Signal Processing 141 (2017) 144–157

Contents lists available at ScienceDirect

Signal Processing

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

Low-resolution face recognition with single sample per person

Yongjie Chu[a,b], Touqeer Ahmad[b], George Bebis[b], Lindu Zhao[a,*]

[a] Institute of Systems Engineering, Southeast University, Nanjing 211096, China
[b] Department of Computer Science and Engineering, University of Nevada, Reno, 89557, NV, USA

ARTICLE INFO

Article history:
Received 5 November 2016
Revised 30 March 2017
Accepted 10 May 2017
Available online 11 May 2017

Keywords:
Low-resolution
Single sample per person
Coupled mappings
Cluster-based regularization

ABSTRACT

As a growing number of low-resolution (LR) face images are captured by surveillance cameras, LR face recognition has been a hot issue for recent years. Previous efforts on LR face recognition typically assume each subject has multiple high-resolution (HR) training samples. However, this assumption may not hold in some special cases such as law-enforcement where only a single HR sample per person exists in the training set. For LR face recognition in SSPP scenario, it often suffers from overfitting and singular matrix problems. In this paper, we are the first to investigate LR face recognition with single sample per person, and propose a cluster-based regularized simultaneous discriminant analysis (C-RSDA) method based on SDA. C-RSDA regularizes the between-class and within-class scatter matrices respectively with inter-cluster and intra-cluster scatter matrices, where the cluster-based scatter matrices are computed from unsupervised clustering. With the cluster-based scatter matrices, not only the singularity problem is resolved, but overfitting problem is overcomed as more variations are exploited from the limited training samples. Thus, the proposed C-RSDA enhances the discriminative power of the feature subspace. We extensively evaluate C-RSDA on recognizing LR face images captured in both controlled and uncontrolled environments. The encouraging experimental results demonstrate the effectiveness of the proposed approach.

1. Introduction

Face recognition for single sample per person (SSPP), i.e., conducting face recognition when there is only one sample of each subject in the training set, has been a topical issue for recent years, because its significant application in law enforcement. Many techniques have been proposed to deal with SSPP face recognition [1–6], and all of them so far are designed for high-resolution (HR) face images. In the last decade, for security and law enforcement purposes, an ever-increased number of surveillance cameras have been installed in public area. This proliferation of cameras as well as the significance for real-world applications raises the new requirement for face recognition algorithms to be competent enough to well recognize the surveillance camera quality face images. Due to the large distances between the cameras and the subjects, face regions in the captured images are usually small, resulting in very low-resolution (LR) face images. Recognizing such low-resolution face images is termed as LR face recognition. Compared with high-resolution face images, as demonstrated in [7], the LR faces involve more noise and lack effective features, which considerably degrades the performance of conventional face recognition techniques developed for HR images. Moreover, HR face images of subjects are usually enrolled in the gallery set, the dimensional mismatch between gallery and probe images makes LR face recognition even more challenging.

The most straightforward way to address this dimensional mismatch is to down-sample the HR gallery images to LR ones, and then perform classification on all the LR images. However, this down-sampling removes some discriminative information embedded in HR gallery images, which results in disappointing recognition accuracy. To enhance the recognition performance on LR images, two types of methods have been developed in the literature: (i) super-resolution based techniques and (ii) coupled mappings based techniques.

1.1. Super-resolution based techniques

Super-resolution based techniques rely on super-resolution methods to first reconstruct HR probe images from corresponding LR ones, and then perform recognition on the reconstructed HR probe images. As an instrument to enhance the resolution of images, super-resolution techniques have been developed steadily. These super-resolution techniques typically focus on learning the relationship between LR and HR image patches. Various learning algorithms have been employed to learn this LR-HR map-

* Corresponding author.
E-mail address: ldzhao@seu.edu.cn (L. Zhao).

http://dx.doi.org/10.1016/j.sigpro.2017.05.012
0165-1684/

ping, such as local linear regression [8,9], sparse representation [10,11], manifold learning [12] and convolutional networks [13,14]. These super-resolution methods are mostly designed to be vision-oriented, which means they are designed to obtain good HR reconstructions from LR images, and are not necessarily optimized for recognition purposes.

Recently, there have been some efforts towards recognition-oriented super-resolution techniques. The recognition-oriented super-resolution is more suitable for LR face recognition, because it takes face recognition into consideration when performing super-resolution. Since the decrease in image resolution results in loss of high frequency components, Bilgazyev et al. [15] propose to build a dictionary of high frequency components in the facial data, and then these high frequency components are added to LR face images to perform super-resolution. Studies on human vision systems have shown that the high-frequency information by itself is not sufficient for recognition of low-quality facial images [16], hence super-resolution in the pixel domain will not help much to improve the recognition performance of LR face images. To overcome this problem, Huang et al. [17] present a method to perform super-resolution on coherent feature domain by establishing a nonlinear mapping between HR and LR features. Compared to the pixel domain super-resolution methods, their approach is computationally efficient and robust. These super-resolution methods are specially devised for LR face recognition, but they perform HR face reconstruction and recognition sequentially. To address this issue, some progress has been made in simultaneous super-resolution and recognition, which embeds the super-resolution process into face recognition. Instead of performing super-resolution and recognition independently as two separate sequential process, Jia et al. [18] propose to integrate these two processes together by directly computing a maximum likelihood identity parameter vector in HR tensor space for recognition. Hennings-Yeomans et al. [19] present a discriminative approach that combines recognition with HR reconstruction simultaneously by introducing the super-resolution constraints and feature constraints in a regularization framework. Zou et al. [20] develop a relationship-learning based algorithm, in which a new data constraint and a discriminative constraint are designed for good visual quality image reconstruction and discriminative feature extraction, respectively. The recognition-oriented super-resolution methods, especially those performing super-resolution and recognition synchronously, improve the performance of LR face recognition. Unfortunately, the computational cost to perform super-resolution is high. In addition, when the resolution of a probe image is very low, for example, 8×8 pixels, the performance of these algorithms is depressing, because it is quite difficult to recover much meaningful information from such limited pixels by super-resolution.

1.2. Coupled mappings based techniques

To avoid super-resolution process, coupled mappings based approaches are also developed to address LR face recognition problem. These methods are inspired from canonical correlation analysis (CCA) [21] and aim to learn two different mappings: one for HR images and one for LR image. These learned mappings are first employed to project images of different resolutions into a unified subspace, followed by the classification in the resultant subspace. Based on the framework of coupled mappings (CMs), Li et al. [22] introduce locality preserving objective into the optimization of CMs model and propose coupled locality preserving mappings (CLPMs). CLMPs learn the two mappings in an unsupervised manner. To incorporate the label information, some supervised CMs models are also developed. Siena et al. [23] extend the principles of marginal fisher analysis and propose coupled marginal fisher analysis (CMFA) by utilizing the local relationship of data. Inspired by maximum margin projection, Zhang et al. [24] propose large margin coupled mapping algorithm, by which two projections are learned to maximize the distance of features with different labels and minimize the distance of features with identical label in the common space. Zhou et al. [25] and Zhang et al. [26] propose simultaneous discriminant analysis (SDA) and coupled marginal discriminant mappings respectively to improve the classification discriminability of CMs. Both methods introduce the within-class scatter and between-class scatter into CMs. To further enhance the recognition performance, the geometry structure information of samples are integrated into CMs models [27–29]. Shi et al. [27] combine local and global geometry structures together to learn the coupled mappings, while both Jiang et al. [28] and Xing et al. [29] propose to obtain the coupled mappings by discriminant manifold analysis. Lately, Biswas et al. [30,31] improve the matching performance of the LR face images based on multi-dimensional scaling. The coupled mappings ensure that the distances between LR and HR images in the latent subspace approximate the distances between two corresponding HR images. Besides, kernel tricks have been introduced into CMs to deal with LR face recognition, so that images can be projected into the unified subspace by coupled non-linear mappings. Such methods are coupled kernel embedding [32], coupled kernel fisher discriminative analysis [33], and kernel coupled distance metric learning [34]. Most recently, domain adaption or transfer learning is also embedded into CMs. These approaches deeply exploit the discriminant information from source domain (HR images) to target domain (LR images), and have reported promising results. Two typical methods using domain adaption or transfer learning technique are proposed by Ren et al. [34] and Kan et al. [35], respectively.

1.3. Motivation

All of the methods discussed so far are designed based on the assumption that there are multiple HR images of each subject in the training/gallery set. This assumption may not always be true as single sample per person (SSPP) may occur in some law-enforcement situations, leading to SSPP face recognition problem. Many efforts [1–4] have been made to address SSPP face recognition of HR probe images. However, to the best of our knowledge, there are no attempts to cope with SSPP face recognition of LR probe images. To fill this gap in face recognition, we work on addressing this problem in this paper.

In SSPP situation, it is more difficult to recognize a LR probe face. Less training HR samples of subjects in the training set lead to weak ability to cover sufficient variations, which further induces overfitting problem. Moreover, when employing Fisher criterion to do discriminative analysis, the singularity of the within-class scatter matrix will be a negative factor. A popular way to deal with singularity is to regularize the within-class scatter matrix by adding an identity matrix to it [23,25,26]. However, the introduction of the identity matrix does not bring in any discriminative information for classification, and is unable to solve overfitting.

As the unsupervised clustering algorithms gather similar samples and separate dissimilar ones, some within-class and between-class variations can be exploited from such clusters. Based on SDA, we propose a cluster-based regularized simultaneous discriminant analysis (C-RSDA) method to address LR face recognition when there is single HR face sample per person in the training set. By employing the cluster-based regularization, the singularity problem is solved; meanwhile, as more variations are sought from the limited training samples, the overfitting problem is also settled, so that more discriminative information is introduced into the optimization of the feature subspace. The proposed C-RSDA aims to learn two mappings that can project the HR face images and LR face images into a unified subspace, in which samples from the

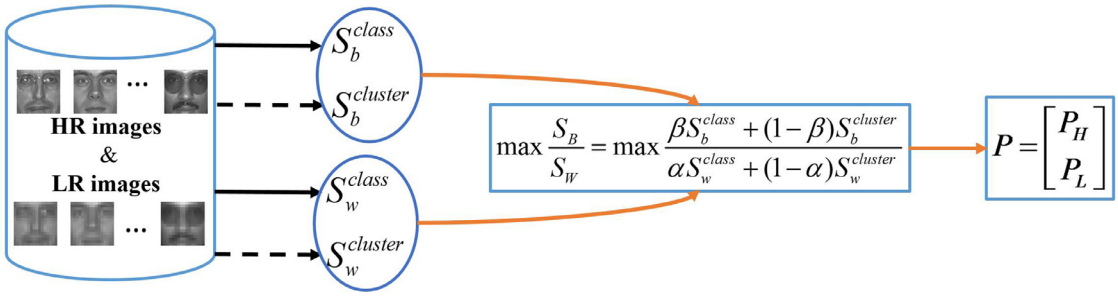

Fig. 1. The pipeline of our proposed C-RSDA.

same class gather as close as possible; meanwhile, samples from different classes disperse as far as possible with a large margin.

The reminder of this paper is organized as follows. In Section 2, we introduce LR face recognition problem. Section 3 provides the details of C-RSDA, and analyzes the significance of the proposed approach; besides, the relationship between the proposed method and SDA is also presented in this section. Section 4 presents the results of experimental evaluation on three popular face databases. We make a brief conclusion in Section 5.

2. Problem statement

In this work, we focus on addressing the low-resolution face recognition problem in which only one HR face image of each subject is enrolled in the training set, and the LR face images are to be recognized.

Assume we have a gallery set of HR face images $\mathrm{H} = [h_1, h_2, \cdots, h_C] \in \mathrm{R}^{M\times C}$, where h_i denotes the HR image, M is the feature dimension, and C is the number of classes, each class has only one HR image. Given a probe LR face image, $l \in \mathrm{R}^m$, m is the feature dimension of LR face image and $m << M$, it is impossible to directly compare the similarity between HR samples and the probe sample since their dimensions are not equal. An alternative way is to project the HR and LR images into a unified subspace by learning a couple of mappings: f_H: $\mathrm{R}^M \rightarrow \mathrm{R}^d$ for HR images and f_L: $\mathrm{R}^m \rightarrow \mathrm{R}^d$ for LR images, where R^d denotes the unified subspace and $d < m$. Thus, the similarity among HR gallery projections and probe LR projection can be calculated in the unified subspace, and recognition can be performed. The probe LR is assigned a class label by

$$c(l) = j \quad \text{when} \quad j = \arg\max_j s\big(f_H(h_j), f_L(l)\big), \tag{1}$$

where $c(l)$ is the class label of probe LR sample l, $s(\circ, \circ)$ is the similarity metric. If we define f_H and f_L as linear projections, then $f_H(h_j)$ and $f_L(l)$ can be respectively formulated as $f_H(h_j) = P_H^T h_j$ and $f_L(l) = P_L^T l$, where P_H and P_L are $M \times d$ and $m \times d$ matrices, respectively.

When learning P_H and P_L, we expect $s(f_H(h_j), f_L(l))$ to be large if l belongs to the same class as h_j, and be small otherwise. It means that in the unified subspace, the projections of the same class accumulated and those of different classes decentralized, regardless of their original identity of HR or LR. To achieve this goal, we compute the between-class and within-class scatter matrices by considering the class label information in the unified projection subspace, and propose the following objective function, which is similar to that of SDA:

$$J(P_H, P_L) = \max \frac{J_B(P_H, P_L)}{J_W(P_H, P_L)}, \tag{2}$$

where $J_B(P_H, P_L)$ and $J_W(P_H, P_L)$ are measures of between-class and within-class scatters in the unified subspace. The above objection function can be solved by a generalized Eigen decomposition problem.

3. Proposed C-RSDA

We assume that the number of LR samples equals the number of HR samples for each subject in the training set. This is reasonable because in most cases the training LR samples have to be acquired by smoothly down sampling the HR ones. Note that if there is only one HR sample for each subject in the gallery set, in the unified subspace there will be only two projection samples for each subject, one is from a HR image and the other is from a LR image. As the number of projections in each class is small, it is easy to encounter overfitting or small-sample-size problem. To avoid this, we attempt to regularize SDA by clustering and propose C-RSDA, which enhances the discriminative power of the unified subspace.

The core thought of the proposed C-RSDA is to employ both cluster-based scatter matrices and class-based scatter matrices to more accurately estimate the true scatter matrices with limited training samples. We utilize SDA as the basic framework to demonstrate our idea in this paper, because SDA is easily understood and computed. Fig. 1 illustrates the core idea of the proposed C-RSDA. It is worth pointing out that this idea can be introduced into any methods or frameworks using either the between-class scatter matrix or the within-class scatter matrix or both, to better assess the inter-class or intra-class variations. In order to clearly describe the proposed C-RSDA, we first introduce Cluster SDA, which is the generalization of SDA. When each cluster is right a class, Cluster SDA is the same as SDA.

3.1. Cluster SDA

Let $\mathrm{L} = [l_1, l_2, \cdots, l_C] \in \mathrm{R}^{m\times C}$ be the corresponding LR images of HR samples in the gallery set. With the two linear mappings, the projections of HR and LR samples in the unified subspace can be expressed as: $\tilde{h}_j = P_H^T h_j$, and $\tilde{l}_j = P_L^T l_j$, where $j = 1, 2, \cdots, C$, $\tilde{h}_j$ and $\tilde{l}_j$ are the projections of HR and LR samples, respectively. To get the mappings of Cluster SDA, we first group the training data into K non-overlapping clusters. Compared with LR images, HR images include more discriminative information, hence we conduct clustering on HR face images, and the clustering result can be directly used to group LR images. The clustering can be expressed as

$$\mathrm{H} = \mathrm{H}_1 \cup \mathrm{H}_2 \cup, \cdots, \mathrm{H}_K, \tag{3}$$

$$\mathrm{H}_i \cap \mathrm{H}_j = \varnothing \quad \text{for } \forall i \neq j, \tag{4}$$

where H_i is the ith cluster. The cluster label of a sample h_j is denoted by $c(h_j)$ with $c(h_j) \in (1, 2, \cdots, K)$.

Then we compute the means of samples in each cluster of HR and LR images as

$$\mu_{L_i} = \frac{1}{N_i}\sum_{c(l_j)=i} l_j,\ \mu_{H_i} = \frac{1}{N_i}\sum_{c(h_j)=i} h_j, \tag{5}$$

where N_i is the number of samples in H_i, and $i = 1, 2, \cdots, K$. The total means of LR and HR images are computed by

$$\mu_L = \frac{1}{C}\sum_{i=1}^{K}\sum_{c(l_j)=i} l_j,\ \mu_H = \frac{1}{C}\sum_{i=1}^{K}\sum_{c(h_j)=i} h_j. \tag{6}$$

Regardless of the original identity (HR or LR) of projections, the mean of projection in each cluster, and the total mean of all projections in the common subspace, can be calculated as follows:

$$\mu_i = \frac{1}{2N_i}\left(\sum_{c(h_j)=i}\tilde{h}_j + \sum_{c(l_j)=i}\tilde{l}_j\right) = \frac{1}{2}\left(P_H^T\mu_{H_i} + P_L^T\mu_{L_i}\right), \tag{7}$$

$$\mu = \frac{1}{2C}\left(\sum_{i=1}^{K}\sum_{c(h_j)=i}\tilde{h}_j + \sum_{i=1}^{K}\sum_{c(l_j)=i}\tilde{l}_j\right) = \frac{1}{2}\left(P_H^T\mu_H + P_L^T\mu_L\right). \tag{8}$$

After obtaining the projections of samples and mean vectors, it is trivial to compute the intra-cluster and inter-cluster scatters in the unified subspace, just by following the way to characterize the scatter matrices in LDA. In the unified projection subspace, the intra-cluster scatter is computed by

$$\begin{aligned} J_w^{(K)} &= \sum_{i=1}^{K}\left(\sum_{c(h_j)=i}\left(\tilde{h}_i - \mu_i\right)^2 + \sum_{c(l_j)=i}\left(\tilde{l}_i - \mu_i\right)^2\right) \\ &= \begin{bmatrix} P_H^T & P_L^T \end{bmatrix}\sum_{i}^{K}\begin{bmatrix} W_{HH}^i & W_{HL}^i \\ W_{LH}^i & W_{LL}^i \end{bmatrix}\begin{bmatrix} P_H \\ P_L \end{bmatrix} \\ &= P^T S_w^{(K)} P, \end{aligned} \tag{9}$$

where $P = [P_H^T \quad P_L^T]^T$ is the concatenation of the mappings P_H and P_L, $S_w^{(K)}$ is the intra-cluster scatter matrix of the projections of HR and LR images when the HR samples are grouped in K clusters. The blocks in $S_w^{(K)}$ are

$$\begin{aligned} W_{HH}^i &= \sum_{c(h_j)=i}\left(h_j - \frac{1}{2}\mu_{H_i}\right)\left(h_j - \frac{1}{2}\mu_{H_i}\right)^T + \frac{1}{4}N_i\mu_{H_i}\mu_{H_i}^T \\ W_{HL}^i &= -\frac{1}{2}N_i\mu_{H_i}\mu_{L_i}^T \\ W_{LH}^i &= -\frac{1}{2}N_i\mu_{L_i}\mu_{H_i}^T \\ W_{LL}^i &= \sum_{c(l_j)=i}\left(l_j - \frac{1}{2}\mu_{L_i}\right)\left(l_j - \frac{1}{2}\mu_{L_i}\right)^T + \frac{1}{4}N_i\mu_{L_i}\mu_{L_i}^T, \end{aligned} \tag{11}$$

where N_i is the number of samples in the ith cluster of HR images, and $C = N_1 + N_2 + \cdots + N_K$.

Similarly, the inter-cluster scatter in the projected subspace can be computed by

$$\begin{aligned} J_b^{(K)} &= 2\sum_{i=1}^{K} N_i(\mu_i - \mu)^2 \\ &= \begin{bmatrix} P_H^T & P_L^T \end{bmatrix}\sum_{i}^{K}\begin{bmatrix} B_{HH} & B_{HL} \\ B_{LH} & B_{LL} \end{bmatrix}\begin{bmatrix} P_H \\ P_L \end{bmatrix} \\ &= P^T S_b^{(K)} P, \end{aligned} \tag{12}$$

where $S_b^{(K)}$ is the inter-cluster scatter matrix of all the projections when HR images are grouped in K clusters. The blocks in $S_b^{(K)}$ are

$$\begin{aligned} B_{HH} &= \frac{1}{2}\sum_{i=1}^{K} N_i\left(\mu_{H_i} - \mu_H\right)\left(\mu_{H_i} - \mu_H\right)^T \\ B_{HL} &= \frac{1}{2}\sum_{i=1}^{K} N_i\left(\mu_{H_i} - \mu_H\right)\left(\mu_{L_i} - \mu_L\right)^T \\ B_{LH} &= \frac{1}{2}\sum_{i=1}^{K} N_i\left(\mu_{L_i} - \mu_L\right)\left(\mu_{H_i} - \mu_H\right)^T \\ B_{LL} &= \frac{1}{2}\sum_{i=1}^{K} N_i\left(\mu_{L_i} - \mu_L\right)\left(\mu_{L_i} - \mu_L\right)^T. \end{aligned} \tag{13}$$

If we want to find out which cluster a probe LR face belongs to in the projected subspace, the Fisher discriminant criterion is helpful to find the couple of mappings. According to Eq. (2), the optimal projections are obtained by

$$\begin{aligned} \begin{bmatrix} P_H & P_L \end{bmatrix}_{opt} &= \arg\max J_{Fisher}(P) \\ &= \arg\max \frac{J_b^{(K)}}{J_w^{(K)}} = \arg\max \frac{Tr\left(P^T S_b^{(K)} P\right)}{Tr\left(P^T S_w^{(K)} P\right)}. \end{aligned} \tag{14}$$

3.2. Cluster-based regularization

As our goal is to find out the class to which the probe LR face belongs, the easiest way is to set $K = C$ in Eq. (14), which indicates that no clustering is conducted. Then the optimal projections can be acquired by solving the following problem:

$$\begin{bmatrix} P_H & P_L \end{bmatrix}_{opt} = \arg\max \frac{J_b^{(C)}}{J_w^{(C)}} = \arg\max \frac{Tr\left(P^T S_b^{(C)} P\right)}{Tr\left(P^T S_w^{(C)} P\right)}, \tag{15}$$

where $S_w^{(C)}$ and $S_b^{(C)}$ are the within-class and between-class scatter matrices, respectively.

Because there is only one HR sample and one LR sample in each class, there will be only two projection samples for each subject in the unified subspace. With the limited number of samples in each class, it is difficult to accurately measure the within-class scatter matrix $S_w^{(C)}$ and the between-class scatter matrix $S_b^{(C)}$, which gives rise to overfitting and thus lowers the discriminative power of the feature subspace. To overcome this problem, we propose to regularize $S_w^{(C)}$ and $S_b^{(C)}$ respectively by using cluster-based scatter matrices $S_w^{(K)}$ and $S_b^{(K)}$. This is based on the fact that samples in the same cluster share more similarities and thus involve some information of intra-class variations; meanwhile, samples in different clusters are quite distinctive and the mean vectors of various clusters include some discriminative information [36].

In this work, we employ correlation criterion to quantitatively measure the similarity between samples, and K-means method to conduct clustering. As K-means algorithm is sensitive to initial randomly selected cluster centers, we run the clustering algorithm T times with different initializations and obtain diverse clustering results. Corresponding to different initializations, there are T intra-cluster and inter-cluster scatter matrices. Then the within-class and between-class scatter matrices are regularized by the averaged intra-cluster and inter-cluster scatter matrices, respectively. With the regularization, Eq. (15) can be rewritten as

$$\begin{bmatrix} P_H & P_L \end{bmatrix}_{opt} = \arg\max \frac{J_B}{J_W} = \arg\max \frac{Tr\left(P^T S_B P\right)}{Tr\left(P^T S_W P\right)}, \tag{16}$$

where J_W and J_B are the total within-class and between-class scatters, respectively; S_W and S_B are the total within-class and between-class scatter matrices, respectively; and

$$\begin{aligned} S_W &= \alpha S_w^{(C)} + (1-\alpha)\frac{1}{T}\sum_{r=1}^{T} S_{wr}^{(K)} \\ S_B &= \beta S_b^{(C)} + (1-\beta)\frac{1}{T}\sum_{r=1}^{T} S_{br}^{(K)}. \end{aligned} \tag{17}$$

In Eq. (17), $0 \le \alpha \le 1$ and $0 \le \beta \le 1$ are the regularization parameters that balance the weight between class-based scatter matrices and cluster-based scatter matrices. The three parameters involved in Eq. (17), i.e., α, β and K, are usually determined empirically. Because the optimization problem characterized by Eq. (16) is similar to that in SDA, meanwhile the cluster-based scatter matrices are used to regularize the class-based scatter matrices, we call the proposed approach Cluster-based Regularized SDA (C-RSDA for short). The solution of the optimization problem in (17) can be obtained by conducting the standard Eigen decomposition:

$$S_W^{-1} S_B = \lambda p, \tag{18}$$

where λ is the eigenvalue and p is the corresponding eigenvector. The optimal solution of (17) is the eigenvectors associated with the first d largest eigenvalues.

Using the cluster-based scatter matrices to regularize the class-based scatter matrices has two advantages. First, it helps to avoid the matrix singularity problem of S_W, so C-RSDA can produce more stable results. Second, it is good for mining more intra-class and inter-class variations from limited number of samples, thus the within-class and between-class scatter matrices can be more accurately estimated, which overcomes overfitting, making the unified projection subspace more discriminative.

3.3. Matching

During matching, the feature vectors of both the LR probe and HR gallery images are first projected into the unified subspace by the learned two mappings. If x_p and x_g respectively denote the feature vectors corresponding to an LR probe and an HR gallery image, the feature vectors in the common subspace are given by

$$\tilde{\mathrm{x}}_g = P_H^T \mathrm{x}_g, \ \ \tilde{\mathrm{x}}_p = P_L^T \mathrm{x}_p. \tag{19}$$

The similarity between the probe and gallery image is measured by *cosine* criterion between their projected features:

$$s = \frac{\tilde{\mathrm{x}}_g \tilde{\mathrm{x}}_p^T}{\sqrt{\left(\tilde{\mathrm{x}}_g \tilde{\mathrm{x}}_g^T\right)\left(\tilde{\mathrm{x}}_p \tilde{\mathrm{x}}_p^T\right)}}. \tag{20}$$

3.4. Relationship with SDA

When α=1 and β=1, the optimization problem in (16) is identical to that of SDA, which is formulated by Eq. (15). Therefore, SDA can be viewed as a special case of the proposed C-RSDA. SDA works well when there are multiple training images for each subject in the HR image set. If there is only one HR training sample for each subject, the number of projection samples of each class in the common subspace is small, it is difficult for SDA to truthfully characterize the within- and between-class scatter matrices. Therefore, conducting SDA in this situation will be prone to overfitting, and the within-scatter matrix will be often singular.

To avoid this problem, SDA was improved by its regularized version:

$$\begin{bmatrix} P_H & P_L \end{bmatrix}_{opt} = \arg\max \frac{J_b^{(C)}}{J_w^{(C)}} = \arg\max \frac{Tr\left(P^T S_b^{(C)} P\right)}{Tr\left(P^T \left(S_w^{(C)} + \gamma R\right) P\right)}, \tag{21}$$

in which R is the regularization term and γ is the regularization coefficient. In [25], two kinds of regularizations were proposed: R is an identity matrix or a specially designed matrix by considering local consistency [37]. On the one hand, when R is an identity matrix, it is the l_2-norm regularization, this regularization term does not include any intra-class information that will be useful for classification. On the other hand, when R is a specially designed matrix by considering local consistency, it does incorporate some local geometry information about the distribution of HR and LR samples; however, it fails to explore more variations from those local geometries. Besides, compared with the proposed C-RSDA, Eq. (21) chooses to regularize only the within-class scatter matrix, instead of both scatter matrices.

3.5. Why cluster- based regularization?

We are not the first to use clustering results of samples to do regularization. Ahead of us, several works have explored the advantage of cluster-based regularization. Soares et al. [38] introduce a cluster-based regularization term to boost the semi-supervised classifier, Wang et al. [39] apply cluster-based regularization to learn discriminative dictionary. Both two works aim to make full use of the cluster structure in designing their loss functions. To address SSPP for HR face images, Pang et al. [40] propose to replace the within-class scatter matrix with the within-cluster scatter matrix, thus LDA can be applied even if there is only one training sample of each subject. Recently, Pang et al. [36] put forward to learn regularized LDA by clustering to deal with small sample size problem of HR face recognition. These two works intent to exploit more variations from limited samples. Inspired by this, we bring in the cluster-based regularization into coupled mappings framework to tackle SSPP face recognition for LR face images. Though the above works have explained from different perspectives why cluster-based regularization is helpful, we next offer our insight into this issue.

3.5.1. Exploiting more variations from clusters

As our method employs cluster-based regularization, we first visualize some clustering results of images. We run K-means algorithm four times, Fig. 2 illustrates four clustering results in which each cluster contains Subject 17. The left subfigure shows the original HR gallery images, and the right one displays the corresponding reconstructed face images using PCA eigenvectors. It is difficult to find similarities in a cluster from Fig. 2(a), but we can figure out some common features in a cluster according to the reconstructed faces. In Fig. 2(b), the faces in the first row show similar expression (the corners of the mouth show some similarity); the faces in the second row have similar left half face; the faces in the fourth row seem to have beard, and have similar textures around the eyes. Though we cannot express what similarity exists among faces in the third row, we believe they do share some likeness since they are clustered together. Consequently, many similarity features, i.e., intra-cluster variations are extracted from clusters.

We can also observe from Fig. 2 that different clustering results share some common faces. By run the clustering algorithm more times, we find that faces in the same cluster as Subject 17 are mostly those in Fig. 2, through various combinations. This implies that Fig. 2 includes a group of similar faces. We exhibit four group of similar faces in Fig. 3. Note that analogous result as Fig. 2 shows will be obtained for each group if we run K-means algorithm several times. We can trivially discover the similarity in each group of reconstructed faces, and different groups present different similarities. Therefore, the inter-cluster variations can be exploited for each clustering results.

Together with LR gallery face images, we can get two parts of cluster-based variations in the common projected subspace, both of them contribute to enhancing the discriminative ability of the subspace. Figs. 6 and 7 illustrate the usefulness of cluster-based variations.

3.5.2. Avoiding overfitting and singularity

We have demonstrated that variations can be explored from clustering results of face images, now we disclose how the cluster-based variations help avoid overfitting and singularity problem. Ac-

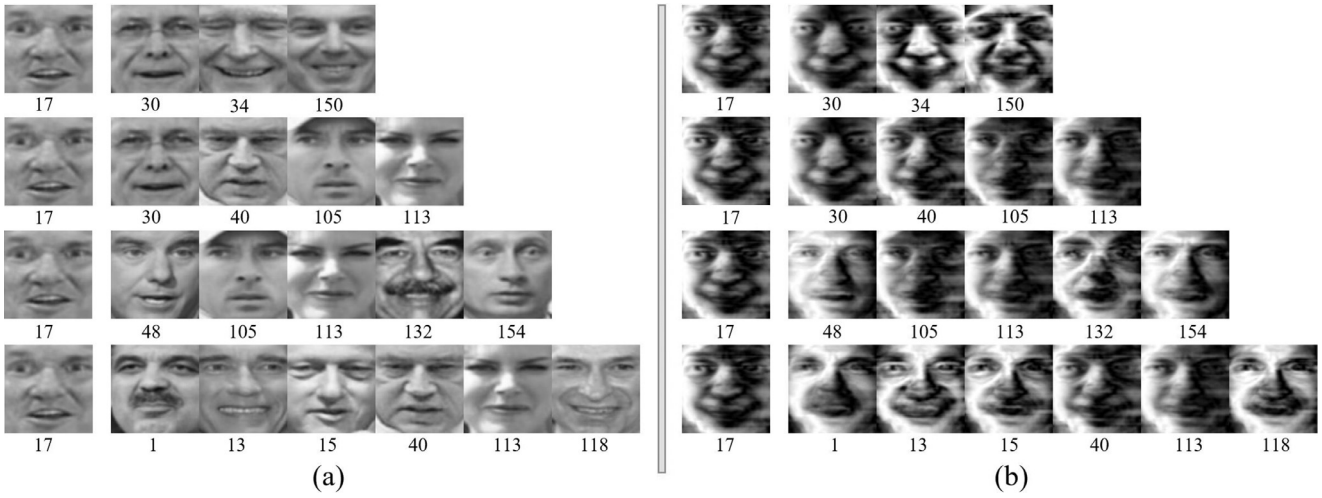

Fig. 2. Four clustering results involving subject 17: (a) the original HR gallery faces, (b) the reconstructed faces. The numbers below the images denote their class labels.

Fig. 3. Four group similar faces (for each group, images in the top row are the original HR gallery faces and those in the bottom row are the reconstructed faces).

cording to Eqs. (9), (16) and (17), we reformulate the total within-class scatter as follows:

$$
\begin{aligned}
J_W &= P^T\left(\alpha S_w^{(C)} + (1-\alpha)\frac{1}{T}\sum_{r=1}^{T} S_{wr}^{(K)}\right)P \\
&= \alpha\left(P^T S_w^{(C)} P + \frac{1-\alpha}{\alpha T}\sum_{r=1}^{T} P^T S_{wr}^{(K)} P\right) \\
&= \alpha\left(\sum_{i=1}^{C}\left(\sum_{c(h_j)=i}\left(\tilde{h}_i-\mu_i\right)^2 + \sum_{c(l_j)=i}\left(\tilde{l}_i-\mu_i\right)^2\right)\right. \\
&\quad \left.+\frac{1-\alpha}{\alpha T}\sum_{r=1}^{T}\sum_{ii=1}^{K}\left(\sum_{c(h_j)=ii}\left(\tilde{h}_{ii}-\mu_{ii}\right)^2 + \sum_{c(l_j)=ii}\left(\tilde{l}_{ii}-\mu_{ii}\right)^2\right)\right) \\
&= \alpha\left(\sum_{i=1}^{C}\sum_{c(h_j)=i}\left(\tilde{h}_i-\mu_i\right)^2 + \sum_{i=1}^{C}\sum_{c(l_j)=i}\left(\tilde{l}_i-\mu_i\right)^2\right. \\
&\quad +\frac{1-\alpha}{\alpha T}\sum_{r=1}^{T}\sum_{ii=1}^{K}\sum_{c(h_j)=ii}\left(\tilde{h}_{ii}-\mu_{ii}\right)^2 \\
&\quad \left.+\frac{1-\alpha}{\alpha T}\sum_{r=1}^{T}\sum_{ii=1}^{K}\sum_{c(l_j)=ii}\left(\tilde{l}_{ii}-\mu_{ii}\right)^2\right) \\
&= \alpha\left(\sum_{i=1}^{C}\left(\sum_{c(h_j)=i}\left(\tilde{h}_i-\mu_i\right)^2 + \frac{1-\alpha}{\alpha TC}J_{w-H}^{cluster}\right)\right. \\
&\quad \left.+\sum_{i=1}^{C}\left(\sum_{c(l_j)=i}\left(\tilde{l}_i-\mu_i\right)^2 + \frac{1-\alpha}{\alpha TC}J_{w-L}^{cluster}\right)\right),
\end{aligned}
\tag{22}
$$

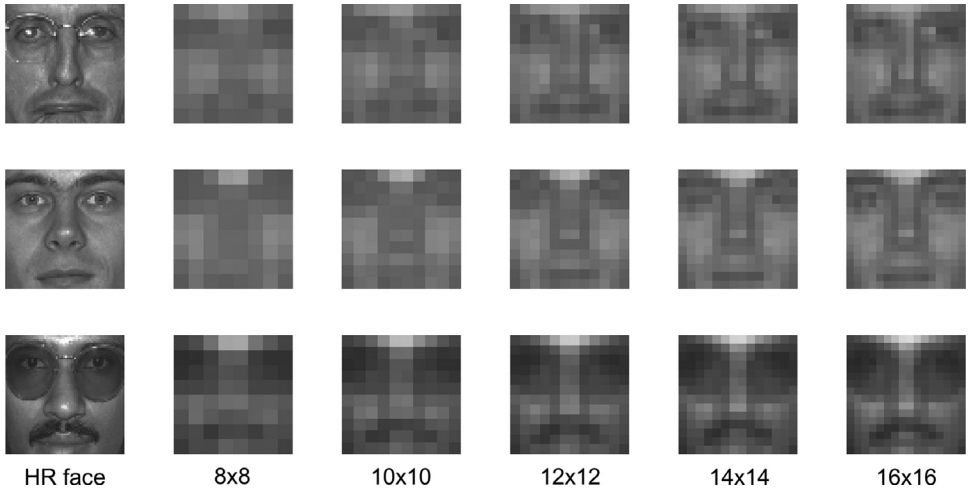

Fig. 4. The illustration of HR and LR images.

where $J_{w-H}^{cluster}=\sum_{r=1}^{T}\sum_{ii=1}^{K}\sum_{c(h_j)=ii}(\tilde{h}_{ii}-\mu_{ii})^2$ $J_{w-L}^{cluster}=\sum_{r=1}^{T}\sum_{ii=1}^{K}\sum_{c(l_j)=ii}(\tilde{l}_{ii}-\mu_{ii})^2$ denote the within-cluster scatter in the common subspace computed by HR images and LR images, respectively. According to Eqs. (12), (16) and (17), we reformulate the total between-class scatter as follows:

$$\begin{aligned} J_B &= P^T\left(\beta S_b^{(C)}+(1-\beta)\frac{1}{T}\sum_{r=1}^{T}S_{br}^{(K)}\right)P \\ &= \beta\left(P^T S_b^{(C)}P+\frac{1-\beta}{\beta T}\sum_{r=1}^{T}P^T S_{br}^{(K)}P\right) \\ &= 2\beta\left(\sum_{i=1}^{C}N_i(\mu_i-\mu)^2+\frac{1-\beta}{\beta T}\sum_{r=1}^{T}\sum_{ii=1}^{K}N_{ii}(\mu_{ii}-\mu)^2\right) \\ &= 2\beta\left(\sum_{i=1}^{C}\left(N_i(\mu_i-\mu)^2+\frac{1-\beta}{\beta TC}J_b^{cluster}\right)\right), \end{aligned} \tag{23}$$

where $J_b^{cluster}=\sum_{r=1}^{T}\sum_{ii=1}^{K}N_{ii}(\mu_{ii}-\mu)^2$ denotes the between-cluster scatter in the common subspace computed by all samples.

Because there is only one HR gallery image and only one LR gallery image for each class, we can seek few class-based variations from them, which leads to overfitting for classification. However, we observe from Eq. (22) that a proportion of the total within-scatter is added to each class, which means that intra-cluster variations exploited from clustering results are used to enrich the intra-class variations for each class; similarly, we observe from Eq. (23) that a proportion of the total between-scatter is summed to each class, that is to say, inter-cluster variations exploited from clustering results are used to enrich the inter-class variations. Therefore, more variations (include class-based and cluster-based variations) are embedded into the common subspace and overfitting is solved.

Note that $S_W=\alpha S_w^{(C)}+(1-\alpha)\frac{1}{T}\sum_{r=1}^{T}S_{wr}^{(K)}$. For every clustering result, there are multiple samples in most of clusters, hence the intra-cluster scatter matrix is always non-singular. As a result, the total within-class scatter matrix S_W will be non-singular, and stable results can be obtained.

4. Experimental evaluation

In this section, we describe the details of extensive experiments, which are performed to evaluate the effectiveness of the proposed approach to match LR probe images with HR gallery face images.

4.1. Data description

Most experiments described in this paper are performed on FERET face data set [41]. As the grayscale FERET database is not available to be downloaded any more, all grayscale images used in our experiments are obtained from the color FERET database currently distributed in the official website. The FERET database consists of 13,539 facial images corresponding to 1565 subjects, who are diverse across ethnicity, gender, and age. In our experiments, a subset from FERET database is chosen to compare our approach with the state-of-the-art methods that can address low-resolution face recognition problem. There are four groups (Fa, Fb, Dup1, and Dup2) in the selected subset. Fa containing 994 frontal images of 994 subjects, is used as Gallery, while Fb (992 images of expression variations), Dup1 (736 images), and Dup2 (228 images) are the Probe sets. It is worth mentioning that we are the first to evaluate the LR face recognition performance on Dup1 and Dup2 probe sets of FERET database.

We also perform matching experiments on two more challenging databases, i.e., labeled faces in the wild (LFW) [42] and Surveillance Cameras Face (SCface) database [43]. The two databases include face images captured in the uncontrolled environment. Brief description of the datasets is provided along with the details of the experiments.

4.2. Experimental settings

All face images are aligned and cropped using the location of their eyes. The HR images are resized to 64 × 64 pixels, and the LR images are down-sampled from the HR ones using a standard bilinear interpolation technique. In our work, we test LR face images of 5 different resolutions, i.e., 8 × 8, 10 × 10, 12 × 12, 14 × 14 and 16 × 16 pixels. Fig. 4 shows three HR face images and their corresponding LR samples in FERET subset. For training, the proposed algorithm needs both HR and LR images of the same subject to learn the coupled mappings. Note that there are only one HR sample and one LR sample for each subject in the training set.

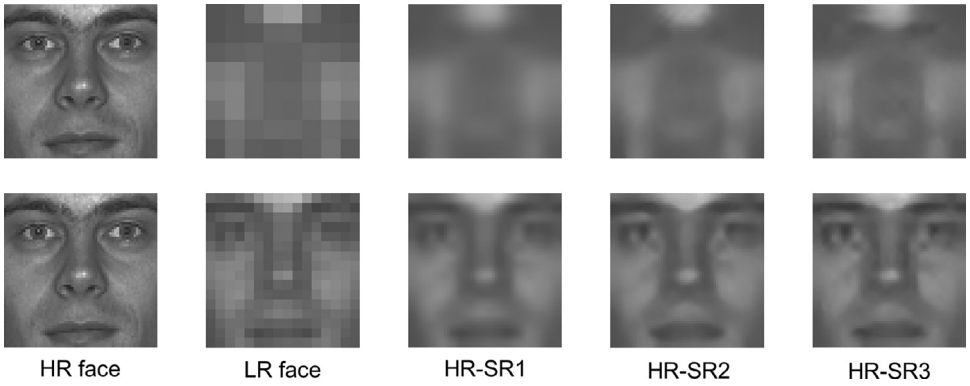

Fig. 5. Comparison of HR images created by different super-resolution algorithms. The 1st column is the original HR faces, and the input LR faces for super-resolution algorithms are 8 × 8 pixels (the 1st row) and 16 × 16 pixels (the 2nd row).

Table 1
The baseline recognition results (%) on FERET dataset.

	L-L					H-H	
	8 × 8	10 × 10	12 × 12	14 × 14	16 × 16	PCA	LBP
Fb	69.46	71.07	71.47	71.77	72.08	74.37	93.00
Dup1	29.76	32.47	35.46	36.55	37.50	42.70	61.00
Dup2	13.60	17.11	20.18	20.61	21.05	30.46	50.00

Notes: 'L-L' means LR images are used in both training and testing process, 'H-H' means HR images are used in both training and testing process.

Table 2
Recognition performance (%) comparison of super-resolution techniques.

		SR1	SR2	SR3	**C-RSDA**	L-L	H-H
8 × 8	Fb	37.30	21.17	27.72	82.36	69.46	74.37
	Dup1	5.71	6.39	8.56	41.85	29.76	42.70
	Dup2	0.88	2.63	6.14	34.65	13.60	30.46
16 × 16	Fb	70.26	73.29	74.29	86.29	72.08	74.37
	Dup1	31.93	38.41	41.58	59.24	37.50	42.70
	Dup2	21.93	28.07	33.77	51.75	21.05	30.46

In all experiments, the standard Principal Component Analysis (PCA) [44] is firstly used to eliminate the useless information in the original HR and LR images, which induces the improvement of recognition accuracy. The number of PCA coefficients used to represent the face images is determined based on the number of eigenvalues preserving 99.99% of the total energy. The parameters in the proposed algorithm, i.e., K, T, α and β are empirically set to 20, 80, 0.6 and 0.3 respectively when conducting experiments on FERET database. The choices of various parameter values are discussed in latter subsection. Unless otherwise stated, all experiments report accuracy in terms of rank-1 recognition performance.

4.3. Experimental results

4.3.1. Baseline results

To show clearly how effective the proposed approach is in recognizing LR faces, we first obtain some baseline recognition results to be compared with. One part of baseline results is obtained by matching LR probe images with down-sampled gallery images, which is denoted by L-L in the reported results. The other part of results are gained by directly matching high-resolution version of probe LR images against HR gallery images, which is denoted by H-H. For experiments in this subsection, we employ PCA technique to extract features and the nearest neighbor classifier to make classification. We also use LBP feature to get better recognition performance in H-H setting. Table 1 displays these baseline recognition results on FERET database. It is obvious that the performance in L-L setting is typically much worse than that in H-H setting, and the recognition accuracy is from bad to worse with the decrease of resolution in L-L setting.

4.3.2. Performance comparison with super-resolution techniques

For matching an LR probe with an HR gallery image, one of the most commonly used approaches is to obtain an HR reconstruction of the LR probe image firstly using a super-resolution technique, and then the reconstructed HR probe images is used for matching with the HR gallery sample. For comparison, we follow the work of [30] and apply three different super-resolution techniques to obtain HR images from LR probe images. The three techniques are Bicubic Interpolation (SR1), Sparse Regression-based super-resolution (SR2) [10] and Sparse Representation-based super-resolution (SR3) [11]. For Bicubic Interpolation, a standard MATLAB interpolation function is utilized; while for the other two methods, we have used the code available from the authors' websites. In this subsection, we choose the probe images of two different resolutions, namely, 8 × 8 and 16 × 16 to evaluate various approaches. Fig. 5 shows examples of HR images created by the three super-resolution techniques described above.

As the proposed C-RSDA is based on discriminant analysis, to provide fair comparisons, we get the recognition rates of super-resolutions based methods by using LDA technique. Specifically, (i) the HR counterparts of gallery LR images are firstly obtained by super-resolution techniques, these virtual HR images together with the original gallery HR images constitute the training set; (ii) each class has two samples in the training set, PCA+ LDA technique is employed to get the discriminative subspace and then to extract training features from only the original HR images; (iii) the HR counterparts of the LR probe images are obtained by super-resolution techniques and then projected into the discriminative subspace to extract testing features; (iv) matching testing features against training features by nearest neighbor classifier with cosine metric. Table 2 presents the rank-1 recognition performance of the proposed approach along with those of the three different super-resolution approaches on FERET data subset.

As can be seen, when the resolution of probe LR faces is quite low, namely 8 × 8 pixels, every super-resolution method performs badly, even much worse than the standard LR matching (L-L), which implies that it is quite difficult to recover meaningful information from images of such a low resolution. When the resolution of probe LR images increases to 16 × 16 pixels, the super-resolution methods perform much better, however, the performance of SR1 is still inferior to that of L-L matching, while the performance of the two elaborated super-resolution techniques is better than L-L matching but still worse than H–H baseline performance. This may be partly attributed to the fact that these super-

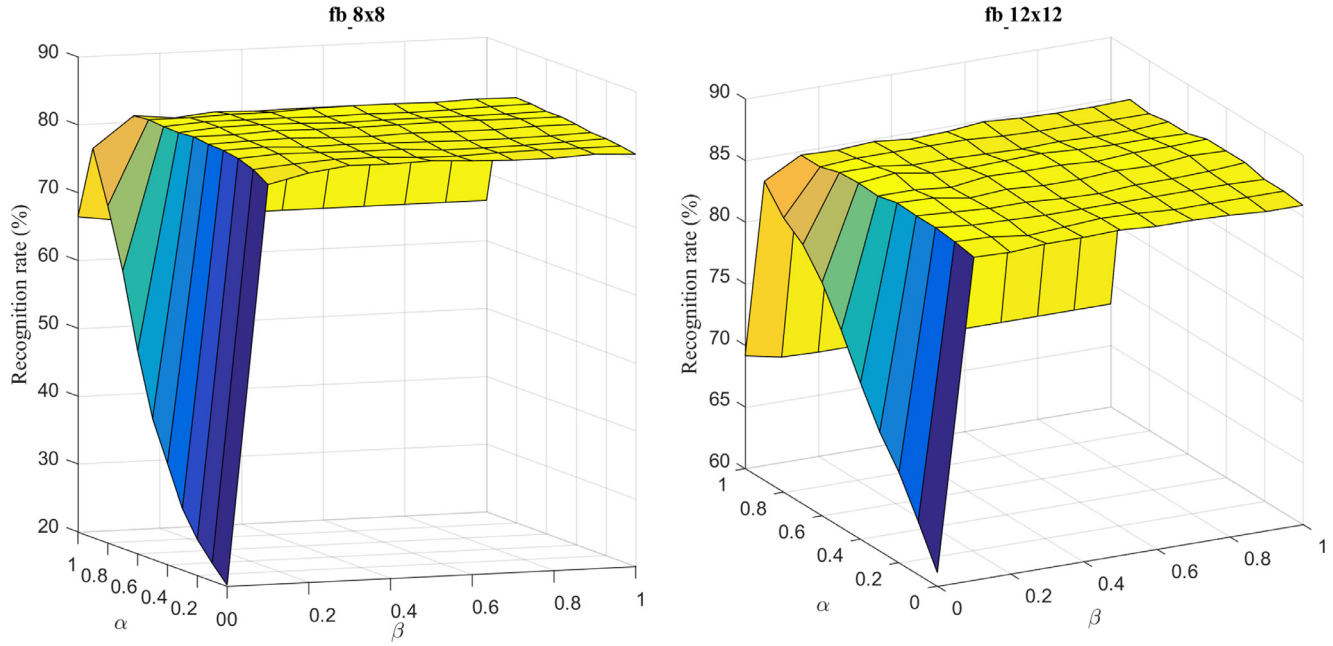

Fig. 6. Recognition accuracy versus various values of α and β for different low resolutions.

resolution techniques are not customized for a face matching application. Not surprisingly, the proposed approach performs significantly better than all the super-resolution techniques in every scenario, and its performance is much better than the HR versus HR setting in most cases, even though the resolution of probe LR faces is quite low.

4.3.3. Performance comparison with coupled mappings approaches

Since C-RSDA learns two mappings, we then compare our proposed method with three different coupled mappings approaches, namely, CLPM [22], CMFA [23] and SDA [25]. To make fair comparisons, we finely tune the parameters for each compared technique. For CLPM, the scale parameter α and the number of nearest neighbors of an HR image $N(i)$ are set to be 11 and 1, respectively; for CMFA, the regularization constant α, the number of the nearest intra-class neighbors k_1 and the number of the nearest inter-class neighbors $k_2$are set to be 10^{-5}, 1 and 20, respectively; for SDA, we employ l_2 – norm to regularize the within-class scatter matrix, as the authors did in their work, and tune two regularization coefficients to be $\gamma=\eta=10^{-6}$. All the approaches are evaluated on the LR probe images of 5 different resolution, and for each experiment the nearest neighbor classifier with cosine metric is used to make classification. The recognition performance of the compared methods, together with the baseline performance, is tabulated in Table 3.

The effectiveness of the our proposed C-RSDA can be demonstrated by the following three facts: (i) among the four coupled mapping methods, the C-RSDA achieves the best performance for each experiment on Fb and Dup1 probe sets, and performs best or nearly best on Dup2 probe set; (ii) according to the average performance over various resolutions, C-RSDA reaches the highest recognition rates of 85.36% for Fb probe set, 51.44% for Dup1 and 43.07% for Dup2; (iii) compared with the baseline performance of HH(PCA), C-RSDA attains either better or much better performance in all experiments; (iv) compared with the baseline results of HH(LBP), the overall performance of C-RSDA is lower by a margin within only 10%.

In addition, it can be observed from Table 3 that C-RSDA largely surpasses SDA in each experiment, which indicates that the cluster-based regularization employed in the proposed method is very efficient in LR face recognition when there is single HR image for each subject in the training set.

Table 3
Recognition performance (%) comparison with couple mapping techniques.

		8×8	10×10	12×12	14×14	16×16	Ave.
Fb	CLPM	79.94	82.56	82.46	84.07	84.48	82.70
	CMFA	72.08	75.40	75.40	74.60	75.60	74.62
	SDA	68.75	72.08	71.77	71.88	72.08	71.31
	C-RSDA	82.36	85.48	86.29	86.39	86.29	85.36
	HH(PCA)	74.37	74.37	74.37	74.37	74.37	74.37
	HH(LBP)	93.00	93.00	93.00	93.00	93.00	93.00
Dup1	CLPM	39.54	44.16	50.95	56.52	59.24	50.08
	CMFA	24.86	28.26	32.88	35.33	36.14	31.49
	SDA	27.45	31.66	33.97	37.50	37.77	33.67
	C-RSDA	42.85	46.60	52.85	56.66	59.24	51.64
	HH(PCA)	42.70	42.70	42.70	42.70	42.70	42.70
	HH(LBP)	61.00	61.00	61.00	61.00	61.00	61.00
Dup2	CLPM	31.14	35.96	42.11	47.81	52.19	41.84
	CMFA	13.60	20.61	23.25	22.81	24.12	20.88
	SDA	14.04	18.42	18.86	20.61	20.18	18.42
	C-RSDA	34.65	37.72	43.86	47.37	51.75	43.07
	HH(PCA)	30.46	30.46	30.46	30.46	30.46	30.46
	HH(LBP)	50.00	50.00	50.00	50.00	50.00	50.00

4.4. Parameter analysis

We now evaluate the robustness of the proposed C-RSDA algorithm with different parametric choices. The proposed method involves four parameters, namely, regularization parameter α and β, the cluster number K and the clustering times T. In the following analysis, we conduct experiments using HR gallery images of resolution 64×64 pixels, LR probe images of 8×8 and 12×12 pixels.

4.4.1. The effect of regularization parameter α and β

The parameter α and β are used to regularize the within-class and between-class scatter matrices, they play a crucial role in the proposed algorithm. Two sets of experiments are conducted by fixing $K=20$ and $T=80$, varying both α and β from 0 to 1 by an increment of 0.1. The first set of experiments are used to examine

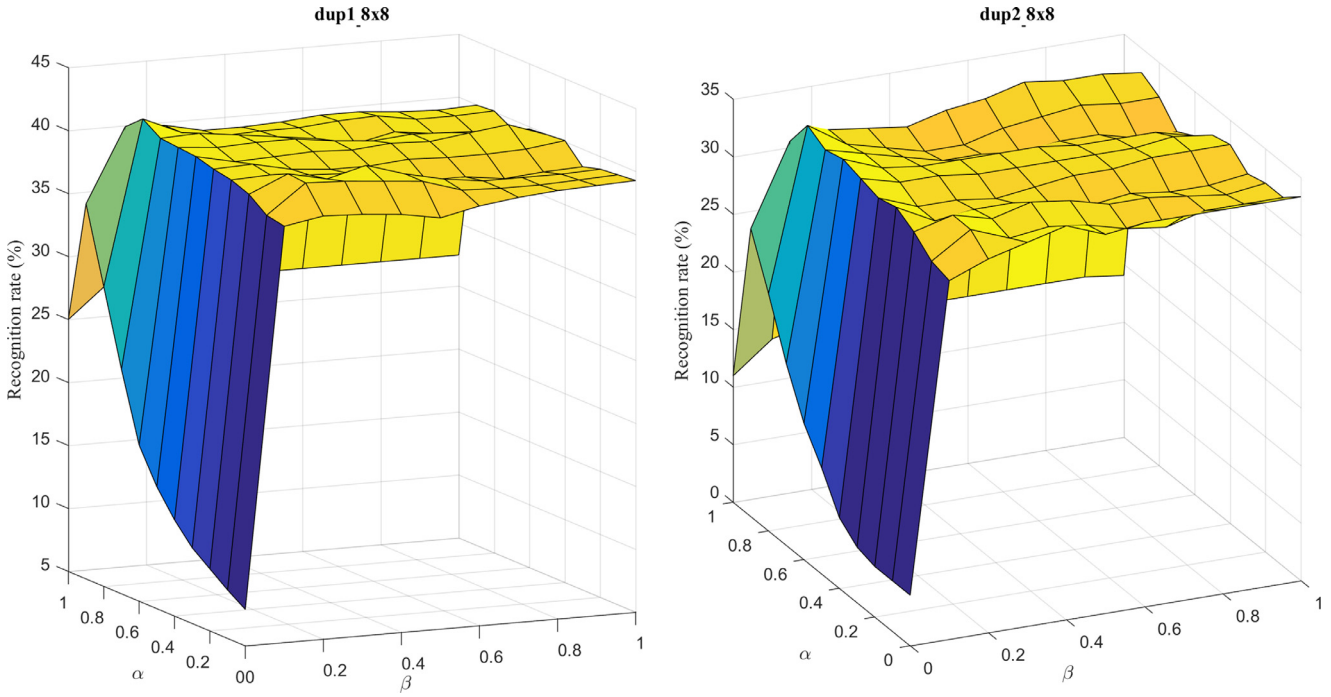

Fig. 7. Recognition accuracy versus various values of α and β on different probe sets.

the effect of parameters when C-RSDA is employed to recognize LR images of different resolutions. Fig. 6 shows the plots of rank-1 recognition rates on Fb probe subset versus various values of α and β, the results in left and right subfigures are obtained when probe LR images in Fb subset are 8×8 pixels and 12×12 pixels, respectively. The second set of experiments are used to examine the effect of parameters when C-RSDA is employed to recognize LR images in different probe sets. Fig. 7 shows the plots of rank-1 recognition rates on Dup1 and Dup2 probe subsets versus various values of α and β, the results in left and right subfigure are acquired when Dup1 and Dup2 are used as probe sets, respectively; LR images in both probe sets in this experiment are 8×8 pixels.

As can be seen, figures in both Figs. 6 and 7 have similar 3D shapes, we describe it in detail in the following. When $\alpha = \beta = 0$, the recognition rate is lowest, it is as expected since the two total scatter matrices are replaced by the cluster-based scatter matrices. When $\alpha = \beta = 1$, which means that no cluster-based regularization is imposed on the scatter matrices, the corresponding recognition accuracy is significantly better than that when $\alpha = \beta = 0$, but much lower than the highest point. When $\alpha = \beta = 0$, though the recognition performance is quite low, it is higher than zero and occupies more than one quarter of the recognition rates when $\alpha = \beta = 1$. This reveals that the cluster-based scatter matrices are beneficial for classification. The highest recognition rate is achieved when $0 < \alpha < 1$ and $0 < \beta < 1$ in all subfigures. When $0 < \alpha < 1$ and $0 < \beta < 1$, the recognition accuracy in each subfigure slightly fluctuates around its highest value. It suggests that the proposed C-RSDA algorithm can perform well with very flexible choices of parameter values. Specially, when $\beta = 0$, the recognition rate ascends with the increase of the value of α, but it drops before reaches the peak point; when $\alpha = 0$, the recognition rate does not show apparent trend with the changes of β. This implies that the intra-cluster scatter matrix plays a more influential role in boosting the performance than the inter-cluster scatter matrix.

4.4.2. The effect of the number of clusters K

We utilize the cluster-based scatter matrices to do regularization in our proposed method, so the number of clusters is another vital parameter in our algorithm. In this subsection, we investigate the effect of cluster number K on the recognition performance. Two sets of experiments are conducted by fixing α=0.6, β=0.3 and $T = 80$, varying K from 5 to 40 by an increment of 5. Fig. 8 illustrates the recognition rates versus various values of K. The left subfigure shows the results on Fb probe set while the left one shows the results on Dup1 and Dup2 probe sets.

As it is shown, when the LR probe sets are fb_12×12, dup1_8×8 and dup1_12×12, the corresponding recognition rate firstly rises and then drops with the increase of K, and reaches their peak value at K=20; when fb_12×12 is the probe set, the recognition rate shows an ascending trend with the increase of K, but it goes up slightly when K is more than 20. Therefore, neither too small nor too large values of K are beneficial, and a moderate value of K can ensure the proposed C-RSDA to achieve satisfying results.

Table 4
The recognition accuracy (%) and standard deviation with respect to various cluster times T.

T	20	40	60	80	100
Ave.	82.31	82.49	82.51	82.61	82.58
Std.	0.0045	0.0032	0.0030	0.0022	0.0016

4.4.3. The effect of clustering times T

To eliminate the influence caused by the instability of K-means algorithm, we run the clustering algorithm for T times to calculate the cluster-based scatter matrices. The effect of T is tested in this part. We use Fb, which contains LR image of 8×8 pixels, as probe set, set α=0.6, β=0.3 and $K = 20$, and vary the value of T to conduct experiments. For each value of T, we run C-RSDA algorithm 10 times and get the average recognition rates. Table 4 displays the recognition accuracy (%) and standard deviation with respect to various clustering times T. From Table 4, we find that the value of T has slight effect on the recognition performance when $T > 20$. Furthermore, the recognition rate is quite stable by run the clustering algorithm for over 20 times since all of the standard devia-

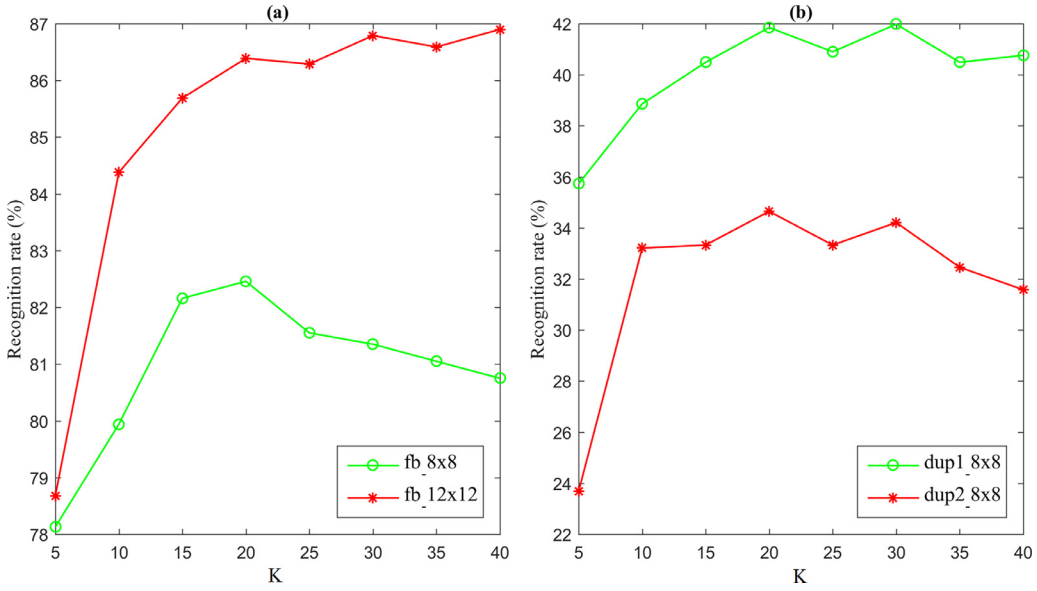

Fig. 8. The recognition rate versus different values of *K*.

Table 5
The recognition results (%) on LFW dataset.

	8×8	12×12	16×16	Ave.
CLPM	4.10	6.58	7.47	6.05
CMFA	9.49	10.38	10.89	10.25
SDA	4.94	6.96	7.85	6.58
C-RSDA	12.28	14.43	15.06	13.92
H–H(PCA)	3.92	3.92	3.92	3.92

tions are less than 0.005, though it is more stable when the value of T is larger. Considering the efficiency and stability, selecting the value of T between 40 to 80 is a favorable choice for the proposed method.

4.5. Evaluation on unconstrained face images

To show the recognition performance of our C-RSDA algorithm on face images captured in unconstrained circumstances, we conduct two series of experiments on LFW database and SCface database.

4.5.1. Evaluation on LFW database

The LFW contains images of 5749 individuals taken under an unconstrained setting. The complex surroundings of image capturing and inaccurate alignment of faces make the LFW data quite challenging for LR face recognition in the SSPP setting. LFW-a is a subset of the LFW dataset, and the images in LFW-a have been aligned with a commercial software tool. In this subsection, we gather the subjects containing no less than ten samples and then get a dataset with 158 subjects from LFW-a database, and further choose one most frontal face image for each subject to construct the HR gallery set, then select the first 5 face images from the remaining ones for every person to construct HR probe set. All HR images are resized to 64×64 pixels, the LR gallery images and LR probe images are obtained by downsampling corresponding HR ones via bilinear interpolation. Our task on this database is to recognize LR face images of three resolutions,8×8, 12×12, 16×16 pixels. Fig. 9 displays the sample images of one subject in LFW subset.

We use the same experimental setting as before to run the proposed algorithm on LFW subset. Table 5 displays the recognition results of various methods on this data set. It clearly shows that the recognition accuracy is quite low (less than 17%) for each method in every case. The reason is that face images in LFW face database are captured under unconstrained circumstances and much more complicated than images in FERET dataset, these face images are hardly frontal and each probe face involves at least one variation of expression, pose, illumination and occlusion. On this challenging dataset, the proposed C-RSDA performs best, it achieves the accuracy larger than 12% for each low resolution and the averaged accuracy is 13.92%; the following is CMFA, whose averaged recognition rate is 10.25%, 2.67% lower than C-RSDA; CLPM and SDA are competitive, both accuracies are much less than 10%. We also observe that each CMs based method performs better than the baseline method PCA, which indicates that the unsupervised PCA technique is inferior to deal with such complicated face images. According to Table 5, we see the advantage of C-RSDA on recognizing LR faces captured in complicated environment.

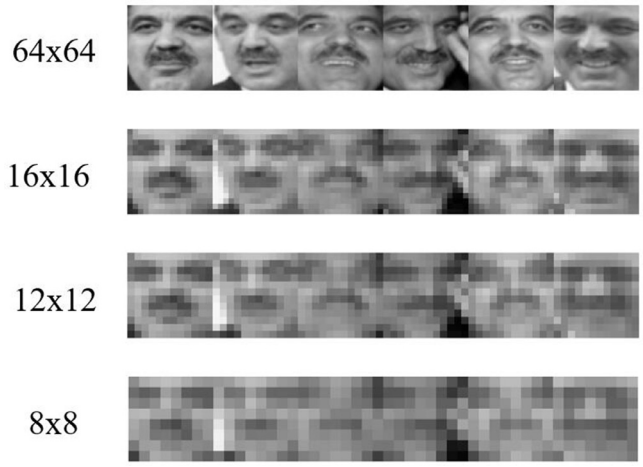

Fig. 9. Sample images in LFW subset (Images in the 1st column are gallery images, and the rest images are probe ones).

4.5.2. Evaluation on SCface database

The SCface database contains images of 130 subjects taken in uncontrolled indoor environment using five video surveillance

Fig. 10. Sample images in SCface database.

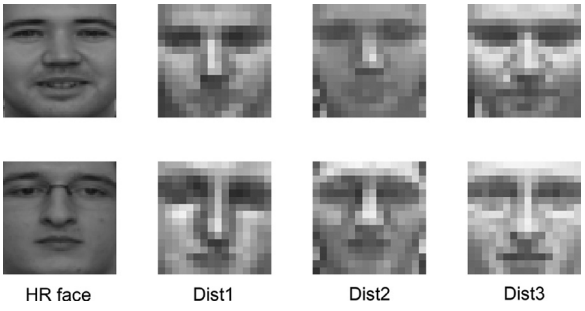

Fig. 11. The cropped grey-scaled images of samples in SCface database.

cameras at three different distances (4.20, 2.60 and 1.00 m). As in typical commercial surveillance systems, the database was collected with the camera placed slightly above the subject's head and also the individuals were not required to look at a fixed point during the recordings, which is most like the real-world circumstance, thus making the dataset much more challenging. In our experiment, a subset containing five sections of images is selected: the first section includes images captured by a high-quality photo camera, images in this section are HR faces; the other four sections include images captured by camera 1–4 at 3 different distances, images in these sections are LR faces. Fig. 10 illustrates some sample images of two subjects in SCface dataset. For HR faces, we align and crop them into 64 × 64 pixels; for all of the LR faces, we align and crop them into 16 × 16 pixels. Fig. 11 shows the cropped images of samples in Fig. 10.

We conduct two sets of experiments on SCface database to evaluate the LR face recognition algorithms. In the first set of experiments, 130 subjects' HR face images and their virtual LR images, which are obtained by downsampling the corresponding HR ones, are used for training purpose. The LR images captured by 4 cameras at 3 distances are considered as probe samples, thus there are 12 probe sets. It worth noting that the training LR images and probe LR ones are not captured in the same condition, which makes the recognition task more difficult. We set $K = 8$ and use the same values of other parameters as before to run the proposed algorithm. Table 6 displays the recognition results of various methods in this strict scenario. We only compare the proposed algorithm to the coupled mapping based methods, because these methods have shown better performance than super-resolution based methods in the previous experiments. As can be

Table 6
The recognition results (%) on SCface dataset when virtual LR images are used for training.

		Cam1	Cam2	Cam3	Cam4	Ave.
Dis1 (4.2 m)	CLPM	1.54	2.31	2.31	2.31	2.12
	CMFA	3.85	3.85	3.08	3.08	3.46
	SDA	1.54	2.31	0.77	3.08	1.92
	C-RSDA	6.92	4.62	5.38	3.85	5.19
Dis2 (2.6 m)	CLPM	1.54	3.85	1.54	2.31	2.31
	CMFA	3.08	3.08	4.62	4.62	3.85
	SDA	1.54	1.54	2.31	5.38	2.69
	C-RSDA	7.69	6.92	3.85	5.38	5.96
Dis3 (1 m)	CLPM	3.08	3.08	1.54	2.31	2.50
	CMFA	4.62	6.15	3.08	3.08	4.23
	SDA	4.62	3.08	3.85	3.85	3.85
	C-RSDA	6.15	7.69	4.62	6.15	6.15

seen from Table 6, all coupled mapping based approaches show extremely low recognition rates, which are less than 8%. This attributes to the fact that the gallery LR images are virtually generated, which is not helpful for extracting discriminative features to recognize those real LR images captured by cameras. Though the experimental setting is quite rigorous, the proposed C-RSDA exhibits better performance than other compared methods.

In the second set of experiments, we choose the HR images of the first 65 subjects to be the HR gallery images, and real LR images (which were captured by cameras) of the first 65 subjects to be the LR gallery images. The real LR images of remaining 65 subjects are elected to be the LR probe images, hence the gallery LR images and probe LR ones are acquired in the same condition, but there is no overlapping subjects in the gallery and probe sets. We set $K = 2$ and use the same values of other parameters as before. Table 7 displays the recognition results of compared methods on SCface database when the real LR face images are used in the training process. It can be seen from Table 7, the recognition performance of all the methods is far from satisfying. However, compared with other methods, the proposed C-RSDA achieves the highest recognition rates on this real-world face dataset, which again demonstrate its effectiveness on LR face recognition. As shown in Table 7, the recognition rates in this experiment of all methods are higher than those in the former experiment. One intuitive reason is that there are 65 subjects in the probe sets, which is the half of that in the previous experiment. But the primary cause leading to the improvement of recognition

Table 7
The recognition results (%) on SCface dataset when real LR images are used for training.

		Cam1	Cam2	Cam3	Cam4	Ave.
Dis1 (4.2 m)	CLPM	3.08	4.62	3.08	3.08	3.46
	CMFA	4.62	4.62	4.62	3.08	4.23
	SDA	12.31	12.31	18.46	12.31	13.85
	C-RSDA	13.85	13.85	20.00	15.38	15.77
Dis2 (2.6 m)	CLPM	3.08	6.15	4.62	3.08	4.23
	CMFA	5.38	5.38	7.69	5.38	5.96
	SDA	15.38	18.64	10.77	23.08	16.97
	C-RSDA	16.92	20.00	10.77	24.62	18.08
Dis3 (1 m)	CLPM	6.15	1.54	1.54	3.08	3.08
	CMFA	10.00	4.62	7.69	8.46	7.69
	SDA	16.92	12.31	20.00	16.92	16.54
	C-RSDA	20.00	10.77	23.08	20.00	18.46

performance is that the real LR face images are used to extract features. By comparing Table 6 and Table 7, we can observe that SDA and C-RSDA significantly promote the recognition rates after using real LR training images, whereas CLPM and CMFA get slight improvement. The proposed approach again presents the best performance over most probe sets, and achieves the best average recognition rates of 15.77%, 18.08% and 18.46% respectively when the distance between subjects and cameras are 4.2 m, 2.6 m and 1 m.

5. Conclusion

In this paper, we propose a novel coupled mappings based method, named cluster-based regularized SDA, to cope with low-resolution face recognition problem where there is only one HR training sample in each class. The main idea is to employ cluster-based scatter matrices to regularize the class-based scatter matrices and then to learn coupled mappings simultaneously to project HR and LR images into a discriminative feature subspace. With limited training samples, we introduce more discriminative information into the feature subspace by exploiting variations from clustering results. We conduct extensive experiments on FERET dataset, and the impressive results demonstrate that our approach outperforms some of the state-of-the-art super-resolution techniques as well as coupled mappings based methods. We then test the proposed approach on two more complicated databases, i.e., labeled faces in the wild (LFW database) and a real surveillance face database (SCface database), the experimental results on the two challenging datasets further show its effectiveness on low-resolution face recognition.

Acknowledgement

This work was partially supported by the National Natural Science Foundation of China (No.71390333), the National Key Technology R&D Program of China during the 12th Five-Year Plan Period (No.2013BAD19B05), the Special Fund for Basic Research in Central University, i.e., Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYZZ15_0072), and the Study Abroad Program for Graduate Studies by China Scholarship Council. Portions of the research in this paper used the SCface database of facial images. Credit is hereby given to the University of Zagreb, Faculty of Electrical Engineering and Computing, for providing the database of facial images. The authors would like to thank the reviewers for their valuable comments and suggestions, which help improve the work.

References

[1] W. Chen, Y. Gao, Recognizing partially occluded faces from a single sample per class using string-based matching, in: Computer Vision–ECCV 2010, Springer, 2010, pp. 496–509.
[2] J. Lu, Y.-P. Tan, G. Wang, Discriminative multimanifold analysis for face recognition from a single training sample per person, Pattern Anal. Mach. Intell. IEEE Trans. 35 (2013) 39–51.
[3] F. Yin, L. Jiao, F. Shang, L. Xiong, S. Mao, Double linear regressions for single labeled image per person face recognition, Pattern Recognit. 47 (2014) 1547–1558.
[4] 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. Vision 111 (2015) 365–383.
[5] S. Chen, J. Liu, Z.-H. Zhou, Making FLDA applicable to face recognition with one sample per person, Pattern Recognit. 37 (2004) 1553–1555.
[6] S. Chen, D. Zhang, Z.-H. Zhou, Enhanced (PC) 2 A for face recognition with one training image per person, Pattern Recognit. Lett. 25 (2004) 1173–1181.
[7] Z. Wang, Z. Miao, Q.J. Wu, Y. Wan, Z. Tang, Low-resolution face recognition: a review, Visual Comput. 30 (2014) 359–386.
[8] R. Timofte, V. De Smet, L. Van Gool, Anchored neighborhood regression for fast example-based super-resolution, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 1920–1927.
[9] R. Timofte, V. De Smet, L. Van Gool, A+: adjusted anchored neighborhood regression for fast super-resolution, in: Asian Conference on Computer Vision, Springer, 2014, pp. 111–126.
[10] K.I. Kim, Y. Kwon, Single-image super-resolution using sparse regression and natural image prior, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 1127–1133.
[11] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (2010) 2861–2873.
[12] H. Chang, D.-Y. Yeung, Y. Xiong, Super-resolution through neighbor embedding, in: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, IEEE, 2004 I–I.
[13] Z. Wang, D. Liu, J. Yang, W. Han, T. Huang, Deep networks for image super-resolution with sparse prior, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 370–378.
[14] C. Dong, C.C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2016) 295–307.
[15] E. Bilgazyev, B.A. Efraty, S.K. Shah, I.A. Kakadiaris, in: Sparse Representation-Based Super Resolution for Face Recognition At a Distance, BMVC, 2011, pp. 1–11.
[16] P. Sinha, B. Balas, Y. Ostrovsky, R. Russell, Face recognition by humans: nineteen results all computer vision researchers should know about, Proc. IEEE 94 (2006) 1948–1962.
[17] H. Huang, H. He, Super-resolution method for face recognition using nonlinear mappings on coherent features, Neural Netw. IEEE Trans. 22 (2011) 121–130.
[18] K. Jia, S. Gong, Multi-modal tensor face for simultaneous super-resolution and recognition, in: Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, IEEE, 2005, pp. 1683–1690.
[19] P.H. Hennings-Yeomans, S. Baker, B. Kumar, Simultaneous super-resolution and feature extraction for recognition of low-resolution faces, in: Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, IEEE, 2008, pp. 1–8.
[20] W.W. Zou, P.C. Yuen, Very low resolution face recognition problem, Image Process. IEEE Trans. 21 (2012) 327–340.
[21] H. Hotelling, Relations between two sets of variates, Biometrika. 28 (1936) 321–377.
[22] B. Li, H. Chang, S. Shan, X. Chen, Low-resolution face recognition via coupled locality preserving mappings, Signal Process. Lett. IEEE 17 (2010) 20–23.
[23] S. Siena, V.N. Boddeti, B.V. Kumar, Coupled marginal fisher analysis for low-resolution face recognition, in: Computer Vision–ECCV 2012. Workshops and Demonstrations, Springer, 2012, pp. 240–249.
[24] J. Zhang, Z. Guo, X. Li, Y. Chen, Large margin coupled mapping for low resolution face recognition, in: Pacific Rim International Conference on Artificial Intelligence, Springer, 2016, pp. 661–672.
[25] C. Zhou, Z. Zhang, D. Yi, Z. Lei, S.Z. Li, Low-resolution face recognition via simultaneous discriminant analysis, in: Biometrics (IJCB), 2011 International Joint Conference on, IEEE, 2011, pp. 1–6.
[26] P. Zhang, X. Ben, W. Jiang, R. Yan, Y. Zhang, Coupled marginal discriminant mappings for low-resolution face recognition, Opt. Int. J. Light Electron Opt. 126 (2015) 4352–4357.
[27] J. Shi, C. Qi, From local geometry to global structure: learning latent subspace for low-resolution face image recognition, IEEE Signal Process Lett. 22 (2015) 554–558.
[28] J. Jiang, R. Hu, Z. Wang, Z. Cai, CDMMA: coupled discriminant multi-manifold analysis for matching low-resolution face images, Signal Process. 124 (2016) 162–172.
[29] X. Xing, K. Wang, Couple manifold discriminant analysis with bipartite graph embedding for low-resolution face recognition, Signal Process. 125 (2016) 329–335.
[30] S. Biswas, K.W. Bowyer, P.J. Flynn, Multidimensional scaling for matching low-resolution face images, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 2019–2030.
[31] S. Biswas, G. Aggarwal, P.J. Flynn, K.W. Bowyer, Pose-robust recognition of low-resolution face images, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 3037–3049.
[32] C.-X. Ren, D.-Q. Dai, H. Yan, Coupled kernel embedding for low-resolution face image recognition, IEEE Trans. Image Process. 21 (2012) 3770–3783.

[33] X. Wang, L. Liu, H. Hu, Coupled Kernel Fisher discriminative analysis for low-resolution face recognition, in: Biometric Recognition, Springer, 2013, pp. 81–88.
[34] C. Ren, D. Dai, K. Huang, Z. Lai, Transfer learning of structured representation for face recognition, IEEE Trans. Image Process. 23 (2014) 5440–5454.
[35] M. Kan, J. Wu, S. Shan, X. Chen, Domain adaptation for face recognition: targetize source domain bridged by common subspace, Int. J. Comput. Vision 109 (2014) 94–109.
[36] Y. Pang, S. Wang, Y. Yuan, Learning regularized LDA by clustering, IEEE Trans. Neural Netw. Learn. Syst. 25 (2014) 2191–2201.
[37] D. Zhou, O. Bousquet, T.N. Lal, J. Weston, B. Schölkopf, Learning with local and global consistency, Adv. Neural Inf. Process. Syst. 16 (2004) 321–328.
[38] R.G. Soares, H. Chen, X. Yao, Semisupervised classification with cluster regularization, Neural Netw. Learn. Syst. IEEE Trans. 23 (2012) 1779.
[39] X. Wang, Gu Yuantao, Out-of-label Suppression Dictionary Learning with Cluster Regularization, IEEE SigPort, 2016.
[40] Y.-W. Pang, J. Pan, Z.-K. Liu, Cluster-based LDA for single sample problem in face recognition, in: 2005 International Conference on Machine Learning and Cybernetics, IEEE, 2005, pp. 4583–4587.
[41] P.J. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, Pattern Anal. Mach. Intell. IEEE Trans. 22 (2000) 1090–1104.
[42] G.B. Huang, M. Mattar, T. Berg, E. Learned-Miller, Labeled faces in the wild: a database forstudying face recognition in unconstrained environments, Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition, 2008.
[43] M. Grgic, K. Delac, S. Grgic, SCface–surveillance cameras face database, Multimedia Tools Appl. 51 (2011) 863–879.
[44] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1991) 71–86.