Introduction

Reading
- Textbook DHS Chapter 1

References:
- DIP Gonzalez and Woods, Textbook, Chapter 12, Object Recognition
About the course

Objectives:
- Learn how to formulate and solve pattern recognition problems
  - Feature extraction, object/event recognition, structure detection, video search and retrieval
- Get insights and experience of recent machine learning techniques
  - Bayesian, SVM, Neural Network, HMM, feature selection, clustering
- Have fun in experimenting with actual visual classification/indexing problems
- Demos:
  - Google Yahoo Image search, Goggle News search, Columbia video search

Intended Audience
- Beginning graduate students or professionals
- familiar with signal/image processing
- comfortable with probability, statistics, linear algebra, and some machine learning concepts

Sample Applications:

Video Indexing and Analysis

- Indexing, search, and retrieval for images and videos
  - See Columbia’s WebSEEK and VideoQ
    http://www.ee.columbia.edu/{webseek, videoq}
  - Goggle image search
  - “find video clips of basketball going through the hoop”
  - “find images containing shape shown in the sketch”
- Automatic annotation of visual content
  - recognition of text, face, scene, vehicle, location, etc
- Automatic parsing of video programs into structures
  - break videos into shots, scenes, and stories
- Event detection
  - sports events, human activities, meetings, medical, and other spatio-temporal patterns
- Summary
  - topic clustering, highlight generation
  - See Columbia’s sports highlight, news topic clustering demo
What do we search in video? (examples)

- **TRECVID 2003**
  - “Find shots of an airplane taking off.”
- **TRECVID 2004**
  - “Find shots of Bill Clinton speaking with a US flag visible behind him.”
- **IBM Speech Group**
  - “Find shots containing monkeys or gorillas.”
- **BBC Logs**
  - “Find shots of the Kremlin.”

Related Activities NIST TRECVID

- Low-level feature detection (motion, shot etc)
- High-level feature detection
  - Image $\rightarrow$ classifier $\rightarrow$ {'people', 'vehicle', 'explosion', etc}
- Story boundary detection
- Search: fully automatic, manual, interactive
- 2005 Data
  - 6 channels in English, Chinese, Arabic
  - >170 hours, 126,000 subshots
  - 39 concepts manually annotated over >80 hours
    (LSCOM-Lite): very valuable resource for researchers!
- Participating groups
  - 62 groups (due: high-level feature 8/22, search 9/21/05)
Examples of object & structure recognition

How to detect and recognize the characters and words? (Demo)

How to detect the boundaries of programs, stories, and commercials?

A multi-modal pattern recognition problem

- Objective: a story boundary at time $\tau_k$?
  - $\tau_k = \{\text{shot boundaries or significant pauses}\}$  
  
  (Hsu and Chang)

$$q(b|x) = \frac{1}{Z(x)} \sum_e e^b(x,y), b \in \{0,1\}$$

EE6887-Chang
Image to Keyword Annotation

- Tag images with category labels
- Organize the photos into hierarchical categories

Unsupervised Video Structure Discovery: Hierarchical Hidden Markov Model

- Inference Multi-Level Markovian Temporal Dependence
  - High-level states represent distinct events
  - Presence of each event produces observations modeled by low-level HMMs

![A Very High-Level Stat. Pattern Recog. Architecture](image)

*Figure 1: Model for statistical pattern recognition.*

(From Jain, Duin, and Mao, SPR Review, ‘99)

---

**Important issues**

- Image/video pre-processing - quality, resolution etc
- Feature extraction - Color, texture, motion, shape, layout, regions, parts, etc
- Feature representation - Discrete vs. continuous, vectorization, dimension
- Invariance to scale, rotation, translation ...
- Feature selection - PCA, MDS, Kernel PCA, etc
- Classification models - Generative vs. discriminative
- Multi-modal fusion, early fusion vs. late fusion
- Size of training/test data and manual supervision efforts
- Validation and evaluation processes
- Complexity
Some examples of feature representation

- Features determine the patterns and their separability
- E.g.,
  - Angular distance for closed shapes
  - Part features for iris flowers

Another example of feature

- Bankers Asso. Font used on personal checks
- Use magnetic ink and reader to simplify segmentation
- Feature: the horizontal scan of the rate of increase/decrease of the character area
- Peaks and zeros are arranged to be located at the vertical grid lines → can be sampled accurately
- Patterns can be easily distinguished

Feature Representation
Extraction/Selection

(Jain et al 99)

PCA

Fischer Analysis

MDS

Kernel PCA

EE6887-Chang
Classification Paradigms

Class 1 vs Class 2

- **Discriminative**
  - Decision Boundary
  - \( f(x) > 0 \)
  - \( f(x) < 0 \)

- **Generative**
  - \( P(x|C=1) > P(x|C=2) \)
  - \( x_0 \)
  - \( C(x_0) = ? \)

Training / Validation / Testing

- Assume the same distribution in different set, otherwise the optimal solution from validation may not be optimal in test data
Training / Validation / Testing (cont.)

- Multiple validation sets can be used for different optimization steps.

- Cross validation, leave-one-out

  1 2 … K
  Training Testing

  Rotate the choice of the test set and average the performance over runs

Curse of Dimensionality and Overfitting

A case of overtraining

Rule of thumb -

\[ \frac{\text{(\# of training patterns per class)}}{\text{(\# of features)}} > 10 \]
Topics (tentative)

- Image/Video feature extraction
- Bayesian Decision Theory
- Parameter Estimation (ML and non-parametric)
- Linear Discriminant Functions
- Support Vector Machines
- Neural Networks
- Fusion, evaluation
- Feature Selection
- Clustering

Course Format

- Weekly Lectures
- Textbook
- Reference Books
  - Related research papers, distributed online
  - The Elements of Statistical Learning, Trevor Hastie, Robert Tibshirani & Jerome Friedman, 2001, Springer Verlag
- Grading
  - 6 homeworks (30%), midterm (25%), final (25%), mini-project (20%)
  - Homeworks include both analytical problems and programming assignments
  - Matlab is recommended for programming
Resources and Matlab

- Links on the class web site
  - Tutorials on paper writing, Matlab, etc
- Software links on web site to
  PR Tools, Spider (SVM), Netlab, HMM, etc
- EE6887 SPR Class Dataset
  - Benchmark data set, a few thousands of images from broadcast news and stock photos
  - Extracted features and labels
  - Will distribute for class project use only
- Matlab is recommended for programming
  - Accessible in Mudd 251 Computer Lab (Need CU ACIS account)
  - Student edition available

Next Lecture

- Image feature extraction
- Bayesian Decision and Bayesian Minimal Risk Classifiers
  Textbook DHS Chap. 2.1 – 2.6