Lecture 14: Introduction to Recognition

- mountain
- tree
- banner
- building
- street lamp
- vendor
- people
Announcements

• Project 2 due Sunday at 11:59pm
  – Voting for artifacts will begin soon after
What do we mean by “object recognition”?

Next 15 slides adapted from Li, Fergus, & Torralba’s excellent short course on category and object recognition.
Verification: is that a lamp?
Detection: are there people?
Identification: is that Potala Palace?
Object categorization

- mountain
- tree
- building
- banner
- street lamp
- vendor
- people
Scene and context categorization

- outdoor
- city
- ...
Object recognition
Is it really so hard?

Find the chair in this image

Output of normalized correlation

This is a chair
Object recognition
Is it really so hard?

Find the chair in this image

Pretty much garbage
Simple template matching is not going to make it
Object recognition
Is it really so hard?

Find the chair in this image

A “popular method is that of template matching, by point to point correlation of a model pattern with the image pattern. These techniques are inadequate for three-dimensional scene analysis for many reasons, such as occlusion, changes in viewing angle, and articulation of parts.” Nivatia & Binford, 1977.
Why not use SIFT matching for everything?

• Works well for object *instances*

• Not great for generic object *categories*
Applications: Computational photography

[Face priority AE] When a bright part of the face is too bright
Applications: Assisted driving

Pedestrian and car detection

Lane detection

- Collision warning systems with adaptive cruise control,
- Lane departure warning systems,
- Rear object detection systems,
Applications: image search

Refine your image search with visual similarity

Similar Images allows you to search for images using pictures rather than words. Click the "Similar images" link under an image to find other images that look like it. Try a search of your own or click on an example below.

**places**
- London
- New York
- Egypt
- Forbidden City

**celebrities**
- Michael Jordan
- Angelina Jolie
- Halle Berry
- Seth Rogan
- Rihanna

**art**
- impressionism
- Keith Haring
- cubism
- Salvador Dali
- pointillism

**shopping**
- evening gown
- necklace
- shoes
How do human do recognition?

• We don’t completely know yet
• But we have some experimental observations.
Observation 1

- We can recognize familiar faces even in low-resolution images
Observation 2:

- High frequency information is not enough

Jim Carrey  Kevin Costner
What is the single most important facial features for recognition?
Observation 4:

- Image Warping is OK
The list goes on

Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About

Let’s start simple

• Today
  – skin detection
  – eigenfaces
Face detection

- Do these images contain faces? Where?
One simple method: skin detection

Skin pixels have a distinctive range of colors
- Corresponds to region(s) in RGB color space
  - for visualization, only R and G components are shown above

Skin classifier
- A pixel $X = (R,G,B)$ is skin if it is in the skin region
- But how to find this region?
Learn the skin region from examples
- Manually label pixels in one or more “training images” as skin or not skin
- Plot the training data in RGB space
  - skin pixels shown in orange, non-skin pixels shown in blue
  - some skin pixels may be outside the region, non-skin pixels inside. Why?

Skin classifier
- Given $X = (R, G, B)$: how to determine if it is skin or not?
Skin classification techniques

Skin classifier
- Given $X = (R,G,B)$: how to determine if it is skin or not?
- Nearest neighbor
  - find labeled pixel closest to $X$
  - choose the label for that pixel
- Data modeling
  - fit a model (curve, surface, or volume) to each class
- Probabilistic data modeling
  - fit a probability model to each class
Basic probability

- **X** is a random variable
- **P(X)** is the probability that **X** achieves a certain value

Conditional probability:  
- \( P(X | Y) \) — probability of **X** given that we already know **Y**

\[
0 \leq P(X) \leq 1
\]

\[
\int_{-\infty}^{\infty} P(X) \, dX = 1 \quad \text{or} \quad \sum P(X) = 1
\]

- continuous **X**
- discrete **X**

- called a PDF
  - probability distribution/density function
  - a 2D PDF is a surface, 3D PDF is a volume
Probabilistic skin classification

Now we can model uncertainty

- Each pixel has a probability of being skin or not skin
  - $P(\sim \text{skin}|R) = 1 - P(\text{skin}|R)$

Skin classifier

- Given $X = (R,G,B)$: how to determine if it is skin or not?
- Choose interpretation of highest probability
  - set $X$ to be a skin pixel if and only if $R_1 < X \leq R_2$

Where do we get $P(\text{skin}|R)$ and $P(\sim \text{skin}|R)$?
Learning conditional PDF’s

We can calculate \( P(R \mid \text{skin}) \) from a set of training images

- It is simply a histogram over the pixels in the training images
  - each bin \( R_i \) contains the proportion of skin pixels with color \( R_i \)

This doesn’t work as well in higher-dimensional spaces. Why not?

Approach: fit parametric PDF functions

- common choice is rotated Gaussian
  - center \( c = \overline{X} \)
  - covariance \( \sum_{X} (X - \overline{X})(X - \overline{X})^T \)

  » orientation, size defined by eigenvcs, eigenvals
Learning conditional PDF’s

\[ P(R|\text{skin}) = \frac{\#\text{skin pixels with color } R}{\#\text{skin pixels}} \]

We can calculate \( P(R \mid \text{skin}) \) from a set of training images

- It is simply a histogram over the pixels in the training images
  - each bin \( R_i \) contains the proportion of skin pixels with color \( R_i \)

But this isn’t quite what we want

- Why not? How to determine if a pixel is skin?
- We want \( P(\text{skin} \mid R) \), not \( P(R \mid \text{skin}) \)
- How can we get it?
Bayes rule

\[ P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)} \]

In terms of our problem:

what we measure (likelihood) \quad domain knowledge (prior)

\[ P(\text{skin}|R) = \frac{P(R|\text{skin})P(\text{skin})}{P(R)} \]

what we want (posterior) \quad normalization term

\[ P(R) = P(R|\text{skin})P(\text{skin}) + P(R|\sim\text{skin})P(\sim\text{skin}) \]

The prior: \( P(\text{skin}) \)

- Could use domain knowledge
  - \( P(\text{skin}) \) may be larger if we know the image contains a person
  - for a portrait, \( P(\text{skin}) \) may be higher for pixels in the center
- Could learn the prior from the training set. How?
  - \( P(\text{skin}) \) could be the proportion of skin pixels in training set
Bayesian estimation

- Goal is to choose the label (skin or ~skin) that maximizes the posterior
  - this is called **Maximum A Posteriori (MAP) estimation**
- Suppose the prior is uniform: \( P(\text{skin}) = P(\sim \text{skin}) = 0.5 \)
  - in this case \( P(\text{skin}|R) = cP(R|\text{skin}) \), \( P(\sim \text{skin}|R) = cP(R|\sim \text{skin}) \)
  - maximizing the posterior is equivalent to maximizing the likelihood
    - \( P(\text{skin}|R) > P(\sim \text{skin}|R) \) if and only if \( P(R|\text{skin}) > P(R|\sim \text{skin}) \)
  - this is called **Maximum Likelihood (ML) estimation**
Figure 25.3. The figure shows a variety of images together with the output of the skin detector of Jones and Rehg applied to the image. Pixels marked black are skin pixels, and white are background. Notice that this process is relatively effective, and could certainly be used to focus attention on, say, faces and hands. Figure from “Statistical color models with application to skin detection,” M.J. Jones and J. Rehg, Proc. Computer Vision and Pattern Recognition, 1999 © 1999, IEEE.
General classification

This same procedure applies in more general circumstances

- More than two classes
- More than one dimension

Example: face detection

- Here, X is an image region
  - dimension = # pixels
  - each face can be thought of as a point in a high dimensional space

http://www-2.cs.cmu.edu/afs/cs.cmu.edu/user/hws/www/CVPR00.pdf
Linear subspaces

Classification can be expensive

- Must either search (e.g., nearest neighbors) or store large PDF’s

Suppose the data points are arranged as above

- Idea—fit a line, classifier measures distance to line

convert $x$ into $v_1$, $v_2$ coordinates

$$x \rightarrow ((x - \bar{x}) \cdot v_1, (x - \bar{x}) \cdot v_2)$$

What does the $v_2$ coordinate measure?
- distance to line
- use it for classification—near 0 for orange pts

What does the $v_1$ coordinate measure?
- position along line
- use it to specify which orange point it is
Dimensionality reduction

- We can represent the orange points with *only* their $v_1$ coordinates
  - since $v_2$ coordinates are all essentially 0
- This makes it much cheaper to store and compare points
- A bigger deal for higher dimensional problems
Linear subspaces

Consider the variation along direction $v$ among all of the orange points:

$$\text{var}(v) = \sum_{x} \|(x - \bar{x})^T \cdot v\|^2$$

What unit vector $v$ minimizes $\text{var}$?

$$v_2 = \min_v \{\text{var}(v)\}$$

What unit vector $v$ maximizes $\text{var}$?

$$v_1 = \max_v \{\text{var}(v)\}$$

Solution: $v_1$ is eigenvector of $A$ with largest eigenvalue

$v_2$ is eigenvector of $A$ with smallest eigenvalue
Suppose each data point is N-dimensional

- Same procedure applies:
  \[
  \text{var}(v) = \sum_x \| (x - \bar{x})^T \cdot v \|
  = v^T A v \quad \text{where} \quad A = \sum_x (x - \bar{x})(x - \bar{x})^T
  \]

- The eigenvectors of A define a new coordinate system
  - eigenvector with largest eigenvalue captures the most variation among training vectors x
  - eigenvector with smallest eigenvalue has least variation

- We can compress the data by only using 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
The space of faces

An image is a point in a high dimensional space

- An $N \times M$ intensity image is a point in $\mathbb{R}^{NM}$
- We can define vectors in this space as we did in the 2D case
Dimensionality reduction

The set of faces is a “subspace” of the set of images

• Suppose it is $K$ dimensional
• We can find the best subspace using PCA
• This is like fitting a “hyper-plane” to the set of faces
  – spanned by vectors $v_1, v_2, \ldots, v_K$
  – any face $x \approx \bar{x} + a_1 v_1 + a_2 v_2 + \ldots + a_k v_k$