Recognition

Introduction to Computer Vision

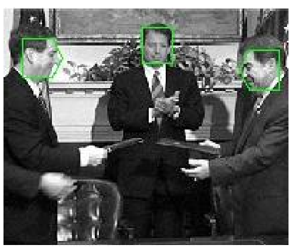
CSE 152

Lecture 14

Announcements

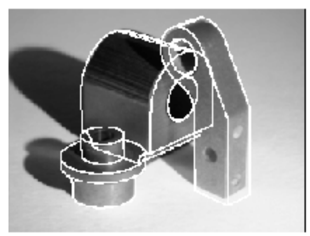
- Homework 4 is due May 24, 11:59 PM
- Reading:
  - Chapter 15: Learning to Classify
  - Chapter 16: Classifying Images
  - Chapter 17: Detecting Objects in Images

Recognition



Given a database of objects and an image determine what, if any of the objects are present in the image.

Recognition

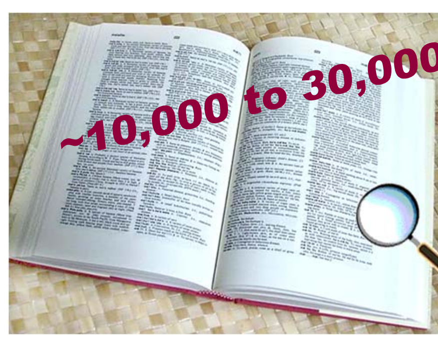


Given a database of objects and an image determine what, if any of the objects are present in the image.

Recognition

Given a database of objects and an image determine what, if any of the objects are present in the image.

How many visual object categories are there?



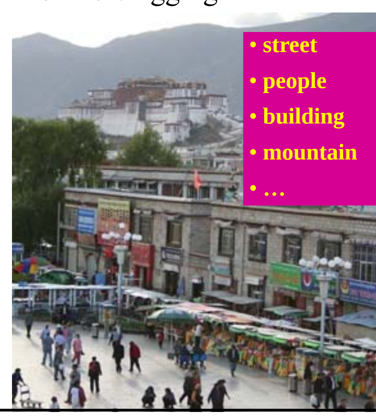
Biederman 1987

Specific recognition tasks

Scene categorization

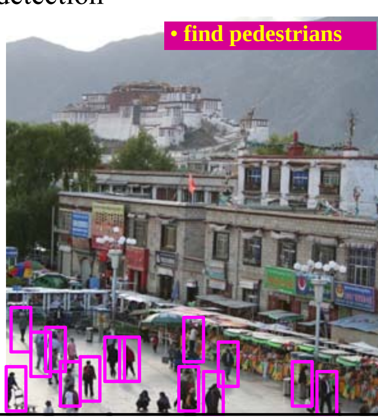
- outdoor/indoor
- city/forest/factory/etc.

Image annotation/tagging



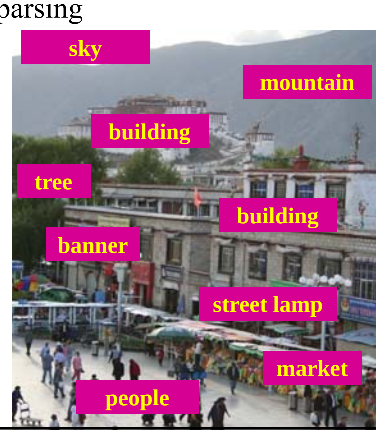
- street
- people
- building
- mountain
- …

Object detection



- find pedestrians

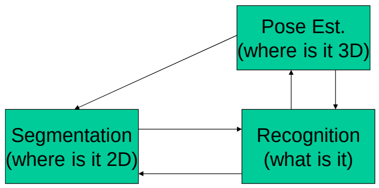
Image parsing



Object Recognition: The Problem

Given: A database D of “known” objects and an image I:

1. Determine which (if any) objects in D appear in I
2. Determine the pose (rotation and translation) of the object

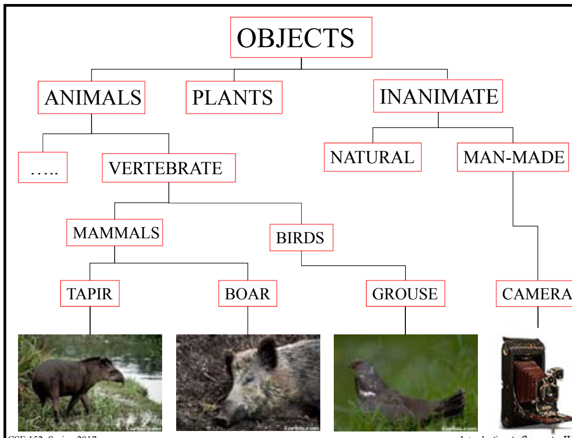


WHAT AND WHERE!!!

Within-class variations

Recognition Challenges

- Within-class variability
  - Different objects within the class have different shapes or different material characteristics
  - Deformable
  - Articulated
  - Compositional
- Pose variability:
  - 2-D Image transformation (translation, rotation, scale)
  - 3-D Pose Variability (perspective, orthographic projection)
- Lighting
  - Direction (multiple sources & type)
  - Color
  - Shadows
- Occlusion – partial
- Clutter in background -> false positives



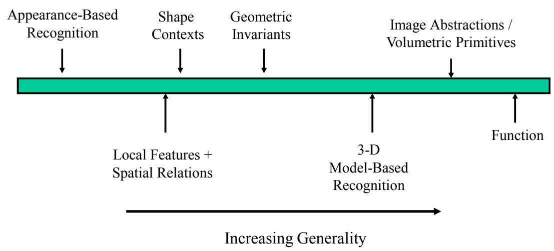
Object Categories (Classes)

- Categories near top of tree (e.g., vehicles) – lots of within class variability
- Fine grain categories (e.g., species of birds) -- Moderate within class variation
- Instance recognition (e.g., person identification) – within class mostly shape articulation, bending, etc.

Pattern Classification

- Supervised vs. Unsupervised: Do we have labels?
- Supervised
  - Nearest Neighbor
  - Bayesian
    - Plug in classifier
    - Distribution-based
    - Projection methods
  - Neural Network
  - Support Vector Machine
  - Kernel methods
- Unsupervised
  - Clustering
  - Reinforcement learning

A Rough Recognition Spectrum



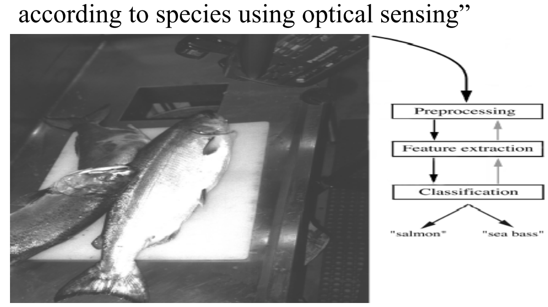
Appearance-Based Recognition

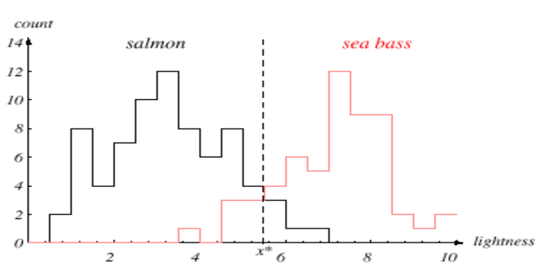
Appearance-Based Vision for Instances Level Recognition

- A Pattern Classification Viewpoint
  1. Bayesian Classification
  2. Appearance Manifolds
  3. Feature Space
  4. Dimensionality Reduction

Bayesian Classification

- Example: "Sorting incoming Fish on a conveyor according to species using optical sensing"

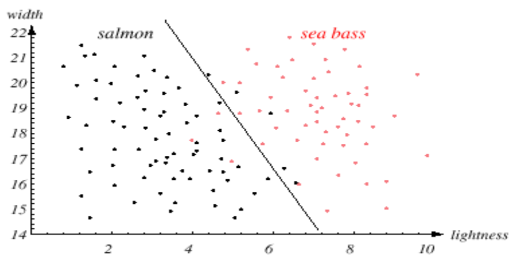


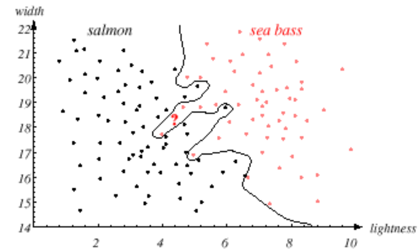


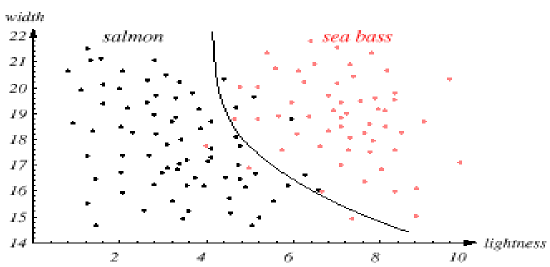
- Adopt the lightness and add the width of the fish

Fish ⇒ $x^T = [x_1, x_2]$

Lightness Width







Basic ideas in classifiers

- Loss
  - Some errors may be more expensive than others
    - e.g., a fatal disease that is easily cured by a cheap medicine with no side-effects -> false positives in diagnosis are better than false negatives
  - We discuss two class classification: L(1->2) is the loss caused by calling 1 a 2
- Total risk of using classifier s

$$R(s) = Pr\left\{1 \to 2 | \text{using } s\right\} L(1 \to 2) + Pr\left\{2 \to 1 | \text{using } s\right\} L(2 \to 1)$$

Basic ideas in classifiers

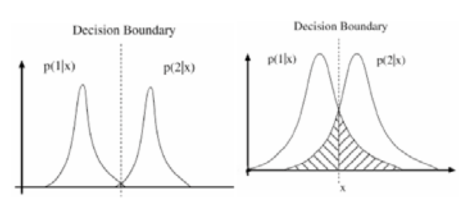
- Generally, we should classify as 1 if the expected loss of classifying as 1 is better than for 2 gives

1 if $p(1|\boldsymbol{x})L(1 \to 2) > p(2|\boldsymbol{x})L(2 \to 1)$

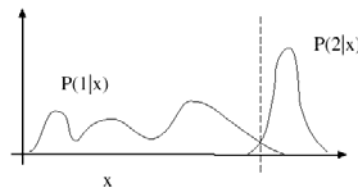
2 if $p(1|\boldsymbol{x})L(1 \to 2) < p(2|\boldsymbol{x})L(2 \to 1)$

- Crucial notion: Decision boundary
  - Points where the loss is the same for either case

Some loss may be inevitable: the minimum risk (shaded area) is called the Bayes risk



Finding a decision boundary is not the same as modeling a conditional density



- Classifier boils down to: choose class that minimizes:

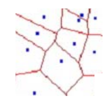
$$\delta\left(\underline{x},\underline{\mu}_k\right)^2 - 2\log \pi_k$$

where

Mahalanobis distance

$$\delta\left(\underline{x},\underline{\mu}_k\right) = \left[\left(\underline{x}-\underline{\mu}_k\right)^T \underline{\underline{\Sigma}}^{-1}\left(\underline{x}-\underline{\mu}_k\right)\right]^{1/2}$$

Because covariance is common, this simplifies to sign of a linear expression (i.e., Voronoi diagram in 2D for Σ=I)



Plug-in classifiers

- Assume that class conditional distributions P(x|ωi) have some parametric form - now estimate the parameters from the data
- Common:
  - Assume a normal distribution with shared covariance, different means; use usual estimates
  - Normal distribution but with different covariances
- Issue: parameter estimates that are "good" may not give optimal classifiers

Example: Finding skin

- Skin has a very small range of (intensity independent) colors and little texture
  - Compute an intensity-independent color measure, check if color is in this range, check if there is little texture (median filter)
  - See this as a classifier
    - We can set up the tests by hand or learn them
  - Get class conditional densities (histograms) and priors from data (counting)
- Classifier is
  - if $p(\text{skin}|\boldsymbol{x}) > \theta$, classify as skin
  - if $p(\text{skin}|\boldsymbol{x}) < \theta$, classify as not skin
  - if $p(\text{skin}|\boldsymbol{x}) = \theta$, choose classes uniformly and at random

Figure from "Statistical color models with application to skin detection," M.J. Jones and J. Rehg, Proc. Computer Vision and Pattern Recognition, 1999 copyright 1999, IEEE

Receiver
Operating
Curve

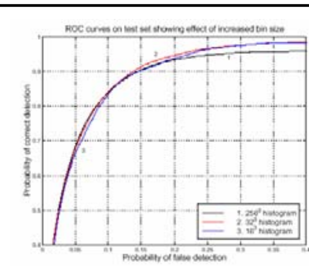
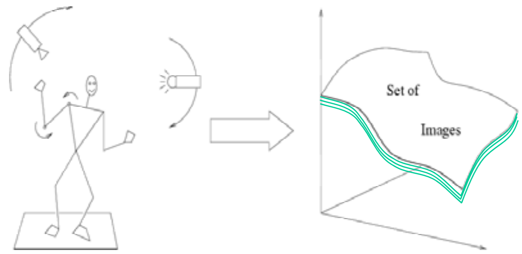


Figure from "Statistical color models with application to skin detection," M.J. Jones and J. Rehg, Proc. Computer Vision and Pattern Recognition, 1999 copyright 1999, IEEE



Variability: Camera position
Illumination
Internal parameters

Within-class variations

Appearance manifold approach

- For every object
  1. Sample the set of viewing conditions
  2. Crop & scale images to standard size
  3. Use as feature vector
- Apply principal component analysis (PCA) over all the images
- Keep the dominant principal components
- Set of views for one object is represented as a manifold in the projected space
- Recognition: What is nearest manifold for a given test image?

(Nayar et al. '96)

Limitations of these approaches

- Object must be segmented from background (How would one do this in non-trivial situations?)
- Occlusion?
- The variability (dimension) in images is large (Is sampling feasible?)
- How can one generalize to classes of objects?

Appearance-Based Vision: Lessons

Strengths

- Posing the recognition metric in the image space rather than a derived representation is more powerful than expected.
- Modeling objects from many images is not unreasonable given hardware developments.
- The data (images) may provide a better representations than abstractions for many tasks.

Appearance-Based Vision: Lessons

Weaknesses

- Segmentation or object detection is still an issue.
- To train the method, objects have to be observed under a wide range of conditions (e.g. pose, lighting, shape deformation).
- Limited power to extrapolate or generalize (abstract) to novel conditions.

Bag-of-features models

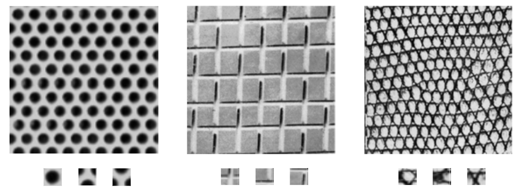
Object → **Bag of 'words'**

Slides from Svetlana Lazebnik who borrowed from others

Bag-of-features models

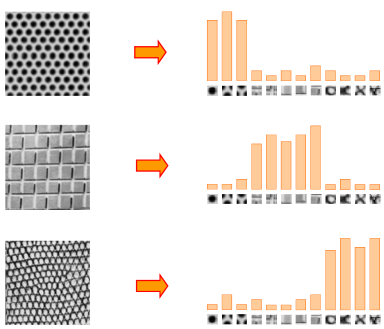
Origin 1: Texture recognition

- Texture is characterized by the repetition of basic elements or *textons*
- For stochastic textures, it is the identity of the textons, not their spatial arrangement, that matters



Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001; Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

Origin 1: Texture recognition



Julesz, 1981; Cula & Dana, 2001; Leung & Malik 2001; Mori, Belongie & Malik, 2001; Schmid 2001; Varma & Zisserman, 2002, 2003; Lazebnik, Schmid & Ponce, 2003

Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

Which US President?
Franklin D. Roosevelt, John F. Kennedy, George W. Bush

Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

Which US President?
Franklin D. Roosevelt, John F. Kennedy, George W. Bush

Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

Which US President?
Franklin D. Roosevelt, John F. Kennedy, George W. Bush

Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

Which US President?
Franklin D. Roosevelt, John F. Kennedy, George W. Bush

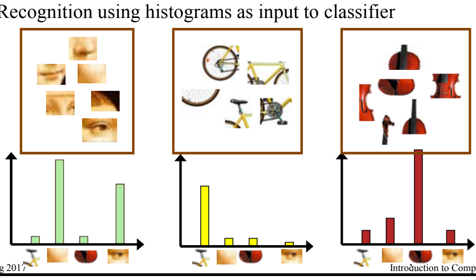
Origin 2: Bag-of-words models

- Orderless document representation: frequencies of words from a dictionary Salton & McGill (1983)

Which US President?
Franklin D. Roosevelt, John F. Kennedy, George W. Bush

Bag-of-features steps

1. Extract features
2. Learn “visual vocabulary”
3. Quantize features using visual vocabulary
4. Represent images by frequencies (histogram) of “visual words”
5. Recognition using histograms as input to classifier

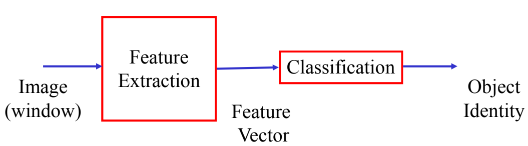


Feature extraction

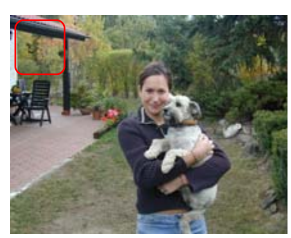
- Regular grid or interest regions

Feature Space

- Sketch of a Pattern Recognition Architecture



Sliding window approaches



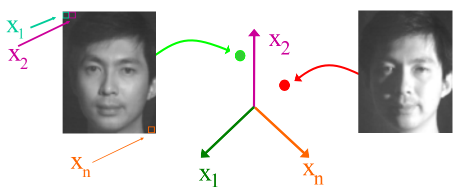
Example: Face Detection

- Scan window over image
- Search over position & scale
- Classify window as either:
  - Face
  - Non-face

Feature Space

- **What are the features?**
- What is the classifier?

The Space of Images



- We will treat an d-pixel image as a point in an d-dimensional space, $\mathbf{x} \in \mathbf{R}^d$.
- Each pixel value is a coordinate of $\mathbf{x}$.

More features

- Filtered image
- Filter with multiple filters (bank of filters)
- Histogram of colors
- Histogram of Gradients (HOG)
- Haar wavelets
- Scale Invariant Feature Transform (SIFT)
- Speeded Up Robust Feature (SURF)

Feature Space

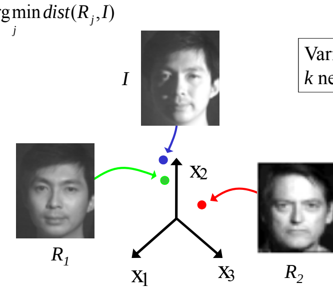
- What are the features?
- **What is the classifier?**

Nearest Neighbor Classifier

{ R_j } are set of training images.

$$ID = \arg\min_j dist(R_j, I)$$

Variation of this:
k nearest neighbors



Comments on Nearest Neighbor

- Sometimes called "Template Matching"
- Variations on distance function (e.g., L_1, robust distances)
- Multiple templates per class - perhaps many training images per class
- Expensive to compute *k* distances, especially when each image is big (*d*-dimensional)
- May not generalize well to unseen examples of class
- No worse than twice the error rate of the optimal classifier (if enough training samples)
- Some solutions:
  - Bayesian classification
  - Dimensionality reduction

Next Lectures

- Recognition, detection, and classification
- Reading:
  - Chapter 15: Learning to Classify
  - Chapter 16: Classifying Images
  - Chapter 17: Detecting Objects in Images