CS231A

Computer Vision:
From 3D reconstruction to Recognition

Professor Silvio Savarese

Computational Vision and Geometry Lab

CS231A

- Instructor
  - Silvio Savarese
  - ssilvio@stanford.edu
  - Office: Gates Building, room: 228
  - Office hour: Tues 3:30-4:30pm or under appoint.

- CAs:
- Kevin Wong
- David Held
- Jiayuan (Mark) Ma
- Chris Lengerich

- Class Time & Location
  - Tu Th 11:00am - 12:15PM – Nvidia Auditorium

CS231A

Prerequisites:

- CS 131 or equivalent; It is encouraged and preferred that you have taken CS221 or CS229, or have equivalent knowledge.

Course assignments:

- 4 problem sets (first problem released next week!)
- 1 mid-term exam (take home, 48 hours)
- 1 project

CS231A

- **Suggested text books:**

- R. Szeliski. *Computer Vision: Algorithms and Applications*. Springer, 2011.
- [FP] D. A. Forsyth and J. Ponce. *Computer Vision: A Modern Approach* (2nd Edition). Prentice Hall, 2011.
- [HZ] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Academic Press, 2002.
- D. Hoiem and S. Savarese. *Representations and Techniques for 3D Object Recognition and Scene Interpretation*, Synthesis lecture on Artificial Intelligence and Machine Learning. Morgan Claypool Publishers, 2011
- Learning OpenCV, by Gary Bradski & Adrian Kaehler, O'Reilly Media, 2008.

CS231A

Grading policy

- Homeworks: 40%
  - 4 homeworks
- Mid term exam: 15%
- Course project: 40%
  - mid term progress report 5%
  - final report 30%
  - presentation 10%
- Attendance and class participation: 5%
  - Questions, answers, remarks…

Grading policy

- Late policy home works:
  - If 1 day late, 50% off the grade for that homework
  - Zero credits if more than one day.
  - A "48-hours one-time late submission bonus" is available; that is, you can use this bonus to submit your HW late after at most 48 hours. This is one time bonus: After you use your bonus, you must adhere to the standard late submission policy.
  - No exceptions will be made.
  - No "late submission bonus" is allowed when submitting your exam or project.

- 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
  - Home works or class project coding policy: using on line code or other students/researchers’ code is not allowed in general. Exceptions can be made and individual cases will be discussed with the instructor.

Course Project

- Replicate an interesting paper
- Comparing different methods to a test bed
- A new approach to an existing problem
- Original research
- 1 or 2 TBA large scale projects (5-10 students each)

- Write a 8-page paper summarizing your results
- Release the final code
- Give a presentation

- We will introduce projects in 1-2 weeks

- Important dates: look up class schedule

Course Project

- Form your team:
  - 1-3 people
  - 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 project presentation (spotlight and/or poster presentation)

For final project due dates please consult webpage

Lecture 1
Introduction

- An introduction to computer vision
- Course overview

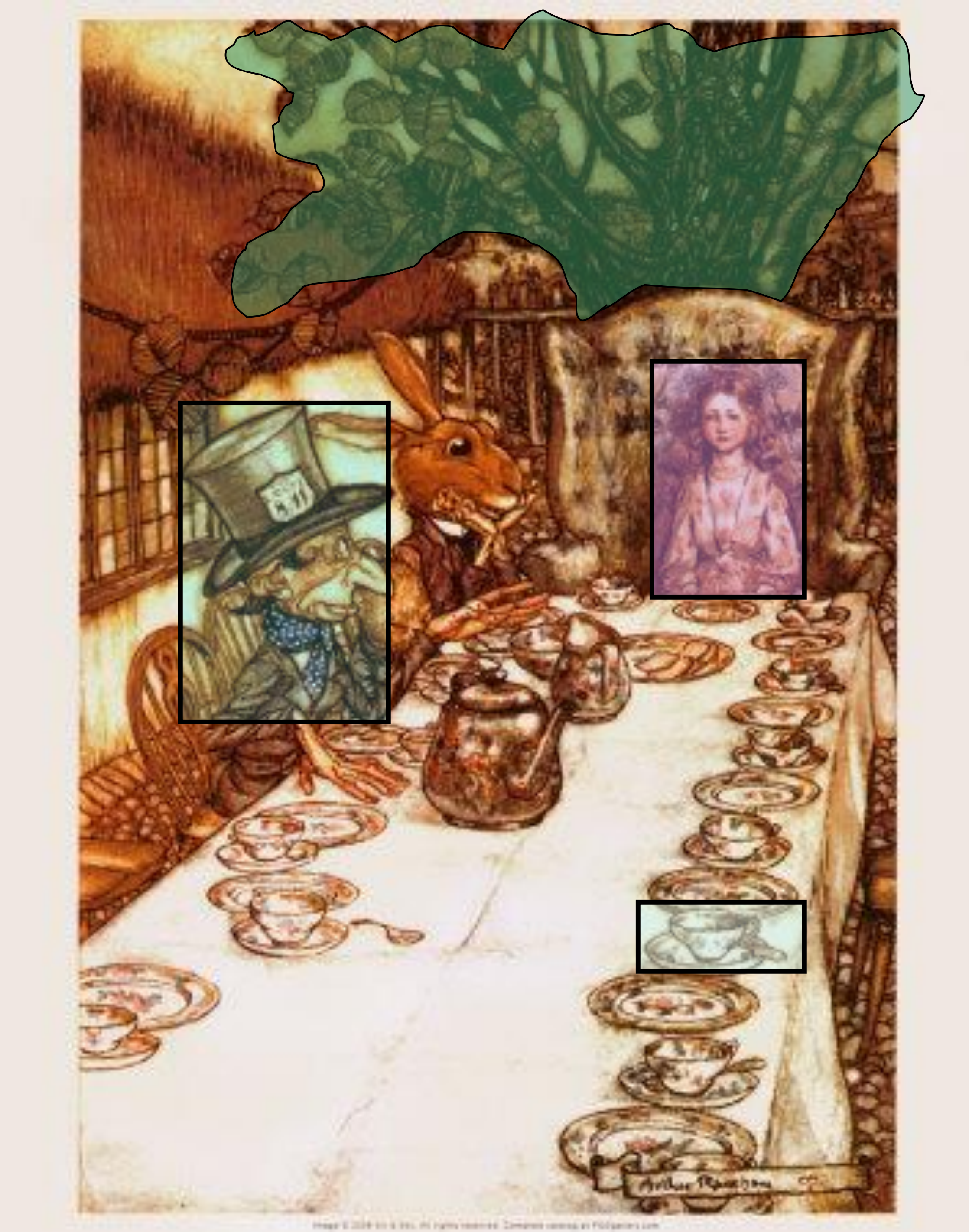
Illustration by Arthur Rackham

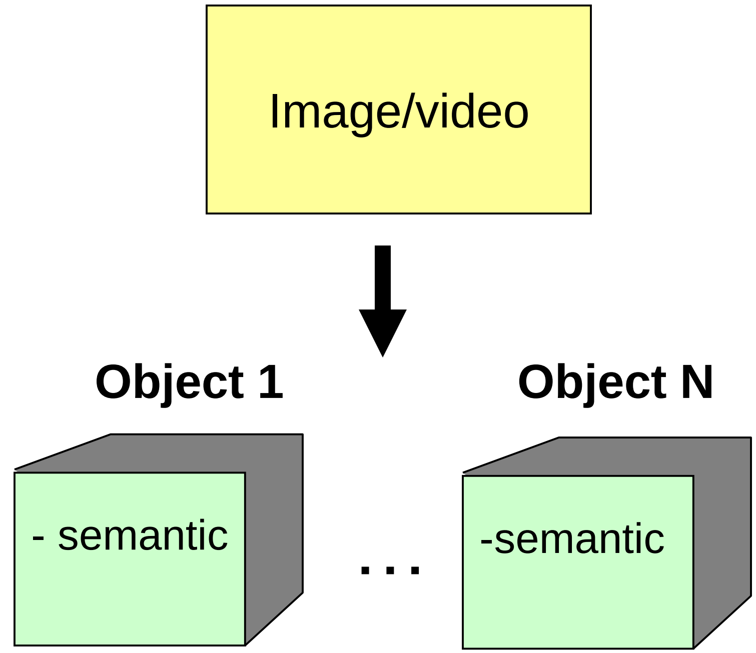
"There was a table set out under a tree in front of the house, and the March Hare and the Hatter were having tea at it."

"The table was a large one, but the three were all crowded together at one corner of it …"

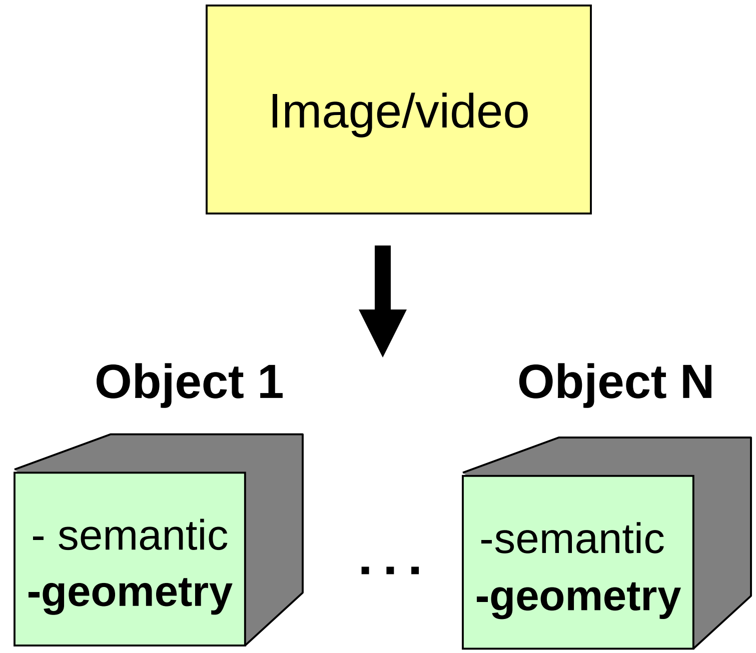
From "A Mad Tea-Party"
Alice's Adventures in Wonderland
by
Lewis Carroll

Computer vision

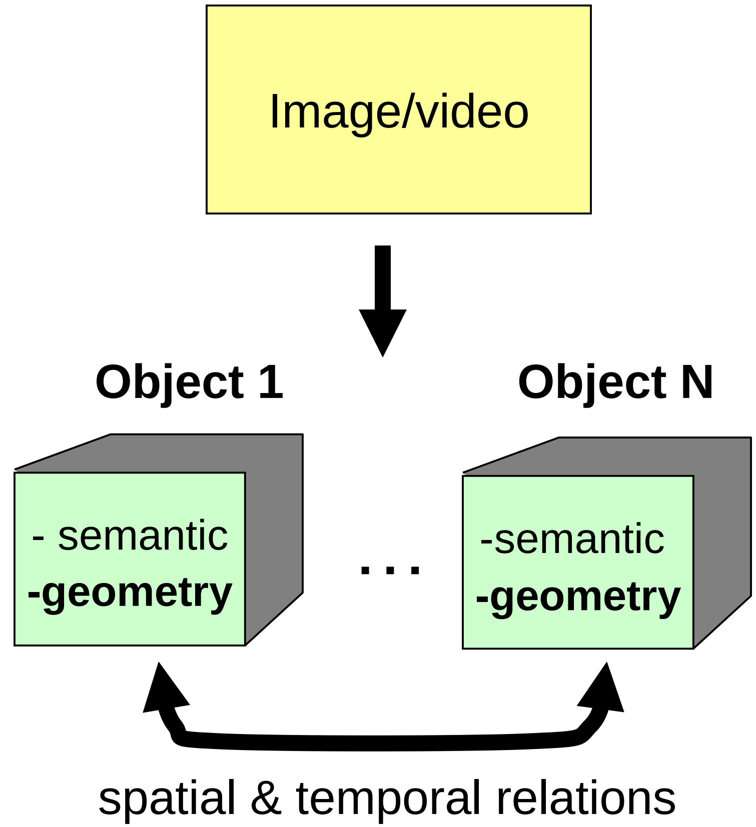




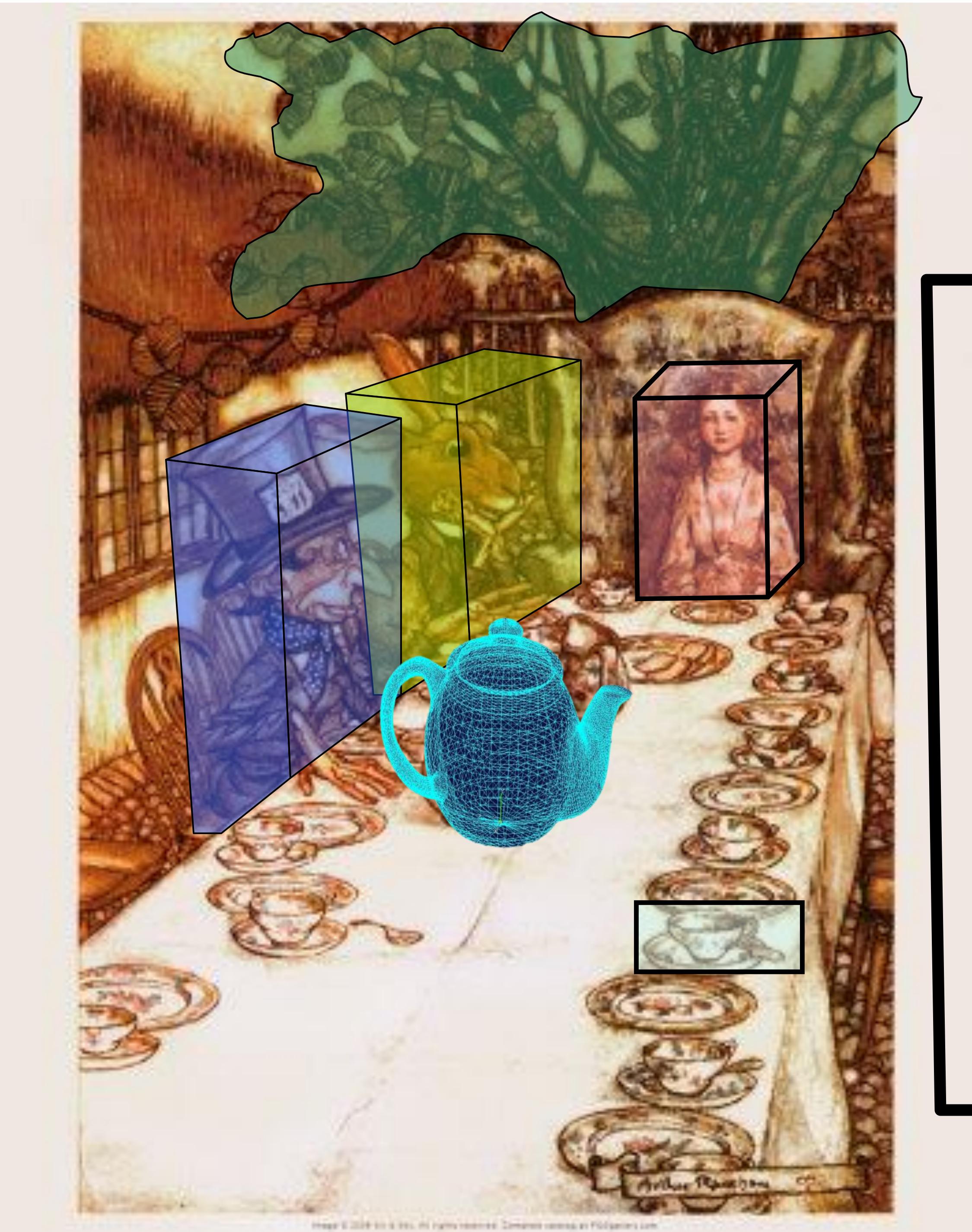
Computer vision

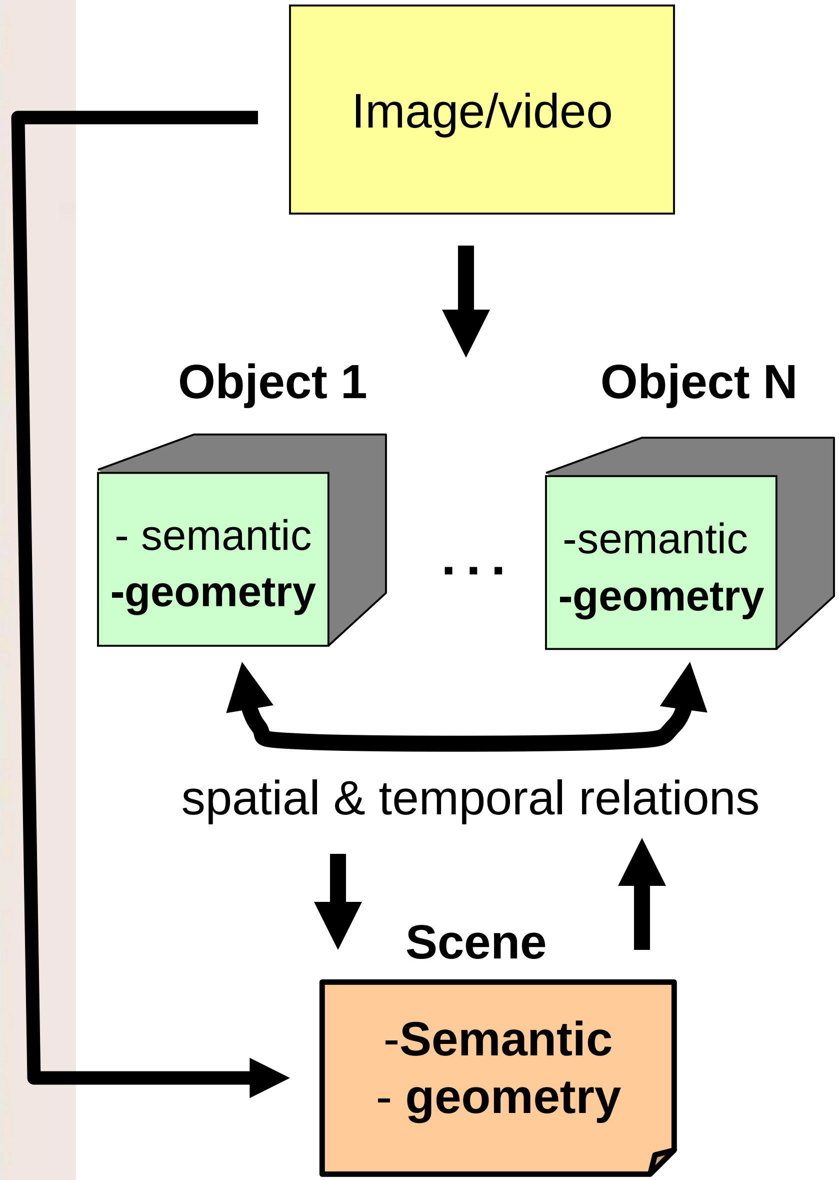


Computer vision



Computer vision





Computer vision

Computer vision studies the **tools and theories** that enable the design of machines that can **extract useful information from imagery data** (images and videos) toward the goal of **interpreting the world**

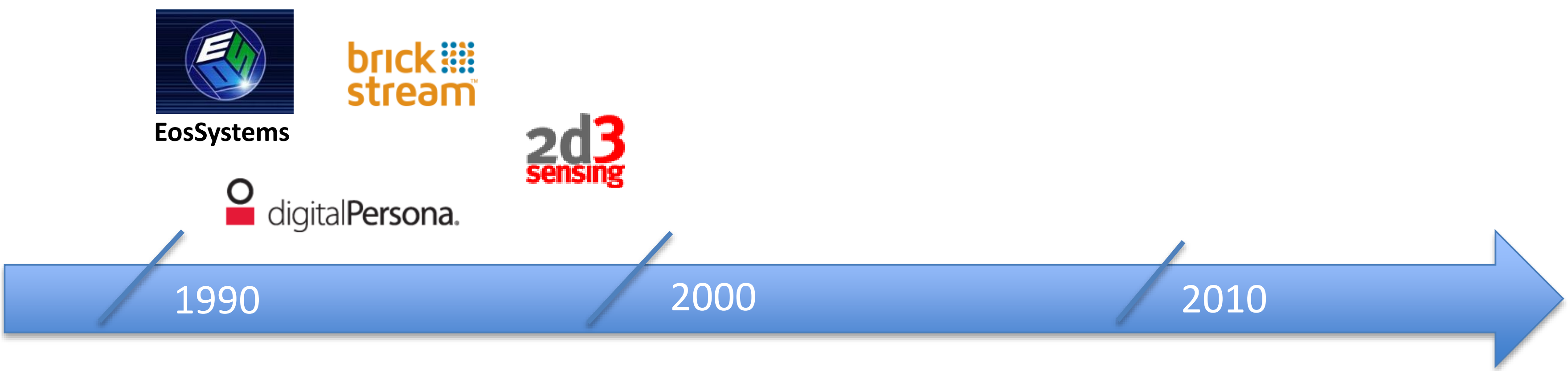
Sensing device

Computational device

- **Extract information**
- **Interpretation**

Information: features, 3D structure, motion flows, etc…
Interpretation: recognize objects, scenes, actions, events

Computer vision and Applications



Fingerprint biometrics

digitalPersona.

Augmentation with 3D computer graphics

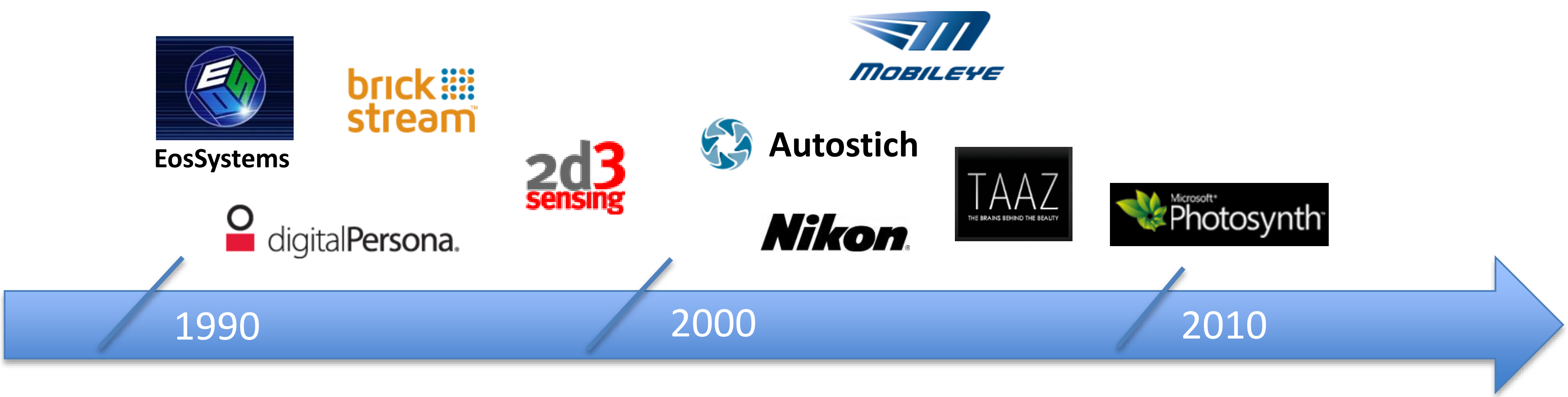
3D object prototyping

EosSystems

Photomodeler

Computer vision and Applications

- New features detector/descriptors
- CV leverages machine learning



Face detection

BBC NEWS

UK version | International version | About the versions | L

- News Front Page
- World
- UK
- England
- Northern Ireland
- Scotland
- Wales
- **Business**
- Market Data
- Your Money
- E-Commerce
- Economy
- Companies
- Politics
- Health
- Education

Last Updated: Monday, 6 February 2006, 14:29 GMT

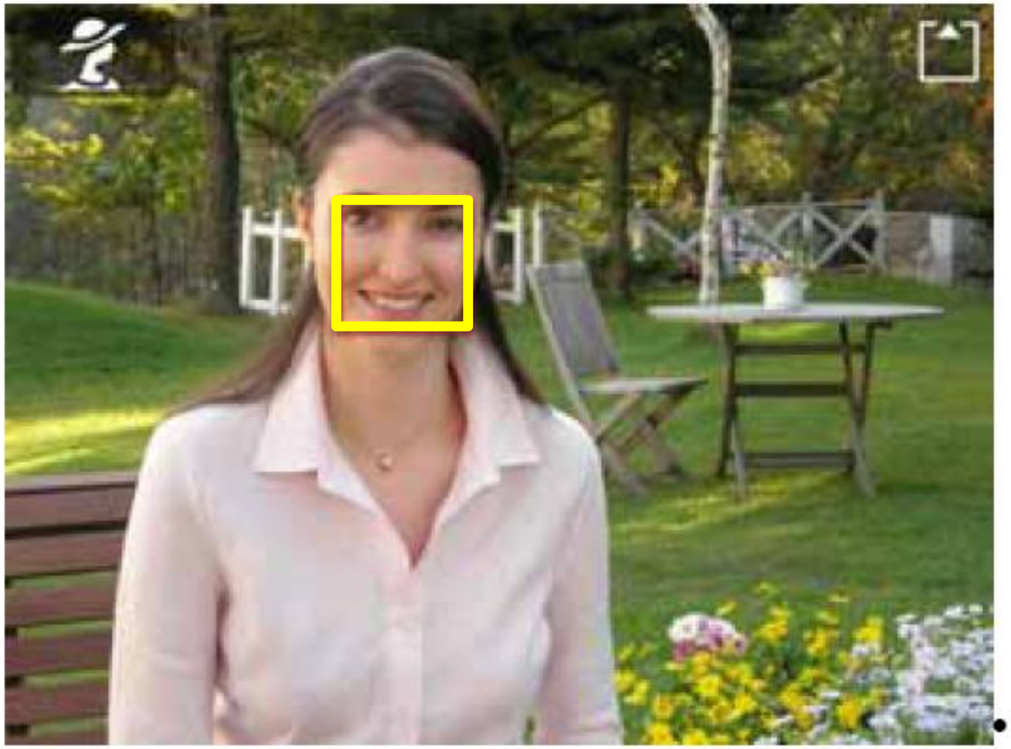
E-mail this to a friend | Printable version

Face-hunting cameras boost Nikon

Japanese camera maker Nikon has tripled its profits on the back of strong sales of digital cameras that automatically focus on human faces.

Face recognition cameras like the Coolpix L1 are popular

Face detection



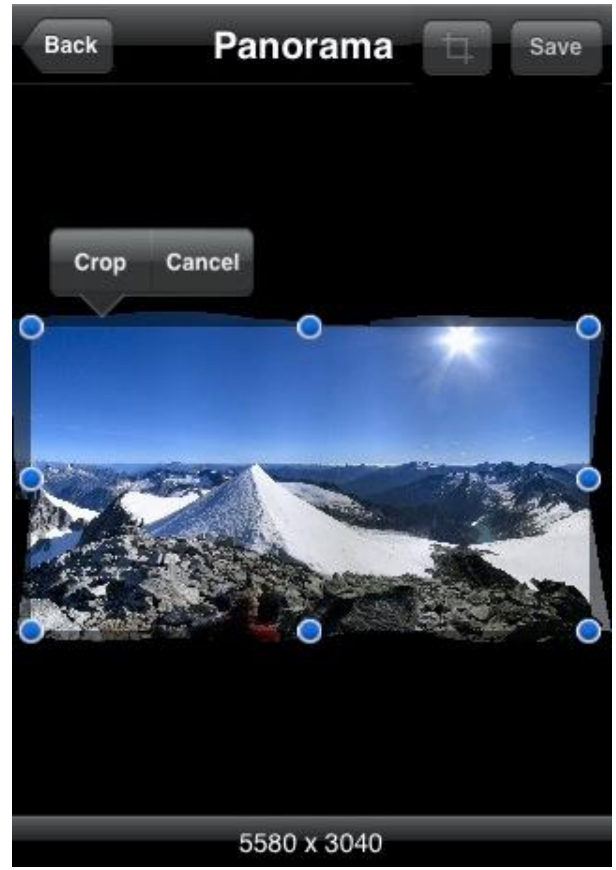
Sample image: Subject as seen on the COOLPIX 5900 camera's color LCD and when using Nikon's Face-priority AF function.

Web applications

TAAZ
THE BRAINS BEHIND THE BEAUTY

Photometria

Panoramic Photography

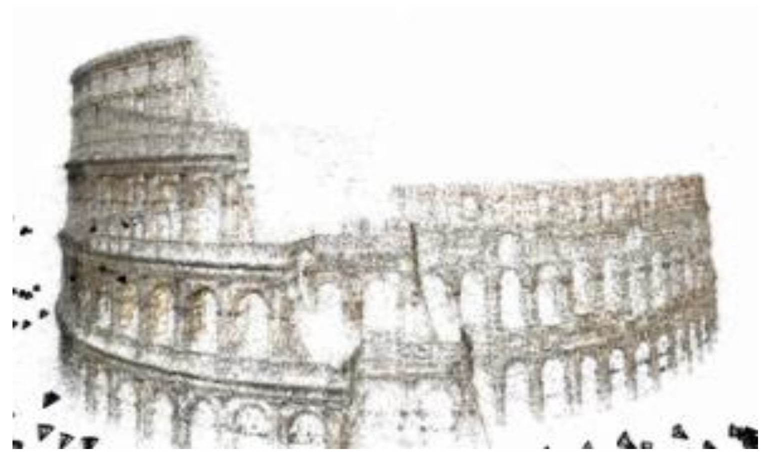


Cloudburst RESEARCH

HumanEyes Print Out of the Box

kolor

3D modeling of landmarks



Microsoft Photosynth

Computer vision and Applications

- Large scale image matching
- Efficient SLAM/SFM
- Better clouds ☺
- More bandwidth
- Increase computational power

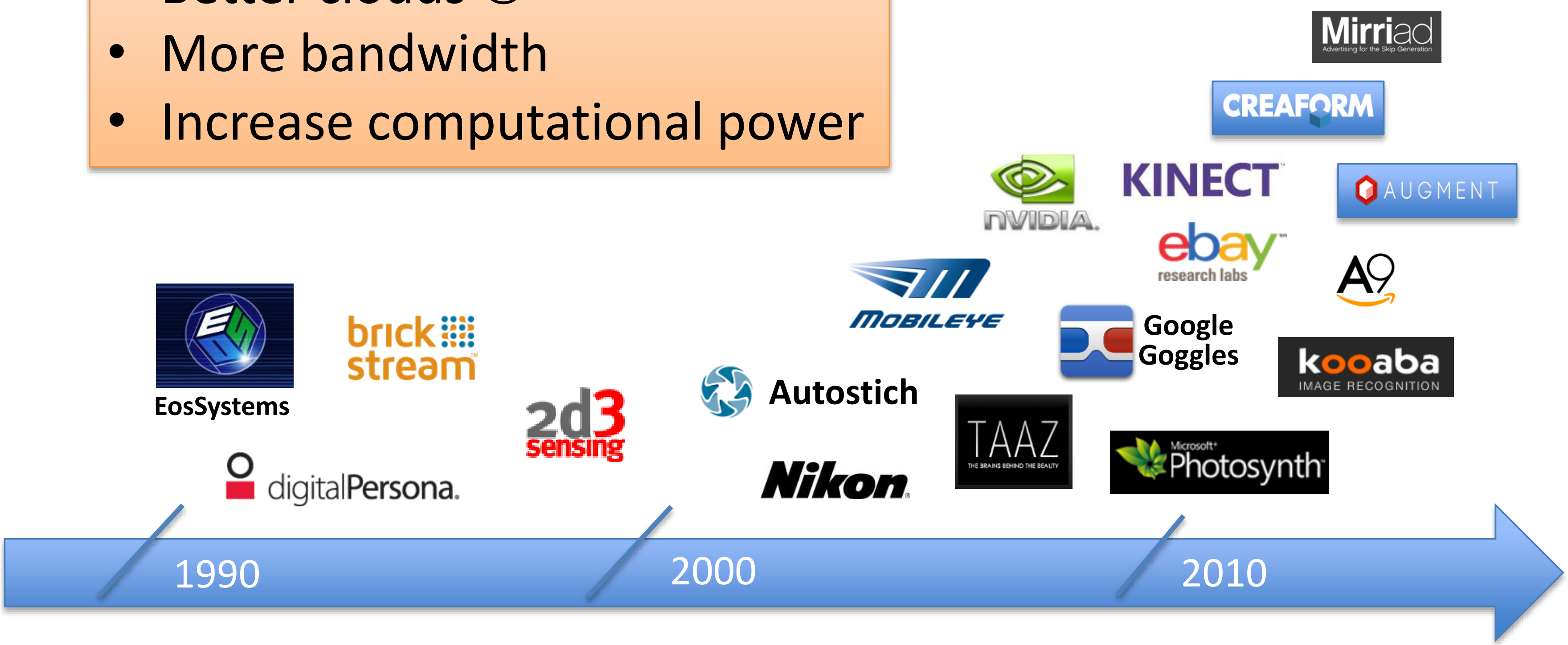


Image search engines

Google Image Search

Picasa

flickr

webshots

bing

YouTube Broadcast Yourself

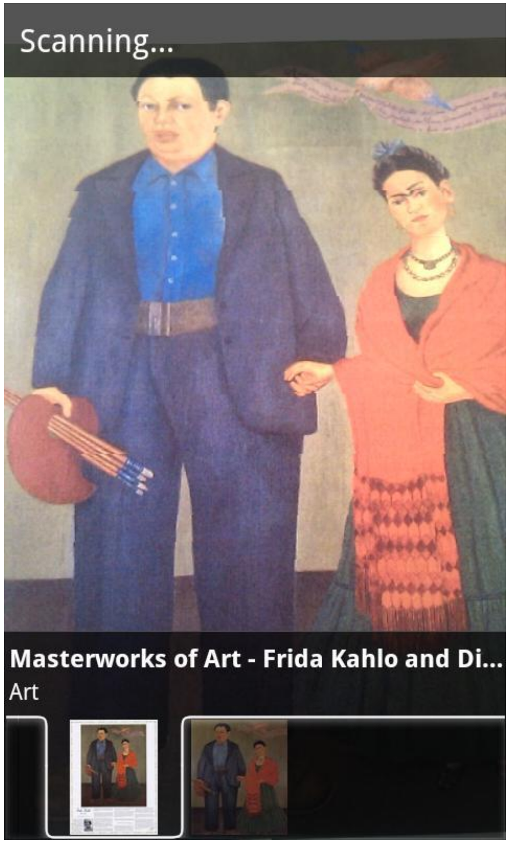
Incogna

LTU technologies

picsearch

YAHOO!

Visual search and landmarks recognition



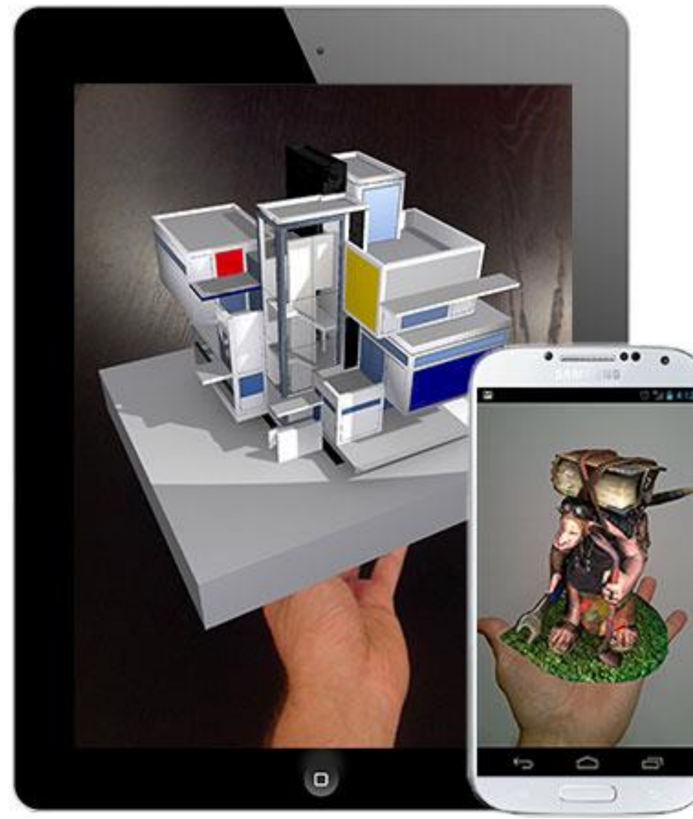
Google Goggles

Visual search and landmarks recognition

kooaba IMAGE RECOGNITION

A9

Augmented reality



Mirriad
Advertising for the Skip Generation

AUGMENT

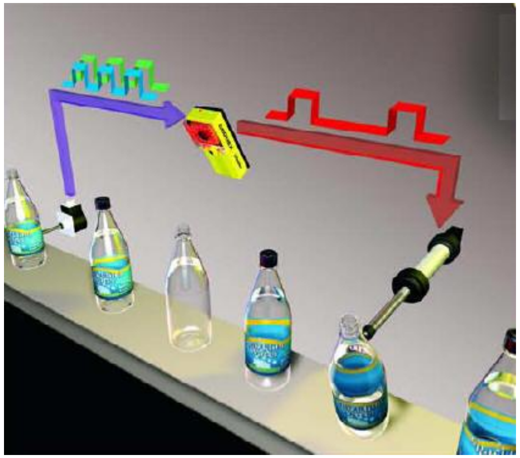
Motion sensing and gesture recognition

Automotive safety

Mobileye: Vision systems in high-end BMW, GM, Volvo models

Source: A. Shashua, S. Seitz

Computer vision and Applications



Factory inspection

Assistive technologies

Surveillance

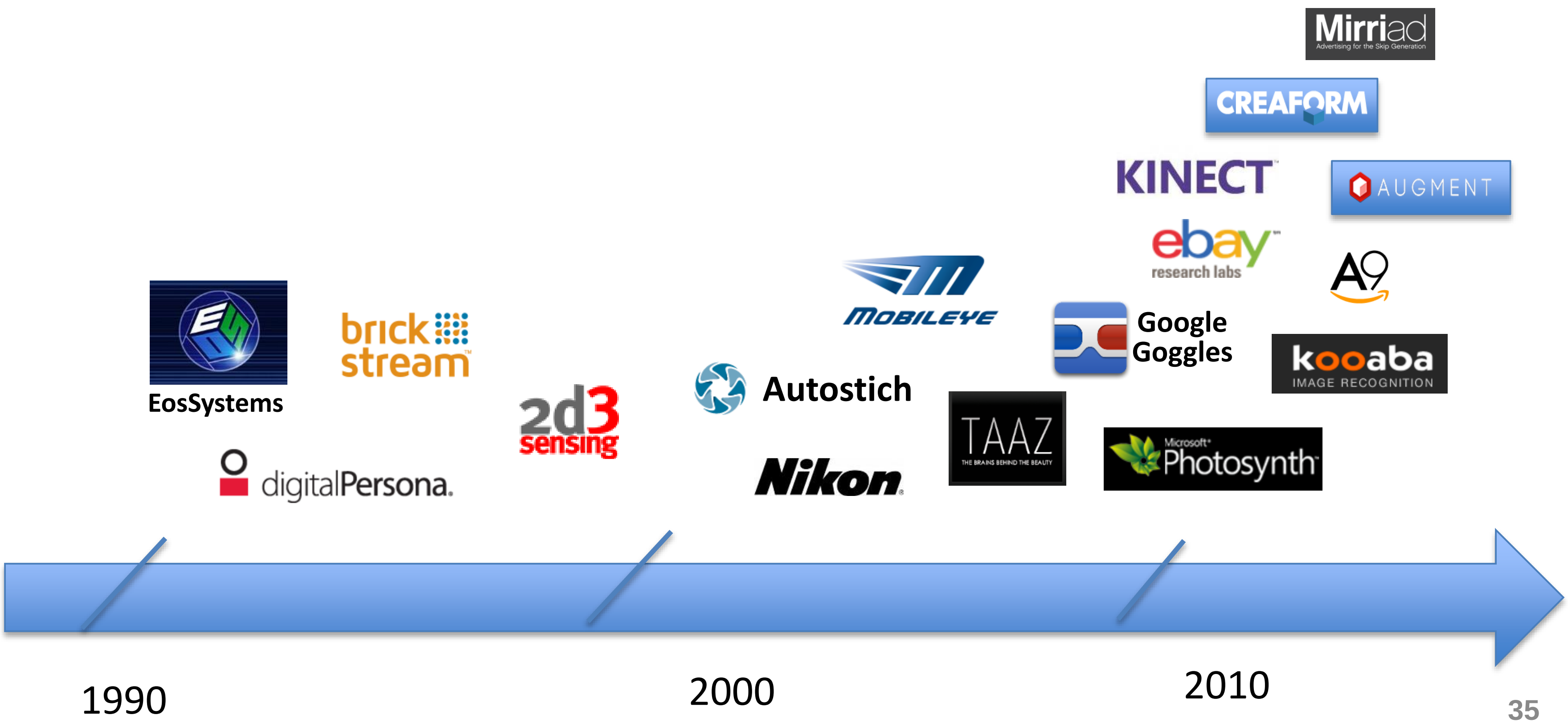
Autonomous driving,
robot navigation

Vision for robotics,
space exploration

Security

Sources: K. Grauman, L. Fei-Fei, S. Laznebick

Computer vision and Applications



Computer vision and Applications

Computer vision and Applications

EosSystems

2d3 sensing

Microsoft Photosynth

CREAFORM

AUGMENT

3D

digitalPersona

brickstream

Nikon

TAAZ

Google Goggles

A9

kooaba IMAGE RECOGNITION

2D

1990

2000

2010

Current state of computer vision

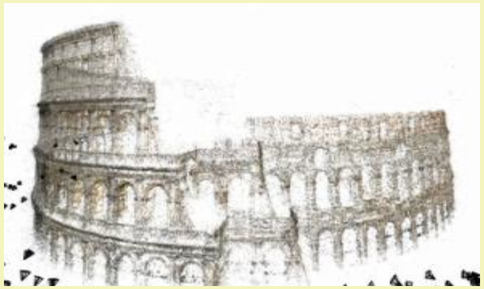
3D Reconstruction

- 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

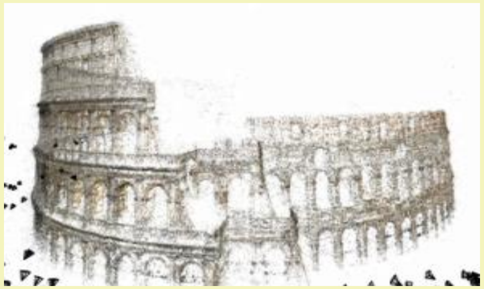
Snavely et al., 06-08

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

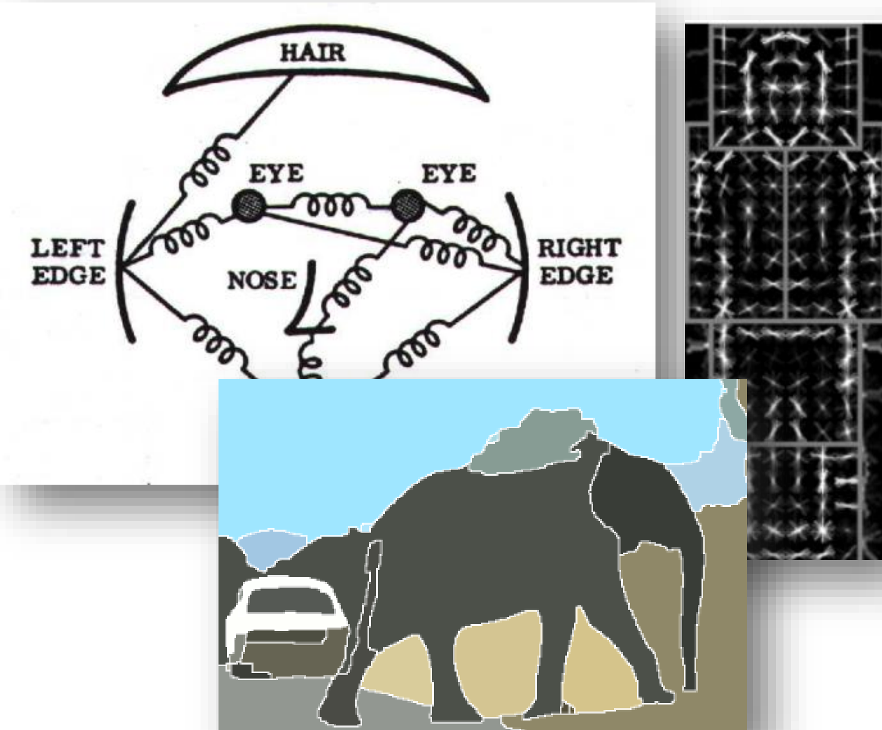
Snavely et al., 06-08

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

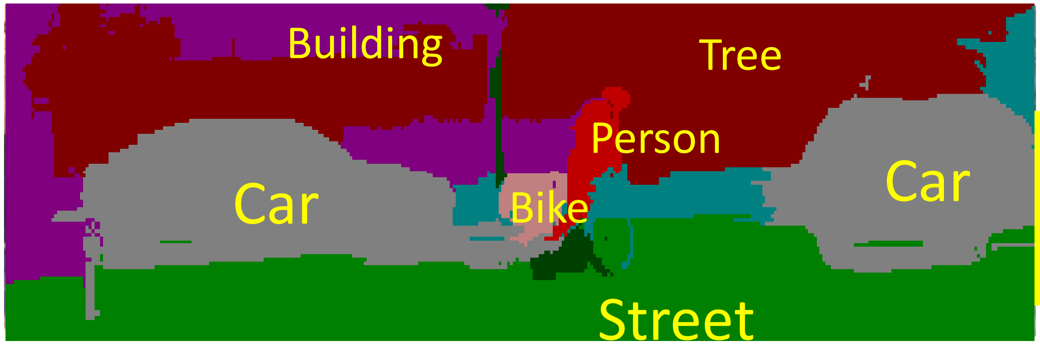
- Object detection
- Texture classification
- Target tracking
- Activity recognition

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

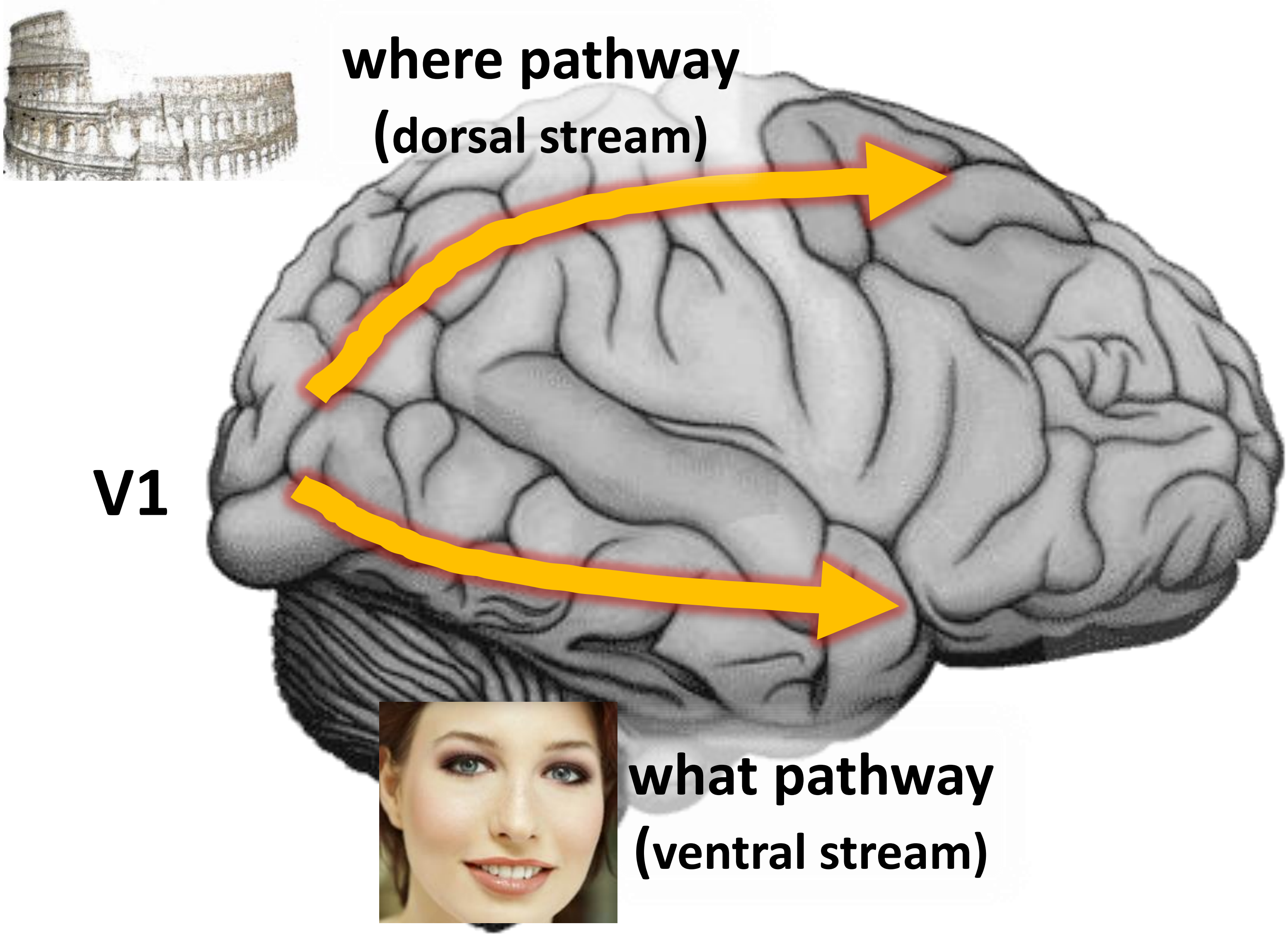
- 3D shape recovery
- 3D scene reconstruction
- Camera localization
- Pose estimation

2D Recognition

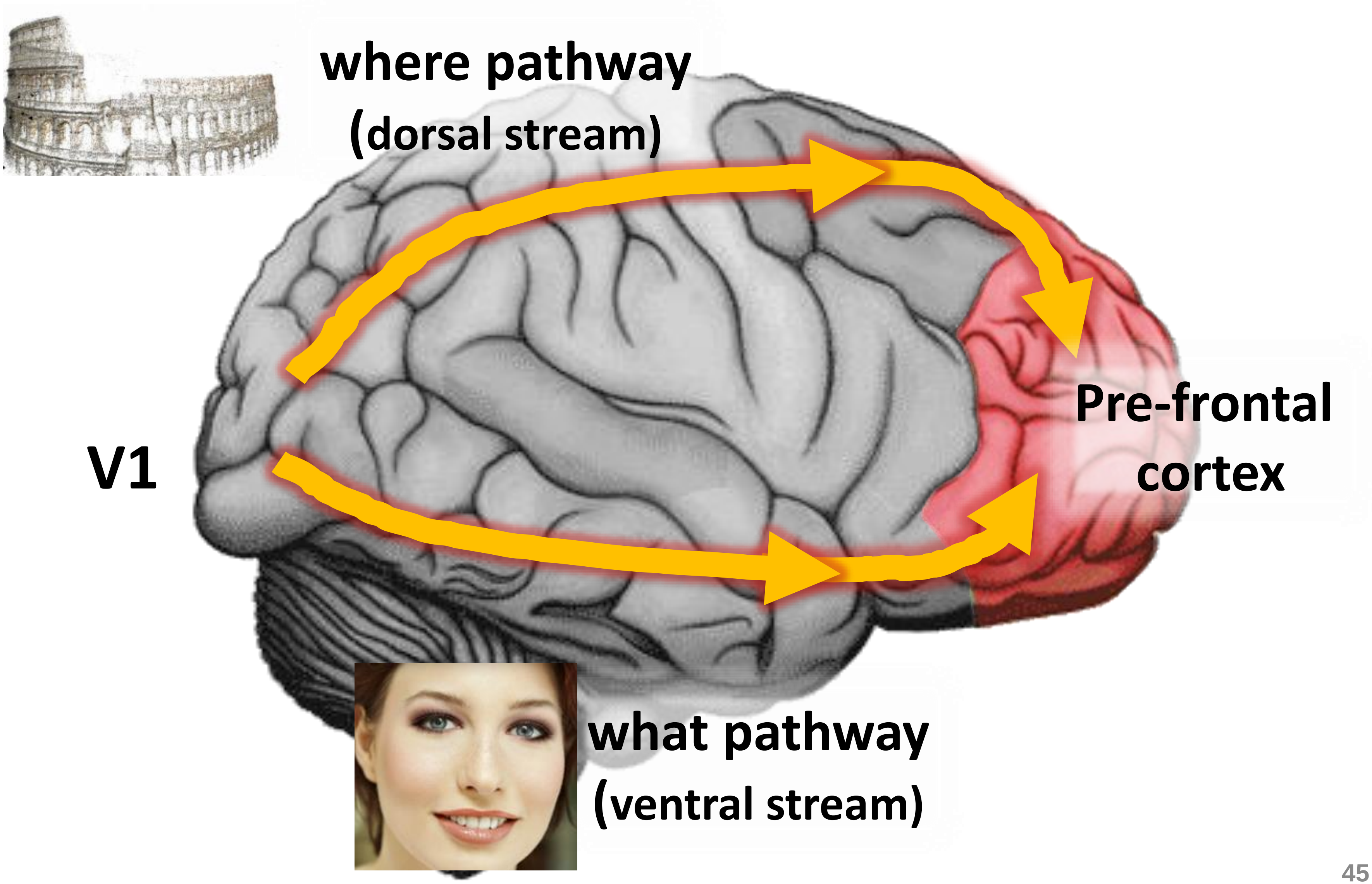
- Object detection
- Texture classification
- Target tracking
- Activity recognition

Perceiving the World in 3D!

Visual processing in the brain



Visual processing in the brain



CS 231A course overview

1. Geometry
2. Semantics

Geometry:

- How to extract 3d information?
- Which cues are useful?
- What are the mathematical tools?

Camera systems

Establish a mapping from 3D to 2D

How to calibrate a camera

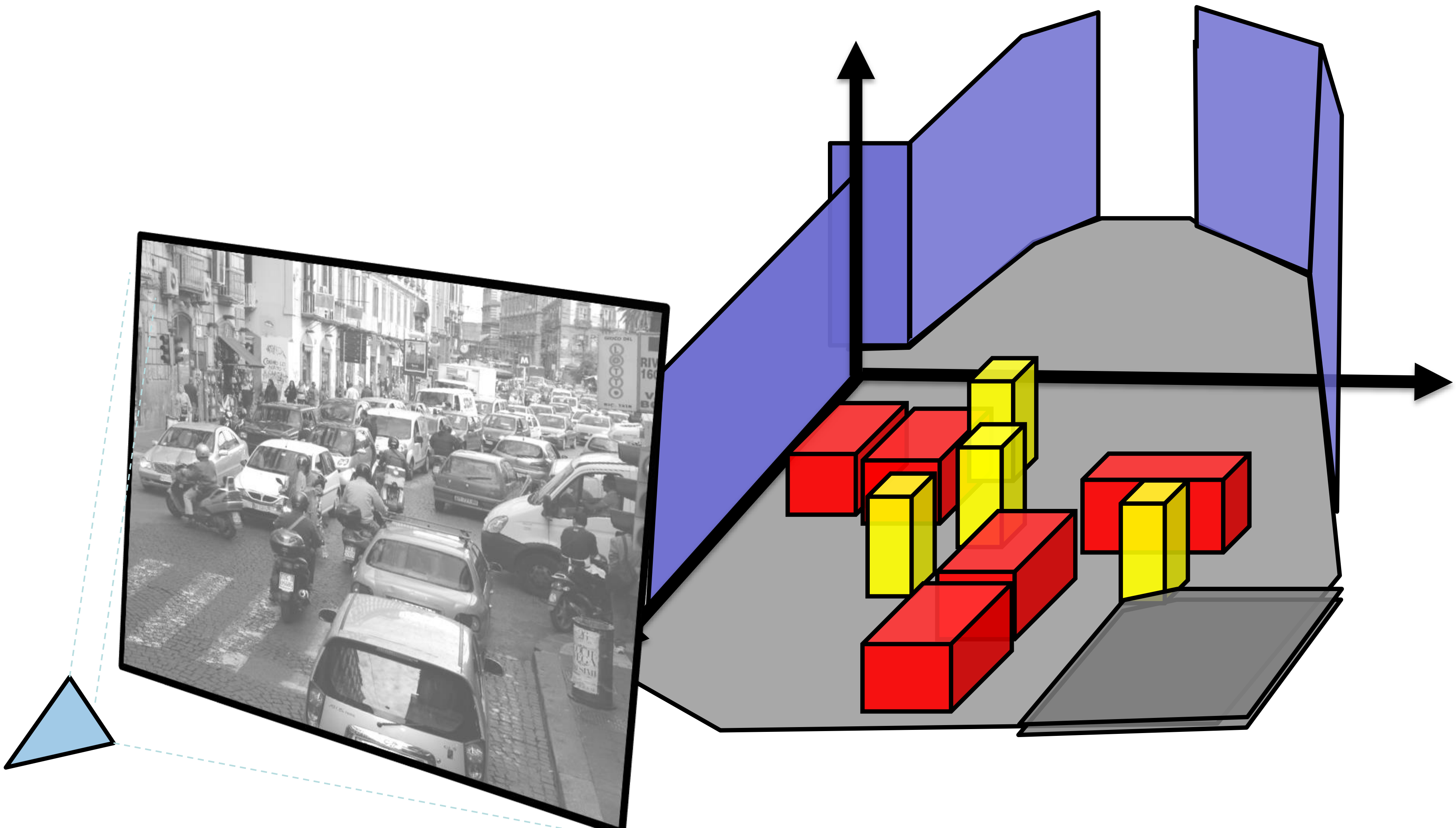
Estimate camera parameters such pose or focal length

Single view metrology

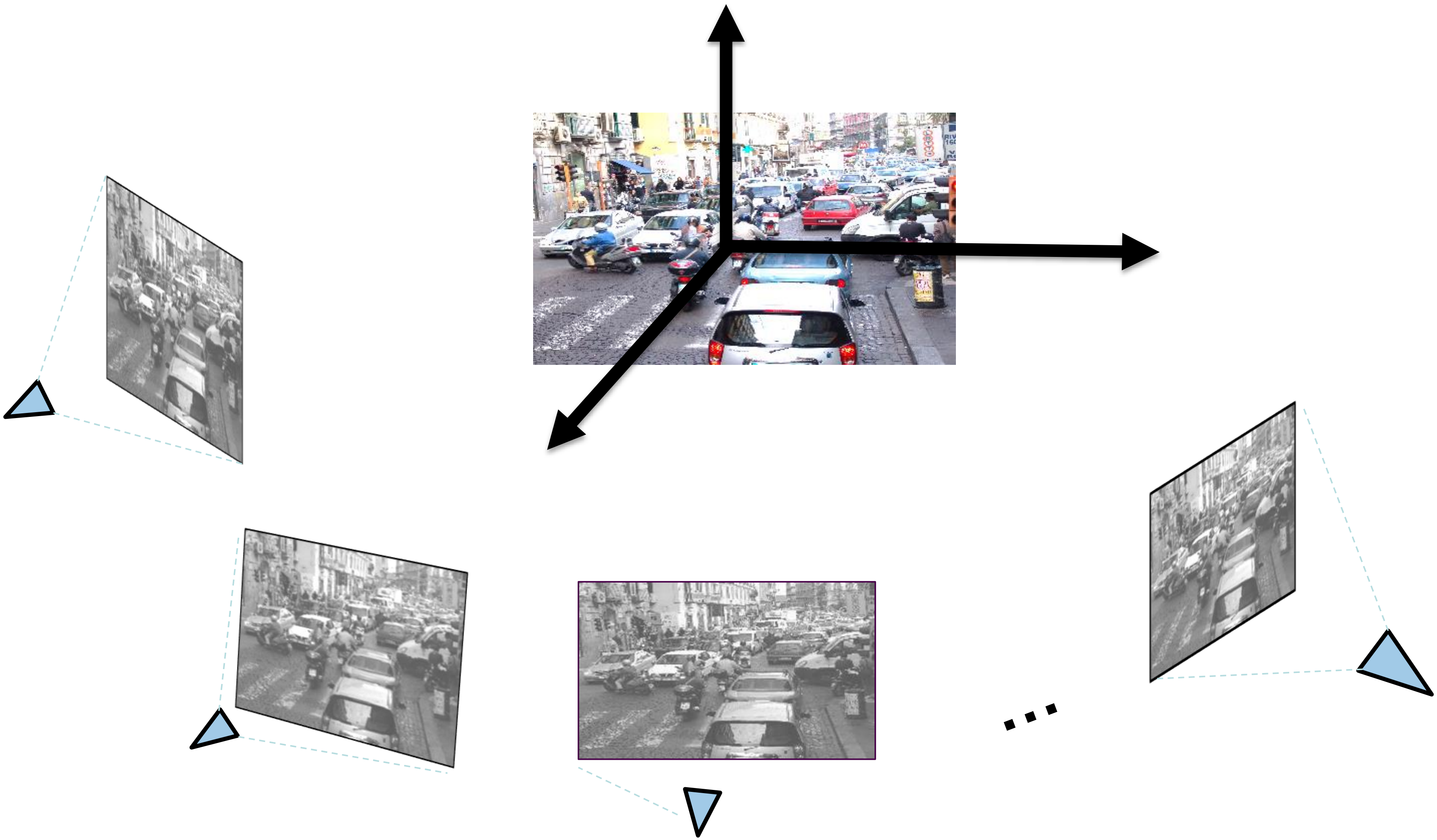
Estimate 3D properties of the world from a single image

Single view metrology

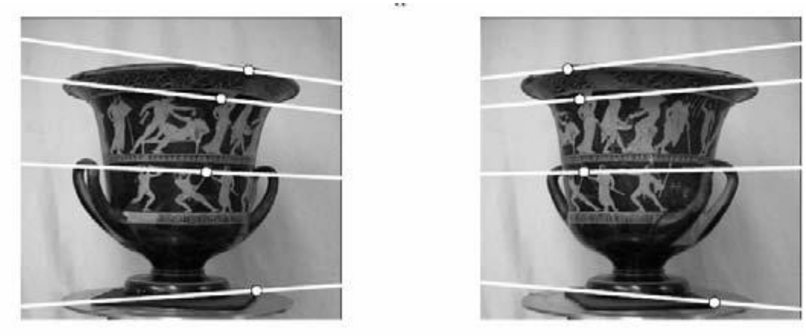
Estimate 3D properties of the world from a single image



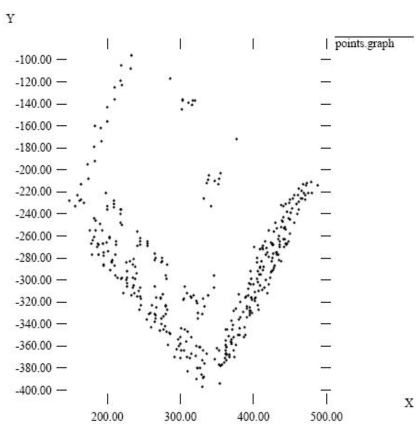
Multiple view geometry

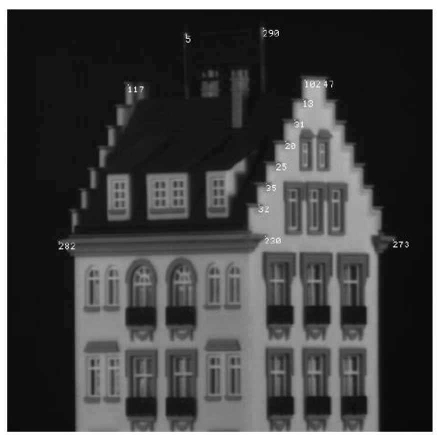


Mathematical tools



Epipolar geometry

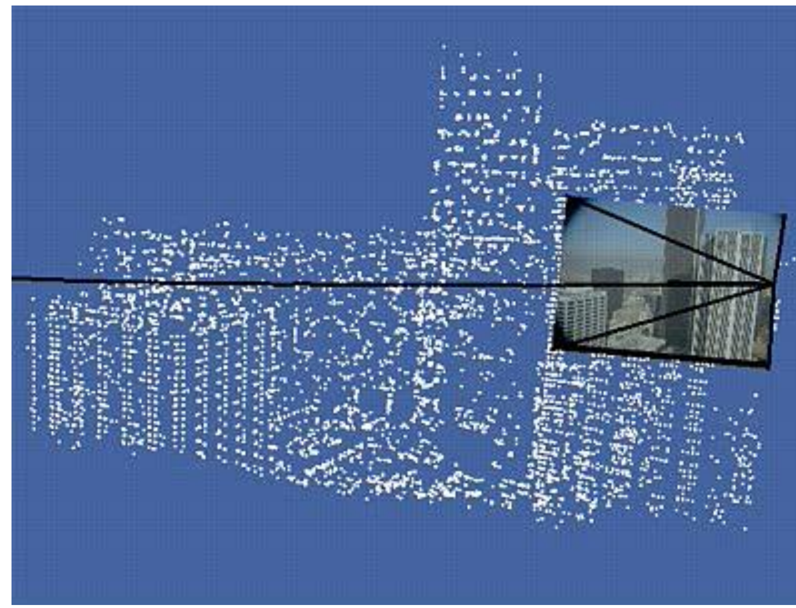




Tomasi & Kanade (1993)

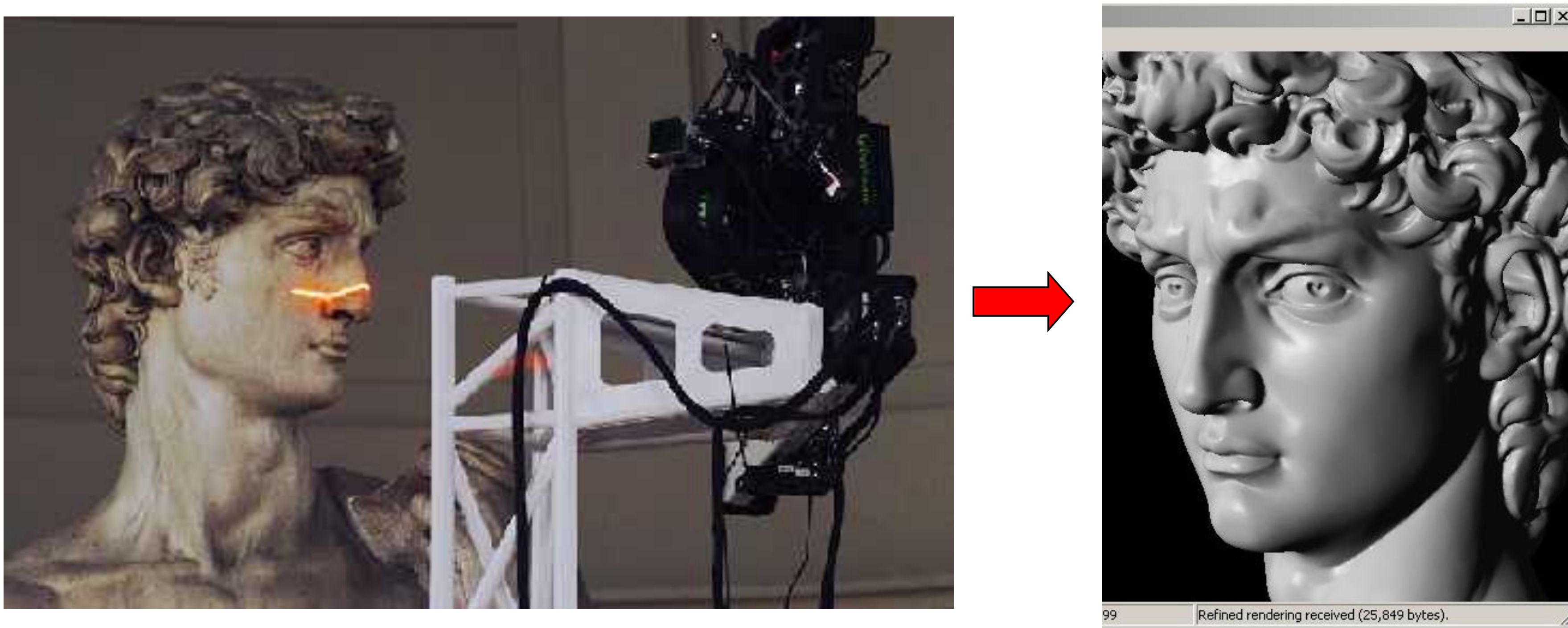
Photoconsistency

Structure from motion



Courtesy of Exford **Visual Geometry Group**

3D Models



Scanning Michelangelo's *"The David"*

- The Digital Michelangelo Project
  - http://graphics.stanford.edu/projects/mich/
- 2 BILLION polygons, accuracy to .29mm

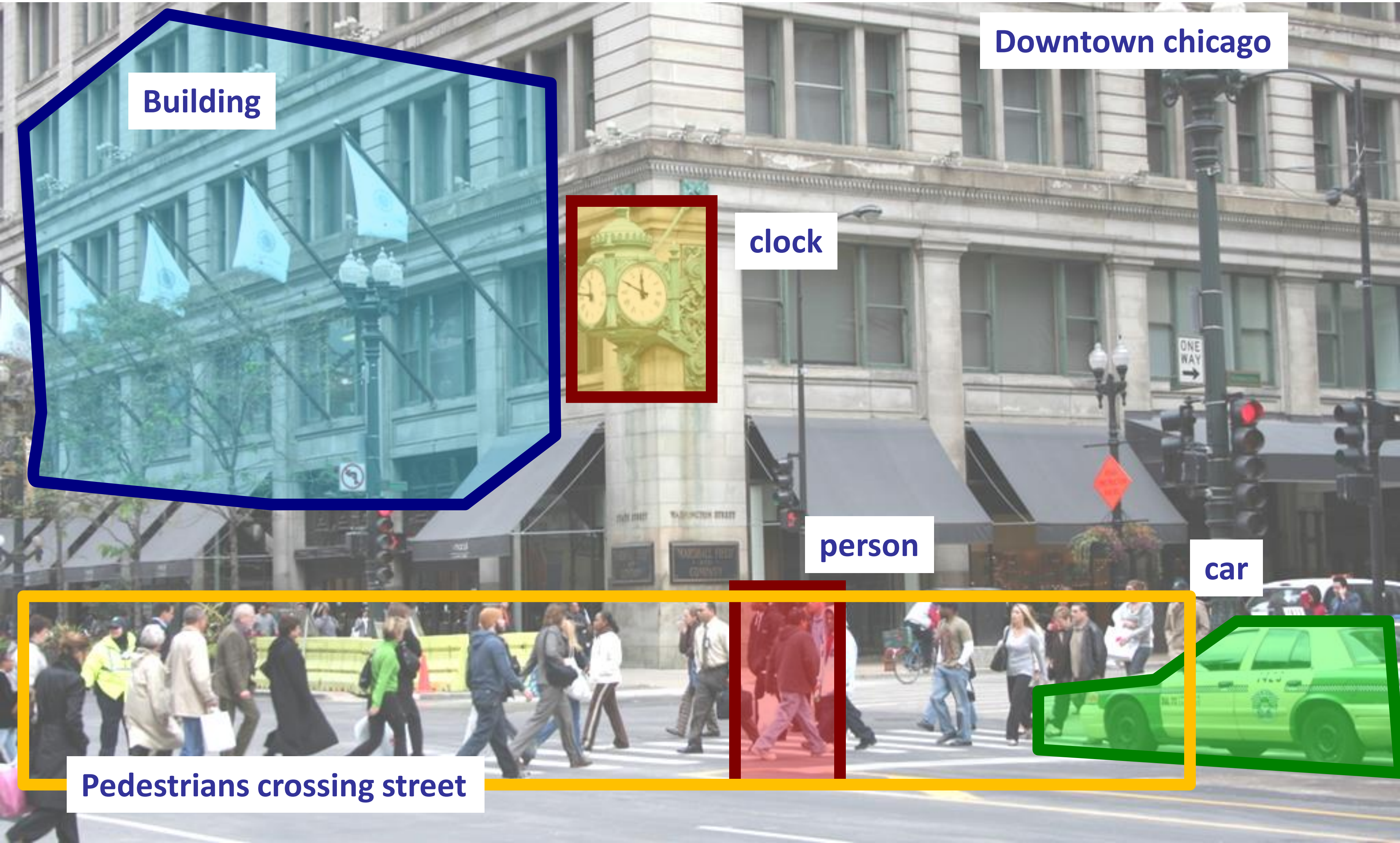
CS 231A course overview

1. Geometry
2. Semantics

Semantics:

- How to recognize objects?
- How to classify images or understand a scene?
- How to recognize what humans are doing?

Object recognition and categorization



Classification:
Is this an forest?

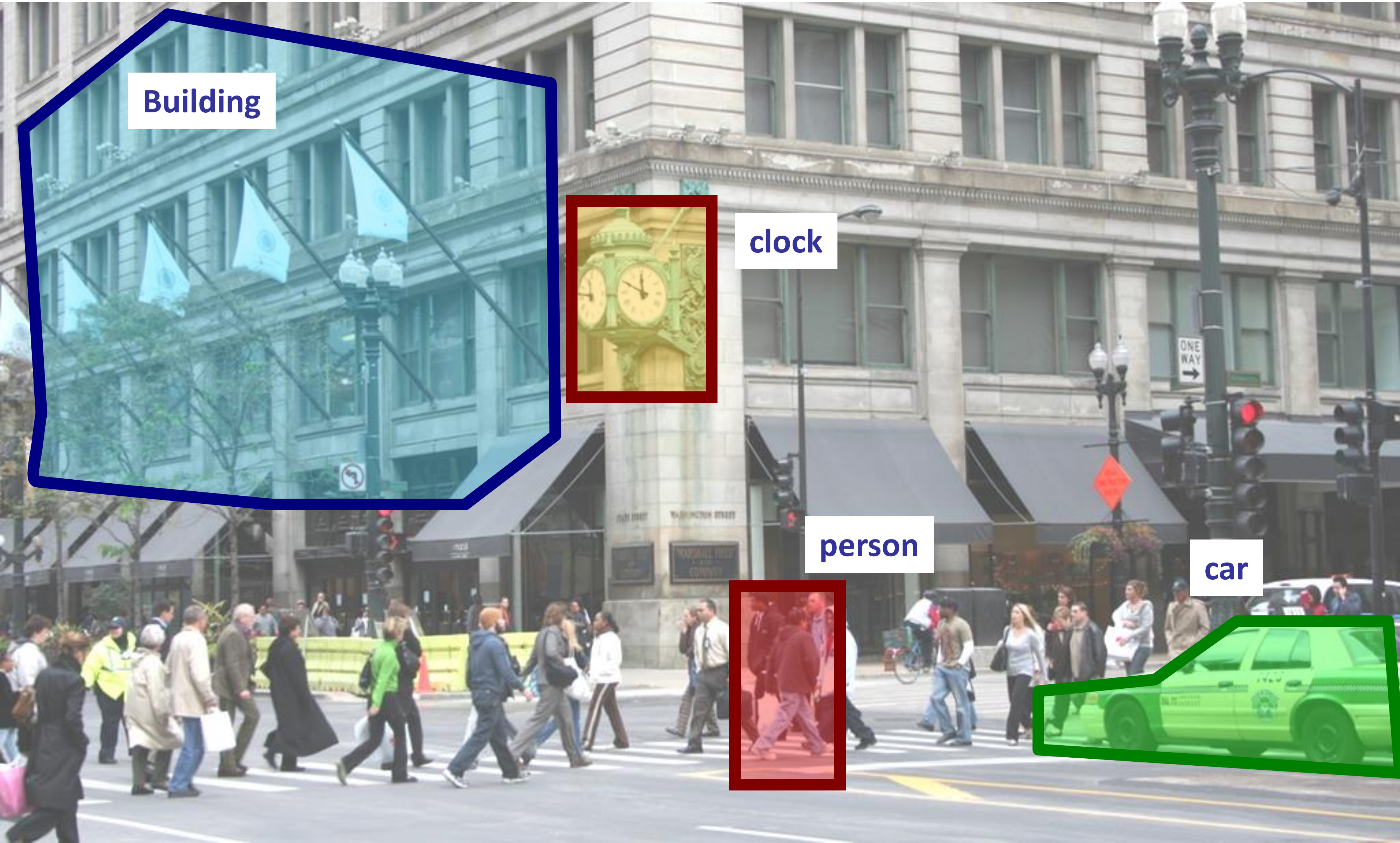
No!

Classification:
Does this image contain a building? [yes/no]

Detection:
Does this image contain a car? [where?]

Detection:

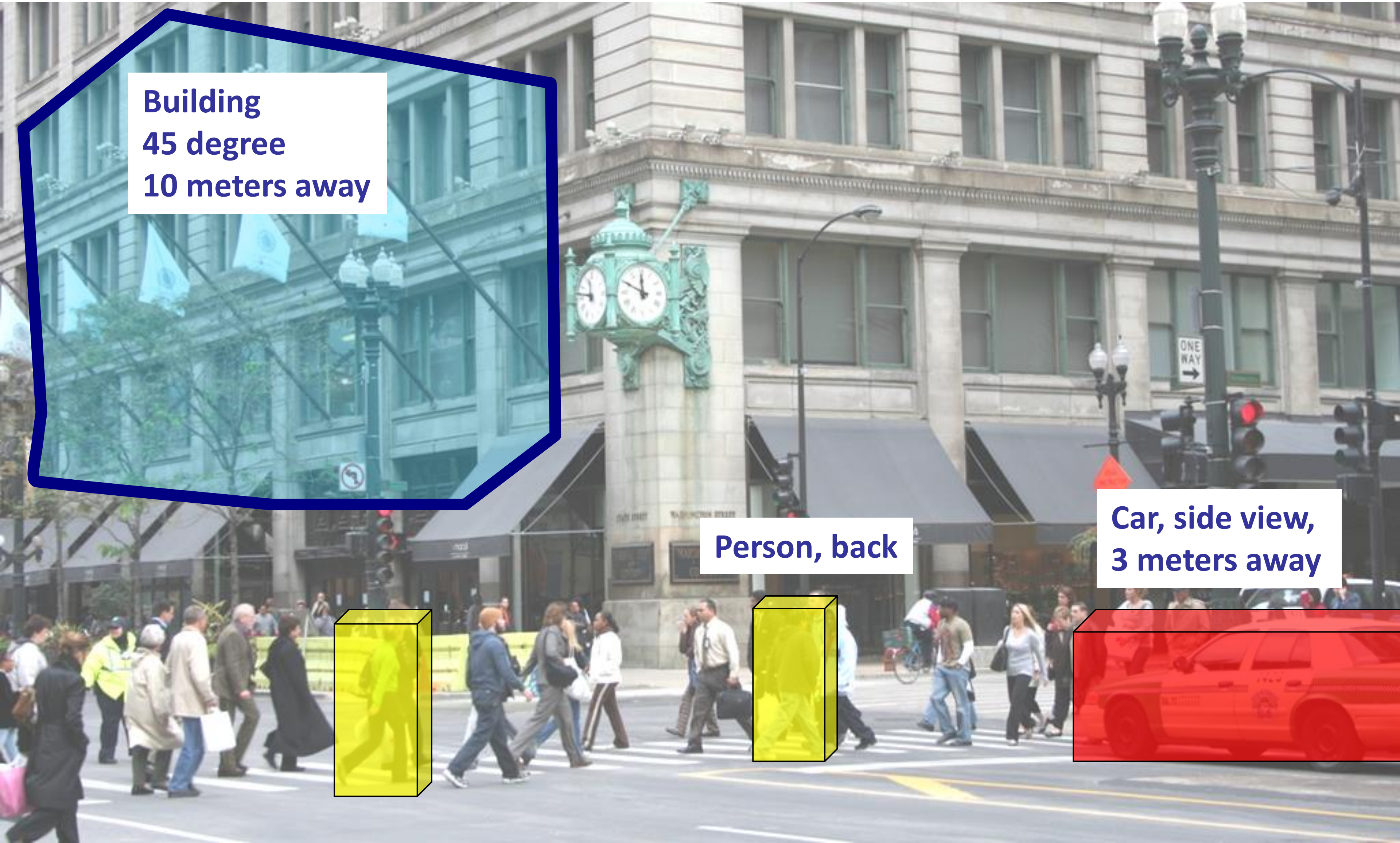
Which objects do this image contain? [where?]



Detection:
Accurate localization (segmentation)

Detection:

Estimating 3D geometrical properties



Challenges: viewpoint variation

slide credit: Fei-Fei, Fergus & Torralba

Challenges: illumination

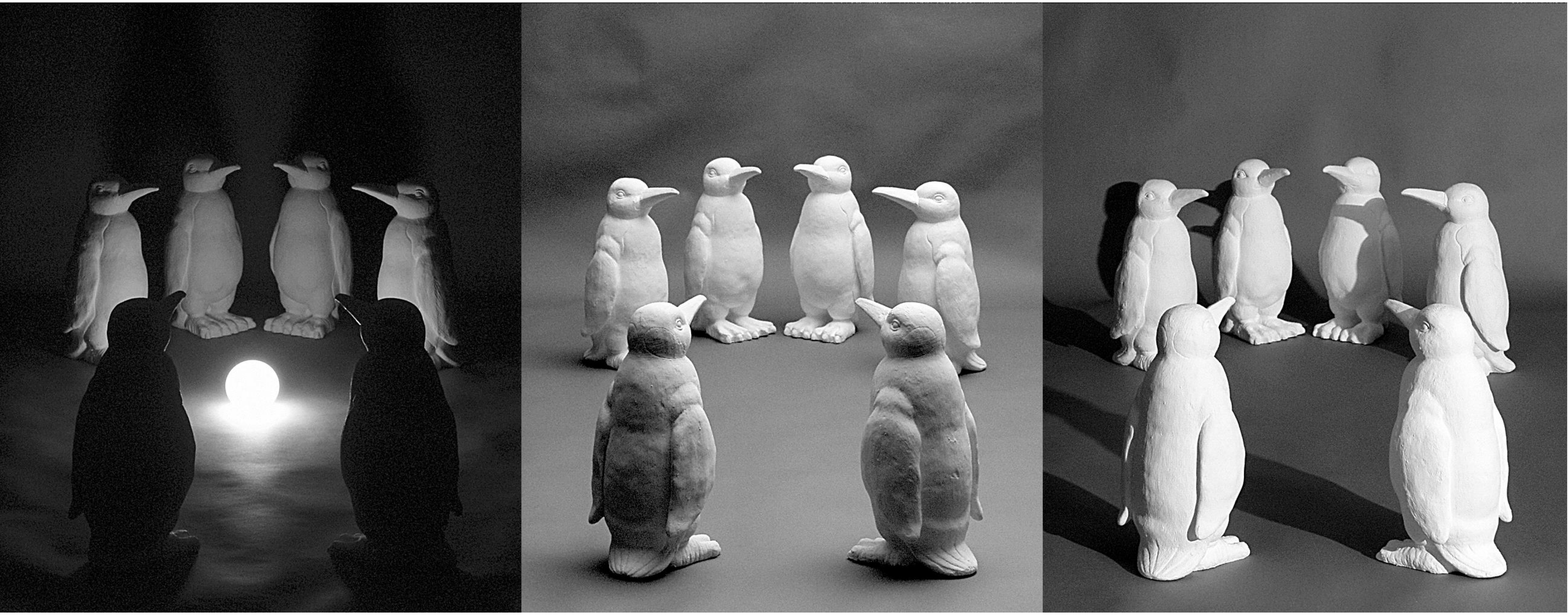


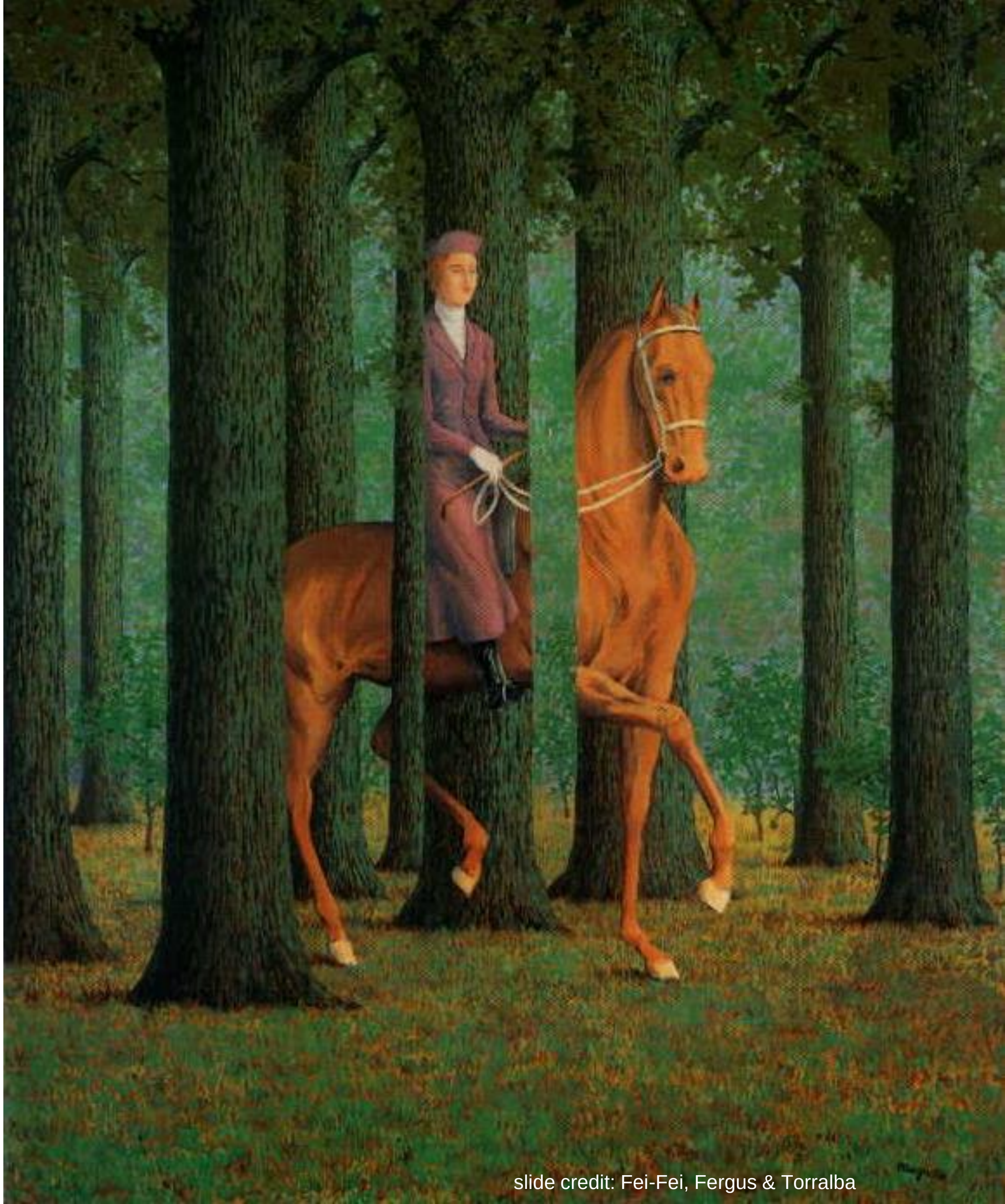
image credit: J. Koenderink

Challenges: scale

slide credit: Fei-Fei, Fergus & Torralba

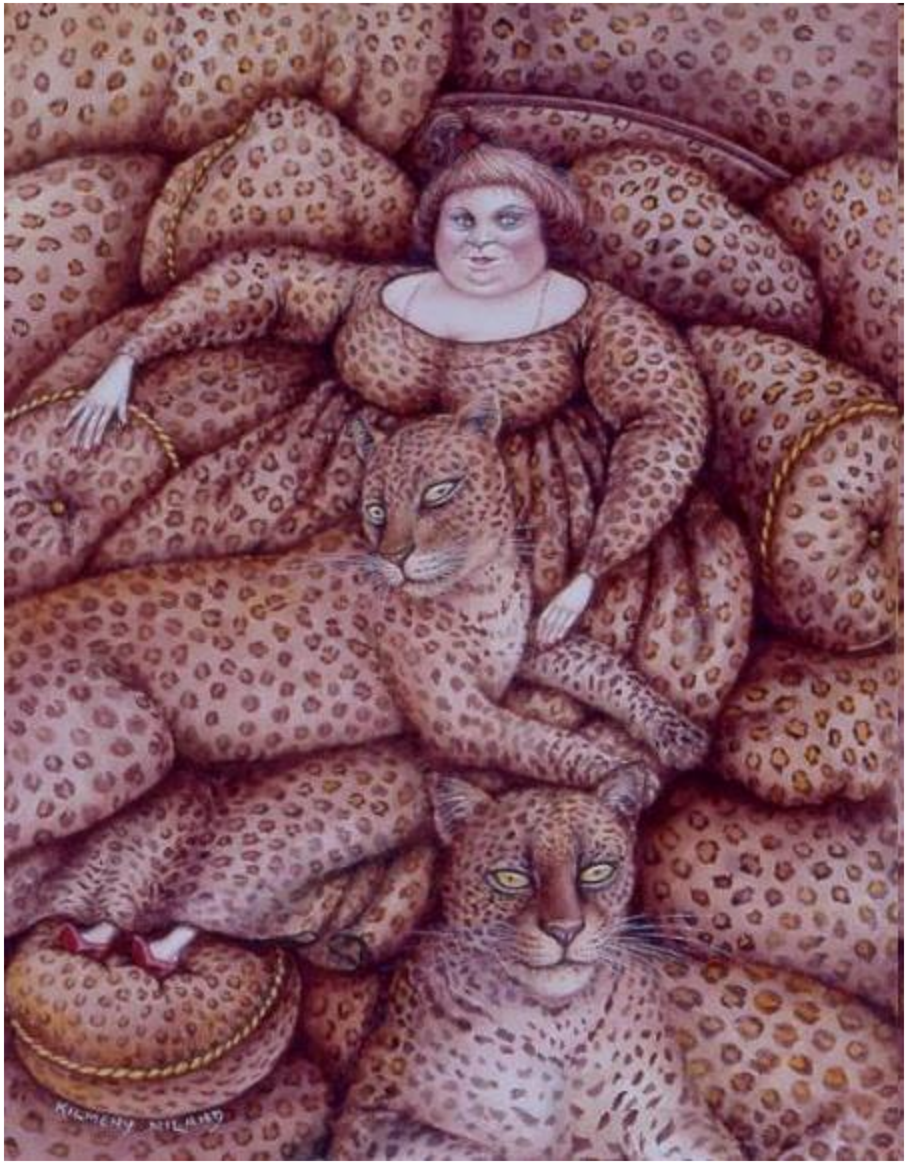
Challenges: deformation

Challenges: occlusion



Magritte, 1957

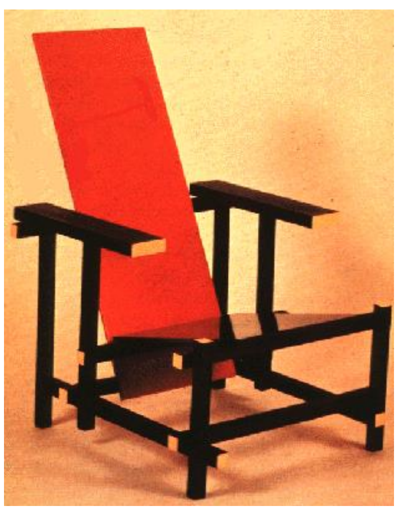
Challenges: background clutter



Kilmeny Niland. 1995

Challenges: object intra-class variation



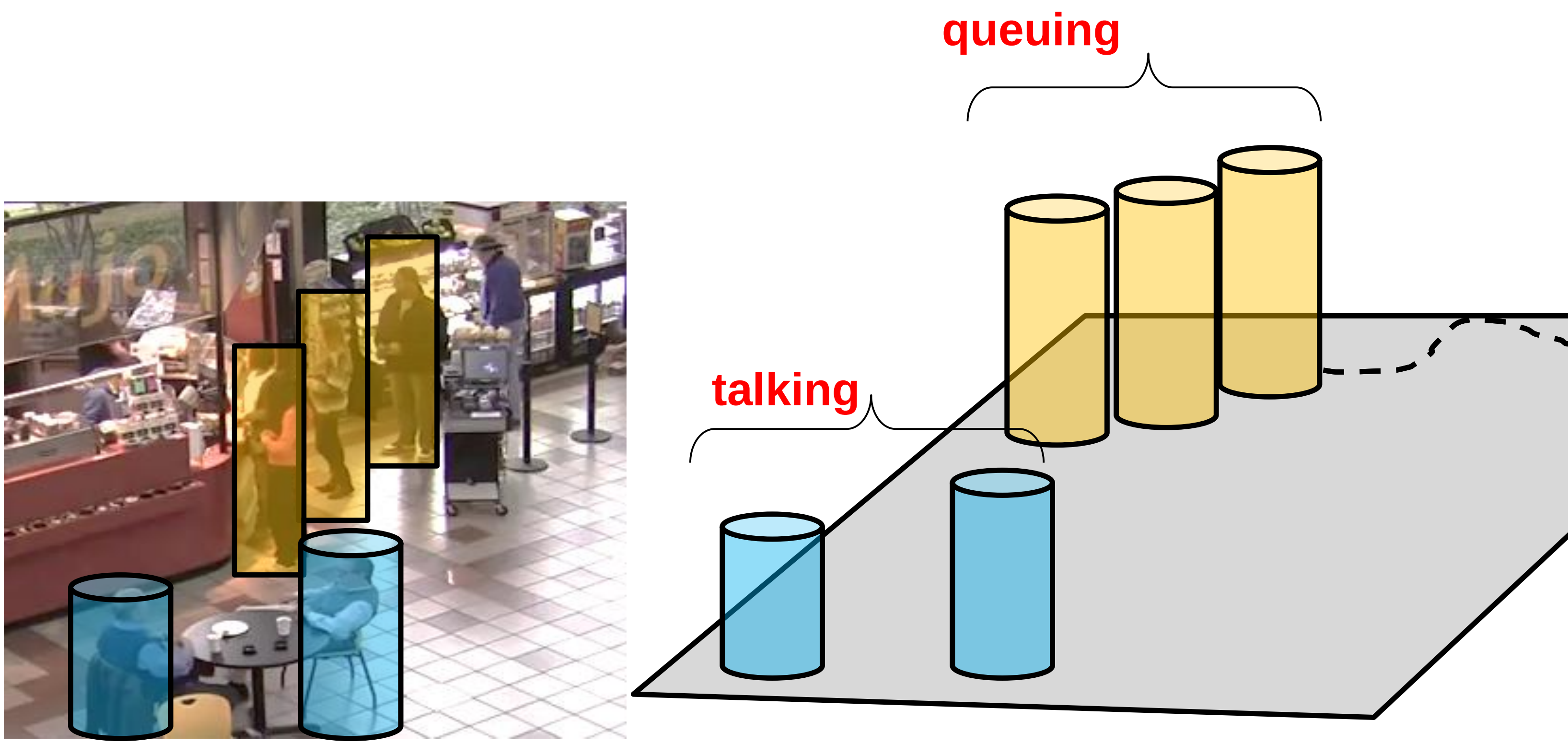



slide credit: Fei-Fei, Fergus & Torralba


~10,000 to 30,000

Activity understanding

Activity understanding



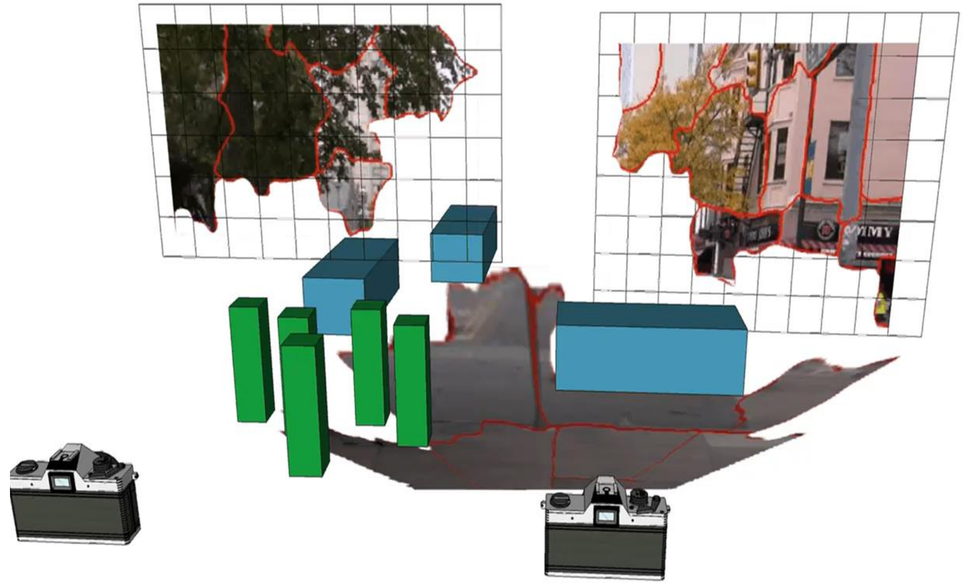
CS 231A course overview

1. Geometry
2. Semantics

Joint recovery of geometry and semantics!

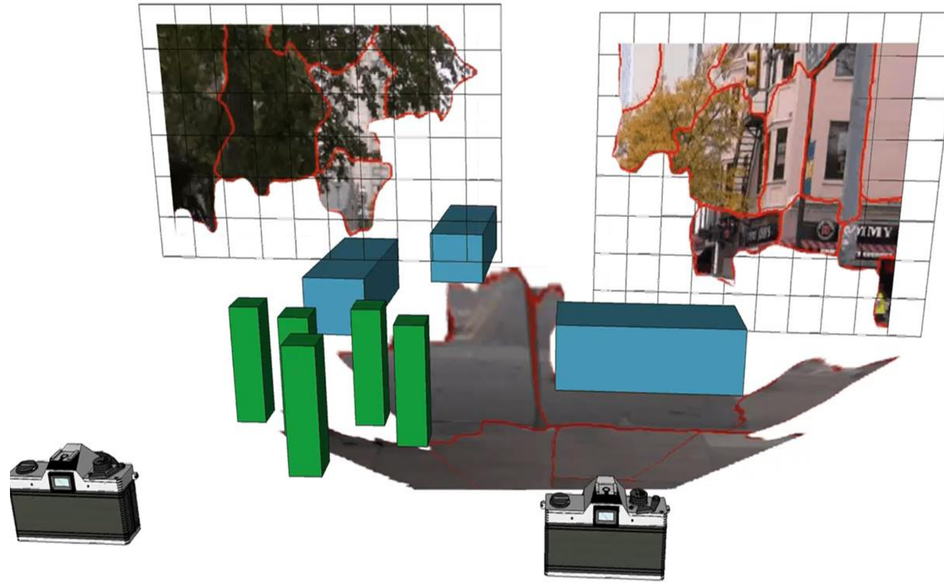
Joint reconstruction and recognition

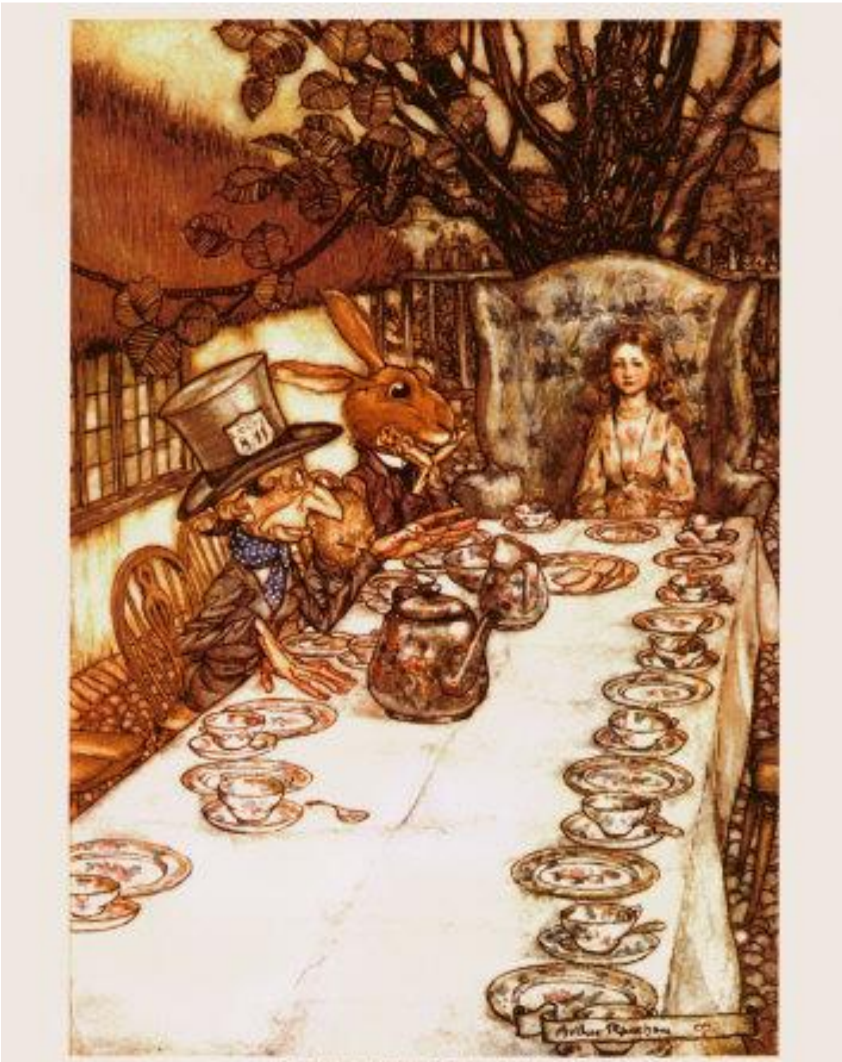
Input images



Joint reconstruction and recognition

Input images





"There was a table set out under a tree in front of the house, and the March Hare and the Hatter were having tea at it."

"The table was a large one, but the three were all crowded together at one corner of it …"

From "A Mad Tea-Party"
Alice's Adventures in Wonderland
by
Lewis Carroll

Syllabus

Lect.	Date	Topic	Link	Notes
1	Tues 1.7	Introduction		
2	Thur 1.9	Camera models		
3	Tues 1.14	Camera calibration		
4	Thur 1.16	Single view metrology		
5	Tues 1.21	Epipolar geometry & Stereo systems		
6	Wed 1.23	Structure from motion		
7	Tues 1.28	Structure from motion/ SLAM		
8	Thur 1.30	Volumetric stereo		
9	Tues 2.4	Fitting and Matching		
10	Thur 2.6	Recognition: intro; bag of words models (I)		
11	Tues 2.11	Visual Classification: bag of words models (II)		
12	Thur 2.13	Visual classification – deep nets		
13	Tues 2.18	Object detection		
14	Thur 2.20	3D Object recognition		
15	Tues 2.25	Scene understanding & segmentation		
16	Thur 2.27	Scene understanding & segmentation		
17	Tues 3.4	3D Scene understanding		
18	Thur 3.6	Activity understanding		
19	Tues 3.11	Project presentations		
20	Thur 3.13	Project presentations		

CS231
Introduction to Computer Vision

Next lecture: Camera systems