Introductions

- **Instructor**: Prof. Kristen Grauman
- **TA**: Wei-Lin Hsiao

Today

- Course overview
- Requirements, logistics

What is computer vision?

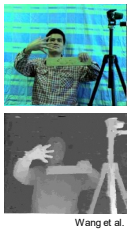
Done?

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

Snavely et al.

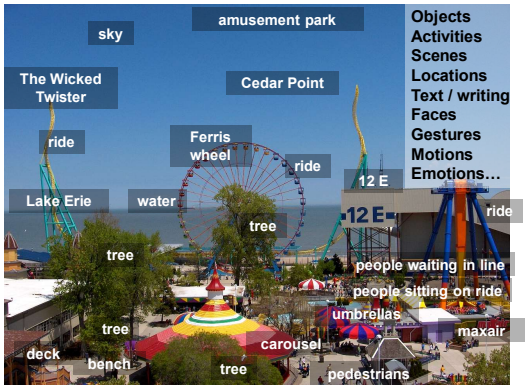
Tracking

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)*

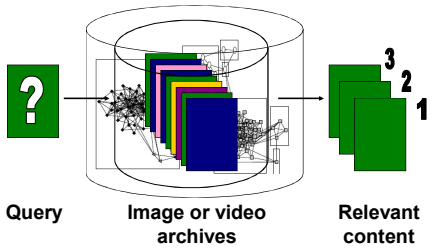
2. Vision for perception, interpretation



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

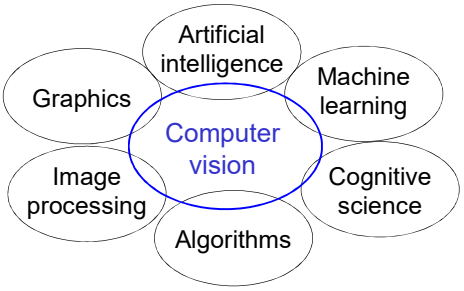


Computer Vision

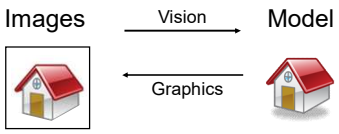
- Automatic understanding of images and video
  1. Computing properties of the 3D world from visual data *(measurement)*
  2. Algorithms and representations to allow a machine to recognize objects, people, scenes, and activities. *(perception and interpretation)*
  3. Algorithms to mine, search, and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

Course focus

Related disciplines

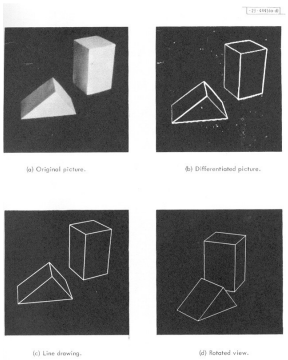


Vision and graphics



Inverse problems: analysis and synthesis.

Visual data in 1963



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 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?

Faces

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

Setting camera focus via face detection

Autonomous agents able to detect objects

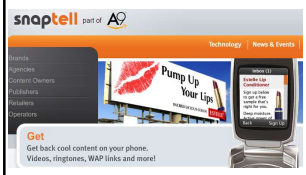
http://www.darpa.mil/grandchallenge/gallery.asp

Posing visual queries

Yeh et al., MIT

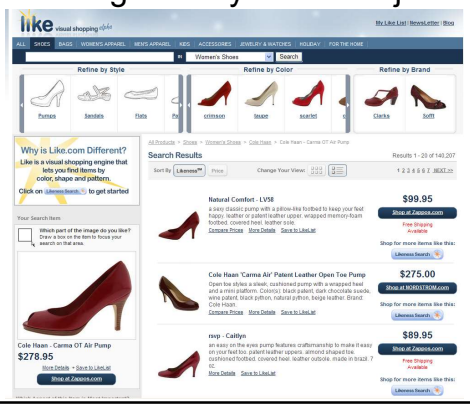
Digital Field Guides Eliminate the Guesswork

Belhumeur et al.



Kooaba, Bay & Quack et al.

Finding visually similar objects



Exploring community photo collections

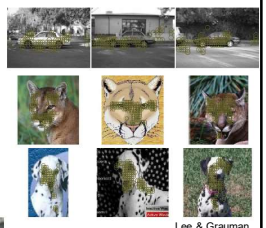
Snavely et al.

Simon & Seitz

Discovering visual patterns

Sivic & Zisserman

Objects



Lee & Grauman

Categories

Wang et al.

Actions

Auto-annotation

Gammeter et al.

T. Berg et al.

Video-based interfaces

Human joystick, NewsBreaker Live

Assistive technology systems
Camera Mouse, Boston College

Microsoft Kinect

What else?

Obstacles?

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

Artificial Intelligence Group
Vision Memo. No. 100.

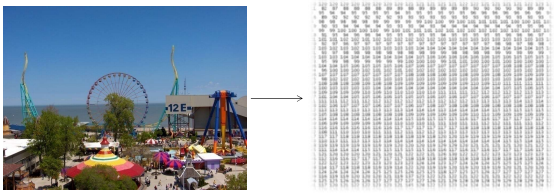
July 7, 1966

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".

What the computer gets



Why is vision difficult?

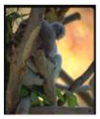
- Ill-posed problem: real world much more complex than what we can measure in images
  - 3D → 2D
- Impossible to literally "invert" image formation process

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

Video credit: Rob Fergus and Antonio Torralba

Challenges: importance of context

Video credit: Rob Fergus and Antonio Torralba

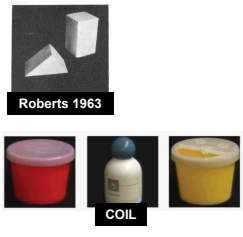
Challenges: importance of context

slide credit: Fei-Fei, Fergus & Torralba

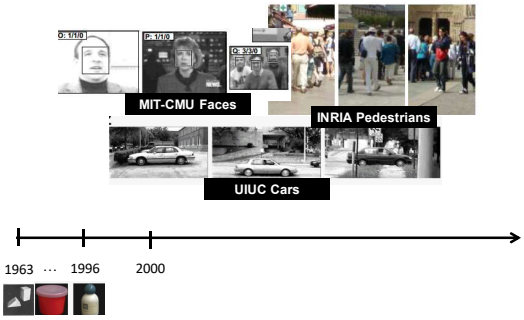
Challenges: complexity

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

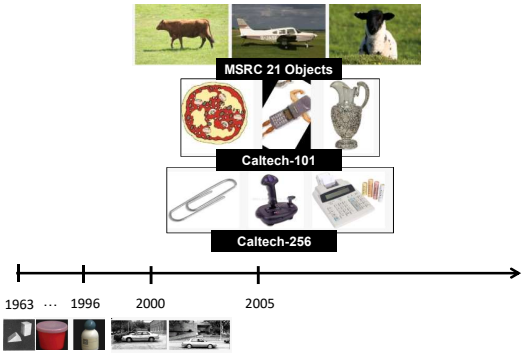
Progress charted by datasets



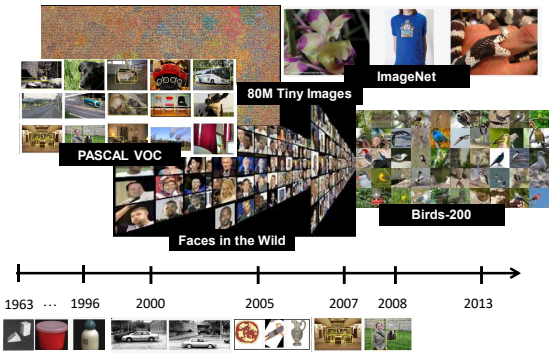
Progress charted by datasets



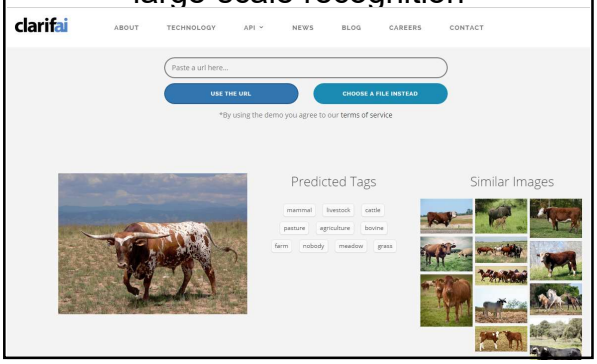
Progress charted by datasets



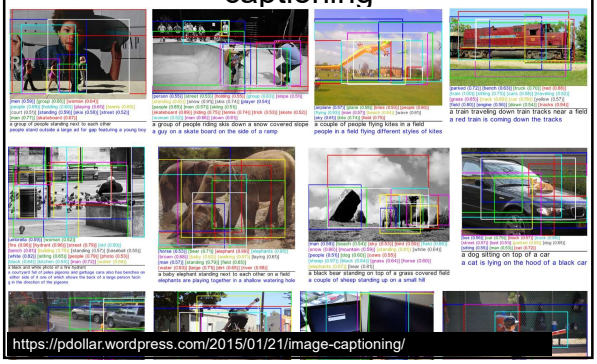
Progress charted by datasets



Expanding horizons: large-scale recognition

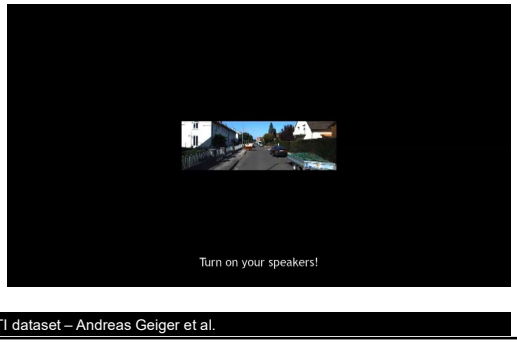


Expanding horizons: captioning

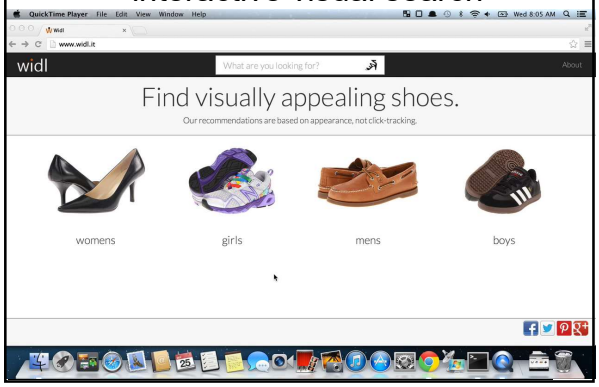


Expanding horizons: visual question answering

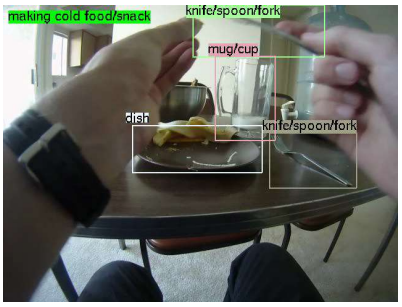
Expanding horizons: vision for autonomous vehicles



Expanding horizons: interactive visual search



Expanding horizons: first-person vision



Activities of Daily Living – Hamed Pirsiavash et al.

This course

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

Goals

- Understand current approaches
- Analyze
- Identify interesting research questions
- Some hands-on experience

Prerequisites

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

Basic format

- Early weeks (1-4):
  - Lectures by instructor
  - CNN tutorial
  - Paper reading
- Later weeks (5-11):
  - Paper discussion
  - Experiment
  - External paper presentation

Overview of requirements

- **Discussions** will center on recent papers in the field
  - Write 2 paper reviews each week, due Mon
  - Serve as proponent/opponent ~twice
- **Student presentations**
  - Present an "external" from syllabus
  - Experiment on an assigned paper
- **2 implementation assignments**
- **Project with a partner**

Workload is fairly high

Assigned vs. external papers

http://vision.cs.utexas.edu/381V-fall2017

Paper reviews

- Each week, review two of the assigned papers.
- Separately, summarize 2-3 "discussion points"
- Post each separately to Piazza following instructions on course "requirements" page.
- Skip reviews the week(s) you are presenting an external paper or experiment.

Paper review guidelines

- Brief (2-3 sentences) summary
- Main contribution
- Strengths? Weaknesses?
- How convincing are the experiments? Suggestions to improve them?
- Extensions? What's inspiring?
- Additional comments, unclear points
- Relationships observed between the papers we are reading
- due 8 pm Monday on Piazza

Discussion point guidelines

- ~2-3 sentences/bullets per reviewed paper
- Recap of salient parts of your reviews
  - Key observations, lingering questions, interesting connections, etc.
- Will be shared to our class via Piazza
- Discussion points required for each class session (due 8 pm Monday)
- All encouraged to browse and post before and after class

External paper(s) presentation guidelines

- Well-organized talk that introduces it to the class
- About 15 minutes
- What to cover?
  - Problem overview, motivation
  - Algorithm explanation, technical details
  - Results summary
  - Relation to assigned reading where relevant
  - Demos, videos, other visuals etc. from authors
- See class webpage for more details.

Experiment guidelines

- Implement/download code for a main idea in the paper and show us toy examples:
  - Show (on a small scale) an example to analyze a strength/weakness of the approach
  - Experiment with different types of thoughtfully chosen data
  - Compare some aspect of assigned papers
- Key to a good experiment:
  - Don't duplicate what we saw in the paper!
  - Not necessary to run whole thing end to end – focus, essentials
- Present in class – about 20 minutes.
  - Don't recap the paper beyond 1-2 slides
- Include links to any tools or data in slides

Timetable and prep

- For external paper or experiment presentation, by the Wednesday **the week before** your presentation is scheduled:
  - Email draft slides to me
  - I'll provide feedback within the next few days
  - Hard deadline: 5 points per day late
- Please **coordinate with other** presenters in advance for your day to avoid duplication of papers
- Please **bring slides** on own laptop and check it prior to class
- Please **email me final slides** pdf after class session <lastname>_paper.pdf / <lastname>_expt.pdf

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

- Work in pairs
- Project proposal due mid-term

Important dates

- Monday, Sept 4: paper topic preferences due to TA
- Monday, Sept 4: first set of 2 reviews due on Piazza
- Friday, Sept 22: first coding assignment due
- Wednesday, Oct 11: second coding assignment due
- Friday, Oct 13: second coding assignment follow-up run due
- Wednesday, Oct 25: project proposal due
- TBD in late Nov: poster printing deadline, 12 pm
- Wednesday, Dec 6: poster session in class, 1-4 pm
- Friday, Dec 8: final papers due

Grades

- Grades will be determined as follows:
  - 25% **participation** (includes attendance, in-class discussions, paper reviews)
  - 15% **coding** assignments
  - 35% **presentations** (includes drafts submitted one week prior, and in-class presentation)
  - 25% **final project** (includes proposal, poster, final paper)

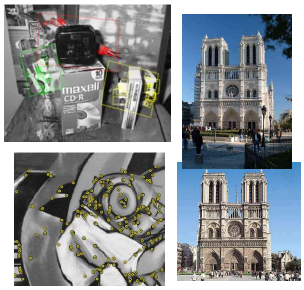
Miscellaneous

- Feedback welcome and useful!
- Slides on class website
- Discussion including assignment questions on Piazza
- No laptops, phones, etc. open in class please.
- Course is restricted to registered students

Syllabus Tour

- Learning objects and image representations
  - Instance recognition
  - Category recognition/detection
  - Self-supervised representation learning
  - ConvNet implementation tutorial
- Recognition on the move
  - Actions and objects in video
  - First-person vision
  - Embodied visual perception
- Potpourri
  - People
  - Visual data mining and discovery
  - Where to look
  - Language and vision

Instance recognition

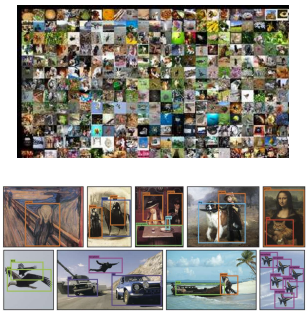


Local invariant features, detection and description

Matching models to images

Indexing specific objects with bag-of-words descriptors

Category recognition/detection



Recognition as an image classification problem

Discriminative methods

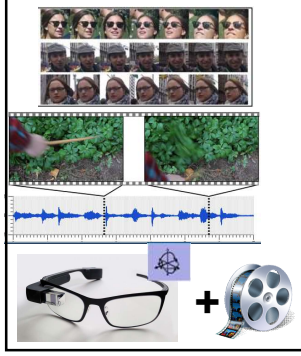
Image descriptors

Convolutional neural networks

Benchmark datasets

Object detection

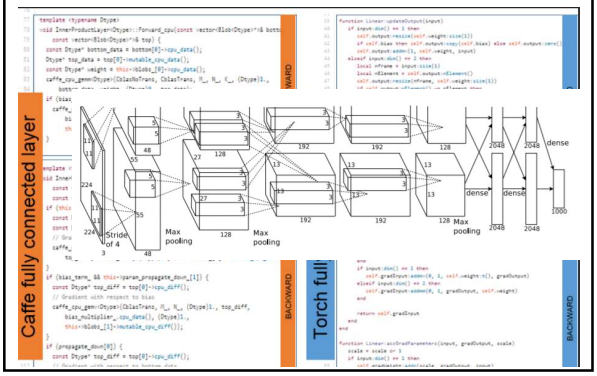
Self-supervised representation learning



Unsupervised feature learning from "free" side information

(tracks in video, spatial layout in images, audio, colorization, ego-motion…)

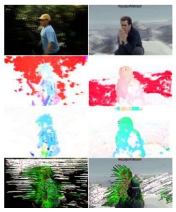
CNN tutorial



Syllabus Tour

- Learning objects and image representations
  - Instance recognition
  - Category recognition/detection
  - Self-supervised representation learning
  - ConvNet implementation tutorial
- Recognition on the move
  - Actions and objects in video
  - First-person vision
  - Embodied visual perception
- Potpourri
  - People
  - Visual data mining and discovery
  - Where to look
  - Language and vision

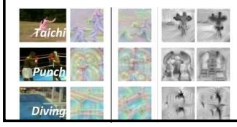
Actions and objects in video



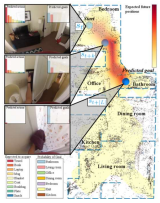
Detecting activities, actions, and events in images or video.

Video descriptors, interactions with objects and scenes.

Video object segmentation



First-person vision



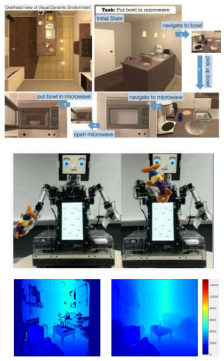
Egocentric wearable cameras

Actions w/manipulated objects

Forecasting future activities

Developmental learning lessons

Embodied visual perception



Learning how to move for recognition, manipulation.

3D objects and next best view

Visual learning grounded in action and physical interaction

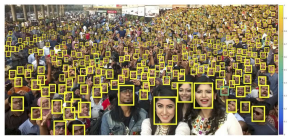
Visual recognition for robotics

Affordances

Syllabus Tour

- Learning objects and image representations
  - Instance recognition
  - Category recognition/detection
  - Self-supervised representation learning
  - ConvNet implementation tutorial
- Recognition on the move
  - Actions and objects in video
  - First-person vision
  - Embodied visual perception
- Potpourri
  - People
  - Visual data mining and discovery
  - Where to look
  - Language and vision

People



Human body pose

Faces

Fashion/clothing

Attributes

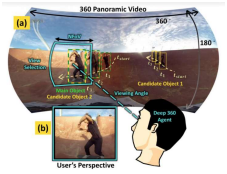
Visual data mining and discovery

Discovering visual patterns in large-scale community photo collections

StreetView, Flickr data

Demograhics, geography, ecology, brands, fashion.

Where to look



Gaze following

Gaze as a learning cue

Saliency

Summarization

Language and vision

Captioning

Referring expressions

Visual question answering

Word-image embeddings

Storytelling

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

- Please read over course requirements online

- Due Monday 8 PM
  - Reading and paper reviews/discussion point posts for instance recognition
  - 6 top topic preferences to Wei-Lin via email