SUPERVISED SPARSE CODING WITH LOCAL GEOMETRICAL CONSTRAINTS

Hanchao Zhang[1] *and Jinhua Xu*[1,2]

[1] Department of Computer Science and Technology
[2] Shanghai Key Laboratory of Multidimensional Information Processing
East China Normal University, Shanghai 200241, China

ABSTRACT

Sparse coding algorithms with geometrical constraints have received much attention recently. However, these methods are unsupervised and might lead to less discriminative representations. In this paper, we propose a supervised locality-constrained sparse coding method for classification. Two graphs are constructed, a labeled graph and an unlabeled graph. Sparse codes with a labeled geometrical constraint will be more discriminative, however we cannot embed test samples with unknown label into a labeled graph. By coupling the two graphs, we aim to make the difference between sparse codes with labeled and unlabeled geometrical constraints as small as possible. As a result, sparse codes of test data can be obtained with the unlabeled geometrical constraint and the discrimination of the labeled geometrical constraint is maintained. Experiments on some benchmark datasets demonstrate the effectiveness of the proposed method.

Index Terms— sparse coding, supervised, manifold learning, geometrical constraint, manifold embedding

1. INTRODUCTION

Sparse coding has been applied successfully in numerous classification tasks [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. In some early research [7, 10], unsupervised sparse coding and dictionary learning were used, without considering the label information. However, the sparse codes and dictionaries learned in this way often lack discrimination as they are optimal for reconstruction but not for classification.

Many algorithms have been proposed to enhance the discrimination of visual dictionaries through supervised learning. In some previous work [5, 11, 12, 13, 14], multiple category-specific dictionaries were learned to promote discrimination between classes. In other work [3, 6, 15], the dictionary learning and classifier training were combined into a single objective function, aiming at enhancing the discrimination of the learned dictionary by solving the unified optimization. Supervised dictionary learning by using backpropagation of the classification error was proposed in [1, 4, 16, 17, 18]. It was indicated that dictionaries learned via backpropagation yield better classification performances [1, 4].

Recently, some research work suggested that image space is actually a smooth low dimensional sub-manifold embedded in a high dimensional ambient space. Standard sparse coding fails to consider geometrical structure, and thus may be inaccurate in modeling the manifold [19]. Meanwhile, the over-completeness of the dictionary and the independent coding process may also result in the instability of sparse coding [20], that is, similar features may be encoded as totally different sparse codes. As suggested in [19], locality was more essential than sparsity. Locality can lead to sparsity but not vice versa. Therefore, some research has been done to address locality preserving or similarity preserving during dictionary learning for image classification [8, 19, 20, 21, 22]. Gao et al. [20] and Zheng et al. [22] proposed locality-constrained sparse coding to preserve the local manifold structure of the instances by embedding the Laplacian matrix into sparse coding algorithm. However these algorithms did not use labels either.

In this paper, we propose a supervised sparse coding with local geometrical-constraint. Our goal is to learn a discriminative dictionary and sparse codes for classification problems. Two graphs are constructed, one is a labeled graph, in which each data point is connected with its nearest neighbors in the same class; The other is an unlabeled graph, in which each data point is connected with its nearest neighbors, which may have different labels. In the labeled graph, a block diagonal similarity matrix is calculated to capture the local geometrical relationship and enhance the discrimination of the dictionary and the sparse codes. Using the diagonal constraint matrix alone, the classification results in the training set can be perfect using the learned representations. But the classification results in the test set is severely constrained by the gap between the labelled training coding process and the unlabelled testing coding process. By coupling the two graphs, a common dictionary is learned, and the difference between sparse codes with labeled and unlabeled geometrical constraints is

This work is supported by the National Natural Science Foundation of China under Project 61175116, Shanghai Knowledge Service Platform for Trustworthy Internet of Things No. ZF1213, and Science and Technology Commission of Shanghai Municipality under research grant No. 14DZ2260800.

narrowed as much as possible. As a result, embedding a test data point into the unlabeled graph can have a discriminative representation.

2. RELATED WORK

In this section, we review the standard sparse coding and the related work on locality preserving sparse coding.

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N]$, and $\mathbf{x}_i \in R^d$ represent observations, where d is the dimension of data points and N is the number of samples. Let $\mathbf{\Phi} = [\phi_1, \phi_2, \cdots, \phi_m]$ be an over-complete dictionary ($m > d$), where m is the size of the dictionary and the columns $\phi_i \in R^d$ are visual words or bases. Let $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_N]$ be the coefficients(sparse codes) of $\mathbf{X}$ and $\mathbf{s}_i = [s_{1i}, \cdots, s_{mi}]^T$.

The sparse coding algorithm can be formulated as:

$$< \hat{\mathbf{S}}, \hat{\mathbf{\Phi}} >= arg\min_{\mathbf{s}_i, \mathbf{\Phi}} \sum_i \{\|\mathbf{x}_i - \mathbf{\Phi}\mathbf{s}_i\|^2 + \gamma\|\mathbf{s}_i\|_1\}, \quad (1)$$
$$s.t.\ \ \|\phi_j\|^2 = 1,\ \ \forall j = 1, \cdots m.$$

Where γ is a regularization parameter to control the tradeoff between reconstruction and sparseness.

After the dictionary is learned, the sparse code of a new signal $\mathbf{x}$ is calculated

$$\hat{\mathbf{s}} = arg\min_{s} \|\mathbf{x} - \mathbf{\Phi}\mathbf{s}\|^2 + \gamma\|\mathbf{s}\|_1. \quad (2)$$

To enhance the smoothness along the manifold, manifold structure is embedded in the sparse coding process as regularization terms [8, 20, 21, 22]. Locality-constrained sparse coding was proposed in [20, 22] by embedding the Laplacian matrix into sparse coding algorithm. The objective function for locality-constrained sparse coding is as follows:

$$\min_{\mathbf{s}_i} \sum_i \{\|\mathbf{x}_i - \mathbf{\Phi}\mathbf{s}_i\|^2 + \gamma\|\mathbf{s}_i\|_1\} + \frac{\alpha}{2}\sum_{ij}\|\mathbf{s}_i - \mathbf{s}_j\|^2 W_{ij}, \quad (3)$$

where $\mathbf{W}$ is the similarity matrix between instances. In [22], $\mathbf{W}$ is defined as:

$$W_{ij} = \begin{cases} 1 & if\ \ \mathbf{x}_j \in \mathcal{N}_{knn}\{\mathbf{x}_i\}\ \ or\ \ \mathbf{x}_i \in \mathcal{N}_{knn}\{\mathbf{x}_j\}; \\ 0 & otherwise. \end{cases} \quad (4)$$

In [20], $\mathbf{W}$ is defined as:

$$W_{ij} = \begin{cases} sim(\mathbf{x}_i, \mathbf{x}_j) & if\ \ \mathbf{x}_j \in \mathcal{N}_{knn}\{\mathbf{x}_i\}; \\ 0 & if\ \ \mathbf{x}_j \notin \mathcal{N}_{knn}\{\mathbf{x}_i\}. \end{cases} \quad (5)$$

Where $\mathcal{N}_{knn}\{\mathbf{x}_i\}$ represents k-nearest neighbors of $\mathbf{x}_i$, and $sim(\mathbf{x}_i, \mathbf{x}_j)$ denotes the similarity of $\mathbf{x}_i$ and $\mathbf{x}_j$.

In [21], the geometrical structures are encoded in two situations. When data points distribute on a single manifold, the topological structures are explicitly modeled by locally linear embedding algorithm combined with k-nearest neighbors:

$$\begin{aligned} \mathbf{x}_i &= \mathbf{X}\mathbf{u}_i, \quad (6) \\ s.t. \quad & u_{ji} = 0,\ \ if\ \ \mathbf{x}_j \notin \mathcal{N}_{knn}\{\mathbf{x}_i\}, \\ & \sum_j u_{ji} = 1. \end{aligned}$$

here $\mathbf{u}_i = [u_{1i}, u_{2i}, \cdots, u_{Ni}]^T$. When data points lie on multiple manifolds, sparse representation algorithm combined with k-nearest neighbors is utilized to construct the topological structures:

$$\begin{aligned} &\min_{\mathbf{u}_i} \sum_i \{\|\mathbf{x}_i - \mathbf{D}\mathbf{u}_i^{knn}\|^2 + \gamma\|\mathbf{u}_i^{knn}\|_1\}, \\ &\mathbf{D} = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \cdots, \mathbf{x}_{in}], \mathbf{x}_{ij} \in \mathcal{N}_{knn}. \end{aligned} \quad (7)$$

After obtaining the local fitting relationship represented by $U = [\mathbf{u}_1, \cdots, \mathbf{u}_N]$,the topological structures are then embedded into sparse coding algorithm as regularization terms to formulate the corresponding objective functions of dictionary learning.

$$\min_{\mathbf{S}, \mathbf{\Phi}} \|\mathbf{X} - \mathbf{\Phi}\mathbf{S}\|_F^2 + \gamma\|\mathbf{S}\|_1 + \alpha tr\{\mathbf{S}\mathbf{G}\mathbf{S}^T\}, \quad (8)$$

where $\mathbf{G} = (\mathbf{I}_N - \mathbf{U})(\mathbf{I}_N - \mathbf{U})^T$, and $\mathbf{I}_N$ is the $N \times N$ identity matrix.

In all these approaches proposed in [8, 20, 21, 22], the label information was not utilized.

3. SPARSE CODING FOR TRAINING SET

Let $\mathbf{X} = [\mathbf{X}^1, \mathbf{X}^2, \cdots, \mathbf{X}^C]$ and C be the number of categories, where $\mathbf{X}^c = [\mathbf{x}_1^c, \cdots, \mathbf{x}_{N_c}^c]$ includes all N_c training samples from the cth category. We construct a labeled graph and an unlabeled graph. In the labeled graph, each training sample $\mathbf{x}_i^c$ is connected with its neighbors $\mathbf{x}_j^c \in \mathcal{N}_{knn}\{\mathbf{x}_i^c\}$. We model the local geometrical structure using the locally linear embedding (LLE).

$$\begin{aligned} \mathbf{x}_i^c &= \mathbf{X}^c\mathbf{u}_i^c, \quad (9) \\ s.t. \quad & u_{ji}^c = 0,\ \ if\ \ \mathbf{x}_j^c \notin \mathcal{N}_{knn}\{\mathbf{x}_i^c\}, \\ & \sum_j u_{ji}^c = 1. \end{aligned}$$

where $\mathbf{u}_i^c = [u_{1i}^c, \cdots, u_{N_c i}^c]^T$. Note that x_i is excluded from $\mathcal{N}_{knn}\{\mathbf{x}_i^c\}$. Let $U^c = [\mathbf{u}_1^c, \cdots, \mathbf{u}_{N_c}^c]$, that is,

$$\mathbf{X}^c = \mathbf{X}^c U^c, \quad (10)$$

then we have

$$\mathbf{X} = \mathbf{X}\mathbf{U}. \quad (11)$$

The similarity matrix of the labeled graph can be defined as $\mathbf{W}^* = \max\{|\mathbf{U}|, |\mathbf{U}^T|\}$, where $|\mathbf{U}|$ denotes the matrix composed of absolute value of elements in $\mathbf{U}$. We can see the geometric structure of the labeled graph is represented by a block diagonal matrix

$$\mathbf{W}^* = \begin{bmatrix} \mathbf{W}^1 & & \\ & \mathbf{W}^2 & \\ & & \mathbf{W}^C \end{bmatrix}. \quad (12)$$

In the unlabeled graph, each training data $\mathbf{x}_i$ is connected with its neighbors $\mathbf{x}_j \in \mathbf{X}$. We model the local geometrical structure using the locally linear embedding as in (6). Let $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_N]$, the similarity matrix can be defined as $\mathbf{W} = \max\{|\mathbf{U}|, |\mathbf{U}^T|\}$. Here W is not block diagonal.

For a training sample $\mathbf{x}_i$, let $\mathbf{s}_i^*$ and $\mathbf{s}_i$ be the codes learned with labeled and unlabeled graph constraints respectively.

$$\min_{\mathbf{s}_i^*, \mathbf{\Phi}} \sum_i \{\|\mathbf{x}_i - \mathbf{\Phi}\mathbf{s}_i^*\|^2 + \gamma\|\mathbf{s}_i^*\|_1\} + \frac{\alpha}{2}\sum_{ij}\|\mathbf{s}_i^* - \mathbf{s}_j^*\|^2 W_{ji}^*, \quad (13)$$

$$\min_{\mathbf{s}_i, \mathbf{\Phi}} \sum_i \{\|\mathbf{x}_i - \mathbf{\Phi}\mathbf{s}_i\|^2 + \gamma\|\mathbf{s}_i\|_1\} + \frac{\beta}{2}\sum_{ij}\|\mathbf{s}_i - \mathbf{s}_j\|^2 W_{ji}. \quad (14)$$

To make the difference of the sparse codes with labeled and unlabeled graph constraints as small as possible, sparse coding for the training set can be formulated as

$$\begin{aligned} \min_{\mathbf{s}_i^*, \mathbf{\Phi}} & \sum_i \{\|\mathbf{x}_i - \mathbf{\Phi}\mathbf{s}_i^*\|^2 + \gamma\|\mathbf{s}_i^*\|_1\} \\ & + \frac{\alpha_1}{2}\sum_{i\neq j}\|\mathbf{s}_i^* - \mathbf{s}_j^*\|^2 W_{ji}^* + \frac{\alpha_2}{2}\sum_{i\neq j}\|\mathbf{s}_i^* - \mathbf{s}_j\|^2 W_{ji}, \end{aligned} \quad (15)$$

$$\begin{aligned} \min_{\mathbf{s}_i, \mathbf{\Phi}} & \sum_i \{\|\mathbf{x}_i - \mathbf{\Phi}\mathbf{s}_i\|^2 + \gamma\|\mathbf{s}_i\|_1\} \\ & + \frac{\beta_1}{2}\sum_{i\neq j}\|\mathbf{s}_i - \mathbf{s}_j\|^2 W_{ji} + \frac{\beta_2}{2}\sum_{i\neq j}\|\mathbf{s}_i - \mathbf{s}_j^*\|^2 W_{ji}. \end{aligned} \quad (16)$$

Since $\mathbf{W}^*$ is a diagonal matrix, $\mathbf{s}_i^*$ calculated by (13) is more discriminative than $\mathbf{s}_i$ obtained by (14). Our aim is to incorporate the label information and narrow the gap between labelled and unlabelled coding process at the same time, so we simply set α_1 equal to β_2 , β_1 equal to α_2 and let the former larger than the latter.

In practice we optimize (15) and (16) alternately. Let $\widetilde{\mathbf{X}} = [\mathbf{X}, \mathbf{X}]$, $\widetilde{\mathbf{S}} = [\mathbf{S}^*, \mathbf{S}]$, and

$$\widetilde{\mathbf{W}} = \begin{pmatrix} \alpha\mathbf{W}^* & \alpha\mathbf{W} \\ \beta\mathbf{W} & \beta\mathbf{W} \end{pmatrix}. \quad (17)$$

Calculating $\mathbf{S}^*$, $\mathbf{S}$ in sequential order according to their positions in $\widetilde{\mathbf{S}}$, the combined objective function of (15) and (16) can be written as follows

$$\min_{\widetilde{\mathbf{S}}} \|\widetilde{\mathbf{X}} - \mathbf{\Phi}\widetilde{\mathbf{S}}\|^2 + \gamma\|\widetilde{\mathbf{S}}\|_1 + tr(\widetilde{\mathbf{S}}\widetilde{\mathbf{G}}\widetilde{\mathbf{S}}^T), \quad (18)$$

where $\widetilde{\mathbf{G}}$ is a Laplacian matrix $\widetilde{\mathbf{G}} = \widetilde{\mathbf{D}} - \widetilde{\mathbf{W}}$, and $\widetilde{\mathbf{D}}$ is a diagonal function $D_{ii} = \sum_j \widetilde{W}_{ji}$.

$\mathbf{\Phi}$ and $\widetilde{\mathbf{S}}$ can be calculated by solving (18) using feature sign algorithm as [22]. Then a classifier can be trained using labels and $\mathbf{S}$.

4. SPARSE CODING FOR TESTING SET

For a test sample $\mathbf{x}_t$, we do not know its label. Therefore we embed it into only the unsupervised graph. Represent $\mathbf{x}_t$ using its nearest neighbors in the training set:

$$\begin{aligned} \mathbf{x}_t &= \mathbf{X}\mathbf{u}_t, \quad (19) \\ s.t. \quad & u_{jt} = 0, \ \ if \ \ \mathbf{x}_j \notin \mathcal{N}_{knn}\{\mathbf{x}_t\}, \\ & \sum_j u_{jt} = 1. \end{aligned}$$

Let $\widetilde{\mathbf{X}}_t = [\mathbf{X}, \mathbf{X}_t\}$, $\widetilde{\mathbf{S}}_t = [\mathbf{S}^*, \mathbf{S}_t]$, here $\mathbf{X}_t = [\mathbf{X}, \mathbf{x}_t]$, $\mathbf{S}_t = [\mathbf{S}, \mathbf{s}_t]$, and the similarity matrix $\widetilde{\mathbf{W}}_t$ can be defined as

$$\widetilde{\mathbf{W}}_t = \begin{pmatrix} \alpha\mathbf{W}^* & \alpha\mathbf{W} & \alpha|\mathbf{u}_t| \\ \beta\mathbf{W} & \beta\mathbf{W} & \beta|\mathbf{u}_t| \\ \beta|\mathbf{u}_t^T| & \beta|\mathbf{u}_t^T| & 0 \end{pmatrix}. \quad (20)$$

Then $\widetilde{\mathbf{S}}_t$ is obtained by solving

$$\min_{\widetilde{\mathbf{S}}_t} \|\widetilde{\mathbf{X}}_t - \mathbf{\Phi}\widetilde{\mathbf{S}}_t\|^2 + \gamma\|\widetilde{\mathbf{S}}_t\|_1 + tr(\widetilde{\mathbf{S}}_t\widetilde{\mathbf{G}}_t\widetilde{\mathbf{S}}_t^T). \quad (21)$$

In fact, we only need to solve $\mathbf{s}_t$ with the learned $\widetilde{\mathbf{S}}$ fixed. The class label of $\mathbf{x}_t$ can be predicted using the trained classifier.

5. EXPERIMENTS

In this section, we tested the proposed model with different recognition tasks, including textures, handwritten digits and faces. The sparse codes were fed to Linear SVM for classification. We used libsvm implementation [23] for SVM.

5.1. Texture classification

Two texture images from the Brodatz dataset were used to build two classes as in [6] and the patch size was 12×12.We randomly selected patches from left and right half of each texture for training and testing set respectively, so that there was no overlap between the training and test sets. We used $\gamma = 0.1$, $\alpha = 1$, and $\beta = 0.1$. The dictionary size was 64. We compared the classification performance of our method with the unsupervised and supervised dictionary learning methods in [6] and Graph sparse coding in [22]. The results of unsupervised and supervised dictionary learning methods were from [6] because we did not implement their methods. It should be noted that we downsampled the two texture images before extracting patches for Graph sparse coding and experiments. We ran 20 times for each training set size. The results for training sets of various size N were shown in Table 1. For all training sets, our results were significantly better than unsupervised dictionary learning (REC), generative supervised dictionary learning (SDL G) and discriminative dictionary learning (SDL D). Compared with GraphSC in [22], our results were also better for all training sets.

Table 1. Error rates for the texture classification task using various methods and sizes N of the training set.

Training	REC(L1) [6]	SDL G[6]	SDL D[6]	GraphSC [22]	Ours
300	48.84	47.34	44.84	20.96	18.28
1500	46.8	46.3	42.0	16.70	14.61
3000	45.17	45.1	40.6	16.23	13.80
6000	45.71	43.68	39.77	15.54	13.43
15000	47.54	46.15	38.99	15.29	13.70

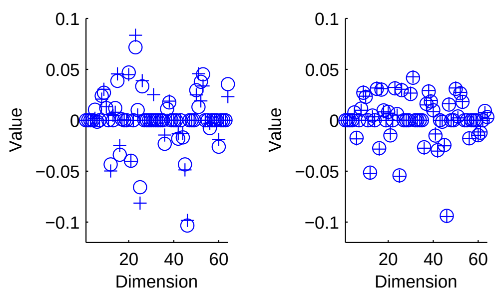

Fig. 1. Codes under different methods

Now we explain why the proposed approach performed better. We compared the sparse codes of a training sample with different geometrical constraints. In the left of Fig.1, we showed the sparse codes learned with a labeled geometrical constraint using (13) (denoted by circle) and with a unlabeled geometrical constraint using (14) (denoted by plus sign) separately. The difference between codes can be seen. In the right of Fig.1, we showed the sparse codes learned with coupled graphs using (15) (denoted by circle) and using (16) (denoted by plus sign). It can be seen that s^* and s were almost overlapped, and our proposed method narrowed the gap between supervised and unsupervised sparse coding process. Therefore, by embedding the test sample into the unlabeled graph, the discrimination of the labeled graph can be achieved.

5.2. Face Recognition

The Extended Yale B database consists of 2,414 near frontal face images from 38 individuals. We randomly selected 20 images for training and used the rest for testing. Images were resized to 32×32 and these original images were used as input. Our settings for GSC and proposal method were different from [13], where images were resized to 54×48 and Eigenface was used. But we also listed their result here as a comparison. The parameters we used here were $\gamma = 0.001$, $\alpha = 0.1$, and $\beta = 0.001$. As showed in Table 2, our method achieved the best performance.

Table 2. The recognition rates for Extended Yale B for different approaches

Approaches	FDDL[13]	GSC[22]	Ours
Recognition rates	0.919	0.917	0.934

Table 3. Error rate for USPS for different approaches

Approaches	UDL [4]	SDL [4]	GSC [22]	Ours
Error rate	4.58	2.84	5.0	1.95

5.3. USPS dataset

The USPS dataset has 7291 training images and 2007 test images of size 16×16. We used regularization parameters $\gamma = 0.1$, $\alpha = 0.1$, and $\beta = 0.01$. The dictionary size is 512.

We compared our results with graph sparse coding (GSC) in [22], Mairal et al.'s unsupervised(UDL) and supervised dictionary learning(SDL) in [4]. The best results of these approaches were shown in Table 3. The error rate of our approach was 1.95%, which was lower than all other approaches compared.

6. CONCLUSIONS

In this article we propose a supervised sparse coding framework. The label information and the geometrical structure are simultaneously incorporated to the sparse coding method through a block diagonal similarity matrix. Through the common learned dictionary, which is stable for labeled and unlabeled geometrical constraints, the discrimination of the dictionary and the sparse codes is improved. Experimental results on different classification problems illustrated the effectiveness of our proposal method.

7. REFERENCES

[1] D. M. Bradley and J. A. Bagnell, "Differential sparse coding," in *Proc. Advances in neural information processing systems(NIPS)*, 2008, pp. 113–120.

[2] K. Huang and S. Aviyente, "Sparse representation for signal classification," in *Proc. Advances in neural information processing systems(NIPS)*, 2006, pp. 609–616.

[3] Z. Jiang, Z. Lin, and L. Davis, "Label consistent k-svd: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, 2013.

[4] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, 2012.

[5] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Discriminative learned dictionaries for local image analysis," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2008, pp. 1–8.

[6] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," in *Proc. Advances in neural information processing systems(NIPS)*, 2009, pp. 1033–1040.

[7] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: transfer learning from unlabeled data," in *Proc. International conference on Machine Learning*. ACM, 2007, pp. 759–766.

[8] J. Wang, J. Yang, K. Yu, F. Lv, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*. IEEE, 2010, pp. 3360–3367.

[9] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210C–227, 2009.

[10] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*. IEEE, 2009, pp. 1794–1801.

[11] J. Mairal, M. Leordeanu, F. Bach, M. Hebert, and J. Ponce, "Discriminative sparse image models for class-specific edge detection and image interpretation," in *Proc. European Conference on Computer Vision(ECCV)*. 2008, pp. 43–56, Springer.

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

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

[14] N. Zhou, Y. Shen, J. Peng, and J. Fan, "Learning inter-related visual dictionay for object recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*. IEEE, 2012, pp. 3490–3497.

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

[16] Y. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 2559–2566.

[17] J. Yang, K. Yu, and T. Huang, "Supervised translation-invariant sparse coding," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*. IEEE, 2010, pp. 3517–3524.

[18] J. Yang and M. Yang, "Top-down visual saliency via a joint crf and dictionary learning," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*. IEEE, 2012, pp. 2296 – 2303.

[19] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Proc. Advances in neural information processing systems(NIPS)*, 2009, pp. 2223–2231.

[20] S. Gao, I. W. Tsang, and L.T. Chia, "Laplacian sparse coding, hypergraph laplacian sparse coding, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 92–104, 2013.

[21] B.D. Liu, Y.X. Wang, Y.J. Zhang, and B. Shen, "Learning dictionary on manifolds for image classification," *Pattern Recognition*, vol. 46, pp. 1879–1890, 2013.

[22] M. Zheng, J. Bu, C. Chen, C. Wang, L. Zhang, G. Qiu, and D. Cai, "Graph regularized sparse coding for image representation," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1327–1336, 2011.

[23] C.C. Chang and C.J. Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 27, pp. 1–27, 2011.