CSE 595 Words and Pictures

Tamara L. Berg

SUNY Stony Brook
Class Info

CSE 595: Words & Pictures

Instructor: Tamara Berg  (tlberg@cs.sunysb.edu)
Office: 1411 Computer Science
Lectures: Tues/Thurs 11:20-12:40pm Rm 2129 CS
Office Hours: Tues/Thurs 3:40-5:10pm

Course Webpage:
http://tamaraberg.com/teaching/Spring_11/wordspics
About Me

• Joined Stony Brook in 2008
  – PhD from UC Berkeley 2007.
  – 2007-2008 Yahoo! Research

• Research in computer vision and natural language processing - combining information from multiple forms of digital media for applications like image search and recognition.
You?
MS/PhD?
Experience in Comp Vision or NLP?
Matlab?
What’s in this picture?
What does the picture tell us?

Green, textured region – maybe tree?

Fuzzy black thing with a face-like part -- maybe an animal?
What do the words tell us?

Tags:

leaves, endangered, green, i love nature, chennai, nilgiri langur, monkey, forest, wildlife, perch, black, wallpaper, ARK OF WILDLIFE, topv111, WeeklySurvivor, top20HallFame, topv333, 100v10f, captive, simian
What do words+picture tell us?

Tags: leaves, endangered, green, i love nature, chennai, nilgiri langur, monkey, forest, wildlife, perch, black, wallpaper, ARK OF WILDLIFE, topv111, WeeklySurvivor, top20HallFame, topv333, 100v10f, captive, simian
Consumer Photo Collections

Flickr – 3+ billion photographs, 3-5 million uploaded per day

Over the hills and far away
Road, Hills, Germany, Hoffenheim, Outstanding Shots, specland, Baden-Wuerttemberg

Heavenly

End of the world - Verdens Ende - The lighthouse 1
Verdens ende, end of the world, norway, lighthouse, ABigFave, vippefyr, wood, coal
Museum and Library Collections

Fine Arts Museum of San Francisco (82,000 images)
- Woman of Head Howard H G Mrs Gift America North bust States United Sculpture marble
- Bowl stemmed small Iridescent glass

New York Public Library Digital Collection
- The new board walk, Rockaway, Long Island
- Part of New England, New York, east New Jersey and Long Island
Web Collections

Billions of Web Pages

Tree

From Wikipedia, the free encyclopedia

http://en.wikipedia.org/wiki/Trees

A tree is a large, perennial, woody plant. Though there is no set definition regarding minimum size, the term generally applies to plants at least 6 m (20 ft) high at maturity having secondary branches supported on a main stem or stems (see strubs for comparison). Most trees exhibit a broad apical dominance, though this is not always the case (Mitchell, 1978). Compared with most other plant forms, trees are long-lived. A few species of trees grow to 115 m (375 ft) tall and some can live for several thousand years.

Trees are an important component of the natural world, providing oxygen and reducing carbon dioxide concentrations, and as well as moderating ground temperatures and improving air quality. These traits could potentially alleviate climate warming.

Contents

- 1 Classification
- 2 Morphology
- 3 Life cycle
- 4 Champion trees
- 5 Giant trees
- 6 Major tree groups
- 7 Water

Tillamook Rock, OR

Description: One mile west of Tillamook Head, a rock rises from the sea. In the shape of a wedge-shaped, it is reached by a stone staircase. A view from the lighthouse shows the sea and the sky. The lighthouse is a popular attraction for tourists.

Flower Picture Gallery

Flowers are important to pollination, and we all have our favourite flowers, whether they are bright or subtle. They can be seen in many places. This is a photo of a flower from the gallery.

Flower Picture Gallery

Flowers are important to pollination, and we all have our favourite flowers, whether they are bright or subtle. They can be seen in many places. This is a photo of a flower from the gallery.

Flower Picture Gallery

Flowers are important to pollination, and we all have our favourite flowers, whether they are bright or subtle. They can be seen in many places. This is a photo of a flower from the gallery.
OUTSIDE IN THE RAIN THE SENATOR WEARING HIS UH BASEBALL CAP A BOSTON RED SOX CAP AS HE TALKED TO HIS SUPPORTERS HERE IN THE RAIN THE UH SENATOR THEY'RE DOING HIS BEST TO TRY TO MAKE HIS CASE THAT HE WILL BE THE MAN FOR THE MIDDLE CLASS AND UH TRY TO CONVINCE HIS SUPPORTERS TO EXPRESS THEIR SUPPORT THROUGH A VOTE ON TUESDAY IN THERE WE ARE TWENTY FOUR HOURS FROM THE GREAT MOMENT THAT THE WORLD IN AMERICA IS WAITING FOR IT I NEED TO YOU IN THESE HOURS TO GO OUT AND DO THE HARD WORK NOT ON THOSE DOORS MAKE THOSE PHONE CALLS TO TALK TO FRIENDS TAKE PEOPLE TO THE POLLS HELP US CHANGE THE DIRECTION OF THIS GREAT NATION FOR THE BETTER CAN YOU IMAGINE A UH SENATOR BEGINNING HIS DAY IN FLORIDA TODAY
Consumer Products

Soft and glossy patent calfskin trimmed with natural vachetta cowhide, open top satchel for daytime and weekends, interior double slide pockets and zip pocket, seersucker stripe cotton twill lining, kate spade leather license plate logo, imported.
2.8" drop length
14"h x 14.2"w x 6.9"d
Katespade.com

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.
bananarepublic.com

Internet retail transactions in 2006, 2007 of $145 billion, $175 billion (Forrester Research).
Lots of Data!
What do we want to do?
What do we want to do?

Organize
Search
Browse
What do we want to do?

Organize

Search

Browse

Fine Arts Museum of San Francisco (82,000 images)

bowl stemmed small Iridescent glass

Woman of Head Howard H G Mrs Gift America North bust States United Sculpture marble
What do we want to do?

Organize
Search
Browse

What do we want to do?

Organize

Search

Browse

Image Search circa 2007
What do we want to do?

Organize

Search

Browse

Image Search now
What do we want to do?

Organize

Search

Browse

The results of the “river” and “tiger” query.

Kobus Barnard and David Forsyth
Learning the Semantics of Words & Pictures,
What do we want to do?

Organize

Search

Browse

Image re-ranking for “monkey”

Tamara L Berg, David A Forsyth, Animals on the Web CVPR 2006
What do we want to do?

Organize

Search

Browse

Visual shopping at like.com
What do we want to do?

Organize
Search
Browse

Visual attribute discovery

Tamara L Berg, Alexander C Berg, Jonathan Shih
Automatic Attribute Discovery and Characterization from Noisy Web Data
ECCV 2010
What do we want to do?

Organize
Search
Browse

Visual attribute discovery

J. Wang, K. Markert, and M. Everingham.
"Learning models for object recognition from natural language descriptions" BMVC 2009.
Types of Words & Pictures
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-reddish 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.
General web pages

Improving Search

Image re-ranking for “monkey”

Tamara L Berg, David A Forsyth, Animals on the Web CVPR 2006
General web pages

Figure 6. Top ranked 36 images of zebra, wristwatch and car using the text+vision algorithm of figure 5. Red boxes indicate false positives.

Mining to build big computer vision data sets.

Harvesting Image Databases from the Web
Schroff, F., Criminisi, A. and Zisserman, A.
General web pages

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-reddish 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: canon, eos, macro, japan, frog, animal, toad, amphibian, pet, eye, feet, mouth, finger, hand, prince, photo, art, light, photo, flickr, blurry, favorite, nice.
Tags or keywords + images

Annotating regions with keywords

Pinar Duygulu, Kobus Barnard, Nando de Freitas, and David Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary" ECCV 2002.
Using tags and similar images for novel image classification

Gang Wang, Derek Hoiem, and David Forsyth,
Building text features for object image classification.
Tags or keywords + images

Frog on a hand

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

Pros?

Cons?
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
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

Captions provide direct information about depiction!
Captioned images for face and pose labeling

(a) Four sets ... **Roger Federer** prepares to **hit a backhand** in a quarter-final match with **Andy Roddick** at the US Open.

(b) US Democratic presidential candidate Senator **Barack Obama** waves to supporters together with his wife **Michelle Obama** standing beside him at his North Carolina and Indiana primary election night rally in Raleigh.

**Who's Doing What: Joint Modeling of Names and Verbs for Simultaneous Face and Pose Annotation**
Jie Luo, Barbara Caputo, Vittorio Ferrari
NIPS 2009
Video with transcripts
Video with transcripts for face labeling

M. Everingham, J. Sivic, and A. Zisserman.
Hello! My name is... Buffy' - Automatic naming of characters in TV video
BMVC 2006.
His Royal Highness from Saudi Arabia wanted to know about the history of the trees

and maybe take out a tree from somewhere and letting in a bit more light or something like that

a 50 ft crane, is attempting tree planting on a mammoth scale in readiness for a grand occasion

P. Buehler, M. Everingham, and A. Zisserman.
"Learning sign language by watching TV (using weakly aligned subtitles)". CVPR 2009.
Videos and text-based webpages

Z. Wang, M. Zhao, Y. Song, S. Kumar and B. Li
YouTubeCat: Learning to Categorize Wild Web Videos
Beyond traditional object class recognition
Traditional Recognition

- person
- car
- shoe
Beyond traditional recognition
Beyond traditional recognition

“It was an arresting face, pointed of chin, square of jaw. Her eyes were pale green without a touch of hazel, starred with bristly black lashes and slightly tilted at the ends. Above them, her thick black brows slanted upward, cutting a startling oblique line in her magnolia-white skin—that skin so prized by Southern women and so carefully guarded with bonnets, veils and mittens against hot Georgia suns” – Scarlett O’Hara, Gone with the Wind.
Attributes

Visual attribute learning from text

Tamara L Berg, Alexander C Berg, Jonathan Shih
Automatic Attribute Discovery and Characterization from Noisy Web Data
ECCV 2010
Object relationships
Object relationships

Car is on the street

Object relationships – prepositions & adjectives

Beyond Nouns: Exploiting prepositions and comparative adjectives for learning visual classifiers
Abhinav Gupta and Larry S. Davis
In ECCV 2008
Visually descriptive language offers:

1) information about the world, especially the visual world.
2) training data for how people construct natural language to describe imagery.

“It was an arresting face, pointed of chin, square of jaw. Her eyes were pale green without a touch of hazel, starred with bristly black lashes and slightly tilted at the ends. Above them, her thick black brows slanted upward, cutting a startling oblique line in her magnolia-white skin— that skin so prized by Southern women and so carefully guarded with bonnets, veils and mittens against hot Georgia suns” – Scarlett O’Hara, Gone with the Wind.
Generating descriptions for images
**Table:**

<table>
<thead>
<tr>
<th>Triplets</th>
<th>Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>(pet, sleep, ground)</td>
<td>see something unexpected.</td>
</tr>
<tr>
<td>(dog, sleep, ground)</td>
<td>Cow in the grassfield.</td>
</tr>
<tr>
<td>(animal, sleep, ground)</td>
<td>Beautiful scenery surrounds a fluffy sheep.</td>
</tr>
<tr>
<td>(animal, stand, ground)</td>
<td>Dog hearding sheep in open terrain.</td>
</tr>
<tr>
<td>(goat, stand, ground)</td>
<td>Cattle feeding at a trough.</td>
</tr>
<tr>
<td>(furniture, place, furniture)</td>
<td>Refrigerator almost empty.</td>
</tr>
<tr>
<td>(furniture, place, room)</td>
<td>Foods and utensils.</td>
</tr>
<tr>
<td>(furniture, place, home)</td>
<td>Eatables in the refrigerator.</td>
</tr>
<tr>
<td>(bottle, place, table)</td>
<td>The inside of a refrigerator apples, cottage cheese, tupperwares and lunch bags.</td>
</tr>
<tr>
<td>(display, place, table)</td>
<td>Squash apenny white store with a hand statue, picnic tables in front of the building.</td>
</tr>
<tr>
<td>(transportation, move, track)</td>
<td>A man stands next to a train on a cloudy day</td>
</tr>
<tr>
<td>(bike, ride, track)</td>
<td>A backpacker stands beside a green train</td>
</tr>
<tr>
<td>(transportation, move, road)</td>
<td>This is a picture of a man standing next to a green train</td>
</tr>
<tr>
<td>(pet, sleep, ground)</td>
<td>There are two men standing on a rocky beach, smiling at the camera.</td>
</tr>
<tr>
<td>(bike, ride, road)</td>
<td>This is a person laying down in the grass next to their bike in front of a strange white building.</td>
</tr>
<tr>
<td>(display, place, table)</td>
<td>A black laptop is connected to a black Dell monitor</td>
</tr>
<tr>
<td>(furniture, place, furniture)</td>
<td>This is a dual monitor setup</td>
</tr>
<tr>
<td>(furniture, place, furniture)</td>
<td>Old school Computer monitor with way to many stickers on it</td>
</tr>
</tbody>
</table>

**Fig. 3.** Generating sentences for images: We show top five predicted triplets in the middle column and top five predicted sentences in the right column.

*Farhadi, A., Hejrati, M., Sadeghi, M.A., Young, P., Rashtchian, C., Hockenmaier, J., Forsyth, D.A.,*  
*Every Picture Tells a Story: Generating Sentences from Images,*  
*ECCV 2010.*
Generating Simple Descriptions for images

Automatically generated description:

“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.”
General knowledge

Computer Vision
Natural Language Processing
Features & Representations
Clustering & EM
Discriminative Models & Classification
Generative & Topic Models
Summary

Enormous amounts of data.

Lots of commercial and academic applications.

We should combine information from words & pictures intelligently.
Your responsibilities

Homework – 3 relatively simple homeworks.

Paper presentations – each student will present 1 paper in class.

Paper summaries – on each paper presentation day turn in 1 paragraph summary of 1 of the assigned papers.

Project – final project including in class updates and final write-up.
Grading

Grading will consist of: Assignments (30%), Project (40%), Paper presentation (10%), Paper summaries (10%), Participation (10%).

You will be allowed 5 free homework/project late days of your choice over the semester. After those are used late assignments/projects will be accepted with a 10% reduction in value per day late.
Class Info

CSE 595: Words & Pictures

Instructor: Tamara Berg  (tlberg@cs.sunysb.edu)
Office: 1411 Computer Science
Lectures: Tues/Thurs 11:20-12:40pm Rm 2129 CS
Office Hours: Tues/Thurs 3:40-5:10pm

Course Webpage:
http://tamaraberg.com/teaching/Spring_11/wordspics