Appearance-Based Recognition

Computer Vision I

CSE252A

Lecture 19

Announcements

- HW 4 is on the web site : stereo, due on Friday.
- HW2 Returned
- Today: From Pattern Recognition to Appearance-based vision

Tracking Modalities

- Color
  - Histogram [Birchfield 1998; Bradski 1998]
  - Volume [Wren *et al*., 1995; Bregler, 1997; Darrell, 1998]
- Shape
  - Deformable curve [Kass *et al*.1988]
  - Template[Blake *etal*. 1993; Birchfield 1998]
  - Example-based [Cootes *et al*., 1993; Baumberg & Hogg, 1994]
- Appearance
  - Correlation [Lucas & Kanade, 1981; Shi & Tomasi, 1994]
  - Photometric variation [Hager & Belhumeur, 1998]
  - Outliers [Black *et al*., 1998; Hager & Belhumeur, 1998]
  - Nonrigidity [Black *et al*., 1998; Sclaroff & Isidoro, 1998]
- Motion
  - Background model [Wren *et al*., 1995; Rosales & Sclaroff, 1999; Stauffer & Grimson, 1999]
  - Optical flow [Cutler & Turk]
  - Egomotion [Sawhney & Ayer, 1996; Irani & Anandan, 1998]
- Stereo
  - Blob correlation [Azarbayejani & Pentland, 1996]
  - Disparity map [Kanade *et al*., 1996; Konolige, 1997; Darrell *et al*., 1998]

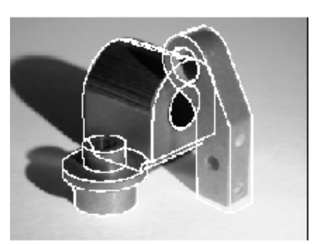
Color Blob tracking

- **Color-based tracker gets lost on white knight: Same Color**

Recognition

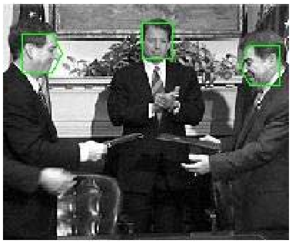
Given a database of objects and an image determine what, if any of the objects are present in the image.

Recognition



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

Recognition

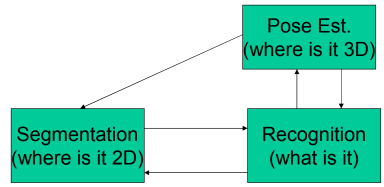


Given a database of objects and an image determine what, if any of the objects are present in the image.

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!!!

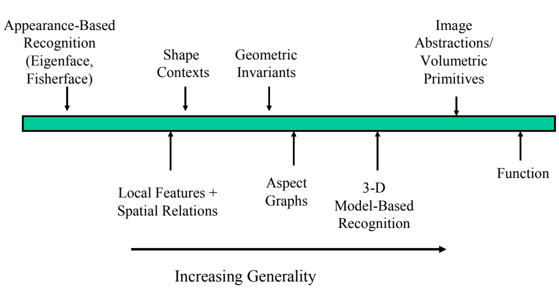
Recognition Challenges

- 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

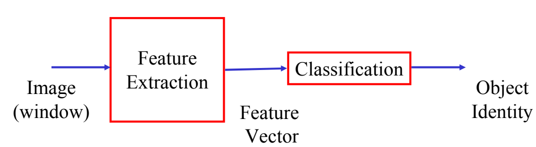
Object Recognition Issues:

- How general is the problem?
  - 2D vs. 3D
  - range of viewing conditions
  - available context
  - segmentation cues
- What sort of data is best suited to the problem?
  - Whole images
  - Local 2D features (color, texture,
  - 3D (range) data
- What information do we have in the database?
  - Collection of images?
  - 3-D models?
  - Learned representation?
  - Learned classifiers?
- How many objects are involved?
  - small: brute force search
  - large: ??

A Rough Recognition Spectrum



Sketch of a Pattern Recognition Architecture



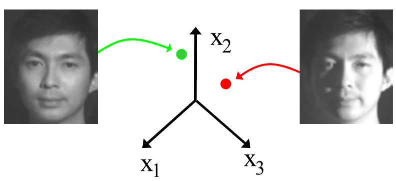
Example: Face Detection

- Scan window over image.
- Classify window as either:
  - Face
  - Non-face

Pattern Classification Summary

- Supervised vs. Unsupervised: Do we have labels?
- Supervised
  - Nearest Neighbor
  - Bayesian
    - Plug in classifier
    - Distribution-based
    - Projection Methods (Fisher's, LDA)
  - Neural Network
  - Support Vector Machine
  - Kernel methods
- Unsupervised
  - Clustering
  - Reinforcement learning

Image as a Feature Vector

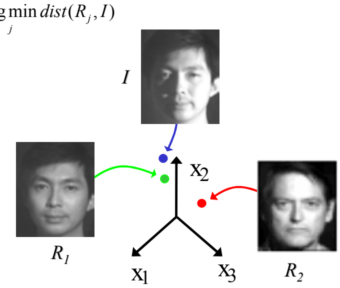


- Consider an n-pixel image to be a point in an n-dimensional space, $\mathbf{x} \in \mathbf{R}^n$.
- Each pixel value is a coordinate of **x**.

Nearest Neighbor Classifier

$\{ R_j \}$ are set of training images.

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



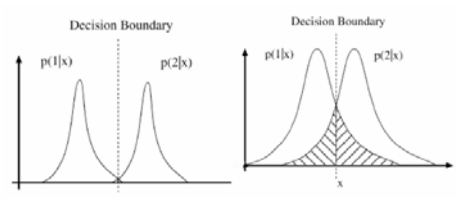
Comments

- 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 (N dimensional).
- May not generalize well to unseen examples of class.
- Some solutions:
  - Bayesian classification
  - Dimensionality reduction

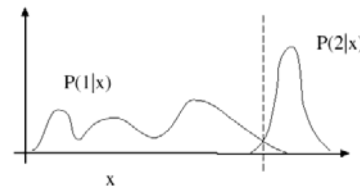
Bayesian Classification

Blackboard

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.

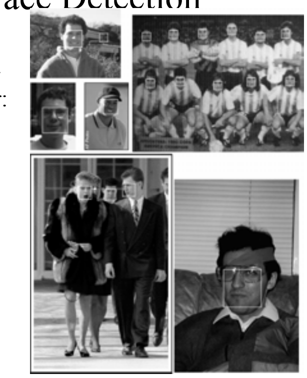


Plug-in classifiers

- Assume that distributions have some parametric form - now estimate the parameters from the data.
- Common:
  - assume a normal distribution with shared covariance, different means; use usual estimates
  - ditto, but different covariances; ditto
- Issue: parameter estimates that are “good” may not give optimal classifiers.

Example: Face Detection

- Scan window over image.
- Classify window as either:
  - Face
  - Non-face



Example: Finding skin
Non-parametric Representation of CCD

- 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), 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

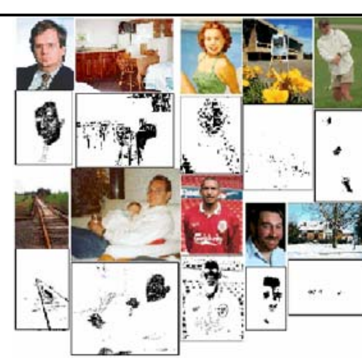


Figure from “Statistical color models with application to skin detection,” M.J. Jones and J. Rehg, Proc. Computer Vision and Pattern Recognition, 1999 copyright 1999, IEEE