Locality-preserving K-SVD Based Joint Dictionary and Classifier Learning for Object Recognition

Yuan-Shan Lee
kg934283@gmail.com

Chien-Yao Wang
102582011@cc.ncu.edu.tw

Seksan Mathulaprangsan
mr.sekk@gmail.com

Jia-Hao Zhao
ross800127@gmail.com

Jia-Ching Wang
jcw@csie.ncu.edu.tw

Department of Computer Science and Information Engineering
National Central University, Taoyuan, Taiwan

ABSTRACT

This paper concerns the development of locality-preserving methods for object recognition. The major purpose is consideration of both descriptor-level locality and image-level locality throughout the recognition process. Two dual-layer locality-preserving methods are developed, in which locality-constrained linear coding (LLC) is used to represent an image. In the learning phase, the discriminative locality-preserving K-SVD (DLP-KSVD) in which the label information is incorporated into the locality-preserving term is proposed. In addition to using class labels to learn a linear classifier, the label-consistent LP-KSVD (LCLP-KSVD) is proposed to enhance the discriminability of the learned dictionary. In LCLP-KSVD, the objective function includes a label-consistent term that penalizes sparse codes from different classes. For testing, additional information about the locality of query samples is obtained by treating the locality-preserving matrix as a feature. The recognition results that were obtained in experiments with the Caltech101 database indicate that the proposed method outperforms existing sparse coding based approaches.

Keywords

K-SVD; D-KSVD; locality-preserving; joint dictionary learning; object recognition

1. INTRODUCTION

Object recognition is that aspect of computer vision that involves recognizing objects from photographs or videos. Although humans can recognize objects with ease, machines face numerous difficulties in so doing: for example the variable orientation of objects relative to a camera may have a huge effect on recognition performance, and the variation of the size of a target object with the distance from the object to the camera can greatly affect the accuracy of recognition. These difficulties make object recognition a highly challenging task.

Generally, the process of object recognition can be divided into two main steps, which are feature extraction and classification. Diverse methods exist for extracting the features from image files. Scale-invariant feature transformation (SIFT) [12] is a method that solves the problem of the rotated image. Another famous approach to resolving scale- and rotation-invariance is based on speeded-up robust features (SURF) [1]; the method creates a SURF descriptor using a precomputed integral image. Spatial pyramid matching (SPM) [8] is a technique that solves the problem of scale variation. SPM splits an image into blocks and computes histograms of low-level descriptors, such as SIFT, for each block. The histograms of each block are then concatenated into a feature vector to represent the image. Originally, a codebook that is generated by vector quantization (VQ) was utilized to compute the histograms of descriptors. To avoid the quantization loss that is associated with VQ, Yang *et al.* [17] proposed a linear SPM that is based on sparse coding (ScSPM). The ScSPM has been proven to be effective with simple linear SVMs. Yu *et al.* [19] indicated that locality is more important than sparsity in nonlinear learning. Based on the work of Yu *et al.* [19], locality-constrained linear coding (LLC) was developed [14]. Zhang *et al.* [22] presented low-rank sparse coding (LRSC), which jointly codes the local region of descriptors for image classification. More recently, locality-constrained low-rank coding (LCLR) [5] which benefits from joint coding and locality, has also been developed. The aforementioned locality-based representations [5, 14, 19, 22] provide favorable classification accuracy, even with linear classifiers.

In the classification step, the extracted image representations are fed into classifiers such as the support vector machine (SVM) [2], the sparse representation-based classification (SRC) [16] and others. Recently, the advent of supervised dictionary-based methods [3, 7, 10, 18, 20, 21] has massively improved object recognition; of these methods, discriminative K-SVD (D-KSVD) [21] is regarded as one of the most representative. D-KSVD jointly learns the dictionary and the linear classifier. The classification error term is incorporated into the reconstruction error term in the objective function. This scheme improves the discriminative power of the dictionary in face recognition task. Label-consistent K-SVD (LC-KSVD) [6] was developed by adding a label-consistent term to D-KSVD to improve the distinguishability of the learned sparse codes. To preserve

MM '16, October 15-19, 2016, Amsterdam, Netherlands
© 2016 ACM. ISBN 978-1-4503-3603-1/16/10...$15.00
DOI: http://dx.doi.org/10.1145/2964284.2967267

the local structure of data in dictionary learning, Wei *et al.* [15] proposed a locality-sensitive SRC (L-SRC), which has a close-form solution throughout the learning process. Liu *et al.* [11] developed a locality-preserving K-SVD (LP-KSVD), which integrates the locality penalty term into the D-KSVD framework. However, to the best of our knowledge, although various locality-based methods have been proposed, most consider either locality on SIFT descriptors [5, 14, 19, 22], or locality on image representations [11, 15].

To exploit fully the potential of the locality-preserving technique for object recognition task, this paper develops two dual-layer locality-preserving methods. First, the discriminative LP-KSVD (DLP-KSVD), which incorporates the label information into locality-preserving term, is proposed. Second, the label-consistent LP-KSVD (LCLP-KSVD) is developed to improve the discriminability of the learned dictionary. By employing the LLC for image representation, the proposed methods take both descriptor-level locality and image-level locality into consideration.

The rest of this paper is organized as follows. Section 2 describes works relevant to the developed model. Section 3 elucidates details of the proposed model. Section 4 presents experimental results. Finally, section 5 draws conclusions and summarizes possibilities for future work.

2. RELATED WORKS

2.1 Locality-preserving Projections

LPP [4] is a linear dimensionality reduction algorithm. It finds a linear transformation that exposes the locality of the input signals. The objective function of the LPP can be presented as,

$$\min \sum_{ij} (x_i - x_j)\mathbf{L}_{ij} \tag{1}$$

where x_i is the low-dimensional representation of input signal y_i. $\mathbf{L}$ is a similarity matrix which is defined as:

$$\mathbf{L}_{ij} = \begin{cases} exp(\frac{-\|y_i - y_j\|^2}{\rho}) & ; e(y_i, y_j) = 1 \\ 0 & ; e(y_i, y_j) = 0 \end{cases} \tag{2}$$

where $e(y_i, y_j) = 1$ for k-nearest y_j according to y_i. ρ denotes a tunable parameter.

2.2 Sparse Representation Classifier

Sparse representation classifier (SRC) [16] is a classifier that exploits sparse representation. The dictionary comprises training data. Following sparse coding, the reconstruction error of each class can be calculated. The recognition results are obtained by minimizing the reconstruction error.

Let $\mathbf{Y}$ be a training dataset that has N signals, each of which has d dimensions, so $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_N] \in R^{d \times N}$. For the reconstruction of $\mathbf{Y}$, the optimization formula of dictionary learning can be conducted by solving the following equation.

$$< \mathbf{D}, \mathbf{X} >= \arg \min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \quad s.t. \forall i, \|\mathbf{x}_i\|_0 \leq T \tag{3}$$

where $\mathbf{X} = [\mathbf{x}_1, ..., \mathbf{x}_N] \in R^{m \times N}$ represents the sparse codes of $\mathbf{Y}$, $\mathbf{D} = [\mathbf{d}_1, ..., \mathbf{d}_m] \in R^{d \times m}$ is the trained overcomplete dictionary. m denotes the total number of dictionary atoms. The sparse constraint T is applied to control the sparseness of nonzero items in each sparse code.

2.3 Joint Dictionary Learning

Pham and Venkatesh [13] developed joint dictionary learning (JDL) as a method that combines dictionary learning and classifier learning. The objective function is defined as,

$$< \mathbf{W}, \mathbf{b} >= \arg \min_{\mathbf{W},\mathbf{b}} \|\mathbf{H} - \mathbf{W}\mathbf{X} - \mathbf{b}\|_2^2 + \gamma \|\mathbf{W}\|_2^2 \tag{4}$$

where $\mathbf{W}$ and $\mathbf{b}$ are parameters of the linear classifier $\mathbf{H} = \mathbf{W}\mathbf{X} + \mathbf{b}$. $\mathbf{H}$ is the target matrix of the linear classifier. Each column in $\mathbf{H}$ is a vector for which $\mathbf{h}_i = [0, ...1, ..., 0, 0]$ is class number that is associated with each input sample. $\|\mathbf{H} - \mathbf{W}\mathbf{X} - \mathbf{b}\|_2^2$ is the recognition error. $\|\mathbf{W}\|_2^2$ is the regularization penalty term. To simplify the objective function, $\mathbf{b}$ is set to zero.

The dictionary learning objective function is based on the combination of l_0-norm and the linear classifier objective function in (4):

$$\begin{aligned} < \mathbf{D}, \mathbf{X}, \mathbf{W} >= \arg \min_{\mathbf{D},\mathbf{X},\mathbf{W}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2 + \beta \|\mathbf{H} - \mathbf{W}\mathbf{X}\|_2^2 \\ + \gamma \|\mathbf{W}\|_2^2 \quad s.t. \forall i, \|\mathbf{x}_i\|_0 \leq T \end{aligned} \tag{5}$$

where the parameters β and γ are used to balance the influence of each term.

2.4 Discriminative K-SVD

Zhang and Li [21] developed a method of simultaneously training the dictionary and the classifier, called discriminative K-SVD. The objective function of D-KSVD is defined as,

$$\begin{aligned} < \mathbf{D}, \mathbf{X}, \mathbf{W} >= \arg \min_{\mathbf{D},\mathbf{X},\mathbf{W}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2 + \beta \|\mathbf{H} - \mathbf{W}\mathbf{X}\|_2^2 \\ s.t. \forall i, \|\mathbf{x}_i\|_0 \leq T \end{aligned} \tag{6}$$

where $\mathbf{H} \in R^{c \times N}$ is the object matrix of c-classes linear classifier. Each column in $\mathbf{H}$ is a vector such that $\mathbf{h}_i = [0, ...1, ..., 0, 0]$ is the class number for each input data. $\mathbf{W}$ is a linear classifier. β is a parameter that modifies the influence between reconstruction error and recognition error. Notably, the regularization penalty term $\|\mathbf{W}\|_2^2$ in joint dictionary learning is discarded to simplify the objective function.

To obtain the dictionary, sparse codes, and linear classifier simultaneously, the objective function in (6) is modified as,

$$\begin{aligned} < \mathbf{D}, \mathbf{X}, \mathbf{W} >= \arg \min_{\mathbf{D},\mathbf{X},\mathbf{W}} \left\| \begin{pmatrix} \mathbf{Y} \\ \sqrt{\beta}\mathbf{H} \end{pmatrix} - \begin{pmatrix} \mathbf{D} \\ \sqrt{\beta}\mathbf{W} \end{pmatrix} \mathbf{X} \right\|_2^2 \\ s.t. \forall i, \|\mathbf{x}_i\|_0 \leq T \end{aligned} \tag{7}$$

K-SVD can be used to obtain dictionary $\mathbf{D}$ and linear classifier $\mathbf{W}$ at the same time. After training, the sparse code $\mathbf{x}'$ of the query $\mathbf{y}'$ can be identified. The recognition result is computed by multiplying the linear classifier $\mathbf{W}$ by the sparse code $\mathbf{x}'$:

$$\mathbf{l} = \mathbf{W}\mathbf{x}' \tag{8}$$

where $\mathbf{l}$ is the recognition result.

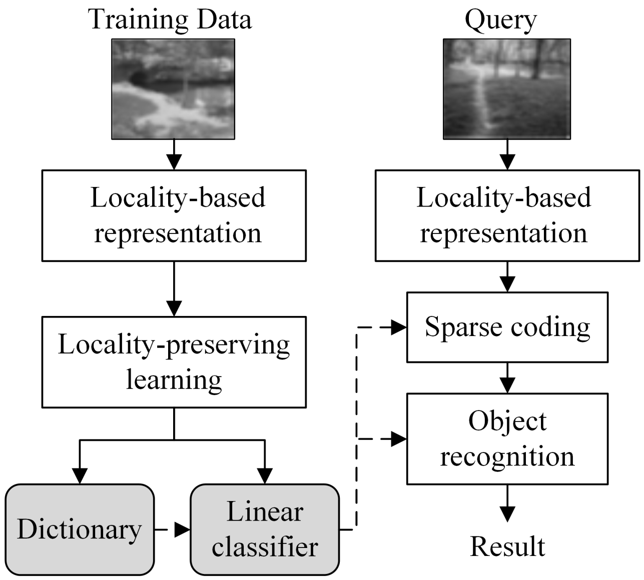

Figure 1: An overview of the proposed method.

3. DUAL-LAYER LOCALITY-PRESERVING K-SVD FOR OBJECT RECOGNITION

This work developed two dual-layer locality-preserving methods for object recognition. Sections 3.1 and 3.2 address the difference between them. Both of the proposed methods comprise two main stages, which are feature extraction and classifier learning. In the first stage, LLC [14] features are extracted from the images. In the second stage, the sparse codes are learned from LLC features based on the proposed methods, DLP-KSVD and LCLP-KSVD. Thereafter, the linear classifier is used with the learned sparse codes to recognize objects. In so doing, both descriptor-level locality and image-level locality are considered. Figure 1 displays an overview of the proposed dual-layer locality-preserving methods.

3.1 Discriminative LP-KSVD (DLP-KSVD)

The proposed DLP-KSVD is derived by combining the preservation of locality with dictionary learning. Unlike the work of Liu *et al.* [11], a discriminative target matrix is proposed to preserve locality. Label information is incorporated into the proposed locality target matrix to further enhance the discrimination of learned sparse codes. Adding the locality-preserving term into the original D-KSVD yields the objective function as,

$$< \mathbf{D},\mathbf{P},\mathbf{W},\mathbf{X} >= \arg\min_{\mathbf{D},\mathbf{P},\mathbf{W},\mathbf{X}} \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_2^2 + \alpha \left\|\check{\mathbf{L}}-\mathbf{P}\mathbf{X}\right\|_2^2 + \beta\|\mathbf{H}-\mathbf{W}\mathbf{X}\|_2^2 \quad s.t. \forall i, \|\mathbf{x}_i\|_0 \leq T \tag{9}$$

where α and β are parameters that fine-tune the effects of reconstruction term, locality term, and recognition term on each other. $\mathbf{P}$ denotes a linear transformation matrix that maps $\mathbf{X}$ into locality-preserving code space. $\mathbf{W}$ denotes the linear classifier. $\check{\mathbf{L}}$ indicates the discriminative locality-preserving matrix, defined as,

$$\check{\mathbf{L}}_{ij} = \begin{cases} exp(\frac{-\|y_i-y_j\|^2}{\rho}) & ; \check{e}(y_i,y_j)=1 \\ -(1-exp(\frac{-\|y_i-y_j\|^2}{\rho})) & ; \check{e}(y_i,y_j)=0 \end{cases} \tag{10}$$

where $\check{e}(y_i,y_j)=1$ for k-nearest y_j, which belongs to the same class as y_i, and $\check{e}(y_i,y_j)=0$ for k-nearest y_j, which belongs to a different class from y_i.

Equation (9) can be rewritten by using K-SVD to minimize the objective function as,

$$< \mathbf{D},\mathbf{P},\mathbf{W},\mathbf{X} >= \arg\min_{\mathbf{D},\mathbf{P},\mathbf{W},\mathbf{X}} \left\| \begin{pmatrix} \mathbf{Y} \\ \sqrt{\alpha}\check{\mathbf{L}} \\ \sqrt{\beta}\mathbf{H} \end{pmatrix} - \begin{pmatrix} \mathbf{D} \\ \sqrt{\alpha}\mathbf{P} \\ \sqrt{\beta}\mathbf{W} \end{pmatrix} \mathbf{X} \right\|_2^2 \quad s.t. \forall i, \|\mathbf{x}_i\|_0 \leq T \tag{11}$$

3.2 Label-consistent LP-KSVD (LCLP-KSVD)

Except for using class labels to learn a linear classifier, the label consistent term is utilized to improve the discriminability of the learned dictionary. Following the work of Zhang and Li [21], the label information is associated with each dictionary atom in the learning process. Adding the label consistent term into the original LP-KSVD yields the objective function as,

$$< \mathbf{D},\mathbf{P},\mathbf{W},\mathbf{A},\mathbf{X} >= \arg\min_{\mathbf{D},\mathbf{P},\mathbf{W},\mathbf{A},\mathbf{X}} \|\mathbf{Y}-\mathbf{D}\mathbf{X}\|_2^2 + \alpha\|\mathbf{L}-\mathbf{P}\mathbf{X}\|_2^2 + \beta\|\mathbf{H}-\mathbf{W}\mathbf{X}\|_2^2 + \gamma\|\mathbf{Q}-\mathbf{A}\mathbf{X}\|_2^2 \quad s.t. \forall i, \|\mathbf{x}_i\|_0 \leq T \tag{12}$$

where α, β and γ are parameters that fine-tune the effects of reconstruction term, locality term, recognition term, and label consistent term on each other. $\mathbf{Q}$ denotes the discriminative sparse codes of input signals [21].

Similarly, Eq. (12) can be rewritten by using K-SVD to minimize the objective function as,

$$< \mathbf{D},\mathbf{P},\mathbf{W},\mathbf{A},\mathbf{X} >= \arg\min_{\mathbf{D},\mathbf{P},\mathbf{W},\mathbf{A},\mathbf{X}} \left\| \begin{pmatrix} \mathbf{Y} \\ \sqrt{\alpha}\mathbf{L} \\ \sqrt{\beta}\mathbf{H} \\ \sqrt{\gamma}\mathbf{Q} \end{pmatrix} - \begin{pmatrix} \mathbf{D} \\ \sqrt{\alpha}\mathbf{P} \\ \sqrt{\beta}\mathbf{W} \\ \sqrt{\gamma}\mathbf{A} \end{pmatrix} \mathbf{X} \right\|_2^2 \quad s.t. \forall i, \|\mathbf{x}_i\|_0 \leq T \tag{13}$$

3.3 Locality-incorporated Dictionary Learning

For a query sample $\mathbf{z}$, the corresponding sparse code can be obtained as,

$$\mathbf{x}' = \arg\min_{\mathbf{x}'} \|\mathbf{z}-\mathbf{D}\mathbf{x}'\|_2^2 \quad s.t. \forall i, \|{\mathbf{x}'}_i\|_0 \leq T \tag{14}$$

Finally, after the training step, the recognition result can be computed using Eq. (8). To incorporate the locality information into the query, the locality-preserving matrix is regarded as a new feature. The objective function is rewritten as,

$$< \mathbf{D}',\mathbf{P},\mathbf{W},\mathbf{X}, >= \arg\min_{\mathbf{D},\mathbf{P},\mathbf{W},\mathbf{X}} \|\mathbf{Y}'-\mathbf{D}'\mathbf{X}\|_2^2 + \beta\|\mathbf{H}-\mathbf{W}\mathbf{X}\|_2^2 \quad s.t. \forall i, \|\mathbf{x}_i\|_0 \leq T \tag{15}$$

Table 1: Accuracy rates of different training data sizes on Caltech101 dataset

# of Tr. Images	5	10	15	20	25	30
SRC	49.00	58.81	63.90	66.96	69.47	71.83
D-KSVD	52.59	62.08	66.94	69.66	72.00	74.21
LC-KSVD	52.49	62.14	66.80	69.35	71.49	73.67
LP-KSVD	52.51	62.09	66.90	69.65	71.99	74.23
LCLP-KSVD	55.69	64.59	**69.95**	71.90	74.50	75.72
LCLP-KSVD+	55.66	**64.78**	69.59	71.86	74.63	75.62
DLP-KSVD	**55.73**	64.75	69.57	72.02	74.80	76.67
DLP-KSVD+	55.67	64.67	69.48	**72.26**	**74.95**	**76.79**

where $\mathbf{Y}' = \begin{pmatrix} \mathbf{Y} \\ \sqrt{\alpha}\hat{\mathbf{L}} \end{pmatrix}$, $\mathbf{D}' = \begin{pmatrix} \mathbf{D} \\ \sqrt{\alpha}\mathbf{P} \end{pmatrix}$. K-SVD was used to solve the objective function as in (11). The locality-preserving feature $\hat{\mathbf{L}}$ is set to $\check{\mathbf{L}}$ in DLP-KSVD, whereas it is set to set to $\mathbf{L}$ in LCLP-KSVD.

When a query sample $\mathbf{z}$ is addressed, the training data $\mathbf{Y}$ were utilized to calculate locality-preserving feature which is defined as,

$$\bar{\mathbf{L}}_j = \begin{cases} exp(\frac{-\left\|z-y_j\right\|^2}{\rho}) & ; e(z, y_j) = 1 \\ 0 & ; e(z, y_j) = 0 \end{cases} \tag{16}$$

where $e(z, y_j) = 1$ for k-nearest y_j according to z.

We combine the query $\mathbf{z}$ and $\bar{\mathbf{L}}$. Extracting sparse code $\bar{\mathbf{x}}$ and then utilize (8) to obtain the recognition result. The objective function is defined as,

$$\bar{\mathbf{x}} = \arg\min_{\bar{\mathbf{x}}} \left\| \begin{pmatrix} \mathbf{z} \\ \sqrt{\alpha}\bar{\mathbf{L}} \end{pmatrix} - \mathbf{D}'\bar{\mathbf{x}} \right\|_2^2 \quad s.t. \forall i, \left\|\bar{\mathbf{x}}_i\right\|_0 \leq T \tag{17}$$

where $\mathbf{D}'$ denotes the locality-incorporated dictionary for coding the query.

4. EXPERIMENTS

Experiments were performed on Caltech101 [9] database. Caltech101 contains 102 classes (101 main classes and a background class) and a total of 9144 images. Each class is associated with 31 to 800 images of various sizes and the sizes vary hugely among classes. SRC [16], D-KSVD [21], LC-KSVD [6], and LP-KSVD [11] were selected as baseline algorithms for comparison. SPM [8] was used to extract the features in the SRC, D-KSVD, LC-KSVD, and LP-KSVD methods. For simplicity, the proposed methods with locality-incorporated dictionary learning are abbreviated as DLP-KSVD+ and LCLP-KSVD+, respectively. LP-KSVD, DLP-KSVD and LCLP-KSVD used 60 nearest neighbors ($k = 60$) in the calculation of the locality-preserving matrix. The parameters α, β, and γ are set to 0.001. To obtain a reliable result, all of the experiments were carried out ten times and the mean value of the results was obtained.

4.1 Comparison of Effects of Variously Sized Training Data

To determine the effect on the size of the training dataset, 5, 10, 15, 20, and 25 images per class were selected at random for training with dictionary size 510, 1020, 1530, 2040, 2550, and 3060 respectively, and the remaining images were used as the testing data.

The experimental results, presented in Table 1, confirm that the proposed DLP-KSVD performed best. Specifically, LCLP-KSVD is 1.5% to 2.5% better than LP-KSVD with 25 and 30 training samples per class, respectively. DLP-KSVD outperforms LP-KSVD by approximate 1% in the same condition. DLP-KSVD+ and LCLP-KSVD+ underperformed DLP-KSVD and LCLP-KSVD when the number of training samples was small but outperformed them when the number of training samples was large. This observation indicates that the incorporated locality information becomes more important as the number of training samples increases.

4.2 Comparison of Effects of Variously Sized Dictionaries

To elucidate the effect of dictionary size on the Caltech101 dataset, 30 images were randomly selected to provide training data. The dictionary size was set to 510, 1020, 1530, 2040, 2550, or 3060. As shown in Table 2, the proposed methods achieve the most accuracy for all dictionary sizes. The proposed DLP-KSVD (76.79%) is much better than the other sparse coding based methods in Table 3, which are SRC (71.83%), D-KSVD (74.21%), LC-KSVD (73.67%), and is approximately 1.5% better than the LPKSVD.

Table 2: Accuracy rates of different dictionary sizes on Caltech101 dataset

Dictionary Size	510	1020	1530	2040	2550	3060
SRC	71.57	72.00	71.75	71.50	71.60	71.83
D-KSVD	72.03	72.42	71.97	70.81	72.00	74.21
LC-KSVD	72.16	72.22	71.19	71.32	71.19	73.67
LP-KSVD	71.89	72.35	71.96	72.01	71.95	74.23
LCLP-KSVD	73.10	74.19	74.17	73.79	74.33	75.72
LCLP-KSVD+	73.02	74.21	74.43	73.76	74.42	75.62
DLP-KSVD	73.11	74.34	**74.48**	74.62	74.91	76.67
DLP-KSVD+	**73.40**	**74.75**	74.47	**74.68**	**75.30**	**76.79**

5. CONCLUSIONS AND FUTURE WORKS

This work developed the dual-layer locality-preserving model, which improves conventional dictionary learning approaches by jointly considering descriptor-level locality and image-level locality. The addition of a locality-preserving term improves the distinguishability of the classifier for object recognition problems. Specifically, label information is incorporated into the locality-preserving matrix. A label-consistent term is utilized to enhance the discrimination of learned sparse codes. Experimental results demonstrate the superiority of the proposed DLP-KSVD and LCLP-KSVD with Caltech101 database. The locality-incorporated dictionary learning has been shown to play an important role when the training dataset is sufficiently large. Possible future work includes extending the presented approach to the online dictionary learning, which is highly applicable to large-scale data processing. Also, treating the locality-preserving matrix as the target of another nonlinear classifier to perform discriminative feature extractions, will be of interest.

6. REFERENCES

[1] H. Bay and A. Ess. Speeded-up robust features (SURF). *Comput. Vision and Image Understanding*, 110:346–359, 2008.

[2] C. Cortes and V. Vapnik. Support-vector networks. *Mach. Learning*, 20:273–297, 1995.

[3] M. J. Gangeh, A. Ghodsi, and M. S. Kamel. Kernelized Supervised Dictionary Learning. *IEEE Trans. on Signal Process.*, 61(19):4753–4767, 2013.

[4] X. He and P. Niyogi. Locality preserving projections. In *Proc. Conf. Advances in Neural Inform. Process. Syst.*, pages 153–160, 2003.

[5] Z. Jiang, P. Guo, and L. Peng. Locality-constrained low-rank coding for image classification. In *Proc. the Twenty-Eighth AAAI Conf. Artificial Intell.*, pages 2780–2786, 2013.

[6] Z. Jiang, Z. Lin, and L. S. Davis. Learning a discriminative dictionary for sparse coding via label consistent. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 1697–1704, 2011.

[7] S. Lazebnik and M. Raginsky. Supervised learning of quantizer codebooks by information loss minimization. *IEEE Trans. on Pattern Anal. and Mach. Intell.*, 31(7):1294–1309, 2009.

[8] S. Lazebnik and C. Schmid. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 2169–2178, 2006.

[9] F. 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 and Image Understanding*, 106(1):59–70, 2007.

[10] X. C. Lian, Z. Li, C. Wang, B. L. Lu, and L. Zhang. Probabilistic models for supervised dictionary learning. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 2305–2312, 2010.

[11] W. Liu, Z. Yu, M. Yang, L. Lu, and Y. Zou. Joint kernel dictionary and classifier learning for sparse coding via locality preserving k-svd. In *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, pages 1–6, 2015.

[12] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. of Comput. Vision*, 60(2):91–110, 2004.

[13] D. S. Pham and S. Venkatesh. Joint learning and dictionary construction for pattern recognition. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, 2008.

[14] 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 and Pattern Recognition (CVPR)*, pages 3360–3367, 2010.

[15] C.-P. Wei, Y.-W. Chao, Y.-R. Yeh, and Y.-C. F. Wang. Locality-sensitive dictionary learning for sparse representation based classification. *Pattern Recognition*, 46(5):1277–1287, 2013.

[16] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. on Pattern Anal. and Mach. Intell.*, 31(2):210–227, 2009.

[17] J. Yang, K. Yu, Y. Gong, and T. Huang. Linear spatial pyramid matching using sparse coding for image classification. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 1794–1801, 2009.

[18] M. Yang, L. Zhang, X. Feng, D. Zhang, and H. Kong. Fisher discrimination dictionary learning for sparse representation. In *Proc. IEEE Conf. Comput. Vision*, pages 543–550, 2011.

[19] K. Yu, T. Zhang, and Y. Gong. Nonlinear learning using local coordinate coding. In *Advances Neural Inf. Proc. Syst. 22*, pages 2223–2231, 2009.

[20] H. Zhang, Y. Zhang, and T. S. Huang. Simultaneous discriminative projection and dictionary learning for sparse representation based classification. *Pattern Recognition*, 46(1):346–354, 2013.

[21] Q. Zhang and B. Li. Discriminative K-SVD for dictionary learning in face recognition. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, pages 2691–2698, 2010.

[22] T. Zhang, B. Ghanem, S. Liu, C. Xu, and N. Ahuja. Low-rank sparse coding for image classification. In *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, pages 281–288, 2013.