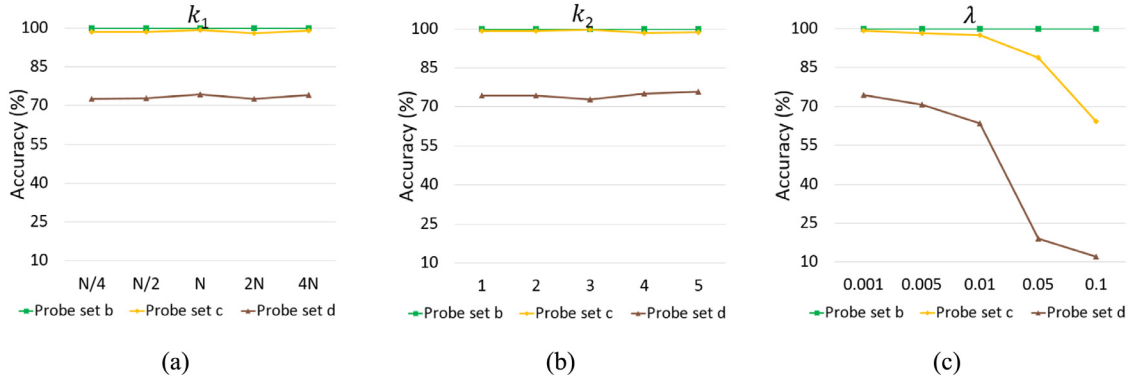

Fig. 11. (a) to (c) are the recognition accuracies of RHDA versus the parameters k_1, k_2 and λ on E-YaleB dataset. k_1 varies from $N/4$ to $4N$ (N indicates the number of gallery subjects), k_2 varies from 1 to 5 and λ varies from 0.001 to 0.1.

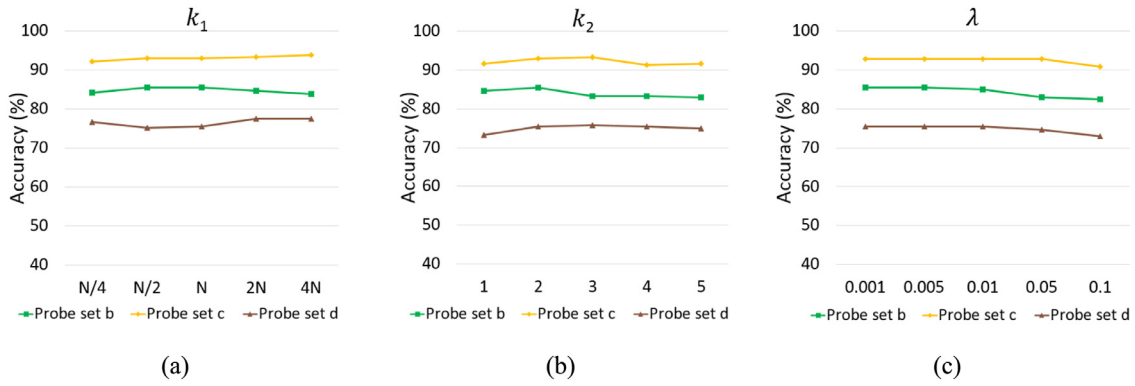

Fig. 12. (a) to (c) are the recognition accuracies of RHDA versus the parameters k_1, k_2 and λ on Multi-PIE dataset. k_1 varies from $N/4$ to $4N$ (N indicates the number of gallery subjects), k_2 varies from 1 to 5 and λ varies from 0.001 to 0.1.

Table 7
Training time (in seconds) used by different methods on AR dataset.

Methods	Training time
CPL [28]	0.9133
DMMA [23]	4.1767
SVDL [20]	75.1430
SDMME [31]	136.8938
RHDA	11.4568

in total. Hence, the time complexity of DSME is $O(td^2N + tdQ + dQ^2 + QM^2)$. Overall, since DSME and DMME in our RHDA model can be executed parallelly, the time complexity of RHDA can be $O(td^2N + tdQ + dQ^2 + Q^3)$. Furthermore, the memory complexity of RHDA is $O(MN(d + MN))$.

We also list the training time of the proposed RHDA method, and compare it with the other four popular SSPP FR methods, including two generic learning methods, i.e., CPL and SVDL, and two patch-based methods, i.e., DMMA and SDMME. Table 7 shows the time spent on the training phase by these methods, where the MATLAB R2016a software and AR dataset were used.

From Table 7, we can observe that, the training time of RHDA is not the most one, and is far less than that of SVDL and SDMME. A plausible reason is twofold: First, compared with SVDL, RHDA needs not perform time-consuming dictionary learning on the generic set. Second, compared with SDMME, the Fisher-like criterion enables RHDA to avoid the computation of the reconstruction-based weights for the numerous between-class image patches, which would save much time.

Please note that, as the training phase is always an offline process in practical applications, the testing time thus becomes the key metric to measure the reality of one method. Hence, we further record the testing time of the proposed RHDA method. On average, the testing time of RHDA is 0.3245 s[1], which is less than the acceptable 0.5 s. In a nutshell, the computational time of RHDA will not limit its applications from the practical viewpoint.

3.5. Evaluation on LFW dataset with deep features

The LFW dataset [62] consists of the faces of 5749 subjects in unconstrained environment. The face images collected under the unconstrained environment with inaccurate alignment make the LFW data extremely challenging for face verification, let alone FR with SSPP. In this experiment, we further evaluate the proposed RHDA with the deep feature on this challenging LFW dataset. We employ the MatConvNet [63] toolbox, with a 37-layer VGG-Face model [41] pre-trained on a very large scale dataset (2.6M images, over 2.6K subjects) being used. Since RHDA is a patch-based method, we thus choose the feature generated from the 32th layer and convert it to 64×64 tensor-based feature for RHDA. For convenience, the proposed RHDA method using VGG-Face deep feature is called RHDA+VGG-Face for short.

In the experiment, we choose two deep learning based methods, i.e., DeepID [40] and **J**oint and **C**ollaborative **R**epresentation with local **A**daptive **C**onvolution **F**eature (JCR-ACF) [47], for comparison. Besides, since DMMA and SDMME are closely related to our RHDA, we thus report the results of DMMA using VGG-Face feature, i.e., DMMA+VGG-Face, and SDMME using VGG-Face feature, i.e., SDMME+VGG-Face, for reference. Moreover, for comparative studies, we also report the recognition results of RHDA and the other 7 comparing methods including SRC, ESRC, PCRC, DMMA, SDMME, SVDL and CPL, using the raw pixels as feature. For the experimental configuration, we followed the protocol in JCR-ACF, and utilized a subset of 158 subjects with no less than 10 images per subject from LFW-a to form the evaluation and generic sets. The first 50 subjects were used for evaluation and the remaining 108 subjects were used for generic learning. The first image of each subject was selected as the gallery sample, and the rest 9 images were used for testing. For the deep learning based DeepID and JCR-ACF, their parameters were set in accordance with [47]. The parameters for the other comparing methods including SRC, PCRC, ESRC, SVDL, CPL, DMMA, SDMME, DMMA+VGG-Face and SDMME+VGG-Face were tuned to achieve their best results. For our RHDA&RHDA+VGG-Face, the parameters k_1, k_2, λ, σ, T_1 and T_2 were empirically set to be 50, 2, 0.001, 1, 4 and 4, respectively. The performance of all the methods on LFW dataset are reported in Table 8.

From Table 8, we can observe that, with raw pixel features, no methods achieve very high accuracy because of the extremely challenging facial variations in uncontrolled setting. Nevertheless, our RHDA still works better than other comparing methods with improvements ranging from 4.67%-17.45%. Moreover, benefiting from the VGG-Face feature, the recognition accuracy of RHDA can be significantly enhanced from 32.89% to 87.56%, and achieves slightly better results than that of JCR-ACF (i.e., 86.00%), the state-of-the-art deep learning based method that addresses the SSPP FR problem. This experiment again demonstrates the discriminating power of the deep VGG-Face feature, and provides a feasible way to address the practical SSPP FR problem by combining our RHDA with deep features.

[1] As the expanded gallery patch dictionary for the patch-to-patch distance can be prepared in advance before the identification stage, we thus not count its time cost in the testing time.

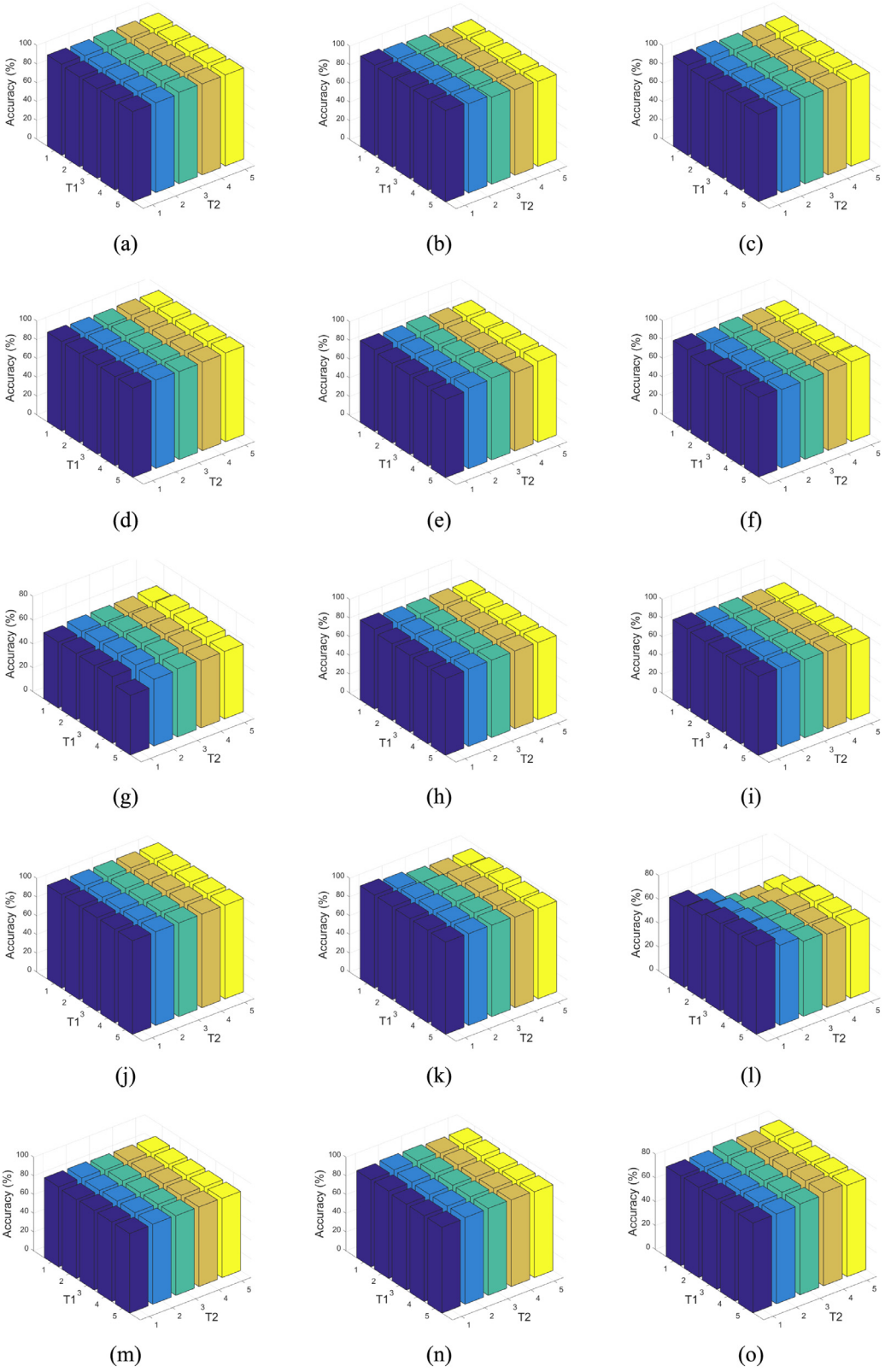

Fig. 13. The recognition accuracies of RHDA versus the combinations of parameters T_1&T_2 on 15 cases over AR, FERET, CAS-PEAL, E-YaleB and Multi-PIE datasets. (a)-(d), (e)-(g), (h) and (i), (j)-(l), and (m)-(o) show the results of prob sets b-e, b-d, b and c, b-d, b-d on AR, FERET, CAS-PEAL, E-YaleB and Multi-PIE, respectively. T_1 and T_2 both vary from 1 to 5.

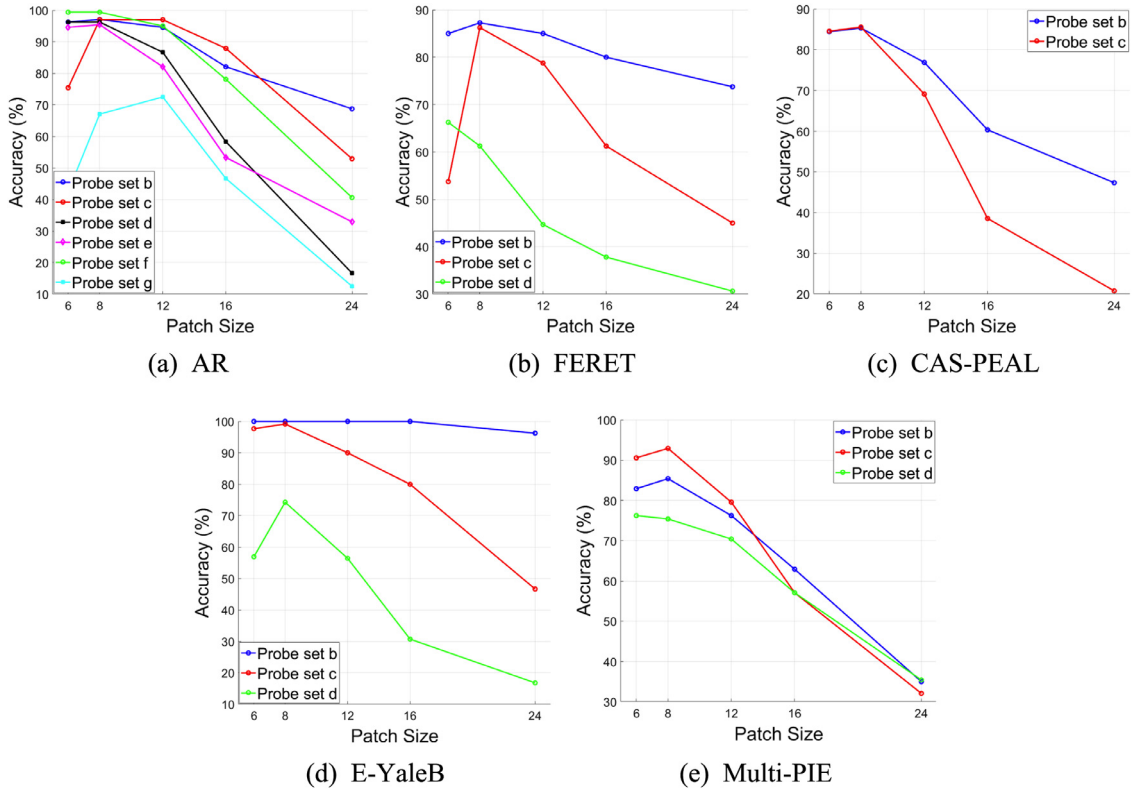

Fig. 14. The performance of RHDA using patches with different sizes on (a) AR, (b) FERET, (c) CAS-PEAL, (d) E-YaleB, and (e) Multi-PIE datasets.

Table 8
Recognition accuracy (%) of different methods on LFW dataset.

Raw pixel based methods	Accuracy
SRC [16]	17.11
DMMA [23]	16.22
SDMME [31]	15.44
PCRC [29]	21.78
ESRC [19]	24.00
SVDL [20]	28.22
CPL [28]	27.56
RHDA	32.89
Deep learning based methods	Accuracy
DeepID [40]	70.70
JCR-ACF [47]	86.00
DMMA+VGG-Face	82.33
SDMME+VGG-Face	81.11
RHDA+VGG-Face	**87.56**

4. Discussions

Although the proposed RHDA is specifically designed for SSPP FR in this work, it can also be applicable to other pattern recognition applications. Two typical examples are *undersampled* FR [64] where each subject contains few training samples (more than single sample), and *imbalanced* FR [65] where some subjects contain sufficient training samples while the other subjects contain very limited training samples or even single training sample. In fact, the two problems can be considered as the simplified cases of SSPP FR, because not all subjects are restricted to contain single sample for training like SSPP FR. Under such circumstances, the training patches of each subject would increase remarkably and more discriminative information can be captured. Hence, It is expected that the proposed RHDA still performs well for the above two FR tasks.

Moreover, RHDA can also be applied to the image set based classification problems such as video-based FR under constrained/unconstrained conditions [66]. In this case, each frame of the video is treated as an independent sample of this subject. Then, the patch-based RHDA can be easily extended to the sample-based RHDA by modeling the whole samples over all videos as a single manifold and multiple manifolds (refer to DSME and DMME), respectively. Thus, the video-based FR can be formulated as a combination of sample-to-sample and sample-to-manifold matching problem.

5. Conclusion

This paper proposes a new patch-based method, i.e., RHDA, for FR with SSPP. The proposed RHDA has two major advantages, so that it shows good robustness against different types of facial variations or occlusions in the query face. The first advantage attributes to the Fisher-like criterion, which is able to extract the hidden discriminant information across two heterogeneous adjacency graphs, and meanwhile improve the discriminative ability of patch distribution in underlying subspaces. The second one is the joint majority voting strategy by considering both the patch-to-patch and patch-to-manifold distances, which can generate complementary information as well as increase the error tolerance for identification. Experimental results on AR, FERET, CAS-PEAL, E-YaleB, Multi-PIE and LFW datasets have demonstrated the effectiveness of the proposed method in comparison with the existing counterparts.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grants 61672444, 61272366 and 61672125, and in part by the SZSTI Grant JCYJ20160531194006833,

and by the Faculty Research Grant of Hong Kong Baptist University under Project FRG2/17-18/082. Besides, we would like to express our sincere gratitude to all the reviewers for their constructive and valuable comments.

References

[1] L. Best-Rowden, H. Han, C. Otto, B.F. Klare, A.K. Jain, Unconstrained face recognition: identifying a person of interest from a media collection, IEEE Trans. Inf. Forensics Secur. 9 (12) (2014) 2144–2157.

[2] L. Wolf, T. Hassner, I. Maoz, Face recognition in unconstrained videos with matched background similarity, in: Proceedings of CVPR, 2011, pp. 529–534.

[3] S. Bashbaghi, E. Granger, R. Sabourin, G.-A. Bilodeau, Dynamic ensembles of exemplar-svms for still-to-video face recognition, Pattern Recognit. 69 (2017) 61–81.

[4] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, Y. Ma, Toward a practical face recognition system: robust alignment and illumination by sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2) (2012) 372–386.

[5] N. Erdogmus, S. Marcel, Spoofing face recognition with 3D masks, IEEE Trans. Inf. Forensics Secur. 9 (7) (2014) 1084–1097.

[6] M. Ye, C. Liang, Y. Yu, Z. Wang, Q. Leng, C. Xiao, J. Chen, R. Hu, Person re-identification via ranking aggregation of similarity pulling and dissimilarity pushing, IEEE Trans. Multimed. 18 (12) (2016) 2553–2566.

[7] J. Li, A.J. Ma, P.C. Yuen, Semi-supervised region metric learning for person re-identification, Int. J. Comput. Vis. (2018) 1–20.

[8] X. Lan, S. Zhang, P.C. Yuen, R. Chellappa, Learning common and feature-specific patterns: a novel multiple-sparse-representation-based tracker, IEEE Trans. Image Process. 27 (4) (2018) 2022–2037.

[9] Z. He, S. Yi, Y.-M. Cheung, X. You, Y.Y. Tang, Robust object tracking via key patch sparse representation, IEEE Trans. Cybern. 47 (2) (2017) 354–364.

[10] W. Zhao, R. Chellappa, P.J. Phillips, A. Rosenfeld, Face recognition: a literature survey, ACM Comput. Surv. 35 (4) (2003) 399–458.

[11] 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.

[12] P.N. Belhumeur, J.P. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.

[13] J. Gui, Z. Sun, W. Jia, R. Hu, Y. Lei, S. Ji, Discriminant sparse neighborhood preserving embedding for face recognition, Pattern Recognit. 45 (8) (2012) 2884–2893.

[14] Y. Zhou, S. Sun, Manifold partition discriminant analysis, IEEE Trans. Cybern. 47 (4) (2017) 830–840.

[15] M. Pang, B. Wang, Y.-M. Cheung, C. Lin, Discriminant manifold learning via sparse coding for robust feature extraction, IEEE Access 5 (2017) 13978–13991.

[16] 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.

[17] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: which helps face recognition? in: Proceedings of ICCV, 2011, pp. 471–478.

[18] S. Gao, K. Jia, L. Zhuang, Y. Ma, Neither global nor local: regularized patch-based representation for single sample per person face recognition, Int. J. Comput. Vis. 111 (3) (2015) 365–383.

[19] 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.

[20] M. Yang, L. Van Gool, L. Zhang, Sparse variation dictionary learning for face recognition with a single training sample per person, in: Proceedings of ICCV, 2013, pp. 689–696.

[21] Y.-F. Yu, D.-Q. Dai, C.-X. Ren, K.-K. Huang, Discriminative multi-scale sparse coding for single-sample face recognition with occlusion, Pattern Recognit. 66 (2017) 302–312.

[22] Y. Gao, J. Ma, A.L. Yuille, Semi-supervised sparse representation based classification for face recognition with insufficient labeled samples, IEEE Trans. Image Process. 26 (5) (2017) 2545–2560.

[23] J. Lu, Y.-P. Tan, G. Wang, Discriminative multimanifold analysis for face recognition from a single training sample per person, IEEE Trans. Pattern Anal. Mach. Intell. 35 (1) (2013) 39–51.

[24] 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.

[25] Q.-x. Gao, L. Zhang, D. Zhang, Face recognition using FLDA with single training image per person, Appl. Math. Comput. 205 (2) (2008) 726–734.

[26] J. Wang, K.N. Plataniotis, J. Lu, A.N. Venetsanopoulos, On solving the face recognition problem with one training sample per subject, Pattern Recognit. 39 (9) (2006) 1746–1762.

[27] W. Deng, J. Hu, J. Guo, In defense of sparsity based face recognition, in: Proceedings of CVPR, 2013, pp. 399–406.

[28] H.-K. Ji, Q.-S. Sun, Z.-X. Ji, Y.-H. Yuan, G.-Q. Zhang, Collaborative probabilistic labels for face recognition from single sample per person, Pattern Recognit. 62 (2017) 125–134.

[29] P. Zhu, L. Zhang, Q. Hu, S.C. Shiu, Multi-scale patch based collaborative representation for face recognition with margin distribution optimization, in: Proceedings of ECCV, 2012, pp. 822–835.

[30] F. Liu, J. Tang, Y. Song, L. Zhang, Z. Tang, Local structure-based sparse representation for face recognition, ACM Trans. Intell. Syst. Technol. 7 (1) (2015) 2.

[31] P. Zhang, X. You, W. Ou, C.P. Chen, Y.-M. Cheung, Sparse discriminative multi-manifold embedding for one-sample face identification, Pattern Recognit. 52 (2016) 249–259.

[32] T. Pei, L. Zhang, B. Wang, F. Li, Z. Zhang, Decision pyramid classifier for face recognition under complex variations using single sample per person, Pattern Recognit. 64 (2017) 305–313.

[33] R. Gottumukkal, V.K. Asari, An improved face recognition technique based on modular PCA approach, Pattern Recognit. Lett. 25 (4) (2004) 429–436.

[34] S. Chen, J. Liu, Z.-H. Zhou, Making FLDA applicable to face recognition with one sample per person, Pattern Recognit. 37 (7) (2004) 1553–1555.

[35] H. Yan, J. Lu, X. Zhou, Y. Shang, Multi-feature multi-manifold learning for single-sample face recognition, Neurocomputing 143 (2014) 134–143.

[36] P. Zhu, M. Yang, L. Zhang, I.-Y. Lee, Local generic representation for face recognition with single sample per person, in: Proceedings of ACCV, 2014, pp. 34–50.

[37] T. Khadhraoui, M.A. Borgi, F. Benzarti, C.B. Amar, H. Amiri, Local generic representation for patch uLBP-based face recognition with single training sample per subject, Multimed. Tools Appl. (2018) 1–20.

[38] M. Belkin, P. Niyogi, Laplacian eigenmaps and spectral techniques for embedding and clustering, in: Proceedings of NIPS, vol. 14, 2001, pp. 585–591.

[39] S. Yan, H. Wang, Semi-supervised learning by sparse representation, in: Proceedings of SDM, 2009, pp. 792–801.

[40] Y. Sun, X. Wang, X. Tang, Deep learning face representation from predicting 10,000 classes, in: Proceedings of CVPR, 2014, pp. 1891–1898.

[41] O.M. Parkhi, A. Vedaldi, A. Zisserman, et al., Deep face recognition., in: Proceedings of BMVC, vol. 1, 2015, p. 6.

[42] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion, J. Mach. Learn. Res. 11 (Dec) (2010) 3371–3408.

[43] Y. Li, G. Wang, L. Nie, Q. Wang, W. Tan, Distance metric optimization driven convolutional neural network for age invariant face recognition, Pattern Recognit. 75 (2018) 51–62.

[44] B. Amos, B. Ludwiczuk, M. Satyanarayanan, et al., Openface: A General-Purpose Face Recognition Library with Mobile Applications, CMU School of Computer Science(2016).

[45] S. Gao, Y. Zhang, K. Jia, J. Lu, Y. Zhang, Single sample face recognition via learning deep supervised autoencoders, IEEE Trans. Inf. Forensics Secur. 10 (10) (2015) 2108–2118.

[46] M. Parchami, S. Bashbaghi, E. Granger, CNNs with cross-correlation matching for face recognition in video surveillance using a single training sample per person, in: Proceedings of AVSS, 2017, pp. 1–6.

[47] M. Yang, X. Wang, G. Zeng, L. Shen, Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person, Pattern Recognit. 66 (2017) 117–128.

[48] M. Pang, Y.-M. Cheung, B. Wang, R. Liu, Robust heterogeneous discriminative analysis for single sample per person face recognition, in: Proceedings of CIKM, 2017, pp. 2251–2254.

[49] D. Cai, X. He, J. Han, Spectral regression: a unified subspace learning framework for content-based image retrieval, in: Proceedings of ACM MM, 2007, pp. 403–412.

[50] L.v.d. Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008) 2579–2605.

[51] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: Proceedings of CVPR, 1991, pp. 586–591.

[52] J. Wu, Z.-H. Zhou, Face recognition with one training image per person, Pattern Recognit Lett 23 (14) (2002) 1711–1719.

[53] J. Yang, D. Zhang, A.F. Frangi, J.-y. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (1) (2004) 131–137.

[54] 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.

[55] A.M. Martinez, The AR Face Database, CVC Technical Report, 1998.

[56] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 22 (10) (2000) 1090–1104.

[57] W. Gao, B. Cao, S. Shan, X. Chen, D. Zhou, X. Zhang, D. Zhao, The CAS-PEAL large-scale chinese face database and baseline evaluations, IEEE Trans. Syst. Man Cybern. 38 (1) (2008) 149–161.

[58] 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.

[59] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-PIE, Image Vis. Comput. 28 (5) (2010) 807–813.

[60] D.L. Donoho, Y. Tsaig, Fast solution of l_1-norm minimization problems when the solution may be sparse, IEEE Trans. Inf. Theory 54 (11) (2008) 4789–4812.

[61] A.Y. Yang, Z. Zhou, A.G. Balasubramanian, S.S. Sastry, Y. Ma, Fast l_1-minimization algorithms for robust face recognition, IEEE Trans. Image Process. 22 (8) (2013) 3234–3246.

[62] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Technical Report, Technical Report 07–49, University of Massachusetts, Amherst, 2007.

[63] A. Vedaldi, K. Lenc, Matconvnet: convolutional neural networks for matlab, in: Proceedings of ACM MM, 2015, pp. 689–692.

[64] C.-P. Wei, Y.-C.F. Wang, Undersampled face recognition via robust auxiliary dictionary learning, IEEE Trans. Image Process. 24 (6) (2015) 1722–1734.

[65] J. Wang, J. You, Q. Li, Y. Xu, Extract minimum positive and maximum negative features for imbalanced binary classification, Pattern Recognit. 45 (3) (2012) 1136–1145.
[66] R. Wang, S. Shan, X. Chen, Q. Dai, W. Gao, Manifold–manifold distance and its application to face recognition with image sets, IEEE Trans. Image Process. 21 (10) (2012) 4466–4479.

Meng Pang received the B.Sc. degree in embedded engineering and the M.Sc. degree in software engineering from Dalian University of Technology, Dalian, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Hong Kong Baptist University, Hong Kong, China. His research interests include image processing, pattern recognition, and data mining.

Yiu-Ming Cheung received the Ph.D. degree from the Department of Computer Science and Engineering, Chinese University of Hong Kong, Hong Kong, in 2000. He is currently a Full Professor with the Department of Computer Science, Hong Kong Baptist University, Hong Kong. His research interests include machine learning, image and video processing, pattern recognition, and optimization. Prof. Cheung is the Founding Chair of the Computational Intelligence Chapter of IEEE Hong Kong Section and the Vice Chair of Technical Committee on Intelligent Informatics of the IEEE Computer Society. He is an IEEE Fellow, IET/IEE Fellow, BCS Fellow, RSA Fellow, and IETI Distinguished Fellow.

Bing-Hui Wang received the B.Sc. degree in network engineering and the M.Sc. degree in software engineering from Dalian University of Technology, Dalian, China, in 2012 and 2015, respectively. He is currently working toward the Ph.D. degree in electrical and computer engineering at Iowa State University, Ames, IA, USA. His research interests include machine learning, big data mining, data-driven security and privacy, and adversarial machine learning.

Ri-Sheng Liu received the B.Sc. and Ph.D. degrees both in mathematics from the Dalian University of Technology, Dalian, China, in 2007 and 2012, respectively. From 2010 to 2012, he was a visiting scholar in the Robotic Institute of Carnegie Mellon University. Since 2016, he has been Hong Kong Scholar Research Fellow at the Hong Kong Polytechnic University, Hong Kong. He is currently an Associate Professor in the International School of Information and Software Technology, Dalian University of Technology, Dalian, China. His research interests include machine learning, optimization, computer vision and multimedia.