

## Advanced Topics in Computer Vision and Robotics

Jana Kosecka  
<http://cs.gmu.edu/~kosecka/cs884/>  
[kosecka@cs.gmu.edu](mailto: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

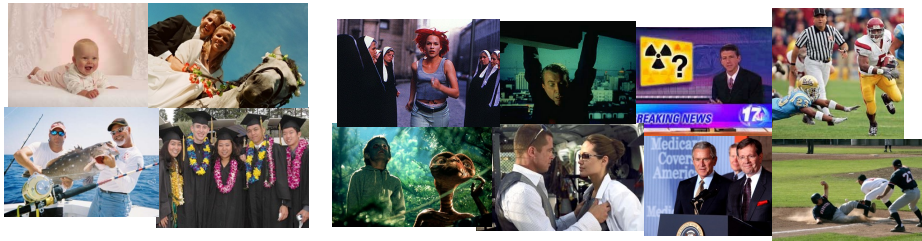
- 2-3 papers for each week discussing selected topic
- 10-15 minute presentation of one paper with slides
- Discuss the main idea of the paper
- The methods used, if the code is available demonstrate the method
- Provide opinion, compare related to the other papers on the same topic
- Presenter should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

## Today's Goals

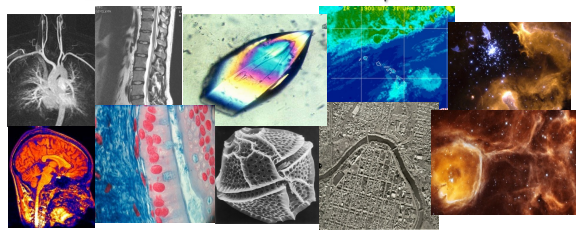
- Brief Overview of Visual Recognition
- Recognition of objects, scenes, human activities
- List of topics
  
- Overview of Image features and basic approaches

## 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

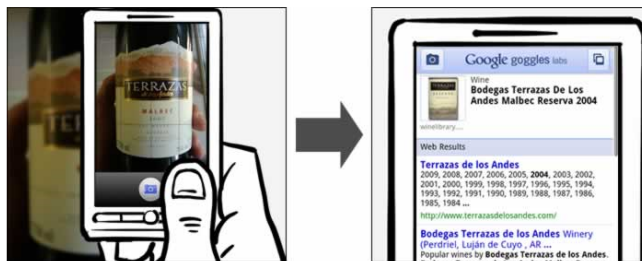
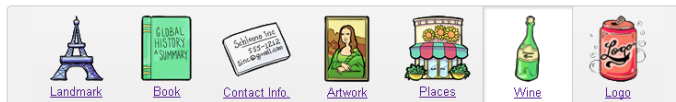


## Everyday applications

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

### Google Goggles in Action

Click the icons below to see the different ways Google Goggles can be used.



## Object instance recognition

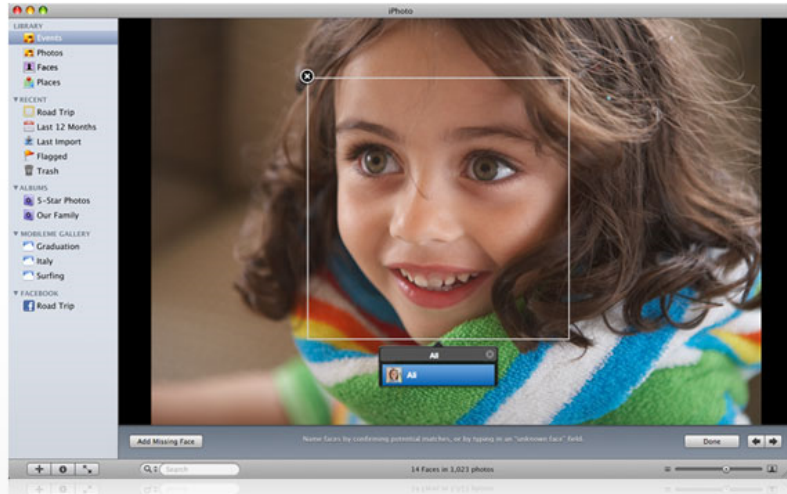


### LaneHawk by EvolutionRobotics

“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

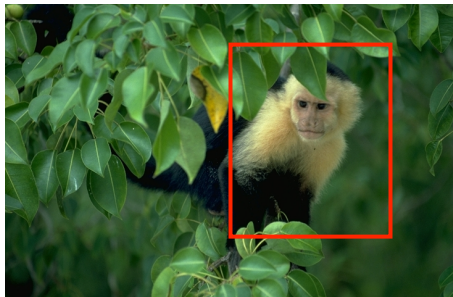


Beach



Mountain

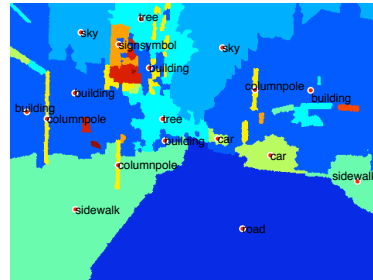
## Types of recognition – Object Detection



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

## Types of recognition - Semantic Segmentation

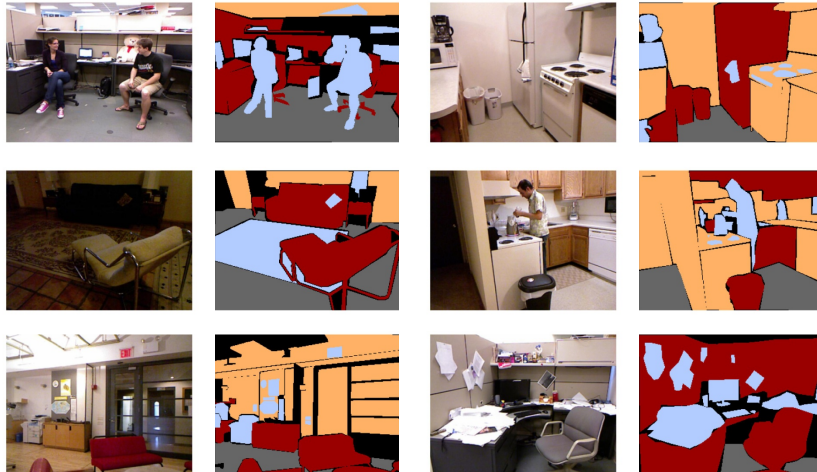
- Simultaneous Segmentation and Categorization



## Types of recognition - Semantic Segmentation

Ground Struct. Furnit. Prop.

NYU v2 - Ground Truth



## Terrain Classification – road/no road pixel classification



Dictionary.com

object  Search

Dictionary Thesaurus Encyclopedia Web

**object** **Perception Key** (ˈɒbjɪkt, -jɛkt)  
n.

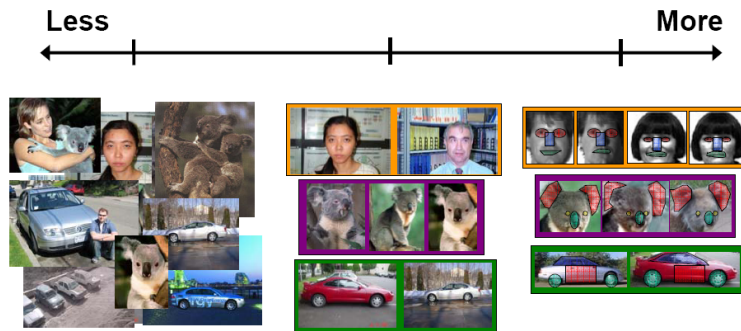
- 1. Something that can be perceived by one or more of the senses, especially sight or touch; a thing that is the object of a sense: *the object of his gaze*.
- 2. A focus of attention, thought, or action: *an object of curiosity*.
- 3. The purpose or goal of a specific action or effort: *the object of the game*.
- 4. Grammar.
  - a. A noun, pronoun, or noun phrase that receives or is affected by the action of a verb within a sentence.
  - b. A noun or substantive governed by a preposition.
- 5. Philosophy. Something intelligible or perceptible by the mind.
- 6. Computer Science. A discrete item that can be selected and maneuvered, such as an onscreen graphic. In object-oriented programming, objects include data and the procedures necessary to operate on that data.

**perceptible** **vision** **material thing**





# Spectrum of supervision



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/Caltech101/)  
[http://www.vision.caltech.edu/Image\\_Datasets/Caltech256/](http://www.vision.caltech.edu/Image_Datasets/Caltech256/)



Griffin, Holub,  
Perona, 2007

Fei-Fei, Fergus, Perona, 2004



## 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)

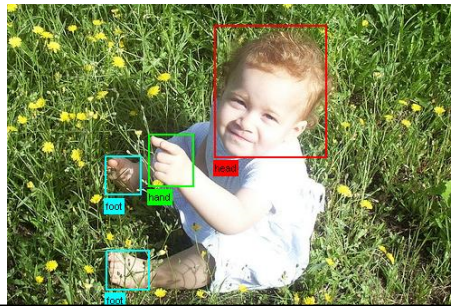
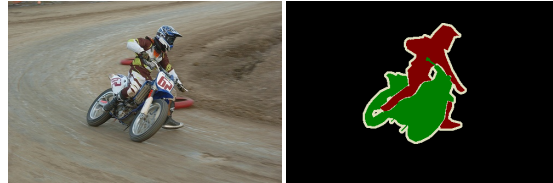
<http://pascallin.ecs.soton.ac.uk/challenges/VOC/>

- 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



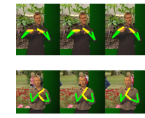
**COIL Objects (1996)**  
S. Nene, S. Nayar, H. Murase



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



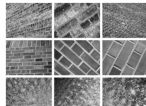
**KTH human action (2004)**  
I. Leptev & B. Caputo



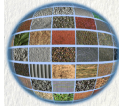
**Sign Language (2008)**  
P. Buehler, M. Everingham, A. Zisserman



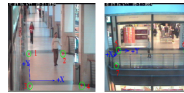
**Segmentation (2001)**  
D. Martin, C. Fowlkes, D. Tal, J. Malik



**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

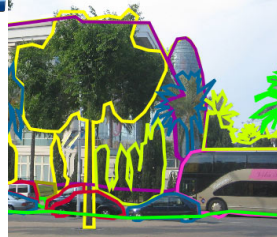
Slide credit Fei-Fei Li

## Example Datasets

### Motorbike



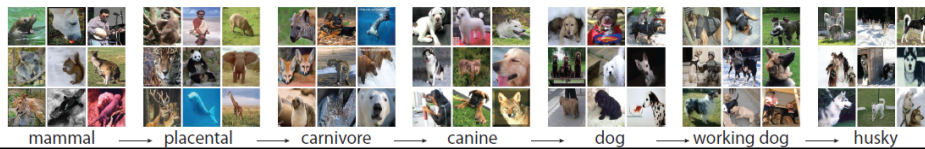
LabelMe  
[Russel et al'05]



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



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

IMAGENET

11,231,732 images, 15589 synsets indexed

[Explore](#) [New!](#) [Download](#) [New!](#) [Challenge](#) [People](#) [Publication](#) [About](#)

Not logged in. [Login](#) | [Signup](#)

ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by hundreds and thousands of images. Currently we have an average of over five hundred images per node. We hope ImageNet will become a useful resource for researchers, educators, students and all of you who share our passion for pictures.

[Click here](#) to learn more about ImageNet. [Click here](#) to join the ImageNet mailing list.

What do these images have in common? *Find out!*

[ImageNet 2010 Spring Release is up!](#) [Click here to check out what's new!](#)

© 2010 Stanford Vision Lab, Stanford University, Princeton University support@image-net.org Copyright infringement

Slide credit Fei-Fei Li

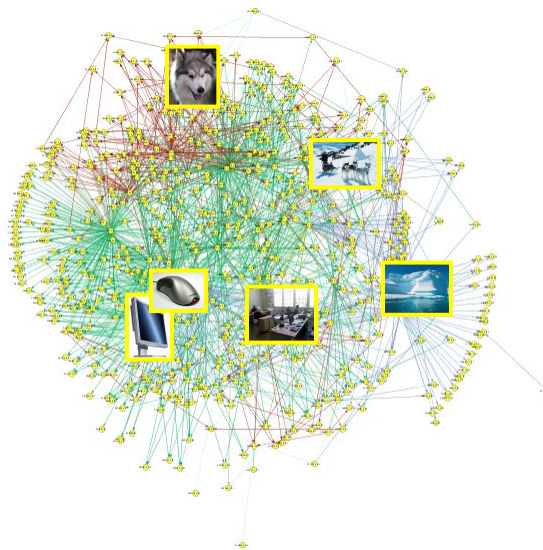
- Taxonomy
- Partonomy
- The “social network” of

5. (n) **car**, **auto**, **automobile**, **machine**, **motorcar** (a motor vehicle with four wheels, usually propelled by an internal combustion engine) "The needs a car to get to work"  
 6. **gear** **mechanism** (all) **mechanism**

- (n) **acceleratory**, **accelerative**, **pedal**, **gas**, **pedal**, **gas**, **throttle**, **gas** (a pedal that controls the throttle valve) "He stepped on the gas"
- (n) **air**, **bag** (a safety restraint in an automobile, the bag inflates on collision and prevents the driver or passenger from being thrown forward)
- (n) **anti**, **accident** (an accessory for an automobile)
- (n) **automobile**, **engine** (the engine that propels an automobile)
- (n) **automobile**, **horn**, **car**, **horn**, **motor**, **horn**, **horn**, **button** (a device on an automobile for making a warning noise)
- (n) **bumper**, **bumper** (a cushion-like device that reduces shock due to an impact)
- (n) **brake** (a mechanical device consisting of bars at either end of a vehicle to absorb shock and prevent serious damage)
- (n) **car**, **door** (the door of a car)
- (n) **car**, **driver** (a person that the driver of a car can use)
- (n) **car**, **seat** (a seat in a car)
- (n) **car**, **steering** (a window in a car)
- (n) **center**, **strip** (a bar that surrounds the wheels of a vehicle to block splashing water or mud) "The driver slip out of fender a wing"
- (n) **first**, **gear**, **first**, **low**, **gear**, **low** (the lowest forward gear ratio in the gear box of a motor vehicle, used to start a car moving)
- (n) **gas**, **pedal**, **gas** (the device of an automobile)
- (n) **gasoline**, **engine**, **internal**, **engine** (an internal-combustion engine that burns gasoline, most automobiles are driven by gasoline engines)
- (n) **gear**, **compartment** (compartment on the dashboard of a car)
- (n) **gear**, **indicator**, **light** (a light that indicates the vehicle's position on a road)
- (n) **high**, **gear**, **high** (a forward gear with a gear ratio that gives the greatest vehicle velocity for a given engine speed)
- (n) **hood**, **hood**, **hood**, **cover** (a protective covering consisting of a metal part that covers the motor) "There are powerful engines under the hood of new vehicles in order to repair the plane's engine"
- (n) **rear**, **view**, **mirror** (a mirror that allows vision out of the back of the car)
- (n) **rear**, **view**, **mirror** (a mirror that allows vision out of the back of the car)
- (n) **roof**, **roof** (a narrow horizontal surface as a step between the doors of some old cars)
- (n) **shock**, **absorber**, **shock**, **absorber** (a coil spring between the front suspension and between the rear suspension of cars and trucks, serves to stabilize the car)
- (n) **spring**, **spring** (a coil of wire of disc-shaped spring above the rear fenders of an automobile)
- (n) **tail**, **light**, **tail** (one of a pair of disc-shaped spring above the rear fenders of an automobile)
- (n) **third**, **gear**, **third** (the third from the lowest forward ratio gear in the gear box of a motor vehicle) "You shouldn't try to start in third gear"
- (n) **window** (a transparent opening in a vehicle that allows vision out of the sides or back, usually is capable of being opened)



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



Slide credit 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 than 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.

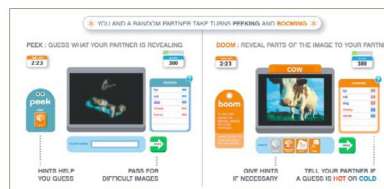
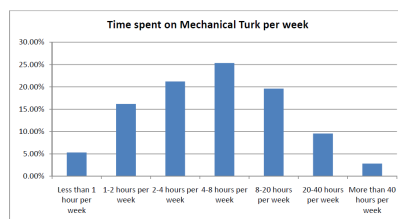


Figure 2. Peekaboom. "Peek" tries to guess the word associated with an image slowly revealed by "Boom."

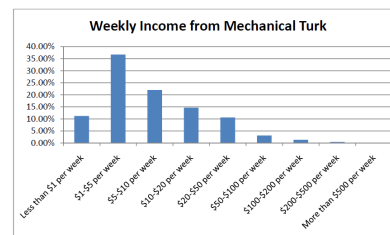
L. von Ahn, L. Dabbish, 2004; L. von Ahn, R. Liu and M. Blum, 2006

## 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



amazonmechanicalturk  
beta 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