CS 4476-B and 6476-A: Computer Vision

Instructor: James Hays
TAs: **Cusuh Ham** (head TA), Otis Smith, Pranav Khorana, Sukriti Bhardwaj, Xueqing Li, Yash Kothari, Yoonwoo Kim, Wei Xiong Toh, Chengde Xu, Vince Li, Nikith Hosangadi

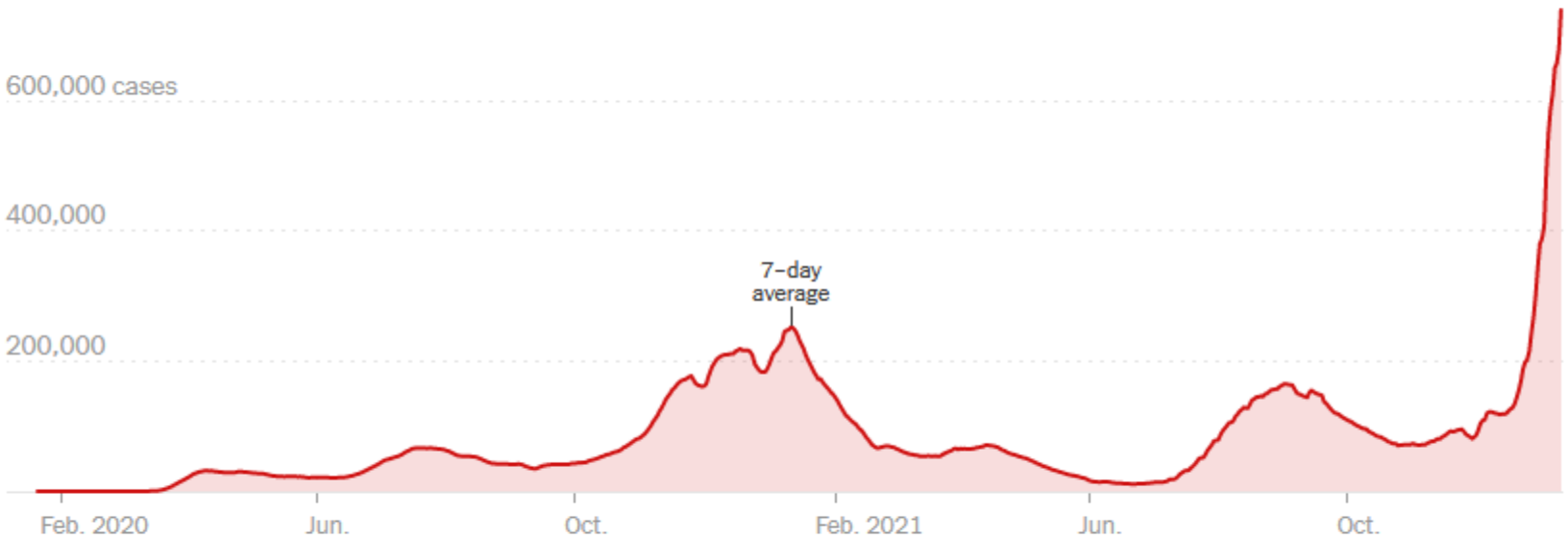
Image by kirkh.deviantart.com

Today's Class

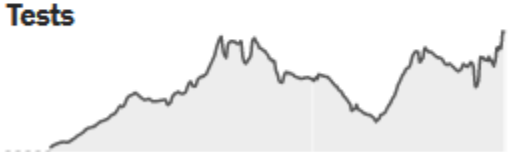
- Covid safety
- Who am I?
- What is Computer Vision?
- Specifics of this course
- Geometry of Image Formation
- Questions

New reported cases

All time | Last 90 days

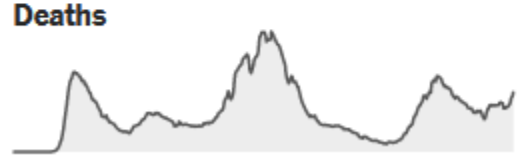


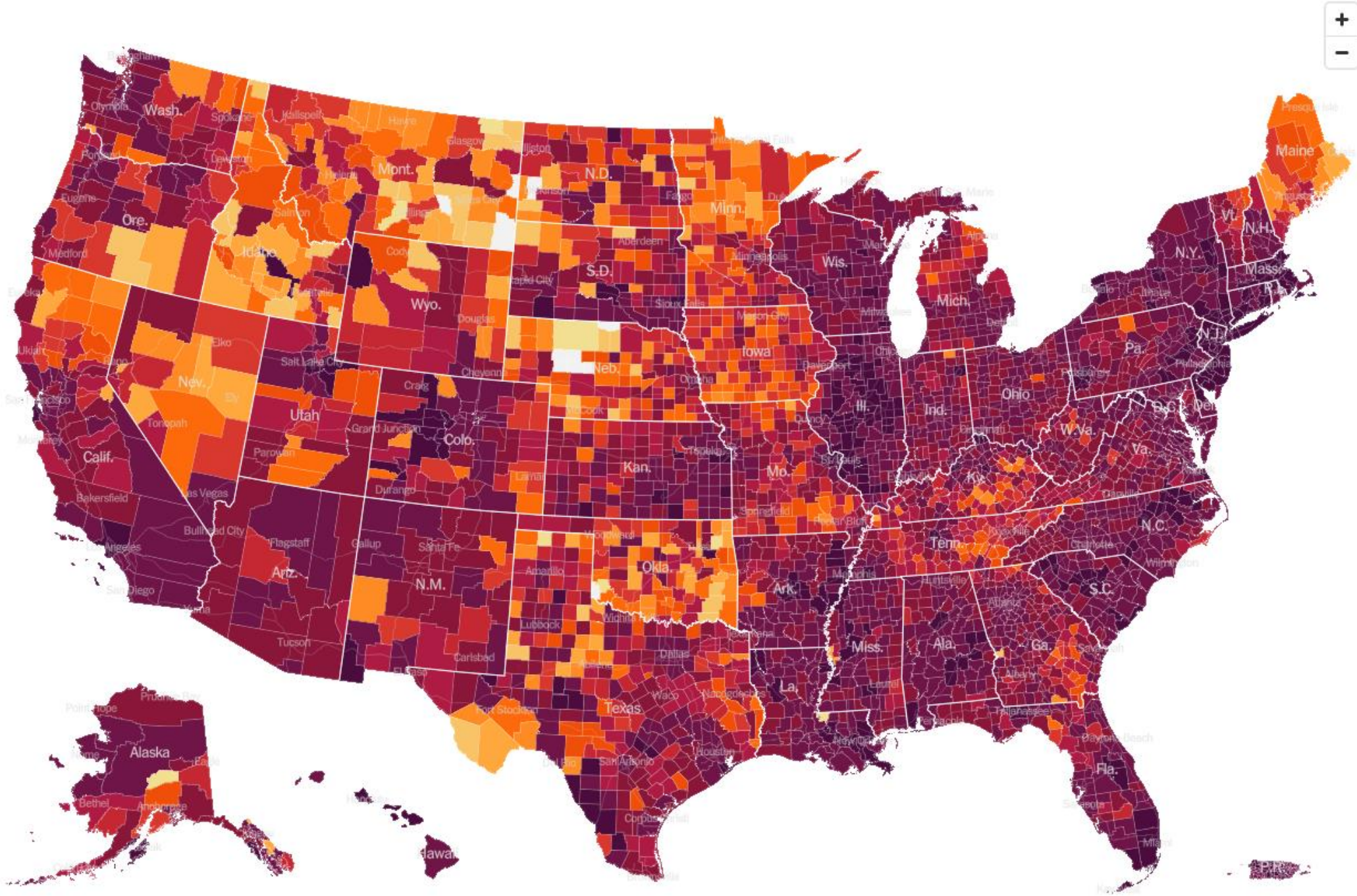
Tests



Hospitalized

Deaths




Wash.
Ore.
Calif.
Idaho
Nev.
Mont.
Wyo.
Utah
Colo.
Ariz.
N.M.
N.D.
S.D.
Neb.
Kan.
Okla.
Texas
Minn.
Iowa
Mo.
Ark.
La.
Wis.
Ill.
Mich.
Ind.
Ohio
Ky.
Tenn.
Miss.
Ala.
Ga.
Fla.
S.C.
N.C.
Va.
W.Va.
Pa.
N.Y.
Maine
Vt.
N.H.
Mass.
R.I.
N.J.
Alaska
Hawaii

Georgia Institute of Technology Covid-19 Data

Count of Positive Cases since March 2020	Testing Count since August 2020
4,342	443,694

Student Count of Positive Cases	Employee/Affiliate Count of Positive Cases	Student Test Count	Employee/Affiliate Test Count
3,018	1,324	317,879 (72%)	124,359 (28%)

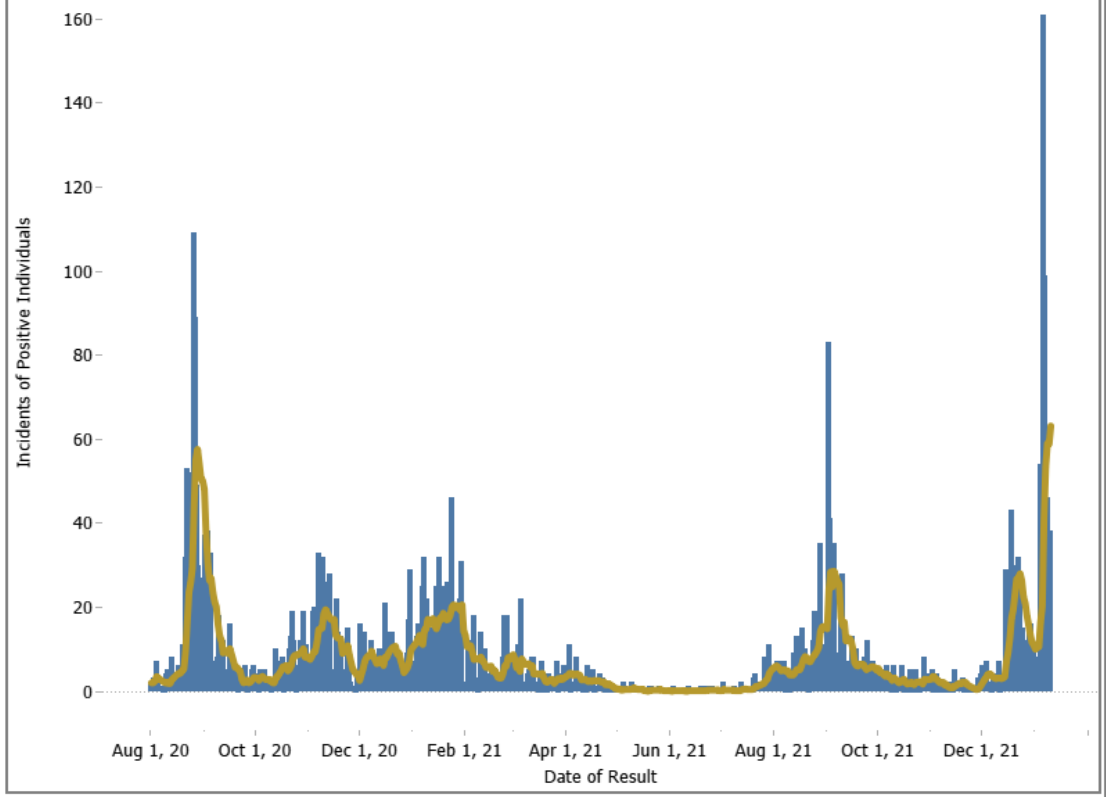
Isolation/Quarantine In Use Bed Count	Total Beds**
15	110

Past Seven Days Rolling Averages

46.71 (Positives) 584.43 (Tests) 6.12% (Surveillance Incident Rate)

****This number will update based on changes to GT contracts for isolation space**

Moving Averages by Incidents of Positive Individuals



1. Population Selection

- ○ Total
- ○ Campus Location Type
- ○ GT Primary Affiliation
- ○ Testing Source
- ◉ Moving Averages

2. Population Filter

- ☑ **(All)**
- ☑

3. Measures

- ◉ Incidents of Positive Individuals
- ○ Cumulative Incidents of Positive Individuals
- ○ Number of Tests Per Day
- ○ Cumulative Number of Tests
- ○ Surveillance Incidence Positivity (0-100)

From Date of Result

8/1/2020

1/11/2022

*Notes days with < 450 total tests may not be indicative of surveillance trends.

Covid-19 Data Table

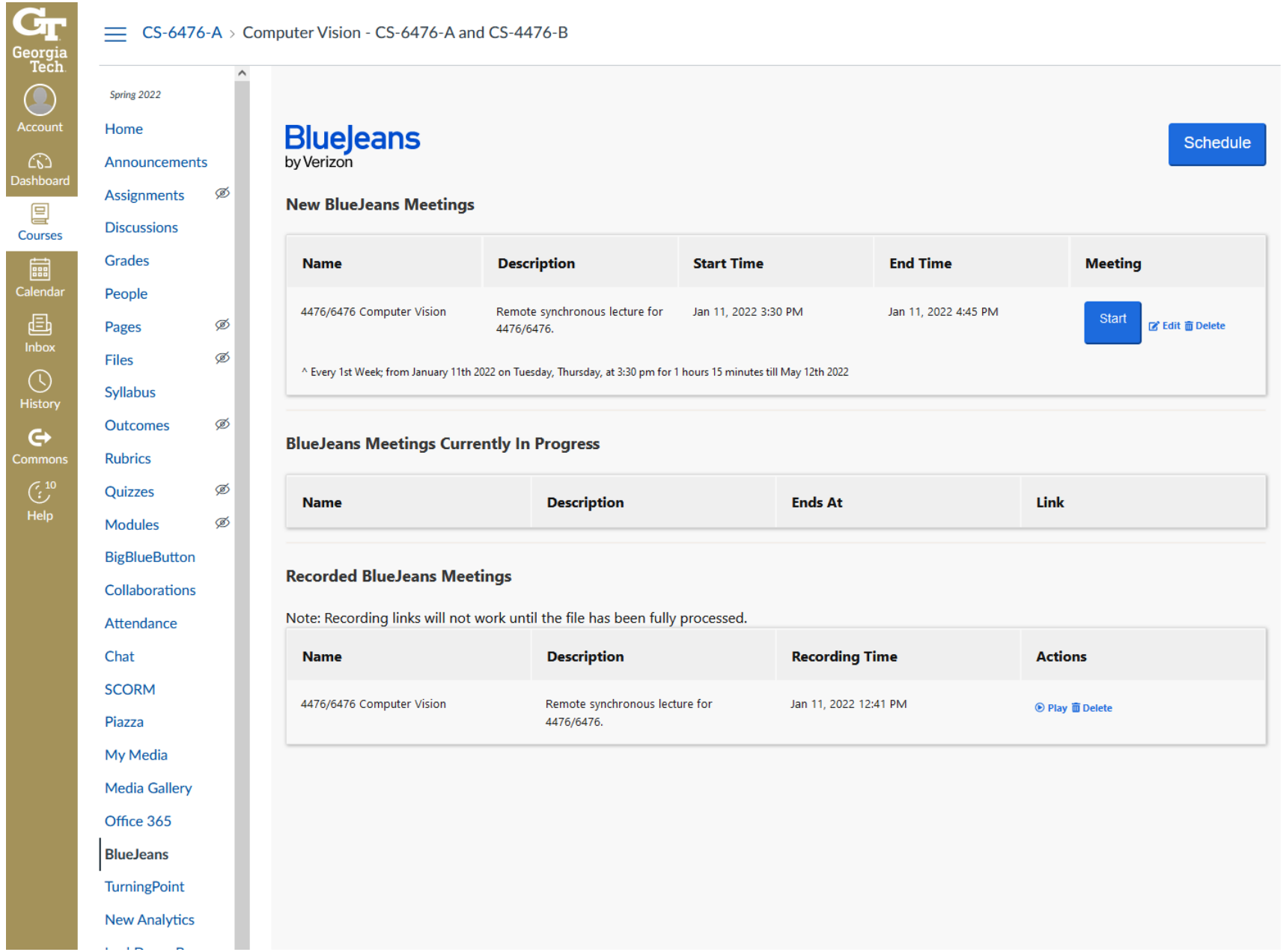
Surveillance Incident Positivity Table

For all of these reasons, in-person attendance is **discouraged** during the start of classes.

If you come to class, please make sure you are

- Masking
- Vaccinated
- Symptom free

Even doing all of those things, there is risk, so I am happy for students to attend lecture through Bluejeans


Georgia Tech
Account
Dashboard
Courses
Calendar
Inbox
History
Commons
Help
CS-6476-A › Computer Vision - CS-6476-A and CS-4476-B
Spring 2022
Home
Announcements
Assignments
Discussions
Grades
People
Pages
Files
Syllabus
Outcomes
Rubrics
Quizzes
Modules
BigBlueButton
Collaborations
Attendance
Chat
SCORM
Piazza
My Media
Media Gallery
Office 365
BlueJeans
TurningPoint
New Analytics
BlueJeans by Verizon
Schedule
New BlueJeans Meetings
Name
Description
Start Time
End Time
Meeting
4476/6476 Computer Vision
Remote synchronous lecture for 4476/6476.
Jan 11, 2022 3:30 PM
Jan 11, 2022 4:45 PM
Start
Edit
Delete
^ Every 1st Week; from January 11th 2022 on Tuesday, Thursday, at 3:30 pm for 1 hours 15 minutes till May 12th 2022
BlueJeans Meetings Currently In Progress
Name
Description
Ends At
Link
Recorded BlueJeans Meetings
Note: Recording links will not work until the file has been fully processed.
Name
Description
Recording Time
Actions
4476/6476 Computer Vision
Remote synchronous lecture for 4476/6476.
Jan 11, 2022 12:41 PM
Play
Delete

Today’s Class

~~Covid safety~~

Who am I?

What is Computer Vision?

Specifics of this course

Geometry of Image Formation

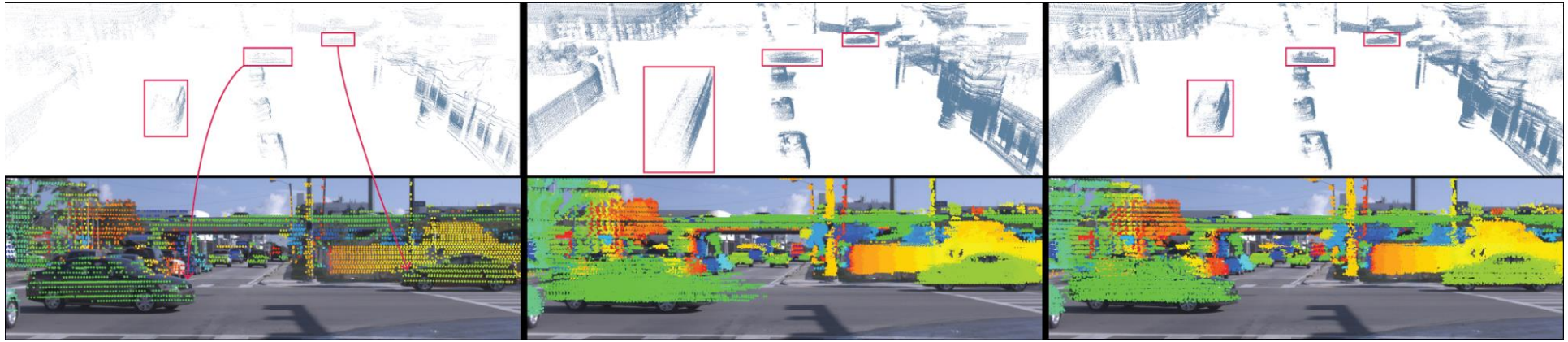
Questions

A bit about me


ARGO AI
Ford
self-driving test vehicle
vehículo de manejo autónomo

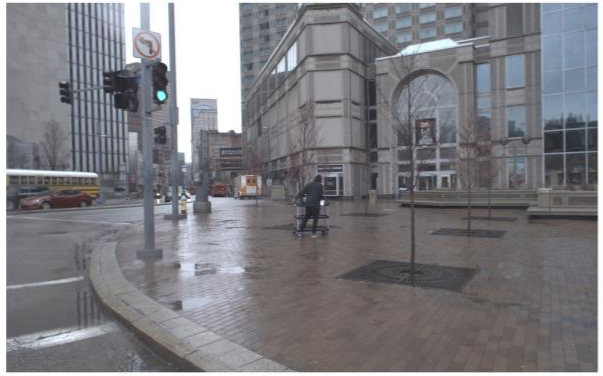
What type of stuff do I work on?

Understanding Lidar

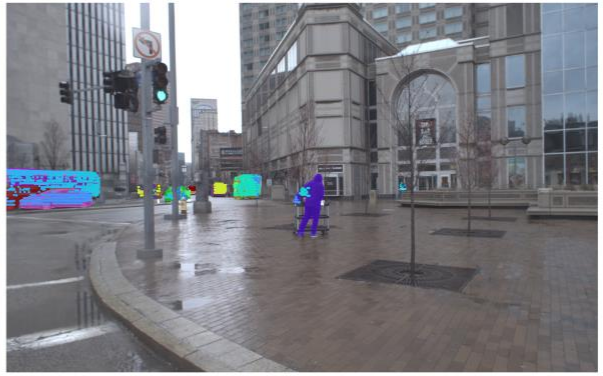


Scene Flow from Point Clouds with or without Learning
Jhony Kaesemodel Pontes, James Hays, Simon Lucey
https://jhonykaesemodel.com/publication/sceneflow-3dv2020/

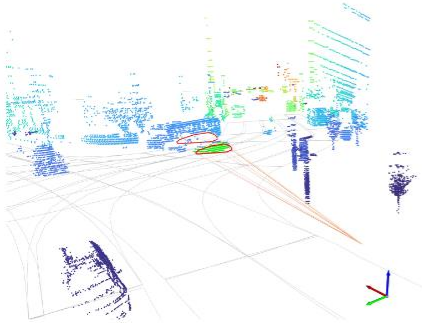
Understanding Lidar



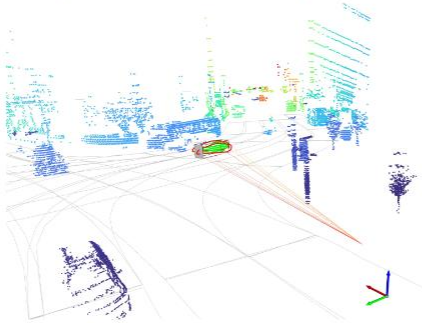
(a) Original camera image



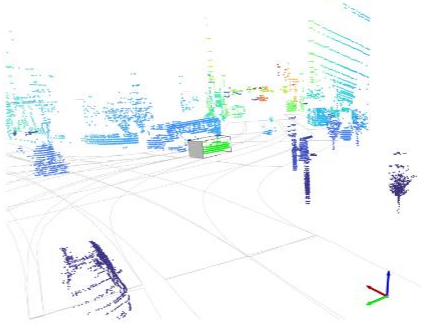
(b) Frustum proposals



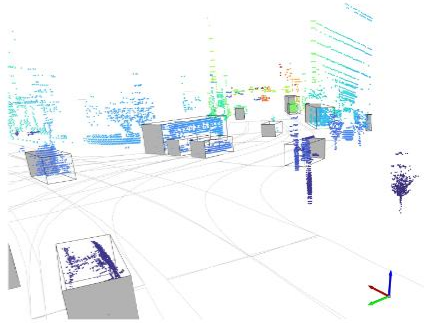
(c) Object frustum proposal



(d) LiDAR instance segmentation



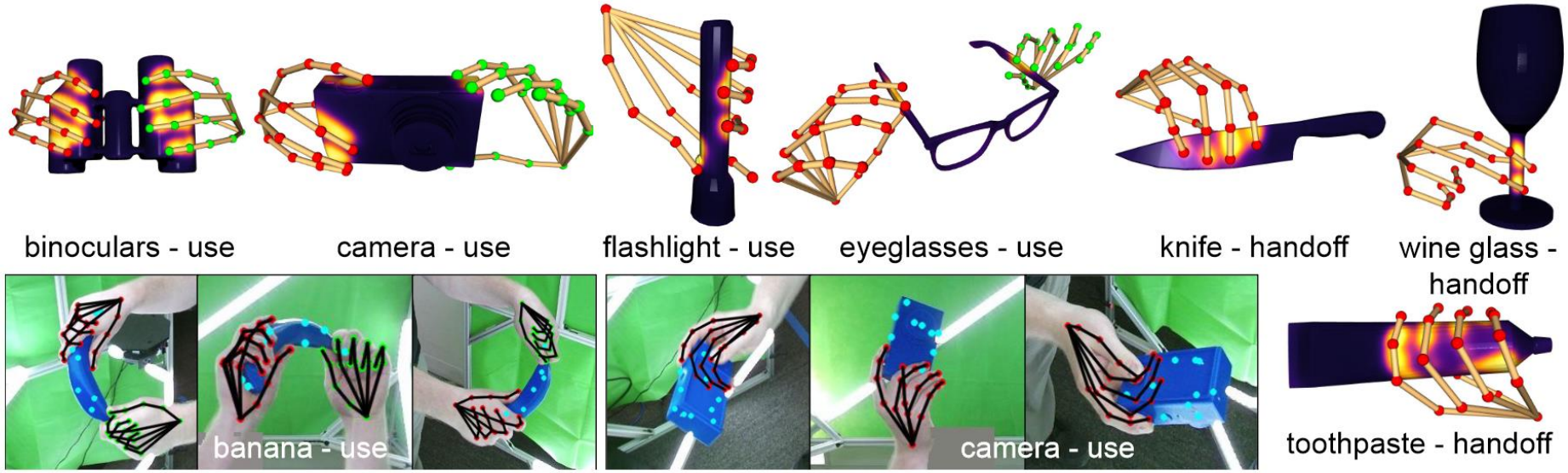
(e) Amodal completion



(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/

Today's Class

- ~~Covid safety~~
- ~~Who am I?~~
- What is Computer Vision?
- Specifics of this course
- Geometry of Image Formation
- Questions

What is Computer Vision?

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

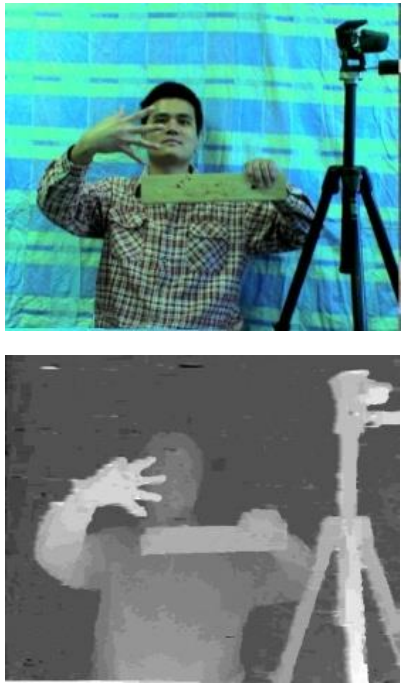
Computer Vision

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

Slide credit: Kristen Grauman

1. Vision for measurement

Real-time stereo

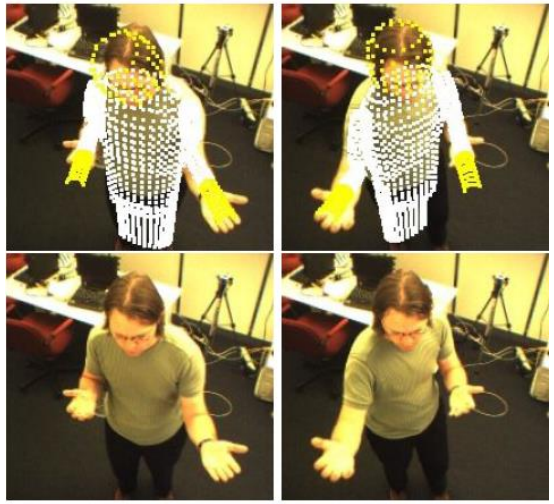


Wang et al.

Structure from motion

Snavely et al.

Tracking



Demirdjian et al.

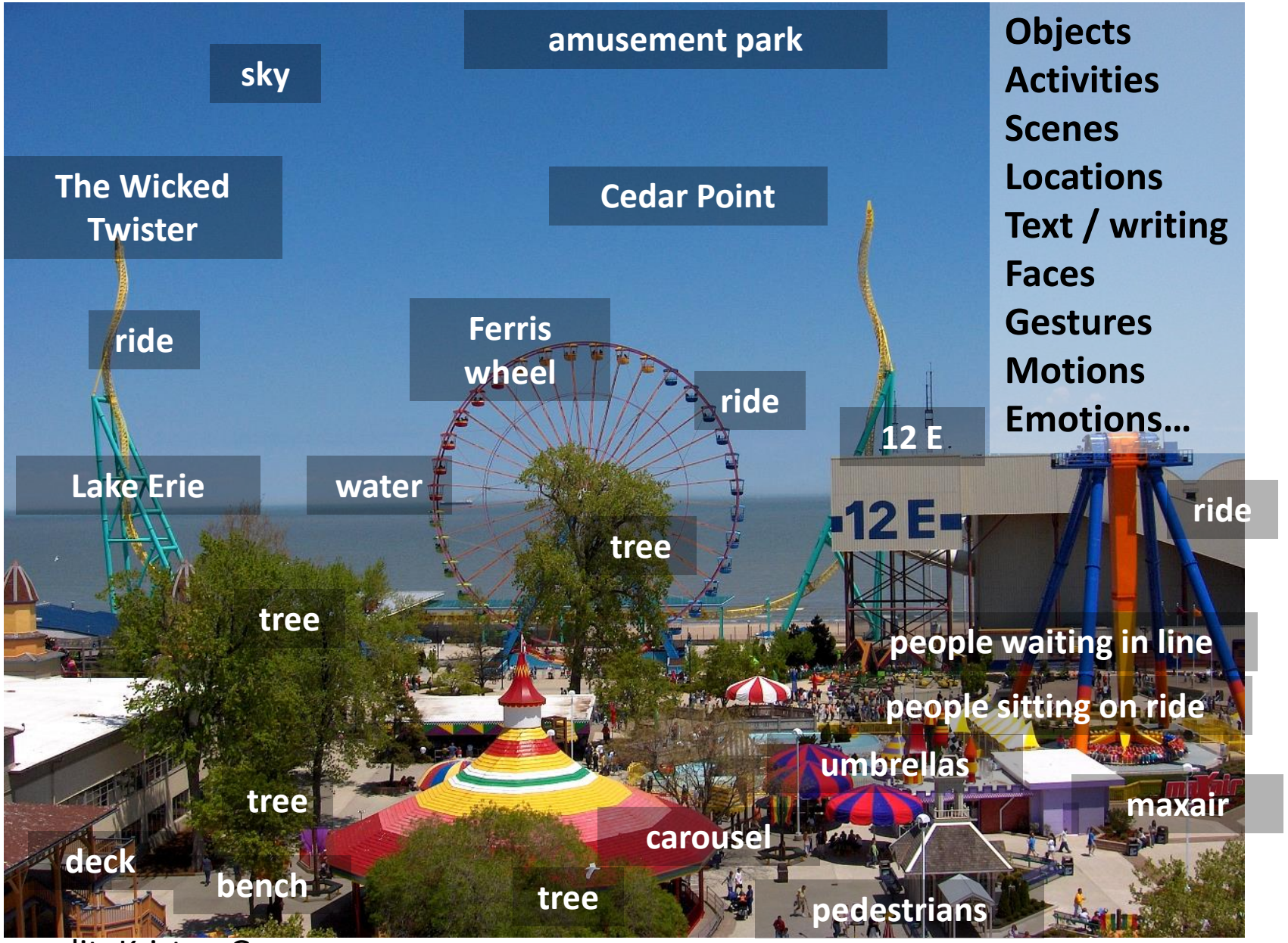
Slide credit: Kristen Grauman

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)*

Slide credit: Kristen Grauman

2. Vision for perception, interpretation



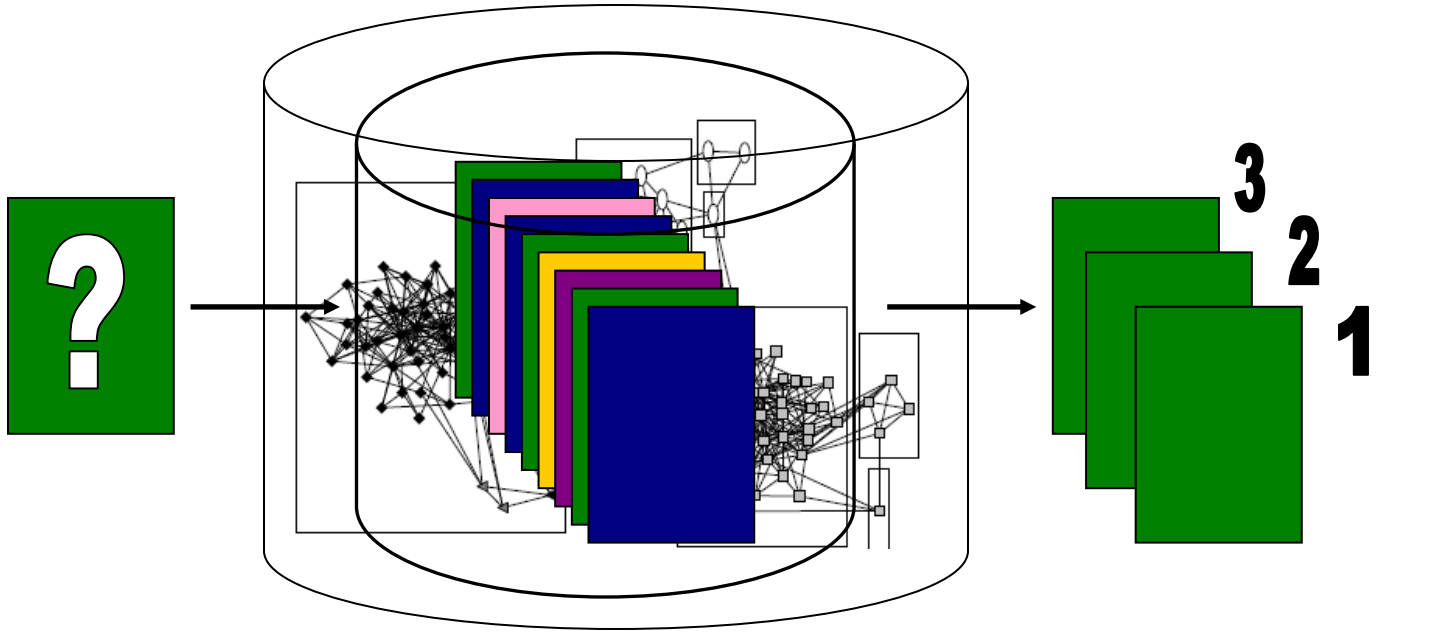
Slide credit: Kristen Grauman

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*)

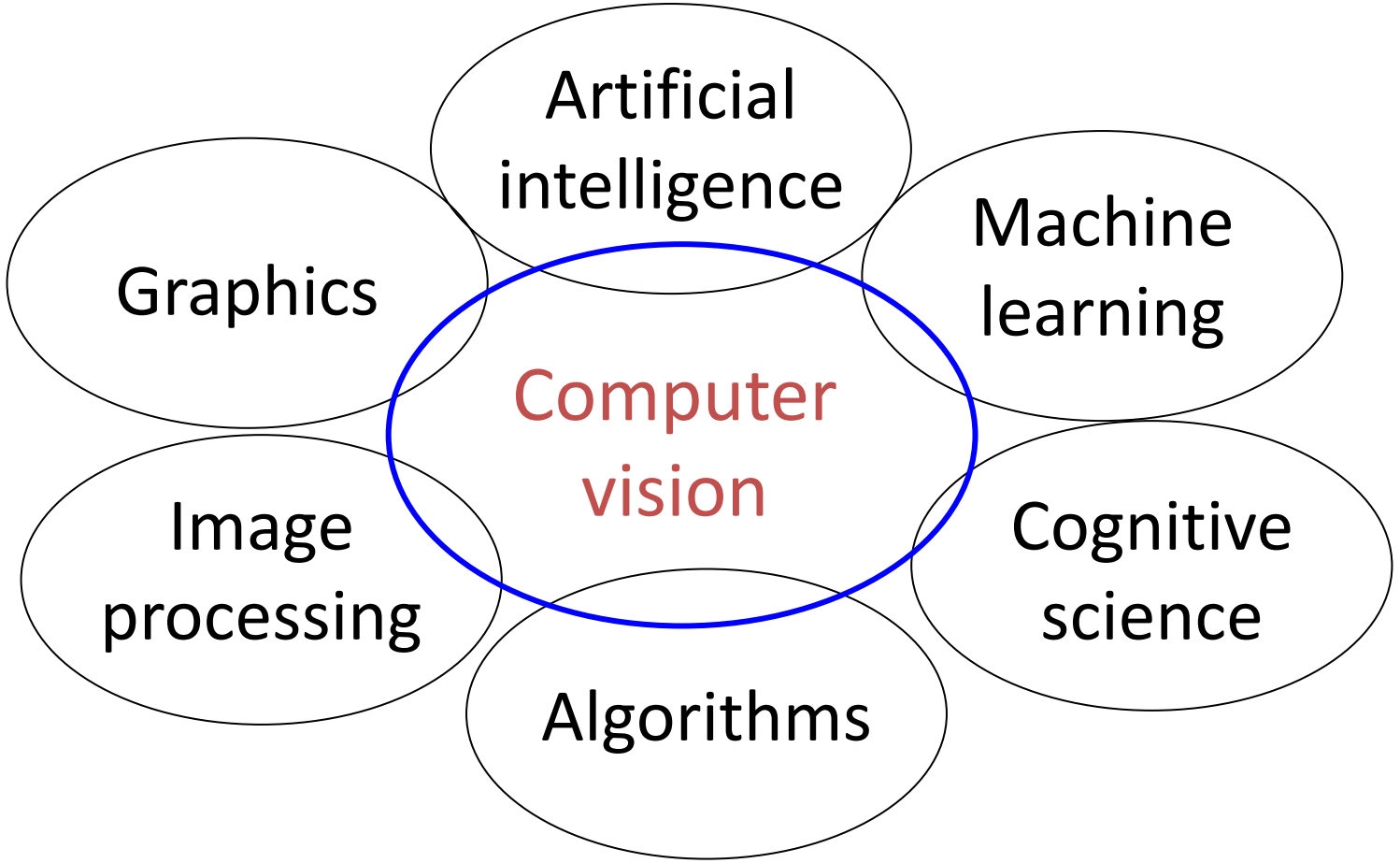
Slide credit: Kristen Grauman

3. Visual search, organization



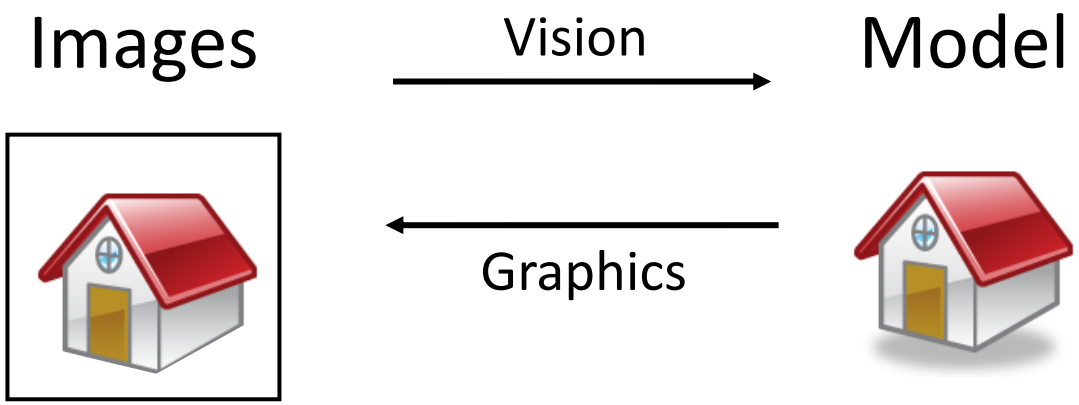
Slide credit: Kristen Grauman

Related disciplines



Slide credit: Kristen Grauman

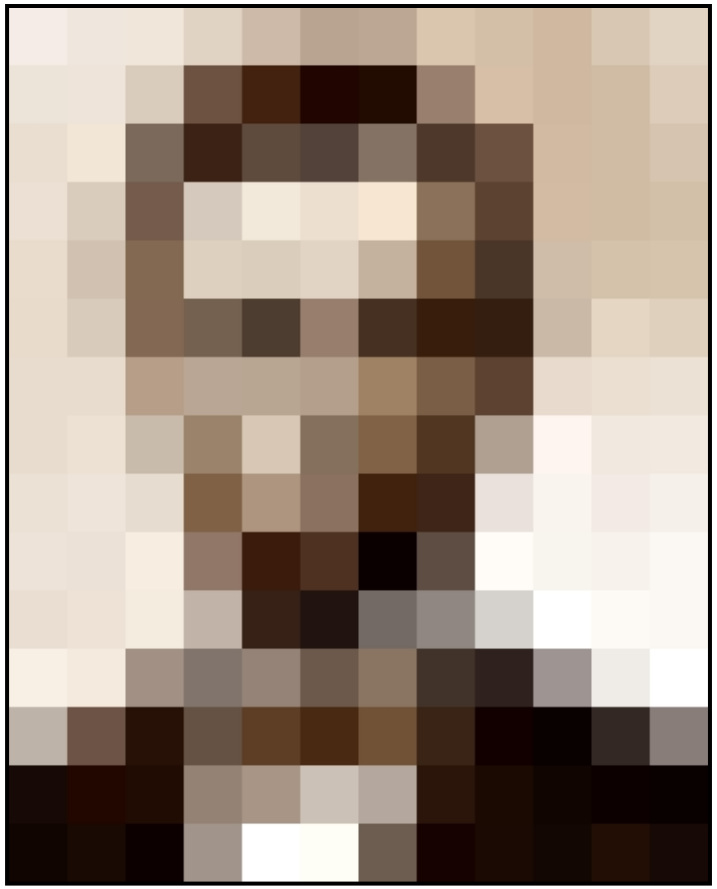
Vision and graphics



Inverse problems: analysis and synthesis.

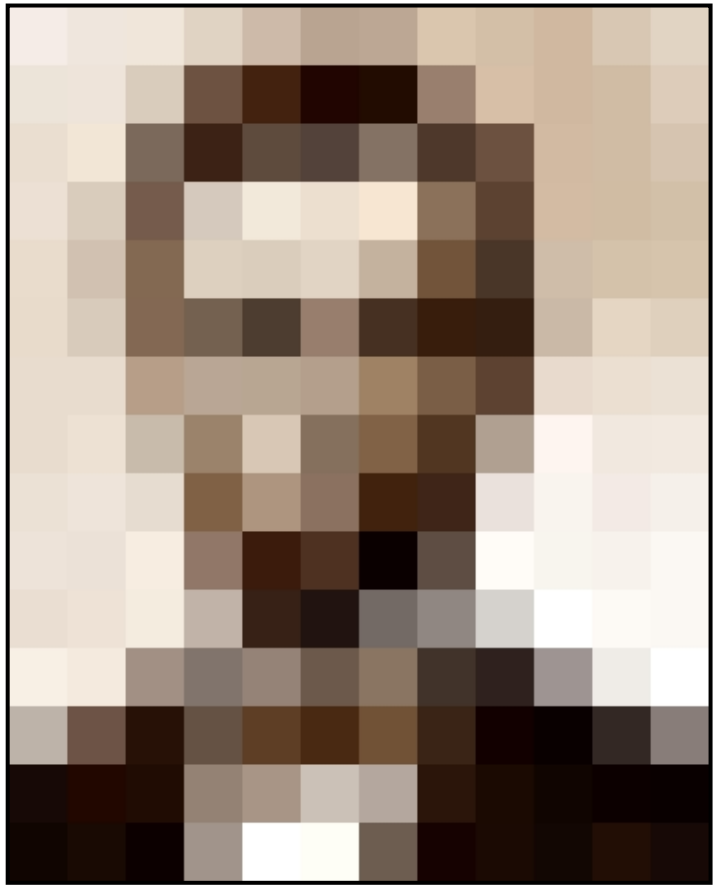
Slide credit: Kristen Grauman

What humans see



Slide credit: Larry Zitnick

What computers see



Slide credit: Larry Zitnick

What do humans see?

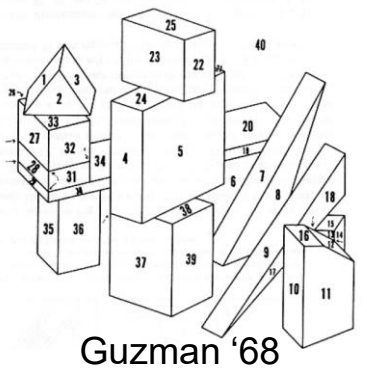
Slide credit: Larry Zitnick

Vision is really hard

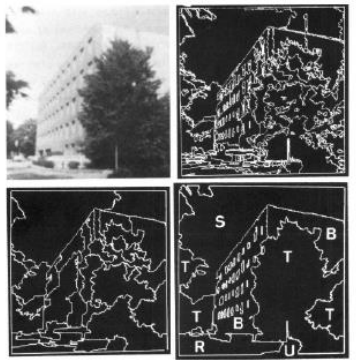
- Vision is an amazing feat of natural intelligence
  - Visual cortex occupies about 50% of Macaque brain
  - One third of human brain devoted to vision (more than anything else)

Ridiculously brief history of computer vision

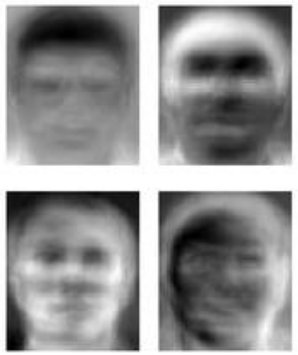
- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis 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?



Guzman '68



Ohta Kanade '78



Turk and Pentland '91

How vision is used now

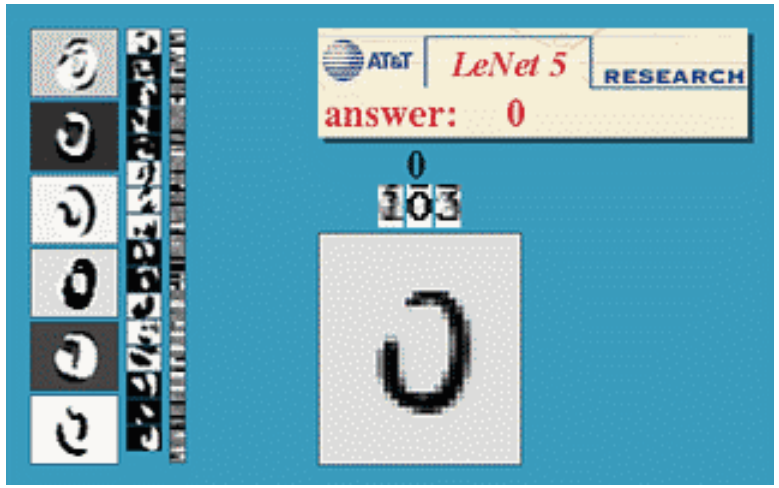
- Examples of real-world applications

Some of the following slides by Steve Seitz

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

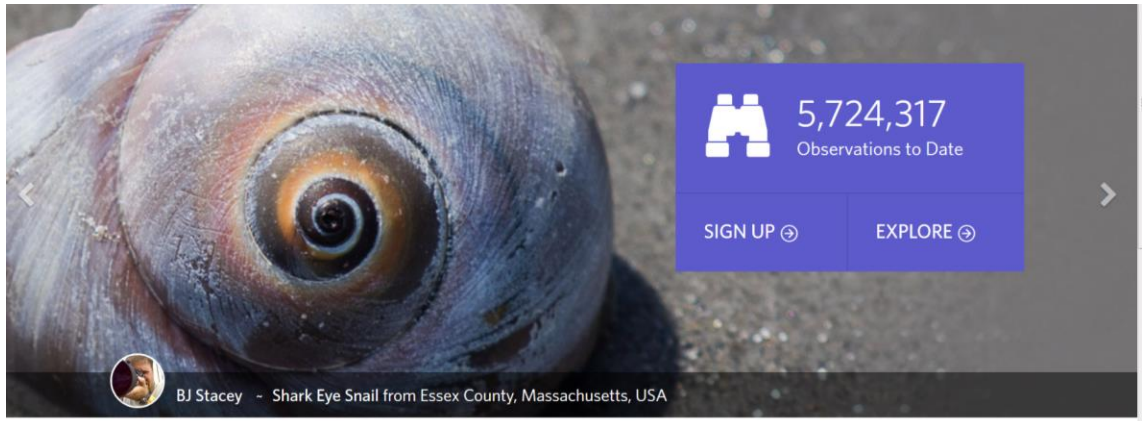
Vision in space

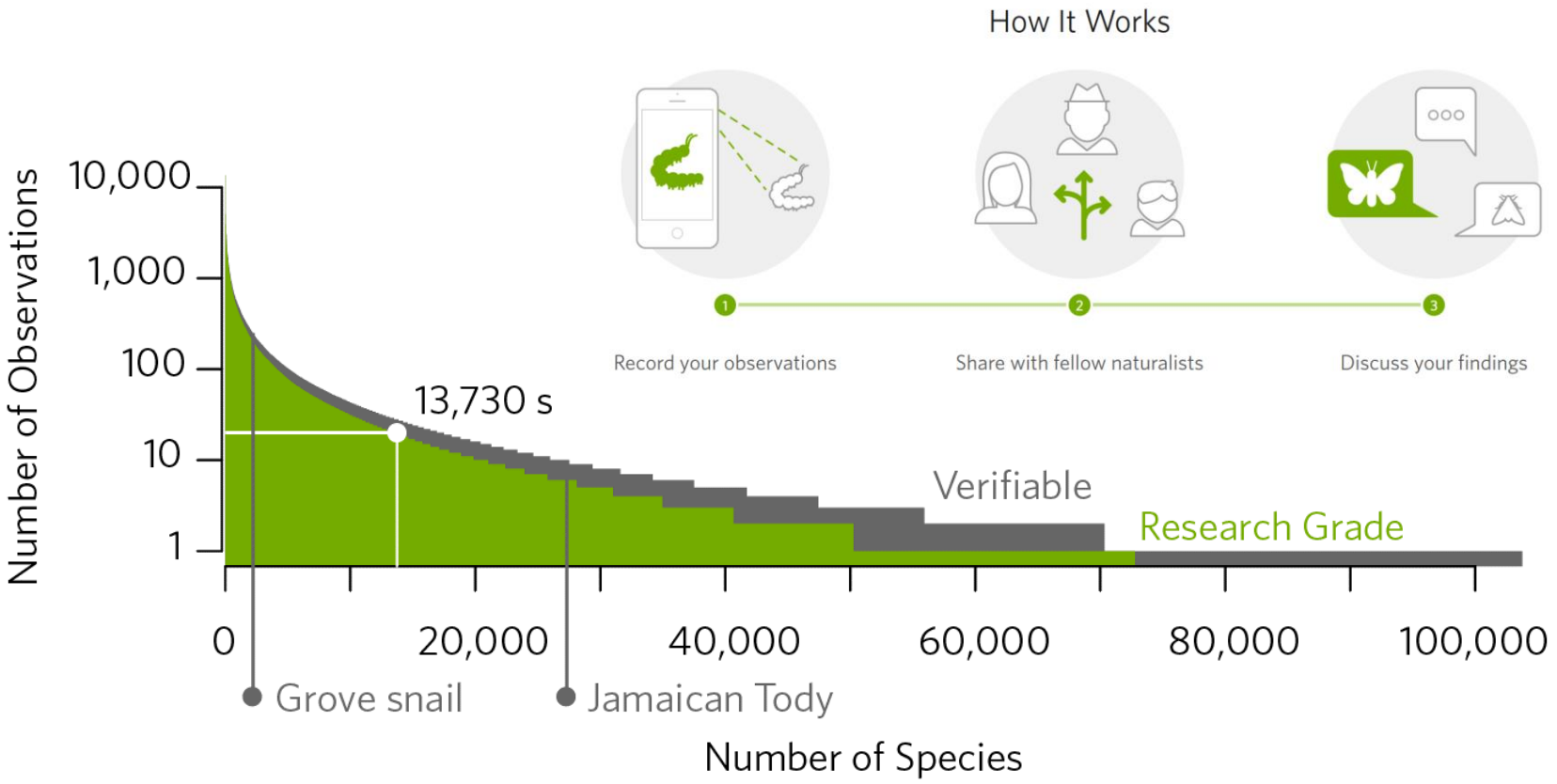
NASA'S Mars Exploration Rover Spirit captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

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



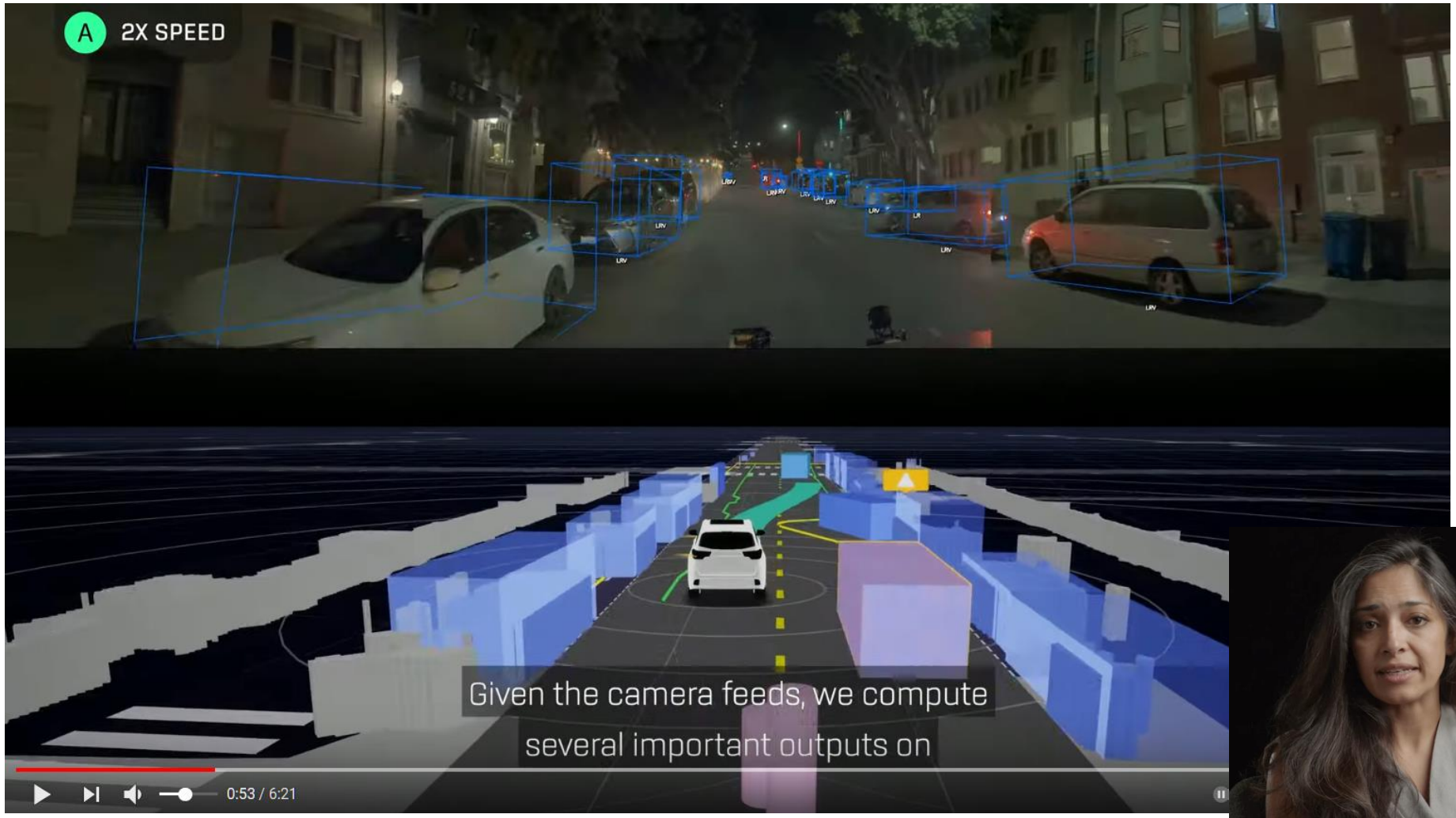


https://www.inaturalist.org/pages/computer_vision_demo

Skydio

https://www.skydio.com/

Zoox Computer Vision Demo

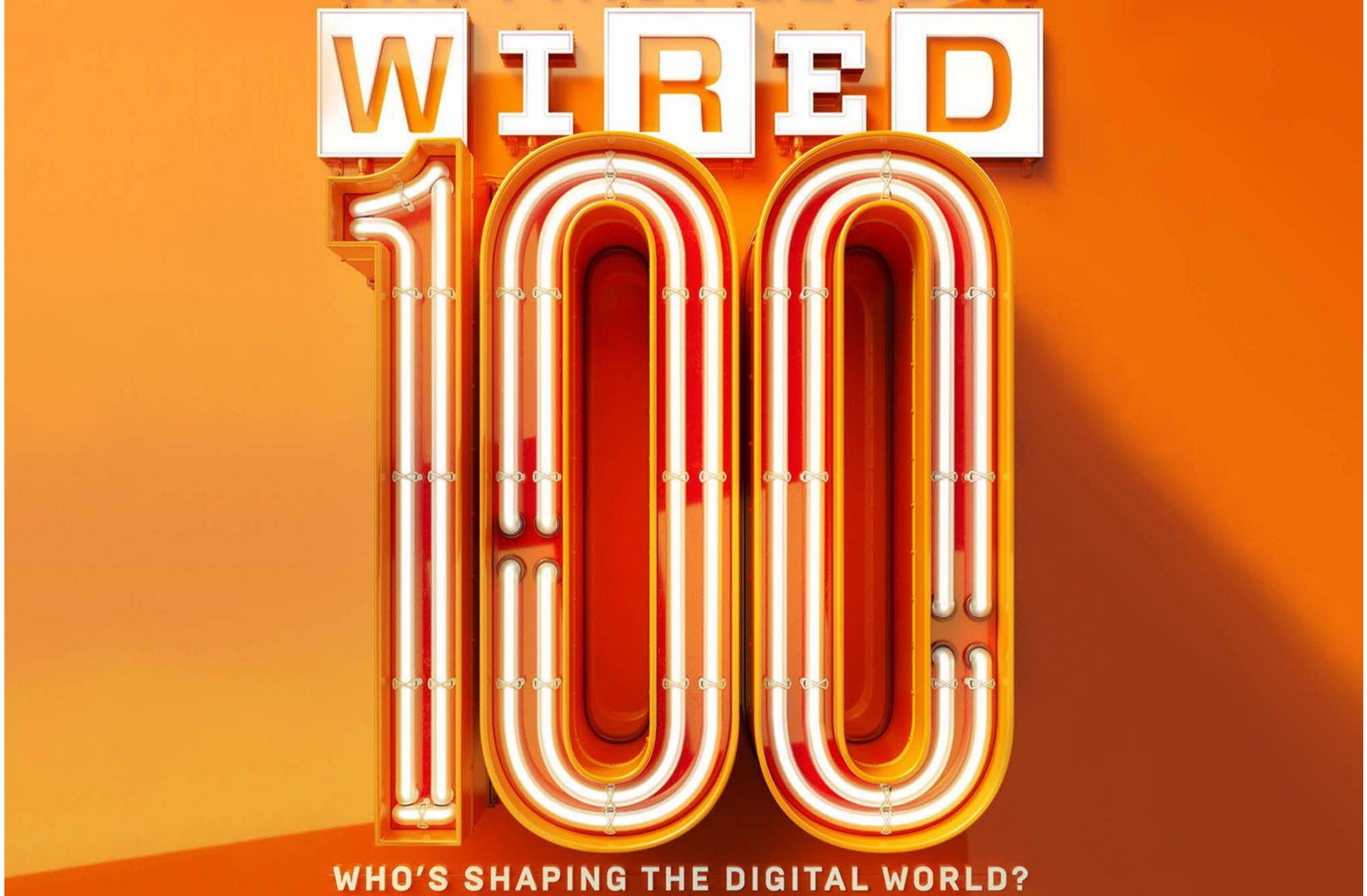


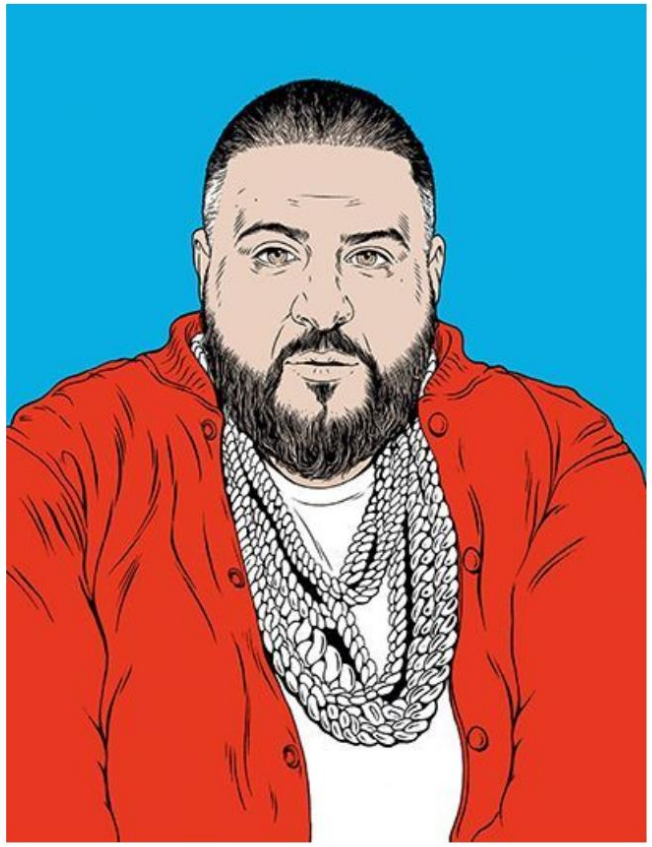
https://www.youtube.com/watch?v=BVRMh9NO9Cs

State of the art today?

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

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


WIRED
100
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

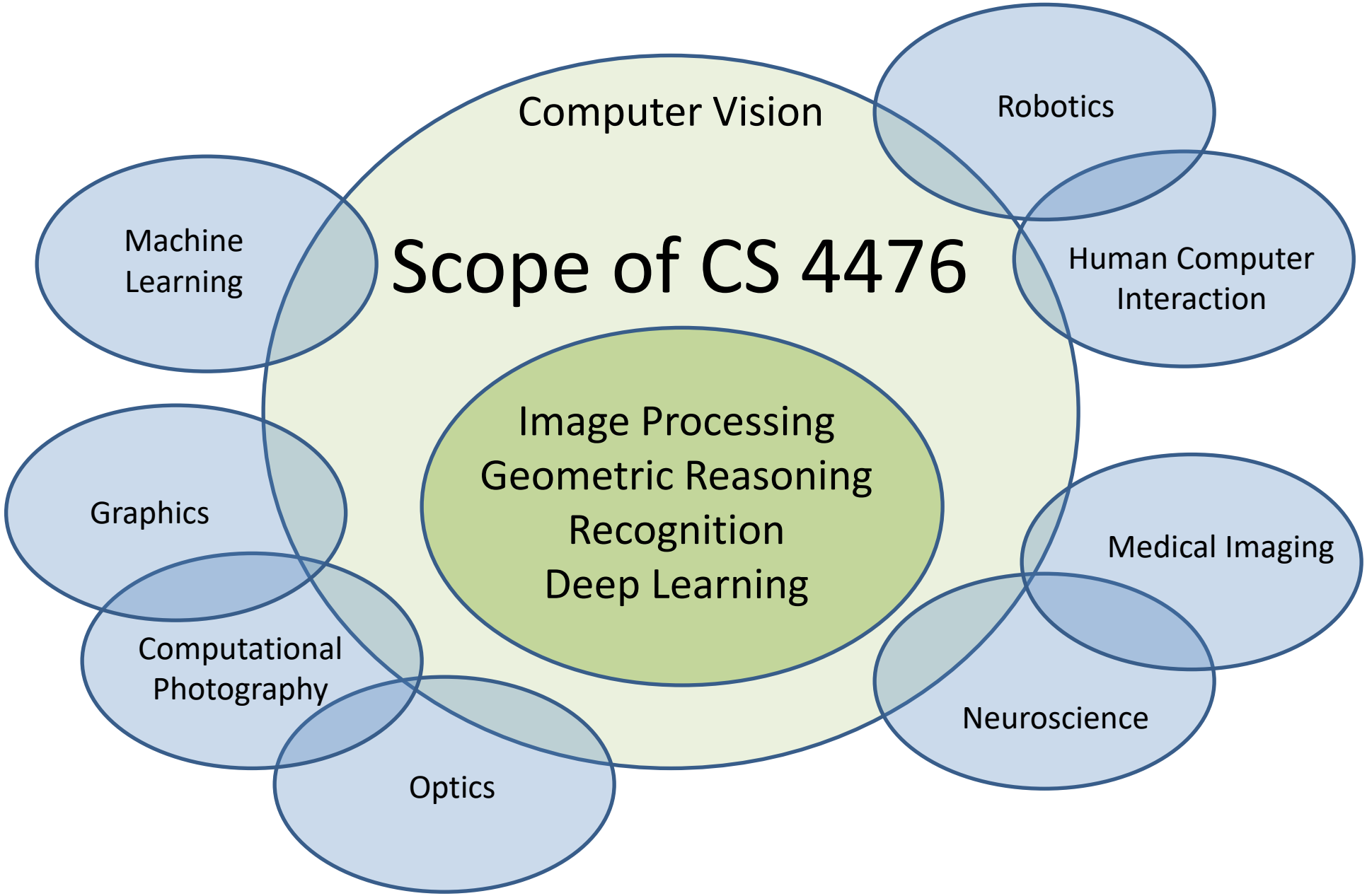
Entertainer, Los Angeles

Today’s Class

- ~~Covid safety~~
- ~~Who am I?~~
- ~~What is Computer Vision?~~
- Specifics of this course
- Geometry of Image Formation
- Questions

Grading

- 80% programming projects (5 total)
- 20% Quizzes or Problem sets


Computer Vision
Scope of CS 4476
Image Processing
Geometric Reasoning
Recognition
Deep Learning
Robotics
Human Computer Interaction
Machine Learning
Graphics
Computational Photography
Optics
Medical Imaging
Neuroscience

Textbook

Computer Vision: Algorithms and Applications, 2nd ed.

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

You need a decent computer

You may want to buy a month of Google Colab Pro near the end of the semester

Tentative Projects

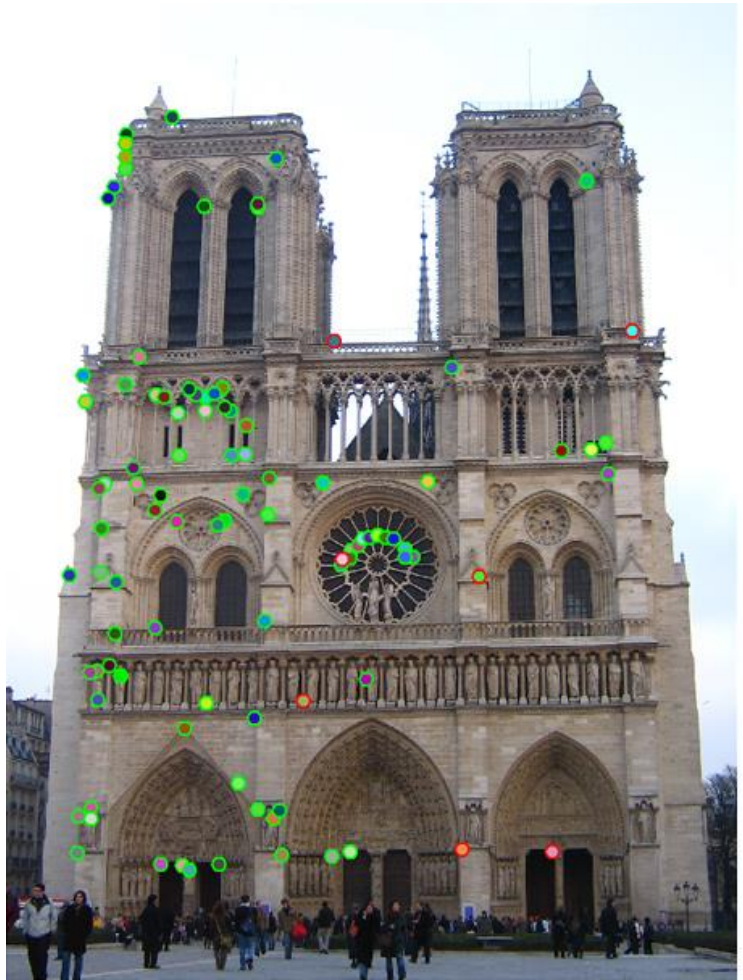
- (project 0 to test environment setup and handin)
- Image Filtering and Hybrid Images
- Local Feature Matching
- Camera Calibration and Fundamental Matrix Estimation with RANSAC
- Image Classification with Deep Learning
- Semantic Segmentation with Deep Learning

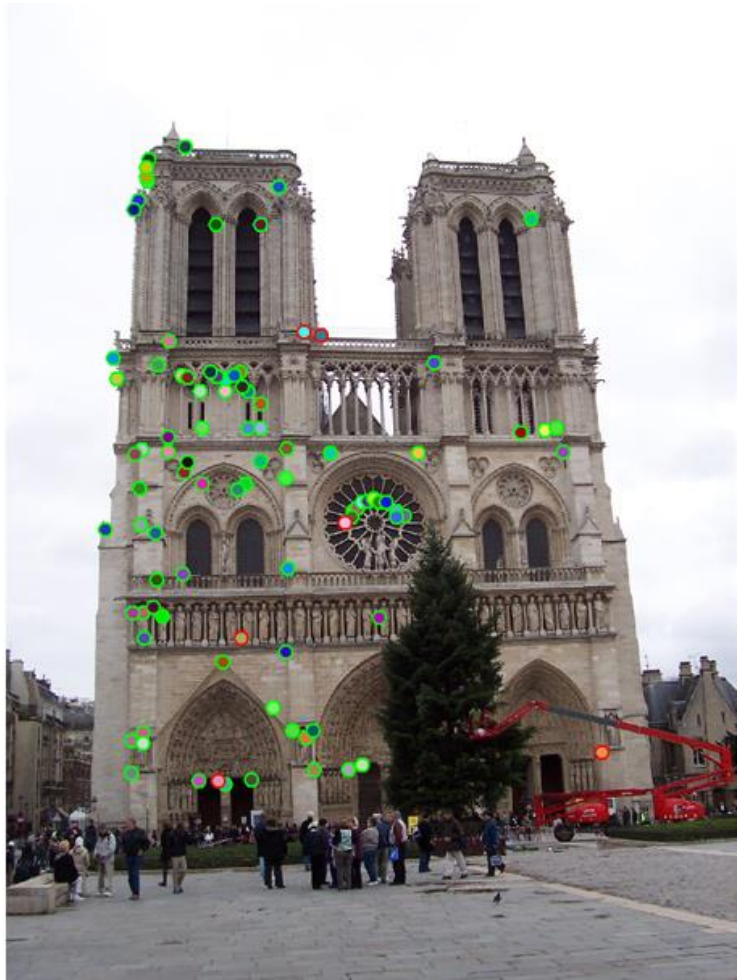
Proj1: Image Filtering and Hybrid Images

- 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.

Today's Class

- ~~Covid safety~~
- ~~Who am I?~~
- ~~What is Computer Vision?~~
- ~~Specifics of this course~~
- Geometry of Image Formation
- Questions


WÓDKA
Cuisinart
SNOW CRYSTALS
FOTOGRAFIE UND DAS UNSICHTBARE
DER KLEINE STOWASSER
MATHEMATICS
MAN AND SPACE
DRUGS
STACK-ON


WÓDKA
VODKA
Cuisinart
SNOW CRYSTALS
MATHEMATICS
MAN AND SPACE
DRUGS
STACK-ON

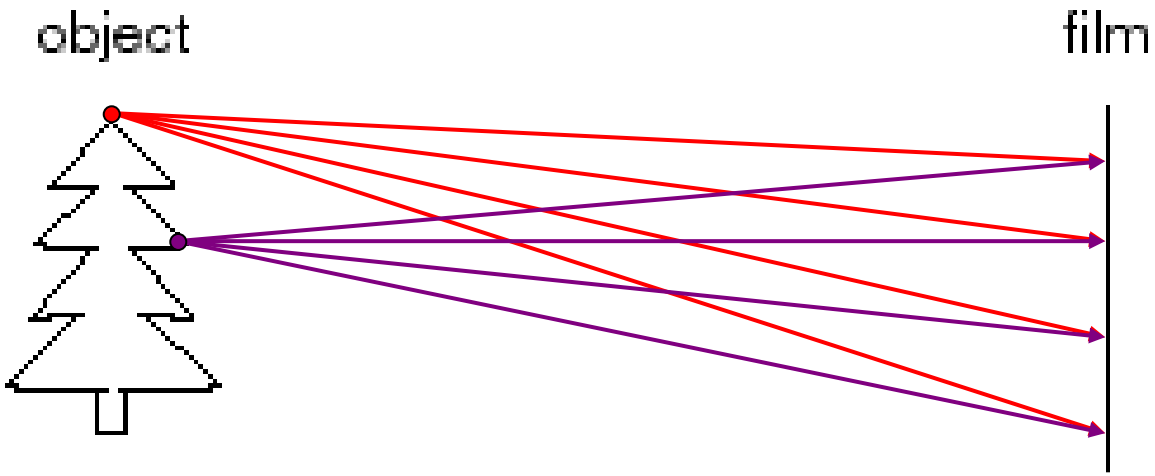
The Geometry of Image Formation

Mapping between image and world coordinates

- Pinhole camera model
- Projective geometry
  - Vanishing points and lines
- Projection matrix

What do you need to make a camera from scratch?

Image formation

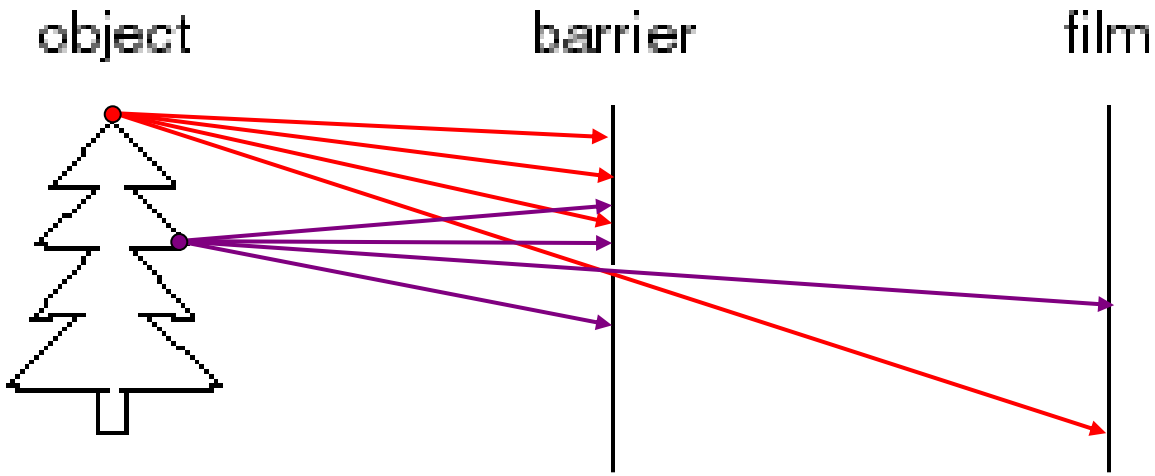


Let’s design a camera

- Idea 1: put a piece of film in front of an object
- Do we get a reasonable image?

Slide source: Seitz

Pinhole camera

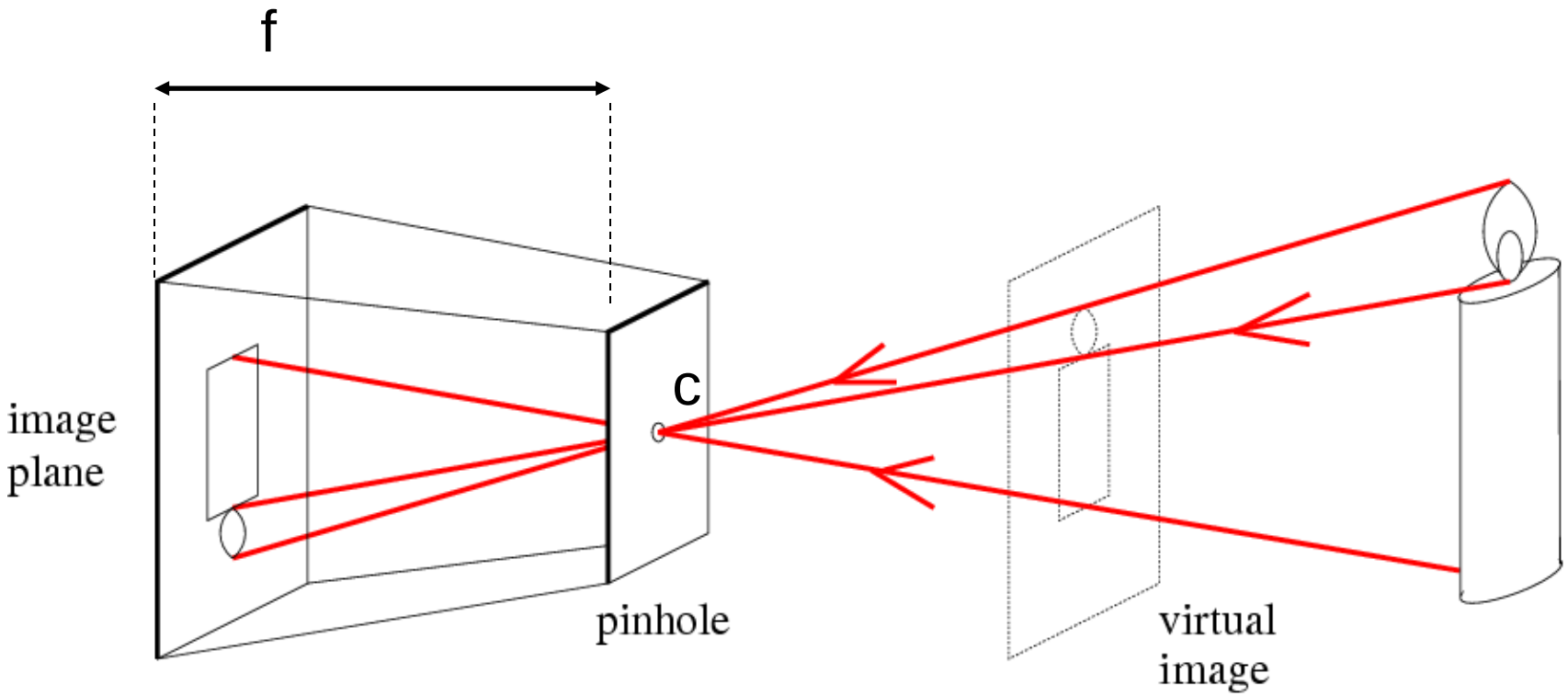


Idea 2: add a barrier to block off most of the rays

- This reduces blurring
- The opening known as the **aperture**

Slide source: Seitz

Pinhole camera



f = focal length
c = center of the camera

Figure from Forsyth

Camera obscura: the pre-camera

- Known during classical period in China and Greece (e.g. Mo-Ti, China, 470BC to 390BC)

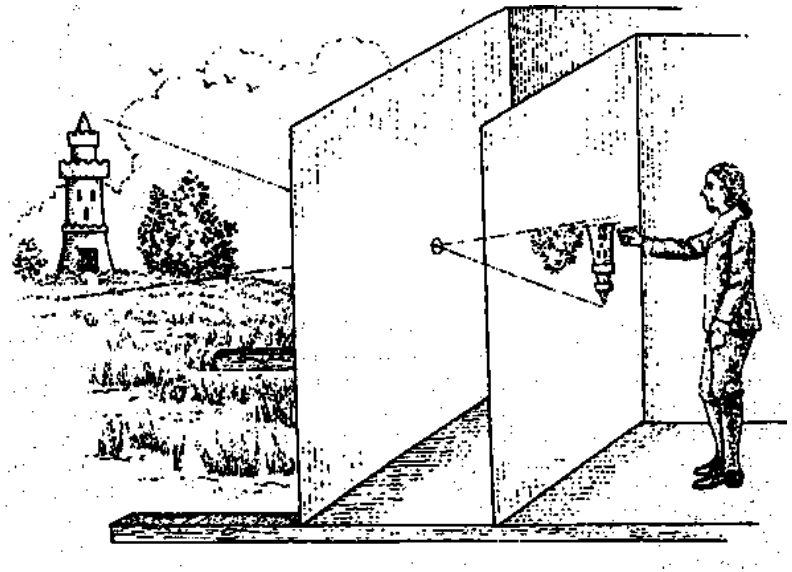
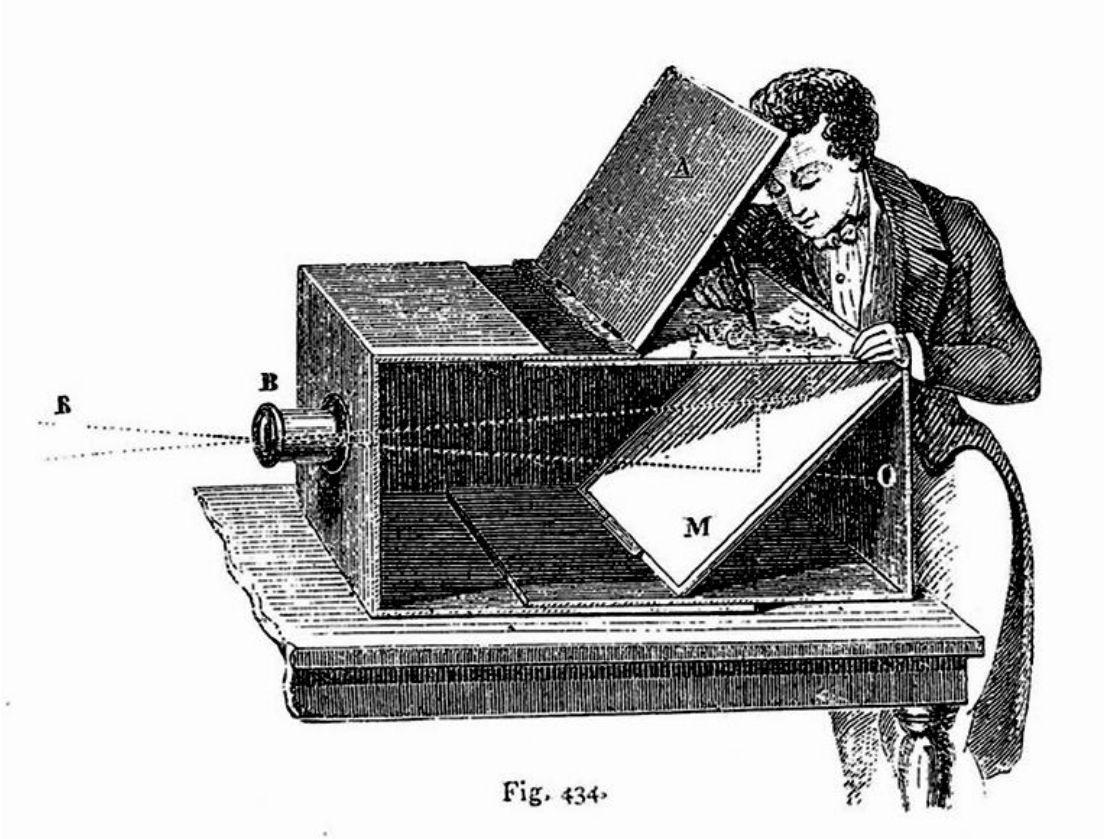


Illustration of Camera Obscura

Freestanding camera obscura at UNC Chapel Hill

Photo by Seth Ilys

Camera Obscura used for Tracing



Lens Based Camera Obscura, 1568

Accidental Cameras

Accidental Pinhole and Pinspeck Cameras
Revealing the scene outside the picture.
Antonio Torralba, William T. Freeman

Accidental Cameras

a) Input (occluder present)

b) Reference (occluder absent)

c) Difference image (b-a)

d) Crop upside down

e) True view

First Photograph

Oldest surviving photograph

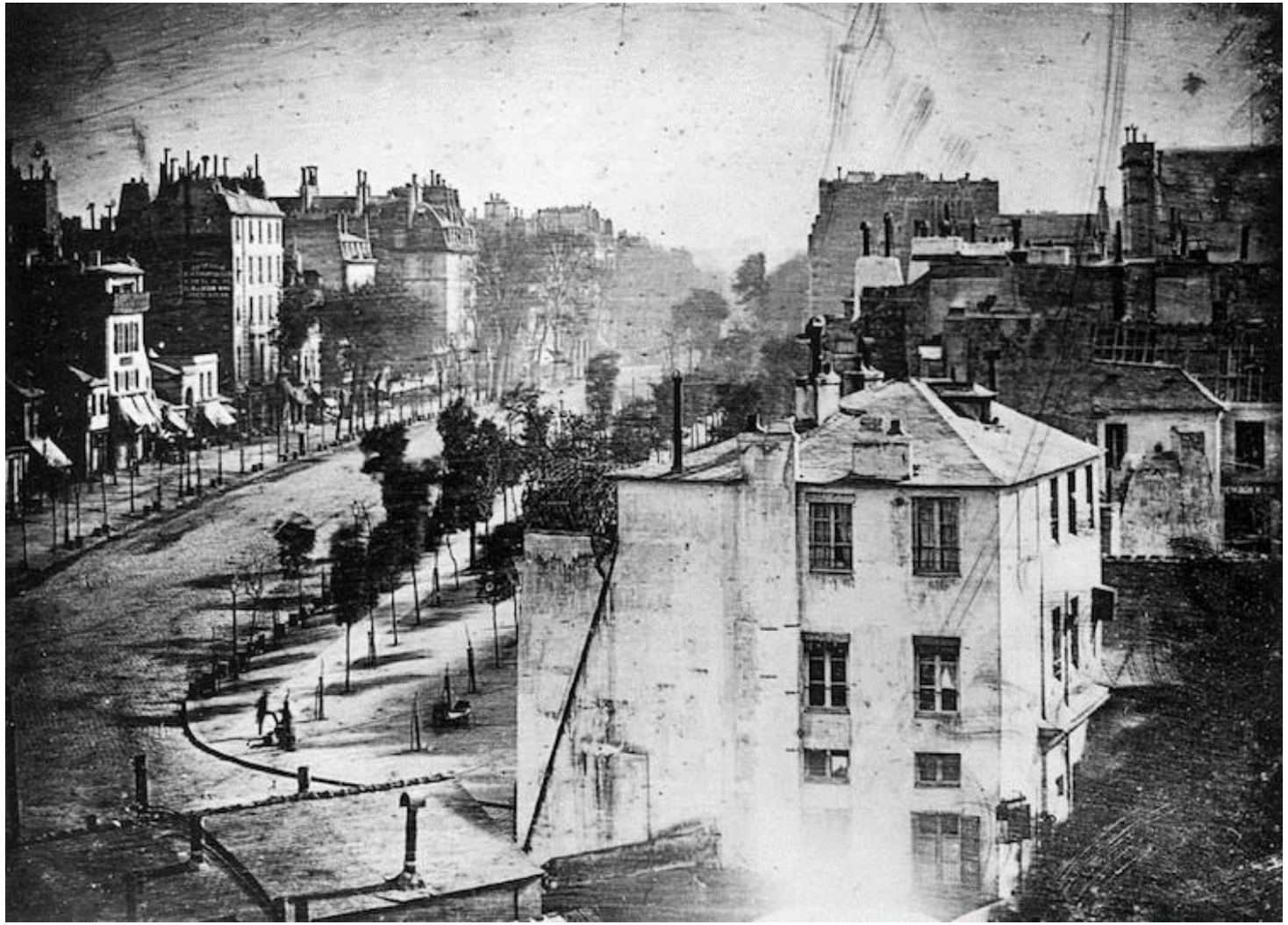
- Took 8 hours on pewter plate

Joseph Niepce, 1826

Photograph of the first photograph

Stored at UT Austin

Niepce later teamed up with Daguerre, who eventually created Daguerrotypes



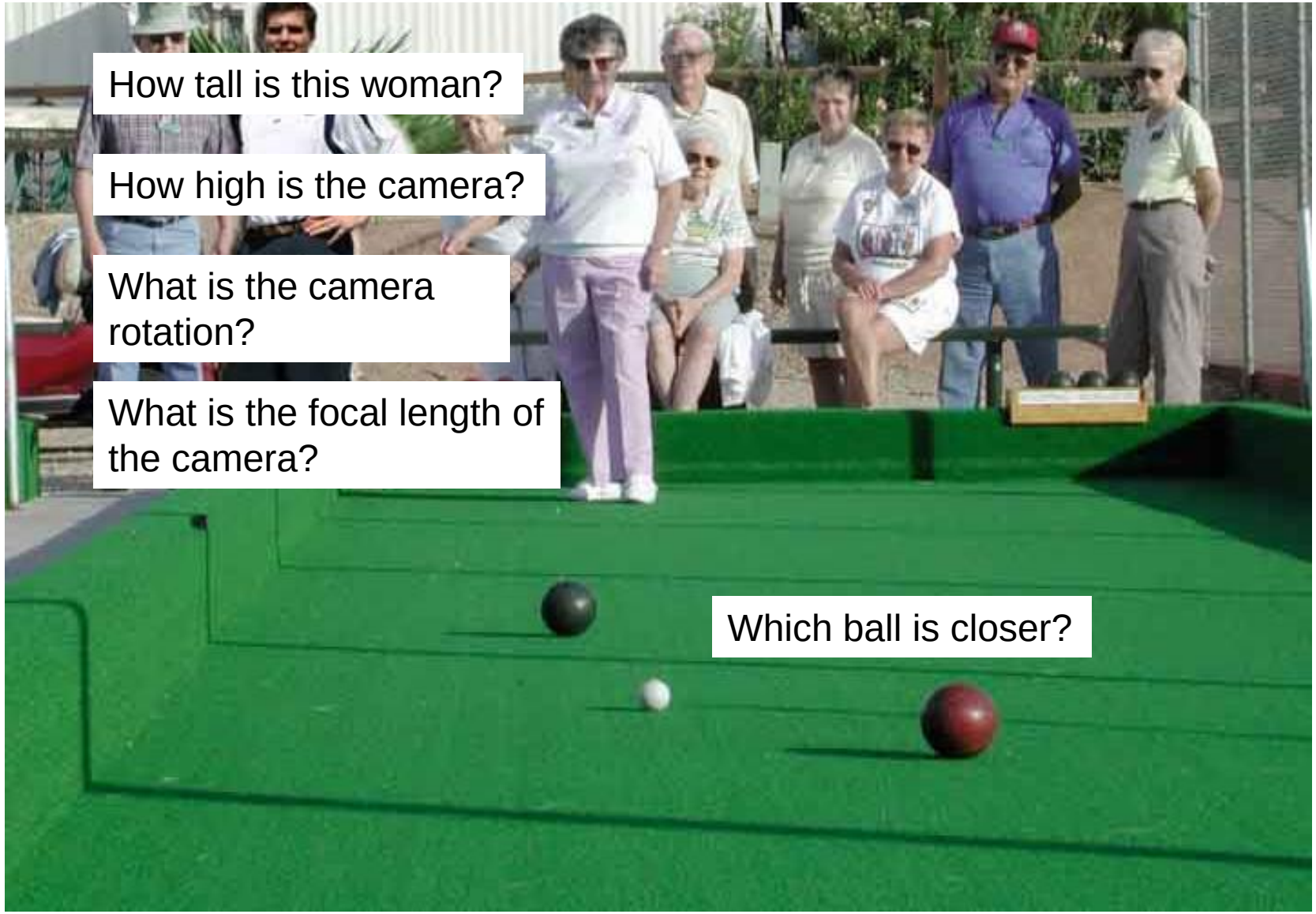
“Louis Daguerre—the inventor of daguerreotype—shot what is not only the world's oldest photograph of Paris, but also the first photo with humans. The 10-minute long exposure was taken in 1839 in Place de la République and it's just possible to make out two blurry figures in the left-hand corner.”

Source

- Great history lesson on the chemistry and engineering challenges of early photography from the “Technology Connections” YouTube channel.

https://www.youtube.com/watch?v=wbbH77rYaa8&list=PLv0jwu7G_DFV6yW240e6CbiwCLaZ0Z6PV

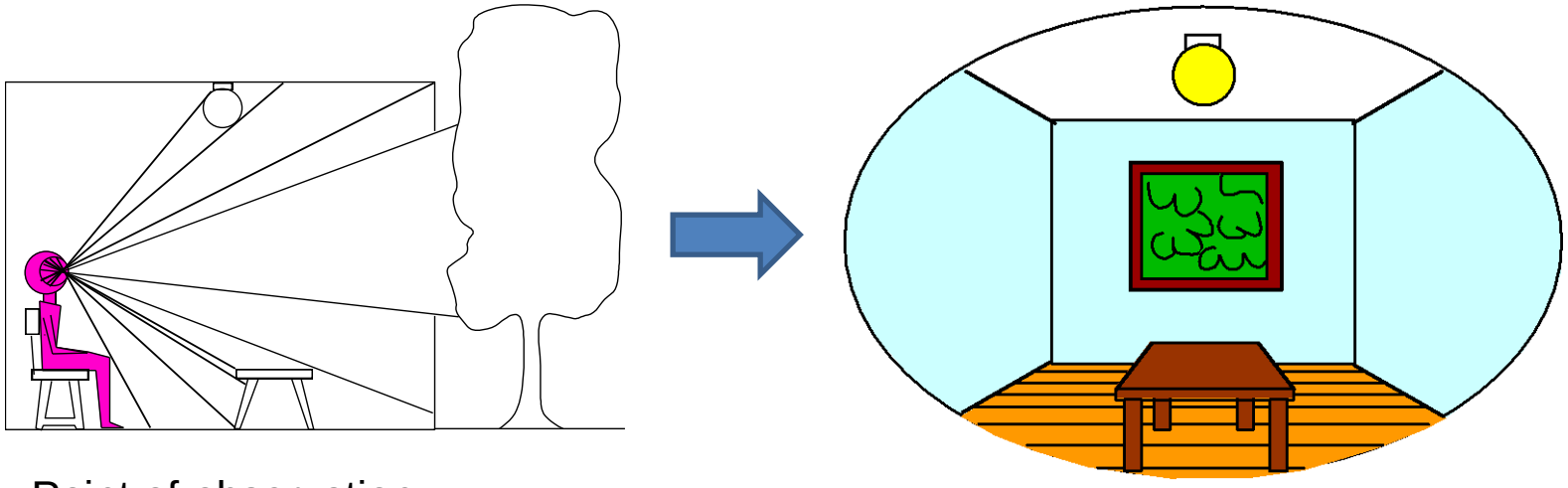
Camera and World Geometry



Dimensionality Reduction Machine (3D to 2D)

3D world

2D image



Point of observation

Slide source: Seitz

Projection can be tricky…

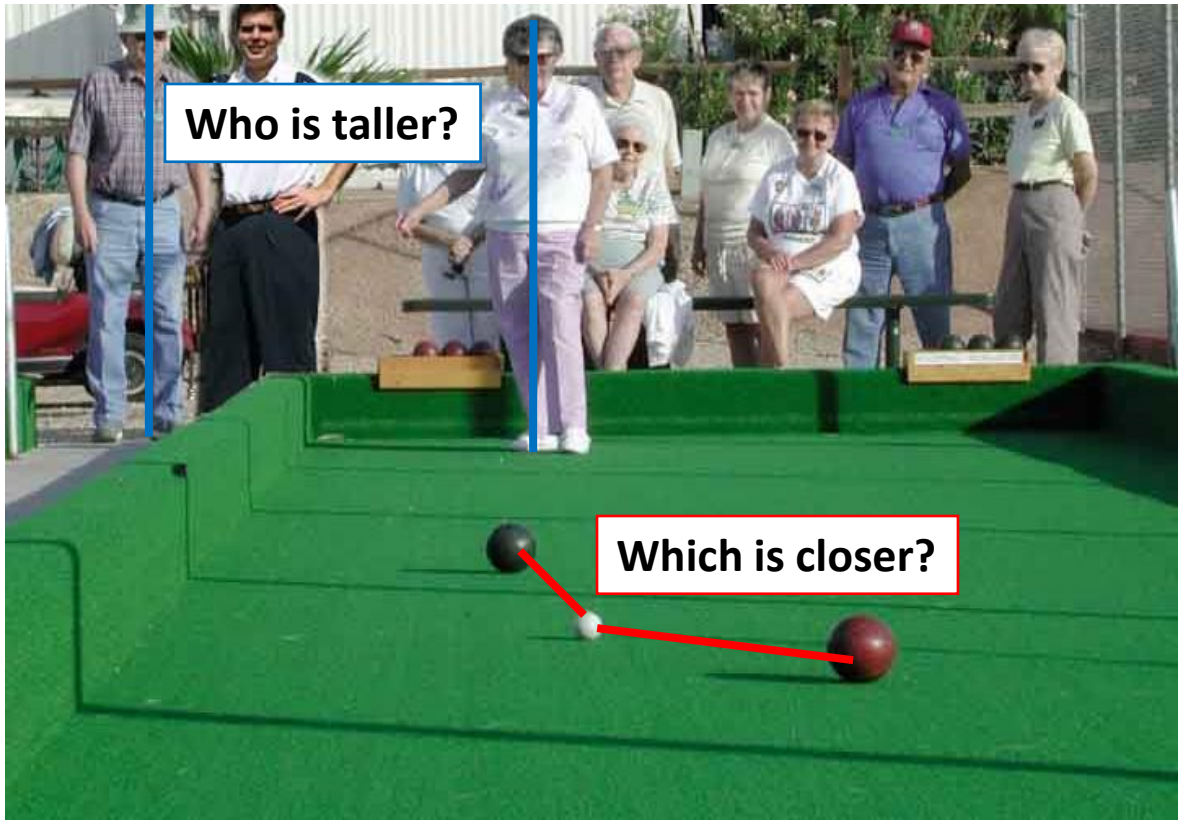
Slide source: Seitz

Projection can be tricky…

Projective Geometry

What is lost?

- Length



Length and area are not preserved

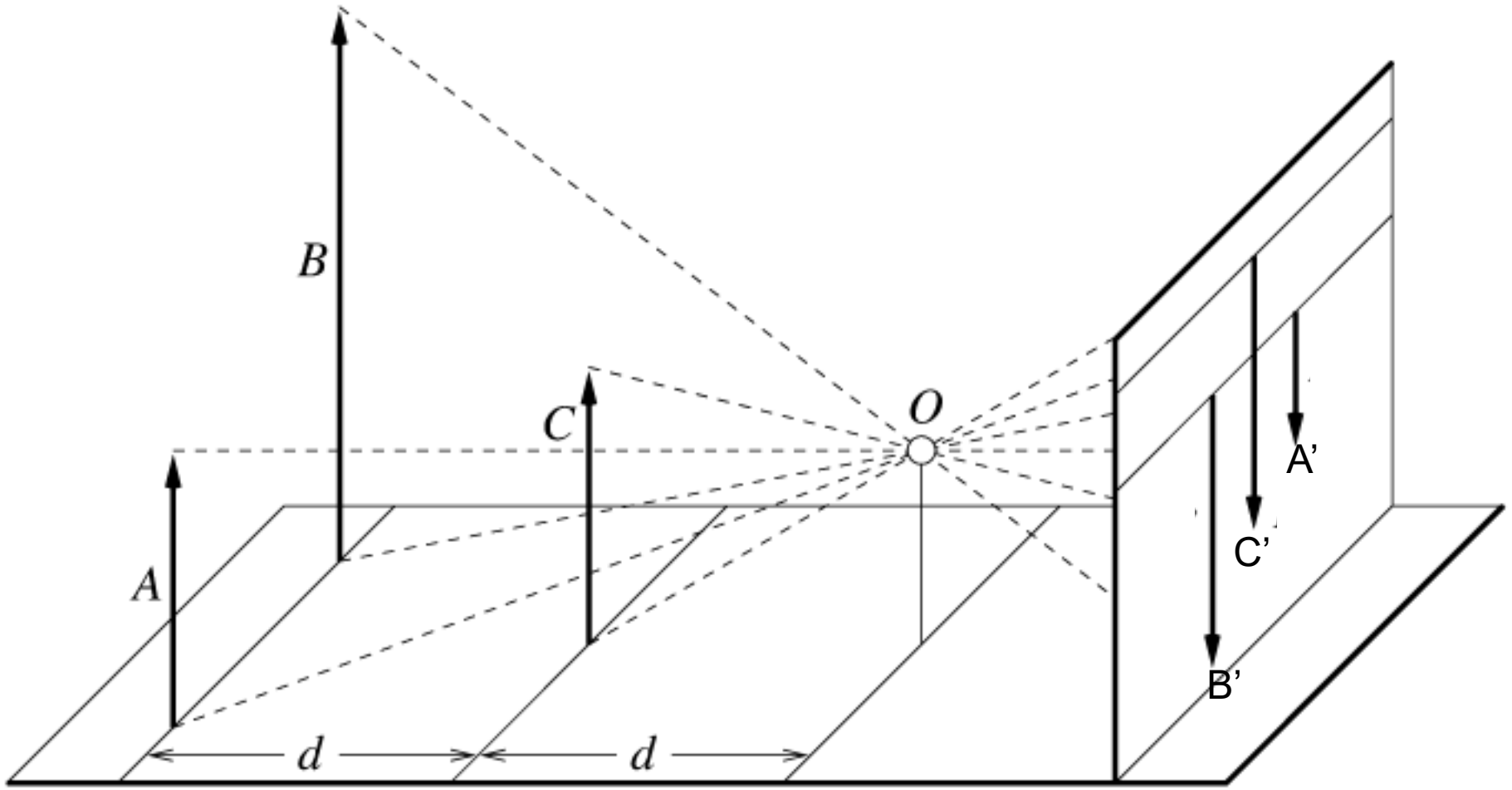
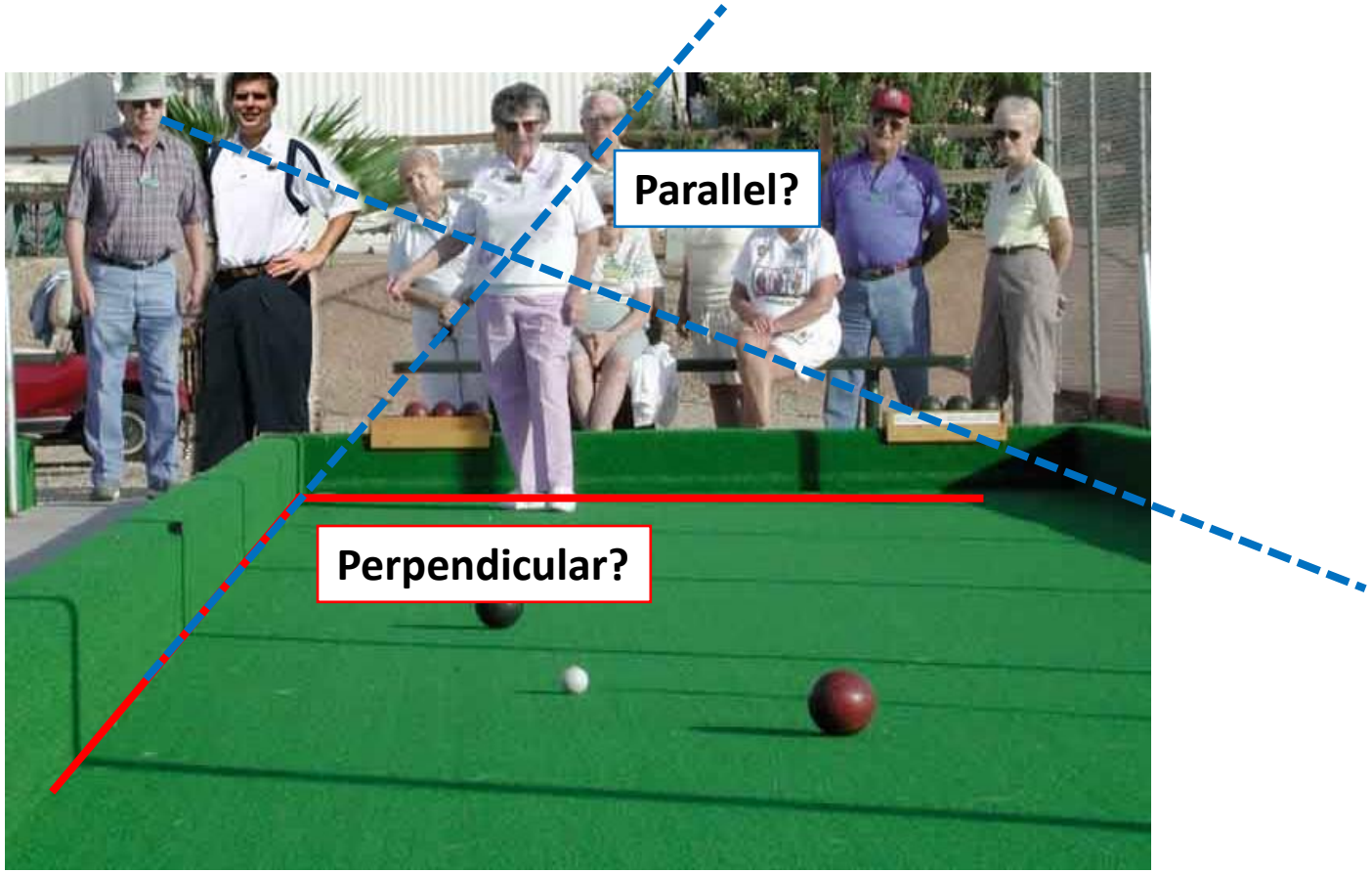


Figure by David Forsyth

Projective Geometry

What is lost?

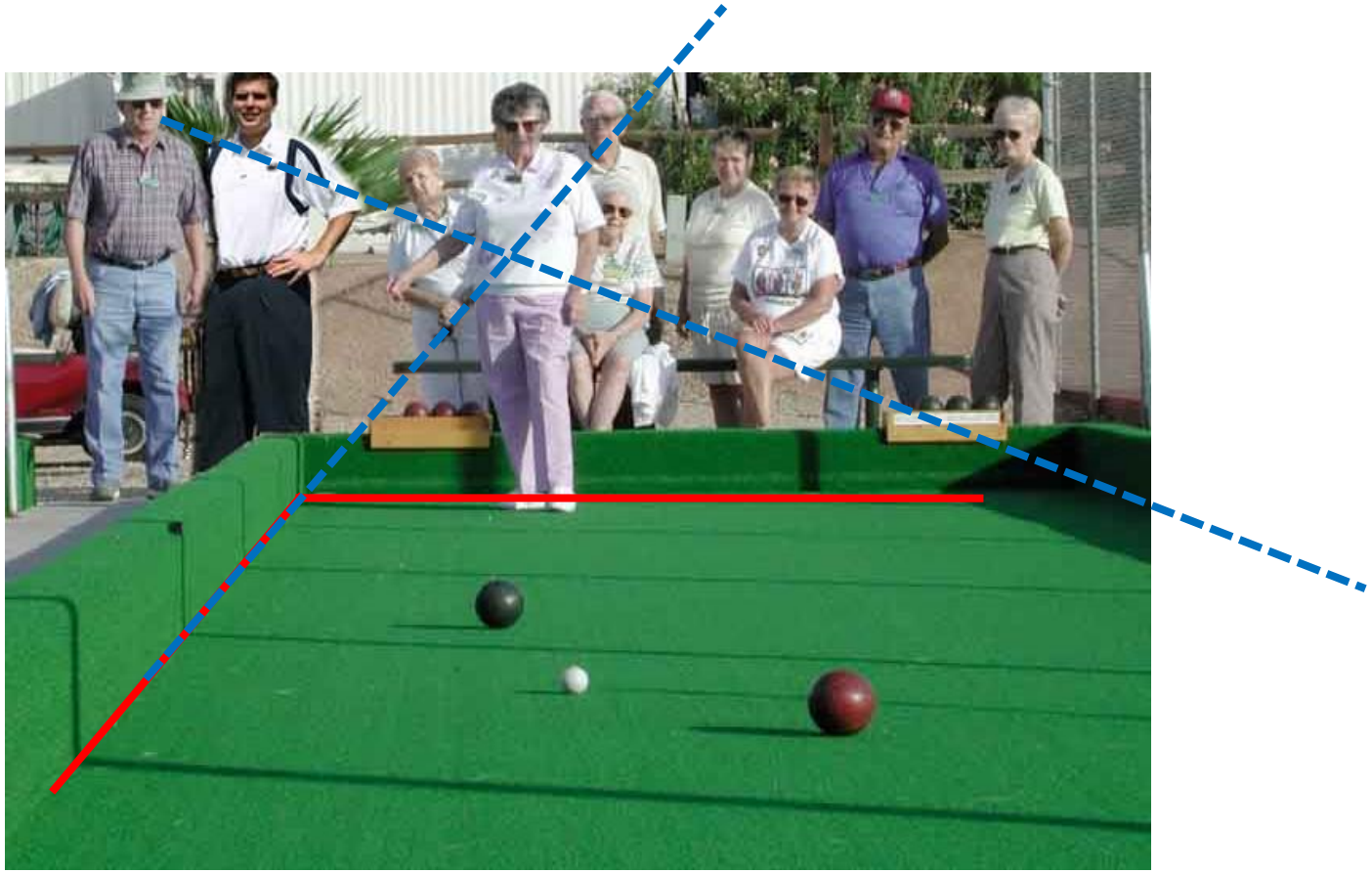
- Length
- Angles



Projective Geometry

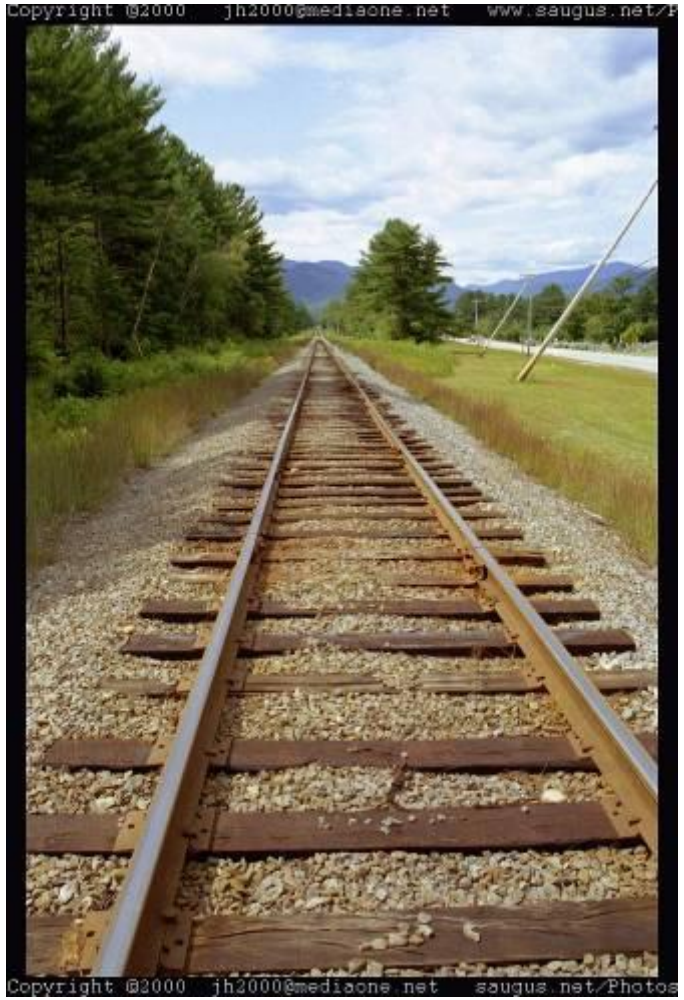
What is preserved?

- Straight lines are still straight

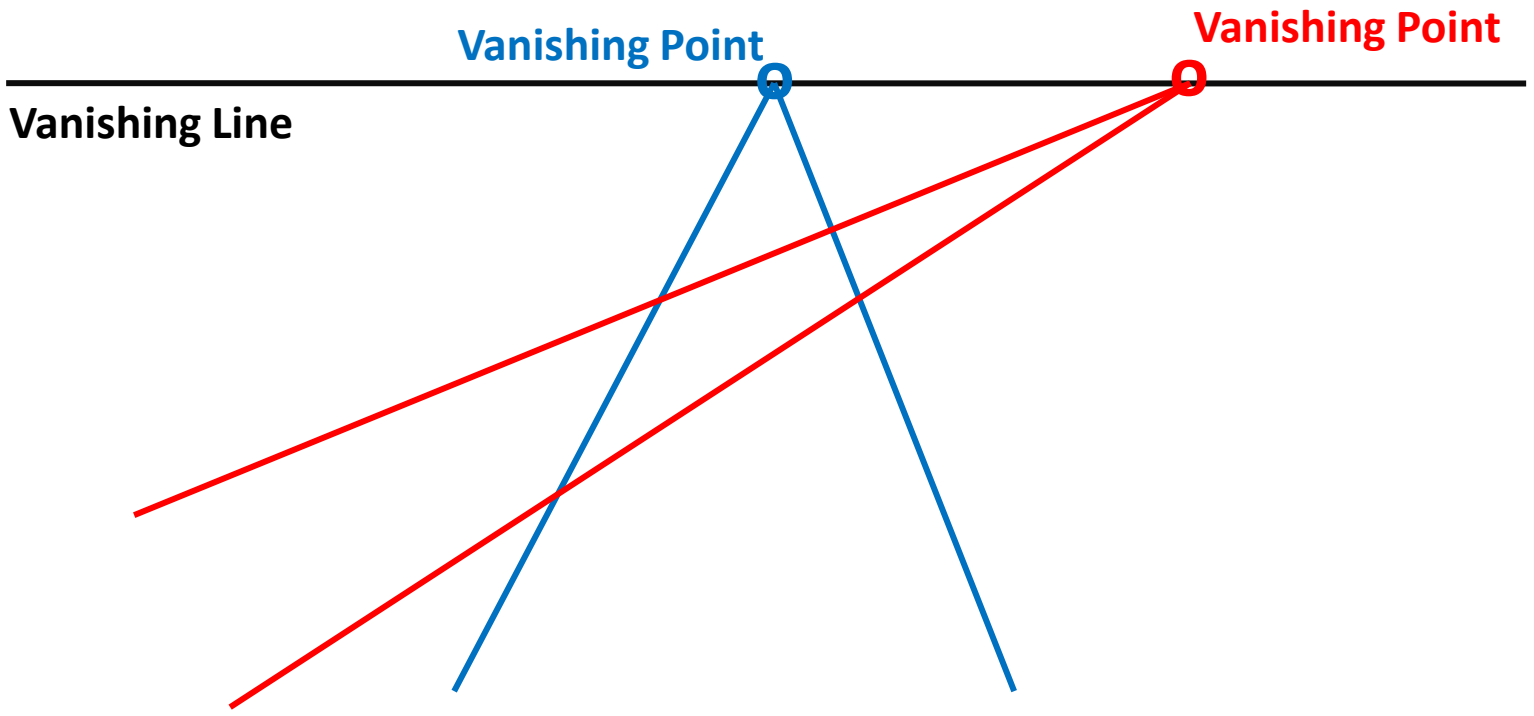


Vanishing points and lines

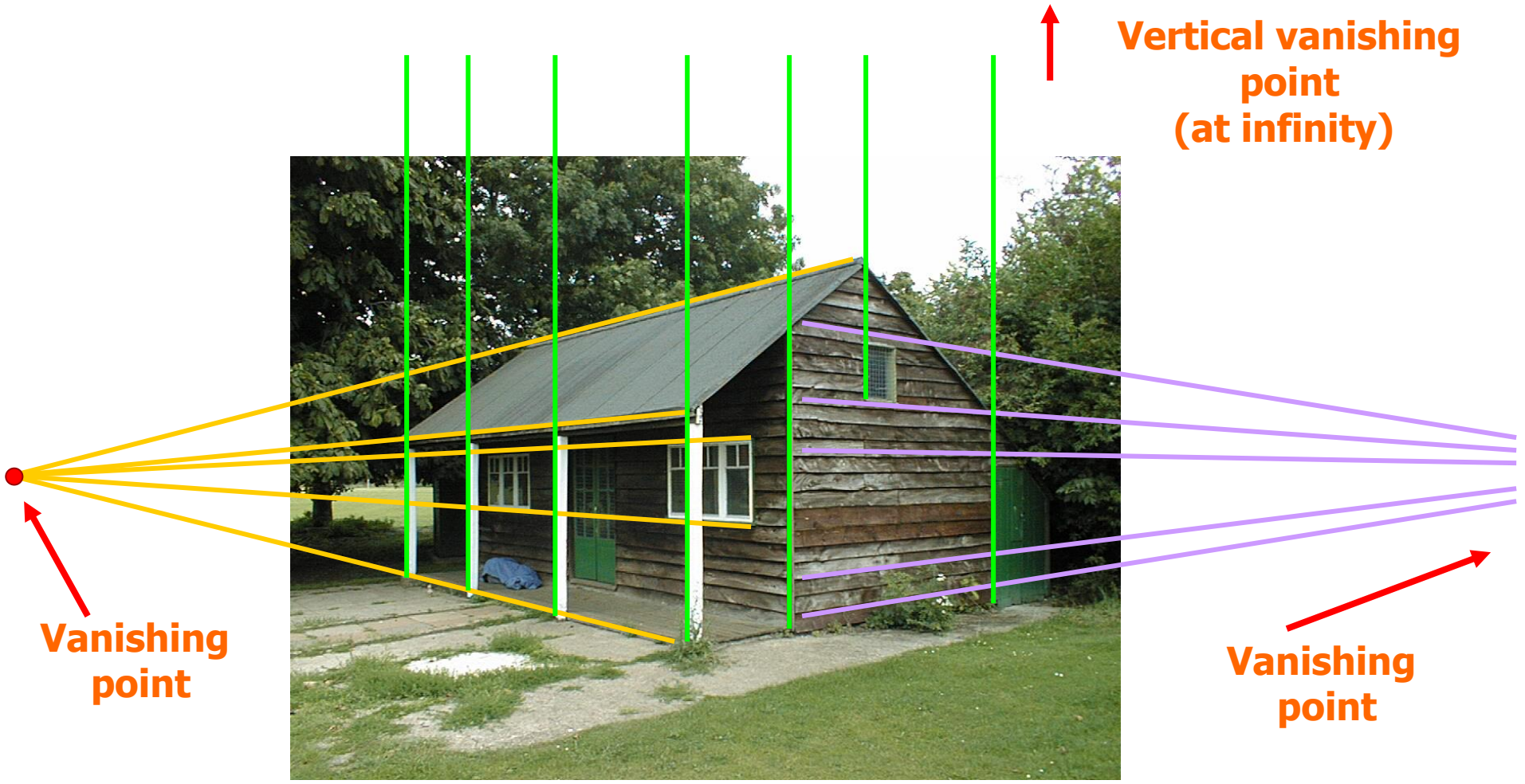
Parallel lines in the world intersect in the image at a “vanishing point”



Vanishing points and lines

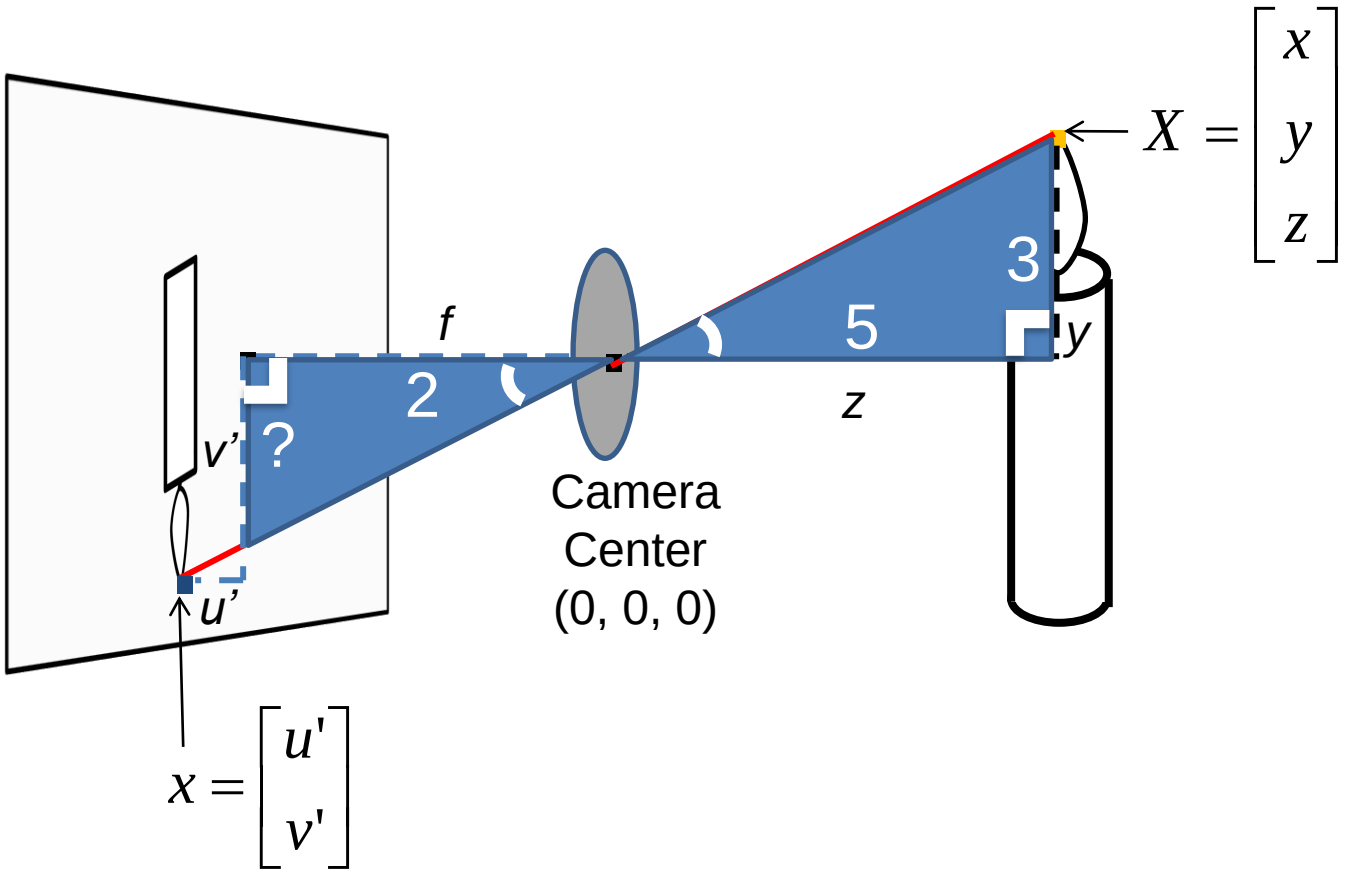


Vanishing points and lines



Slide from Efros, Photo from Criminisi

Projection: world coordinates→image coordinates



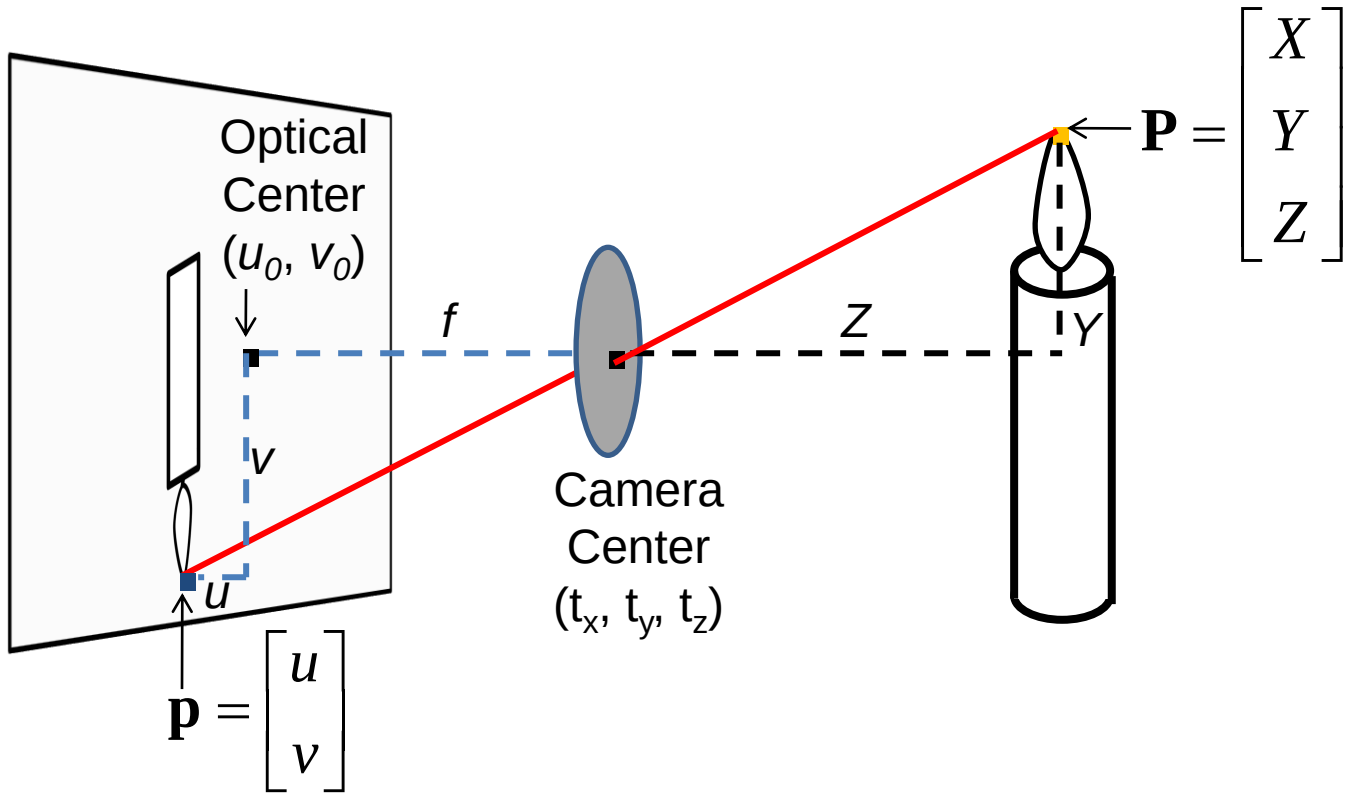
If X = 2, Y = 3,
Z = 5, and f = 2
What are U and V?

$$\frac{v'}{-f} = \frac{y}{z}$$

$$u' = -x * \frac{f}{z} \qquad u' = -2 * \frac{2}{5}$$

$$v' = -y * \frac{f}{z} \qquad v' = -3 * \frac{2}{5}$$

Projection: world coordinates→image coordinates



How do we handle the general case?