Announcements

- In-class presentations next week!
  - Monday March 9th ; 11am-12:15pm (in class)
  - Tuesday March 10th ; 1:30pm-3:30pm (Bishop Auditorium)
  - Wednesday March 11th; 11am-12:15pm (in class)
- 2 ½ minutes for each presentation + Q&A
- It's a team presentation
- In-class or Piazza questions count toward attendance evaluation
- Best presentation award
- See Piazza and website for more information

Announcements

- Thanks for the online evaluation
- Your feedback is extremely important! (see lecture notes ☺)

New course: 231M: Mobile Computer Vision

- Shameless advertisement
- The course surveys recent developments in computer vision, graphics and image processing for mobile applications
  - Three problems sets (each related to a toy mobile application).
  - Course project: Extend one of the toy applications
  - Latest Nvidia Shield Tablets will be assigned to each student
- No midterm; no final

Lecture 16
Closure

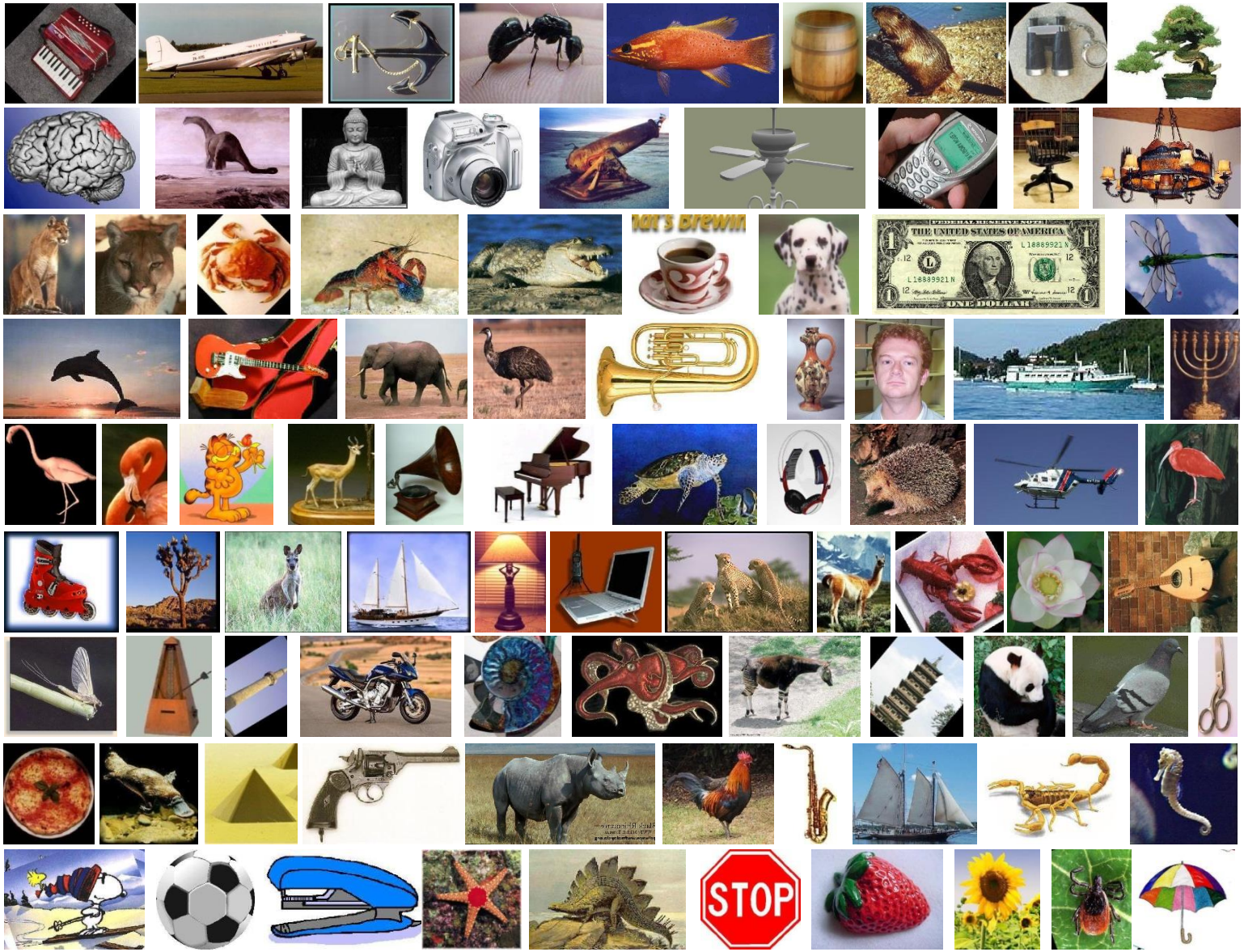
- Datasets in computer vision
- 3D scene understanding

Caltech 101

Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories. L. Fei-Fei, R. Fergus, and P. Perona. CVPR 2004, Workshop on Generative-Model Based Vision. 2004

- Pictures of objects belonging to 101 categories.
- About 40 to 800 images per category. Most categories have about 50 images.
- The size of each image is roughly 300 x 200 pixels.

Caltech 101 images

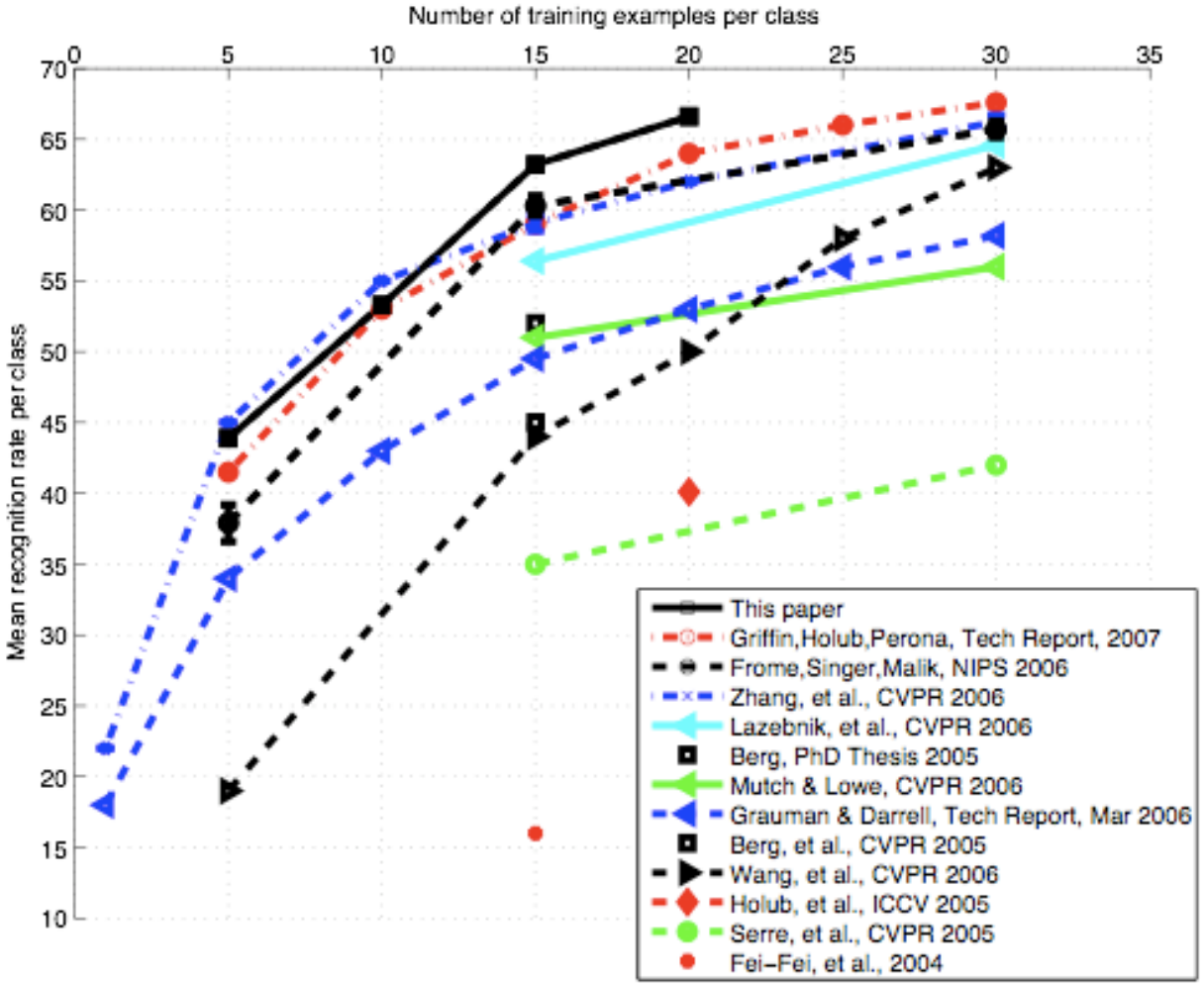


Caltech-101: Drawbacks

- Smallest category size is 31 images: $N_{train} \leq 30$

- Too easy?
  - left-right aligned
  - Rotation artifacts
  - Saturated performance

Caltech-101: Evaluation



Results up to 2007 -- recent methods obtain almost 100%

Caltech-256

Griffin, Gregory and Holub, Alex and Perona, Pietro (2007) *Caltech-256 Object Category Dataset.* California

- Smallest category size now 80 images
- About 30K images

- Harder
  - Not left-right aligned
  - No artifacts
  - More categories

- New and larger clutter category




Caltech 256 images

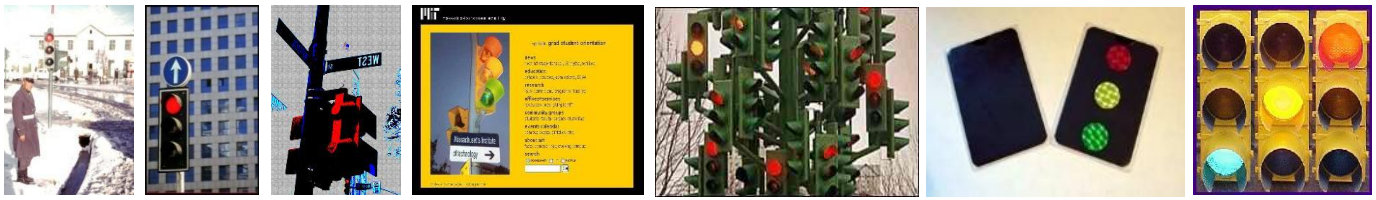
baseball-bat

dog

basketball-hoop

kayac

traffic light



The PASCAL Visual Object Classes (VOC) Dataset and Challenge (2005-2012)

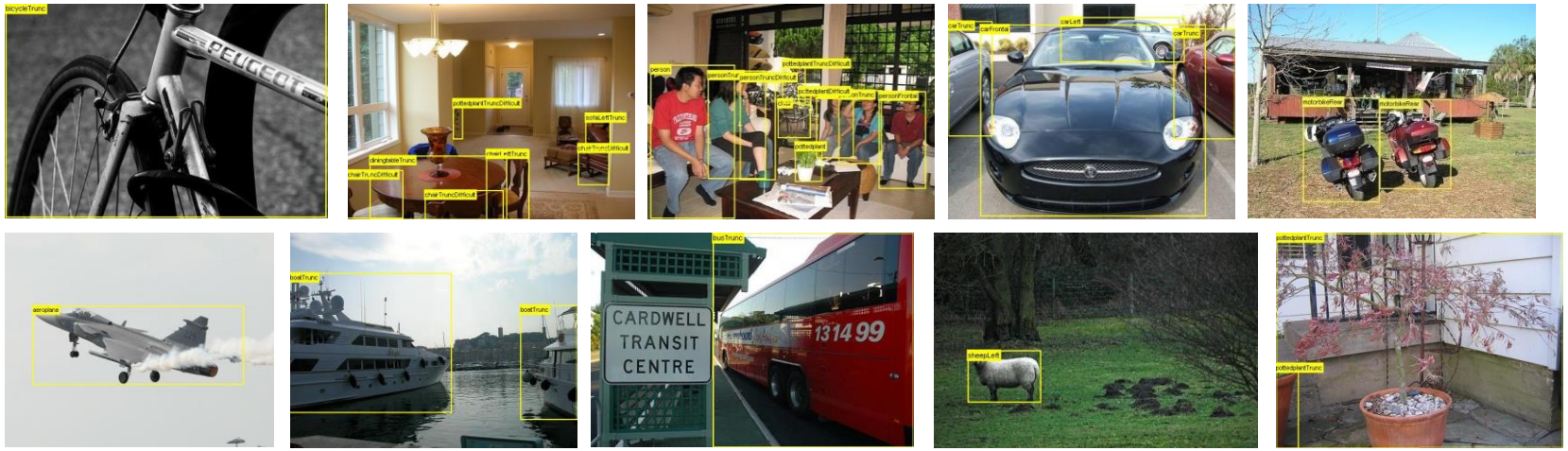
Mark Everingham
Luc Van Gool
Chris Williams
John Winn
Andrew Zisserman

PASCAL
Pattern Analysis, Statistical Modelling and Computational Learning

Dataset Content

- 20 classes: aeroplane, bicycle, boat, bottle, bus, car, cat, chair, cow, dining table, dog, horse, motorbike, person, potted plant, sheep, train, TV

- Real images downloaded from flickr, not filtered for “quality”



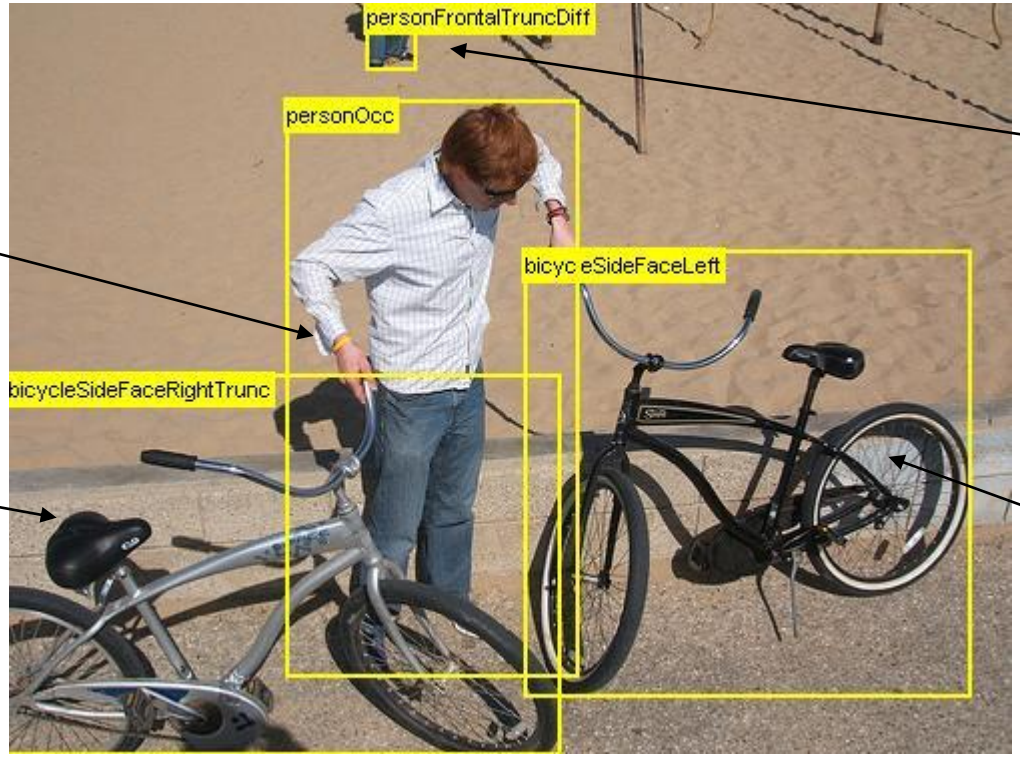
- Complex scenes, scale, pose, lighting, occlusion, ...

Annotations

- Complete annotation of all objects
- Annotated in one session with written guidelines

Occluded
Object is significantly occluded within BB

Truncated
Object extends beyond BB



Difficult
Not scored in evaluation

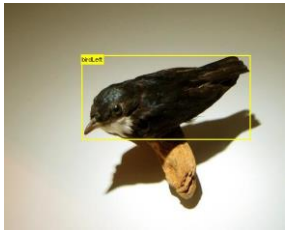
Pose
Facing left

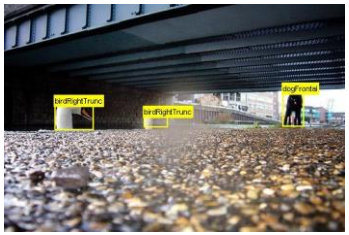
Examples

Aeroplane

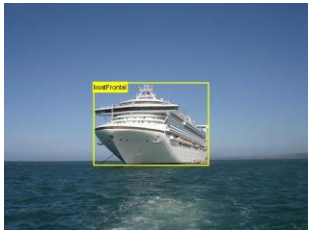
Bicycle

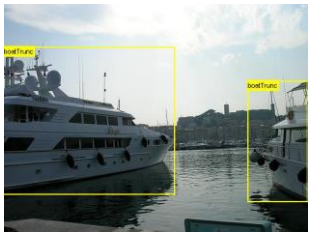
Bird



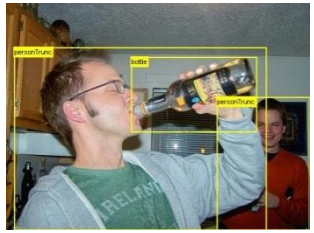


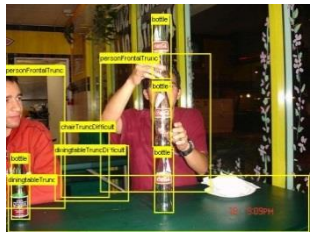
Boat





Bottle

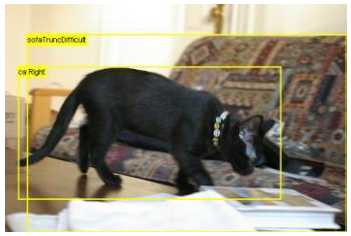




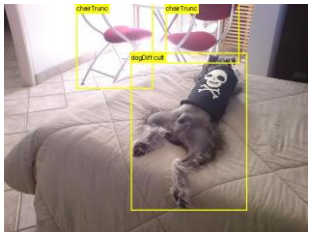
Bus

Car

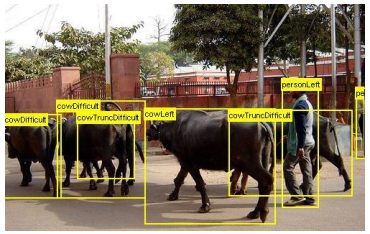
Cat



Chair



Cow



History

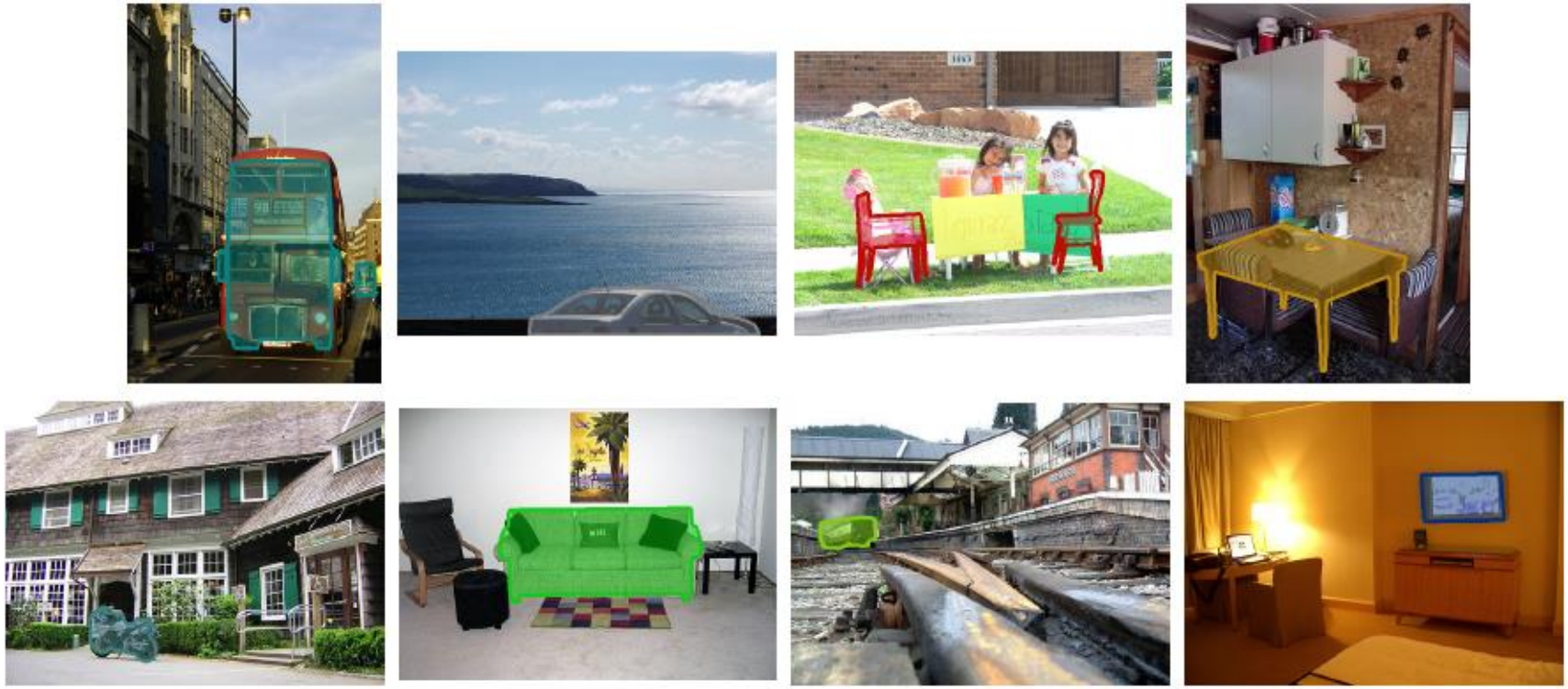
	Images	Objects	Classes	Notes
2005	2,232	2,871	4	*Collection of existing and some new data.*
2006	5,304	9,507	10	*Completely new dataset from flickr (+MSRC)*
2007	9,963	24,640	20	*Increased classes to 20. Introduced tasters.*
2008	8,776	20,739	20	*Added "occlusion" flag."*
⇩				
2012	11,530	27,450	20	*Added segmentation masks*

- Challenge: annotation of test set is withheld until after challenge

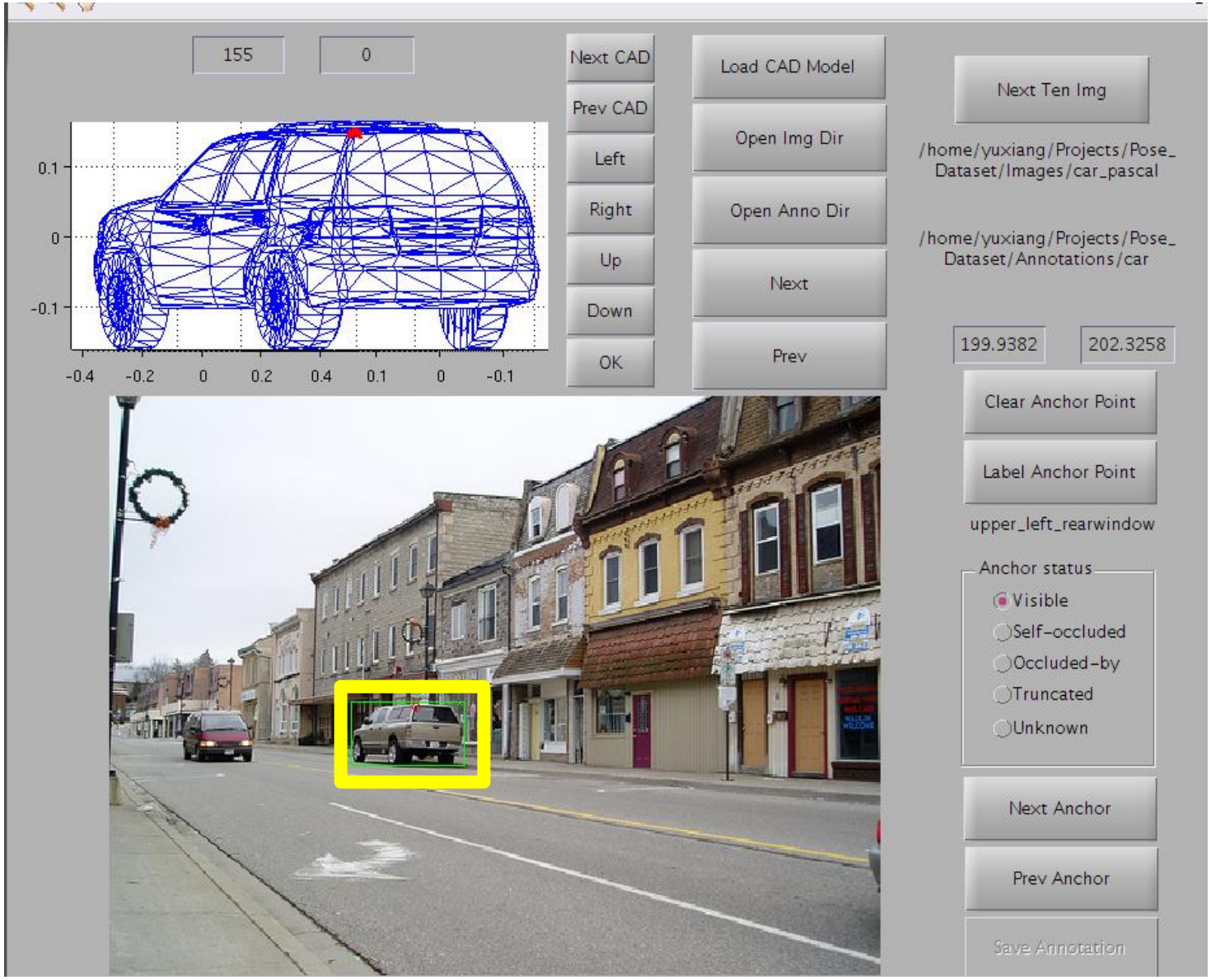
PASCAL 3D+

Xiang, Mottaghi, Savarese WACV 14

- 12 rigid categories from PASCAL VOC are annotated with 3D pose and aligned with 3D cad models (now almost 100 categories!)
- Integrated with images from other repositories
- New benchmark for continuous 3D pose estimation and shape recovery of object categories



PASCAL 3D+


155
0
Next CAD
Prev CAD
Left
Right
Up
Down
OK
Load CAD Model
Open Img Dir
Open Anno Dir
Next
Prev
Next Ten Img
/home/yuxiang/Projects/Pose_Dataset/Images/car_pascal
/home/yuxiang/Projects/Pose_Dataset/Annotations/car
199.9382
202.3258
Clear Anchor Point
Label Anchor Point
upper_left_rearwindow
Anchor status
Visible
Self-occluded
Occluded-by
Truncated
Unknown
Next Anchor
Prev Anchor
Save Annotation

Other recent datasets

ESP
[Ahn et al, 2006]

LabelMe
[Russell et al, 2005]

TinyImage
Torralba et al. 2007

Lotus Hill
[Yao et al, 2007]

MSRC
[Shotton et al. 2006]

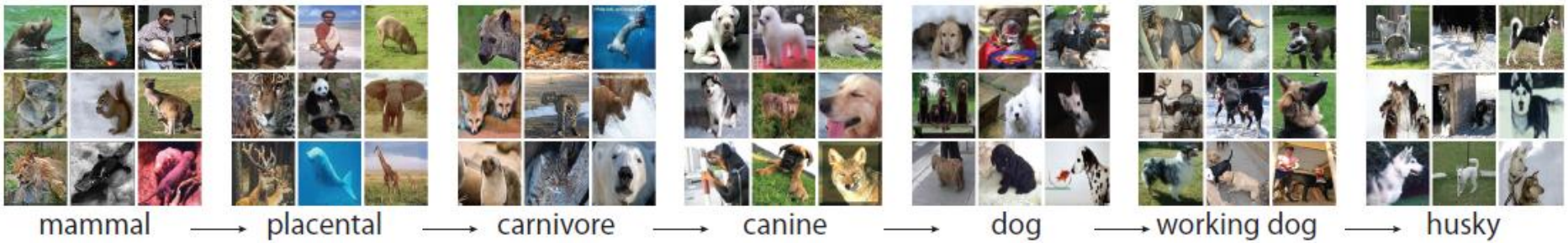
IMAGENET J. Deng, H. Su. K. Li , L. Fei-Fei ,

Largest dataset for object categories up to date

- ~20K categories;
- 14 million images;
- ~700im/categ;
- free to public at **www.image-net.org**

IMAGENET is a knowledge ontology

- Taxonomy



- S: (n) Eskimo dog, **husky** (breed of heavy-coated Arctic sled dog)
  - *direct hypernym* / ***inherited hypernym*** / *sister term*
    - S: (n) working dog (any of several breeds of usually large powerful dogs bred to work as draft animals and guard and guide dogs)
      - S: (n) dog, domestic dog, Canis familiaris (a member of the genus Canis (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds) *"the dog barked all night"*
        - S: (n) canine, canid (any of various fissiped mammals with nonretractile claws and typically long muzzles)
          - S: (n) carnivore (a terrestrial or aquatic flesh-eating mammal) *"terrestrial carnivores have four or five clawed digits on each limb"*
            - S: (n) placental, placental mammal, eutherian, eutherian mammal (mammals having a placenta; all mammals except monotremes and marsupials)
              - S: (n) mammal, mammalian (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for the small subclass of monotremes and nourished with milk)
                - S: (n) vertebrate, craniate (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)
                - S: (n) chordate (any animal of the phylum Chordata having a notochord or spinal column)
                - S: (n) animal, animate being, beast, brute, creature, fauna (a living organism characterized by voluntary movement)
                - S: (n) organism, being (a living thing that has (or can develop) the ability to act or function independently)
                - S: (n) living thing, animate thing (a living (or once living) entity)
                - S: (n) whole, unit (an assemblage of parts that is regarded as a single entity) *"how big is that part compared to the whole?"; "the team is a unit"*
                - S: (n) object, physical object (a tangible and visible entity; an entity that can cast a shadow) *"it was full of rackets, balls and other objects"*
                - S: (n) physical entity (an entity that has physical existence)
                - S: (n) entity (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

OpenSurfaces

Sean Bell, Paul Upchurch, Noah Snavely, Kavita Bala
Cornell University

Materials

Reflectances

Textures

+80K annotated images of materials

COCO dataset

Tsung-Yi Lin Cornell Tech
Michael Maire TTI Chicago
Serge Belongie Cornell Tech
Lubomir Bourdev Facebook AI
Ross Girshick Microsoft Research
James Hays Brown University
Pietro Perona Caltech
Deva Ramanan UC Irvine
Larry Zitnick Microsoft Research
Piotr Dollár Facebook AI

- Microsoft COCO is a new image recognition, segmentation, and captioning dataset.
- Features:
  - Object segmentation
  - Recognition in Context
  - Multiple objects per image
  - More than 300,000 images
  - More than 2 Million instances
  - 80 object categories
  - 5 captions per image

More Datasets….

UIUC Cars (2004)
S. Agarwal, A. Awan, D. Roth

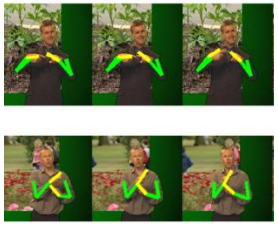
CMU/VASC Faces (1998)
H. Rowley, S. Baluja, T. Kanade

FERET Faces (1998)
P. Phillips, H. Wechsler, J. Huang, P. Raus

COIL Objects (1996)
S. Nene, S. Nayar, H. Murase

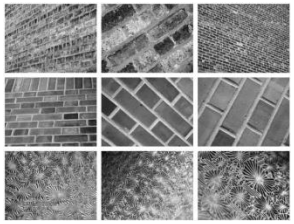
MNIST digits (1998-10)
Y LeCun & C. Cortes

KTH human action (2004)
I. Leptev & B. Caputo

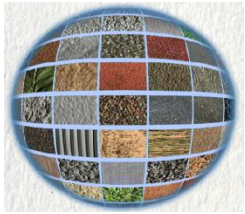


Sign Language (2008)
P. Buehler, M. Everingham, A. Zisserman

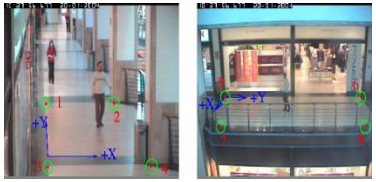
Segmentation (2001)
D. Martin, C. Fowlkes, D. Tal, J. Malik.



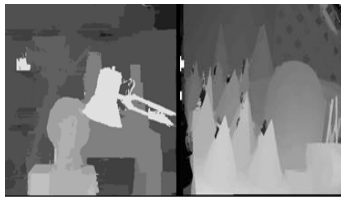
3D Textures (2005)
S. Lazebnik, C. Schmid, J. Ponce



CuRRET Textures (1999)
K. Dana B. Van Ginneken S. Nayar J. Koenderink



CAVIAR Tracking (2005)
R. Fisher, J. Santos-Victor J. Crowley



Middlebury Stereo (2002)
D. Scharstein R. Szeliski

Links to datasets

The next tables summarize some of the available datasets for training and testing object detection and recognition algorithms. These lists are far from exhaustive.

Databases for object localization

CMU/MIT frontal faces	vasc.ri.cmu.edu/idb/html/face/frontal_images cbcl.mit.edu/software-datasets/FaceData2.html	Patches	Frontal faces
Graz-02 Database	www.emt.tugraz.at/~pinz/data/GRAZ_02/	Segmentation masks	Bikes, cars, people
UIUC Image Database	l2r.cs.uiuc.edu/~cogcomp/Data/Car/	Bounding boxes	Cars
TU Darmstadt Database	www.vision.ethz.ch/leibe/data/	Segmentation masks	Motorbikes, cars, cows
LabelMe dataset	people.csail.mit.edu/brussell/research/LabelMe/intro.html	Polygonal boundary	>500 Categories

Databases for object recognition

Caltech 101	www.vision.caltech.edu/Image_Datasets/Caltech101/Caltech101.html	Segmentation masks	101 categories
Caltech 256	http://www.vision.caltech.edu/Image_Datasets/Caltech256/	Bounding Box	256 Categories
COIL-100	www1.cs.columbia.edu/CAVE/research/softlib/coil-100.html	Patches	100 instances
NORB	www.cs.nyu.edu/~ylclab/data/norb-v1.0/	Bounding box	50 toys

On-line annotation tools

ESP game	www.espgame.org	Global image descriptions	Web images
LabelMe	people.csail.mit.edu/brussell/research/LabelMe/intro.html	Polygonal boundary	High resolution images

Collections

PASCAL	http://www.pascal-network.org/challenges/VOC/	Segmentation, boxes	various

Lecture 16
Closure

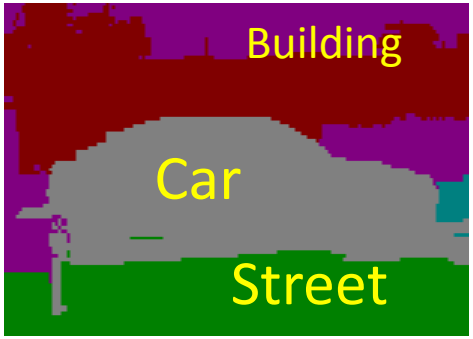
- Datasets in computer vision
- 3D scene understanding

What does it mean to understand a scene?

Image-to-labels paradigm

image

labels



Is computer vision solved??

Future generation of computer vision students

Oh my gosh!

Sky

Chocolate bar

Tree

Road

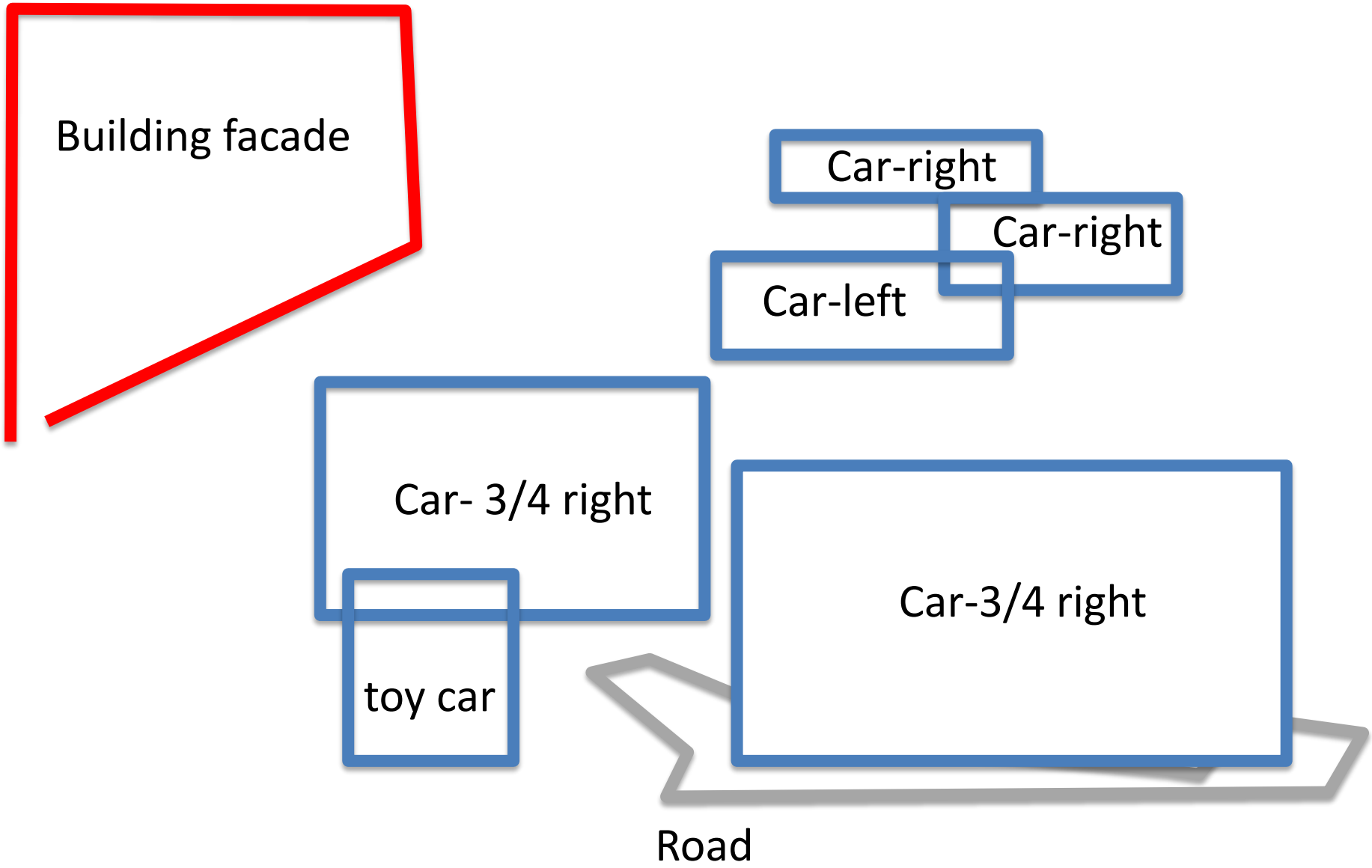
Sky

Chocolate bar

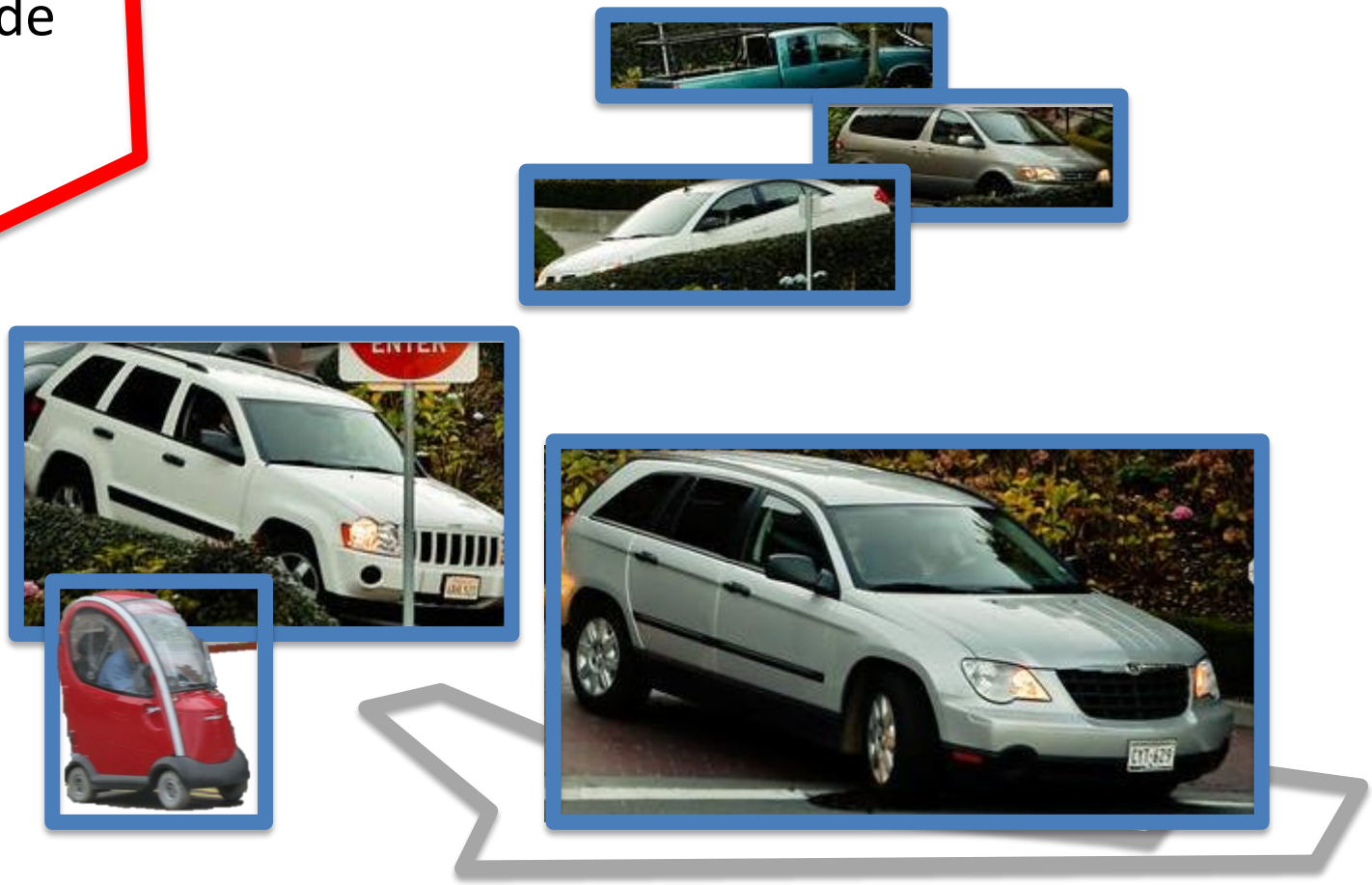
Tree

Road


Mars
TRUCK SIZE

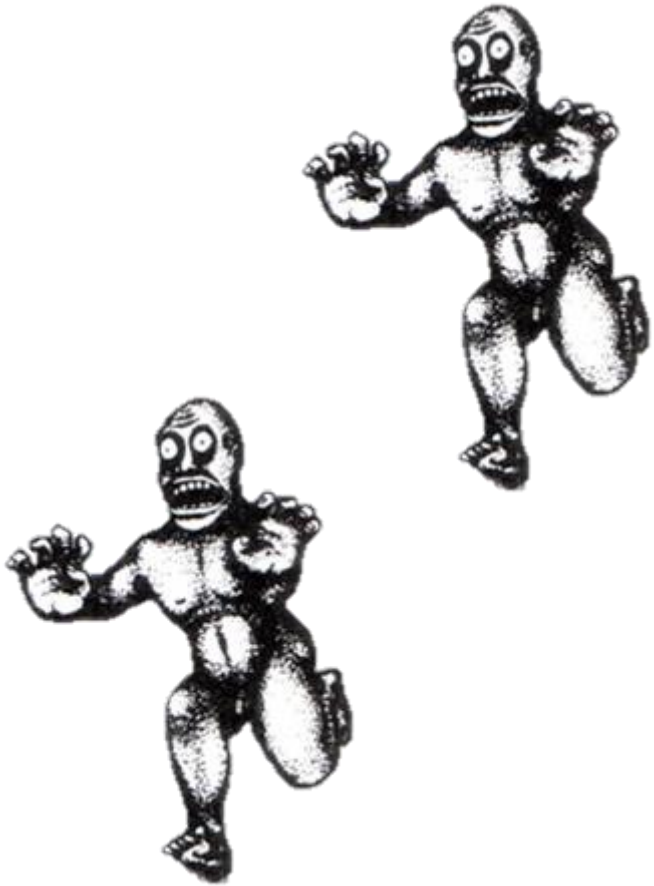

Building facade
Car-right
Car-right
Car-left
Car- 3/4 right
toy car
Car-3/4 right
Road

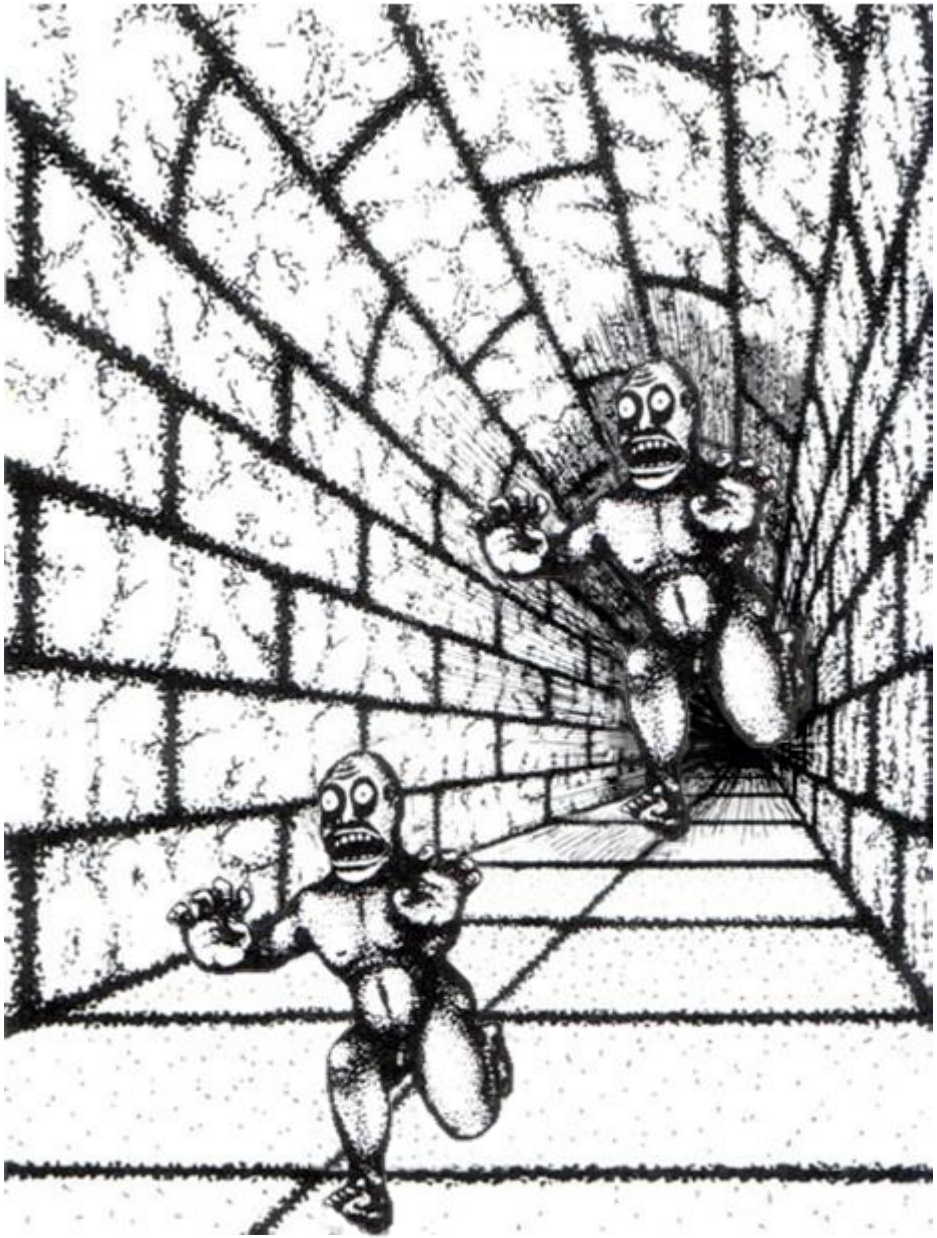
Building facade



Road


Lombard Street, San Francisco (2)
DO NOT
ENTER
WorldOnaBike.com
(c) Harry Kikstra, WorldOnaBike.co



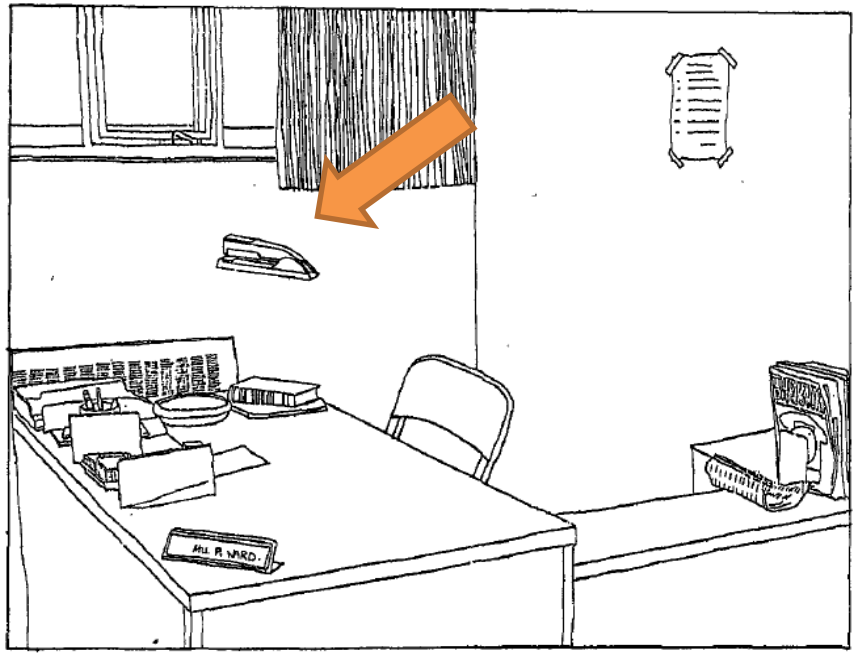


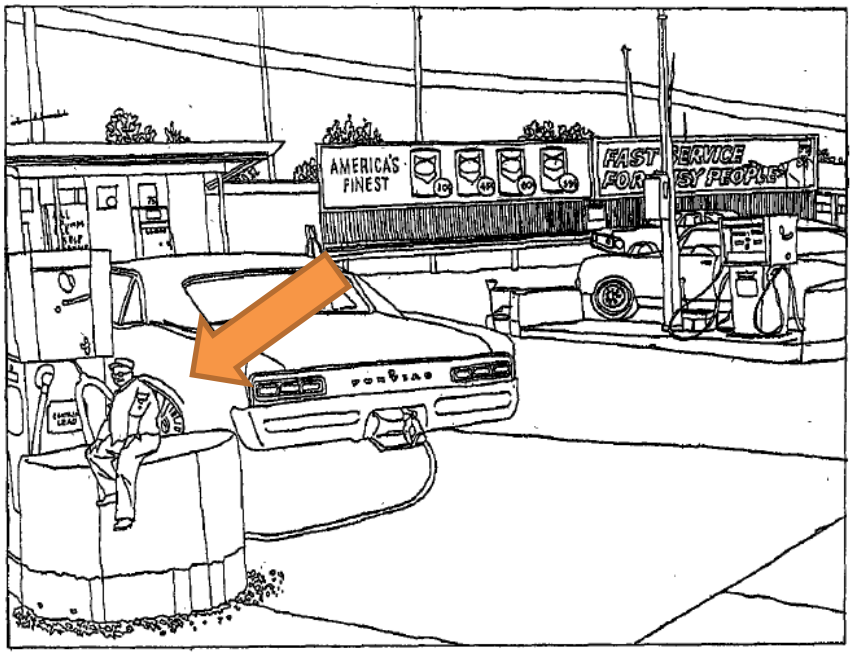
Objects are constrained by the 3D space

The 3D space is shaped by its objects

Modeling this interplay is critical for 3D perception!

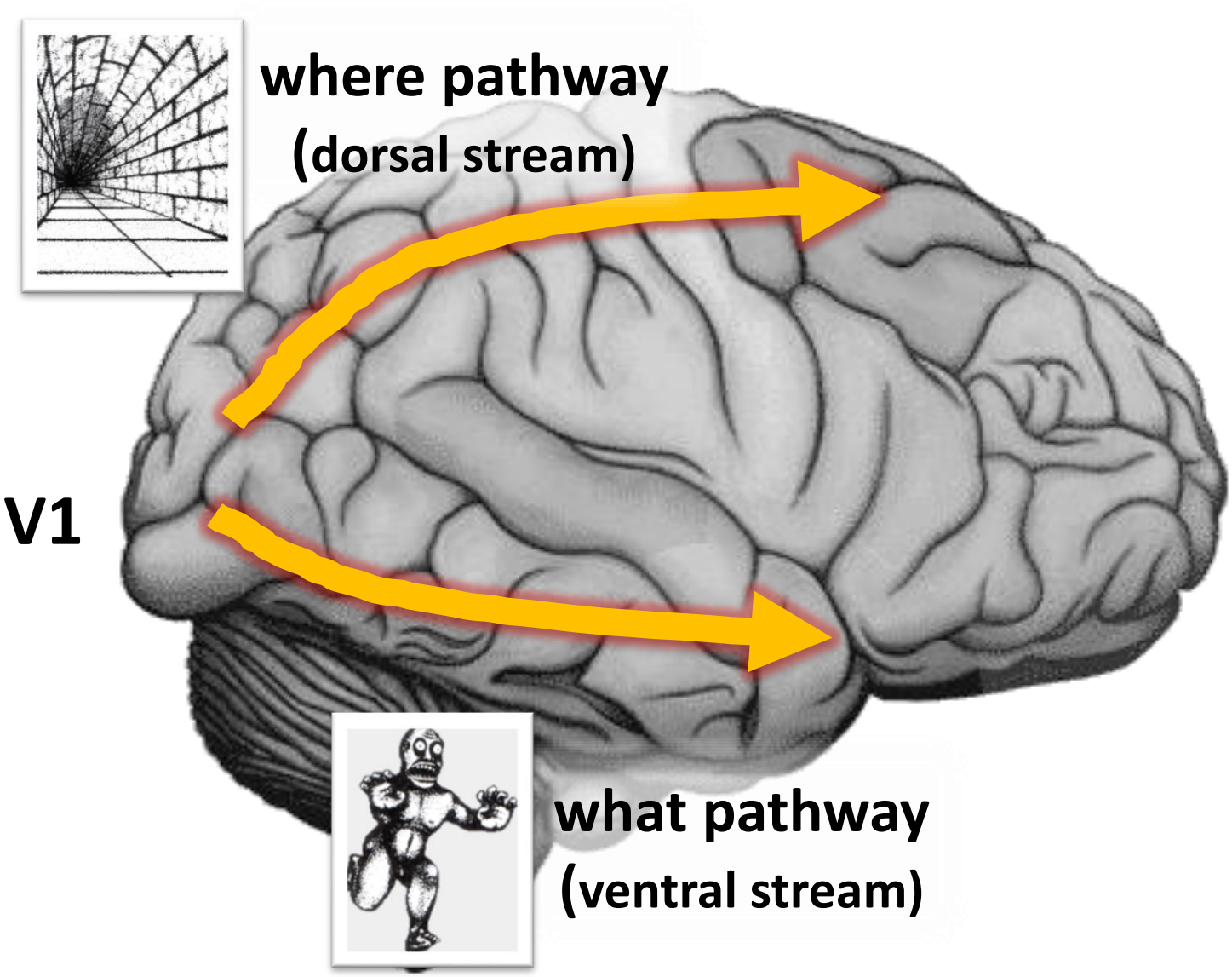
Humans perceive the world in 3D



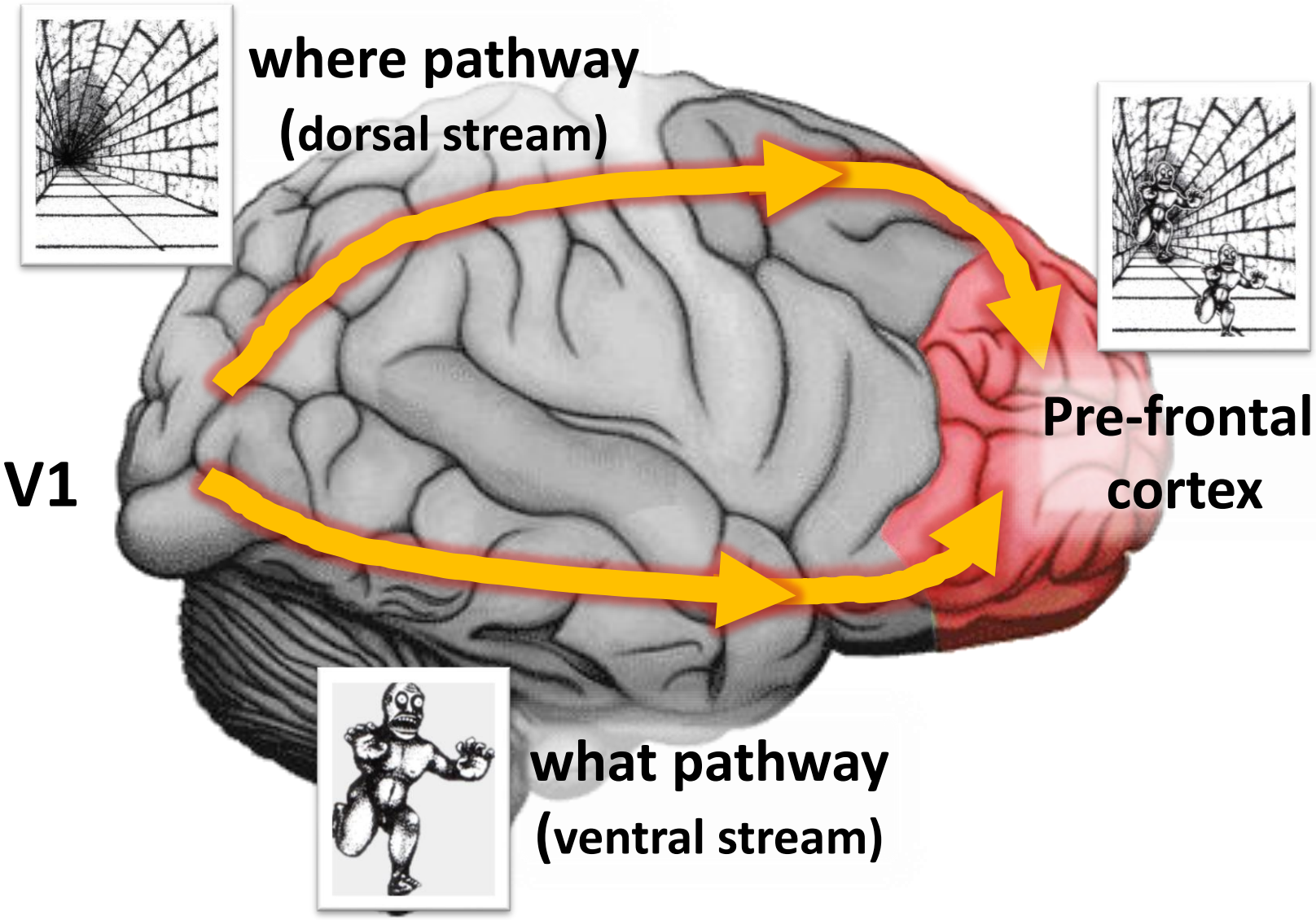


Biederman, Mezzanotte and Rabinowitz, 1982

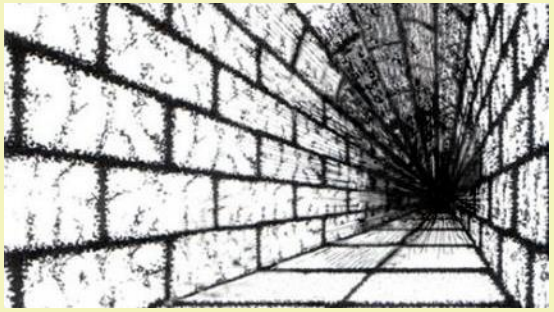
Visual processing in the brain



Visual processing in the brain



Current state of computer vision



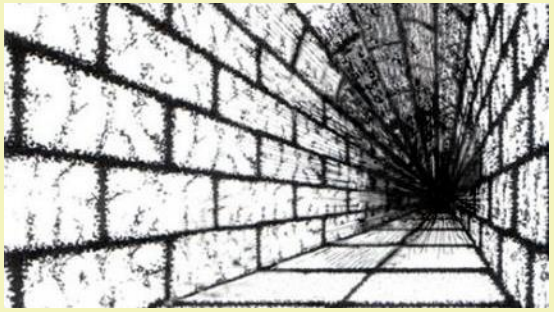
3D Reconstruction

- 3D shape recovery
- 3D scene reconstruction
- Camera localization
- Pose estimation

2D Recognition

- Object detection
- Texture classification
- Target tracking
- Activity recognition

Current state of computer vision



3D Reconstruction

- 3D shape recovery
- 3D scene reconstruction
- Camera localization
- Pose estimation

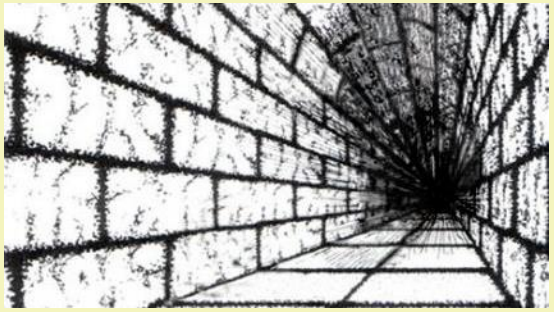
Snavely et al., 06-08

Lucas & Kanade, 81
Chen & Medioni, 92
Debevec et al., 96
Levoy & Hanrahan, 96
Fitzgibbon & Zisserman, 98
Triggs et al., 99
Pollefeys et al., 99
Kutulakos & Seitz, 99

Levoy et al., 00
Hartley & Zisserman, 00
Dellaert et al., 00
Rusinkiewic et al., 02
Nistér, 04
Brown & Lowe, 04
Schindler et al, 04
Lourakis & Argyros, 04
Colombo et al. 05

Golparvar-Fard, et al. JAEI 10
Pandey et al. IFAC , 2010
Pandey et al. ICRA 2011
Savarese et al. IJCV 05
Savarese et al. IJCV 06
Microsoft's PhotoSynth
Snavely et al., 06-08
Schindler et al., 08
Agarwal et al., 09
Frahm et al., 10

Current state of computer vision



3D Reconstruction

- 3D shape recovery
- 3D scene reconstruction
- Camera localization
- Pose estimation

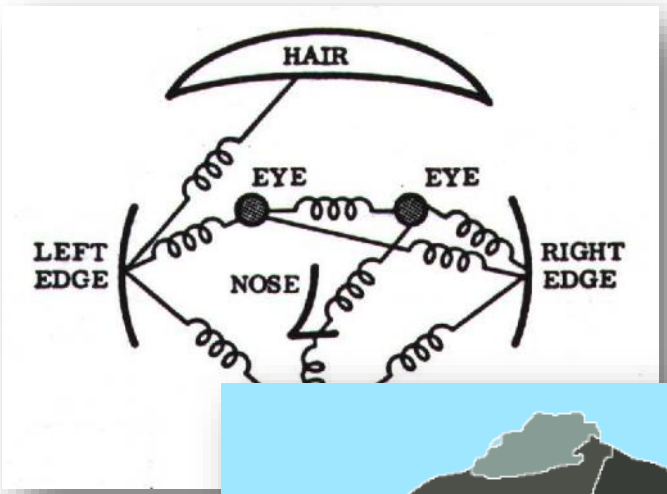
Snavely et al., 06-08

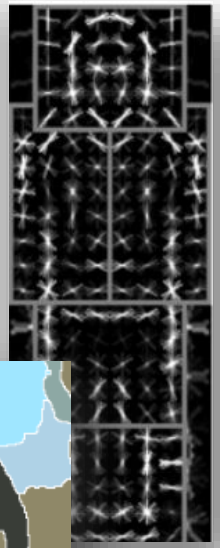
Lucas & Kanade, 81
Chen & Medioni, 92
Debevec et al., 96
Levoy & Hanrahan, 96
Fitzgibbon & Zisserman, 98
Triggs et al., 99
Pollefeys et al., 99
Kutulakos & Seitz, 99

Levoy et al., 00
Hartley & Zisserman, 00
Dellaert et al., 00
Rusinkiewic et al., 02
Nistér, 04
Brown & Lowe, 04
Schindler et al, 04
Lourakis & Argyros, 04
Colombo et al. 05

Golparvar-Fard, et al. JAEI 10
Pandey et al. IFAC , 2010
Pandey et al. ICRA 2011
Savarese et al. IJCV 05
Savarese et al. IJCV 06
Microsoft's PhotoSynth
Snavely et al., 06-08
Schindler et al., 08
Agarwal et al., 09
Frahm et al., 10

Current state of computer vision





2D Recognition

- Object detection
- Texture classification
- Target tracking
- Activity recognition

Turk & Pentland, 91
Poggio et al., 93
Belhumeur et al., 97
LeCun et al. 98
Amit and Geman, 99
Shi & Malik, 00
Viola & Jones, 00
Felzenszwalb & Huttenlocher 00
Belongie & Malik, 02
Ullman et al. 02

Argawal & Roth, 02
Ramanan & Forsyth, 03
Weber et al., 00
Vidal-Naquet & Ullman 02
Fergus et al., 03
Torralba et al., 03
Vogel & Schiele, 03
Barnard et al., 03
Fei-Fei et al., 04
Kumar & Hebert '04

He et al. 06
Gould et al. 08
Maire et al. 08
Felzenszwalb et al., 08
Kohli et al. 09
L.-J. Li et al. 09
Ladicky et al. 10,11
Gonfaus et al. 10
Farhadi et al., 09
Lampert et al., 09

Current state of computer vision

2D Recognition

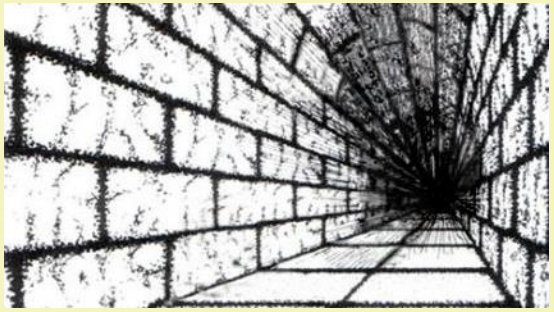
- Object detection
- Texture classification
- Target tracking
- Activity recognition

Turk & Pentland, 91
Poggio et al., 93
Belhumeur et al., 97
LeCun et al. 98
Amit and Geman, 99
Shi & Malik, 00
Viola & Jones, 00
Felzenszwalb & Huttenlocher 00
Belongie & Malik, 02
Ullman et al. 02

Argawal & Roth, 02
Ramanan & Forsyth, 03
Weber et al., 00
Vidal-Naquet & Ullman 02
Fergus et al., 03
Torralba et al., 03
Vogel & Schiele, 03
Barnard et al., 03
Fei-Fei et al., 04
Kumar & Hebert '04

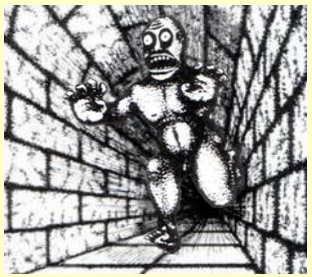
He et al. 06
Gould et al. 08
Maire et al. 08
Felzenszwalb et al., 08
Kohli et al. 09
L.-J. Li et al. 09
Ladicky et al. 10,11
Gonfaus et al. 10
Farhadi et al., 09
Lampert et al., 09

Current state of computer vision



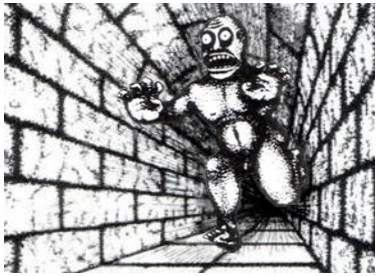
3D reconstruction

2D recognition



Perceiving the World in 3D

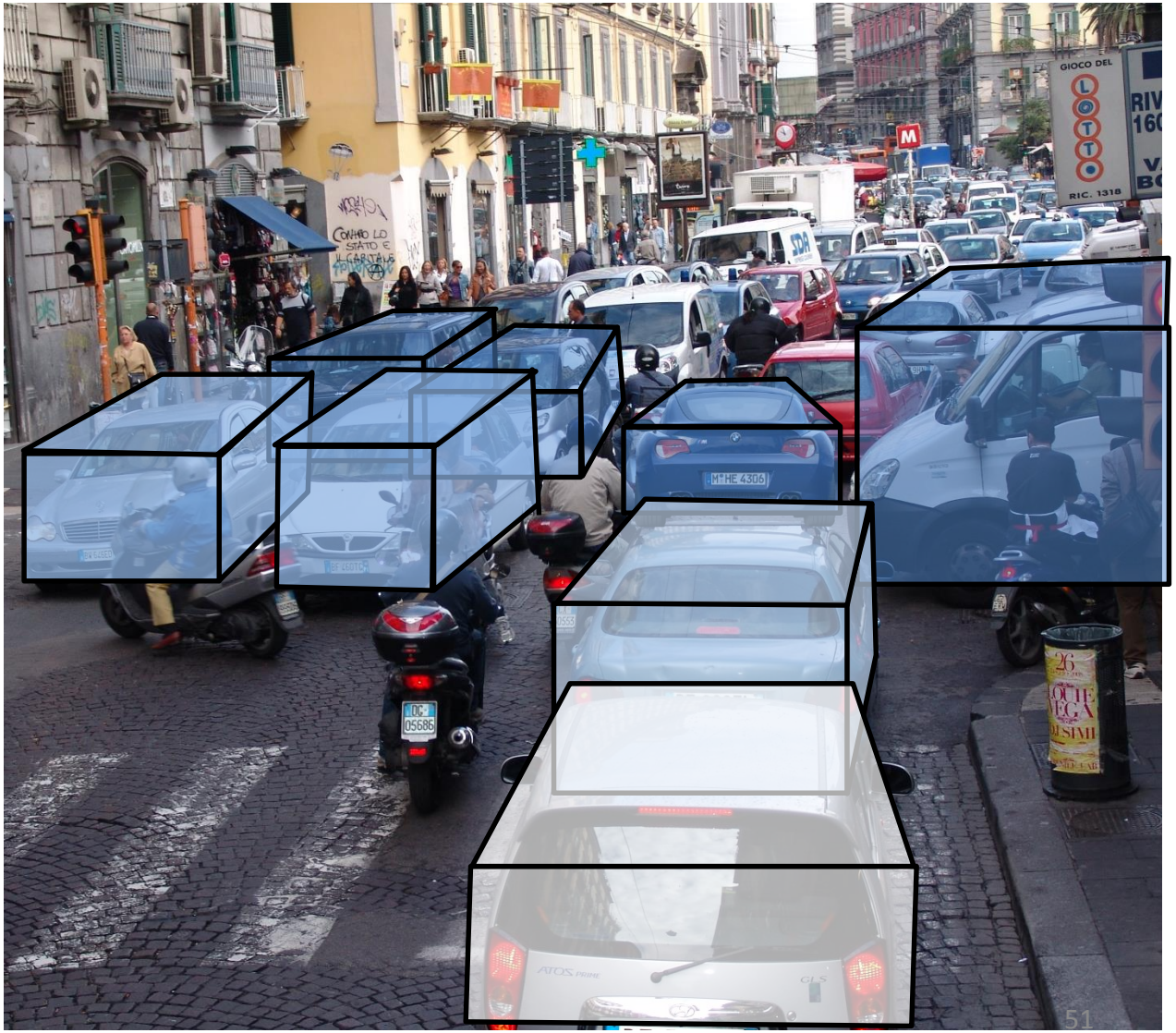
- Modeling objects and their 3D properties
- Modeling interaction among objects and space
- Modeling relationships of object/space across views



Outline

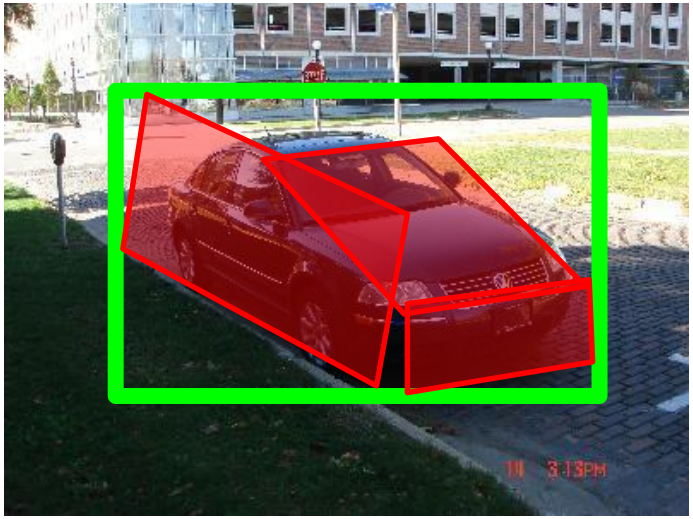
- Modeling objects and their 3D properties
- Modeling interaction among objects and space
- Modeling relationships of objects across views

Detecting objects and estimating their 3D properties

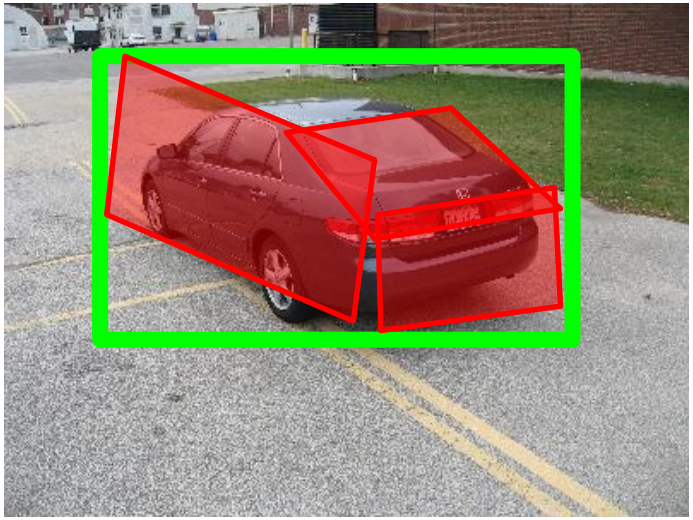


Results

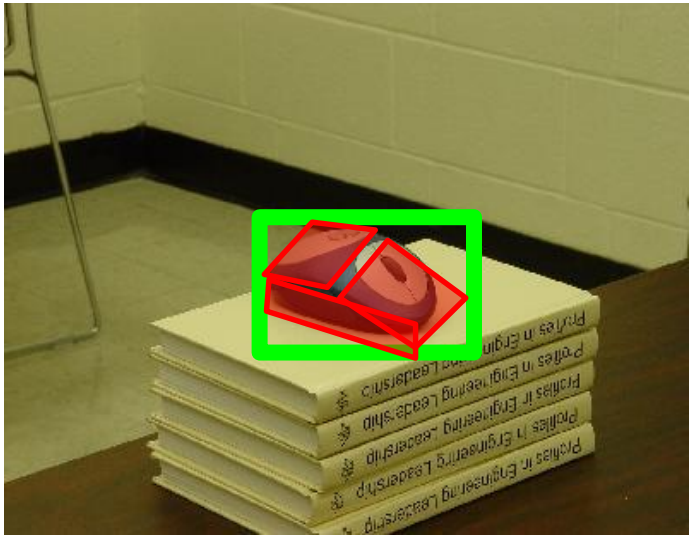
CAR a=330 e=15 d=7



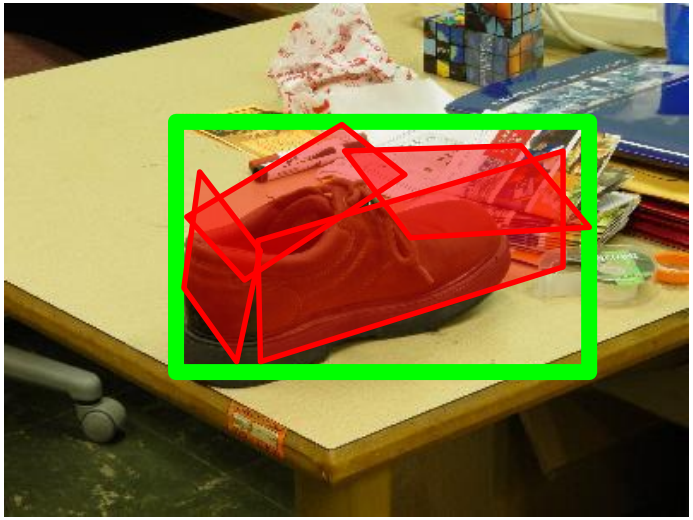
CAR a=150 e=15 d=7



MOUSE a=300 e=45 d=23



SHOE a=240 e=45 d=11



3D object dataset [Savarese & Fei-Fei 07]

Results

CHAIR a=0 e=30 d=7

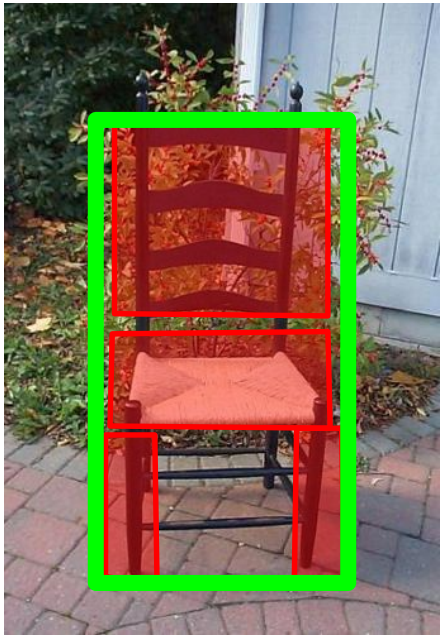
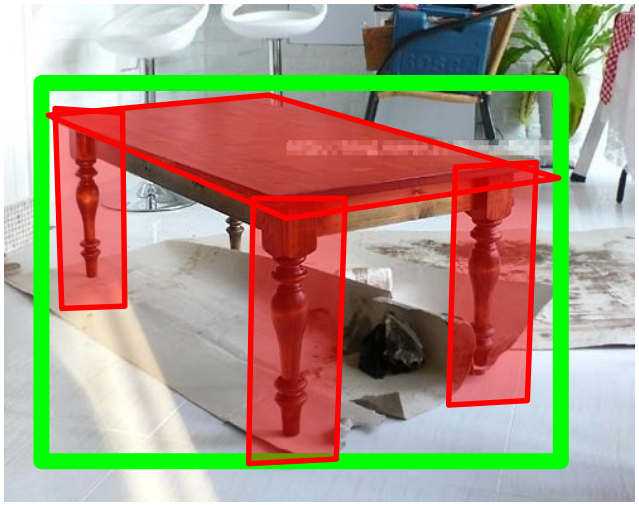
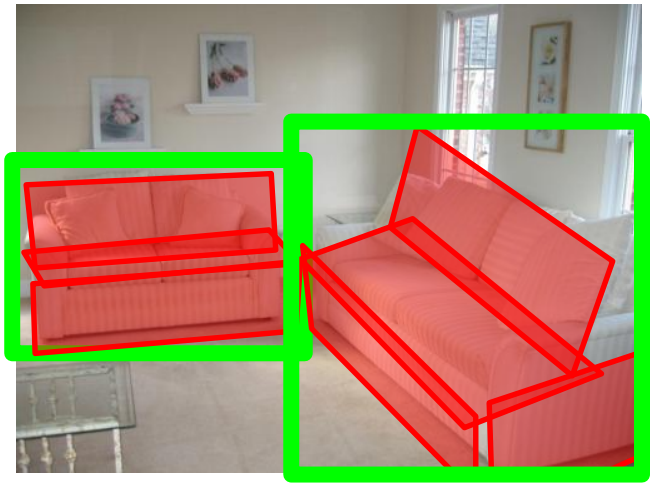


TABLE a=60 e=15 d=2



SOFA a=345 e=15 d=3.5
a=60 e-30 d=2.5

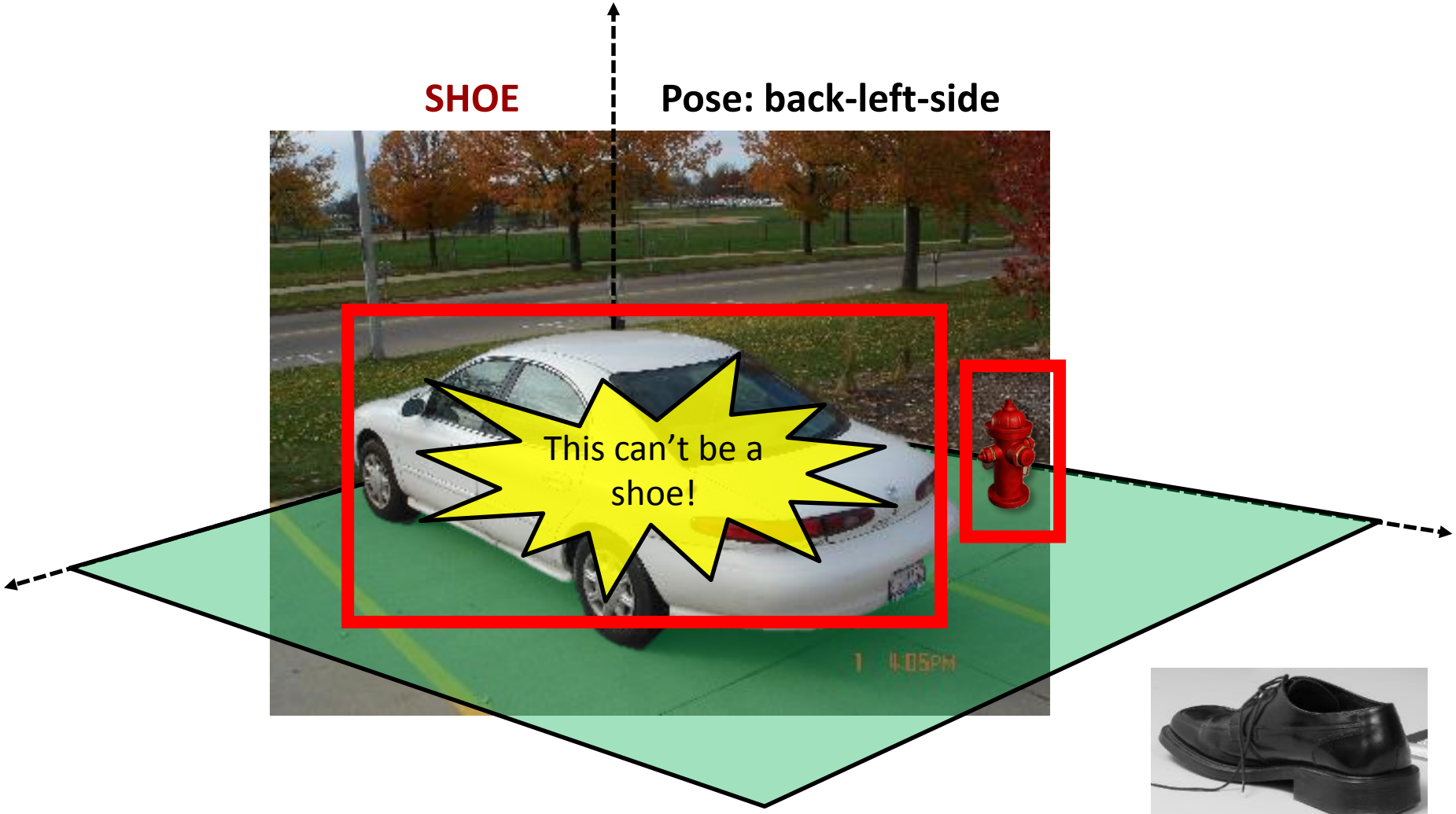


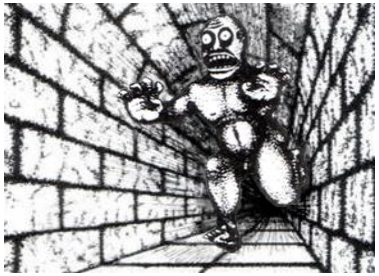
BED a=30 e=15 d=2.5

ImageNet dataset [Deng et al. 2010]

Results

Examples of failure (wrong category)

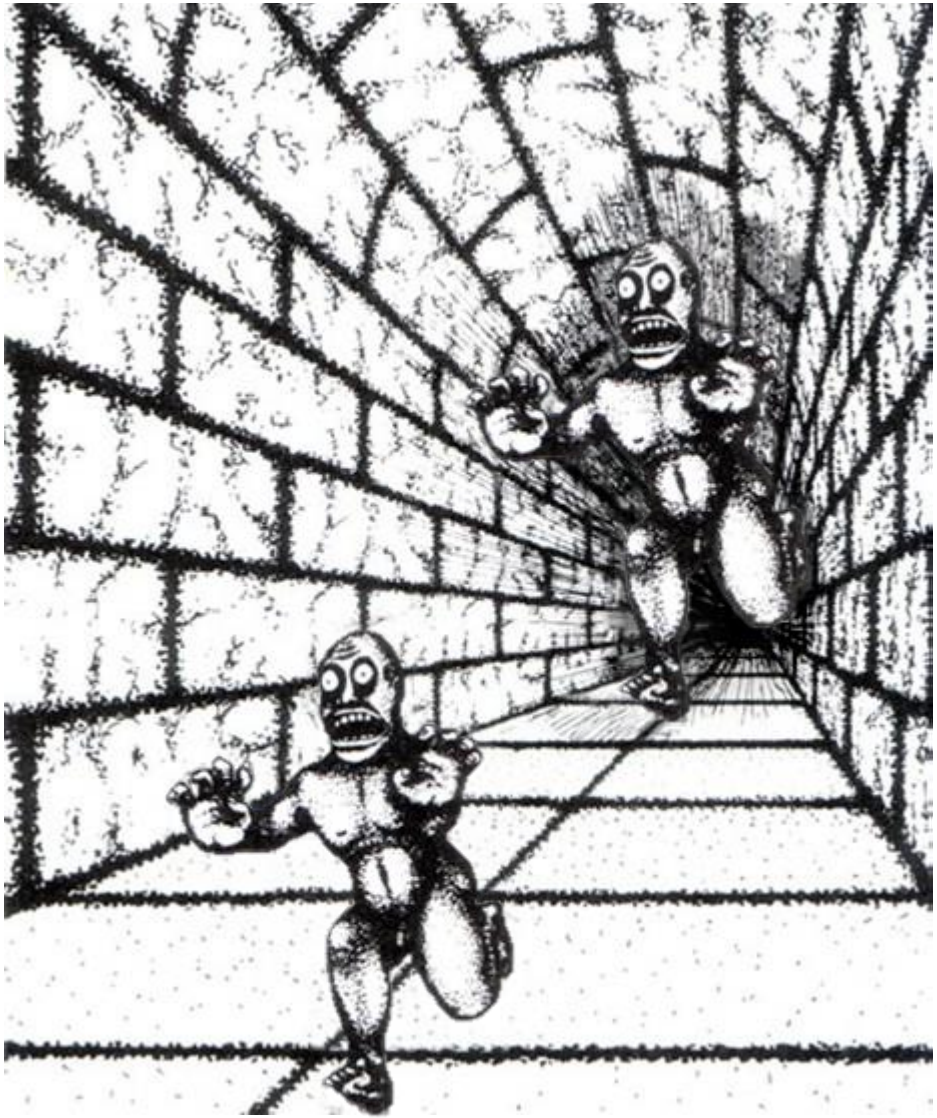




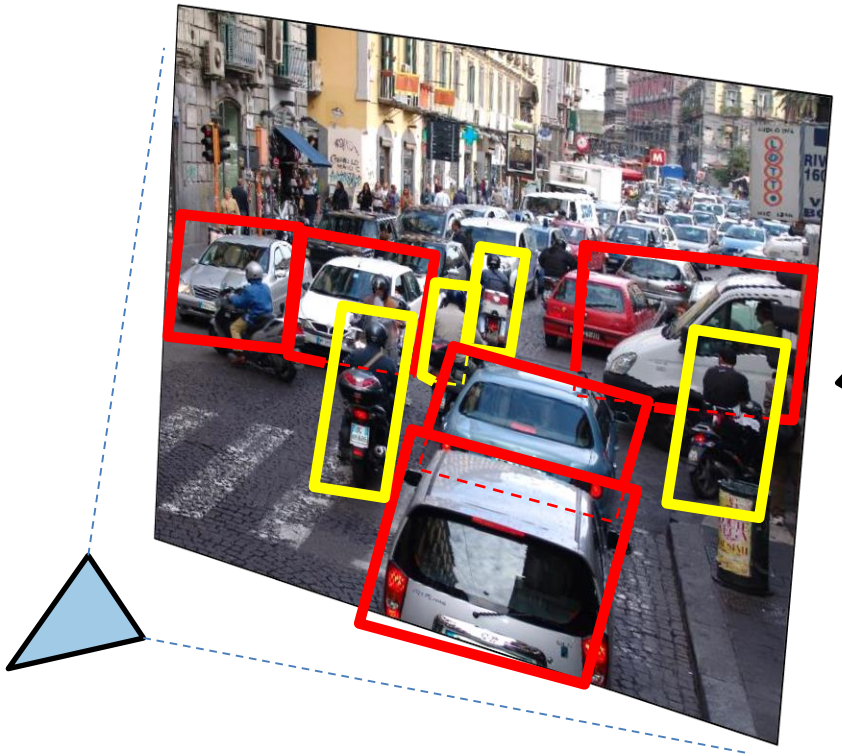
Outline

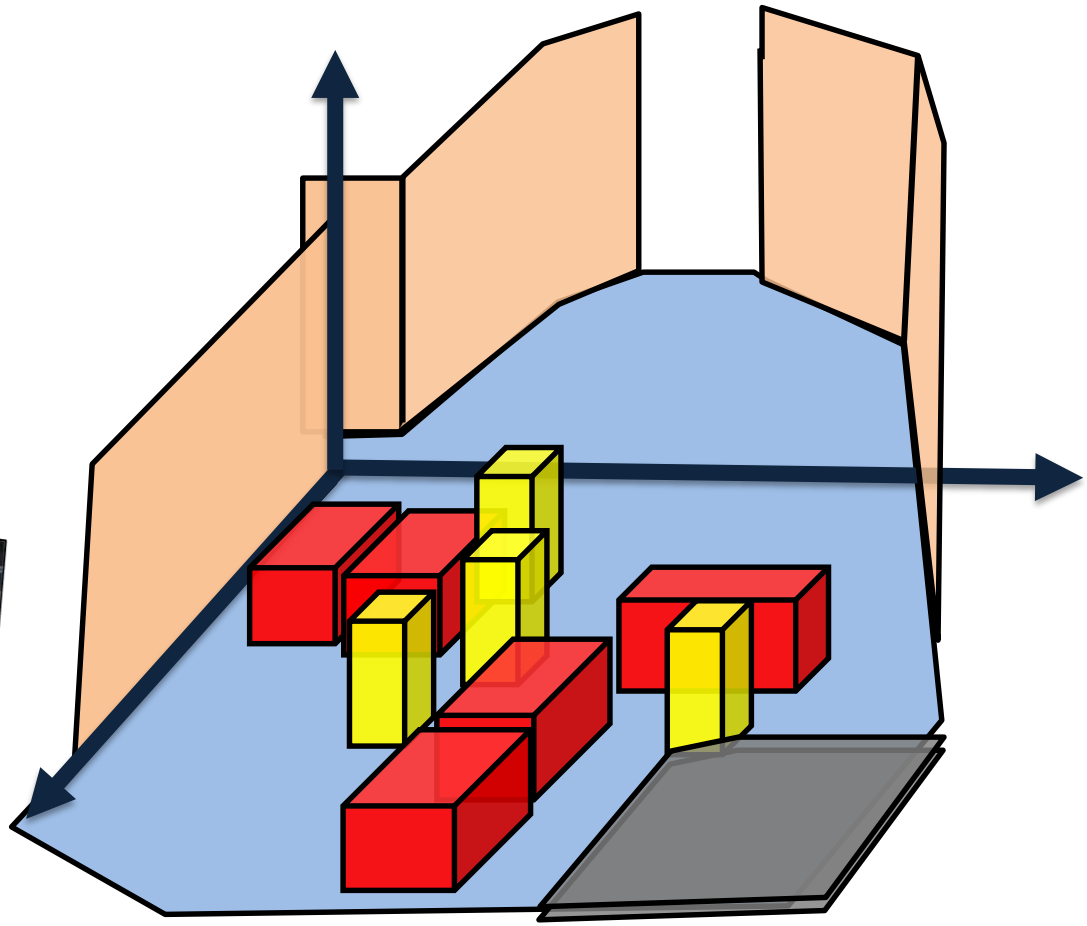
- Modeling objects and their 3D properties
- Modeling interaction among objects and space
- Modeling relationships of objects across views

Scene understanding is an interplay between objects and space

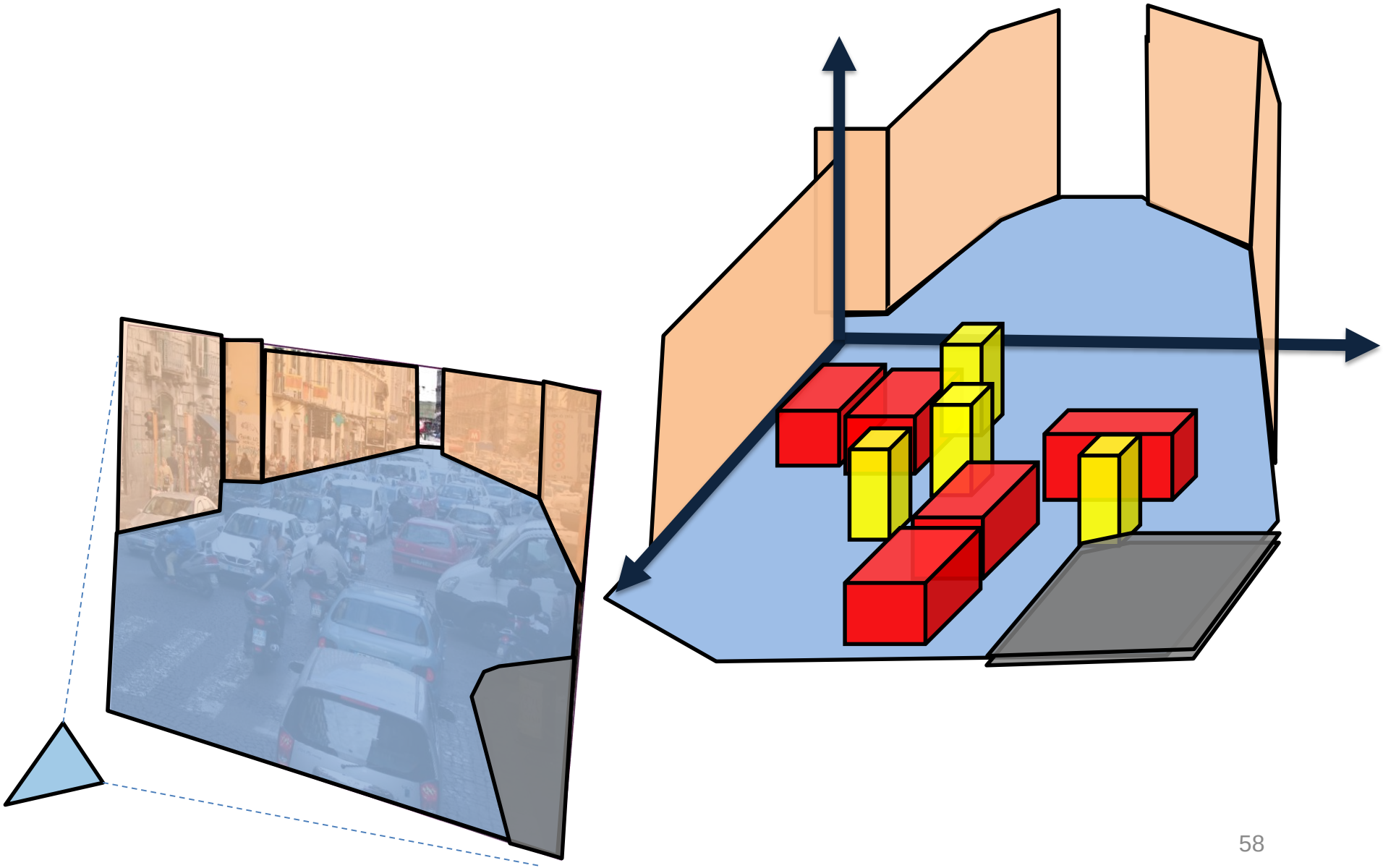


3D space is shaped by its objects

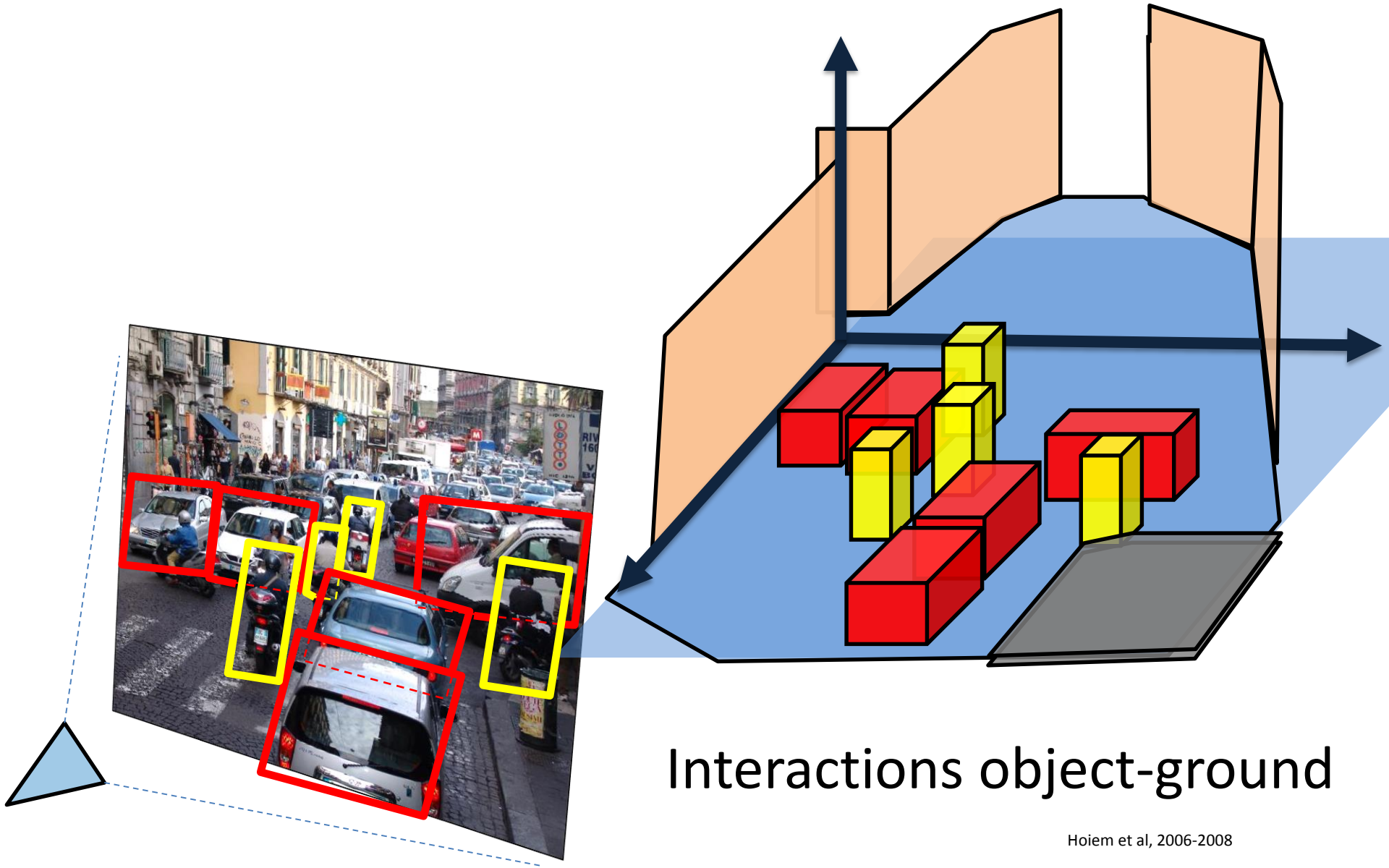




Objects are placed into 3D space



A first attempt….

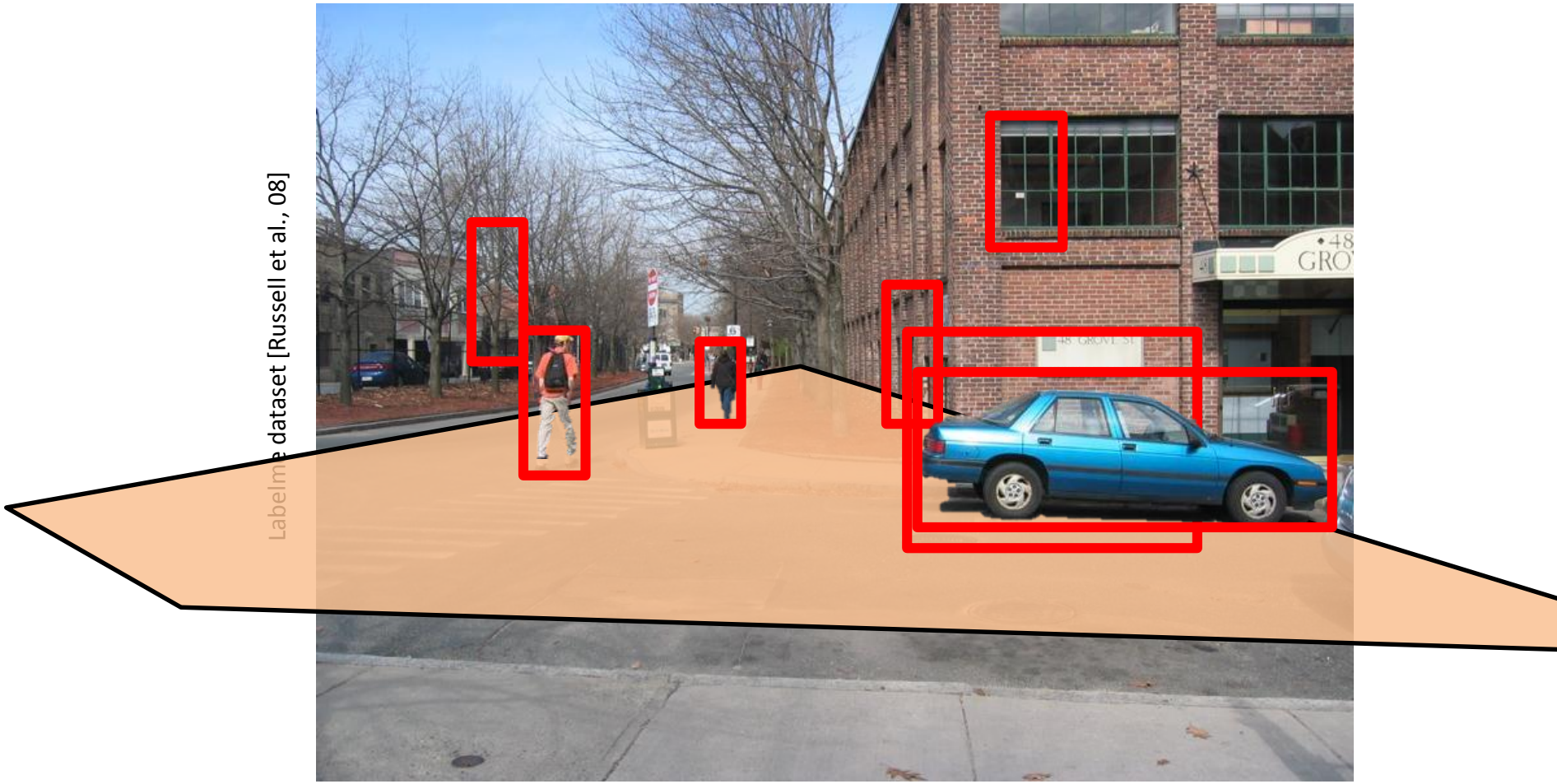


Interactions object-ground

Hoiem et al, 2006-2008

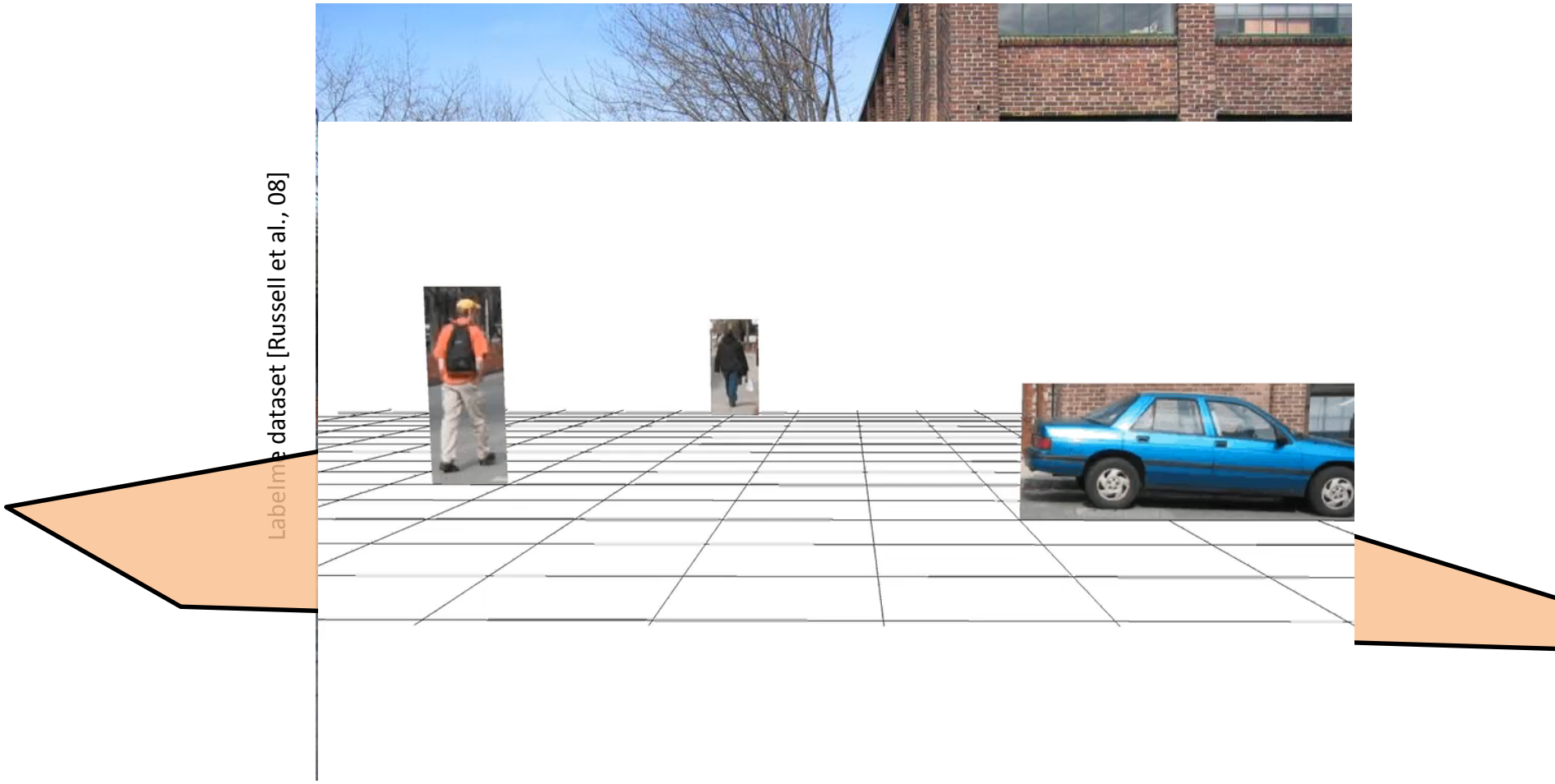
A first attempt….

Bao et al. CVPR 2010;
BMVC 2010; CIVC 2011; IJCV 2012



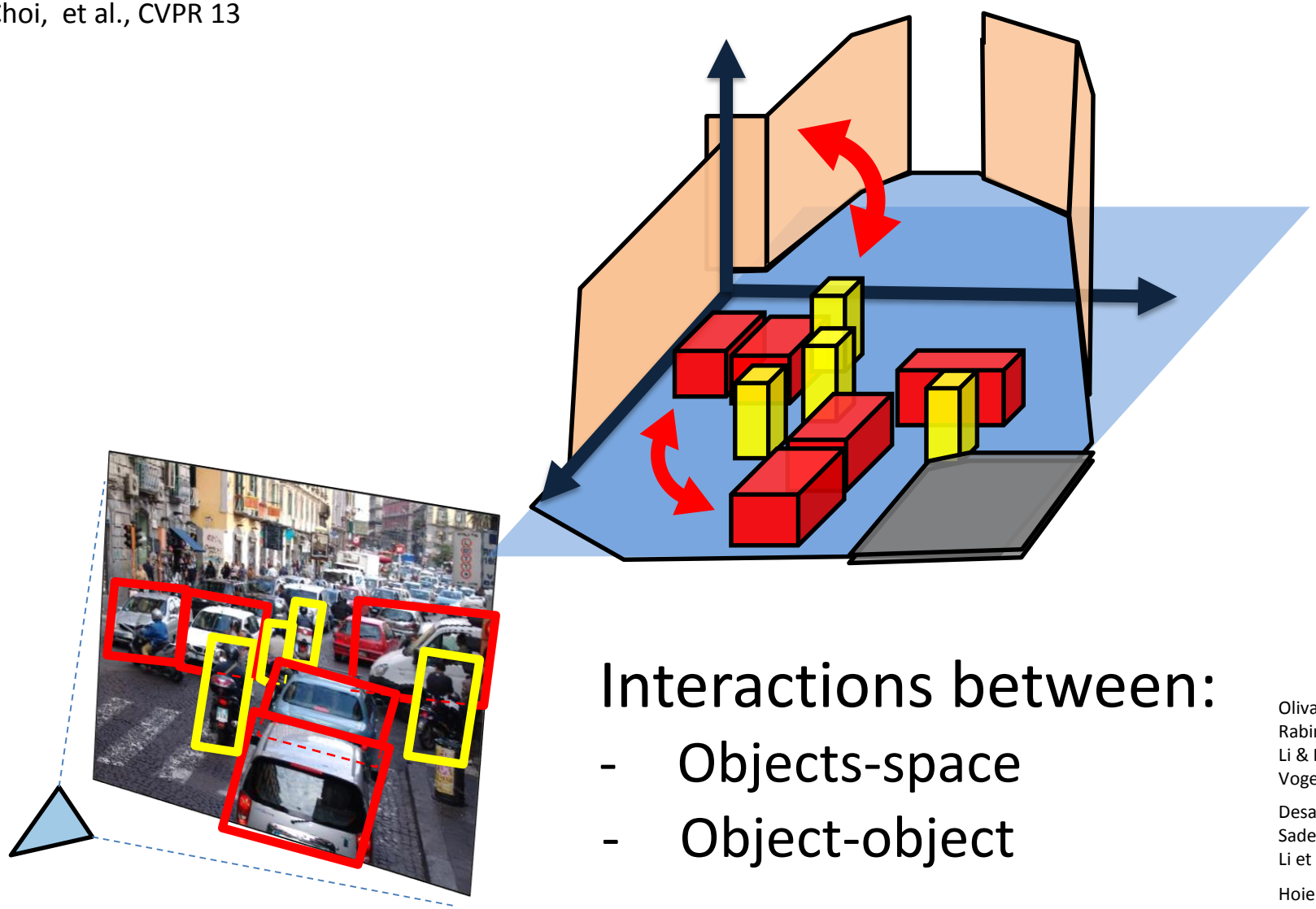
A first attempt….

Bao, et al. CVPR 2010;
BMVC 2010;
CIVC 2011 **(editor choice)**
IJCV 2012



Generalization #1

Choi, et al., CVPR 13



Interactions between:

- Objects-space
- Object-object

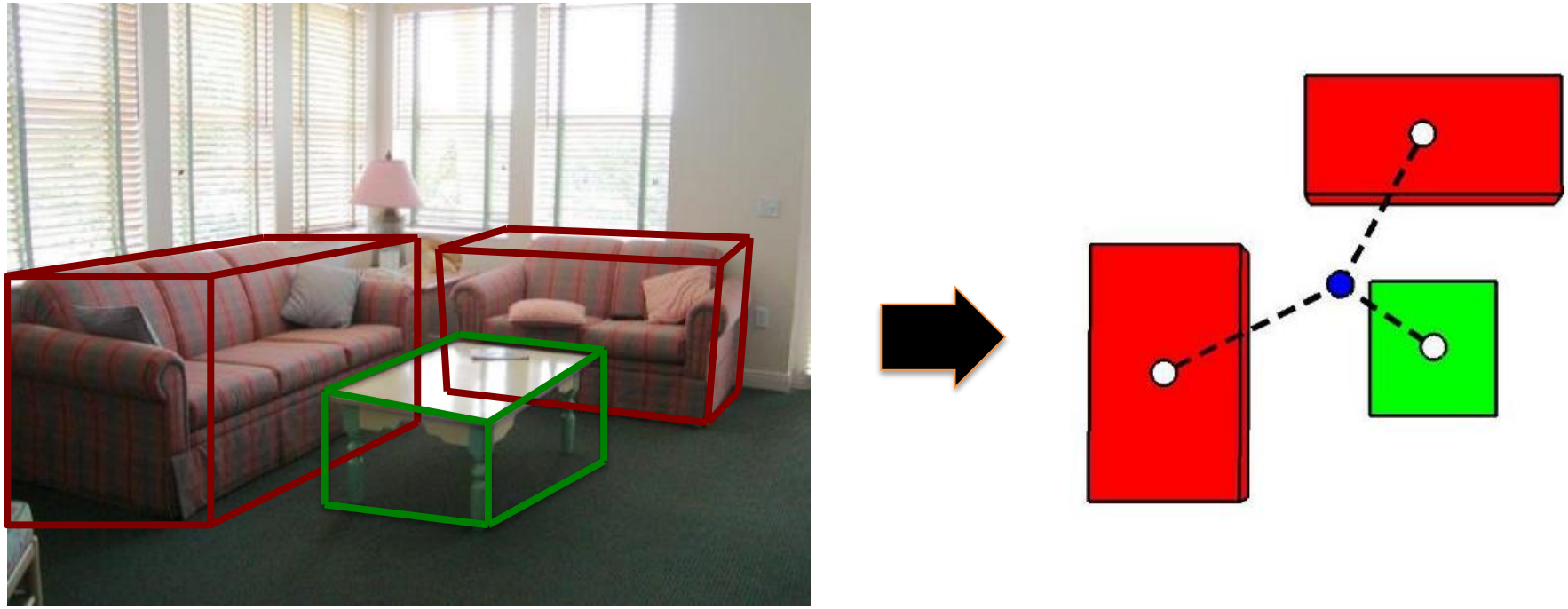
Oliva & Torralba, 2007
Rabinovich et al, 2007
Li & Fei-Fei, 2007
Vogel & Schiele, 2007

Desai et al, 2009
Sadeghi & Farhardi, 2011
Li et al, 2012

Hoiem et al, 2006
Herdau et al., 2009
Gupta et al, 2010
Fouhey et al, 2012

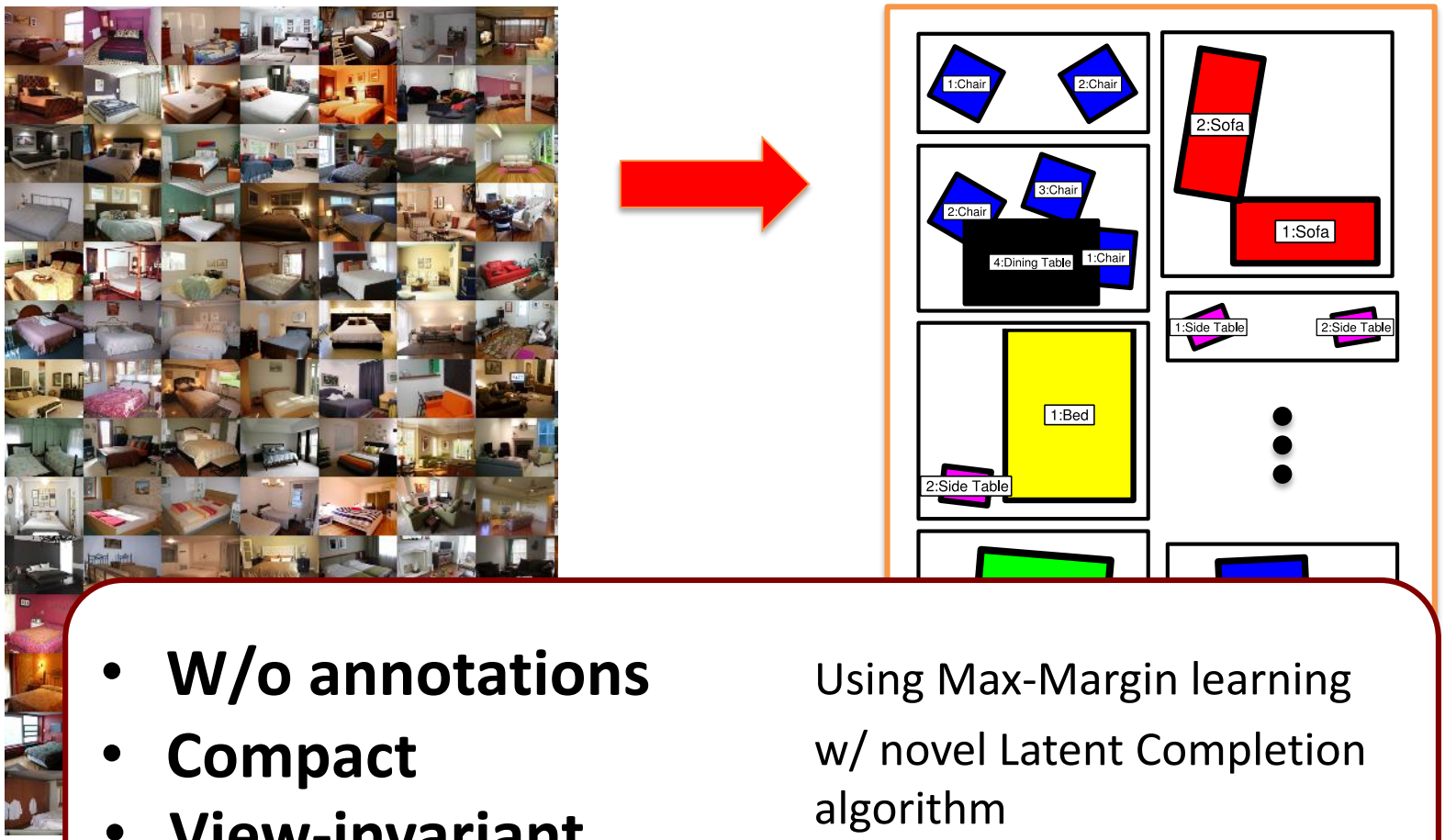
3D Geometric Phrases

A ***3DGP*** encodes **geometric** and **semantic** relationships between groups of objects and space elements which frequently co-occur in **spatially consistent configurations**.



3D Geometric Phrases

Choi, Chao, Pantofaru, Savarese, CVPR 13



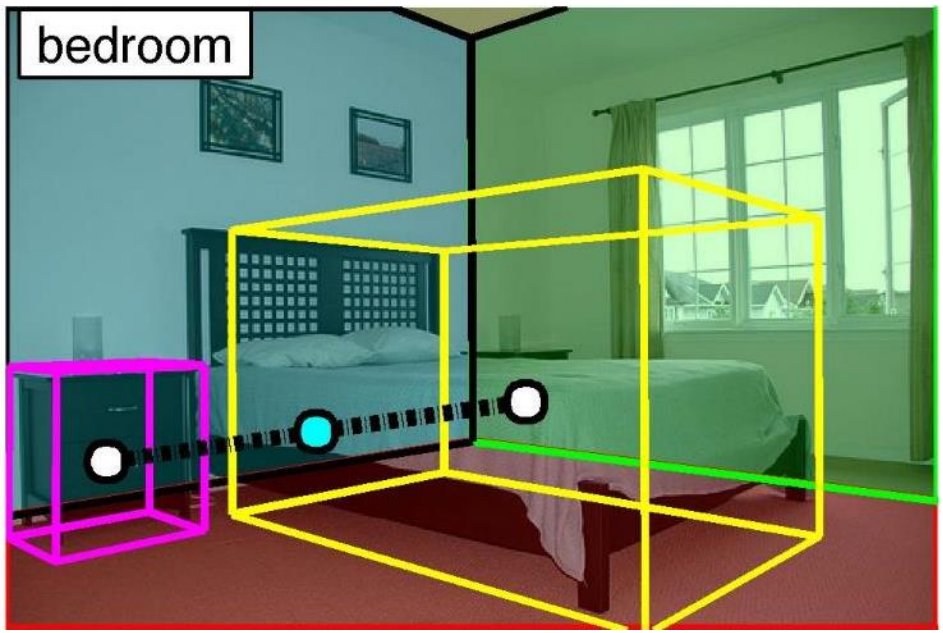
- **W/o annotations**
- **Compact**
- **View-invariant**

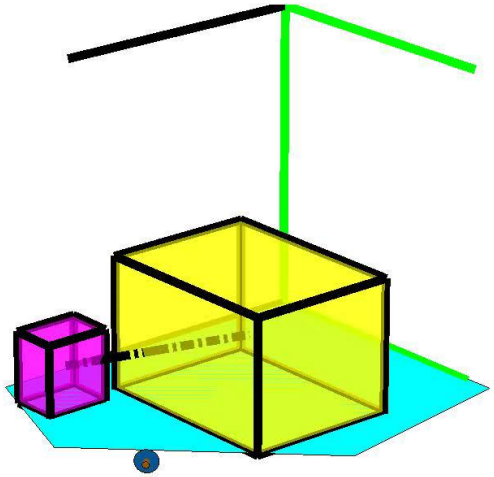
Using Max-Margin learning

w/ novel Latent Completion algorithm

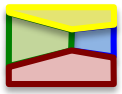
Results

Indoor scene dataset [Choi et al., 12]

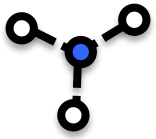




Sofa, **Coffee Table**, **Chair**, **Bed**, **Dining Table**, **Side Table**



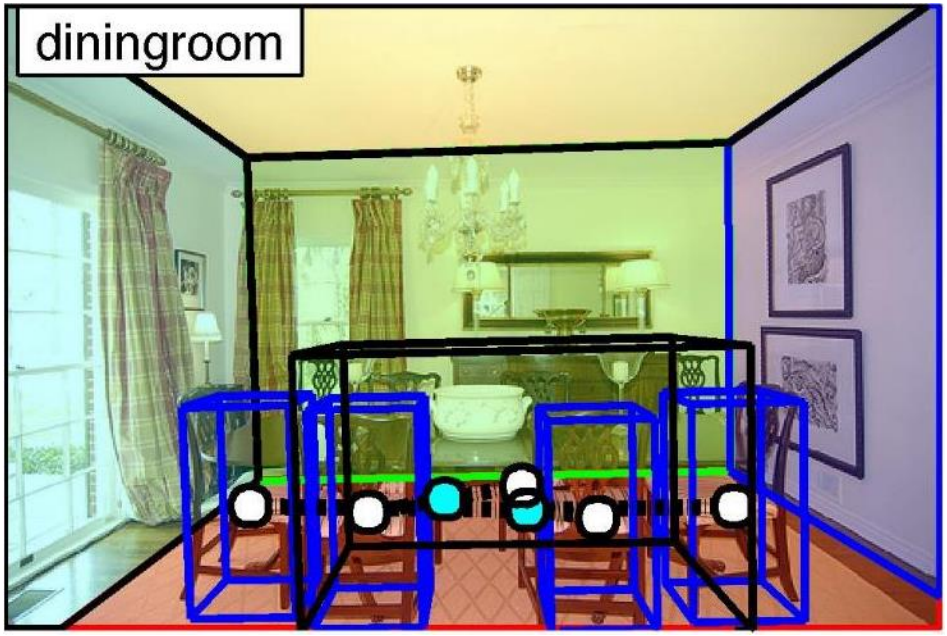
Estimated Layout

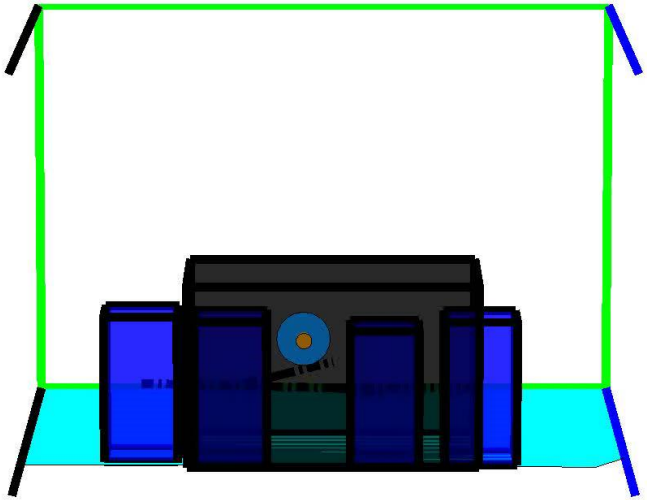


3D Geometric Phrases

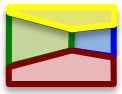
Results

Indoor scene dataset [Choi et al., 12]

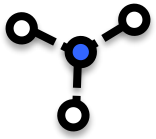




Sofa, **Coffee Table**, **Chair**, **Bed**, **Dining Table**, **Side Table**

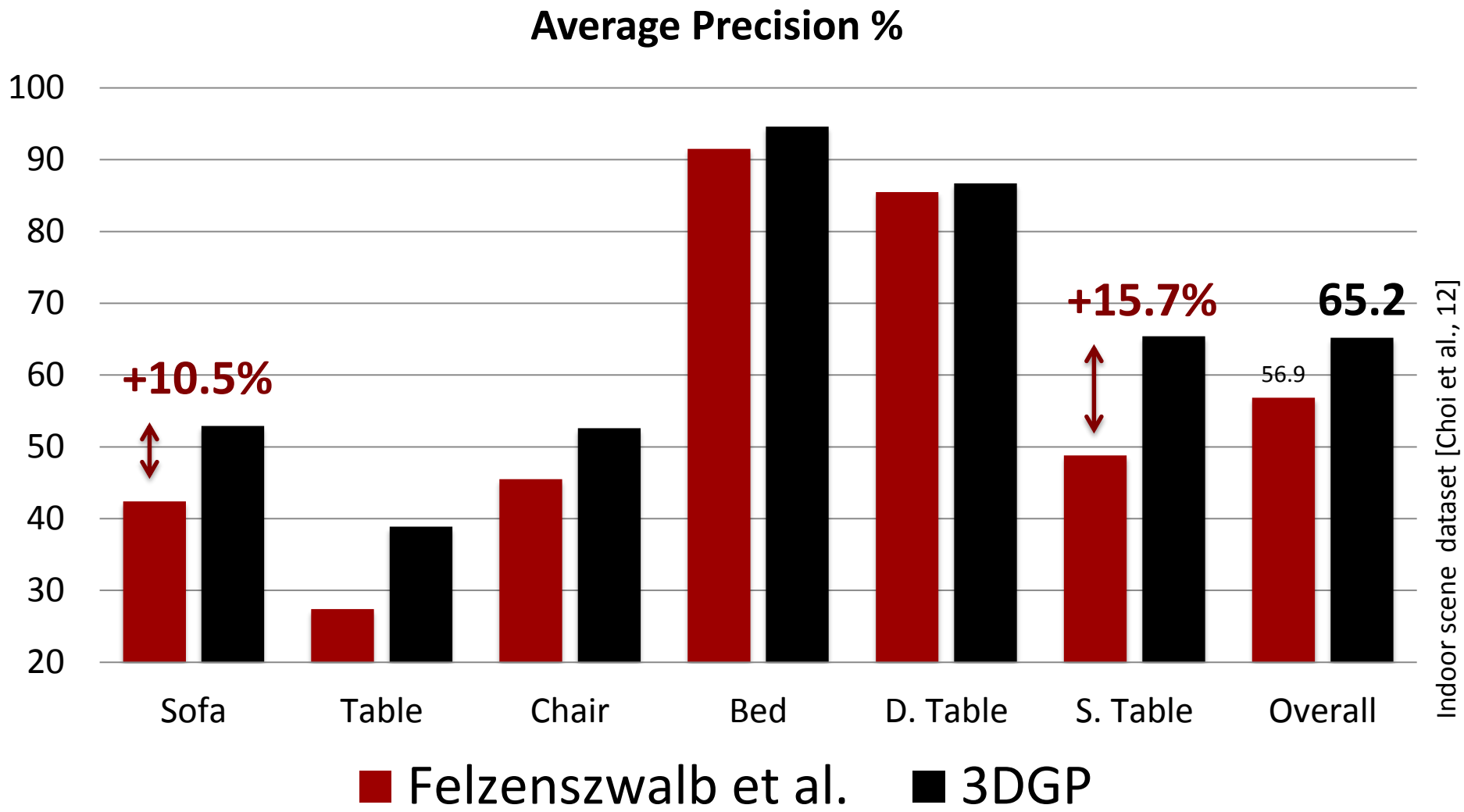


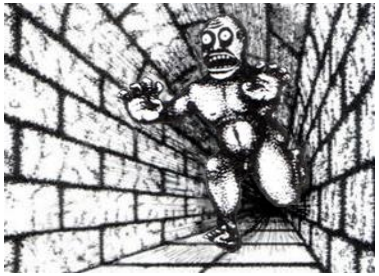
Estimated Layout



3D Geometric Phrases

Results: Object Detection

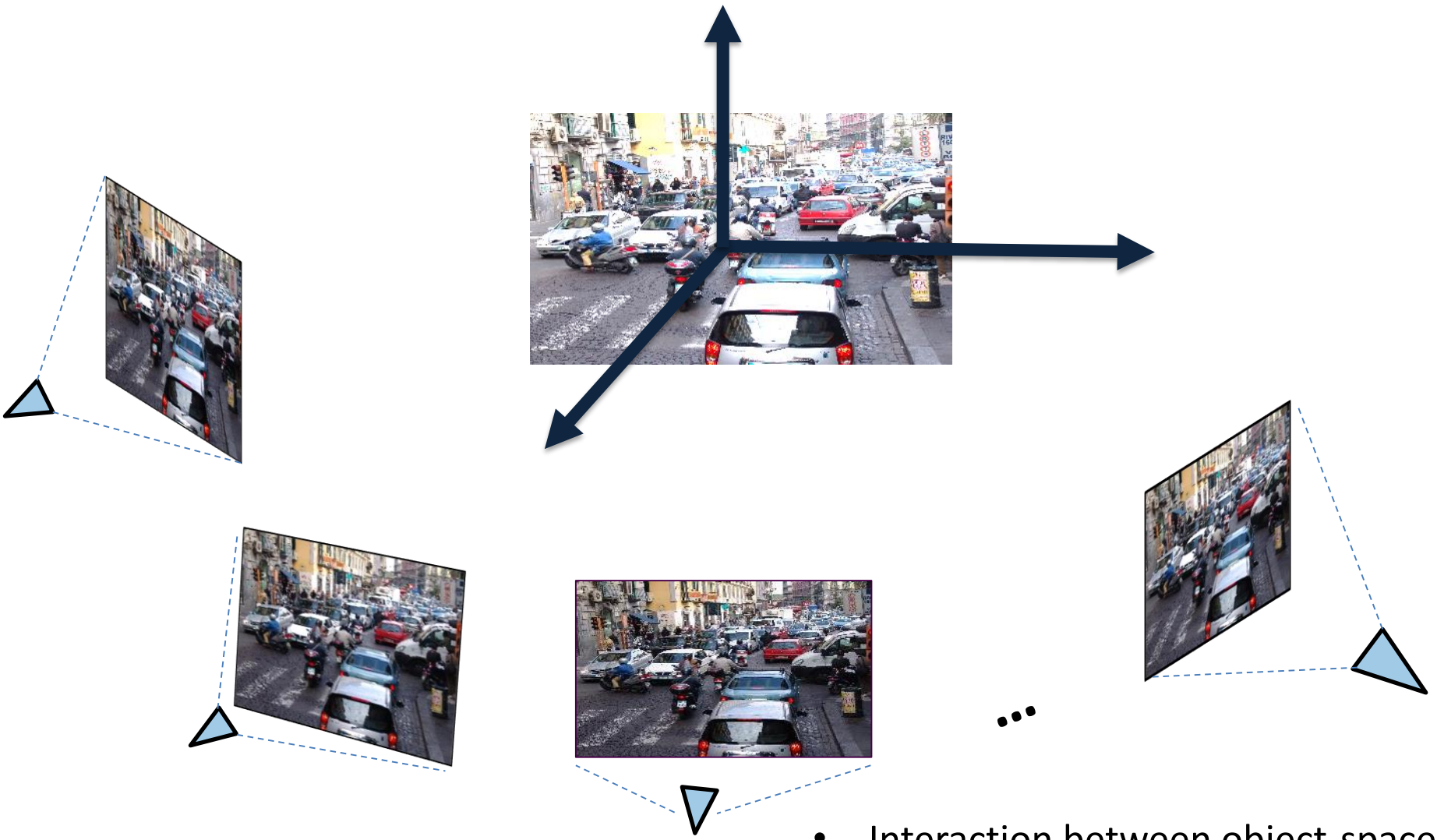




Outline

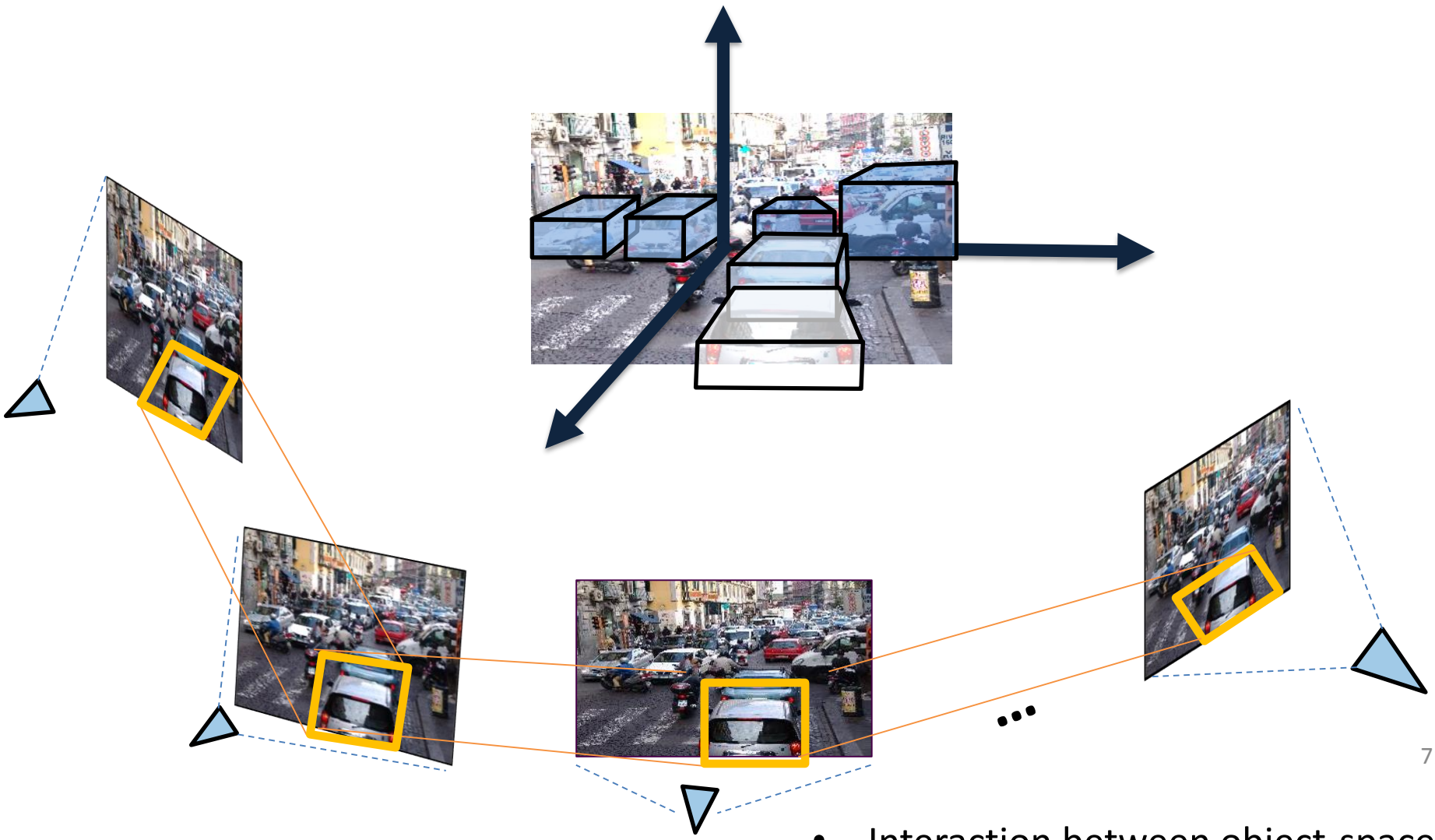
- Modeling objects and their 3D properties
- Modeling interaction among objects and space
- Modeling relationships of objects across views

Modeling relationships of objects across views



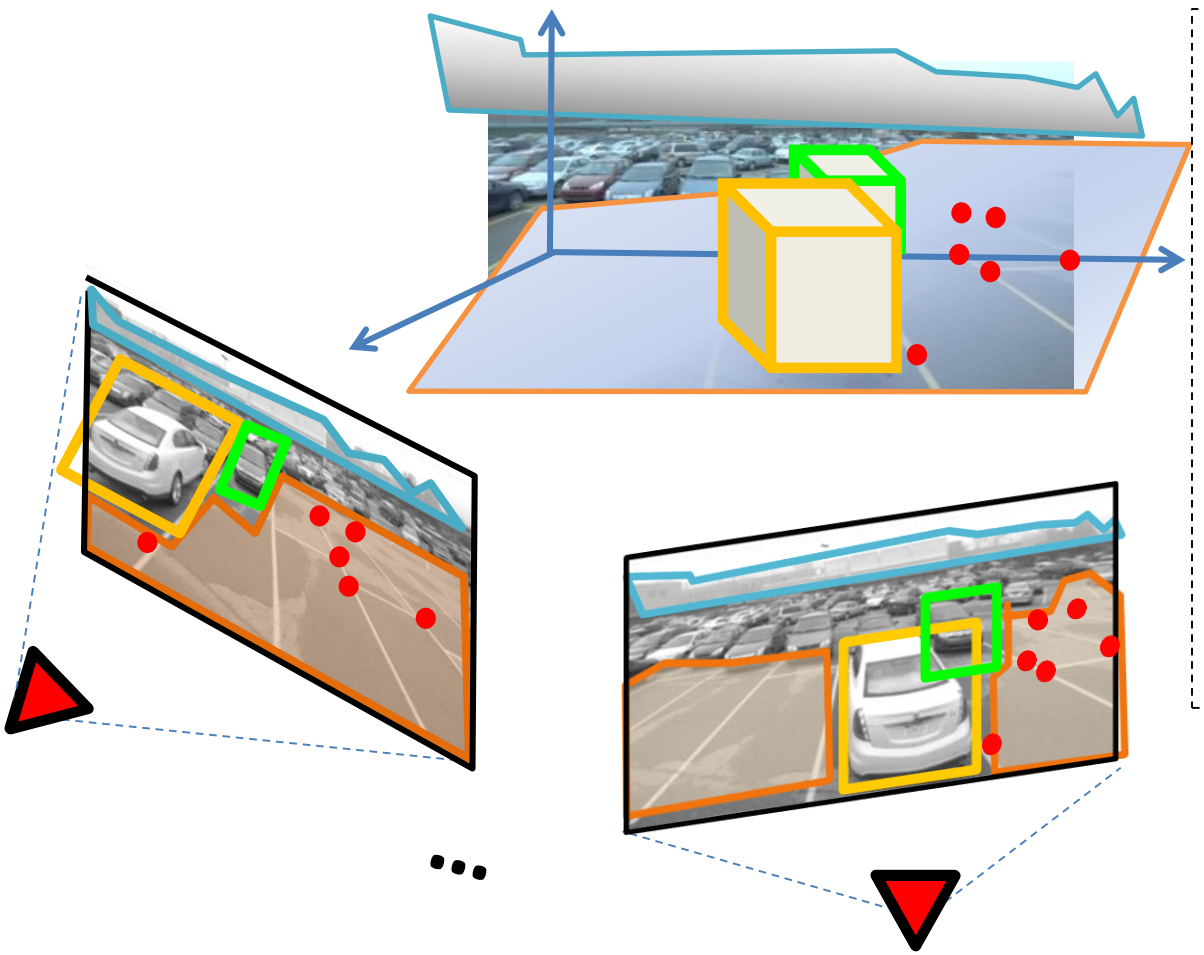
- Interaction between object-space
- Interaction among objects
- **Transfer semantics across views**

Modeling relationships of objects across views



- Interaction between object-space
- Interaction among objects
- **Transfer semantics across views**

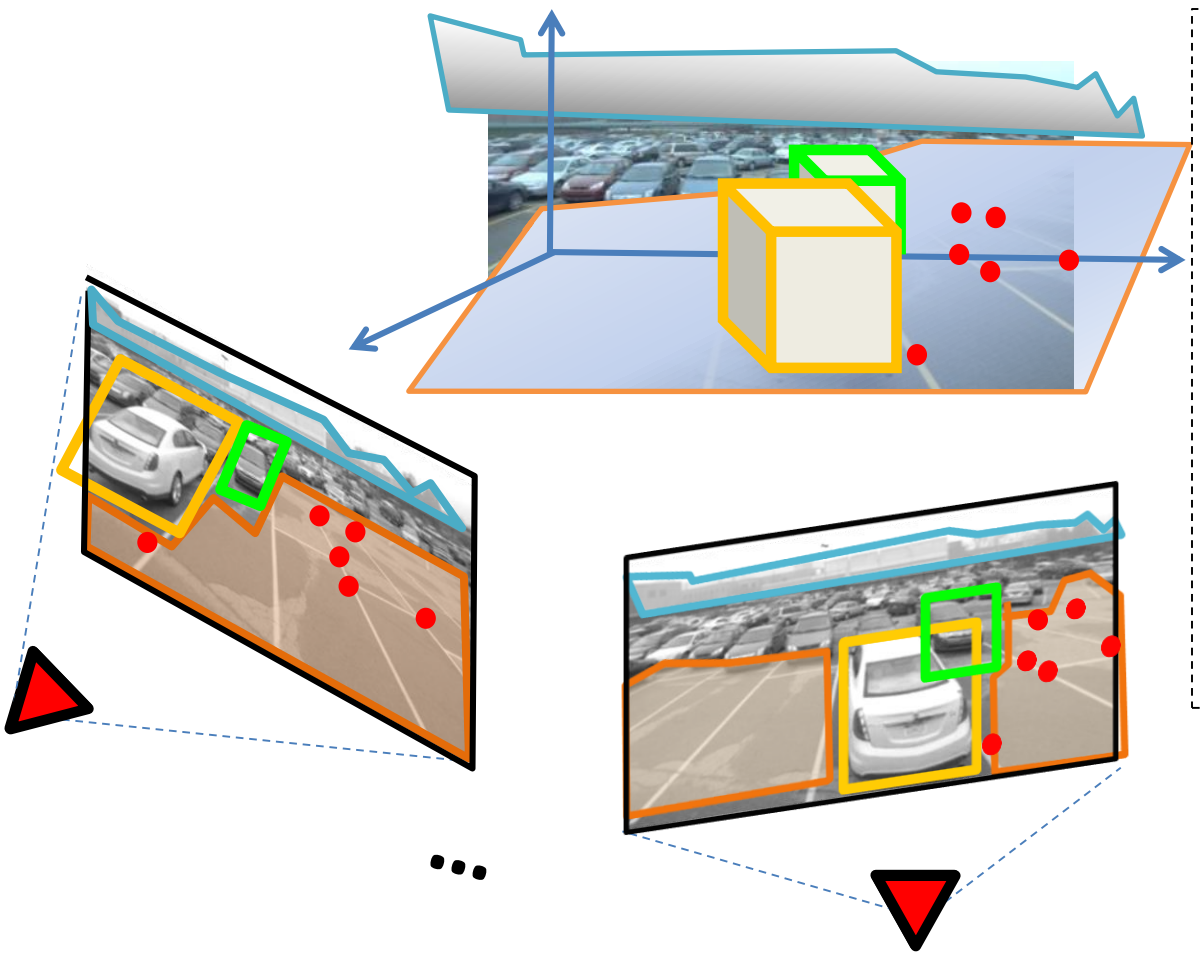
Semantic structure from motion



- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

Semantic structure from motion

$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg\max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \Psi(\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}; \mathbf{I})$$

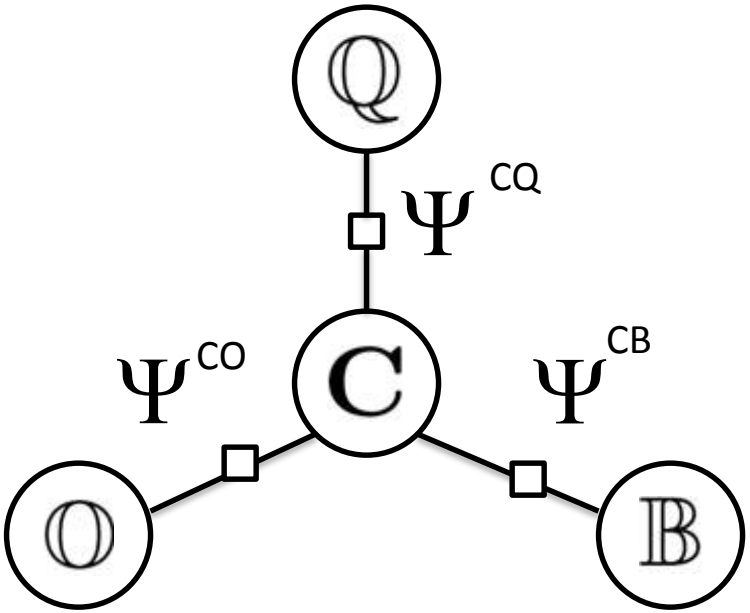


- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

Semantic structure from motion

$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg\max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \prod_s \Psi_s^{CQ} \prod_t \Psi_t^{CO} \prod_r \Psi_r^{CB}$$

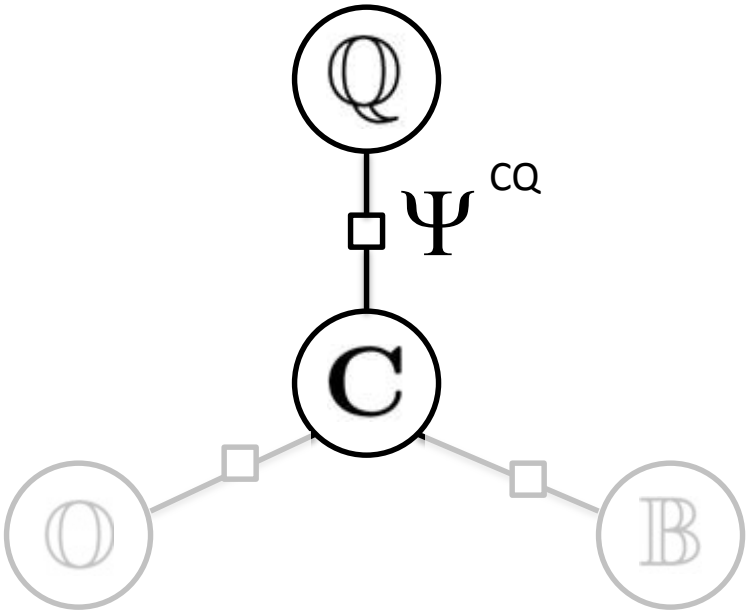
Factor graph



- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

SSFM: point-level compatibility

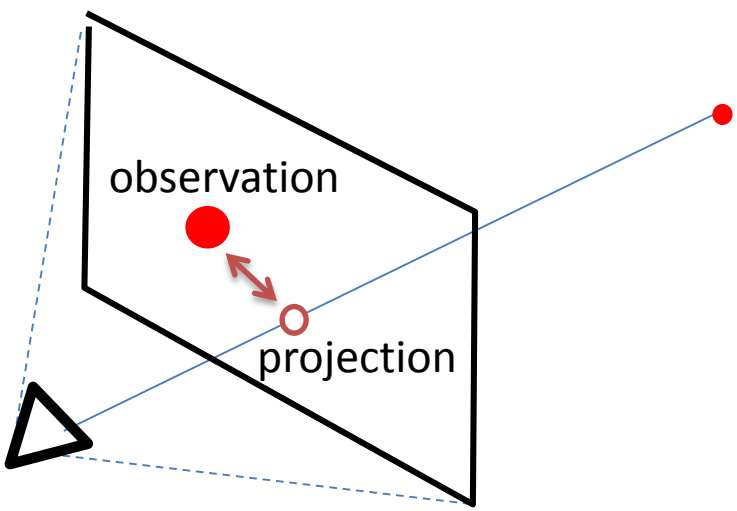
$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg\max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \boxed{\prod_s \Psi_s^{CQ}} \prod_t \Psi_t^{CO} \prod_r \Psi_r^{CB}$$



- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

SSFM: point-level compatibility

$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg\max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \boxed{\prod_s \Psi_s^{CQ}} \prod_t \Psi_t^{CO} \prod_r \Psi_r^{CB}$$



- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

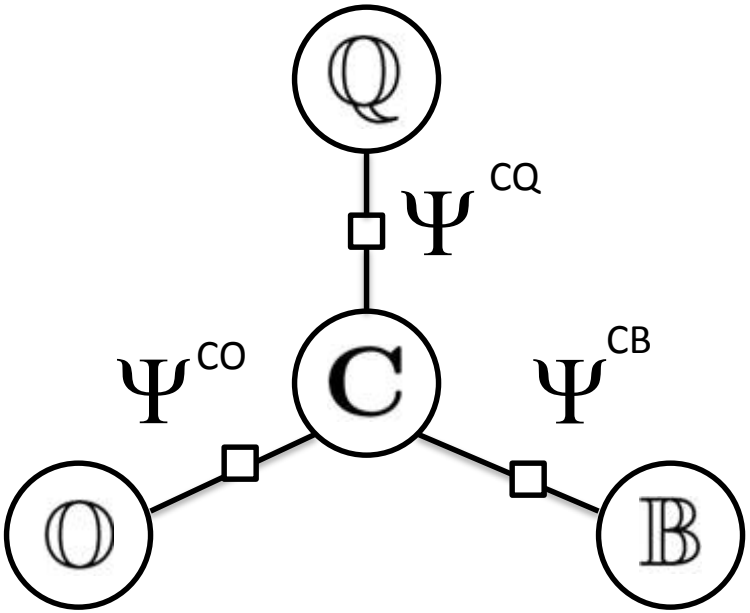
Point re-projection error

$$\prod_s \Psi_s^{CQ} \propto \prod_i^{N_Q} \prod_k^{N_k} \exp(-(q_i^k - q_{u_i^k}^k)^2/\sigma_q)$$

- Tomasi & Kanade '92
- Triggs et al '99
- Soatto & Perona 99
- Hartley & Zisserman 00
- Dellaert et al. 00
- Pollefeys & V. Gool 02
- Nister 04
- Brown & Lowe 07
- Snavely et al. 08

SSFM: Object-level compatibility

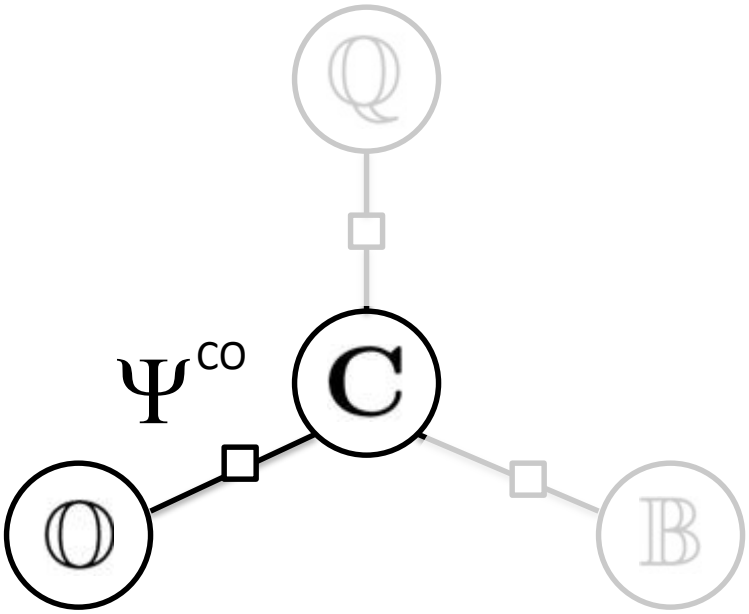
$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg\max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \prod_s \Psi_s^{CQ} \boxed{\prod_t \Psi_t^{CO} \prod_r \Psi_r^{CB}}$$



- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

SSFM: Object-level compatibility

$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg\max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \prod_s \Psi_s^{CQ} \boxed{\prod_t \Psi_t^{CO}} \prod_r \Psi_r^{CB}$$

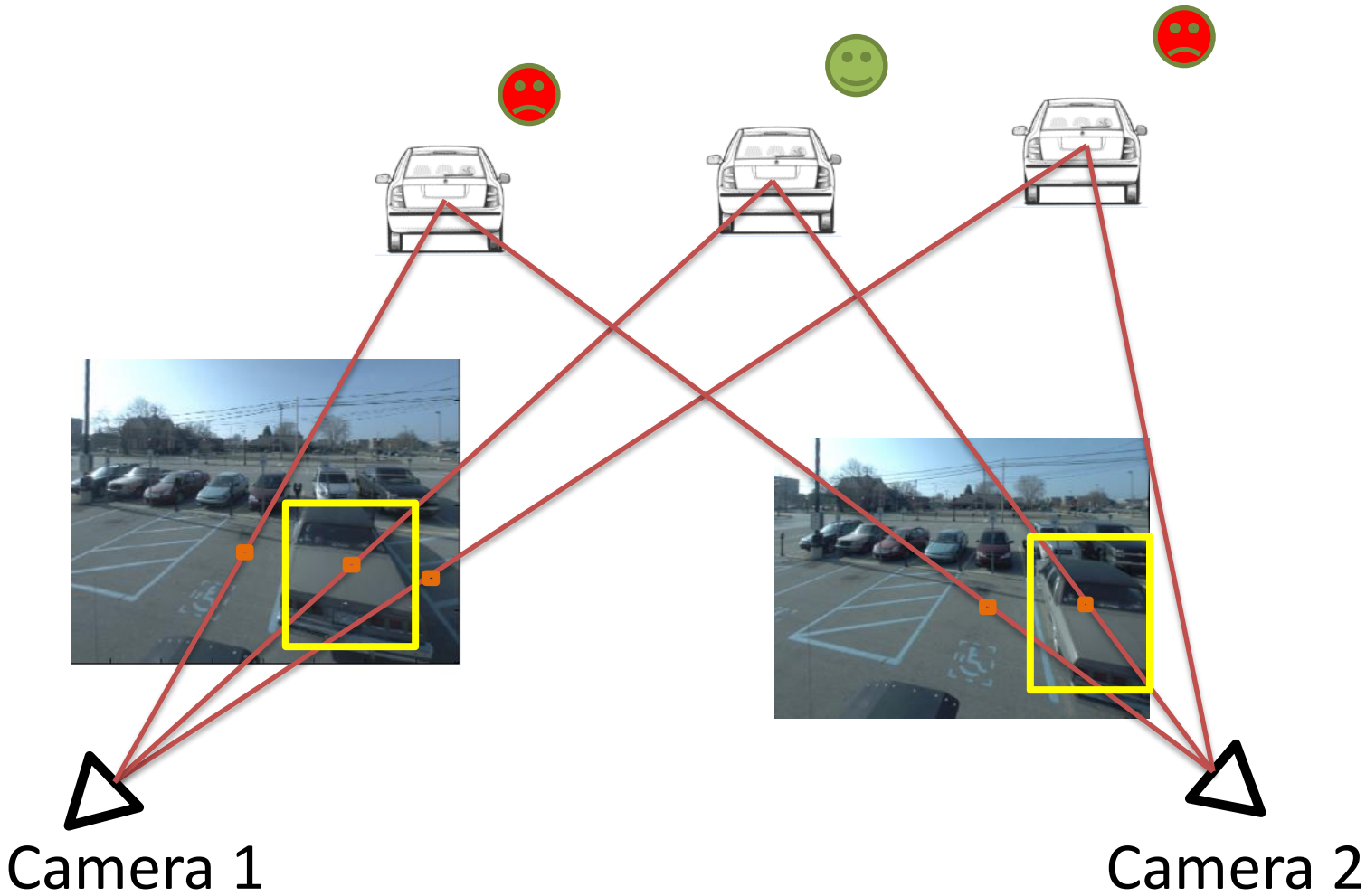


- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

Object "re-projection" error

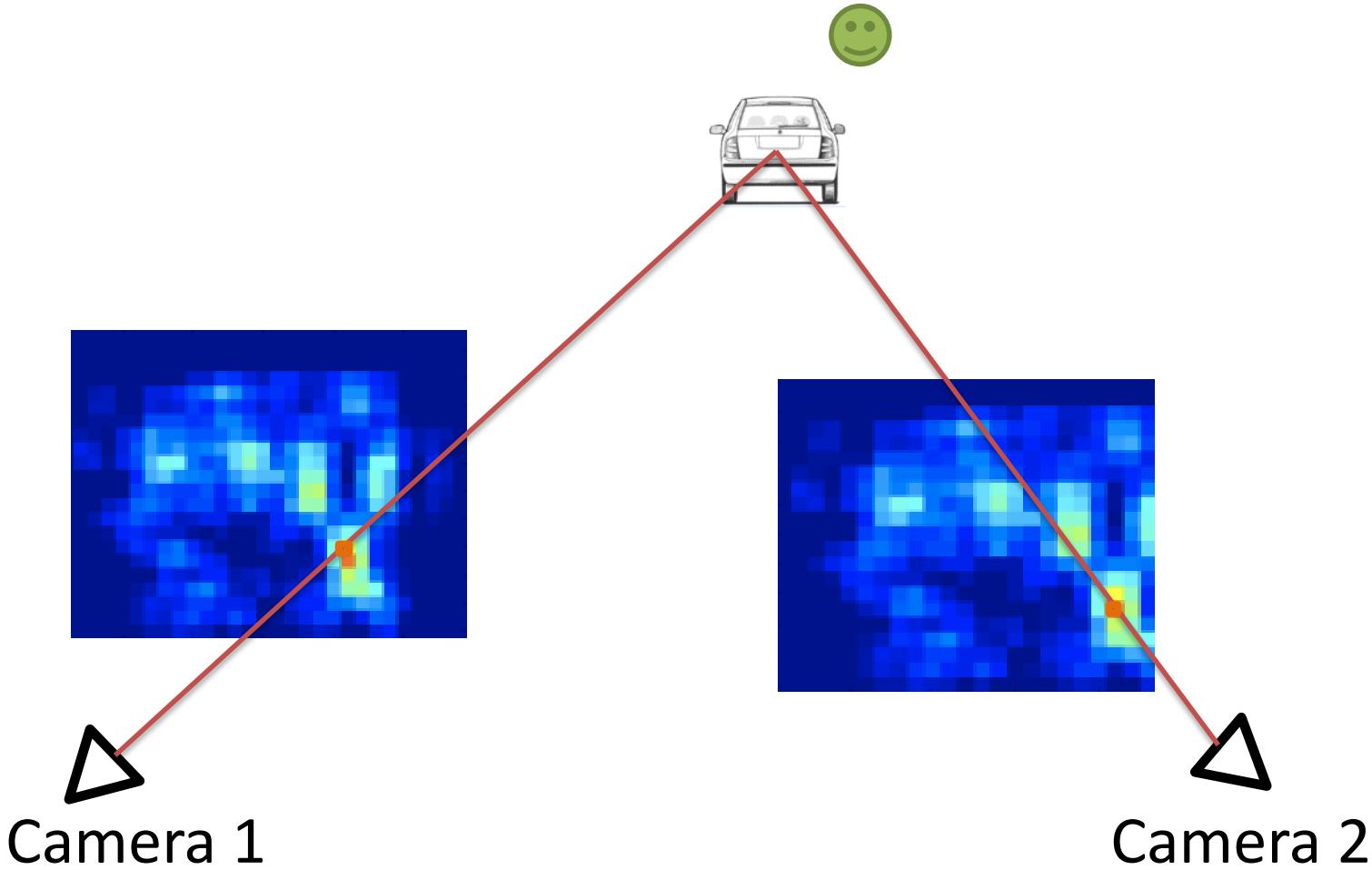
$$\Psi_t^{CO} \propto \prod_t^{N_t} (1 - \prod_k^{N_k} (1 - \Pr(o|O_t, C^k)))$$

SSFM: Object-level compatibility



- Agreement with measurements is computed using position, pose and scale

SSFM: Object-level compatibility

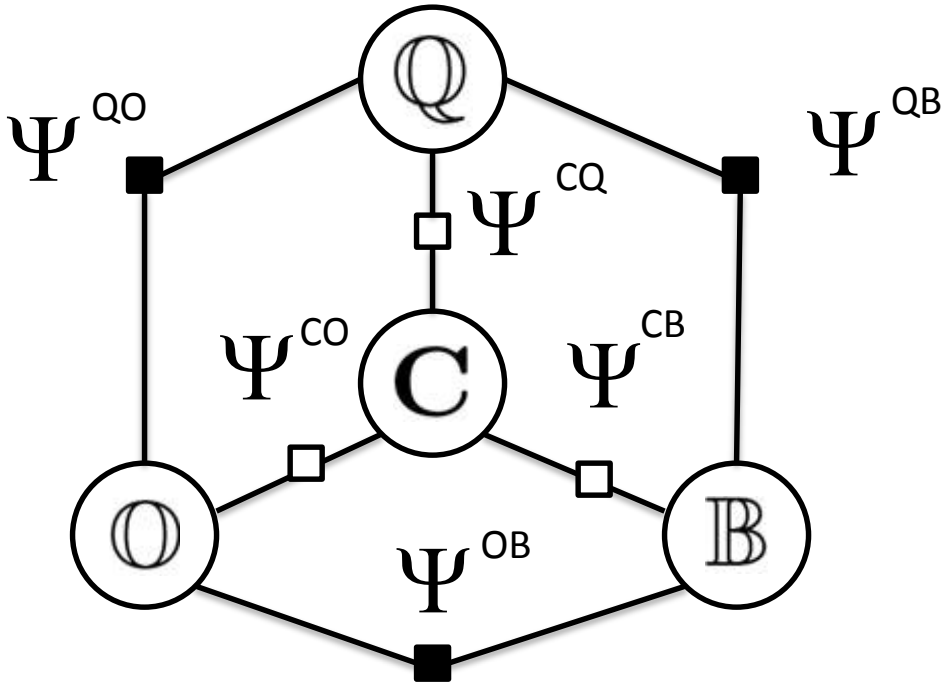


- Agreement with measurements is computed using position, pose and scale

SSFM with interactions

Bao, Bagra, Chao, Savarese
CVPR 2012

$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg\max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \prod_s \Psi_s^{CQ} \prod_t \Psi_t^{CO} \prod_r \Psi_r^{CB} \prod_{t,s} \Psi_{t,s}^{OQ} \prod_{t,r} \Psi_{t,r}^{OB} \prod_{r,s} \Psi_{r,s}^{BQ}$$



- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

- Interactions of points, regions and objects across views
- Interactions among object-regions-points

SSFM with interactions

$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg\max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \prod_s \Psi_s^{CQ} \prod_t \Psi_t^{CO} \prod_r \Psi_r^{CB} \prod_{t,s} \Psi_{t,s}^{OQ} \boxed{\prod_{t,r} \Psi_{t,r}^{OB}} \prod_{r,s} \Psi_{r,s}^{BQ}$$

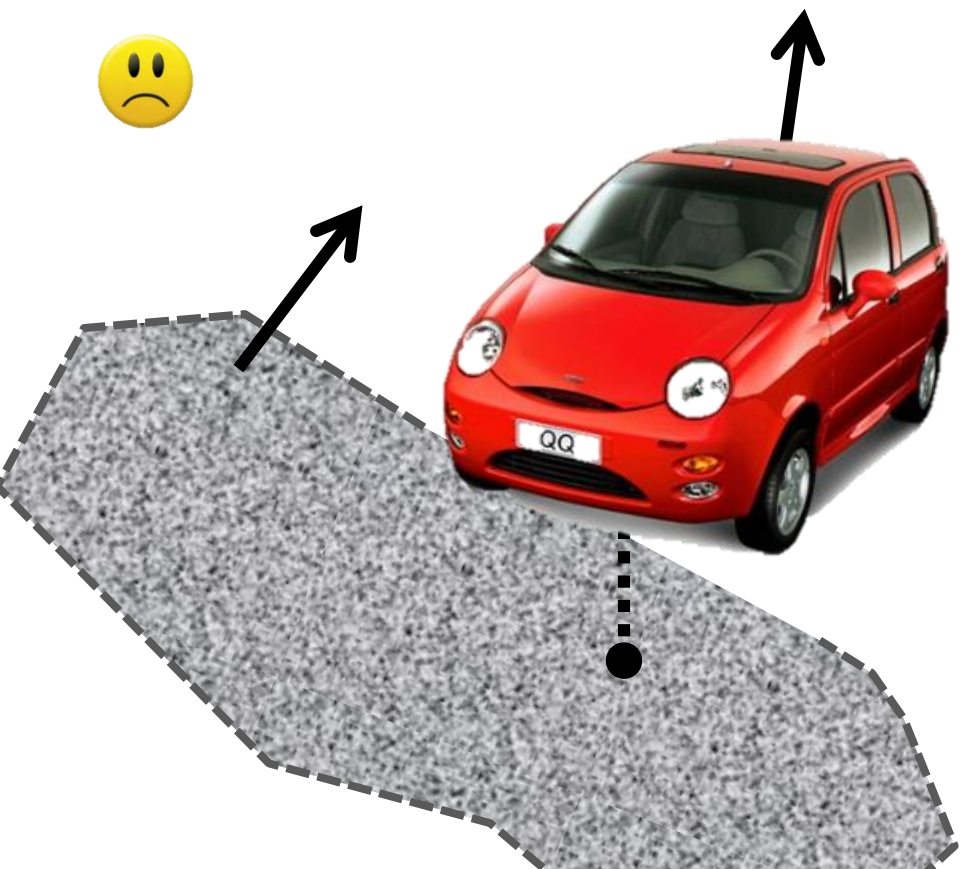
Object-Region Interactions:

- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

SSFM with interactions

$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg\max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \prod_s \Psi_s^{CQ} \prod_t \Psi_t^{CO} \prod_r \Psi_r^{CB} \prod_{t,s} \Psi_{t,s}^{OQ} \boxed{\prod_{t,r} \Psi_{t,r}^{OB}} \prod_{r,s} \Psi_{r,s}^{BQ}$$

Object-Region Interactions:



- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

SSFM with interactions

$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg\max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \prod_s \Psi_s^{CQ} \prod_t \Psi_t^{CO} \prod_r \Psi_r^{CB} \boxed{\prod_{t,s} \Psi_{t,s}^{OQ}} \prod_{t,r} \Psi_{t,r}^{OB} \prod_{r,s} \Psi_{r,s}^{BQ}$$

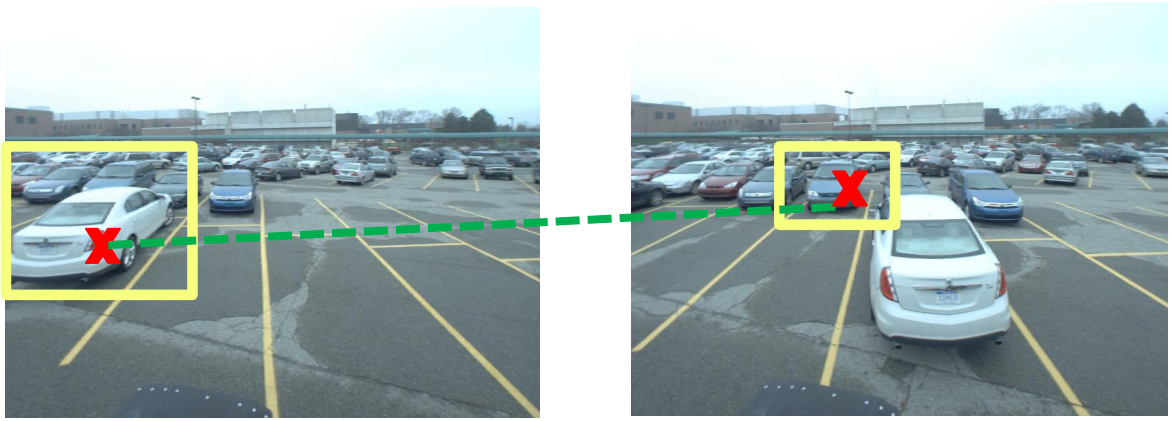
Object-point Interactions:

- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

SSFM with interactions

$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg\max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \prod_s \Psi_s^{CQ} \prod_t \Psi_t^{CO} \prod_r \Psi_r^{CB} \boxed{\prod_{t,s} \Psi_{t,s}^{OQ}} \prod_{t,r} \Psi_{t,r}^{OB} \prod_{r,s} \Psi_{r,s}^{BQ}$$

Object-point Interactions:



- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

Solving the SSFM problem

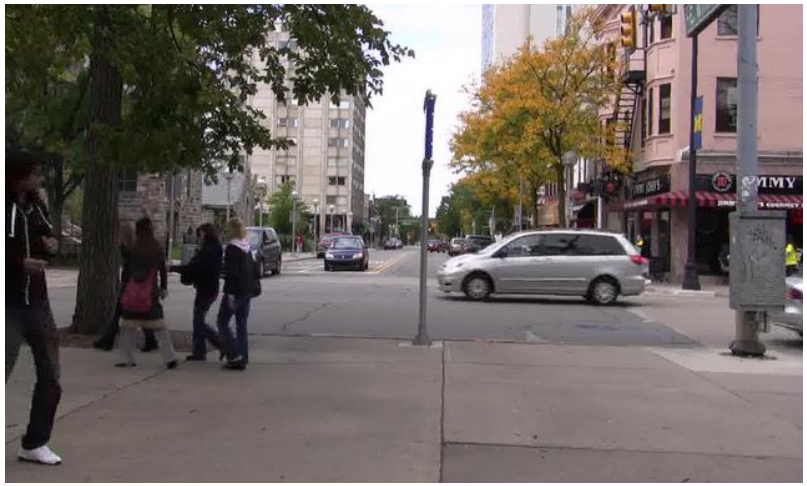
$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg \max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \Psi(\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}; \mathbf{I})$$

- Modified Reversible Jump Markov Chain Monte Carlo (RJ-MCMC) sampling algorithm [Dellaert et al., 2000]
- Initialization of the cameras, objects, and points are critical for the sampling
- Initialize configuration of cameras using:
  - SFM
  - consistency of object/region properties across views

Results

Input images

FORD CAMPUS dataset [Pandey et al., 09]

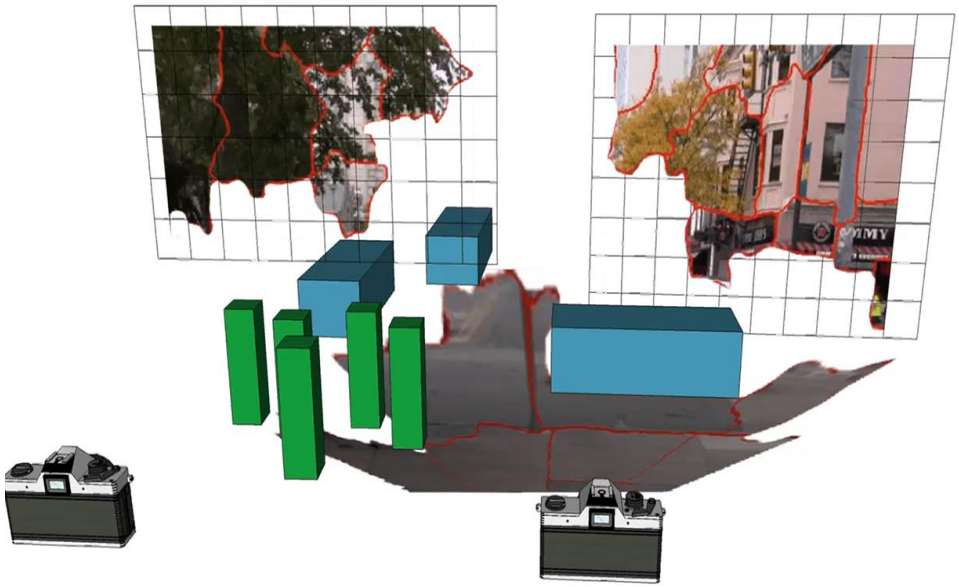


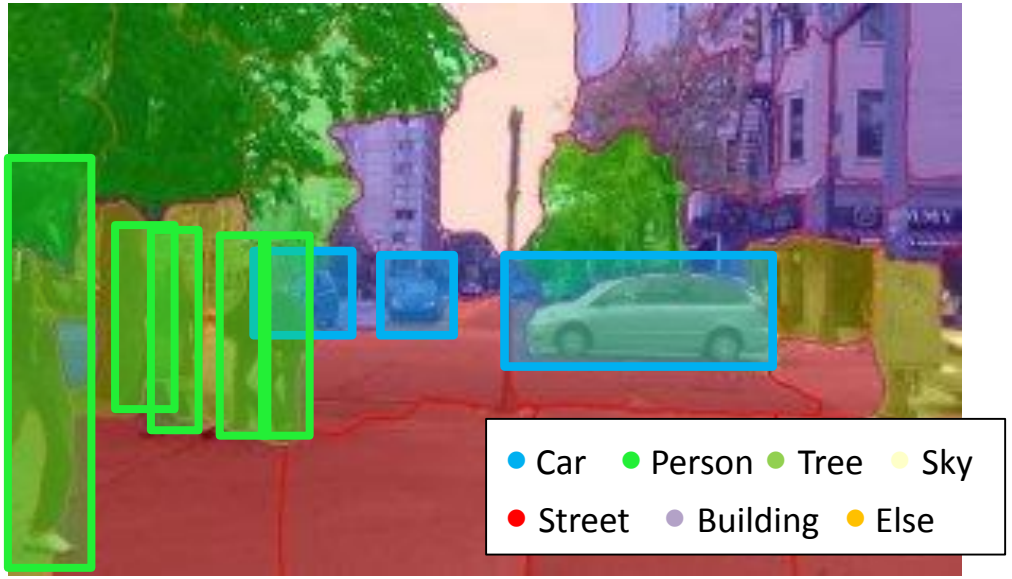
⋮

- Wide baseline
- Background clutter
- Limited visibility
- Un-calibrated cameras

Results

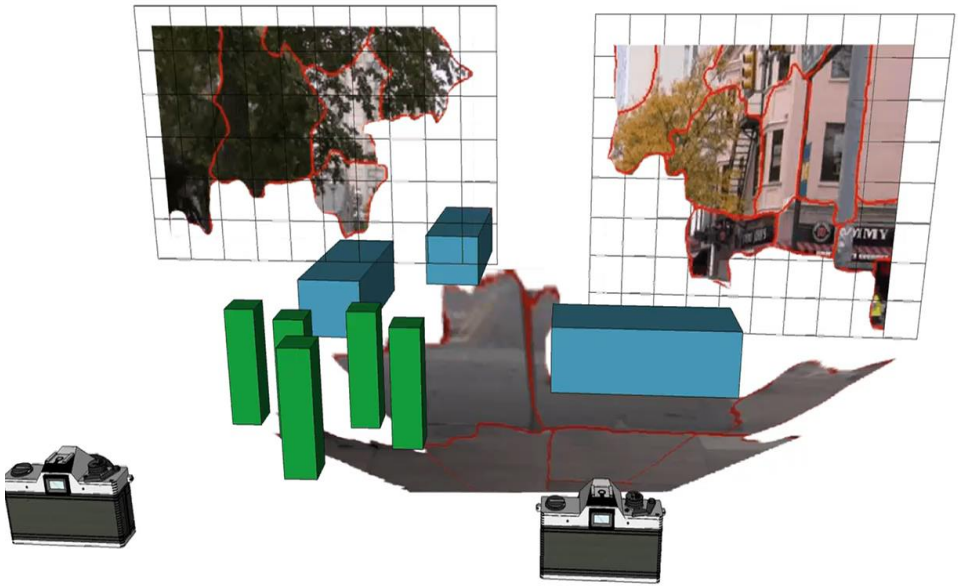
Input images





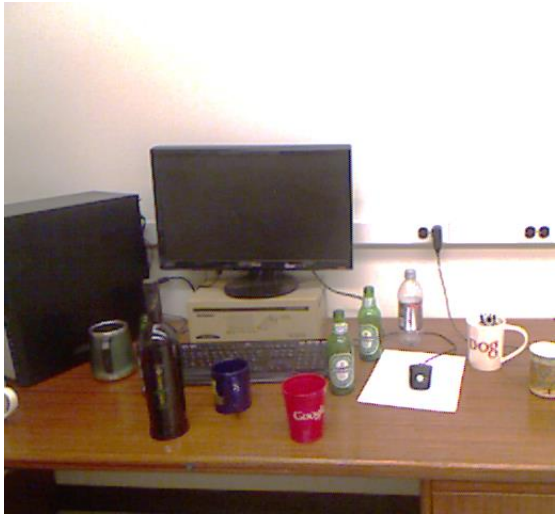
Results

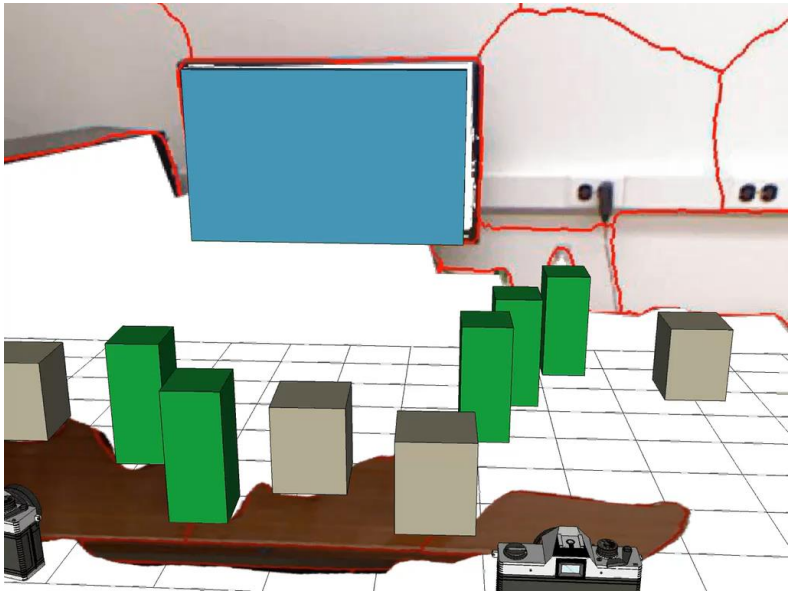
Input images

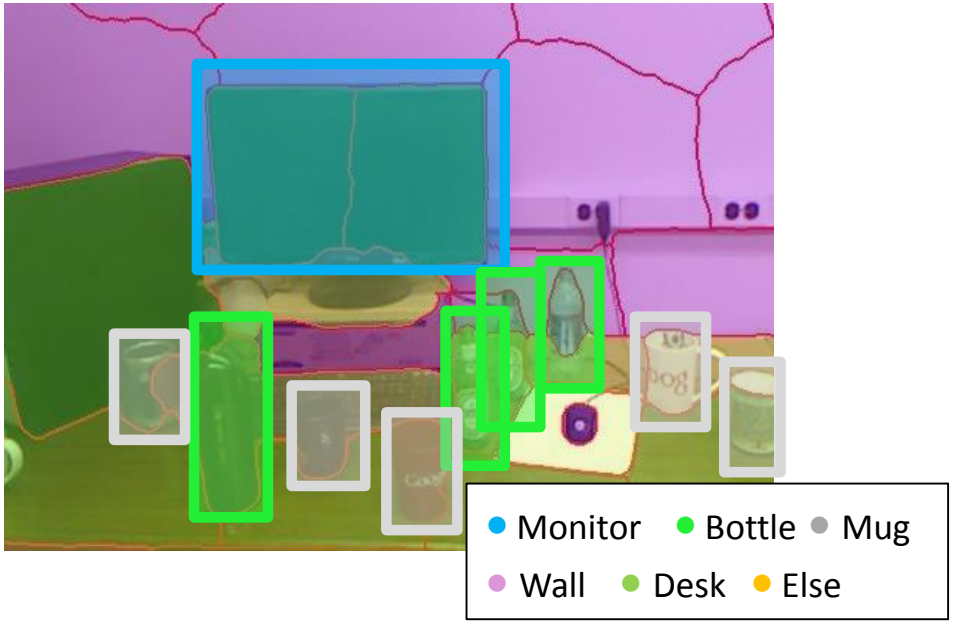


Results

Input images



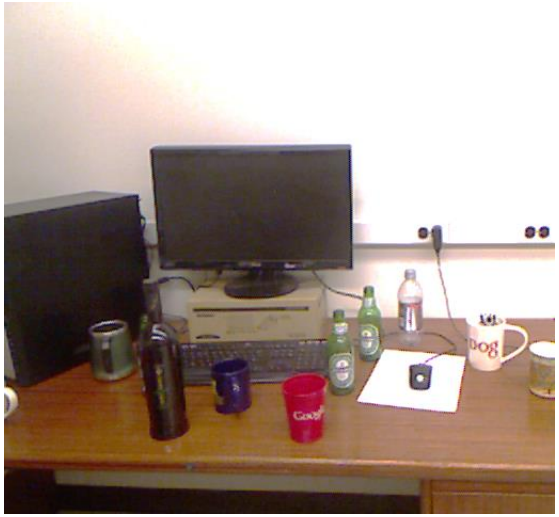


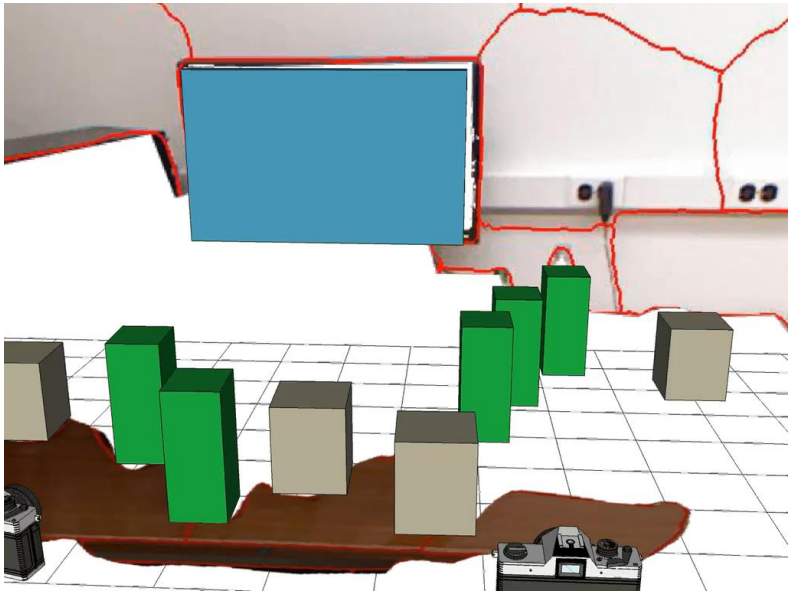


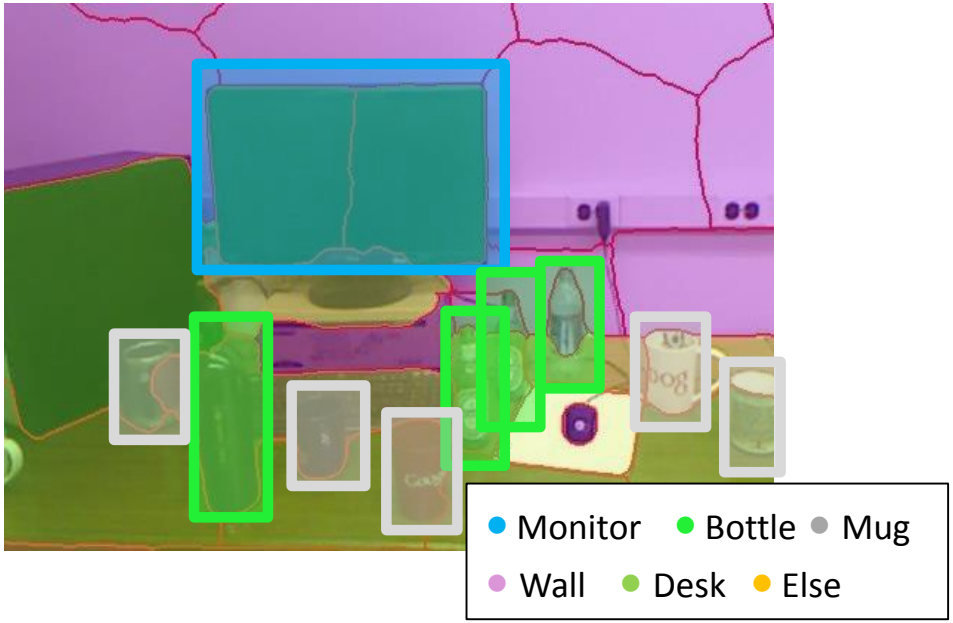
From the office dataset [Bao et al., 11]

Results

Input images





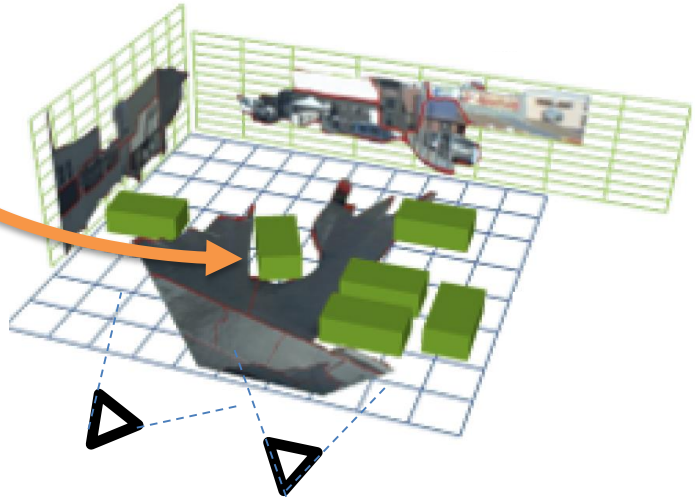


From the office dataset [Bao et al., 11]

Results

Average precision in localizing objects in the 3D space

	Hoiem et al. 2011	SSFM no int.	SSFM
FORD CAMPUS	21.4%	32.7%	**43.1%**
OFFICE	15.5%	20.2%	**21.6%**



Average precision in detecting objects in the 2D image

DPM [1]	SSFM 2 views no int.	SSFM 2 views	SSFM 4 views
54.5%	61.3%	62.8%	**66.5%**

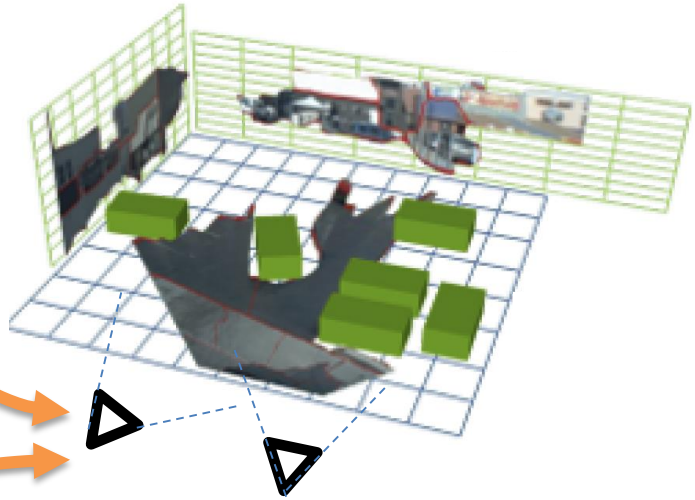
FORD CAMPUS dataset [Pandey et al., 09]
Office dataset [Bao et al., 11]

[1] Felzenszwalb et al. 2008

Results

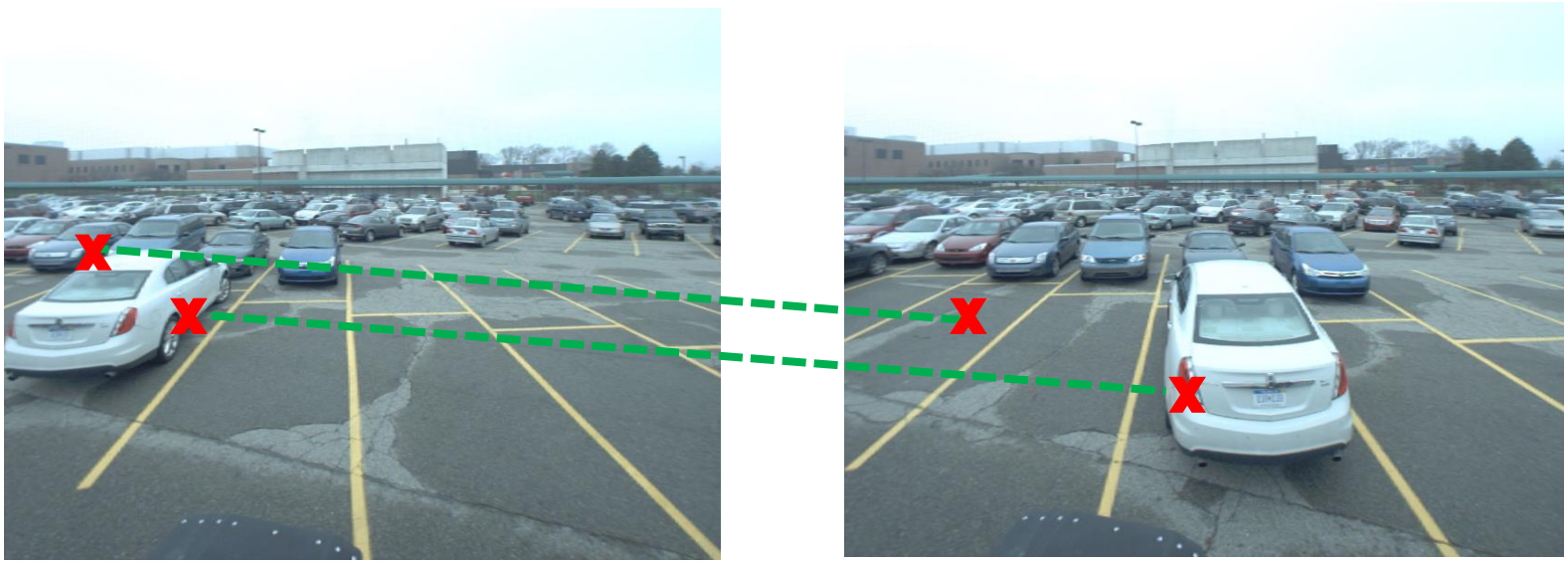
	Camera translation error		
	SFM Snavely et al., 08	**SSFM** no int.	**SSFM**
FORD CAMPUS	26.5°	19.9°	**12.1°**
OFFICE	8.5°	4.7°	**4.2°**
STREET	27.1°	17.6°	**11.4°**

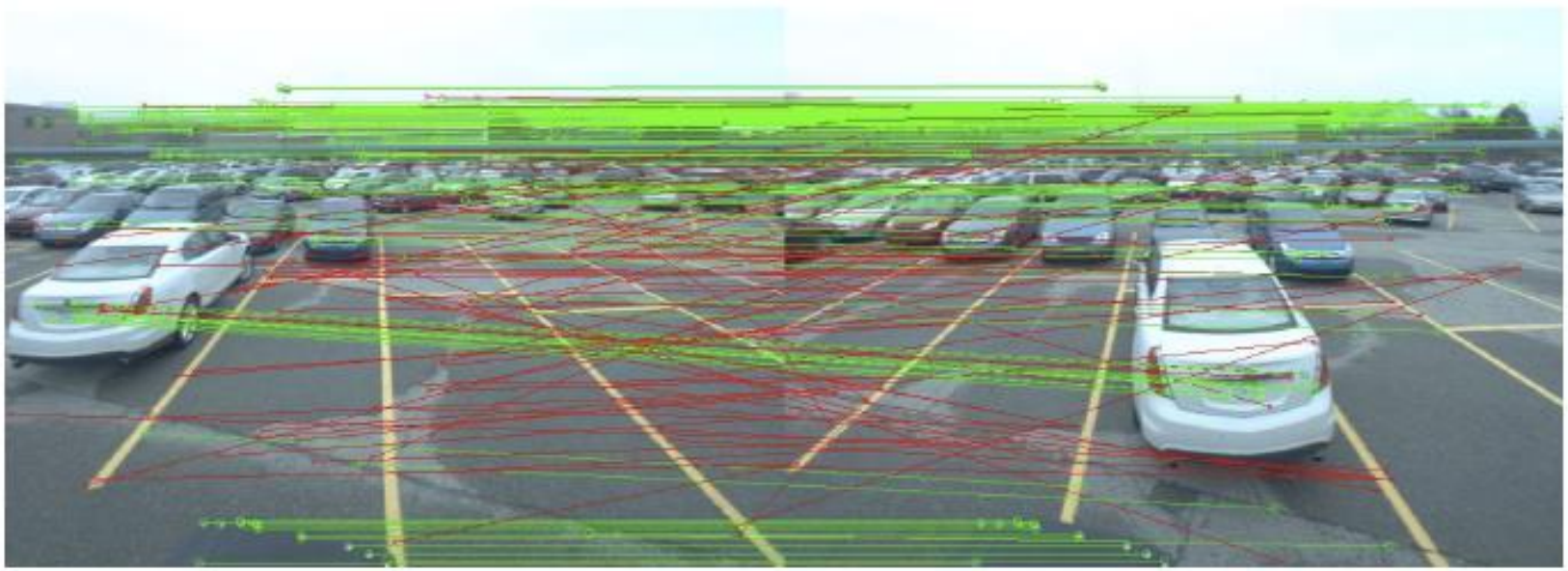
Camera rotation error		
SFM Snavely et al., 08	**SSFM** no int.	**SSFM**
<1°	<1°	**<1**
9.6°	4.2°	**3.5°**
21.1°	3.1°	**3.0°**



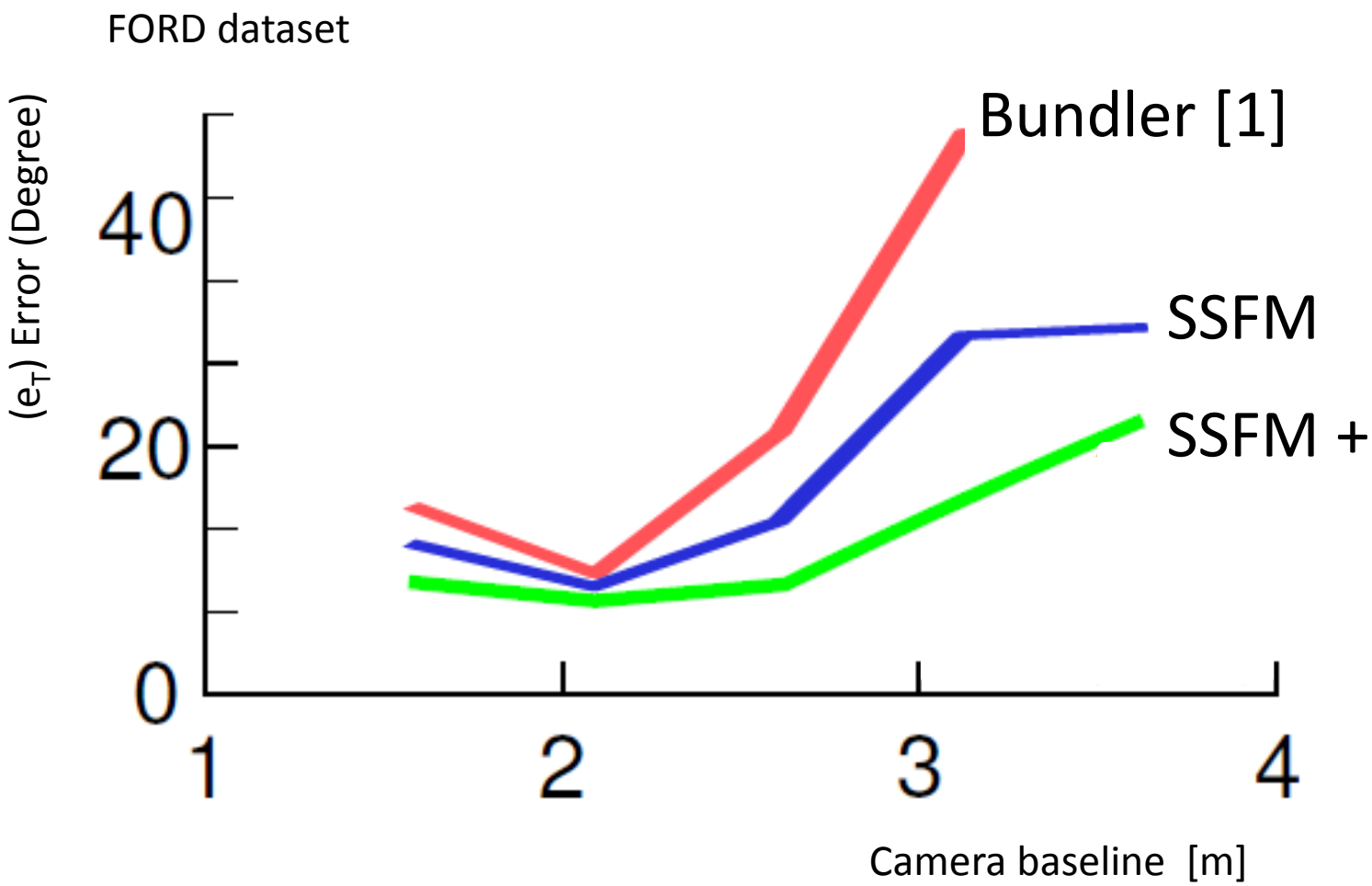
FORD CAMPUS dataset [Pandey et al., 09]
Office dataset [Bao et al., 11]
Street dataset [Bao et al., 11]

Wide-baseline feature correspondence



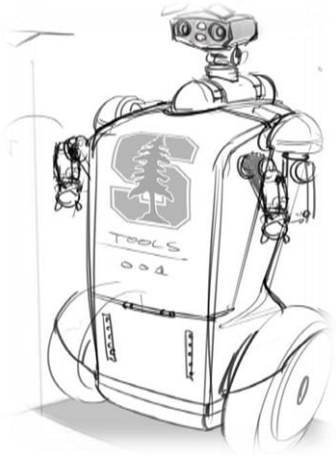


Camera Pose Estimation v.s. Base Line Width

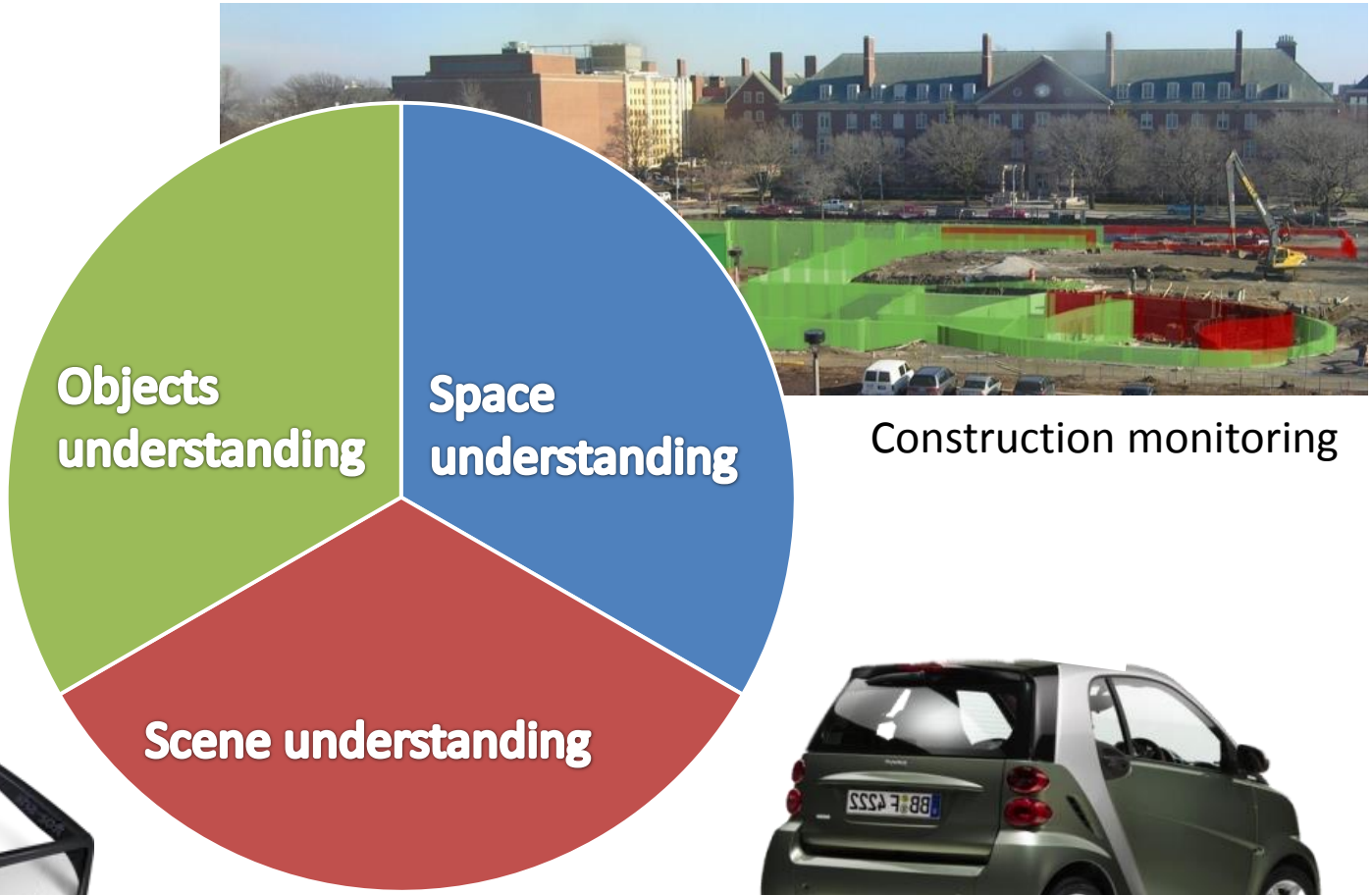


SSFM Source code available!
Please visit: http://www.eecs.umich.edu/vision/research.html

Applications



Robotic navigation



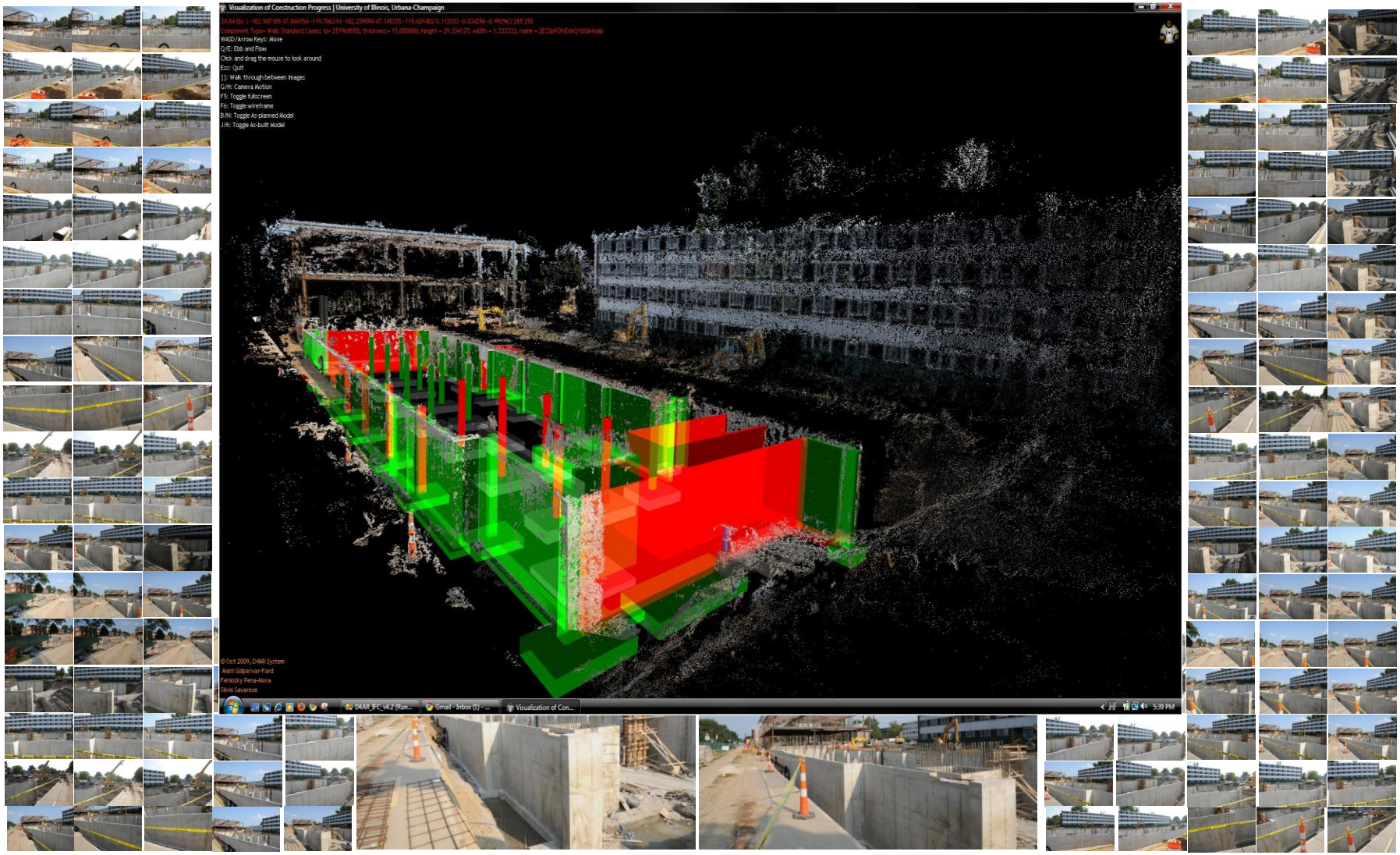
Construction monitoring

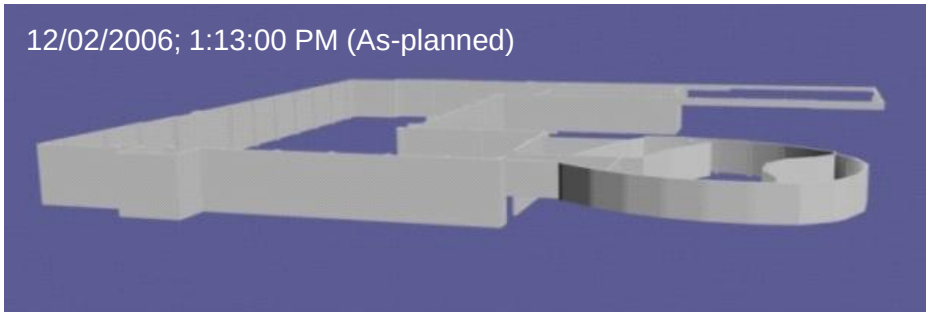
Mobile vision

Safe driving

Visual intelligence and large scale information management

Golparvar-Fard, Pena-Mora, Savarese , 2008-2012

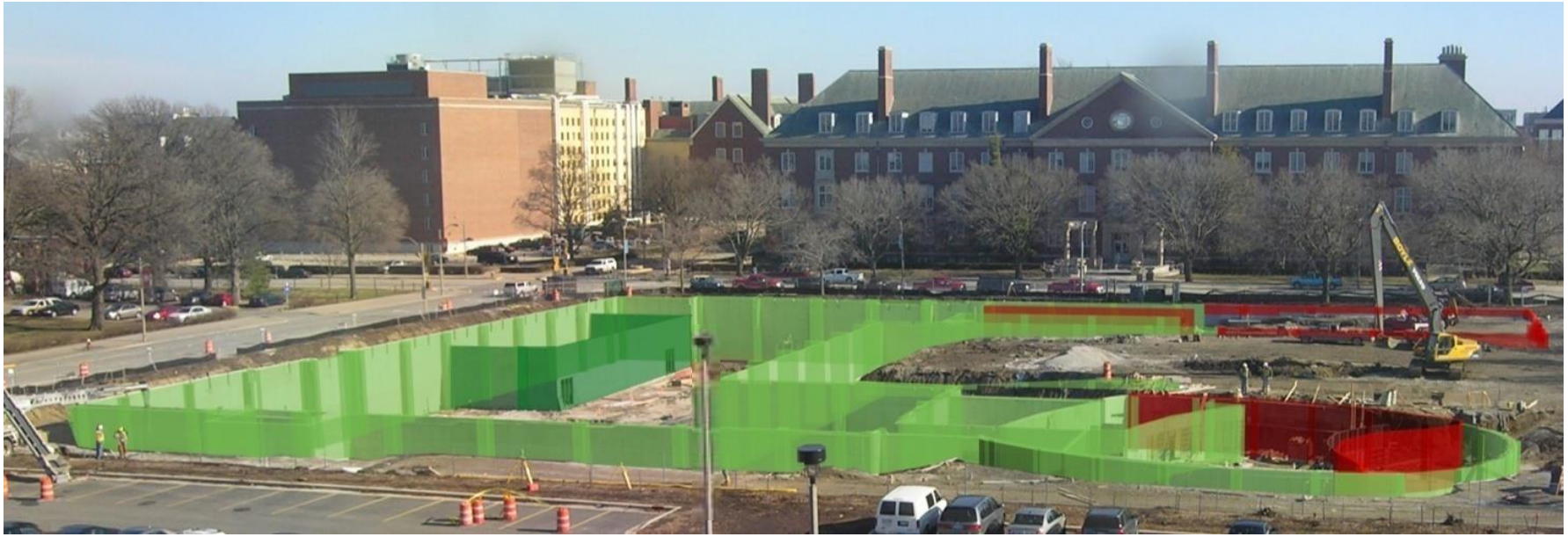

Visualization of Construction Progress | University of Illinois, Urbana-Champaign
WASD / Arrow Keys: Move
Q/E: Ebb and Flow
Click and drag the mouse to look around
Esc: Quit
[]: Walk through between images
G/H: Camera Motion
F5: Toggle fullscreen
F6: Toggle wireframe
B/N: Toggle As-planned Model
J/K: Toggle As-built Model
© Oct 2009, D4AR System
Avani Golparvar-Fard
Feniosky Peña-Mora
Silvio Savarese

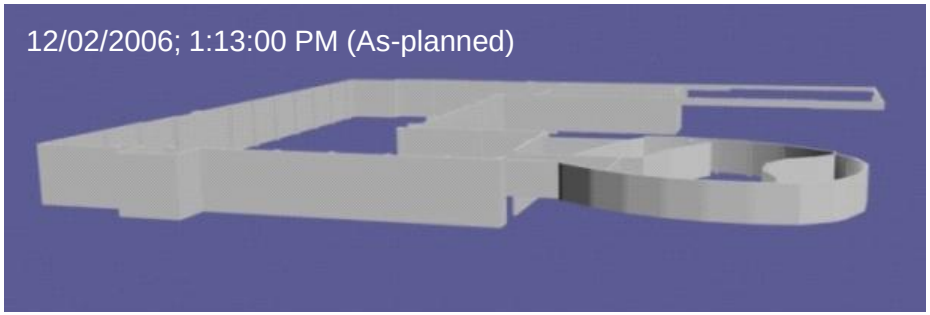


Ahead of Schedule

On Schedule

Behind Schedule



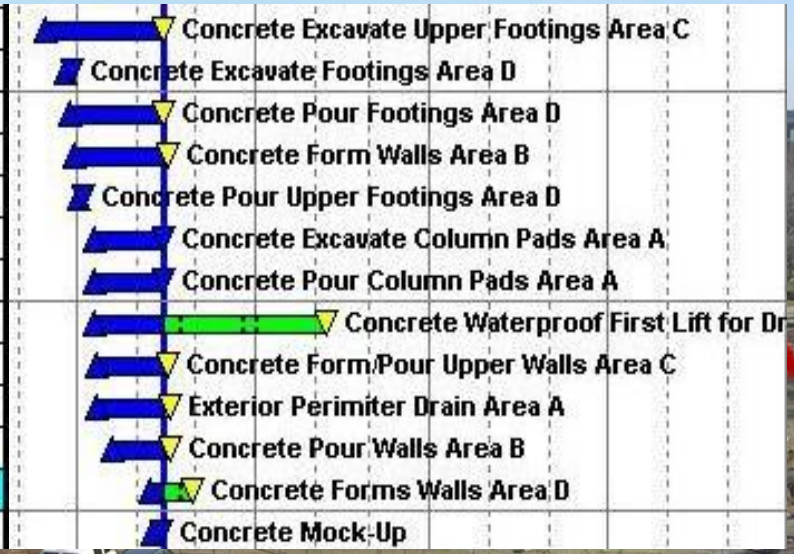


Ahead of Schedule

On Schedule

Behind Schedule

Concrete Excavate Upper Footings Area C	2	15SEP06A	14NOV06
Concrete Excavate Footings Area D	2	25SEP06A	27SEP06A
Concrete Pour Footings Area D	6	27SEP06A	14NOV06
Concrete Form Walls Area B	19	30SEP06A	16NOV06
Concrete Pour Upper Footings Area D	6	02OCT06A	03OCT06A
Concrete Excavate Column Pads Area A	1	10OCT06A	14NOV06A
Concrete Pour Column Pads Area A	3	10OCT06A	14NOV06A
Concrete Waterproof First Lift for Drain Tile	95	10OCT06A	06FEB07
Concrete Form/Pour Upper Walls Area C	6	11OCT06A	16NOV06
Exterior Perimiter Drain Area A	25	11OCT06A	17NOV06
Concrete Pour Walls Area B	20	19OCT06A	17NOV06
Concrete Forms Walls Area D	11	08NOV06A	29NOV06
Concrete Mock-Up	3	09NOV06A	08NOV06A



Hope you enjoyed this course

Good luck on your presentations on next week!