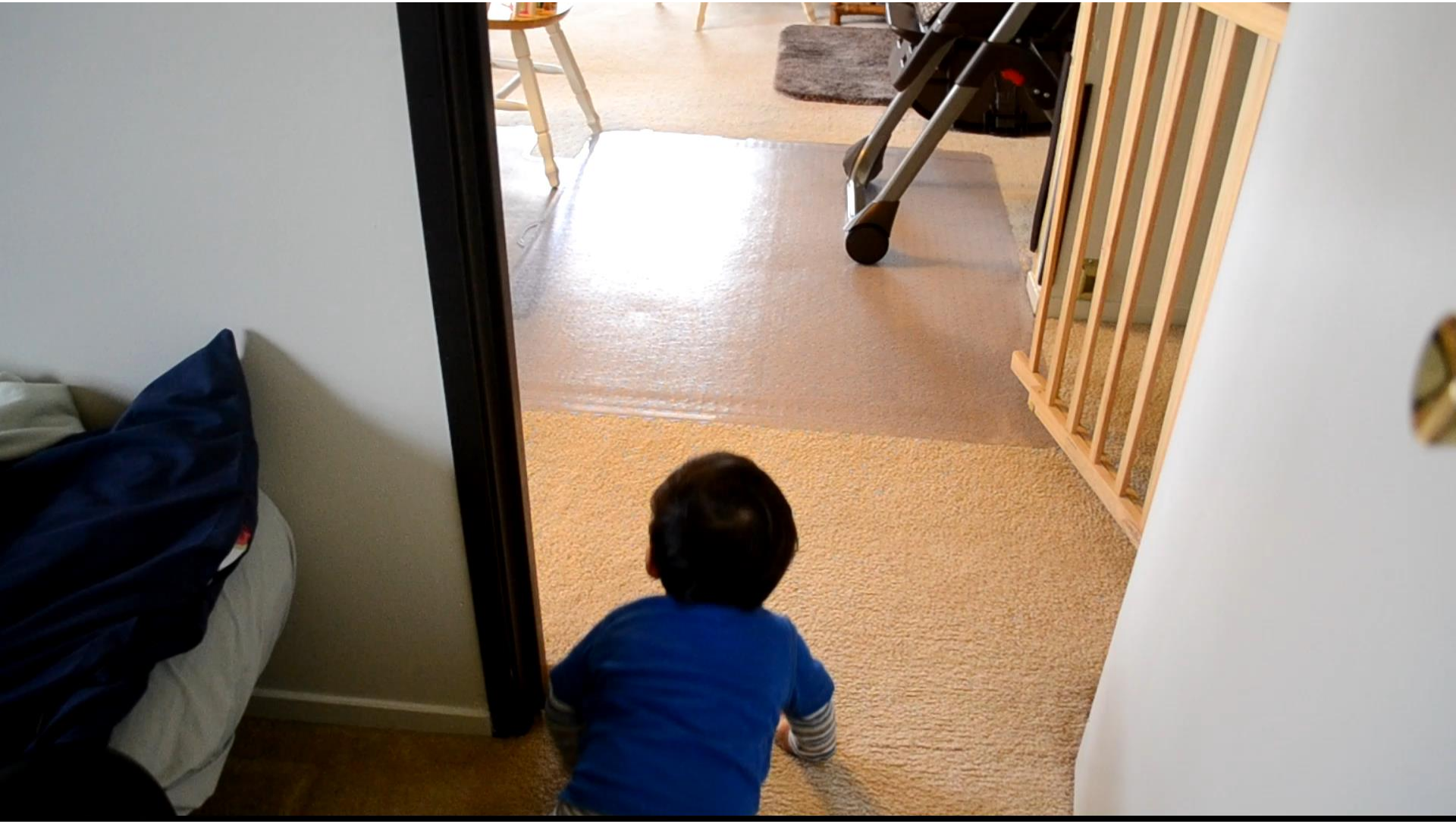
Perceiving the 3D World from Images and Videos

Silvio Savarese

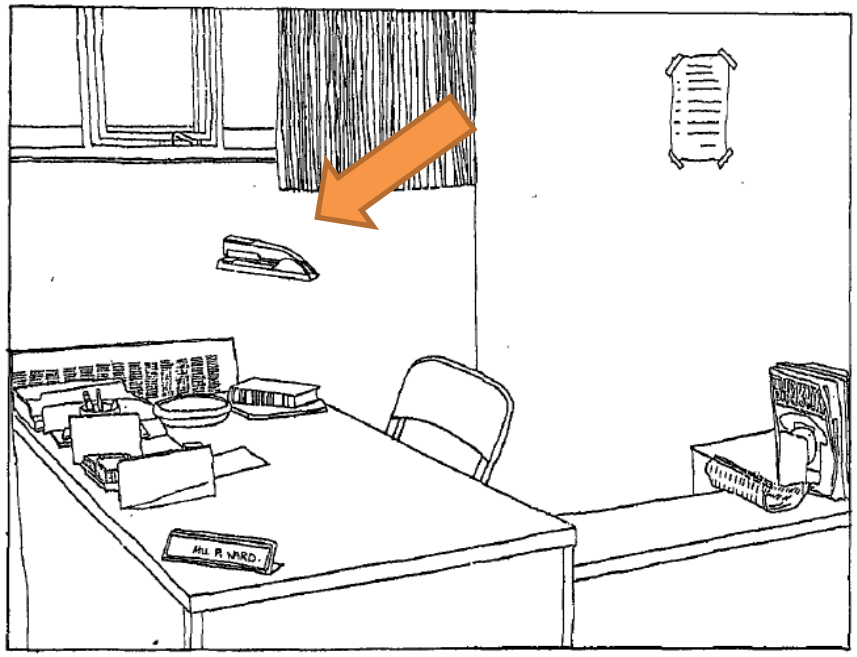
Stanford University
Computer Science Department

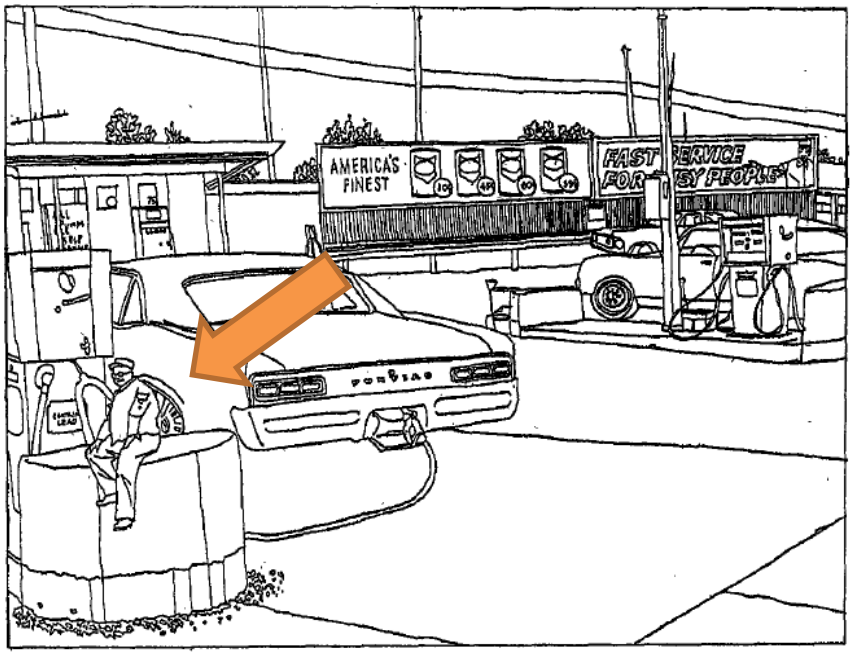
April 15th, 2014



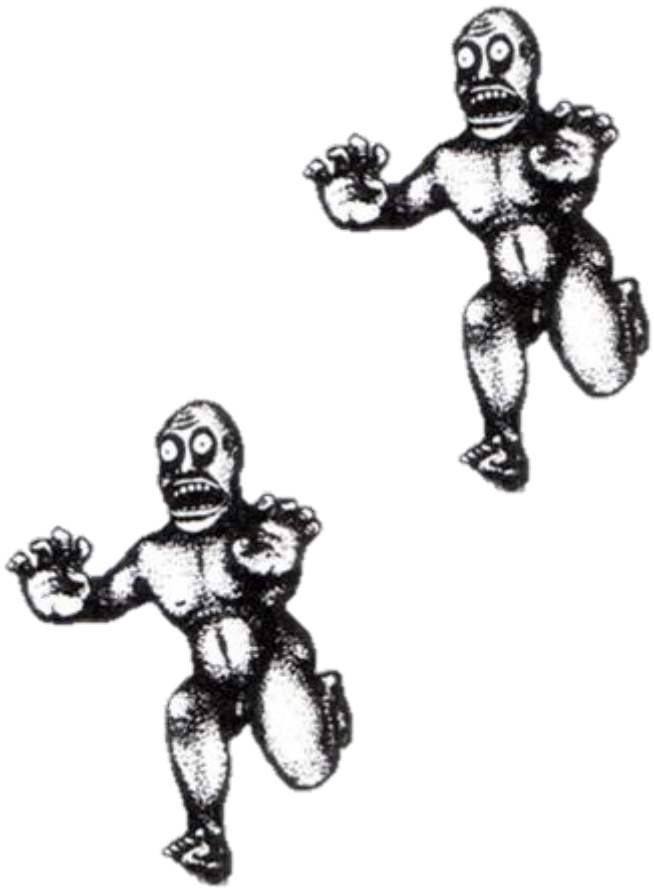
My is goal to build intelligent visual machines that can perceive and understand the 3D world

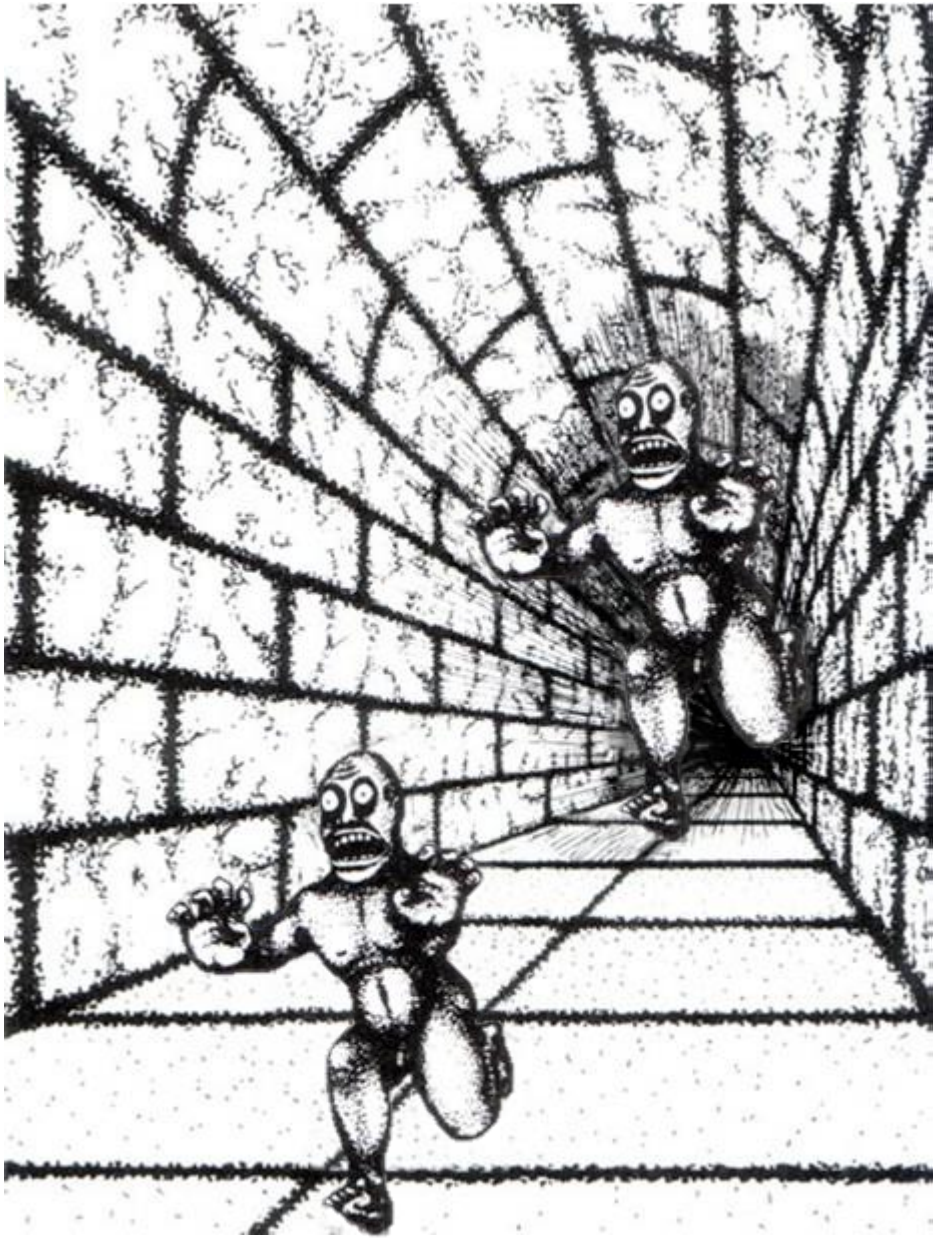
Humans perceive the world in 3D





Biederman, Mezzanotte and Rabinowitz, 1982



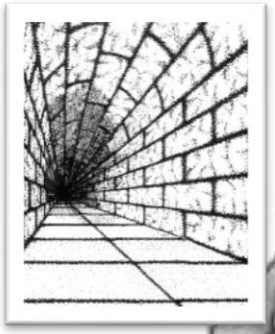


Objects are constrained by the 3D space

The 3D space is shaped by its objects

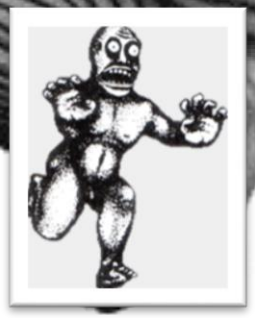
Modeling this interplay is critical for 3D perception!

Visual processing in the brain



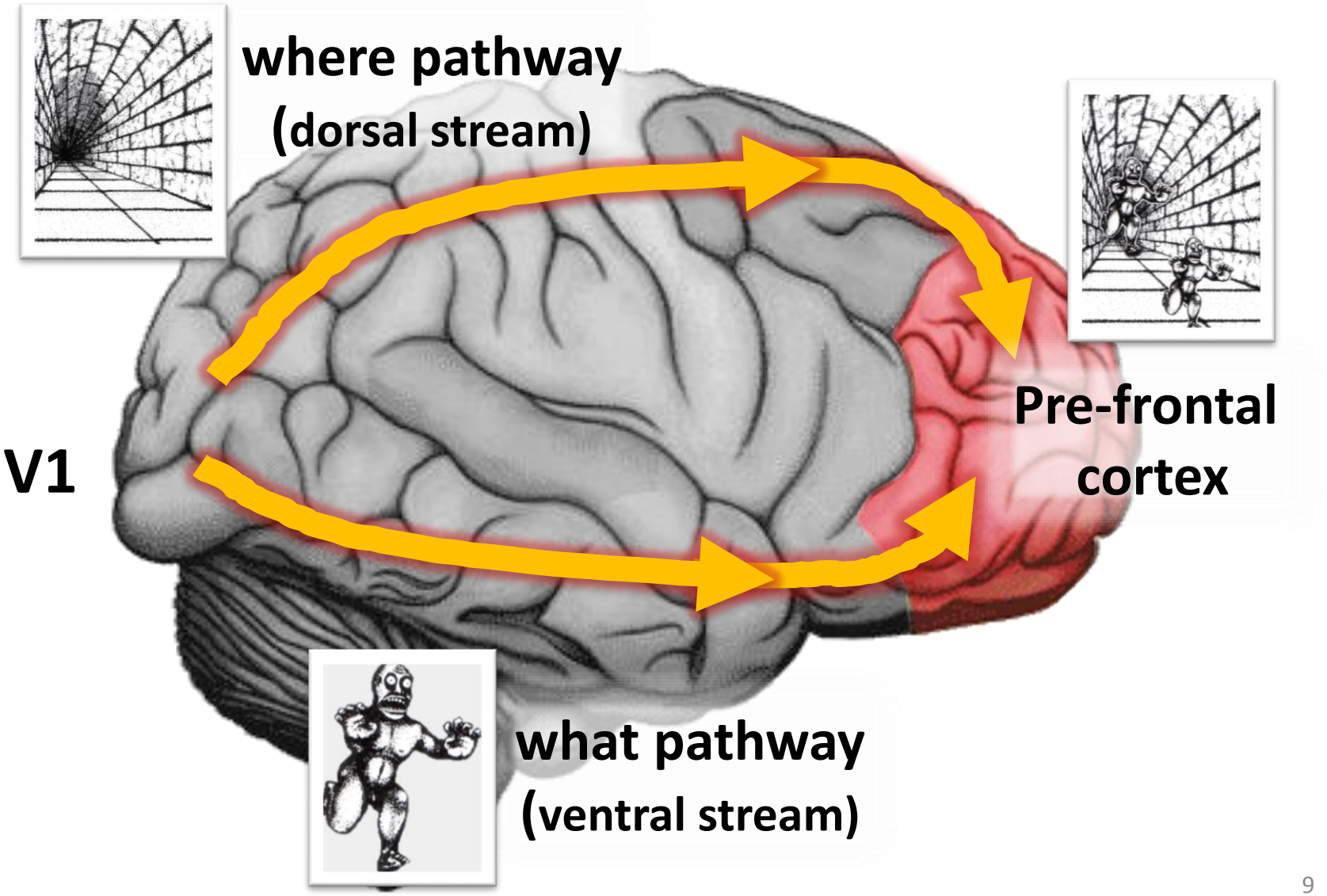
where pathway
(dorsal stream)

V1

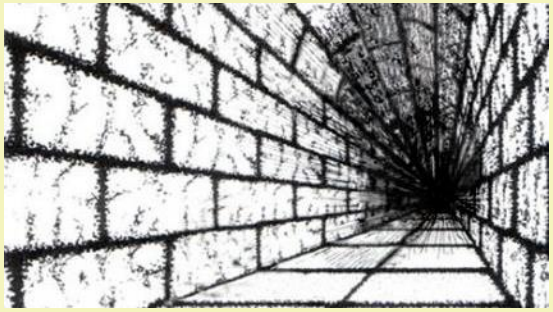


what pathway
(ventral stream)

Visual processing in the brain



Current state of computer vision



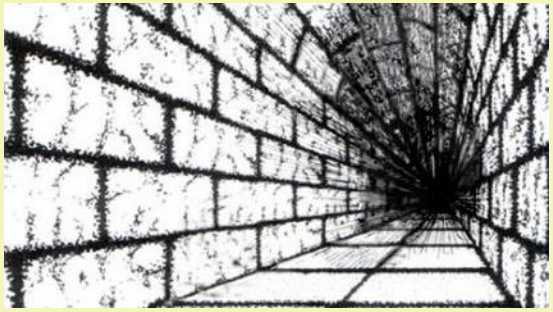
3D Reconstruction

- 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

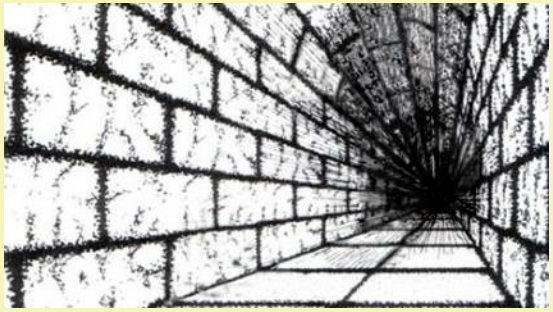
Snavely et al., 06-08

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

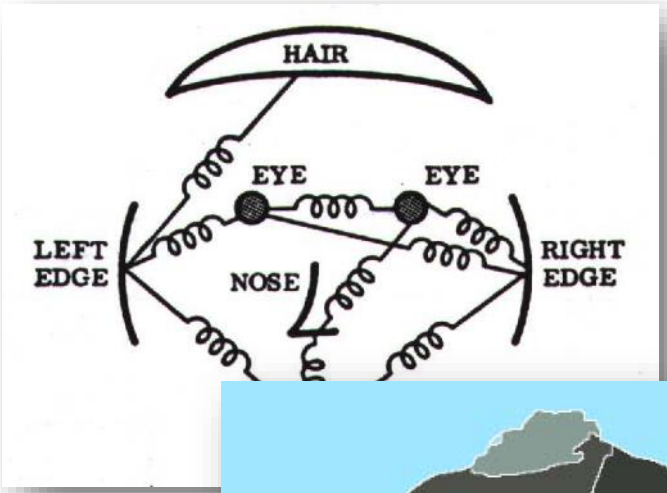
Snavely et al., 06-08

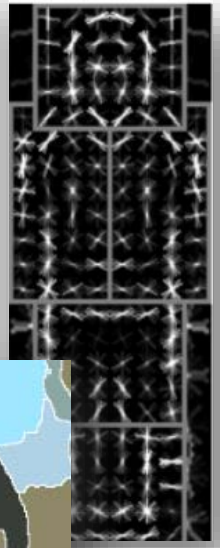
Lucas & Kanade, 81
Chen & Medioni, 92
Debevec et al., 96
Levoy & Hanrahan, 96
Fitzgibbon & Zisserman, 98
Triggs et al., 99
Pollefeys et al., 99
Kutulakos & Seitz, 99

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

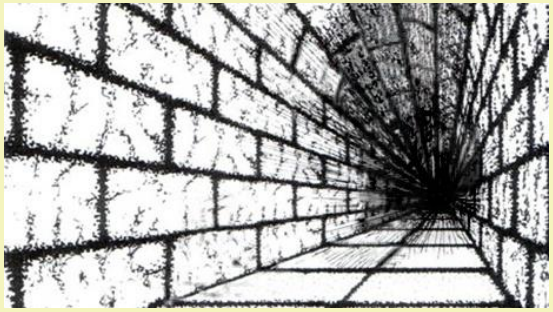
- Object detection
- Texture classification
- Target tracking
- Activity recognition

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

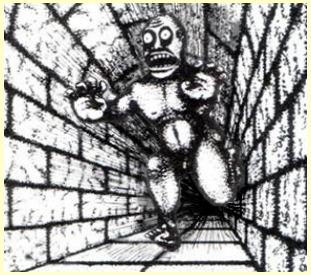
He et al. 06
Gould et al. 08
Maire et al. 08
Felzenszwalb et al., 08
Kohli et al. 09
L.-J. Li et al. 09
Ladicky et al. 10,11
Gonfaus et al. 10
Farhadi et al., 09
Lampert et al., 09

Current state of computer vision



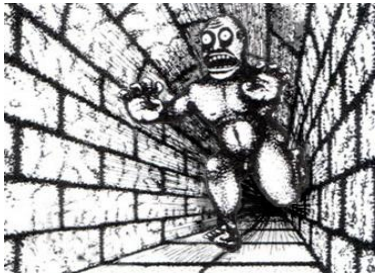
3D reconstruction

2D recognition



Perceiving the World in 3D

- 3D Object detection
- 3D Scene Understanding



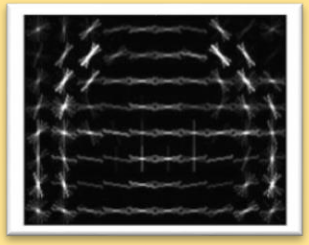
Outline

- 3D Object detection
- 3D Scene Understanding

Modeling objects and their 3D properties

State of the art object detection

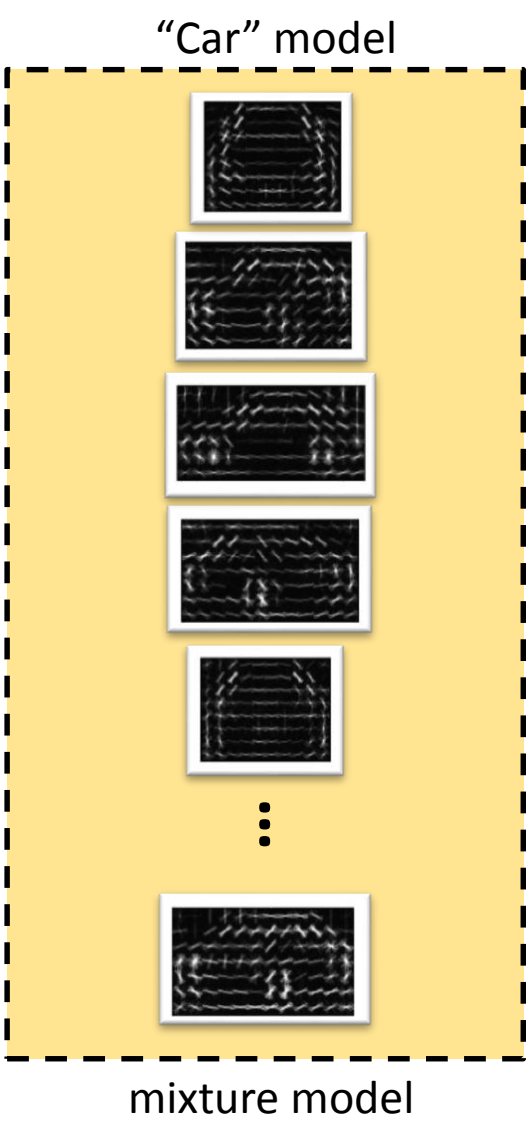
"Car" model

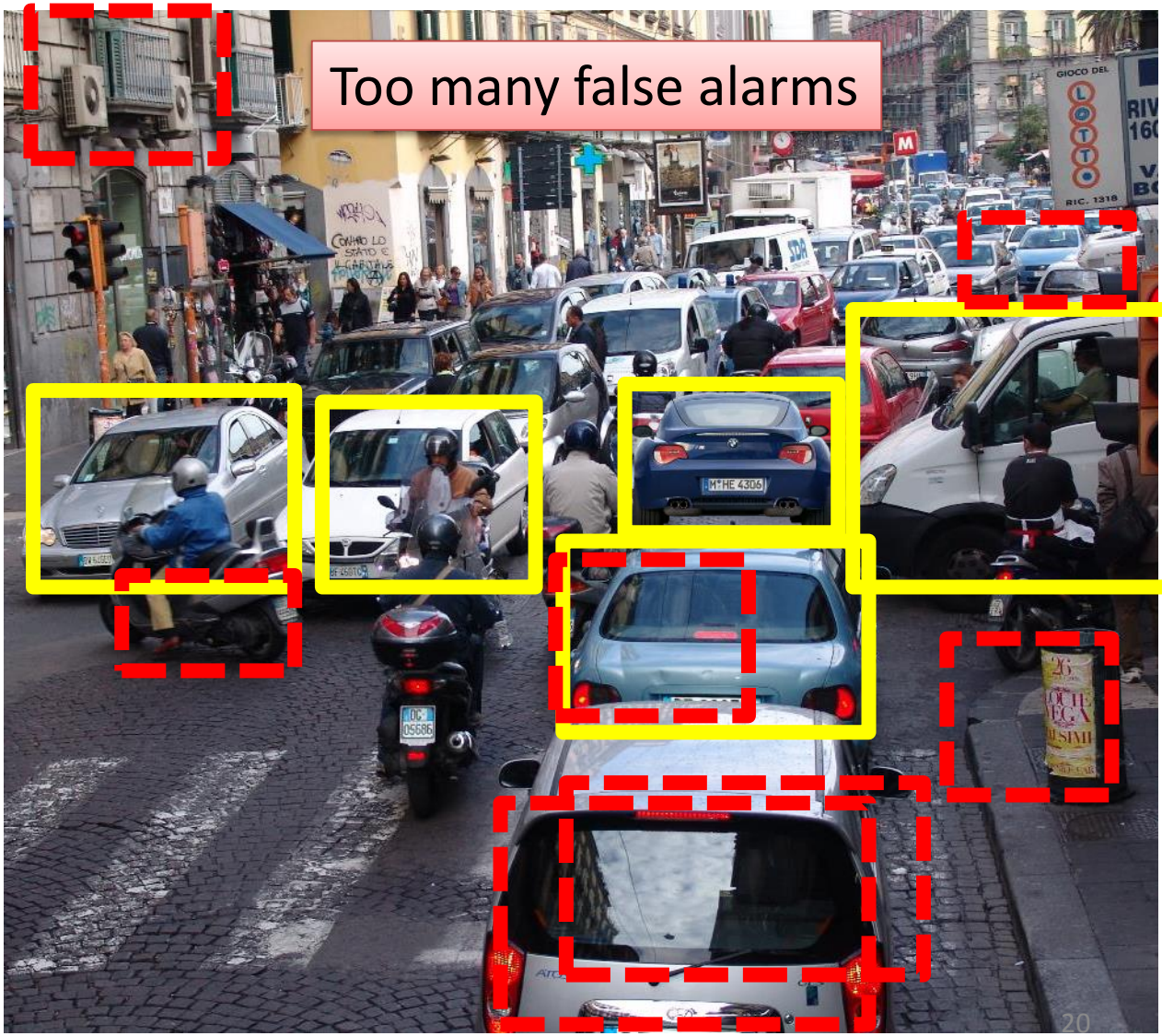


Too many missed detections!
Some false alarms

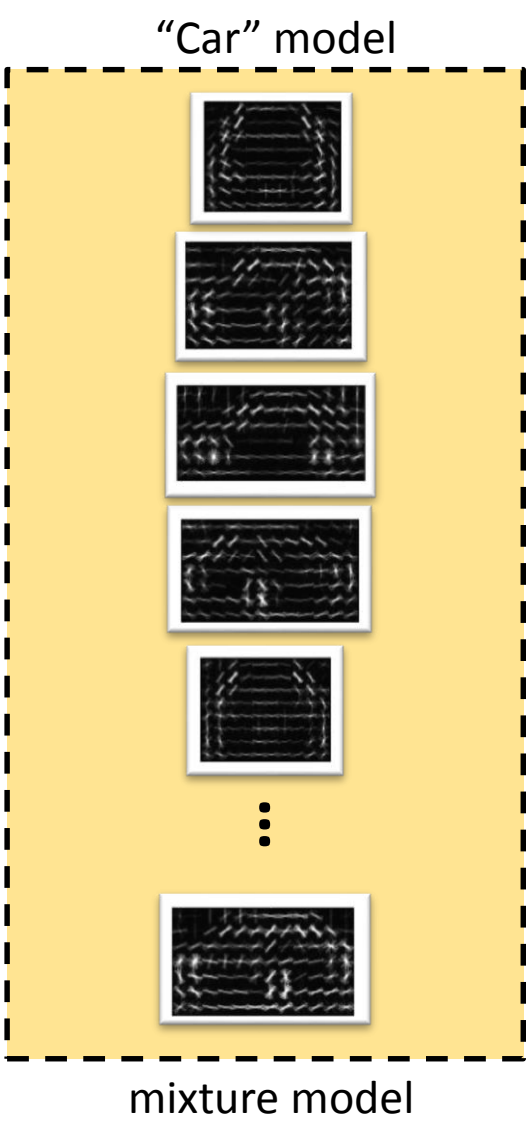
Turk & Pentland, 91
Poggio et al., 93
LeCun et al. 98
Amit and Geman, 99
Shi & Malik, 00
Viola & Jones, 00
Vasconcelos '00
Felzenszwalb & Huttenlocher 00
Belongie & Malik, 02
Ullman et al. 02
Argawal & Roth, 02
Weber et al., 00
Fergus et al., 03
Torralba et al., 03
Fei-Fei et al., 04
Leibe et al., 04
Dalal & Triggs, 05
Savarese et al., CVPR 06
Felzenszwalb et al., 08
Lampert et al., 09

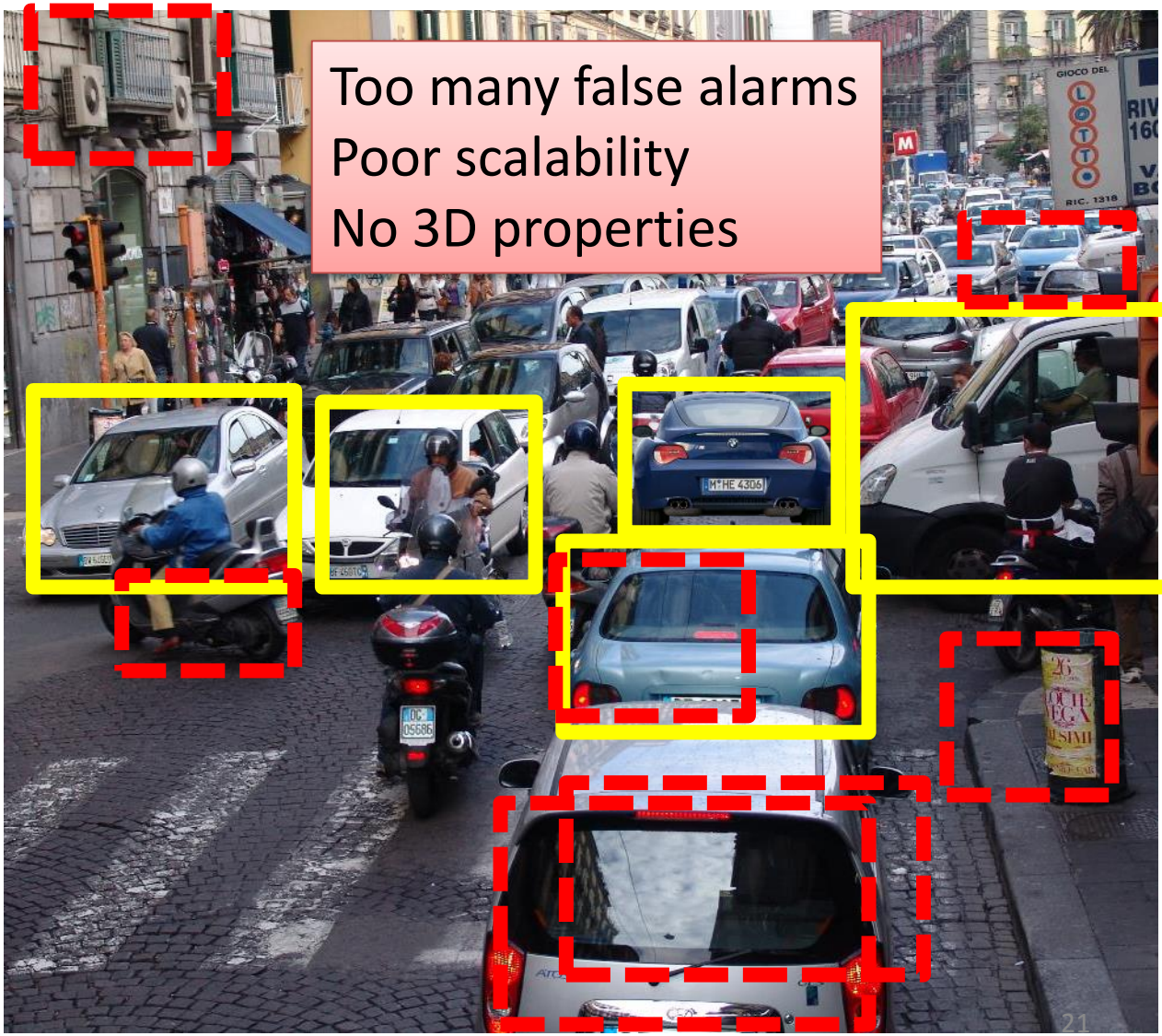
State of the art object detection





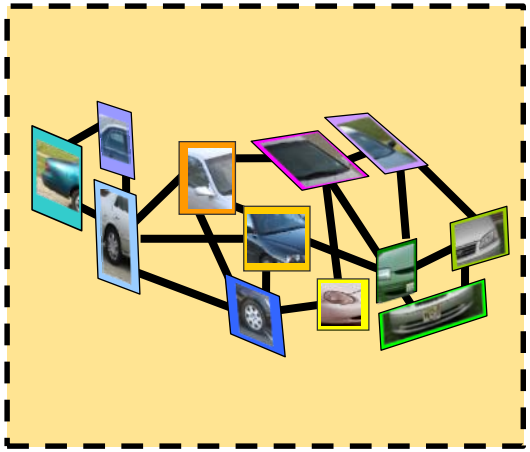
State of the art object detection





3D object detection

"Car" model

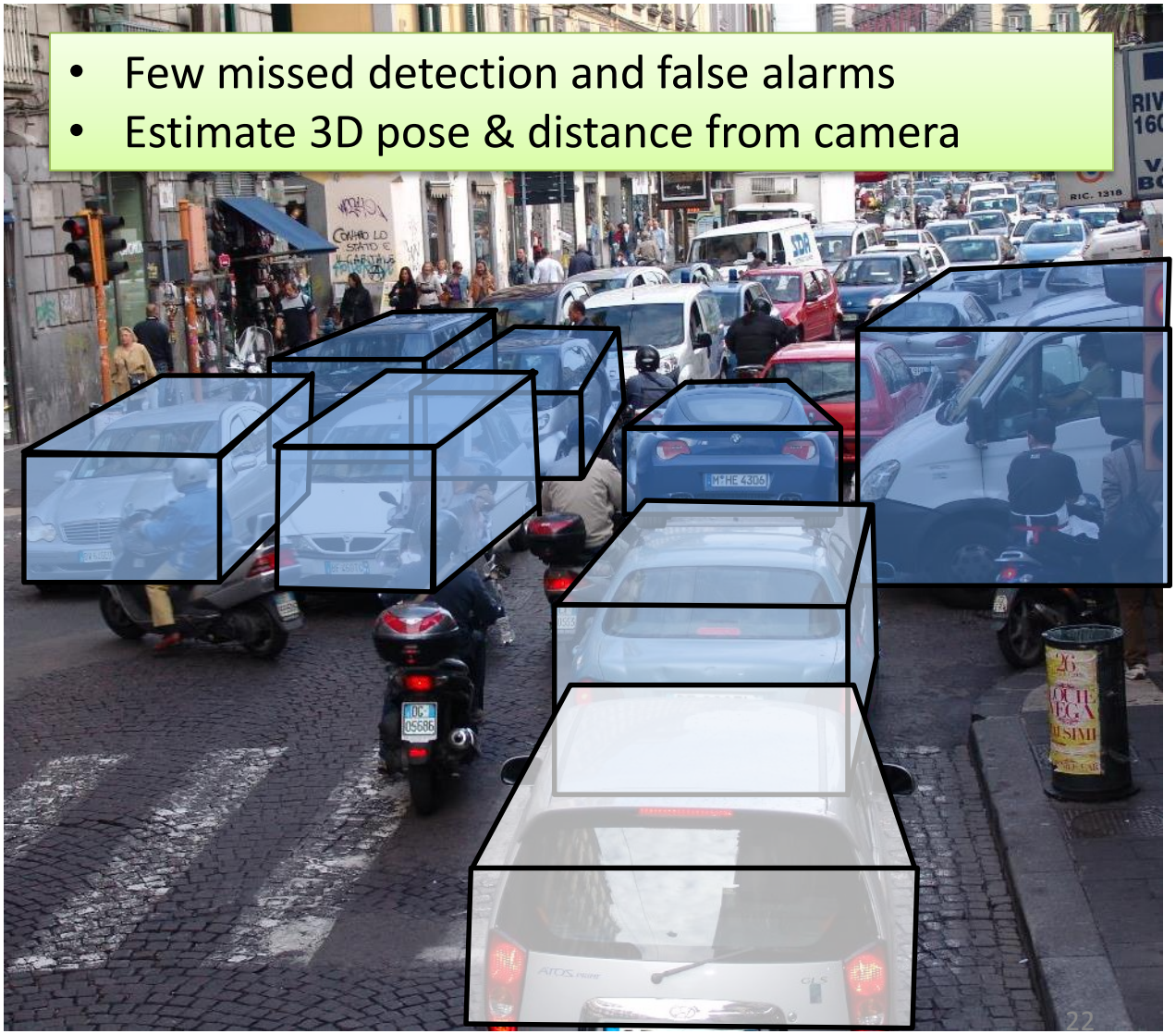


Savarese et al., ICCV 07
Su et al., ICCV 2009
Sun, et al., CVPR 2009
Yu & Savarese, CVPR 2012

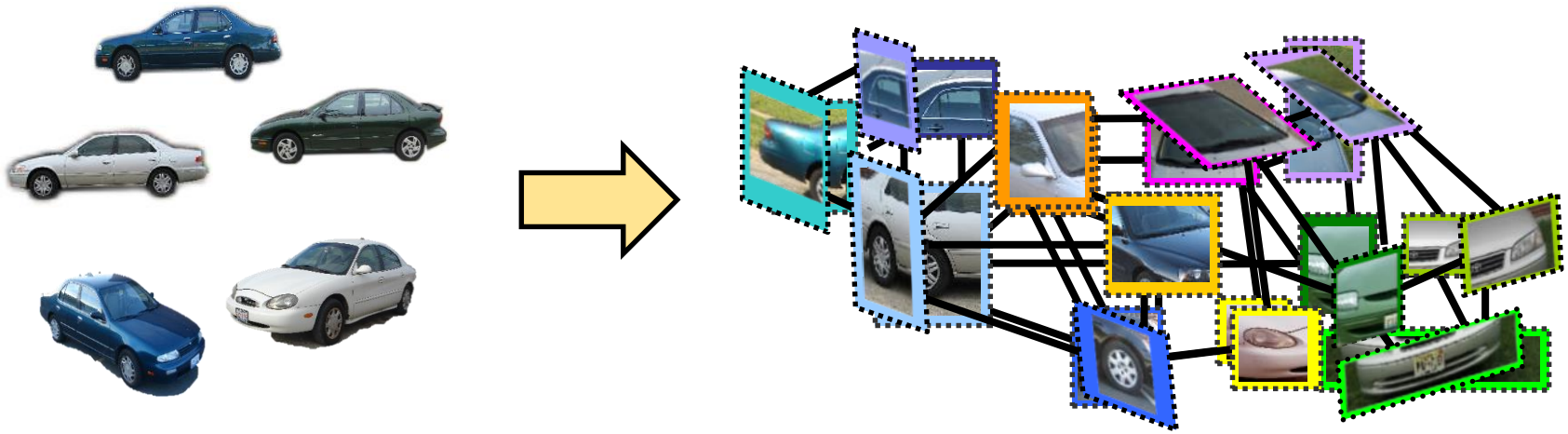
- Thomas et al. '06-09
- Yan et al., '07
- Kushal et al., '07
- Hoiem et al., 07
- Chiu et al '07
- Liebelt et al 08, 10
- Xiao et al 08
- Arie-Nachimson & Barsi '09
- Sandhu et al '09

⋮

- Farhadi '09
- Zhu et al. '09
- Ozuysal et al. '10
- Stark et al.'10
- Payet & Todorovic, 11
- Glasner et al., '11
- Zia et al. 11
- Pepik et al. '12



3D object representation

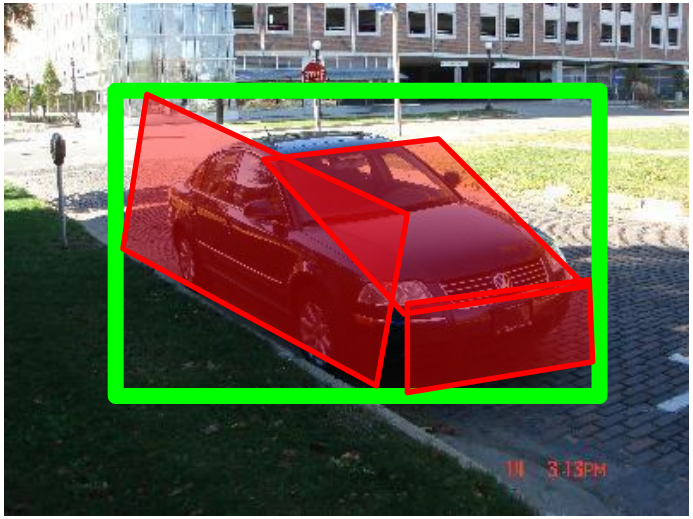


- Object is represented by a collection of parts
- Parts relationship are learnt from training images
- Inference by a novel algorithm based on variational EM
- Part configuration is modeled as a 3D conditional random fields

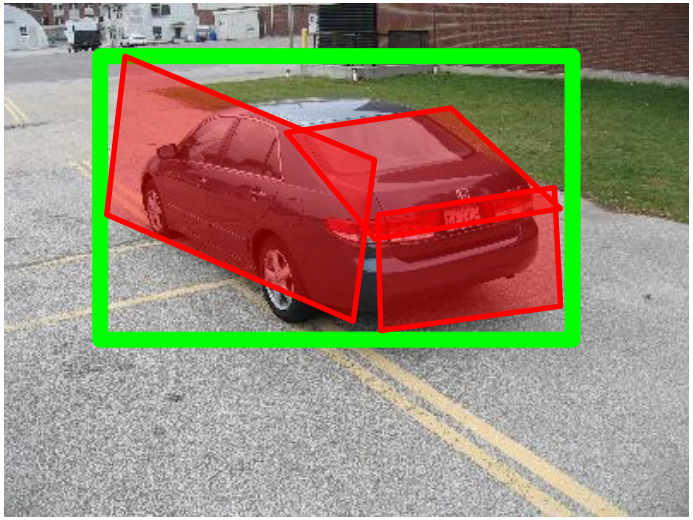
Savarese et al., ICCV 2007
Su et al., ICCV 2009
Sun, et al., CVPR 2009
Yu & Savarese, CVPR 2012

Results

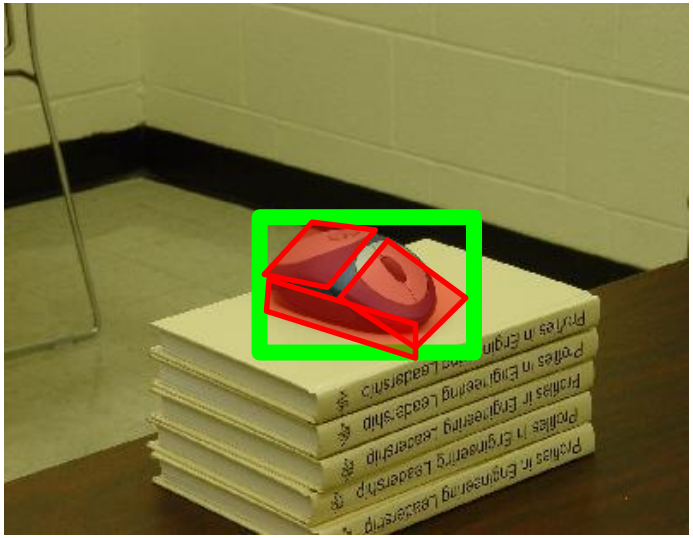
CAR a=330 e=15 d=7



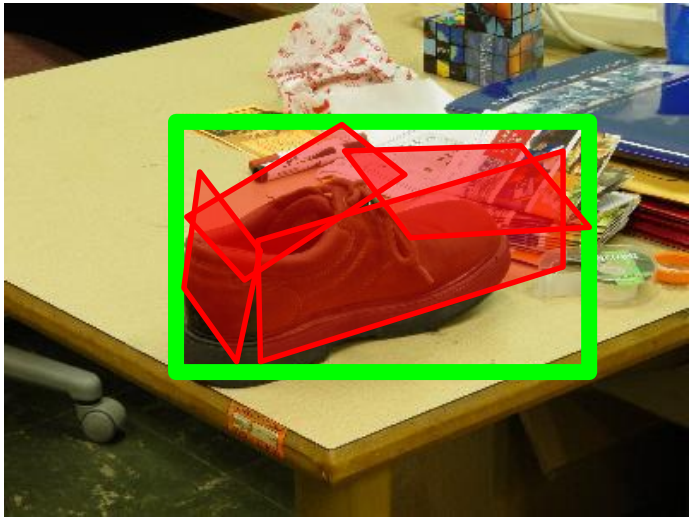
CAR a=150 e=15 d=7



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

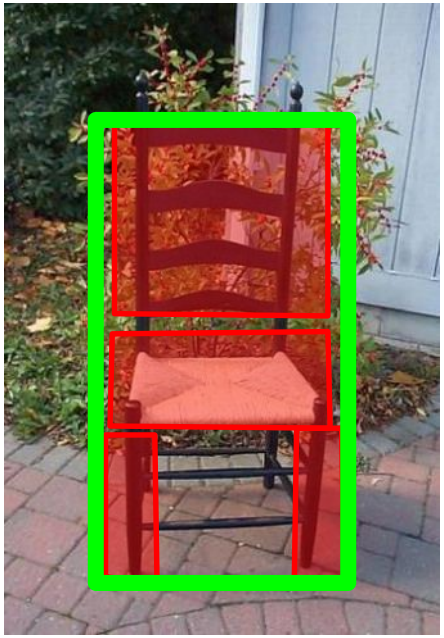
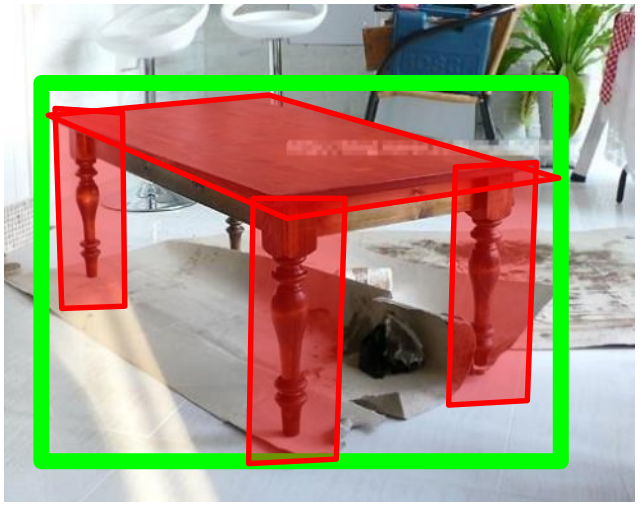
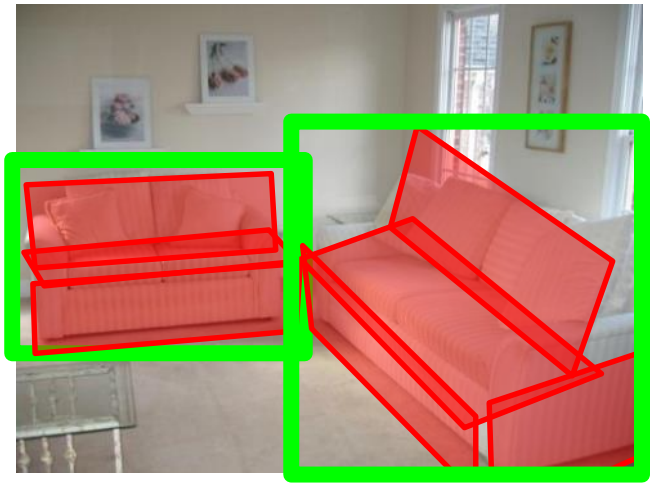


TABLE a=60 e=15 d=2



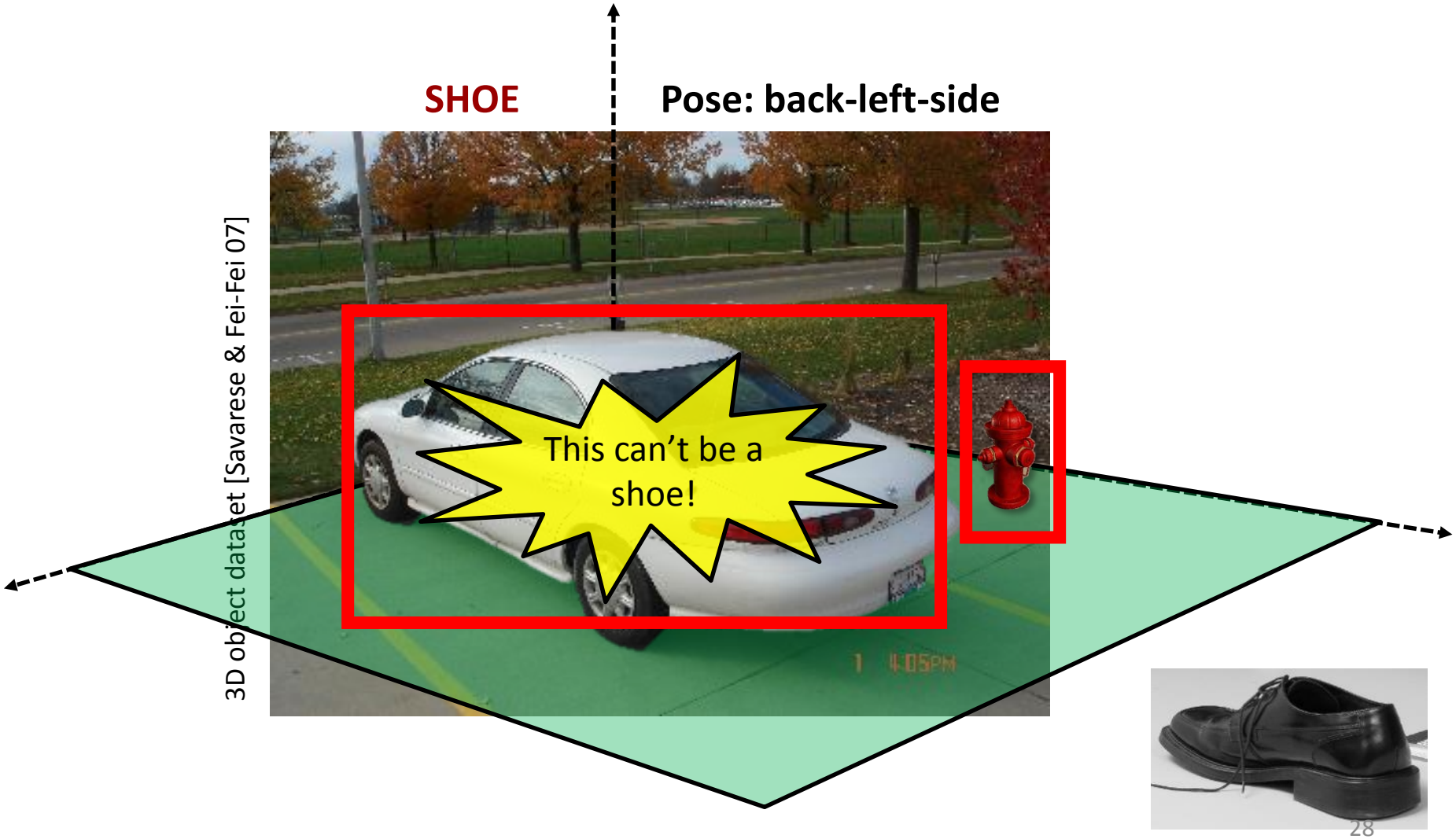
BED a=30 e=15 d=2.5

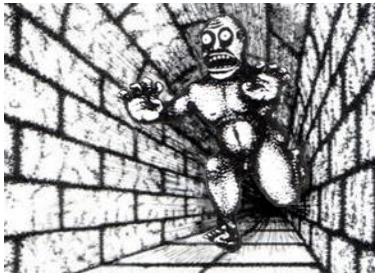
SOFA a=345 e=15 d=3.5
a=60 e-30 d=2.5



ImageNet dataset [Deng et al. 2010]

Results
Examples of failure (wrong category)

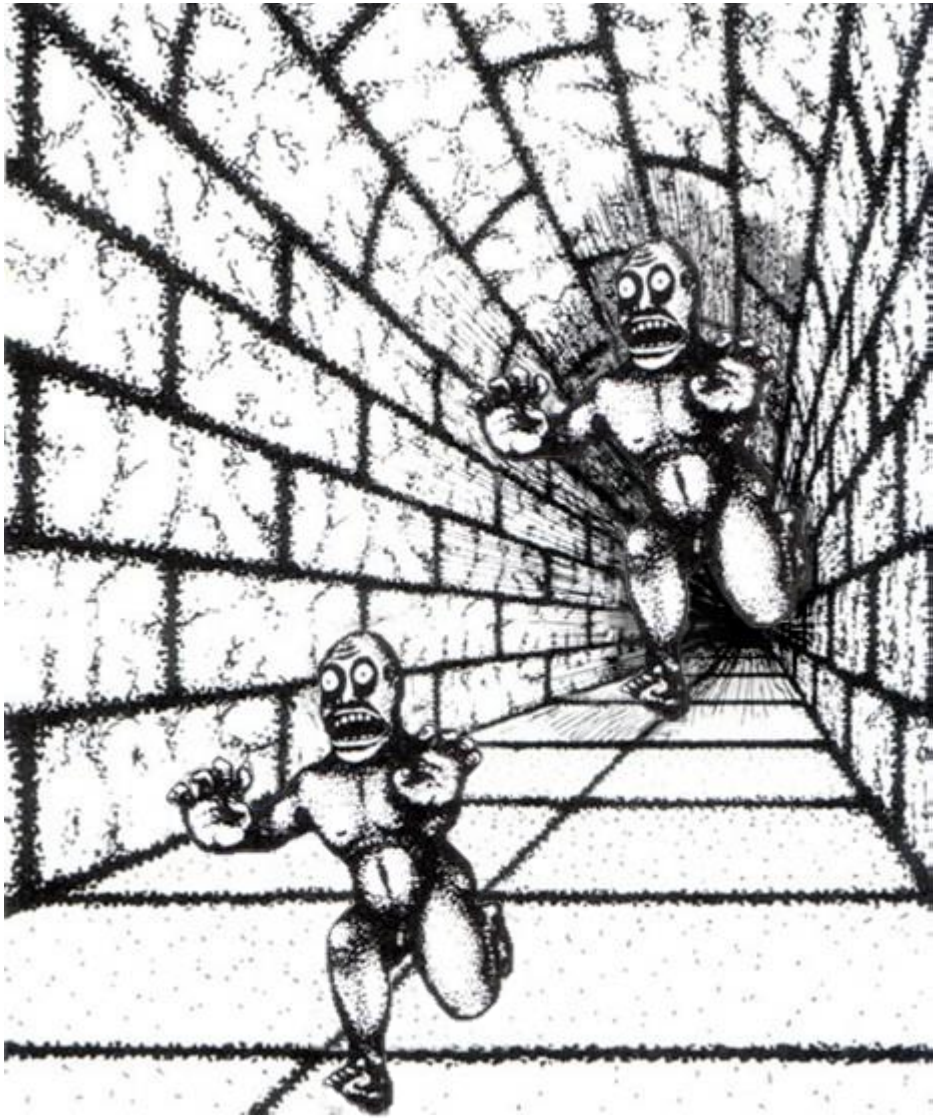




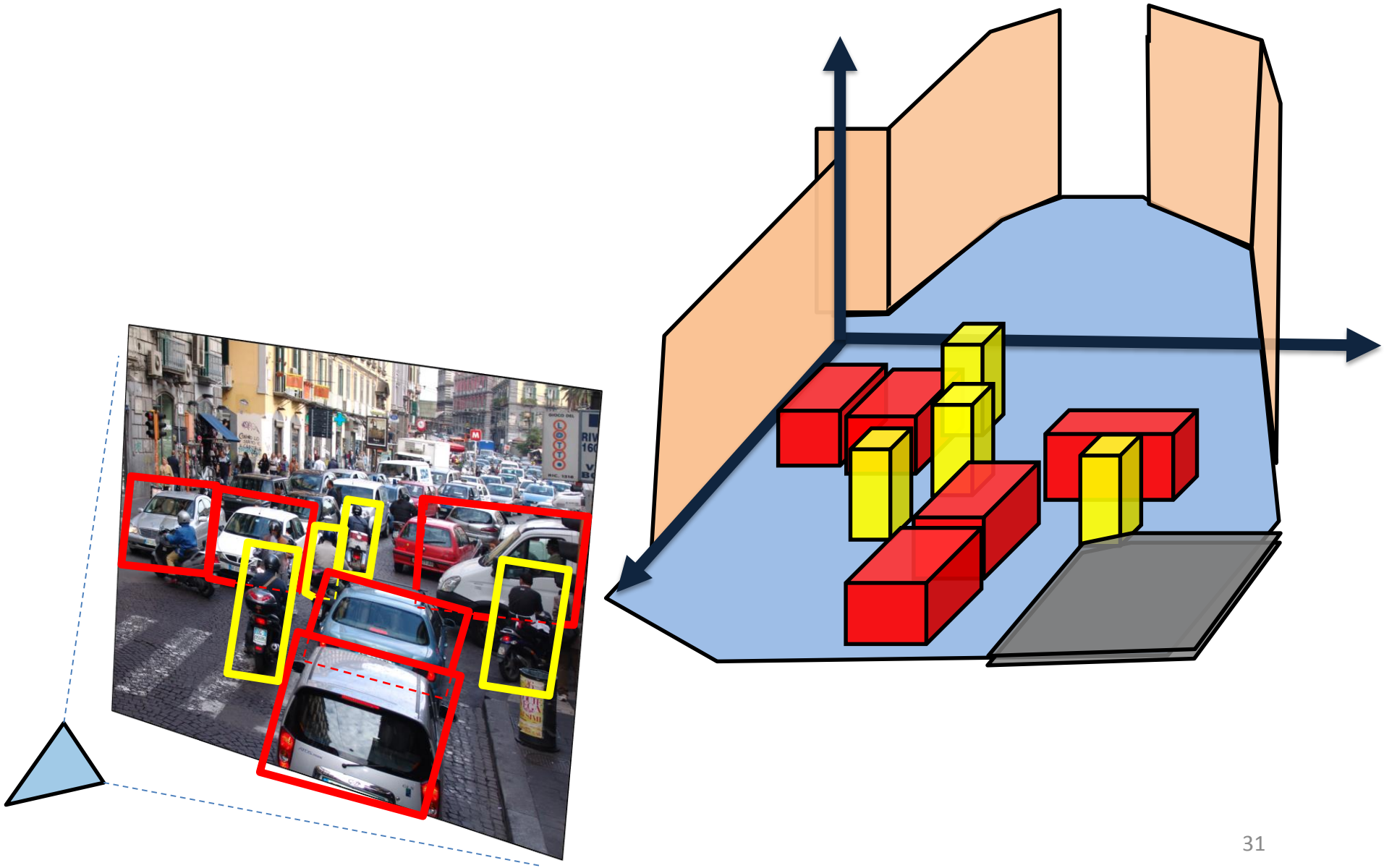
Outline

- 3D Object detection
- 3D Scene Understanding

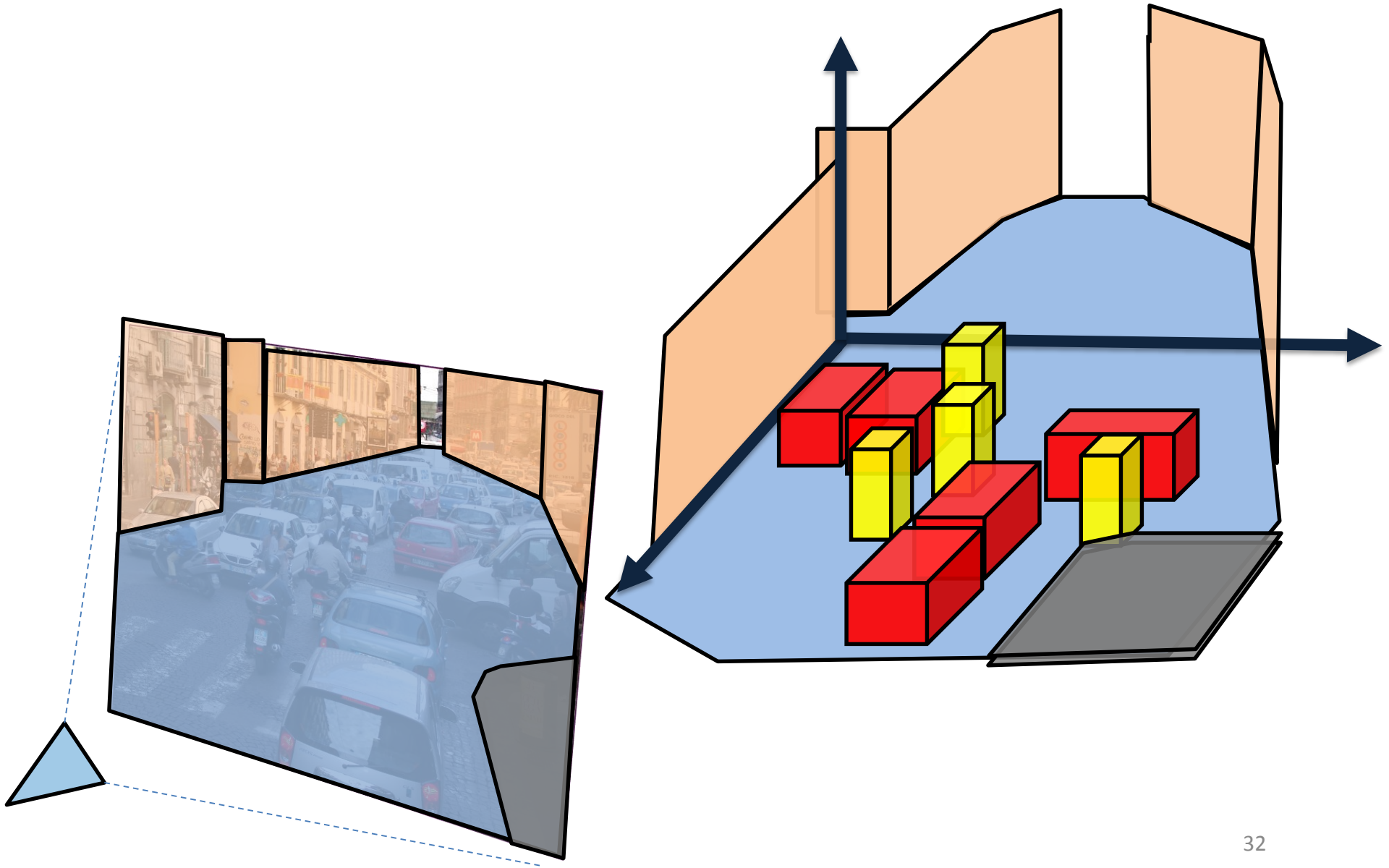
Scene understanding is an interplay between objects and space



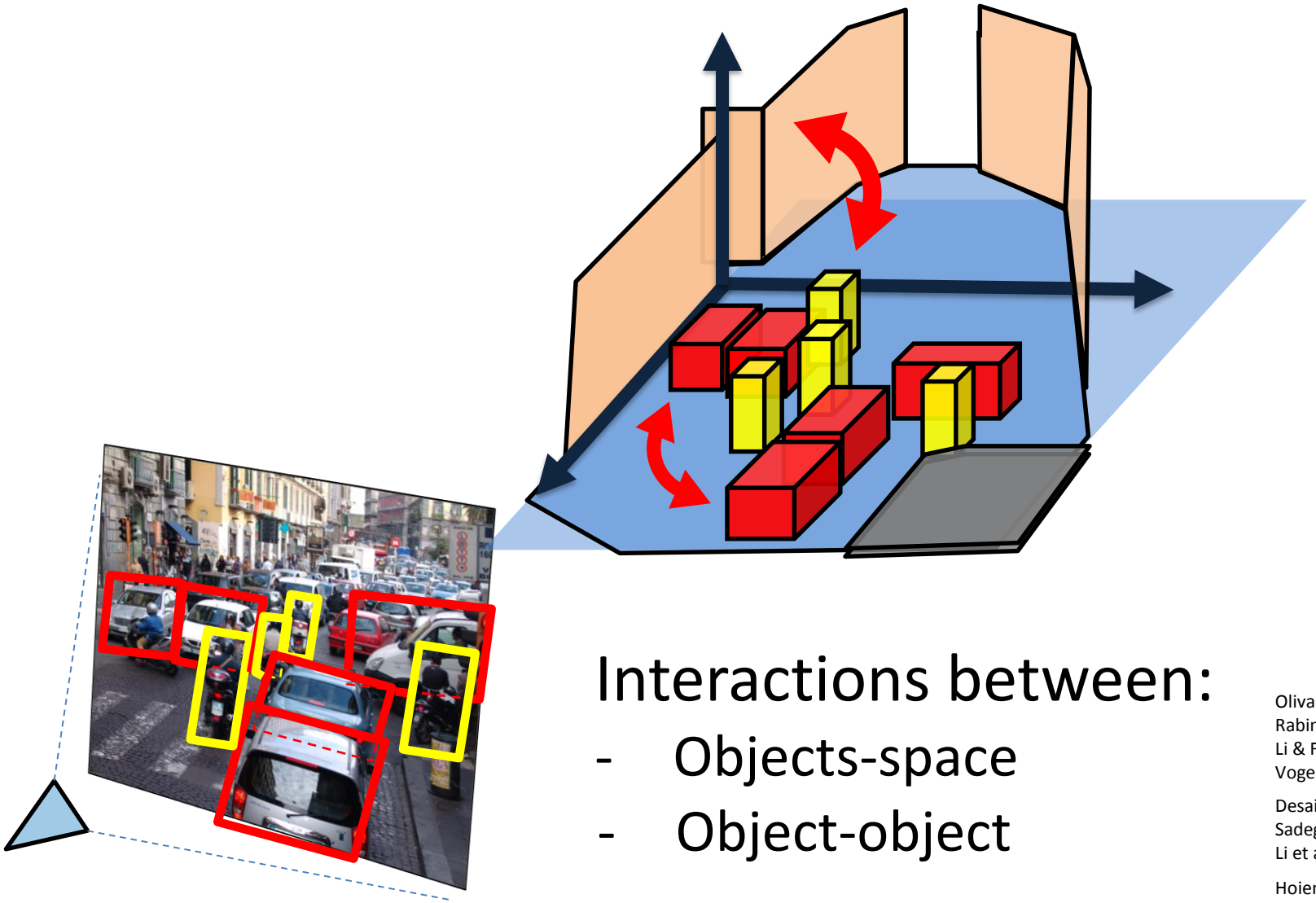
3D space is shaped by its objects



Objects are placed into 3D space



Interplay between objects and space



Interactions between:

- Objects-space
- Object-object

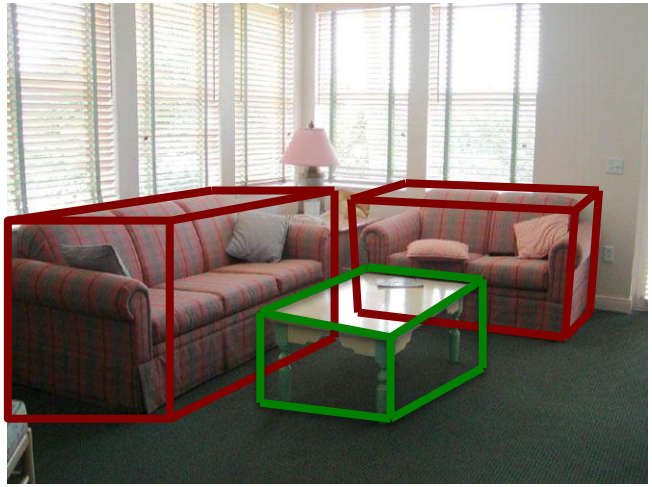
Oliva & Torralba, 2007
Rabinovich et al, 2007
Li & Fei-Fei, 2007
Vogel & Schiele, 2007

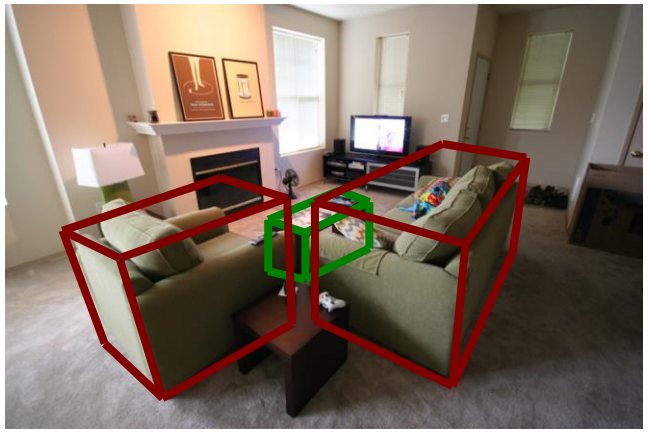
Desai et al, 2009
Sadeghi & Farhardi, 2011
Li et al, 2012

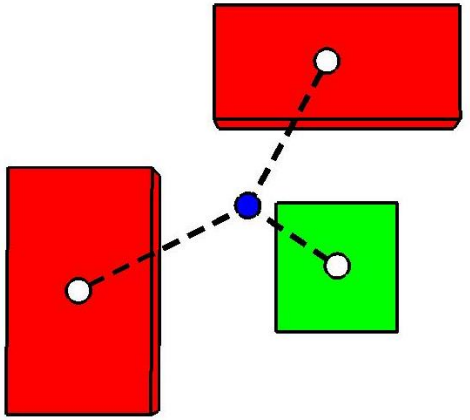
Hoiem et al, 2006
Herdau et al., 2009
Gupta et al, 2010
Fouhey et al, 2012

3D Geometric Phrases

Choi, Chao, Pantofaru, Savarese, CVPR 13



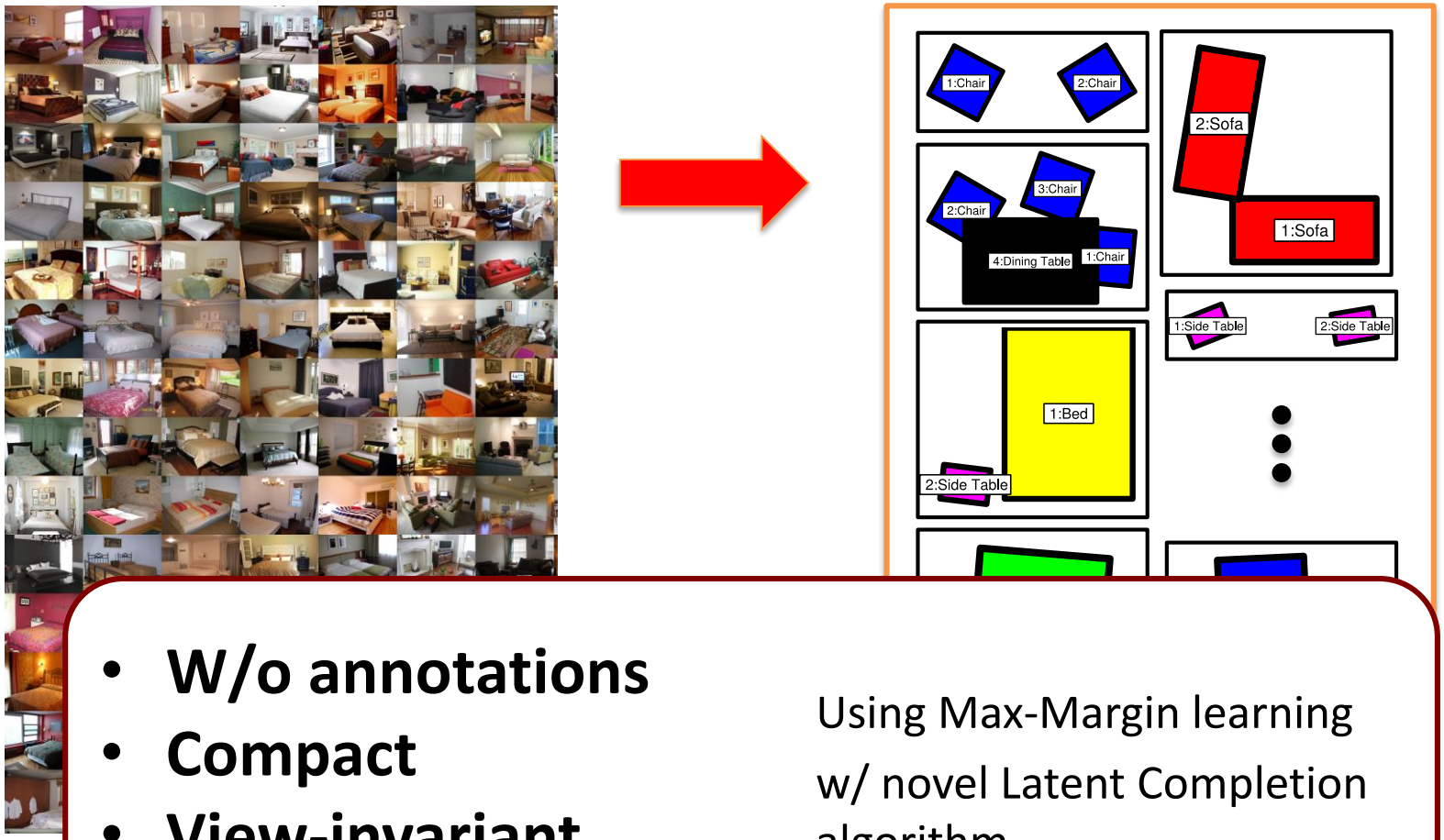




A ***3DGP*** encodes **geometric** and **semantic** relationships between groups of objects and space elements which frequently co-occur in **spatially consistent configurations**.

3D Geometric Phrases

Choi, Chao, Pantofaru, Savarese, CVPR 13

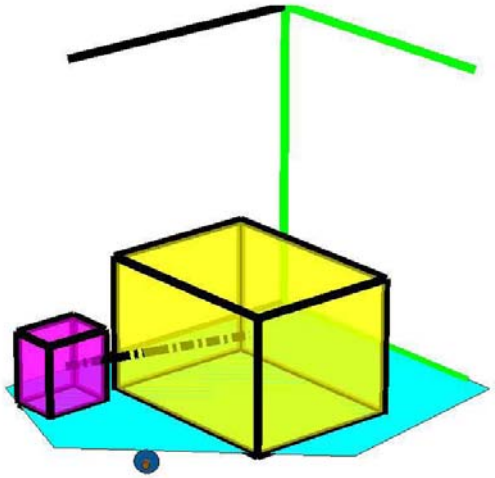


- **W/o annotations**
- **Compact**
- **View-invariant**

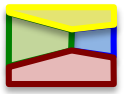
Using Max-Margin learning

w/ novel Latent Completion algorithm

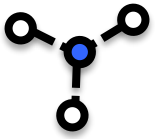
Results



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

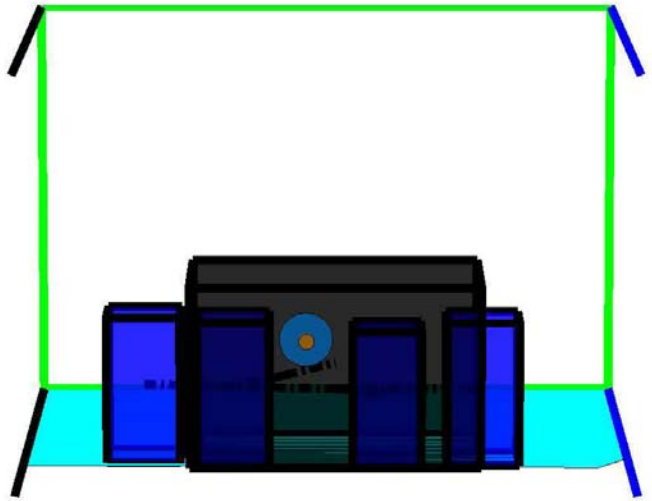


Estimated Layout

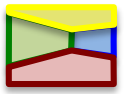


3D Geometric Phrases

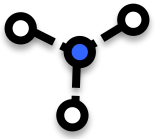
Results



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

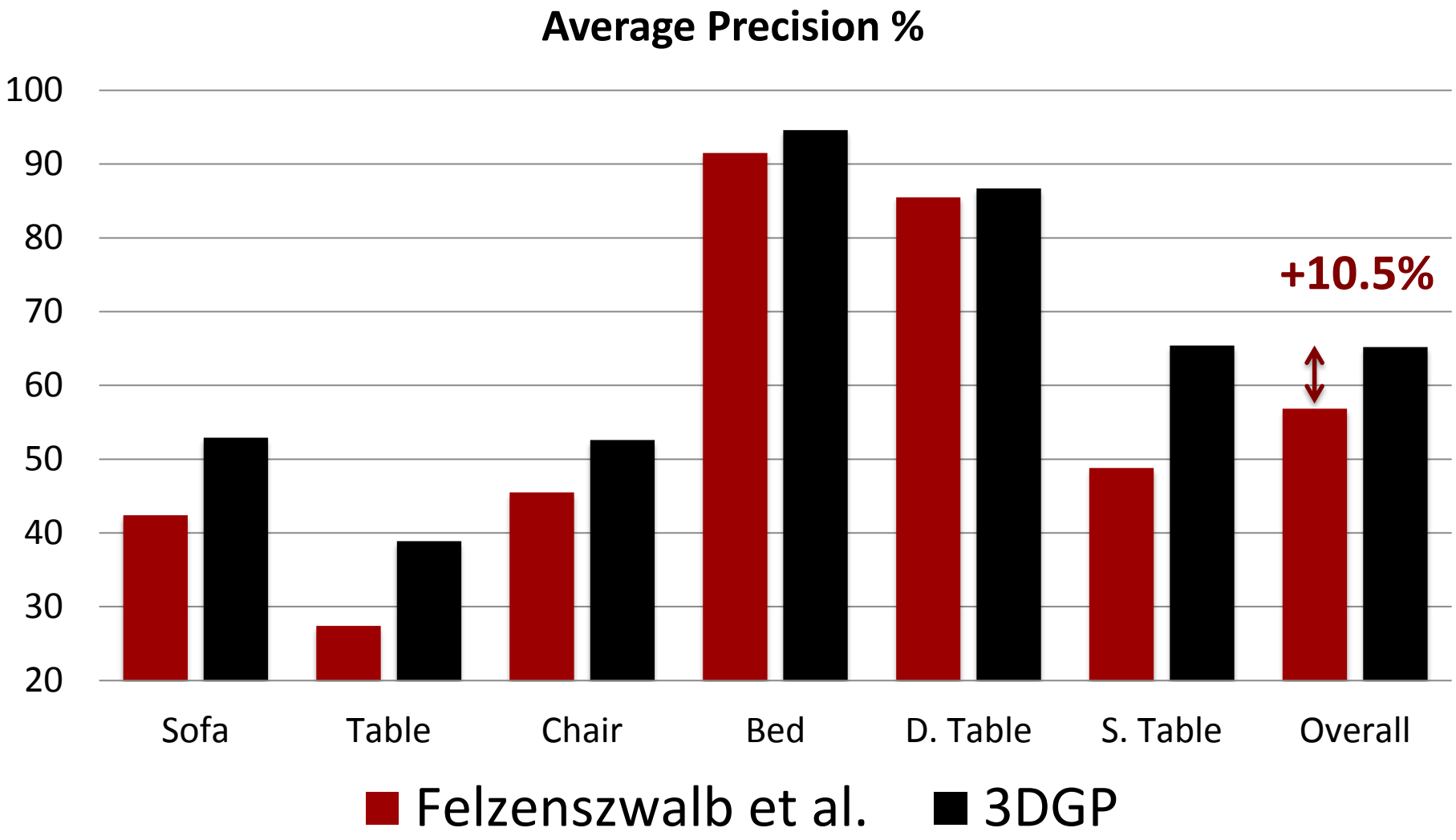


Estimated Layout

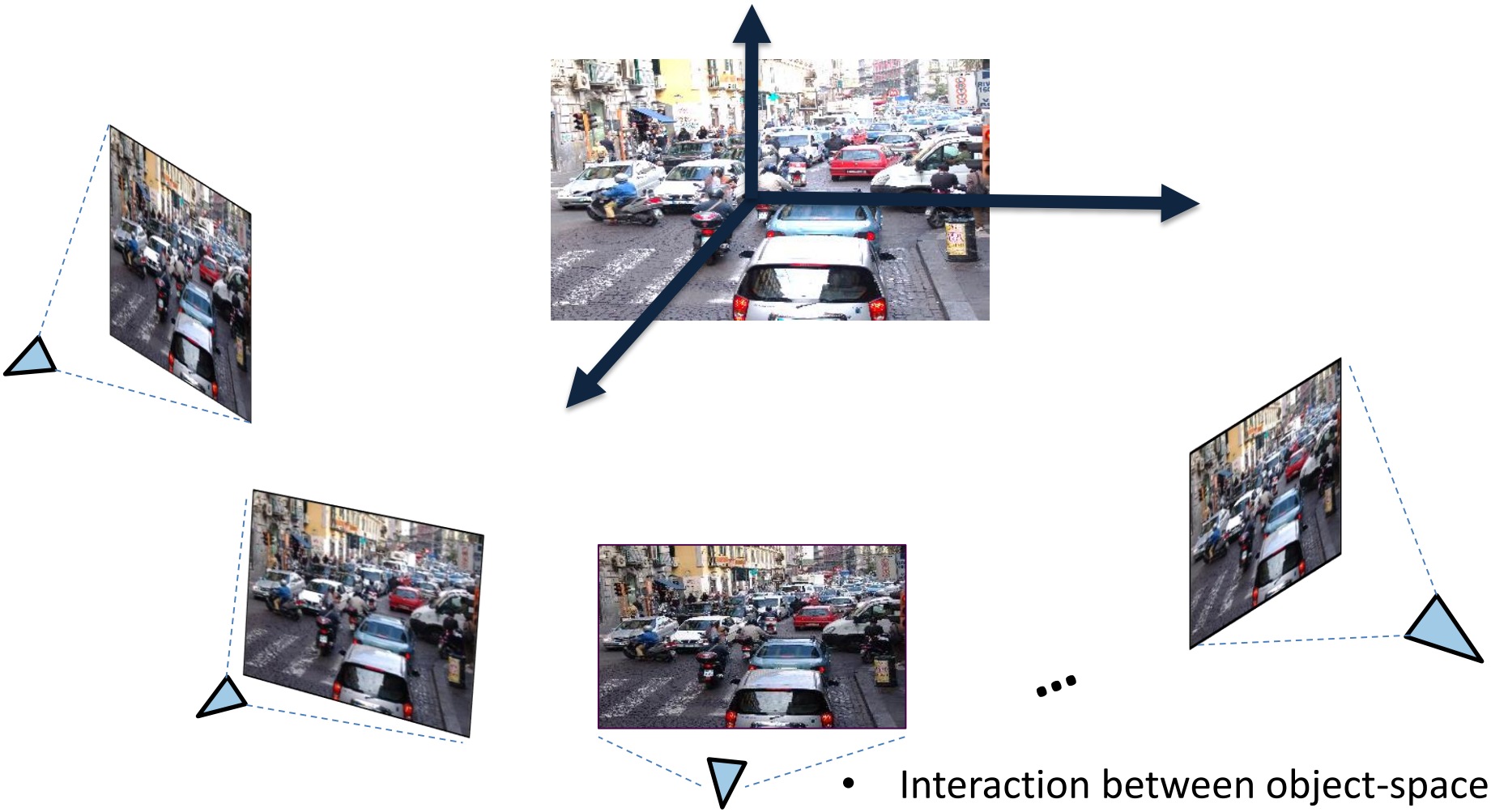


3D Geometric Phrases

Results

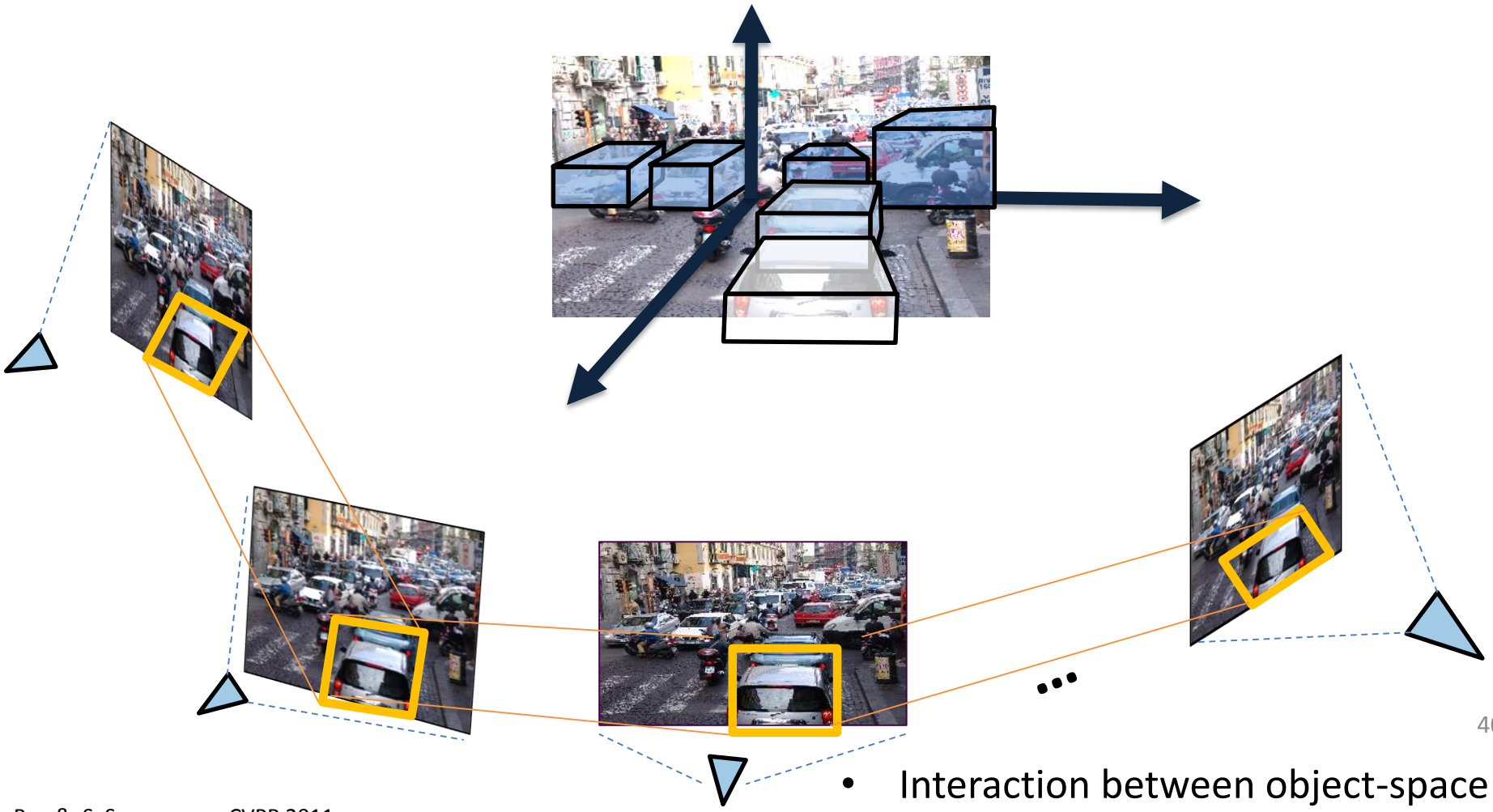

Average Precision %
100
90
80
70
60
50
40
30
20
+10.5%
Sofa
Table
Chair
Bed
D. Table
S. Table
Overall
Felzenszwalb et al.
3DGP

Scene understanding from images



- Interaction between object-space
- Interaction among objects

Scene understanding from images

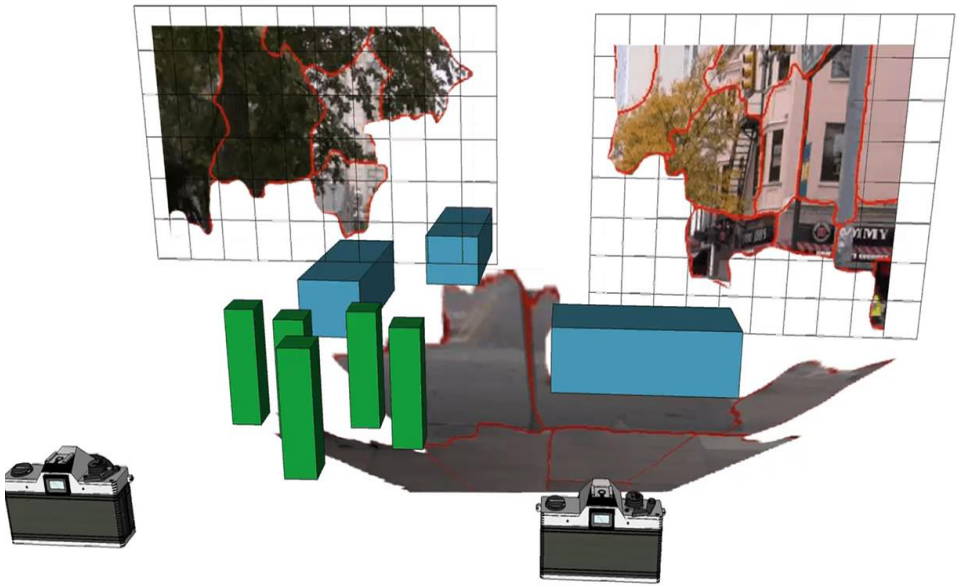


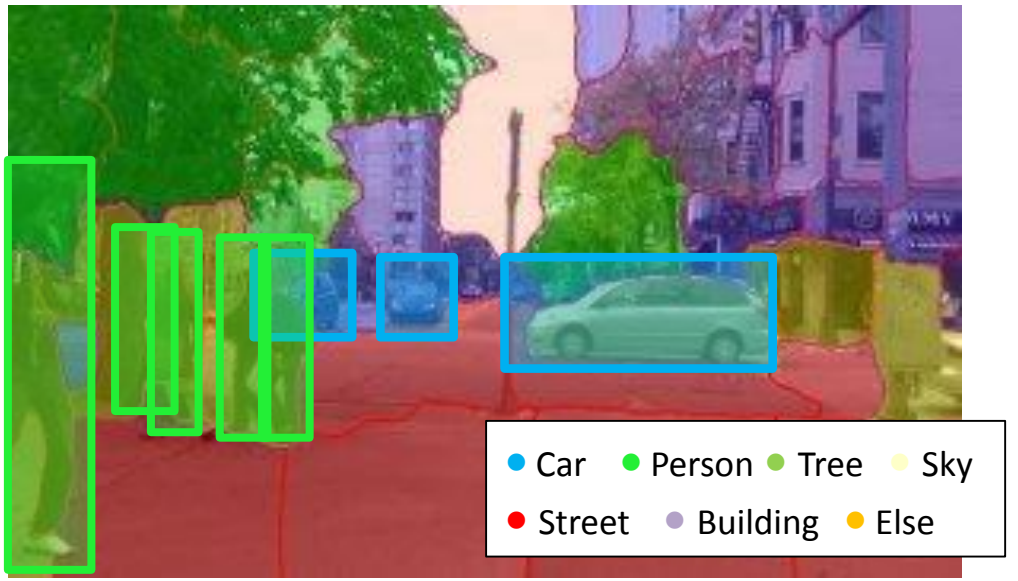
- Interaction between object-space
- Interaction among objects
- **Transfer semantics across views**

Bao & S. Savarese, CVPR 2011
Bao, Bagra, Savarese . CORP – ICCV 2011 **(Best student paper award!)**
Bao, Bagra, Chao, Savarese, CVPR 2012
Bao, Xiang, Savarese, ECCV 2012

Results

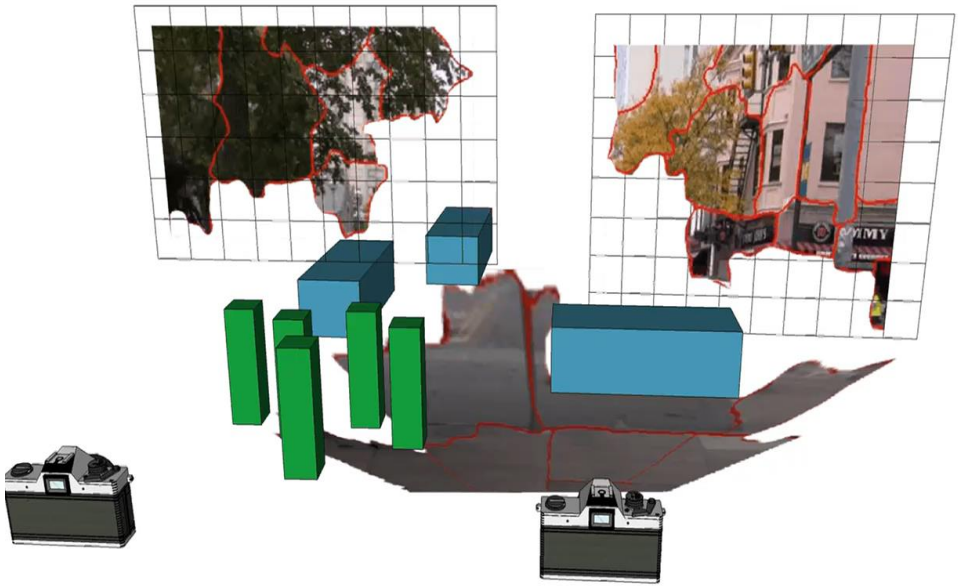
Input images

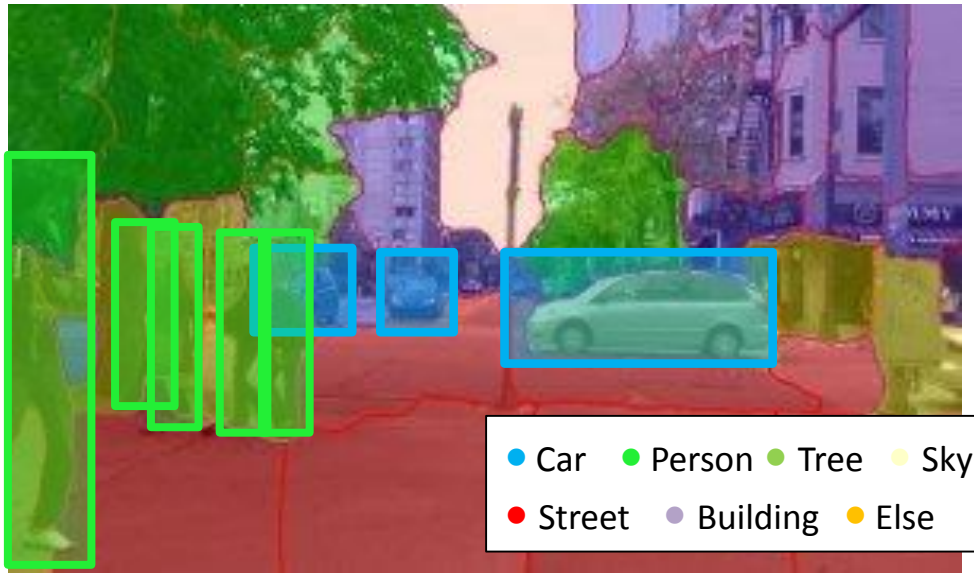




Results

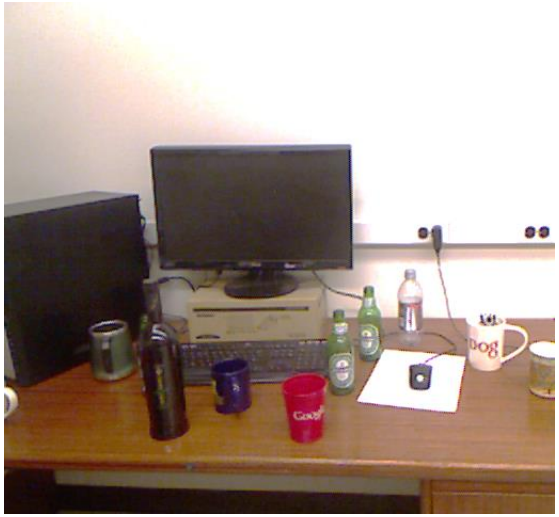
Input images

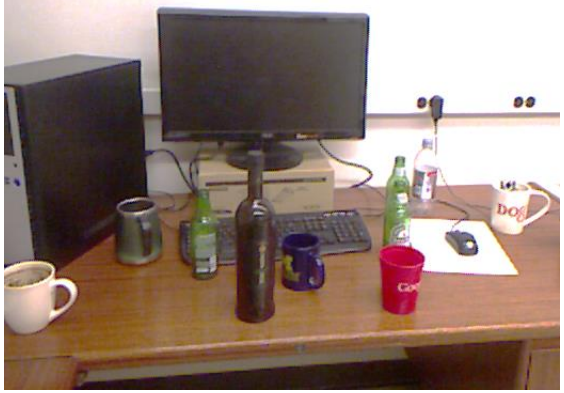




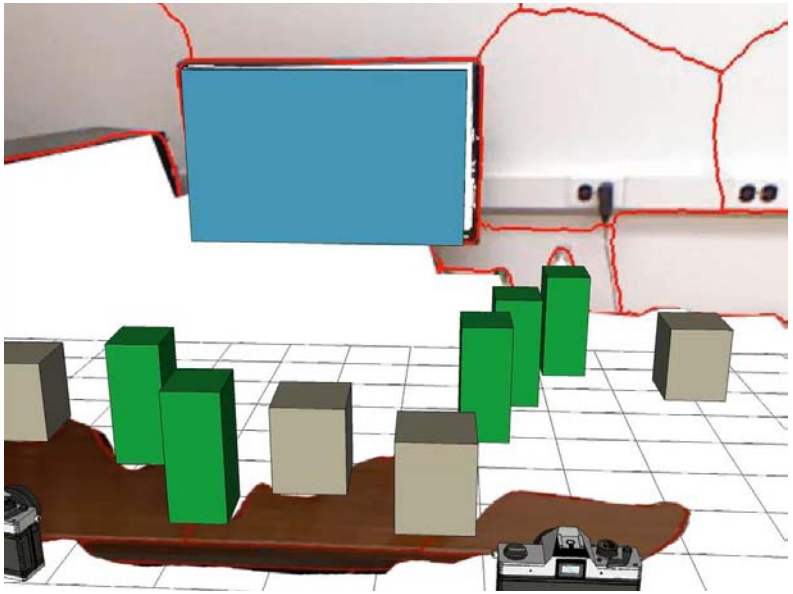
Results

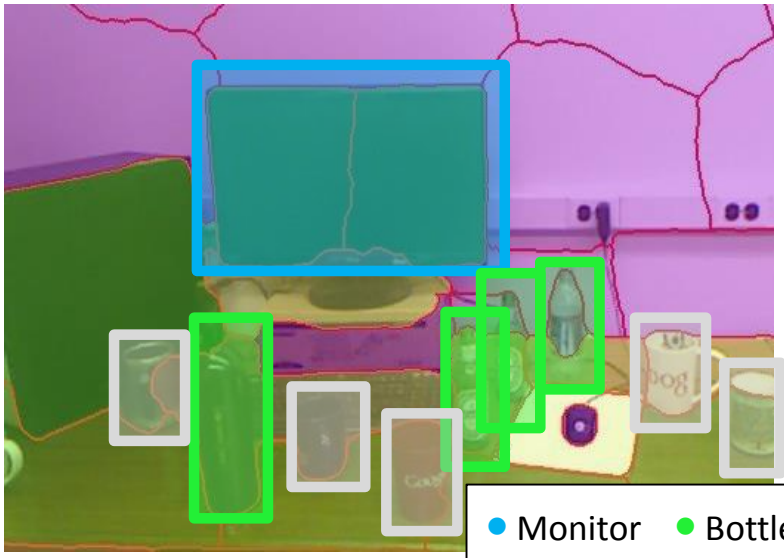
Input images





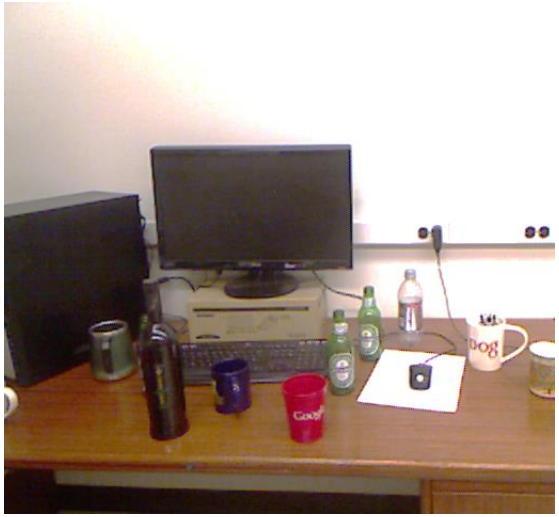
From the office dataset [Bao et al., 11]

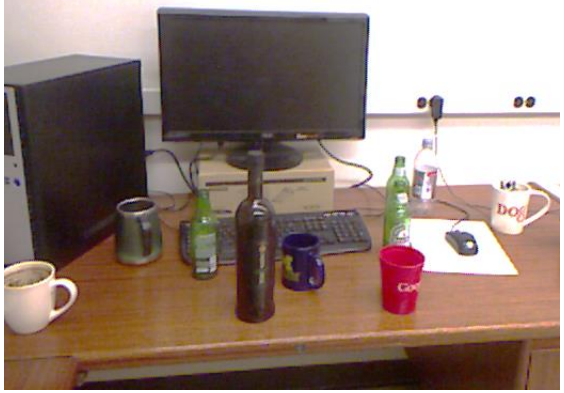




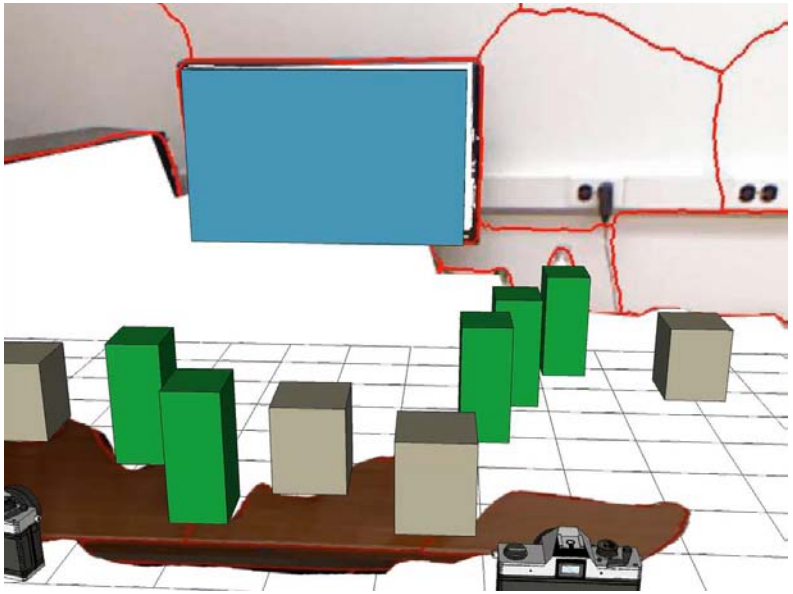
Results

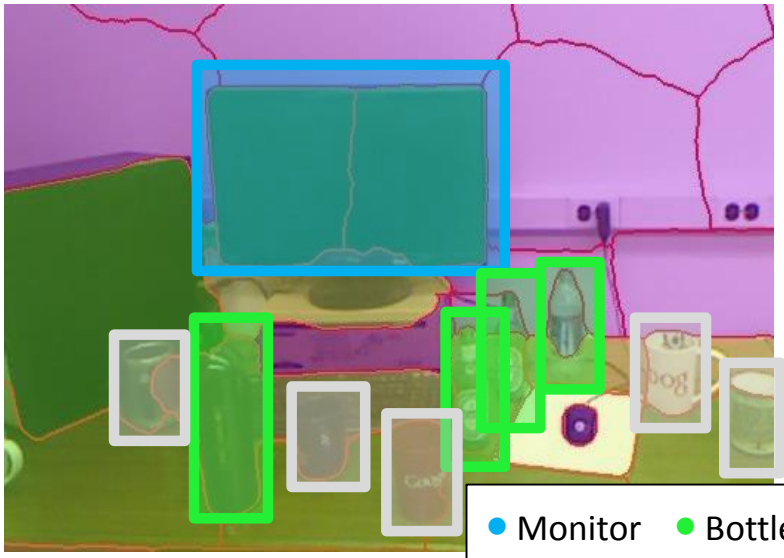
Input images





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

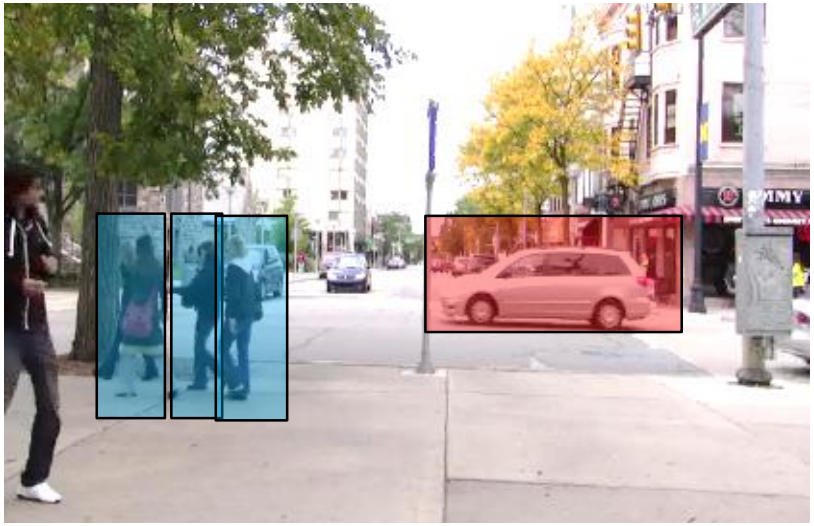


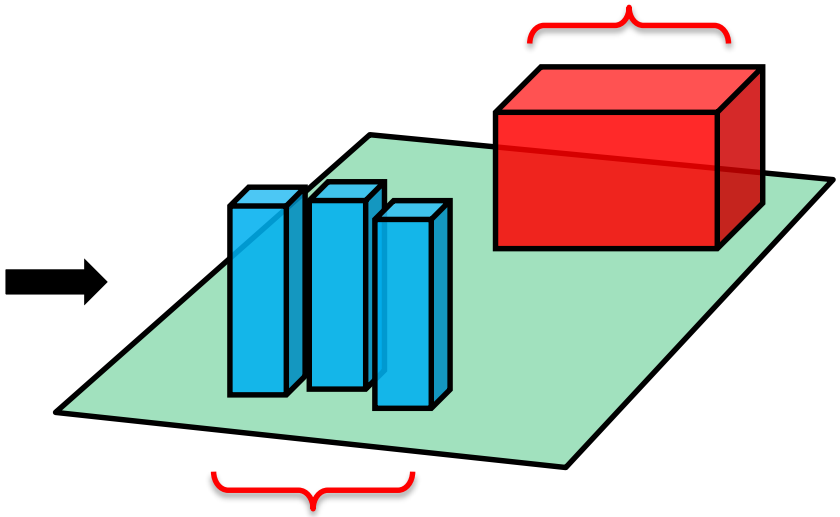


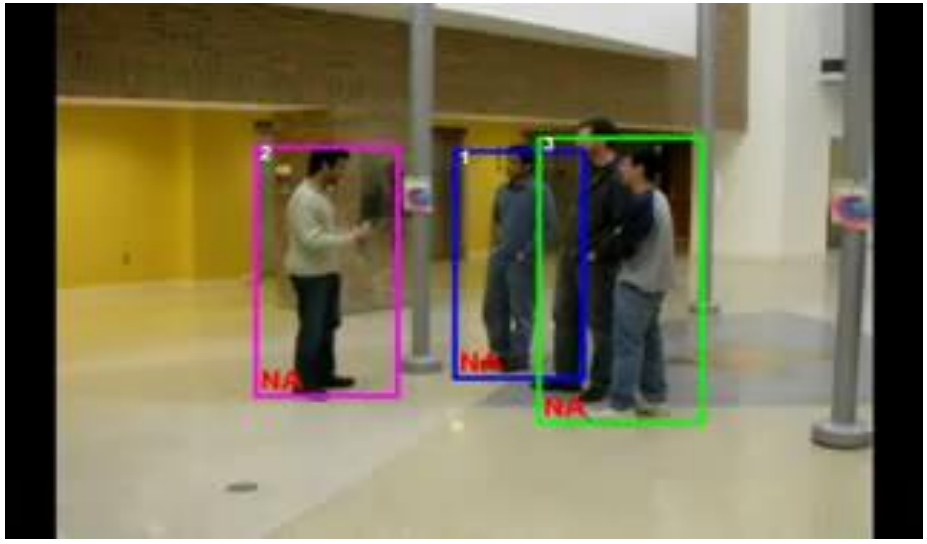
From objects to activities

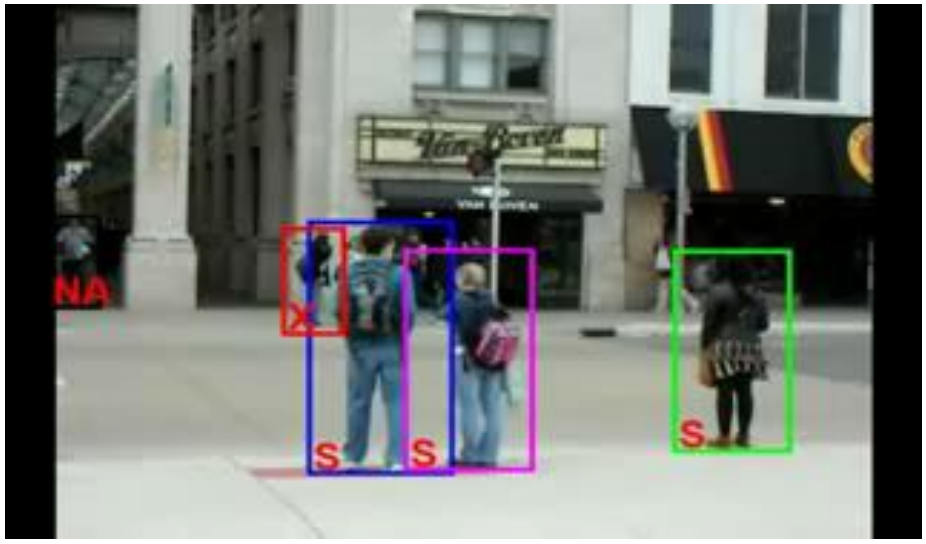
Choi & Savarese , ECCV 2010
Choi, Pantofaru, Savarese, CORP 2011
Choi, Pantofaru, Savarese, PAMI 2013

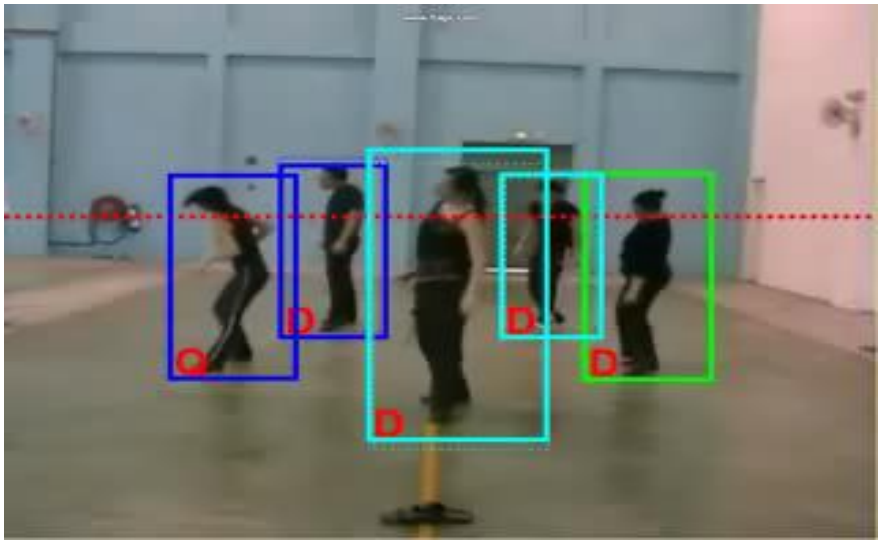
Choi et al., VSWS 09
Choi et al., CVPR 11
Choi & Savarese, ECCV 2012 (oral)







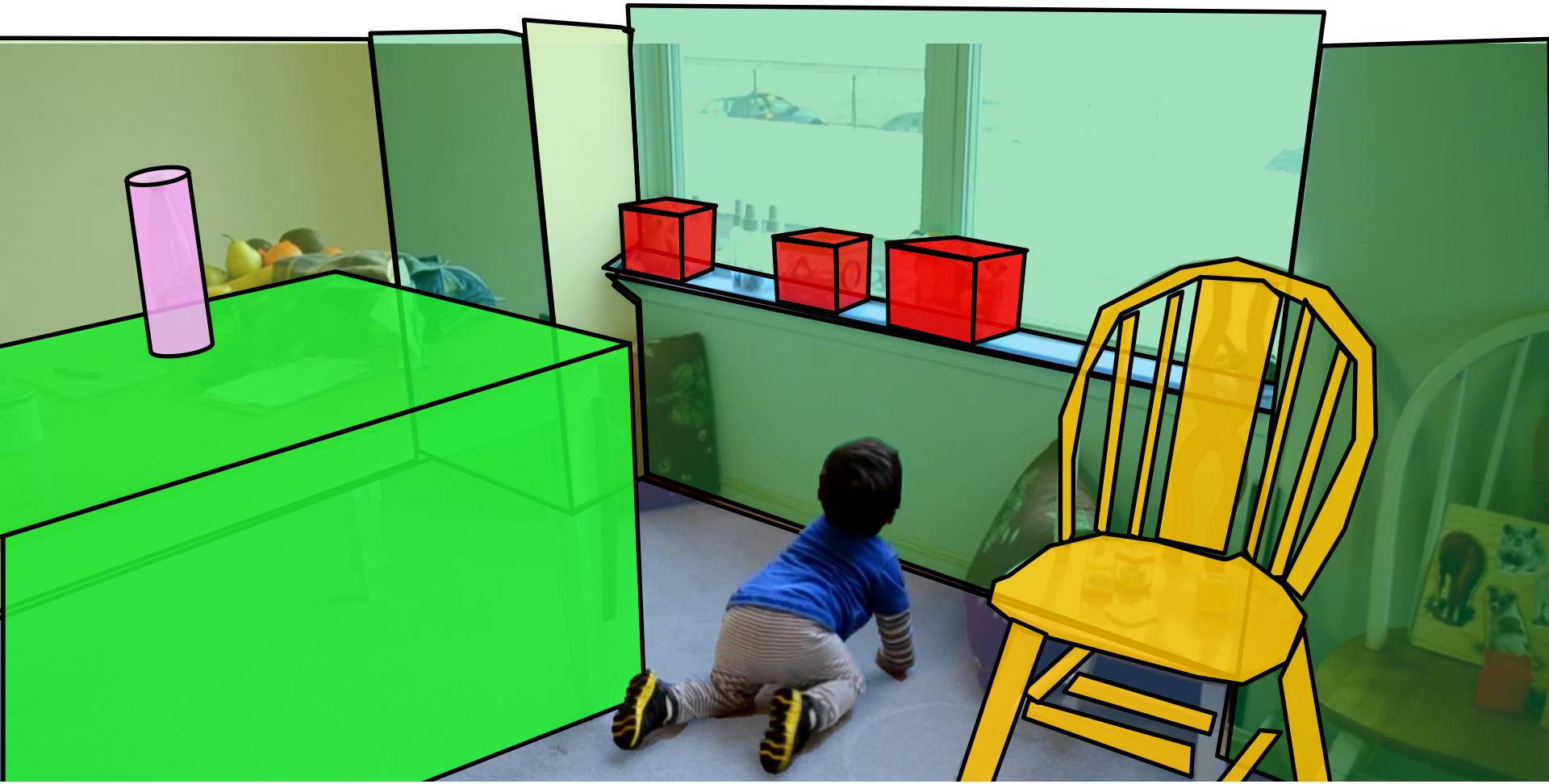




Choi & Savarese, CVPR 11
Choi & Savarese, ECCV 2012

X: Crossing, S: Waiting, Q: Queuing,
W: Walking, T: Talking, D: Dancing

Conclusions

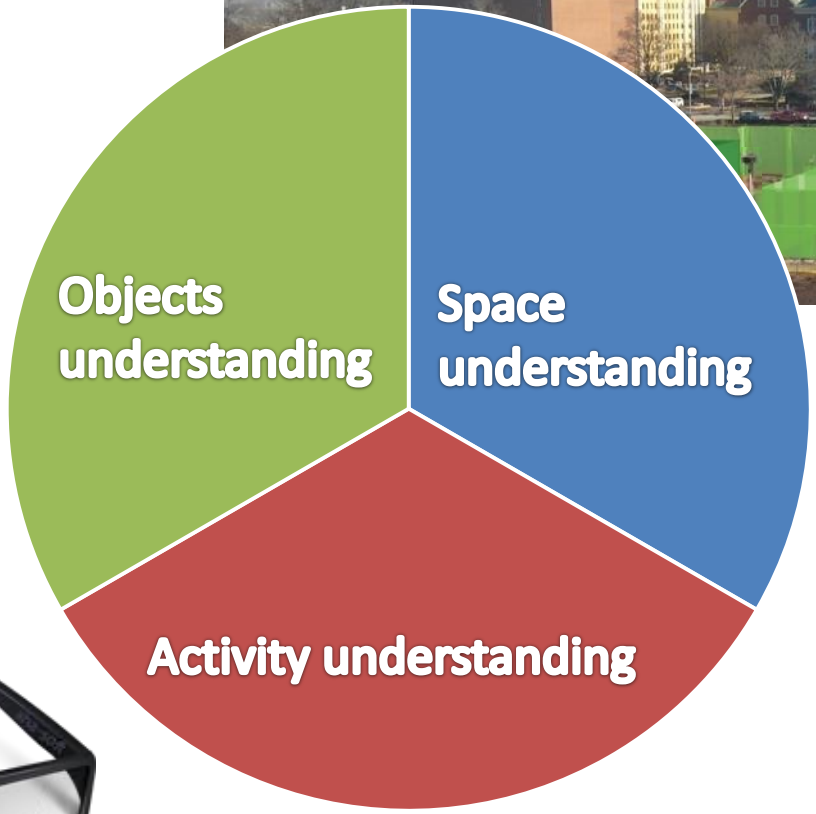


- From images to the 3D physical world
- Interplay between space and semantics

Applications

Vision for the blinds

Construction monitoring



Mobile vision

Safe driving

New intelligent digital interfaces between us and the 3D world

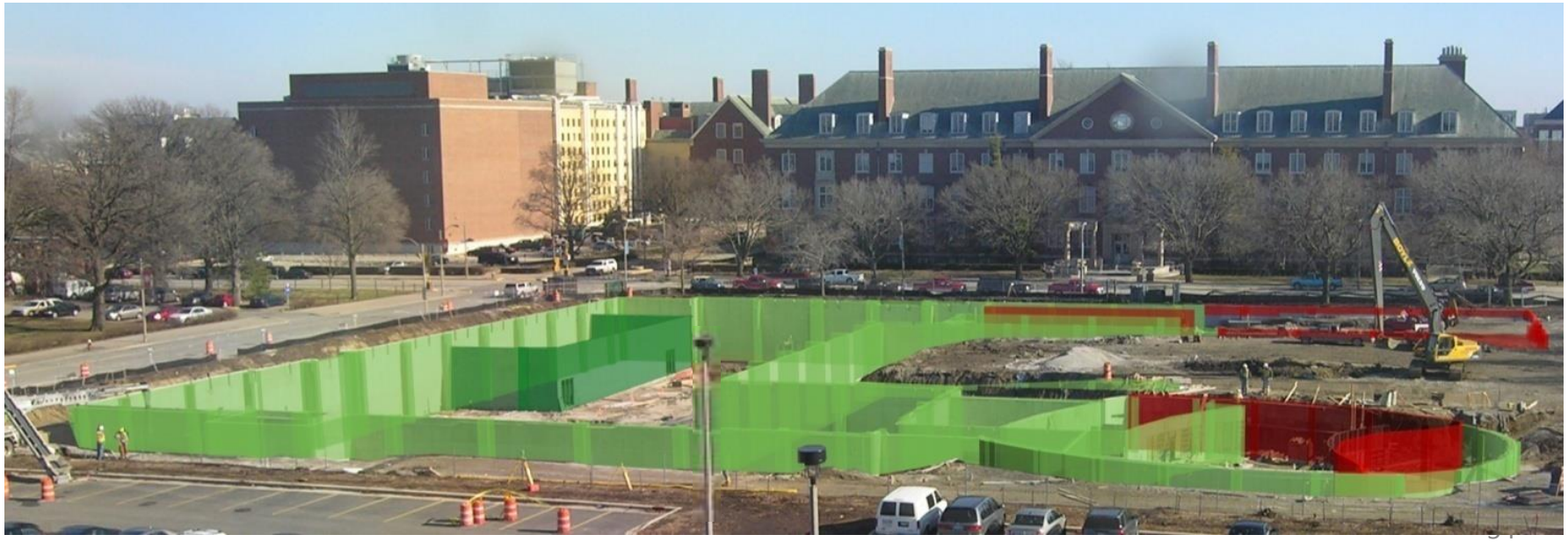
New generation of autonomous agents that can operate safely alongside humans in dynamic environments

Automatizing large scale information and environmental management tasks

Ahead of Schedule

On Schedule

Behind Schedule



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

Computer vision meets civil engineering

San Francisco-Oakland Bay Bridge

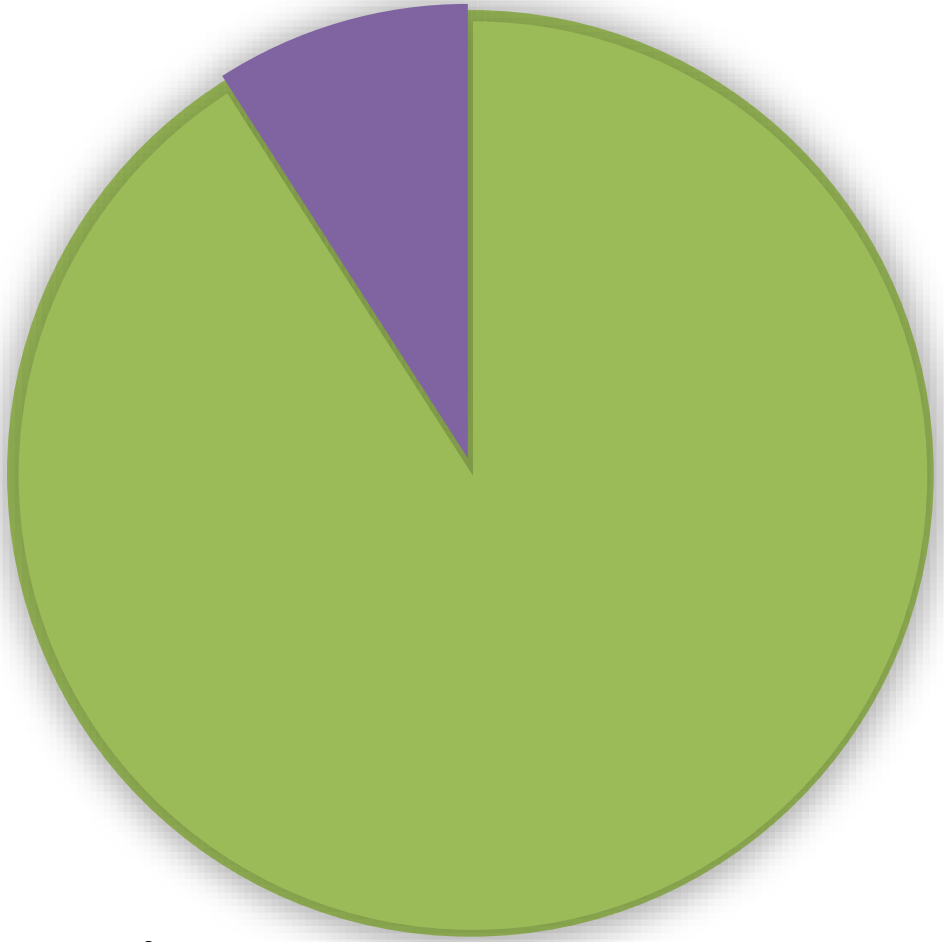
It came in:

- At 5 times the estimated original cost!
- 6 years behind schedule

San Francisco-Oakland Bay Bridge

Volume of construction industry in US

Loss: ~10 billions USD/year!

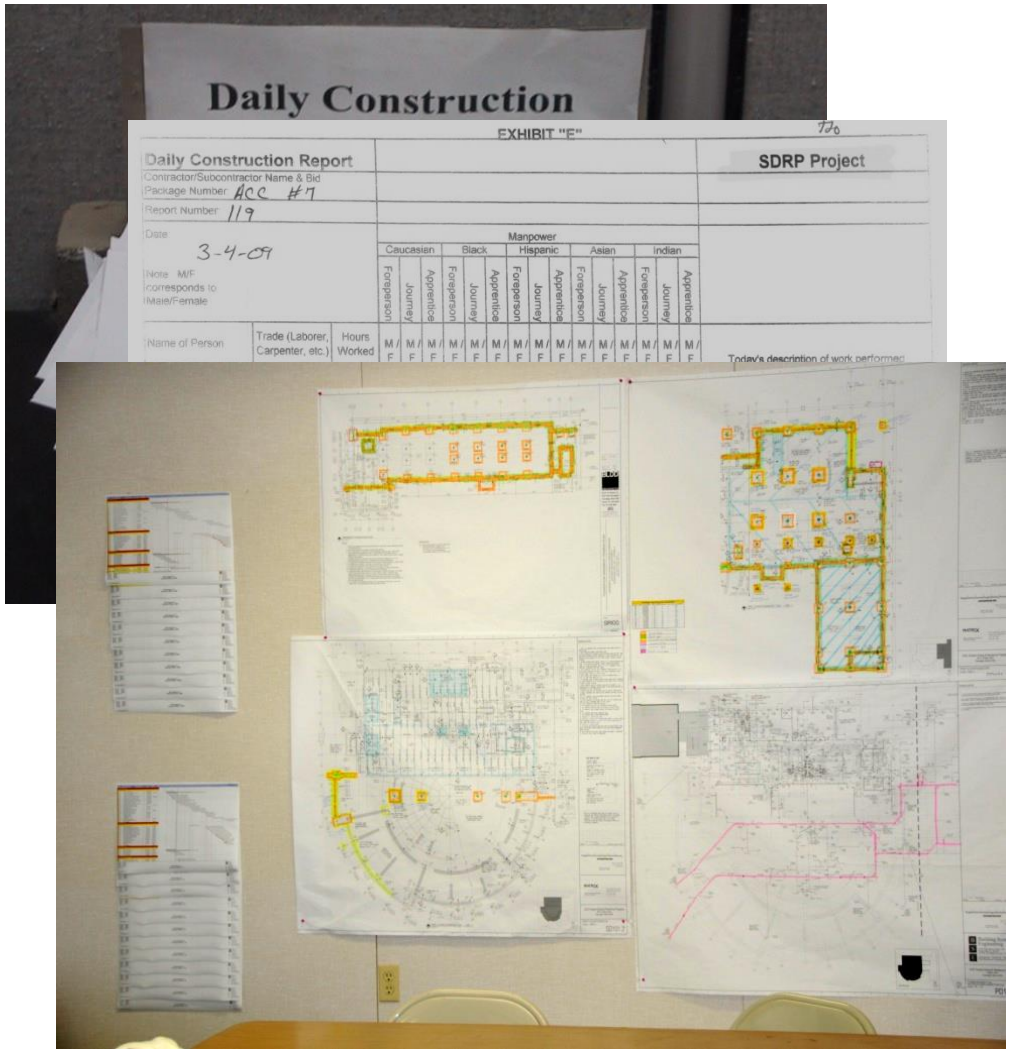


900 billions USD/year

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

It’s not a surprise!

- Manual
- Time consuming
- Non-systematic
- Error prone



“An improvement of as little as 1% can lead to up to 900 millions USD in savings in construction business”
[Census Bureau, www.census.gov, 2007]

Turner Construction trailer, Student Dining and Residential Hall Project, Champaign, IL (Image courtesy of Turner Construction, used by permission).

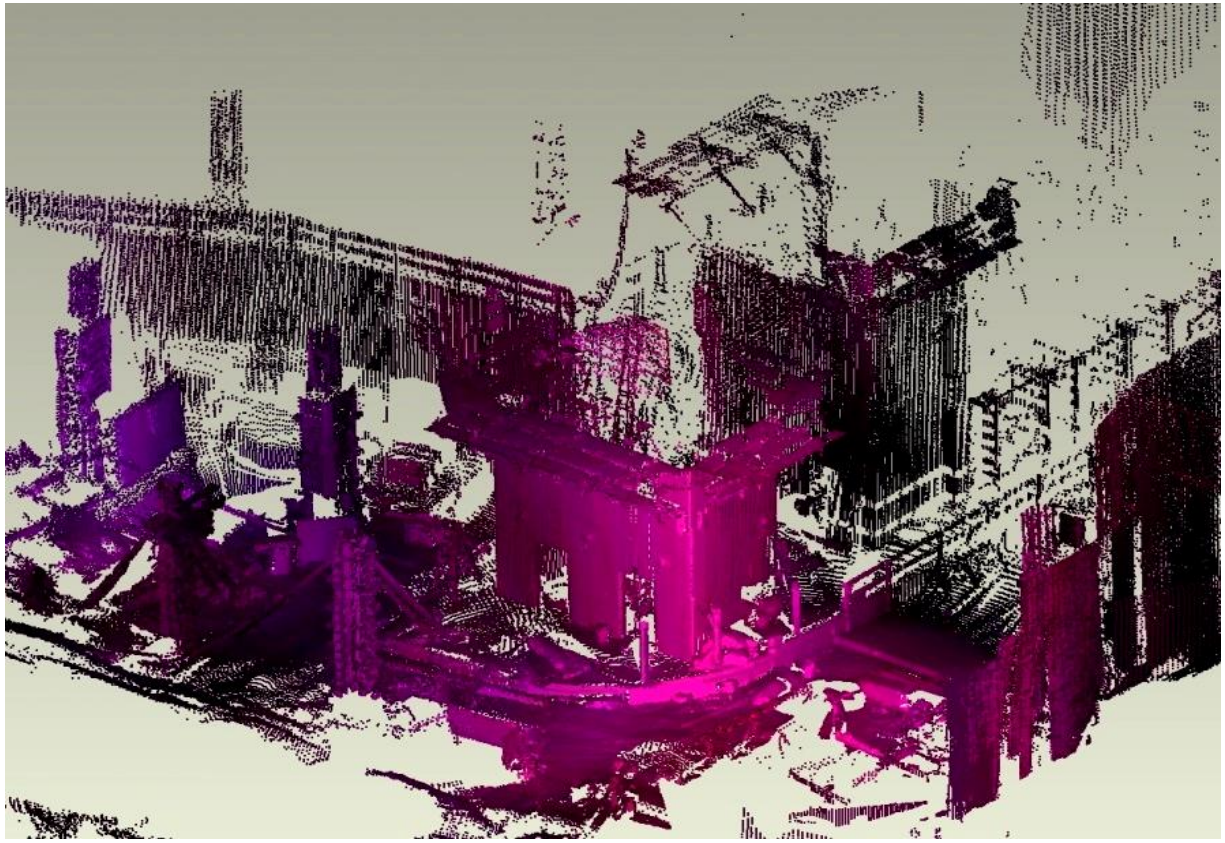
Opportunity to modernize age-old process in a profound way freeing up critical human resources

Towards modernity

(Kiziltas et al. 2008, Navon and Sacks 2007, Ergen et al. 2007, Jaselskis and El-Misalami 2003, Echeverry and Beltran 1997)

- Barcodes & RFID tags 

Towards modernity



(Bosche 2009, Bosche and Haas 2009, El-Omari and Moselhi 2008, Kiziltas et al. 2008, Akinci et al. 2006, Jaselkis et al. 2006, Su et al. 2006, Gordon et al. 2004, Teizer et al. 2005, Bosche and Haas 2008)

- Barcodes & RFID tags
- 3D laser scans

Still time consuming and expensive!

Our solution

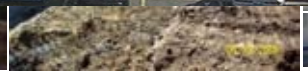
Golparvar-Fard, Peña-Mora & Savarese , 2008-2012

- Thousands of images
- Computer vision

Our solution

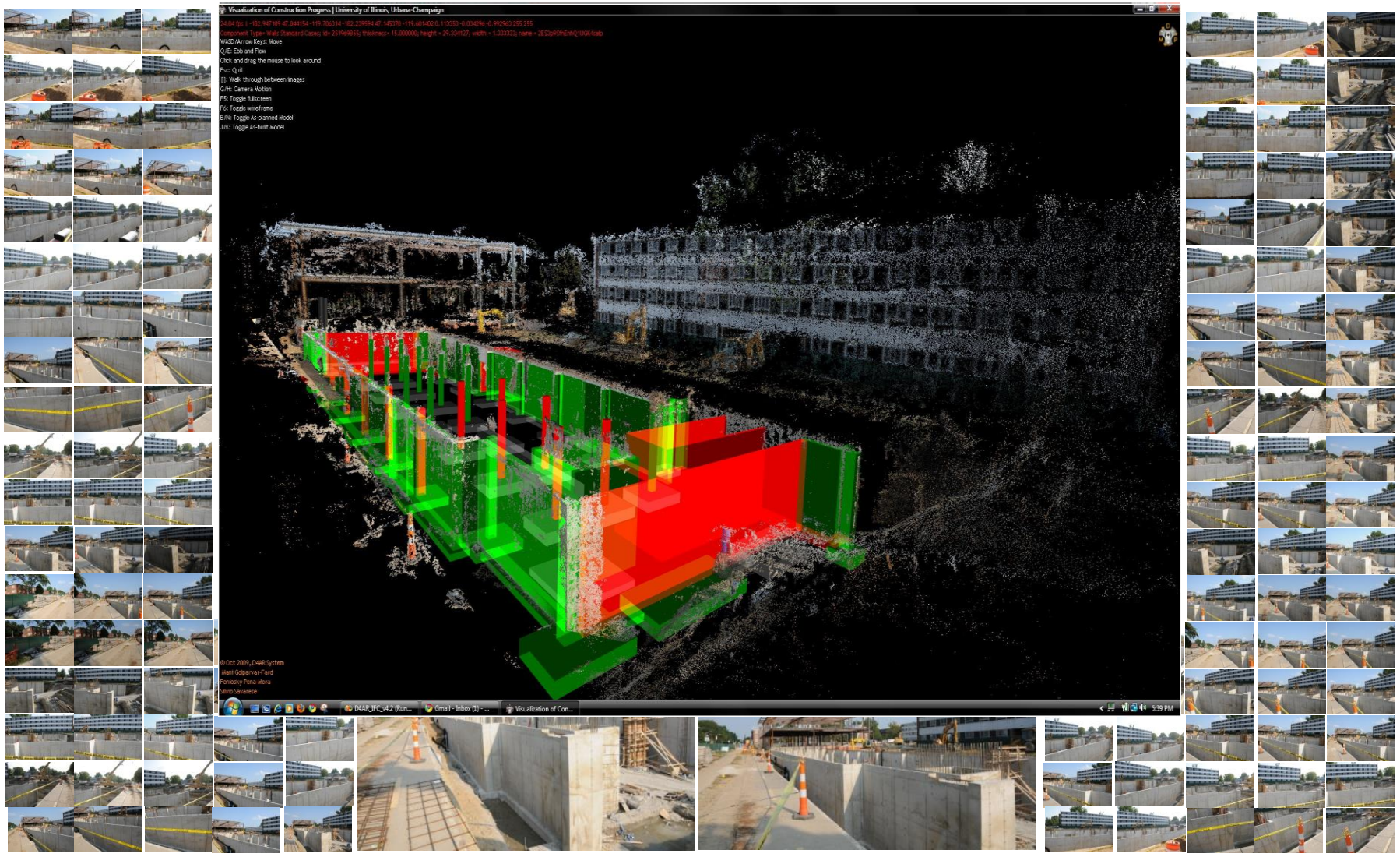
Images are cheap!

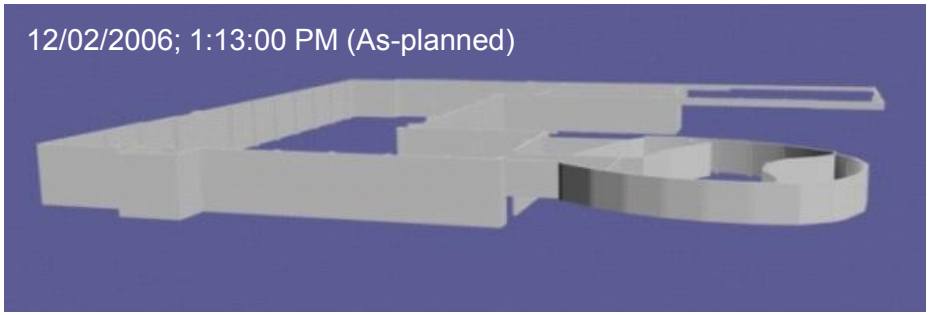
Our solution



Our solution

Our solution

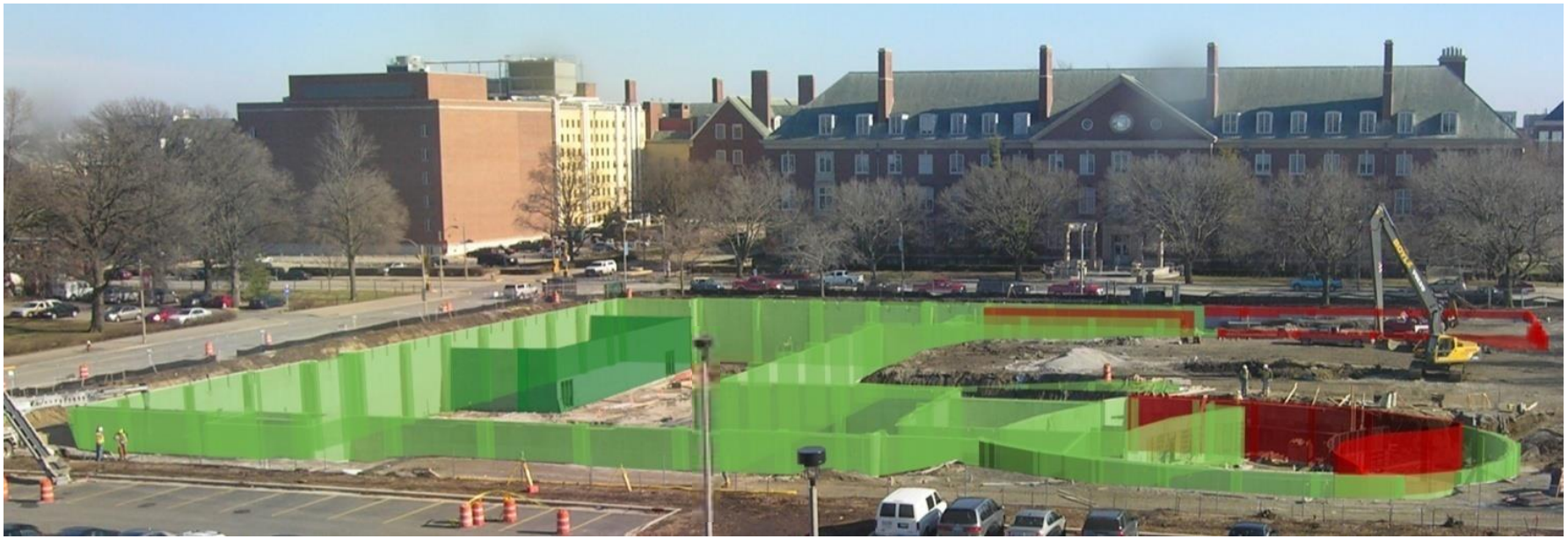


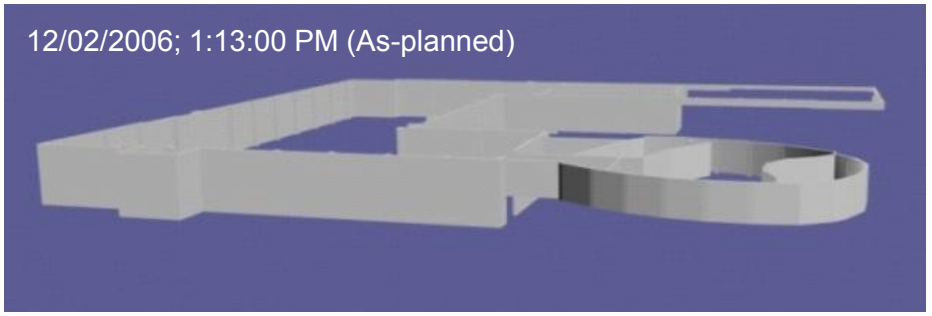


Ahead of Schedule

On Schedule

Behind Schedule



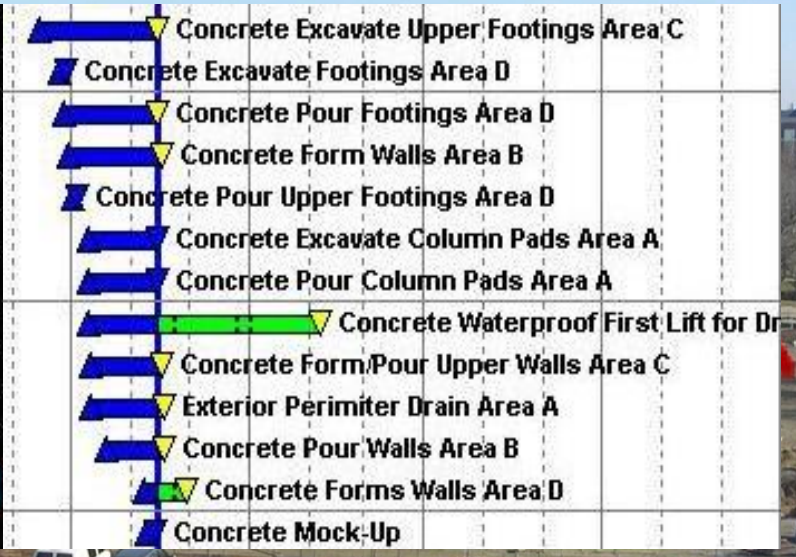


Ahead of Schedule

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



Summary

- Automate communication of performance deviations
- Reduction in delivery time
- Potential to identify unsafe locations/components
- Large impact in the civil engineering community

- **James R. Croes Medal, October 2013 (from the American Society of Civil of Engineers)**
- Best paper award from journal of CEM, 2011
- Best paper award at *AEC/FM 2010*
- Best paper award at Construction Research Congress 2009

Thank you!

TOYOTA

Performance analysis of construction projects

Project: College of Business Instructional Facility, Photograph Courtesy of Facilities & Services, UIUC.

Performance analysis of construction projects

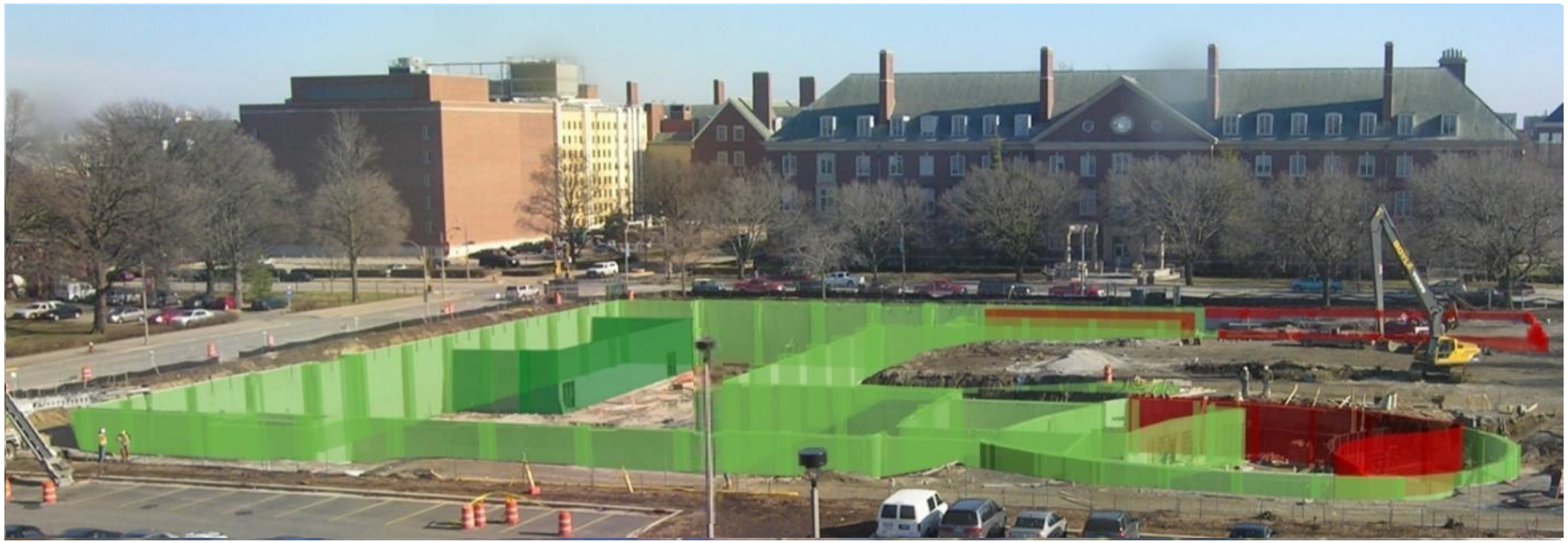
- Manual progress monitoring:
  - Time consuming
  - Non-systematic

- Improvement of as little as 1% can lead to up to 900 millions USD in savings in construction business
  [Census Bureau, www.census.gov, 2007]

Collecting; analyzing; reporting; recording

Performance analysis of construction projects

- Golparvar-Fard, Pena-Mora, Savarese , 2008-2012
- **James R. Croes Medal 2013**
- Best paper award from journal of CEM, 2011
- Best paper award at *AEC/FM 2010*
- Best paper award at Construction Research Congress 2009



Project: College of Business Instructional Facility, Photograph Courtesy of Facilities & Services, UIUC.

- Large impact in the civil engineering community
- Opportunity to modernize age-old labor-intensive process in a profound way, freeing up critical human resources

What's ahead

Granularity

- Representation
- Learning
- Computational demands

Scale

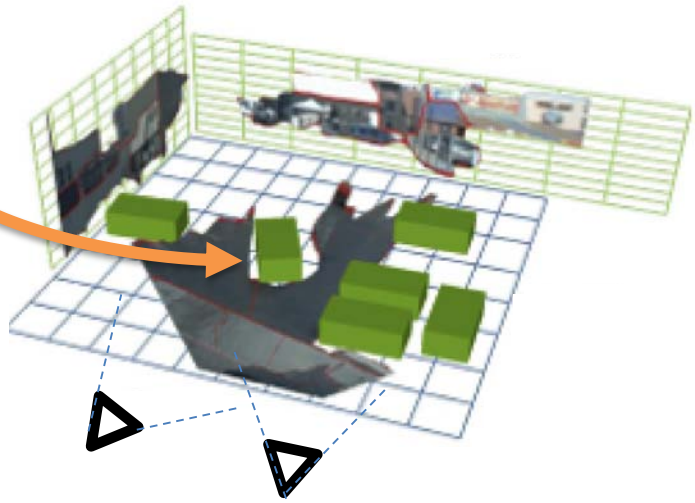
Thank you!

TOYOTA

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%**

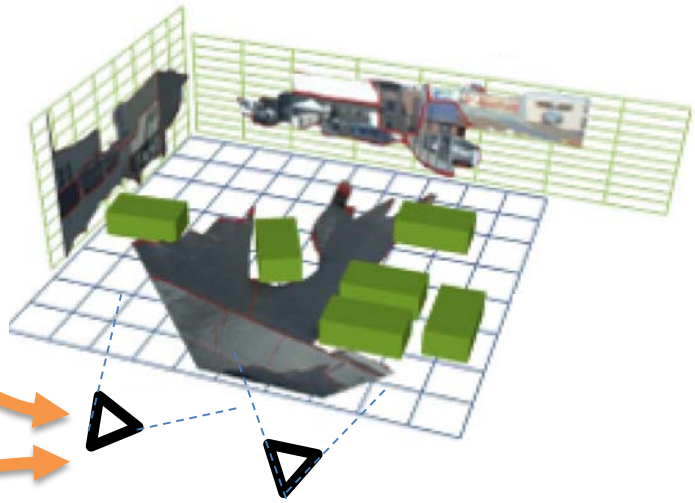
FORD CAMPUS dataset [Pandey et al., 09]
Office dataset [Bao et al., 11]

[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°**

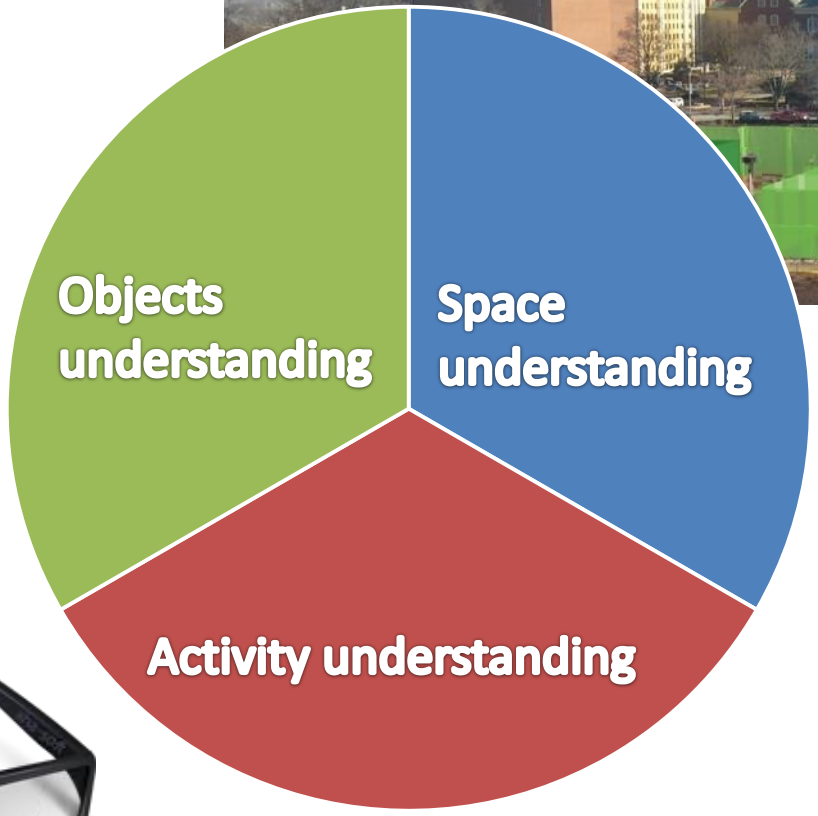


FORD CAMPUS dataset [Pandey et al., 09]
Office dataset [Bao et al., 11]
Street dataset [Bao et al., 11]

Applications

Vision for the blinds

Construction monitoring



Mobile vision

Safe driving