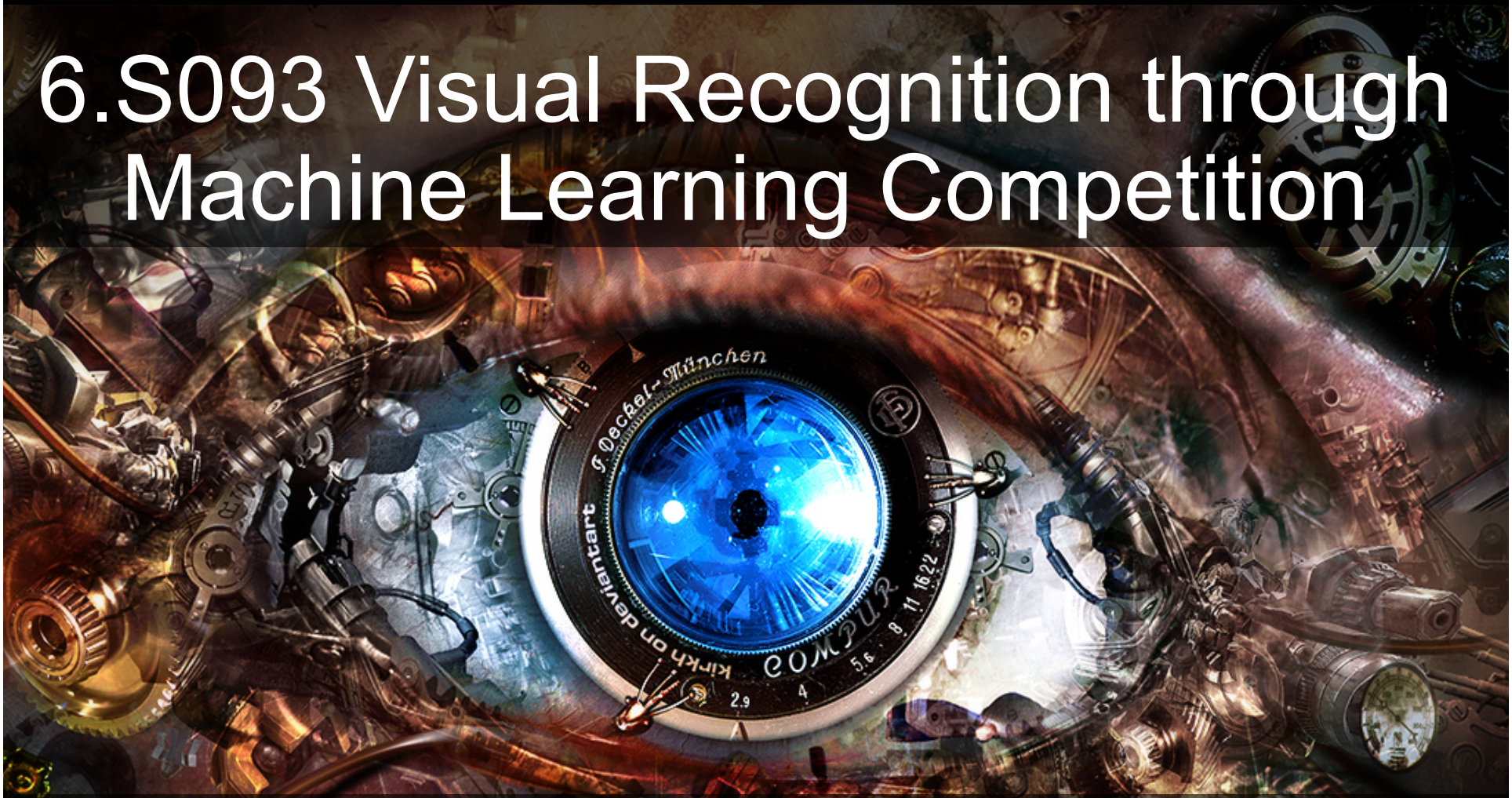
6.S093 Visual Recognition through Machine Learning Competition



Instructors:

Aditya Khosla

Joseph J. Lim

Image by kirkh.deviantart.com

Today's class

- Part 1: Intro lecture
  - Course overview
  - What is computer vision?
  - Why study computer vision?
  - Why is computer vision difficult?
  - Common problems in computer vision

- Part 2: Hands-on MATLAB Tutorial

Course overview: who?

- Non-experts in computer vision
- People who want to develop vision applications

- Pre-requisites:
  - Basic programming experience (MATLAB preferred)
  - Basic linear algebra, probability, calculus
  - No prior knowledge of computer vision required
  - Laptop is required

Course overview: what?

Week	Date	Topic	Presenter
1	January 13	Introduction to Computer Vision	Aditya
	January 14	Image Representations	Joseph
	January 15	Image Representations (continued)	Aditya
	January 16	Introduction to Machine Learning	Joseph
	January 17	Introduction to Deep Learning	Aditya
2	January 21	Guest lecture	
	January 22	Competition prep	
	January 23	Competition prep	
	January 24	Competition prep	
3	January 27	Student presentations + Awards Ceremony	You!

Course website:
http://viscomp.csail.mit.edu

Competition Sponsor:

Course overview: where?

- Here! (56-154 from 1-3pm)
- Please bring your laptops

What is computer vision?

Done?

Every image tells a story

- Goal of computer vision: perceive the "story" behind the picture
- Compute properties of the world
  - 3D shape
  - Names of people or objects
  - What happened?

The goal of computer vision

what we see

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 computers see

The goal of computer vision

what we see

what computers see

What is computer vision?

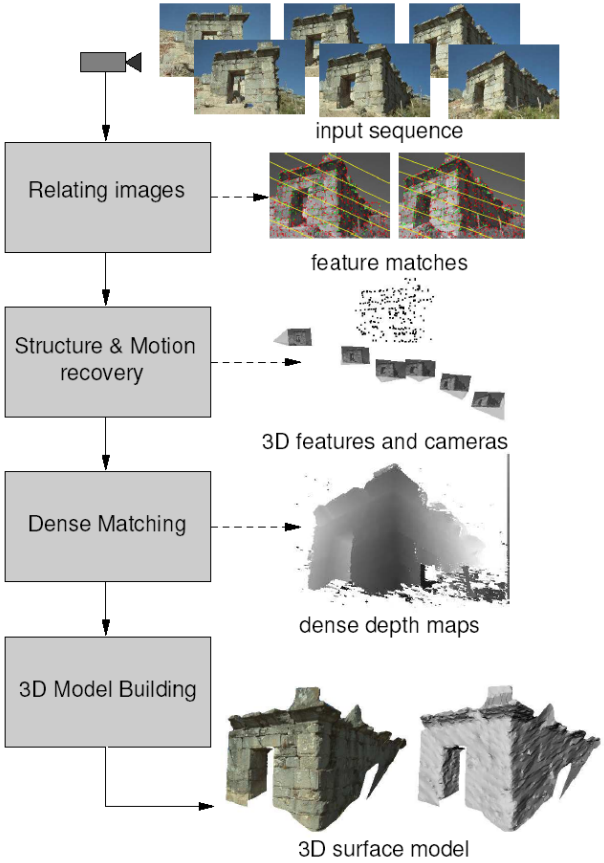
- Automatic understanding of images and video
  - Computing properties of the 3D world from visual data *(measurement)*
  - Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. *(perception and interpretation)*

Vision for measurement

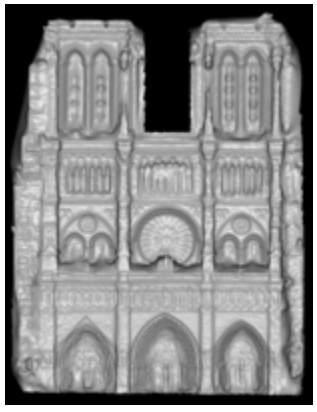
Real-time stereo

Pollefeys et al.

Structure from motion



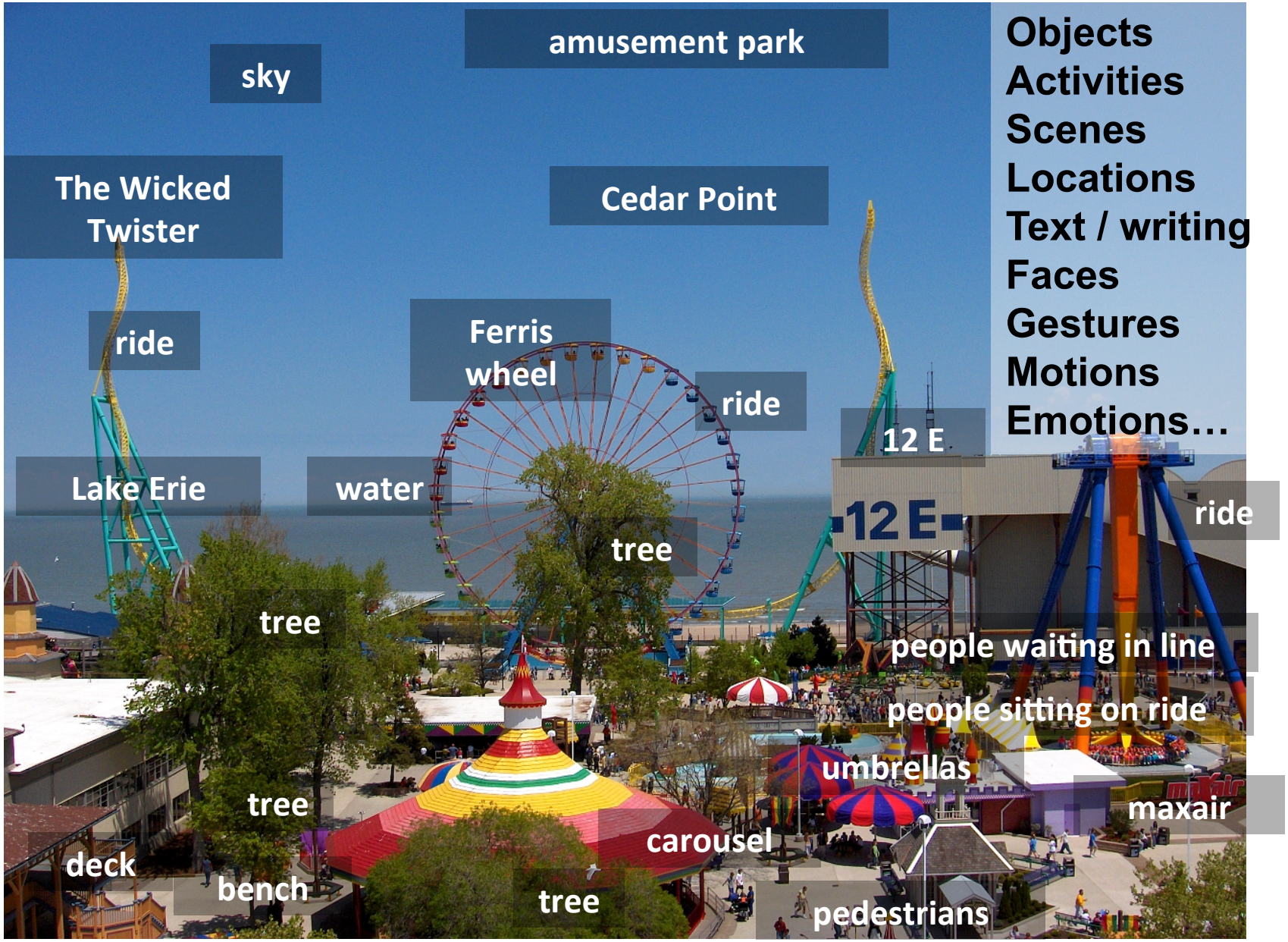
Multi-view stereo for community photo collections



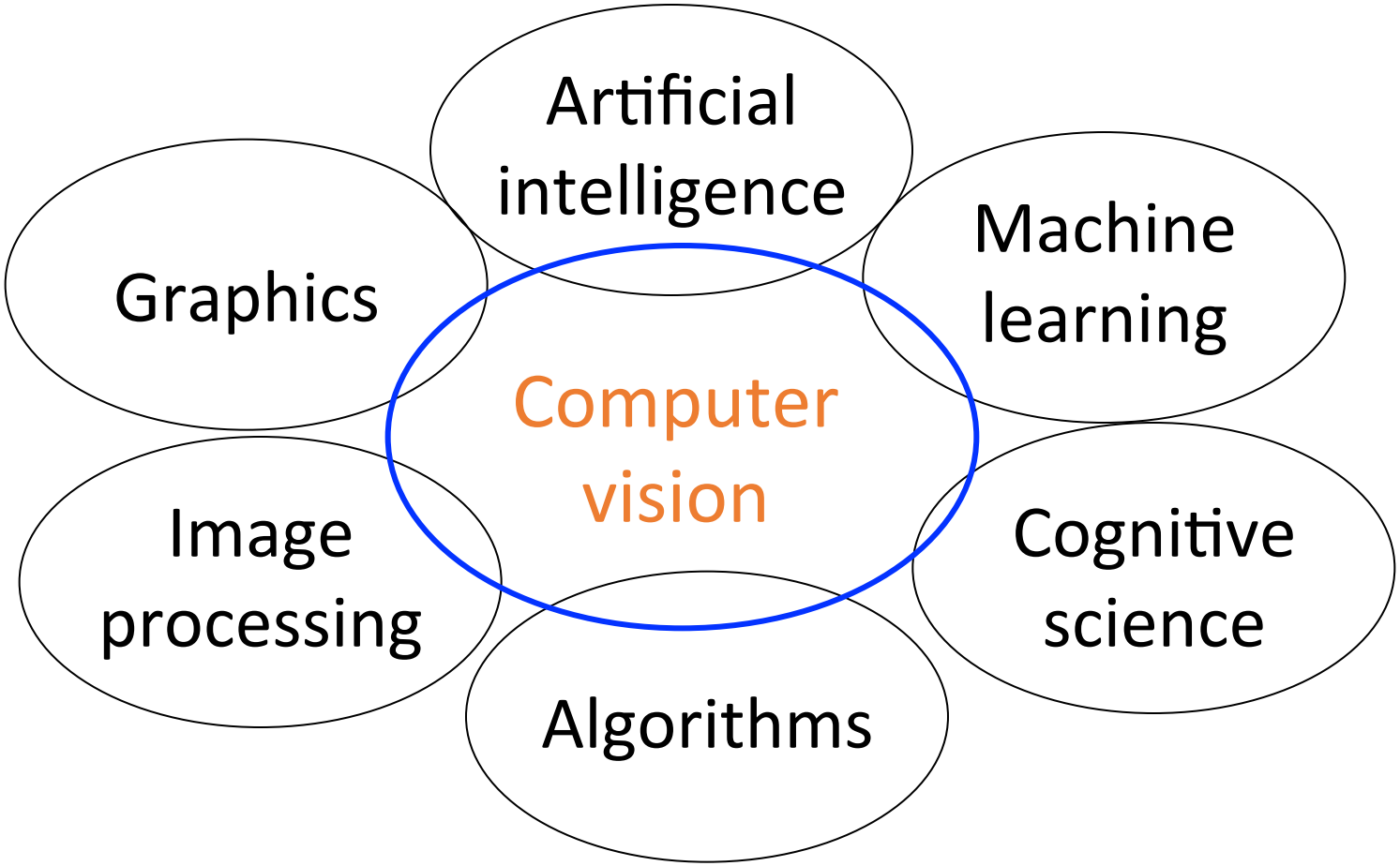
Goesele et al.

Slide credit: L. Lazebnik

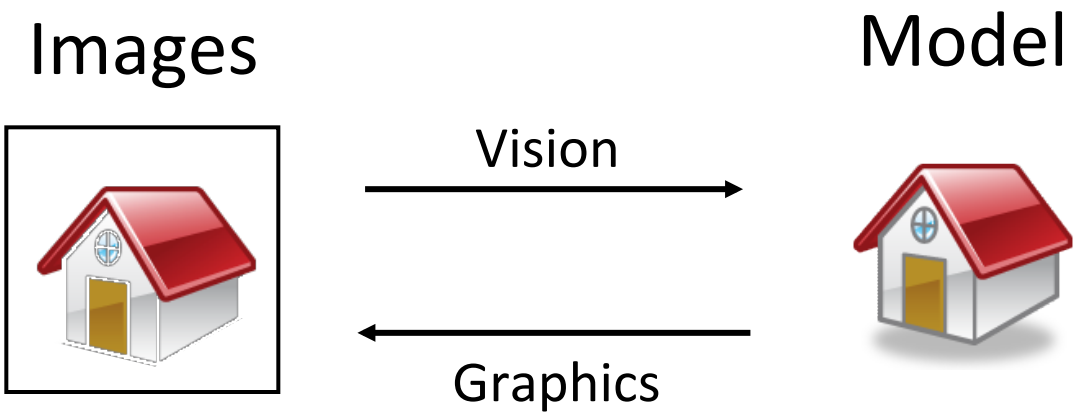
Vision for perception, interpretation



Related disciplines



Vision and graphics



Inverse problems: analysis and synthesis.

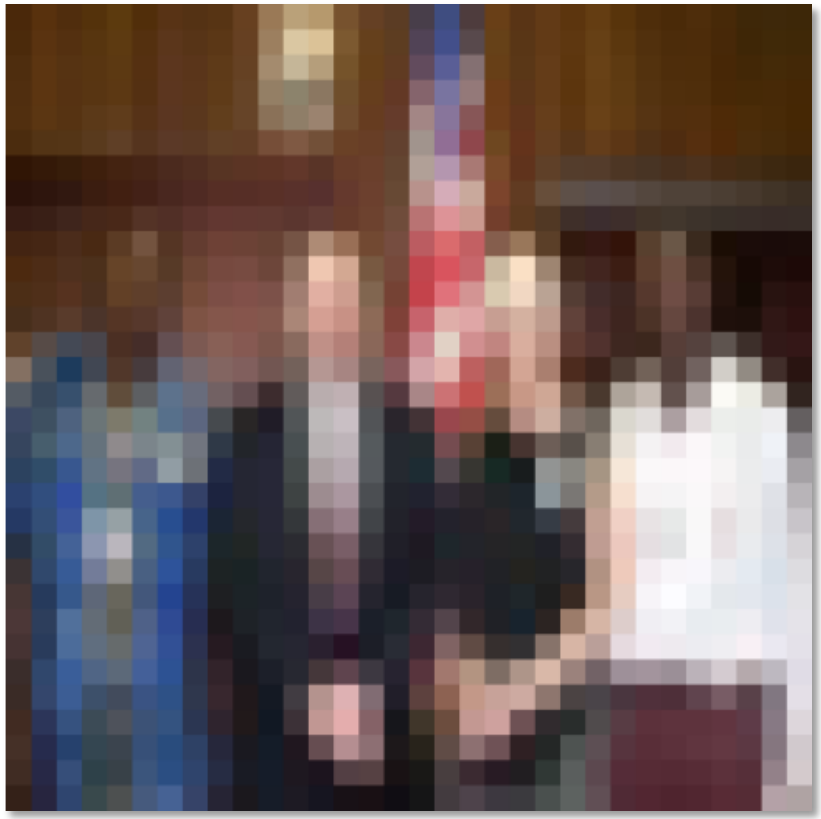
Can the computer match human perception?

- Yes and no (mainly no)
  - computers can be better at "easy" things
  - humans are much better at "hard" things
- But huge progress has been made
  - Especially in the last 10 years
  - What is considered "hard" keeps changing

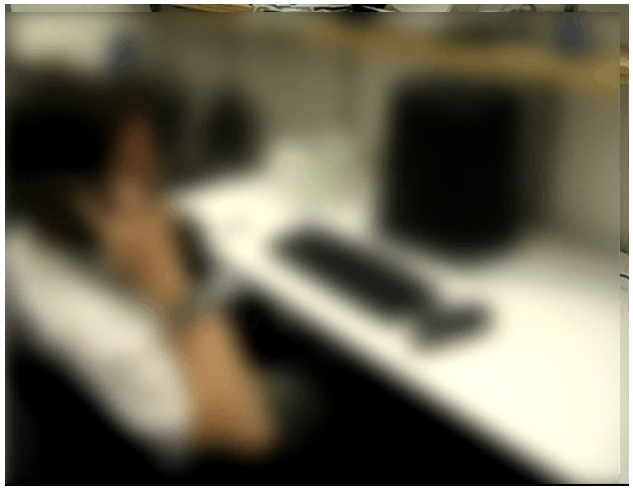
Human perception has its shortcomings

Sinha and Poggio, *Nature*, 1996

But humans can tell a lot about a scene from very little information…



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

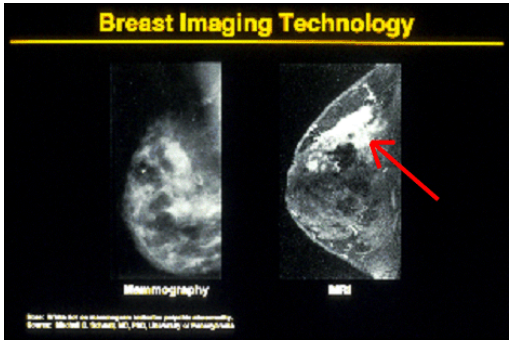


Why study computer vision?


Global Consumer Internet Traffic
(Petabyte usage per month)
2010-2015
40,000
35,000
30,000
25,000
20,000
15,000
10,000
5,000
-
4,672
8,079
12,146
17,583
24,357
33,620
2,393
3,113
4,146
5,325
6,769
8,592
308
659
905
1,251
1,736
49
68
95
133
187
290
2010
2011
2012
2013
2014
2015
Internet video
Web, email, and data
Video calling
Online gaming
MediaBuyerPlanner.com
Source: Cisco

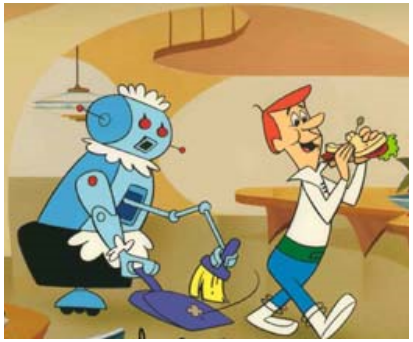
Computer vision matters

Safety



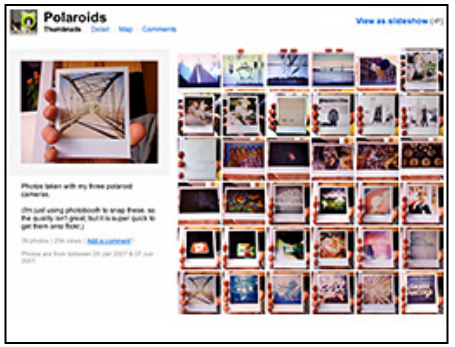
Health

Security



Comfort

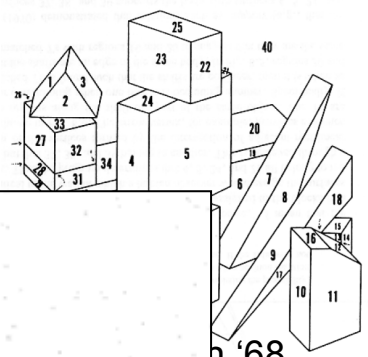
Fun



Access

Ridiculously brief history of computer vision

- 1966: Minsky assigns computer vision as an MIT undergrad summer project
- 1960
- 1970
- 1980
- 1990
- 2000
- 2030



n ‘68

anade ‘78

Turk and Pentland ‘91

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

PROJECT MAC

Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

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 particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

How vision is used now

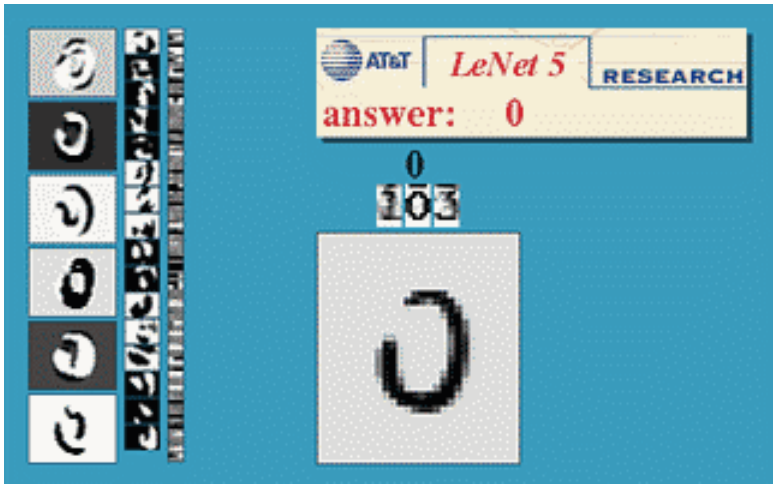
- Examples of state-of-the-art

Some of the following slides by Steve Seitz

Optical character recognition (OCR)

Technology to convert scanned docs to text

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



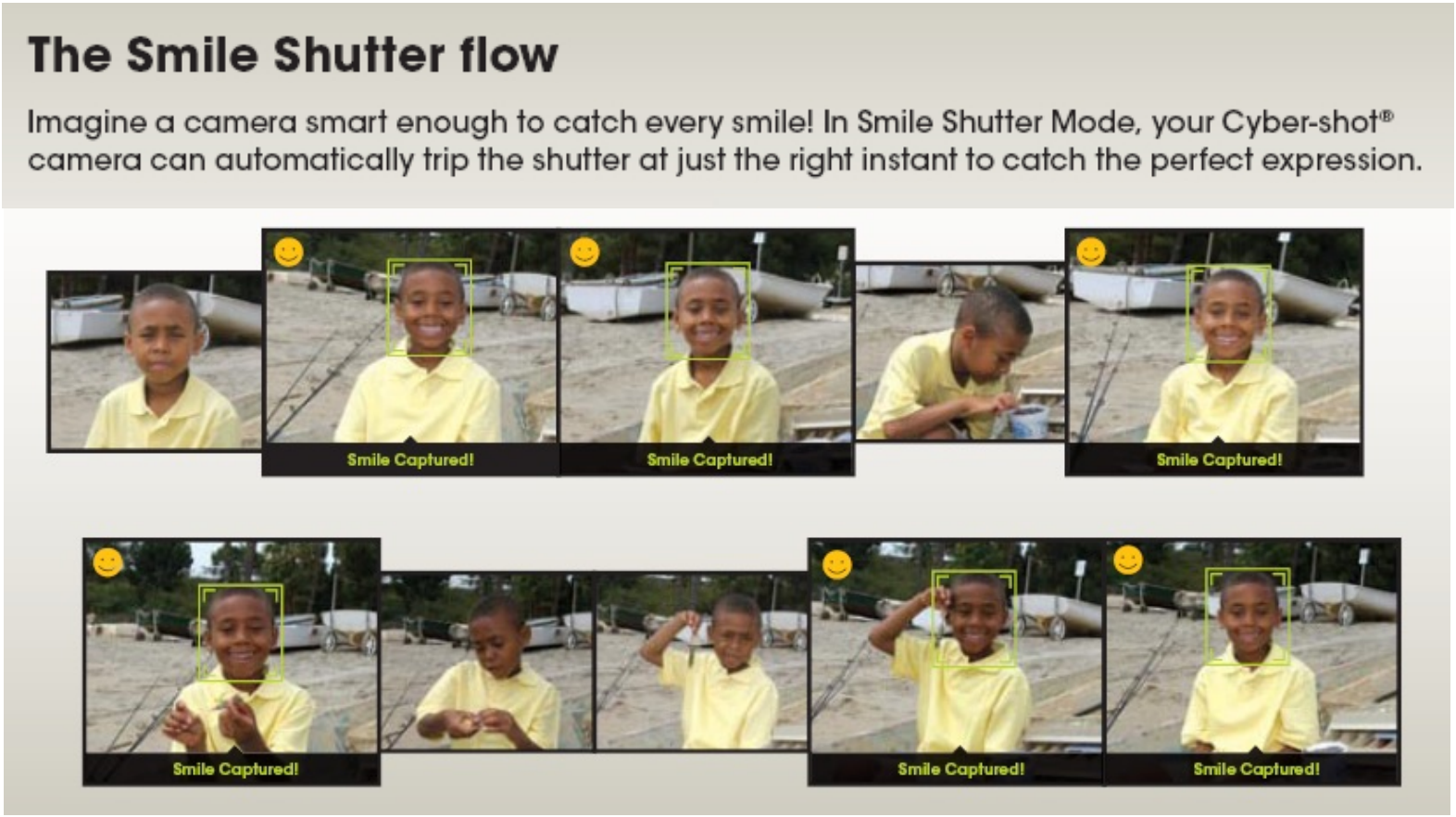
Digit recognition, AT&T labs
http://yann.lecun.com/exdb/lenet/

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

Face detection

- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, …

Smile detection



[Sony Cyber-shot® T70 Digital Still Camera]

3D from thousands of images

Building Rome in a Day, Agarwal et al. 2009

Indoor 3D Reconstruction

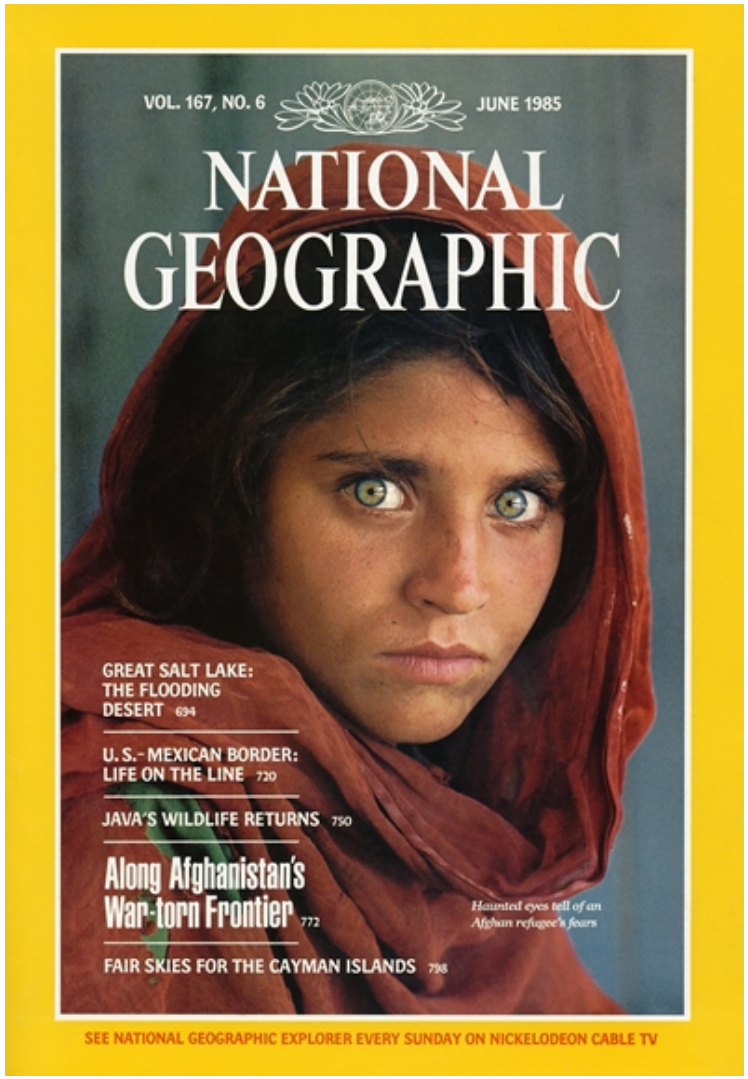
Reconstructing the World's Museums, Xiao & Furukawa, 2012

Object recognition (in supermarkets)

LaneHawk by EvolutionRobotics

“A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk, you are assured to get paid for it… “

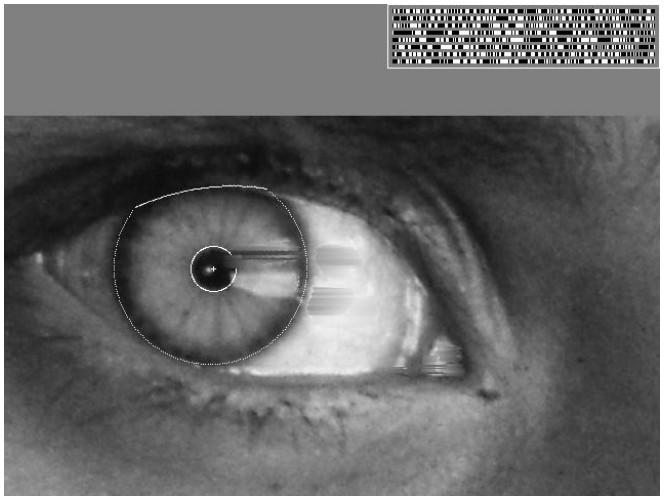
Face recognition

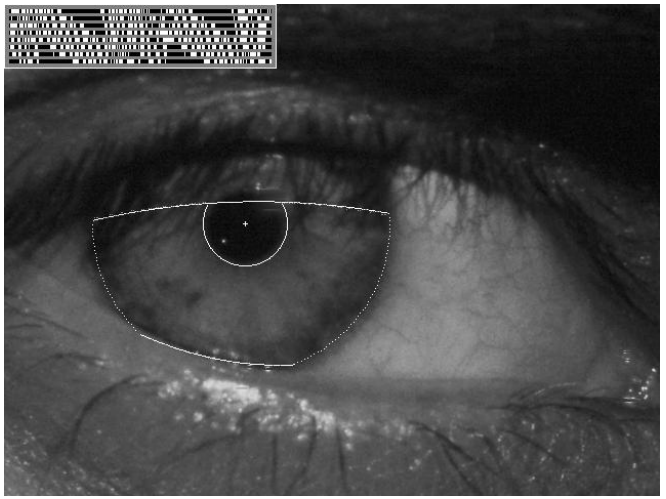


Who is she?

Vision-based biometrics

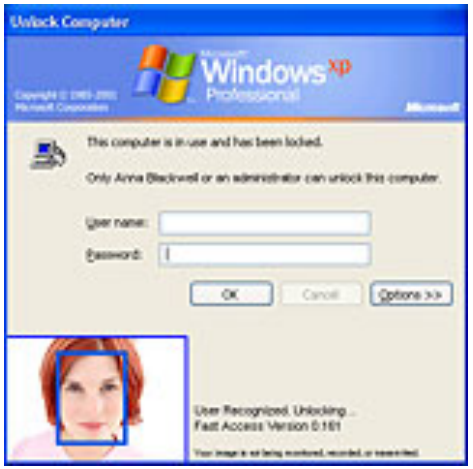
"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story





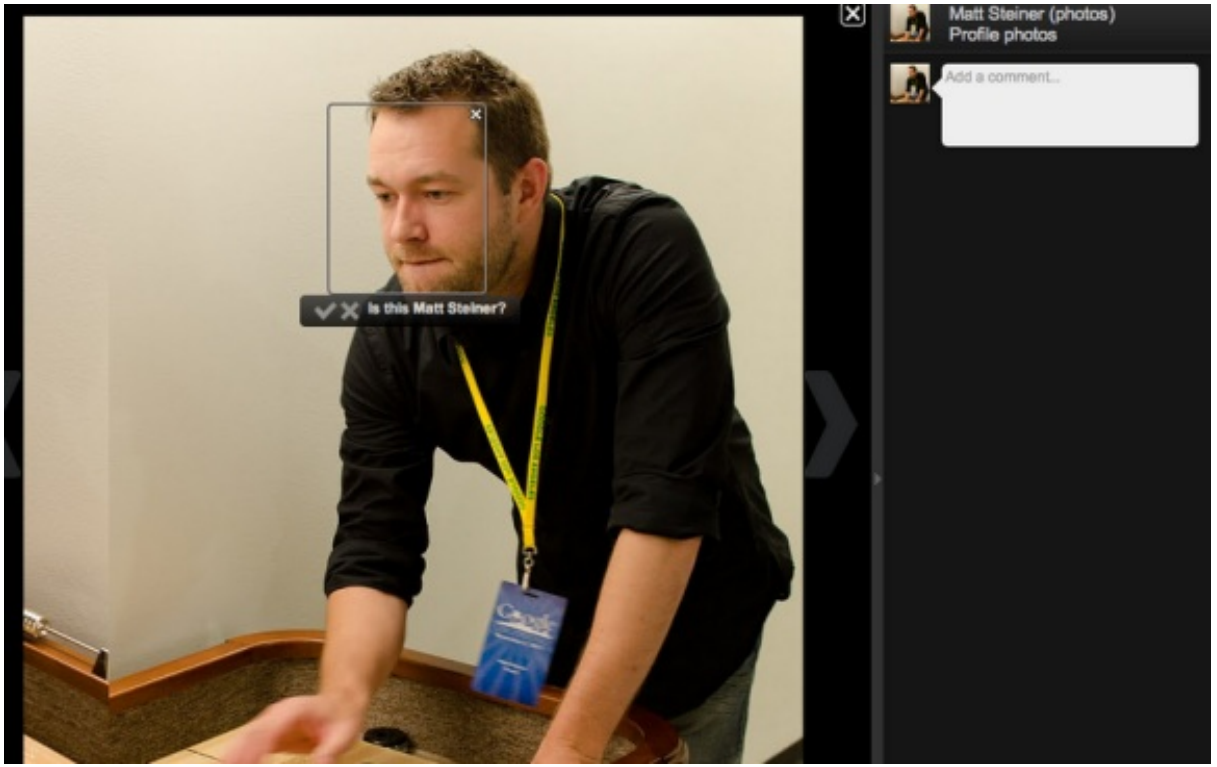
Login without a password…

Fingerprint scanners on many new laptops, other devices



Face recognition systems now beginning to appear more widely
http://www.sensiblevision.com/

Vision-based autotagging



Social networking websites have started automatically tagging faces

- Google Plus, Facebook, etc

Object recognition (in mobile phones)

[Point & Find](), [Nokia]()

[Google Goggles]()

Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Special effects: motion capture

Pirates of the Carribean, Industrial Light and Magic

Sports

Sportvision first down line
Nice explanation on www.howstuffworks.com

http://www.sportvision.com/video.html

Smart cars

- Mobileye
  - Vision systems currently in high-end BMW, GM, Volvo models
  - By 2010: 70% of car manufacturers.

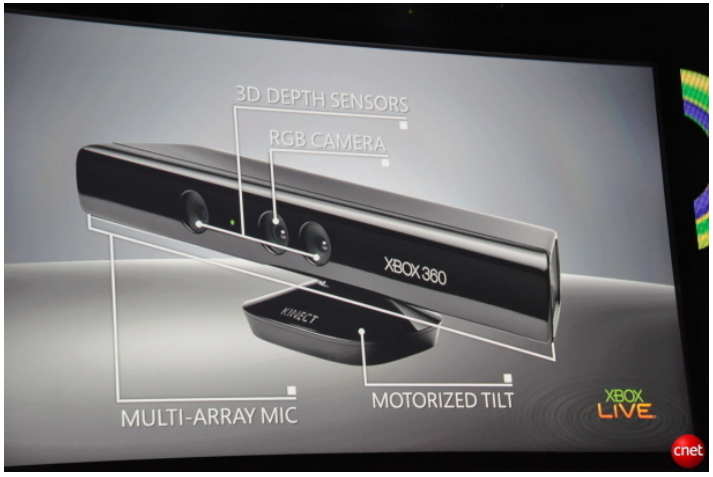
Slide content courtesy of Amnon Shashua

Google self-driving cars

Oct 9, 2010. "Google Cars Drive Themselves, in Traffic". *The New York Times*. John Markoff
June 24, 2011. "Nevada state law paves the way for driverless cars". *Financial Post*. Christine Dobby
Aug 9, 2011, "Human error blamed after Google's driverless car sparks five-vehicle crash". *The Star* (Toronto)

Interactive Games: Kinect

- Object Recognition: http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o
- Mario: http://www.youtube.com/watch?v=8CTJL5lUjHg
- 3D: http://www.youtube.com/watch?v=7QrnwoO1-8A
- Robot: http://www.youtube.com/watch?v=w8BmgtMKFbY



Vision in space

NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

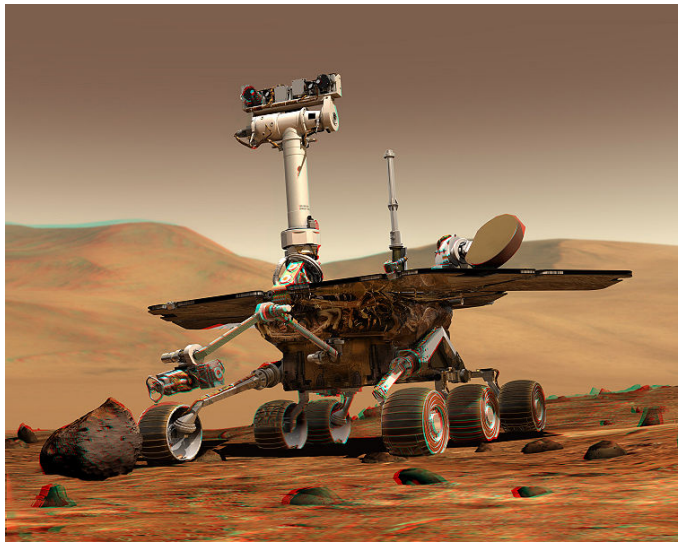
Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

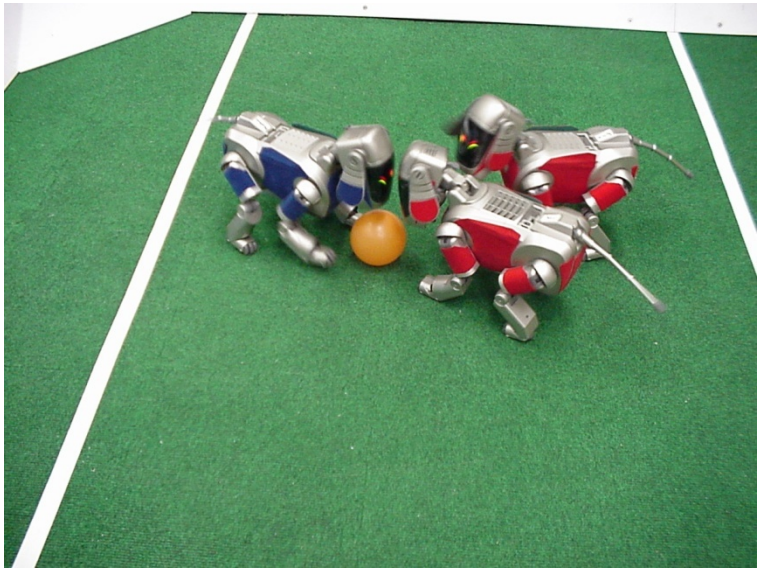
Industrial robots

Vision-guided robots position nut runners on wheels

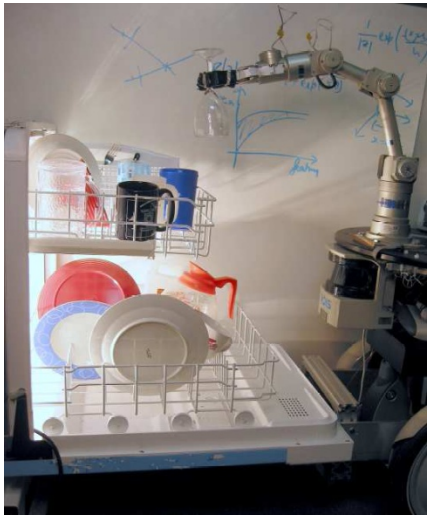
Mobile robots



NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover

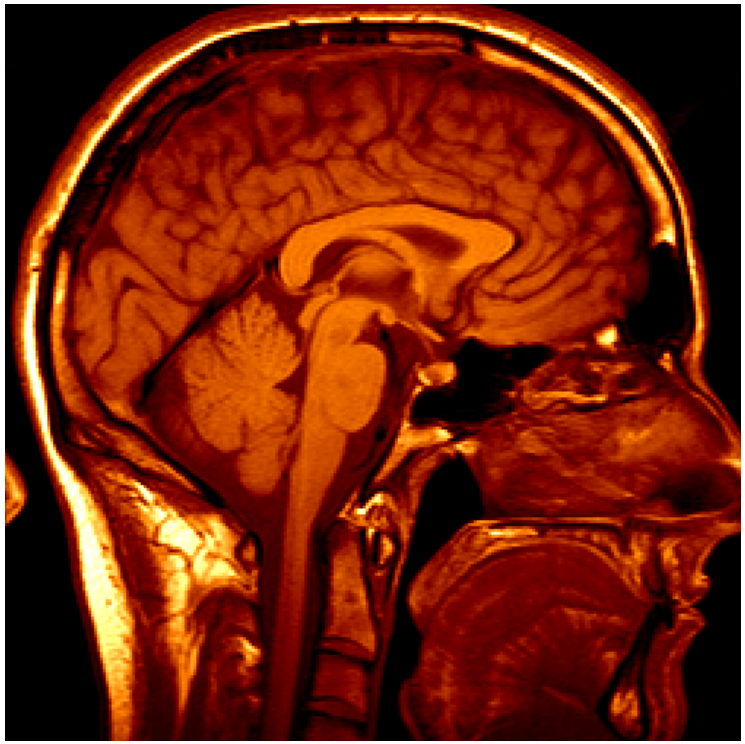


http://www.robocup.org/



Saxena et al. 2008
STAIR at Stanford

Medical imaging



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

Current state of the art

- You just saw examples of current systems.
  - Many of these are less than 5 years old

- This is a very active research area, and rapidly changing
  - Many new apps in the next 5 years

- To learn more about vision applications and companies
  - David Lowe maintains an excellent overview of vision companies: http://www.cs.ubc.ca/spider/lowe/vision.html

Why is computer vision difficult?

Viewpoint variation

Illumination

Scale

Why is computer vision difficult?

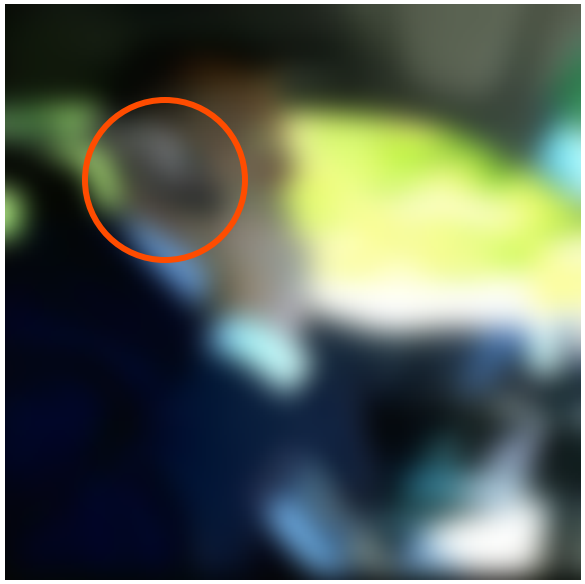
Intra-class variation

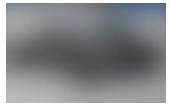
Motion (Source: S. Lazebnik)

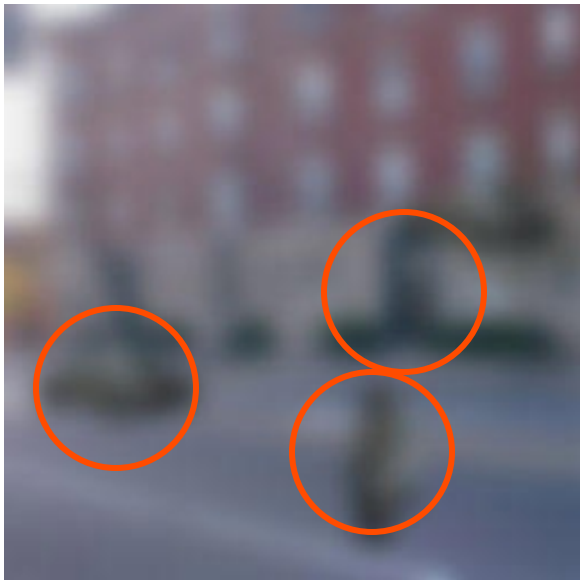
Background clutter

Occlusion

Challenges: local ambiguity

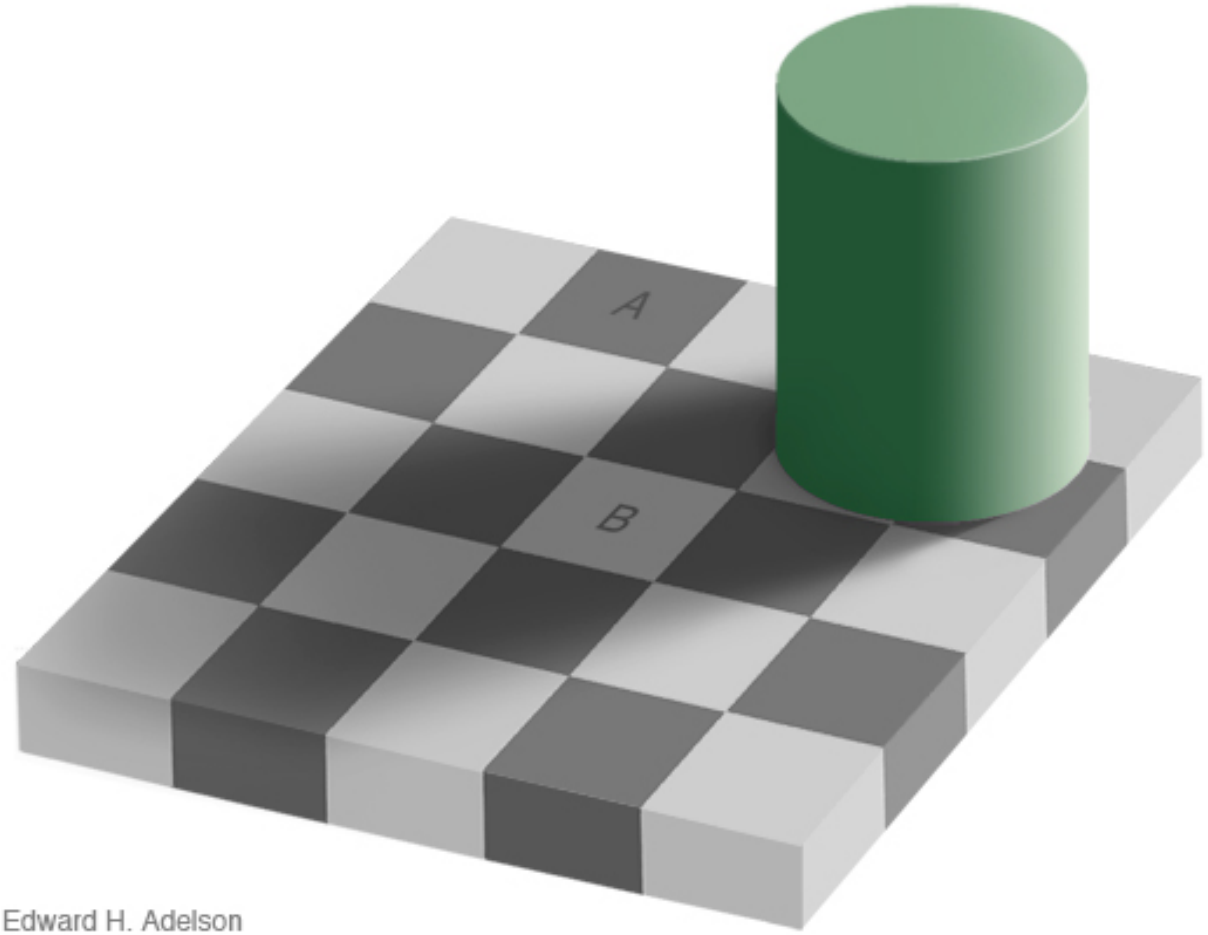






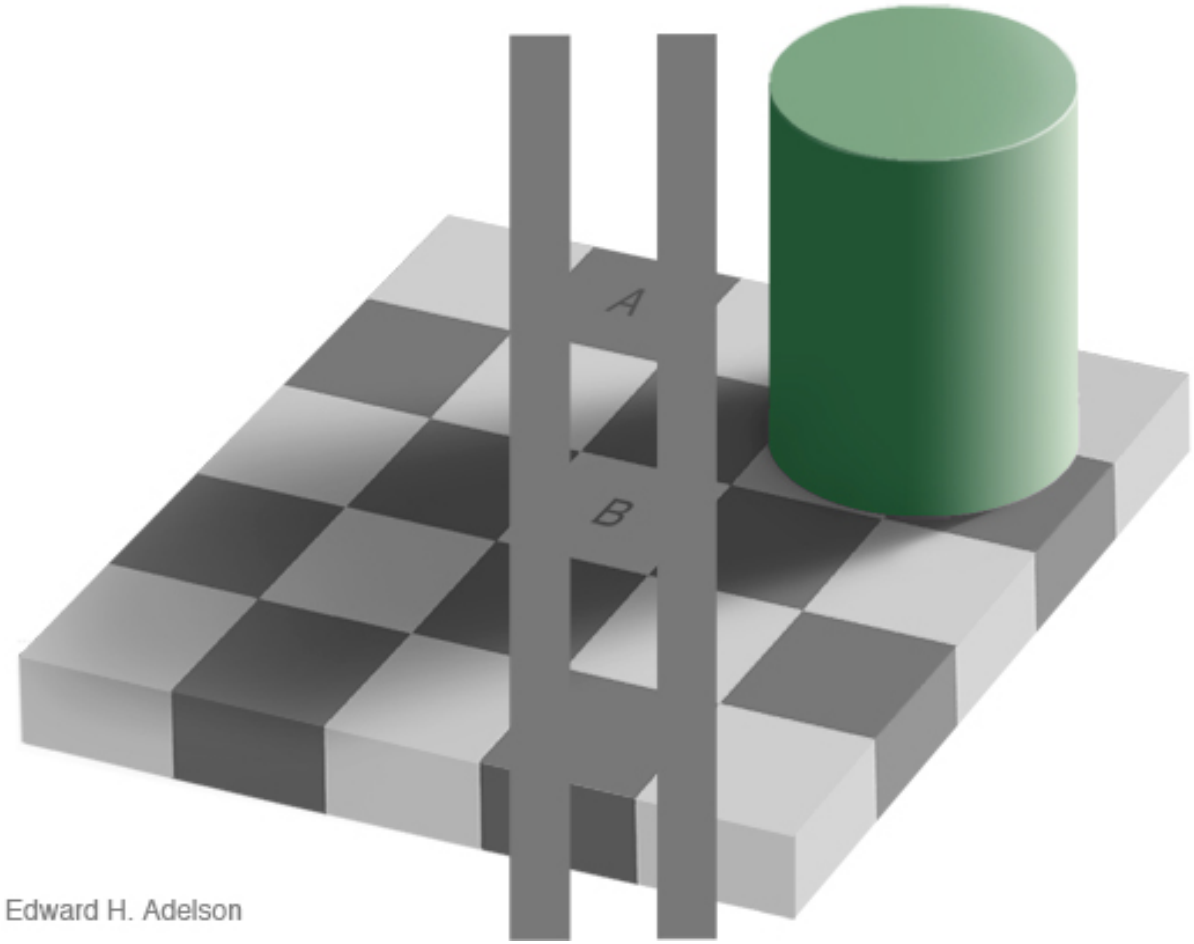
slide credit: Fei-Fei, Fergus & Torralba

Challenges: computers don’t see what we see



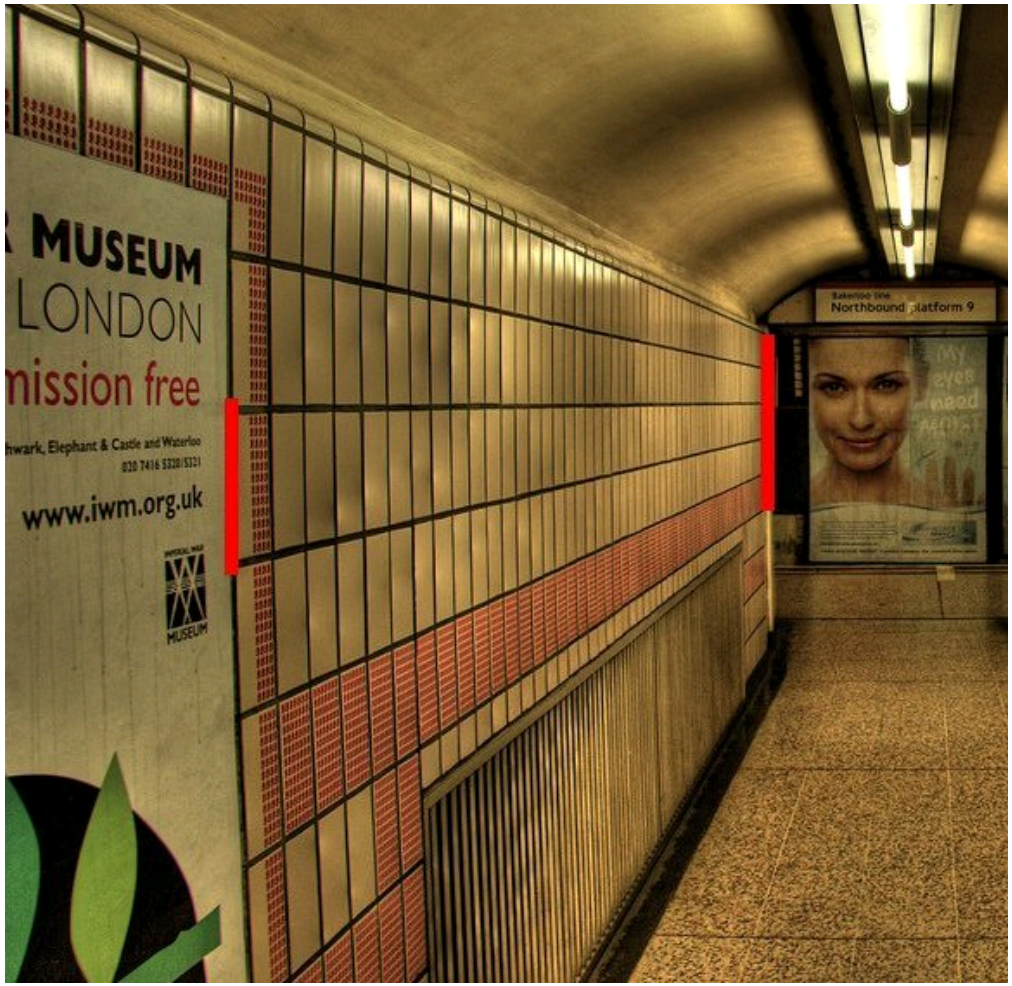
Checker Shadow Illusion – [E. H. Adelson]

Challenges: computers don’t see what we see



Checker Shadow Illusion – [E. H. Adelson]

Challenges:
computers don’t see what we see



Challenges:
computers don’t see what we see

But there are lots of cues we can exploit…

Source: S. Lazebnik

Bottom line

- Perception is an inherently ambiguous problem
  - Many different 3D scenes could have given rise to a particular 2D picture

  - We often need to use prior knowledge about the structure of the world

Object recognition

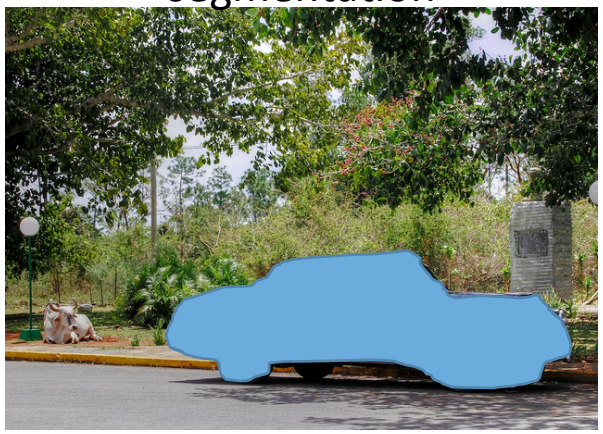
classification

is there a car in this image?

detection

is there a car in this image, and if so, where?

segmentation



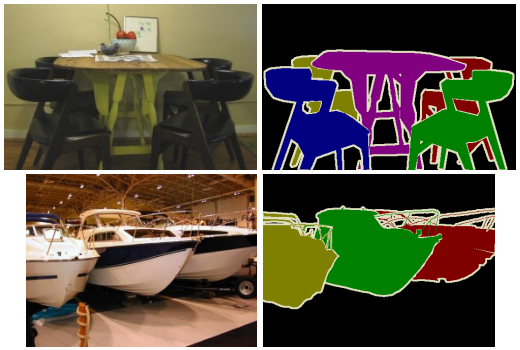
which pixels in this image correspond to a car?

Object recognition datasets

31K images, 256 objects
classification only

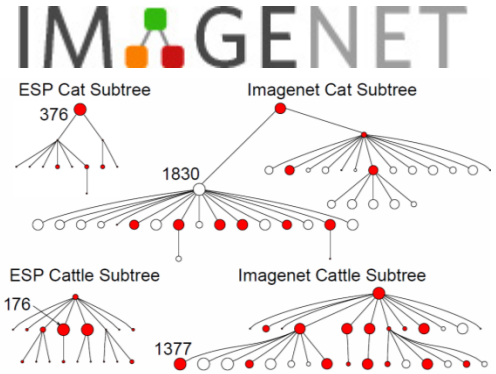
[Caltech - 256](), 2007

20K images, 20 objects
classification, detection, segmentation



[PASCAL VOC](), 2005-2012

15M images, 21K objects
classification, detection

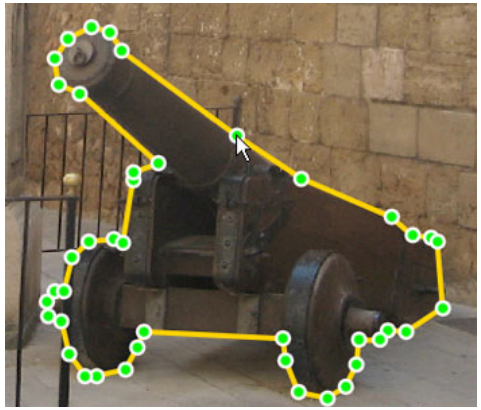


[ImageNet](), 2009-2013

[Caltech - 101](), 2004

6K images, 101 objects
classification, detection

[LabelMe](), 2008



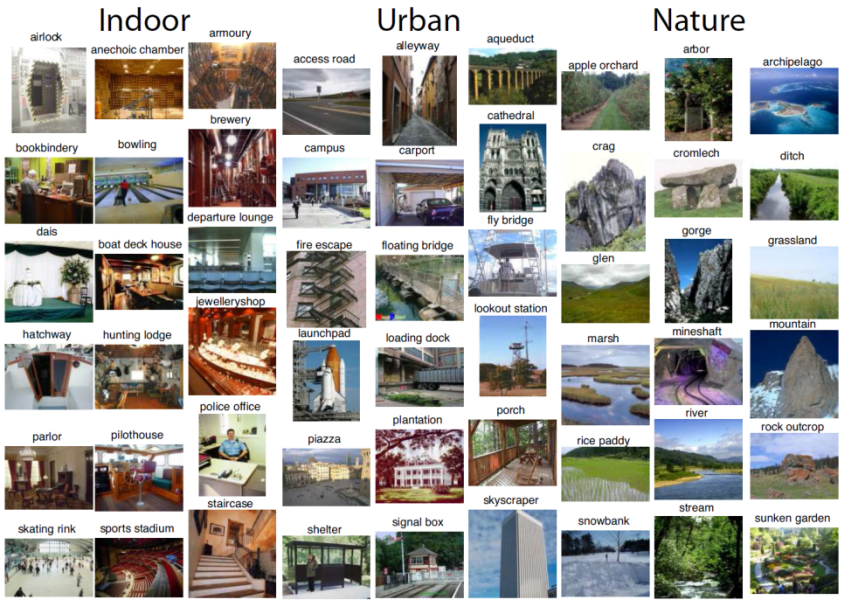
208K images, 4K objects
classification, detection, segmentation

[SUN](), 2010

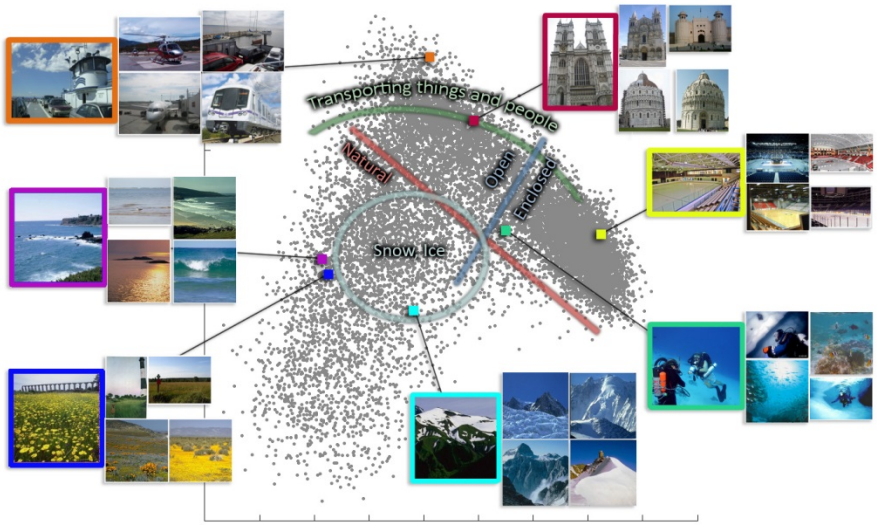
131K images, 3819 objects
classification, detection, segmentation

Scene understanding

Scene classification



Scene attributes



[SUN Attribute Database](#)
14K images, 102 attributes

Datasets

- [SUN Database](#)

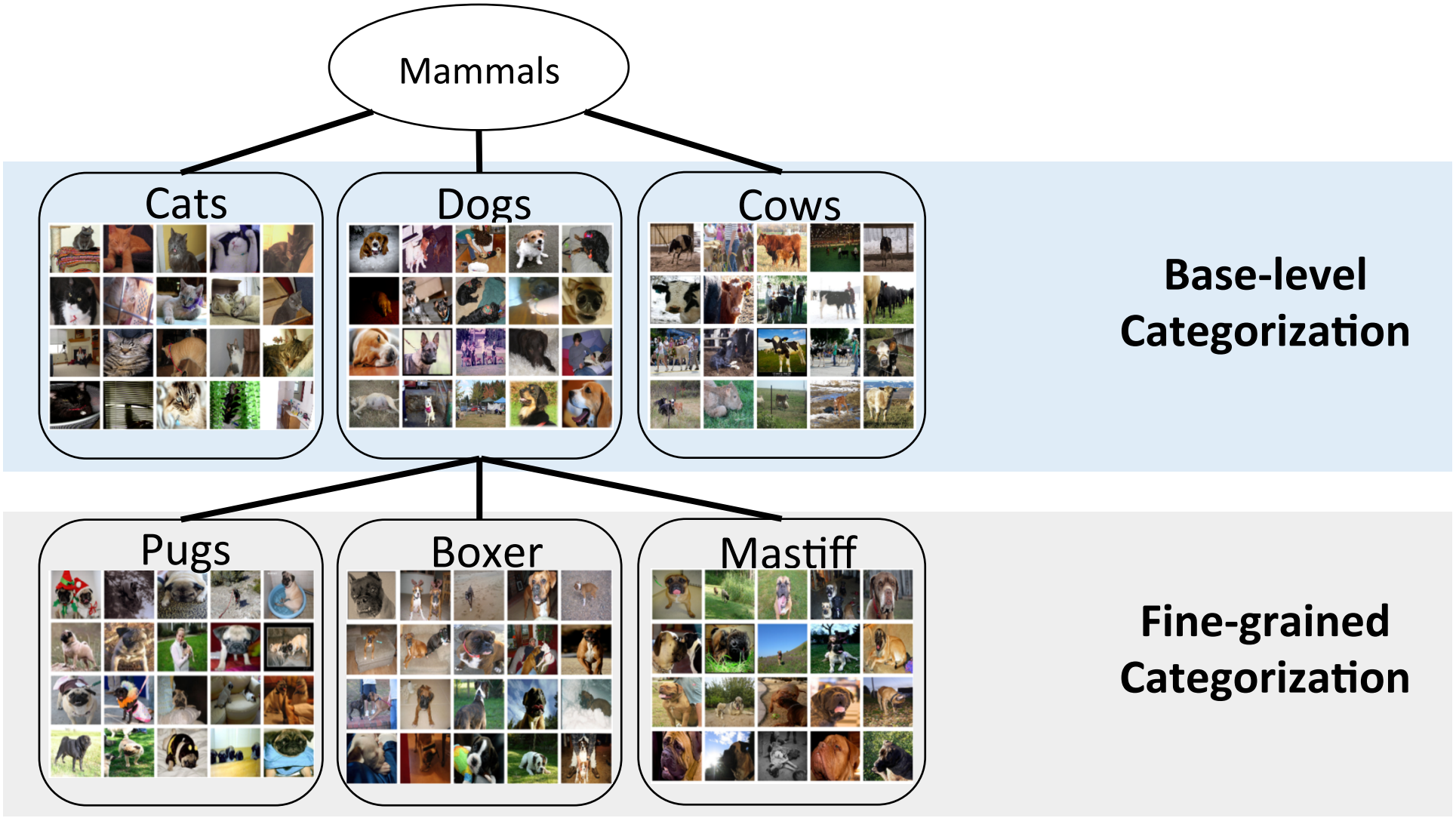
131K images, 908 scene categories

- [15 scene dataset](#)

Algorithms:

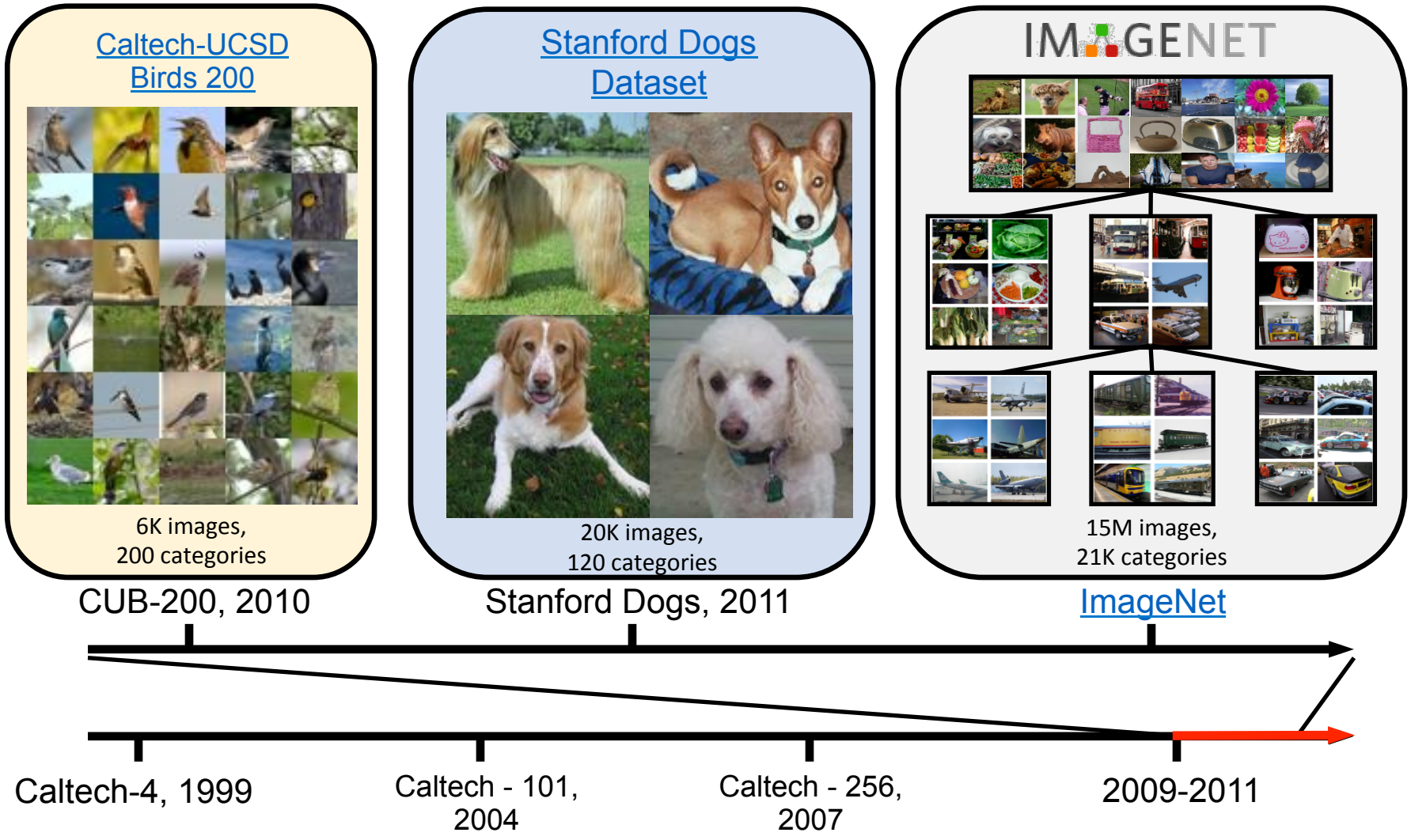
- [Spatial pyramid matching](#)
- [ObjectBank](#)

Fine-grained Categorization



[Rosch, 1979]

Recently Emerging Fine-Grained Datasets

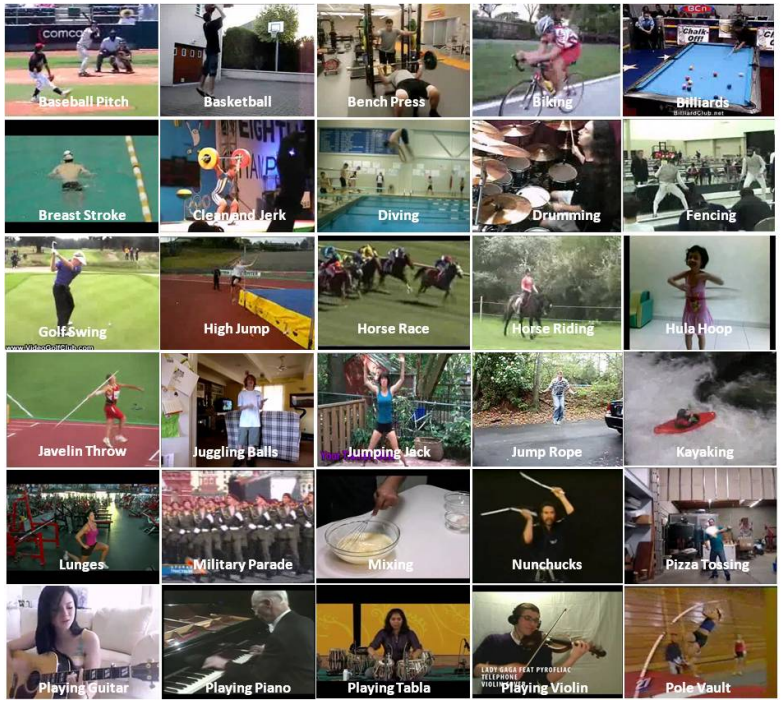


Action recognition

still images

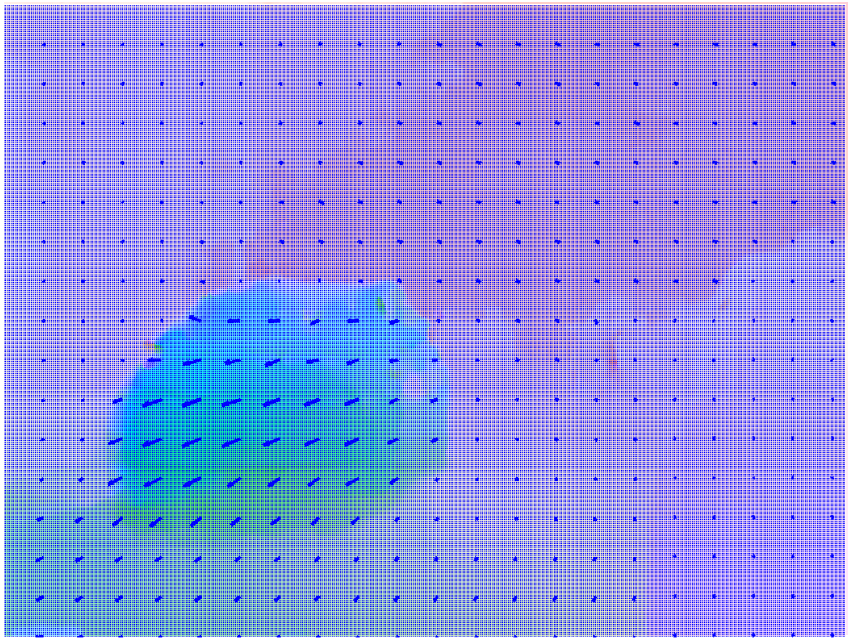
- Stanford 40 Actions
- People-playing-Musical Instruments (24 classes)
- PASCAL VOC (10 classes)

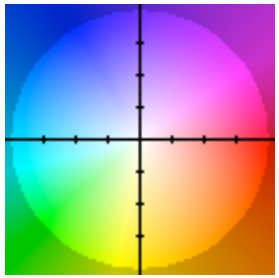
videos



- UCF50 Action recognition dataset
- Hollywood2 dataset (12 classes)
- KTH dataset (6 classes)

Optical flow





Optical flow

Datasets

- Middlebury dataset
- MPI Sintel dataset

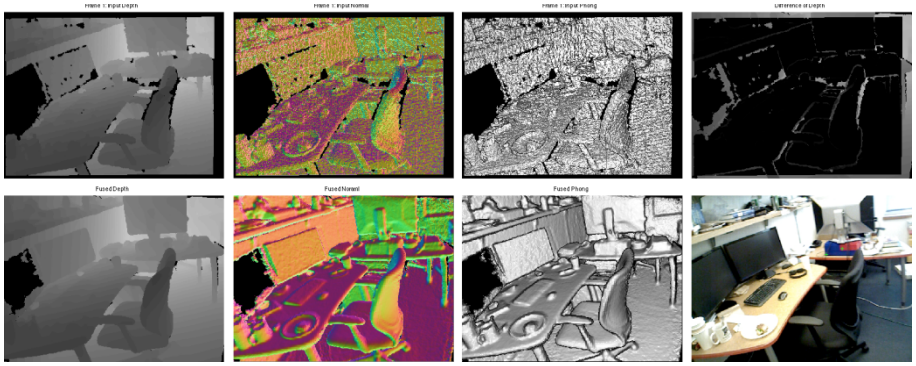
Algorithms

- Lucas-Kanade and Horn-Schunck (PMT)
- `Secrets of optical flow'
- Motion detail preserving flow

3D Reconstruction/SfM

without kinect

with kinect



- CMVS
- Bundler
- MeshLab

- KinectFusion
- SiftFu
- ReconstructMe

Questions?