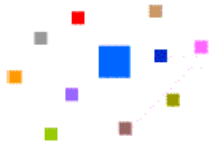


EE 6882 Statistical Methods for Video Indexing and Analysis

Fall 2003

Prof. Shih-Fu Chang
http://www.ee.columbia.edu/~sfchang

Lecture 1 (9/3/03)

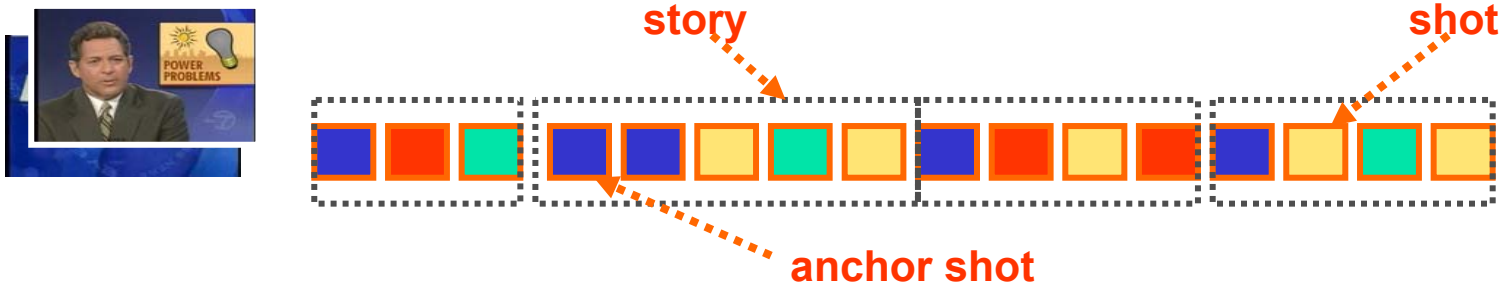
Research Problems in Video Indexing and Analysis

- Object detection and recognition
  (e.g., face, text, vehicles)
- Structure parsing
  (e.g., breaking videos into shots, scenes, and stories)
- Event detection
  (e.g., sports events, human activities, meetings, medical)
- Search and retrieval
  (e.g., interactive search with feedback)
- Synthesis
  (e.g., personal summaries, highlight generation)

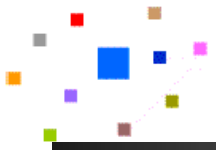
Object recognition and structure parsing

Text in video

Text with different styles



dvmm DIGITAL VIDEO • MULTIMEDIA LAB



Statistical Methods

- Emerging mature tools and promising performance
- Increasing computing resources
- More challenging, interesting problems
- Increasing benchmark data (e.g., NIST TREC Video)

Why this course?

- Learn insights of different tools and models
- Understand match between tools and problems in this field
- Get some experience on tools publicly available and from DVMM Lab
- Related hard-core courses, see web site

Papers to Study

- Problems
  - Image/video classification
  - Interactive image retrieval
  - Video structure parsing
  - Multimedia data mining
- Techniques
  - Bayesian, factor graph, graphical model
  - HMM and variations
  - SVM
  - Hierarchical Mixture
  - others

SPR System Architecture

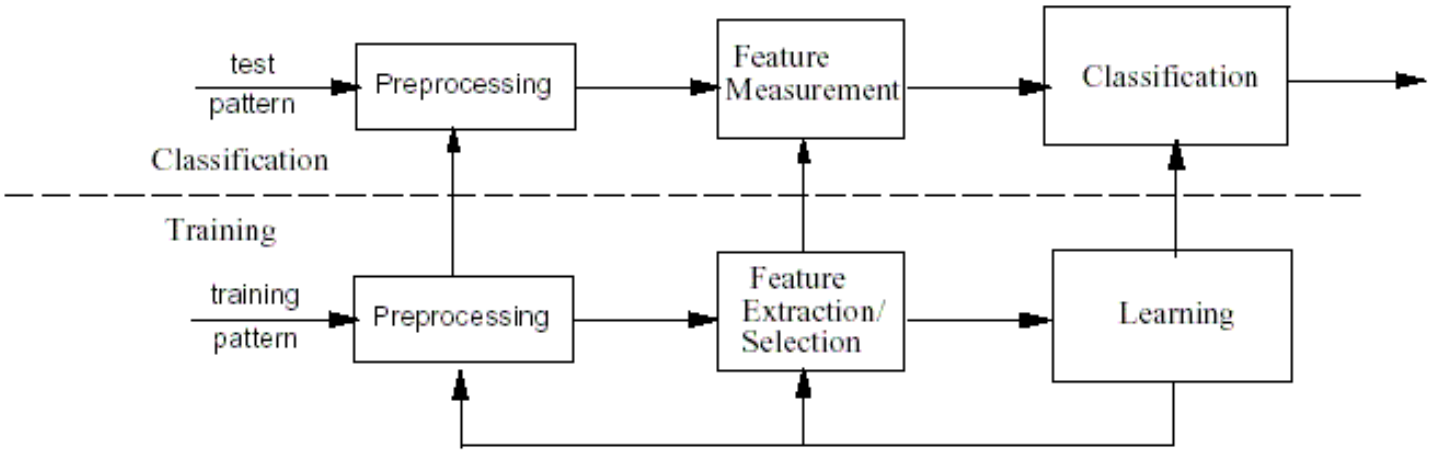
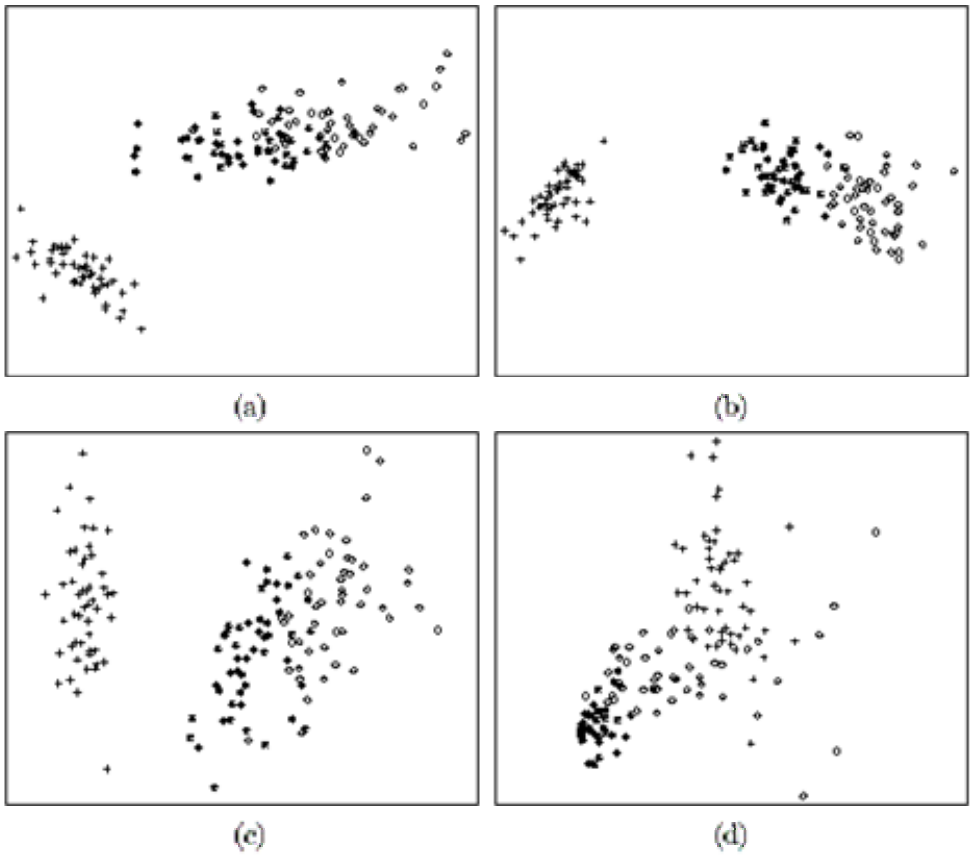


Figure 1: Model for statistical pattern recognition.

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

Feature Representation Extraction/Selection



(Jain et al 99)

PCA

Fischer Analysis

MDS

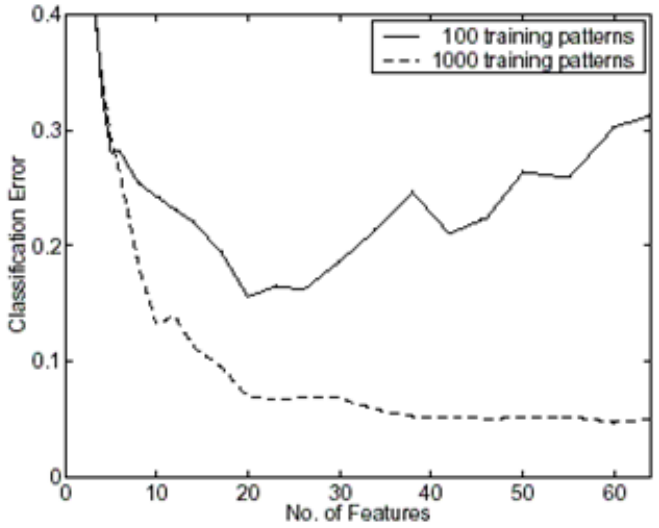
Kernel PCA

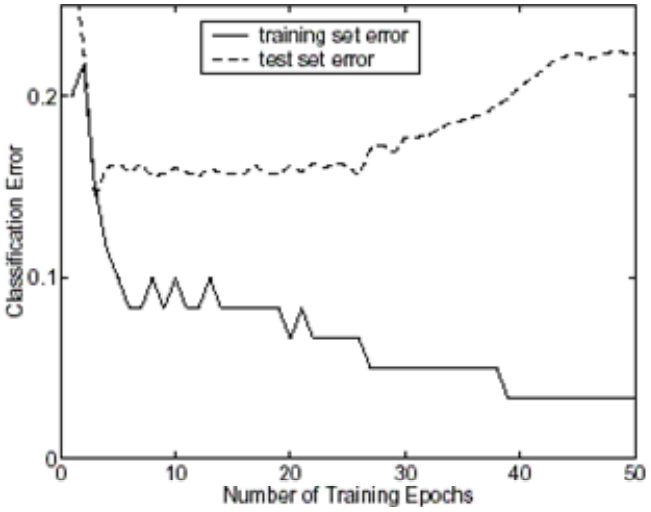
Issues to Consider

- There are no universally optimal classifiers!

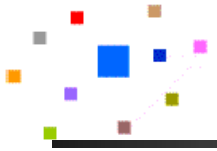
- Statistical structures of problems and models (dependence, features, scale, etc)
- Generation vs. discrimination
- Feature representation and selection
- Amount of training/test data
- Performance estimation and comparison
- Online vs. offline
- User supervision/feedback

Curse of Dimensionality and Overtraining





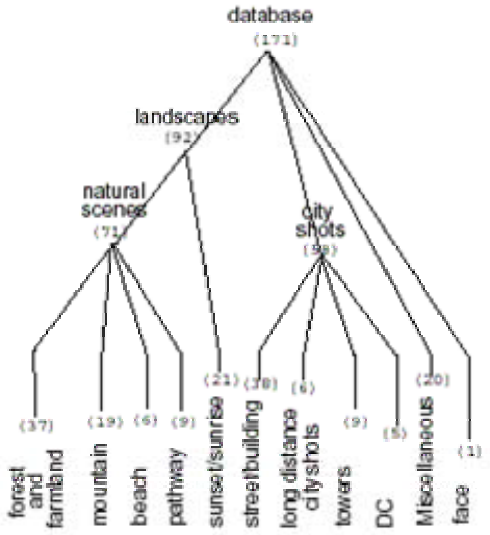
Rule of thumb -- # of training patterns per class / # of features > 10

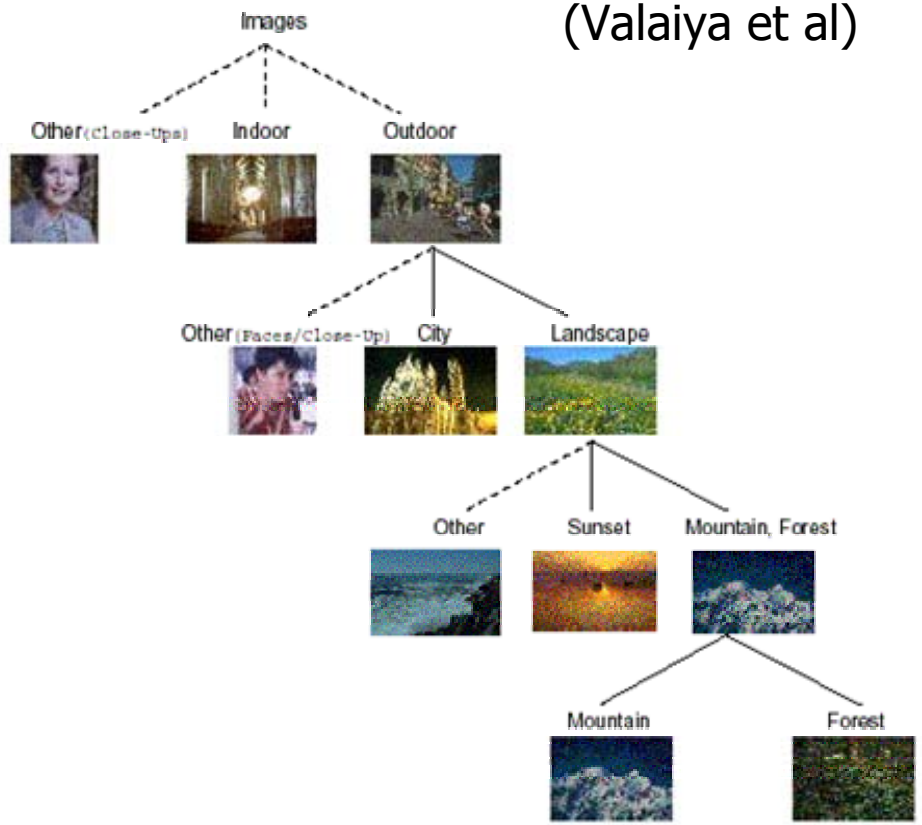


- A few examples from paper list

Bayesian Image Classification

(Valaiya et al)





dvmm DIGITAL VIDEO • MULTIMEDIA LAB

Bayesian Image Classification

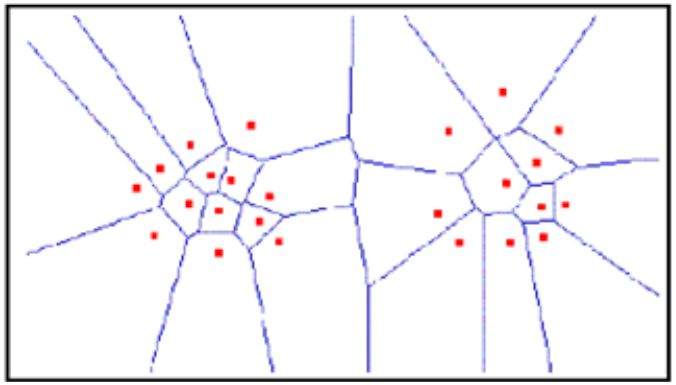
Feature independence

$$f_{\mathbf{X}}(\boldsymbol{x} \mid \omega) \equiv f_{\mathbf{Y}}(\boldsymbol{y} \mid \omega) = \prod_{i=1}^{M} f_{\mathbf{Y}^{(i)}}(\boldsymbol{y}^{(i)} \mid \omega).$$

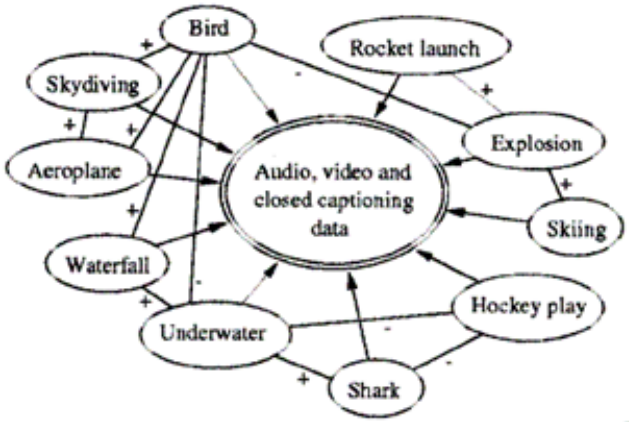
MAP Classification

$$\hat{\omega} = \delta(\boldsymbol{x}) = \arg\max_{\omega \in \Omega} \{p(\omega \mid \boldsymbol{y})\} = \arg\max_{\omega \in \Omega} \{f_{\mathbf{Y}}(\boldsymbol{y} \mid \omega)\, p(\omega)\}.$$

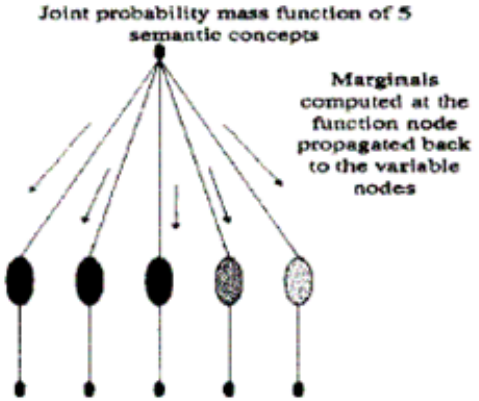
VQ as distribution estimator

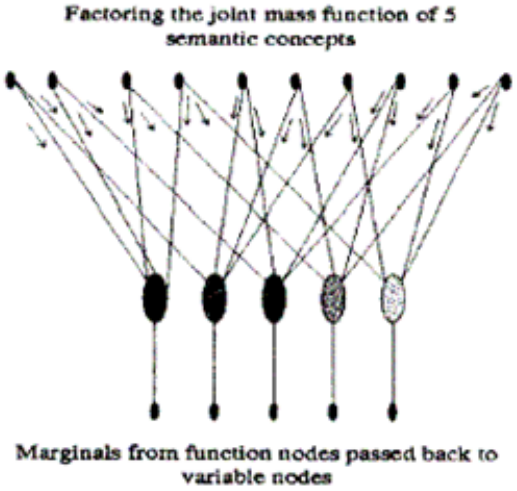


Concept (In)Dependence



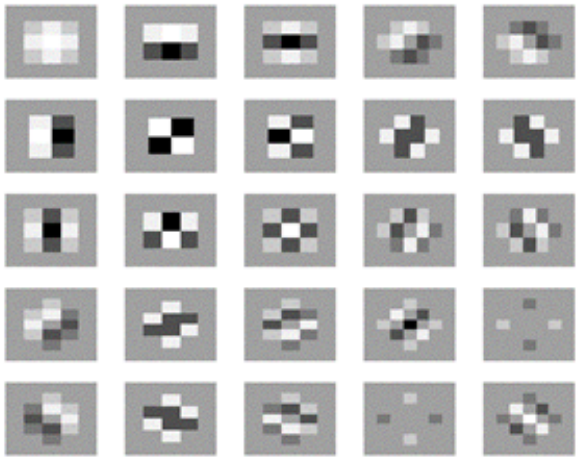
(Naphade et al)



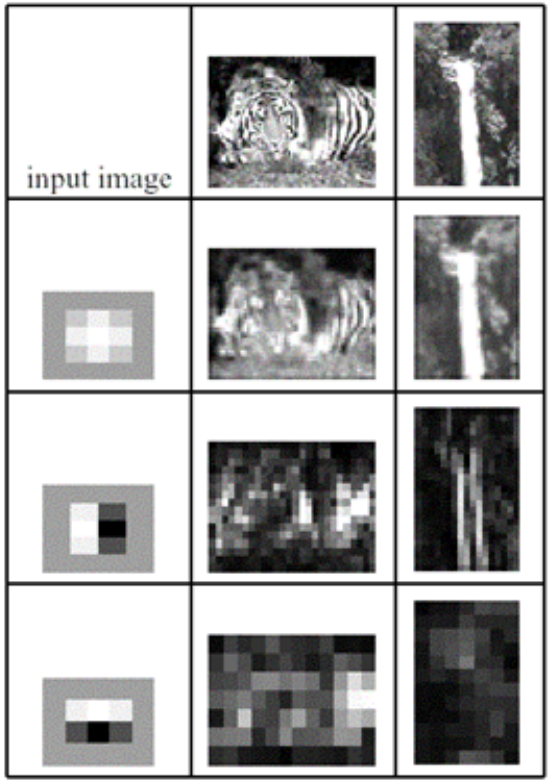


Boosting

(Tieu and Viola)



Extract > 45K selective efficient features by multi-scale filtering

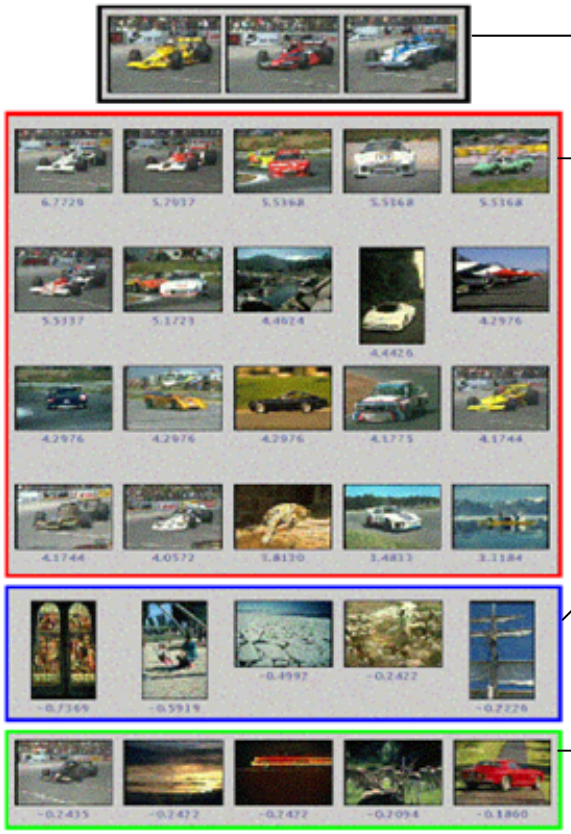


Classifier combination and sample re-weighting

$$w_{t+1,i} = w_{t,i}\beta_t^{1-e_i}$$

$$h(x) = \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2}\sum_{t=1}^{T} \alpha_t$$

Boosting retrieval interface



User selected examples

20 retrieval results

Real-time evaluation of
20 features over millions of images

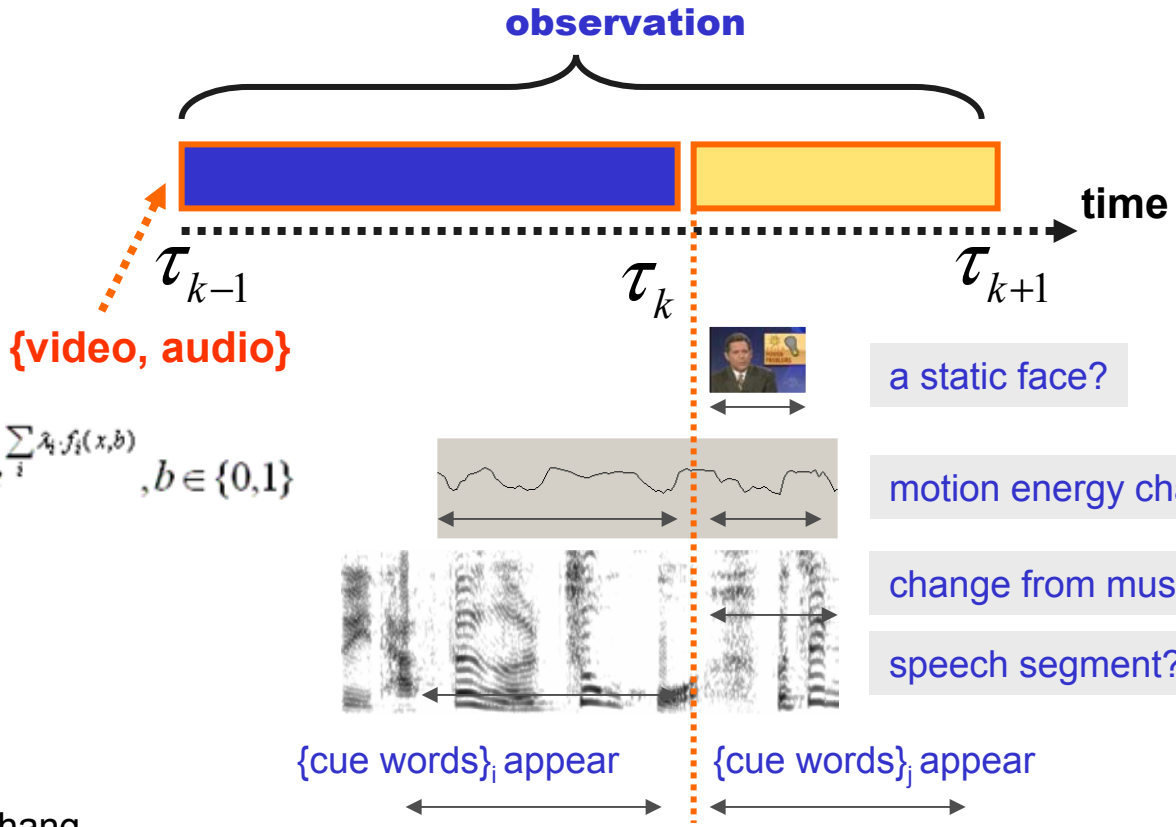
Negative images in the training set close to decision boundary

Images in the testing set close to the decision boundary

Maximum Entropy Fusing

- Objective: a boundary at time τ_k?
  - τ_k= { shot boundaries or significant pauses}

(Hsu and Chang)



$$q(b \mid x) = \frac{1}{Z_\lambda(x)} e^{\sum_i \lambda_i f_i(x,b)}, b \in \{0,1\}$$

dvmm DIGITAL VIDEO • MULTIMEDIA LAB

Object-Word Correspondence

(Duygulu et al)

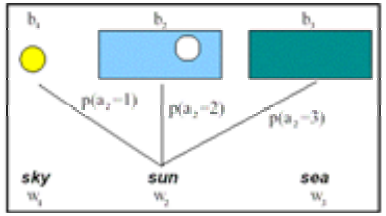


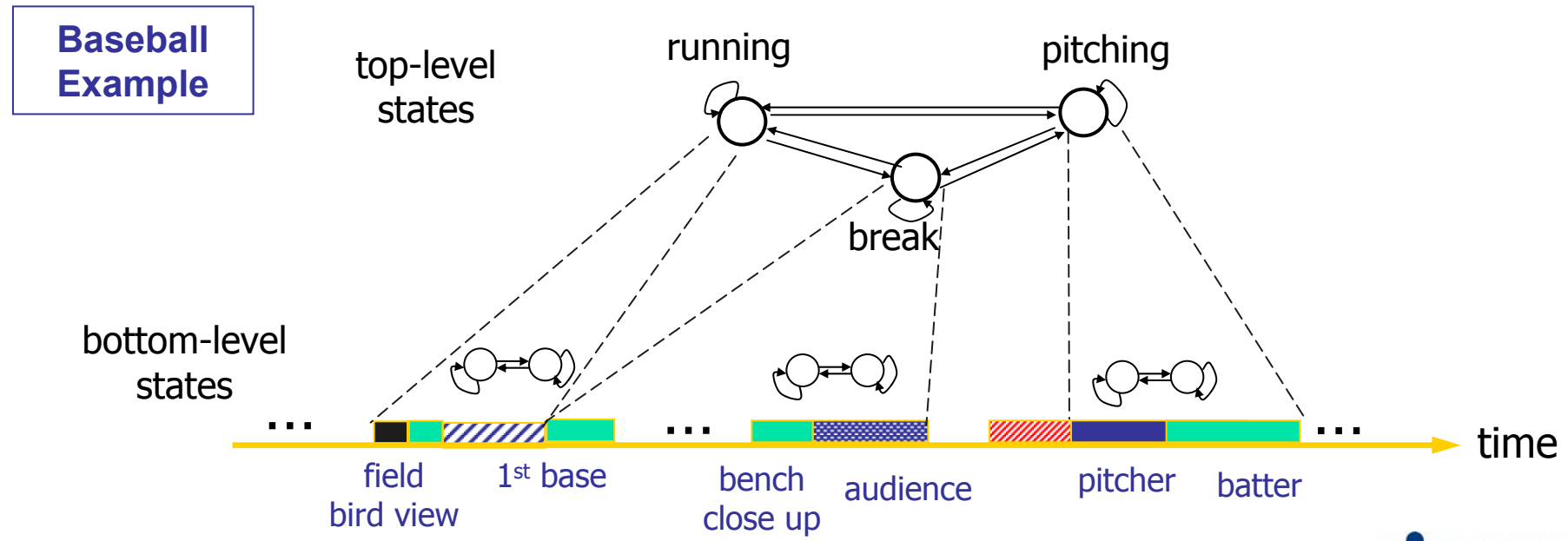
Fig. 3. *Example : Each word is predicted with some probability by each blob, meaning that we have a mixture model for each word. The association probabilities provide the correspondences (assignments) between each word and the various image segments. Assume that these assignments are known; then computing the mixture model is a matter of counting. Similarly, assume that the association probabilities are known; then the correspondences can be predicted. This means that EM is an appropriate estimation algorithm.*

Fig. 8. *Some examples of the labelling results. The words overlaid on the images are the words predicted with top probability for corresponding blob. We are very successful in predicting words like* **sky**, **tree** *and* **grass** *which have high recall. Sometimes, the words are correct but not in the right place like* **tree** *and* **buildings** *in the center image.*

Fig. 9. *Some test results which are not satisfactory. Words that are wrongly predicted are the ones with very low recall values. The problem mostly seen in the third image is since green blobs ooocur mostly with* **grass**, **plants** *or* **leaf** *rather than the under water plants.*

Unsupervised Video Structure Discovery: Hierarchical Hidden Markov Model (Xie et al)

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



Course Format

- Reading seminar
- 2 papers reviewed and demonstrated each week (class size will be limited)
- Each student assigned one paper
  → assignments determined 2-3 weeks in advance
- Everyone writes comments before and after class on personal web sites
- Term project at the end of course (12/10/03)
  -- target at conference paper submission

Paper review and demo

- Each paper allocated 60 mins total
- Discuss paper and plan demos with me and TA before class
- Prepare copies of slide handouts before class, or make them available online
- Computer demo of the reviewed method using toy data set

Paper Review and Demo (2)

- Review
  - Background review and examples
  - Problem addressed and main ideas
  - Insights about why it works
  - Limitation, generality, and repeatability
  - Alternatives and comparisons
- Demo
  - Software and data available and repeatable?
  - Reconstruct the method and try on toy data set? (from some publicly available generic toolkit)
  - Analysis of results (not just accuracy numbers, offer explanations and verifiable theories about observations)
  - Demo code archived on class site and shared with others

Required background

- Familiarity with
  - Image processing or computer vision
  - Statistical pattern recognition or machine learning
  - Computer programming (e.g., Matlab)
- Background assessment given in the first class
  - video representation, features, and statistical concepts

Grading and Credit

- 25% paper review,
  25% demo,
  25% class participation, and
  25% term project
- Auditing permitted only
  - for non-students
  - with active, continuous class participation

Class Resources

- How to read/present/write a research paper? (see links on web site)
- Software links on web site to HMM, Netlab, SVM, and Bayesian Network
- Image/video data and features from DVMM lab

Schedule

- Available on the web site
- Next 2 lectures (need volunteers)
  - Image classification (9/10, work with me and TA)
    - Bayesian Methods (Vailaya, Jain, and Zhang)
    - Factor Graph (Naphade and Huang)
  - Boosting (9/24)
    - Freund & Schapire, Tieu and Viola

Goals

- Everyone learns insights and experience in this emerging field
- Accumulate tools and reports
  →
  Construct a self-contained reading and experimentation learning set for statistical video indexing/analysis