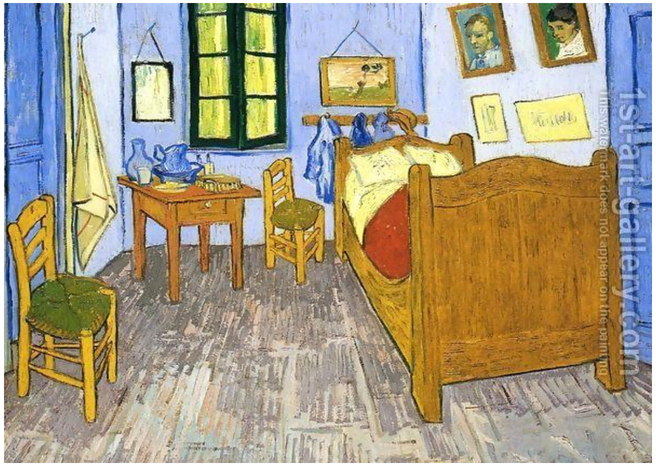


CS331B (3 units)

3dRR: Representation and Recognition

Instructor: Silvio Savarese

Email: ssilvio@stanford.edu

Office: Gates, room: 228

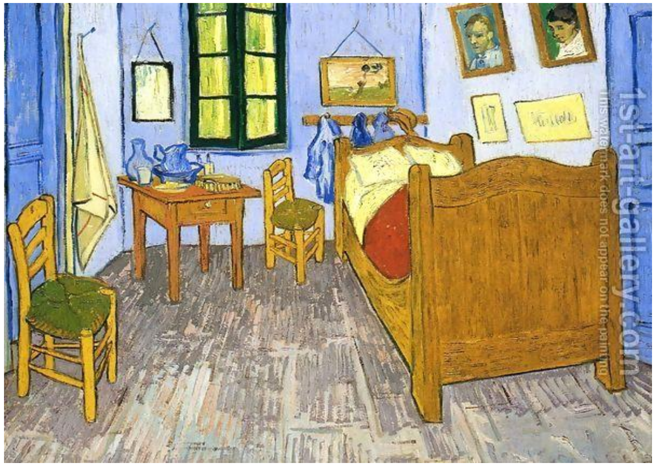
Office hour: by appointment

TA: Scott Chung

scottc52@gmail.com

Class Time & Location

- Monday & Tuesday 2:15-3:30pm



CS331B (3 units)

3dRR: Representation and Recognition

- Course inspired by a series of workshops called 3dRR in conjunction with ICCV 2007, 2009, 2011 (and the upcoming 2013)
- The last workshop had hundred of attendees
- Has been the venue of influential papers. E.g., Make 3D by A Saxena and A. Ng from Stanford (best paper award 2007)

Agenda

- **Administrative**
  - **Requirements**
  - **Grading policy**
- Overview of this course

Prerequisites

- Some experience in research with one of the following fields: computer vision (CS 231) & machine learning (CS 229)

References:

- R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2011.

- D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach* (2nd Edition). Prentice Hall, 2011.

- D. Hoiem and S. Savarese. *Representations and Techniques for 3D Object Recognition and Scene Interpretation*, Morgan Claypool Publishers, 2011

- Links to papers and supplementary material from syllabus page

Requirements

- Co-Present once or twice during the course
  - Each lecture will have theme
  - 2 students share one theme and should coordinate in presenting the material
  - The presentation must include:
    - Shared goal & motivation, prev. work review
    - Two technical presentations (one student each)
    - Shared conclusion and comparison of the two works
  - Some themes are presented by domain experts

- Read papers and participate at class discussion during paper presentations
  - During the lecture be prepared to ask questions. At the end of each lecture, two discussion leaders are randomly selected: the discussion leader will ask questions to the presenters and lead a 5-minute discussion panel; the quality of the questions & discussion panel will be used for evaluating class participation.
  - The more questions you ask during each lecture, the better!

- Course project [see later]

Course Project Evaluation

- Form your team:
  - 1-2 people per team
  - the quality is judged regardless of the number of people on the team
  - be nice to your partner: do you plan to drop the course?
- Evaluation
  - Quality of the project (including writing)
  - Final ~20 minutes project presentation in class – students will vote your presentation!

Grading policy

- Course project: 50%
  - progress report 5%
  - final report 35%
  - presentation 10%

- Attendance and class participation: 20%
  - See class participation protocol

- Paper presentation (quality, clarity, depth, etc.): 30%

- Late policy project:
  - If 1 day late, 25% off the grade for the project
  - If 2 days late, 50% off the grade for the project
  - Zero credits if more than 2 days

- Collaboration policy
  - Read the student code book, understand what is ‘collaboration’ and what is ‘academic infraction’.
  - Discussing project assignment with each other is allowed, but coding must be done individually
  - Using on line presentation material (slides, etc…) is not allowed in general. Exceptions can be made and individual cases will be discussed with the instructor.

syllabus

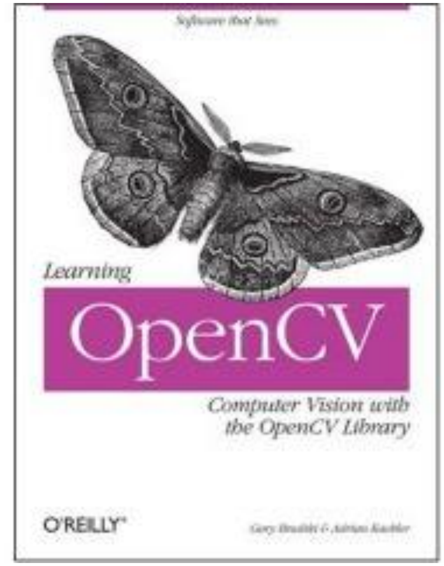
- Syllabus contains the schedule of the course with the list of papers to present:

http://www.stanford.edu/class/archive/cs/cs331b/cs331b.1142/cgi-bin/mediawiki/index.php/Main_Page

- Look at the syllabus page for important dates (e.g., reports due dates) and updates;

- NOTE: the syllabus page is still under construction

Open Computer Vision



The Open Computer Vision Library has > 500 algorithms, documentation and sample code for real time computer vision.

Tutorial documentation is in O'Reilly Book: Learning OpenCV
http://www.amazon.com/Learning-OpenCV-Computer-Vision-Library/dp/0596516134

http://sourceforge.net/projects/opencvlibrary/

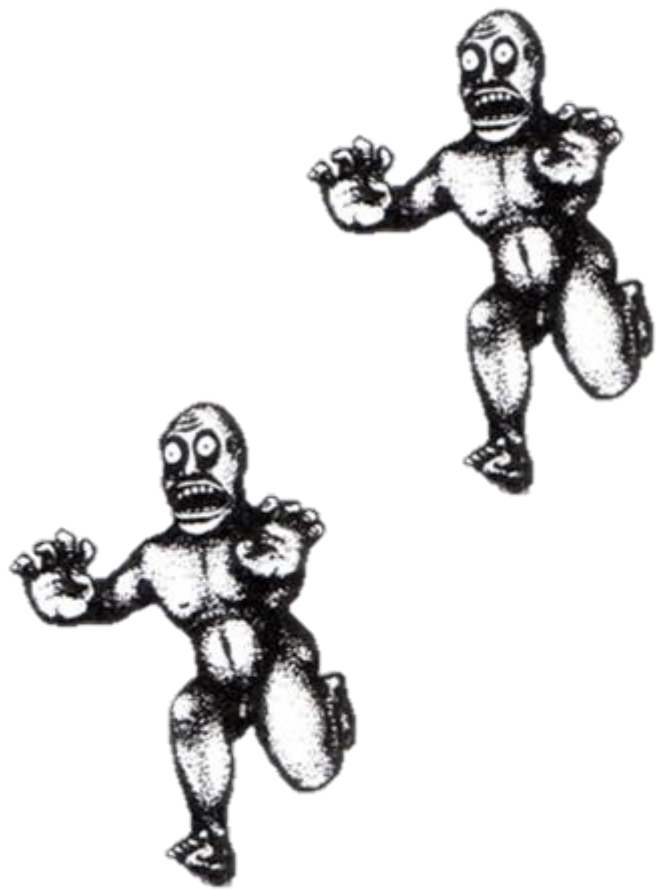
Agenda

- Administrative
  - Requirements
  - Grading policy

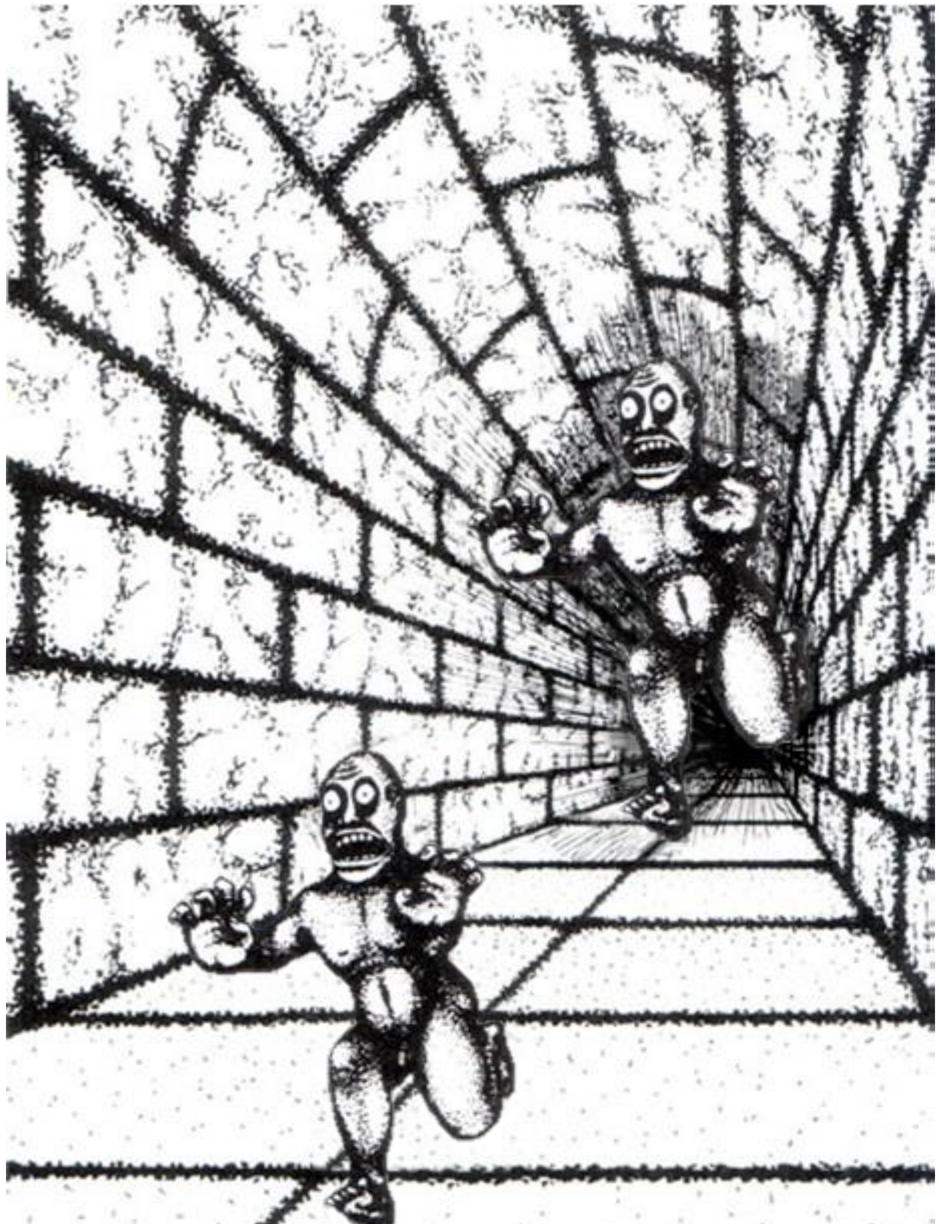
- **Overview of this course**

V. Van Gogh, Bedroom at Arles

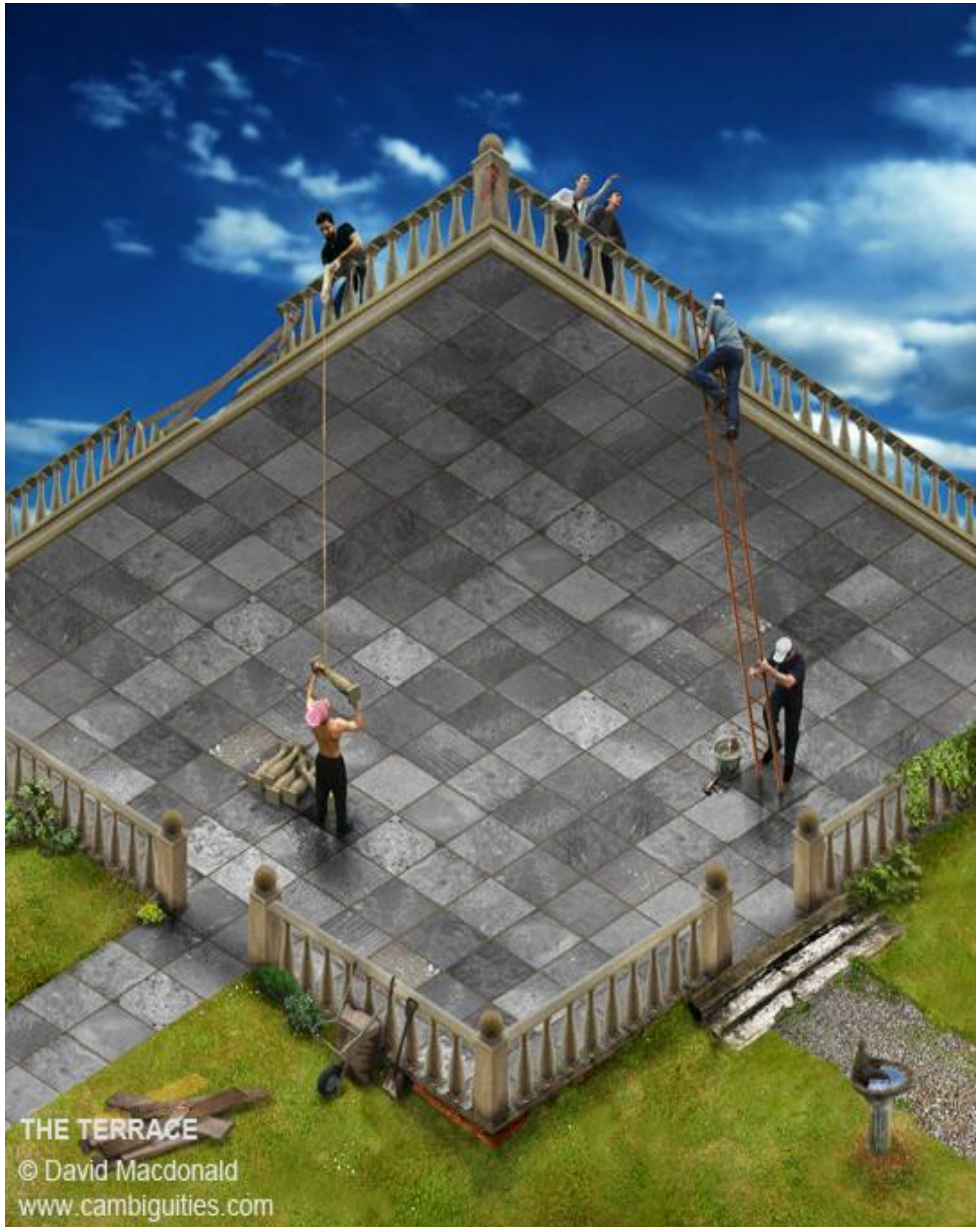
Understanding the 3D world



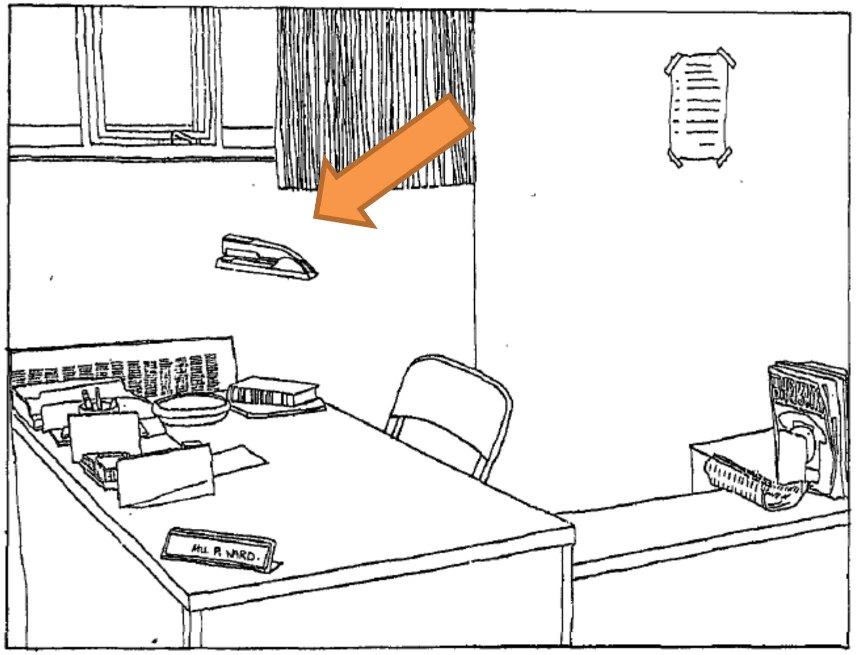
Understanding the 3D world

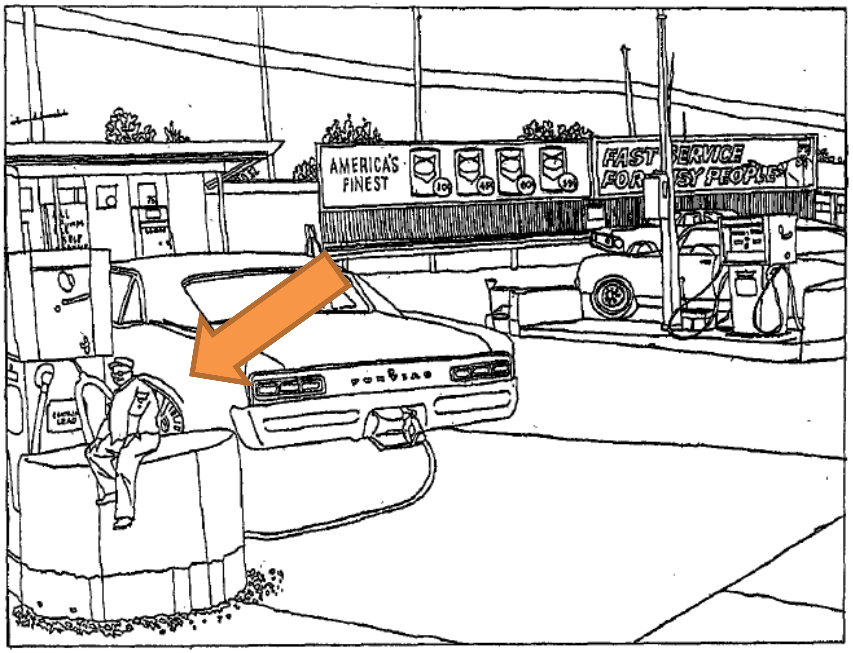


Understanding the 3D world



Understanding the 3D world



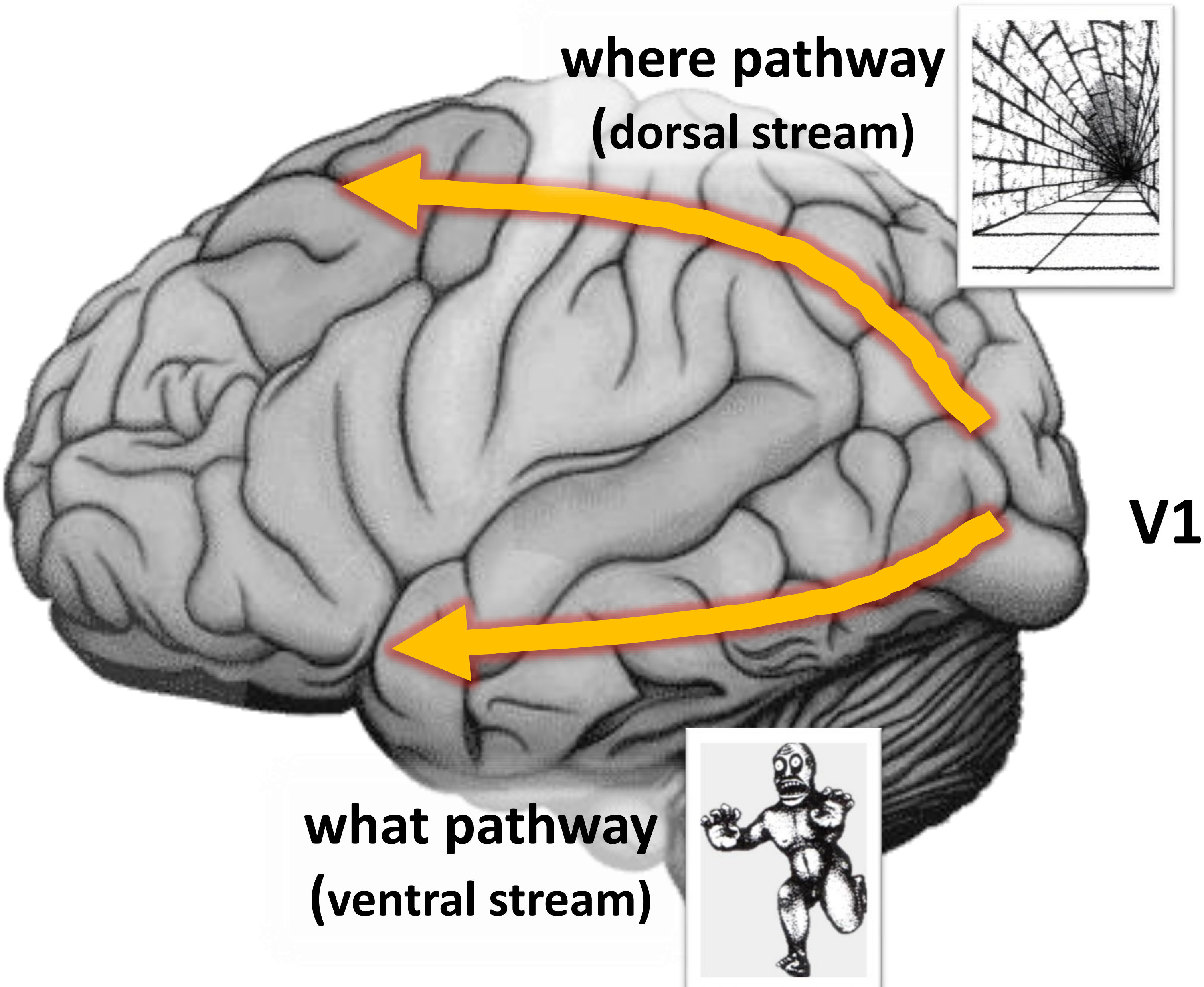


Biederman, Mezzanotte and Rabinowitz, 1982

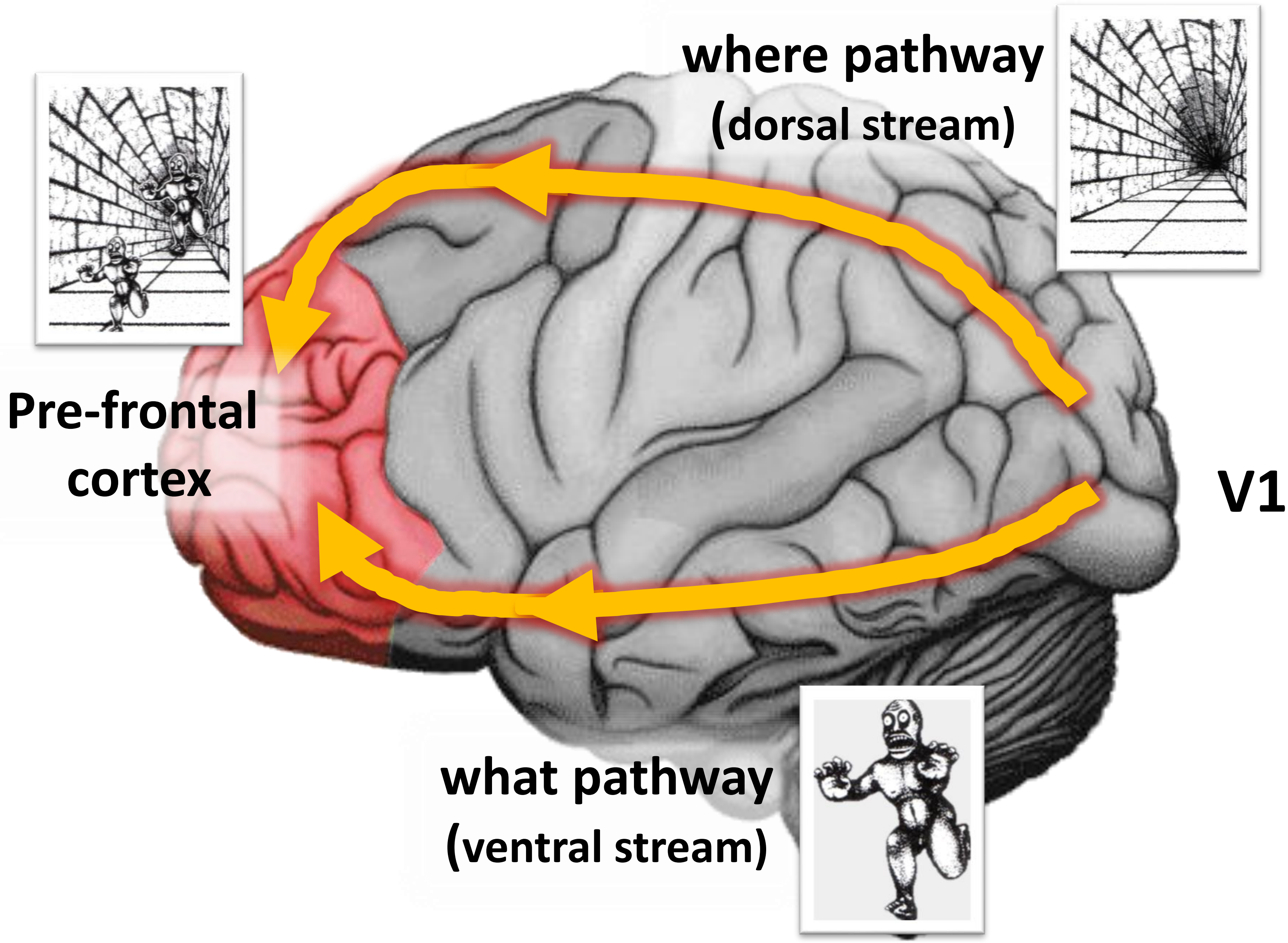
Familiar relationships:

- Location
- Size
- Support
- …

Understanding the 3D world



Understanding the 3D world



From computer vision

WHERE

- 3D shape modeling
- 3D scene reconstruction
- Camera localization

WHAT

- Material classification
- Object recognition
- Scene classification
- Target tracking
- Activity recognition

From computer vision

WHERE

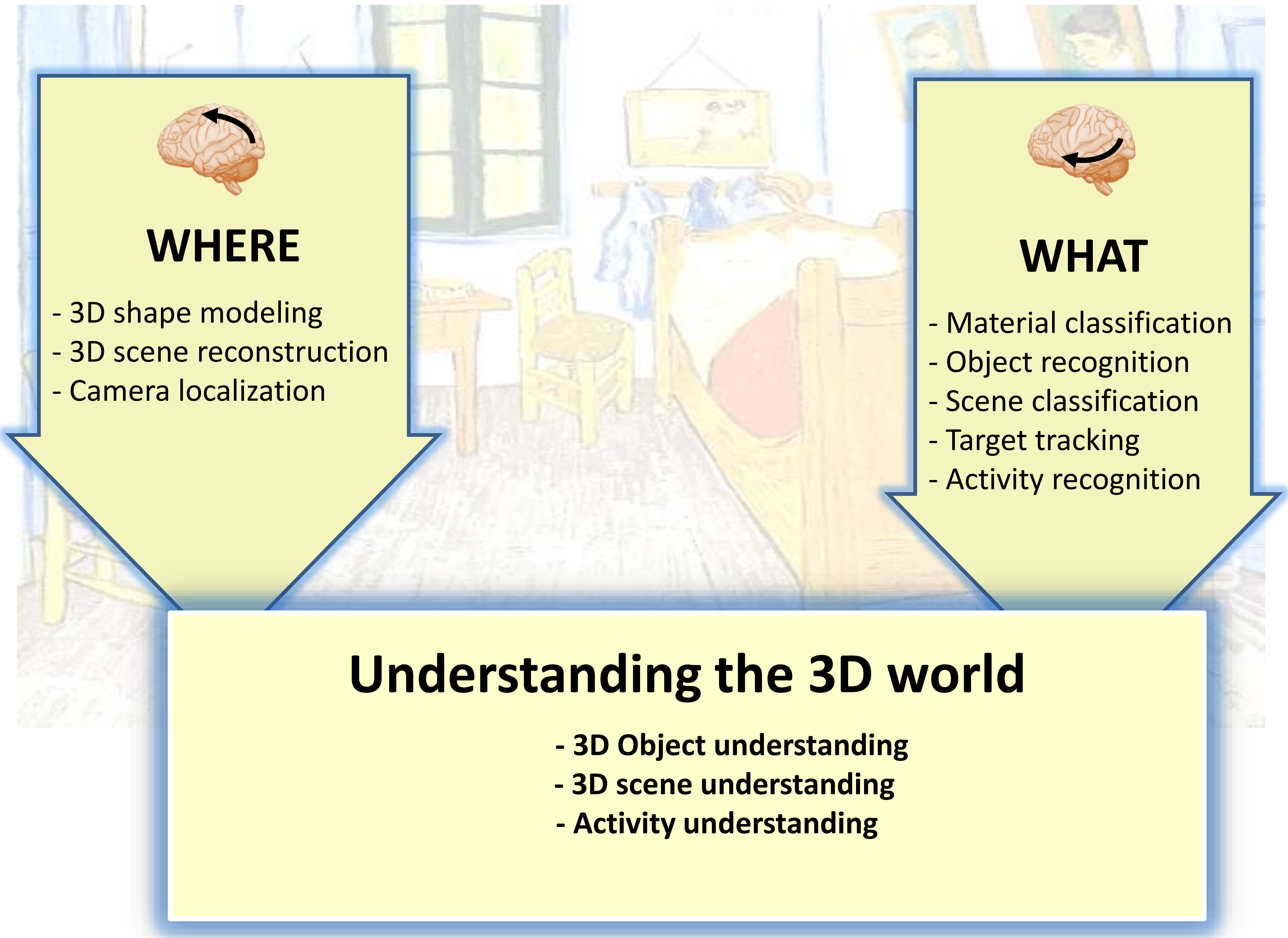
- 3D shape modeling
- 3D scene reconstruction
- Camera localization

WHAT

- Material classification
- Object recognition
- Scene classification
- Target tracking
- Activity recognition

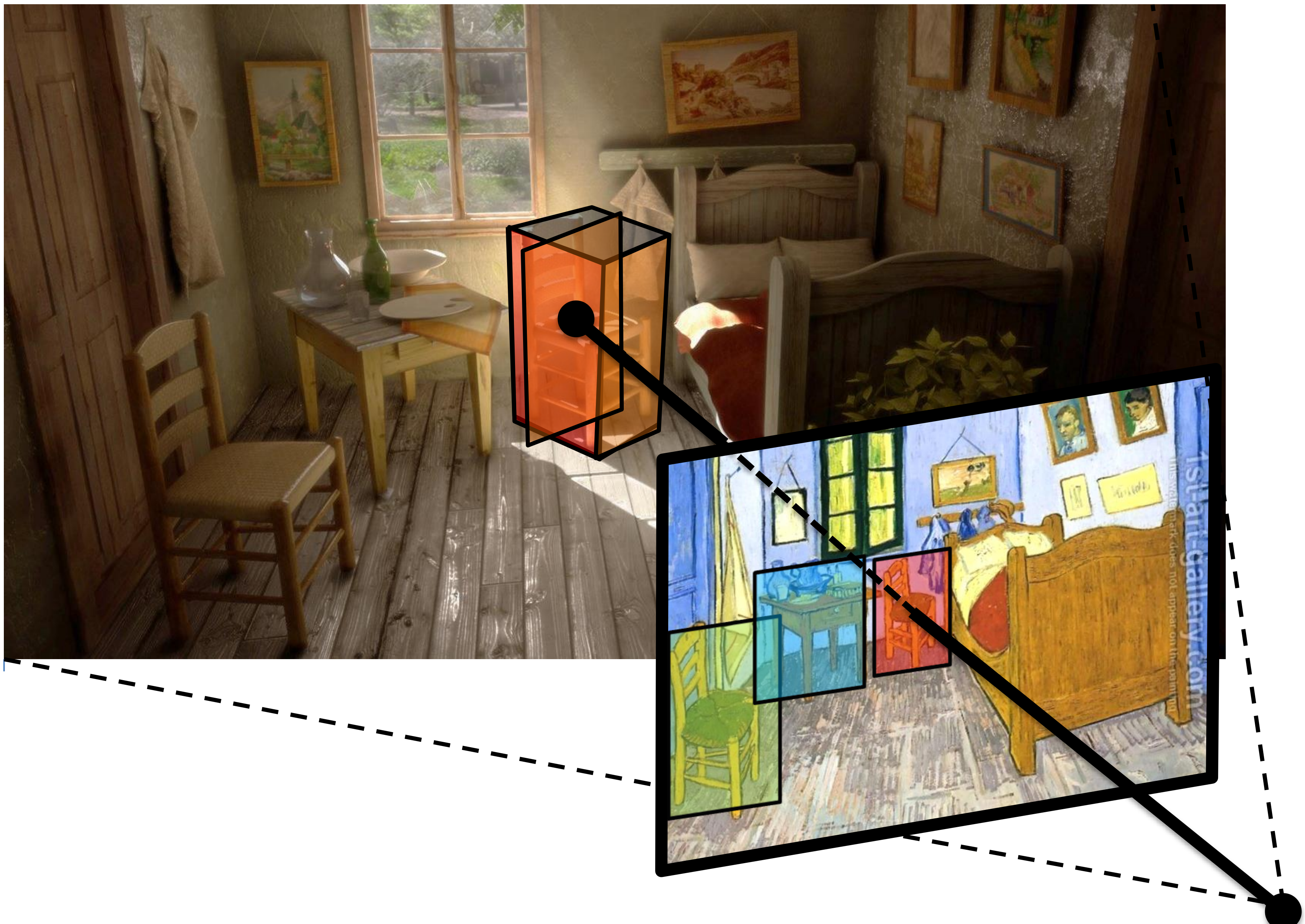
Understanding the 3D world

- **3D Object understanding**
- **3D scene understanding**
- **Activity understanding**

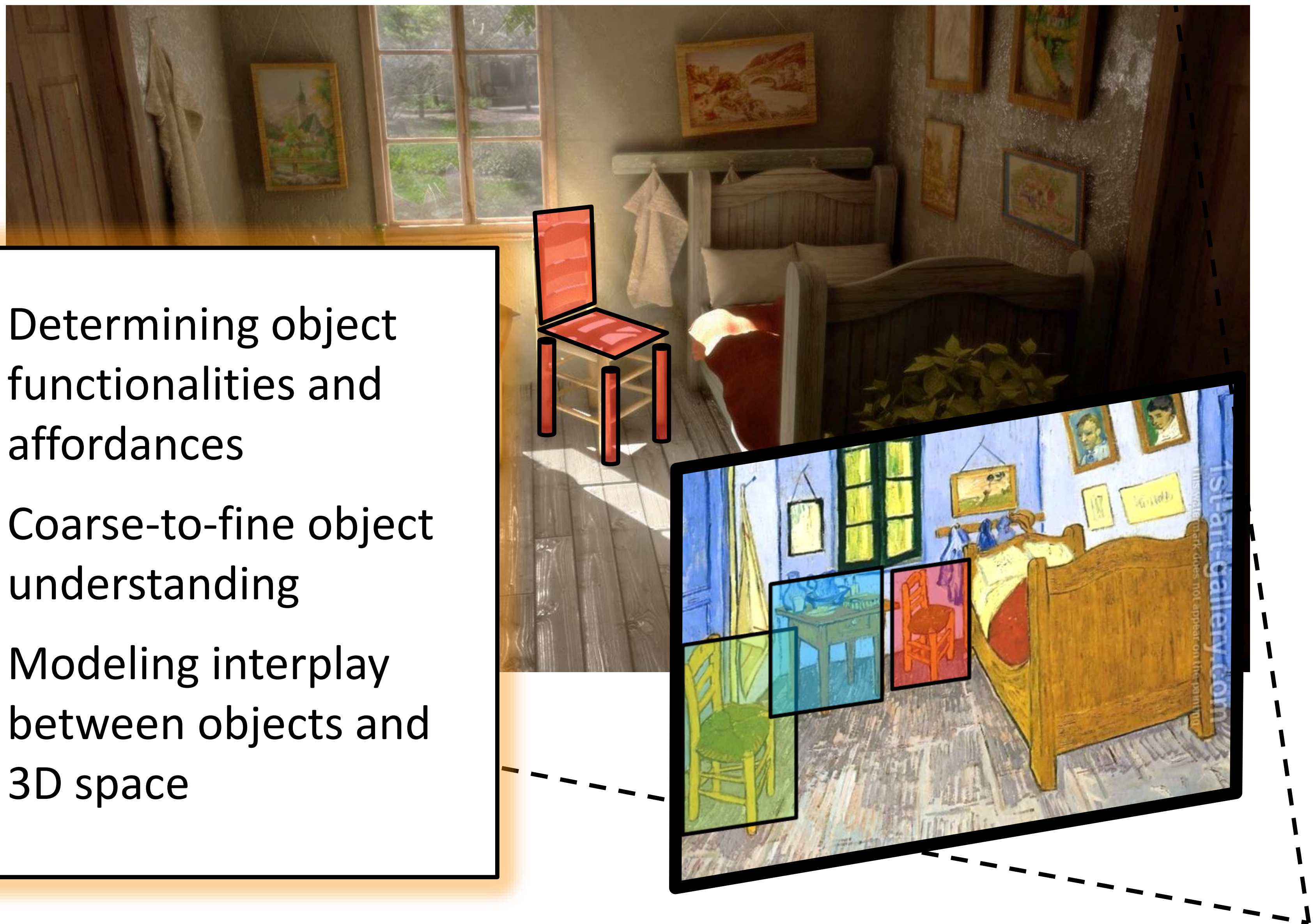


WHERE
- 3D shape modeling
- 3D scene reconstruction
- Camera localization
WHAT
- Material classification
- Object recognition
- Scene classification
- Target tracking
- Activity recognition
Understanding the 3D world
- 3D Object understanding
- 3D scene understanding
- Activity understanding

3D object understanding



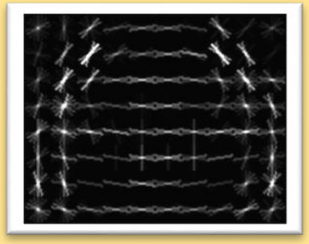
3D object understanding



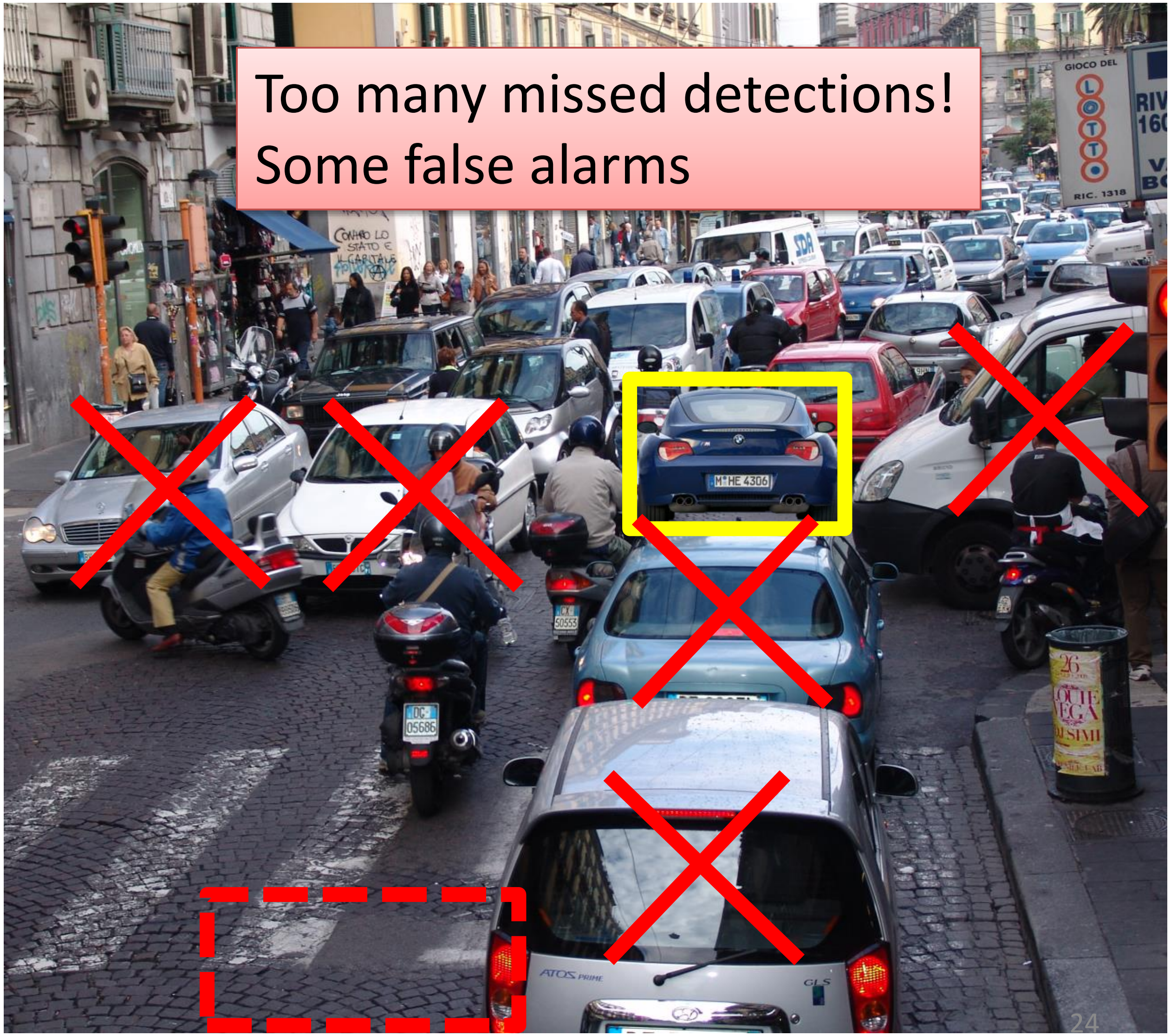
- Determining object functionalities and affordances
- Coarse-to-fine object understanding
- Modeling interplay between objects and 3D space

State of the art object detection

“Car” model

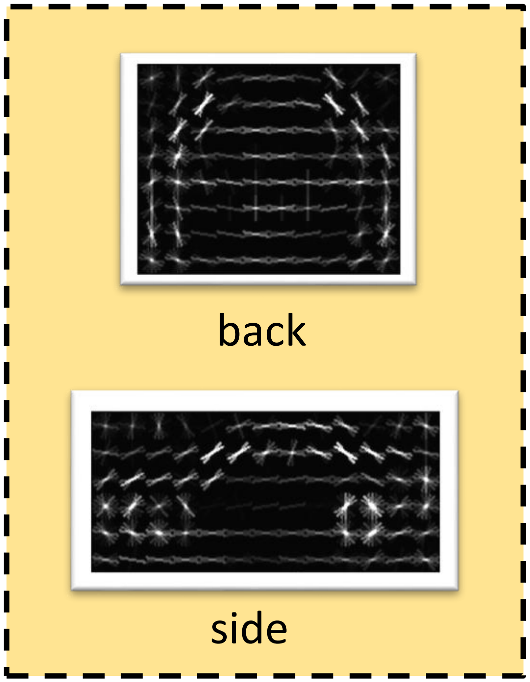


Turk & Pentland, 91
Poggio et al., 93
LeCun et al. 98
Amit and Geman, 99
Shi & Malik, 00
Viola & Jones, 00
Vasconcelos ‘00
Felzenszwalb & Huttenlocher 00
Belongie & Malik, 02
Ullman et al. 02
Argawal & Roth, 02
Weber et al., 00
Fergus et al., 03
Torralba et al., 03
Fei-Fei et al., 04
Leibe et al., 04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



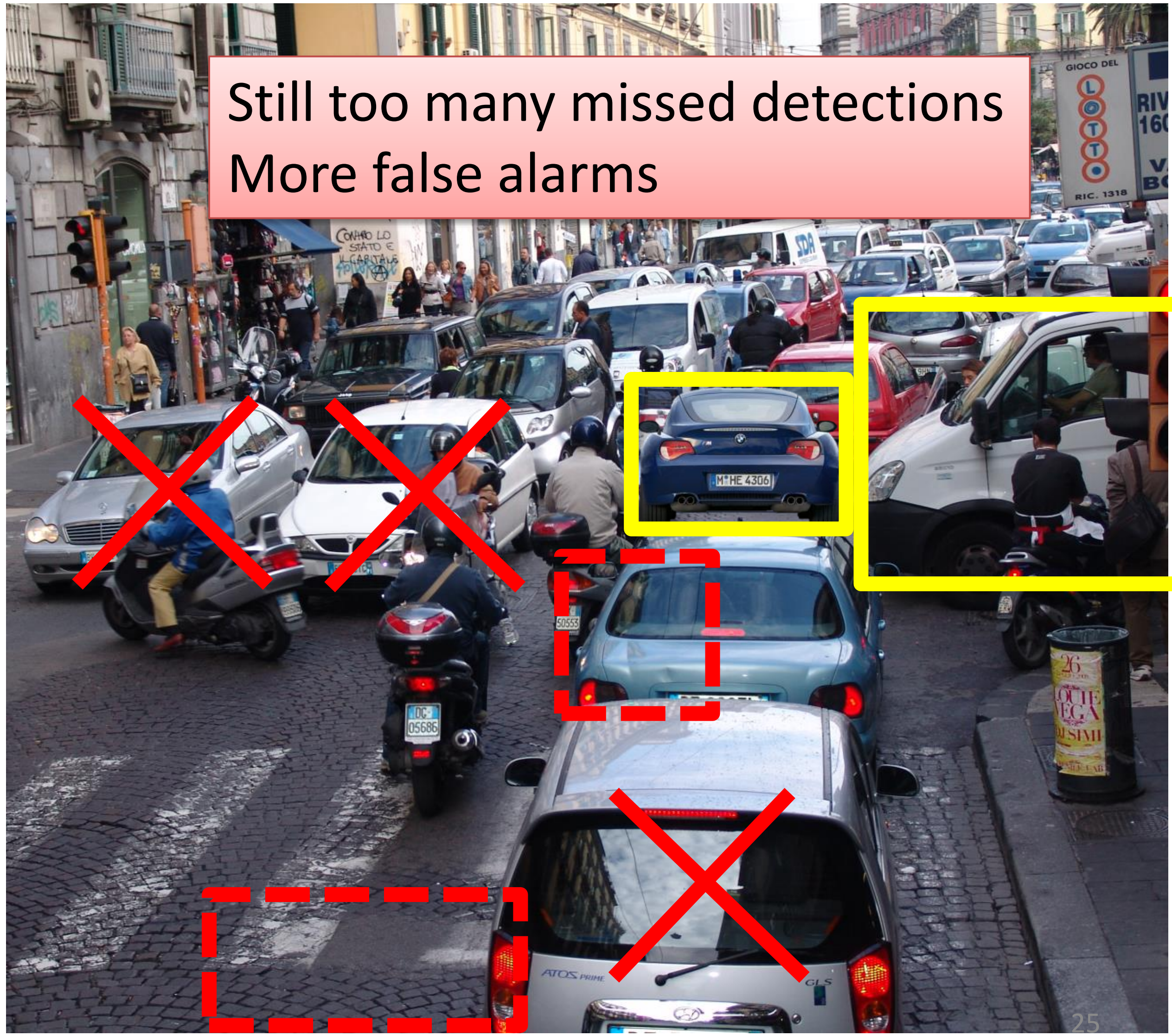
State of the art object detection

"Car" model



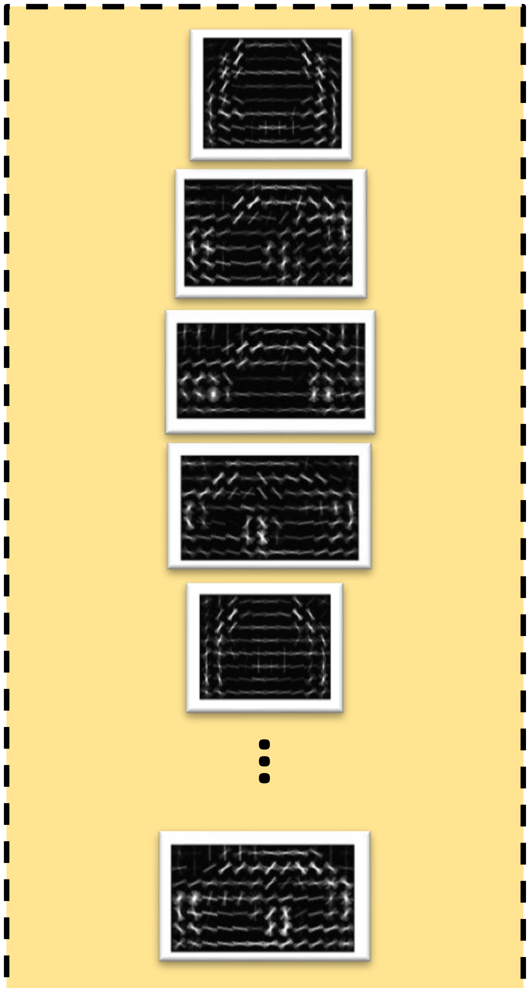
mixture model

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

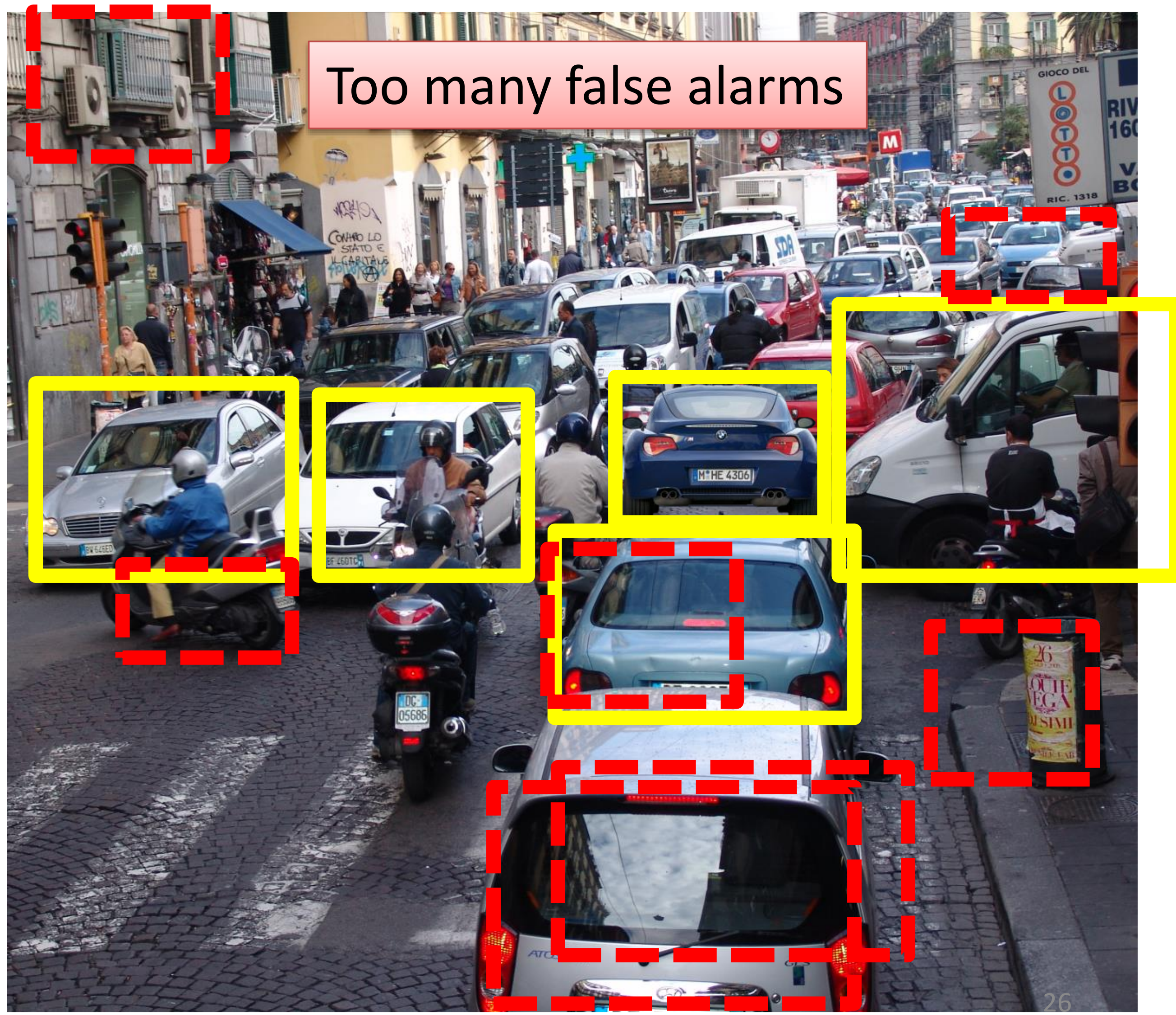


State of the art object detection

“Car” model

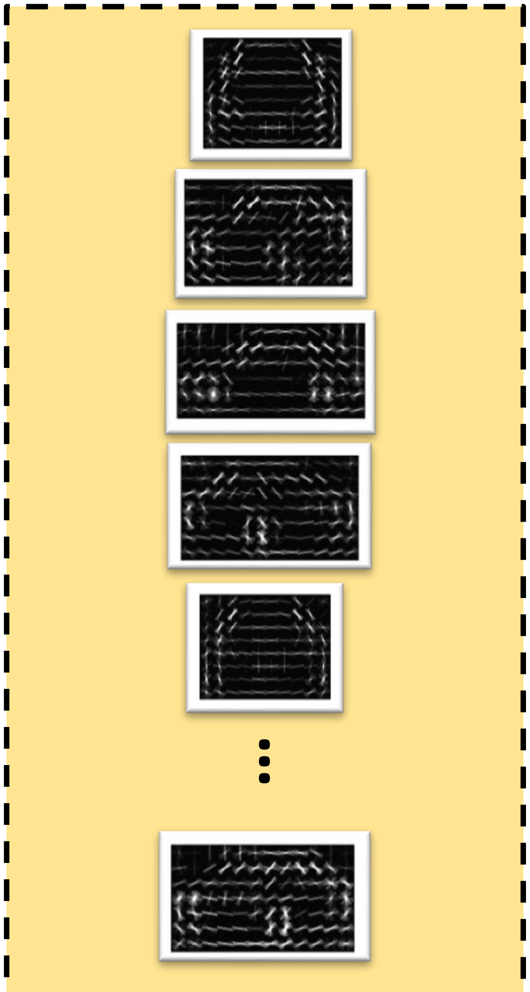


mixture model

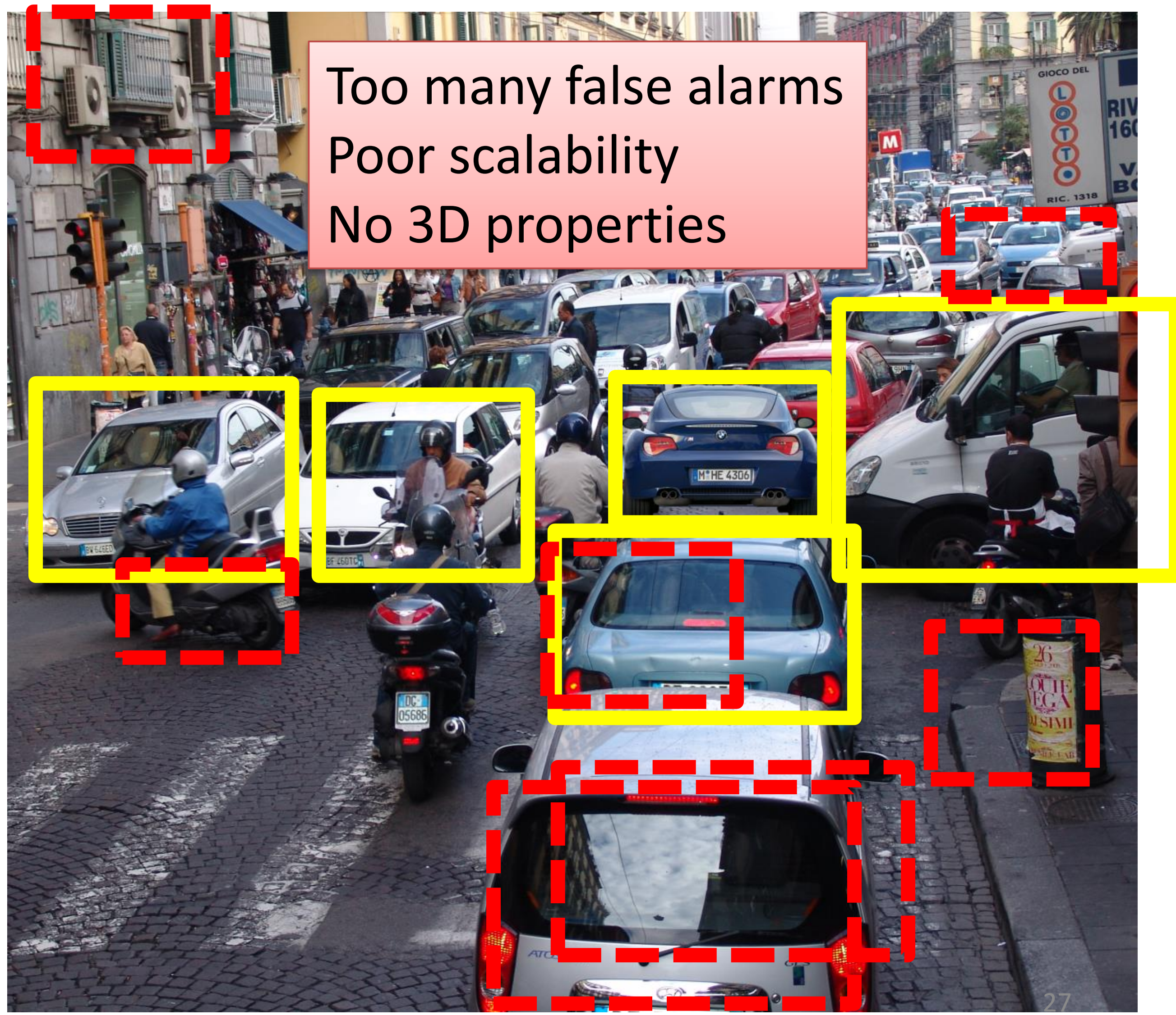


State of the art object detection

“Car” model

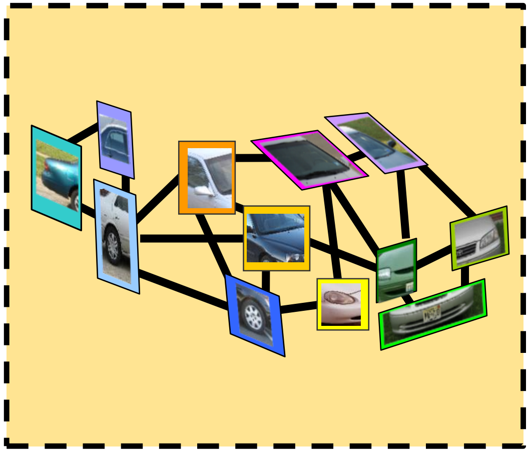


mixture model



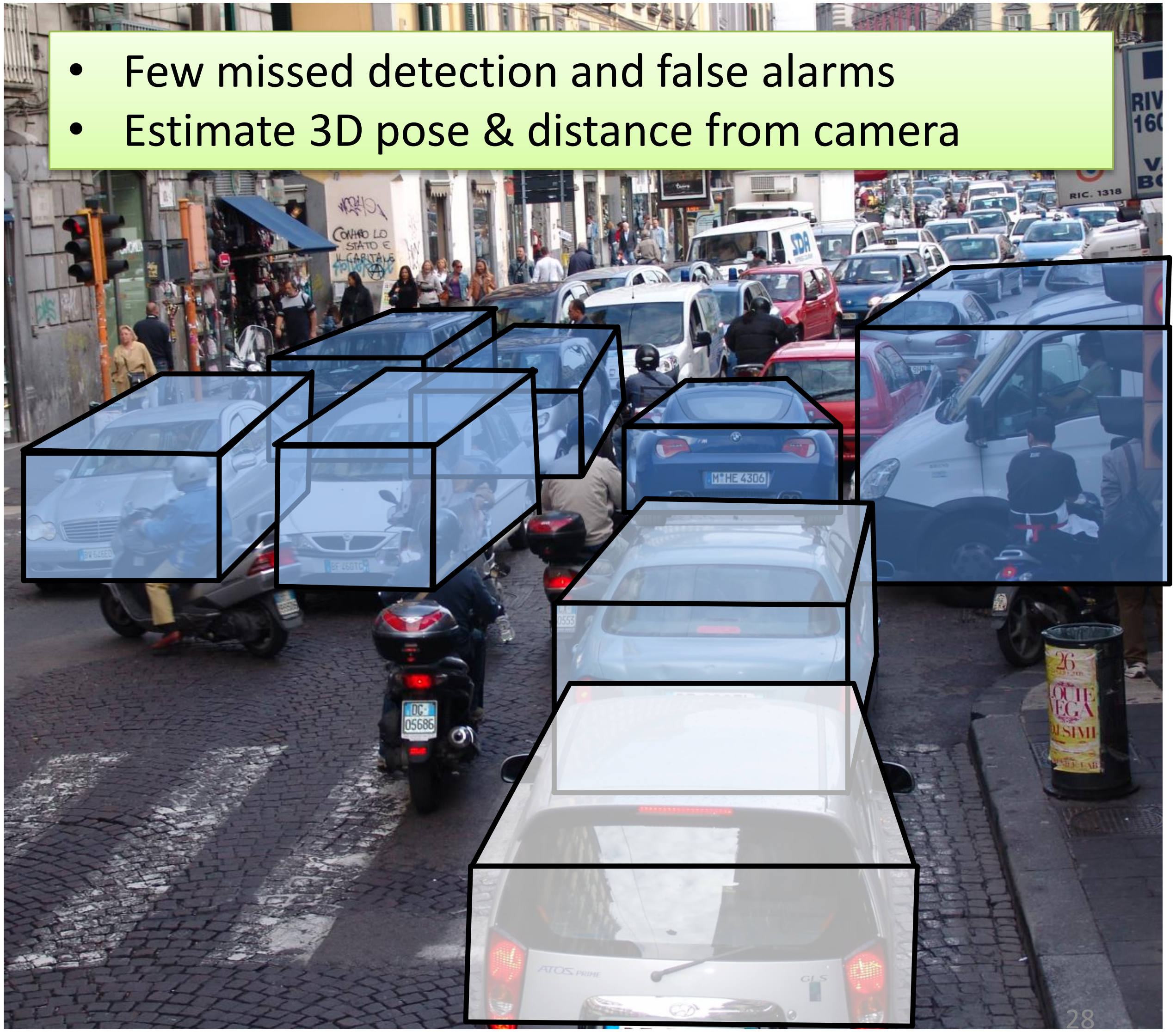
3D object detection

"Car" model

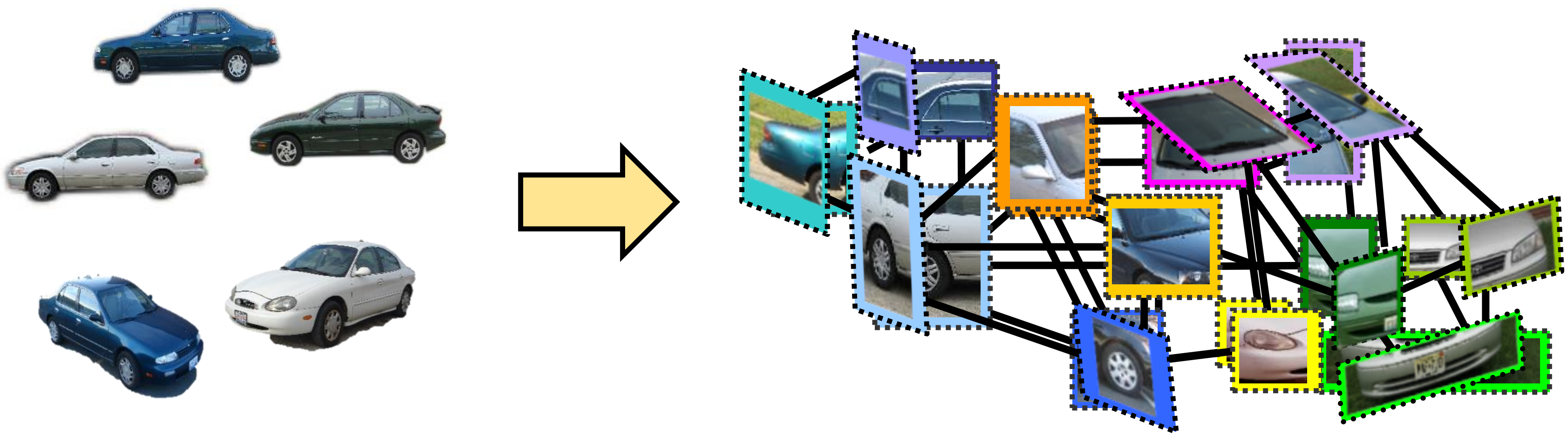


- Savarese et al., ICCV 07
- Su et al., ICCV 2009
- Sun, et al., CVPR 2009
- Yu & Savarese, CVPR 2012

- 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



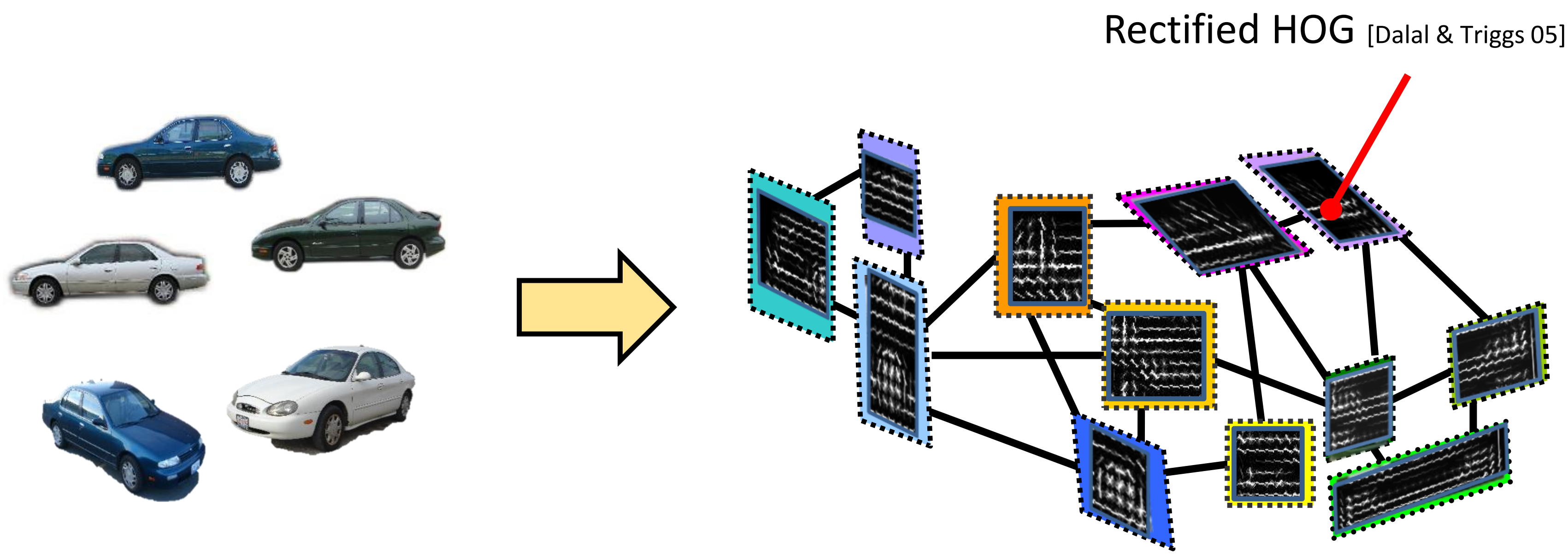
3D object representation



Savarese et al., ICCV 07
Sun et al., ICCV 09

- Object is represented by a collection of parts
- Parts relationship are learnt from training images

3D object representation

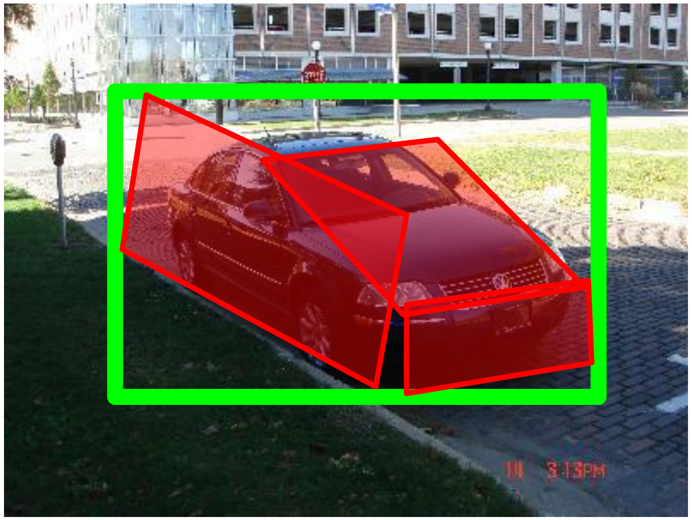


- Object is represented by a collection of parts
- Parts relationship are learnt from training images

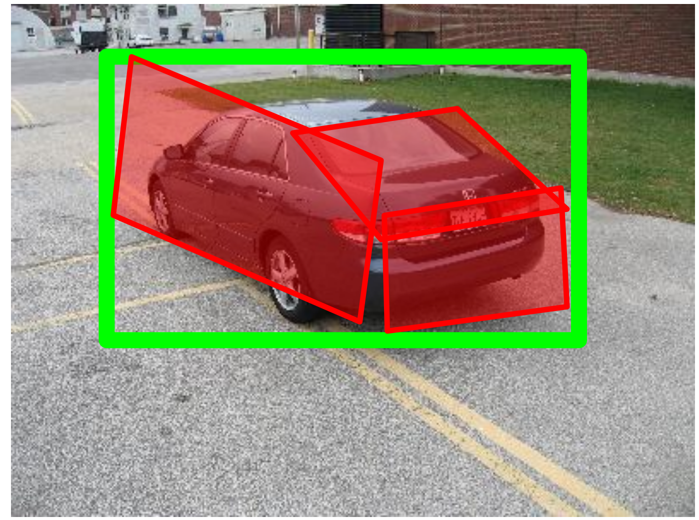
Results

Yu & Savarese, CVPR 2012

CAR a=330 e=15 d=7

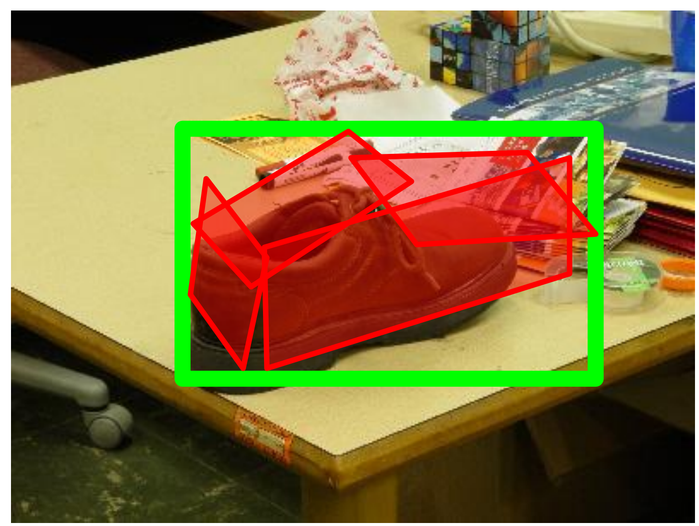


CAR a=150 e=15 d=7



MOUSE a=300 e=45 d=23

SHOE a=240 e=45 d=11



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

Results

Yu & Savarese, CVPR 2012

CHAIR a=0 e=30 d=7

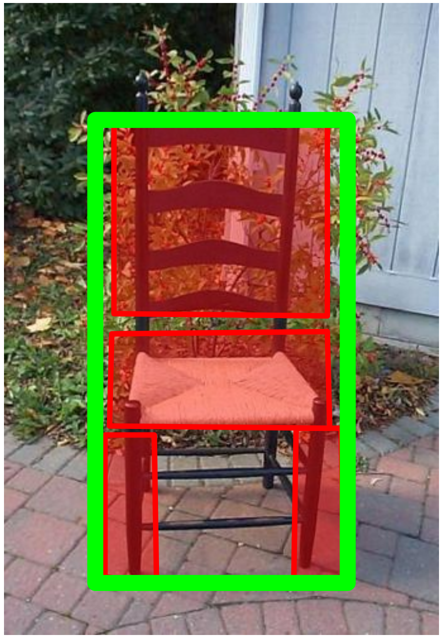
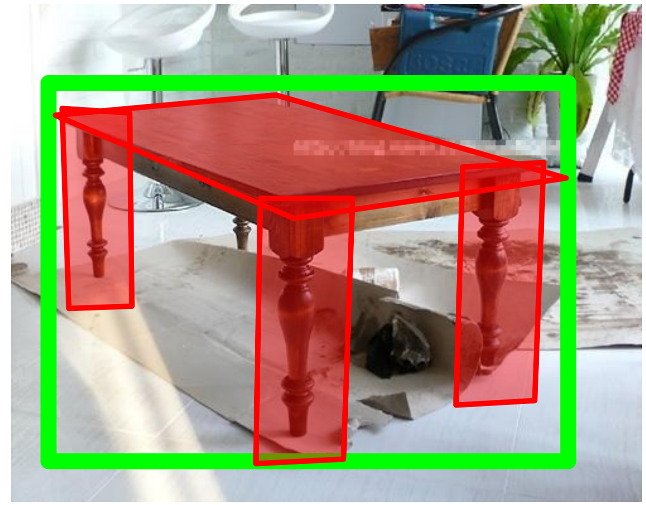
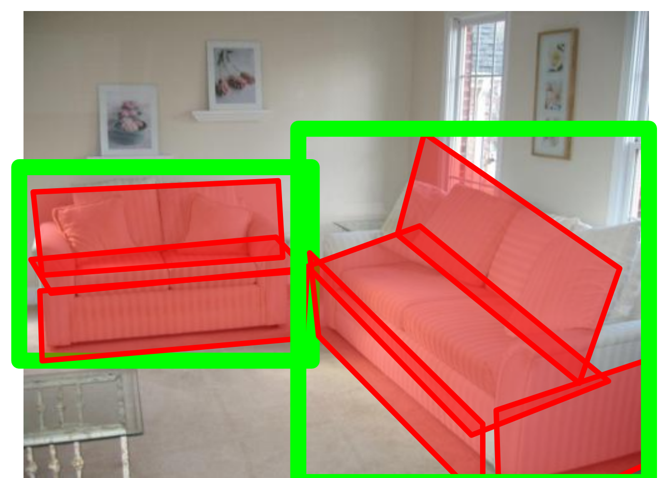


TABLE a=60 e=15 d=2



BED a=30 e=15 d=2.5

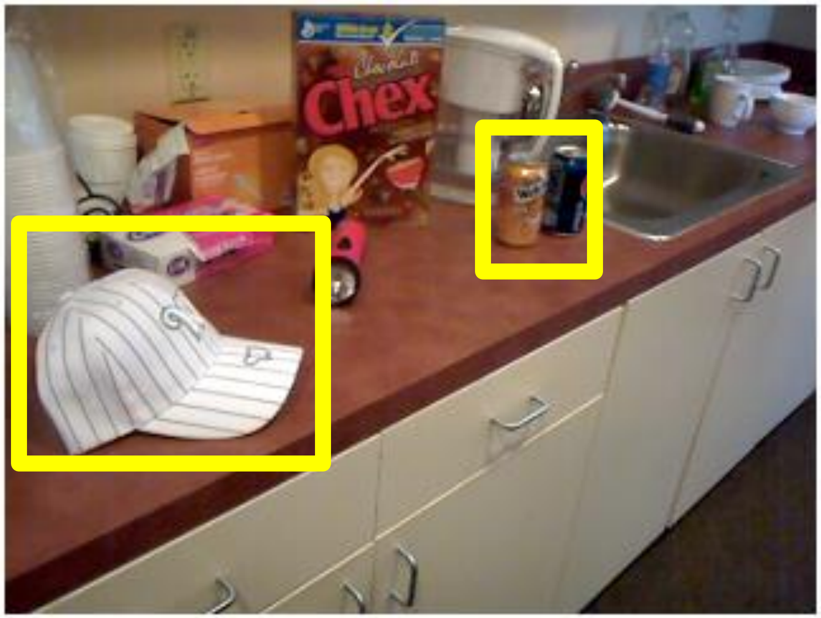
SOFA a=345 e=15 d=3.5
a=60 e-30 d=2.5

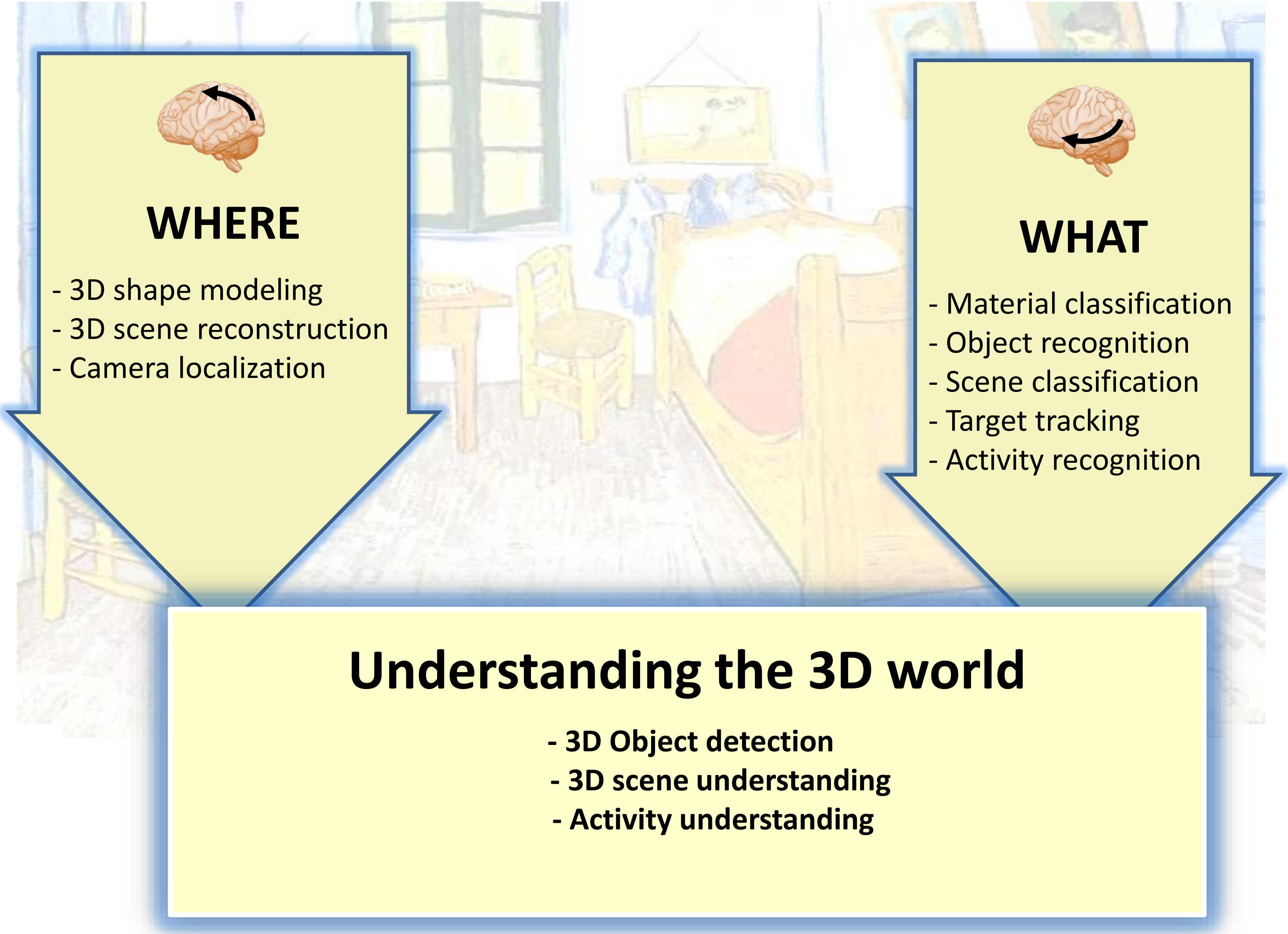


ImageNet dataset [Deng et al. 2010]

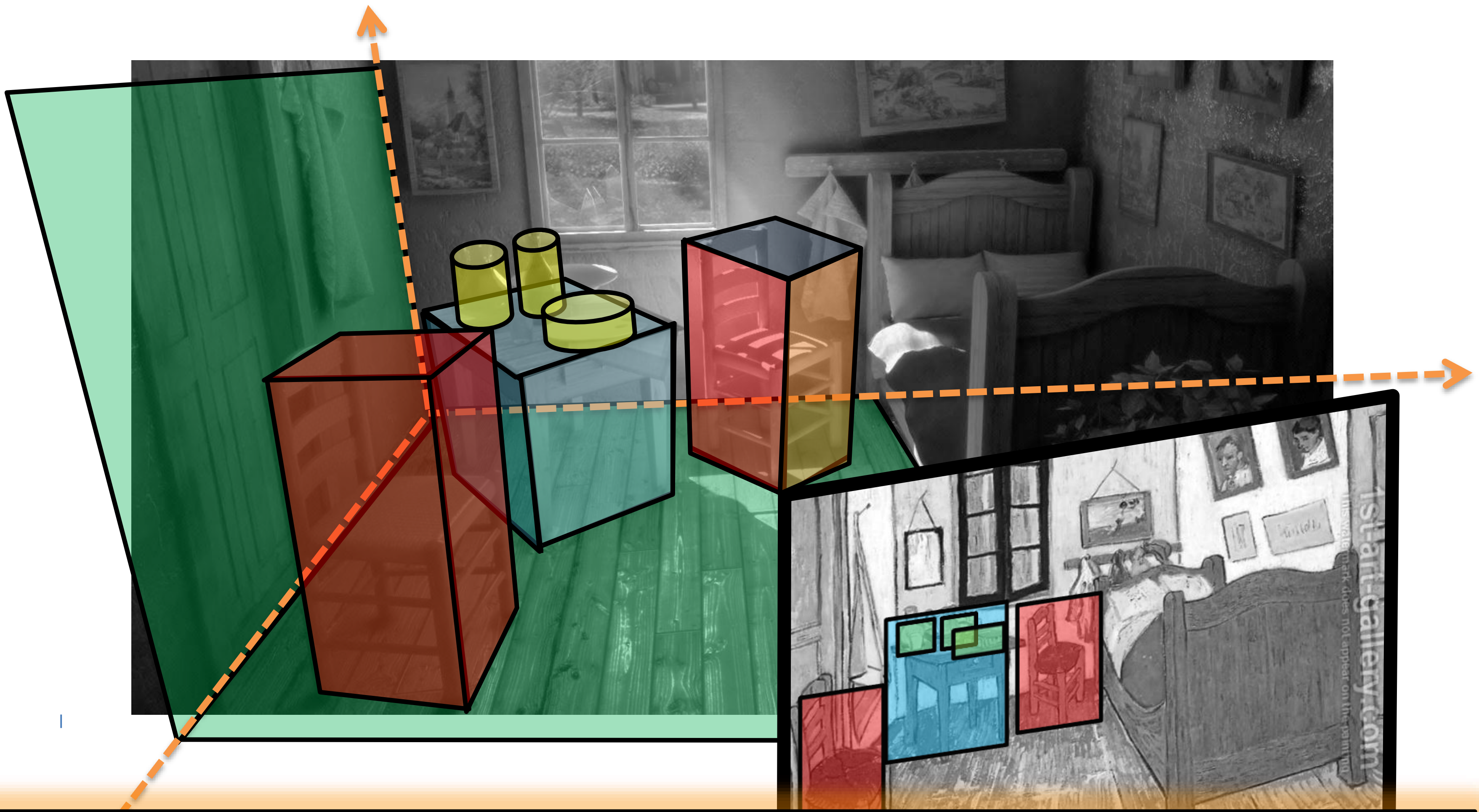
Object detection from D-RGB data

"Object Recognition with Hierarchical Kernel Descriptors", Liefeng Bo, Kevin Lai, Xiaofeng Ren, and Dieter Fox, CVPR, 2011




WHERE
- 3D shape modeling
- 3D scene reconstruction
- Camera localization
WHAT
- Material classification
- Object recognition
- Scene classification
- Target tracking
- Activity recognition
Understanding the 3D world
- 3D Object detection
- 3D scene understanding
- Activity understanding

Coherent 3D scene understanding



- Objects affects perception of 3D space
- 3D space affects perception of objects
- Interplay between objects and 3D space

Look-Alikes by Joan Steiner

Credit slide by A. Torralba

Understanding the 3D space

- 3d point clouds

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

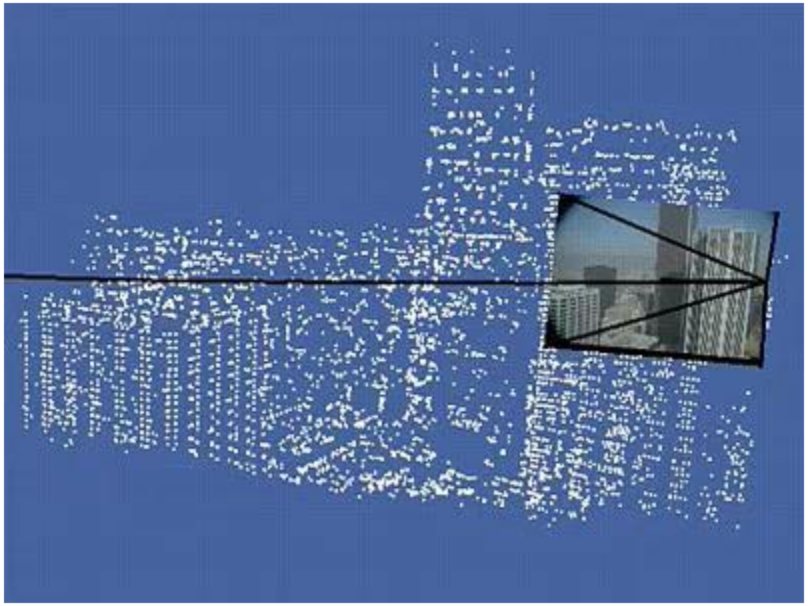
Levoy et al., 00
Hartley & Zisserman, 00
Dellaert et al., 00
Savarese & Perona, 01
Savarese et al., 02
Rusinkiewic et al., 02
Nistér, 04
Brown & Lowe, 04

Savarese et al., 04
Schindler et al., 04
Lourakis & Argyros, 04
Colombo et al., 05
Savarese et al., IJCV 05
Savarese et al., IJCV 06
Saxena et al., 07-09

Snavely et al., 06-08
Schindler et al., 08
Agarwal et al., 09
Frahm et al., 10
Golparvar-Fard, et al. JAEI 10
Pandey et al. IFAC , 2010
Pandey et al. ICRA 2011

Understanding the 3D space

- 3d point clouds



Courtesy of Oxford Visual Geometry Group

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
Savarese & Perona, 01
Savarese et al., 02
Rusinkiewic et al., 02
Nistér, 04
Brown & Lowe, 04

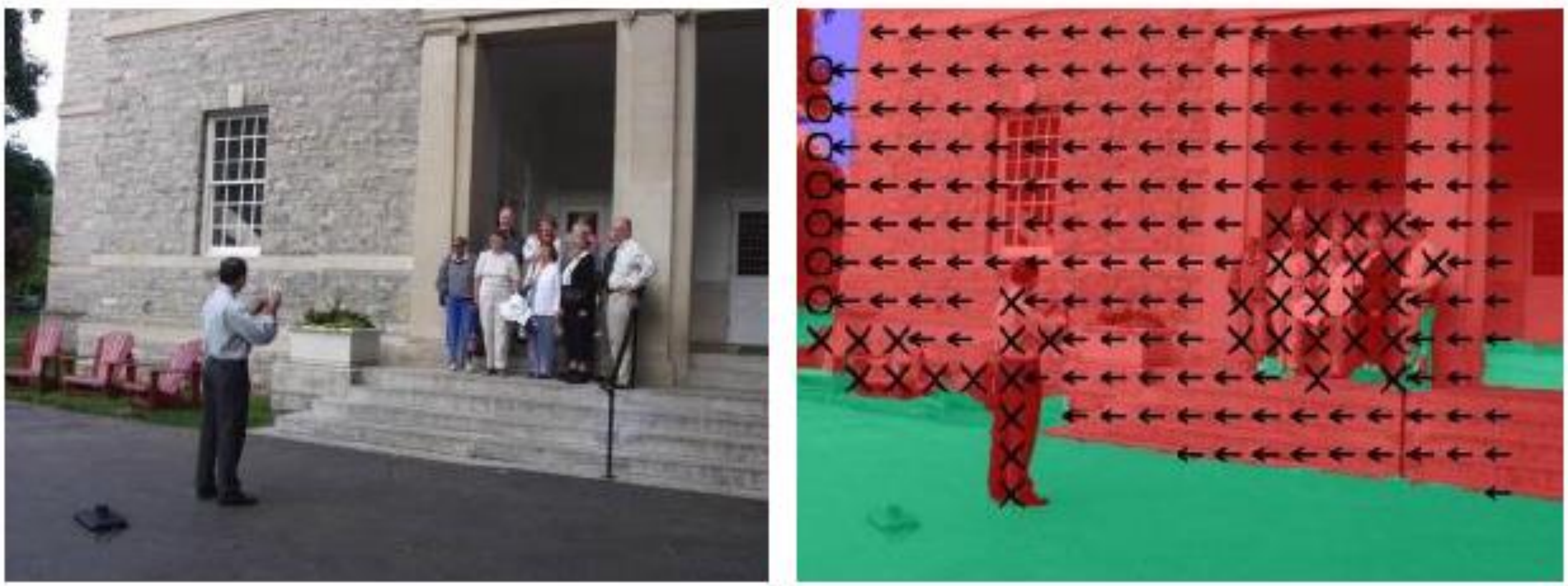
Savarese et al., 04
Schindler et al., 04
Lourakis & Argyros, 04
Colombo et al., 05
Savarese et al., IJCV 05
Savarese et al., IJCV 06
Saxena et al., 07-09

Snavely et al., 06-08
Schindler et al., 08
Agarwal et al., 09
Frahm et al., 10
Golparvar-Fard, et al. JAEI 10
Pandey et al. IFAC , 2010
Pandey et al. ICRA 2011

Understanding the 3D space

- 3d point clouds
- Retinotopic maps

Hoiem et al. 05



Understanding the 3D space

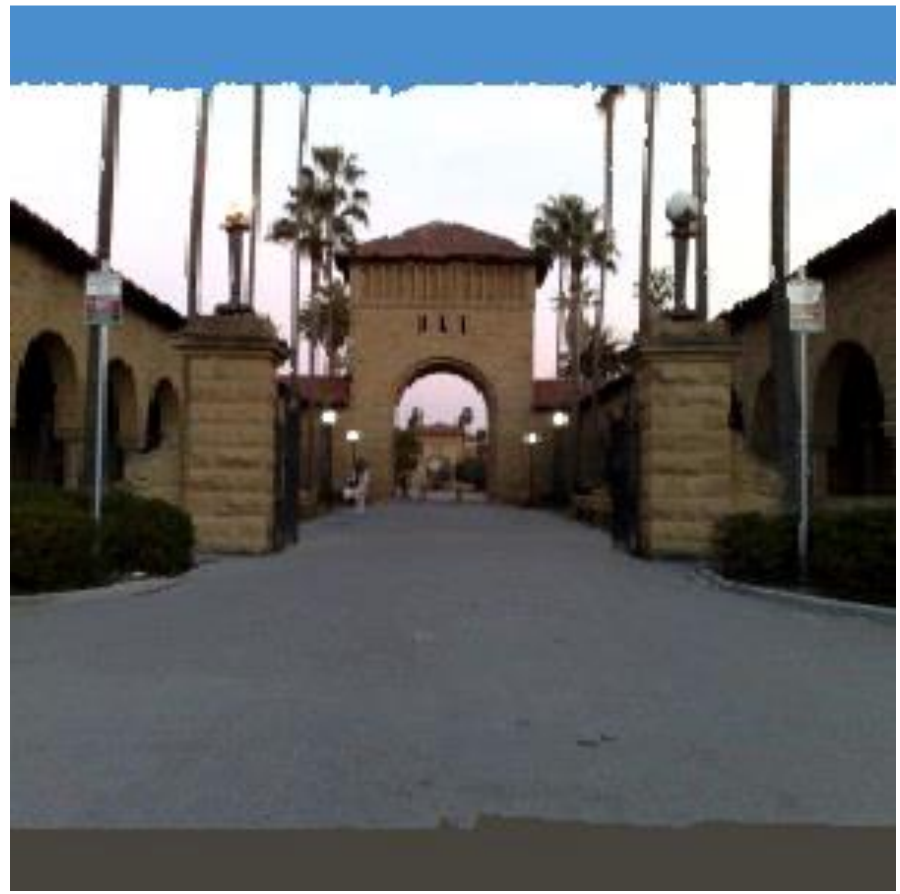
- 3d point clouds
- Retinotopic maps

Hoiem et al. 05

Understanding the 3D space

- 3d point clouds
- Retinotopic maps

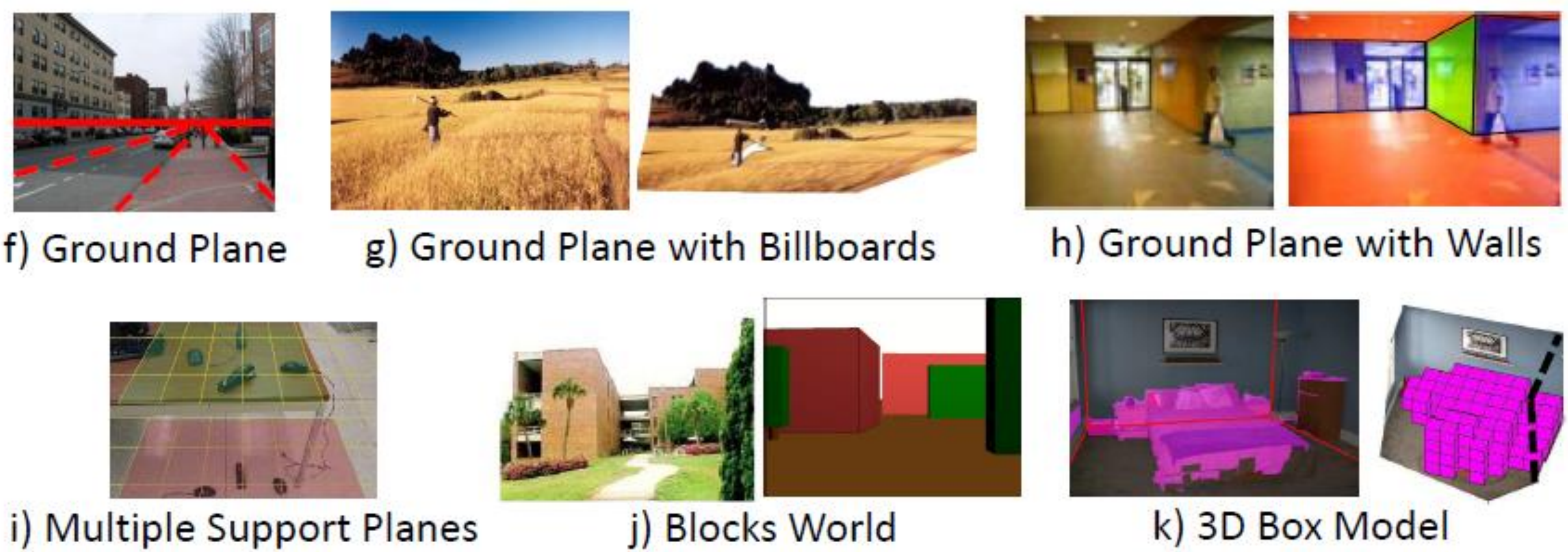
Saxena et al., 07-09



Make3D

Understanding the 3D space

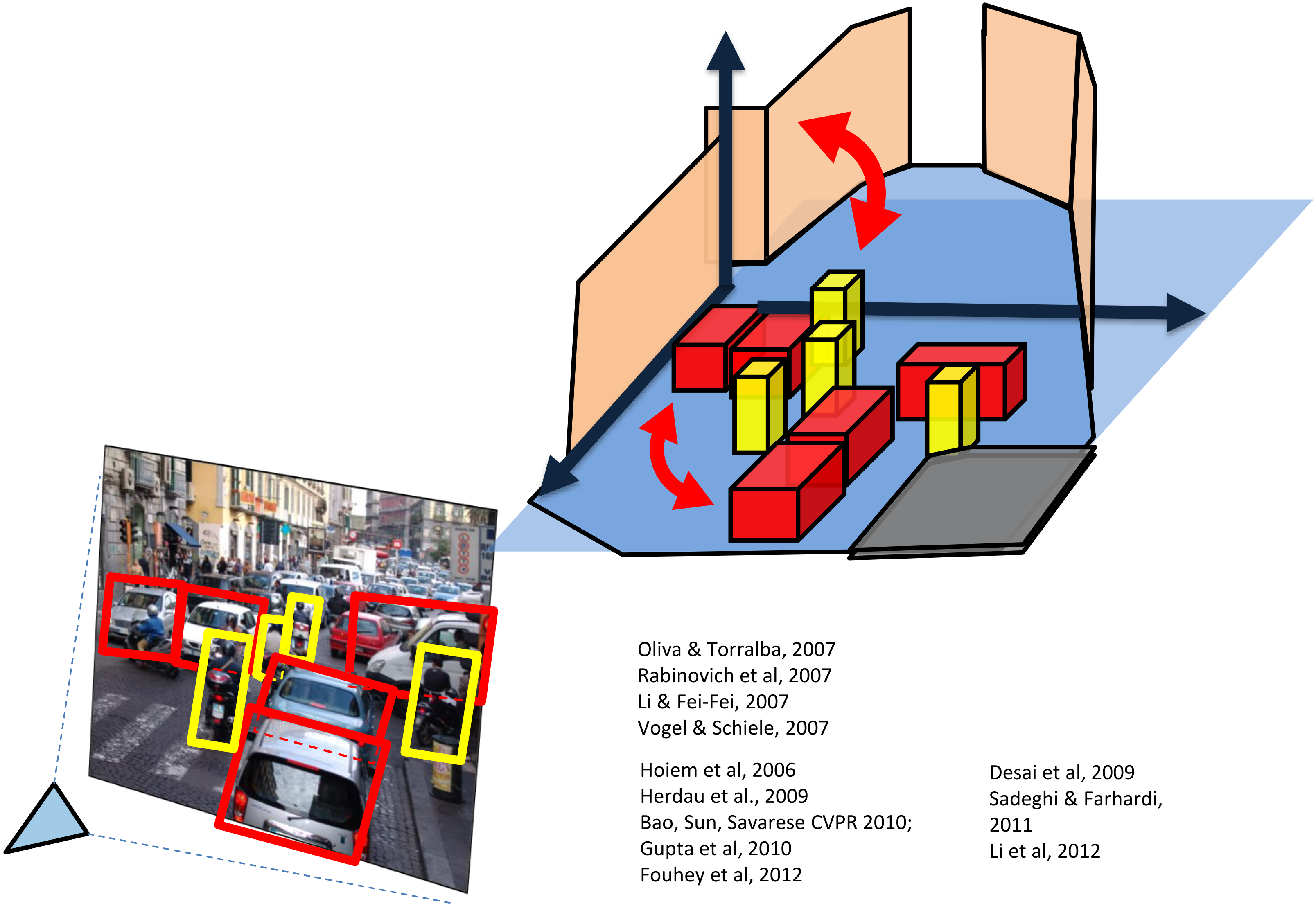
- 3d point clouds
- Retinotopic maps
- Complex models



Examples of box models

Vedau et al. 07

Modeling interaction objects-3D space

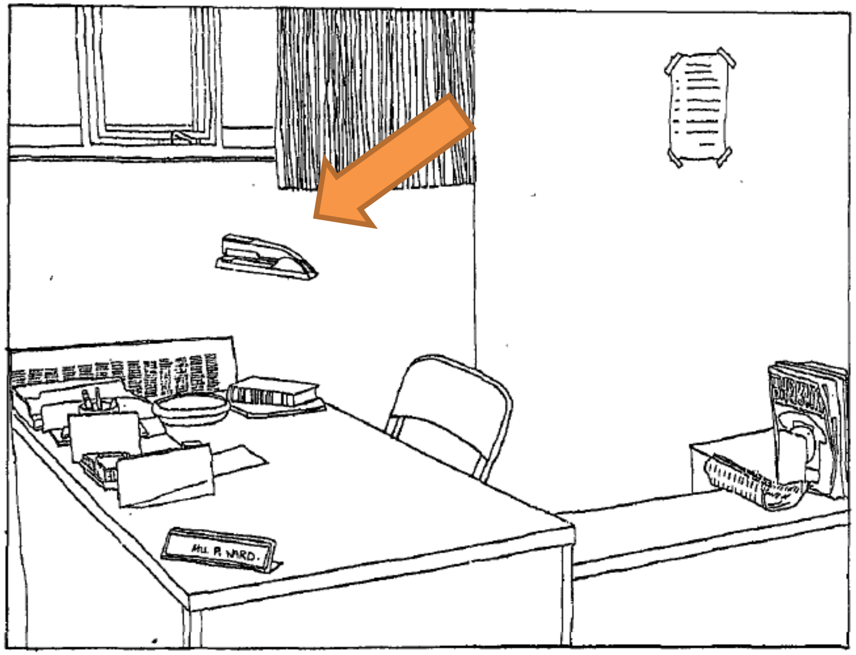


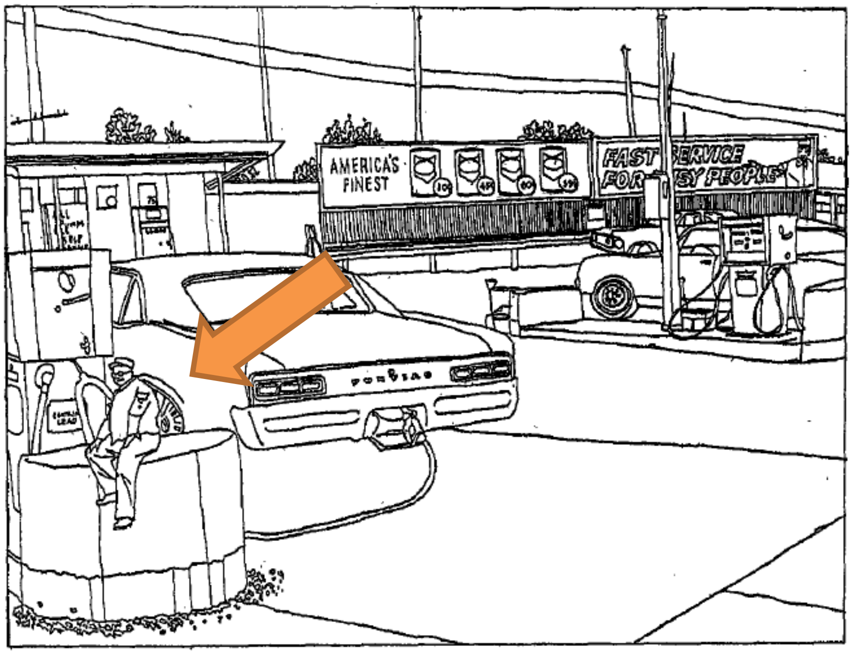
Oliva & Torralba, 2007
Rabinovich et al, 2007
Li & Fei-Fei, 2007
Vogel & Schiele, 2007

Hoiem et al, 2006
Herdau et al., 2009
Bao, Sun, Savarese CVPR 2010;
Gupta et al, 2010
Fouhey et al, 2012

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

Modeling interaction objects-3D space



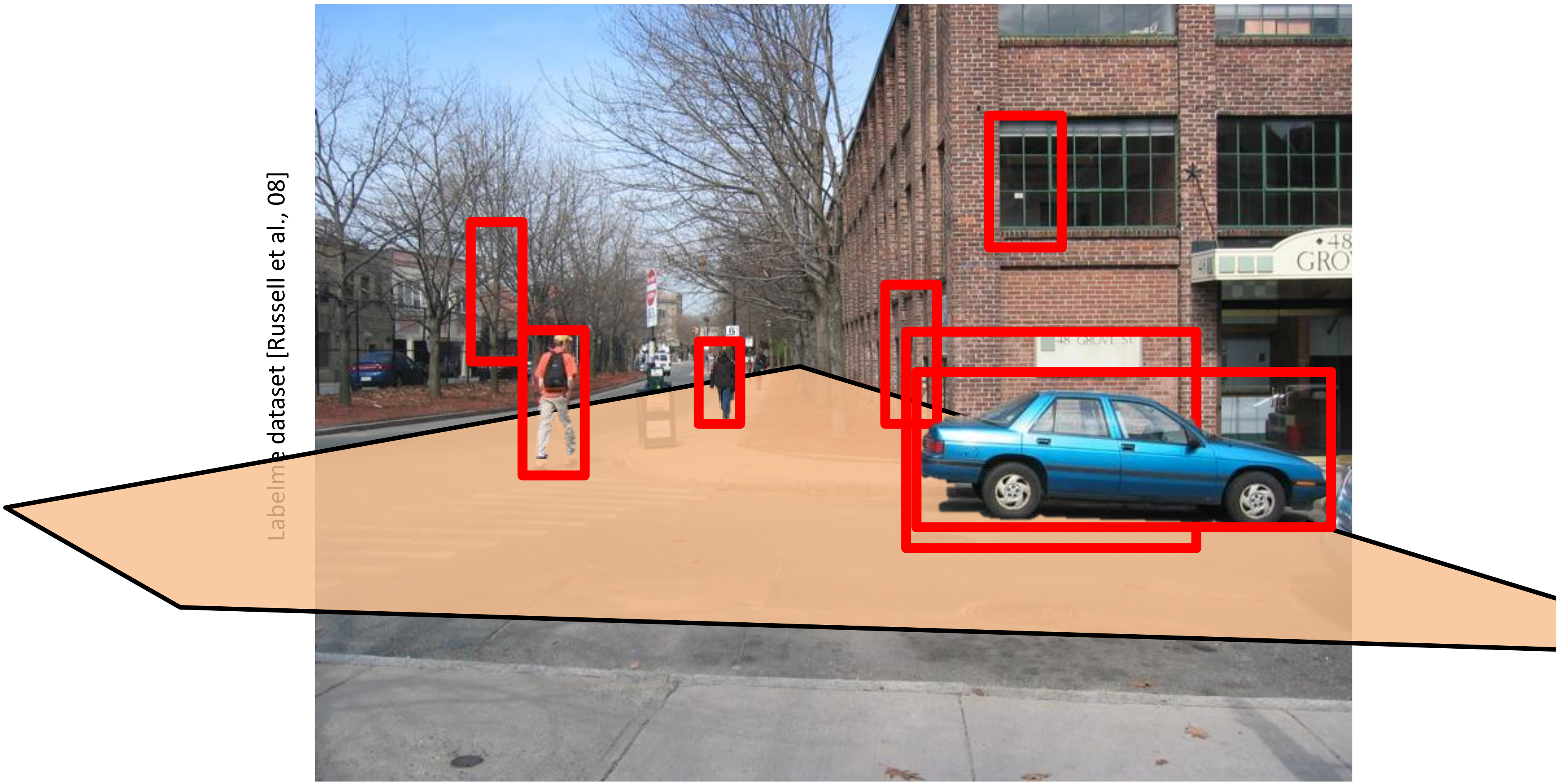


Biederman, Mezzanotte and Rabinowitz, 1982

Modeling interaction objects-3D space

Bao, Sun, Savarese CVPR 2010;

Interactions object-supporting plane

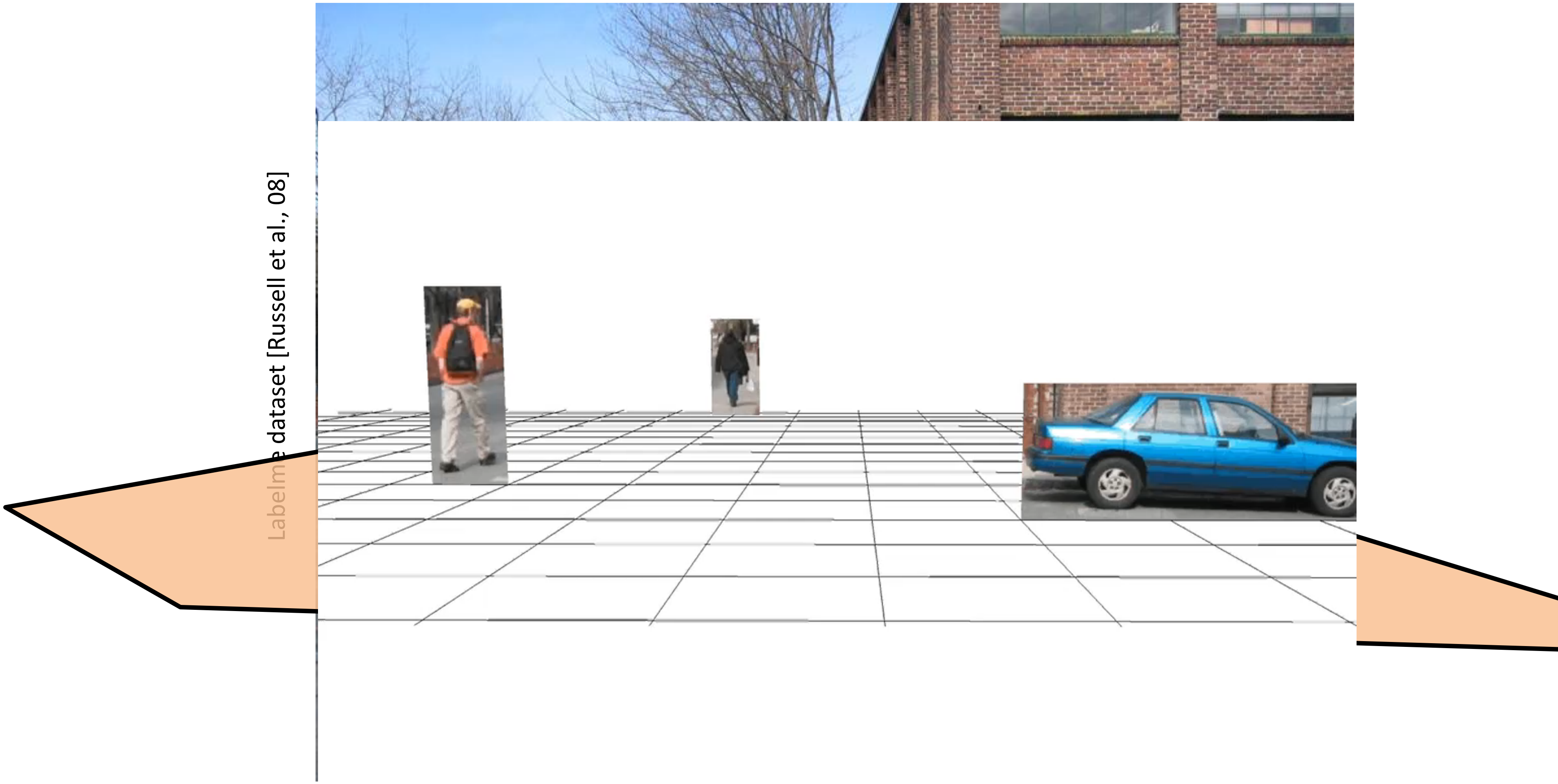


Labelme dataset [Russell et al., 08]

Modeling interaction objects-3D space

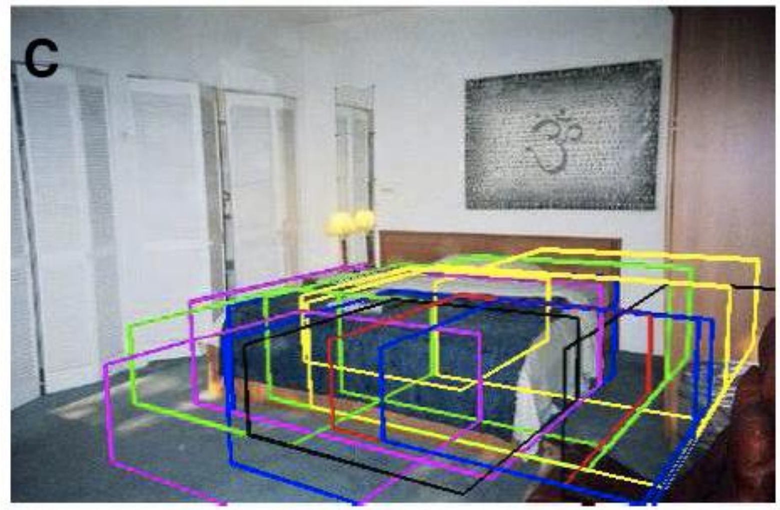
Bao, Sun, Savarese CVPR 2010;

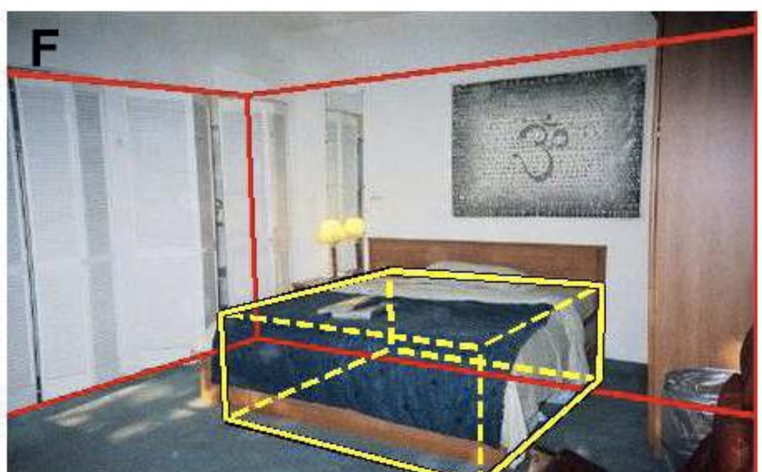
Interactions object-supporting plane



Modeling interaction objects-3D space

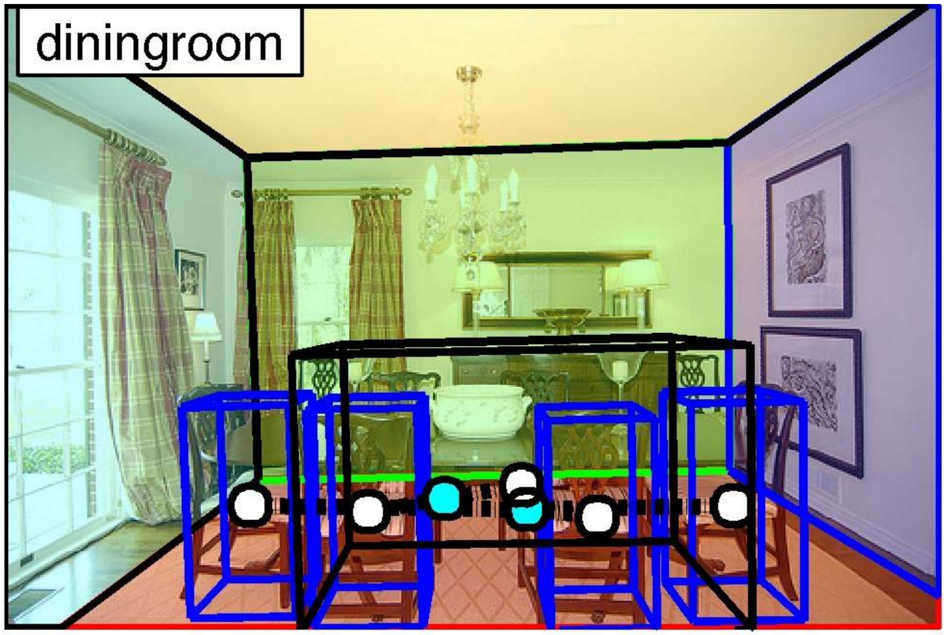
V. Hedau, D. Hoiem, and D.A. Forsyth *ECCV* 2010

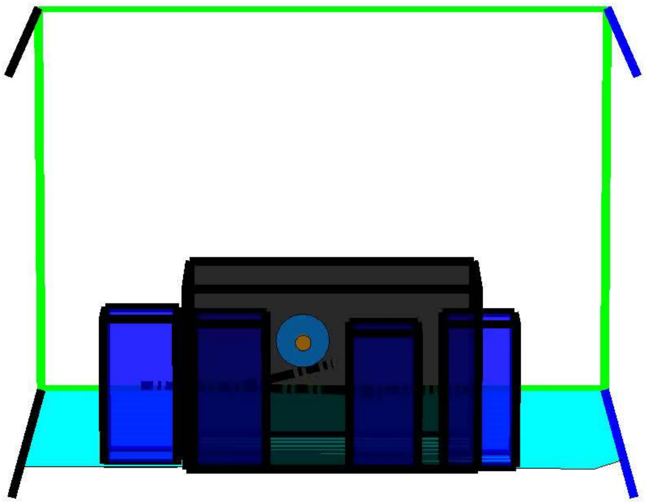




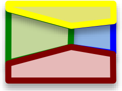
Modeling interaction objects-3D space

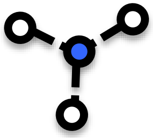
Choi, Chao, Pantofaru, Savarese, CVPR 13



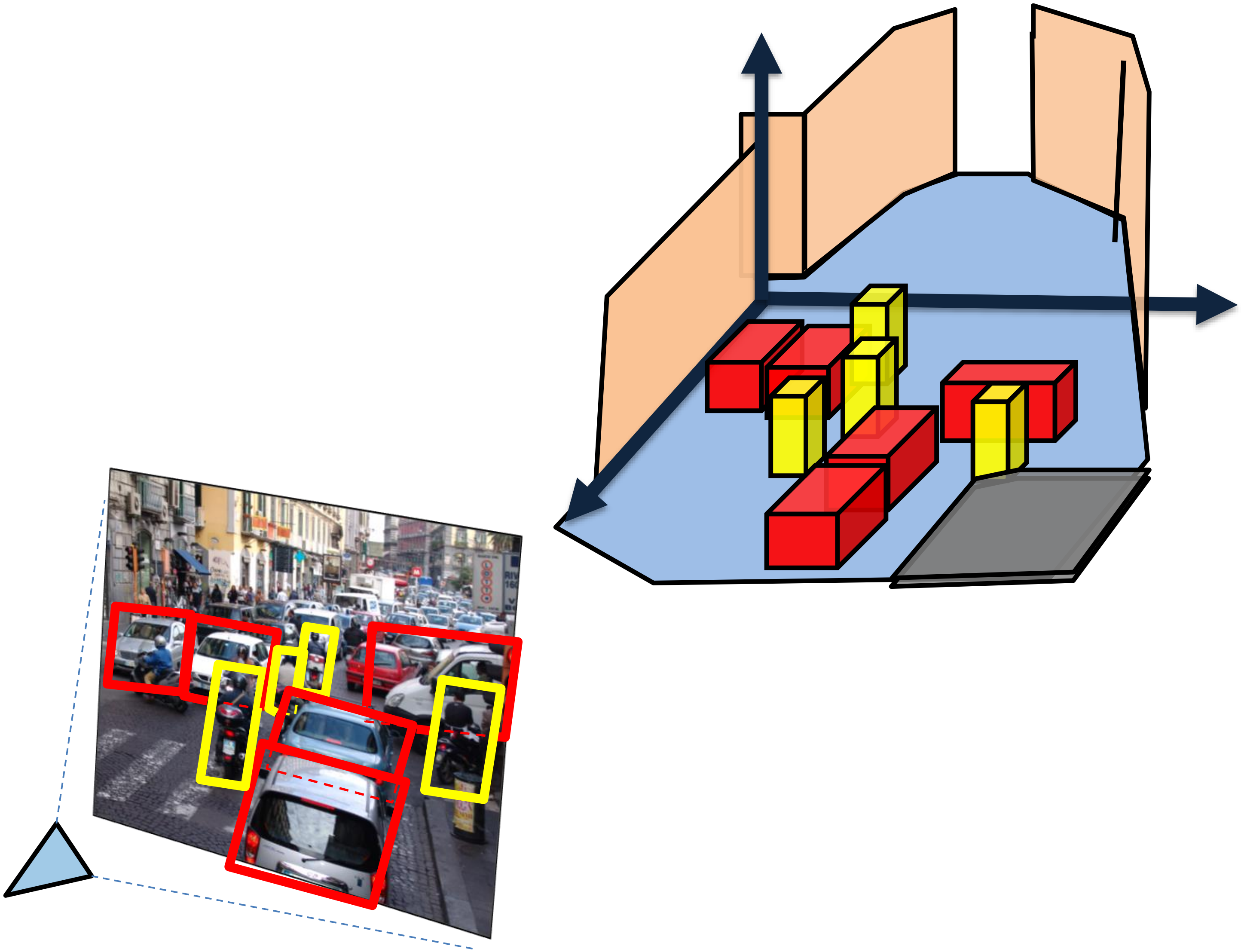


Sofa, **Coffee Table**, **Chair**, **Bed**, **Dining Table**, **Side Table**

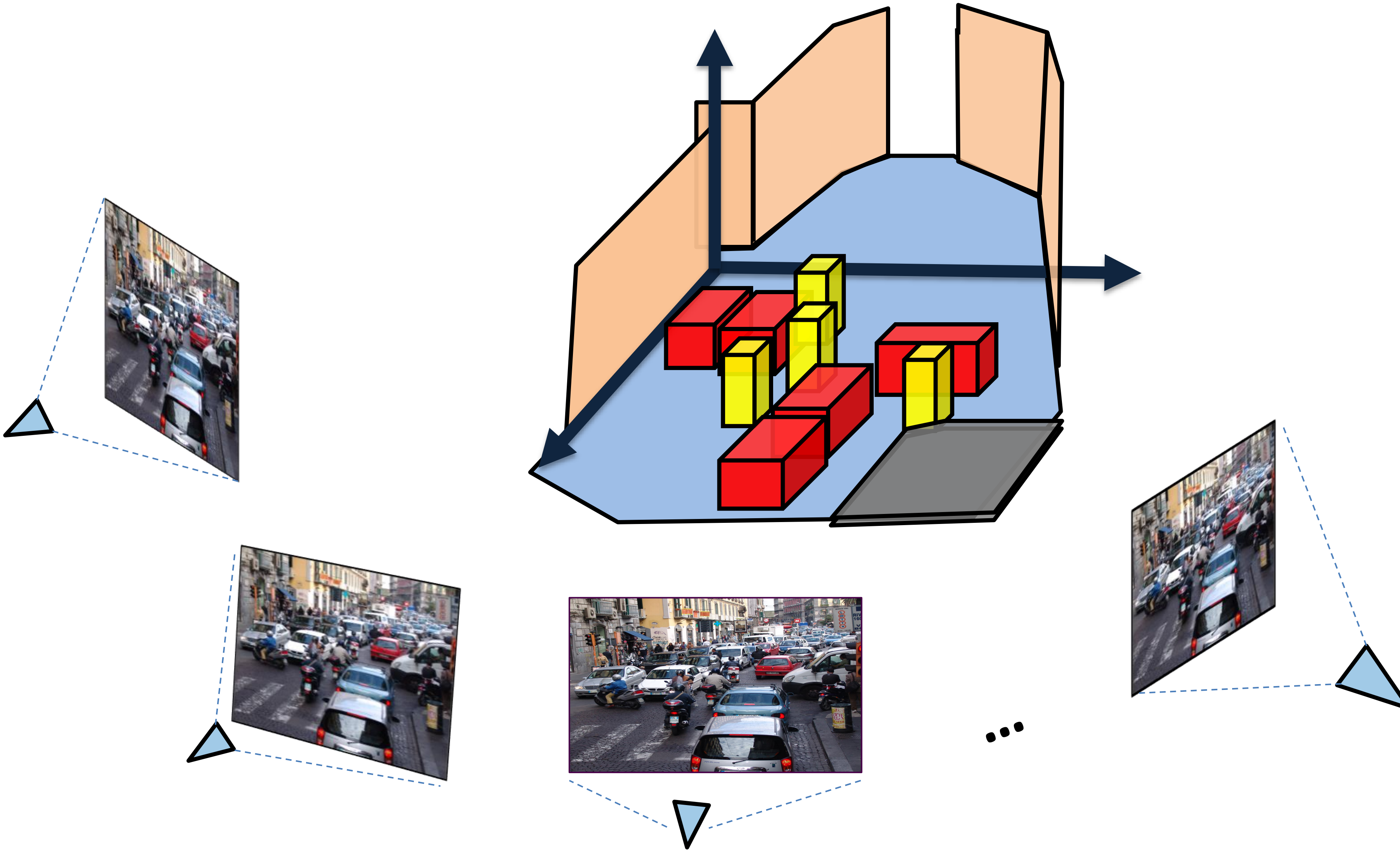

Estimated Layout


3D Geometric Phrases

Understanding the objects and space from multiple images



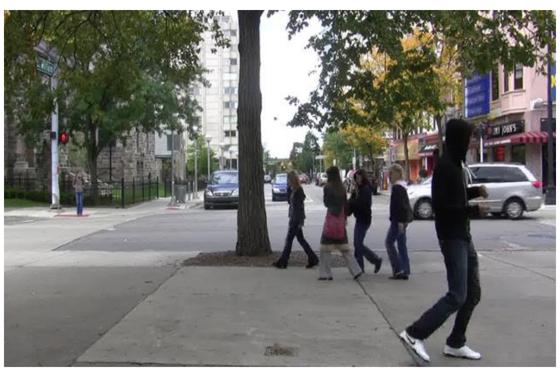
Understanding the objects and space from multiple images

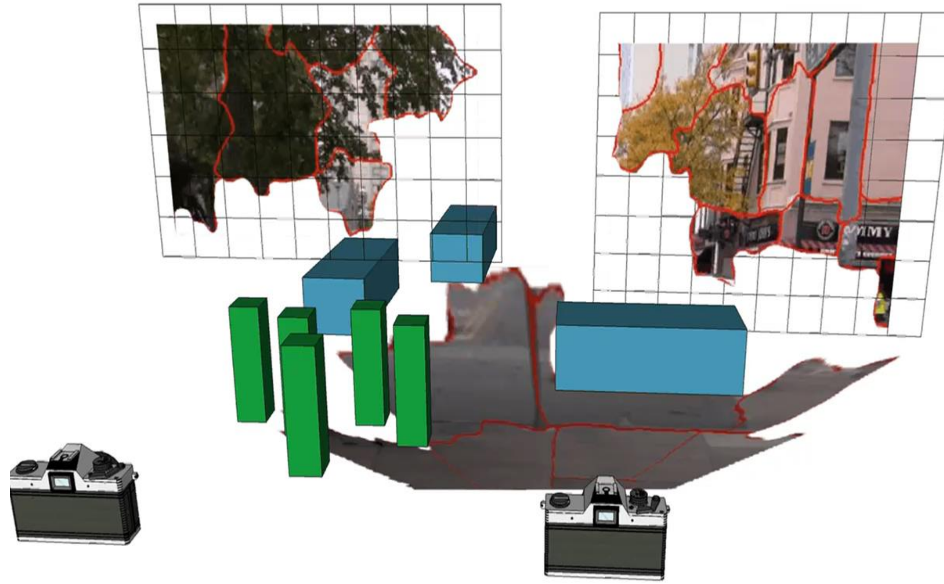


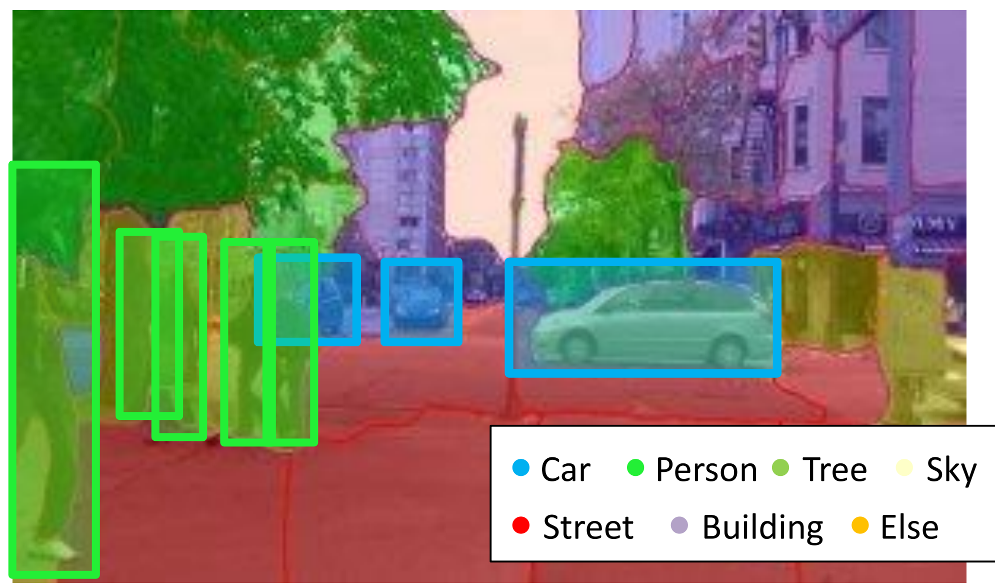
Understanding the objects and space from multiple images

Bao & S. Savarese, CVPR 2011

Input images

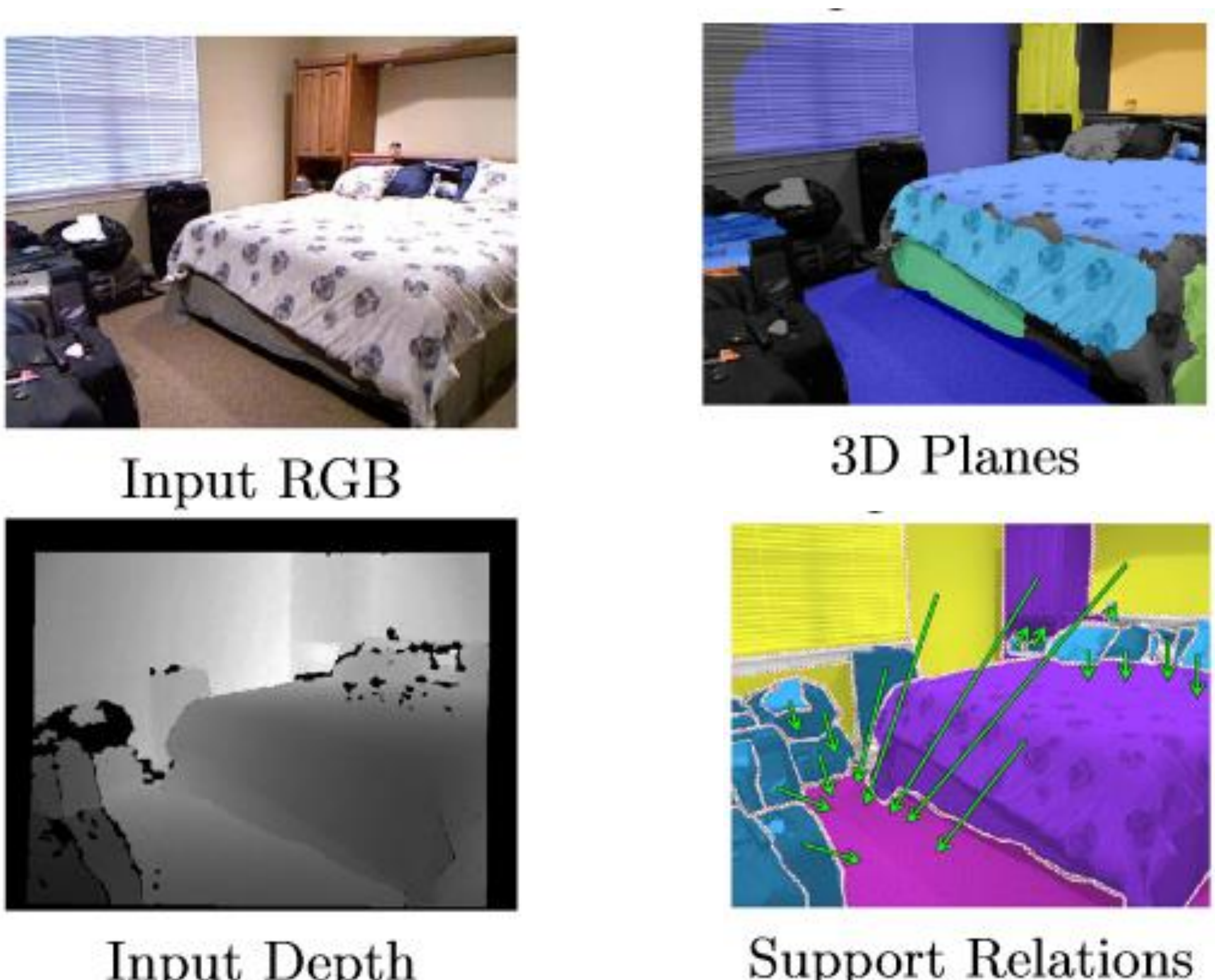






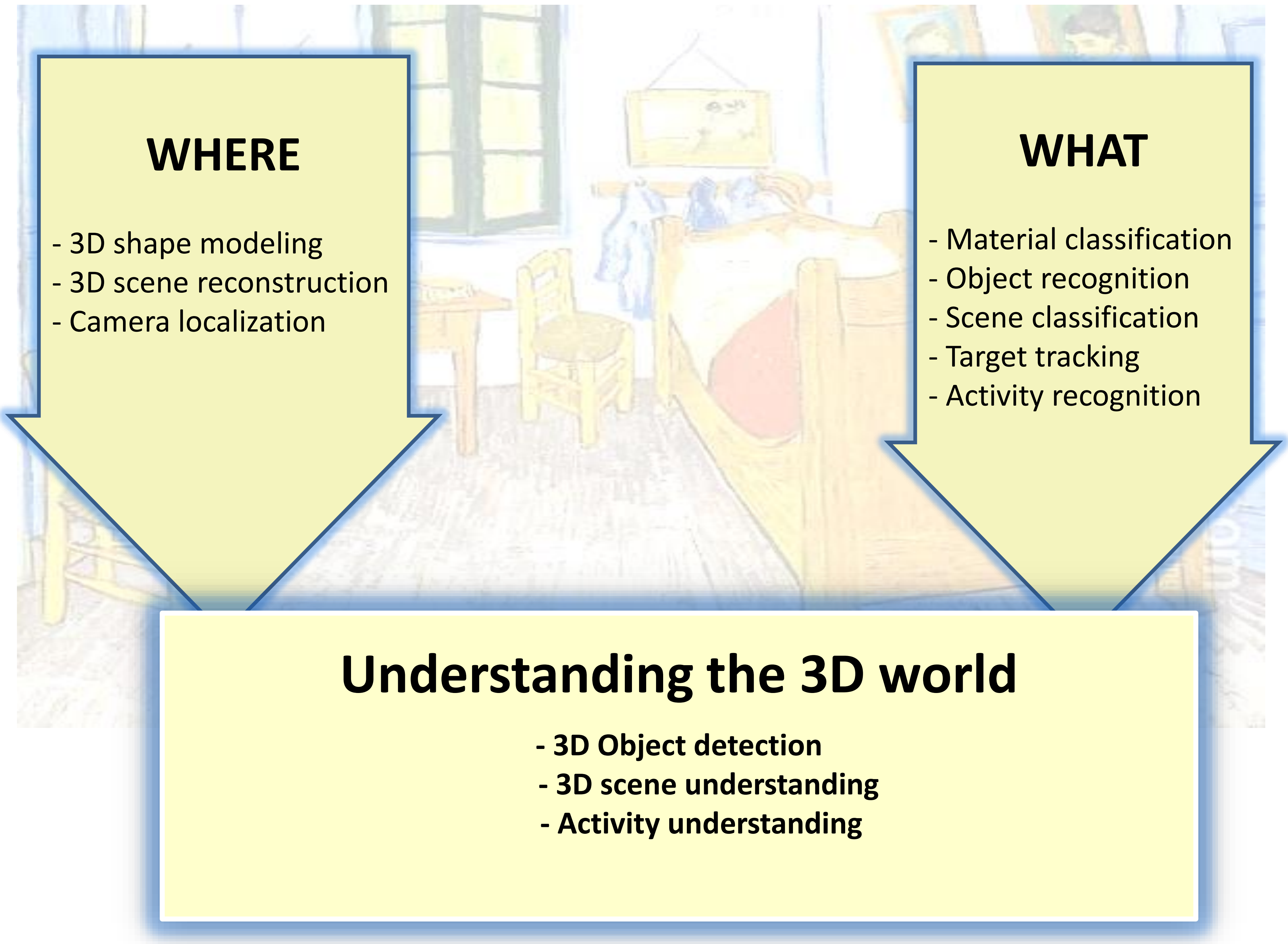
Understanding the objects and space from D-RGB images

Indoor Segmentation and Support Inference from RGBD Images
Silberman, N., Hoiem, D., Kolhi, P. and Fergus, R.

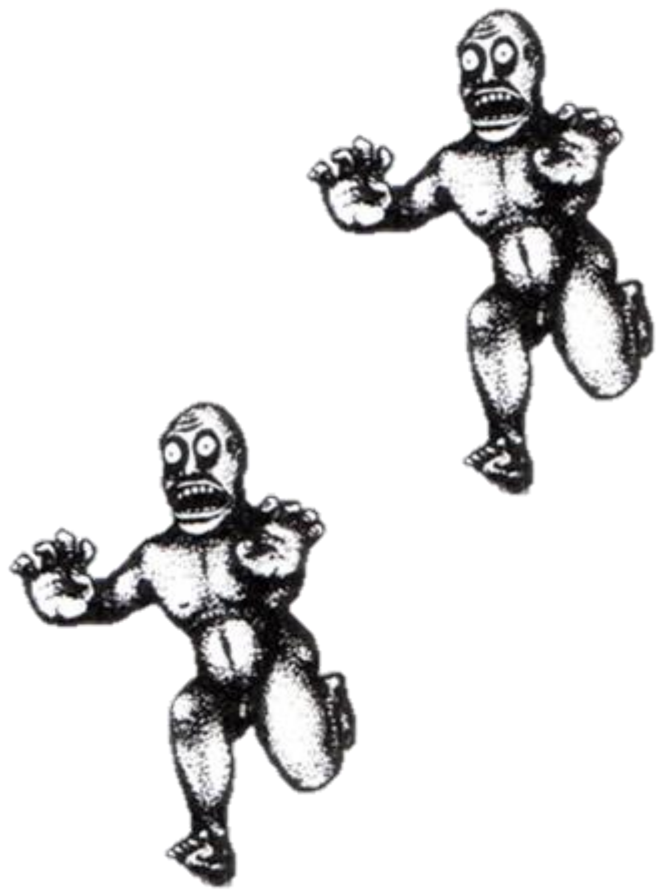


Understanding the objects and space from D-RGB images

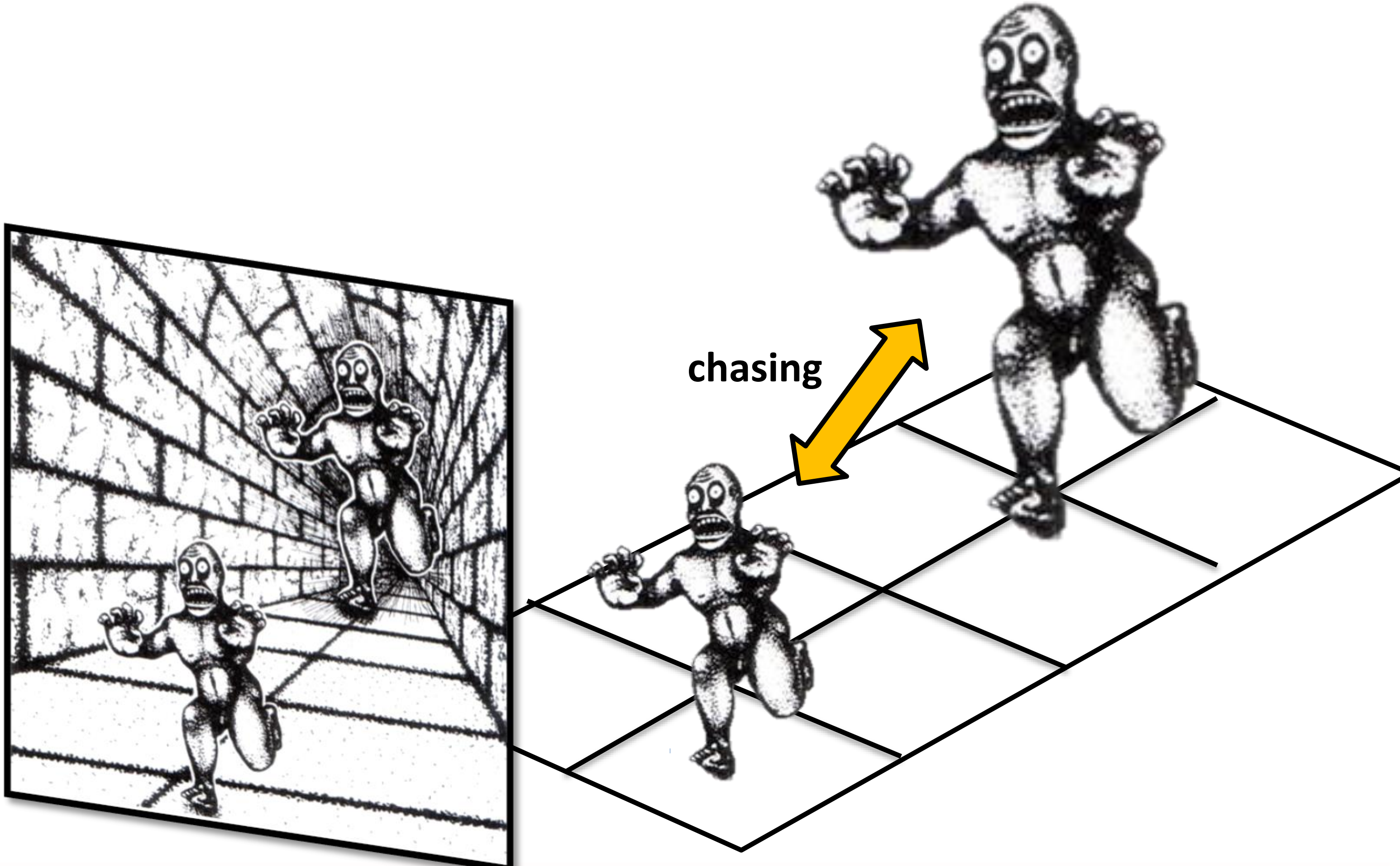
Co-inference for Multi-modal Scene Analysis
D. Munoz, J. A. Bagnell, M. Hebert, ECCV 2012


WHERE
- 3D shape modeling
- 3D scene reconstruction
- Camera localization
WHAT
- Material classification
- Object recognition
- Scene classification
- Target tracking
- Activity recognition
Understanding the 3D world
- 3D Object detection
- 3D scene understanding
- Activity understanding

Activity understanding



Activity understanding

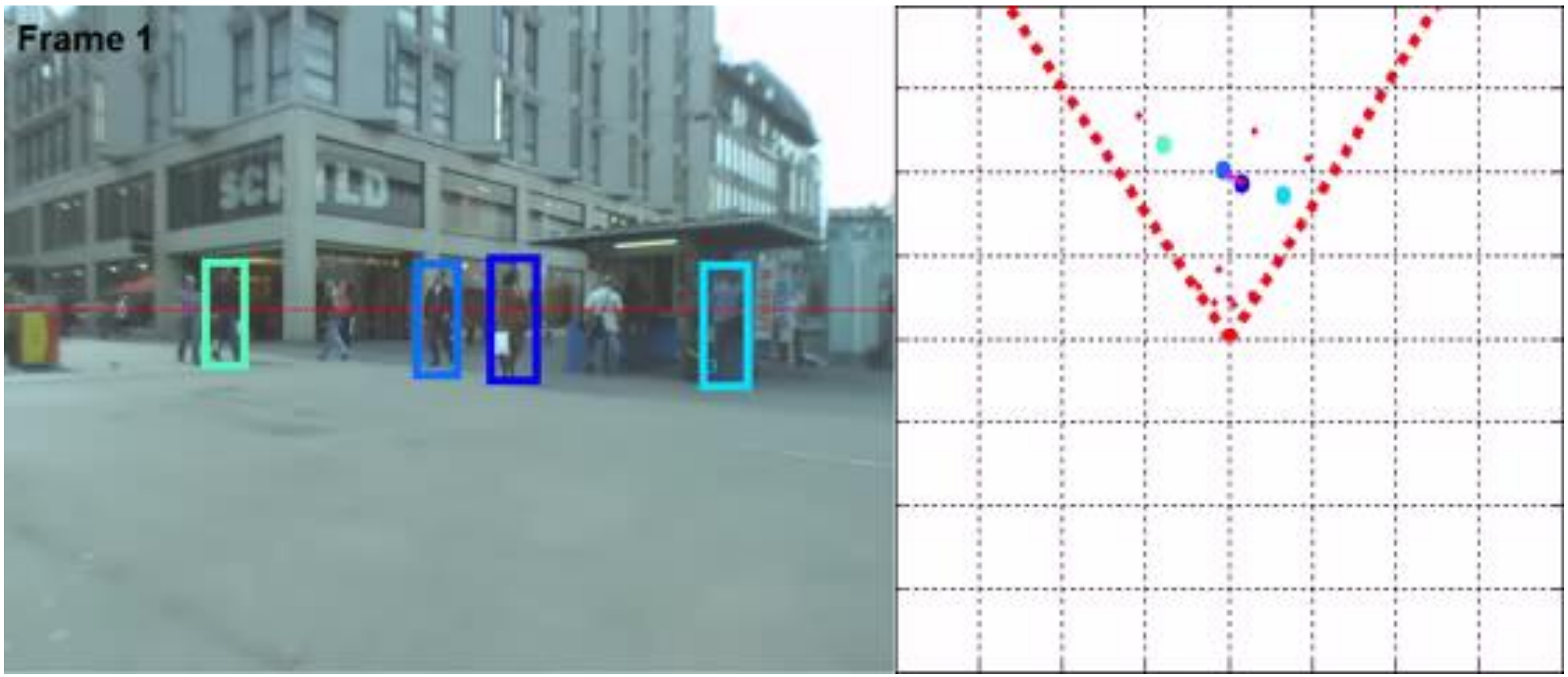


- Interplay between humans, objects and 3D space helps activity understanding

Modeling interaction human-space

Choi & Savarese , ECCV 2010
Choi, Pantofaru, Savarese, CORP-ICCV 2011

Zhao *et al.* 04
Wu *et al.* 07,
Breitenstein *et al.* 09,
Dollár *et al.* 2009
Ess *et al. 09*

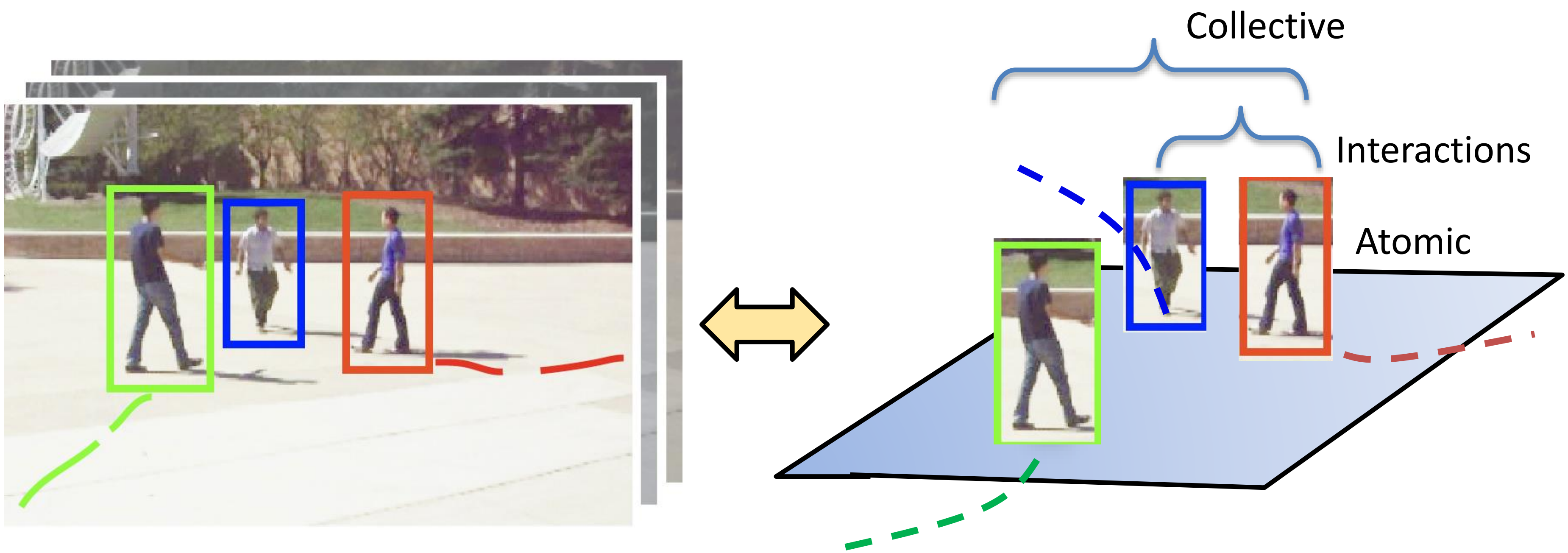


- Monocular cameras
- Un-calibrated cameras
- Arbitrary motion

- Highly cluttered scenes
  - Occlusion
  - Background clutter
- Moving targets

Activity understanding in 3D

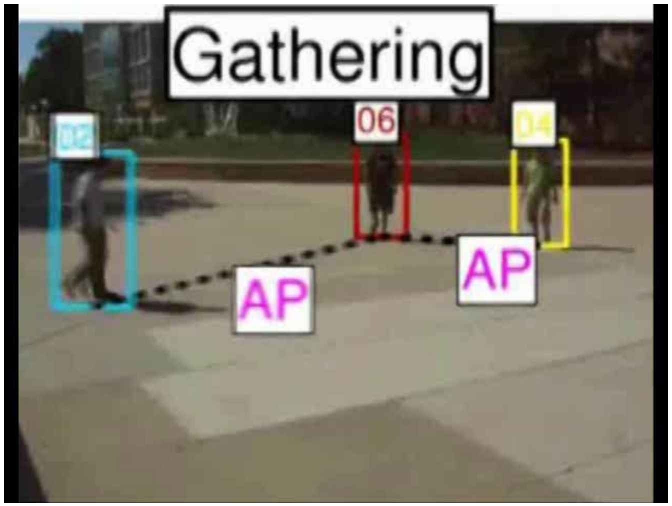
Choi & Savarese , ECCV 2012 (oral)
Gupta et al 09, 11
Li et al. 09
Ryoo & Aggarwal '09
Lan et al. '09
Delaitre et al. '12

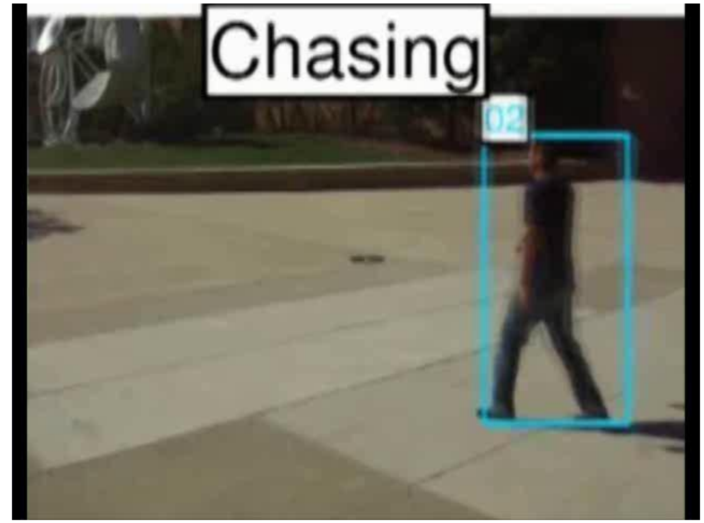


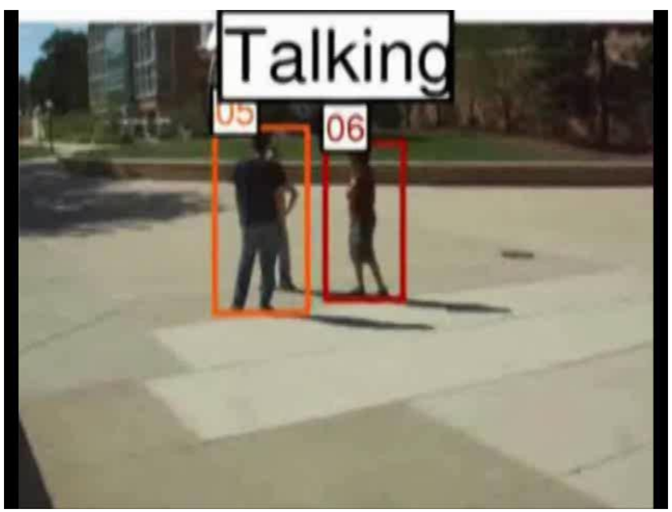
- 3D trajectories enables activity recognition:
  - Atomic ⇔ Interactions ⇔ Collective
- High-level social behavior helps 3D tracking

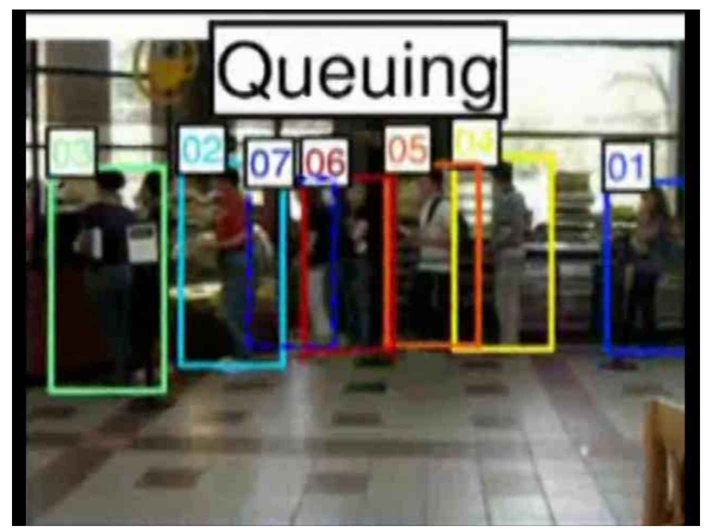
Activity understanding in 3D

Choi & Savarese , ECCV 2012 (oral)









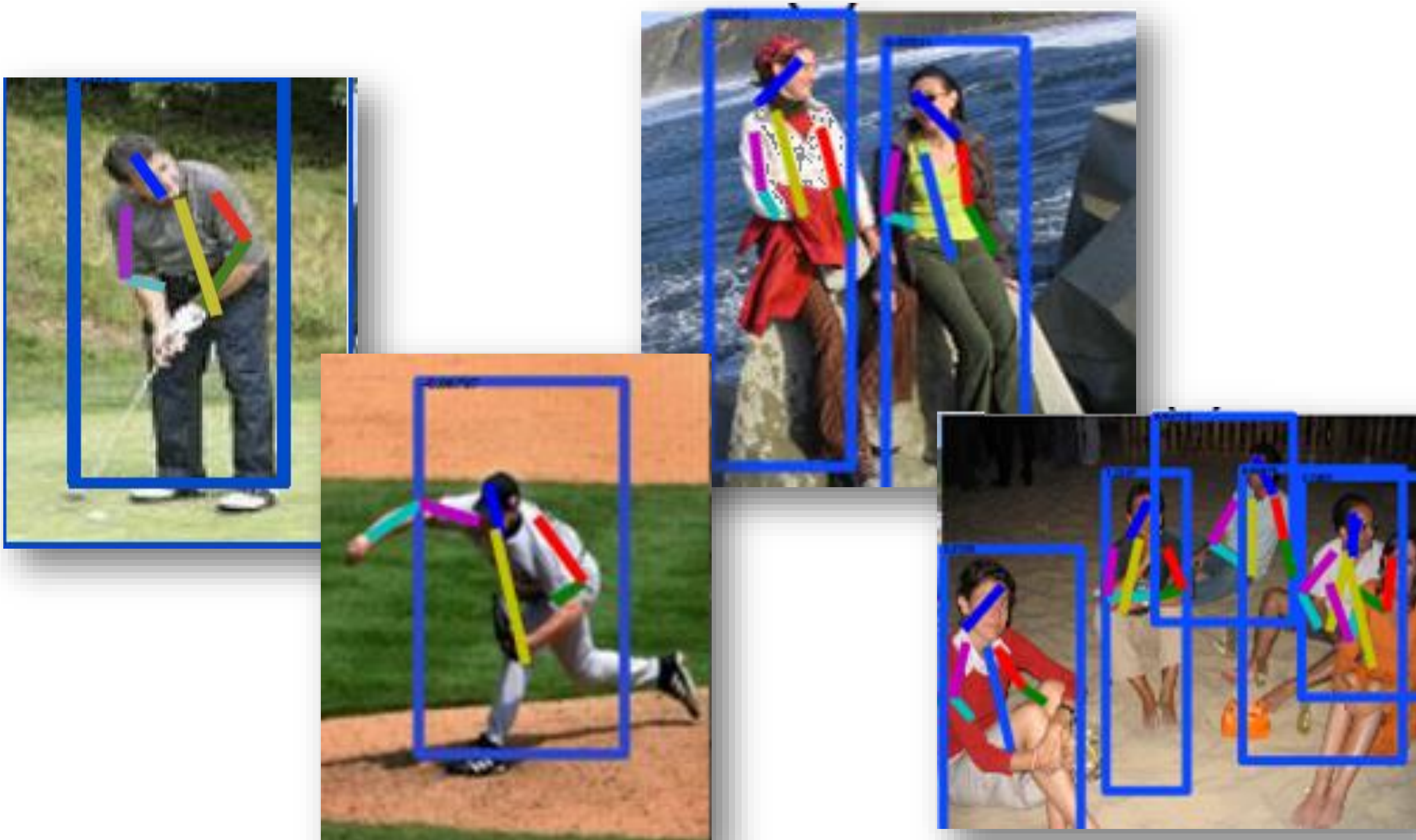
AP = approaching;   FE = face each other; SR= standing in a row
WS = walking-side-by-side   RR= run after each other;

Activity understanding and prediction in 3D

Activity Forecasting. Kris M. Kitani, Brian D. Ziebart, Drew Bagnell and Martial Hebert. European Conference on Computer Vision (ECCV 2012).

Understanding human pose from images

Sun & Savarese, ICCV'11
Bourdev et al. '10
Yang & Ramanan, '11

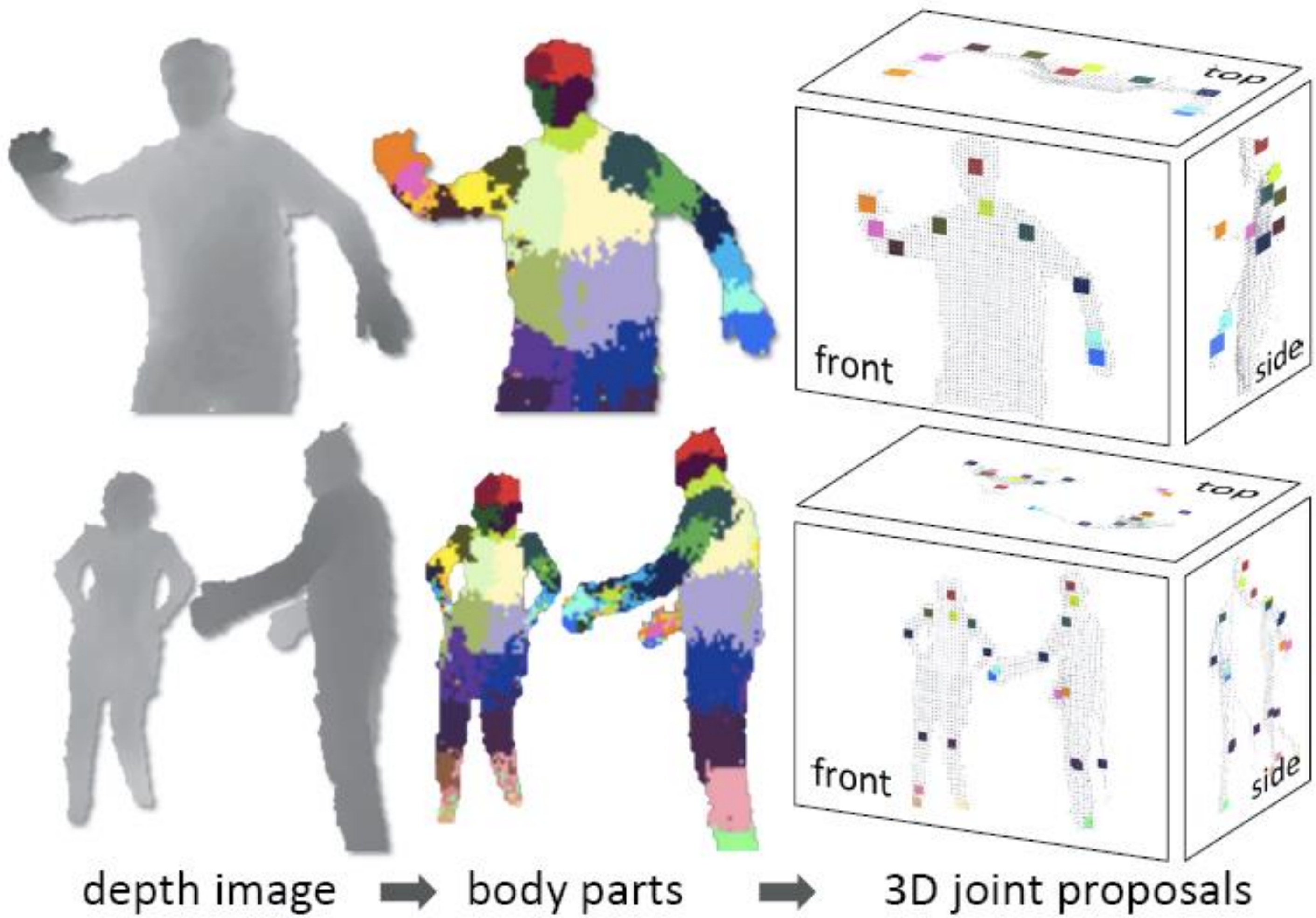


Understanding human pose 3D-RGB

Microsoft Kinect

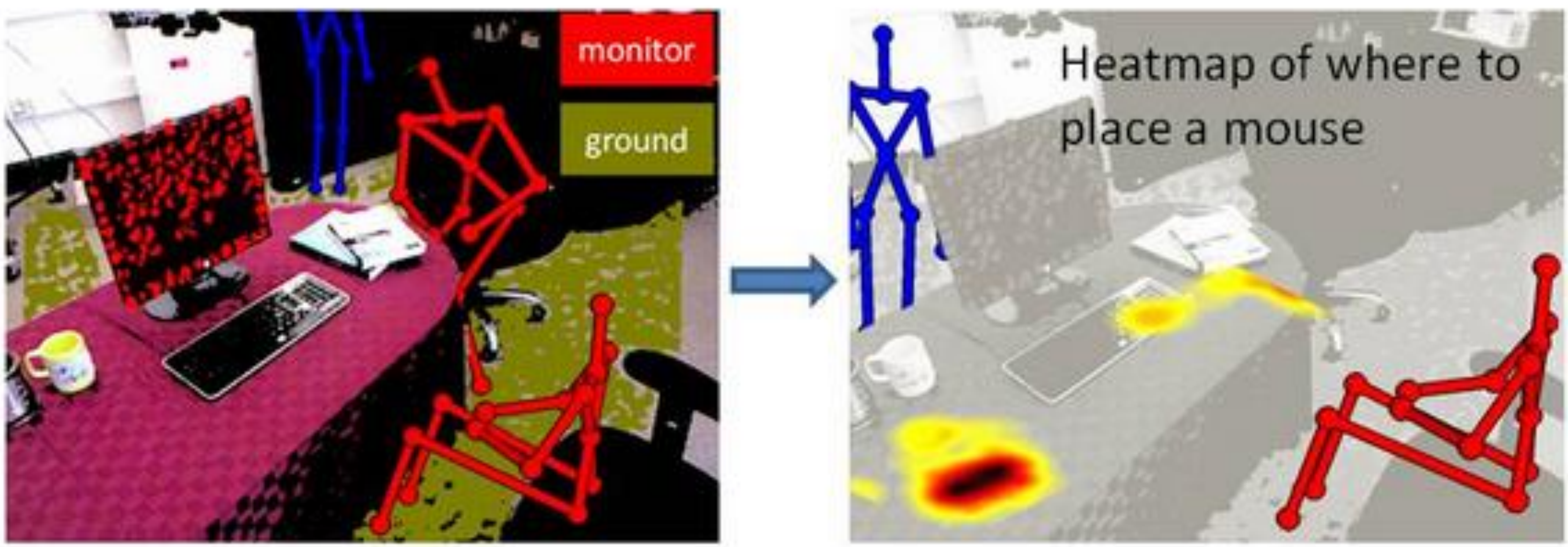
Understanding human pose 3D-RGB

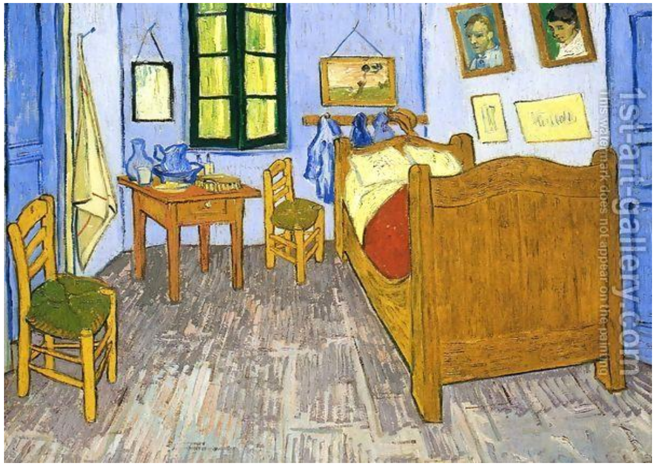
Jamie Shotton , Andrew Fitzgibbon , Mat Cook , Toby Sharp , Mark Finocchio , Richard Moore , Alex Kipman , Andrew Blake In CVPR, 2011



Modeling human-object interactions and affordances

Hallucinated Humans as the Hidden Context for Labeling 3D Scenes, Yun Jiang, Hema S. Koppula, Ashutosh Saxena. In Computer Vision and Pattern Recognition (CVPR), 2013





Take home message from this course

- 3D representations for modeling, learning and detecting objects (humans) from images
- 3D representation for modeling and estimating 3D space
- Jointly reasoning about objects (humans) and 3D space provide powerful tools for
  - Coherent 3D scene understanding
  - Coherent activity understanding

syllabus