Announcements

- No class on Tuesday (moved to Wed)
- In-class Presentations on:
  - Wed March 12 at
  - Thursday March 13 from 11-12:15pm
- Posters Thursday from 12:15-2pm
- Attendance is highlight recommended
- Best presentation award sponsored by Yelp
- Free food!
- See Piazza and website for more information about the format

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
- No midterm; no final

Lecture 17
Closure

- Datasets in computer vision
- 3D scene understanding

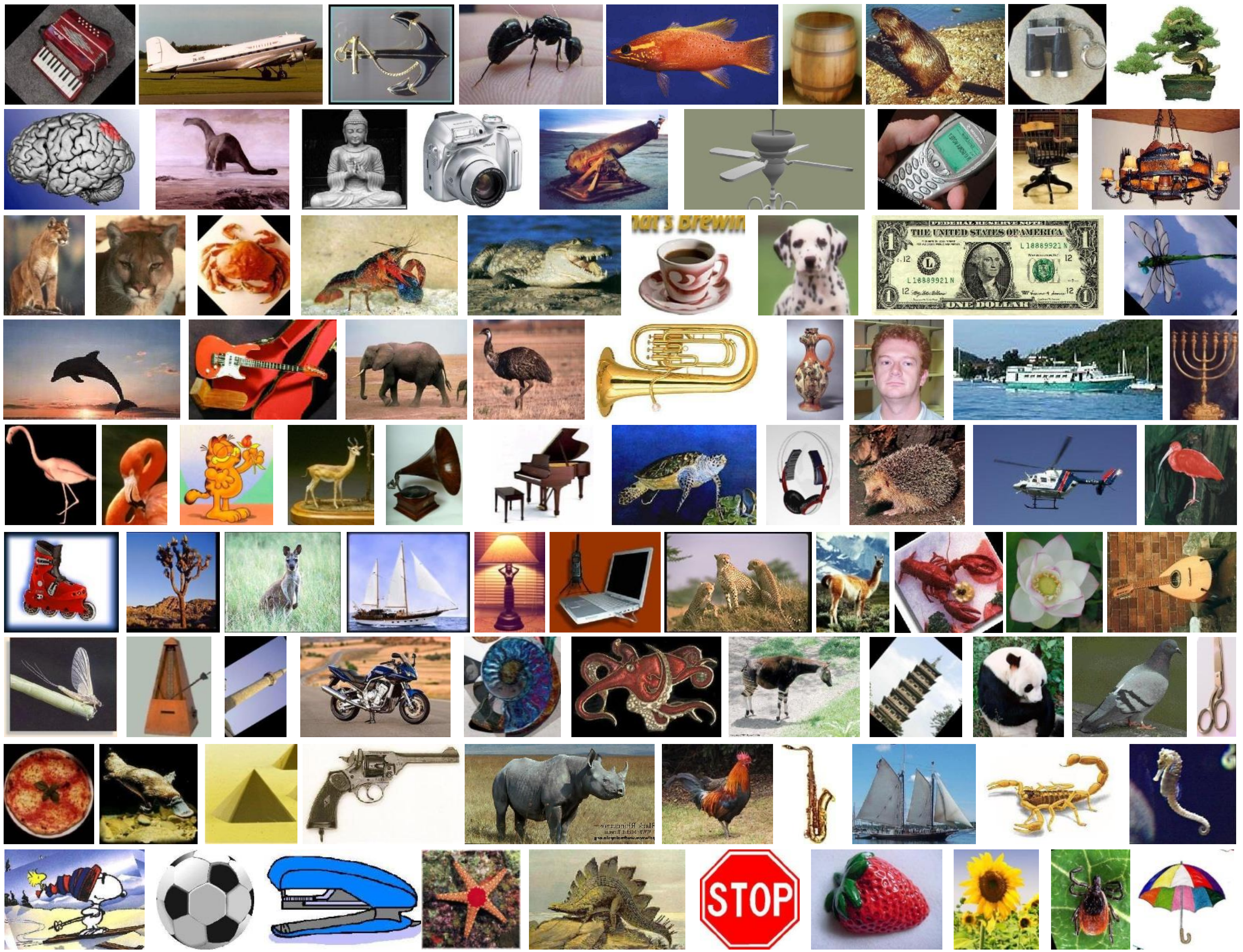
Caltech 101

- 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



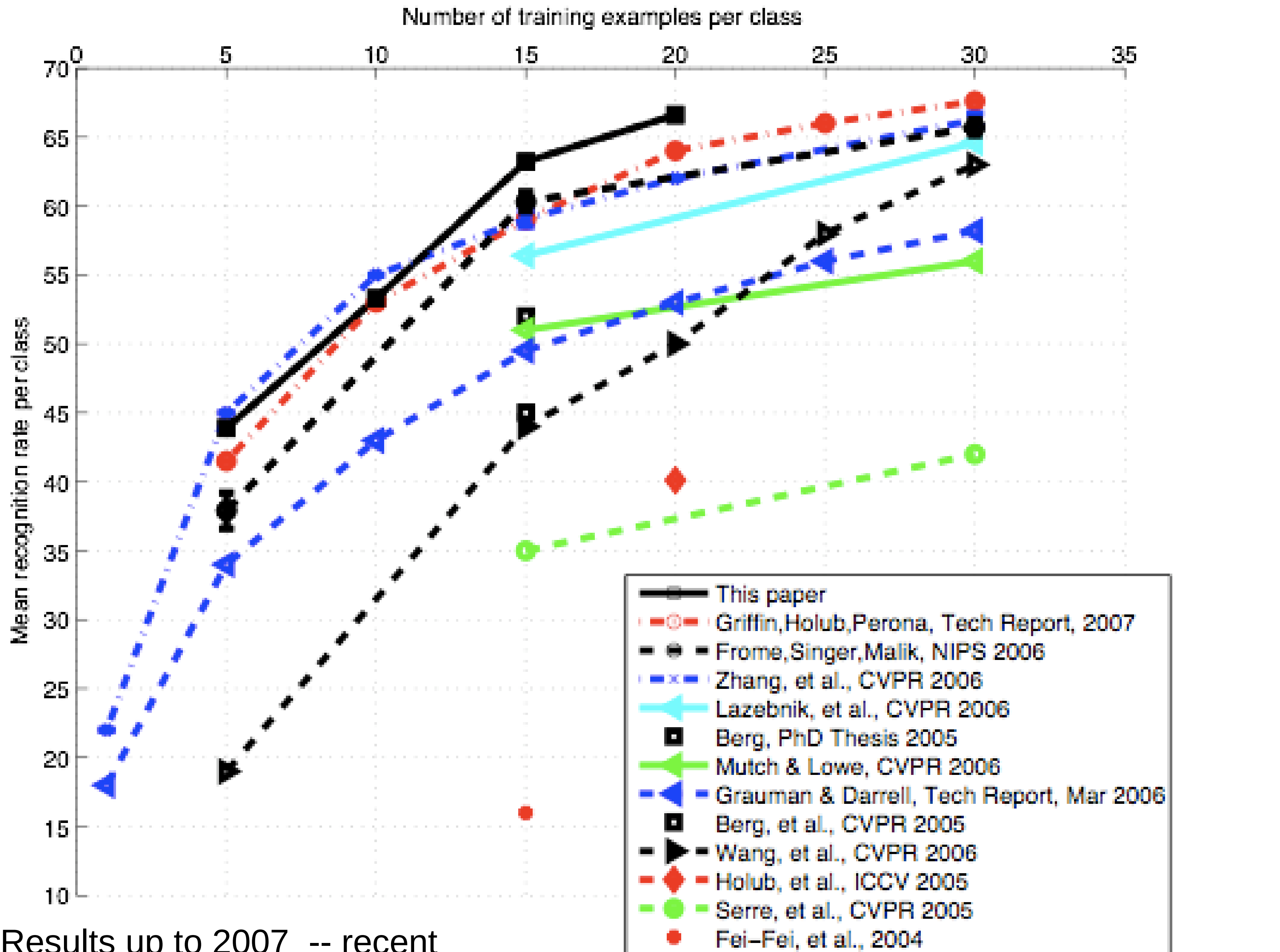
Caltech-101: Drawbacks

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

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





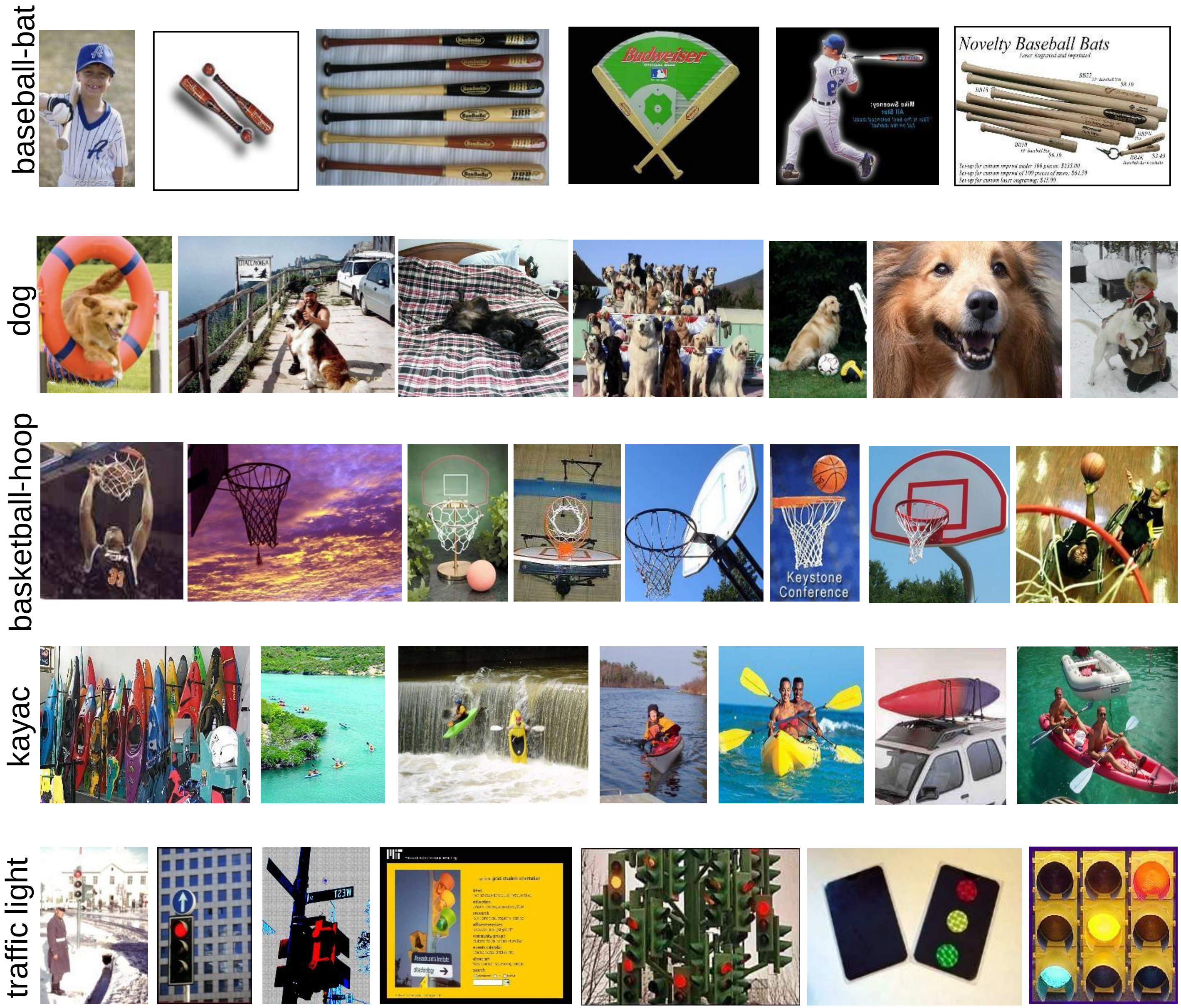


Results up to 2007 -- recent methods obtain almost 100%

Caltech-256

- Smallest category size now 80 images
- About 30K images
- Harder
  - Not left-right aligned
  - No artifacts
  - Performance is halved
  - More categories
- New and larger clutter category

Caltech 256 images



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

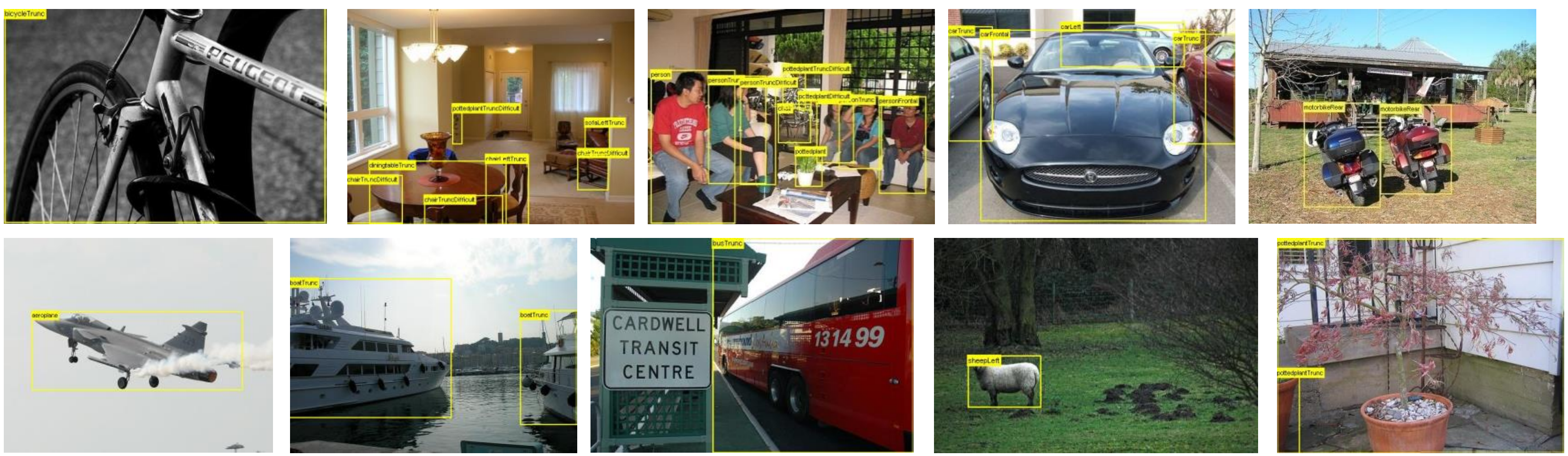
Mark Everingham
Luc Van Gool
Chris Williams
John Winn
Andrew Zisserman

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”



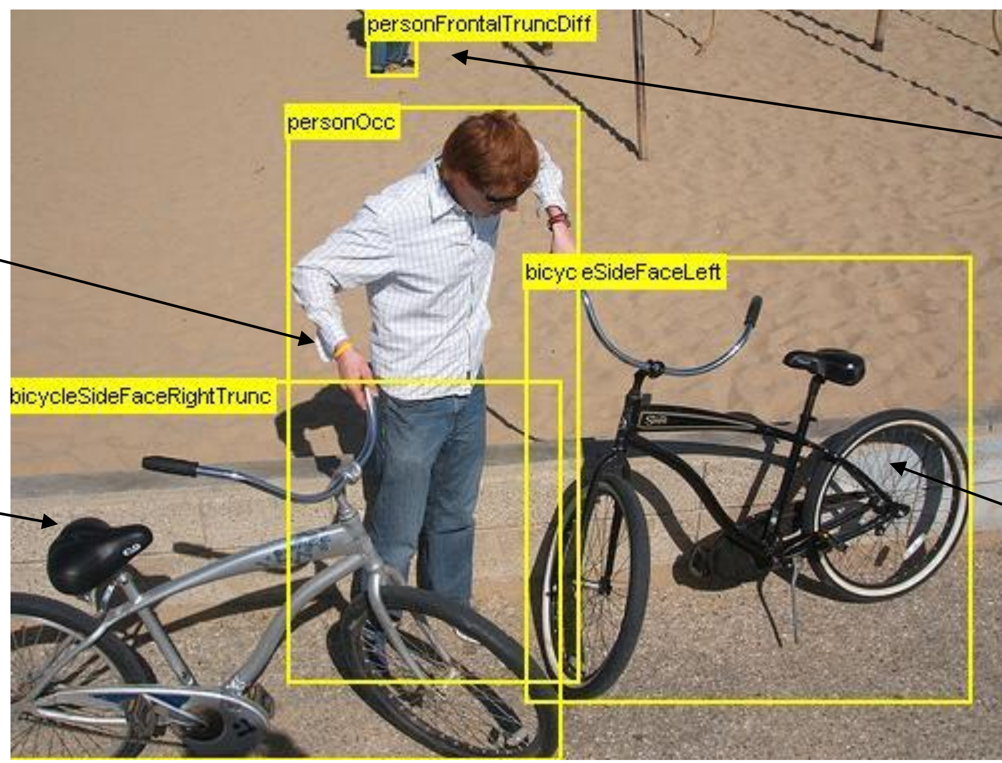
- Complex scenes, scale, pose, lighting, occlusion, ...

Annotation

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

Occluded
Object is significantly occluded within BB

Difficult
Not scored in evaluation



Truncated
Object extends beyond BB

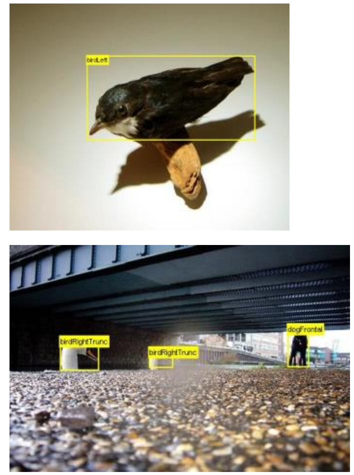
Pose
Facing left

Examples

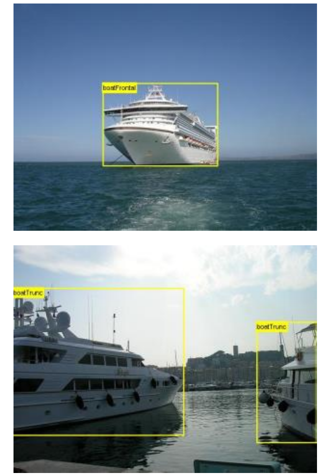
Aeroplane

Bicycle

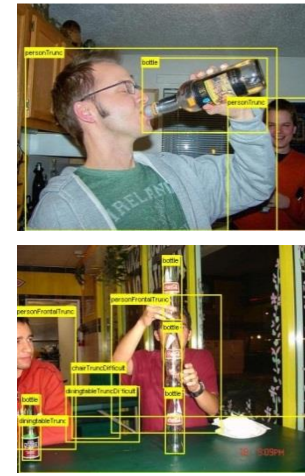
Bird



Boat



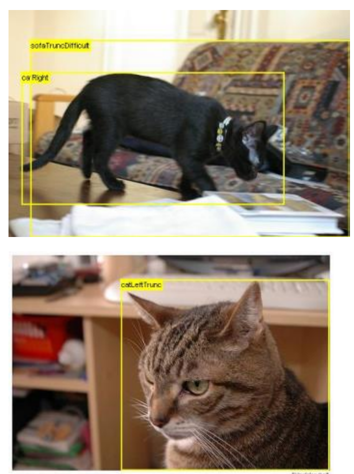
Bottle



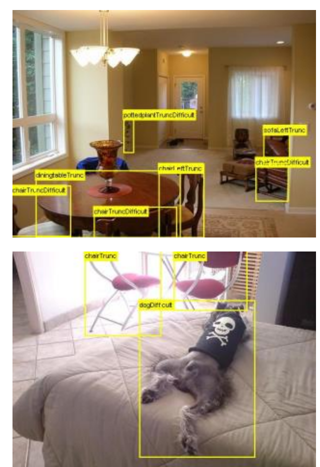
Bus

Car

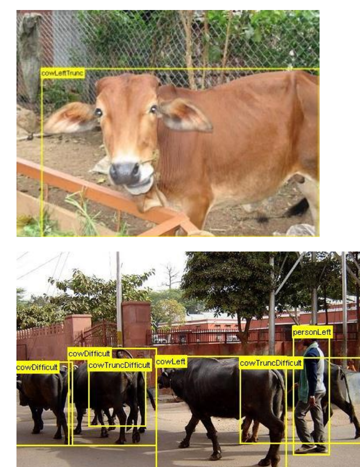
Cat



Chair



Cow



History

	Images	Objects	Classes	Entries	
2005	2,232	2,871	4	12	*Collection of existing and some new data.*
2006	5,304	9,507	10	25	*Completely new dataset from flickr (+MSRC)*
2007	9,963	24,640	20	28	*Increased classes to 20. Introduced tasters.*
2008	**8,776**	**20,739**	20		*Added "occlusion" flag. Reuse of taster data. Release detailed results to support "meta-analysis"*
2012					

- New dataset annotated annually
  - Annotation of test set is withheld until after challenge

Other recent datasets

ESP

[Ahn et al, 2006]

LabelMe

[Russell et al, 2005]

TinyImage

Torralba et al. 2007

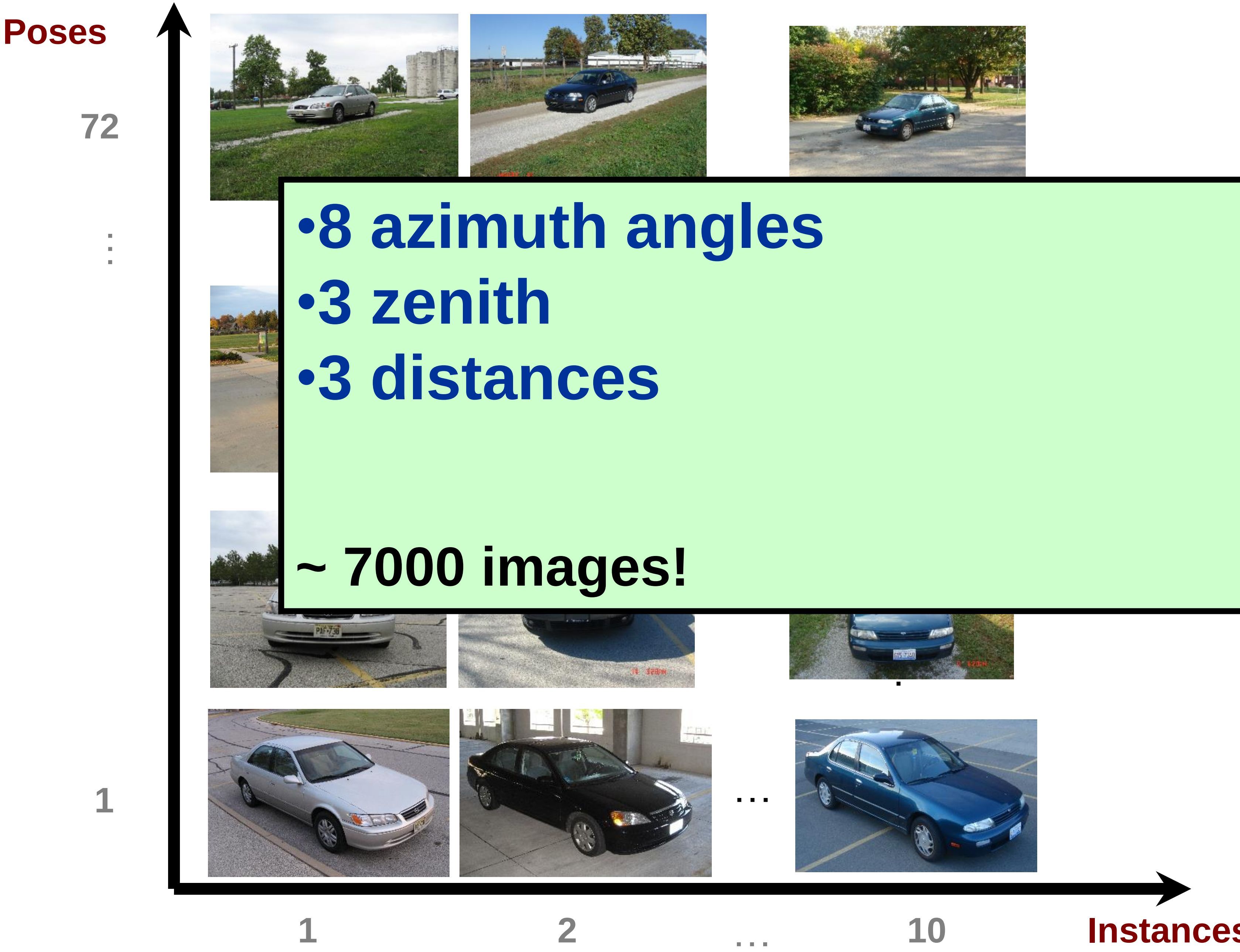
Lotus Hill

[Yao et al, 2007]

MSRC

[Shotton et al. 2006]

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



- 8 azimuth angles
- 3 zenith
- 3 distances

~ 7000 images!

Largest dataset for object categories up to date

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

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

http://www.image-net.org

IMAGENET

9,956,478 images, 14841 synsets indexed

Explore Download Challenge New! People Publication Sponsors About

Not logged in. Login | Signup

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.
Click Here to learn more about ImageNet, Click Here to join ImageNet mailing list.

SEARCH

What do these images have in common? *Find out!*

Update Notice: ImageNet 2010 Spring Version will be released in April, 2010

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))

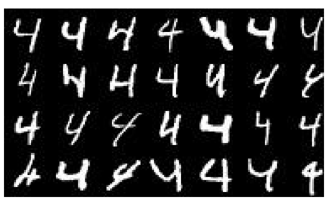
More Datasets....

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

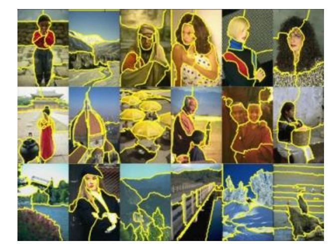
COIL Objects (1996)
S. Nene, S. Nayar, H. Murase



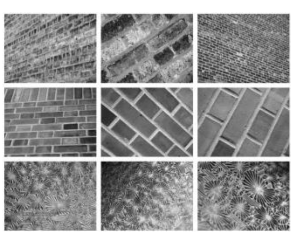
MNIST digits (1998-10)
Y LeCun & C. Cortes

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

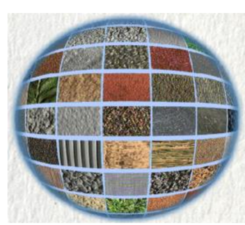
Sign Language (2008)
P. Buehler, M. Everingham, A. Zisserman



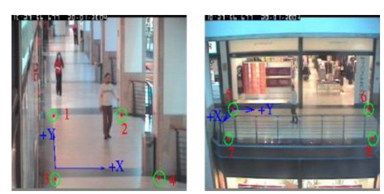
Segmentation (2001)
D. Martin, C. Fowlkes, D. Tal, J. Malik.



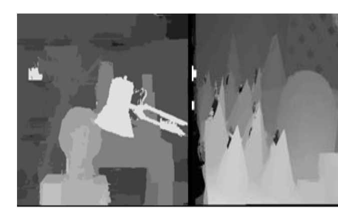
3D Textures (2005)
S. Lazebnik, C. Schmid, J. Ponce



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

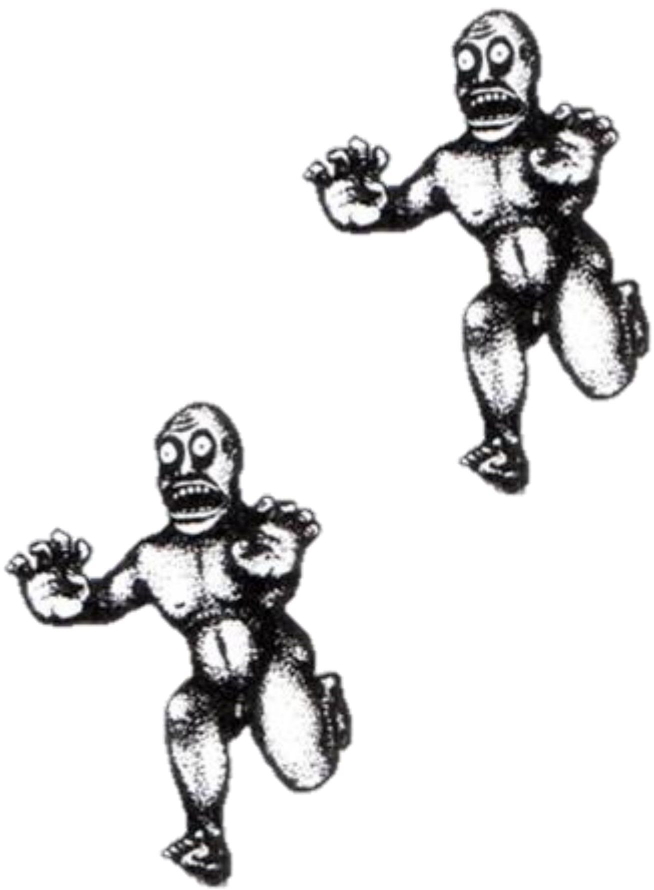
ESP game	www.espgame.org	Global image descriptions	Web images
LabelMe	people.csail.mit.edu/brussell/research/LabelMe/intro.html	Polygonal boundary	High resolution images

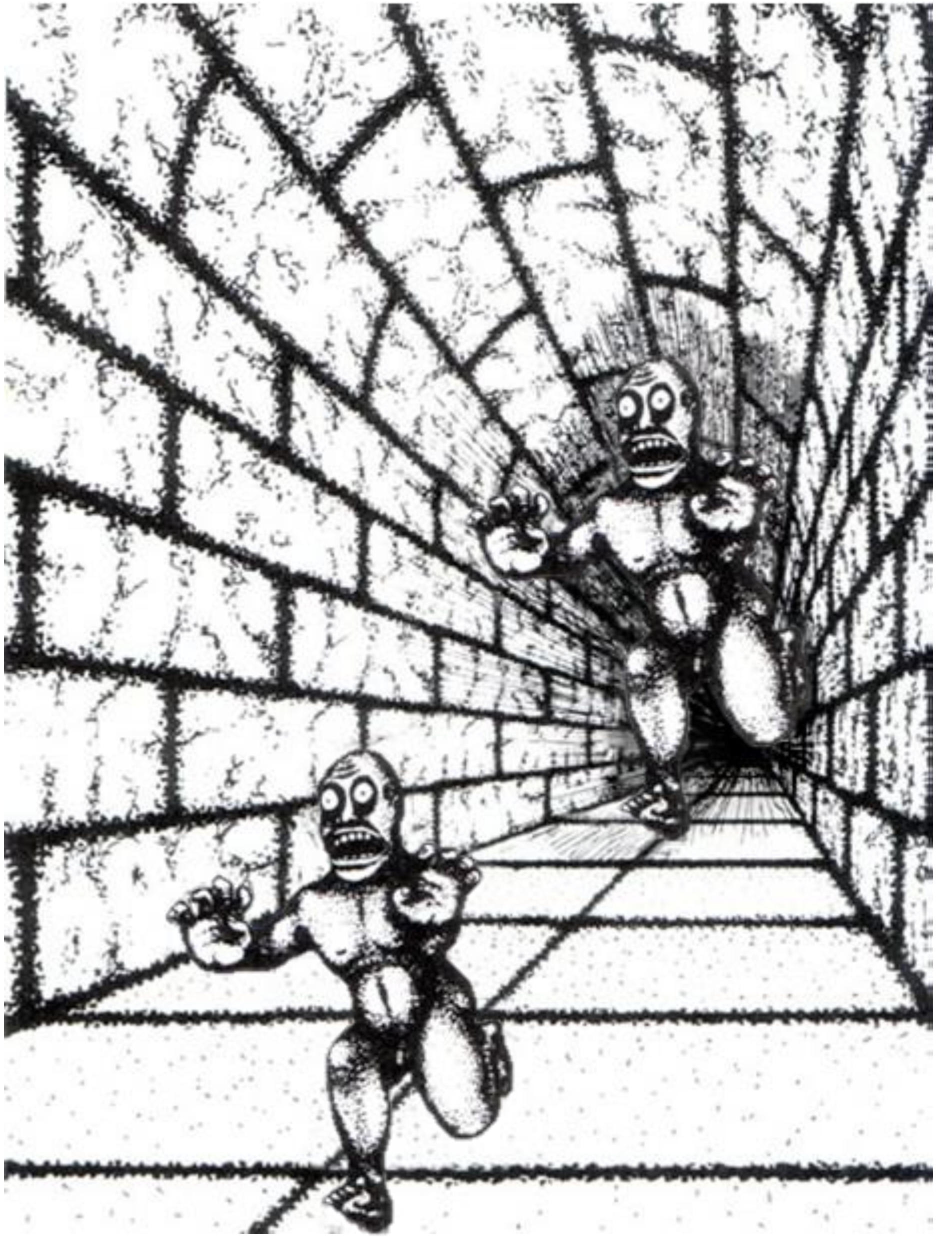
Collections

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

Lecture 16
Closure

- Datasets in computer vision
- 3D scene understanding



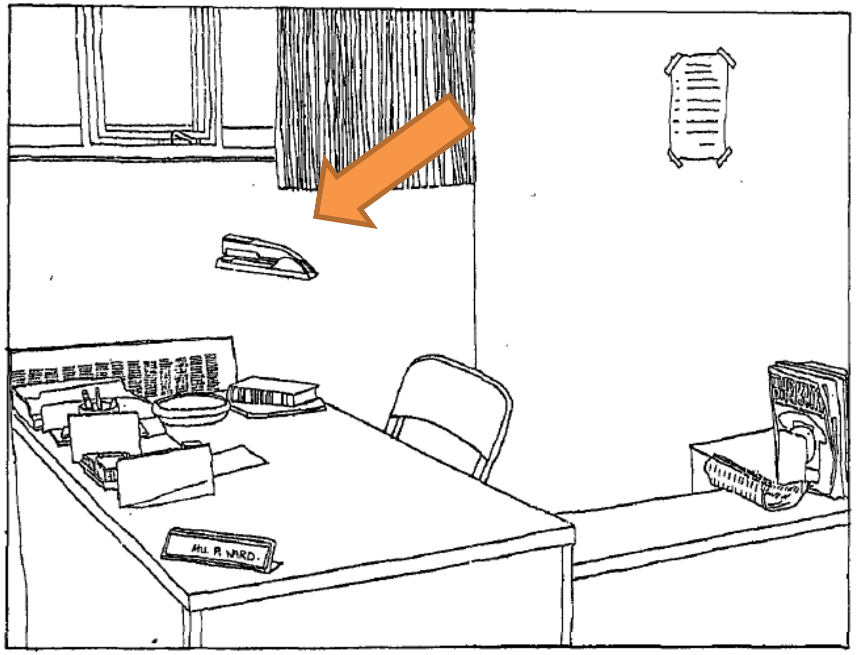


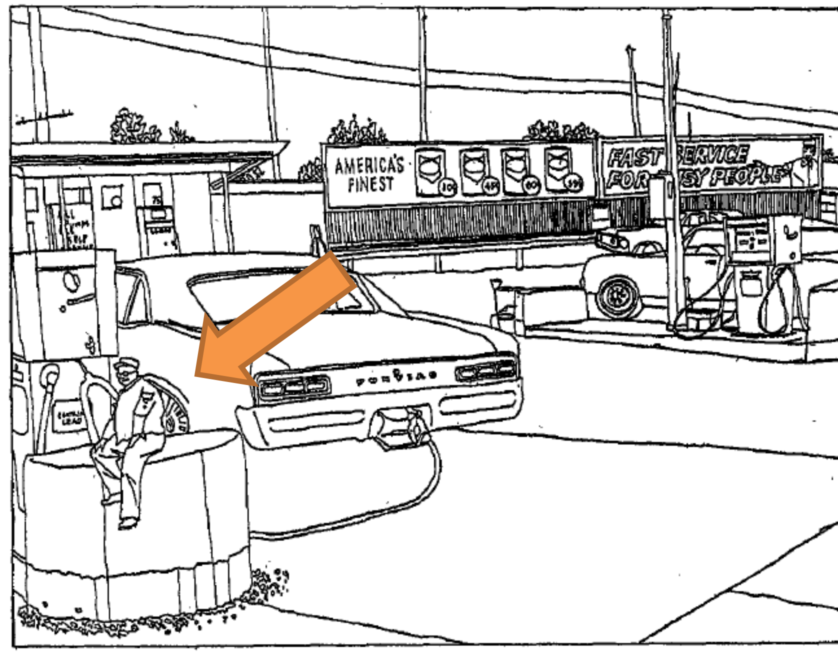
Objects are constrained by the 3D space

The 3D space is shaped by its objects

Modeling this interplay is critical for 3D perception!

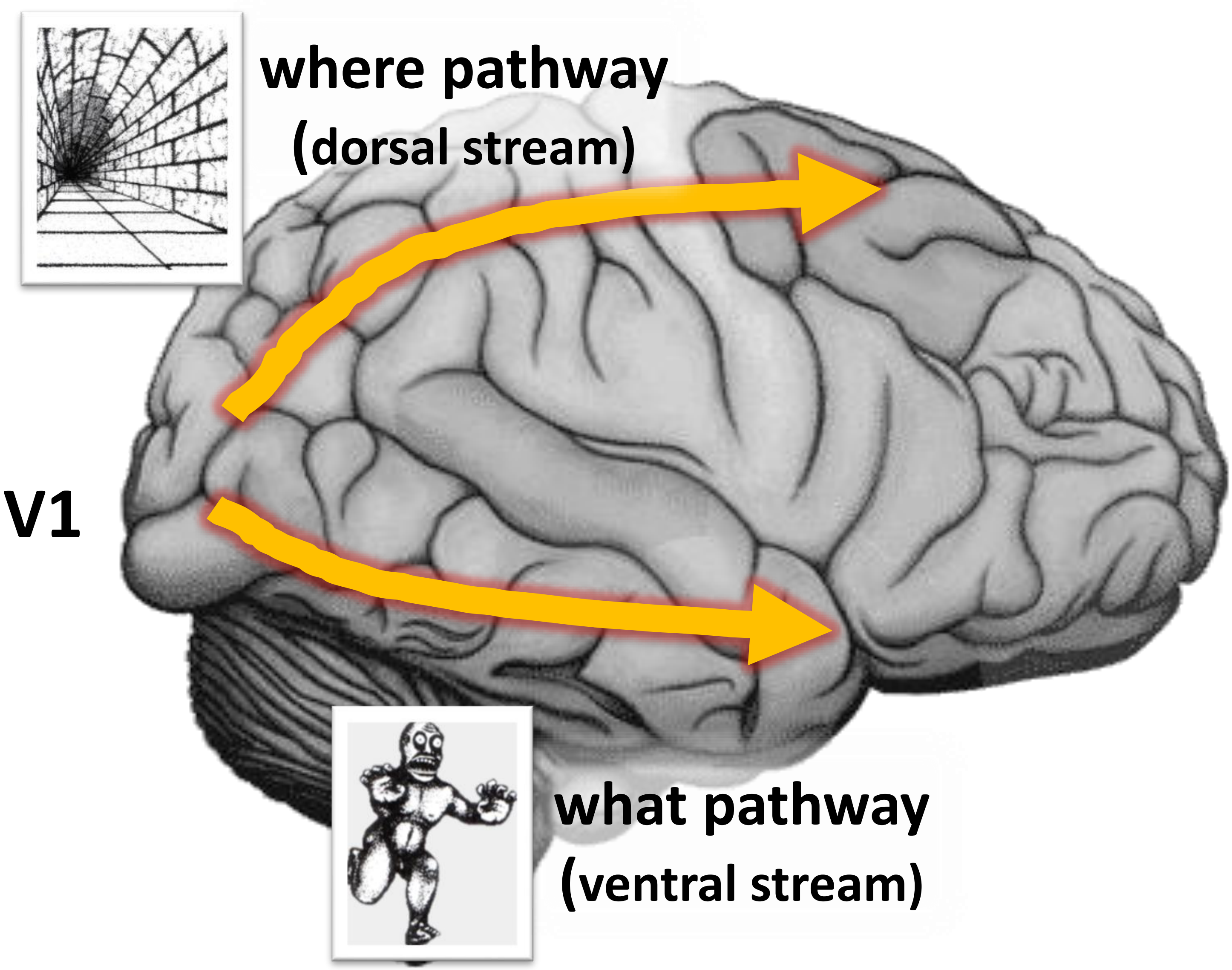
Humans perceive the world in 3D



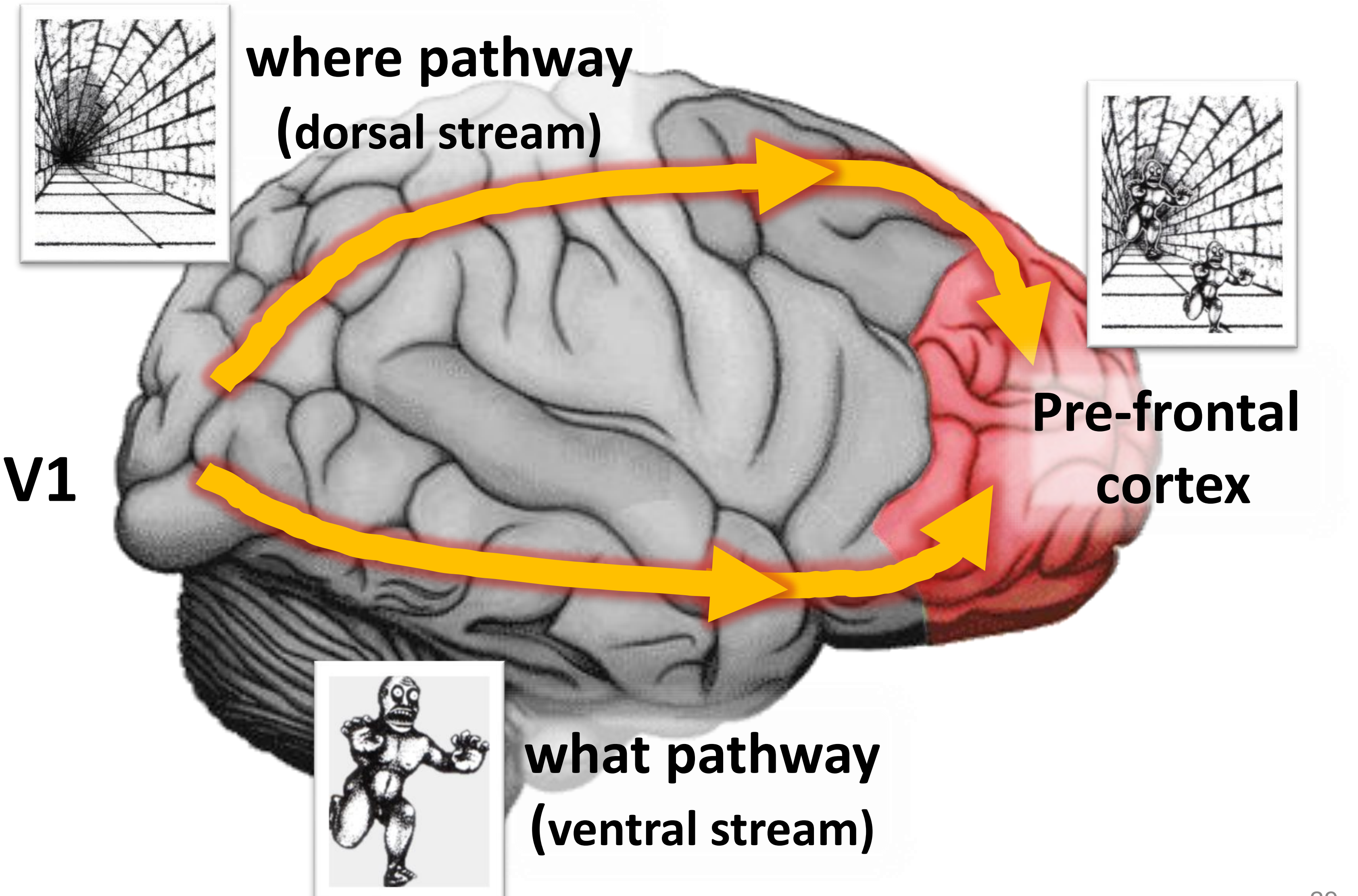


Biederman, Mezzanotte and Rabinowitz, 1982

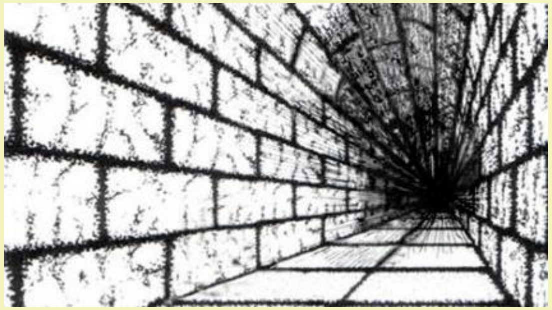
Visual processing in the brain



Visual processing in the brain



Current state of computer vision



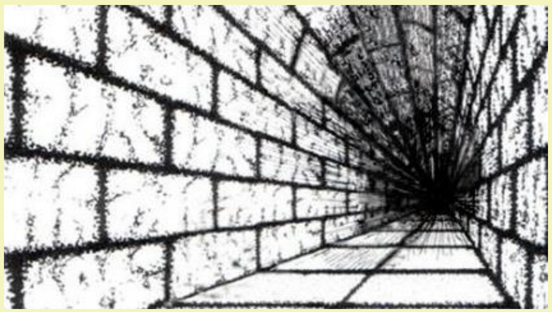
3D Reconstruction

- 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

Microsoft Photosynth

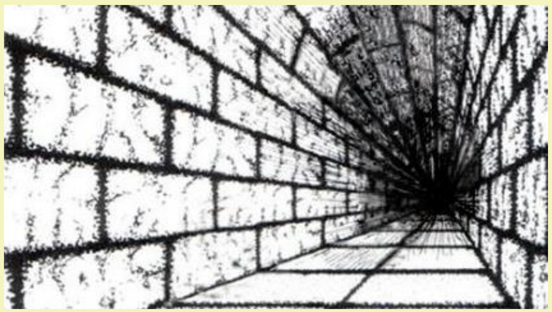
Snavely et al., 06-08

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

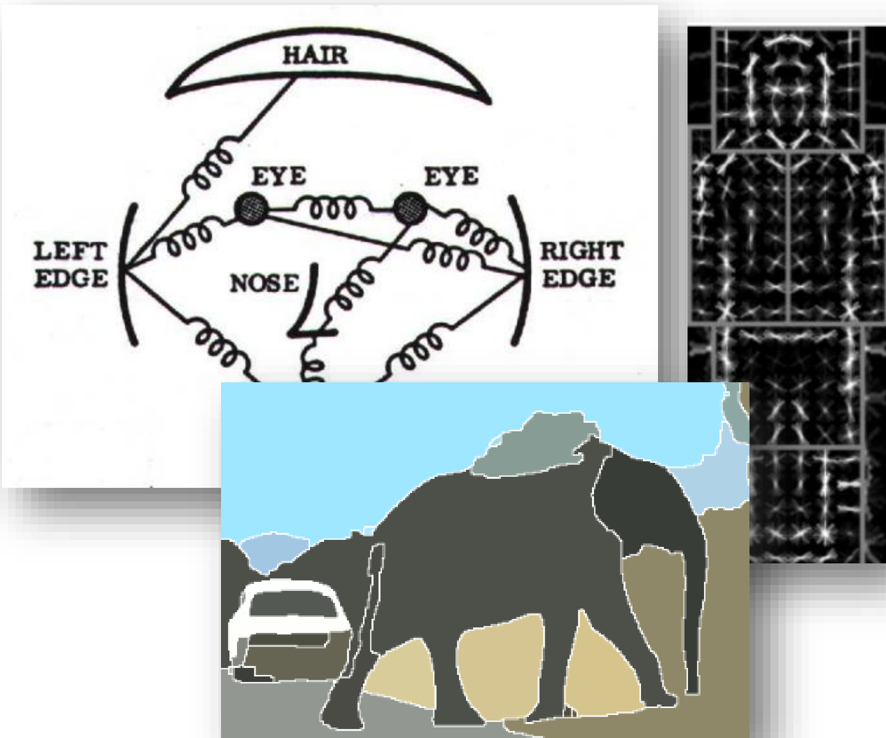
Snavely et al., 06-08

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

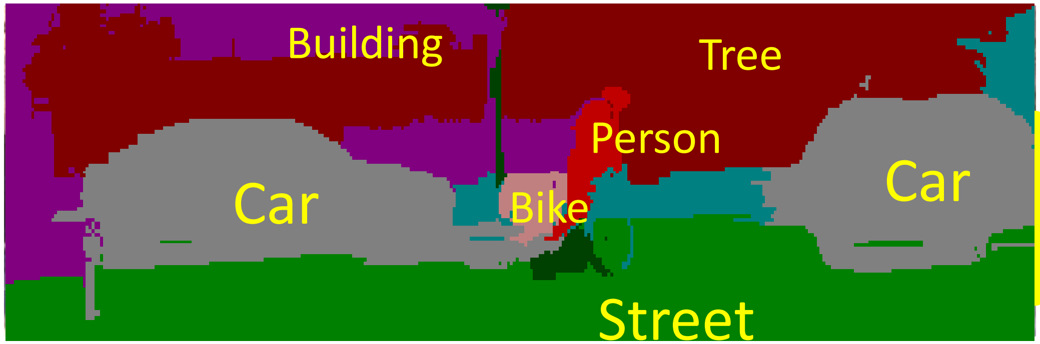
- Object detection
- Texture classification
- Target tracking
- Activity recognition

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

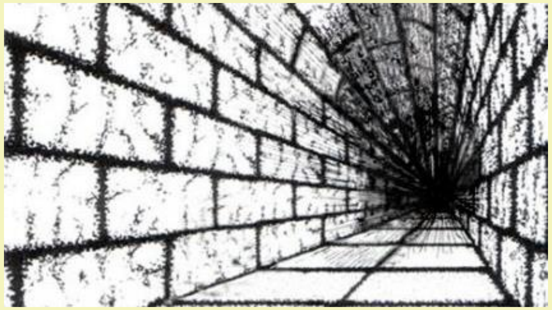
- Object detection
- Texture classification
- Target tracking
- Activity recognition

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

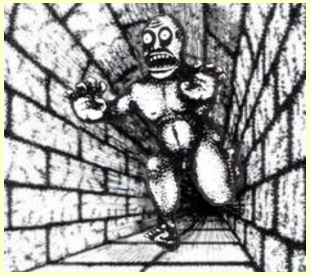
He et al. 06
Gould et al. 08
Maire et al. 08
Felzenszwalb et al., 08
Kohli et al. 09
L.-J. Li et al. 09
Ladicky et al. 10,11
Gonfaus et al. 10
Farhadi et al., 09
Lampert et al., 09

Current state of computer vision



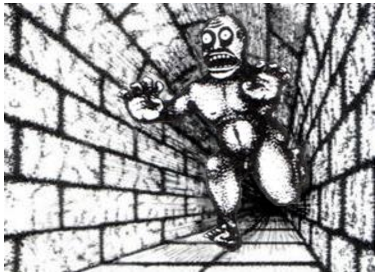
3D reconstruction

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



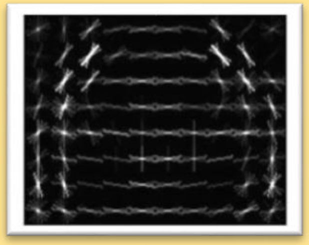
Outline

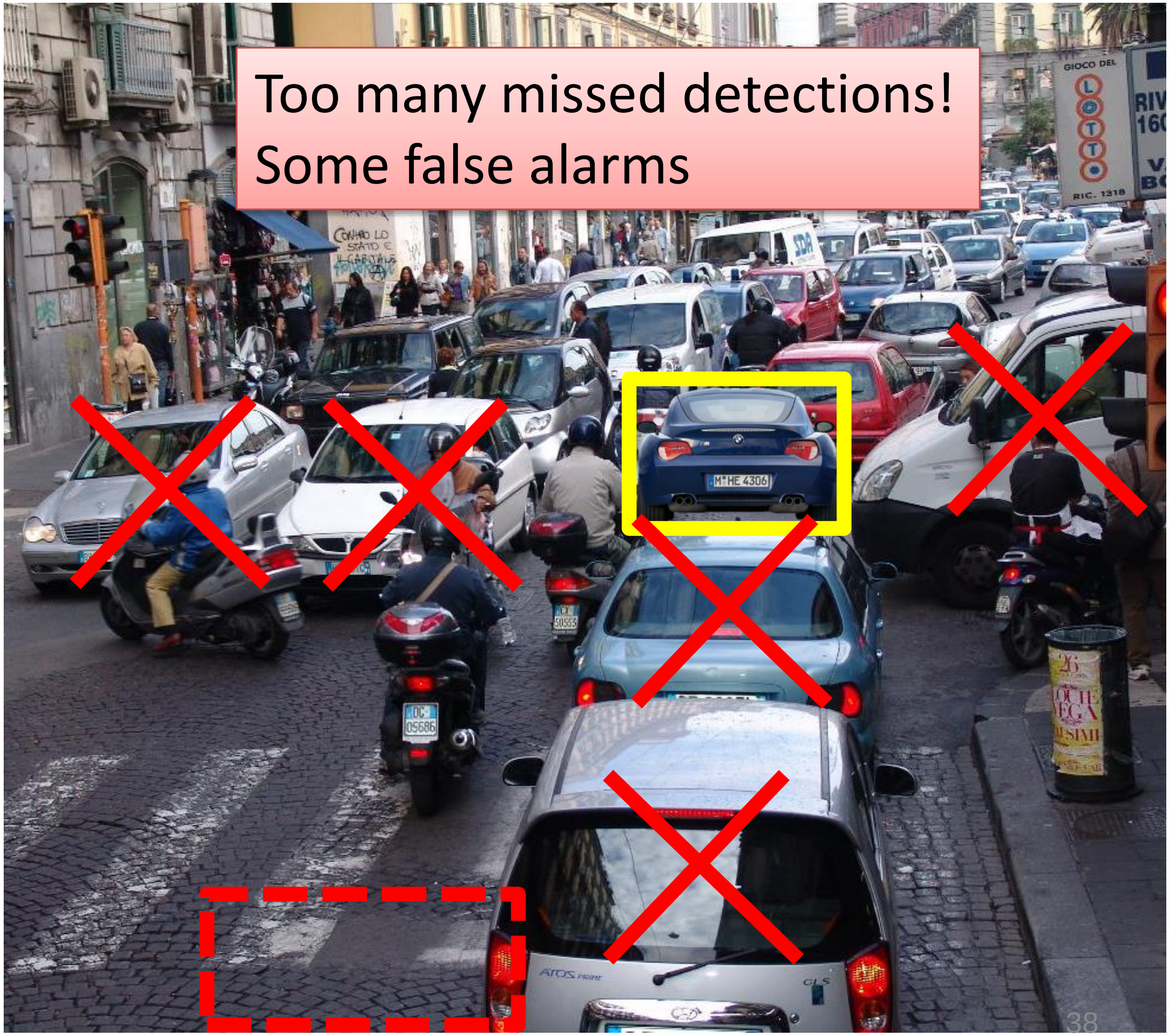
- Modeling objects and their 3D properties
- Modeling interaction among objects and space
- Modeling relationships of objects across views

Modeling objects and their 3D properties

State of the art object detection

"Car" model

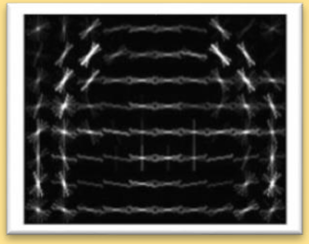




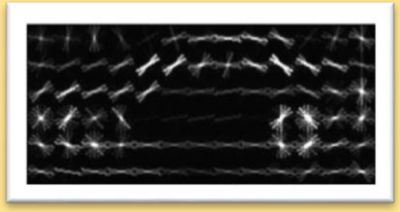
Turk & Pentland, 91
Poggio et al., 93
LeCun et al. 98
Amit and Geman, 99
Shi & Malik, 00
Viola & Jones, 00
Vasconcelos '00
Felzenszwalb & Huttenlocher 00
Belongie & Malik, 02
Ullman et al. 02
Argawal & Roth, 02
Weber et al., 00
Fergus et al., 03
Torralba et al., 03
Fei-Fei et al., 04
Leibe et al., 04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

State of the art object detection

“Car” model

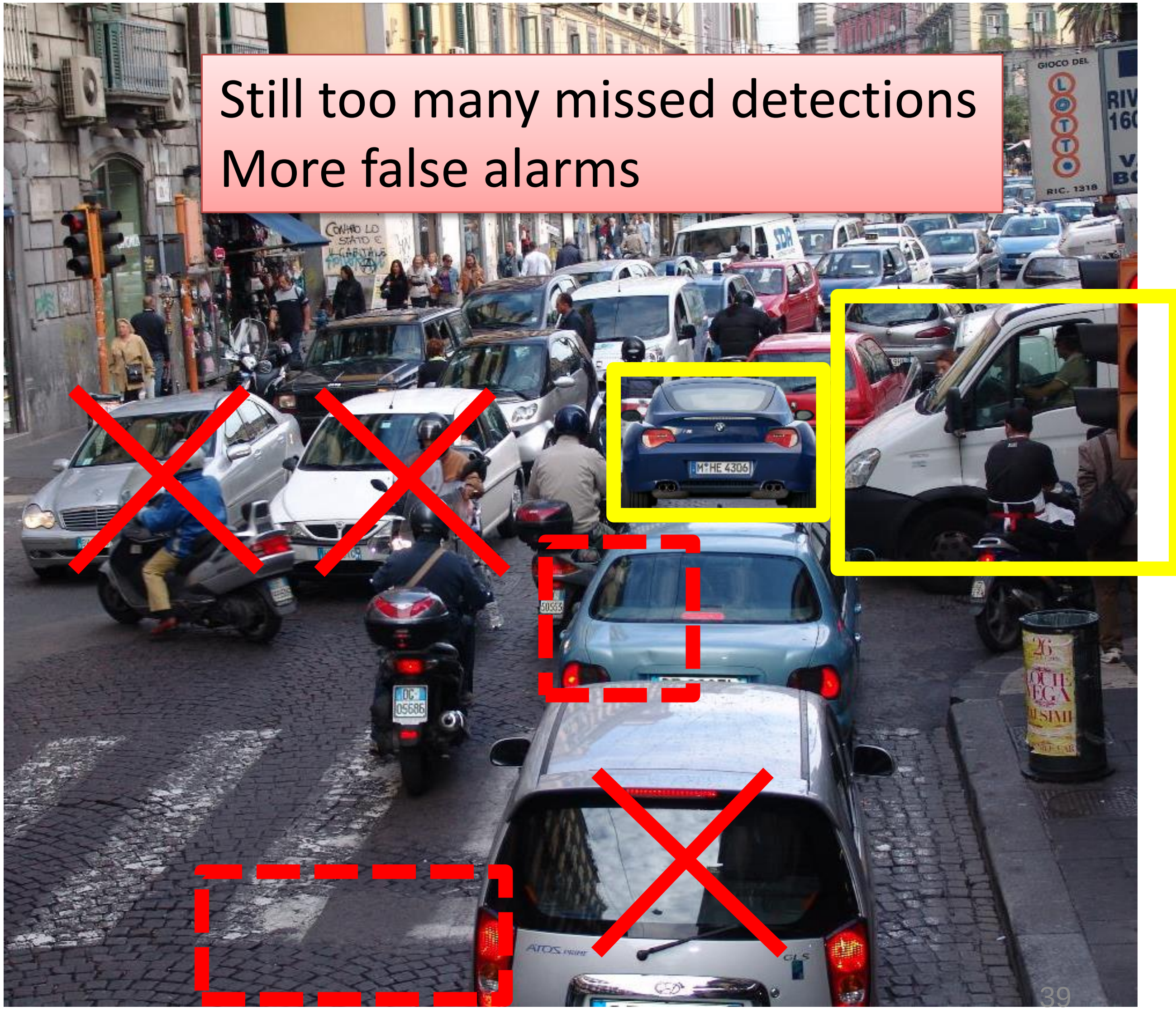


back



side

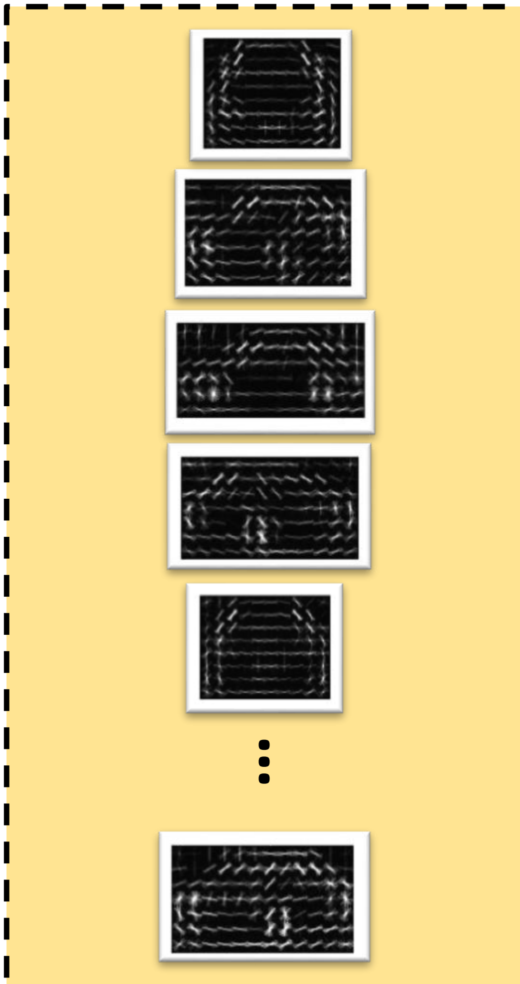
mixture model



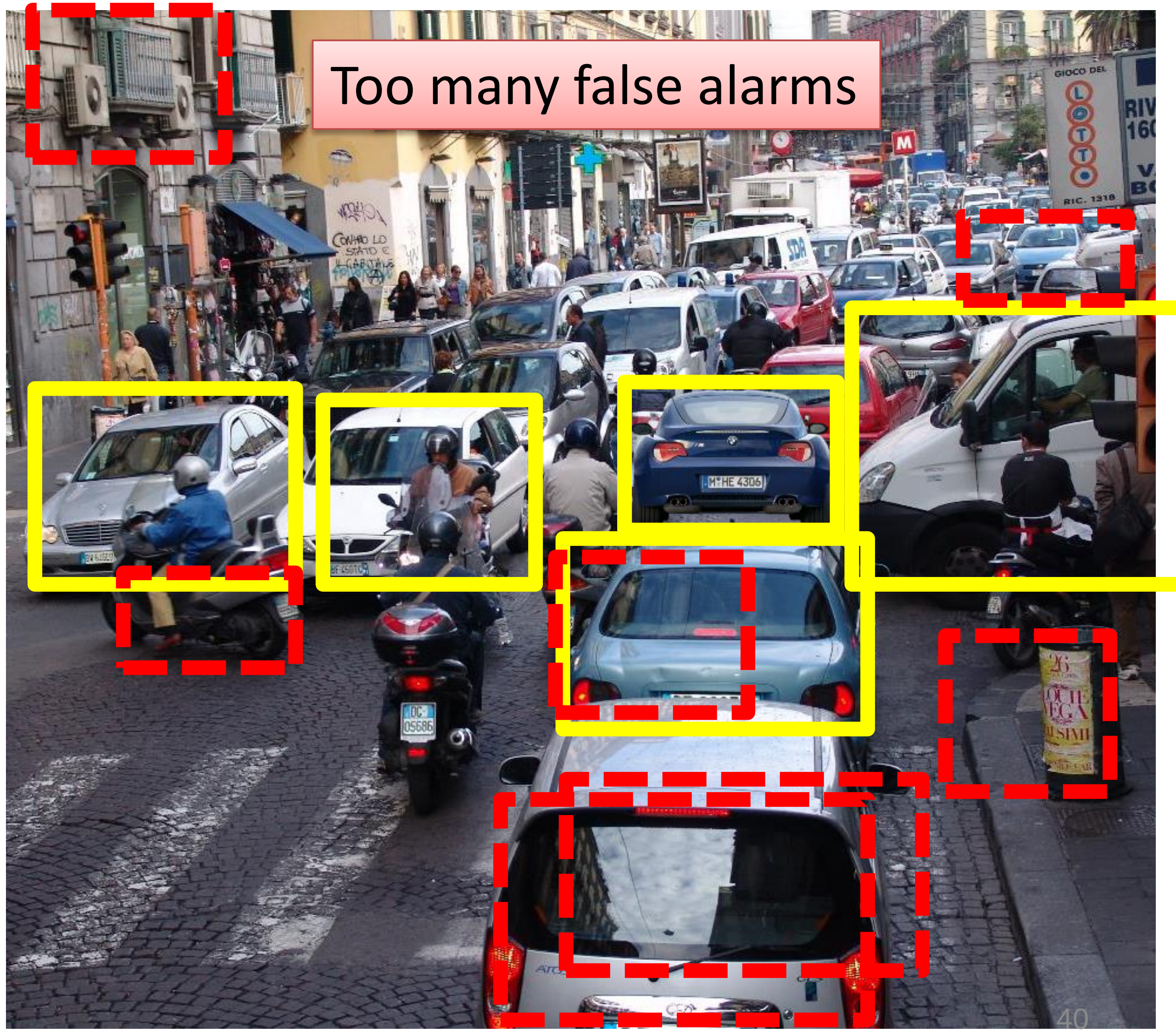
Weber et al. ’00
Schneiderman et al. ’01
Ullman et al. 02
Felzenszwalb et al., 08
Gu & Ren, ‘10

State of the art object detection

“Car” model

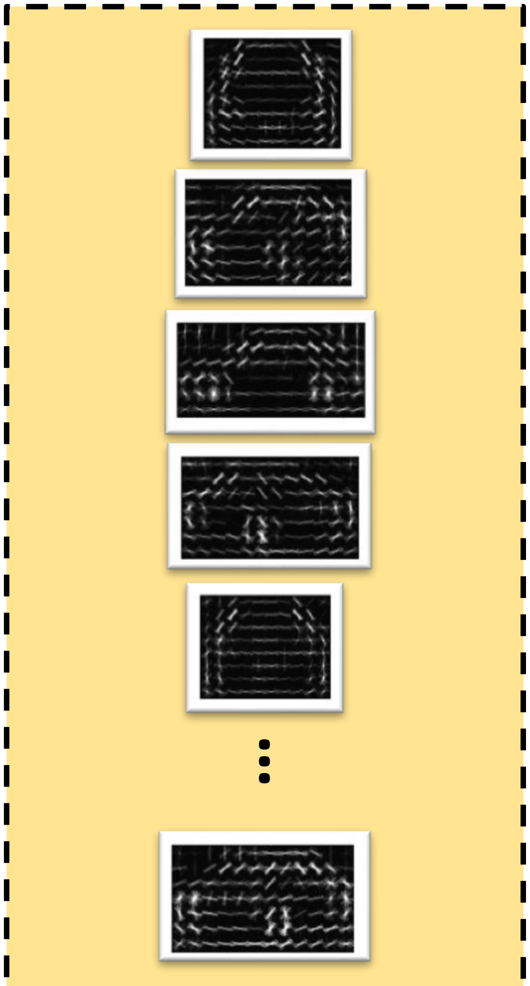


mixture model

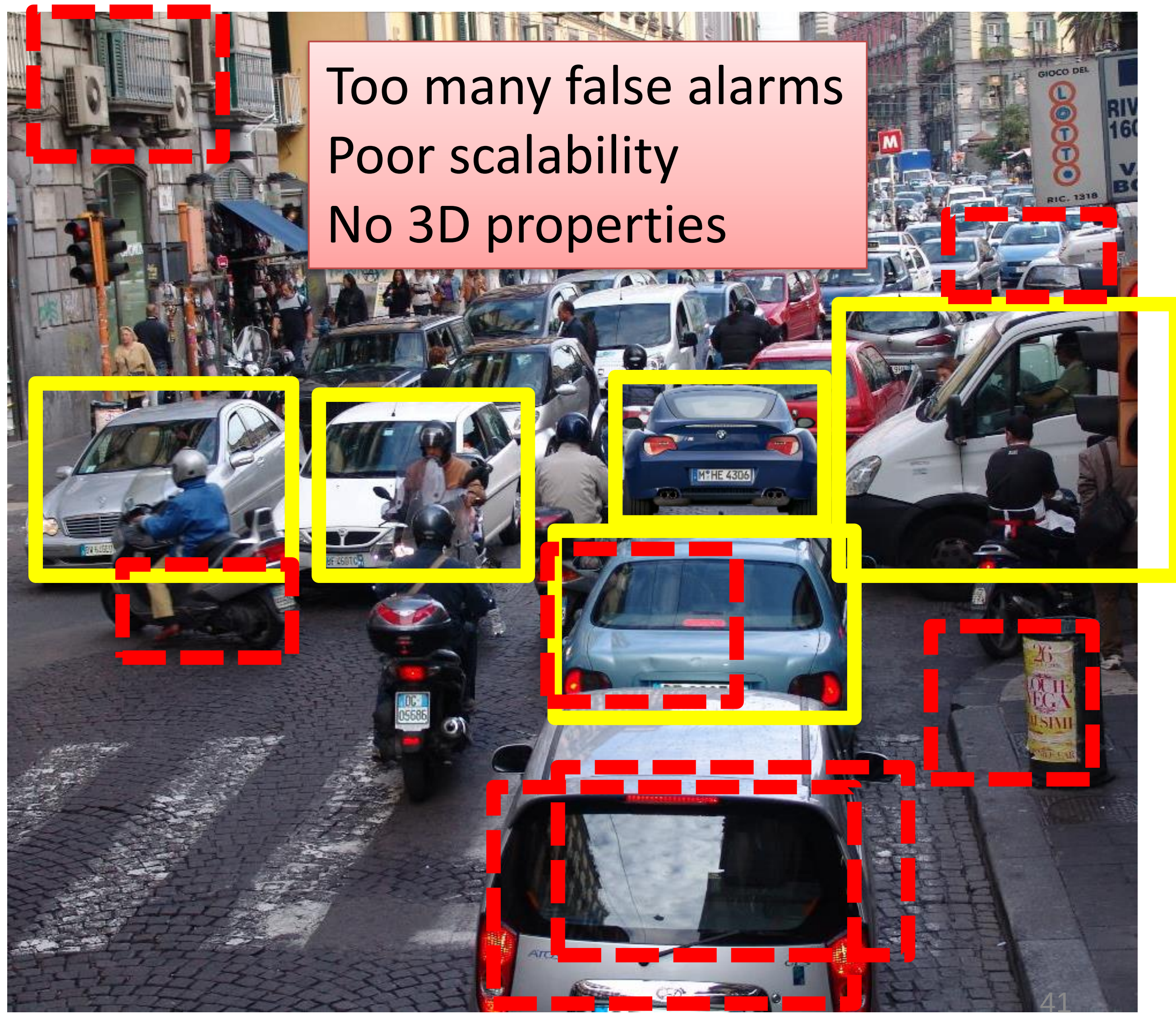


State of the art object detection

“Car” model

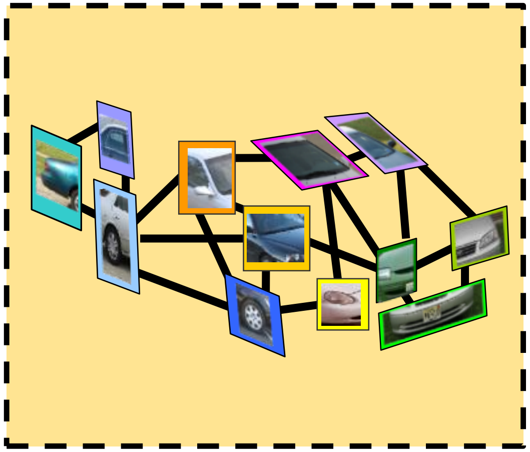


mixture model



3D object detection

"Car" model



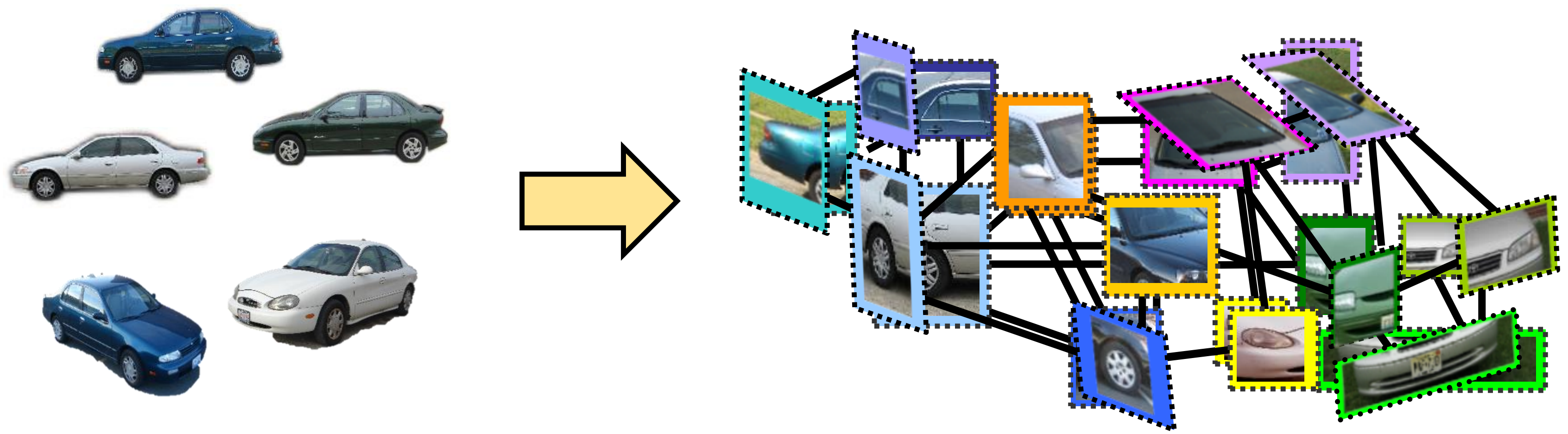
Savarese et al., ICCV 07
Su et al., ICCV 2009
Sun, et al., CVPR 2009
Yu & Savarese, CVPR 2012

- 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

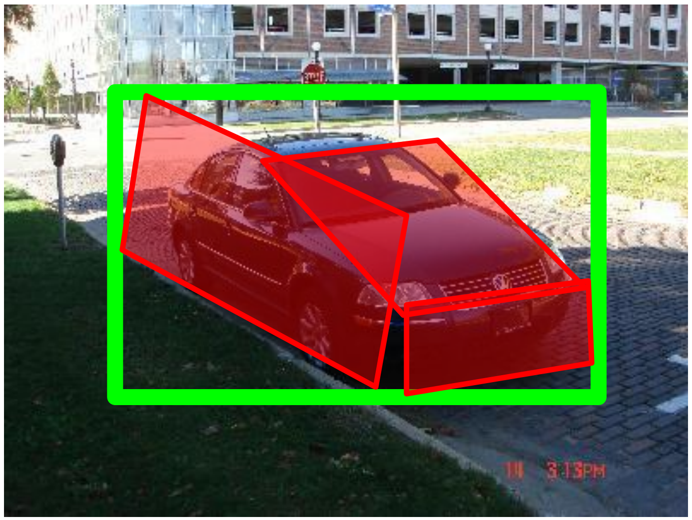
Savarese et al., ICCV 2007

Su et al., ICCV 2009
Sun, et al., CVPR 2009

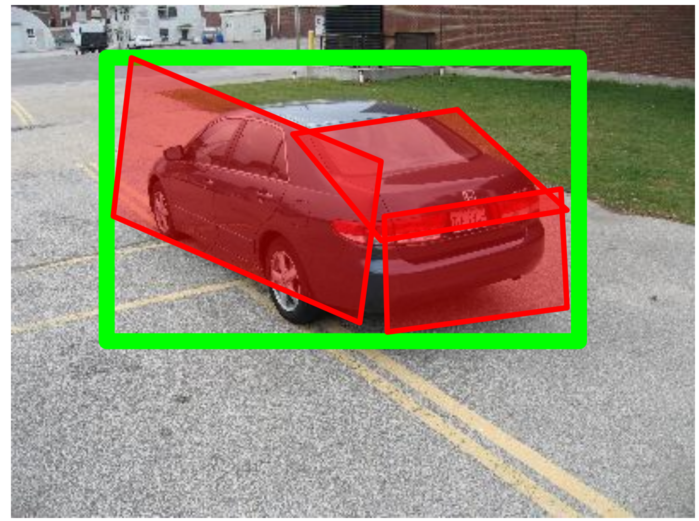
Yu & Savarese, CVPR 2012

Results

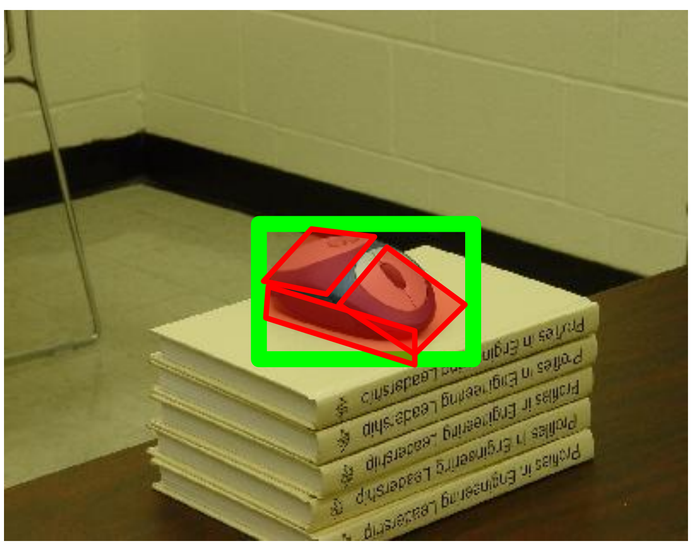
CAR a=330 e=15 d=7



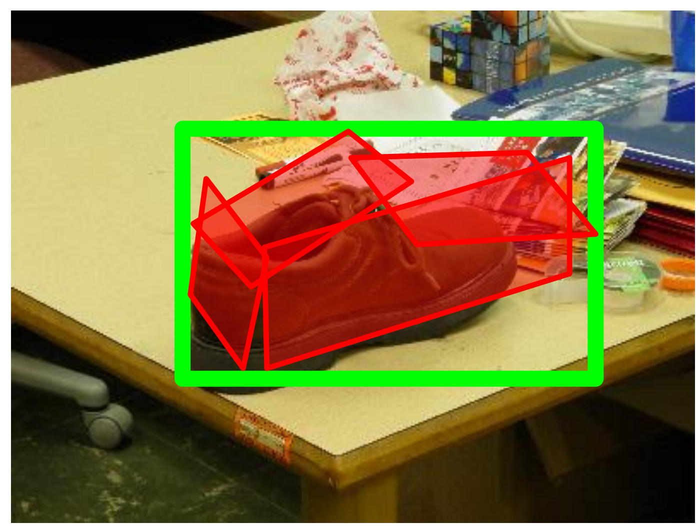
CAR a=150 e=15 d=7



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

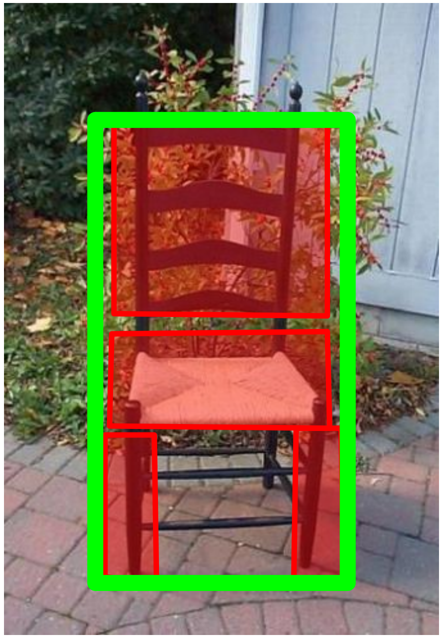
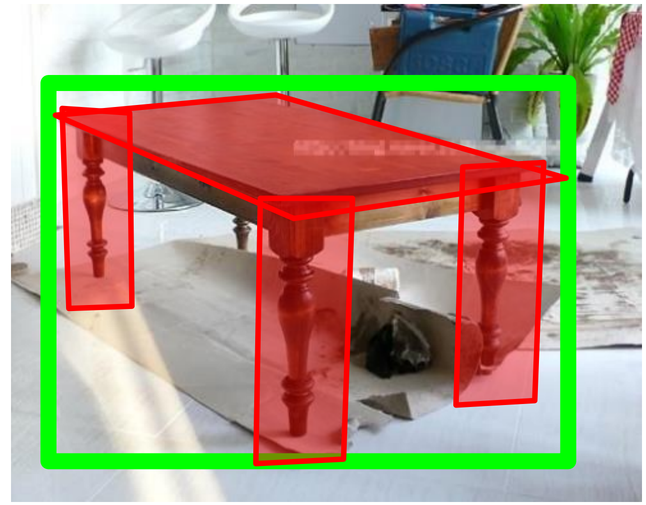


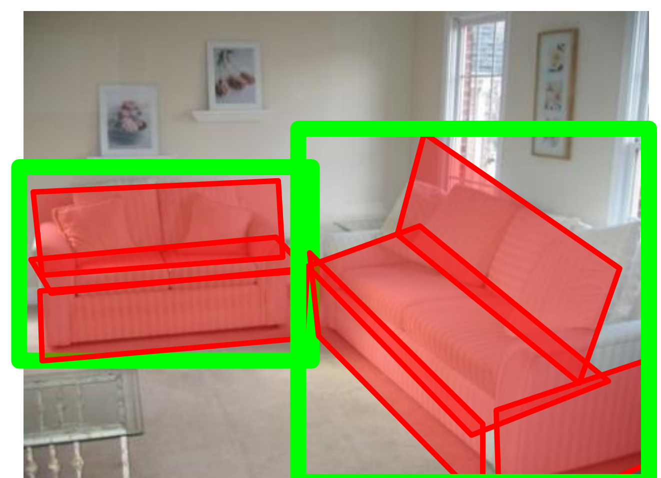
TABLE a=60 e=15 d=2



BED a=30 e=15 d=2.5

SOFA a=345 e=15 d=3.5

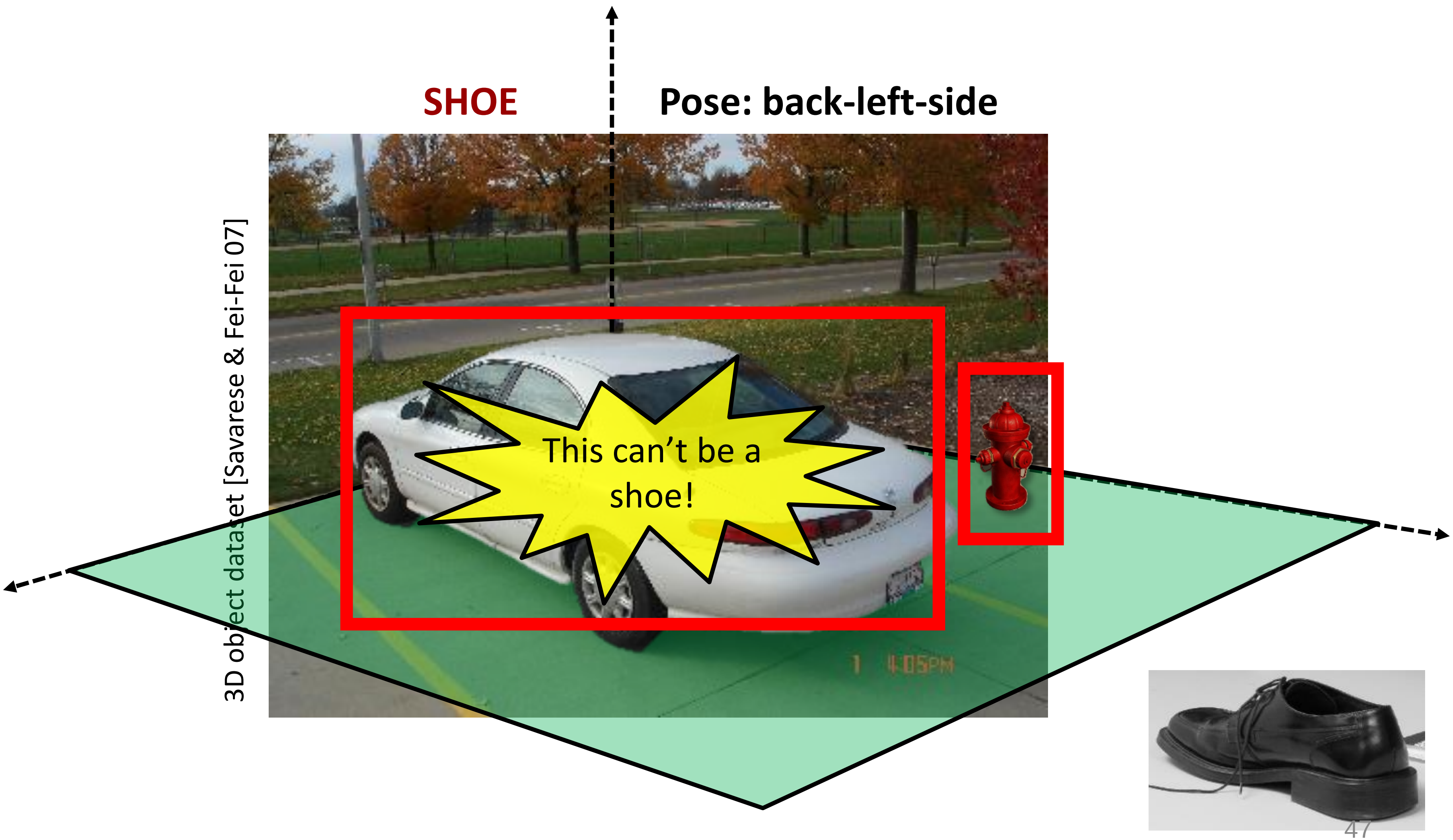
a=60 e-30 d=2.5

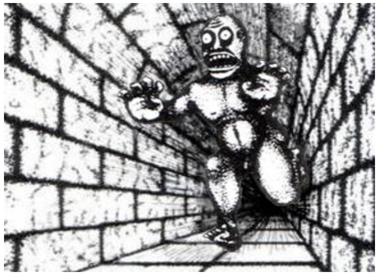


ImageNet dataset [Deng et al. 2010]

Results

Examples of failure (wrong category)

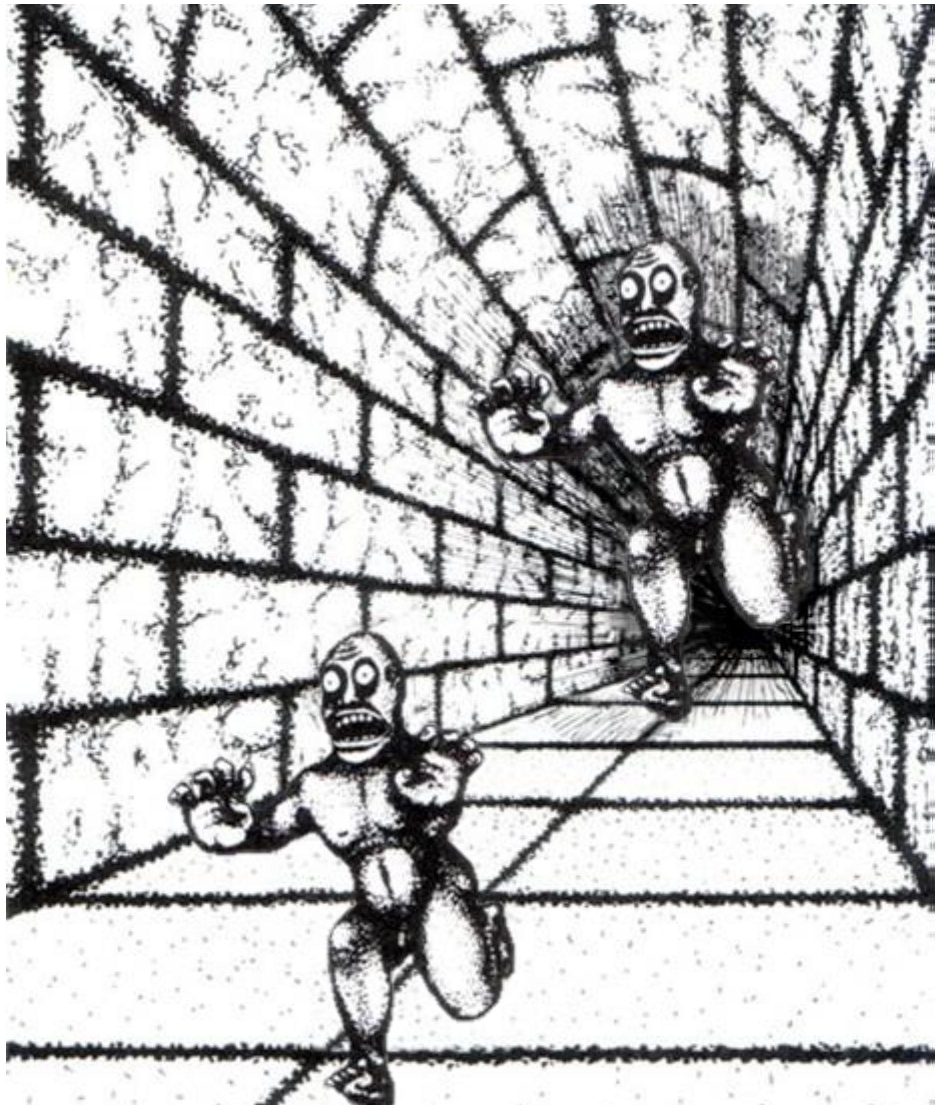




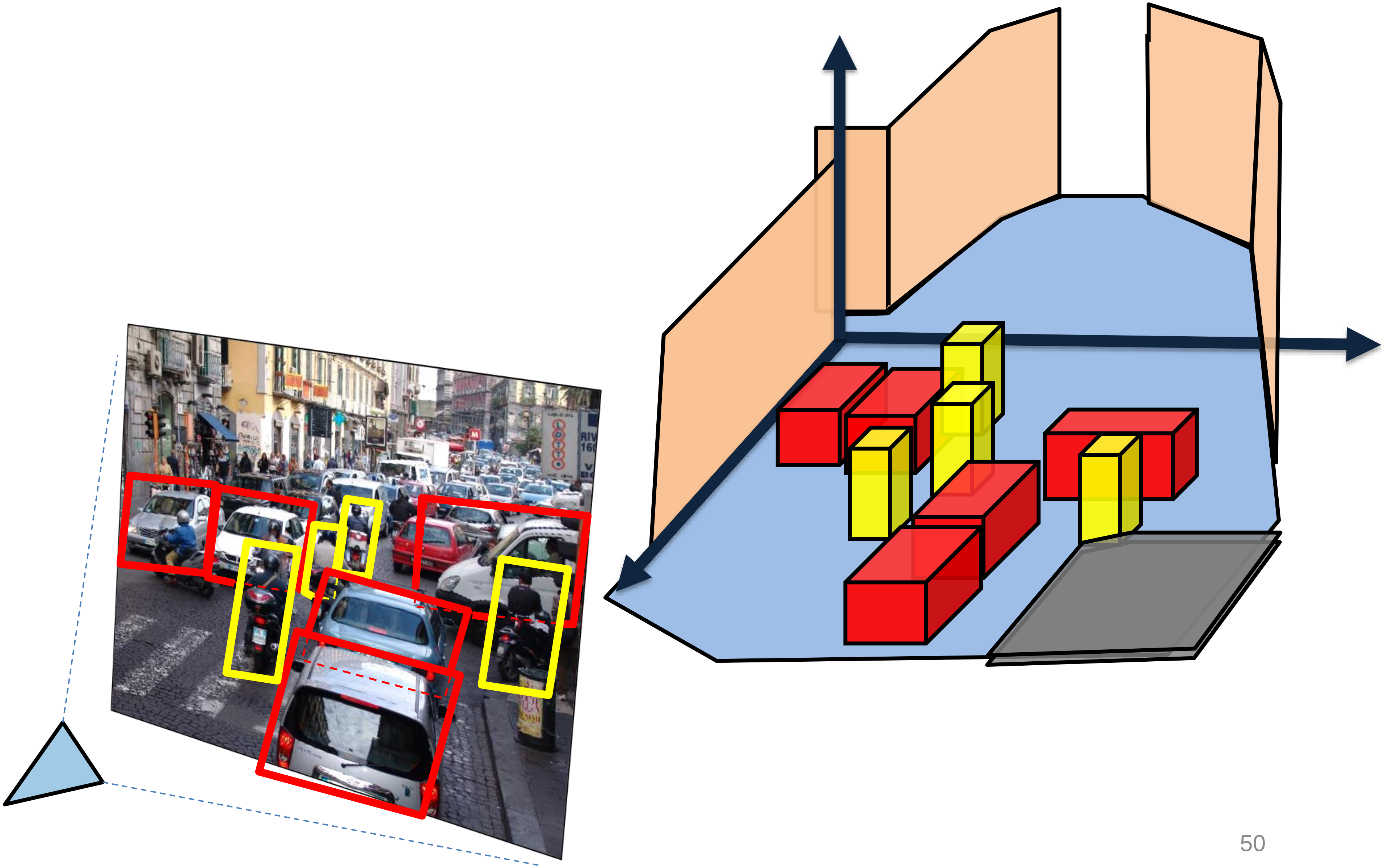
Outline

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

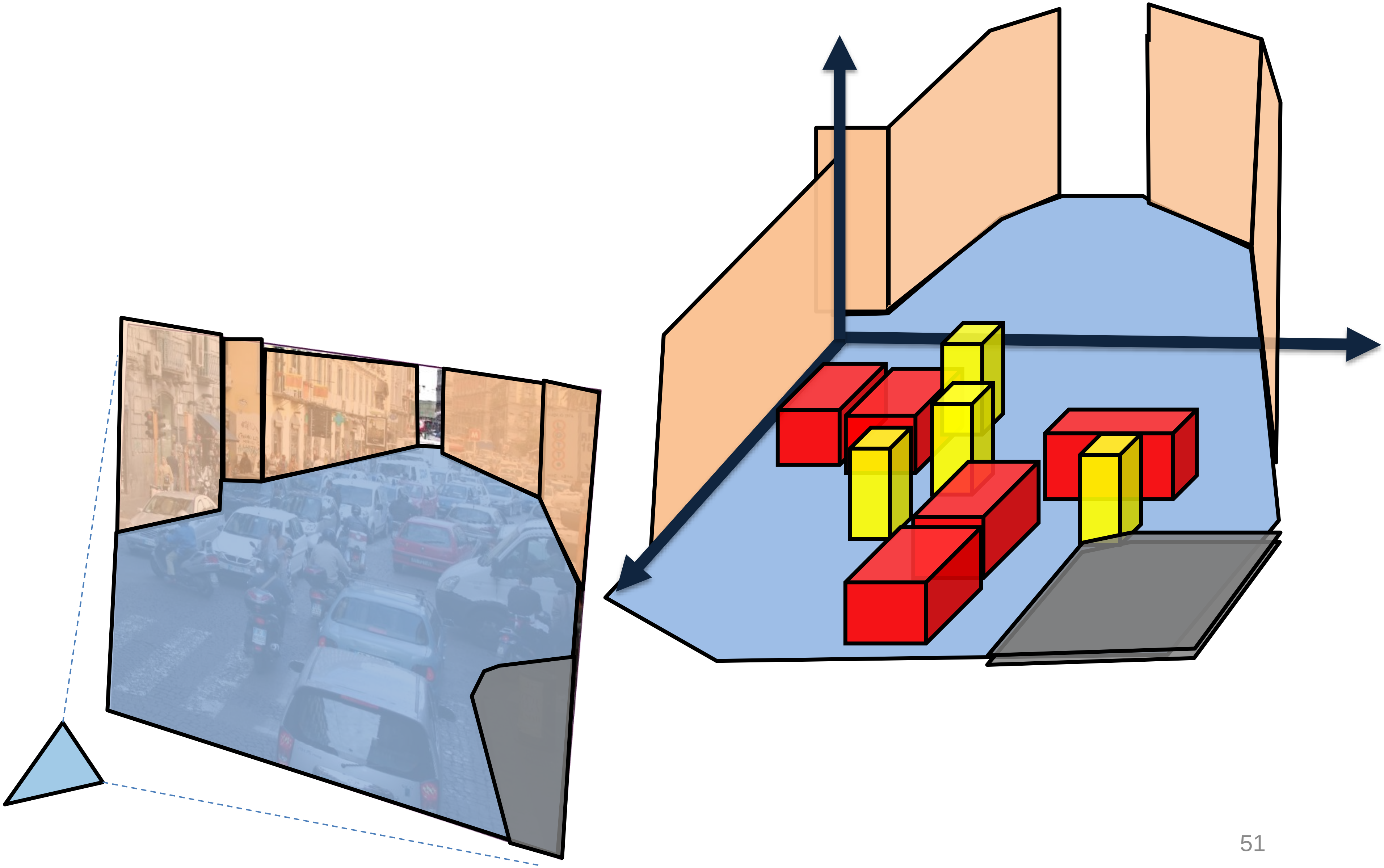
Scene understanding is an interplay between objects and space



3D space is shaped by its objects

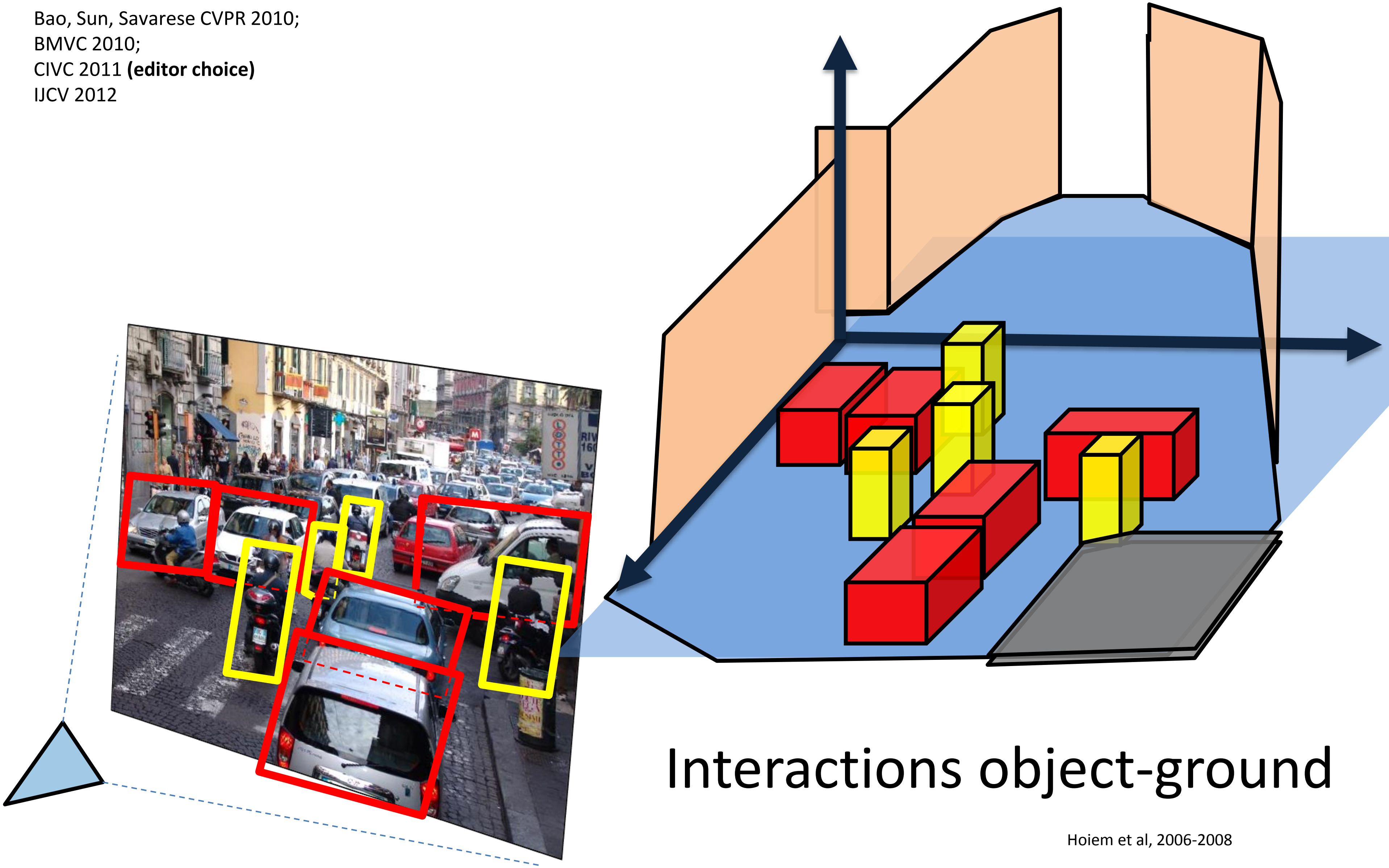


Objects are placed into 3D space



A first attempt….

Bao, Sun, Savarese CVPR 2010;
BMVC 2010;
CIVC 2011 **(editor choice)**
IJCV 2012

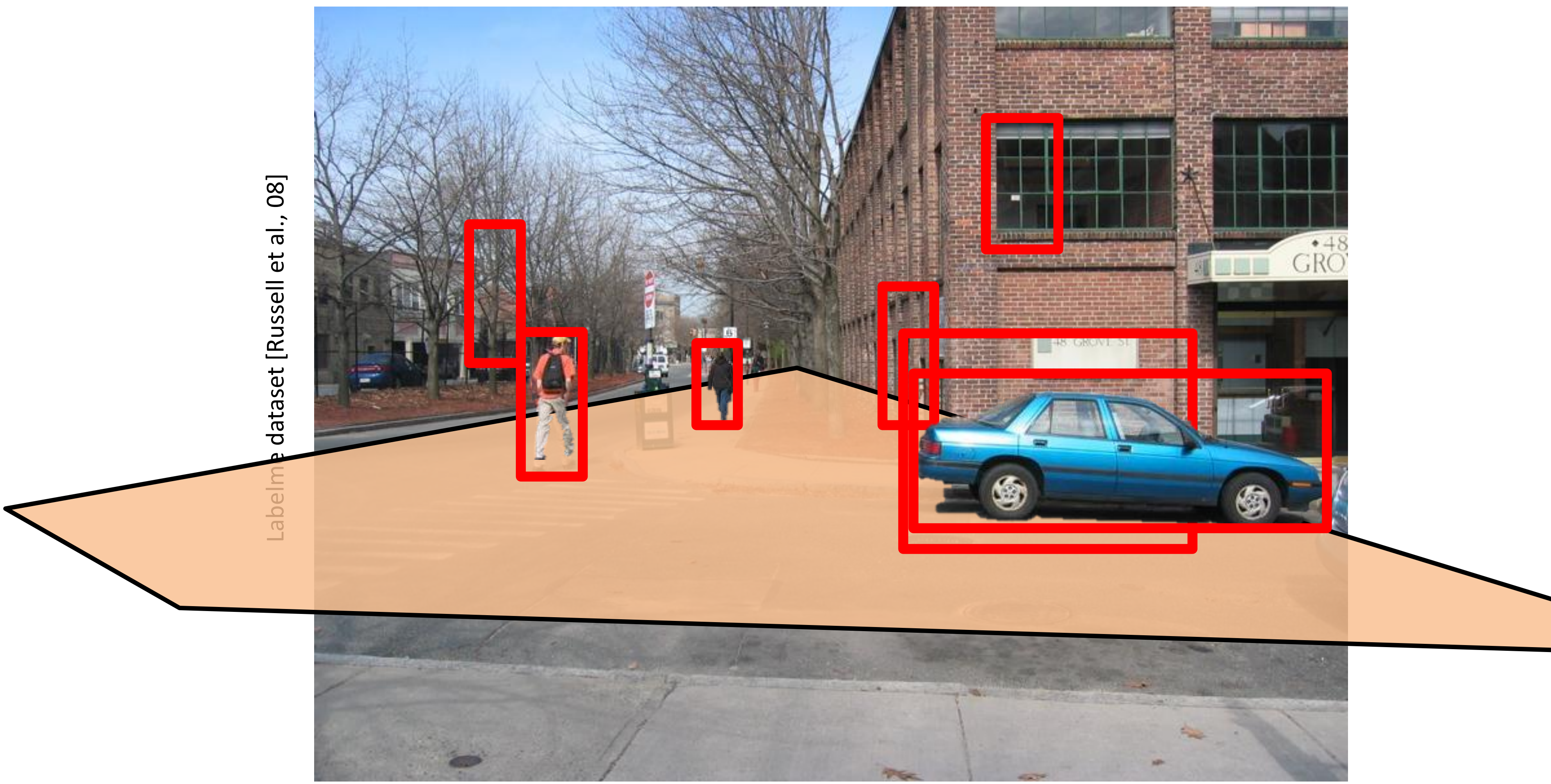


Interactions object-ground

Hoiem et al, 2006-2008

A first attempt....

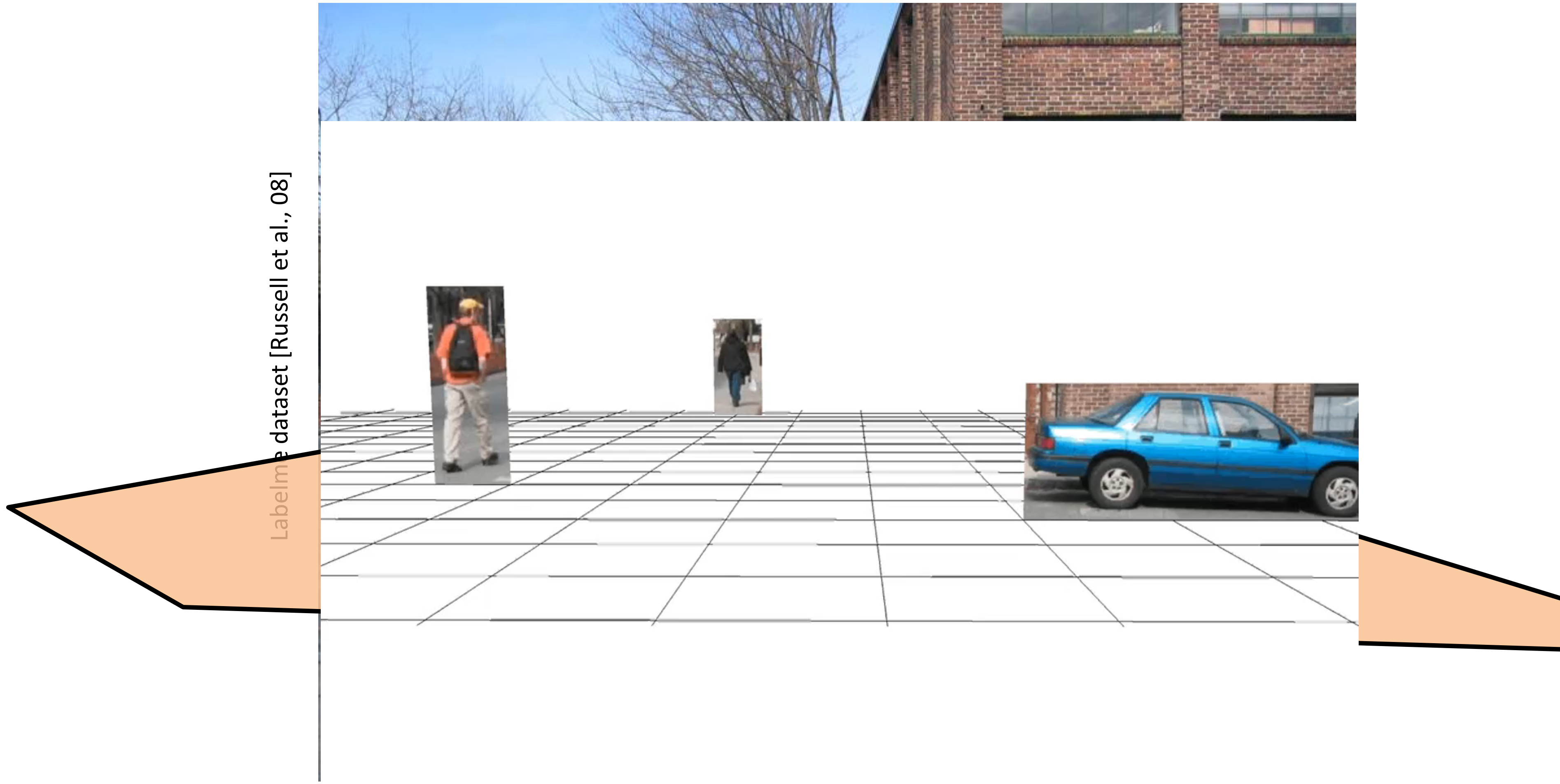
Bao et al. CVPR 2010;
BMVC 2010;
CIVC 2011 **(editor choice)**
IJCV 2012



LabelMe dataset [Russell et al., 08]

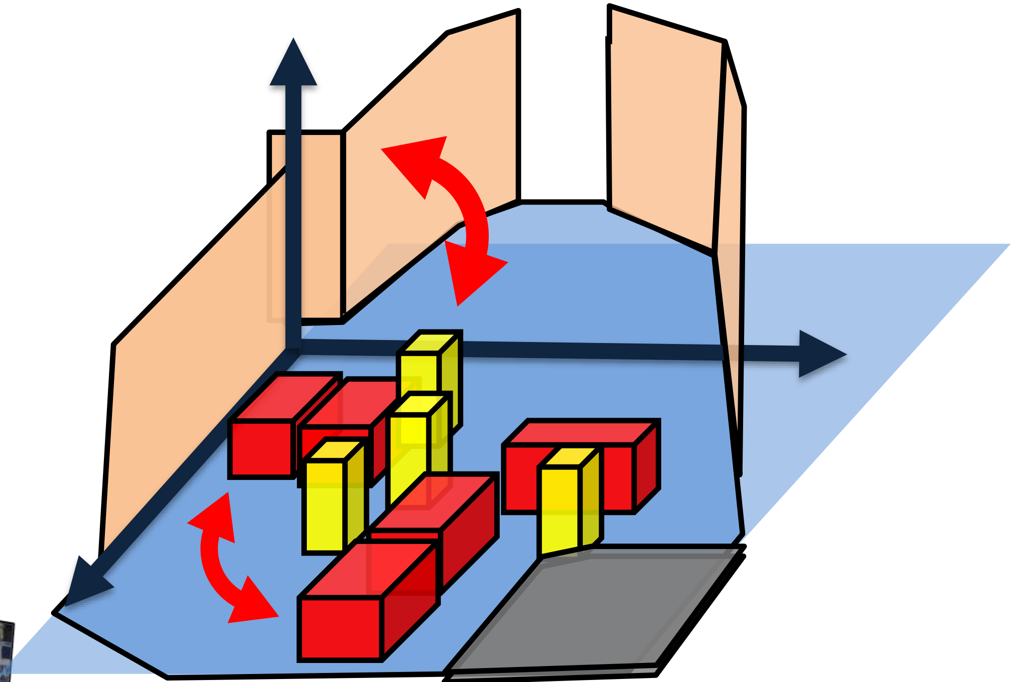
A first attempt….

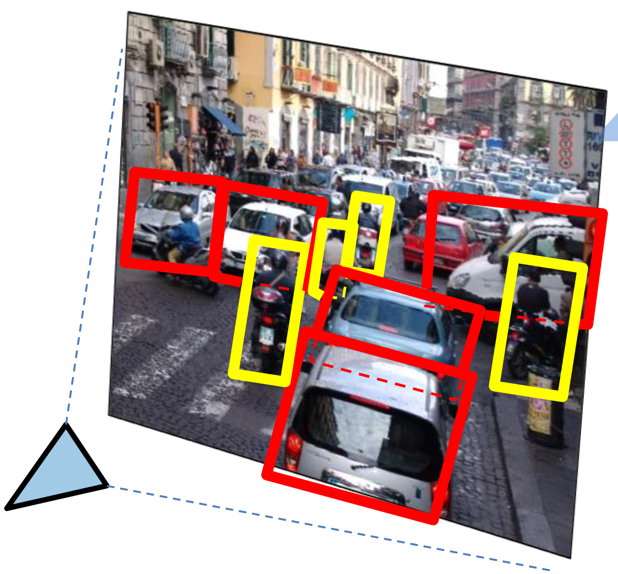
Bao, et al. CVPR 2010;
BMVC 2010;
CIVC 2011 **(editor choice)**
IJCV 2012



Generalization #1

Choi, et al., CVPR 13





Interactions between:

- Objects-space
- Object-object

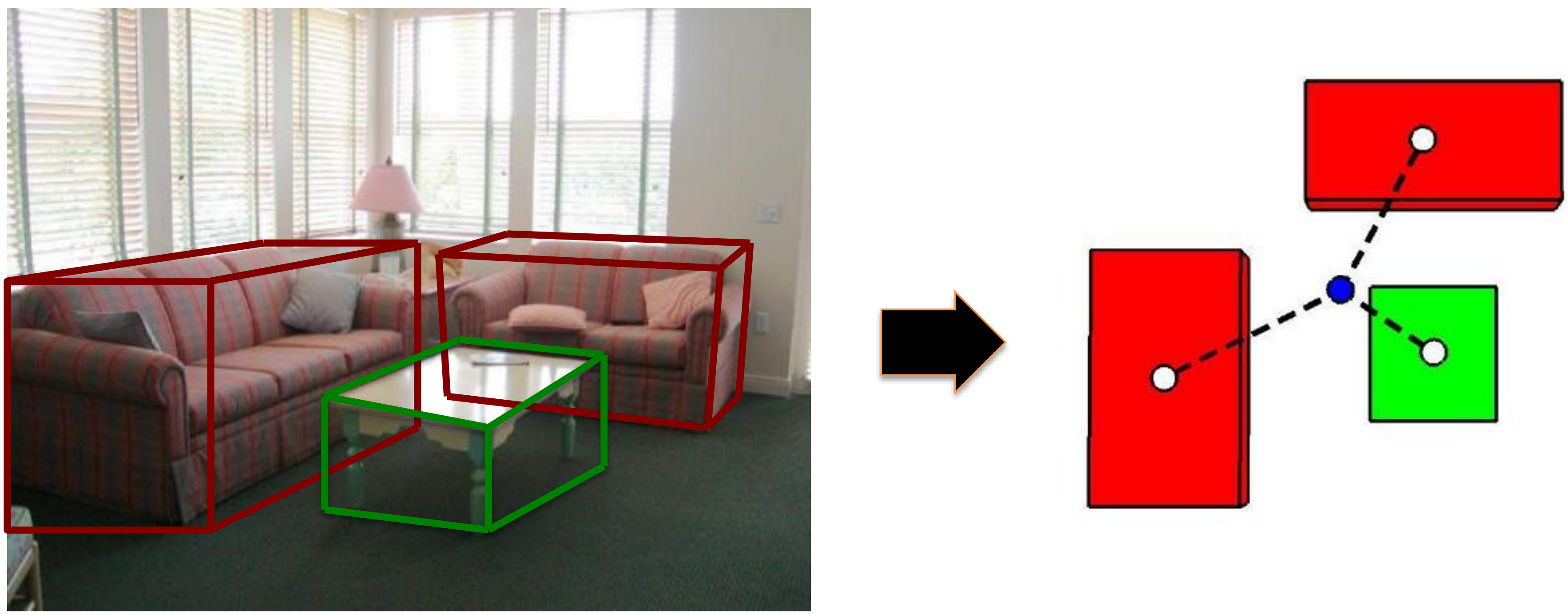
Oliva & Torralba, 2007
Rabinovich et al, 2007
Li & Fei-Fei, 2007
Vogel & Schiele, 2007

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

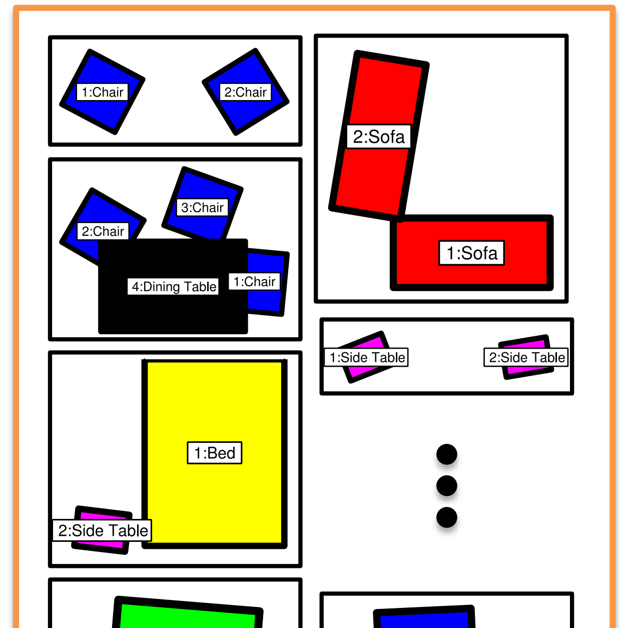
3D Geometric Phrases

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

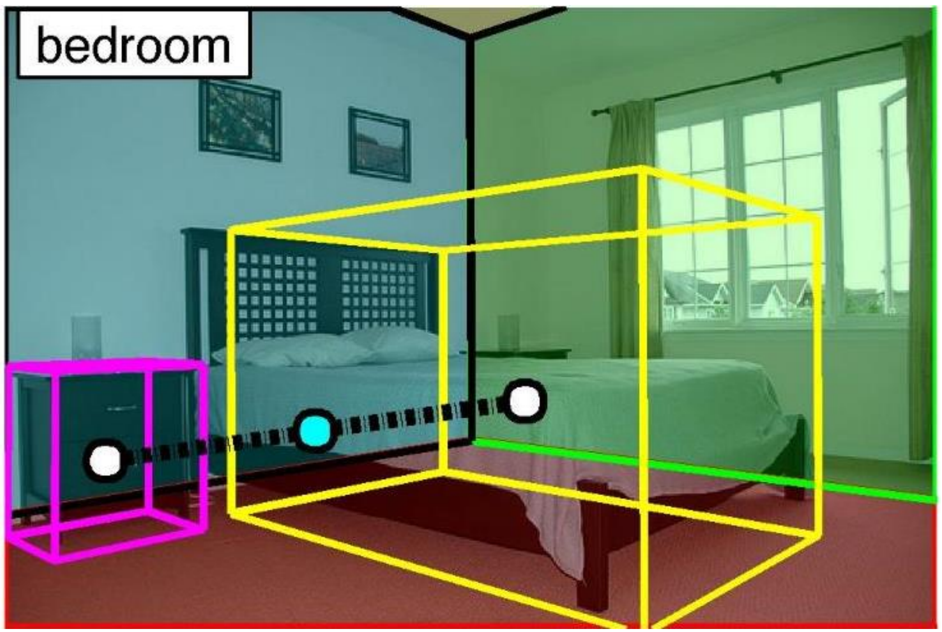


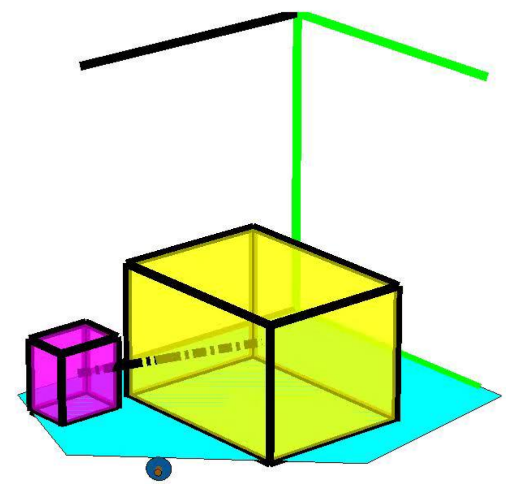
- **W/o annotations**
- **Compact**
- **View-invariant**

Using Max-Margin learning w/ novel Latent Completion algorithm

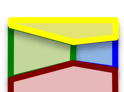
Results

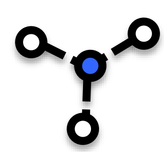
Indoor scene dataset [Choi et al., 12]





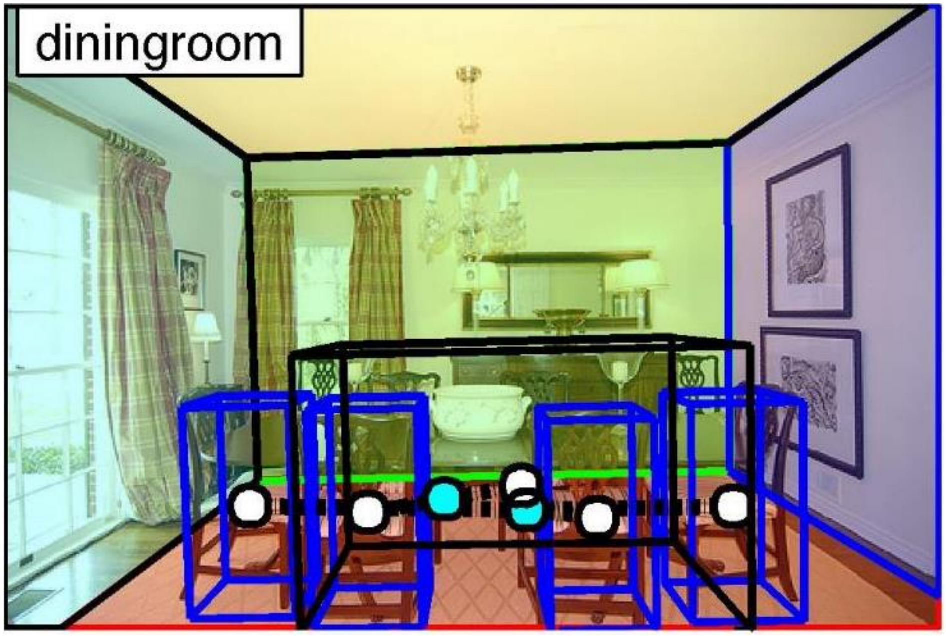
Sofa, Coffee Table, Chair, Bed, Dining Table, Side Table

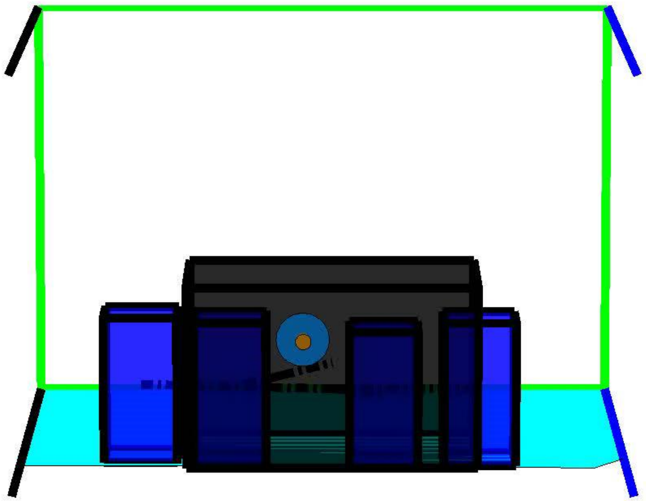
 **Estimated Layout**

 **3D Geometric Phrases**

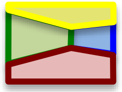
Results

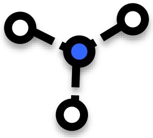
Indoor scene dataset [Choi et al., 12]



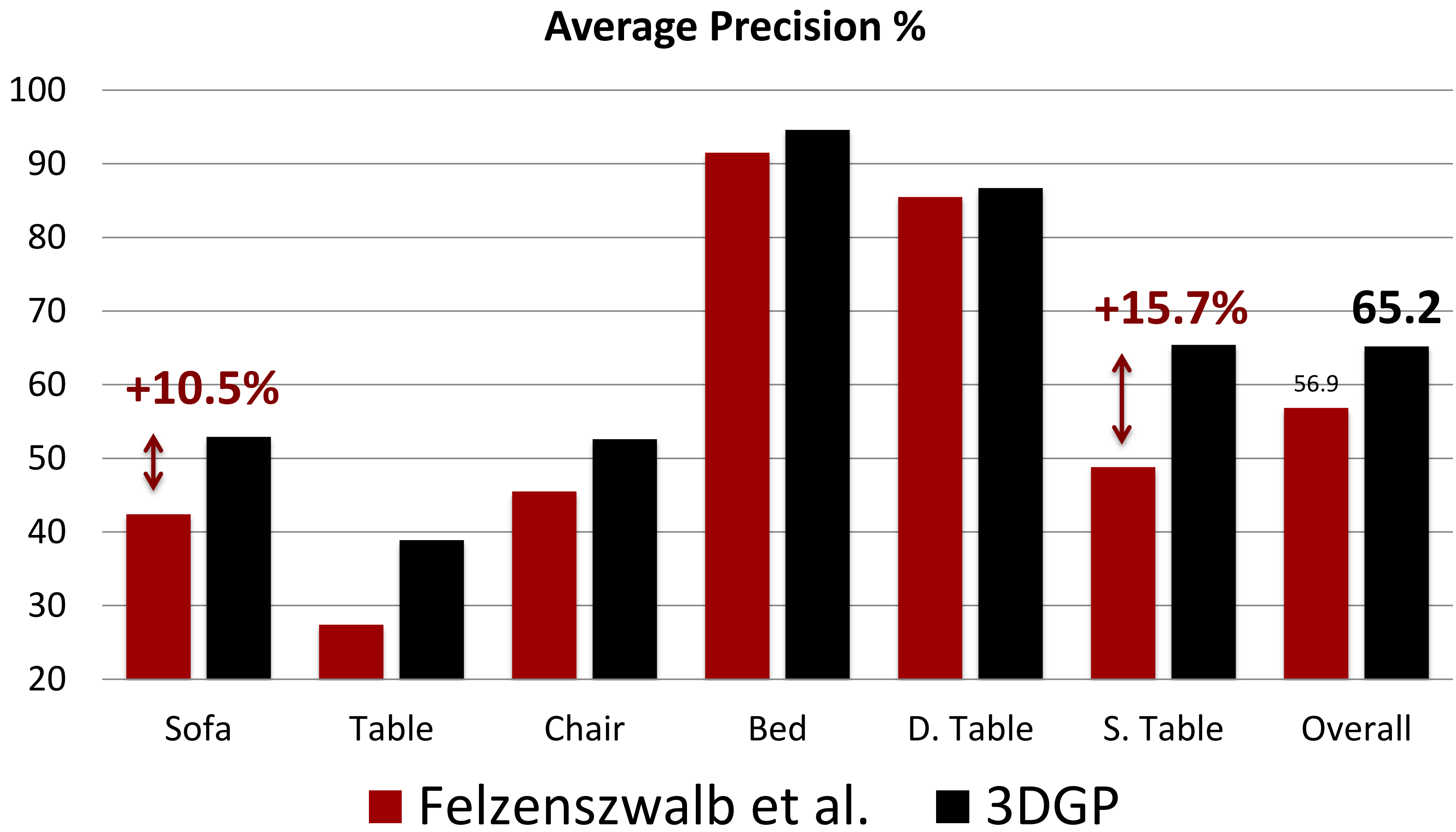


Sofa, **Coffee Table**, **Chair**, **Bed**, **Dining Table**, **Side Table**

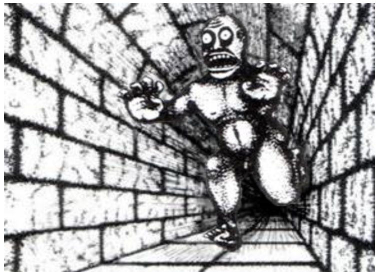

Estimated Layout


3D Geometric Phrases

Results: Object Detection



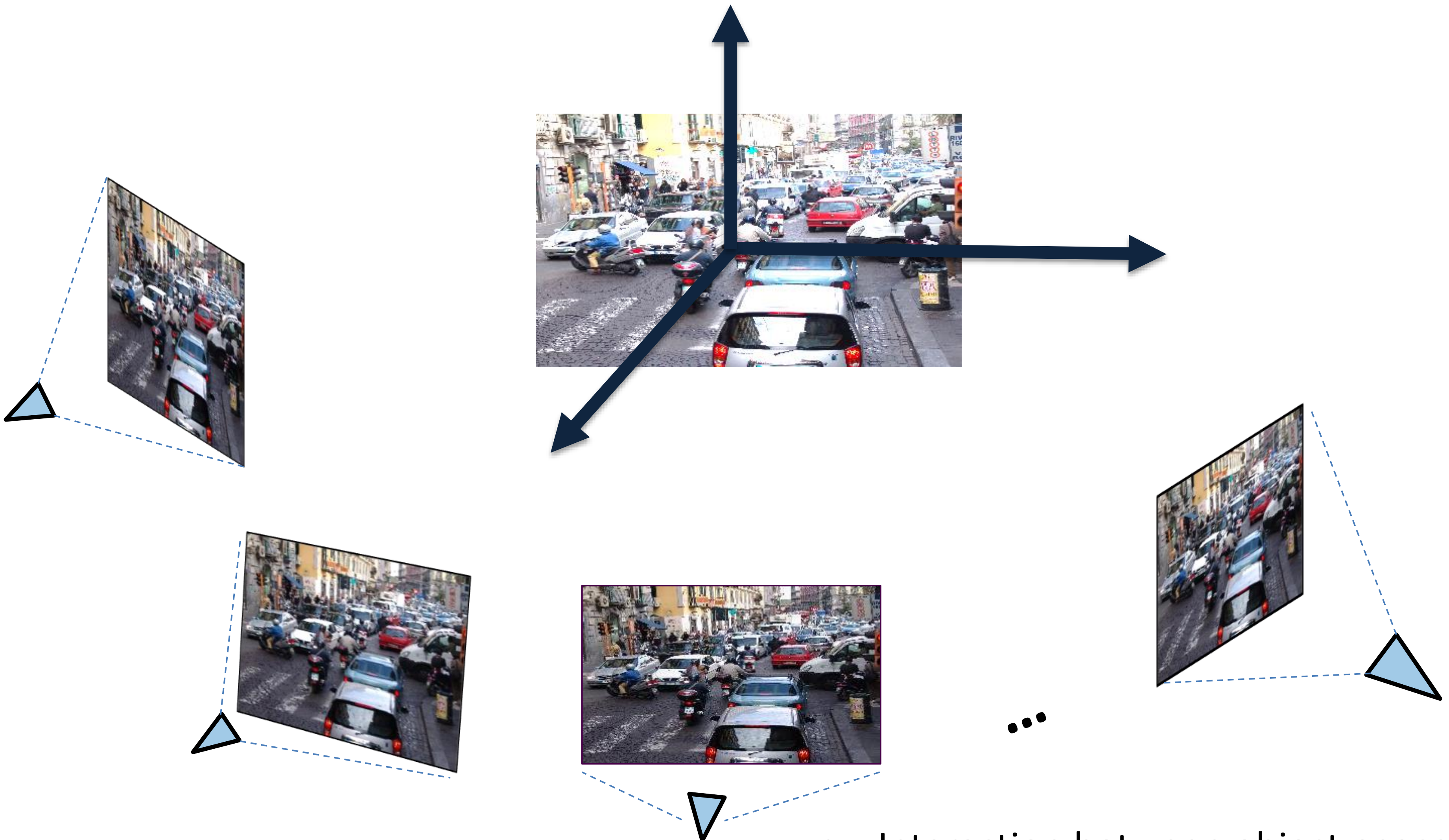
Indoor scene dataset [Choi et al., 12]



Outline

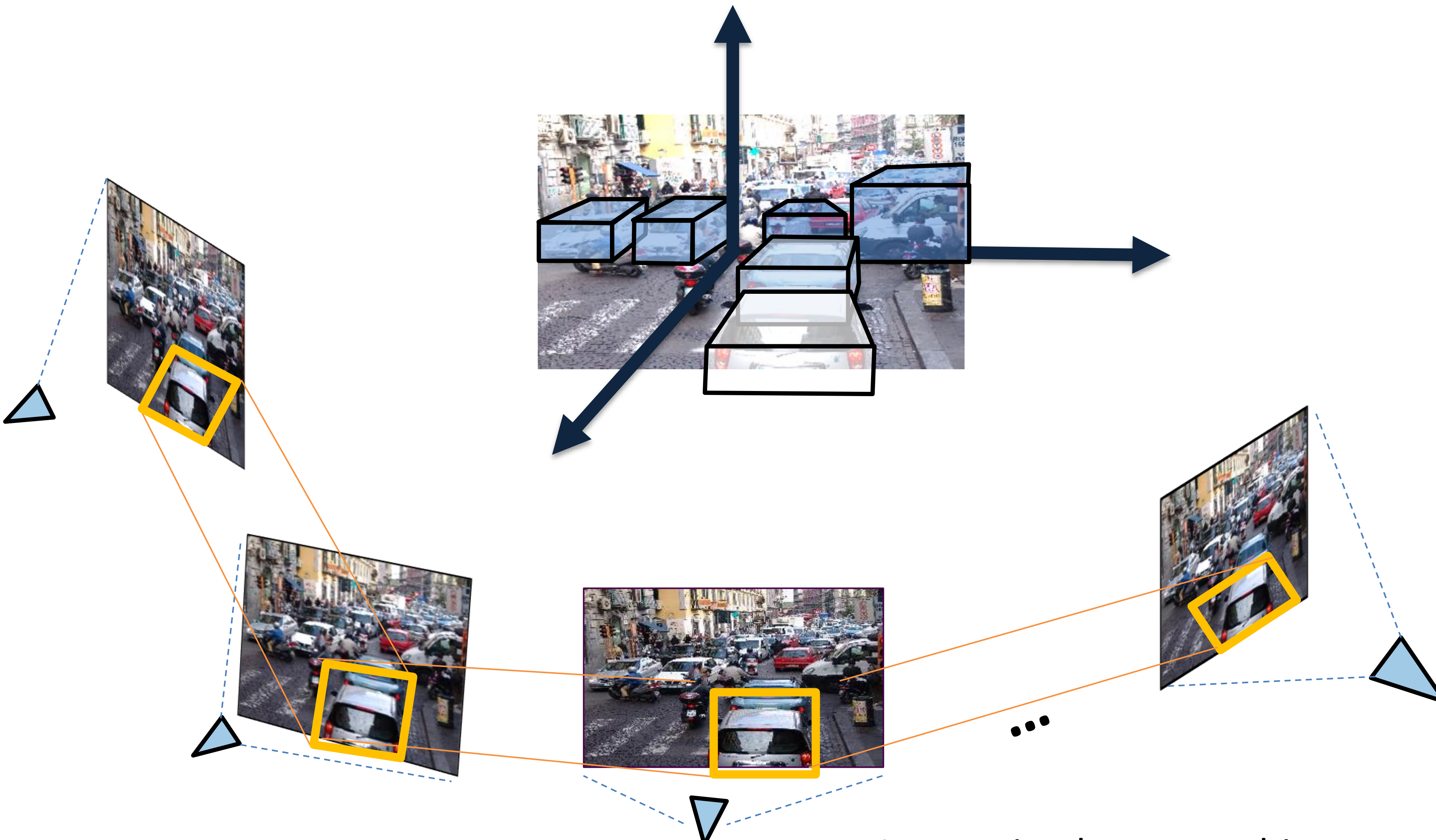
- Modeling objects and their 3D properties
- Modeling interaction among objects and space
- Modeling relationships of objects across views

Modeling relationships of objects across views



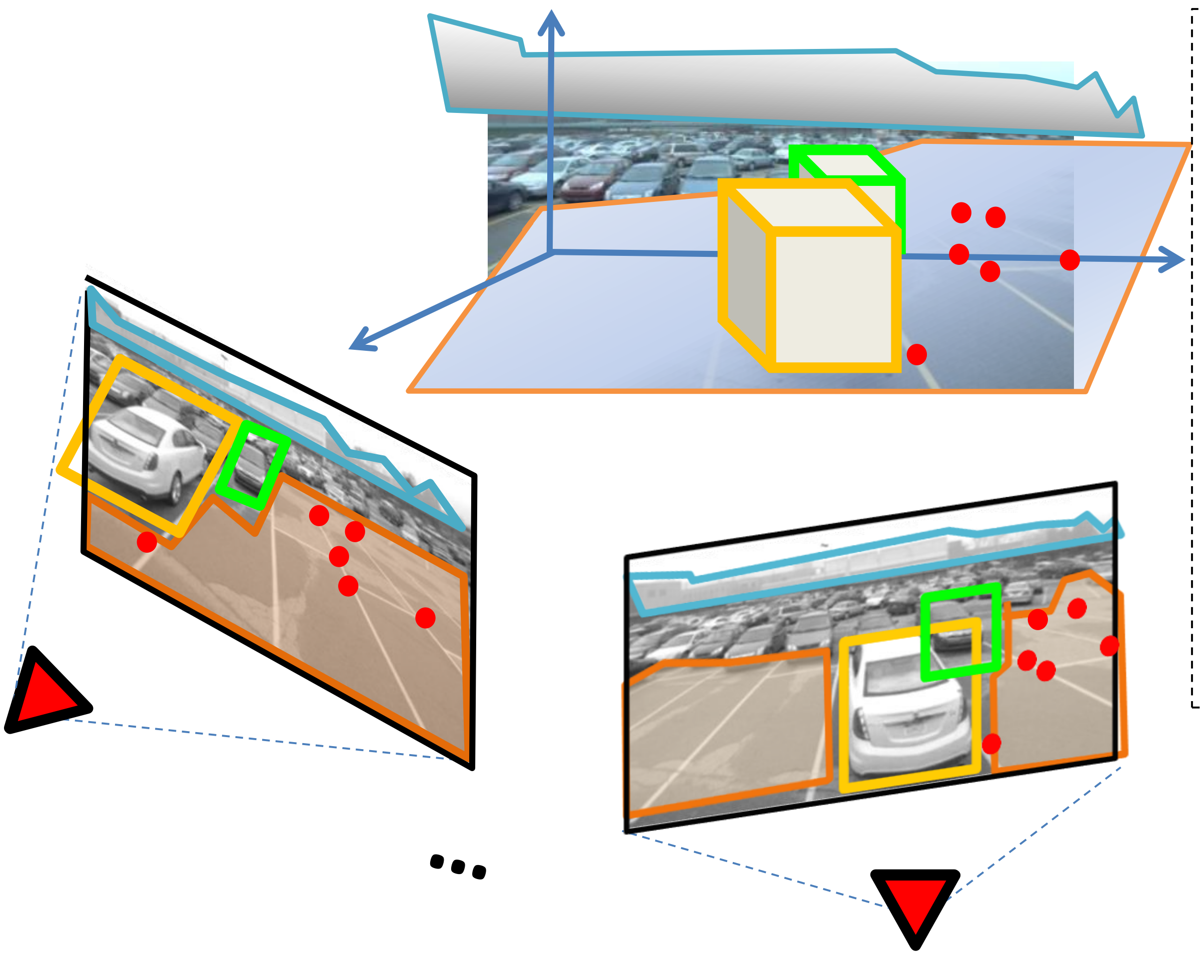
- Interaction between object-space
- Interaction among objects
- **Transfer semantics across views**

Modeling relationships of objects across views



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

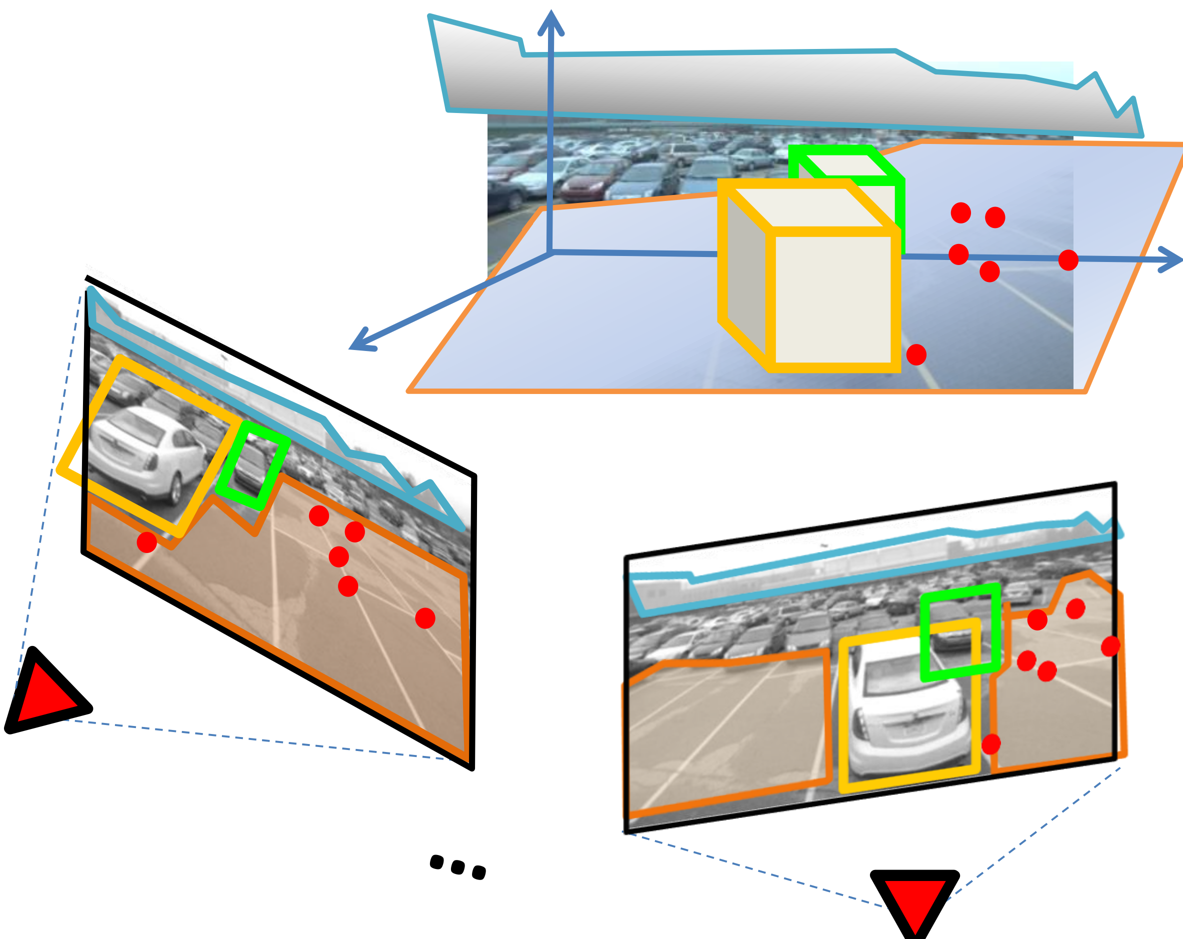
Semantic structure from motion



- 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

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

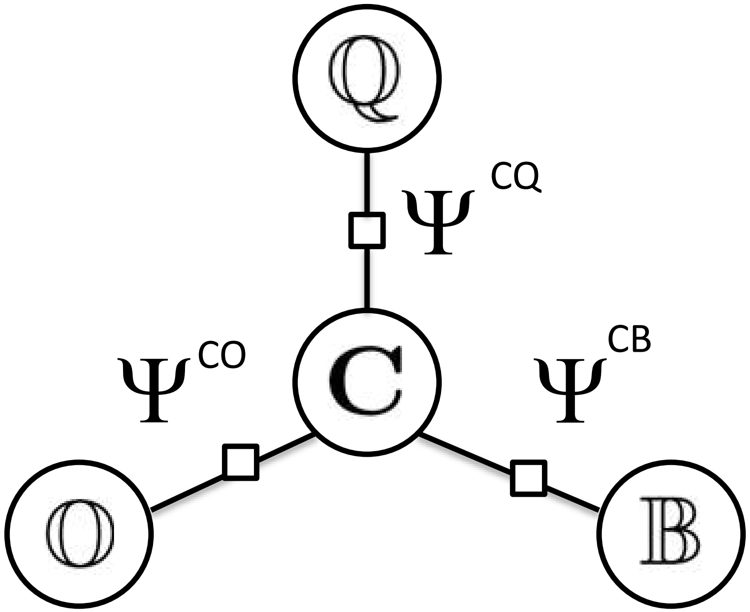


- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

Semantic structure from motion

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

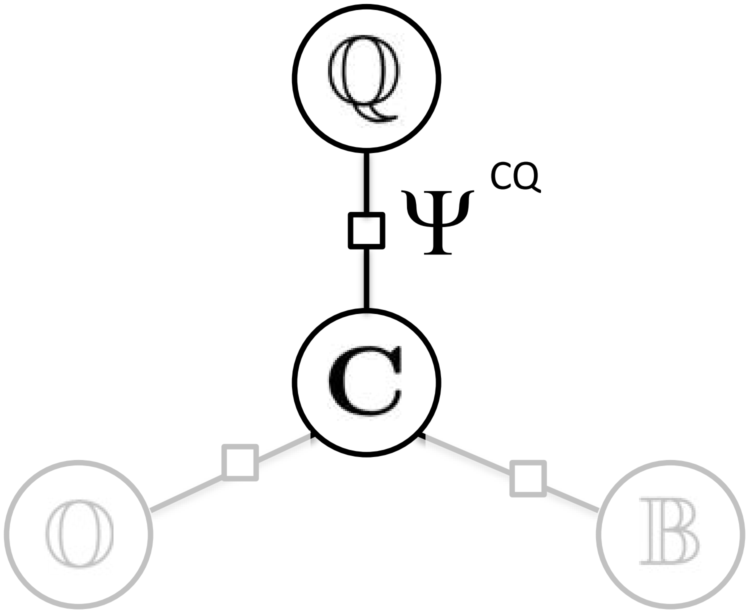
Factor graph



- 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

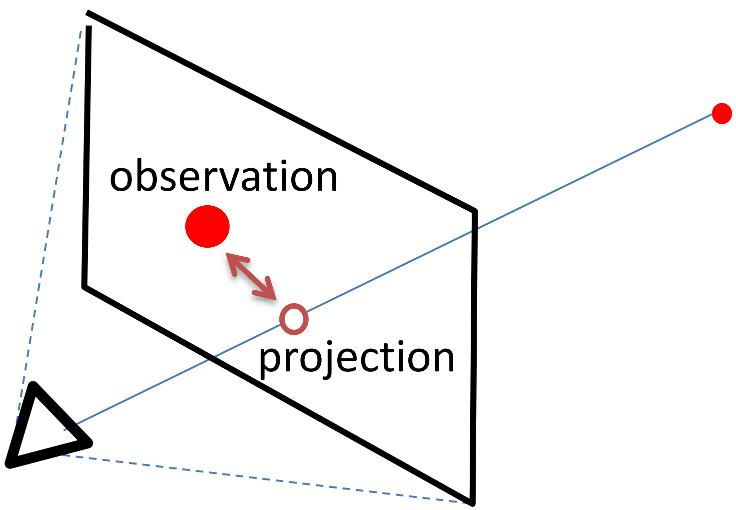
$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg \max_{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}} \boxed{\prod_s \Psi_s^{CQ}} \prod_t \Psi_t^{CO} \prod_r \Psi_r^{CB}$$



- 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

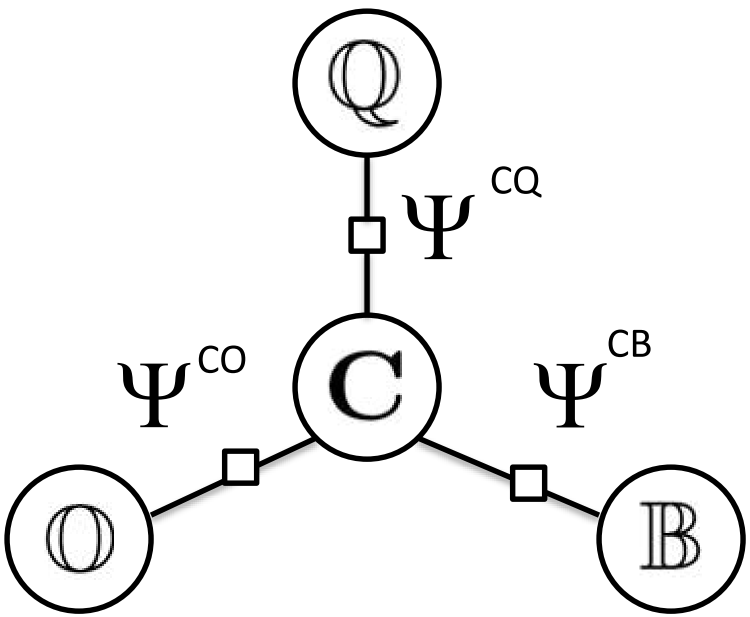
Point re-projection error

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

- 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

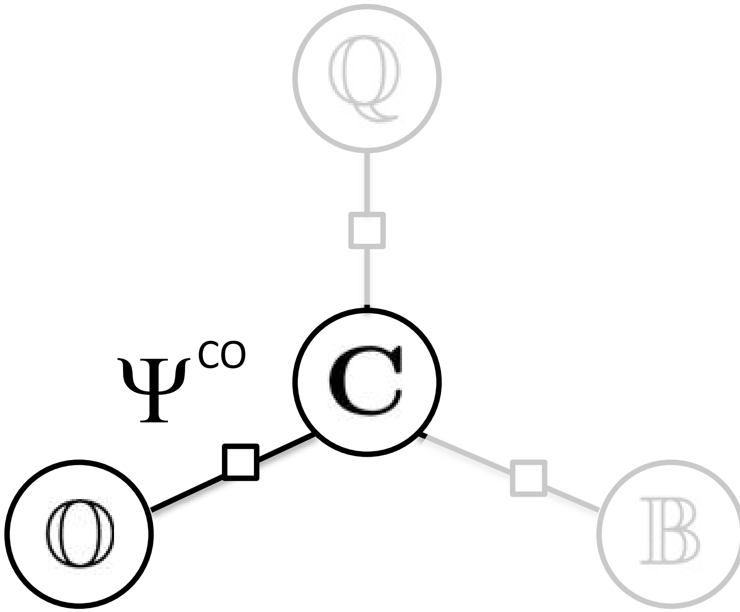
$$\{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}\} = \arg\max_{\mathbb{Q}, \mathbb{O}, \mathbb{B}, \mathbf{C}} \prod_s \Psi_s^{CQ} \boxed{\prod_t \Psi_t^{CO} \prod_r \Psi_r^{CB}}$$



- 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

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

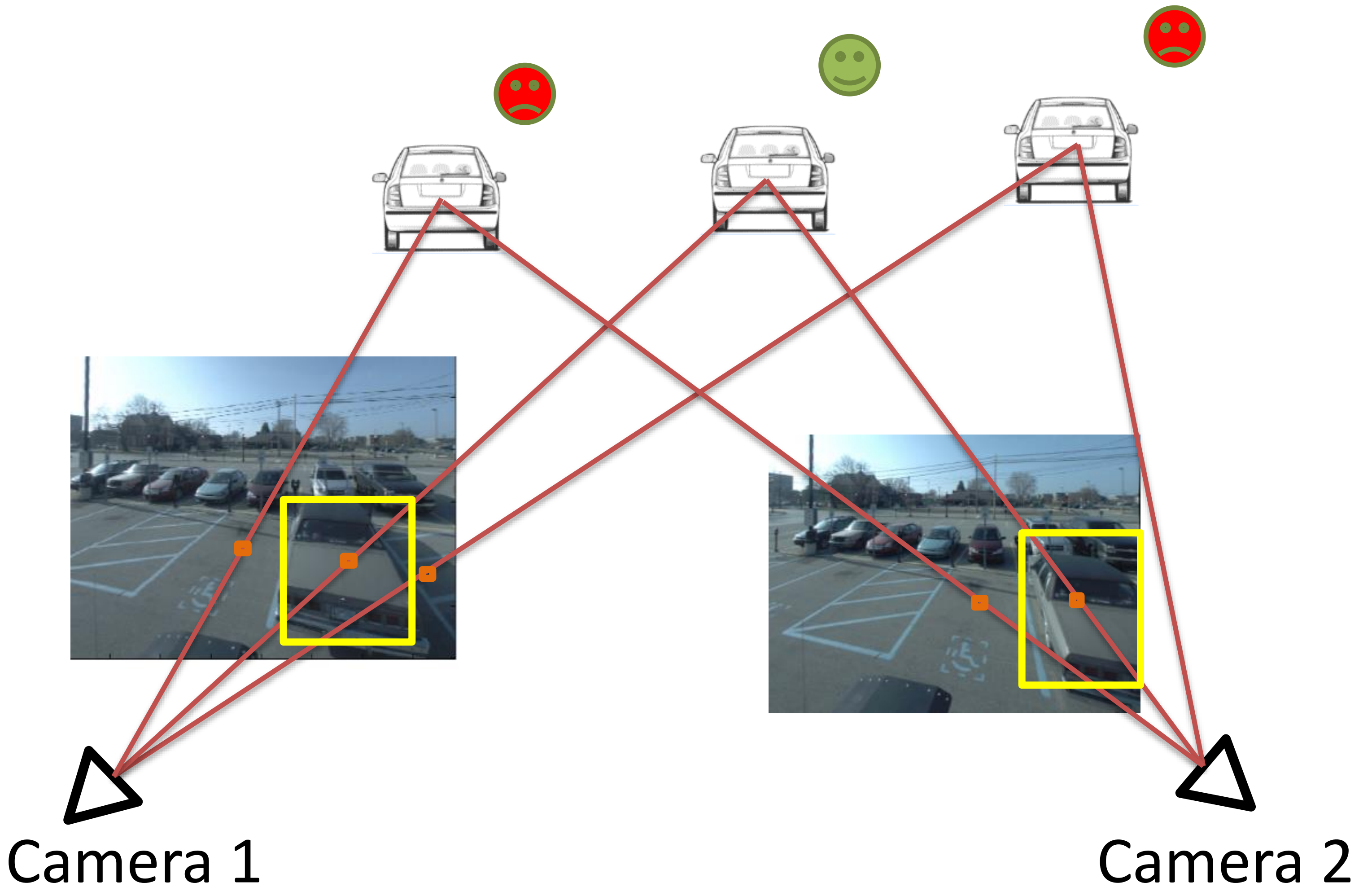


- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

Object "re-projection" error

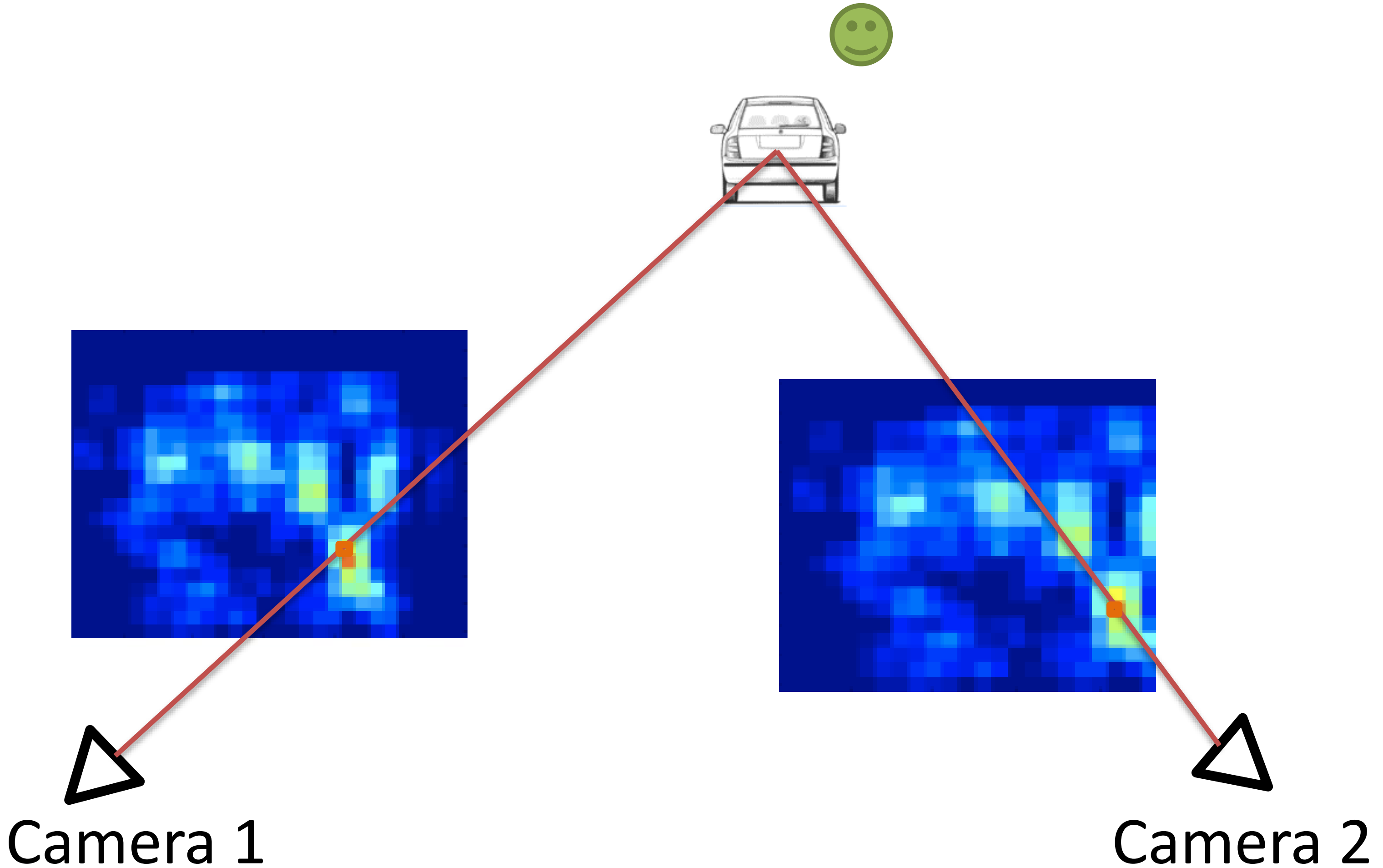
$$\Psi_t^{CO} \propto \prod_t^{N_t} (1 - \prod_k^{N_k} (1 - \Pr(o|O_t, C^k)))$$

SSFM: Object-level compatibility



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

SSFM: Object-level compatibility

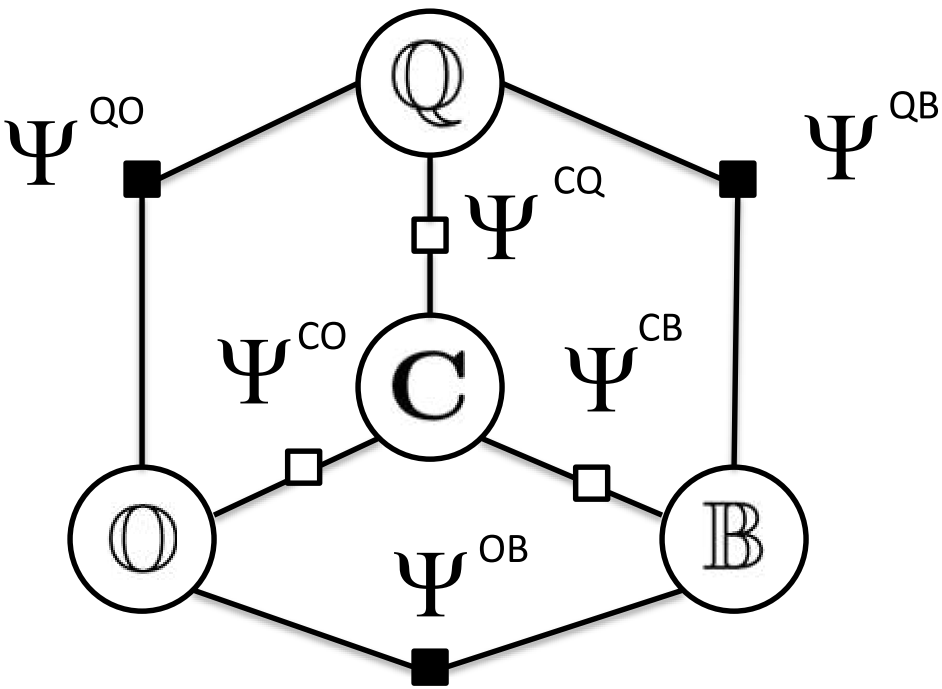


- Agreement with measurements is computed using position, pose and scale

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

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

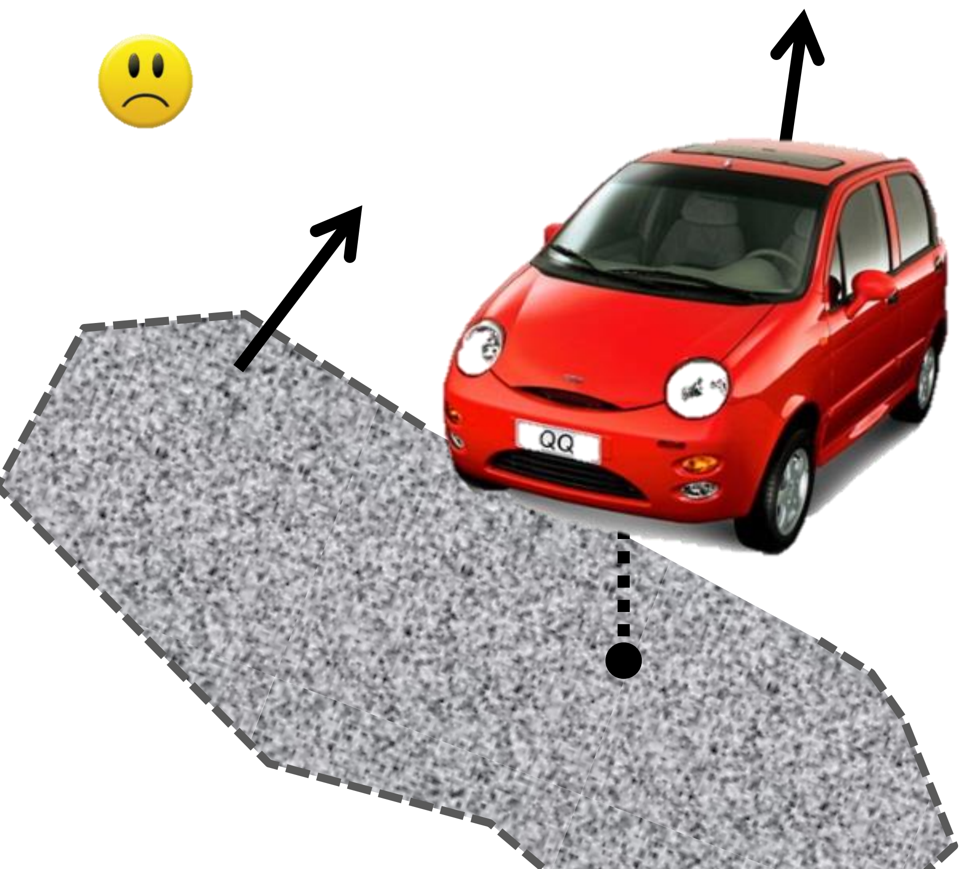
Object-Region Interactions:

- 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:



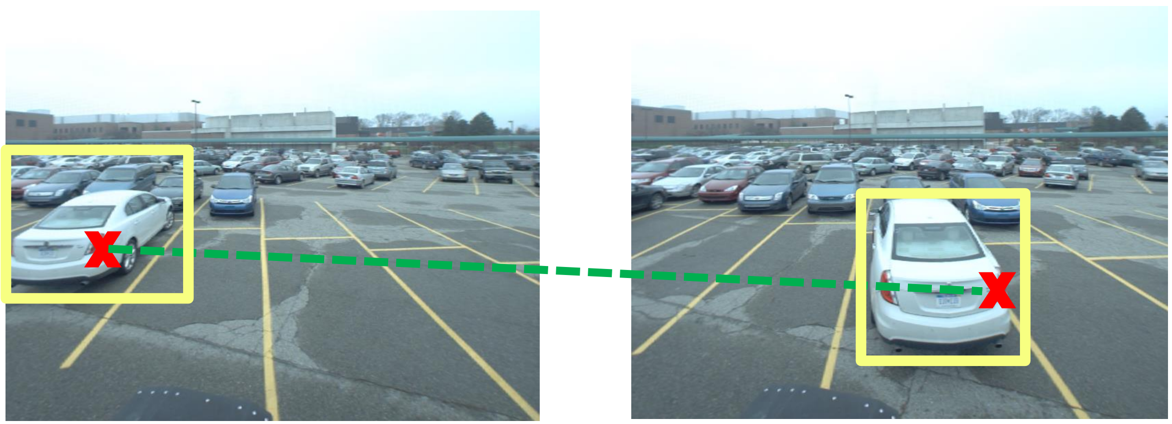
- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)

- 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:

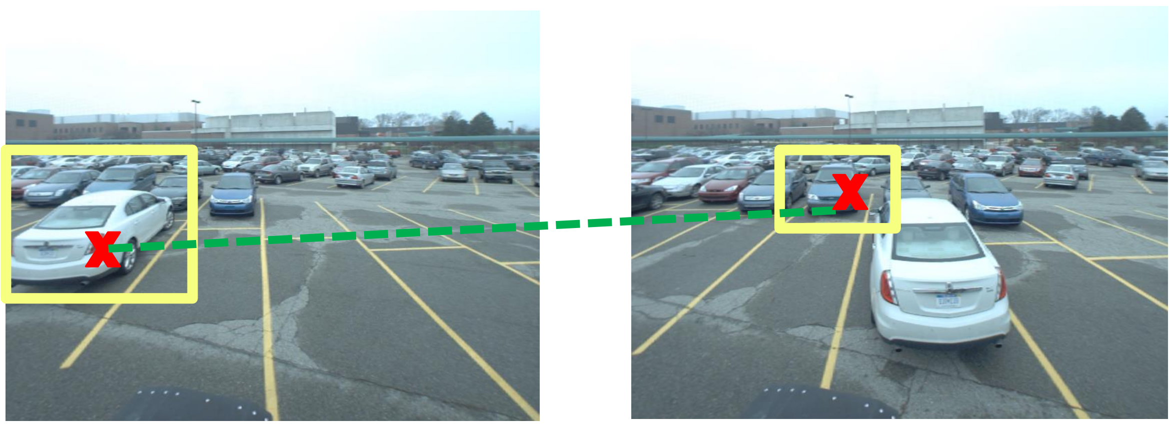


- Measurements I
  - Points (x,y,scale)
  - Objects (x,y, scale, pose)
  - Regions (x,y, pose)
- Model Parameters:
  - $\mathbb{Q}$ = 3D points
  - $\mathbb{O}$ = 3D objects
  - $\mathbb{B}$ = 3D regions
  - $\mathbf{C}$ = cam. prm. K, R, T

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

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

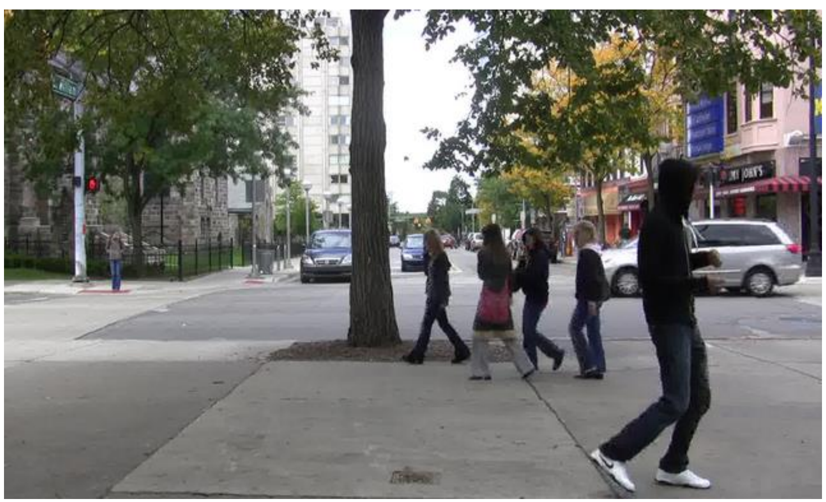
- Modified Reversible Jump Markov Chain Monte Carlo (RJ-MCMC) sampling algorithm [Dellaert et al., 2000]
- Initialization of the cameras, objects, and points are critical for the sampling
- Initialize configuration of cameras using:
    - SFM
    - consistency of object/region properties across views

Results

Input images

FORD CAMPUS dataset [Pandey et al., 09]

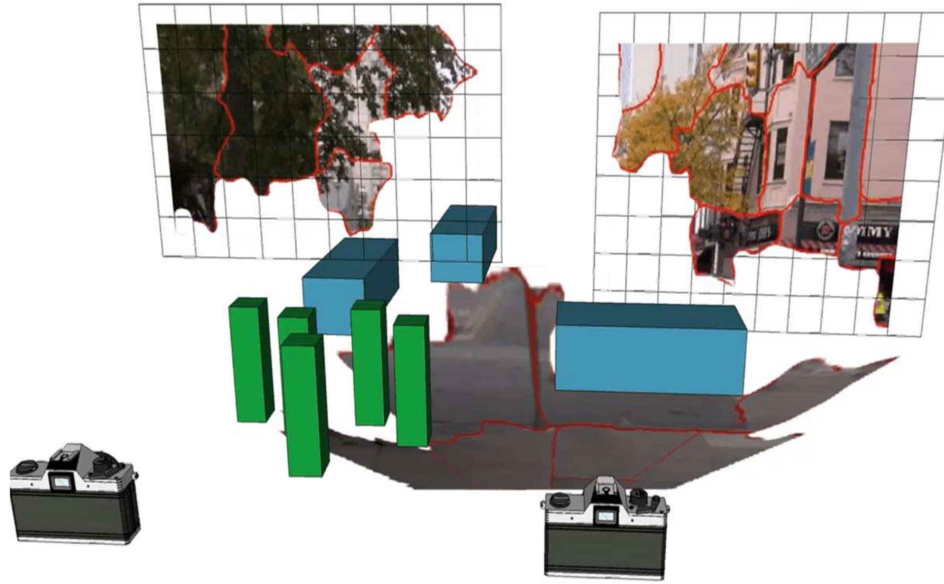
⋮

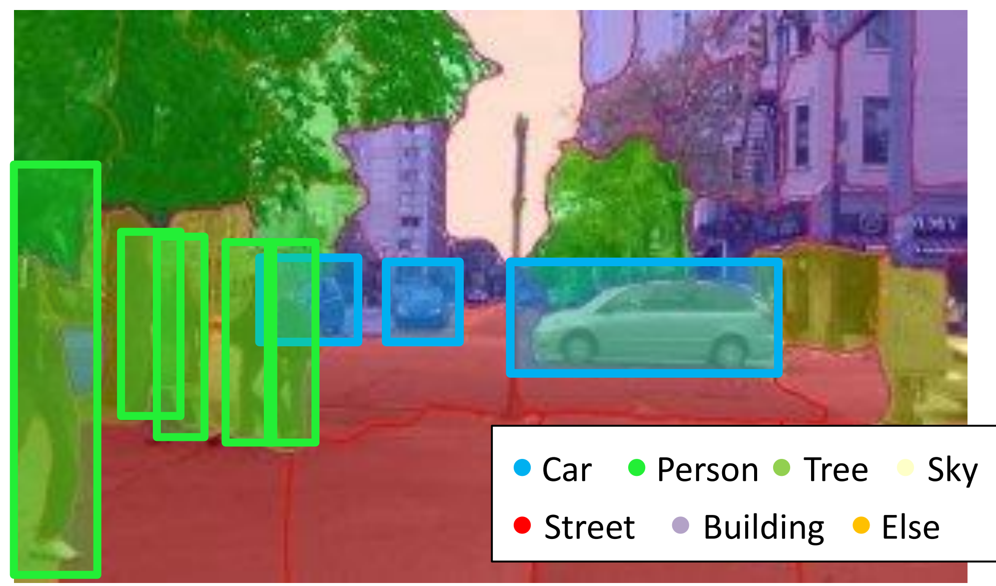


- Wide baseline
- Background clutter
- Limited visibility
- Un-calibrated cameras

Results

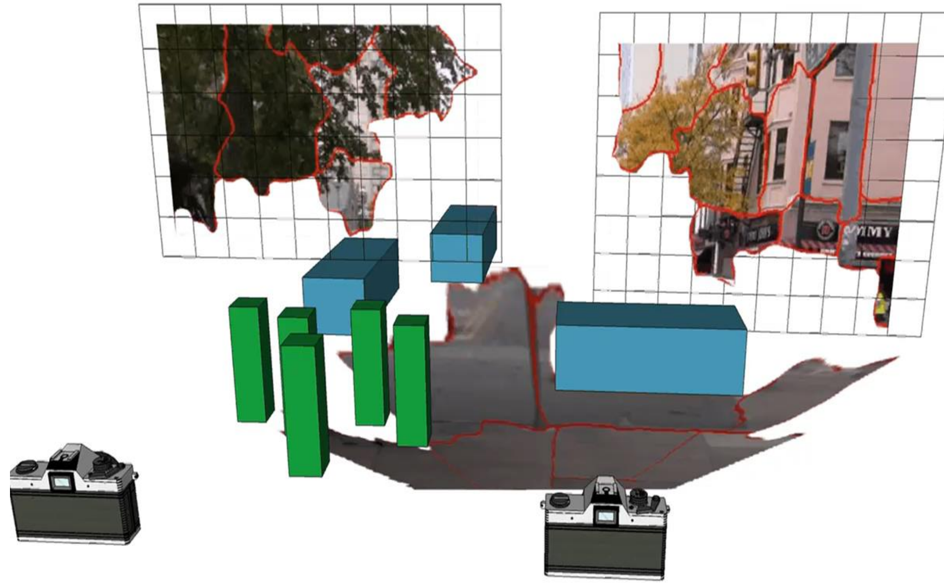
Input images





Results

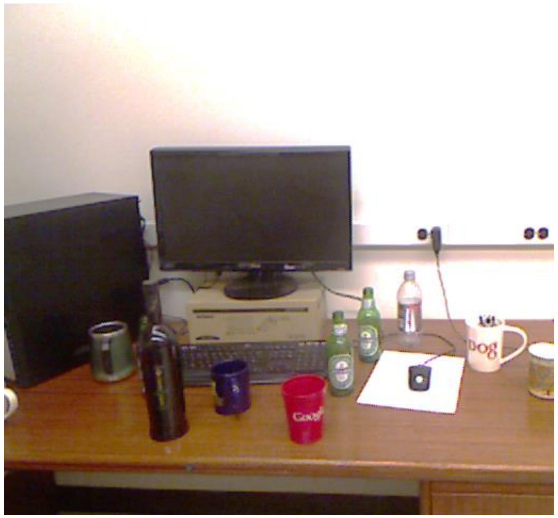
Input images



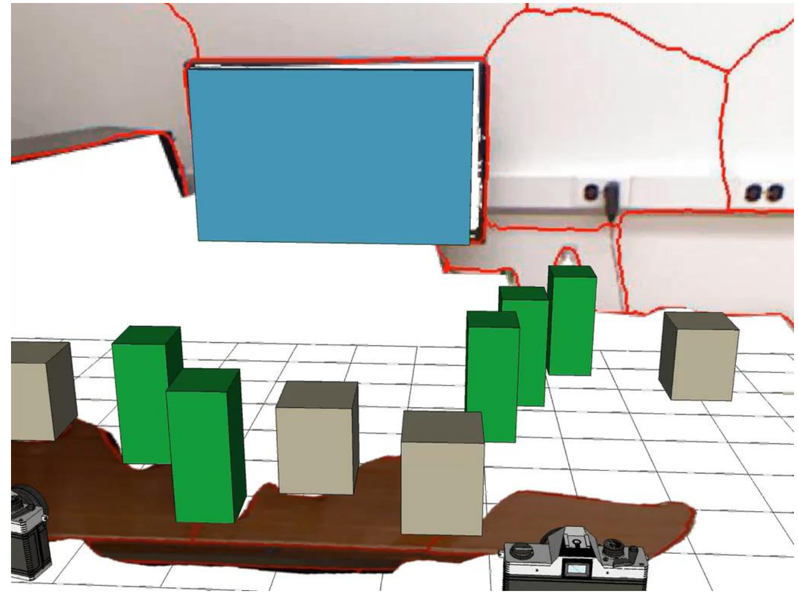
Results

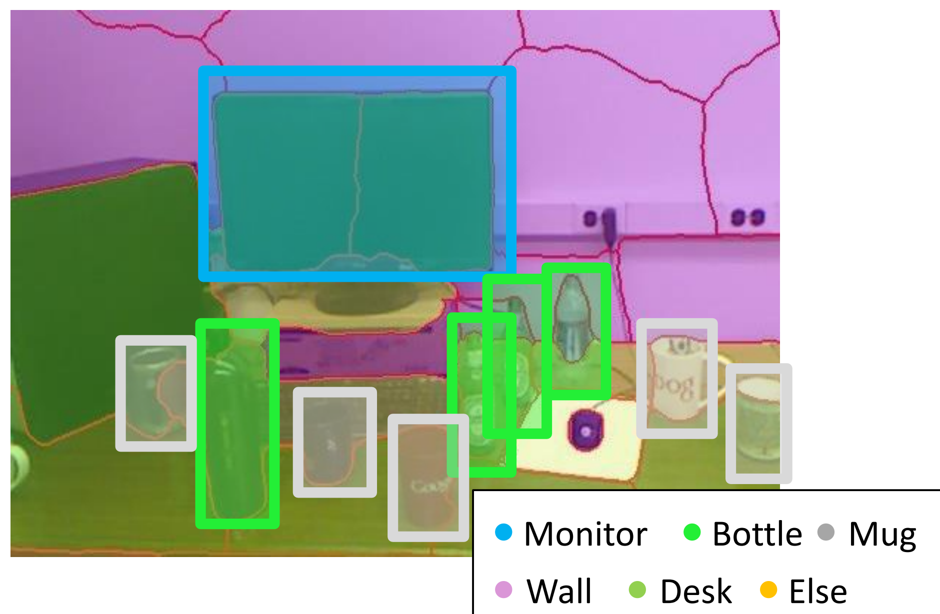
Input images

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



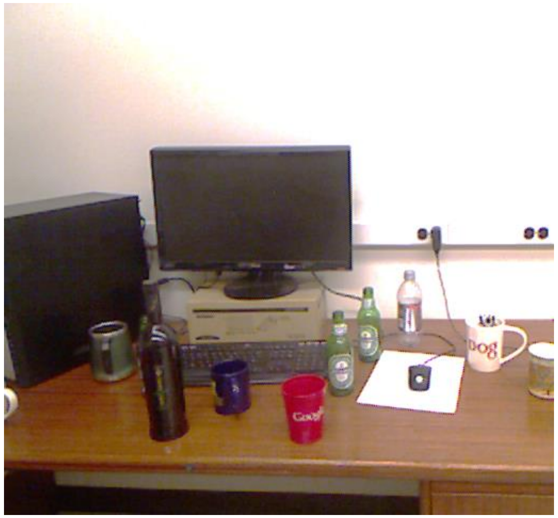
⋮



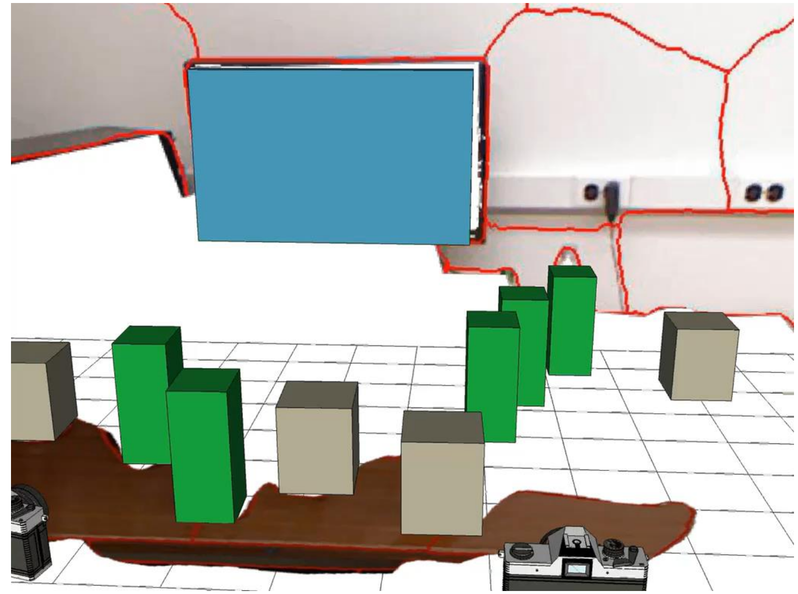


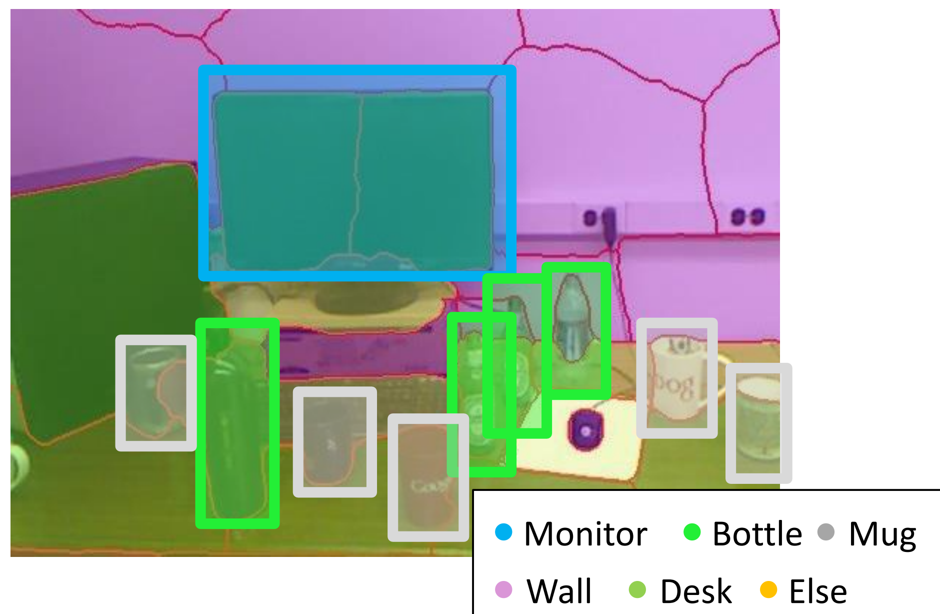
Results

Input images



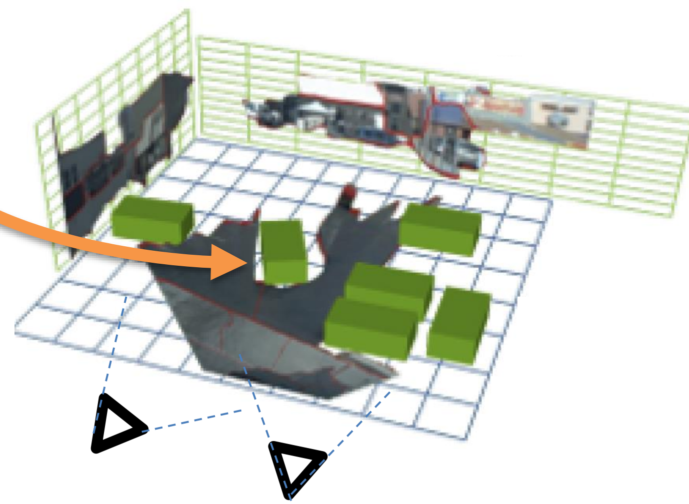
From the office dataset [Bao et al., 11]





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%**

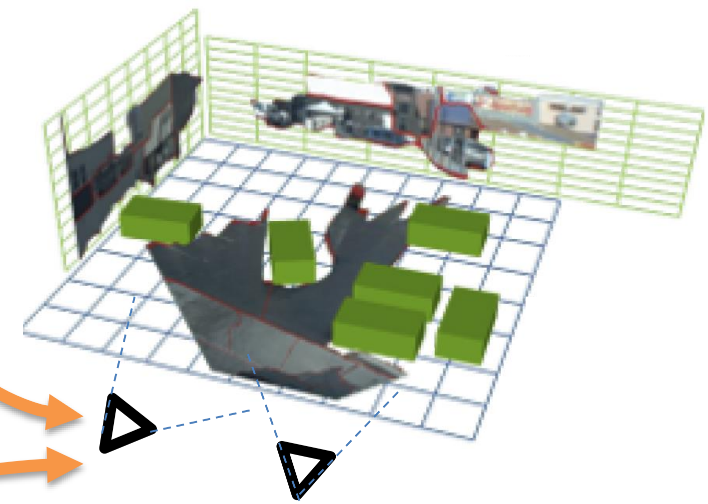
FORD CAMPUS dataset [Pandey et al., 09]
Office dataset [Bao et al., 11]

[1] Felzenszwalb et al. 2008

Results

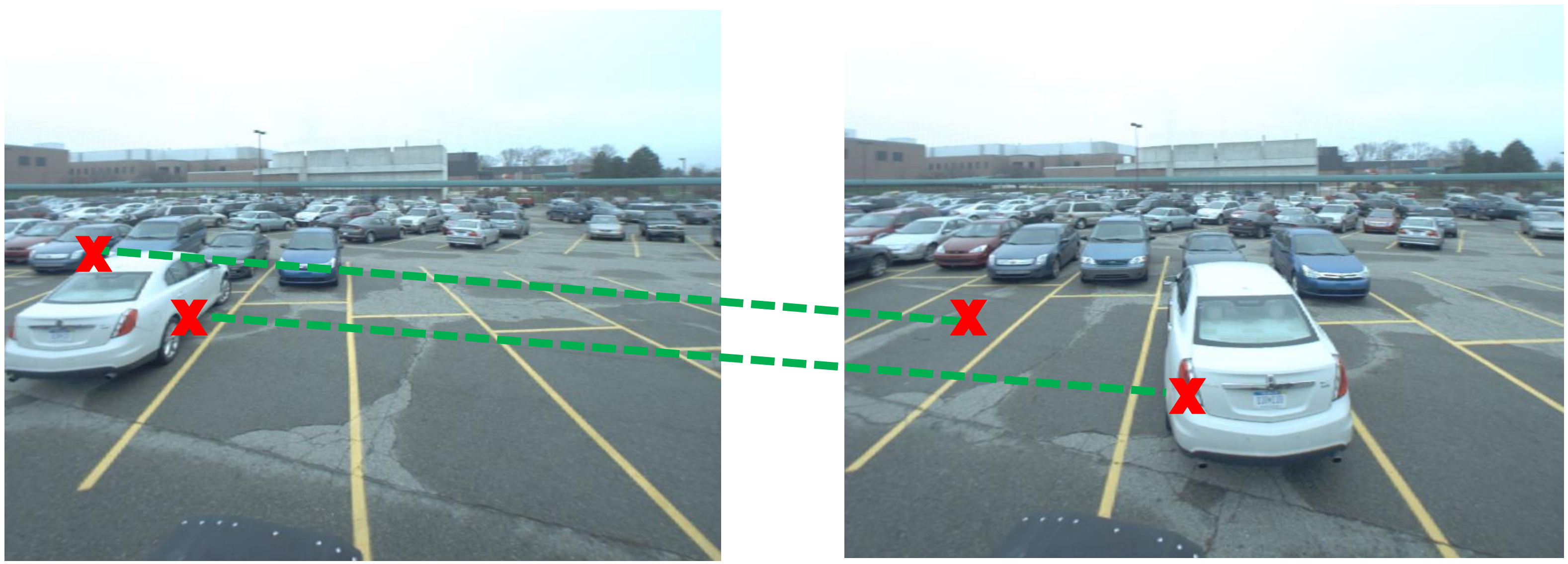
	Camera translation error		
	SFM Snavely et al., 08	**SSFM** no int.	**SSFM**
FORD CAMPUS	26.5°	19.9°	**12.1°**
OFFICE	8.5°	4.7°	**4.2°**
STREET	27.1°	17.6°	**11.4°**

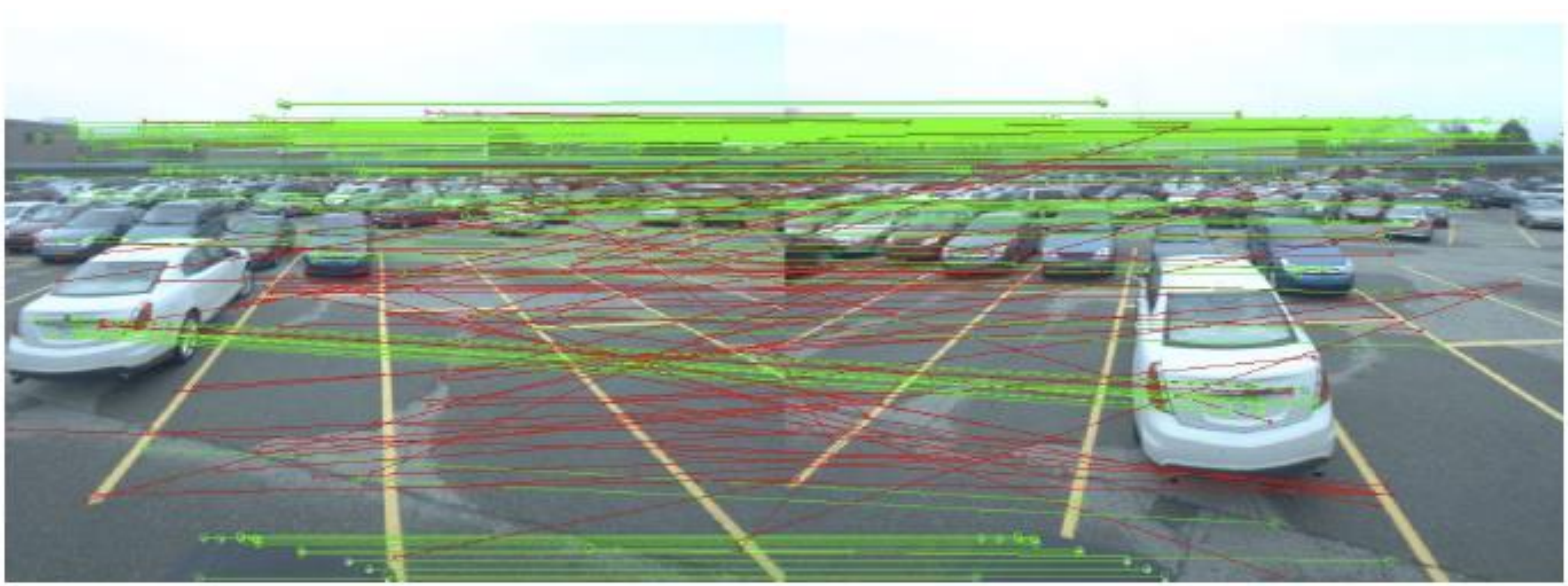
Camera rotation error		
SFM Snavely et al., 08	**SSFM** no int.	**SSFM**
<1°	<1°	**<1**
9.6°	4.2°	**3.5°**
21.1°	3.1°	**3.0°**



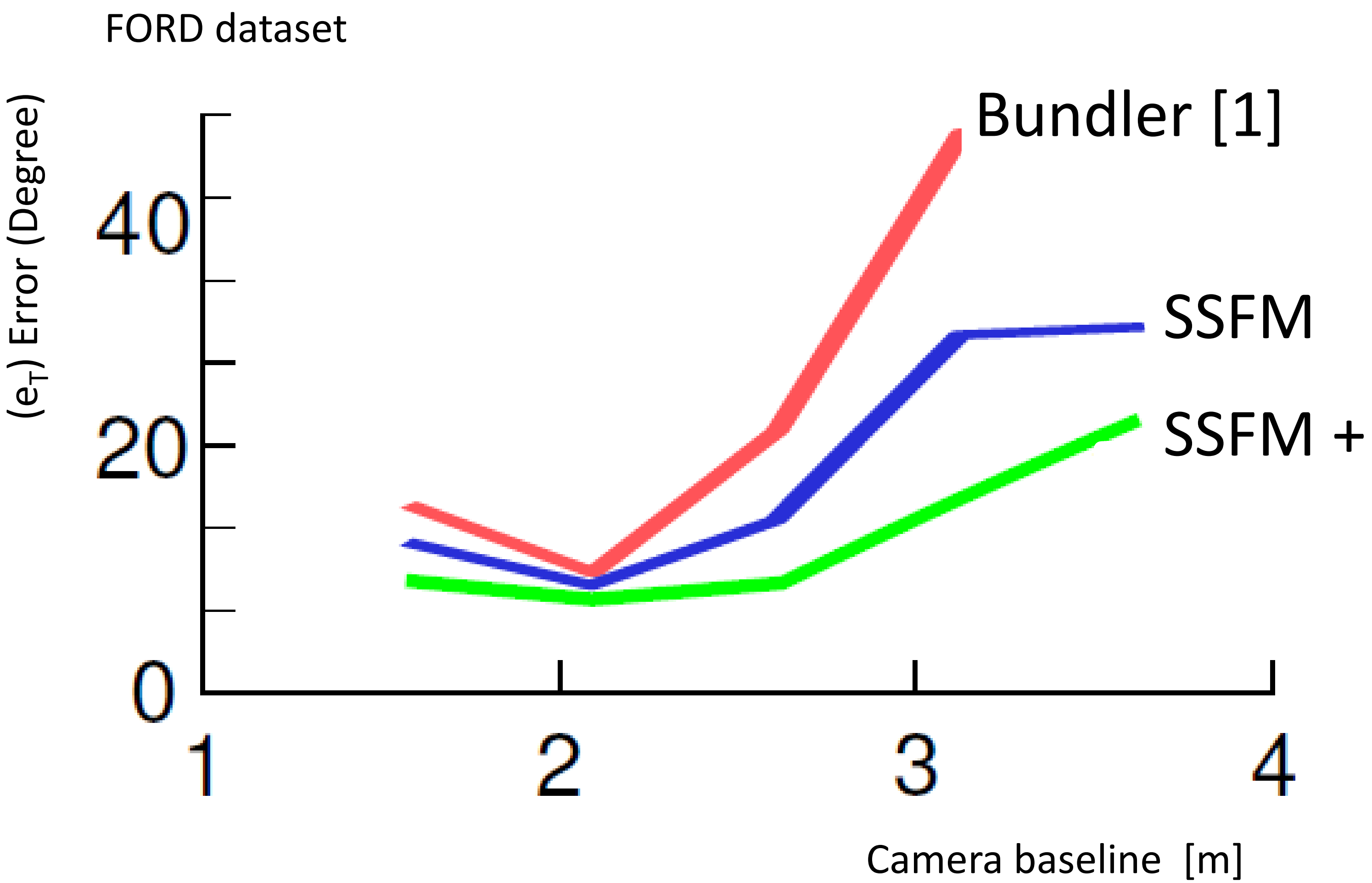
FORD CAMPUS dataset [Pandey et al., 09]
Office dataset [Bao et al., 11]
Street dataset [Bao et al., 11]

Wide-baseline feature correspondence





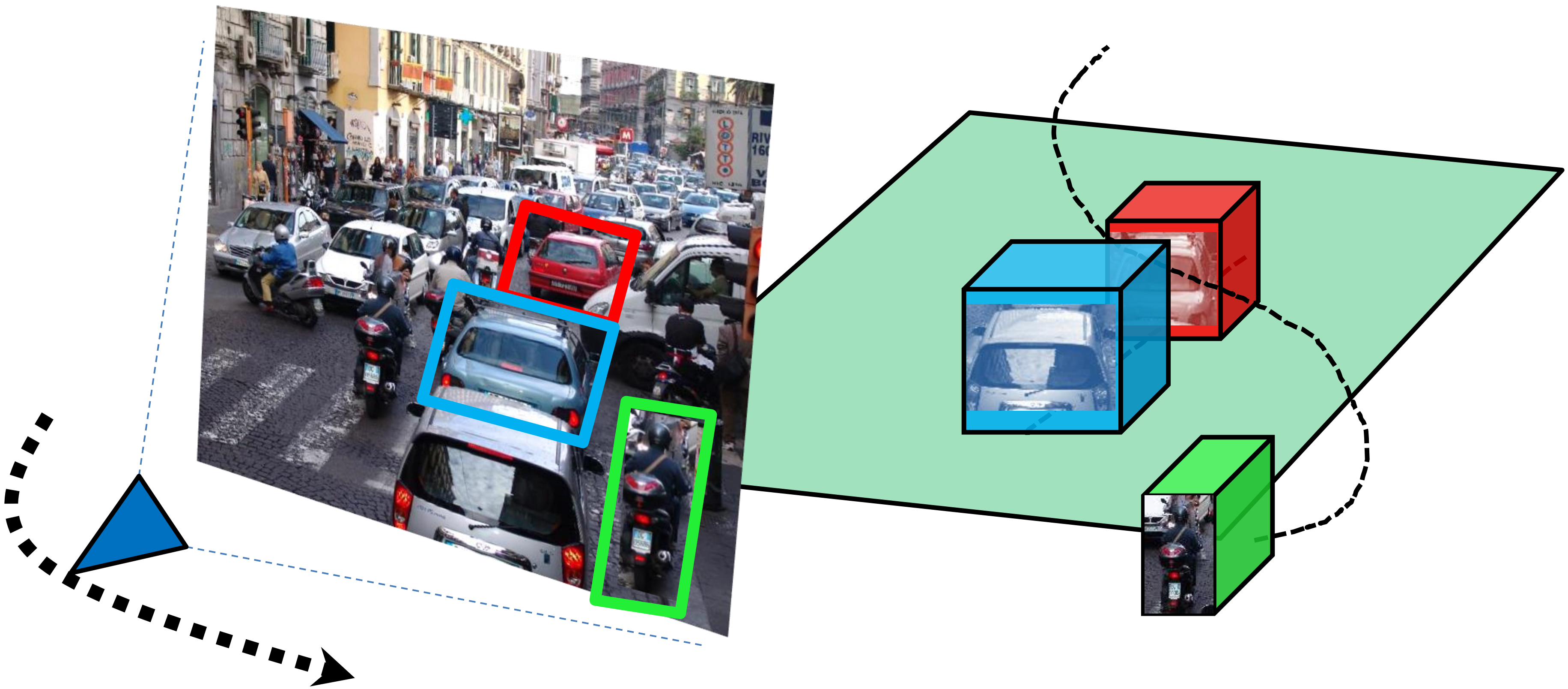
Camera Pose Estimation v.s. Base Line Width



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

Scene understanding from multiple views

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

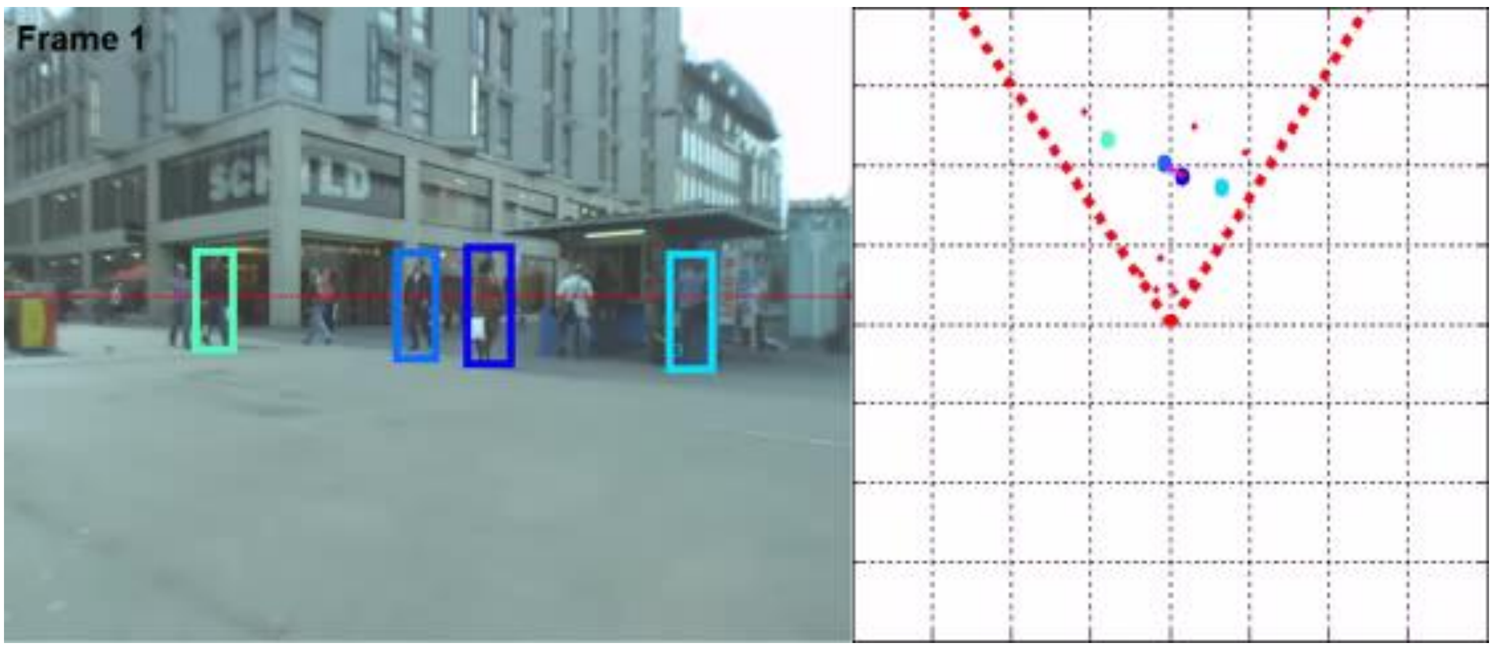


Interactions between:

- Moving targets and space

Multi-target tracking from moving cameras

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



- Monocular cameras
- Un-calibrated cameras
- Arbitrary motion

- Highly cluttered scenes
    - Occlusion
    - Background clutter
- Almost in real time!

Conclusions

- Joint reconstruction and recognition enables:
  - Rich characterization of the scene
  - Accurate recognition and reconstruction results
- 3 ingredients for enabling 3D perception: Ability to model:
  - Objects and their 3D properties
  - Interaction among objects and space
  - Relationships of objects across views

Summary

- Enable better tools for visualization
- Automate communication of performance deviations
- Reduction in delivery time
- Safety management
- 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

What's ahead

Granularity

- Representation
- Learning
- Computational demands

Scale

Good luck on your presentations on Wednesday & Thursday!