Visual Recognition & Search

January 22, 2009

Introductions

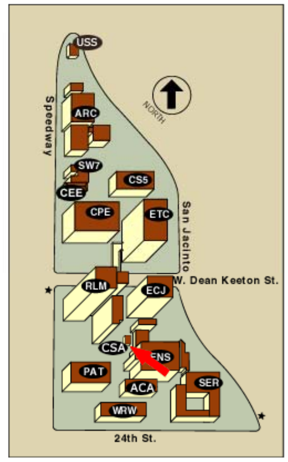
- **Class**: Thursday 3:30-6:30 PM
- **Instructor**: Kristen Grauman
  grauman at cs.utexas.edu
  CSA 114
- **Office hours**: by appointment
- TA Harshdeep Singh

- **Class page**: link from
  http://www.cs.utexas.edu/~grauman/
  Check for updates to schedule.

CSA

Computer Science Annex - CSA

Engineering Area



My office : CSA 114

Plan for today

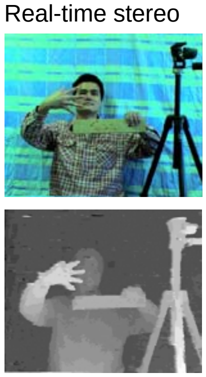
- Topic overview: What is visual recognition and search? Why are these hard problems? What sorta works?
- Course overview: Requirements, syllabus tour

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*)

Vision for measurement

Real-time stereo



Wang et al.

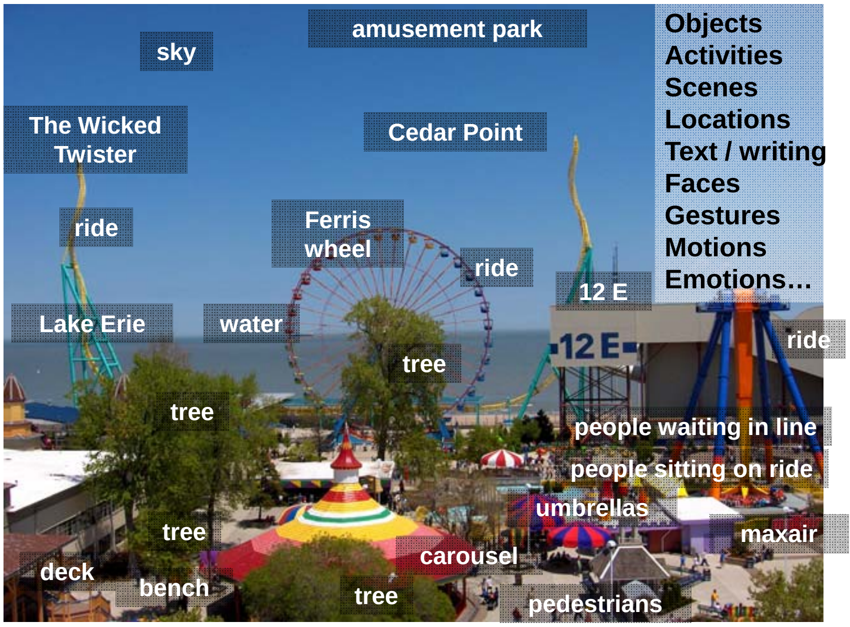
Structure from motion

Snavely et al.

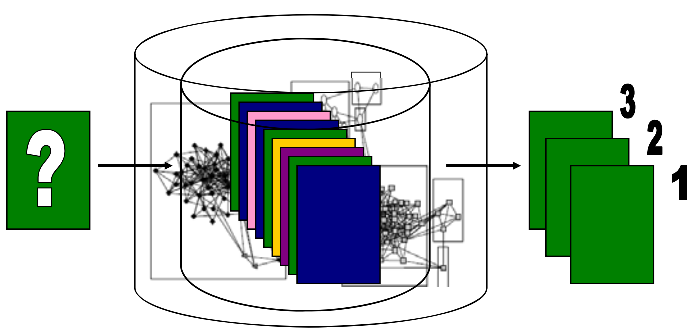
Tracking

Demirdjian et al.

Vision for perception, interpretation



Visual search, organization



Query

Image or video archives

Relevant content

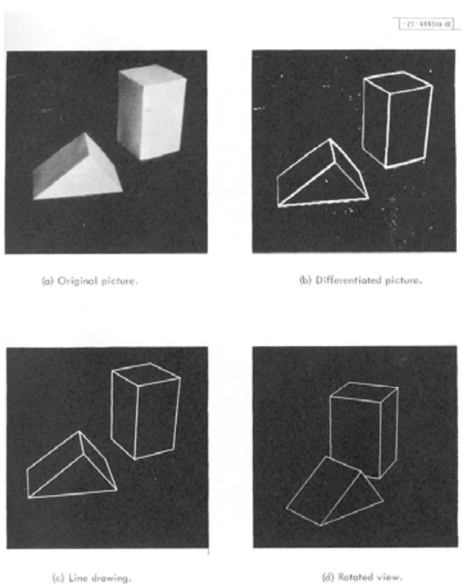
Why recognition and search?

– 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?

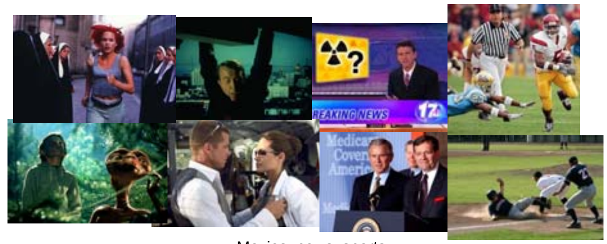
Vision in 1963



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

Today: visual data in the wild

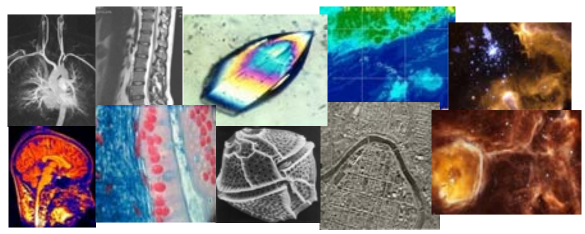
Personal photo albums



Movies, news, sports

Google Image Search Picasa flickr webshots picsearch YouTube

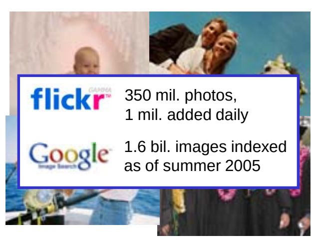
Surveillance and security

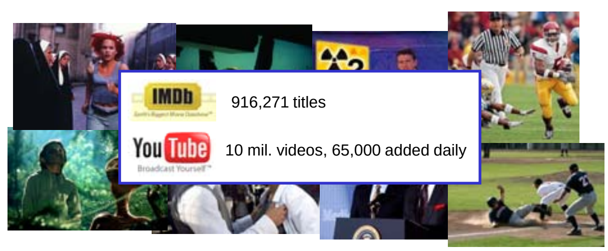


Medical and scientific images

Slide credit; L. Lazebnik

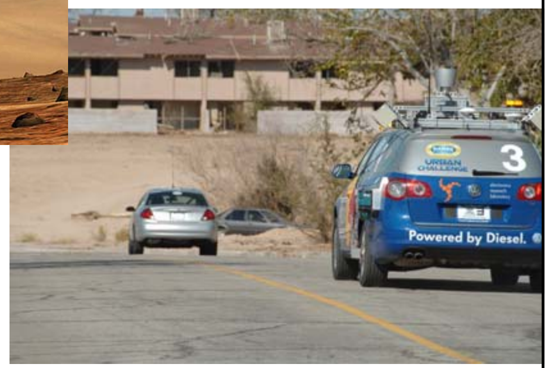
Today: visual data in the wild





Slide by Lana Lazebnik

Autonomous agents able to detect objects



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

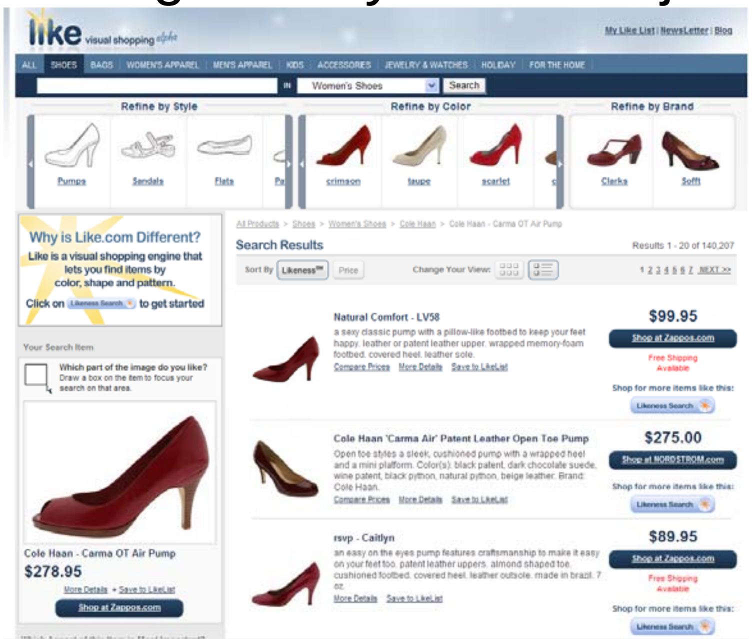
Linking to info with a mobile device

Situated search
Yeh et al., MIT

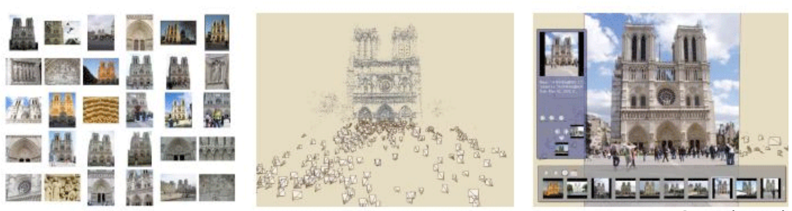
MSR Lincoln

kooaba

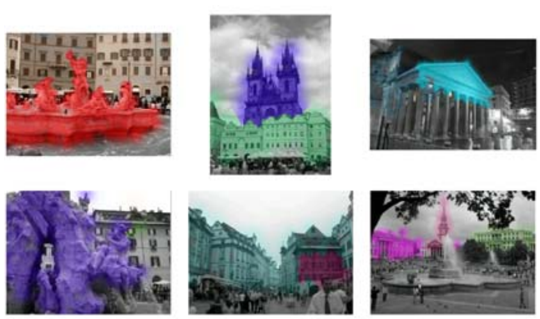
Finding visually similar objects



Exploring community photo collections

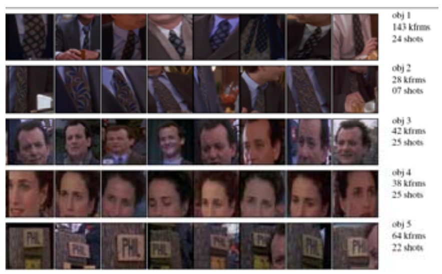


Snavely et al.

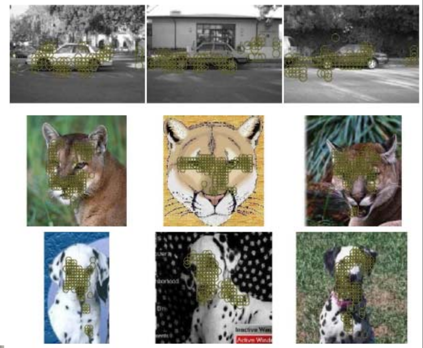


Simon & Seitz

Discovering visual patterns



Objects Sivic & Zisserman



Lee & Grauman
Categories

Actions Wang et al.

Plan for today

- Topic overview: **What is visual recognition and search? Why are these hard problems?** What sorta works?
- Course overview: Requirements, syllabus tour

The Instance-Level Recognition Problem

John’s car

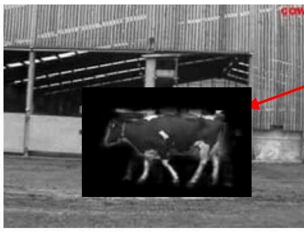
The Categorization Problem

- How to recognize ANY car





Levels of Object Categorization



"cow"

"car"

"motorbike"

- Different levels of recognition
  - *Which* object class is in the image? ⇒ Obj/Img classification
  - *Where* is it in the image? ⇒ Detection/Localization
  - *Where exactly* — which pixels? ⇒ Figure/Ground segmentation

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

Visual Object Categories

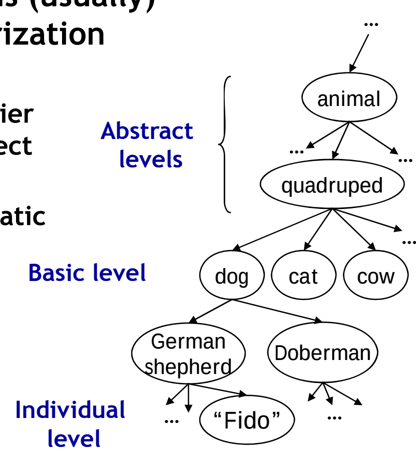
- Basic Level Categories in human categorization [Rosch 76, Lakoff 87]
  - The highest level at which category members have similar perceived shape
  - The highest level at which a single mental image reflects the entire category
  - The level at which human subjects are usually fastest at identifying 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

K. Grauman, B. Leibe

Visual Object Recognition Tutorial

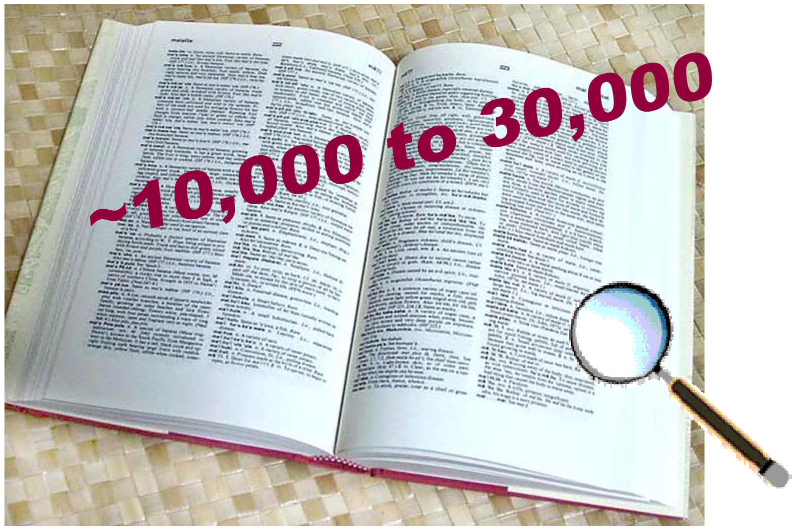
Visual Object Categories

- 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

Other Types of Categories

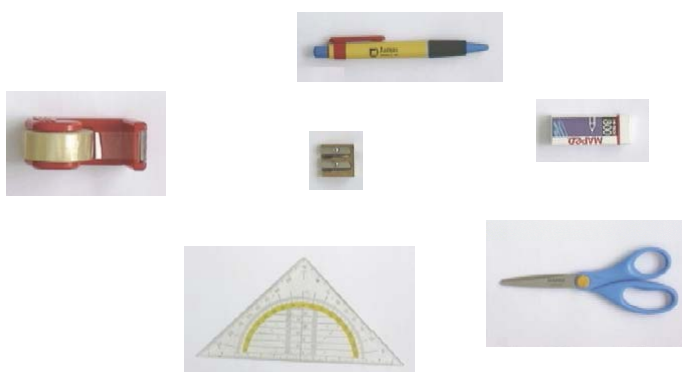
- **Functional Categories**
  - **e.g. chairs =** *"something you can sit on"*

Visual Object Recognition Tutorial

K. Grauman, B. Leibe

Other Types of Categories

- **Ad-hoc categories**
  - **e.g.** *"something you can find in an office environment"*



Visual Object Recognition Tutorial

K. Grauman, B. Leibe

Challenges: robustness

Illumination

Object pose

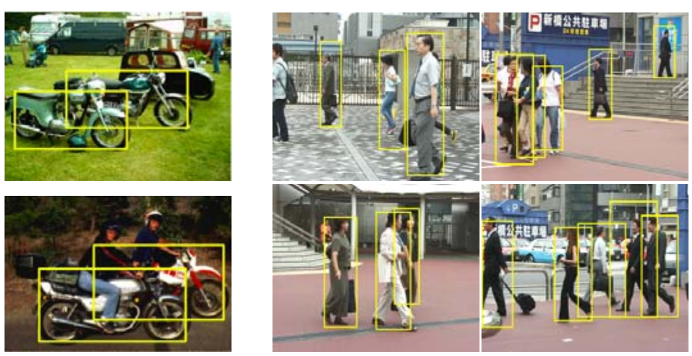
Clutter

Occlusions

Intra-class appearance

Viewpoint

Challenges: robustness

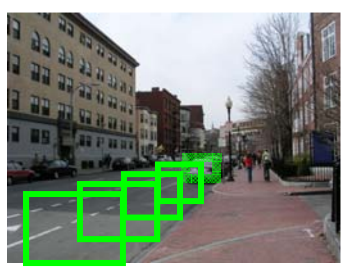


Realistic scenes are crowded, cluttered, have overlapping objects.

Challenges: importance of context

slide credit: Fei-Fei, Fergus & Torralba

Challenges: importance of context



Challenges: complexity

- Thousands to millions of pixels in an image
- 3,000-30,000 human recognizable object categories
- 30+ degrees of freedom in the pose of articulated objects (humans)
- Billions of images indexed by Google Image Search
- 18 billion+ prints produced from digital camera images in 2004
- 295.5 million camera phones sold in 2005
- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991]

Challenges: learning with minimal supervision

What “works” today

- Reading license plates, zip codes, checks

Source: Lana Lazebnik

What “works” today

- Reading license plates, zip codes, checks
- Fingerprint recognition

Source: Lana Lazebnik

What “works” today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection

[Face priority AE] When a bright part of the face is too bright

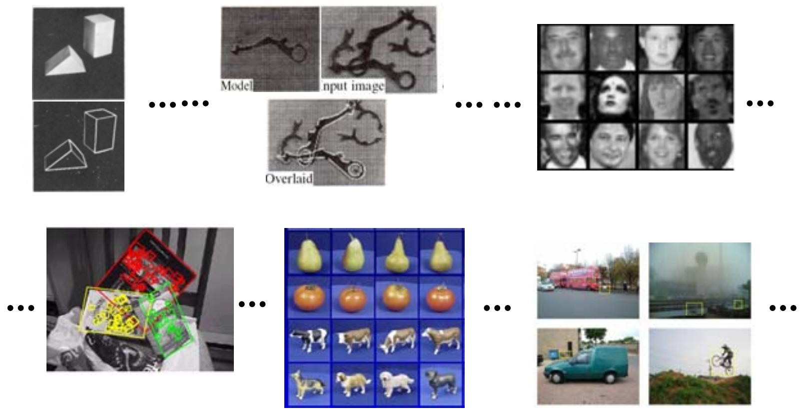
Source: Lana Lazebnik

What “works” today

- Reading license plates, zip codes, checks
- Fingerprint recognition
- Face detection
- Recognition of flat textured objects (CD covers, book covers, etc.)

Source: Lana Lazebnik

- Active research area with exciting progress!



Today's challenge

This course

- Focus on current research in
  - visual category and object recognition
  - image/video retrieval
  - organization, exploration, interaction with visual content
- 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
- **Student presentations**
  - Papers and background reading
  - Demos
- **Projects**
  - Research-oriented
- **Workload = reasonably high**

Prerequisites

- Courses in:
  - Computer vision
  - Machine learning
  - Basic probability
  - Linear algebra
- Ability to analyze high-level conference papers

Paper reviews

- For each class, review two of the assigned papers.
- Post by Wed night 10 PM on Google docs (instructions are on Blackboard)

- Don’t review papers 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
- ½ page to 1 page.

Presentation guidelines

- Read 3-4 selected papers in topic area
- Well-organized talk, about 30 minutes

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

- See class webpage for more details.

Demo 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 in practice that highlights a strength/weakness of the approach
- Present in class – about 20-30 minutes.
- Post webpage with links to any tools or data.

Timetable for presenters

- By the Thursday **the week before** your presentation is scheduled:
  - Email draft slides to me, and schedule a time to meet and discuss.
- The week of your presentation:
  - Refine slides, practice presentation, know about how long each part requires.
- The day of your presentation:
  - Send final slides (and, for demos, pointer to webpage) to me.

Presenter feedback

- Preparedness
- Coverage of topic
- Organization and clarity of presentation
- Enthusiasm, use of engaging examples
- Serves to start discussion, quality of discussion points raised

Demo feedback

- Preparedness
- Clarity of message and organization
- Technical detail and relevance to reading
- Enthusiasm, use of engaging examples

Projects

Possibilities:

- Extend a technique studied in class
- Analysis and empirical evaluation of a technique
- Comparison between two approaches
- Design and evaluate a novel approach

- Work in pairs

Grading policy

- 20% participation
  - includes attendance and paper reviews
- 20% demo
- 20% paper presentation
- 40% project

Important dates

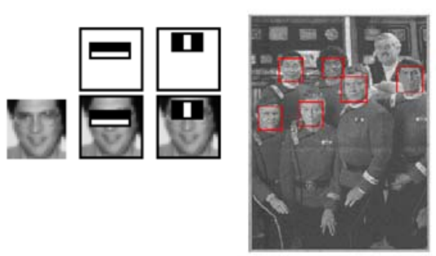
- March 26 : project proposals due (tentative)
- April 16 : project progress report / draft (tentative)
- May 7 : Final project papers due
- May 7 and May 8 : Final presentations

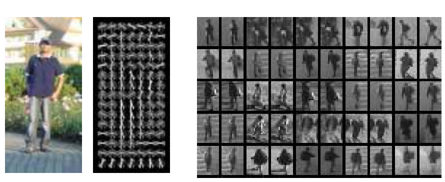
  - May 8 is Friday after last class.

Syllabus tour

I. <u>Categorizing and matching objects</u>
II. Surrounding cues
III. Data-driven visual learning
IV. Searching and browsing visual content

Sliding windows and global representations





- Sliding window protocol for detection
- Good features for “patch” appearance, global descriptors
- Building detectors with discriminative classifiers
- Faces, pedestrians as case studies

- (Next week)

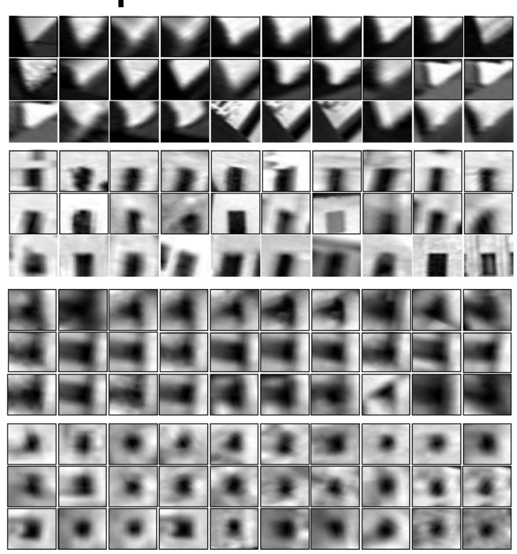
Distances and kernels, bags of words representations

Local features: interest operators and descriptors

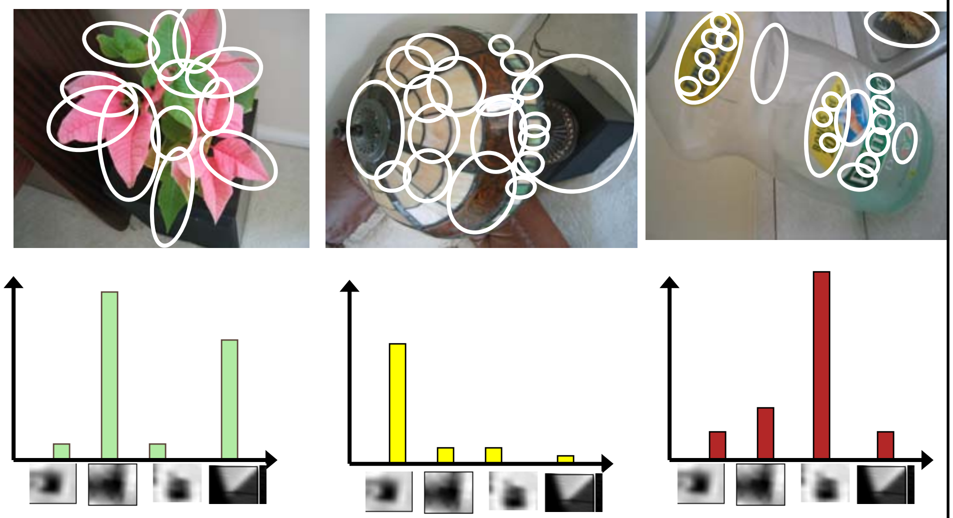
How to summarize local content?
How to *match* or compare images with local descriptors?

Distances and kernels, bags of words representations

- Constructing a visual "vocabulary"



Distances and kernels, bags of words representations



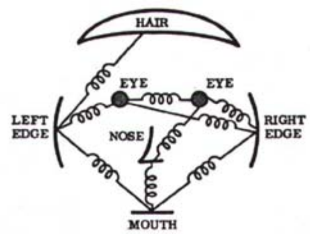
Distances and kernels, bags of words representations

Local features: interest operators and descriptors

Correspondence kernels
- how to compute matches efficiently?

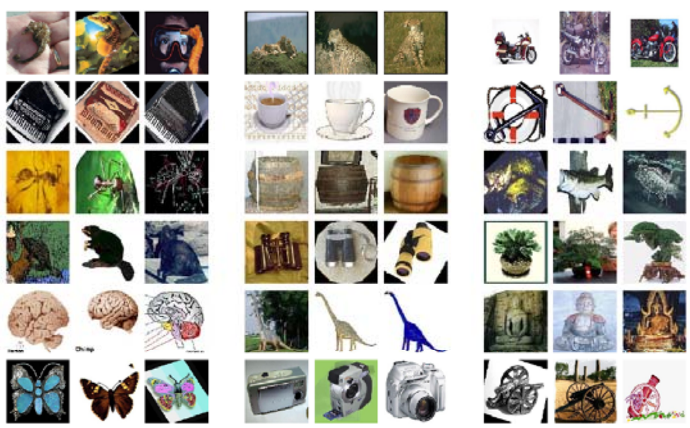
Learning feature significance
- which features are most discriminative?

Part-based models



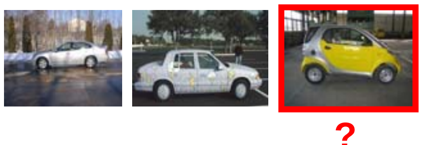
- Representing part appearance plus structure
- Summarizing repeated parts
- Efficient matching

Image annotation process



Classifiers can be trained from labeled data…

Image annotation process

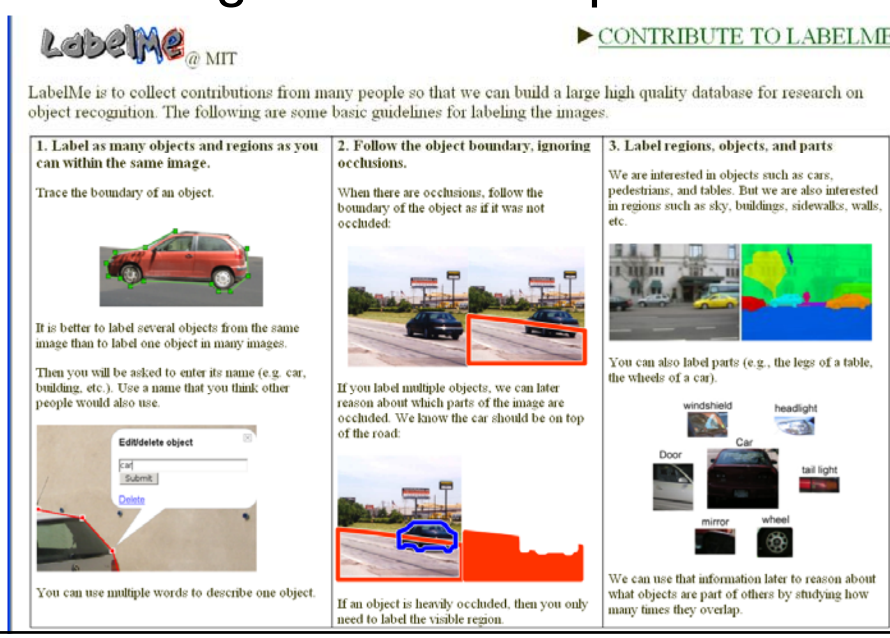


Von Ahn et al.

Rother et al.

- What data should be labeled?
- How can the task be streamlined with semi-automatic tools?
- How can it be more enticing?
- What makes an image dataset useful/not so useful?

Image annotation process

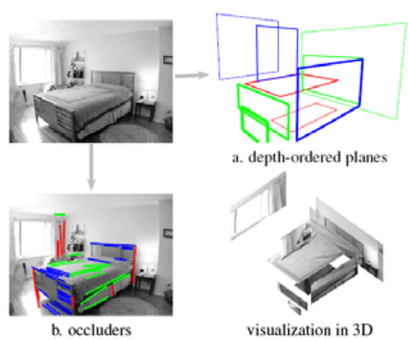


Syllabus tour

I. Categorizing and matching objects
II. Surrounding cues
III. Data-driven visual learning
IV. Searching and browsing visual content

Inferring 3D cues from single images

Hoiem et al.



Yu et al.

Geometric context is important to scene understanding.

- What are the primary surfaces and their orientations?
- How can this be inferred with a single snapshot?

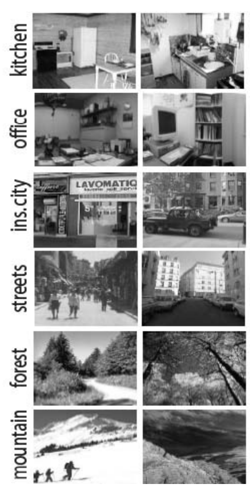
Scene recognition



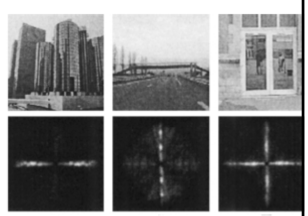
Scene recognition

Many objects occur only in certain scenes, and scene types are a useful summary of a shot.

- What kind of scene is it? Indoor/outdoor, city/mountain?
- Holistic representations for scenes



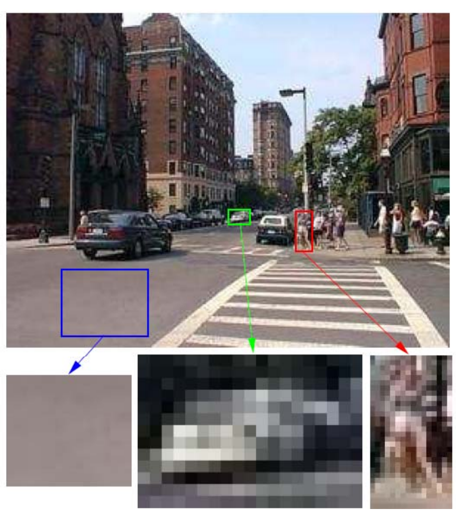
FeiFei & Perona


Oliva & Torralba

Context

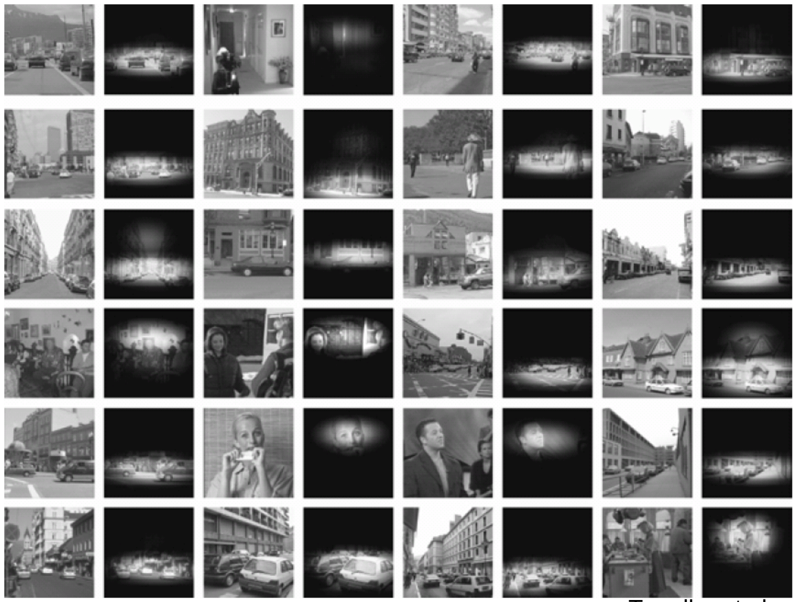
- The context of the scene, the other objects, and the spatial layout could tell us a lot about what is reasonable to detect.

Context



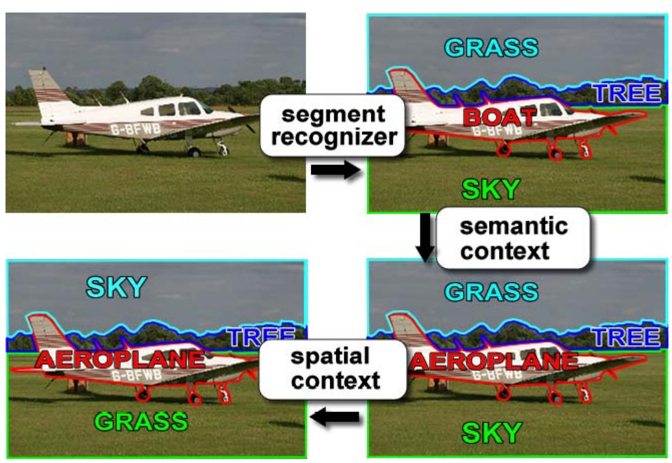
Hoiem et al.

Context



Torralba et al.

Context



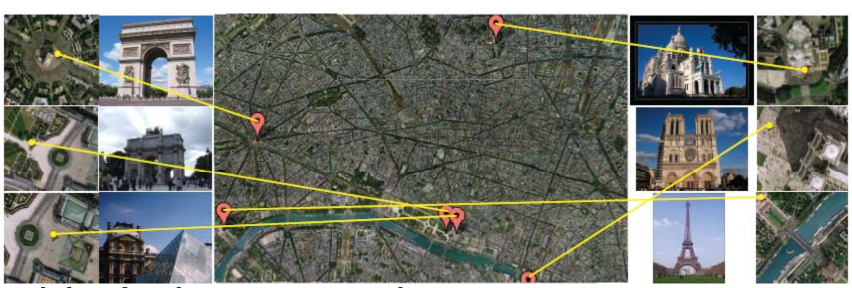
Galleguillos et al.

Syllabus tour

I. Categorizing and matching objects
II. Surrounding cues
III. Data-driven visual learning
IV. Searching and browsing visual content

Leveraging internet data

- The internet offers unprecedented access to lots of data: both images and surrounding cues.

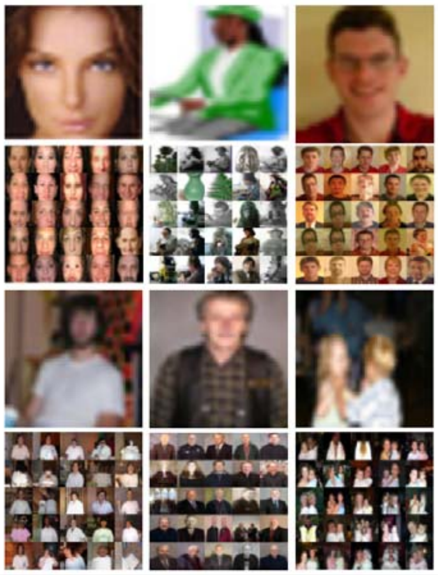


Mining for themes, connecting to geo-tags

Quack et al.

Leveraging internet data

The value of volume



Torralba et al.

Dealing with noisy sources

Text, language, and imagery

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

World number one **Lleyton Hewitt** of Australia hits a return to **Nicolas Massu** of Chile at the Japan Open tennis championships in Tokyo October 3, 2002. REUTERS/Eriko Sugita

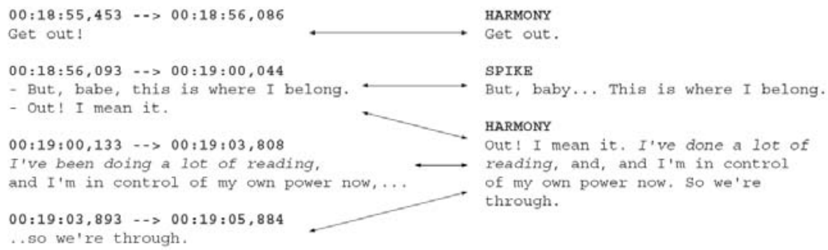
British director **Sam Mendes** and his partner actress **Kate Winslet** arrive at the London premiere of 'The Road to Perdition', September 18, 2002. The films stars **Tom Hanks** as a Chicago hit man who has a separate family life and co-stars **Paul Newman** and Jude Law. REUTERS/Dan Chung

German supermodel **Claudia Schiffer** gave birth to a baby boy by Caesarian section January 30, 2003, her spokeswoman said. The baby is the first child for both Schiffer, 32, and her husband, British film producer **Matthew Vaughn**, who was at her side for the birth. Schiffer is seen on the German television show 'Bet It...?!' ('Wetten Dass...?!') in Braunschweig, on January 26, 2002. (Alexandra Winkler/Reuters)

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)

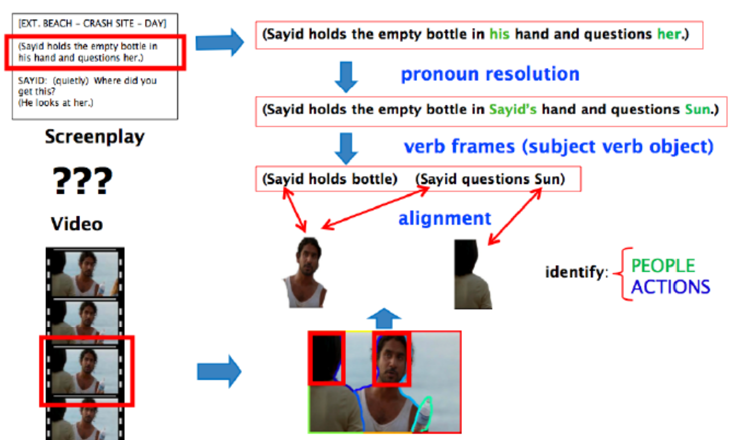
US **President George** W. Bush (L) makes remarks while Secretary of **State Colin Powell** (R) listens before signing the US Leadership Against HIV /AIDS , Tuberculosis and Malaria Act of 2003 at the Department of State in Washington, DC. The five-year plan is designed to help prevent and treat AIDS, especially in more than a dozen African and Caribbean nations(AFP/Luke Frazza)

Text, language, and imagery



Everingham et al.

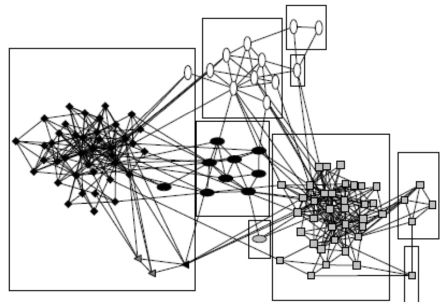
Text, language, and imagery



Cour et al.

Unsupervised learning and discovery

- What are common visual patterns?
- What is unusual, or salient?



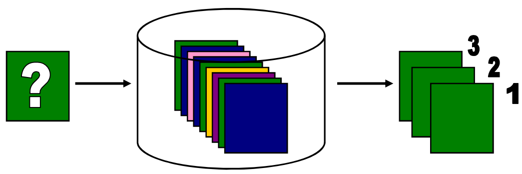
Boiman et al.

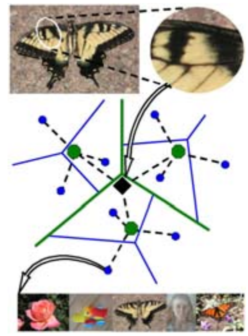
Syllabus tour

I. Categorizing and matching objects
II. Surrounding cues
III. Data-driven visual learning
IV. Searching and browsing visual content

Fast indexing and search

- With large archives, how to access the relevant content rapidly with good image metrics?

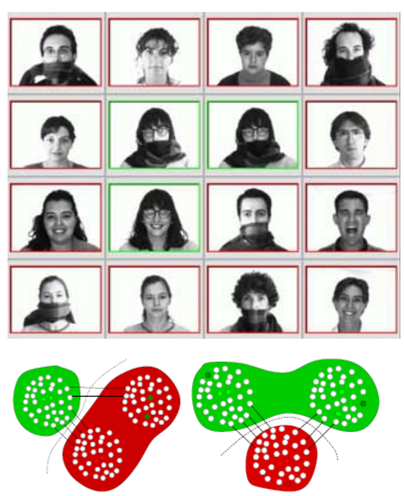




Nister et al.

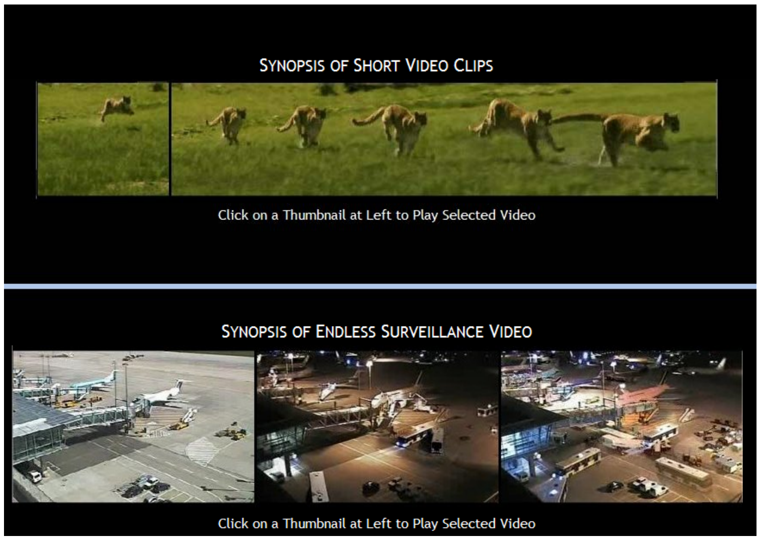
Browsing: query refinement and summarization

- How will a user peruse resulting content efficiently?
- How can a user intervene in the search process?
- Visualizing the aggregation of multiple users' photos

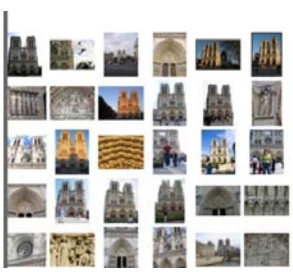


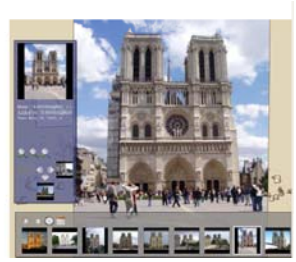
Sahbi et al.

Browsing: query refinement and summarization



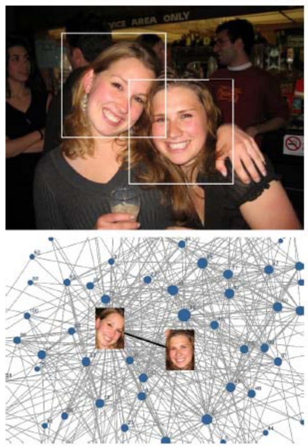
Browsing: query refinement and summarization





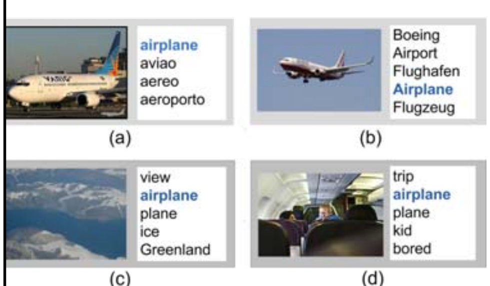
Snavely et al.

Social networks and image tagging



- What information (helpful for recognition) does a community of users provide?
- Why and how do people contribute tags?
- When do they agree? What is objective?

Social networks and image tagging



- What information (helpful for recognition) does a community of users provide?
- Why and how do people contribute tags?
- When do they agree? What is objective?

Not covered in this course

- Low-level processing
- Basic machine learning methods

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

Schedule

22-Jan	**Introduction**	
29-Jan	**Categorizing and matching objects**	Global appearance, window-based recognition
5-Feb		Distances and kernels
12-Feb		Part-based models
19-Feb		Image annotation process
26-Feb	**Surrounding cues**	Inferring 3d cues from a single image
5-Mar		Scene recognition
12-Mar		Context
19-Mar	***Spring break - no class***	
26-Mar	**Data-driven visual learning**	Leveraging internet data
2-Apr		Text, language, and imagery
9-Apr		Unsupervised learning and discovery
16-Apr	**Searching and browsing visual content**	Fast indexing and search
23-Apr		Browsing: query refinement and summarization
30-Apr		Social networks and image tagging
7-May	**Final project presentations**	
8-May	**Final project presentations**	

For next week

- Read and review:
  - Viola & Jones, CVPR 2001
  - Dalal & Triggs, CVPR 2005
  - **Review syllabus, select topic preferences (3 for demo, 3 for paper topics)**
    - **Email me by Monday.**

- First student presenters will be on Feb 5.