2014 IEEE International Conference on Robotics & Automation (ICRA)
Hong Kong Convention and Exhibition Center
May 31 - June 7, 2014. Hong Kong, China

Discriminative Dictionary Learning via Shared Latent Structure for Object Recognition and Activity Recognition

Hongcheng Wang[1], Hongbo Zhou[2], and Alan Finn[1]

***Abstract*— We propose a novel low-dimensional discriminative dictionary learning approach for multi-class classification tasks, Latent Structure based Discriminative Dictionary Learning (LS-DDL). Our approach first projects features and class labels onto a shared latent structure space, and then generates a discriminative and low-dimensional input to a discriminative dictionary learning framework. LS-DDL learns a more discriminative and lower-dimensional dictionary than existing dictionary learning methods. Therefore we obtain high recognition accuracy with a small number of low-dimensional dictionary atoms. The low dimensionality also improves the efficiency in storage and testing. In addition, the latent structure projection eliminates the classifier weighting parameter in existing discriminative dictionary learning approaches. We validate the effectiveness and efficiency of the proposed approach through a series of experiments on image-based face recognition and video-based activity recognition. Our results show that the proposed approach obtains much higher recognition accuracy with a small number of dictionary atoms, and costs much less computational time than state-of-the-art methods.**

I. INTRODUCTION

We consider the following linear decomposition-based dictionary learning,

$$X \cong DW, \tag{1}$$

where X is a given dataset, D is a dictionary, and W is a coefficient matrix or representation of data X on D. Many criteria or constraints could be introduced to make Equation (1) well-defined. A classical model, K-SVD [1], is designed to minimize a Frobenius norm loss and a sparsity constraint $||W||_0$ simultaneously. Therefore, K-SVD improves the data's interpretability and expressive power with sparse coefficients while eliminating noise and outliers.

Despite many successful applications, K-SVD is unsuitable for classification tasks, where desirable dictionaries should be not only expressive, but also discriminative. This problem has been partially addressed in recent work, such as Discriminative K-SVD (D-KSVD) [25] and label consistent K-SVD (LC-KSVD) [7], where additional constraints are added to a K-SVD cost function to account for the classifier training cost. Taking D-KSVD as an example, the optimization problem is formulated as

$$\min_{D,P,W} ||X - DW||_F^2 + \lambda_c ||Y - PW||_F^2 + \lambda_s ||W||_0, \tag{2}$$

where Y encodes training label information and P is a projection matrix [25]. The second term extends K-SVD with a classifier construction loss where λ_c is the classifier weighting parameter. This parameter balances the expressive power and discriminative power of the learned dictionary. By setting $\lambda_c = 0$, Equation (2) results in the original K-SVD formulation. The parameter λ_s is the sparsity weighting parameter. For LC-KSVD, there is an additional parameter controlling the relative contribution of a label consistent term.

The discriminative models of dictionary learning have generally shown superior performance in classification problems compared to generative models (e.g., K-SVD). However, existing state-of-the-art dictionary learning models such as D-SVD and LC-KSVD have several practical disadvantages. First, the feature vector X is usually high dimensional, and therefore the dimensionality of the learned dictionary remains very high. This makes this approach less efficient for data storage, and algorithm training and testing, especially for "big data" and high dimensionality applications. Second, several free parameters are involved including the classifier weighting parameter, the sparsity parameter, and/or the label-consistent parameter. These parameters demand considerable effort to tune for the best performance.

To address the aforementioned problems, we introduce a novel formulation for dictionary learning as, $X \cong GDW$, where G is a matrix for dimensionality reduction. Based on this, we develop a simple yet effective discriminative dictionary learning approach for general multi-class classification tasks. Figure 1 illustrates the scheme of the proposed LS-DDL method. It optimizes the dictionary and classifier parameters simultaneously, and has the advantages over previous work in the following aspects:

- The dictionary learned from LS-DDL is more compact. Our dictionary learning approach achieves reductions in both sample size and dimensionality, while conventional dictionary learning methods only achieve reduction in sample size. The low-dimensional dictionary learning not only alleviates the effects of noise in the input data to avoid over-fitting, but also obtains more efficient data storage, as well as efficiency in dictionary learning and effective object classification.
- The dictionary learned from LS-DDL is more discriminative for classification tasks. Our discriminative dictionary takes input from a shared latent space through use of the original feature vectors and the class labels. The latent structure itself is very powerful in classification. On the other hand, the D-KSVD and its variants simply take the original feature vectors as input. Therefore, our method is highly discriminative in classification such that we can achieve very good classification accuracy

[1]United Technologies Research Center (UTRC), East Hartford, Connecticut, USA. [2]now at Southern Illinois University, Carbondale, Illinois, USA. Corresponding author: Hongcheng Wang wangh1@utrc.utc.com

978-1-4799-3684-7/14/$31.00 ©2014 IEEE

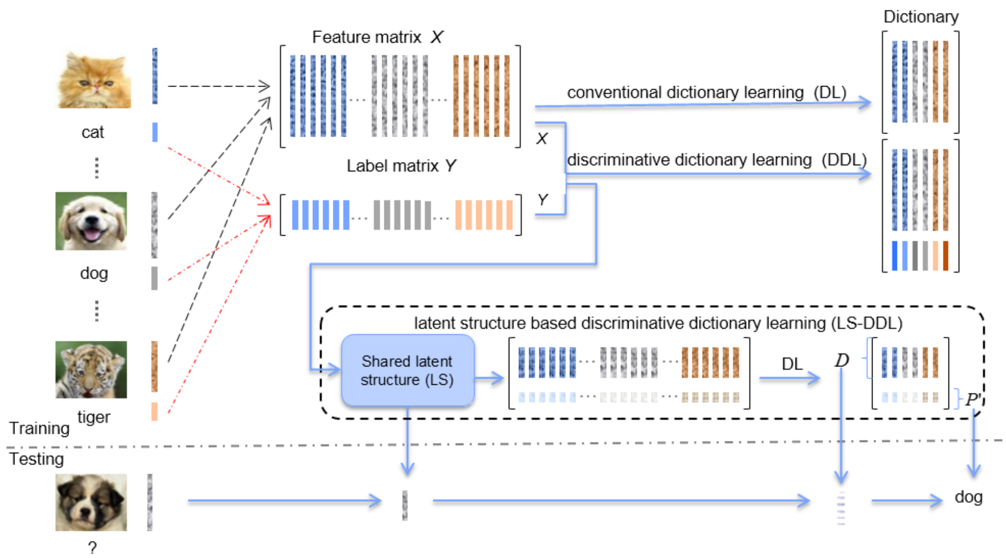

Fig. 1. *Illustration of the shared Latent Structure based Discriminative Dictionary Learning (LS-DDL) scheme, which results in a low-dimensional and discriminative dictionary for classification. D and P' are the learned dictionary and projection matrix respectively as will be explained in Section III. Conventional dictionary learning methods, such as K-SVD, do not consider label information, and reduce sample size but not sample dimensionality. Though discriminative dictionary learning (DDL) approach (such as D-KSVD) uses label matrix Y, it only reduces sample size. Our method (LS-DDL) not only incorporates label information, but also achieves reductions in both sample size and dimensionality.*

with a small number of dictionary atoms.

- Our approach is insensitive to the classifier weighting parameter. The classifier weighting parameter essentially weights the contributions of the original data and the class labels, which are projected onto a canonical latent space in LS-DDL. The classifier weighting parameter is automatically embedded in the latent structure projection.

The rest of this paper is organized as follows. We start by reviewing closely related work in dictionary learning in Section II, and then proceed to formulate and discuss our discriminative dictionary learning approach via shared latent structure in Section III. Section IV presents experimental verification of the proposed method. Finally, we conclude this paper in Section V.

II. Related Work

We roughly categorize various dictionary learning methods into two families, generative methods and discriminative methods. Generative dictionary learning methods are designed to extract or build a compact set of exemplars, while discriminative dictionary learning methods also consider incorporating discriminative information for successive classification tasks.

Representatives of generative dictionary learning methods include K-means [11] and K-SVD [1]. Successful applications have been reported that use K-means to construct visual dictionaries for image-based object recognition [10], video-based activity recognition [8], [16], etc. However, this approach usually requires a very large dictionary. As a generalization of K-means, K-SVD [1] was developed to build dictionaries from linear decompositions of the training data. More recently, work by Nguyen et al. presents an unsupervised dimensionality reduction method for dictionary learning [15]. These generative dictionary learning methods focus on the expressive power of the learned dictionary without explicitly considering the discriminative power. In general, dictionaries learned using generative methods are shown to be inferior to these learned using discriminative methods in many classification tasks [25].

Discriminative dictionary learning methods have attracted considerable attention. One research direction was to learn multiple complementary dictionaries [23], [13], and the other was to construct a single discriminative dictionary for all classes [25], [18]. In each direction, various criteria, such as Fisher's discriminant analysis [6], [24] and maximizing mutual information [18], have been explored to incorporate discriminative information. Besides the difference in discriminative criteria, some methods trained dictionary learning and classifier parameters separately [13], [17], while others optimized dictionary and classifier parameters simultaneously [17], [23], [14], [25]. More recently, work by Guo et al. [5] introduced a pairwise sparse code error into the discriminative dictionary learning framework. Despite the advances in incorporating discriminativeness in classifier construction, none of the aforementioned work addresses the tuning problem of the classifier weighting parameter or considers simultaneous dimensionality reduction for robust and efficient classification.

III. Discriminative Dictionary Learning via Shared Latent Structure (LS-DDL)

A. Notation and Setting

In a general multi-class classification setting, we are given a training dataset that consists of n examples from M classes. Denote the training dataset by $\{(x_i, y_i)\}_{i=1}^n$, where $x_i \in \mathcal{R}^p$ is the ith example, and $y_i \in \{1, \cdots, M\}$ is the class label for the corresponding feature vector, where c is the number of classes. A data (observation) matrix is formed from $X = [x_1, \cdots, x_n]$ of size $p \times n$; a basic label matrix is denoted as $y = [y_1, \cdots, y_n]$ of size $1 \times n$. A dummy coding matrix Y for label information can be constructed from y by replacing each label y_i with a standard basis of $\mathcal{R}^M$, for example, $y_i = [0, 0, ..., 0, 1, 0, 0]^T$ has 1 at entry j denoting that the sample belongs to the j^{th} class.

Given a test example $x_t \in \mathcal{R}^p$, the task is to identify its label. Since these feature vectors might be of high dimensionality, it is usually problematic to directly define a proper distance metric for classification. A plausible alternative is to approximate x_t by a linear combination of atoms from a dictionary [22].

To learn a dictionary from a training dataset X, we consider the following approximation,

$$X \cong GDW, \tag{3}$$

where $G \in \mathcal{R}^{p \times m}$ is a matrix for dimensionality reduction, $D \in \mathcal{R}^{m \times k}$ is the dictionary of k atoms ($m << p, D$ is called the low-dimensional dictionary), $W \in \mathcal{R}^{k \times n}$ is the coefficient matrix.

B. Algorithm

Our goal is to build an effective discriminative learning algorithm by exploring label information from the dummy coding matrix Y. Generally, an effective learning algorithm requires expressing or transforming the data and constructing a learning formulation by minimizing an empirical cost.

We decompose Equation (3) into two sub-problems,

$$\left\{ \begin{array}{ll} X \cong GT^T & \text{(a)} \\ T^T \cong DW & \text{(b)} \end{array} \right. , \tag{4}$$

where T links the two sub-problems: (a) is used for data representation and dimensionality reduction, and (b) is the sparse dictionary learning approximation. In the following sections, we show how to incorporate class label information into each sub-problem.

To solve the problem in Equation (4)(a), we adopt a type of shared latent structure decomposition, Partial Least Square (PLS) regression [21]. PLS is a projection-based regression method, and the curse of dimensionality is avoided. In addition, it accounts for class labels in the process, in contrast to traditional dimensionality reduction techniques such as Principal Component Analysis (PCA).

The basic idea of PLS is to construct new latent variables as a linear combination of the original variables. The decomposition allows us to develop the following approximation,

$$\left\{ \begin{array}{l} X^T = LG^T + E_1 \\ Y^T = HQ^T + E_2 \end{array} \right. , \tag{5}$$

given a condition that the following approximation also holds,

$$H = LB + E_3, \tag{6}$$

where L and H are score matrices[1], G and Q are loading matrices, B is a weighting matrix, and E_1, E_2 and E_3 are residual matrices.

By maximizing the covariance of X^T and Y^T and enforcing L as an orthogonal matrix, a standard partial least squares (PLS) procedure can be readily employed to compute G and Q, for example, by SIMPLS [21] or the nonlinear iterative partial least squares (NIPALS) algorithm [4]. From G and Q, the PLS procedure further evaluates L and H, which are projections of X^T and Y^T respectively. More details of PLS procedures can be found in the statistics literature [20].

Both L^T and H^T are projections on a canonical latent space. As the next step we will incorporate the latent structures L and H into the dictionary learning framework in Equation (4)(b). By setting $P' = \sqrt{\lambda_c} P$ (P is the dictionary corresponding to Y. Since P is a target parameter, the re-scaling has no effect for classification purposes), we re-write Equation (2) as,

$$\min_{D,P,W} ||X - DW||_F^2 + \lambda_c ||Y - P'W||_F^2 + \lambda_s ||W||_0, \tag{7}$$

One good property of PLS is that the re-scaling of X or Y does not affect either L or H. Therefore, we can simply replace X and $\sqrt{\lambda_c} Y$ by L^T and H^T respectively, and eliminate the classifier weighting parameter λ_c. Accordingly, a latent structure based discriminative dictionary learning method can be formulated as follows,

$$\min_{D,P,W} ||L^T - DW||_F^2 + ||H^T - P'W||_F^2 + \lambda_s ||W||_0, \tag{8}$$

where λ_s is the sparsity parameter.

The problem in Equation (8) is reformulated as follows, and solved approximately by a standard Orthogonal Matching Pursuit algorithm [1].

$$\min_{\bar{D},W} ||\bar{X}_{LS} - \bar{D}W||_F^2 + \lambda_s ||W||_0, \tag{9}$$

where

$$\bar{X}_{LS} = \left(\begin{array}{c} L^T \\ H^T \end{array} \right), \bar{D} = \left(\begin{array}{c} D \\ P' \end{array} \right). \tag{10}$$

C. Properties of LS-DDL

The PLS-based shared latent structure decomposition leads to several important properties for dictionary learning. LS-DDL computes a much more compact dictionary than conventional dictionary learning methods. Assuming s latent vectors in the PLS estimation and k atoms for the dictionary learning,

Property 1: LS-DDL learns a compact dictionary D of size $s \times k$ and its computational complexity is $\mathcal{O}(s^2)$, while

[1]Terminology such as score, loading, weighting, and residual are standard nomenclature from the partial least squares literature [20].

the dictionary learned from conventional dictionary learning or discriminative dictionary learning is of size $p \times k$ and its computational complexity is $\mathcal{O}(p^2)$. Here $s << p$.

Property 2: LS-DDL formulation eliminates the classifier weighting parameter being present in D-KSVD and its variants.

For Property 1, LS-DDL performs dimensionality reduction by retaining only a few informative and discriminative latent vectors while eliminating the noise in the data. Therefore the learned dictionary via shared latent structure is more robust in classification. The main driving factor of the computational complexity of dictionary learning is the dimension of the signal space [12]. The latent structures are of much lower dimensionality, and thus LS-DDL alleviates the computational bottleneck by feeding much smaller data into the subsequent dictionary learning procedures, and greatly improves the efficiency in dictionary storage as well as dictionary learning and object classification. Compared with a computationally demanding dictionary learning procedure such as K-SVD, the time cost of NIPALS or SIMPLS in LS-DDL is negligible. Therefore LS-DDL is much more efficient.

For Property 2, LS-DDL learns a discriminative dictionary from L and H, which are projections of X^T and Y^T respectively onto a canonical latent space. L and H are linearly related by a weighting matrix B as in Equation (6), and a re-scaling of one will automatically yield a commensurate re-scaling of the other. By directly integrating L and H into the discriminative dictionary learning framework, we can eliminate the classifier weighting parameter present in a traditional formulation. Compared with D-KSVD and its variants, this property implies reduced effort in tuning parameters.

D. Implementation Details

One important implementation issue is how to determine the number of latent vectors. This topic has been extensively studied in the statistics literature, especially for partial least squares and its variants [20]. Besides the commonly used cross-validation strategy [20], the percentage of expressed variance is a good measure. We choose expressed variance here. That is, for s latent vectors, we denote $L = [l_1, ..., l_s], G = [g_1, ..., g_s]$, then the Cumulative Percentage of expressed Variance (CPV) for X^T is defined as

$$CPV_X = \frac{\sum_{j=1}^{s} {l_j}^T l_j {g_j}^T g_j}{tr(XX^T)}, \tag{11}$$

where tr denotes the trace. We suggest that for datasets with sharp transitions in their CPV curves, it usually works well to use the number of latent vectors at around the 97% threshold on the CPV curve. This is similar to the empirical threshold used in PCA, which takes the smallest number of components needed to explain 97% of the total variance.

A SIMPLS [4] procedure is applied to estimate G and Q, then we compute the shared latent structure L and H from Equation (5), and then D and P are evaluated from Equation (9). For each class, we also compute a class center h_c from H^T and let $H_c = [h_1, ..., h_M]$. We then construct a new matrix $\hat{H}$ to replace H, by replacing each atom of H with its corresponding class center h_c. This reduces noise in H. We call this strategy centralized LS-DDL. The intuition for this centralized version is based on the observation that columns of H from different classes are almost orthogonal and the intrinsic rank of H is approximately M. This also suggests that a PCA procedure could be applied to reduce noise in H. For a given test example x_t, we project x_t onto the latent structure, compute its sparse code w, and evaluate a projected vector $\nu := Pw$. Finally, a label $\hat{c}$ is predicted by maximizing the correlation between ν and H_c,

$$\hat{c} = \underset{c \in \{1,...,M\}}{\text{argmin}} \ \nu^T h_c. \tag{12}$$

IV. Experiments

In this section, we design a set of experiments to verify the effectiveness and efficiency of the proposed LS-DDL method on two typical applications, image-based face recognition and video-based activity recognition.

We compare our results with the Sparse Representation Classifier (SRC) [22], K-SVD [1], D-KSVD [25], and LC-KSVD [7]. We obtained the original implementations for LC-KSVD directly from the authors [7]. D-KSVD was implemented by eliminating the label consistent term in LC-KSVD; similarly, K-SVD was obtained by omitting the classification cost term in D-KSVD. For K-SVD, D-KSVD and LS-DDL, we evenly selected atoms from each class as an initial dictionary. For all these methods, we adopted the same orthogonal matching pursuit (OMP) package to approximate the sparse representation. Also, the same K-SVD implementation was used for all methods[2]. SRC is implemented as in [22]. For fair comparison, we used OMP to replace the convex programming techniques [22] in SRC. The sparsity parameters (10 and 30) were chosen based on previous work [25], [7]. All implementations were in Matlab except that the core components of K-SVD and OMP were compiled from C routines.

A. Image-based Face Recognition

We used the extended YaleB database, which consists of 32 people and each person has about 64 images varying in illumination, expression, angle, with/without glasses, etc. In total, the database has 2414 cropped images of 192 × 168 pixels. Following previous settings [22], we projected all images onto a 504-dimensional vector using a random matrix of zero-mean normal distribution. A split was generated by uniformly sampling 32 images per person for training and taking the rest for testing. This sampling process was repeated to get 10 splits.

First, we investigated the performance as a function of the number of latent vectors. By fixing a dictionary of 570 atoms for the first split, we evaluated the recognition accuracy curve and the cumulative percentage of expressed variance curve for X. The recognition accuracy curve showed that the best

[2]http://www.cs.technion.ac.il/˜ronrubin/software.html

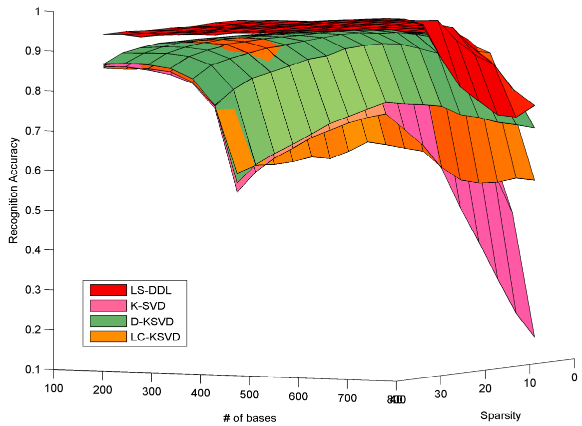

Fig. 2. *Illustration of the recognition accuracy surface w.r.t. dictionary sizes and sparsity for the compared methods.*

results were obtained at around $150 \sim 200$ latent vectors. The CPV_X curve confirmed that the 150 latent vectors contribute to about 97.2% of the variance.

Second, we investigated the performance as a function of dictionary size and sparsity for all methods (except SRC) on the same splits. The number of latent vectors was set to 150 for LS-DDL. The parameters settings for LC-KSVD, D-KSVD were the same as those in [7] and [25], respectively. Figure 2 shows the recognition accuracy surface w.r.t. dictionary size and sparsity for the compared methods. It shows that the LS-DDL recognition accuracy is consistently higher than the other methods for all dictionary sizes and sparsity.

Figure 3 shows the recognition accuracy curves w.r.t. dictionary size with sparsity of 10 and 30. The error bar is calculated as the standard deviation of results from the 10 splits on each side of the curve. Both sparsity parameters are used in [7] and [25]. At sparsity of 10, the recognition performance for LS-DDL peaks at around dictionary size of 570. This dictionary size is also used in [7] and [25]. Above a dictionary size of about 600, the performance degrades for all compared methods due to over-fitting. At sparsity of 30, the recognition performance for LS-DDL has much smaller variation w.r.t dictionary size compared to the other methods. This shows that the dictionary learned by LS-DDL is very compact and discriminative. This is because that LS-DDL is built on a shared latent structure rather than the original data. Figure 4 shows the recognition accuracy curves w.r.t. sparsity at dictionary sizes of 270 and 570 for the compared methods. It shows that LS-DDL is relatively insensitive to the sparsity parameter too. From Figure 3 and Figure 4, it is obvious that with much lower sparsity and much smaller dictionary size, LS-DDL outperforms the compared methods in recognition accuracy.

Table 1 shows the recognition accuracy and computational time at sparsity of 10 and 30, and dictionary size of 570. LS-DDL has much higher recognition accuracy at sparsity of 10, and slightly higher recognition accuracy at sparsity of 30. It

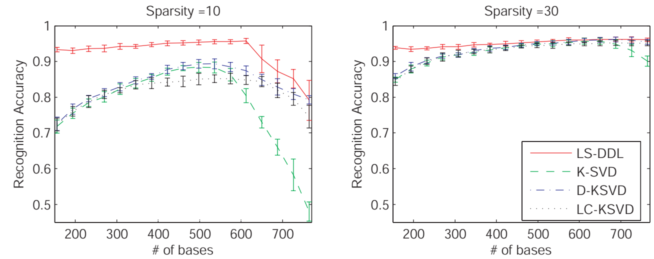

Fig. 3. *The recognition accuracy curves w.r.t. dictionary sizes (# of bases) at sparsity of 10 and 30 for the compared methods.*

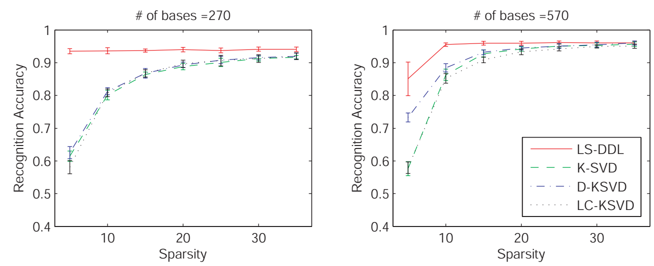

Fig. 4. *The recognition accuracy curves w.r.t. sparsity at dictionary sizes of 270 and 570 for the compared methods.*

is much faster in the training phase than the other compared methods because of its low-dimensional dictionary, and it is comparable with other methods in testing time. As the dataset becomes larger, LS-DDL's advantage in the testing phase will become more significant.

B. *Video-based Activity Recognition*

For video-based activity recognition, we propose a new Bag-of-Discriminative-Words (BoDW) model to represent different classes of activities, which is based on discriminative dictionary learning by LS-DDL and the traditional Bag-of-Words (BoW) model. We denote this as LS-DDL-BoDW. The Bag-of-Words model has been widely used in computer vision for scene classification [9] and activity

Method	Sparsity = 10 Accuracy ($\pm std$)%	Sparsity = 10 $t_{train/test}$ (s)	Sparsity = 30 Accuracy ($\pm std$)%	Sparsity = 30 $t_{train/test}$ (s)
SRC	84.9±0.9	- /5.4	90.9±0.9	- /5.8
K-SVD	86.3±1.8	5.1/0.1	95.3±0.8	9.8/0.5
D-KSVD	88.5±1.3	5.2/0.1	95.5±.7	9.9/0.5
LC-KSVD	85.2±1.5	97.0/0.1	95.0±0.5	117.6/0.5
LS-DDL	**95.6±0.5**	**4.5/0.1**	**96.1±0.6**	**8.9/0.5**

TABLE I

PERFORMANCE OF THE COMPARED METHODS USING A SPARSITY OF 10 AND 30. DICTIONARY SIZE IS FIXED TO 570 FOR ALL METHODS. SRC DOES NOT REQUIRE A TRAINING PHASE SO THE TRAINING TIME COST IS UNAVAILABLE. THE TESTING TIME IS FOR CLASSIFICATION OF ALL TESTING EXAMPLES IN A SPLIT. LS-DDL METHODS OUTPERFORM OTHER COMPARED METHODS IN TERMS OF EFFICIENCY AND ACCURACY. ALL TIME COSTS REPORTED IN THIS PAPER ARE BASED ON A COMPUTER WITH DUAL CORE 2.5G CPU AND 3G MEMORY.

Method	Accuracy (%)	$t_{training/testing}$ (s/split)
K-SVD-BoW	84.6	16.8/6.4
D-KSVD-BoDW	92.7	16.8/6.4
LC-KSVD-BoDW	86.8	70.0/6.4
LS-DDL-BoDW	**93.2**	**15.8/1.8**

TABLE II
PERFORMANCE OF THE COMPARED METHODS ON UCF SPORTS DATASET IN A FIVE-FOLD CROSS-VALIDATION SETUP.

recognition [16]. Instead of generating the "words" using k-means clustering in the BoW model, our BoDW model takes the atoms learned by LS-DDL as "words", such that all training data are represented as histograms of discriminative words from the learned dictionary.

For a comparison with previous work, we also extend the K-SVD, D-KSVD, and LC-KSVD methods by incorporating the BoW model, and the corresponding methods are denoted as K-SVD-BoW, D-KSVD-BoDW and LC-KSVD-BoDW respectively. We start by learning a dictionary, and then treat the dictionary atoms as a bag of visual words. Frequency histograms are built based on Euclidean distances. Finally, a non-linear SVM with a Chi-Squares kernel is employed to classify a video clip. We tune the classifier weighting parameters and the sparsity parameter for D-KSVD and LC-KSVD for the best results. SRC has only a testing phase and thus cannot be extended in a similar way. We also compare our results with other state-of-the-art methods in similar settings. We extract Histogram of Oriented Gradient (HOG) [3] descriptors and Histogram of Optical Flow (HOF) [2] descriptors. Corresponding HOG and HOF vectors are concatenated to form a HOGHOF vector.

We test our proposed method on several publicly available activity recognition datasets, and consistently obtain superior performance than existing dictionary learning based methods. Here we report the results for one representative dataset, the UCF Sports Action dataset [19]. The UCF Sports dataset consists of 10 different actions. In total, it has 150 videos extracted from real sports broadcasts, with a wide range of viewpoints and scene backgrounds. We follow the steps in [18] to extract tracks from ground-truth annotations and compute the HOGHOF descriptors. Using the same features, we apply K-SVD-BoW, D-KSVD-BoDW, LC-KSVD-BoDW and LS-DDL-BoDW (with 20 latent vectors) to learn a dictionary of 110 atoms from an initial dictionary of 7817 atoms. The performance is reported as the accuracy over all classes. We use the same five-fold cross-validation setting for all methods. Detailed comparison results are given in Table 2, which shows that LS-DDL-BoDW achieves an accuracy of 93.2% and it outperforms other compared methods in terms of accuracy and efficiency. The performance boosting is mainly due to two factors: 1) The low-dimensional dictionary after dimensionality reduction makes LS-DDL more efficient and more robust to the noisy and redundant HOGHOF features; and 2) LS-DDL learns a discriminative dictionary in the discriminative space constructed by PLS.

V. CONCLUSION

We have presented a simple yet effective discriminative dictionary learning method using shared latent structures in a general multi-class classification setting. The LS-DDL approach learns a discriminative and low-dimensional dictionary in a shared latent structure space, which shows superior recognition accuracy and efficiency in classification tasks.

REFERENCES

[1] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing over-complete dictionaries for sparse representation. *IEEE T. SP*, 54(11):4311–4322, 2006.

[2] R. Chaudhry, A. Ravichandran, G. Hager, and R. Vidal. Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In *CVPR*, 2009.

[3] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005.

[4] S. de Jong. Simpls: An alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18:251–263, 1993.

[5] H. Guo, Z. Jiang, and L. S. Davis. Discriminative dictionary learning with pairwise constraints. In *ACCV*, 2012.

[6] K. Huang and S. Aviyente. Sparse representation for signal classification. In *NIPS*, 2007.

[7] Z. Jiang, Z. Lin, and L. S. Davis. Learning a discriminative dictionary for sparse coding via label consistent k-svd. In *CVPR*, 2011.

[8] I. Laptev and T. Lindeberg. Space-time interest points. In *ICCV*, 2003.

[9] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR*, 2006.

[10] T. Leung and J. Malik. Representing and recognizing the visual appearance of materials using three-dimensional textons. *IJCV*, 43(1):29–44, 2001.

[11] J. B. MacQueen. Some methods for classification and analysis of multivariate observations. *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–297, 1967.

[12] B. Mailhé, R. Gribonval, F. Bimbot, and P. Vandergheynst. A low complexity orthogonal matching pursuit for sparse signal approximation with shift-invariant dictionaries. In *ICASSP*, 2009.

[13] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Discriminative learned dictionaries for local image analysis. In *CVPR*, 2008.

[14] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman. Supervised dictionary learning. In *NIPS*, 2009.

[15] H. Nguyen, V. Patel, N. Nasrabadi, and R. Chellappa. Sparse embedding: A framework for sparsity promoting dimensionality reduction. In *ECCV'12*, 2012.

[16] J. C. Niebles, H. Wang, and L. Fei-Fei. Unsupervised learning of human action categories using spatial-temporal words. *IJCV*, 79:299–318, 2008.

[17] D. Pham and S. Venkatesh. Joint learning and dictionary construction for pattern recognition. In *CVPR*, 2008.

[18] Q. Qiu, Z. Jiang, and R. Chellappa. Sparse dictionary-based representation and recognition of action attributes. In *ICCV*, 2011.

[19] M. Rodriguez, J. Ahmed, and M. Shah. Action mach a spatio-temporal maximum average correlation height filter for action recognition. In *CVPR*, 2008.

[20] V. E. Vinzi, W. W. Chin, and J. Henseler. *Handbook of partial least squares: concepts, methods and applications*. Springer, 2010.

[21] S. Wold, M. Sjstrm, and L. Eriksson. Pls-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58:109–130, 2001.

[22] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE T. PAMI*, 31(2):210–227, 2009.

[23] L. Yang, R. Jin, R. Sukthankar, and F. Jurie. Unifying discriminative visual codebook genearation with classifier training for object category recognition. In *CVPR*, 2008.

[24] M. Yang, L. Zhang, X. Feng, and D. Zhang. Fisher discrimination dictionary learning for sparse representation. In *ICCV*, 2011.

[25] Q. Zhang and B. Li. Discriminative k-svd for dictionary learning in face recognition. In *CVPR*, 2010.