ECS 289H: Visual Recognition
Fall 2014

Yong Jae Lee
Department of Computer Science
Plan for today

• Topic overview
• Introductions
• Course requirements and administrivia
• Syllabus tour (if time permits)
Computer Vision

• Let computers see!

• Automatic understanding of visual data (i.e., 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*)
What does recognition involve?
Detection: are there people?
Activity: What are they doing?
Object categorization

- mountain
- building
- tree
- banner
- street lamp
- vendor
- people
Instance recognition

Potala Palace

A particular sign
Scene categorization

- outdoor
- city
- ...

[Image of a cityscape with a mountainous background and a busy streetscape]
Attribute recognition

- gray
- made of fabric
- crowded
- flat
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?
Posing visual queries

Yeh et al., MIT

Digital Field Guides Eliminate the Guesswork

Belhumeur et al.

Kooaba, Bay & Quack et al.

Kristen Grauman
Interactive systems

Shotton et al.
Autonomous agents

Mars rover

Google self-driving car
Exploring community photo collections

Snavely et al.

Kristen Grauman

Simon & Seitz
Discovering visual patterns

Actions

Wang et al.

Categories

Doersch et al.

Lee & Grauman

Characteristic elements
What kinds of things work best today?

Recognizing flat, textured objects (e.g. books, CD covers, posters)

Reading license plates, zip codes, checks

Frontal face detection

Fingerprint recognition

Kristen Grauman
What are the challenges?
Challenges: robustness

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

Kristen Grauman
Challenges: context and human experience

Context cues

Kristen Grauman
Challenges: context and human experience

Context cues

Function

Dynamics

Kristen Grauman
Challenges: context and human experience

Fei Fei Li, Rob Fergus, Antonio Torralba
Challenges: context and human experience

Fei Fei Li, Rob Fergus, Antonio Torralba
Challenges: context and human experience

Fei Fei Li, Rob Fergus, Antonio Torralba
Challenges: scale, efficiency

• Half of the cerebral cortex in primates is devoted to processing visual information
Challenges: scale, efficiency

flickr: 6 billion images

Facebook: 70 billion images

Imgur: 1 billion images served daily

YouTube: 10 billion images

100 hours uploaded per minute

From Cisco: Almost 90% of web traffic is visual!
Challenges: scale, efficiency

~10,000 to 30,000 object categories

Fei Fei Li, Rob Fergus, Antonio Torralba
Challenges: learning with minimal supervision

Less

Unlabeled, multiple objects

Classes labeled, some clutter

More

Cropped to object, parts and classes

Kristen Grauman
This is a pottopod
Find the pottopod
Inputs in 1963...

... and inputs today

Understand and organize and index all this data!!

Surveillance and security

Medical and scientific images

Svetlana Lazebnik
Introductions
Course Information

• Website: https://sites.google.com/a/ucdavis.edu/ecs-289h-visual-recognition/

• Office hours: by appointment (email)

• Email: start subject with “[ECS289H]”
This course

• Survey state-of-the-art research in computer vision

• High-level vision and learning problems
Goals

• Understand, analyze, and discuss state-of-the-art techniques

• Identify interesting open questions and future directions
Prerequisites

• Interest in computer vision!

• Courses in computer vision, machine learning, image processing is a plus (but not required)
Requirements

• Class Participation [25%]

• Paper Reviews [25%]

• Paper Presentations (1-2 times) [25%]

• Final Project [25%]
Class participation

• Actively participate in discussions
• Read all assigned papers before each class
Paper reviews

• Detailed review of one paper before each class
• One page in length
  – A summary of the paper (2-3 sentences)
  – Main contributions
  – Strengths and weaknesses
  – Experiments convincing?
  – Extensions?
  – Additional comments, including unclear points, open research questions, and applications

• Email by 11:59 pm the day before each class
• Email subject “[ECS 289H] paper review”
Paper presentations

• 1 or 2 times during the quarter
• ~20 minutes long, well-organized, and polished:
  – Clear statement of the problem
  – Motivate why the problem is interesting, important, and/or difficult
  – Describe key contributions and technical ideas
  – Experimental setup and results
  – Strengths and weaknesses
  – Interesting open research questions, possible extensions, and applications
• Slides should be clear and mostly visual (figures, animations, videos)
• Look at background papers as necessary
Paper presentations

• Search for relevant material, e.g., from the authors' webpage, project pages, etc.

• **Each slide that is not your own must be clearly cited**

• Meet with me **one week before your presentation** with prepared slides

• Email slides on day of presentation

• Email subject "[ECS 289H] presentation slides"

• Skip paper review the day you are presenting
Final project

• One of:
  – Design and evaluation of a novel approach
  – An extension of an approach studied in class
  – In-depth analysis of an existing technique
  – In-depth comparison of two related existing techniques

• Work in pairs
• Talk to me if you need ideas
Final project

• Final Project Proposal (5%) due 10/31
  – 1 page

• Final Project Presentation (10%) due 12/9, 12/11
  – 15 minutes

• Final Project Paper (10%) due TBD
  – 6-8 pages
Local Features and Matching

- Local invariant features; detectors and descriptors
- Matching, indexing, bag-of-words
Image classification

- Object and scene classification
- Global and object-based representations
Object detection

- Discriminative methods
- Faces, people, objects
Unsupervised/weakly-supervised discovery

- Clustering, context
- Mid-level visual elements
Segmentation

- Grouping pixels
Human-in-the-loop

- Interactive systems, crowdsourcing
- Active learning
Activity recognition

• What are the humans in the image or video doing?
Human pose

- Finding people and their poses
Attributes

- Describing properties of objects, people, scenes
Image search and mining

• Large-scale instance-based recognition
• Discovering connections
Language and images

- Generating image descriptions

A two girls in the store.

A small herd of animals with a calf in the grass.
Big data

- Data-driven techniques
- Dataset bias
First-person vision

- Recognition from the first-person view
- Social scene understanding
Deep convolutional neural networks

**R-CNN: Regions with CNN features**

1. Input image
2. Extract region proposals (~2k)
3. Compute CNN features
4. Classify regions

- aeroplane? no.
- person? yes.
- tvmonitor? no.
Coming up

• Carefully read class webpage
• Sign-up for papers you want to present
• Next class: overview of my research 😊