Locality Sensitive Discriminative Dictionary Learning (ICIP 2015)

Jun Guo*, Yanqing Guo, Yi Li, Bo Wang, Ming Li

School of Information and Communication Engineering,
Dalian University of Technology, China

guojun@mail.dlut.edu.cn
Outline

- Introduction
- Motivation of Our Work
- The Proposed Method
- Experimental Results
- Summarization and Future Work
Outline

• Introduction
• Motivation of Our Work
• The Proposed Method
• Experimental Results
• Summarization and Future Work
Dictionary Learning (DL)

- A basic framework of DL:

$$
\min_{D,X} \sum_{i=1}^{N} \left( \| y_i - Dx_i \|_2^2 + \tau \| x_i \|_p \right)
$$

$$
\min_{D,X} \| Y - DX \|_F^2 + \tau \| X \|_p
$$

- Training data: $Y = [y_1, \ldots, y_N] \in \mathbb{R}^{n \times N}$
- Dictionary: $D = [d_1, \ldots, d_K] \in \mathbb{R}^{n \times K}$
- Coding vectors: $X = [x_1, \ldots, x_N] \in \mathbb{R}^{K \times N}$
- Regularization parameter: $\tau \geq 0$

Emphasize representation rather than discrimination!
Discriminative Dictionary Learning (DDL)

• A general formula of DDL:

\[
\min_{D,X} \|Y - DX\|_F^2 + \tau \|X\|_p + \alpha f(Y, D, X, H)
\]

– Structured incoherence of dictionary
– Fisher discrimination on the dictionary and codes
– Label consistency of codes
– Transform-invariance of dictionary
– Joint dictionary learning and subspace clustering
– ......
Outline

• Introduction
• Motivation of Our Work
• The Proposed Method
• Experimental Results
• Summarization and Future Work
Motivation of Our Work

• Illustration:

- DDL + $k$NN classifier
- In the coding space:
  - same-label neighbors are orderly preserved
  - neighbors with different labels are repelled

Maximize the local margin between different classes.
Motivation of Our Work

• Illustration:

– Integrate two significant characters into DL
  ➢ preserve ordinal locality
  ➢ strengthen discriminability

Locality Sensitive Discriminative Dictionary Learning (LSDDL)
Outline

• Introduction
• Motivation of Our Work
• The Proposed Method
• Experimental Results
• Summarization and Future Work
The Proposed Method

• Construct two graphs via Gaussian kernel:
  – a within-class graph \( G_w \)
    ➢ the corresponding weight matrix
    \[
    W_{w,ij} = \begin{cases} 
    w(y_i, y_j), & \text{if } y_i \in N_w(y_j) \text{ or } y_j \in N_w(y_i) \\
    0, & \text{otherwise.}
    \end{cases}
    \]
  – a between-class graph \( G_b \)
    ➢ the corresponding weight matrix
    \[
    W_{b,ij} = \begin{cases} 
    1 - w(y_i, y_j), & \text{if } y_i \in N_b(y_j) \text{ or } y_j \in N_b(y_i) \\
    0, & \text{otherwise.}
    \end{cases}
    \]
The Proposed Method

• Determine a “good” coding process:
  – same-label neighbors are orderly preserved
    \[ \min_x \sum_{i=1}^{N} \sum_{j=1}^{N} \| x_i - x_j \|_2^2 W_{w,ij} \]
  – neighbors with different labels are repelled
    \[ \max_x \sum_{i=1}^{N} \sum_{j=1}^{N} \| x_i - x_j \|_2^2 W_{b,ij} \]

• Rewrite the objective:
  \[ \min_x \sum_{i=1}^{N} \sum_{j=1}^{N} \| x_i - x_j \|_2^2 (W_{w,ij} - \lambda W_{b,ij}) \rightarrow \min_x \text{Tr}(X^T XL) \]
The Proposed Method

• The final objective function:

\[
\min_{D,X} \left\| Y - DX \right\|_F^2 + \alpha \text{Tr}(X^T XL) + \tau \left\| X \right\|_F^2
\]

• Alternating Optimization

– Compute \( X \) column by column with fixed \( D \)

\[
x_i^* = \arg \min_{x_i} \left\| y_i - Dx_i \right\|_2^2 + \tau \left\| x_i \right\|_2^2 + \alpha \left[ 2x_i^T(XL_i) - x_i^T x_i L_{ii} \right]
\]

– Update \( D \) with fixed \( X \)

\[
D^* = \arg \min_D \left\| Y - DX \right\|_F^2
\]

– Alternatively minimized until convergence
The Proposed Method

• Alternating Optimization

  – Compute $X$ column by column with fixed $D$

  \[
  x_i^* = \arg\min_{x_i} \| y_i - Dx_i \|^2 + \tau \| x_i \|^2 + \alpha [2x_i^T(XL_i) - x_i^T x_i L_{ii}]
  \]

  ➤ First derivative:
  \[
  2D^T(Dx_i - y_i) + 2\tau x_i + 2\alpha XL_i
  \]

  ➤ Second derivative:
  \[
  2\left[ D^TD + (\tau + \alpha L_{ii}) I \right]
  \]

  ➤ The objective function is convex for $x_i$

  – Optimal solution

  \[
  x_i^* = \left[ D^TD + (\tau + \alpha L_{ii}) I \right]^{-1} \left( D^Ty_i - \alpha \sum_{m \neq i} x_m L_{mi} \right)
  \]
The Proposed Method

• Alternating Optimization

  – Update $D$ with fixed $X$
    \[
    D^* = \arg \min_D \| Y - DX \|^2_F
    \]

  ➢ First derivative:
    \[
    2 \left( DX - Y \right) X^T
    \]
  ➢ Second derivative:
    \[
    2 \left[ I \otimes \left( XX^T \right) \right]
    \]
  ➢ The objective function is convex for $D$

  – Optimal solution
    \[
    D^* = YX^T \left( XX^T + \eta I \right)^{-1}
    \]
Outline

• Introduction
• Motivation of Our Work
• The Proposed Method
• Experimental Results
• Summarization and Future Work
Experimental Setup

• Datasets and Features
  – Extended YaleB: 504-dimensional random-face features
  – AR: 540-dimensional random-face features
  – Caltech 101: 3000-dimensional BoVW+SPM features

• Comparing Algorithms
  – the baseline support vector machine (SVM)
  – the classical SRC [PAMI 2009] and CRC [ICCV 2011]
  – the novel locality-sensitive SRC (LSRC) [PR 2013]
  – the other famous DL methods: DLSI [CVPR 2010], FDDL [ICCV 2011], LC-KSVD [PAMI 2013], DDL-PC [ACCV 2012]
  – the recently proposed LPDDL [ICIP 2014]
Results

<table>
<thead>
<tr>
<th></th>
<th>Extended YaleB</th>
<th>AR</th>
<th>Caltech101</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>95.6</td>
<td>96.5</td>
<td>64.6</td>
</tr>
<tr>
<td>SRC</td>
<td>96.5</td>
<td>97.5</td>
<td>70.7</td>
</tr>
<tr>
<td>CRC</td>
<td>97.0</td>
<td>98.0</td>
<td>68.2</td>
</tr>
<tr>
<td>LSRC</td>
<td>95.7</td>
<td>97.4</td>
<td>73.4</td>
</tr>
<tr>
<td>DLSI</td>
<td>97.0</td>
<td>97.5</td>
<td>73.1</td>
</tr>
<tr>
<td>FDDL</td>
<td>96.7</td>
<td>97.5</td>
<td>73.2</td>
</tr>
<tr>
<td>LC-KSVD</td>
<td>96.7</td>
<td>97.8</td>
<td>73.6</td>
</tr>
<tr>
<td>DDL-PC</td>
<td>95.3</td>
<td>96.0</td>
<td>73.2</td>
</tr>
<tr>
<td>LPDDL</td>
<td>96.4</td>
<td>97.3</td>
<td>73.3</td>
</tr>
<tr>
<td><strong>Ours</strong></td>
<td><strong>97.0</strong></td>
<td><strong>98.0</strong></td>
<td><strong>73.6</strong></td>
</tr>
</tbody>
</table>
Outline

• Introduction
• Motivation of Our Work
• The Proposed Method
• Experimental Results
• Summarization and Future Work
Summarization

• Conclusion
  – Performance is competitive with previous arts.
  – The locality sensitive objective function is useful to
    DDL + $k$NN classifier.

• Main Contribution
  – Preserve local relationship of same-label points and
    induce a margin between points from different classes.
  – Utilize analytical solutions in both dictionary learning
    and coding phases.
Future Work

• Reducing the consumed time for cross-validation in the training phase.

• Generalize our work to analysis dictionary learning framework:

\[
\begin{align*}
\min_{\Omega, X} & \quad \|X - \Omega Y\|_F^2 \\
\text{s.t.} & \quad \|X\|_0 \leq T_0, \\
& \quad \|\omega_i\|_2 = 1, \ i = 1, 2, \ldots
\end{align*}
\]

Thank you!

Questions please?