CSE 591
Recognizing People, Objects, & Actions

Tamara Berg
Course Intro & Overview
Course Information

Instructor: Tamara Berg
Email: tlberg@cs.sunysb.edu
Office: 1411 Computer Science, 214 CEWIT
Webpage: http://tamaraberg.com/

Course: Tues/Thurs 11:20-12:40
Office Hours: Tues/Thurs 1:00-2:00 and by appt.
Course Webpage: http://tamaraberg.com/teaching/Spring_12/cse591
About Me

PhD from University of California, Berkeley 2007
Research Scientist at Yahoo! Research 2007-2008
Assistant Professor SBU 2008-present

Research

Digital Media, Computer Vision, Natural Language Processing, Web Scale projects, Integrating Words & Pictures
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
Automatically Created Face Dictionary

Browse the news according to the individuals present

Face Dictionary

Actress Jennifer Lopez was nominated for a Golden Raspberry or Razzie award as "the year’s worst actress" for "Enough" and "Maid in Manhattan" on February 10, 2003. Lopez is shown at the premiere of "Maid in Manhattan" on Dec. 8 and is joined by Madonna, Britney Spears, Winona Ryder and Angelina Jolie for the dubious honor. (Jeff Christensen/Reuters)
Baby Talk: Understanding and Generating Simple Image Descriptions

Girish Kulkarni, Visruth Premraj, Sagnik Dhar, Siming Li, Yejin Choi, Alexander C Berg, Tamara L Berg

CVPR 2011
This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant.
“This picture shows one person,
“This picture shows one person, one grass,
“This picture shows one person, one grass, one chair,
“This picture shows one person, one grass, one chair, and one potted plant."
“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass,
“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair."
“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair,
“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant.”
“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant.”
Us vs Humans

“This picture shows one person, one grass, one chair, and one potted plant. The person is near the green grass, and in the chair. The green grass is by the chair, and near the potted plant.”

H1: A Lemonade stand is manned by a blonde child with a cookie.
H2: A small child at a lemonade and cookie stand on a city corner.
H3: Young child behind lemonade stand eating a cookie.

• Sounds unnatural

UIUC pascal sentence dataset
Rashtchian, Young, Hodosh and Hockenmaier
NAACL HLT 2010
Im2Text: Describing Images Using 1 Million Captioned Photographs

Vicente Ordonez, Girish Kulkarni, Tamara L. Berg
Stony Brook University
NIPS 2011

An old bridge over dirty green water.
One of the many stone bridges in town that carry the gravel carriage roads.
A stone bridge over a peaceful river.

Generate Natural Sounding Descriptions
Results

**Good**

- Amazing colours in the sky at sunset with the orange of the cloud and the blue of the sky behind.
- A female Mallard duck in the lake at Luukki Espoo.
- Fresh fruit and vegetables at the market in Port Louis Mauritius.

**Bad**

- The cat in the window.
- The boat ended up a kilometre from the water in the middle of the airstrip.
Consortium for Digital Arts, Culture & Technology (cDACT)

http://emedia.art.sunysb.edu/cDACT/index.html

Christa Erickson, cDACT Director and Associate Professor of Art

Zabet Patterson, Assistant Professor of Art

Raiford Guins, Assistant Professor of Comparative Studies

Margaret Schedel, Assistant Professor of Music

Tamara Berg, Assistant Professor of Computer Science
About You?

• Name?
• Year?
• MS/PhD?
• Research Focus?
• Any related background in Computer Vision, Machine Learning, Graphics, AI?
Lecture Topics

Recognition of:

Objects - single instance or category based
People - faces, pedestrians, pose, and actions
Scenes - global image features, recognition in context, surfaces and geometry
Multi-modal Recognition – images + text, location, social context
Human-Centric recognition outputs – recognize “good” pictures, describing images with language
Objects
Finding Things and Observing Them

Image classification: Are there any dogs?

Photo credit: iansand – flickr.com

Slide credit: Derek Hoiem
Finding Things and Observing Them

Object Localization: Where are the dog(s)?
Finding Things and Observing Them

Verification: Is this a dog?

Slide credit: Derek Hoiem
Finding Things and Observing Them

Instance categorization: Is this my dog fluffy?
Bag-of-features models

Object → Bag of ‘words’

source: Svetlana Lazebnik
Spatial Models

Model: Parts and Structure
Spatial Models

• Evaluate correspondence based on:
  – Similarity of *appearance* near feature points
  – Similarity in *configuration* of the feature points

Spatial Models

Object Recognition: applications

Photo Search

Robots

Slide credit: Derek Hoiem
People
Face Detection

Sliding window approach: for each possible rectangular region in the image asks the question, is there a face here?

Now even integrated into many consumer digital cameras.

Face Recognition

Pedestrian detection

Bastian Leibe, Edgar Seemann, and Bernt Schiele
Pose Recognition

Mori, Ren, Efros and Malik
Recognizing Actions

KTH Dataset

Figure from Laptev et al. 2008

Slide credit: Derek Hoiem
Recognizing Actions

Figure from Laptev et al. 2008
Reading Emotions
People: applications

Video Search

Security

Slide credit: Derek Hoiem
Scenes
Scene Recognition

office

kitchen

living room

bedroom

store

industrial

tall building*

inside city*

street*

highway*

coast*

open country*

mountain*

forest*

suburb

Lazebnik, Schmid, Ponce
Recognizing stuff/surfaces

SKY

WATER

SAND

Slide credit: Derek Hoiem
Recognizing stuff/surfaces

Automatic Photo Popup
Derek Hoiem, Alexei A. Efros, Martial Hebert
From recognition to rough geometry

Automatic Photo Popup
Derek Hoiem, Alexei A. Efros, Martial Hebert
Putting it together
Putting it together

Recovering the Spatial Layout of Cluttered Rooms
Varsha Hedau, Derek Hoiem, David Forsyth
Recognition in Context

(a) Isolated object  (b) Object in context  (c) Low-res Object

Torralba, Murphy, Freeman and Rubin
Recognition in Context

Rabinovich, Vedaldi, Galleguillos, Wiewiora, Belongie
Finding Things and Observing Them

Description: Furry, small, nice, side view

Slide credit: Derek Hoiem
Describing objects by attributes

Ali Farhadi, Ian Endres, Derek Hoiem, David Forsyth
Recognition with multi-modal data
**Giraffe**

Learn about giraffe teeth on our video [Giraffe Treat Jar Enrichment](#).
Test your knowledge - take our [quiz](#).

**SWAHILI NAME:** Twiga

**PHYSICAL CHARACTERISTICS:**

The Giraffe is the tallest animal in the world. Males may be 16-18 feet tall and weigh up to 2,000 pounds. Females are usually lighter and about two feet shorter. The giraffe's front legs are only slightly longer than the back ones, the height of the fore part of the body being largely due to the heavy muscular development of the base of the neck.

The long neck has the usual seven vertebrae of most mammals, although each is greatly elongated. The giraffe's soup-plate-sized hooves are used as offensive weapons, usually in the defense of the calves. The powerful kick from the front feet can kill a lion.

The giraffe is one of the few ruminant born with horns. Both sexes have horns which are covered with skin. The horns of males are thicker and heavier and are used in fights between males.

The giraffe's canine teeth are splayed out in two or three lobes to comb the leaves off shoots. Their long black tongue, which can be extended 18 inches, is used to gather food into the mouth. To compensate for the sudden increase in blood pressure when the head is lowered, the giraffe has very elastic blood vessels and valves in the venous system of the neck.

The Nubian giraffe's color scheme varies considerably in pattern, but consists of dark-redish to chestnut brown blotches of various shapes and sizes on a buff ground color. The coloration darkens with age.

The reticulated giraffe has large, four-sided, liver-colored spots separated by a sharply defined network of narrow white lines.

**DISTRIBUTION:**

Giraffes inhabit savanna country where there are Acacia and climbing leguminous plants. They are found in Africa south of the Sahara to Cape Town. The reticulated giraffe is found in East Africa.
Tags or keywords + images

Flickr
>5 billion photos
3-5 million new/day

Tags: canon, eos, macro, japan, vacation, frog, animal, toad, amphibian, pet, eye, feet, mouth, finger, hand, prince, photo, art, light, photo, flickr, blurry, favorite, nice.
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.
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.
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.

*It's the perfect party dress. With distinctly feminine details such as a wide sash bow around an empire waist and a deep scoopneck, this linen dress will keep you comfortable and feeling elegant all evening long.*

* Measured 38" from center back, hits at the knee.
* Scoop neck, full skirt.
* Hidden side zip, fully lined.
* 100% Linen. Dry clean.
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.

It's the perfect party dress. With distinctly feminine details such as a wide sash bow around an empire waist and a deep scoopneck, this linen dress will keep you comfortable and feeling elegant all evening long.
* Measures 38" from center back, hits at the knee.
* Scoopneck, full skirt.
* Hidden side zip, fully lined.
* 100% Linen. Dry clean.
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
Words & Pictures for Clustering

- Higher level nodes emit more general words and blobs (e.g. sky)
- Moderately general words and blobs (e.g. sun, sea)
- Lower level nodes emit more specific words and blobs (e.g. waves)

Examples:
- Sun
- Sky
- Sea
- Waves
Words & Pictures for Retrieval

Words

Present range of the Red-legged Frog in British Columbia
Present range of the Tailed Frog in British Columbia
Present range of the Wood Frog in British Columbia

Words + Pictures

© Will Brown
© Ed Nied Jr.
© Tom Jankie-Mills Ignition

Freddo's Frog Gallery

Find Birthday Presents for your Kids at the Zebra Toy Shop! Click Here!
Location & Pictures

What can you say about where these photos were taken?
Results
Sierra and Patrick

Mildred and Lisa

Who’s who?
Human-centric recognition outputs
Automatically describing pictures...

**Good**

- Amazing colours in the sky at sunset with the orange of the cloud and the blue of the sky behind.
- A female Mallard duck in the lake at Luukki Espoo.
- Fresh fruit and vegetables at the market in Port Louis Mauritius.

**Bad**

- The cat in the window.
- The boat ended up a kilometre from the water in the middle of the airstrip.
A quick experiment

You will see a mask, then image, then mask.

What do you see?
Finding good photos
(separating the wheat from the chaff)
What’s “good”?

Highly representative examples for a class

Aesthetically pleasing images
Course Organization

- Lectures
- Paper Presentations
- 3 HWs
- Course Project
Lectures & Papers

• Some lectures to introduce concepts and start off the course

• Each paper reading day will cover about 2 research papers.

• Students are expected to read each research paper (there will be some simple in-class quizzes to ensure you read the papers).
There will be 3 assignments during the first 1-2 months of class.

This will be relatively simple assignments related to recognition to expose students to computer vision and get you started on implementation.

Homework should be completed in matlab and done individually. *Academic honesty* – all code submitted for assignments should be your own.
Programming

We will be implementing the assignments in Matlab - a software package with useful built in functionality for working with image and video data.

If you haven’t used Matlab please complete a tutorial before next class – some links are provided on the course webpage, but there are lots of tutorials available online.

Matlab is available in all departmental computing labs (2126,2114,2129).

Can be purchased from Mathworks for home use for $99: http://www.mathworks.com/academia/student_version/
Student Paper Presentations

- You be the professor.
- Prepare a presentation and lead a discussion on a research paper including your opinion on pros/cons etc.
- Feel free to run your presentation by me in office hours beforehand.
You will pick a project idea about 1.5 months into semester - I can help suggest or refine project ideas in office hours.

Projects can be implementing/modifying one of the papers we discuss or original research.

There will be a few status updates over the semester where you will briefly describe progress and we can provide suggestions.

Projects can be completed alone or in pairs.
Grading

This will be a very hands-on course and project focused. Therefore, there will be no formal exams.

Grading will be based on:
- 35% assignments, 35% project, 10% participation, 10% paper quizzes, 10% paper presentation.

For each assignment and project milestone you should turn in your code + a web page describing your results and displaying any related images or video (by email to cse591@gmail.com).

For assignments each day late will subtract 10% from the assignment’s grade.
Worried about the material?

• I will present an overview of computer vision next class.
• We will discuss other technical topics as we go along.
• Try to grasp the main ideas (novel contribution, what’s interesting, what do you think). Don’t sweat the small stuff.
• Questions are always welcome!
• Good reference – Rick Szeliski’s book (online draft version here: http://szeliski.org/Book/)
For Next Time

• Read—Sections 1.1-1.2 of Rick Szeliski’s book (http://szeliski.org/Book/)

• Get access to matlab (I recommend installing it on your own machine if possible). Do a tutorial if you haven’t used matlab before.

• Look over the paper list – choose one you want to present (and a back-up)