Computer Vision

Jana Kosecka
http://cs.gmu.edu/~kosecka/cs482/
kosecka@gmu.edu

Some slides thanks to S. Lazebnik, Fei-Fei Li,H. Farid, K. Grauman and others

Topics of the class

- Image formation process
- Image processing techniques for color and gray level images: edge detection, corner detection, segmentation
- Video processing, motion computation and 3D vision and geometry
- Basics of image classification, object detection and recognition
- Implement basic vision algorithms in Python/OpenCV (open source computer vision library)

Logistics

- **Grading:** Homeworks 50%,
- Exam 30% Final project: 20%
- **Prerequisites:** linear algebra, calculus, probability and statistics
- **Lectures:** Introduction by an instructor, homeworks every two weeks
- **Projects:** up to teams of 2 people
- **Dates**
  - Project proposals due March 24th
  - May week of finals final report due
  - Project presentations

Visual Perception

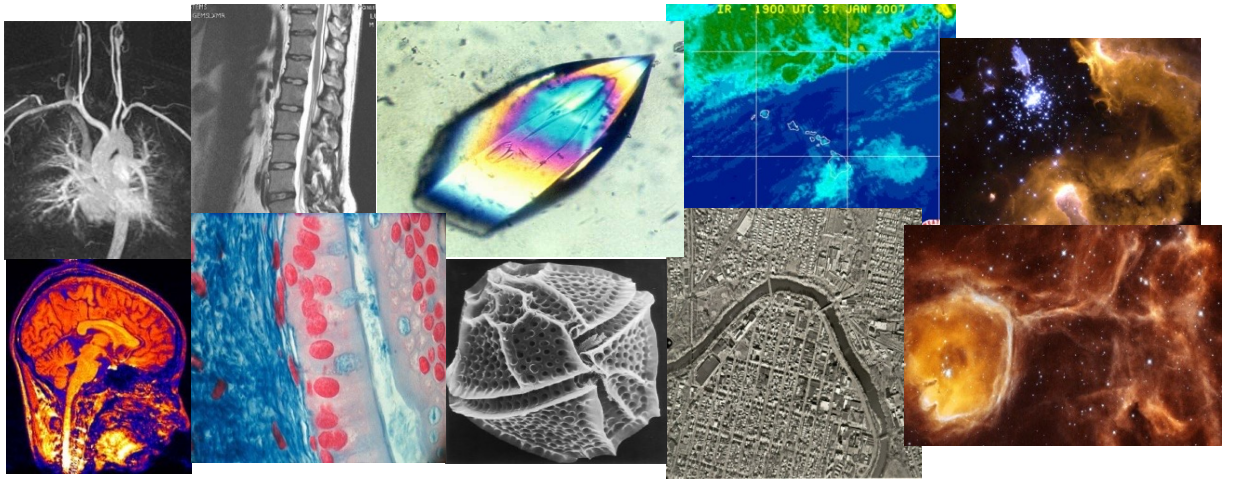
- There are 1.8 billion images uploaded to Internet every day
- Every autonomous car, delivery robot, laptop and phone is equipped with cameras
- The opportunities and challenges of visual perception

Why study computer vision?

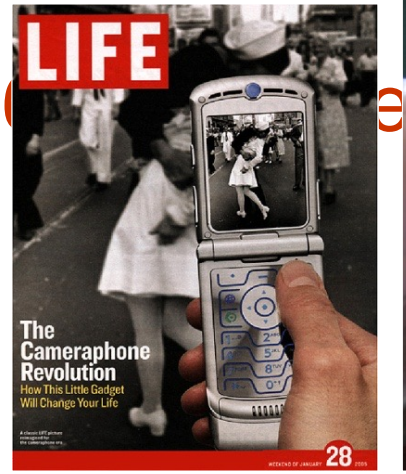
- Vision is useful: Images and video are everywhere!

Google Image Search™ Picasa™ flickr™ webshots™ picsearch™ YouTube Broadcast Yourself™

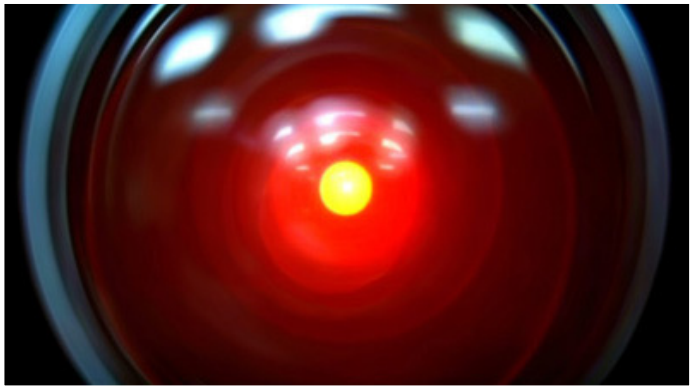
Surveillance and security



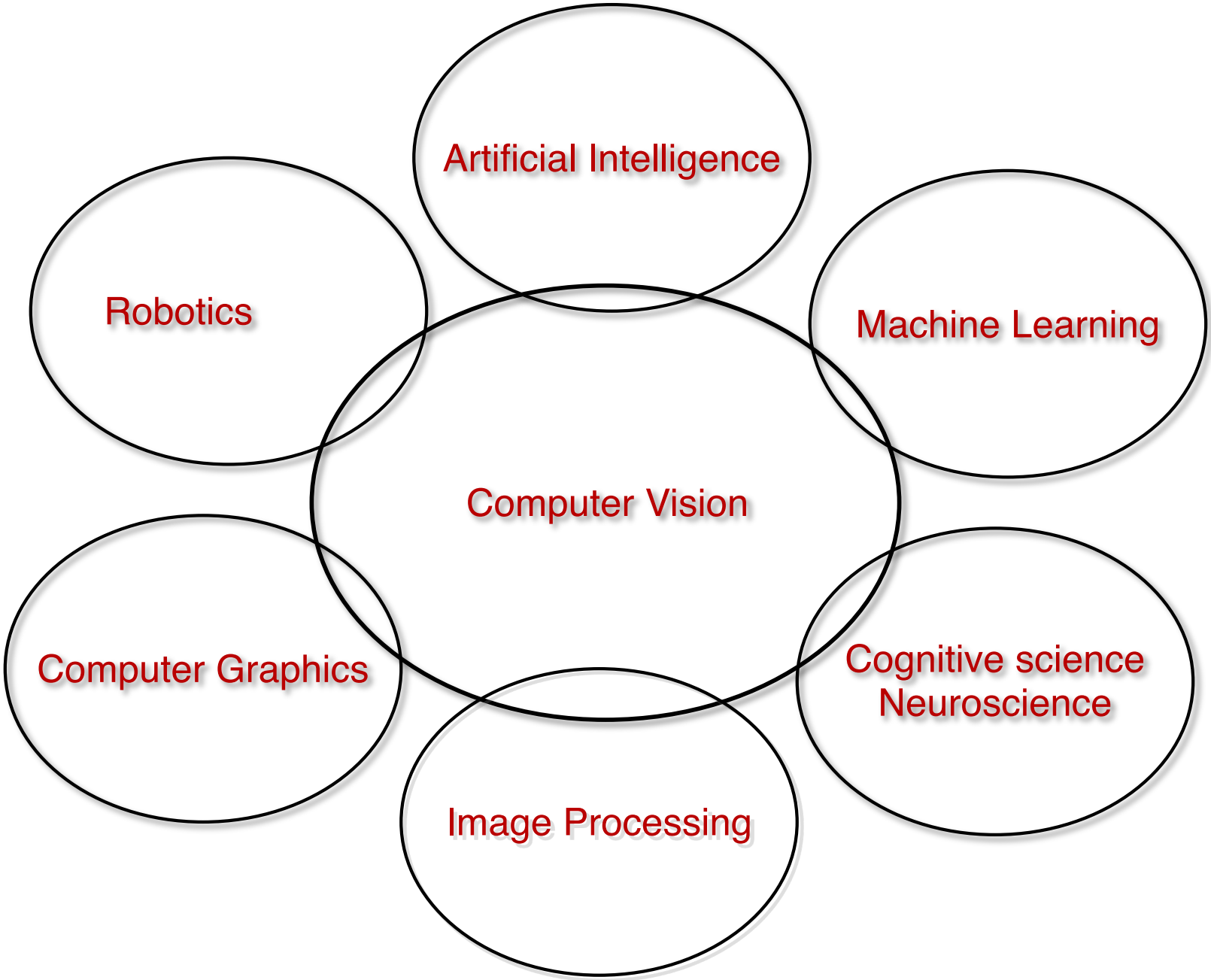
Medical and scientific images


LIFE
The Cameraphone Revolution
How This Little Gadget Will Change Your Life
28


Detector Camera



Connections to other disciplines



The goal of computer vision

- To extract “meaning” from pixels

What we see

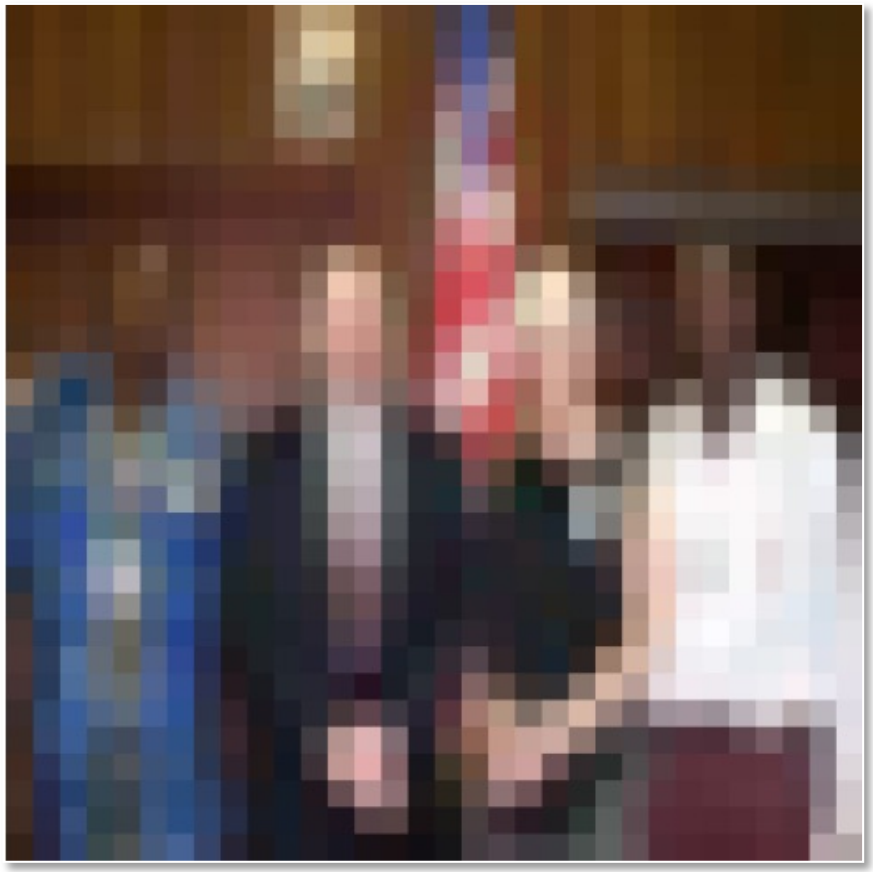
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

Source: S. Narasimhan

The goal of computer vision

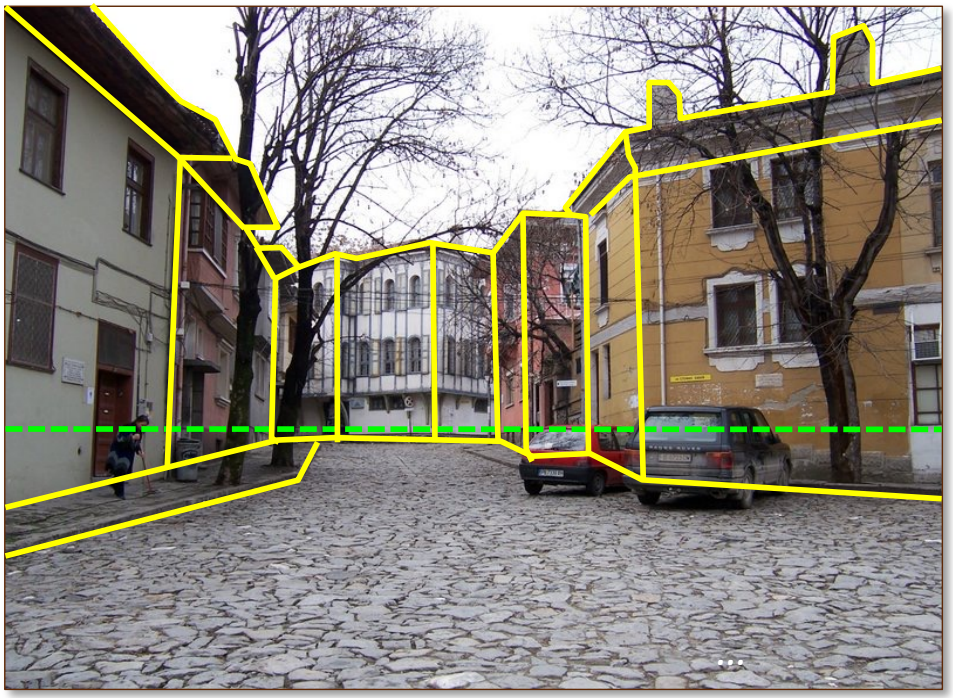
- To extract “meaning” from pixels



Humans are remarkably good at this…

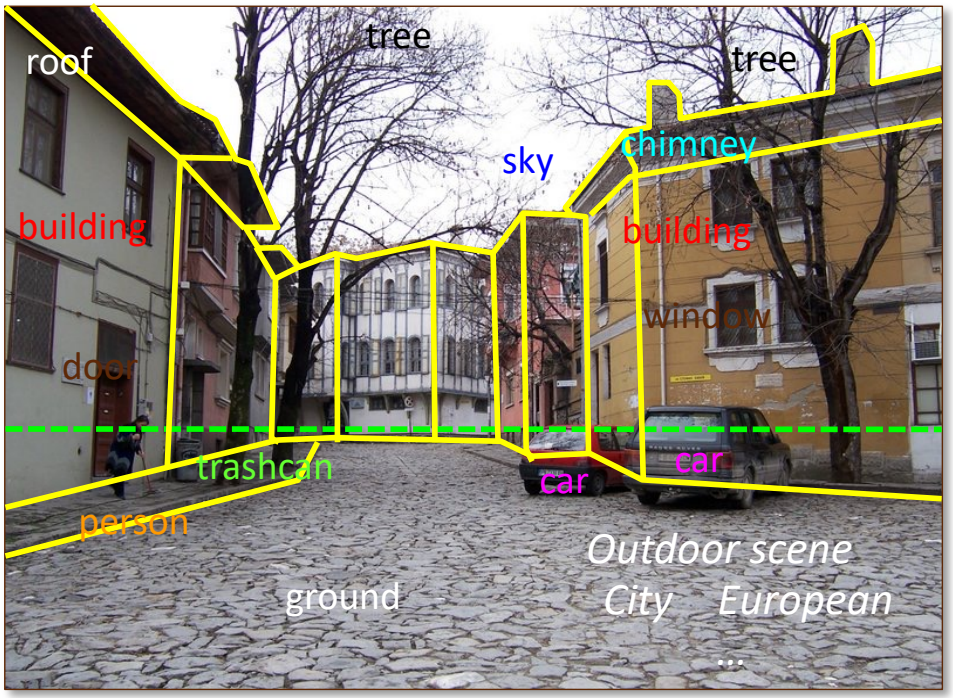
Source: “80 million tiny images” by Torralba et al.

What kind of information can be extracted from an image?



Geometric information

What kind of information can be extracted from an image?

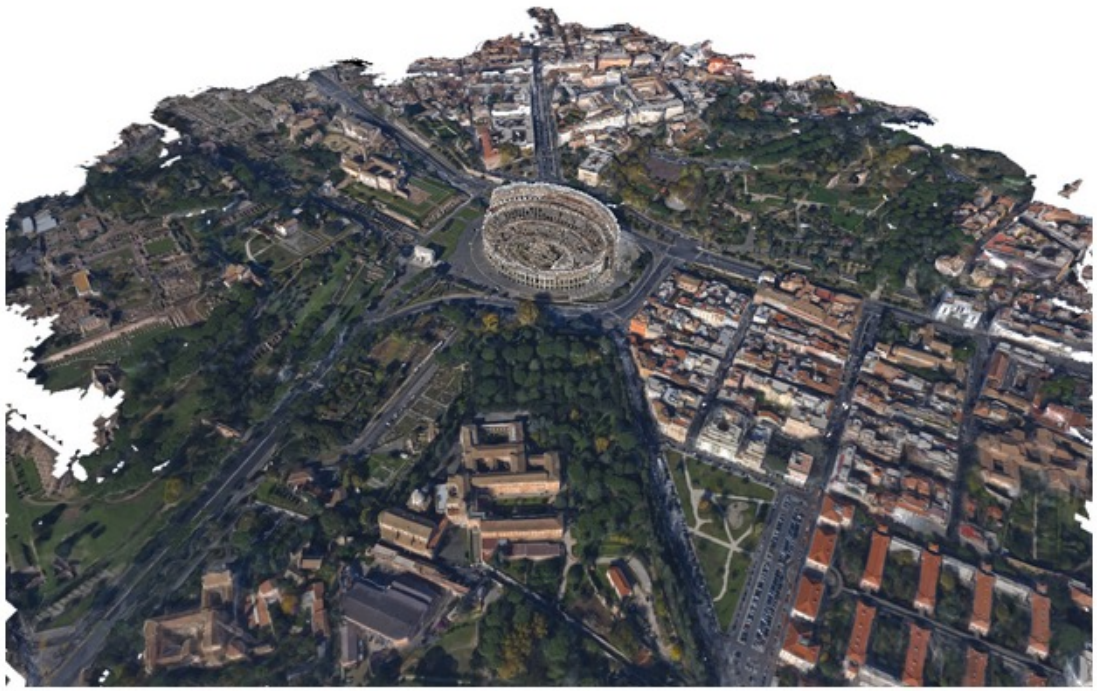


Geometric information

Semantic information

Reconstruction: 3D from photo collections

Colosseum, Rome, Italy



San Marco Square, Venice, Italy

Q. Shan, R. Adams, B. Curless, Y. Furukawa, and S. Seitz, The Visual Turing Test for Scene Reconstruction, 3DV 2013

YouTube Video

Reconstruction: 4D from depth cameras

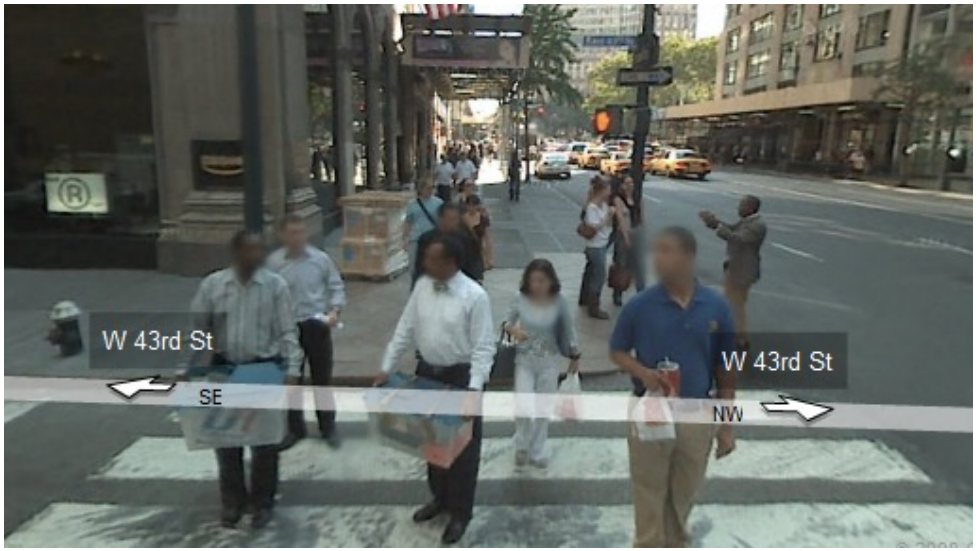
Figure 1: Real-time reconstructions of a moving scene with DynamicFusion; both the person and the camera are moving. The initially noisy and incomplete model is progressively denoised and completed over time (left to right).

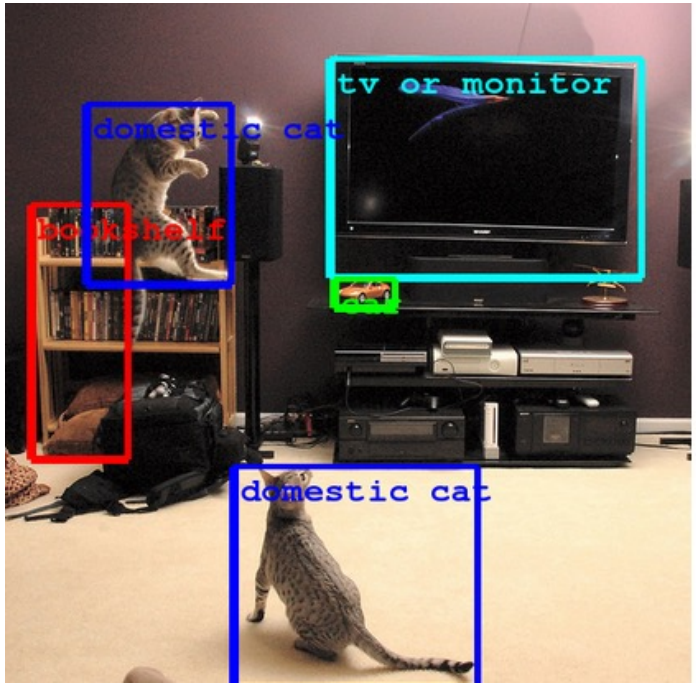
R. Newcombe, D. Fox, and S. Seitz, DynamicFusion: Reconstruction and Tracking of Non-rigid Scenes in Real-Time, CVPR 2015

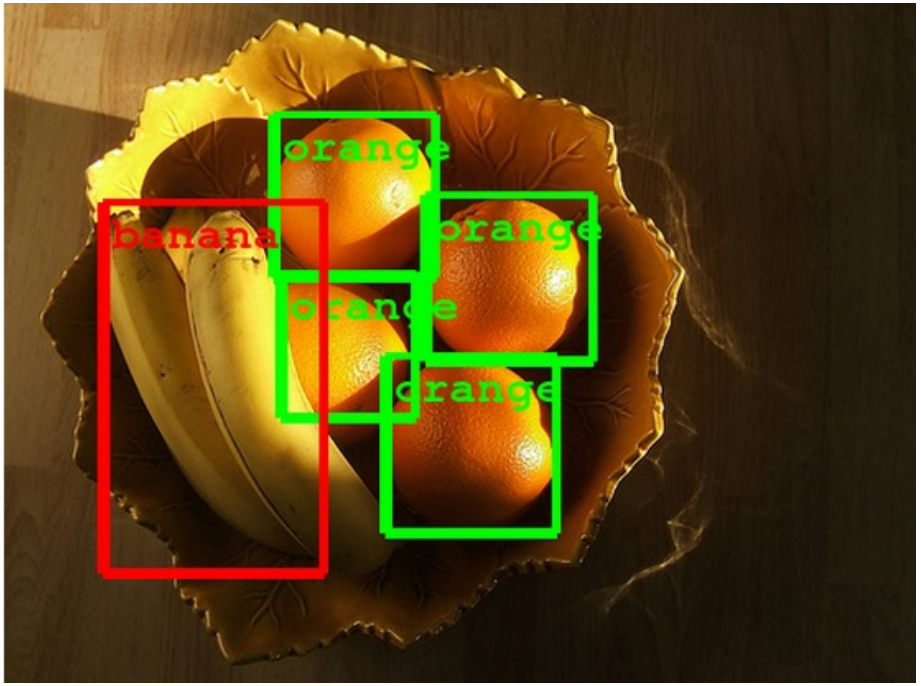
YouTube Video

Recognition: “Simple” patterns

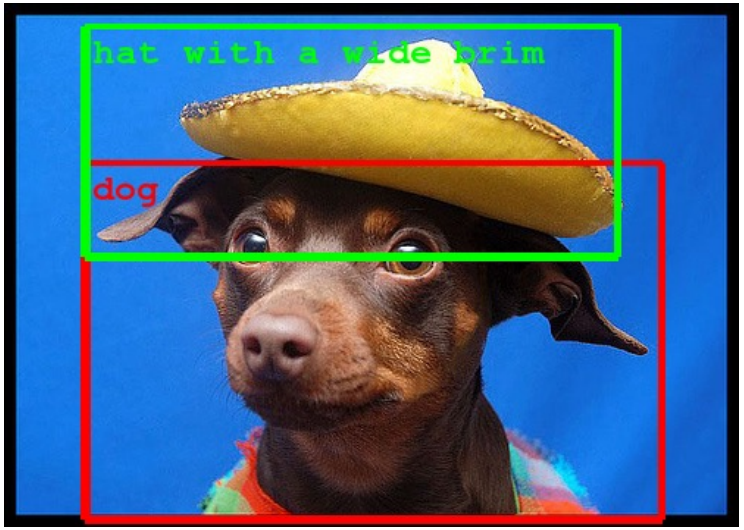
Recognition: Faces





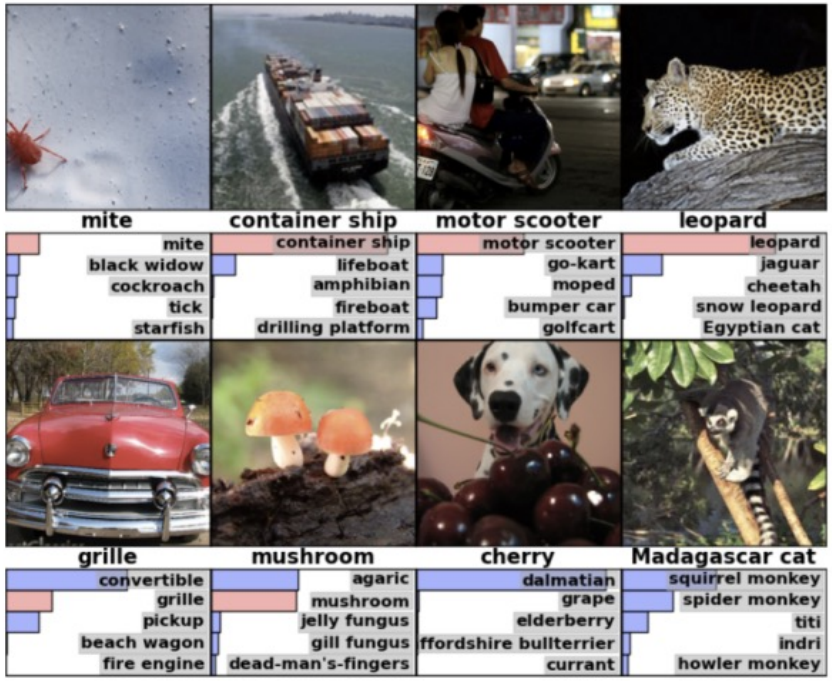


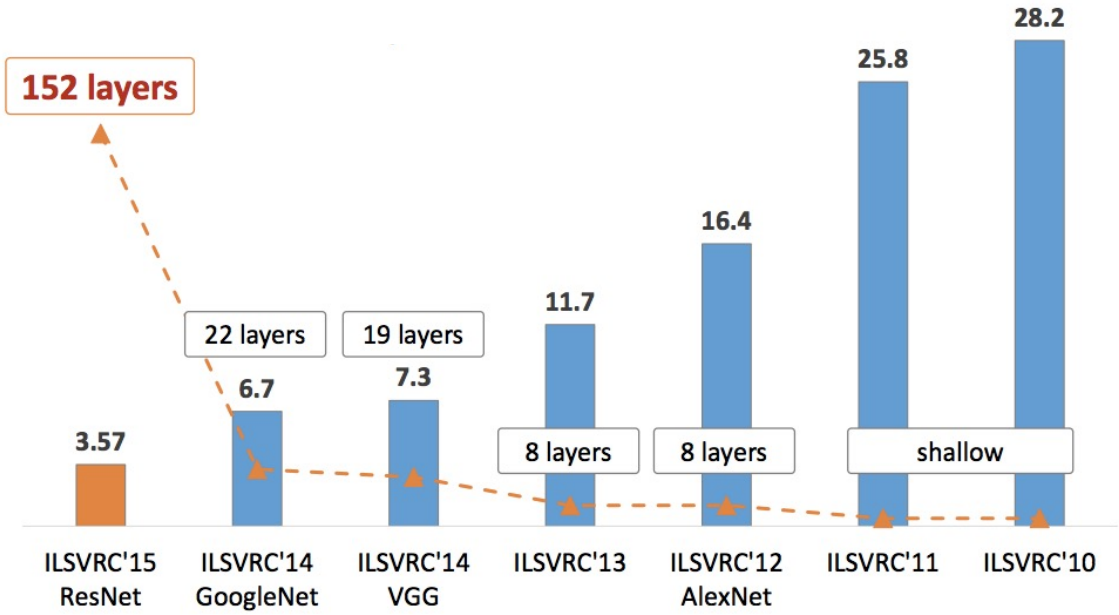
- Computer Eyesight Gets a Lot More Accurate, NY Times Bits blog, August 18, 2014
- Building A Deeper Understanding of Images, Google Research Blog, September 5, 2014



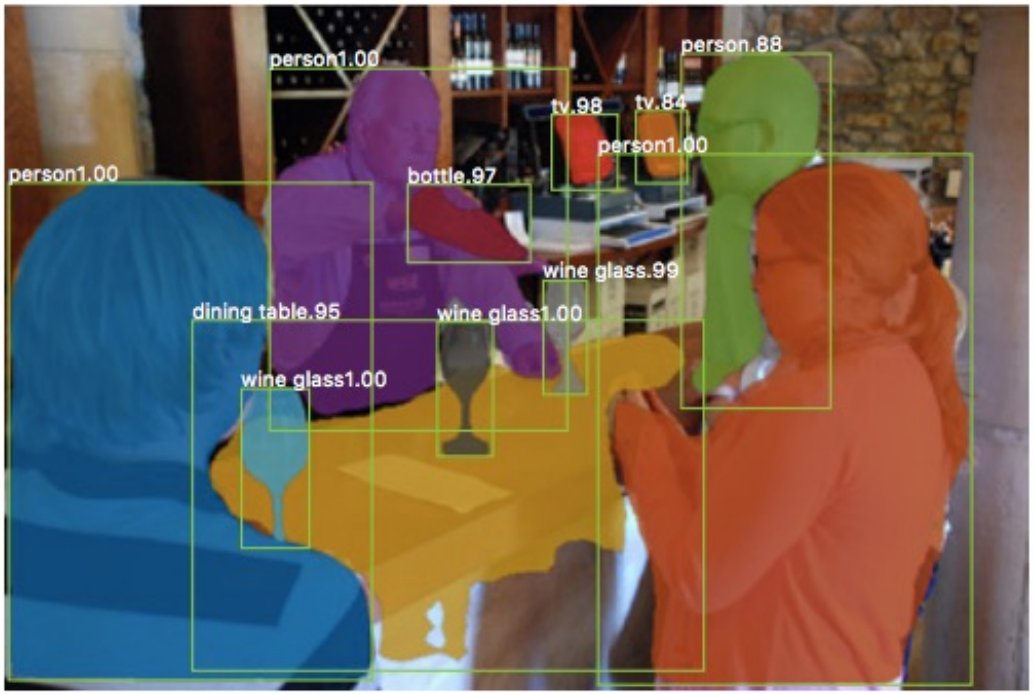
Recognition: General categories

- ImageNet challenge





Object detection, instance segmentation



K. He, G. Gkioxari, P. Dollar, and R. Girshick, Mask R-CNN,
ICCV 2017 (Best Paper Award)

Image generation

- BigGAN: 512 x 512 resolution, ImageNet

A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis, arXiv 2018

Image generation

- BigGAN: 512 x 512 resolution, ImageNet

Easy classes

Difficult classes

A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis, arXiv 2018

DeepFakes

- *"A quiet wager has taken hold among researchers who study artificial intelligence techniques and the societal impacts of such technologies. They're betting whether or not someone will create a so-called Deepfake video about a political candidate that receives more than 2 million views before getting debunked by the end of 2018"* – IEEE Spectrum, 6/22/2018

DeepFakes

DEPT. OF TECHNOLOGY NOVEMBER 12, 2018 ISSUE

THE NEW YORKER

IN THE AGE OF A.I., IS SEEING STILL BELIEVING?

Advances in digital imagery could deepen the fake-news crisis—or help us get out of it.

By Joshua Rothman

As synthetic media spreads, even real images will invite skepticism.

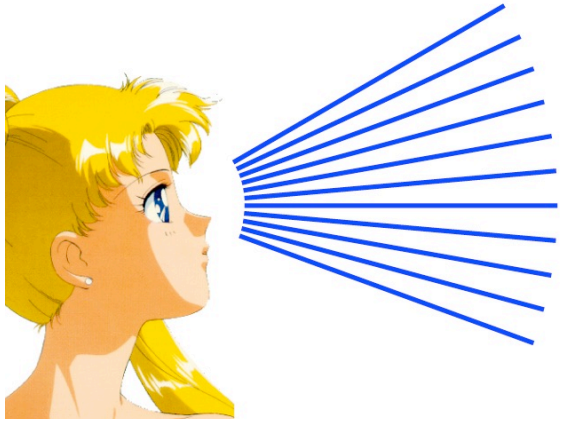
Illustration by Javier Jaén; photograph by Svetikd / Getty

https://www.newyorker.com/magazine/2018/11/12/in-the-age-of-ai-is-seeing-still-believing

Course overview

I. Early vision: Image formation and processing
II. Mid-level vision: Grouping and fitting
III. Multi-view geometry
IV. Recognition
V. Additional topics

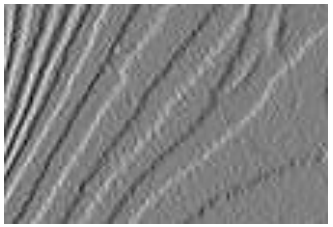
I. Early vision



nation an

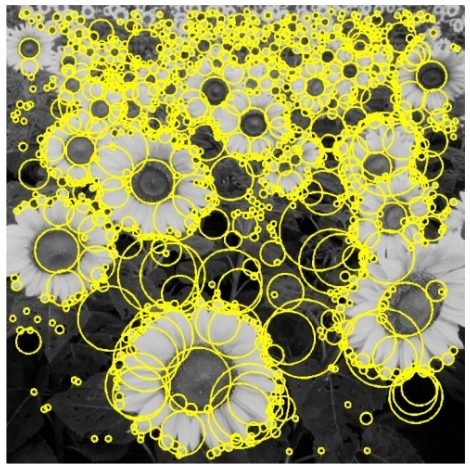
ng

=

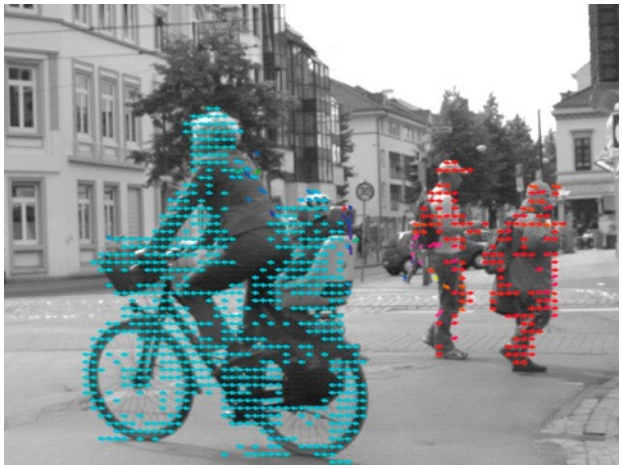


Cameras and sensors
Light and color

Linear filtering
Edge detection

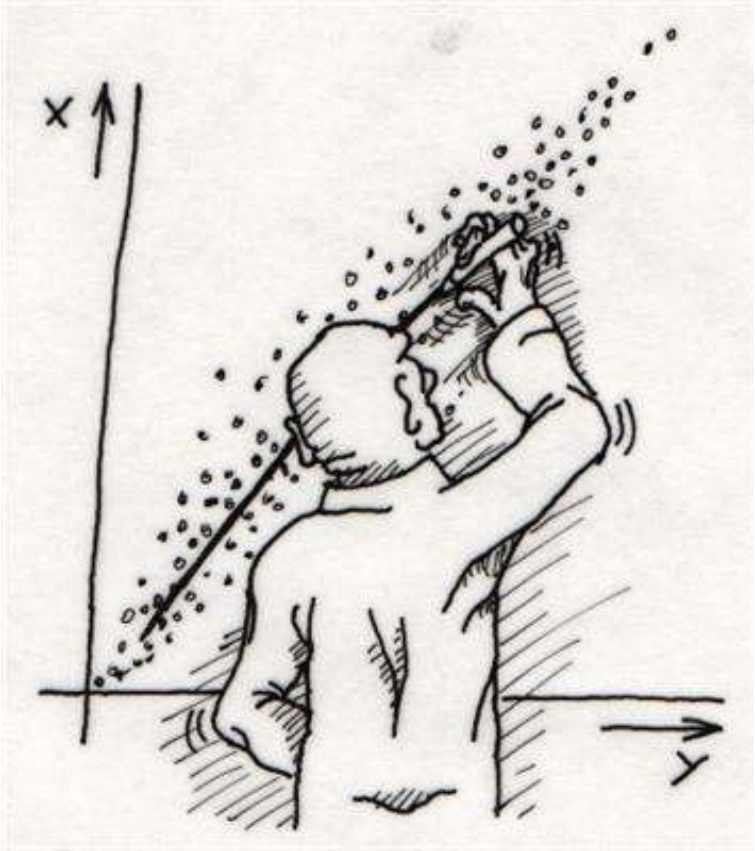


Feature extraction



Optical flow

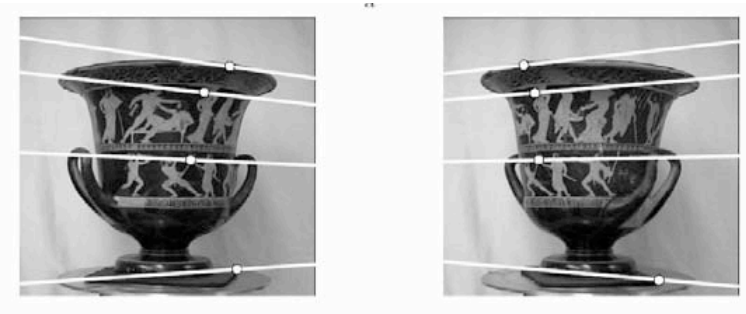
II. “Mid-level vision”



Fitting: Least squares
Voting methods

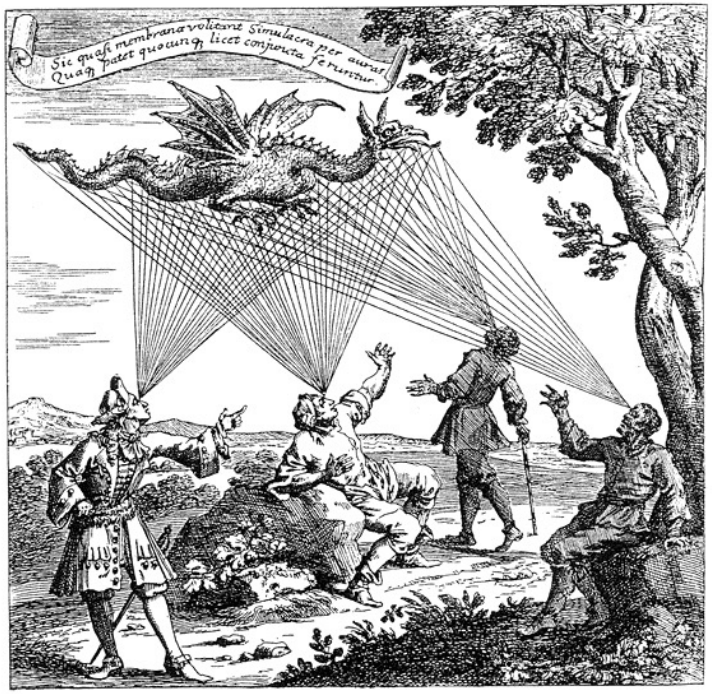
Alignment

III. Multi-view geometry



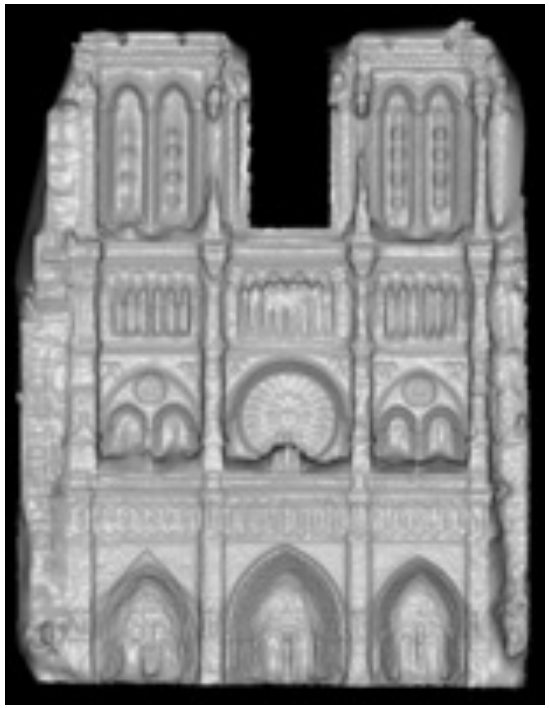
Epipolar geometry

Two-view stereo



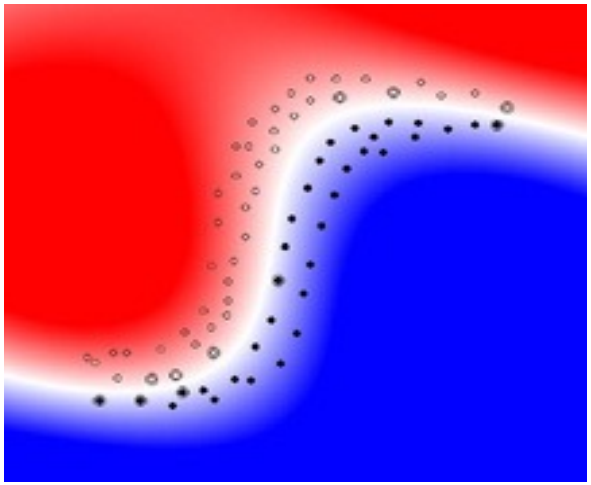
Драконъ, видимый подъ различными углами зрѣнія
По гравюрѣ на мѣди изъ „Oculus artificialis telediopticus" Цана. 1702 года.

Structure from motion

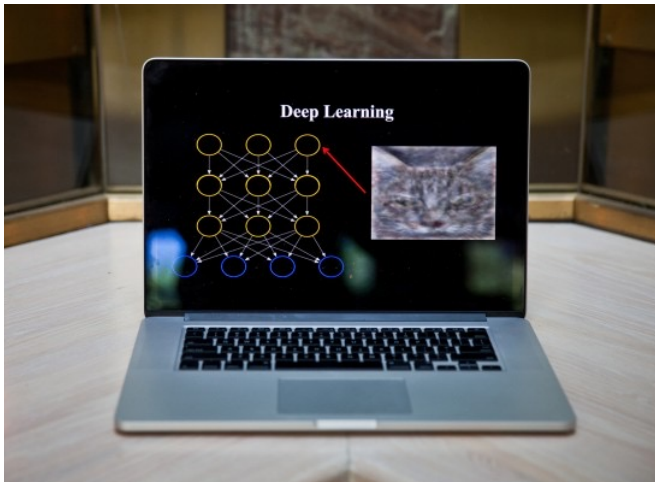


Multi-view stereo

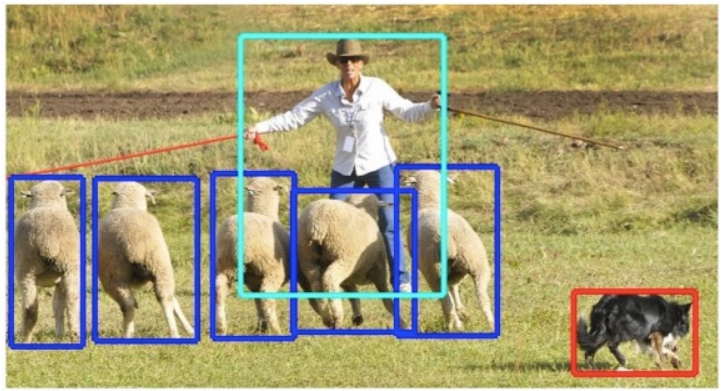
IV. Recognition



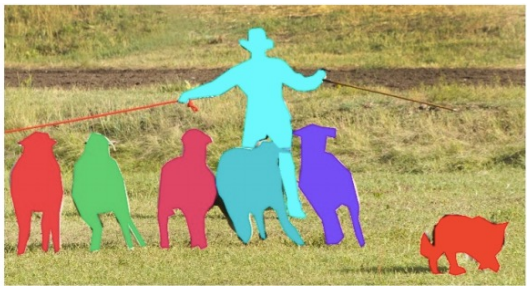
Basic classification



Deep learning



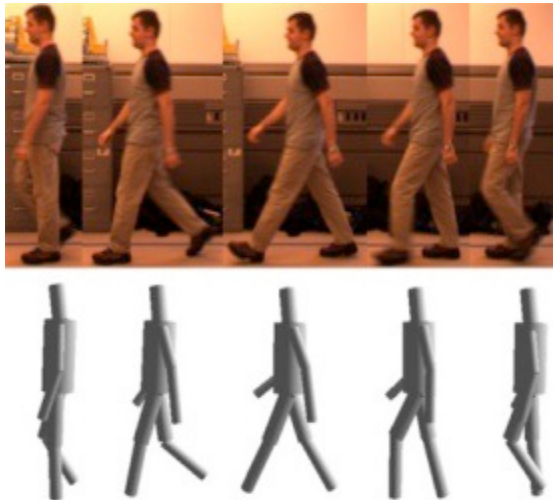
Object detection



Segmentation

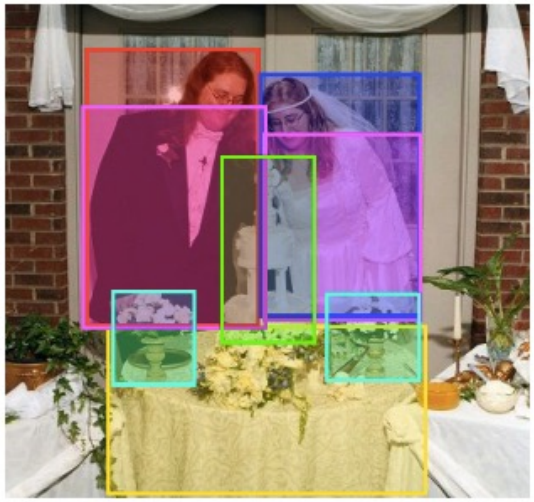
V. Additional Topics (time permitting)

Generation



Video

3D scene understanding



Images and text

Vision-based interaction (and games)

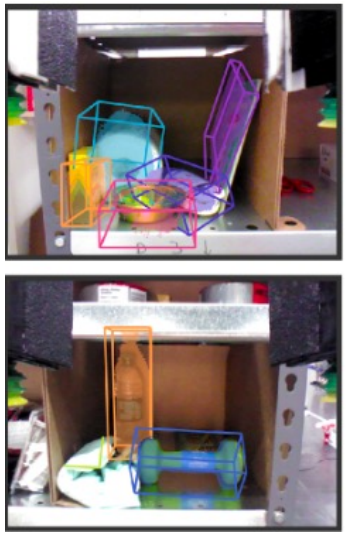
- Human pose estimation
- Activity Recognition

Xbox and Kinect sensor

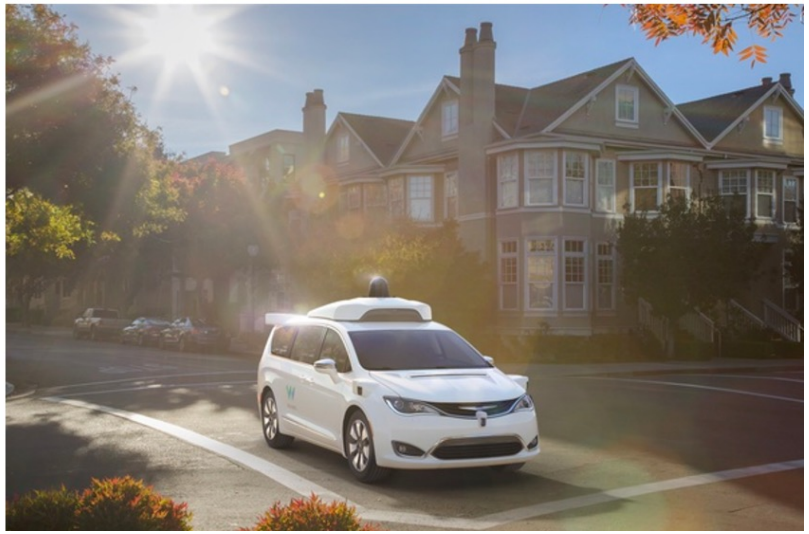
Sony EyeToy

Assistive technologies

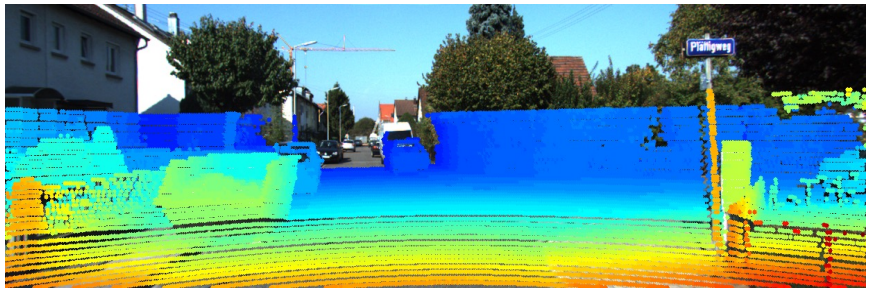
Dexnet

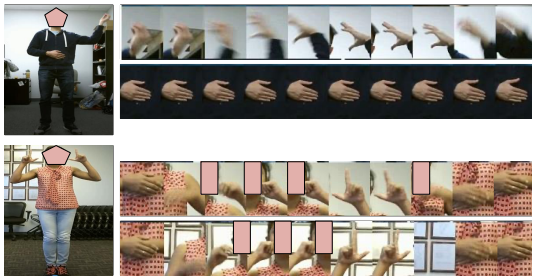


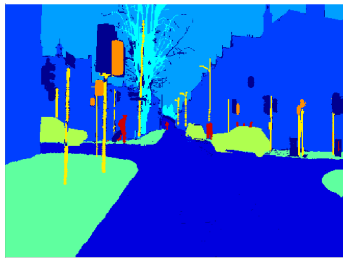
Amazon Picking Challenge

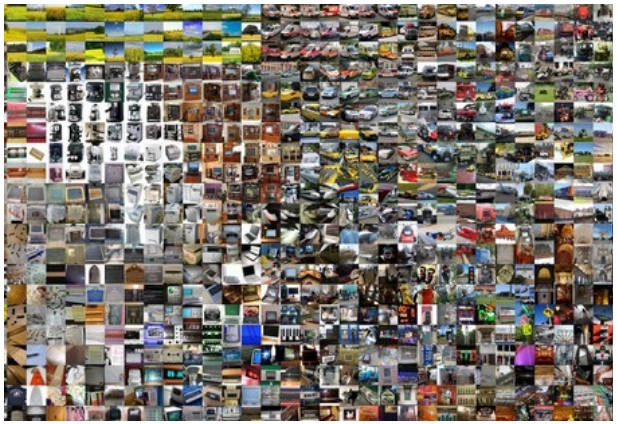


Waymo





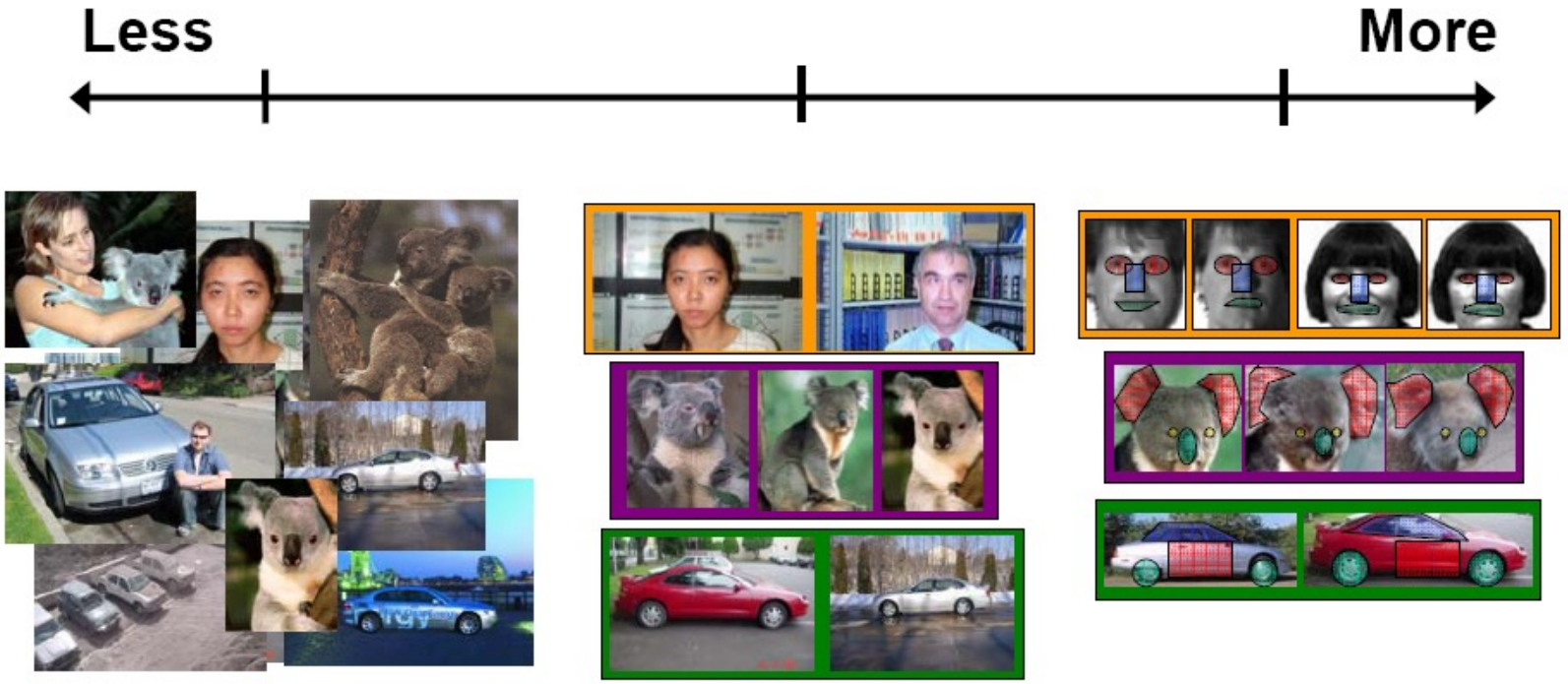




ImageNet

Coco

Spectrum of supervision



Learning approaches proceed in supervised way: need some labeled data

Unsupervised

"Weakly" supervised

Supervised

Definition depends on task

Caltech 5, Caltech 101

Slide credit Fei-Fei Li

Example Datasets

Motorbike

LabelMe
[Russel et al'05

Tiny Images [Torralba et al'07],
80 million tiny images

ImageNet [Fei-Fei, 2008] 10K object categories, sync sets, ontology & word hierarchy

Object Recognition

Dog

Classification

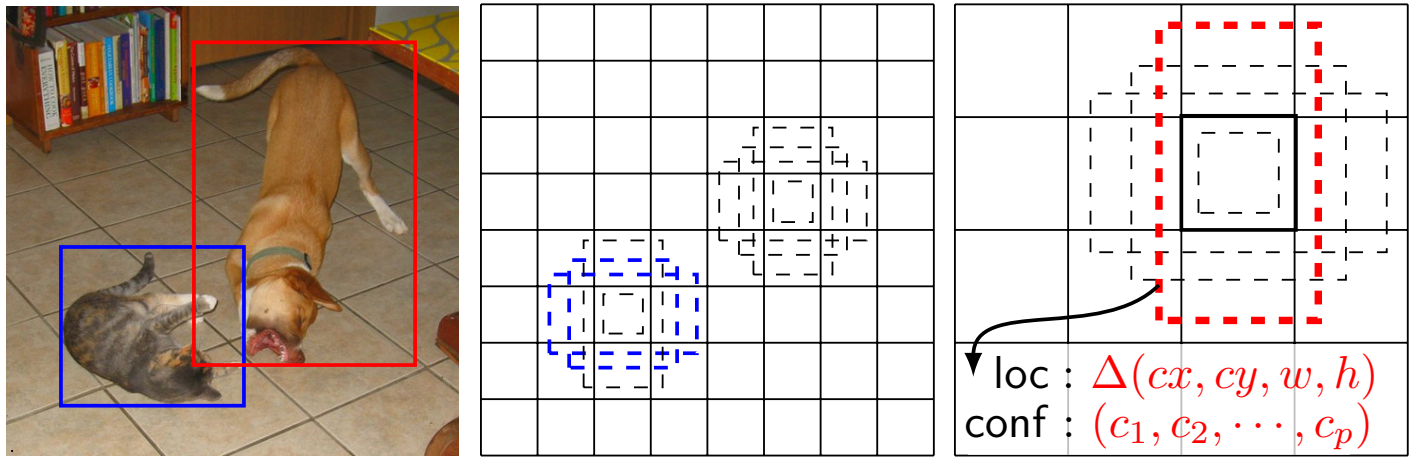
Detection

Segmentation

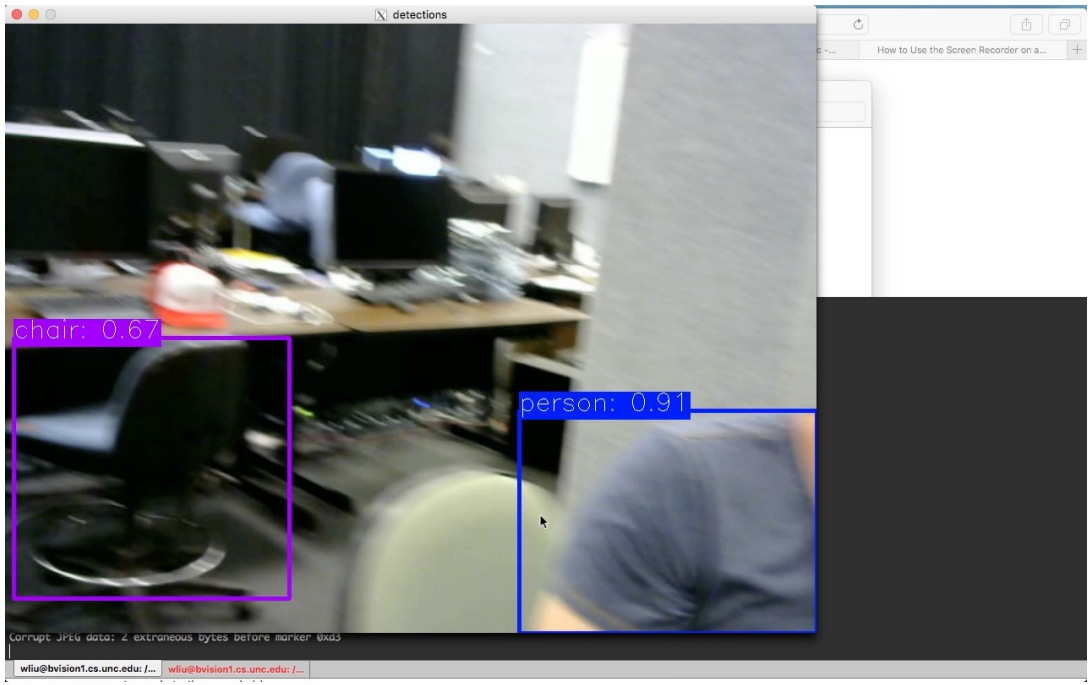
COCO Common Objects in Context

Coco 80 object categories
500K + 2K + 5K
Train instances-validate-text

Mapilary Vista 37 object categories
18K + 2K + 5K
train-validate-text



Multi-object detection 52fps

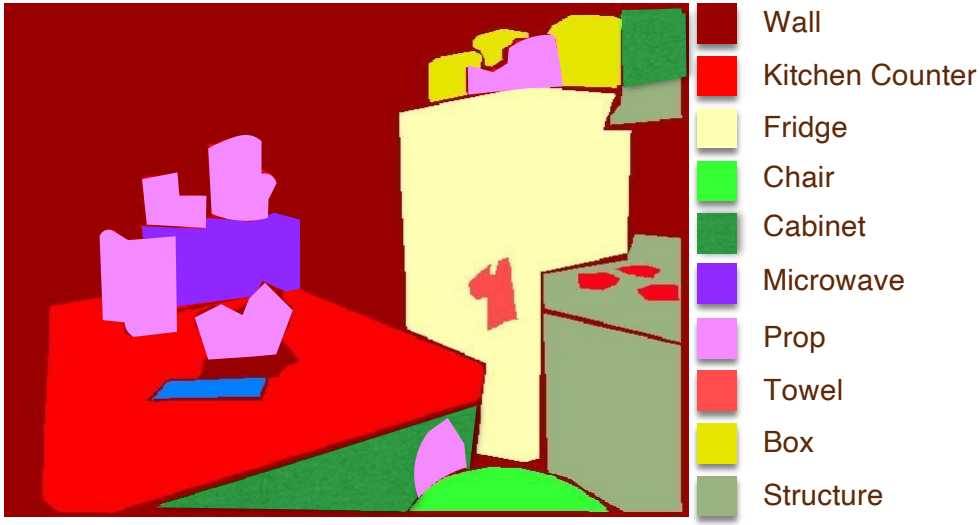


- SSD object detector single pass
- confidences for 80 categories
- and their bounding boxes
- Overall mAP ~ 50%

Fast Single Shot Detection and Pose Estimation P. Poirson, Philip Ammirato, Cheng-Yang Fu W. Liu, J Kosecka, A. Berg

Semantic Segmentation

- Definition: Assigning a label to each pixel.
- Labels can be object categories such as closet, fridge, chair or can be structural categories such as wall and structure.



Semantic Segmentation

Large Scale Image Categorization

11 million images, 10,000 image categories 15,000+ synsets

Slide cerdit Fei-Fei Li

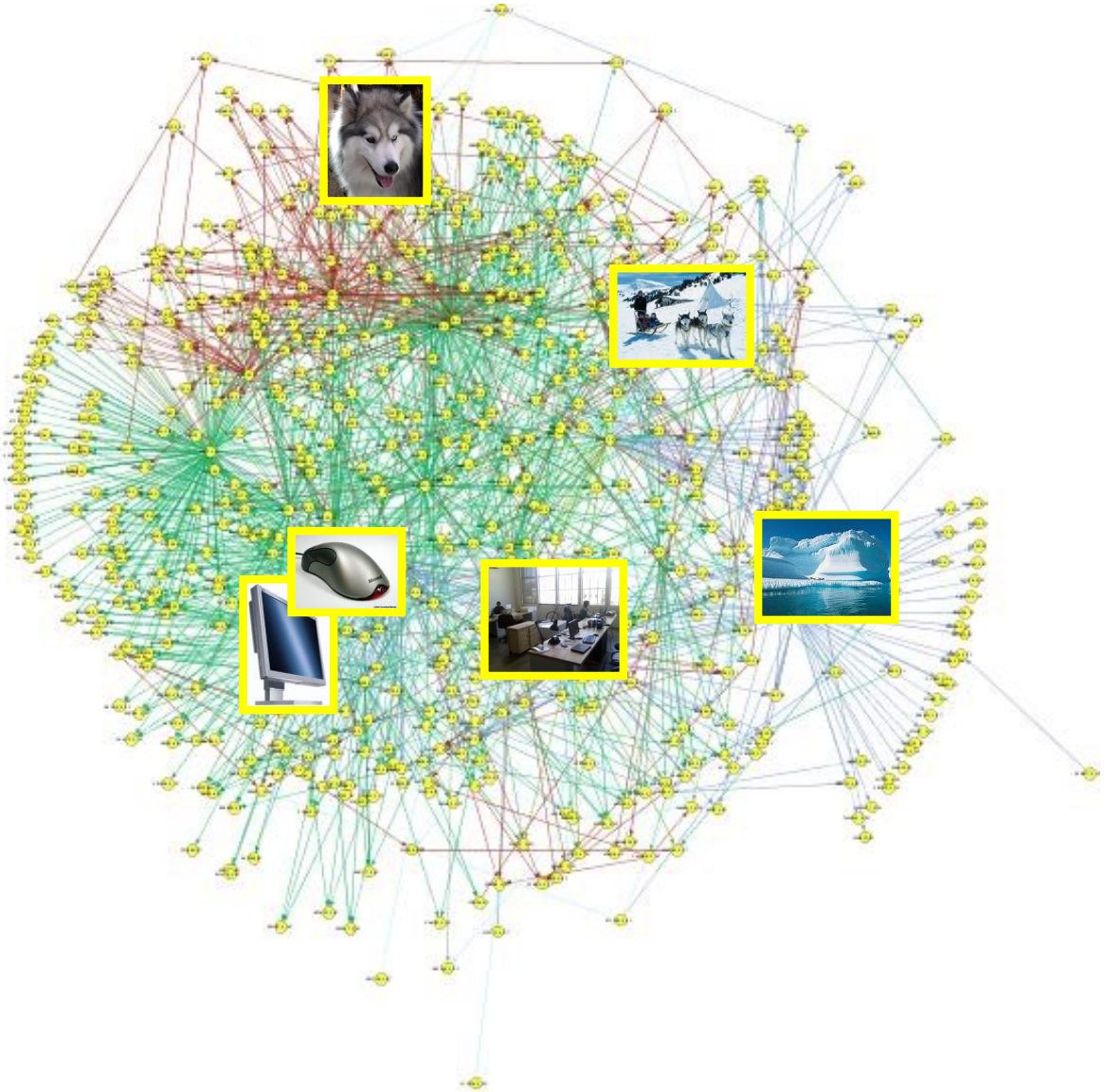
IMAGENET is a knowledge ontology

- Taxonomy
- Partonomy
- The “social network” of

- S: (n) **car**, auto, automobile, machine, motorcar (a motor vehicle with four wheels; usually propelled by an internal combustion engine) *"he needs a car to get to work"*
  - *direct hyponym* / *full hyponym*
  - ***part meronym***
    - S: (n) accelerator, accelerator pedal, gas pedal, gas, throttle, gun (a pedal that controls the throttle valve) *"he stepped on the gas"*
    - S: (n) air bag (a safety restraint in an automobile; the bag inflates on collision and prevents the driver or passenger from being thrown forward)
    - S: (n) auto accessory (an accessory for an automobile)
    - S: (n) automobile engine (the engine that propels an automobile)
    - S: (n) automobile horn, car horn, motor horn, horn, hooter (a device on an automobile for making a warning noise)
    - S: (n) buffer, fender (a cushion-like device that reduces shock due to an impact)
    - S: (n) bumper (a mechanical device consisting of bars at either end of a vehicle to absorb shock and prevent serious damage)
    - S: (n) car door (the door of a car)
    - S: (n) car mirror (a mirror that the driver of a car can use)
    - S: (n) car seat (a seat in a car)
    - S: (n) car window (a window in a car)
    - S: (n) fender, wing (a barrier that surrounds the wheels of a vehicle to block splashing water or mud) *"in Britain they call a fender a wing"*
    - S: (n) first gear, first, low gear, low (the lowest forward gear ratio in the gear box of a motor vehicle; used to start a car moving)
    - S: (n) floorboard (the floor of an automobile)
    - S: (n) gasoline engine, petrol engine (an internal-combustion engine that burns gasoline; most automobiles are driven by gasoline engines)
    - S: (n) glove compartment (compartment on the dashboard of a car)
    - S: (n) grille, radiator grille (grating that admits cooling air to car's radiator)
    - S: (n) high gear, high (a forward gear with a gear ratio that gives the greatest vehicle velocity for a given engine speed)
    - S: (n) hood, bonnet, cowl, cowling (protective covering consisting of a metal part that covers the engine) *"there are powerful engines under the hoods of new cowling in order to repair the plane's engine"*
    - S: (n) luggage compartment, automobile trunk, trunk (compartment in an automobile that carries luggage or shopping or tools) *"he put his golf bag in the trunk*
    - S: (n) rear window (car window that allows vision out of the back of the car)
    - S: (n) reverse, reverse gear (the gears by which the motion of a machine can be reversed)
    - S: (n) roof (protective covering on top of a motor vehicle)
    - S: (n) running board (a narrow footboard serving as a step beneath the doors of some old cars)
    - S: (n) stabilizer bar, anti-sway bar (a rigid metal bar between the front suspensions and between the rear suspensions of cars and trucks; serves to stabilize the cl
    - S: (n) sunroof, sunshine-roof (an automobile roof having a sliding or raisable panel) *"'sunshine-roof' is a British term for 'sunroof'"*
    - S: (n) tail fin, tailfin, fin (one of a pair of decorations projecting above the rear fenders of an automobile)
    - S: (n) third gear, third (the third from the lowest forward ratio gear in the gear box of a motor vehicle) *"you shouldn't try to start in third gear"*
    - S: (n) window (a transparent opening in a vehicle that allow vision out of the sides or back; usually is capable of being opened)

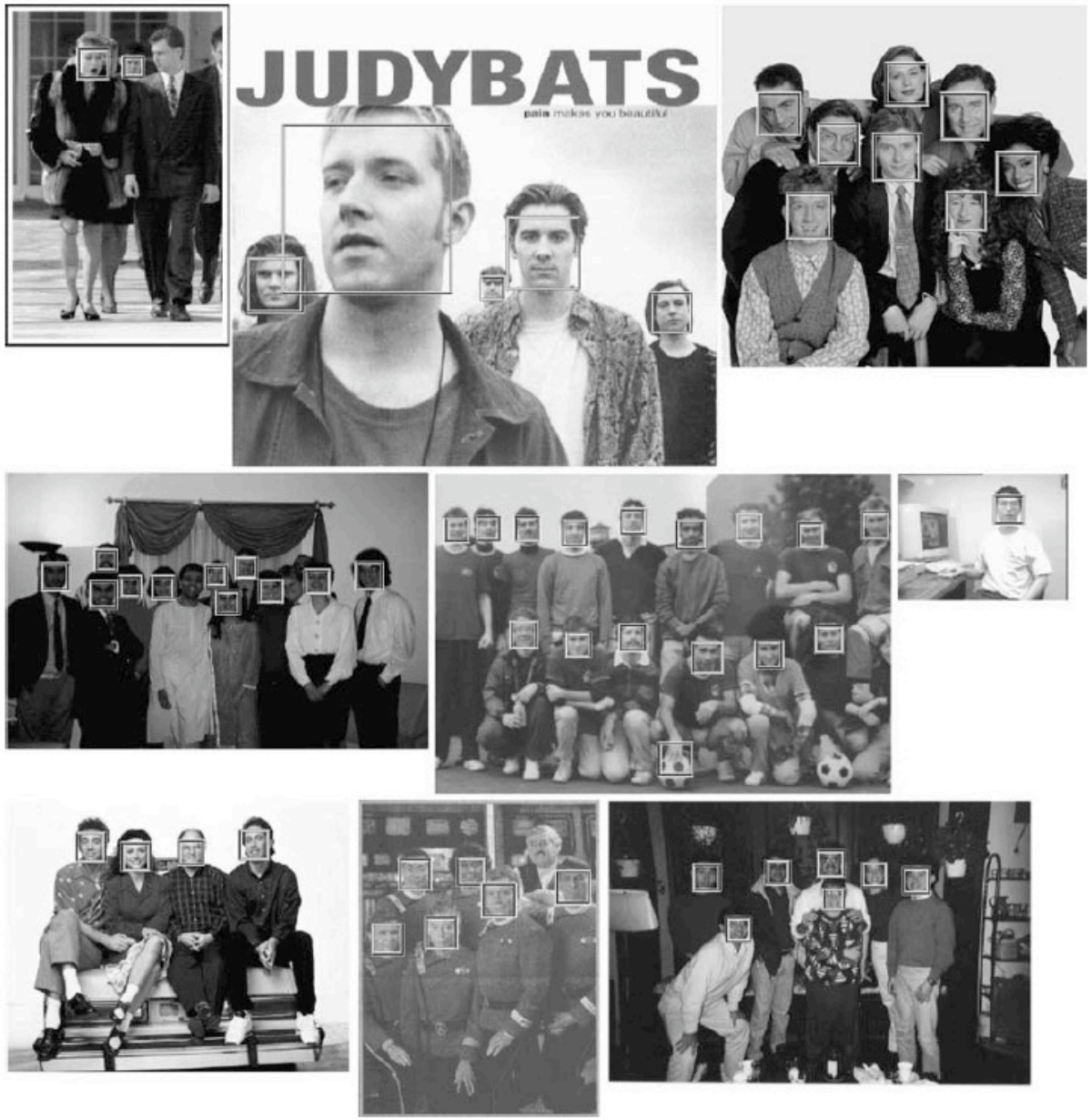
IMAGENET is a knowledge ontology

- Taxonomy
- Partonomy
- The “social network” of visual concepts
  - Prior knowledge
  - Context
  - Hidden knowledge and structure among visual concepts



Slide cerdit Fei-Fei Li

Face Detection

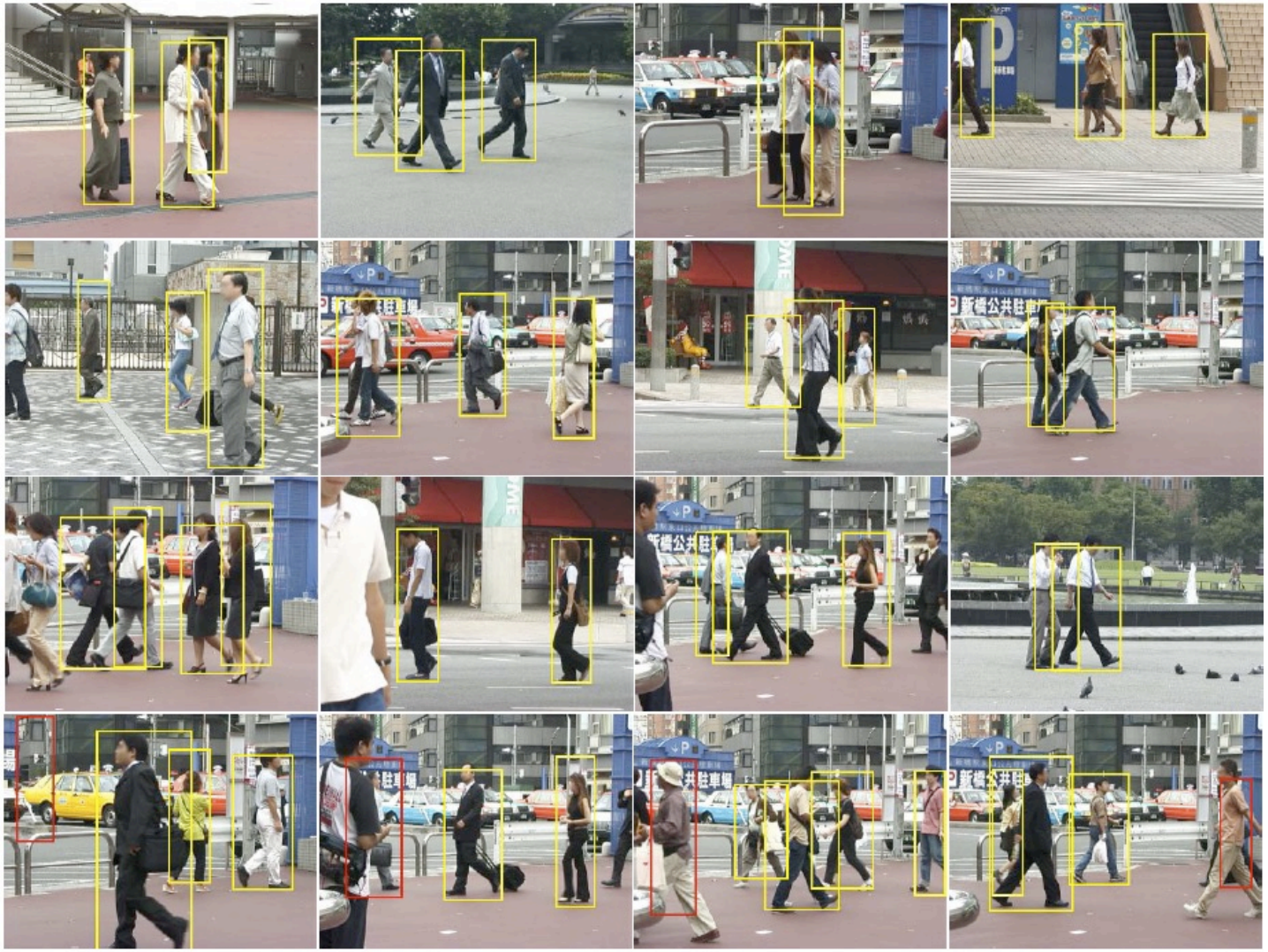


Sliding window approach:
for each possible rectangular region in the image asks the question, is there a face here?

Now even integrated into many consumer digital cameras.

Image from: Viola & Jones, 2004.

Person detection



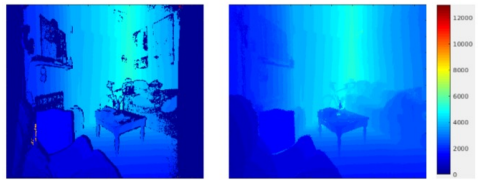
Bastian Leibe, Edgar Seemann, and Bernt Schiele

Scene Recognition

Lazebnik, Schmid, F

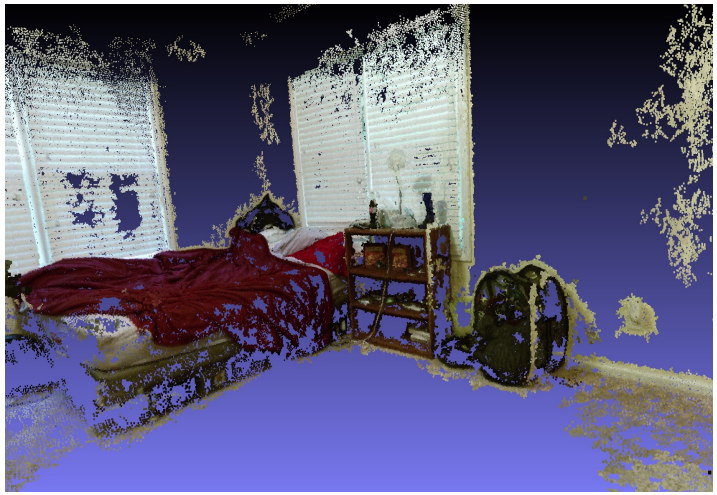
Active Vision Dataset

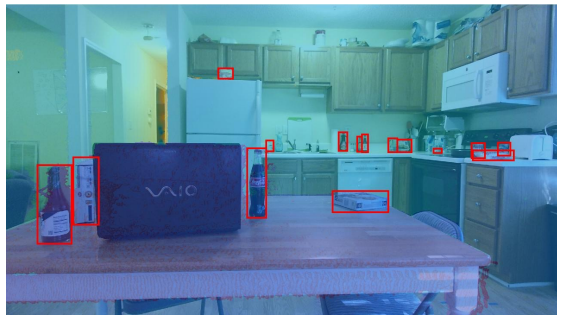
UNC DEPARTMENT OF COMPUTER SCIENCE

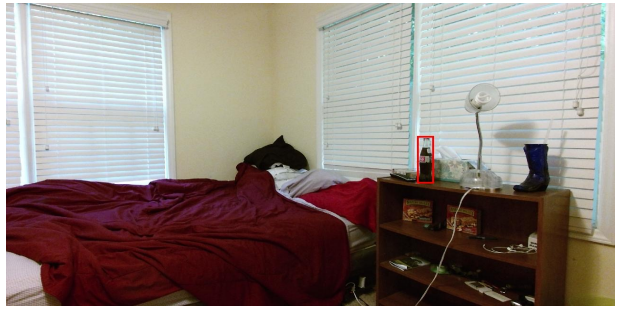


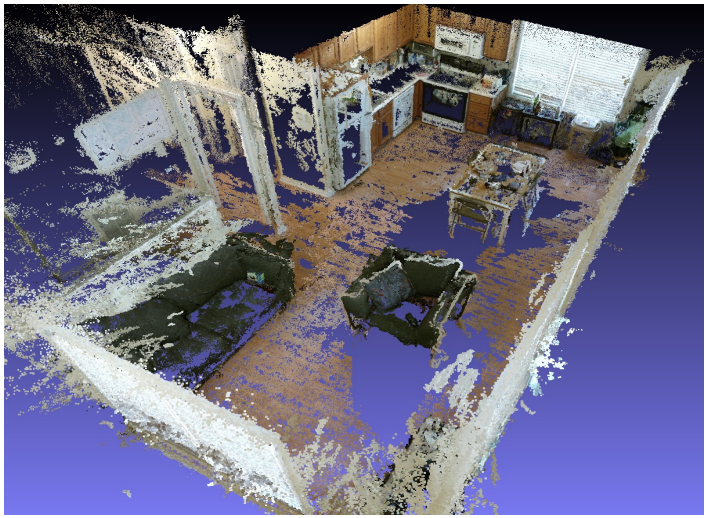
9 indoors scenes, 17 scans, 21K images

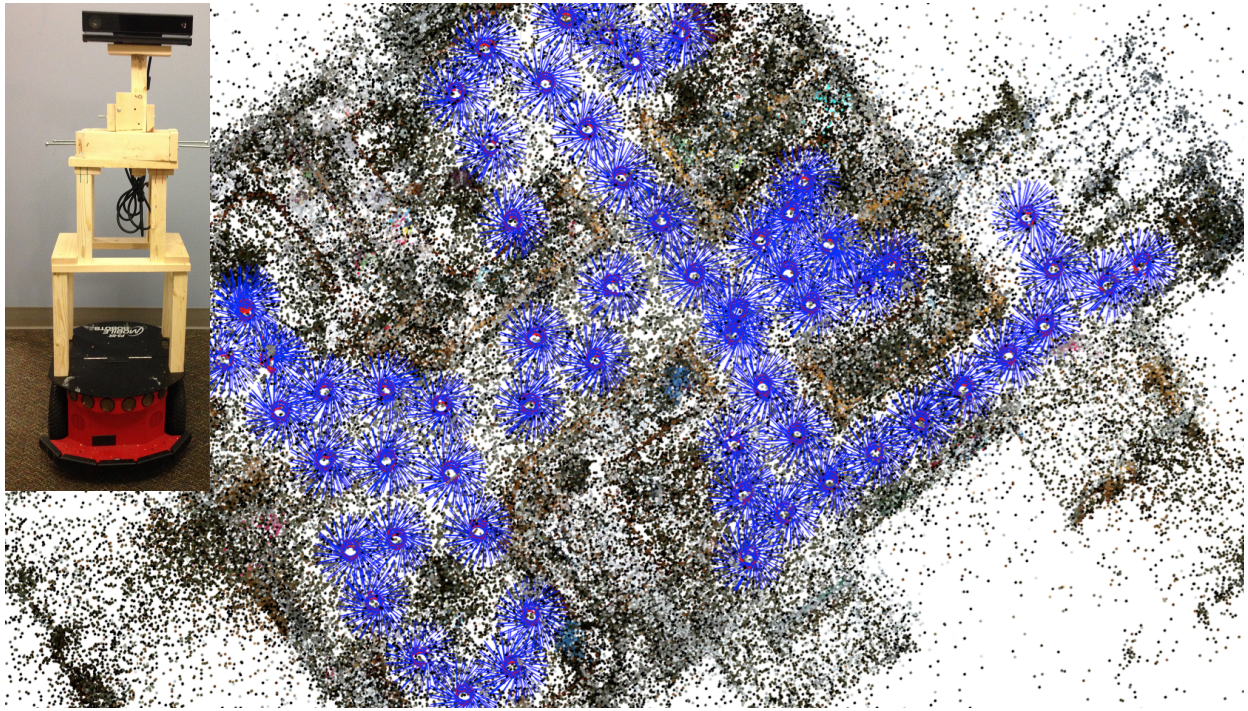
P. Ammirato, P. Poirson, E. Park, A. Berg, J.Kosecka





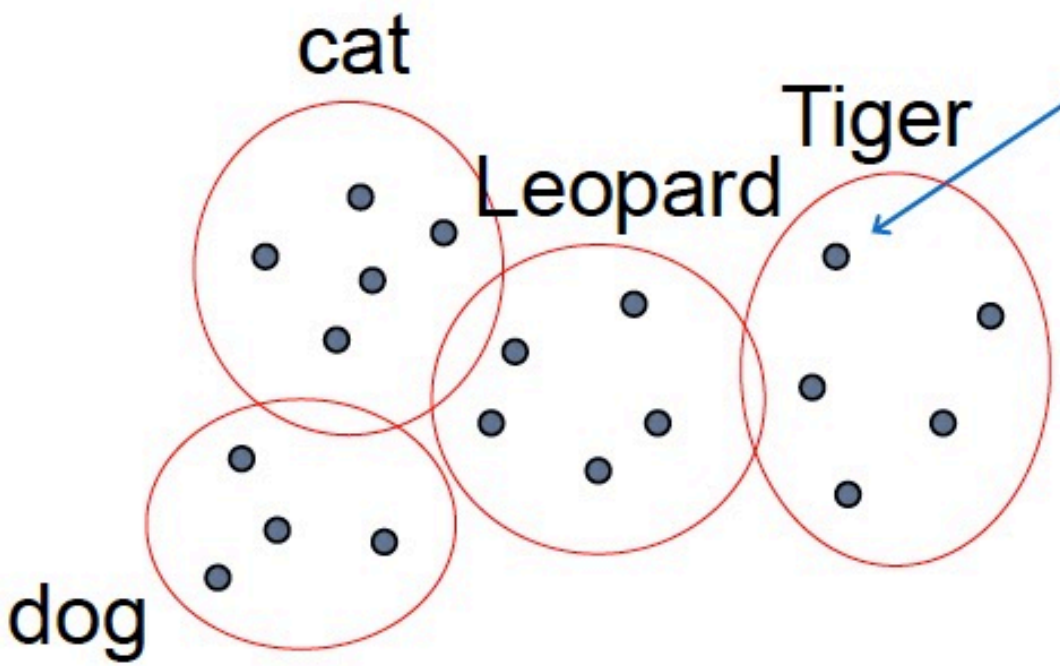






Why Categorize?

1. Knowledge Transfer
2. Communication



Slide credit: A. Efros

Problems with Visual Categories

- A lot of categories are functional

Chair

- World is too varied

- Categories are 3D, but images are 2D

car

Slide credit: A. Efros