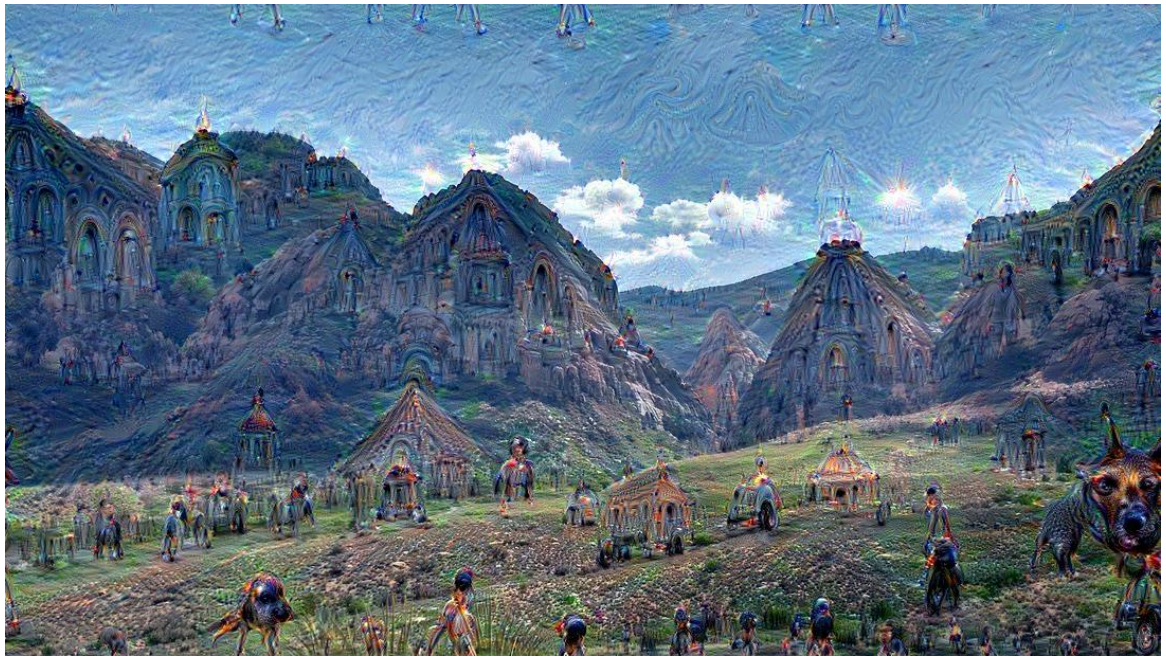
CS585 Image and Video Computing Spring 2025



Deep Dream

Instructor: Mahir Patel

Course Website

Example Piazza question/responses

Q: Is this right?

A: Well, let's think about the expected results if it was right… (i.e., we won't answer this question directly)

Q: What is *some concept*?

A: This is discussed in the book in section X and/or lecture Y.

Better question: Think about what part is confusing you. Be as specific as possible. Ask follow-up questions if necessary.

Q: Some logistics related question that has been asked before

A: *point you to the previously asked question*

You will get your questions answered sooner if you search for if anyone has already asked it

slido

Please download and install the Slido app on all computers you use

What do you think of when you hear the words "Computer Vision"?

ⓘ Start presenting to display the poll results on this slide.

Outline

- Logistics, requirements
- Goal of computer vision and why it is hard
- History of computer vision
- Recent work in vision
- Topics covered in class

slides adapted from Svetlana Lazebnik

Goal: To extract meaning from pixels

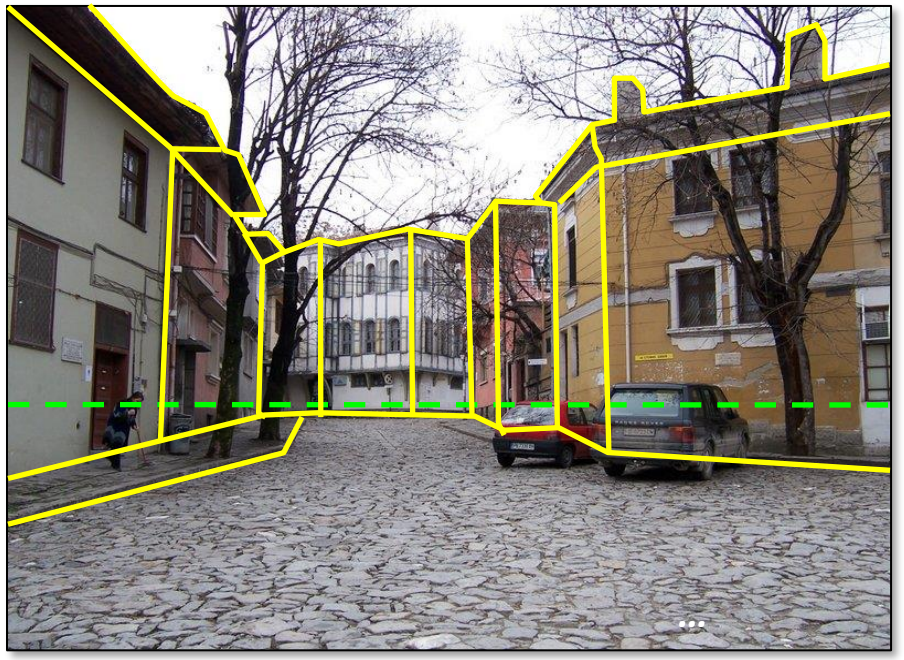
What we see

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

What kind of information can be extracted from an image?

What kind of information can be extracted from an image?



Difficulty

Geometric information

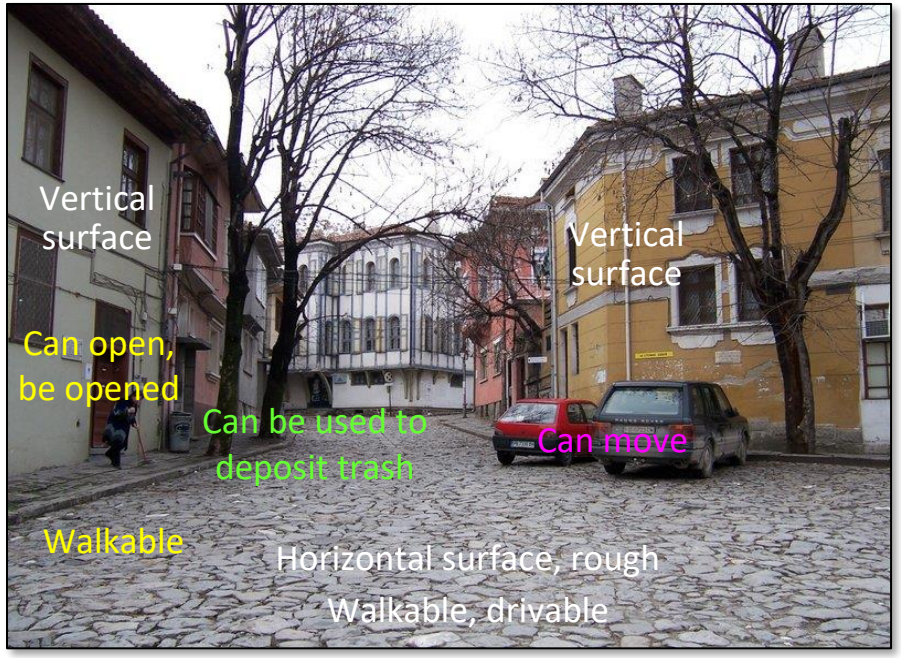
What kind of information can be extracted from an image?

Difficulty

Geometric information

Semantic information

What kind of information can be extracted from an image?



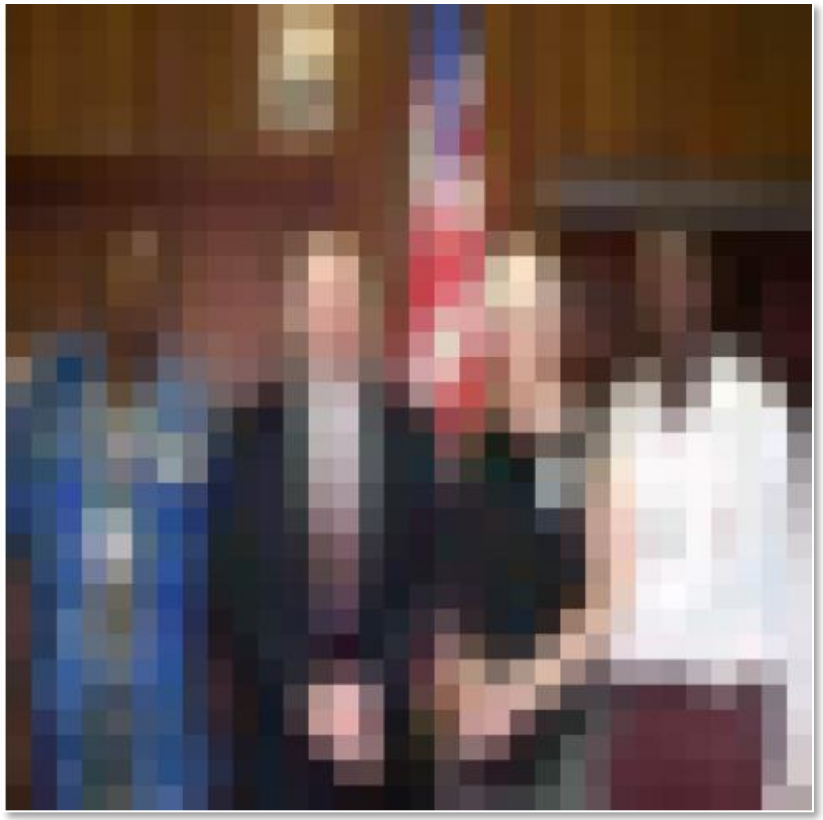
Difficulty

Geometric information

Semantic information

Affordances

Humans are remarkably good at vision…



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

…still, vision is hard even for humans

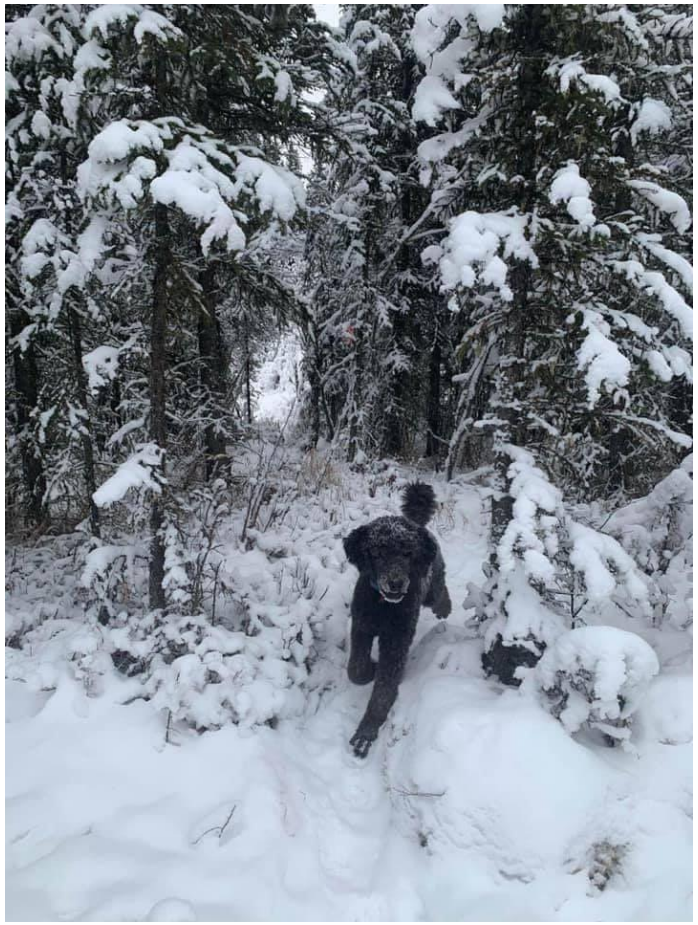


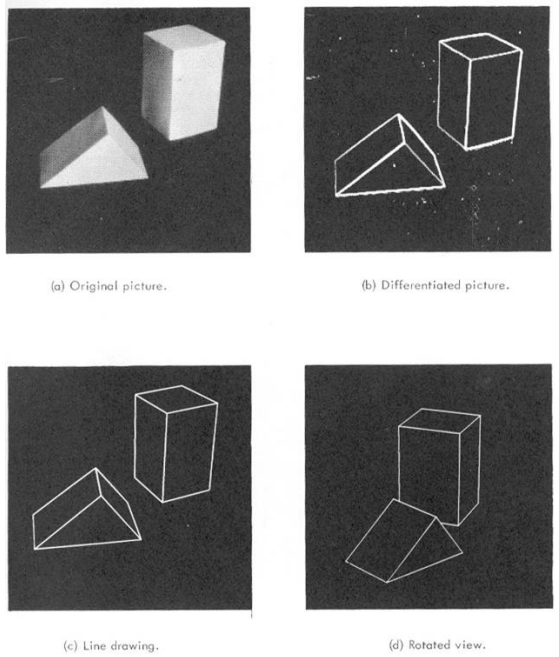
Image source

Images are fundamentally ambiguous!

Outline

- Logistics, requirements
- Goal of computer vision and why it is hard
- History of computer vision

How it started



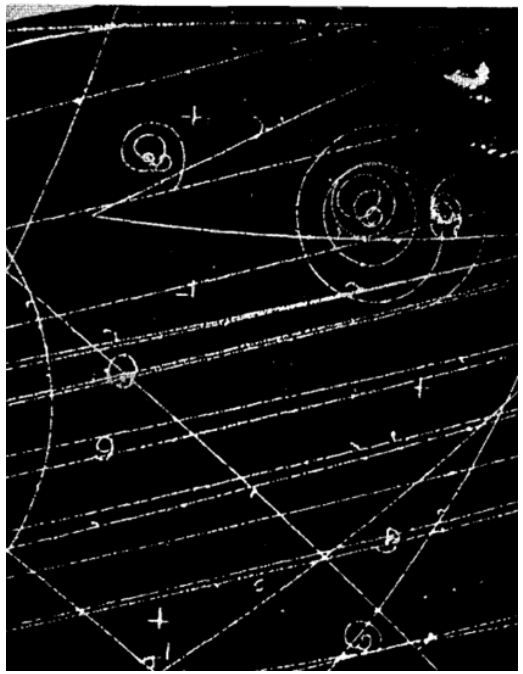
L. G. Roberts, 1963

How it's going

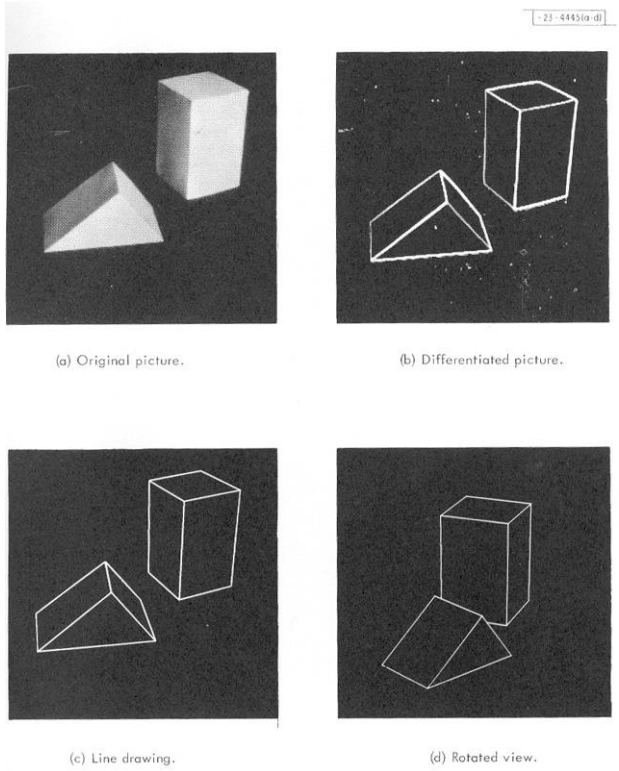
Prompt: a painting of a fox sitting in a field at sunrise in the style of Claude Monet

OpenAI DALL-E-2, 2022

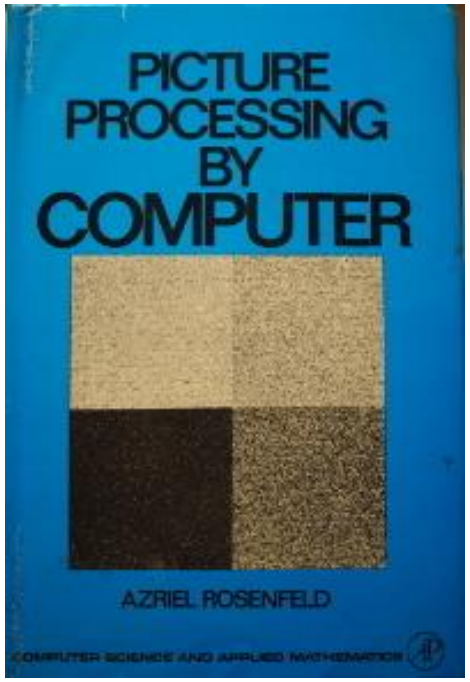
Origins



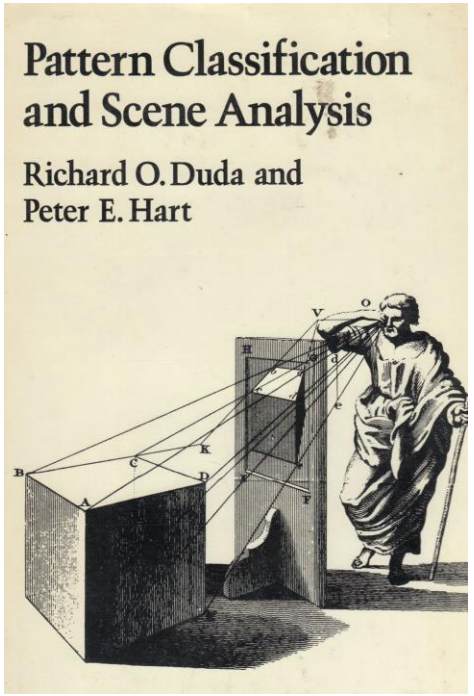
Hough, 1959



Roberts, 1963



Rosenfeld, 1969



Duda & Hart, 1972

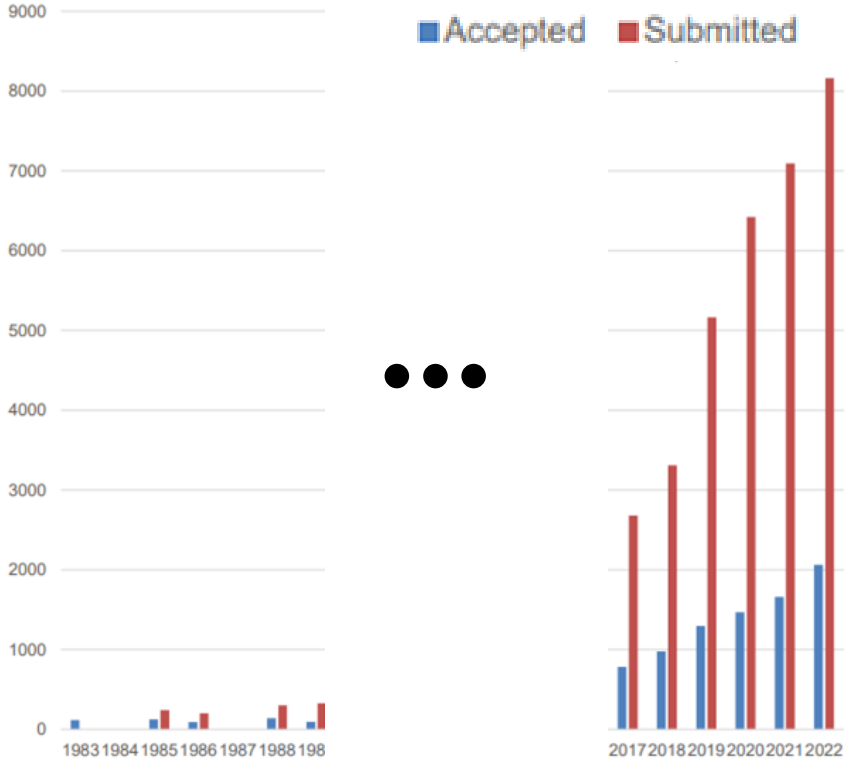
Decade by decade

- **1960s**: Blocks world, image processing and pattern recognition
- **1970s**: Key recovery problems defined: structure from motion, stereo, shape from shading, color constancy. Attempts at knowledge-based recognition
- **1980s**: Fundamental and essential matrix, multi-scale analysis, corner and edge detection, optical flow, geometric recognition as alignment
- **1990s**: Multi-view geometry, statistical and appearance-based models for recognition, first approaches for (class-specific) object detection
- **2000s**: Local features, generic object recognition and detection
- **2010s**: Deep learning, big data

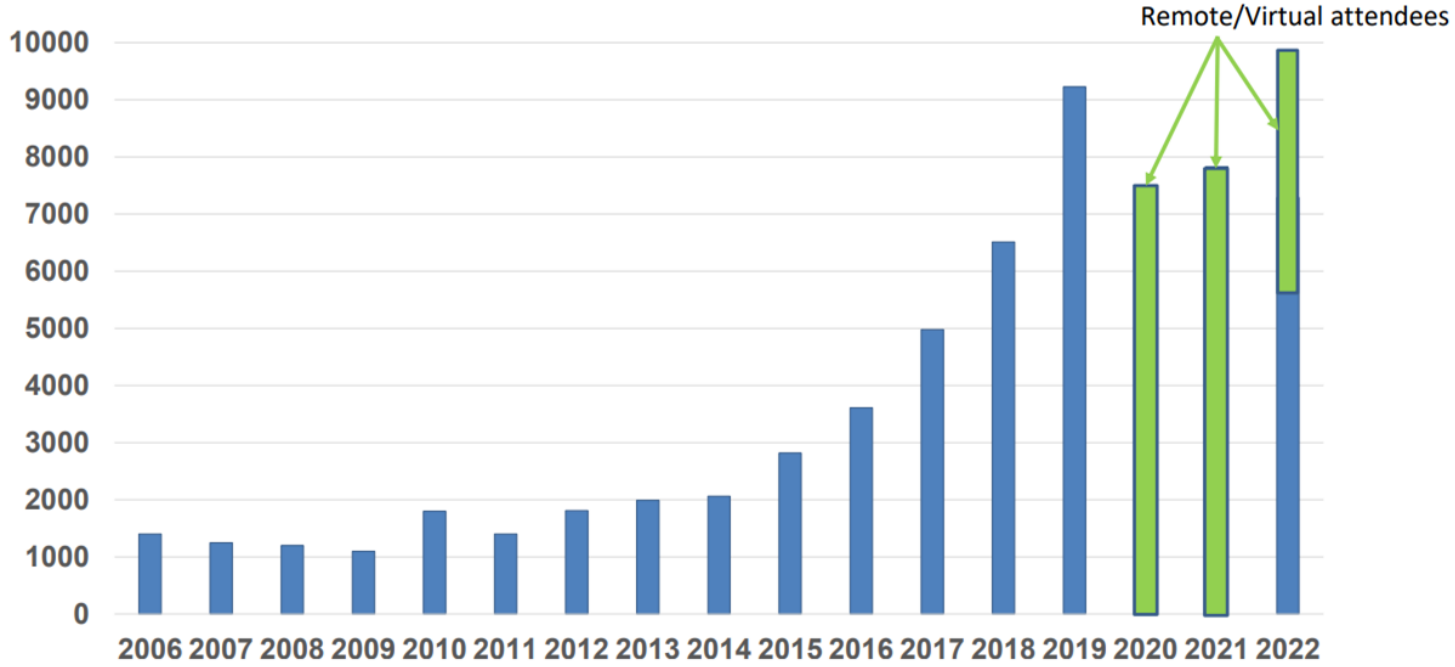
Adapted from J. Malik

Growth of the field

CVPR Papers



CVPR Attendees

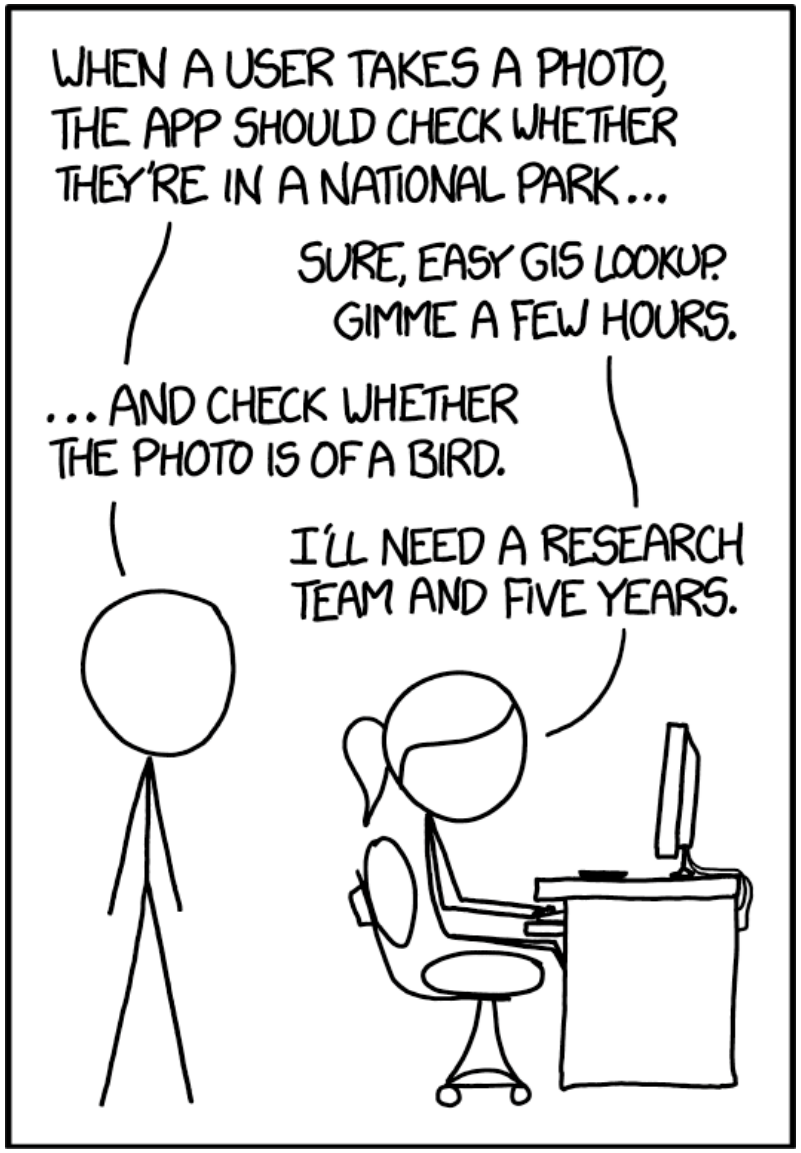


Source: CVPR 2022 opening sides

Outline

- Logistics, requirements
- Goal of computer vision and why it is hard
- History of computer vision
- Recent work in vision

What can computer vision do today?



In the 60s, Marvin Minsky assigned a couple of undergrads to spend the summer programming a computer to use a camera to identify objects in a scene. He figured they'd have the problem solved by the end of the summer. Half a century later, we're still working on it.

https://xkcd.com/1425/

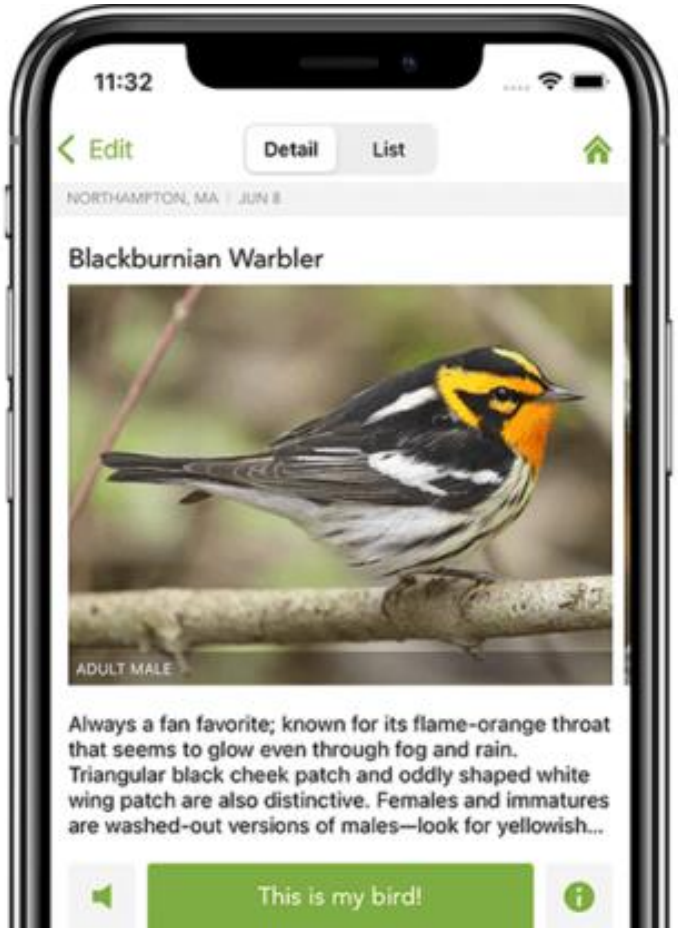
(September 24, 2014)

What can computer vision do today?

- It’s 2025 now…

TheCornellLab

Merlin®



https://merlin.allaboutbirds.org/

Reconstruction: 3D from photo collections

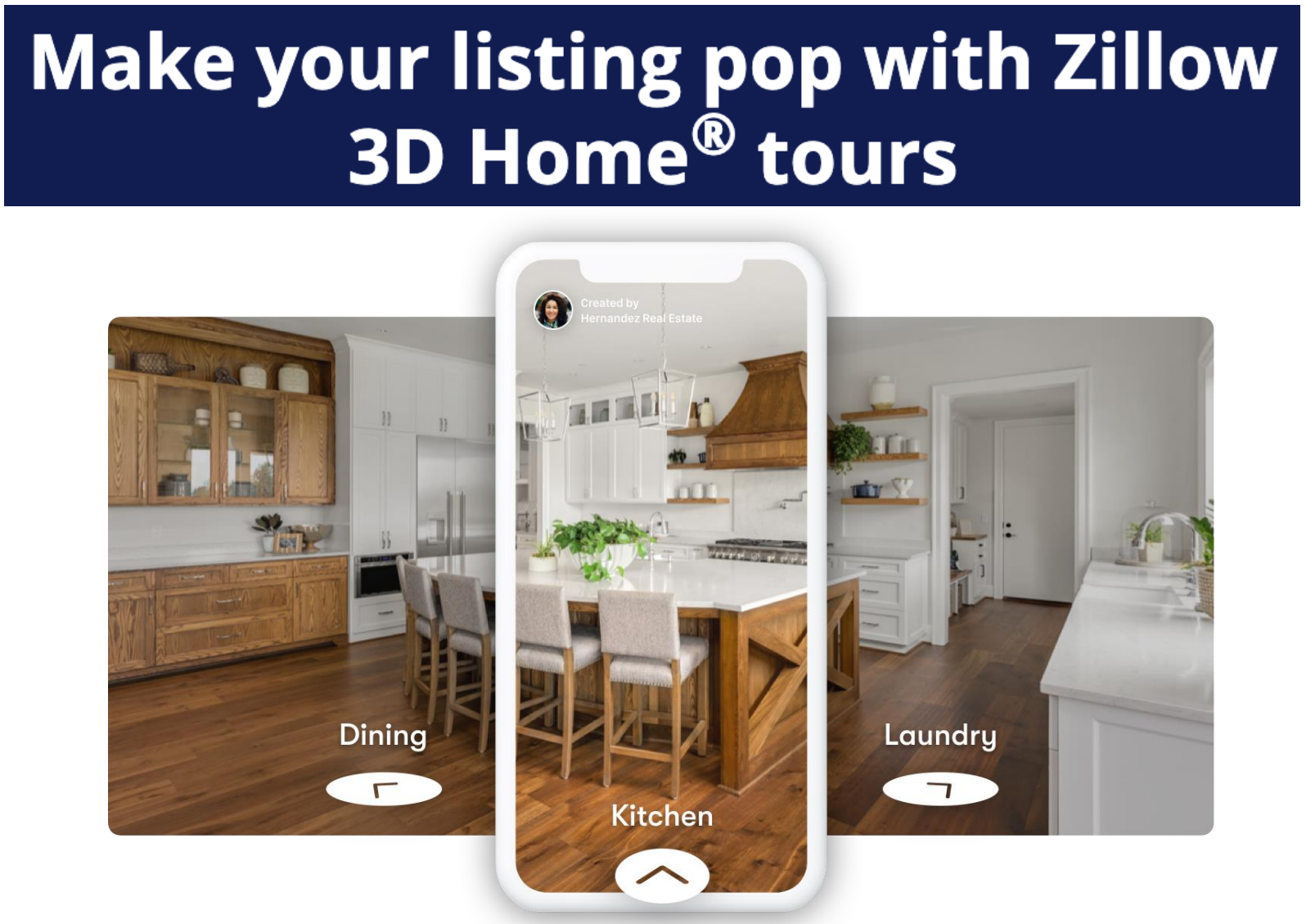
Colosseum, Rome, Italy

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: Commercial applications



https://www.zillow.com/z/3d-home/

Reconstruction: Commercial applications

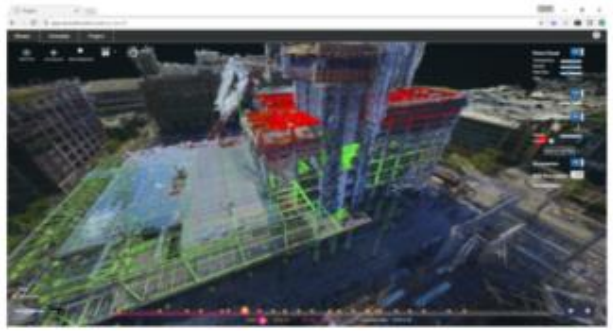
RECONSTRUCT INTEGRATES REALITY AND PLAN

Visual Asset Management

Reconstruct 4D point clouds and organize images and videos from smartphones, time-lapse cameras, and drones around the project schedule. View, annotate, and share anywhere with a web interface.

4D Visual Production Models

Integrate 4D point clouds with 4D BIM, review "who does what work at what location" on a daily basis and improve coordination and communication among project teams.



Predictive Visual Data Analytics

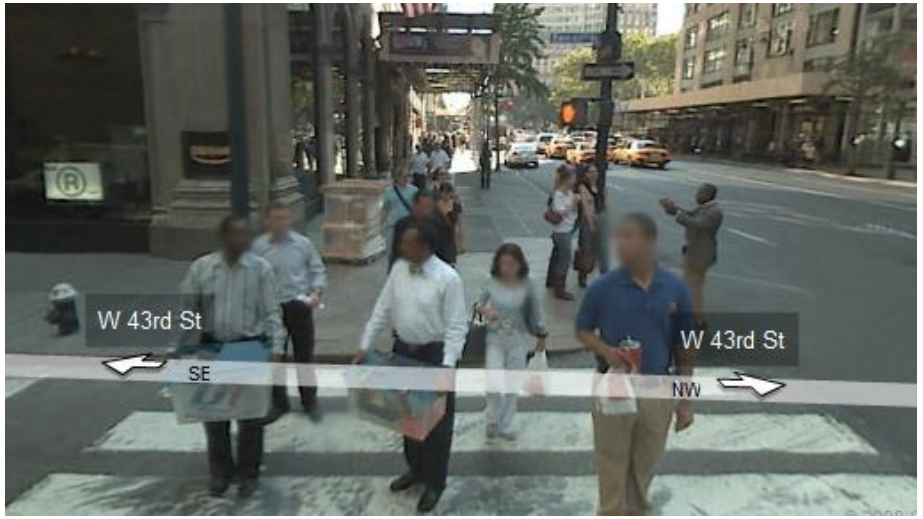
Analyze actual progress deviations by comparing Reality and Plan and predict risk with respect to the execution of the look-ahead schedule for each project location, to offer your project team with an opportunity to tap off potential delays before they surface on your jobsite.

reconstructinc.com

Source: D. Hoiem

Recognition: “Simple” patterns

Recognition: Faces



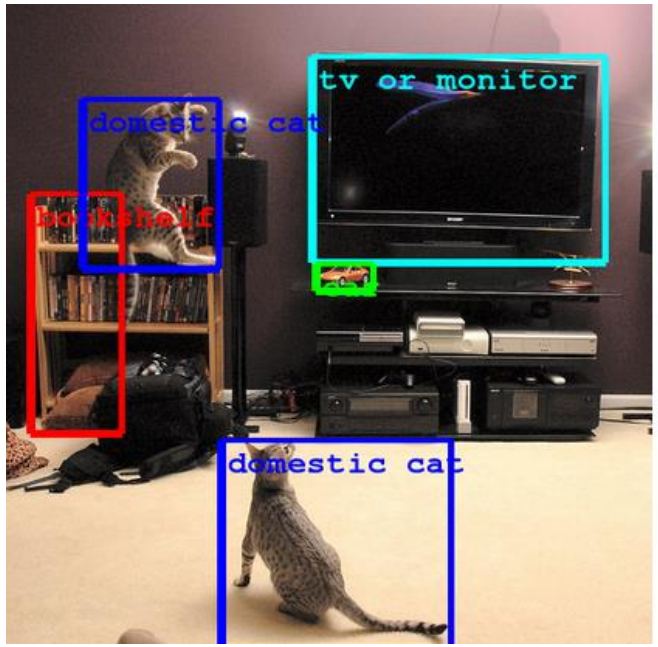
Recognition: Faces

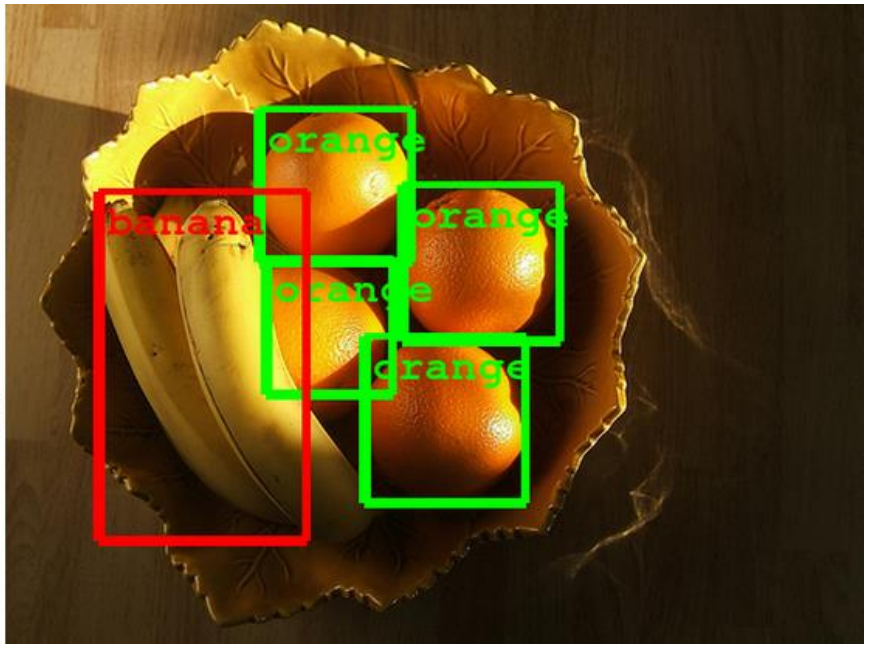
[How China Uses High-Tech Surveillance to Subdue Minorities] – New York Times, 5/22/2019

[The Secretive Company That Might End Privacy As We Know It] – New York Times, 1/18/2020

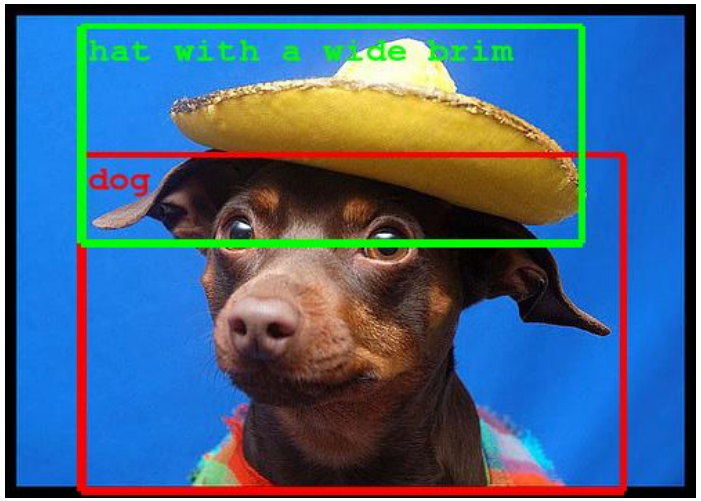
[Wrongfully Accused by an Algorithm] – New York Times, 6/24/2020

Recognition: General categories





- 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



Object detection, instance segmentation

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

DeepFakes

Harrison Ford Is Young Han In Solo Deepfake Video

Thanks to deepfake technology, the maligned Solo: A Star Wars Story now stars Harrison Ford instead of Alden Ehrenreich as the young Han.

BY DAN ZINSKI
2 DAYS AGO

Just a random example…

https://screenrant.com/star-wars-han-solo-movie-harrison-ford-video-deepfake/

https://www.youtube.com/watch?v=bC3uH4Xw4Xo

https://en.wikipedia.org/wiki/Deepfake

Image generation: General categories

- BigGAN: Synthesize ImageNet images, conditioned on class label, up to 512 x 512 resolution

Difficult classes

A. Brock, J. Donahue, K. Simonyan, Large scale GAN training for high fidelity natural image synthesis, ICLR 2019

Text-to-image generation: OpenAI DALL-E

- Learn a joint sequential transformer model that can be used to generate image based on text prompt

(a) a tapir made of accordion. a tapir with the texture of an accordion.

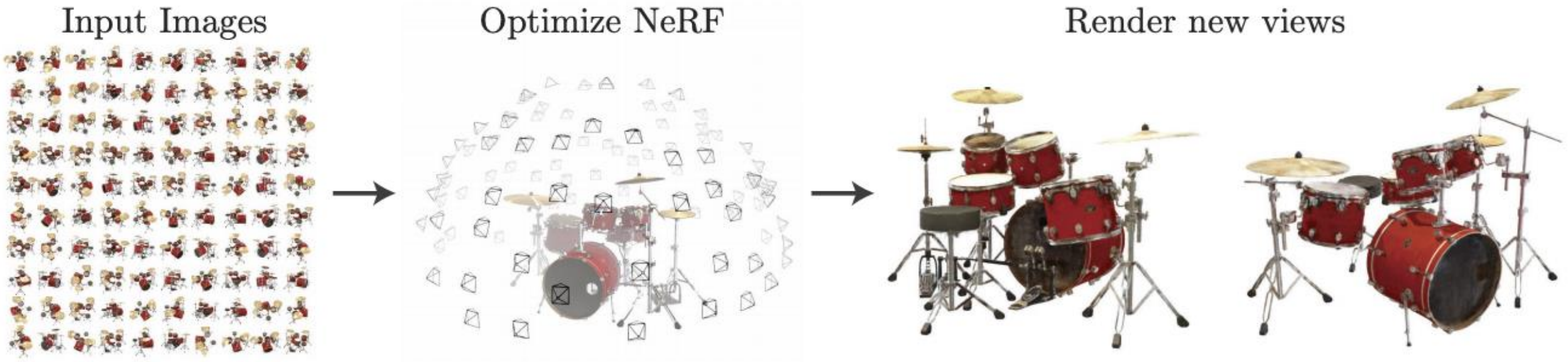
(b) an illustration of a baby hedgehog in a christmas sweater walking a dog

(c) a neon sign that reads "backprop". a neon sign that reads "backprop". backprop neon sign

A. Ramesh et al., Zero-Shot Text-to-Image Generation, arXiv 2021

https://openai.com/blog/dall-e/

3D scene understanding: NERFs



B. Mildenhall et al., Representing Scenes as Neural Radiance Fields for View Synthesis, ECCV 2020

Follow-up work

Physical scene understanding: Learning skills from video

Fig. 1. Simulated characters performing highly dynamic skills learned by imitating video clips of human demonstrations. **Left:** Humanoid performing cartwheel B on irregular terrain. **Right:** Backflip A retargeted to a simulated Atlas robot.

Video

X. B. Peng, A. Kanazawa, J. Malik, P. Abbeel, S. Levine, SFV: Reinforcement Learning of Physical Skills from Videos, SIGGRAPH Asia 2018

Outline

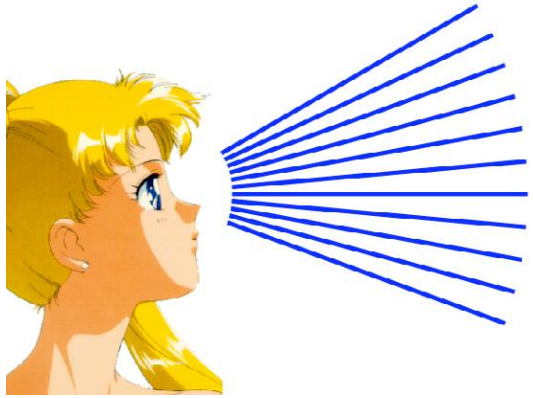
- Logistics, requirements
- Goal of computer vision and why it is hard
- History of computer vision
- Recent work in vision
- Topics covered in class

Topics covered in class

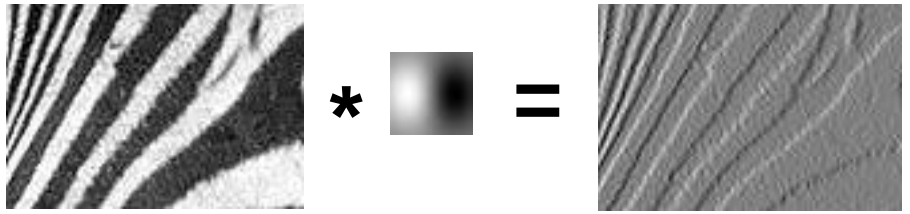
I. Early vision: Image formation and processing
II. Recognition
III. Mid-level vision: Grouping and fitting
IV. Multi-view geometry

I. Early vision

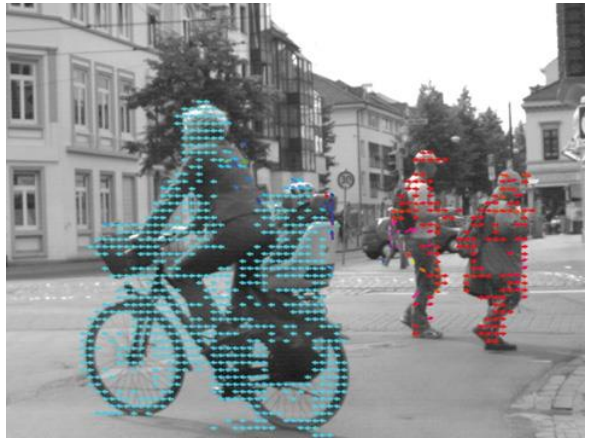
Basic image formation and processing



Cameras and sensors
Light and color

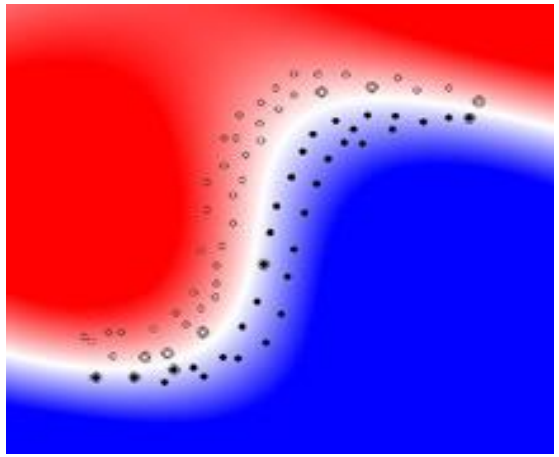


Linear filtering
Edge detection

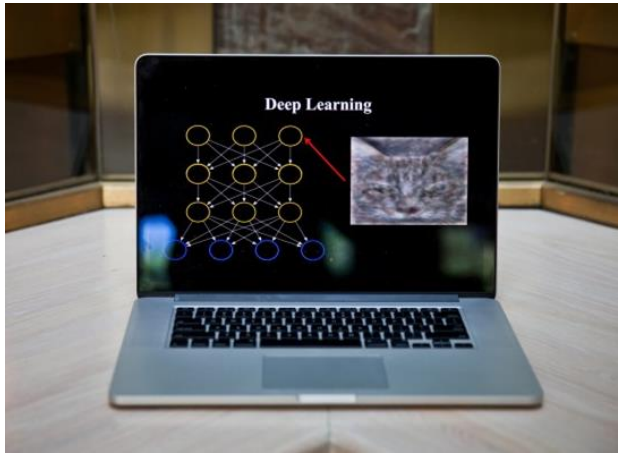


Optical flow

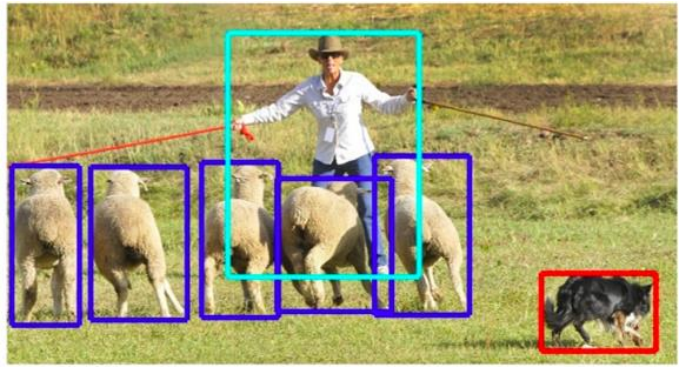
II. Recognition



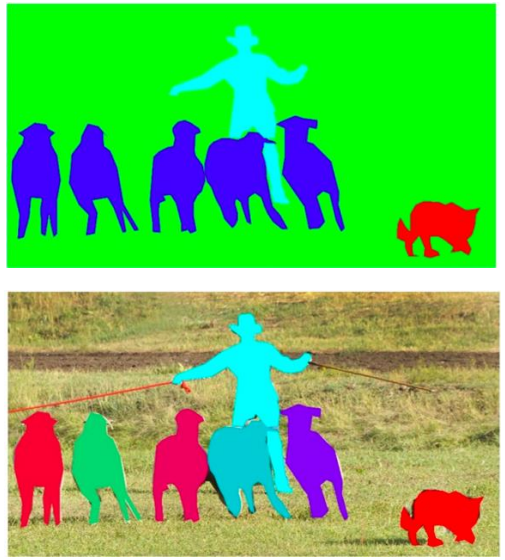
Basic classification



Deep learning



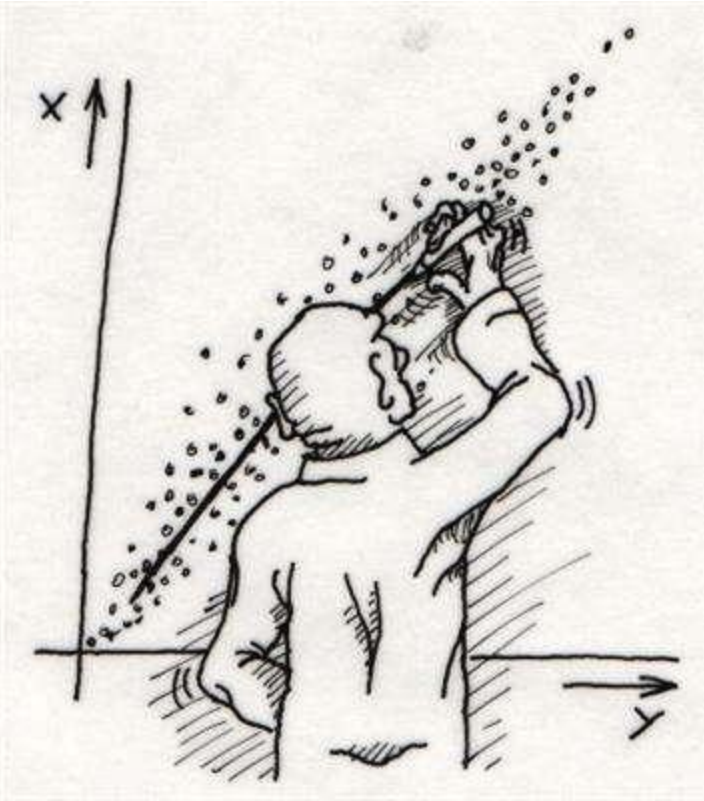
Object detection



Segmentation

III. “Mid-level vision”

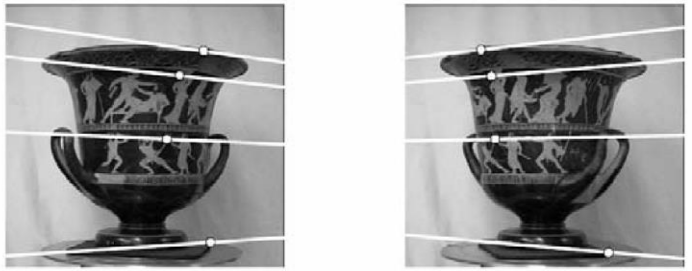
Fitting and grouping



Fitting: Least squares
Voting methods

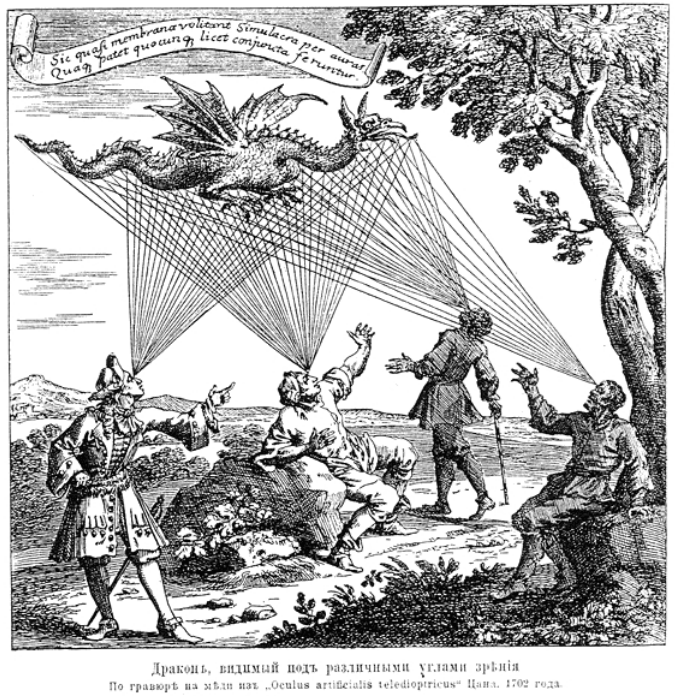
Alignment

IV. Multi-view geometry



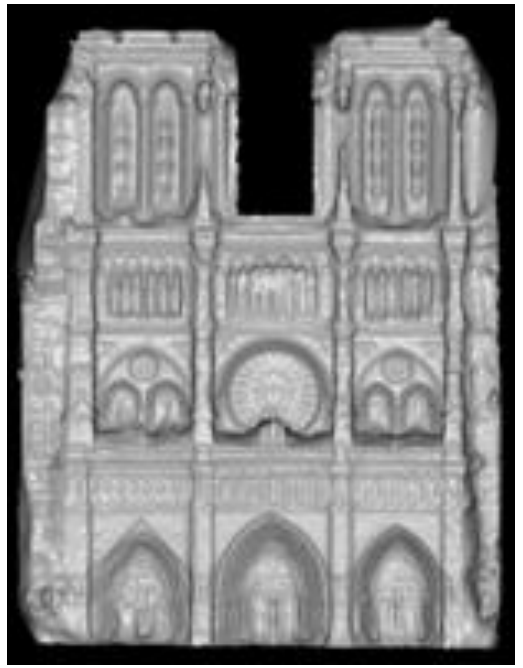
Epipolar geometry

Two-view stereo



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

Structure from motion



Multi-view stereo

Next Class

Image formation: geometric primitives and transformations, photometric image formation, digital camera

Reading: SZ Ch 2