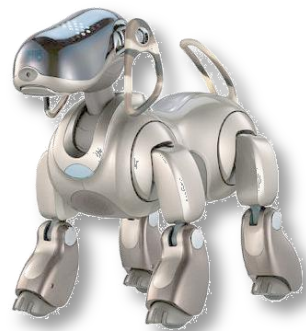
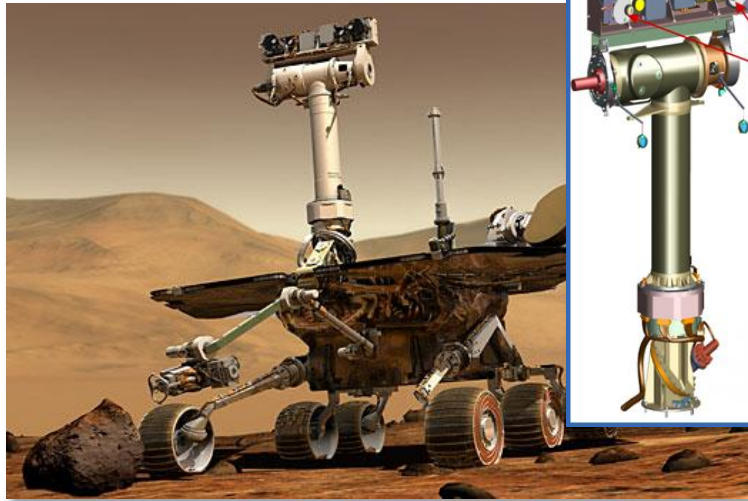


# CS4670: Intro to Computer Vision

Noah Snaveley



# Instructor

- Noah Snavely ([snavely@cs.cornell.edu](mailto:snavely@cs.cornell.edu))
- Office hours:
  - Wednesdays 1:30 – 3pm (tentative), or by appointment
- Research interests:
  - Computer vision and graphics
  - 3D reconstruction and visualization of Internet photo collections

# Important personnel

- TA: Kevin Matzen
  - Office hours TBA

# Other details

- Textbook:  
Richard Szeliski, *Computer Vision:  
Algorithms and Applications*  
online at:  
<http://szeliski.org/Book/>
- Course webpage (content coming soon):  
<http://www.cs.cornell.edu/courses/cs4670/2010fa/>
- Announcements/grades via CMS  
<https://cms.csuglab.cornell.edu/>

# Today

1. Introduction to computer vision
2. Course overview
3. Basic image processing

# Today

- Readings
  - Szeliski, CV: A&A, Ch 1.0 (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?



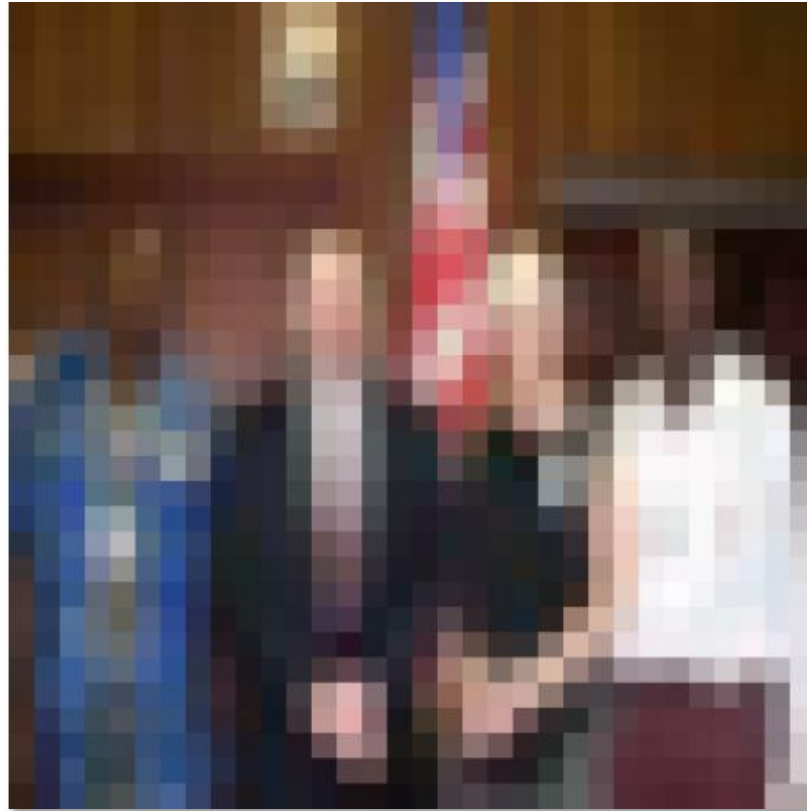
- Yes and no (but mainly no, so far)
  - computers can be better at “easy” things
  - humans are much better at “hard” things

# 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.

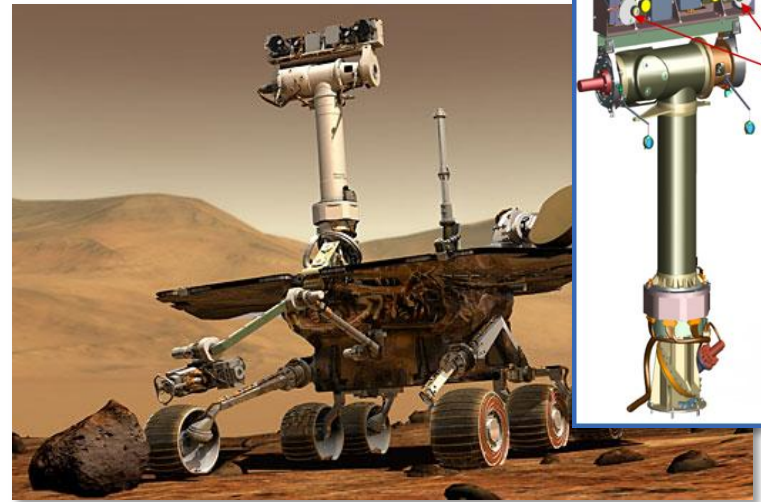
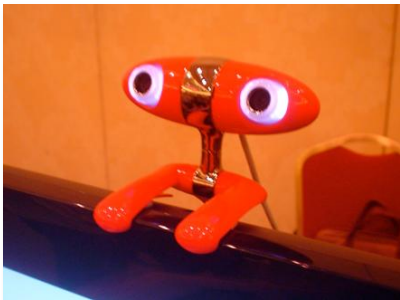


# The goal of computer vision



# The goal of computer vision

- Computing the 3D shape of the world



# The goal of computer vision

- Recognizing objects and people







sky

building

flag

face

banner

wall

street lamp

bus

bus

cars



slide credit: Fei-Fei, Fergus & Torralba

# The goal of computer vision

- “Enhancing” images





# The goal of computer vision

- “Enhancing” images



Super-resolution / denoising  
(source: 2d3)



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

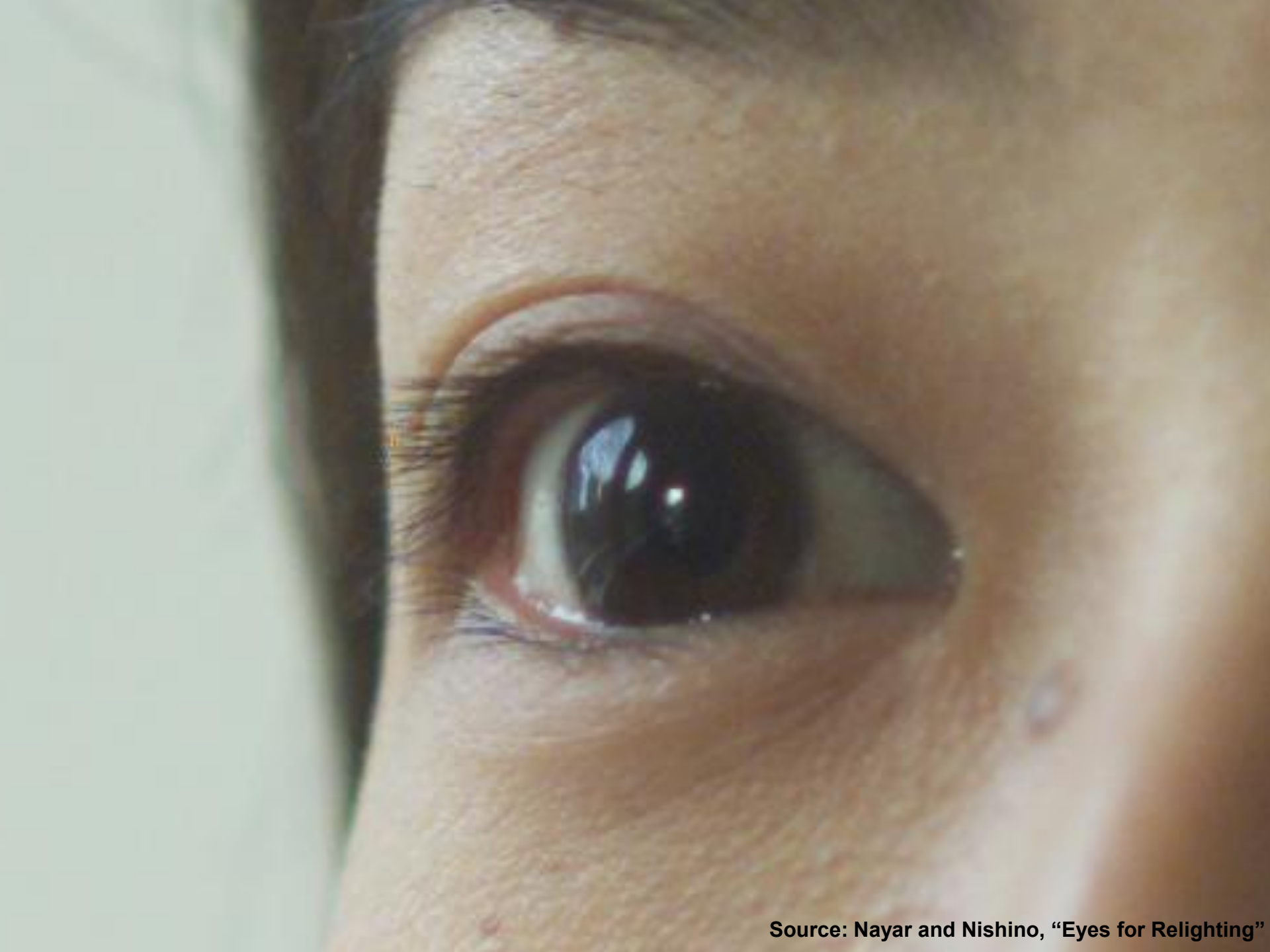


Inpainting / image completion  
(image credit: Hays and Efros)

# The goal of computer vision

- Forensics



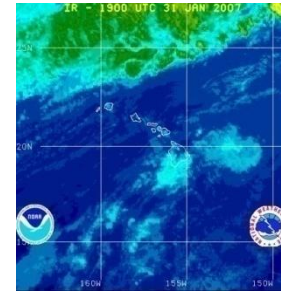
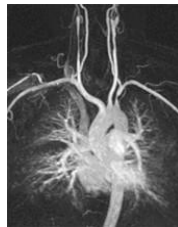


Source: Nayar and Nishino, "Eyes for Relighting"



# Why study computer vision?

- Millions of images being captured all the time



- Lots 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/>

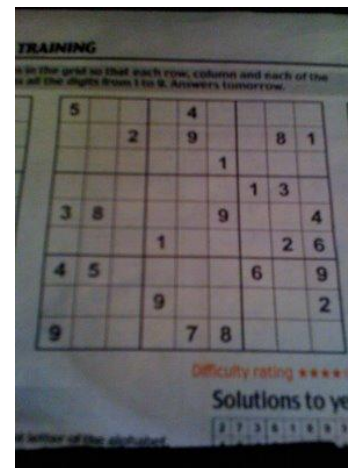


License plate readers

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



Automatic check processing



Sudoku grabber

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

Source: S. Seitz

# Face detection

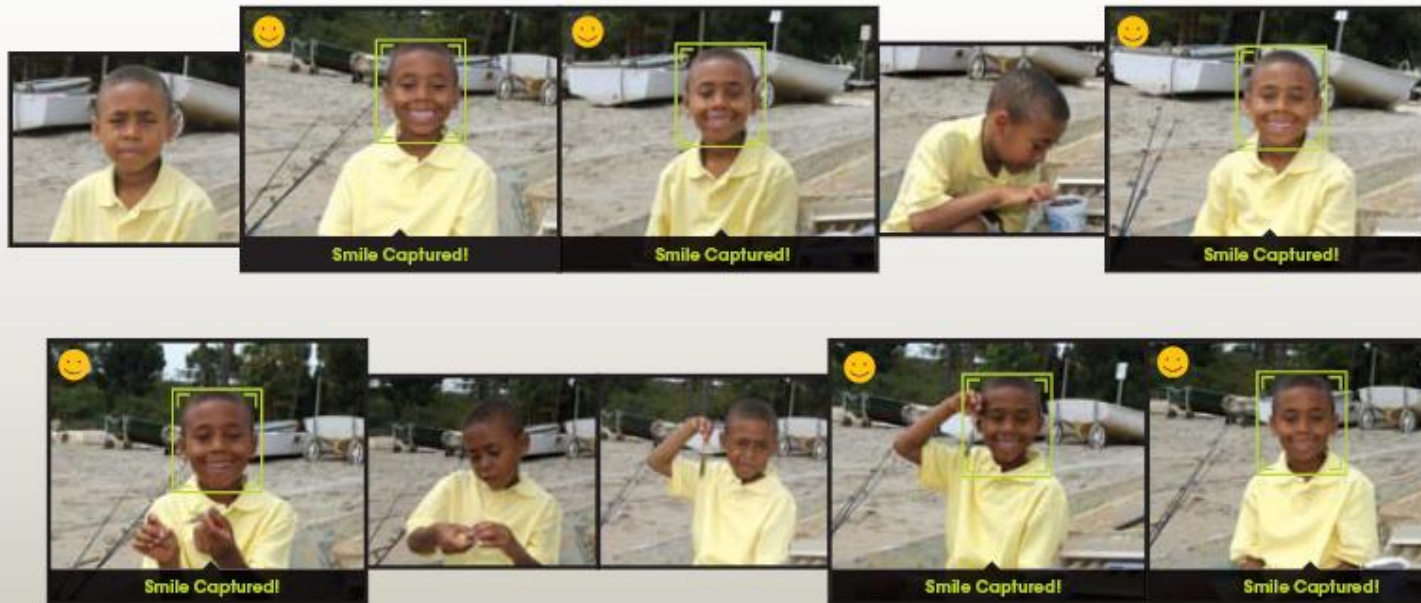


- Many new digital cameras now detect faces
  - Canon, Sony, Fuji, ...

# Smile detection?

## The Smile Shutter flow

Imagine a camera smart enough to catch every smile! In Smile Shutter Mode, your Cyber-shot® camera can automatically trip the shutter at just the right instant to catch the perfect expression.



[Sony Cyber-shot® T70 Digital Still Camera](#)

Source: S. Seitz

# Face recognition

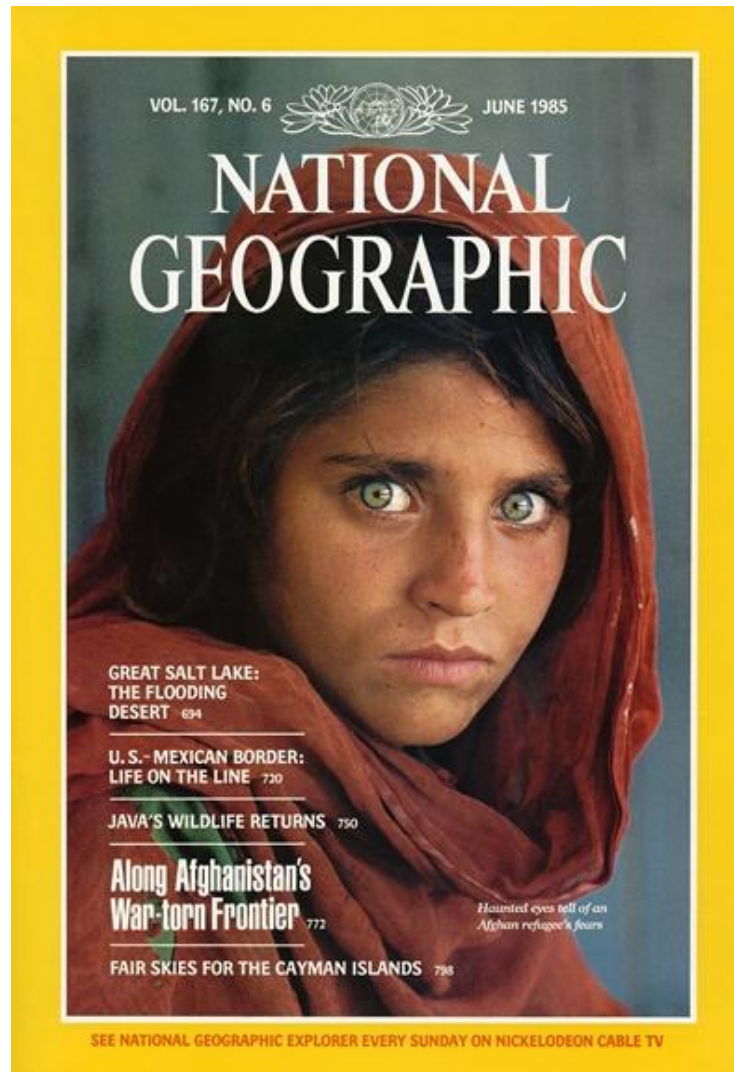
The screenshot displays the Picasa 3 application window. The title bar reads "Picasa 3" and the menu bar includes "File", "Edit", "View", "Album", "Picture", "Create", "Tools", and "Help". The status bar at the top right shows the user "jmartha@gmail.com" and links for "Web Albums" and "Sign Out".

The main interface is titled "Noah Snavelly" and shows a grid of 24 face recognition suggestions. The suggestions are arranged in a 4x6 grid. The first row contains 6 small, low-resolution thumbnails. The second and third rows each contain 6 larger, medium-resolution images. The fourth row contains 6 images, including one where Noah is wearing a black beanie. In the top right corner of the suggestion area, there is a "5 suggestions" label and two buttons: "Confirm all" and "Remove".

On the left side, there is a sidebar with a "People" section containing 28 entries. The entries are: "Unnamed (5,449)", "Me (419)", "Emily Snavelly (16)", "Haitao (2)", "Jeff Stensrud (6)", "Jeremy Snavelly (2)", "Jialin Xie (14)", "John Snavelly (2)", "Li Getao (55)", "Lijun (3)", "Patti Wylie (2)", "Phyllis Orient (3)", "Poppa (2)", and "Pravin Bhat (3)". Below the "People" section is a "Folders (120)" section.

At the bottom of the Picasa window, there is a toolbar with icons for "Selection", "Upload", "Email", "Print", "Export", "Shop", "BlogThis!", "Collage", "Movie", and "Geo-Tag". The status bar at the very bottom shows the file name "IMG\_1873.JPG", the date and time "3/26/2007 1:38:51 PM", the resolution "406x485 pixels", and the file size "3.8MB". The Windows taskbar at the bottom of the screen shows various open applications and the system clock at "10:35 PM".

# Face recognition



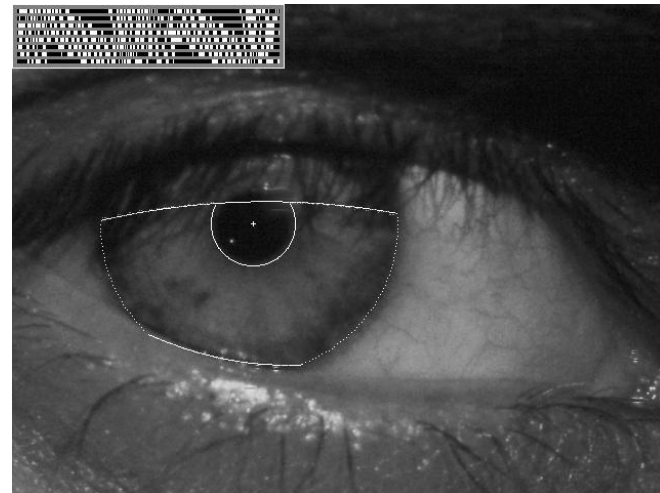
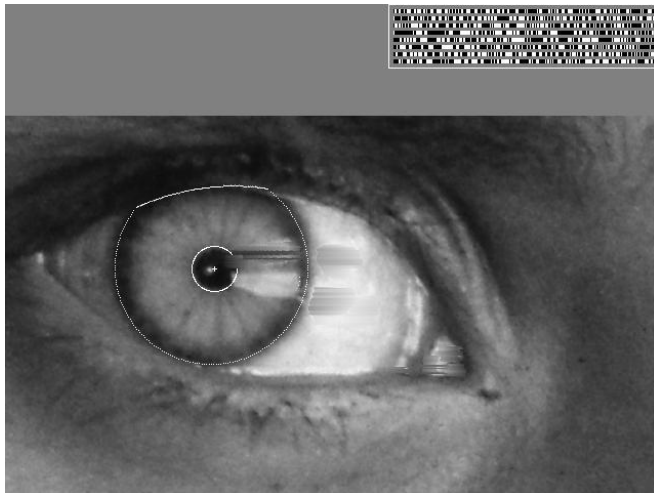
Who is she?

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  
<http://www.sensiblevision.com/>

# 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... “



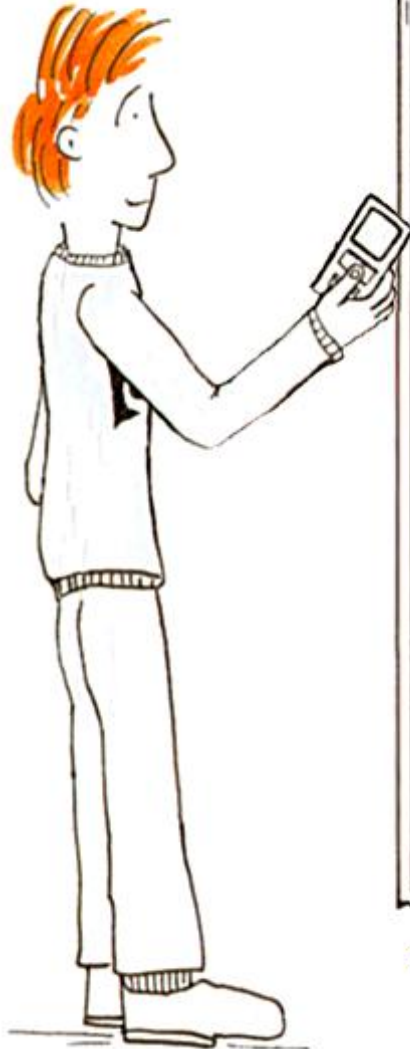
# Object recognition (in mobile phones)



- This is becoming real:
  - **Lincoln** Microsoft Research
  - [Point & Find](#)

# iPhone Apps: kooaba (www.kooaba.com)

MOBILE IMAGE RECOGNITION?  
TRY IT OUT NOW!!!



[Show another poster](#)

Movie data provided by:



1. **POINT**  
YOUR MOBILE  
PHONE CAMERA TO  
THE MOVIE  
POSTER.

2. **SNAP** A  
PICTURE AND SEND  
IT:

IN SWITZERLAND:  
MMS TO 5555 (OR  
079 394 57 00  
FOR ORANGE  
CUSTOMERS)

IN GERMANY:  
MMS TO 84000

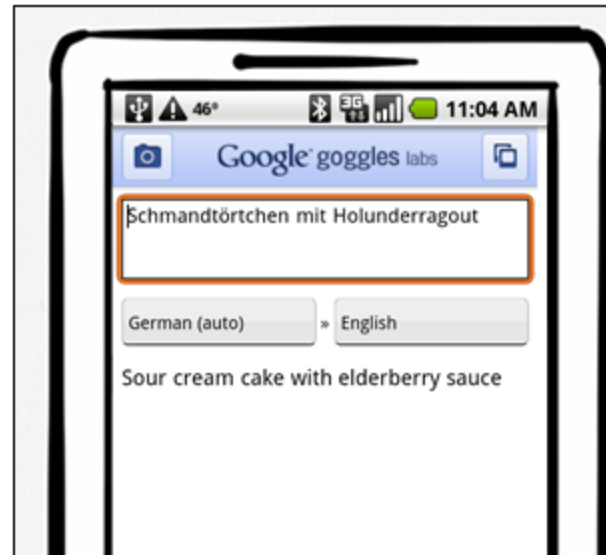
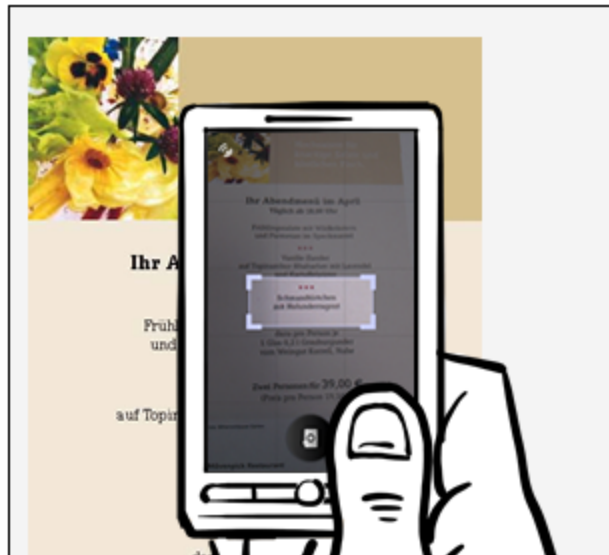
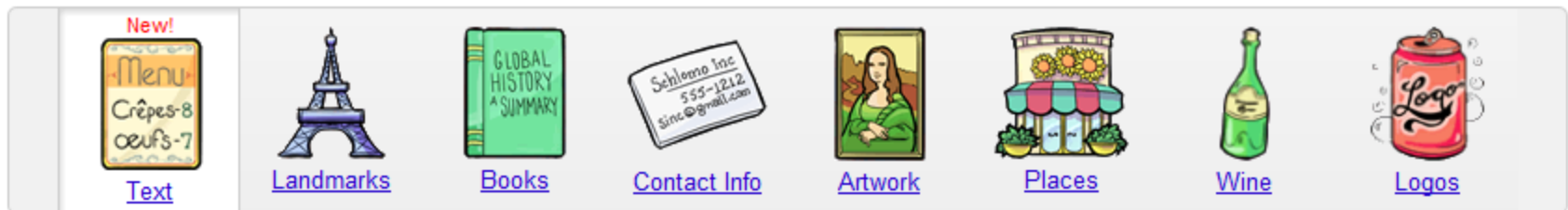
EVERYWHERE:  
EMAIL TO  
M@KOOABA.COM

3. **FIND** ALL  
RELEVANT INFOR-  
MATION ABOUT THE  
MOVIE ON YOUR  
MOBILE PHONE

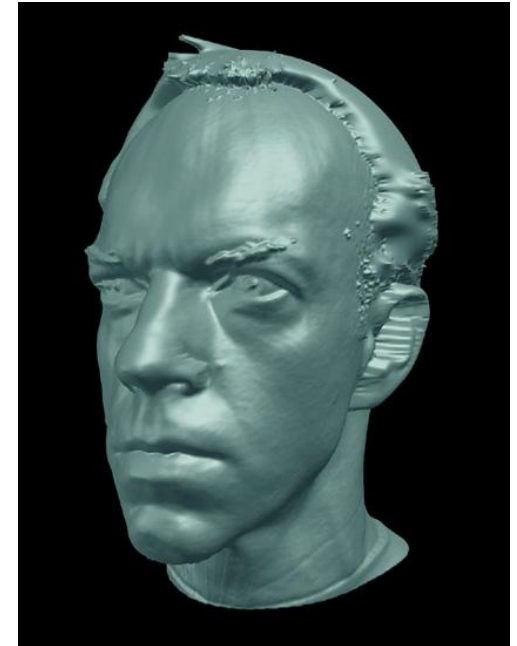
# Google Goggles

## Google Goggles in action

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



# Special effects: shape capture



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

# Special effects: motion capture



*Pirates of the Caribbean*, Industrial Light and Magic

Source: S. Seitz

# Special effects: camera tracking



Boujou, 2d3

# Sports



*Sportvision* first down line  
Nice [explanation](http://www.howstuffworks.com) on [www.howstuffworks.com](http://www.howstuffworks.com)

# Smart cars

The screenshot displays the Mobileye website interface. At the top, there are navigation tabs for 'manufacturer products' and 'consumer products'. The main headline reads 'Our Vision. Your Safety.' Below this, a top-down view of a car is shown with three camera fields of view: 'rear looking camera', 'side looking camera', and 'forward looking camera'. The bottom section features three product highlights: 'EyeQ Vision on a Chip' with an image of the chip, 'Vision Applications' showing a pedestrian detection box, and 'AWS Advance Warning System' with a circular display showing a car icon and a distance of 0.8. On the right side, there are sections for 'News' and 'Events', each with a list of recent articles and a 'read more' link.

- [Mobileye](#)

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

- By 2010: 70% of car manufacturers.

Sources: A. Shashua, S. Seitz



# Vision-based interaction (and games)



Nintendo Wii has camera-based IR tracking built in. See [Lee's work at CMU](#) on clever tricks on using it to create a [multi-touch display](#)!



Sony EyeToy



Assistive technologies



Xbox Kinect ("Project Natal")

# Vision in space

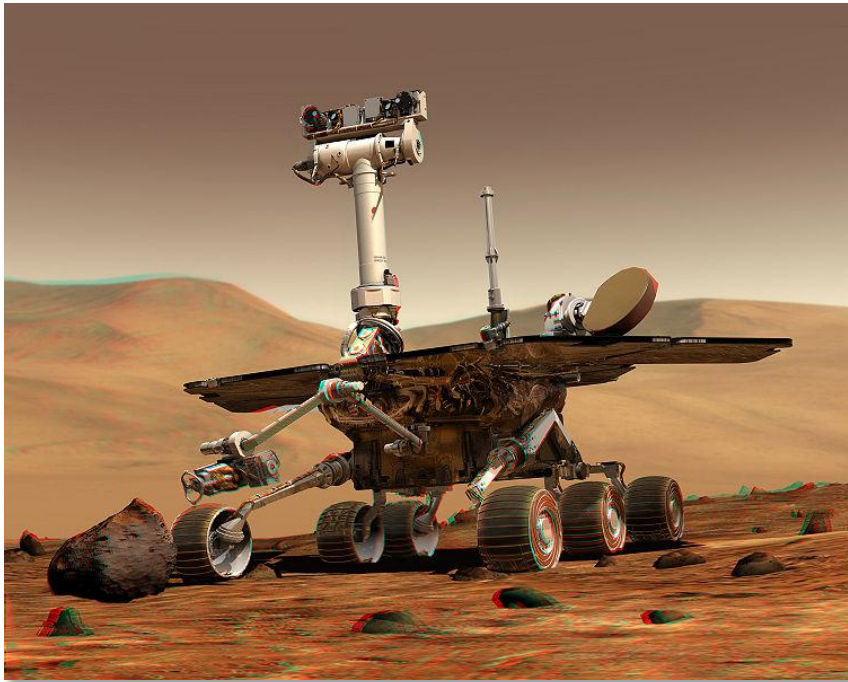


[NASA'S Mars Exploration Rover Spirit](#) captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

## 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.

# Robotics



NASA's Mars Spirit Rover

[http://en.wikipedia.org/wiki/Spirit\\_rover](http://en.wikipedia.org/wiki/Spirit_rover)

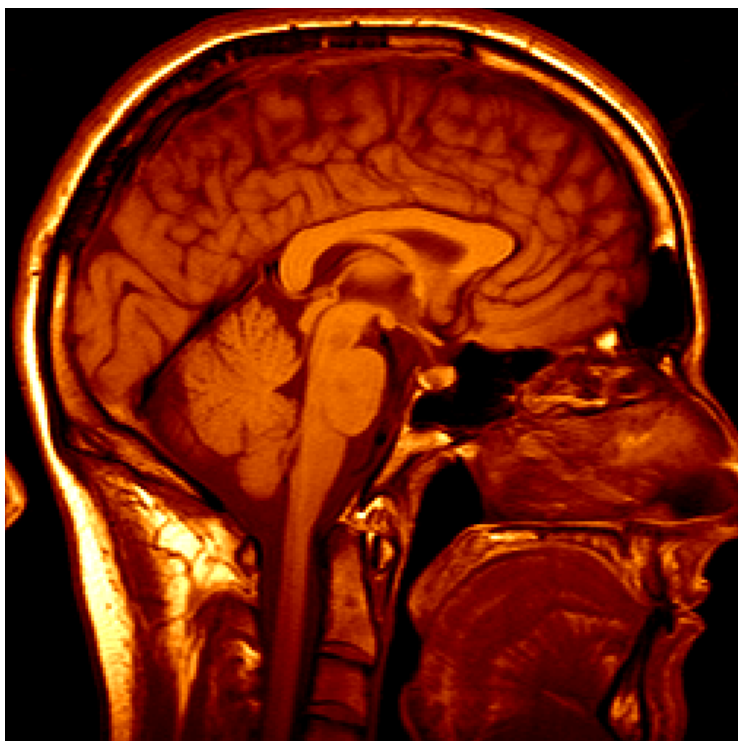


Autonomous RC Car

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

# Medical imaging

---



3D imaging  
MRI, CT



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

# My own work

- Automatic 3D reconstruction from Internet photo collections

“Statue of Liberty”



Flickr photos

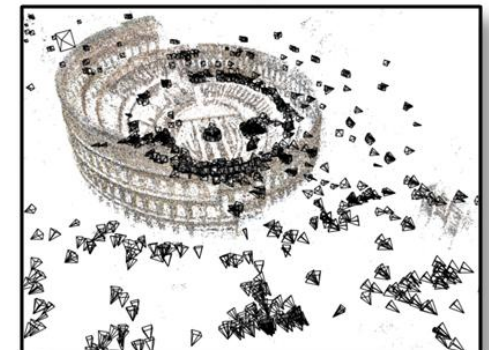
“Half Dome, Yosemite”



“Colosseum, Rome”



3D model



# Photosynth

Microsoft® Live Labs™



Photosynth™



# City-scale reconstruction



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

# 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](http://www.cs.ubc.ca/spider/lowe/vision.html) 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)

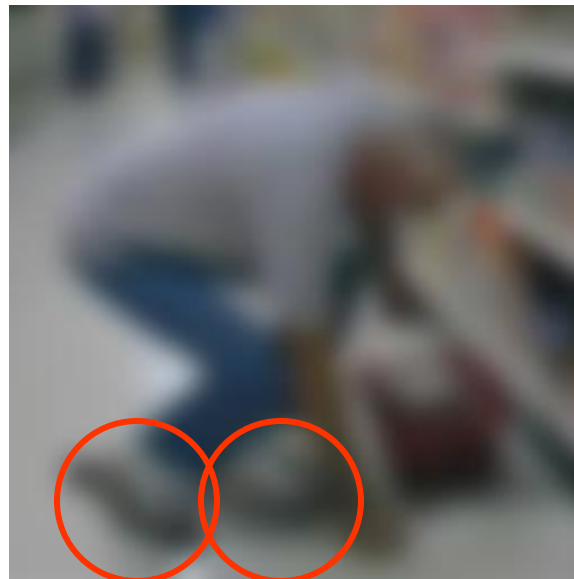
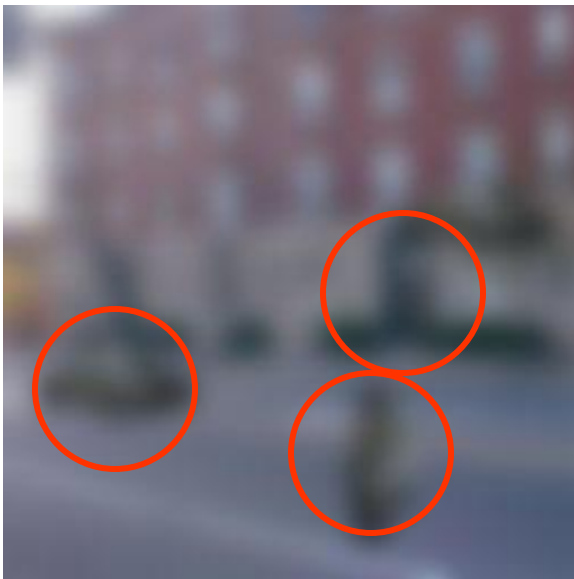
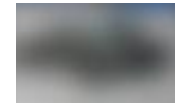
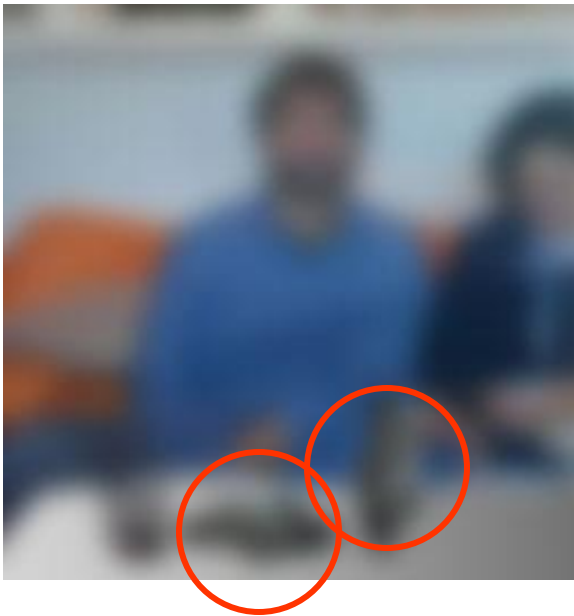


Background clutter



Occlusion

# Challenges: local ambiguity



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

# Projects

- We have 14 Nokia N900s



- Linux-based, WiFi, touch screen, 5MP camera, easy API for camera programming

# Projects

- Some of the projects will involve computer vision on a mobile device (in groups)
  - Image filtering
  - Feature detection and matching
  - Face/object recognition

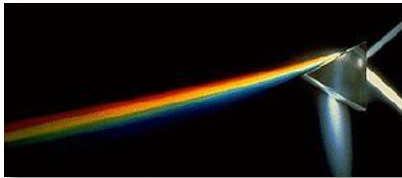
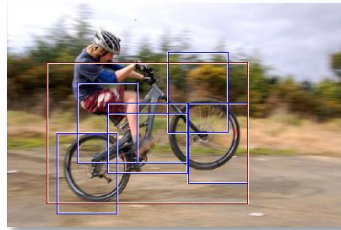
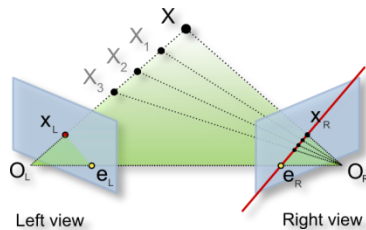
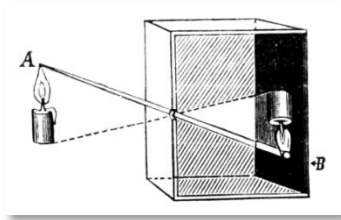


# Course requirements

- Prerequisites—*these are essential!*
  - Data structures
  - A good working knowledge of C/C++ programming
    - (or willingness/time to pick it up quickly!)
  - Linear algebra recommended
- 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

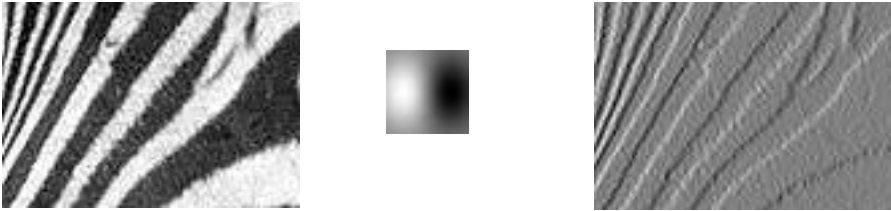
- face detection / recognition, category recognition, segmentation

## 4. Light, color, and reflectance

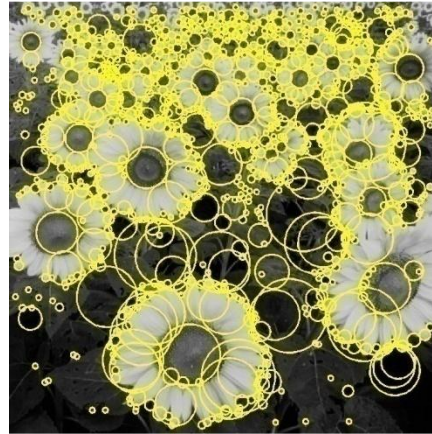
## 5. Advanced topics

# 1. Low-level vision

- Basic image processing and image formation



Filtering, edge detection



Feature extraction

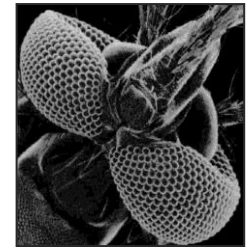
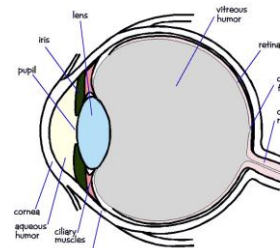
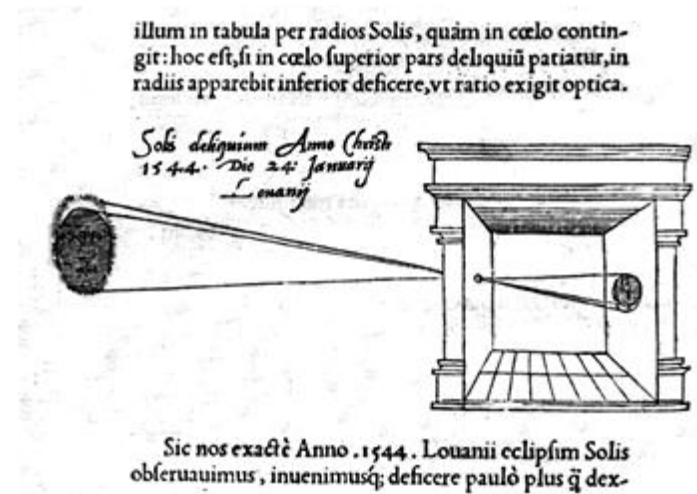
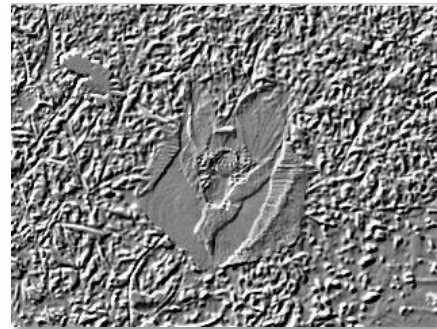
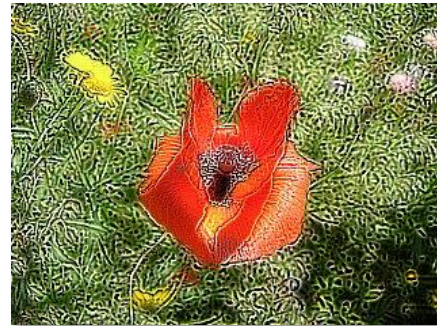
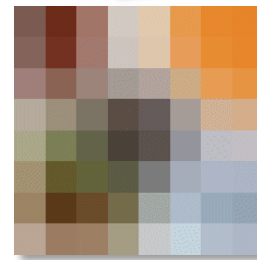
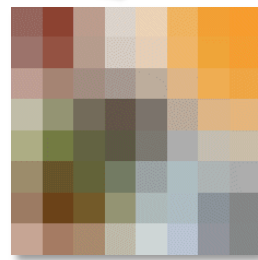
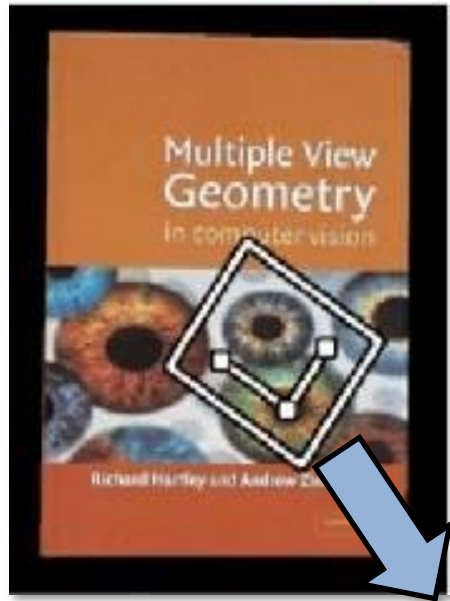


Image formation

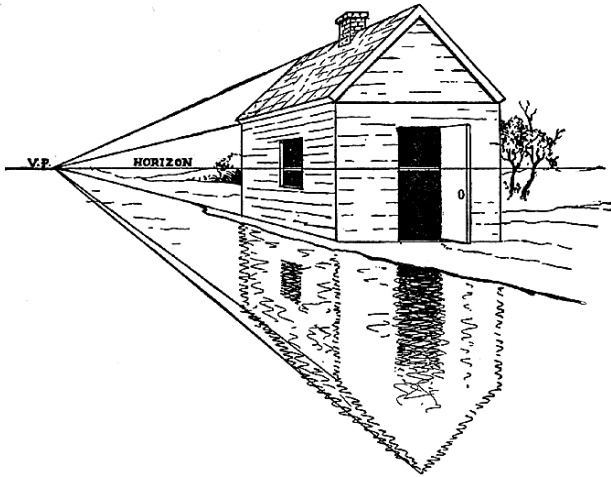
# Project: Image filtering



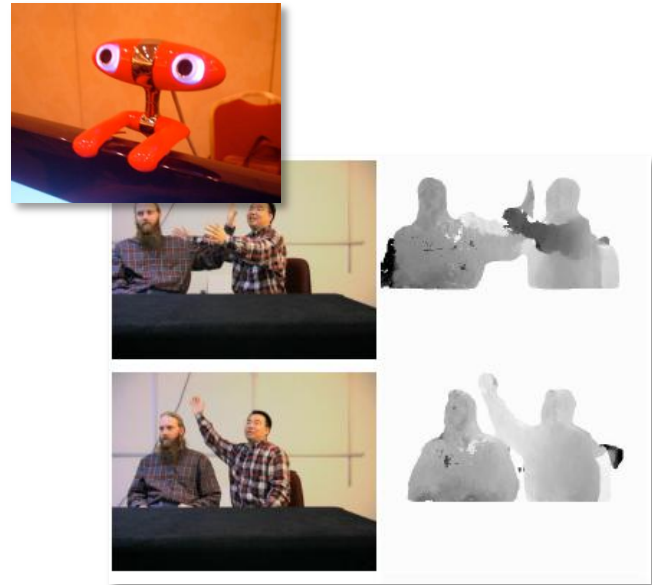
# Project: Feature detection and matching



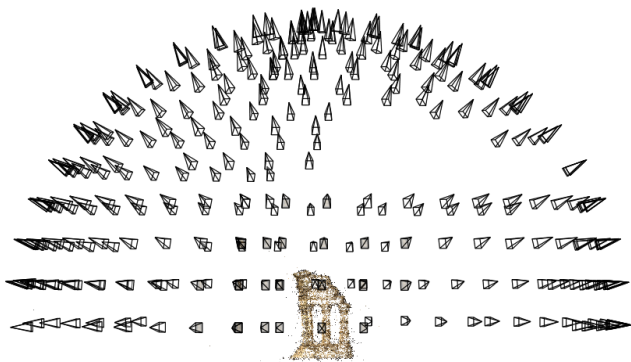
# 2. Geometry



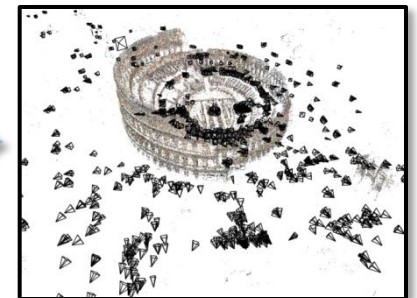
Projective geometry



Stereo



Multi-view stereo



Structure from motion

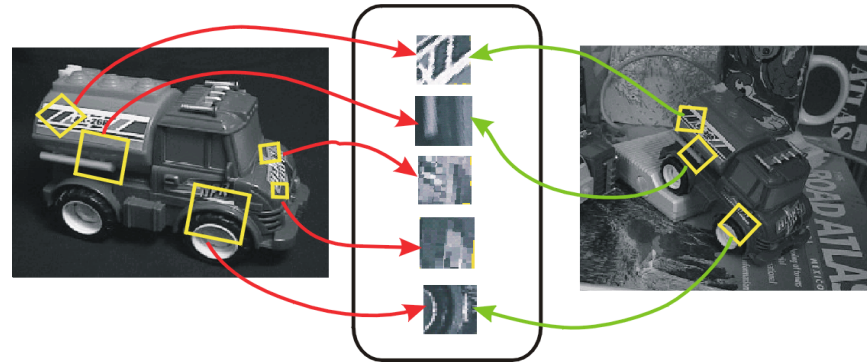
# Project: Creating panoramas



# 3. Recognition



Face detection and recognition



Single instance recognition

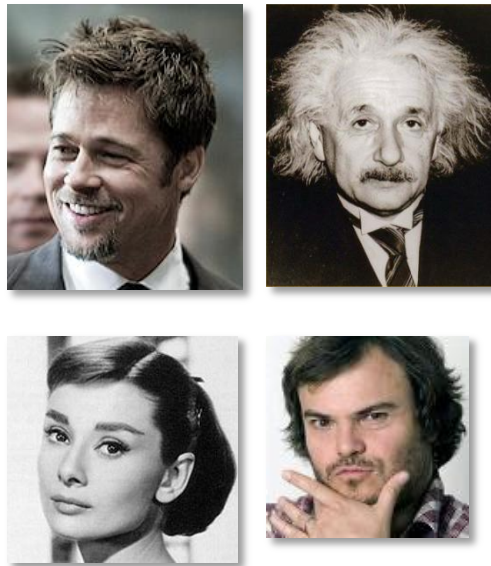


Category recognition

# Project: Recognition



Location recognition



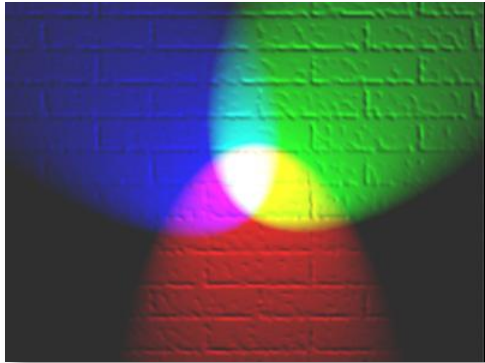
Face recognition



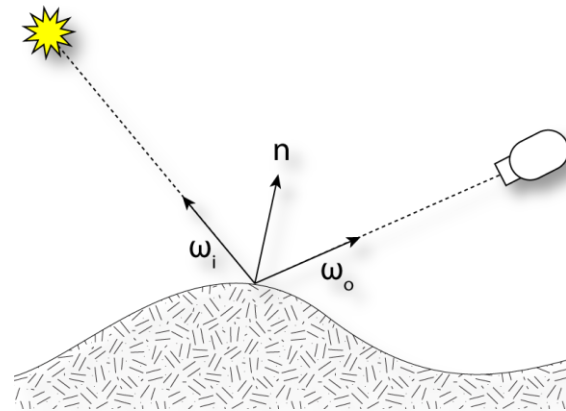
Object category recognition



# 4. Light, color, and reflectance



Light & Color



Reflectance

# 5. Advanced topics: Internet Vision



Human-aided computer vision

amazonmechanical turk  
Artificial Intelligence

## Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task → Work → Earn money

[Find HITs Now](#)



Turning the camera around

IMAGENET

3,298,242 Images, 5247 synsets indexed  
[About](#) [Explore](#) [Download](#)  
Not logged in. [Login](#) | [Signup](#)

ImageNet is an image database organized according to the **WordNet** hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.  
[Click Here](#) to learn more about ImageNet.



What do these images have in common? Find out!

Internet datasets

# Final project

- Do something interesting on a mobile device

# Grading

- Occasional quizzes (at the beginning of class)
- One prelim, possibly a final
  
- Quizzes: ~5%
- Midterm: ~15%
- Programming projects: ~ 50%
- Final project/exam: ~ 30%

Questions?