Admin

- Schedule for the rest of the quarter
  - Wed. class
  - 11/29 status report #2
  - 12/1 Exam #2 (a week from Wednesday?)
  - After exam: focus will be your projects

- No more TA office hours (unless needed?)

Tons of other NLP problems

- Information extraction
  - find entities in text
  - find how they relate to each other

- Information retrieval (search)

- Speech recognition

- Speech generation

- Dialogue/conversational agents
  - http://www.chatbots.org.uk
  - http://www.questioncreation.com/qChat/

- Summarization

- Text simplification/compression
Tons of other NLP problems

- Language recognition (play with Google translate)
- Word sense disambiguation
- Text understanding
- Question answering
- Text coherence
- Coreference resolution
- Text segmentation
- Document classification
- Document clustering
- ...

Some resources

- Books
  - Speech and Language Processing. 2009. Jurafsky and Martin.
- Software
  - Stanford has a good tool collection: http://nlp.stanford.edu/software/index.shtml
  - Berkeley has a few: http://nlp.cs.berkeley.edu/Main.html#Software
  - Natural Language Toolkit (NLTK)
  - python
  - http://www.nltk.org/

Every picture tells a story...

What is going on in this picture? How did you figure it out?

Computer Vision

- the goal of computer vision is to write computer programs that can interpret images (and videos)
- What are some of the challenges?
- Applications?
Long-term work on digits, AT&T labs

Optical character recognition
License plate readers and ways to get around them!

Sports
What are some of the problems that these two systems need to handle?

- The system has to know the orientation of the field with respect to the camera so that its perspective can be seen.
- The system has to know that some perspective change occurs as the camera moves.
- Given that the perspective can change, the system has to be able to reassemble the perspective at a rate of 30 frames per second as the camera moves.
- A football field is not flat; it creates very little in the middle to help the referee run on. So the field is described by the system.
- The system has to adapt for the field easily.
- The system also has to be aware of superimposed graphics that the network might overlay on the scene.

Challenges
- Orientation of the field with respect to the camera
- High-resolution encoders on the cameras
- Detailed field model
- Color palettes: in and out
- Slight delay in the network feed
- SportsVision
- Sportvision
- Dedicated Vision
- Dedicated Vision
- Dedicated Vision
Medical imaging

3D imaging: MRI, CT

Image guided surgery: Eric Grimson @ MIT

Face detection

Recognition is more difficult, but products are pushing that way!

http://www.youtube.com/watch?v=N1WC_00L0b0

Face Recognition

Smile detection?
Fingerprint scanners can be vision-based devices.

Face recognition systems now beginning to appear more widely.
www.sensiblevision.com

Security

Entertainment

Motion capture

ESC Entertainment, NBC, ABC

Shape capture

ESC Entertainment, NBC, ABC

MobilEye vision systems currently in high-end BMW, GM, Volvo model: ~70% of car manufacturers use cameras for safety
courtesy of Amnon Shashua

Safety

LaneHawk
simpler recognition: products passing by...

"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..." ~ Evolution Robotics, Pasadena.
Games and vision-based interaction

- WiiMotes: infrared images
- XBOX 360 Kinect

Digimask: put your face on a 3D avatar.

Camera tracking for crowd interactions...

Exploration in hostile environments (OK, robots!)

NASA's MER "Spirit" captured this westward view from atop a low plateau where Spirit spent the closing months of 2007.

Vision systems (JPL) used for several tasks
- Panorama stitching
- 3D terrain modeling
- Obstacle detection, position tracking
- Slip detection on uncertain terrain

Object recognition with mobile phones

Lincoln

Microsoft research

also: Point & Find, Nokia

Object recognition with mobile phones

Lincoln

Microsoft research

pretty much sums up the state-of-the-art!
How is an image represented?

• Images are made up of pixels
• For a color image, each pixel corresponds to an RGB value (i.e. three numbers)

Image file formats

- Bitmap
- JPEG
- TIFF
- Gif
- Png
- ...

Bitmap
Quantizer: Weights the various spectral coefficients according to their importance, with respect to the human visual system.
JPEG Compression

Object Recognition

What are these?

Different kinds of object recognition?

Object Recognition

Do you recognize these people?

11/22/10
Identification: is that Potala Palace?

Detection: are there people (or faces)?

Object and scene categorization

Verification: is that a lamp?
Recognition Question(s)

Identification: Where is this particular object?
Detection: Locate all instances of a given class
Content-based image retrieval: Find something similar
Categorization: What kind of object(s) is(are) present?
Verification: Is this what I think it is?

How might you arrange these, in order of difficulty?

[Csurka et al. 2006]

Recognition Questions

More accessible
Verification: Is this what I think it is?
Identification: Where is this particular object?
Content-based image retrieval: Find something similar
Detection: Locate all instances of a given class
Categorization: What kind of object(s) is(are) present?

More challenging
Certainly arguable!

Today: face recognition

Face recognition?

Verification: Is this what I think it is?
Identification: Where is this particular object?
Content-based image retrieval: Find something similar
Detection: Locate all instances of a given class
Categorization: What kind of object(s) is(are) present?
Face recognition?

- Verification: Is this what I think it is?
- Identification: Where is this particular object?
- Content-based image retrieval: Find something similar
- Detection: Locate all instances of a given class
- Categorization: What kind of object(s) is(are) present?

Eigenfaces: how do people do it?

The "Margaret Thatcher Illusion", by Peter Thompson
Matthew Turk and Alex Pentland
J. Cognitive Neuroscience, 1991

Eigenfaces for recognition

Image features

- We'd like to represent an image as vector of features
- good for machine learning techniques
- distance/similarity measures
- etc.
- What are possible features?
Color

How can we represent color?

Which is more similar?

L*a*b*

L – lightness (white to black)
a – red-greeness
b – yellowness-blueness

L*a*b* was designed to be uniform in that perceptual “closeness” corresponds to Euclidean distance in the space.

Texture

How is texture different than color?

Is color useful for face detection/verification?
Texture

- Texture is not pointwise like color
- Texture involves a local neighborhood

How can we capture texture?

Local “response” to feature functions

- A “feature” is a particular low-resolution image (intensities)
- “correlation”
- matrix dot product
- image portions with similar intensities will have high values
- Lots of possible features!

Example: Gabor Filters

Gabor filters are Gaussians modulated by sinusoids
They can be tuned in both the scale (size) and the orientation

Scale: 3 at 72°  Scale: 4 at 108°  Scale: 5 at 144°

Gabor filters

What would the response look like to a vertical filter?
Given a face, we can then calculate its response to a number of these filters generating a feature vector.

First-thoughts for detection?
First-thoughts for identification?

~10,000 dimensional space

Idea: faces have distinctive appearance. There is some intra-class variation. But nowhere near the inter-class variation with “everything else.”

Only a few dimensions needed
Only a few dimensions needed

Learning a projection

- We saw data projection when we were looking at machine learning techniques... where?

- How did we figure out the projection for clustering?

Dimensionality reduction

- How can we find the data's natural coordinate system?

Principal component analysis

- Suppose each data point is N-dimensional
  - What directions maximize variance?
    \[ \text{var}(v) = \sum_i (x_i - \bar{x})^T v \]
    \[ = v^T A v \quad \text{where} \quad A = \sum_i (x_i - \bar{x})(x_i - \bar{x})^T \]

- Solution: the eigenvectors of the variance matrix \( A \)
  - Eigenvector with largest eigenvalue captures the most variation among training vectors \( x \)
  - Eigenvector with smallest eigenvalue has least variation
  - We can use only the top few eigenvectors
  - Corresponds to choosing a "linear subspace"
  - Represent points on a line, plane, or "hyper-plane"
  - These eigenvectors are known as the **principal components**
Eigenfaces: pictures!

What do each of these mean?

Eigenfaces (plus the average face)

Projecting onto low-d eigenspace

- The eigenfaces $v_1, \ldots, v_K$ span the space of faces
  - A face is converted to eigenface coordinates by
  $$x \rightarrow [(x - x) \cdot v_1, (x - x) \cdot v_2, \ldots, (x - x) \cdot v_K]$$
  $$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_K \end{bmatrix}$$
  $$x \approx x + a_1 v_1 + a_2 v_2 + \ldots + a_K v_K$$

Eigenfaces (without the average face)

Progressive reconstructions...

How many dimensions?

- How many eigenfaces to use?
  - Look at the decay of the eigenvalues
  - the eigenvalue tells you the amount of variance “in the direction” of that eigenface
  - ignore eigenfaces with low variance

The hope...
How many dimensions?

- Total variance captured vs. number of eigenfaces used
- \( K = 10 \) captures about 36% of the variance
- \( K = 25 \) captures about 56%

In practice

Eigenfaces: recognition (id)

Novel image on left; best-matching image on right

Different lighting conditions
Different facial expressions

On which set do you think eigenfaces will perform better?

32 test cases

9/16 for lighting changes … 23/26 for expression changes
Eigenfaces: detection

How can we do this using eigenfaces?

Eigenfaces: detection

Difficult to avoid false positives...
Top 4

Eigenfaces: detection

Difficult to avoid false positives...
Top 3

Eigenfaces: detection

Top few
Receiver-operating curve

What's wrong with detection?

What's missing?

What parts are important?

Reasonable once we have an image of a face (recognition)
Not so good at finding faces (detection)
What parts are important?

- eyes vs. eyebrows
- who are these two people?

What parts are important?

- Nixon
- Winona Ryder
- who are these two people?

Robust real-time face detection

- Paul A. Viola and Michael J. Jones
- *Intl. J. Computer Vision*
- 57(2), 137–154, 2004

Learn which “parts” are most important…

Image features

- “Rectangle/box filters”
  - Similar to Haar wavelets
- Differences between sums of pixels in adjacent rectangles
- Simple thresholding

$$h(x) = \begin{cases} +1 & \text{if } f(x) > 0, \\ -1 & \text{otherwise} \end{cases}$$

24x24

- each box filter is present or absent
- gray regions are subtracted (after summing)
- white regions are added (after summing)
Huge library of filters

Constructing the classifier

- **For each round of boosting**: *(AdaBoost)*
  - Evaluate each rectangle filter on each example
  - Sort examples by filter values
  - Select best threshold for each filter (min error)
    - Use sorting to quickly scan for optimal threshold
  - Select best filter/threshold combination
  - Reweight examples
    - (There are many tricks to make this more efficient.)

Characteristics of algorithm

- Feature set (…is huge about 16M features)
- Efficient feature selection using AdaBoost
- New image representation
- Cascaded Classifier combining simple weak classifiers for rapid detection
- Fastest known face detector for gray scale images

Viola and Jones: Results
First two filters

- First classifier:
  - 2 features
  - 100% detection
  - 40% false detection

- The whole cascade:
  - 38 stages
  - 6000 features in total
  - On dataset with 807 faces and 78 millions sub-windows, faces are detected using 10 feature evaluations on average.
  - On average, 10 feature evals/sub-window

Receiver - operating curve for best 200 features shows true positive vs. false positive rate

Summary (Viola-Jones)

- Fastest known face detector for gray images
- Three contributions with broad applicability:
  - Cascaded classifier yields rapid classification
  - AdaBoost as an extremely efficient feature selector
  - Rectangle Features + Integral Image can be used for rapid image analysis

But, there are better ones out there...
Happy face!