International Journal of Machine Learning and Cybernetics (2019) 10:1051–1064
https://doi.org/10.1007/s13042-017-0782-5

ORIGINAL ARTICLE

Image-set based face recognition using K-SVD dictionary learning

Jingjing Liu[1] · Wanquan Liu[2] · Shiwei Ma[1] · Meixi Wang[1] · Ling Li[2] · Guanghua Chen[1]

Received: 14 January 2017 / Accepted: 26 December 2017 / Published online: 12 January 2018

Abstract

With rapid development of digital imaging and communication technologies, image set based face recognition (ISFR) is becoming increasingly important and popular. On one hand, easy capture of large number of samples for each subject in training and testing makes us have more information for possible utilization. On the other hand, this large size of data will eventually increase training and classification time and possibly reduce the recognition rate if they are not used appropriately. In this paper, a new face recognition approach is proposed based on the K-SVD dictionary learning to solve this large sample problem by using joint sparse representation. The core idea of this proposed approach is to learn variation dictionaries from gallery and probe face images separately, and then we propose an improved joint sparse representation, which employs the information learned from both gallery and probe samples effectively. Finally, the proposed method is compared with some related methods on several popular face databases, including YaleB, AR, CMU-PIE, Georgia and LFW databases. The experimental results show that the proposed method outperforms several related face recognition methods.

Keywords Image set · Face recognition · K-SVD dictionary learning · Improved joint sparse representation

1 Introduction

Face recognition (FR) is an active research topic in computer vision community [1–3]. Over past two decades, various of subspace learning approaches have been proposed, such as principle component analysis (PCA) [4, 5] or its variants [6, 7], linear discriminant analysis (LDA) [8], independent component analysis (ICA)[9], sparse representation classification (SRC) [10], kernel sparse representation (KSR) [4], linear regression (LR) [5], collaborative representation classification (CRC) [6], locality constrained collaborative representation (LCCR) [7], structured regularized robust coding (SRRC) [11], multi-step linear representation-based classification (MLRC) [12] and so on. All these methods are based on an important hypothesis that the face image of one person lies on a special subspace and the training set contains multi images of one object. These approaches can achieve satisfactory results in controlled environment when training samples are sufficiently large. We call this category as the set to image face recognition (SIFR).

Different from the SIFR, there is usually only a single gallery sample per person (SSPP) enrolled in many practical face recognition systems. And many existing face recognition methods may fail to work in this case. For conventional SSPP face recognition [13, 14], there is only one image per person for training and one sample per person for probing. We call this category as image to image face recognition (IIFR). Generally, a gallery sample is an frontal view photo and natural expression, and a probe sample is a captured photo which may be affected by many factors, including background illumination, pose, and facial corruption/disguise (such as makeup, beard and glasses)in some situations. Therefore, there is a large gap between a normal gallery sample and a probe sample for the same person. This is a critical issue for IIFR to bridge such gap. To address this IIFR problem, several kinds of efficient methods [15–22] were presented recently. These methods are based on generic learning, which assumes that there is a generic training set, and this set and gallery set share some similar variation information for both inter-class and intra-class. Particularly, dictionary learning methods (such as extended

✉ Wanquan Liu
W.Liu@curtin.edu.au

Jingjing Liu
Liu.jingjing@shu.edu.cn

1 School of Mechatronical Engineering and Automation, Shanghai University, Shanghai, China

2 Department of Computing, Curtin University, Perth, WA, USA

sparse representation-based classification (ESRC) [17, 18], sparse variation dictionary learning (SVDL) [19], sparse illumination learning and transfer (SILT) [20], variational feature representation-based classification (VFRC) [21] can learn a dictionary from an additional generic set in order to offer some extra information which was mentioned above. However, these IIFR methods ignore possible collection of multiple probe samples, which may have potentially useful information to improve performance of FR. In 2013, Lu et. al. [23] developed a new face recognition problem, named as image to set face recognition (ISFR). In this ISFR framework, there are multiple probe samples per person in testing phase, and only one gallery sample per person in training phase. This method is more suitable for many applications, but for the multiple face images, it is difficult to improve the performance.

With rapid development of digital imaging and communication technologies, more information can be provided to describe the person with image sets, image set based face recognition is still challenging because there are usually large intra-class variations within a set, especially when they are captured in unconstrained environments. In this case, the set to set face recognition (SSFR) becomes a very important research topic with broad applications and has attracted much intention in research community. Recently, a number of approaches [8, 9, 24–29] have been proposed to solve SSFR problem in a video-based framework classification. Being different from conventional SIFR where the probe is a single image, SSFR assumes that both gallery set and probe set have multi samples per subject. Also, all images are captured from different poses, illuminations, expressions and resolutions, these face nuisances will eventually affect classification performance. In fact, some key issues in SSFR include how to model an image set for each subject and compute distance or similarity between probe and gallery sets effectively. In order to solve these problems, researchers have proposed different approaches for SSFR including subspace [30–32], manifold [8, 24, 28], affine or convex hull [11, 12, 25–27], nearest points distance [26, 27, 33] and dictionary learning [1, 4, 5, 9, 15, 18, 19, 21, 29, 34–38]. However, most existing dictionary based image set based face recognition methods are unsupervised, which are not discriminative enough to classify face sets. Moreover, these methods learn dictionaries using the original raw pixels, which may contain some noisy components that are irrelevant to dictionary learning. To make the learned dictionary meet the demands of classification tasks, some supervised dictionary learning methods [39, 40] have been proposed recently. In [39], learning a discriminative analysis-synthesis dictionary pair which combined with a linear classifier for FR is presented, the main drawback is the high computational cost and it is not suitable for large scale learning problems. In order to reduce the computational complexity, Li et al. [40] proposed a label that embedded within the atoms to improve the discriminative ability of the shared dictionary. Since face images usually lie on a low-dimensional manifold, it is desirable to seek the most discriminative features in a low-dimensional subspace and suppress the useless information to promote learning dictionaries for image sets.

In this paper, our aim is to investigate the problem of SSFR. Inspired by works of sparse representation and dictionary learning methods, we propose a new method, named as the K-SVD dictionary learning based face recognition with improved joint sparse representation (KSVD-IJSR), to solve the SSFR problem. First, the gallery and probe dictionary are learned from given samples for each subject. The learned dictionary for each subject would contain variation features about the uncontrolled variations, such as pose, illumination and expression. Then, we use the improved joint spares representation classification (JSRC) approach, which utilizes the information both from dictionary and gallery samples to classify probe samples. In summary, our main contributions can be summarized as:

1. Being different from conventional dictionary learning methods [17, 19, 20] for SIFR problem, the dictionaries in this paper are learned from gallery samples and probe samples respectively. In this case, the significant performance improvement is benefit from linear regression with the generic learning method for SSFR.
2. The variation dictionary is learned by using the K-SVD model, which is employed to solve this optimization problem via the ℓ_1 optimization in each step. This adoption is more suitable in practice than the models using the regularization term with ℓ_0 norm. Hence, the optimization process can be implemented efficiently as it is a convex problem.
3. The improved joint sparse representation (IJSR) model is used. It not only takes advantage of the learned variation dictionary, which represents the intra-class variation between the gallery and probe samples, but also utilizes group structure to enhance the performance for recognition.
4. Extensive experimental results demonstrate that the proposed approach can achieve a new state-of-the-art performance for FR under various complex scenarios in comparison to other related methods.

The rest of this paper is organized as follows. The related works are reviewed in Sect. 2. The proposed approach is presented in Sect. 3. Extensive experiments for validating efficacy of the proposed approach are shown in Sect. 4. We conclude this paper in Sect. 5.

2 Related work

In recent years, there are considerable interests in developing new methods for image set based face recognition. These methods can be classified into two categories: linear Regression based approaches [5], and generic learning based approaches [36]. For first category, gallery image samples for one object are assumed to span a linear subspace, in which its probe samples should locate [5, 41]. We use this concept to develop class-specific models for registered users, thereby face recognition is defined as a problem of linear regression. For second category, an additional generic training set with multiple samples per person is employed to extract discriminative features. Even though these methods work well for SSFR problem, the performance of this type of methods is affected by the selection of generic training set, and such selection is extraordinary. However, most generic learning methods mentioned above share a common characteristic that a learned dictionary added into the ESRC framework to finish the FR process. To address this, Chen et al. [4–6] presented a dictionary-based approach for image set based face recognition by building one dictionary for each face image set and then using these dictionaries to measure similarity among different face image sets. While reasonably good recognition rates can be obtained, their approach is generative and the dictionaries are learned from the original raw pixels, which may contain some noisy and irrelevant components. Here, these three types of methods will be described with the following notations.

Let $X = [X_1, X_2, \cdots, X_C] \in \mathbb{R}^{d\times n}$ denote gallery sample matrix, which contains C different classes; let $X_i = [x_{i1}, x_{i2}, \cdots, x_{in_i}] \in \mathbb{R}^{d\times n_i}$ denote samples of the i-th class, where $n_i(i = 1, 2, \cdots, C)$ is the number of samples in the i-th class, and $n = \Sigma_(i = 1)^C n_i$; $x_{ij} \in \mathbb{R}^d (i = 1, 2, \cdots, C, j = 1, 2, \cdots, n_i)$ is the j-th face image of the i-th special class, d is the number of pixels of a face image. All of the face images in this paper are vectorized.

2.1 Linear regression model

The linear regression model has been extensively studied in [5, 6, 10, 42–47]. Many extended methods are also developed. In fact, the typical linear regression model can be described as follows:

$$\min_{a} \| W(y - X\alpha) \|_p^p + \lambda \| \alpha \|_q^q \tag{1}$$

where $p \in \{1, 2\}$, $q \in \{1, 2\}$, and $\lambda \geq 0$; W is a weighting matrix; and $\lambda \| \alpha \|_q^q$ is the regularization term. Several popular models are the special cases of model 1. First, when $\lambda = 0$, $p = 2$ and W is an identity matrix, model 1 degenerates to LRC [5] model.

$$\min_{\alpha_i} \| y - X_i\alpha_i \|_2^2, i = 1, 2, \cdots, C$$

Secondly, if $\lambda > 0$, $p = 2$, $q = 1$ and W is an identity matrix, SRC [10] model is obtained as follows

$$\min_{\alpha} \| A\alpha - y \|_2^2 + \lambda \| \alpha \|_1$$

Finally, $\lambda > 0$, $p = 2$, $q = 2$ and W is an identity matrix, model 1 is CRC [6] model:

$$\min_{\alpha} \| X\alpha - y \|_2^2 + \lambda \| \alpha \|_2^2$$

Usually, all these models need sufficiently large gallery samples to build a complete space. For undersampled case or SSFR, performance of these linear regression methods would decline significantly.

2.2 Generic learning model

In recent years, methods based on generic learning have been extensively investigated by many researchers [48–50]. All these methods are based on a fundamental assumption that a selected generic set and gallery set share similar intra-class variational information. An additional generic set with multi-samples per person is used to extract variational information, which can enrich the diversity of gallery set. The strategy that combines generic learning methods with linear regression methods is used in many face recognition systems. The generic learning model is described as follows:

$$\min_{\alpha,\beta} \| y - [X \quad V][\begin{matrix}\alpha \\ \beta\end{matrix}] \|_p^p + \lambda \| [\begin{matrix}\alpha \\ \beta\end{matrix}] \|_q^q \tag{2}$$

where $X \in \mathbb{R}^{d\times n}$ denotes gallery sample matrix; $V \in \mathbb{R}^{d\times m}$ denotes generic variational dictionary. p, q and λ are the same as the parameters in model 2. $[\alpha^T, \beta^T]^T$ denotes the representation coefficients of probe sample over gallery sample matrix and general variational dictionary. The dimensions of α and β equal to the column numbers of the matrix A and V, respectively. When $p = 2$ and $q = 1$, model 2 degenerates to the ESRC [17, 18] model:

$$\min_{\alpha,\beta} \| y - [X \quad V][\begin{matrix}\alpha \\ \beta\end{matrix}] \|_2^2 + \lambda \| [\begin{matrix}\alpha \\ \beta\end{matrix}] \|_1$$

This method uses an additional generic set to supplement the variational information of gallery set. The variant bases for each class are usually obtained by extracting images from some essential images in the same class. For example, in [17], frontal faces in each class are chosen as the essential images. It has been shown that the ESRC approach can improve the performance in the case of undersampled FR. In addition, there are other typical methods, such as the SVDL [19] and SILT [20], which have embedded the learned sparse dictionaries into the ESRC framework.

2.3 Dictionary learning model

Dictionary learning methods aim to learn a good dictionary from original training samples such that it can properly represent the original samples in feature extraction process. General Sparse Representation based Classification (SRC) and Collaborative Representation based Classification (CRC) techniques can create a dictionary from the original samples without any sample modification [6, 10, 31], therefore, performance of those algorithms directly depends upon availability and quality of samples instead of learned dictionary. In 2006, Aharon et al. proposed the K-SVD based dictionary learning algorithm, which can create an over-complete dictionary [52], and this new dictionary learning algorithm has promoted the sparse representation based approaches. But due to its unsupervised nature of this KSVD learning method, this particular algorithm suffers from low classification accuracy. On the other hand, supervised dictionary learning methods can take advantages of class label information to make the learned dictionary more discriminative and thus can improve classification performance. Mathematically, the KSVD can be formulated as follows:

$$\min_{D,X}\{\| Y - DX \|_F^2\}, s.t.\forall i, \| x_i \|_0 \leq T_0 \tag{3}$$

where $D \in \mathbb{R}^{n\times K}$ is the dictionary, $X = \{x_i\}_{i=1}^N$ is the coefficient matrix. We will revise this approach in our proposed model in next section.

3 The proposed approach

Based on previous works, different models on dictionary learning use different regularization terms including ℓ_1-norm and ℓ_2-norm. These regularization terms in linear regression models and generic learning models are usually assumed to follow Gaussian or Laplacian distribution. The ℓ_2-regularization term corresponds to a Gaussian prior, and the ℓ_1 -regularization term corresponds to Laplacian (or double exponential) prior. However, in the real world, face image has much more complex structures with non-negative intensity values, which implies that it is not appropriate to simply use ℓ_1-norm or ℓ_2-norm to regularize the linear regression model. One key question is whether we can take both advantages of ℓ_1 and ℓ_2 regularization. In this paper, we propose a model with $\ell_{2,1}$-norm for probe samples, which can produce a solution that achieves among group-sparsity with the level of ℓ_1. The proposed model in fact improved the JSRC framework such that the learned dictionary can offset the intra-class variation dictionary between gallery and probe samples. The group Lasso structure is still retained

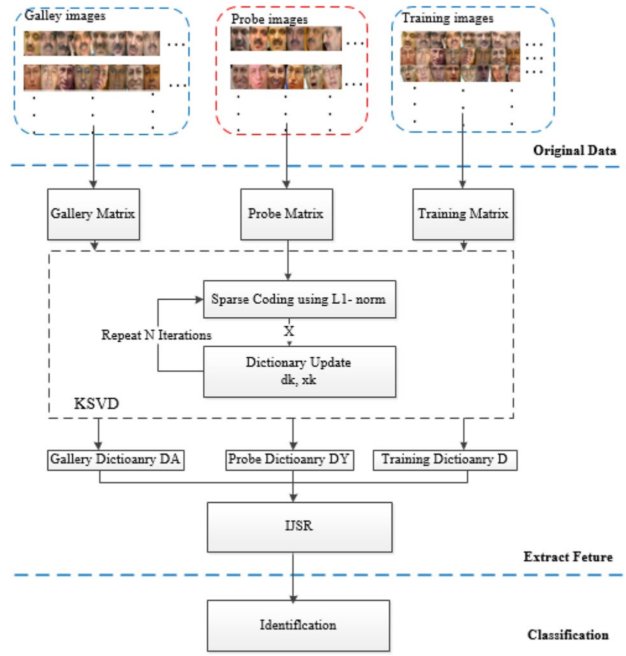

Fig. 1 The basic idea of our proposed approach. In second stage, the variation dictionaries are learned by using K-SVD, to encode pose, illumination, expression and occlusion information in the original image sets. Then in third stage, this customized dictionary is used in the IJSR model to supply intra-class variation. Finally, we recognize its label by using the reconstruction error

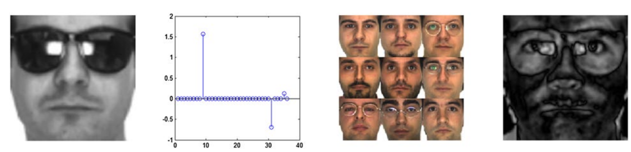

Fig. 2 The linear representation of a face image in the AR database. The test image (left), which is (**a**) potentially occluded, as a sparse linear combination of all the training images (middle) plus sparse errors (right) due to occlusion or corruption. Blue (darker) coefficients correspond to training images of the correct individual

to enhance recognition performance. The overview of our proposed approach is shown in Fig. 1.

3.1 K-SVD dictionary learning method

For convenience, based on the idea of linear representation, the probe images Y can be represented as

$$Y = AB + e.$$

where A and B denote the gallery samples and coefficient matrix, and e denotes the error term. Fig. 2 illustrates the above representation. However, It is well known that appearance of the captured face images is affected by background illumination, pose, and facial corruption/disguise (such as scarf, beard, and glasses) in the world. Such recognition performance will be affected by the different appearances in testing process, that means there is still a gap between a gallery sample and any probe samples for the same person. Inspired by the work [51], any face image can be linearly represented by its dictionary images in the same subspace. Obviously, we can use a similar assumption that variations on one face are located in a common variation subspace.

Assume there are totally P subjects and n images per person in the training set, we can write A as $A = [A_1, A_2, \cdots, A_p]$. Hence, A can be rewritten as $a = [a_1, a_2, \cdots, a_n] \in \mathbb{R}^{d \times Pn}$, where a_i is a d-dimensional vector of cropped face image. In order to extract more discriminative and robust information from this training set, K-SVD aims to simultaneously learn a discriminative structured dictionary for all individual image sets for each subject, under which each image frame is encoded by a discriminative coefficient. To achieve this, we formulate the following optimization problem:

$$\min_{D_i, B_i} \| A_i - D_i B_i \|_F^2 + \lambda \| B_i \|_1^2 . \tag{4}$$

where $A \in \mathbb{R}^{d \times Pn}$ is a group of samples from the training subject, each column of A denotes a face vector, B_i is the coefficient vector of A_i, which is the sparse representation of training samples in A, and $\| \cdot \|_1$ is defined by the sum of the ℓ_1 norm of all rows of a matrix. Here, the representation coefficients B_i are simultaneously penalized by the same parameter λ in this paper. $D = [D_1, D_2, \cdots, D_p]$ is a structured dictionary learned from the K-SVD.

In general, the difference of sparse codes for two face images should be minimized if they are from the same class as they should look similar, the difference of sparse codes for two face images should be maximized if they are from different classes as they should be different. In this case, the images of per subject (including gallery, generic and probe) are separately trained by the K-SVD to obtain different dictionaries. It is reasonable that combining all dictionaries forms an integrated dictionary which has the standard class labels corresponding to the image labels before trained. So, after the group sparse coding and dictionary updating, we can archive a collection of B and D with class labels, where $B = [B_1, B_2, \cdots, B_p]$ and $D = [D_1, D_2, \cdots, D_p]$. In detail, the update of k-th column of each subject is done by rewriting the penalty term as

$$\begin{aligned} \| A_i - D_i B_i \|_F^2 &= | \left(A_i - \sum_{j \neq k} d_j b_T^j \right) - d_k b_T^k |_F^2 \\ &= \| E_k - d_k b_T^k \|_F^2 \end{aligned} \tag{5}$$

where b_T^k denotes the k-th row of B_i. After this step, we preserve the matrix Ω_k with a size $N \times |\{i|1 \leq i \leq K, b_T^k(i) \neq 0 \bigcap b_T^k(i) \leq T\}|$ according to the sparsity [52]. The same effect happens with $E_k^R = E_k \Omega_k$, implying a selection of error columns that correspond to examples that use the atom d_k. So the minimization problem as mentioned before becomes $\| E_k^R - d_k b_T^k \Omega_k \|_F^2$, which can be solved into $U \Delta V^T$ by using the SVD. Then we can obtain that, $d_k = U(:, 1)$ and the coefficient vector $b_k = b_T^k \Omega_k$ is actually the first column of $V \times \Delta(1, 1)$. Once the entire dictionary is updated, the sparse coding process will be invoked again and then we can update D consequently. The details of the solution of K-SVD on a manifold is described as below:

Algorithm 1 The K-SVD Algorithm

Input
Training group samples, $A = [A_1, A_2, \cdots, A_p] \in \mathbb{R}^{d \times Pn}$ for P classes, $A_i \in \mathbb{R}^{d \times n}$
.
Initialize $D_0 \in \mathbb{R}^{d \times K}$, set $J = 1$
repeat
  Sparse Coding Step: Using the Lasso technique in [**?**] to compute the representation vectors $b = [b_0, b_1, \cdots, b_N]$ for each subject of A_i.
  Dictionary Update Step: For each column $k = 1, 2, \cdots, N$ in D^{J-1},update it by
  - Calculate the error matrix $E_k = A_i - \sum_{j \neq K}^{K} d_j b_T^j$,
    Restrict E_k by choosing only the columns corresponding to $|\{i|1 \leq i \leq K, b_T^k(i) \neq 0 \bigcap b_T^k(i) \leq T\}|$, and obtain E_k^R
  - Using SVD decomposition $E_k^R = U \triangle V^T$
  - Update the dictionary column $d_k = U(:, 1)$
  - Update the cloned vector $b_k = V(:, 1) * \triangle(1, 1)$
until Set $J = J + 1$.

In above algorithm, we have the update scheme in each iteration such that the process of learning the dictionary can be rapidly implemented. The computation time for dictionary learning will be shown in Sect. 3.4. Fig. 3 shows that the original images and the learned dictionaries of the first gallery subject on LFW databases. We will show how the different dictionaries have impact on the performance for

Fig. 3 Examples of the K-SVD learned dictionaries for one subject in LFW database. The images in each sub-figure contain a group of gallery or probe images from the same subject. The bottom of the two images in each sub-figure are dictionaries which are learned from the top images. These K-SVD learned dictionaries have different sizes

FR in Subsection 3.5. In the following algorithm, only the learned dictionaries are used from the gallery set, training set and probe set, respectively.

3.2 Improved joint sparse representation model based on the K-SVD

In the generic learning model, there are multi samples per class in gallery set, a probe sample usually can not be approximated effectively by gallery samples from the same class. If an auxiliary variational information is provided, the probe sample can be linearly represented by the samples from same class in gallery set and the variational information covered in auxiliary dictionary set. By combining the learning dictionaries from the K-SVD, we can formulate a new learning model. For such purpose, let D_Y be probe dictionary matrix that contains all the samples in probe set, Similarly, D_A is the dictionary matrix that learned from the gallery set, and J is a variational dictionary represent the intra-class relationship between gallery and its normalized samples, and it is created by each gallery samples subtracting its standard images for each subject, such as $J = K - SVD[A_1^- - a_1^*, \cdots, A_p^- - a_p^*]$, where a_i represent the normalized samples of i-th subjects, and A_i^- represent the rest gallery samples. Then we can have the following model

$$D_Y = D_A X + JB.$$

The nonzero part in X of representation coefficient should be sparse, and the nonzero elements lie in the location representing samples of the same class. The variation part B should be sparse yet, and its nonzero elements correspond to the special variational bases. Many researchers have revealed that X and B can be solved by ℓ_1-minimization. For example, in the ESRC model, the representation coefficients X and B are simultaneously penalized by the same parameter λ. However, in our problem of this paper, we have multiple testing samples for same subject. Thus, identifying a group of samples at the same time is an important requirement in this paper. As an extension of the ESRC, we have multiple tasks in this situation. For dictionary learning task, we need to use the K-SVD to obtain dictionaries. For recognition task, the joint sparse representation classification (JSRC)[44] is used to exploit the shared information from all the samples. The proposed approach based on the K-SVD dictionary learning and improved JSRC model (named as KSVD-IJSR in this paper) is formulated now as follows.

$$\min_{X,B} \frac{1}{2} \parallel D_Y - JB - D_A X \parallel_F^2 + \lambda \parallel X \parallel_{2,1} + \mu \parallel B \parallel_{2,1} . \tag{6}$$

or

$$\min_{X,B} \frac{1}{2} \parallel D_Y - [J \quad D_A][\begin{matrix} B \\ X \end{matrix}] \parallel_F^2 + \parallel [\begin{matrix} \mu B \\ \lambda X \end{matrix}] \parallel_{2,1} . \tag{7}$$

where B and X denote the representation coefficient on J and D_A, respectively. The variation dictionary J represents the intra-class variation between gallery and probe samples. It joints with gallery samples to confront the complex variation on face images in testing phase. For integration of some variables, we denote $\widehat{J} = \frac{\gamma}{\mu} J$, $\widehat{D}_A = \frac{\gamma}{\lambda} D_A$, $\widehat{X} = \frac{\lambda}{\gamma} X$, and $\widehat{B} = \frac{\mu}{\gamma} B$, where $\gamma = \lambda + \mu$, and then can express model (7) as

$$\min_{X,B} \frac{1}{2} \parallel D_Y - [\widehat{J} \quad \widehat{D}_A][\begin{matrix} \widehat{B} \\ \widehat{X} \end{matrix}] \parallel_F^2 + \gamma \parallel [\begin{matrix} \widehat{B} \\ \widehat{X} \end{matrix}] \parallel_{2,1} . \tag{8}$$

In fact, we can obtain the corresponding variation dictionary $\widehat{J}$ from this model, and obtain the optimal solutions for $\widehat{B}$, $\widehat{X}$. Then, the identity of the probe subject can be obtained via computing the following

$$Identity(D_Y) = \arg\min_i \parallel [D_Y - \widehat{J}\widehat{B} - D_{A_i}\widehat{X}^*] \parallel_{2,1}^2, \tag{9}$$

where D_{A_i} is the i-th column of the gallery dictionary set D_A, and $\widehat{X}^*$ is the i-th row of the coefficient matrix $\widehat{X}^*$.

In order to solve above problem, let $S = [\widehat{J} \quad \widehat{D}_A]$ and $H = [\begin{matrix} \widehat{B}^T \\ \widehat{X}^T \end{matrix}]^T$, we can obtain a simplified model

$$\min_H \frac{1}{2} \parallel D_Y - SH \parallel_F^2 + \gamma \parallel H \parallel_{2,1} . \tag{10}$$

In fact, some existing optimization methods can be used to solve above simplified problem. Here, the diffusion process [53] is used by exploiting the contextual affinities. Given the affinity matrices V and T, which are obtained from

$$\begin{aligned} &\min_{H,V,T} \frac{1}{2} \parallel V \parallel_F^2 + \gamma \parallel T \parallel_{2,1}, \\ &s.t. \quad SH - D_Y - V = 0, \quad H - T = 0. \end{aligned} \tag{11}$$

In order to solve above problem, the augmented Lagrangian function is as follows,

$$\begin{aligned} &L(H, V, T, W_1, W_2) \\ &= \frac{1}{2} \parallel V \parallel_F^2 + \gamma \parallel T \parallel_{2,1} \\ &\quad + tr[W_1^T(SH - D_Y - V)] + \frac{\sigma_1}{2} \parallel SH - D_Y - V \parallel_F^2 \\ &\quad + tr[W_2^T(H - T)] + \frac{\sigma_2}{2} \parallel H - T \parallel_F^2 \end{aligned} \tag{12}$$

where $\sigma_1, \sigma_2 > 0$ are penalty parameters, W_1, W_2 are the Lagrange multipliers associated with the linear constraints, and $tr[\cdot]$ denotes the trace of a matrix. Under the ADMM framework, variables H, V and T can be easily computed by the following three steps,

Step 1: For the variable H and V, we have the update form

$$
\begin{aligned}
&(\sigma_1 I + \sigma_2 I)H = \\
&\sigma_1 S^T(D_Y + V^k - W_1^k/\sigma_1) + \sigma_2(T^k - W_2^k/\sigma_2)
\end{aligned}
\tag{13}
$$

And one can employ the sylvester function(MATLAB package) to solve it. Hence the solution is given by

$$
H^{k+1} = sylvester(\sigma_1 I + \sigma_2 I, \tag{14}
$$

$$
\begin{aligned}
&\sigma_1 S^T(D_Y + V^k - W_1^k/\sigma_1) + \sigma_2(T^k - W_2^k/\sigma_2)) \\
&V^{k+1} = \frac{\sigma_1}{2+\sigma_1}(SH^k - D_Y + W_1^k/\sigma_1)
\end{aligned}
\tag{15}
$$

By using the first-order necessary conditions for unconstrained optimization problem, these two updates of H and V can be easily obtained.

Step 2: For the variable T, the update form is give by:

$$
T^{k+1} = \arg\min_T r \parallel T \parallel_{2,1} + \frac{\sigma_2}{2} \parallel H^k - T + W_2^k/\sigma_2 \parallel_F^2 \tag{16}
$$

Let $H^k + W_2^k/\sigma_2 = Z^k$, and t_i and z_i^k denote the i-th row vectors of T and Z^k, respectively. Then problem 16 can be decomposed as

$$
T^{k+1} = \arg\min_{t_i, i=1,2,\cdots} \Sigma_i[r \parallel t_i \parallel_2 + \frac{\sigma_2}{2} \parallel Z^k - t_i \parallel_2^2]. \tag{17}
$$

Thus, we can find each row of T separately by observing the following result. As proved in [54], the optimal solution of the following problem

$$
\min_t r \parallel t \parallel_2 + \frac{\sigma_2}{2} \parallel z - t \parallel_2^2 . \tag{18}
$$

where is $t = kz$, $k = \max\{1 - \gamma/\sigma_2 \parallel z \parallel_2, 0\}$ if $\parallel z \parallel_2 > 0$, and $k = 0$ if $\parallel z \parallel_2 = 0$. Based on above result, the update formulation of T is given by

$$
T^{k+1} == [t_1^{k+1}, t_2^{k+1}, \cdots, t_{m+l}^{k+1}]^T. \tag{19}
$$

Step 3: The updates for dual variables are:

$$
\begin{aligned}
&W_1^{k+1}/\sigma_1 = W_1^k/\sigma_1 + SH^k - D_Y - V^k. \\
&W_2^{k+1}/\sigma_2 = W_2^k/\sigma_2 + H^k - T^k
\end{aligned}
\tag{20}
$$

where the constant is independent of the variables H, V, T. Through the above updates, the following steps are performed to obtain the solution of model (6).

Algorithm 2 - KSVD-IJSR representation based on the ADMM

Input
Probe face dictionary $D_Y \in \mathbb{R}^{d\times n}$, variational dictionary $J \in \mathbb{R}^{d\times k}$, Galley dictionary set $D_A \in \mathbb{R}^{d\times m}$, regularization parameters λ and μ, Penalty parameters σ_1 and σ_2.
Pre-compute $V \in \mathbb{R}^{d\times n}$, $T \in \mathbb{R}^{(m+l)\times n}$, $W_1/\sigma_1 \in \mathbb{R}^{d\times n}$ $W_2/\sigma_2 \in \mathbb{R}^{(m+l)\times n}$, $\gamma = \lambda + \mu$, $S = [(\gamma/\lambda)J \quad (\gamma/\mu)D_A]$
repeat
Update variable H amd V with

$$
\begin{aligned}
&H^{k+1} = sylvester(\sigma_1 I + \sigma_2 I, \\
&\sigma_1 S^T(D_Y + V^k - W_1^k/\sigma_1) + \sigma_2(T^k - W_2^k/\sigma_2)) \\
&V^{k+1} = \frac{\sigma_1}{2+\sigma_1}(SH^k - D_Y + W_1^k/\sigma_1)
\end{aligned}
$$

Update T with

$$
T^{k+1} = \arg\min_{t_i, i=1,2,\cdots} \Sigma_i[r \parallel t_i \parallel_2 + \frac{\sigma_2}{2} \parallel Z^k - t_i \parallel_2^2]
$$

Variables update

$$
\begin{aligned}
&W_1^{k+1}/\sigma_1 = W_1^k/\sigma_1 + SH^k - D_Y - V^k. \\
&W_2^{k+1}/\sigma_2 = W_2^k/\sigma_2 + H^k - T^k
\end{aligned}
$$

until Some stopping criterion

4 Experiments

In this section, the effectiveness of the proposed method in different situations is evaluated. Several popular face databases, including AR [55], YaleB [41], CMU-PIE [56], Georgia [35], and LFW [33], are used and some samples of the images in these databases are shown in Fig. 4. In order to make the comparison fair among all adopted methods, all the images are cropped into the size 32×32. In each database, 50% of the total number of images per person is employed as gallery samples, and the rest of the images are employed as the probe samples. We perform experiments under 4 different categories to show the performance of our proposed method, including: (1) parameter setting, (2) computational time, (3) different variational dictionaries, (4) performance evaluation on different databases. All experiments are conducted on a PC platform with 64-bit windows 7 operating system, Intel Core i5-3550S CPU, and 8G of RAM.

In the subsequent experiments, first some suitable parameters of the proposed model are selected, and the computation time of K-SVD dictionary learning is investigated along with the influence of utilizing various dictionary sizes. In order to improve the performance of SSPP problem, we compare the proposed method with several related sparse coding based methods such as JSRC [57], SRC [10] and

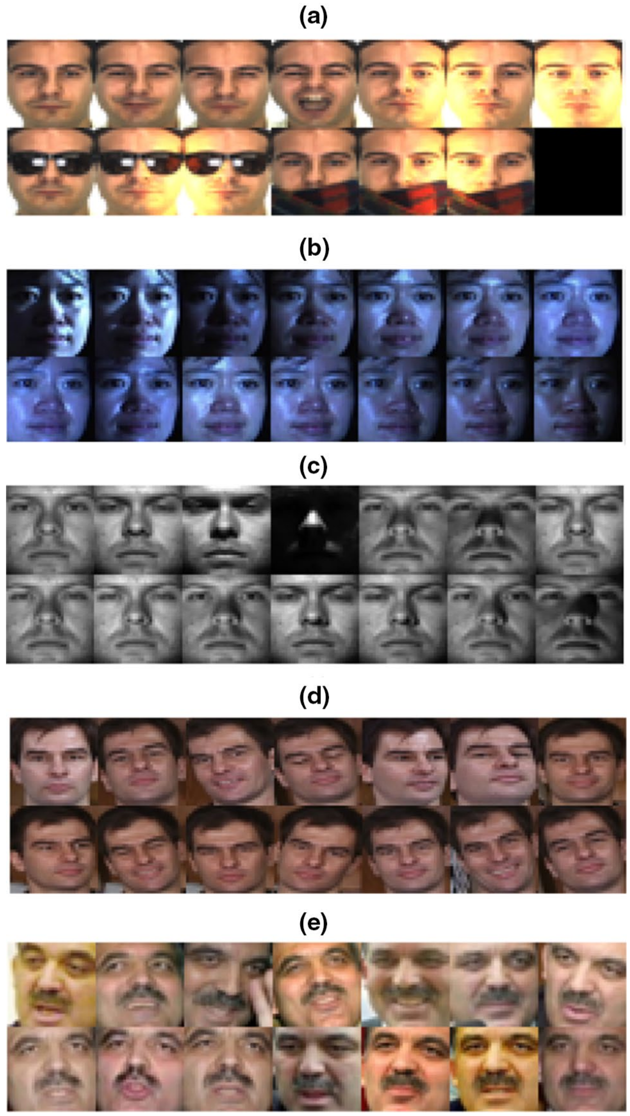

Fig. 4 Face image samples in the AR, CMU-PIE, Yale, Georgia and LFW databases

related dictionary learning methods, including ESRC [17], SVDL [19], SILT [20], customized sparse representation model CSR-MN [58]. The ℓ_1-regularized minimization in SRC and SVDL is solved by $\ell_1 - \ell_s$ algorithm [34]; the ℓ_1 -regularized minimization in ESRC and SILT is solved by Homotopy algorithm [38], a generalization of the extended sparse representation-based classification model with mixed norm in CSR-MN, and the ADMM framework is employed to solve JSRC and our method. The recognition rates of SRC, ESRC, SVDL, SILT are obtained by the majority voting strategy. In the following experiments, the parameters of these related methods are the same as those in their original papers.

4.1 Parameter setting

In order to select appropriate parameters, the values of parameters λ and μ are estimated via using grid search. The regularization parameters λ is in the grid range of $[0.005, 0.001, \ldots, 0.0001]$ and the coefficient μ is investigated in the grid region $[0.1, 0.2, \cdots, 1.0]$. We follow the cross validation strategy to decide these optimal parameters. We used fifteen-fold cross validation (each person have 15 images) to find the available combination schemes of relevant parameters for our method. The KSVD-IJSR method was evaluated on the LFW database and we only consider the KSVD dictionary learning model in the case ($L = size(A, 2)$). The average recognition rates are recorded in Table 1. As shown in Table 1, $\lambda = 0.0005$, $\mu = 0.4$ are the best choice and they will used for the following experiments.

4.2 Reconstruction error and computational time

Since the proposed model requires to learn a variational dictionary, we use K-SVD with ℓ_1 norm to obtain such dictionary. This problem is a convex problem which can be iteratively solved quickly. We can observe the reconstruction error since it is an important indicator. In Fig. 5, the different number of coefficients and iterations are considered. It show the reconstruction error of the YaleB database based K-SVD algorithm with the sparsity level 5 to 35, and it is clear that with the increasing sparsity, the error is decreased, and we also can see that all the cases reach the reconstruction error threshold (20.5) after 12 iterations. These show that K-SVD algorithm always converges rapidly regardless of the dictionary size.

In addition, the time of recognition process is also an important indicator for a face recognition system. In the sequent section, the computation time of learning dictionary on the database of LFW and YaleB is calculated when

Table 1 Parameter setting of the KSVD dictionary learning model

λ	μ									
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.005	49.84	47.94	49.20	47.30	51.10	48.57	49.20	47.94	50.47	49.20
0.001	50.47	50.84	47.94	48.57	47.94	48.57	49.20	49.20	49.20	47.94
0.0005	47.30	46.41	50.47	**56.47**	47.94	49.84	49.20	47.30	49.84	48.20
0.0001	52.10	51.10	48.30	49.20	48.57	47.30	46.67	48.57	50.47	48.57

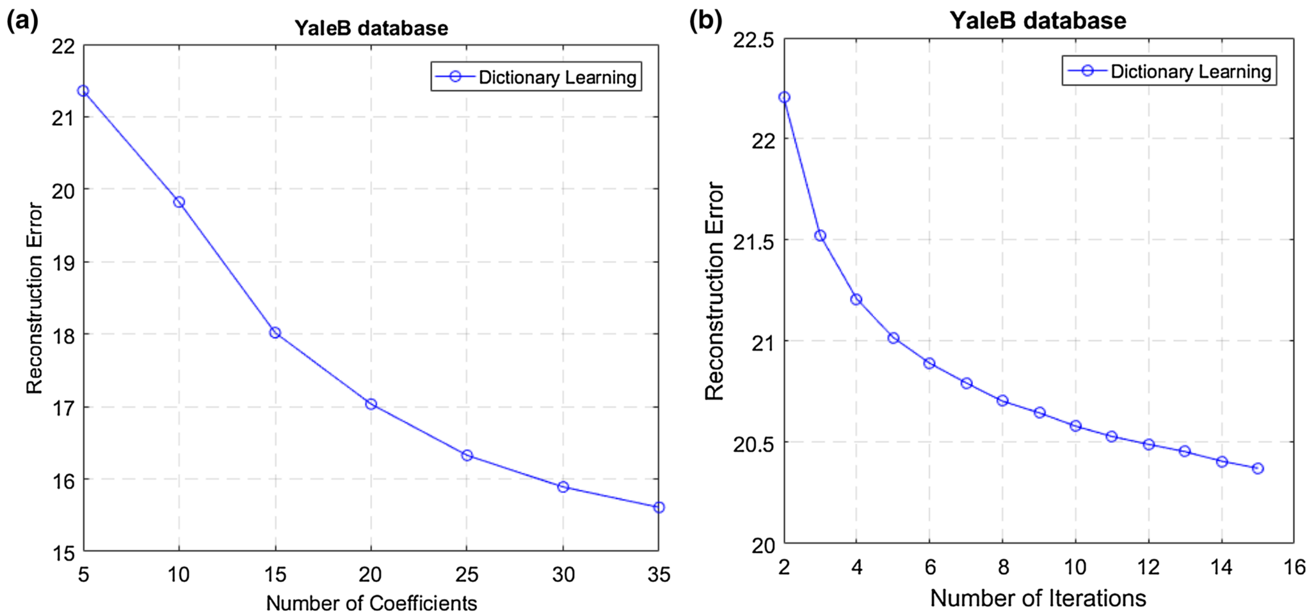

Fig. 5 An example of KSVD process on the YaleB face database. **a** The reconstruction error via same sparsity. **b** The reconstruction error via same iteration

it has 16 images on the LFW database, 64 images on the YaleB database for each person. For a fair and thorough comparisons, per 100 persons are recorded on each database respectively for the computational time of dictionary learning. We implement the proposed method with respect to the various dictionary sizes on each database. The final results are shown in Fig. 6, with the increase of the dictionary size, the time of the dictionary learning gradually increases.

4.3 Size of variation dictionary

In this subsection, we will discuss the impact of variation dictionary size on recognition perfromance. For instance, how many subjects are suitable in the gallery set and how to create a suitable variational dictionary. The experimental setup is the same as the those in the previous subsection. The results on the two databases (LFW, Georgia) are presented in Fig. 7. It can be seen that the recognition rates are unstable

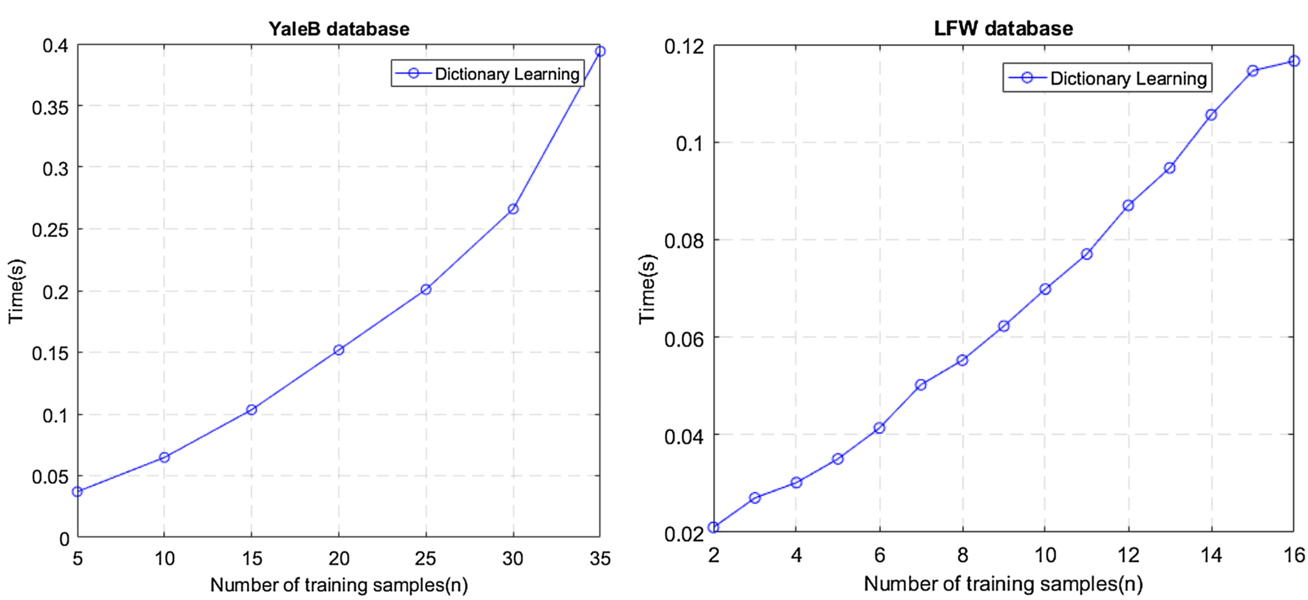

Fig. 6 Illustration of the time cost of the KSVD dictionary learning on YaleB and LFW database

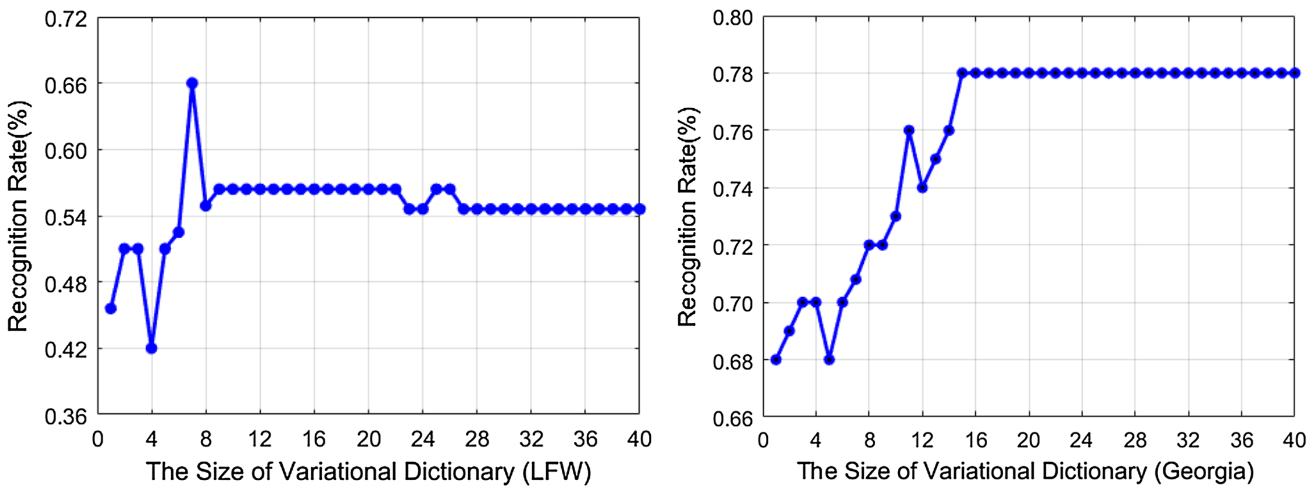

Fig. 7 Illustration how the dictionary size effects on the recognition rate

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 dictionary size is larger than 8 and 16 on LFW and Georgia databases, respectively. In summary, the size of the dictionary should be large enough in order to achieve better performance.

One can see from Fig. 7 that the recognition rates are stable when the size of the training set is more than 16 on these databases.

4.4 Evaluation on different databases

4.4.1 YaleB database

The YaleB database has 2432 images of 38 adults, 64 images per person. The face images have variations with respect to facial expressions (as normal, sad, happy, sleepy, surprised, and winking) and illuminations. In our experiment, the first 30 subjects are employed to build the gallery set and probe set, the remainder 8 subjects are employed to build the variational dictionary training set. We use half of samples per person for training and set three cases for number of probe samples. In the first case, 30 samples per person of the probe samples are used. In the second case, we choose randomly 20 probe samples to evaluate the performance of the related methods, and in the last case, 10 samples are selected randomly. Both of the case are compared with JSRC, SRC, SILT, SVDL, ESRC method. The recognition rates of the 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 is better than all other related methods on this database. In particular, the proposed KSVD-IJSR method possesses the excellent performance in the case with 30 probe images per person, which achieves the best recognition rates 100%, while the recognition rates of the other methods are less than KSVD-IJSR method.

4.4.2 CMU-PIE database

The CMU-PIE database is employed to evaluate robustness of the proposed method for variational illumination. 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. In this subsection, we select 43 images per person from the camera c05, which contains 68 subjects and can be seen from the Fig. 4, each image is cropped to 32×32 and it only contains the illumination changes. The first 50 subjects are used to build the gallery set and probe set, the remainder 18 subjects are used to build the variational dictionary training set. The 50% images per person are selected as the gallery set, and the rest images are for the probe set, which are randomly selected k ($k = 20, 10$) images per person. For each k, the experiments are repeated for 10 times to obtain the average results. As shown in Fig. 8, the proposed KSVD-IJSR method outperforms other related methods regardless of the number of the probe images.

4.4.3 Georgia database

The Georgia database contains 650 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 a lot of changes in illumination, expression and pose. All the images are manually cropped from the original ones based on the locations of the two eyes, and each cropped image is resized to the size

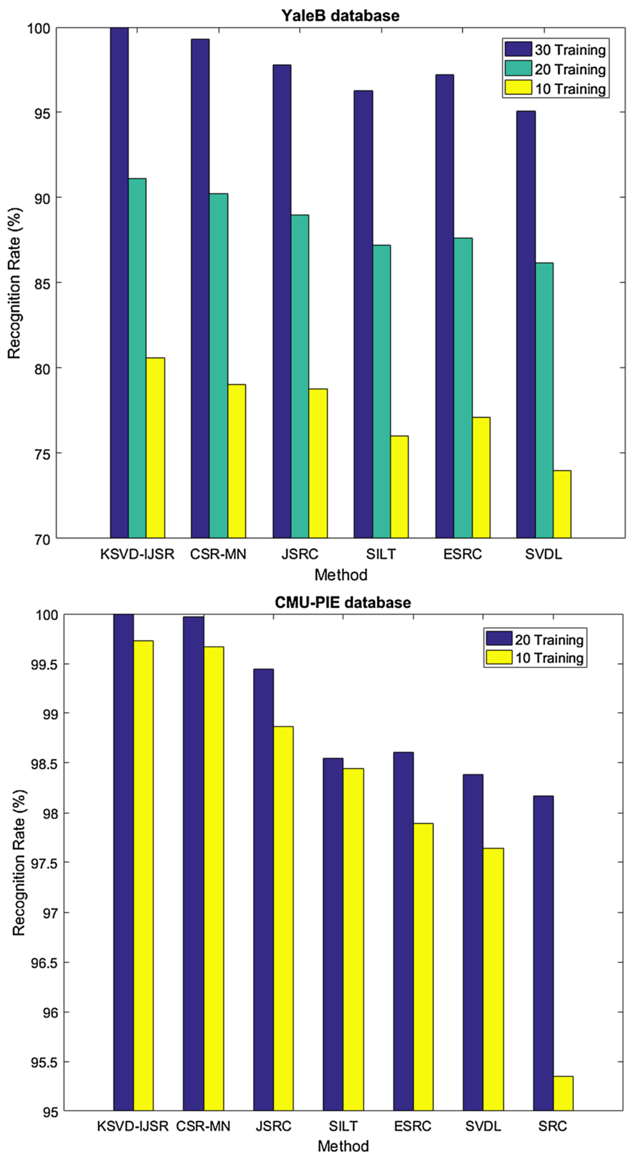

Fig. 8 The recognition rates on the YaleB and CMU-PIE database

of 32×32 pixels, and the first 35 subjects are taken to build the gallery and probe set, the rest are for the training set. For each person, 5 samples are selected for the gallery set, and three cases with 5, 10 probe samples are considered. The experimental results are shown in Fig. 9, and it can see that the proposed method has achieved the highest recognition rates than all other methods in all cases. In particular, for the case with 10 probe samples per person, KSVD-IJSR has an obvious advantage than other methods.

4.4.4 LFW database

The LFW database contains 5749 images of different individuals captured in unconstrained environments. It contains 85 individuals and each person has 23 different images. The

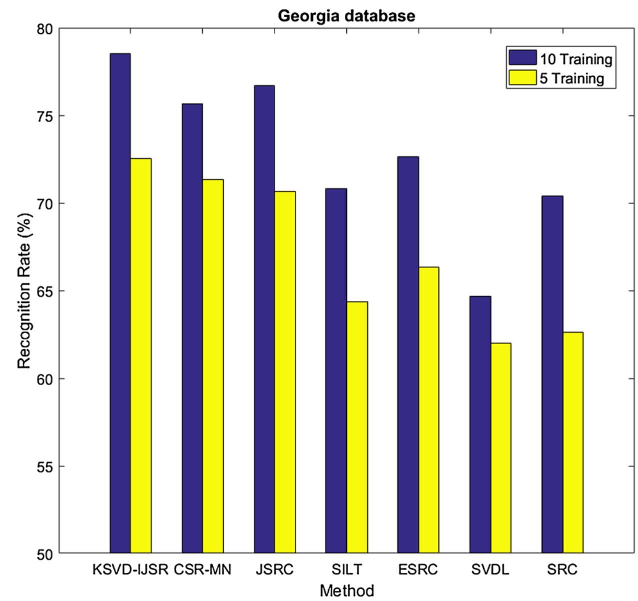

Fig. 9 The recognition rates on the Georgia database

variances in illumination, pose, occlusion, and expression among these images make SSFR face recognition extremely challenging. In this subsection, 16 frontal images per person are selected for our experiments, and the rest are for the training set. The 50% images per person are selected as the gallery set, and the rest images are for the probe set. Firstly, 8 probe samples per person are randomly and the average values are calculated, in Fig. 10, the proposed method achieves the highest recognition rate 56.47% as well as those obtained by the related methods. Secondly, in the case with 6 probe samples per person, KSVD-IJSR method also gets better results than the related methods, except for the SRC 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.

4.4.5 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, 120 individuals (65 men and 55 women) with 12 images are selected and converted into grayscaling images for recognition in case of occlusions testing, so the last 20 subjects are for the training set, the rest for the gallery and probe set. We choose the 6 images (4 front images, 1 glass image and 1 scarf image) for the gallery and the rest images for probe. We randomly choose k ($k = 4, 6$) images per person

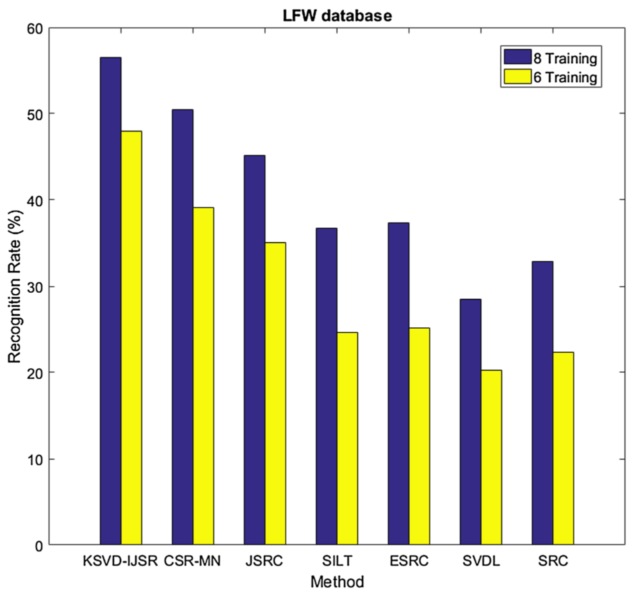

Fig. 10 The recognition rates on the LFW database

as the probe samples. The experimental results about different probe number are shown in Fig. 11. We can see that all the selected methods achieve the state-of-the-art results for the experiments with occlusions. However, our method achieves the best recognition rates. Considering these factors, our method is the best one of all approaches for the experiments on the AR database.

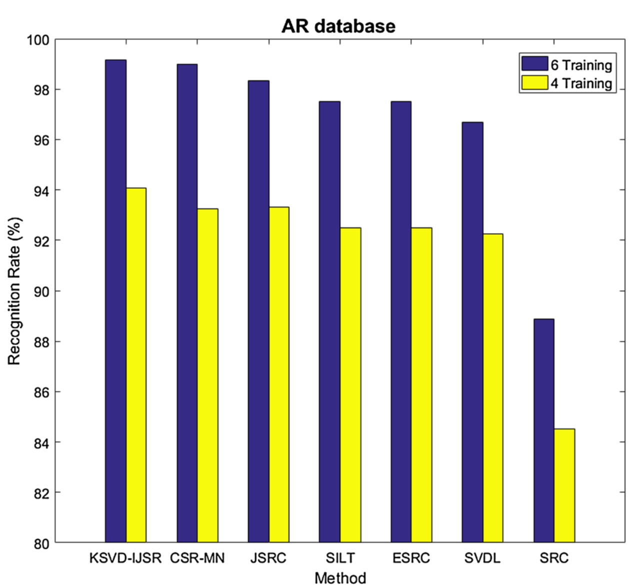

Fig. 11 The recognition rates on the AR database

4.4.6 Comparison with related dictionary learning methods

In this subsection, the performance of different dictionary learning methods for the same face recognition problem is discussed. For fair comparison, the gallery set needs to be as sufficient large as possible. In fact, the training dictionary set is smaller than the gallery set for stable recognition rate. Some related methods about dictionary learning for the SSFR problem, such as SILT, SVDL, ESRC with the basic variation dictionary (BVD), are selected for comparison with our proposed KSVD-IJSR method on several databases (LFW, Georgia, AR, YaleB and CMU-PIE databases). For each database, the 50% samples for per person are selected for the gallery set, all the rest samples are selected for the probe set.

The experimental results are shown in Table 2. For each database, our method achieves the best results among these dictionary learning methods. Although training dictionary contains more related variation feature information and can affect the recognition, the learned gallery and probe dictionaries by the K-SVD will have little information of face images and avoids naturally the interference of the extra information, such as noise and residue errors, which has negative influence for the recognition results. So the performance of the KSCD-IJSR is better than others in these experiments.

5 Conclusion

In this paper, a novel K-SVD dictionary-based face recognition approach with the improved joint sparse representation (KSVD-IJSR) have been proposed for image set based face recognition. The structure of proposed KSVD-IJSR has three phases. In the first phase, the variation dictionary

Table 2 The recognition rates about the different dictionary learning methods on several different database

Training Set Method	LFW	Georgia	AR	YaleB	CMU-PIE
KSVD-IJSR	**56.47**	**78. 54**	**99. 17**	**100**	**100**
CSR-MN	52.51	75.67	98.25	99.28	99.97
JSRC	50.00	76.67	98.33	97.77	99.44
SILT	36.71	70.80	97.50	96.25	98.55
ESRC	37.34	72.62	97.50	97.20	98.61
SVDL	28.48	64.66	96.67	90.10	98.38
SRC	32.91	70.40	88.88	87.91	98.17

The number represents facial recognition accuracy (max 100%). The bold number is the highest accuracy achieved among all the compared techniques. Our proposed approach is the one achieving highest accuracy is learned directly from the training set by K-SVD, which aims to extract more discriminative information from the variation gallery and probe subjects, such as the variation features (pose, illumination and expression). In the second phase, the ℓ_1 is embedded into the K-SVD to solve the optimization problem, the experiment results show a significant speed improvement per iteration. In the third phase, the improved joint sparse representation (IJSR) is presented, which utilizes the dictionary information from both the gallery and probe samples to classify the probe samples. The IJSR model does not only takes advantage of the K-SVD dictionary learning, but also utilizes the group structure to enhance the recognition performance. The computational time and size of the dictionary are discussed in the experiments. Extensive numerical experiments are implemented on five databases to verify the performance of KSVD-IJSR for set-to-set face recognition under complication conditions, and experimental results show our approach outperforms the related methods in terms of accuracy, computational costs and robustness.

Acknowledgements This work was partially supported by the National Natural Science Foundation of China with Grant Nos. 61671285 and 61363066.

References

1. Zhao W, Chellappa R, Phillips PJ, Rosenfeld A (2003) Face recognition: a literature survey. ACM Comput Surv (CSUR) 35(4):399–458
2. Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Chang J, Hoffman K, Marques J, Min J, Worek W (2005) Overview of the face recognition grand challenge. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol 1. IEEE, pp. 947–954
3. Tan X, Chen S, Zhou Z-H, Zhang F (2006) Face recognition from a single image per person: A survey. Pattern Recogn 39(9):1725–1745
4. Gao S, Tsang IW-H, Chia L-T (2010) Kernel sparse representation for image classification and face recognition. In: European Conference on Computer Vision, Springer, New York, pp 1–14
5. Naseem I, Togneri R, Bennamoun M (2010) Linear regression for face recognition. IEEE Trans Pattern Anal Mach Intell 32(11):2106–2112
6. Zhang L, Yang M, Feng X (2011) Sparse representation or collaborative representation: which helps face recognition? In: 2011 International Conference on Computer Vision, IEEE, pp 471–478
7. Peng X, Zhang L, Yi Z, Tan KK (2014) Learning locality-constrained collaborative representation for robust face recognition. Pattern Recogn 47(9):2794–2806
8. Hadid A, Pietikainen M (2004) From still image to video-based face recognition: an experimental analysis. In: Automatic Face and Gesture Recognition. Proceedings Sixth IEEE International Conference on IEEE, pp 813–818
9. Arandjelovic O, Shakhnarovich G, Fisher J, Cipolla R, Darrell T (2005) Face recognition with image sets using manifold density divergence. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol 1. IEEE, pp 581–588
10. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009) Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 31(2):210–227
11. Wang X, Yang M, Shen L (2016) Structured regularized robust coding for face recognition. Neurocomputing 216:18–27
12. Mi J-X, Liu T (2016) Multi-step linear representation-based classification for face recognition. IET Comput Vision 10(8):836–841
13. Chen S, Liu J, Zhou Z-H (2004) Making flda applicable to face recognition with one sample per person. Pattern Recogn 37(7):1553–1555
14. Gao Q-X, Zhang L, Zhang D (2008) Face recognition using flda with single training image per person. Appl Math Comput 205(2):726–734
15. Su Y, Shan S, Chen X, Gao W (2010) Adaptive generic learning for face recognition from a single sample per person. CVPR 2010:2699–2706
16. Majumdar A, Ward RK (2008) Pseudo-fisherface method for single image per person face recognition. In: 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, pp 989–992
17. Deng W, Hu J, Guo J (2012) Extended src: undersampled face recognition via intraclass variant dictionary. IEEE Trans Pattern Anal Mach Intell 34(9):1864–1870
18. Deng W, Hu J, Guo J (2013) In defense of sparsity based face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 399–406
19. Yang M, Van Gool L, Zhang L (2013) Sparse variation dictionary learning for face recognition with a single training sample per person. In: Proceedings of the IEEE international conference on computer vision, pp 689–696
20. Zhuang L, Chan T-H, Yang AY, Sastry SS, Ma Y (2015) Sparse illumination learning and transfer for single-sample face recognition with image corruption and misalignment. Int J Comput Vision 114(2–3):272–287
21. Ding R-X, Du DK, Huang Z-H, Li Z-M, Shang K (2015) Variational feature representation-based classification for face recognition with single sample per person. J Vis Commun Image Represent 30:35–45
22. Li Z-M, Huang Z-H, Shang K A customized sparse representation model with mixed norm for undersampled face recognition
23. Lu J, Tan Y-P, Wang G, Yang G (2013) Image-to-set face recognition using locality repulsion projections and sparse reconstruction-based similarity measure. IEEE Trans Circuits Syst Video Technol 23(6):1070–1080
24. Wang R, Shan S, Chen X, Gao W (2008) Manifold-manifold distance with application to face recognition based on image set. In: Computer Vision and Pattern Recognition. CVPR IEEE Conference on IEEE, pp 1–8
25. Cevikalp H, Triggs B (2010) Face recognition based on image sets. In: Computer vision and pattern recognition (CVPR), IEEE Conference on IEEE, pp 2567–2573
26. Hu Y, Mian AS, Owens R (2011) Sparse approximated nearest points for image set classification. In: Computer vision and pattern recognition (CVPR), IEEE conference on IEEE, pp 121–128
27. Hu Y, Mian AS, Owens R (2012) Face recognition using sparse approximated nearest points between image sets. IEEE Trans Pattern Anal Mach Intell 34(10):1992–2004
28. Wang R, Guo H, Davis LS, Dai Q (2012) Covariance discriminative learning: a natural and efficient approach to image set classification. In: Computer vision and pattern recognition (CVPR), 2012 IEEE Conference on IEEE, pp 2496–2503
29. Huang L, Lu J, Tan Y-P (2014) Co-learned multi-view spectral clustering for face recognition based on image sets. IEEE Signal Process Lett 21(7):875–879

30. Kim T-K, Kittler J, Cipolla R (2007) Discriminative learning and recognition of image set classes using canonical correlations. IEEE Trans Pattern Anal Mach Intell 29(6):1005–1018
31. Zhang H, Nasrabadi NM, Zhang Y, Huang TS (2012) Joint dynamic sparse representation for multi-view face recognition. Pattern Recogn 45(4):1290–1298
32. Zhang X, Pham D-S, Venkatesh S, Liu W, Phung D (2015) Mixed-norm sparse representation for multi view face recognition. Pattern Recogn 48(9):2935–2946
33. Huang GB, Ramesh M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: a database for studying face recognition in unconstrained environments, Tech. rep., Technical Report 07-49, University of Massachusetts, Amherst
34. Kim S-J, Koh K, Lustig M, Boyd S, Gorinevsky D (2007) An interior-point method for large-scale-regularized least squares. IEEE J Sel Topics Sig Process 1(4):606–617
35. Nefian A (2013) Georgia tech face database
36. Chen Y-C, Patel VM, Phillips PJ, Chellappa R (2012) Dictionary-based face recognition from video. European conference on computer vision, Springer, New York, pp 766–779
37. Martinez AM, The ar face database, CVC Technical Report 24
38. Donoho D, Tsaig Y (2006) Fast solution of l1-norm minimization problems when the solution may be sparse, Preprint 1(2)
39. Yang M, Chang H, Luo W (2017) Discriminative analysis-synthesis dictionary learning for image classification. Neurocomputing 219:404–411
40. Li Z, Lai Z, Xu Y, Yang J, Zhang D (2017) A locality-constrained and label embedding dictionary learning algorithm for image classification. IEEE Trans Neural Netw Learn Syst 28(2):278–293
41. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720
42. Fan Z, Ni M, Zhu Q, Liu E (2015) Weighted sparse representation for face recognition. Neurocomputing 151:304–309
43. Zhu P, Zhang L, Hu Q, Shiu SC (2012) Multi-scale patch based collaborative representation for face recognition with margin distribution optimization. European Conference on Computer Vision, Springer, New York, pp 822–835
44. Huang S-M, Yang J-F (2013) Linear discriminant regression classification for face recognition. IEEE Signal Process Lett 20(1):91–94
45. Yang M, Zhang L, Yang J, Zhang D (2013) Regularized robust coding for face recognition. IEEE Trans Image Process 22(5):1753–1766
46. Lai J, Jiang X (2012) Modular weighted global sparse representation for robust face recognition. IEEE Signal Process Lett 19(9):571–574
47. Zhu P, Zuo W, Zhang L, Shiu SC-K, Zhang D (2014) Image set-based collaborative representation for face recognition. IEEE Trans Inf Forensics Secur 9(7):1120–1132
48. Kan M, Shan S, Su Y, Chen X, Gao W (2011) Adaptive discriminant analysis for face recognition from single sample per person, In: Automatic Face and Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on IEEE, pp. 193–199
49. Kan M, Shan S, Su Y, Xu D, Chen X (2013) Adaptive discriminant learning for face recognition. Pattern Recogn 46(9):2497–2509
50. Gao S, Jia K, Zhuang L, Ma Y (2015) Neither global nor local: regularized patch-based representation for single sample per person face recognition. Int J Comput Vision 111(3):365–383
51. Jin Y, Lu J, Ruan Q (2015) Coupled discriminative feature learning for heterogeneous face recognition. IEEE Trans Inf Forensics Secur 10(3):640–652
52. Aharon M, Elad M, Bruckstein A (2006) *rmk*-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 54(11):4311–4322
53. Shekhar S, Patel VM, Nasrabadi NM, Chellappa R (2014) Joint sparse representation for robust multimodal biometrics recognition. IEEE Trans Pattern Anal Mach Intell 36(1):113–126
54. Chen C, He B, Ye Y, Yuan X (2016) The direct extension of admm for multi-block convex minimization problems is not necessarily convergent. Math Program 155(1–2):57–79
55. Martinez A, Benavente R (1998) The ar face database, 1998. Computer Vision Center, Technical Report, p 3
56. Sim T, Baker S, Bsat M (2002) The cmu pose, illumination, and expression (pie) database. In: Automatic Face and Gesture Recognition. Proceedings Fifth IEEE International Conference on IEEE, pp. 46–51
57. Rakotomamonjy A (2011) Surveying and comparing simultaneous sparse approximation (or group-lasso) algorithms. Sig Process 91(7):1505–1526
58. Li Z-M, Huang Z-H, Shang K (2016) A customized sparse representation model with mixed norm for undersampled face recognition. IEEE Trans Inf Forensics Secur 11(10):2203–2214