Robust Heterogeneous Discriminative Analysis for Single Sample Per Person Face Recognition

Meng Pang
Hong Kong Baptist University
mengpang@comp.hkbu.edu.hk

Yiu-ming Cheung*
Hong Kong Baptist University
ymc@comp.hkbu.edu.hk

Binghui Wang
Iowa State University
binghuiw@iastate.edu

Risheng Liu
Dalian University of Technology
rsliu@dlut.edu.cn

ABSTRACT

Single sample face recognition is one of the most challenging problems in face recognition (FR), where only one single sample per person (SSPP) is enrolled in the gallery set for training. Although patch-based methods have achieved great success in FR with SSPP, they still have significant limitations. In this work, we propose a new patch-based method, namely Robust Heterogeneous Discriminative Analysis (RHDA), to tackle FR with SSPP. Compared with the existing patch-based methods, RHDA can enhance the robustness against complex facial variations from two aspects. First, we develop a novel 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 neighboring patches from different persons. Second, we present 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 joint majority voting for identification. Experimental results on the AR and FERET benchmark datasets demonstrate the efficacy of the proposed method.

KEYWORDS

Single sample face recognition, representation learning, heterogeneous subspace analysis, joint majority voting.

1 INTRODUCTION

In many practical face recognition (FR) systems, e.g., law enforcement and ID card identification, there is only single sample per person (SSPP) when considering their limited storage and privacy policy. As a result, it becomes particularly intractable for FR with SSPP when within-class information is not available to predict facial variations in the query face. Therefore, a variety of existing Fisher-based subspace learning methods [4, 15], e.g., linear discriminant analysis (LDA), fail to work in this scenario. Moreover, the emerging representation-based classifiers, i.e., sparse representation classifier (SRC) [11] and collaborative representation classifier (CRC) [14], also suffer from heavy performance degeneration because these classifiers still require multiple within-class training samples to reasonably represent the query face.

To address the SSPP problem in FR, there have been some attempts in the literature, which can be roughly classified into two categories [3]: holistic generic learning methods and patch-based methods. Holistic generic learning methods introduce an auxiliary generic set with multiple samples per person to supplement the original gallery set. For example, Wang *et al.* developed a generic learning framework [10] to estimate approximated within-class scatter from generic set provided that different sets of person share similar within-class variations. Representative methods under this framework include extended SRC (ESRC) [2], sparse variation dictionary learning (SVDL) [13], collaborative probabilistic labels (CPL) [5], etc. Although this kind of methods can alleviate the SSPP problem to some extent, their performance depends heavily on the elaborative selection of auxiliary generic set, which is a tough task in practical applications.

Patch-based methods [6, 8, 16, 17] recognize a query face by leveraging its partitioned patches, and each partitioned patch of a face image is treated as an independent sample of this person (i.e., class). As a result, researchers attempted to extend conventional subspace learning and representation-based methods, e.g., LDA, SRC and CRC, to the corresponding patch-based versions, i.e., modular LDA [6], PSRC [11] and PCRC [17], and conducted FR via integrating the recognition outputs of all partitioned patches. More recently, Lu *et al.* [6] have developed a discriminative multi-manifold analysis (DMMA) method provided that the partitioned patches of each person lie in an individual manifold, thus converting FR to a manifold-manifold matching problem. Furthermore, Zhang *et al.* [16] modified DMMA and proposed a sparse discriminative multi-manifold embedding (SDMME) method by leveraging sparse graph embedding.

Nevertheless, we emphasize that patch-based methods still have two major drawbacks. First, for feature extraction,

*Yiu-ming Cheung is the corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CIKM'17, November 6–10, 2017, Singapore.
© 2017 ACM. ISBN 978-1-4503-4918-5/17/11...$15.00
DOI: https://doi.org/10.1145/3132847.3133096

the Fisher criteria applied in patch-based methods, such as modular LDA, DMMA, and SDMME, cannot generate representations (i.e., features) that are discriminant enough, because they only conduct discriminant analysis in the same feature space, while ignoring vital discriminant information across different feature spaces. Second, for identification, it is believed that, given a patch from a query face, 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. However, all above patch-based methods only consider one of the two observations (i.e., distance metrics), which is inadequate when handling complex facial variations. **Our work:** We propose a new patch-based method, called Robust Heterogeneous Discriminative Analysis (RHDA), to address the above drawbacks.

To address the first issue, we develop a novel Fisher-like criterion in RHDA model, based on graph embedding, to extract sufficient discriminant information from two heterogeneous feature spaces. One feature space preserves the reconstruction relationship of neighboring patches from the same person, and the other suppresses neighboring patches from different persons.

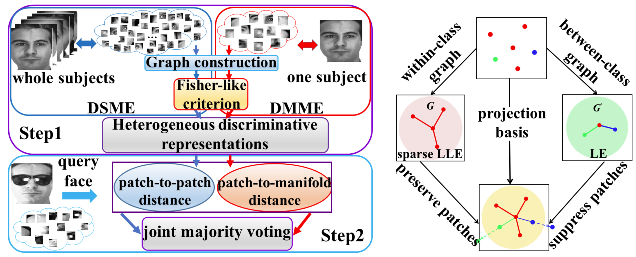
Regarding the second issue, we present two discriminative manifold embeddings, namely discriminative single-manifold embedding (DSME) and discriminative multi-manifold embedding (DMME). The two embeddings, respectively, model the whole partitioned patches over all persons as a single manifold and multiple manifolds, and are then incorporated into above Fisher-like criterion to generate heterogeneous discriminative representations for image patches. Subsequently, we design 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 two distance metrics and combining their recognition outputs via joint majority voting to identify each query face.

2 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 develop two discriminative manifold embeddings (i.e., DSME and DMME), and finally leverage a Fisher-like criterion to generate heterogeneous discriminative subspace representations for patches. For identification, we develop a fusion strategy to exploit the heterogeneous subspace representations and identify each query face via joint majority voting. The pipeline of the RHDA method is illustrated in Figure 1(a).

2.1 Heterogeneous Feature Extraction

2.1.1 Graph construction and weight matrix definition. Suppose $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N] \in \Re^{D\times N}$ is a gallery set with N persons. We first partition each $\mathbf{x}_i$ into M non-overlapping local patches with an equal size d, and concatenate the patches column by column. For the ith person, we define its patch set as $\mathbf{X}_i = [\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \cdots, \mathbf{x}_{i,M}] \in \Re^{d\times M}$. Following [12] and [1], we then construct an intrinsic graph G in sparse LLE feature space and penalty graph G' in Laplacian eigenmap (LE) feature space, respectively. $\mathbf{S}^w$ and $\mathbf{S}^b$ denote the corresponding reconstruction weight and affinity weight matrix, respectively.



(a) (b)

Figure 1: (a) Pipeline of RHDA. (b) Illustration of the Fisher-like criterion. The points with the same color indicate the patches from the same class.

In intrinsic graph G, we aim to measure the representation capability of patches from the same person. Similar with l_1-graph [12], we first design a within-class dictionary for each patch (e.g., $\mathbf{x}_{i,j}$): $\mathbf{A}_{i,j} = [\mathbf{x}_{i,1}, \cdots, \mathbf{x}_{i,j-1}, \mathbf{x}_{i,j+1}, \cdots, \mathbf{x}_{i,M}]$. Then, the representation coefficients of remaining within-class patches for $\mathbf{x}_{i,j}$ can be calculated as $\boldsymbol{\alpha}_{i,j} = \arg\min_{\boldsymbol{\alpha}_{i,j}} ||\mathbf{x}_{i,j} - \mathbf{A}_{i,j}\boldsymbol{\alpha}_{i,j}||_F^2 + ||\boldsymbol{\alpha}_{i,j}||_1$. Therefore, the within-class reconstruction weight matrix $\mathbf{W}_i$ for the ith person is defined as $\mathbf{W}_i = [\mathbf{W}_{i,1}, \cdots, \mathbf{W}_{i,j}, \cdots, \mathbf{W}_{i,M}] \in \Re^{M\times M}$. $\mathbf{W}_{i,j}^p$ denotes the pth element of $\mathbf{W}_{i,j}$, we define $\mathbf{W}_{i,j}^p = \boldsymbol{\alpha}_{i,j}^p$ if $0 < p < j$, $\mathbf{W}_{i,j}^p = 0$ if $p = j$, and $\mathbf{W}_{i,j}^p = \boldsymbol{\alpha}_{i,j}^{p-1}$ if $j < p < M$. Hence, $\mathbf{S}^w$ for whole patches over all persons can be defined as $\mathbf{S}^w = \mathrm{diag}(\mathbf{W}_1, \cdots, \mathbf{W}_N) \in \Re^{MN\times MN}$.

In 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: $\widehat{\mathbf{W}}_{i,j}^p = exp(-||\mathbf{x}_{i,j} - \mathbf{x}_{i,j}^p||^2/\sigma^2)$ if $\mathbf{x}_{i,j}^p \in N_{k_1}(\mathbf{x}_{i,j})$, and $\widehat{\mathbf{W}}_{i,j}^p = 0$ otherwise. $N_{k_1}(\mathbf{x}_{i,j})$ denote the k_1-nearest between-class patches of $\mathbf{x}_{i,j}$. Hence, the affinity weight matrix $\mathbf{S}^b$ for whole patches in graph G' is set as $\mathbf{S}^b = \widehat{\mathbf{W}} \in \Re^{MN\times MN}$.

2.1.2 Discriminative manifold embeddings. **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 the one hand, we need to preserve the reconstruction relationship of neighboring within-class patches in sparse LLE feature space. On the other hand, we need to suppress neighboring patches of different classes in LE feature space. Formally, we can achieve it by learning a shared projection basis $\mathbf{U} \in \Re^{d\times r}$ for all patches and *optimizing the following two objective functions*: $\min_{\mathbf{U}} \Phi^w(\mathbf{U}) = \sum_i \|\mathbf{U}^T\widehat{\mathbf{x}}_i - \Sigma_j \mathbf{S}_{ij}^w \mathbf{U}^T\widehat{\mathbf{x}}_j\|^2$ and $\max_{\mathbf{U}} \Phi^b(\mathbf{U}) = \sum_{i,j} \|\mathbf{U}^T\widehat{\mathbf{x}}_i - \mathbf{U}^T\widehat{\mathbf{x}}_j\|^2 \mathbf{S}_{ij}^b$.
DMME: It models the whole patch set as a collection of multiple manifolds, and assumes 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 optimize the following two objective functions*: $\max_{\mathbf{V}} J_1(\mathbf{V}_i) = \sum_{j=1}^{M}\sum_{p=1}^{k_1} \|\mathbf{V}_i^T\mathbf{x}_{i,j} - \mathbf{V}_i^T\mathbf{x}_{i,j}^p\|^2\widehat{\mathbf{W}}_{i,j}^p$ and $\min_{\mathbf{V}} J_2(\mathbf{V}_i) = \sum_{j=1}^{M}\|\mathbf{V}_i^T\mathbf{x}_{i,j} - \mathbf{V}_i^T\mathbf{X}_i\mathbf{W}_{i,j}\|^2$. J_1 is generated from G' to ensure that if $\mathbf{x}_{i,j}$ and $\mathbf{x}_{i,j}^p$ are close but from different persons, they should be separated as far as possible after projection. On the other hand, J_2 from G is to preserve the reconstruction relationship of within-class neighboring patches after projection.

2.1.3 Feature extraction via a Fisher-like criterion. We design a Fisher-like criterion to extract discriminative features across the two heterogeneous feature spaces, i.e., sparse LLE and LE. Specifically, it aims to simultaneously preserve the reconstruction relationship of neighboring within-class patches in sparse LLE feature space, while suppressing neighboring patches of different classes in LE feature space. The illustration of the Fisher-like criterion is shown in Figure 1(b).

Then for DSME, by incorporating the Fisher-like criterion, the final objective function becomes:

$$\max_{\mathbf{U}} \frac{\Phi^b(\mathbf{U})}{\Phi^w(\mathbf{U})} = \frac{tr(\mathbf{U}^T\widehat{\mathbf{X}}\mathbf{L}^b\widehat{\mathbf{X}}^T\mathbf{U})}{tr(\mathbf{U}^T\widehat{\mathbf{X}}\mathbf{M}^w\widehat{\mathbf{X}}^T\mathbf{U})}, \tag{1}$$

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. Thus, the above maximization problem can be transformed to the following generalized eigen-problem: $\widehat{\mathbf{X}}\mathbf{L}^b\widehat{\mathbf{X}}^T\mathbf{U} = \lambda\widehat{\mathbf{X}}\mathbf{M}^w\widehat{\mathbf{X}}^T\mathbf{U}$, where λ and $\mathbf{U}$ are the eigenvalues and eigenvectors of the above generalized eigen-problem.

For DMME, we define the Fisher-like criterion as

$$\max_{\mathbf{V}} J(\mathbf{V}) = \sum_{i=1}^{N}(J_1(\mathbf{V}_i) - J_2(\mathbf{V}_i)). \tag{2}$$

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-problem: $(\mathbf{H}_1 - \mathbf{H}_2)\mathbf{v} = \lambda\mathbf{v}$, where $\mathbf{H}_1 = \sum_{j=1}^{M}\sum_{p=1}^{k_1}(\mathbf{x}_{i,j} - \mathbf{x}_{i,j}^p)(\mathbf{x}_{i,j} - \mathbf{x}_{i,j}^p)^T\widehat{\mathbf{W}}_{i,j}^p$ and $\mathbf{H}_2 = \sum_{j=1}^{M}(\mathbf{x}_{i,j} - \mathbf{X}_i\mathbf{W}_{i,j})(\mathbf{x}_{i,j} - \mathbf{X}_i\mathbf{W}_{i,j})^T$. 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 $\lambda_1 \geq \cdots \geq \lambda_{d_i} \geq 0$. Then the projection basis for the ith class is indicated as $\mathbf{V}_i = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{d_i}]$.

2.2 Face Identification

Given a query face $\mathbf{y}$, we partition it into M non-overlapping local patches $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_M$. Next, we design our *first distance metric* as patch-to-patch distance, 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 dictionary as $\mathbf{D}_j = [\mathbf{U}^T\mathbf{x}_{1,j}, \cdots, \mathbf{U}^T\mathbf{x}_{i,j} \cdots, \mathbf{U}^T\mathbf{x}_{N,j}] \in \Re^{r\times N}$. $\mathbf{U}^T\mathbf{x}_{i,j}$ denotes the subspace representation for the jth patch of the ith person in the gallery set.

Hence, for $\mathbf{U}^T\mathbf{y}_j$, its representation coefficients over $\mathbf{D}_j$ are computed by $\widehat{\boldsymbol{\rho}}_j = \arg\min_{\boldsymbol{\rho}_j}\{\|\mathbf{U}^T\mathbf{y}_j - \mathbf{D}_j\boldsymbol{\rho}_j\|^2 + \lambda\|\boldsymbol{\rho}_j\|^2\}$, where $\widehat{\boldsymbol{\rho}}_j = [\widehat{\boldsymbol{\rho}}_{j,1}; \widehat{\boldsymbol{\rho}}_{j,2}; \cdots; \widehat{\boldsymbol{\rho}}_{j,N}]$. Thus, the identification output of the query patch $\mathbf{y}_j$ is defined as:

$$L^f(\mathbf{y}_j) = \arg\min_k\{\|\mathbf{U}^T\mathbf{y}_j - \mathbf{D}_{j,k}\widehat{\boldsymbol{\rho}}_{j,k}\|^2/\|\widehat{\boldsymbol{\rho}}_{j,k}\|^2\}. \tag{3}$$

Furthermore, we design our *second distance metric* as patch-to-manifold distance, which measures the reconstruction capability of the reference manifold. As depicted in DMME, the patch set $\mathbf{X}_i$ for the ith person is regarded as an individual manifold. Hence, the distance between the query patch $\mathbf{y}_j$ and $\mathbf{X}_i$ can be computed by $d(\mathbf{y}_j, \mathbf{X}_i) = \min\|\mathbf{V}_i^T\mathbf{y}_j - \Sigma_{p=1}^{k_2}\mathbf{c}_pG_{k_2}^p(\mathbf{V}_i^T\mathbf{y}_j)\|^2$, 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 obtained by

$$L^s(\mathbf{y}_j) = \arg\min_k\ d(\mathbf{y}_j, \mathbf{X}_k). \tag{4}$$

In the final stage, we aim to identify the unlabeled query face by exploiting the identification outputs of all query patches. Please note that $L^f(\mathbf{y}_j)$ and $L^s(\mathbf{y}_j)$ obtained by two distance metrics may be different. Therefore, it is difficult to decide which output is correct. To this end, we present a fusion strategy by leveraging both outputs of two distance metrics and determine the final label of the query face via a joint majority voting. Specifically, we define $\mathbf{vote}^f, \mathbf{vote}^s \in \Re^N$ as two zero initial vectors. Then, we update their values by applying the following formula: $\mathbf{vote}^f(L^f(\mathbf{y}_j)) = \mathbf{vote}^f(L^f(\mathbf{y}_j))+1$ and $\mathbf{vote}^s(L^s(\mathbf{y}_j)) = \mathbf{vote}^s(L^s(\mathbf{y}_j)) + 1$, $j = 1, \cdots, M$.

Thus, the label of the query face $\mathbf{y}$ can be determined using the following joint majority voting:

$$L(\mathbf{y}) = \arg\min_j(\mathbf{vote}^f(j) + \mathbf{vote}^s(j)). \tag{5}$$

3 EXPERIMENTAL RESULTS

This section evaluates RHDA on AR and FERET databases.

Comparing Algorithms: We compare RHDA with 8 representative methods that are used to address the SSPP problem, including PCA, SRC, CRC, ESRC, DMMA, and state-of-the-art PCRC, SDMME and SVDL.

Parameter Setting: In experiments, the face images were resized to 48×48 on AR and FERET databases. For patch-based methods including PCRC, DMMA, SDMME and RHDA, the non-overlapping patch size was fixed as 8×8 for a fair comparison. Furthermore, the other parameters in comparing algorithms were also tuned to achieve the best results. As to the proposed RHDA, the values of the parameters k_1, k_2 and σ were empirically set as 100, 3, and 100, respectively.

3.1 Evaluation on AR Database

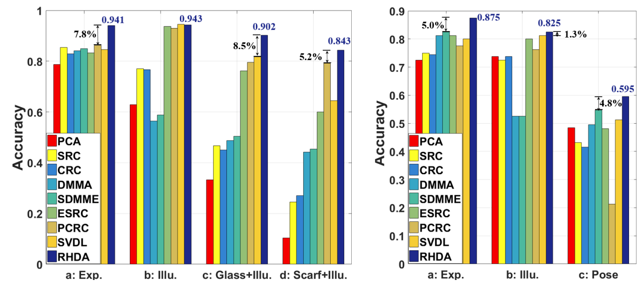
The AR database [7] contains over 4,000 face images of 126 people from two sessions, and each session has 13 face images per subject. Following the work in [3], the first 80 subjects from session-1 were used for evaluation, while the remaining 20 subjects were randomly selected from the remaining set in the same session as the generic set for generic learning methods. The frontal faces taken under normal illuminations and neutral expressions were used as the gallery images, while the remaining 12 images of each subject were arranged to form 4 probe sets (i.e., expression, illumination, sunglass disguise+illumination and scarf disguise+illumination). Figure 2(a) shows the recognition results of involved methods on AR database. It is clear to see that RHDA achieves promising performance among comparing methods. Compared with the state-of-the-art holistic generic learning method SVDL, our RHDA obtains comparable recognition accuracy in prob set b (i.e., illumination) and boosts the recognition rates by a margin as large as 8.5-20 percent for the variances in expression (prob set a) and disguises (prob set c-d). Moreover, RHDA consistently outperforms the state-of-the-art patch-based methods, i.e., PCRC and SDMME, in all cases.

3.2 Evaluation on FERET Database

In this sub-section, we aim to test the robustness of all the methods to the facial variations of expressions, illuminations and poses on FERET database [9]. To this end, we selected 700 face images of 100 subjects from seven galleries (ba, bj, bk, bd, be, bf and bg) on FERET. Following the strategy on AR database, we also utilized the first 80 subjects for evaluation, while the rest 20 subjects were chosen as the generic set. Figure 2(b) shows the performances of all the methods on FERET database, where RHDA also performs the best in all cases. Furthermore, we found that the performance of PCRC degrades seriously in prob set c (i.e., pose variation). A plausible reason is that the pose variations always result in mismatch of corresponding patches. Simply considering the patch-to-patch distance may lead PCRC to make misjudgment when identifying query patch. By contrast, RHDA exhibits greater robustness against pose variations as well as other facial variations compared with PCRC and other comparing methods owing to two important factors. First, the Fisher-like criterion in RHDA can extract highly discriminant information hidden in partitioned patches, and meanwhile improving the discriminative ability of patch distribution in underlying subspaces. On the other hand, RHDA considers both the patch-to-patch and patch-to-manifold distances for identification, which can greatly increase the error tolerance when handling complex facial variation situations.

4 CONCLUSION

This paper has proposed a new patch-based method, i.e. RHDA, for FR with SSPP. RHDA possesses two major advantages, so that it shows great robustness against different types of facial variations or occlusions. The first advantage attributes to the Fisher-like criterion, which is able to extract hidden discriminant information across heterogeneous feature spaces. The other one is the fusion strategy by leveraging both the patch-to-patch and patch-to-manifold distances, which can generate complementary information and increase the error tolerance for identification. Note that RHDA has been directly applied on the original pixel intensity, therefore its performance can be further improved towards practical FR with SSPP applications. One potential direction is to leverage features learnt via deep learning methods. We will leave it as our future work.



(a) AR (b) FERET

Figure 2: The performances of different methods.

5 ACKNOWLEDGMENT

This work was supported in part by NSFC (Nos. 61272366, 61672444, 61672125, 61300086, 61572096, 61432003 and 61632019), by the Faculty Research Grant (No. FRG2/16-17/051) and KTO Grant (No. MPCF-004-2017/18) of HKBU, and by the SZSTI Grant (No. JCYJ20160531194006833).

REFERENCES

[1] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In *NIPS'01*.

[2] W. Deng, J. Hu, and J. Guo. Extended SRC: Undersampled face recognition via intraclass variant dictionary. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34(9):1864–1870, 2012.

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

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

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

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

[7] A. M. Martinez. The AR face database. *CVC Tech. Rep.*, 24, 1998.

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

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

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

[11] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(2):210–227, 2009.

[12] S. Yan and H. Wang. Semi-supervised learning by sparse representation. In *SDM'09*.

[13] M. Yang, L. Van Gool, and L. Zhang. Sparse variation dictionary learning for face recognition with a single training sample per person. In *ICCV'13*.

[14] L. Zhang, M. Yang, and X. Feng. Sparse representation or collaborative representation: Which helps face recognition? In *ICCV'11*.

[15] L. Zhang, P. Zhu, Q. Hu, and D. Zhang. A linear subspace learning approach via sparse coding. In *ICCV'11*.

[16] P. Zhang, X. You, W. Ou, C. P. Chen, and Y.-m. Cheung. Sparse discriminative multi-manifold embedding for one-sample face identification. *Pattern Recognit.*, 52:249–259, 2016.

[17] P. Zhu, L. Zhang, Q. Hu, and S. C. Shiu. Multi-scale patch based collaborative representation for face recognition with margin distribution optimization. In *ECCV'12*.