

Computer Vision Overview

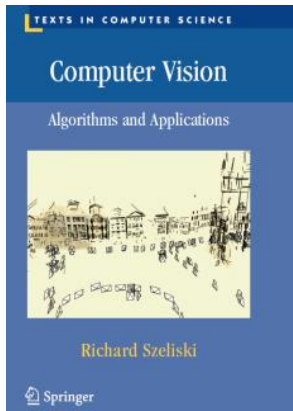
Samuel Cheng

Slide credits: Noah Snavely

Instructor

- Samuel Cheng
- Office hours:
 - By reservation
 - Please contact me through Discord (signup link on Canvas)
- Research interests:
 - Computer vision
 - Machine learning/AI
 - Signal processing

Other administrative details



- Reference:
Rick Szeliski, *Computer Vision: Algorithms and Applications*
online at: <http://szeliski.org/Book/>
- Course webpage:
http://www.samuelcheng.info/computer_vision_2023
- Discussion forum (Discord signup link on Canvas)

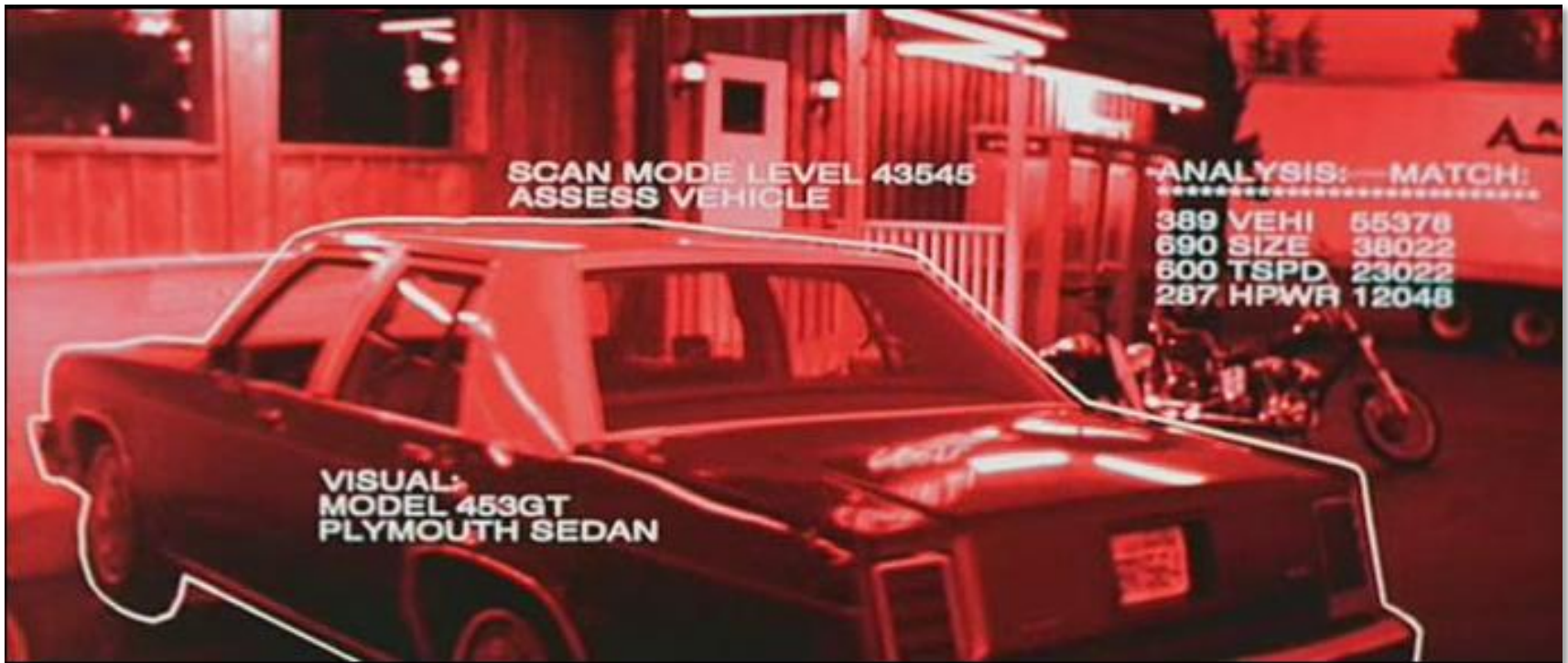
Course requirements

- Prerequisites
 - Python
 - Linear algebra
 - Vector calculus (a little bit)
 - Will review some materials when used
- Course does ***not*** assume prior imaging experience
 - computer vision, image processing, graphics, etc.

Today

1. What is computer vision?
2. Course overview

What is computer vision (CV)?



Every image tells a story



- CV aims to perceive the “story” behind the picture
 - Who?
 - What?
 - What happened?

CV is difficult in nature



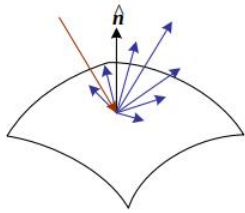
0	3	2	5	4	7	6	9	8
3	0	1	2	3	4	5	6	7
2	1	0	3	2	5	4	7	6
5	2	3	0	1	2	3	4	5
4	3	2	1	0	3	2	5	4
7	4	5	2	3	0	1	2	3
6	5	4	3	2	1	0	3	2
9	6	7	4	5	2	3	0	1
8	7	6	5	4	3	2	1	0

Can the computer match human perception?



- Yes and no (mainly no)
 - computers can be better at “easy” things
 - humans are much better at “hard” things
- But huge progress has been made
 - Especially in the last 10 years
 - What is considered “hard” keeps changing

CV covers lots of topics



2. Image Formation



3. Image Processing



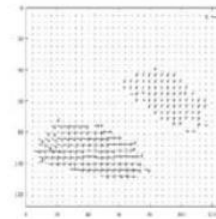
4. Features



5. Segmentation



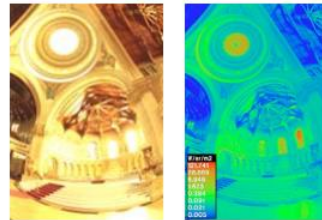
6-7. Structure from Motion



8. Motion



9. Stitching



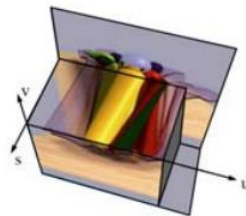
10. Computational Photography



11. Stereo



12. 3D Shape



13. Image-based Rendering

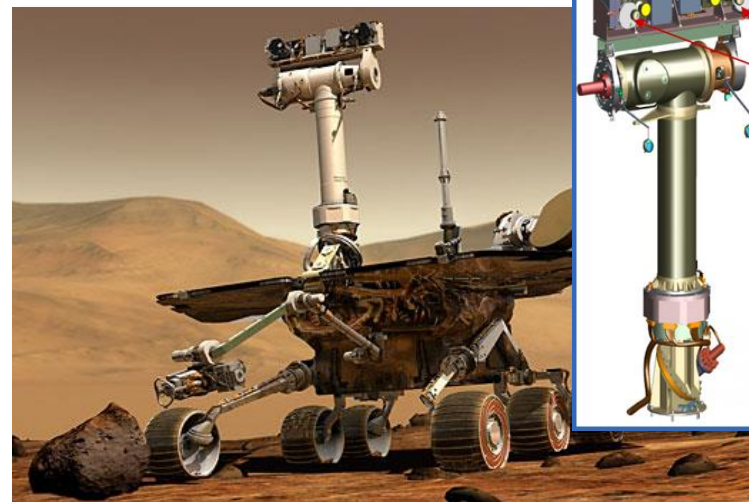


14. Recognition

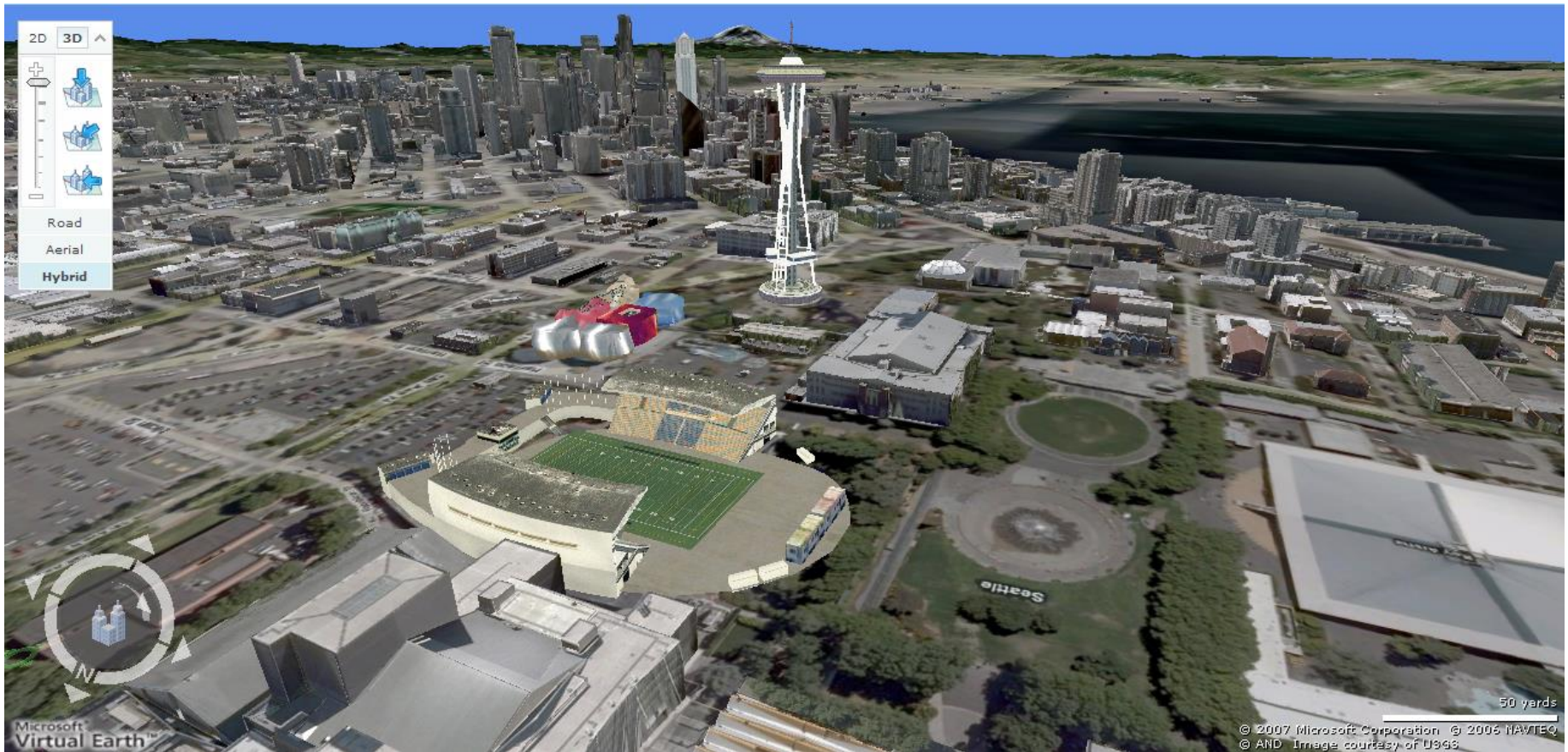
Some CV applications



3D Modeling



3D urban modeling



[Bing maps](#), Google Streetview

Source: S. Seitz

Some CV applications

- Recognizing objects and people





sky

building

flag

face

banner

wall

street lamp

bus

bus

cars

slide credit: Fei-Fei, Fergus & Torralba



Some CV applications

- “Enhancing” images





Some CV applications

- “Enhancing” images (c.f. Computational Photography)



Super-resolution / denoising
(source: 2d3)



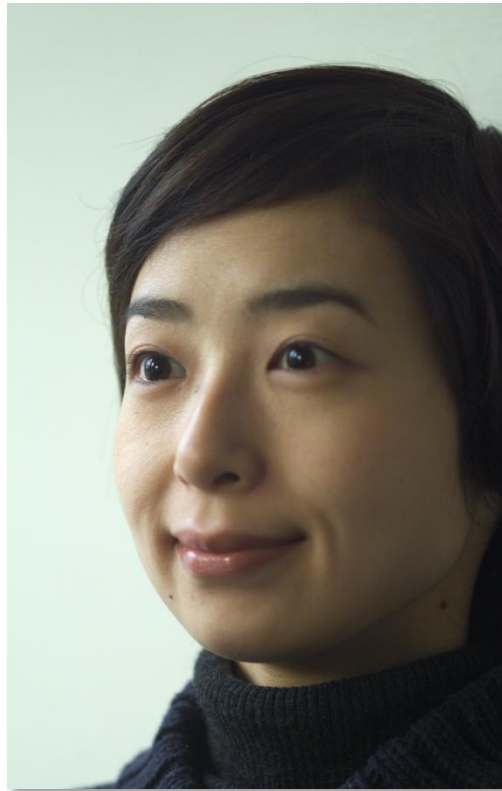
Texture synthesis / increased field of view (uncropping)
(image credit: Efros and Leung)



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

Some CV applications

- Forensics





Source: Nayar and Nishino, "Eyes for Relighting"



COMPUTER VISION PROBLEMS

Optical character recognition (OCR)

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



Digit recognition, AT&T labs

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

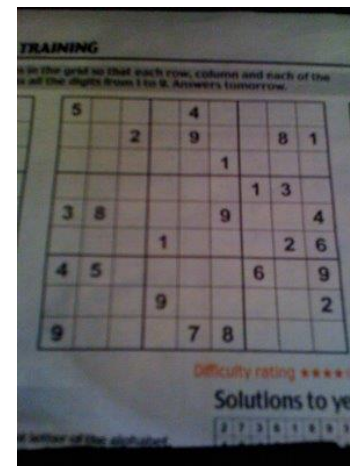


License plate readers

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



Automatic check processing



Sudoku grabber

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

Face detection

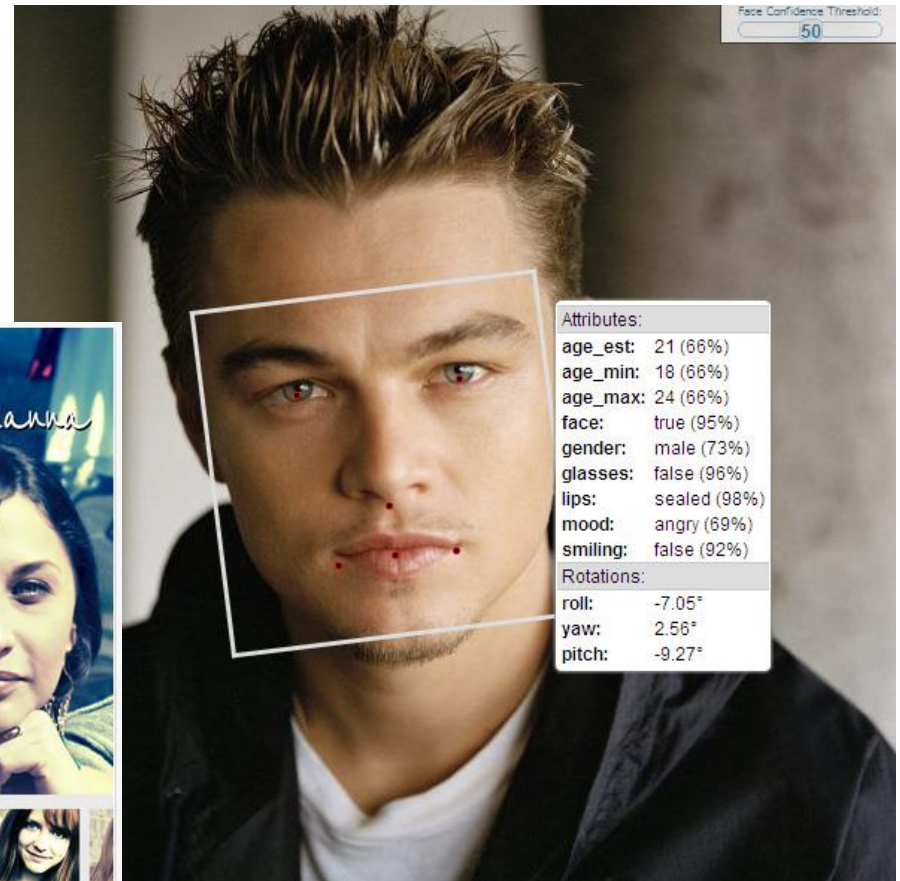


- Many new digital cameras now detect faces
 - Canon, Sony, Fuji, ...

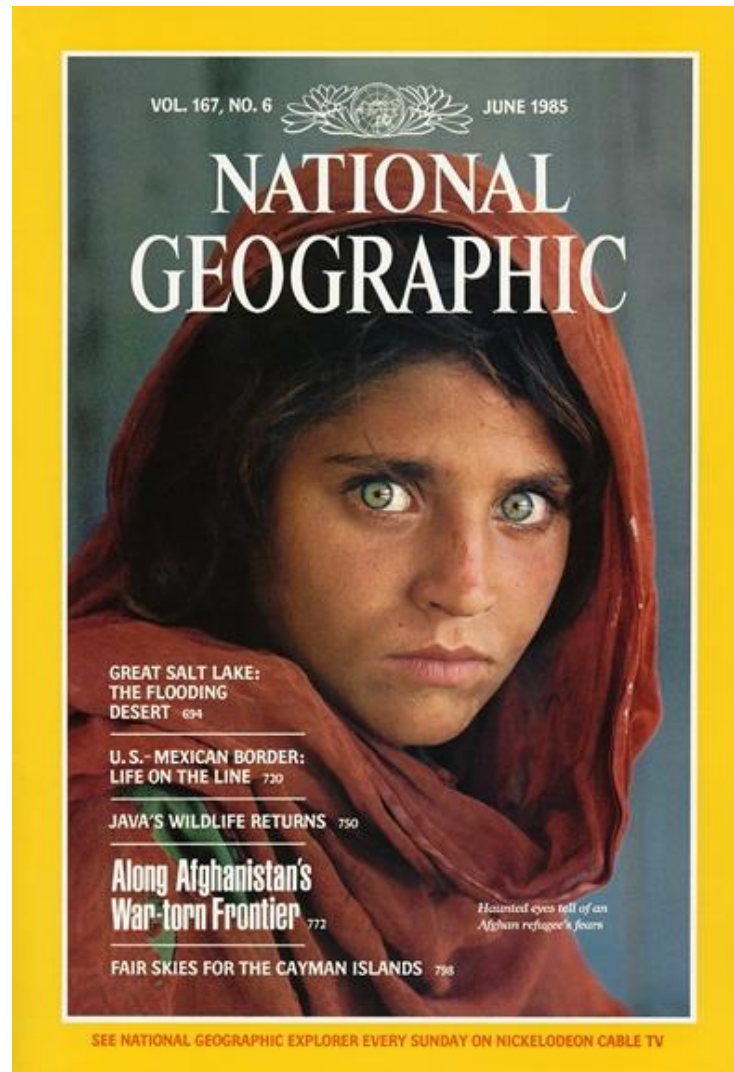
Face Recognition



<http://developers.face.com/tools/>



Face recognition



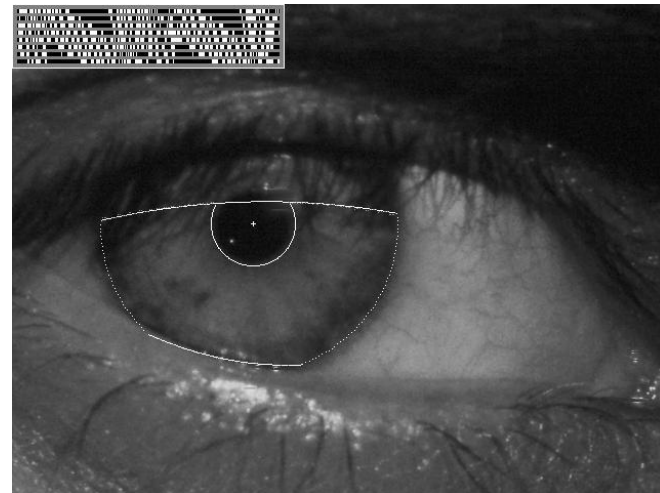
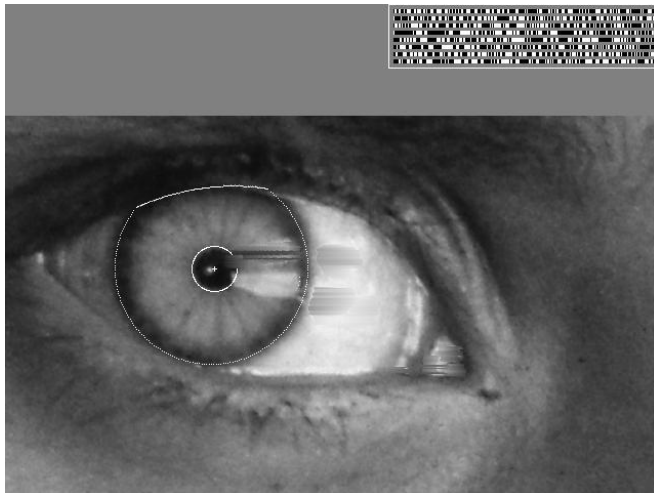
Who is she?

Source: S. Seitz

Vision-based biometrics



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



Login without a password...



Fingerprint scanners on many new laptops, other devices



Face recognition systems now beginning to appear more widely

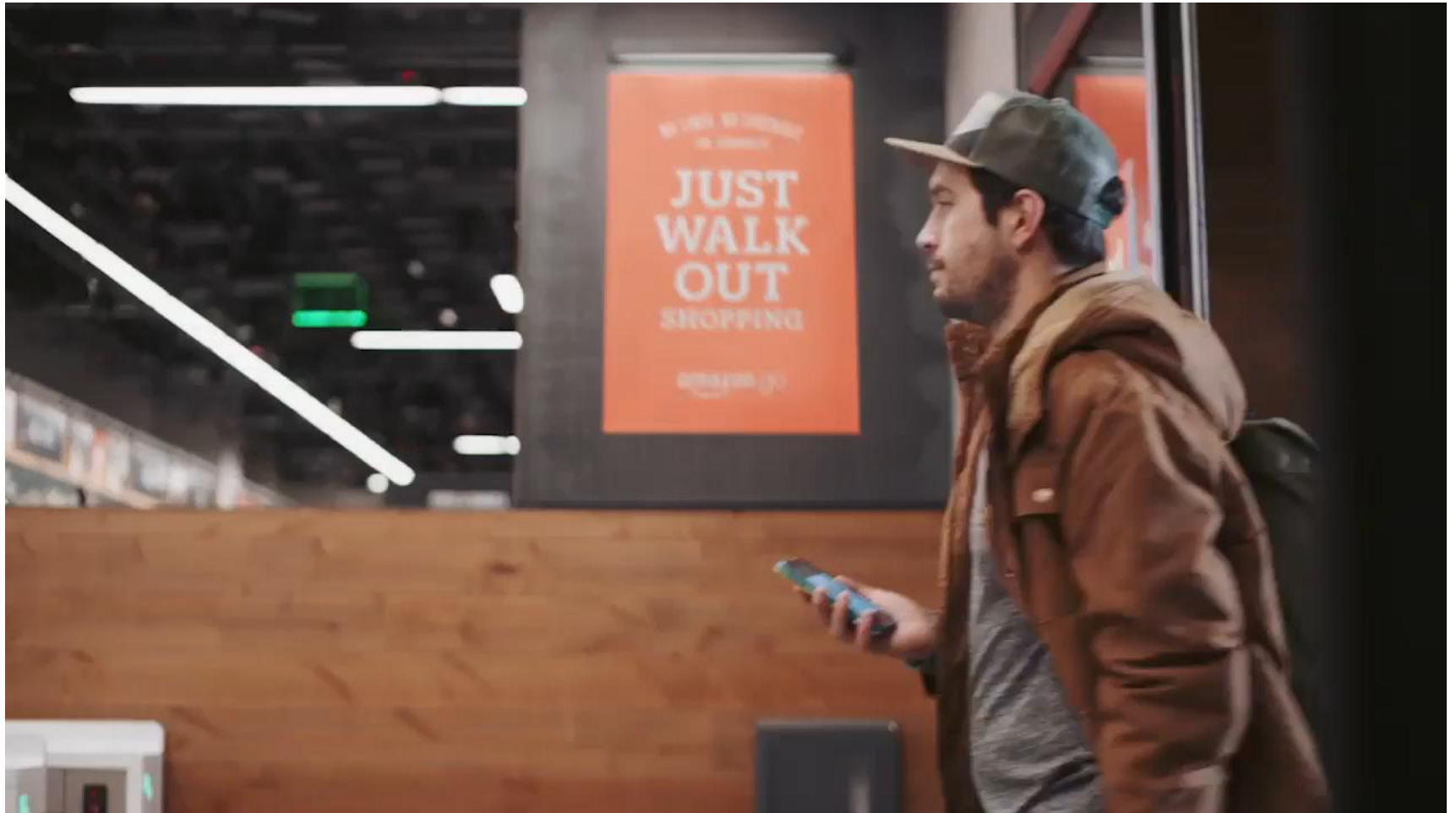
Object recognition (in supermarkets)



[LaneHawk by EvolutionRobotics](#)

“A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk, you are assured to get paid for it...”

Object recognition (in supermarkets)



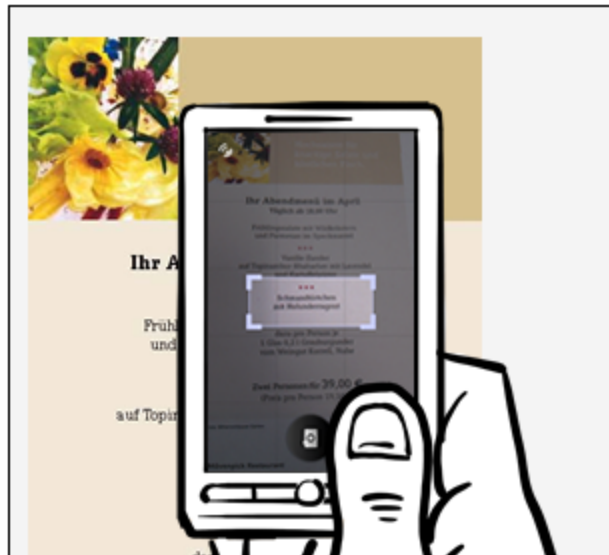
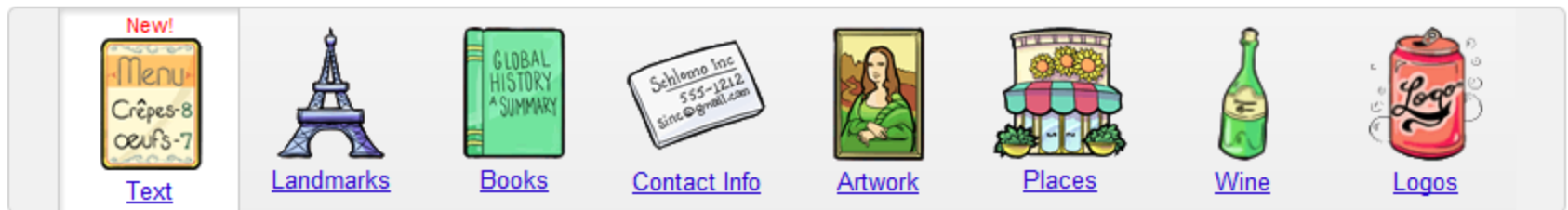
Object recognition (in mobile phones)



Google Goggles

Google Goggles in action

Click the icons below to see the different kinds of objects and places you can search for using Google Goggles.



Google Search by Image

- <https://images.google.com>

Leaf of the Bottlebrush Buckeye

Leafsnap: An Electronic Field Guide

Leafsnap is the first in a series of electronic field guides being developed by researchers from [Columbia University](#), the [University of Maryland](#), and the [Smithsonian Institution](#). This free mobile app uses visual recognition software to help identify tree species from photographs of their leaves.

Leafsnap contains beautiful high-resolution images of leaves, flowers, fruit, petiole, seeds, and bark. Leafsnap currently includes the trees of the Northeast and will soon grow to include the trees of the entire continental United States.

This website shows the tree species included in Leafsnap, the collections of its users, and the team of research volunteers working to produce it.

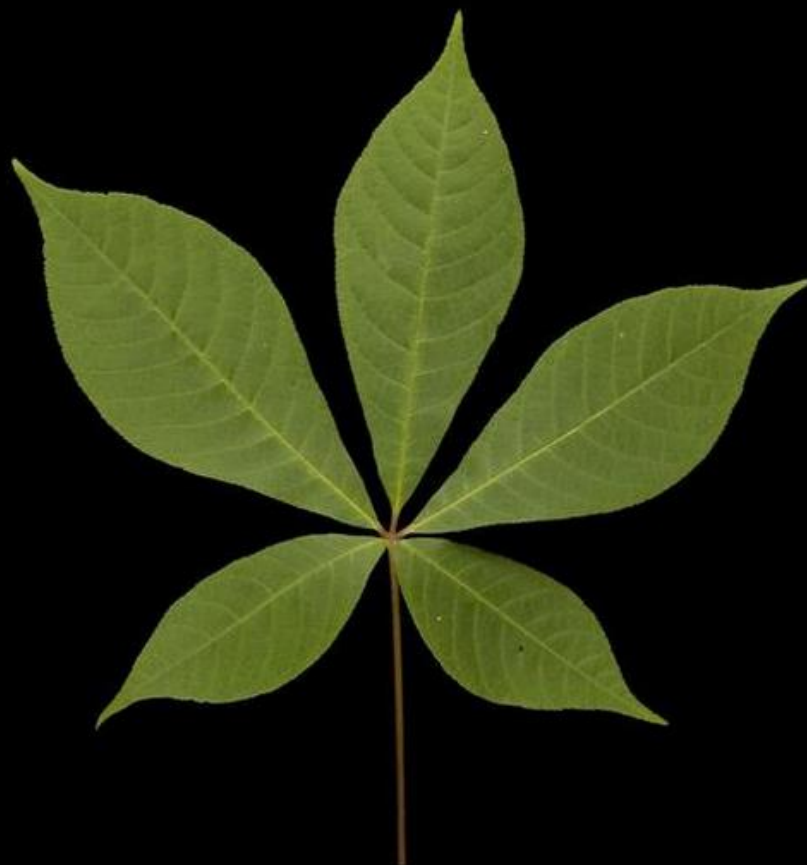
Free for iPhone:



and iPad:



guardian.co.uk



3D Mapping



Smart and self-driving cars

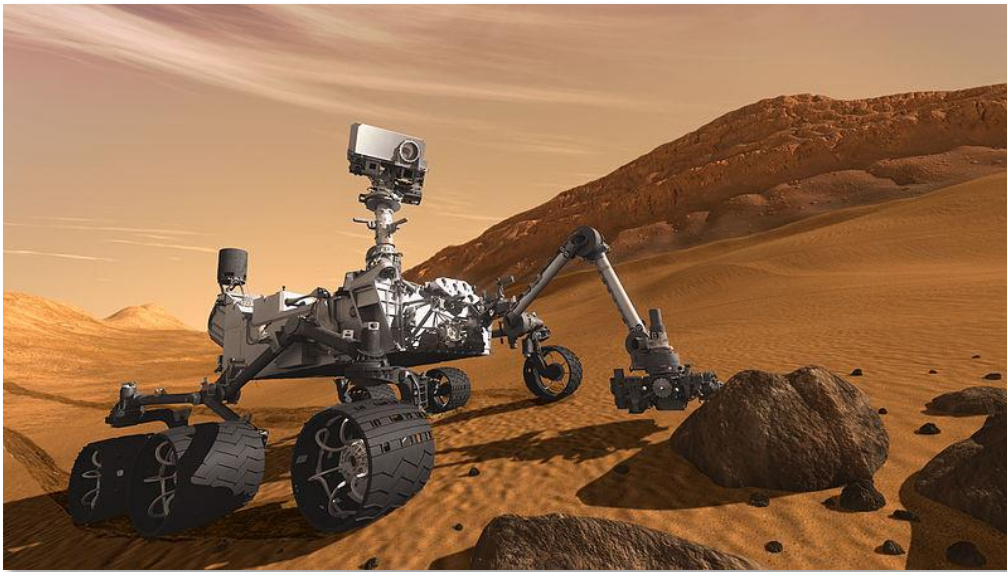
The screenshot displays the Mobileye website with the following content:

- Navigation:** "manufacturer products" (selected) and "consumer products".
- Header:** "Our Vision. Your Safety."
- Car Diagram:** A top-down view of a car with three camera fields of view: "rear looking camera", "side looking camera", and "forward looking camera".
- Product Highlights:**
 - EyeQ Vision on a Chip:** Includes an image of the EyeQ chip and a "read more" link.
 - Vision Applications:** Shows a pedestrian in a bounding box with the text "Road, Vehicle, Pedestrian Protection and more" and a "read more" link.
 - AWS Advance Warning System:** Shows a car icon on a circular display with "0.8" below it and a "read more" link.
- News Section:**
 - Article 1: "Mobileye Advanced Technologies Power Volvo Cars World First Collision Warning With Auto Brake System"
 - Article 2: "Volvo: New Collision Warning with Auto Brake Helps Prevent Rear-end"
 - Link: "> all news"
- Events Section:**
 - Event 1: "Mobileye at Equip Auto, Paris, France"
 - Event 2: "Mobileye at SEMA, Las Vegas, NV"
 - Link: "> read more"

- [Mobileye](#)

- Vision systems currently in high-end BMW, GM, Volvo models

Robot vision



NASA's Mars Curiosity Rover (Mars Science Laboratory)

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



Autonomous RC Car

<http://www.cs.cornell.edu/~asaxena/rccar/>

Robot vision in space



The Heights of Mount Sharp

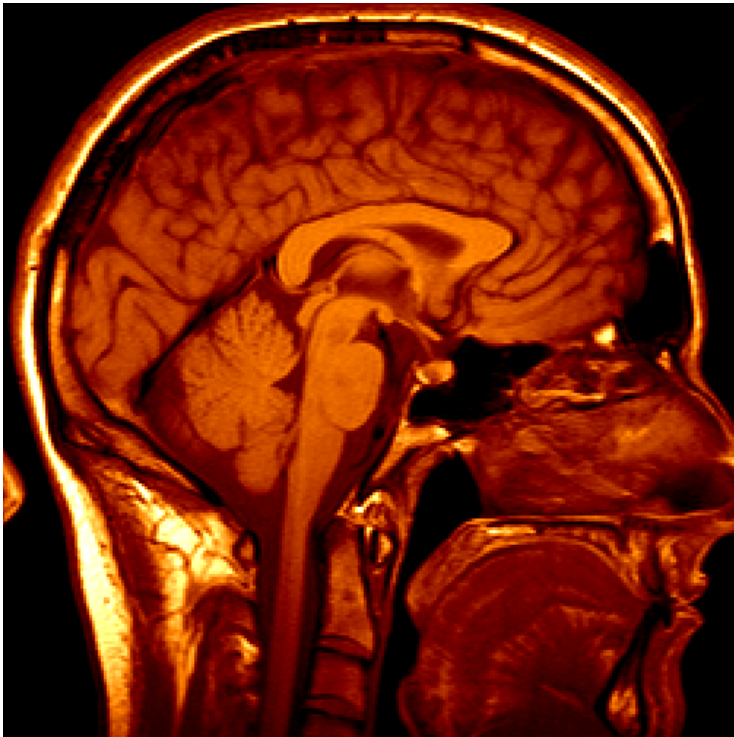
http://www.nasa.gov/mission_pages/msl/multimedia/pia16077.html

Panorama captured by Curiosity Rover, August 18, 2012 (Sol 12)

Vision systems (JPL) uses for several tasks

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

Medical imaging



3D imaging
MRI, CT

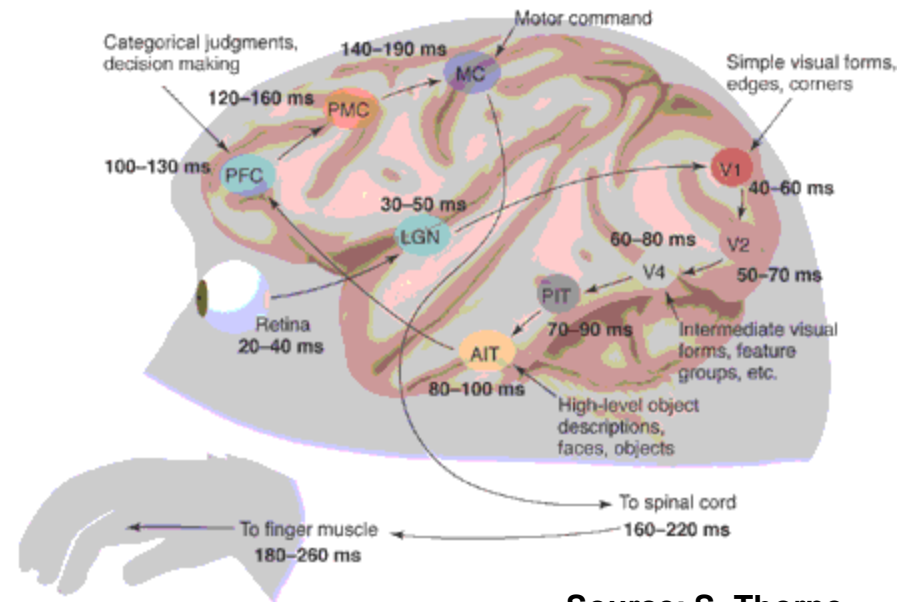


Image guided surgery
[Grimson et al., MIT](#)

WHY IS CV DIFFICULT?

Discover

SCIENCE FOR THE CURIOUS



FROM THE JUNE 1993 ISSUE

Source: S. Thorpe

The Vision Thing: Mainly in the Brain

The machinery that accomplishes these tasks is by far the most powerful and complex of the sensory systems. The retina, which contains 150 million light-sensitive rod and cone cells, is actually an outgrowth of the brain. In the brain itself, neurons devoted to visual processing number in the hundreds of millions and take up about 30 percent of the cortex, as compared with 8 percent for touch and just 3 percent for hearing. Each of the two optic nerves, which carry signals from the retina to the brain, consists of a million fibers; each auditory nerve carries a mere 30,000.

Why is computer vision difficult?



Viewpoint variation



Illumination



Scale

Why is computer vision difficult?



Intra-class variation



Motion (Source: S. Lazebnik)

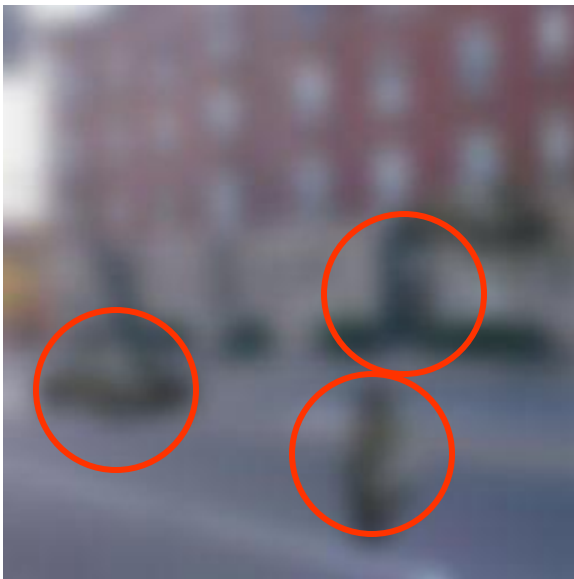
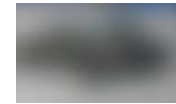
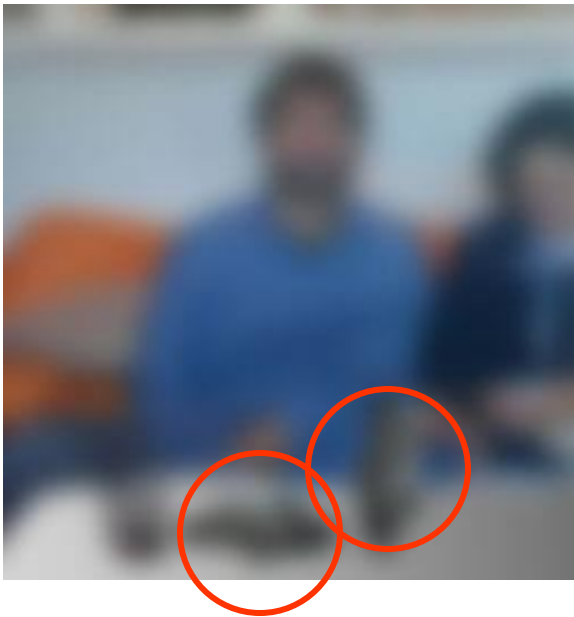


Background clutter



Occlusion

Challenges: local ambiguity





But there are lots of cues we can exploit...



Bottom line

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

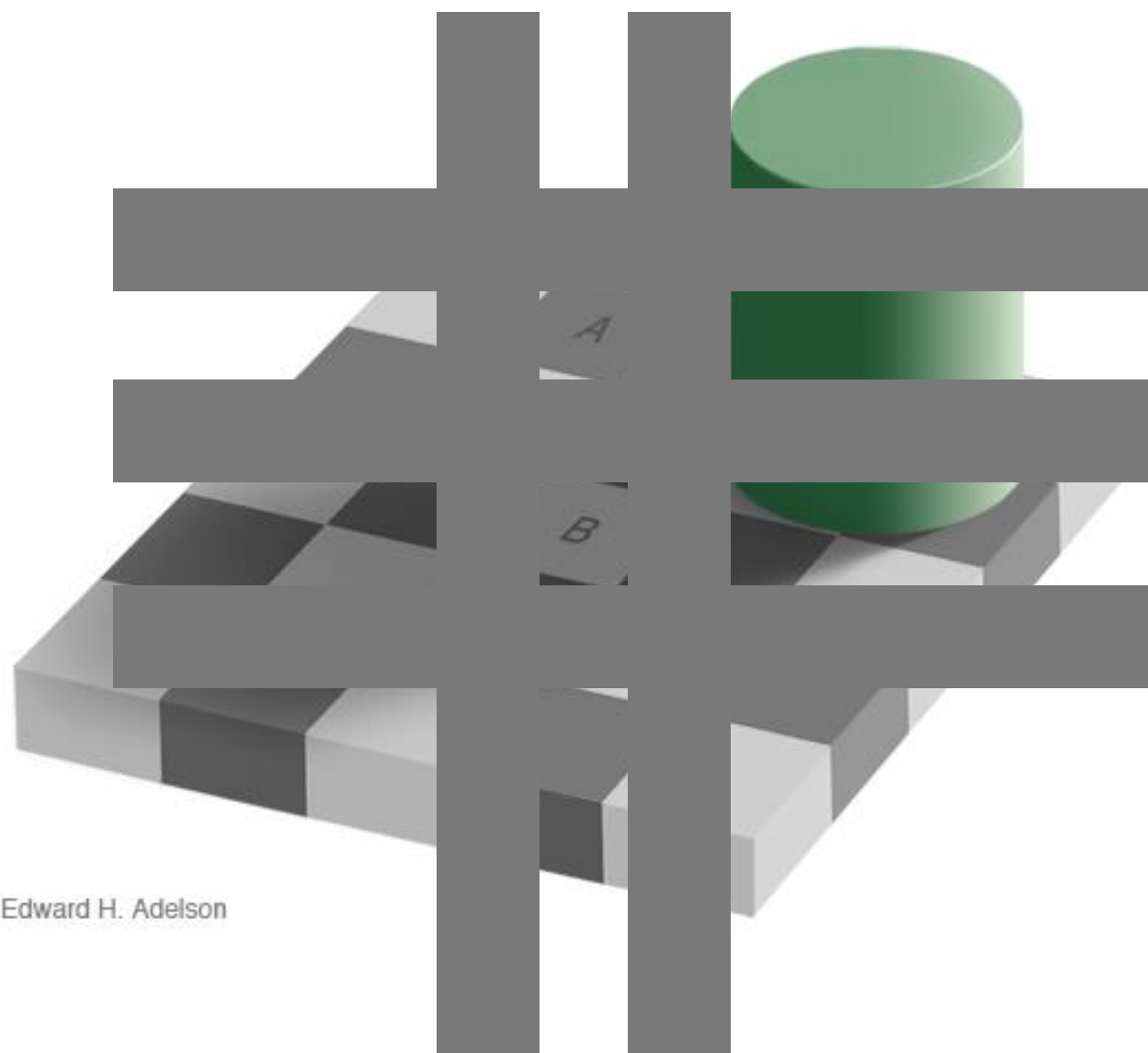


- We often need to use prior knowledge about the structure of the world

Human perception has its “shortcomings”

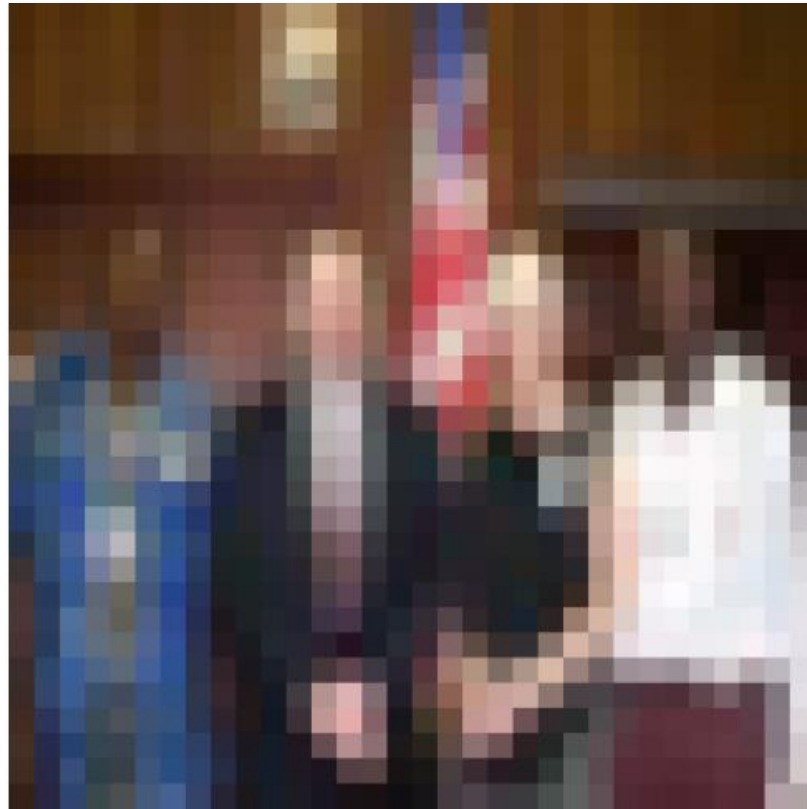


[Sinha and Poggio, *Nature*, 1996](#)



Edward H. Adelson

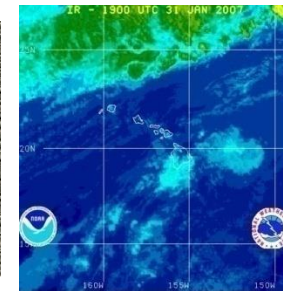
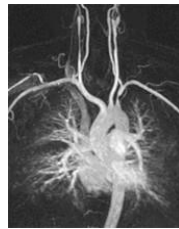
But humans can tell a lot about a scene from a little information...



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

Why study CV?

- Millions of images being captured all the time



- Loads of useful applications

Expected outcomes

- Able to implement simple low-level operations on images and videos (manipulate them directly as matrices and tensors)
- Able to use OpenCV create simple applications
- Expose to sufficient theoretical background to understand current research

Course overview

- Project (HW) based. Expected quite a few programming assignments
 - Bottom-up: implement simple filter
 - Top-down: camera calibration
- Final project of your choice
- Presentations if taking graduate credits
- Extra credit: in-class participation
- No exam

Grading

- Please see course website for detail breakdown

Late policy

- Late projects will be penalized by 5% for each day it is late, and penalty capped after 10 days
 - Penalty is mild and so no exception will be granted

Academic Integrity

- Common sense
- Discussion okay (and encouraged) but please do your own work for the final solution. That is the only way to learn
 - Take advantage of Discord
- You may try to get pointers from AI (chatgpt) and Google. But don't take them as truth. Verify it yourself



Questions?

Sample projects

- [Invisibility cloak](#)
- [Face mask detection](#)
- [Ergonomic risk analysis](#)
- [Air canvas](#)
- [Cup full or not](#)
- [Video summarization](#)