Visual Object Recognition and Image Search

Monday, June 27

Kristen Grauman
UT-Austin
Intro

• Topic overview:
  – What does the visual recognition problem entail?
  – Why are these hard problems?
  – What works today?

• Course plan
  – Format and syllabus overview
  – Requirements, grading
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?
Verification: is that a lamp?
Detection: where are there people?
Identification: is that Potala Palace?
Object categorization

- mountain
- tree
- building
- banner
- street lamp
- vendor
- people
Scene and context categorization

- outdoor
- city
- ...

![Outdoor city scene](image)
Instance-level recognition problem

John’s car
Generic categorization problem
How many object categories are there?

~10,000 to 30,000

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

Biederman 1987
Other Types of Categories

- Functional Categories
  - e.g. chairs = “something you can sit on”
Other Types of Categories

• Ad-hoc categories
  - e.g. “something you can find in an office environment”
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

Posing visual queries

Yeh et al., MIT

Belhumeur et al.

Kooaba, Bay & Quack et al.
Finding visually similar objects
Exploring community photo collections

Snavely et al.

Kristen Grauman

Simon & Seitz
Discovering visual patterns

Objects - Sivic & Zisserman

Categories - Lee & Grauman

Actions - Wang et al.

Kristen Grauman
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.

T. Berg 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 film 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 size) 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 Raynor/Reuters)

Kristen Grauman
Challenges
Challenges: robustness

- Illumination
- Object pose
- Clutter
- Occlusions
- Intra-class appearance
- Viewpoint

Kristen Grauman
Challenges: robustness

Realistic scenes are crowded, cluttered, have overlapping objects.
Challenges: importance of context
Challenges: importance of context
Challenges: complexity

- Thousands to millions of pixels in an image
- 3,000-30,000 human recognizable object categories
- ~20 hours of video/minute added to YouTube
- ~5,000 new tagged photos/minute added to Flickr
- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman and van Essen 1991]
Challenges: learning with minimal supervision

- Less supervision
  - Unlabeled, multiple objects

- More supervision
  - Classes labeled, some clutter
  - Cropped to object, parts and classes

Kristen Grauman
What works most reliably today

- Reading license plates, zip codes, checks

Source: Lana Lazebnik
What works most reliably today

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

Source: Lana Lazebnik
What works most reliably today

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

Source: Lana Lazebnik
What works most reliably today

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

Source: Lana Lazebnik
Inputs in 1963...

... and inputs today

- Personal photo albums
- Movies, news, sports
- Surveillance and security
- Medical and scientific images

Slide credit: L. Lazebnik
introductions
This course

• Basic background
• Focus on current research in
  – Object recognition and categorization
  – Image/video retrieval, annotation

• High-level vision and learning problems, innovative applications.
Welcome to the Web site for my computer vision textbook, which you can now purchase at a variety of locations, including Springer, Amazon, and Barnes & Noble.

This book is largely based on the computer vision courses that I have co-taught at the University of Washington (2008, 2005, 2001) and Stanford (2003) with Steve Seitz and David Fleet.

You are welcome to download the PDF from this Web site for personal use, but not to repost it on any other Web site. Please post a link to this URL instead. An electronic manuscript will continue to be available even after the book is published. Note, however, that while the content of the electronic and hardcopy versions are the same, the page layout (pagination) electronic version is optimized for online reading.

The PDFs should be enabled for commenting directly in your viewer. Also, hyper-links to sections, equations, and references are enabled. To get back to where you were, use Alt-Left-Arrow or
Monday

- Low-level feature extraction
  - Filters, edges, local invariant features
Tuesday

- Mid-level representation
  - Segmentation, grouping, fitting
Wednesday

• Instance recognition
  – Retrieving and matching same objects
Thursday

- Category recognition
  - Learning and detection generic categories
Friday

• Visual search and mining
This course

• Monday: Low-level feature extraction
  – Filters, edges, local invariant features
• Tuesday: Mid-level representation
  – Segmentation, grouping, fitting
• Wednesday: Instance recognition
  – Retrieving and matching same objects
• Thursday: Category recognition
  – Learning and detection generic categories
• Friday: Visual search and mining
Background assumed

- Basic linear algebra, probability, familiarity with machine learning, algorithms, programming.

- Image processing and vision experience helpful, but I will cover the basics.
Requirements and grades

• participation, exercises in class (15%)
• 2 paper reviews due Monday through Thursday evening 9 pm via email (25%)
• one coding assignment (25%) due Friday July 8.
• final exam Friday July 1 in class (35%)
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
- See course website for examples
Miscellaneous

• Feedback welcome and useful
• Check class website
• Put your email on the list
• No open laptops/phones/etc. in class please