SHAPE-INCLUDED LABEL-CONSISTENT DISCRIMINATIVE DICTIONARY LEARNING: AN APPROACH TO DETECT AND SEGMENT MULTI-CLASS OBJECTS IN IMAGES

Authors: Mahdi Marsousi, Xingyu Li, Konstantinos Plataniotis

The Edward S. Rogers Sr. Department of Electrical & Computer Engineering

Multimedia Lab. Group Presentation
September, 2016
Introduction:

- Objective: to detect and segment objects of interest
- Applications: robotic vision, surveillance systems, …

- Requirements:
  - Independent from operator’s interaction or manual initialization
  - Flexible to add new objects
Prior art

- Sparse representation has been used for image classification.
- Label-consistent discriminative dictionary learning (LC-DDL or LC-DKSVD):

Dictionary learning problem:
\[
\langle D^P, X, A, W \rangle = \arg \min_{D, X, A, W} \sqrt{\alpha}Q - \sqrt{\beta}H X \quad \text{s.t. } \|x_i\|_0 \leq T \quad \forall i
\]

\( Q \): Label-consistency assigns dictionary atoms to classes.

- Pros: discriminating images of different classes
- Cons: not powerful for detecting objects inside images.
Contributions

1) Adding shape information in dictionary learning and sparse coding

2) A new recursive segmentation process, including the following steps:
   a. sparse coding,
   b. class-specific segmentation,
   c. refinement.
Methodology
Connection of image and patch domains

\[ P = [p_1, p_2, \ldots, p_{NP}] \]

\[ M = [m_1, m_2, \ldots, m_{NP}] \]

\[ \kappa: \text{pixel overlapping of patches} \]
\[ r: \text{patch radius} \]

Image domain \[ \Rightarrow \text{Patch domain} \]

Vectorization

\[ \text{ext}(I, \kappa, r): \text{Extract patches} \]
\[ \text{ext}(I_M, \kappa, r): \text{Extract patches} \]

\[ \text{rec}(P, \kappa, r, s_x, s_y) \]
Proposed dictionary learning

1) Creating dictionary of atoms from training patches
2) In image segmentation context: adding texture and shape prior in segmentation
3) **Shape-included** Label-consistent Discriminative Dictionary learning (SI-LC-DDL):

\[
\langle D^p, D^m, A, W \rangle = \arg\min_{D^p,D^m,X,A,W} \left\| \begin{bmatrix} P \\ \sqrt{\alpha}Q \\ \sqrt{\beta}H \end{bmatrix} - \begin{bmatrix} D^p \\ \sqrt{\alpha}A \\ \sqrt{\beta}W \end{bmatrix} X \right\|_F^2 + \lambda \| M - D^m X \|_F^2 \quad \text{s.t. } \| x_i \|_0 \leq T \quad \forall i
\]

<table>
<thead>
<tr>
<th>Patches</th>
<th>Dictionary</th>
<th>Multiplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Texture information</td>
<td>( P )</td>
<td>( D^p )</td>
</tr>
<tr>
<td><strong>Shape information</strong></td>
<td>( M )</td>
<td>( D^m )</td>
</tr>
<tr>
<td>Object class information</td>
<td>( H )</td>
<td>( W )</td>
</tr>
<tr>
<td>Label-consistency information</td>
<td>( Q )</td>
<td>( A )</td>
</tr>
</tbody>
</table>

\( H \): works like a linear classifier

\( Q \): assigns atoms to object classes

Hint: Some patches of \( M \) are either mostly zero or mostly one. Thus, mixing the shape constraint with the other ones results in imbalanced dictionary learning.
Optimization algorithm

Set $P^{new} = [P^T, \sqrt{\alpha}Q^T, \sqrt{\beta}H^T]^T$,

$D^{new} = [D^p_{(0)}^T, \sqrt{\alpha}A^T_{(0)}, \sqrt{\beta}W^T_{(0)}]^T$;

for $iter \in 1, \ldots, N_{Itr}$ do

Sparse Coding $X = OMP(P^{new}, D^{new})$;

for $k \in 1, \ldots, K$ do

Calculate: $E^p_k = (Y - \sum_{j \neq k} d^{new}_j X^R_j)$;

Hold non-zero entries of $X^R_j$: $\tilde{E}^p_k$;

Calculate: $U \Sigma V^T = SVD(\tilde{E}^p_k)$;

Update: $d^{new}_k = U(:, 1), \tilde{X}^R_k = \Sigma (1, 1) V(:, 1)$;

Calculate: $E^m_k = (M - \sum_{j \neq k} d^m_j X_j)$;

Hold non-zero entries of $X^R_j$: $\tilde{E}^m_k$;

Update: $d^m_k = \tilde{E}^m_k \times \tilde{X}^R_k$.
Dictionary initialization

**Question:** how many atoms is required for each class?

**Answer:** depends on the complexity of texture and shape prior.

**Solution:** using adaptive size dictionary learning (DLENE) to separately generate initial dictionaries for all classes of objects.

\[ cl \in \{1, 2, \cdots, N_C\} \quad \text{where } N_C : \text{the number of object classes} \]

\[
Y^{cl} = \begin{bmatrix} P^{cl} \\ M^{cl} \end{bmatrix} \xrightarrow{\forall cl} \begin{bmatrix} D^{cl,p} \\ D^{cl,m} \end{bmatrix} = DLENE(Y^{cl})
\]

\[
D^{p}_{(0)} = [D^{1,p}, D^{2,p}, \ldots, D^{N_C,p}]
\]

\[
D^{m}_{(0)} = [D^{1,m}, D^{2,m}, \ldots, D^{N_C,m}]
\]

Initial dictionaries
Contribution 2: Object detection and segmentation

**Objective:** to use learnt-based dictionaries, containing texture and shape prior information of object classes to detect and segment objects.

\[
p = \left[ p_1, p_2, \ldots, p_{N_P} \right]
\]

Only known info

Sparse coding \( OMP(P) \)

\[
X^e = \text{OMP}(P)
\]

Estimating classes & shapes

Recursive process

\[
H^{rd} = [h_1, h_2, \ldots, h_{N_P}]
\]

Refined Estimation of classes

\[
H^e = WX^e
\]

Class-Specific Segmentation

\[
I^e_{M,cl} = I^e_M \times I^e_{H,cl}
\]

Transferring into image domain

\[
\begin{align*}
I^e_M &= \text{rec}(M^e, \kappa, r, s_x, s_y) \\
I^e_{H,cl} &= \text{rec}(v_{cl}H^e, \kappa, s_x, s_y) \quad \forall cl \in \{1,2, \ldots, N_c\}
\end{align*}
\]
Experimental setup

- **Objective**: to evaluate the proposed method for
  - Single-subject segmentation
  - Multi-subject segmentation

- Method for comparison:
  - D-KSVD
  - LC-KSVD

- **Experimental setup**
  - 40 car images & 40 motorbike images from Caltech-101 database [1]
  - Hold-out evaluation protocol [2]
  - Evaluation metrics:
    - Single-subject setting: segmentation accuracy $ACC_{sg}$
    - Multi-subject setting: segmentation accuracy $ACC_{sg}$
      detection accuracy $ACC_{dt}$

[1] “Learning generative visual models from few training examples: An incremental Bayesian approach tested on 1-1 object categories”, Fei-Fei et. al., 2007.
Comparison with state-of-the-art

- Segmentation accuracy $ACC_{sg}$ in both single/multi-class setting

<table>
<thead>
<tr>
<th>Methods</th>
<th>Car</th>
<th>MotorB</th>
<th>Car&amp;MotorB</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-KSVD</td>
<td>0.859±0.102</td>
<td>0.696±0.082</td>
<td>0.814±0.112</td>
</tr>
<tr>
<td>LC-KSVD</td>
<td>0.824±0.145</td>
<td>0.728±0.068</td>
<td>0.801±0.103</td>
</tr>
<tr>
<td>SI-LC-KSVD</td>
<td>0.928±0.025</td>
<td>0.816±0.039</td>
<td>0.820±0.138</td>
</tr>
</tbody>
</table>

- Detection accuracy $ACC_{dt}$ in multi-class (Car&MotorB) setting

<table>
<thead>
<tr>
<th>Methods</th>
<th>$N_{cr}$</th>
<th>$N_{mb}$</th>
<th>$ACC_{dt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>D-KSVD</td>
<td>19</td>
<td>6</td>
<td>0.63</td>
</tr>
<tr>
<td>LC-KSVD</td>
<td>18</td>
<td>14</td>
<td>0.80</td>
</tr>
<tr>
<td>SI-LC-KSVD</td>
<td>13</td>
<td>17</td>
<td>0.75</td>
</tr>
</tbody>
</table>

Ground-truth  Segmentation First iteration  Segmentation Second iteration
Conclusion

• Summary
  ✓ A novel multi-object segmentation approach based on DL+SR
    ▪ Prior shape information included in DL
    ▪ Complete dictionary update algorithm SI-LC-KSVD
    ▪ Recursive class-specific segmentation process

• Future work
  ✓ Conducting more evaluation and analysis
  ✓ Optimal parameter setting
  ✓ Providing an online-DL process to be used in self-training of robots.
Thank you!