LOCALITY SENSITIVE DISCRIMINATIVE DICTIONARY LEARNING

Jun Guo, Yanqing Guo, Yi Li, Bo Wang, and Ming Li

School of Information and Communication Engineering, Dalian University of Technology, Dalian, China

E-mail: {guojun, liyi}@mail.dlut.edu.cn; {guoyq, bowang, mli}@dlut.edu.cn

ABSTRACT

Discriminative dictionary learning (DDL) has been applied to various pattern classification problems. Despite satisfying experimental results, most existing discriminative dictionary learning methods emphasize too much on the role of l_0 or l_1-norm sparsity, while the underlying local structure of original data is totally ignored. In this paper, we present a novel dictionary learning method, named Locality Sensitive Discriminative Dictionary Learning (LSDDL), which combines basic dictionary learning scheme and locality relationship of original data which is propagated to the coding vectors. The learned discriminative dictionary can map the original data points into a new space in which the nearby points with the same label are close to each other while the nearby points with different labels are far apart. Experiments clearly show that our method has very competitive performance in contrast to previous discriminative dictionary learning methods.

Index Terms— Discriminative dictionary learning, pattern classification, supervised learning, locality.

1. INTRODUCTION

Dictionary learning (DL) aims to learn a series of bases, namely atoms, so that a given data point can be well approximated by linearly combining these bases [1]. However, the traditional dictionary learning framework is not widely applied to pattern classification at prime tense, since the dictionary is merely used to reconstruct signals. To overcome this obstacle, researchers have proposed a lot of approaches to obtain a classification-oriented dictionary by utilizing the category label information.

In [2] and [3], a given testing sample is first coded over a fixed dictionary which is made up of all the training samples, then the classification is conducted by checking which class yields the least reconstruction error. These novel classification algorithms perform well in human face recognition and boost the research of dictionary learning in the field of pattern classification. By learning a dictionary instead of exploiting a pre-defined dictionary, a class-wise residual can still be used as a decision value for label assignment [4], [5]. Recently, more and more attention has been paid on making the sparse coefficients discriminative by enforcing the discriminability of dictionary [6]-[10]. Researchers utilize the coding coefficients as new features and apply various classification schemes according to multifarious learning tasks. In [11]-[14], the local similarity of coding coefficients is taken into account during the dictionary learning process.

However, those previous works emphasize too much on the role of sparsity and consider only discriminativeness for classification, while the underlying local structure of original input data is totally ignored. In many applications, high-dimensional features from the same class often exhibit degenerate structure [15] and lie on or near low-dimensional subspaces, submanifolds, or stratifications. The dependence information among the local data points is mostly lost during the reconstruction process with a learned dictionary.

To exploit this underlying local structure of original input data and encourage discriminability simultaneously, we present a novel Locality Sensitive Discriminative Dictionary Learning (LSDDL) method, which combines basic dictionary learning framework and locality relationship of original data which is propagated to the coding vectors. The learned discriminative dictionary can map the data points into a new space where the nearby points from the same class are close to each other while the nearby points from different classes are far apart. The other contribution of our paper is that the LSDDL method utilizes analytical solutions in both dictionary learning and coding phases, which can improve the efficiency. Experiments in famous visual classification datasets show that our proposed LSDDL outperforms state-of-the-art DL methods.

The rest of this paper is organized as follows. Section 2 provides a brief review of the basics of dictionary learning. In Section 3, the proposed Locality Sensitive Discriminative Dictionary Learning (LSDDL) method is introduced. Extensive experiments are conducted to verify the effectiveness of our proposed method in Section 4. Finally, we conclude this paper as well as discuss future works in Section 5.

This work is supported by the National Natural Science Foundation of China (Grant No. 61402079), and China Postdoctoral Science Foundation (Grant No. 20110490343 and 2013T60090), and the Fundamental Research Funds for the Central Universities (Grant No. DUT14RC(3)103).

2. DICTIONARY LEARNING

Let $\boldsymbol{Y}=[\boldsymbol{y}_1,...,\boldsymbol{y}_N]\in\mathbb{R}^{n\times N}$ be the training data matrix, where $\boldsymbol{y}_i\in\mathbb{R}^n$ denotes the ith sample with n-dimensional feature description. Dictionary learning scheme aims to obtain an optimized dictionary $\boldsymbol{D}=[\boldsymbol{d}_1,...,\boldsymbol{d}_K]\in\mathbb{R}^{n\times K}$ that provides a succinct and effective representation for most training data. The procedure of dictionary learning can be formulated as

$$\left(\boldsymbol{D}^*,\boldsymbol{X}^*\right)=\arg\min_{\boldsymbol{D},\boldsymbol{X}}\sum_{i=1}^{N}\left(\|\,\boldsymbol{y}_i-\boldsymbol{D}\boldsymbol{x}_i\,\|_2^2+\tau\,\|\,\boldsymbol{x}_i\,\|_p\right),\quad (1)$$

where $\boldsymbol{X}=[\boldsymbol{x}_1,...,\boldsymbol{x}_N]\in\mathbb{R}^{K\times N}$ is the coding coefficient matrix of $\boldsymbol{Y}$ over $\boldsymbol{D}$. Constant $\tau\geq 0$ is a regularization parameter and $\|\cdot\|_p$ denotes the l_p-norm. If $p=1$, we obtain a LASSO model [21], which can produce a sparse representation. If $p=2$, we have a model called Ridge Regression [21], which can shrink all the elements. The basic dictionary learning frame in (1) can be rewritten in an equivalent matrix form:

$$\left(\boldsymbol{D}^*,\boldsymbol{X}^*\right)=\arg\min_{\boldsymbol{D},\boldsymbol{X}}\left\|\boldsymbol{Y}-\boldsymbol{D}\boldsymbol{X}\right\|_F^2+\tau\left\|\boldsymbol{X}\right\|_p.\qquad (2)$$

The optimization problem in (2) is convex in either $\boldsymbol{D}$ or $\boldsymbol{X}$, but not convex in both simultaneously. A two-step strategy is commonly used by alternatively optimizing with respect to one variable while holding the other fixed.

Note that (2) is concentrated on the reconstruction error and ignores the discriminatablity of dictionary for classification tasks. Hence, discriminative dictionary learning (DDL) methods have been proposed, which integrate the category label information of training data into the objective function of dictionary learning by introducing discrimination promotion functions to (2) to ensure the discrimination power of $\boldsymbol{D}$. Inspired by these attempts, we utilize the local structure of training data, which is totally ignored in most DDL methods.

3. PROPOSED METHOD

In this section, we introduce our proposed Locality Sensitive Discriminative Dictionary Learning method, which respects both discriminative and geometrical structure of the original data $\boldsymbol{Y}$. We begin with our motivation for Locality Sensitive Discriminative Dictionary Learning.

3.1. Motivation

As described previously, the original features from the same class often tend to lie on or near low-dimensional subspaces, submanifolds, or stratifications. We wish to learn a dictionary in order to map original features to a new space where several nearest same-label neighbors of each training data point are preserved while the points with different labels are repelled. In other words, we hope to propagate the locality relationship of training data $\boldsymbol{Y}$ to corresponding coding vectors $\boldsymbol{X}$ by learning a discriminative dictionary $\boldsymbol{D}$ which can maximize the local margin between different classes.

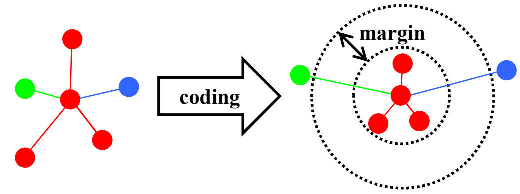

Fig. 1: Illustration of our proposed LSDDL method.

3.2. Locality Sensitive and Discriminative Objective

To discover both geometrical and discriminantive structure of the training data $\boldsymbol{Y}=[\boldsymbol{y}_1,...,\boldsymbol{y}_N]\in\mathbb{R}^{n\times N}$, we construct two adjacency graphs. For each data point $\boldsymbol{y}_i$, we use its k nearest same-label neighbors to create a set $N_w(\boldsymbol{y}_i)$ and put an edge between $\boldsymbol{y}_i$ and these neighbors to construct the within-class graph G_w. Similarly, we obtain a set $N_b(\boldsymbol{y}_i)$ containing the k nearest neighbors with different labels from $\boldsymbol{y}_i$ and construct the between-class graph G_b. Here, k is a pre-defined positive integer. Clearly, $N_b(\boldsymbol{y}_i)\cap N_w(\boldsymbol{y}_i)=\emptyset$. Let $\boldsymbol{W}_b$ and $\boldsymbol{W}_w$ denote the weight matrices of G_b and G_w, respectively. We define:

$$\boldsymbol{W}_{b,ij}=\begin{cases}1-w(\boldsymbol{y}_i,\boldsymbol{y}_j),\ \mathit{if}\ \boldsymbol{y}_i\in N_b(\boldsymbol{y}_j)\,\mathit{or}\,\boldsymbol{y}_j\in N_b(\boldsymbol{y}_i)\\ 0\qquad\qquad\quad,\ \mathit{otherwise}.\end{cases}\qquad (3)$$

$$\boldsymbol{W}_{w,ij}=\begin{cases}w(\boldsymbol{y}_i,\boldsymbol{y}_j),\ \mathit{if}\ \boldsymbol{y}_i\in N_w(\boldsymbol{y}_j)\,\mathit{or}\,\boldsymbol{y}_j\in N_w(\boldsymbol{y}_i)\\ 0\qquad\qquad,\ \mathit{otherwise}.\end{cases}\qquad (4)$$

Specifically, we use Gaussian kernel to define the weights

$$w(\boldsymbol{y}_i,\boldsymbol{y}_j)=\exp\left\{-\left[\frac{dist(\boldsymbol{y}_i,\boldsymbol{y}_j)}{t}\right]^2\right\},\qquad (5)$$

where $dist(\boldsymbol{y}_i,\boldsymbol{y}_j)$ represents the Euclidean distance between $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ in this paper. Denominator t adjusts the weight decay speed, which can be set to 1 without loss of generality.

Now, we consider the problem of propagating the locality relationship of the training data $\boldsymbol{Y}$ to the coding vectors $\boldsymbol{X}$, while maximizing local margin between different categories. A reasonable criterion for determining a "good" coding is to optimize the two objective functions

$$\max_{\boldsymbol{X}}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\,\boldsymbol{x}_i-\boldsymbol{x}_j\,\|_2^2\,\boldsymbol{W}_{b,ij},\qquad (6)$$

$$\min_{\boldsymbol{X}}\sum_{i=1}^{N}\sum_{j=1}^{N}\|\,\boldsymbol{x}_i-\boldsymbol{x}_j\,\|_2^2\,\boldsymbol{W}_{w,ij},\qquad (7)$$

where $\boldsymbol{X}=[\boldsymbol{x}_1,...,\boldsymbol{x}_N]\in\mathbb{R}^{K\times N}$.

3.3. Proposed LSDDL Method

In this subsection, we introduce our proposed Locality Sensitive Discriminative Dictionary Learning (LSDDL) method, which combines the essential dictionary learning framework and the two locality sensitive and discriminative objectives together. The final objective function is defined:

$$\left(\boldsymbol{D}^*, \boldsymbol{X}^*\right) = \arg\min_{\boldsymbol{D},\boldsymbol{X}} \sum_{i=1}^{N} \left(\| \boldsymbol{y}_i - \boldsymbol{D}\boldsymbol{x}_i \|_2^2 + \tau \| \boldsymbol{x}_i \|_2^2 \right) + \frac{\alpha}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \| \boldsymbol{x}_i - \boldsymbol{x}_j \|_2^2 \left(\boldsymbol{W}_{w,ij} - \lambda \boldsymbol{W}_{b,ij}\right), \tag{8}$$

where the first summation is the basic objective function for dictionary learning, which is similar to (1) in Section 2. Here, we utilize a simple l_2-norm constraint on $\boldsymbol{x}_i$. Constant $\alpha > 0$ and $\lambda > 0$ are regularization parameters that control the relative contribution of corresponding terms.

Let $\boldsymbol{S}_b$ be a diagonal matrix whose entries are column (or row, since it can be seen from the previous description that $\boldsymbol{W}_b$ is symmetric) sum of $\boldsymbol{W}_b$, so $\boldsymbol{S}_{b,ii} \triangleq \sum_j \boldsymbol{W}_{b,ij}$. We define $\boldsymbol{L}_b \triangleq \boldsymbol{S}_b - \boldsymbol{W}_b$ is the Laplacian matrix of between-class graph G_b. Similarly, we can obtain the Laplacian matrix $\boldsymbol{L}_w$ of within-class graph G_w. It can be derived that the objective function in (8) is equivalent to

$$\left(\boldsymbol{D}^*, \boldsymbol{X}^*\right) = \arg\min_{\boldsymbol{D},\boldsymbol{X}} \left\| \boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X} \right\|_F^2 + \alpha Tr(\boldsymbol{X}^T \boldsymbol{X}\boldsymbol{L}) + \tau \left\| \boldsymbol{X} \right\|_F^2, \tag{9}$$

where $\boldsymbol{L} \triangleq \boldsymbol{L}_w - \lambda \boldsymbol{L}_b$, and $Tr(\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{L})$ returns the trace of matrix product $\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{L}$.

3.4. Optimization Procedure

Similar to most discriminative dictionary learning methods, we solve (9) iteratively with first respect to each $\boldsymbol{x}_i$ and then $\boldsymbol{D}$ with all other variables fixed. These steps are continued until convergence, which is summarized in Algorithm 1.

To compute the coding coefficient matrix $\boldsymbol{X}$ with fixed $\boldsymbol{D}$, we optimize $\boldsymbol{x}_i$ alternately and fix all other $\boldsymbol{x}_j$ ($j \neq i$). Optimizing (9) is equivalent to

$$\boldsymbol{x}_i^* = \arg\min_{\boldsymbol{x}_i} \| \boldsymbol{y}_i - \boldsymbol{D}\boldsymbol{x}_i \|_2^2 + \tau \| \boldsymbol{x}_i \|_2^2 + f(\boldsymbol{x}_i), \tag{10}$$

where $f(\boldsymbol{x}_i) = \alpha[2\boldsymbol{x}_i^T(\boldsymbol{X}\boldsymbol{L}_i) - \boldsymbol{x}_i^T\boldsymbol{x}_i\boldsymbol{L}_{ii}]$ and $\boldsymbol{L}_i$ is the ith column of matrix $\boldsymbol{L}$, and $\boldsymbol{L}_{ii}$ is the (i, i)th element of matrix $\boldsymbol{L}$. We denote $g(\boldsymbol{x}_i)$ as the sum of the previous two terms in (10), then

$$\frac{\partial \left[g(\boldsymbol{x}_i) + f(\boldsymbol{x}_i) \right]}{\partial \boldsymbol{x}_i} = 2\boldsymbol{D}^T \left(\boldsymbol{D}\boldsymbol{x}_i - \boldsymbol{y}_i \right) + 2\tau \boldsymbol{x}_i + 2\alpha \boldsymbol{X}\boldsymbol{L}_i, \tag{11}$$

$$\frac{\partial^2 \left[g(\boldsymbol{x}_i) + f(\boldsymbol{x}_i) \right]}{\partial \boldsymbol{x}_i \partial \boldsymbol{x}_i^T} = 2\left[\boldsymbol{D}^T\boldsymbol{D} + \left(\tau + \alpha \boldsymbol{L}_{ii}\right)\boldsymbol{I} \right], \tag{12}$$

where $\boldsymbol{I}$ is an identity matrix. It can be seen that the Hessian matrix of objective function can be positive semi-definite if we choose proper τ, α and λ in practice, hence (10) is convex for $\boldsymbol{x}_i$. So we can set the first derivative (11) to zero and obtain the analytic solution

$$\boldsymbol{x}_i^* = \left[\boldsymbol{D}^T\boldsymbol{D} + (\tau + \alpha \boldsymbol{L}_{ii})\boldsymbol{I} \right]^{-1} \left(\boldsymbol{D}^T \boldsymbol{y}_i - \alpha \sum_{m \neq i} \boldsymbol{x}_m \boldsymbol{L}_{mi} \right). \tag{13}$$

To update dictionary $\boldsymbol{D}$ with fixed $\boldsymbol{X}$, we need to optimize the following objective function

---

Algorithm 1: Locality Sensitive Discriminative Dictionary Learning (LSDDL)

Input: Training data $\boldsymbol{Y}$ and corresponding label matrix $\boldsymbol{H}$; graph parameter k; regularization parameters τ, α, λ.
Initialization: Compute $\boldsymbol{L}$[1]. Initialize $\boldsymbol{D}_0$ and $\boldsymbol{X}_0$ referring to [9]. Normalize each atoms of $\boldsymbol{D}_0$ to have unit l_2-norm.
1: **while** not convergence **do**
2: Compute $\boldsymbol{X}$ via (13) column by column;
3: Update $\boldsymbol{D}$ via (17) or referring to [17];
4: **end while**
Output: The learned dictionary $\boldsymbol{D}$.

$$\boldsymbol{D}^* = \arg\min_{\boldsymbol{D}} \left\| \boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X} \right\|_F^2. \tag{14}$$

It can be derived that

$$\frac{\partial \left\| \boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X} \right\|_F^2}{\partial \boldsymbol{D}} = 2\left(\boldsymbol{D}\boldsymbol{X} - \boldsymbol{Y} \right) \boldsymbol{X}^T, \tag{15}$$

$$\frac{\partial^2 \left\| \boldsymbol{Y} - \boldsymbol{D}\boldsymbol{X} \right\|_F^2}{\partial \boldsymbol{D} \partial \boldsymbol{D}^T} = 2\left[\boldsymbol{I} \otimes \left(\boldsymbol{X}\boldsymbol{X}^T \right) \right], \tag{16}$$

where $\otimes$ denotes the Kronecker product. It can be demonstrated that the Hessian matrix of objective function (14) is positive semi-definite, so it is convex for $\boldsymbol{D}$. The analytical solution can also be computed by setting the first derivative (15) to zero. Then, we can obtain

$$\boldsymbol{D}^* = \boldsymbol{Y}\boldsymbol{X}^T \left(\boldsymbol{X}\boldsymbol{X}^T + \eta \boldsymbol{I} \right)^{-1}, \tag{17}$$

where $\eta > 0$ is very small to guarantee that the matrix $(\boldsymbol{X}\boldsymbol{X}^T + \eta\boldsymbol{I})$ is nonsingular. In practice, we can constrain the energy of each $\boldsymbol{d}_i$ to make the LSDDL more stable, and utilize an effective algorithm [17] to update $\boldsymbol{D}$.

3.5. Classification

With the given training data $\boldsymbol{Y}$ and corresponding label matrix $\boldsymbol{H}$, an optimized dictionary $\boldsymbol{D}$ can be learned by LSDDL. Then, we can obtain an optimized coding vector $\boldsymbol{x}_i$ for each training sample $\boldsymbol{y}_i$ over the learned discriminative dictionary $\boldsymbol{D}$ by solving Ridge Regression

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} \left\| \boldsymbol{y} - \boldsymbol{D}\boldsymbol{x} \right\|_2^2 + \tau \left\| \boldsymbol{x} \right\|_2^2, \tag{18}$$

whose solution can be analytically derived as $\boldsymbol{x} = \boldsymbol{P}\boldsymbol{y}$, where $\boldsymbol{P} = (\boldsymbol{D}^T\boldsymbol{D} + \tau\boldsymbol{I})^{-1}\boldsymbol{D}^T$, and $\tau > 0$ could ensure the invertibility of $(\boldsymbol{D}^T\boldsymbol{D} + \tau\boldsymbol{I})$. Clearly, $\boldsymbol{P}$ is independent of $\boldsymbol{y}$, so it can be pre-calculated.

When a testing sample arrives, we code it via (18), and utilize simple classifiers (*e.g.*, Nearest Neighbor Classifier) to estimate its class label. The only thing to note here is that we use the coding vectors of both training and testing samples, as well as the corresponding label matrix $\boldsymbol{H}$ of training samples, to perform classification.

---

[1] In practice, we use the normalized Laplacian matrix $\boldsymbol{L}_{\text{norm}} = \boldsymbol{I} - \boldsymbol{S}^{-1/2}\boldsymbol{W}\boldsymbol{S}^{-1/2}$ instead of $\boldsymbol{L}$, where $\boldsymbol{I}$ is an identity matrix, $\boldsymbol{S} \triangleq \boldsymbol{S}_w - \lambda \boldsymbol{S}_b$ and $\boldsymbol{W} \triangleq \boldsymbol{W}_w - \lambda \boldsymbol{W}_b$.

Table 1: LSDDL parameters, obtained by cross-validation.

	Extended YaleB	AR	Caltech101
k	7	7	5
τ	1×10^{-6}	1×10^{-8}	1×10^{-8}
α	5×10^{-4}	2×10^{-4}	7×10^{-6}
λ	0.20	0.10	0.06

4. EXPERIMENTS

We evaluate the proposed LSDDL method on famous visual classification datasets: Two face databases (Extended YaleB [18] and AR [19]), and one object categorization database (Caltech101) [20]. These datasets are widely used in previous works to evaluate various dictionary learning methods. We compare our proposed method with the following methods: the baseline support vector machine (SVM), the classical sparse-representation-based classification (SRC) [2] and collaborative-representation-based classification (CRC) [3], the novel locality-sensitive SRC (LSRC) [11] motivated by the locality-constrained linear coding (LLC) algorithm [12], and the state-of-the-art dictionary learning algorithms DLSI [5], FDDL [8], LC-KSVD [9], DDL-PC [13], and the recently proposed LPDDL [14]. For all the above competing methods, we follow the experimental settings in [9] for fair comparison. We utilize the features of the three famous databases provided by Jiang *et al.* [9][2] and tune our parameters by cross validation on each dataset. The best parameters of our LSDDL method are listed in Table 1.

The Extended YaleB [18] database includes 2,414 face images from 38 human subjects under 64 illumination conditions. All the images are cropped and scaled to 192×168. The Extended YaleB database is challenging due to varying illumination conditions and plentiful expressions. We randomly choose 32 images per person for training and the rest for testing. We repeated 20 times such a sampling process and report their average as the classification accuracy. The dimension of the random-face feature is 504, and we fix the dictionary size of 1,216 atoms. As shown in Table 2, our approach has higher accuracy than SVM, SRC, LSRC, and state-of-the-art dictionary learning algorithms FDDL, LC-KSVD, DDL-PC, and the recent LPDDL. It can be seen that our method is comparable to CRC and DLSI.

The AR face database [19] is comprised of more than 4,000 color face images of 126 persons. Each human subject has 26 frontal face images taken during two sessions. The main characteristic of the AR face database is that it includes frontal views of faces with different occlusion conditions, lighting conditions and facial expressions. All the original images are cropped to 165×120 pixels. Following the standard evaluation protocol, a set of 2,600 images from

Table 2: Classification accuracies (%) on three databases.

	Extended YaleB	AR	Caltech101
SVM	95.6	96.5	64.6
SRC	96.5	97.5	70.7
CRC	97.0	98.0	68.2
LSRC	95.7	97.4	73.4
DLSI	97.0	97.5	73.1
FDDL	96.7	97.5	73.2
LC-KSVD	96.7	97.8	73.6
DDL-PC	95.3	96.0	73.2
LPDDL	96.4	97.3	73.3
Ours	**97.0**	**98.0**	**73.6**

50 males and 50 females are extracted. For each person, we randomly select 20 images for training and the rest for testing. The results are reported from the average of 20 such random splits. The features used here are 540-dimensional random face features. We utilize a dictionary of 2,000 atoms. It can be seen from Table 2 that our LSDDL obtains an improvement over other algorithms.

The Caltech101 database [20] consists of 9,144 images from 101 common object classes and a background class. The number of samples in each category ranges from 31 to 800. Following the experimental settings in [9], 30 samples per category are selected for training and the rest for testing. We use the bag-of-visual-words (BoVW) + spatial pyramid matching (SPM) framework for feature extraction. Dense SIFT descriptors are extracted on three grids of sizes 1×1, 2×2, and 4×4 to obtain the SPM features. For a fair comparison, we utilize the vector quantization (VQ) based coding method to extract the mid-level features and utilize the max pooling approach to obtain the high dimension pooled features. Finally, we use Principal Component Analysis (PCA) to reduce the 21,504 dimensional data to 3,000 dimensions. The experimental results are summarized in Table 2. Our LSDDL method achieves the best performance again.

5. CONCLUSION

In this paper, we proposed a novel Locality Sensitive Discriminative Dictionary Learning (LSDDL) method, which focuses on making full use of the discriminative and geometrical structure of original feature. The main contribution of LSDDL is that it makes the corresponding coding coefficients over the learned dictionary preserve local relationship of original data points with the same class label and induce a large margin between points belonging to different classes. Performance evaluation on three publically visual classification datasets reveals the effectiveness of our method. Future researches will mainly focus on reducing the consumed time for cross-validation in the training phase.

[2] Online: http://www.umiacs.umd.edu/~zhuolin/projectlcksvd.html

6. REFERENCES

[1] S. Kong and D. Wang, "A brief summary of dictionary learning based approach for classification," *CoRR*, vol. abs/1205.6391, 2012.

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

[3] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, Barcelona, Spain, Nov. 2011, pp. 471-478.

[4] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Metaface learning for sparse representation based face recognition," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Hong Kong, China, Sep. 2010, pp. 1601-1604.

[5] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition (CVPR)*, San Francisco, CA, Jun. 2010, pp. 3501-3508.

[6] R. Raina, A. Battle, H. Lee, B. Packer, and A. Y. Ng, "Self-taught learning: transfer learning from unlabeled data," in *Proc. ACM Int. Conf. Mach. Learning (ICML)*, New York, NY, Jun. 2007, pp. 759-766.

[7] Q. Zhang and B. Li, "Discriminative k-svd for dictionary learning in face recognition," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition (CVPR)*, San Francisco, CA, Jun. 2010, pp. 2691-2698.

[8] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, Barcelona, Spain, Nov. 2011, pp. 543–550.

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

[10] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Supervised dictionary learning," in *Proc. Annu. Conf. Neural Inform. Process. Syst. (NIPS)*, Vancouver, Canada, Dec. 2008, pp. 1033-1040.

[11] C. Wei, Y. Chao, Y. Yeh, and Y. F. Wang, "Locality-sensitive dictionary learning for sparse representation based classification," *Pattern Recognition (PR)*, vol. 46, no. 5, pp. 1277-1287, May 2013.

[12] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," in *Proc. IEEE Conf. Comput. Vision Pattern Recognition (CVPR)*, San Francisco, CA, Jun. 2010, pp. 3360-3367.

[13] H. Guo, Z. Jiang, and L. S. Davis, "Discriminative dictionary learning with pairwise constraints," in *Proc. Asian Conf. Comput. Vision (ACCV)*, Daejeon, Korea, Nov. 2012, pp. 328-342.

[14] S. Haghiri, H. R. Rabiee, A. Soltani-Farani, S. A. Hosseini, and M. Shadloo, "Locality preserving discriminative dictionary learning," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Paris, France, Oct. 2014, pp. 5242-5246.

[15] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031-1044, Jun. 2010.

[16] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," in *Proc. Int. Joint Conf. Artificial Intell. (IJCAI)*, Hyderabad, India, Jan. 2007, pp. 708-713.

[17] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," in *Proc. Annu. Conf. Neural Inform. Process. Syst. (NIPS)*, Vancouver, Canada, Dec. 2007, pp. 801–808.

[18] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643-660, Jun. 2001.

[19] A. Martinez and R. Benavente, "The ar face database," CVC, U. A. B., Barcelona, Spain, *Tech. Rep.* 24, 1998.

[20] F. Li, R. Fergus and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," *Comput. Vision Image Understanding*, vol. 106, no. 1, pp. 59-70, Apr. 2007.

[21] T. Hastie, R. Tibshirani, and J. Friedman, "Linear methods for regression," in *The Elements of Statistical Learning*, 2nd ed. NY: Springer, 2009, pp. 43-73.

[22] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1-27, May 2011.