Discriminative Dictionary Learning with Pairwise Constraints

Huimin Guo*, Zhuolin Jiang*, and Larry S. Davis

University of Maryland, College Park, MD, 20742
{hmguo,zhuolin,lsd}@umiacs.umd.edu

Abstract. In computer vision problems such as pair matching, only binary information - 'same' or 'different' label for pairs of images - is given during training. This is in contrast to classification problems, where the category labels of training images are provided. We propose a unified discriminative dictionary learning approach for both pair matching and multiclass classification tasks. More specifically, we introduce a new discriminative term called 'pairwise sparse code error' for the discriminativeness in sparse representation of pairs of signals, and then combine it with the classification error for discriminativeness in classifier construction to form a unified objective function. The solution to the new objective function is achieved by employing the efficient feature-sign search algorithm. The learned dictionary encourages feature points from a similar pair (or the same class) to have similar sparse codes. We validate the effectiveness of our approach through a series of experiments on face verification and recognition problems.

1 Introduction

Different from many classification problems where the specific class label of each image is given during training, only binary information such as same/different or relevant/irrlevant is provided for training data in applications such as face verification (given a *target* and a *query* image, determine whether they are from the same person), pair matching, image retrieval, etc. Typically, a discriminative similarity measure is learned through metric learning [1–4] from pairs of training images labeled as 'same' or 'different'; this provides less specific information than known classes - category labels. In this paper, we propose a framework to learn a discriminative dictionary satisfying pairwise constraints. The learned dictionary is suitable for pair matching problems with the pairwise constraints from the binary similarity or dissimilarity information; in addition, it is also suitable for classification problems given pairwise constraints about category information.

Sparse coding [5] approximates a signal y as a linear combination of a few atoms from a learned dictionary A, *i.e.*, $y = Ax$, and leads to good performance in numerous applications. The learned dictionary A is critical to performance. K-SVD [6] minimizes a reconstruction error to learn an over-complete dictionary. However, despite its many successful applications, K-SVD is not suitable for classification, where the dictionary should be not only representative, but also

* Indicates equal contributions.

K.M. Lee et al. (Eds.): ACCV 2012, Part I, LNCS 7724, pp. 328–342, 2013.

discriminative. Hence, some supervised dictionary learning approaches incorporate classification error into the objective function to construct a dictionary with discriminative power. However, such frameworks consider only discriminativeness in the classifier construction, but do not guarantee the discriminativeness in the sparse representations of input signals. The discriminative capability of a dictionary usually comes from category label information. We will show that considering the pair similarity/dissimilarity constraints without category labels during dictionary learning can also improve the discriminative power of a dictionary; no existing dictionary learning approach has fully explored this property. Our dictionary learning approach explicitly integrates pairwise constraints for sparse codes of input signals and a linear predictive classifier into one objective function. The learned dictionary encourages signals from the same class (or a similar pair) to have similar sparse codes, and signals from different classes (or a dissimilar pair) to have dissimilar sparse codes, illustrated in Figures 1 and 2. The similarity can be thresholded to yield a binary decision of same/different (face verification), or it can be used to find the most similar face in a gallery (face recognition). The main contributions of this paper are:

- We present a dictionary learning framework with explicit pairwise constraints, which unifies the discriminative dictionary learning for pair matching and classification problems.
- Our framework furthermore integrates the pairwise constraints for sparse codes of input signals and a linear predictive classifier into the objective function for dictionary learning, which addresses the desirable properties of discriminativeness in the sparse representations of signals, and the discriminativeness in classifier construction.
- The objective function can be optimized via the efficient feature-sign search algorithm [7].
- Our approach is validated on various public face verification and recognition benchmarks.

1.1 Related Work

Metric learning (ML) aims at learning a discriminative similarity measure between different images [1–4]. An appropriate distance metric plays a very important role in many learning problems. Most work in metric learning, including LDML [1], MkNN [1], ITML [2], CSML [3], etc, relies on learning a Mahalanobis distance to map the feature space into a target space [4]. Less work, however, has been done for face verification using dictionary learning with pairwise similarity and dissimilarity constraints on input training examples.

[5] used sparse representations for face recognition (1:N matching problem which finds a nearest neighbor of a given *probe* in a *gallery* face set) by relating the problem of finding the most similar face to noiseless signal reconstruction. Since then, many other researchers have developed methods for face recognition using sparse representations or dictionary learning [5, 9–14]. Although many of these existing algorithms have been shown to perform well in classification

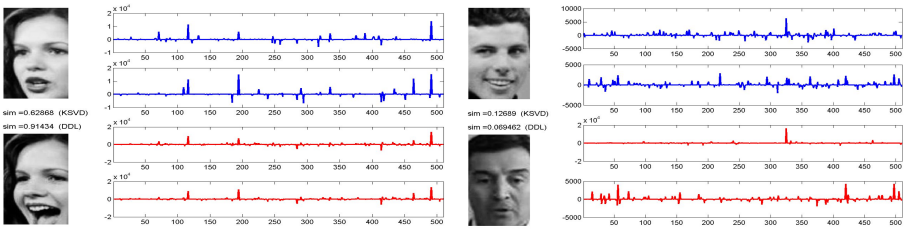


Fig. 1. An example of sparse codes (HoG feature) and similarity scores obtained by K-SVD dictionary learning and our proposed discriminative dictionary learning with pairwise constraints. Image pairs are from test set 1 of the LFW [8] dataset. (a) Original faces of the 'same' pair and their similarity scores obtained by 'K-SVD' and 'DDL'. (b) Sparse codes for the 'same' pair obtained from 'K-SVD'(blue) and 'DDL'(red), respectively. (c) Original faces of a 'different' pair. (d) Sparse codes for the 'different' pair. It can be seen that our dictionary encourages a pair from 'same' person to have similar sparse codes while a pair from 'different' persons to have dissimilar sparse codes.

(e.g. face recognition) applications, most of them do not explicitly deal with dictionary learning with pairwise constraints - when only binary information such as same/different or relevant/irrelevant is given in the training stage (e.g. face verification). Our dictionary learning framework is more general since it deals with face verification and face recognition problems simultaneously.

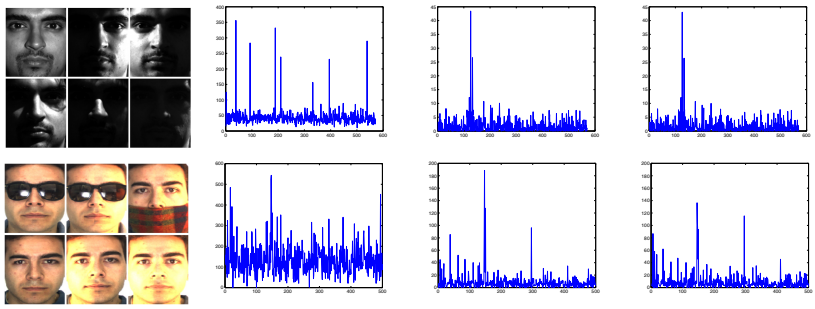
To enhance discrimination power, our dictionary learning framework explicitly integrates pairwise constraints for sparse codes of input signals and a linear predictive classifier into the objective function during training. Most previous approaches treat dictionary learning and classifier training as two separate processes, such as [15–20]. In these approaches, a dictionary is typically learned first and then a classifier is trained based on it. There are also sophisticated approaches [13, 21–23] combining dictionary learning and classifier training in a mixed reconstructive and discriminative formulation. Our approach falls into this category. We learn a single dictionary and an optimal classifier jointly.

Laplacian Sparse Coding [24] explicitly introduces a locality preserving constraint among similar local features in the sparse coding step to preserve the consistence of the sparse codes. This is different since our approach is to learn a dictionary which encourages signals from a similar pair (or the same class) to have similar sparse codes. Furthermore, our approach integrates a linear predictive classifier into the objective function to learn the dictionary and the classifier simultaneously while [24] learns the dictionary and the classifier separately.

2 Sparse Coding and Dictionary Learning

2.1 Sparse Coding

Let $Y = [\boldsymbol{y}_1, \boldsymbol{y}_2, ...\boldsymbol{y}_N] \in \mathbb{R}^{n \times N}$ be the data matrix of N input signals, where $y_i \in \mathbb{R}^n$ denotes the i-th input signal with n-dimensional feature description.



(a) Sample images (b) K-SVD (c) DDL-PC1(ours) (d) DDL-PC2(ours)

Fig. 2. Examples of sparse codes using dictionaries learned by K-SVD and our approaches on the Extended YaleB [25] and AR [26] databases. X axis indicates the dimensions of sparse codes. Y axis indicates the average of absolute sparse codes for different testing images from the same class. The first and second row correspond to class 9 in Extended YaleB (32 images) and class 30 in AR database (6 images), respectively. The consistency of sparse codes of signals from the same class should have low entropy (*i.e.*, less high values) of these average sparse codes.

Given a dictionary $A = [\boldsymbol{a}_1, \boldsymbol{a}_2, ..., \boldsymbol{a}_K] \in \mathbb{R}^{n \times K}$, where $\boldsymbol{a}_i$ is the i-th dictionary atom (l_2-normalized), sparse coding [5] with l_1 regularization computes the sparse representations $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N] \in \mathbb{R}^{K \times N}$ of the input signals Y, through solving the following l_1-minimization problem,

$$X^* = \arg\min_X \sum_{i=1}^{N} (\|\boldsymbol{y}_i - A\boldsymbol{x}_i\|_2^2 + \gamma \|\boldsymbol{x}_i\|_1) \tag{1}$$

where constant γ is a sparsity constraint factor and the term $\|\boldsymbol{y}_i - A\boldsymbol{x}_i\|_2^2$ denotes the reconstruction error. Each input signal $\boldsymbol{y}_i$ can be represented as a sparse linear combination of a few dictionary atoms. The feature-sign search algorithm [7] is an efficient algorithm that can be used to solve (1).

2.2 Dictionary Learning

The goal of dictionary learning is to find optimized dictionaries that provides a succinct representation for most statistically representative input signals. The learning procedure can be formulated as solving the following problem [7],

$$< A^*, X^* >= \arg\min_{A,X} \sum_{i=1}^{N} (\|\boldsymbol{y}_i - A\boldsymbol{x}_i\|_2^2 + \gamma \|\boldsymbol{x}_i\|_1) \tag{2}$$

The optimization problem is convex in A (while holding X fixed) and convex in X (while holding A fixed), but not convex in both simultaneously. Usually, the above objective is iteratively optimized in a two stage manner, by alternatively optimizing with respect to A (bases) and X (coefficients) while holding the other fixed. The formulation (2) only focuses on minimizing the reconstruction error

and does not consider the discriminative power of a dictionary for classification tasks. Hence, some supervised approaches [12, 13, 21–23] have been proposed to improve the discriminative power of dictionary, by integrating the category label information into the objective function of dictionary learning. However, most of them do not explicitly deal with dictionary learning with pairwise constraints.

3 Discriminative Dictionary Learning with Pairwise Constraints (DDL-PC)

In this section, we present our Discriminative Dictionary Learning with Pairwise Constraints algorithm which takes into account the relationships of each pair of learned sparse codes $(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Here, the intuition is to encourage signals from a similar pair to have similar sparse codes. We subsequently focus on the effects of adding a discriminative term, and a classification error term into the objective function in (2). We refer to them as DDL-PC1 and DDL-PC2, respectively.

3.1 DDL-PC1

To obtain discriminative sparse codes $\boldsymbol{x}$ with the pairwise constrained dictionary A, the objective function for dictionary construction is defined as:

$$\begin{aligned} < A^*, X^* > &= \arg\min_{A,X} \sum_{i=1}^{N} (\|\boldsymbol{y}_i - A\boldsymbol{x}_i\|_2^2 + \gamma\|\boldsymbol{x}_i\|_1) + \frac{\beta}{2} \sum_{i,j=1}^{N} (\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 M_{ij}) \\ &= \arg\min_{A,X} \sum_{i=1}^{N} (\|\boldsymbol{y}_i - A\boldsymbol{x}_i\|_2^2 + \gamma\|\boldsymbol{x}_i\|_1) + \beta(Tr(X^T X D) - Tr(X^T X M)) \\ &= \arg\min_{A,X} \sum_{i=1}^{N} (\|\boldsymbol{y}_i - A\boldsymbol{x}_i\|_2^2 + \gamma\|\boldsymbol{x}_i\|_1) + \beta(Tr(X^T X L)) \end{aligned} \tag{3}$$

where the constants γ and β control the relative contribution of the corresponding terms. The first term $\|\boldsymbol{y}_i - A\boldsymbol{x}_i\|_2^2$ is the *reconstruction error term*, which evaluates the reconstruction error of the approximation to the input signals. The second term $\|\boldsymbol{x}_i\|_1$ is the *regularization term* for sparsity. The last term, which is new and proposed here, is the *discrimination term* called 'pairwise sparse code error' based on pairwise constraints which are encoded in matrix M. $D = diag\{d_1, ...d_N\}$ is a diagonal matrix whose diagonal elements are the sums of the row elements of M (see below), $d_i = \sum_{j=1}^{N} M_{ij}$. $L = D - M$ is the Laplacian matrix. Matrix M has different forms depending on the problems being considered. For example, in face verification, the relationship of a pair $(\boldsymbol{y}_i, \boldsymbol{y}_j)$ is given as same/different. Thus, given the sets of 'same' and 'different' pairs $\mathcal{S}$ and $\mathcal{D}$, we define matrix M to encode the (dis)similarity information as

$$M_{ij} = \begin{cases} +1, & \text{if } (\boldsymbol{y}_i, \boldsymbol{y}_j) \in \mathcal{S} \\ -1, & \text{if } (\boldsymbol{y}_i, \boldsymbol{y}_j) \in \mathcal{D} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

3.2 DDL-PC2

Although (3) can already be used for classification by defining M based on the pairwise similarity constraints with category labels (see Sec. 3.4), the classification error can be further included as an additional term in the objective function in (3). Here we use a linear predictive classifier $f(\boldsymbol{x}; W) = W\boldsymbol{x}$. The objective function for learning a pairwise constrained dictionary A with both reconstructive and discriminative power can then be defined as follows:

$$< A^*, X^*, W^* >= \arg\min_{A,X,W} \sum_{i=1}^{N} (\|\boldsymbol{y}_i - A\boldsymbol{x}_i\|_2^2 + \gamma\|\boldsymbol{x}_i\|_1)$$
$$+\frac{\beta}{2} \sum_{i,j=1}^{N} (\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 M_{ij}) + \alpha \sum_{i=1}^{N} (\|\boldsymbol{h}_i - W\boldsymbol{x}_i\|_2^2 + \lambda\|W\|_2^2) \quad (5)$$

The new term $\|\boldsymbol{h}_i - W\boldsymbol{x}_i\|_2^2 + \lambda\|W\|_2^2$, where $\|\boldsymbol{h}_i - W\boldsymbol{x}_i\|_2^2$ represents the classification error and $\|W\|_2^2$ is the regularization penalty term, supports learning an optimal linear predictive classifier. $\boldsymbol{h}_i = [0, 0, ...1...0, 0]^T \in \mathbb{R}^m$ (m: number of classes) is a label vector corresponding to an input signal $\boldsymbol{y}_i$, where the non-zero position indicates the class label of $\boldsymbol{y}_i$.

3.3 Optimization Procedure

In this section, we only describe the optimization procedure for DDL-PC2 since DDL-PC1 utilizes the same procedure except that $\alpha = 0$ in (6)(7)(8) and the classifier W update step is not considered during dictionary learning. Solving (5) is a challenging task because the objective function is not convex for A, X and W simultaneously; but fortunately, it is convex in one variable when the other two variables are fixed. In [7], (2) was solved by an efficient feature-sign search algorithm. Motivated by [7], we optimize A, X and W alternatively. Algorithm 1 presents the pseudocode of algorithm DDL-PC2.

Computing Sparse Codes X with Fixed A and W. When A and W are fixed, we optimize $\boldsymbol{x}_i$ alternately and fix other $\boldsymbol{x}_j (j \neq i)$ for other signals. Optimizing (5) is equivalent to:

$$\min_{\boldsymbol{x}_i} \mathcal{L}(\boldsymbol{x}_i) + \gamma\|\boldsymbol{x}_i\|_1 \quad (6)$$

where $\mathcal{L}(\boldsymbol{x}_i) = \|\boldsymbol{y}_i - A\boldsymbol{x}_i\|_2^2 + \beta(2\boldsymbol{x}_i^T(XL_i) - \boldsymbol{x}_i^T\boldsymbol{x}_i L_{ii}) + \alpha(\boldsymbol{x}_i^T W^T W\boldsymbol{x}_i - 2\boldsymbol{x}_i^T W^T\boldsymbol{h}_i)$, L_i is the i^{th} column of L and L_{ii} is the (i, i) element of L. (6) is exactly the problem that the feature-sign search algorithm in [7] solves. [7] iteratively searches for the coefficient sign vector $\boldsymbol{\theta}$ for $\boldsymbol{x}_i$, then (6) reduces to a standard, unconstrained quadratic optimization problem (QP). To compute the analytical solution, we calculate the gradient of $\mathcal{L}(\boldsymbol{x}_i)$ with respect to $\boldsymbol{x}_i$:

$$\frac{\partial \mathcal{L}(\boldsymbol{x}_i)}{\partial \boldsymbol{x}_i} = 2A^T(A\boldsymbol{x}_i - \boldsymbol{y}_i) + 2\beta(XL_i) + 2\alpha(W^T W\boldsymbol{x}_i - W^T\boldsymbol{h}_i) + \gamma\boldsymbol{\theta} \quad (7)$$

Finally the analytic solution of $\boldsymbol{x}_i$ can be obtained when we have $\frac{\partial \mathcal{L}(\boldsymbol{x}_i)}{\partial \boldsymbol{x}_i} = 0$:

$$\boldsymbol{x}_i^* = (A^T A + 2\beta L_{ii} I + 2\alpha W^T W)^{-1}(A^T \boldsymbol{y}_i + 2\alpha W^T \boldsymbol{h}_i - 2\beta \sum_{k \neq i} \boldsymbol{x}_k L_{ki} - \gamma \boldsymbol{\theta}) \quad (8)$$

In practice, a very small β is chosen to guarantee the Hessian matrix $(A^T A + 2\beta L_{ii} I)$ to be positive semidefinite, hence (3) is convex.

Updating Dictionary A with Fixed X and W. Given X and W, we use the Lagrange dual in [7] to optimize the following objective function:

$$\min_A \sum_{i=1}^{N} \|\boldsymbol{y}_i - A\boldsymbol{x}_i\|_2^2 \quad s.t. \|\boldsymbol{a}_j\|_2^2 \leq c, \quad \forall j = 1...K. \quad (9)$$

The analytical solution of A can be computed as: $A^* = YX^T(XX^T + \Lambda)^{-1}$, where Λ is a diagonal matrix constructed from all the dual variables.

Updating Classifier W with Fixed X and A. Given X and A, we employ the multivariate ridge regression model [22] to update W, with the quadratic loss and l_2 norm regularization:

$$\min_W \sum_{i=1}^{N} \|\boldsymbol{h}_i - W\boldsymbol{x}_i\|_2^2 + \lambda \|W\|_2^2, \quad (10)$$

which yields the following solution: $W^* = HX^T(XX^t + \lambda I)^{-1}$.

Algorithm 1. Discriminative Dictionary Learning with Pairwise Constraints-2 (DDL-PC2)

Input: input signals Y, Laplacian matrix L, label matrix H, regularization constant γ, β and α, iteration number $\hat{T}$
Output: learned dictionary A, classifier W and sparse code X.
Initialization: Compute initial A_0 via K-SVD, initial X_0, W_0 using (1), (10)
for $t = 1, 2, ..., \hat{T}$ **do**
  Sparse Coding: compute sparse code X using (6);
  Dictionary Update: update dictionary A using (9);
  Classifier Update: update classifier W using (10).
end for

3.4 Matching Approach

Face Verification. In face verification or pair matching problems, a similarity measure is typically learned from pairs of training images labeled as 'same' or 'different'; this provides less specific information than known identities - image labels. Given a training set of pairs, we first construct matrix M with their pairwise relationships. For example, suppose three pairs of feature vectors are given - $(\boldsymbol{y}_1, \boldsymbol{y}_2)$ are features vectors from the same person, $(\boldsymbol{y}_3, \boldsymbol{y}_4)$ are also

features vectors from the same person and $(\boldsymbol{y}_5, \boldsymbol{y}_6)$ are features vectors from different persons. Matrix M would then be:

$$M = \begin{bmatrix} & y1 & y2 & y3 & y4 & y5 & y6 \\ y1 & 0 & 1 & 0 & 0 & 0 & 0 \\ y2 & 1 & 0 & 0 & 0 & 0 & 0 \\ y3 & 0 & 0 & 0 & 1 & 0 & 0 \\ y4 & 0 & 0 & 1 & 0 & 0 & 0 \\ y5 & 0 & 0 & 0 & 0 & 0 & -1 \\ y6 & 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix}$$

With the given training set of pairs and the corresponding matrix M, an optimized discriminative dictionary A (initialized by K-SVD algorithm [6]) can be learned using DDL-PC1. Then, when a new test pair $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ comes in, we can compute the optimized sparse codes $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ with dictionary A by solving (1). Finally, the cosine similarity [3, 27] of the two sparse codes is used as the similarity metric between the image pair. This similarity is thresholded to yield a binary decision of same/different.

Face Recognition. In face recognition, class labels are given for each image in the training set. The pair relationships are derived from the category labels. If $\boldsymbol{y}_i$ and $\boldsymbol{y}_j$ belong to the same class, we define M_{ij} as 1; otherwise we set it to 0. Matrix M encoding the (dis)similarity information can be defined as

$$M_{ij} = \begin{cases} 1, \text{ if } (\boldsymbol{y}_i, \boldsymbol{y}_j) \in c_k, k = 1...m \\ 0, \text{ otherwise} \end{cases} \tag{11}$$

There are two ways to construct the classifier W here. For DDL-PC1, we obtain A and X first and then the matrix W is trained separately using (10). For DDL-PC2, we obtain A and W jointly using Algorithm 1.

Then, when a new test sample $\boldsymbol{y}_i$ comes in, we compute its sparse code $\boldsymbol{x}_i$ with respect to A by solving (1). Finally we simply use W to estimate a class label vector for $\boldsymbol{y}_i$: $l = W\boldsymbol{x}_i$, where $l \in \mathbb{R}^m$. The label of $\boldsymbol{y}_i$ is assigned as the index j where l_j is the largest element of l.

4 Experimental Results

We evaluate the proposed algorithm on the LFW dataset [8] for face verification task, and the Extended YaleB database [25] and AR face database [26] for face recognition task.

4.1 Face Verification

LFW Database. The Labeled Faces in the Wild (LFW) dataset was recently introduced as a challenging benchmark for face verification in unconstrained environments. Real-world images in the LFW dataset exhibit visual variations

caused by pose, facial appearance, age, lighting, expression, occlusion, scale, camera, misalignment hairstyle, etc.

The dataset comes with a division of 10 splits/folds (disjoint subject identities) for cross validation with three evaluation protocols: unsupervised, image-restricted, and image-unrestricted protocols [8]. We only consider the most common protocol called 'image-restricted': in this setting, it is known whether an image pair belongs to the same person or not, but identity information of the images is not provided. The aligned version lfw-a is used in all experiments.

In our evaluations, for each independent fold, we randomly choose 500 pairs of 'same' and 500 pairs of 'different' from the training set (other 9 splits, 5400 image pairs) to learn an optimal dictionary through DDL-PC1. The learned dictionary consists of 510 atoms. γ is set to be 30 and β is set to be 0.1.

Experimental Setup. All the faces are cropped and rescaled to 80×148. According to [28–31], combining multiple similarities from different descriptors usually boosts performance. In our experiments, the intensity, HoG, LBP, and Gabor features are used. Finally, the four scores for different features are fused by averaging (no training) or training SVM. For extracting HoG and LBP features, we divide the faces into blocks of 20×20 and extract the 16-bin HoG feature and the 59-bin uniform LBP feature for each block. For Gabor features, we adopt five scales and eight orientations of the Gabor filters. The final Gabor feature vector is obtained by concatenating the responses at every 10 pixels in order to reduce the dimensionality of the feature vector to manageable size.

Fig.3 shows some examples (5 'same' and 5 'different') of testing image pairs from the LFW dataset. The similarity scores obtained from KSVD dictionary learning and our DDL-PC1 are listed under each pair. As it shows, compared to KSVD, higher similarity scores for the 'same pairs' and lower similarity scores for 'different' pairs are obtained by our discriminative dictionary learning.

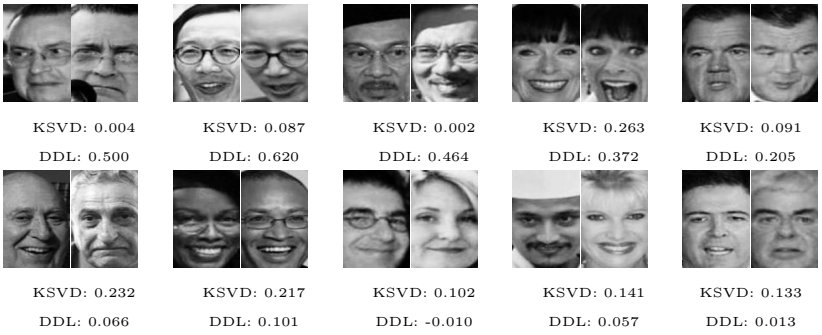


Fig. 3. Examples of some image pairs from the LFW dataset and the similarity scores obtained from KSVD dictionary learning and proposed DDL-PC1 respectively. Top row: Five examples of '**same**' pairs; Bottom row: Five examples of '**different**' pairs.

Table 1 summarizes the performances of our method with individual feature and their fusion. The first column shows the face verification accuracy (at equal error rate) obtained from using the Euclidean distance of the original feature vector pairs as similarity measure. The second column shows the accuracy from the dictionary learned by K-SVD (followed by the l_1 based sparse coding) and the third column shows those from the proposed DDL-PC1. The combined scores are the results from fusing the four scores for all features by averaging (no training) or training SVM. Clearly, DDL-PC1 works best in all situations comparing to 'Euclidean' and 'K-SVD'.

Table 1. Mean (± standard error) verification accuracy at equal error rate of different feature descriptors and their fused scores on LFW dataset. Euclidean, dictionaries learned by K-SVD and the proposed DDL-PC1 are compared.

Descriptor	Euclidean	K-SVD	DDL-PC1
Intensity	0.7140±0.0056	0.7424±0.0051	0.7870±0.0048
HoG	0.6803±0.0046	0.7524±0.0049	0.8030±0.0037
LBP	0.6763±0.0054	0.7433±0.0052	0.7876±0.0032
Gabor	0.6920±0.0041	0.7646±0.0047	0.7996±0.0052
Combined (Avg)	0.7013±0.0045	0.8056±0.0045	0.8410±0.0041
Combined (SVM)	0.7216±0.0047	0.8196±0.0036	0.8603±0.0033

Comparison with the State-of-the-art Methods. Table 2 shows the face verification accuracy of our method compared with recent methods with the Image-Restricted protocol. The 'flip' means that when comparing image pair I and J, we also compare I and the horizontally flipped image of J to reduce the effects of pose variation. Then, the average of the two scores is taken as the final similarity score. Figure 4 contains the ROC curve of our approach (dotted red line), along with the ROC curves of selected recent state-of-the-art methods with the Image-Restricted protocol for presentation clarity.

The results show that the verification accuracy of our approach is comparable with the state-of-the-art methods on the LFW benchmark in the challenging image-restricted protocol. Moreover, the methods marked by '$*$' use training data outside of LFW for facial point detection or pose/illumination classification and so on. Those can have a significant impact on verification accuracy, thus not directly comparable. Kumar [32] achieved excellent results, marginally lower than ours. However, the work of Kumar requires expensive training of high-level classifiers incorporating a huge volume of images outside of the LFW dataset. The LE method [30] relies on facial feature point detectors. Predict-Associate [33] not only relies on facial feature point detectors, but also uses the Multi-PIE dataset with identities covering 7 poses and 4 illumination conditions as prior knowledge. For other methods we are in the same category with, [31] is most comparable. Wolf [31] also combines multiple descriptors; their method adds up several layers of information and leverages metric learning [33]. Moreover, one disadvantage of Wolf's method is that it requires background samples (a fixed set

Table 2. Mean (± standard error) verification accuracy on the LFW dataset, image-restricted protocol using the proposed DDL-PC1, and the same model except the addition of the 'flipped' image idea. '∗' denotes methods using outside training data.

Method	Accuracy
LDML [1]	0.7927±0.0060
Hybrid [29]	0.8398±0.0035
Combined b/g samples based [31]	0.8683±0.0034
*Attribute and Simile classifiers [32]	0.8529±0.0123
Single LE + holistic [30]	0.8122±0.0053
*Multiple LE + comp [30]	0.8445±0.0046
*Predict-Associate [33]	0.9057 ±0.0056
LARK + OSS [34]	0.8512 ±0.0037
DDL-PC1	**0.8603 ±0.0033**
DDL-PC1 (flip)	**0.8710 ±0.0035**

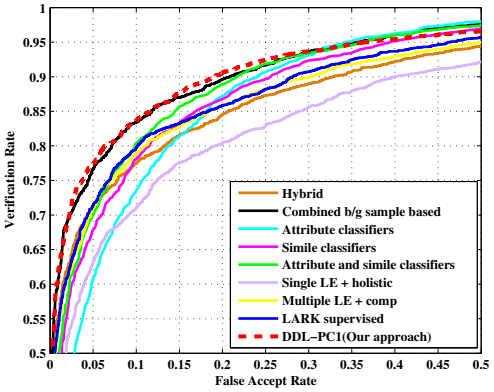


Fig. 4. ROC curves for View 2 of the LFW dataset (Image-Restricted protocol). Only shown with the selected **best** results that recently reported for clarity.

of 'negative' examples) that have similar properties as the faces being compared and do not contain faces from any person who might subsequently appear in a pair to be compared. It learns models for each pair being compared on-the-fly, which might not be desirable in practical applications. Overall, our DDL-PC1 achieves competitive accuracy without local feature identification or any other additional information.

4.2 Face Recognition

Extended YaleB Database. The Extended YaleB database [25] contains 38 persons under 64 illumination conditions, 2, 414 frontal-face images. The original images are cropped to 192×168. We used the random face features [5, 13] to represent the face images. Following [12, 13], we project each face image into a 504-dimensional feature vector using a random matrix of zero-mean normal distribution. Each row of the random matrix is l_2 normalized. We randomly sample 32 images per person for training and taking the rest as testing. We

repeated 10 times such this sampling process and report their average as the recognition accuracy. The parameter γ is set to 20; β and α are set to 2.0 and λ is 1.0 here.

We fix the dictionary size of 570 atoms as in [12, 13] and evaluate our approach. We compare the recognition accuracy with K-SVD [6], D-KSVD [13], SRC [5], LLC [35] and recently proposed LC-KSVD [12]. We obtain the original implementations of LC-KSVD [1] from the authors [12]. A D-KSVD is implemented by eliminating the label consistent term in LC-KSVD. For SRC, we randomly select the average of dictionary size per person from each person and report the best result we achieved. For LLC, we perform the experiment with 30 local bases, which determines the sparsity of the LLC codes. The results are summarized in Table 3. Our approaches achieve better results than K-SVD, D-KSVD, SRC and LLC and are comparable to LC-KSVD.

We also evaluate our approach using random-face features and dictionary sizes 190, 380, 570 and 760. Then we compare the classification accuracy with state-of-art approaches including LC-KSVD, D-KSVD, K-SVD, SRC and LLC which use the same features and dictionary sizes. As shown in Figure 5, our approach has higher accuracy than K-SVD, D-KSVD, SRC and LLC, and is comparable to LC-KSVD.

Table 3. Recognition results using random-face features on the Extended YaleB.

Method	K-SVD[6]	D-KSVD[13]	SRC[5]	LLC[35]	LC-KSVD[12]	**DDL-PC1**	**DDL-PC2**
Acc. (%)	90.5	94.1	88.6	82.3	95.0	**94.5**	**95.3**

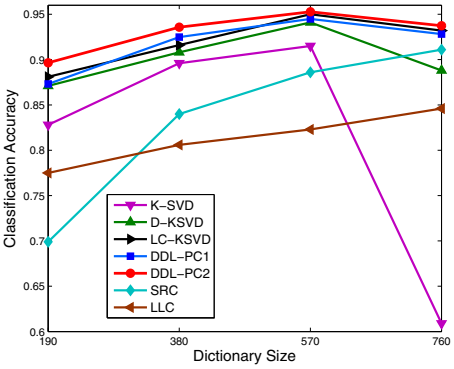


Fig. 5. Recognition performance on the Extended YaleB with varying number of dictionary sizes

AR Face Database. The AR face database [26] contains over 4,000 color face images of 126 persons taken during two sessions, with 26 images per person. The main characteristic of the AR database is that it includes frontal views of faces

[1] LC-KSVD here is the approach LC-KSVD2 in [12].

with different facial expressions, lighting conditions and occlusion conditions. All the faces are cropped to 165×120. Following the standard evaluation protocol, we use a subset of the database consisting of $2,600$ images from 50 males and 50 females. For each person, we randomly select 20 images for training and the other six for testing. We report the results from the average of ten such random splits. Each face image is projected into the 540-dimensional feature vector with a randomly generated matrix as in [12, 13]. The feature descriptors used here are random face features. The parameter γ is set to be 30, β is 0.5, α and λ are 1.0.

We evaluate our approach with a dictionary of size 500 and compare with state-of-art approaches [5, 6, 12, 13, 35]. As shown in Table 4, both DDL-PC1 and DDL-PC2 obtain better results than K-SVD, D-KSVD, SRC, LLC and LC-KSVD. DDL-PC2 obtains a 2% improvement over DDL-PC1.

Table 4. Recognition results using random face features on the AR face database

Method	K-SVD[6]	D-KSVD[13]	SRC[5]	LLC[35]	LC-KSVD[12]	**DDL-PC1**	**DDL-PC2**
Acc. (%)	87.2	88.8	74.5	88.7	93.7	**94.0**	**96.0**

5 Conclusions

We presented a novel dictionary learning approach that tackles the pair matching and classification problem in a unified framework. We introduced a discriminative term called 'pairwise sparse code error' based on pairwise constraints and combined it with the classification error term to form the objective function of dictionary learning for better discriminating power. The objective function can be optimized by employing the efficient feature-sign search algorithm. The effectiveness of our approach was evaluated on both face verification and face recognition tasks. Experimental results on face verification demonstrated that our approach is competitive with existing techniques without using facial feature point detectors or other additional information. We also compared our approach with several recently proposed dictionary learning methods on two well-known face databases. Our approach can obtain comparable face recognition performance to state-of-art on both databases.

Acknowledgement. This work was supported by the Army Research Office MURI Grant W911NF-09-1-0383.

References

1. Guillaumin, M., Verbeek, J., Schmid, C.: Is that you? metric learning approaches for face identification. In: ICCV (2009)
2. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: ICML (2007)

3. Nguyen, H.V., Bai, L.: Cosine Similarity Metric Learning for Face Verification. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV 2010, Part II. LNCS, vol. 6493, pp. 709–720. Springer, Heidelberg (2011)
4. Nowak, E., Jurie, F.: Learning visual similarity measures for comparing never seen objects. In: CVPR (2007)
5. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. 31 (2009)
6. Aharon, M., Elad, M., Bruckstein, A.: K-svd: An algorithm for designing overcomplete dictionries for sparse representation. IEEE Trans. on Signal Processing 54, 4311–4322 (2006)
7. Lee, H., Battle, A., Raina, R., Ng, A.Y.: Efficient sparse coding algorithms. In: NIPS (2007)
8. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, University of Massachusetts, Amherst (2007)
9. Nagesh, P., Li, B.: A compressive sensing approach for expression-invariant face recognition. In: CVPR (2009)
10. Yang, M., Zhang, L.: Gabor Feature Based Sparse Representation for Face Recognition with Gabor Occlusion Dictionary. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 448–461. Springer, Heidelberg (2010)
11. Gao, S., Tsang, I.W.-H., Chia, L.-T.: Kernel Sparse Representation for Image Classification and Face Recognition. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 1–14. Springer, Heidelberg (2010)
12. Jiang, Z., Lin, Z., Davis, L.: Learning a discriminative dictionary for sparse coding via label consistent k-svd. In: CVPR (2011)
13. Zhang, Q., Li, B.: Discriminative k-svd for dictionary learning in face recognition. In: CVPR (2010)
14. Yang, M., 0006, L.Z., Feng, X., Zhang, D.: Fisher discrimination dictionary learning for sparse representation. In: ICCV (2011)
15. Huang, K., Aviyente, S.: Sparse representation for signal classification. In: NIPS (2007)
16. Boureau, Y., Bach, F., LeCun, Y., Ponce, J.: Learning mid-level features for recognition. In: CVPR (2010)
17. Grosse, R., Raina, R., Kwong, H., Ng, A.Y.: Shift-invariant sparse coding for audio classification. In: Conf. on Uncertainty in AI (2007)
18. Mairal, J., Leordeanu, M., Bach, F., Hebert, M., Ponce, J.: Discriminative Sparse Image Models for Class-Specific Edge Detection and Image Interpretation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 43–56. Springer, Heidelberg (2008)
19. Zhang, W., Surve, A., Fern, X., Dietterich, T.: Learning non-redundant codebooks for classifying complex objects. In: ICML (2009)
20. Rodriguez, F., Sapiro, G.: Sparse representations for image classification: Learning discriminative and reconstructive non-parametric dictionaries (2007); IMA Preprint 2213
21. Yang, J., Yu, K., Huang, T.: Supervised translation-invariant sparse coding. In: CVPR (2010)
22. Pham, D., Venkatesh, S.: Joint learning and dictionary construction for pattern recognition. In: CVPR (2008)
23. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Supervised dictionary learning. In: NIPS (2009)

24. Gao, S., Tsang, I.W., Chia, L.T., Zhao, P.: Local features are not lonely - laplacian sparse coding for image classification. In: CVPR (2010)
25. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intelligence 23, 643–660 (2001)
26. Martinez, A.M., Benavente, R.: The AR Face Database. Technical report (1998)
27. Yan, S., Wang, H., Tang, X., Huang, T.: Exploring feature descritors for face recognition. In: ICASSP (2007)
28. Wolf, L., Hassner, T., Taigman, Y.: Descriptor based methods in the wild. In: ECCV (2008)
29. Taigman, Y., Wolf, L., Hassner, T.: Multiple one-shots for utilizing class label information. In: BMVC (2009)
30. Cao, Z., Yin, Q., Tang, X., Sun, J.: Face recognition with learning-based descriptor. In: CVPR (2010)
31. Wolf, L., Hassner, T., Taigman, Y.: Similarity Scores Based on Background Samples. In: Zha, H., Taniguchi, R.-i., Maybank, S. (eds.) ACCV 2009, Part II. LNCS, vol. 5995, pp. 88–97. Springer, Heidelberg (2010)
32. Kumar, N., Berg, A.C., Belhumeur, P.N., Nayar, S.K.: Attribute and Simile Classifiers for Face Verification. In: ICCV (2009)
33. Yin, Q., Tang, X., Sun, J.: An associate-predict model for face recognition. In: CVPR (2011)
34. Seo, H.J., Milanfar, P.: Face verification using the lark representation. IEEE Transactions on Information Forensics and Security 6, 1275–1286 (2011)
35. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: CVPR (2010)