CS4670/5670: Intro to Computer Vision

Instructor: Bharath Hariharan

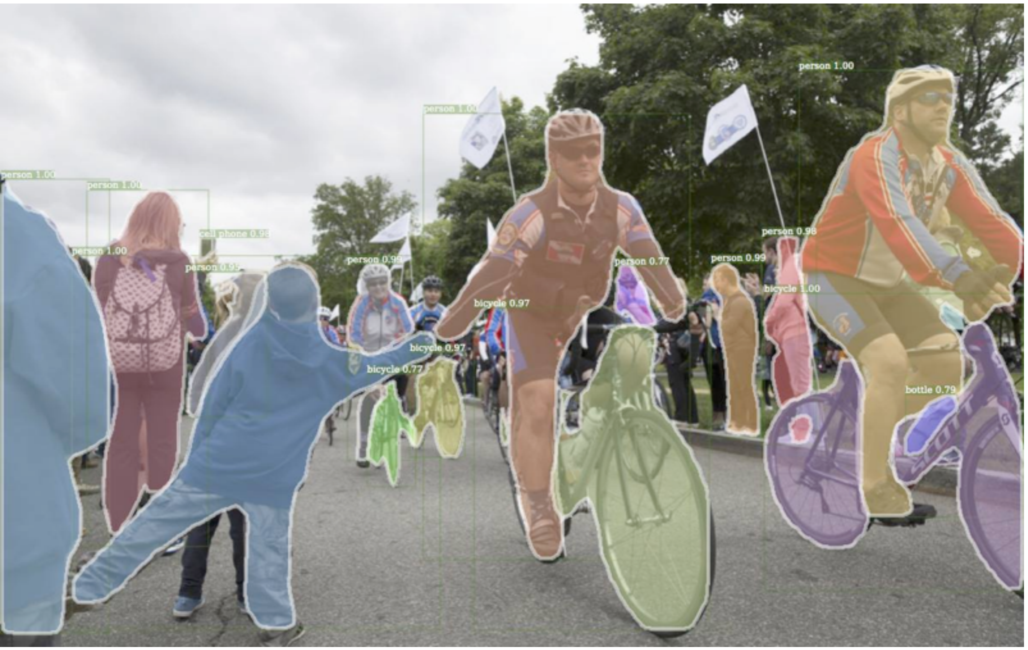


Image credit Ross Girshick

Image credit Ira Kemelmacher-Shlizerman

Instructor

- Bharath Hariharan (bharathh@cs.cornell.edu)
- Office hours:

  M/W/Thur: 10:00-11:00 am, or by appointment
- Research interests:
  - Computer vision
  - Machine learning

Today

1. What is computer vision?

2. Course overview

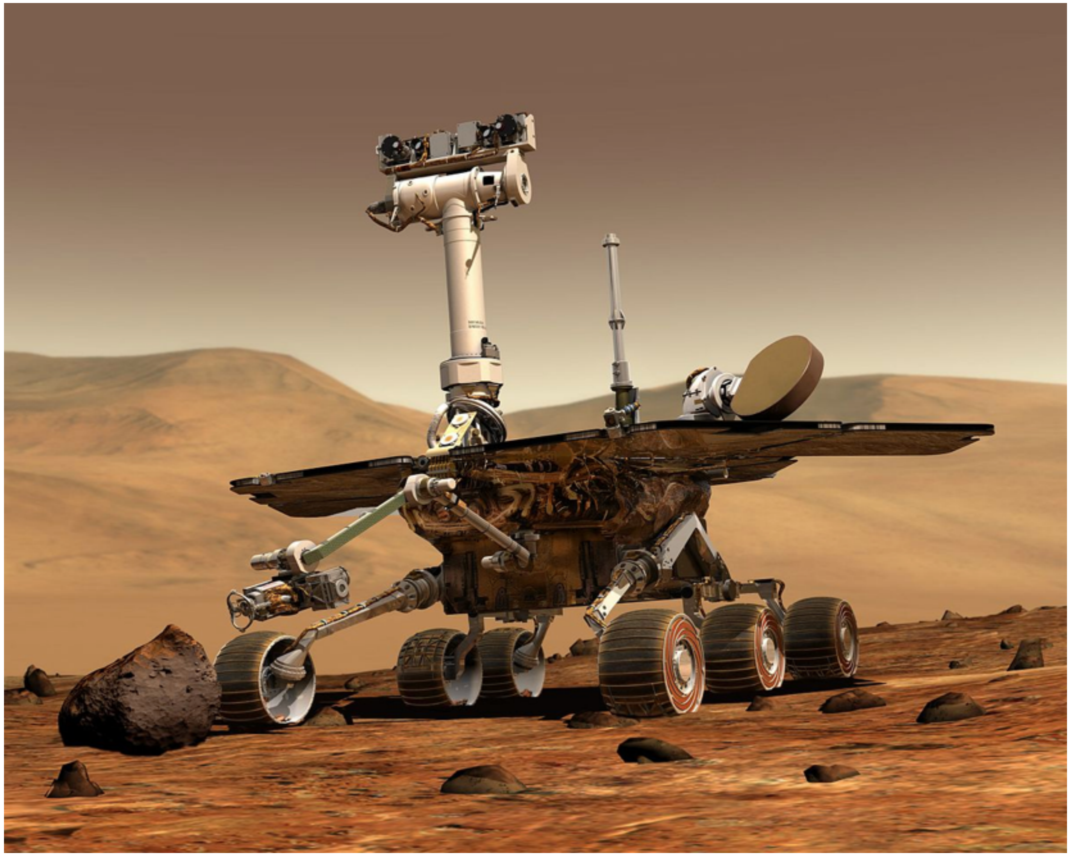
Today

- Readings
  - Szeliski, Chapter 1 (Introduction)

Every image tells a story

- Goal of computer vision: perceive the "story" behind the picture
- But what does "story" mean?
- Depends on what we want to do with it!

Why Computer Vision?
Example 1: Robotics



Why Computer Vision?
Example 1: Robotics

http://theoatmeal.com/blog/google_self_driving_car

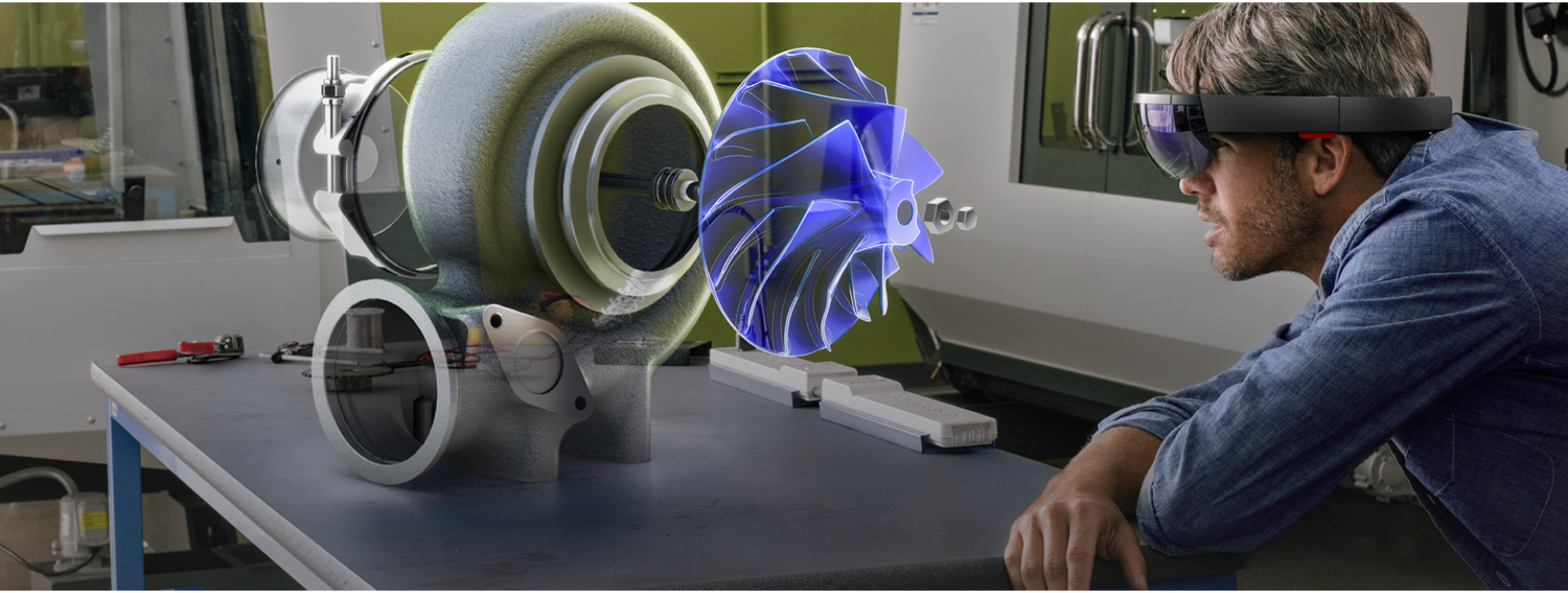
Why Computer Vision?
Example 2: Internet Vision

Facebook Users Are Uploading 350 Million New Photos Each Day

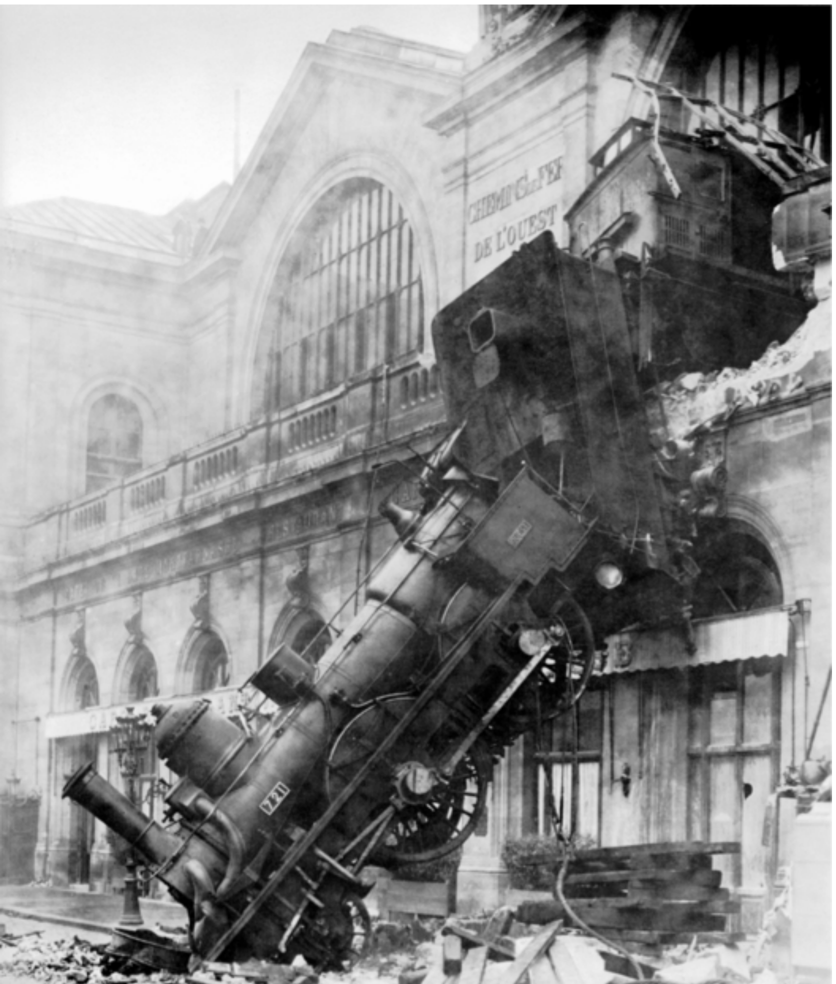
Cooper Smith

Sep. 18, 2013, 8:00 AM 23,351

Why Computer Vision?
Example 3: AR



The goal(s) or computer vision



- What is the image about?
- What objects are in the image?
- Where are they?
- How are they oriented?
- What is the layout of the scene in 3D?
- What is the shape of each object?

Vision is easy for humans

Vision is easy for humans

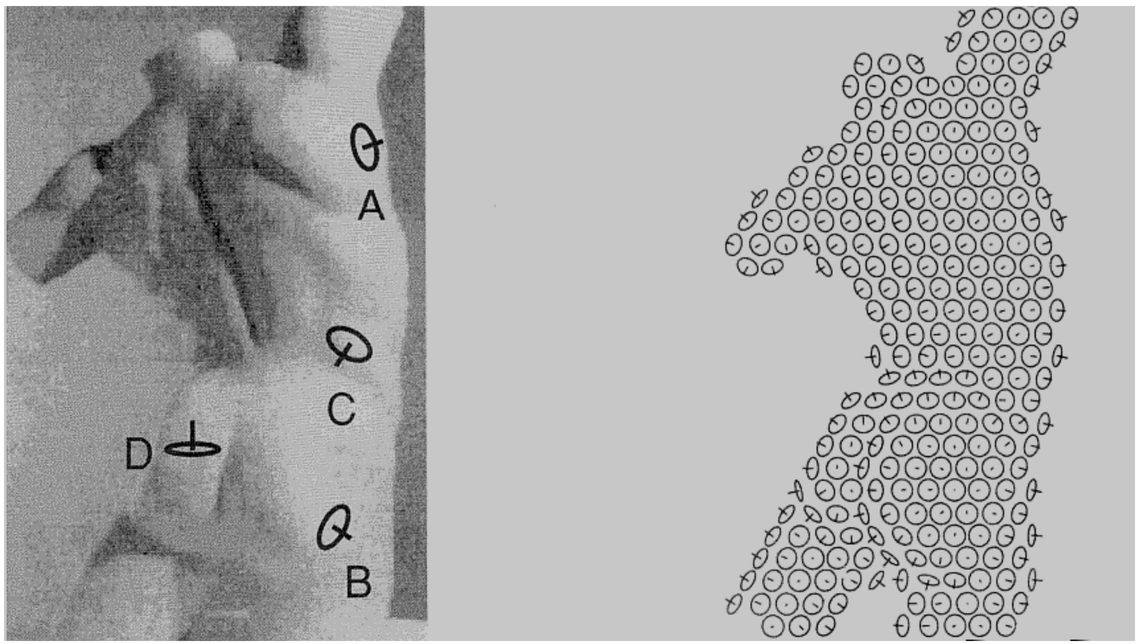
Vision is easy for humans

- Attneave's cat

Vision is easy for humans

- Mooney faces

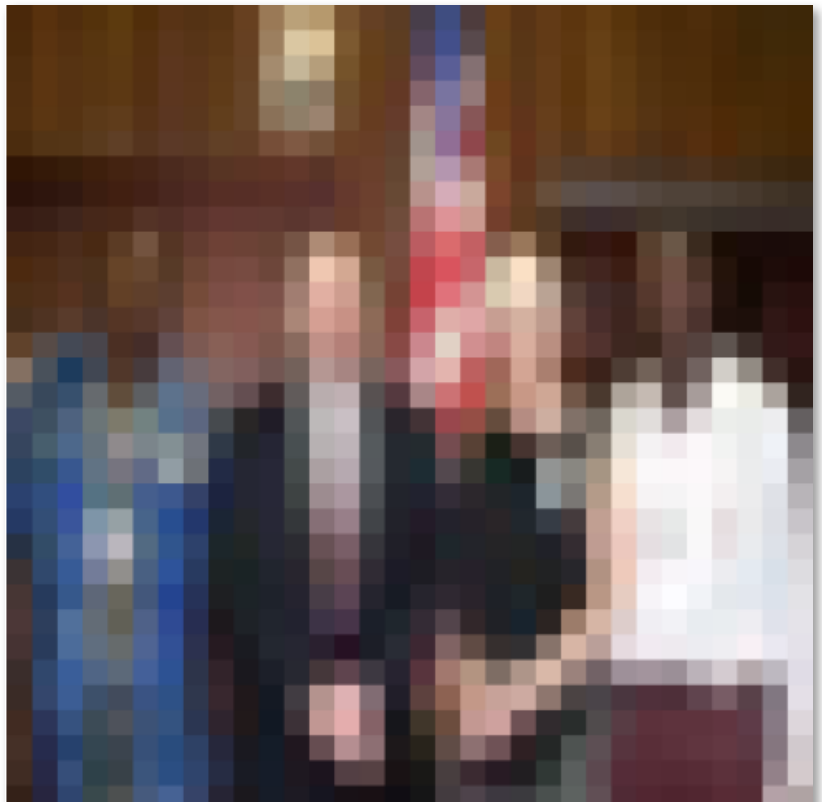
Vision is easy for humans



Surface perception in pictures. Koenderink, van Doorn and Kappers, 1992

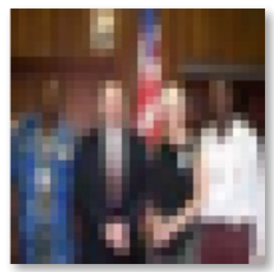
Slide credit: Jitendra Malik

Vision is easy for humans



Source: “80 million tiny images” by Torralba, et al.

Vision is easy for humans



Source: “80 million tiny images” by Torralba, et al.

…but not always

[Sinha and Poggio, *Nature*, 1996]

Vision is hard: Images are ambiguous

Vision is hard: Objects blend together

Vision is hard: Objects blend together

Vision is hard: Concepts have variance

The many faces of intra-class variance

Viewpoint variation

Illumination

Scale

The many faces of intra-class variance

Shape variation

Occlusion

Background clutter

The many faces of intra-class variation

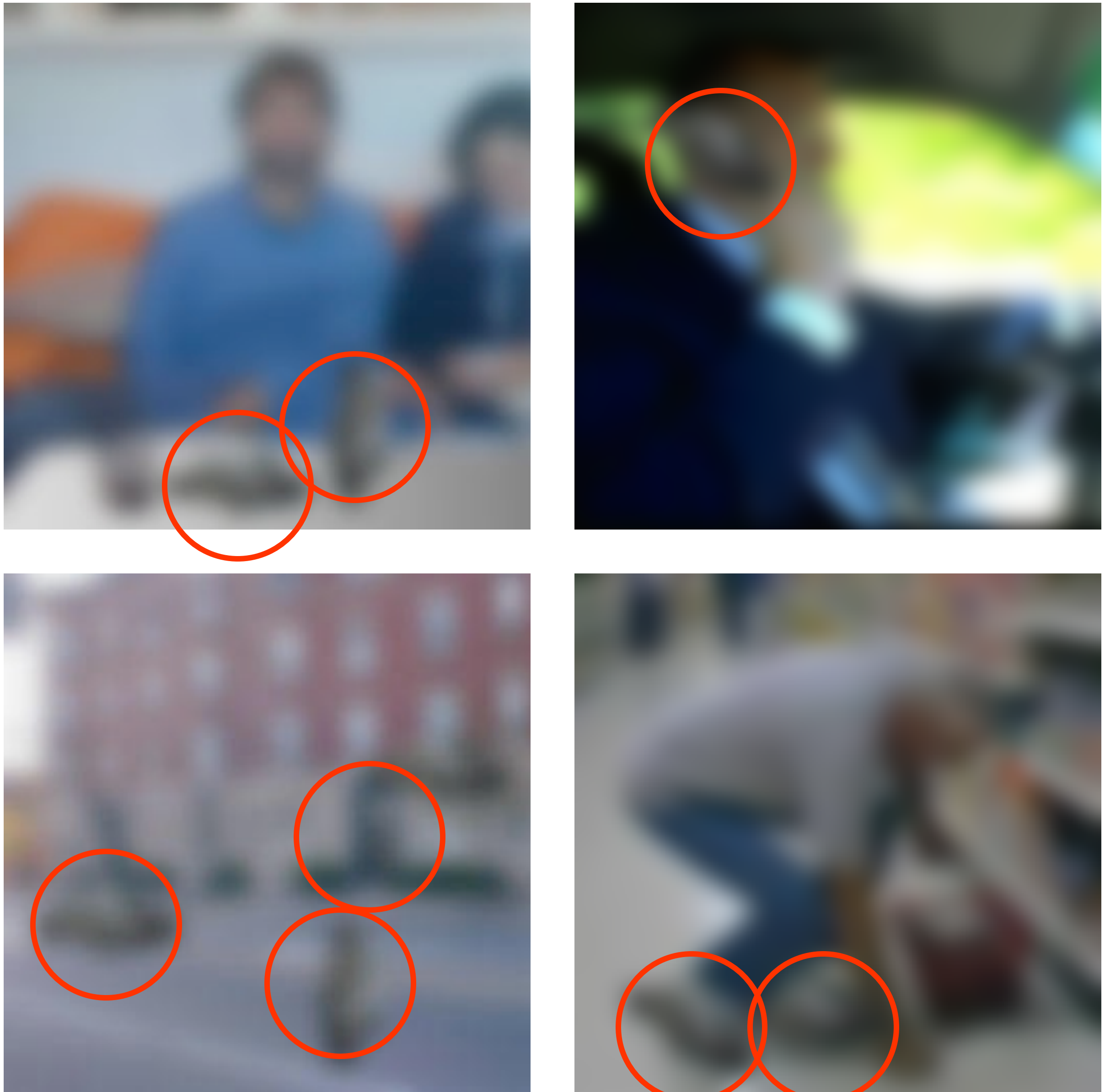
Vision is hard: Concepts are subtle

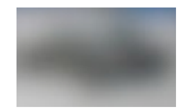
Tenessee Warbler

Orange Crowned Warbler

https://www.allaboutbirds.org

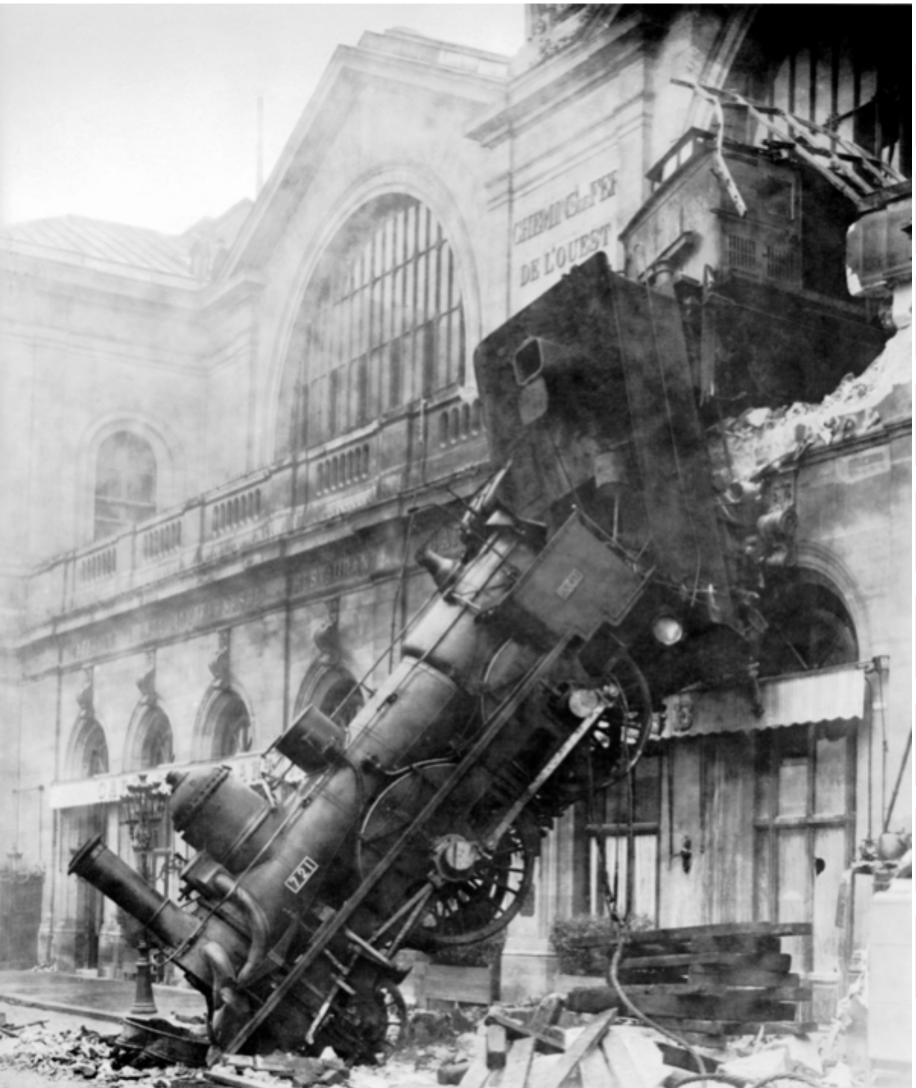
Vision is hard: local ambiguity





slide credit: Fei-Fei, Fergus & Torralba

What the input looks like



0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

The "summer vision project"

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system.

The "summer vision project"

Goals - General

The primary goal of the project is to construct a system of programs which will divide a vidisector picture into regions such as

likely objects

likely background areas

chaos.

We shall call this part of its operation FIGURE-GROUND analysis.

It will be impossible to do this without considerable analysis of shape and surface properties, so FIGURE-GROUND analysis is really insepar-able in practice from the second goal which is REGION DESCRIPTION.

The final goal is OBJECT IDENTIFICATION which will actually name objects by matching them with a vocabulary of known objects.

The big reveal

"… in the 1960s almost no one realized that machine vision was difficult… The common and almost despairing feeling of the early investigators like B.K.P. Horn and T.O. Binford was that practically anything could happen in an image and furthermore that practically everything did."
--- Marr, 1982

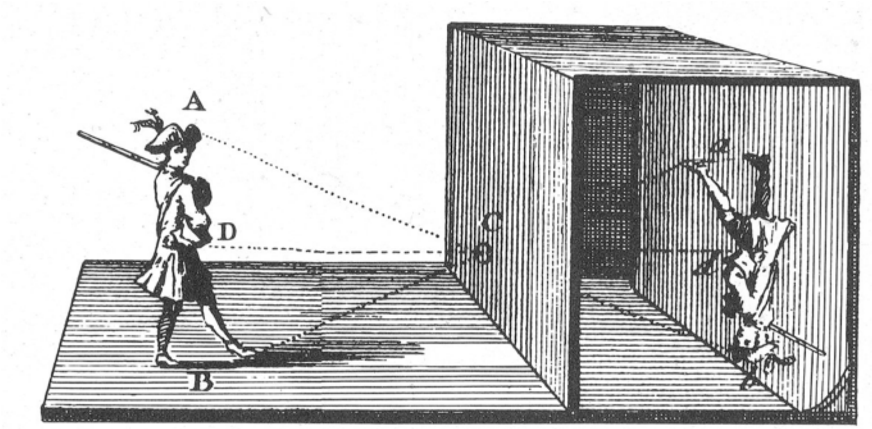
Perception is the big problem

Our first foray into Artificial Intelligence was a program that did a credible job of solving problems in college calculus. Armed with that success, we tackled high school algebra; we found, to our surprise, that it was much harder. Attempts at grade school arithmetic provide problems of current research interest. An exploration of the child's world of blocks proved insurmountable, except under the most rigidly constrained circumstances.

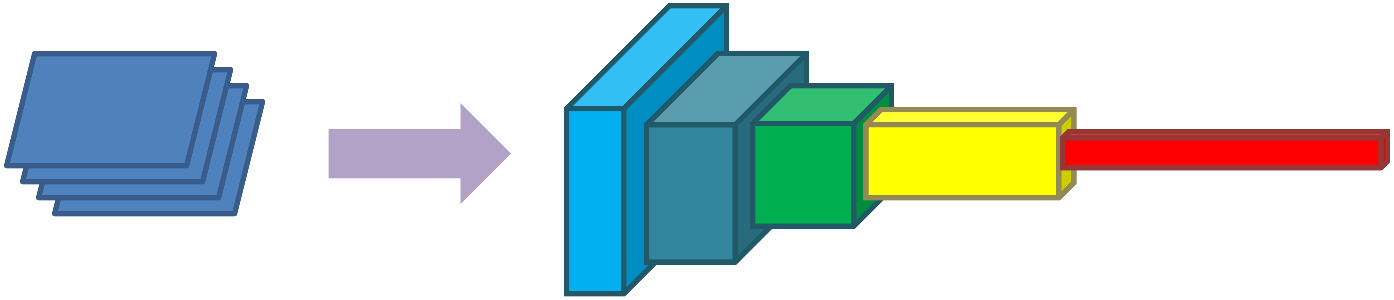
It finally dawned on us that the overwhelming majority of what we call intelligence is developed by the end of the first year of life."

– Marvin Minsky, 1977

Cues to help us



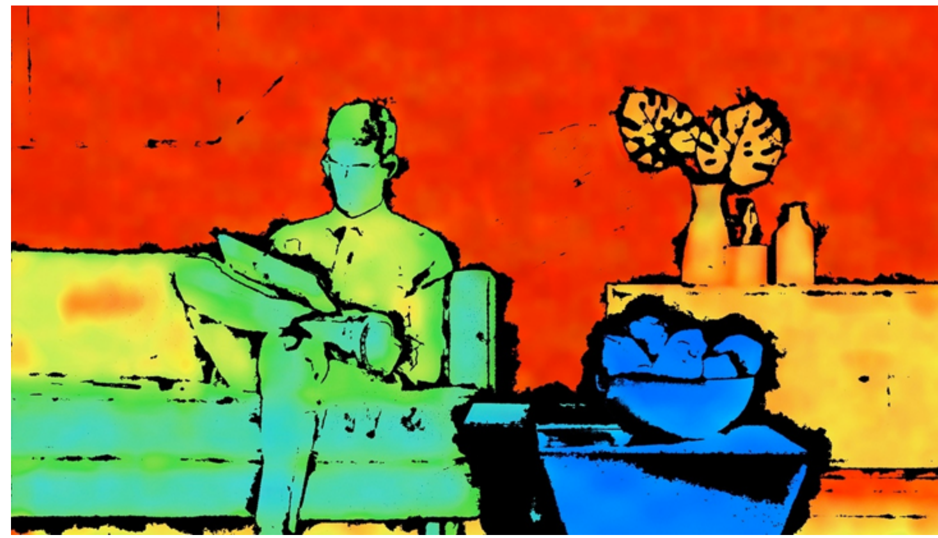
The physics of image formation



Statistics and machine learning

WHERE ARE WE NOW?

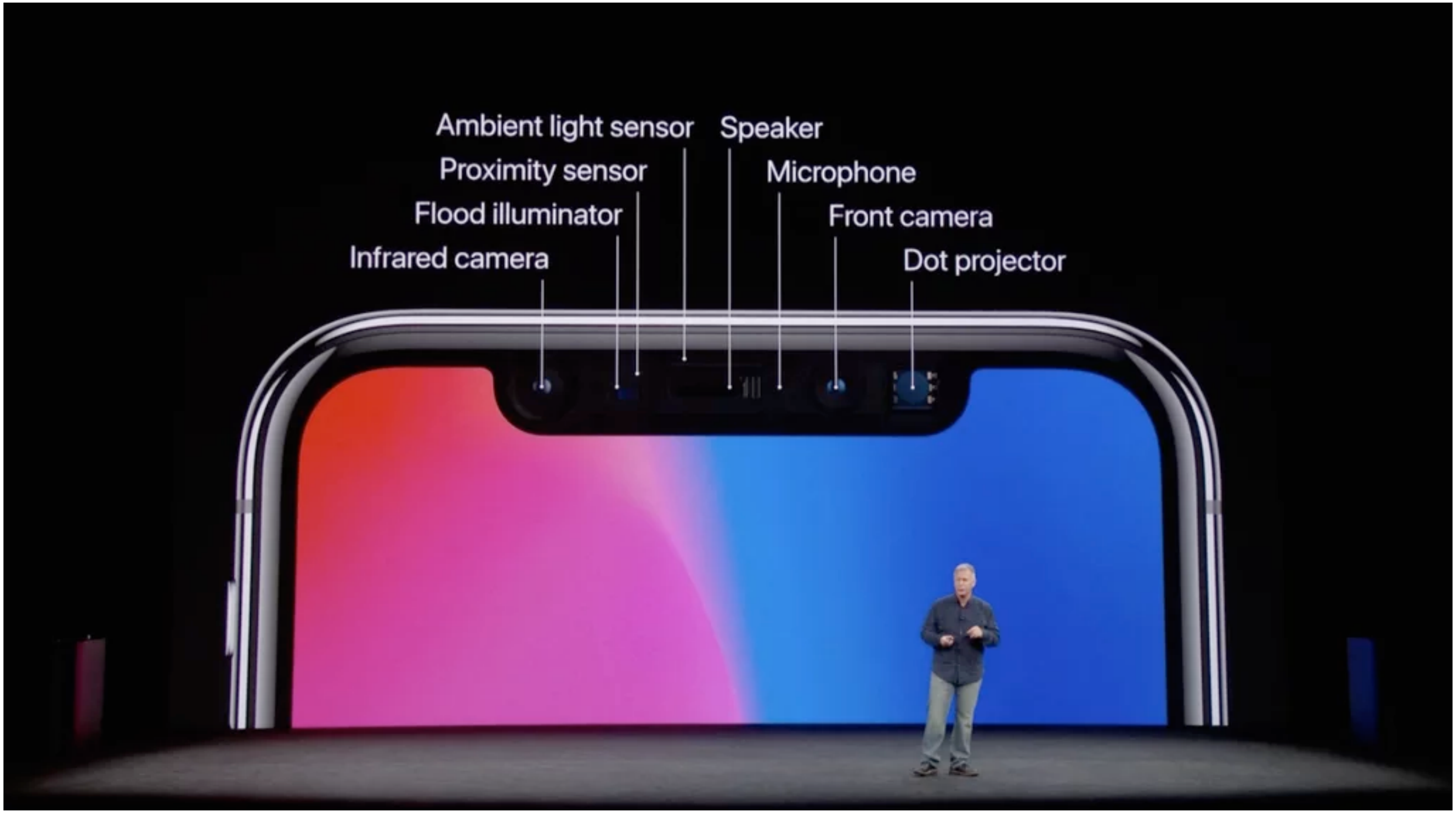
Deployed: depth cameras



https://realsense.intel.com/stereo/

Microsoft Kinect

Deployed: depth cameras



Iphone TrueDepth

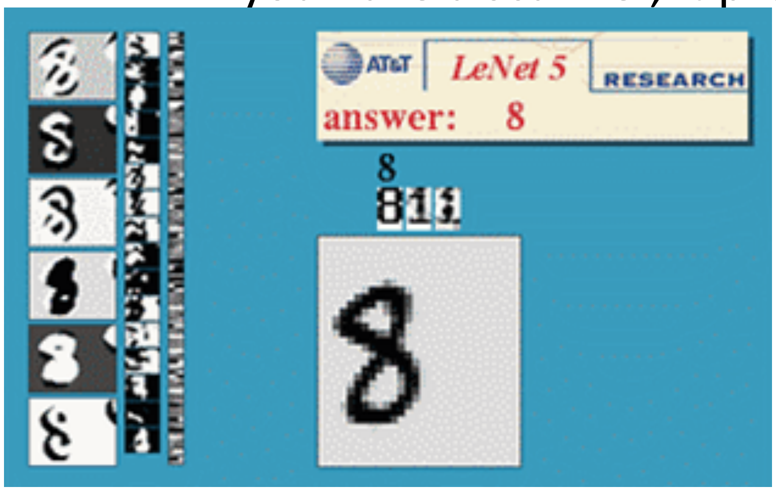
Deployed: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Source: S. Seitz

Deployed: Optical character recognition (OCR)

- If you have a scanner, it probably came with OCR software



Digit recognition, AT&T labs
http://www.research.att.com/~yann/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Automatic check processing

Source: S. Seitz

Deployed: Face detection

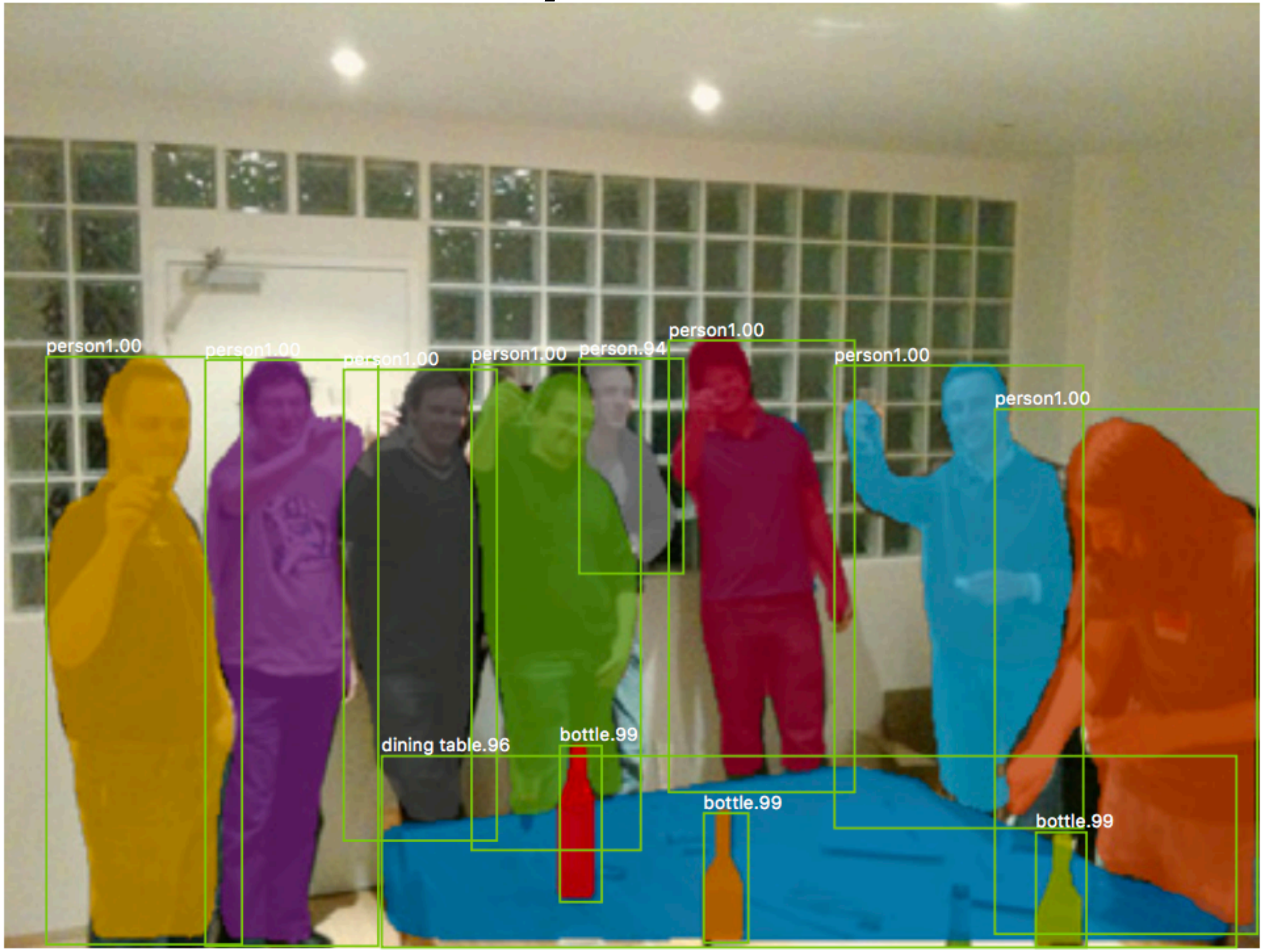
- Cameras now detect faces
  - Canon, Sony, Fuji, …

Source: S. Seitz

Established technology: 3D Models of the world

Building Rome in a Day.
Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski.
ICCV, 2009, Kyoto, Japan.

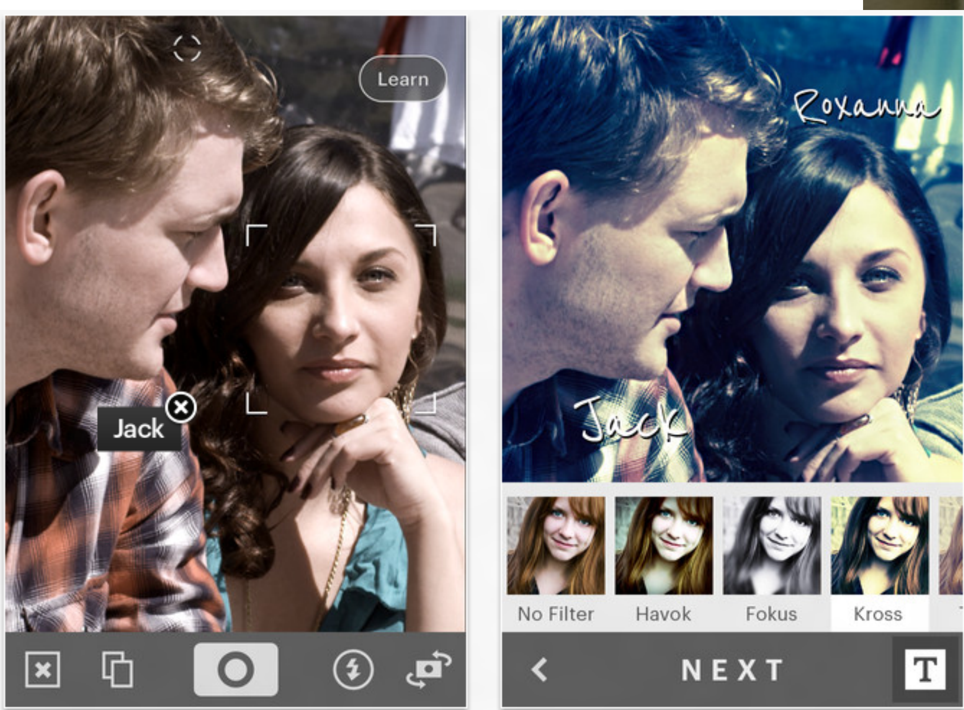
Significant progress: Recognizing objects

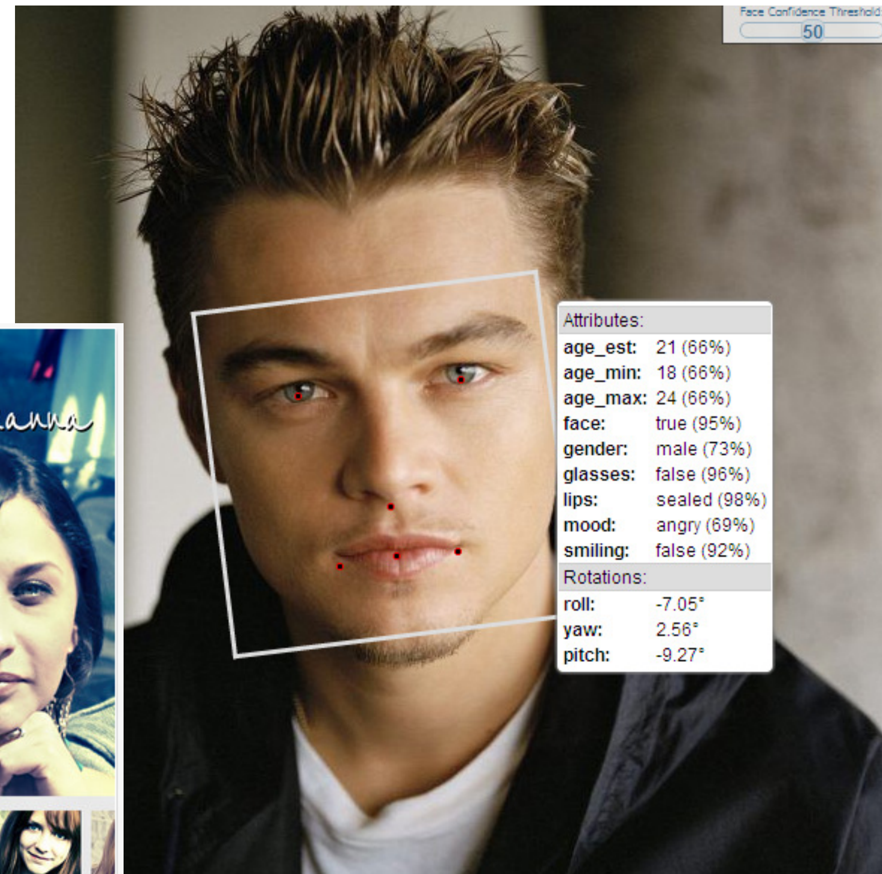


Mask R-CNN. Kaiming He, Georgia Gkioxari, Piotr Dollar, Ross Girshick. ICCV 2017

Significant progress: Face Recognition

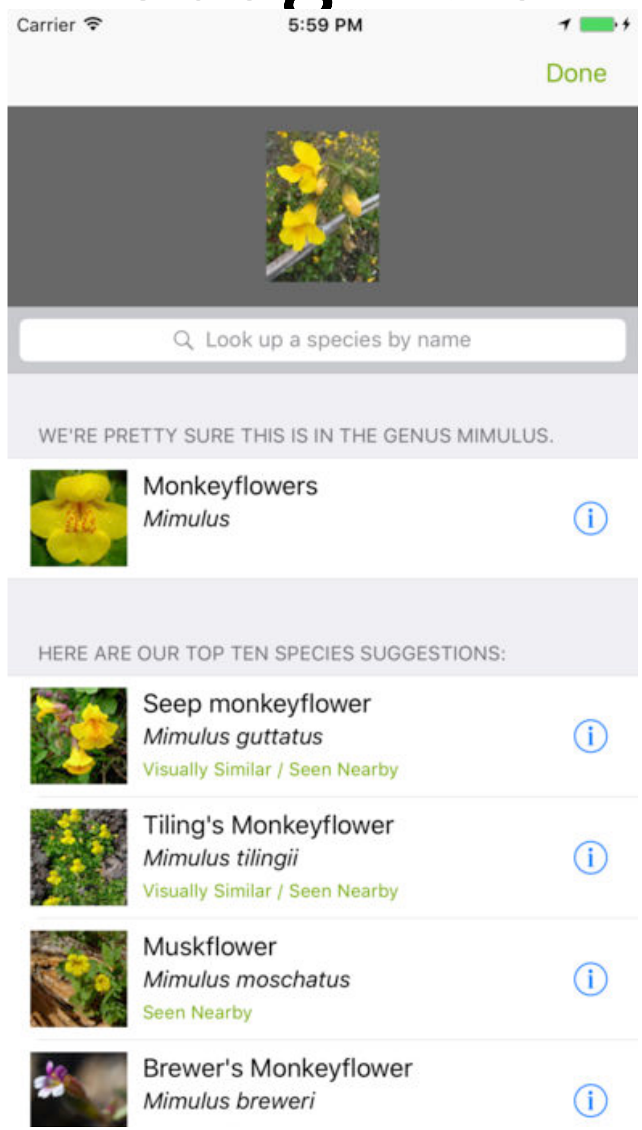
FACE++





Recognition-based product search

Significant progress: Species recognition



[iNaturalist]

Challenges: recognizing rare concepts

Aye-Aye

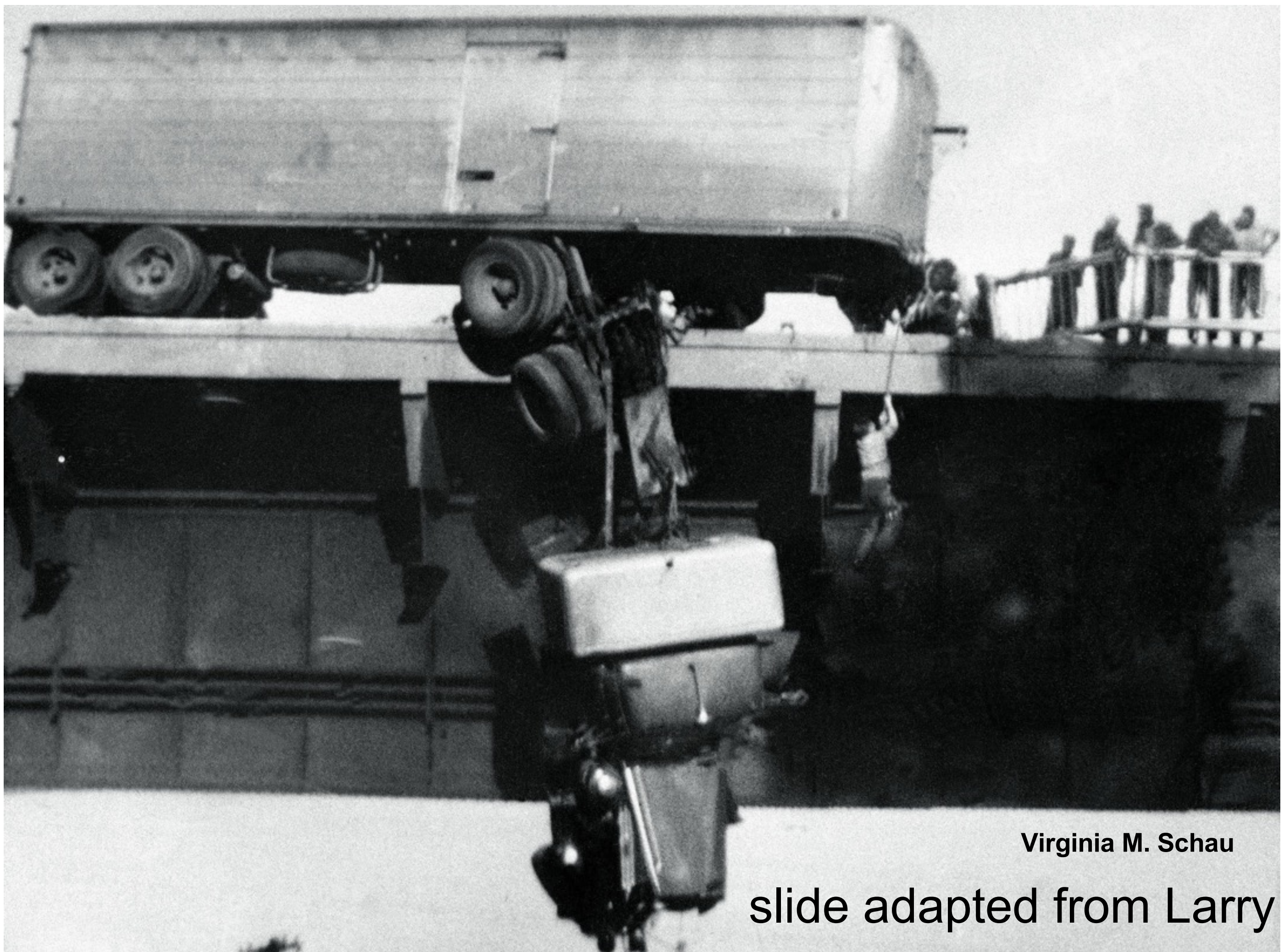
Challenges: recognizing rare concepts

Challenges: recognizing rare concepts

Challenges: Recovering 3D structure from limited views

Challenges: Reasoning

What is going to happen next?



Virginia M. Schau

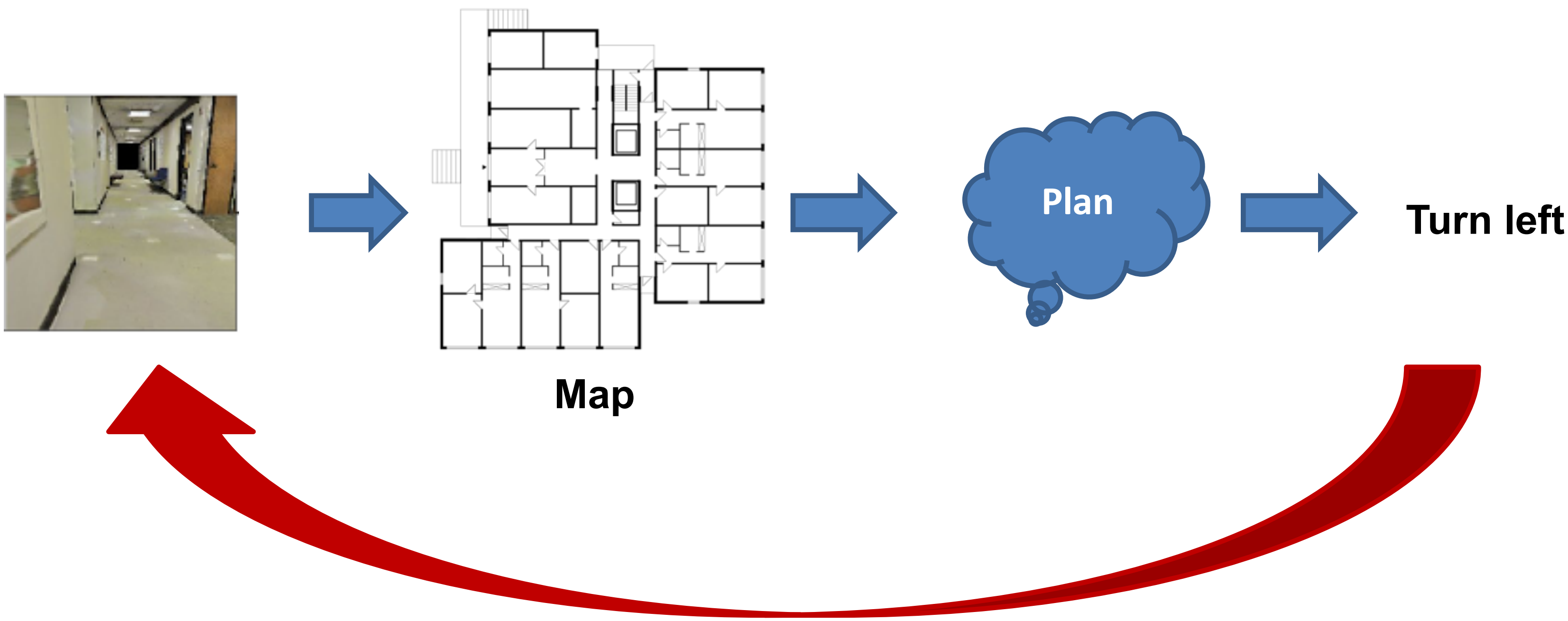
slide adapted from Larry Zitnick

Why is this funny?

The picture above is funny.

Andrej Karpathy

Challenges: Integrating Vision and Action



Cognitive Mapping and Planning for Visual Navigation
Saurabh Gupta, James Davidson, Sergey Levine, Rahul Sukthankar, Jitendra Malik
***CVPR* 2017**

Challenges: Other imaging domains

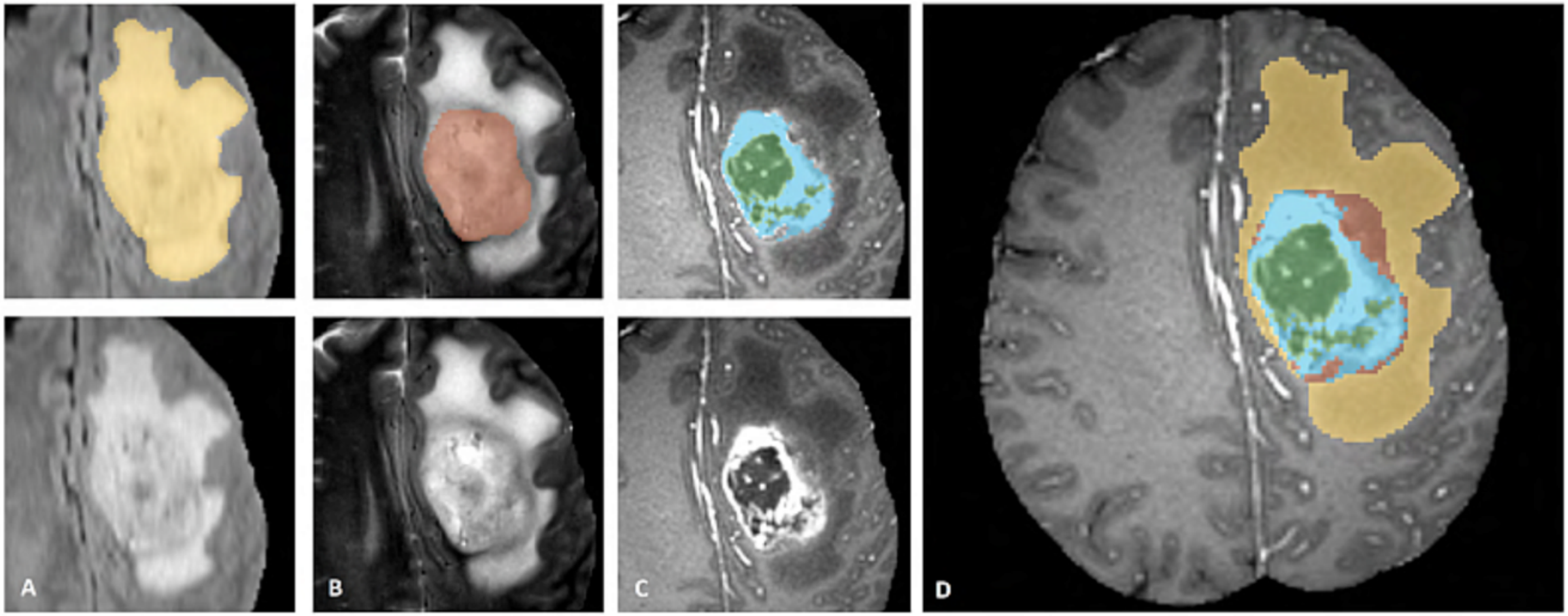


Fig.1: Glioma sub-regions. Shown are image patches with the tumor sub-regions that are annotated in the different modalities (top left) and the final labels for the whole dataset (right). The image patches show from left to right: the whole tumor (yellow) visible in T2-FLAIR (Fig.A), the tumor core (red) visible in T2 (Fig.B), the enhancing tumor structures (light blue) visible in T1Gd, surrounding the cystic/necrotic components of the core (green) (Fig. C). The segmentations are combined to generate the final labels of the tumor sub-regions (Fig.D): edema (yellow), non-enhancing solid core (red), necrotic/cystic core (green), enhancing core (blue). (Figure taken from the BraTS IEEE TMI paper.)

Our Course

Instructor

- Bharath Hariharan (bharathh@cs.cornell.edu)

- Office hours:

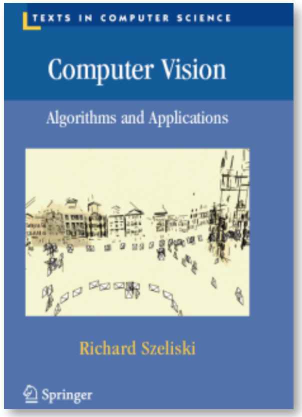
  M/W/Thu: 10:00-11:00, or by appointment

- Research interests:
  - Computer vision: Object recognition
  - Machine learning: Deep learning

Important personnel

- TAs:
  - Jimmy Briggs
  - Danlu (Athena) Huang
  - Alvin Zhu
  - Yiwei Ni
  - Karun Singh
  - Zhiqiu (Douglas) Lin

Other administrative details



- Textbook:

  Rick Szeliski, *Computer Vision: Algorithms and Applications*

  online at: http://szeliski.org/Book/

- Course webpage (lectures, assignments, OH):

  http://www.cs.cornell.edu/courses/cs4670/2018sp/

- Announcements/grades via Piazza/CMS

  Sign up on piazza

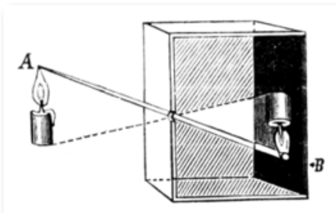
  https://cmsx.cornell.edu/

Course requirements

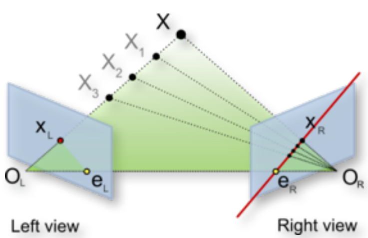
- Prerequisites—*these are essential*!
  - Data structures
  - A good working knowledge of python programming
  - Linear algebra
  - Calculus (plus basic multivariate calculus)

- Course does ***not*** assume prior imaging experience
  - computer vision, image processing, graphics, etc.

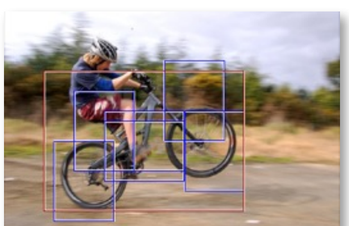
Course overview (tentative)



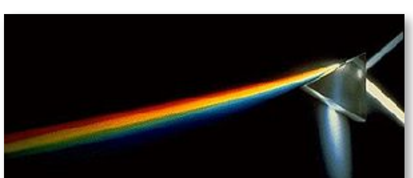
1. Low and mid-level vision
   - basic image formation
   - image processing, segmentation



2. Reconstruction
   - cameras, geometry and physics of image formation
   - stereo, structure from motion

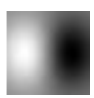
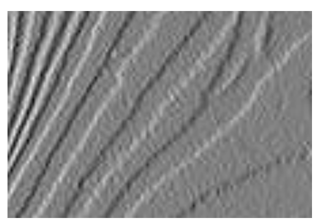


3. Recognition
   - primer on machine learning, convolutional networks
   - classification, detection, segmentation

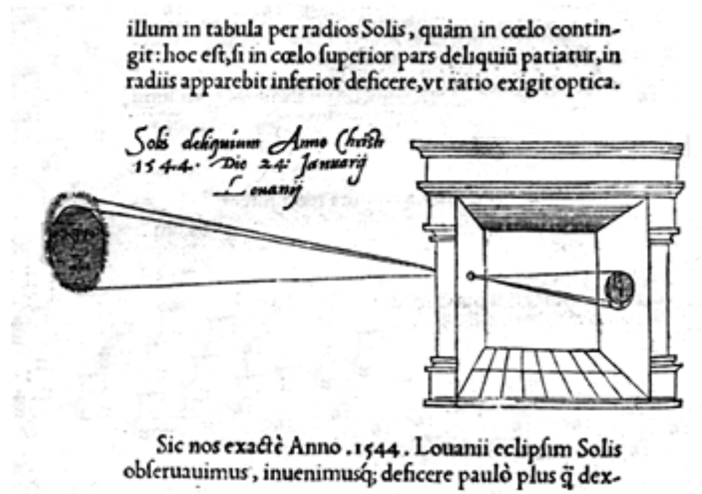


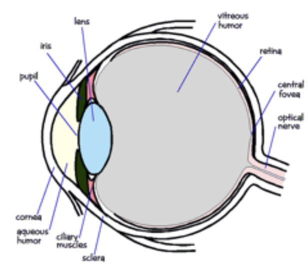
1. Low-level vision

- Basic image processing and image formation

Filtering, edge detection





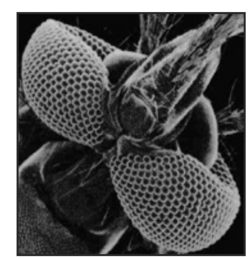
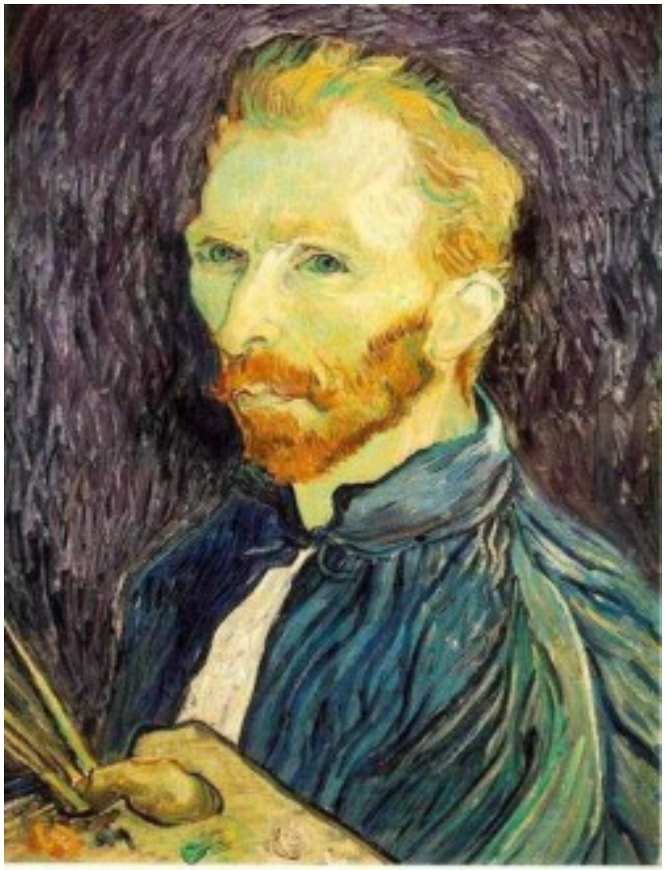


Image formation

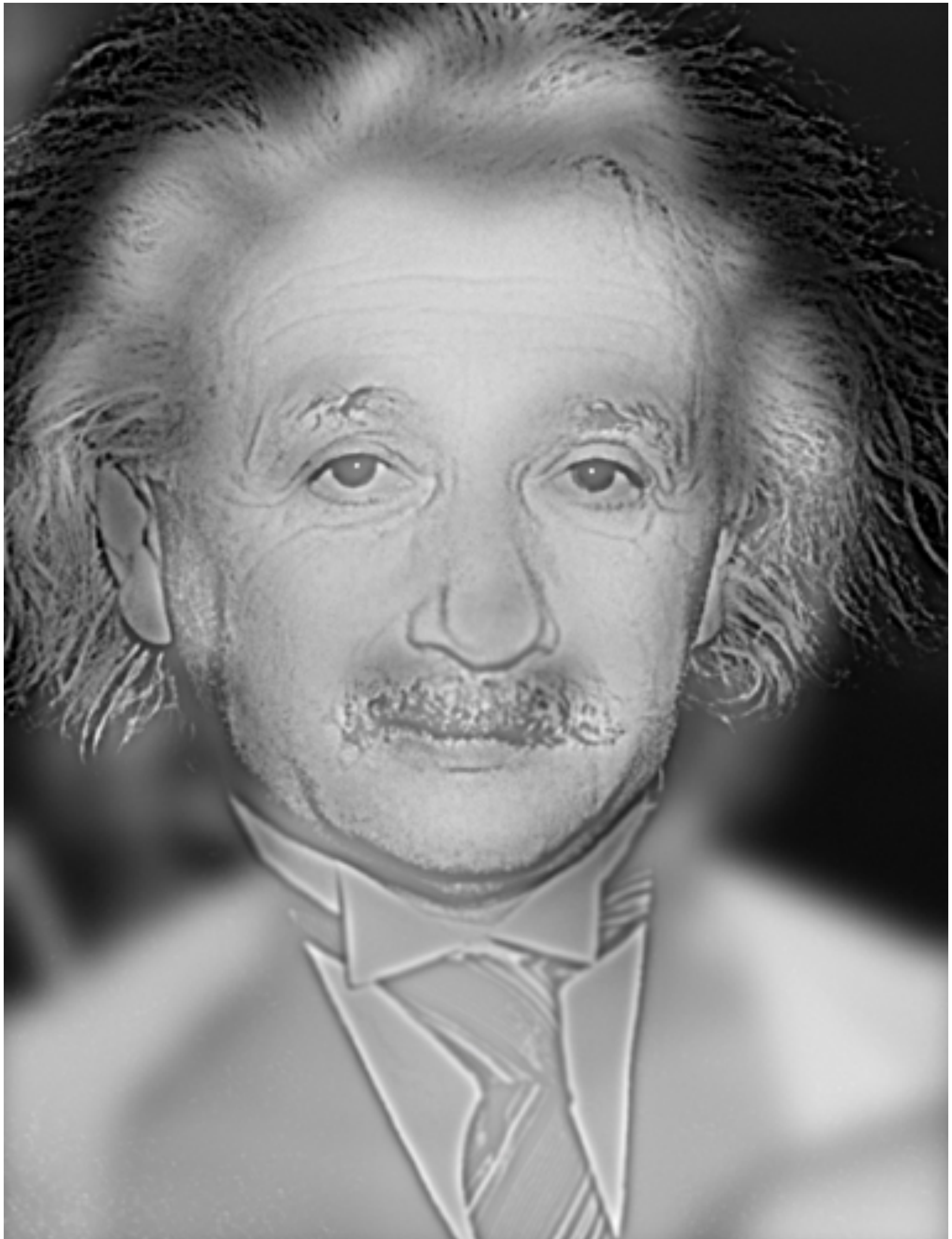
Project: Multiscale pyramids

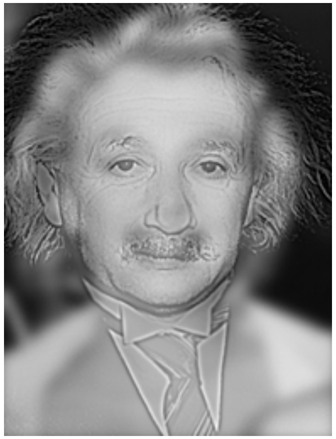


Gaussian 1/2

G 1/4

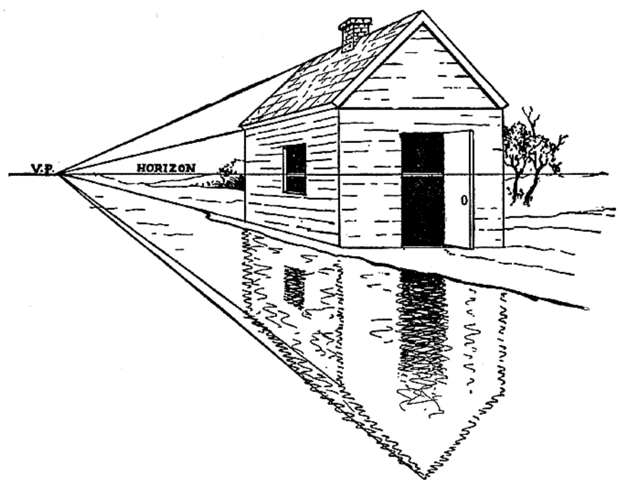
G 1/8



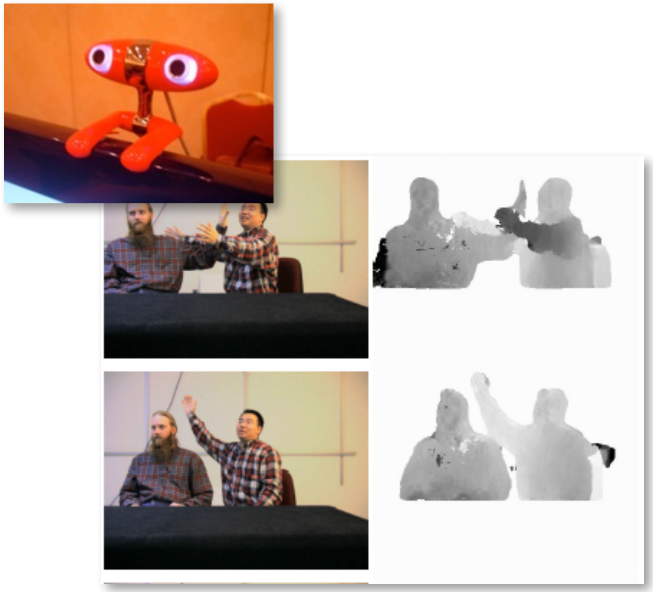


Project: Grouping and segmentation

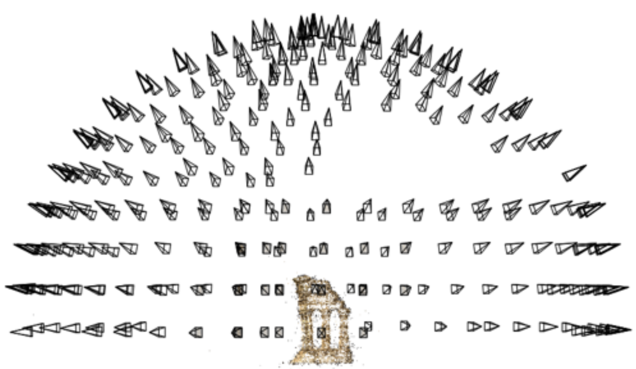
2. Reconstruction



Projective geometry



Stereo



Multi-view stereo

Structure from motion

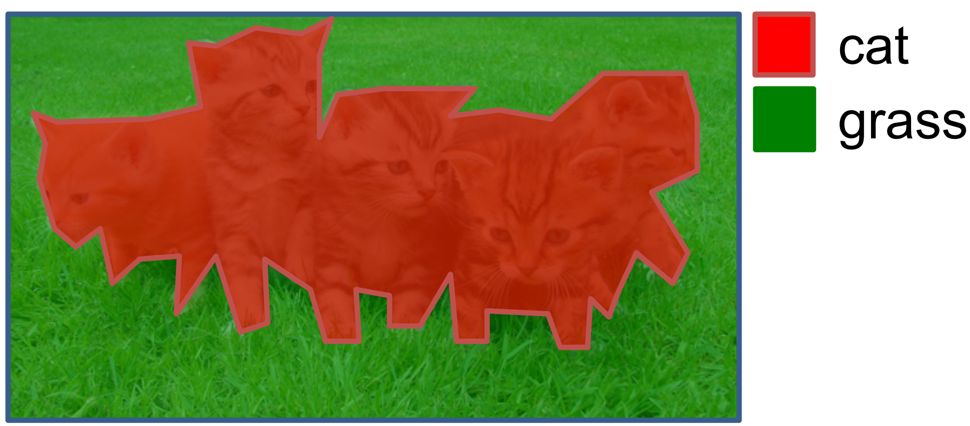
Project: Feature detection and matching

Project: Stereo and photometric stereo

3. Recognition

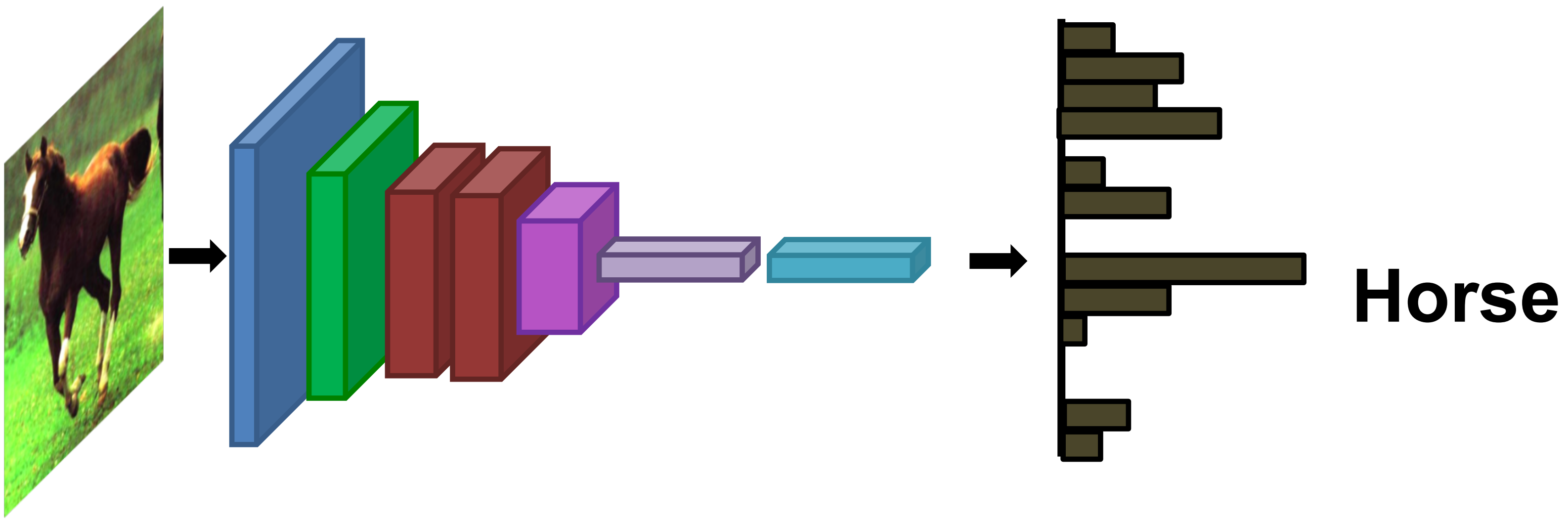
Image classification

Object detection



Semantic segmentation

Project: Deep learning for classification



Grading

- One prelim, one final exam
- Rough grade breakdown:
  - Midterm: 15-20%
  - Homeworks: 10-20%
  - Programming projects: 40-50%
  - Final exam: 15-20%

Late policy

- Five free “slip days” will be available for the semester

Academic Integrity

- Homeworks have to be done alone
  - No discussing
- Assignments in pairs

Questions?