

## Research Article

# A Dictionary Learning Algorithm Based on Dictionary Reconstruction and Its Application in Face Recognition

Shijun Zheng, Yongjun Zhang , Wenjie Liu, Yongjie Zou, and Xuexue Zhang

Key Laboratory of Intelligent Medical Image Analysis and Precise Diagnosis of Guizhou Province,  
College of Computer Science and Technology, Guizhou University, Guiyang, China

Correspondence should be addressed to Yongjun Zhang; [zyj6667@126.com](mailto:zyj6667@126.com)

Received 30 October 2019; Accepted 23 May 2020; Published 10 June 2020

Academic Editor: Matteo Gaeta

Copyright © 2020 Shijun Zheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In recent years, dictionary learning has received more and more attention in the study of face recognition. However, most dictionary learning algorithms directly use the original training samples to learn the dictionary, ignoring noise existing in the training samples. For example, there are differences between different images of the same subject due to changes in illumination, expression, etc. To address the above problems, this paper proposes the dictionary relearning algorithm (DRLA) based on locality constraint and label embedding, which can effectively reduce the influence of noise on the dictionary learning algorithm. In our proposed dictionary learning algorithm, first, the initial dictionary and coding coefficient matrix are directly obtained from the training samples, and then the original training samples are reconstructed by the product of the initial dictionary and coding coefficient matrix. Finally, the dictionary learning algorithm is reapplied to obtain a new dictionary and coding coefficient matrix, and the newly obtained dictionary and coding coefficient matrix are used for subsequent image classification. The dictionary reconstruction method can partially eliminate noise in the original training samples. Therefore, the proposed algorithm can obtain more robust classification results. The experimental results demonstrate that the proposed algorithm performs better in recognition accuracy than some state-of-the-art algorithms.

## 1. Introduction

In recent years, dictionary learning has been widely applied in various fields due to its excellent performance, such as face recognition [1–3], image denoising [4, 5] and blurring [6, 7], image segmentation [8, 9], and image recognition [10]. For face recognition [11, 12], the conventional dictionary learning method first learns a dictionary through the training samples. Then, given a test image, the image is represented by the atoms in the dictionary. Finally, the image is classified according to the result of the representation.

According to previous studies, using a dictionary obtained by the training samples to represent and classify test samples can lead to a higher accuracy than directly using training samples to represent and classify the test samples [13–15]. The dictionary learning method has achieved very significant performance in face recognition applications so

that researchers have proposed various dictionary learning methods. Luo et al. proposed a multiresolution dictionary learning method [16], which mainly resolves the problem that the conventional dictionary learning method only focuses on a single resolution. The method can effectively reduce the influence of noise and has better robustness. At the same time, the performance of face recognition has also been greatly improved. The multiview-like multiple vector representations can provide supplementary information for the representation object, and a robust dictionary learning method is proposed in [17]. Wang et al. used the low-rank structure of training data to construct a dictionary and proposed a new dictionary learning method, namely, Discriminative and Common hybrid Dictionary Learning (DCDL) [18]. The method overcomes two problems with the Sparse Representation-based Classification (SRC). One is that there is no major damage to the training samples, and the other is that each subject should have enough training

samples. Foroughi et al. proposed a multimodal structured low-rank dictionary learning method [19], which also has a good recognition rate and robustness under severe illumination changes and occlusions. In addition, sparse representation has achieved significant performance in face recognition. For example, Xu et al. proposed an improved collaborative representation [20]. The novel image representation and score fusion scheme is effective for classifying deformable objects, such as face images. Qian also proposed a method for merging virtual images and training samples [21]. Tang et al. proposed a novel algorithm, Distance Weighted Regression Classifier (DWLRC) [22], which solves the problem that current algorithms cannot fully utilize sample information.

To improve the discriminative ability of the dictionary, the researchers proposed various constraint models, and the most important ones are the local structure constraint model and the label constraint model [23]. In the field of sparse coding and dictionary learning, the use of the local structure of data is considered to be an important way to improve performance. In fact, the locality of the data can lead to the sparseness of the data, but not vice versa [24]. Therefore, many dictionary learning methods have added local constraints. For example, to maintain the structural characteristics of the training samples, Zheng et al. [25] proposed a Graph-SC algorithm to generate a Laplace graph of the training samples and used it to design the discriminant of the dictionary. Haghiri et al. [26] also proposed a discriminative dictionary learning method that preserves the local structure of the training samples. Zhang et al. [27] proposed the Locality-Constrained Projective Dictionary Learning (LC-PDL) algorithm, which adds the local constraint of atoms to maintain local information. Similarly, the label constraint also plays an important role in dictionary classification. More and more dictionary learning algorithms improve performance by adding the label constraint. For example, Shrivastava et al. [28] proposed a nonlinear discriminative dictionary learning method. The method can learn a dictionary by using both marked and unmarked data. The D-KSVD [29] algorithm adds the class label of the training samples to the discriminant, which ensures that the coding coefficients corresponding to the training samples of the same class also have a similarity. The algorithm greatly improves the discriminative ability of the dictionary. At the same time, the class label of the atom has also received more attention from researchers. The K-SVD [30] algorithm associates each atom with label information and introduces a new label consistency constraint. Considering the importance of the local structure and the label constraint, Li et al. [31] proposed the Locality-Constrained and Label Embedding Dictionary Learning Algorithm (LCLE-DL).

At present, most dictionary learning algorithms use the original training samples to generate a dictionary directly. However, the original training samples are generally collected at different angles, illuminations, and facial expressions. In other words, there are differences between different face images of the same object, and we can think that there is noise in the original image, which may reduce the discriminative ability of the dictionary. Actually, noise almost

always exists [32]. Therefore, this paper proposes the dictionary relearning algorithm (DRLA) based on Locality-Constrained and Label Embedding Dictionary Learning Algorithm (LCLE-DL), which can partially eliminate noise in the original image. The dictionary reconstruction method is as follows: first, assume that we have obtained the initial dictionary and encoding coefficient matrix. Then, the product of the initialized dictionary and the coding coefficient matrix replaces the original training samples matrix. Finally, the dictionary learning algorithm uses the updated training samples matrix to obtain the reconstructed dictionary and coding coefficient matrix, and the reconstructed result is used for subsequent image classification. We will explain the proposed method in detail in the subsequent chapters. The results of face recognition experiments show that the proposed method is feasible, which can reduce the influence of noise in the image and improve the accuracy of face recognition.

The other parts of the paper are organized as follows. Section 2 introduced the work related to the proposed algorithm. Section 3 explains the complete steps of the dictionary relearning algorithm (DRLA). Section 4 presents and analyzes the experimental results of the face database. Section 5 provides the conclusions.

## 2. Related Works

In this section, we introduce the work related to the dictionary relearning algorithm proposed in this paper. It mainly explains how to construct the local structure and the label embedding constraint. Besides, we introduce a discriminant dictionary model with both atomic local structure and label embedding constraints [31]. We also define some important symbols and definitions. We represent the training samples as  $Y = [Y_1, Y_2, \dots, Y_C] = [y_1, y_2, \dots, y_N] \in \mathfrak{R}^{n \times N}$ .  $C$  represents the number of classes of training samples.  $N$  is the number of training samples, and the dimension of each training sample is  $n$ . The dictionary is represented as  $D = [d_1, d_2, \dots, d_K] \in \mathfrak{R}^{n \times K}$ ,  $K$  is the number of atoms. The coding coefficient matrix is recorded as  $X = [x_1, x_2, \dots, x_N] \in \mathfrak{R}^{K \times N}$ ,  $x_i$  represents the coding coefficient corresponding to the training sample  $y_i$ , and  $x_i = [x_{i1}, x_{i2}, \dots, x_{iK}]^T$ . We refer to the transpose of the coefficient matrix  $X$  as the profiles matrix, denoted as  $P = X^T \in \mathfrak{R}^{N \times K}$ . That is, each row of the coding coefficient matrix represents a profile [33].

*2.1. Locality Constraint of Atoms.* Previous research [27, 30, 31] results show that the local constraint of atoms can effectively improve the discriminative ability of the dictionary and increase the robustness of the dictionary. There is a similar relationship between atoms and profiles. If two atoms are similar, then their corresponding profiles are similar, and vice versa. Moreover, all elements in the same profile correspond to the same atom. If the values of the  $m$ th and  $k$ th elements of the vector  $x_i$  are not zero, then the training sample  $y_i$  is reconstructed by the atoms  $d_m$  and  $d_k$ , and other atoms do not participate in the process. If profiles

$p_i$  is a nonzero vector and the  $s$ th and  $t$ th elements are not zero, this means that the training samples  $y_s$  and  $y_t$  are represented by the atom  $d_i$ . To maintain the local structural characteristics of the atom and inherit the structural features of the training samples, the paper [31] designed a discriminant dictionary model of the local constraint of atoms. Since the graph can effectively represent the similar relationship between the atoms and the profiles, a neighbor graph  $M$  can be constructed by the atoms. We can calculate the weight matrix of graph  $M$  by equation (1) [34], denoted as  $G$ :

$$g(i, j) = \begin{cases} \exp\left(-\frac{\|d_i - d_j\|}{\delta}\right), & \text{if } d_j \in \text{kNN}(d_i), \\ 0, & \text{else,} \end{cases} \quad (1)$$

where  $g(i, j)$  refers to the weight between the atoms  $d_i$  and  $d_j$ , which represents the degree of similarity between the two atoms,  $\text{kNN}(d_i)$  represents the  $k$  atoms closest to the atom  $d_i$ , and  $\delta$  is a parameter and is equal to 4 in all experiments. Through the previously constructed neighboring graph  $M$ , we can generate a Laplacian graph  $L$ . The equation for constructing a Laplacian graph is as follows:

$$L = A - G, \quad (2)$$

where  $A$  is a diagonal matrix, denoted as  $A = \text{diag}(a_1, a_2, \dots, a_k)$ , and  $a_i = \sum_{j=1}^k g(i, j)$ .

As the Laplace graph  $L$  is generated by the dictionary, it has better robustness and can better reflect the similarity of atoms [31]. According to the similar relationship between atoms and profiles, the profile matrix  $P = X^T$  and Laplacian graph  $L$  can be used to construct the discriminant of the dictionary, which can effectively maintain the local structure of atoms. The construction method of the discriminant dictionary model based on the locality constraint of atoms is as follows [31]:

$$\frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k \|p_i - p_j\|_2 g(i, j) = \text{Tr}(X^T L X). \quad (3)$$

**2.2. Label Embedding of Atoms.** The label constraint of atoms can improve the discriminative ability of the dictionary, which has been proved in previous studies [30, 31]. The atoms in the dictionary should have a different reconstruction performance when reconstructing the training samples. If some atoms reconstruct only one class of the training samples, then these atoms can be considered to belong to the same class [35]. The K-SVD algorithm can generate a specific class dictionary for each class of the training samples. Therefore, using K-SVD on the training samples can generate a dictionary containing the atoms of the  $C$  classes, denoted as  $D = [D_1, D_2, \dots, D_C]$ . If the atom  $d_i$  belongs to  $D_r$ , then the atom can be assigned a class label of  $r_i = (0, 0, \dots, 0, 1, 0, \dots, 0) \in \mathfrak{R}^C$ ,  $r_i$  is a  $C$ -dimensional row vector, and the value of the  $i$ th element is one, and the

values of the other elements are zero. Finally, we can get the class label vector for each atom and construct the label matrix of dictionary  $D$ , denoted as  $R = [r_1, r_2, \dots, r_K]^T \in \mathfrak{R}^{K \times C}$ , where  $K$  is the number of atoms in the dictionary.

After obtaining label matrix  $R$ , the weighted label matrix  $H$  of dictionary  $D$  can be obtained according to the following equation [36]:

$$H = R(R^T R)^{(-1/2)} \in \mathfrak{R}^{K \times C}. \quad (4)$$

According to [31], the profiles matrix  $P = V^T$  and the weighted label matrix  $H$  are used to construct the label embedding constraint model of the atom as follows:

$$\min_V \text{Tr}(V^T H H^T V) = \min_V \text{Tr}(V^T U V), \quad (5)$$

where  $V$  is the coding coefficient matrix and  $U$  is called the scaled label matrix. From equation (5), we can see that  $U = H H^T \in \mathfrak{R}^{K \times K}$ . Furthermore,  $U$  is a block diagonal matrix, which causes matrix  $V$  to be a similar structure. This structure increases the discrimination of coding coefficients. The label embedding constraint of atoms promotes the similarity of profiles corresponding to atoms of the same class.

Through the above locality and label embedding constraint, a discriminant dictionary model with both constraints can be constructed as follows:

$$\begin{aligned} \min_{D, X, V, L} & \|Y - DX\|_2^2 + \alpha \text{Tr}(X^T L X) + \|Y - DV\|_2^2 + \beta \text{Tr}(V^T U V) \\ & + \gamma \|X - V\|_2^2 \\ \text{s.t. } & d_i^2 = 1, \quad i = 1, \dots, K, \end{aligned} \quad (6)$$

where  $\alpha$ ,  $\beta$ , and  $\gamma$  are parameters. The second and fourth terms correspond to the locality constraint and the label embedding constraint of atoms, respectively. The fifth term ensures that the coding coefficient matrices  $X$  and  $V$  should be as close as possible so that the structural features and the discriminating information can be converted to each other.

### 3. The Dictionary Relearning Algorithm

In this section, we will explain in detail the proposed method, i.e., the dictionary relearning algorithm (DRLA).

The specific steps of the algorithm are as follows:

Step 1: the K-SVD algorithm is exploited to obtain the initial dictionary  $D$  and the coding coefficient matrix  $X$  based on the training samples  $Y$ . In particular, we apply the K-SVD algorithm to training samples of each class to generate a specific class dictionary  $D_i$ . Therefore, dictionary  $D$  is defined as  $D = [D_1, D_2, \dots, D_C] = [d_1, d_2, \dots, d_K]$ ,  $C$  is the number of classes, and  $K$  is the number of atoms. Meanwhile, the coding coefficient matrix obtained by the K-SVD algorithm can be expressed as  $X = [X_1, X_2, \dots, X_C] = [x_1, x_2, \dots, x_N]$ , where  $N$  is the number of the training samples.

Step 2: we record the label matrix of the training samples as  $T = [t_1, t_2, \dots, t_N]$ , where  $t_i$  is a  $C$ -dimensional column vector and  $N$  is the number of the training samples.  $t_i$  represents the class label vector of training sample  $y_i$ , a member of the training sample set of the  $i$ th class, then  $t_i = [0, \dots, 0, 1, 0, \dots, 0]^T$ . Only the value of the  $i$ th element of  $t_i$  is one, and the values of other elements are zero. When using the K-SVD algorithm to generate a specific class dictionary for each class of the training samples, we construct the class label of the atom of this class dictionary through the matrix  $T$ . Finally, we can get the label matrix  $R$  of dictionary  $D$ , and Section 2.2 describes the structure of  $R$ .

Step 3: we can construct the weight label matrix  $H$  of the dictionary  $D$  according to the label matrix  $R$ . The construction formula is equation (4), that is,  $H = R(R^T R)^{(-1/2)}$ . At the same time, the scaled label matrix  $U$  is obtained, and the calculation method is  $U = HH^T$ .

Step 4: we construct the Laplacian graph  $L$  using the initialized dictionary  $D$ . The construction methods are equations (1) and (2).

Step 5: in this step, we explain how to update the dictionary  $D$  and the sparse coding matrices  $X$  and  $V$ . According to the objective function (6) of the locality constraint and the label embedding constraint, we only regard the coding coefficient matrix  $V$  as a variable, and other variables in the equation are regarded as constants. Therefore, the formula for calculating  $V$  can be solved as follows:

$$V = (D^T D + \beta U + \gamma I)^{-1} (D^T Y + \gamma X). \quad (7)$$

Here,  $I$  is an identity matrix.

Similarly, we can get the update formula of the dictionary  $D$  and the coding coefficient matrix  $X$  as follows:

$$D = Y(X^T + V^T)(XX^T + VV^T)^{-1}, \quad (8)$$

$$X = (D^T D + \alpha L + \gamma I)^{-1} (D^T Y + \gamma V). \quad (9)$$

After obtaining the updated dictionary  $D$ , the Laplacian graph  $L$  is updated by equations (1) and (2). The step should be performed continuously until the termination condition is satisfied.

Step 6 (reconstruction of the training samples): the reconstruction method is implemented as follows:

$$Z = DX. \quad (10)$$

Step 7: using reconstructed training sample matrix  $Z$ , we reobtain dictionary  $D$  and coding coefficient matrices  $X$  and  $V$  by equations (7)–(9), respectively.

Step 8: classification parameter  $B$  is calculated and the test samples are classified. We record the label matrix of the test set as  $H_t$  and use it and the reobtained coding coefficient matrix  $X$  above to calculate classification parameter  $B$ . The calculation equation is as follows:

$$B = H_t X^T (X X^T + I)^{-1}. \quad (11)$$

For each test sample  $y$ , we use the OMP algorithm [37] to obtain its corresponding coding coefficient, denoted as  $x'$ . The classification result vector  $F$  is calculated by the following equation:

$$F = B x'. \quad (12)$$

Suppose that  $F = [f_1, \dots, f_k]$ . If  $f_t$  is the maximum value of  $F$ , then the test sample is classified into the  $t$ th class. The algorithm only needs to iterate from step 1 to step 8 once to obtain the classification result.

The scheme presented above for reconstruction of the training samples is inspired by the algorithm proposed in [38]. However, the algorithm in [38] is proposed for rerepresentation of a test sample by virtue of all training samples, whereas our scheme is devised for reconstruction of the training samples. In other words, the algorithm proposed in [38] uses the rerepresentation of a test sample for sparse representation, while the reconstruction of training samples in our algorithm is used for dictionary learning.

## 4. Experimental and Results

In this section, we conducted experiments on several widely used face databases. These face databases are AR face database (AR) [39], Extended Yale B face database [40], Labeled Faces in the Wild database (LFW) [41], and CMU PIE face database (PIE) [42]. To better reflect the advantages of dictionary reconstruction for the algorithm, we compare the proposed algorithm with other excellent algorithms, including classical dictionary learning and sparse representation algorithms. In all experiments, the proposed algorithm was compared with seven algorithms. They are LCLE-DL [31], LC-KSVD2 [30], D-KSVD [29], K-SVD [43], LLC [44], Sparse Representation-based Classification algorithm (SRC) [45], and Linear Regression Classification algorithm (LRC) [46]. The following is the experimental results and analysis of face databases.

In our experiments, all algorithms were executed ten times and then the average recognition rate and standard deviation were calculated. In the experiment table, the number in parentheses following the algorithm indicates the number of atoms, and the symbol  $\pm$  represents the standard deviation of the average recognition rate.



FIGURE 1: Examples of images in the AR face database.

TABLE 1: Average recognition rates on the AR face database.

Algorithm	Average recognition rates (%)
DRLA (960)	$83.1 \pm 0.011$
DRLA (840)	$82.6 \pm 0.010$
LCLE-DL (960)	$79.4 \pm 0.065$
LCLE-DL (840)	$79.3 \pm 0.065$
LC-KSVD2 (960)	$74.2 \pm 0.066$
LC-KSVD2 (840)	$78.5 \pm 0.065$
D-KSVD (960)	$74.1 \pm 0.067$
D-KSVD (840)	$69.3 \pm 0.056$
K-SVD (960)	$78.8 \pm 0.063$
K-SVD (840)	$25.5 \pm 0.048$
SRC (960)	$72.2 \pm 0.072$
LLC	$71.1 \pm 0.06$
LRC (960)	$69.7 \pm 0.074$

**4.1. Experiments on the AR Face Database.** The AR face database contains 126 people with more than 4,000 color images. The images are obtained under different angles, lighting, and facial expressions. Besides, the image of each subject is divided into occlusion and no occlusion, in other words, whether the subject has glasses or a scarf. In this experiment, we selected a subset of the AR face database for experimentation. The subset contains a total of 3120 images of 120 subjects, and each subject selects 26 images. Figure 1 shows examples of face images of a subject in the AR database.

In the experiment, for each object, we randomly selected eight images (including the first five images) as training samples. The remaining images of each subject are used as test samples. We set the parameters of the algorithm as  $\alpha = 1e - 4$ ,  $\beta = 1e - 3$ , and  $\gamma = 1e - 3$ . The experimental results of the average recognition rates of different algorithms are shown in Table 1. Here, we give the case where the number of atoms is 840 and 960.

From Table 1, we can see that the DRLA achieves the best classification performance when the number of atoms is 960 or 840. For example, when the number of atoms is 960, the average recognition rate of the DRLA is 83.1%, which is 3.7% and 4.3% higher than LCLE-DL and K-SVD, respectively. Similarly, the DRLA algorithm is superior to other algorithms in terms of average recognition rate. When the

number of atoms is 840, the average recognition rate of the DRLA is 82.6%. However, the average recognition rates of the LCLE-DL, LC-KSVD2, D-KSVD, and K-SVD algorithms were 79.3%, 78.5%, 69.3%, and 25.5%, respectively. It can be seen that the DRLA achieves better classification performance. Although DRLA can be viewed as one improvement to the locality-constrained and label embedding dictionary learning (LCLE-DL) algorithm, DRLA attains better recognition results.

Figure 2 shows the average recognition rates of the DRLA, LCLE-DL, LC-KSVD2, D-KSVD, and K-SVD algorithms as the number of atoms increases ( $K = 120, 240, \dots, 600, 720$ ). It can be clearly seen that the DRLA algorithm has obtained the best recognition accuracy.

**4.2. Experiments on the LFW Face Database.** In this section, we select a subset from the LEW database as the experimental dataset. The subset contains 1215 images with a total of 86 people and approximately 11–20 images per person. We adjusted each image to a  $32 \times 32$  pixel size. Figure 3 shows examples of face images in the LEW database.

In the experiment, we selected the first five images of each subject as training samples and randomly selected three images from the remaining images of each subject to join the training samples. In other words, each subject has eight training samples and the remaining samples as the test image. We set the parameters of the algorithm as  $\alpha = 1e - 2$ ,  $\beta = 1e - 2$ , and  $\gamma = 1e - 1$ . The experimental results on the LFW database are shown in Table 2, which shows the average recognition rates of different algorithms. Table 2 mainly shows the results when the number of atoms is 688. We also give the average recognition rate of the proposed algorithm when the number of atoms is 602.

From Table 2, we can see that when the number of atoms is 688, the average recognition rate of the DRLA is higher than other algorithms. For example, the average recognition rate of the DRLA is 38.9%, which is slightly higher than 38.8% of the LCLE-DL algorithm. Moreover, the recognition rate of the DRLA is significantly higher than 32.2% of LC-KSVD2, 33.4% of D-KSVD, and 34.8% of LLC. In addition, when the number of atoms is 602, the average recognition

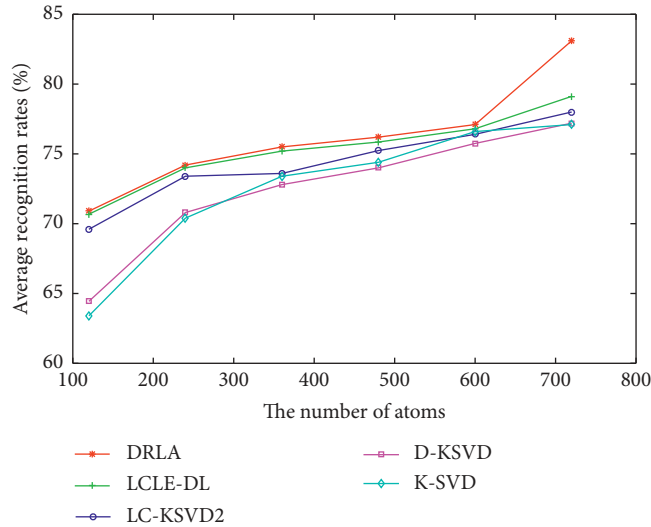


FIGURE 2: The average recognition rates with different numbers of atoms on the AR database.



FIGURE 3: Examples of images in the LFW face database.

TABLE 2: Average recognition rates on the LFW face database.

Algorithm	Average recognition rates (%)
DRLA (602)	$38.4 \pm 0.012$
DRLA (688)	$38.9 \pm 0.016$
LCLE-DL (602)	$36.8 \pm 0.013$
LCLE-DL (688)	$38.8 \pm 0.009$
LC-KSVD2 (688)	$32.2 \pm 0.012$
D-KSVD (688)	$33.4 \pm 0.016$
K-SVD (688)	$32.4 \pm 0.020$
SRC (688)	$38.1 \pm 0.011$
LLC	$34.8 \pm 0.011$
LRC (688)	$37.1 \pm 0.014$

rate of the DRLA is 38.4%, which is 1.6% higher than LCLE-DL. The above results show that the DRLA has good classification performance.

When the number of atoms ( $K=86, 172, \dots, 430, 516$ ) changes, the average recognition rates of the DRLA, LCLE-DL, LC-KSVD2, D-KSVD, and K-SVD algorithms are

shown in Figure 4. Obviously, the DRLA algorithm performs better than the other four algorithms.

**4.3. Experiments on the Yale B Face Database.** The images of The Extended Yale B database are collected under different lighting and facial expressions. The database contains 38 people with a total of 2,414 images and approximately 59–64 images per person. Each image is adjusted to a  $32 \times 32$  pixel size. Figure 5 shows examples of face images in the Yale B database.

In the experiment, for each subject, we randomly selected 20 images (including the first five images) as training samples. The remaining images of each subject are used for testing. We set the parameters of the algorithm as  $\alpha = 1e - 3$ ,  $\beta = 1e - 5$ , and  $\gamma = 1e - 3$ . On the Yale B database, the average recognition rates of different algorithms are shown in Table 3. We mainly compare the case when the number of atoms is 760.

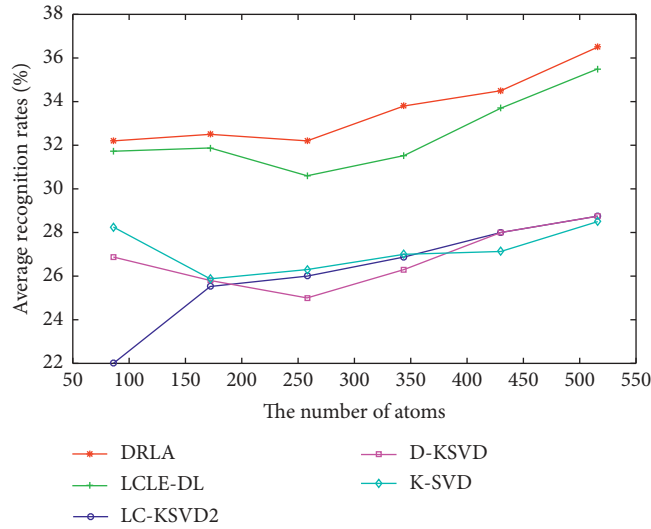


FIGURE 4: The average recognition rates with different numbers of atoms on the LFW database.



FIGURE 5: Examples of images in the Yale B face database.

TABLE 3: Average recognition rates on the Yale B face database.

Algorithm	Average recognition rates (%)
DRLA (760)	96.0 ± 0.008
DRLA (722)	95.6 ± 0.006
LCLE-DL (760)	95.8 ± 0.005
LCLE-DL (722)	95.4 ± 0.005
LC-KSVD2 (760)	92.7 ± 0.008
LC-KSVD2 (608)	92.9 ± 0.008
D-KSVD (760)	83.0 ± 0.026
D-KSVD (456)	94.3 ± 0.005
K-SVD (760)	95.3 ± 0.016
K-SVD (456)	94.0 ± 0.005
SRC (760)	95.3 ± 0.005
LLC	88.9 ± 0.010
LRC (760)	92.4 ± 0.008

From Table 3, when the number of atoms is 760 and 722, the average recognition rate of DRLA is improved by 0.2% compared with LCLE-DL. Similarly, the recognition rate of the DRLA is still the highest compared to other algorithms. For example, when the number of atoms is 760, the DRLA is 0.7% and 3.6% higher than SRC and LRC, respectively.

Figure 6 shows the average recognition rates of the DRLA, LCLE-DL, LC-KSVD2, D-KSVD, and K-SVD algorithms with different numbers of atoms ( $K = 38, 76, \dots, 760$ ). When the number of atoms increases, the average recognition rates of the DRLA, LCLE-DL, and LC-KSVD2 algorithms also gradually increases, but K-SVD performs unstable.

4.4. Experiments on the CMU PIE Face Database. We select the C05, C07, C09, C27, and C29 subsets of the PIE database as experimental data. The face database subset contains a total of 68 people, each with 170 images. The images of each subject were collected under different lighting and facial expressions. We resized all images to  $32 \times 32$  size. Figure 7 shows examples of face images in the PIE database.

In the experiment, we randomly selected ten images (including the first 5 images) of each subject as training samples, and the rest as test samples. We set the parameters of the algorithm as  $\alpha = 1e - 2$ ,  $\beta = 1e - 6$ , and  $\gamma = 1e - 2$ . The average recognition rates of different algorithms are shown in Table 4. When the number of atoms is 680, the DRLA algorithm achieves the best performance, which improves the average recognition rate by 5.4% compared to LCLE-DL.

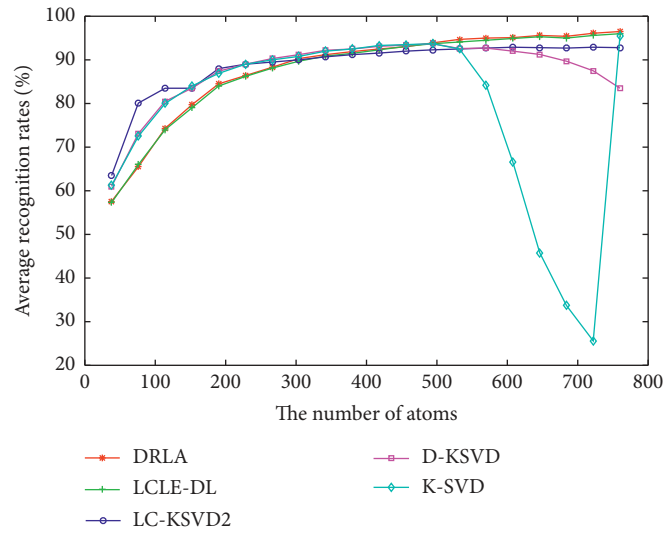


FIGURE 6: The average recognition rates with different numbers of atoms on the Yale B database.



FIGURE 7: Examples of images in the PIE face database.

TABLE 4: Average recognition rates on the PIE face database.

Algorithm	Average recognition rates (%)
DRLA (680)	81.0 ± 0.006
DRLA (544)	79.2 ± 0.006
LCLE-DL (680)	75.6 ± 0.009
LCLE-DL (544)	73.9 ± 0.012
LC-KSVD2 (680)	72.3 ± 0.009
D-KSVD (680)	71.9 ± 0.008
K-SVD (680)	72.4 ± 0.010
SRC (680)	72.1 ± 0.008
LLC	53.7 ± 0.016
LRC (680)	61.6 ± 0.021

As the number of atoms ( $K=68, 136, \dots, 408, 476$ ) increases, the average recognition rates of the DRLA, LCLE-DL, LC-KSVD2, D-KSVD, and K-SVD algorithms are shown in Figure 8. Compared with the other four algorithms, the performance of the DRLA algorithm is very superior.

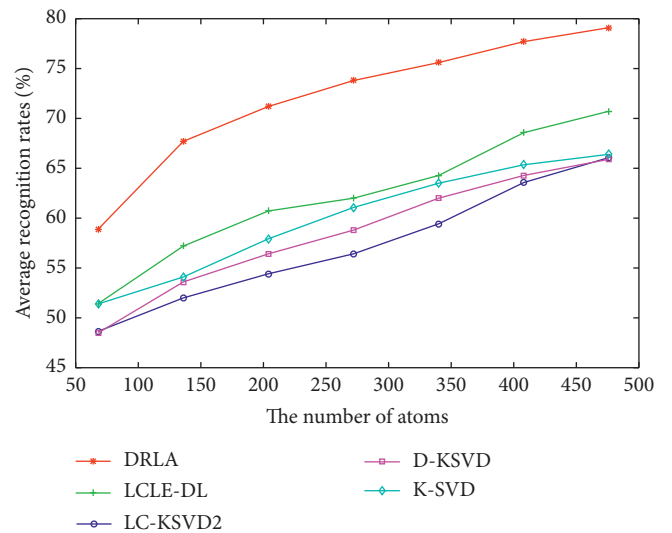


FIGURE 8: The average recognition rates with different numbers of atoms on the PIE database.



## 5. Conclusions

There are always noises in the original training samples. For example, differences in face images of the same object, which are caused by changes in lighting, facial expressions, and posture can be viewed as noise too in terms of viewpoint of pattern classification. We believe that these noises may reduce the discriminative ability of the dictionary. To reduce the influence of noise in the original training samples on dictionary performance, this paper proposes a dictionary relearning algorithm, which is implemented based on the locality constraint and the label embedding constraint. A large number of face experiments show that our proposed algorithm can eliminate some noise in the original training samples. This algorithm not only can effectively improve the discriminative ability of the dictionary but also enhance the robustness of the dictionary learning algorithm. In addition, we believe that the idea of dictionary reconstruction has a certain versatility, and it can be applied to a variety of different dictionary learning algorithms. The method is also an effective way to reduce the impact of image noise on the performance of the algorithm.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the Research Foundation for Advanced Talents of Guizhou University under Grant (2016) no. 49, Key Disciplines of Guizhou Province-Computer Science and Technology (no. ZDXK [2018]007), Key Supported Disciplines of Guizhou Province-Computer Application Technology (no. QianXueWeiHeZi ZDXK[2016]20), and National Natural Science Foundation of China (nos. 61462013 and 61661010).

## References

- [1] V. M. Patel, T. Wu, S. Biswas, P. J. Phillips, and R. Chellappa, "Dictionary-based face recognition under variable lighting and pose," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 954–965, 2012.
- [2] H. Liang, R. Liang, M. Song, and X. He, "Coupled dictionary learning for the detail-enhanced synthesis of 3-D facial expressions," *IEEE Transactions on Cybernetics*, vol. 46, no. 4, pp. 890–901, 2016.
- [3] G. Zhang, H. Sun, Z. Ji, Y.-H. Yuan, and Q. Sun, "Cost-sensitive dictionary learning for face recognition," *Pattern Recognition*, vol. 60, pp. 613–629, 2016.
- [4] R. Giryes and M. Elad, "Sparsity-based Poisson denoising with dictionary learning," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5057–5069, 2014.
- [5] Y. Fu, A. Lam, I. Sato, and Y. Sato, "Adaptive spatial-spectral dictionary learning for hyperspectral image denoising," in *Proceeding of the 2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 343–351, Santiago, Chile, December 2015.
- [6] L. Ma, L. Moisan, J. Yu, and T. Zeng, "A dictionary learning approach for Poisson image deblurring," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1277–1289, 2013.
- [7] S. Xiang, G. Meng, Y. Wang, C. Pan, and Z. Zhang, "Image deblurring with coupled dictionary learning," *International Journal of Computer Vision*, vol. 114, no. 2-3, pp. 248–271, 2015.
- [8] K. Cao, E. Liu, and A. K. Jain, "Segmentation and enhancement of latent fingerprints: a coarse to fine ridge structure dictionary," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 9, pp. 1847–1859, 2014.
- [9] S. Yang, Y. Lv, Y. Ren, L. Yang, and L. Jiao, "Unsupervised images segmentation via incremental dictionary learning based sparse representation," *Information Sciences*, vol. 269, pp. 48–59, 2014.
- [10] J. Wen, Y. Xu, Z. Li, Z. Ma, and Y. Xu, "Inter-class sparsity based discriminative least square regression," *Neural Networks*, vol. 102, pp. 36–47, 2018.
- [11] Y. Chen and J. Su, "Sparse embedded dictionary learning on face recognition," *Pattern Recognition*, vol. 64, pp. 51–59, 2017.
- [12] G. Lin, M. Yang, J. Yang, L. Shen, and W. Xie, "Robust, discriminative and comprehensive dictionary learning for face recognition," *Pattern Recognition*, vol. 81, pp. 341–356, 2018.
- [13] Q. Qiu, V. M. Patel, and R. Chellappa, "Information-theoretic dictionary learning for image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 11, pp. 2173–2184, 2014.
- [14] M. Harandi, R. Hartley, C. Shen, B. Lovell, and C. Sanderson, "Extrinsic methods for coding and dictionary learning on Grassmann manifolds," *International Journal of Computer Vision*, vol. 114, no. 2-3, pp. 113–136, 2015.
- [15] A. Fawzi, M. Davies, and P. Frossard, "Dictionary learning for fast classification based on soft-thresholding," *International Journal of Computer Vision*, vol. 114, no. 2-3, pp. 306–321, 2015.
- [16] X. Luo, Y. Xu, and J. Yang, "Multi-resolution dictionary learning for face recognition," *Pattern Recognition*, vol. 93, pp. 283–292, 2019.
- [17] Y. Xu, Z. Li, C. Tian, and J. Yang, "Multiple vector representations of images and robust dictionary learning," *Pattern Recognition Letters*, vol. 128, pp. 131–136, 2019.
- [18] C.-P. Wang, W. Wei, J.-S. Zhang, and H.-B. Song, "Robust face recognition via discriminative and common hybrid dictionary learning," *Applied Intelligence*, vol. 48, no. 1, pp. 156–165, 2017.
- [19] H. Foroughi, M. Shakeri, N. Ray, and H. Zhang, "Face recognition using multi-modal low-rank dictionary learning," in *Proceeding of the 2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, China, September 2017.
- [20] Y. Xu, B. Zhang, and Z. Zhong, "Multiple representations and sparse representation for image classification," *Pattern Recognition Letters*, vol. 68, pp. 9–14, 2015.
- [21] R. Qian, "Inverse transformation based weighted fusion for face recognition," *Multimedia Tools and Applications*, vol. 77, no. 21, pp. 28441–28456, 2018.
- [22] L. Tang, H. Lu, Z. Pang, Z. Li, and J. Su, "A distance weighted linear regression classifier based on optimized distance calculating approach for face recognition," *Multimedia Tools and Applications*, vol. 77, no. 22, pp. 32485–32501, 2019.

- [23] Y. Xu, Z. Li, J. Yang, and D. Zhang, "A survey of dictionary learning algorithms for face recognition," *IEEE Access*, vol. 5, pp. 8502–8514, 2017.
- [24] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Proceeding of the Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing Systems 2009*, pp. 2223–2231, Vancouver, Canada, December 2009.
- [25] M. Zheng, J. Bu, C. Chen et al., "Graph regularized sparse coding for image representation," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1327–1336, 2011.
- [26] S. Haghiri, H. R. Rabiee, A. Soltani-Farani, S. A. Hosseini, and M. Shadloo, "Locality preserving discriminative dictionary learning," in *Proceeding of the 2014 IEEE International Conference on Image Processing (ICIP)*, pp. 5242–5246, Paris, France, October 2014.
- [27] Z. Zhang, W. Jiang, Z. Zhang, S. Li, G. Liu, and J. Qin, "Scalable block-diagonal locality-constrained projective dictionary learning," in *Proceedings of the 2019 International Joint Conference on Artificial Intelligence*, China, August 2019.
- [28] A. Shrivastava, V. M. Patel, and R. Chellappa, "Non-linear dictionary learning with partially labeled data," *Pattern Recognition*, vol. 48, no. 11, pp. 3283–3292, 2015.
- [29] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proceeding of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2691–2698, San Francisco, CA, USA, June 2010.
- [30] Z. Jiang, Z. Lin, and L. S. Davis, "Learning a discriminative dictionary for sparse coding via label consistent K-SVD," in *Proceeding of the IEEE Computer Vision and Pattern Recognition 2011*, pp. 1697–1704, Providence, RI, USA, June 2011.
- [31] Z. Li, Z. Lai, Y. Xu, J. Yang, and D. Zhang, "A locality-constrained and label embedding dictionary learning algorithm for image classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 2, pp. 278–293, 2017.
- [32] C. Tian, Y. Xu, and W. Zuo, "Image denoising using deep CNN with batch renormalization," *Neural Networks*, vol. 121, pp. 461–473, 2020.
- [33] M. Sadeghi, M. Babaie-Zadeh, and C. Jutten, "Learning overcomplete dictionaries based on atom-by-atom updating," *IEEE Transactions on Signal Processing*, vol. 62, no. 4, pp. 883–891, 2014.
- [34] A. Shaban, H. R. Rabiee, M. Farajtabar, and M. Ghazvininejad, "From local similarity to global coding: an application to image classification," in *Proceeding of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2794–2801, Portland, OR, USA, June 2013.
- [35] Y. Zhang, Z. Jiang, and L. S. Davis, "Learning structured low-rank representations for image classification," in *Proceeding of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 676–683, Portland, OR, USA, June 2013.
- [36] J. Ye, Z. Zhao, and M. Wu, "Discriminative K-means for clustering," in *Proceeding of the Advances in Neural Information Processing Systems 20: 21st Annual Conference on Neural Information Processing Systems 2007*, pp. 1649–1656, Vancouver, Canada, December 2007.
- [37] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [38] Y. Xu, D. Zhang, J. Yang, and J.-Y. Yang, "A two-phase test sample sparse representation method for use with face recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 9, pp. 1255–1262, 2011.
- [39] A. Martínez and R. Benavente, "The AR face database," Technical report 24, Computer Vision Center, Barcelona, Spain, 1998.
- [40] A. S. Georghiadis, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.
- [41] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: a database for studying face recognition in unconstrained environments," Technical report 07-49, College Computer Science, University of Massachusetts, Amherst, MA, USA, 2007.
- [42] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression database," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1615–1618, 2003.
- [43] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4311–4322, 2006.
- [44] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proceeding of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3360–3367, San Francisco, CA, USA, June 2010.
- [45] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [46] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2106–2112, 2010.