Information Sciences 375 (2017) 171–182

Contents lists available at ScienceDirect

Information Sciences

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

Sample diversity, representation effectiveness and robust dictionary learning for face recognition

Yong Xu[a,b,*], Zhengming Li[a,c], Bob Zhang[d], Jian Yang[e], Jane You[f]

[a] Bio-Computing Research Center, Shenzhen Graduate School, Harbin Institute of Technology, Shenzhen, China
[b] Shenzhen Key Laboratory of Network Oriented Intelligent Computation, Shenzhen, China
[c] Industrial Training Center, Guangdong Polytechnic Normal University, Guangzhou 510665, China
[d] Department of Computer and Information Science, University of Macau, Avenida da Universidade, Taipa, Macau, China
[e] School of Computer Science & Technology, Nanjing University of Science & Technology, Nanjing, China
[f] Biometrics Researcher Centre, Department of Computing, Hong Kong Polytechnic University, Hong Kong, China

ARTICLE INFO

Article history:
Received 27 September 2015
Revised 19 September 2016
Accepted 27 September 2016
Available online 28 September 2016

Keyword:
Dictionary learning
Sparse coding
Face recognition

ABSTRACT

Conventional dictionary learning algorithms suffer from the following problems when applied to face recognition. First, since in most face recognition applications there are only a limited number of original training samples, it is difficult to obtain a reliable dictionary with a large number of atoms from these samples. Second, because the face images of the same person vary with facial poses and expressions as well as illumination conditions, it is difficult to obtain a robust dictionary for face recognition. Thus, obtaining a robust and reliable dictionary is a crucial key to improve the performance of dictionary learning algorithms for face recognition. In this paper, we propose a novel dictionary learning framework to achieve this. The proposed algorithm framework takes training sample diversities of the same face image into account and tries to obtain more effective representations of face images and a more robust dictionary. It first produces virtual face images and then designs an elaborate objective function. Based on this objective function, we obtain a mathematically tractable and computationally efficient algorithm to generate a robust dictionary. Experimental results demonstrate that the proposed algorithm framework outperforms some previous state-of-the-art dictionary learning and sparse coding algorithms in face recognition. Moreover, the proposed algorithm framework can also be applied to other pattern classification tasks.

1. Introduction

Dictionary learning is an important branch of sparse representation, and it is widely used in pattern recognition [13,47] and image processing [41]. Sparse representation has the key idea that samples can be sparsely represented by a large number of "atoms", and it is widely used for face recognition [35,43]. In order to produce competent "atoms", researchers have proposed various dictionary learning algorithms [5,45].

Although dictionary learning has exhibited promising performance in face recognition [4,5,12,44,45], previous dictionary learning algorithms suffer from the following problems. First, face recognition is a typical small sample size problem, and insufficient available samples have severe negative effects on dictionary learning algorithms for face recognition. Actually,

* Corresponding author.
E-mail address: yongxu@ymail.com (Y. Xu).

http://dx.doi.org/10.1016/j.ins.2016.09.059
0020-0255/

learning an over-complete dictionary from just a limited number of high-dimensional images is a challenge for all image classification issues. Second, because the face images of the same person vary with facial poses and expressions as well as illumination conditions, it is hard to obtain a very robust dictionary for face recognition. For face recognition tasks, even if the dictionary is not very sensitive to variations in poses, expressions, and illumination conditions, it is able to obtain relatively stable descriptions of the face image with a high accuracy. Exploring and tackling the above two problems represent a significant step in the study of dictionary learning for face recognition. In order to achieve robust face recognition performance via dictionary learning, researchers have proposed novel algorithms with valuable ideas. For example, learning an “occlusion dictionary” is proposed for better recognition of occluded face images [19,44]. However, it seems that an “occlusion dictionary” is just a special-problem-associated algorithm and cannot perform very well when one is dealing with a general face recognition issue. Learning a robust dictionary [14] is also very important for other fields such as background modeling [25] and tracking [34]. In addition, dictionary learning has been applied to a wide array of image-associated tasks [17,24].

This paper proposes a new dictionary learning framework, focusing on enhancing the diversity of training samples of the same face and obtaining more effective representation of face images. The robust dictionary learning framework designed in this study is not only applicable to face recognition but can also be applied to other pattern classification tasks. For face recognition tasks, the proposed algorithm framework achieves robustness by exploiting an elaborated algorithm and by generating virtual face images that convey new possible poses and illuminations of the face. For other pattern classification tasks, the designed framework can be applied with either of the two following schemes. (1) Virtual training samples are first obtained by corrupting original training samples. Then the designed elaborated algorithm is applied to the original and virtual training samples. (2) The set of original training samples is divided into two, and the designed elaborated algorithm is applied to these two halves. In other words, the first and second halves are viewed as the original and virtual training samples, respectively. This scheme may be highly effective when there is a large number of original training samples. In particular, compared with previous robust dictionary learning algorithms on the basis of l_1 regularization such as those in [25,34], the elaborated algorithm designed in this paper has a much lower computational cost.

The remainder of this paper is organized as follows. Section 2 presents the related work of dictionary learning algorithms. Section 3 provides the proposed algorithm framework. Section 4 supplies the experimental results and analysis. Section 5 presents the conclusion.

2. Related work

Dictionary learning algorithms can be roughly categorized 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 are mainly designed based on the reconstruction error and constraint on labels. Dictionaries obtained in a supervised manner usually have strong discriminative power, and they have the potential to achieve excellent classification performance for classification problems. Studies on supervised dictionary learning have been extensively performed [7,10,21,31,32,40,42,46]. In order to improve the discriminative ability of the dictionary, the Fisher discrimination criterion [42] was used to implement the supervision. Moreover, Wang et al. [32] proposed learning class-specific dictionaries and a global dictionary shared by all categories. However, if the number of categories is large, the global dictionary may be not suitable for dealing with the complex category correlation. Zhang and Li [45] proposed the discriminative K-SVD algorithm (D-KSVD) based on the classification error. Jiang et al. [10] proposed the label consistent K-SVD algorithm (LC-KSVD) by constructing the discriminative sparse code error term to improve the discriminative ability of the learned dictionary. Recently, Cai et al. [5] proposed a support vector-guided dictionary learning algorithm (SVGDL) for image classification by using the weighted summation of the squared distances between all pairs of coding vectors to construct the discrimination term. Gu et al. [7] proposed a projective dictionary pair learning algorithm (PDPL) for pattern classification. Yang et al. [40] presented a unified model by integrating the analysis-synthesis dictionary learning and universality-particularity representation. Tang et al. [31] proposed an efficient method to learn a compact and discriminative dictionary for visual categorization, in which dictionary learning was formulated as a problem of graph partition. When dealing with high-complexity data due to the use of simple supervised techniques, many dictionary learning algorithms usually suffer from insufficient discrimination. Quan et al. [21] proposed a supervised dictionary learning algorithm by integrating multiple classifier training into dictionary learning to improve the discrimination. Moreover, in order to transform a dictionary learned from one visual domain to the other, Zhu and Shao [46] utilized weakly labeled data from other visual domains as the auxiliary source data for enhancing the original learning system. These supervised dictionary learning algorithms have achieved excellent performance in classification tasks. However, in many pattern classification problems, accessibility to a large set of labeled data may not be possible because labeling data is expensive and very time consuming. Thus, insufficient labeled training data are adverse to supervised dictionary learning algorithms.

2.2. Semi-supervised dictionary learning algorithms

Unlabeled data are easily available and also very useful. This has motivated researchers to design semi-supervised dictionary learning algorithms which exploit unlabeled data, along with labeled data, to build better dictionary for classification tasks. Shrivastava et al. [26] proposed a semi-supervised dictionary learning algorithm using labeled training samples to learn specific-class dictionaries, and then used them to predict the label of unlabeled training samples. However, the algorithm did not take into account the underlying geometrical structure of both labeled and unlabeled data, and generally could not preserve the locality structure and thus the obtained dictionary may not be optimal for classification tasks. In particular, in the case where the data lies in the nonlinear manifold embedded in a very high-dimensional space [22,28], the classification performance of the above algorithm will be degraded. In order to address this problem, Babagholami-Mohamadabadi et al. [2] proposed a semi-supervised dictionary learning algorithm by allowing different categories to share a single dictionary. Recently, Wang et al. [33] proposed a semi-supervised dictionary learning algorithm to automatically optimize the dictionary size. Meanwhile, Jian and Jung [9] proposed a semi-supervised dictionary learning algorithm using the smooth representation-based label propagation method.

The main shortcoming of semi-supervised dictionary learning algorithms is that they are usually sensitive to the number of labeled training samples.

2.3. Unsupervised dictionary learning algorithms

Unsupervised dictionary learning algorithms are usually designed based on the reconstruction error of training samples. One of the most well-known unsupervised dictionary learning algorithms is the K-SVD algorithm [1]. In order to take advantage of the local geometric structure of training samples, Wang et al. [36] proposed a locality-constrained linear coding algorithm (LLC) by using distances between the bases (atoms) and training samples to select the k nearest neighbor atoms for coding, and to set the coding coefficients of other bases to zero. Moreover, Jenatton et al. [11] encoded the dependencies between atoms of the dictionary by using a tree-structured sparsity. Yu et al. [39] proposed a hierarchical sparse coding algorithm by using the spatial neighborhood dependency of local patches. However, when the number of training samples is small, these algorithms are not very suitable for classification tasks. Obtaining virtual training samples is a feasible way to improve the performance of a classification task [38]. We see that even a simple procedure to obtain virtual training samples such as the mean-face procedure proposed in [37] may be beneficial for better recognition of faces.

3. The proposed algorithm framework

Given a training sample set $Y=[y_1, \ldots, y_N] \in \Re^{n\times N}$, let $D=[d_1, \ldots, d_K] \in \Re^{n\times K}$ be the obtained dictionary matrix, where each d_i represents an atom in dictionary. D, N and K are the numbers of all original training samples and atoms, respectively. $X=[x_1, \ldots, x_N] \in \Re^{K\times N}$ is the coding coefficients matrix. We also assume that the training sample set contains all training samples from C categories.

3.1. The objective function

In order to make the learned dictionary robust to variations in facial expressions and poses of the same person, we can obtain alternative training samples using a special scheme. The set of alternative training samples has the same size and structure as the set of original training samples. Hereafter the same structure means that the J-th ($J=1,\ldots, N$) entries of the set of alternative training samples and set of original training samples have the same label. N is the number of all original training samples. Therefore, it is reasonable to require that these two sets have the same dictionary. Thus, the proposed algorithm framework is based on objective function

$$\begin{array}{l} \min\limits_{D,X} \|Y-DX\|_2^2+\alpha\|Y_{alter}-DX\|_2^2+\beta\|X\|_2^2 \\ s.t. \|d_i\|^2=1, i=1,\cdots,K \end{array}, \tag{1}$$

where Y_{alter} is the matrix consisting of all alternative training samples. In order to enable the effect of alternative training samples to be flexible, the proposed algorithm framework introduces parameter α. Under the condition that the original training samples are more reliable than alternative training samples, α can be set to a positive constant less than 1. Under the condition that these two kinds of training samples are equally important, α can be set to about 1. β is a regularization parameter. The following three specific features distinguish our proposed algorithm framework from previous dictionary learning algorithms.

(1) The proposed algorithm framework takes training sample diversities of the same face image into account and tries to obtain more effective representations of face images and a more robust dictionary.
(2) Instead of using the L0-norm or L1-norm, we use the L2-norm to constrain the coding coefficients in dictionary learning, and it can obtain a mathematically tractable and computationally efficient algorithm.

(3) The designed algorithm is indeed a framework which allows various schemes to produce "alternative training sample" to be used. It is not only applicable for face recognition but can also be applied to other pattern classification issues. Actually, once reasonable alternative training samples are available, our framework can be directly applied to other pattern classification issues.

The proposed algorithm framework first obtains alternative training samples using a special scheme. For a classification task with sufficient training samples, we can also simply divide the set of naive training samples into two subsets with the same size and take the first subset and second subset as the original and virtual training samples, respectively. In this paper, we use three methods to generate alternative training samples and apply them to the proposed algorithm framework. Because the proposed algorithm framework requires original face images and alternative training samples to produce the same dictionary, the algorithm is robust to deformation of face images of the same person. As a result, the proposed dictionary learning framework is very useful because of its applicability and flexibility to face recognition. The procedures to generate alternative training samples are presented as follows.

(1) We treat mirror face images of original training samples as alternative training samples and refer to the corresponding implementation of the proposed algorithm framework as MFI-DL algorithm. According to Xu et al. [38], for original training sample y in the form of face image, its mirror face image is defined as

$$y^m(p,q) = y(p, Q-q+1)(p=1,\cdots,P; q=1,\cdots,Q) \tag{2}$$

where P and Q denote the numbers of rows and columns of the face image matrix, respectively. $y(p, q)$ and $y^m(p, q)$ denote the pixels located in the p-th row and q-th column of y and y^m, respectively.

(2) We use the corrupted images of original training samples as alternative training samples, and the noise images are produced by applying the Matlab function "imnoise" to the original face images. When we obtain alternative training samples via corrupting the original face images by using the Gaussian noise, and Salt & Pepper noise, we refer to the corresponding implementation of the proposed algorithm framework as GN-DL and SPN-DL algorithms, respectively.

(3) We use the eigenface images of original training samples as alternative training samples and refer to the corresponding implementation of the proposed algorithm framework as EI-DL algorithm. The eigenface images are generated using the procedure described in [30].

3.2. Optimization of the objective function

In general, a dictionary can be obtained using two kinds of algorithms. The first kind of algorithms updates the atoms one by one, and the K-SVD algorithm is a typical example of this kind. In the second kind of algorithms, the learned dictionary is updated as a whole [13]. For our algorithm, when we update one variable, we fix the others. As a result, the objective function of our proposed algorithm framework has a closed form solution. We present the procedure to obtain the solution of Eq. (1) as follows.

If we fix D, we can calculate X using

$$X = \left(D^TD + \alpha D^TD + \beta I\right)^{-1}\left(D^TY + \alpha D^TY_{alter}\right). \tag{3}$$

where I is an identity matrix.

Under the condition that X is fixed, then D can be calculated using

$$D = \left(YX^T + \alpha Y_{alter}X^T\right)\left(XX^T + \alpha XX^T\right)^{-1}. \tag{4}$$

If dictionary D is initialized, then Eqs. (3) and (4) can be iteratively updated to obtain the ultimate solution. In the proposed algorithm framework, we use all original training samples and the K-SVD algorithm to obtain initial dictionary D.

We analyze the computational complexity of the iteration procedure to solve our objective function below. Suppose that each sample is an n-dimensional column vector. Let Y and D be an n by N matrix and n by K matrix, respectively, and X is a K by N matrix. The main computations are from the loop procedure in Table 1. In this procedure, the computational complexity to update X is $O(K^2n)+O(K^3)+O(KNN)+O(K^2n)$. The computational complexity to update D is $O(KnN)+O(K^2N)+O(K^2n)$. As a result, the proposed algorithm framework is computationally very efficient.

3.3. The classification procedure

Following [20], we adopt a linear classification procedure as follows.

First, a classifier $f(X)$ can be obtained by determining its model parameters $W \in \Re^{C\times K}$ with constraint

$$W = \min_W \Gamma\{H, f(X,W)\} + \lambda\|W\|_F^2, \tag{5}$$

where $H=[h_1, \ldots, h_n] \in \Re^{C\times n}(h_i=[0, \ldots, 1, \ldots 0]^T \in \Re^C)$ is the label matrix of Y, and the non-zero position of column vector indicates the class of training samples. Γ is the loss function and λ is a regularization parameter. Therefore, based on coding coefficients matrix X and label matrix H, we calculate classifier parameter matrix $W \in \Re^{C\times K}$ by using

$$W = HX^T(XX^T + I)^{-1} \tag{6}$$

Table 1
The proposed algorithm framework.

Input: Training sample matrix Y (and Y_{alter}), Parameters α and β, Test sample T, Label matrix H of training sample set Y.
Output: Dictionary D, Coding coefficients matrix X and representation coefficient V, Label k of test sample T.
Initialization: The K-SVD algorithm is used to obtain initial dictionary D;
while not converged **do**
Update X using Eq. (3).
Update D using Eq. (4).
end while
calculate W using Eq. (6)
The OMP algorithm is exploited to calculate representation coefficient V of test sample T.
Let $g = WV$. If $k = \underset{j}{Max}(g_j)$, then test sample T is assigned to the k-th class.

Fig. 1. Sample images from the PIE face database.

The representation coefficient $V \in \Re^{K\times 1}$ of test sample T is obtained by using the orthogonal matching pursuit (OMP) algorithm [29]. The label vector of test sample T is calculated using $g = WV$. Let g_j denote the j-th entry of g. Suppose that $k = \max(g_j)_j$, $j = 1, \cdots, C$, then test sample T is assigned to the k-th class. The proposed algorithm framework is described in Table 1.

4. Experimental results and analysis

In order to evaluate the performance of our proposed algorithm framework, we compare it with the K-SVD [1], D-KSVD [45], LC-KSVD [10], SRC [35], LRC [18], CRC [48], LLC [36], SVGDL [5], PDPL [7] and DNFC [37] algorithms via experiments on the CMU PIE face database (PIE) [23], the Labeled Faces in the Wild (LFW) database [8], the AR face database (AR) [16] and the Extended Yale B face database [6]. Moreover, in order to better show the performance of our proposed algorithms, we also compare our proposed algorithms with the ten comparison algorithms on the Fifteen Scene Category database [15].

4.1. Experimental setting

In this section, we give the experimental details. The mirror images are produced by using the algorithm proposed in [38]. The noise images are produced by using the Matlab function "imnoise" to impose noise on the original face images. The first parameter of "imnoise" should be an original face image. If the second parameter of "imnoise" is set to 'Gaussian' or 'Salt & Pepper', the Gaussian noise or Salt & Pepper noise will be added to the original image. For the Gaussian noise, the mean (i.e., the third parameter of "imnoise") is always set to zero and the standard deviation (i.e., the fourth parameter of "imnoise") is set to 0.1. For the Salt & Pepper noise, the density of noise (i.e., the third parameter of "imnoise") is set to 0.1.

Following [10], we use sparsity factor $\psi = 30$ in the K-SVD, D-KSVD, and LC-KSVD algorithms as well as our proposed algorithm framework. Moreover, the LLC algorithm obtains the coding coefficients by using the approximated LLC. Then, the distances between the test sample and each class are calculated. Thus, the test sample is assigned by using the minimum distance. In order to make a fair comparison, the number of local bases of the LLC algorithm is identical to sparsity factor ψ. In the experiments on K-SVD, D-KSVD, and LC-KSVD algorithms and our proposed algorithm framework, a linear classifier method [20] is used to classify the test samples.

In addition, there are two parameters in our objective function. We use fixed parameter values $\alpha = \beta = 10^{-3}$ in the experiments on the four face databases. For the Fifteen Scene Category database, α and β are set to 1 and 10^{-4}, respectively.

4.2. Experimental results on the PIE face database

The PIE face database consists of 41,368 front-face images of 68 persons, and the face images of each person are captured under 13 different poses, 43 different illumination conditions, and 4 different facial expressions. Several sample images from the PIE face database are shown in Fig. 1.

Following [3], we choose the five near-frontal poses (C05, C07, C09, C27, C29) of each subject and use all the images under different illumination conditions and facial expressions. Thus we get 170 images for each individual. Every image is normalized to the size of 32×32. We randomly select ten images of each person as training samples (the first five images of each person are always used, and the other five images of each person are randomly selected from the rest of the images)

Table 2
The average recognition rates and computing time for training a dictionary and classifying a test sample on the PIE face database.

Algorithm	Average recognition rates (%)	Training time(s)	Testing time(s)
LRC [18]	61.6 ± 0.021	–	1.45e-2
CRC [48]	74.1 ± 0.012	–	3.53e-3
LLC [36]	53.7 ± 0.016	–	2.83e-2
SRC [35]	72.1 ± 0.008	–	8.50
K-SVD [1]	72.0 ± 0.013	8.39	1.32e-3
D-KSVD [45]	71.9 ± 0.008	235	1.31e-3
LC-KSVD [10]	72.3 ± 0.009	28.8	1.31e-3
DNFC [37]	71.4 ± 0.009	–	6.53e-2
PDPL [7]	77.5 ± 0.011	25.0	0.71e-4
SVGDL [5]	76.4 ± 0.007	634	0.7e-5
MFI-DL	77.2 ± 0.010	15.80	1.31e-3
GN-DL	77.0 ± 0.008	15.80	1.31e-3
SPN-DL	78.0 ± 0.008	15.80	1.31e-3
EI-DL	77.8 ± 0.012	15.80	1.31e-3

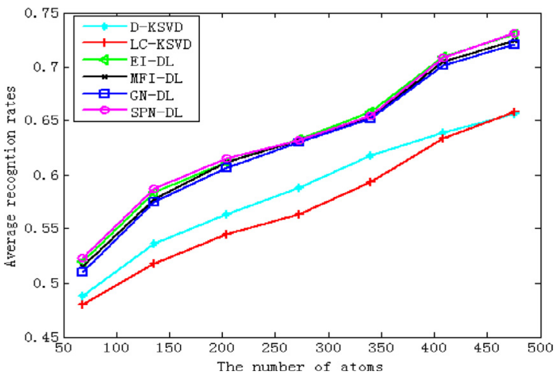

Fig. 2. The average recognition rates versus different numbers of atoms on the PIE face database.

and use the remaining samples as test samples. The proposed algorithm and ten comparison algorithms are carried out ten times, and the average recognition rates are reported in Table 2. Moreover, the average computing times for training a dictionary and classifying a test sample are also reported in Table 2. The number in the bracket in Table 2 is the number of atoms.

Table 2 shows that the proposed algorithms achieve higher average recognition rates than the ten comparison algorithms in most cases. In particular, the SPN-DL algorithm achieves the highest average recognition rate.

In order to further test the proposed algorithms, we also compare the average recognition rates of the D-KSVD, LC-KSVD, MF-DL, GN-DL, SPN-DL and EI-DL algorithms with different numbers of atoms (K=68, 136, ..., 476). The experimental results are shown in Fig. 2. It shows that the proposed algorithms outperform the D-KSVD and LC-KSVD algorithms.

Fig. 2 shows that the average recognition rates of the D-KSVD and LC-KSVD algorithms and the proposed algorithms all increase with the increase in the number of atoms. This is mainly because the reconstruction ability and discriminative ability of the dictionary improve with the increase of the number of atoms. Thus, better classification performance can be obtained.

4.3. Experimental results on the LFW database

The LFW database contains more than 13,000 images of faces collected from the web, and all of them labeled with the name of the person pictured. The main goal is to study the problem of unconstrained face recognition. In the database, 1680 of the persons have two or more distinct photographs. Following [27], we use a cropped version (LFW crop) of the Labeled Faces in the Wild (LFW) dataset, which keeps only the center portion of each image (i.e., the face), and almost all of the background is omitted. The LFW crop face database was created due to concerns about the misuse of the original LFW dataset, where face matching accuracy can be unrealistically boosted through the use of background parts of images (i.e. exploitation of possible correlations between faces and backgrounds). For each LFW image, the area inside a fixed bounding box was extracted. The bounding box was at the same location for all images, with the upper-left and lower-right corners being (83,92) and (166,175), respectively. The extracted area was then scaled to a size of 64×64 pixels. The selection of bounding box location was based on the positions of 40 randomly selected LFW faces. As the location and size of faces in the LFW database were determined through the use of an automatic face locator (detector), the cropped faces in the

Fig. 3. Sample images from the LFW crop database.

Table 3
The average recognition rates and computing time for training a dictionary and classifying a test sample on the LFW database.

Algorithm	Average recognition rates (%)	Training time (s)	Testing time (s)
LRC [18]	20.8 ± 0.011	–	11.8
CRC [48]	27.3 ± 0.008	–	6.89e-3
LLC [36]	21.7 ± 0.013	–	22.6
SRC [35]	27.2 ± 0.012	–	561
K-SVD [1]	17.1 ± 0.032	6.99	1.9e-3
D-KSVD [45]	23.6 ± 0.015	206	1.8e-3
LC-KSVD [10]	22.3 ± 0.012	9.61	1.76e-3
DNFC [37]	25.6 ± 0.015	–	26.6
PDPL [7]	26.5 ± 0.007	37.6	1.46e-3
SVGDL [5]	24.7 ± 0.015	554	2.49e-4
MFI-DL	27.8 ± 0.014	7.86	1.95e-3
GN-DL	28.3 ± 0.011	7.86	1.95e-3
SPN-DL	28.1 ± 0.020	7.86	1.95e-3
EI-DL	28.4 ± 0.012	7.86	1.95e-3

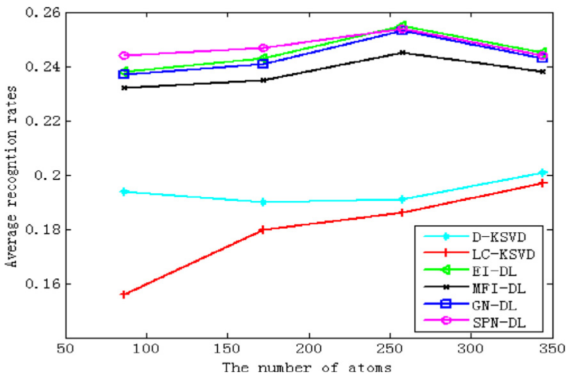

Fig. 4. The average recognition rates versus different numbers of atoms on the LFW database.

LFW crop database exhibit real-life conditions, including misalignment, scale variations, in-plane as well as out-of-plane rotations. In this experiments, we select a subset of the LFW crop face database consisting of 1215 images of 86 persons. In this subset, each person has around 11–20 images. Each image is resized to a 32×32 image. Sample images from the LFW crop database are shown in Fig. 3.

We randomly select six images of each person as training samples and take the remaining as test samples. The proposed algorithm framework and ten comparison algorithms are repeatedly carried out ten times and the average recognition rates are reported in Table 3. Moreover, the computing times for training a dictionary and classifying a test sample are also reported in Table 3.

Table 3 shows that the proposed algorithms still achieve higher average recognition rates than the ten comparison algorithms.

We also compare the average recognition rates of the D-KSVD, LC-KSVD, MFI-DL, GN-DL, SPN-DL and EI-DL algorithms with different numbers of atoms ($K = 86$, 172, 258, 344). The experimental results are presented in Fig. 4. This shows again that the proposed algorithms outperform the D-KSVD and LC-KSVD algorithms.

4.4. Experimental results on the AR face database

The AR face database contains more than 4000 images of 126 persons. Each person has 26 face images captured in two sessions, and each face image is captured under various lighting conditions. We use a subset of the AR face database

Fig. 5. Sample images of one person from the AR face database. The images in the first two lines are captured in Session 1, and images in the last two lines are captured in Session 2.

Table 4
The average recognition rates and computing time for training a dictionary and classifying a test sample on the AR face database.

Algorithm	Average recognition rates (%)	Training time (s)	Testing time (s)
LRC [18]	69.7 ± 0.074	–	3.42e-2
CRC [48]	71.4 ± 0.061	–	7.75e-3
LLC [36]	71.1 ± 0.060	–	7.71e-2
SRC [35]	72.2 ± 0.072	–	3.41
K-SVD [1]	78.8 ± 0.063	25.9	1.81e-3
D-KSVD [45]	74.1 ± 0.067	1137	1.63e-3
LC-KSVD [10]	74.2 ± 0.066	80.9	1.71e-3
DNFC [37]	77.7 ± 0.064	–	2.21e-1
PDPL [7]	79.5 ± 0.059	70.3	5.30e-4
SVGDL [5]	78.0 ± 0.059	2139	3.68e-4
MFI-DL	79.9 ± 0.062	45.3	1.84e-3
GN-DL	79.7 ± 0.066	45.3	1.84e-3
SPN-DL	79.8 ± 0.067	45.3	1.84e-3
EI-DL	79.6 ± 0.066	45.3	1.84e-3

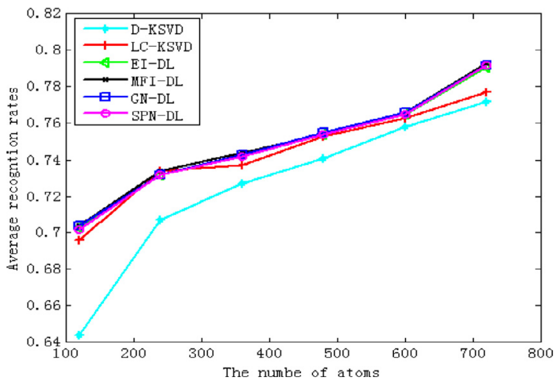

Fig. 6. The average recognition rates versus different numbers of atoms on the AR face database.

consisting of 3120 images from 120 persons (65 men and 55 women). The resolution of the AR images is 40×50. For each person, 12 face images with sunglasses and scarves were captured in two sessions. The images of one person from the AR face database are shown in Fig. 5.

We choose seven neutral face images captured in session 1 and one occluded face image of each person from the used subset as training samples (the first face image of each person with sunglasses captured in sessions 1 and 2 are not used as training samples). Thus, sixteen images of each person (seven neutral images at session 2 plus the remaining nine occluded images) are used for testing. Since each person has ten available occluded face images, the experiments are carried out ten times, and the average recognition rates are given in Table 4. The computing times for training a dictionary and classifying a test sample are also reported in Table 4.

As shown in Table 4, the average recognition rates of the proposed algorithms are higher than those of the other algorithms when the number of atoms is 960. In particular, the MFI-DL algorithm achieves the highest average recognition rate among all algorithms.

In order to further test the proposed algorithm framework, we also compare the average recognition rates of the D-KSVD, LC-KSVD, MFI-DL, GN-DL, SPN-DL and EI-DL algorithms with different numbers of atoms ($K = 120, 240, \ldots, 600, 720$). The experimental results are shown in Fig. 6.

Fig. 7. Sample images from the Extended Yale B face database.

Table 5
The average recognition rates and computing time for training a dictionary and classifying a test sample on the Extended Yale B face database.

Algorithm	Average recognition rates (%)	Training time(s)	Testing time(s)
LRC [18]	92.4 ± 0.008	–	1.65e-2
CRC [48]	95.0 ± 0.009	–	2.79e-3
LLC [36]	88.9 ± 0.010	–	3.04e-2
SRC [35]	95.3 ± 0.005	–	1.84
K-SVD [1]	94.0 ± 0.005	10.1	1.59e-3
D-KSVD [45]	94.3 ± 0.005	398	1.316e-3
LC-KSVD [10]	92.7 ± 0.008	34.9	1.43e-3
DNFC [37]	92.5 ± 0.010	–	1.07e-1
PDPL [7]	96.0 ± 0.005	45.8	4.30e-4
SVGDL [5]	93.8 ± 0.006	843	1.52e-4
MFI-DL	95.9 ± 0.007	33.1	1.05e-3
GN-DL	96.0 ± 0.006	33.1	1.05e-3
SPN-DL	96.2 ± 0.009	33.1	1.05e-3
EI-DL	96.2 ± 0.004	33.1	1.05e-3

We see that with the increase in the number of atoms, the average recognition rates of the D-KSVD and LC-KSVD algorithms and the proposed algorithms all increase. Moreover, it is obvious that in terms of the average recognition rates, the proposed algorithms outperform the D-KSVD and LC-KSVD algorithms.

4.5. Experimental results on the Extended Yale B face database

The Extended Yale B face database contain 2414 front-face images of 38 persons, and the face images are obtained under various illumination conditions and expressions. Every person has 59–64 images and each image is normalized to the size of 32×32. Several sample images from the Extended Yale B face database are shown in Fig. 7.

We randomly select twenty images of each person (the first five images of each person are always used, and the other fifteen images of each person are randomly selected from the rest of the images) to form the training samples and use the remaining samples as test samples. We repeatedly run the proposed algorithms and ten comparison algorithms ten times and report the average recognition rates in Table 5. The computing times for training a dictionary and classifying a test sample are also reported in Table 5. When the number of atoms is 760, the proposed algorithms still achieve a higher average recognition rate than the ten comparison algorithms.

In order to further evaluate the performance of the proposed algorithms, we also compare the average recognition rates of the D-KSVD, LC-KSVD, MFI-DL, GN-DL, SPN-DL and EI-DL algorithms with different numbers of atoms (K=38, 76, ..., 722, 760). The experimental results are shown in Fig. 8. When the number of atoms is larger than 532, the average recognition rates of the proposed algorithms are higher than those of the D-KSVD and LC-KSVD algorithms.

4.6. Experimental results on the Fifteen Scene Category database

For a classification task with sufficient training samples, we can also simply divide the set of naive training samples into two subsets with the same size and take the first subset and second subset as original and alternative training samples, respectively. In order to show the classification performance of our proposed algorithms, we use the same setting as the LC-KSVD algorithm on the Fifteen Scene Category database. Actually, the Fifteen Scene Category is a database of fifteen natural scene categories, and it contains fifteen scenes such as bedroom, kitchen and country scenes. Each category has 200 to 400 images, and the average image size is about 250×300 pixels. Several sample images from the Fifteen Scene Category database are shown in Fig. 9.

The spatial pyramid feature of the Fifteen Scene Category is extracted by using a four-level spatial pyramid and a SIFT-descriptor codebook with a size of 200. The final spatial pyramid features are reduced to 3000-dimension features using the PCA method. We randomly select 100 images per category as training samples, and the rest as testing samples. For the training samples, we use the first 50 images per category as original training samples, and the rest as alternative training

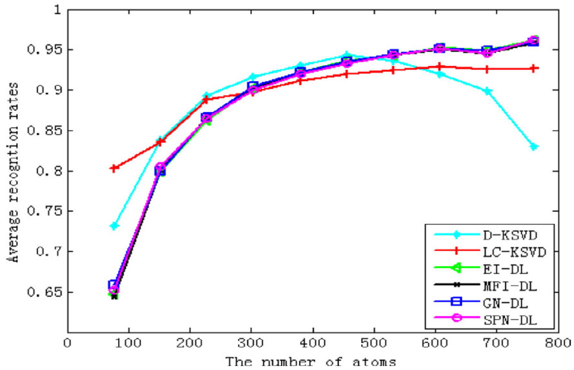

Fig. 8. The average recognition rates versus different numbers of atoms on the Extended Yale B face database.

Fig. 9. Sample images from five categories in the Fifteen Scene Category database.

Table 6
The average recognition rates on the Fifteen Scene Category database.

Algorithm	Average recognition rates (%)
LRC [18]	96.1
CRC [48]	95.8
LLC [36]	89.2
SRC [35]	91.8
K-SVD [1]	86.7
D-KSVD [45]	89.1
LC-KSVD [10]	92.9
DNFC [37]	95.9
PDPL [7]	97.0
SVGDL [5]	95.4
Our proposed algorithm	97.3

samples. In this experiment, the number of atoms is 450. The experimental results are shown in Table 6. It can be seen that our proposed algorithms can achieve higher average recognition rates than the ten comparison algorithms.

4.7. Analysis of experimental results

We summarize the experimental results below.

(1) When the numbers of atoms and original training samples are the same, Tables 2–6 show that the proposed algorithm framework achieves higher average recognition rates than the SRC, LRC, LLC, DNFC and CRC algorithms, which directly use only original training samples to perform pattern recognition. This demonstrates that the obtained dictionaries have more discriminative ability than the original training samples.
(2) When the numbers of atoms and original training samples are the same, Tables 2–6 show that the average recognition rate of the proposed algorithm framework is superior to those of the K-SVD, D-KSVD and LC-KSVD algorithms. This demonstrates that the proposed algorithm framework has more power discriminative ability than the K-SVD, D-KSVD and LC-KSVD algorithms, although they are all dictionary learning algorithms.
(3) When the number of atoms is equal to that of training samples, the average recognition rate of the proposed algorithm framework is superior to those of two compared specific-class dictionary learning algorithms, the PDPL and SVGDL algorithms. This is mainly because all atoms of the specific-class dictionaries have more redundancy which may degrade the classification performance. Our proposed algorithm framework combines the original training samples and alternative training samples to learn a dictionary for all the classes. It can overcome the above shortcoming to some extent.

(4) When we integrate the mirror face images, the face images corrupted by the Gaussian noise, Salt &Pepper noise and eigenface images of the original training samples to our proposed algorithm framework, Tables 2–5 show that the average recognition rates of our proposed algorithm framework have little difference for the same database.
(5) Figs. 2, 4 and 6 show that the average recognition rates of the proposed algorithm framework are superior to those of the D-KSVD and LC-KSVD algorithms. In particular, both the D-KSVD and LC-KSVD algorithms are supervised algorithms and use the label information in the dictionary learning process. Our proposed algorithm framework uses only the original training samples to generate virtual face images, noised face images or eigenface images for improving the robustness of the dictionary. However, when the number of atoms is smaller than 228, Fig. 8 shows that the average recognition rates of the proposed algorithm framework are lower than that of the LC-KSVD algorithm. When the number of atoms is smaller than 456, Fig. 8 shows that the average recognition rates of the proposed algorithm framework are lower than that of the D-KSVD algorithm. This partly shows that the proposed algorithm framework can outperform conventional dictionary learning algorithms under the condition of a small scale training sample set and a large scale dictionary, but in other cases it may not perform so well.
(6) Tables 2–5 show that our proposed algorithm framework requires less training time than the D-KSVD, LC-KSVD, PDPL and SVGDL algorithms, and takes slightly more training time than the K-SVD algorithm. The K-SVD, D-KSVD and LC-KSVD algorithms and our proposed algorithms have shorter testing times than those of the SRC, LRC, CNRC and LLC algorithms. Since the K-SVD, D-KSVD and LC-KSVD algorithms and our own proposed algorithms use the same linear classification method, the testing time was nearly equal. Moreover, the testing time of our proposed algorithm framework is slightly longer than that of the PDPL and SVGDL algorithms.

5. Conclusions and discussions

In this paper, a novel idea for robust and mathematically tractable dictionary learning algorithm is proposed. Moreover, the paper designs a framework that allows various schemes to produce “alternative training samples” to be used. Extensive experiments demonstrate that the proposed algorithm framework is robust and outperforms some previous state-of-the-art dictionary learning and sparse coding algorithms in classification accuracy.

Another advantage of the proposed algorithm framework is that it is not only applicable for face recognition but also can be applied to other pattern classification issues. Actually, once reasonable alternative training samples are available, our framework can be directly applied to other pattern classification issues. For example, the following scheme can also be applied to general pattern classification issues: alternative training samples are first obtained by corrupting the original training samples. Then the proposed algorithm is applied to the original and alternative training samples. As presented earlier, among potential possible schemes, one very simple scheme is to divide the set of original training samples into two groups and to regard one group as original training samples and the other group as alternative training samples. It seems that this scheme is highly suitable for cases where there is a large number of original training samples.

Acknowledgements

The work reported in this paper is partly supported by the Natural Science Foundation of China under grants nos. 61370163, 61233011 and 61332011, the Science and Technology Development Fund of Macao SAR (FDCT/128/2013/A) as well as the Shenzhen Municipal Science and Technology Innovation Council under grants Nos. JCYJ20150330155220591. This work is also partly supported by the Foundation for Young Talents in Higher Education of Guangdong, grants nos. 2015KQNCX08.

References

[1] M. Aharon, M. Elad, A.M. Bruckstein, K-SVD: an algorithm for designing of over-complete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (11) (2006) 4311–4322.
[2] B. Babagholami-Mohamadabadi, A. Zarghami, M. Zolfaghari, M.S. Baghshah, Probabilistic semi-supervised dictionary learning, in: Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 2013, pp. 192–207.
[3] D. Cai, X. He, J. Han, H. Zhang, Orthogonal Laplacian faces for face recognition, IEEE Trans. Image Process. 15 (11) (2006) 3608–3614.
[4] Y. Chen, J. Yang, L. Luo, H. Zhang, J. Qian, Y. Tai, J. Zhang, Adaptive noise dictionary construction via IRRPCA for face recognition, Pattern Recognit. 59 (2016) 26–41.
[5] S. Cai, W. Zuo, L. Zhang, X. Feng, P. Wang, Support vector guided dictionary learning, in: Proceedings of European Conference Computer Vision, 2014, pp. 624–639.
[6] A. Georghiades, P. Belhumeur, D. 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.
[7] S. Gu, L. Zhang, W. Zuo, X. Feng, Projective dictionary pair learning for pattern classification, in: Proceedings of Advances in Neural Information Processing Systems, 2014, pp. 793–801.
[8] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: a Database for Studying Face Recognition in Unconstrained Environments, University of Massachusetts, Amherst, 07-49, Oct. 2007 Technical Report.
[9] M. Jian, C. Jung, Semi-supervised bi-dictionary learning for image classification with smooth representation-based label propagation, IEEE Trans. Multimed. 18 (3) (2016) 458–473.
[10] 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.
[11] R. Jenatton, J. Mairal, G. Obozinski, F. Bach, Proximal methods for sparse hierarchical dictionary learning, in: Proceedings of International Conference on Machine Learning, 2010, pp. 487–494.

[12] X. Jing, F. Wu, X. Zhu, X. Dong, F. Ma, Z. Li, Multi-spectral low-rank structured dictionary learning for face recognition, Pattern Recognit. 59 (2016) 14–25.
[13] Z. Li, Z. Lai, Y. Xu, J. Yang, D. Zhang, A locality-constrained and label embedding dictionary learning algorithm for image classification, IEEE Trans. Neural Netw. Learn. Syst. (2015), doi:10.1109/TNNLS.2015.2508025.
[14] C. Lu, J. Shi, J. Jia, Robust online dictionary learning, in: Proceedings of IEEE International Conference on Computer Vision, 2013, pp. 415–422.
[15] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2011, pp. 1697–1704.
[16] A.M. Martinez, R. Benavente, The AR Face Database, CVC Technical Report #24, Jun.1998.
[17] Y. Ma, L. Wang, P. Liu, R. Ranjan, Towards building a data-intensive index for big data computing – a case study of remote sensing data processing, Inf. Sci. 319 (2015) 171–188.
[18] I. Naseem, R. Togneri, M. Bennamoun, Linear regression for face recognition, IEEE Trans. Neural Netw. Learn. Syst. 32 (11) (2010) 2106–2112.
[19] W. Ou, X. You, D. Tao, P. Zhang, Y. Tang, Z. Zhu, Robust face recognition via occlusion dictionary learning, Pattern Recognit. 47 (4) (2014) 1559–1572.
[20] D. Pham, S. Venkatesh, Joint learning and dictionary construction for pattern recognition, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
[21] Y. Quan, Y. Xu, Y. Sun, Y. Huang, Supervised dictionary learning with multiple classifier integration, Pattern Recognit. 55 (2016) 247–260.
[22] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326.
[23] T. Sim, S. Baker, M. Bsat, The CMU pose, illumination, and expression database, IEEE Trans. Pattern Anal. Mach. Intell. 25 (12) (2003) 1615–1618.
[24] M. Song, C. Chen, J. Bu, T. Sha, Image-based sketch-to-photo synthesis via online coupled dictionary learning, Inf. Sci. 193 (2012) 233–246.
[25] R. Sivalingam, A. D'Souza, M. Bazakos, R. Miezianko, V. Morellas, N. Papanikolopoulos, Dictionary learning for robust background modeling, in: Proceedings of IEEE International Conference on Robotics and Automation, 2011, pp. 4234–4239.
[26] A. Shrivastava, K.P. Jaishanker, V.M. Patel, R. Chellappa, Learning discriminative dictionaries with partially labeled data, in: Proceedings of IEEE International Conference on Image Processing, 2012, pp. 3113–3116.
[27] C. Sanderson, B.C. Lovell, Multi-region probabilistic histograms for robust and scalable identity inference, in: Proceedings of the 3rd IAPR/IEEE International Conference on Biometrics, 2009, pp. 199–208.
[28] J. Tenebaum, V. De Silva, J. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323.
[29] J. Tropp, A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Trans. Inf. Theory 53 (12) (2007) 4655–4666.
[30] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1) (1991) 71–86.
[31] J. Tang, L. Shao, X. Li, Efficient dictionary learning for visual categorization, Comput. Vis. Image Und. 124 (2014) 91–98.
[32] 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.
[33] H. Wang, F. Nie, W. Cai, H. Huang, Semi-supervised robust dictionary learning via efficient $\ell 2$, 0+-norms minimization, in: Proceedings of IEEE International Conference on Computer Vision, 2013, pp. 1145–1152.
[34] N. Wang, J. Wang, D.-Y. Yeung, Online robust non-negative dictionary learning for visual tracking, in: Proceedings of IEEE International Conference on Computer Vision, 2013, pp. 657–664.
[35] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 210–227.
[36] J. Wang, J. Yang, K. Yu, F. Lv, Locality-constrained linear coding for image classification, in: Proceedings of IEEE International Conference on Computer Vision, 2010, pp. 3360–3367.
[37] Y. Xu, X. Fang, X. Li, J. Yang, J. You, H. Liu, S. Teng, Data Uncertainty in Face Recognition, IEEE Trans. Cybern. 44 (10) (2014) 1950–1961.
[38] Y. Xu, X. Li, J. Yang, Z. Lai, D. Zhang, Integrating conventional and inverse representation for face recognition, IEEE Trans. Cybern. 44 (10) (2013) 1738–2746.
[39] K. Yu, Y. Lin, J. Lafferty, Learning image representations from the pixel level via hierarchical sparse coding, in: Proceedings of IEEE International Conference on Computer Vision, 2011, pp. 1713–1720.
[40] M. Yang, W. Liu, W. Luo, L. Shen, Analysis-synthesis dictionary learning for universality-particularity representation based classification, in: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, 2016, pp. 2251–2257.
[41] S. Yang, Y. Lv, Y. Ren, L. Yang, L. Jiao, Unsupervised images segmentation via incremental dictionary learning based sparse representation, Inf. Sci. 269 (2014) 48–59.
[42] M. Yang, L. Zhang, X. Feng, D. Zhang, Fisher discrimination dictionary learning for sparse representation, in: Proceedings of IEEE International Conference on Computer Vision, 2011, pp. 543–550.
[43] M. Yang, P. Zhu, F. Liu, L. Shen, Joint representation and pattern learning for robust face recognition, Neurocomputing 168 (2015) 70–80.
[44] M. Yang, L. Zhang, S.C.K. Shiu, D. Zhang, Gabor feature based robust representation and classification for face recognition with Gabor occlusion dictionary, Pattern Recognit. 46 (7) (2013) 1865–1878.
[45] Q. Zhang, B. Li, Discriminative K-SVD for dictionary learning in face recognition, in: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 2010, pp. 2691–2698.
[46] F. Zhu, L. Shao, Weakly-supervised cross-domain dictionary learning for visual recognition, Int. J. Comput. Vis. 109 (1) (2014) 42–59.
[47] F. Zhu, L. Shao, M. Yu, Cross-modality submodular dictionary learning for information retrieval, in: Proceedings of ACM International Conference on Information and Knowledge Management, 2014, pp. 1479–1488.
[48] L. Zhang, M. Yang, X. Feng, Sparse representation or collaborative representation: which helps face recognition? in: Proceedings of IEEE International Conference on Computer Vision, 2011, pp. 471–478.