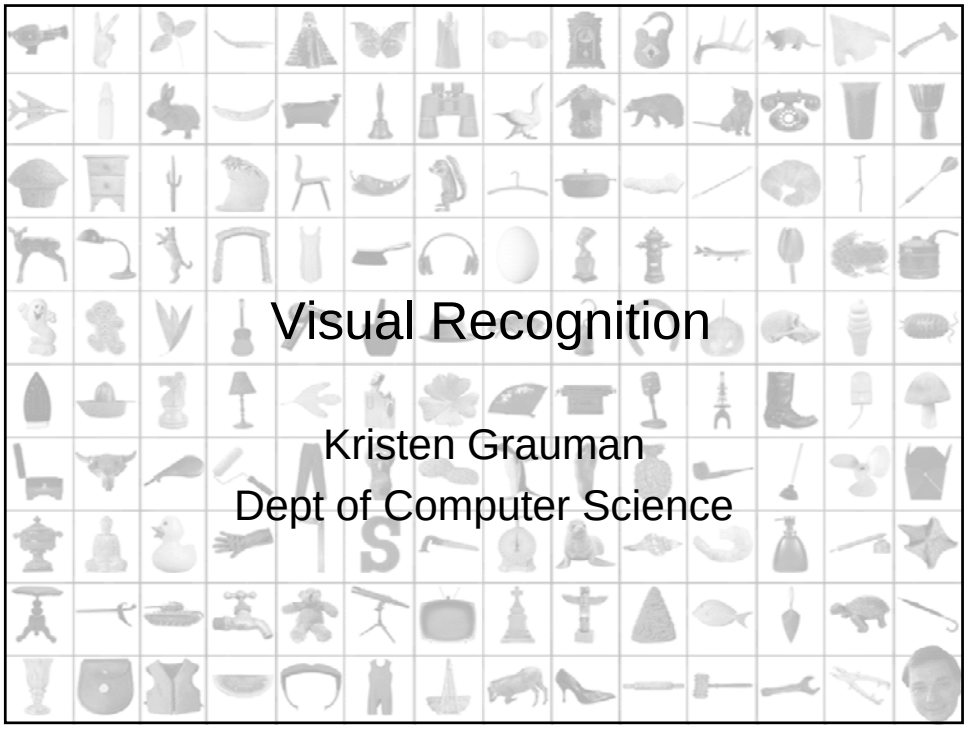


Visual Recognition

Kristen Grauman
Dept of Computer Science

Plan for today

- Topic overview:
  - What does the visual recognition problem entail?
  - Why are these hard problems?
  - What works today?
- Course overview:
  - Requirements
  - Syllabus tour

Computer Vision

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

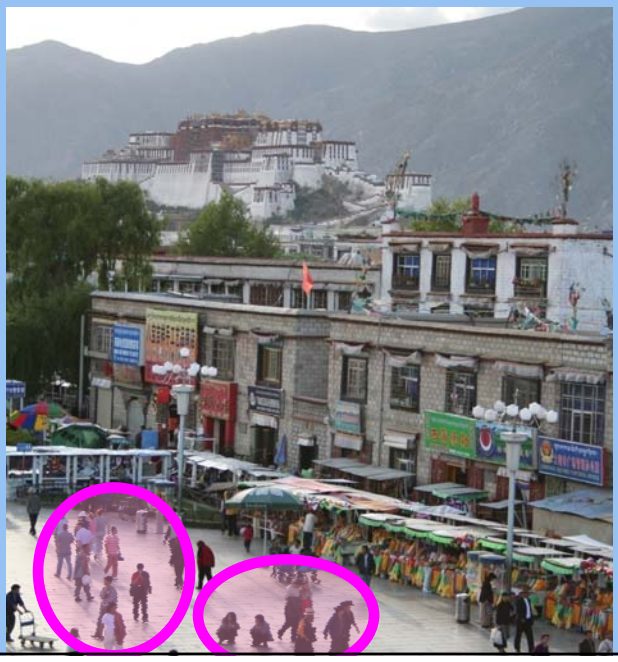
What does recognition involve?

Slide by Fei-Fei Li

Detection: are there people?

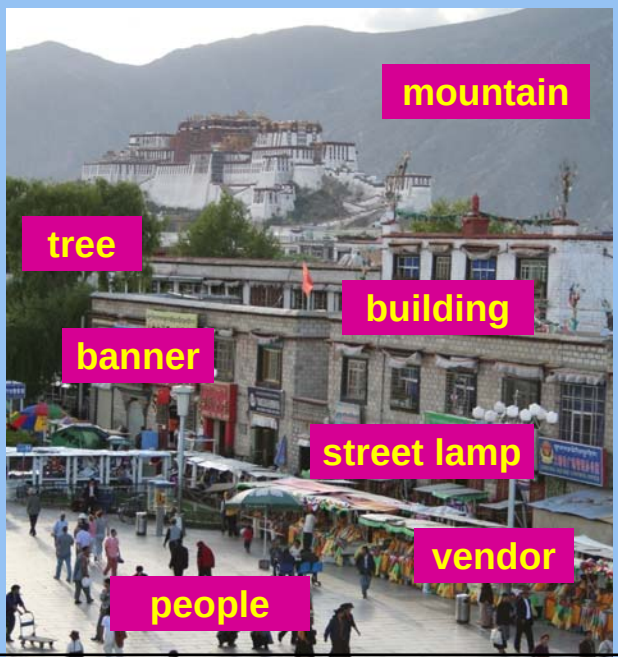
Slide by Fei-Fei Li

Activity: What are they doing?



Slide by Fei-Fei Li

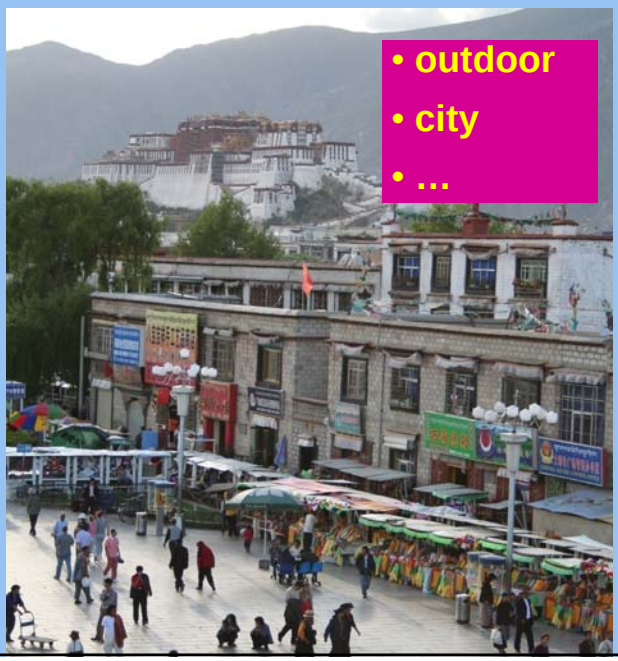
Object categorization



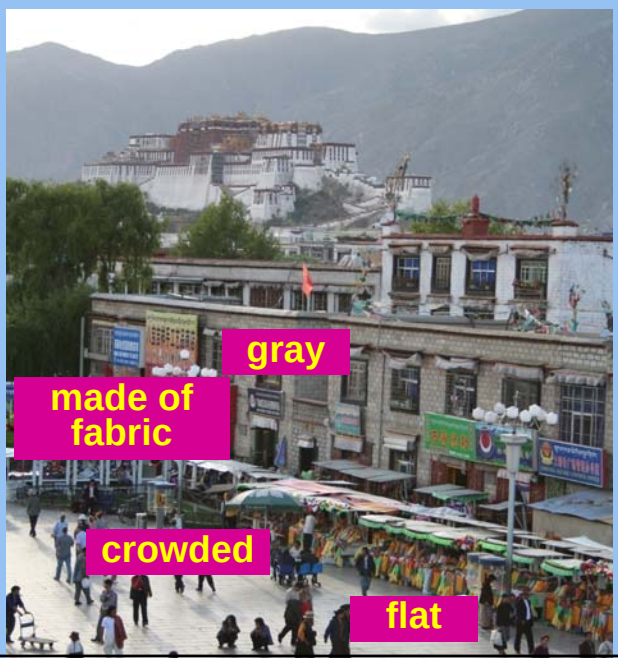
Slide by Fei-Fei Li

Instance recognition

Scene and context categorization



Attribute recognition



Object Categorization

- Task Description
  - "Given a small number of training images of a category, recognize a-priori unknown instances of that category and assign the correct category label."
- Which categories are feasible visually?

"Fido" — German shepherd — dog — animal — living being

Visual Object Recognition Tutorial

K. Grauman, B. Leibe

Visual Object Categories

- Basic Level Categories in human categorization
  [Rosch 76, Lakoff 87]
  - The highest level at which category members have similar perceived shape
  - The highest level at which a single mental image reflects the entire category
  - The level at which human subjects are usually fastest at identifying category members
  - The first level named and understood by children
  - The highest level at which a person uses similar motor actions for interaction with category members

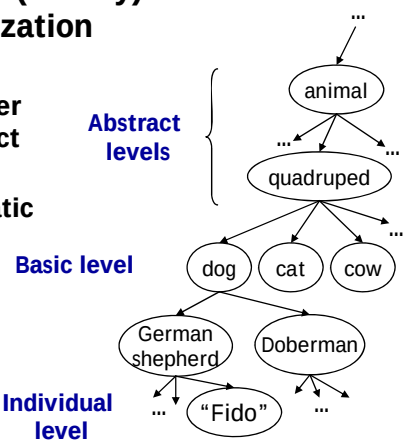
Visual Object Recognition Tutorial

K. Grauman, B. Leibe

Visual Object Recognition Tutorial

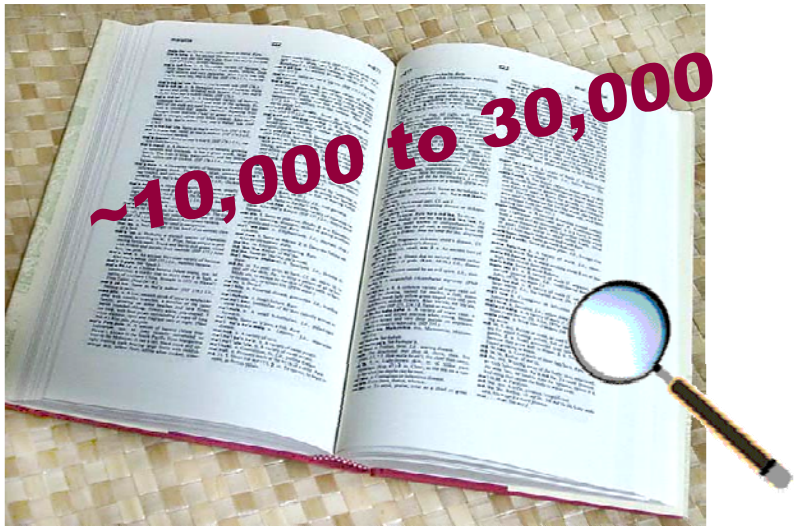
Visual Object Categories

- **Basic-level categories in humans seem to be defined predominantly visually.**
- **There is evidence that humans (usually) start with basic-level categorization *before* doing identification.**
  - ⇒ **Basic-level categorization is easier and faster for humans than object identification!**
  - ⇒ **How does this transfer to automatic classification algorithms?**



K. Grauman, B. Leibe

How many object categories are there?



Source: Fei-Fei Li, Rob Fergus, Antonio Torralba.

Biederman 1987

Visual Object Recognition Tutorial

Other Types of Categories

- **Functional Categories**
  - **e.g. chairs =** *"something you can sit on"*







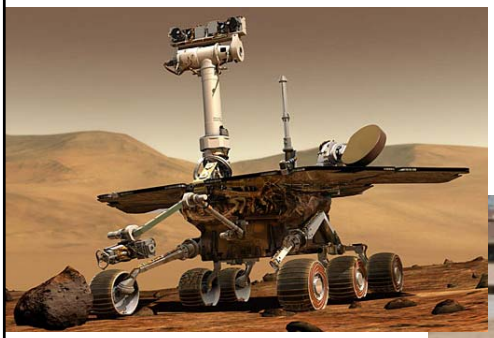
K. Grauman, B. Leibe

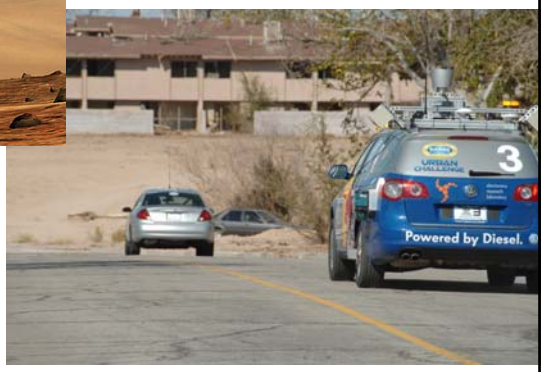
Why recognition?

- Recognition a fundamental part of perception
  - e.g., robots, autonomous agents
- Organize and give access to visual content
  - Connect to information
  - Detect trends and themes

- Why now?

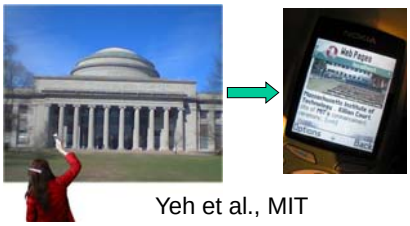
Autonomous agents able to detect objects





http://www.darpa.mil/grandchallenge/gallery.asp

Posing visual queries



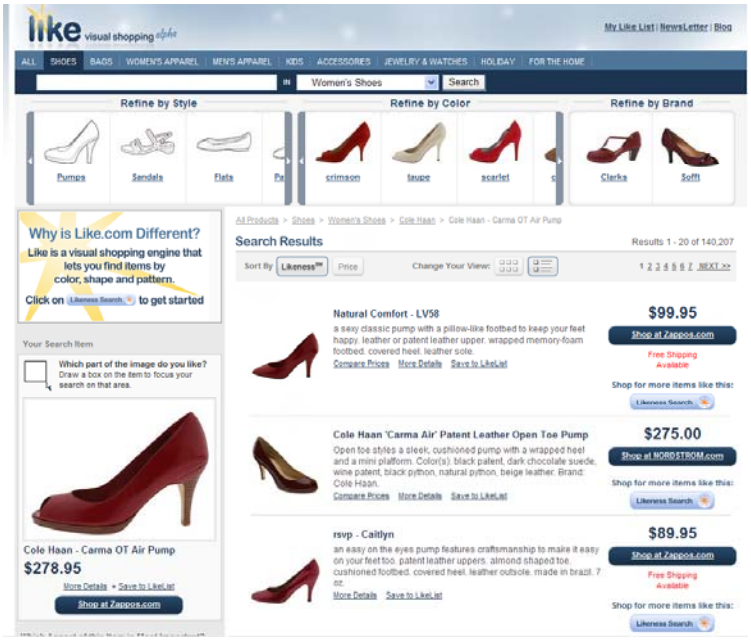
Yeh et al., MIT

Digital Field Guides Eliminate the Guesswork

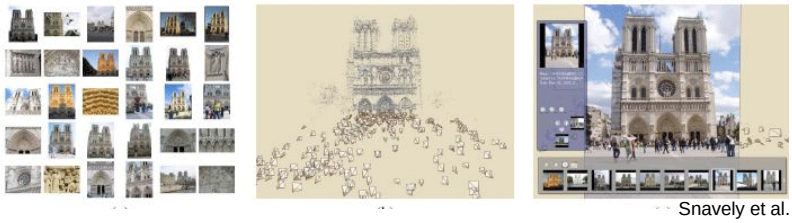
Belhumeur et al.

Kooaba, Bay & Quack et al.

Finding visually similar objects



Exploring community photo collections



Snavely et al.

Simon & Seitz

Discovering visual patterns

Objects Sivic & Zisserman

Lee & Grauman

Categories

Wang et al.

Actions

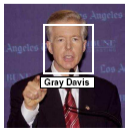
Auto-annotation

Figure 9. Results of automatic object-level annotation with bounding boxes. Groundtruth annotation is shown with dashed lines, correct detection with solid green lines, false detections with solid red lines. Auto-annotation with related Wikipedia articles is also shown. All results are also labeled with their GPS position and estimated tags (not shown here).

Gammeter et al.

President George W. Bush makes a statement in the Rose Garden while Secretary of Defense Donald Rumsfeld looks on, July 23, 2003. Rumsfeld said the United States would release graphic photographs of the dead sons of Saddam Hussein to prove they were killed by American troops. Photo by Larry Downing/Reuters

British director Sam Mendes and his partner actress Kate Winslet arrive at the London premiere of 'The Road to Perdition', September 18, 2002. The films stars Tom Hanks as a Chicago hit man who has a separate family life and co-stars Paul Newman and Jude Law. REUTERS/Dan Chung



Incumbent California Gov. Gray Davis (news - web sites) leads Republican challenger Bill Simon by 10 percentage points – although 17 percent of voters are still undecided, according to a poll released October 22, 2002 by the Public Policy Institute of California. Davis is shown speaking to reporters after his debate with Simon in Los Angeles, on Oct. 7. (Jim Ruymen/Reuters)

T. Berg et al.

Challenges

Challenges: robustness

Illumination

Object pose

Clutter

Occlusions

Intra-class appearance

Viewpoint

Challenges: context and human experience

Context cues

Challenges: context and human experience

Context cues

Function

Dynamics

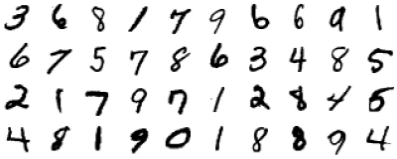
Video credit: J. Davis

Challenges: scale, efficiency

- Half of the cerebral cortex in primates is devoted to processing visual information
- ~20 hours of video added to YouTube per minute
- ~5,000 new tagged photos added to Flickr per minute
- Thousands to millions of pixels in an image
- 30+ degrees of freedom in the pose of articulated objects (humans)
- 3,000-30,000 human recognizable object categories

Challenges: learning with minimal supervision

What kinds of things work best today?



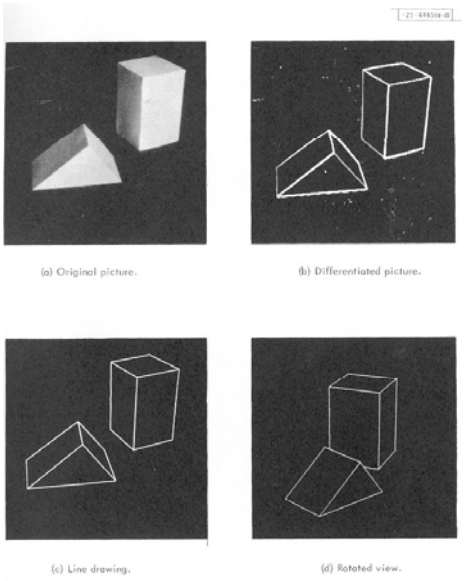
Reading license plates, zip codes, checks

Frontal face detection

Recognizing flat, textured objects (like books, CD covers, posters)

Fingerprint recognition

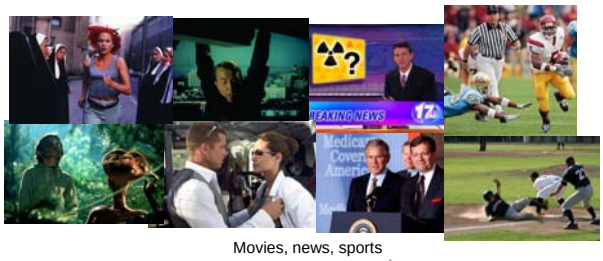
Inputs in 1963…



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

… and inputs today

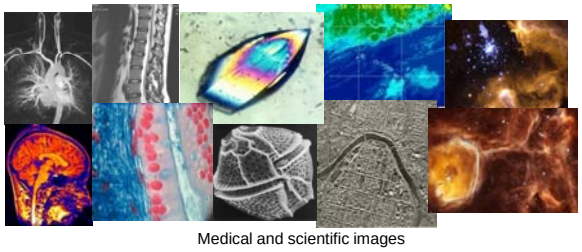
Personal photo albums



Movies, news, sports

Google Image Search Picasa™ flickr™ webshots picsearch™ YouTube Broadcast Yourself™

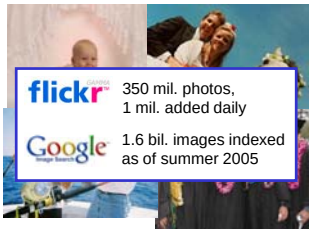
Surveillance and security

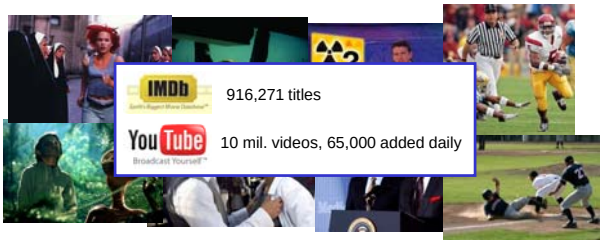


Medical and scientific images

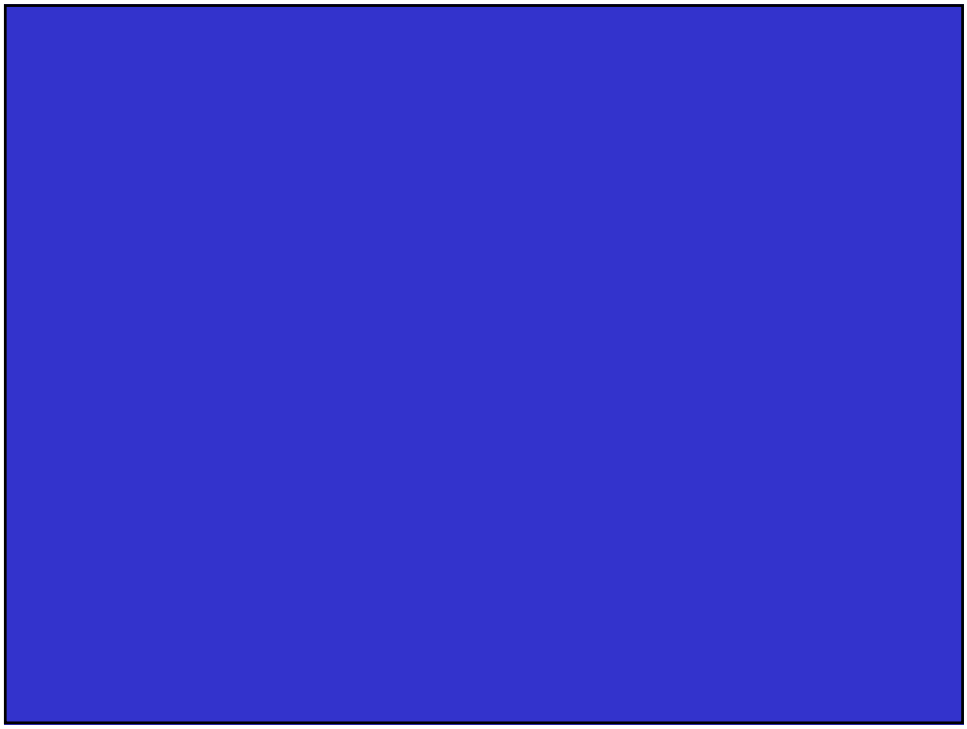
Slide credit; L. Lazebnik

… and inputs today





introductions



This course

- Focus on current research in
  - Object recognition and categorization
  - Image/video retrieval, annotation
  - Activity recognition
- High-level vision and learning problems, innovative applications.

Goals

- Understand current approaches
- Analyze
- Identify interesting research questions

Expectations

- **Discussions** will center on recent papers in the field
  - Paper reviews each week
- **Student presentations**
  - Papers and background reading
  - Experiment presentation
- **2 implementation assignments**
- **Project**

Workload is fairly high

Prerequisites

- Courses in:
  - Computer vision
  - Machine learning
- Ability to analyze high-level conference papers

Paper reviews

- Each week, review two of the assigned papers.
- Email me and TA by Thurs 9 PM
- Skip reviews the week(s) you are presenting.

Paper review guidelines

- Brief (2-3 sentences) summary
- Main contribution
- Strengths? Weaknesses?
- How convincing are the experiments? Suggestions to improve them?
- Extensions?
- Additional comments, unclear points
- Relationships observed between the papers we are reading

Paper presentation guidelines

- Read 3 selected papers in topic area
- Well-organized talk, about 30-45 minutes

- What to cover?
  - Problem overview, motivation
  - Algorithm explanation, technical details
  - Any commonalities, important differences between techniques covered in the papers.

- See handout and class webpage for more details.

Experiment guidelines

- Implement/download code for a main idea in the paper and show us toy examples:
  - Experiment with different types of (mini) training/testing data sets
  - Evaluate sensitivity to important parameter settings
  - Show (on a small scale) an example to analyze a strength/weakness of the approach
- Present in class – about 30 minutes.
- Share links to any tools or data.

Timetable for presenters

- For papers or experiments, by the Friday **the week before** your presentation is scheduled:
  - Email draft slides to me, and schedule a time to meet, do dry run, discuss.
  - This is a hard deadline: 5 points off automatically per day late
- See course webpage for examples of good reviews, presentations.

Projects

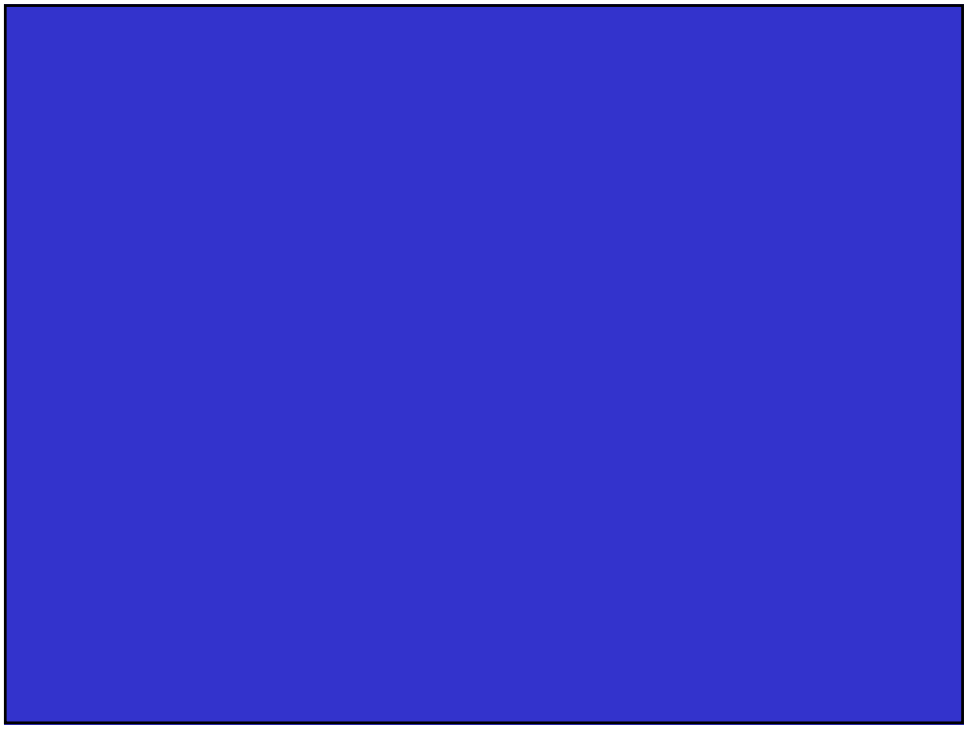
Possibilities:

- Extend a technique studied in class
- Analysis and empirical evaluation of an existing technique
- Comparison between two approaches
- Design and evaluate a novel approach
- Thorough survey / review paper

- Work in pairs, except for survey.

Miscellaneous

- Feedback welcome and useful
- No laptops, phones, etc. in class please
- Check class website
- I'll use Blackboard to email class



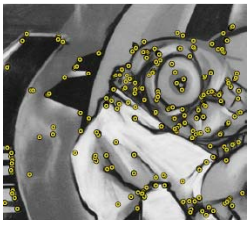
Syllabus tour

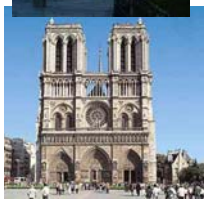
I. Object recognition fundamentals
II. Beyond modeling individual objects
III. Human-centered recognition

Syllabus tour

I. Object recognition fundamentals
   A. Local features and matching object instances
   B. Large-scale search and mining
   C. Classification and detection of categories
   D. Mid-level representations

Local features and matching object instances





Local invariant features, detection and description

Matching models to images

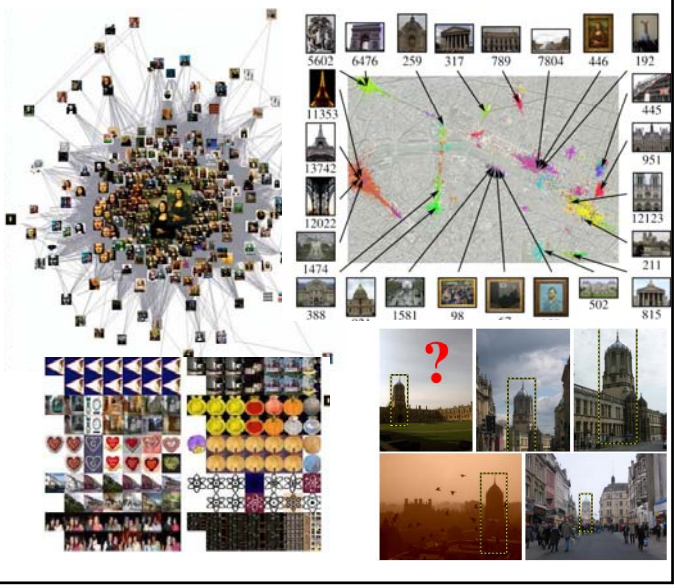
Indexing specific objects with bag-of-words descriptors

Large-scale image/object search and mining

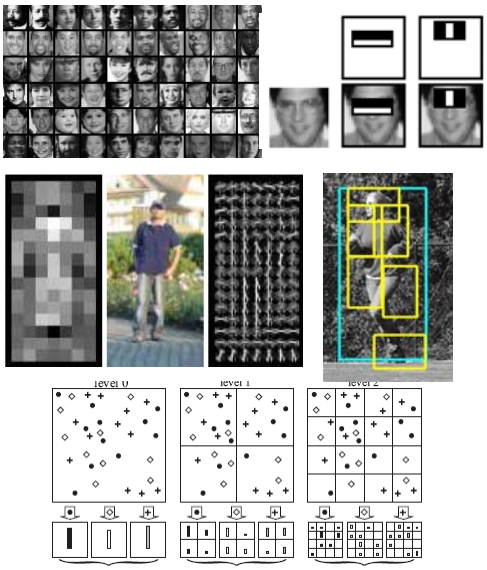
Using instance recognition for large-scale search

Scalable hashing algorithms

Adopting text retrieval insights



Classification and detection for object categories



Detection as classification problem

Discriminative methods

Global representations with rigid spatial

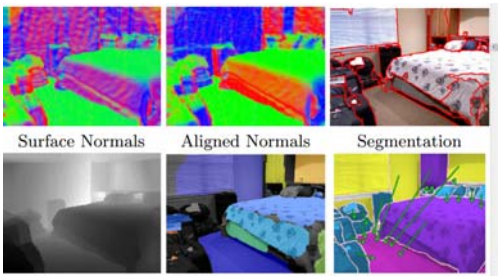
Faces and pedestrians as case studies

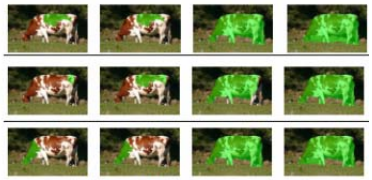
Mid-level representations

Segmentation

Category-independent region ranking

Surface estimation



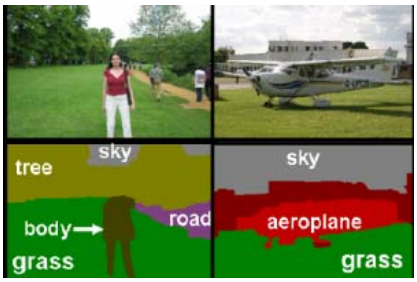


Syllabus tour

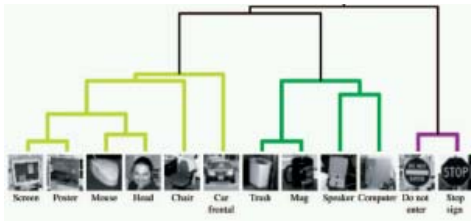
II. Beyond modeling individual objects
   - A. Context and scenes
   - B. Dealing with many categories
   - C. Describing objects with attributes
   - D. Importance and saliency

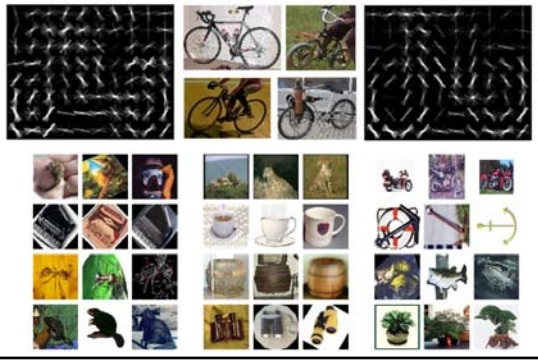
Context and scenes

The scene, the other objects, the spatial layout, geometry of surfaces --- all tell us more about what is reasonable to detect.



Dealing with many categories





Sharing features between classes

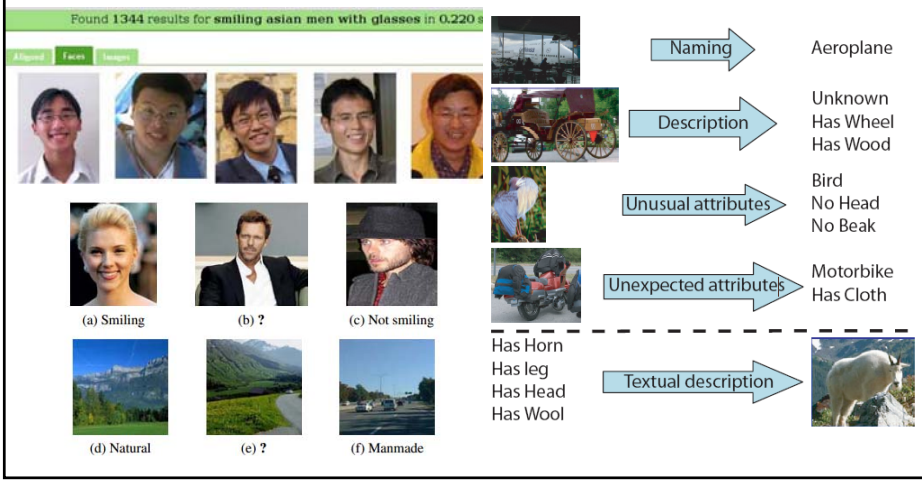
Transfer learning

Learning from few examples

Category hierarchies

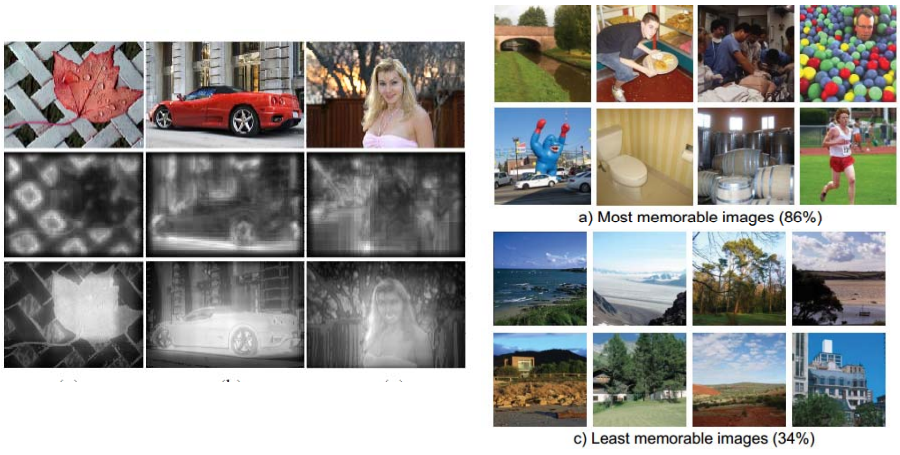
Describing objects with attributes

Beyond naming object by category, we should be able to describe their properties, or use descriptions to understand novel objects.



Saliency and importance

Among all items in the scene, which deserve attention (first)? What makes images interesting or memorable?



Syllabus tour

III. Human-centered recognition

A. Pictures of people

B. Activity recognition

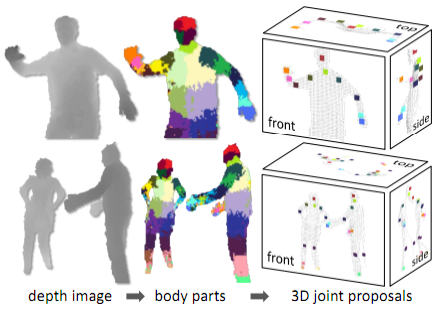
C. Egocentric cameras

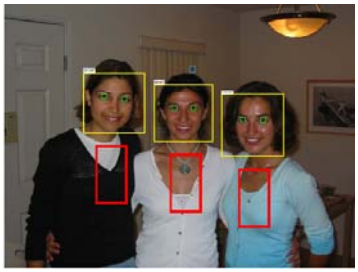
D. Human-in-the-loop interactive systems

Pictures of people

Finding people and their poses

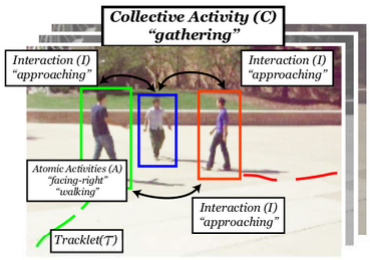
Automatic face tagging





Activity recognition

Recognizing human actions in images and video



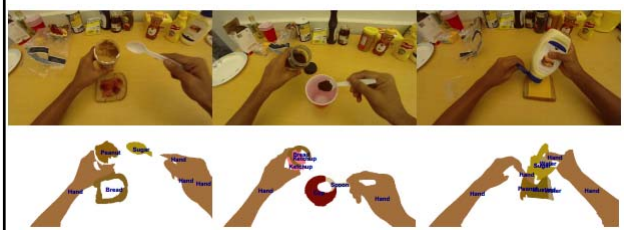
Egocentric cameras

Recognizing objects and actions from a first person point of view

Summarization

Input: *Egocentric video of the camera wearer's day*

Output: *Storyboard summary of important people and objects*



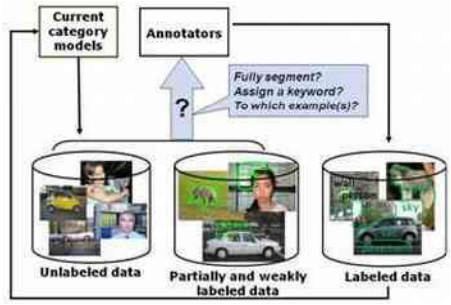
Human-in-the-loop interactive systems

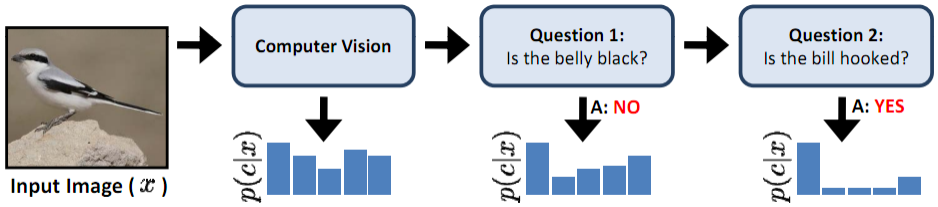
Human-in-the-loop learning

Active annotation collection

Crowdsourcing

amazonmechanical turk
Artificial Artificial Intelligence





Not covered

- Low-level image processing
- Basic machine learning methods

- I will assume you already know these, or are willing to pick them up on your own.

Coming up

- Talk next Friday at 11:30 am in ACES 2.402:
  Silvio Savarese, Univ. of Michigan
  "Understanding the 3d world from images"
- Review syllabus, select 4 topic preferences
  - Email to Austin (TA) by Wed Sept 5 at 5 pm
- Read assigned papers for "local features and matching for object instances", and review the Sivic and Lowe papers.