Computer Vision

CS 543 / ECE 549

University of Illinois

Instructor: Derek Hoiem

TAs: Tanmay Gupta

Bryan Plummer

Chuhang Zou

photo source

Today’s class

- A little about me and TAS
- Intro to computer vision
- Course logistics
- Questions

About me

Raised in “upstate” NY

About me

1998-2002
Undergrad at SUNY Buffalo
B.S., EE and CSE

2002-2007
Grad at Carnegie Mellon
Ph.D. in Robotics

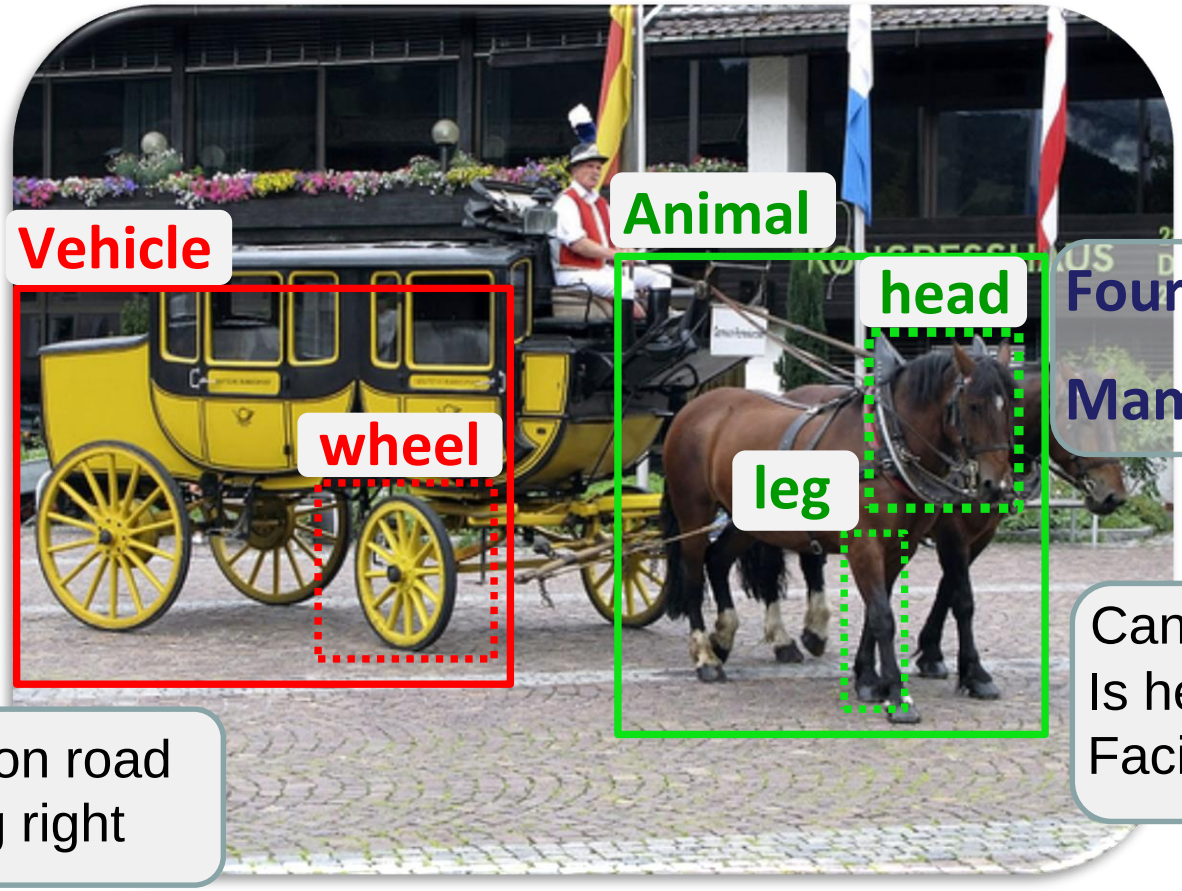
2007-2008
Postdoc at Beckman Institute

2009-
Prof in CS at UIUC

My research: single-view 3D reconstruction

with Efros, Hebert

My Research: objects and attributes



with Farhadi, Endres

My Research: full 3D shape from depth

Input Depth Image

Match to exemplar,
Predict complete shape

Estimated Shape

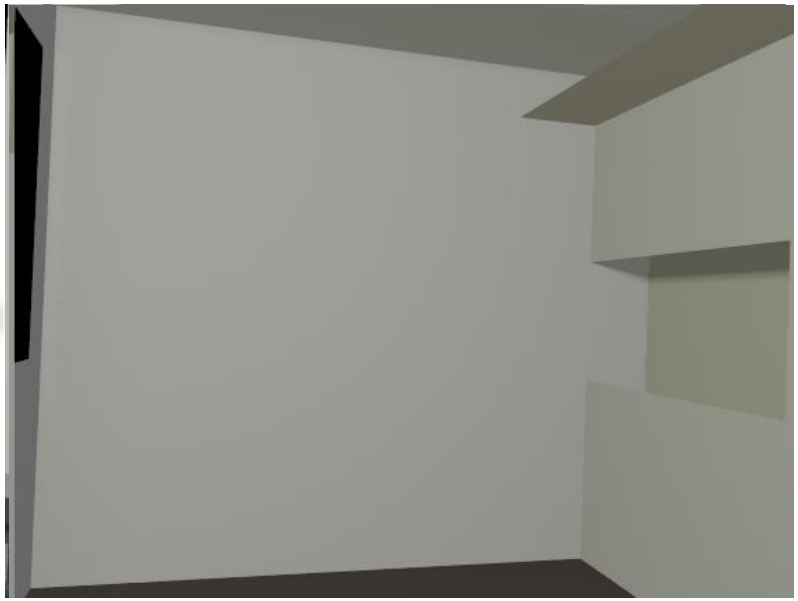
True
Shape

with Rock, Shin, et al.

My Research

3D scene model from RGB+D image

RGBD Image



3D Model

with Zou, Guo

My Research

My Research

Editing images as if they were 3D scenes

(video)

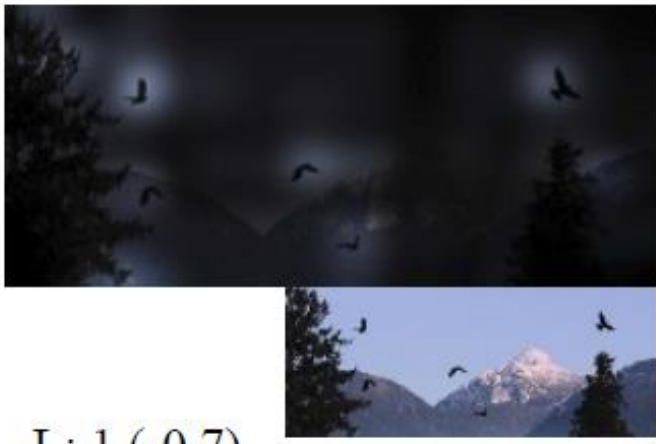
with Karsch, Hedau, Forsyth

My Research: visual question answering

What color is the street sign?

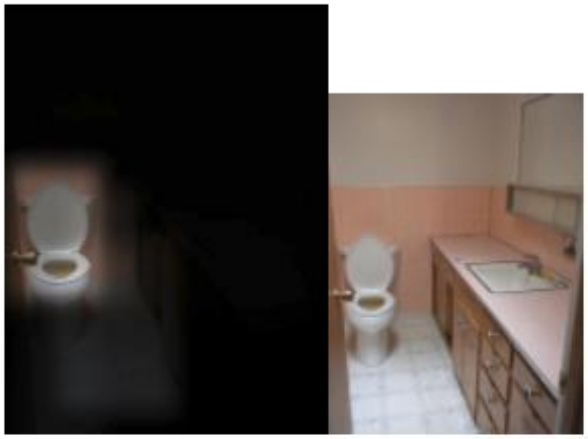
L: gray (-0.2)
I: gray (-0.4)
R: yellow (0.4)
Ans: yellow

How many birds are in the sky?



L: 1 (-0.7)
I: several (-0.1)
R: 9600 (-0.2)
Ans: 5

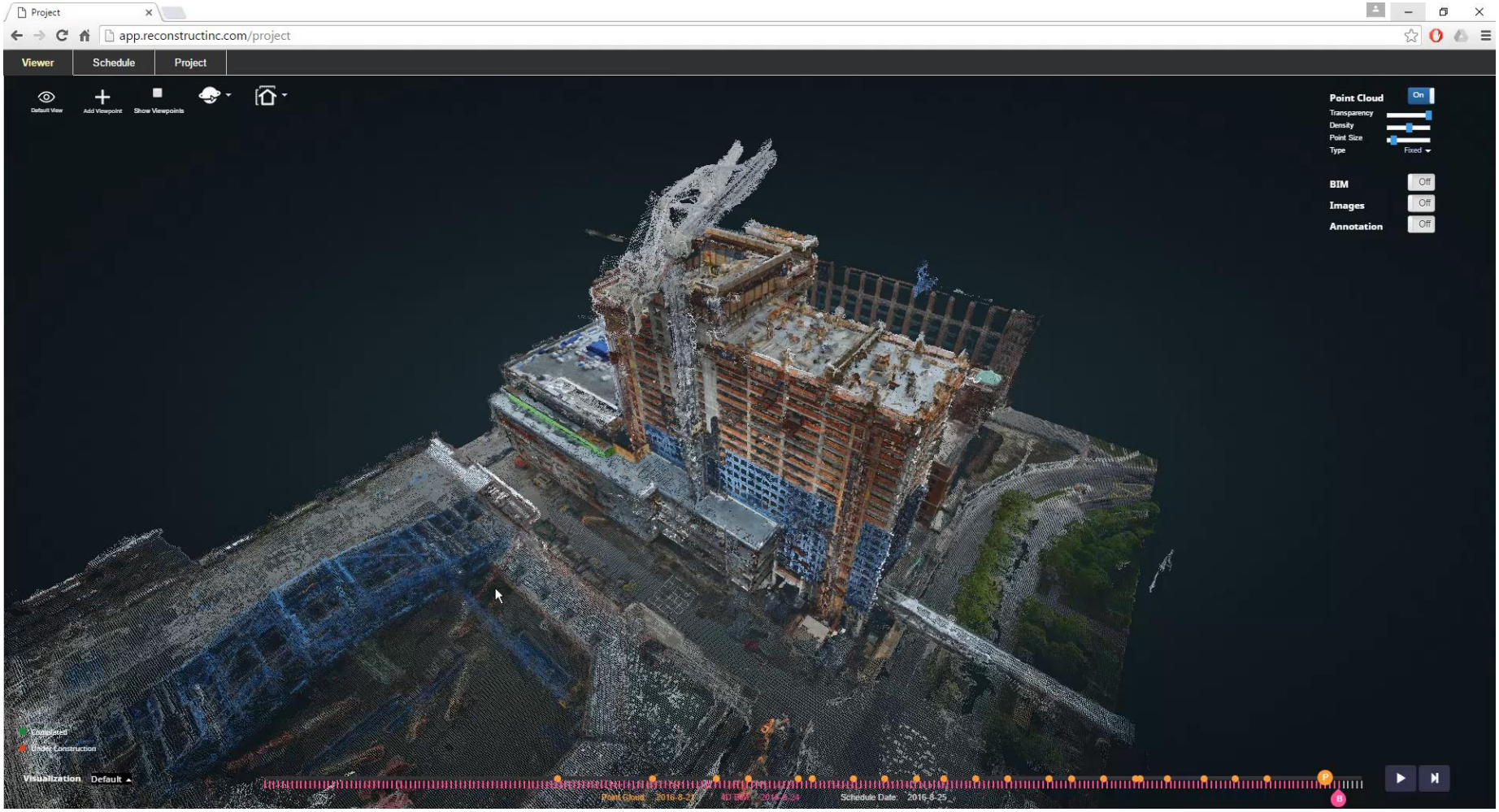
What room is this?



L: bathroom(0.1)
I: bathroom (2.6)
R: bathroom (6.8)
Ans: bathroom

with Shih, Gupta, Singh

Reconstruct, Inc: vision for construction

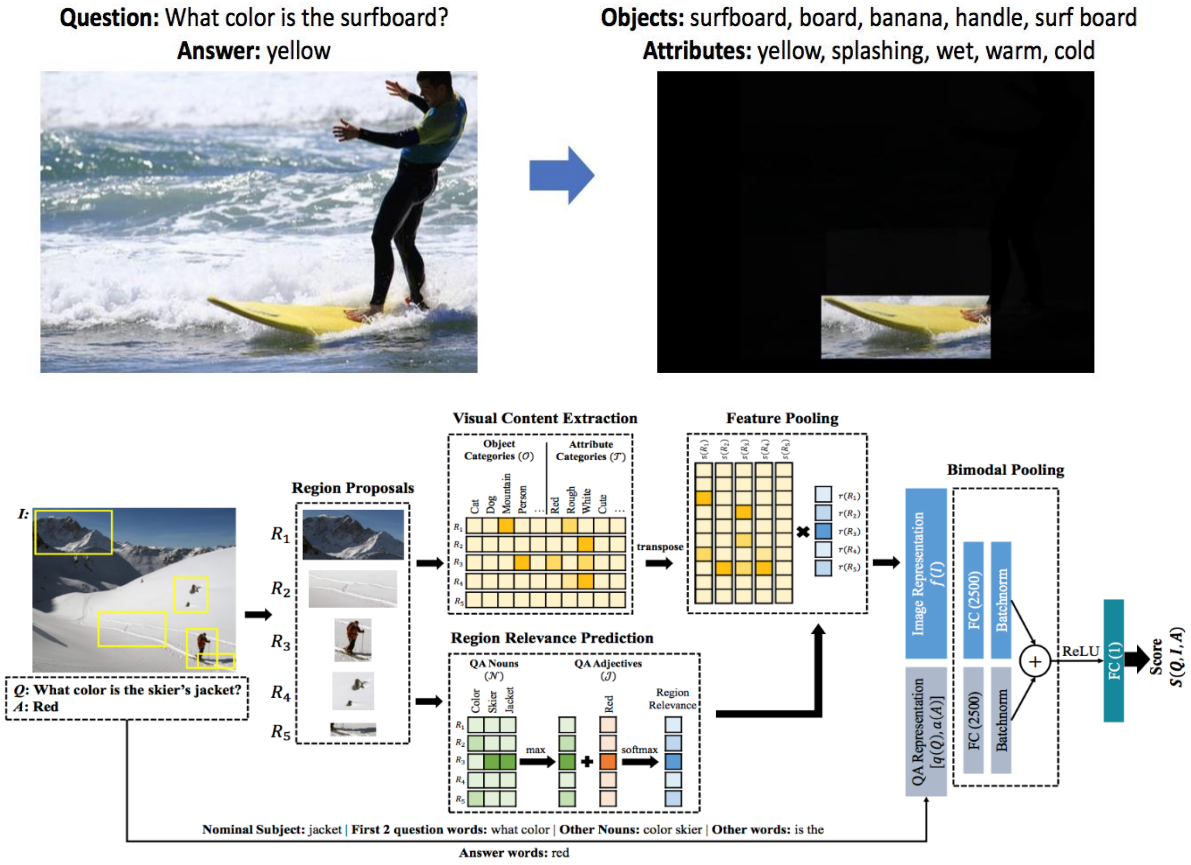


https://www.reconstructinc.com/

My daughter :)

Isla

Tanmay Gupta (TA)



Hometown

Bokaro Steel City, India

Undergrad

IIT Kanpur, India

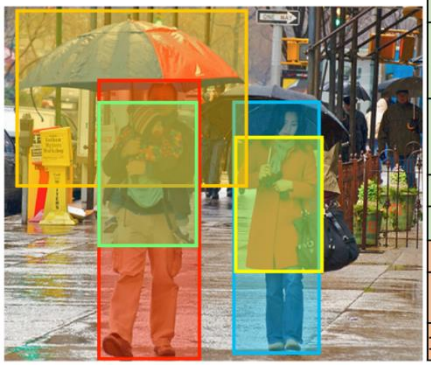
Research

Understanding the Visual World through Language

- Visual Question Answering
- Visual Relationships and their role in Vision-Language Tasks

Bryan Plummer (TA)

A man carries a baby under a red and blue umbrella next to a woman in a red jacket



Cues	Examples
1) Entities	man, baby, umbrella, woman, jacket
2) Candidate Box Position	—
3) Candidate Box Size	—
4) Common Object Detectors	man → person baby → person woman → person
5) Adjectives	umbrella → red umbrella → blue jacket → red
6) Subject - Verb	(man, carries)
7) Verb – Object	(carries, baby)
8) Verbs	(man, carries, baby)
9) Prepositions	(baby, under, umbrella) (man, next to, woman)
10)Clothing & Body Parts	(woman, in, jacket)

- Grew up in Angeles National Forest in Southern California
- Undergrad: 2 years at Mesa Community College, 2 years at UIUC
- Current Research: vision-language (language grounding, bidirectional retrieval, caption generation), video summarization

Chuhang Zou (TA)

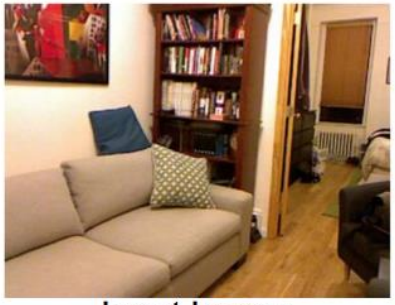
Location in China
Coordinates: 30°15′N 120°10′E

- Grew up in Hangzhou, China

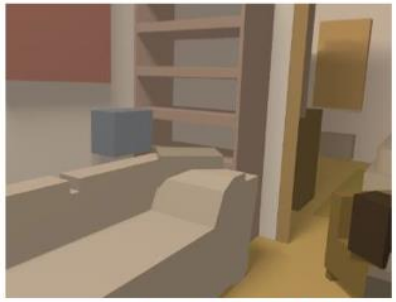
- 2010-2014 CS Undergrad at Zhejiang University, China

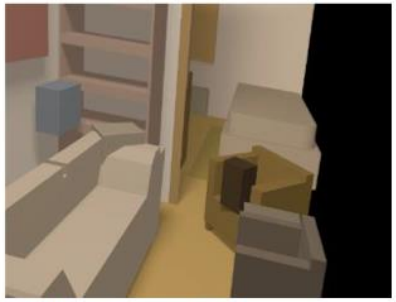
  2014- CS PhD at UIUC

- Research: 3D scene understanding, 3D shape & deep learning



Input Image





Annotated 3D Model (two views)

Computer Vision

Make computers understand images and video.

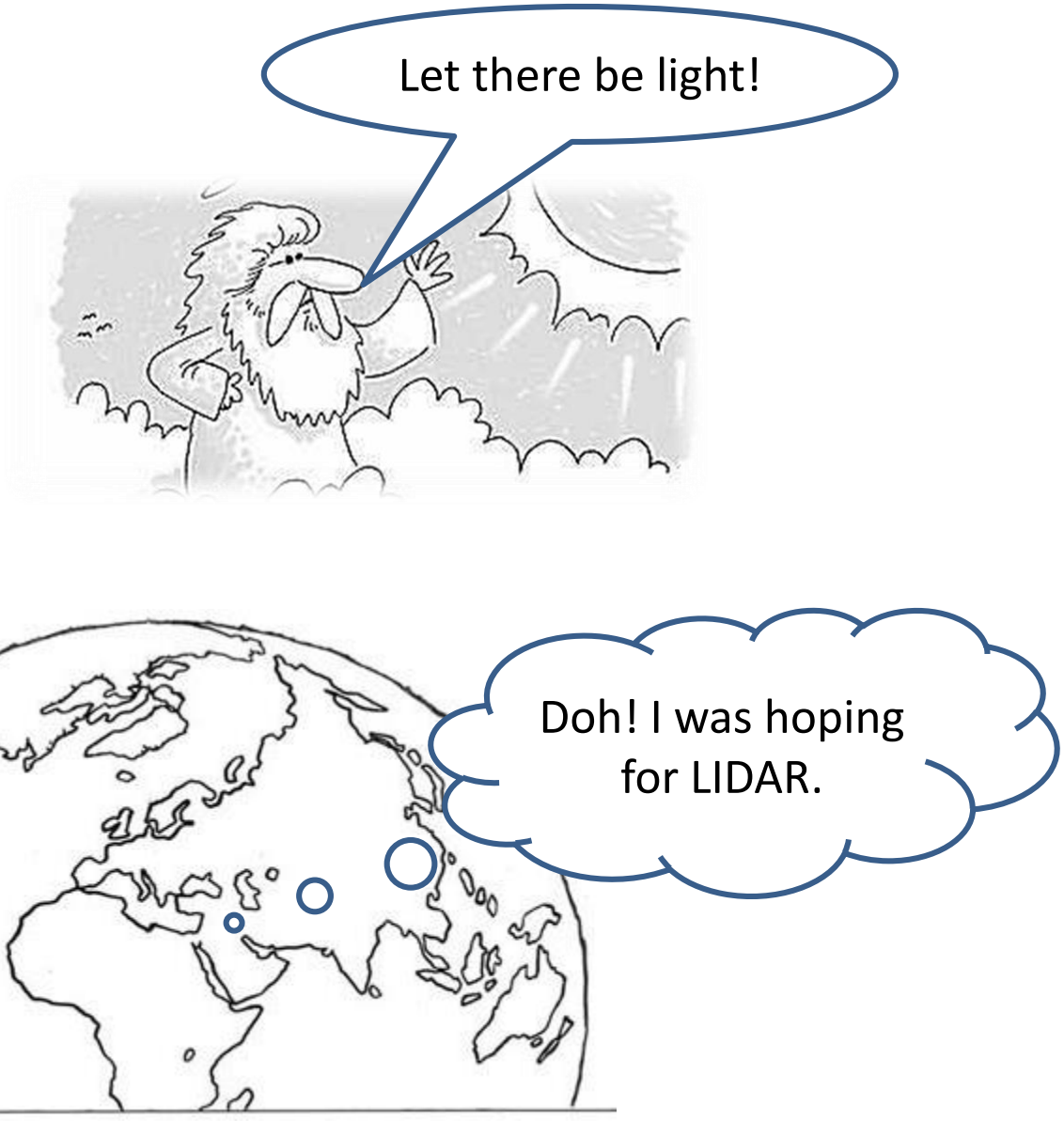
What kind of scene?

Where are the cars?

How far is the building?

…

The miracle of vision

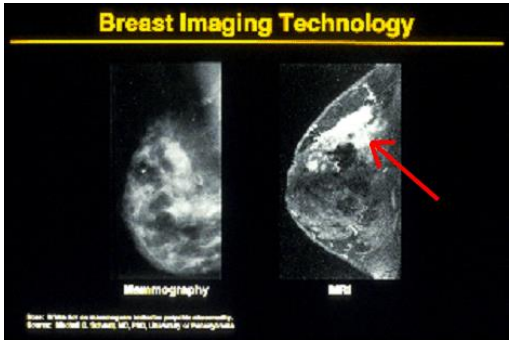


Vision is really hard

- Vision is an amazing feat of natural intelligence
  - More human brain devoted to vision than anything else

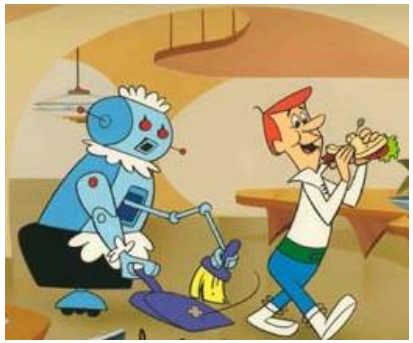
Computer vision matters

Safety



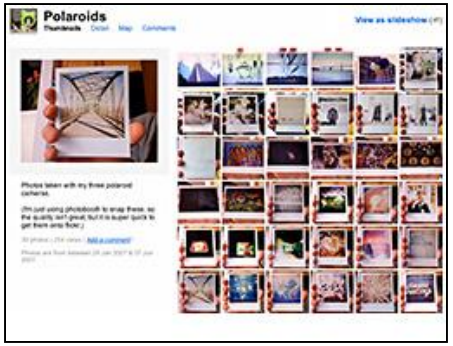
Health

Security



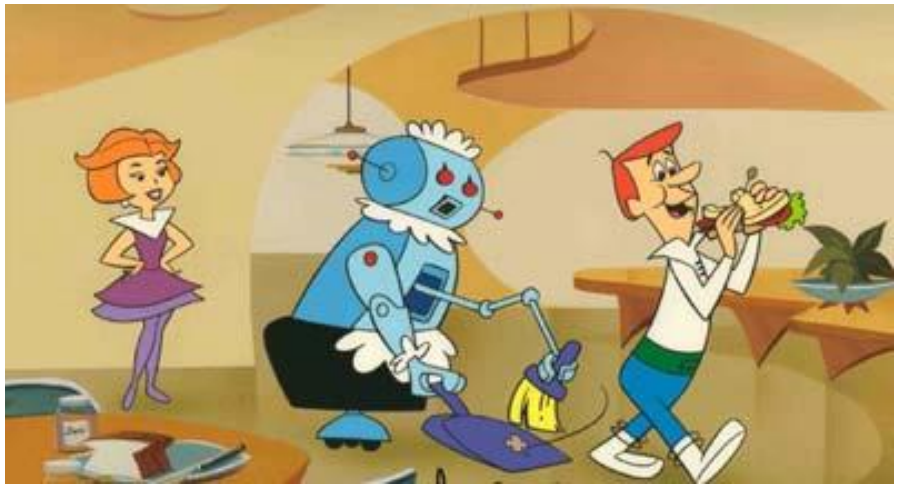
Comfort

Fun



Access

Two reasons for computer vision

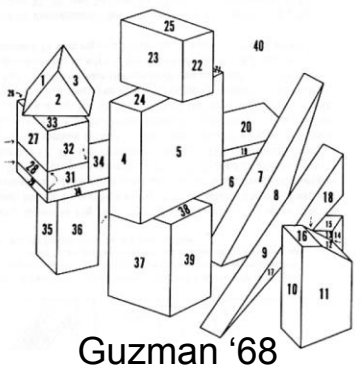


Household Robots

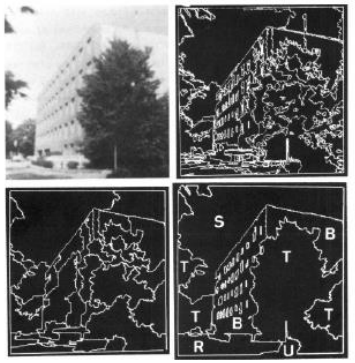
Assisted Driving

Ridiculously brief history of computer vision

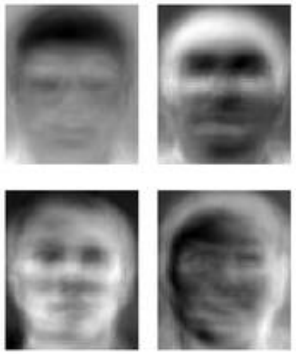
- 1966: Minsky assigns computer vision as an undergrad summer project
- 1960's: interpretation of synthetic worlds
- 1970's: some progress on interpreting selected images
- 1980's: ANNs come and go; shift toward geometry and increased mathematical rigor
- 1990's: face recognition; statistical analysis in vogue
- 2000's: broader recognition; large annotated datasets available; video processing starts
- 2010's: ANNs are back for big improvements in recognition; likely large increase in deployed vision systems
- 2020's: autonomous vehicles, the great robot rebellion?



Guzman '68



Ohta Kanade '78



Turk and Pentland '91

How vision is used now

- Examples of state-of-the-art

Some of the following slides by Steve Seitz

Earth viewers (3D modeling)

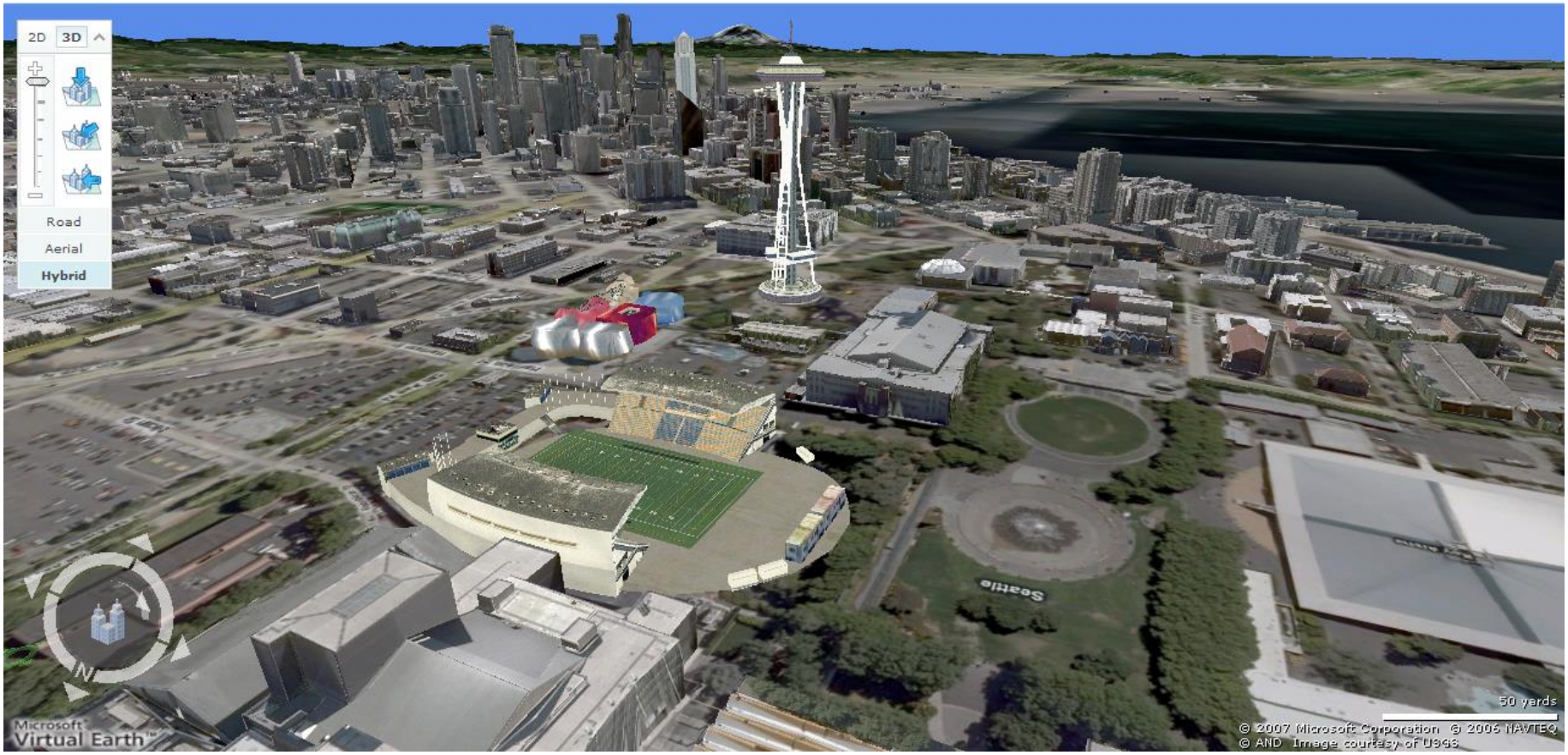


Image from Microsoft's Virtual Earth
(see also: Google Earth)

Vision in construction

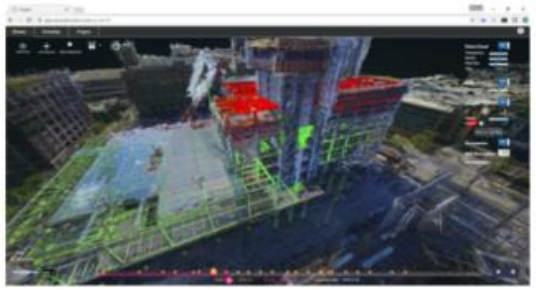
RECONSTRUCT INTEGRATES REALITY AND PLAN

Visual Asset Management

Reconstruct 4D point clouds and organize images and videos from smartphones, time-lapse cameras, and drones around the project schedule. View, annotate, and share anywhere with a web interface.

4D Visual Production Models

Integrate 4D point clouds with 4D BIM, review "who does what work at what location" on a daily basis and improve coordination and communication among project teams.



Predictive Visual Data Analytics

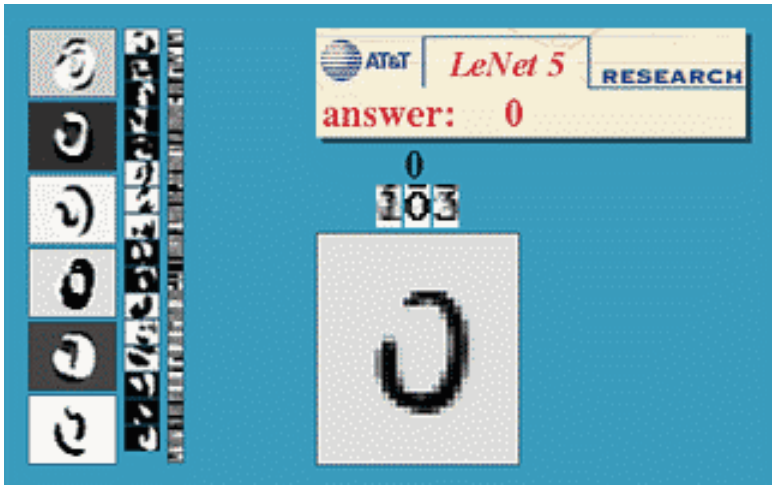
Analyze actual progress deviations by comparing Reality and Plan and predict risk with respect to the execution of the look-ahead schedule for each project location, to offer your project team with an opportunity to tap off potential delays before they surface on your jobsite.

reconstructinc.com

Optical character recognition (OCR)

Technology to convert scanned docs to text

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



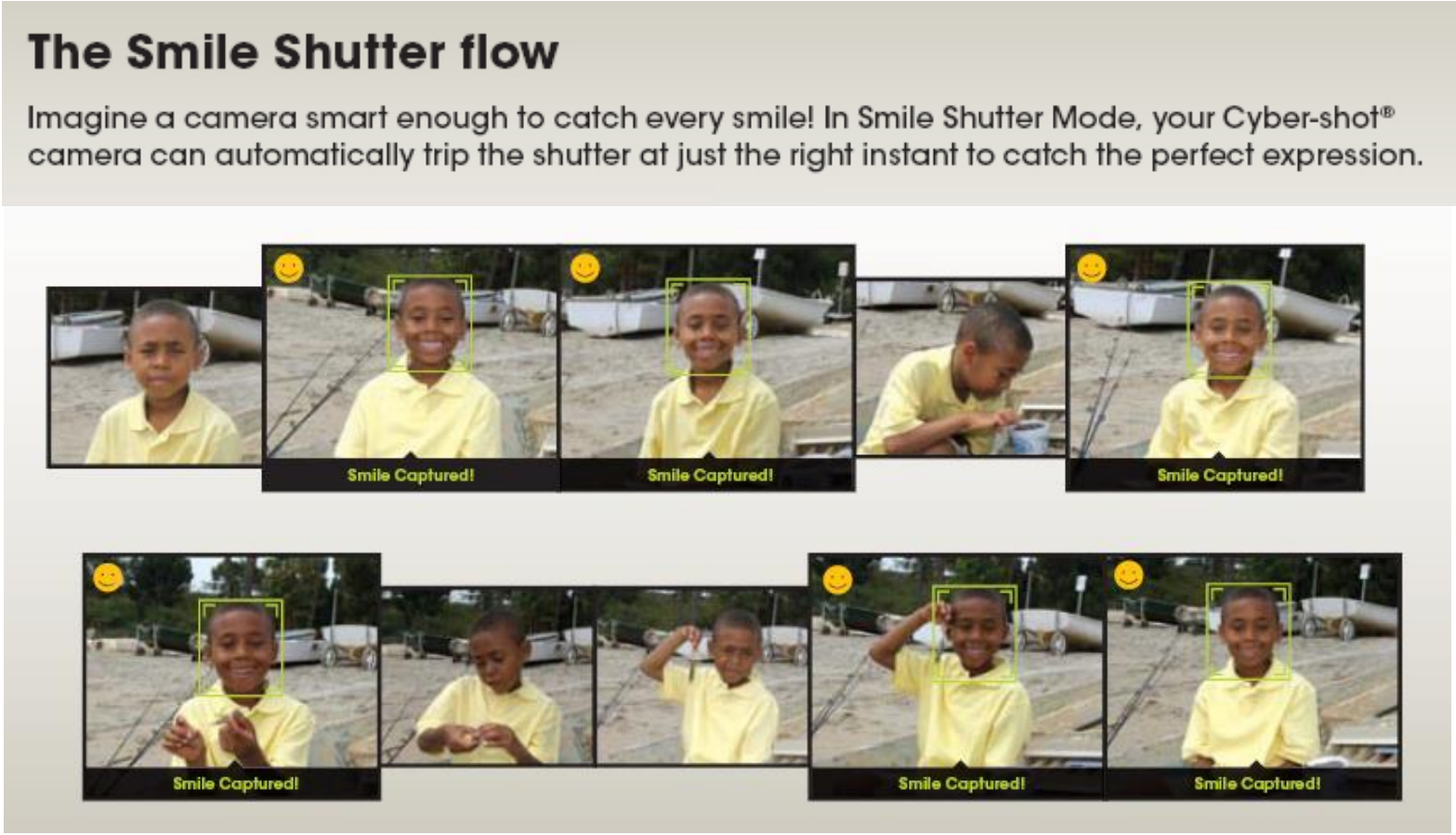
Digit recognition, AT&T labs
http://www.research.att.com/~yann/

License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Face detection

- Most digital cameras detect faces (and more)
  - Canon, Sony, Fuji, …

Smile detection



[Sony Cyber-shot® T70 Digital Still Camera]

Object recognition (in supermarkets)

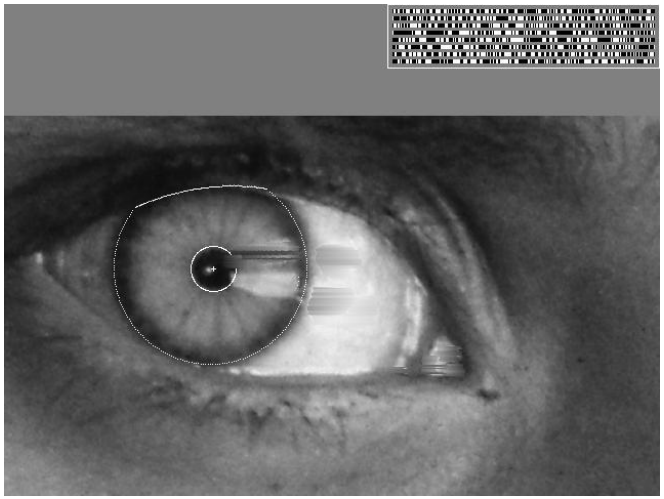
LaneHawk by EvolutionRobotics

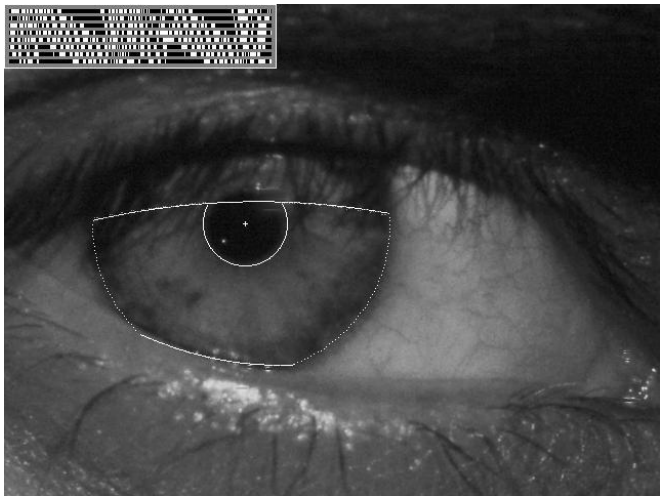
“A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you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

Shopping without checkout

Vision-based biometrics

"*How the Afghan Girl was Identified by Her Iris Patterns*" Read the story wikipedia





Object recognition (in mobile phones)

[Point & Find](), [Nokia]()

[Google Goggles]()

Special effects: shape capture

The Matrix movies, ESC Entertainment, XYZRGB, NRC

Special effects: motion capture

Pirates of the Carribean, Industrial Light and Magic

Sports

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

http://www.sportvision.com/video.html

Smart Cars: Tesla

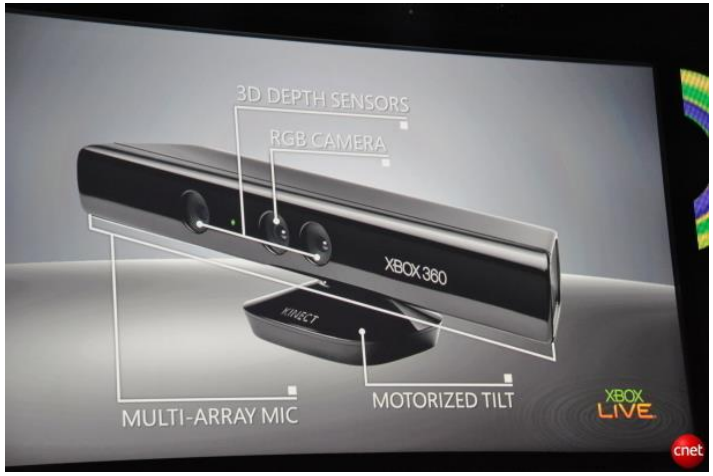
Google cars

Google in talks with Ford, Toyota and Volkswagen to realise driverless cars

http://www.theatlantic.com/technology/archive/2014/05/all-the-world-a-track-the-trick-that-makes-googles-self-driving-cars-work/370871/

Interactive Games: Kinect

- Object Recognition: http://www.youtube.com/watch?feature=iv&v=fQ59dXOo63o
- Mario: http://www.youtube.com/watch?v=8CTJL5lUjHg
- 3D: http://www.youtube.com/watch?v=7QrnwoO1-8A
- Robot: http://www.youtube.com/watch?v=w8BmgtMKFbY



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

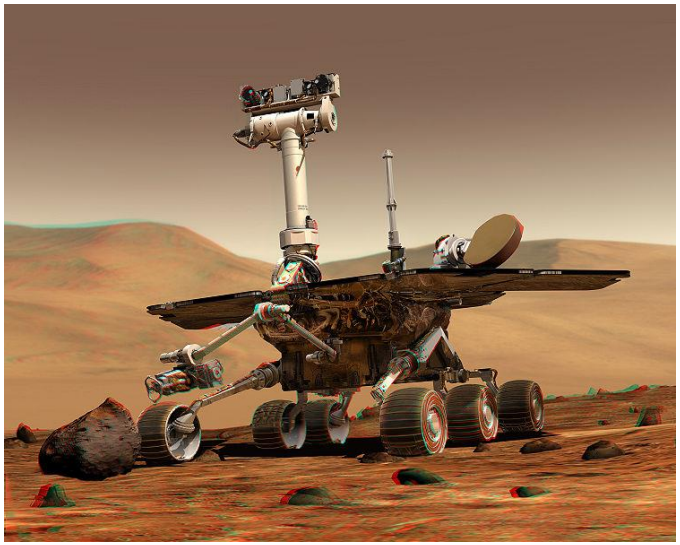
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- For more, read "Computer Vision on Mars" by Matthies et al.

Industrial robots

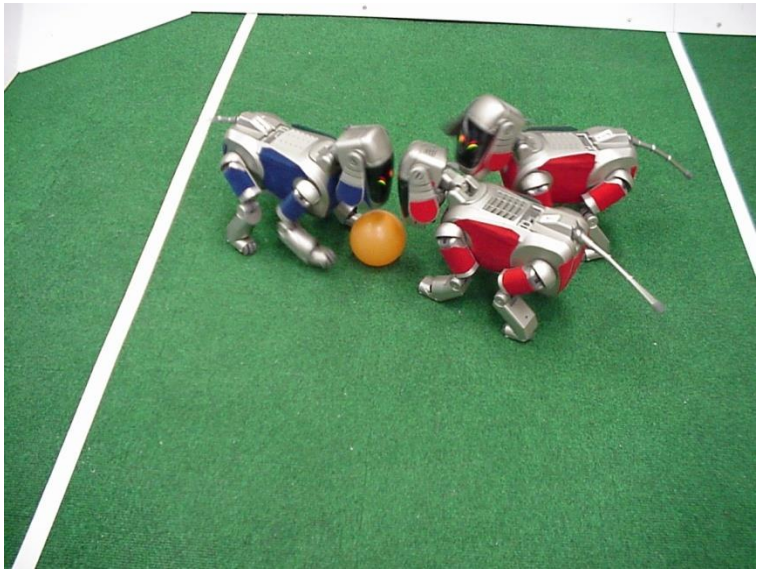
Vision-guided robots position nut runners on wheels

http://www.automationworld.com/computer-vision-opportunity-or-threat

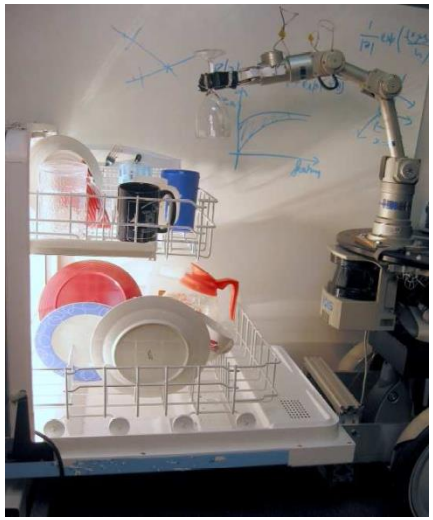
Mobile robots



NASA's Mars Spirit Rover
http://en.wikipedia.org/wiki/Spirit_rover



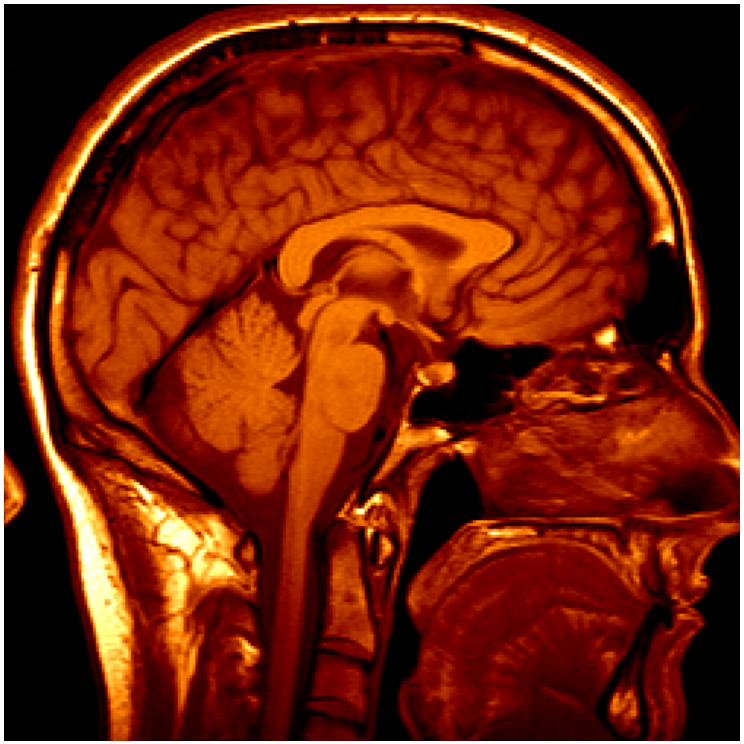
http://www.robocup.org/



Saxena et al. 2008
STAIR at Stanford

http://www.youtube.com/watch?v=DF39Ygp53mQ

Medical imaging



3D imaging
MRI, CT

Image guided surgery
Grimson et al., MIT

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

Course outline

Prof: Derek Hoiem (dhoiem@illinois.edu), SC3312

TAs: Tanmay Gupta, tgupta6

Bryan Plummer, bplumme2

Chuhang Zou, czou4

Web page:

https://courses.engr.illinois.edu/cs543/

Waitlist

- Fill out this form and keep coming to class
  https://goo.gl/forms/p3jQZ4CX3exc5QNm2

- Will create priority list for 25 overflow seats and any drops by early next week

- I don’t know what ECE 549 ADD is ¯_(ツ)_/¯

Grades

- Homeworks (75%)
- Final project (25%)
- Attendance

Late policy

- 10 points per day
- 5 late days forgiven

Academic Integrity

- Can discuss hw with peers, but don't copy
- Carefully document any sources within hw hand-in
- Don't use code from Internet unless you have permission
  - If you're not sure, ask
- Minimum penalty for any violation is zero on full assignment (15% of total grade)

Getting help outside of class

Office hours

- Time, see website
- Otherwise, just stop by. If I'm not there, send me an e-mail.

Discussion board:
https://piazza.com/class/i4ohqjpy49s6ga

TAs: Tanmay Gupta (tgupta6@illinois.edu)
Bryan Plummer (bplumme2@illinois.edu)
Chuhang Zou (czou4@illinois.edu)

Readings/Textbook

- Computer Vision: A Modern Approach (2nd edition) by David Forsyth and Jean Ponce (2011)
- See syllabus for other useful books

What to expect from this course

- Broad coverage (geometry, image processing, recognition, multiview, video)
- Background to delve deeper into any computer vision-related topic
- Practical experience
- Lots of work, tough material, fast pace, but hopefully lots of learning too!

Topics

- Interpreting Intensities
  - What determines the brightness and color of a pixel?
  - How can we use image filters to extract meaningful information from the image?
- Correspondence and Alignment
  - How can we find corresponding points in objects or scenes?
  - How can we estimate the transformation between them?
- Perspective and 3D Geometry
  - How can we map between the 3D world and the 2D image?
  - How can we recover 3D coordinates from images or video?
- Grouping and Segmentation
  - How can we group pixels into meaningful regions?
- Categorization and Object Recognition
  - How can we represent images and categorize them?
  - How can we recognize categories of objects?
- Advanced Topics
  - Action recognition, 3D scenes and context, VQA, …

Prerequisites

- **Linear algebra**, basic calculus, and probability
- Programming will mostly be in Matlab, may be some Python
- Experience with image processing or Matlab will help but is not necessary

Goals and Expectations

- My goal: maximize the learning effectiveness of your time

- What I expect from you
  - Attend and participate, when possible
  - Start assignments well before deadline
  - Tell me what's working and suggest improvements

Final comments

- To do
  - Sign up for newsgroup:
    https://piazza.com/class/i4ohqjpy49s6ga
  - Sign up for Matlab tutorial with Bryan (if interested):
    http://doodle.com/poll/cwg3qeiuh2yq89hs
  - Read syllabus, etc.

- Next class: light and color

- Questions?