CS543 / ECE549
Computer Vision
Spring 2024

Course webpage URL:

http://saurabhg.web.illinois.edu/teaching/ece549/sp2024/

Image as generated using Stable Diffusion 2:
https://huggingface.co/stabilityai/stable-diffusion-2-1?text=Logo+for+a+Computer+Vision+Class

Plan for today

- Course Introduction
  - What is computer vision
  - Vision is easy for humans, yet has proven hard for computers
  - What can computer vision do today?
  - The six decades of computer vision
- Course Overview
- Logistics

Computer Vision

To extract “meaning” from pixels

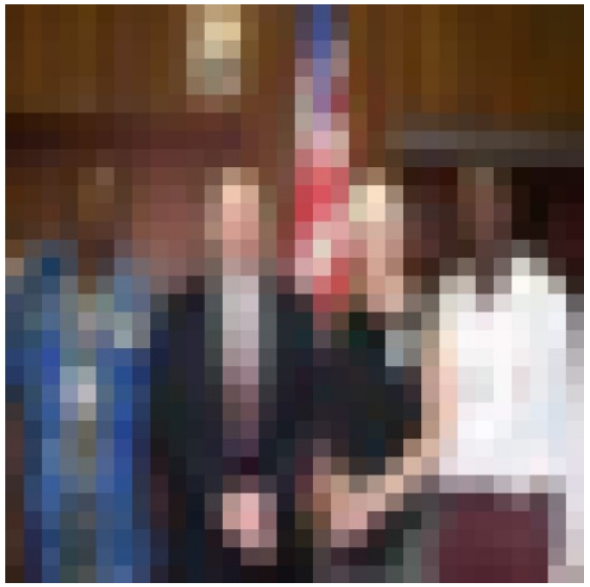
person, motorcycle, car, chair

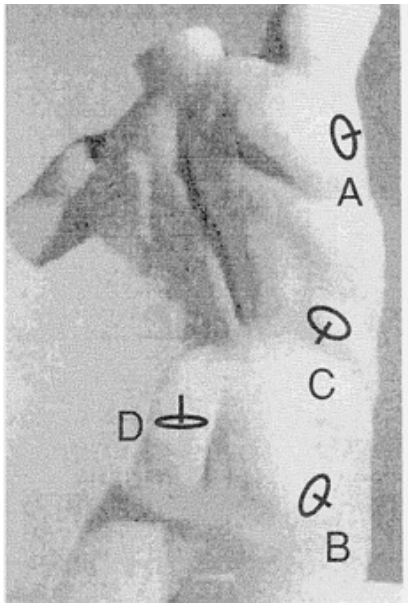
Meaning can take different forms:

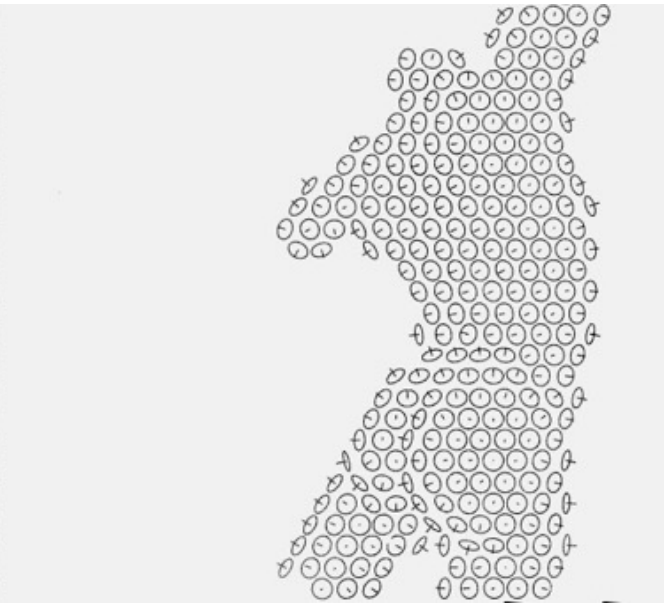
- Geometric Inferences
- Semantic Inferences
- Inferences about actions
- …

Computer vision is easy for humans

- Effortlessly analyze images for a variety of tasks
- Infer semantics even from severely ablated input
- Can also make precise inference about certain geometric properties







Yet has proven very hard for computers

- Computer vision research easily goes back 60 years …

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

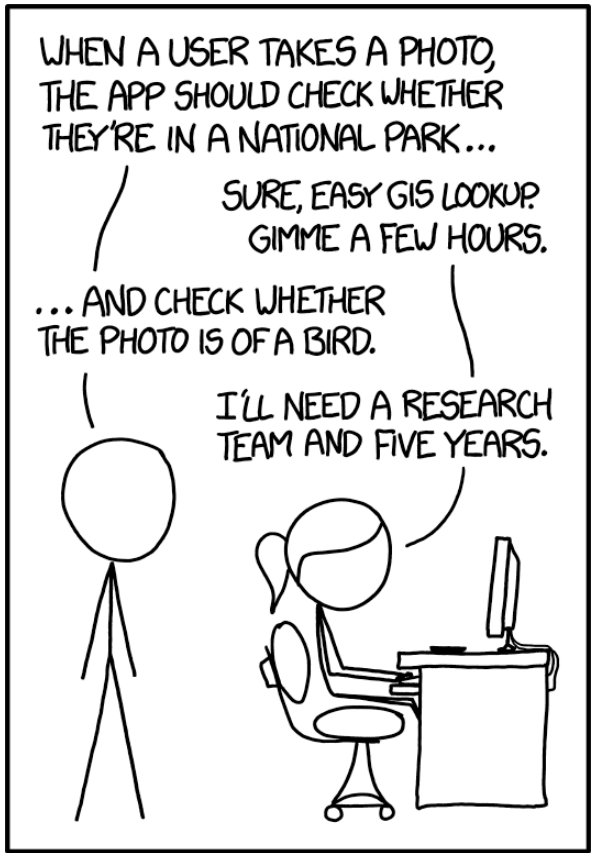
Artificial Intelligence Group
Vision Memo. No. 100.

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 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".

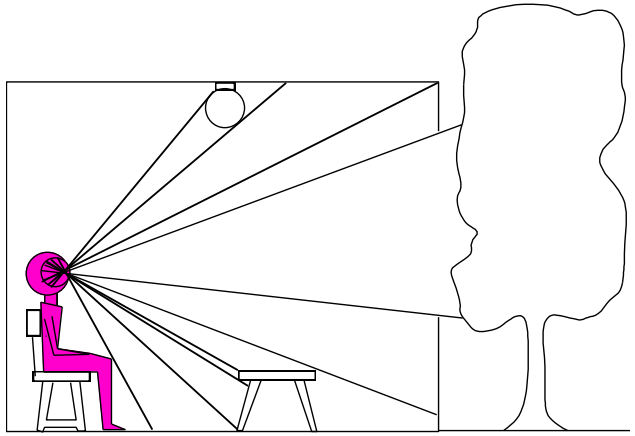


- Entirely true as of 2014 (or so) when this xkcd was published

Why is computer vision hard?

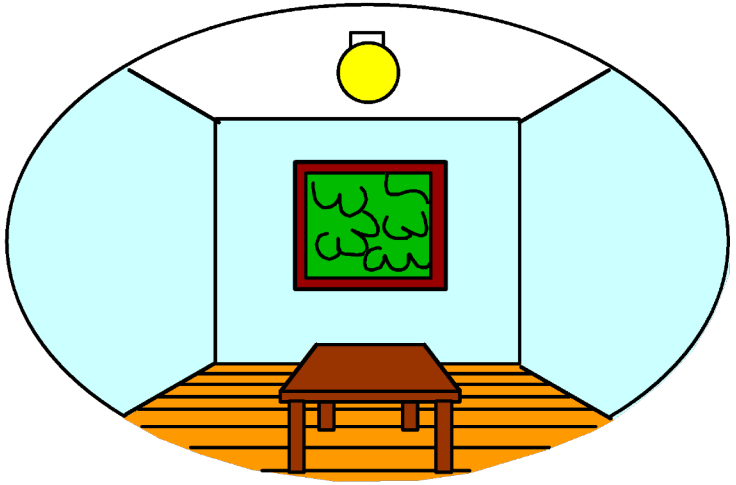
- Images are a lossy projection of the world

3D world



Point of observation

2D image



Geometry information is lost

Why is computer vision hard?

- Images are a lossy projection of the world

What color is the dress?

A) Black and blue

B) White and gold?

Appearance information is also lost

Slide by L. Lazebnik

https://www.wired.com/2015/02/science-one-agrees-color-dress/

Why is computer vision hard?

- Images are a lossy projection of the world

Might cause objects to blend

Why is computer vision hard?

- Images are a lossy projection of the world (geometry, appearance, … are lost)
- Visual world is diverse

Viewpoint variation

Shape variation

Why is computer vision hard?

- Images are a lossy projection of the world (geometry, appearance, … are lost)
- Visual world is diverse

Background clutter

Occlusion

Why is computer vision hard?

- Images are a lossy projection of the world (geometry, appearance, … are lost)
  - need some priors to interpret what you are seeing
- Visual world is diverse
  - can’t write down these priors by hand

John's Diner with John's Chevelle, 2007

What can computer vision do today?

Reconstruction: 3D from photo collections

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

YouTube Video See also: NYTimes Article

Source: L. Lazebnik

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

Also see: NeRF

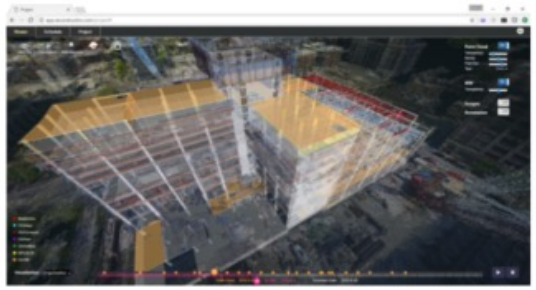
Source: L. Lazebnik

Reconstruction in construction industry

RECONSTRUCT INTEGRATES REALITY AND PLAN

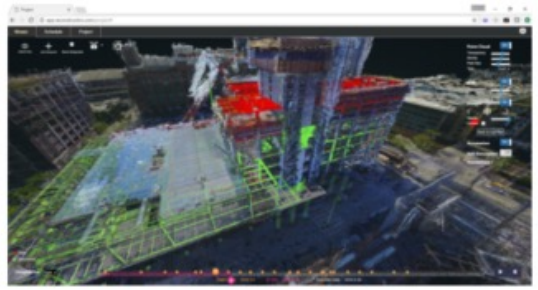
Visual Asset Management

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

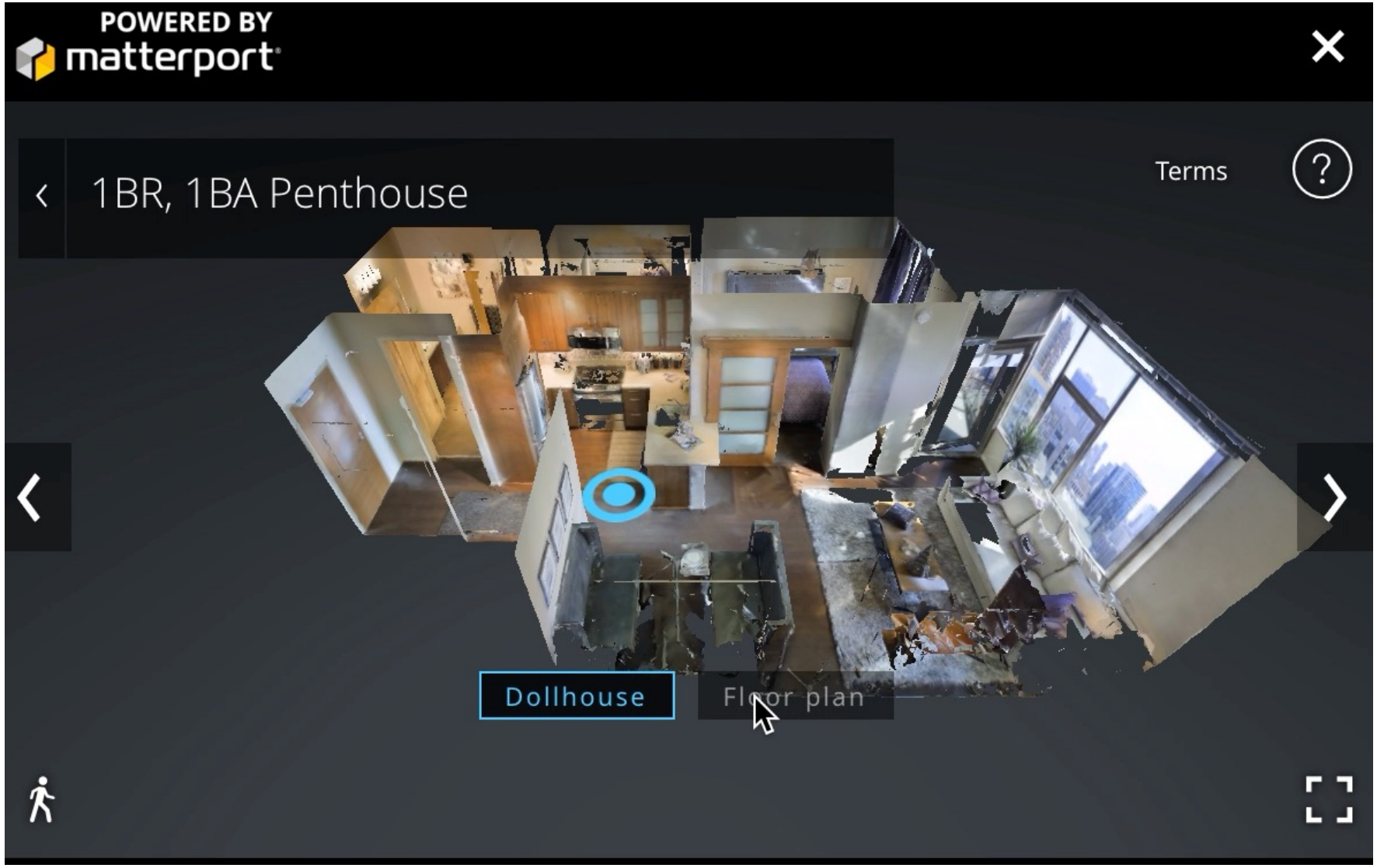
Analyze actual progress deviations by comparing Reality and Plan and predict risk with respect to the execution of the look-ahead schedule for each project location, to offer your project team with an opportunity to tap off potential delays before they surface on your jobsite.

reconstructinc.com

Source: L. Lazebnik

Source: D. Hoiem

Applications

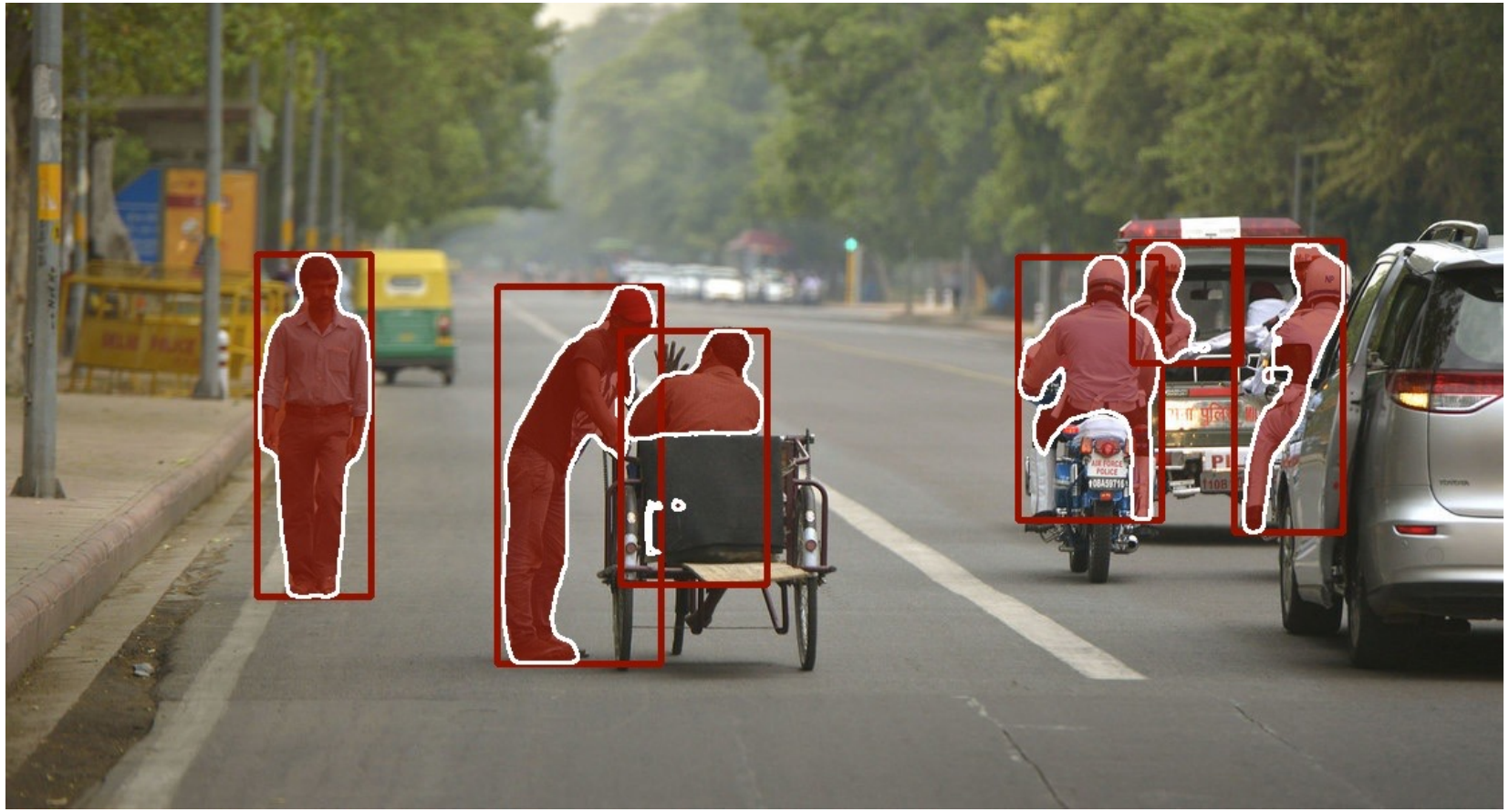


Source: N. Snavely

Novel View Synthesis

B. Mildenhall*, P. Srinivasan*, M. Tancik*, J. Barron, R. Ramamoorthi, R. Ng. Representing Scenes as Neural Radiance Fields for View Synthesis, ECCV 2020

Image Labeling Tasks



person, motorcycle, car, chair

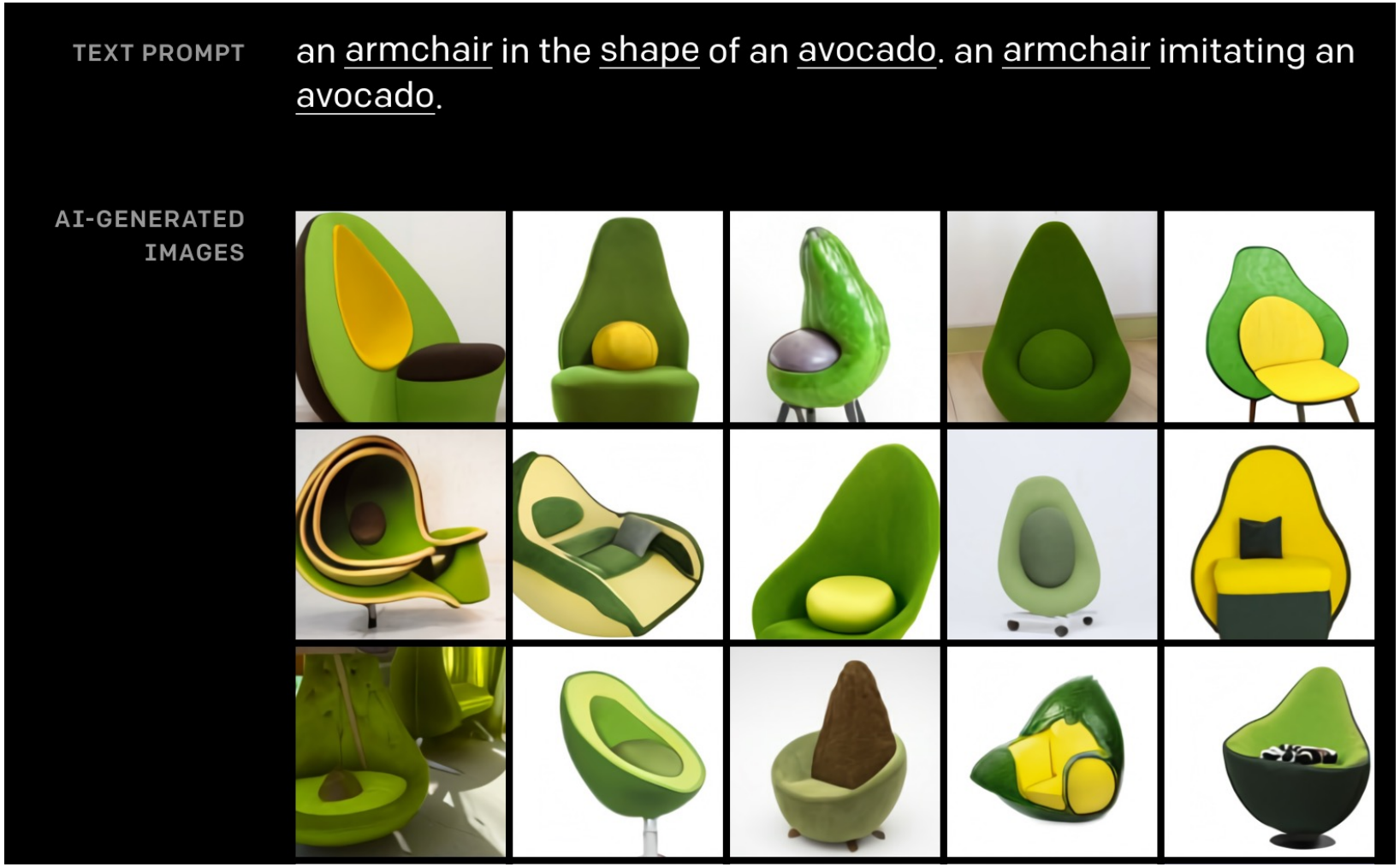
K. He et al. **Mask R-CNN** ICCV 2017

Full 3D from a single image

Shape and Pose Estimation for Objects and Humans

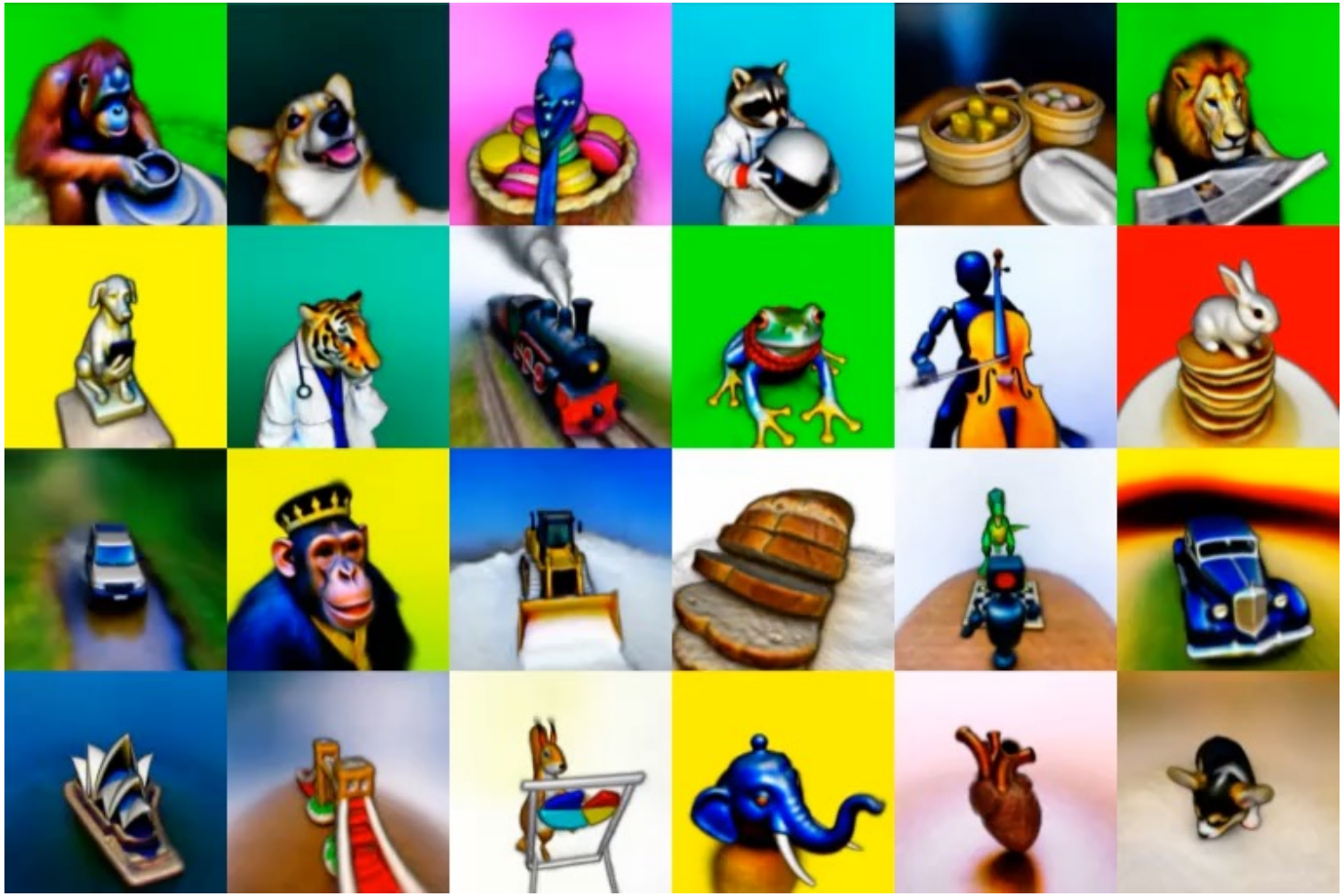
S. Goel et al. **Shape and Viewpoint without Keypoints.** ECCV 2020
A. Kanawaza et al. **End-to-end Recovery of Human Shape and Pose.** CVPR 2018

Image Generation



A. Ramesh et al. **Zero-Shot Text-to-Image Generation.** ICML 2021.

3D Scene Generation



B. Poole, A. Jain, J. Barron, B. Mildenhall. DreamFusion: Text-to-3D using 2D Diffusion. arXiv 2022

A. Ramesh et al. **Zero-Shot Text-to-Image Generation.** ICML 2021.

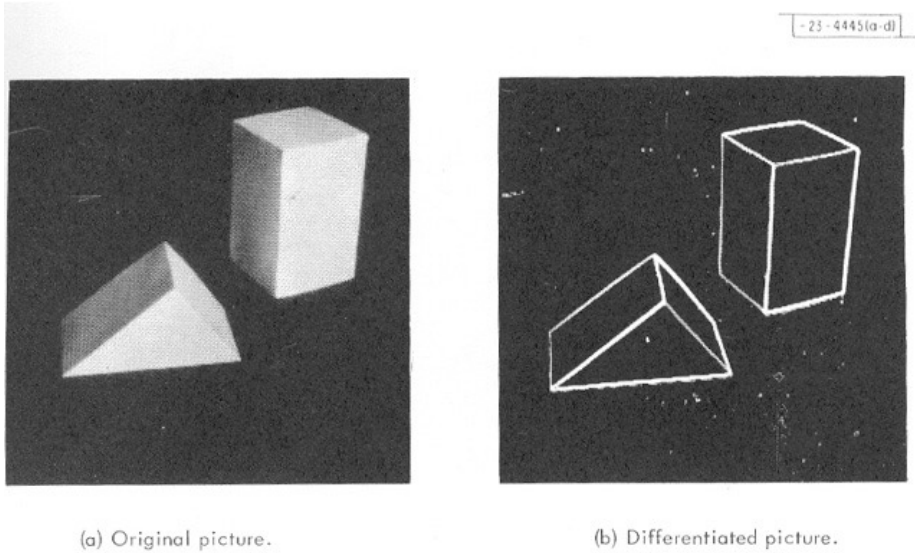
Video Generation

A teddy bear washing dishes

Jonathan Ho*, William Chan*, Chitwan Saharia*, Jay Whang*, Ruiqi Gao, Alexey Gritsenko, Diederik P. Kingma, Ben Poole, Mohammad Norouzi, David Fleet, Tim Salimans*. Imagen Video: High-definition video generation with diffusion models arXiv 2022

A. Ramesh et al. **Zero-Shot Text-to-Image Generation.** ICML 2021.

Origins of computer vision



(a) Original picture. (b) Differentiated picture.

(c) Line drawing. (d) Rotated view.

L. G. Roberts *Machine Perception of Three Dimensional Solids*

Source: L. Lazebnik

Origins of computer vision

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".

Source: L. Lazebnik

Six decades of computer vision

1960s: Beginnings in artificial intelligence, image processing and pattern recognition

1970s: Foundational work on image formation: Horn, Koenderink, Longuet-Higgins …

1980s: Vision as applied mathematics: geometry, multi-scale analysis, probabilistic modeling, control theory, optimization

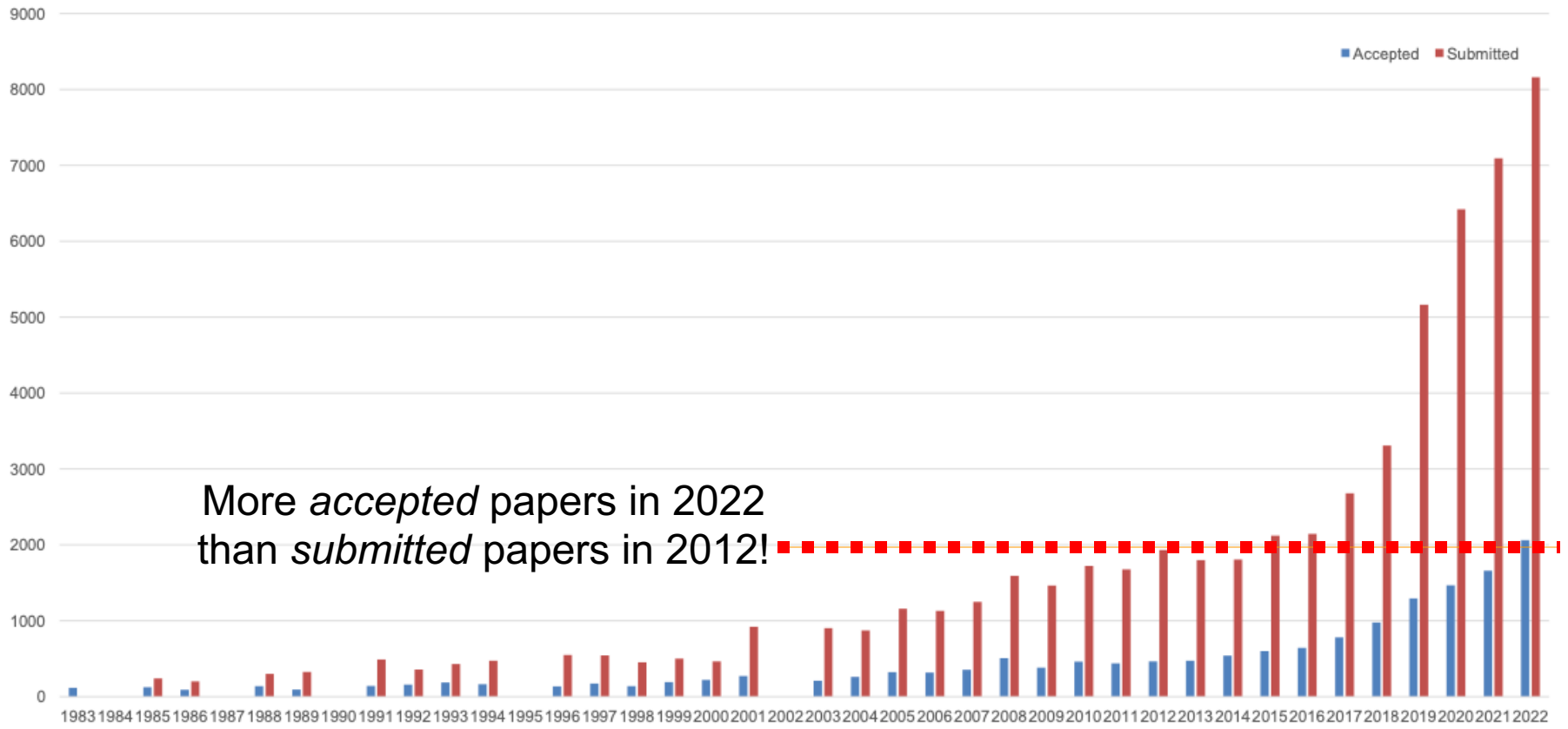
1990s: Geometric analysis largely completed, vision meets graphics, statistical learning approaches resurface

2000s: Significant advances in visual recognition

2010s: Progress continues, aided by the availability of large amounts of visual data and massive computing power. Deep learning has become pre-eminent

Source: L. Lazebnik

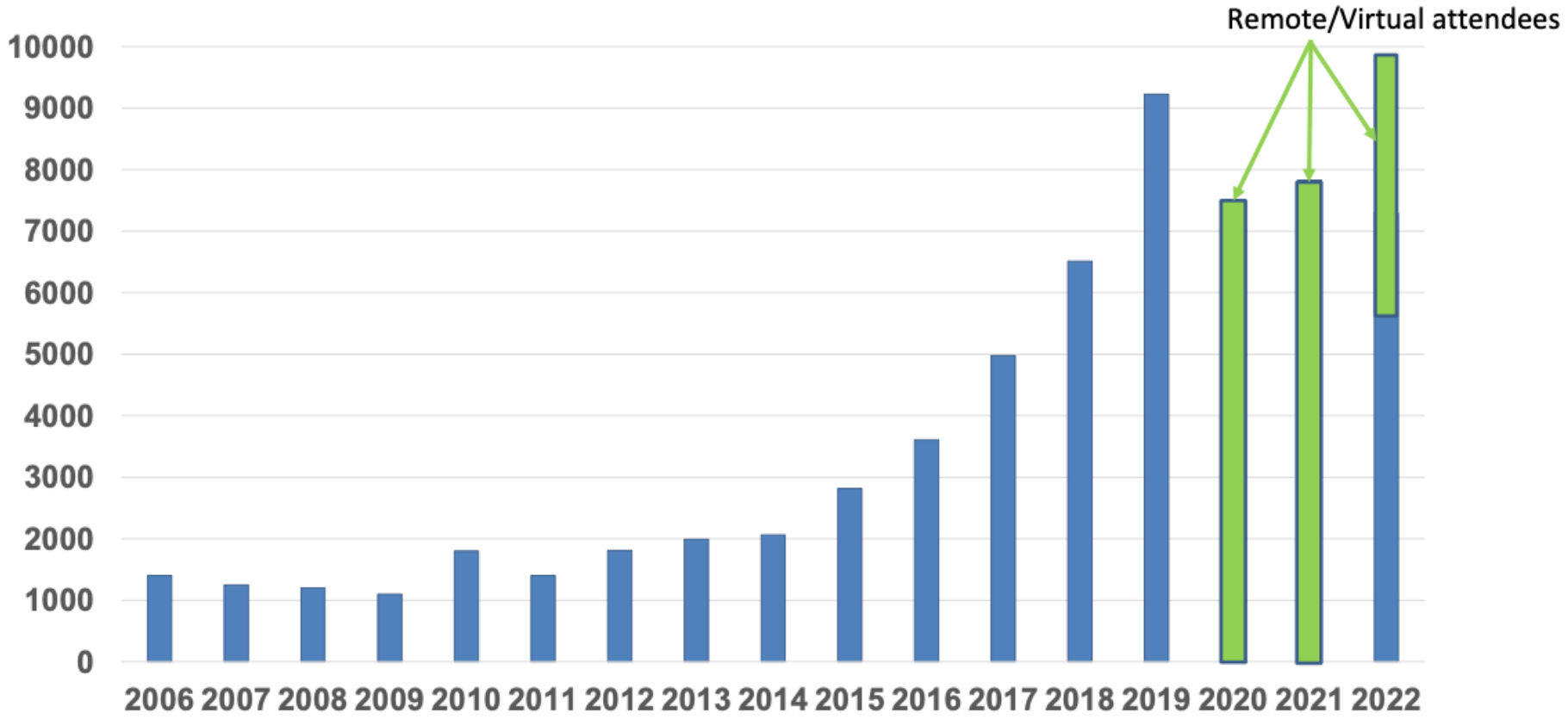
Growth of the field: CVPR papers



Source: L. Lazebnik

Source: CVPR 2022 opening sides

Growth of the field: CVPR attendance



Source: L. Lazebnik

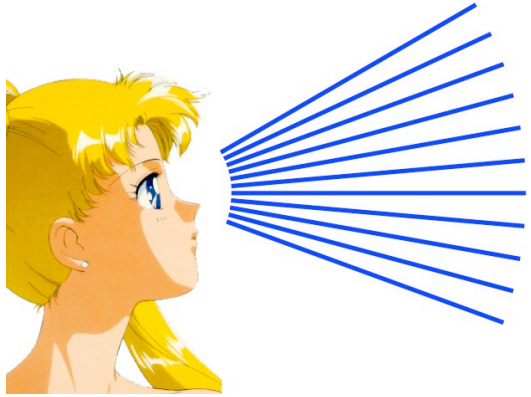
Source: CVPR 2022 opening sides

Course overview

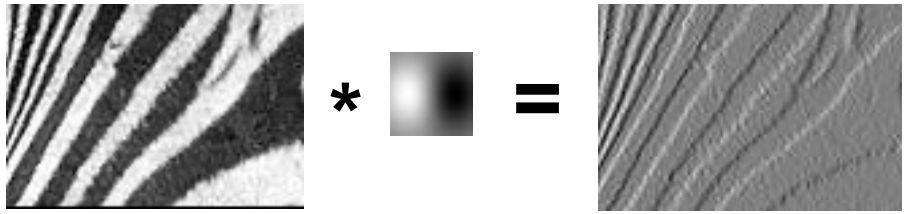
I. Early vision: Image formation and processing
II. Mid-level vision: Grouping and fitting
III. Multi-view geometry
IV. Recognition
V. Additional topics

I. Early vision

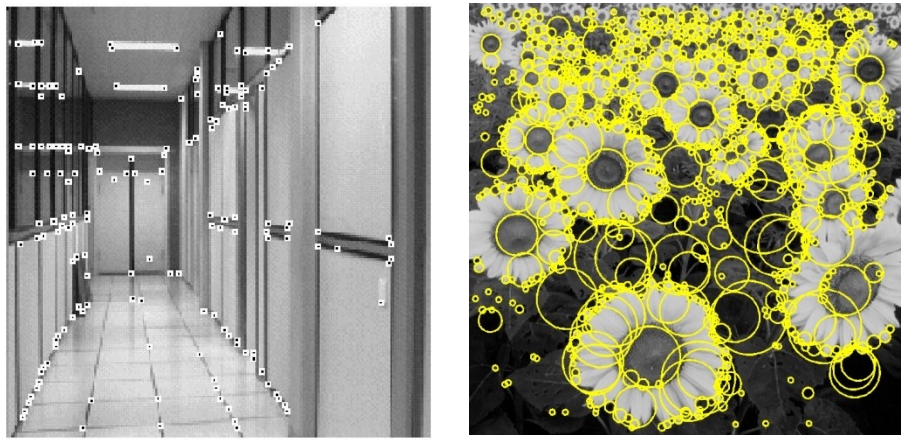
Basic image formation and processing



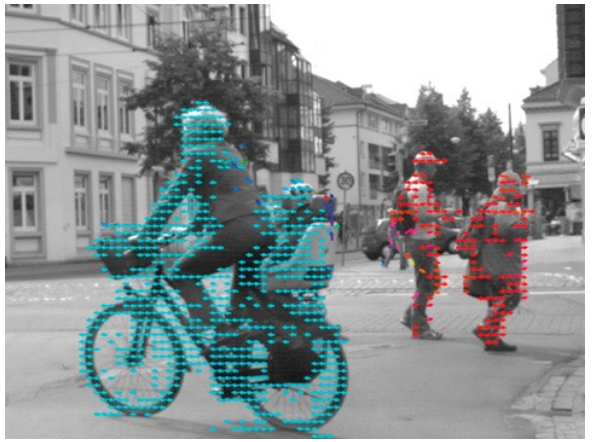
Cameras and sensors
Light and color



Linear filtering
Edge detection



Feature extraction

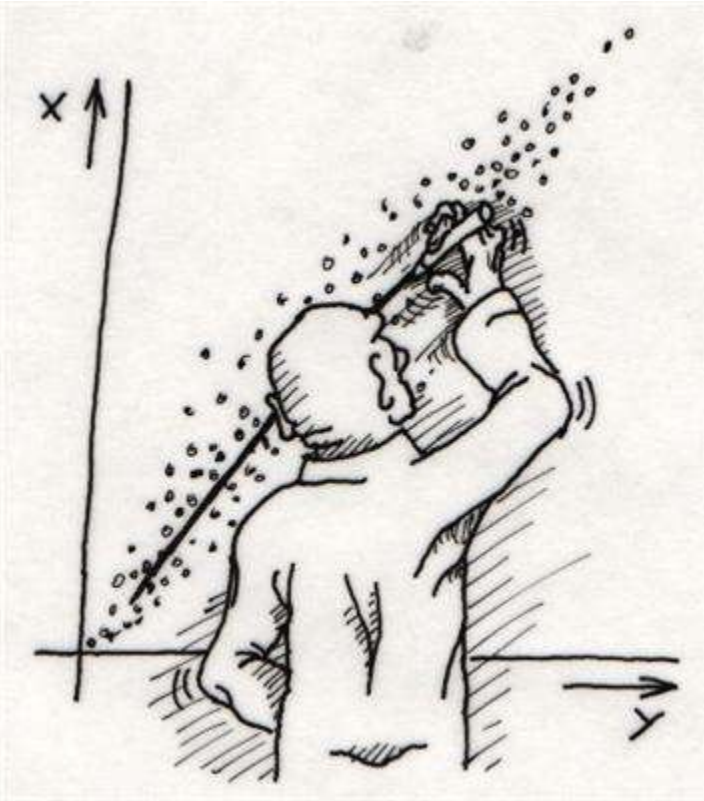


Optical flow

Source: L. Lazebnik

II. “Mid-level vision”

Fitting and grouping

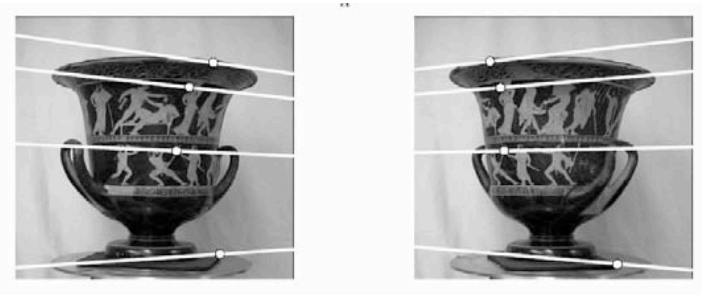


Fitting: Least squares
Voting methods

Alignment

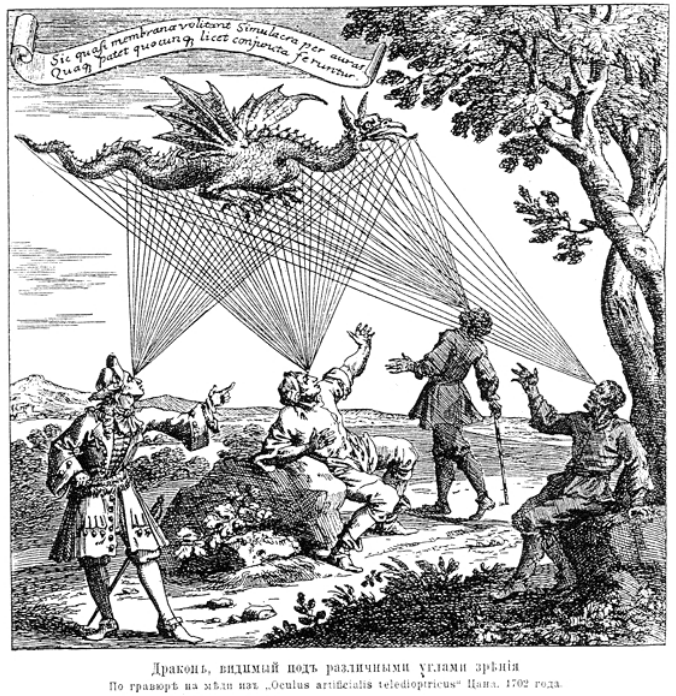
Source: L. Lazebnik

III. Multi-view geometry



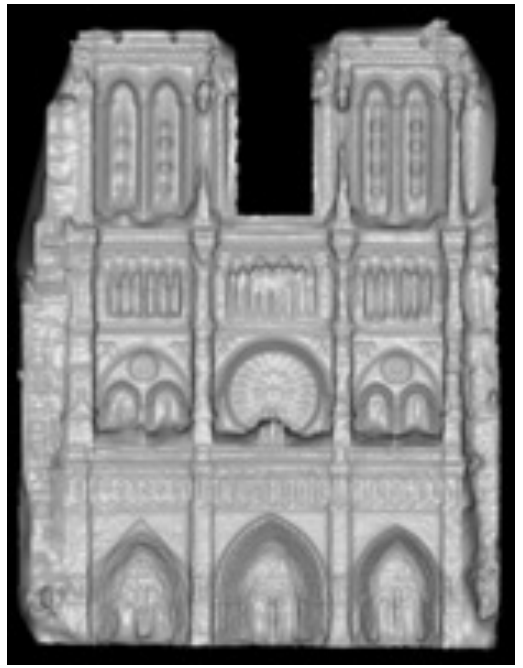
Epipolar geometry

Two-view stereo



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

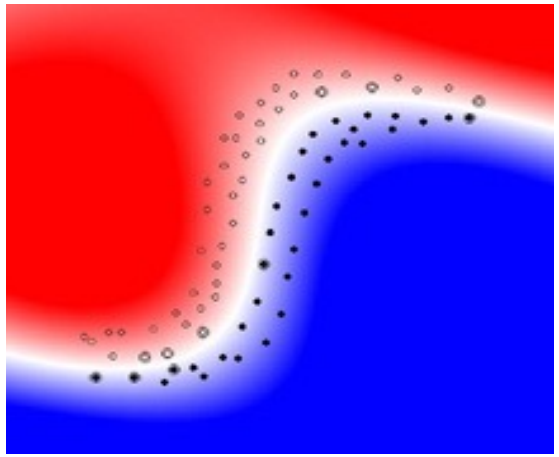
Structure from motion



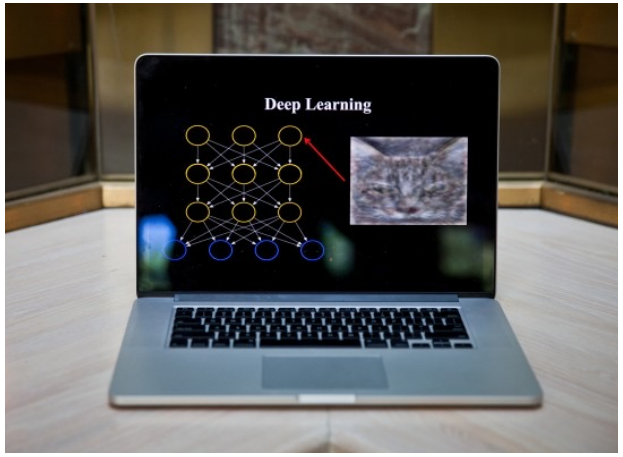
Multi-view stereo

Source: L. Lazebnik

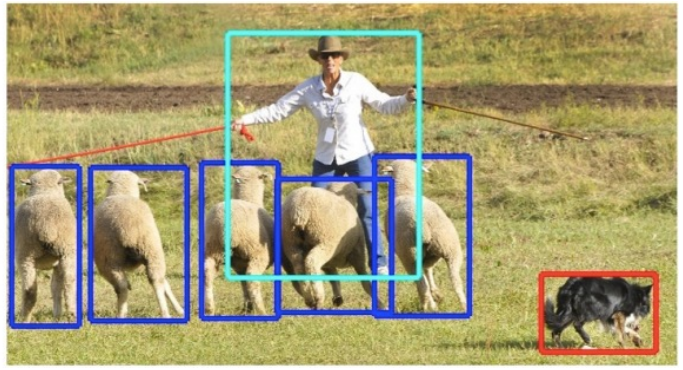
IV. Recognition



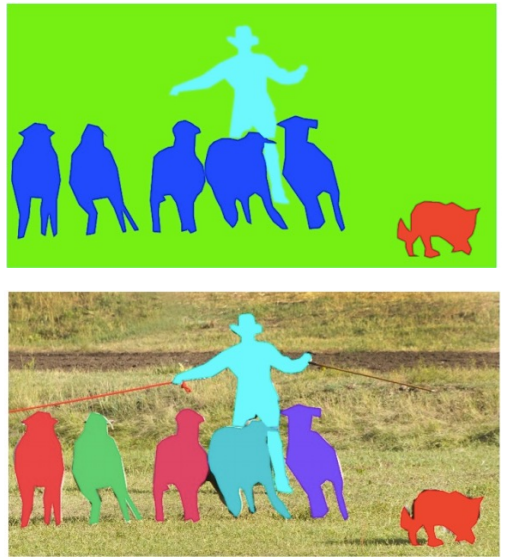
Basic classification



Deep learning



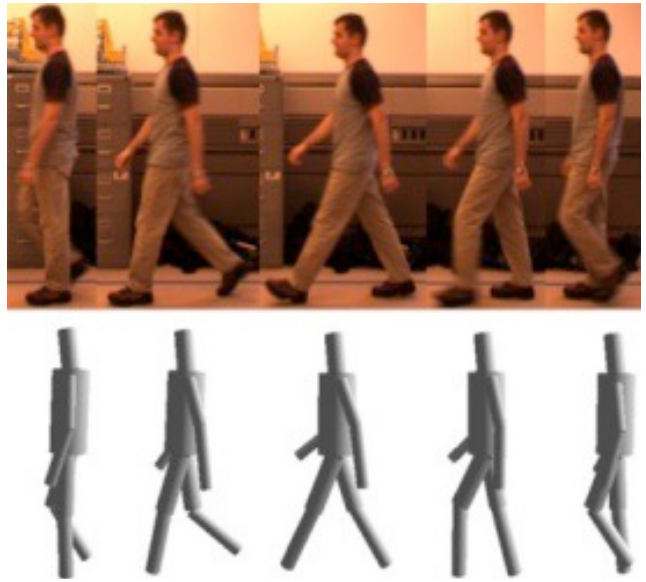
Object detection



Segmentation

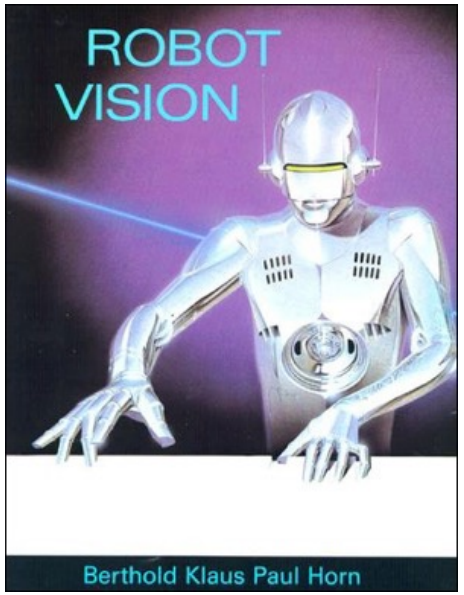
Source: L. Lazebnik

V. Additional Topics (time permitting)

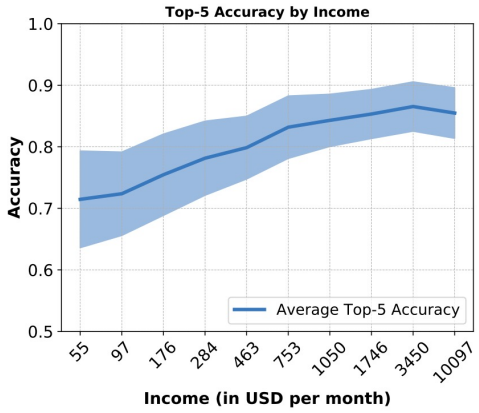


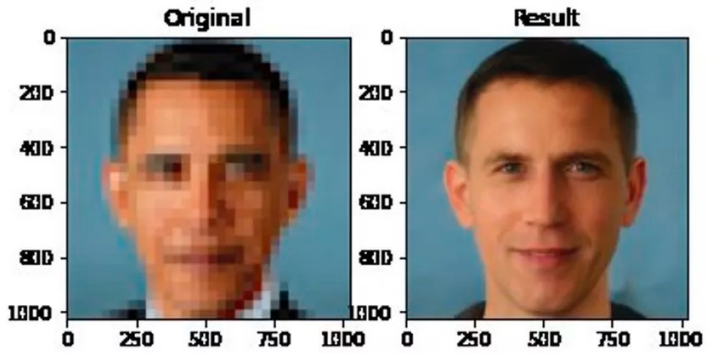
Video

Generation



Vision and Robotics





Bias and Ethical Considerations

Source: L. Lazebnik

Logistics

- Course TAs

Refik Mert Cam

Shun-Hsinag Hsu

Xiaodan Hu

David Yao

- Class website: http://saurabhg.web.illinois.edu/teaching/ece549/sp2024/