CS4670/5670: Intro to Computer Vision

Instructor: Bharath Hariharan

Today

1. Administrivia

2. What is computer vision?

3. Course overview

(Today not to be tested)

Course web page

- http://www.cs.cornell.edu/courses/cs4670/2020sp/

Instructor

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

- Office hours:

  M/F: 2:30 – 3:30 pm, or by appointment, Gates 311

- Research interests:
  - Computer vision
  - Machine learning

Grading

- **Scribing lectures (2%)**
  - Take notes and share with class
  - Sign up here
- **Programming assignments 1 - 4 (40% total)**
  - These have to be done in groups of 2. Find a partner!
- **Written homeworks (10% total)**
  - These have to be done individually (no discussion)
- **Prelim (20%)**
- **Final (28%)**

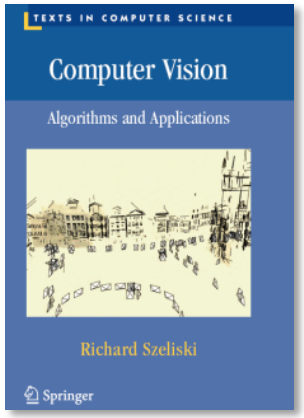
Late policy

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

Important personnel

- Lindsay Fei
- Apoorv Khandelwal
- Jiwon Kim
- Eric Landgrebe
- Frank Li
- Mason Liu
- Alisha Mithal
- Riley Niu
- Kane Tian
- Albert Tsao
- Ziyang Wu
- Xiaokai (Steven) Ye
- Stephanie Chang

Other administrative details



- Reference for more details *(not 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/cs4670/2020sp/

- 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
  - Calculus (plus basic multivariate calculus)
- Pre-requisites (highly encouraged)
  - Linear algebra (tutorial)
- Course does ***not*** assume prior imaging experience
  - computer vision, image processing, graphics, etc.

Questions?

Today

- What is computer vision?
  - What is the input?
  - What should be the output?
- How easy or hard is it?
- What have we done so far?
- Our course

WHAT IS COMPUTER VISION?

What is computer vision?

- What is the input?

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

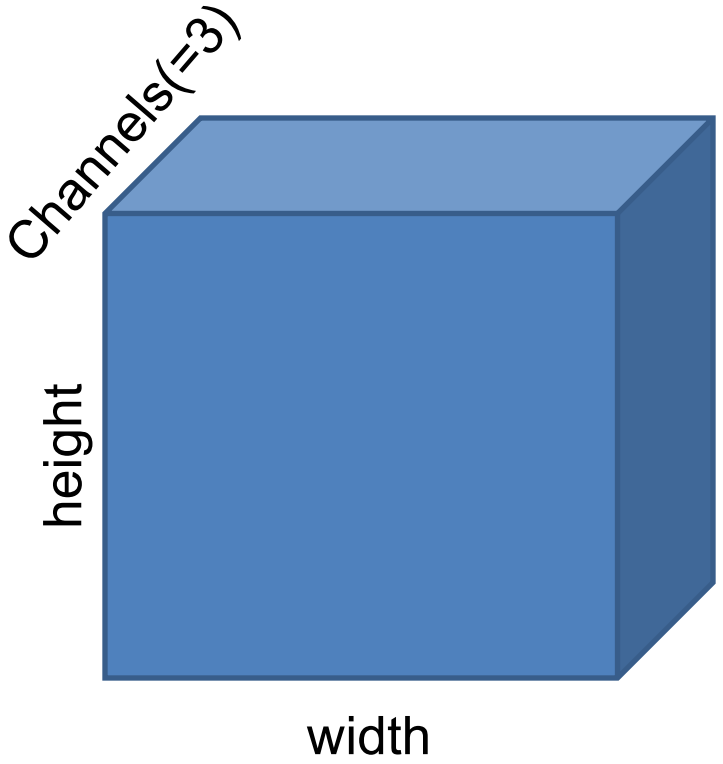
What is the input?

- A (gray-scale) image is a 2D array
- For purposes of this course, array of real numbers
  - *Any 2D array of real numbers can be treated as an image*
  - *Typically white = high, black = low*

What is the input?

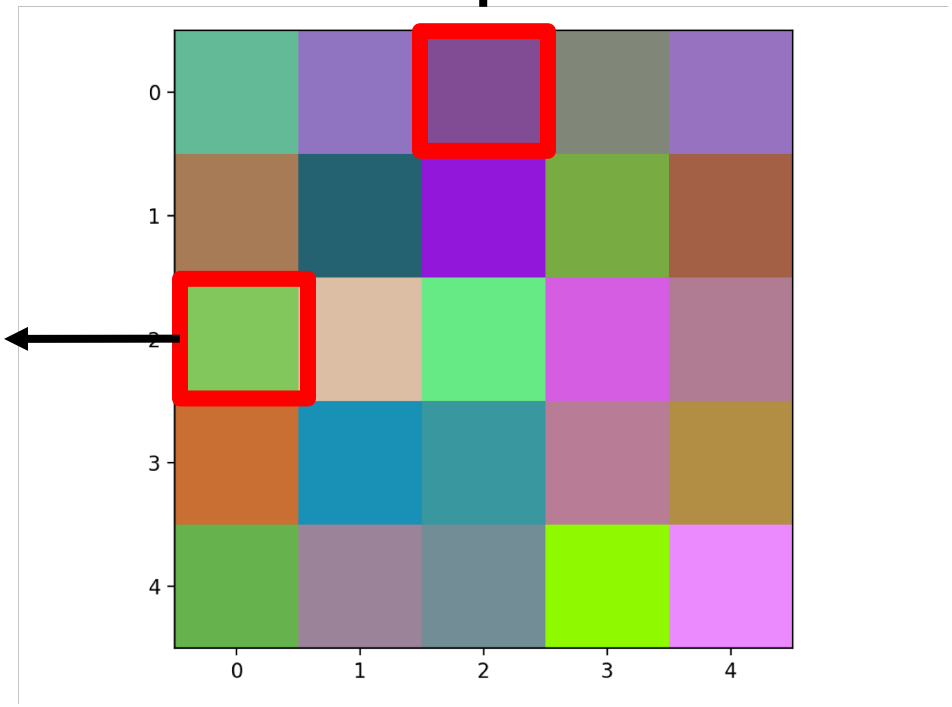
- An image is a 2D array
- For purposes of this course, array of real numbers
  - *Any 2D array of real numbers can be treated as an image*
- What about color images?

What is the input



What is the input

[130,76,149]

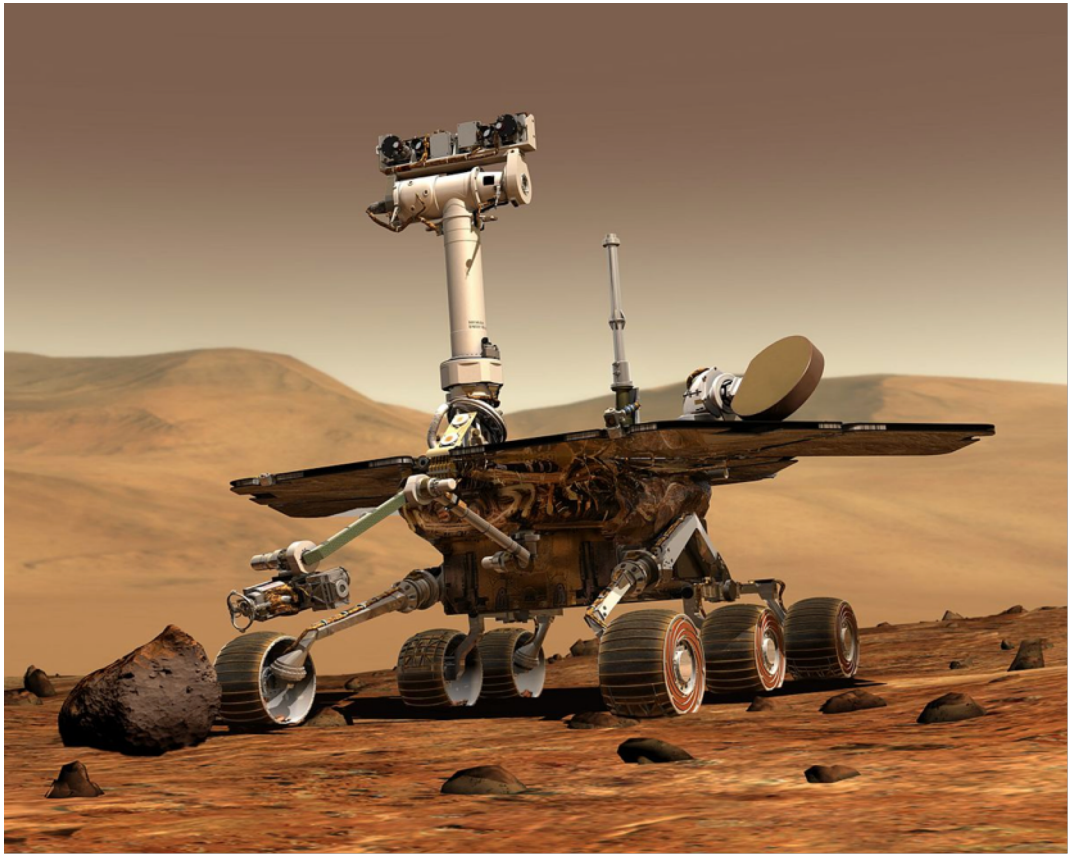


[130,199,92]

What is the output?

- Depends on what we want to do with the image

What do we do with images? Example 1: Robotics



- Understanding terrain and identifying obstacles

What do we do with images? Example 1: Robotics

- Understanding terrain and identifying obstacles
- Identifying people and understanding their intentions

http://theoatmeal.com/blog/google_self_driving_car

What do we do with images? Example 2: Internet Vision

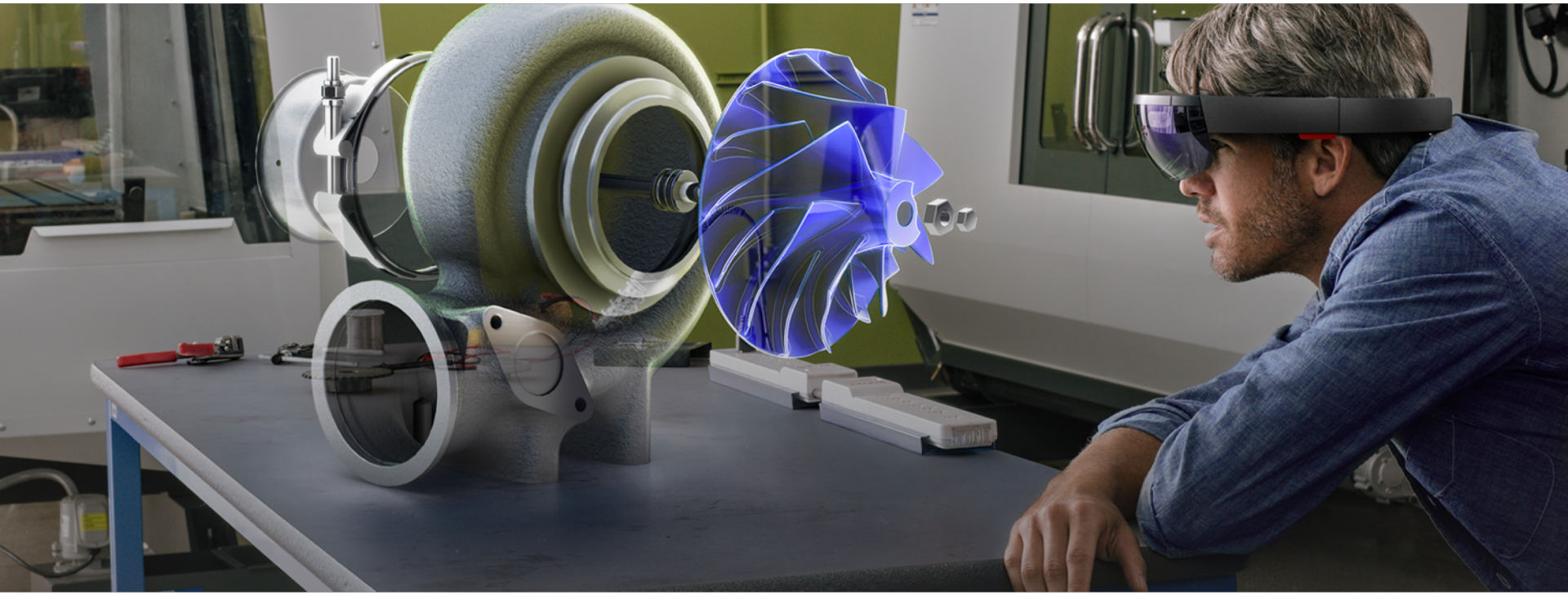
Facebook Users Are Uploading 350 Million New Photos Each Day

Cooper Smith

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

- Recognizing obscene/ violent content
- Creating new content (image editing)

What do we do with images?
Example 3: AR/VR



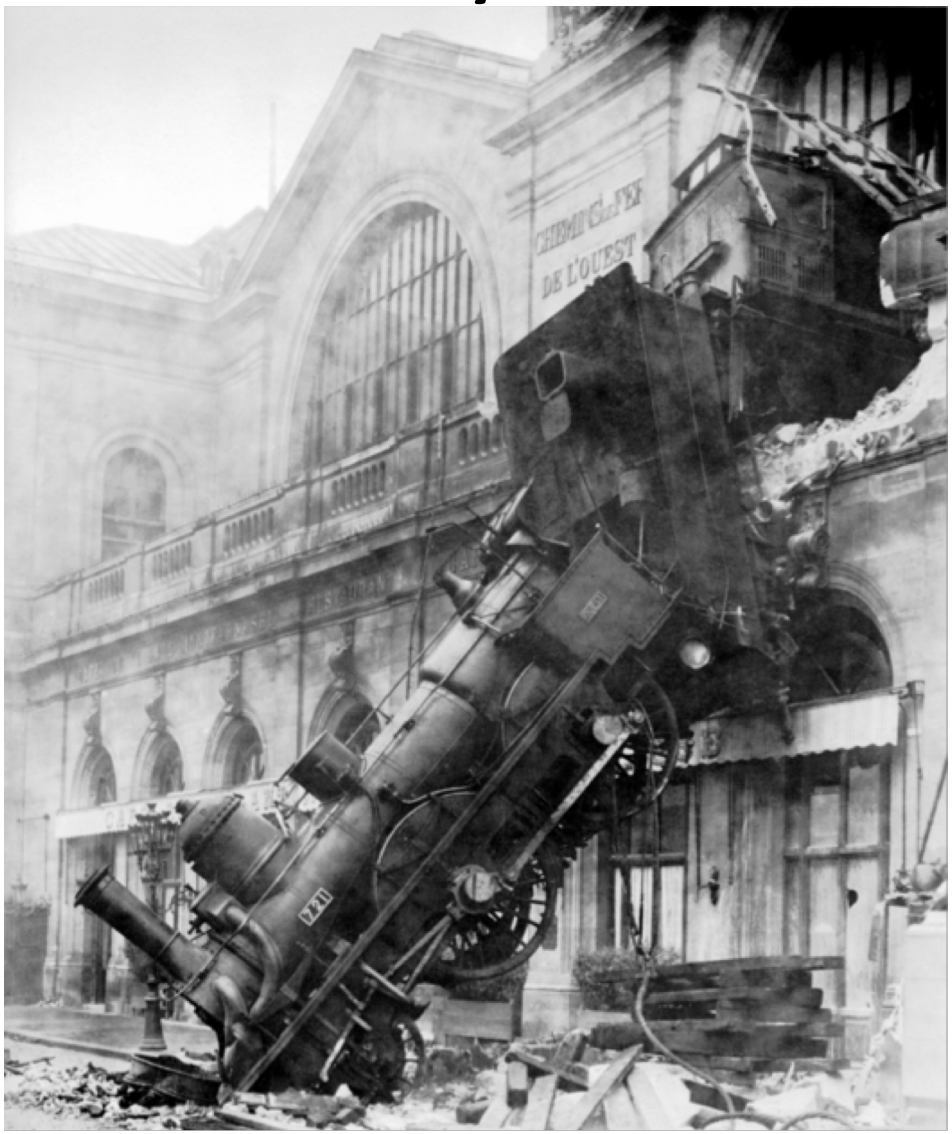
- Understand 3D structure of the world

The goal(s) or computer vision

- Reconstruction
  - Understanding 3D structure of the world
- Grouping
  - Group pixels into objects
- Recognition
  - Classify objects, scenes, actions…

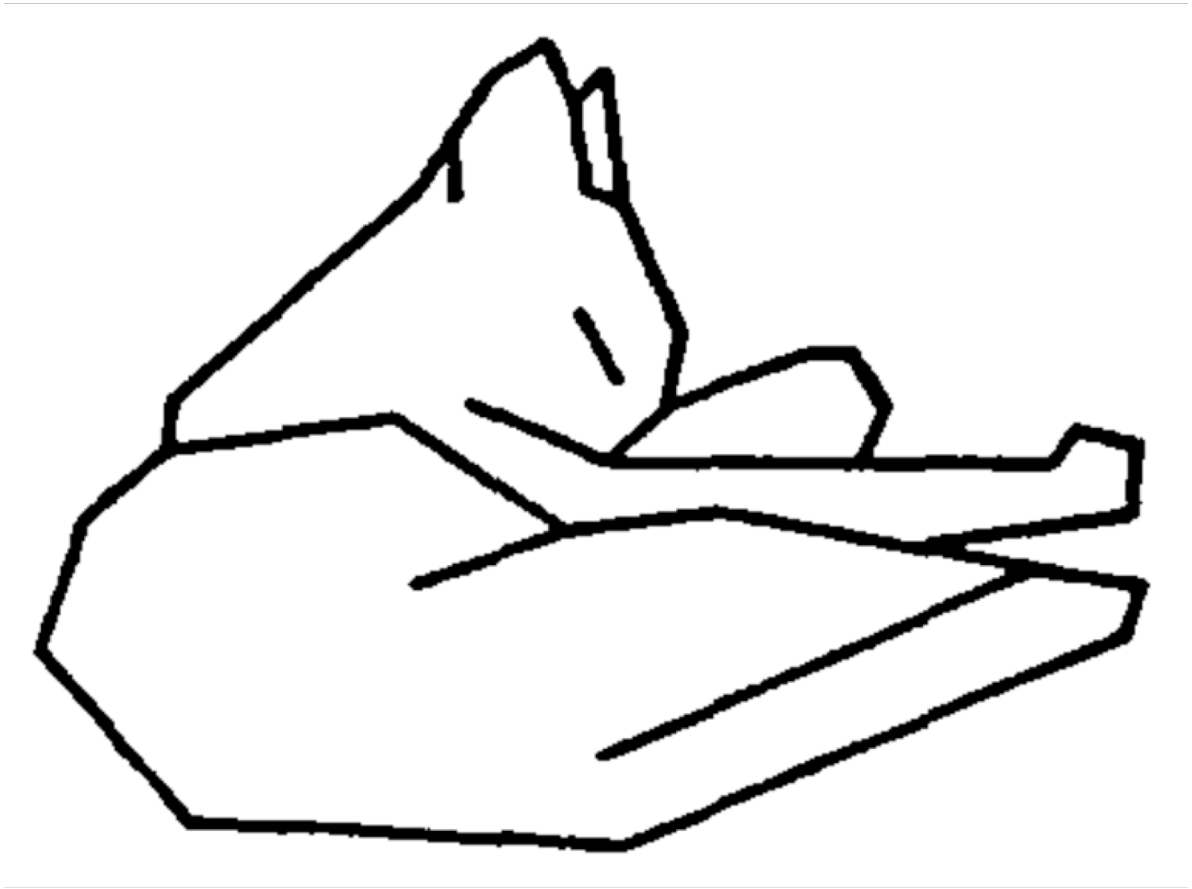
HOW EASY/HARD IS IT?

Vision is easy for humans



Vision is easy for humans

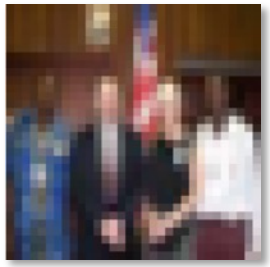
- Attneave's cat



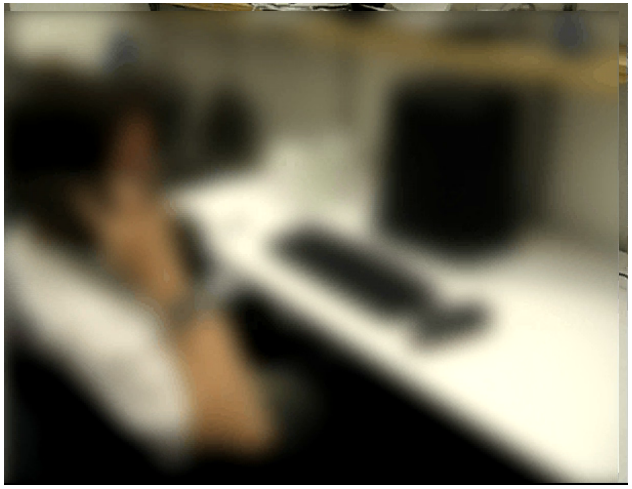
Vision is easy for humans

- Mooney faces

Vision is easy for humans



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



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

Vision is hard: Concepts are subtle

Tenessee Warbler

Orange Crowned Warbler

https://www.allaboutbirds.org

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 inseparable 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

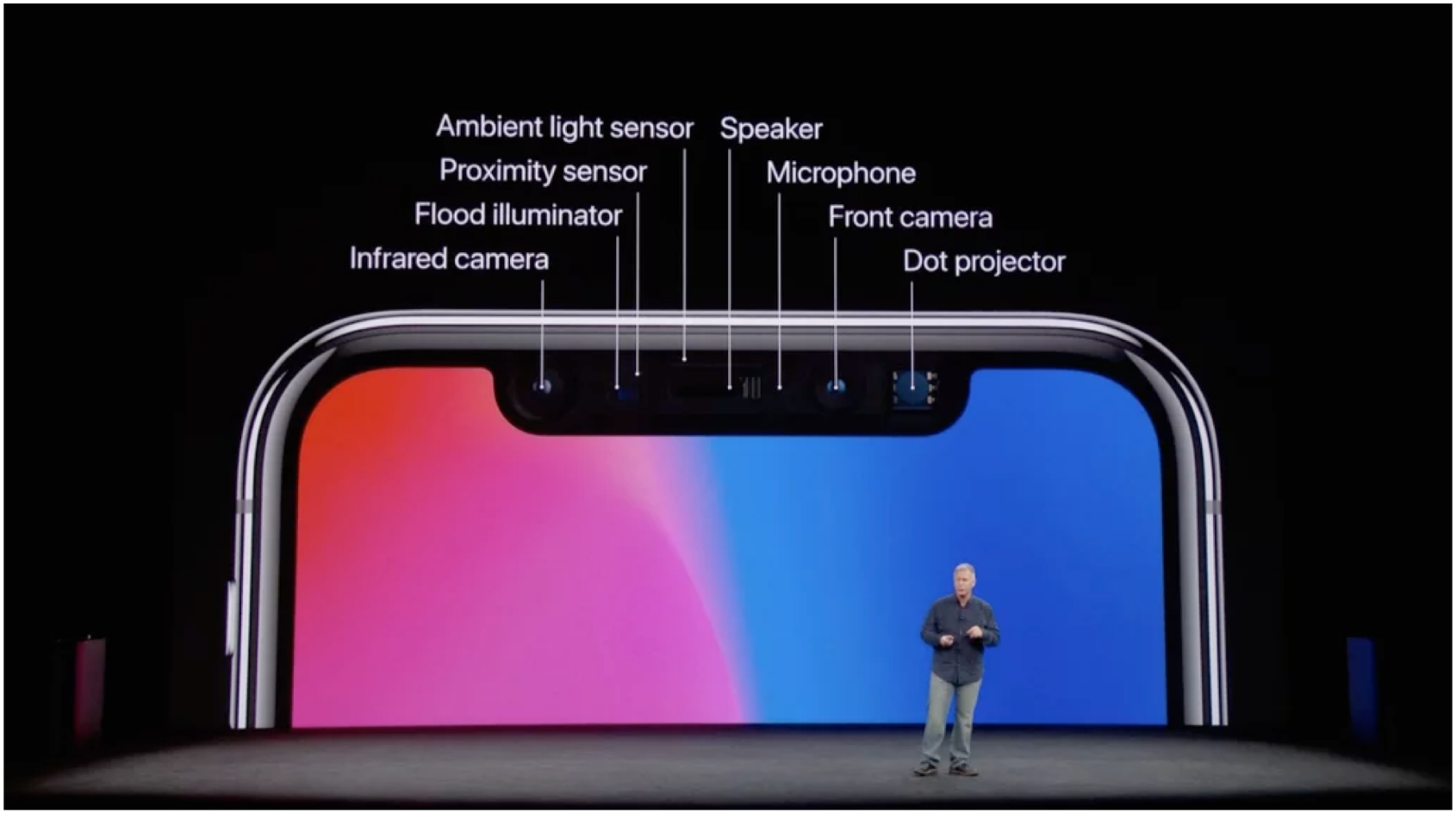
WHERE ARE WE NOW?

Deployed: depth cameras

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

Microsoft Kinect

Deployed: depth cameras



Iphone TrueDepth

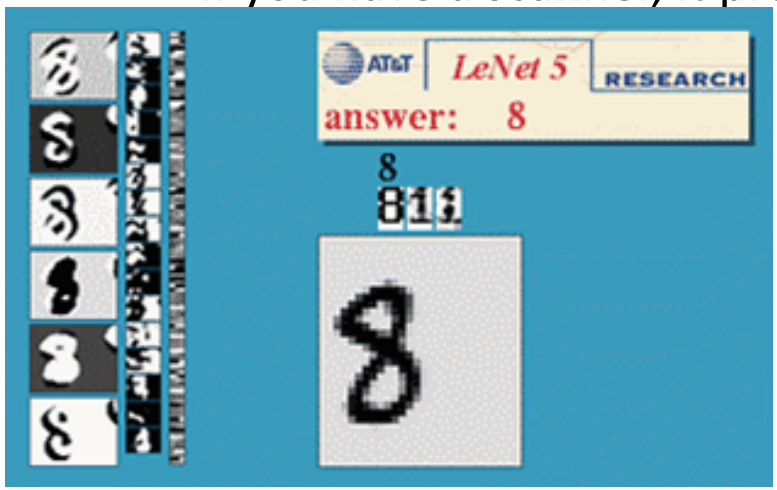
Deployed: shape capture

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

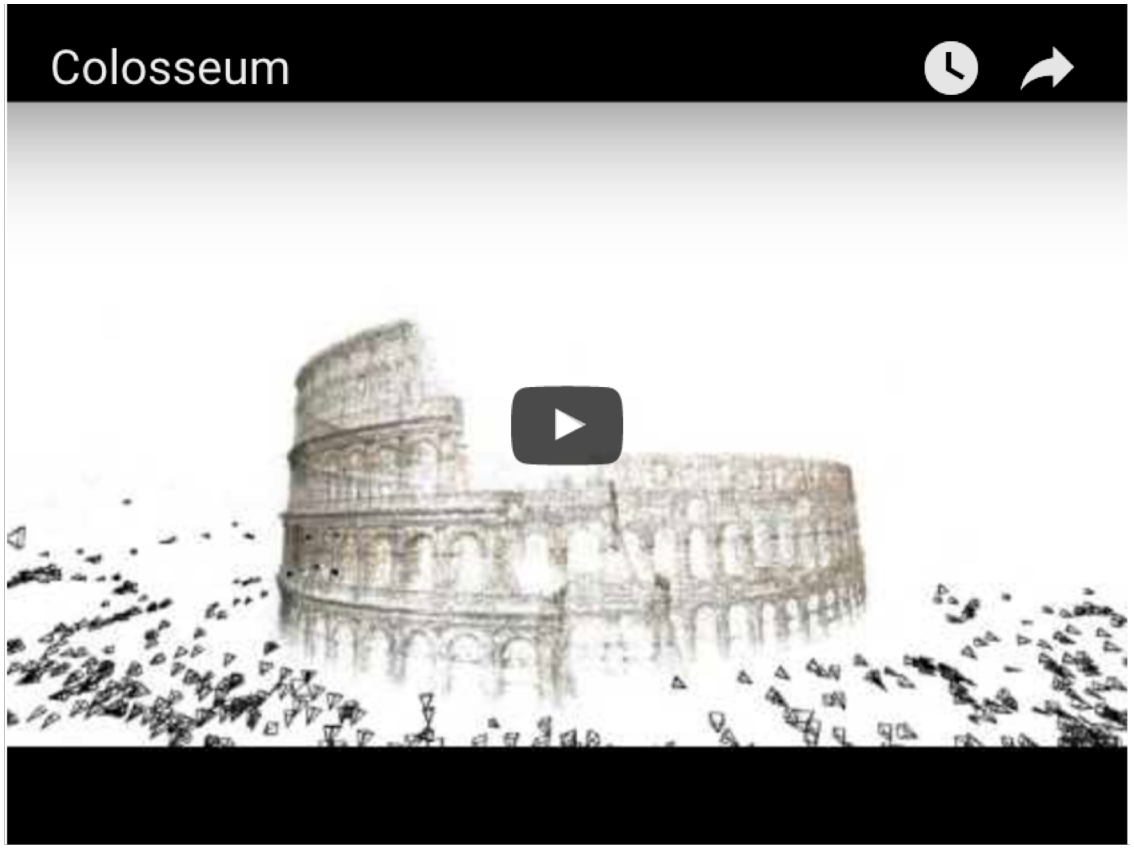
Source: S. Seitz

Deployed: Face detection

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

Source: S. Seitz

Established technology: 3D Models of the world



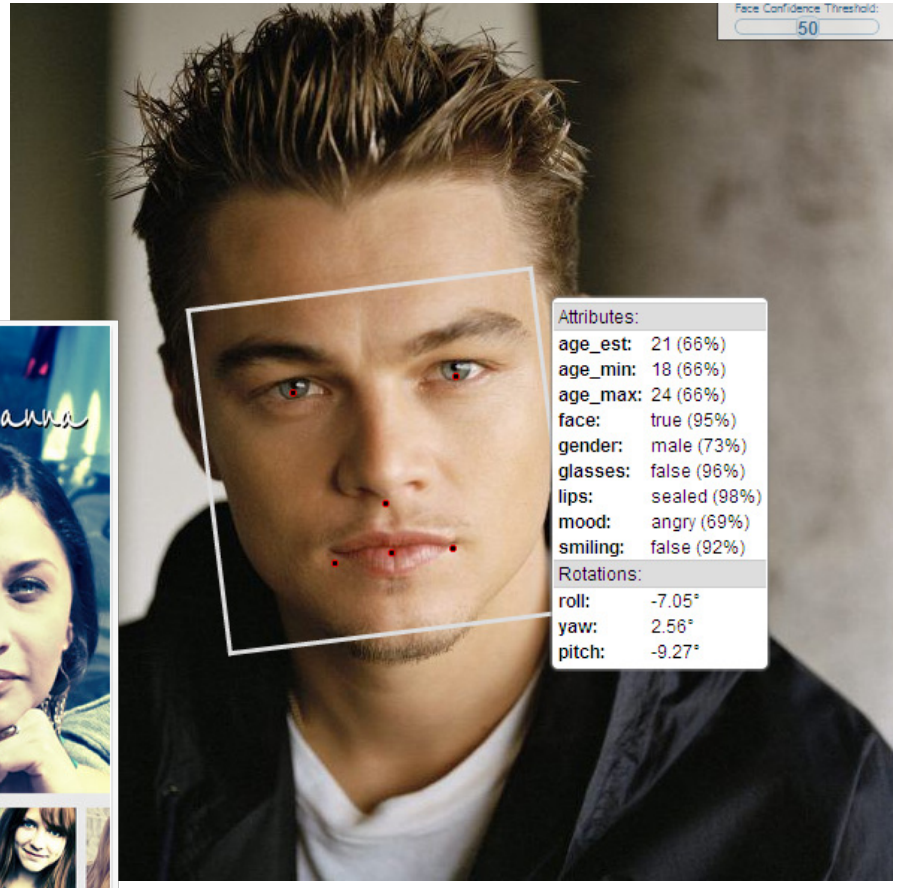
Building Rome in a Day.
Sameer Agarwal, Noah Snavely, Ian Simon, Steven M. Seitz and Richard Szeliski.
ICCV, 2009, Kyoto, Japan.

Significant progress: Recognizing objects

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

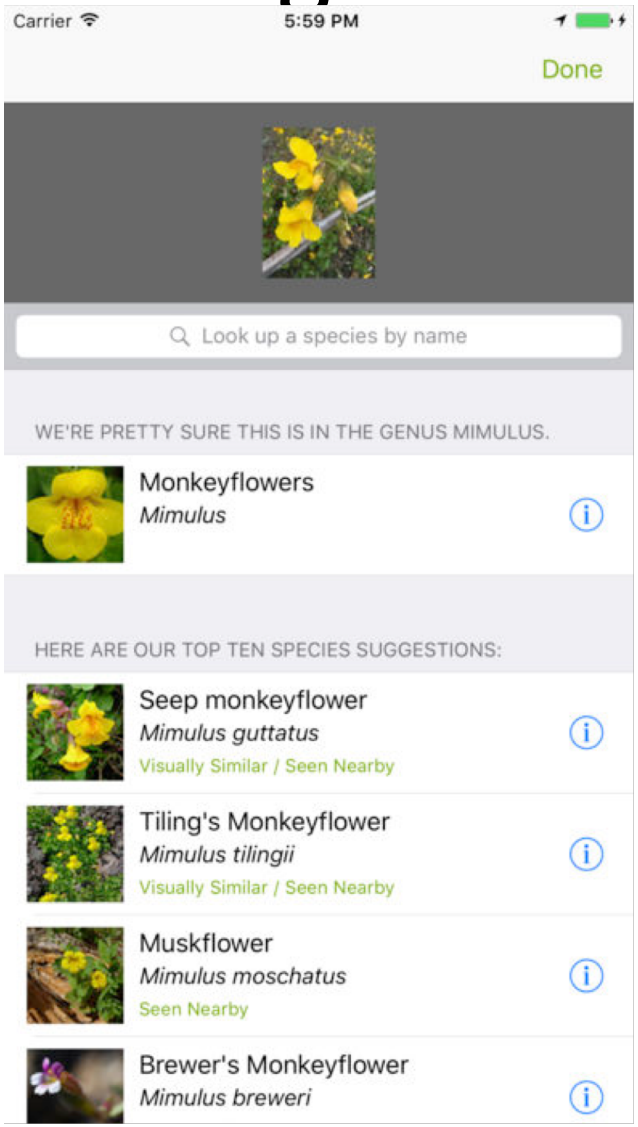
Significant progress: Face Recognition

FACE++



Recognition-based product search

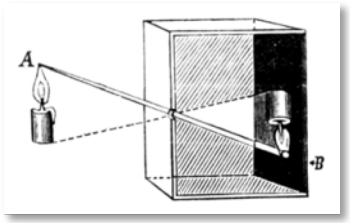
Significant progress: Species recognition

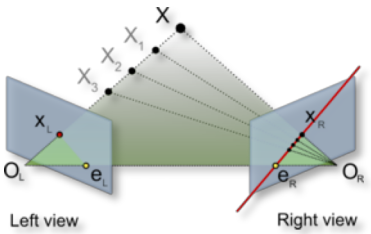


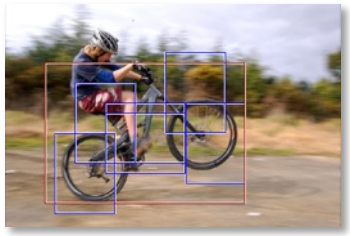
[iNaturalist]

OUR COURSE

Course overview (tentative)



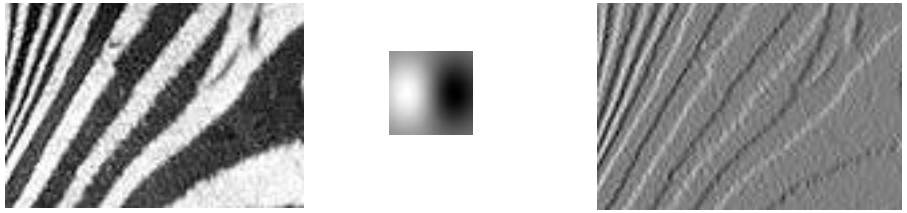




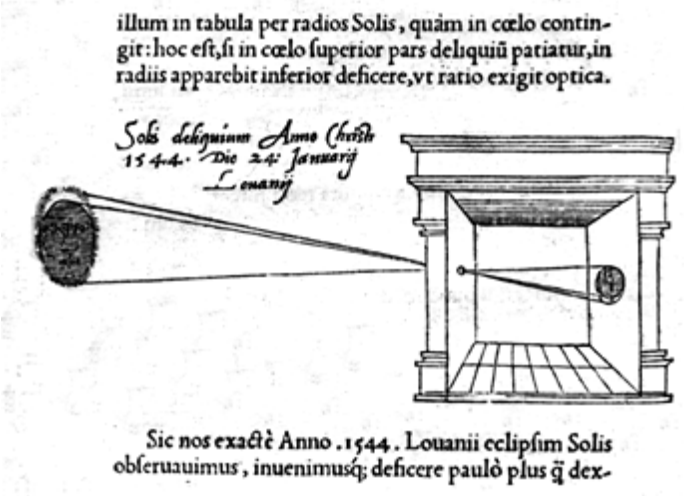
1. Low and mid-level vision
   - basic image formation
   - image processing, segmentation
2. Reconstruction
   - cameras, geometry and physics of image formation
   - stereo, structure from motion
3. Recognition
   - primer on machine learning, convolutional networks
   - classification, detection, segmentation

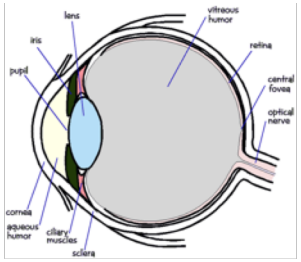
1. Low-level vision

- Basic image processing and image formation



Filtering, edge detection





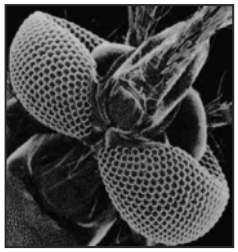
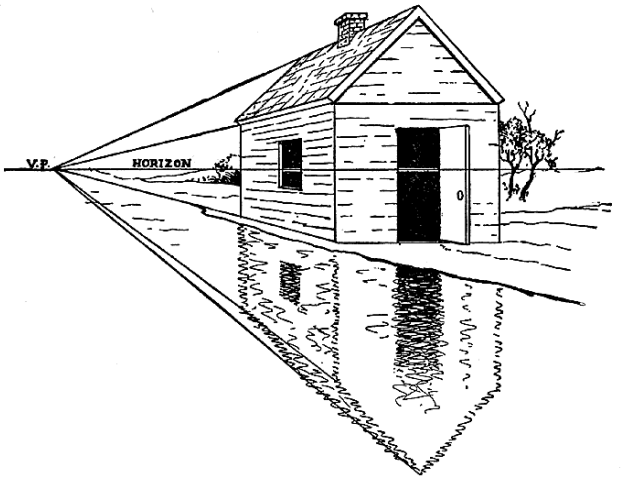
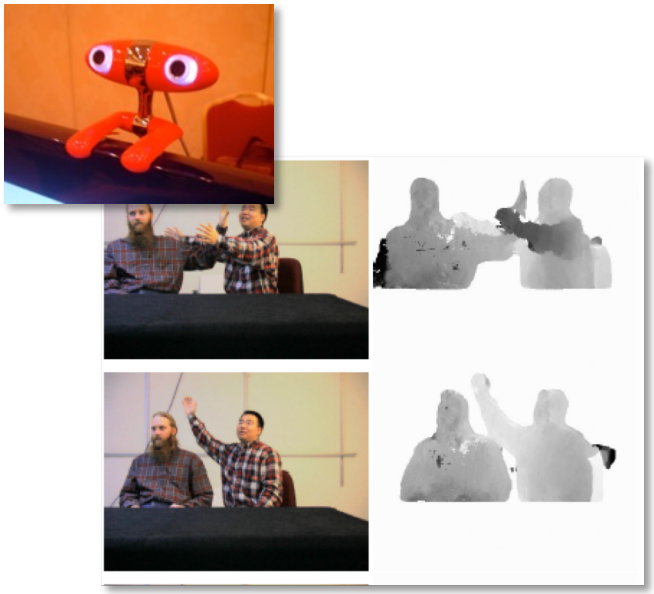


Image formation

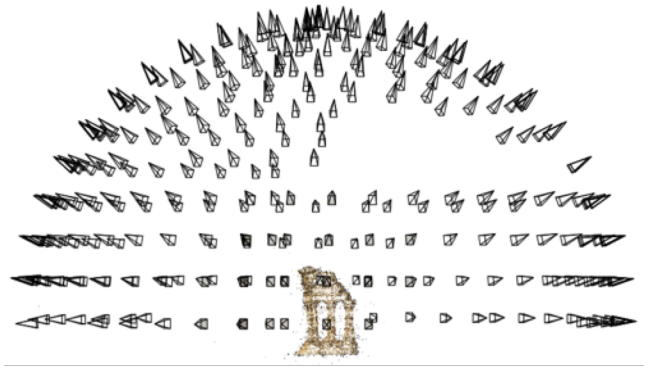
2. Reconstruction



Projective geometry



Stereo



Multi-view stereo

Structure from motion

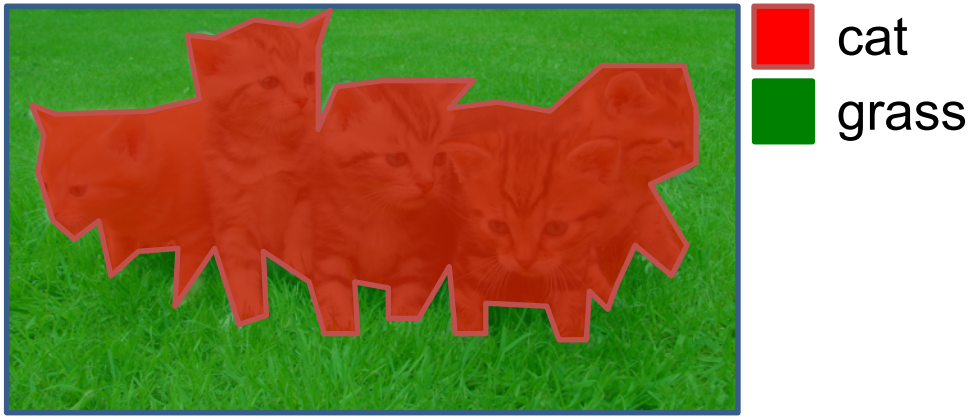
Project: Feature detection and matching

Project: Stereo and photometric stereo

3. Recognition

Image classification

Object detection



Semantic segmentation

Project: Deep learning for classification