

CS 6476: Computer Vision

Instructor: James Hays

TAs: Cusuh Ham (head TA), Yijun Gu, Anant Joshi, Arvind Krishnakumar, John Lambert, Vijay Upadhyaya, Jing Wu

Today's Class

- Course enrollment
- Who am I?
- What is Computer Vision?
- Specifics of this course
- Geometry of Image Formation
- Questions

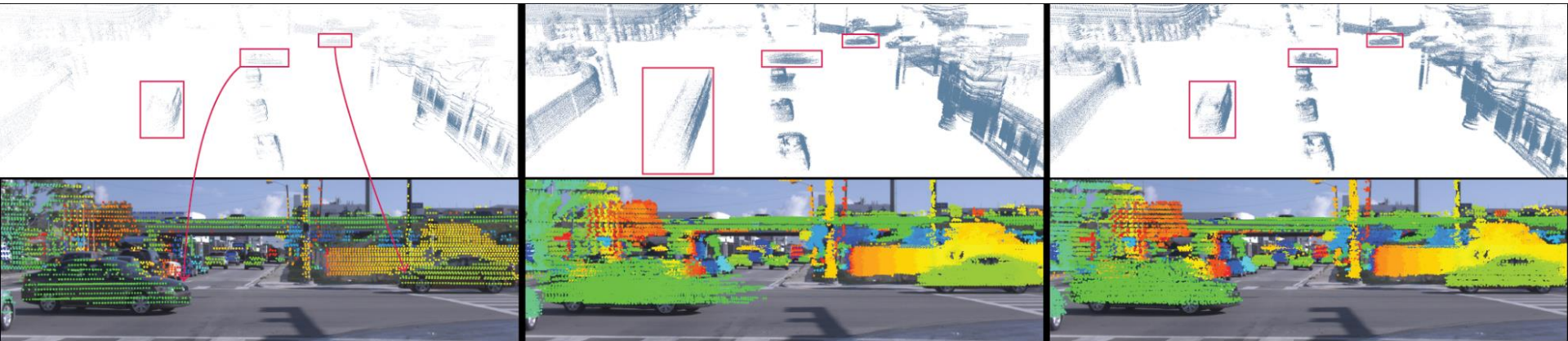
A bit about me





What type of stuff do I work on?

Understanding Lidar

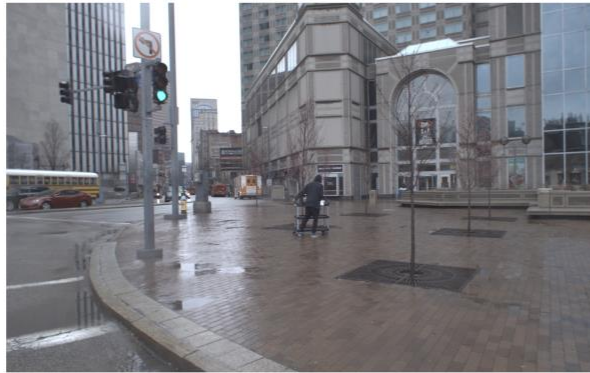


Scene Flow from Point Clouds with or without Learning

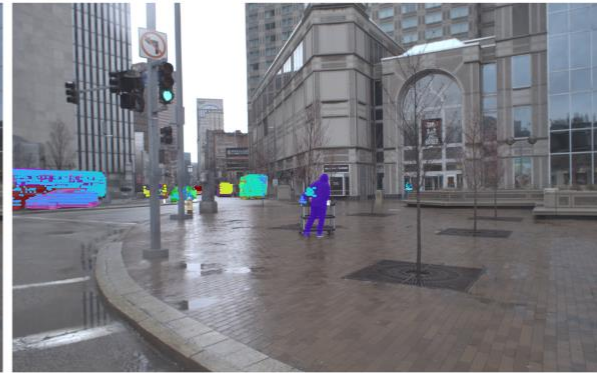
[Jhony Kaesemodel Pontes](#), [James Hays](#), [Simon Lucey](#)

<https://jhonykaesemodel.com/publication/sceneflow-3dv2020/>

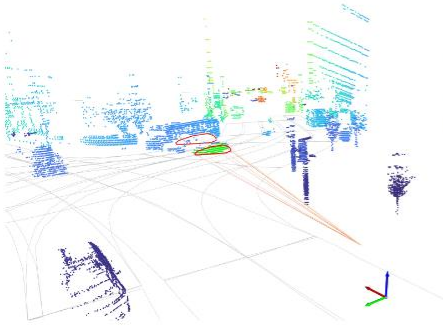
Understanding Lidar



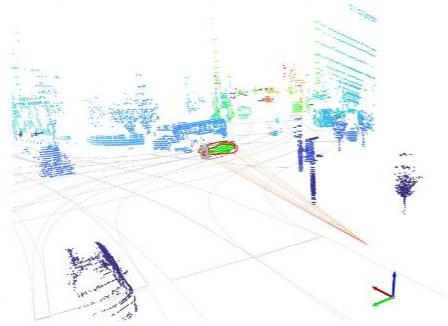
(a) Original camera image



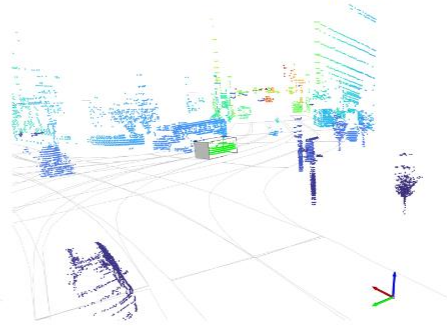
(b) Frustum proposals



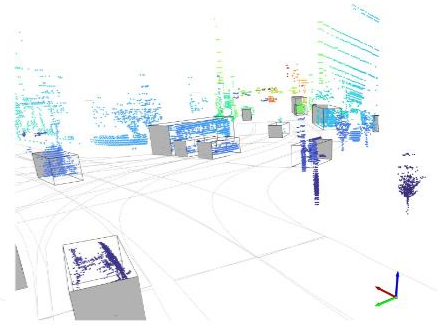
(c) Object frustum proposal



(d) LiDAR instance segmentation



(e) Amodal completion



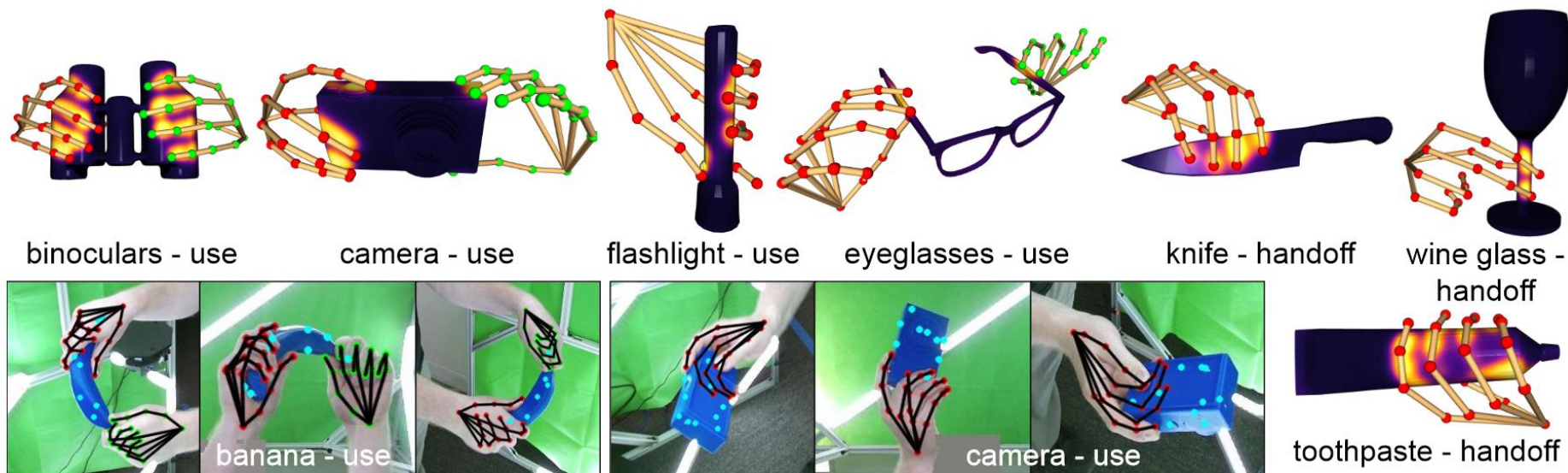
(f) Final cuboids

3D for Free: Crossmodal Transfer Learning using HD Maps

[Benjamin Wilson](#), [Zsolt Kira](#), [James Hays](#)

<https://arxiv.org/abs/2008.10592>

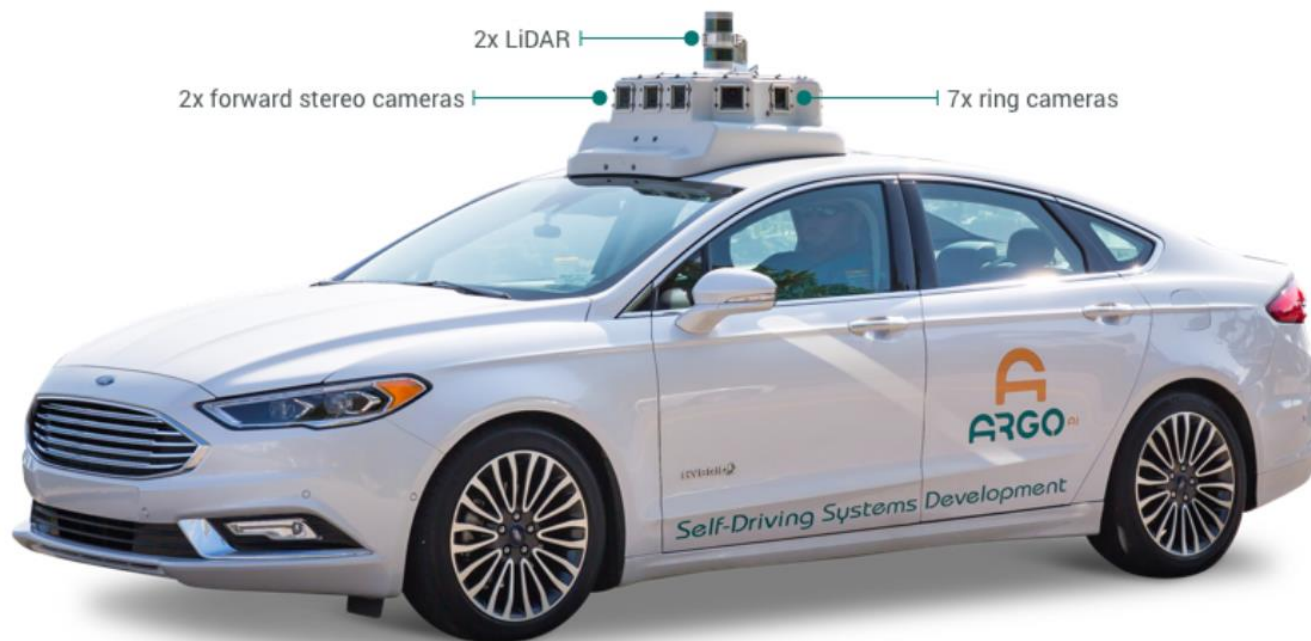
Exploring new data sources



ContactPose: A Dataset of Grasps with Object Contact and Hand Pose

[Samarth Brahmbhatt](#), [Chengcheng Tang](#), [Christopher D. Twigg](#), [Charles C. Kemp](#), [James Hays](#) ECCV 2020

Exploring new data sources



LiDAR

- 2 roof-mounted LiDAR sensors
- Overlapping 40° vertical field of view
- Range of 200m
- On average, our LiDAR sensors produce a point cloud with ~ 107,000 points at 10 Hz

Cameras

- Seven high-resolution ring cameras (1920 x 1200) recording at 30 Hz with a combined 360° field of view
- Two front-view facing stereo cameras (2056 x 2464) sampled at 5 Hz

Localization

We use a city-specific coordinate system for vehicle localization. We include 6-DOF localization for each timestamp, from a combination of GPS-based and sensor-based localization methods.

Calibration

Sensor measurements for each driving session are stored in "logs." For each log, we provide intrinsic and extrinsic calibration data for LiDAR and all nine cameras.

<https://www.argoverse.org/>

What is Computer Vision?

Derogatory summary of computer vision:
Machine learning applied to visual data

Computer Vision

- Automatic understanding of images and video
 1. Computing properties of the 3D world from visual data (*measurement*)

1. Vision for measurement

Real-time stereo



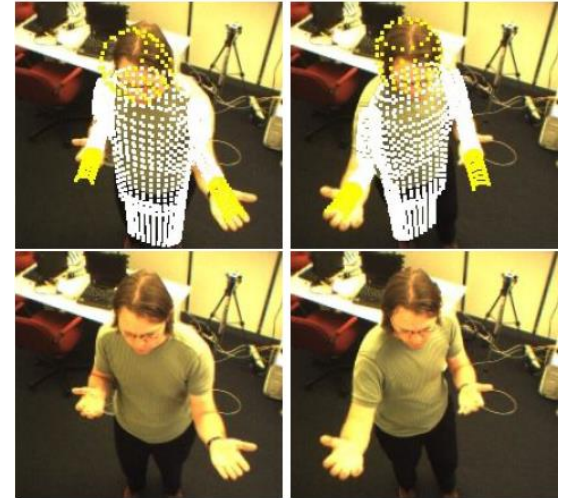
Wang et al.

Structure from motion



Snavey et al.

Tracking

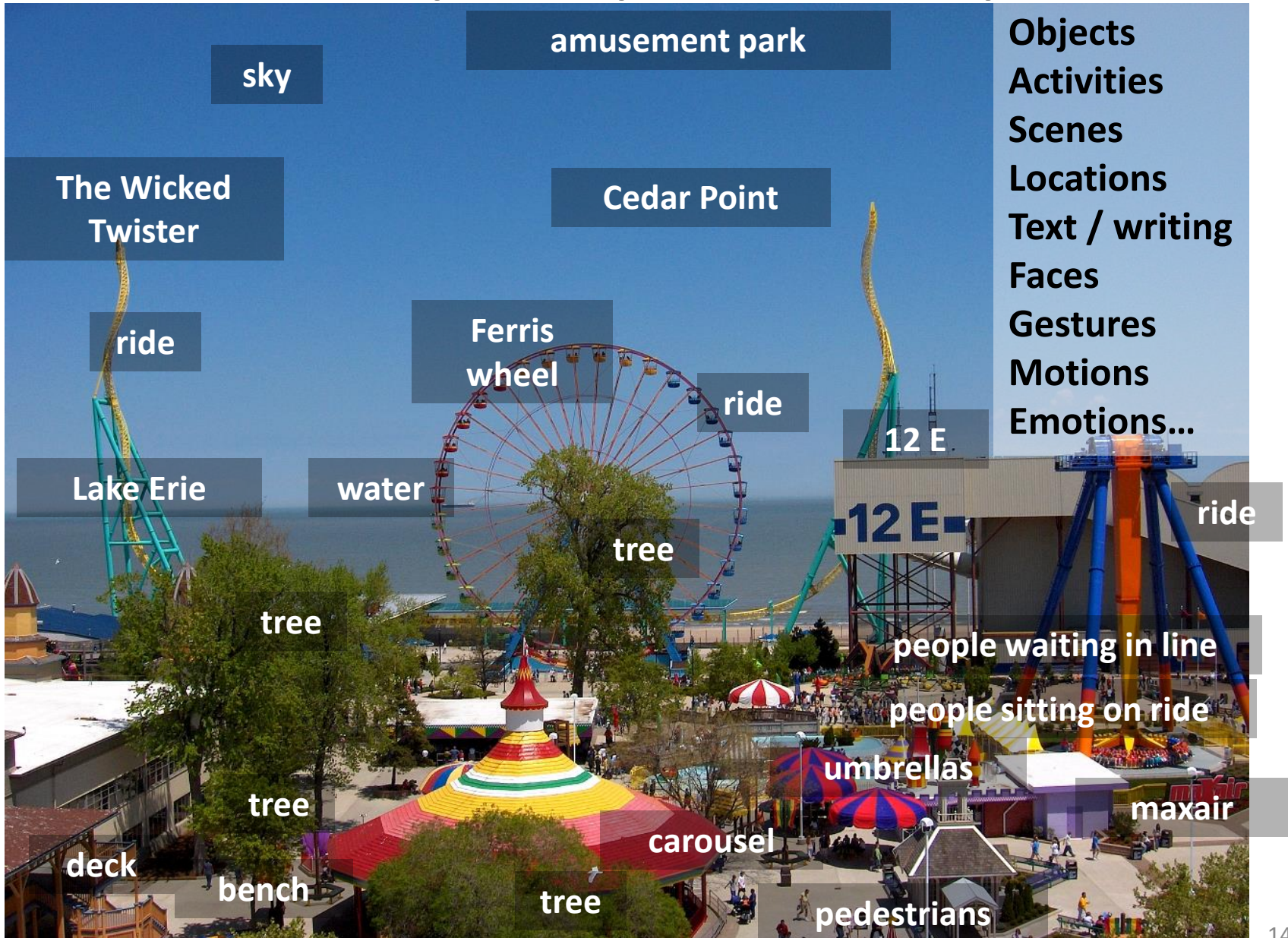


Demirdjian et al.

Computer Vision

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

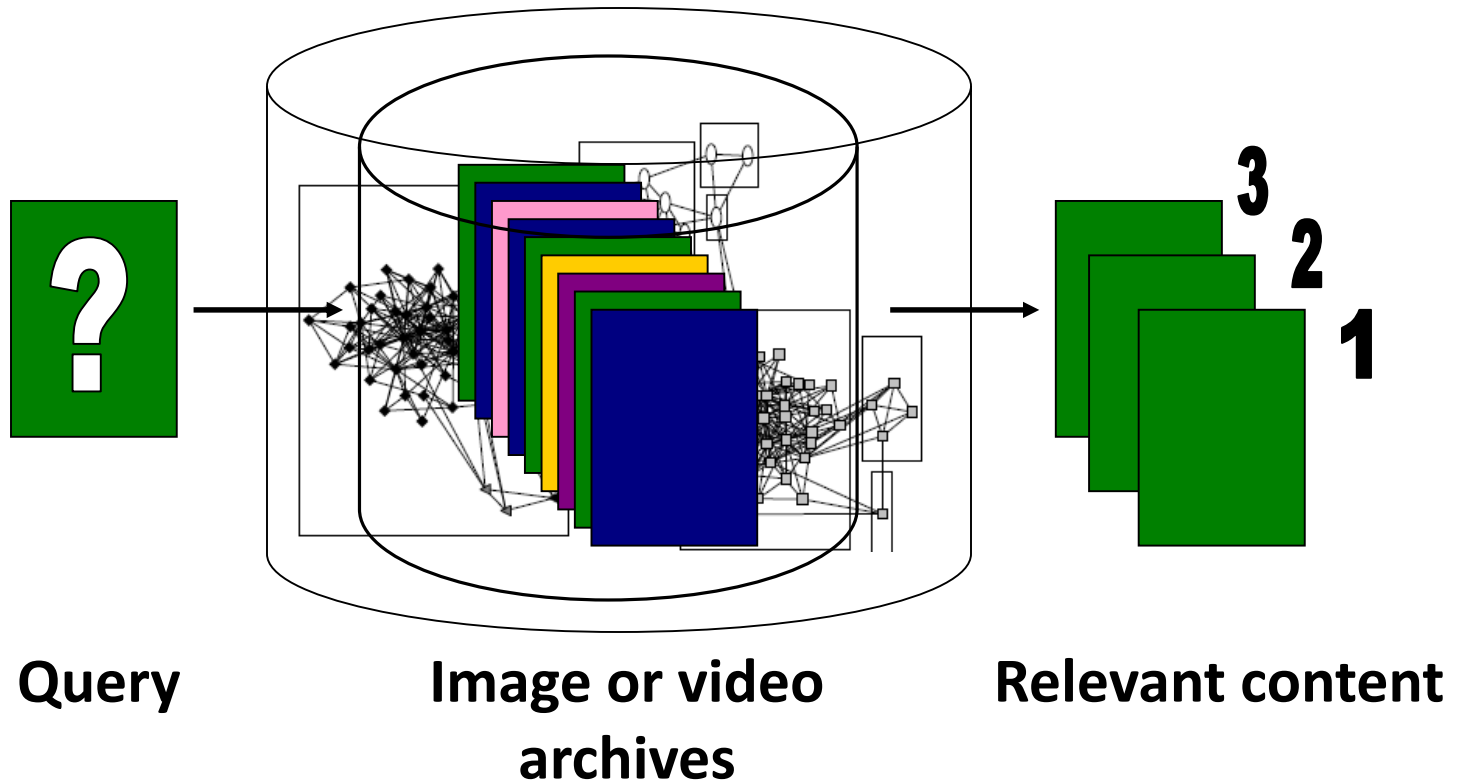
2. Vision for perception, interpretation



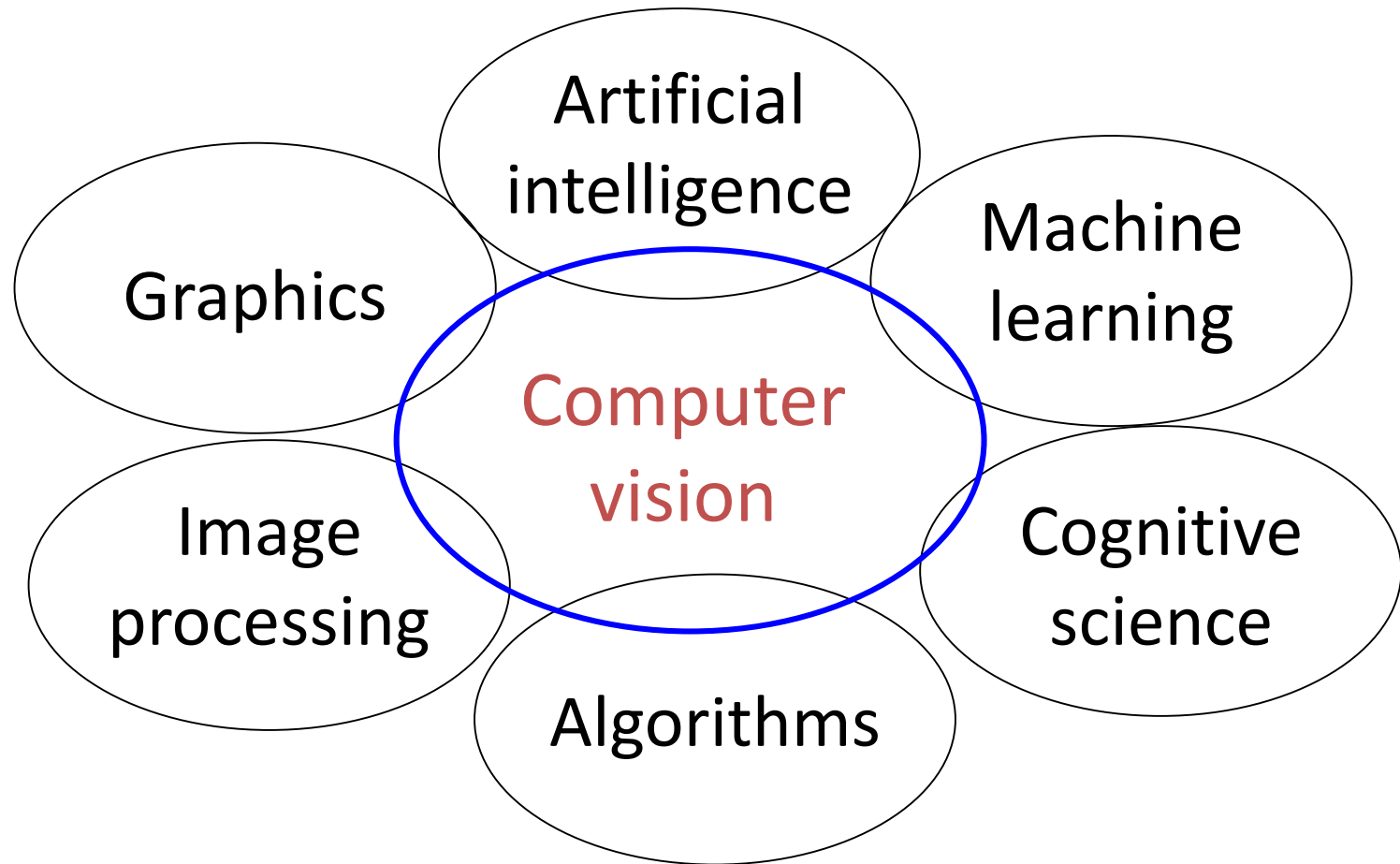
Computer Vision

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

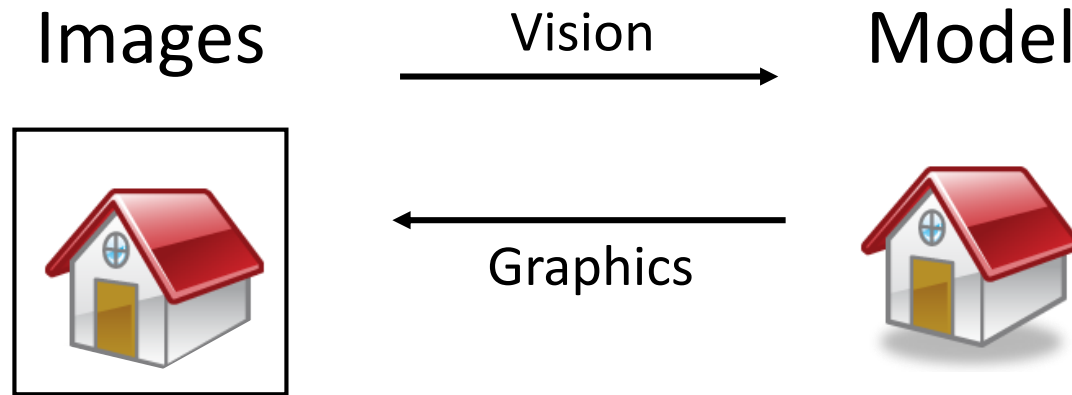
3. Visual search, organization



Related disciplines

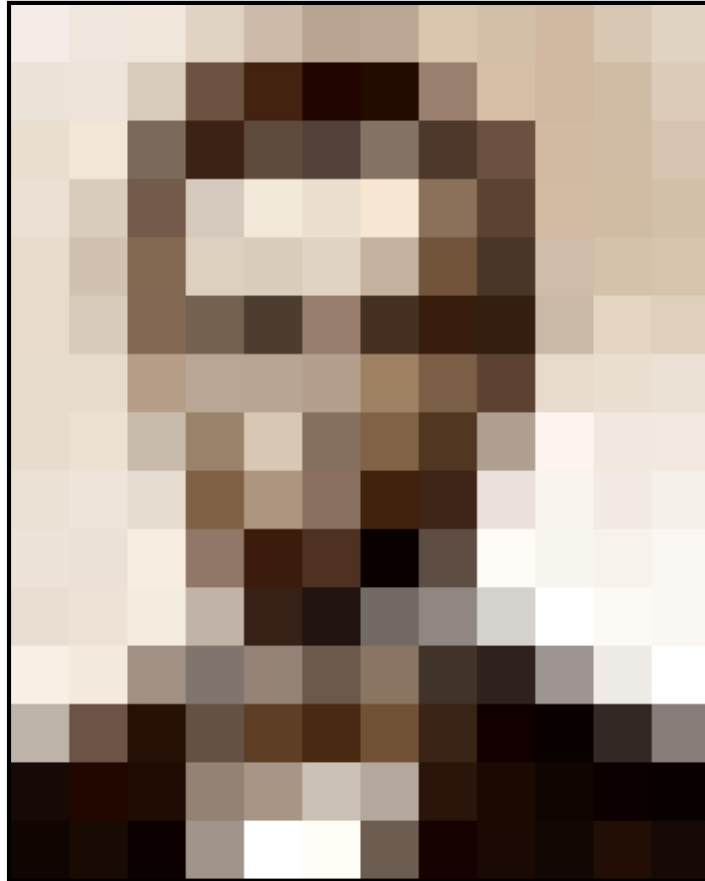


Vision and graphics



Inverse problems: analysis and synthesis.

What humans see



What computers see

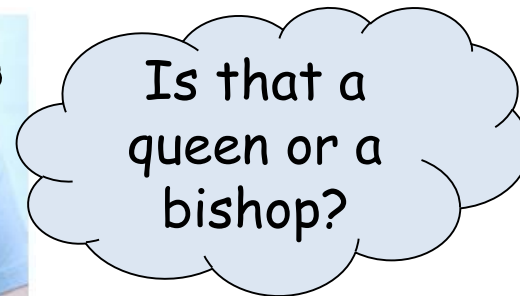


What do humans see?



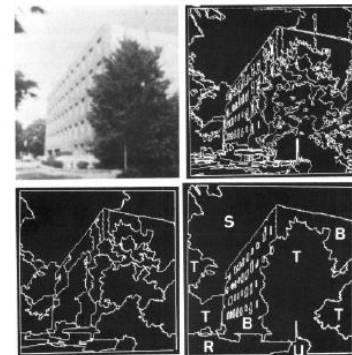
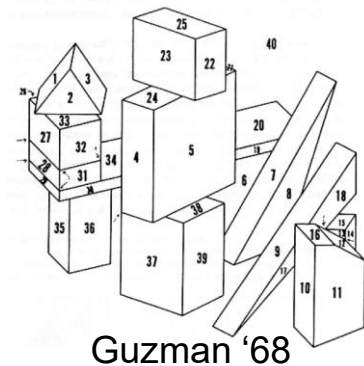
Vision is really hard

- Vision is an amazing feat of natural intelligence
 - Visual cortex occupies about 50% of Macaque brain
 - One third of human brain devoted to vision (more than anything else)



Ridiculously brief history of computer vision

- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts
- 2010's: Deep learning with ConvNets
- 2020's: Widespread autonomous vehicles?
- 2030's: robot uprising?



Ohta Kanade '78



Turk and Pentland '91

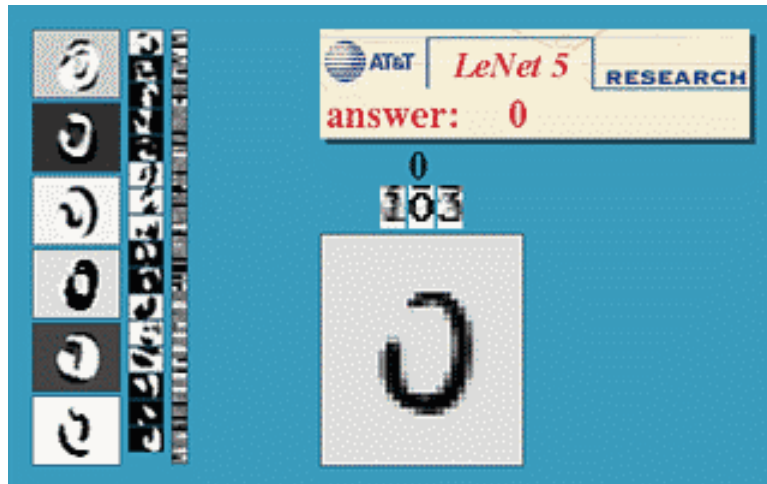
How vision is used now

- Examples of real world applications

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://www.research.att.com/~yann/>



License plate readers

http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Face detection



- Digital cameras detect faces

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.

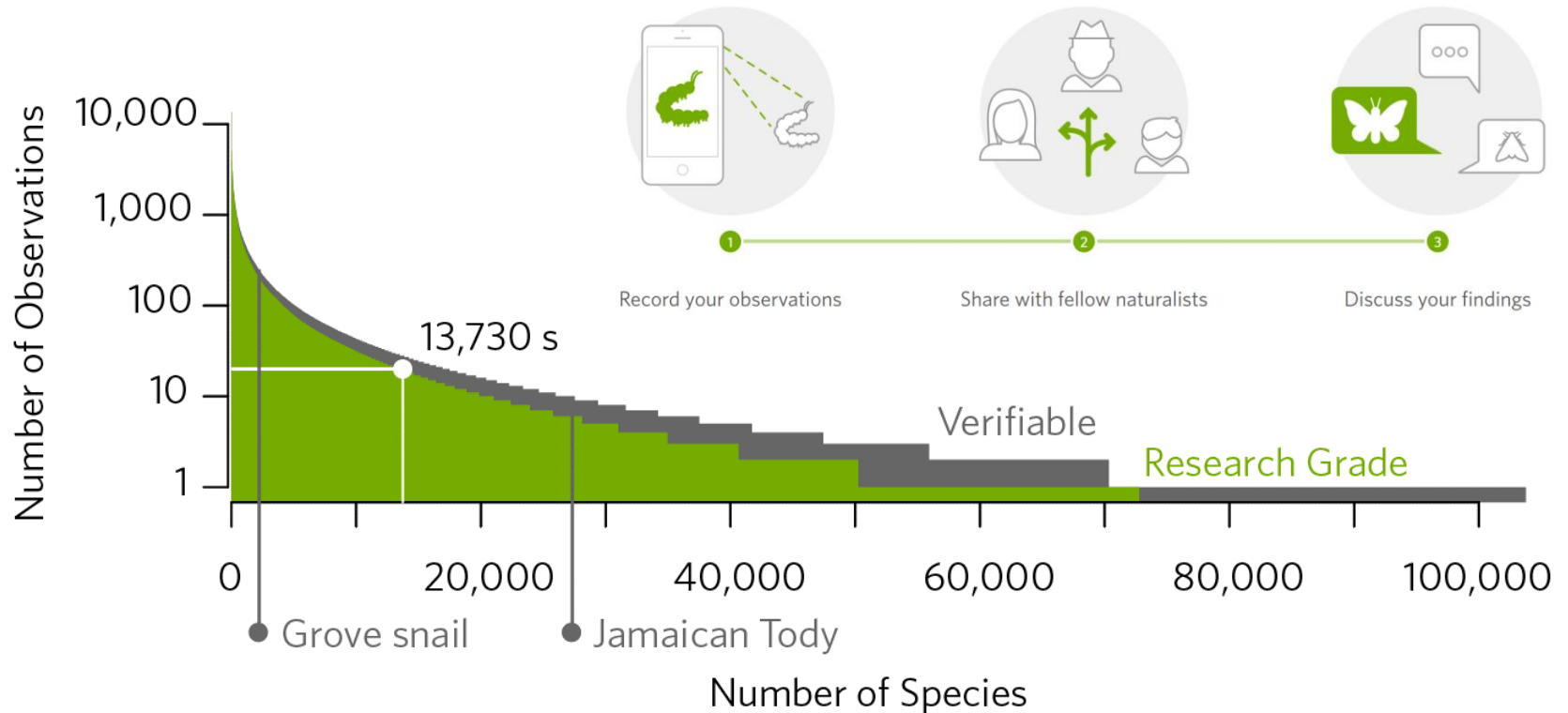
Vision systems (JPL) used for several tasks

- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read “[Computer Vision on Mars](#)” by Matthies et al.

iNaturalist



How It Works



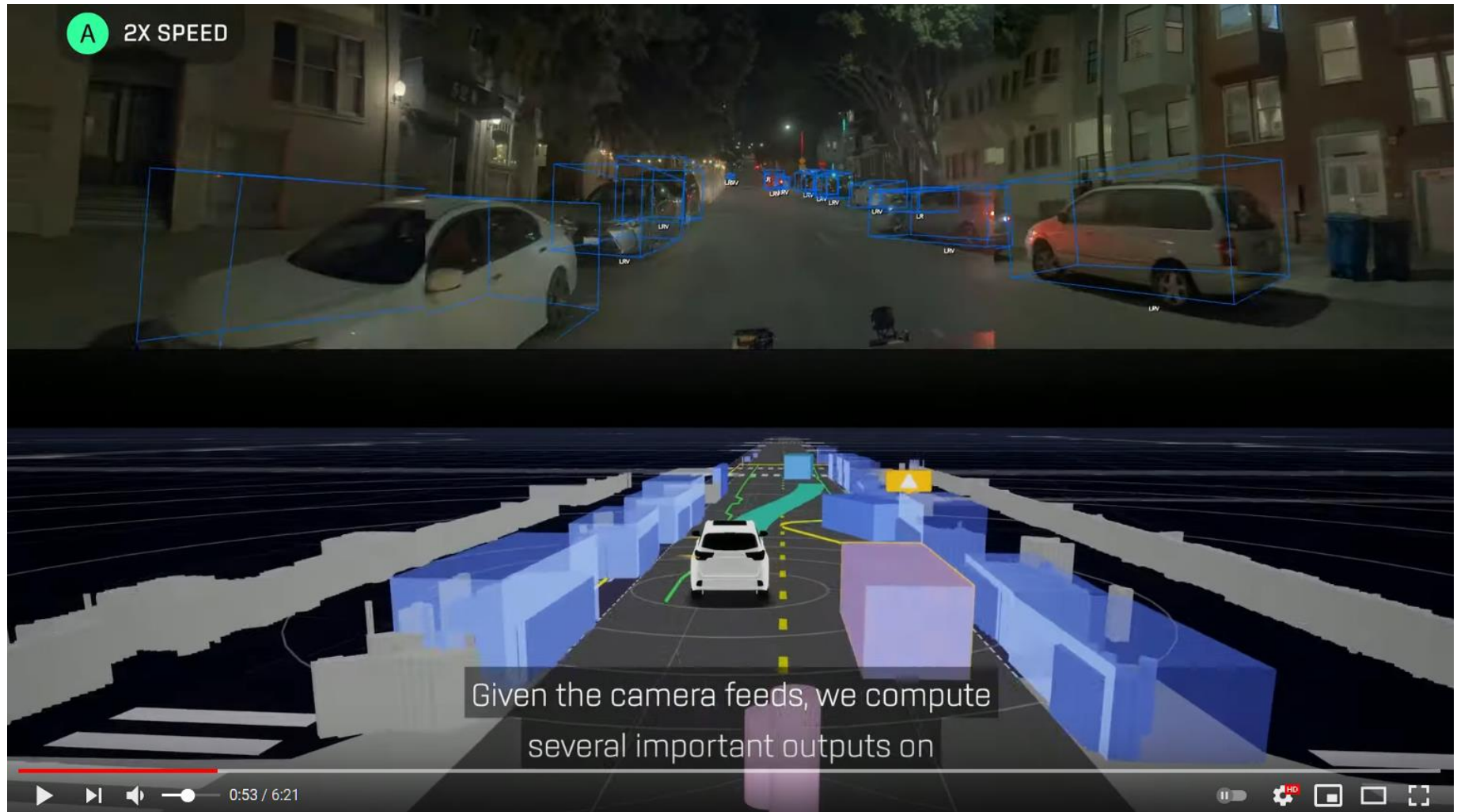
https://www.inaturalist.org/pages/computer_vision_demo

Skydio



<https://www.skydio.com/>

Zoox Computer Vision Demo

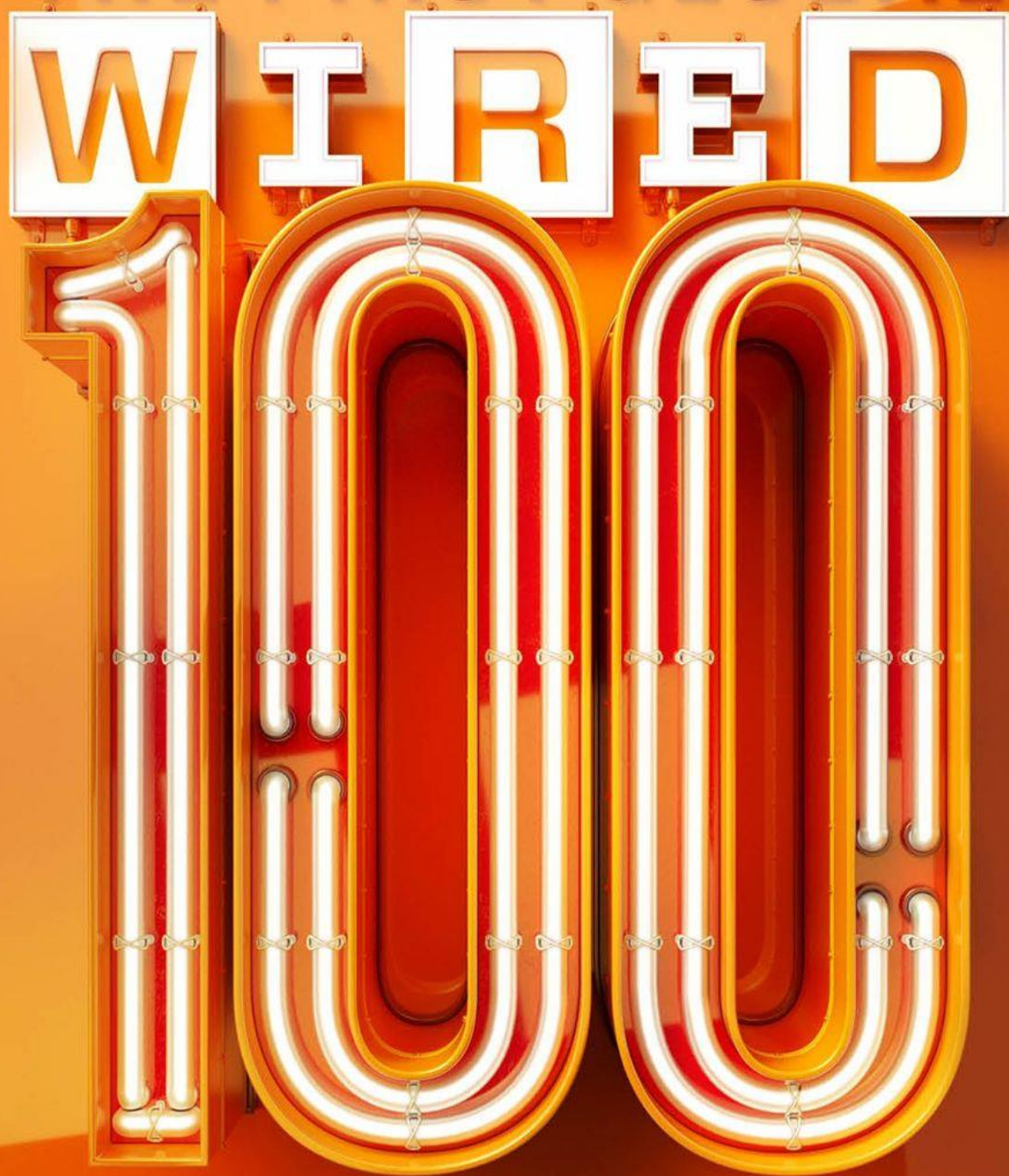


<https://www.youtube.com/watch?v=BVRMh9NO9Cs>

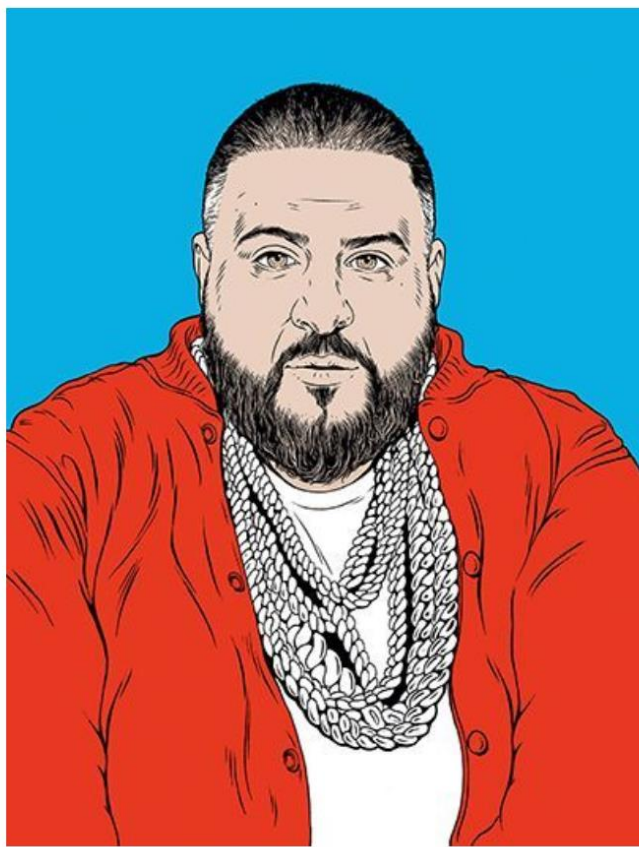
State of the art today?

With enough training data, computer vision ~~near~~ matches human vision at most recognition tasks

Deep learning has been an enormous disruption to the field. More and more techniques are being “deepified”.



WHO'S SHAPING THE DIGITAL WORLD?



DJ Khaled

Credit **Louise Zergaeng Pomeroy**

73. DJ Khaled

Snapchat icon; DJ and producer

Louisiana-born Khaled Mohamed Khaled, aka DJ Khaled, cut his musical chops in the early 00s as a host for Miami urban music radio WEDR. He proceeded to build a solid if not dazzling career as a mixtape DJ and music producer (he founded his label We The Best Music Group in 2008, and was appointed president of Def Jam South in 2009).

69. Geoffrey Hinton

Psychologist, computer scientist; researcher, Google Toronto

British-born Hinton has been dubbed the "godfather of deep learning". The Cambridge-educated cognitive psychologist and computer scientist started being an ardent believer in the potential of neural networks and deep learning in the 80s, when those technologies enjoyed little support in the wider AI community.

But he soldiered on: in 2004, with support from the Canadian Institute for Advanced Research, he launched a University of Toronto programme in neural computation and adaptive perception, where, with a group of researchers, he carried on investigating how to create computers that could behave like brains.

Hinton's work – in particular his algorithms that train multilayered neural networks – caught the attention of tech giants in Silicon Valley, which realised how deep learning could be applied to voice recognition, predictive search and machine vision.

The spike in interest prompted him to launch a free course on neural networks on e-learning platform Coursera in 2012. Today, 68-year-old Hinton is chair of machine learning at the University of Toronto and moonlights at Google, where he has been using deep learning to help build internet tools since 2013.

63. Yann Lecun

Director of AI research, Facebook, Menlo Park

LeCun is a leading expert in deep learning and heads up what, for Facebook, could be a hugely significant source of revenue: understanding its user's intentions.

62. Richard Branson

Founder, Virgin Group, London

Branson saw his personal fortune grow £550 million when Alaska Air bought Virgin America for \$2.6 billion in April. He is pressing on with civilian space travel with [Virgin Galactic](#).

61. Taylor Swift

Entertainer, Los Angeles



Grading

- 90% programming projects (6 total)
- 10% Open book problem sets (2 total)

Scope of CS 4476

Computer Vision

Robotics

Human Computer
Interaction

Medical Imaging

Neuroscience

Optics

Computational
Photography

Graphics

Machine
Learning

Image Processing
Geometric Reasoning
Recognition
Deep Learning

Textbook

Computer Vision: Algorithms and Applications, 2nd ed.

© 2020 [Richard Szeliski](#), Facebook



<http://szeliski.org/Book/>

Prerequisites

- **Linear algebra**, basic calculus, and probability
- Experience with image processing will help but is not necessary
- Experience with Python or Python-like languages will help

Projects (tentative)

- Image Filtering and Hybrid Images
- Local Feature Matching
- Camera Calibration and Fundamental Matrix Estimation with RANSAC
- Stereo
- Image Classification with Deep Learning
- Semantic Segmentation with Deep Learning

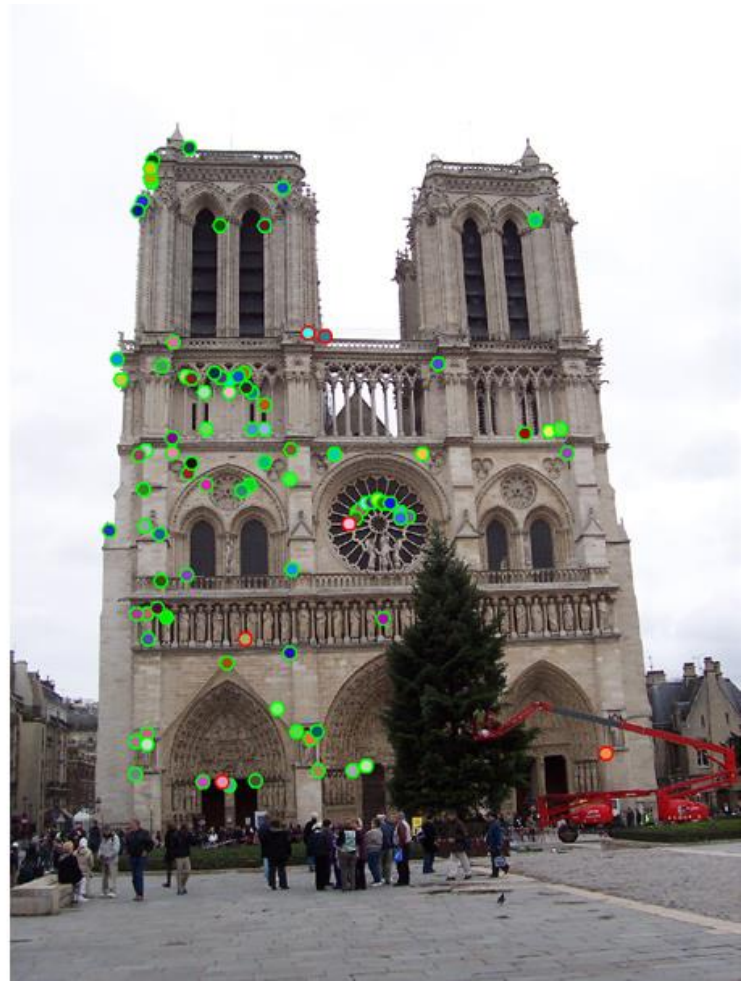
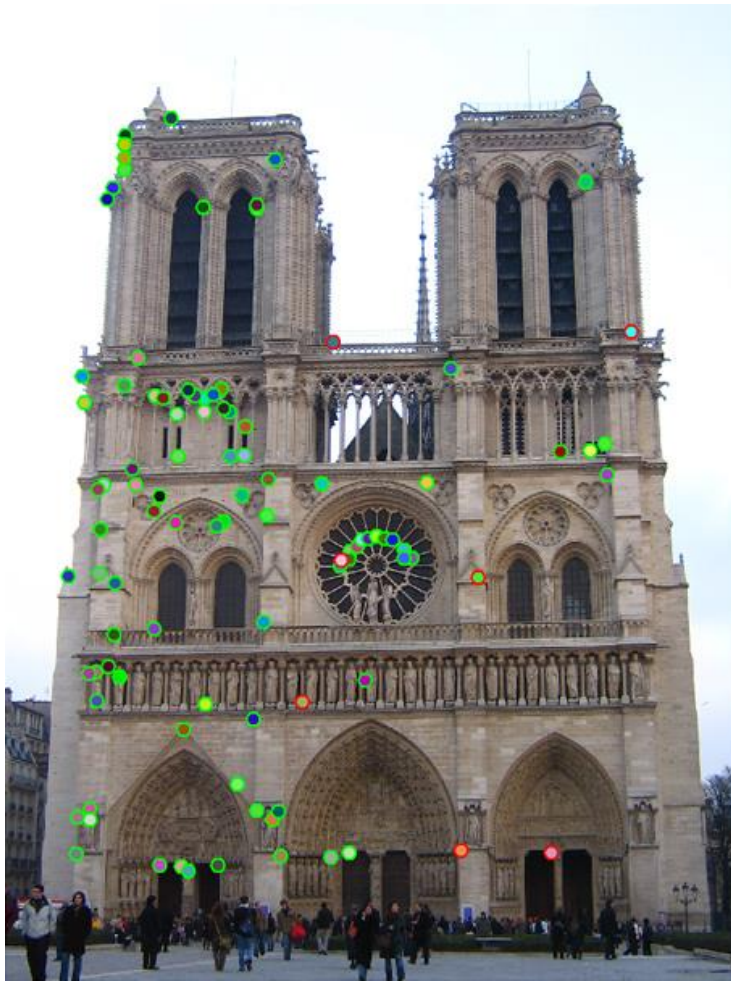
Proj1: Image Filtering and Hybrid Images

- Implement image filtering to separate high and low frequencies
- Combine high frequencies and low frequencies from different images to create an image with scale-dependent interpretation



Proj2: Local Feature Matching

- Implement interest point detector, SIFT-like local feature descriptor, and simple matching algorithm.



Course Syllabus (tentative)

<http://www.cc.gatech.edu/~hays/compvision>

Code of Conduct

Your work must be your own. We'll look for cheating. Don't talk at the level of code with other students.