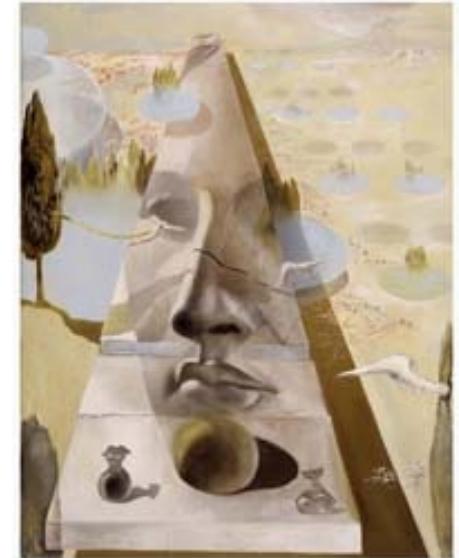


# CS231A

## Computer Vision: From 3D Reconstruction to Recognition



Class Time & Location:

M-W; 1:30–2:50PM – **Skilling Auditorium**

# CS231A

## Instructor



[Silvio Savarese](#)

- [ssilvio@stanford.edu](mailto:ssilvio@stanford.edu)
- Office: Gates Building, room: 154
- Office hour: Friday 2-3pm or under appoint.

## CAs:

- Amir Sadeghian
- Kuan Fang
- Danfei Xu
- Julian Gao
- Fei Xia

# Prerequisites

- This course requires knowledge of linear algebra, probability, statistics, machine learning and computer vision, as well as decent programming skills.
- Though not an absolute requirement, it is encouraged and preferred that you have at least taken either CS221 or CS229 or CS131A or have equivalent knowledge.
- We will leverage concepts from low-level image processing (CS131A) (e.g., linear filters, edge detectors, corner detectors, etc...) and machine learning (CS229) (e.g., **SVM**, **basic Bayesian inference**, **clustering**, neural networks, etc...) which we won't cover in this class.
- We will provide links to background material related to CS131A and CS229 (or discuss during TA sessions) so students can refresh or study those topics if needed.

# Text books

## Required:

- [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.

## Recommended:

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

# Course assignments

- 1 warm up problem set (HW-0)
  - 4 problem sets
  - 1 mid-term exam
  - 1 project
- 
- Look up class schedule for release and due dates.
  - Problems will be released through the [schedule page](#) and must be submitted through [Gradescope](#) (Use code MB5ZB9).

# Midterm Exam

- The exam will be held in class and you will have 80 minutes to complete it.
- You will be updated with more details, e.g., material to be covered, review sessions etc., as we approach the midterm.

# Course Projects

- Replicate an interesting paper
  - Comparing different methods to a test bed
  - A new approach to an existing problem
  - Original research
- 
- Write a 10-page paper summarizing your results
  - Release the final code
  - Give a final in-class presentation
  - SCPD students can send videos instead.
- 
- We will introduce projects in 1-2 weeks
  - Important dates: look up class schedule

# Course Projects

- Form your team:
  - 1-3 people
  - The larger is the team, the more work we expect from the team
  - Be nice to your partner: do you plan to drop the course?
- Evaluation
  - Quality of the project (including writing)
  - Final project in-class presentation (~ TBA minutes spotlight presentations)

# Grading policy

- Homeworks: **37%**
  - 1% for HW0
  - 9% for HW1, HW2, HW3, HW4 (each)
- Mid term exam: **20%**
- Course project: **38%**
  - Project proposal 1%
  - mid term progress report 5%
  - final report 25%
  - presentation 7%
- Attendance and class participation: **5%**
  - Questions, answers, remarks, piazza posts,...
  - Class participation are waived for SCPD students. For the project presentation, SCPD students can send videos instead.

# Grading policy (HWs)

- 25% will be deducted per day late.
- Two 48-hours one-time late submission “bonuses” are available; that is, you can use this bonus to submit your HW late after at most 48 hours. This is one time deal: After you use all your bonuses, you must adhere to the standard late submission policy.
- No exceptions will be made.

# Grading 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
- No "late submission bonus" is allowed when submitting your progress report or project report

# 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.

# Lecture 1

## Introduction



- An introduction to computer vision
- Course overview

"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**



"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."

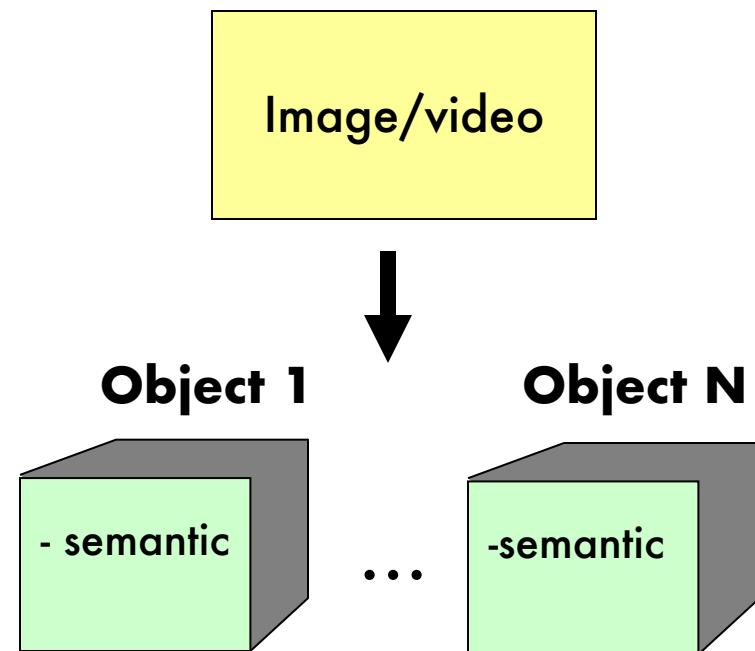
"The table was a large one, but the three were all crowded together at one corner of it ..."

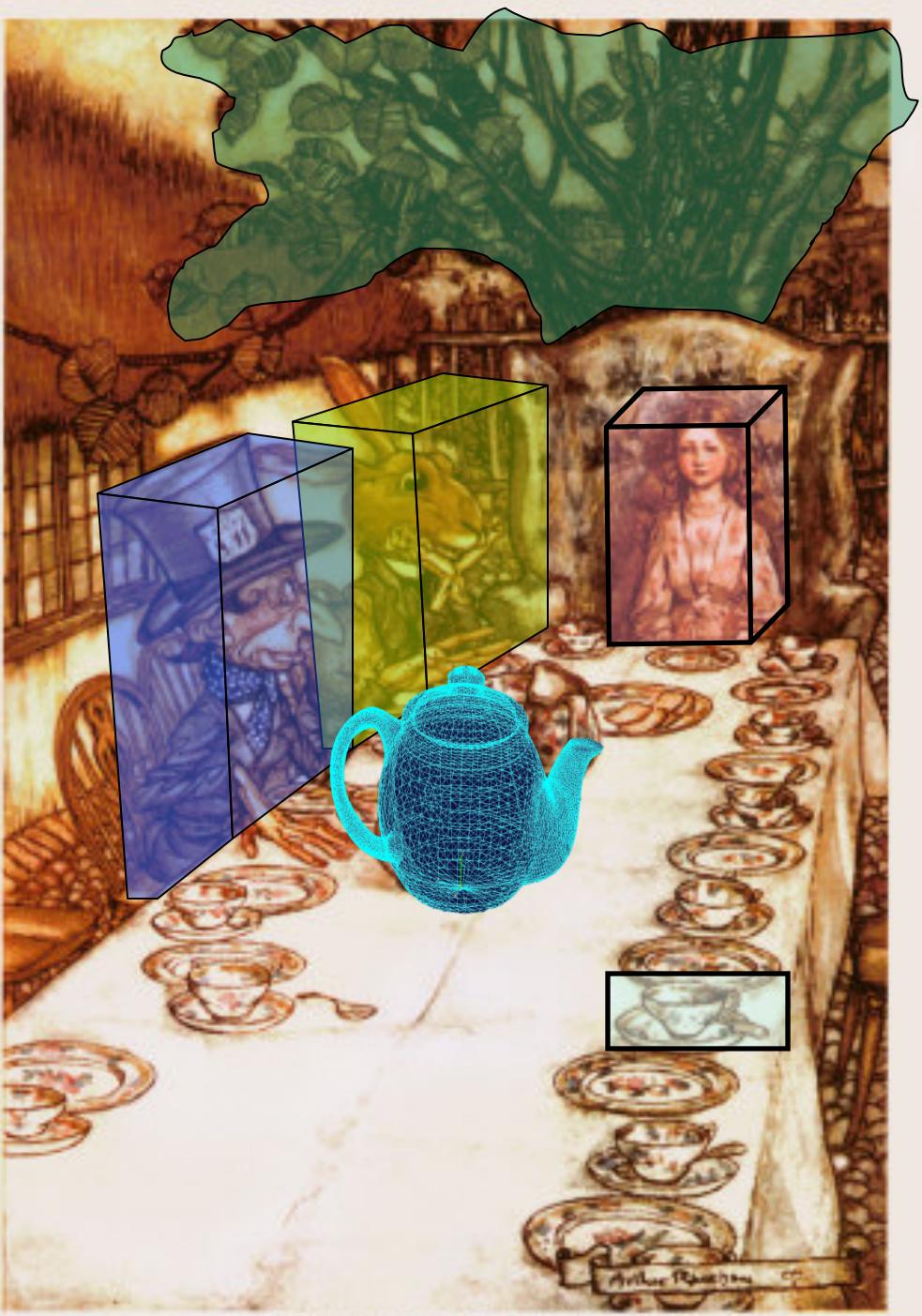
**From "A Mad Tea-Party"  
Alice's Adventures in Wonderland  
by  
Lewis Carroll**

**Illustration by Arthur Rackham**

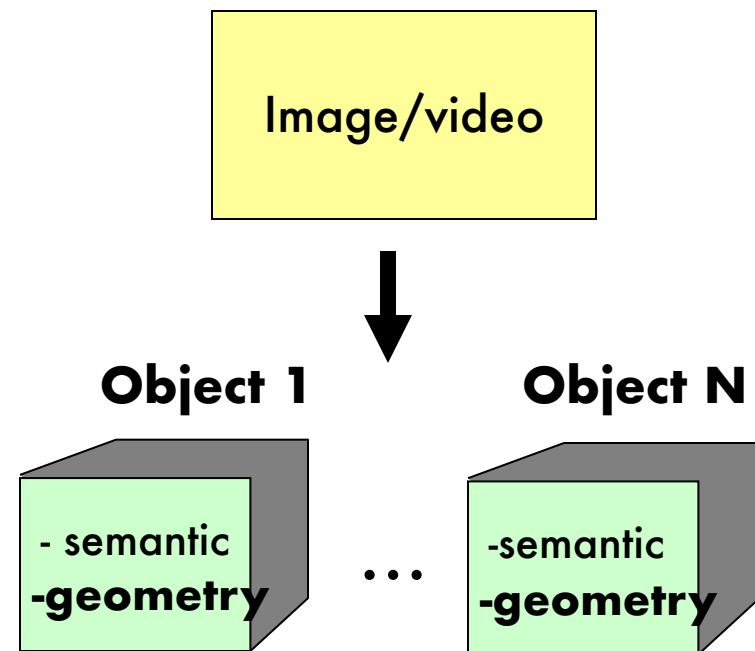


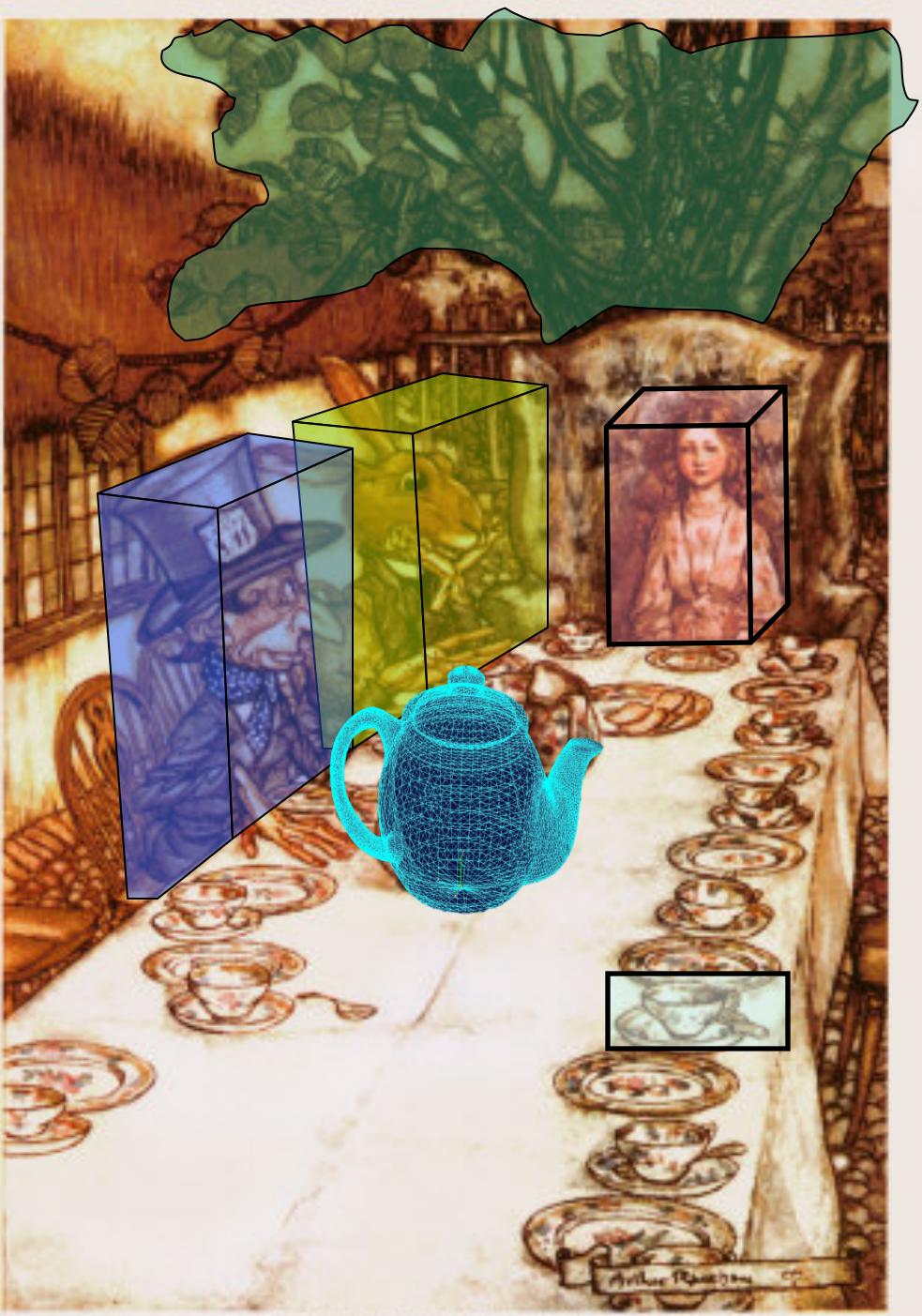
# Computer vision



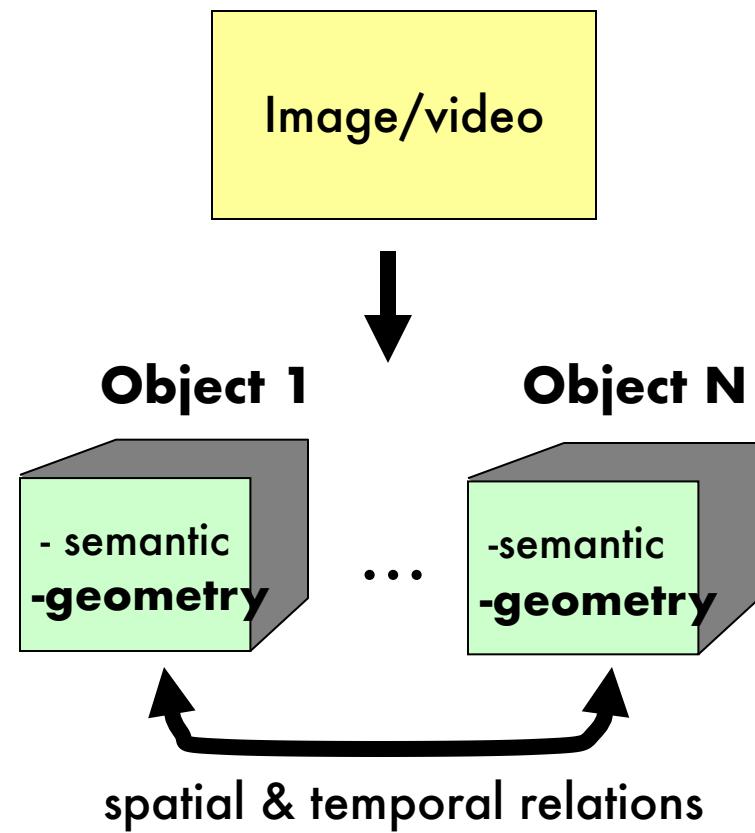


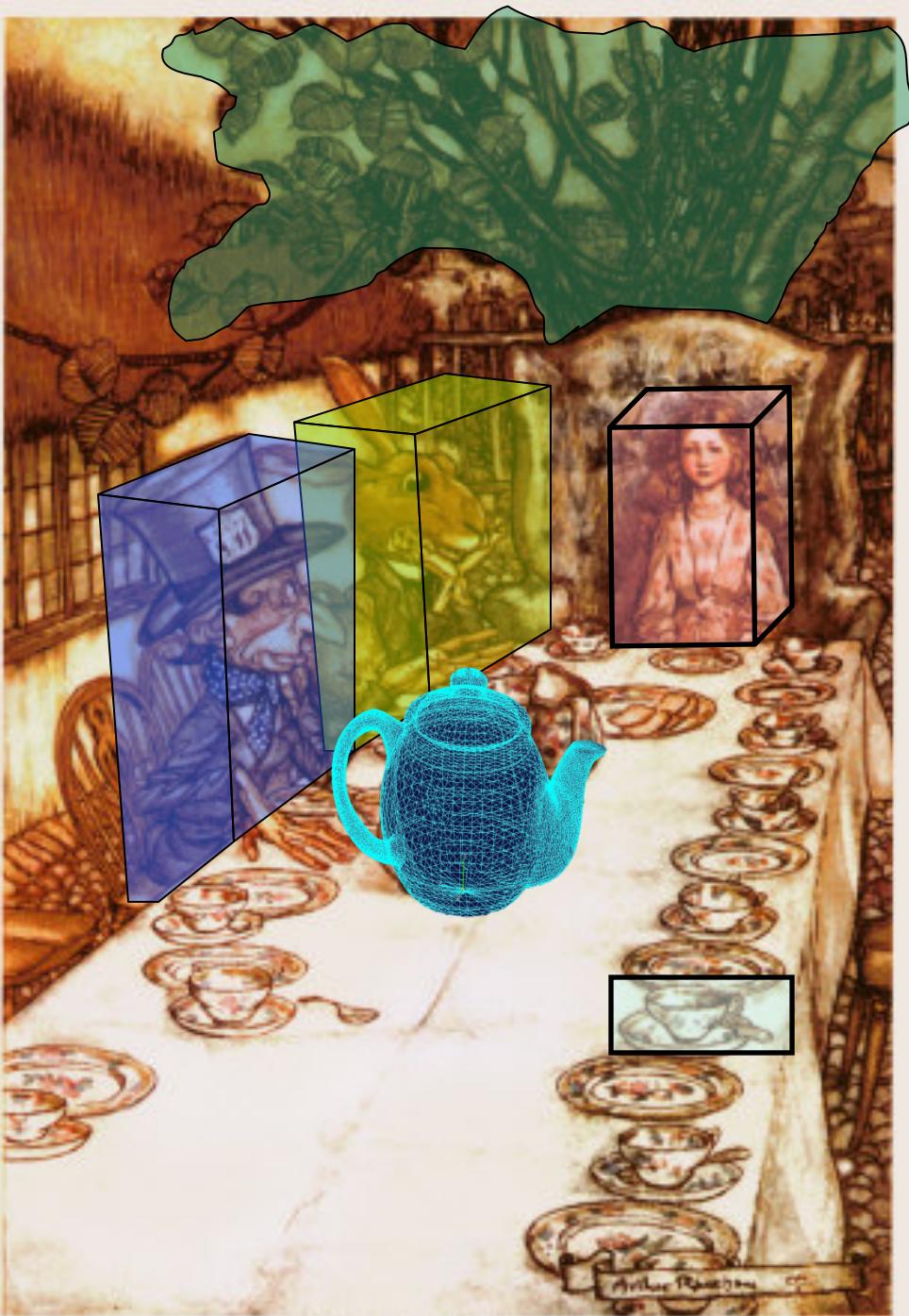
# Computer vision



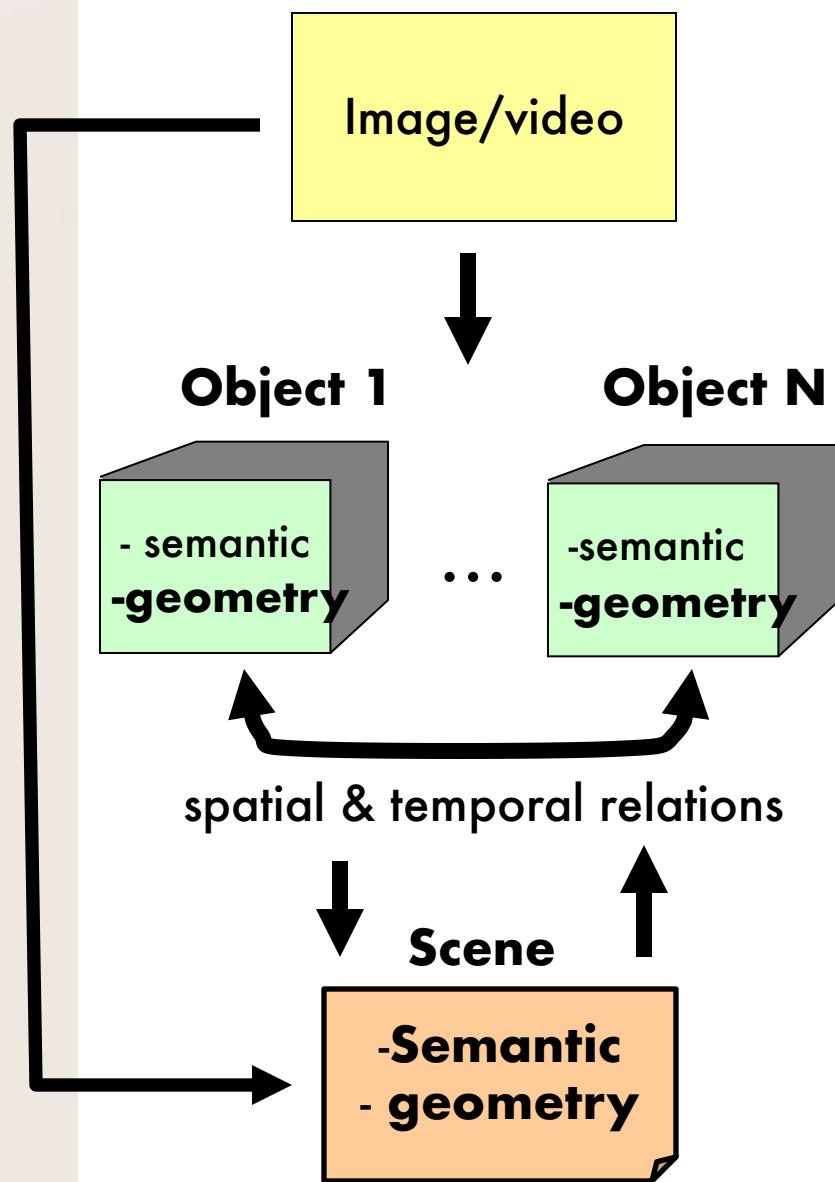


# Computer vision

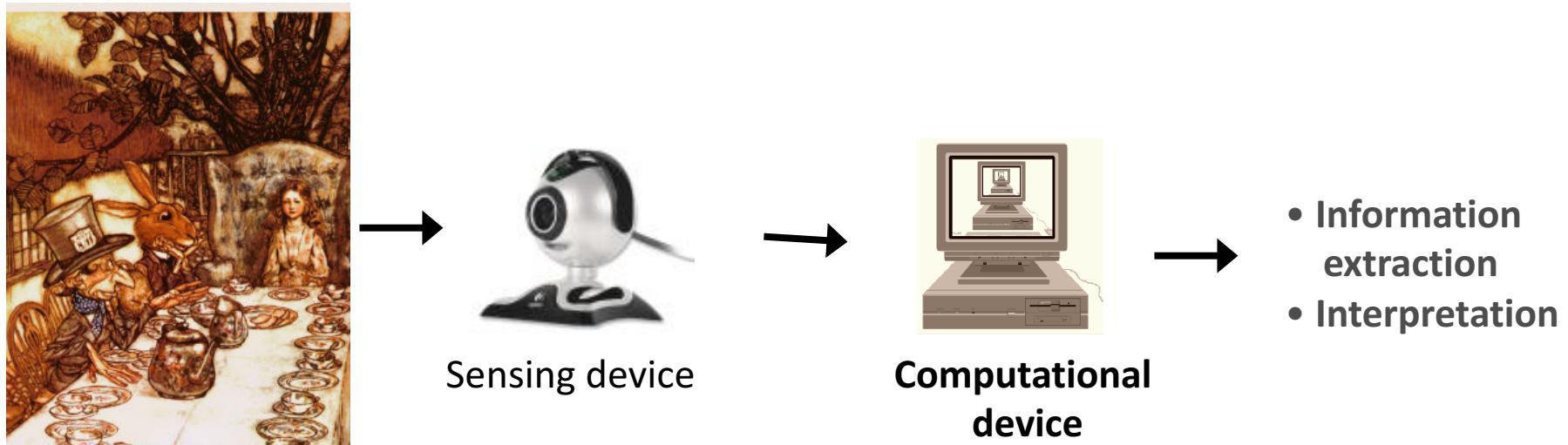




# Computer vision



# Computer vision



- 1. Information extraction:** features, 3D structure, motion flows, etc...
- 2. Interpretation:** recognize objects, scenes, actions, events

# Computer vision and Applications



# Fingerprint biometrics



 digitalPersona.



# Augmentation with 3D computer graphics



# 3D object prototyping



# Computer vision and Applications

- New features detector/descriptors
- CV leverages machine learning



EosSystems



brick stream™



1990



2d3  
sensing



MOBILEYE



Autostich



2000



TAAZ  
THE BRAINS BEHIND THE BEAUTY



Microsoft®  
Photosynth

2010

# 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](#)

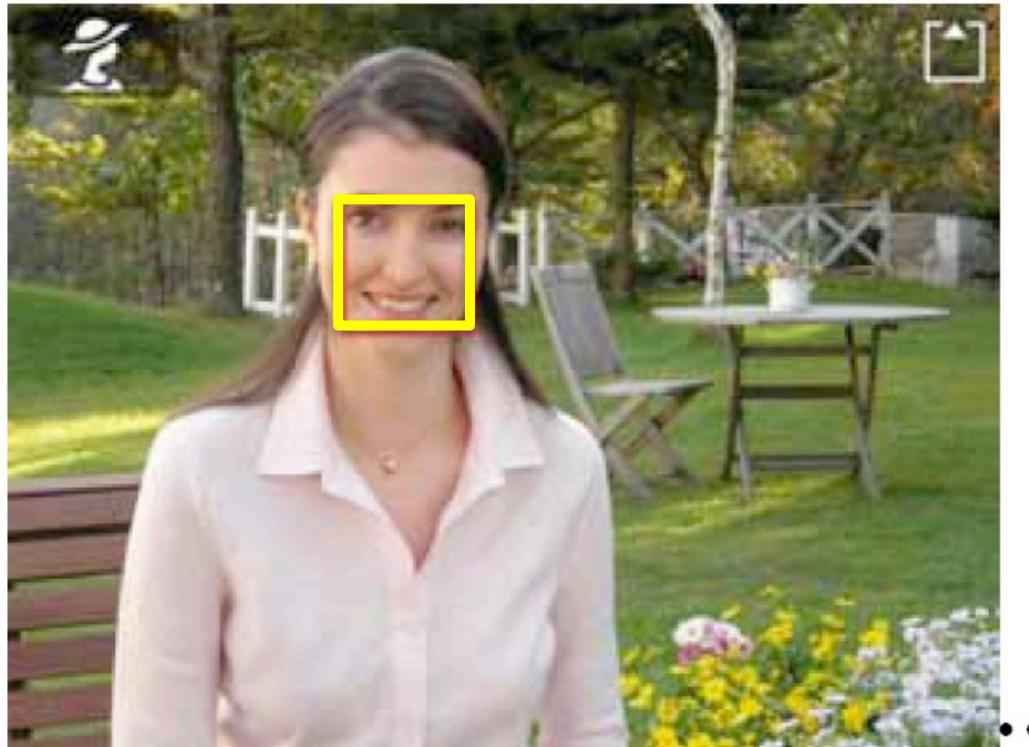
## Face-hunting cameras boost Nikon

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



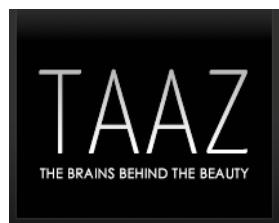
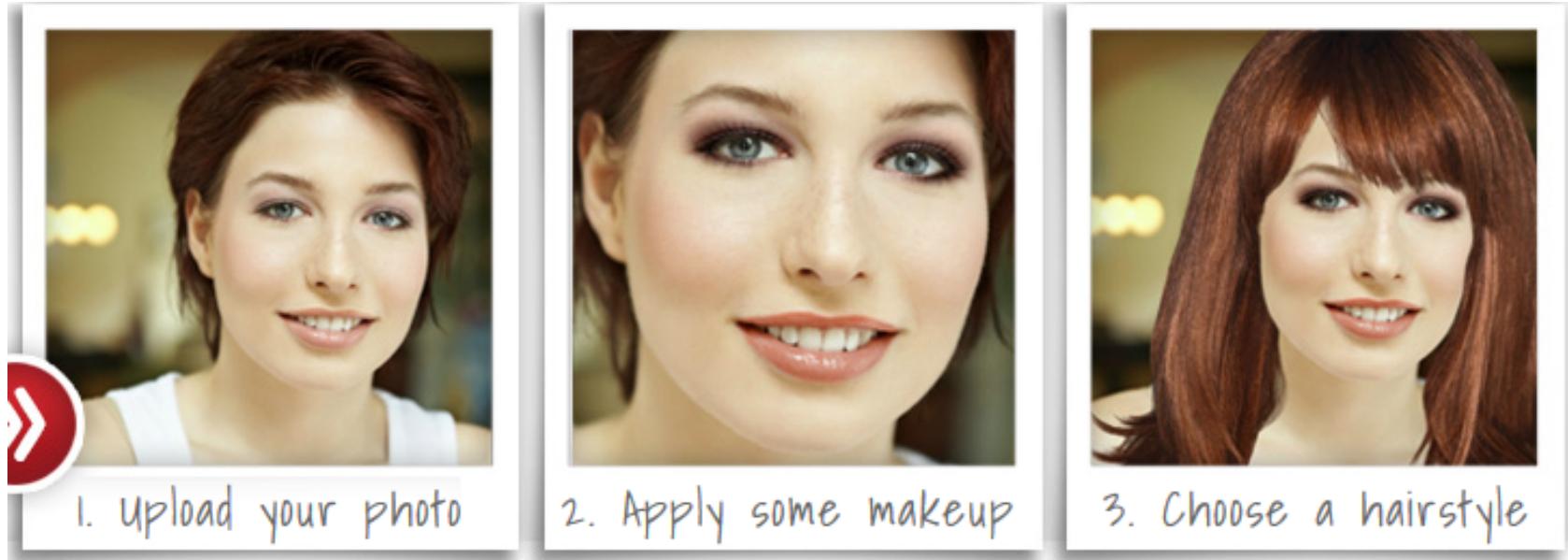
Face recognition cameras like the Coolpix L1 are popular

# Face detection



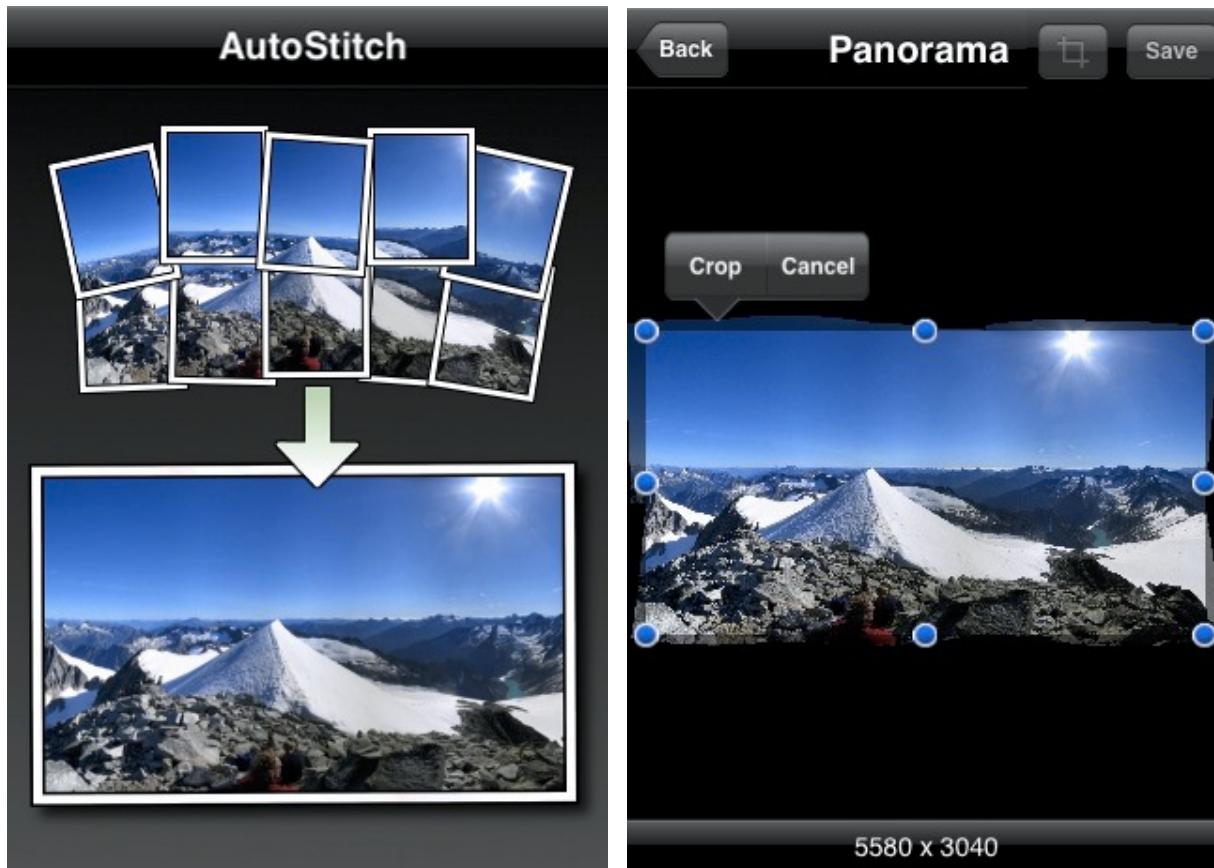
**Sample image:** Subject as seen on the COOLPIX 5900 camera's color LCD and when using Nikon's Face-priority AF function.

# Web applications

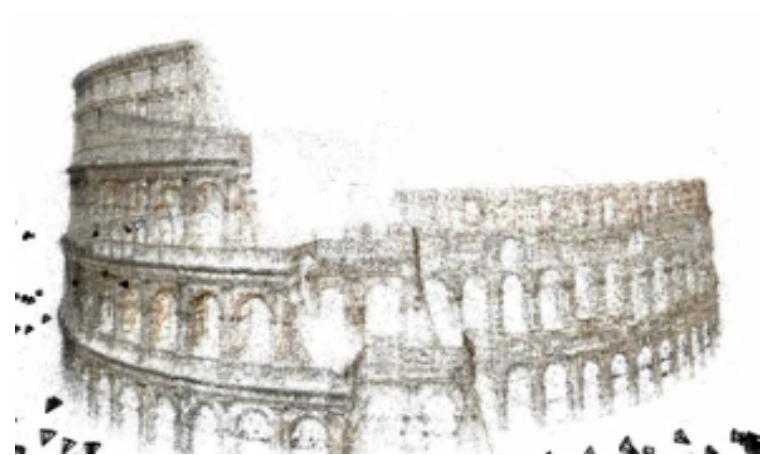
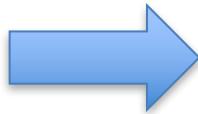


Photometria

# Panoramic Photography



# 3D modeling of landmarks



# Computer vision and Applications

- Efficient SLAM/SFM
- Large scale image repositories
- Deep learning (e.g. ImageNet)



EosSystems



brick  
stream™



1990



2d3  
sensing



Autostich



2000



TAAZ  
THE BRAINS BEHIND THE BEAUTY



Microsoft®  
Photosynth

2010

# Computer vision and Applications

- Efficient SLAM/SFM
- Large scale image repositories
- Deep learning (e.g. ImageNet)
- Better clouds ☺
- More bandwidth
- Increase computational power



Autostich



2000

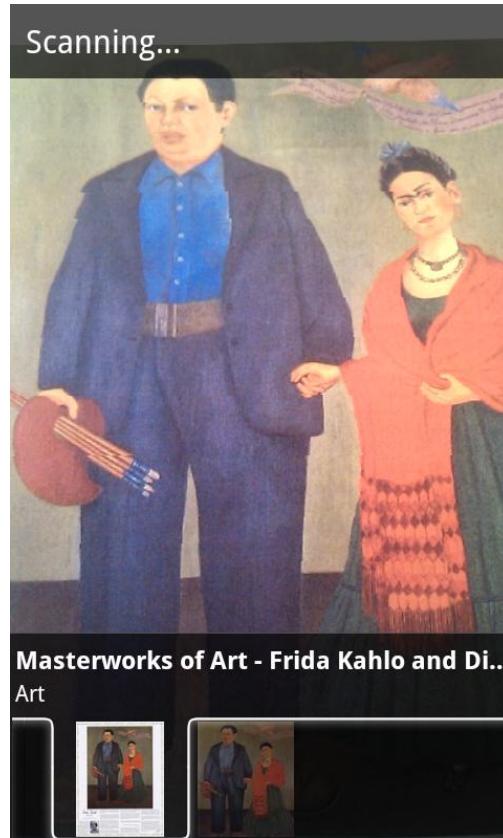
1990

2010

# Image search engines



# Visual search and landmarks recognition



Google Goggles

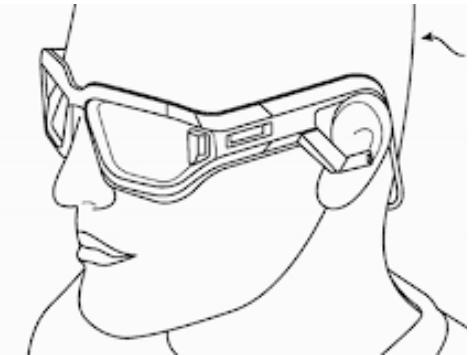
# Visual search and landmarks recognition



# Augmented reality



**Mirriad**  
Advertising for the Skip Generation



- Magic leap
- Daqri
- Meta
- Etc...

# Motion sensing and gesture recognition



# Autonomous navigation and safety

► manufacturer products      consumer products ◀◀

## Our Vision. Your Safety.

rear looking camera      forward looking camera      side looking camera

**EyeQ** Vision on a Chip

News

- Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System
- Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end

> all news

Events

- Mobileye at Equip Auto, Paris, France
- Mobileye at SEMA, Las Vegas, NV

> read more

Mobileye: Vision systems in high-end BMW, GM, Volvo models  
But also, Toyota, Google, Apple, Tesla, Nissan, Ford, etc....

# Personal robotics

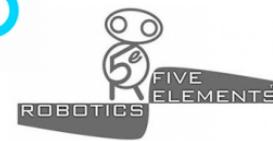


© Robodynamics/SchultzeWORKS/REX



jibo

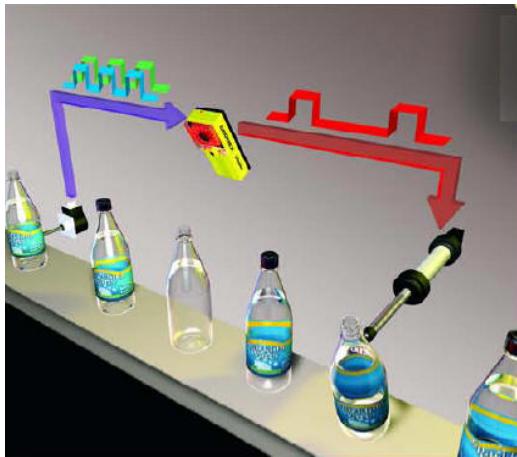
a



ECOVACS

BLUE FROG  
robotics

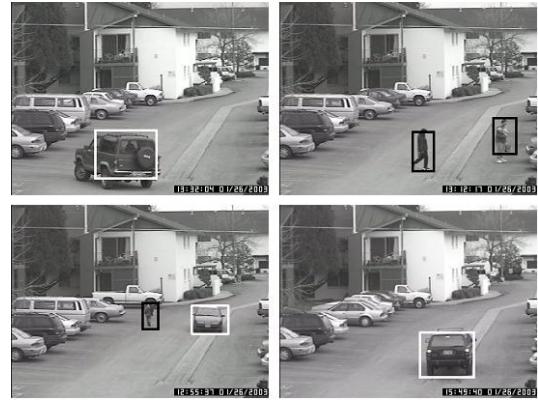
# Computer vision and Applications



Factory inspection



Assistive technologies



Surveillance



Vision for robotics, space exploration



Security

# Computer vision and Applications



EosSystems



Autostich



MOBILEYE



2000



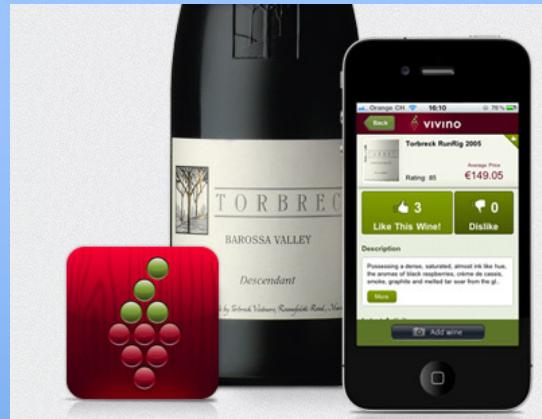
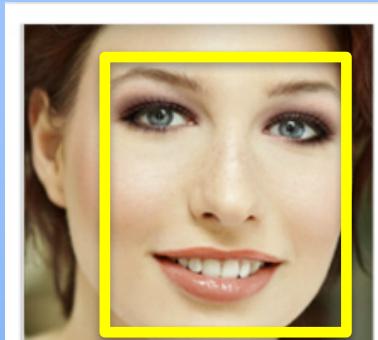
Google  
Goggles



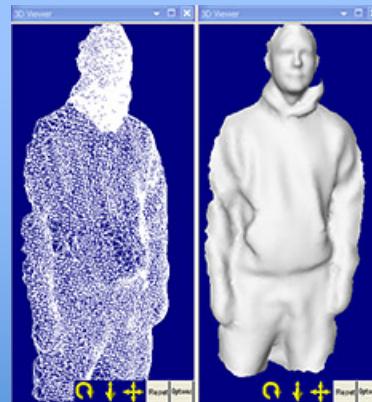
1990

2010

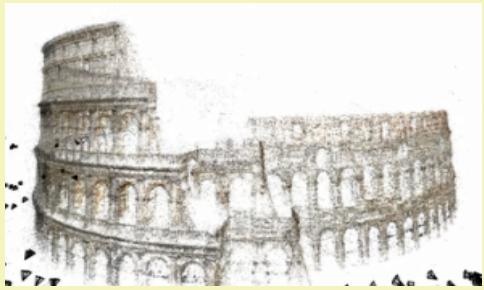
# Computer vision and Applications



# Computer vision and Applications



# Current state of computer vision



## 3D Reconstruction

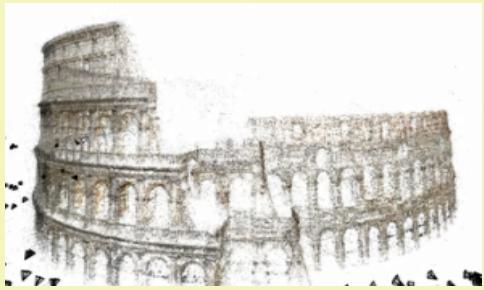
- 3D shape recovery
- 3D scene reconstruction
- Camera localization
- Pose estimation



## 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



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

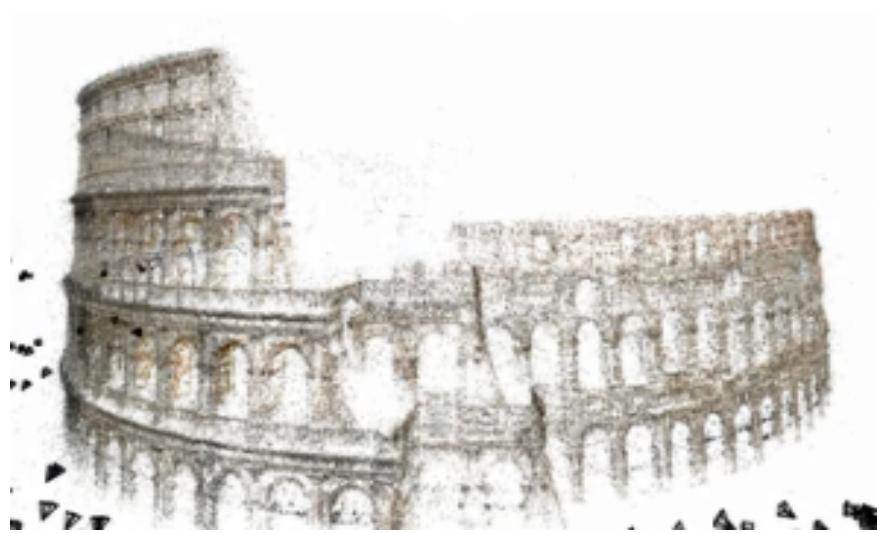
Golparvar-Fard, et al. JAEI 10  
Pandey et al. IFAC , 2010  
Pandey et al. ICRA 2011  
Savarese et al. IJCV 05  
Savarese et al. IJCV 06  
Microsoft's PhotoSynth  
Snavely et al., 06-08  
Schindler et al., 08  
Agarwal et al., 09 45  
Frahm et al., 10

# Current state of computer vision



## 3D Reconstruction

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

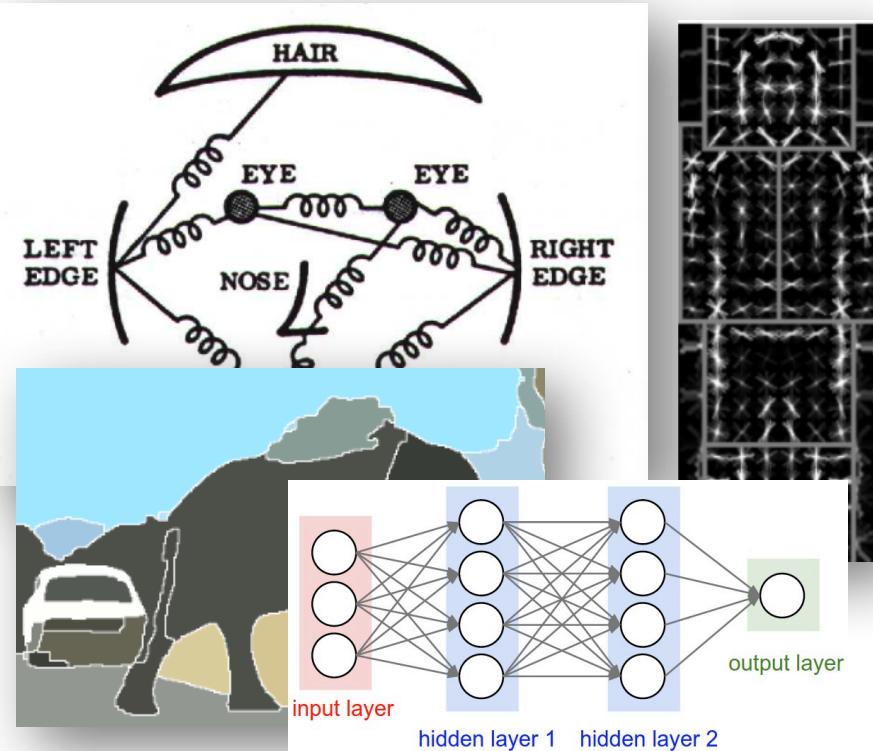


Lucas & Kanade, 81  
Chen & Medioni, 92  
Debevec et al., 96  
Levoy & Hanrahan, 96  
Fitzgibbon & Zisserman, 98  
Triggs et al., 99  
Pollefeys et al., 99  
Kutulakos & Seitz, 99

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

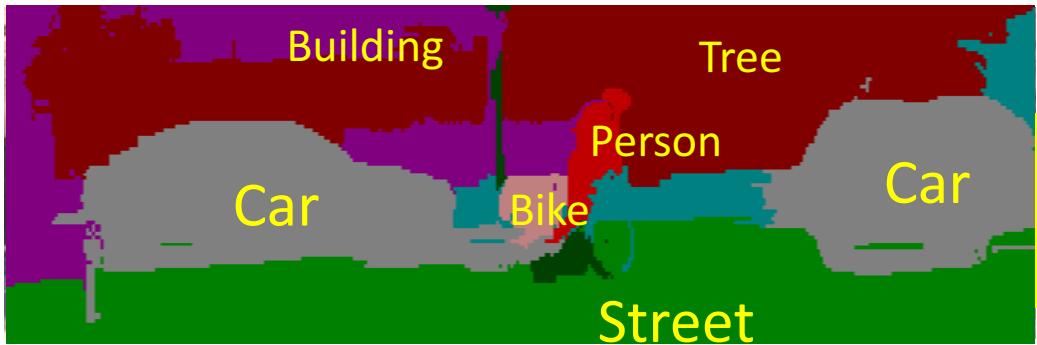
- Object detection
- Texture classification
- Target tracking
- Activity recognition

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

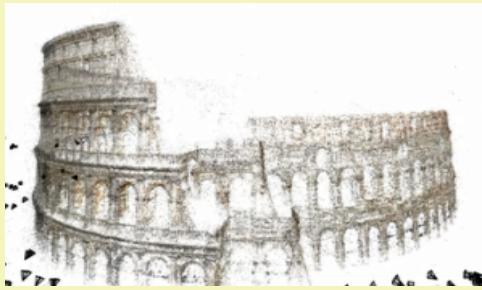
- Object detection
- Texture classification
- Target tracking
- Activity recognition

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

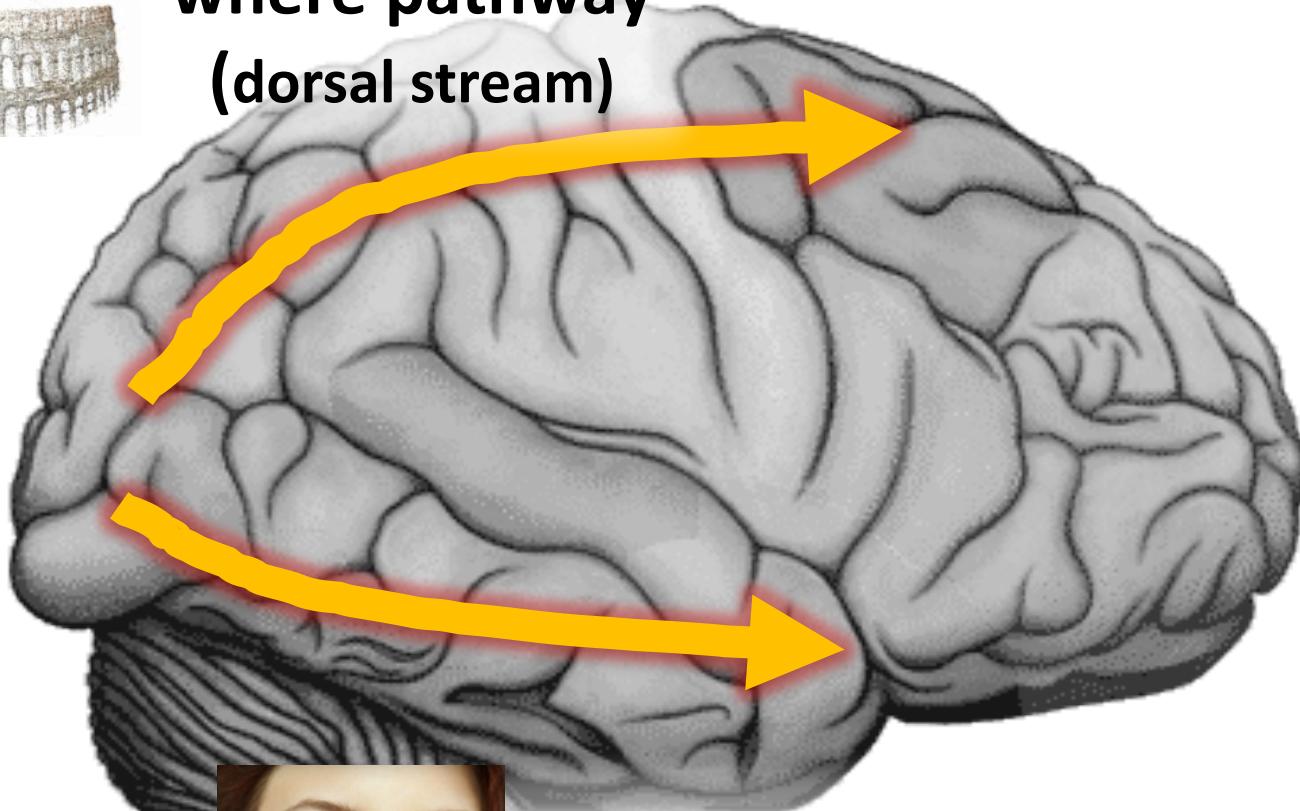
- Object detection
- Texture classification
- Target tracking
- Activity recognition

**Perceiving the World in 3D!**

# Visual processing in the brain



V1

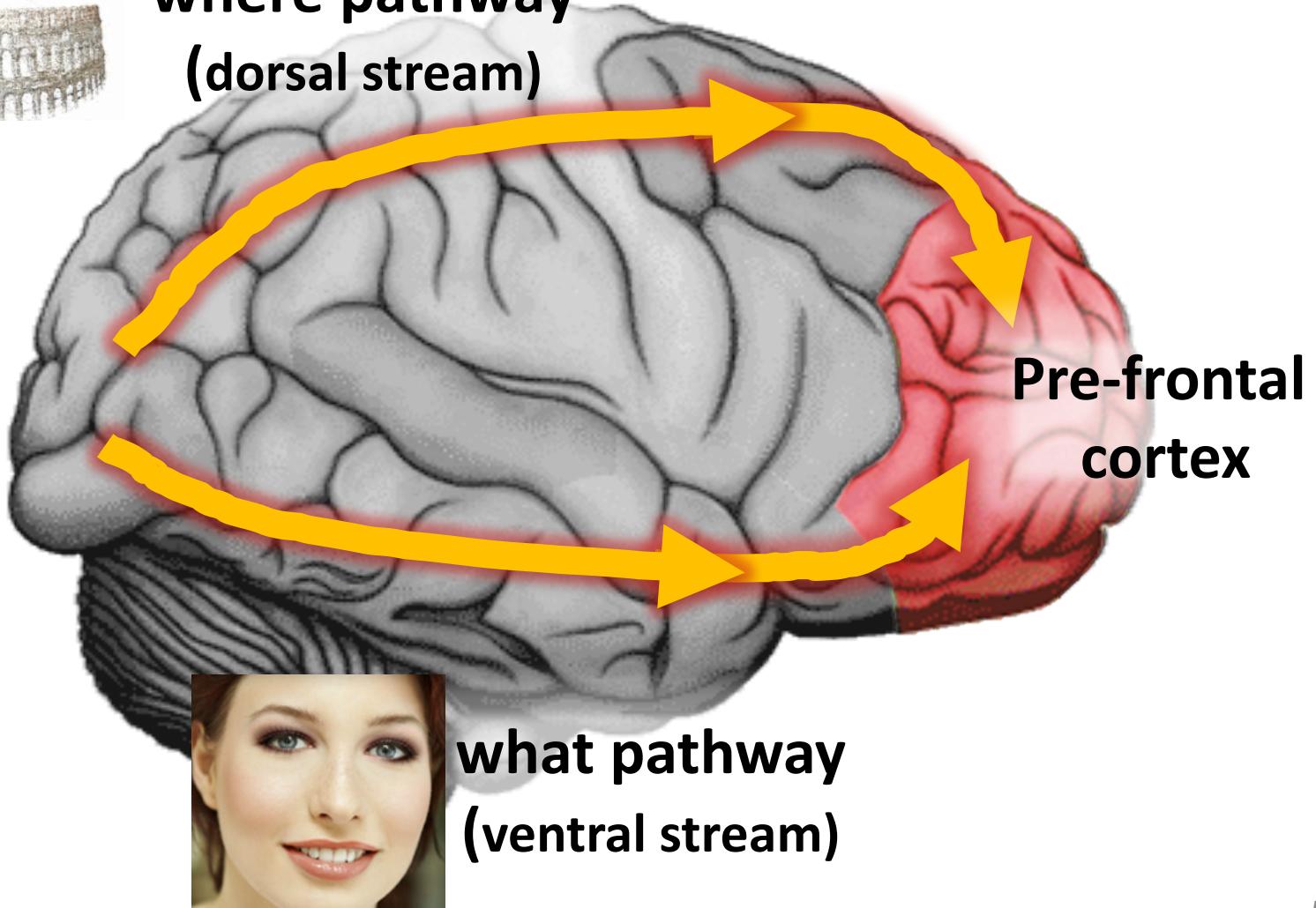


what pathway  
(ventral stream)

# Visual processing in the brain



V1



# 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



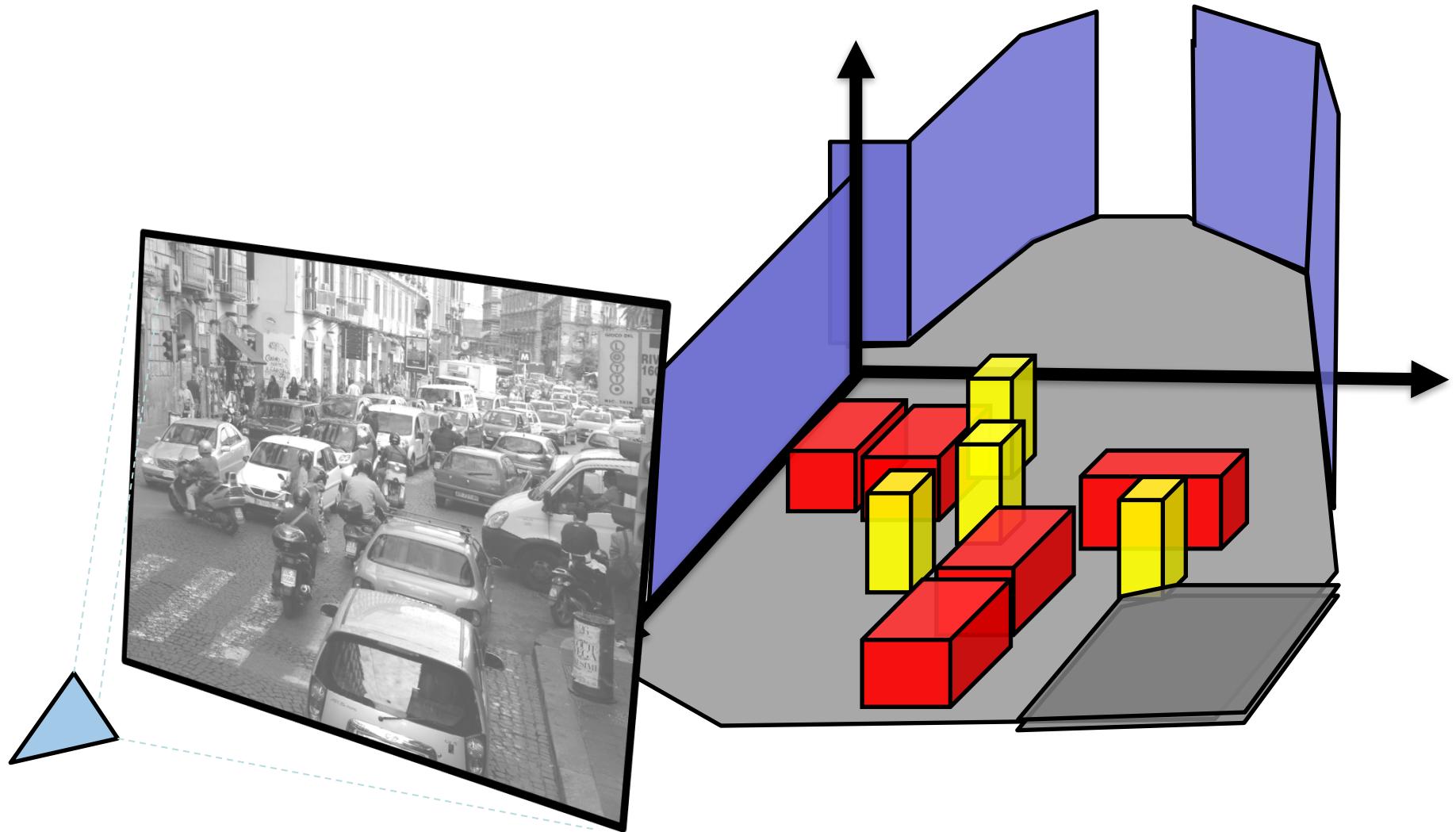
# Single view metrology

Estimate 3D properties of the world from a single image



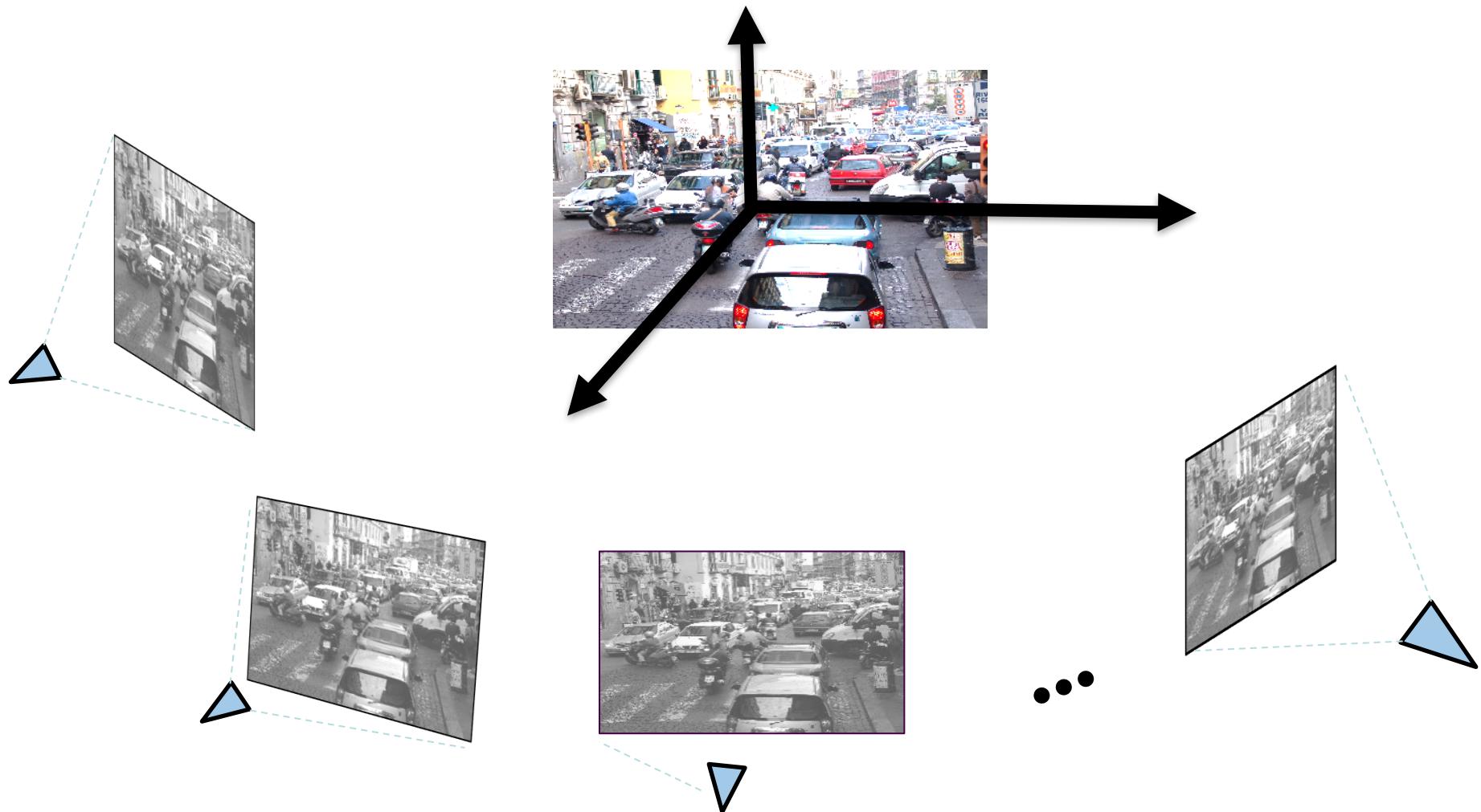
# Single view metrology

Estimate 3D properties of the world from a single image



# Multiple view geometry

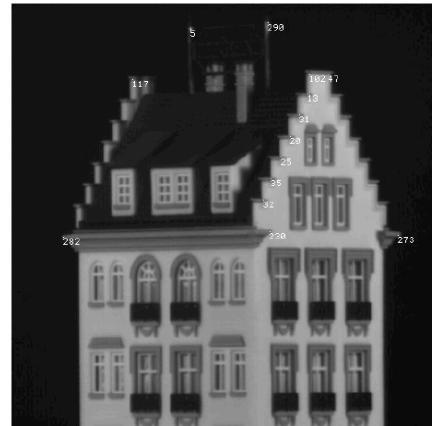
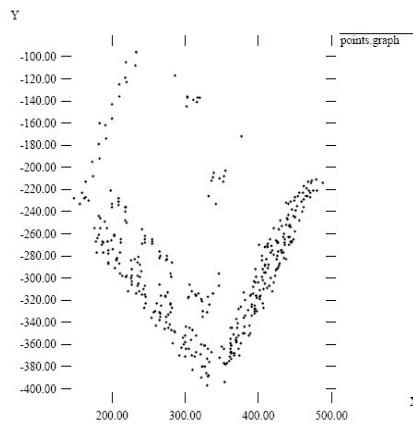
Estimate 3D properties of the world from multiple views



# Mathematical tools



**Epipolar geometry**



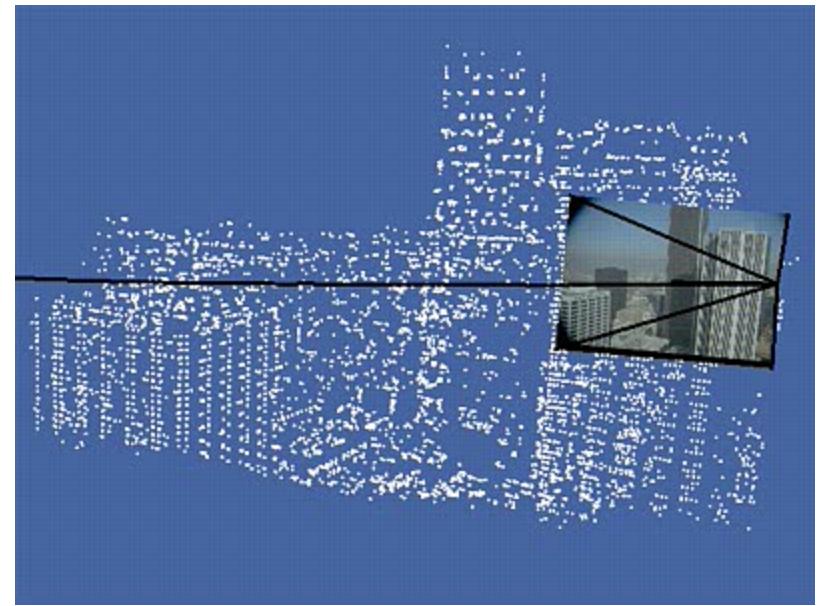
**Tomasi & Kanade (1993)**



Драконъ, видимый подъ различными углами зреинъ  
По гравюре из книги из „Oculus artificialis teledioptricus“ Ганса 1702 года.

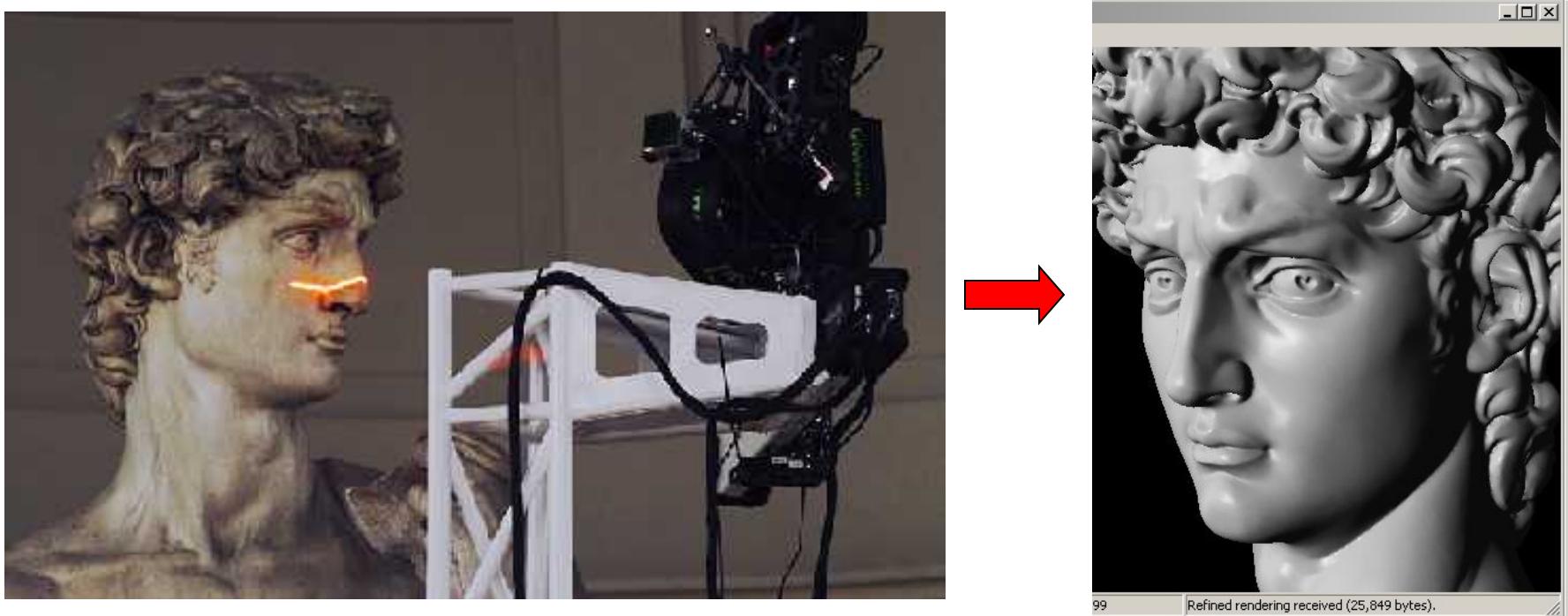
**Photoconsistency**

# Structure from motion



Courtesy of Oxford **Visual Geometry Group**

# Structure lighting and volumetric stereo



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 segment out critical semantics
- How to estimate 3D properties (pose, size, shape...)

# Object recognition and categorization



# Classification:

Is this an forest?



# Classification:

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



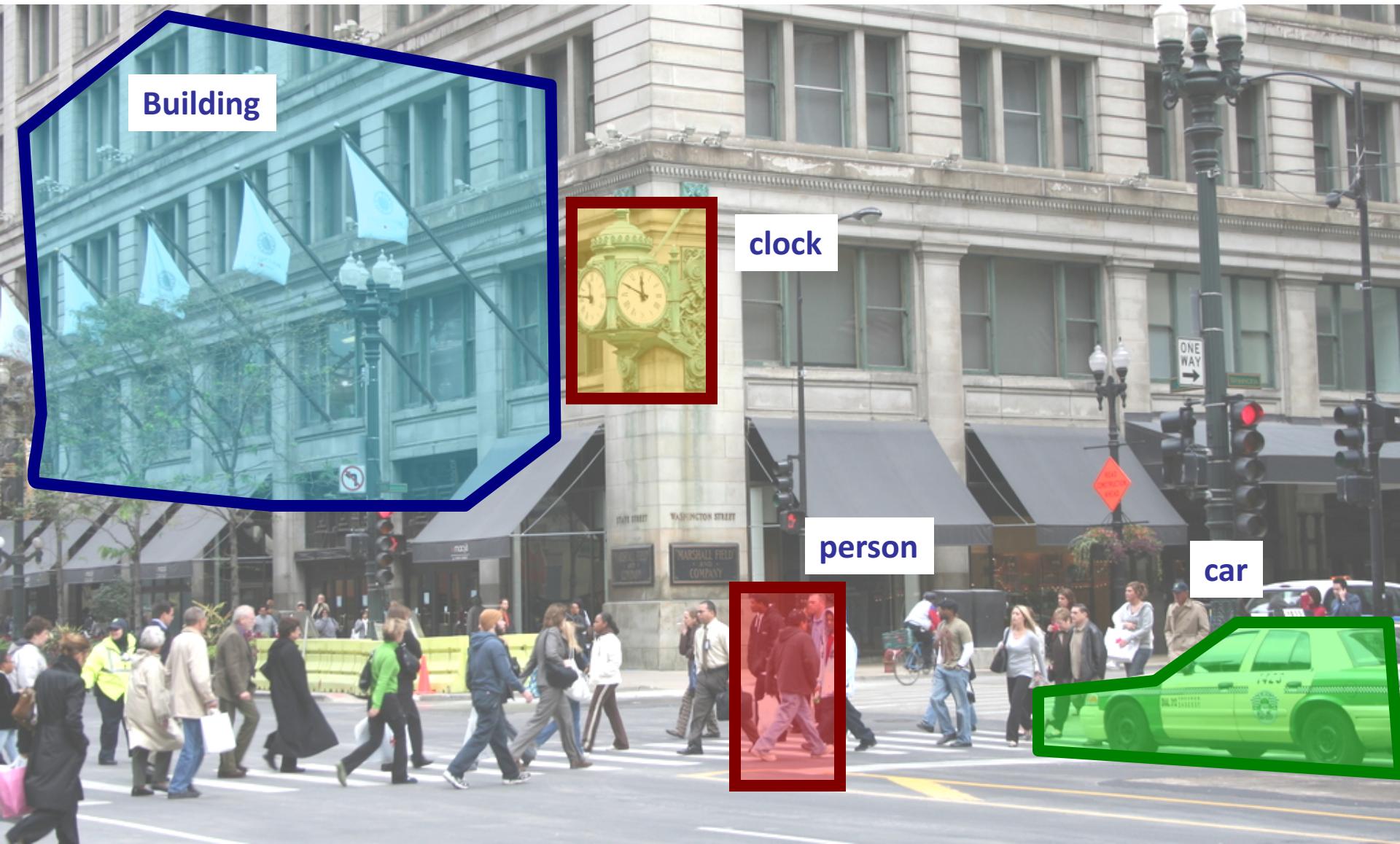
# Detection:

Does this image contain a car? [where?]



# Detection:

Which objects do this image contain? [where?]



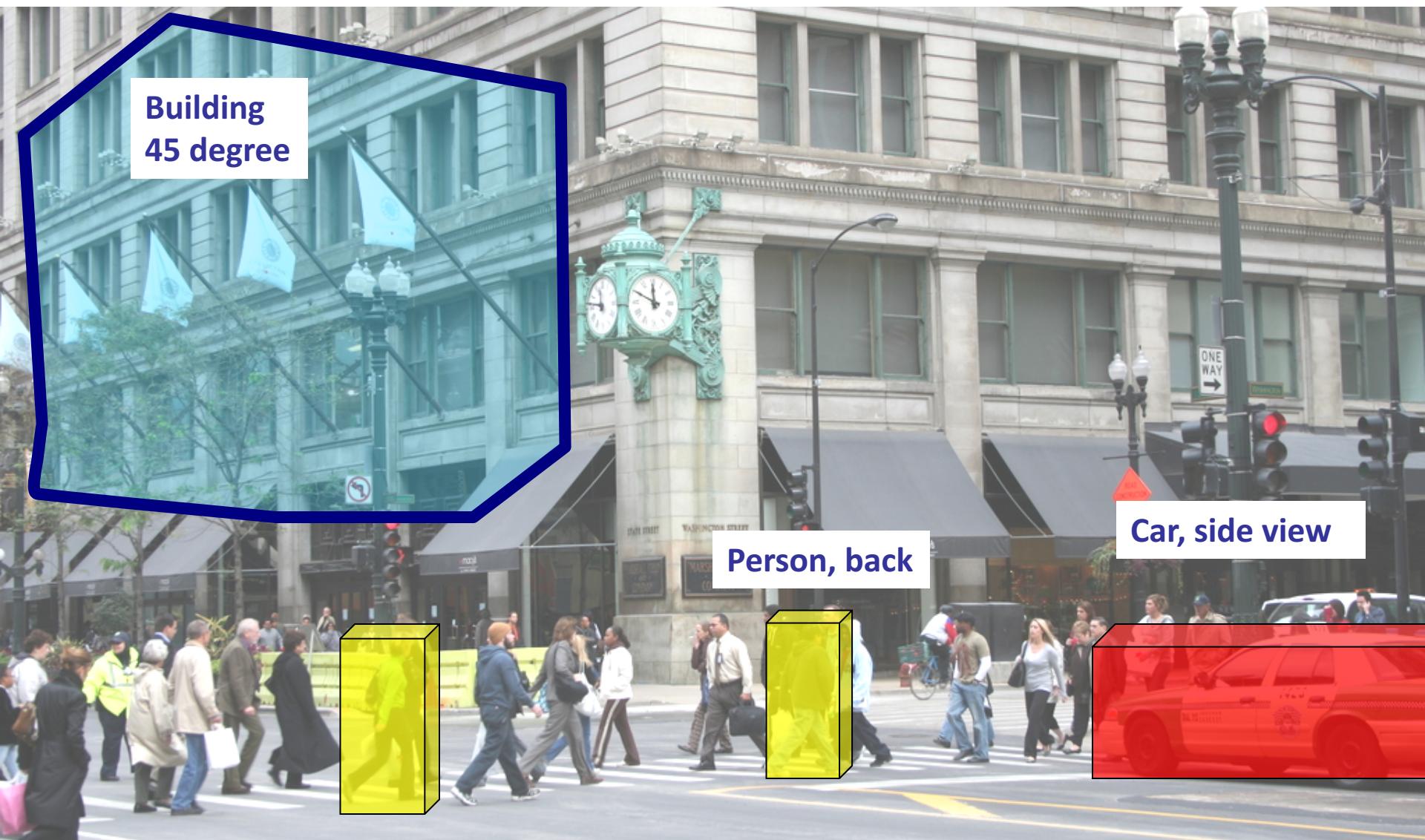
# Detection:

## Accurate localization (segmentation)



# Detection:

## Estimating 3D geometrical properties



# Challenges: viewpoint variation



slide credit: Fei-Fei, Fergus & Torralba

# Challenges: illumination

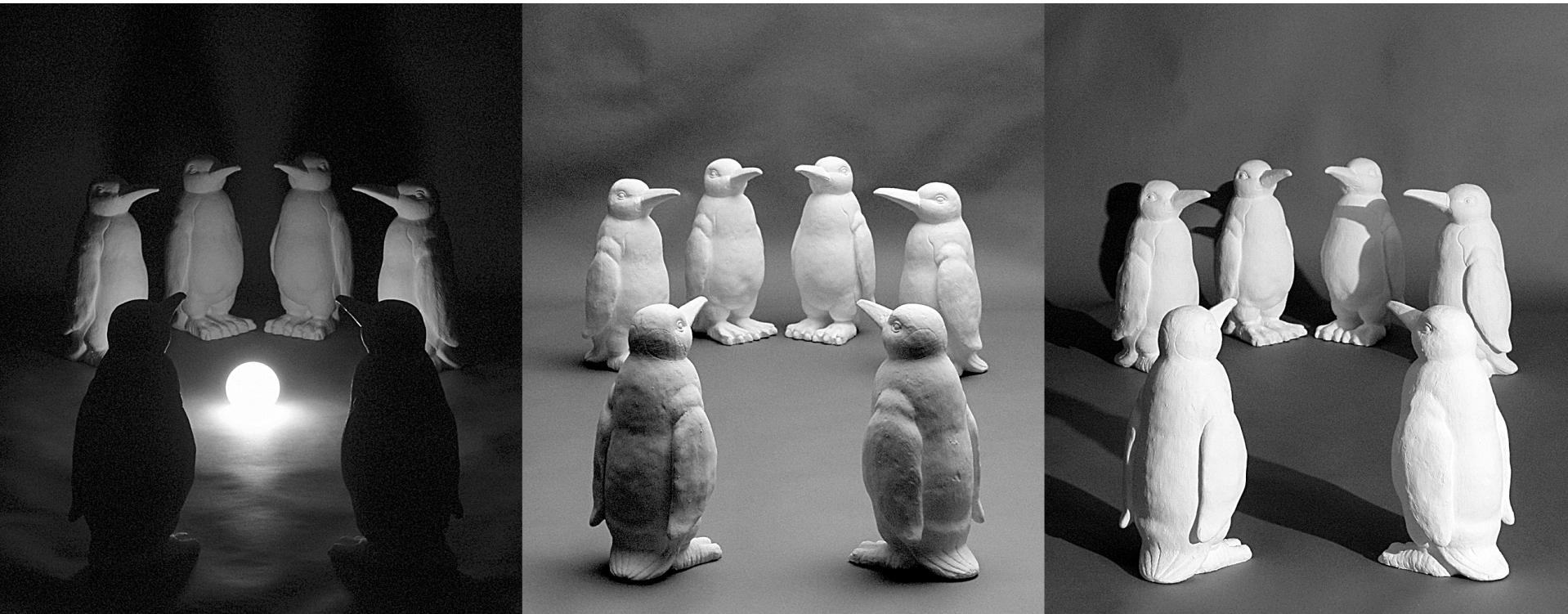


image credit: J. Koenderink

# Challenges: scale



slide credit: Fei-Fei, Fergus & Torralba

# Challenges: deformation



# Challenges: occlusion



Magritte, 1957

slide credit: Fei-Fei, Fergus & Torralba

# Challenges: background clutter



Kilmeny Niland. 1995

# Challenges: object intra-class variation



slide credit: Fei-Fei, Fergus & Torralba



~10,000 to 30,000



# CS 231A course overview

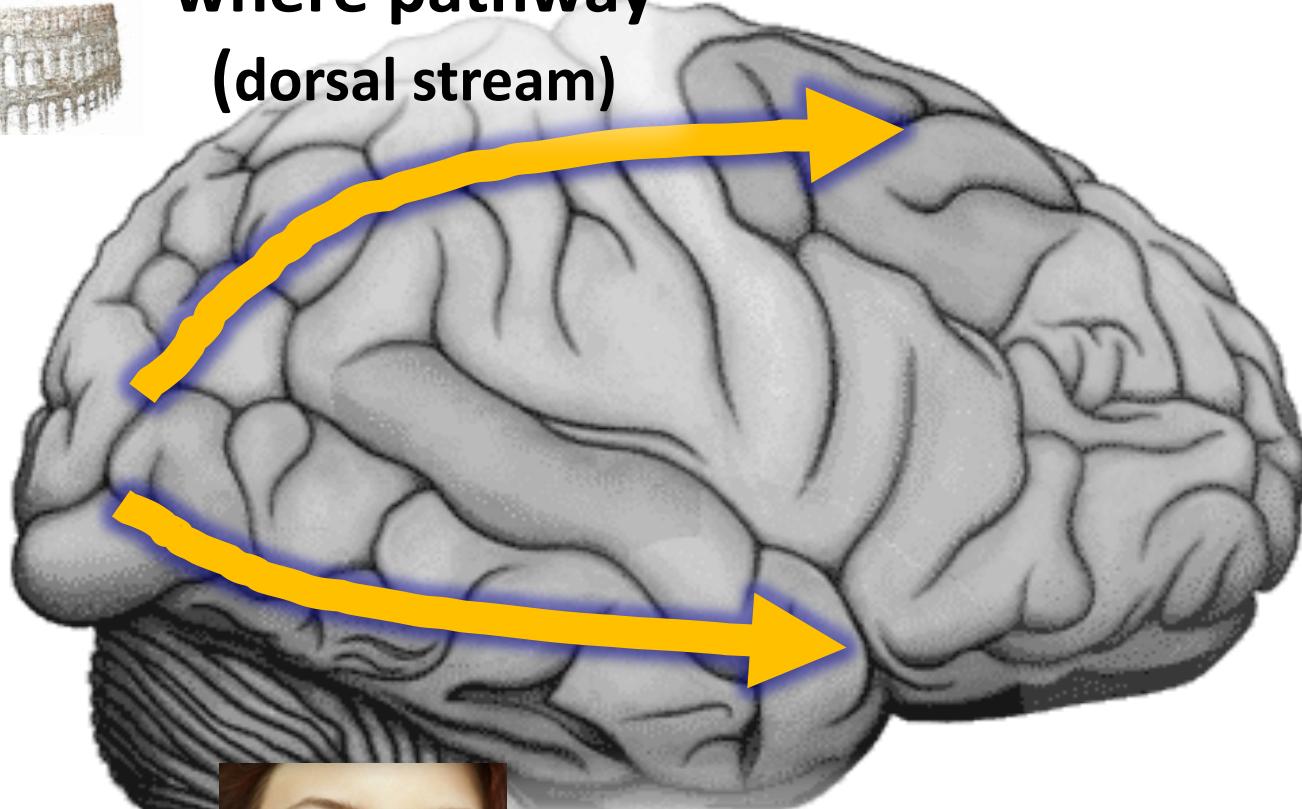
1. Geometry
2. Semantics

Joint recovery of geometry and semantics!

# Visual processing in the brain



V1

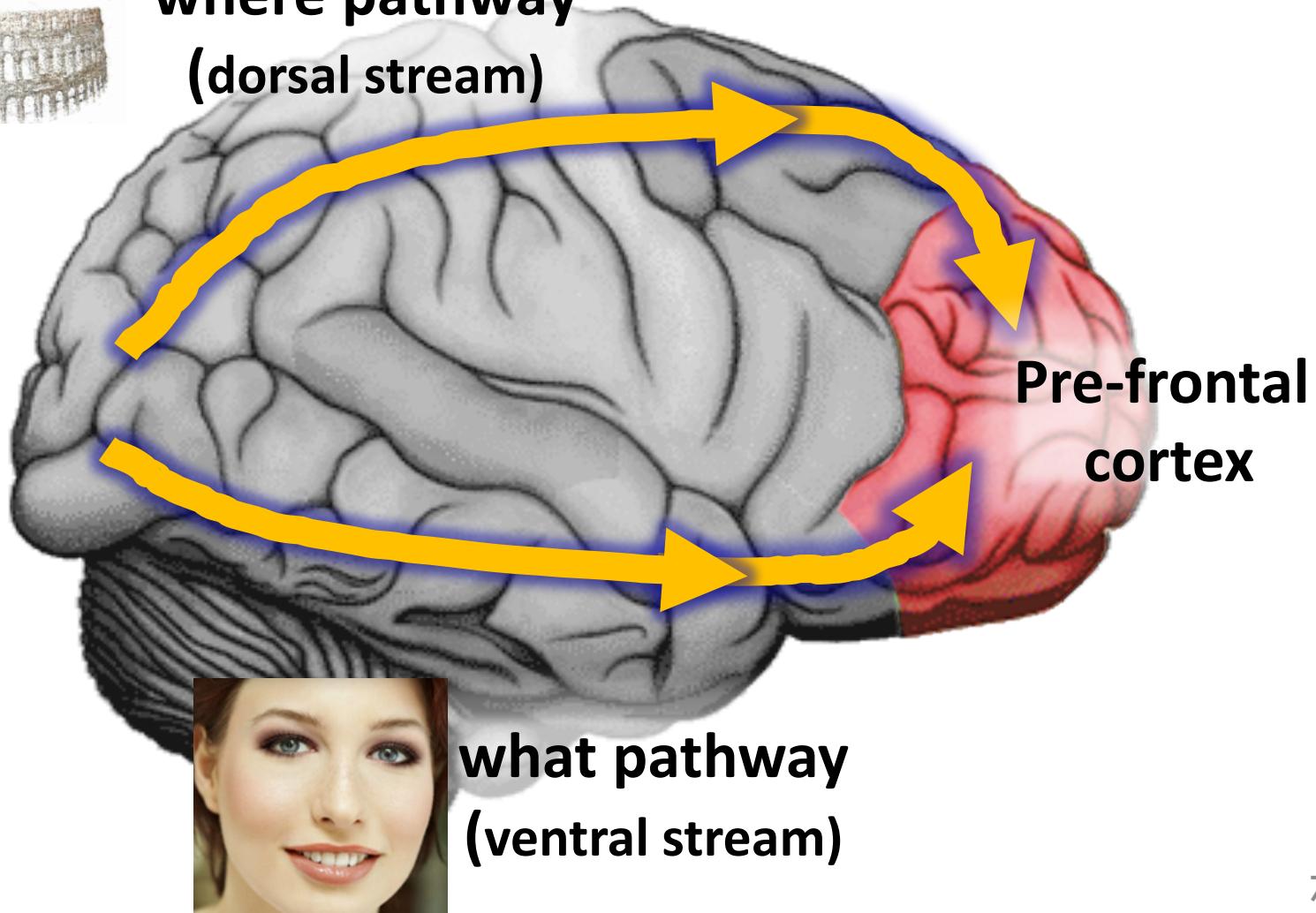


what pathway  
(ventral stream)

# Visual processing in the brain



V1

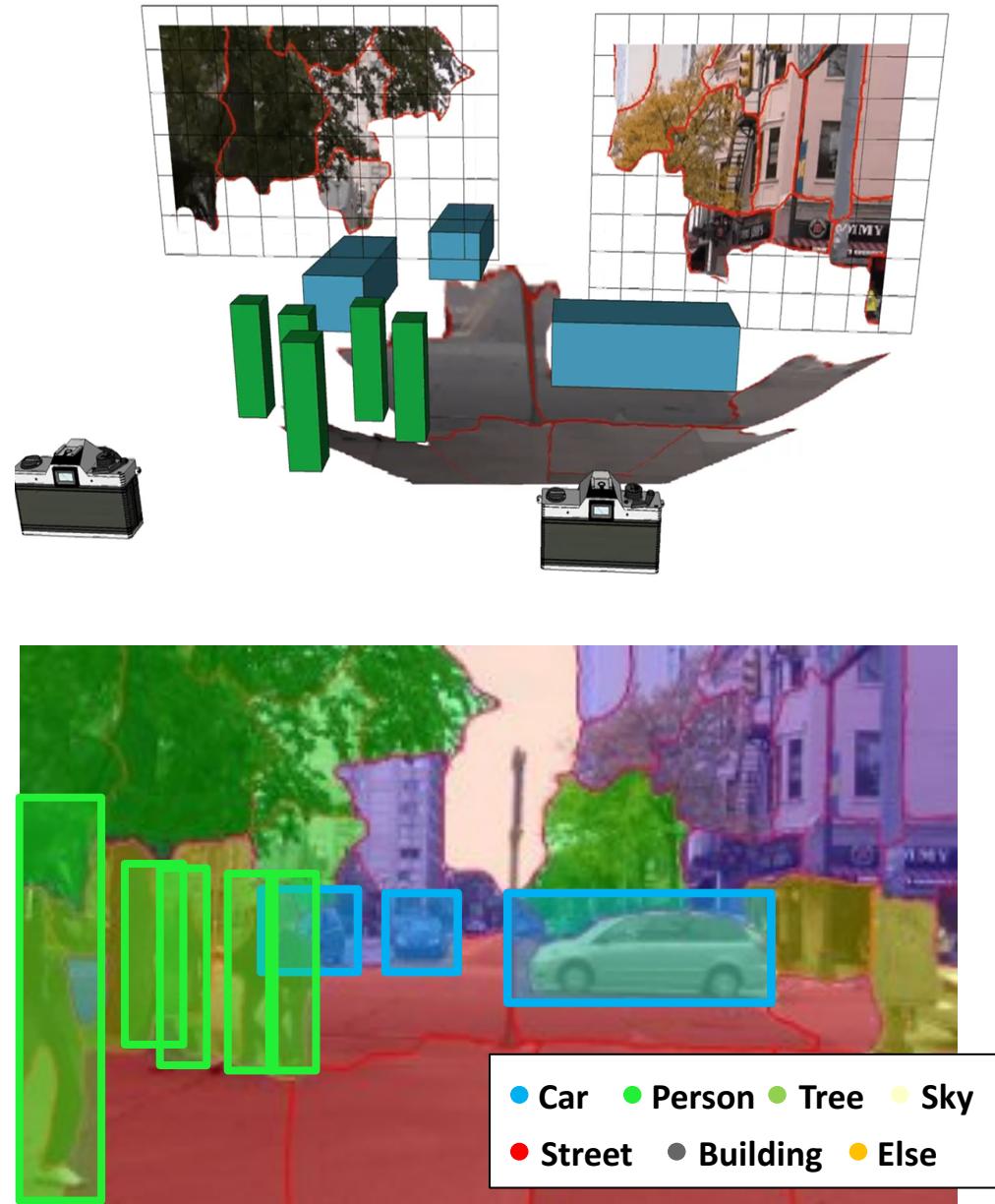


# Joint reconstruction and recognition

Input images



⋮

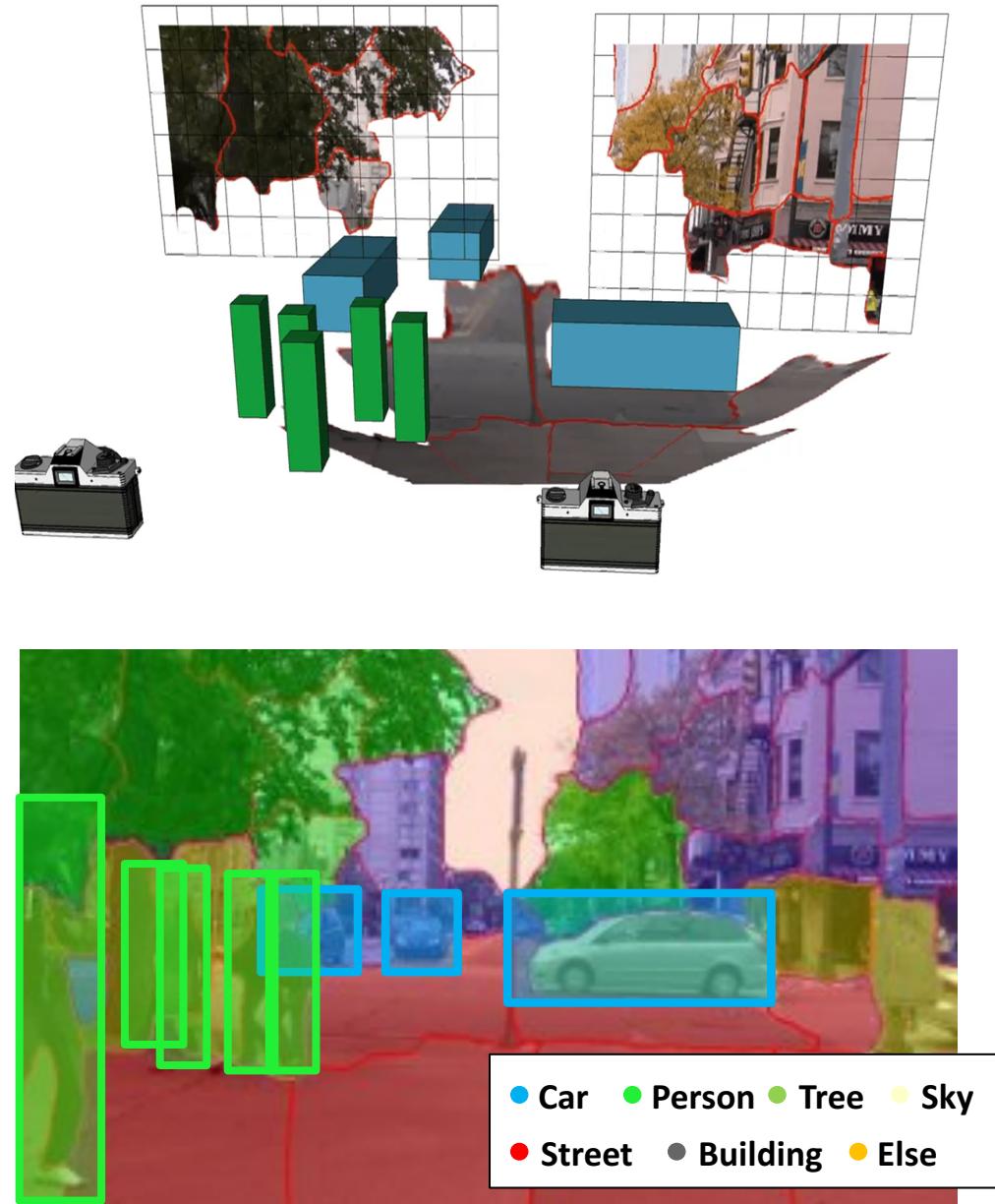
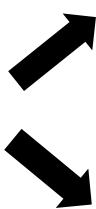


# Joint reconstruction and recognition

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 ..."**

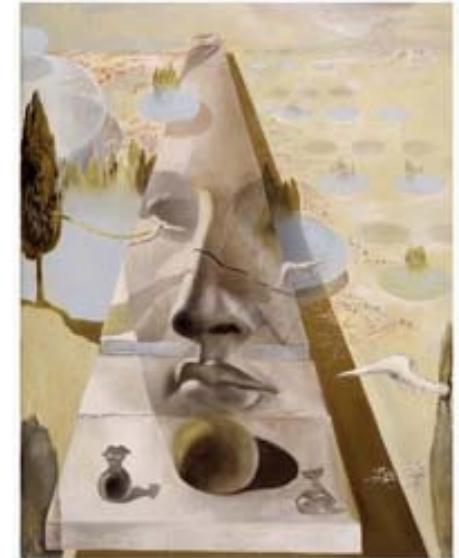
**From "A Mad Tea-Party"  
Alice's Adventures in Wonderland  
by  
Lewis Carroll**

# Syllabus

Lecture	Topic
1	Introduction
2	Camera models
3	Camera calibration
4	Single view metrology
5	Epipolar geometry
6	Multi-view geometry
7	Structure from motion/ SLAM
8	Volumetric stereo
9	Fitting and Matching
10	Detector and Descriptors
11	Intro to Recognition; Object classification I
12	Object classification II
13	Scene understanding & segmentation
14	Visual Representation Learning by NNs
15	3D Object recognition
16	3D Scene understanding

# CS231

## Introduction to Computer Vision



Next lecture: Camera systems