A Single Gallery-based Face Recognition using Extended Joint Sparse Representation

Kun Shang[a], Zheng-Hai Huang[b], Wanquan Liu[c], Zhi-Ming Li[d,*]

[a] *Center for Applied Mathematics of Tianjin University, Tianjin, PR China*
[b] *School of Mathematics, Tianjin University, Tianjin, PR China*
[c] *Department of Computing, Curtin University, Perth, WA, Australia.*
[d] *Faculty of Applied Math, Shanxi University of Finance and Economics, Taiyuan, PR China*

Abstract

For many practical face recognition problems, such as law enforcement, e-passport, ID card identification, and video surveillance, there is usually only a single sample per person enrolled for training, meanwhile the probe samples can usually be captured on the spot, it is possible to collect multiple face images per person. This is a new face recognition problem with many challenges, and we name it as the single-image-to-image-set face recognition problem (ISFR). In this paper, a customized dictionary-based face recognition approach is proposed to solve this problem using the extended joint sparse representation. We first learn a customized variation dictionary from the on-location probing face images, and then propose the extended joint sparse representation, which utilizes the information of both the customized dictionary and the gallery samples, to classify the probe samples. Finally we compare the proposed method with the related methods on several popular face databases, including Yale, AR, CMU-PIE, Georgia, Multi-PIE and LFW databases. The experimental results show that the proposed method outperforms most of these popular face recognition methods for the ISFR problem.

Keywords: Face recognition, single image to image set, customized

*Corresponding author.
Email addresses: skun@tju.edu.cn (Kun Shang), huangzhenghai@tju.edu.cn (Zheng-Hai Huang), w.liu@curtin.edu.au (Wanquan Liu), li_zm@sxufe.edu.cn (Zhi-Ming Li)

dictionary, dictionary learning, extended joint sparse representation.

1. Introduction

Face recognition (FR) is an active research topic in computer vision and pattern recognition [1, 2, 3, 4, 5, 6]. Various demands of applications, such as law enforcement, e-passport, ID card identification, video surveillance, access control, social network, photo management, criminal investigation, etc., lead to a wide range of solutions for FR. Over the past decades, many appearance-based methods were proposed to improve the performance of face recognition. With the increasing attention from researchers, many methods have been proposed in the literature, such as principle component analysis (PCA)[7, 8], linear discriminant analysis (LDA) [9], independent component analysis (ICA) [10], sparse representation classification (SRC) [11], kernel sparse representation (K-SR) [12], linear regression (LR) [13], collaborative representation classification (CRC) [14], locality-constrained collaborative representation (LCCR) [15], manifold constraints transfer (MCT) [16] and so on. All these methods are in one framework that many face samples per person are used for training and a face sample is used for testing. These approaches can achieve state-of-the-art results when the training samples are as large as possible, especially with deep learning technique. We call this category as the image-set-to-image face recognition (SIFR).

With the rapid development of digital imaging and communication technologies, the image-set-to-image-set face recognition (SSFR) becomes a very important research topic for video surveillance and has attracted much intention in research community. Recently, a number of approaches [17, 18, 19, 20, 4, 21, 22, 23] were proposed to solve the SSFR problem. Different from conventional SIFR where the probe is single, SSFR assumes that the gallery set and the probe set both have multi samples. All the samples are captured with different poses, illuminations and expressions. These face nuisances will affect the classification in the SSFR problem. Therefore, the key issues in SSFR include how to model

a set and compute the distance/similarity between probe and gallery sets effectively. Researchers have proposed subspace [24, 25, 26], manifold [17, 19, 22], affine or convex hull [20, 4, 21] and dictionary learning [27, 28] with attempt to achieve a satisfactory solution.

Unfortunately, sometimes there is only a single sample per person (SSPP) for training due to difficulty of collecting the sample with ID information. And in this case, many existing face recognition methods (both SIFR and SSFR methods) may fail to work because there are not sufficient samples for training. For the conventional SSPP face recognition [29, 30], there is one sample per person for training and one sample per person for probing. We call this category as single-image-to-single-image face recognition (IIFR). This IIFR problem has attracted much attention in computer vision community due to its difficulty and several kinds of efficient methods [31, 32, 33, 34, 35, 36, 37, 38] were proposed in the past. These methods are based on the generic learning, which assumes that the generic training set and the gallery set share similar variation information of both inter-class and intra-class. Especially, the dictionary learning methods, such as extended sparse representation-based classification (ESRC) [33, 34], sparse variation dictionary learning (SVDL) [35], sparse illumination learning and transfer (SILT) [36], variational feature representation-based classification (VFRC)[37], would learn a dictionary from an additional generic set to offer the extra information, including illumination, expression, occlusion, and pose.

What may be less obvious is that, in the real world, the probe samples usually can be captured easily on the spot, and it is possible to collect multiple face images per person. The IIFR methods ignore the collection of the multiple probe samples, which would have potentially useful information to improve the performance of FR. This is a special SSPP face recognition problem, named as the single-image-to-image-set face recognition (ISFR). In this case, there are multiple probe samples per person in the testing phase, and only one gallery sample per person in the training phase. This framework is new and more suitable for many practical applications. In the case of ISFR, it arises an essential question for this application scenario: how can we use the multiple testing face images

to improve the performance in ISFR? In 2013, Lu et. al. [39] proposed the locality repulsion projections and sparse reconstruction-based similarity measure (LRP-SRSM) method to solve it. From the metric learning perspective, Zhu et. al. [5] proposed the point-to-set distance metric learning method to learn a proper metric between a single image and image set in Euclidean space with an aim to achieve more accurate classification. From the manifold learning perspective, Huang et. al. [40] proposed the learning Euclidean-to-Riemannian metric method. They think the single image is a point lying in Euclidean space, and the image set reside on certain Riemannian manifolds, and build a bridge between them. But all the above methods ignore the specific contents (the various uncontrolled variations, such as pose, illumination and expression) of the face images on the shot. They do not make full use of the information in the observation data.

Inspired by the works of the sparse representation methods and the dictionary learning methods, we propose a new method, named customized dictionary-based face recognition with extended joint sparse representation (CD-EJSR), to solve the ISFR problem. First, each customized dictionary is obtained on the shot by using the samples corresponding the same probe subject. In other words, every probe subject would have a special dictionary. The learned dictionary contains the variation features about the uncontrolled variations (pose, illumination and expression). Then, we propose the extended joint sparse representation (EJSR), which utilizes the information of both the customized dictionary and the gallery samples to classify the probe samples. In summary, we can highlight the contributions of this work as follows:

- Different from the conventional dictionary learning methods [33, 35, 36] for the SSPP problem (e.g. IIFR), in which the dictionaries are learned from the gallery samples and generic samples, the variation dictionary in CD-EJSR is learned directly from the observed probe samples without identity information.

- The variation dictionary is learned by using a new optimization model,

which can be solved by the alternating direction method of multipliers (ADMM) approach. Also the closed-form solution is obtained in each step, which makes the proposed algorithm converge fast.

- We propose the extended joint sparse representation (EJSR) model. The EJSR model not only takes advantage of the learned variation dictionary which represents the intra-class variation between the gallery and probe samples, but also utilizes the group structure to enhance the performance for recognition.

The rest of this paper is organized as follows. Section 2 discusses the proposed customized dictionary learning and the extended joint sparse representation in detail. In Section 3, the experiments on several face databases are presented. The final section gives our conclusions for this paper.

2. Models

For the ISFR problem, there is only one single image for training and multiple samples for testing per person. The great difference (intra-class variation) between one training image and variational testing images becomes a huge barrier to recognize the identity of testing set. In order to reduce the barrier, we firstly propose a novel dictionary learning model to represent the intra-class variations of each probe subject. In fact, the dictionary just relies on the observation data (the probe face images of a subject on the shot), and it doesn't need to learn the additional auxiliary information from another face set. Furthermore, we do not need the ID information in this stage. Secondly, we propose a model with $l_{2,1}$-norm for fitting the probe samples and hope the group structure is still retained in the learned dictionary to enhance the performance for recognition. The overview of our approach is shown in Fig 1.

Next we will explain the framework in detail. For such purpose, we first list the notations.

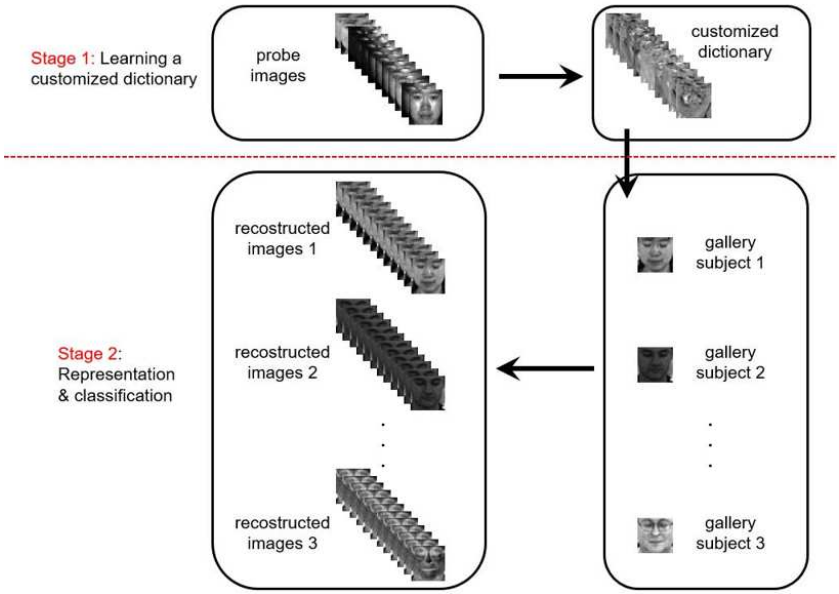


Fig. 1. The basic idea of our proposed ISFR approach. In the first stage, the variation dictionaries are learned, by using the proposed dictionary learning model, to encode the pose, illumination, expression and occlusion information in the probe images. Then in the second stage, this customized dictionary is used in the EJSR model to supply the intra-class variation. The reconstructed images are represented by the variation dictionary and the gallery samples. Finally, we recognize its label by using the smallest reconstruction error.

2.1. Notations

Let $A = [a_1, a_2, \cdots, a_L]$ be a gallery set, where a_l $(l = 1, 2 \cdots, L)$ is a single gallery sample of the l-th person. Assume that $\{Y_1, \cdots, Y_L\}$ denotes the collected probe face data of L different persons, where $Y_l = [y_{l1}, \cdots, y_{ln_l}] \in \mathbb{R}^{d \times n_l}$ is the face data including n_l images of the l-th person, $y_{li} \in \mathbb{R}^d$ $(i = 1, 2, \cdots, n_l)$ denotes a probe face image, d is the dimension of an image, and $n = \sum_{l=1}^{L} n_l$. The set $\{D_1, \cdots, D_L\}$ is a variation dictionary set learned from L probe persons, where $D_l \in \mathbb{R}^{d \times m_l}$ is a dictionary corresponding to the l-th probe person. Vectorization of all the samples is used in this paper.

2.2. The Customized Dictionary Learning Model

Based on the idea of linear representation, a face image I can be represented as three parts,

$$I = N + V + \varepsilon$$

where N denotes the normal face feature, V denotes the variation feature on the face, and ε denotes the error term. Fig. 2 illustrates the above representation. For the ISFR problem, the probe face data Y_l of the l-th person is captured

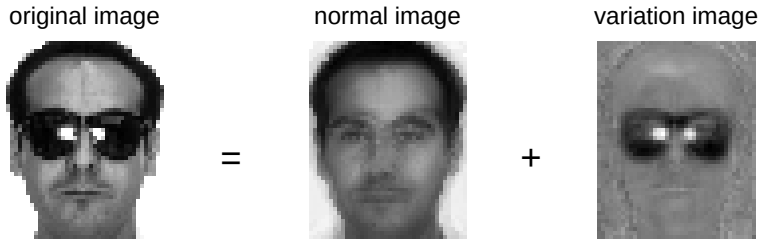


Fig. 2. The linear representation of a face image in the AR database, The original image contains the variation of occlusion. The key idea of linear representation is that this original image is the combination of the normal image and variation image.

on the spot. It is well known that the appearance of the captured face images are affected by many face images nuisances, including illumination, pose, and facial corruption/disguise (such as makeup, beard and glasses). Those noisy appearance would affect the recognition performance in testing phase. Inspired by the work [11], any face image can be linearly represented by the other face images in the same subspace (the same ID). Intuitively speaking, we can obtain an assumption that the variations on the face are located in a common variation subspace, and a face image in fact belongs to the sum of face subspace and the variation subspace. Based on the above analysis, there is a similar linear representation for the probe set Y_l of l-th person in the following:

$$Y_l = D_l X_l + N_l \otimes \mathbf{1} + E_l,$$

where X_l is the coefficient matrix corresponding the variation dictionary D_l, E_l is the error matrix, N_l denotes the normal face data (an ideal face or the average face of the probe face images) of the l-th person, $\otimes$ is kronecker product and $\mathbf{1}$ is full-one row vector. For brevity, $N_l \otimes \mathbf{1}$ denotes still by N_l in the rest of this article. A key problem is how to obtain a suitable variation dictionary D_l to remove the face images nuisances. To solve the problem, we design a new model to learn a customized dictionary from the given probe samples instead of generic face set. The model is written as the following optimization problem:

$$\begin{aligned} \min_{D_l, X_l, E_l} \quad & \tfrac{\lambda}{2}\|X_l\|_F^2 + \|E_l\|_F^2 \\ \text{s.t.} \quad & Y_l = D_l X_l + N_l + E_l, \end{aligned} \tag{1}$$

where λ is a regularization parameter. We omit the subscripts such as l in the above model for notational simplicity, then the simplified version is as follows

$$\begin{aligned} \min_{D, X, E} \quad & \tfrac{\lambda}{2}\|X\|_F^2 + \|E\|_F^2 \\ \text{s.t.} \quad & Y = DX + N + E. \end{aligned} \tag{2}$$

Note that this problem is a non-convex optimization problem. However, when E is fixed, this problem is convex for X, and there exists a global minimum. The situation is similar for E when X is fixed. Therefore, we can utilize the alternating direction method of multipliers (ADMM) framework [41] to solve (2). For variable separation, we introduce an additional variable C, then the problem (2) can be expressed as follows

$$\begin{aligned} \min_{D, C, X, E} \quad & \tfrac{\lambda}{2}\|C\|_F^2 + \|E\|_F^2 \\ \text{s.t.} \quad & Y - DX - N - E = 0,\ C - X = 0. \end{aligned}$$

In this case, we can consider the augmented Lagrangian

$$\begin{aligned} \mathcal{L} = & \tfrac{\lambda}{2}\|C\|_F^2 + \|E\|_F^2 \\ & + \mathrm{tr}[\Lambda_1^\top (C - X)] + \tfrac{\mu_1}{2}\|C - X\|_F^2 \\ & + \mathrm{tr}[\Lambda_2^\top (Y - DX - N - E)] + \tfrac{\mu_2}{2}\|Y - DX - N - E\|_F^2, \end{aligned}$$

where $\mathrm{tr}[\cdot]$ denotes the trace of a matrix. The updates for the variables can be easily derived under the ADMM framework. The complete process consists of the following five steps:

step 1. For variable C, we have the updated form

$$C^{k+1}=\arg\min_{C}\frac{\lambda}{2}\|C\|_F^2+\mathrm{tr}\left[\Lambda_1^{k\top}(C-X^k)\right]+\frac{\mu_1}{2}\left\|C-X^k\right\|_F^2.$$

We consider the partial derivative of the object function of the above problem, and can obtain the following,

$$\lambda C+\Lambda_1^k+\mu_1(C-X^k)=0,$$

which gives the explicit iterative formula as

$$C^{k+1}=(\mu_1X^k-\Lambda_1^k)/(\lambda+\mu_1). \tag{3}$$

step 2. For variable D, consider the following optimization problem,

$$D^{k+1}=\arg\min_{D}\mathrm{tr}\left[\Lambda_2^{k\top}(P^k-DX^k)\right]+\frac{\mu_2}{2}\left\|P^k-DX^k\right\|_F^2.$$

where $P^k=Y-N-E^k$. Similarly, by the first-order necessary conditions for unconstrained optimization problem, we have

$$-\Lambda_2^kX^{k\top}-\mu_2\left(P^k-DX^k\right)X^{k\top}=0.$$

Then, the iterative formula of the variable D is given by

$$D^{k+1}=\left[\frac{1}{\mu_2}\Lambda_2^k+P^k\right]X^{k\top}\left(X^kX^{k\top}\right)^{-1}. \tag{4}$$

step 3. For variable X, we consider the following optimization problem

$$X^{k+1}=\arg\min_{X}\begin{Bmatrix}\mathrm{tr}\left[\Lambda_1^{k\top}(C^{k+1}-X)\right]+\frac{\mu_1}{2}\left\|C^{k+1}-X\right\|_F^2+\\ \mathrm{tr}\left[\Lambda_2^{k\top}(P^k-D^{k+1}X)\right]+\frac{\mu_2}{2}\left\|P^k-D^{k+1}X\right\|_F^2\end{Bmatrix},$$

where $P^k=Y-N-E^k$. We use the first-order necessary conditions for unconstrained optimization problem to obtain the following equation

$$-\Lambda_1^k-\mu_1(C^{k+1}-X)-D^{k+1\top}\Lambda_2^k-\mu_2D^{k+1\top}(P^k-D^{k+1}X)=0.$$

Furthermore, the update formula is derived as

$$X^{k+1}=\left(\mu_1I+\mu_2D^{k+1\top}D^{k+1}\right)^{-1}\left(\Lambda_1^k+D^{k+1\top}\Lambda_2^k+\mu_1C^{k+1}+\mu_2D^{k+1\top}P^k\right), \tag{5}$$

where I denotes an identity matrix.

step 4. For the final variable E, we give the following iterative formula

$$E^{k+1}=\arg\min_{E}\|E\|_F^2+\mathrm{tr}\left[\Lambda_2^{k\top}(Q^{k+1}-E)\right]+\frac{\mu_2}{2}\left\|Q^{k+1}-E\right\|_F^2,$$

where $Q^{k+1}=Y-D^{k+1}X^{k+1}-N$. By solving the following equation

$$2E-\Lambda_2^k-\mu_2\left(Q^{k+1}-E\right)=0,$$

we can obtain the following closed-form iterative formula for E:

$$E^{k+1}=\left(\Lambda_2^k+\mu_2 Q^{k+1}\right)/(2+\mu_2). \tag{6}$$

step 5. The updates for dual variables are as follows:

$$\begin{aligned}\Lambda_1^{k+1}&=\Lambda_1^k+\mu_1\left(C^{k+1}-X^{k+1}\right),\\ \Lambda_2^{k+1}&=\Lambda_2^k+\mu_2\left(Y-D^{k+1}X^{k+1}-N-E^{k+1}\right).\end{aligned} \tag{7}$$

The above process is a Gauss-Seidel based ADMM Algorithm. We summarize it in the following Algorithm 1.

Algorithm 1 Customized dictionary learning

Input: Probe face data $Y\in\mathbb{R}^{d\times n}$, penalty parameters μ_1 and μ_2, regularization parameter λ, dictionary scale m;

Initialization: $X=[1/m]_{i,j}\in\mathbb{R}^{m\times n}$, $E=O\in\mathbb{R}^{d\times n}$, $\Lambda_1=O\in\mathbb{R}^{m\times n}$, $\Lambda_2=O\in\mathbb{R}^{d\times n}$;

for k = 0,1,2, ...

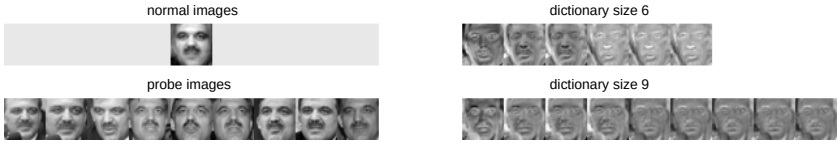
1. update C by (3);
2. update D by (4);
3. update X by (5);
4. update E by (6);
5. update the dual variables Λ_1 and Λ_2 by (7);

end for

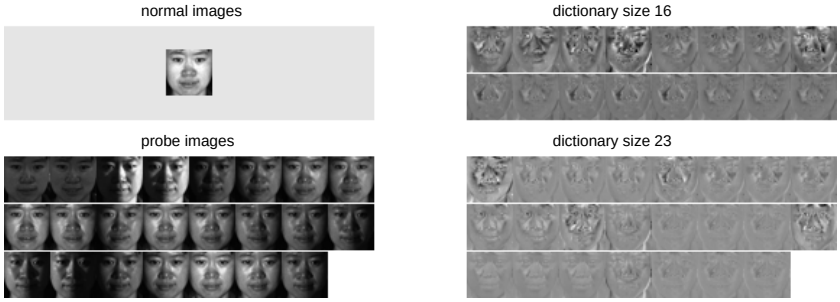
Output: the dictionary D.

In Algorithm 1, we have the closed-form update formula in each iteration and this will make the process of learning the customized variation dictionary

be rapidly implemented online. The computation time for dictionary learning will be discussed in Subsection 3.4. Fig. 3 shows that the original images and the learned variation dictionaries of the first probe subject on LFW and CMU-PIE databases. It is shown that the dictionaries contain the intra-class variations, and two different scale dictionaries have the obvious differences. We will show how the different dictionaries have impact on the performance for FR in Subsection 3.5.



(a) the LFW database



(b) the CMU-PIE database

Fig. 3. Examples of the customized variation dictionaries of two subjects on LFW and CMU-PIE databases. The left images in each sub-figure contain a normal image and a group of probe images from the same subject. The right images in each sub-figure are the customized dictionaries which are learned from the left probe images. These customized dictionaries have the different sizes.

2.3. The Extended Joint Sparse Representation Model

In the classical SRC [11], determining the identity of L probe samples needs to solve L sparse representation problems. In this paper, we have access to multiple samples of the same subject in the probing phase, and our aim is to

identifying a group of image samples in the testing stage. As an extension of SRC under the multi-task situation, the joint sparse representation classification (JSRC) [42] exploits the shared information from all the samples to make a joint decision for recognition task and it can be used in this paper. The JSRC model is in the following form

$$\min_{S} \|Y - AS\|_F^2 + \mu\|S\|_{2,1}, \tag{8}$$

where $Y \in \mathbb{R}^{d\times n}$ is a group of the samples from a probe subject, $A \in \mathbb{R}^{d\times L}$ denotes the gallery set, each column of A denotes a normal face vector (in general, an ideal face), and $\|\cdot\|_{2,1}$ is defined by the sum of the l_2-norm of all rows of a matrix. For the SSPP problem, there is still a gap between a normal gallery sample and any probe samples for the same person. Inspired by the ESRC model [33], we insert the customized variation dictionary into the JSRC model, and then obtain a novel model, named the extended joint sparse representation (EJSR) model, which is rewritten as follows:

$$\min_{X,S} \|Y - DX - AS\|_F^2 + \tau\|X\|_{2,1} + \mu\|S\|_{2,1}$$

or

$$\min_{X,S} \left\| Y - [D\ A] \begin{bmatrix} X \\ S \end{bmatrix} \right\|_F^2 + \left\| \begin{bmatrix} \tau X \\ \mu S \end{bmatrix} \right\|_{2,1}, \tag{9}$$

where X and S denote the representation coefficient on D and A, respectively; The customized variation dictionary D obtained in (2) represents the intra-class variation between the gallery and probe samples. It joints with the gallery sample to confront the complex variation on the face in the probe phase. For integration of the variables, we denote $\bar{D} = \frac{\gamma}{\tau}D$, $\bar{A} = \frac{\gamma}{\mu}A$, $\bar{X} = \frac{\tau}{\gamma}X$, and $\bar{S} = \frac{\mu}{\gamma}S$, where $\gamma = \tau + \mu$, and express the model (9) as

$$\min_{\bar{X},\bar{S}} \left\| Y - [\bar{D}\ \bar{A}] \begin{bmatrix} \bar{X} \\ \bar{S} \end{bmatrix} \right\|_F^2 + \gamma \left\| \begin{bmatrix} \bar{X} \\ \bar{S} \end{bmatrix} \right\|_{2,1}.$$

Again, let $B = \left[\bar{D}\ \bar{A}\right]$ and $H = \left[\bar{X}^\top\ \bar{S}^\top\right]^\top$, we obtain a simplified model

$$\min_{H} \|Y - BH\|_F^2 + \gamma\|H\|_{2,1}. \tag{10}$$

So, by a simple transform, one can see that the EJSR model would degenerate into the JSRC model. Next, we would like to use some existing optimization methods to solve this EJSR problem. In this paper, we use the ADMM framework to solve the problem. For details, we introduce the additional variable V and T, and the model can be transformed into

$$\begin{aligned} \min_{H,V,T} \quad & \|V\|_F^2 + \gamma\|T\|_{2,1} \\ \text{s.t.} \quad & BH - Y - V = 0,\ H - T = 0. \end{aligned}$$

The augmented Lagrangian function is as follows:

$$\begin{aligned} \mathcal{L} = & \|V\|_F^2 + \gamma\|T\|_{2,1} \\ & + \text{tr}\left[W_1^\top (BH - Y - V)\right] + \tfrac{\eta_1}{2}\|BH - Y - V\|_F^2 \\ & + \text{tr}\left[W_2^\top (H - T)\right] + \tfrac{\eta_2}{2}\|H - T\|_F^2. \end{aligned}$$

We now can scale dual variables $U_i = W_i/\eta_i$, $i = 1, 2$, and obtain an explicit form

$$\begin{aligned} \mathcal{L} = & \|V\|_F^2 + \gamma\|T\|_{2,1} + \tfrac{\eta_1}{2}\|U_1\|_F^2 \\ & + \eta_1 \text{tr}\left[U_1^\top (BH - Y - V)\right] + \tfrac{\eta_1}{2}\|BH - Y - V\|_F^2 \\ & + \tfrac{\eta_2}{2}\|U_2\|_F^2 + \eta_2 \text{tr}\left[U_2^\top (H - T)\right] + \tfrac{\eta_2}{2}\|H - T\|_F^2 \\ & - \tfrac{\eta_1}{2}\|U_1\|_F^2 - \tfrac{\eta_2}{2}\|U_2\|_F^2 \\ = & \|V\|_F^2 + \gamma\|T\|_{2,1} \\ & + \tfrac{\eta_1}{2}\|BH - Y - V + U_1\|_F^2 \\ & + \tfrac{\eta_2}{2}\|H - T + U_2\|_F^2 + \text{const}, \end{aligned}$$

where the constant is independent of the primal variables H, V, T. We perform the following steps to obtain the solution of the model (10).

step 1. For the variables H and V, the iterative schemes are obtained by solving the following two problems as

$$\begin{aligned} H^{k+1} &= \arg\min_H \tfrac{\eta_1}{2}\left\|BH - Y - V^k + U_1^k\right\|_F^2 + \tfrac{\eta_2}{2}\left\|H - T^k + U_2^k\right\|_F^2, \\ V^{k+1} &= \arg\min_V \|V\|_F^2 + \tfrac{\eta_1}{2}\left\|BH^k - Y - V + U_1^k\right\|_F^2. \end{aligned}$$

By the first-order necessary conditions for unconstrained optimization problem,

we can easily get these two updates of H and V

$$\begin{aligned} H^{k+1} &= \left(\eta_1 B^\top B + \eta_2 I\right)^{-1} \left[\eta_1 B^\top (Y + V^k - U_1^k) + \eta_2 (T^k - U_2^k)\right], \\ V^{k+1} &= \tfrac{\eta_1}{2+\eta_1} \left(BH^k - Y + U_1^k\right), \end{aligned} \tag{11}$$

where I denotes an identity matrix.

step 2. For the last primal variable T, the update steps are only slightly different. Firstly, we have

$$T^{k+1} = \arg\min_T \gamma \|T\|_{2,1} + \frac{\eta_2}{2} \left\|H^k - T + U_2^k\right\|_F^2. \tag{12}$$

Let $H^k + U_2^k = Z^k$, and t_i and z_i^k denote the i-th row vectors of T and Z^k, respectively. Then we decompose the problem (12) as

$$T^{k+1} = \arg \min_{t_i, i=1,2,\ldots} \sum_i \left[\gamma \|t_i\|_2 + \frac{\eta_2}{2} \left\|z_i^k - t_i\right\|_2^2\right].$$

Thus, we can find each row of T separately by exploiting the following result. The optimal solution of the problem

$$\min_t \gamma \|t\|_2 + \frac{\eta_2}{2} \|z - t\|_2^2$$

is $t = \kappa z$, where $\kappa = \max\{1 - \frac{\gamma}{\eta_2}\|z\|_2, 0\}$ if $\|z\|_2 > 0$, and $\kappa = 0$ if $\|z\|_2 = 0$. Hence, the update formula of T is given by

$$T^{k+1} = \left[{t_1^{k+1}}^\top, {t_2^{k+1}}^\top, \cdots, {t_{m+L}^{k+1}}^\top\right]^\top, \tag{13}$$

where

$$t_i^{k+1} = \begin{cases} \max\{1 - \frac{\gamma}{\eta_2}\|z_i^k\|_2, 0\} \cdot z_i^k, & \text{if } \|z_i^k\|_2 > 0, \\ \mathbf{0}, & \text{if } \|z_i^k\|_2 = 0. \end{cases}$$

step 3. The updates for dual variables are

$$\begin{aligned} U_1^{k+1} &= U_1^k + BH^k - Y - V^k, \\ U_2^{k+1} &= U_2^k + H^k - T^k. \end{aligned} \tag{14}$$

Through the above updates, H can be found. Accordingly, the solution of model (9) can be achieved from the following formulas.

$$\begin{aligned} X &= \tfrac{\gamma}{\tau}[I_{m\times m} \ O_{m\times n}] \cdot H, \\ S &= \tfrac{\gamma}{\mu}[O_{n\times m} \ I_{n\times n}] \cdot H. \end{aligned} \tag{15}$$

The method described above is summarized the following Algorithm 2.

Algorithm 2 Representation

Input: Probe face data $Y \in \mathbb{R}^{d\times n}$, variation dictionary $D \in \mathbb{R}^{d\times m}$, gallery set $A \in \mathbb{R}^{d\times L}$, regularization parameters τ and μ, penalty parameters η_1 and η_2;

Initialization: $V = O \in \mathbb{R}^{d\times n}$, $T = O \in \mathbb{R}^{(m+L)\times n}$, $U_1 = O \in \mathbb{R}^{d\times n}$, $U_2 = O \in \mathbb{R}^{(m+L))\times n}$, $\gamma = \tau + \mu$, $B = \left[\frac{\gamma}{\tau}D \ \frac{\gamma}{\mu}A\right]$;

for k = 0,1,2, ...

1. update H, V by (11);
2. update T by (13);
3. update the dual variables U_1 and U_2 by (14);

end for

Output: X and S by (15).

2.4. Classification

Given the probe face data Y of the same subject, we can indirectly get the corresponding variation dictionary D^* from the model (2), and obtain the optimal solution X^*, S^* of the model (9). Then, the identity of the probe subject is obtained via

$$\text{identity}(Y) = \arg\min_i \|Y - D^*X^* - a_i s_i^*\|_F^2,$$

where a_i is the i-th column of the gallery set A, and s_i^* is the i-th row of the coefficient matrix S^*.

3. Experiments

In this section, we evaluate the effectiveness of our method on different datasets, including AR [43], Yale [9], CMU-PIE [44], Georgia [45], Multi-PIE[46] and LFW [47]. Some images in these datasets are shown in Fig. 4. Our experiments are conducted on a PC platform with 64-bit win 8 operating system with Intel Core i5-3550S CPU, and 8 G memory.

3.1. Experiment Setup

As we focus on ISFR in this paper, we will establish a standard for fair comparison with other approaches. For such purpose, all images are cropped into the

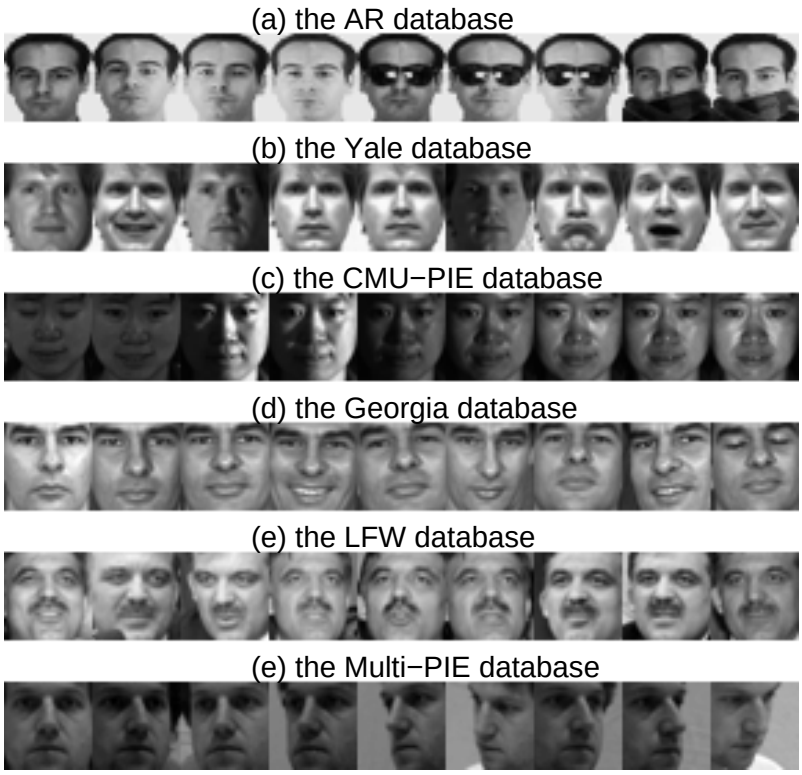


Fig. 4. Face image samples in the AR, Yale, CMU-PIE, Georgia, LFW and Multi-PIE databases.

size 32×32. On each database, the first image per person is used as the gallery sample, and the rest images are used as the probe samples in our experiments, except for Multi-PIE. We will detail the experiment setup on the Multi-PIE database separately in subsection 3.7. In the subsequent experiments, we first select suitable parameters of the proposed model, and investigate the convergence, computation time of our algorithm, and the influence of the dictionary size. After that, we will compare the proposed method with several related sparse coding based methods, including JSRC [42], MNSRC [26], LRP_SRSM [39], SRC [11], LCCR [15], and related dictionary learning methods, including ESRC [33, 34], SVDL [35], SILT [36], VFRC [37]. The l_1-regularized minimization in SRC, SVDL and LRP-SRSM is solved by l_1-l_s algorithm [48]; the

l_1-regularized minimization in ESRC and SILT is solved by Homotopy algorithm [49]; the ADMM framework is used to solve JSRC, MNSRC and our method. The final recognition results for SRC, ESRC, SVDL, SILT and VFRC are obtained by the majority voting strategy, except for the single-image-to-image-set methods: LRP-SRSM, JSRC and MNSRC. In the following experiments, the parameters of these related methods are the same as those in their original papers.

3.2. Parameter Tuning

Cross validation is a popular method to select parameters. We used the five-fold cross validation to find the optimal combination schemes of relevant parameters for our method. In order to present the complete experimental process of parameters tuning, the CD-EJSR is implemented on LFW database. The parameters of the EJSR model is the same as those of the JSRC model. Here we only consider the customized dictionary learning model with the case ($\mu_1 = \mu_2 = \mu$). The mean recognition rates are recorded in Table 1. We can

Table 1. Parameter setting of the customized dictionary learning model

λ \ μ	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00	1.10	1.20
0.0500	47.47	44.94	45.57	45.57	47.47	46.84	45.57	45.57	43.67	48.10	44.94	48.10
0.0100	44.94	46.84	43.67	44.94	44.30	44.30	48.10	45.57	45.57	44.94	44.94	44.30
0.0050	46.84	44.94	46.20	44.30	48.10	45.57	46.20	44.94	47.47	46.20	44.30	43.67
0.0010	47.47	46.84	44.94	45.57	44.94	45.57	46.20	46.20	46.20	44.94	46.20	48.10
0.0005	44.30	42.41	47.47	**50.00**	44.94	46.84	46.20	44.30	46.84	46.20	43.67	44.30
0.0001	48.10	48.10	44.30	46.20	45.57	44.30	43.67	45.57	47.47	45.57	45.57	45.57

see from this table that the best choice is $\lambda = 0.0005$, $\mu = 0.4$. So we will use these parameters in the following experiments.

3.3. Illustration of Algorithm Convergence

In this subsection, we will illustrate the convergence of Gauss-Seidel based ADMM algorithm for the customized dictionary learning model (2). Two face databases (LFW and CMU-PIE) are used in these experiments. On each face

database, different sizes of dictionaries are considered, such as the sizes 5, 10, $\cdots$, 40. The convergence threshold ϵ is set as 0.001 for all the dictionary sizes. Fig. 5 shows the convergence of the Gauss-Seidel based ADMM algorithm. It

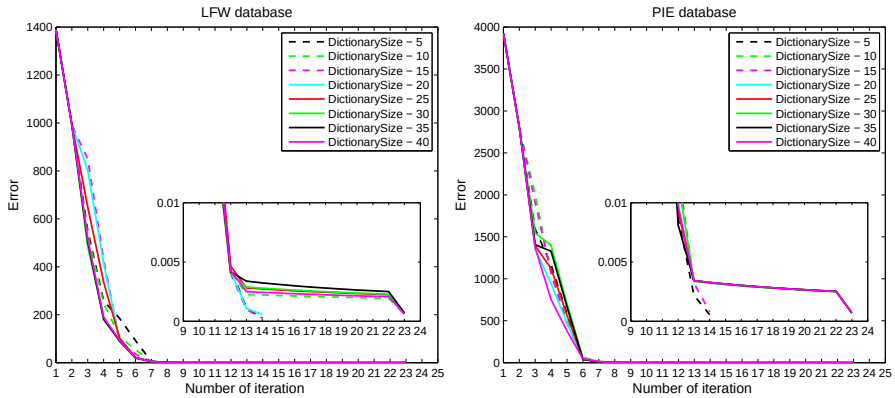


Fig. 5. Illustration of the convergence of the proposed Gauss-Seidel based ADMM algorithm for the customized dictionary learning model.

is clear to see from above figures that one can reach the convergence threshold after 23 iterations in all cases. When the dictionary sizes are 5, 15 and 20, the convergence threshold is reached after 14 iterations on the LFW database. When the dictionary sizes are 5 and 15, the convergence threshold can be reached after 14 iterations on the CMU-PIE database. These figures show that our algorithm always converges rapidly regardless of the dictionary size.

3.4. Computation Time

Based on the requirements that the process of learning variation dictionary needs to be implemented online, we design the model with Frobenius norm and employ the Gauss-Seidel based ADMM algorithm to solve the proposed model. Fortunately, the closed-form solutions have been derived in each iteration. Therefore, it is possible to rapidly learn variation dictionary online. In addition, the time of recognition process is also an important indicator for a face recognition system. In the sequent section, we would show the computation time of learning dictionary and recognition process with respect to the various

dictionary sizes on the databases (LFW, Georgia, CMU-PIE). These databases have different number of persons, and the number of images per person is also different. Each person has 10 images on the LFW database, 15 images on the Georgia database and 24 images on the CMU-PIE database. For a fair and thorough comparisons, we would record the time of dictionary learning and recognition per 100 persons. We implement the proposed method with respect to the various dictionary sizes on each database. The computation time of the dictionary learning and recognition processes are recorded on each database, respectively.

The final results are shown in Fig. 6. We can see that the recognition time

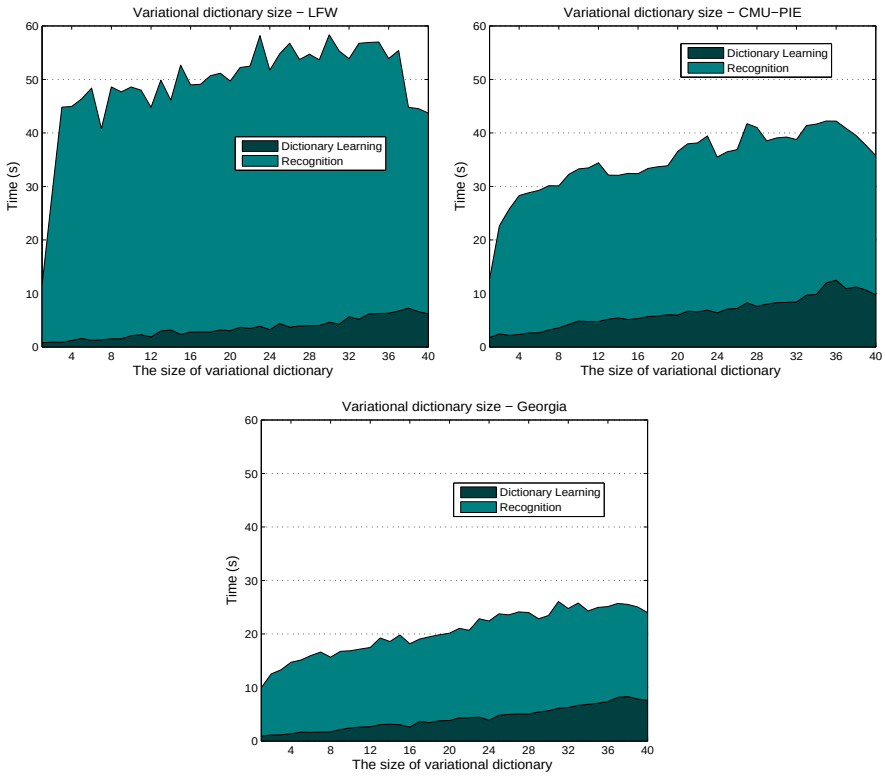


Fig. 6. Illustration of the time costs of the proposed dictionary learning and recognition processes with testing 100 persons on each face database.

is always longer than the time of learning dictionary regardless of the used

databases. With the increasing of dictionary size, the time of the dictionary learning gradually increases, and the recognition time first increases and then decreases on each database. Furthermore, the computation time in the dictionary learning phase is less than 10s for each database, which implies that the computation time of learning a dictionary from a probe subject is about 0.1s. Overall, it takes about 0.4s to identify each person, which is reasonable for practical applications.

3.5. Size of Variation Dictionary

In this subsection, we discuss how the dictionary size will affect the recognition rate. The experimental setup is the same as the those in the previous subsection. The results on the three databases (LFW, Georgia, CMU-PIE) are

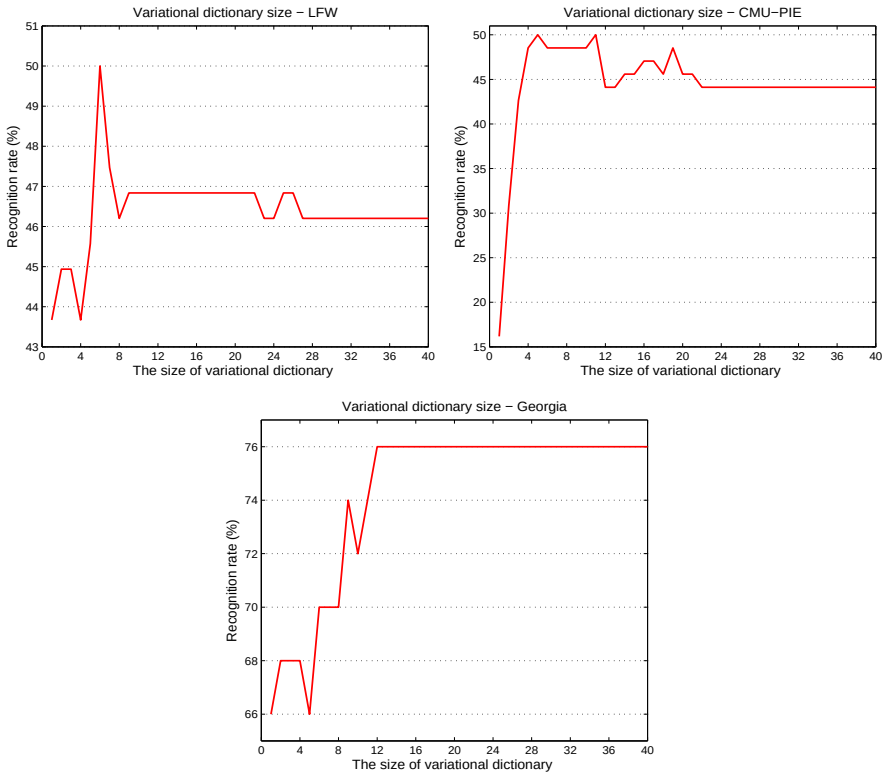


Fig. 7. Illustration how the dictionary size effects on the recognition rate.

presented in Fig. 7. We can see that the recognition rates increase rapidly when the dictionary size is small. Nevertheless, the recognition rates will not increase anymore when the dictionary size is large enough. As shown in Fig. 7, the recognition rates will be stable when the size of the dictionary size is larger than 8, 8 and 12 on LFW, CMU-PIE and Georgia databases, respectively. We will take the size around 8 in our experiments.

3.6. Evaluations on Different Databases

3.6.1. Yale Database

The Yale database has 165 images of 15 adults, 11 images per person. The face images have variations with respect to facial expressions (as normal, sad, happy, sleepy, surprised, and winking) and illuminations (where the position of the light source is at the center, left and right). We conduct two groups of tests on this database. In the first group, all the probe samples (10 samples per person) are used. In the second one, we choose randomly half part of the probe samples to evaluate the performance of the related methods, including JSRC, MNSRC, LRP_SRSM, SRC, LCCR and our proposed CD-EJSR method. The recognition rates of the second test are the average values of ten times results. The experimental results of the related methods are shown in Fig. 8. It is clear that our method outperforms all other related methods on this database. In particular, the proposed CD-EJSR method possesses the excellent performance in the case with 10 probe images per person. Our method obtains the best recognition rates 100%, while the recognition rate of the other methods are less than 90%. For the case with 5 probe images per person, the CD-EJSR is also better than other methods.

3.6.2. CMU-PIE database

In this subsection, we will evaluate the robustness of the proposed method for variational illumination on the CMU-PIE database. The database consists of 41368 images of 68 people. Each person has many images captured under 13 poses and 43 illumination conditions and with 4 expressions. We select 24

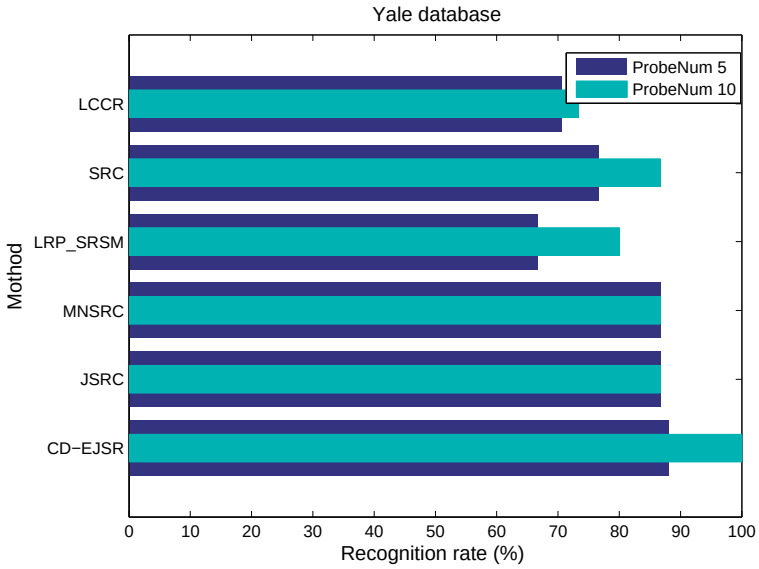


Fig. 8. The recognition rates on the Yale database

frontal images per person from the camera No. 05 for the experiments in this subsection. These images involve abundant illumination changes. The first image per person is selected for the gallery set, and the rest images consist of the probe set. We randomly choose k ($k = 5, 10, 15, 20, 23$) images per person as the probe samples. For each k, the experiments are repeated for 10 times to obtain the average results. As shown in Table 2, the proposed CD-EJSR method outperforms all other related methods regardless of the number of the probe images. In details, for each number of the probe samples, the results of

Table 2. The recognition rates (%) about the different numbers of the probe images on the CMU-PIE databases

Method	23	20	15	10	5
CD-EJSR	**50.00**	**48.97**	**45.74**	**34.56**	**24.26**
JSRC	17.65	17.06	17.94	17.94	17.06
MNSRC	17.65	17.06	17.94	17.94	17.06
LRP_SRSM	10.29	9.71	9.56	9.56	9.71
SRC	10.29	10.15	10.00	9.85	9.12
LCCR1	14.71	14.71	15.88	14.71	13.38

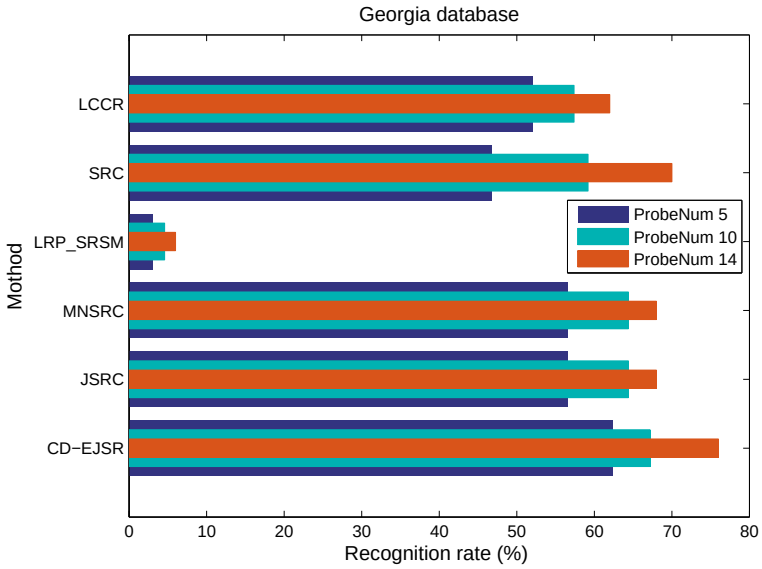


Fig. 9. The recognition rates on the Georgia database

CD-EJSR are higher than all other methods by $7\% - 40\%$. Particularly, when the number of the probe samples is 23, the advantage of the proposed method is the most outstanding. The experimental results on the CMU-PIE database show that our method is much more effective than other related methods on variational illumination conditions.

3.6.3. Georgia Database

The Georgia database contains 750 images of 50 people taken at the Center for Signal and Image Processing at Georgia Institute of Technology. Each people in the database has 15 color images with cluttered background taken in two or three sessions within half a year at resolution 640×480 pixels. The images of this database have abundant changes of illumination, expression and pose. Particularly, many of these pictures are rotated clockwise or anticlockwise. For each person, we consider three cases with 5, 10, 14 probe samples. For the first two cases, ten groups of samples are randomly selected to test from all probe samples per person. The experimental results are shown in Fig. 9. We can see that the proposed method obtains higher performance rates than all

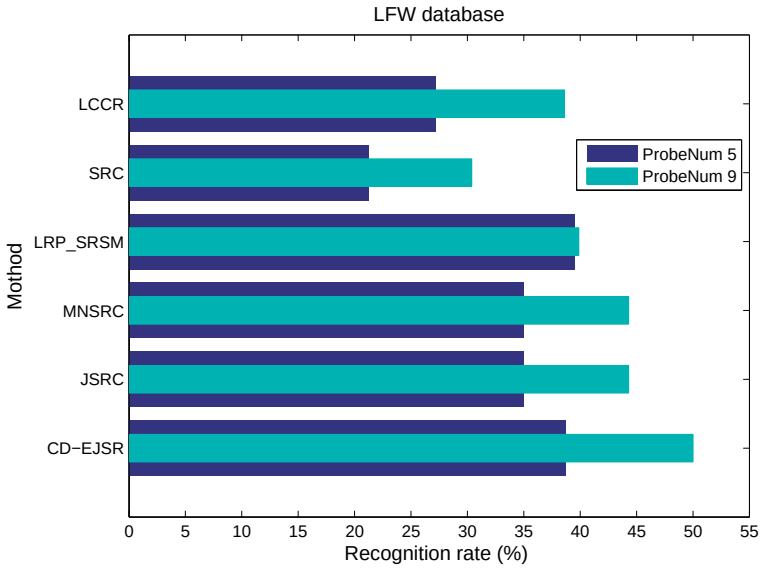


Fig. 10. The recognition rates on the LFW database

other methods for all cases. In particular, for the case with 14 probe samples per person, the CD-EJSR has the obvious advantage. In a word, the proposed method is also effective on the Georgia database.

3.6.4. LFW Database

The LFW database contains images of 5749 different individuals in unconstrained environments. We create a sub-database which includes the subjects no less than ten samples. It contains 158 individuals and each person has just 10 different images. The variances in illumination, pose, occlusion, and expression between these images make SSPP face recognition extremely challenging. We also make two groups of tests. The first one is that all the probe samples are used. The other is that five samples randomly selected per person are recognized. The results of 10 trials are recorded and the average values are calculated. Fig. 10 shows the compared results with related methods. For the first test with 9 probe samples per person, the proposed method obtains the highest recognition rate 50% among all compared methods. Our method gets better results than the related methods in the case with 5 probe samples per

person, except for the LRP_SRSM method with slightly difference. In addition, by observing the results of all compared methods, we have the following conclusion: the more the probe samples per person we have, the higher recognition rates we can obtain for the concerned methods.

3.6.5. AR Database

The AR database contains over 4000 color face images of 126 people (70 men and 56 women). All images are frontal views of faces with different facial expressions, lighting conditions and occlusions. In our experiments, we choose 120 individuals (65 men and 55 women) with 26 images and convert them into grayscale images for verifying the occlusions testing. We choose the first front image for training and the rest images for testing. We randomly choose k ($k = 20, 25$) images per person as the probe samples. The experimental results about different probe number are shown in Table 3.

Table 3. The recognition rates (%) about the different probe number on the AR database

Method	25		20	
	rate(%)	time(s)	rate(%)	time(s)
CD-EJSR	99.17	28.3	99.08	24.8
JSRC	98.33	5.4	97.75	4.6
MNSRC	98.33	22.1	97.75	19.2
LRP_SRSM	99.17	2000.1	99.08	1915.1
SRC	97.50	166.6	96.83	133.2
LCCR	98.33	8.4	98.17	6.8

From Table 3, all the selected methods achieve the state-of-the-art results for the experiments with occlusions. Nonetheless, our method achieves the best recognition rates as LRP_SRSM. From another point of view, LRP_SRSM takes lots of time to get the same best recognition rates. It takes 2000s to identify 25 samples per person, and 1915s to identify 20 samples per person. However, our method takes less than 30s to achieve the best recognition. It saves about 70 times to identify the probe samples than LRP_SRSM. Considering these two

factors, our method is the best of all approaches for the experiments on the AR database.

3.7. Multi-PIE Database

The final database used in our experiment is the Multi-PIE database. The Multi-PIE database contains 755370 images from 337 different subjects. These images were taken from four different seasons with 15 different poses under 20 illumination conditions. In order to evaluate the performance of our CD-EJSR method under the different illumination and the multi-view (or the different poses) conditions, we select two subsets of the Multi-PIE database, denoted by sub-M1 and sub-M2. Sub-M1 collects all frontal images (from the camera 05_1) with 20 illumination conditions from the first seasons. In the illumination experiment, we use the sub-M1 database to investigate the effect of different illumination conditions on the gallery samples. So there should be 20 groups of sub-experiment. For each group, all gallery samples are with the same illumination condition, and the the rest images are used for probe. The average value of 20 recognition rates is calculated. In the multi-view experiment, we use the Sub-M2 database which contains 9 images per person from the cameras (05_1, 05_0, 14_0, 04_1, 13_0, 19_0 and 08_0) which are located at head height and spaced in $\pm 15^\circ$ intervals. Simultaneously, two additional cameras (08_1, 19_1), which were located above the subject, simulating a typical surveillance camera view, are also selected. All the images are with the same illumination condition (without flash). For each subject, the image recorded by the camera 05_1 is used for training, and other rest images are used for testing. We also implement several related methods on the sub-M1 and sub-M2 databases. All the experiment results are listed in the Table 4.

From Table 4, we can see that the proposed method achieves better performance than all other compared methods in terms of recognition rate and cost time. On two sub-M1 and sub-M2 databases, the recognition rates of the CD-EJSR and LRP_SRSM methods are higher than those with other methods. Particularly, in the case with multi-view (using sub-M2), these two methods

Table 4. The recognition rates (%) about the different probe number on the Multi-PIE database

Method	sub-M1		sub-M2	
	rate(%)	time(s)	rate(%)	time(s)
CD-EJSR	99.96	733.7	73.29	284.1
JSRC	95.13	246.8	17.51	107.7
MNSRC	95.13	251.5	17.51	108.0
LRP_SRSM	99.88	23027.8	73.89	14561.8
SRC	52.92	2541.1	38.58	288.5
LCCR	75.37	28.5	57.57	12.5

have the obvious advantage. In details, the recognition rates of the CD-EJSR and LRP_SRSM methods are higher about 20% $\sim$ 50% than those with other methods. Both the CD-EJSR and LRP_SRSM method have the outstanding performance with respect to the recognition accuracy rate. However, the proposed CD-EJSR method takes much less time than the LRP_SRSM method. Accurately speaking, the latter spends 50 times time of the former. From the perspective of both recognition rate and time, our CD-EJSR method is the best one of these related methods on the multi-PIE database.

3.8. Comparisons with other Dictionary Learning Methods

In this subsection, we will discuss the performance of dictionary learning methods for the SSPP problem (ISFR is a special case of SSPP problem). For most of the dictionary learning methods, the training set needs to be as sufficient large as possible. However, the training set is the same as the gallery set for the SSPP problem. Generally, people will take a generic set to extract useful variation information. Some related methods about dictionary learning for SSPP problem, such as SILT [36], SVDL [35], ESRC [33] with the basic variation dictionary (BVD) and VFRC[37], are selected for comparison with our proposed CD-EJSR method on several databases (LFW, Georgia, AR, Yale and CMU-PIE databases). For each database, the first sample for each person is selected to be as gallery set, all the rest samples is selected to be the probe

set.

In order to demonstrate the effectiveness of each related approach, we provide two generic sets in the experiments of the subsection. The first one is LFWs which contains 250 persons with 3 to 9 face images per person from the LFW database. The second one is FRGCs which consists of 30 persons with 20 images per person from the whole FRGC database [2]. Note that the dictionary learning methods with SILT and SVDL demand the same number of images per person in the generic set. Therefore, the SILT and SVDL methods use only the generic set FRGCs. The experimental results are shown in Table 5.

Table 5. The recognition rates (%) about the different dictionary learning methods on several different databases

GenericSet	Method	LFW	Georgia	AR	Yale	PIE
LFWs	IS-ESRC	37.34	54.00	98.33	80.00	17.65
FRGCs	IS-ESRC	32.91	62.00	97.50	53.33	16.18
LFWs	IS-SILT	–	–	–	–	–
FRGCs	IS-SILT	36.71	58.00	97.50	80.00	13.24
LFWs	IS-SVDL	–	–	–	–	–
FRGCs	IS-SVDL	28.48	56.00	96.67	46.67	13.24
LFWs	BVD-JSRC	21.52	36.00	96.67	26.67	23.53
FRGCs	BVD-JSRC	19.62	46.00	96.67	40.00	20.59
LFWs	SILT-JSRC	–	–	–	–	–
FRGCs	SILT-JSRC	37.34	64.00	98.33	93.33	25.00
LFWs	BVD-VFRC	43.67	54.00	98.33	60.00	20.59
FRGCs	BVD-VFRC	30.38	56.00	96.67	60.00	16.18
	CD-EJSR	**50.00**	**74.00**	**99.17**	**100.0**	**48.53**

For each database, one can see that our method achieves the best results among these dictionary learning methods. One of the reasons is that all the conventional dictionary learning methods need a generic set to obtain a variation dictionary. Their performance of recognition is sensitive to the selections of the generic set. Another reason is that the facial information of the generic set will have residue errors in the dictionary, which may have negative influence for the recognition results. Usually, if the variation dictionary contains more related variation feature information, the performance will be more accurate.

As the customized dictionary of the proposed method is learned directly from the observation data (the probe samples), the inference information (such as ID information of the generic samples) from the generic set does not possess. In other words, the extra information from the generic set can affect the recognition in ESRC, SILT, SVDL and VFRC. However, our CD-EJSR method learns the customized dictionary from the current probe subject and avoids naturally the interference of the extra information. Thus, the performance of CD-EJSR is better than others in this group of experiments.

4. Conclusion

In this paper, we propose a new customized dictionary-based face recognition approach with the extended joint sparse representation (CD-EJSR). The proposed CD-EJSR has two phases. In the first phase, the customized variation dictionary is learned directly from the observation data (probe samples) instead of a generic set such that it avoids the interference information (such as ID information of the generic samples) from the generic set. The customized dictionary, which contains the uncontrolled variation feature (pose, illumination and expression) corresponding to the intra-class variation of the current probe subject, is learned by using a new model. The alternating direction method of multipliers (ADMM) is employed to solve the optimization problem, and the closed-form solution can be found in each iteration. In the second phase, we present the extended joint sparse representation (EJSR), which utilizes the information of both the customized dictionary and the gallery samples to classify the probe samples. The new EJSR model not only takes advantage of the customized variation dictionary, which represents the intra-class variation between the gallery and probe samples, but also utilizes the group structure to enhance the recognition performance.

The convergence, computation time and size of the customized variation dictionary are discussed in this paper. Extensive numerical experiments are implemented on six databases to verify the performance of CD-EJSR for single-

image-to-image-set face recognition under the complication conditions in illumination, pose and expressions, respectively. Our approach surpasses the related methods in terms of accuracy and computational cost.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Grant 11431002).

References

[1] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, Face Recognition: A Literature Survey, ACM Computing Surveys 35 (4) (2003) 399–458.

[2] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the Face Recognition Grand Challenge, in: IEEE computer society conference on Computer Vision and Pattern Recognition, 2005. CVPR 2005., Vol. 1, IEEE, 2005, pp. 947–954.

[3] X. Tan, S. Chen, Z. H. Zhou, F. Zhang, Face Recognition from A Single Image Per Person: A Survey, Pattern Recognition 39 (9) (2006) 1725–1745.

[4] Y. Hu, A. S. Mian, R. Owens, Sparse Approximated Nearest Points for Image Set Classification, in: IEEE conference on Computer vision and pattern recognition, IEEE, 2011, pp. 121–128.

[5] P. Zhu, L. Zhang, W. Zuo, D. Zhang, From Point to Set: Extend the Learning of Distance Metrics, in: Proceedings of the IEEE International Conference on Computer Vision, 2013, pp. 2664–2671.

[6] B. Zhang, Z. Li, X. Cao, Q. Ye, C. Chen, L. Shen, A. Perina, R. Jill, Output Constraint Transfer for Kernelized Correlation Filter in Tracking, IEEE Transactions on Systems, Man, and Cybernetics: Systems 47 (4) (2017) 693–703.

[7] I. Jolliffe, Principal Component Analysis, Wiley Online Library, 2005.

[8] M. Turk, A. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.

[9] P. N. Belhumeur, J. P. Hespanha, D. Kriegman, Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (7) (1997) 711–720.

[10] M. S. Bartlett, Independent Component Representations for Face Recognition, in: Face Image Analysis by Unsupervised Learning, Springer, 2001, pp. 39–67.

[11] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, Y. Ma, Robust Face Recognition via Sparse Representation, IEEE Transactions on Pattern Analysis and Machine Intelligence 31 (2) (2009) 210–227.

[12] S. Gao, I. W. H. Tsang, L. T. Chia, Kernel Sparse Representation for Image Classification and Face Recognition, in: European Conference on Computer Vision (ECCV), Springer, 2010, pp. 1–14.

[13] I. Naseem, R. Togneri, M. Bennamoun, Linear Regression for Face Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 32 (11) (2010) 2106–2112.

[14] L. Zhang, M. Yang, X. Feng, Sparse Representation or Collaborative Representation: Which Helps Face Recognition?, in: IEEE International Conference on Computer Vision (ICCV), IEEE, 2011, pp. 471–478.

[15] X. Peng, L. Zhang, Z. Yi, K. K. Tan, Learning Locality-Constrained Collaborative Representation for Robust Face Recognition, Pattern Recognition 47 (9) (2014) 2794–2806.

[16] B. Zhang, A. Perina, V. Murino, A. Del Bue, Sparse Representation Classification with Manifold Constraints Transfer, in: The IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4557–4565.

[17] A. Hadid, M. Pietikäinen, From Still Image to Video-based Face Recognition: An Experimental Analysis, in: Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on, IEEE, 2004, pp. 813–818.

[18] O. Arandjelović, G. Shakhnarovich, J. Fisher, R. Cipolla, T. Darrell, Face Recognition with Image Sets Using Manifold Density Divergence, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, Vol. 1, IEEE, 2005, pp. 581–588.

[19] R. Wang, S. Shan, X. Chen, W. Gao, Manifold-Manifold Distance with Application to Face Recognition based on Image Set, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8.

[20] H. Cevikalp, B. Triggs, Face Recognition based on Image Sets, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2010, pp. 2567–2573.

[21] Y. Hu, A. S. Mian, R. Owens, Face Recognition Using Sparse Approximated Nearest Points between Image Sets, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (10) (2012) 1992–2004.

[22] R. Wang, H. Guo, L. S. Davis, Q. Dai, Covariance Discriminative Learning: A Natural and Efficient Approach to Image Set Classification, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 2496–2503.

[23] L. Huang, J. Lu, Y. P. Tan, Co-Learned Multi-View Spectral Clustering for Face Recognition based on Image Sets, Signal Processing Letters, IEEE 21 (7) (2014) 875–879.

[24] T. K. Kim, J. Kittler, R. Cipolla, Discriminative Learning and Recognition of Image Set Classes Using Canonical Correlations, IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (6) (2007) 1005–1018.

[25] H. Zhang, N. M. Nasrabadi, Y. Zhang, T. S. Huang, Joint Dynamic Sparse Representation for Multi-View Face Recognition, Pattern Recognition 45 (4) (2012) 1290–1298.

[26] X. Zhang, D. S. Pham, S. Venkatesh, W. Liu, D. Phung, Mixed-Norm Sparse Representation for Multi View Face Recognition, Pattern Recognition 48 (9) (2015) 2935–2946.

[27] Y. C. Chen, V. M. Patel, P. J. Phillips, R. Chellappa, Dictionary-based Face Recognition from Video, in: European Conference on Computer Vision (ECCV), Springer, 2012, pp. 766–779.

[28] J. Lu, G. Wang, W. Deng, P. Moulin, Simultaneous Feature and Dictionary Learning for Image Set based Face Recognition, in: European Conference on Computer Vision (ECCV), Springer, 2014, pp. 265–280.

[29] S. Chen, J. Liu, Z. H. Zhou, Making FLDA Applicable to Face Recognition with One Sample Per Person, Pattern Recognition 37 (7) (2004) 1553–1555.

[30] Q. X. Gao, L. Zhang, D. Zhang, Face Recognition Using FLDA with Single Training Image Per Person, Applied Mathematics and Computation 205 (2) (2008) 726–734.

[31] Y. Su, S. Shan, X. Chen, W. Gao, Adaptive Ggneric Learning for Face Recognition from A Single Sample Per Person, in: IEEE computer society conference on Computer Vision and Pattern Recognition, 2010, pp. 2699–2706.

[32] A. Majumdar, R. K. Ward, Pseudo-Fisherface Method for Single Image Per Person Face Recognition, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008, 2008, pp. 989–992.

[33] W. Deng, J. Hu, J. Guo, Extended SRC: Undersampled Face Recognition via Intraclass Variant Dictionary, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (9) (2012) 1864–1870.

[34] W. Deng, J. Hu, J. Guo, In Defense of Sparsity based Face Recognition, in: IEEE computer society conference on Computer Vision and Pattern Recognition, 2013, pp. 399–406.

[35] M. Yang, L. Van, L. Zhang, Sparse Variation Dictionary Learning for Face Recognition with A Single Training Sample Per Person, in: IEEE International Conference on Computer Vision (ICCV), 2013, pp. 689–696.

[36] L. Zhuang, T. H. Chan, A. Y. Yang, S. S. Sastry, Y. Ma, Sparse Illumination Learning and Transfer for Single-Sample Face Recognition with Image Corruption and Misalignment, International Journal of Computer Vision (2014) 1–16.

[37] R. X. Ding, D. K. Du, Z. H. Huang, Z. M. Li, K. Shang, Variational Feature Representation-based Classification for Face Recognition with Single Sample Per Person, Journal of Visual Communication and Image Representation 30 (2015) 35–45.

[38] Z. M. Li, Z. H. Huang, K. Shang, A Customized Sparse Representation Model with Mixed Norm for Undersampled Face Recognition, IEEE Transactions on Information Forensics and Security 11 (10) (2016) 2203–2214.

[39] J. Lu, Y. P. Tan, G. Wang, G. Yang, Image-to-Set Face Recognition Using Locality Repulsion Projections and Sparse Reconstruction-Based Similarity Measure., IEEE Transactions on Circuits and Systems for Video Technology 23 (6) (2013) 1070–1080.

[40] Z. Huang, R. Wang, S. Shan, X. Chen, Learning Euclidean-to-Riemannian Metric for Point-to-Set Classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 1677–1684.

[41] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers, Foundations and Trends® in Machine Learning 3 (1) (2011) 1–122.

[42] A. Rakotomamonjy, Surveying and Comparing Simultaneous Sparse Approximation (or Group-Lasso) Algorithms, Signal Processing 91 (7) (2011) 1505–1526.

[43] A. M. Martinez, R. Benavente, The AR Face Database, Technical report, CVC Technical Report (1998).

[44] T. Sim, S. Baker, M. Bsat, The CMU Pose, Illumination, and Expression Database, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (12) (2003) 1615–1618.

[45] Georgia Tech Face Database, ftp://ftp.ee.gatech.edu/pub/users/hayes/facedb/.

[46] R. Gross, I. Matthews, J. Cohn, T. Kanade, S. Baker, Multi-pie, Image and Vision Computing 28 (5) (2010) 807–813.

[47] G. B. Huang, R. Manu, B. Tamara, L. M. Erik, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Tech. Rep. 07–49, University of Massachusetts, Amherst (October 2007).

[48] S. J. Kim, K. Koh, M. Lustig, S. Boyd, D. Gorinevsky, An Interior-Point Method for Large-Scale l_1-Regularized Least Squares, IEEE Journal of Selected Topics in Signal Processing 1 (4) (2007) 606–617.

[49] D. L. Donoho, Y. Tsaig, Fast Solution of l_1-Norm Minimization Problems When the Solution May be Sparse, IEEE Transactions on Information Theory 54 (11) (2008) 4789–4812.