Class Relatedness Oriented Discriminative Dictionary Learning

Pengju Liu, Hongzhi Zhang, Kai Zhang, Changchun Luo, and Wangmeng Zuo(✉)

Center of Computational Perception and Cognition, School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
cswmzuo@gmail.com

Abstract. Discriminative dictionary learning (DDL) has recently attracted intensive attention due to its representative and discriminative power in various classification tasks. However, most of the existing DDL methods fall into two extreme cases, i.e., they either learn a global dictionary for all classes or train a class-specific dictionary, leading to less discriminative dictionary as the former do not consider correspondence between dictionary atoms and class labels while the latter ignore dictionary relatedness between different classes. To tackle this issue, in this paper we propose a well-principled DDL method which adaptively builds the relationship between dictionary and class labels. To be specific, we separatively impose a joint sparsity constraint on the coding vectors of each class to learn the class correspondence and relatedness for the dictionary. Experimental results on object classification and face recognition demonstrate that our proposed method can outperform many state-of-the-art DDL methods with more powerful and discriminative dictionary.

Keywords: Dictionary learning · Joint sparsity · $\ell_{1,\infty}$-norm · Support vector machine · Class relatedness

1 Introduction

Discriminative dictionary learning (DDL), with the goal of learning a dictionary to linearly represent the training data while enforcing the coding vectors or/and reconstruction error to be discriminative, has been successfully applied in pattern recognition applications such as image classification [1,2] and face recognition [3]. The success of DDL lies in that there usually exists a compact dictionary which can be learned from the available training data for more effective and efficient classification.

Different from unsupervised dictionary learning methods which only require the dictionary to faithfully represent training data, the DDL methods concentrate on discriminative classification capability of the dictionary as its goal is to assign correct class labels to test data. To enrich such capability, how to design relationship between dictionary atoms and class labels plays a vital role in dictionary training stage. Based on relationship between dictionary atoms and class

H. Zha et al. (Eds.): CCCV 2015, Part I, CCIS 546, pp. 335–343, 2015.
DOI: 10.1007/978-3-662-48558-3_34

labels, existing DDL methods can be divided into two main categories: one is global dictionary learning methods which associate each dictionary atom to all classes, the other is class-specific dictionary learning methods which assign each dictionary atom to only a single class. For global dictionary learning methods, the coding vectors are generally explored for classification and are usually jointly optimized with a classifier. Mairal *et al.* [4] proposed a DDL method by training a classifier of coding vectors for digging recognition and texture classification. Zhang and Li [5] proposed a joint learning algorithm base on KSVD for face recognition. Pham *et al.* [6] proposed to jointly train the dictionary and classifier for face recognition and object categorization. Cai *et al.* [7] introduced linear support vector machines (SVM) to jointly optimize the dictionary and classifier and thus making the coding vectors and dictionary more adaptive and flexible. Even though a global dictionary with small size can be powerful enough to represent training data and thereby the testing phase is very efficient, all the above methods fail to consider correspondence between dictionary atoms and class labels.

In the class-specific DDL methods, each dictionary atom is assigned to a single class and the dictionary atoms associated with different classes are encouraged to be as independent as possible. Ramirez *et al.* [1] proposed a structured dictionary learning scheme by promoting the discriminative ability between different class-specific sub-dictionaries. Castrodad and Sapiro [8] learned a set of class-specific sub-dictionaries with non-negative penalty on both dictionary atoms and coding vectors. Yang *et al.* [3] proposed a DDL framework which employs Fisher discrimination criterion to learn a class-specific dictionary. Since each dictionary atom has a single label, the reconstruction error with respect to each class could be used for classification. However, those methods ignored the dictionary relatedness across different classes, e.g., one dictionary atom can be helpful for the reconstruction of samples from different classes. Consequently, when there are numerous classes, the size of dictionary would be very large which will increase the memory and computational complexity for real applications.

As a matter of fact, the two DDL categories build relationship between dictionary atoms and class labels in two extreme manners. In order to make a trade-off to adaptively build the relationship, we propose a well-principled DDL scheme in which a joint sparsity constraint is separatively imposed on the coding vectors of each class by applying $\ell_{1,\infty}$-norm regularizer to the coding vectors of each class. Since the $\ell_{1,\infty}$-norm is a matrix norm that encourages entire rows of the matrix to be zeros, the resultant row sparsity of coding vectors of a specific class would build relationship between the specific class and the whole dictionary. Therefore, some samples can be sparsely represented by the dictionary atoms from the same and different classes. To make the coding vectors more discriminative, as in [7] we also add a discrimination term to the objective function which is formulated as sum of the weighted Euclidean distances between all pairs of coding vectors. What is more, a multi-class linear SVM classifier is incorporated into the DDL scheme to learn a dictionary in training phase and classify input samples in the testing phase.

The remainder of this paper is organized as follows. In Section 2, we briefly introduce the DDL scheme. In Section 3, we present the proposed DDL model and corresponding optimization procedure. To verify the efficiency of the proposed DDL method on classification problems, some experiments are conducted and the results are analyzed in Section 4. Finally, we conclude the paper in Section 5.

2 A Brief Review of the DDL Models

Suppose that $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_C]$ is a set of training samples with C classes, where $\mathbf{X}_c$ is the subset containing n_c samples from the c-th class. Correspondingly, let $\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_C]$ be the coding matrix of $\mathbf{X}$ over the dictionary. A general DDL model can be described as follows

$$\langle \mathbf{D}, \mathbf{A} \rangle = \arg\min_{\mathbf{D},\mathbf{A}} \mathcal{R}(\mathbf{X}, \mathbf{D}, \mathbf{A}) + \lambda_1 \|\mathbf{A}\| + \lambda_2 \mathcal{L}(\mathbf{A}), \tag{1}$$

where $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, ..., \mathbf{d}_K]$ is the dictionary, λ_1 and λ_2 are the trade-off parameters, $\mathcal{R}(\mathbf{X}, \mathbf{D}, \mathbf{A})$ is the reconstruction error term, $\|\mathbf{A}\|$ denotes a certain norm for $\mathbf{A}$, and $\mathcal{L}(\mathbf{A})$ denotes the discrimination term for $\mathbf{A}$.

In general, $\|\mathbf{A}\|$ is set to be $\|\mathbf{A}\|_1$ to ensure sparsity of the coding vectors since it tends to produce better classification results [9]. However, ℓ_1-norm sparse coding suffers from high computation burden. To tackle this problem, some researchers attempt to use ℓ_2-norm regularizer and their results can be very competitive with well-designed classification rule or classifier.

3 Main Results

Instead of learning a global dictionary without class specific property or a class-specific dictionary without class relatedness property, we propose to adaptively learn the relationship between dictionary atoms and class labels. In our proposed method, we simply assume a training sample can be sparsely represented by the dictionary atoms from the same and different classes. Under this assumption, we take advantage of $\ell_{1,\infty}$-norm which penalizes the sum of maximum absolute values of each row for a matrix. To be specific, the $\ell_{1,\infty}$-norm encourages entire rows of a matrix to be zeros and can be utilized for a joint sparse regularization. Thus, we replace $\|\mathbf{A}\|$ by $\sum_{c=1}^{C} \|\mathbf{A}_c\|_{1,\infty}$, where $\|\mathbf{A}_c\|_{1,\infty}$ forces that $\mathbf{X}_c$ should be jointly and sparsely represented by the dictionary and thus regularize the dictionary to have class relatedness property.

For the discrimination term, to enlarge similarity of coding vectors from same class and dissimilarity of coding vectors from different classes, we adopt sum of weighted Euclidean distances between all pairs of coding vectors to indicate the discrimination capability,

$$\mathcal{L}(\mathbf{A}) = \sum_{i,j} w_{ij} \|\mathbf{a}_i - \mathbf{a}_j\|_2^2, \tag{2}$$

where $\mathbf{a}_i$ and $\mathbf{a}_j$ denote the coding vectors of i-th and j-th sample, respectively, and w_{ij} is the associated weight which plays a key role in the discrimination term. It has been pointed out that with the symmetry, consistency and balance constraints on the weights, a multi-class linear SVM can be fused into the discrimination term. According to [7], Eq. (2) can be further rewritten as

$$\mathcal{L}(\mathbf{A}) = 2\sum_{c=1}^{C} \mathcal{L}(\mathbf{A}, \mathbf{y}^c, \mathbf{u}_c, b_c), \tag{3}$$

where $\mathbf{y}^c = [y_1^c, y_2^c, \ldots, y_n^c]$, n is the number of training samples, $y_i^c = 1$ if $y_i = c$ and otherwise $y_i^c = -1$, $\mathbf{u}_c$ is the normal to the c-th classs hyperplane of SVM, b_c is the corresponding bias. To be specific, $\mathcal{L}(\mathbf{A}, \mathbf{y}^c, \mathbf{u}_c, b_c) = \|\mathbf{u}_c\|_2^2 + \theta\ell(\mathbf{A}, \mathbf{y}^c, \mathbf{u}_c, b_c)$, where $\ell(\mathbf{A}, \mathbf{y}^c, \mathbf{u}_c, b_c)$ is the hinge loss function and θ is a predefined constant. For the reconstruction error term, we formulate it as $\mathcal{R}(\mathbf{X}, \mathbf{D}, \mathbf{A}) = \|\mathbf{X} - \mathbf{DA}\|_F^2$. Note that $\|\mathbf{X} - \mathbf{DA}\|_F^2 = \sum_{c=1}^{C} \|\mathbf{X}_c - \mathbf{DA}_c\|_F^2$. As a result, our model can be formulated as follows

$$\min_{\mathbf{D},\mathbf{A},\mathbf{U},\mathbf{b}} \sum_{c=1}^{C} \|\mathbf{X}_c - \mathbf{DA}_c\|_F^2 + \lambda_1 \sum_{c=1}^{C} \|\mathbf{A}_c\|_{1,\infty} + 2\lambda_2 \sum_{c=1}^{C} \mathcal{L}(\mathbf{A}, \mathbf{y}^c, \mathbf{u}_c, b_c), \tag{4}$$

where $\mathbf{U}$ and $\mathbf{b}$ are a collection of $\mathbf{u}_c$ and b_c, $c = 1, 2, \ldots, C$, respectively.

3.1 Model Training

Eq. (4) is a joint optimization problem and can be solved in an alternative minimization scheme [7]. Thus, we alternatively optimize the objective function with respect to $\mathbf{D}$, $\mathbf{A}$ and $\langle \mathbf{U}, \mathbf{b} \rangle$ as follows.

By fixing $\mathbf{D}$ and $\langle \mathbf{U}, \mathbf{b} \rangle$, we can separately calculate $\mathbf{A}_c$ by solving the following problem:

$$\langle \mathbf{A}_c \rangle = \arg\min_{\mathbf{A}_c} \|\mathbf{X}_c - \mathbf{DA}_c\|_F^2 + \lambda_1 \|\mathbf{A}_c\|_{1,\infty} + 2\lambda_2 \sum_{c=1}^{C} \mathcal{L}(\mathbf{A}_c, \mathbf{y}^c). \tag{5}$$

To efficiently solve Eq. (5), we introduce an auxiliary variable $\mathbf{A}'_c$, resulting in an equivalent problem as follows

$$\langle \mathbf{A}_c, \mathbf{A}'_c \rangle = \arg\min_{\mathbf{A}_c, \mathbf{A}'_c} \|\mathbf{X}_c - \mathbf{DA}_c\|_F^2 + \lambda_1 \|\mathbf{A}'_c\|_{1,\infty} + 2\lambda_2 \sum_{c=1}^{C} \mathcal{L}(\mathbf{A}_c, \mathbf{y}^c) + \frac{\mu}{2} \|\mathbf{A}_c - \mathbf{A}'_c\|_F^2, \tag{6}$$

where μ is a positive penalty parameter. We then use the augmented Lagrangian method to alternatively optimize $\mathbf{A}_c$ and $\mathbf{A}'_c$ until convergence as follow. (i) When $\mathbf{A}'_c$ is fixed, let $\mathbf{a}_i^c$ denotes the coding vector of i-th sample from c-th class and $\mathbf{a}_i'^c$ is the corresponding auxiliary variable, we can optimize $\mathbf{A}_c$ in columns,

$$\langle \mathbf{a}_i^c \rangle = \arg\min_{\mathbf{a}_i^c} \|\mathbf{x}_i^c - \mathbf{Da}_i^c\|_2^2 + 2\lambda_2\theta \sum_{c=1}^{C} \ell(\mathbf{a}_i^c, y_i^c) + \frac{\mu}{2} \|\mathbf{a}_i^c - \mathbf{a}_i'^c\|_2^2, \tag{7}$$

Algorithm 1 Algorithm of the proposed DDL model.

Input: $\mathbf{D}_{init}, \mathbf{A}_{init}, \mathbf{A}'_{init}, \mathbf{U}_{init}, \mathbf{b}_{init}, \lambda_1, \lambda_2, \theta, \mu$.
Output: $\mathbf{D}, \mathbf{U}, \mathbf{b}$

while not converged **do**
  for $c = 1$ **to** C **do**
    while not converged **do**
      for $i = 1$ **to** n_c **do**
        $\mathbf{a}_i^c \leftarrow \arg\min_{\mathbf{a}_i^c} \|\mathbf{x}_i^c - \mathbf{D}\mathbf{a}_i^c\|^2 + 2\lambda_2\theta\sum_{c=1}^{C} \ell(\mathbf{a}_i^c, y_i^c) + \frac{\mu}{2}\|\mathbf{a}_i^c - \mathbf{a}_i'^c\|_2^2$
      end for
      $\mathbf{A}'_c \leftarrow \arg\min_{\mathbf{A}'_c} \lambda_1\|\mathbf{A}'_\mathbf{c}\|_{1,\infty} + \frac{\mu}{2}\|\mathbf{A}'_c - \mathbf{A}_c\|_F^2$
    end while
  end for
  $\mathbf{D} \leftarrow \arg\min_{\mathbf{D}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 \;\; s.t. \;\; \|\mathbf{d}_k\|^2 \le 1, \; k = 1, 2, \ldots, K$.
  for $c = 1$ **to** C **do**
    $\mathbf{u}_c, b_c \leftarrow$ by multi-class linear SVM
  end for
end while

According to [7], $\mathbf{a}_i^c$ has a closed-form solution. (ii) When $\mathbf{A}_c$ is fixed, we get the following problem which can be solved by the projected gradient method [10],

$$\langle \mathbf{A}'_c \rangle = \arg\min_{\mathbf{A}'_c} \lambda_1\|\mathbf{A}'_\mathbf{c}\|_{1,\infty} + \frac{\mu}{2}\|\mathbf{A}'_c - \mathbf{A}_c\|_F^2. \tag{8}$$

By fixing $\mathbf{A}$ and $\langle \mathbf{U}, \mathbf{b} \rangle$, the optimization problem with respect to $\mathbf{D}$ is as follows

$$\langle \mathbf{D} \rangle = \arg\min_{\mathbf{D}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 \quad s.t. \quad \|\mathbf{d}_k\|_2^2 \le 1, \quad k = 1, 2, \ldots, K, \tag{9}$$

where the constraint is to avoid the scaling issue of the atoms. This problem can be solved effectively by the Lagrange dual method [11].

By fixing $\mathbf{D}$ and $\mathbf{A}$, we can update $\langle \mathbf{U}, \mathbf{b} \rangle$ by solving the following problem

$$\langle \mathbf{U}, \mathbf{b} \rangle = \arg\min_{\mathbf{U},\mathbf{b}} \sum_{c=1}^{C} \mathcal{L}(\mathbf{A}_\mathbf{c}, \mathbf{y}^c, \mathbf{u}_c, b_c). \tag{10}$$

Eq. (10) is actually a multi-class linear SVM problem and can be solved by the SVM solver in [12]. Algorithm 1 summarizes the optimization procedure.

3.2 Model Testing

Once the dictionary $\mathbf{D}$ and the classifier parameters $\langle \mathbf{U}, \mathbf{b} \rangle$ are learned, we perform classification as follows: for a test sample $\mathbf{x}$, we first calculate the sparse coding vector. As $\ell_{1,\infty}$-norm is a matrix norm and thus cannot tackle with a single vector, we instead use ℓ_1-norm regularization to get the coding vector, resulting in the following tractable problem [13,14]

$$\langle \mathbf{a} \rangle = \arg\min_{\mathbf{a}} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda_1\|\mathbf{a}\|_1. \tag{11}$$

We then apply learned SVM classifier to identify the test sample as follows

$$Label(\mathbf{x}) = \arg\max_{c} \mathbf{u}_c^T \mathbf{a} + b_c. \tag{12}$$

4 Experiments

To verify effectiveness of the proposed DDL model, extensive experiments on Caltech-101 object database [15], AR [16] and Extended Yale B [17] face database are carried out and performance of the proposed model are compared with the base-line sparse representation based classification (SRC) [18] method and state-of-the-art DDL methods including DKSVD [5], LC-KSVD [19], dictionary learning with structure incoherence (DLSI) [1], Fisher discrimination dictionary learning (FDDL) [3] and SVGDL [7]. All the experiments are carried out in Matlab (R2014a) environment running on a modern computer with Intel(R) Xeon(R) CPU 3.30 GHz and 32 GB memory. Note that we fix $\theta = 0.2$ and set $\mu = \lambda_1$ in all the experiments.

4.1 Object Classification

Caltech-101 object database contains 9,144 images from 102 object classes (101 common object classes and a background class). For each class, its number varies from 31 to 800. Following [7], we randomly select 5, 10, 15, 20, 25 and 30 images per object for training and the rest are used for testing. The number of dictionary atoms is set to be 510 in all the cases. We set $\lambda_1 = 5$ and $\lambda_2 = 0.1$. As shown in Table 1, SRC achieves the worst accuracy which is possibly attributed to lack of discriminative dictionary learning. With a class-specific dictionary, FDDL outperforms K-SVD and LCKSVD, however, when the training number is high (say 25, 30) per class, there is no significant gain over K-SVD and LCSVD. By learning a discriminative dictionary under the guidance of SVM, SVGDL has a better classification accuracy than LCKSVD and FDDL. However, our proposed DDL method outperforms SVGDL which indicates that the adaptive class relatedness learning can lead to more discriminative dictionary.

Table 1. The classification accuracy results on Caltech-101 database.

Training number	5	10	15	20	25	30
SRC	0.488	0.601	0.649	0.677	0.692	0.707
K-SVD	0.498	0.598	0.652	0.687	0.710	0.732
DKSVD	0.496	0.595	0.651	0.686	0.711	0.730
LCKSVD	0.540	0.631	0.677	0.705	0.723	0.736
FDDL	0.536	0.636	0.668	0.698	0.717	0.731
SVGDL	0.553	0.643	0.696	0.723	0.751	0.767
Proposed	**0.576**	**0.668**	**0.709**	**0.750**	**0.770**	**0.788**

Fig. 1. Resized face images in the AR database.

4.2 Face Recognition

We also apply our algorithm to face recognition (FR) on two widely used databases: AR and Extended Yale B. The features are reduced to 300 dimensions by PCA for all FR experiments. Note that here we also make comparisons with the SVM method.

(1) The AR database consists of over 4,000 face images from 126 individuals. As in [18], we use a set containing 1,400 face images from 50 female and 50 male subjects. For each subject, there are 7 images for training and 7 images for testing. The face images are resized to 60×43 as shown in Fig. 1. Here the number of dictionary atoms is set to be 500. In this experiment, we set $\lambda_1 = 0.02$ and $\lambda_2 = 0.00005$. The experimental results of different methods are listed in Table 2. To the best of our knowledge, the recognition accuracy of 0.951 achieved by the proposed DDL method is the best result ever reported on this database with the same training and testing samples.

Table 2. The recognition accuracy results on AR database.

Methods	SRC	SVM	DKSVD	LC-KSVD	DLSI	FDDL	SVGDL	**Proposed**
Accuracy	0.888	0.871	0.854	0.897	0.898	0.920	0.946	**0.951**

(2) The Extended Yale B database consists of 2,414 face images from 38 persons. All the images are cropped into the size of 54×48. We randomly select 20 images of each person for training and the rest are used for testing. In this experiment, the number of the dictionary atoms is 380, $\lambda_1 = 0.1$ and $\lambda_2 = 0.0005$. As one can see from Table 3, our proposed DDL method achieves the best recognition accuracy.

Table 3. The recognition accuracy results on Extended Yale B database.

Methods	SRC	SVM	DKSVD	LC-KSVD	DLSI	FDDL	SVGDL	**Proposed**
Accuracy	0.900	0.888	0.753	0.906	0.890	0.919	0.961	**0.971**

5 Conclusions

In this paper, a novel DDL method which adaptively builds relationship between dictionary and class labels is presented. Instead of learning a global dictionary which lacks of correspondence between the dictionary atoms and class labels or a class-specific dictionary which misses the dictionary relatedness between different classes, we learn a dictionary which not only preserves correspondence between the dictionary atoms and class labels but also remains class relatedness between different classes, leading to a more powerful and discriminative dictionary. Experimental results on object classification and face recognition demonstrate superiority of our proposed method over many state-of-the-art DDL methods. Thus, we argue that our proposed method can provide a new insight to current dictionary learning based pattern recognition methods.

References

1. Ramirez, I., Sprechmann, P., Sapiro, G.: Classification and clustering via dictionary learning with structured incoherence and shared features. In: 23rd IEEE Conference on Computer Vision and Pattern Recognition, pp. 3501–3508 (2010)
2. Tošić, I., Frossard, P.: Dictionary Learning, What is the right representation for my signal? IEEE Signal Proc. Mag. **28**, 27–38 (2011)
3. Yang, M., Zhang, D., Feng, X.: Fisher discrimination dictionary learning for sparse representation. In: IEEE International Conference on Computer Vision, pp. 543–550 (2011)
4. Mairal, J., Bach, F., Ponce, J.: Task-driven dictionary learning. IEEE Trans. Pattern Anal. Mach. Intell. **34**, 791–804 (2012)
5. Zhang, Q., Li, B.: Discriminative K-SVD for dictionary learning in face recognition. In: 23rd IEEE Conference on Computer Vision and Pattern Recognition, pp. 3501–3508 (2010)
6. Pham, D., Svetha, V.: Joint learning and dictionary construction for pattern recognition. In: 26th IEEE Conference on Computer Vision and Pattern Recognition (2008)
7. Cai, S., Zuo, W., Zhang, L., Feng, X., Wang, P.: Support vector guided dictionary learning. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014, Part IV. LNCS, vol. 8692, pp. 624–639. Springer, Heidelberg (2014)
8. Castrodad, A., Sapiro, G.: Sparse modeling of human actions from motion imagery. Int. J. Comput. Vis. **100**, 1–15 (2012)
9. Gao, S., Tsang, I.W.H., Chia, L.T.: Laplacian sparse coding, hypergraph laplacian sparse coding, and applications. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**(1), 92–104 (2013)
10. Quattoni, A., Carreras, X., Collins, M., Darrell, T.: An efficient projection for $\ell_{1,\infty}$ regularization. In: The 26th Annual International Conference on Machine Learning, vol. 382 (2009)
11. Lee, H., Battle, A.: Efficient sparse coding algorithms. In: 20th Annual Conference on Neural Information Processing Systems, pp. 801–808 (2007)
12. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1794–1801 (2009)

13. Yang, A.Y., Sastry, S.S., Ganesh, A., Ma, Y.: Fast ℓ1-minimization algorithms and an application in robust face recognition: a review. In: IEEE International Conference on Image Processing, pp. 1849–1852 (2010)
14. Zuo, W., Meng, D., Zhang, L., Feng, X., Zhang, D.: A generalized iterated shrinkage algorithm for non-convex sparse coding. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 217–224 (2013)
15. Fei-Fei, L., Fergus, R., Perona, P.: Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 178–186 (2004)
16. Martinez, A., Benavente, R.: The AR face database. CVC Technical report (1998)
17. Lee, K.C., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. IEEE Transactions on Pattern Analysis and Machine Intelligence **27**, 684–698 (2005)
18. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence **31**, 210–227 (2009)
19. Jiang, Z., Lin, Z., Davis, L.: Label consistent K-SVD: learning a discriminative dictionary for recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence **35**, 2651–2664 (2013)