Computer Vision
CS 682

Jana Kosecka
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%,
- Midterm 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

Logistics

- **Homeworks Due:** 11:59pm submitted on Blackboard
  Homeworks 50%,
- **Late policy**: 3 days late budget, notify instructor or TA
- **Prerequisites:** linear algebra, calculus, probability and statistics
- **Lectures:** Introduction by an instructor, homeworks every two weeks
- **Projects:** up to teams of 2-3 people

  apply vision to a problem you care about; re-implement a paper,

  extend some approach.

- **Dates**
  - Project proposals due March 22th
  - May week of finals final report due, project presentations instead of finals

Grade Scale

A	>96
A-	92-96
B+	88-92
B	84-88
B-	80-84
C+	76-80
C	72-76
C-	68-72
F	< 68

Computer Vision

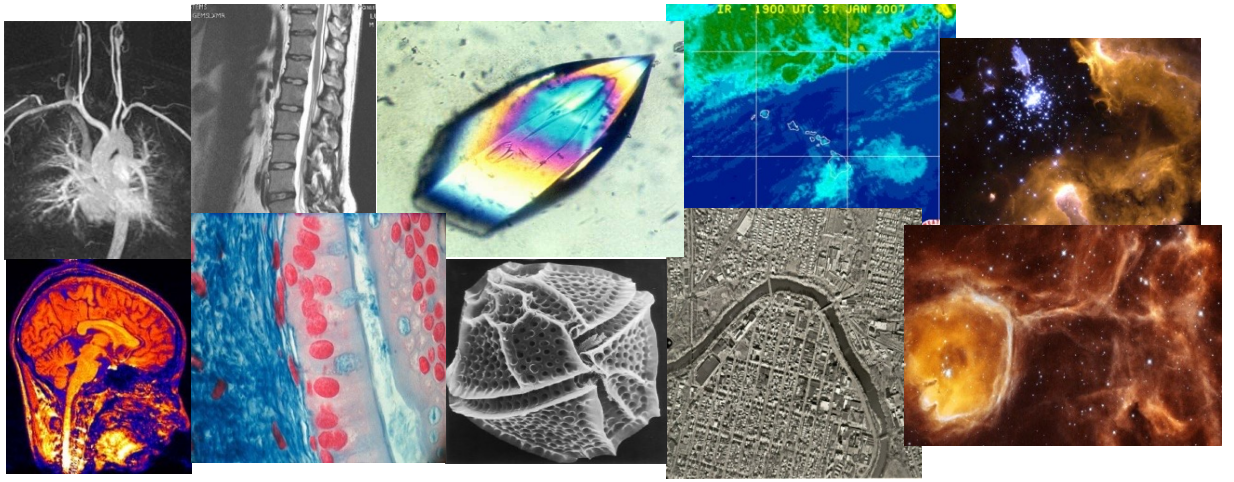
- There are 1.8 billion images uploaded to Internet every day
- Every autonomous car, delivery robot, laptop and phone is equipped with cameras
- The opportunities and challenges of visual perception

Why study computer vision?

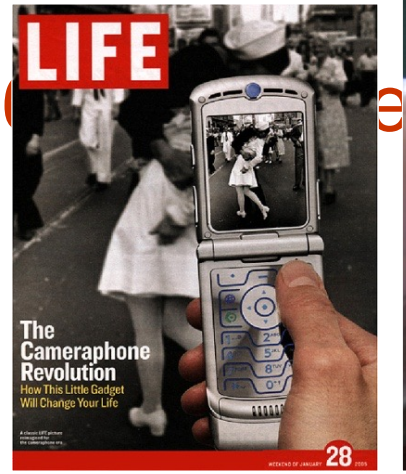
- Images and video are everywhere!

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

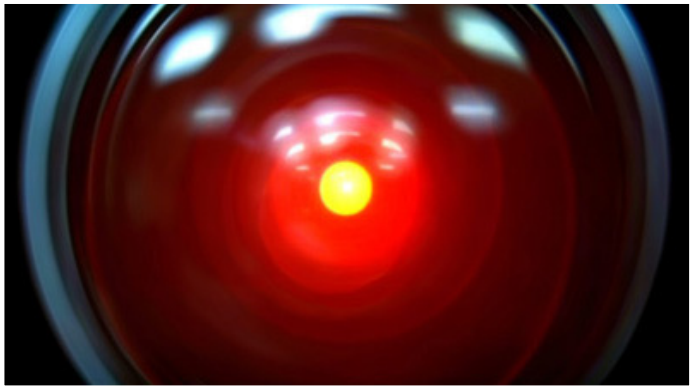
Surveillance and security



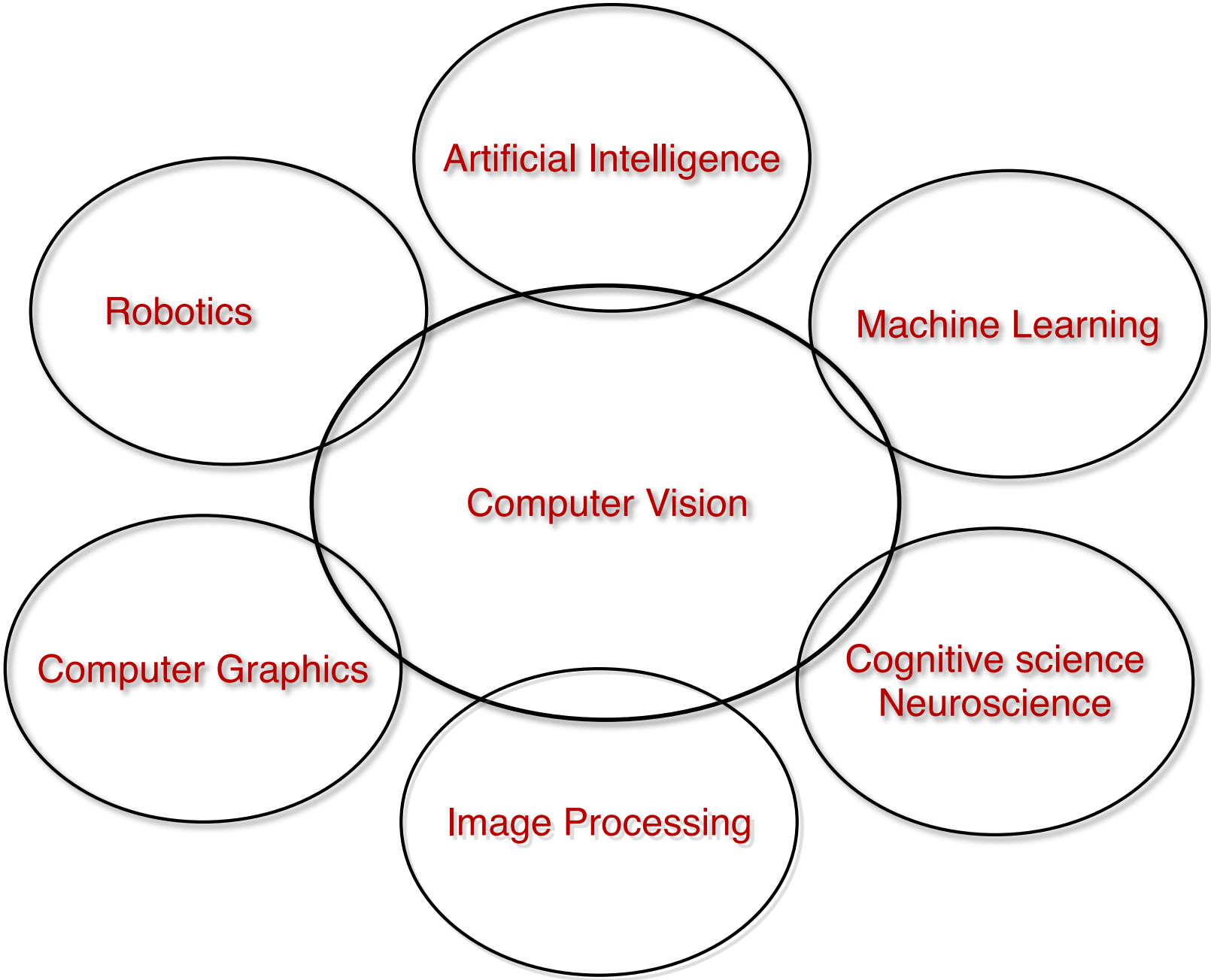
Medical and scientific images


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

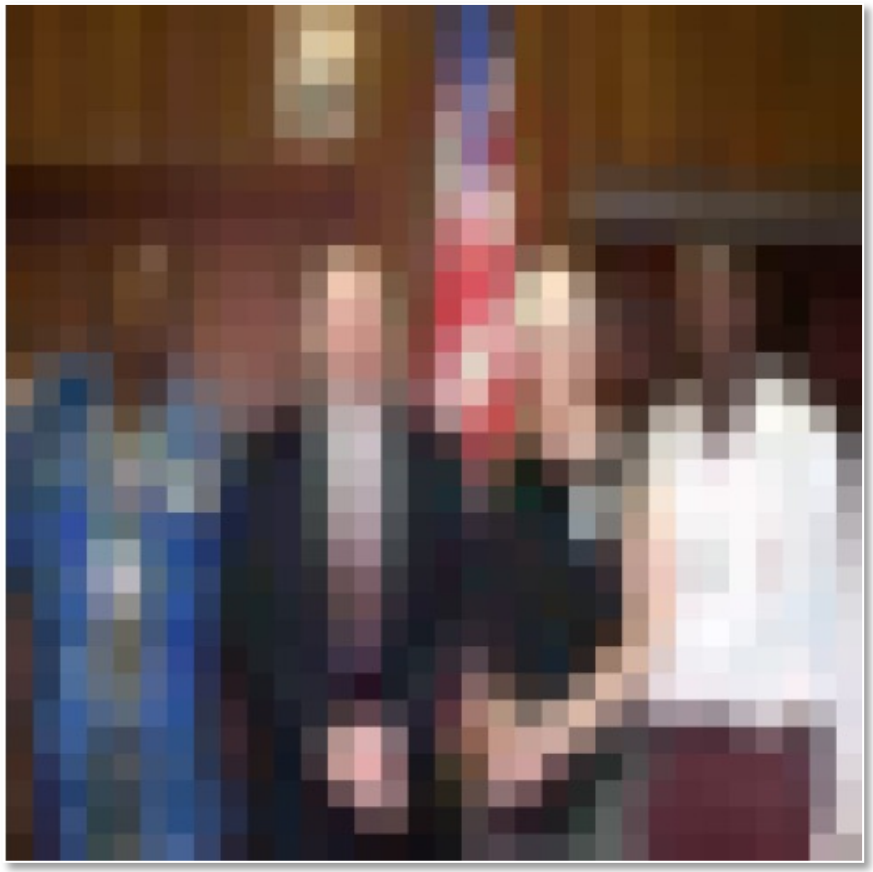
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

What a computer sees

Source: S. Narasimhan

The goal of computer vision

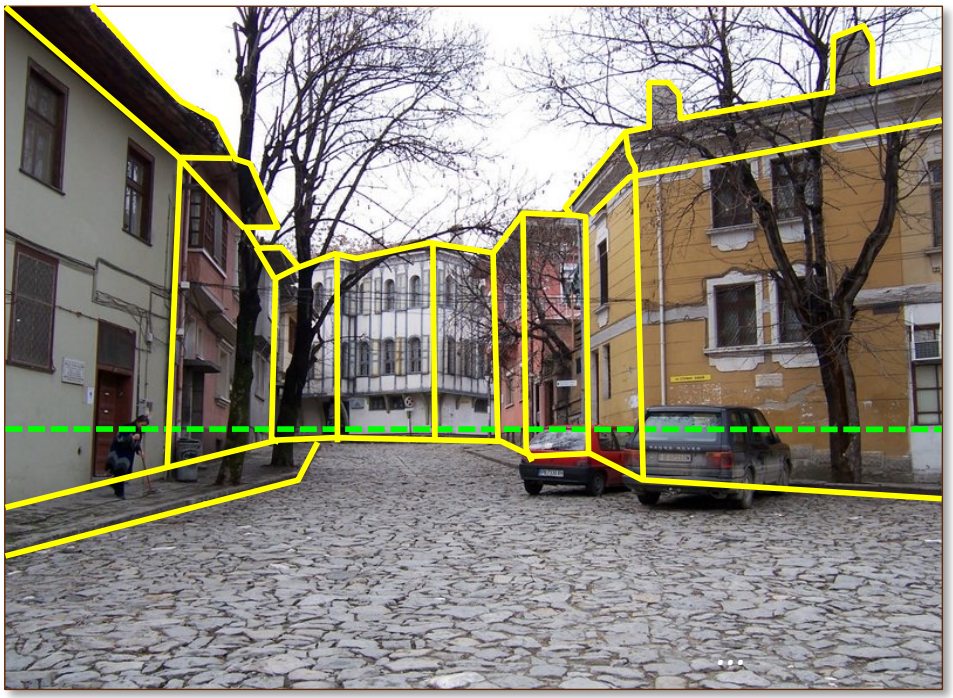
- To extract information from images



Humans are remarkably good at this…

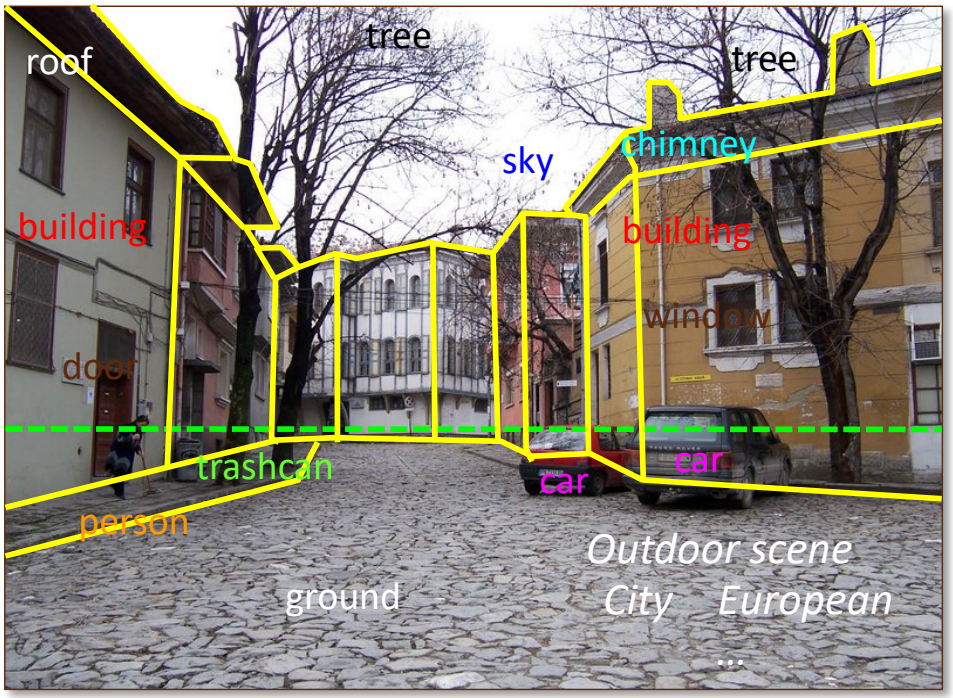
Source: “80 million tiny images” by Torralba et al.

What kind of information can be extracted from an image?



Geometric information

What kind of information can be extracted from an image?

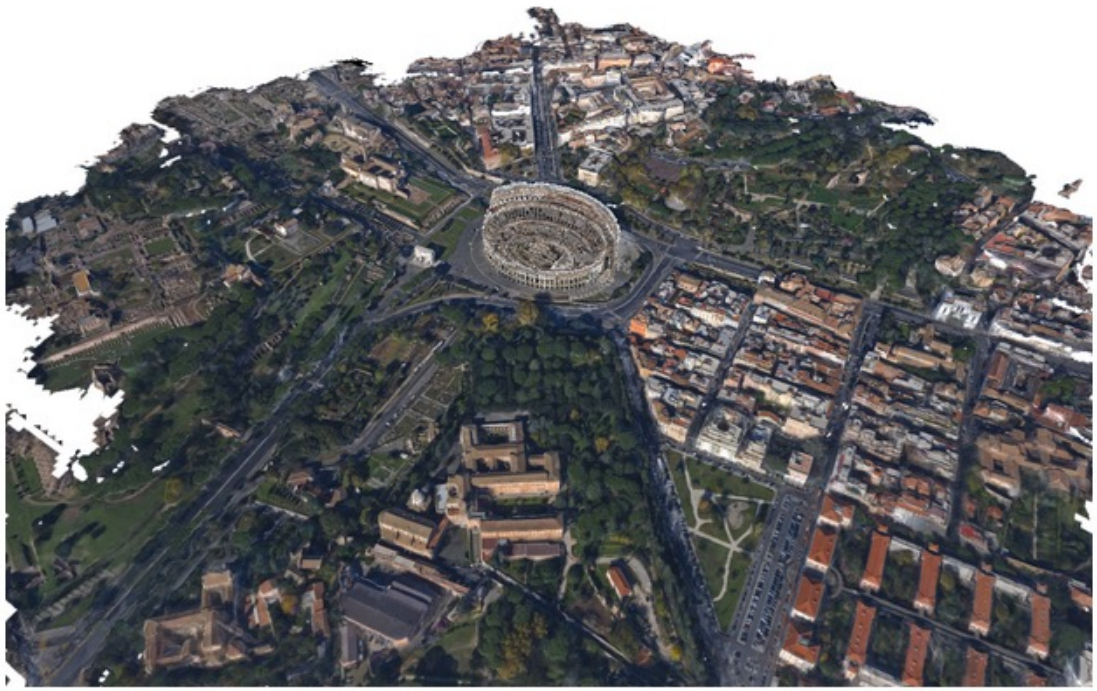


Geometric information

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 https: //www.youtube.com/watch?v=NdeD4cjLI0c

Neural Scene Rendering
3D scene representations

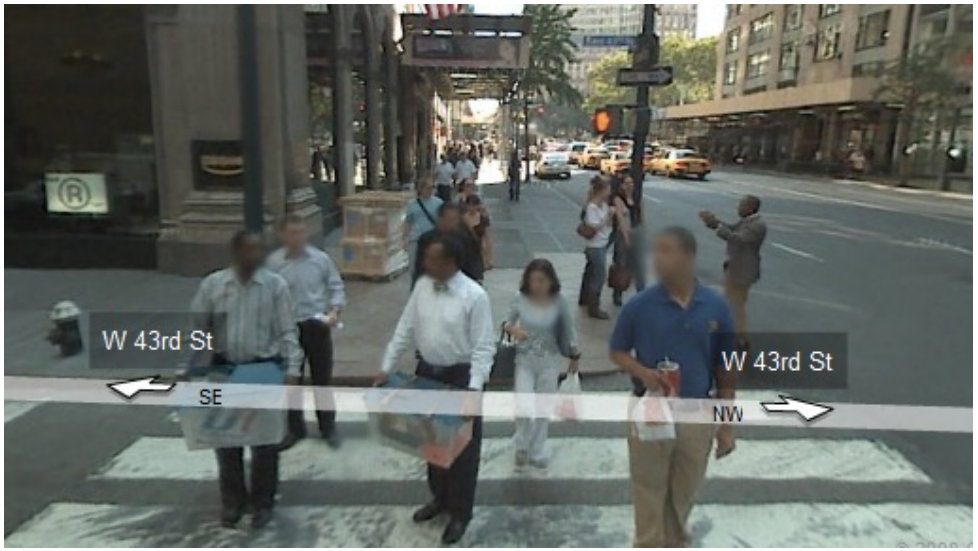
Courtesy NERF Studio API https://docs.nerf.studio/en/latest/

Representing Scenes as Neural Radiance Fields for View Synthesis

Ben Mildenhall* Pratul P. Srinivasan* Matthew Tancik* Jonathan T. Barron
Ravi Ramamoorthi Ren Ng

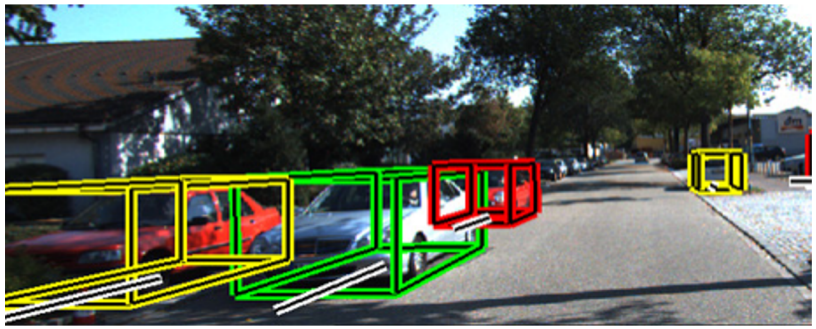
Recognition: “Simple” patterns

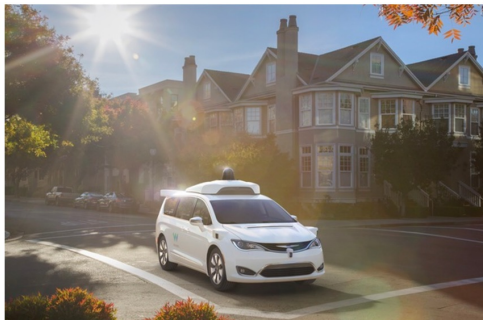
Recognition: Faces

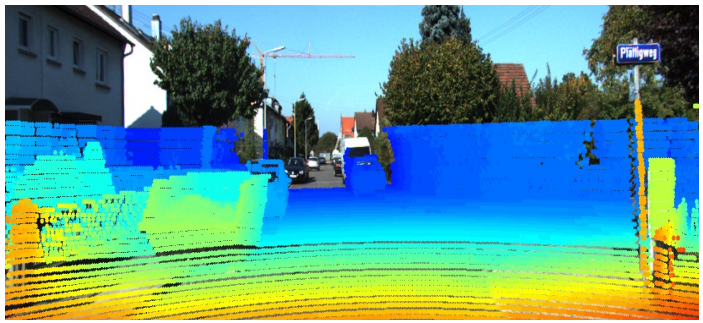


Computer Vision in Robot Perception

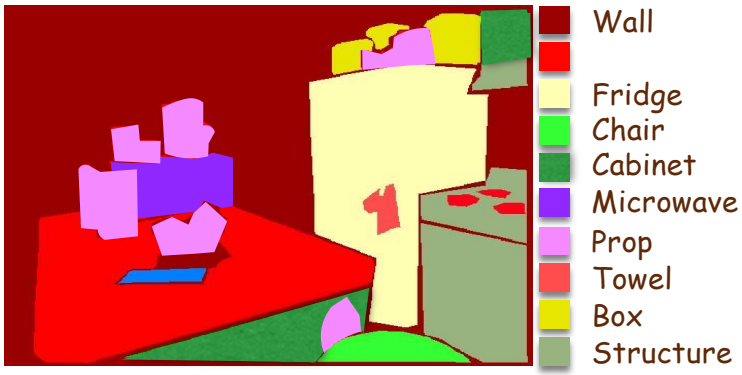
- Autonomous Driving







- Manufacturing and Service Robotics



Vision in supermarkets

LaneHawk by EvolutionRobotics

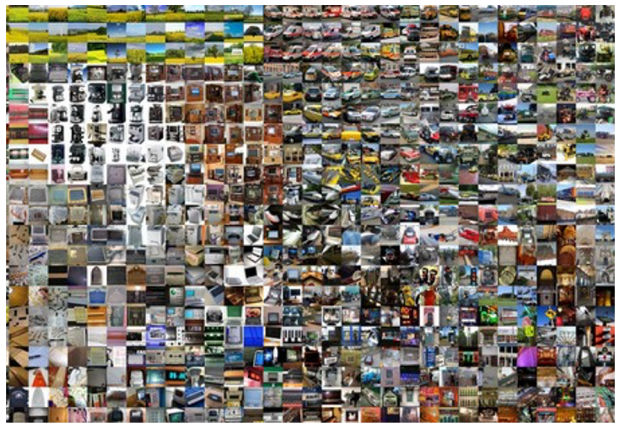
"A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you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

Source: S. Seitz

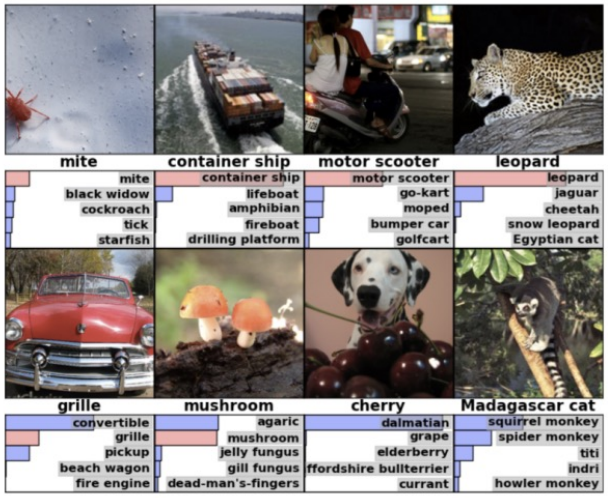
Deep Learning

- ~14 million images, 20k classes
- Images gathered from Internet
- Human labels via Amazon Turk

IMAGENET



[NIPS 2012]

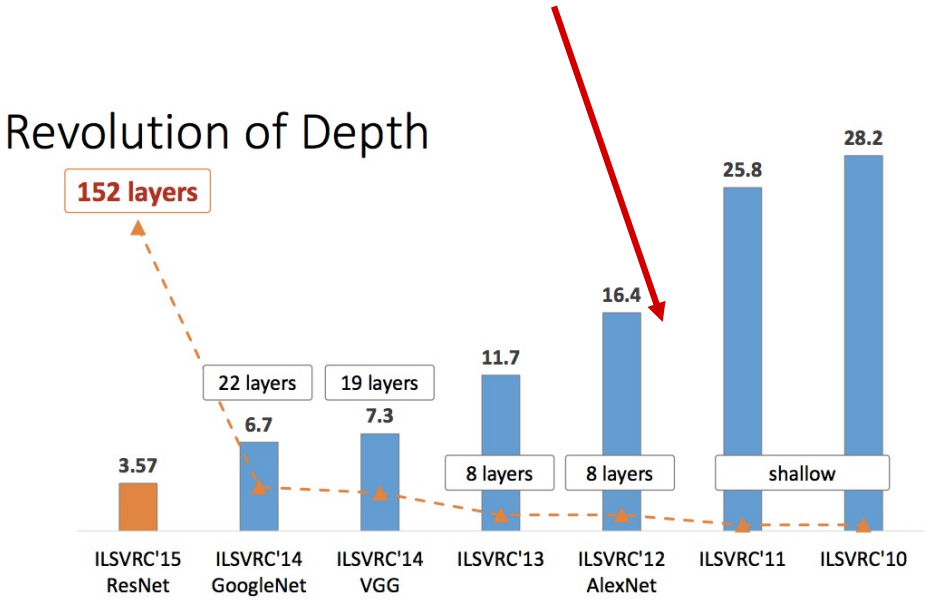


ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

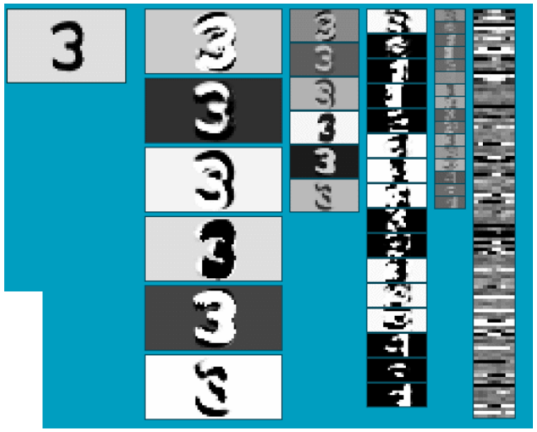
Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

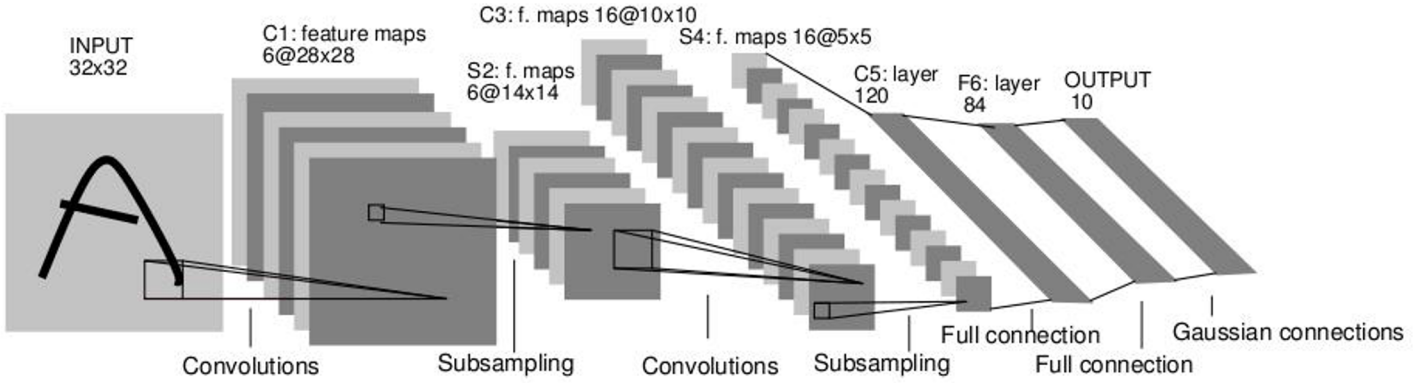
Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca



Convolutional neural networks

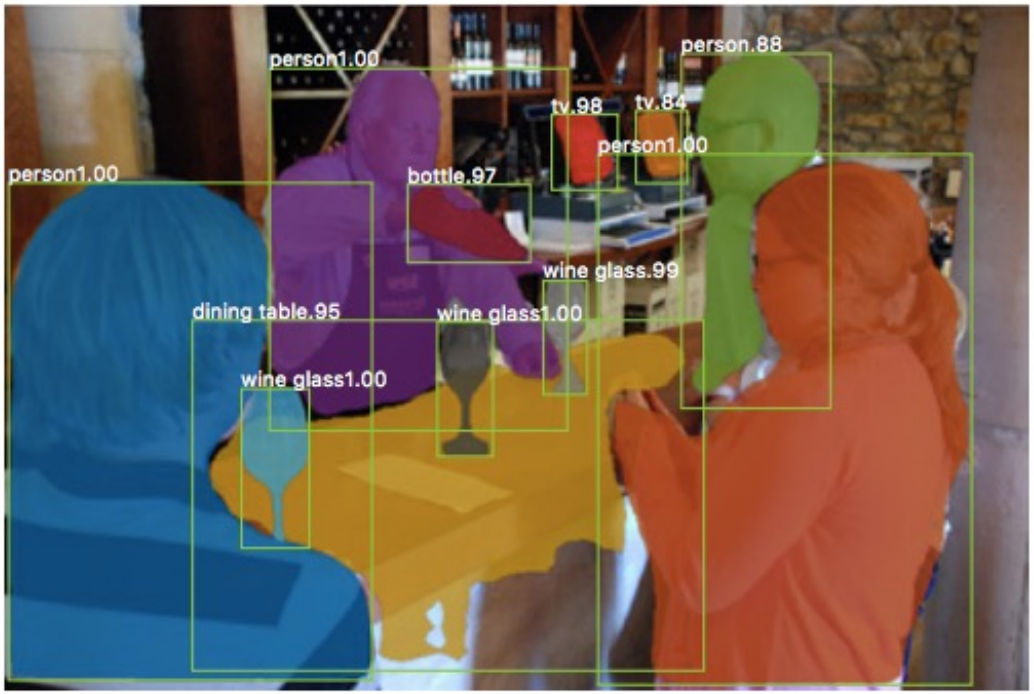
Beginnings: [LeCun *et al*. 1998]





Traditional supervised approach – Neural Networks Back-propagation, requires lots of labeled data.

Object detection, instance segmentation



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

Vision: Image generation

- Faces: 1024x1024 resolution, CelebA-HQ dataset

T. Karras, T. Aila, S. Laine, and J. Lehtinen, Progressive Growing of GANs for Improved Quality, Stability, and Variation, ICLR 2018 Follow-up work

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

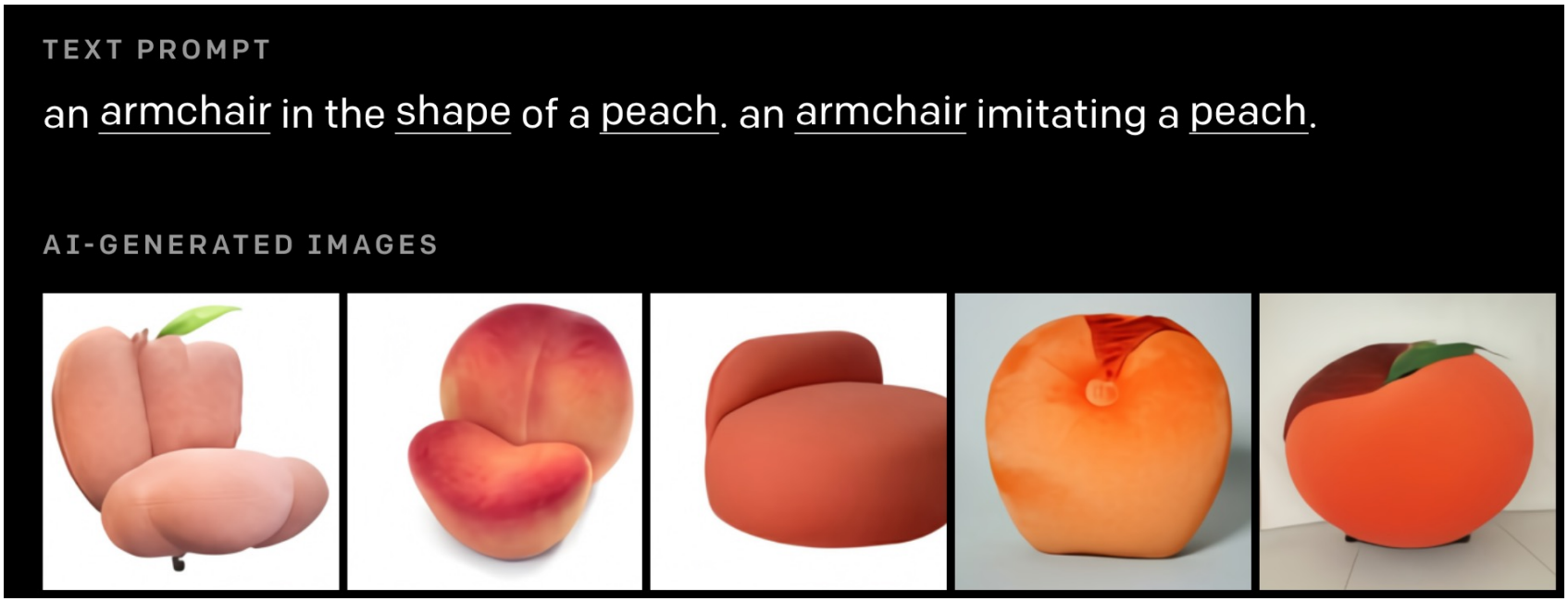
As synthetic media spreads, even real images will invite skepticism.

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

Other exciting developments

- Vision and language models: DALL-E, CLIP



Other exciting developments

- Vision and language models: DALL-E, CLIP

An astronaut Teddy bears A bowl of soup

riding a horse lounging in a tropical resort in space playing basketball with cats in space

in a photorealistic style in the style of Andy Warhol as a pencil drawing

→

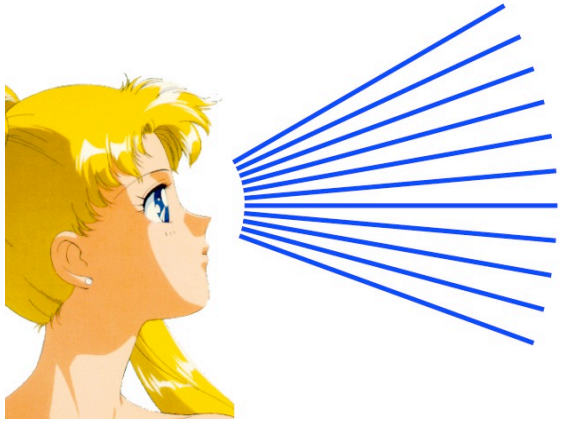
Course overview

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

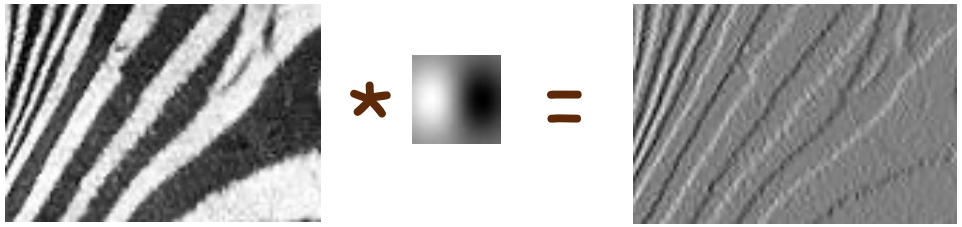
Advanced topics: Deep Learning CS 747

I. Early vision

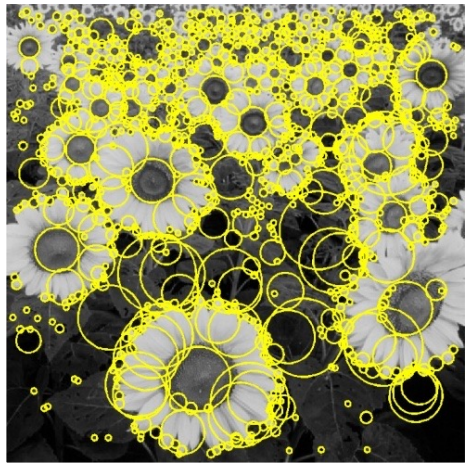
- Basic image formation and processing



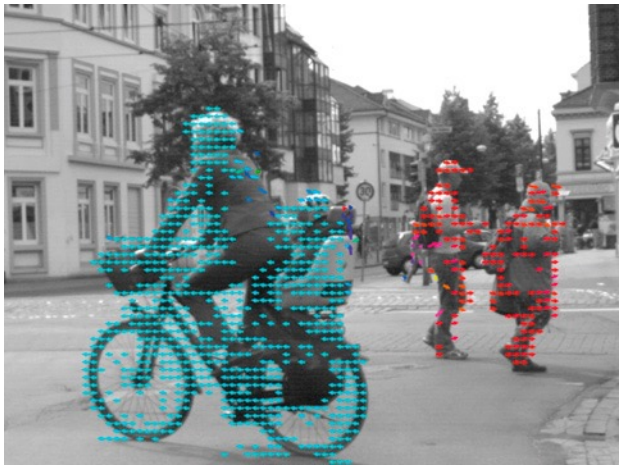
Cameras and sensors
Light and color



Linear filtering
Edge detection

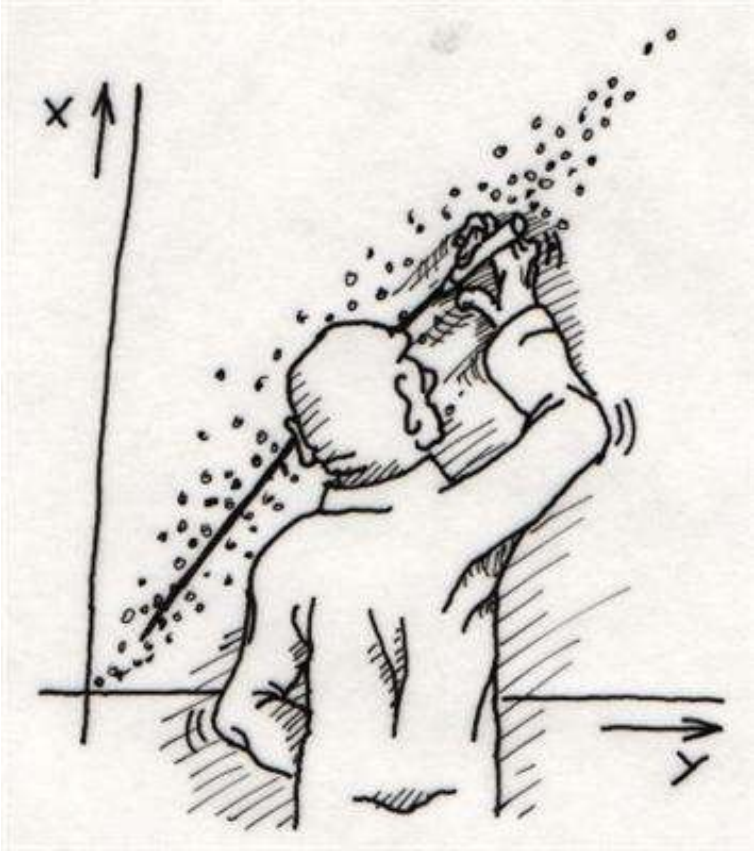


Feature extraction



Optical flow

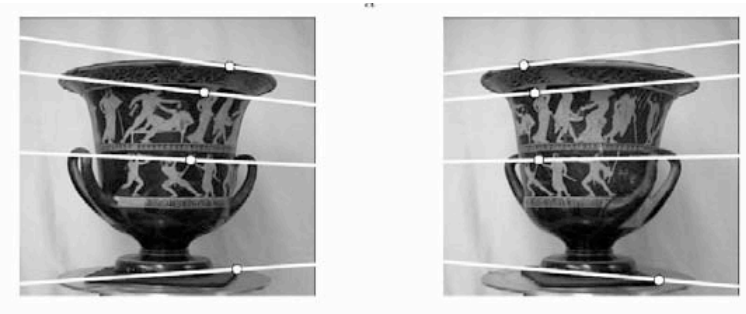
II. “Mid-level vision”



Fitting: Least squares
Voting methods

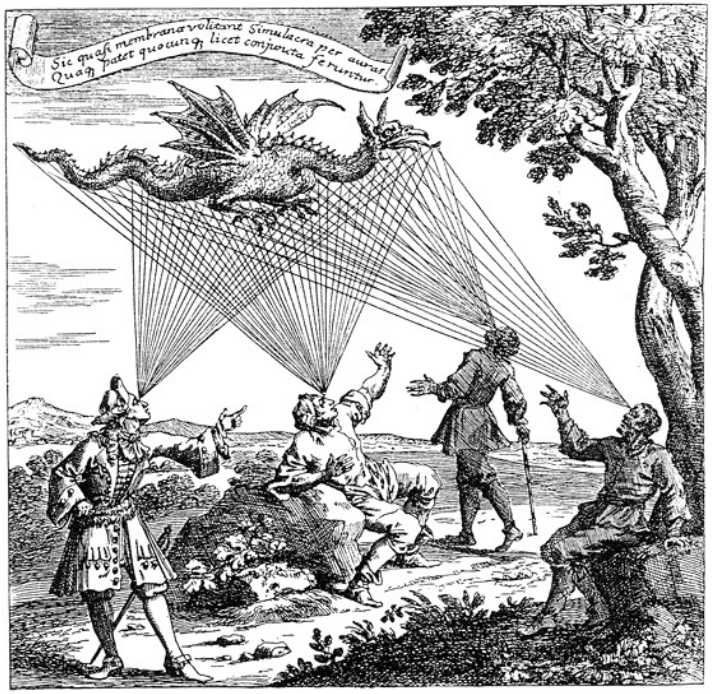
Alignment

III. Multi-view geometry



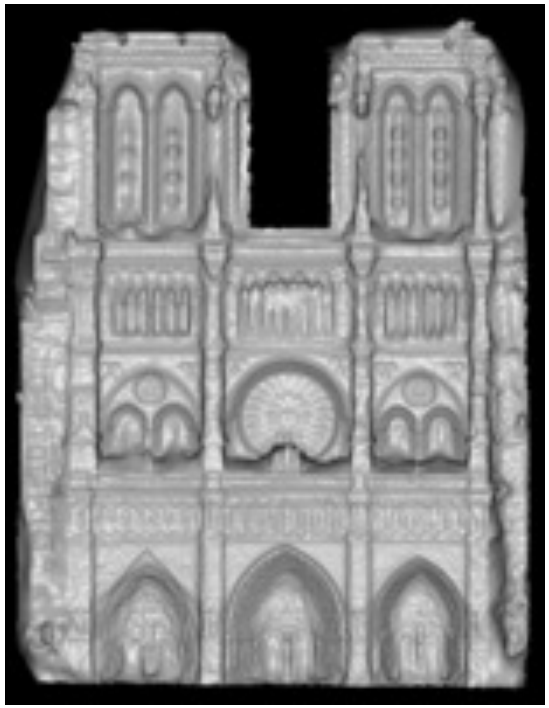
Epipolar geometry

Two-view stereo



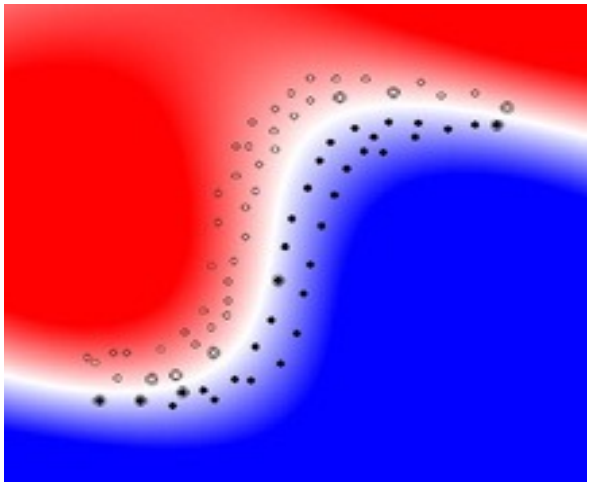
Драконъ, видимый подъ различными углами зрѣнія
По гравюрѣ на мѣди изъ „Oculus artificialis teledioptricus" Цана. 1702 года.

Structure from motion

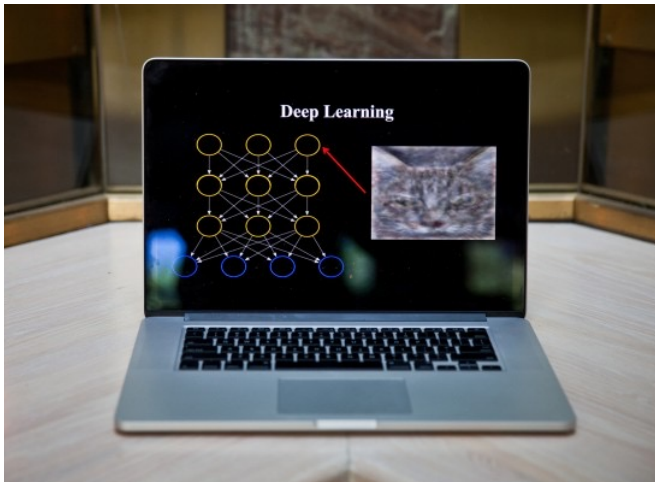


Multi-view stereo

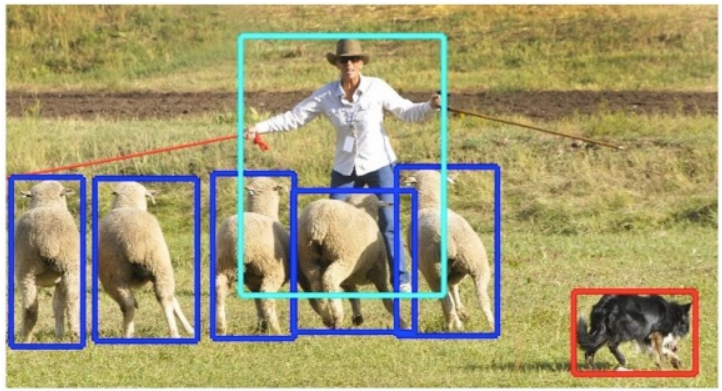
IV. Recognition



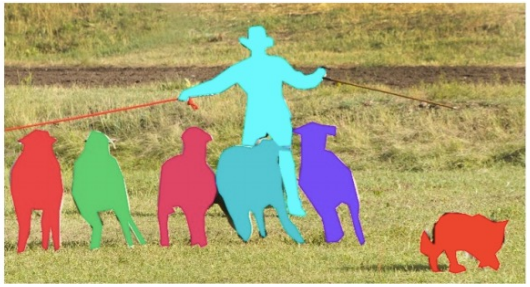
Basic classification



Deep learning



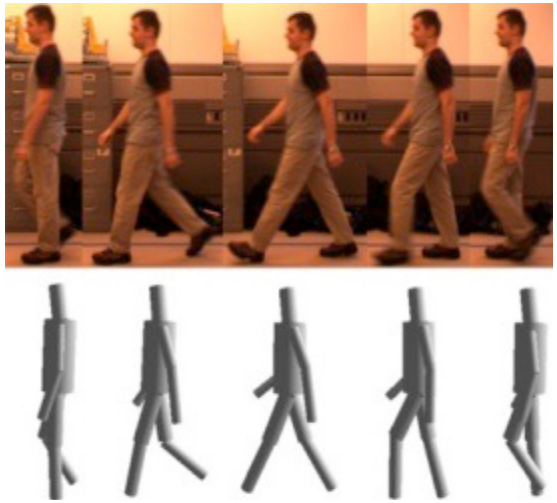
Object detection



Segmentation

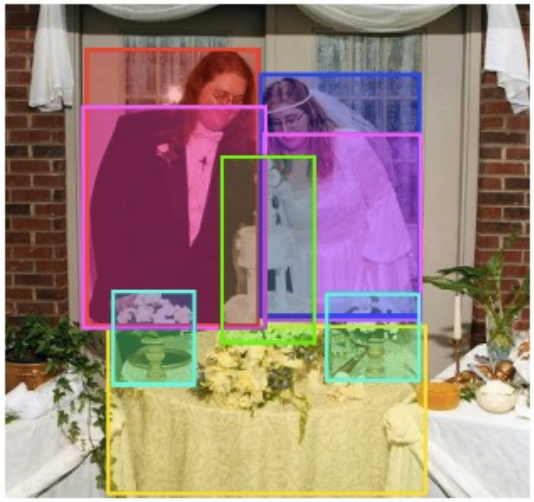
V. Additional Topics (time permitting)

Generation



Video

3D scene understanding



Images and text