Pattern Recognition 93 (2019) 283–292

Contents lists available at ScienceDirect

Pattern Recognition

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

Multi-resolution dictionary learning for face recognition

Xiaoling Luo[a], Yong Xu[a,c,*], Jian Yang[b]

[a] Bio-Computing Research Center, Harbin Institute of Technology, Shenzhen 518055, China
[b] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China
[c] Peng Cheng Laboratory, Shenzhen 518055, China

ARTICLE INFO

Article history:
Received 10 September 2018
Revised 22 March 2019
Accepted 24 April 2019
Available online 25 April 2019

Keywords:
Dictionary learning
Multi-resolution
Face recognition

ABSTRACT

In recent years, there has been a growing interest in the study of dictionary learning for face recognition. Most of the conventional dictionary learning methods focus only on a single resolution, which ignores the variability of resolutions of real-world face images. In order to address the above issue, this paper proposes a novel multi-resolution dictionary learning method that provides multiple dictionaries each being associated with a resolution. Especially, to enhance the robustness of the model, our method adds a relatively strong constraint to keep the similarity of representations obtained using different dictionaries in the training phase. We compare the proposed method to several state-of-the-art dictionary learning methods by applying this method to multi-resolution face recognition. The experimental results demonstrate that our method outperforms many recently proposed dictionary learning methods. The MATLAB codes of the proposed method will be available at http://www.yongxu.org/lunwen.html.

1. Introduction

Dictionary learning, as a vital branch of sparse representation, has been widely used in various fields of image processing [1,2], such as image denoising [3], super-resolution imaging [4], image recognition [5] and object detection [6]. For face recognition, the conventional dictionary learning method [7,8] mainly contains two steps. The first step is the representation learning. Given an image, a dictionary learning model exploits some or all training images to represent this image based on a learned dictionary. And the second step is to classify the test samples according to the representation results.

Learning a dictionary from the training set for sparse coding or feature representation has achieved significant improvement in image classification and face recognition. Various dictionary learning methods have been devised. Due to the difficulty of completely covering the important components of signal representation (e.g. commonality, particularity, and disturbance), Lin et al. [8] proposed a novel robust, discriminative and comprehensive dictionary learning method (RDCDL). They first trained a robust dictionary on comprehensive training sample diversities which were extracted or generated from facial variations. Then class-shared, class-specific and disturbance dictionary atoms were learned to extract features from different classes. Finally, discriminative regularizations and the representation coefficients were used to improve the classification capability of the dictionary. To alleviate the limitation that most existing dictionary learning methods only learn a linear dictionary, Hu et al. [9] proposed a nonlinear dictionary learning (NDL) method, in which a feed-forward neural network was employed to seek hierarchical feature projection matrices and a dictionary simultaneously. Then, a class-specific dictionary was obtained to exploit the discriminative information. Recently, the joint dictionary learning algorithm has been wildly used in super-resolution imaging [4] and multispectral change detection [10]. Yang et al. [4] sought a sparse representation for image patches of the low-resolution input, and then utilized the coefficients of this representation to generate high-resolution outputs. Finally, the jointly trained dictionaries for the low-resolution and high-resolution image patches obtained a more compact representation and reduced the computational cost substantially. Lu et al. [10] designed a joint dictionary composed of two coupled dictionaries, which could provide adequate descriptive power for bitemporal multispectral images. Joint dictionary learning which has the capability to combine low-resolution and high-resolution feature spaces and mining the inner associations between dictionaries of different resolutions, provides the theoretical basis for our proposed method.

Recent years we have witnessed an increase in the use of deep learning in various research domains especially the field of image and video analysis. Deep learning has received much attention in the field of image analysis owing to its outstanding performance in extracting discriminative features from samples [11,12].

[*] Corresponding author.
E-mail addresses: luoxiaoling@stu.hit.edu.cn (X. Luo), yongxu@ymail.com, laterfall@hit.edu.cn (Y. Xu), csjyang@mail.njust.edu.cn (J. Yang).

https://doi.org/10.1016/j.patcog.2019.04.027
0031-3203/© 2019 Elsevier Ltd. All rights reserved.

In this branch, deep learning for face recognition is one of the most representative works. For example, Sun et al. [13] proposed a hybrid convolutional network (ConvNet)-Restricted Boltzmann Machine (RBM) model for face recognition. To learn rich identity similarity information, this model concatenates the features of different face region pairs extracted by different deep ConvNets. However, it needs a large-scale training set to obtain enough features. In addition, because of its complex network structure, it consumes a significant amount of computing resources to keep running. Compared with the methods based on deep learning, conventional methods are more suitable for the tasks with small-scale databases and are more efficient in computation. In this work, we mainly focus on conventional methods for face recognition.

Although conventional methods are promising, they still have some disadvantages. For example, for face recognition, it is difficult for a conventional dictionary learning method to obtain a reliable and robust dictionary due to the small-sample-size problem. The other significant issue is that joint dictionary learning algorithms can only generate high-resolution face images from low-resolution face images by super-resolution imaging [4,10], rather than obtain the dictionary for multi-resolution face recognition. To address the above problems, we propose a novel robust multi-resolution dictionary learning method.

We propose a multi-resolution dictionary learning method, which uses multi-resolution images to get a robust dictionary. To demonstrate the robustness of the proposed method, we trained our model on several multi-resolutions face recognition datasets, which was constructed by the resolution pyramid method. For each resolution, we learn one corresponding dictionary, and we concatenate them with proper normalization. Besides, we combine the coefficient matrix of different dictionaries to form features of different resolutions. Experimental results in face recognition tasks show that our proposed method achieves substantial improvements in comparison with other dictionary learning methods.

2. Related works

In this section, we briefly introduce some related algorithms. For convenience, we first roughly divide the dictionary learning algorithms into three types: supervised dictionary learning algorithms, semi-supervised dictionary learning algorithms, and unsupervised dictionary learning algorithms.

2.1. Supervised dictionary learning algorithms

Supervised dictionary learning algorithms aim at minimizing the reconstruction error of training samples based on the constraint on labels. As it can make full use of the potential classification information in the training samples, it is able to reconstruct the original data excellently. Mairal et al. [14] added category labels into supervised dictionary learning for extracting the information implicit in the data. In order to improve the pattern classification performance, the Fisher discrimination criterion was used to learn a structured dictionary, whose dictionary atoms had correspondence to the class labels [15]. And a Fisher discrimination dictionary learning (FDDL) model based on the Fisher discrimination criterion was proposed in [16]. Moreover, Wang et al. [17] proposed a discriminative dictionary learning method for image classification and found that global coding classifier (GC) was more effective when the number of training samples of each class was relatively small, or the learned class-specific dictionaries were small in size. Zhang et al. [18] proposed the discriminative K-SVD (D-KSVD) method based on extending K-SVD by appending the classification error into the objective function. Jiang et al. [19] proposed a label consistent K-SVD (LC-KSVD), which combined discriminative sparse-code error with the reconstruction error and the classification error to form an integral objective function. Cai et al. [20] proposed a support vector guided dictionary learning (SVGDL) model. Compared with FDDL, SVGDL can automatically allocate suitable weights to coding vector pairs and adaptively select only a few essential pairs to allocate non-zero weights.

The supervised dictionary learning method introduces the supervision information of training data into the dictionary learning process, and the learned dictionary has inherent advantages in data classification. However, the practical problem is that large-scale labeled sample data sets are difficult to obtain, which limits the development of supervised dictionary learning.

2.2. Semi-supervised dictionary learning algorithms

Compared with labeled data, unlabeled data are more easily available and more numerous. Such advantages have promoted researchers to design semi-supervised dictionary learning algorithms that use both unlabeled data and labeled data to generate better representations for classification tasks. As the performance of dictionary learning is mostly limited by the scale of training samples, Shrivastava et al. [21] proposed a discriminative dictionary learning algorithm, which exploited both labeled and unlabeled data to address the disadvantages of insufficient samples. However, it ignores the preservation of the local structure, which may affect the classification accuracy. To adequately address this problem, Behnam et al. [22] introduced a semi-supervised dictionary learning algorithm with a probabilistic framework, which used the geometric characteristics of the marker. In addition, Wang et al. [23] proposed a semi-supervised robust dictionary learning model to automatically optimize the dictionary size and fix the problem of the sensitivity to noisy and outlier samples. Moreover, Jian and Jung [24] proposed a semi-supervised bi-dictionary learning algorithm for image classification with smooth representation-based label propagation (SRLP). However, a limitation of the semi-supervised dictionary learning algorithm is its sensitivity to labeled samples.

2.3. Unsupervised dictionary learning algorithms

Unsupervised dictionary learning algorithm aims at minimizing the reconstruction error of training samples, and it learns the dictionary without the class information of training samples. The optimization goal of unsupervised dictionary learning algorithms mainly focuses on the reconstruction of original samples and the sparsity of coding. Making the algorithm generate sparse coding in the learned dictionary can benefit to the reconstruction of the original signal. One of the most well-known unsupervised dictionary learning algorithms is K-SVD [25], which updates the dictionary atom by atom until satisfying the sparsity condition. As Wang et al. [26] thought that locality was more essential than sparsity, they proposed a locality-constrained linear coding algorithm (LLC) by using a locality constraint. This method selected similar basis of local image descriptors from sparse representations and obtained a linear combination weight of these basis to reconstruct each descriptor. Moreover, Jenatton et al. [27] designed an algorithm to obtain dictionaries embedded in a hierarchy by using a tree-structured sparsity. According to [28], obtaining multiple dictionaries is a feasible way to improve the performance of a classification task. We see that even a simple joint double dictionary learning procedure can also achieve impressive performance [29].

Our proposed method is close to joint dictionary learning algorithms [4,10], with some differences: on the one hand, as an unsupervised dictionary learning algorithm, it aims at minimizing the reconstruction error of multi-resolution training samples; on the other hand, it utilizes more than two dictionaries to represent and classify the multi-resolution face images, which directly solve the problem of multi-resolution face images recognition.

3. Proposed method

A basic problem in face recognition is to use a model learned from training samples to correctly determine the class of a test sample. We implement this process with a multi-resolution dictionary learning model. Assume that we have the training set $Y = [Y_1, \dots, Y_i, \dots, Y_k]$ with k resolutions, where Y_i represents the *ith* subset of Y. Y_i contains N training samples, in which all images have the same resolution. k is the number of resolutions. Specifically, $Y_i = [y_i^1, \dots, y_i^s, \dots, y_i^N] \in \mathbb{R}^{n\times N}$ is a matrix, and y_i^s is corresponding to the training sample s under the *ith* resolution. Let $D = [D_1, \dots, D_k]$ be the dictionary matrix of k resolutions, and $D_i = [d_i^1, \dots, d_i^m] \in \mathbb{R}^{n\times m}$. $X = [x_1, \dots, x_s, \dots, x_N] \in \mathbb{R}^{m\times N}$ is the coding coefficient matrix, where x_s represents the coefficient of the *sth* sample. In addition, we assume the training set of each resolution contains all categories.

3.1. Motivations and model of multi-resolution dictionary learning

Intuitively, images captured by different cameras vary in resolutions. Therefore, when the resolution of the image is uncertain, we are encouraged to train multiple resolution images to ensure that this algorithm is capable of adapting various resolutions. We design a robust and flexible dictionary learning model to effectively solve the multi-resolution face image recognition problem. This model produces multiple dictionaries from multiple resolution training samples, and it enforces that these dictionaries are associated with the same coefficient matrix. All sub training sets have the same size and structure. Here, the same structure means that the *sth* $(s = 1, \dots, N)$ item in the training set for each resolution has the same label. Moreover, we expect to learn the multiple dictionaries simultaneously, such that the representations of all resolutions can be integrated into a framework to promote the learning of dictionaries. To this end, we propose the following dictionary learning model,

$$\langle D_1, \dots, D_k, X\rangle = \arg\min_{D_1,\dots,D_k,X} \|Y_1 - D_1X\|_F^2 + \dots + \|Y_k - D_kX\|_F^2 + \beta\ \|X\|_F^2 \qquad (1)$$

where the regularization parameter β can be tuned using a validation set. The $\beta\ \|X\|_F^2$ regularization is designed to smooth the decision boundary and alleviate the problem of local minimum and overfitting. Moreover, $\beta\ \|X\|_F^2$ is helpful for changing the algorithm framework from ill-condition to well-condition when β is not equal to 0.

Besides, we minimize the reconstruction error between the training set and the reconstructed set to ensure that the algorithm is adaptive to multi-resolution training samples. Thus, we can obtain a proper representation for each sample in the training set, and a more robust dictionary for each resolution. Particularly, the obtained dictionaries are also suitable for representing images with unknown resolutions [30].

Our solution mainly focuses on the minimization of the reconstruction error, and we minimize the expression in Eq. (1) iteratively. First, we fix $D_1, \dots, D_k$ and calculate the best coefficient matrix X. Then, we update the dictionaries $D_1, \dots, D_k$ respectively, by fixing coefficient matrix X. For example, when we update D_1 according to the objective function we will fix $D_2, \dots, D_k$ and X. By optimizing each variable iteratively, we can obtain the optimal solution on the face recognition task. The framework of the method is illustrated in Fig. 1.

In this paper, we use resolution pyramid methods to generate multi-resolution training samples and apply them to the proposed algorithm. In the following section, we describe the solution for the objective function in detail. During training, X will be computed first by Eq. (2) using initialized $D_1, \dots, D_k$, and then $D_1, \dots, D_k$ are updated using Eq. (3). X and $D_1, \dots, D_k$ are updated alternately in the above way until the end condition is satisfied.

3.2. Solution of multi-resolution dictionary learning model

Generally, there are two different ways to learn and optimize the dictionary. On the one hand, like K-SVD, the algorithm updates dictionary atom by atom. On the other hand, in some algorithms, such as MOD [31], the learned dictionary is updated as a whole. In our algorithm, we update the whole dictionary at once. When we update one variable, we fix the others. As a result, the objective function of our proposed algorithm framework has an approximately optimal solution. We present the procedure to obtain the solution of objective function as follows.

We should first initialize the dictionaries $D_1, \dots, D_i, \dots, D_k$. In this paper, we initialize dictionaries one by one. For the training samples with the *ith* resolution, we employ PCA for each class and then concatenate all the outputs (i.e., dictionary atoms learned from each class) to form the initialized D_i. We update coefficient matrix X using Eq. (2), if we fix $D_1, \dots, D_k$.

$$X = \left(D_1^TD_1 + \ldots + D_k^TD_k + \beta I\right)^{-1}\left(D_1^TY_1 + \ldots + D_k^TY_k\right) \qquad (2)$$

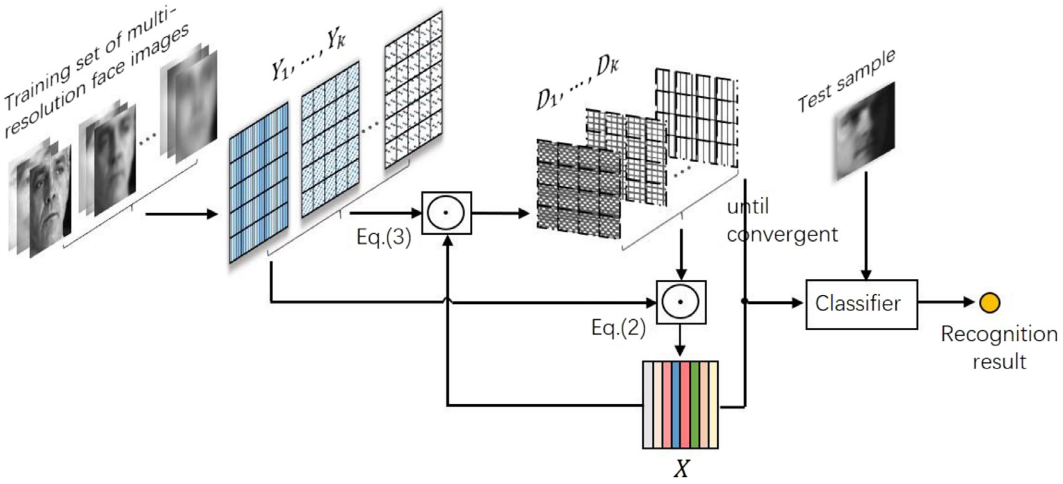

Fig. 1. The basic framework of the proposed method.

Algorithm 1
Algorithm of the proposed method.

```
Task: Find the best dictionaries to represent the data samples $\boldsymbol{y_i^s}$ ($\boldsymbol{i} = 1, \ldots, \boldsymbol{k}, \boldsymbol{s} = 1, \ldots, \boldsymbol{N}$), by solving
        $\langle \boldsymbol{D}_1, \ldots, \boldsymbol{D_k}, \boldsymbol{X} \rangle = \arg\min_{\boldsymbol{D}_1,\ldots,\boldsymbol{D_k},\boldsymbol{X}} \|\boldsymbol{Y}_1 - \boldsymbol{D}_1\boldsymbol{X}\|_F^2 + \cdots + \|\boldsymbol{Y_k} - \boldsymbol{D_k}\boldsymbol{X}\|_F^2 + \boldsymbol{\beta}\|\boldsymbol{X}\|_F^2$.
Input: $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y_k}$,   $\boldsymbol{\beta}$
Initialization: Initialize the dictionaries $\boldsymbol{D}_1, \ldots, \boldsymbol{D_k}$ for multi-resolution training sets using PCA.
Do until convergence:
Update coefficient matrix $\boldsymbol{X}$:
        $\boldsymbol{X} = (\boldsymbol{D}_1^T\boldsymbol{D}_1 + \ldots + \boldsymbol{D}_k^T\boldsymbol{D_k} + \boldsymbol{\beta I})^{-1}(\boldsymbol{D}_1^T\boldsymbol{Y}_1 + \ldots + \boldsymbol{D}_k^T\boldsymbol{Y_k})$.
Update dictionaries: For dictionary   $\boldsymbol{i}$, update it by
        $\boldsymbol{D_i} = (\boldsymbol{Y_i}\boldsymbol{X}^T)(\boldsymbol{X}\boldsymbol{X}^T)^{-1}$.
End do
Output: $\boldsymbol{D}_1, \ldots, \boldsymbol{D_k}$,   $\boldsymbol{X}$
```

β is a positive constant much less than 1 even close to zero, and I is an identity matrix.

We can obtain dictionaries $D_1, \ldots,\ D_i, \ldots, D_k$, if we fix coefficient matrix X. When updating D_i, all the other dictionaries are fixed. Then the formula of optimizing D_i can be written as:

$$D_i = \left(Y_i X^T\right)\left(XX^T\right)^{-1} \tag{3}$$

From these two steps above, the dictionaries $D_1, \ldots, D_k$ and coefficient matrix X can be updated alternately, and we should repeat these steps until convergence. Finally, we will obtain the optimal $D_1, \ldots, D_k$, and X. Algorithm 1 gives a more detailed description of these steps.

3.3. The classification procedure

When $D_1, \ldots, D_k$ are available, a testing sample can be classified via coding it over these dictionaries. Based on the employed dictionaries, the difference of coefficients can be utilized for the classification task. We can find coefficients for a given testing sample y by solving the following problem.

$$x_{test\_1} = \left(D_1^T D_1\right)^{-1} D_1^T y,\ \ \ldots,\ \ x_{test\_k} = \left(D_k^T D_k\right)^{-1} D_k^T y. \tag{4}$$

The difference between y and the *sth* training sample is defining by

$$dist_s = \|x_{test\_1} - x_s\| + \cdots + \|x_{test\_k} - x_s\| \tag{5}$$

where $x_{test\_1}, \ldots, x_{test\_k}$ denote the coefficients of y and they are represented by dictionaries $D_1, \ldots, D_k$ respectively. In addition, x_s denotes the coefficient of the *sth* training sample on multi-resolution dictionary, which is the *sth* column of X.

In the classification task, the labels of all training examples are known. If the difference between y and the *rth* training sample is smallest, the *rth* training sample and y are considered to belong to the same category. Based on Eq. (5), the classification function is:

$$R = \arg\min_s \{dist_s\}, s = 1 \ldots N \tag{6}$$

and then the label of the *Rth* training sample is assigned to y.

4. Experimental results and analysis

In order to well show the advantage of our proposed method, we compare it with conventional methods like K-SVD [25], D-KSVD [18], LC-KSVD [19], SRC [32], and DLSPC [17]. To further demonstrate the effectiveness of multi-dictionary learning, we also compare it with the methods based on deep learning. We evaluate our model by virtue of several public face recognition datasets, including the Extended Yale B face database [33], the ORL face database [34], the AR face database [35], the CMU PIE face database (PIE) [36] and the Labeled Faces in the Wild database (LFW) [37].

4.1. Experimental setting

In this section, we give the experimental details. First of all, to train the multi-resolution dictionary learning model, we construct several multi-resolution datasets. Specifically, we adopt the resolution pyramid method to reduce the resolution of the original samples. For instance, before training model by using the Extended Yale B face database (64 × 64 pixels), we first divide the whole dataset into a training set and a test set in the 1:1 ratio, and then we convert them to multi-resolution datasets respectively. For the training set, we reduce the resolution of all the images in the original training set to 32 × 32 pixels and 16 × 16 pixels respectively. For the test set, we split the test samples of each class into three equal pieces, and then we convert samples of two of them into 32 × 32 pixels and 16 × 16 pixels. In other words, the test samples have three different resolutions.

Similar to the converting process on the Extended Yale B face database, we build multi-resolution datasets on the other benchmark datasets. In this paper, the count of resolution types is 3, and the parameter β is set to 0.0001. PCA is using to initialize the multiple dictionaries in our method. We train distinct dictionaries for each resolution and integrate them through the learned coefficients matrix X. Therefore, the model can predict the category of test sample straightly, without considering the resolution of it. Besides, we repeatedly train our proposed algorithm and the benchmark methods 10 times for each database and then calculate the average recognition rates and the average computational time for classifying a test sample.

4.2. Compared with conventional methods

4.2.1. Experimental results on the Extended Yale B face database

The Extended Yale B face database is captured from 38 individuals under various lighting conditions, contains 2414 front face images. The resolutions of samples in the processed Extended Yale B face database are 64 × 64, 32 × 32 and 16 × 16. Examples with different resolutions are shown in Fig. 2. For each person, we randomly select 32 images to form the training set and use the remaining samples as the test set. Empirical results and average testing time are shown in Table 1.

Besides, to further verify the advantage of the proposed method, we compare it with the baseline methods with a various number of atoms (76, 114, ..., 266). Experimental results are shown in Fig. 3, and it is obvious to see that our method outperforms the others.

4.2.2. Experimental results on the ORL face database

The ORL dataset is captured from 40 persons, and each person provides ten face images. Some images are taken at different times, and some are different with varying lighting, facial expressions (opening or closing eyes, smiling or not smiling), and facial details (glasses or no glasses). The resolution of original images is

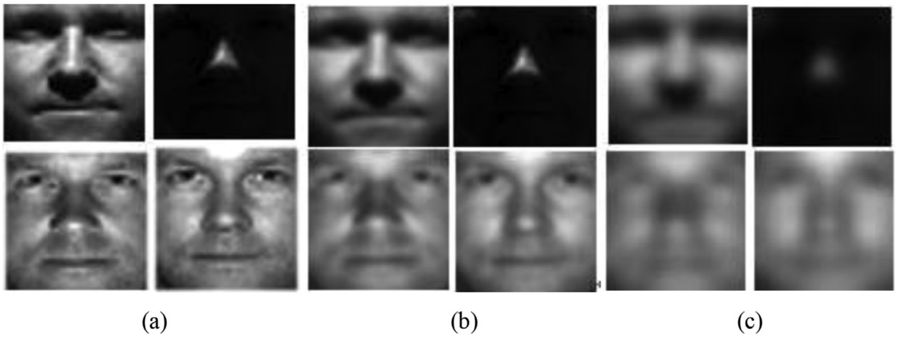

Fig. 2. Sample images from the Extended Yale B face database. (a) Sample images with size of 64 × 64 pixels; (b) Sample images with size of 32 × 32 pixels; (c) Sample images with size of 16 × 16 pixels.

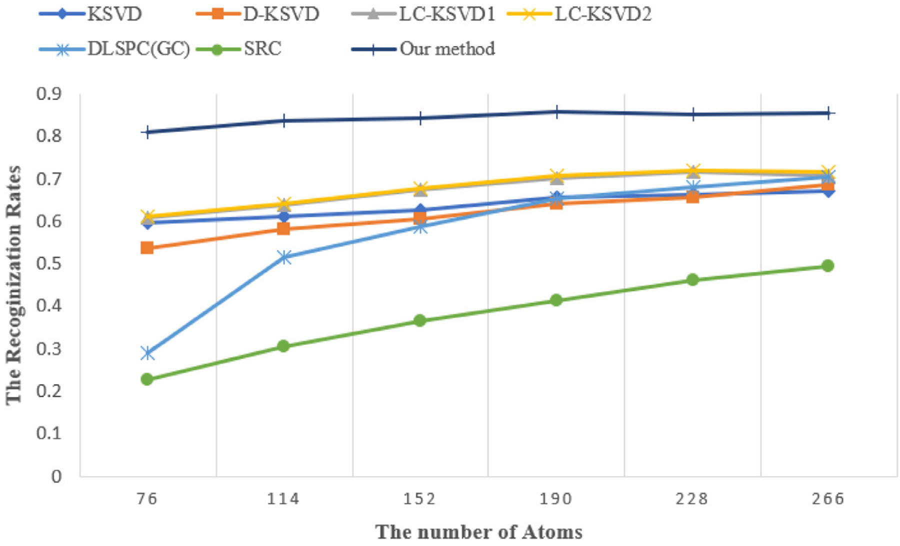

Fig. 3. The average recognition rates with different numbers of atoms on the Extended Yale B face database.

Table 1
The average recognition rates and computational time for classifying a test sample of different methods on the Extended Yale B face database.

Algorithm	Recognition rates (%)	Time(s)
KSVD [25]	72.69 ± 0.67	1e-5
D-KSVD [18]	74.88 ± 1.60	7e-6
LC-KSVD1 [19]	74.95 ± 1.44	1.4e-5
LC-KSVD2 [19]	76.70 ± 1.64	1.5e-5
DLSPC(GC) [17]	77.59 ± 0.78	0.048
SRC [32]	81.85 ± 1.12	2.590
Our method	88.58 ± 1.53	0.045

Table 2
The average recognition rates and computational time for classifying a test sample of different methods on the ORL face database.

Algorithm	Recognition rates (%)	Time(s)
KSVD [25]	89.15 ± 2.14	1e-5
D-KSVD [18]	87.45 ± 2.73	1.6e-5
LC-KSVD1 [19]	87.45 ± 1.83	8e-6
LC-KSVD2 [19]	88.85 ± 2.61	2.4e-5
DLSPC(GC) [17]	84.60 ± 2.98	0.013
SRC [32]	90.01 ± 1.96	0.073
Our method	92.15 ± 1.51	0.020

64 × 64 pixels, and we employ the resolution pyramid method to convert them into 32 × 32 and 16 × 16 images. Fig. 4 shows different resolution images converted from the ORL face database. We randomly select 5 face images of each person for training and the rest of them for testing. The average recognition rates and computational time of different methods are presented in Table 2.

We also compare the average recognition rates of KSVD, D-KSVD, LC-KSVD, DLSPC, and SRC with different numbers of atoms (80, 120, 160, 200). The experimental results are presented in Fig. 5.

4.2.3. Experimental results on the PIE face database

In this section, we evaluate our method on the PIE face database. The PIE face database contains the face images of 68 persons, and it captures each person's facial images in 13 different poses, 43 different illumination conditions and 4 different facial expressions. We choose face images from pose 05, which including 68 individuals and total 3332 images. Original images are normalized to the size of 64 × 64 pixels, and then they are converted into 32 × 32 pixels and 16 × 16 pixels. We show the samples of different resolutions in Fig. 6.

We randomly select 25 images of each person for training and the remaining samples for testing. The average recognition rates and the computational time are shown in Table 3. Experimental results of different recognition methods on PIE face database with different numbers of atoms (408, 476, ..., 748) are shown in Fig. 7.

4.2.4. Experimental results on the AR face database

The AR face database is composed of face images of 126 persons including 70 males and 56 Females, and each person has

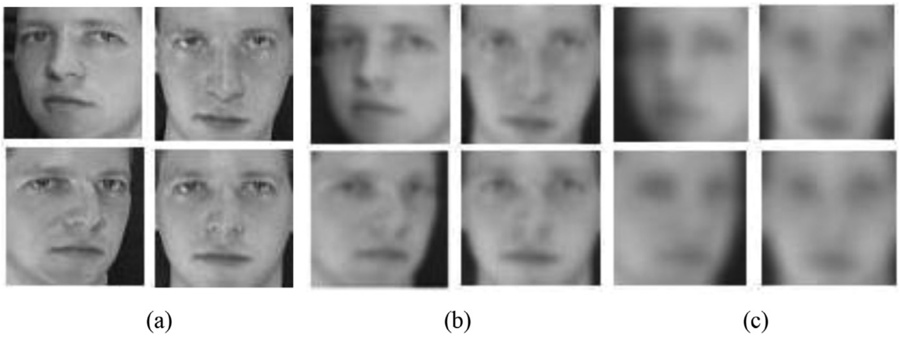

Fig. 4. Sample images from the ORL face database. (a) Sample images with size of 64 × 64 pixels; (b) Sample images with size of 32 × 32 pixels; (c) Sample images with size of 16 × 16 pixels.

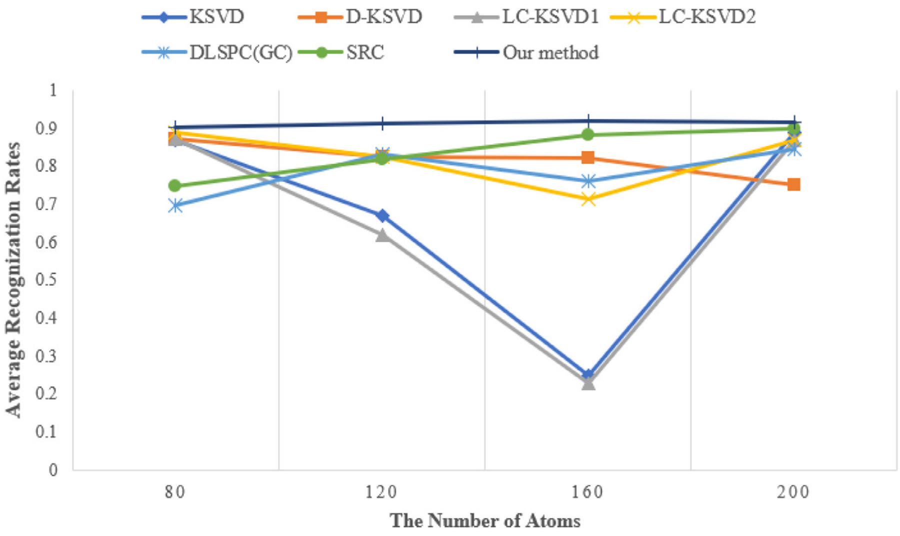

Fig. 5. The average recognition rates with different numbers of atoms on the ORL face database.

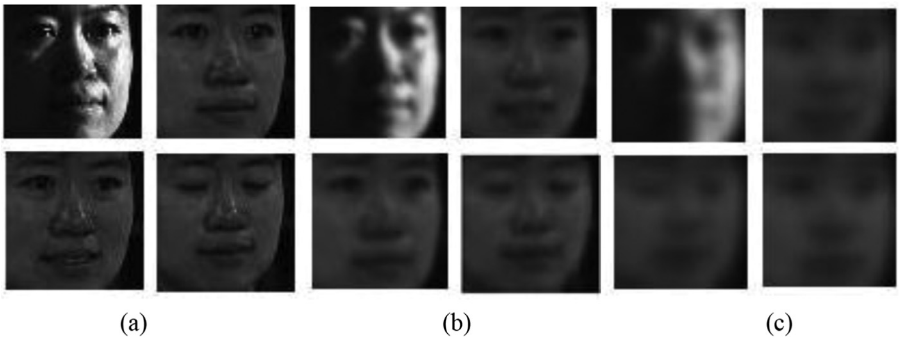

Fig. 6. Sample images from the PIE face database. (a) Sample images with size of 64 × 64 pixels; (b) Sample images with size of 32 × 32 pixels; (c) Sample images with size of 16 × 16 pixels.

26 face images. These images vary in illumination conditions (left light on, right light on or all side lights on), facial expression (neutral expression, smile, anger or scream) and dresses (wearing sunglasses or wearing a scarf). In our experiment, we use a subset of the database consisting of 120 persons and total 3120 images. For each person, we randomly select 13 images for training and the remaining ones for testing. Each original face image is resized to 50 × 40 pixels, and then use the resolution pyramid method to reduce the resolution of the images to 25 × 20 and 12 × 10 pixels. Fig. 8 show these three types of resolution images based on the AR face database.

Table 3
The average recognition rates and computational time for classifying a test sample of different methods on the PIE face database.

Algorithm	Recognition rates (%)	Time(s)
KSVD [25]	67.03 ± 1.02	1e-5
D-KSVD [18]	60.52 ± 1.07	1.5e-5
LC-KSVD1 [19]	67.02 ± 0.90	2.4e-6
LC-KSVD2 [19]	67.74 ± 0.85	1.9e-5
DLSPC(GC) [17]	63.31 ± 1.67	0.031
SRC [32]	70.66 ± 1.15	2.086
Our method	95.81 ± 0.50	0.027

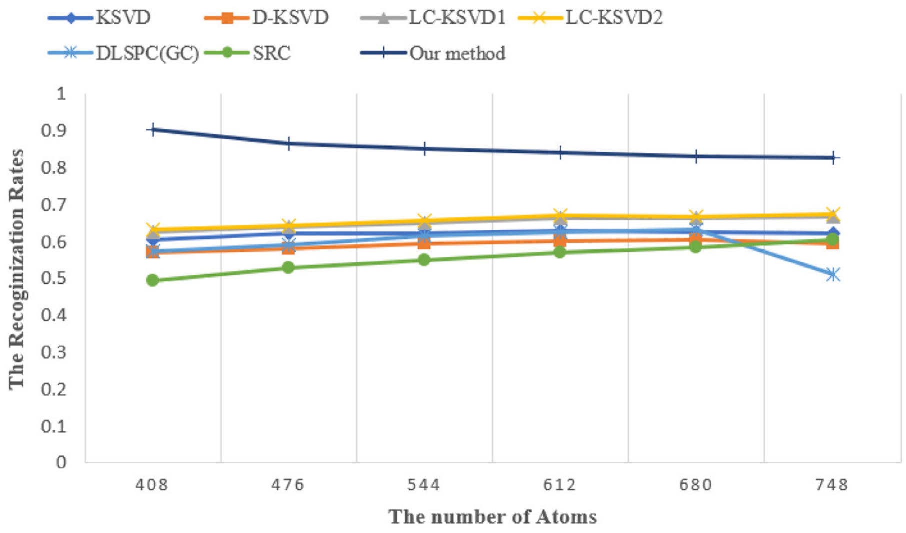

Fig. 7. The average recognition rates with different numbers of atoms on the PIE face database.

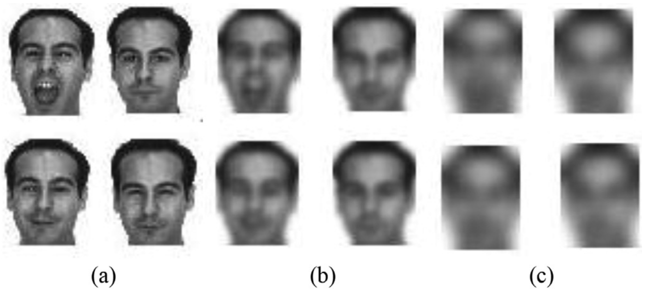

Fig. 8. Sample images from the AR face database. (a) Sample images with size of 50×40 pixels; (b) Sample images with size of 25×20 pixels; (c) Sample images with size of 12×10 pixels.

Table 4
The average recognition rates and computational time for classifying a test sample of different methods on the AR face database.

Algorithm	Recognition rates (%)	Time(s)
KSVD [25]	69.99 ± 1.08	3e-5
D-KSVD [18]	65.17 ± 1.11	1e-5
LC-KSVD1 [19]	74.58 ± 1.40	2.2e-5
LC-KSVD2 [19]	75.30 ± 1.24	2.1e-5
DLSPC(GC) [17]	68.38 ± 1.15	0.013
SRC [32]	78.18 ± 1.45	0.802
Our method	82.19 ± 1.54	0.031

While training, each dictionary is trained with shared parameters but different size. And we compare the average recognition rates on different numbers of dictionary atoms in a range of (240, 360, …, 840). Empirical results and recognizing time for one test sample are summarized in Table 4.

4.2.5. Experimental results on the LFW face database

The LFW database contains more than 13,000 images of faces collected in an unconstrained environment, and they are labeled with the names of different individuals. LFW is more challenging than the above databases since it includes various uncontrolled variations of pose and misalignment, etc. Following the experiment setting in [38], we use a cropped version (LFW crop) of a subset in the LFW database, which contains 158 subjects with 10 images per person. Each face image only retains the center portion of the image and almost all of the background is omitted. The training set collects 5 images from the face images of each subject randomly and the test set contains the remaining ones. We resize each original face image to 32×32 pixels and then convert it into 16×16 and 8×8 pixels via the resolution pyramid method. Fig. 10 show the images generated from LFW face database with three types of resolutions.

Table 5
The average recognition rates and computational time for classifying a test sample of different methods on the LFW face database.

Algorithm	Recognition rates (%)	Time(s)
KSVD [25]	11.08 ± 1.43	1e-5
D-KSVD [18]	7.70 ± 0.98	1.1e-5
LC-KSVD1 [19]	9.86 ± 0.88	1.8e-5
LC-KSVD2 [19]	11.54 ± 1.19	1.9e-5
DLSPC(GC) [17]	9.47 ± 0.61	0.072
SRC [32]	14.79 ± 1.21	0.158
Our method	16.63 ± 0.63	0.062

Similar to the experiments in the above databases, we trained the model on the different number of atoms in a range of (158, 316, …, 790). Experimental results on the LFW dataset are shown in Table 5. Table 5 indicates that our method outperforms the KSVD, D-KSVD, LC-KSVD, DLSPC, and SRC under the condition that the computational time is within the acceptable range.

4.3. Compared with deep learning methods

Recently, due to the competitive ability of representation and feature extraction, Convolutional Neural Networks (CNNs) have been widely used in many image processing tasks, especially in face recognition [12,13,39]. In this section, we employ several CNN based models (AlexNet [40], VGG [41], ResNet [42]) for comparison. As we all know, the training process of deep learning methods requires large-scale dataset. However, the scale of our benchmark databases is small. Therefore, to objectively compare the performance of deep learning methods with our proposed method, we trained them in two ways: without pre-training and with pre-training. Experimental results are shown in Table 6.

From the experimental results, we can see that deep learning based methods pre-trained on ImageNet [43] achieve better performance than methods without pre-training. The reason is that it is difficult for deep learning methods to obtain optimal parameter in the small-scale dataset. In addition, although pre-trained ResNet18 outperforms the proposed method on the Extended Yale

Table 6
The average recognition rates of different deep learning methods on the Extended Yale B, ORL, PIE, AR, and LFW face database.

Algorithm	Extended Yale B	ORL	PIE	AR	LFW
AlexNet	7.85%	2.50%	6.74%	5.51%	0.63%
VGG16	78.55%	5.50%	66.73%	77.12%	1.90%
ResNet18	86.48%	84.00%	69.00%	86.60%	15.19%
AlexNet with pre-training	51.84%	64.50%	46.08%	53.46%	6.84%
VGG16 with pre-training	82.97%	79.50%	69.00%	86.35%	8.61%
ResNet18 with pre-training	92.07%	92.00%	71.81%	90.38%	19.49%
Our method	88.58%	92.15%	95.81%	82.192%	16.63%

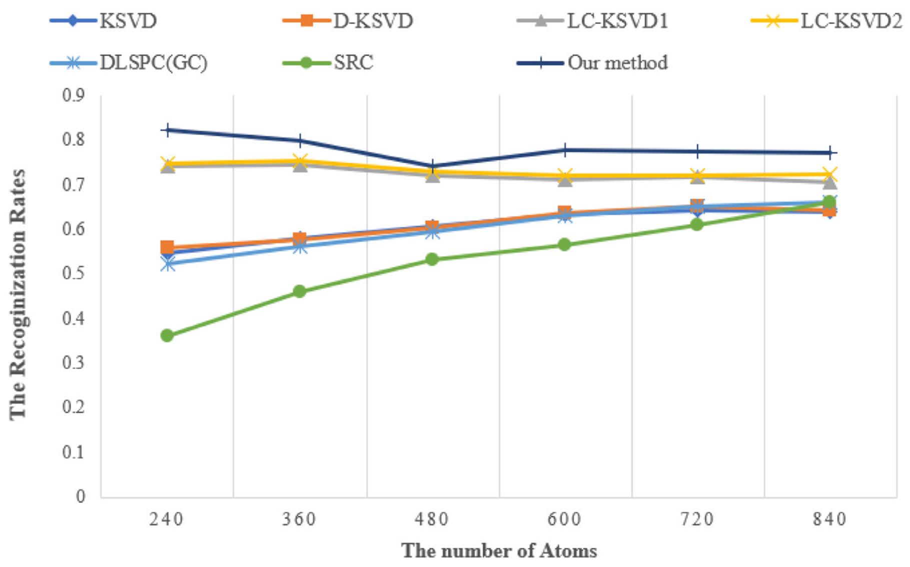

Fig. 9. The average recognition rates with different numbers of atoms on the AR face database.

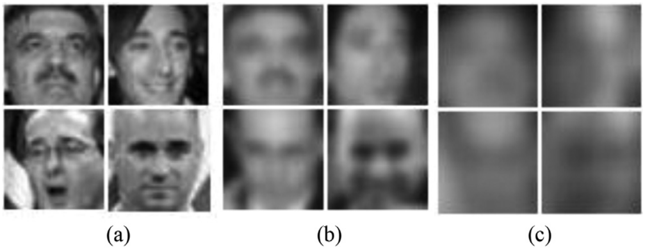

Fig. 10. Sample images from the LFW face database. (a) Sample images with size of 32 × 32 pixels; (b) Sample images with size of 16 × 16 pixels; (c) Sample images with size of 8 × 8 pixels.

B, AR, and LFW databases, our proposed method achieves better results on the ORL and PIE databases. And it should be mentioned that our proposed method is more flexible and needs less computing resource.

4.4. Experimental analysis

From the above experiments, we have the following observations.

(1) Our proposed method is suitable for face recognition tasks. In spites of the data sets we used contain different illumination conditions, facial expressions, poses and dresses, our method can still achieve satisfactory performance.

(2) When we use the same training samples and test samples on different methods, Tables 1–5 and Figs. 3, 5, 7, 9, and 11 show that the proposed method is superior to K-SVD, D-KSVD, LC-KSVD, DLSPC, SRC and achieves a better performance in almost all cases. It demonstrates that dictionaries obtained from the proposed method are stronger and more robust than the dictionaries of other state-of-art approaches.

(3) From Figs. 3, 5, 7, 9, and 11, we can see that our proposed method always achieves a higher average recognition rate despite the number of dictionary atom. It indicates that the proposed method has better ability to reconstruct multi-resolution images, even if the obtained dictionaries have a small size. Therefore, the proposed method is able to efficiently reduce the computational data redundancy and running time.

(4) The proposed method achieves better performance than KSVD and D-KSVD. It mainly because these atom-by-atomic optimization methods do not consider the overall dictionary optimization, and their obtained solutions are only the local optimal solutions in a certain sense. Our proposed method updates the dictionary as a whole, which can alleviate the above problems to some extent.

(5) Tables 1–5 show that, within the acceptable range of computational time, our proposed method requires more computational time than KSVD, D-KSVD, and LC-KSVD, but our recognition accuracy is far better than theirs. SRC and DLSPC always have higher accuracy than KSVD, D-KSVD, and LC-KSVD, but our proposed method takes shorter computational time than SRC and DLSPC. Overall, our proposed method has the best comprehensive performance among the above algorithms.

(6) From Table 6, we can see that the pre-trained deep learning methods usually outperform our method. In this sense, deep learning methods have advantages in accuracy. However, our method is computationally more efficient than deep learning methods. Moreover, our method is suitable for both small-scale

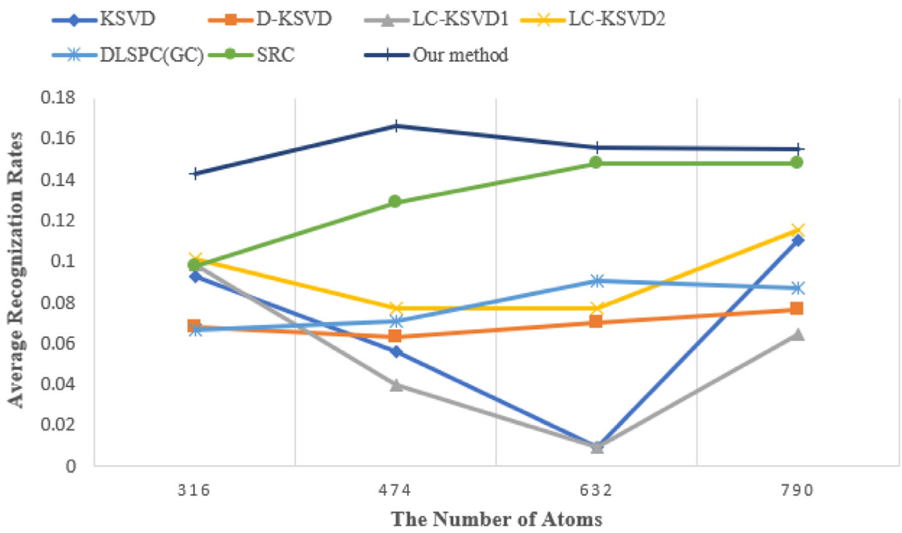

Fig. 11. The average recognition rates with different numbers of atoms on the LFW face database.

and large-scale datasets, but deep learning methods usually depend on large-scale datasets to achieve good and stable performance.

5. Conclusion

In this paper, we propose a novel multi-resolution dictionary learning method for face recognition. As far as we know, no similar algorithm has been proposed for dictionary learning. It is well known that when images are captured by different cameras, the obtained images usually have different resolutions. Moreover, the resolution of the images used for training of dictionary learning directly influence the performance. However, previous dictionary learning algorithms always exploit images of the same resolution for training. As a consequence, the obtained dictionary and features of the samples is not very suitable for the real case where the samples have different resolutions. Compared to previous dictionary learning algorithms, our proposed method not only provides dictionaries that associated with each resolution, but also adds a relatively strong constraint to keep the similarity of the representations obtained using different dictionaries in the training phase. Therefore, the learned dictionaries of the proposed algorithm are robust to different resolutions and not sensitive to noise. Though this work is focused on only dictionary learning, the proposed idea and scheme of multi-resolution learning might be also feasible for other kinds of methods. In the future, we will explore the issue to extend them to other methods.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under grant no. 61876051 and Science and Technology Innovation Committee of Shenzhen Municipality under grant no. JCYJ20180306172101694.

References

[1] J. Wen, X. Fang, J. Cui, L. Fei, K. Yan, K. Yan, Y. Chen, Y. Xu, Robust sparse linear discriminant analysis, IEEE Trans. Circuits .Syst. Video Technol. 29 (2) (2018) 1 –1.

[2] C. Tian, Q. Zhang, G. Sun, Z. Song, S. Li, FFT consolidated sparse and collaborative representation for image classification, Arab. J. Sci. Eng. 43 (2) (2018) 741–758.

[3] H. Li, X. He, D. Tao, Y. Tang, R. Wang, Joint medical image fusion, denoising and enhancement via discriminative low-rank sparse dictionaries learning, Pattern Recognit. 79 (2018) 130–146.

[4] J. Yang, J. Wright, T.S. Huang, Y. Ma, Image super-resolution via sparse representation, IEEE Trans. Image Process. 19 (24) (2010) 2861–2873.

[5] J. Wen, Y. Xu, Z. Li, Z. Ma, Y. Xu, Inter-class sparsity based discriminative least square regression, Neural Netw. 102 (2018) 36–47.

[6] Z. Yuan, T. Lu, C.L. Tan, Learning discriminated and correlated patches for multi-view object detection using sparse coding, Pattern Recognit. 69 (2017) 26–38.

[7] Y. Chen, J. Su, Sparse embedded dictionary learning on face recognition, Pattern Recognit. 64 (2017) 51–59.

[8] G. Lin, M. Yang, J. Yang, L. Shen, W. Xie, Robust, discriminative and comprehensive dictionary learning for face recognition, Pattern Recognit. 81 (2018) 341–356.

[9] J. Hu, Y. Tan, Nonlinear dictionary learning with application to image classification, Pattern Recognit. 75 (2018) 282–291.

[10] X. Lu, Y. Yuan, X. Zheng, Joint Dictionary Learning for Multispectral Change Detection, IEEE Trans. Cybernet. 47 (4) (2017) 884–897.

[11] Z. Jian, Y. Jun, T. Dacheng, Local deep-feature alignment for unsupervised dimension reduction, IEEE Trans. Image Process. 27 (5) (2018) 2420–2432.

[12] H. Chaoqun, Y. Jun, Z. Jian, J. Xiongnan, L. Kyong-Ho, Multi-modal face pose estimation with multi-task manifold deep learning, IEEE Trans. Ind. Inform. (2018) 1-1.

[13] Y. Sun, X. Wang, X. Tang, Hybrid Deep Learning for Face Verification, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2016) 1997–2009.

[14] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, F.R. Bach, Supervised dictionary learning, in: D. Koller, D. Schuurmans, Y. Bengio, L. Bottou (Eds.), Advances in Neural Information Processing Systems 21, Curran Associates, Inc., 2009, pp. 1033–1040.

[15] M. Yang, L. Zhang, X. Feng, D. Zhang, Fisher discrimination dictionary learning for sparse representation, in: 2011 International Conference on Computer Vision, IEEE, 2011, pp. 543–550.

[16] M. Yang, L. Zhang, X. Feng, D. Zhang, Sparse representation based fisher discrimination dictionary learning for image classification, Int. J. Comput. Vision 109 (3) (2014) 209–232.

[17] D. Wang, S. Kong, A classification-oriented dictionary learning model: explicitly learning the particularity and commonality across categories, Pattern Recognit. 47 (2) (2014) 885–898.

[18] Q. Zhang, B. Li, Discriminative K-SVD for dictionary learning in face recognition, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2010, pp. 2691–2698.

[19] Z. Jiang, Z. Lin, L.S. Davis, Label consistent K-SVD: learning a discriminative dictionary for recognition, IEEE Trans. Pattern Anal. Mach. Intell. 35 (11) (2013) 2651–2664.

[20] S. Cai, W. Zuo, L. Zhang, X. Feng, P. Wang, Support vector guided dictionary learning, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision-ECCV 2014, 2014, pp. 624–639.

[21] A. Shrivastava, J.K. Pillai, V.M. Patel, R. Chellappa, Learning discriminative dictionaries with partially labeled data, in: 2012 19th IEEE International Conference on Image Processing, 2012, pp. 3113–3116.

[22] B. Babagholami-Mohamadabadi, A. Zarghami, M. Zolfaghari, M.S. Baghshah, PSSDL: probabilistic semi-supervised dictionary learning, in: ECML PKDD 2013, Berlin, Heidelberg, Springer Berlin Heidelberg, 2013, pp. 192–207.

[23] H. Wang, F. Nie, W. Cai, H. Huang, Semi-supervised robust dictionary learning via efficient l-norms minimization, in: 2013 IEEE International Conference on Computer Vision, IEEE, 2013, pp. 1145–1152.

[24] M. Jian, C. Jung, Semi-supervised bi-dictionary learning for image classification with smooth representation-based label propagation, IEEE Trans. Multimedia 18 (3) (2016) 458–473.

[25] M. Aharon, M. Elad, A. Bruckstein, K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (2006) 4311–4322.

[26] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2010, pp. 3360–3367.

[27] R. Jenatton, J. Mairal, G. Obozinski, F. Bach, Proximal methods for sparse hierarchical dictionary learning, in: 27th International Conference on Machine Learning, 2010, Haifa, ICML, 2010, pp. 487–494.

[28] L.N. Smith, M. Elad, Improving dictionary learning: multiple dictionary updates and coefficient reuse, IEEE Signal Process. Lett. 20 (1) (2013) 79–82.

[29] J. Junjun, Y. Yi, W. Zheng, L. Xianming, M. Jiayi, Graph-regularized locality-constrained joint dictionary and residual learning for face sketch synthesis, IEEE Trans. Image Process. 28 (2) (2019) 628–641.

[30] J. Wen, B. Zhang, Y. Xu, J. Yang, N. Han, Adaptive weighted nonnegative low-rank representation, Pattern Recognit. 81 (2018) 326–340.

[31] K. Engan, S.O. Aase, J. Hakon Husoy, Method of optimal directions for frame design, in: 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258), 5, IEEE, 1999, pp. 2443–2446.

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

[33] A.S. Georghiades, P.N. Belhumeur, D.J. Kriegman, From few to many: illumination cone models for face recognition under variable lighting and pose, IEEE Trans. Pattern Anal. Mach. Intell. 23 (6) (2001) 643–660.

[34] F.S. Samaria, A.C. Harter, Parameterisation of a stochastic model for human face identification, in: Proceedings of 1994 IEEE Workshop on Applications of Computer Vision, IEEE Comput. Soc. Press, 1994, pp. 138–142.

[35] A.M. Martinez, The AR face database, CVC Technical Report24, (1998).

[36] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression (PIE) database, in: Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition, 2002, pp. 46–51.

[37] G.B. Huang, M. Mattar, T. Berg, E. Learned-Miller, Labeled faces in the wild: a database forstudying face recognition in unconstrained environments, Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, 2008.

[38] Y. Xu, Z. Li, B. Zhang, J. Yang, J. You, Sample diversity, representation effectiveness and robust dictionary learning for face recognition, Inform. Sci. 375 (2017) 171–182.

[39] H. Chaoqun, Y. Jun, W. Jian, T. Dacheng, W. Meng, Multimodal deep autoencoder for human pose recovery, IEEE Trans. Image Process. 24 (12) (2015) 5659–5670.

[40] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 25 (NIPS 2012), Curran Associates, Inc., 2012, pp. 1097–1105.

[41] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, CoRR, abs/1409.1556 (2014).

[42] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778.

[43] J. Deng, W. Dong, R. Socher, L. Li, K. Li, L. Fei-Fei, ImageNet: a large-scale hierarchical image database, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009), Los Alamitos, CA, USA, IEEE Computer Society, 2009, pp. 248–255.

Xiaoling Luo received her B.S. degree in Software Engineering at South China Normal University (SCNU) in 2017. She is currently pursuing the Ph.D. degree in computer science and technology at Harbin Institute of Technology, Shenzhen, China. Her research interests include pattern recognition, medical image processing and deep learning.

Yong Xu was born in Sichuan, China, in 1972. He received the Ph.D. degree in Pattern recognition and Intelligence System at the Nanjing University of Science and Technology (NUST) in 2005. Now, he works at Harbin Institute of Technology, Shenzhen, China. His current interests include pattern recognition, biometrics, machine learning and video analysis. More information please refer to http://www.yongxu.org/lunwen.html.

Jian Yang received the B.S. degree in Mathematics from Xuzhou Normal University, Xuzhou, China, in 1995, the M.S. degree in applied mathematics from Changsha Railway University, Changsha, China, in 1998, and the Ph.D. degree in the subject of pattern recognition and intelligence systems from the Nanjing University of Science and Technology (NUST), Nanjing, China, in 2002. In 2003, he was a Post-Doctoral Researcher with the University of Zaragoza, Zaragoza, Aragon, Spain. From 2004 to 2006, he was a Post-Doctoral Fellow with the Biometrics Center, Hong Kong Polytechnic University, Kowloon, Hong Kong. From 2006 to 2007, he was a Post-Doctoral Fellow with the Department of Computer Science, New Jersey Institute of Technology, Newark. Currently, he is a Professor with the School of Computer Science and Technology, NUST. He is the author of more than 50 scientific papers on pattern recognition and computer vision. His journal papers have been cited more than 1200 times on the ISI Web of Science, and 2000 times on Google Scholar. His current research interests include pattern recognition, computer vision, and machine learning.