Uncorrelated Multi-View Discrimination Dictionary Learning for Recognition

Xiao-Yuan Jing[1,3], Rui-Min Hu[2,*], Fei Wu[3], Xi-Lin Chen[4], Qian Liu[3], Yong-Fang Yao[3]

[1]State Key Laboratory of Software Engineering, School of Computer, Wuhan University, China

[2]National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, China

[3]College of Automation, Nanjing University of Posts and Telecommunications, China

[4]Institute of Computing Technology, Chinese Academy of Sciences, China

*Corresponding author: hrm1964@163.com

Abstract

Dictionary learning (DL) has now become an important feature learning technique that owns state-of-the-art recognition performance. Due to sparse characteristic of data in real-world applications, DL uses a set of learned dictionary bases to represent the linear decomposition of a data point. Fisher discrimination DL (FDDL) is a representative supervised DL method, which constructs a structured dictionary whose atoms correspond to the class labels. Recent years have witnessed a growing interest in multi-view (more than two views) feature learning techniques. Although some multi-view (or multi-modal) DL methods have been presented, there still exists much room for improvement. How to enhance the total discriminability of dictionaries and reduce their redundancy is a crucial research topic. To boost the performance of multi-view DL technique, we propose an uncorrelated multi-view discrimination DL (UMD^{2}L) approach for recognition. By making dictionary atoms correspond to the class labels such that the obtained reconstruction error is discriminative, UMD^{2}L aims to jointly learn multiple dictionaries with totally favorable discriminative power. Furthermore, we design the uncorrelated constraint for multi-view DL, so as to reduce the redundancy among dictionaries learned from different views. Experiments on several public datasets demonstrate the effectiveness of the proposed approach.

Introduction

Sparse representation based classification has led to interesting object recognition results (Wright et al., 2009), while the dictionary used for sparse coding plays a key role in it (Yang et al., 2011). Due to sparse characteristic of data in real-world applications, dictionary learning (DL) uses a set of learned dictionary bases to represent the linear decomposition of a data point. Dictionary learning has now become an important feature learning technique that owns state-of-the-art recognition performance.

Most DL methods have been addressed to solve single or two views based DL problem (Lin, Liu, and Zha 2012; Feng et al., 2013; Mailhe et al., 2012; Li, Li, and Fu 2013). Low-rank DL for sparse representation method (Ma et al., 2012) stacks data from the same pattern as column vectors of a dictionary to learn class-specific sub-dictionaries for face recognition. Incoherent DL method (Barchiesi and Plumbley 2013) learns dictionaries that exhibit a low mutual coherence while providing a sparse approximation with favorable signal-to-noise ratio. Semi-coupled dictionary learning (SCDL) (Wang et al., 2012), which learns a pair of dictionaries and a mapping function simultaneously, was presented to solve cross-style image synthesis problems. Fisher discrimination DL (FDDL) is a representative supervised DL method (Yang et al., 2011), which is based on the commonly-used fisher discriminant criterion and constructs a structured dictionary whose atoms correspond to the class labels. From the perspective of multi-view (more than two views) feature learning, above DL methods are not designed for multi-view data.

Multi-view feature learning (Han et al., 2012; Guo 2013) has attracted a lot of research interests, because there exists more useful information for recognition in multiple views than that in a single view. In this field, multi-view subspace learning is an important research direction. Under this direction, canonical correlation analysis (CCA) based and discriminant analysis based multi-view subspace learning are two representative techniques. Multi-view CCA (MCCA, Li et al., 2009) was presented to analyze linear relationships between multiple sets of variables. Multiple discriminant CCA (MDCCA, Gao et al., 2012) investigates the supervised correlation across different views to make

full use of available class information. By iteratively learning multiple subspaces as well as a global discriminative subspace, multiple principal angle (MPA, Su et al., 2012) jointly computes both local and global canonical correlations. A generalized multi-view linear discriminant analysis (GMLDA) is addressed (Sharma et al., 2012). Multi-view discriminant analysis (MvDA, Kan et al., 2012) can maximize the between-class variations and minimize the within-class variations of samples in the learning common space from both intra-view and inter-view.

Recently, some multi-view or multi-modal dictionary learning methods have been presented, such as multi-view DL methods (Zheng et al., 2011; Memisevic 2012) and multimodal DL methods (Monaci et al., 2007; Irie et al., 2013; Cao et al., 2013; Wu et al., 2014). Sparse multimodal biometrics recognition (SMBR) method (Shekhar et al., 2014) uses original training sample as dictionary and exploits the joint sparsity of coding coefficients from different biometric modalities to make a joint decision. Literature (Tosic and Frossard 2011) provides a multi-view DL method to learn overcomplete dictionaries for representing stereo images. SliM^2 (*s*upervised coupled dictionary *l*earn*i*ng with group structures for *m*ulti-*m*odal retrieval) method (Zhuang et al., 2013) introduces coupled dictionary learning into supervised sparse coding and learns a set of mapping functions across different modalities for multi-modal retrieval. For classifying lung needle biopsy images, multimodal sparse representation-based classification (MSRC) method (Shi et al., 2013) aims to select the topmost discriminative samples for each individual modality as well as to guarantee the large diversity among different modalities.

Motivation and Contribution

Most multi-view (or multi-modal) DL methods mainly focus on the reconstructive accuracy, whereas enhancing the total discriminability of dictionaries and reducing the redundancy between multiple dictionaries have not been investigated comprehensively and thoroughly. The key of multi-view DL technique is how to utilize the complementary information among different dictionaries, learn more useful features for recognition and reduce the redundancy between dictionaries.

Information redundancy in original multi-view data will lead to redundancy in the learned dictionaries, which will bring trouble to subsequent classification. On the one hand, several single-view based works (Chen et al., 2013; Lin et al., 2012) have taken dictionary atom de-correlation into consideration. On the other side, discrimination dictionary learning has demonstrated to be effective in classification (Yang et al., 2011). Inspired by these two aspects, we propose an uncorrelated multi-view discrimination DL (UMD^{2}L) approach for recognition. We summarize the contributions of our work as following points:

(1) By making dictionary atoms correspond to the class labels such that the obtained reconstruction error is discriminative, we aim to jointly learn multiple dictionaries with totally favorable discriminative power.

(2) We design the uncorrelated constraint for multi-view DL, so as to reduce the redundancy among dictionaries learned from different views.

The proposed UMD^{2}L approach is verified on the Multi-PIE (Cai et al., 2006), AR (Martinez and Benavente 1998), COIL-20 (Murase and Nayar 1995) and MNIST (LeCun et al., 1998) datasets. Experimental results demonstrate its effectiveness as compared with several related methods.

Organization

The rest of the paper is organized as follows. We first briefly review the related multi-view supervised DL models. Then we describe our approach. We then discuss the differences between the proposed UMD^{2}L approach and related multi-view feature learning works, followed by experiments and conclusion.

Related Supervised Multi-view DL Models

SliM^2

SliM^2 (Zhuang et al., 2013) learns a set of dictionaries for M modality data respectively, i.e., $D=\left\{D^{(1)},D^{(2)},\cdots,D^{(M)}\right\}$ and their corresponding reconstruction coefficients $A=\left\{A^{(1)},A^{(2)},\cdots,A^{(M)}\right\}$. To do multi-modal retrieval, SliM^2 assumes that there exists a set of linear mappings $W=\left\{W^{(1)},W^{(2)},\cdots,W^{(M)}\right\}$. The objective function of SliM^2 is formulated as:

$$\begin{aligned}&\min \sum_{m=1}^{M}\left\|X^{(m)}-D^{(m)}A^{(m)}\right\|_F^2+\sum_{m=1}^{M}\sum_{l=1}^{J}\lambda_m\left\|A^{(m)}_{:,\Omega_l}\right\|_{1,2}\\&+\beta\sum_{m=1}^{M}\sum_{n\neq m}\left\|A^{(n)}-W^{(m)}A^{(m)}\right\|_F^2+\gamma\sum_{m=1}^{M}\left\|W^{(m)}\right\|_F^2 \quad , \qquad (1)\\&s.t.\ \ \left\|d_k^{(m)}\right\|\le 1,\ \ \forall k,\ \forall m\end{aligned}$$

where J is the number of classes, Ω_l represents the indices of the samples that belong to the l^{th} class, $A_{:,\Omega_l}$ is the coefficient matrix associated to those intra-modality data belonging to the l^{th} class. β, γ and $\lambda_m\left(m=1,\cdots,M\right)$ are tuning parameters denoting the weights of each term in Formula (1).

MSRC

MSRC (Shi et al., 2013) uses the binary sample selectors $\left\{\beta^S,\beta^C,\beta^T\right\}$ to select samples from original sample sets

$\{D^S, D^C, D^T\}$ for constructing dictionaries $\{\beta^S(D^S), \beta^C(D^C), \beta^T(D^T)\}$. Here, S, C and T separately denote shape, color, and texture modalities. The k^{th} cell nuclei can be denoted as a tuple $(x_k^S, x_k^C, x_k^T, y_k)_{k=1}^N$, where N denotes the number of training cell nuclei and y_k is the label of the k^{th} cell nuclei. For the k^{th} labeled training cell nuclei, MSRC denotes $f(x_k^m, \beta^m), m \in \{S, C, T\}$ as the mapping function, which is the label of x_k^m predicted by sub-classifier trained on the learned sub-dictionary $\beta^m(D^m)$. The selection criteria are that 1) each sub-dictionary after dictionary learning can train a good classifier independently and 2) the diversity among different sub-dictionaries after dictionary learning is encouraged to be large. Then, the objective function of MSRC is formulated as:

$$\min_{\beta} \exp\left\{ \begin{array}{l} -\sum_{k=1}^{N} \sum_{m\in\{S,C,T\}} \langle y_k, f(x_k^m, \beta^m)\rangle \\ -\sum_{k=1}^{N} \lambda \left[3 - \sum_{\substack{m,\hat{m}\in\{S,C,T\} \\ m\neq\hat{m}}} \langle f(x_k^m, \beta^m), f(x_k^{\hat{m}}, \beta^m)\rangle \right] \end{array} \right\}, \quad (2)$$

where $\langle a, b\rangle$ is the Kronecker delta function and λ is the parameter to control influence caused by the second term.

The Model of UMD^{2}L

In this section, we first briefly review the model of FDDL (Yang et al., 2011). Then we provide the formulation of UMD^{2}L, followed by the optimization and classification scheme of UMD^{2}L.

FDDL

Let $A = \{A_1, A_2, \cdots A_C\}$ be the training sample set, where A_i is the sub-set of training samples from class i and C is the number of classes. Instead of learning a shared dictionary to all classes, FDDL aims to learn a structured dictionary $D = [D_1, D_2, \cdots, D_C]$, where D_i is the class-specified sub-dictionary associated with class i. Suppose that X is the coding coefficient matrix of A over D, X can be written as $X = [X_1, X_2, \cdots, X_C]$, where X_i is the sub-matrix containing the coding coefficients of A_i over D. X_i^i and X_i^j are coding coefficients of A_i over the sub-dictionaries D_i and D_j, respectively. The objective function of FDDL is defined as:

$$J_{(D,X)} = \arg\min_{(D,X)} \left\{ \begin{array}{l} \sum_{i=1}^{C} r(A_i, D, X_i) + \lambda_1 \|X\|_1 \\ +\lambda_2 \left(tr(S_W(X) - S_B(X)) + \eta \|X\|_F^2 \right) \end{array} \right\}, \quad (3)$$

where $S_W(X)$ and $S_B(X)$ are separately the within-class scatter and between-class scatter of X, $r(A_i, D, X_i) = \|A_i - DX_i\|_F^2 + \|A_i - D_i X_i^i\|_F^2 + \sum_{\substack{j=1 \\ j\neq i}}^{C} \|D_j X_i^j\|_F^2$ is called discriminative fidelity term, and λ_1, λ_2 and η are scalar parameters. With the learned dictionary D, FDDL uses the reconstruction error for classification.

The Formulation of UMD^{2}L

In this subsection, we describe the model of the proposed UMD^{2}L approach. For the given M views (datasets) $A_k (k = 1, \cdots, M)$, we jointly learn discriminant dictionaries $D_k (k = 1, \cdots, M)$ with each corresponding to one view, and we design uncorrelated constraint for the multi-dictionary learning procedure to reduce the redundancy of dictionaries.

For DL task from the k^{th} view, let A_k^i denote the training sample subset from the i^{th} class of A_k, D_k^i and D_k^j separately denote the class-specified sub-dictionaries associated with class i and class j. Suppose that X_k^i (the coding coefficients of A_k^i over D_k) can be written as $X_k^i = [X_k^{i1}, X_k^{i2}, \cdots, X_k^{iC}]$, where X_k^{ij} is the coding coefficient of A_k^i over the sub-dictionary D_k^j. We require that D_k should have powerful reconstruction capability of A_k, and we also require that D_k should have powerful discriminative capability of samples in A_k. In other words, dictionary D_k should be able to represent A_k^i. In addition, since D_k^i is associated with the i^{th} class, it is expected that A_k^i should be represented by D_k^i but not by $D_k^j (i \neq j)$. Therefore, we can define the discriminative fidelity term for the k^{th} view as:

$$q(A_k^i, D_k, X_k^i) = \|A_k^i - D_k X_k^i\|_F^2 + \|A_k^i - D_k^i X_k^{ii}\|_F^2 + \sum_{\substack{j=1 \\ j\neq i}}^{C} \|D_k^j X_k^{ij}\|_F^2 . \quad (4)$$

To reduce redundancy of learned dictionaries, we design uncorrelated constraint for multi-view DL. The correlation coefficient between D_k and D_l (the dictionary corresponding to l^{th} view and $l \neq k$) can be defined as:

$$Corr(D_k, D_l) = \frac{\sum_{i=1}^{N_k} \sum_{j=1}^{N_l} Corr(d_k^i, d_l^j)}{N_k \cdot N_l}, \quad (5)$$

where d_k^i and d_l^j separately denote the i^{th} dictionary atom in D_k and the j^{th} dictionary atom in D_l, N_k and N_l separately denote the numbers of dictionary atoms in D_k and D_l. Here,

$$Corr\left(d_k^i, d_l^j\right) = \frac{\left(d_k^i - \overline{d_k^i}\cdot I\right)^T \left(d_l^j - \overline{d_l^j}\cdot I\right)}{\left\|d_k^i - \overline{d_k^i}\cdot I\right\| \bullet \left\|d_l^j - \overline{d_l^j}\cdot I\right\|}$$

indicates the correlation coefficient between d_k^i and d_l^j, where $\overline{d_k^i}$ and $\overline{d_l^j}$ are mean values of these two atoms, I is a vector with all elements being equal to one. Thus, the objective function of UMD^{2}L can be formulated as:

$$\begin{aligned} &\min \sum_{k=1}^{M}\sum_{i=1}^{C} q\left(A_k^i, D_k, X_k^i\right) + \lambda \sum_{k=1}^{M} \left\|X_k\right\|_1 \\ &s.t.\ \ Corr\left(D_k, D_l\right) = 0,\ l \neq k \end{aligned} . \quad (6)$$

Note that the discriminative coefficient term in FDDL, namely $tr\left(S_W(X) - S_B(X)\right) + \eta\left\|X\right\|_F^2$, is not used in the model of UMD^{2}L because we found this term has small influence on classification result in experiment. Although the objective function in Eq. (6) is not jointly convex to $\left(D_k, X_k\right)_{k=1}^{M}$, it is convex with respect to each of D_k and X_k when the other variables are fixed.

The Optimization of UMD^{2}L

The variables in Eq. (6) can be optimized by using a two-level optimization strategy: 1) updating variables of the k^{th} view by fixing variables of other views; 2) for the k^{th} view, updating X_k by fixing D_k and then updating D_k by fixing X_k.

When variables corresponding to the k^{th} view are optimized, $D_l\left(l \neq k\right)$, $X_l\left(l \neq k\right)$ and D_k are supposed to be fixed. Then the objective function in Eq. (6) is reduced to a sparse coding problem to compute $X_k = \left[X_k^1, X_k^2, \cdots, X_k^C\right]$. Here we compute X_k^i class by class. When computing X_k^i, all $X_k^j\left(j \neq i\right)$ are fixed. Thus the objective function in Eq. (6) is reduced to:

$$J_{(X_k)} = \arg\min_{(X_k)} \left\{ \begin{array}{l} \left\|A_k^i - D_k X_k^i\right\|_F^2 + \left\|A_k^i - D_k^i X_k^{ii}\right\|_F^2 \\ + \sum_{\substack{j=1 \\ j\neq i}}^{C} \left\|D_k^j X_k^{ij}\right\|_F^2 + \lambda_1 \left\|X_k\right\|_1 \end{array} \right\} . \quad (7)$$

The sparse coding problem of Eq. (7) can be solved by using the Iterative Projection Method (IPM, Rosasco et al., 2009). Please refer to FDDL for the detailed derivation.

When X_k is fixed, we update D_k^i class by class. When updating D_k^i, all $D_k^j\left(j \neq i\right)$ are fixed. Then the objective function in Eq. (6) is reduced to:

$$\begin{aligned} &J_{(D_k^i)} = \arg\min_{(D_k^i)} \left\{ p\left(D_k^i\right) \right\} \\ &s.t.\ Corr\left(D_k^i, D_l\right) = 0 \end{aligned} , \quad (8)$$

where

$$\begin{aligned} p\left(D_k^i\right) = &\left\| A_k - D_k^i X_k^i - \sum_{\substack{j=1 \\ j\neq i}}^{C} D_k^j X_k^j \right\|_F^2 \\ &+ \left\| A_k^i - D_k^i X_k^{ii} \right\|_F^2 + \sum_{\substack{j=1 \\ j\neq i}}^{C} \left\| D_k^i X_k^{ij} \right\|_F^2 \end{aligned} ,$$

$$Corr\left(D_k^i, D_l\right) = \frac{1}{N_k^i \cdot N_l} \sum_{n=1}^{N_k^i} \sum_{j=1}^{N_l} Corr\left(d_k^{in}, d_l^j\right)$$

and

$$= \frac{1}{N_k^i \cdot N_l} \sum_{n=1}^{N_k^i} \sum_{j=1}^{N_l} \frac{\left(d_k^{in} - \overline{d_k^{in}}\cdot I\right)^T \left(d_l^j - \overline{d_l^j}\cdot I\right)}{\left\|d_k^{in} - \overline{d_k^{in}}\cdot I\right\| \bullet \left\|d_l^j - \overline{d_l^j}\cdot I\right\|} .$$

Here, N_k^i denotes the number of sub-dictionary atoms in D_k^i, d_k^{in} is the n^{th} sample of D_k^i and $\overline{d_k^{in}}$ denotes mean value of d_k^{in}. The uncorrelated constraint in Eq. (8) is simplified by using

$$s.t.\ \sum_{n=1}^{N_k^i} \sum_{j=1}^{N_l} \left(d_k^{in} - \overline{d_k^{in}}\cdot I\right)^T \left(d_l^j - \overline{d_l^j}\cdot I\right) = 0 . \quad (9)$$

In general, we require that d_k^{in} is a unit vector. Then we can solve the quadratic programming problem in Eq. (8) according to the literature (Yang et al., 2010), which updates D_k^i atom by atom.

Algorithm 1 realizes the proposed UMD^{2}L approach. UMD^{2}L converges since the two alternative optimizations for each view in it are both convex.

Algorithm 1. UMD^{2}L

1. Initialization $\left\{D_1, D_2, \cdots, D_M\right\}$.

We initialize all the atoms of each D_i as random vectors with unit l_2-norm.

2. For $k = 1$ **to** M **do**

2.1 Update the sparse coding coefficients X_k.

Fix D_k and solve X_k^i class by class by solving Eq. (7) with the IPM algorithm.

2.2 Update the dictionary D_k.

Fix X_k and solve D_k^i class by class by solving Eqs. (8-9) with the optimization algorithm in the literature (Yang et al., 2010).

End

3. Iterative learning.

Repeat **2** until the values of objective function in adjacent iterations are close enough, or the maximum number of iterations is reached.

4. Output $\left\{X_1, X_2, \cdots, X_M\right\}$ **and** $\left\{D_1, D_2, \cdots, D_M\right\}$.

The Classification Scheme

When $\{D_1, D_2, \cdots, D_M\}$ is available, a testing sample can be classified via coding it over these dictionaries. For the given testing sample $y = \{y_1, y_2, \cdots, y_M\}$, we code y_k over D_k for $k = 1:M$. The sparse coding coefficients are obtained by solving:

$$(\hat{\alpha}_1, \cdots \alpha_M) = \arg\min_{\alpha_1, \cdots \alpha_M} \left\{ \sum_{k=1}^{M} \left(\|y_k - D_k \alpha_k\|_2^2 + \gamma \|\alpha_k\|_1 \right) \right\}, \quad (10)$$

where γ is a constant.

The variables in Eq. (10) can be optimized by using alternate optimization strategy, which is updating α_k for the $k^{th}(k = 1, \cdots, M)$ view by fixing coding coefficients corresponding to the other views. For the $k^{th}(k = 1, \cdots, M)$ view, $\hat{\alpha}_k$ can be obtained by using the sparse representation based classification (SRC) method (Wright et al. 2009). $\hat{\alpha}_k(k = 1, \cdots, M)$ can be written as $\hat{\alpha}_k = [\alpha_k^1; \alpha_k^2; \cdots; \alpha_k^C]$, where $\hat{\alpha}_k^i$ is the coefficient vector associated with sub-dictionary D_k^i. We define the metric for final classification as:

$$e_i = \sum_{k=1}^{M} \|y_k - D_k^i \hat{\alpha}_k^i\|_2^2, \quad (11)$$

where $\|y_k - D_k^i \hat{\alpha}_k^i\|_2^2$ denotes the reconstruction error of class i from the k^{th} view. Then we do classification via $identity(y) = \arg\min_i \{e_i\}$.

Discussion

Comparison with Multi-view Subspace Learning Methods

Multi-view subspace learning is an important research direction of multi-view feature learning. CCA based and discriminant analysis based multi-view subspace learning (such as MCCA and MvDA, respectively) are two representative techniques. CCA based multi-view subspace learning methods are dedicated to learn features depicting intrinsic correlation among multiple views. Discriminant analysis based multi-view subspace learning methods usually aim to achieve multiple linear transformations, with which the between-class variations of low-dimensional embeddings are maximized and the within-class variations of low-dimensional embeddings are minimized. The proposed UMD^{2}L approach differs from these multi-view subspace learning methods, because we aim to fully extract complementary discriminant information from multiple views by learning multiple uncorrelated discrimination dictionaries for helping recognition.

Comparison with Existing Multi-view Dictionary Learning Methods

SMBR (Shekhar et al., 2014), SliM^2 (Zhuang et al., 2013) and MSRC (Shi et al., 2013) are three representative multi-view dictionary learning methods. Specifically, SMBR uses original training samples as dictionary. In order to conduct multi-model retrieval, besides multiple dictionaries, SliM^2 learns a set of linear mappings which characterize connections of sparse codes corresponding to different views. MSRC employs a simple comparison strategy to encourage large diversity among the learning dictionaries of different views. Our approach makes dictionary atoms correspond to the class labels, which can acquire totally favorable discriminative power. Furthermore, to effectively combine multiple views for recognition task, we design the uncorrelated constraint to reduce the redundancy among dictionaries learned from different views.

Experiments

In this section, we compare the proposed UMD^{2}L approach with multi-view subspace learning methods including MCCA and MvDA, and multi-view DL methods including SMBR, SliM^2 and MSRC on the Multi-PIE, AR, COIL-20 and MNIST datasets.

In all experiments, the tuning parameters in UMD^{2}L (λ in dictionary learning phase, and γ in classification phase) and the parameters of all compared methods are evaluated by 5-fold cross validation to avoid over-fitting. Concretely, the parameters of UMD^{2}L are set as $\lambda = 0.005$ and $\gamma = 0.001$ for three datasets. In addition, the default dictionary atoms number for each view in UMD^{2}L is set as the number of training samples.

Experiments on the Multi-PIE Dataset

Multi-PIE dataset contains more than 750,000 images of 337 people under various views, illumination and expressions. More introductions about this dataset can be referred to the literature (Cai et al., 2006). Here, a subset about 1632 images from 68 peoples (24 images for each people) with 5 different poses (C05, C07, C09, C27, C29) is selected for experiment. The image size is 64×64 pixels. Figure 1 shows demo images (with 5 different poses) of one subject. PCA transformation (Turk and Pentland 1991) is used to reduce the dimension of samples to 100. And we randomly select 8 samples per class for training, use the remainder for testing, and run all compared methods 20 times.

Figure 1: Demo images of one subject in the Multi-PIE dataset.

Figure 2 shows average inter-view correlation coefficients of the dictionaries learned by SMBR, SliM^2, MSRC and UMD^{2}L, where the absolute values of coefficients are given. Compared with SMBR that uses original training samples to construct dictionary, UMD^{2}L effectively reduce correlation of dictionaries learned from different views. In addition, as compared with MSRC which encourages diversity of dictionaries corresponding to different views, UMD^{2}L obtains better de-correlation effect. All these demonstrate effectiveness of the proposed uncorrelated constraint.

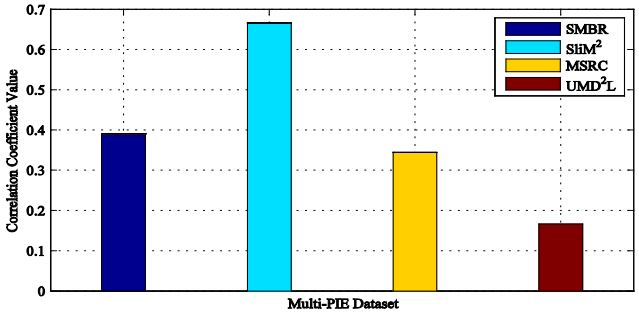

Figure 2: Average inter-view correlation coefficients of the learned dictionaries by all the compared DL methods on Multi-PIE dataset.

Experiments on the AR Dataset

The AR face dataset (Martinez and Benavente 1998) contains images of 119 individuals (26 images for each people), including frontal view of faces under different lighting conditions and with various occlusions. Each image is scaled to 60×60. All image samples of one subject are shown in Figure 3. We extract Gabor transformation features (Grigorescu, Petkov, and Kruizinga 2002), Karhunen-Loeve (KL) transformation features (Fukunaga and Koontz 1970) and Local Binary Patterns (LBP) (Ahonen, Hadid, and Pietikainen 2006) to construct three feature sets for experiment. The process of multiple feature sets construction is illustrated in Figure 4. Similar to experimental setting for the Multi-PIE dataset, we employ PCA transformation to reduce the dimension of these feature samples to 100. We randomly select 8 samples per class for training, use the remainder for testing, and run all compared methods 20 times.

Figure 3: Demo images of one subject in the AR dataset.

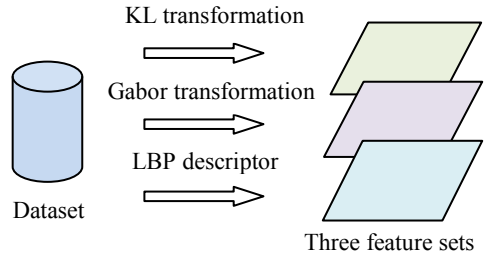

Figure 4: Construction of multiple views.

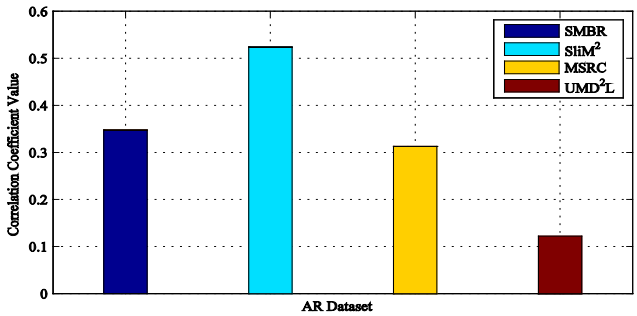

Figure 5: Average inter-view correlation coefficients of the learned dictionaries by all the compared DL methods on AR dataset.

Figure 5 shows average inter-view correlation coefficients of the dictionaries learned by SMBR, SliM^2, MSRC and UMD^{2}L. Due to Figure 5, the correlation coefficient corresponding to UMD^{2}L is smaller than those of multi-view DL methods including SMBR, SliM^2 and MSRC, which demonstrates effectiveness of the proposed uncorrelated constraint.

Experiments on the COIL-20 Dataset

The COIL-20 object dataset (Murase and Nayar 1995) contains 1440 grayscale images of 20 objects (72 images per object) under various poses. The objects are rotated through 360 degrees and taken at the interval of 5 degrees. The size of each image is 64×64 pixels. Image samples of one subject are shown in Figure 6. We also extract Gabor transformation features, Karhunen-Loeve transformation features and Local Binary Patterns to build three feature sets for experiment. PCA transformation is employed to reduce the dimension of these feature samples to 100. 36 samples per class are randomly chosen to form the training set, while the remaining samples are regarded as the testing set. The random selection process is performed 20 times, and we record the average experimental results for all compared methods.

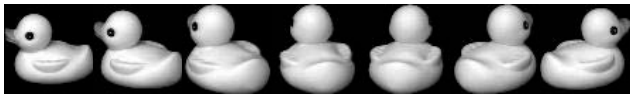

Figure 6: Demo images of one object in the COIL-20 dataset.

Figure 7 shows average inter-view correlation coefficients of the dictionaries learned by SMBR, SliM^2, MSRC and UMD^{2}L. Due to Figure 7, our approach achieves the smallest multi-view dictionary correlation, which demonstrates effectiveness of the proposed uncorrelated constraint.

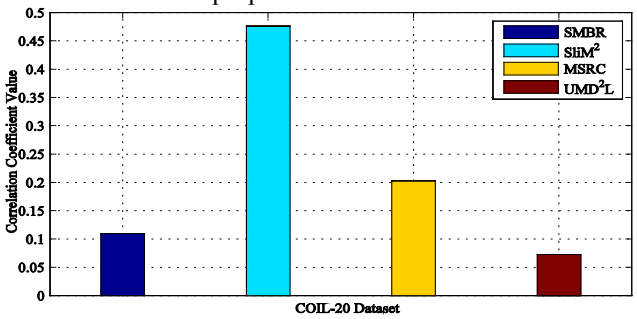

Figure 7: Average inter-view correlation coefficients of the learned dictionaries by all the compared DL methods on COIL-20 dataset.

Experiments on the MNIST Dataset

The MNIST dataset (LeCun et al. 1998) used in our experiment contains 1000 handwritten digit images (100 images for each digit). The image size is 28×28 pixels. Figure 8 shows demo images of ten digits. Gabor transformation features, Karhunen-Loeve transformation features and Local Binary Patterns are extracted to build three feature sets for experiment. And the dimension of feature samples is reduced to 100 by using the PCA transformation. We randomly select 40 samples per class for training, use the remainder for testing, and run all compared methods 20 times.

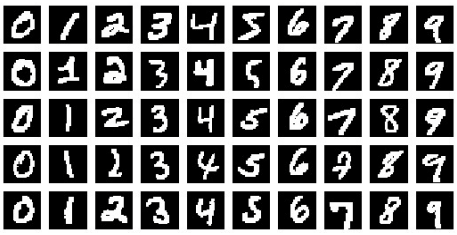

Figure 8: Demo images in the MNIST database.

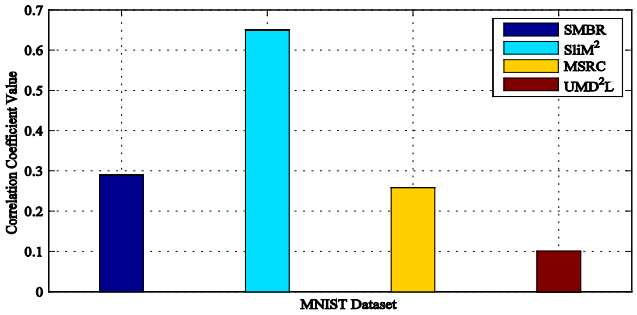

Figure 9: Average inter-view correlation coefficients of the learned dictionaries by all the compared DL methods on MNIST dataset.

Figure 9 shows average inter-view correlation coefficients of the dictionaries learned by SMBR, SliM^2, MSRC and UMD^{2}L. Due to Figure 9, the correlation coefficient corresponding to UMD^{2}L is much smaller than those of multi-view DL methods including SMBR, SliM^2 and MSRC, which demonstrates effectiveness of the proposed uncorrelated constraint.

Table 1 shows the average recognition rates of all compared methods across 20 random running on the Multi-PIE, AR, COIL-20 and MNIST datasets. The mean value of each method's average recognition rates corresponding to four different datasets is also listed in Table 1. As compared with related multi-view feature learning methods including MCCA, MvDA, SMBR, SliM^2 and MSRC, UMD^{2}L improves the average recognition rates at least by 1.67% (=95.10-93.43).

Table 1: Average recognition rates on three datasets.

Method	Average recognition rates (%)				
	Multi-PIE	AR	COIL-20	MNIST	Mean
MCCA	94.06	93.61	95.64	86.76	92.51
MvDA	94.28	94.19	95.33	87.42	92.80
SMBR	91.57	91.83	92.85	82.30	89.63
SliM^2	94.47	94.22	95.67	88.51	93.21
MSRC	93.02	94.64	96.56	89.52	93.43
UMD^{2}L	95.85	96.37	97.74	90.44	**95.10**

To statistically analyze the recognition rates given in Table 1, we conduct a statistical test, i.e., Mcnemar's test (Draper, Yambor, and Beveridge 2002). This test can provide statistical significance between UMD^{2}L and other methods. Here, the Mcnemar's test uses a significance level of 0.05, that is, if the p-value is below 0.05, the performance difference between two compared methods is considered to be statistically significant. Table 2 shows the p-values between UMD^{2}L and other compared methods on 4 datasets. According to Table 2, the proposed approach indeed makes

a statistically significant difference in comparison with the related methods.

Table 2: P-values between UMD^{2}L and other compared methods on four datasets.

Datasets	UMD^{2}L vs. MCCA	UMD^{2}L vs. MvDA	UMD^{2}L vs. SMBR	UMD^{2}L vs. SliM^2	UMD^{2}L vs. MSRC
Multi-PIE	1.29×10^{-16}	3.15×10^{-23}	2.50×10^{-18}	1.41×10^{-21}	1.57×10^{-23}
AR	1.76×10^{-25}	5.31×10^{-13}	1.17×10^{-17}	2.73×10^{-11}	1.66×10^{-14}
COIL-20	3.66×10^{-17}	2.30×10^{-19}	4.52×10^{-12}	1.34×10^{-16}	2.29×10^{-13}
MNIST	2.23×10^{-16}	3.39×10^{-12}	2.83×10^{-16}	2.34×10^{-28}	5.39×10^{-14}

Conclusion

Due to state-of-the-art recognition performance of dictionary learning (DL), multi-view DL has now become an interesting multi-view feature learning technique. For this technique, how to enhance the total discriminability of dictionaries and reduce their redundancy is a crucial research topic. In this paper, we propose a novel approach called uncorrelated multi-view discrimination dictionary learning (UMD^{2}L). By making dictionary atoms correspond to the class labels, it jointly learns multiple dictionaries and acquires totally favorable discriminative power. Moreover, UMD^{2}L designs the uncorrelated constraint to reduce the redundancy among dictionaries learned from different views.

By employing four multi-view datasets, experiments demonstrate that the proposed approach achieves better recognition results than two representative multi-view subspace learning methods and several representative multi-view DL methods. In addition, experimental results show effectiveness of the designed uncorrelated constraints for multi-view DL.

Acknowledgements

The work described in this paper was supported by the National Nature Science Foundation of China under Project Nos. 61231015, 61172173, 61272273 and 61233011, the Major Science and Technology Innovation Plan of Hubei Province (No. 2013AAA020), the Guangdong-Hongkong Key Domain Breakthrough Project of China (No. 2012A090200007).

References

Ahonen, T.; Hadid, A.; and Pietikainen, M. 2006. Face description with local binary patterns: application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28(12):2037-2041.

Barchiesi, D.; and Plumbley, M. D. 2013. Learning incoherent dictionaries for sparse approximation using iterative projections and rotations. *IEEE Transactions on Signal Processing* 61(8): 2055-2065.

Cao, T.; Jojic, V.; Modla, S.; Powell, D.; Czymmek, K.; and Niethammer, M. 2013. Robust multimodal dictionary learning. In *Proceedings of the 16th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 259-266.

Cai, D.; He, X.; Han, J.; and Zhang, H. J. 2006. Orthogonal laplacianfaces for face recognition. *IEEE Transactions on Image Processing* 15(11):3608-3614.

Chen, S.; Ding, C.; Luo, B.; and Xie, Y. 2013. Uncorrelated lasso. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, 166-172.

Draper, B. A.; Yambor, W. S.; and Beveridge, J. R. 2002. Analyzing PCA-based face recognition algorithms: eigenvector selection and distance measures. *Empirical Evaluation Methods in Computer Vision*, 1-15.

Feng, Z.; Yang, M.; Zhang, L.; Liu, Y.; and Zhang, D. 2013. Joint discriminative dimensionality reduction and dictionary learning for face recognition. *Pattern Recognition* 46(8):2134-2143.

Fukunaga, K.; and Koontz, W. L. 1970. Application of the Karhunen-Loeve expansion to feature selection and ordering. *IEEE Transactions on Computers* 100(4):311-318.

Gao, L.; Qi, L.; Chen, E.; and Guan, L. 2012. Discriminative multiple canonical correlation analysis for multi-feature information fusion. In *IEEE International Symposium on Multimedia*, 36-43.

Grigorescu, S. E.; Petkov, N.; and Kruizinga, P. 2002. Comparison of texture features based on Gabor filters. *IEEE Transactions on Image Processing* 11(10):1160-1167.

Guo, Y. 2013. Convex subspace representation learning from multi-view data. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, 387-393.

Han, Y.; Wu, F.; Tao, D.; Shao, J.; Zhuang, Y.; and Jiang, J. 2012. Sparse unsupervised dimensionality reduction for multiple view data. *IEEE Transactions on Circuits and Systems for Video Technology* 22(10):1485-1496.

Irie, G.; Liu, D.; Li, Z.; and Chang, S. F. 2013. A bayesian approach to multimodal visual dictionary learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 329-336.

Kan, M.; Shan, S.; Zhang, H; Lao, S.; and Chen, X. 2012. Multi-view discriminant analysis. In *Proceedings of the 12th European Conference on Computer Vision (ECCV)*, 808-821.

LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11):2278-2324.

Li, L.; Li, S.; and Fu, Y. 2013. Discriminative dictionary learning with low-rank regularization for face recognition. In *IEEE Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, 1-6.

Lin, T.; Liu, S.; and Zha, H. 2012. Incoherent dictionary learning for sparse representation. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR)*, 1237-1240.

Li, Y. O.; Adali, T.; Wang, W.; and Calhoun, V. D. 2009. Joint blind source separation by multiset canonical correlation analysis. *IEEE Transactions on Signal Processing* 57(10):3918-3929.

Mailhe, B.; Barchiesi, D.; and Plumbley, M. D. 2012. INK-SVD: Learning incoherent dictionaries for sparse representations. In *IEEE Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 3573-3576.

Ma, L.; Wang, C.; Xiao, B.; and Zhou, W. 2012. Sparse representation for face recognition based on discriminative low-rank dictionary learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2586-2593.

Martinez, A. M.; and Benavente, R. 1998. The AR Face Database, CVC Technical Report, 24, Ohio State University, Columbus, OH.

Memisevic, R. 2012. On multi-view feature learning. In *Proceedings of the 29th International Conference on Machine Learning (ICML)*, 161-168.

Monaci, G.; Jost, P.; Vandergheynst, P.; Mailhe, B.; Lesage, S.; and Gribonval, R. 2007. Learning multimodal dictionaries. *IEEE Transactions on Image Processing* 16(9):2272-2283.

Murase, H.; and Nayar, S. K. 1995. Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision* 14(1):5-24.

Rosasco, L.; Verri, A.; Santoro, M.; Mosci, S.; and Villa, S. 2009. Iterative projection methods for structured sparsity regularization, MIT Technical Reports, MIT-CSAIL-TR-2009-050, CBCL-282, Massachusetts Institute of Technology.

Sharma, A.; Kumar, A.; Daume, H.; and Jacobs, D. W. 2012. Generalized multiview analysis: a discriminative latent space. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2160-2167.

Shekhar, S.; Patel, V. M.; Nasrabadi, N. M.; and Chellappa, R. 2014. Joint sparse representation for robust multimodal biometrics recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36(1):113-126.

Shi, Y.; Gao, Y.; Yang, Y.; Zhang, Y.; and Wang, D. 2013. Multi-modal sparse representation-based classification for lung needle biopsy images. *IEEE Transactions on Biomedical Engineering* 60(10):2675-2685.

Su, Y.; Fu, Y.; Gao, X; and Tian, Q. 2012. Discriminant learning through multiple principal angles for visual recognition. *IEEE Transactions on Image Processing* 21(3): 1381-1390.

Tosic, I.; and Frossard, P. 2011. Dictionary learning for stereo image representation. *IEEE Transactions on Image Processing* 20(4):921-934.

Turk, M.; and Pentland, A. 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3(1):71-86.

Wang, S.; Zhang, L.; Liang, Y.; and Pan, Q. 2012. Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2216-2223.

Wright, J.; Yang, A. Y.; Ganesh, A.; Sastry, S. S.; and Ma, Y. 2009. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31(2):210-227.

Wu, F.; Yu, Z.; Yang, Y.; Tang, S.; Zhang, Y.; and Zhuang, Y. 2014. Sparse multi modal hashing. *IEEE Transactions on Multimedia* 16(2):427-439.

Yang, M.; Zhang, L.; Feng, X.; and Zhang, D. 2011. Fisher discrimination dictionary learning for sparse representation. In *IEEE Conference on Computer Vision (ICCV)*, 543-550.

Yang, M.; Zhang, L.; Yang, J.; and Zhang, D. 2010. Metaface learning for sparse representation based face recognition. In *Proceedings of the 17th International Conference on Image Processing (ICIP)*, 1601-1604.

Zheng, S.; Xie, B.; Huang, K.; and Tao, D. 2011. Multi-view pedestrian recognition using shared dictionary learning with group sparsity. In *Proceedings of the 18th International Conference* on *Neural Information Processing (ICONIP)*, 629-638.

Zhuang, Y.; Wang, Y.; Wu, F.; Zhang, Y.; and Lu, W. 2013. Supervised coupled dictionary learning with group structures for multi-modal retrieval. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, 1070-1076.