Computational Vision and Geometry Lab

Object Detection by 3D Aspectlets and Occlusion Reasoning

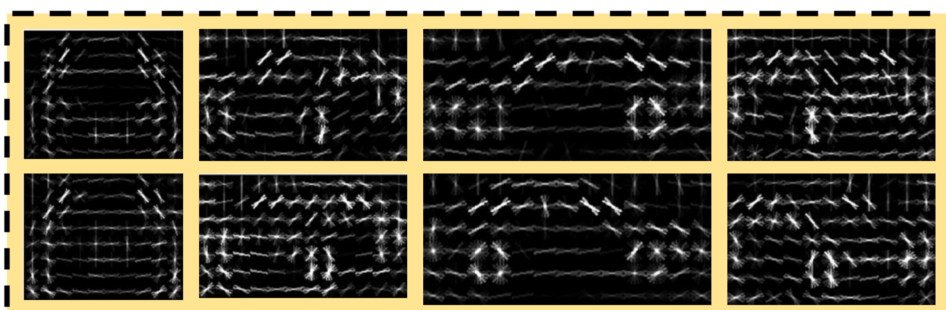
Yu Xiang
University of Michigan

Silvio Savarese
Stanford University

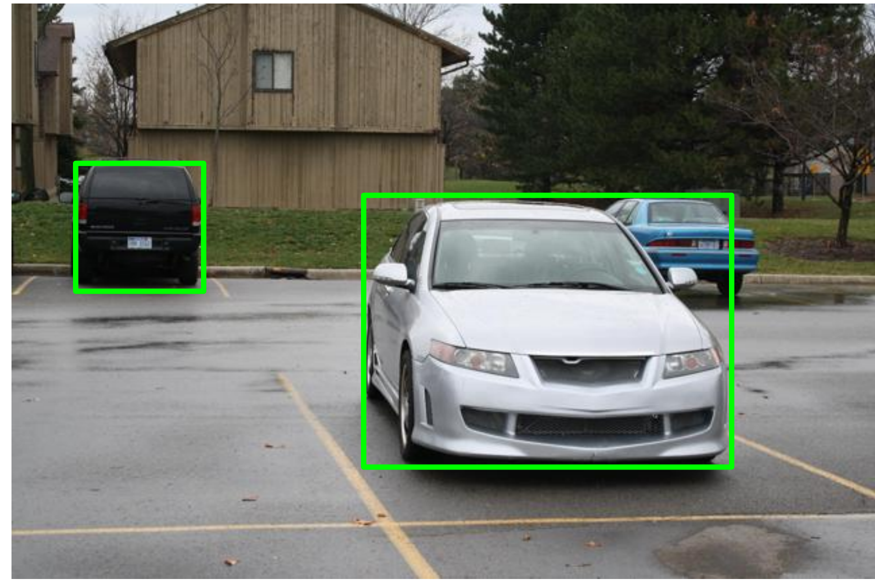
In the 4th International IEEE Workshop on 3D Representation and Recognition (**3dRR**), 2013.

Occlusions in Object Detection

training images



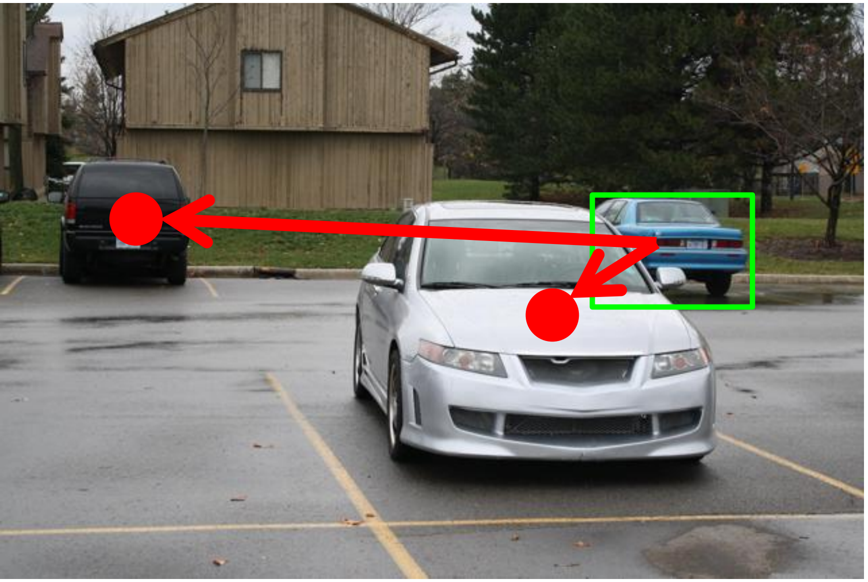
car mixture detector



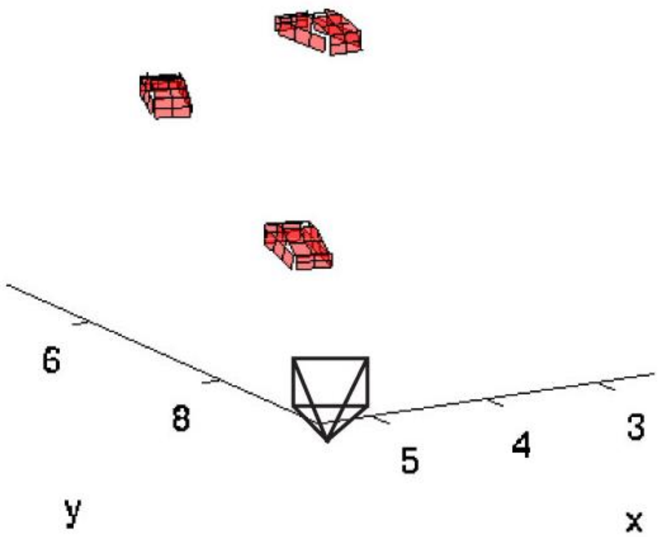
test image

Occlusion changes the appearances of objects.

Object Context for Occlusion Reasoning



test image

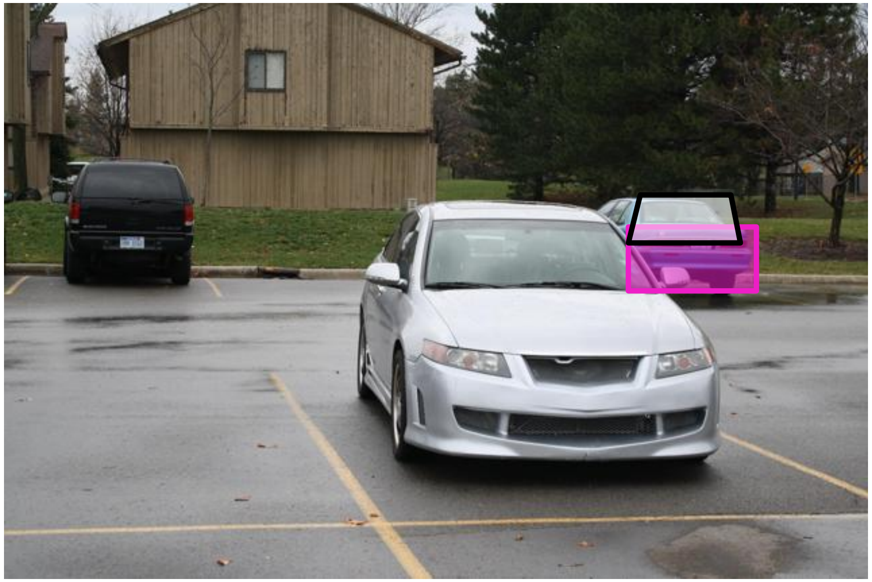


3D spatial layout

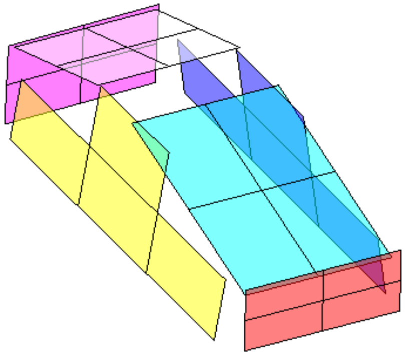
Consider all the objects in the scene jointly by estimating their 3D spatial layout.

3D Parts for Handling Occlusion

training images



test image

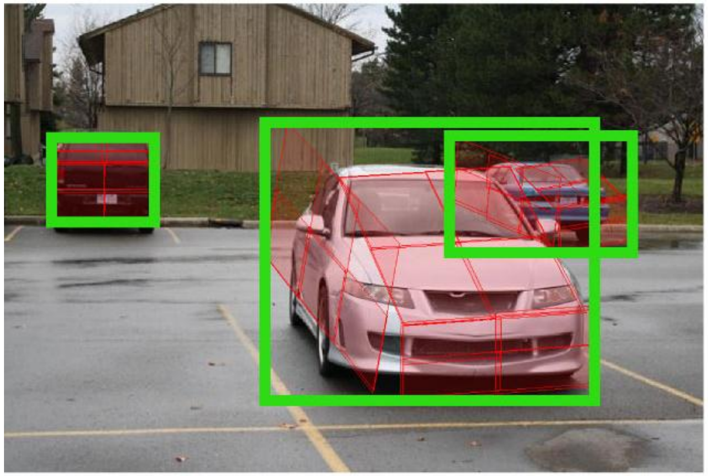


3D object detector

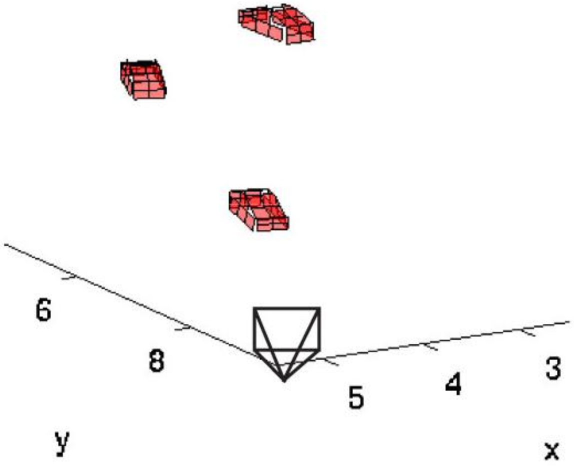
3D Parts provides evidences of partial observations from different views.

Our Method

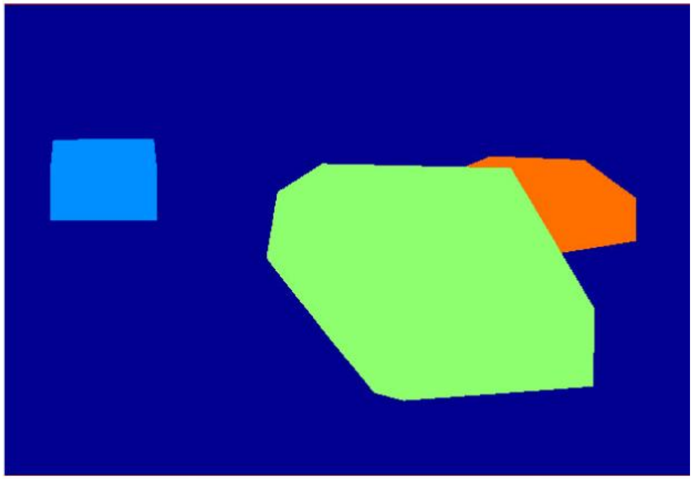
(a) input image



(b) 2D detection



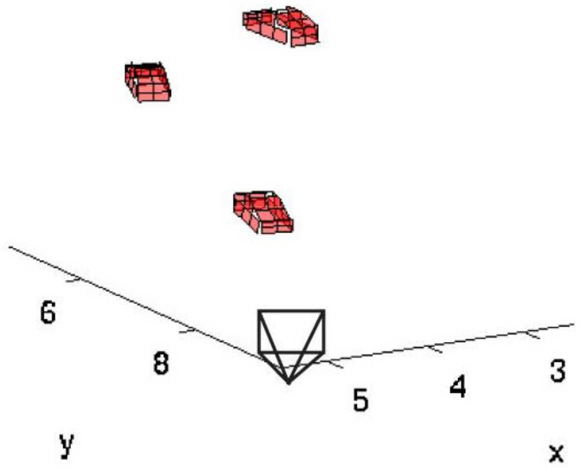
(c) 3D spatial layout



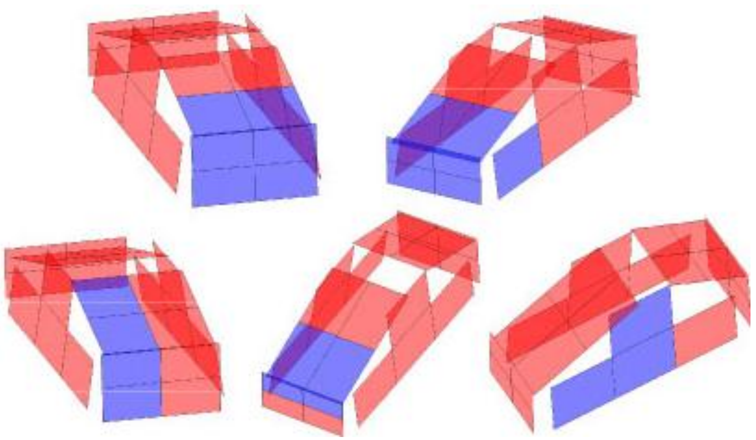
(d) 2D object mask

Our Method

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



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



Outline

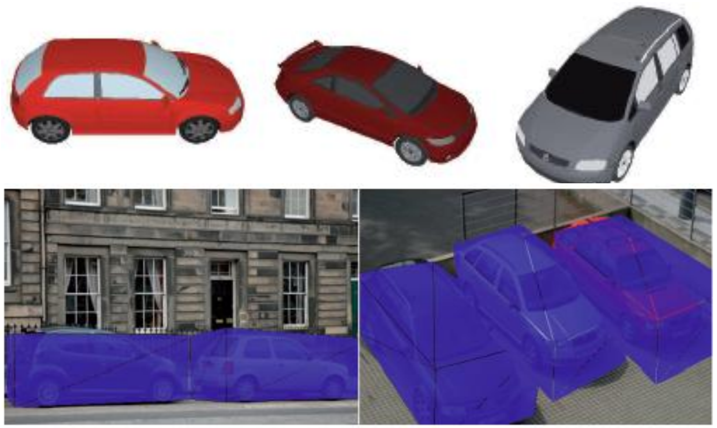
- Related work
- 3D aspectlets
- Spatial layout model
- Experiments
- Conclusion

Outline

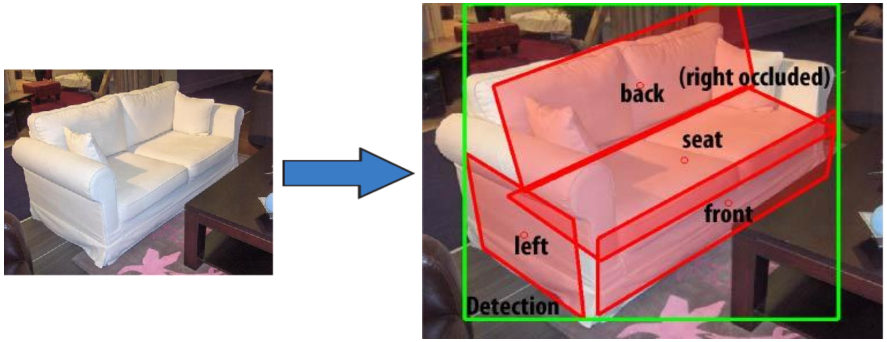
- Related work
- 3D aspectlets
- Spatial layout model
- Experiments
- Conclusion

Related Work: 3D Object Detection

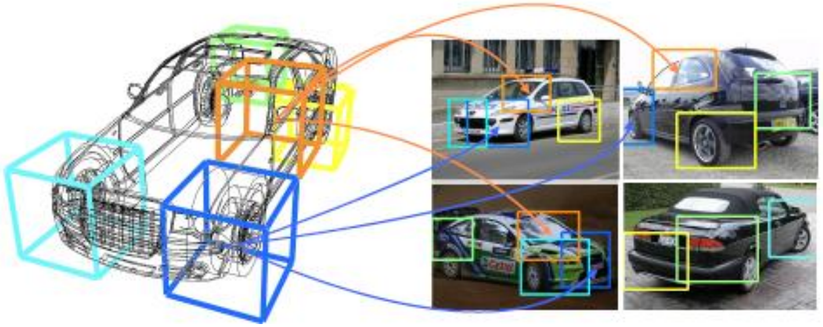
- Use 3D models, learn object appearances from training images for robust 2D matching
- Hard to handle complicated scenes with occlusions and truncations



From Liebelt et al. CVPR'08



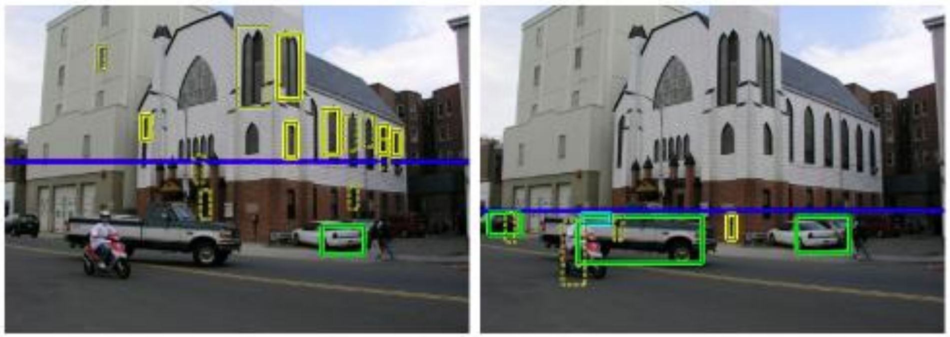
From Xiang & Savarese, CVPR'12



From Pepik et al. CVPR'12

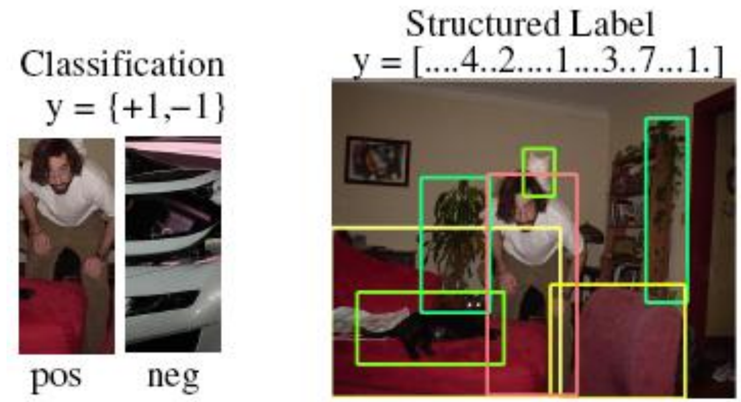
From Fidler et al., NIPS'12

Related Work: Object Context for Object Detection

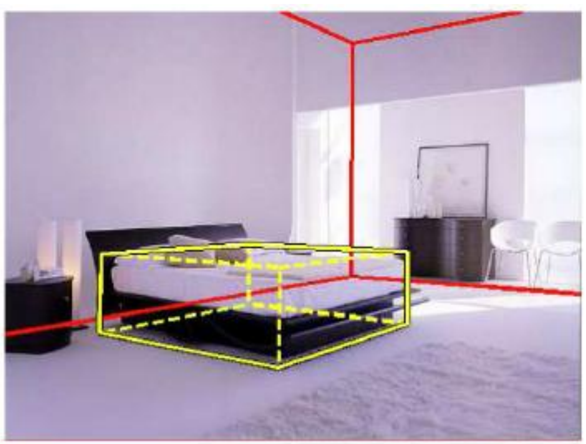


(b) Local Detection (b) Full Model Detection

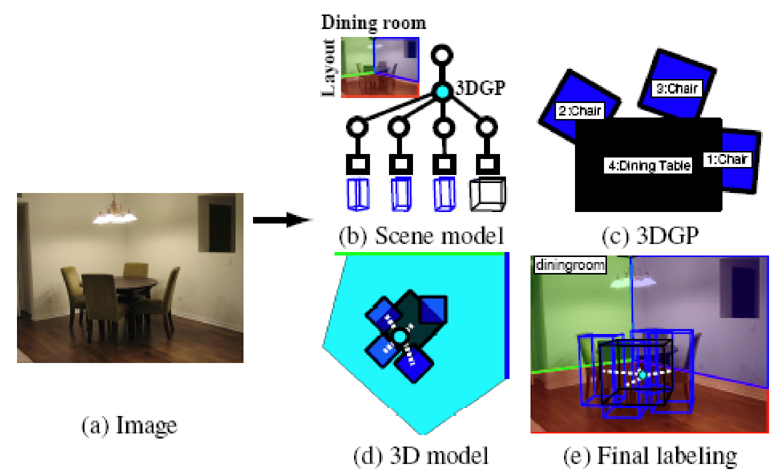
Hoiem et al. use 3D scene geometry, CVPR'06



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



Hedau et al. use room layout, ECCV'10

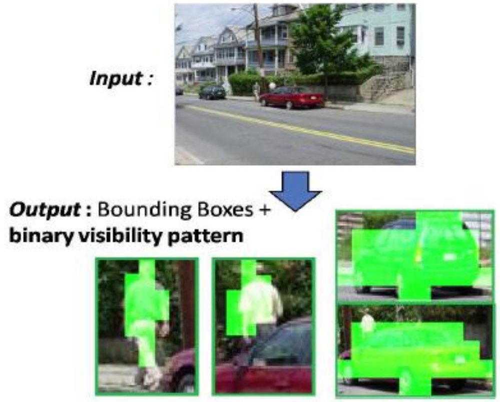


Choi et al. use geometry phases, CVPR'13

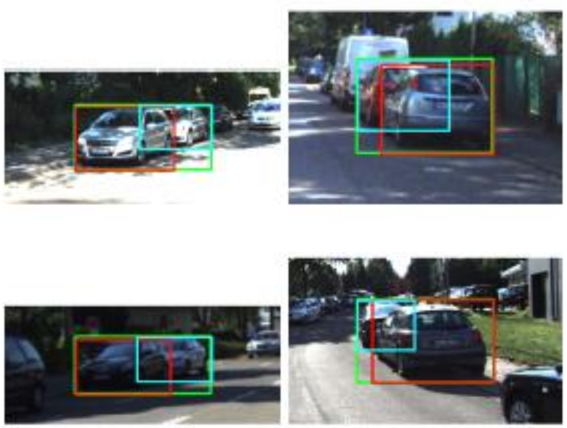
Related work: 2D Occlusion Reasoning

Occlusion boundaries recovery,
Hoiem et al. ICCV'07

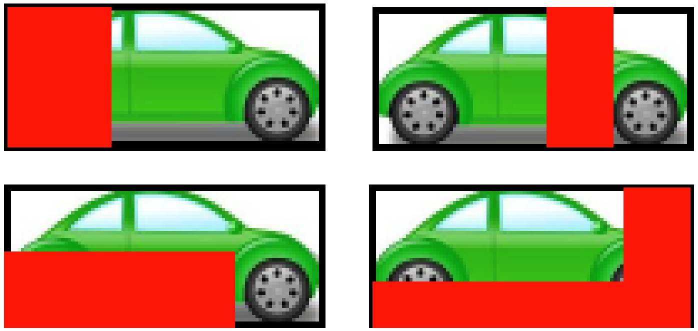
HOG-LBP human detector,
Wang et al. ICCV'09



Segmentation-aware detector,
Gao et al. CVPR'11

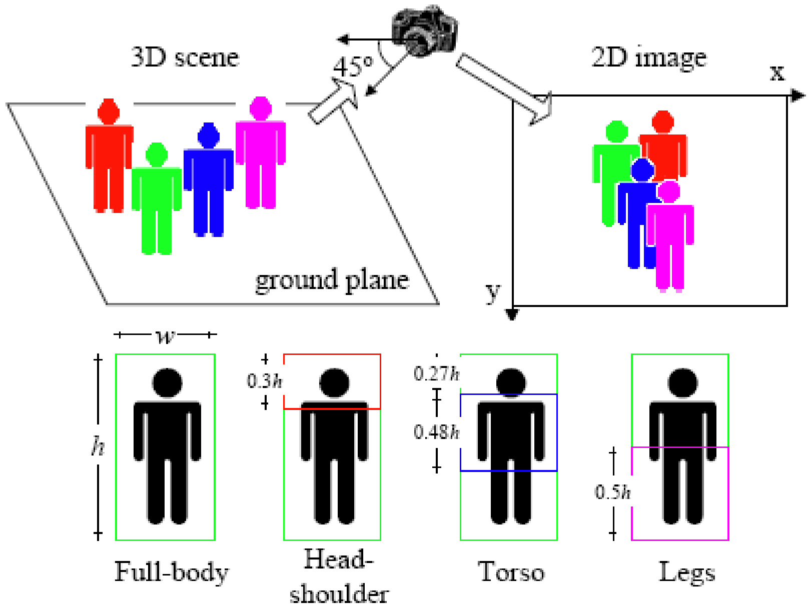


Occlusion Patterns, Pepik et al. CVPR'13

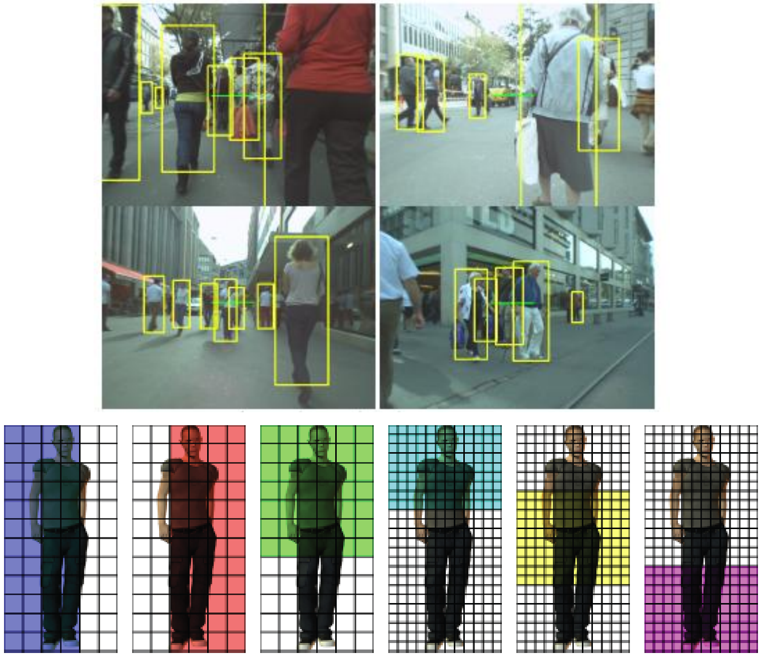


Occlusion masks, Zia et al. CVPR'13

Related work: 3D occlusion reasoning in object detection



Wu and Nevatia, ICCV'05



Wojek et al. CVPR'11

Difference:

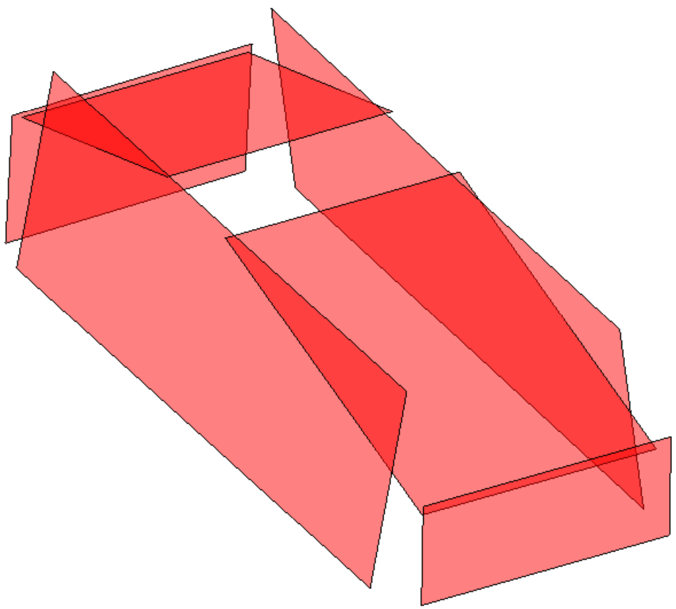
Instead of a simpilifed 2.5D structure of depth layers, we handle occlusion using a true 3D representation of object.

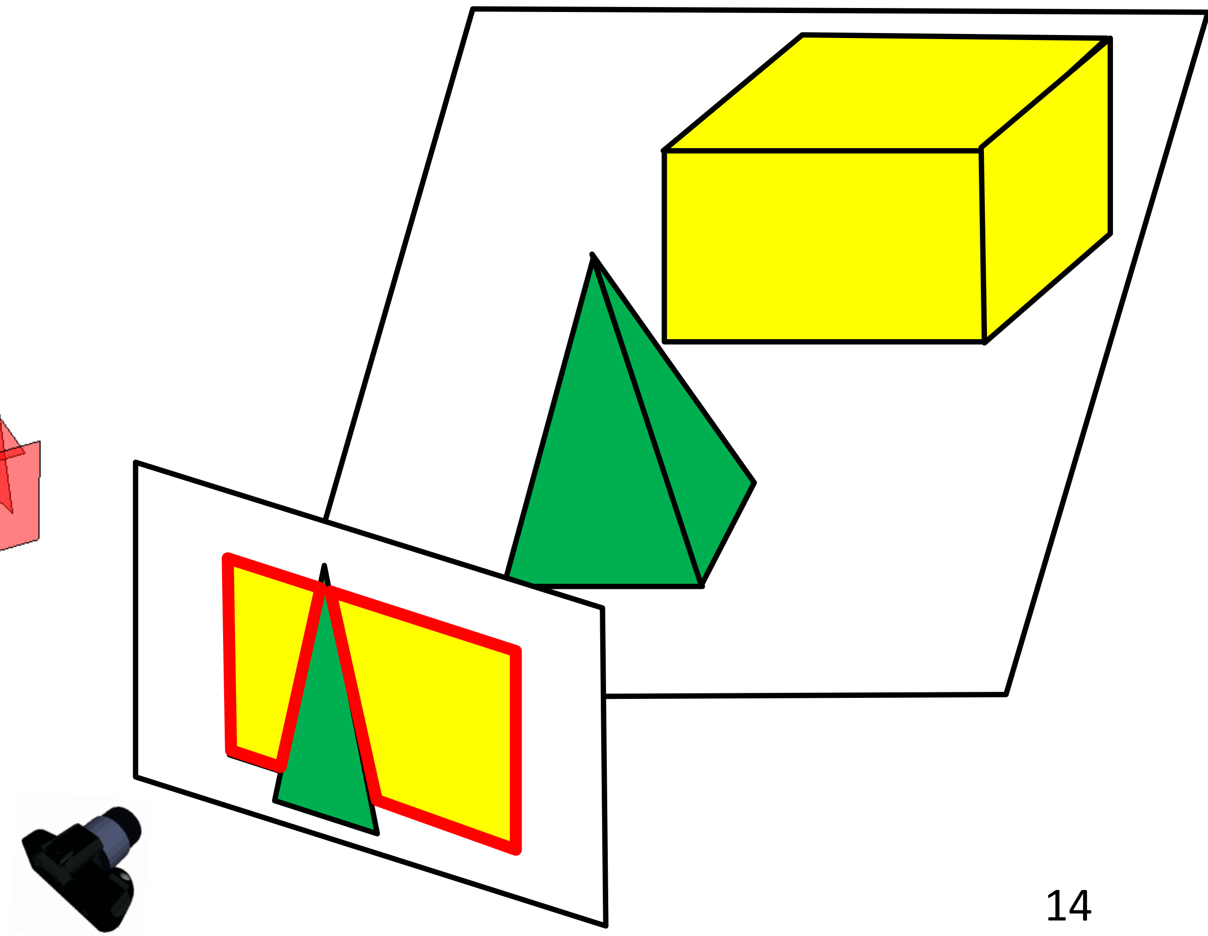
Outline

- 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

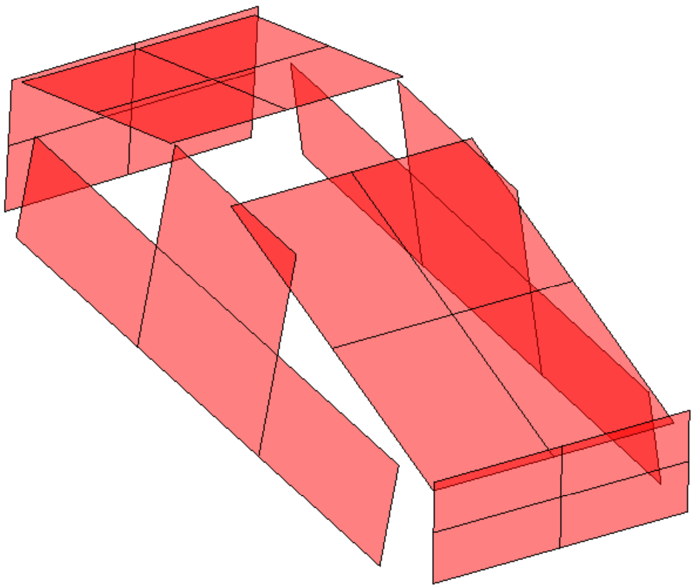


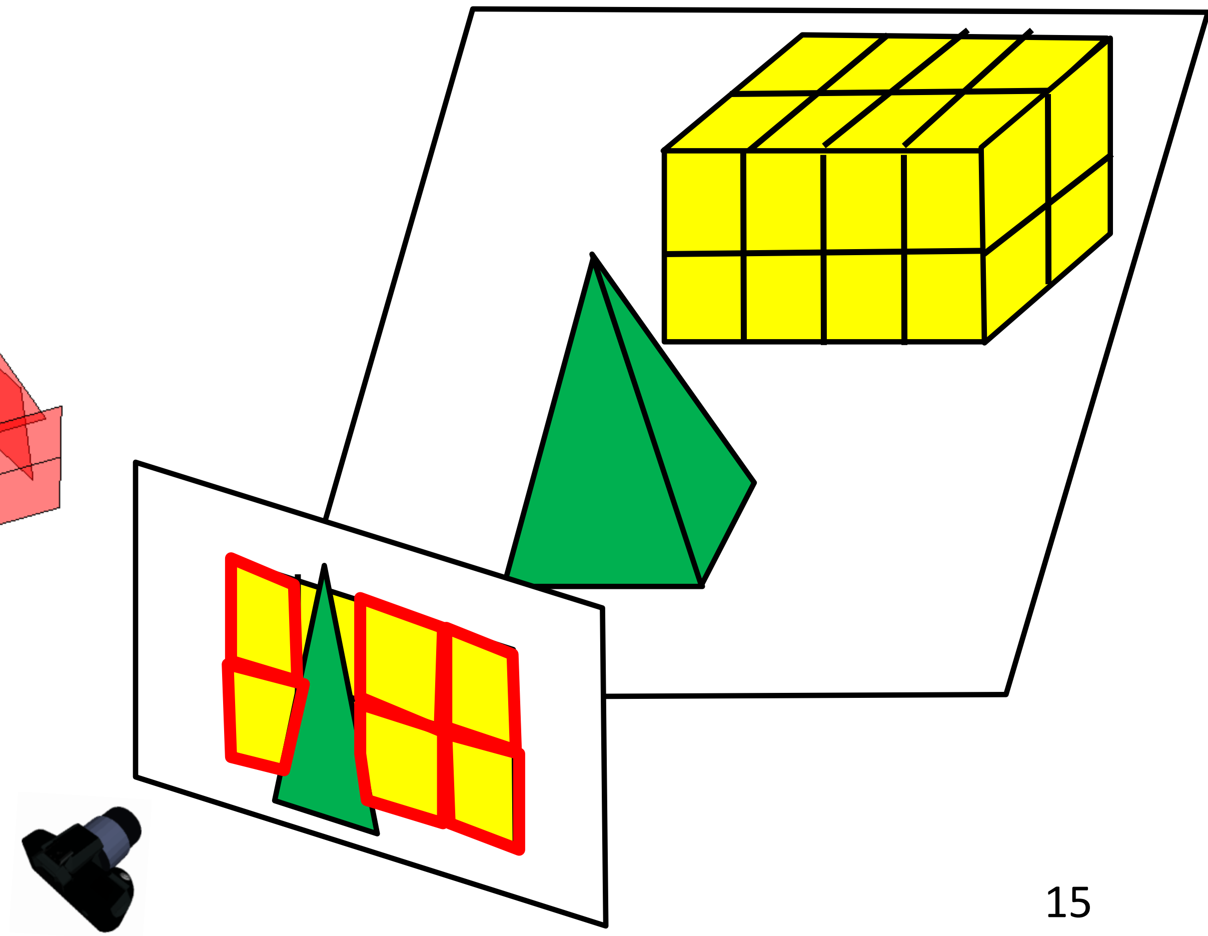


[Xiang & Savarese, CVPR'12]

3D Aspectlet

- Atomic aspect part for handling occlusion

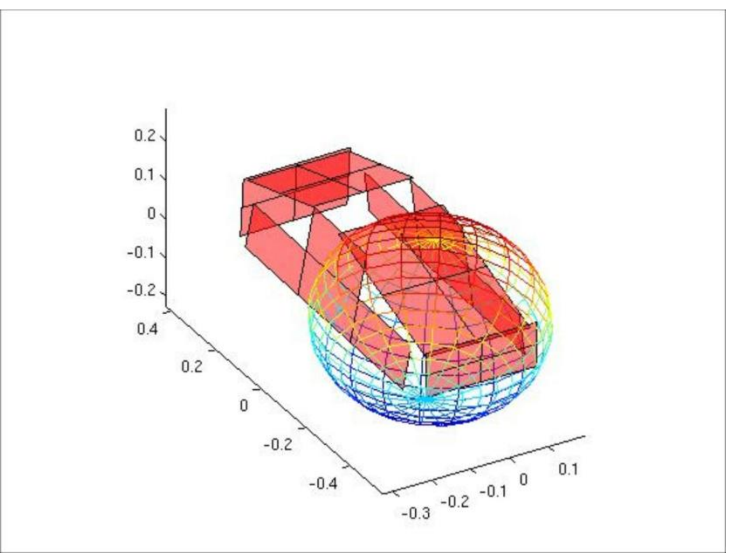


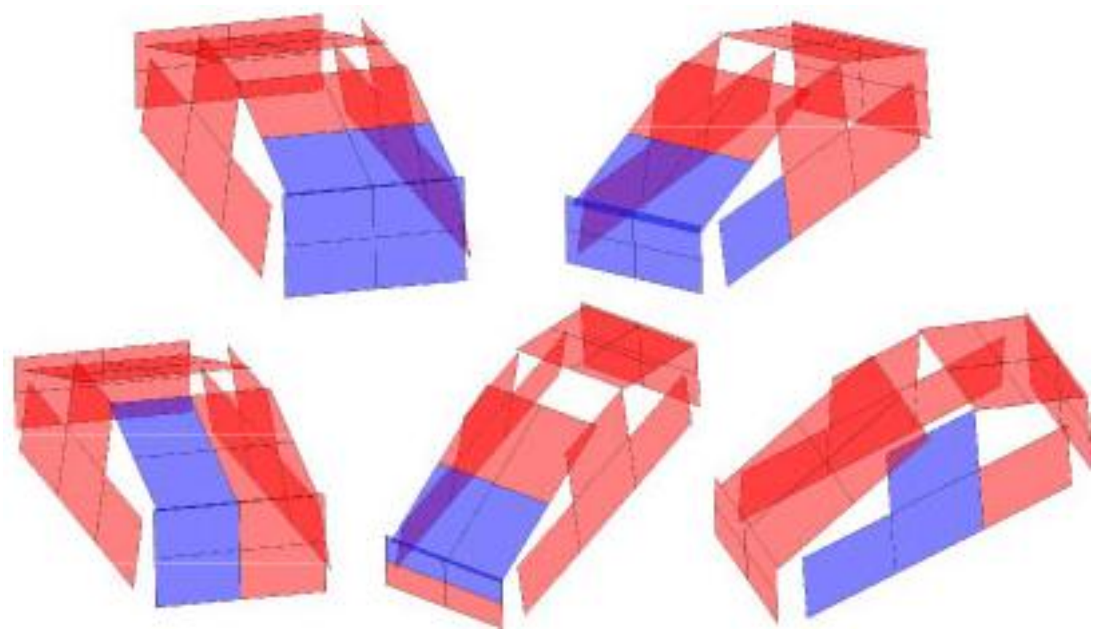


[Xiang & Savarese, 3dRR'13]

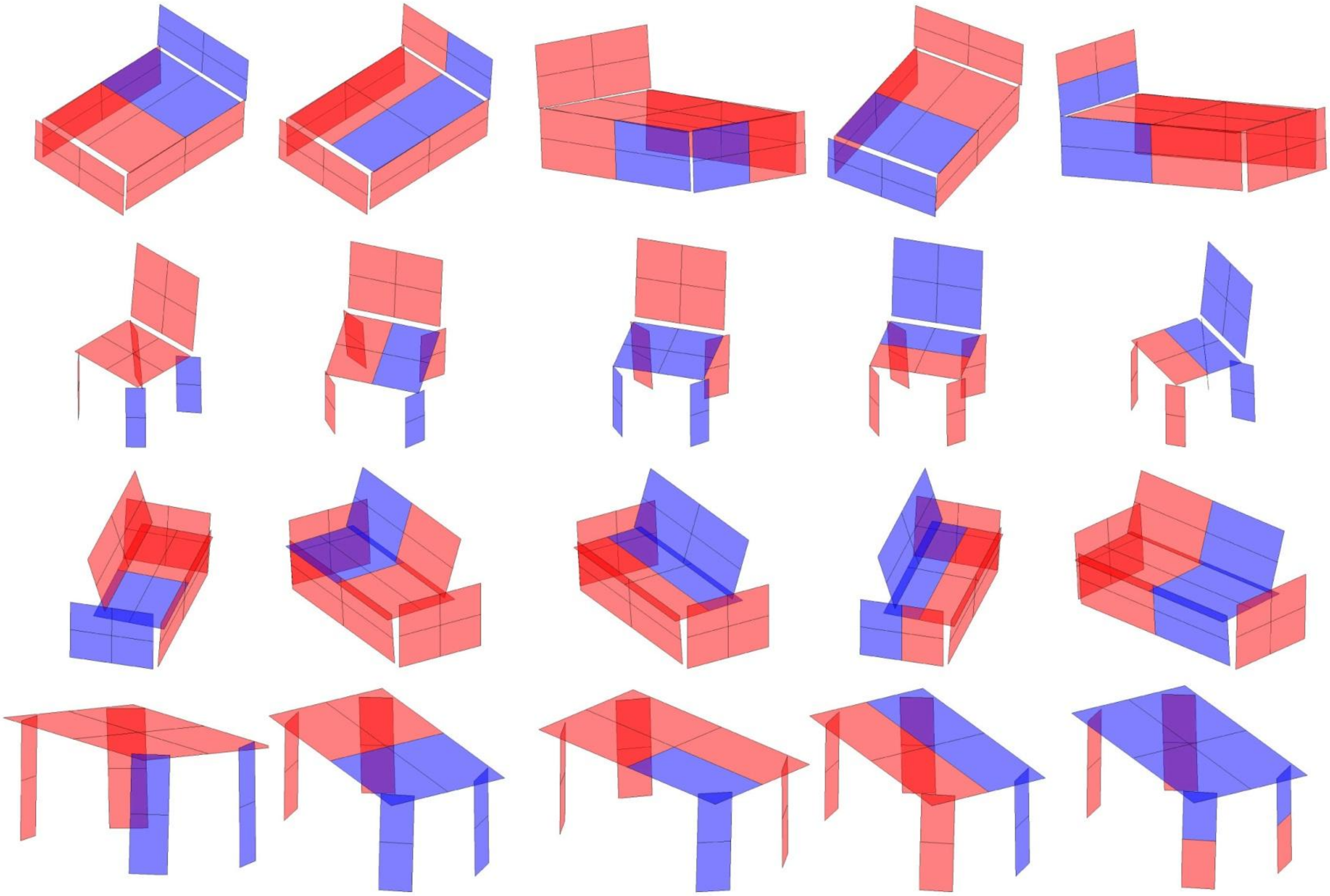
3D Aspectlets

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





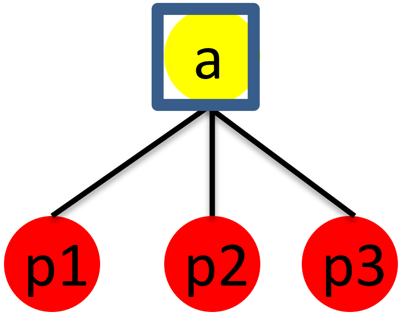
3D Aspectlets

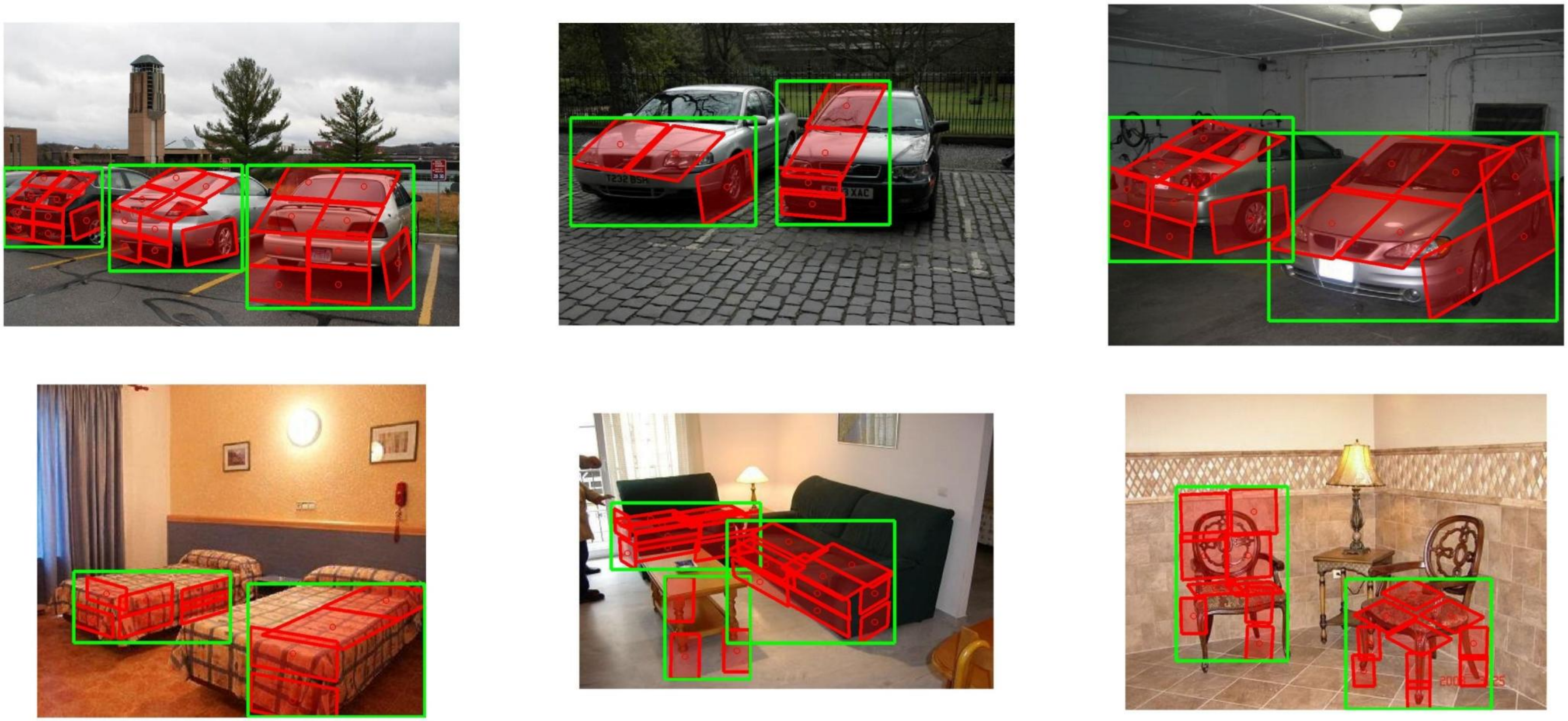


Yu Xiang and Silvio Savarese. 3dRR'13

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

2D projections | 3D objects | camera | a single input image

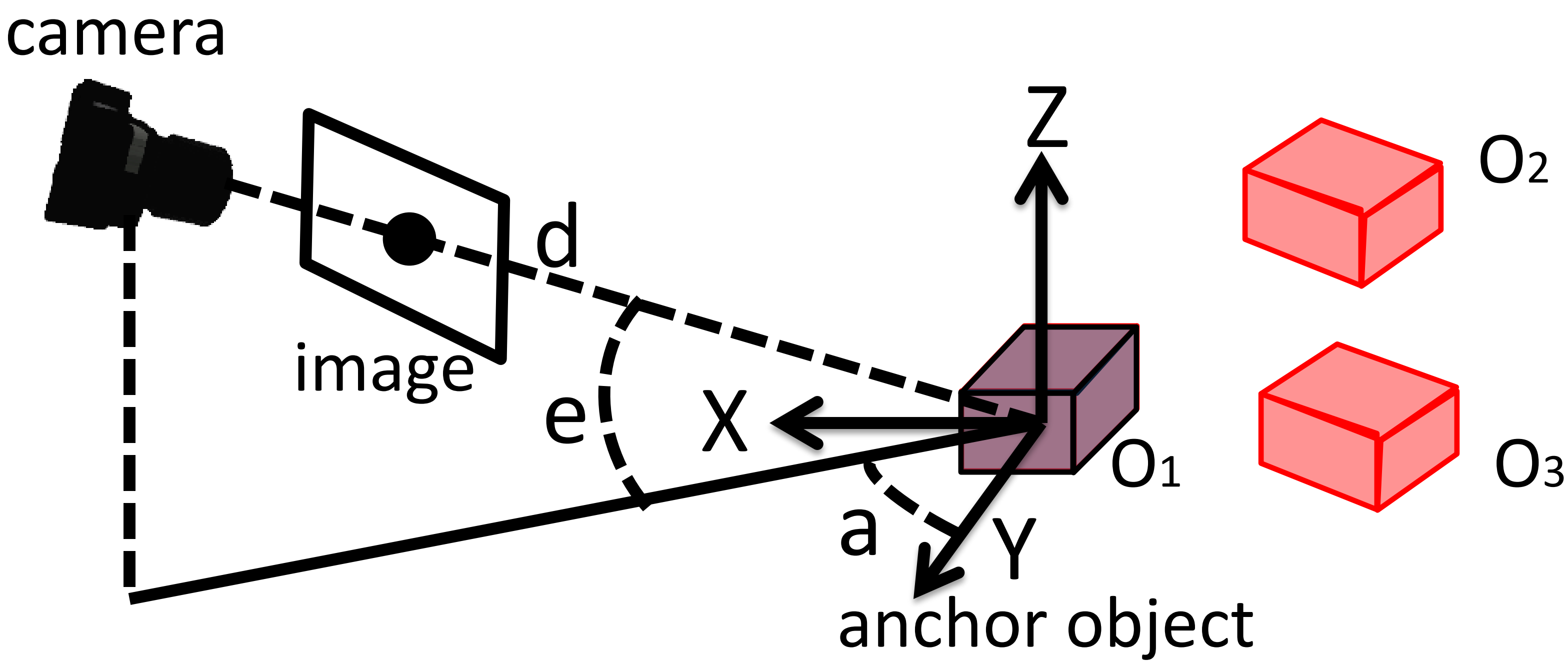
$$P(\mathbf{o}, \mathbf{O}, C \mid I)$$

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

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

Spatial Layout Model

- Camera prior
  - Virtual intrinsic camera matrix



$$P(C) = P(a)P(e)P(d)$$

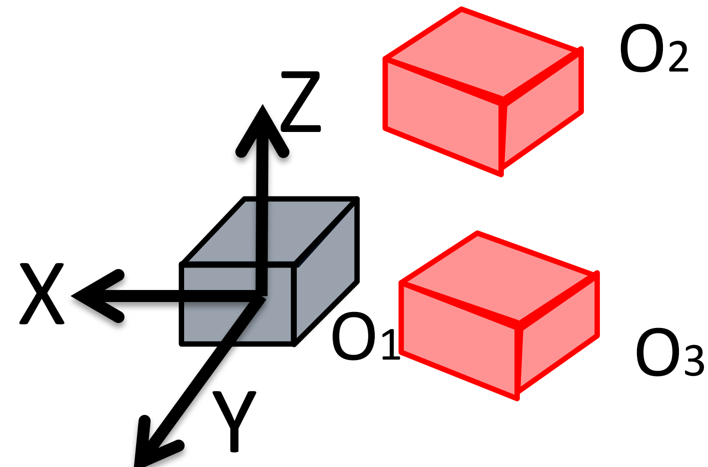
Spatial Layout Model

- 3D objects prior

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

  - "ground plane" constraint

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



  - 3D space constraint

$$V_2(O_i, O_j) = -\rho \frac{O_i \bigcap O_j}{O_i \bigcup O_j}$$

Spatial Layout Model

- Unary 2D projection likelihood

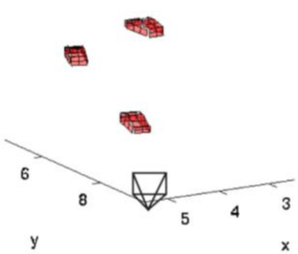
$$P(o_i \,|\, \mathbf{O}, C, I) \propto P_0(o_i \,|\, O_i, C, I) + \sum_{k=1}^{N} w_k(\mathbf{O}, C)\, P_k(o_i \,|\, O_i, C, I)$$

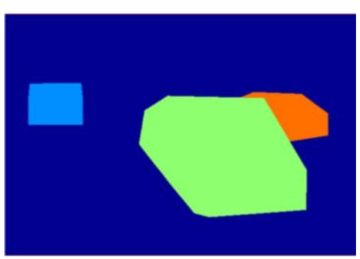
$$\text{s.t. } \sum_{k=1}^{N} w_k(\mathbf{O}, C) = 1$$

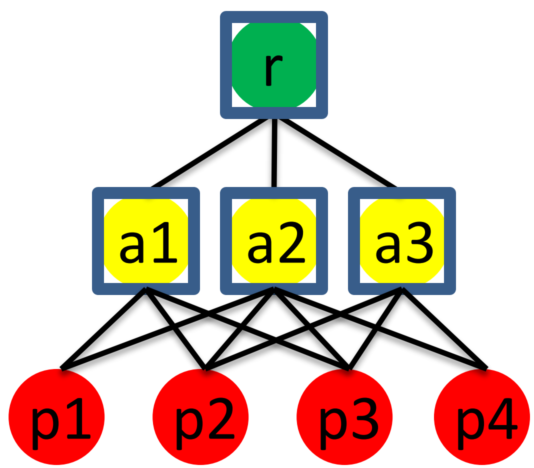
likelihood from the full-object model

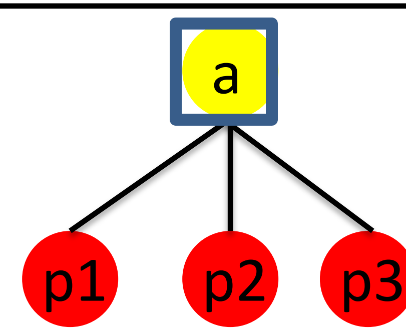
likelihood from the kth 3D aspectlet

weights proportional to the number of visible atomic aspect parts





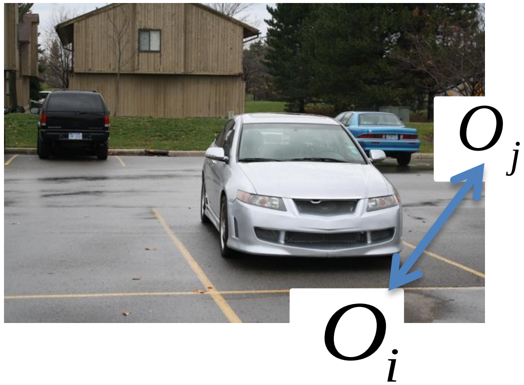




Yu Xiang and Silvio Savarese. 3dRR'13

Spatial Layout Model

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



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

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

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]

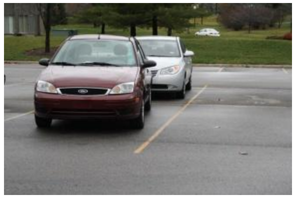
- Bed, Chair, Sofa and Table: Subset of ImageNet Dataset [Xiang & Savarese, CVPR'12]

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)

Category	Car	Bed	Chair	Sofa	Table
#objects	659	202	235	273	222
#occluded	235	81	112	175	61
#truncated	135	86	41	99	80






Detection APs

Category	Car	Bed	Chair	Sofa	Table
ALM [1]	46.6	28.9	14.2	41.1	19.2
DPM [2]	57.0	34.8	14.4	38.3	15.1
SLM Aspectlets	59.2	35.8	15.9	45.5	24.3
SLM Full	63.0	39.1	19.0	48.6	28.6

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

[1] Y. Xiang and S. Savarese. Estimating the aspect layout of object categories. In CVPR, 2012.
[2] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.

Detection APs

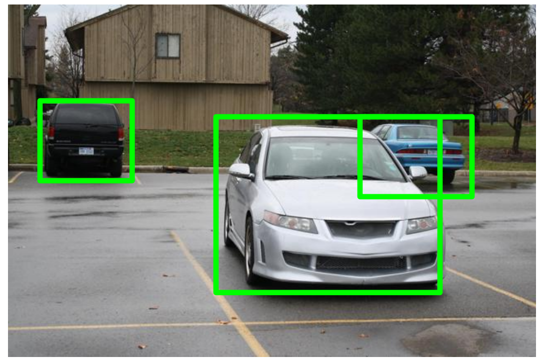
Dataset	Outdoor-scene			Indoor-scene		
% occlusion	< 0.3	0.3 – 0.6	> 0.6	<0.2	0.2-0.4	>0.4
# images	66	68	66	77	111	112
ALM [1]	72.3	42.9	35.5	38.5	25.0	20.2
DPM [2]	75.9	58.6	44.6	38.0	22.9	21.9
SLM Aspectlets	78.7	59.7	47.7	41.9	30.8	24.8
SLM Full	80.2	63.3	52.9	45.9	34.5	28.0

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

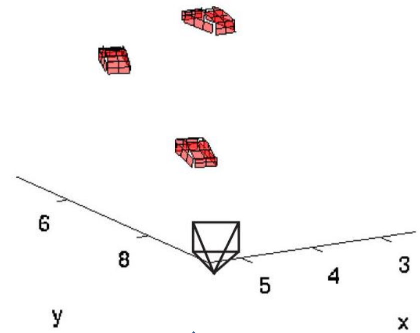
[1] Y. Xiang and S. Savarese. Estimating the aspect layout of object categories. In CVPR, 2012.
[2] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.

3D Localization Evaluation

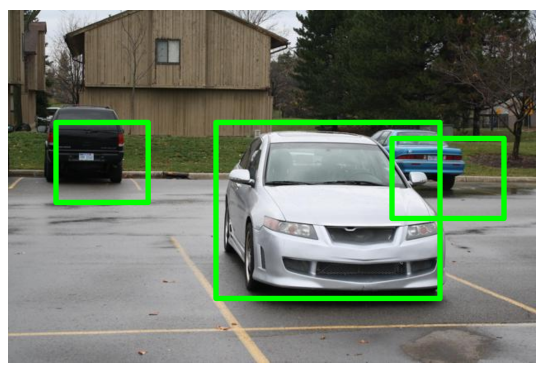
Ground truth 2D annotations



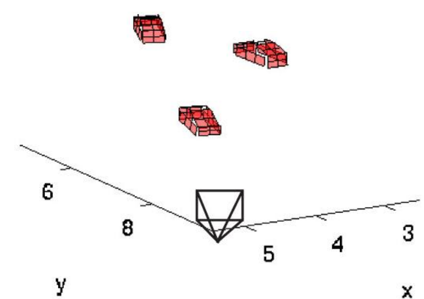
Ground truth 3D spatial layout



Mean absolute difference in pairwise distances



2D object detections



Predicted 3D spatial layout

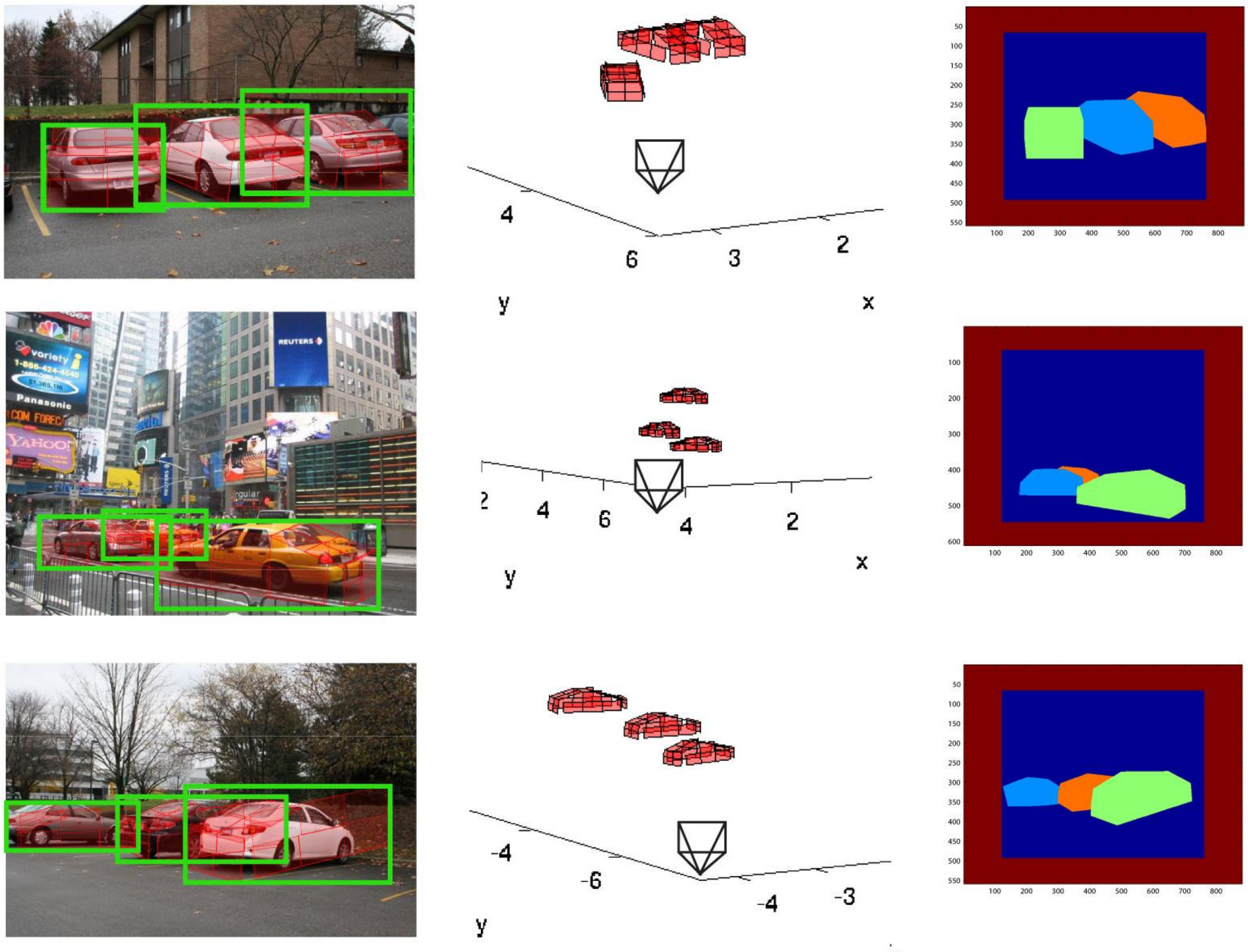
3D Localization

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

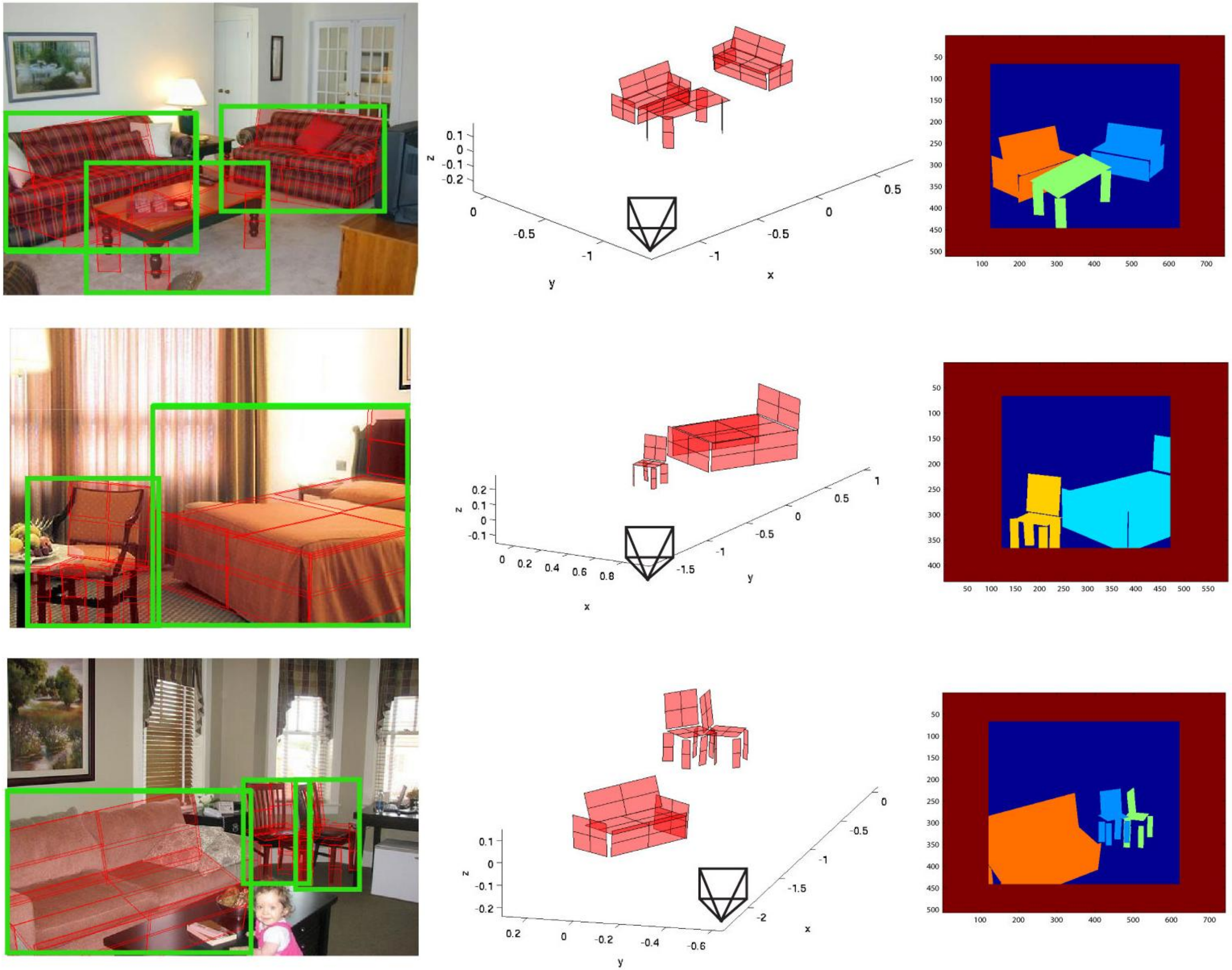
Recall	54.8	64.4	76.8
ALM [1]	1.90	-	-
DPM [2]	2.07	2.39	-
SLM	1.64	1.86	2.33

[1] Y. Xiang and S. Savarese. Estimating the aspect layout of object categories. In CVPR, 2012.
[2] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. TPAMI, 2010.

Anecdotal Results



Anecdotal Results



Outline

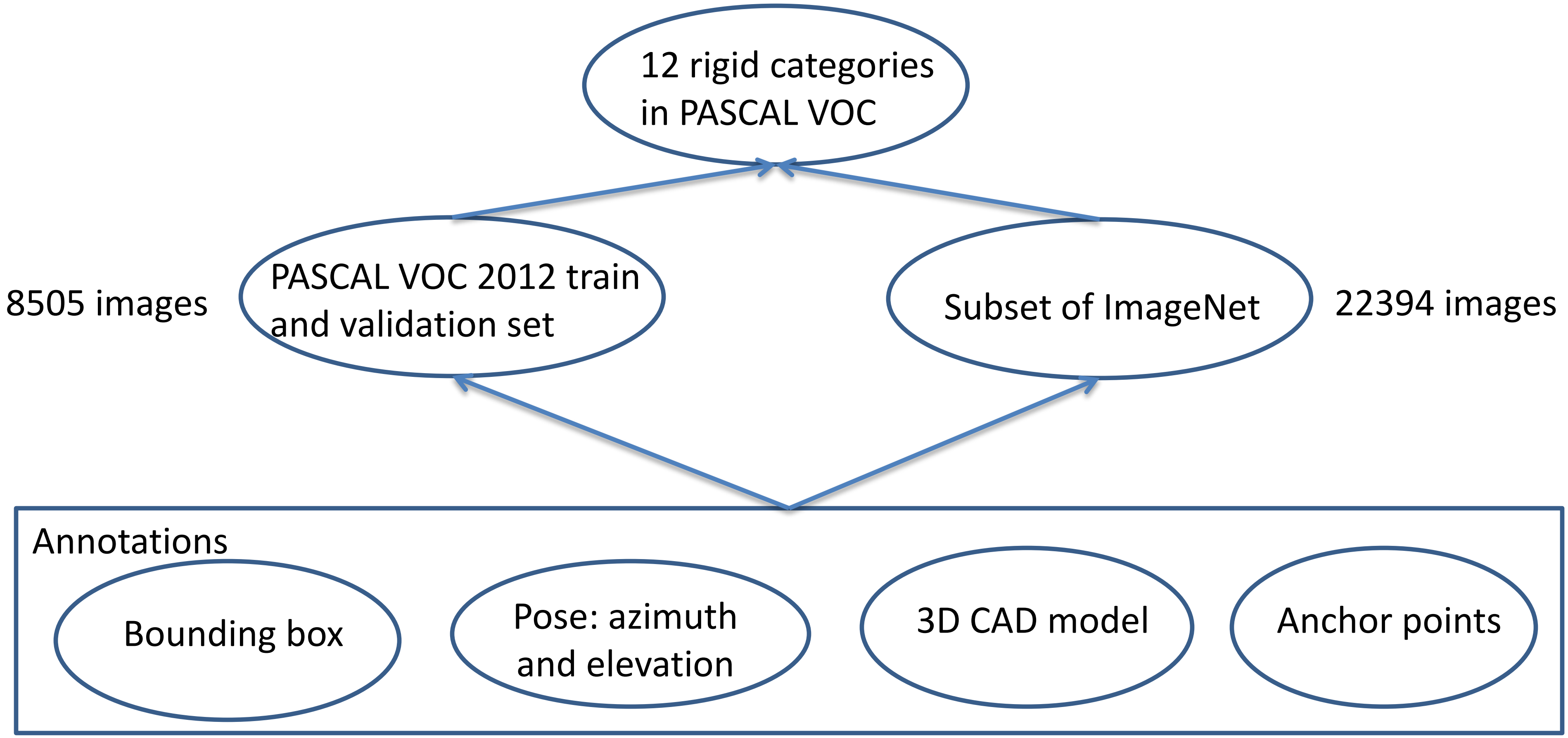
- Related work
- 3D aspectlets
- Spatial layout model
- Experiments
- Conclusion

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



Acknowledgments

Thank you!