DISCRIMINATIVE DICTIONARY PAIR LEARNING FOR IMAGE CLASSIFICATION

NGUYEN HOANG VU[1,*], TRAN QUOC CUONG[1], TRAN THANH PHONG[2]

[1]*Faculty of Industrial Engineering, Tien Giang University*

[2]*Office of Scientific Research and Technology & International Cooperation,*

Tien Giang University

Crossref
Similarity Check
Powered by iThenticate

Abstract. Dictionary learning (DL) for sparse coding has been widely applied in the field of computer vision. Many DL approaches have been developed recently to solve pattern classification problems and have achieved promising performance. In this paper, to improve the discriminability of the popular dictionary pair learning (DPL) algorithm, we propose a new method called discriminative dictionary pair learning (DDPL) for image classification. To achieve the goal of signal representation and discrimination, we impose the incoherence constraints on the synthesis dictionary and the low-rank regularization on the analysis dictionary. The DDPL method ensures that the learned dictionary has a powerful discriminative ability and signals are more separable after coding. We evaluate the proposed method on benchmark image databases in comparison with existing DL methods. The experimental results demonstrate that our method outperforms many recently proposed dictionary learning approaches.

Keywords. Dictionary learning; Synthesis and analysis dictionary; Incoherent dictionary; Classification; Face recognition.

1. INTRODUCTION

Dictionary learning (DL) for sparse coding has attracted a lot of attention in recent years and achieved great success in various application areas. Many previous studies used the original training samples as a dictionary to reconstruct the test samples, and achieved impressive results in comparison with many well-known image classification algorithms [11, 14, 27, 28]. However, research has demonstrated that learning a desired dictionary from training samples can well represent the given signal and it has led to state-of-the-art results in many practical applications, such as image de-noising [7], face recognition [13], and image classification [14]. Most of the existing supervised dictionary learning approaches could be mainly divided into three categories: synthesis dictionary learning [1], analysis dictionary learning [22], and analysis-synthesis dictionary pair learning [10]. Synthesis dictionary represents an input signal by using a linear combination of dictionary atoms, while analysis dictionary directly transforms a signal to a sparse feature space by multiplying the signal, which provides a complementary view of data representation. Analysis-synthesis dictionary representation can reconstruct a signal with analysis coding coefficient, which can be fast computed by a linear projection. In DL model, the discriminating ability of the dictionary atoms will determine the accuracy of the linear reconstruction over the atoms. Therefore,

---

*Corresponding author.

E-mail addresses: nguyenhoangvu@tgu.edu.vn (N.H. Vu); tranquoccuong@tgu.edu.vn (T.Q. Cuong); tranthanhphong@tgu.edu.vn (T.T. Phong).

many discriminative dictionary learning (DDL) methods have been proposed based on basic sparse model. The general DDL method is designed to learn the dictionary by combining reconstruction and discrimination term into objective function to improve the discriminative power of the learned dictionary for the classification tasks. One popular strategy is the structured discriminative DL that aims to learn a dictionary shared by all classes while forcing the resulting coding coefficients to be discriminative [12, 34]. The other one is to learn a class-specific dictionary and encourage each sub-dictionary to correspond to a single class label so that the class-specific reconstruction error could be used for classification [20, 21, 26, 29].

For the above dictionary learning methods, the dictionary learned with low coherence or incoherence between atoms in the dictionary is an important condition for sparse signal recovery. Several techniques have been proposed to improve the incoherence of a learned dictionary. Mailh *et al.* [18] proposed the incoherent K-SVD (IK-SVD) based on the addition of a decorrelation step to the K-SVD algorithm. Lin *et al.* [17] proposed an incoherent dictionary learning (IDL) model by incorporating the mutual incoherence between any two basis atoms into the learning process, which aims to increase the discrimination capacity of the learned dictionary. Ramirez *et al.* [21] proposed dictionary learning with structured incoherence (DLSI) method by encouraging each sub-dictionary to be as independent as possible. Chen *et al.* [5] adopts the low-rank recovery technique and a structural incoherence term to enforce the resulting low-rank dictionary for each class to be independent. To remove sparse noises like illumination changes and occlusions in corrupted face images, Yin *et al.* [33] presented low rank matrix recovery with structural incoherence and low rank projection (LRSI-LRP) model. Although various algorithms have been proposed to improve the efficiency of the DL method, the optimization for sparse coding is still a big computation burden for training the dictionary and testing the query sample.

Recently, Gu *et al.* [10] proposed a projective dictionary pair learning (DPL) algorithm which jointly learned a synthesis dictionary and an analysis dictionary for image classification. In the DPL model, the discrimination sparse code is replaced by the multiplication of the analysis dictionary and the input data. Compared with the traditional synthesis supervised dictionary learning methods, the DPL method could achieved higher recognition rate with lower time complexity. Based on the DPL model, Chen and Gao [6] proposed discrimination projective dictionary pair learning (DPDPL) for face recognition. Yang *et al.* [32] proposed a shared and specific-class analysis-synthesis dictionary learning algorithm for image classification. Yang *et al.* [31] proposed the Fisher discrimination dictionary pair learning (FDDPL) for image classification. By jointly learning a classifier with the dictionary pair, Yang *et al.* [30] explored a discriminative analysis-synthesis dictionary learning (DASDL) model. Chen *et al.* [4] proposed discriminative dictionary pair learning method based on differentiable support vector function (DPL-SV) for visual recognition. Li *et al.* [15] proposed a discriminative low-rank analysis-synthesis dictionary learning (LR-ASDL) algorithm with the adaptively ordinal locality preserving (AOLP) term and low-rank model for object classification. To preserve the locality property of learned atoms in the synthesis dictionary, Zhang *et al.* [35] proposed a locality constrained projective dictionary learning (LC-PDL) method. To achieve powerful representation ability of the available samples, Sun *et al.* [24] proposed a structured robust adaptive dictionary pair learning (RA-DPL) framework. Although those dictionary learning algorithms could achieve promising performance

in classification tasks, the discriminative ability of the learned dictionary still needs to be improved.

In this paper, we focus on improving the discriminative ability of analysis-synthesis dictionary and propose a discriminative dictionary pair learning (DDPL) algorithm for face recognition. The major contributions of this paper include:

(1) We propose a discriminative dictionary pair learning (DDPL) approach, DDPL integrates the synthesis discriminative dictionary learning, analysis representation into a unified model and learns a pair of synthesis-analysis dictionaries for each class.

(2) We consider the inter-class and intra-class incoherence constraints of the synthesis dictionary, which aim to minimize similarity between the dictionary atoms associated with different classes. As a result, class-specific dictionaries can be learnt from the optimizations and the sub-dictionaries are independent as much as possible.

(3) We design a low-rank regularization term, which requires that the learned analysis dictionary for each class should be low-rank, and therefore the obtained coding coefficients of samples from the same class are low-rank. This means that samples from the same class can have similar representations by using the learned analysis dictionary, which is beneficial to the following classification.

The DDPL method not only preserves the advantage of low computational complexity of DPL model, but also can learn a pair of dictionary with more discriminative power. Various comparisons between the proposed method and other dictionary learning methods in face recognition will be given to demonstrate the effectiveness of the proposed method. The remainder of this paper is organized as follows: Section 2 gives a brief review of some related work. Section 3 introduces a discriminative dictionary pair learning model. Experimental results are given in Section 4. Finally, Section 5 concludes this paper.

2. RELATED WORKS

As the goal of this paper is to develop an efficient dictionary learning algorithm for image classification, we first give the mathematical expression of the DDL method to illustrate the bases of supervised dictionary learning methods. Then, we will present the formulation of the DPL method which is related to our work. Given matrix $X = [X_1, X_2, \ldots, X_C]$ a set of m-dimensional training samples from C classes, where each is the training samples set of class ith, and n is the number of training samples of each class. Most of the state-of-the-art discriminative dictionary learning methods aim to learn an effective data representation model from X for classification tasks by exploiting the class label information of training data under the following framework

$$\min_{D,A} \|X - DA\|_F^2 + \lambda \|A\|_p + \Psi(D, A, Y), \tag{1}$$

where $\lambda \geq 0$ is a scalar constant, $D = [D_1, D_2, \ldots, D_C]$ is the synthesis dictionary to be learned $D_i \in R^{m \times p}$, and $A = [A_1, A_2, \ldots, A_C]$ is the coding coefficient matrix of X over D. In the training model (1), the data term $\|X - DA\|_F^2$ is the reconstruction residual of D;

$\|A\|_p$ is the l_p-norm regularizer on A; Y represents the class label matrix of samples in X, and $\Psi(D, A, Y)$ stands for some discrimination function, which ensures the discrimination power of D and A. Based on the structure of D, current discriminative DL models can be categorized into three main types: the shared dictionary learning method, the class-specific dictionary learning method, and the hybrid dictionary learning method.

To achieve good performance of sparse coding in various classification tasks requires imposing some additional constraints of the dictionary. One of such essential dictionary properties is the so-called mutual coherence. The mutual coherence $\mu(D)$ of a dictionary D is defined as the maximum absolute inner product between two distinct atoms [25]

$$\mu(D) = \max_{i \neq j} |\langle d_i, d_j \rangle|, \tag{2}$$

where d_i and d_j represents two different normalized dictionary atoms, $\mu(D) \in [0, 1]$. The value of $\mu(D)$ can reflect the similarity between the atoms in a certain extent. If it is bigger, then the similarity between the atoms will be stronger. Otherwise, the similarity is weaker. In the incoherent dictionary learning [17], an incoherence promoting term is introduced to make the atoms of the learned dictionary as independent as possible. Hence, it contributes to the increasing of the discrimination capacity of the learned dictionary. The incoherent promoting term is defined as a correlation measure between the atoms of D

$$\mathrm{cor}(D) = \left\| D^T D - I \right\|_F^2, \tag{3}$$

where I is an identity matrix. The dictionary D is said to be most incoherent if the correlation measure is zero, *i.e.*, all the atoms of D are orthonormal to each other. Minimizing the incoherent term guarantees that the dictionary can efficiently represent the input samples and achieve higher accuracies for classification tasks. However, most of the DDL models utilized the l_p-norm ($p = 0 \ or \ 1$) sparsity regularizer on the representation coefficients to obtain the robust classification results, the minimization of l_0 or l_1 norm is very complicated.

Different from the conventional discriminative dictionary learning model, Gu *et al.* [10] extended the conventional problem (1) into the DPL model by learns a synthesis dictionary D and an analysis dictionary P such that the code A can be analytically obtained as $A = PX$, thus the representation of X would become very efficient. The DPL model is defined as follows

$$\min_{P,D} \sum_{i=1}^{C} \|X_i - D_i P_i X_i\|_F^2 + \lambda \left\| P_i \overline{X_i} \right\|_F^2 \quad s.t \quad \|d_j\|_2^2 \leq 1, \tag{4}$$

where $\bar{X}_i$ denotes the complementary data matrix of X_i in the whole training set X; $D = [D_1, D_2, ..., \ D_C], D_i \in R^{m \times p}$ represents the i-th class of synthesis dictionary D, and $P = [P_1, P_2, ..., P_C], P_i \in R^{p \times m}$ is the corresponding analysis sub-dictionary in analysis dictionary P. The matrices D_i and P_i were used for classification. The classification scheme of the DPL model is decided by the reconstruction residual. For a test image y, the label of y is decided by

$$\mathrm{identity}(y) = \arg \min_i \|y - D_i P_i y\|_2. \tag{5}$$

Although the DPL model can reduce the computational complexity than the conventional discriminative DL model and has better classification accuracy, it ignored the discrimination in the synthesis dictionary representation and analysis dictionary representation. In order

to better obtain the discrimination in the learned dictionary, we introduced the incoherent constraints and low-rank regularization into dictionary pair learning model.

3. DISCRIMINATIVE DICTIONARY PAIR LEARNING

3.1. Formulation of DDPL

The learned synthesis analysis dictionaries are used for image classification task, and thus they should own favorable discriminability. In order to improve the discriminative ability of the synthesis dictionary, we impose the incoherent constraints on each synthesis sub-dictionary D_i to minimize the correlation between the atoms of D_i. Without this constraint, each dictionary of every class only needs to best encode samples of its own class. Furthermore, we design a low-rank regularization term on the analysis dictionary, which ensures that each analysis sub-dictionary P_i is low-rank. Therefore, the obtained coding coefficients for each class have high similarity, which will facilitate the image classification. An intuitive explanation of the proposed method is shown in Figure 1. Figure 1 shows that the synthesis dictionaries of similar features have high coherence atoms. So, minimizing $\left\|D_i^T D_j\right\|_F^2$ results in improving discrimination ability. With the structured analysis dictionary, we desire that each sub-dictionary P_i is low-rank to make them more compact and encourage them to be as independent as possible.

Based on the above analysis, the objective function of our approach is designed as follows

$$\min_{P,D} \sum_{i=1}^{C} \|X_i - D_i P_i X_i\|_F^2 + \lambda \left\|P_i \overline{X_i}\right\|_F^2 \\ + \mu \|P_i\|_* + \eta_1 \sum_{i=1,i \neq j}^{C} \left\|D_i^T D_j\right\|_F^2 + \eta_2 \left\|D_i^T D_i - I\right\|_F^2 \quad s.t \quad \|d_j\|_2^2 \leq 1, \quad (6)$$

where $\mu \geq 0$, $\eta_1 \geq 0$, $\eta_2 \geq 0$ is a scalar constant; $\|P_i\|_*$ is the low-rank regularization term, here $\|.\|_*$ represents the nuclear norm of a matrix; $\sum_{i \neq j}^{C} \left\|D_i^T D_j\right\|_F^2$ is the incoherence term to encourage inter-class sub-dictionaries to be independent, (i.e., $D_i^T D_j \approx 0$, $\forall i \neq j$); the $\left\|D_i^T D_i - I\right\|_F^2$ term make to stabilize the learned dictionary for each class; $I \in R^{p \times p}$ is an identity matrix.

3.2. Optimization strategy of DDPL

The objective function in (6) is generally non-convex. We introduce a variable matrix A and relax (6) to the following problem

$$\{A^*, P^*, D^*\} = \arg \min_{P,D} \sum_{i=1}^{C} \|X_i - D_i A_i\|_F^2 + \tau \|P_i X_i - A_i\|_F^2 \\ + \lambda \left\|P_i \overline{X_i}\right\|_F^2 + \mu \|P_i\|_* + \eta_1 \sum_{i=1,i \neq j}^{C} \left\|D_i^T D_j\right\|_F^2 + \eta_2 \left\|D_i^T D_i - I\right\|_F^2 \quad s.t \quad \|d_j\|_2^2 \leq 1,$$

$$(7)$$

where τ is a scalar constant. Here, we optimize A, P and D class by class. First, we initialize D and P as random matrices with unit Frobenius norm for each column vector and
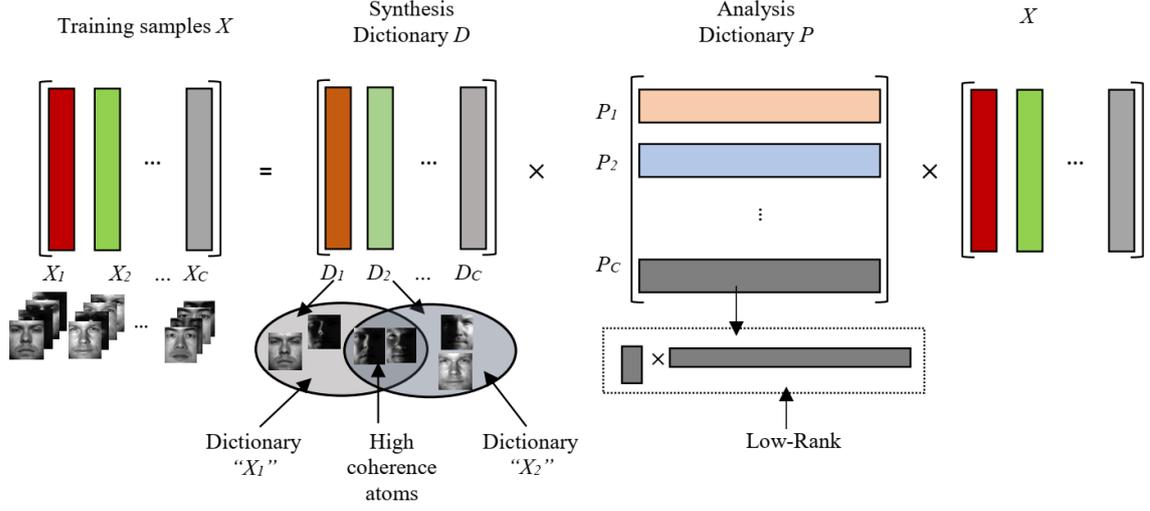
Figure 1. Illustration of DDPL model

then alternatively update A and $\{D, P\}$. The minimization can be alternated between the following two steps:

(1) Fixing D and P update A: When D and P are fixed, the objective function related to A can be written as

$$A^* = \arg\min_P \sum_{i=1}^{C} \|X_i - D_i A_i\|_F^2 + \tau \|P_i X_i - A_i\|_F^2 \tag{8}$$

This is a standard least squares problem, the closed-form solution for (8) can be obtained by taking the derivative and equating to zero.

$$A_i^* = (D_i^T D_i + \tau I)^{-1}(\tau P_i X_i + D_i^T X_i) \tag{9}$$

(2) Fixing A update P and D. When A are fixed, P and D can be updated by

$$P^* = \arg\min_P \sum_{i=1}^{C} \tau \|P_i X_i - A_i\|_F^2 + \lambda \|P_i \overline{X_i}\|_F^2 + \mu\|P_i\|_* \tag{10}$$

$$D^* = \arg\min_D \sum_{i=1}^{C} \|X_i - D_i A_i\|_F^2 + \eta_1 \sum_{i=1, i \neq j}^{C} \|D_i^T D_j\|_F^2 + \eta_2 \|D_i^T D_i - I\|_F^2 \; s.t \; \|d_j\|_2^2 \leq 1 \tag{11}$$

To address the optimization of problem (10), we transform it into the same minimization problem by introducing a relaxing variable Z

$$\{P^*, Z^*\} = \arg\min_{P,Z} \sum_{i=1}^{C} f(P_i) + \mu\|Z\|_* \quad s.t \quad P_i = Z \tag{12}$$

where $f(P_i) = \tau \|P_i X_i - A_i\|_F^2 + \lambda \|P_i \overline{X_i}\|_F^2$. Problem (12) can be addressed by solving the following Augmented Lagrange Multiplier problem

$$\min_{P_i, Z} f(P_i) + \mu\|Z\|_* + \langle T_1, P_i - Z \rangle + \frac{\varepsilon}{2} \|P_i - Z\|_F^2 \tag{13}$$

where $\varepsilon > 0$ is a penalty paramater whilst T_1 is the Lagrange multiplier. The optimal solution of (13) can be obtained by the ADMM algorithm [2]

$$\begin{cases} P_i^* = \arg\min_{P_i} \frac{1}{\varepsilon} f(P_i) + \frac{1}{2} \left\| P_i - Z + \frac{T_1}{\varepsilon} \right\|_F^2 \\ Z^* = \arg\min_{Z} \frac{\mu}{\varepsilon} \|Z\|_* + \frac{1}{2} \left\| Z - P_i + \frac{T_1}{\varepsilon} \right\|_F^2 \\ T_1 = T_1 + \varepsilon(P_i - Z) \end{cases} \quad (14)$$

A closed-form solution for P_i^* can be achieved by setting the derivative to zero. The optimization of Z^* can be solved with Singular Value Thresholding (SVT) [3].

The problem (11) can translated into the following form by introducing a variable T

$$\min_{D,T} \sum_{i=1}^{C} \|X_i - D_i A_i\|_F^2 + \eta_1 \sum_{i=1, i\neq j}^{C} \left\| D_i^T D_j \right\|_F^2 + \eta_2 \left\| D_i^T D_i - I \right\|_F^2 s.t. D = T, \|t_j\|_2^2 \leq 1 \quad (15)$$

where t_i denotes the ith column of T. The columns of T were normalized to avoid trivial solutions. The optimal solution of (15) can be obtained by the ADMM algorithm [2]:

$$\begin{cases} D_i^{k+1} = \min_{D_i} \|X_i - D_i A_i\|_F^2 + \eta_1 \left\| D_i^T D_j \right\|_F^2 + \eta_2 \left\| D_i^T D_i - I \right\|_F^2 + \rho \left\| D_i - T_i^k - S_i^k \right\|_F^2 \\ T_i^{k+1} = \min_{T_i} \rho \left\| D_i^{k+1} - T_i^k + S_i^k \right\|_F^2 \quad s.t. \quad \|t_j\|_2^2 \leq 1 \\ S_i^{k+1} = S_i^k + D_i^{k+1} - T_i^{k+1}, \text{update } \rho \text{ if appropriate} \end{cases}$$
$$(16)$$

where k is the iteration index and $0 < \rho < 1$ is a scalar that gradually increases at rate $\rho_{rate} \geq 1$. Closed-form solutions for (16) can be obtained by taking the derivatives of every sub-dictionary and equating to zero.

In each step of optimization, we have closed form solutions for variables A and P, and the ADMM based optimization of D converges rapidly. The DDPL algorithm is summarized in Algorithm 1. In Algorithm 1, when the difference between the energy in two adjacent iterations is less than 0.01 or the iteration limit reached, the iteration stops. The analysis dictionary P and the synthesis dictionary D are then output for classification.

Algorithm1: Discriminative dictionary pair learning (DDPL)

Input: Training sample $X = [X_1, X_2, \ldots, X_C]$, parameter $\lambda, \tau, \mu, \eta_1, \eta_2$;
1: Initialize D^0 and P^0 as random matrixes with unit Frobenious norm, $t = 0$;
2: while not converge do
3:         $t \leftarrow t + 1$;
4:         for $i = 1 : C$ do
5:                 update A_i by (9);
6:                 update P_i by (14);
7:                 update D_i by (16);
8:         end for
9:     end while

Output: Analysis dictionary P, synthesis dictionary D.

3.3. Classification scheme of DDPL

After the dictionary pair (D^*, P^*) are learned, we can perform the face recognition task as follows. Let y be a test image, if y belongs to class i, then $\|y - D_i^* P_i^* y\|_2^2$ would be smallest

$$\text{identity}(y) = \arg \min_i \|y - D_i^* P_i^* y\|_2^2. \qquad (17)$$

4. EXPERIMENTAL RESULTS

In this section, the performance of our proposed DDPL is evaluated on five image databases: Extended YaleB [9], AR [19] ORL [23], UMIST [16] and Caltech101 [8]. We compare our DDPL algorithm with some state-of-the-art dictionary learning algorithms including: Sparse Representation based Classification (SRC) [1], Discriminative K-SVD (DKSVD) [34], Fisher Discrimination Dictionary Learning for sparse representation (FDDL) [29], Dictionary Learning with Structured Incoherence (DLSI) [21], Label Consistent K-SVD (LC-KSVD) [12], LC-PDL [35], and DPL [10]. For SRC, DKSVD and DLSI, we implement them by ourselves. For the other algorithms, we use their published codes directly. All methods are programmed by Matlab.

4.1. Datasets

The extended YaleB database [9] contains 2414 frontal face images of 38 individuals, with images of each person taken under 64 different controlled lighting conditions. Some sample images of the Extended YaleB database are illustrated in Figure 2a. Random half of the images per class are selected for training and the other half for testing. The 504 dimensions feature provided by [12] is used to represent the face image. The dictionary contains 570 items, corresponding to an average of 15 items of each class.

The AR database [19] consists of over 4000 frontal images from 126 individuals. For each individual, 26 pictures were taken in two separate sessions, including different illumination conditions, different expressions and different facial disguises (sunglasses and scarves). Some sample images of AR database are illustrated in Figure 2b. Following the experimental setting of AR in [12], a set of 2600 face images of 50 female and 50 male classes is extracted. We randomly select 20 images of each class for training and the rest 6 images for testing. The feature dimensions is 540. The learned dictionary has 500 items, corresponding to an average of 5 items per category.

The ORL database [23] contains 400 images of 40 individuals (about 10 images per subject) taken under different lighting conditions, facial expression and accessories (see Figure 2c for example). We randomly select 6 images for each individual in the dataset for training and the remaining images for testing. In these experiments, we use random face features descriptors by [1] and set the dimension to 300. The learned dictionary has 240 atoms, or 6 atoms in each sub-dictionary.

The UMIST face database [16] consists of 564 cropped gray scale images of 20 subjects, each subject is taken in a range of poses from profile to frontal views as well as rate, gender and appearance. Figure 2d shows several sample images for one subject in the UMIST face database. We randomly chose 15 images for each individual for training set and the remaining images for testing set. Each face image is projected onto a 540- dimensional vector with a

(a) Extended Yale B

(b) AR
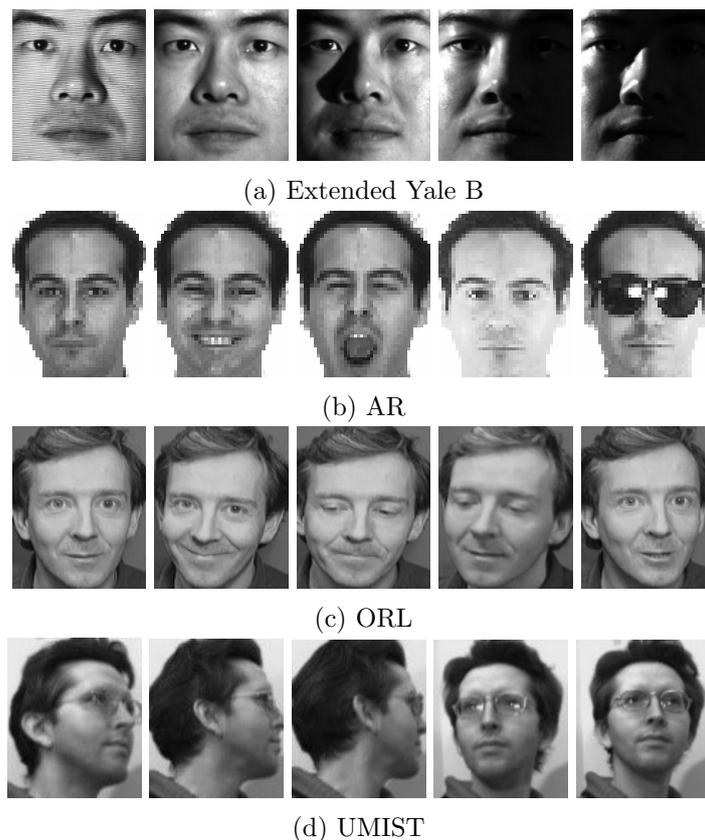
(c) ORL

(d) UMIST

Figure 2. Some sample images in (a) Extended Yale B, (b) AR, (c) ORL, and (d) UMIST

randomly generated matrix [1]. The number of dictionary atoms is set to be the number of training images, *i.e.*, the dictionary contains 300 items, corresponding to an average of 15 items of each class.

The Caltech101 database [8] contains 9144 images from 102 classes (*i.e.* 101 object classes and a background class) including animals, vehicles, flowers, etc. Some image samples of this dataset are shown in Fig. 3. The samples from each category have significant shape variability. The number of images in each category varies from 31 to 800. Following the common experimental settings, 30 samples per category are used for training and the rest are used for testing.

Figure 3. Some sample objects from the Caltech 101 database

4.2. Parameter settings

As shown in Eq. (7), there are five parameters (namely, $\lambda, \tau, \mu, \eta_1, \eta_2$) to be determined in the proposed DDPL model. First, we set the number of dictionary atoms as the number

of training images for all experiments. Experiments show that the parameters τ and λ have stable values for different experiments. So, we fix τ =0.05 and λ =3e-3 in all experiments. For the Extend Yale B, we fix $\mu = 0.001$. The impact of the parameter η_1 and η_2 on the classification accuracy is showed on Table 1. We can see that it achieves the best classification accuracy when η_1 =0.05 and η_2 =0.001. On the other hand, when η_1 =0.05 and η_2 =0.001 are fixed, the impact of the parameter μ on the classification accuracy is also showed on Table 2. It achieves the best classification accuracy when μ =0.001. So, for the Extend Yale B, the values of every parameter are set as follows μ =0.001, η_1 =0.05 and η_2 =0.001. Similarly, the parameters of DDPL on different databases are as follows: for the AR μ =0.005, η_1 =0.03 and η_2 =0.001; for the ORL μ =0.01, η_1 =0.01 and η_2 =0.005; and for the UMIST μ =0.003, η_1 =0.001 and η_2 =0.01.

Table 1. Impact of the parameters η_1 and η_2 on the classification accuracy when $\mu = 0.001$

η_1	0.001	0.01	0.05	0.1	0.15	0.2
η_2	0.00001	0.0001	0.001	0.01	0.1	0.15
Accuracy	97.7	97.9	98.1	95.6	91.5	80.9

Table 2. Impact of the parameter μ when $\eta_1 = 0.05$ and $\eta_2 = 0.001$

η_2	0.00001	0.0001	0.001	0.005	0.01	0.015
Accuracy	97.4	97.8	98.1	96.9	94.6	90.5

4.3. Convergence of DDPL model

Although the objective function in (7) is not jointly convex, it is convex when the others are fixed, *i.e.*, in each step of the optimization, the sub-problem is convex. We analyze the convergence behavior by describing the objective function values on four datasets. For YaleB and AR databases, we select 20 images from each subject for the training set and set the number of atoms corresponding to an average of 5 items per person. For the ORL and UMIST, we select 5 and 10 images from each subject for the training set and set the dictionary size as the number of training samples.

The convergence curves are shown in Figure 4. It can be seen that the values of the objective function can efficiently converge in limited iterations, usually within 20 iterations. That is, the proposed optimization algorithm has a good convergence property. Therefore, we set the number of iterations equal to 20 in all experiments of this paper.

4.4. Face recognition

The recognition results by different algorithms are showed in Table 3. From this table, we can observe that the accuracy of DDPL model is higher than others (*e.g.*, for Extended Yale B database, the accuracy of our proposed DDPL method is higher than others about 0.6%-4%, for AR is about 1%-10.1%, for ORL is about 0.9%-4.1%, and UMIST is about 2.8%-5.3%). We can also see that the accuracy of the DDPL is higher than the others in large datasets. Especially for UMIST dataset, when the number of classes in the training sample

(a) Extended Yale B

(b) AR
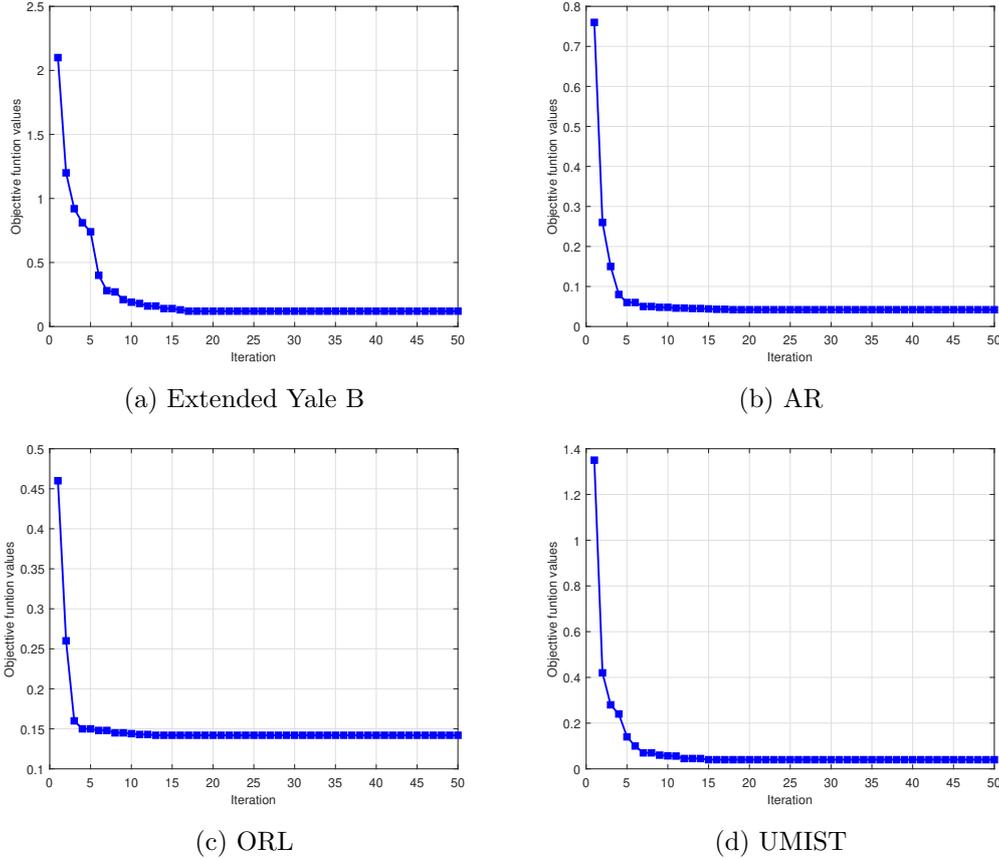
(c) ORL

(d) UMIST

Figure 4. The convergence curves of DDPL on the (a) Extended YaleB, (b)AR, (c) ORL and (d) UMIST

was small, DDPL clearly behaved more efficiently than all other recognition techniques, which proves the robustness of proposed method in this case. This demonstrates that by learning a synthesis dictionary and an analysis dictionary, the accuracy can be improved obviously. The most important reason of the higher accuracy of our DDPL method than the DPL model is that we both make the synthesis dictionary D and the analysis dictionary P to be more discriminative. The experimental result demonstrates the significant advantage of the DDPL method on large dataset.

Table 3. Recognition results (%) on the face datasets

Data set	SRC	DKSVD	LC-KSVD	DLSI	FDDL	DPL	LC-PDL	DDPL
Yale B	96.5	94.1	96.7	96.5	96.7	97.5	97.8	98.1
AR	97.5	88.8	97.8	97.5	97.5	98.3	98.6	98.9
ORL	94.6	93.6	95.6	96.2	96.3	96.8	96.9	97.7
UMIST	91.3	90.6	92.4	92.5	92.8	93.1	93.4	95.9

To evaluate the performance of the proposed algorithm, we observe the effect of dictionary size (the number of atoms) by comparing DDPL with DPL model. We fix the number

(a) Extended Yale B

(b) AR
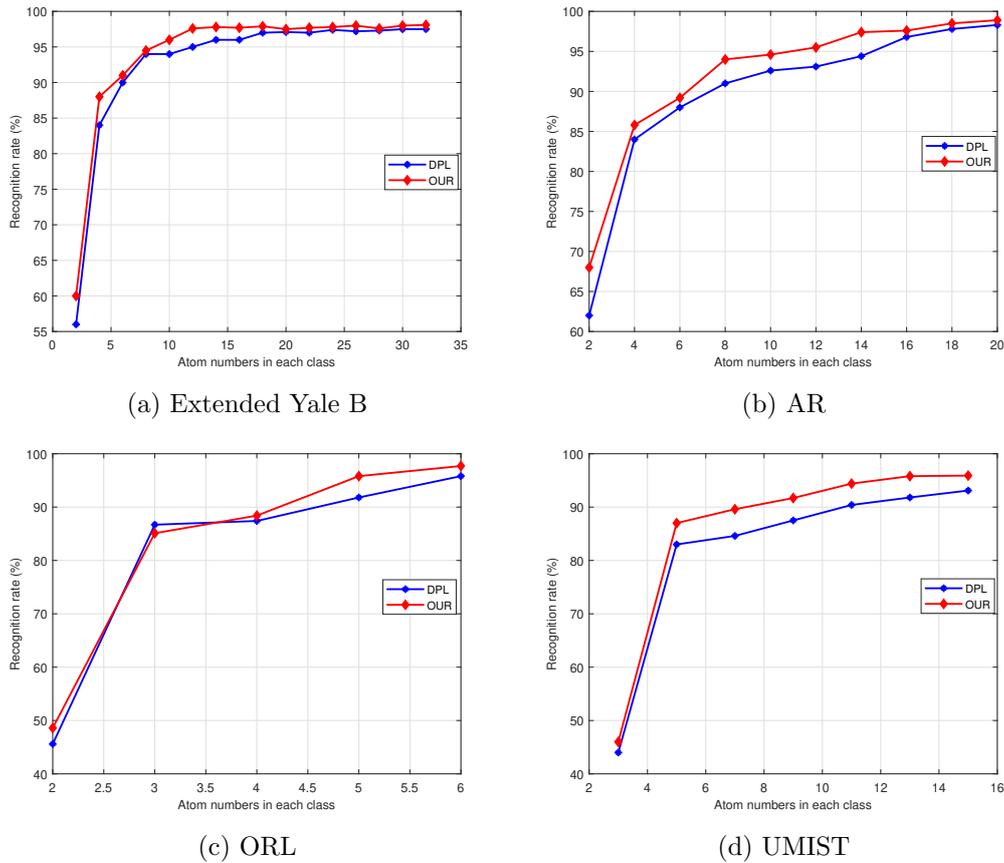
(c) ORL

(d) UMIST

Figure 5. The recognition rates (%) with different number of atoms on the (a) Extended YaleB, (b) AR, (c) ORL, and (d) UMIST

of the training samples, and then the atom numbers are varied with an integral multiple of the class number C. For the Extended Yale B database and the AR database, the atom numbers in each class are varied from 2 to 32 and from 2 to 20 with an interval of 2, respectively. For the ORL database, the atom numbers in each class are varied from 2 to 6 with an interval of 1. For the UMIST database, the atom numbers are varied from 3 to 15 with an interval of 3. Recognition rates versus different numbers of atoms are shown in Figure 5. It can be seen that the recognition rates of our DDPL algorithm are getting better with the increasing number of the atoms on the four databases. We could draw the conclusion that the accuracies of the proposed DDPL and DPL methods are improved with increasing number of the atoms, and the proposed DDPL method performs better than DPL method.

In order to further evaluate the effect of the low-rank regularization term and the incoherent constraint to our approach, we conduct DDPL with or without the low-rank term and the incoherent term. We call the version of DDPL without low-rank term as DDPL1 (*i.e.*, $\mu = 0$, and the parameters η_1, η_2 are varied in the range of [1e-5, 1]) and the version of DDPL without incoherent term as DDPL2 (*i.e.*, $\eta_1 = \eta_2 = 0$, and the parameter μ is varied in the range of [1e-5, 1]). Table 4 shows the comparison of recognition results on all datasets.

Table 4. Recognition accuracy (%) of DDPL, DDPL1, DDPL2, DLSI, and DPL

Datasets	DDPL	DDPL1	DDPL2	DLSI	DPL
Extended YaleB	98.1	97.9	97.6	96.5	97.5
AR	98.9	98.6	98.4	97.5	98.3
ORL	97.7	97.1	96.9	96.2	96.8
UMIST	95.9	94.6	93.8	92.5	93.1

We can see that both DDPL1 and DDPL2 achieve better results than DPL, which proves that the low-rank regularization and the incoherent term are meaningful and valuable. We can see that DDPL outperforms DDPL1 and DDPL2 at least by 0.5%, which means that our approach can obtain more favorable discriminative capability by employing the low-rank regularization term and the incoherent constraint. We evaluate the mutual coherence values of our method and other competing methods by measuring the coherence $\mu(D)$ of learned synthesis dictionary D on the UMIST database. The coherence $\mu(D)$ can be calculated using equation (18) as the maximal correlation of any two atoms from various classes

$$\mu(D) = \max_{d_i \in D_i, d_j \in D_j, i \neq j} \left| \left\langle \frac{d_i}{\|d_i\|_2}, \frac{d_j}{\|d_j\|_2} \right\rangle \right| \tag{18}$$

The mutual coherence values are illustrated in Figure 6. From Figure 6, it can see that, DLSI, DPL and DDPL algorithms have smaller coherence values than SRC, DKSVD, and LC-KSVD. Both DLSI and DDPL methods achieve the smallest coherence values, because both DLSI and our DDPL can learn the most independent sub-dictionaries. But since our DDPL method jointly learn a low-rank analysis dictionary and a synthesis discriminative for classification, it achieves higher recognition rate than DLSI.
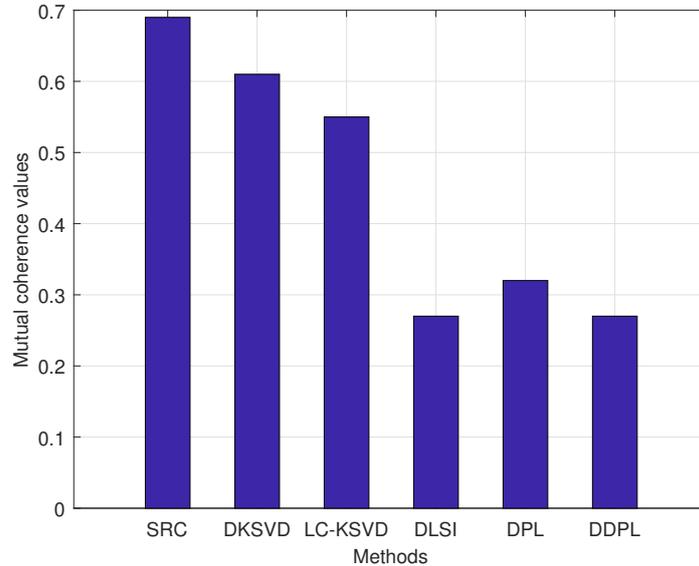
Figure 6. Coherence comparison of the algorithms on the UMIST database

4.5. Object recognition

In this section, we test DPL on object categorization by using the Caltech101 database [8]. Following the common experimental settings, 30 samples per category are used for training and the rest are used for testing. The experimental results are listed in Table 5. We can find that our presented DDPL algorithm can deliver better accuracies than its competitors on the used databases under the same setting.

Table 5. Recognition accuracy (%) on the Caltech101 database

Method	SRC	DKSVD	LC-KSVD	DLSI	FDDL	DPL	LC-PDL	DDPL
Accuracy	70.7	71.2	73.6	73.1	73.2	73.9	74.1	75.6

5. CONCLUSIONS

This paper presented a novel discrimination dictionary pair learning (DDPL) based dictionary learning method for face recognition. With the designed incoherence term and the low-rank regularization term, our model improved the representation and the discrimination abilities of existing projective dictionary pair learning. The advantage of the DDPL algorithm is that it combines incoherence constraints on the synthesis dictionary to minimize similarity between the dictionary atoms associated with different classes and the low-rank regularization on the analysis dictionary to improve the similarity between coding coefficients from the same class. Therefore, DDPL can ensure that the learned dictionary pair owns favorable discriminability. Experimental results on the public image database are given to demonstrate the superiority of the proposed model compared with other DL algorithms. The proposed DDPL not only can be used for face recognition but also can be applied to other pattern classification.

REFERENCES

[1] 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.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, pp. 1–122, 01 2011.

[3] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010. [Online]. Available: https://doi.org/10.1137/080738970

[4] B. Chen, J. Li, B. Ma, and G. Wei, "Discriminative dictionary pair learning based on differentiable support vector function for visual recognition," *Neurocomputing*, vol. 272, pp. 306 – 313, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S092523121731233X

[5] C. Chen, C. Wei, and Y. F. Wang, "Low-rank matrix recovery with structural incoherence for robust face recognition," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2618–2625.

[6] X. Chen and J. Gao, "Discrimination projective dictionary pair methods in dictionary learning1," in *2015 8th International Congress on Image and Signal Processing (CISP)*, 2015, pp. 204–208.

[7] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Transactions on Image Processing*, vol. 15, no. 12, pp. 3736–3745, 2006.

[8] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 59 – 70, 2007, special issue on Generative Model Based Vision. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077314206001688

[9] A. S. Georghiades, 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.

[10] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Projective dictionary pair learning for pattern classification," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 793–801. [Online]. Available: http://papers.nips.cc/paper/5600-projective-dictionary-pair-learning-for-pattern-classification.pdf

[11] R. He, W. Zheng, B. Hu, and X. Kong, "Two-stage nonnegative sparse representation for large-scale face recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 1, pp. 35–46, 2013.

[12] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent k-svd: Learning a discriminative dictionary for recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2651–2664, 2013.

[13] X.-Y. Jing, F. Wu, X. Zhu, X. Dong, F. Ma, and Z. Li, "Multi-spectral low-rank structured dictionary learning for face recognition," *Pattern Recognition*, vol. 59, pp. 14 – 25, 2016, compositional Models and Structured Learning for Visual Recognition. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320316000443

[14] 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.

[15] Z. Li, Z. Zhang, J. Qin, S. Li, and H. Cai, "Low-rank analysis?synthesis dictionary learning with adaptively ordinal locality," *Neural Networks*, vol. 119, pp. 93 – 112, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608019302011

[16] R. F. Li Fei-Fei and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 59 – 70, 2007, special issue on Generative Model Based Vision. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1077314206001688

[17] T. Lin, S. Liu, and H. Zha, "Incoherent dictionary learning for sparse representation," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, 2012, pp. 1237–1240.

[18] B. Mailhe, D. Barchiesi, and M. D. Plumbley, "Ink-svd: Learning incoherent dictionaries for sparse representations," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 3573–3576.

[19] A. Martinez and R. Benavente, "The ar face database," *Tech. Rep. 24 CVC Technical Report*, 01 1998.

[20] H. Nguyen, W. Yang, B. Sheng, and C. Sun, "Discriminative low-rank dictionary learning for face recognition," *Neurocomputing*, vol. 173, pp. 541 – 551, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231215010115

[21] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3501–3508.

[22] R. Rubinstein, T. Peleg, and M. Elad, "Analysis k-svd: A dictionary-learning algorithm for the analysis sparse model," *IEEE Transactions on Signal Processing*, vol. 61, no. 3, pp. 661–677, 2013.

[23] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, 1994, pp. 138–142.

[24] Y. Sun, Z. Zhang, W. Jiang, Z. Zhang, L. Zhang, S. Yan, and M. Wang, "Discriminative local sparse representation by robust adaptive dictionary pair learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2020.

[25] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, 2004.

[26] T. H. Vu and V. Monga, "Fast low-rank shared dictionary learning for image classification," *IEEE Transactions on Image Processing*, vol. 26, no. 11, pp. 5160–5175, 2017.

[27] 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.

[28] Y. Xu, D. Zhang, J. Yang, and J. 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.

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

[30] M. Yang, H. Chang, and W. Luo, "Discriminative analysis-synthesis dictionary learning for image classification," *Neurocomputing*, vol. 219, pp. 404 – 411, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231216310657

[31] M. Yang, H. Chang, W. Luo, and J. Yang, "Fisher discrimination dictionary pair learning for image classification," *Neurocomputing*, vol. 269, pp. 13 – 20, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231217309724

[32] M. Yang, W. Liu, W. Luo, and L. Shen, "Analysis-synthesis dictionary learning for universality-particularity representation based classification," in *AAAI*, 2016.

[33] H. Yin and X. Wu, "Face recognition based on structural incoherence and low rank projection," in *Intelligent Data Engineering and Automated Learning – IDEAL 2016*, H. Yin, Y. Gao, B. Li, D. Zhang, M. Yang, Y. Li, F. Klawonn, and A. J. Tallón-Ballesteros, Eds. Cham: Springer International Publishing, 2016, pp. 68–78.

[34] Q. Zhang and B. Li, "Discriminative k-svd for dictionary learning in face recognition," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2691–2698.

[35] 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 Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*. International Joint Conferences on Artificial Intelligence Organization, 7 2019, pp. 4376–4382. [Online]. Available: https://doi.org/10.24963/ijcai.2019/608