Discriminative Dictionary Learning with Pairwise Constraints

Huimin Guo
Zhuolin Jiang
LARRY S. DAVIS
UNIVERSITY OF MARYLAND

Nov. 6^{th}, 2012

Outline

- Introduction/motivation
- Dictionary Learning
- Discriminative Dictionary Learning with Pairwise Constraints
- Experiments
  - Face verification
  - Face recognition
- Summary

Applications

- Pair-matching type problems, only binary class information
  - Face Verification (same/different)
  - Pair-matching (same/different, similar/dissimilar)
  - Image Retrieval (relevant/irrelevant)

- Classification problems, category labels provided
  - Face Recognition
  - Image Classification
  - …

Motivations

- Pair matching problems are common in many practical applications; we can use provided pairwise constraints explicitly
- DDL-PC1: the learned dictionary encourages feature points from the same class (or a similar pair) to have similar sparse codes, discriminative+
- DDL-PC2: furthermore add in a classification error term in classifier construction for a unified objective function, discriminative++

Dictionary Learning

- find optimized dictionaries A* that provides a succinct representation for most statistically representative input signals
- Solving l1-minimization

$$< A^*, X^* >= \arg\min_{A,X} \sum_{i=1}^{N} (\boxed{\|y_i - Ax_i\|_2^2} + \boxed{\gamma\|x_i\|_1})$$

Reconstruction Term   Regularization Term

$(y_1 \ldots y_N)$: training signals; $(x_1 \ldots x_N)$: sparse codes for $(y_1 \ldots y_N)$

DDL-PCI

- The objective function of Dictionary Learning

$$< A^*, X^* >= \arg\min_{A,X} \sum_{i=1}^{N} \left(\| y_i - Ax_i \|_2^2 + \gamma \| x_i \|_1 \right) + \beta/2 \sum_{i,j=1}^{N} \| x_i - x_j \|_2^2 \, M_{ij}$$

$$= \arg\min_{A,X} \sum_{i=1}^{N} \left(\| y_i - Ax_i \|_2^2 + \gamma \| x_i \|_1 \right) + \beta \left(Tr(X^T XD) - Tr(X^T XM) \right)$$

$$= \arg\min_{A,X} \sum_{i=1}^{N} \left(\boxed{\| y_i - Ax_i \|_2^2} \boxed{+\gamma \| x_i \|_1} \right) + \boxed{\beta \left(Tr(X^T XL) \right)}$$

Reconstruction Term Regularization Term Discrimination Term

$(y_1 \ldots y_N)$: training signals; $(x_1 \ldots x_N)$: sparse codes for $(y_1 \ldots y_N)$

M: Adjacency (weight) matrix;

$D = diag(d_1 .. d_N)$: degree matrix, where $d_i = \sum_{j=1}^{N} M_{ij}$

L=*D*-*M* : Laplacian matrix

Optimization

- The objective function is not convex for *A* and *X* simultaneously, but fortunately, it is convex in *A* (while holding *X* fixed) and convex in *X* (while holding *A* fixed) .
- When *A* is fixed, we optimize each $\boldsymbol{x}_i$ alternately and fix the other $\boldsymbol{x}_j\ (\boldsymbol{j} \neq \boldsymbol{i})$ for other signals. Optimizing the objective function is equivalent to

$$\min_{\boldsymbol{x}_i} \boldsymbol{L}(\boldsymbol{x}_i) = \|\boldsymbol{y}_i - \boldsymbol{A}\boldsymbol{x}_i\|_2^2 + \gamma \|\boldsymbol{x}_i\|_1 + \beta/2 \left(2\boldsymbol{x}_i^T(\boldsymbol{X}\boldsymbol{L}_i) - \boldsymbol{x}_i^T \boldsymbol{x}_i \boldsymbol{L}_{ii}\right)$$

Here we modify *feature sign search algorithm** to solve this convex problem.

*H. Lee, A. Batte, R. Raina and A. Y. Ng, Efficient Sparse Coding Algorithm. NIPS2006

Optimization (cont.)

- Given all the sparse codes X, Optimizing the objective function is equivalent to

$$\min_{\boldsymbol{A}} \boldsymbol{L}(\boldsymbol{A}) = \sum_{\boldsymbol{i}=1}^{\boldsymbol{N}} || \boldsymbol{y}_{\boldsymbol{i}} - \boldsymbol{A}\boldsymbol{x}_{\boldsymbol{i}} ||_2^2, \qquad \boldsymbol{s.t.} \quad \boldsymbol{a}_{\boldsymbol{i}}^{\boldsymbol{T}} \boldsymbol{a}_{\boldsymbol{i}} \leq 1$$

This is L2 constrained least square problem. We can optimize it using Newton's method or conjugate gradient.

DDL-PC2

- The objective function of Dictionary Learning

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

$$+\frac{\beta}{2}\sum_{i,j=1}^{N}(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 M_{ij}) \boxed{+ \alpha \sum_{i=1}^{N}(\|\boldsymbol{h}_i - W\boldsymbol{x}_i\|_2^2 + \lambda\|W\|_2^2)}$$

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. $\boldsymbol{h}_i = [0, 0, \ldots 1 \ldots 0, 0]^T \in \mathbb{R}^m$ (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 $\boldsymbol{y}_i$.

Matching approach

- Face Verification (given same/not same)
- y1, y2 are the same person, y3, y4 are the same person, y5 y6 are different person

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

Matching approach

- Face Recognition
  - class labels are given for each image in the training set. The pair relationships are derived from the category labels
  - Matrix *M* encoding the (dis)similarity information can be defined as

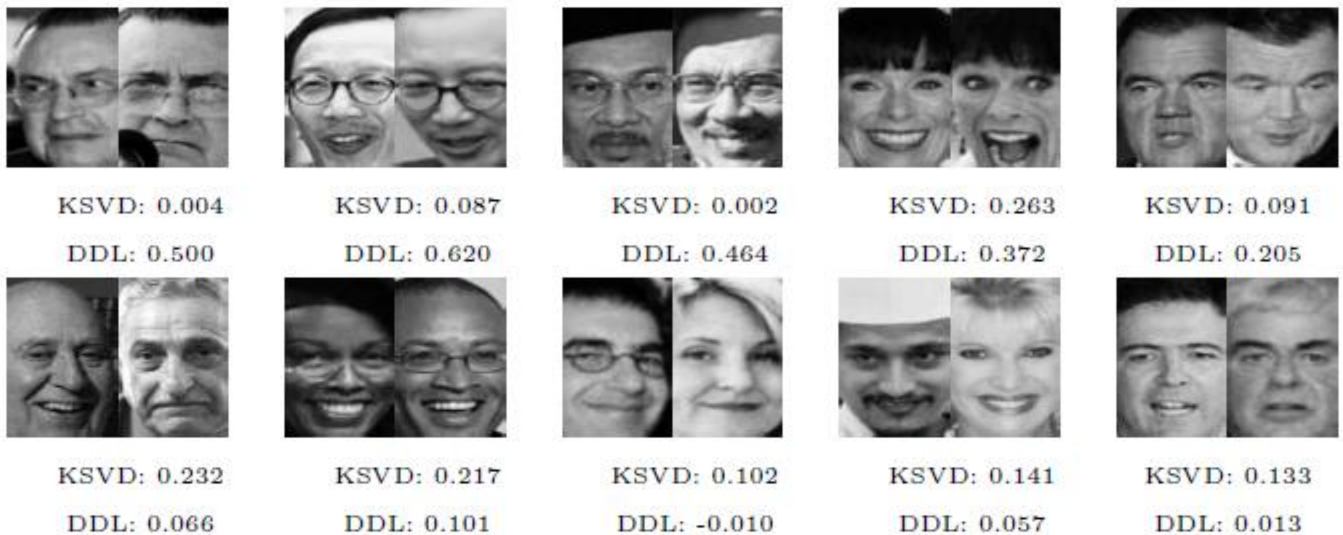
$$M_{ij} = \begin{cases} 1, \text{ if } (y_i, y_j) \in c_k, k = 1...m \\ 0, \text{ otherwise} \end{cases}$$

Experiments: Face Verification

- LFW (Labeled Faces in the Wild) dataset
  - Remarkable variations caused by
    - Pose, facial appearance, age, lighting, expression,
    - occlusion, scale, camera, misalignment, hairstyle, etc.
- 13233 images
- 5749 people

Experimental Results

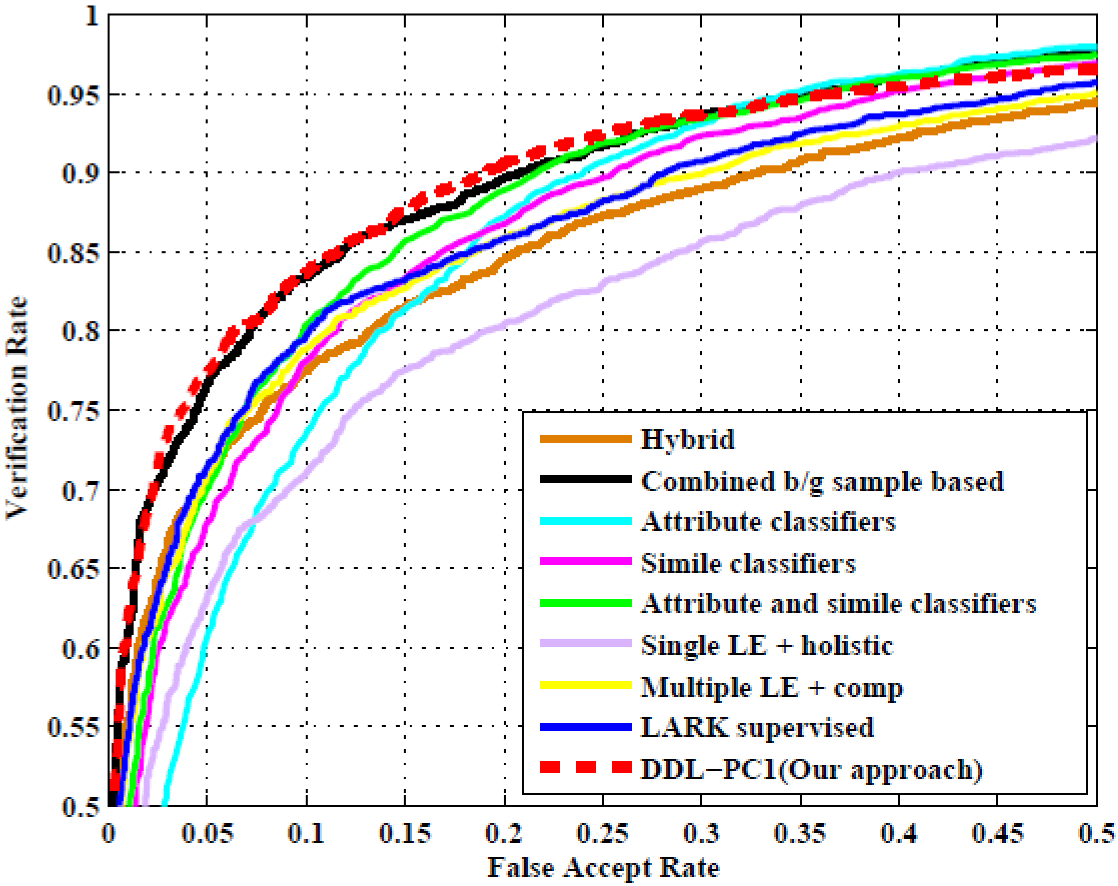
- Face Verification on LFW



- 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.

Evaluation on LFW

- ROC curve



Experiments: Face Recognition

- Extended Yale-B
  - Recognition results using random-face features on the Extended YaleB.

Method	K-SVD[6]	D-KSVD[13]	SRC[5]	LLC[34]	LC-KSVD[12]	**DDL-PC1**	**DDL-PC2**
Acc. (%)	90.5	94.1	88.6	82.3	95.0	**94.5**	**95.3**

- AR face database
  - Recognition results using random-face features on the Extended AR.

Method	K-SVD[6]	D-KSVD[13]	SRC[5]	LLC[34]	LC-KSVD[12]	**DDL-PC1**	**DDL-PC2**
Acc. (%)	87.2	88.8	74.5	88.7	93.7	**94.0**	**96.0**

Summary

- a novel dictionary learning approach that tackles the pair matching and classification problem in a unified framework
- a discriminative term called 'pairwise sparse code error' based on pairwise constraints
- + the classification error term for better discriminating power.

Thanks!
Q&A