3D Detection from D-RGB data

Presented by
Caleb Jordan
Olga Russakovsky
Outline

• Generalized view of feature kernels
  – L. Bo, X. Ren and D. Fox “Kernel descriptors for visual recognition” NIPS 2010
• Large-scale RGBD dataset
  – K. Lai, L. Bo, X. Ren and D. Fox “A Large-scale hierarchical multi-view RGB-D object detection” ICRA 2010
• Applying hierarchical feature kernels to RGBD data
  – L. Bo, K. Lai, X. Ren and D. Fox “Object recognition with hierarchical kernel descriptors” CVPR 2011
Outline

• Generalized view of feature kernels
  – L. Bo, X. Ren and D. Fox “Kernel descriptors for visual recognition” NIPS 2010

• Large-scale RGBD dataset
  – K. Lai, L. Bo, X. Ren and D. Fox “A Large-scale hierarchical multi-view RGB-D object detection” ICRA 2010

• Applying hierarchical feature kernels to RGBD data
  – L. Bo, K. Lai, X. Ren and D. Fox “Object recognition with hierarchical kernel descriptors” CVPR 2011
Kernel descriptors for visual recognition (NIPS ‘10)

• Kernel view of low-level image descriptors based on orientation histograms (HOG, SIFT)
  – plus cool tricks and approximations
• Features validated on RGB image classification
  – Scene-15, Caltech-101, CIFAR10, CIFAR10-ImageNet
Prior work: orientation histograms (SIFT, HOG)
Prior work: orientation histograms (SIFT, HOG)
Prior work: orientation histograms (SIFT, HOG)

\[ F_h(P) = \sum_{z \in P} \tilde{m}(z)\delta(z) \]

- **Image patch**
- **Patch P**
- **Gradient vectors**
- **Features** $F_h(P)$
- **SIFT or HOG**
- **d gradient orientation bins**
- **Sum over pixels**
- **d-dimensional indicator vector for d gradient orientation bins**
- **Gradient magnitude**

**Features**
- Patch P
- Sum over pixels
- Gradient magnitude
- d-dimensional indicator vector for d gradient orientation bins

**Prior work:** orientation histograms (SIFT, HOG)
Kernelizing orientation histograms

Image patch → gradient vectors → Features $F_h(P)$ → d gradient orientation bins

Similarity between patch $P$ and patch $Q$:

$$K_h(P, Q) = F_h(P)^\top F_h(Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}(z) \tilde{m}(z') \delta(z)^\top \delta(z')$$

Gradient magnitude

Orientation
Kernelizing orientation histograms

Similarity between patch $P$ and patch $Q$

\[
K_{\text{grad}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}(z) \tilde{m}(z') k_o(\tilde{\theta}(z), \tilde{\theta}(z')) k_p(z, z')
\]

Gradient magnitude

Orientation

Sum over every pair of pixels

$d$ gradient orientation bins

Patch $P$

Features $F_h(P)$

Image patch

gradient vectors

SIFT or HOG
Kernelizing orientation histograms

\[ K_{\text{grad}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}(z)\tilde{m}(z')k_o(\tilde{\theta}(z), \tilde{\theta}(z'))k_p(z, z') \]

- **Image patch**
- **Patch P**
- **Gradient vectors**
- **Features** \( F_h(P) \)
- **d gradient orientation bins**
- **SIFT or HOG**
- **Position**
- **Orientation**
- **Similarity between patch P and patch Q**
- **Sum over every pair of pixels**
- **Gradient magnitude**
Pixel-level similarity measures

$$k_p(z, z') = \exp(-\gamma_p \| z - z' \|^2)$$
Pixel-level similarity measures

\[ k_p(z, z') = \exp(-\gamma_p \| z - z' \|^2) \]

\[ k_o(\tilde{\theta}(z), \tilde{\theta}(z')) = \exp(-\gamma_o \| \tilde{\theta}(z) - \tilde{\theta}(z') \|^2) \]
Pixel-level similarity measures

$$k_p(z, z') = \exp(-\gamma_p \| z - z' \|^2)$$

$$k_o(\tilde{\theta}(z), \tilde{\theta}(z')) = \exp(-\gamma_o \| \tilde{\theta}(z) - \tilde{\theta}(z') \|^2)$$

$$\tilde{\theta} = (\sin(\theta), \cos(\theta))$$
Pixel-level similarity measures

\[ k_p(z, z') = \exp(-\gamma_p \|z - z'\|^2) \]

\[ k_o(\tilde{\theta}(z), \tilde{\theta}(z')) = \exp(-\gamma_o \|\tilde{\theta}(z) - \tilde{\theta}(z')\|^2) \]
Pixel-level similarity measures

\[ k_p(z, z') = \exp(-\gamma_p \|z - z'\|^2) \]

\[ k_o(\tilde{\theta}(z), \tilde{\theta}(z')) = \exp(-\gamma_o \|\tilde{\theta}(z) - \tilde{\theta}(z')\|^2) \]

\[ k_c(c(z), c(z')) = \exp(-\gamma_c \|c(z) - c(z')\|^2) \]
Pixel-level similarity measures

\[ k_p(z, z') = \exp(-\gamma_p \|z - z'\|^2) \]
\[ k_o(\tilde{\theta}(z), \tilde{\theta}(z')) = \exp(-\gamma_o \|\tilde{\theta}(z) - \tilde{\theta}(z')\|^2) \]
\[ k_c(c(z), c(z')) = \exp(-\gamma_c \|c(z) - c(z')\|^2) \]
\[ k_b(b(z), b(z')) = \exp(-\gamma_b \|b(z) - b(z')\|^2) \]
Approximating pixel-level similarity measures

$$k_o(\tilde{\theta}(z), \tilde{\theta}(z')) = \exp(-\gamma_o \|\tilde{\theta}(z) - \tilde{\theta}(z')\|^2)$$
Approximating pixel-level similarity measures

$$k_o(\tilde{\theta}(z), \tilde{\theta}(z')) = \exp(-\gamma_o \| \tilde{\theta}(z) - \tilde{\theta}(z') \|^2)$$

$$= \phi_o(\tilde{\theta}(z))^\top \phi_o(\tilde{\theta}(z'))$$
Approximating
pixel-level similarity measures

\[ k_o(\tilde{\theta}(z), \tilde{\theta}(z')) = \exp(-\gamma_o \|\tilde{\theta}(z) - \tilde{\theta}(z')\|^2) \]
\[ = \phi_o(\tilde{\theta}(z))^\top \phi_o(\tilde{\theta}(z')) \]
Pixel-level similarity measures

\[ k_p(z, z') \]

\[ k_o(\tilde{\theta}(z), \tilde{\theta}(z')) \]

\[ k_c(c(z), c(z')) \]

\[ k_b(b(z), b(z')) \]
Patch-level similarity measures

\[ K_{\text{grad}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}(z) \tilde{m}(z') k_o(\tilde{\theta}(z), \tilde{\theta}(z')) k_p(z, z') \]

\[ K_{\text{col}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} k_c(c(z), c(z')) k_p(z, z') \]

\[ K_{\text{shape}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{s}(z) \tilde{s}(z') k_b(b(z), b(z')) k_p(z, z') \]
Approximating patch-level similarity measures

\[ K_{\text{grad}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} m(z)m(z')k_o(\tilde{\theta}(z), \tilde{\theta}(z'))k_p(z, z') \]
Approximating patch-level similarity measures

\[ K_{\text{grad}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}(z) \tilde{m}(z') k_{o}(\tilde{\theta}(z), \tilde{\theta}(z')) k_{p}(z, z') \]

\[ = F_{\text{grad}}(P)^\top F_{\text{grad}}(Q) \]
Approximating patch-level similarity measures

\[ K_{\text{grad}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}(z)\tilde{m}(z') k_o(\tilde{\theta}(z), \tilde{\theta}(z')) k_p(z, z') \]

\[ = F_{\text{grad}}(P)^\top F_{\text{grad}}(Q) \]

\[ k_o(\tilde{\theta}(z), \tilde{\theta}(z')) = \phi_o(\tilde{\theta}(z))^\top \phi_o(\tilde{\theta}(z')) \]

\[ k_p(z, z') = \phi_p(z)^\top \phi_p(z') \]
Approximating patch-level similarity measures

\[
K_{\text{grad}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}(z)\tilde{m}(z') k_o(\tilde{\theta}(z), \tilde{\theta}(z')) k_p(z, z')
\]

\[
= F_{\text{grad}}(P)^\top F_{\text{grad}}(Q)
\]

\[
k_o(\tilde{\theta}(z), \tilde{\theta}(z')) = \phi_o(\tilde{\theta}(z))^\top \phi_o(\tilde{\theta}(z'))
\]

\[
k_p(z, z') = \phi_p(z)^\top \phi_p(z')
\]

\[
F_{\text{grad}}(P) = \sum_{z \in P} \tilde{m}(z) \phi_o(\tilde{\theta}(z)) \otimes \phi_p(z)
\]
Approximating patch-level similarity measures

\[ K_{\text{grad}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}(z) \tilde{m}(z') k_o(\tilde{\theta}(z), \tilde{\theta}(z')) k_p(z, z') \]

\[ = F_{\text{grad}}(P)^\top F_{\text{grad}}(Q) \]

\[ k_o(\tilde{\theta}(z), \tilde{\theta}(z')) = \phi_o(\tilde{\theta}(z))^\top \phi_o(\tilde{\theta}(z')) \]

\[ k_p(z, z') = \phi_p(z)^\top \phi_p(z') \]

\[ F_{\text{grad}}(P) = \sum_{z \in P} \tilde{m}(z) \phi_o(\tilde{\theta}(z)) \otimes \phi_p(z) \]

... KPCA fancy math \textbf{(with code online!)}
Approximating patch-level similarity measures
Efficiency (in 2010)

On 300x300 image with MATLAB CPU implementation:

- SIFT: 0.4s
- Gradient kernel: 1.5s
- Shape kernel: 4s
Experiments (on RGB image classification)

- Free parameters in kernel descriptors are optimized on a subset of ImageNet.
- The resulting values are fixed in the following experiments.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Scene-15</th>
<th>Caltech-101</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KDES:</td>
<td>KDES:</td>
</tr>
<tr>
<td></td>
<td>86.7%</td>
<td>76.4%</td>
</tr>
<tr>
<td></td>
<td>SIFT:</td>
<td>SPM\textsuperscript{[1]}:</td>
</tr>
<tr>
<td></td>
<td>82.2%</td>
<td>64.4%</td>
</tr>
<tr>
<td>CIFAR10</td>
<td>mcRBM-DBN\textsuperscript{[3]}:</td>
<td>76.0%</td>
</tr>
<tr>
<td></td>
<td>71.0%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TCNN\textsuperscript{[5]}:</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{[1]} Lazebnik, Schmid, Ponce, CVPR ‘06
\textsuperscript{[2]} Lee, Grosse, Ranganath, Ng, ICML ‘09
\textsuperscript{[3]} Ranzato, Hinton, CVPR ‘10
\textsuperscript{[4]} Yu, Zhang, ICML ‘10
\textsuperscript{[5]} Le, Ngiam, Chen, Chia, Koh, Ng, NIPS ‘10
Kernel descriptors for visual recognition (NIPS ‘10)

• Kernel view of low-level image descriptors based on orientation histograms (HOG, SIFT)
  – plus cool tricks and approximations
• Features validated on RGB image classification
  – Scene-15, Caltech-101, CIFAR10, CIFAR10-ImageNet
• Psst: code is online
Outline

• Generalized view of feature kernels
  – L. Bo, X. Ren and D. Fox “Kernel descriptors for visual recognition” NIPS 2010

• Large-scale RGBD dataset
  – K. Lai, L. Bo, X. Ren and D. Fox “A Large-scale hierarchical multi-view RGB-D object detection” ICRA 2010

• Applying hierarchical feature kernels to RGBD data
  – L. Bo, K. Lai, X. Ren and D. Fox “Object recognition with hierarchical kernel descriptors” CVPR 2011
A new RGB-D dataset

• RGB-D – RGB + depth values per frame
• 300 distinct objects, video sequence of ~250 frames for full revolutions at 30, 45, 60 degree elevation.

• (Previous was 3D Object Category Dataset by Savarese et al: 8 categories x 10 objects x 24 distinct views.)

http://www.cs.washington.edu/rgbd-dataset
Object categories

![Image of various object categories](image-url)
Category hierarchies

51 leaf nodes, 3-14 object instances per leaf category
RGB-D extras

- Images segmented using depth and color information

Segmented using RGB
Using depth
Combined
Outline

• Generalized view of feature kernels
  – L. Bo, X. Ren and D. Fox “Kernel descriptors for visual recognition” NIPS 2010

• Large-scale RGBD dataset
  – K. Lai, L. Bo, X. Ren and D. Fox “A Large-scale hierarchical multi-view RGB-D object detection” ICRA 2010

• Applying hierarchical feature kernels to RGBD data
  – L. Bo, K. Lai, X. Ren and D. Fox “Object recognition with hierarchical kernel descriptors” CVPR 2011
Hierarchical Kernel Descriptors (HKDES)

• Applies the concept of kernel descriptors recursively to create image-level features from pixel-level features.
Related Work

• Models with hierarchies or levels: deep belief nets, convolutional deep belief nets, convolutional neural networks
• Also sparse coding, spatial pyramid
• Parallel approach works on optimizing low-level match kernels
Recall first level kernels

• Gradient kernel

\[ K_{\text{grad}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}_z \tilde{m}_{z'} k_o(\tilde{\theta}_z, \tilde{\theta}_{z'}) k_p(z, z') \]

• Color kernel

\[ K_{\text{col}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} k_c(c_z, c_{z'}) k_p(z, z') \]
Add a second level kernel

- Gradient kernel
  \[ K_{\text{grad}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}_z \tilde{m}_{z'} k_o(\tilde{\theta}_z, \tilde{\theta}_{z'}) k_p(z, z') \]

- Color kernel
  \[ K_{\text{col}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} k_c(c_z, c_{z'}) k_p(z, z') \]

- Patch-level color + gradient kernel
  \[ K(P, Q) = \sum_{A \in P} \sum_{A' \in Q} \tilde{W}_A \tilde{W}_{A'} k_F(F_A, F_{A'}) k_C(C_A, C_{A'}) \]

Kernel over patch-level gradient features
Kernel over patch-level color features
Results for CIFAR-10

HKDES results for RGB data on CIFAR-10 dataset

<table>
<thead>
<tr>
<th>Features</th>
<th>KDES [1]</th>
<th>HKDES (this work)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color</td>
<td>53.9</td>
<td>63.4</td>
</tr>
<tr>
<td>Shape</td>
<td>68.2</td>
<td>69.4</td>
</tr>
<tr>
<td>Gradient</td>
<td>66.3</td>
<td>71.2</td>
</tr>
<tr>
<td>Combination</td>
<td>76.0</td>
<td>80.0</td>
</tr>
</tbody>
</table>

Comparison to non-hierarchical

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic regression [25]</td>
<td>36.0</td>
</tr>
<tr>
<td>GIST [25]</td>
<td>54.7</td>
</tr>
<tr>
<td>SIFT [1]</td>
<td>65.6</td>
</tr>
<tr>
<td>fine-tuning GRBM [24]</td>
<td>64.8</td>
</tr>
<tr>
<td>GRBM two layers [24]</td>
<td>56.6</td>
</tr>
<tr>
<td>mcRBM [25]</td>
<td>68.3</td>
</tr>
<tr>
<td>mcRBM-DBN [25]</td>
<td>71.0</td>
</tr>
<tr>
<td>Tiled CNNs [16]</td>
<td>73.1</td>
</tr>
<tr>
<td>improved LCC [31]</td>
<td>74.5</td>
</tr>
<tr>
<td>KDES + EMK + linear SVMs [1]</td>
<td>76.0</td>
</tr>
<tr>
<td>Convolutional RBM [4]</td>
<td>78.9</td>
</tr>
<tr>
<td>K-means (Triangle, 4k features) [4]</td>
<td>79.6</td>
</tr>
<tr>
<td>HKDES + linear SVMs (this work)</td>
<td>80.0</td>
</tr>
</tbody>
</table>

Comparison to state-of-the-art
Results for RGB-D!

Comparing RGB to depth for HKDES

<table>
<thead>
<tr>
<th>Method</th>
<th>Category</th>
<th>Instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color HKDES (RGB)</td>
<td>60.1±2.1</td>
<td>58.4</td>
</tr>
<tr>
<td>Shape HKDES (RGB)</td>
<td>72.6±1.9</td>
<td>74.6</td>
</tr>
<tr>
<td>Gradient HKDES (RGB)</td>
<td>70.1±2.9</td>
<td>75.9</td>
</tr>
<tr>
<td>Combination of HKDES (RGB)</td>
<td>76.1±2.2</td>
<td>79.3</td>
</tr>
<tr>
<td>Color HKDES (depth)</td>
<td>61.8±2.4</td>
<td>28.8</td>
</tr>
<tr>
<td>Shape HKDES (depth)</td>
<td>65.8±1.8</td>
<td>36.7</td>
</tr>
<tr>
<td>Gradient HKDES (depth)</td>
<td>70.8±2.7</td>
<td>39.3</td>
</tr>
<tr>
<td>Combination of HKDES (depth)</td>
<td>75.7±2.6</td>
<td>46.8</td>
</tr>
<tr>
<td>Combination of all HKDES</td>
<td>84.1±2.2</td>
<td>82.4</td>
</tr>
</tbody>
</table>

Comparing other approaches using a combination of color and depth information

<table>
<thead>
<tr>
<th>Approaches</th>
<th>Category</th>
<th>Instance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear SVMs [15]</td>
<td>81.9±2.8</td>
<td>73.9</td>
</tr>
<tr>
<td>Nonlinear SVMs [15]</td>
<td>83.8±3.5</td>
<td>74.8</td>
</tr>
<tr>
<td>Random Forest [15]</td>
<td>79.6±4.0</td>
<td>73.1</td>
</tr>
<tr>
<td>Combination of all HKDES</td>
<td>84.1±2.2</td>
<td>82.4</td>
</tr>
</tbody>
</table>
Outline

• Generalized view of feature kernels
  – L. Bo, X. Ren and D. Fox “Kernel descriptors for visual recognition” NIPS 2010

• Large-scale RGBD dataset
  – K. Lai, L. Bo, X. Ren and D. Fox “A Large-scale hierarchical multi-view RGB-D object detection” ICRA 2010

• Applying hierarchical feature kernels to RGBD data
  – L. Bo, K. Lai, X. Ren and D. Fox “Object recognition with hierarchical kernel descriptors” CVPR 2011