Object Detection by 3D Aspectlets and Occlusion Reasoning

Yu Xiang
University of Michigan

Silvio Savarese
Stanford University

In the 4th International IEEE Workshop on 3D Representation and Recognition (3dRR), 2013.
Occlusions in Object Detection

Yu Xiang and Silvio Savarese. 3dRR'13
Object Context for Occlusion Reasoning

Consider all the objects in the scene jointly by estimating their 3D spatial layout.
3D Parts for Handling Occlusion

3D Parts provides evidences of partial observations from different views.

Yu Xiang and Silvio Savarese. 3dRR'13
Our Method

Yu Xiang and Silvio Savarese. 3dRR'13
Our Method

• Top-down occlusion reasoning by contextualizing objects in 3D

• Bottom-up evidences provided by part-based 3D object detectors (3D Aspectlets).
Outline

• Related work

• 3D aspectlets

• Spatial layout model

• Experiments

• Conclusion
Outline

• Related work

• 3D aspectlets

• Spatial layout model

• Experiments

• Conclusion
Related Work: 3D Object Detection

• Use 3D models, learn object appearances from training images for robust 2D matching

• Hard to handle complicated scenes with occlusions and truncations

From Liebelt et al. CVPR’08

From Pepik et al. CVPR’12

From Xiang & Savarese, CVPR’12

From Fidler et al., NIPS’12
Related Work: Object Context for Object Detection

Hoiem et al. use 3D scene geometry, CVPR’06

Desai et al. use object co-occurrences, ICCV’09

Hedau et al. use room layout, ECCV’10

Choi et al. use geometry phases, CVPR’13
Related work: 2D Occlusion Reasoning

- Occlusion boundaries recovery, Hoiem et al. ICCV’07
- HOG-LBP human detector, Wang et al. ICCV’09
- Segmentation-aware detector, Gao et al. CVPR’11
- Occlusion Patterns, Pepik et al. CVPR’13
- Occlusion masks, Zia et al. CVPR’13
Related work: 3D occlusion reasoning in object detection

Wu and Nevatia, ICCV’05

Wojek et al. CVPR’11

Difference:
Instead of a simplified 2.5D structure of depth layers, we handle occlusion using a true 3D representation of object.
Outline

• Related work

• 3D aspectlets

• Spatial layout model

• Experiments

• Conclusion
3D Aspectlet

• Aspect part [Xiang & Savarese, CVPR’12] is good at handling self-occlusion, but not good for occlusion between objects
3D Aspectlet

• Atomic aspect part for handling occlusion

[Xiang & Savarese, 3dRR’13]
3D Aspectlets

- Atomic aspect parts are hard to detect, group them to form “bigger parts” – 3D aspectlets
  - Geometrically close to each other in 3D
  - Discriminative
3D Aspectlets

Yu Xiang and Silvio Savarese. 3dRR'13
3D Aspectlets

• Each 3D aspectlet is modeled by a two level tree structure as in Aspect Layout Model [Xiang & Savarese, CVPR’12]
Outline

• Related work

• 3D aspectlets

• Spatial layout model

• Experiments

• Conclusion
Spatial Layout Model

• Posterior distribution

\[ P(o, O, C \mid I) \]

\[ \propto P(C)P(O)\prod_{i=1}^{M} P(o_i \mid O, C, I) \prod_{(i, j)} P(o_i, o_j \mid O, C, I) \]

- camera prior
- 3D objects prior
- unary 2D projection likelihood
- pairwise 2D projection likelihood

Yu Xiang and Silvio Savarese. 3dRR'13
Spatial Layout Model

• Camera prior
  – Virtual intrinsic camera matrix

\[ P(C) = P(a)P(e)P(d) \]

Yu Xiang and Silvio Savarese. 3dRR'13
Spatial Layout Model

• 3D objects prior

\[ P(O) \propto \exp \left( \sum_{i=1}^{M} V_1(O_i) + \sum_{(i,j)} V_2(O_i, O_j) \right) \]

– “ground plane” constraint

\[ V_1(O_i) = -\frac{Z_i^2}{2\sigma^2} \]

– 3D space constraint

\[ V_2(O_i, O_j) = -\rho \frac{O_i \cap O_j}{O_i \cup O_j} \]
Spatial Layout Model

• Unary 2D projection likelihood

\[ P(o_i \mid O, C, I) \propto P_0(o_i \mid O_i, C, I) + \sum_{k=1}^{N} w_k(O, C) P_k(o_i \mid O_i, C, I) \]

s.t. \[ \sum_{k=1}^{N} w_k(O, C) = 1 \]

weights proportional to the number of visible atomic aspect parts
Spatial Layout Model

- Pairwise 2D projection likelihood
  - Penalizes wrong occlusion order
  - Reduces false alarms

\[ P(o_i, o_j \mid O, C, I) \propto \exp\left(- \frac{P(o_j \mid O, C, I)}{P(o_i \mid O, C, I)}\right) \]

if \( O_i \) occludes \( O_j \) and \( P(o_i \mid O, C, I) > \text{threshold} \)

Yu Xiang and Silvio Savarese. 3dRR’13
Spatial Layout Model

• Training
  – Unsupervised learning for selecting 3D aspectlets
  – Structural SVM for parameter estimation of 3D aspectlets

• Inference
  – RJMCMC sampling
  – Object hypotheses from unary 2D projection likelihood without occlusion reasoning
  – Add moves, delete moves, switch moves
  – Log-odds ratios from MAP as 2D detection scores
Outline

• Related work

• 3D aspectlets

• Spatial layout model

• Experiments

• Conclusion
Training Datasets

• Car: 3DObject Dataset [Savarese & Fei-Fei, ICCV’07]

• Bed, Chair, Sofa and Table: Subset of ImageNet Dataset [Xiang & Savarese, CVPR’12]

Yu Xiang and Silvio Savarese. 3dRR'13
Test Datasets

Two new datasets with occlusion (online)
- An outdoor-scene dataset with cars (200 images)
- An indoor-scene dataset with beds, chairs, sofas and tables (300 images)

<table>
<thead>
<tr>
<th>Category</th>
<th>Car</th>
<th>Bed</th>
<th>Chair</th>
<th>Sofa</th>
<th>Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>#objects</td>
<td>659</td>
<td>202</td>
<td>235</td>
<td>273</td>
<td>222</td>
</tr>
<tr>
<td>#occluded</td>
<td>235</td>
<td>81</td>
<td>112</td>
<td>175</td>
<td>61</td>
</tr>
<tr>
<td>#truncated</td>
<td>135</td>
<td>86</td>
<td>41</td>
<td>99</td>
<td>80</td>
</tr>
</tbody>
</table>

Yu Xiang and Silvio Savarese. 3dRR'13
## Detection APs

<table>
<thead>
<tr>
<th>Category</th>
<th>Car</th>
<th>Bed</th>
<th>Chair</th>
<th>Sofa</th>
<th>Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALM [1]</td>
<td>46.6</td>
<td>28.9</td>
<td>14.2</td>
<td>41.1</td>
<td>19.2</td>
</tr>
<tr>
<td>DPM [2]</td>
<td>57.0</td>
<td>34.8</td>
<td>14.4</td>
<td>38.3</td>
<td>15.1</td>
</tr>
<tr>
<td>SLM Aspectlets</td>
<td>59.2</td>
<td>35.8</td>
<td>15.9</td>
<td>45.5</td>
<td>24.3</td>
</tr>
<tr>
<td>SLM Full</td>
<td>63.0</td>
<td>39.1</td>
<td>19.0</td>
<td>48.6</td>
<td>28.6</td>
</tr>
</tbody>
</table>

SLM Aspectlets: using 3D aspectlets in Hough voting without occlusion reasoning
SLM Full: our full model using 3D aspectlets and occlusion reasoning

Detection APs

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Outdoor-scene</th>
<th></th>
<th>Indoor-scene</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% occlusion</td>
<td>&lt; 0.3</td>
<td>0.3 – 0.6</td>
<td>&gt; 0.6</td>
</tr>
<tr>
<td># images</td>
<td>66</td>
<td>68</td>
<td>66</td>
<td>77</td>
</tr>
<tr>
<td>ALM [1]</td>
<td>72.3</td>
<td>42.9</td>
<td>35.5</td>
<td>38.5</td>
</tr>
<tr>
<td>DPM [2]</td>
<td>75.9</td>
<td>58.6</td>
<td>44.6</td>
<td>38.0</td>
</tr>
<tr>
<td>SLM Aspectlets</td>
<td>78.7</td>
<td>59.7</td>
<td>47.7</td>
<td>41.9</td>
</tr>
<tr>
<td>SLM Full</td>
<td>80.2</td>
<td>63.3</td>
<td>52.9</td>
<td>45.9</td>
</tr>
</tbody>
</table>

% occlusion: percentage of occluded area of the object computed from ground truth annotation.

3D Localization Evaluation

Ground truth 2D annotations

2D object detections

Mean absolute difference in pairwise distances

Ground truth 3D spatial layout

Predicted 3D spatial layout
3D Localization

• 3D localization errors on the outdoor-scene dataset according to the best recalls of ALM, DPM and SLM.

<table>
<thead>
<tr>
<th></th>
<th>Recall</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ALM [1]</td>
<td>1.90</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DPM [2]</td>
<td>2.07</td>
<td>2.39</td>
<td>-</td>
</tr>
<tr>
<td>SLM</td>
<td>1.64</td>
<td>1.86</td>
<td>2.33</td>
</tr>
</tbody>
</table>

Anecdotal Results

Yu Xiang and Silvio Savarese. 3dRR’13
Anecdotal Results

Yu Xiang and Silvio Savarese. 3dRR'13
Outline

• Related work

• 3D aspectlets

• Spatial layout model

• Experiments

• Conclusion
Conclusion

• 3D object representation
  – Atomic aspect part
  – 3D aspectlet

• 3D object recognition
  – Spatial Layout Model (SLM)
  – Top-down occlusion reasoning
  – Bottom-up evidence from 3D aspectlets
3DObjectPose: Large Scale Dataset with 3D Poses

- 12 rigid categories in PASCAL VOC
- PASCAL VOC 2012 train and validation set
- Subset of ImageNet
- 8505 images
- 22394 images
- Annotations:
  - Bounding box
  - Pose: azimuth and elevation
  - 3D CAD model
  - Anchor points

Yu Xiang and Silvio Savarese. 3dRR'13
Thank you!

Yu Xiang and Silvio Savarese. 3dRR'13