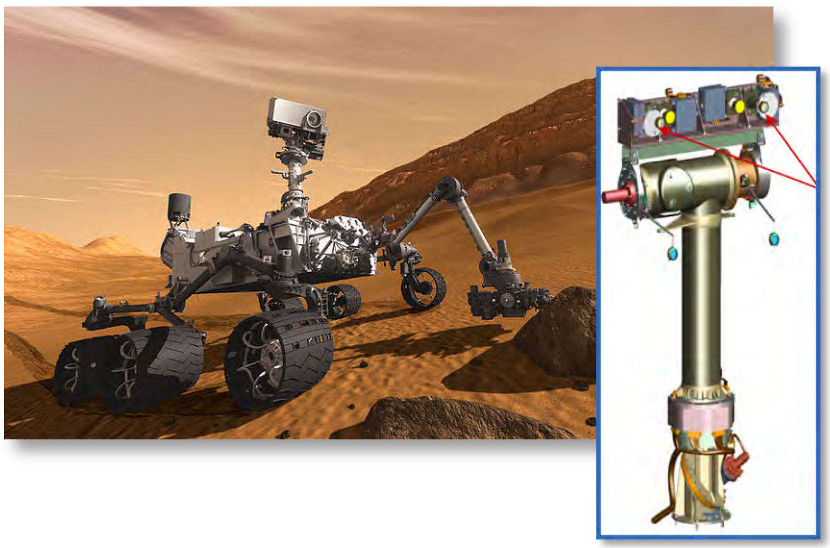
CS4670/5670: Intro to Computer Vision

Instructor: Kavita Bala







Instructor

- Kavita Bala (kb@cs.cornell.edu)

- Office hours:

  F: 10-11, or by appointment

- Research interests:
  - Computer graphics and vision

Today

1. What is computer vision?

2. Course overview

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

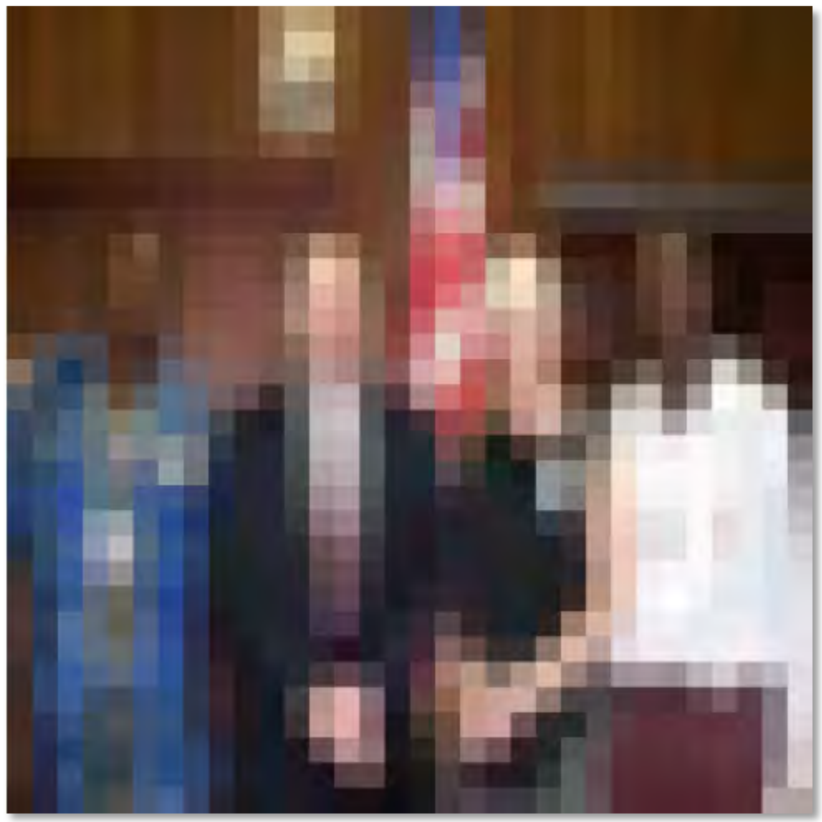
Can the computer match human perception?

- 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

Human perception has its shortcomings

[Sinha and Poggio, *Nature*, 1996]

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

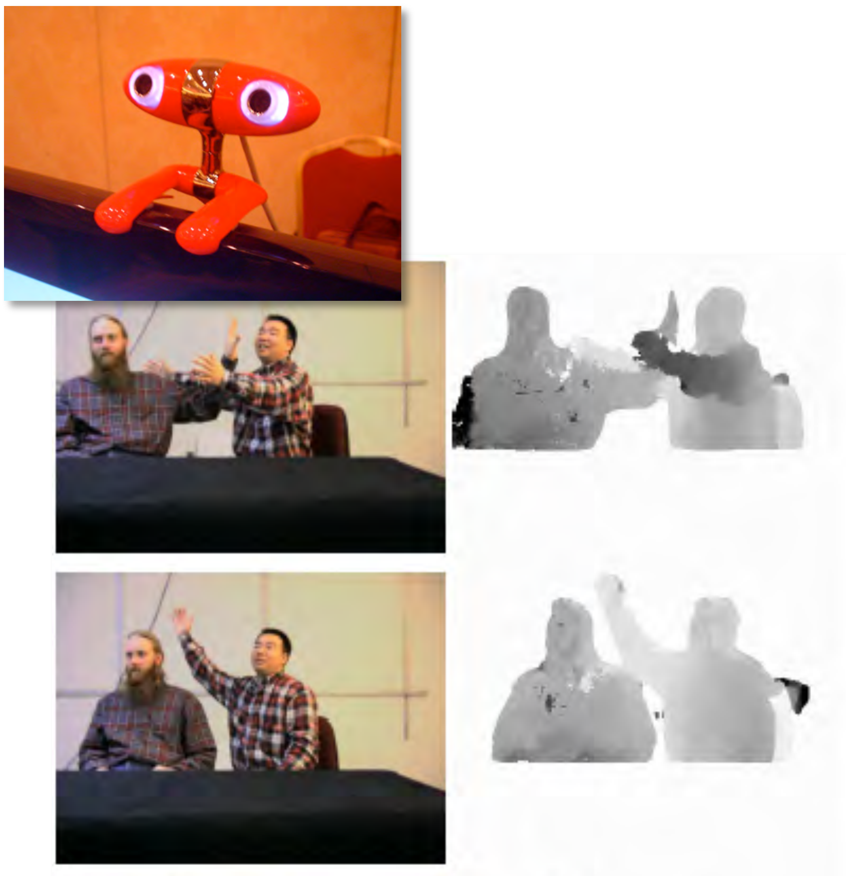


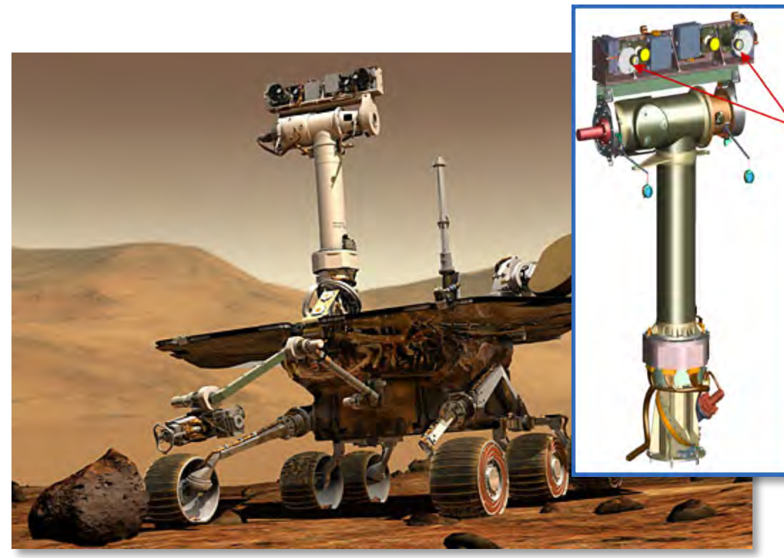
Source: "80 million tiny images" by Torralba, et al.

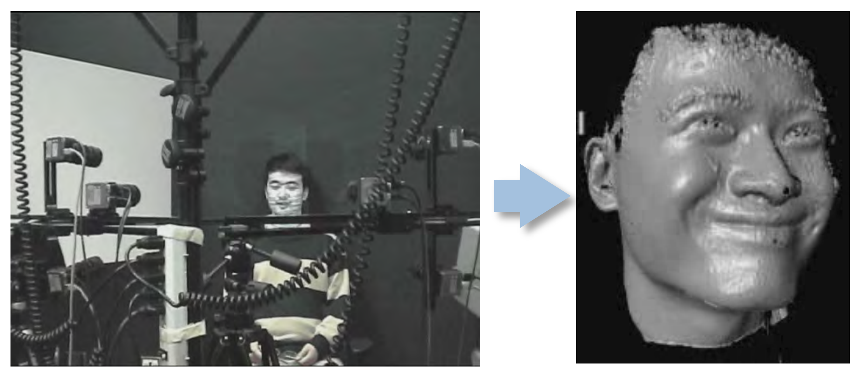
The goal of computer vision

The goal of computer vision

- Computing the 3D shape of the world

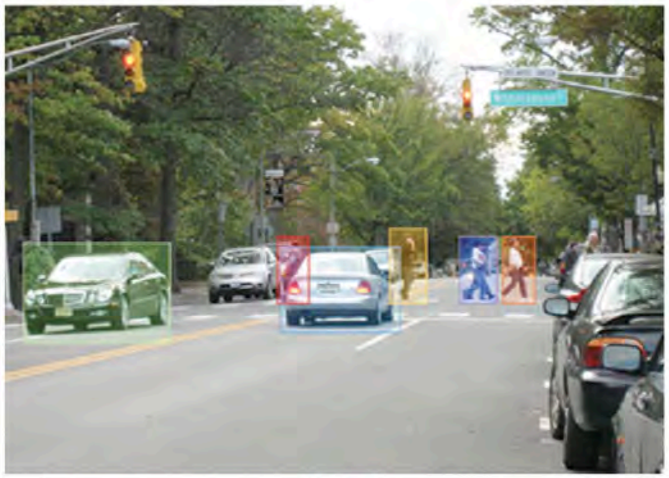




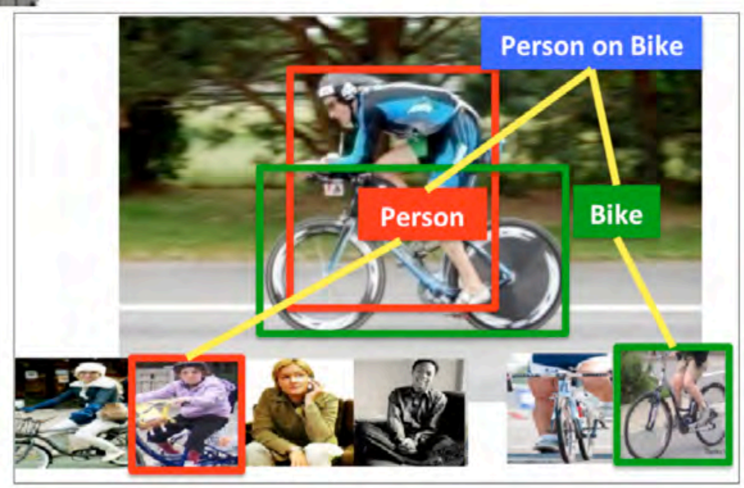


The goal of computer vision

- Recognizing objects and people



- Object detection
- Action classification
- Image captioning
- …



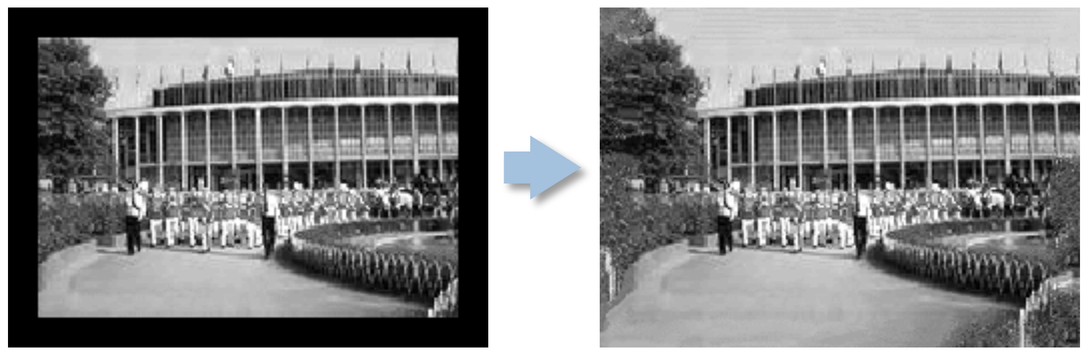
中华人
和国万岁
世界
民大团结万岁


sky
building
flag
face
banner
wall
street lamp
bus
bus
cars
slide credit: Fei-Fei, Fergus & Torralba

The goal of computer vision

- "Enhancing" images (c.f. Computational Photography)

Super-resolution / denoising
(source: 2d3)



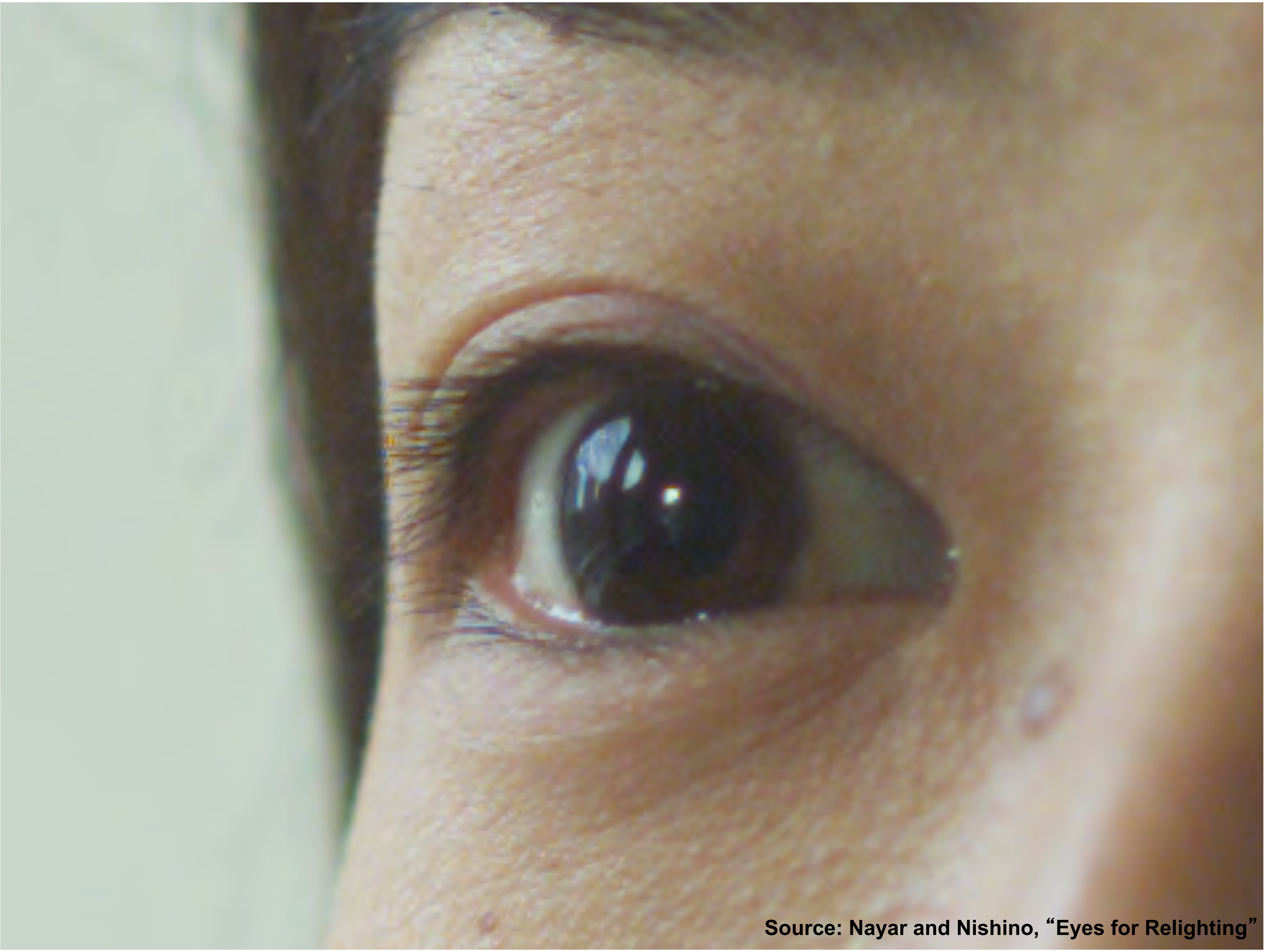
Texture synthesis / increased field of view (uncropping)
(image credit: Efros and Leung)

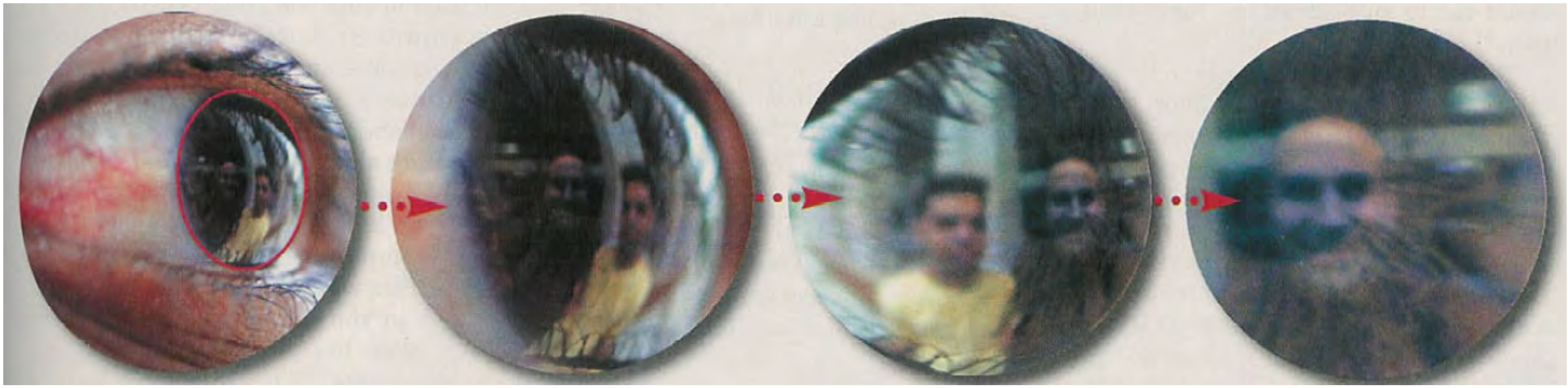
Inpainting / image completion
(image credit: Hays and Efros)

The goal of computer vision

- Forensics

Source: Nayar and Nishino, "Eyes for Relighting"

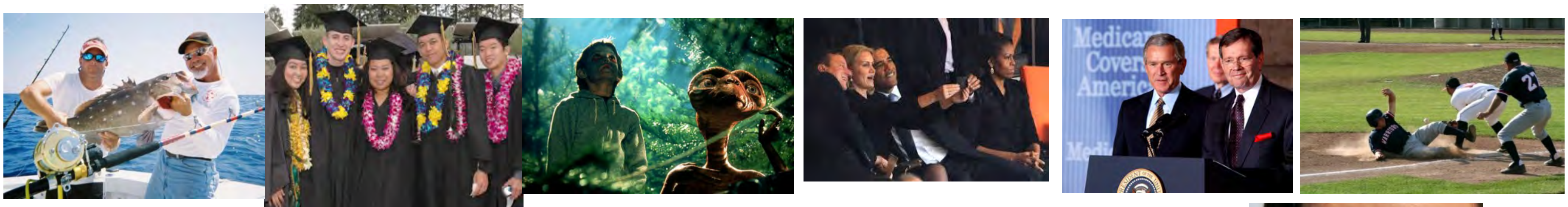

Source: Nayar and Nishino, “Eyes for Relighting”

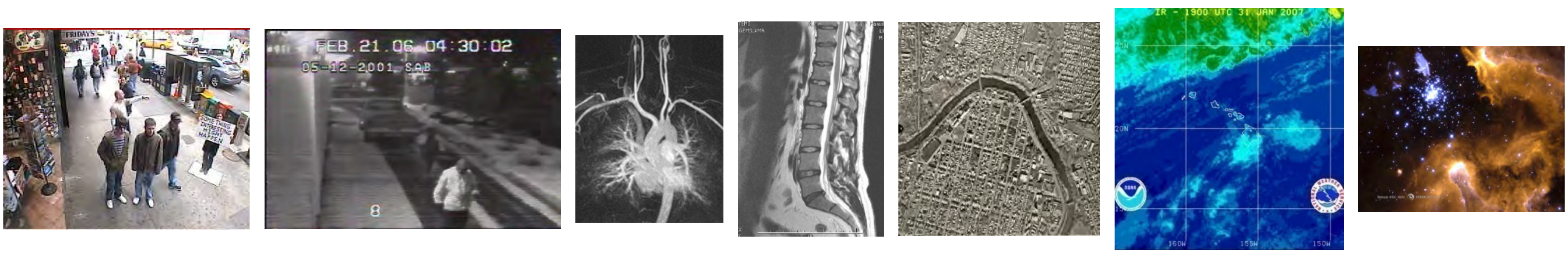


Source: Nayar and Nishino, “Eyes for Relighting”

Why study computer vision?

- Millions of images being captured all the time



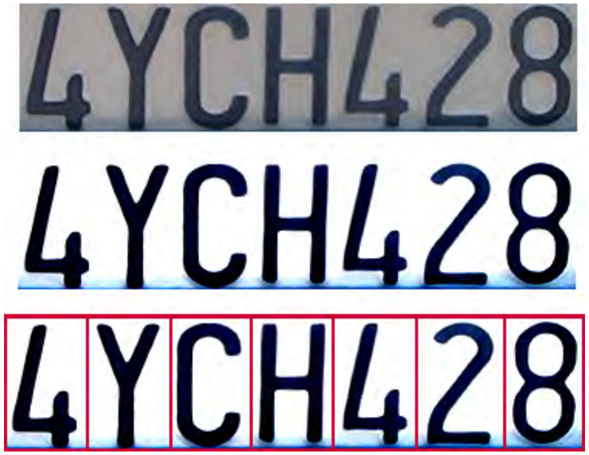


- Loads of useful applications
- The next slides show the current state of the art

Optical character recognition (OCR)

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

Digit recognition, AT&T labs
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/

Source: S. Seitz

Face detection

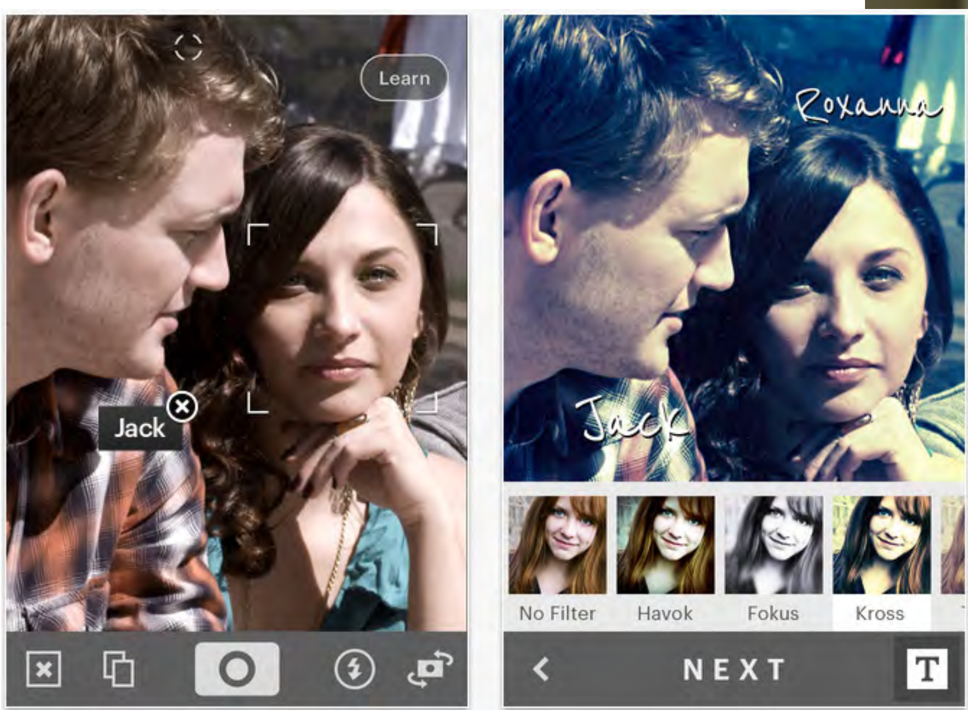
- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, …

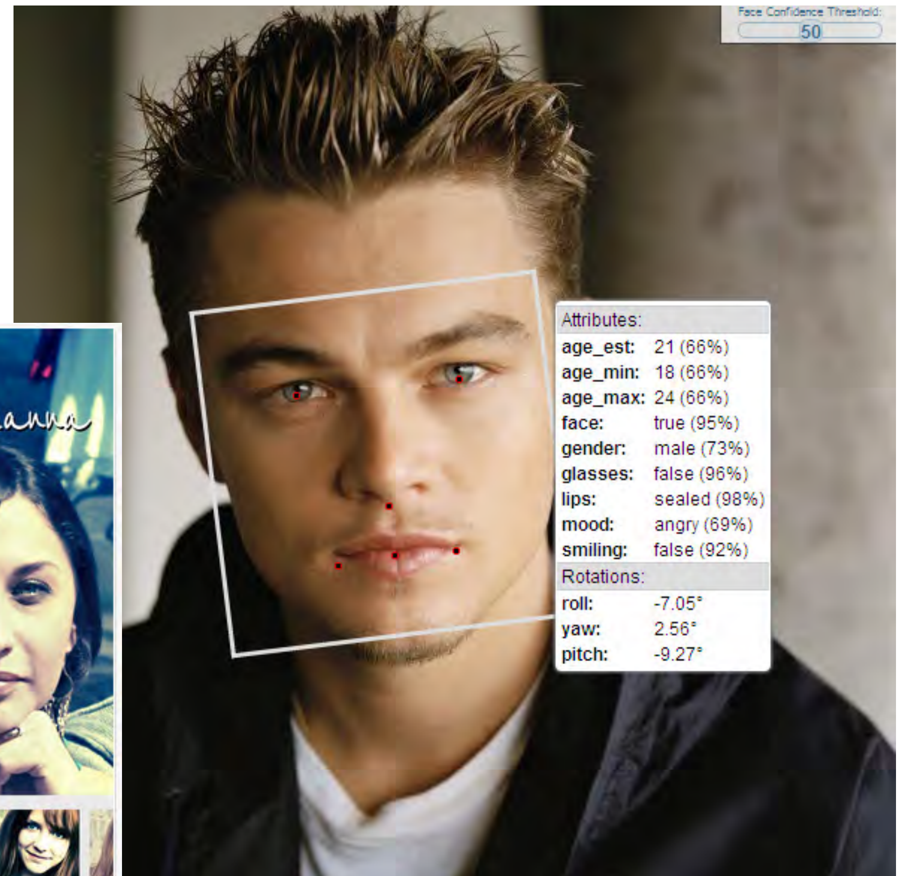
Source: S. Seitz

Face Recognition

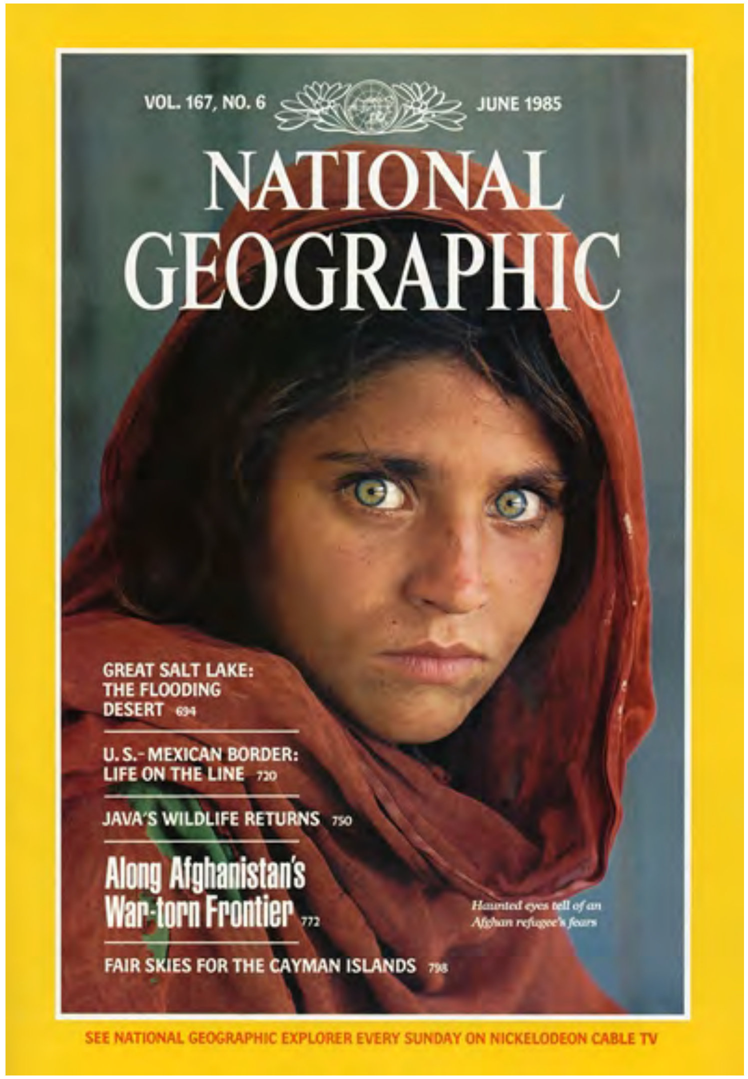
face.com

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





Face recognition

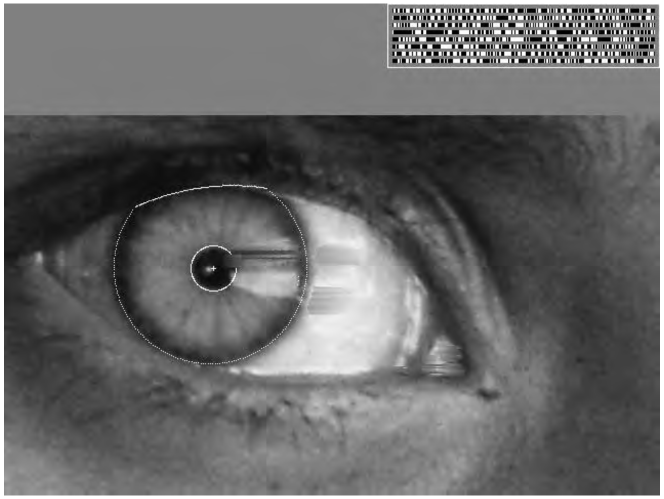


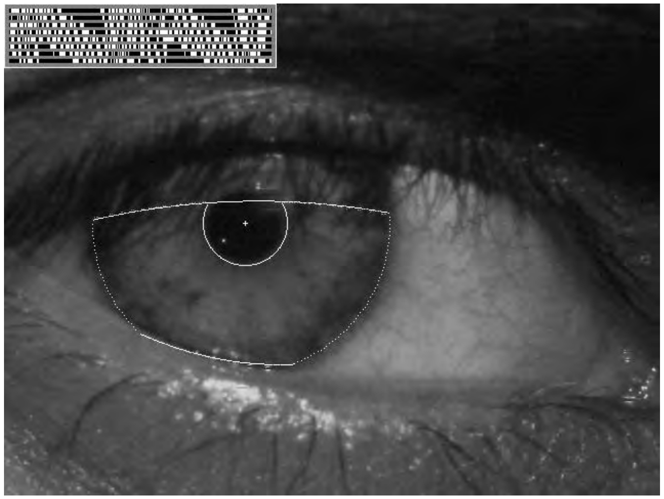
Who is she?

Source: S. Seitz

Vision-based biometrics

“*How the Afghan Girl was Identified by Her Iris Patterns*” Read the story





Source: S. Seitz

Login without a password…

Fingerprint scanners on
many new laptops,
other devices

Face recognition systems now
beginning to appear more widely
http://www.sensiblevision.com/

Source: S. Seitz

Object recognition (in supermarkets)

LaneHawk by EvolutionRobotics

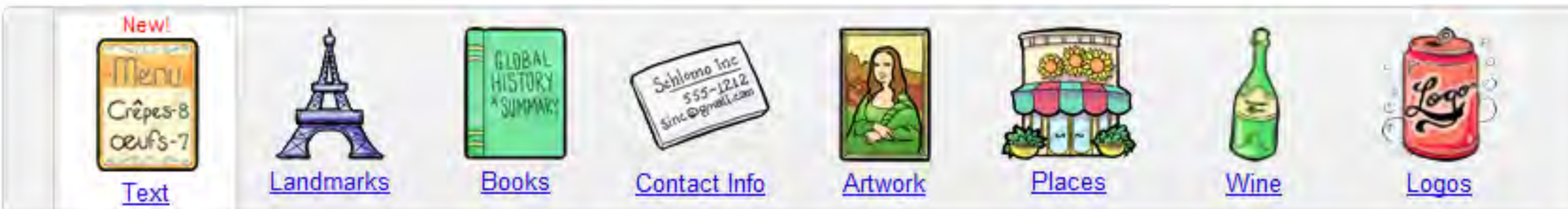
"A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you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

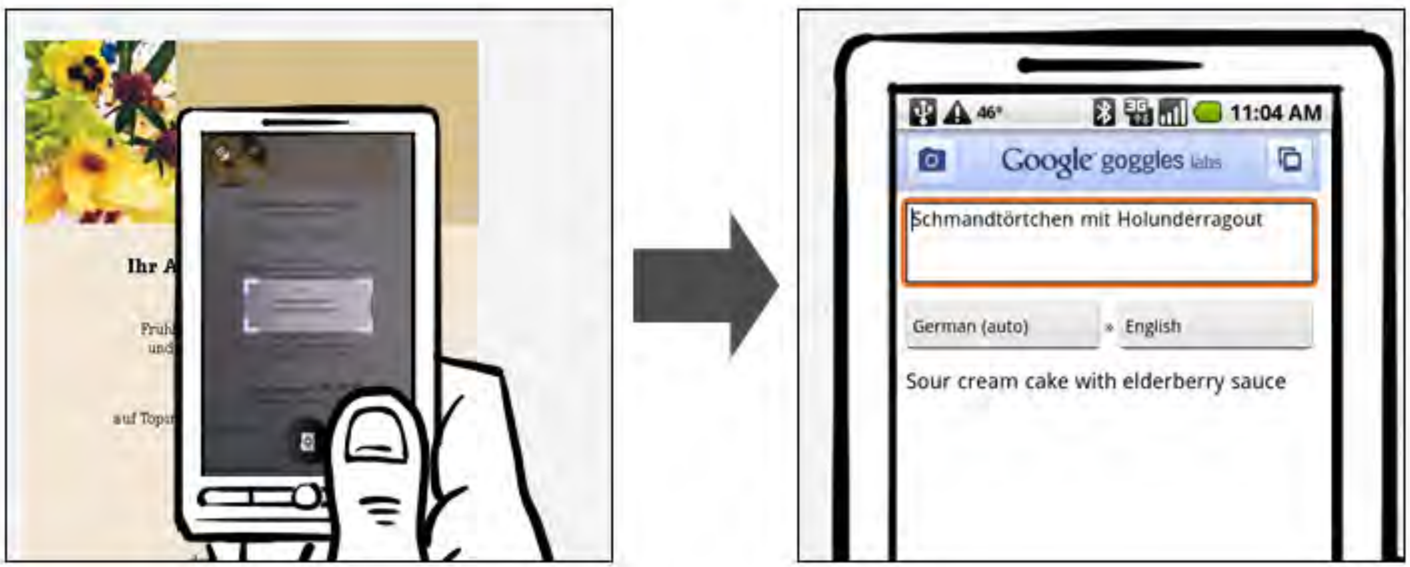
Source: S. Seitz

Google Goggles

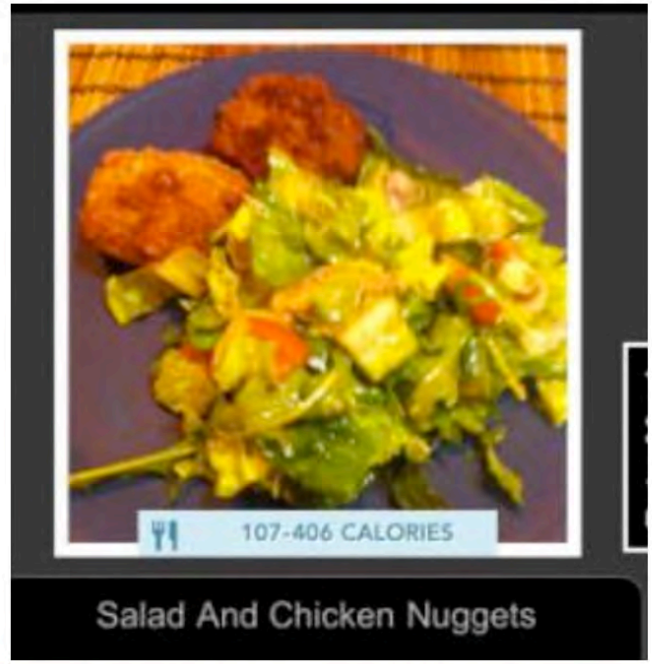
Google Goggles in action

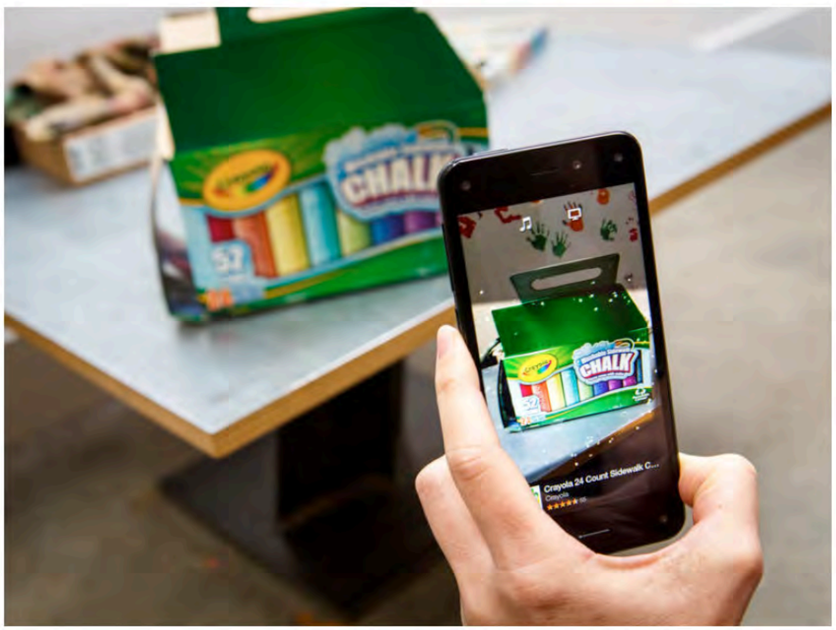
Click the icons below to see the different kinds of objects and places you can search for using Google Goggles.

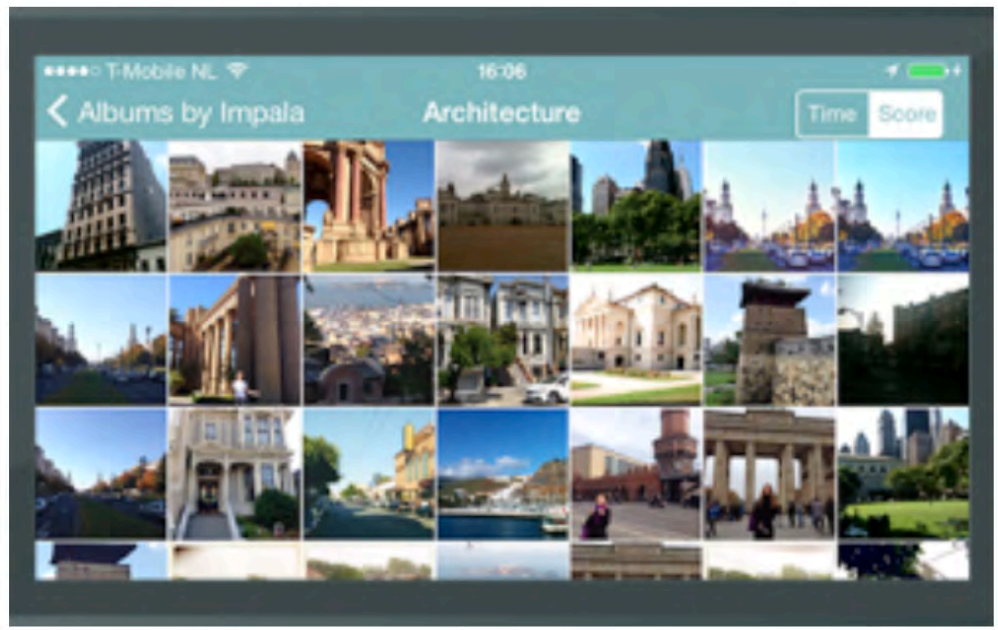





2:53 PM
Google goggles labs
Landmark
Golden Gate Bridge
Golden Gate Bridge
Web Results
Golden Gate Bridge - Wikipedia, the free encyclopedia
The Golden Gate Bridge by night, with part of downtown San Francisco ... Golden Gate Bridge is the most popular place to commit suicide in the United States ...
http://en.wikipedia.org/wiki/Golden_Gate_Bridge
Seacliff Webcam - Weather Seacliff, Golden Gate Bridge (Seacliff ...


107-406 CALORIES
Salad And Chicken Nuggets


CHALK


T-Mobile NL
16:06
Albums by Impala
Architecture
Time
Score

Google Search by Image

leafsnap

Home | Species | Collectors | About

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: Available on the App Store **and iPad:** Available on the iPad App Store

You Tube

guardian.co.uk

Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Source: S. Seitz

Special effects: motion capture

Pirates of the Carribean, Industrial Light and Magic

Source: S. Seitz

Sports

Sportvision first down line
Nice explanation on www.howstuffworks.com

Source: S. Seitz

Vision-based interaction (and games)

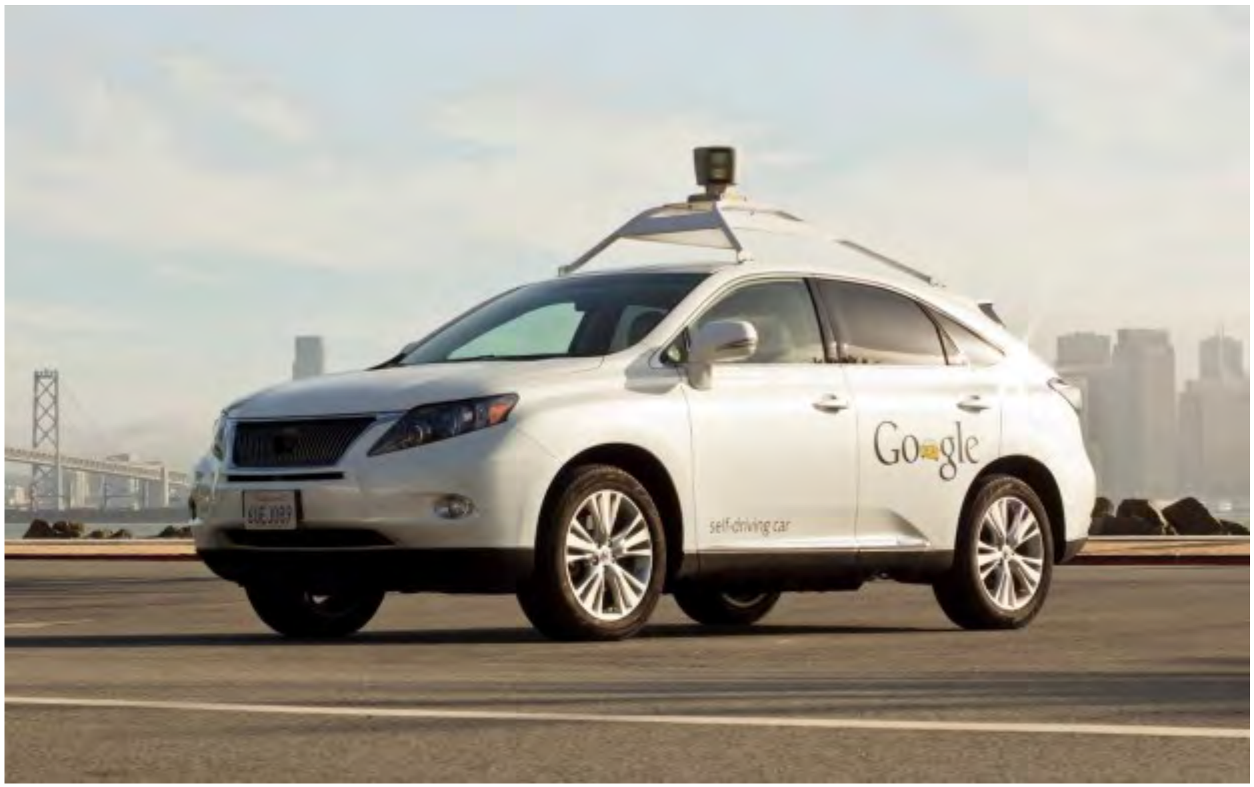
Kinect

Smart cars

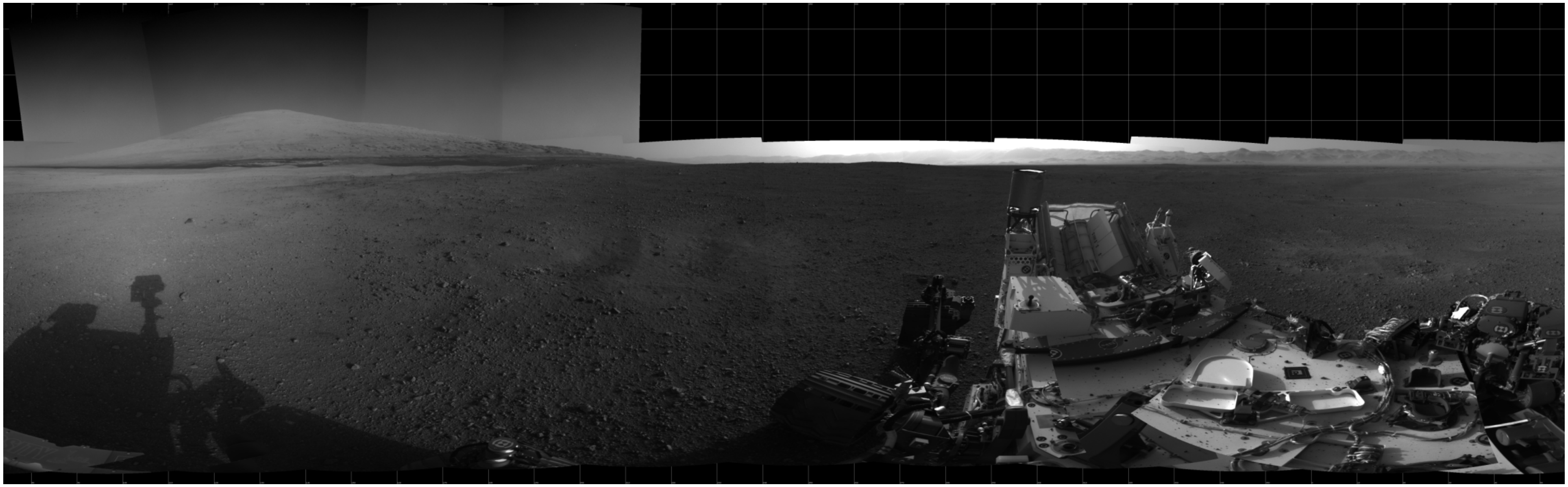
- Mobileye
  - Vision systems currently in high-end cars

Sources: A. Shashua, S. Seitz

Smart cars



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

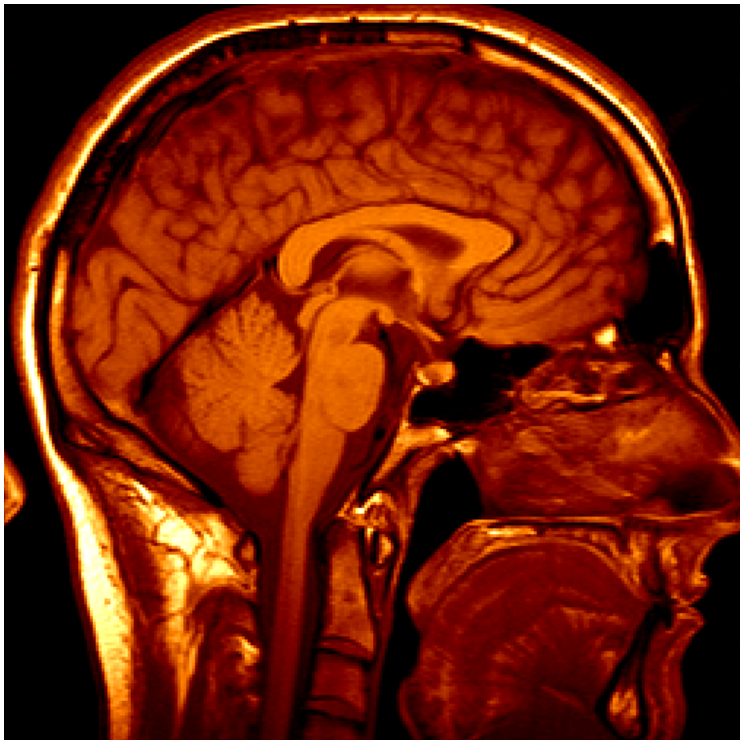
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

Robotics

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/

Medical imaging



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

Source: S. Seitz

Large-scale 3D reconstruction

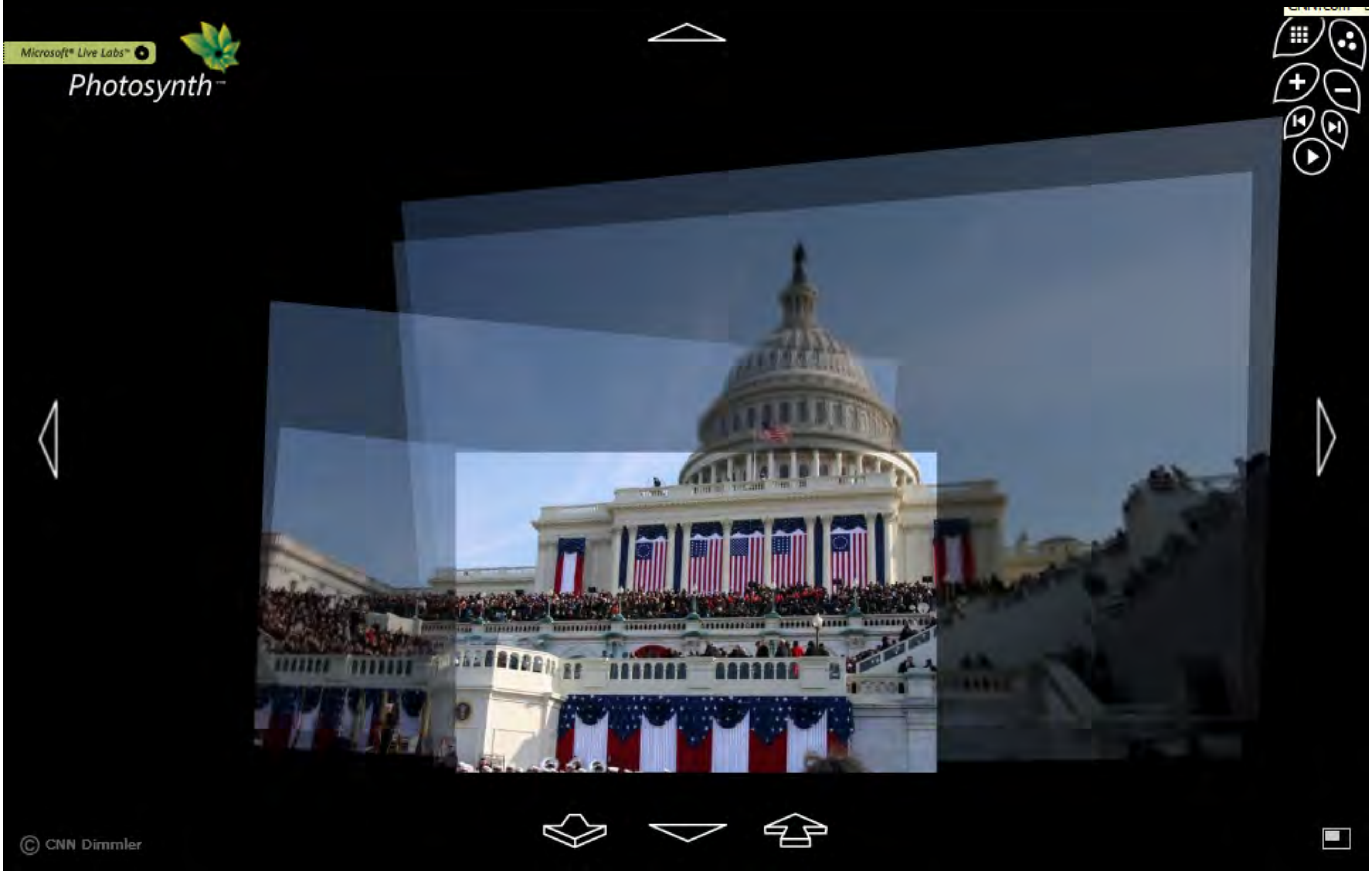
- Automatic 3D reconstruction from Internet photo collections

"Statue of Liberty" "Half Dome, Yosemite" "Colosseum, Rome"

Flickr photos

3D model

Photosynth



Current state of the art

- You just saw examples of current systems.
  - Many of these are less than 5 years old
- This is a very active research area, and rapidly changing
  - Many new apps in the next 5 years
- To learn more about vision applications and companies
  - David Lowe maintains an excellent overview of vision companies
    - http://www.cs.ubc.ca/spider/lowe/vision.html

Why is computer vision difficult?

Viewpoint variation

Illumination

Scale

Why is computer vision difficult?

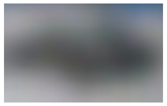
Intra-class variation

Motion (Source: S. Lazebnik)

Background clutter

Occlusion

Challenges: local ambiguity



slide credit: Fei-Fei, Fergus & Torralba

But there are lots of cues we can exploit…

Source: S. Lazebnik

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

Andrej Karpathy blog

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.

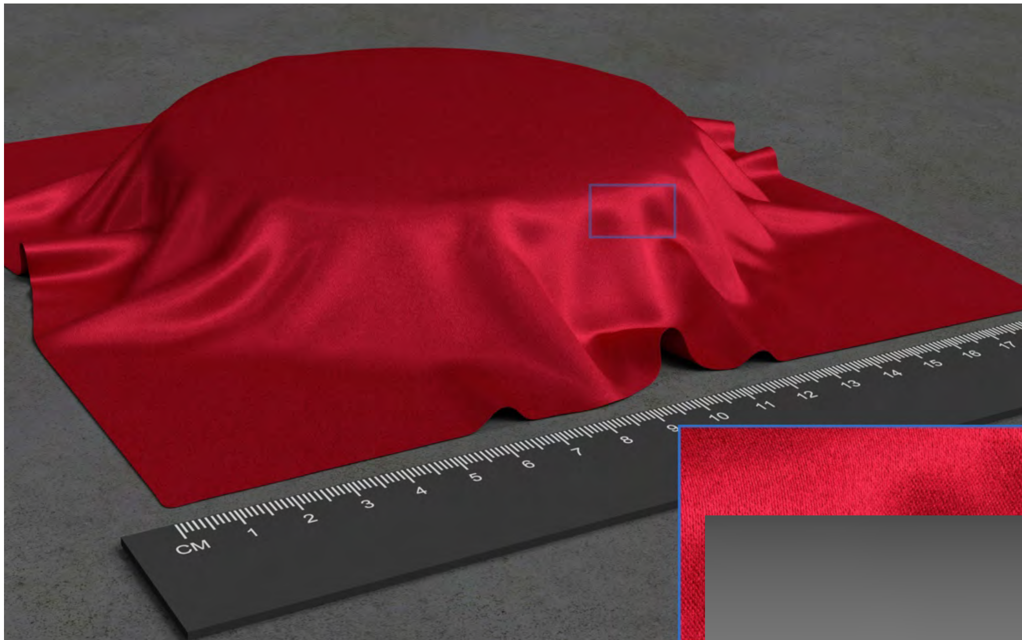
Instructor

- Kavita Bala (kb@cs.cornell.edu)
- Office hours:

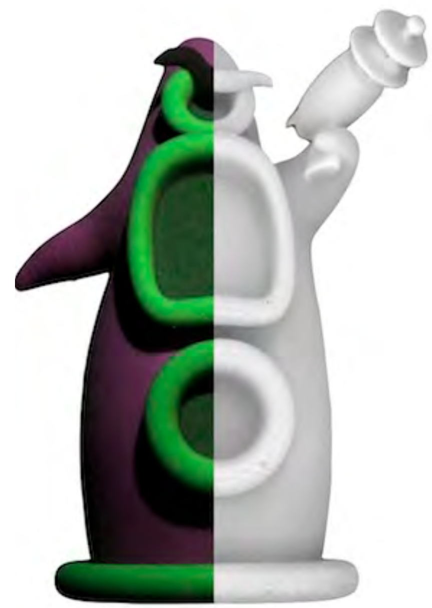
  F: 10-11, or by appointment

- Research interests:
  - Computer graphics and vision
  - Rendering, Perception, Material modeling and recognition

Autodesk 360 Cloud Render


CM 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17

Current Interests

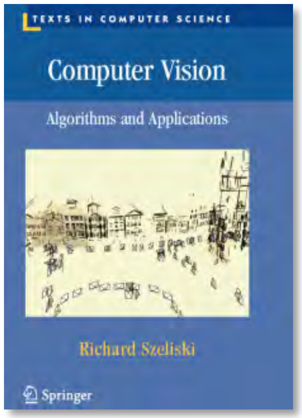


- Material Perception
- Recognition
  - Visual Search
- Scene Reconstruction

Important personnel

- TAs:
  - Senior TAs
    - Scott Wehrwein, Sean Bell
  - PhD TAs
    - Balazs Kovacs, Andreas Veit
  - Meng TA
    - Alec Regulinski
  - Undergrad TAs
    - Akhila Ananthram, Daniel Carpenter, Sheroze Sheriffdeen, Dhruv Singhal, Raghav Subramaniam
- Office hours TBA

Other administrative details



- Textbook:

  Rick Szeliski, *Computer Vision: Algorithms and Applications*

  online at: http://szeliski.org/Book/

- Course webpage (lectures, assignments):

  http://www.cs.cornell.edu/courses/cs4670/2015sp/

- Announcements/grades via Piazza/CMS

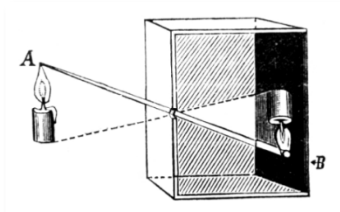
  Sign up on piazza

  https://cms.csuglab.cornell.edu/

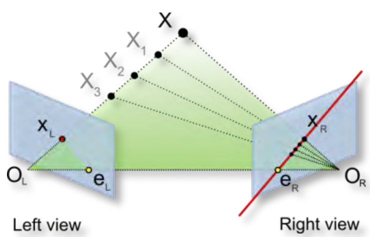
Course requirements

- Prerequisites—*these are essential*!
  - Data structures
  - A good working knowledge of C/C++ programming
  - Linear algebra
  - Vector calculus
- Course does ***not*** assume prior imaging experience
  - computer vision, image processing, graphics, etc.

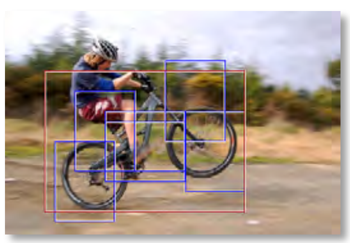
Course overview (tentative)



1. Low-level vision
   – image processing, edge detection, feature detection, cameras, image formation



2. Geometry and algorithms
   – projective geometry, stereo, structure from motion, Markov random fields

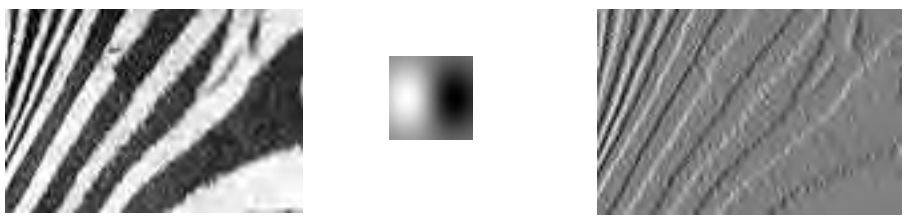


3. Recognition
   – detection / recognition, category recognition, segmentation

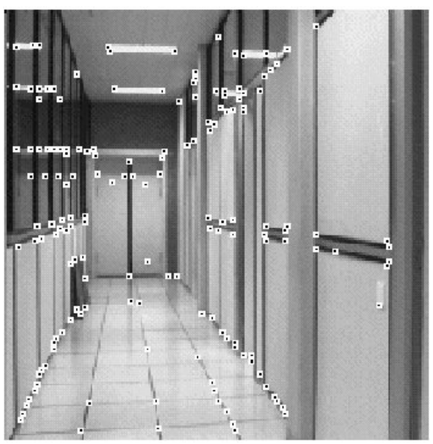
4. Light, color, and reflectance

1. Low-level vision

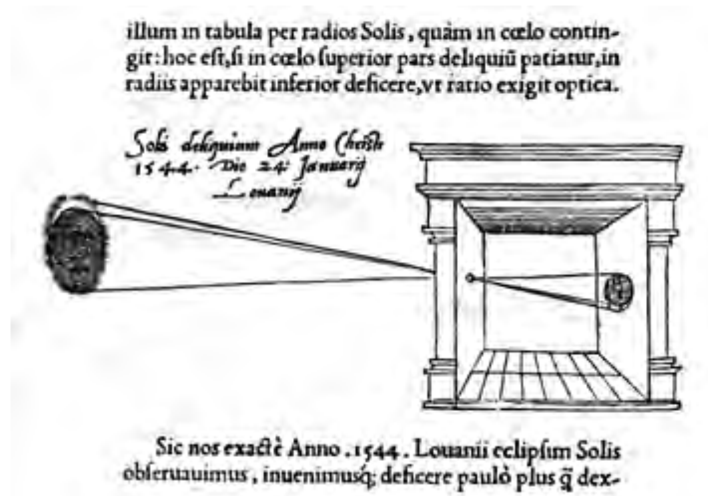
- Basic image processing and image formation



Filtering, edge detection



Feature extraction



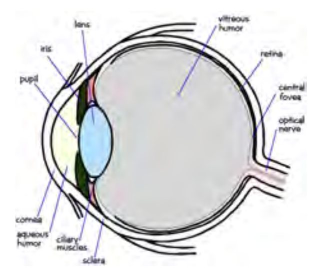
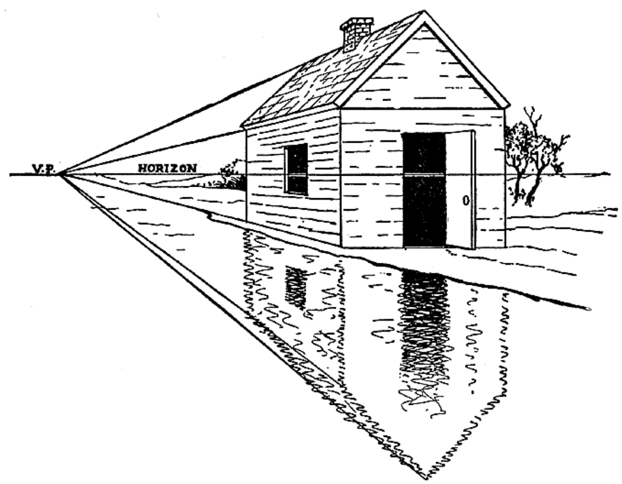


Image formation

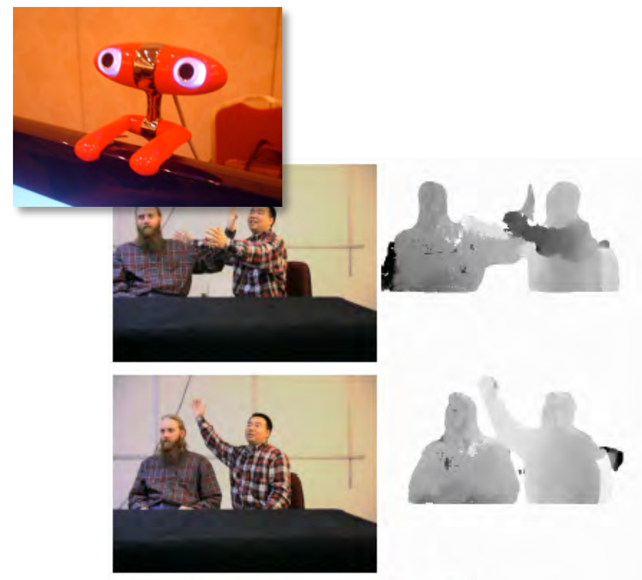
Project: Image Scissors

Project: Feature detection and matching

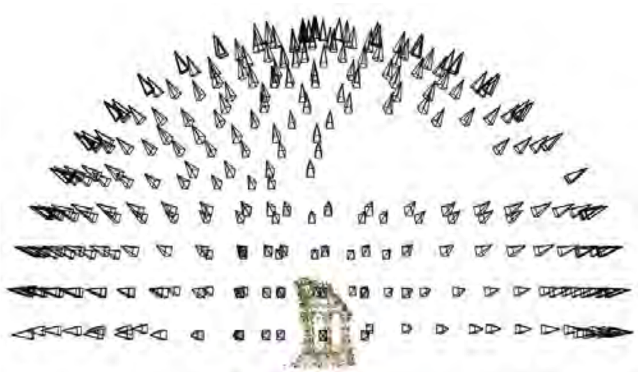
2. Geometry



Projective geometry



Stereo



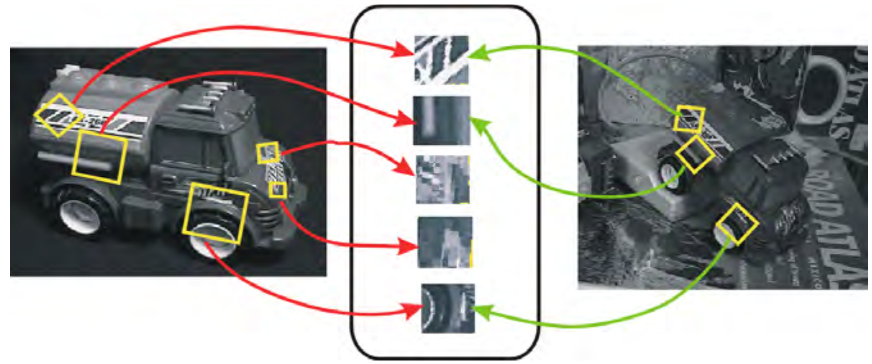
Multi-view stereo

Structure from motion

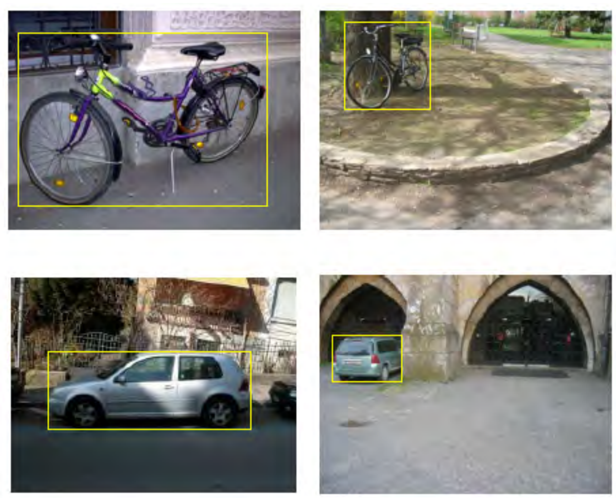
Project: Creating panoramas

3. Recognition

Face detection and recognition



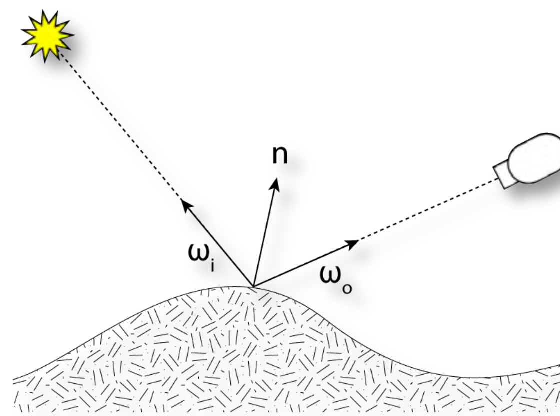
Single instance recognition



Category recognition

4. Light, color, and reflectance

Light & Color



Reflectance

Grading

- Occasional quizzes (at the beginning of class)
- One prelim, one final exam
- Rough grade breakdown:
  - Quizzes: 2-5%
  - Midterm: 15-20%
  - Homeworks: 10-20%
  - Programming projects: 40-50%
  - Final exam: 15-20%

Late policy

- Three “late days” will be available for the semester
- Late projects will be penalized by 25% for each day it is late, and no extra credit will be awarded.

Academic Integrity

Questions?