

Honors Machine Vision
Jan 17, 2017

Kristen Grauman, University of Texas at Austin

Introductions

- **Instructor:** Prof. Kristen Grauman
- **TA:** Dongguang You

Today

- Course overview
- Requirements, logistics

What is computer vision?



Done?

Computer Vision

- Automatic understanding of images and video
 1. Computing properties of the 3D world from visual data (*measurement*)

1. Vision for measurement

Real-time stereo



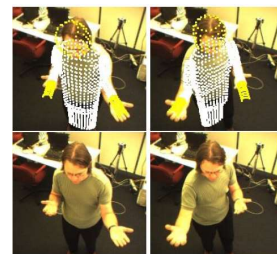
Wang et al.

Structure from motion



Snaveley et al.

Tracking



Demirdjian et al.

Computer Vision

- Automatic understanding of images and video
 1. Computing properties of the 3D world from visual data (*measurement*)
 2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. (*perception and interpretation*)

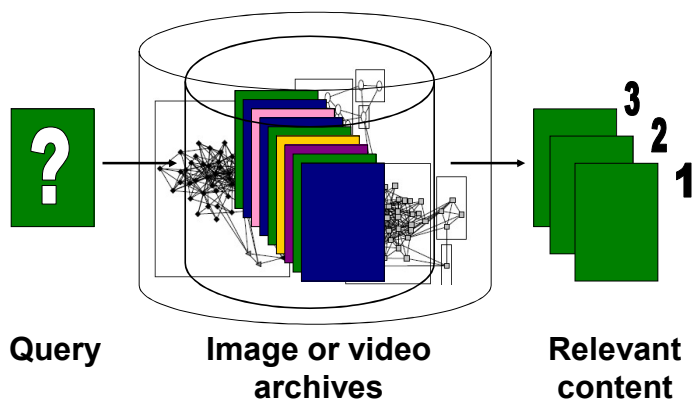
2. Vision for perception, interpretation



Computer Vision

- Automatic understanding of images and video
 1. Computing properties of the 3D world from visual data (*measurement*)
 2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. (*perception and interpretation*)
 3. Algorithms to mine, search, and interact with visual data (*search and organization*)

3. Visual search, organization

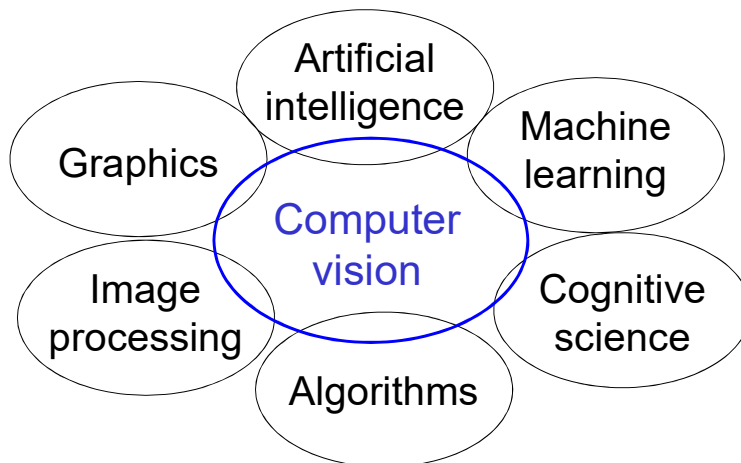


Computer Vision

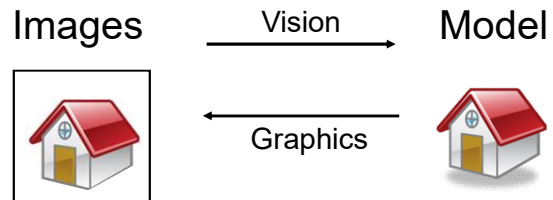
- Automatic understanding of images and video
 1. Computing properties of the 3D world from visual data (*measurement*)
 2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. (*perception and interpretation*)
 3. Algorithms to mine, search, and interact with visual data (*search and organization*)

Course focus

Related disciplines

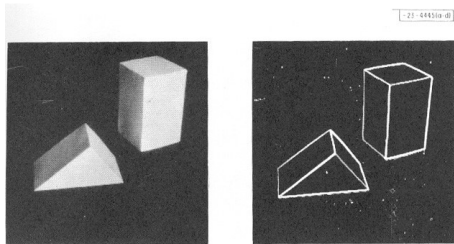


Vision and graphics



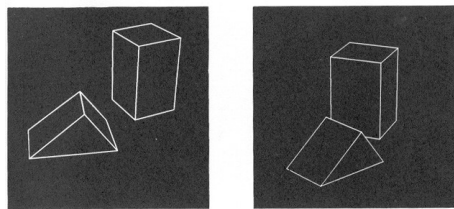
Inverse problems: analysis and synthesis.

Visual data in 1963



(a) Original picture.

(b) Differentiated picture.

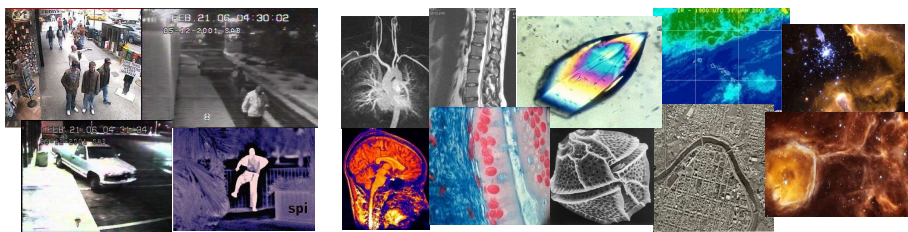
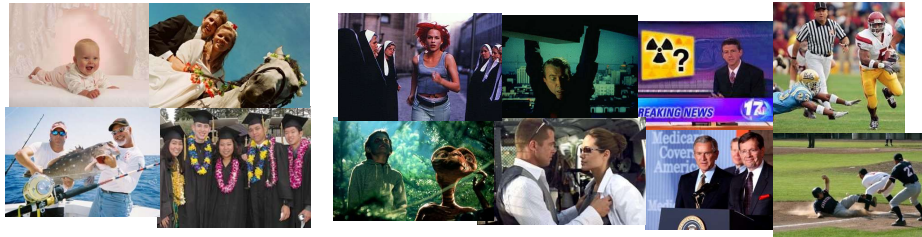


(c) Line drawing.

(d) Rotated view.

L. G. Roberts, *Machine Perception of Three Dimensional Solids*, Ph.D. thesis, MIT Department of Electrical Engineering, 1963.

Visual data in 2017



Slide credit: L. Lazebnik

Why vision?

- As image sources multiply, so do applications
 - Relieve humans of boring, easy tasks
 - Enhance human abilities
 - Advance human-computer interaction, visualization
 - Perception for robotics / autonomous agents
 - Organize and give access to visual content

Faces and digital cameras

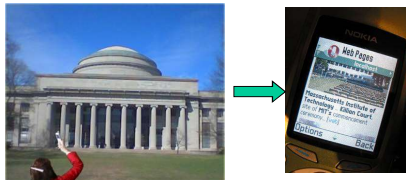


Camera waits for everyone to smile to take a photo [Canon]

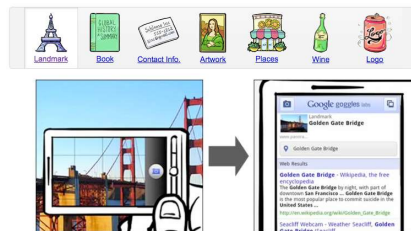


Setting camera focus via face detection

Linking to info with a mobile device



Situated search
Yeh et al., MIT



Google Goggles



MSR Lincoln



kooba

Video-based interfaces



Human joystick, NewsBreaker Live



Assistive technology systems
Camera Mouse, Boston College



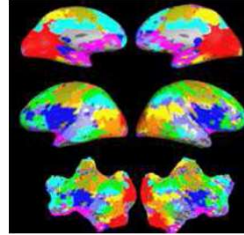
Microsoft Kinect

What else?

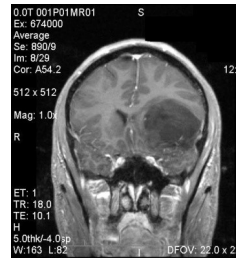
Vision for medical & neuroimages



Image guided surgery
MIT AI Vision Group



fMRI data
Golland et al.



Special visual effects



The Matrix



What Dreams May Come

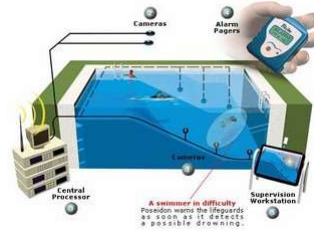


Mocap for *Pirates of the Caribbean*,
Industrial Light and Magic
Source: S. Seitz

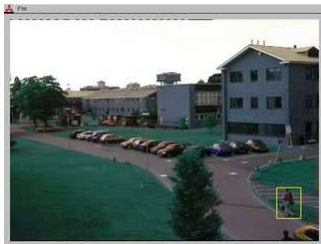
Safety & security



Navigation,
driver safety



Monitoring pool
(Poseidon)



Pedestrian detection
MERL, Viola et al.



Surveillance

Obstacles?

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

July 7, 1966

THE SUMMER VISION PROJECT

Seymour Papert

The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the construction of a system complex enough to be a real landmark in the development of "pattern recognition".

Challenges: many nuisance parameters



Illumination



Object pose



Clutter



Occlusions



Intra-class appearance



Viewpoint

Challenges: intra-class variation



slide credit: Fei-Fei, Fergus & Torralba

Challenges: importance of context



Challenges: importance of context



Challenges: importance of context

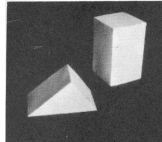


slide credit: Fei-Fei, Fergus & Torralba

Challenges: complexity

- Millions of pixels in an image
- 30,000 human recognizable object categories
- 30+ degrees of freedom in the pose of articulated objects (humans)
- Billions of images online
- 144K hours of new video on YouTube daily
- ...
- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991]

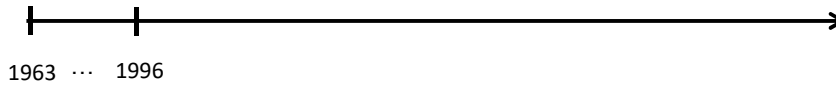
Progress charted by datasets



Roberts 1963



COIL



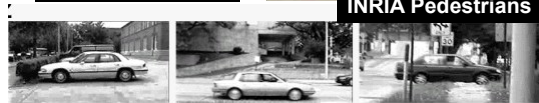
Progress charted by datasets



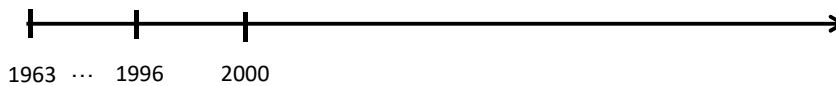
MIT-CMU Faces



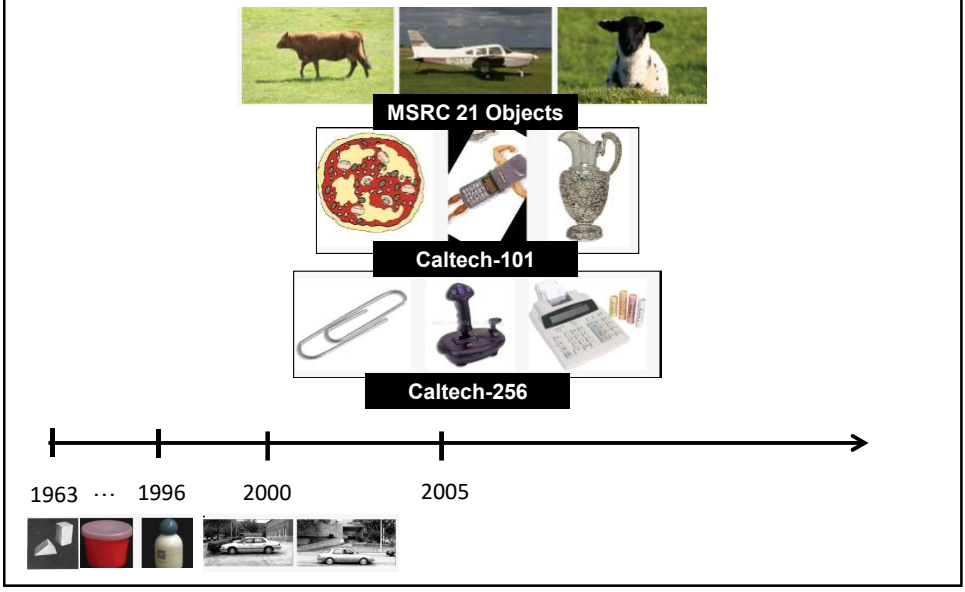
INRIA Pedestrians



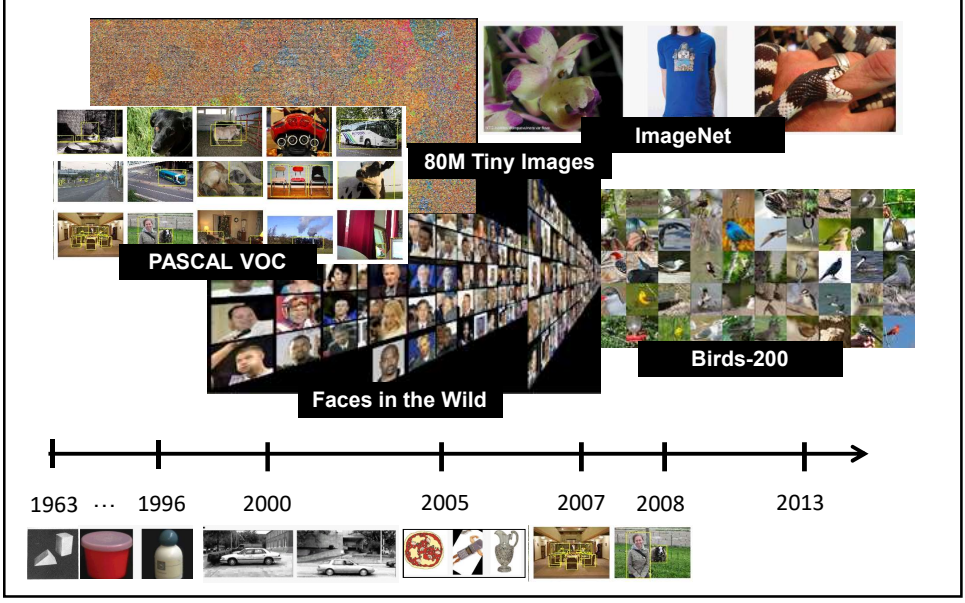
UIUC Cars



Progress charted by datasets



Progress charted by datasets



Expanding horizons: large-scale recognition

clarifai

ABOUT TECHNOLOGY API NEWS BLOG CAREERS CONTACT

Paste a url here...

USE THE URL

CHOOSE A FILE INSTEAD

*By using the demo you agree to our terms of service



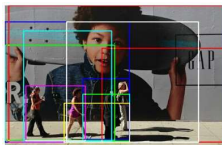
Predicted Tags

mammal livestock cattle
pasture agriculture bovine
farm nobody meadow grass

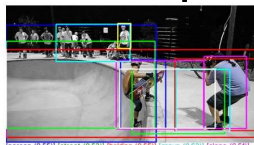
Similar Images



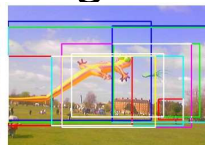
Expanding horizons: captioning



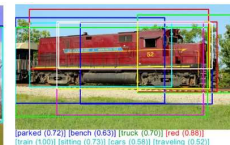
men (0.56) group (0.66) woman (0.64)
people (0.99) holding (0.60) playing (0.64) news (0.89)
court (0.51) standing (0.69) like (0.58) forest (0.52)
man (0.77) skateboard (0.67)
a group of people standing next to each other
people stand outside a large ad for gap featuring a young boy



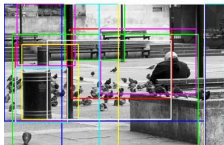
person (0.55) [street (0.53) building (0.55) ground (0.63) slope (0.59)
downhill (0.51) snow (0.91) like (0.74) [river (0.54)
people (0.89) men (0.57) asking (0.54)
skateboard (0.81) riding (0.76) news (0.74) trick (0.53) skate (0.52)
woman (0.52) men (0.86) down (0.69)
a group of people riding skis down a snow covered slope
a guy on a skate board on the side of a ramp



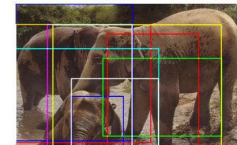
airplane (0.57) plane (0.58) kite (0.83) people (0.80)
thing (0.51) men (0.57) person (0.61) leave (0.6)
sky (0.61) like (0.74) head (0.75)
a couple of people flying kites in a field
people in a field flying different styles of kites



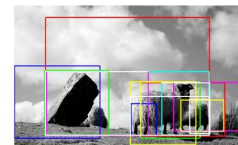
parked (0.72) bench (0.63) truck (0.70) road (0.88)
train (0.99) battery (0.73) Sears (0.63) traveling (0.52)
grass (0.65) track (0.69) car (0.59) yellow (0.57)
field (0.60) people (0.50) down (0.54) tracks (0.84)
a train traveling down train tracks near a field
a red train is coming down the tracks



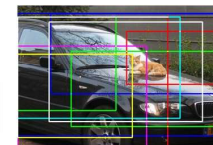
umbrella (0.59) woman (0.52)
fire (0.86) hydrant (0.96) forest (0.79) cat (0.50)
bench (0.81) building (0.71) baseball (0.57) baseball (0.55)
white (0.82) sitting (0.63) people (0.79) photo (0.53)
bank (0.81) kitchen (0.54) man (0.72) water (0.52)
a black and white photo of a tea house
a restaurant full of people prepare and garbage cans also has benches on
either side of it one of which shows the back of a large person facing
in the direction of the picture



horse (0.53) bear (0.71) [elephant (0.92) elephants (0.99)
brown (0.88) baby (0.62) walking (0.57) laying (0.61)
man (0.57) standing (0.79) field (0.65)
water (0.63) large (0.70) dirt (0.65) river (0.58)
a baby elephant standing next to each other on a field
elephants are playing together in a shallow watering hole



man (0.59) beach (0.54) sky (0.53) bird (0.50) [tree (0.58)
river (0.60) mountain (0.56) recording (0.61) white (0.64)
people (0.91) dog (0.60) cows (0.55)
elephant (0.92) black (0.84) grass (0.64) horse (0.60)
elephants (0.71) bear (0.61)
a black bear standing on top of a grass covered field
a couple of sheep standing up on a small hill



bee (0.56) car (0.79) back (0.57) truck (0.86)
person (0.87) head (0.61) person (0.61) dog (0.65)
sitting (0.55) man (0.53) sun (0.72)
a dog sitting on top of a car
a cat is lying on the hood of a black car



<https://pdollar.wordpress.com/2015/01/21/image-captioning/>

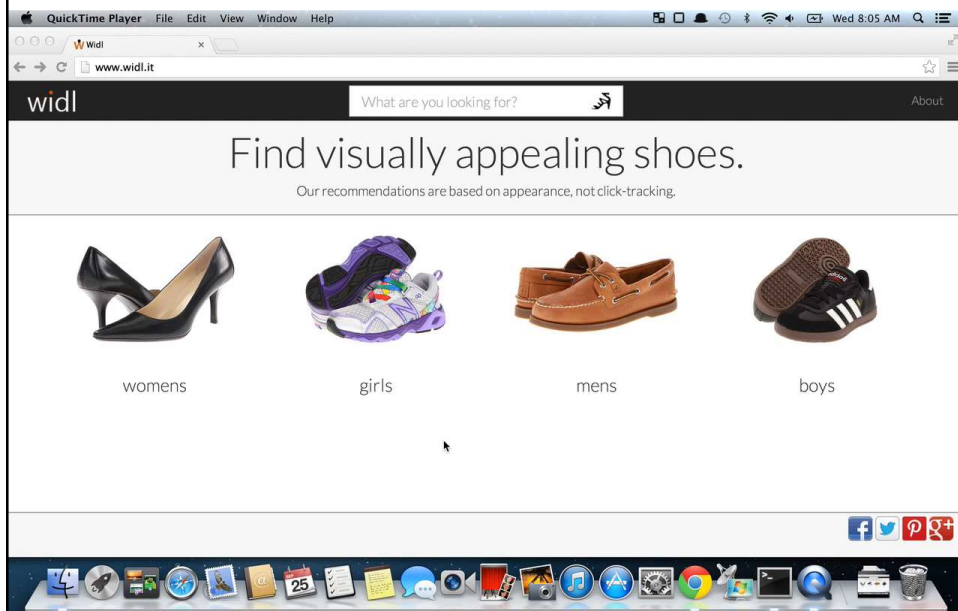
Expanding horizons:



Turn on your speakers!

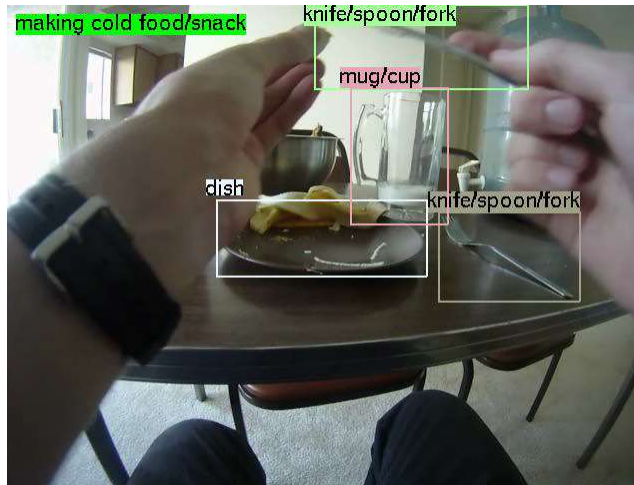
KITTI dataset – Andreas Geiger et al.

Expanding horizons: interactive visual search



The screenshot shows a web browser window displaying the website 'widl'. The browser's address bar shows 'www.widl.it'. The website has a search bar with the placeholder text 'What are you looking for?' and an 'About' link. Below the search bar, the main heading reads 'Find visually appealing shoes.' followed by the subtext 'Our recommendations are based on appearance, not click-tracking.' There are four categories of shoes displayed: 'womens' (black high-heeled shoes), 'girls' (purple and white sneakers), 'mens' (brown boat shoes), and 'boys' (black sneakers with white stripes). At the bottom of the page, there are social media icons for Facebook, Twitter, and Google+, and a Mac OS X dock with various application icons.

Expanding horizons: first-person vision



Activities of Daily Living – Hamed Pirsiavash et al.

Brainstorm

Pick an application or task among any of those we've described so far.

1. What functionality should the system have?
2. Intuitively, what are the technical sub-problems that must be solved?

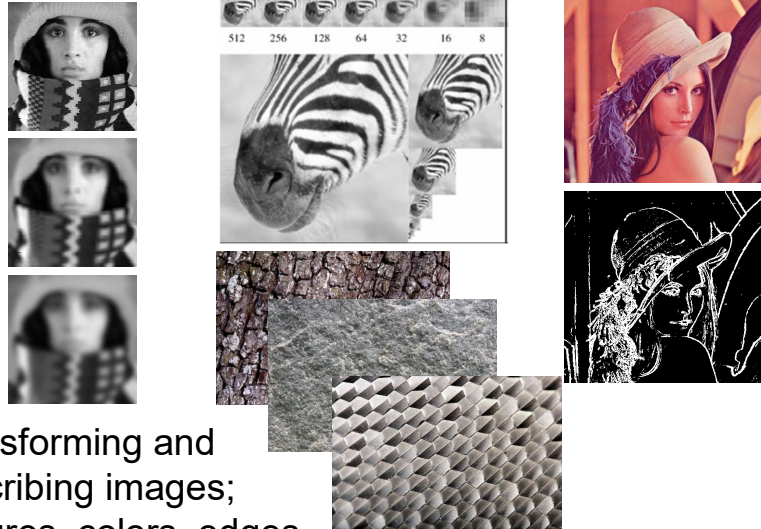
Goals of this course

- Upper division honors undergrad course
- Introduction to primary topics
 - Fundamentals of computer vision – image processing, grouping, multiple views
 - Recognition - emphasis on *learning* (~last third of the course)
- Hands-on experience with algorithms
- Views of vision as a research area

Topics overview

- Features & filters
- Grouping & fitting
- Multiple views
- Recognition

Features and filters



Transforming and describing images; textures, colors, edges

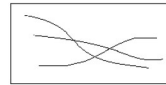
Grouping & fitting



Parallelism



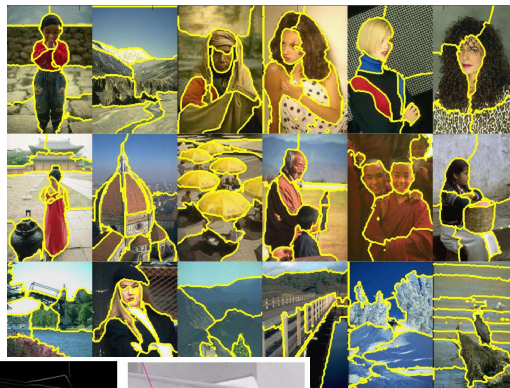
Symmetry



Continuity

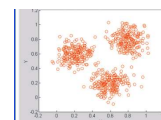
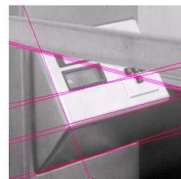


Closure

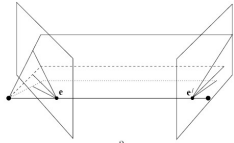


[fig from Shi et al]

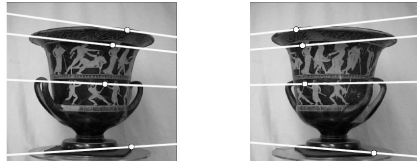
Clustering, segmentation, fitting; what parts belong together?



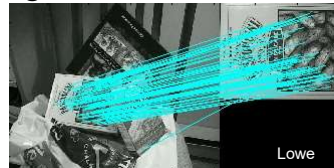
Multiple views



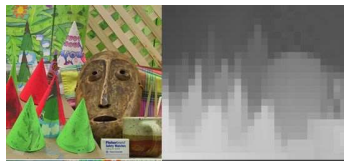
Matching, invariant features, stereo vision, instance recognition



Hartley and Zisserman

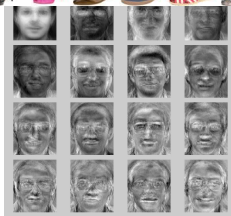
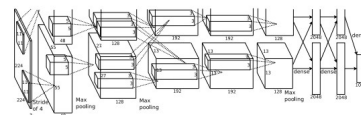
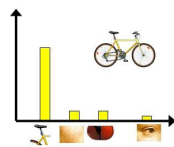


Lowe

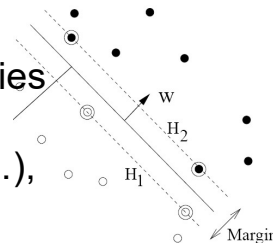


Fei-Fei Li

Recognition and learning

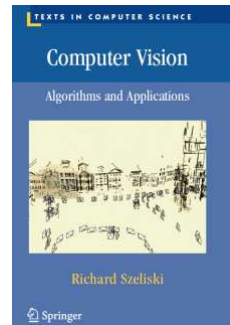


Recognizing categories (objects, scenes, activities, attributes...), learning techniques



Textbooks

- Recommended book:
 - Computer Vision: Algorithms and Applications
 - By Rick Szeliski
 - <http://szeliski.org/Book/>



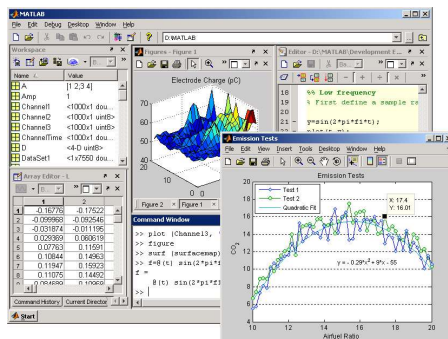
Requirements / Grading

- Problem sets (50%)
 - Midterm exam (15%)
 - Final exam (25%)
 - Class participation, including attendance (10%)
 - Check grades on Canvas
- A quote from a prior student evaluation:
“To be honest, I think without going to class, the course would be very hard. “

Assignments

- Majority - Programming problem
 - Implementation
 - Explanation, results
- Code in Matlab – available on CS Unix machines (see course page)
- Optional Latex templates
- Most of these assignments take significant time to do. We recommend starting early.

Matlab



- Built-in toolboxes for low-level image processing, visualization
- Compact programs
- Intuitive interactive debugging
- Widely used in engineering

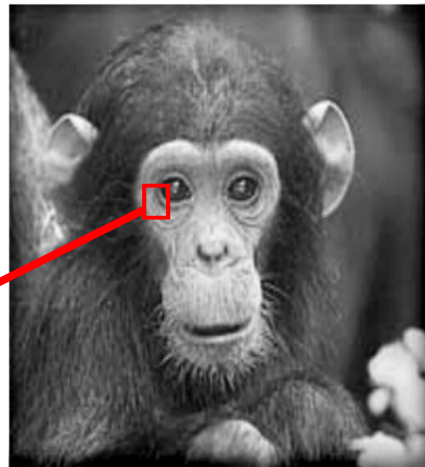
Assignment 0

- A0: Matlab warmup + basic image manipulation
- Out today, due Fri Jan 27

- Verify CS account and Matlab access
- Look at the tutorial online

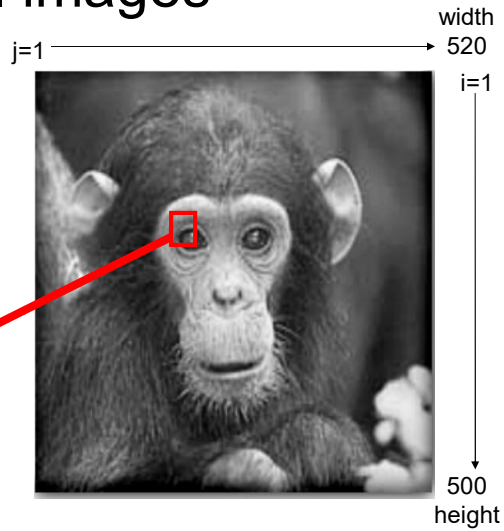
Digital images

Images as matrices



Digital images

Intensity : [0,255]



$im[176][201]$ has value 164

$im[194][203]$ has value 37

Color images,
RGB color
space



R

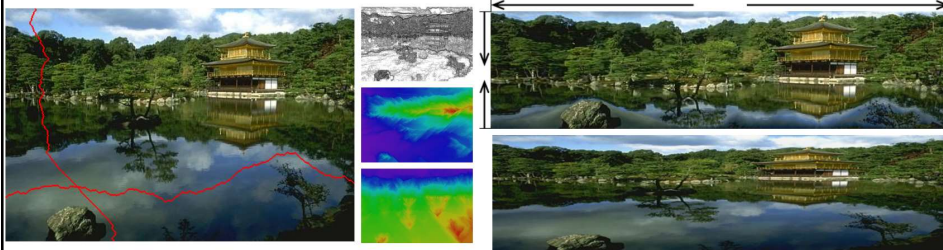


G



B

Preview of assignments



Seam carving

Preview of assignments



Grouping for segmentation

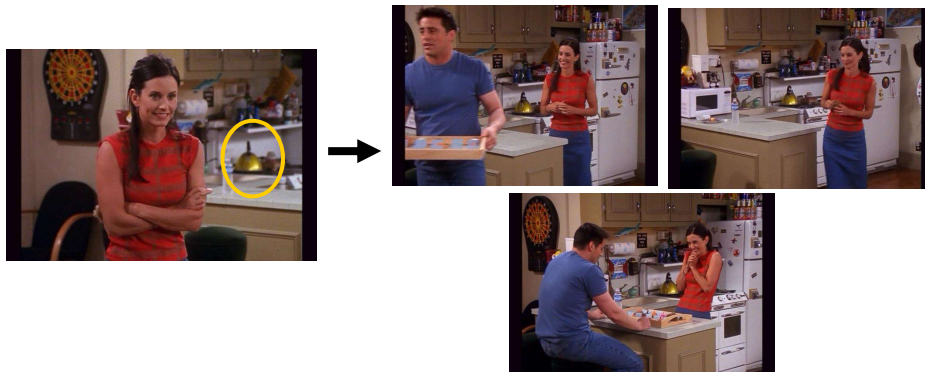
Preview of assignments



Image mosaics / stitching

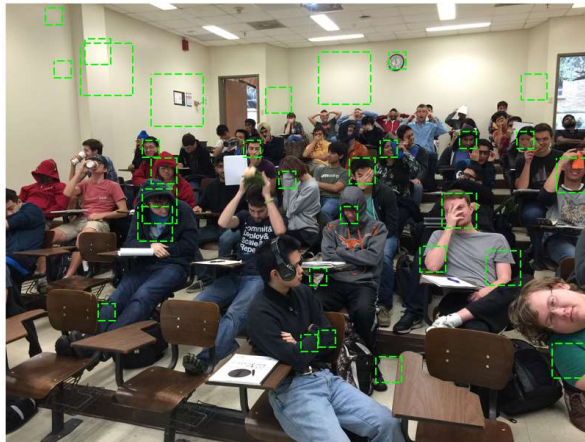
Image from Fei-Fei Li

Preview of assignments



Matching and recognition

Preview of assignments



Object detection

Collaboration policy

All responses and code must be written individually unless otherwise specified.

Students submitting answers or code found to be identical or substantially similar (due to inappropriate collaboration) risk failing the course.

Assignment deadlines

- Due about every two weeks
 - tentative deadlines posted online but could slightly shift depending on lecture pace
- Assignments in by 11:59 PM on due date
 - Submit on Canvas, following submission instructions given in assignment.
 - Deadlines are firm. We'll use timestamp.
- Use Piazza, office hours for questions

Miscellaneous

- Slides, announcements via class website
- No laptops, phones, tablets, etc. open in class please.

Coming up

- Now: check out Matlab tutorial online
- A0 due Fri Jan 27
- Textbook reading posted for next week