Advanced Topics in Computer Vision and Robotics

Jana Kosecka
http://cs.gmu.edu/~kosecka/cs884/
kosecka@cs.gmu.edu

Some slides thanks to S. Lazebnik, T. Berg, Fei-Fei Li, K. Grauman and others

Topics of the class

- State of the art of scene understanding
- Object, Scene, Human Activity Recognition

With applications to:

- Image Based Retrieval
- Image tagging
- Robot Perception, environment understanding

Logistics

- **Grading:** Homeworks, Presentations, Class Participation 60% Final exam/ project: 40%
- **Prerequisites:** Computer Vision, Robotics, AI, Data Mining, Pattern Recognition
- **Related Resources:** Material covered in CS682, CS685 and textbooks and recommended materials there
- **Lectures:** Introduction by an instructor, 3 paper presentations per class, discussions, each student will present one paper every second week; all students should read all papers to participate in discussion; programming homeworks every second week
- **Projects:** up to teams of 2 people
- **Dates**
  - Project proposals due
  - May week of finals final report due
  - Project presentations

- Required Software MATLAB (with Image Processing toolbox)
- Open CV library

Student Participation, Presentation

- 2-3 papers for each week discussing selected topic
- 10-15 minute presentation of one paper with slides
- Discuss the main idea of the paper
- The methods used, if the code is available demonstrate the method
- Provide opinion, compare related to the other papers on the same topic

- Presenter should stimulate the discussion on the paper

Today's Goals

- Brief Overview of Visual Recognition
- Recognition of objects, scenes, human activities
- List of topics

- Overview of Image features and basic approaches

Why study computer vision?

- Vision is useful: Images and video are everywhere!

Surveillance and security

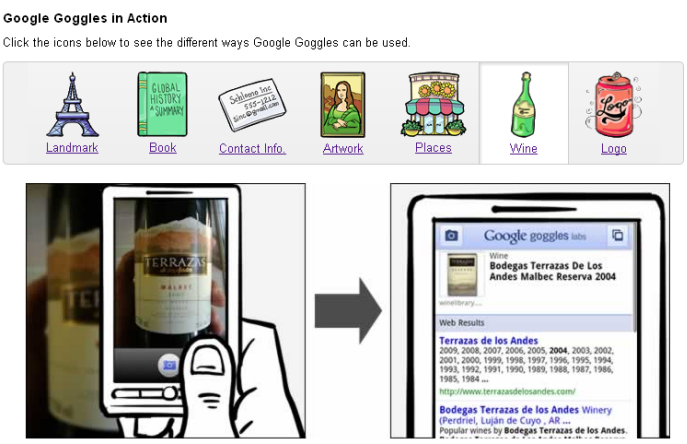
Medical and scientific images

- Associating semantic information, descriptions with images, can be done at different levels

- urban environment
- city
- street scene
- green and yellow buses

Everyday applications

- Everyday applications which exploit visual recognition technologies
- Product Search: Google Goggles, Mobile Visual Search



Object instance recognition

LaneHawk by EvolutionRobotics

"A smart camera is flush-mounted in the checkout lane, continuously watching for items. When an item is detected and recognized, the cashier verifies the quantity of items that were found under the basket, and continues to close the transaction. The item can remain under the basket, and with LaneHawk,you are assured to get paid for it… "

Source: S. Seitz

Face recognition: Apple iPhoto software

http://www.apple.com/ilife/iphoto/

Vision-based interaction (and games)

- Human pose estimation
- Activity Recognition

Xbox and Kinect sensor

Sony EyeToy

Assistive technologies

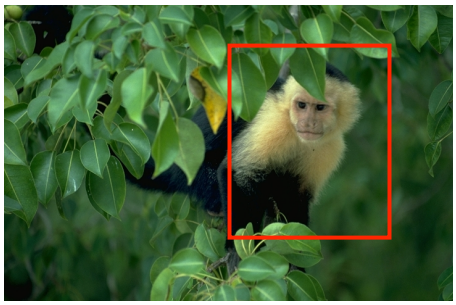
Types of recognition - Scene categorization

Street scene

Beach

Mountain

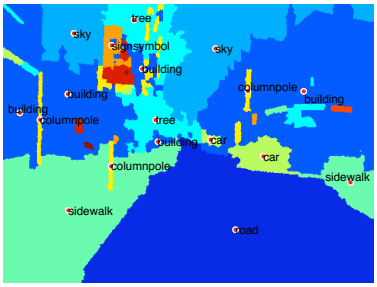
Types of recognition – Object Detection



- Object present in image
- Background may be correlated
- Localize object within the frame
- Bounding box or pixel-level segmentation

Types of recognition - Semantic Segmentation

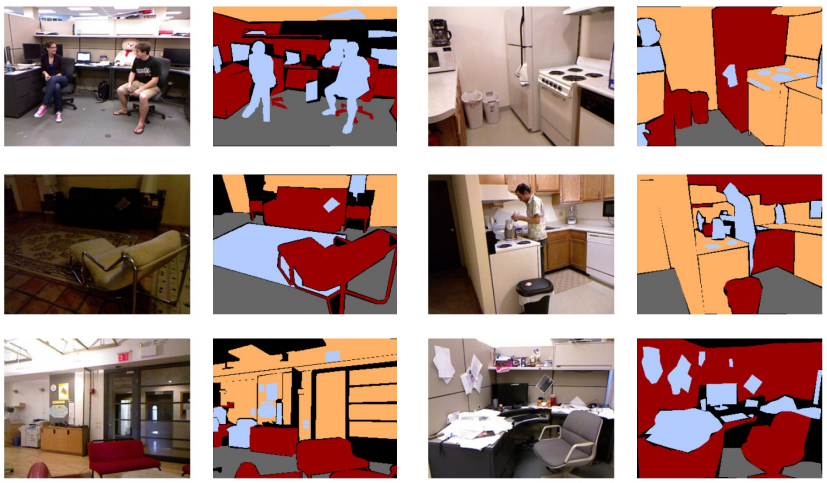
- Simultaneous Segmentation and Categorization



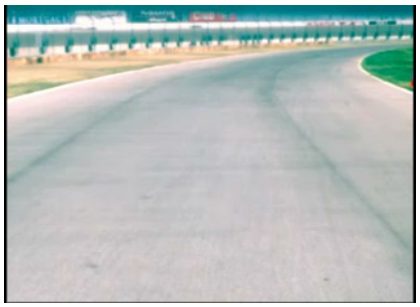
Types of recognition - Semantic Segmentation

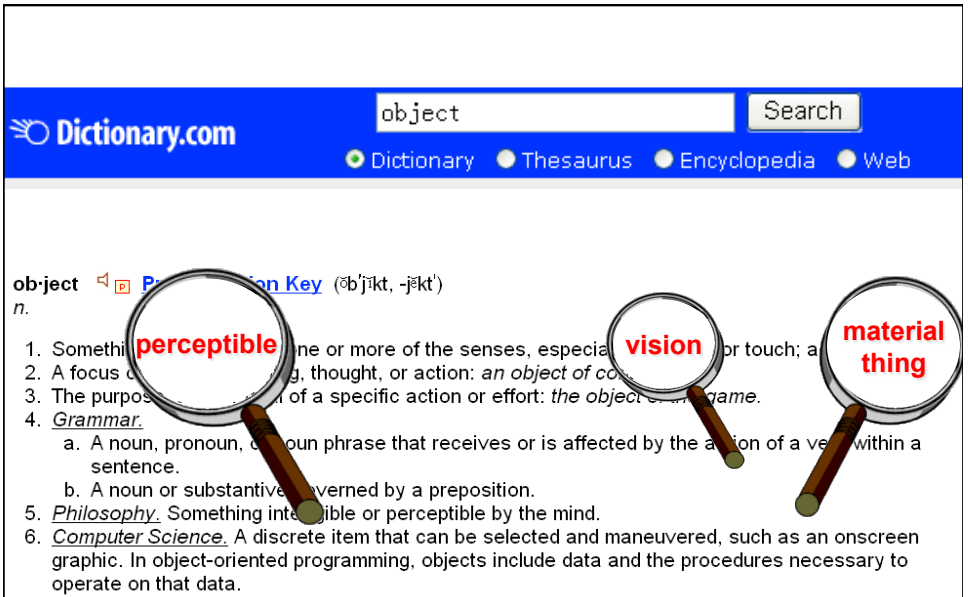
NYU v2 - Ground Truth

Ground
Struct.
Furnit.
Prop.



Terrain Classification – road/no road pixel classification





Types of recognition - Object categorization

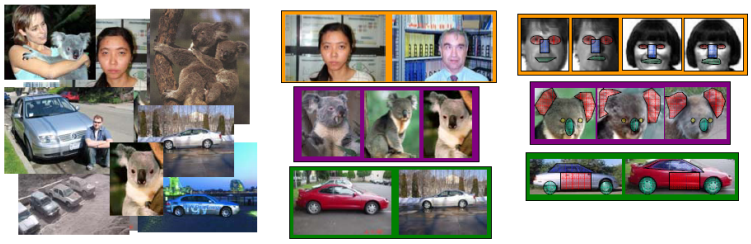
10,000 – 30,000 object categories

General Recognition Pipeline and Beyond

- Formulate Visual Recognition as Classification Problem
- Machine Learning, Pattern Recognition
- Given some number of training examples, learn the model of the class (category); test on the held out part of the dataset
- Earlier research has been driven by the creation of datasets, existing models had often bias
- How to move beyond the closed ‘dataset’ world
- Explore novel ways of supervision and/or use of unsupervised methods

Spectrum of supervision

Less — More



Learning approaches proceed in supervised way: need some labeled data

Unsupervised — "Weakly" supervised — Supervised

Definition depends on task

Caltech 101 & 256

http://www.vision.caltech.edu/Image_Datasets/Caltech101/
http://www.vision.caltech.edu/Image_Datasets/Caltech256/

Fei-Fei, Fergus, Perona, 2004

Griffin, Holub, Perona, 2007

The PASCAL Visual Object Classes Challenge (2005-2009)

http://pascallin.ecs.soton.ac.uk/challenges/VOC/

2008 Challenge classes:
Person: person
Animal: bird, cat, cow, dog, horse, sheep
Vehicle: aeroplane, bicycle, boat, bus, car, motorbike, train
Indoor: bottle, chair, dining table, potted plant, sofa, tv/monitor

The PASCAL Visual Object Classes Challenge (2005-2009)

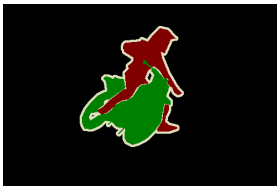
http://pascallin.ecs.soton.ac.uk/challenges/VOC/

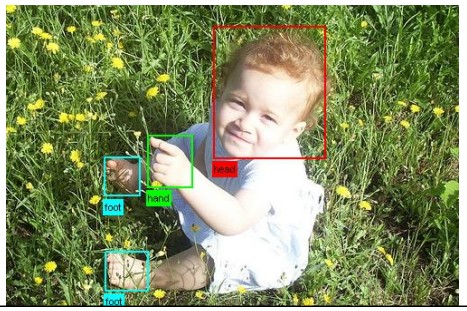
- Main competitions
  - **Classification:** For each of the twenty classes, predicting presence/absence of an example of that class in the test image
  - **Detection:** Predicting the bounding box and label of each object from the twenty target classes in the test image

The PASCAL Visual Object Classes Challenge (2005-2009)

http://pascallin.ecs.soton.ac.uk/challenges/VOC/

- "Taster" challenges
  - **Segmentation:** Generating pixel-wise segmentations giving the class of the object visible at each pixel, or "background" otherwise
  - **Person layout:** Predicting the bounding box and label of each part of a person (head, hands, feet)





Example Datasets

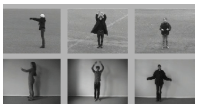
UIUC Cars (2004)
S. Agarwal, A. Awan, D. Roth

CMU/VASC Faces (1998)
H. Rowley, S. Baluja, T. Kanade

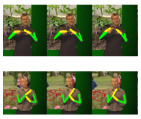
FERET Faces (1998)
P. Phillips, H. Wechsler, J. Huang, P. Raus

COIL Objects (1996)
S. Nene, S. Nayar, H. Murase

MNIST digits (1998-10)
Y LeCun & C. Cortes

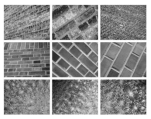


KTH human action (2004)
I. Leptev & B. Caputo

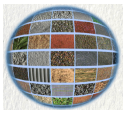


Sign Language (2008)
P. Buehler, M. Everingham, A. Zisserman

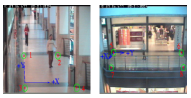
Segmentation (2001)
D. Martin, C. Fowlkes, D. Tal, J. Malik.



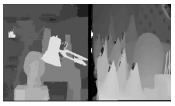
3D Textures (2005)
S. Lazebnik, C. Schmid, J. Ponce



CuRRET Textures (1999)
K. Dana B. Van Ginneken S. Nayar J. Koenderink



CAVIAR Tracking (2005)
R. Fisher, J. Santos-Victor J. Crowley



Middlebury Stereo (2002)
D. Scharstein R. Szeliski

Slide credit Fei-Fei Li

Caltech 5, Caltech 101

Slide credit Fei-Fei Li

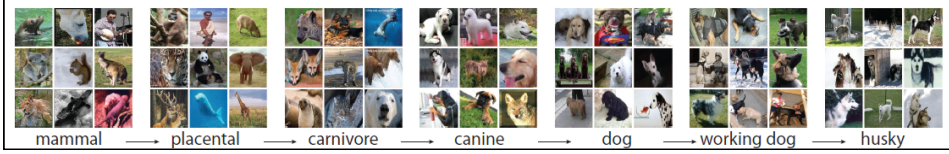
Example Datasets

Motorbike

Tiny Images [Torralba et al'07],
80 million tiny images

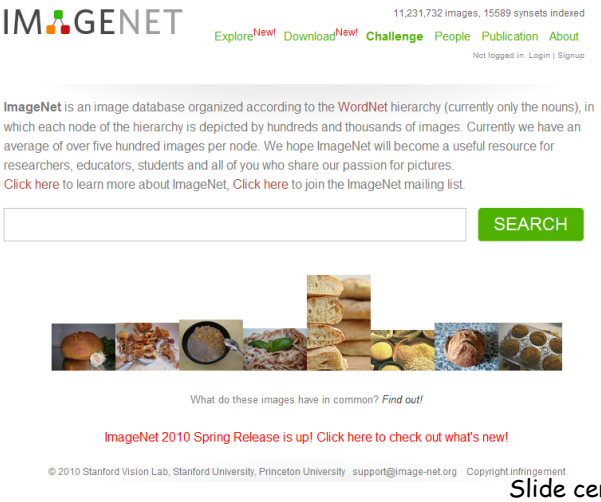
LabelMe
[Russel et al'05

ImageNet [Fei-Fei, 2008] 10K object categories, sync sets, ontology & word hierarchy



Large Scale Image Categorization

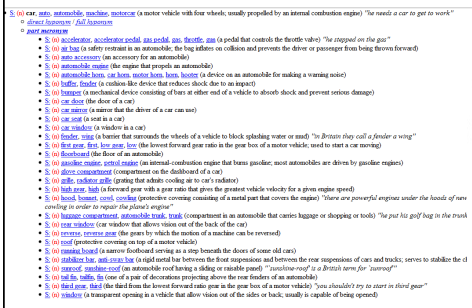
11 million images, 10,000 image categories 15,000+ synsets



Slide cerdit Fei-Fei Li

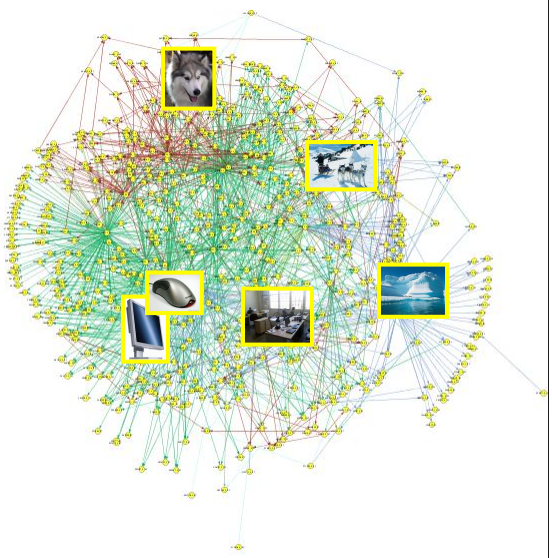
IMAGENET is a knowledge ontology

- Taxonomy
- Partonomy
- The “social network” of



IMAGENET is a knowledge ontology

- Taxonomy
- Partonomy
- The “social network” of visual concepts
  - Prior knowledge
  - Context
  - Hidden knowledge and structure among visual concepts



Slide cerdit Fei-Fei Li

Challenges of Semantic Understanding and Categorization in Robot Perception

- Images from Internet vs Real-World environments

Challenges

- Large amount of occlusions
- Large variations in size/scale
- Large variation in viewpoint
- Large variation in lighting
- Large Amount of Clutter
- Statistics of the data relevant in robotics setting differs

Opportunities

- Availability of video, 3D sensing
- Active sensing and exploration strategies

Robot Perception Opportunities

- Opportunities availability of 3D and video streams
- Large amounts of data
- Use of geometric cues
- Capabilities of active perception
- Exploiting better the statistics of environments where robots reside
- In many tasks object instance recognition is of bigger importance then object categorization
- Develop methods which separate the environments and context

Labeling with games

http://www.gwap.com/gwap/

Figure 1. Partners agreeing on an image in the ESP Game. *Neither player can see the other's guesses.*

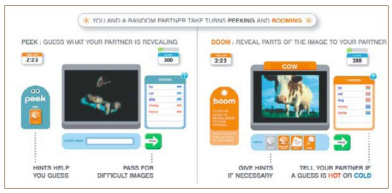
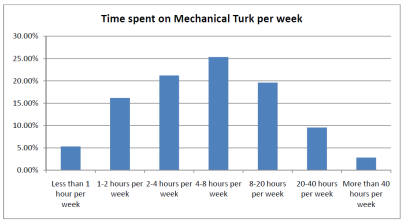


Figure 2. Peekaboom. *"Peek" tries to guess the word associated with an image slowly revealed by "Boom."*

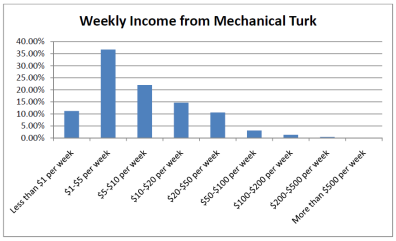
L. von Ahn, L. Dabbish, 2004; L. von Ahn, R. Liu and M. Blum, 2006

Knowledge Acquisition and Labeling Efforts

- Crowd Sourcing: Mechanical Turk
- Games, Web-Based Labeling tools
- July 2008: 0 images
- Dec 2008: 3 million images, 6000+ synsets
- April 2010: 11 million images, 15,000+ synsets



amazonmechanical turk
beta
Artificial Artificial Intelligence



Panos Ipeirotis, NYU, Feb, 2010 Slide credit Fei-Fei Li

Course Outline

- Local and Global Features, Overview of basic ML
- Object Instance Recognition, Object detection
- Part based Models, Region based Models
- Scenes and Image Context
- Attributes
- Action Recognition
- Saliency, Search, Scalability
- Images and text
- Domain Adaptation
- Active Learning
- Active Vision
- Unsupervised Methods, Applications