Visual Recognition
Fall 2016

Introductions

- **Instructor**: Prof. Kristen Grauman
- **TA**: Kai-Yang Chiang

Today

- Course overview
- Requirements, logistics

What is computer vision?

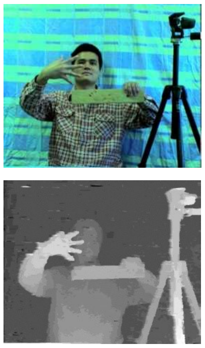
Done?

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

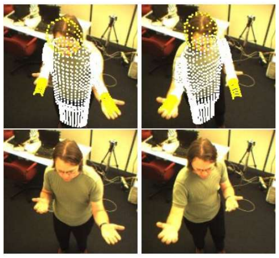


Wang et al.

Structure from motion

Snavely et al.

Tracking

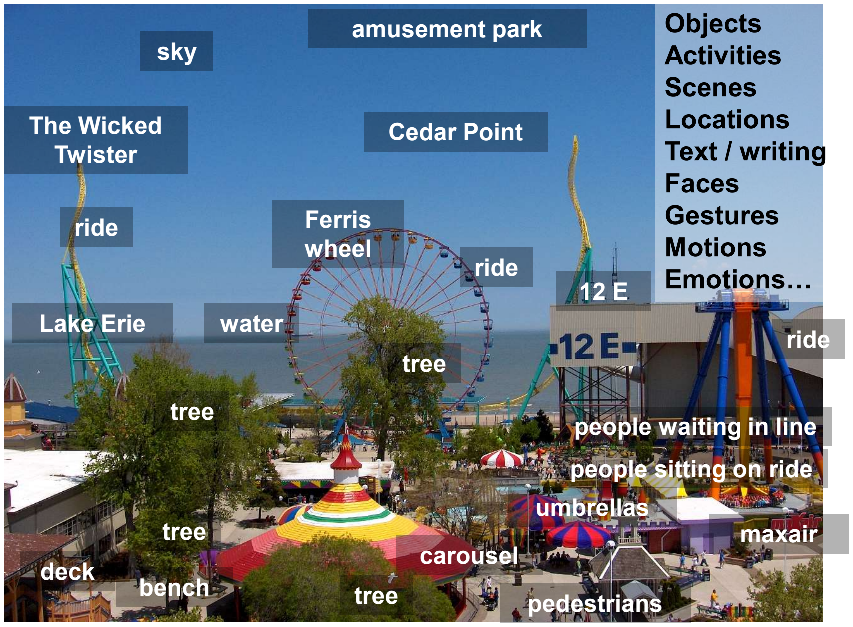


Demirdjian et al.

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

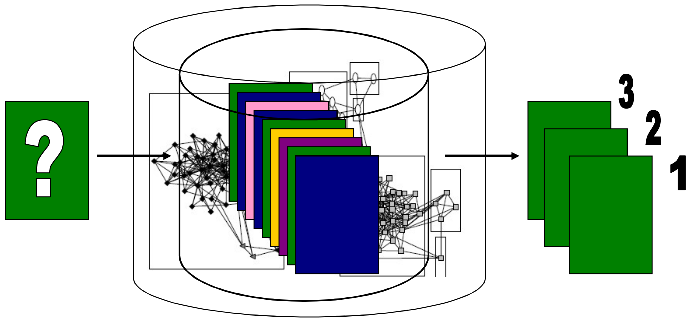


Objects
Activities
Scenes
Locations
Text / writing
Faces
Gestures
Motions
Emotions…

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



Query

Image or video archives

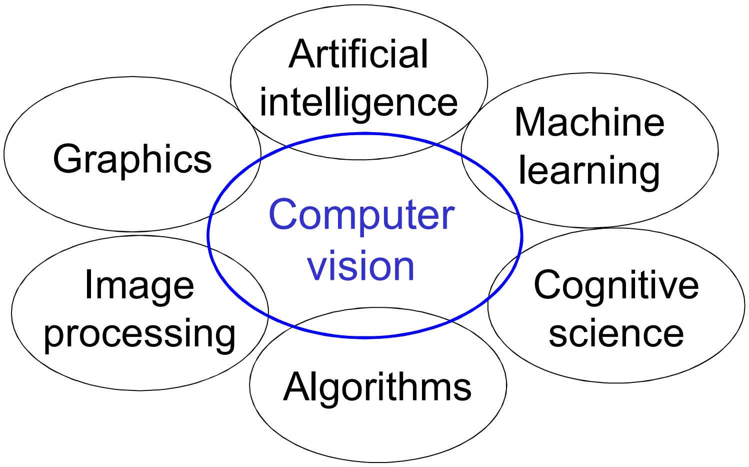
Relevant content

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



Artificial intelligence

Graphics

Machine learning

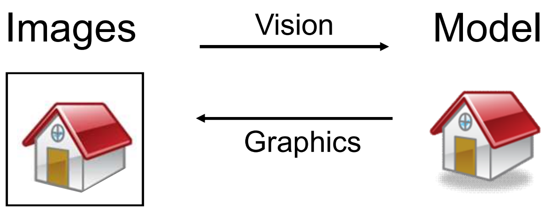
Computer vision

Image processing

Cognitive science

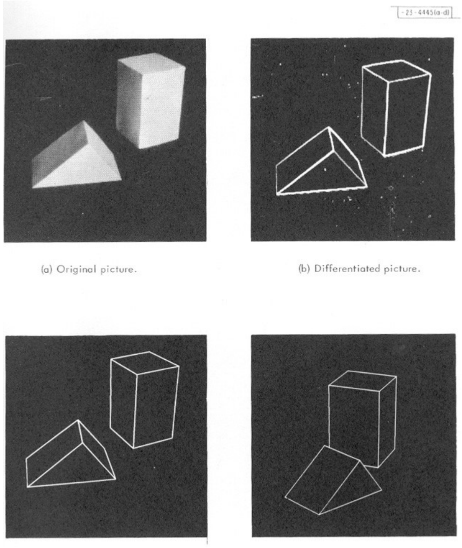
Algorithms

Vision and graphics



Inverse problems: analysis and synthesis.

Visual data in 1963



(a) Original picture.

(b) Differentiated picture.

(c) Line drawing.

(d) Rotated view.

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

Visual data in 2016

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

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

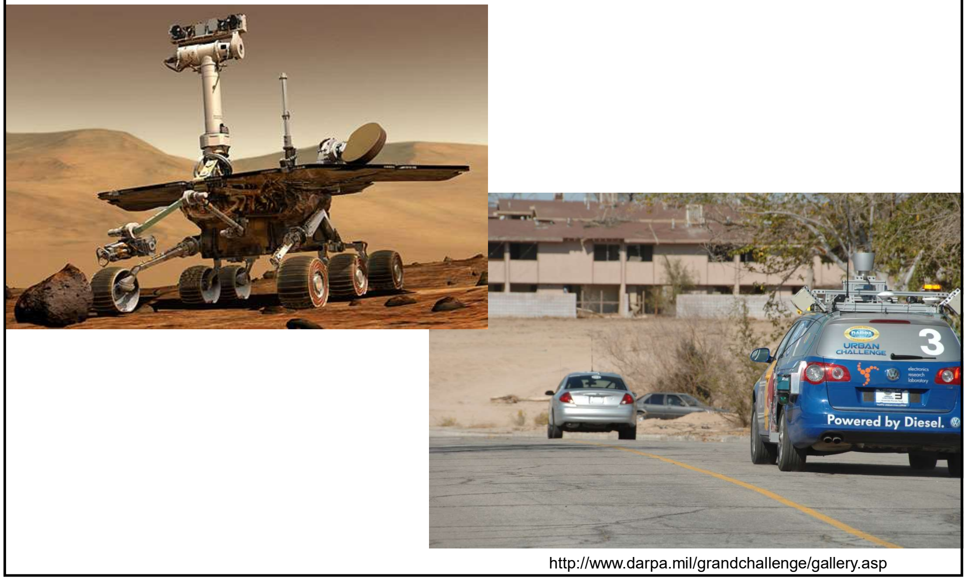
- Why now?

Faces

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

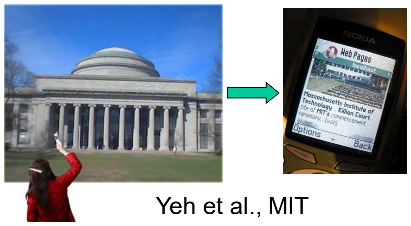
Setting camera focus via face detection

Autonomous agents able to detect objects



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

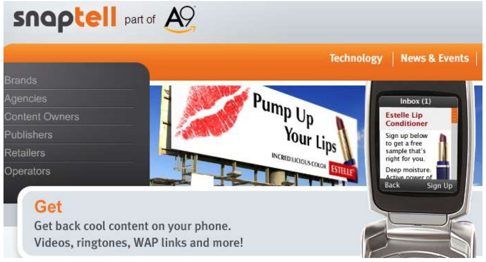
Posing visual queries



Yeh et al., MIT

Digital Field Guides Eliminate the Guesswork

Belhumeur et al.



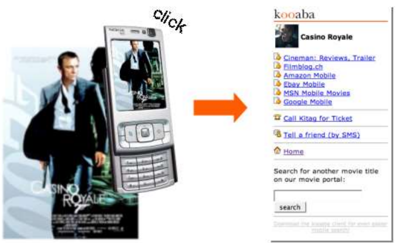
Estelle Lip Conditioner

Sign up below to get a free sample that's right for you.

Deep moisture. Active power of

Get

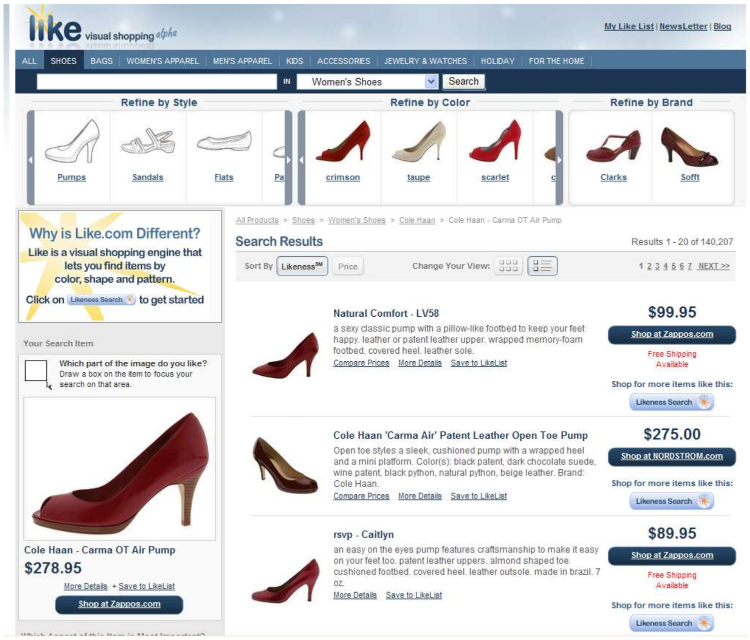
Get back cool content on your phone. Videos, ringtones, WAP links and more!



Search for another movie title on our movie portal:

Kooaba, Bay & Quack et al.

Finding visually similar objects



All Products > Shoes > Women's Shoes > Cole Haan > Cole Haan - Carma OT Air Pump

Why is Like.com Different?

Like is a visual shopping engine that lets you find items by color, shape and pattern.

Click on Likeness Search to get started

Your Search Item

Which part of the image do you like? Draw a box on the item to focus your search on that area.

Cole Haan - Carma OT Air Pump

$278.95

More Details + Save to LikeList

Shop at Zappos.com

Search Results

Results 1 - 20 of 140,207

Sort By: Likeness℠ | Price

Change Your View:

1 2 3 4 5 6 7 NEXT >>

Natural Comfort - LV58

a sexy classic pump with a pillow-like footbed to keep your feet happy. leather or patent leather upper. wrapped memory-foam footbed. covered heel. leather sole.

Compare Prices More Details Save to LikeList

$99.95

Shop at Zappos.com

Free Shipping Available

Shop for more items like this: Likeness Search

Cole Haan 'Carma Air' Patent Leather Open Toe Pump

Open toe styles a sleek, cushioned pump with a wrapped heel and a mini platform. Color(s): black patent, dark chocolate suede, wine patent, black python, natural python, beige leather. Brand: Cole Haan.

Compare Prices More Details Save to LikeList

$275.00

Shop at NORDSTROM.com

Shop for more items like this: Likeness Search

rsvp - Caitlyn

an easy on the eyes pump features craftsmanship to make it easy on your feet too. patent leather uppers. almond shaped toe. cushioned footbed. covered heel. leather outsole. made in brazil. 7 oz.

More Details Save to LikeList

$89.95

Shop at Zappos.com

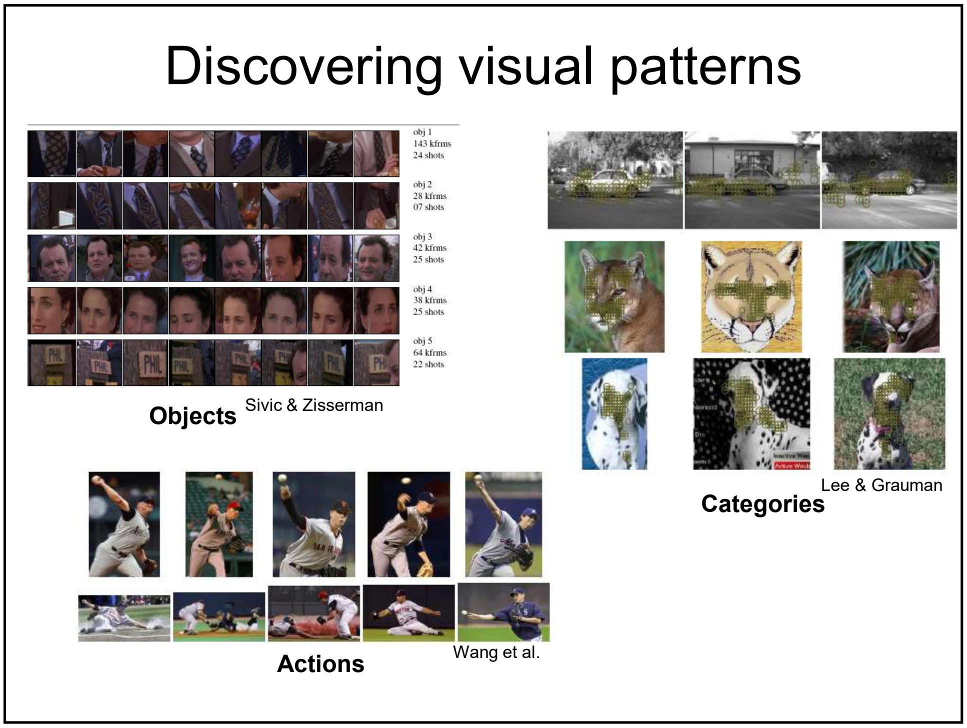
Free Shipping Available

Shop for more items like this: Likeness Search

Exploring community photo collections

Snavely et al.

Simon & Seitz



Discovering visual patterns

Objects Sivic & Zisserman

Categories Lee & Grauman

Actions Wang et al.

Auto-annotation

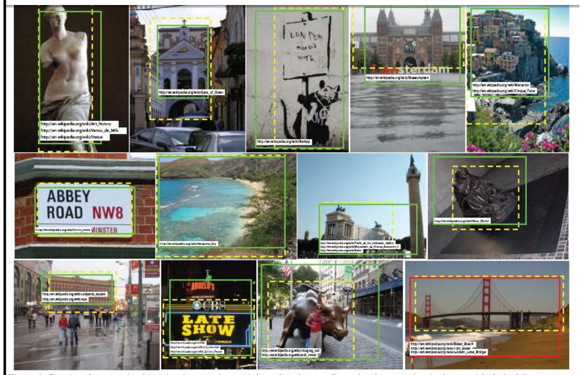
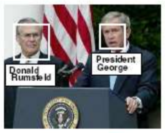
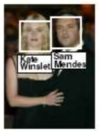


Figure 9. Results of automatic object-level annotation with bounding boxes. Groundtruth annotation is shown with dashed lines, correct detection with solid green lines, false detections with solid red lines. Auto-annotation with related Wikipedia articles is also shown. All results are also labeled with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

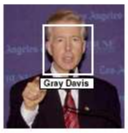
Gammeter et al.



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.

Video-based interfaces

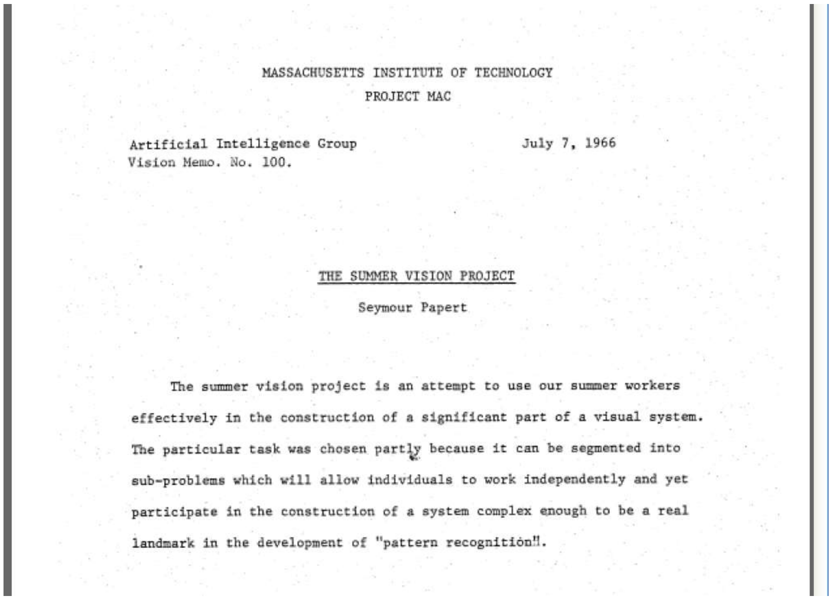
Human joystick, NewsBreaker Live

Assistive technology systems
Camera Mouse, Boston College

Microsoft Kinect

What else?

Obstacles?



MASSACHUSETTS INSTITUTE OF TECHNOLOGY
PROJECT MAC

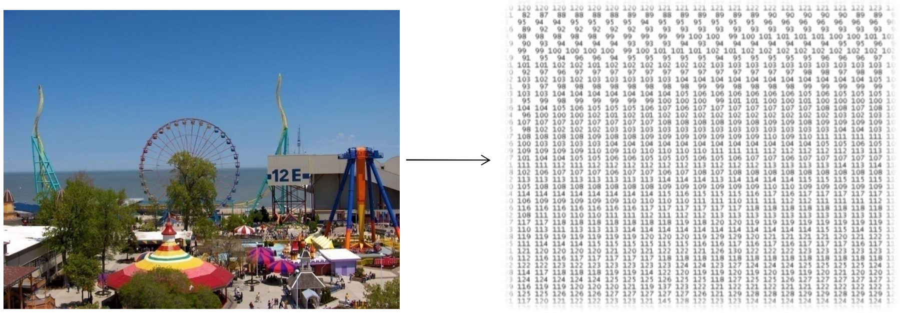
Artificial Intelligence Group July 7, 1966
Vision Memo. No. 100.

THE SUMMER VISION PROJECT

Seymour Papert

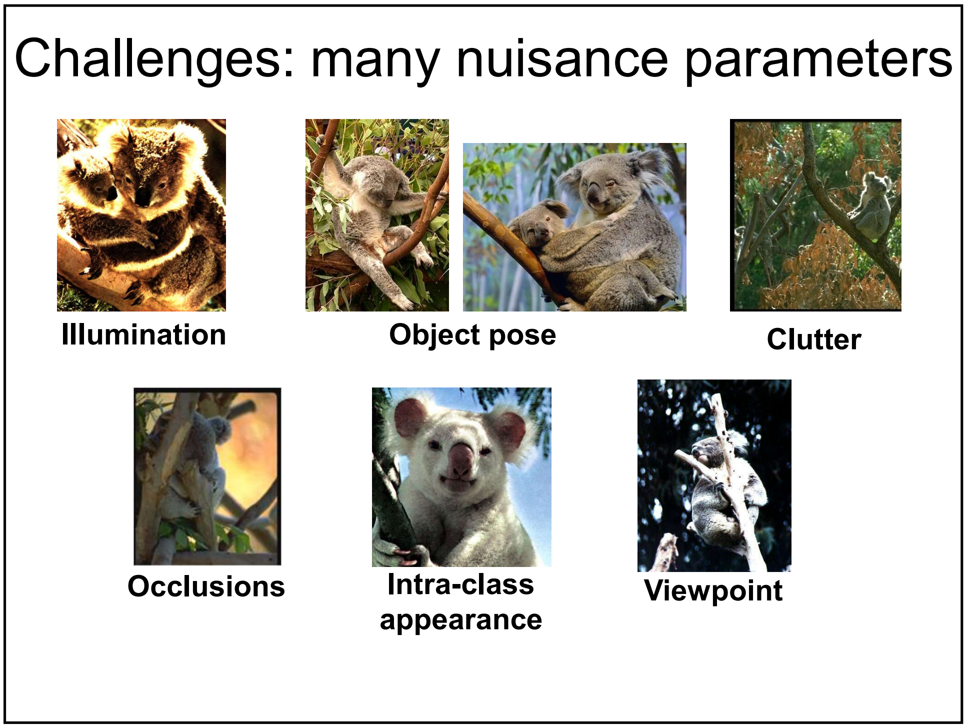
The summer vision project is an attempt to use our summer workers effectively in the construction of a significant part of a visual system. The particular task was chosen partly because it can be segmented into sub-problems which will allow individuals to work independently and yet participate in the 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

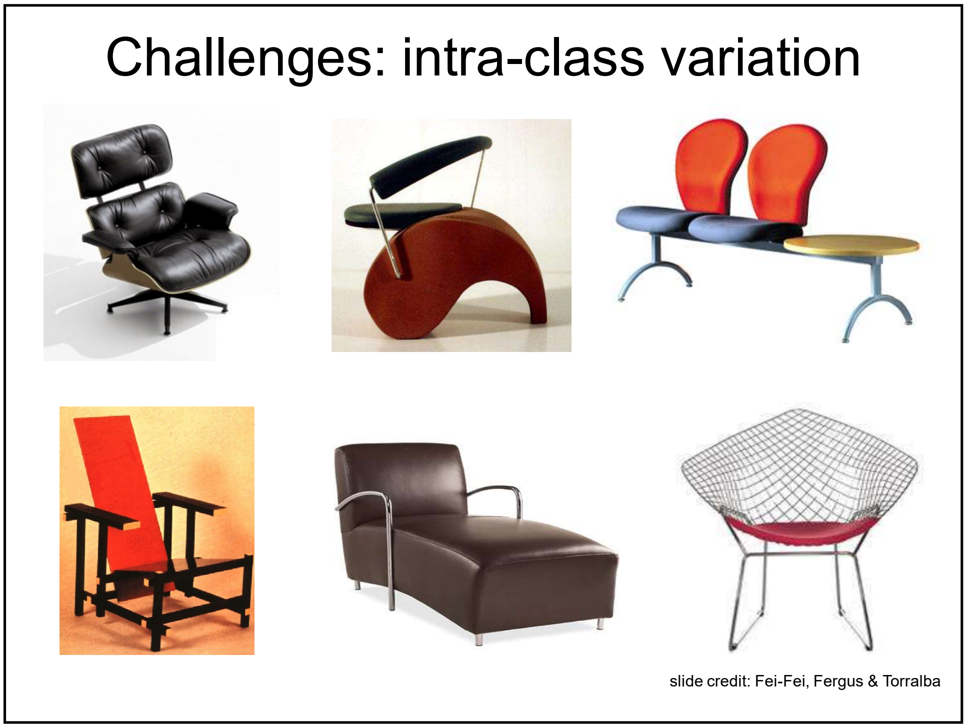
Object pose

Clutter

Occlusions

Intra-class appearance

Viewpoint



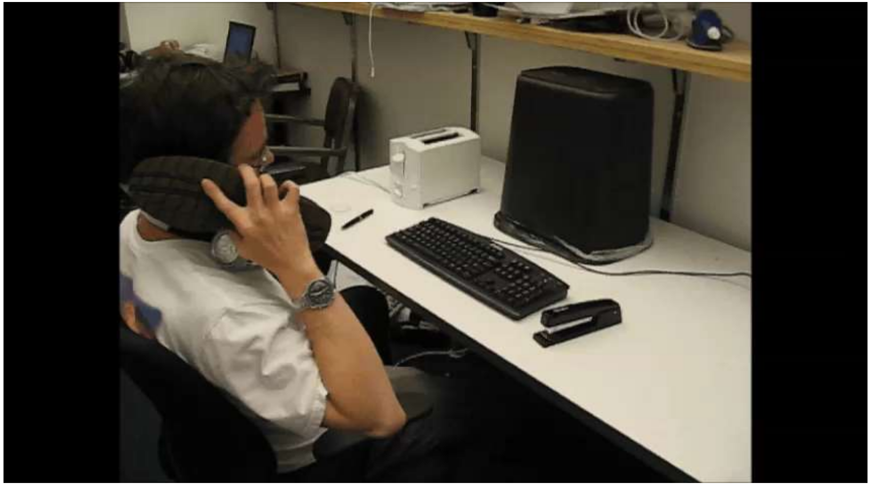
Challenges: intra-class variation

slide credit: Fei-Fei, Fergus & Torralba

Challenges: importance of context

Video credit: Rob Fergus and
Antonio Torralba

Challenges: importance of context



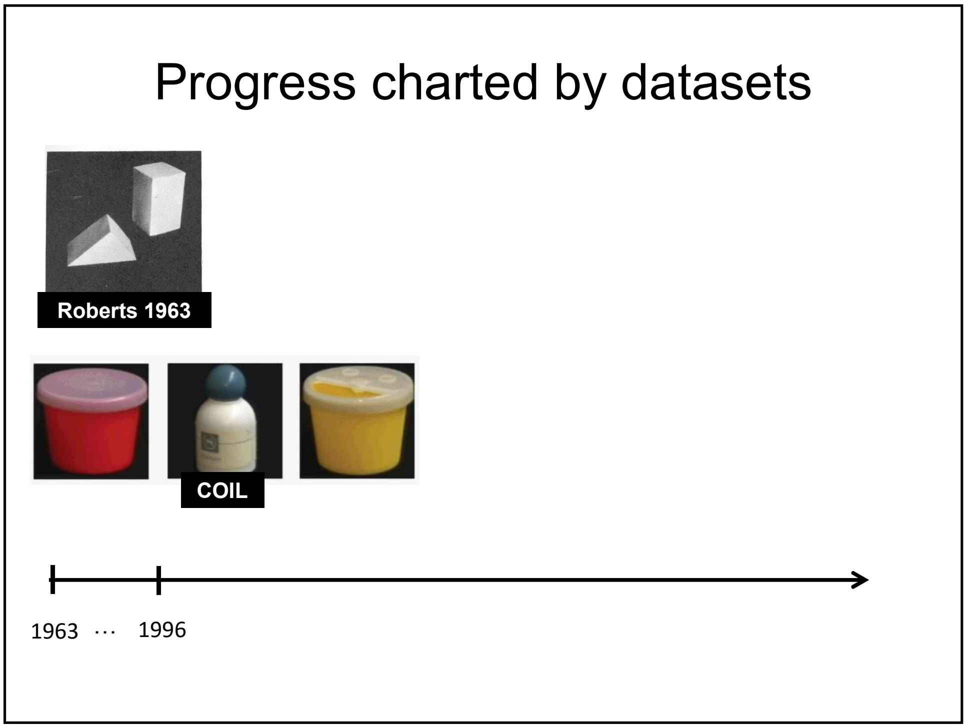
Video credit: Rob Fergus and
Antonio Torralba

Challenges: importance of context

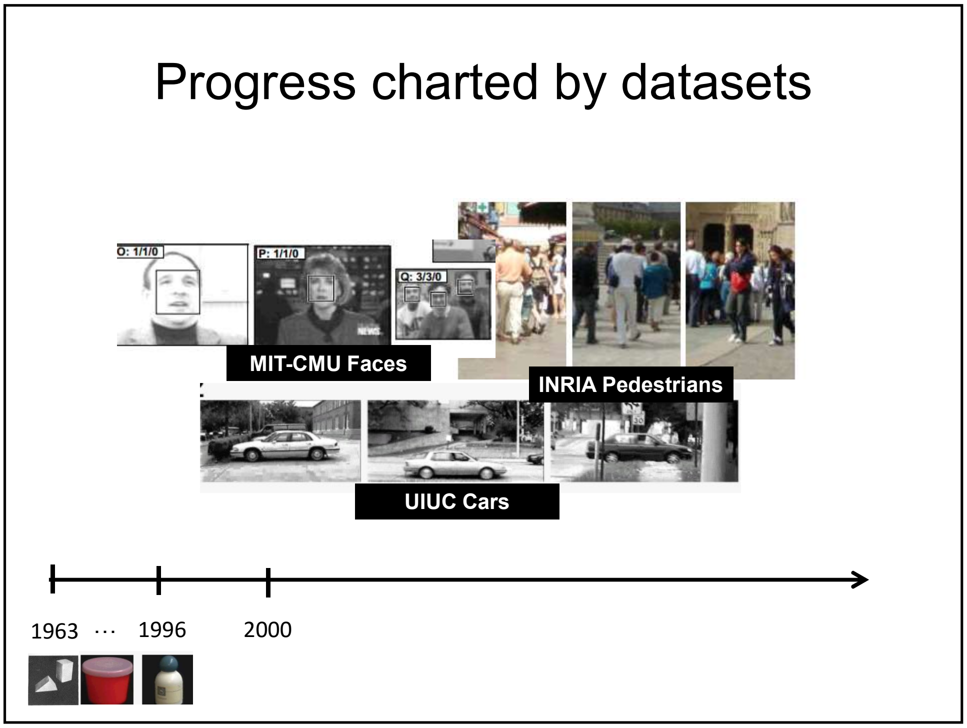
slide credit: Fei-Fei, Fergus & Torralba

Challenges: complexity

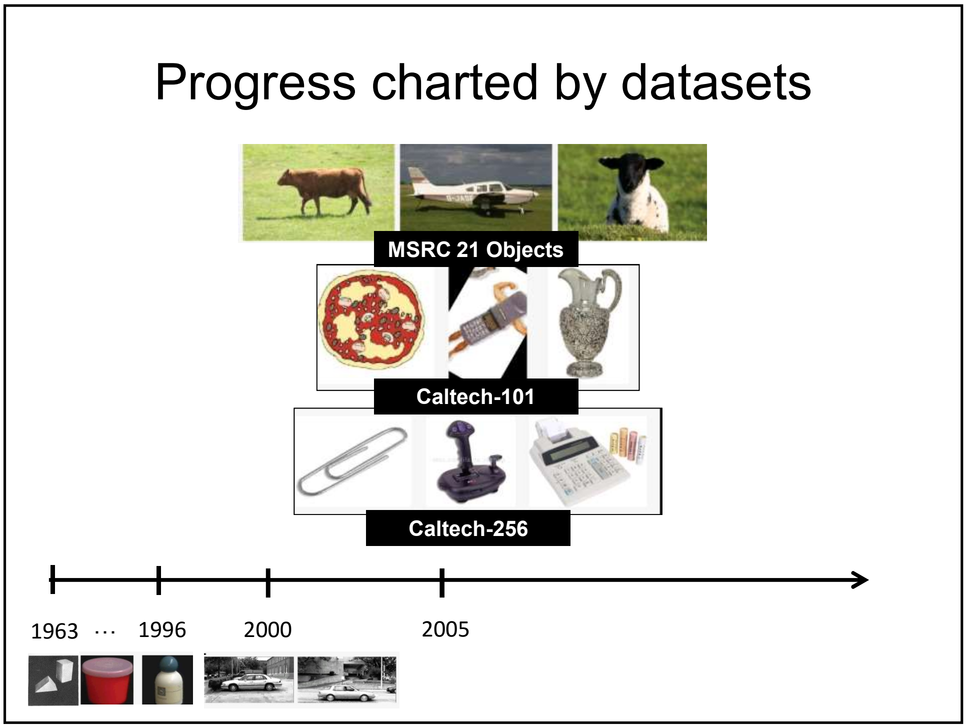
- Millions of pixels in an image
- 30,000 human recognizable object categories
- 30+ degrees of freedom in the pose of articulated objects (humans)
- 300 hours of new video on YouTube per minute
- …
- About half of the cerebral cortex in primates is devoted to processing visual information [Felleman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



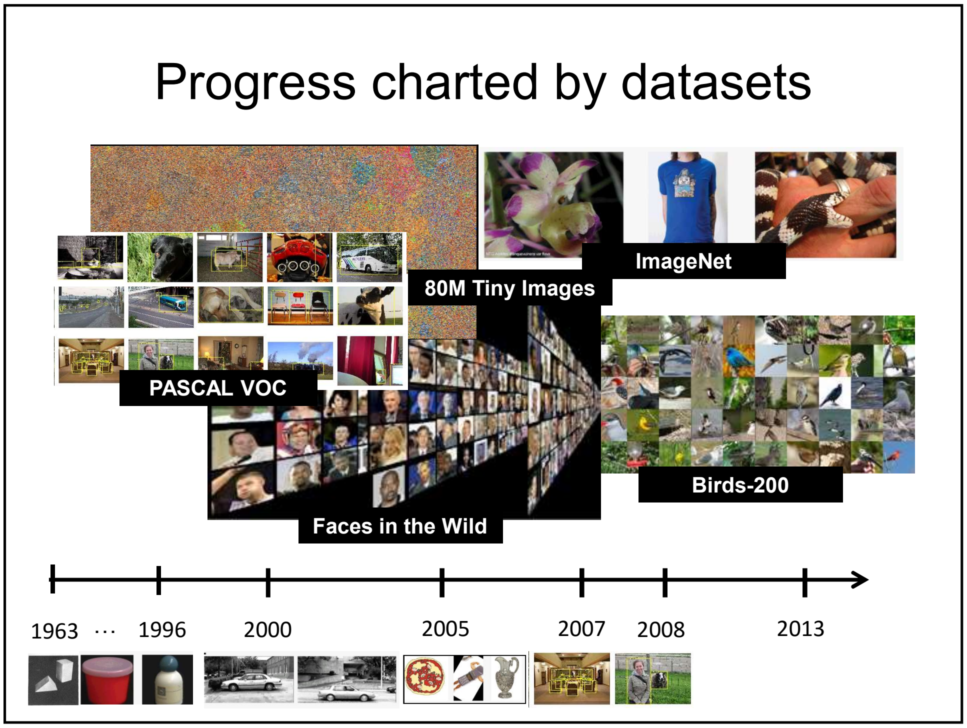
Progress charted by datasets



Progress charted by datasets

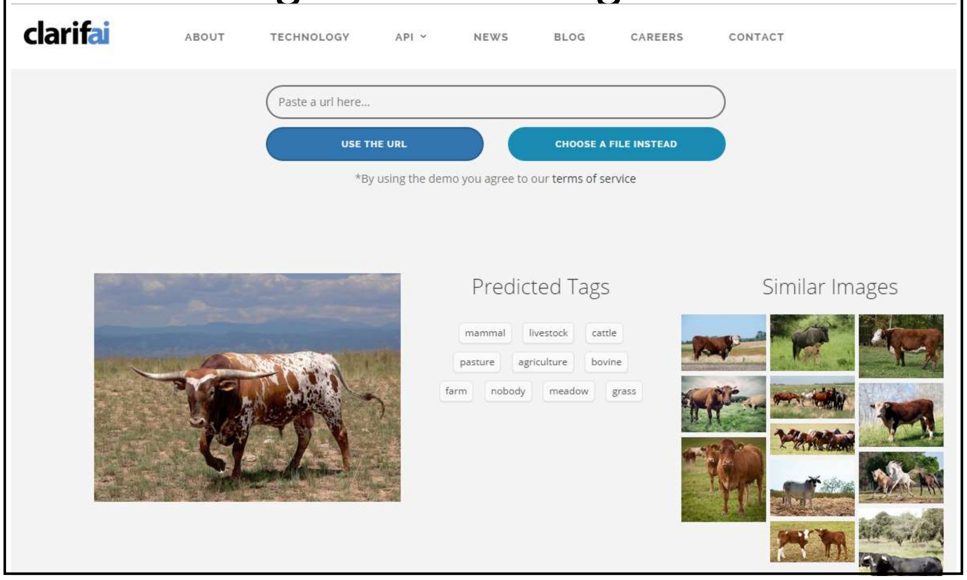


Progress charted by datasets



Progress charted by datasets

Expanding horizons: large-scale recognition



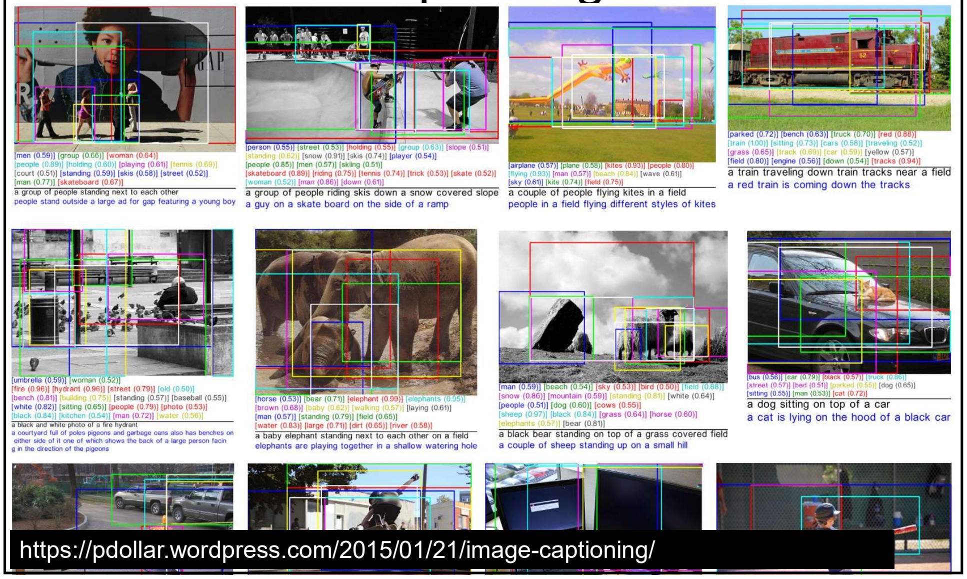
*By using the demo you agree to our terms of service

Predicted Tags

mammal livestock cattle pasture agriculture bovine farm nobody meadow grass

Similar Images

Expanding horizons: captioning



[men (0.59)] [group (0.66)] [woman (0.64)] [people (0.89)] [holding (0.60)] [playing (0.61)] [tennis (0.69)] [court (0.51)] [standing (0.59)] [skis (0.58)] [street (0.52)] [man (0.77)] [skateboard (0.67)]

a group of people standing next to each other

people stand outside a large ad for gap featuring a young boy

[person (0.55)] [street (0.53)] [holding (0.55)] [group (0.63)] [slope (0.51)] [standing (0.62)] [snow (0.91)] [skis (0.74)] [player (0.54)] [people (0.85)] [men (0.57)] [skiing (0.51)] [skateboard (0.89)] [riding (0.75)] [tennis (0.74)] [trick (0.53)] [skate (0.52)] [woman (0.52)] [man (0.86)] [down (0.61)]

a group of people riding skis down a snow covered slope

a guy on a skate board on the side of a ramp

[airplane (0.57)] [plane (0.58)] [kites (0.93)] [people (0.80)] [flying (0.93)] [man (0.57)] [beach (0.84)] [wave (0.61)] [sky (0.61)] [kite (0.74)] [field (0.75)]

a couple of people flying kites in a field

people in a field flying different styles of kites

[parked (0.72)] [bench (0.63)] [truck (0.70)] [red (0.88)] [train (1.00)] [sitting (0.73)] [cars (0.58)] [traveling (0.52)] [grass (0.65)] [track (0.69)] [car (0.59)] [yellow (0.57)] [field (0.80)] [engine (0.56)] [down (0.54)] [tracks (0.94)]

a train traveling down train tracks near a field

a red train is coming down the tracks

[umbrella (0.59)] [woman (0.52)] [fire (0.96)] [hydrant (0.96)] [street (0.79)] [old (0.50)] [bench (0.81)] [building (0.75)] [standing (0.57)] [baseball (0.55)] [white (0.82)] [sitting (0.65)] [people (0.79)] [photo (0.53)] [black (0.84)] [kitchen (0.54)] [man (0.72)] [water (0.56)]

a black and white photo of a fire hydrant

a courtyard full of poles pigeons and garbage cans also has benches on either side of it one of which shows the back of a large person facing in the direction of the pigeons

[horse (0.53)] [bear (0.71)] [elephant (0.99)] [elephants (0.95)] [brown (0.68)] [baby (0.62)] [walking (0.57)] [laying (0.61)] [man (0.57)] [standing (0.79)] [field (0.65)] [water (0.83)] [large (0.71)] [dirt (0.65)] [river (0.58)]

a baby elephant standing next to each other on a field

elephants are playing together in a shallow watering hole

[man (0.59)] [beach (0.54)] [sky (0.53)] [bird (0.50)] [field (0.88)] [snow (0.86)] [mountain (0.59)] [standing (0.81)] [white (0.64)] [people (0.51)] [dog (0.60)] [cows (0.55)] [sheep (0.97)] [black (0.84)] [grass (0.64)] [horse (0.60)] [elephants (0.57)] [bear (0.81)]

a black bear standing on top of a grass covered field

a couple of sheep standing up on a small hill

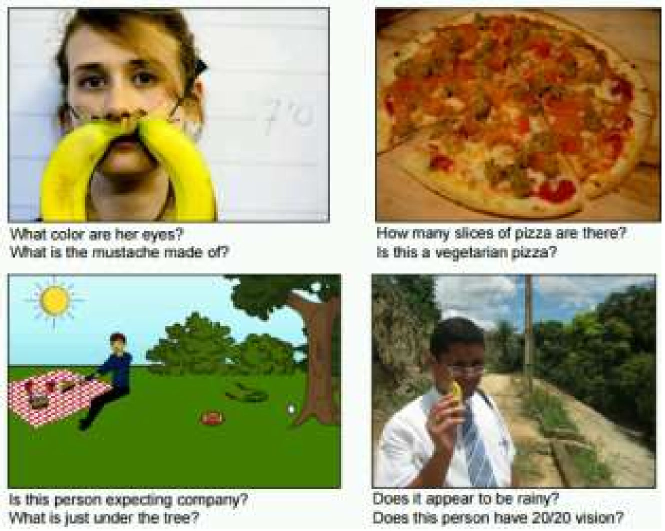
[bus (0.56)] [car (0.79)] [black (0.57)] [truck (0.86)] [street (0.57)] [bed (0.51)] [parked (0.55)] [dog (0.65)] [sitting (0.55)] [man (0.53)] [cat (0.72)]

a dog sitting on top of a car

a cat is lying on the hood of a black car

https://pdollar.wordpress.com/2015/01/21/image-captioning/

Expanding horizons: question answering



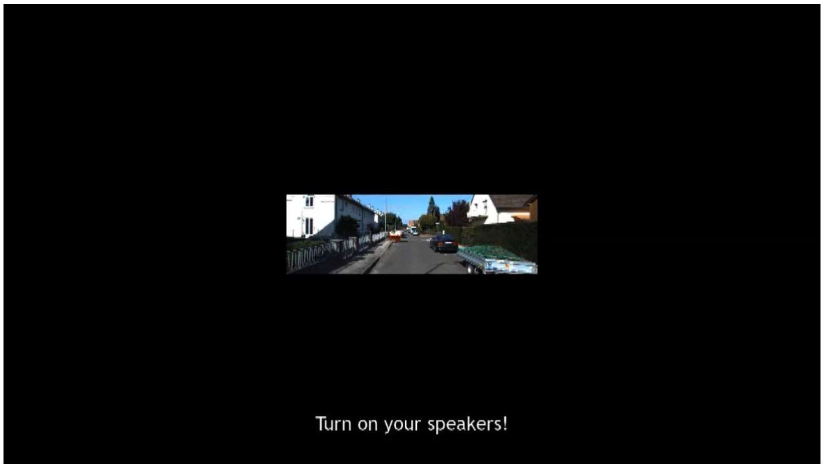
What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

Is this person expecting company?
What is just under the tree?

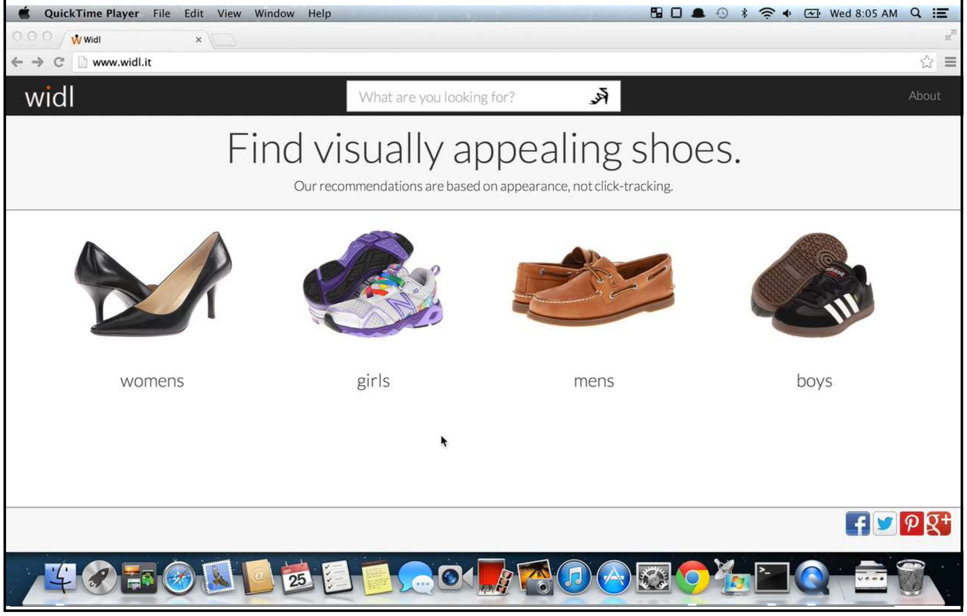
Does it appear to be rainy?
Does this person have 20/20 vision?

Expanding horizons: vision for autonomous vehicles



KITTI dataset – Andreas Geiger et al.

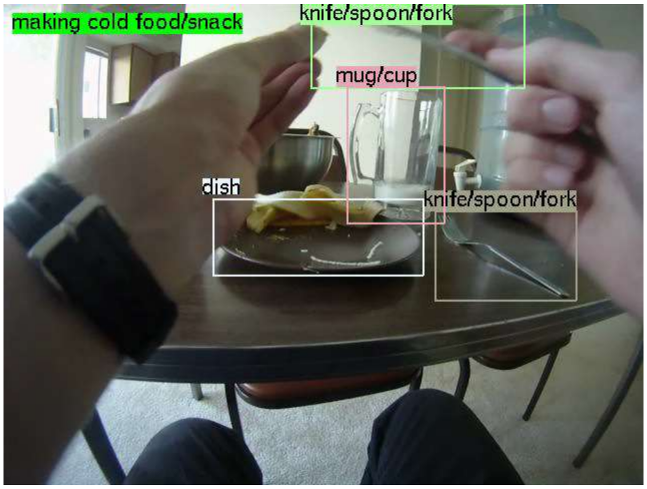
Expanding horizons: interactive visual search


QuickTime Player File Edit View Window Help
Wed 8:05 AM
Widl
www.widl.it
widl
What are you looking for?
About

Find visually appealing shoes.

Our recommendations are based on appearance, not click-tracking.

Expanding horizons: first-person vision



Activities of Daily Living – Hamed Pirsiavash et al.

Brainstorm

Pick an application or task among any of those we’ve described so far.

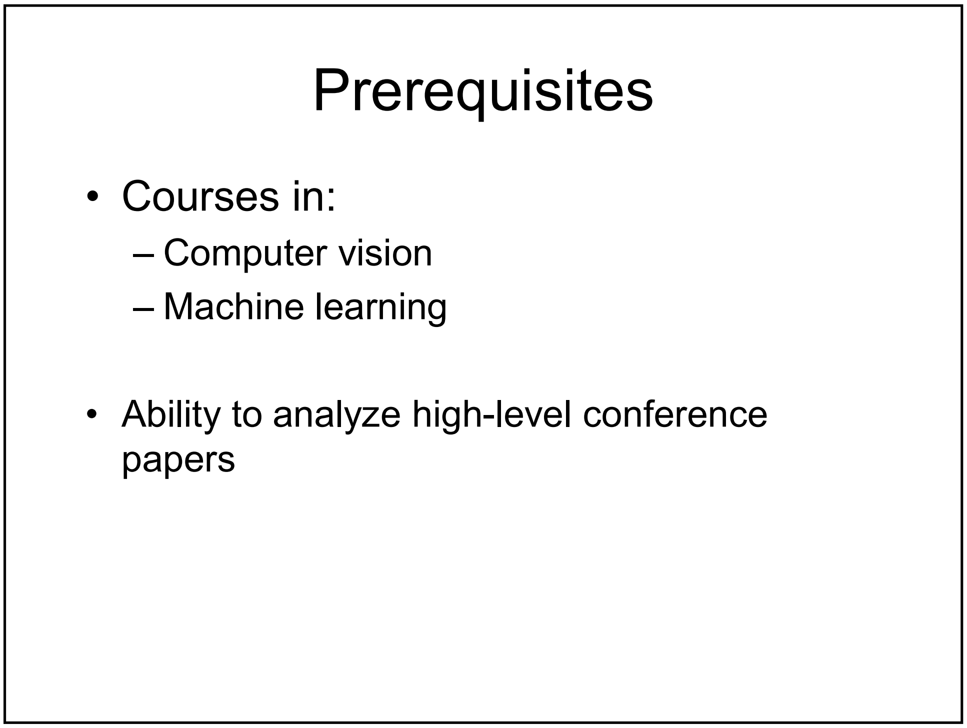
1. What functionality should the system have?
2. Intuitively, what are the technical sub-problems that must be solved?

This course

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

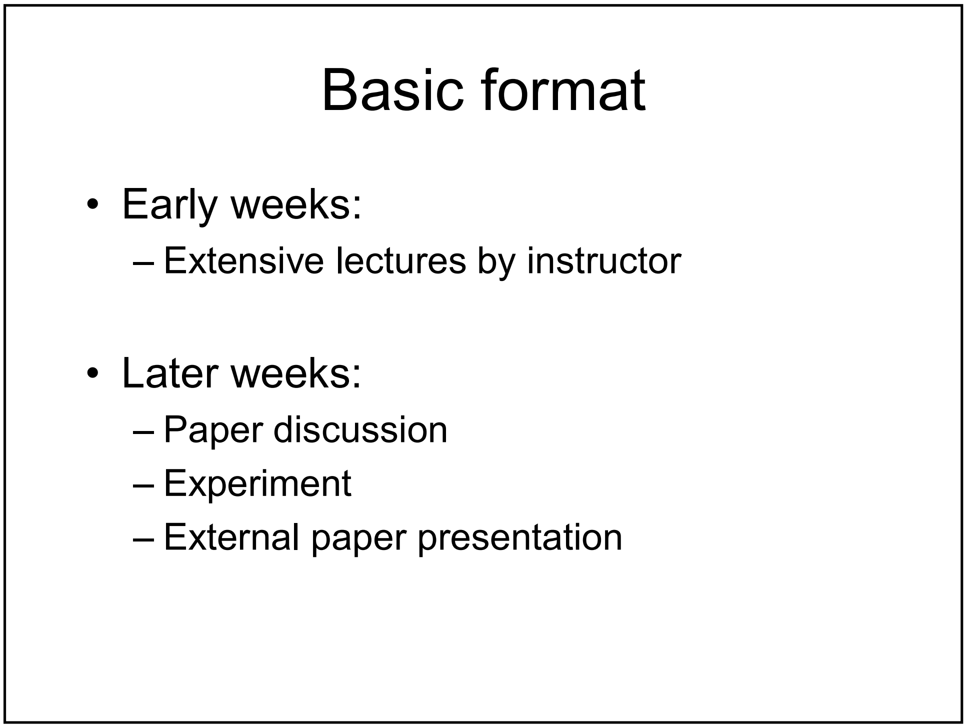
Goals

- Understand current approaches
- Analyze
- Identify interesting research questions



Prerequisites

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



Basic format

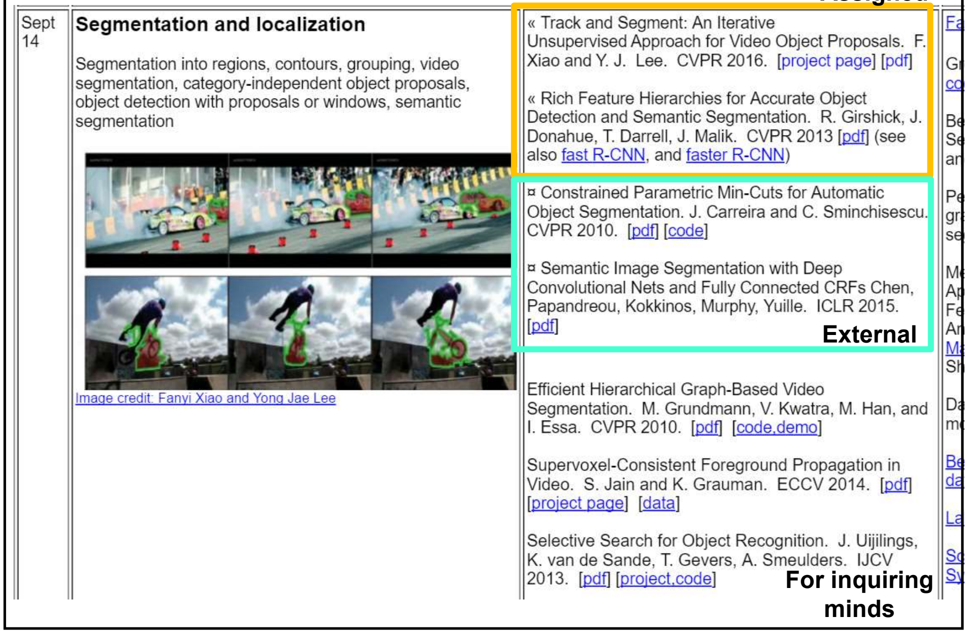
- Early weeks:
  - Extensive lectures by instructor
- Later weeks:
  - Paper discussion
  - Experiment
  - External paper presentation

Expectations

- **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 and external papers



Sept 14	**Segmentation and localization** Segmentation into regions, contours, grouping, video segmentation, category-independent object proposals, object detection with proposals or windows, semantic segmentation Image credit: Fanyi Xiao and Yong Jae Lee	« Track and Segment: An Iterative Unsupervised Approach for Video Object Proposals. F. Xiao and Y. J. Lee. CVPR 2016. [project page] [pdf] « Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. R. Girshick, J. Donahue, T. Darrell, J. Malik. CVPR 2013 [pdf] (see also fast R-CNN, and faster R-CNN) ¤ Constrained Parametric Min-Cuts for Automatic Object Segmentation. J. Carreira and C. Sminchisescu. CVPR 2010. [pdf] [code] ¤ Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs Chen, Papandreou, Kokkinos, Murphy, Yuille. ICLR 2015. [pdf] Efficient Hierarchical Graph-Based Video Segmentation. M. Grundmann, V. Kwatra, M. Han, and I. Essa. CVPR 2010. [pdf] [code,demo] Supervoxel-Consistent Foreground Propagation in Video. S. Jain and K. Grauman. ECCV 2014. [pdf] [project page] [data] Selective Search for Object Recognition. J. Uijilings, K. van de Sande, T. Gevers, A. Smeulders. IJCV 2013. [pdf] [project,code]

Assigned

External

For inquiring minds

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

Discussion point guidelines

- ~2-3 sentences 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 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
- 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 couple 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:

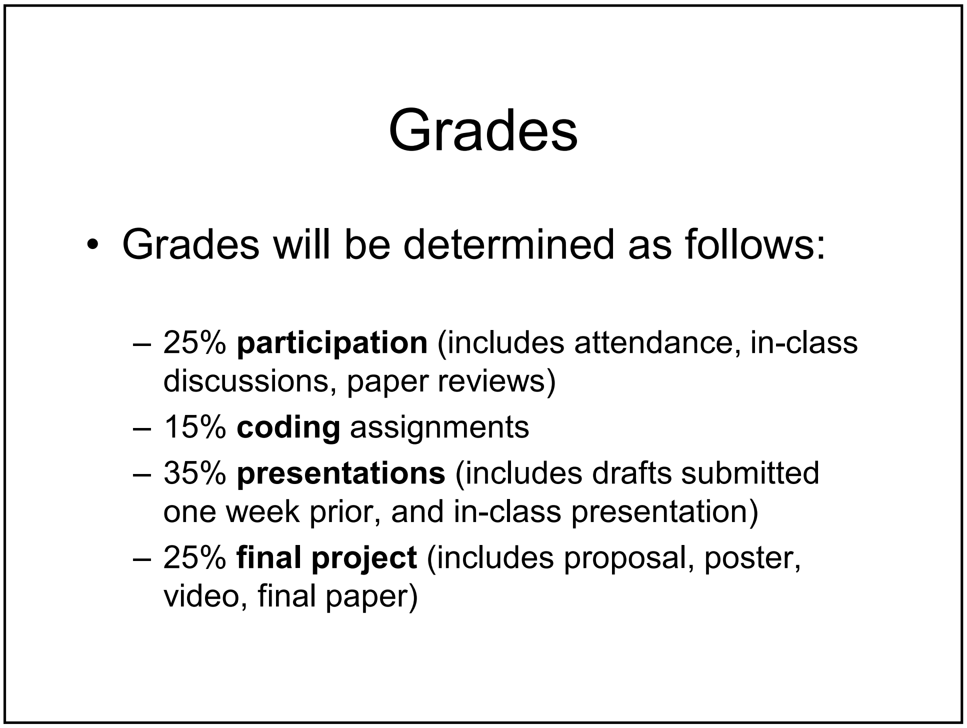
- Extend a technique studied in class
- Analysis and empirical evaluation of an existing technique
- Comparison between two approaches
- Design and evaluate a novel approach

- Work in pairs

- Project proposal due mid-term

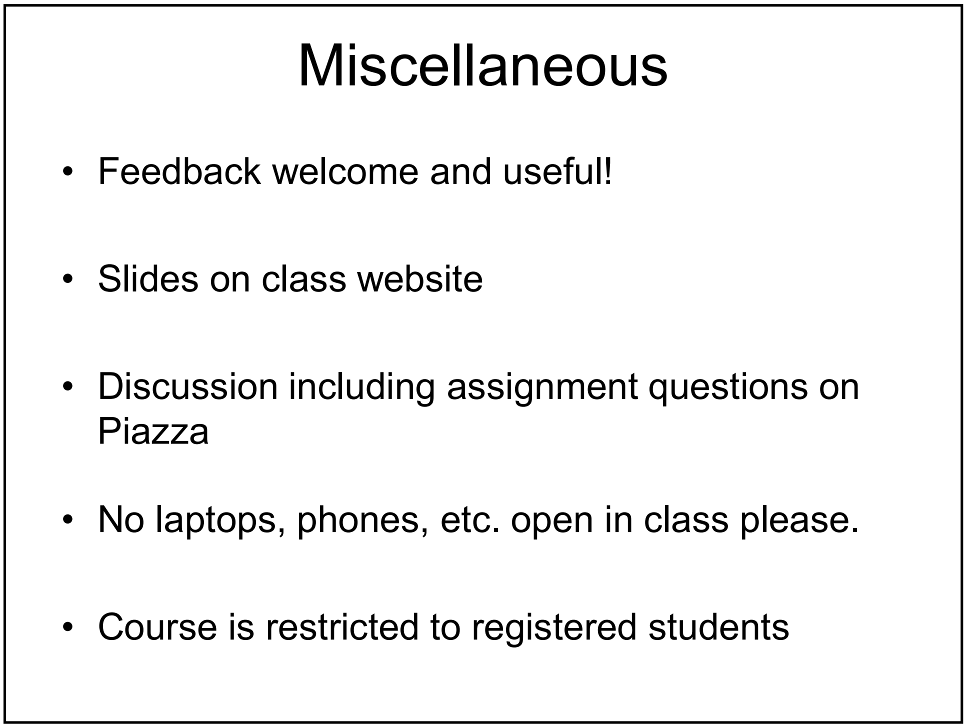
Important dates

- Monday, Aug 28: paper topic preferences due
- Monday, Aug 28: first set of 2 reviews due on Piazza
- Monday, Sept 12: hands-on CNN tutorial, 5-7 pm
- Friday, Sept 16: first coding assignment due
- Friday, Sept 30: second coding assignment due
- Monday, Oct 3: second coding assignment follow-up run due
- Wednesday, Oct 19: project proposal due
- Tuesday, Nov 22: poster printing deadline, 12 pm
- Wednesday, Nov 30: poster session in class, 1-4 pm
- Friday, Dec 2: final papers and poster reviews 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, video, 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

A. Foundations
   1. Instance recognition
   2. Category recognition
   3. Segmentation and localization

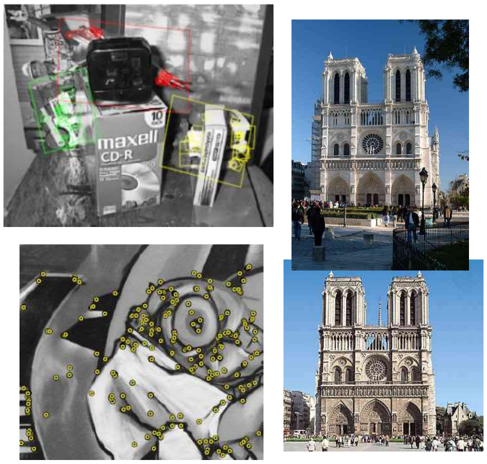
B. Advanced representations
   1. Self-supervised representation learning
   2. Attributes

C. Activity and acting
   1. Actions and events
   2. First-person vision
   3. Active perception

D. People
   1. People looking at scenes
   2. People in scenes

E. More modalities
   1. Sketch
   2. 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

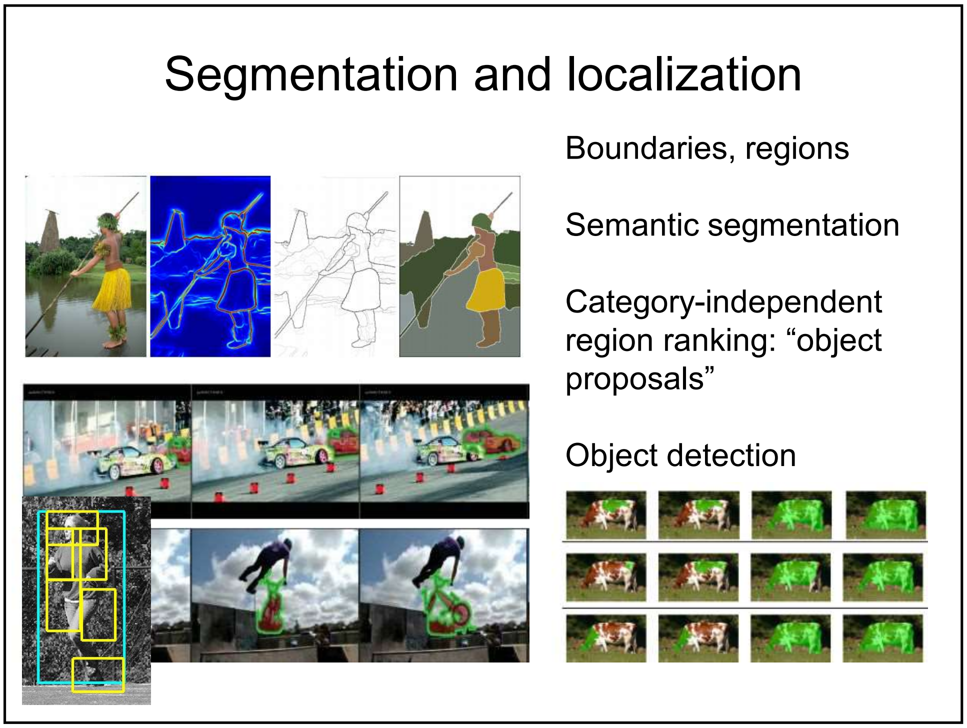
Recognition as an image classification problem

Discriminative methods

Image descriptors

Convolutional neural networks

Large-scale image collections



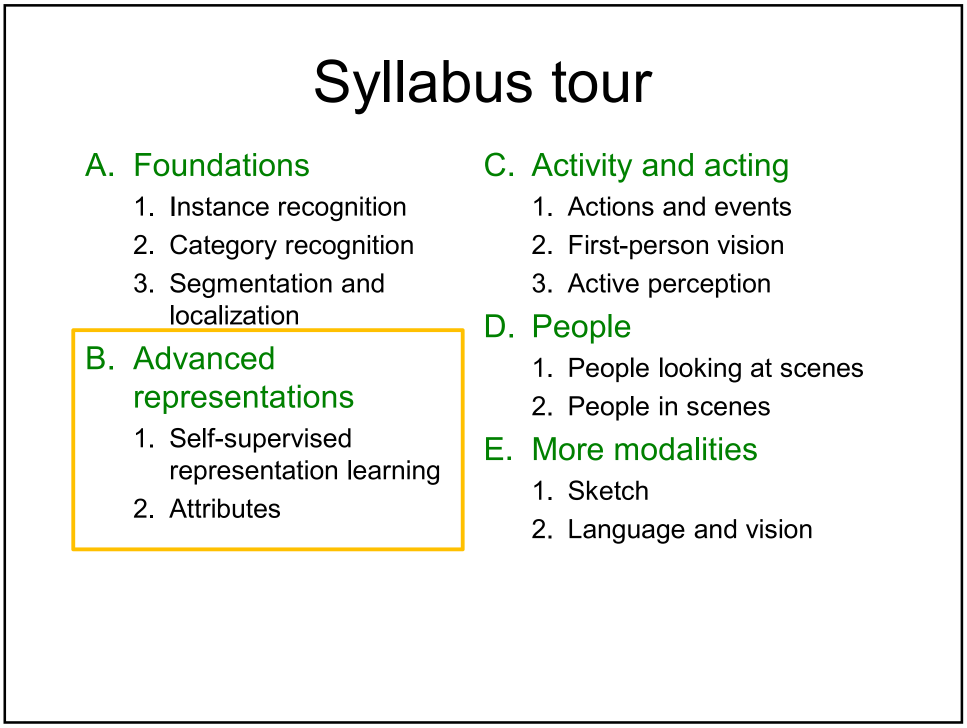
Segmentation and localization

Boundaries, regions

Semantic segmentation

Category-independent region ranking: “object proposals”

Object detection



Syllabus tour

A. Foundations
   1. Instance recognition
   2. Category recognition
   3. Segmentation and localization

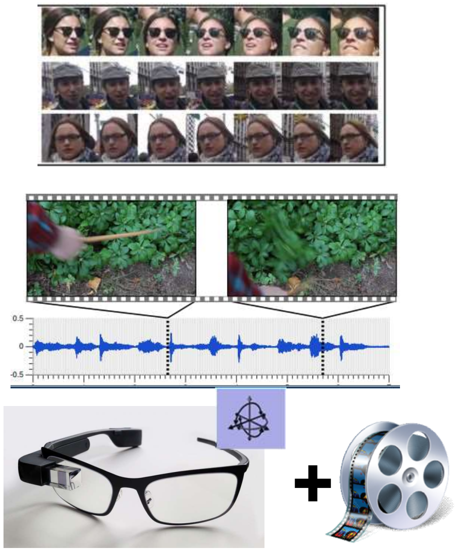
B. Advanced representations
   1. Self-supervised representation learning
   2. Attributes

C. Activity and acting
   1. Actions and events
   2. First-person vision
   3. Active perception

D. People
   1. People looking at scenes
   2. People in scenes

E. More modalities
   1. Sketch
   2. Language and vision

Self-supervised representation learning



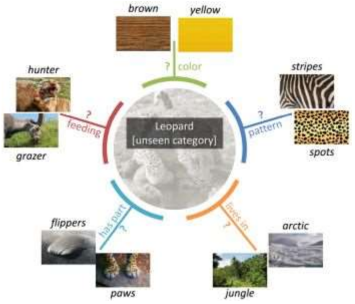
Unsupervised feature learning from "free" side information

(tracks in video, spatial layout in images, other modalities, ego-motion…

Attributes

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

Predicted Attribute
Ground Truth
Discovered Visual Elements
MORE
LESS
SF Violent Crime (Trained in SF)
SF Violent Crime (Ground Truth)



(a) Smiling (b) ? (c) Not smiling

(d) Natural (e) ? (f) Manmade

Syllabus tour

A. Foundations
   1. Instance recognition
   2. Category recognition
   3. Segmentation and localization

B. Advanced representations
   1. Self-supervised representation learning
   2. Attributes

C. Activity and acting
   1. Actions and events
   2. First-person vision
   3. Active perception

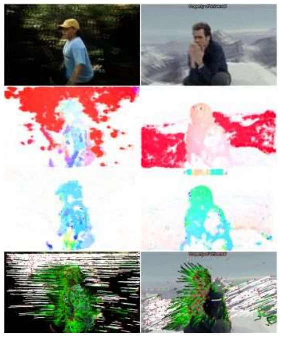
D. People
   1. People looking at scenes
   2. People in scenes

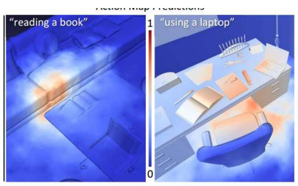
E. More modalities
   1. Sketch
   2. Language and vision

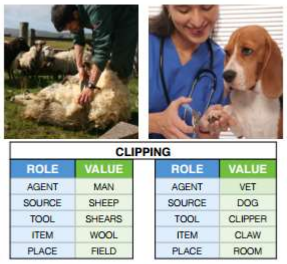
Actions and events

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

Video descriptors, interactions with objects and scenes.



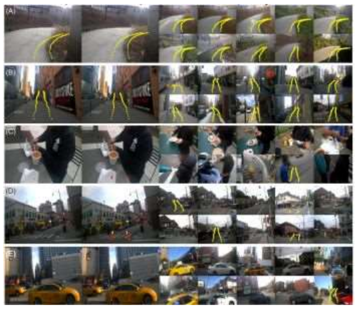




CLIPPING

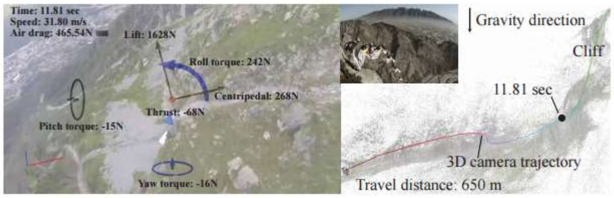
ROLE	VALUE	ROLE	VALUE
AGENT	MAN	AGENT	VET
SOURCE	SHEEP	SOURCE	DOG
TOOL	SHEARS	TOOL	CLIPPER
ITEM	WOOL	ITEM	CLAW
PLACE	FIELD	PLACE	ROOM

First-person vision



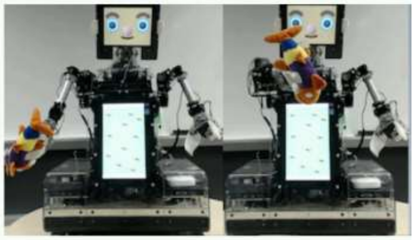
Egocentric wearable cameras.

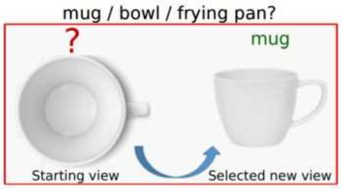
Actions and manipulated objects, gaze, discovering patterns and anomalies, temporal segmentation

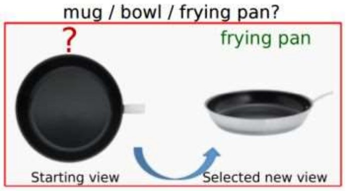


Active perception

- Learning how to move for recognition, manipulation. 3D objects and the next best view. Cost-sensitive recognition







Syllabus tour

A. Foundations
   1. Instance recognition
   2. Category recognition
   3. Segmentation and localization

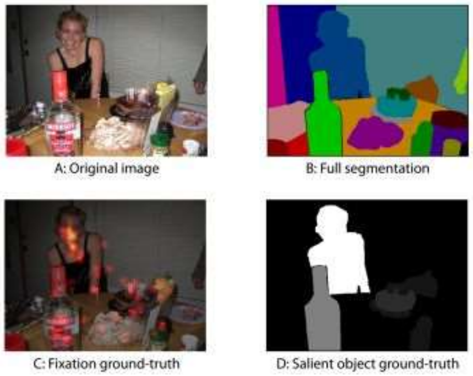
B. Advanced representations
   1. Self-supervised representation learning
   2. Attributes

C. Activity and acting
   1. Actions and events
   2. First-person vision
   3. Active perception

D. People
   1. People looking at scenes
   2. People in scenes

E. More modalities
   1. Sketch
   2. Language and vision

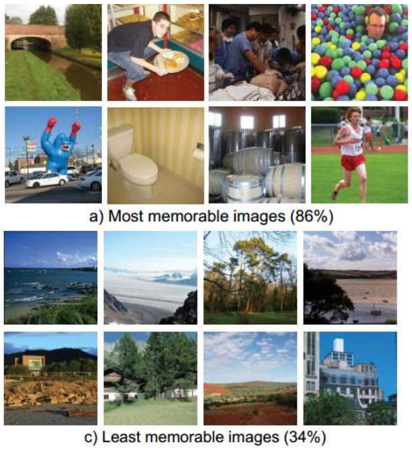
People looking at scenes


A: Original image

B: Full segmentation

C: Fixation ground-truth

D: Salient object ground-truth


a) Most memorable images (86%)

c) Least memorable images (34%)

- Predicting what gets noticed or remembered in images and video. Gaze, saliency, importance, memorability, mentioning biases.

People in scenes

(a) Alejandra
(b) Heather

- Analyzing people in the scene. Re-identification, attributes, gaze following, crowds.

Syllabus tour

A. Foundations
   1. Instance recognition
   2. Category recognition
   3. Segmentation and localization

B. Advanced representations
   1. Self-supervised representation learning
   2. Attributes

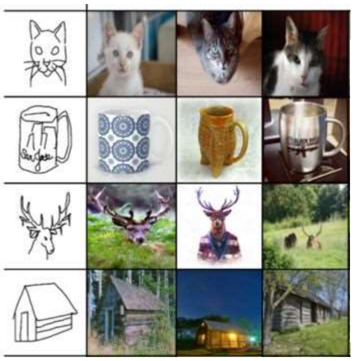
C. Activity and acting
   1. Actions and events
   2. First-person vision
   3. Active perception

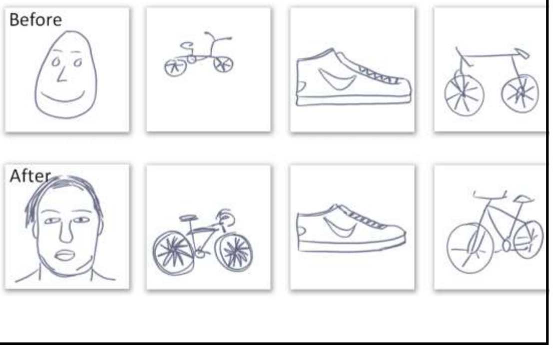
D. People
   1. People looking at scenes
   2. People in scenes

E. More modalities
   1. Sketch
   2. Language and vision

Sketches

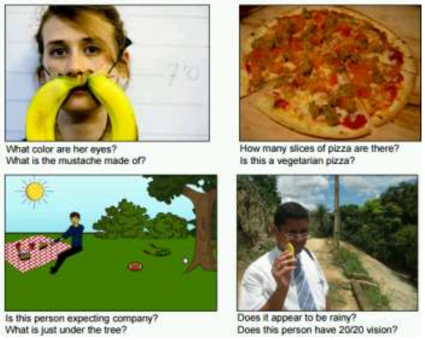
- Hand-drawn sketches and recognition. Retrieving natural images matching a sketch, forensics, interactive drawing, fine-grained retrieval.





Language and vision

- Connecting language and vision. Captioning, referring expressions, question answering, word-image embeddings, storytelling

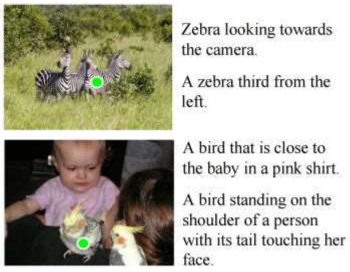


What color are her eyes?
What is the mustache made of?

How many slices of pizza are there?
Is this a vegetarian pizza?

Is this person expecting company?
What is just under the tree?

Does it appear to be rainy?
Does this person have 20/20 vision?



Zebra looking towards the camera.

A zebra third from the left.

A bird that is close to the baby in a pink shirt.

A bird standing on the shoulder of a person with its tail touching her face.

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

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