Lecture 16

3D scene understanding

Announcements

- No class on Wed (ECCV deadline!)

- In-class presentations next week, two parallel two-tracks sessions
  - 12:30pm - 2:30pm, March 19
  - Room 1: Oshman 125
  - Room 2: 450 Serra Mall, 300-300
- 3.5 minutes for each presentation including Q&A
- It's a team presentation
- In-class or Piazza questions count toward attendance evaluation

Announcements

- See Piazza and website for more information on the format for presentation and write up
- CA session on Friday to discuss expectations for final report
- Thanks for the online evaluations!
- Your feedback is extremely important!

What does it mean to understand a scene?

Computers can reconstruct 3D spaces better than humans!

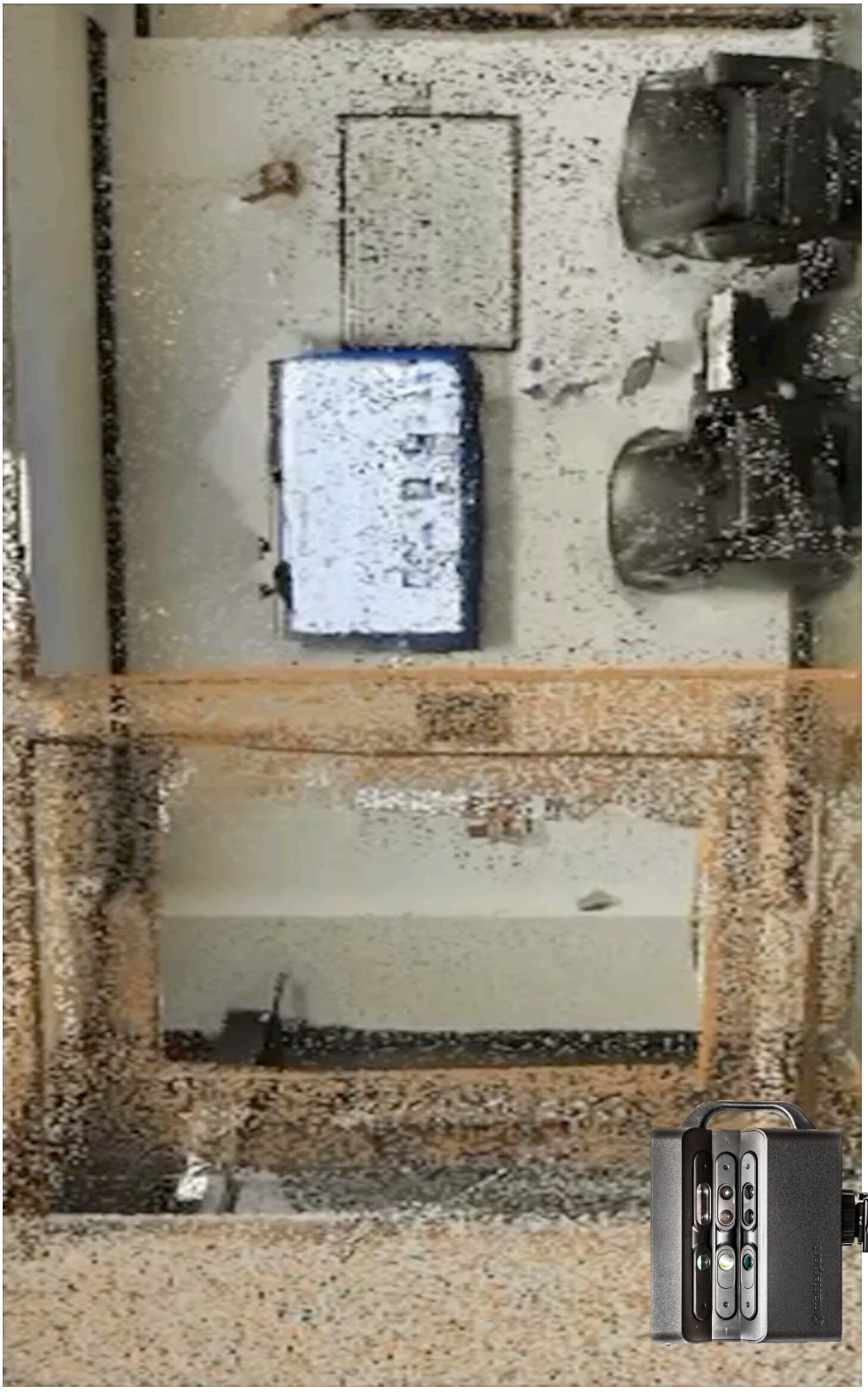
Snavely et al., 06-08

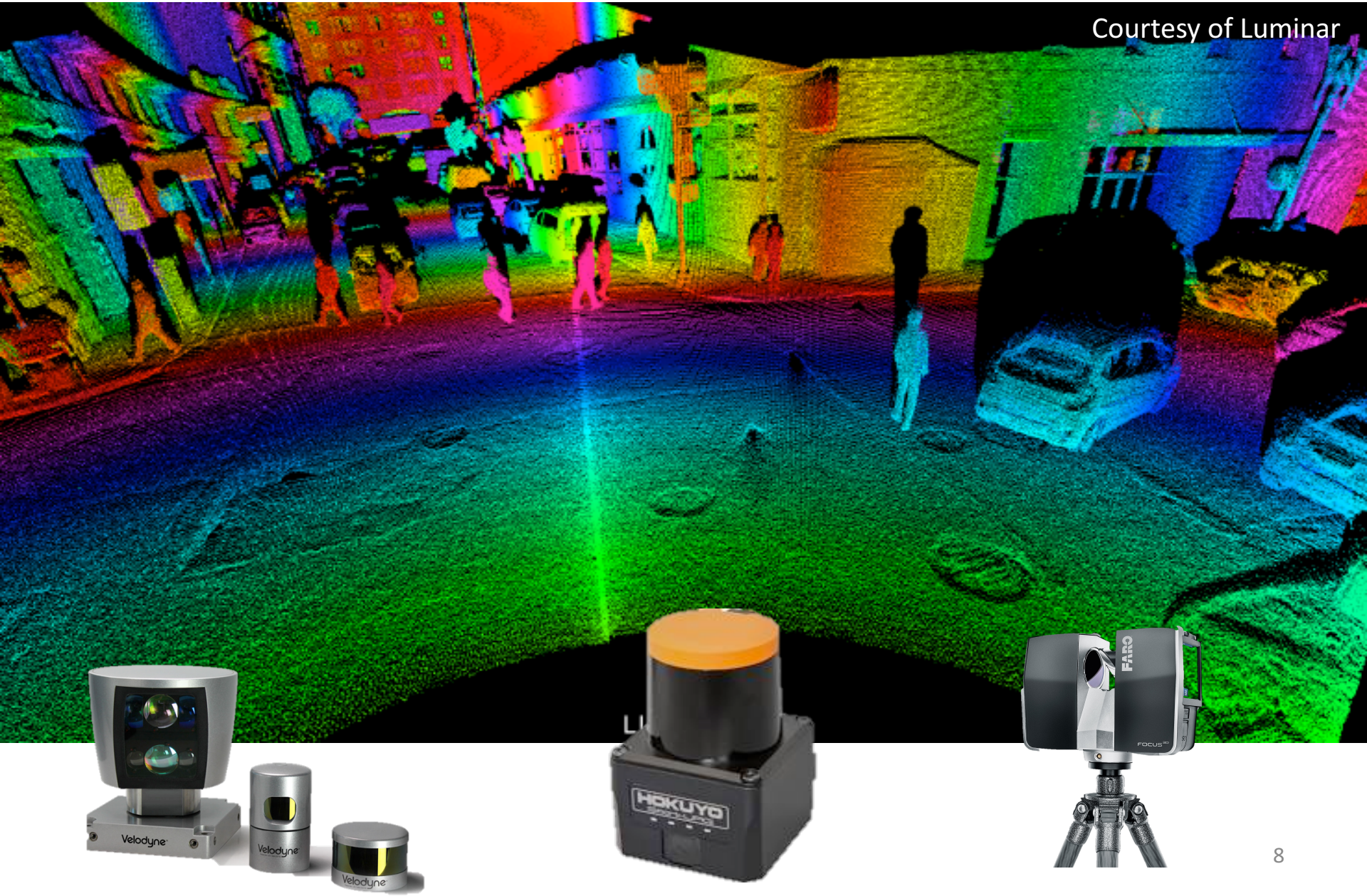
Microsoft Photosynth

Computers can reconstruct 3D spaces better than humans!

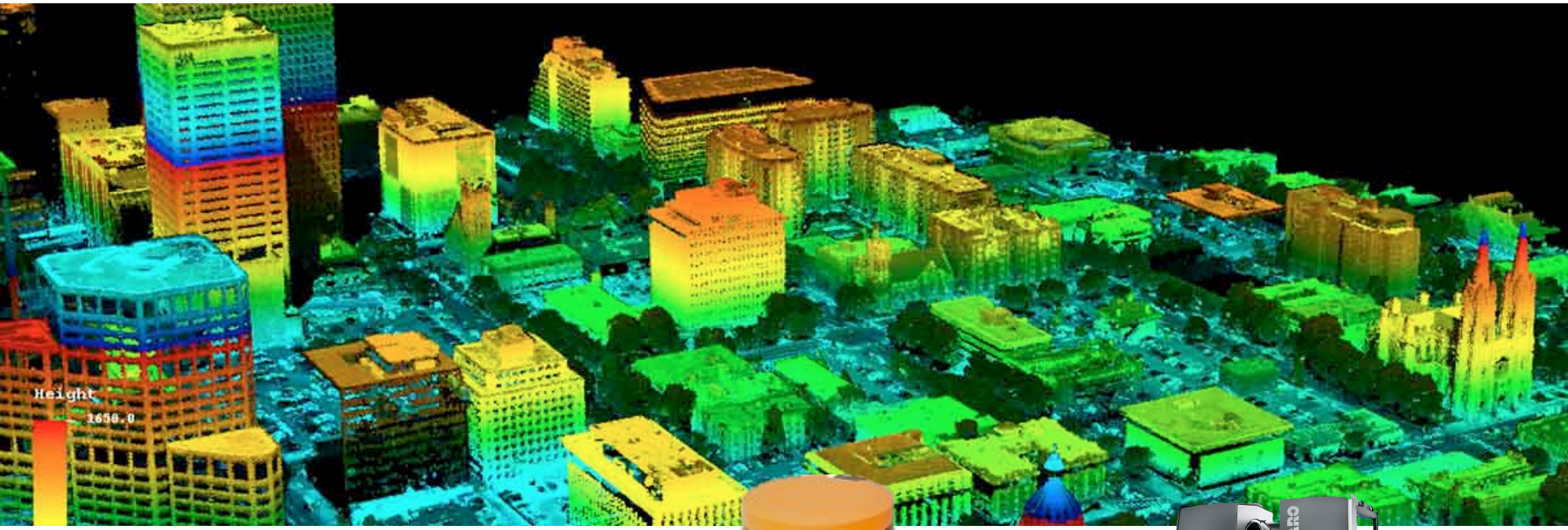
Snavely et al., 06-08

Armeni et al. 2016




Courtesy of Luminar
Velodyne
Velodyne
Velodyne
HOKUYO
FARO
FOCUS





Height
1650.0

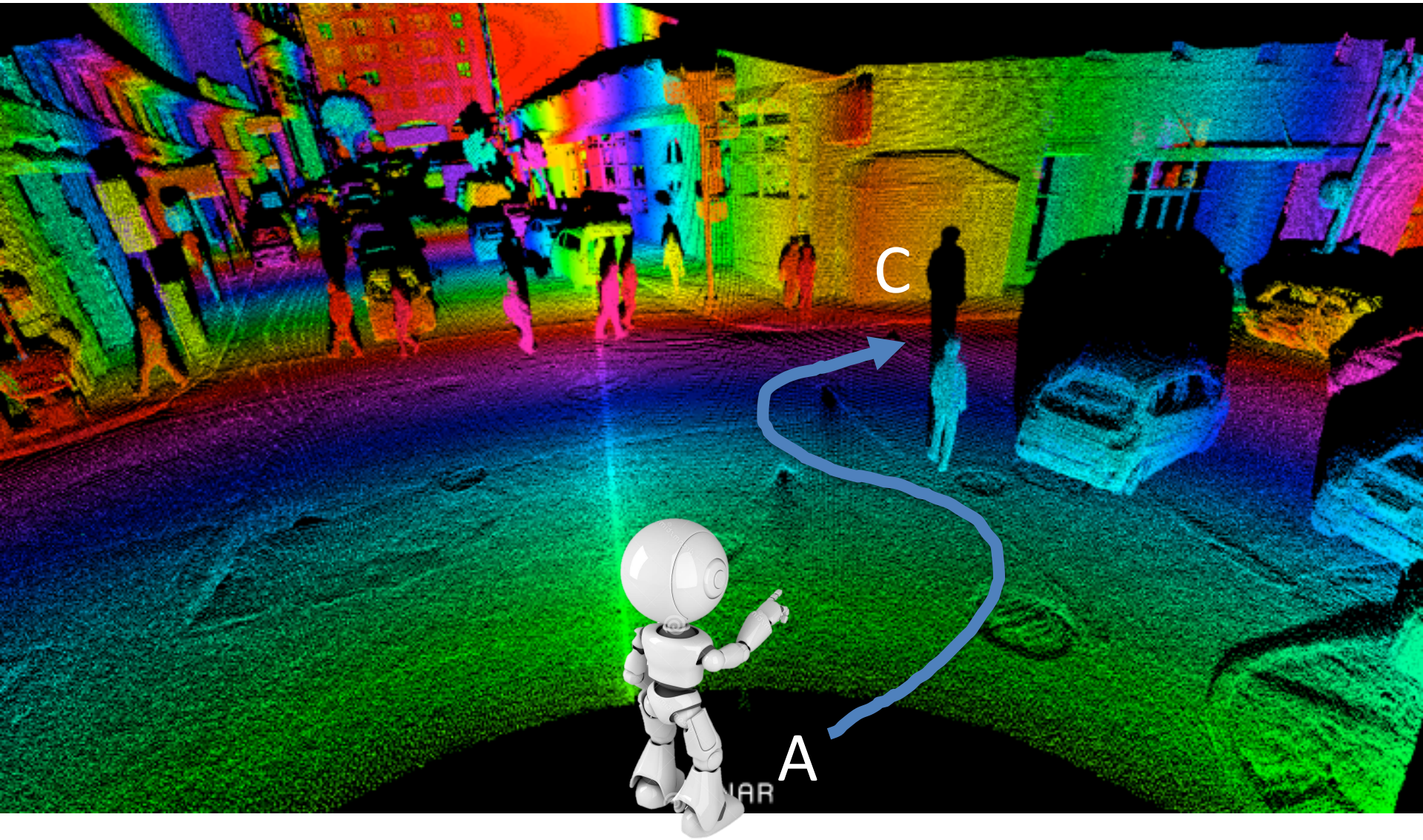
Velodyne
Velodyne
Velodyne

HOKUYO

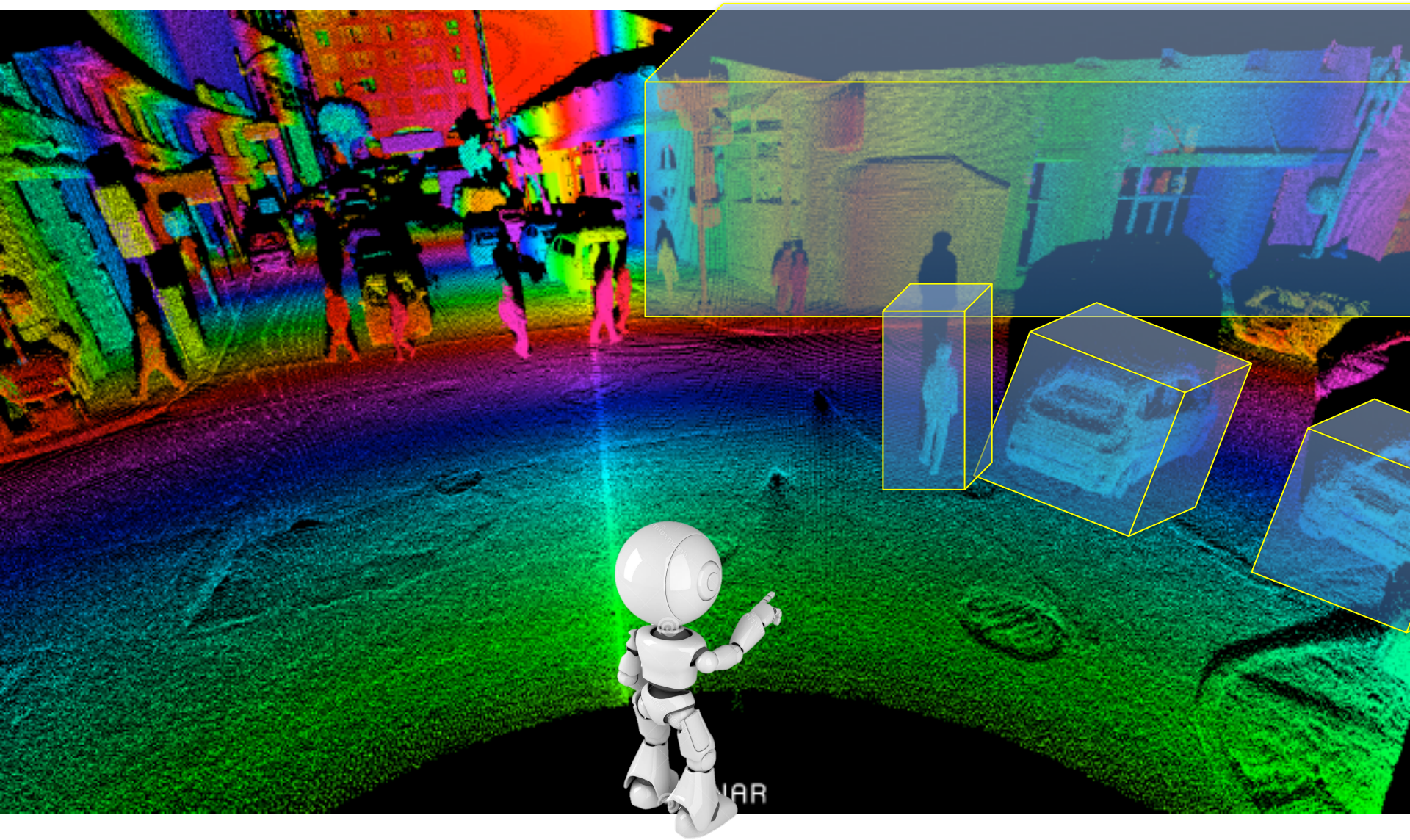
FARO
FOCUS

It this useful?

It is about “where” things are…



Is this sufficient?



Machine vision is…

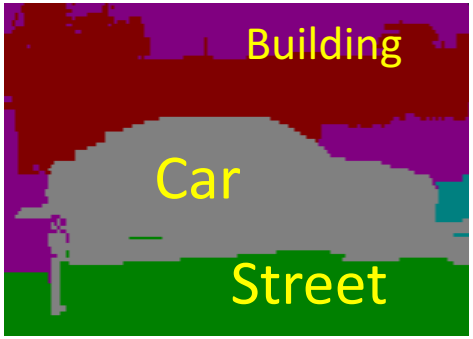
…not just about “where”,

also about “what”

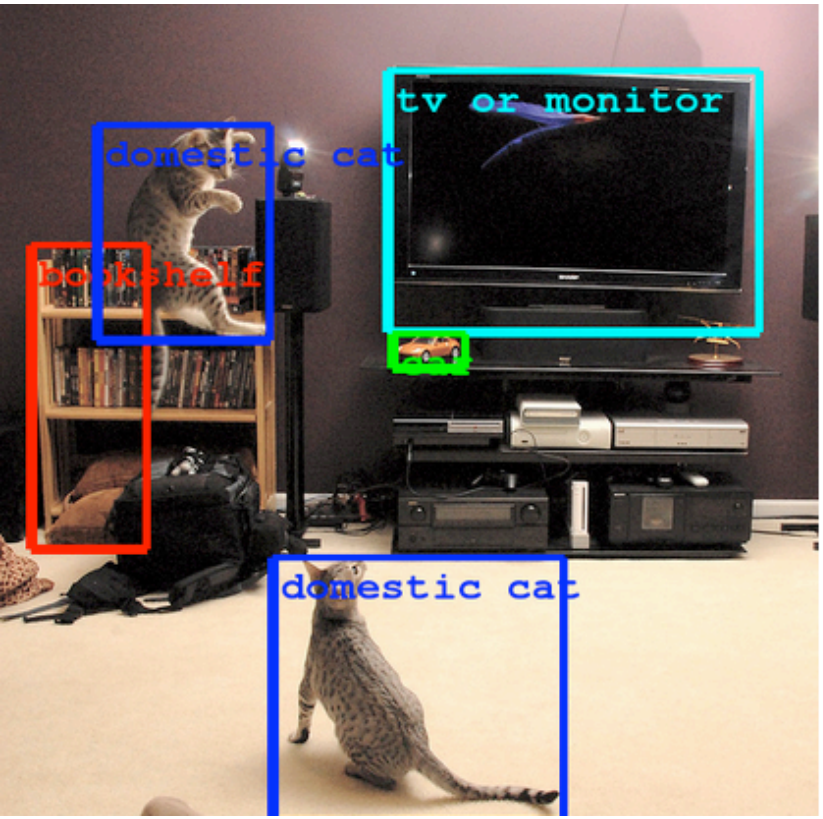
Image-to-labels paradigm

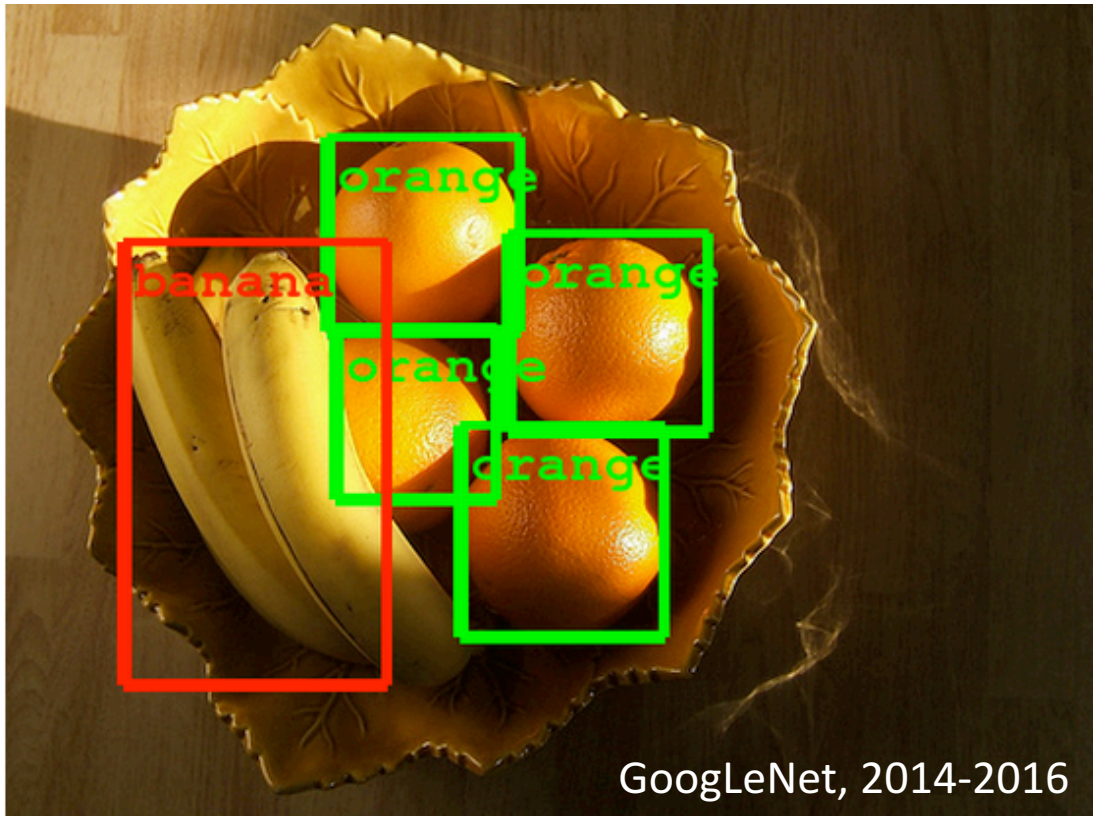
image

labels

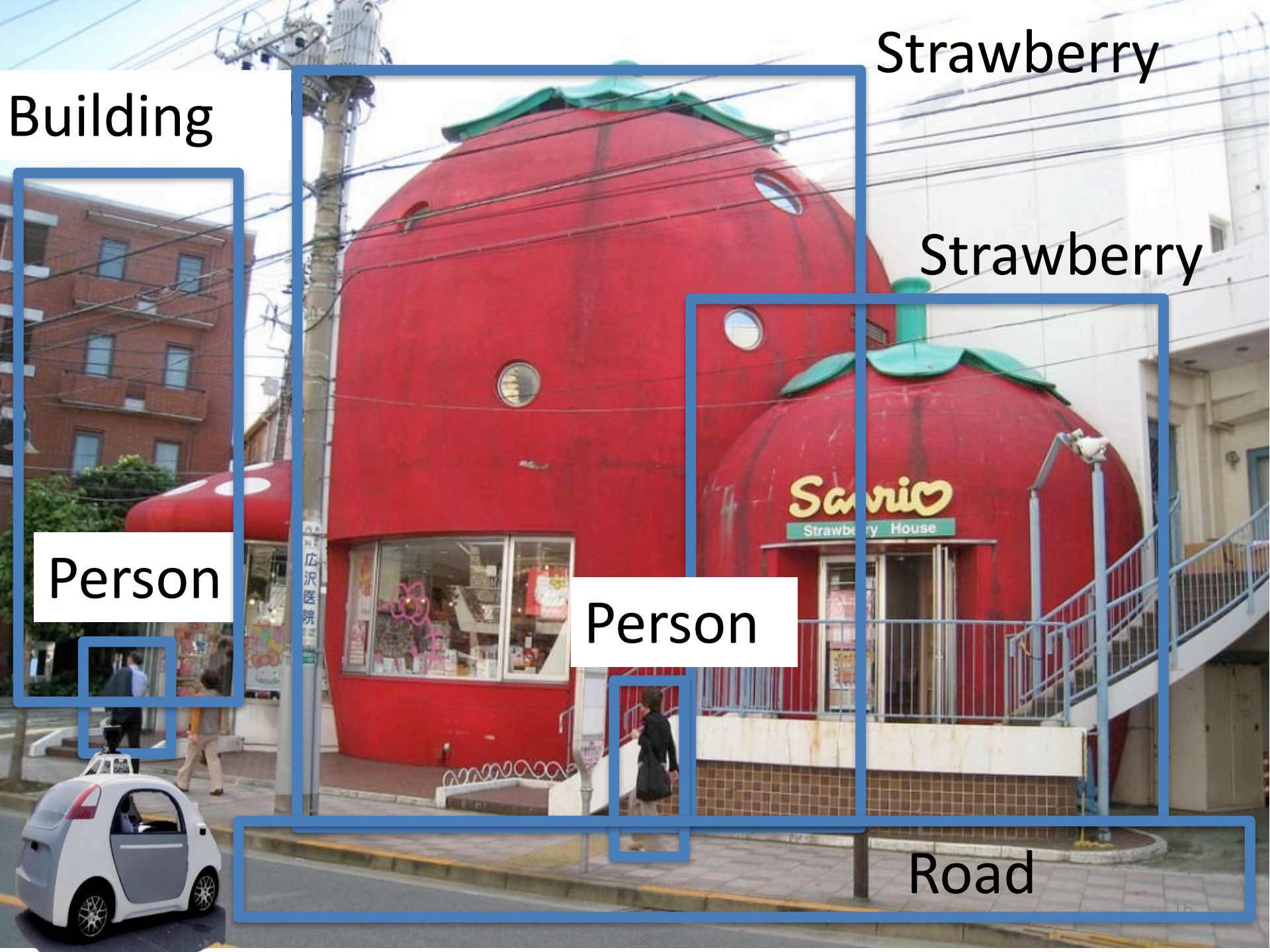


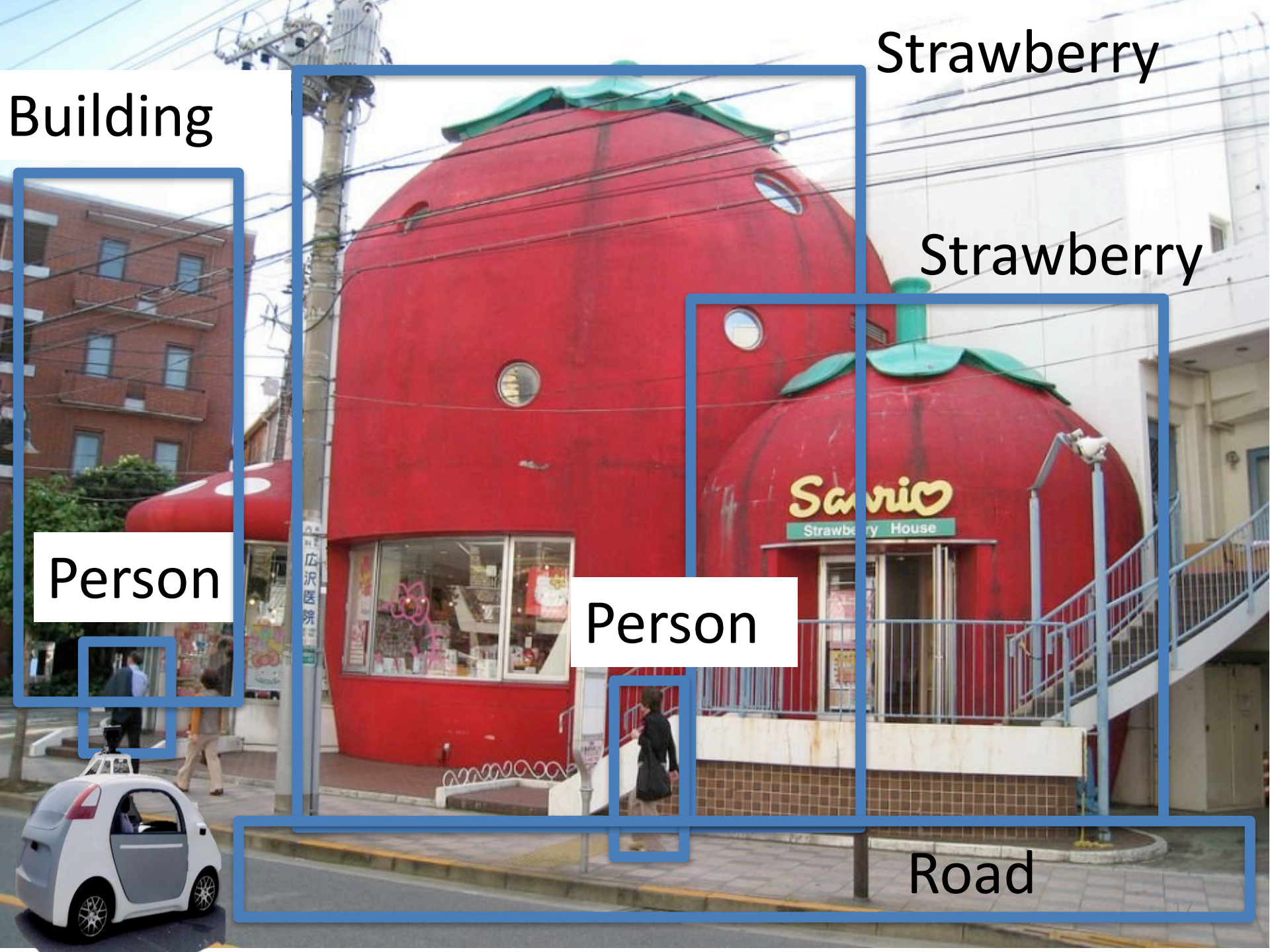
Computers can recognize objects better than humans! (*)

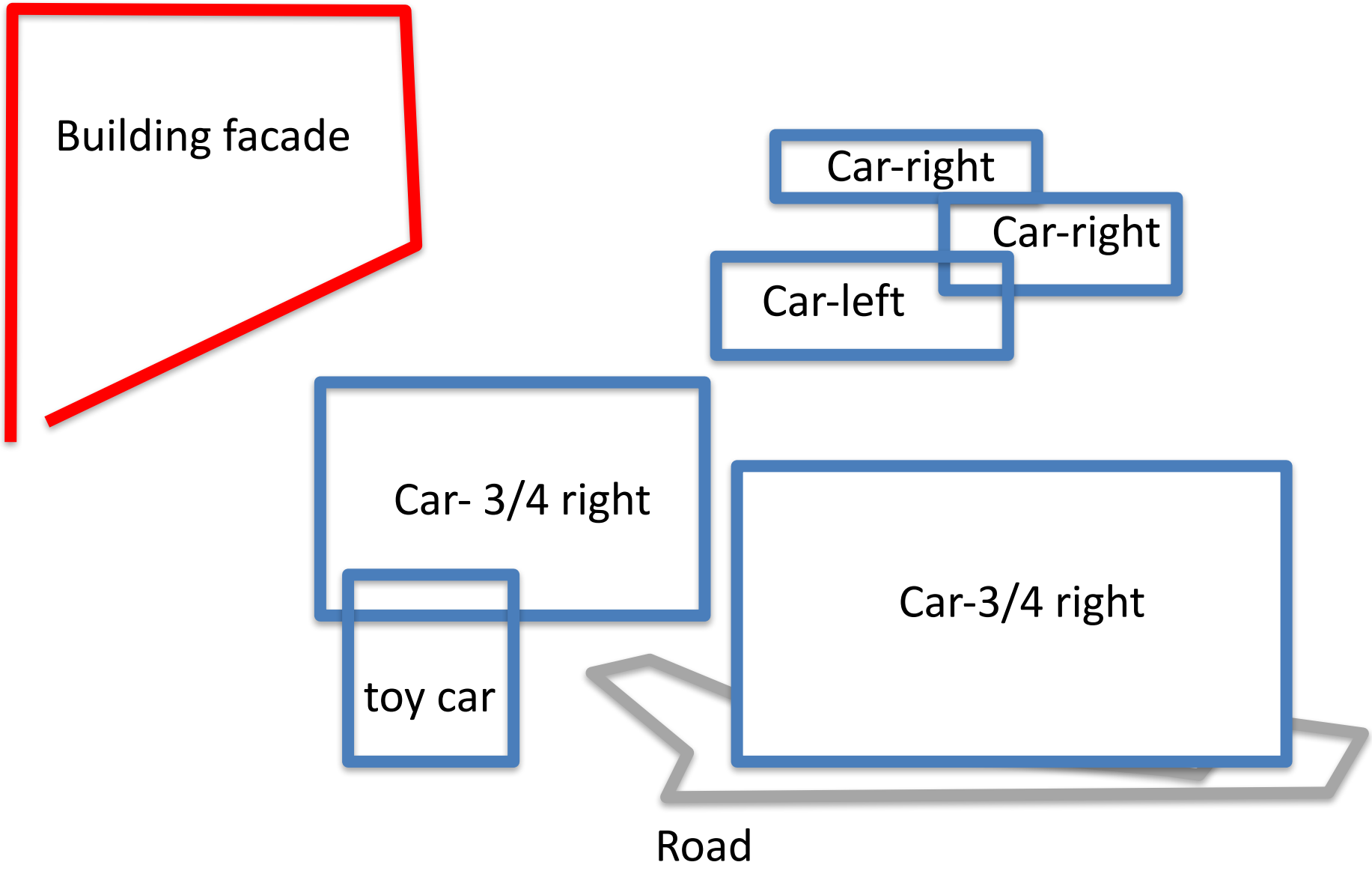




(*) Ensamble Res Nets trained on image Imagenet, 2016, for certain categories.
Russakovsky, et al. (2016)


Strawberry
Building
Strawberry
Person
Person
Road

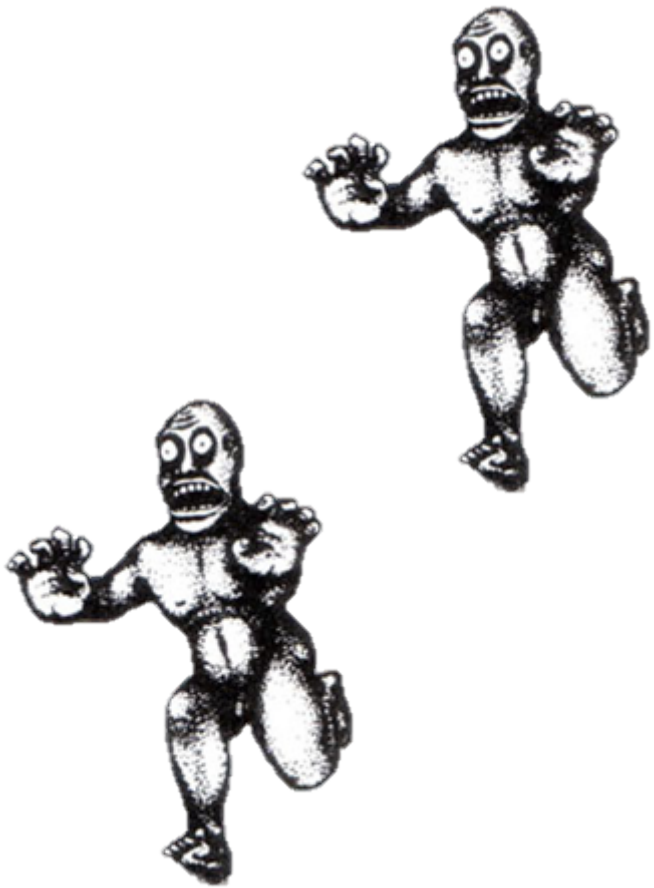

Strawberry
Building
Strawberry
Person
Person
Road

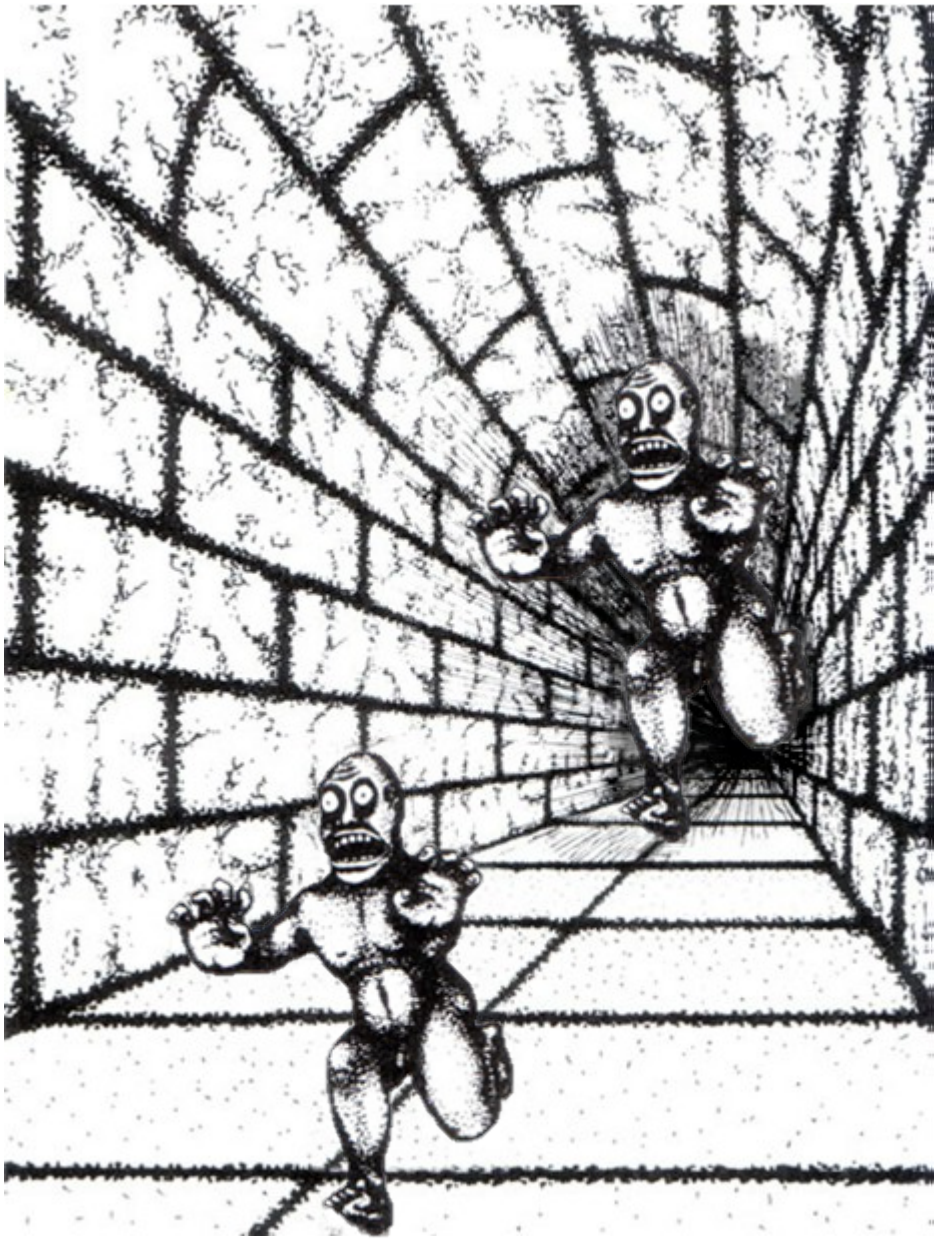

Building facade
Car-right
Car-right
Car-left
Car- 3/4 right
toy car
Car-3/4 right
Road

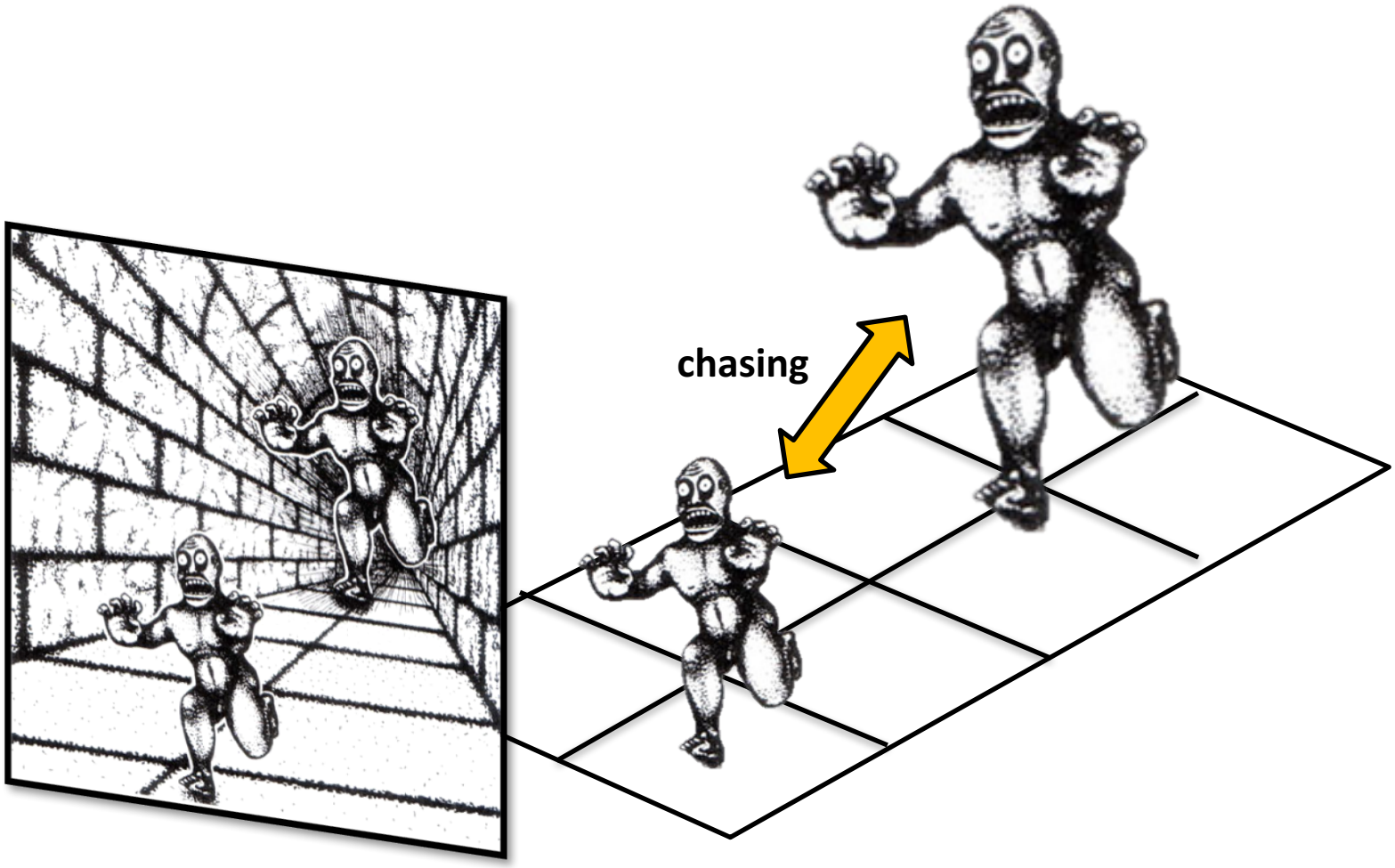
Building facade

Road


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





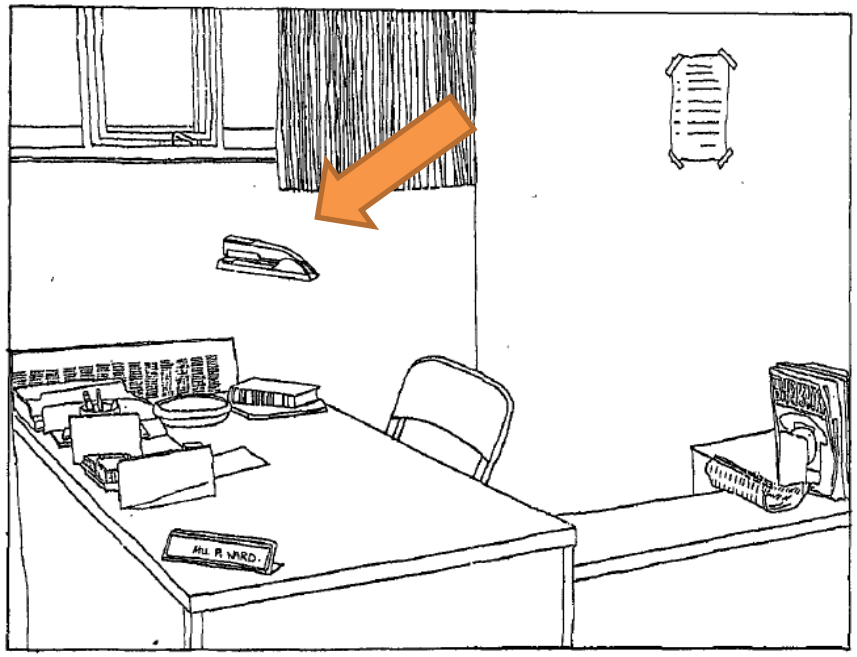

chasing

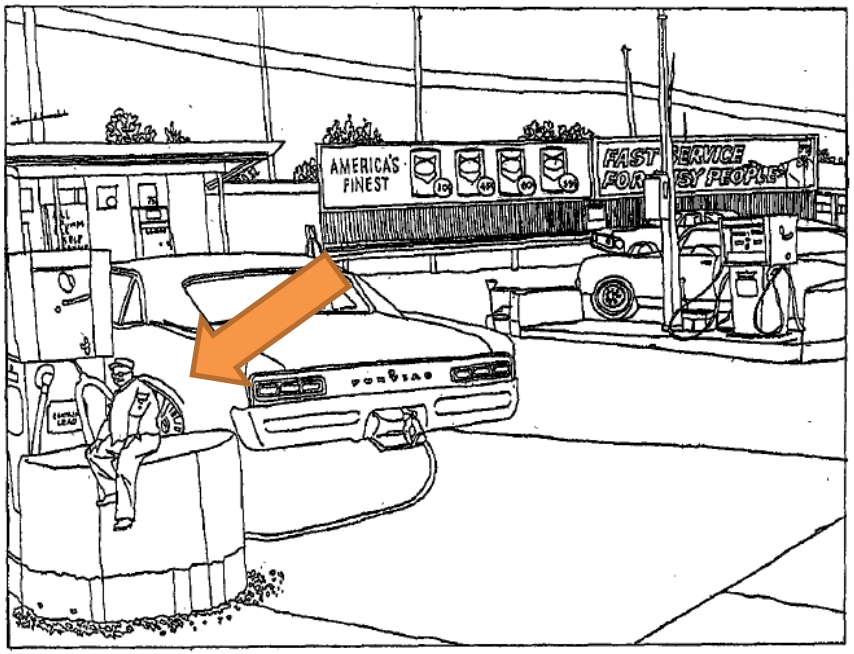
Objects are constrained by the 3D space

The 3D space is shaped by its objects

Modeling this interplay is critical for 3D perception!

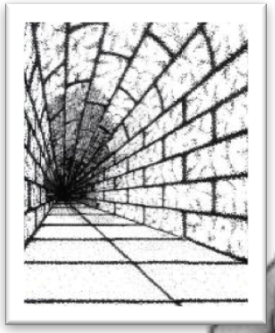
Humans perceive the world in 3D





Biederman, Mezzanotte and Rabinowitz, 1982

Visual processing in the brain

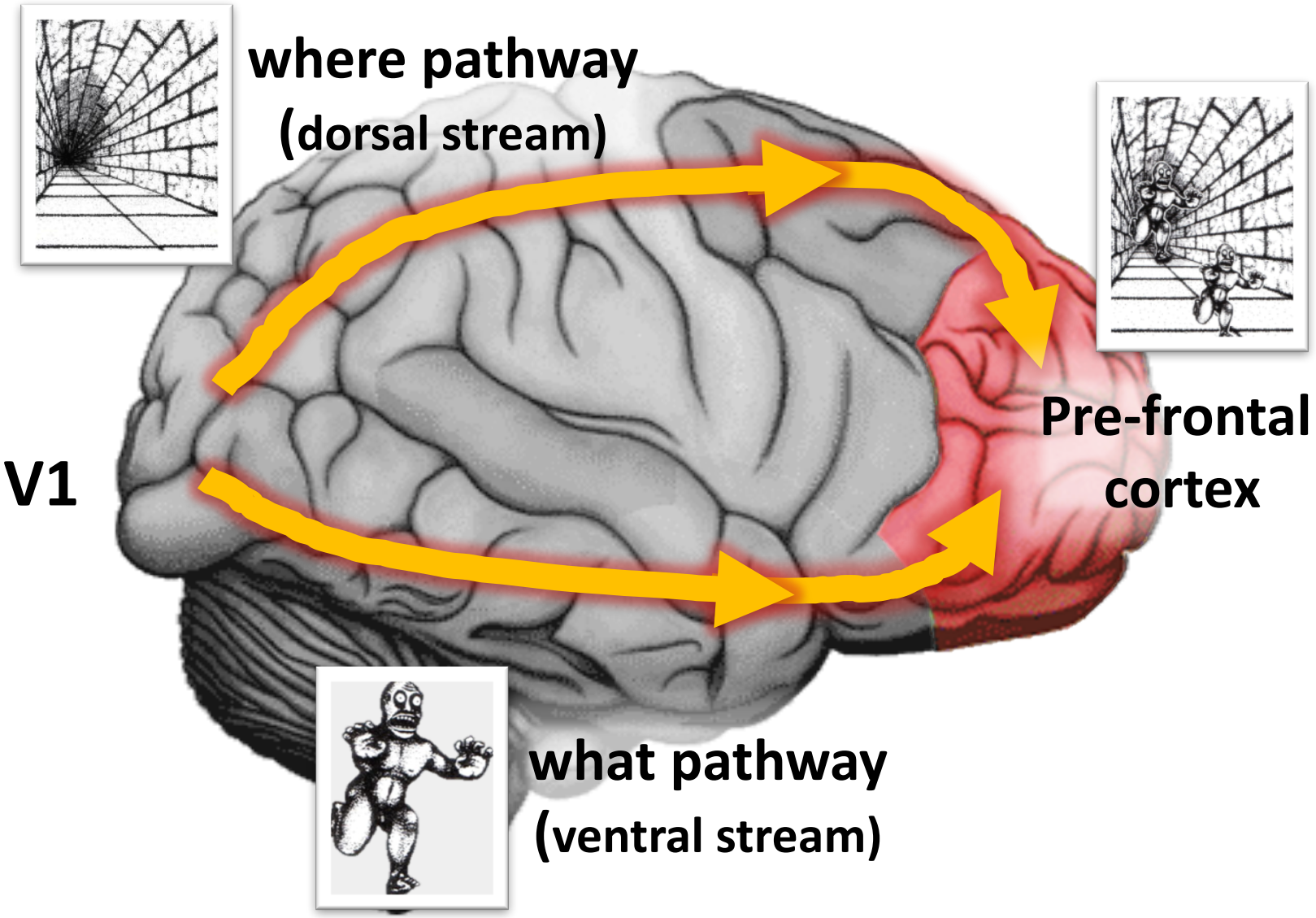


where pathway
(dorsal stream)

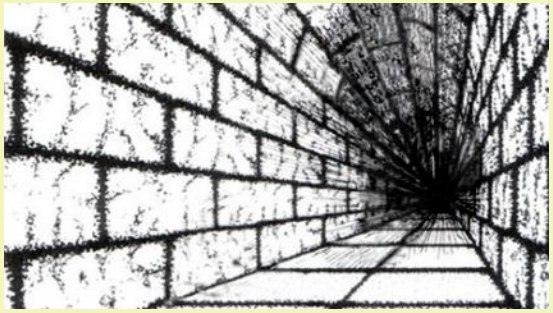
V1

what pathway
(ventral stream)

Visual processing in the brain



Two sides of one coin



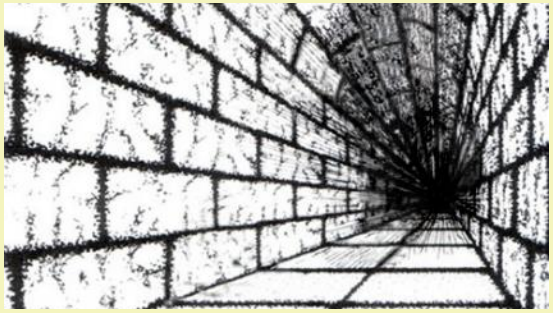
3D Reconstruction

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

2D Recognition

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

Two sides of one coin

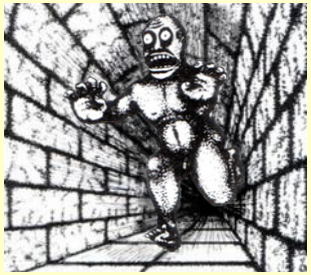


3D Reconstruction

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

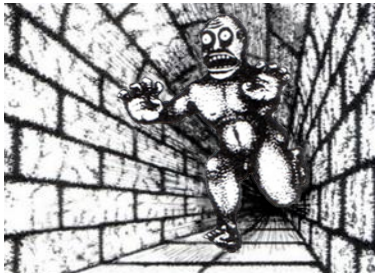
2D Recognition

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



Perceiving the World in 3D

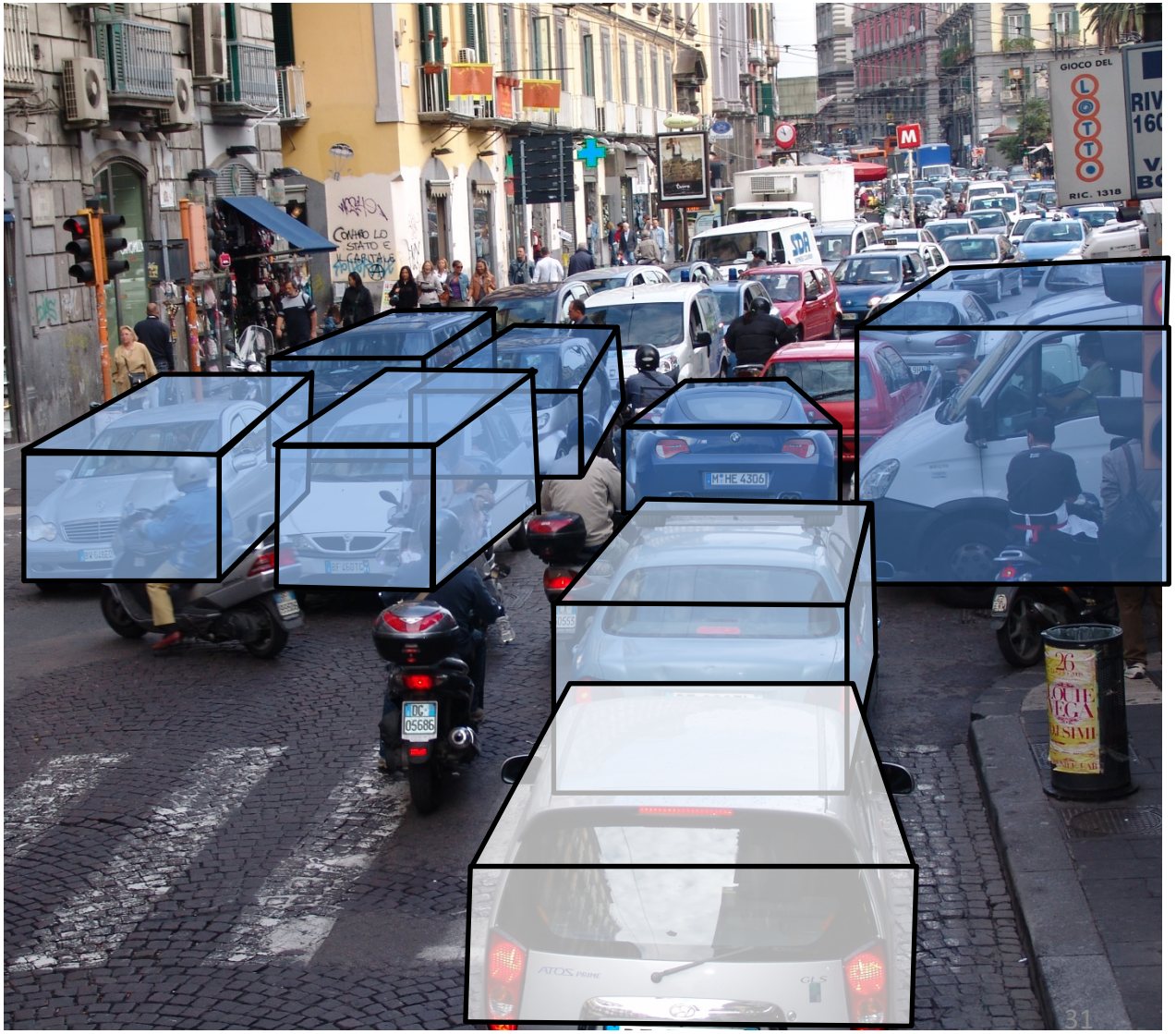
- Modeling objects and their 3D properties
- Modeling interaction among objects and space
- Modeling relationships of object/space across views



Outline

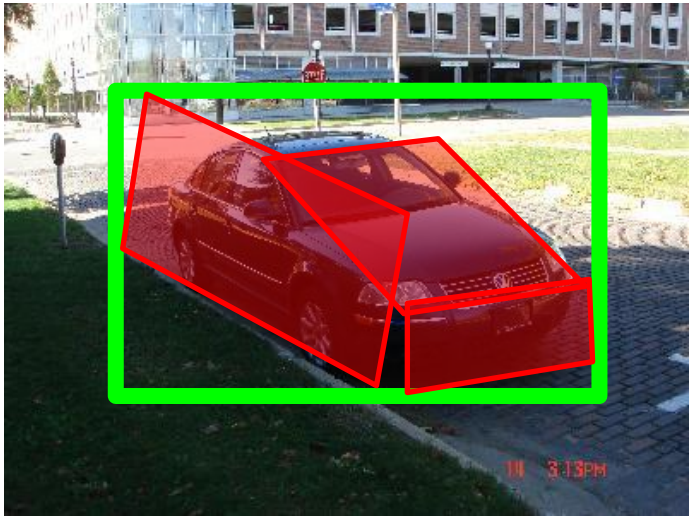
- Modeling objects and their 3D properties
- Modeling interaction among objects and space
- Modeling relationships of objects across views

Detecting objects and estimating their 3D properties

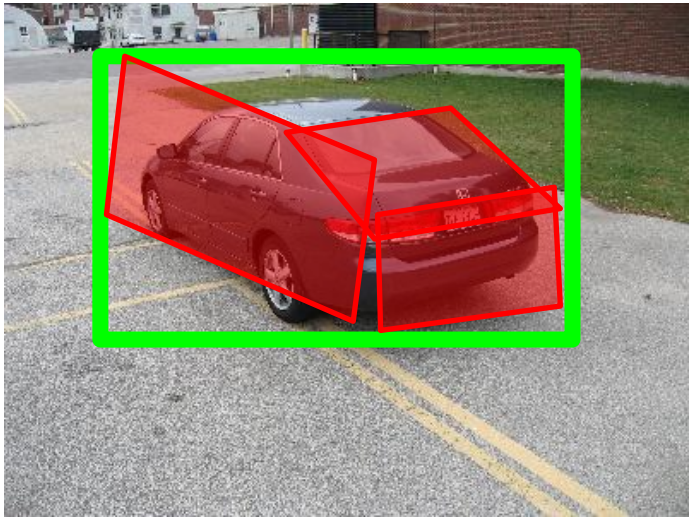


Results

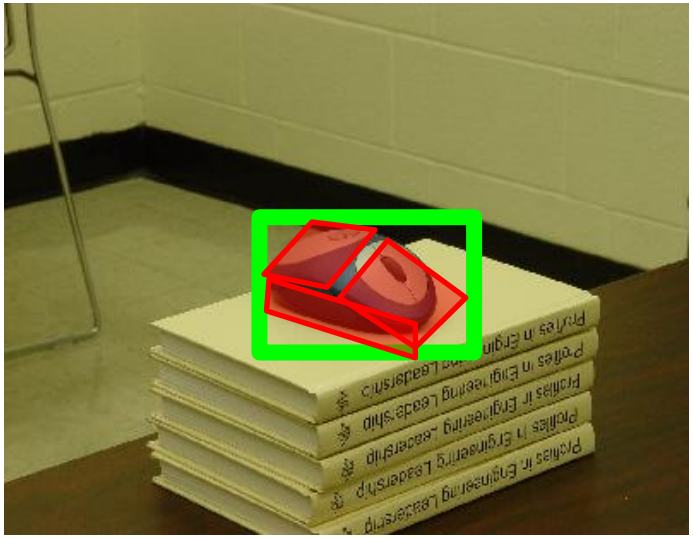
CAR a=330 e=15 d=7



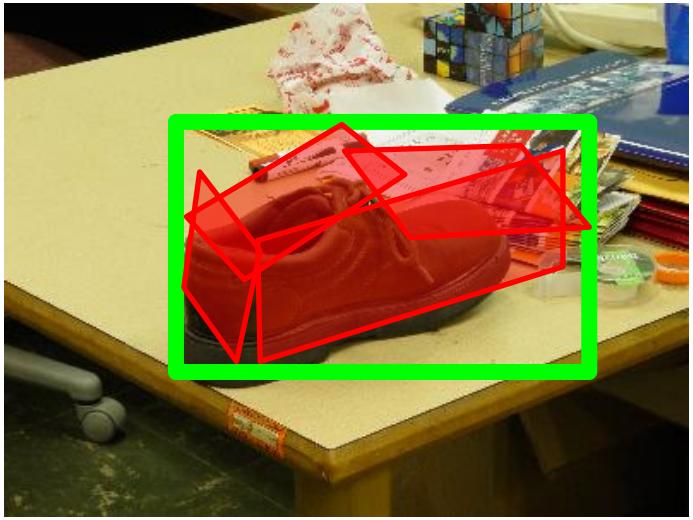
CAR a=150 e=15 d=7



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

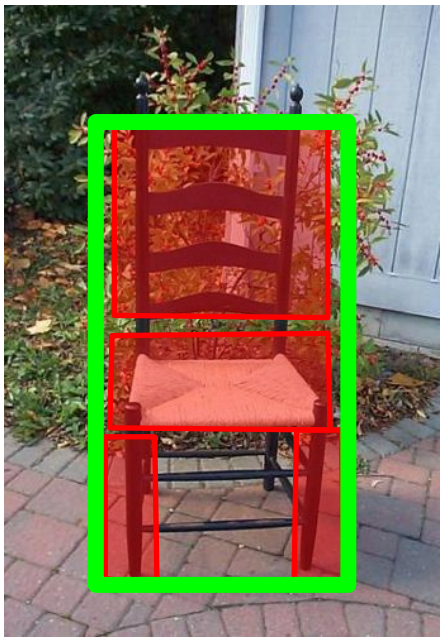
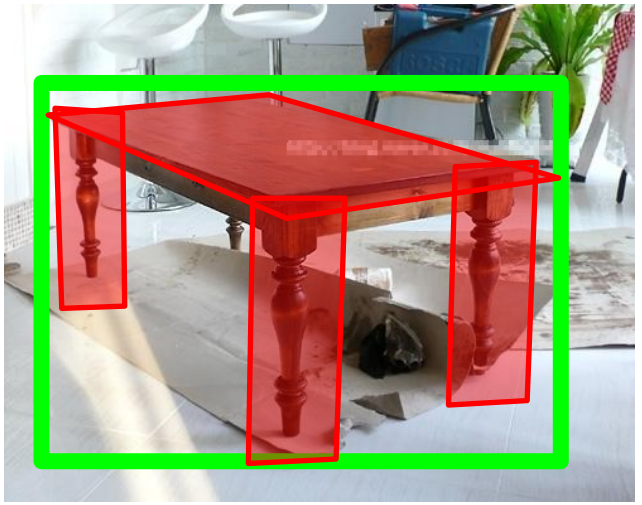
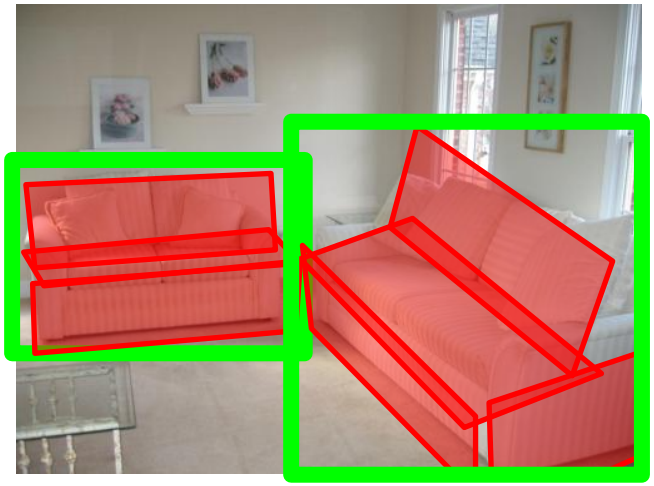


TABLE a=60 e=15 d=2



BED a=30 e=15 d=2.5

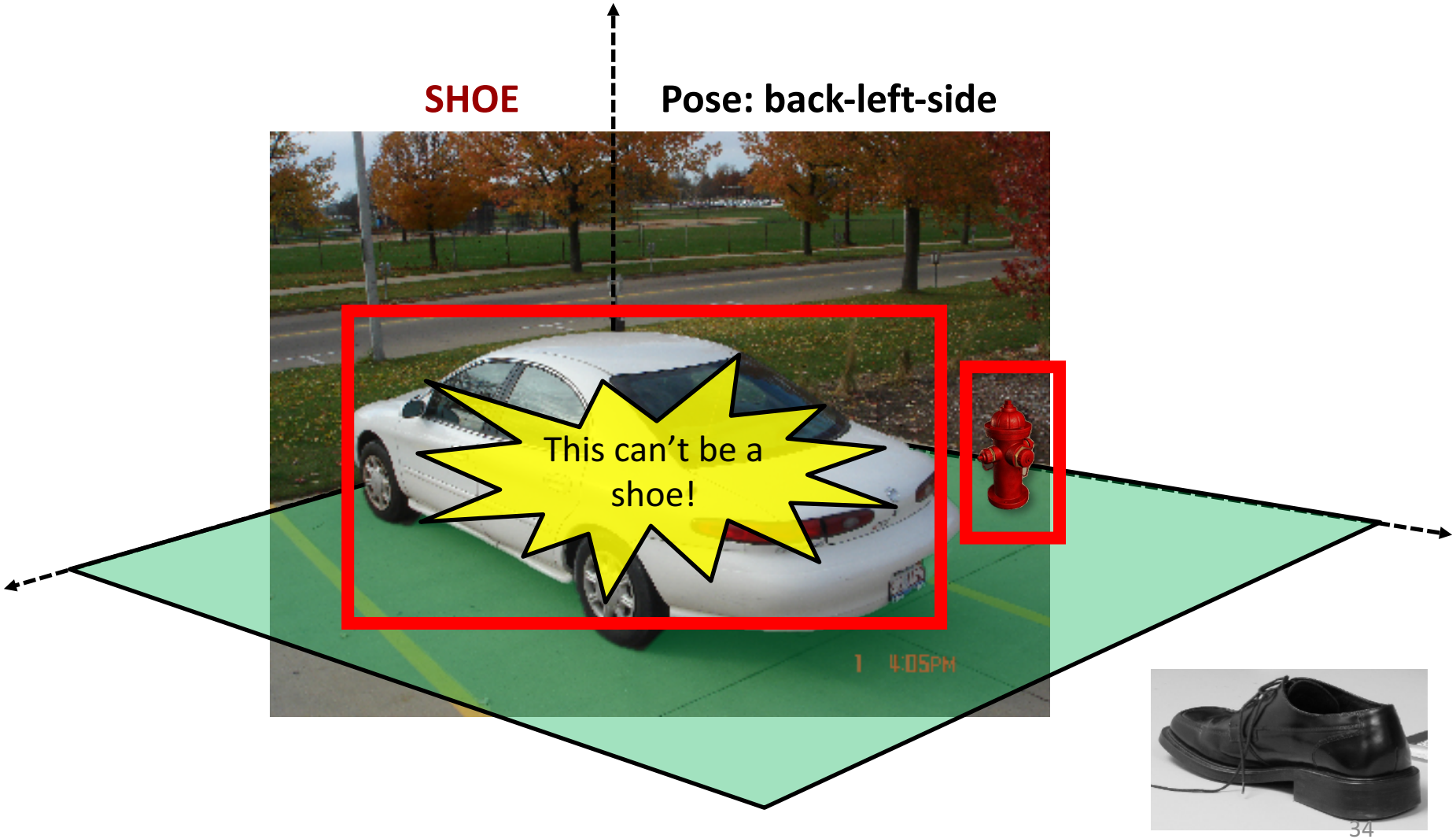
SOFA a=345 e=15 d=3.5
a=60 e-30 d=2.5

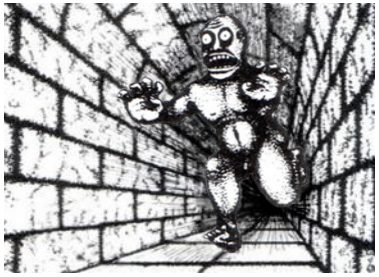


ImageNet dataset [Deng et al. 2010]

Results

Examples of failure (wrong category)

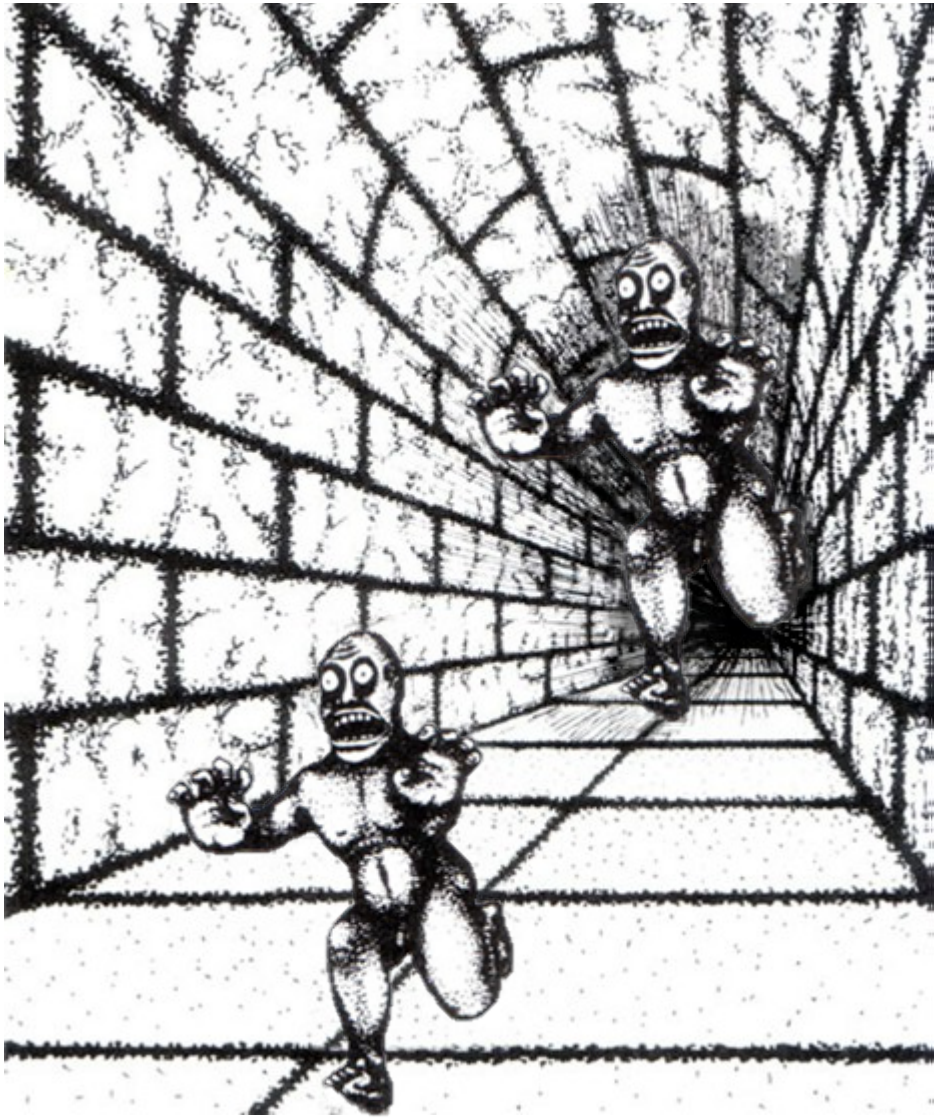




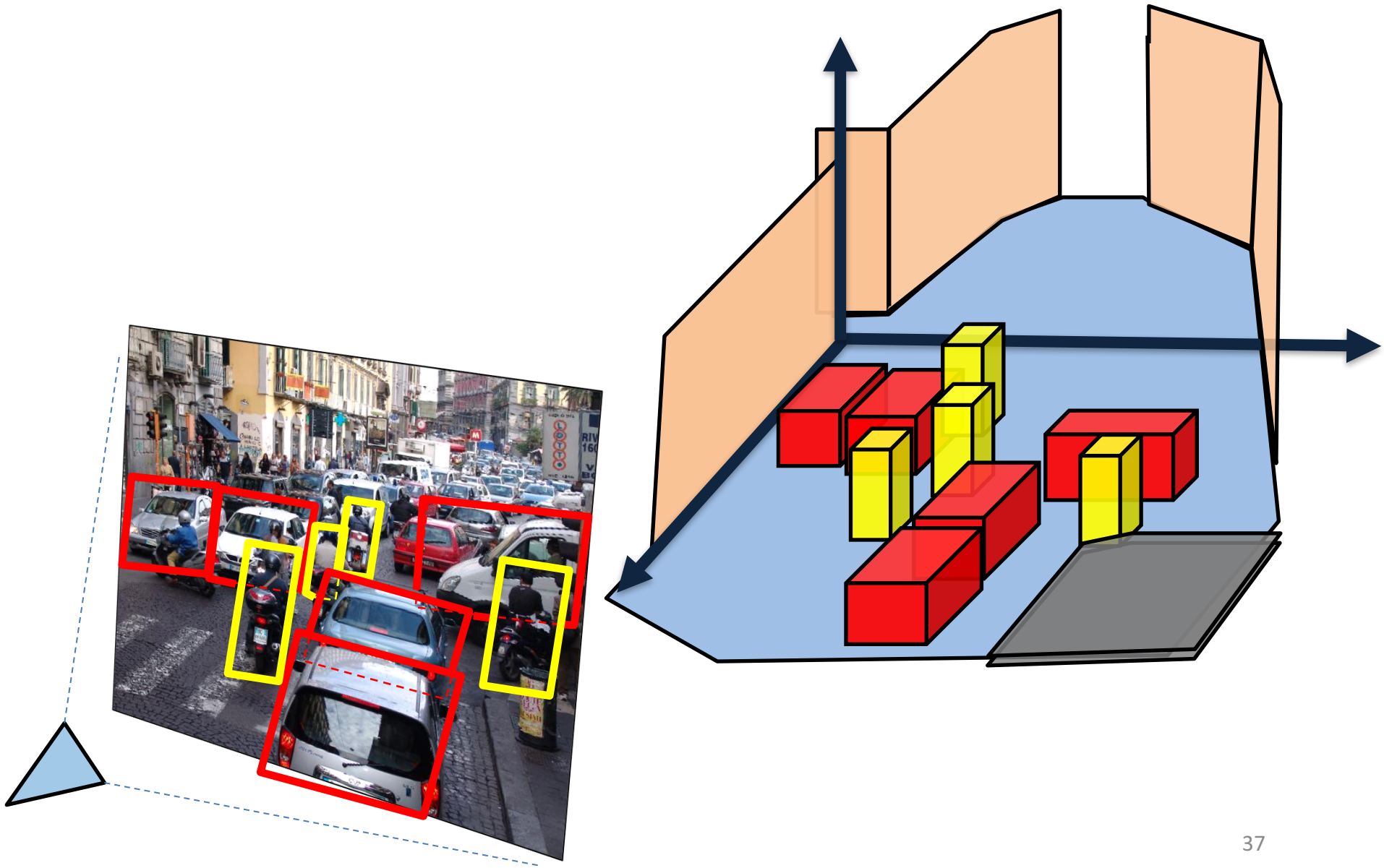
Outline

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

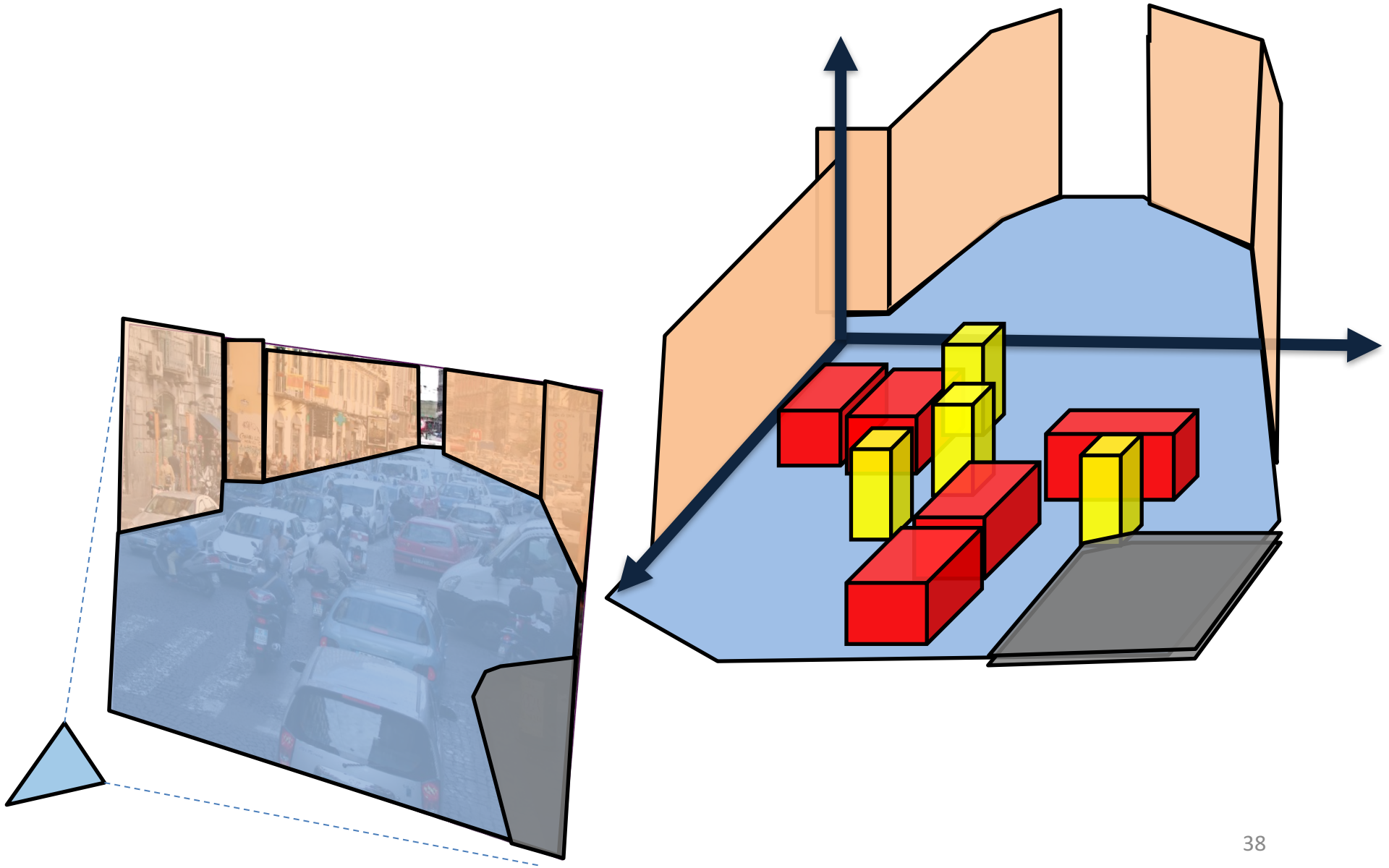
Scene understanding is an interplay between objects and space



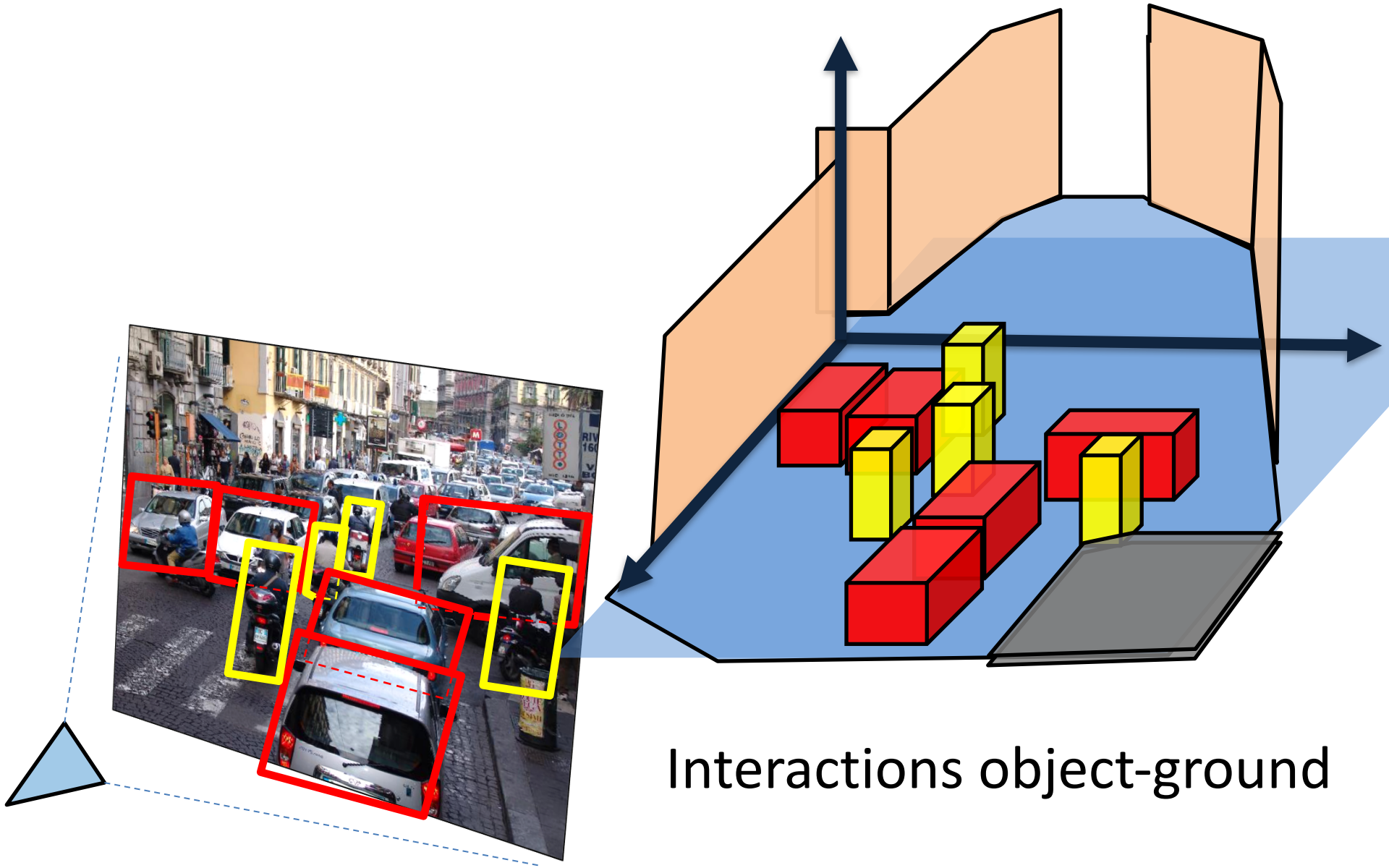
3D space is shaped by its objects



Objects are placed into 3D space



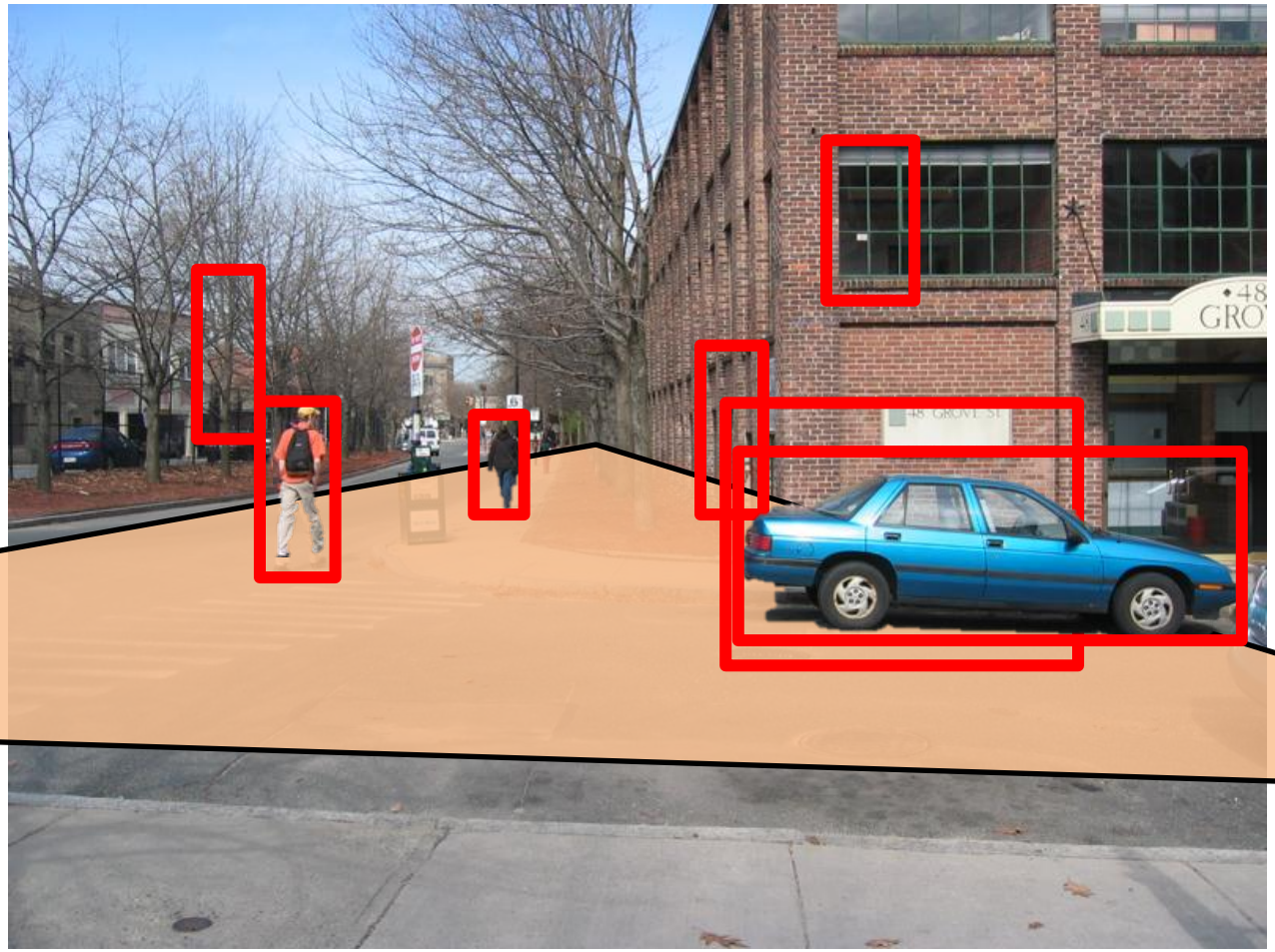
A first attempt….



Interactions object-ground

A first attempt….

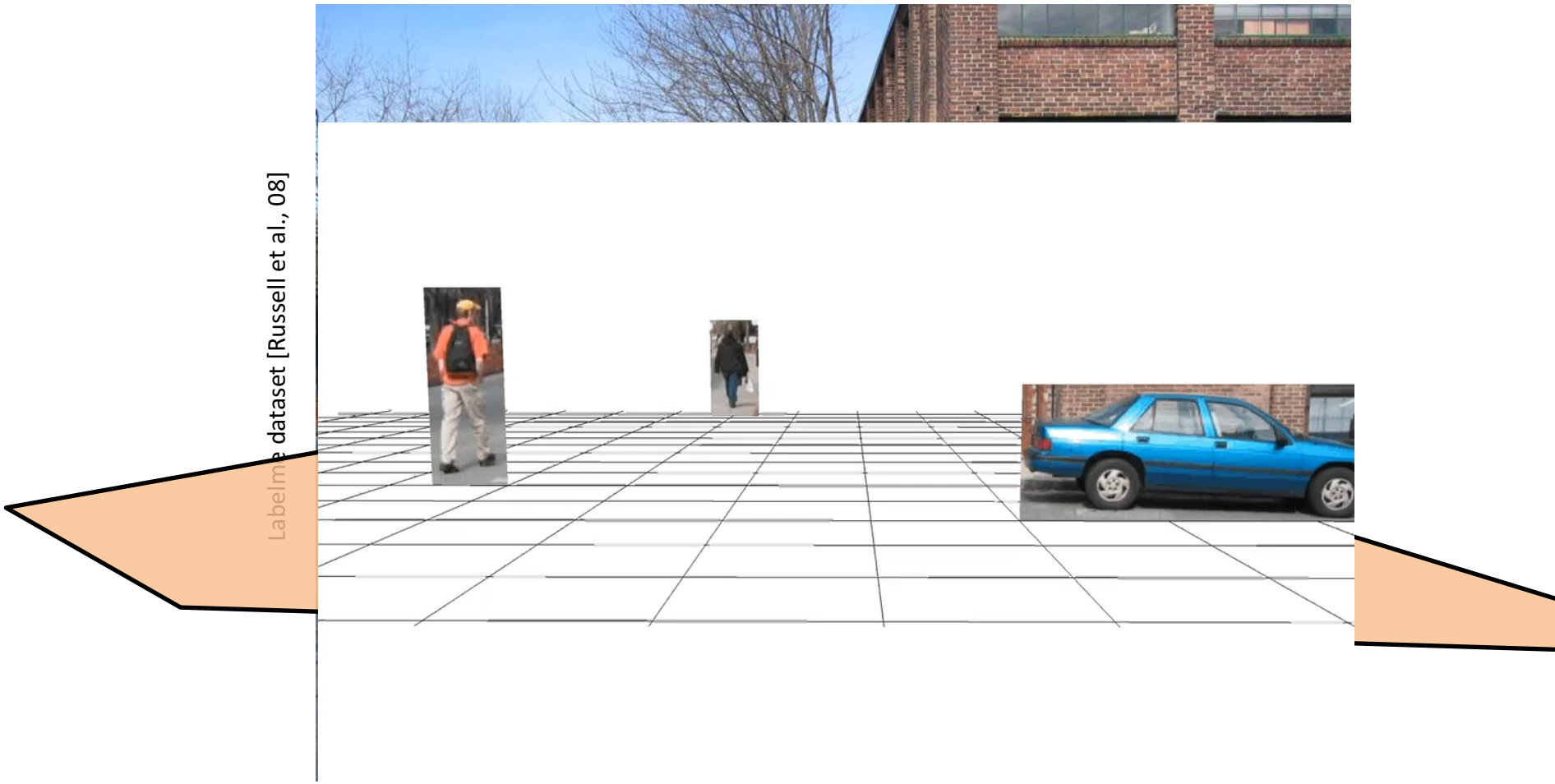
Hoiem et al, 2006-2008
Bao & Savarese, 2008-2012



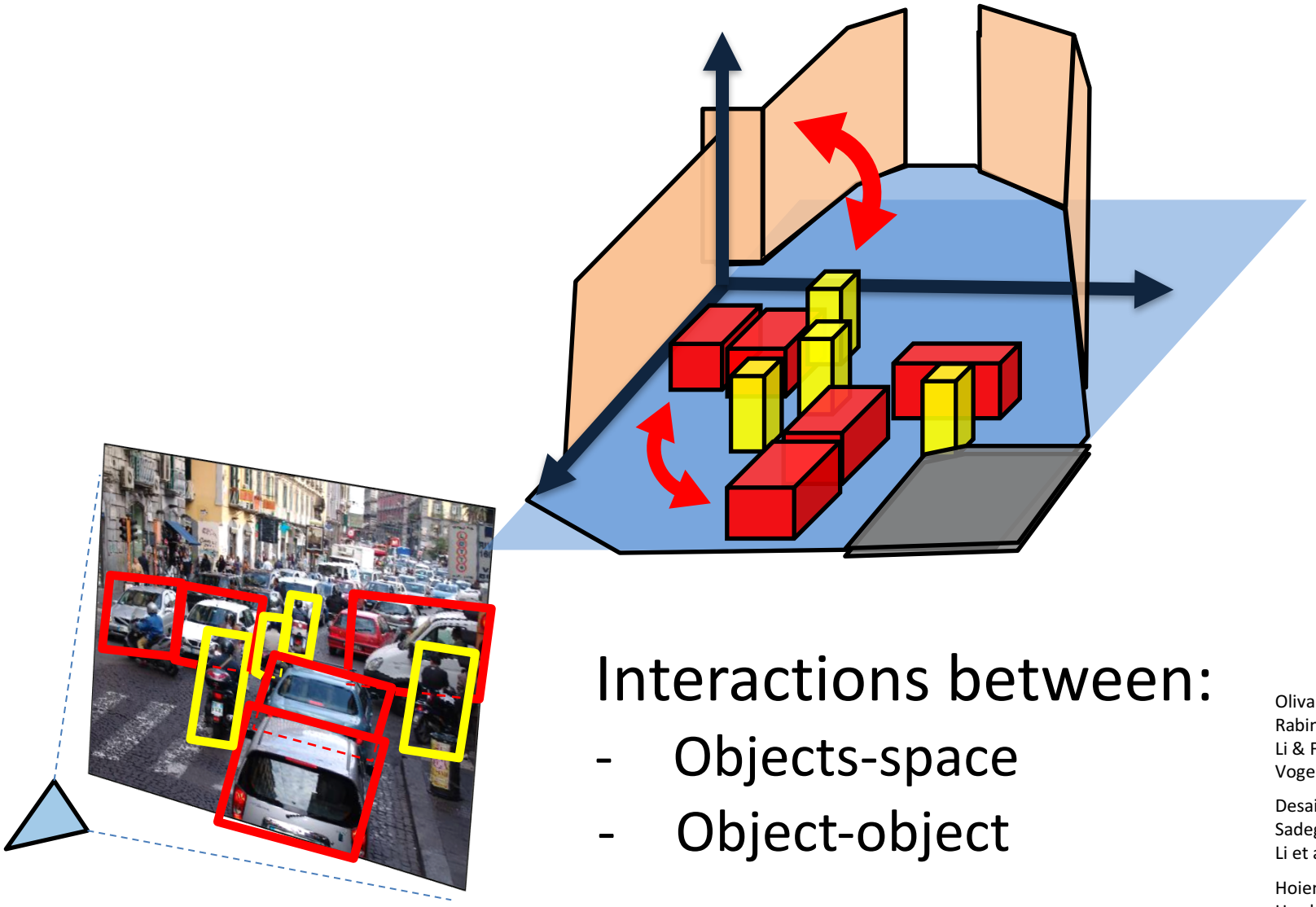
LabelMe dataset [Russell et al., 08]

A first attempt….

Hoiem et al, 2006-2008
Bao & Savarese, 2008-2012



Generalization #1



Interactions between:

- Objects-space
- Object-object

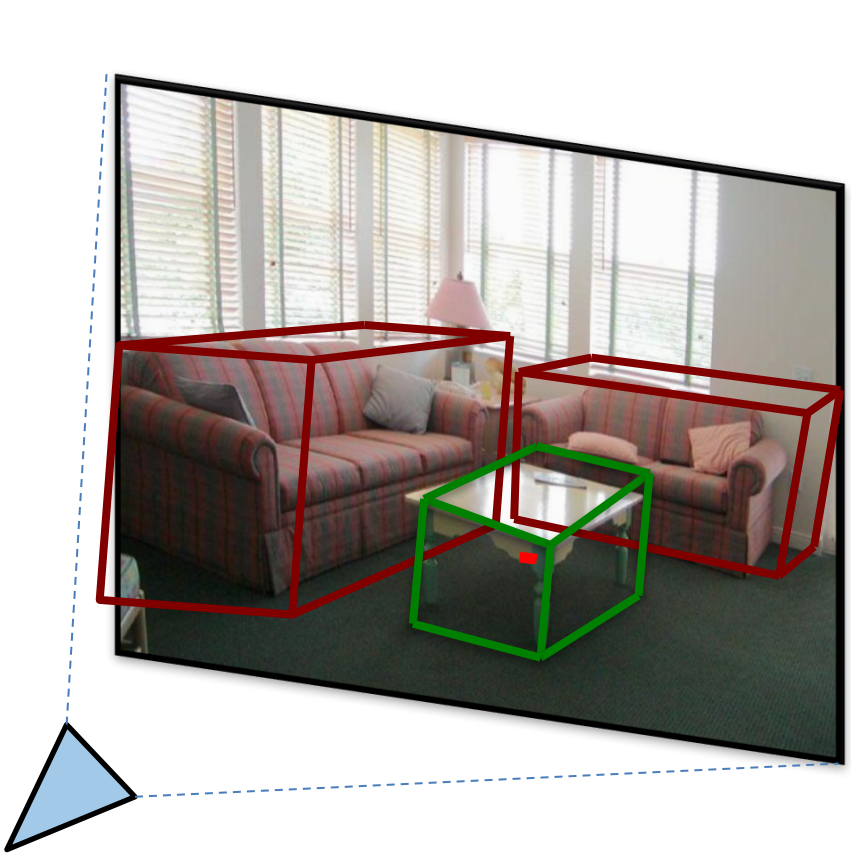
Oliva & Torralba, 2007
Rabinovich et al, 2007
Li & Fei-Fei, 2007
Vogel & Schiele, 2007

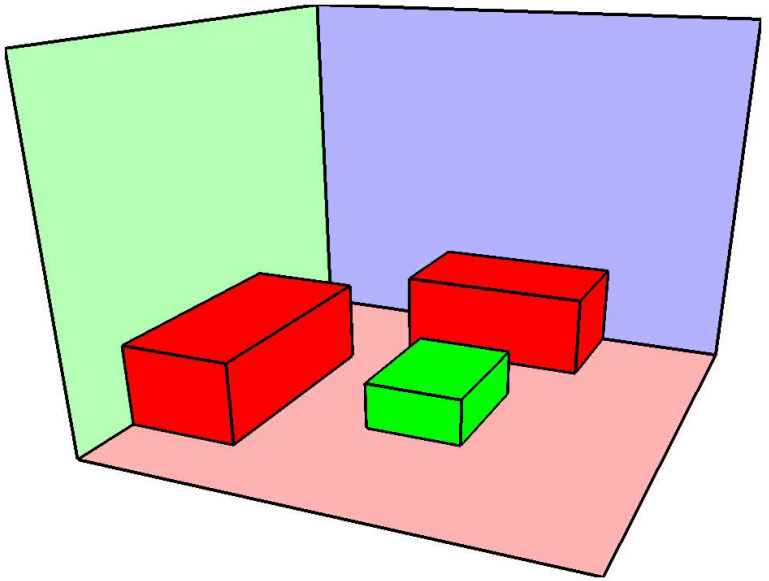
Desai et al, 2009
Sadeghi & Farhardi, 2011
Li et al, 2012

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

3D geometric phrases

Choi, Chao, Pantofaru, Savarese, CVPR 13 , IJCV 15

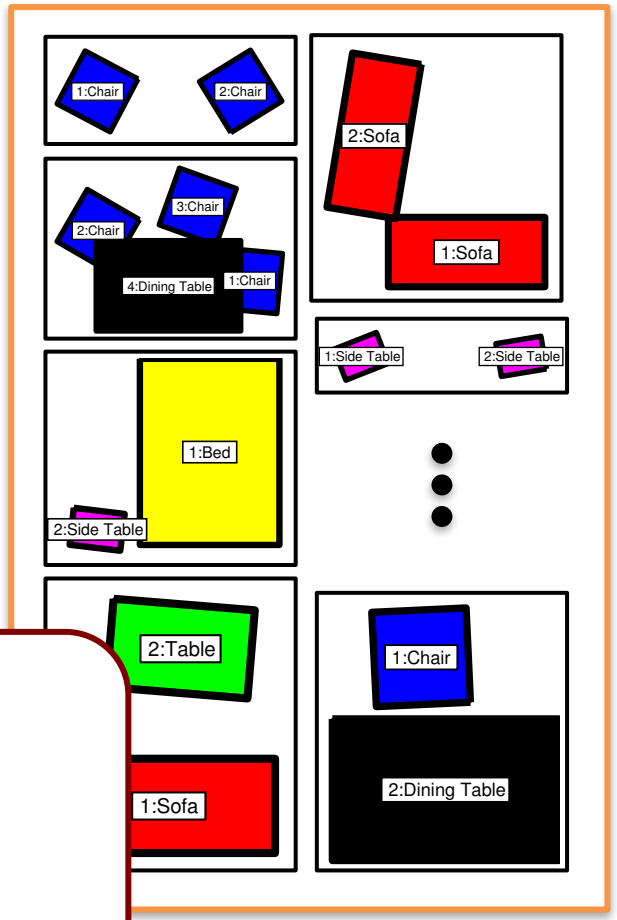




3D Geometric Phrases

Choi, Chao, Pantofaru, Savarese, CVPR 13 , IJCV 15

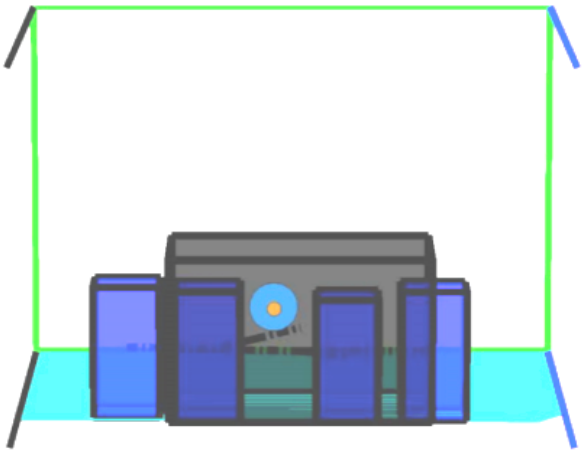
Trainin



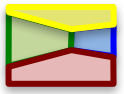
3DGPs

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

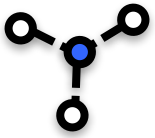
Enable contextual awareness!



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

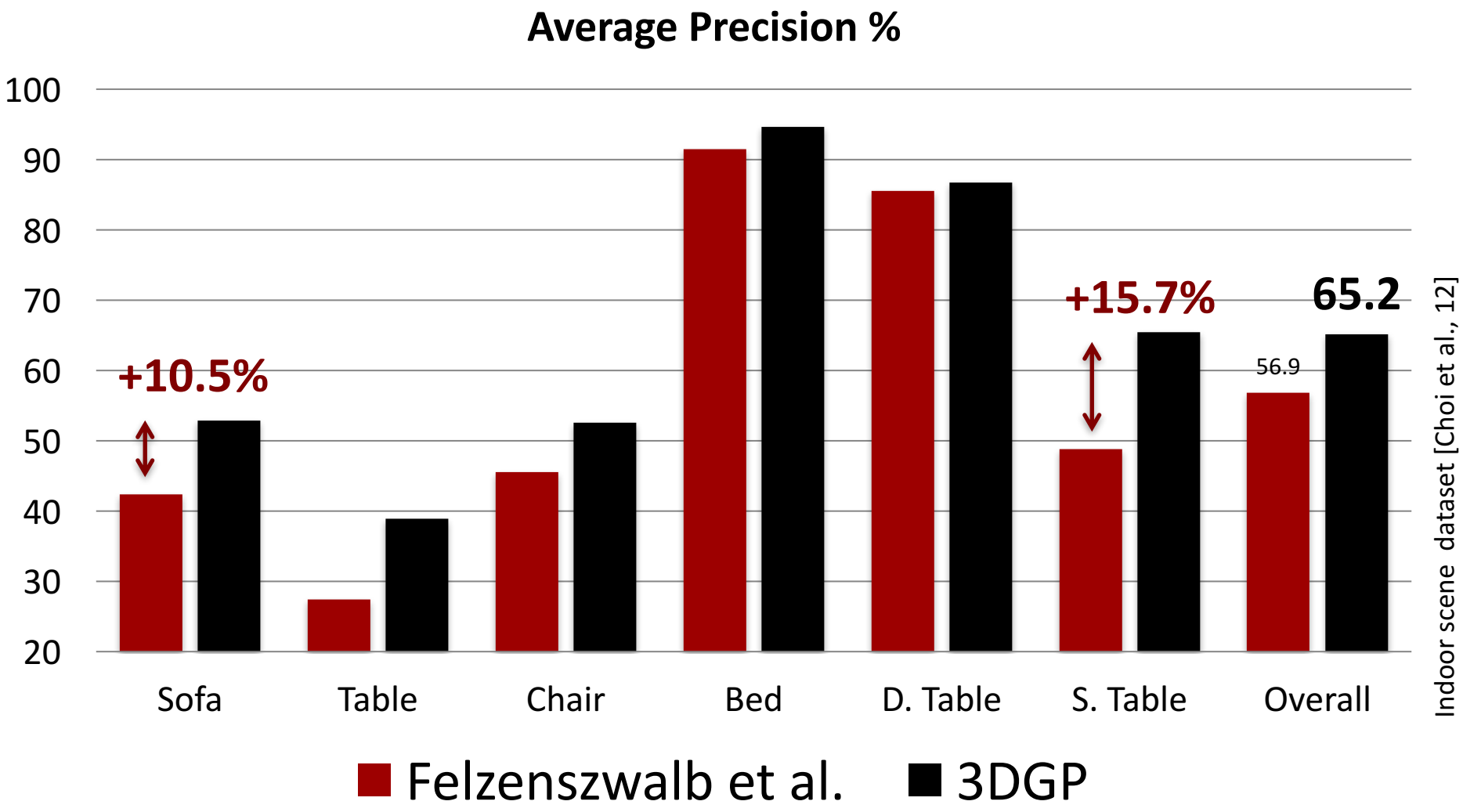


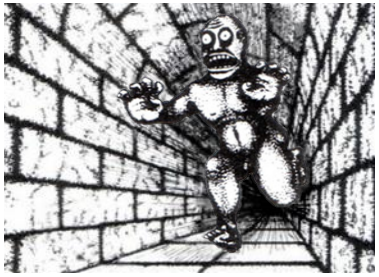
Estimated Layout



3D Geometric Phrases

Results: Object Detection

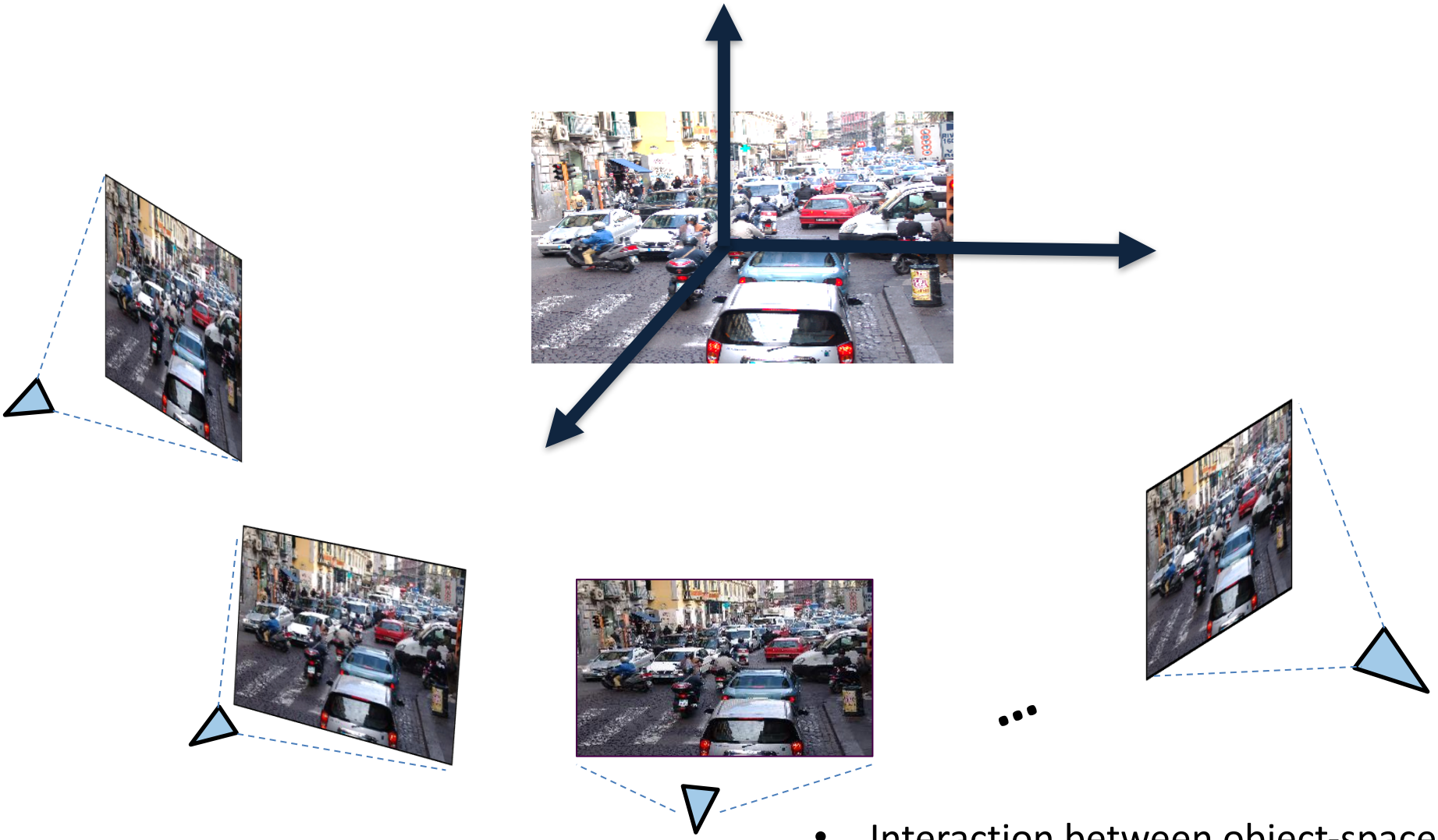




Outline

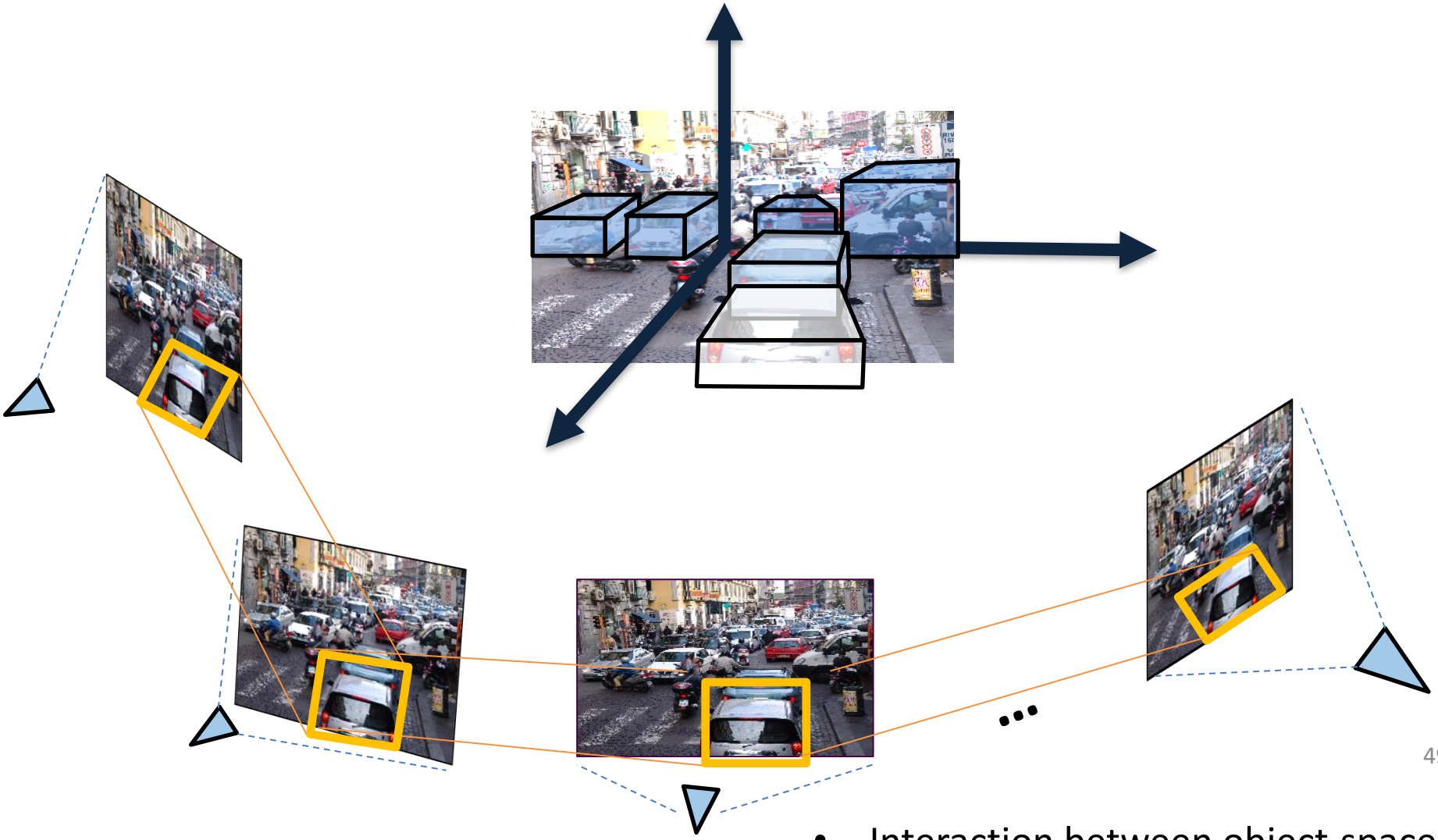
- Modeling objects and their 3D properties
- Modeling interaction among objects and space
- Modeling relationships of objects across views

Modeling relationships of objects across views



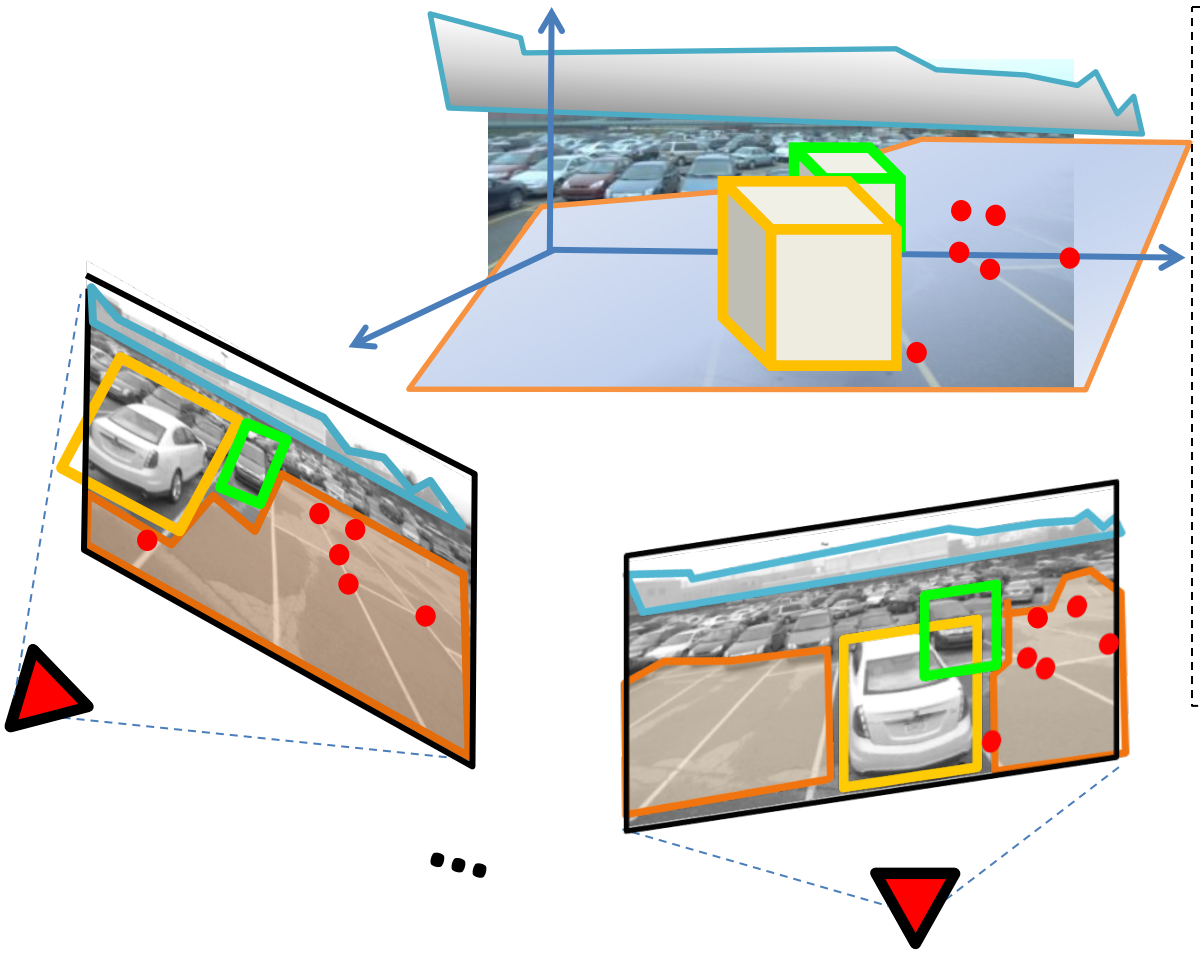
- Interaction between object-space
- Interaction among objects
- **Transfer semantics across views**

Modeling relationships of objects across views



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

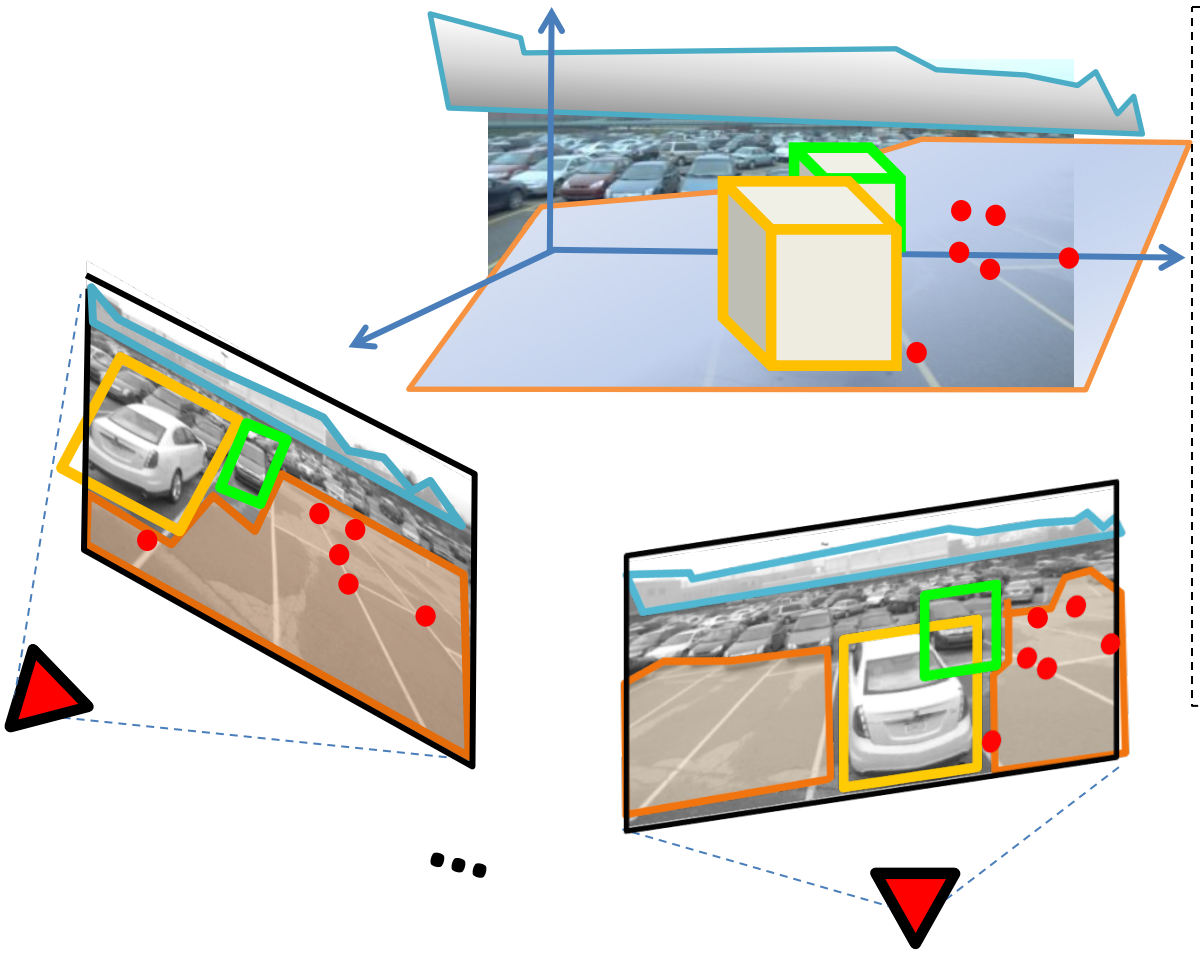
Semantic structure from motion



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

Semantic structure from motion

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

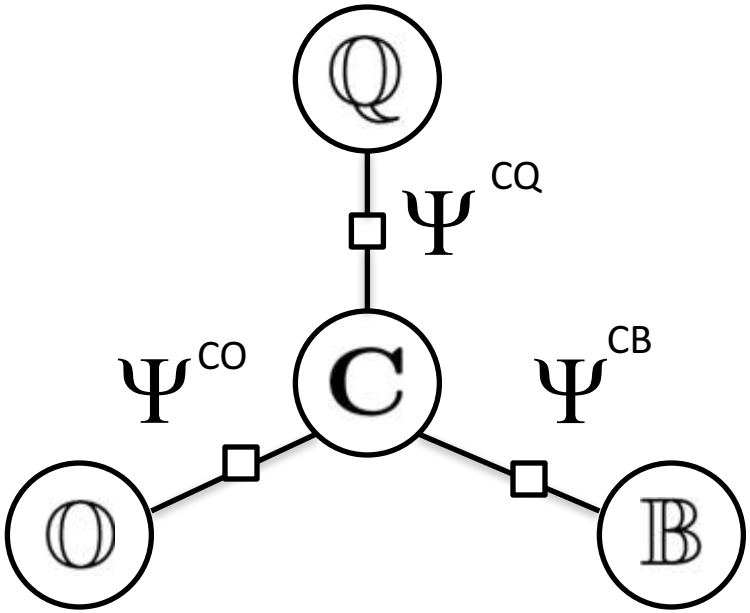


- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

Semantic structure from motion

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

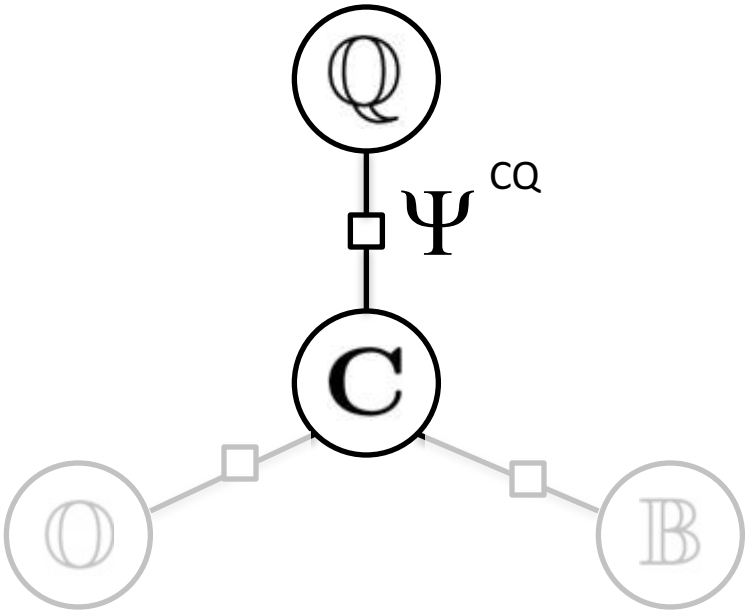
Factor graph



- 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

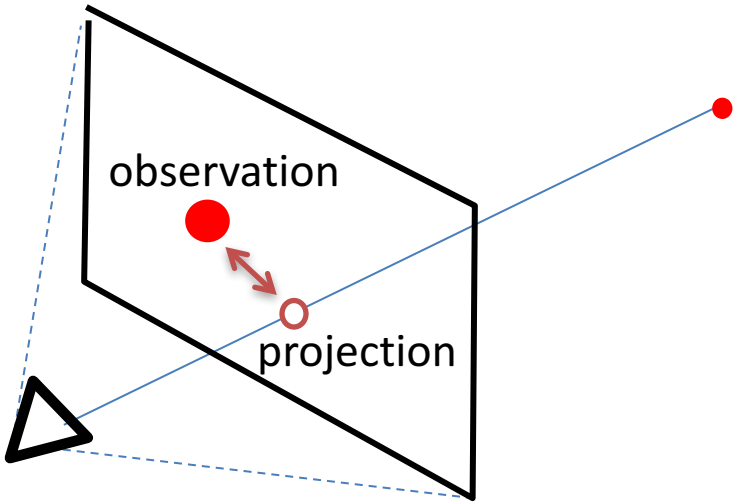
$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg \max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \prod_s \Psi_s^{CQ} \prod_t \Psi_t^{CO} \prod_r \Psi_r^{CB}$$



- 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

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



- 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

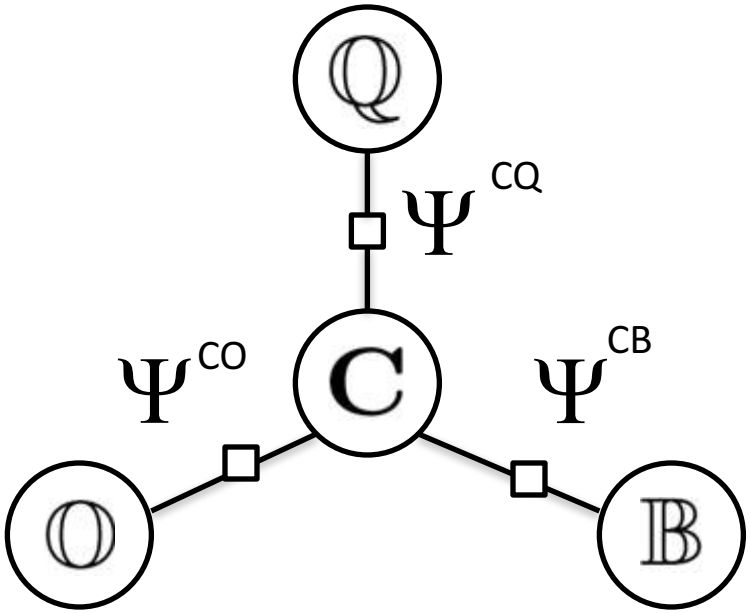
Point re-projection error

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

- 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

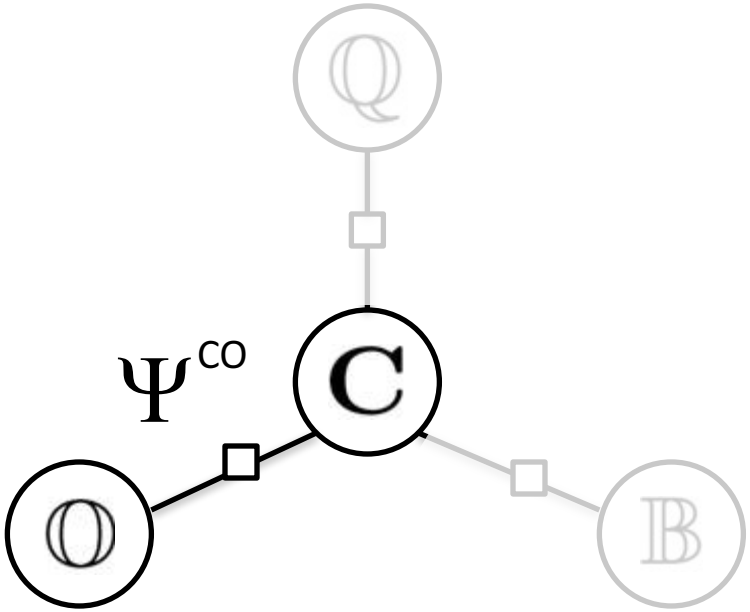
$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg \max_{\mathbb{Q},\mathbb{O},\mathbb{B},\mathbf{C}} \prod_s \Psi_s^{CQ} \prod_t \Psi_t^{CO} \prod_r \Psi_r^{CB}$$



- 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

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

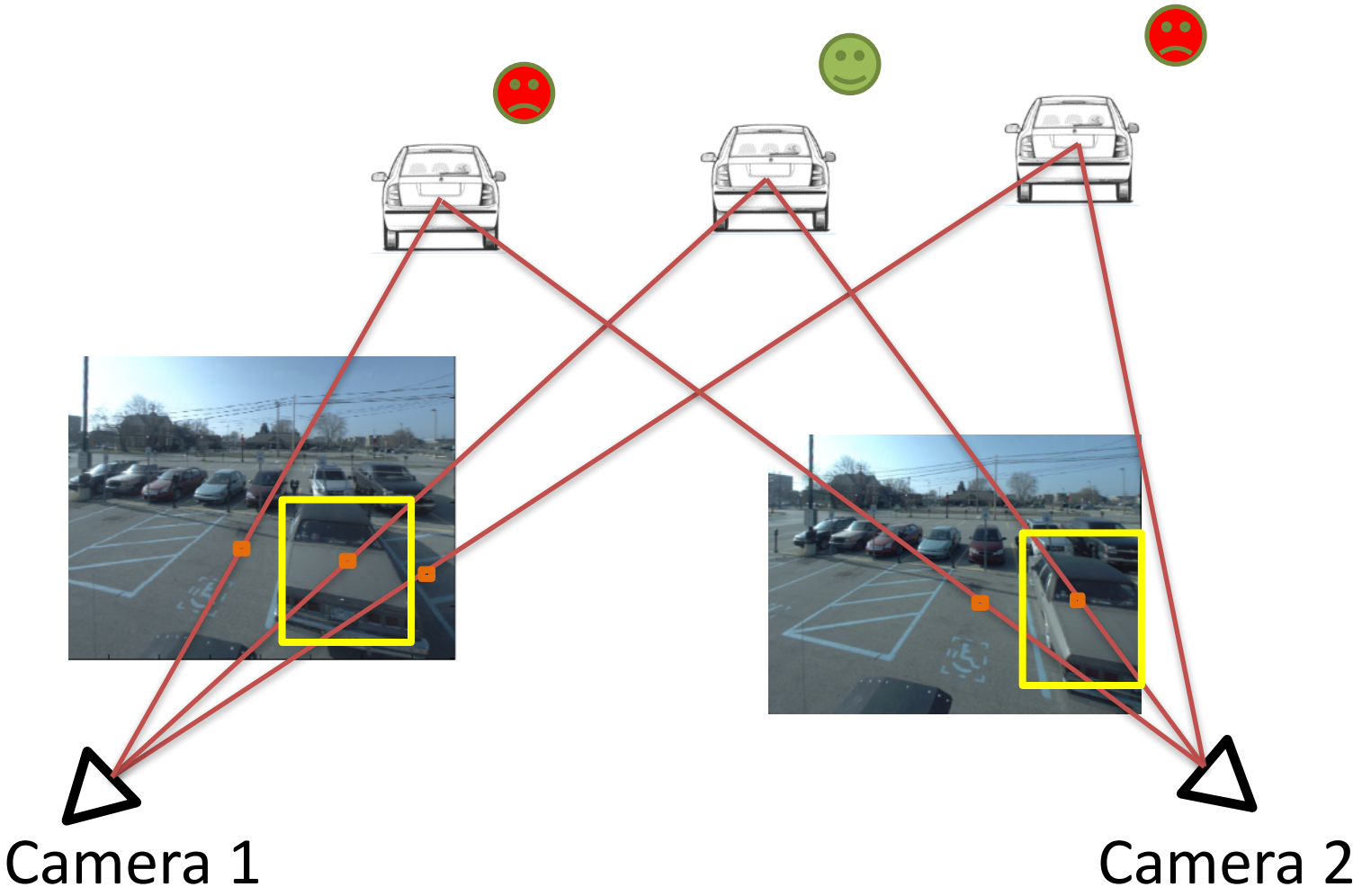


- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

Object "re-projection" error

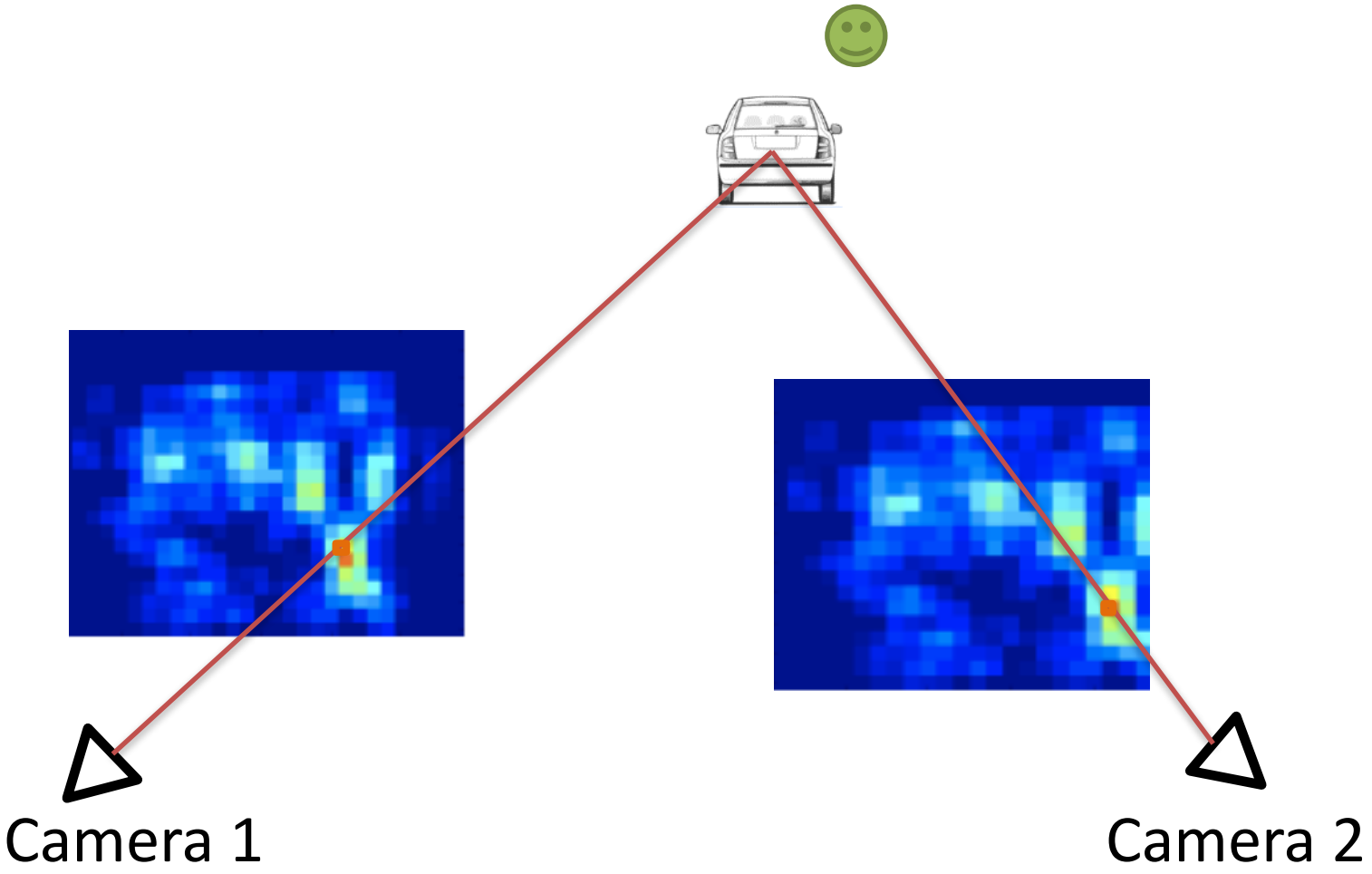
$$\Psi_t^{CO} \propto \prod_t^{N_t} (1 - \prod_k^{N_k} (1 - \Pr(o|O_t, C^k)))$$

SSFM: Object-level compatibility



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

SSFM: Object-level compatibility

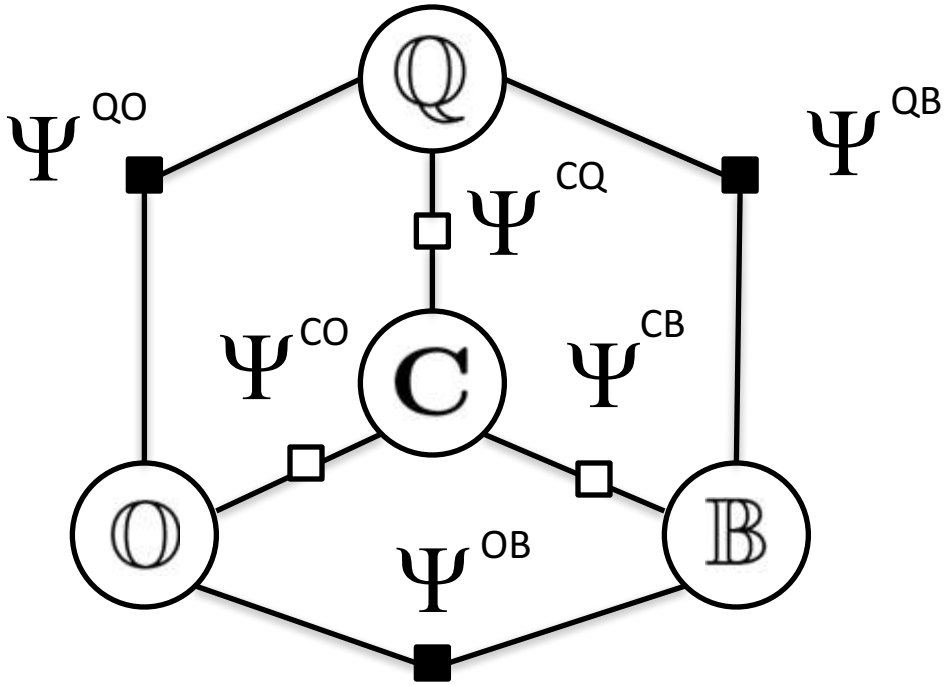


- Agreement with measurements is computed using position, pose and scale

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

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

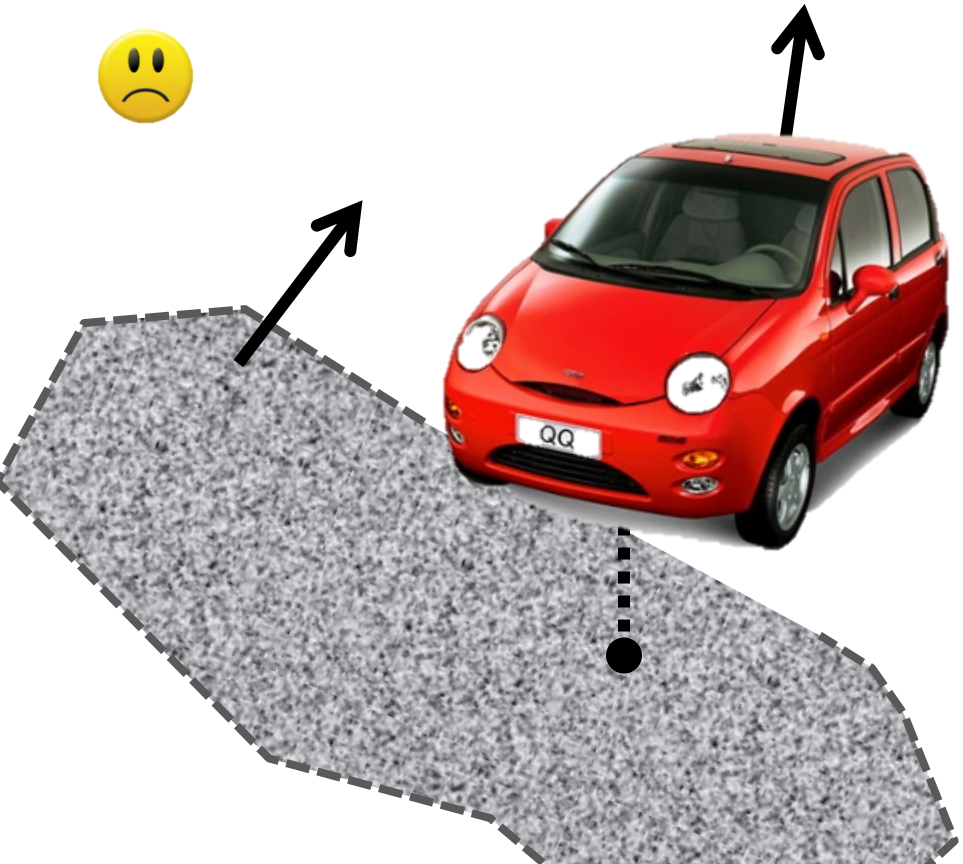
Object-Region Interactions:

- 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

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

Object-Region Interactions:

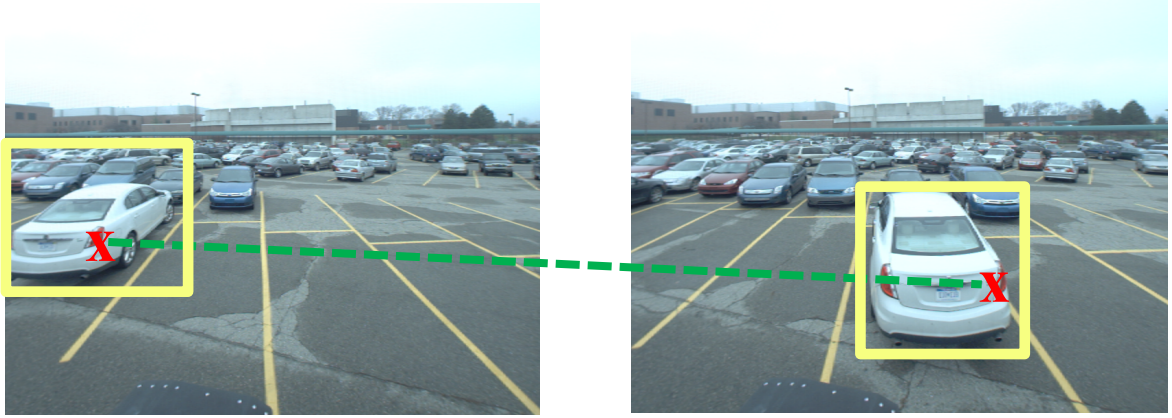


- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

SSFM with interactions

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

Object-point Interactions:

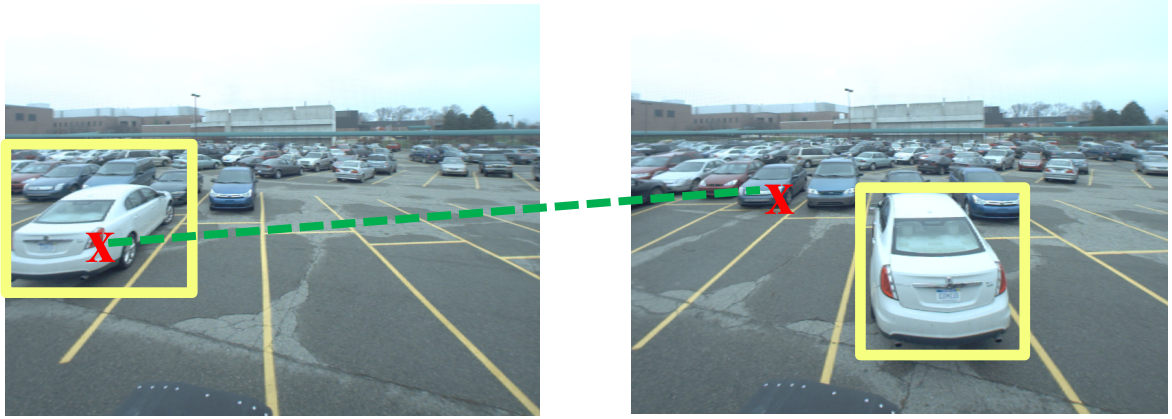


- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

SSFM with interactions

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

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

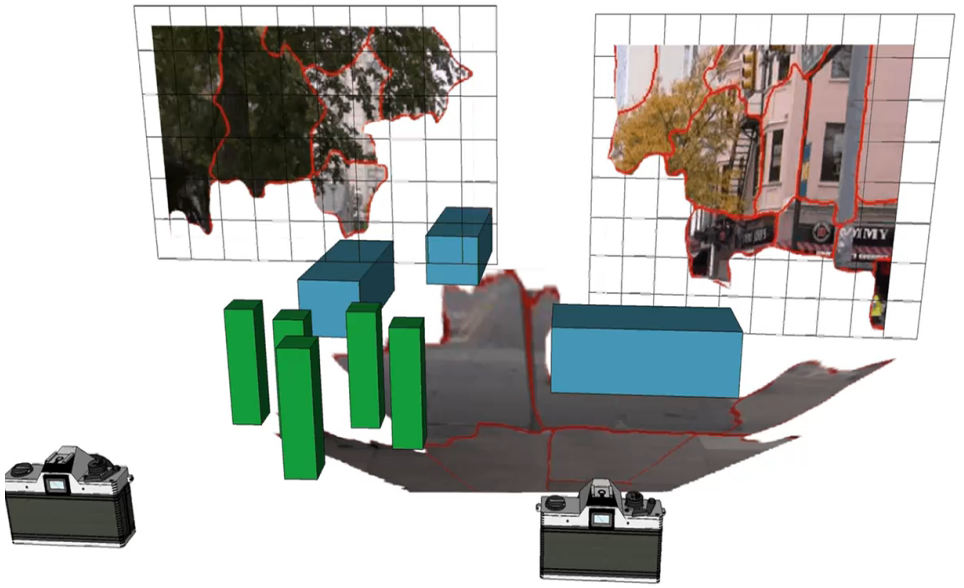
Input images

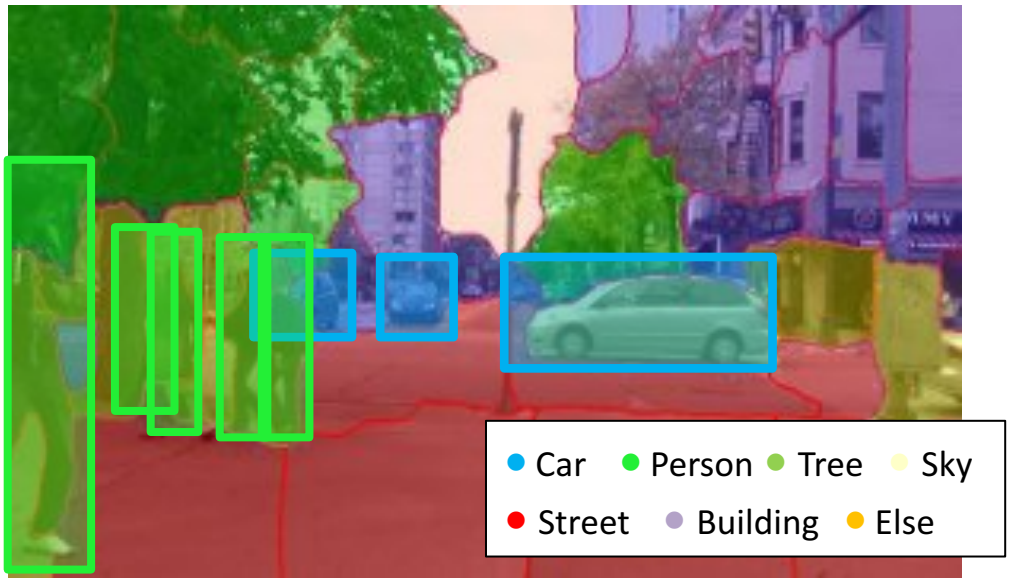
FORD CAMPUS dataset [Pandey et al., 09]

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

Results

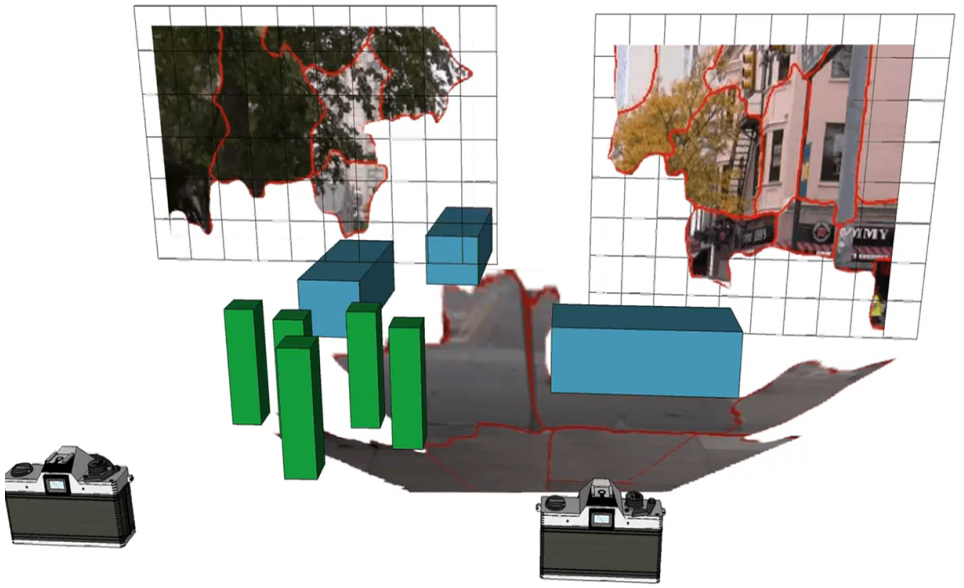
Input images

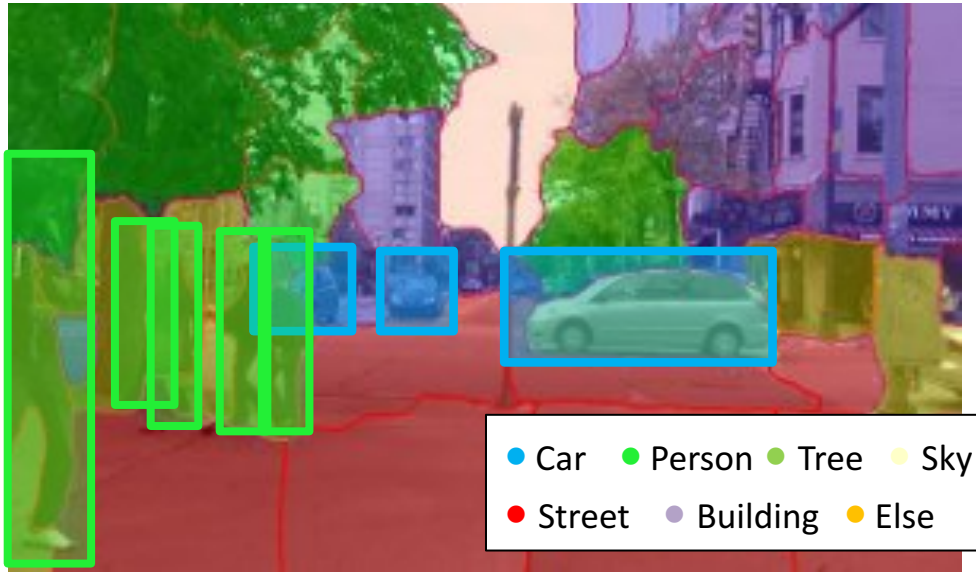




Results

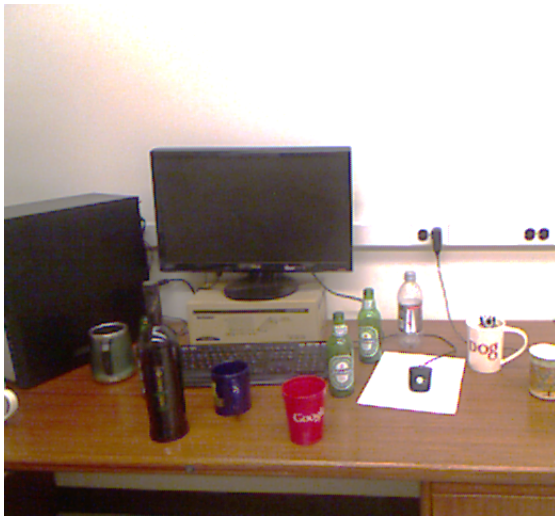
Input images

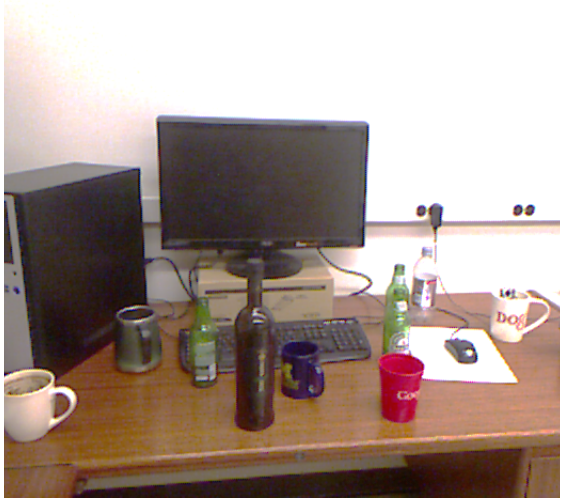




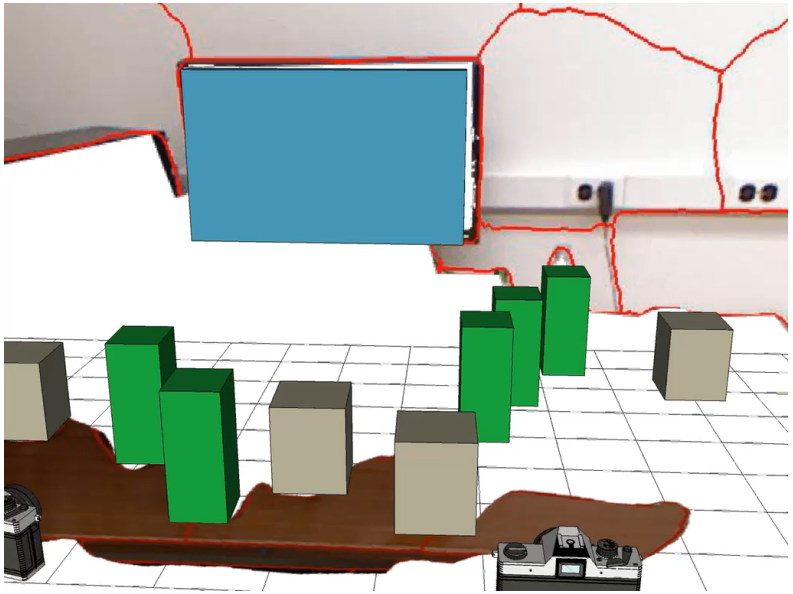
Results

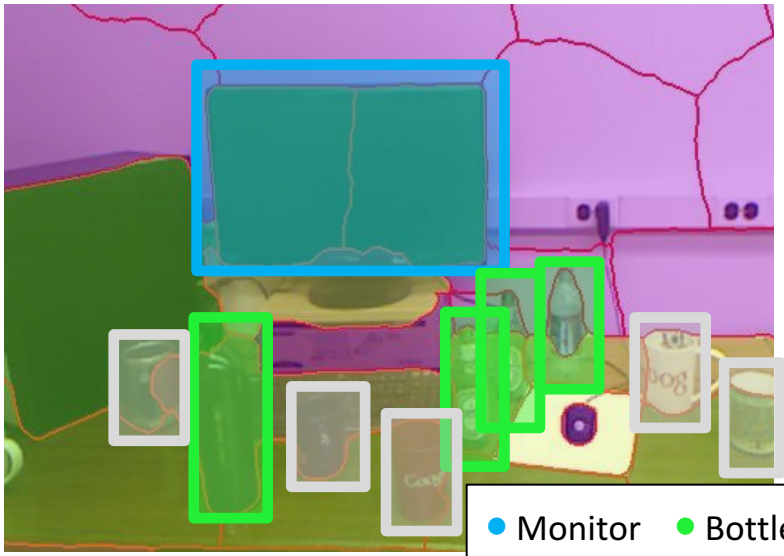
Input images





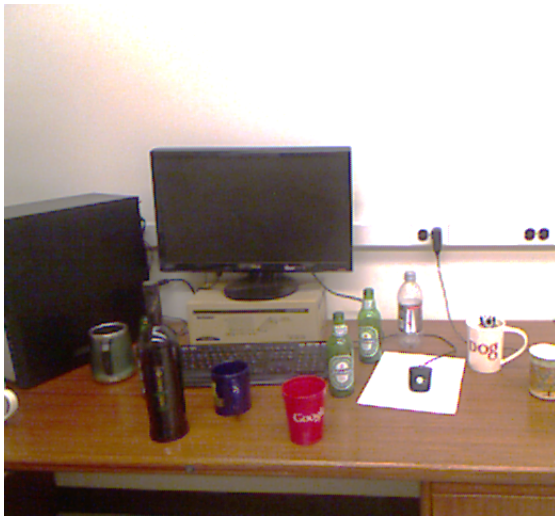
From the office dataset [Bao et al., 11]

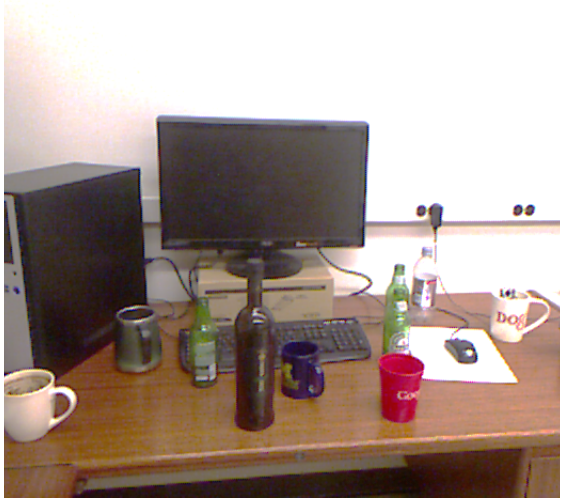




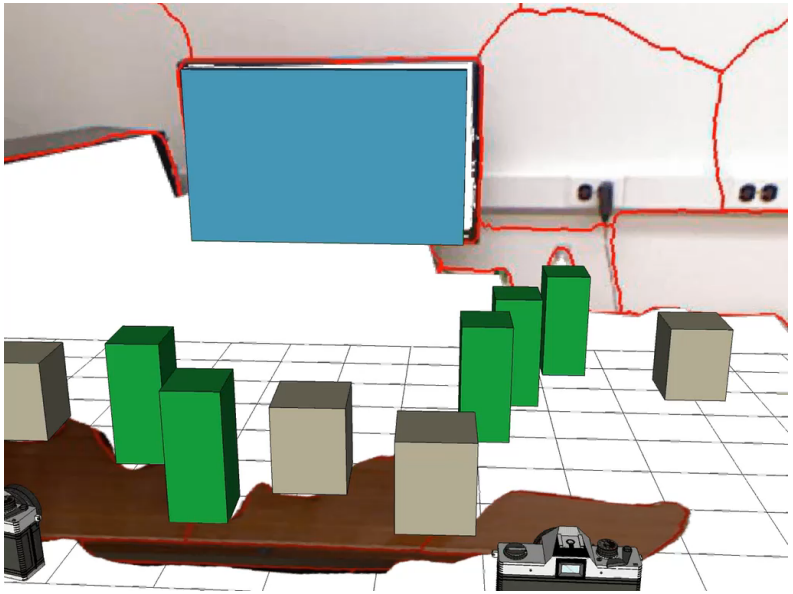
Results

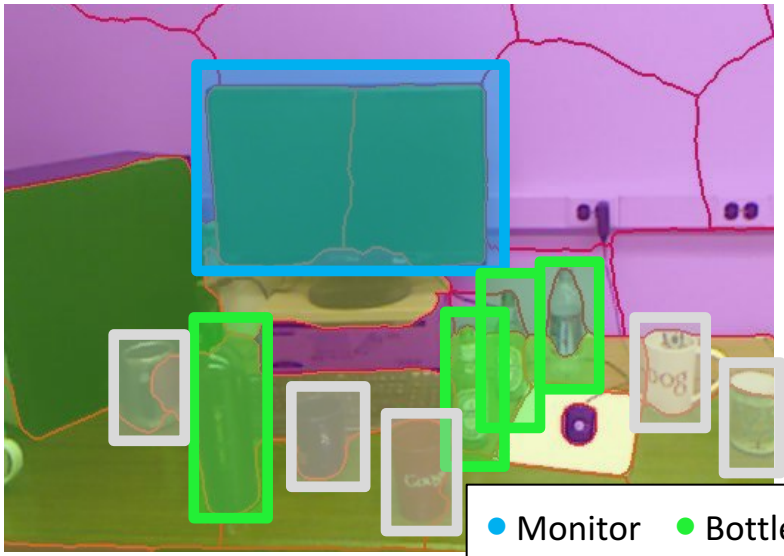
Input images





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

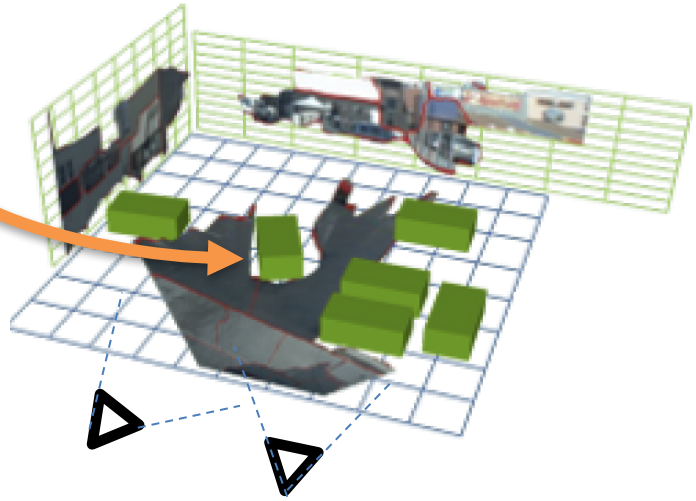




Results

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%**

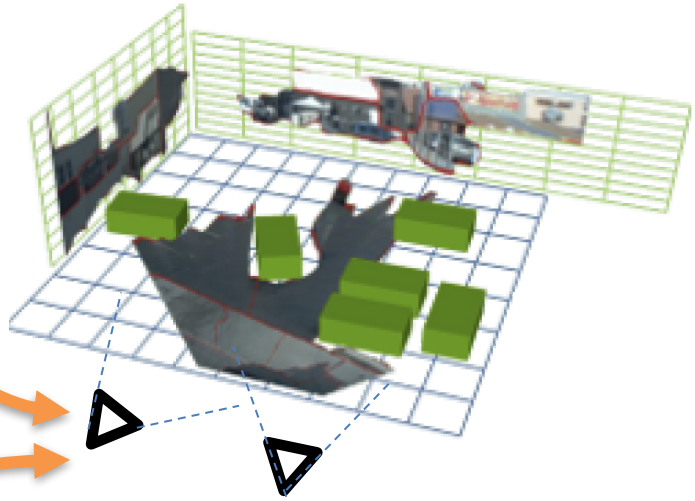
FORD CAMPUS dataset [Pandey et al., 09]
Office dataset [Bao et al., 11]

[1] Felzenszwalb et al. 2008

Results

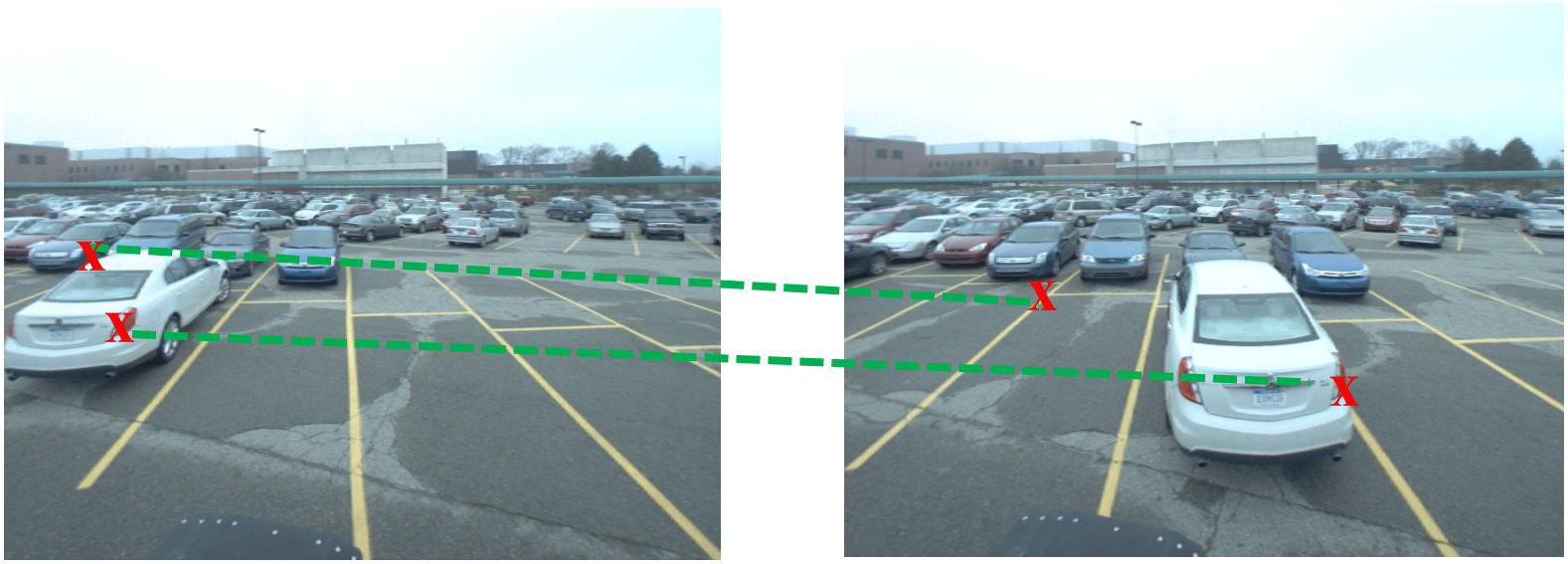
	Camera translation error		
	SFM Snavely et al., 08	**SSFM** no int.	**SSFM**
FORD CAMPUS	26.5°	19.9°	**12.1°**
OFFICE	8.5°	4.7°	**4.2°**
STREET	27.1°	17.6°	**11.4°**

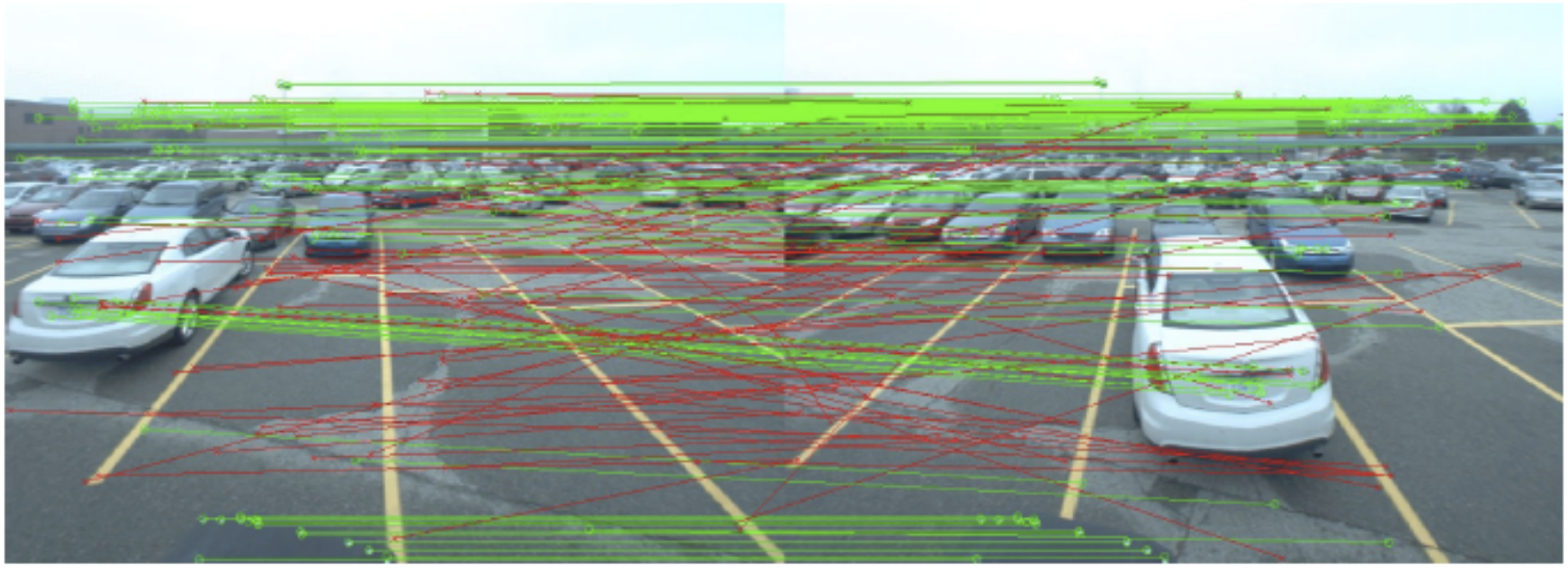
Camera rotation error		
SFM Snavely et al., 08	**SSFM** no int.	**SSFM**
<1°	<1°	**<1**
9.6°	4.2°	**3.5°**
21.1°	3.1°	**3.0°**



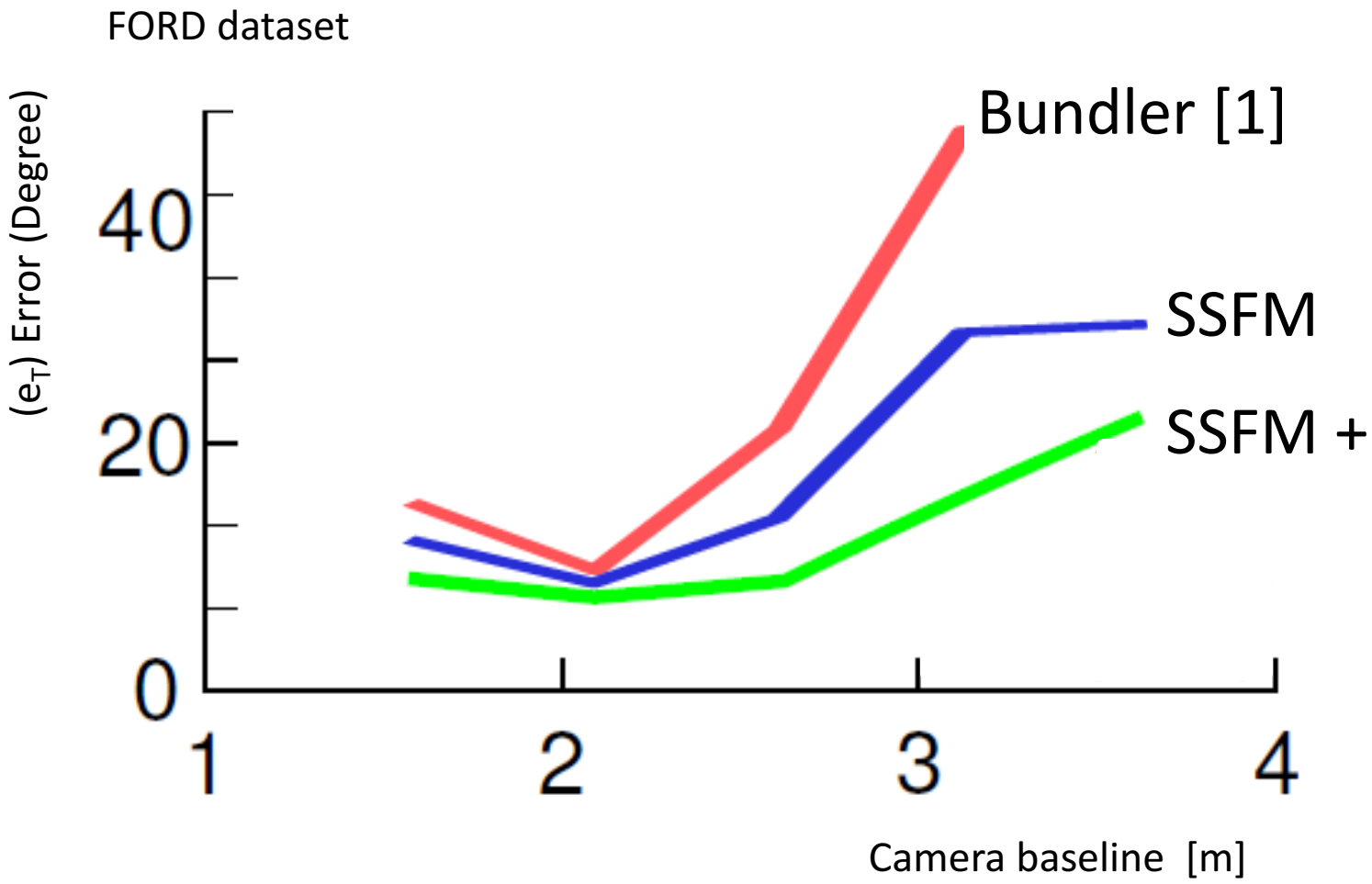
FORD CAMPUS dataset [Pandey et al., 09]
Office dataset [Bao et al., 11]
Street dataset [Bao et al., 11]

Wide-baseline feature correspondence





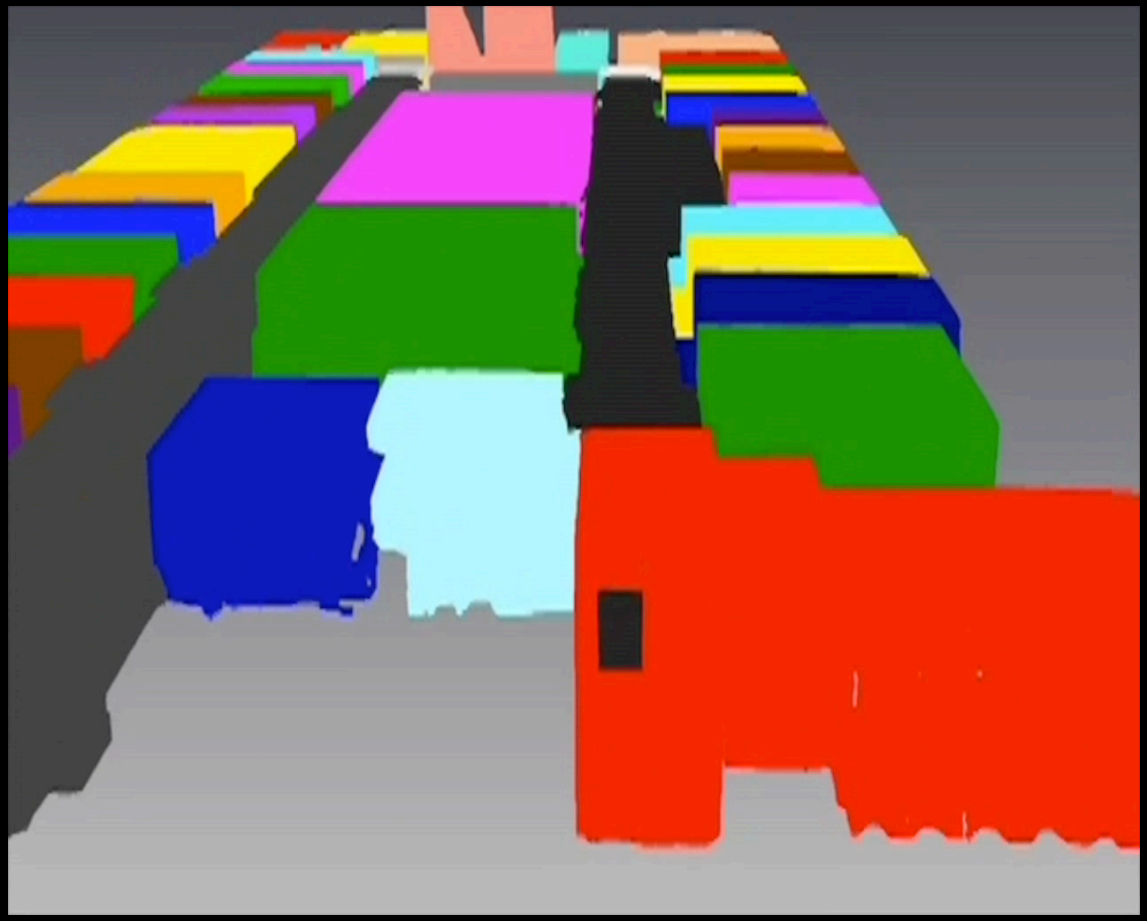
Camera Pose Estimation v.s. Base Line Width



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

Large-scale scene parsing

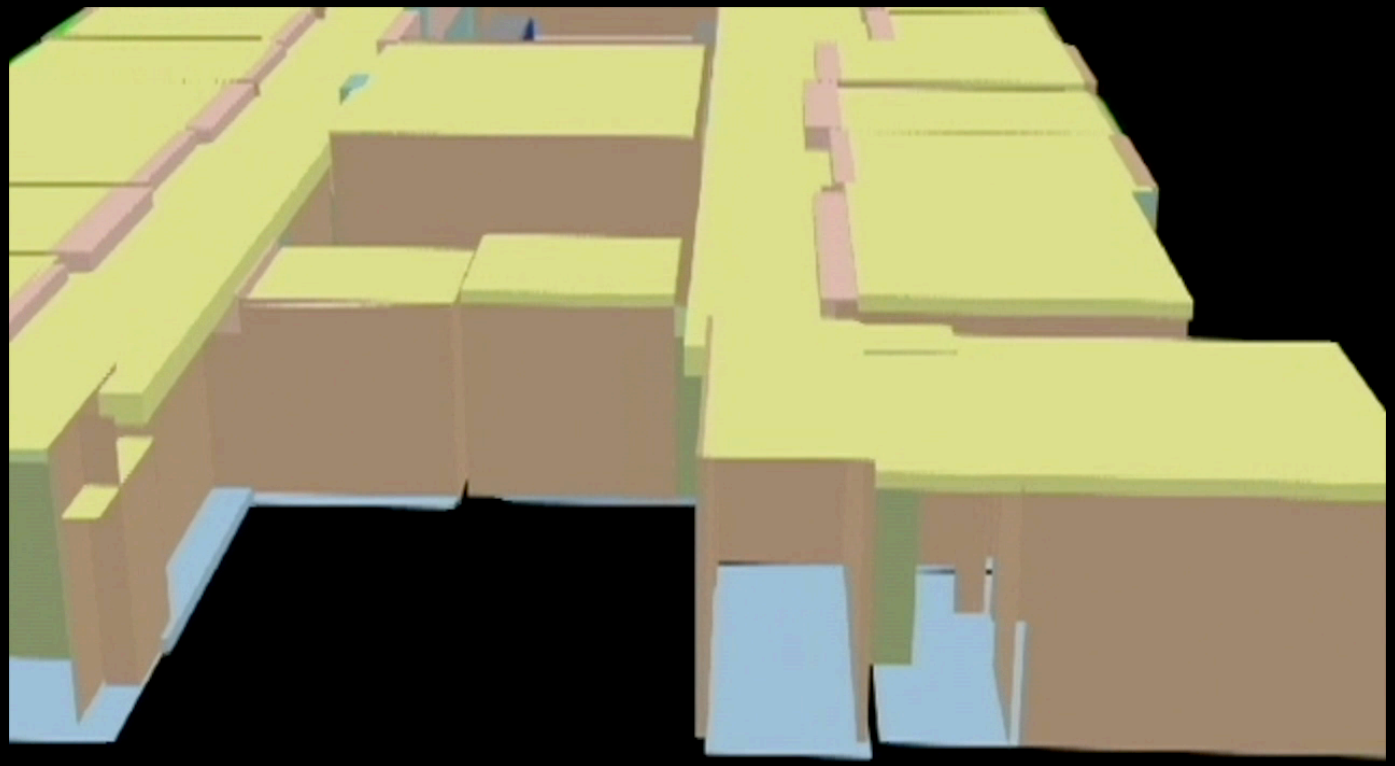
Armeni, Sener, Zamir, Jiang, Brilakis, Fischer, Savarese, 2016



ceiling floor wall column beam window door table chair bookcase sofa board

Large-scale scene parsing

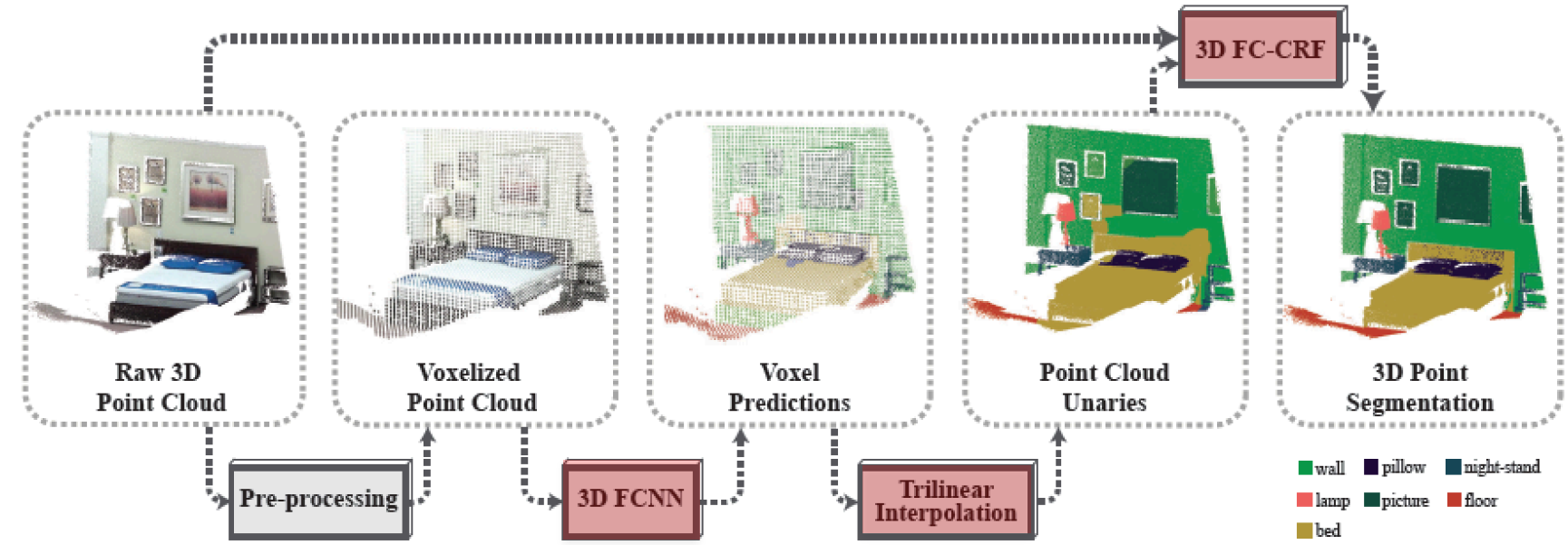
Armeni, Sener, Zamir, Jiang, Brilakis, Fischer, Savarese, 2016



ceiling floor wall column beam window door table chair bookcase sofa board

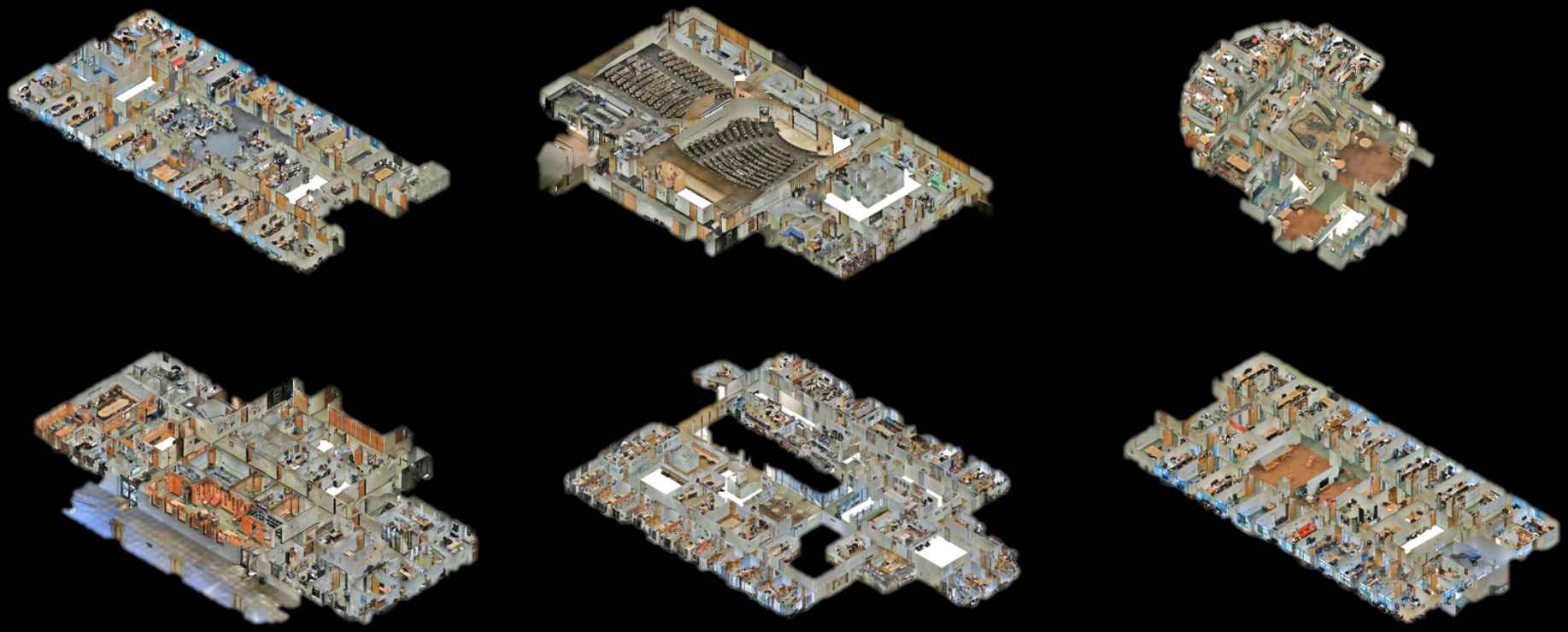
Toward fine-grained 3D scene parsing

Tchapmi et al., 3DV, 2017



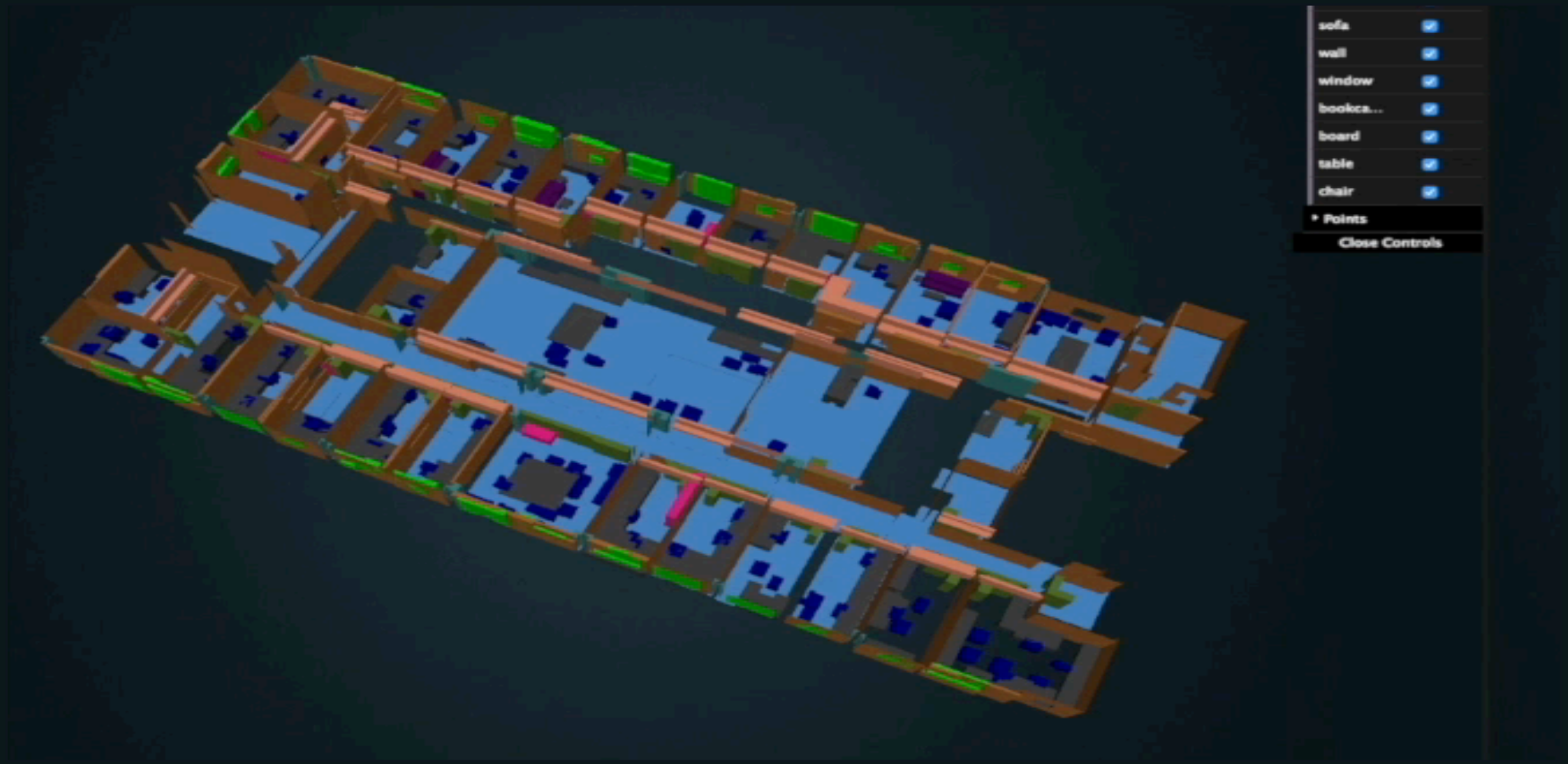
Stanford Large-Scale Indoor dataset

Armeni, Sener, Zamir, Jiang, Brilakis, Fischer, Savarese, 2016



6 buildings ~500 rooms ~6000m^2 area ~6000 Building Elements

Building information management

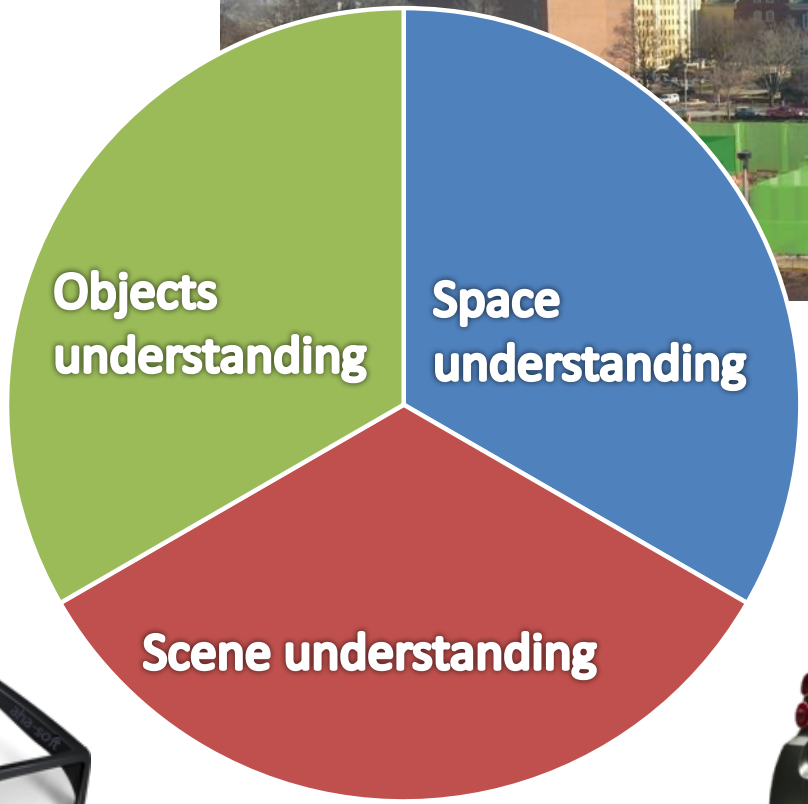


http://buildingparser.stanford.edu

Applications

Robotic navigation

Construction monitoring



Mobile vision

Safe driving

Visual intelligence and large scale information management

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

James R. Croes Medal, October 2013 (from the American Society of Civil of Engineers)

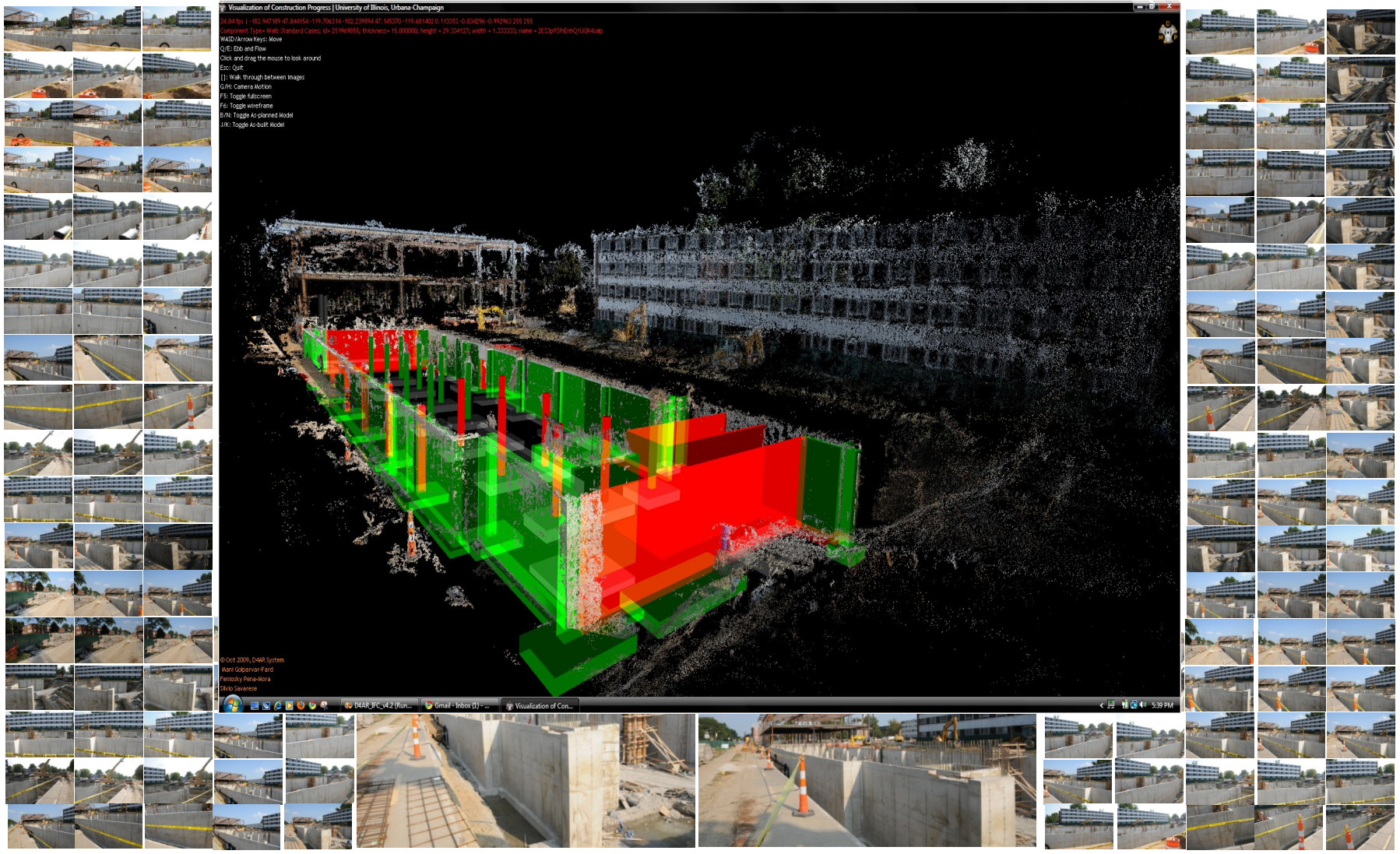
Automatic coordination of construction progress can lead to huge savings (10 billions USD/year) in construction business!

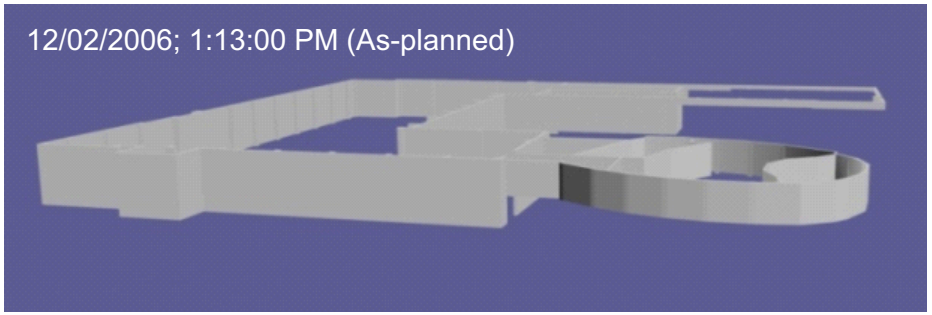
[Census Bureau, www.census.gov, 2007]

Images are cheap!

Our revolution

Our revolution

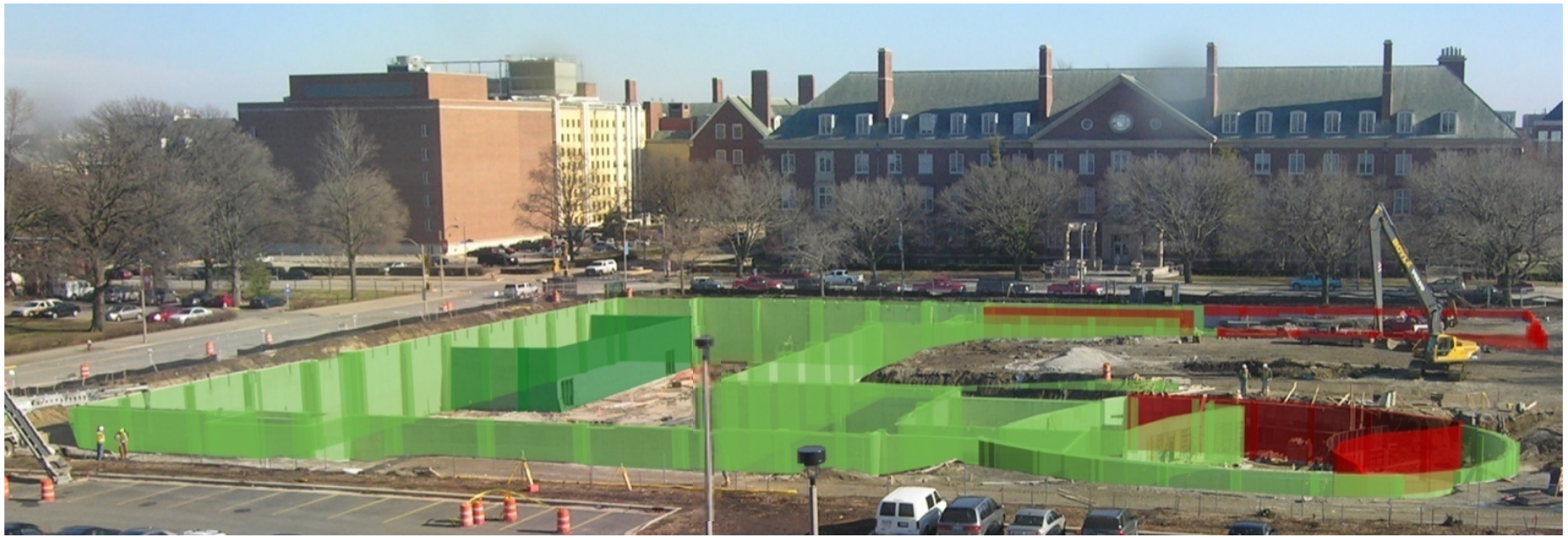


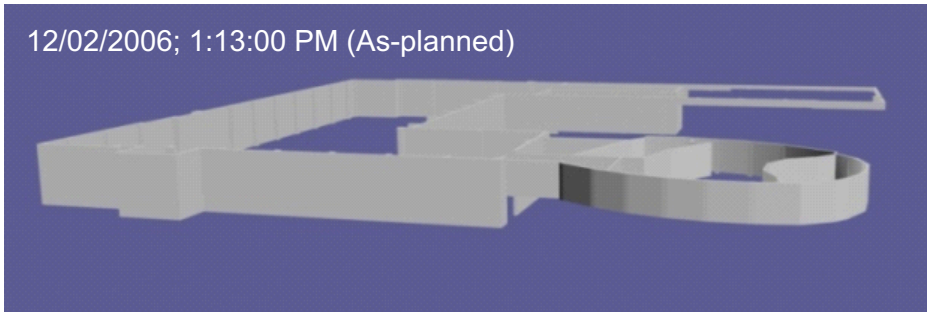


Ahead of Schedule

On Schedule

Behind Schedule



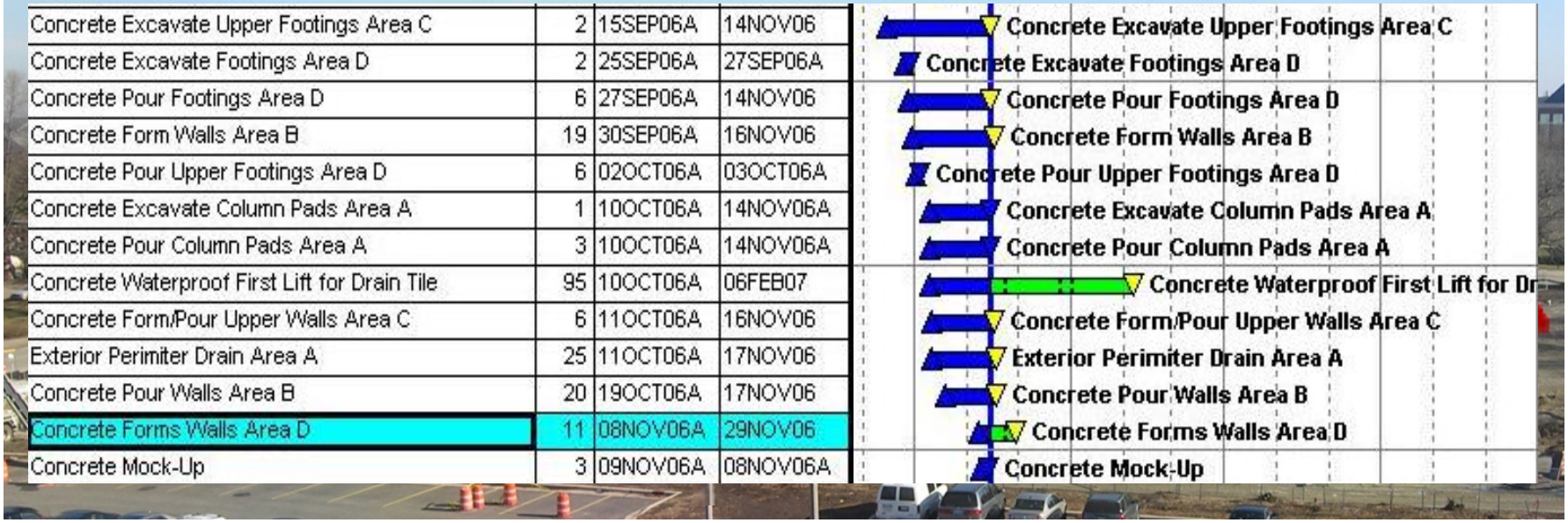


Ahead of Schedule

On Schedule

Behind Schedule

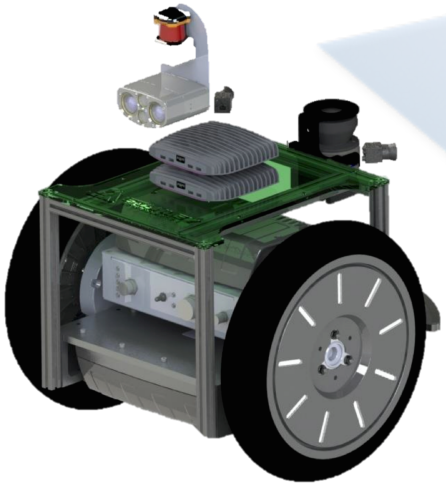
Concrete Excavate Upper Footings Area C	2	15SEP06A	14NOV06
Concrete Excavate Footings Area D	2	25SEP06A	27SEP06A
Concrete Pour Footings Area D	6	27SEP06A	14NOV06
Concrete Form Walls Area B	19	30SEP06A	16NOV06
Concrete Pour Upper Footings Area D	6	02OCT06A	03OCT06A
Concrete Excavate Column Pads Area A	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

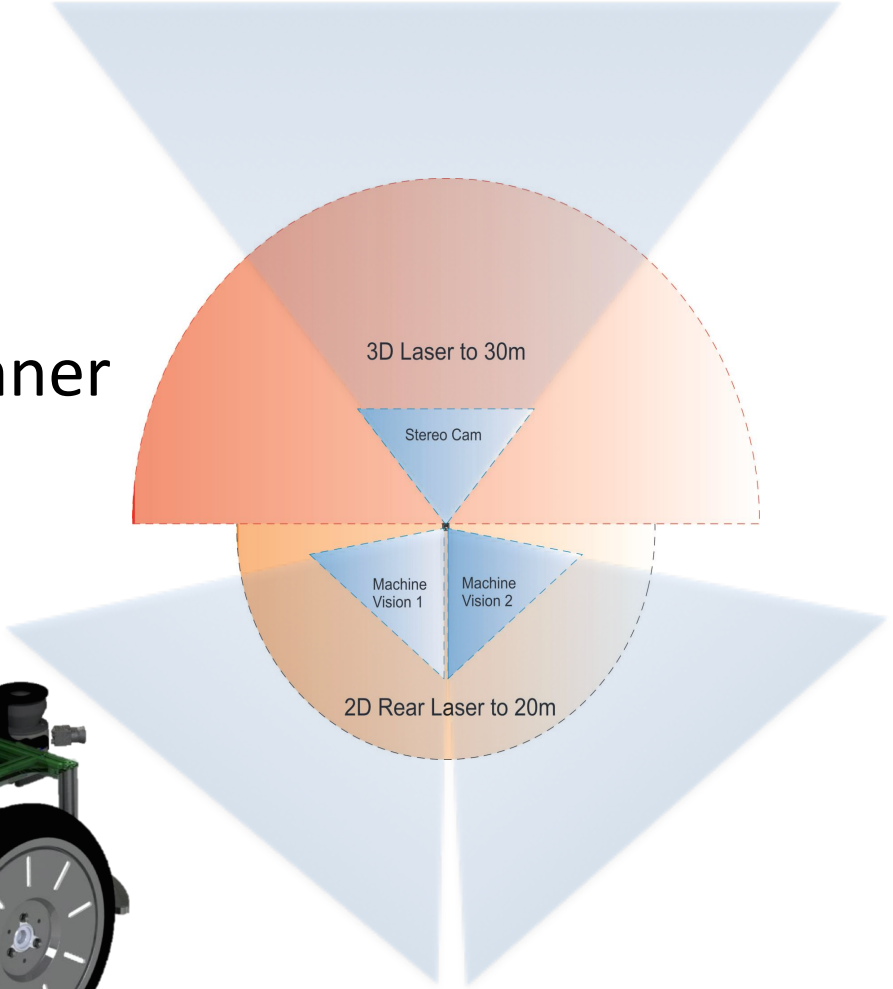


A new experimental platform: The JackRabbot

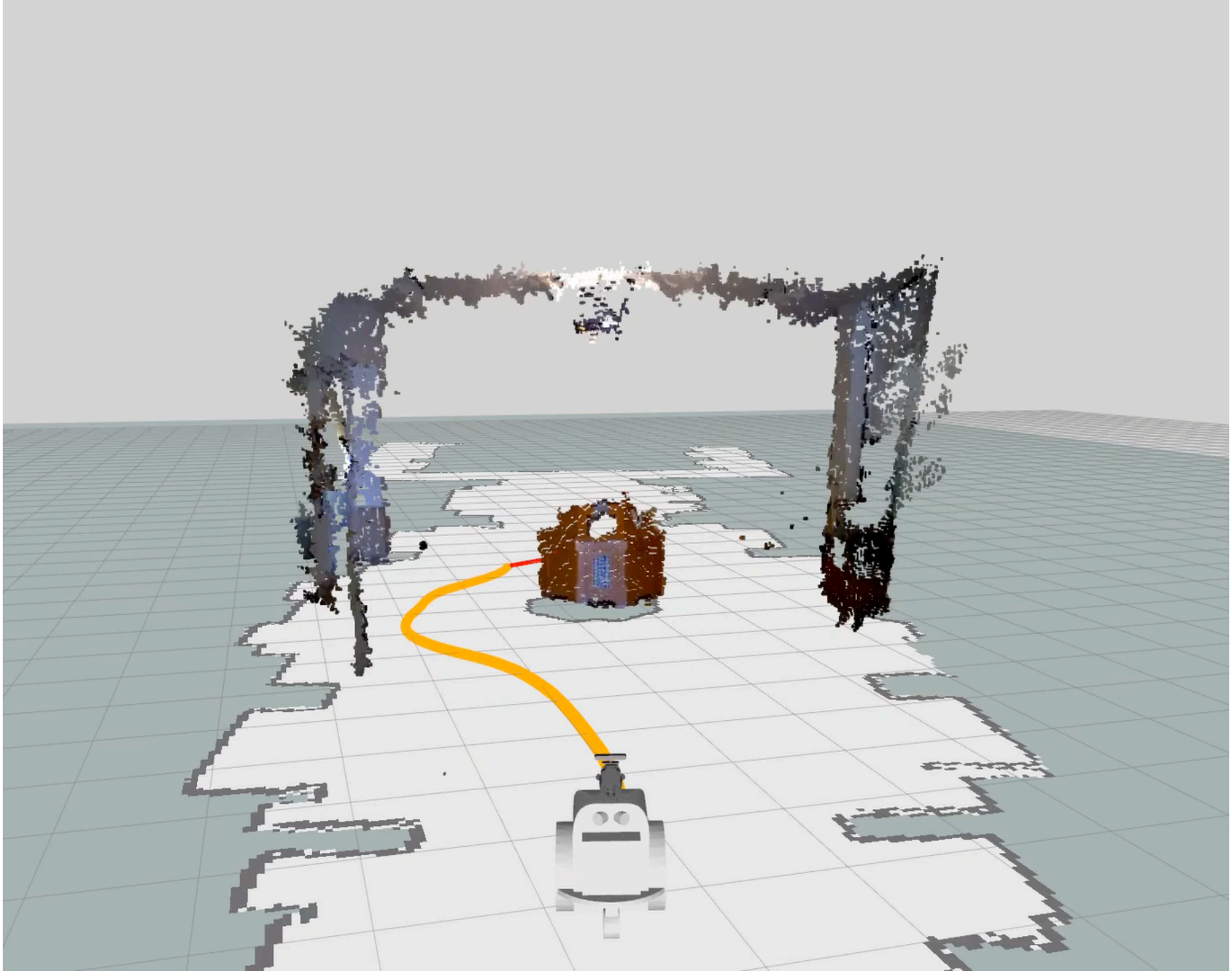
A new experimental platform: The JackRabbot

- Sensors
  - Planar laser scanner
  - 3D stereo vision/laser scanner
  - Cameras
  - IMU+GPS





A new experimental platform: The JackRabbot



Hope you enjoyed this course

Good luck on your presentations next week!