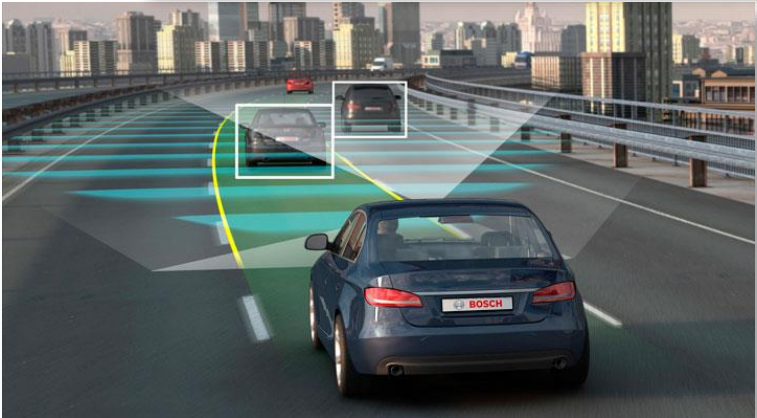
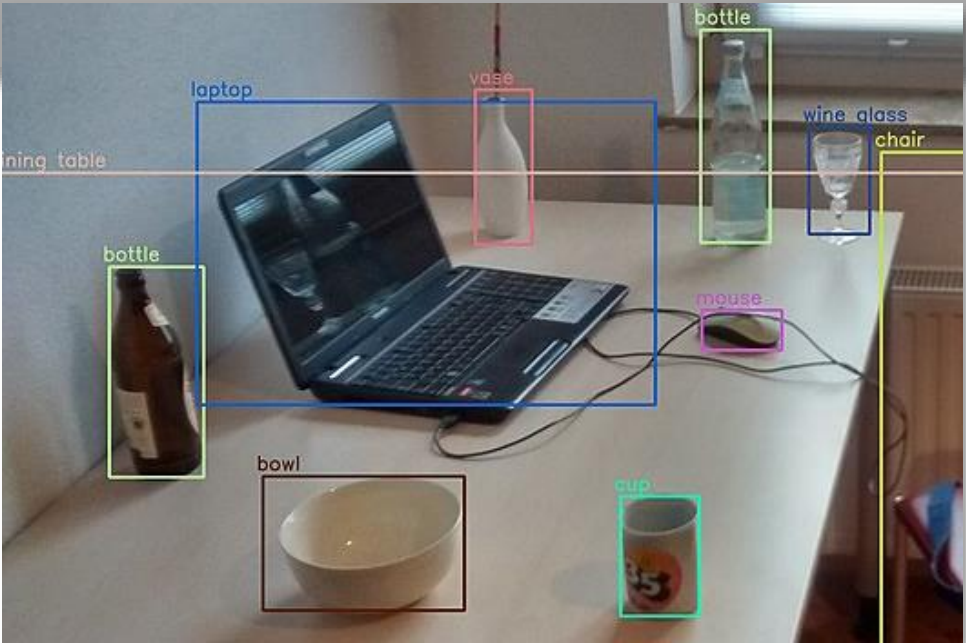


# CS5670: Intro to Computer Vision (Cornell Tech)

Instructor: Noah Snavely

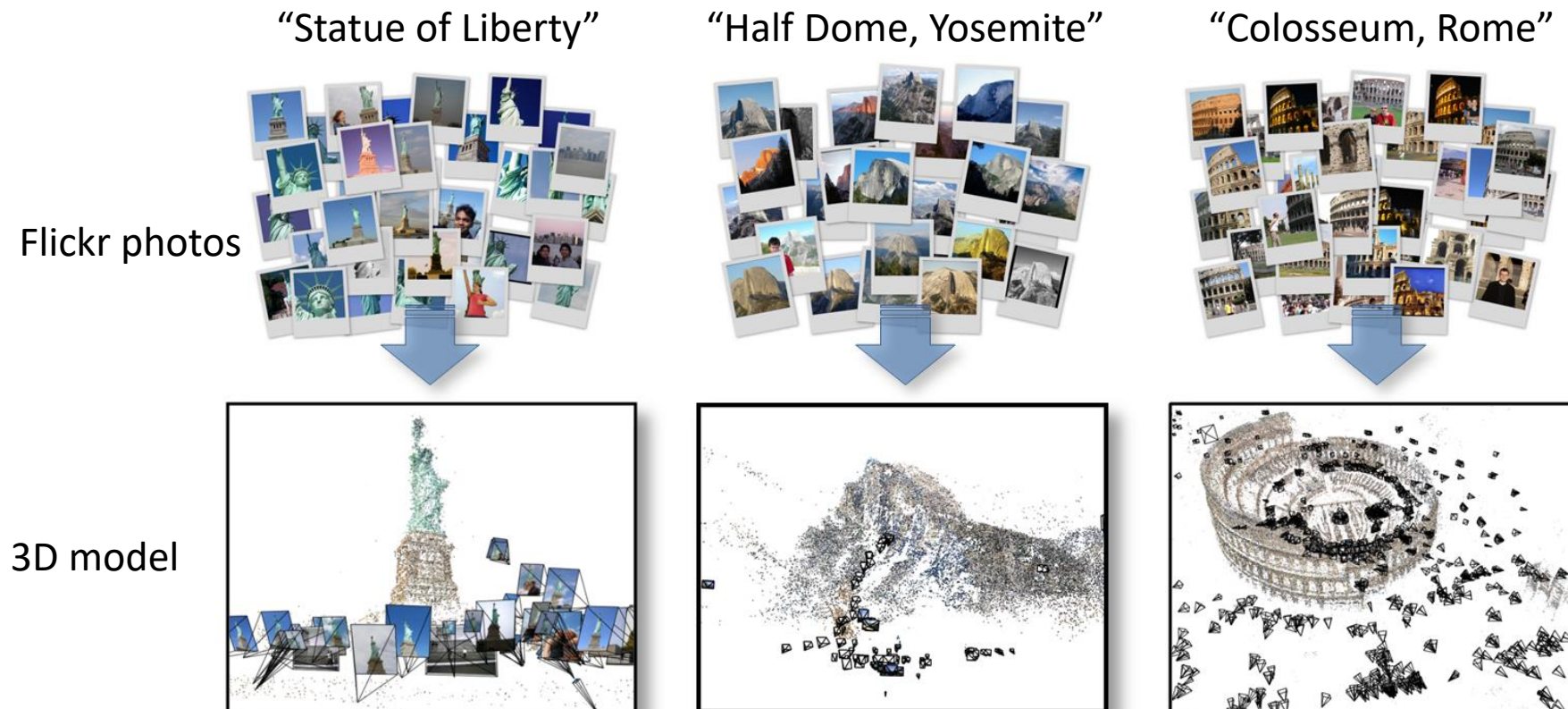


# Instructor

- Noah Snavely ([snavely@cs.cornell.edu](mailto:snavely@cs.cornell.edu))
- Research interests:
  - Computer vision and graphics
  - 3D reconstruction and visualization of Internet photo collections
  - Deep learning for computer graphics
  - Virtual and augmented reality

# Noah's work

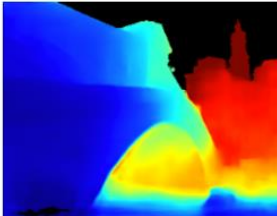
- Automatic 3D reconstruction from Internet photo collections



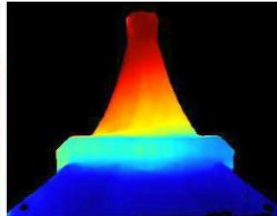
# City-scale 3D reconstruction

Reconstruction of Dubrovnik, Croatia, from ~40,000 images

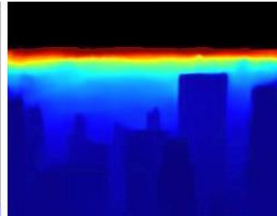
# Depth from a single image



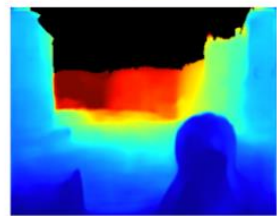
Rialto Bridge, Venice



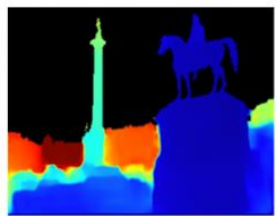
Eiffel Tower, Paris



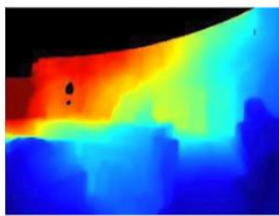
Central Park, NYC



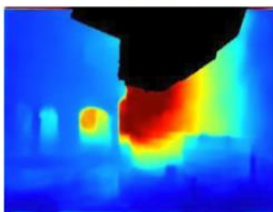
Grand Canal, Venice



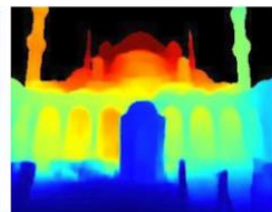
Trafalgar Square, London



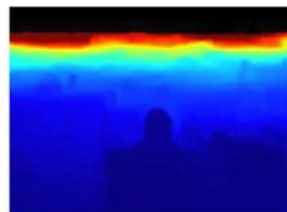
Colosseum, Rome



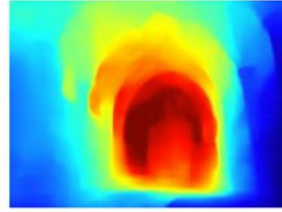
Venetian Hotel, Las Vegas



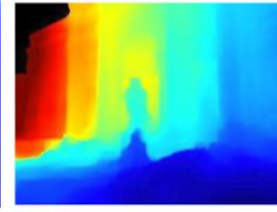
Sultan Ahmed Mosque, Mosque



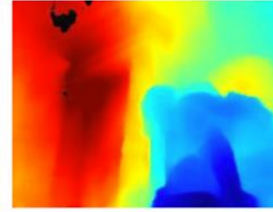
Seville Cathedral, Seville



Notre-Dame Basilica, Montreal



Trevi Fountain, Rome



Medici Fountain, Paris

# Visualizing scenes from tourist photos



# Teaching assistants



Zekun Hao

[hz472@cornell.edu](mailto:hz472@cornell.edu)



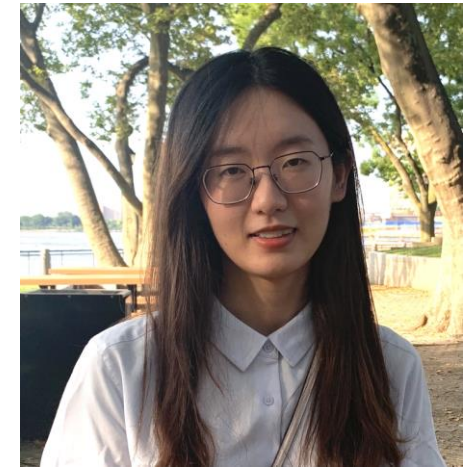
Wenqi Xian

[wx97@cornell.edu](mailto:wx97@cornell.edu)



Ruojin Cai

[rc844@cornell.edu](mailto:rc844@cornell.edu)



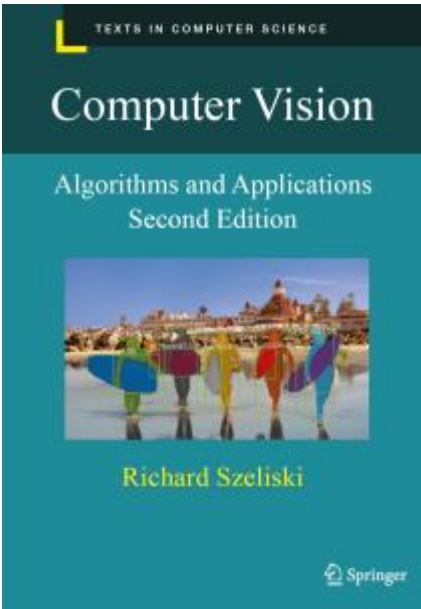
Qianqian Wang

[qw246@cornell.edu](mailto:qw246@cornell.edu)

- Please check back on course webpage for office hours

<https://www.cs.cornell.edu/courses/cs5670/2022sp/>

# Important information



- Textbook:  
Rick Szeliski, *Computer Vision: Algorithms and Applications* online at:  
<http://szeliski.org/Book/>
- Course webpage:  
<http://www.cs.cornell.edu/courses/cs5670/2022sp/>
- Canvas Page:  
<https://canvas.cornell.edu/courses/37687>
- Announcements/discussion via Ed Discussions (via Canvas)
- Assignment turnin via CMSX:  
<https://cmsx.cs.cornell.edu>



# Today

1. What is computer vision?
2. Why study computer vision?
3. Course overview
4. Images & image filtering [time permitting]

# Today

- Readings
  - Szeliski, Chapter 1 (Introduction)

# Every image tells a story



- Goal of computer vision: perceive the "story" behind the picture
- Compute properties of the world
  - 3D shape
  - Names of people or objects
  - What happened?

# The goal of computer vision



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

# Can computers match human perception?



- Yes and no (mainly no)
  - computers can be better at “easy” things
  - humans are better at “hard” things
- But huge progress
  - Accelerating in the last five years due to deep learning
  - What is considered “hard” keeps changing

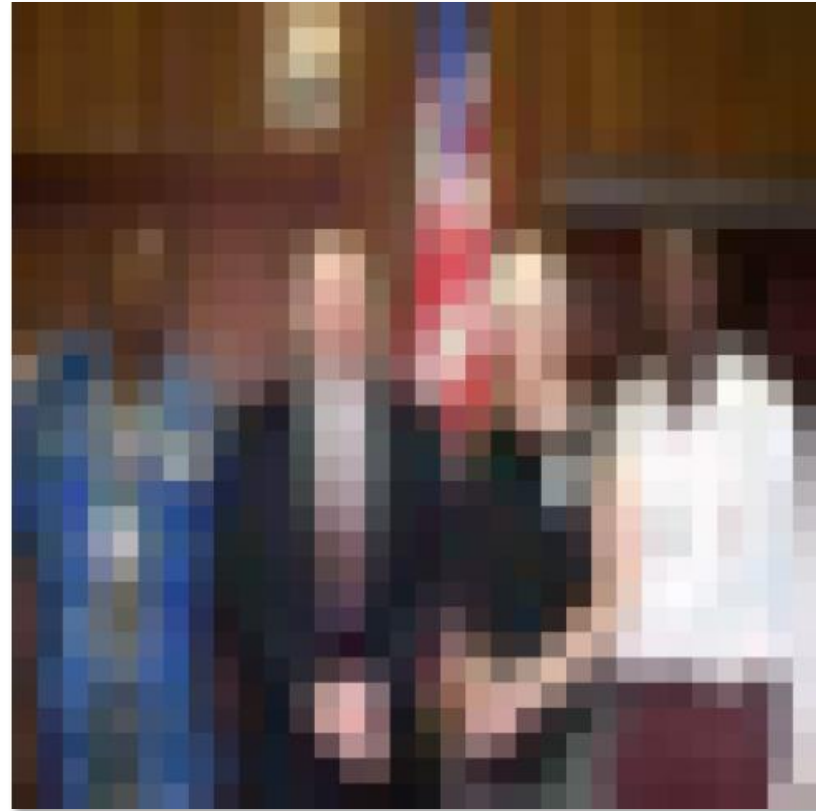
# Human perception has its shortcomings



[Sinha and Poggio, \*Nature\*, 1996](#)

("The Presidential Illusion")

**But humans can tell a lot about a scene from a little information...**



Source: "80 million tiny images" by Torralba, et al.





# The goal of computer vision

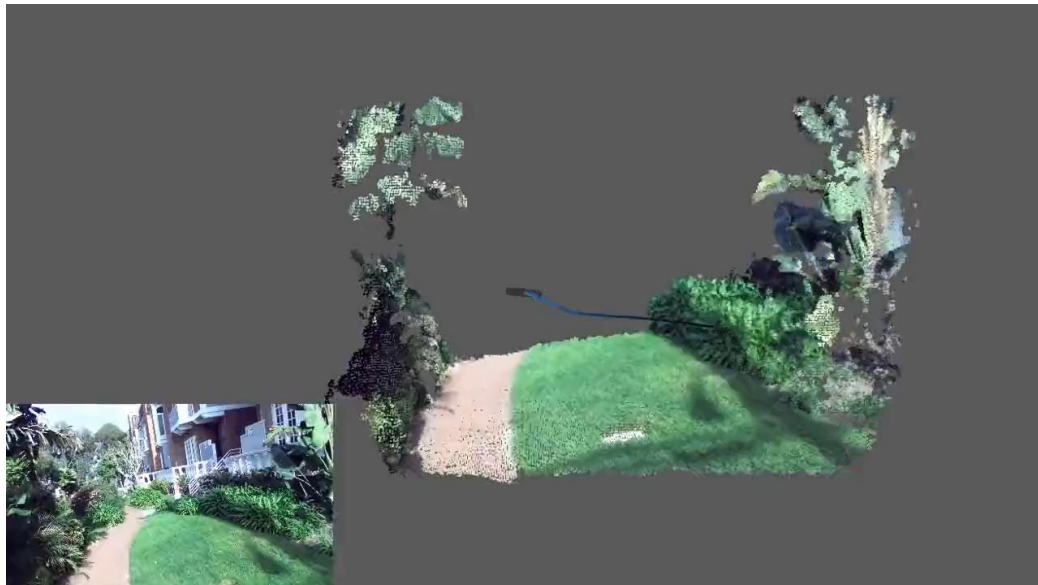
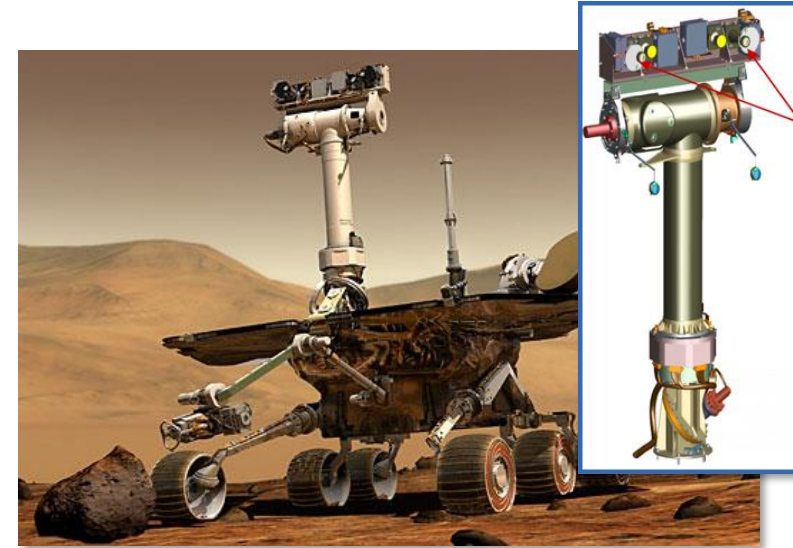


# The goal of computer vision

- Compute the 3D shape of the world

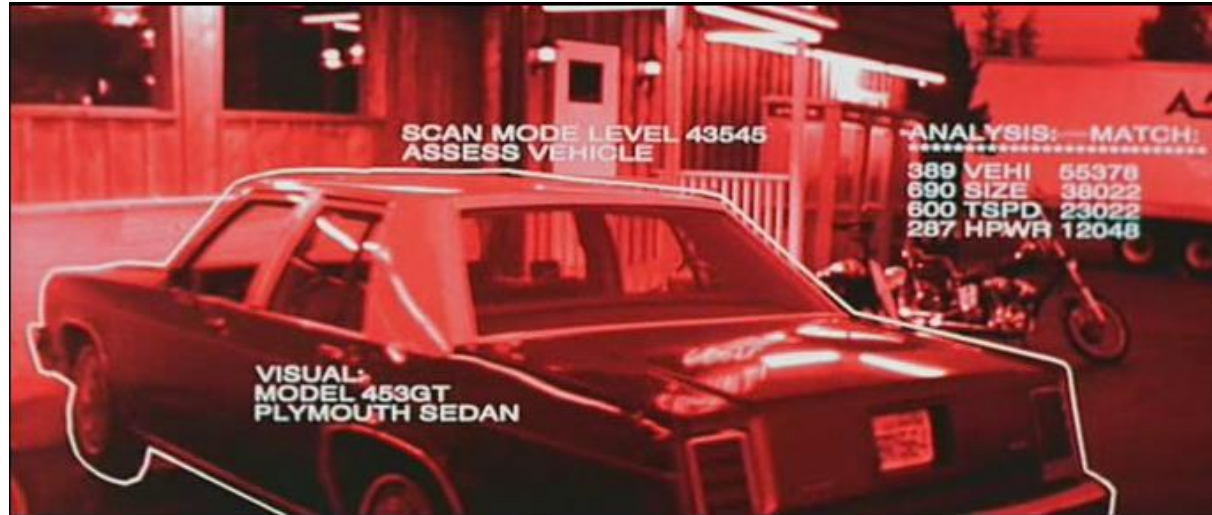


ZED 2i Camera



# The goal of computer vision

- Recognize objects and people



*Terminator 2, 1991*





sky

building

flag

face

banner

wall

street lamp

bus

bus

cars

slide credit: Fei-Fei, Fergus & Torralba

# The goal of computer vision

- “Enhance” images





# The goal of computer vision

- Forensics



Source: Nayar and Nishino, "Eyes for Relighting"





Source: Nayar and Nishino, "Eyes for Relighting"



Source: Nayar and Nishino, "Eyes for Relighting"

# The goal of computer vision

- Improve photos (“Computational Photography”)



Super-resolution (source: 2d3)



Low-light photography  
(credit: [Hasinoff et al., SIGGRAPH ASIA 2016](#))



Depth of field on cell phone camera  
(source: [Google Research Blog](#))



Inpainting / image completion  
(image credit: Hays and Efros)

# Why study computer vision?

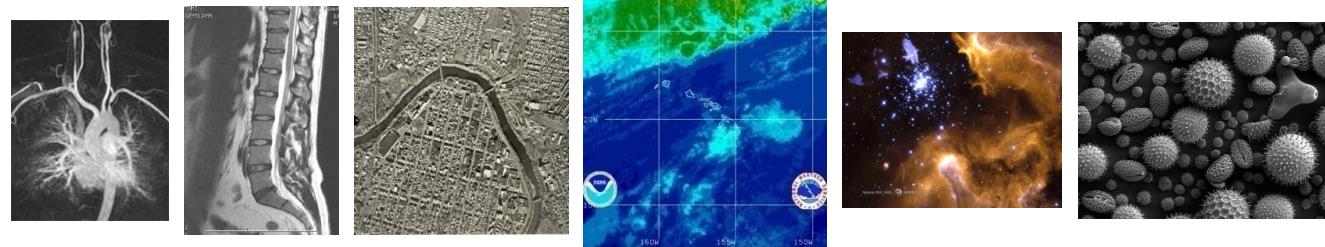
- Billions of images/videos captured per day



flickr



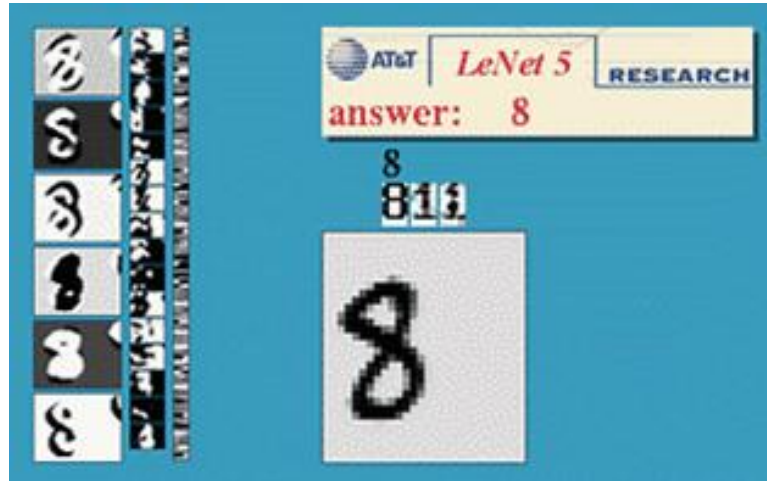
YouTube



- Huge number of potential applications
- The next slides show the current state of the art

# Optical character recognition (OCR)

- If you have a scanner, it probably came with OCR software



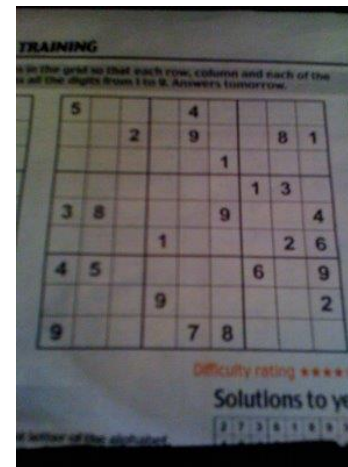
Digit recognition, AT&T labs (1990's)  
<http://yann.lecun.com/exdb/lenet/>



License plate readers  
[http://en.wikipedia.org/wiki/Automatic\\_number\\_plate\\_recognition](http://en.wikipedia.org/wiki/Automatic_number_plate_recognition)



Automatic check processing



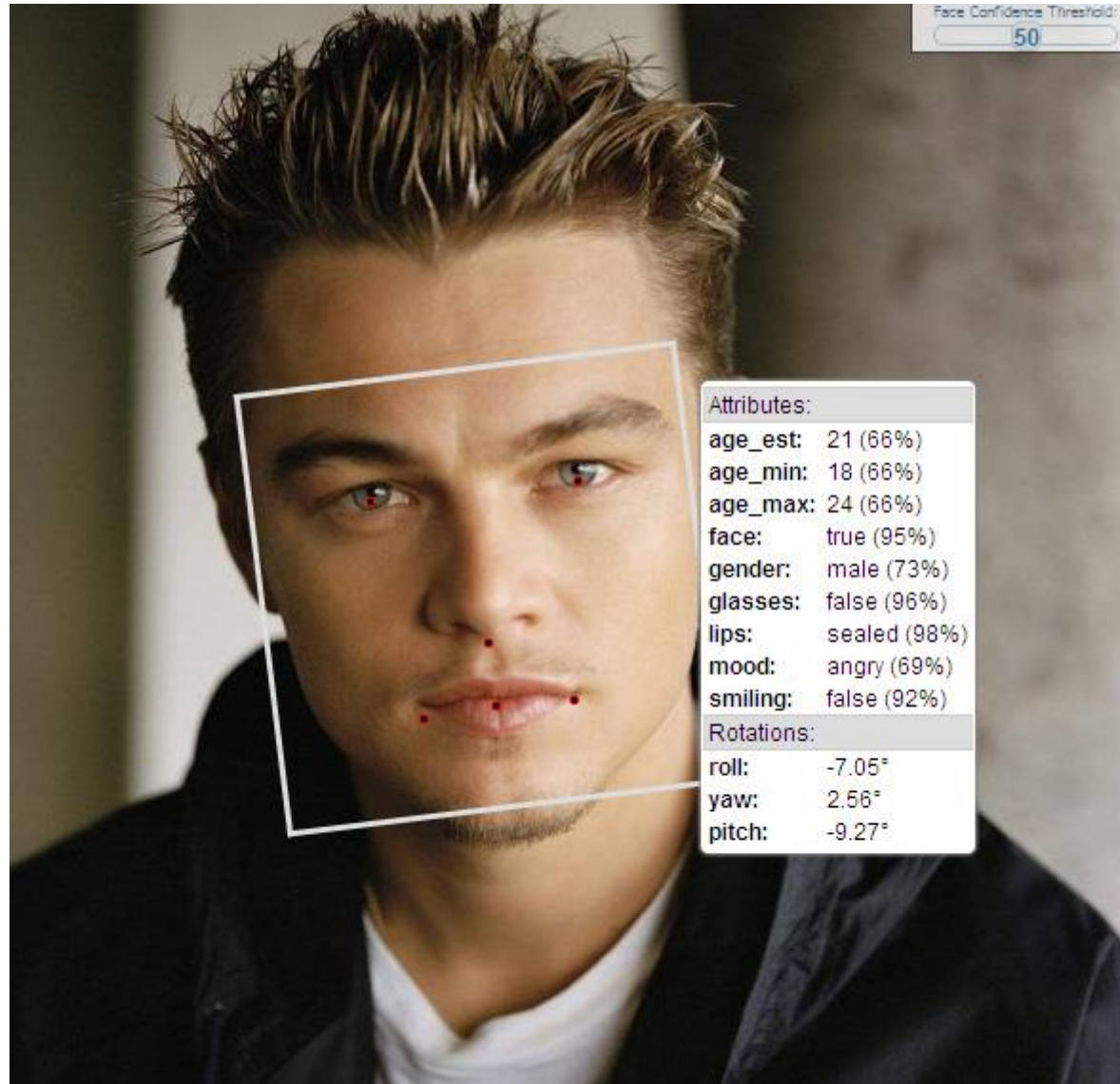
Sudoku grabber  
<http://sudokugrab.blogspot.com/>

# Face detection

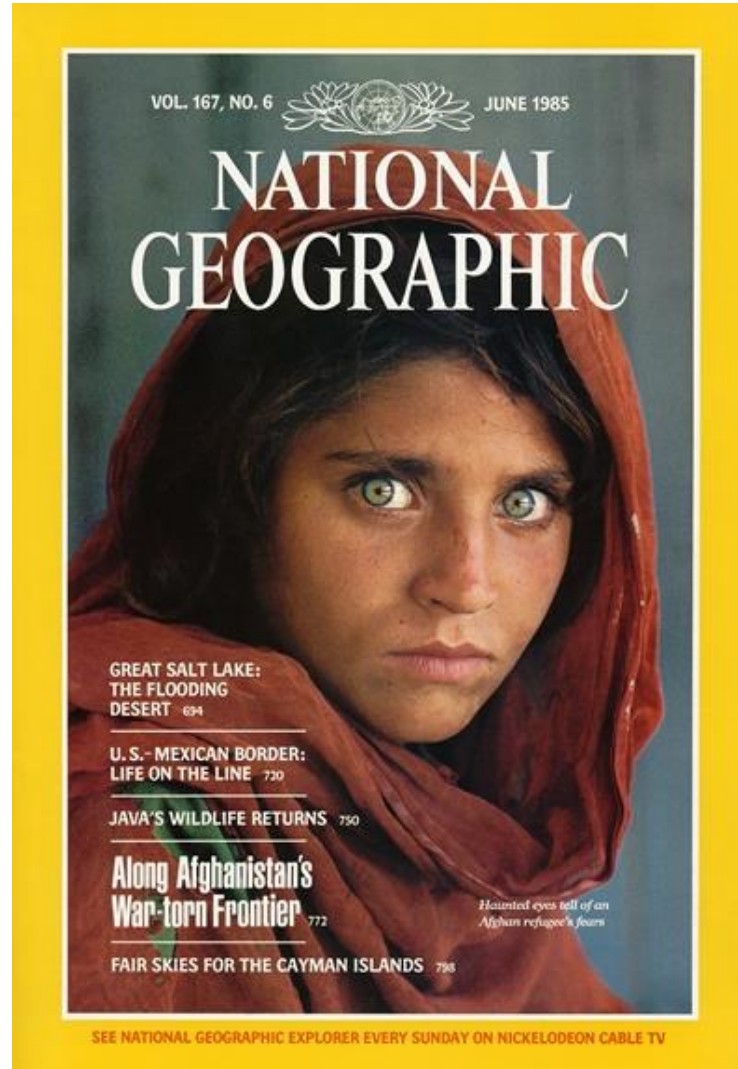


- Nearly all cameras detect faces in real time
  - (Why?)

# Face analysis and recognition



# Vision-based biometrics



Who is she?

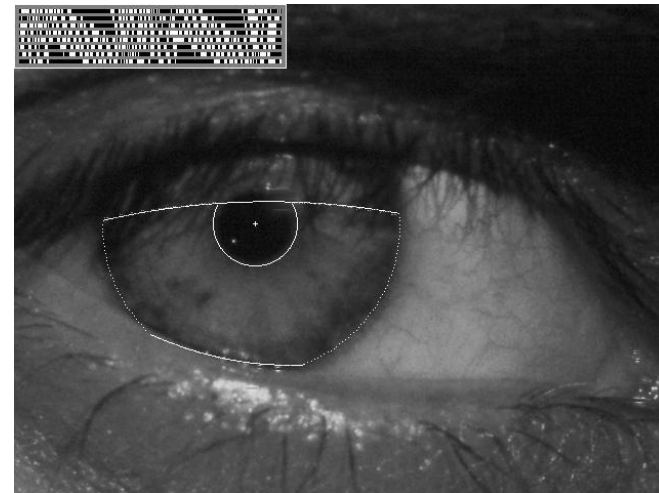
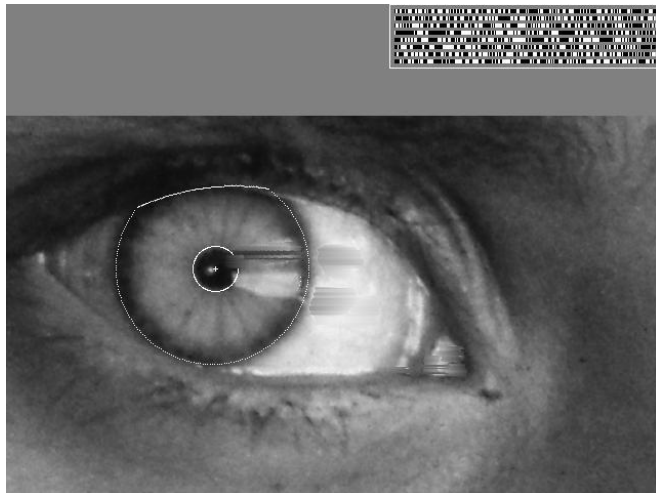
Source: S. Seitz



# Vision-based biometrics

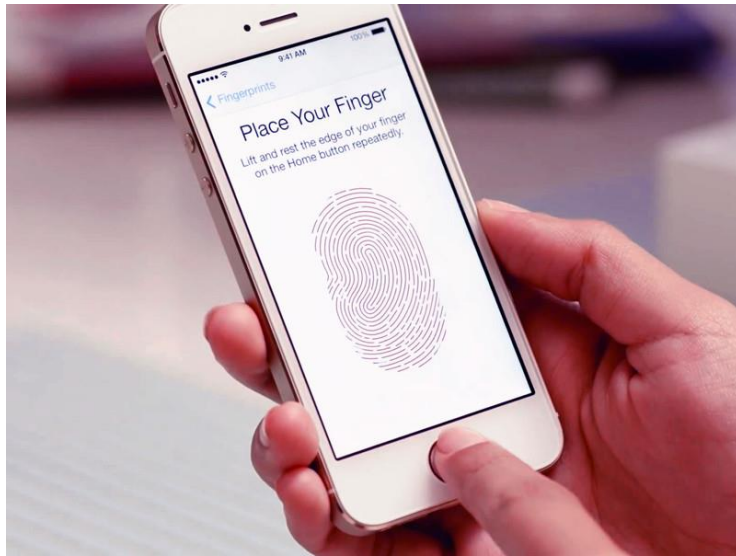


*"How the Afghan Girl was Identified by Her Iris Patterns"* Read the [story](#)



Source: S. Seitz

# Login without a password



Fingerprint scanners on many new smartphones and other devices

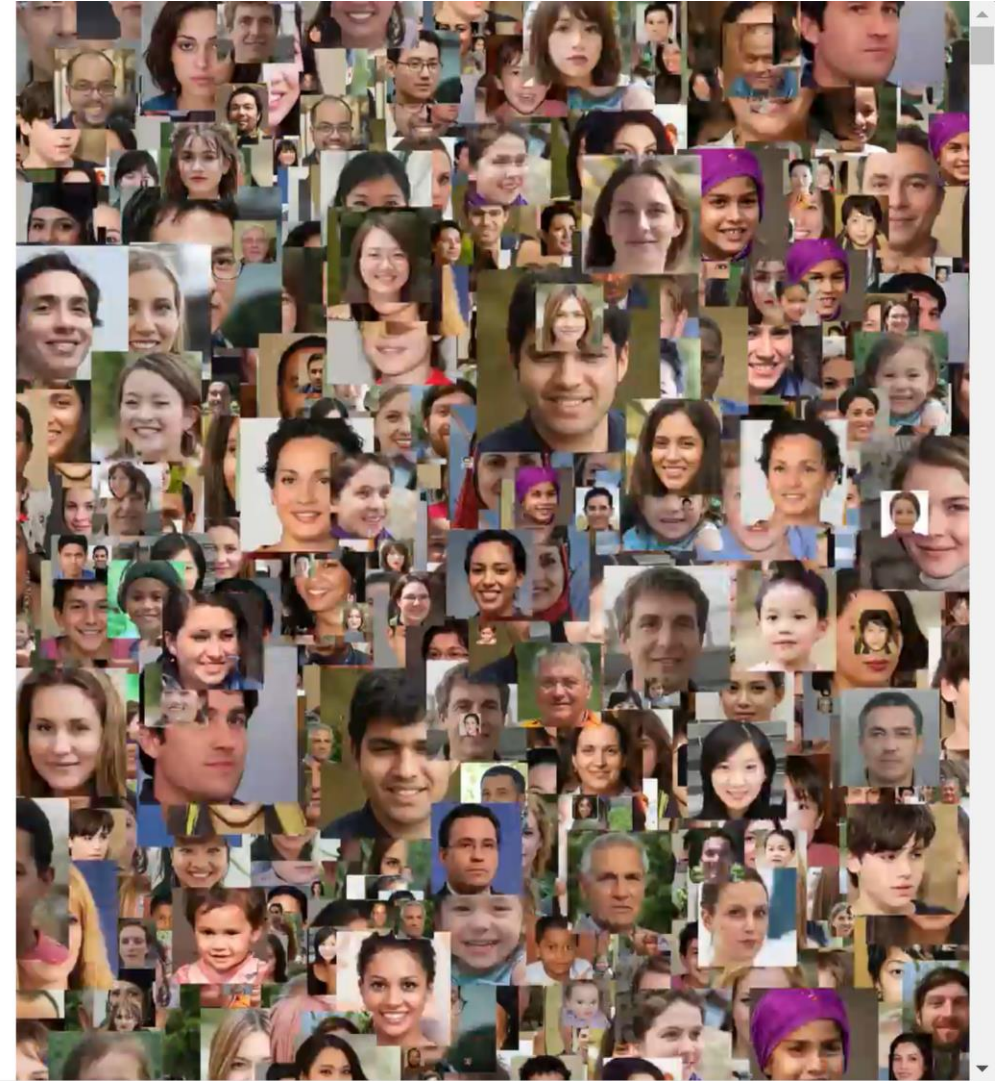


Face unlock on Apple iPhone X  
See also <http://www.sensiblevision.com/>



# The Secretive Company That Might End Privacy as We Know It

A little-known start-up helps law enforcement match photos of unknown people to their online images — and “might lead to a dystopian future or something,” a backer says.



New York Times, Jan. 18, 2020

by Kashmir Hill

---

# Researchers warn peace sign photos could expose fingerprints

But the likelihood of anyone actually using images to recreate prints is pretty slim.



Jamie Rigg, @jmerigg  
01.13.17 in Security

Comments

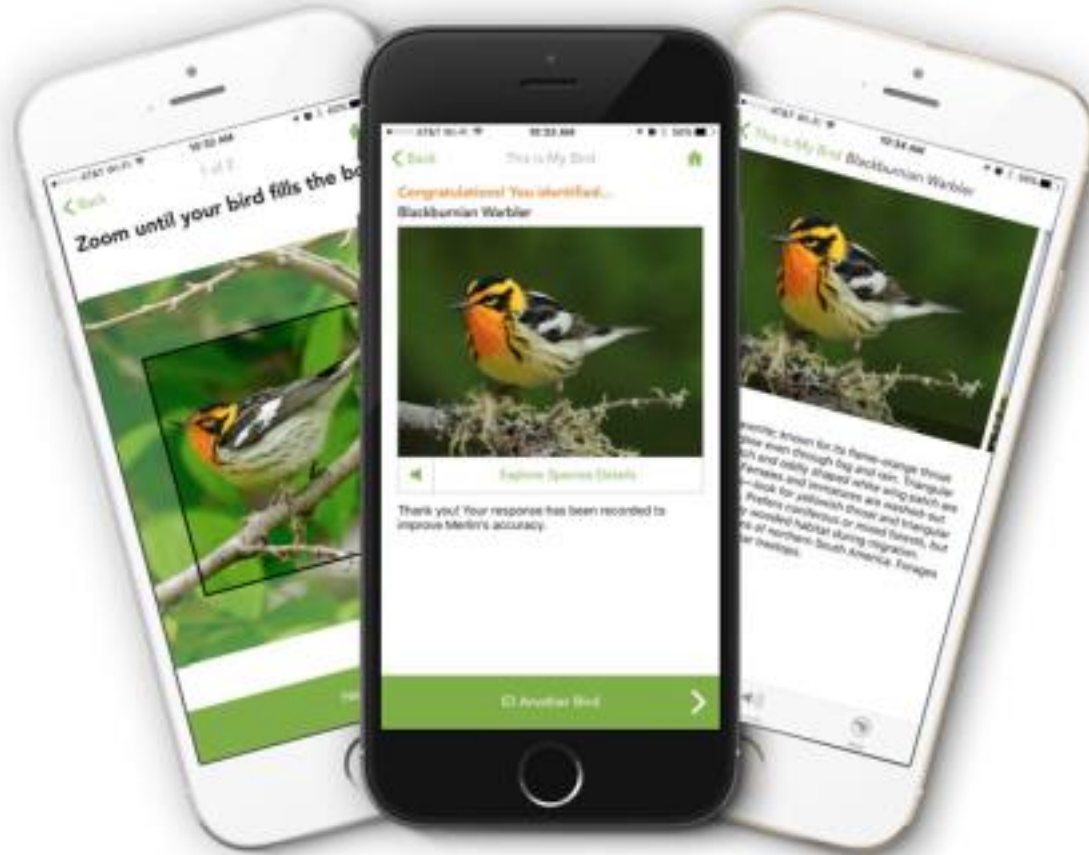
1721  
Shares



Getty



# Bird identification



Merlin Bird ID (based on Cornell Tech technology!)

# Special effects: shape capture



*The Matrix* movies, ESC Entertainment, XYZRGB, NRC

Source: S. Seitz

# Special effects: motion capture



*Pirates of the Caribbean*, Industrial Light and Magic

Source: S. Seitz

MOVIES



## Robert De Niro said no green screen. No face dots. How ‘The Irishman’s’ de-aging changes Hollywood



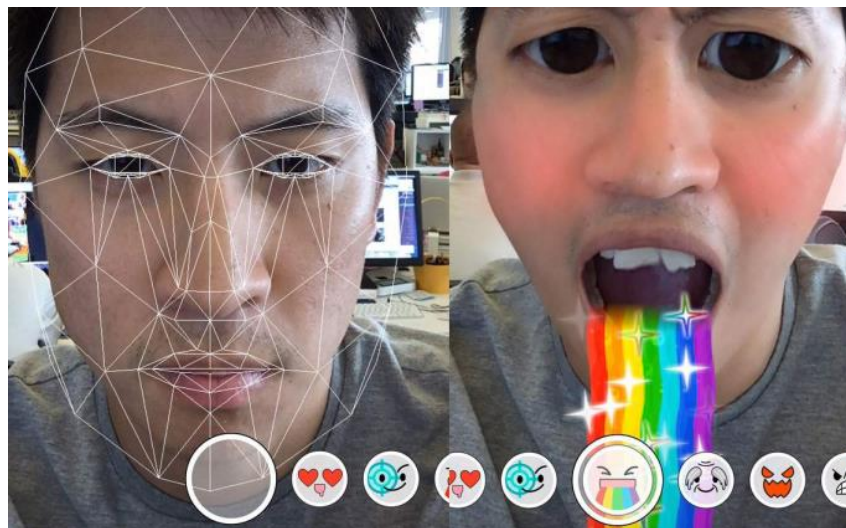
Makeup and wig work got Robert De Niro partway to his character, Frank Sheeran, at 41, left. It took a specially built camera and visual artists to get all the way there, as before-and-after images show. (Netflix)



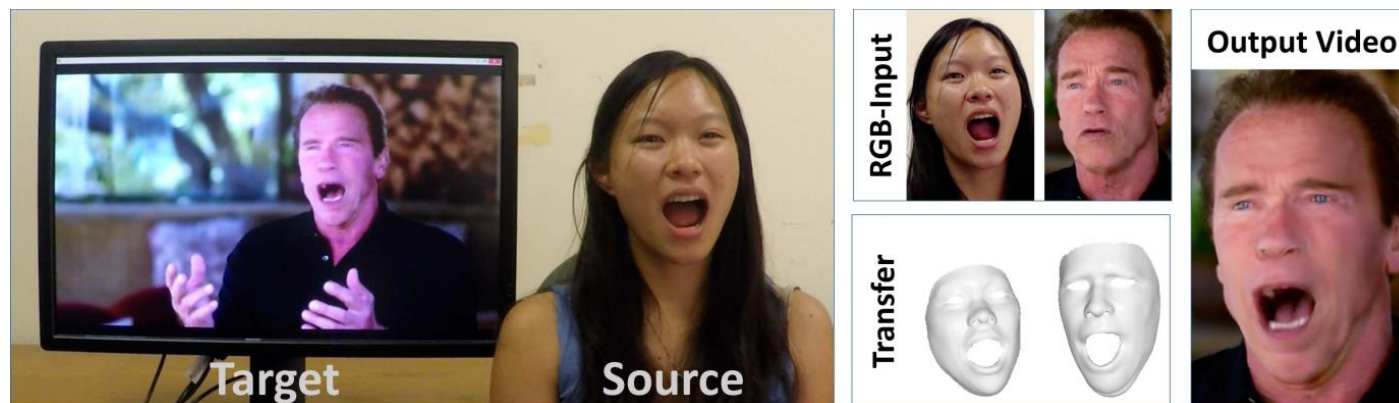
# Los Angeles Times



# 3D face tracking w/ consumer cameras



Snapchat Lenses



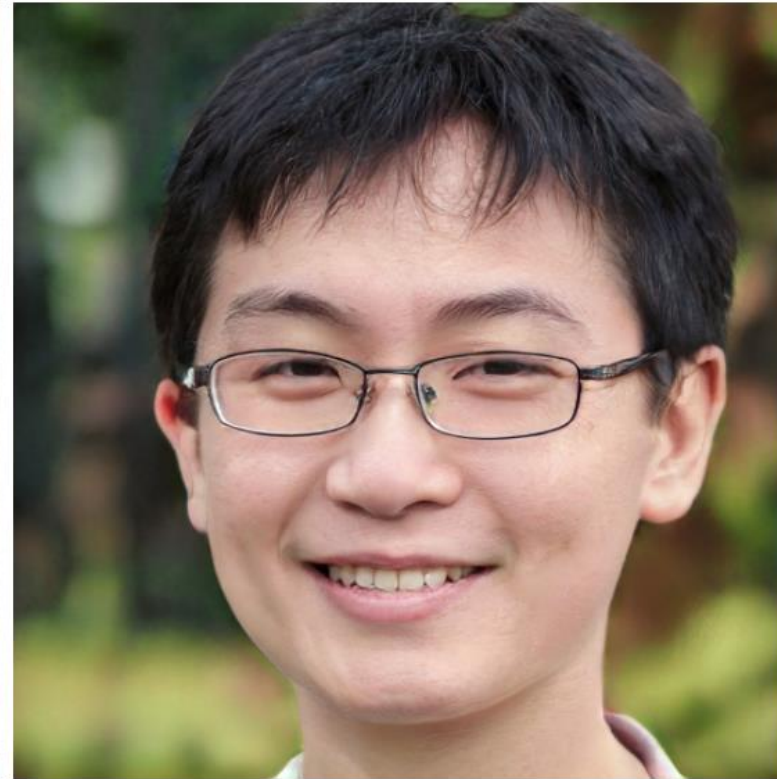
[Face2Face system](#) (Thies et al.)

# Image synthesis



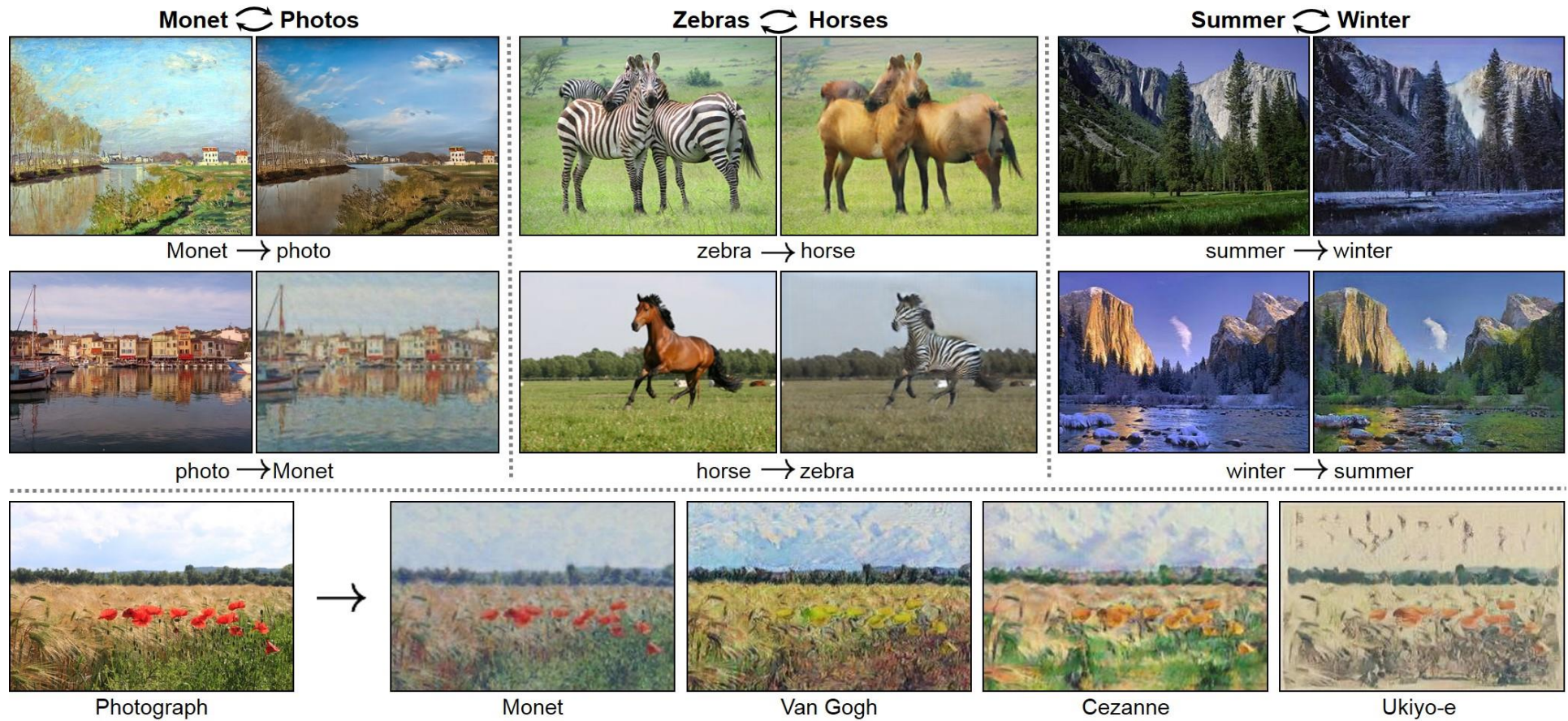
# Which face is real?

Click on the person who is real.



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

# Image synthesis



# Sports



*Sportvision* first down line

[Explanation](http://www.howstuffworks.com) on [www.howstuffworks.com](http://www.howstuffworks.com)

Day 5: Swimming - Men's 4X200M Final



Highlights of the men's 4x200m relay final on Day 5.

Source: S. Seitz

# Smart cars

manufacturer products | consumer products

## Our Vision. Your Safety.

rear looking camera | forward looking camera | side looking camera

↳ **EyeQ** Vision on a Chip

↳ **Vision Applications**  
Road, Vehicle, Pedestrian Protection and more

↳ **AWS** Advance Warning System

News

↳ Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System

↳ Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end

↳ all news

Events

↳ Mobileye at Equip Auto, Paris, France

↳ Mobileye at SEMA, Las Vegas, NV

↳ read more

- [Mobileye](#)
- Tesla Autopilot
- Safety features in many cars

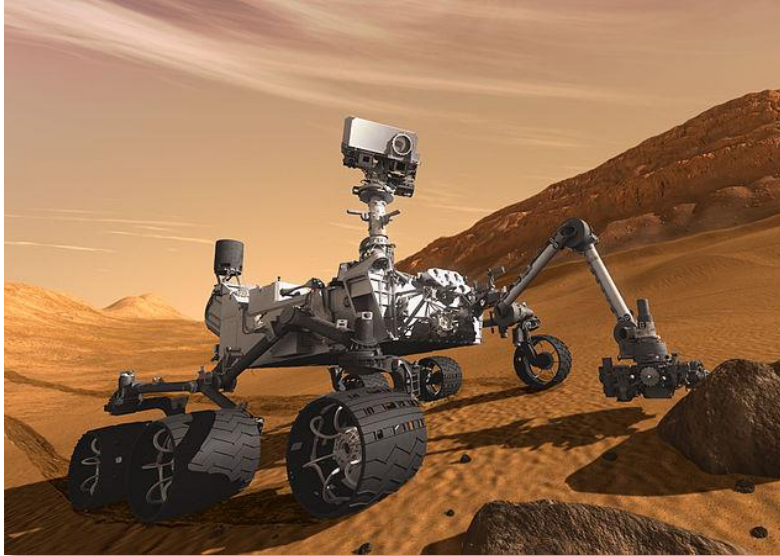
# Self-driving cars



Waymo



# Robotics



NASA's Mars Curiosity Rover  
[https://en.wikipedia.org/wiki/Curiosity\\_\(rover\)](https://en.wikipedia.org/wiki/Curiosity_(rover))



Amazon Picking Challenge  
<http://www.robocup2016.org/en/events/amazon-picking-challenge/>

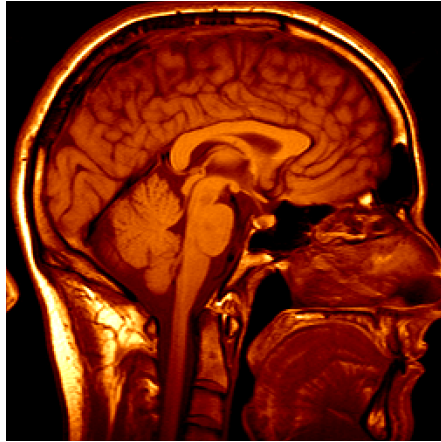


Amazon Prime Air

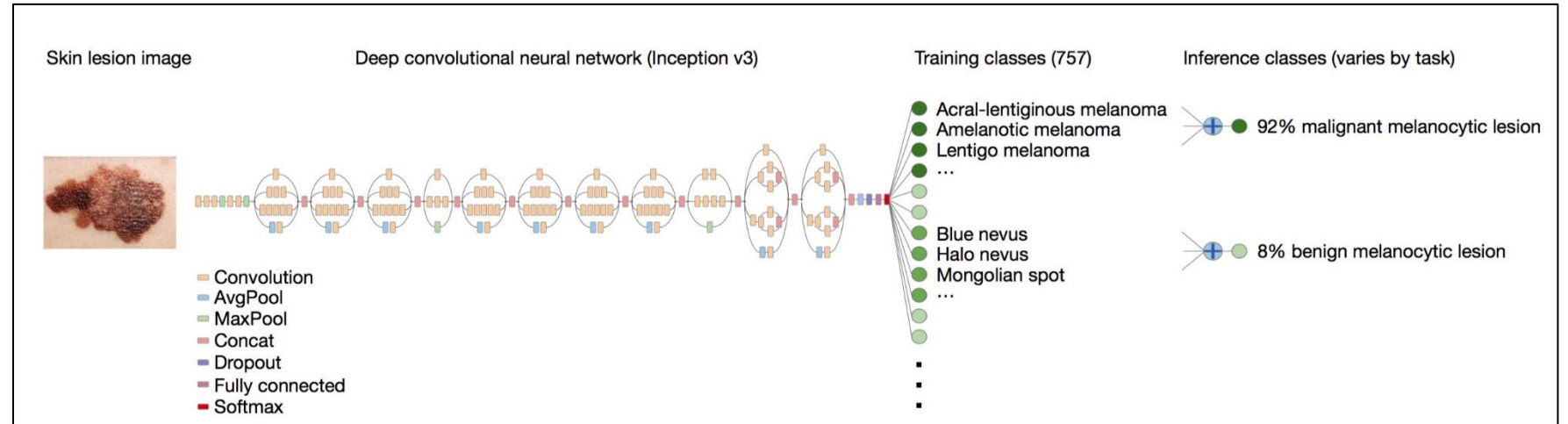


Amazon Scout

# Medical imaging



3D imaging  
(MRI, CT)



Skin cancer classification with deep learning  
<https://cs.stanford.edu/people/esteva/nature/>

INVESTING

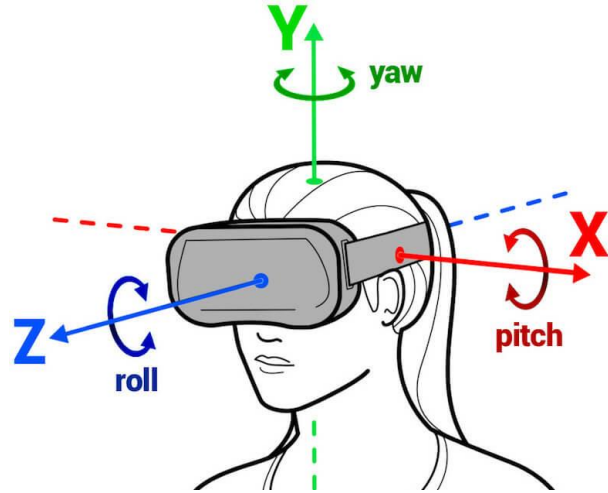
3/25/2014 @ 5:43PM | 70,399 views

# Facebook Buys Oculus, Virtual Reality Gaming Startup, For \$2 Billion

[+ Comment Now](#) [+ Follow Comments](#)



# Virtual & Augmented Reality



6DoF head tracking



Hand & body tracking



3D scene understanding



3D-360 video capture

# Current state of the art

- You just saw many examples of current systems.
  - Many of these are less than 5 years old
- Computer vision is an active research area, and rapidly changing
  - Many new apps in the next 5 years
  - Deep learning powering many modern applications
- Many startups across a dizzying array of areas
  - Deep learning, robotics, autonomous vehicles, medical imaging, construction, inspection, VR/AR, ...

# Why is computer vision difficult?



Viewpoint variation



Illumination



Credit: Flickr user michaelpaul

Scale

# Why is computer vision difficult?



Intra-class variation



Motion (Source: S. Lazebnik)

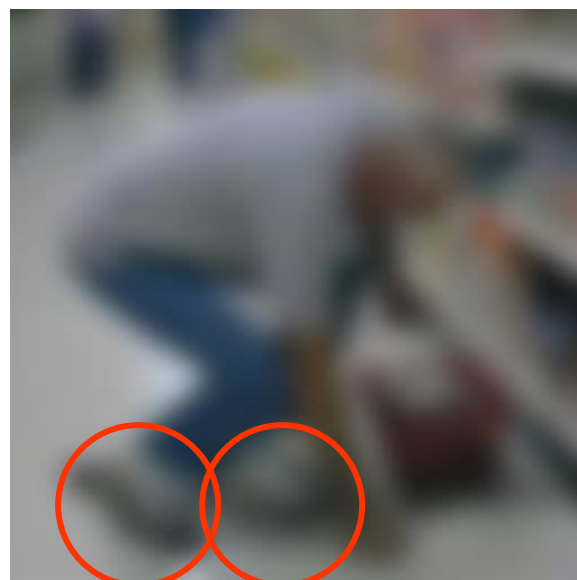
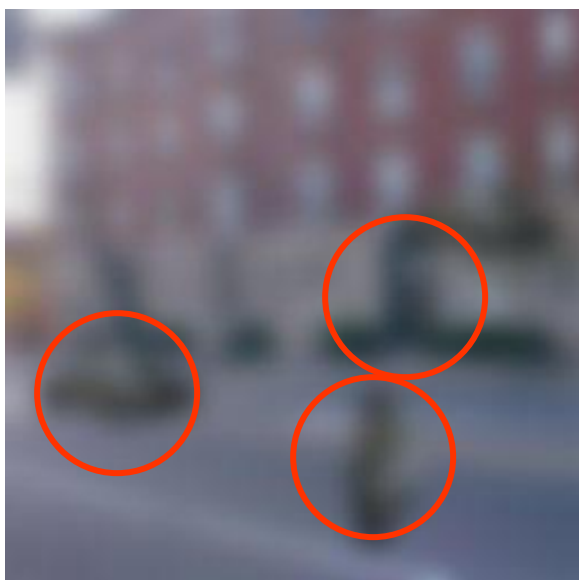
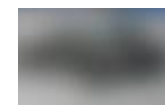
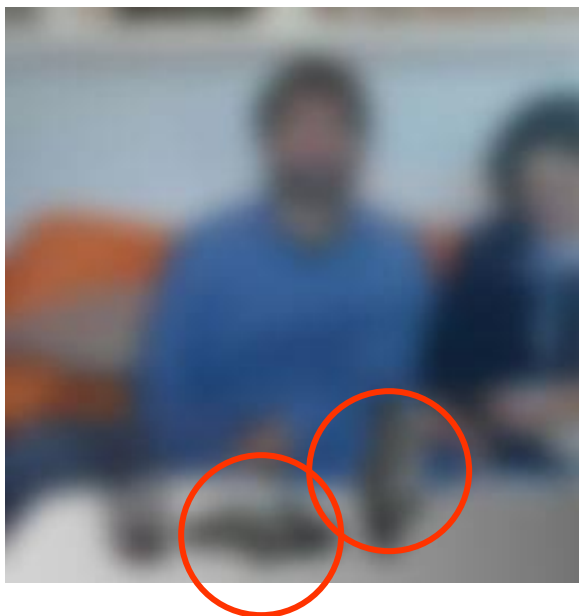


Background clutter



Occlusion

# Challenges: local ambiguity





# But there are lots of visual cues we can use...



Source: S. Lazebnik

# Bottom line

- Perception is an inherently ambiguous problem
  - Many different 3D scenes could have given rise to a given 2D image



Artist Julian Beever with his anamorphic Coke bottle

- We often must use prior knowledge about the world's structure



The picture above is funny.

But for me it is also one of those examples that make me sad about the outlook for AI and for Computer Vision. What would it take for a computer to understand this image as you or I do? I challenge you to think explicitly of all the pieces of knowledge that have to fall in place for it to make sense. Here is my short attempt:

- You recognize it is an image of a bunch of people and you understand they are in a hallway
- You recognize that there are 3 mirrors in the scene so some of those people are "fake" replicas from different viewpoints.
- You recognize Obama from the few pixels that make up his face. It helps that he is in his suit and that he is surrounded by other people with suits.
- You recognize that there's a person standing on a scale, even though the scale occupies only very few white pixels that blend with the background. But, you've used the person's pose and knowledge of how people interact with objects to figure it out.
- You recognize that Obama has his foot positioned just slightly on top of the scale. Notice the language I'm using: It is in terms of the 3D structure of the scene, not the position of the leg in the 2D coordinate system of the image.
- You know how physics works: Obama is leaning in on the scale, which applies a force on it. Scale measures force that is applied on it, that's how it works => it will over-estimate the weight of the person standing on it.
- The person measuring his weight is not aware of Obama doing this. You derive this because you know his pose, you understand that the field of view of a person is finite, and you understand that he is not very likely to sense the slight push of Obama's foot.
- You understand that people are self-conscious about their weight. You also understand that he is reading off the scale measurement, and that shortly the over-estimated weight will confuse him because it will probably be much higher than what he expects. In other words, you reason about implications of the events that are about to unfold seconds after this photo was taken, and especially about the thoughts and how they will develop inside people's heads. You also reason about what pieces of information are available to people.
- There are people in the back who find the person's imminent confusion funny. In other words you are reasoning about state of mind of people, and their view of the state of mind of another person. That's getting frighteningly meta.
- Finally, the fact that the perpetrator here is the president makes it maybe even a little more funnier. You understand what actions are more or less likely to be undertaken by different people based on their status and identity.

# The state of Computer Vision and AI: we are really, really far.

Oct 22, 2012



The picture above is funny.

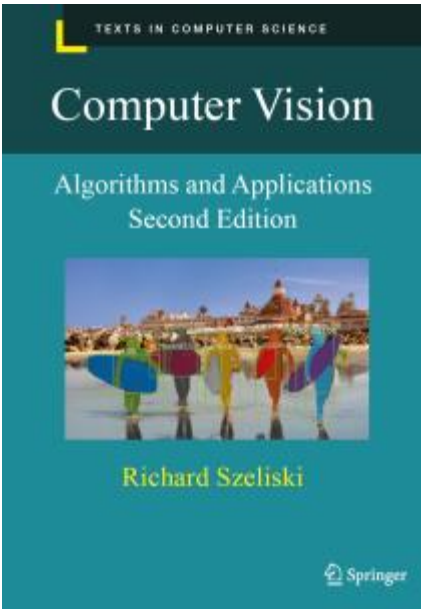
But for me it is also one of those examples that make me sad about the outlook for AI and for Computer Vision. What would it take for a computer to understand this image as you or I do? I challenge you to think explicitly of all the pieces of knowledge that have to fall in place for it to make sense. Here is my short attempt:

- You recognize it is an image of a bunch of people and you understand they are in a hallway
- You recognize that there are 3 mirrors in the scene so some of those people are "fake" replicas from different viewpoints.
- You recognize Obama from the few pixels that make up his face. It helps that he is in his suit and that he is surrounded by other people with suits.
- You recognize that there's a person standing on a scale, even though the scale occupies only very few white pixels that blend with the background. But, you've used the person's pose and knowledge of how people interact with objects to figure it out.
- You recognize that Obama has his foot positioned just slightly on top of the scale. Notice the language I'm using: It is in terms of the 3D structure of the scene, not the position of the leg in the 2D coordinate system of the image.
- You know how physics works: Obama is leaning in on the scale, which applies a force on it. Scale measures force that is applied on it, that's how it works => it will over-estimate the weight of the person standing on it.
- The person measuring his weight is not aware of Obama doing this. You derive this because you know his pose, you understand that the field of view of a person is finite, and you understand that he is not very likely to sense the slight push of Obama's foot.
- You understand that people are self-conscious about their weight. You also understand that he is reading off the scale measurement, and that shortly the over-estimated weight will confuse him because it will probably be much higher than what he expects. In other words, you reason about implications of the events that are about to unfold seconds after this photo was taken, and especially about the thoughts and how they will develop inside people's heads. You also reason about what pieces of information are available to people.
- There are people in the back who find the person's imminent confusion funny. In other words you are reasoning about state of mind of people, and their view of the state of mind of another person. That's getting frighteningly meta.
- Finally, the fact that the perpetrator here is the president makes it maybe even a little more funnier. You understand what actions are more or less likely to be undertaken by different people based on their status and identity.

# CS5670: Introduction to Computer Vision

- Project-based course whose goal is to teach you the basics of computer vision – image processing, geometry, recognition – in a hands-on way

# Important information



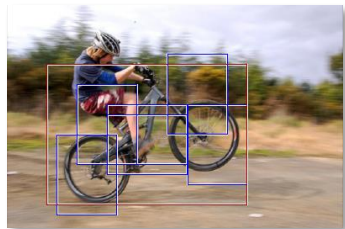
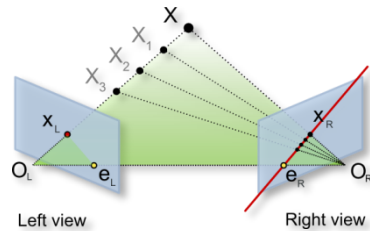
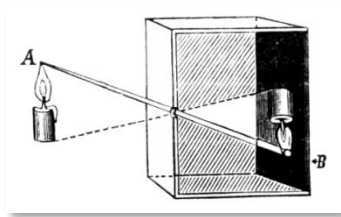
- Textbook:  
Rick Szeliski, *Computer Vision: Algorithms and Applications* online at:  
<http://szeliski.org/Book/>
- Course webpage:  
<http://www.cs.cornell.edu/courses/cs5670/2022sp/>
- Canvas Page:  
<https://canvas.cornell.edu/courses/37687>
- Announcements/discussion via Ed Discussions (via Canvas)
- Assignment turnin via CMSX:  
<https://cmsx.cs.cornell.edu>

# Course requirements

- Prerequisites
  - Data structures
  - Good working knowledge of Python programming
  - Linear algebra
  - Vector calculus
- Course does ***not*** assume prior imaging experience
  - computer vision, image processing, graphics, etc.



# Course overview (tentative)



## 1. Low-level vision

- image processing, edge detection, feature detection, cameras, image formation

## 2. Geometry and algorithms

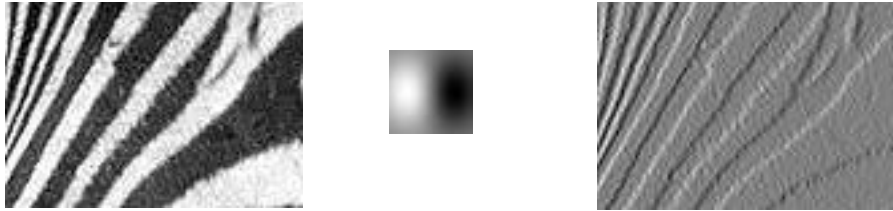
- projective geometry, stereo, structure from motion, optimization

## 3. Recognition

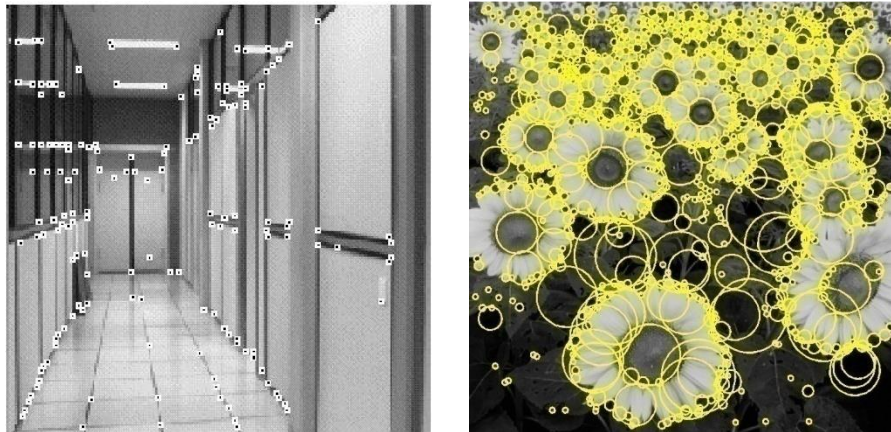
- face detection / recognition, category recognition, segmentation

# 1. Low-level vision

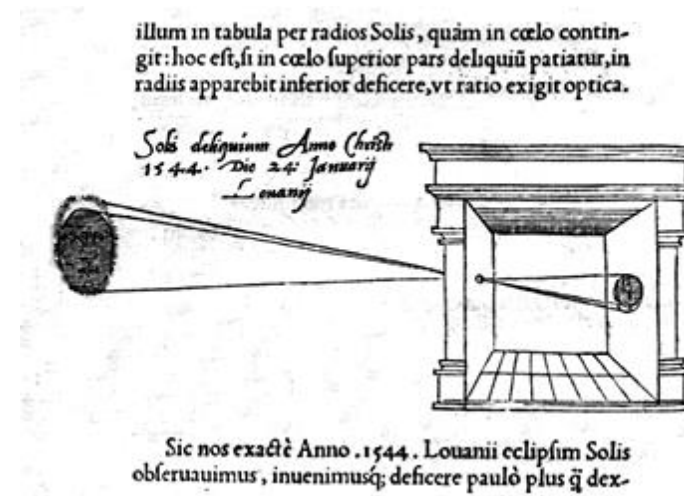
- Basic image processing and image formation



Filtering, edge detection



Feature extraction



Sic nos exactè Anno .1544. Louanii eclipsim Solis obseruauimus, inuenimusq; deficere paulò plus q̄ dex-

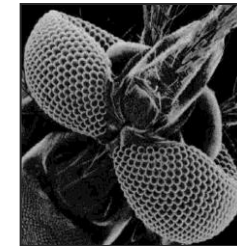
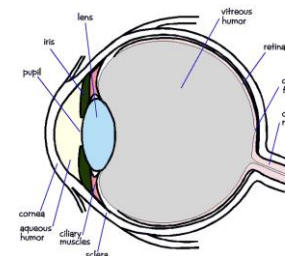
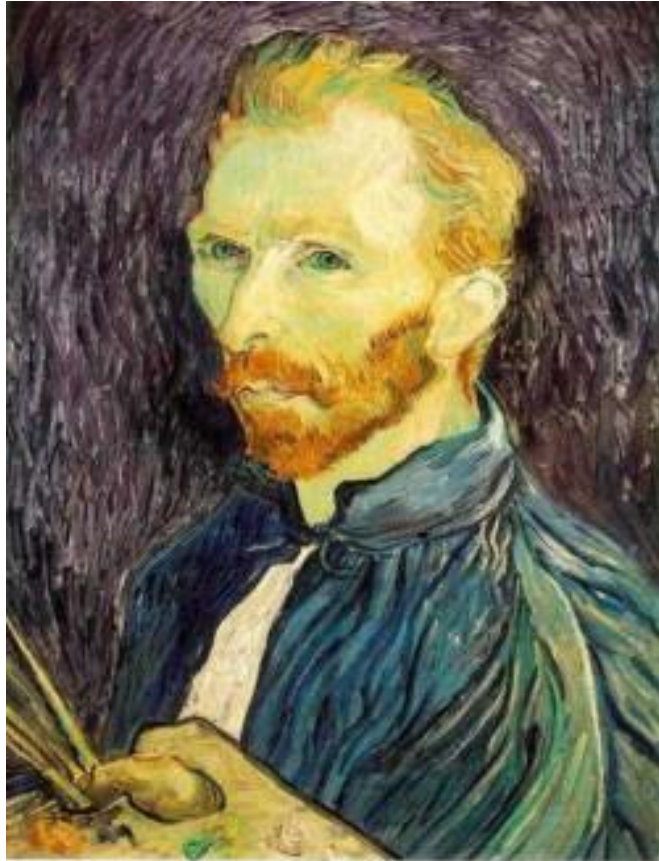


Image formation

# Project: Hybrid images from image pyramids



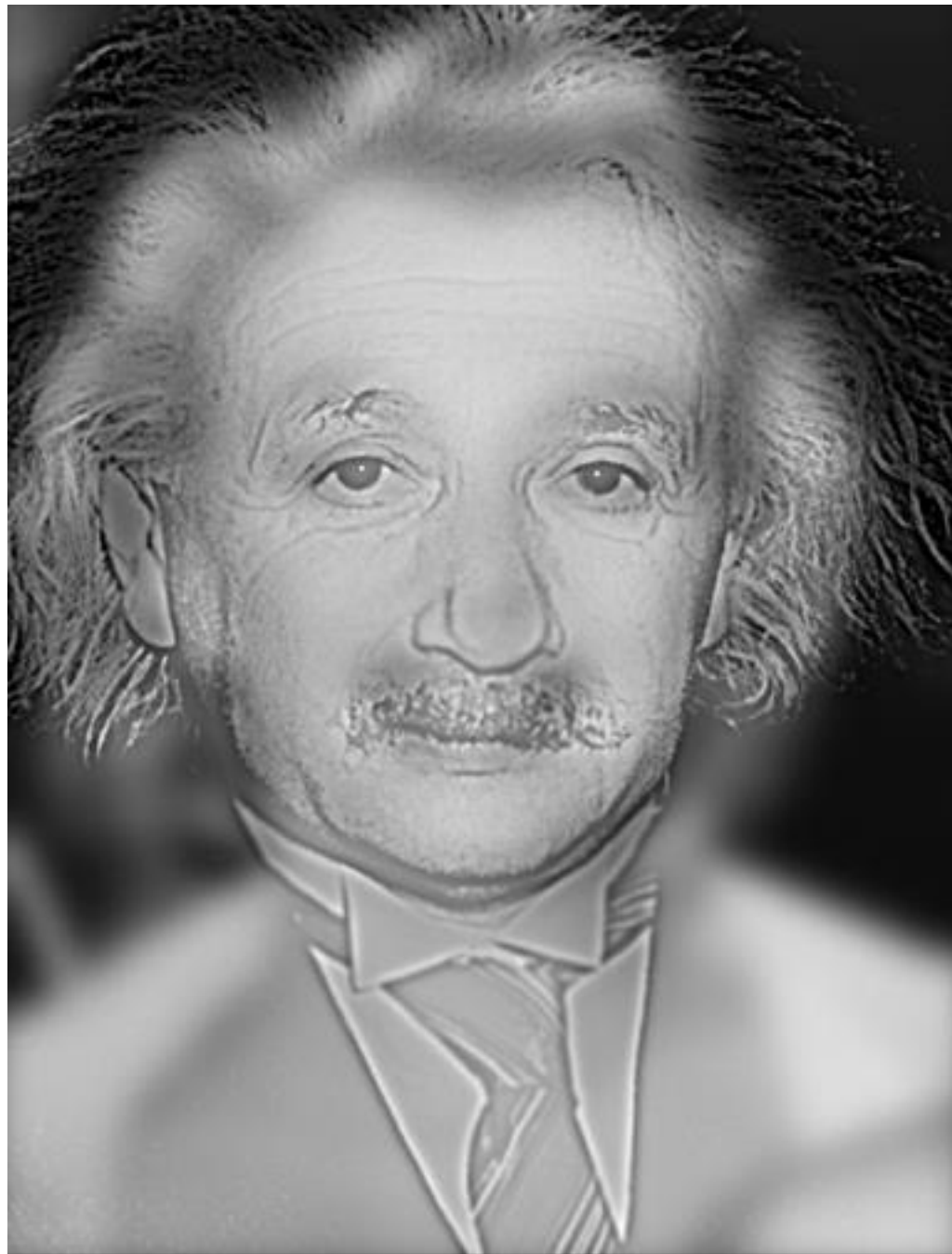
Gaussian 1/2



G 1/4

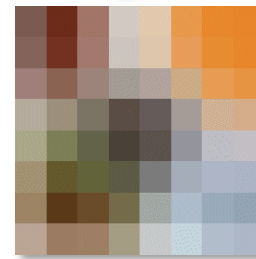
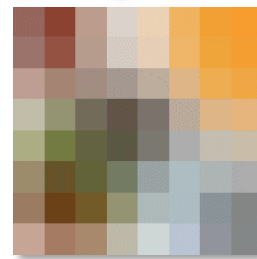
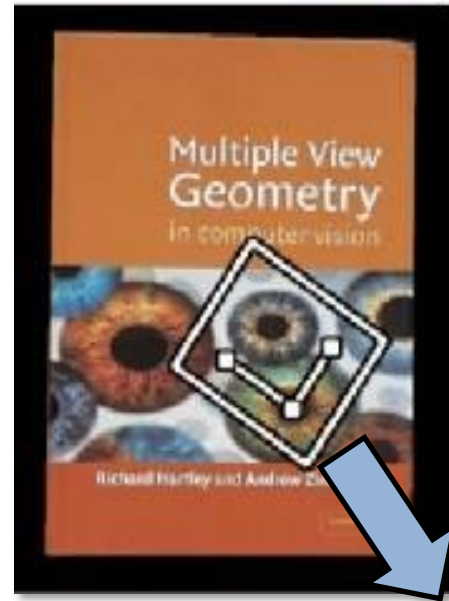


G 1/8

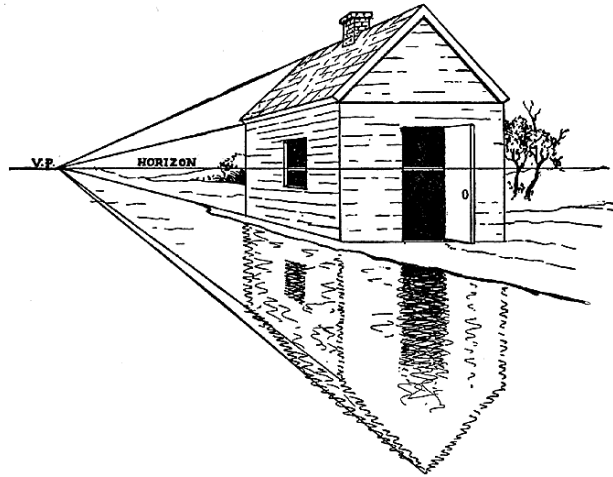




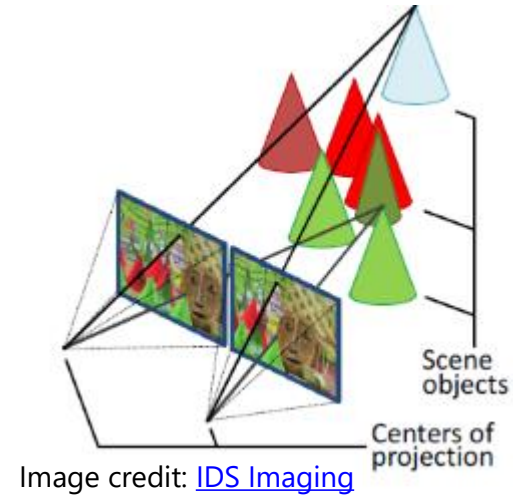
# Project: Feature detection and matching



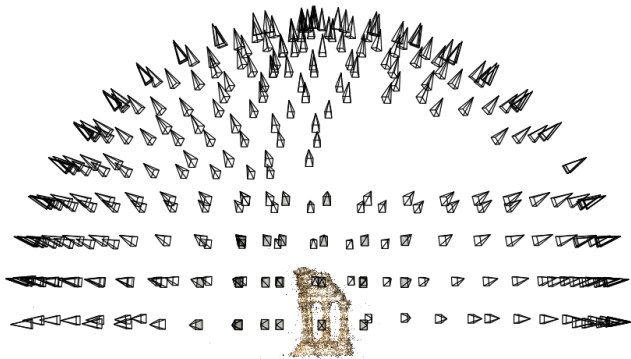
# 2. Geometry



Projective geometry



Stereo vision



Multi-view stereo



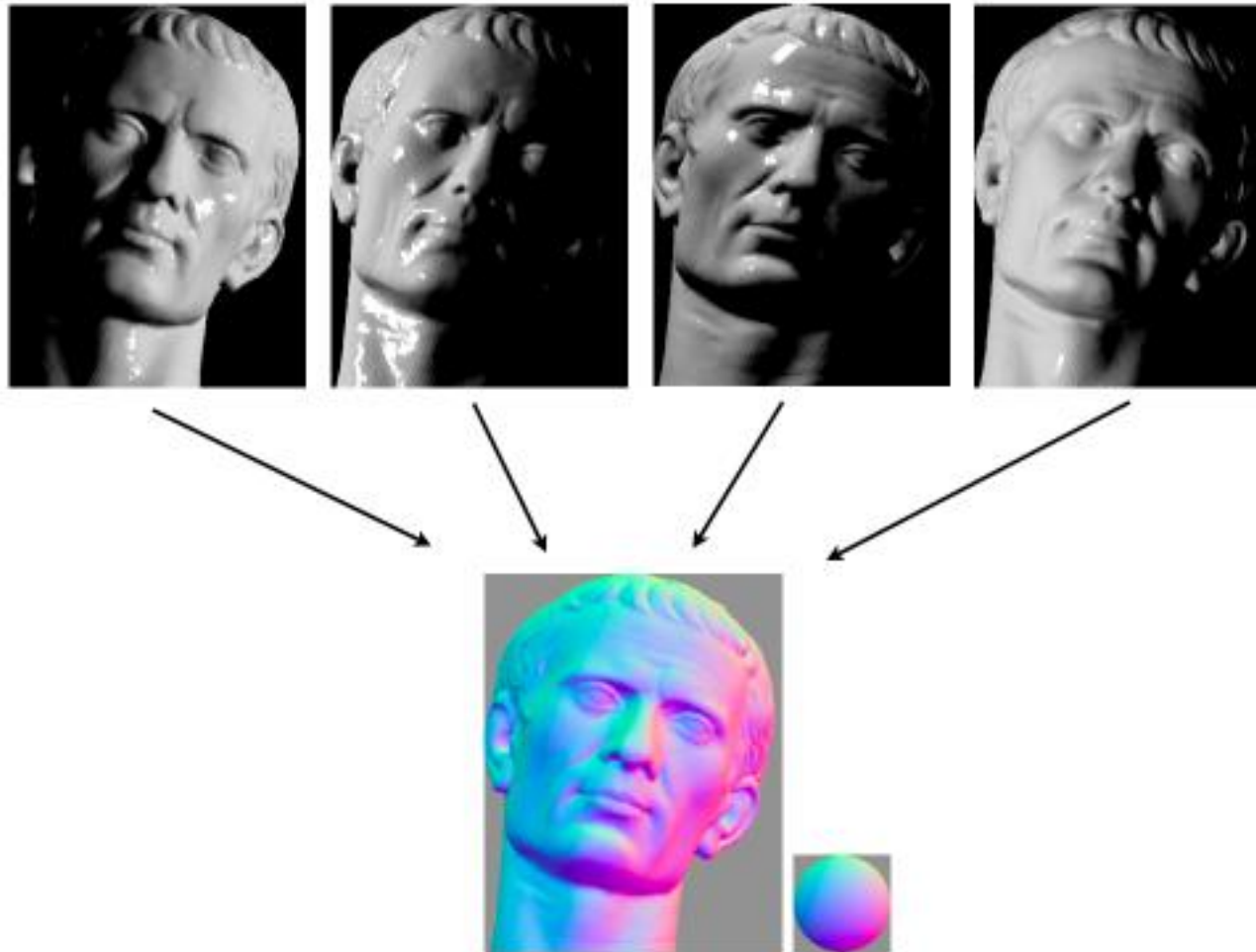
Structure from motion

# Project: Creating panoramas





# Project: 3D reconstruction



# 3. Recognition

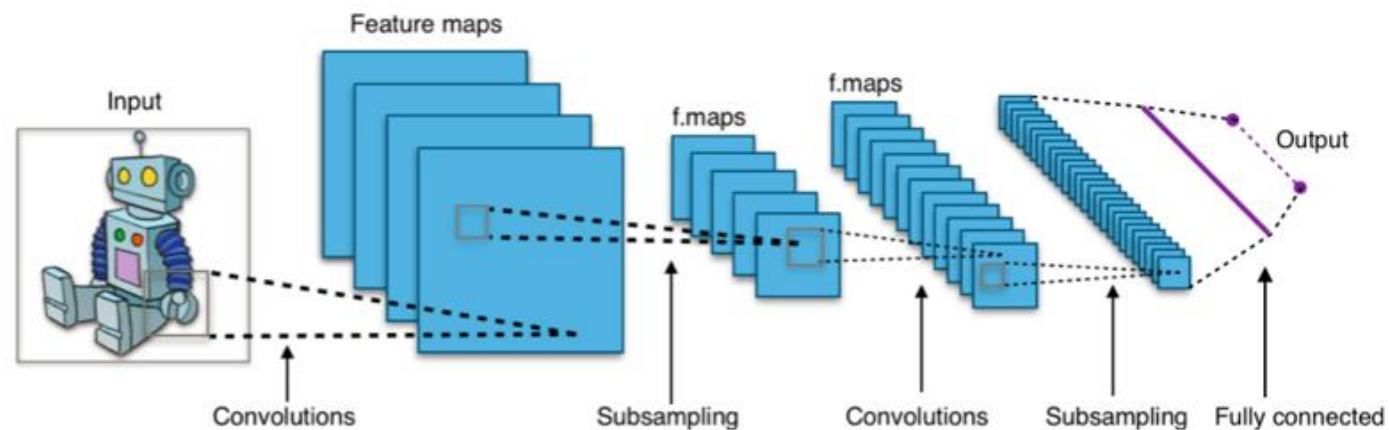


→ "dog"

Image classification

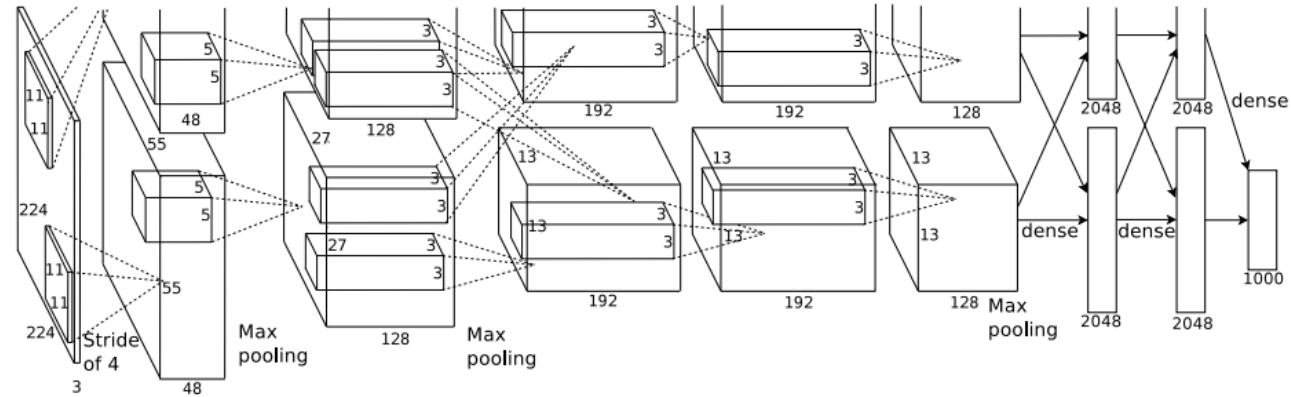


Object detection

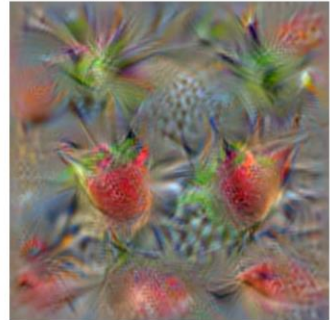


Convolutional Neural Networks

# Project: Convolutional Neural Networks



strawberry



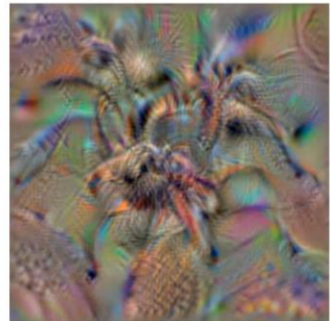
throne



mushroom



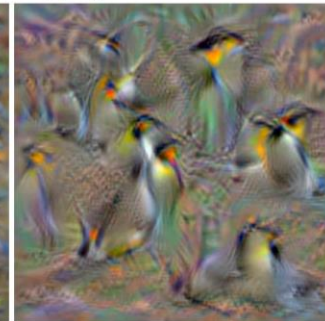
tarantula



flamingo



king penguin



# Grading

- Occasional quizzes (at the beginning of class)
- One midterm (take-home), one final exam (in class vs. take-home TBA)
- Grade breakdown (subject to minor tweaks):
  - Quizzes: 5% (lowest quiz grade dropped)
  - Midterm: 16%
  - Programming projects: 63%
  - Final exam: 16%

# Late policy

- Four free “slip days” will be available for the semester
- A late project will be penalized by 10% for each day it is late (excepting slip days), and no extra credit will be awarded

# Academic Integrity

- Assignments will be done solo or in pairs (we'll let you know for each project)
- Please do not leave any code public on GitHub (or the like) at the end of the semester!
- We will follow the Cornell Code of Academic Integrity (<http://cuinfo.cornell.edu/aic.cfm>)
- We reserve the right to run MOSS (automated code copying service) on submitted code

**Questions?**